

## Comments on conference papers

Reading **Ann Huthwaite's** posted paper, I was much encouraged by her mention of the proposal that collation be restored for nonphysical entities, and the special material designations (smd) called for in ISBD(ER) to appear after edition (256), appear rather in their traditional collation placement (300). Smd after general material designation (gmd) seems a more logical order to me.

Our customers already demand collation smds (e.g., website) for remote electronic resources, as they earlier demanded that gmd appear after the first title proper in all cases. It was good to have the rule change to agree with the practice our customers demanded in that case, and it would be good to have it happen in this one as well.

For a vendor such as ourselves, paying customers can't be wrong, rule or no rule. Fortunately rules do tend to evolve in the direction they wish, if slowly.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

John Byrum wrote:

>> *I am pleased to report that "The Library Catalogue in a Networked Environment" by Tom Delsey is now available at [lcweb.loc.gov/catdir/bibcontrol](http://lcweb.loc.gov/catdir/bibcontrol) for your review and comment.*<<

I've not yet read it in full, but I notice that one matter **Delsey** addresses is the virtual union catalogue.

As a regular user of four facilities that function as union catalogues (OCLC, the Australian service Kinetica, RLIN (through Kinetica), and the British CURL database COPAC) I am constantly falling foul of the apparently limitless amount of duplication of records.

I am very well aware of how these duplications arise and why they remain in the databases, even if there are attempts to resolve duplication either mechanically or by human intervention. In some cases records which are in fact duplicates appear to be different, until one consults them with the item itself in hand, when one can see how different cataloguing codes, and/or different local practices, fullness of data, and so on, have produced records which appear to be for different items.

I've attempted to use two virtual union catalogue service, the German service and the Bradford experimental catalogue (BOPAC). From this side of the world, BOPAC is hardly usable because the transmission delays are unbearable. This is a pity, because it attempts to apply algorithms to consolidate duplicate retrieved records into a single result. It too, on the few occasions when I've had useful results, fell foul of the duplication problem due to differences in the data in the various bibliographic records.

It seems to me that the problems of duplication cannot be resolved if there is no coherent attempt at standardization of retrospective files, and an effective means of getting such records into local databases.

The cataloguing code has a part to play here. While core-level or full records to AACR2 are less likely to engender duplication, it is still possible to encounter situations where a book (to take the simple case) is published jointly by publishers in England, the U.S., Canada and Australia (in a single joint edition) and the application of rule 1.4D5 of AACR2, which enjoins selective recording of publication data, results in records from each country which look different, and therefore don't match as duplicates.

In other words, apart from discrepancies due to past practices (including hasty and unrevised retrospective conversion of cards), the codes by which we catalogue still don't produce uniformity, which is surely the key to a successful virtual union catalogue.

I can't see this being solved any time soon.

Hal Cain, Joint Theological Library, Parkville, Victoria, Australia

\* \* \* \*

I have been reading the draft of **Ann Huthwaite's** conference paper, "AACR2 and its place in the digital world".

She refers to the opinion of the Australian Committee on Cataloguing that the proposed reorganisation of Part 1 of AACR2 according to ISBD areas of description could be complex and unwieldy and that it would still not address the multiple characteristics problem.

Having examined Appendix B of the "CC:DA Task Force on Rule 0.24:Final Report", in which an attempt is made to demonstrate what a reorganisation by ISBD areas might look like, I agree with the Australian Committee.

Despite the problems shown by Tom Delsey regarding the present arrangement by class of material, I believe that it should be retained, but that a definite order of precedence should be set for describing items with multiple characteristics. I do not, however, necessarily recommend the order suggested in Option 3, to deal with multiple characteristics, of the CC:DA Task Force report.

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

Commenting briefly, **Tom Delsey's** paper, The Library Catalogue in a Networked Environment, is very carefully written. It frames the big issues surrounding the contemporary internet catalogue. From cool recesses of the theoretical pond Delsey casts the major terms and concepts knowing readers and conference discussants will hungrily attempt to resolve somewhat the ongoing debate on the issues. His paper is well-positioned early in the conference agenda.

Andrea Morris Gruhl, Member, IFLA Section on Cataloguing

\* \* \* \*

I agree with Ms **Huthwaite** that we do not fully exploit the potential of the electronic version of AACR2. I believe it would be ideal for the cataloguer to be able to arrange the rules in the order which makes most sense to him/her. In order to do this successfully the cataloguer would need a thorough understanding of the principles and structure underlying AACR2; I would also like to see new graduates equipped with this understanding.

It would also be of great practical help to cataloguers if a publication such as "Notes in the catalog record" could also be linked to the electronic version of AACR2. This would enable the cataloguer to make quicker decisions on the phrasing of notes and hopefully speed up the cataloguing process.

I look forward to seeing how AACR2 will adapt to our ever-changing environment.

Susan Battison, Project Leader: SANB, National Library of South Africa, Pretoria Division

\* \* \* \*

I read **Tom Delsey's** paper and was interested in a number of points that he raised -- but particularly his discussion of strategically positioning the library catalog as one information source among many. Our investment in developing and extending this tool will likely have a higher payoff if we focus on its particular strengths, as compared with trying to make it not fully distinguishable from other search tools. This passage is particularly pertinent here, I think:

*>> Equally important from a strategic perspective is the fact that the library catalogue functions as a guide to a collection of resources professionally searched, selected and maintained for the purpose of supporting the research and information needs of a defined community of users. With the exponential growth that characterizes the Internet, the selectivity and pre-determination of relevance that are reflected implicitly in the library catalogue take on even greater value. <<*

I find this contrasts significantly with the attitude, now gradually fading but still prominent, that we have some mandate to attempt to "catalog the Web" -- that is, find -some- array of mechanisms that will allow us to somehow keep up with the "billions and billions" of Web pages being created. Does anyone actually say that we have to bring all Web resources under bibliographic control? Not really. But we do regularly read and hear statements like these:

- 1) That yes, we have cataloged X thousand resources, but what is that compared with the X billion pages now existing?
- 2) That are you aware there is a new domain name registered every three minutes?
- 3) That the growth in existing Web pages continues at a rate of X, and what are we going to do about that?

None of these statements tell us anything about the size of the set of -valuable- web resources, never mind what might constitute value in any given context. I increasingly feel that these statements have little, if any, inherent meaning relative to the work we have to do -- and that their propagation is almost obfuscating.

This isn't a hidden argument for exclusively applying "traditional cataloging" techniques (using Priscilla Caplan's phrase) to Web resources. There are certainly Web resources of differing degrees of value and permanence that may be brought under different types of control. It is a suggestion that we gradually let go of making global statements about the total size of the Web, as if that had some inherent relationship to the resources to which we need to give attention. (These statistics may be interesting in themselves, of course.)

David Miller, Levin Library, Curry College, Milton, MA

\* \* \* \*

I have been reading **Tom Delsey's** paper for the conference. I strongly agree that there is a need to retain and enhance those features of the catalogue, including quality control, adherence to standards, selectivity and the retrospective dimension, that have served over time to make it an effective tool for its users and that give it the potential to outperform other tools in the new environment.

The idea of creating a direct link between the catalogue record and the networked resource described is intriguing. It would be a useful feature, provided, of course, that sufficient terminals were retained for traditional quick-reference consultation.

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

**Lois Mai Chan** in her paper "Exploring LCSH, LCC, and DDC to Retrieve Networked Resources ..." for the upcoming conference at DLC sees traditional library resources in stark contrast to web resources in that the former are stable while the latter are volatile. This distinction serves her well in making the points she wishes to make, but I find the characteristics of some traditional library materials, loose-leaf services for example, not to be that different from what one finds in a website. Wasn't it just this week that an Autocatter was asking about an annual which has radically

changed title every few years? I see an overlap in characteristics, rather than a dichotomy.

Solutions we have already arrived at may be more applicable than some suppose. Web searching using search engines differs not that much from keyword searching of MARC records. One broadens the net, and catches items not wanted along with more than a controlled vocabulary search might have provided.

Chan, quite accurately I think, states that "Controlled vocabulary most likely will not replace keyword searching, but it can be used to supplement and complement keyword searching to enhance retrieval results." The move toward providing subject term and even class keys within websites is laudable. It seems to me we need to be more persuasive in presenting the advantages of the established tools: LCSH, DDC, and LCC. We need also to make greater use of class numbers for subject retrieval of all materials.

The reasons cited for the use of LCSH as opposed to a new vocabulary are persuasive, in addition to the fact the its use allows the simultaneous search of traditional and newer information sources.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

I have been reading **Michael Kaplan's** conference paper, "Exploring partnerships".

He says, "*The bottom line in all this is the need for widespread cooperation and for standards*".

I favour cooperation with producers and vendors of electronic resources and there are certainly areas in which this could usefully be developed, as he points out.

I also favour adherence to standards. It is, however, probably unrealistic to suppose that creators will provide metadata for their work, since they will see such provision as the province of the cataloguer. It is also unrealistic to think that cataloguers will be willing to abandon tried and tested standards such as AACR and MARC in favour of something like the Dublin Core, which they will see as a crude attempt to re-invent the wheel.

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

>> *It is, however, probably unrealistic to suppose that creators will provide metadata for their work, ...* <<

I remember when this was said of CIP (Cataloguing in Publication) in the early days. But publishers soon learned that including CIP helped their sales. If our cataloguing agencies could help provide the metadata, I suspect they would include it. In its early stages, provision of CIP in both Brazil and Canada was decentralized. If the national cataloguing agencies can't take on this task, such a pattern might work. A library willing to take on 1/5th of the websites of a region, might be reducing their cataloguing of electronic resources by 4/5ths through sharing.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

I have been reading **Lois Mai Chan's** conference paper, "Exploiting LCSH, LCC, and DDC to retrieve networked resources".

I found the section "Recent research on subject access systems" of particular interest. She here discusses, for example, mapping terms and data from different sources and integrating different subject access tools.

It is pleasing that she finds "How to combine the salient features of LCSH and classification schemes such as LCC and DDC to improve retrieval of networked resources remains a fertile field for research and exploration".

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

I don't know if it was happy coincidence or an insightful choice, but having **Ann Huthwaite's** paper the first one posted to the Bibcontrol website was very appropriate. I enjoyed learning of some of the background which in part prompted this conference, as well as adding to my slowly-growing understanding of the process of change to AACR. Although this change may be slow (even "glacial" at times), I for one am happy that such careful consideration is given, and that input from so many different constituencies is sought.

But the part of the paper I'd like to comment on is the section in which Huthwaite outlines three possible scenarios of what AACR might look like ten years down the road. We're seeing changes being made to AACR today that are driven by changes in the ways in which people access information. I doubt that libraries (or whatever you'd like to call them if you'd rather avoid the "L" word) will have any influence at all over these technological changes. They will happen, and they will either catch on (i.e., somehow be profitable or at least \*seem\* profitable) or they won't catch on (in which case, the "glacial" process of AACR change will truly be a blessing).

But we can and must change AACR to adapt to the changes in the ways in which people access information. The Internet has had a profound impact on so many different aspects of the world today, but it won't be the last innovation capable of this kind of impact. We must remain as flexible in our thinking as AACR must remain flexible in its application.

It wasn't until I had the great pleasure of attending a workshop on cataloging realia (conducted by Nancy Olson at the '98 OLAC Conference) that I realized that one really \*could\* catalog ANYTHING using AACR. It might not be pretty at times, but it works. And I think the challenge now and in the future is to take a very good product and make it better.

Gene Kinnaly, Senior Cataloger, Library of Congress

\* \* \* \*

I have been reading **Priscilla Caplan's** conference paper, "International metadata initiatives".

She says "In all these cases, but perhaps most intriguingly with DCMES, what we have been seeing, if we've been paying attention, is the re-invention of cataloging". "On the negative side, we can see those communities slowly, painfully and with many false starts rediscover principles that librarians have understood all along. On the positive side, it will be constructive to learn from what they retain and what they throw away, because they are directly confronting what is necessary and feasible to meet the needs of users in the Internet environment".

She draws the important conclusion "that the key question as we enter the new millennium is not bibliographic control of Web resources, but rather bibliographic control of both digital and non-digital resources in the Web environment".

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

I have been reading the draft of **Barbara Tillet's** conference paper, "Authority control on the Web".

It might have been thought that the ultimate objective of cataloguers in the sphere of headings would have been to establish a single form of heading for each name and to have that form accepted throughout the world. One is led by

this paper, however, to the shocking realisation that this is not a good idea and that it would be much better to link the diverse forms which may exist and allow the individual user to choose the form most suitable to them.

The notion of interoperability has occurred before in the conference papers so far made available. For example, Lois Mai Chan speaks of mapping subject terms and data from different sources. William E. Moen, discussing Z39.50, says "There is little doubt interoperability is a key issue in the networked environment". I think it could be an important feature of the conference.

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

I admire **Priscilla Caplan's** courage in tackling the alphabet soup of metadata initiatives, and I appreciate the clarity she brought to the descriptions. I found the information to be very helpful.

Since we are a CORC participant, I paid particular attention to the section on the Dublin Core. It seems to me that the greatest strength of DCMES is the diverse communities, outside the traditional library, which it can serve. And in serving these communities, the number of elements within DCMES was limited and kept fairly general. Now we have "qualifiers" being added to bring about more specificity, and I wonder .... if enough is added to DCMES, will we end up with ... MARC?

I found the section on metadata schemes created by the publishing community to be very interesting. Is there a possible connection here with LC's Electronic Cataloging In Publication (ECIP) Program?

Gene Kinnaly, Senior Cataloger, Library of Congress

\* \* \* \*

I agree with Gene. The **Caplan** paper provided clear explanations of metadata schemes and their evolution, aiding in comparison of resource discovery options.

I was curious about the possible impact that exchange and use of publisher's commodity metadata could have on hierarchical descriptive practices, especially for materials that are unpublished. Caplan suggests that we should "begin thinking about basic bibliographic metadata as a commodity, produced and exchanged by a number of communities in order to serve a number of purposes." What impact would thinking of our metadata as a commodity have on content rules for hierarchical description, if any?

Marcy Flynn, Silver Image Management, Scottsdale, AZ

\* \* \* \*

I have been reading **Karen Calhoun's** conference paper, "Redesign of library workflows".

Michael Gorman once said "The rift between catalogers and reference librarians (technical and public services) has been one of the more damaging aspects of modern librarianship". I agree with this and with the writer of the present paper when she says "*Our workflow sought to put back together again a group that has been artificially separated by organization*".

Cooperation of the kind envisaged, however, which has been seen to work well in the context of a research project, would require particularly close attention, if pursued on a regular basis, to ensure that full treatment was given, if required, to items at first entered with minimal information, and to ensure that sites which had gone out of existence were removed from the record.

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

The new paper for the upcoming conference on cataloguing internet resources, "Metadata for Web Resources: How Metadata Works of the Web" <<http://lcweb.loc.gov/catdir/bibcontrol/whatsnew.html>> by **Martin Dillon** is well worth reading for an overview of efforts to describe Web resources. He comes down strongly on the need of description as opposed to just relying of searching.

His feeling of in importance of digital information as opposed to print is a greater than mine. Just as photography did not replace painting, nor television replace motion pictures, I don't see digital information replacing print. Rather I see it as one more form of information to be incorporated.

He calls strongly for greater cooperative effort in Web description. It is unfortunate that lack of funding prevents our national cataloguing agencies from being more active in this field. With the advent of DLC printed unit cards in he early 1900s, we in North America all benefitted from not having to repeat that cataloguing effort for each book acquired. That has not happened for electronic resources. But it did not happen for print until some centuries had passed.

He points to the Dublin Core as the best we have as a standard, and suggests "that a MARC version of the Dublin Core be developed with appropriate instructions and examples" without delay.

Dillon embeds URLs related to the subjects he discusses in the body of his text, as opposed to references at the end. This allows reading the paper on the Web to be an interactive, almost nonlinear, experience as one decides which references to follow up. The experience gives credence to his assertions concerning the qualitative difference of electronic information.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

I agree with Mac that **Martin Dillon's** paper is well worth reading. I found his comparative summary of several current metadata initiatives, and his more lengthy assessment of the current state of the Dublin Core, quite valuable. But I was brought up quite short by the following, near the end of the paper:

*>> How should libraries provide access to document-like knowledge resources on the Web? If the library community decides that it is necessary to establish a form of bibliographic control for such objects, three paths are open:*

- 1. Use or adapt MARC/AACR2*
- 2. Start fresh creating a library metadata system with the same aims as the Dublin Core*
- 3. Use or adapt the Dublin Core*

*I will discuss each of these briefly.*

*Use or Adapt MARC/AACR2*

*There may have been a time when this was a useful direction to take but it is long past. The result of such an exercise would have many of the attractive attributes of the Dublin Core, particularly its simplicity and flexibility. <<*

Briefly indeed! That's it for point 1. Now, I'm aware that this general opinion is shared, more or less, by a number of respected colleagues -- and that a possibly comparable number of equally respected colleagues disagree, and have been working quite actively on bringing these two major standards into our new environment. But a thumbs up/down verdict isn't what I'm after here.

What strikes me is that, first, this breathtakingly brief dismissal stands in stark contrast to the extent of discussion in evidence throughout the rest of the paper. It's almost as if Dr. Dillon can't be bothered to elaborate: is that really true?

But second, not two paragraphs later there's a call for a "MARC version of the Dublin Core"! Now I really don't understand -- how is this not an adaptation of MARC? Additionally, he repeatedly insists not only on the importance of authority control in the Web environment, but on distinguishing personal names from geographic names and corporate bodies. What else, at the heart of it, is Part II of AACR2 about?

What I'm asking for is an elaboration of his point 1 above -- I also need to understand the relation of the second sentence to the first. It seems as though the "exercise" of adapting AACR2 and MARC would provide benefits parallel to further Dublin Core development, and yet simultaneously it's a direction not to be pursued. Again, a fairly stark contradiction.

I appreciate any enlightenment that Dr. Dillon would be willing to provide. It's more than possible that I'm just being dense.

David Miller, Levin Library, Curry College, Milton, MA

\* \* \* \*

#### CORE DESCRIPTION: BETTER SURROGATES VIA TECHNOLOGY?:

These are comments inspired primarily by **Michael Kaplan's** paper "Exploring partnerships."

*>> The future record, whether you prefer to call it a catalog record or a metadata record, will not be one-dimensional or static. Rather, it will be multi-faceted and dynamic. It will be composed on-the-fly of a variety of different "metadata-lets": the traditional bibliographic description at its core, but with a number of concentric circles associated with it and including such information as citations, reviews, dust jacket illustrations, author information, links for delivery or ordering, etc. ... What do I mean by a dynamic record? It is a record that has a core consisting of a traditional bibliographic description, but it is regularly enhanced and/or refreshed by a series of elements that we have not routinely associated with the record. ... Engines and transmissions are essential to automobiles just as a core bibliographic description is essential to the bibliographic record. <<*

OK, but is it possible that the "enhancements" might be able to replace elements of the traditional core description?

On a Webpac, wouldn't a scanned image of the chief source be preferable to bibliographic transcription of AACR2 areas 1,2, 4, 6, and parts of 7? Wouldn't the user benefit from not having to "translate" catalogese? Wouldn't catalogers benefit from not having to argue about or decide on whether to apply standard abbreviations, capitalization rules for different languages, when to use diacritics, how much to transcribe, etc.? (Gotta have that perfect record, the Perfect being the enemy of Good Enough, per MK. But maybe a good sharp image of the chief source(s) is really closer to Platonic Perfection than the best textual description.)

In the core descriptive record, why couldn't the cataloger simply enter the address of the chief source image file(s), and be done with those areas vis-a-vis the descriptive part of the job? Since the image would function as a surrogate of the bibliographic object (in those situations where there is such a distinction), MARC 4xx could be eliminated & MARC 245-260 would no longer need to function as containers for AACR2 areas, i.e. as textual surrogates for the bib. object. To function as a textual surrogate for an object understandably requires highly standardized conventions, detailed cataloging rules. But if the surrogate is already present as an image, textual description is to a large extent superfluous, and MARC 245-260 could function instead as metadata labels for indexable fields that do not necessarily need to be displayed in the Webpac & do not require the level of rigor applied in textual descriptive transcription.

The cataloger might then be able to spend more time on the standard access points and the labor of getting them under authority control. We have already started to accept "non-standard" intrusions into the integrity of the bibliographic



description by accepting vendor supplied TOCs, which generally follow the contours of the published table of contents layout rather than AACR2 or the LCRIs. Why couldn't vendors also supply a database of scanned t.p. images and catalogers accept these as a reasonable, good enough (or rather, better) substitute for the areas cited above? Why couldn't this be a CIP requirement?

Objections?

1. Let's assume the image of the chief source won't be indexable. (More technical folks are welcome to offer a workaround; would e-book title pages be indexable?) Indexing the title: Retain 245 subfields a & b. But re-define 245 as primarily an indexing field rather than a field for transcribing the AACR2 title area; the true transcription is the image. Therefore, don't get exercised about diacritics, capitalization, including the initial article. No need to translate the chief source into catalogese; let publishers be publishers. Put it all in capital letters or lower case, the computer doesn't care. The field doesn't need to be seen by the OPAC user, anymore than field 653. Incidentally, why do alternate title access fields need to be displayed in Webpacs? Alternate forms of established names are generally not displayed with the established name.

2. What were luxuries have become necessities: Have other descriptive elements become access points with online systems?

-If you really want to index the publisher area, how about a translation table for the ISBN? (Keyword search on place/publisher where system checks against ISBN table to retrieve appropriate bibliographic records.) In the absence of ISBN, enter publication area as an indexed field rather than as a transcription area (transcribe what is there with no diacritics, abbreviations, worries about capitalization, etc. Unless someone wants to change the rules and interpret publication area as an access point, placing it under authority control!).

-No 4xx (scan the series t.p. <see 4>, just 830.

-I can't see any way to avoid textual physical description, unless e-books become the norm (in that case physical description would be superfluous).

3. Resources: Might be more economical to have a master database of title page images, to which the catalog records would be linked. Cooperative efforts as individual libraries contribute their scans to the master database? Does an image work for all chief sources, all types of material? Title pages you've seen on amazon; CD ROM title frames relatively easy to capture; video might require multiple frames; graphics & three dimensional would be better served with images than with text; maps, economically problematic.

4. Add ons: "Include an image of the back cover, jacket, etc if it aids in identification." Part one of AACR2 more a set of rules about what to scan?

Acknowledgement: Ake Koel, formerly head of Yale's Sterling Library Technical Services, often brought up the scanned title page idea (it was title pages in those days) when we would occasionally chat about the Rules. I just thought it might be interesting to re-introduce the concept now that Webpacs are coming into their own and rule revision wants to account for such changes.

VENDOR SUPPORT:

*>> Librarians often pay lip service to cooperation and to trust. Rather, they maintain lists of acceptable and non-acceptable libraries and are constantly re-inventing the bibliographic record to meet their own, internal and internalized standards. ... We cannot afford such arbitrary distinctions. The days of golden records are long gone. ... 'What good is a bibliographic record if it is not there when we need it and at a price we can afford?' <<*

I'd be the first to admit that revising cataloging copy & the attitude of suspicion can be neurotic, but there is such a thing as Gresham's Law, and I do think we still need to distinguish between Good Enough & Unacceptable. One way for vendors to help is to provide better training documentation. To paraphrase: "What good is an integrated electronic library management system if it doesn't have decent training documentation when we implement it?" In addition to providing

on time training documentation, that documentation should be customizable for local practices (if we aren't going to McDonaldize our libraries).

Further, by training documentation I don't simply mean how to operate your system, but training in application of bibliographic standards that would support staff in the creation of acceptable, good enough cataloging copy. One reason behind the attitude of suspicion I believe <I initially wrote "I suspect" but then I'm a cataloger> is that the number of thoroughly trained catalogers contributing copy to the utilities is diminishing. Vicious circle: fewer catalogers also means less time for training; fewer catalogers means there may be only one designated cataloger at the processing unit who may lack a human mentor. Where are the electronic "cataloging wizards"? This might be a useful cooperative venture among the utilities, LC, & vendors: a standard, continuously updated "How to catalog" software program that can accommodate different cataloging levels, with vendor adjustments for the idiosyncrasies of their particular system, and local customization for local practices. Maybe library technical services could help by publishing what they consider "unacceptable" in copy (and what might be more interesting, what might be deviant but acceptable). Such checklists would need to be realistic also about what lengths the library is willing to go to eliminate the unacceptable. If enough tech services posted their quality standards maybe some of the subcommunities (big research libraries, not so big, big public libraries, one-person libraries, etc.) could come to a more general consensus?

PS on exploring partnerships in authority control (B. Tillett's paper):

I realize the paper is not yet complete, but it seems to me that most of the scenarios apply to the standard library catalog (i.e. one which is not extended to the Internet). The scenarios would be most helpful for Z39.50 searching of standard library bibliographic records across national utilities and/or national databases. This strikes me as a not inexpensive project that would primarily benefit multilingual academic researchers, a small minority in comparison to users of the Internet. If some kind of authority control could be applied to the Internet, the expense would be easier to justify, but how this would work was not brought out for me, at least so far.

Steven Arakawa, Catalog Dept., Sterling Memorial Library, Yale University, New Haven, CT

\* \* \* \*

>> *OK, but is it possible that the "enhancements" might be able to replace elements of the traditional core description?*  
<<

I certainly hope not! It is hard to imagine how we would print a set of catalogue cards or a printed book catalogue from such records. Too much discussion tends to assume that card and printed catalogues are totally a thing of the past.

To enrich records with links to images of the title page and table of contents, and to enrich them with access points in the original non-roman scripts of the text, are wonderful ideas. But the basic bibliographic description as launched by Panizzi needs to remain.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

>> *To enrich records with links to images of the title page and table of contents, and to enrich them with access points in the original non-roman scripts of the text, are wonderful ideas. But the basic bibliographic description as launched by Panizzi needs to remain.* <<

Indeed it does. Or how would we know what the work was? The paper copies aren't going to go away, even if libraries turn into sort of museums containing real books as sort of artifacts (which I hope they won't - in my lifetime anyway!)

Helen Buhler, Classification Coordinator, The University of Kent at Canterbury, Canterbury, Kent

\* \* \* \*

I have been reading **Martin Dillon's** conference paper, "Metadata for Web resources".

He believes that paper products will be replaced by digital ones, that AACR is by its very nature incapable of dealing adequately with the situation, that efforts to adapt AACR for use with Web resources should not be undertaken and that the Dublin Core should be adopted.

There is a different view, which is well presented in the section headed "Relationship between AACR2 and metadata schemes" in the paper written for the conference by Ann Huthwaite. I commend it.

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

**Sarah E. Thomas's** "The Catalog as Portal to the Internet" just posted on the website for the upcoming conference begins with a rapid survey of the history of catalogues and cataloguing which would make excellent assigned reading for a beginning cataloguing course.

She makes a clear distinction between the catalogue (which lists a selected and controlled material) and Internet Portals or Gateways (which access a far larger body of material of very mixed character).

*>> Instead of striving from comprehensiveness, the goal of the catalog as portal must be to increase the ability of a community of users to meet their information needs by doing as much 'one-stop shopping' as possible.<<*

*>> Michael Gorman has suggested a tiered approach to the description of publications that takes into account the quality of the material being described with a progression from AACR2 through Dublin Core to keyword search indices ... This is a sensible counsel, and provides a path from the present to the future. <<*

She favours reducing the amount of time devoted to book cataloguing (more acceptance of found records) in order to allow time to increase the scope and coverage of material, as well as cooperating with other libraries and partnering with developers of portals and search engines to share expertise.

Having her use "her" as an inclusive pronoun makes me realize how the more common "his" must grate on some others. Her dropping the "u" from the full title of AACR is, I suppose, the fault of her spell checker.

Although she would probably not agree, this paper leaves me with the thought that perhaps access to specialized portals and gateways in the catalogue might be better than individual websites (apart from websites licensed by the library) being catalogued there, just as we point to periodical indexes but do not catalogue (usually) the individual journal articles. She does not feel we should allow the distance to develop between catalogues and Internet portals which earlier developed between catalogues and journal indexing.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

Considering the quality of the derived records we are finding, I have my reservations about **Thomas'** suggestion that we save time in cataloguing by accepting records as found. Even a recent DLC record had a 710 as a 110, and sometimes they don't catch changes in 245 between CIP and publication. Another recently had a poetry collection classed as a novel in DDC. But I decided not to do a riff on that in my first comments since I agreed with so much else she had to say. There *\*are\** ways of speeding cataloguing of print materials to make way for the new media, and it needs to be done, but I'm not sure giving up editing catalogue copy is the way to do it.

The more I consider her excellent characterizations of portals and gateways (the best I've seen), the more I am thinking that the OPAC as bibliography of bibliographies of remote electronic resources, as opposed to listing each of them, might be a good solution to cataloguing the Web.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

I agree with the person who suggested we spend less time on book cataloging (or cataloguing) as long as she meant that we should not be changing things that are already correct, just to suit some local whim or system, unless absolutely necessary. But I agree with Mac that we can't just accept anyone's records without checking them. I recently had a children's book with a subtitle "linen and lace" which was about fashion in the early part of the century and had been given subject headings for both linen and lace which were totally unsupported by the content of the book. And then there was the handbook for the movie "Dinosaur" which had a heading of Dinosaurs--Handbooks, manuals, etc. but no heading for the movie! At the time, I checked the online catalogs of some other libraries to see if they had caught either of these, and they had not.

Jane Myers, Westlake Porter Public Library, Westlake, Ohio

\* \* \* \*

I have been reading **Sarah Thomas's** conference paper, "The catalog as portal to the internet".

She says "*We should decisively reduce the amount of time we devote to the cataloging of books in order to reallocate the time of our bibliographic control experts to provide access to other resources, especially internet resources, but also unique primary resources and other analog format materials*".

All this represents an inspiring vision. My only objection is to one item in the list of recommendations for reducing the time spent cataloguing books. If "accepting copy with little or no modification from other cataloging agencies, including vendors" means not checking and, if necessary, amending entries, I know from personal experience that this is not a good idea and would act against the aim of providing a quality resource.

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

On Sun, 17 Sep 2000, Philip Davis wrote in part, quoting **Sarah Thomas's** paper, that one of her recommendations is that:

>> *We should decisively reduce the amount of time we devote to the cataloging of books in order to reallocate the time of our bibliographic control experts to provide access to other resources ...* <<

Which seems practically based in a world of shrinking budgets (in itself a contradictory assumption when looking at other events in the information world). However, that recommendation seems in contradiction to the following one:

>> *The catalog can serve as a portal to the internet if the catalog is re-interpreted to be an information service which registers in a systematic arrangement those publications and documents of interest to a particular community, ...* <<

It's the process of registering in a systematic arrangement that concerns me. By decisively reducing our time spent on one resource (a still valuable one) to spend more time on another, I believe that much of our arrangement processes will be scaled back -- subject access (not only assignment of headings, but their creation as well), name authority work, descriptive work, etc., are important to fulfilling our objectives, regardless of format.

Perhaps a call for ramping up what we do, across the board, is at least something we should consider. A burgeoning process of production of materials that we might wish to catalog (10,000,000 items per week just on the Web) in itself calls for an increase in the number of people involved in the process of information management. New programming tools will help, yes, but I think they will not be enough.

Also, if our experience with vendors is analogous, programming tools frequently go only so far. We've seen a perfectly reasonable piece of our arrangement process (the MARC code indicator value for secondary entry) made obsolete not because of difficulty in its assignment or its lack of value in the organization of materials, but because of a lack of implementation by vendors. And, our own acquiescence.

One last bit on this. Even though the theme of the conference regards networked resources, to make procedural changes and philosophical shifts based on those resources can leave out many of our cataloging constituents for whom networked resources remain an esoteric lot.

Philip Davis also wrote:

*>> If "accepting copy with little or no modification from other cataloging agencies, including vendors" means not checking and, if necessary, amending entries, I know from personal experience that this is not a good idea ... <<*

To which my experience, too, leads me to agree.

Daniel CannCasciato, Head of Cataloging, Central Washington University Library  
\* \* \* \*

As someone who has worked on three of the five projects reviewed in Ms. **Greenberg's** paper, I should first declare an interest. :-) Anyway, here are some comments (with apologies for the length) on this interesting paper:

On Mon, 2 Oct 2000, J. McRee Elrod wrote:

*>> Nordic Metadata, unidentified and the url takes one to ROADS listed below. <<*

Information on the Nordic Metadata projects can be found at:

<http://www.lib.helsinki.fi/meta/>

The link to the Nordic Metadata Dublin Core Metadata Template also doesn't work. Try:

<http://www.lub.lu.se/cgi-bin/nmdc.pl>

*>> ROADS, resource organization and discovery in subject-based services, which according to the linked pages is a discontinued set of software tools to enable the set up and maintenance of Web based subject gateways. <<*

The ROADS software tools are NOT discontinued. Although the ROADS project itself has ended, the ROADS software tools are still being developed on open-source principles. The software can still be downloaded, and some support is available from the initial developers and from others who have implemented parts of the toolkit. The software continues to be used by a wide range of Web based information services:

<http://www.roads.lut.ac.uk/>

*>> CORC seems to emerge as the project which best meets the stated criteria, and has gone beyond experiment to a fully operational project. <<*

On the criteria used in Ms. Greenberg's paper, this may indeed be the case. However, if one looks beyond the initial timescales of the other four projects, it becomes clear that they too have contributed quite a lot to the development of operational services (and other things).

So, for example, the ROADS software toolkit was developed to support the development of Internet information gateway services that were funded by the UK Electronic Libraries Programme (eLib). The software, therefore, continues to be used by services like the Social Science Information Gateway (SOSIG) and the OMNI health and medicine gateway. The ROADS software has also been adopted by a wide range of non-eLib services like the Finnish Virtual Library (FVL).

SOSIG: <http://www.sosig.ac.uk/>

OMNI: <http://omni.ac.uk/>

Finnish Virtual Library: <http://www.jyu.fi/library/virtuaalikirjasto/>

In addition to the Nordic Metadata project's Metadata Template creator, there is UKOLN's DC-dot metadata editor. Software tools developed as part of DESIRE (e.g. Combine) have also been used in the development of metadata-aware Web index services.

DC-dot: <http://www.ukoln.ac.uk/metadata/dcdot/>

Combine: <http://www.lub.lu.se/combine/>

Projects like ROADS and DESIRE were not just concerned with the development of software for standalone gateway services. Rather, the project partners wanted to develop tools that would promote interoperability and allow the easy cross-searching of one or more distributed gateways. Services where this type of approach is being implemented are FVL and the UK's Resource Discovery Network (RDN).

RDN: <http://www.rdn.ac.uk/>

The four projects have also led to further research and development work. Ms. Greenberg has already referred to the Nordic Metadata II project. The DESIRE project itself had a second phase (finishing in June this year). Also, a new three-year project called Renardus started in January 2000. Renardus is funded by the European Commission (as part of the Information Societies Technology (IST) Programme) to establish an academic subject gateway service that will provide users with integrated access, through a single interface, to a range of distributed information gateways and other services. Another project, the IMesh Toolkit (funded under the NSF/JISC International Digital Libraries Initiative), is attempting to develop a configurable, reusable and extensible toolkit for subject gateway service providers. Projects like these, therefore, are building upon the tools (and architectural concepts) developed by earlier projects, including those reviewed in Ms. Greenberg's paper.

Renardus: <http://www.renardus.org/>

IMesh Toolkit: <http://www.imesh.org/toolkit/>

*>> Ms. Greenberg calls for the participation of the bibliographic control community in ongoing > research in this area, including partnerships with commercial enterprises, increased efforts to make such Web control projects known, as well as increased interoperability and transition to fully operational projects. She proposes cooperation among us in sharing research, the development of an official list of considerations for improvements, and the development of a master Request for Proposal for Web resource access experimentation.<<*

There are some good ideas here. However, it is important to realise that there was quite good communication between

all the projects reviewed in Ms. Greenberg's paper. All four European-based projects were consortia; several partners featuring in more than one project, most notably the Institute for Learning and Research Technology (ILRT) at the University of Bristol, NetLab (the development department of Lund University Library), the National Library of the Netherlands and UKOLN. In addition, most of the project partners have had some involvement in the development of the Dublin Core, which might provide an existing forum for sharing research outcomes ... as could a project like SCHEMAS, a forum for metadata implementers:

SCHEMAS: <http://www.schemas-forum.org/>

In her paper, Jane Greenberg wrote:

*>>The various experiments in this area could have an even greater impact on Web resource access if they communicated not only on a goal level, but also on an operational level. In other words, a framework is needed that will improve interoperability and permit these experiments to talk to each other more than is currently practiced. <<*

The importance of this was realised by most of the projects themselves - who - as I've pointed out - did communicate with each other reasonably well. One result was the formation of a international collaboration known as IMesh, who held their first workshop last year:

International information gateway collaboration: report of the first IMesh Framework Workshop:  
<http://www.dlib.org/dlib/december99/12dempsey.html>

Technical frameworks for interoperability are part of the continuing research effort being undertaken by Renardus, IMesh Toolkit and other projects, including SCHEMAS.

Finally, a small (but important) point. Ms. Greenberg refers to three of the projects (BIBLINK, DESIRE and ROADS) as "UKOLN experiments." I think that it is only fair to mention that all three projects were based on consortia with (in total) a large number of partners. ROADS was a collaboration between three HE organisations (ILRT, UKOLN and the Department of Computer Science at Loughborough University). The European-funded projects were variously led by the British Library (BIBLINK), ILRT (DESIRE II) and SURFnet (DESIRE I).

Brief information on some of the projects mentioned in Ms. Greenberg's paper can be found at:

<http://www.ukoln.ac.uk/metadata/publications/vine-117/>

Michael Day, Research Officer, UKOLN The UK Office for Library and Information Networking

\* \* \* \*

I'm on my way to the Dublin Core conference so I can't respond in detail right now, but I wanted to send a quick thank you to you for your feedback and all the useful points.

I think *\*we\** as a community may be further along w/ bibliographic control experimentation than we think in some areas, and I will review and incorporate your comments into the paper.

To me, the pressing issues is that of communication. How do we better publicize these developments on a national and international level? The continuing impact of these various projects, whether completed or still operational, is indeed significant.

Let us work together as a community to identify these contributions and share them on a global scale. I believe you can help me with my list!

Jane Greenberg, Assistant Professor, School of Information and Library Science, UNC at Chapel Hill

\* \* \* \*

Michael Day wrote:

*>>As someone who has worked on three of the five projects reviewed in Ms. Greenberg's paper, I should first declare an interest. :-) Anyway, here are some comments (with apologies for the length) on this interesting paper:<<*

As someone who has knowledge of all of the projects compared in Ms. **Greenberg's** paper, and most of those mentioned by Michael Day, because I worked at another related project, the Internet Scout Project, <http://scout.cs.wisc.edu/index.html>

-- Scout publishes 4 electronic newsletters, the Scout Reports, which review and describe "new and newly discovered" Web sites of interest to the academic community. These descriptions then serve as the basis for metadata records, which are in a Web accessible database, Signpost. Currently, Signpost includes over 10,000 records. For its cataloging, Scout uses a number of library and library-like standards, including Dublin Core, LCSH, and LCC, and human librarians (!) Scout also maintains a US mirror of SOSIG, we participated in the IMesh meetings, and have other UK connections; Scout also has started to develop an across-metadata- repositories-searcher, called ISAAC, which is not too different from Renardus -- I also have some type of vested interest ;->

I have to agree with Jane that:

*>> ... pressing issues is that of communication. How do we better publicize these developments on a national and international level? The continuing impact of these various projects, whether completed or still operational, is indeed significant.<<*

But I also have to say, that after a quick reading of the paper, that CORC is the most library-related of the lot, and wonder if Jane had thought of that? And I think that is why CORC "wins" in her comparison, because it is the most integrated of all the efforts mentioned, in terms of numbers of collaborators, (hundreds of libraries) and in terms of built-in community. Despite the library standards it uses, Signpost is essentially a stand-alone database, and a Web user has to know it is there. CORC records can accessed through CORC itself, via WorldCat, or downloaded into a local OPAC, and accessed that way.

Also, as libraries try to "catalog" the Web, we are all attempting to create access to resources we do not own. CORC is an attempt to cooperatively catalog the Web -- last summer ('99), when CORC still was a demonstration project, I found it a terrific boon to be able to search it and see if another library had already created a record for the Web resource I had "in-hand" for cataloging.

Someone in one of the other papers at "Bibliographic Control ..." (Lois Mai Chan, I think ?) said it was occasionally painful to watch our library resource description strategies slowly being re-invented by other communities for use on Web resources, instead of taking advantage of the accumulated knowledge of the library community. Not to slight the talented library collaborators at BIBLINK, DESIRE and ROADS, I think that CORC's edge in Ms. Greenberg's comparison may at least in part be that it is positioned more to the center of the US library community, than the other efforts.

Debra Shapiro, Continuing Education Services, UW-Madison SLIS

\* \* \* \*

The **Greenberg** paper was enlightening and fascinating. I was intrigued by the many projects that had or are still taking place in some form. I suspect CORC and Dublin Core resonate with many of us and also the remaining projects in some



fashion. I was most pleased to know that each of these projects in some way were for the intent to improve web access. Though those in the non-bibliographic community and the general public seemed to have less knowledge of these efforts they did seem to notice the commercial web. And that is where I say that we in the bibliographic control community and others in the library community should look at communicating our efforts to improve web access not only inside of our own community, but the outside as well. It was mentioned in the paper that Northern Lights was using a form of Dublin Core. With this being the case it would be most interesting to see if an engine like Northern Lights would be interested in doing some cooperative work with the library community and vice versa. Communication as stated by Greenberg is important. It is especially true of those in the library profession where we have a tendency not to communicate with those outside of our sphere. And why do we usually feel that others have taken the lead in experimenting with improving web access? Well, maybe we do not communicate as we should in order to better our situation and bring others up to speed on what we would like to do or are doing. And yes, CORC serves as a point of good communication and involvement with others within the library sphere.

All in all, this paper focused on the things using these projects as a backdrop as a set of suggestions and recommendations that can be put forth.

Gail A. Spears, Catalog Librarian/Head of Cat, College of Law, Georgia State University

\* \* \* \*

**Matthew Beacom's** paper "Crossing the Divide: AACR2 and Unaddressed Problems of Networked Resources" contains the clear thoughts well expressed we have come to know through Beacom's posts on Autocat.

He examines four interrelated changes in our knowledge universe in terms of their relation to AACR2, and makes twelve recommendations for changes to AACR2.

I miss a call to make collation smds more reflective of the media being described, using patron known terms and acronyms; and a call for the restoration of smds for intangible media. He does assert than intangible media *\*do\** have a physical manifestation.

4) *Use the concepts "work", "expression", "manifestation", and "item" as more relevant to digital material than "edition", "impression", "original", and "copy", which are more relevant to analog material.*

[T]o adopt the terminology of "work", etc., as can be seen from recommendation 11 above, is not a call for the adoption of complex "work" records as advanced at the Toronto Conference, but rather a suggestion that relationships be recognized and expressed. In perhaps the most innovated paragraph of the paper, Beacom in discussing manifestation vs. work records, states that "There are other ways to split and lump. We could split at the point of record creation and lump at the point of display. Our rules could dictate that we split at the record level by cataloging each manifestation, and lump at the display level by linking each manifestation record into an integrated display of expression level and manifestation level information. ..."

Bravo. It's great when an insight such as this advances our perception of possibilities.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

By comparing the papers by **Reynolds** and **Lagoze**, one may discover some differences in the authors' perspectives.

As Reynolds points out, a bibliographic item traditionally may pass through several states: copyright application, CIP, full cataloging as a book, translations, editions, microfilming, publication on the WEB, etc. The data can quite easily differ from one state to another; for example, the author's name might not be recorded in its authoritative form (suitable

for indexing) until full cataloging is done. Interestingly enough, the form of "Figure 3 - Example of ABC event-aware model" in Lagoze's paper could be utilized to show the relationship of some of these states, and serve as a means of clarifying the relative reliability of the data at each state. Current data-gathering processes are, in part, shaped by the need to clarify which state a given item was in when an author, copyright examiner, journal editor, dissertation adviser, CIP recipient, cataloger, or other person encountered it and decided it was worth saving.

As a result, cataloging can enable the researcher and librarian to perform useful tasks. During the past week, I had to do some of these tasks while doing quite ordinary musicological research:

-- to obtain the item or its surrogate when casual access is insufficient for our purposes (perhaps there are more than 3,000 bibliographic items, and we only want one, and the item is not clearly identified); when browsing items is not feasible (as in LC); when the item is someplace else (in Charlottesville, Virginia or stored at a remote site).

-- to ensure that it's the right one when it would be time-consuming or otherwise troublesome to get it (that is, I wanted the 1983 printed edition of a specific musical work, not a recording, film, or children's book). A microfilm or scanned image might have sufficed if the quality were acceptable, but a significant change of content would be a more serious matter - and the cataloger would have noted that change.

-- to refer to a responsible inventory of items and their surrogates; in this case, a shelflist.

-- to trust that the item or surrogate can be found in future years, when it is needed again. I would really prefer not to buy all of the books and music which I might need to refer to later - for one thing, my personal library is approaching 3,000 items, and I don't want to catalog them.

As a person who has also made a career in the computer field, I would respectfully submit that these tasks change when the research is done in computer science.

-- Computer science research is often more easily available through the Web or CD-ROM than through traditional libraries - for example, IEEE journals. Special libraries for computer science can advantageously be supported through the techniques which Lagoze describes. Such libraries may have relatively little need for traditional cataloging, since their usage follows a different pattern. It is not at all helpful to limit items to traditional states. For example, they could include interactive on-line simulations or raw data files which can be used for computational purposes by the researcher. I am deeply appreciate this expanded functionality when I have to obtain such information.

-- The item's state might not affect its content; that is, a WEB version, a printed version, and a scanned image might have identical content. However, the researcher might care about the revision date and whether the item had been carefully edited for publication - in other words, states which affect content and usability.

-- The field of computer science mutates rapidly, and the value of older materials tends to decrease rapidly and dramatically. I surmise that inventories of older material are probably most valuable to authors who must sustain their careers on their publishing record than to other readers. I surmise that authors would prefer to maintain private archives of their publications rather than relying on a library to do so - it seems like a safer approach. Then, if somebody wants a copy of an article which is no longer in the library, the author can provide it. He can usually be contacted through professional organizations - unless, of course, he leaves the profession or dies.

-- Since older items have decreased value, perhaps no researcher will need to find them after a lapse of some years. I have heard that the air traffic controller system replacement effort required finding information on the early programming language JOVIAL (Jules' Own Version of the Integrated Algebraic Language) - but that is, of course, a very rare case.

Is it possible that these two papers represent not only different perspectives but different landscapes?

Judith Fiehler, Library of Congress

\* \* \* \*

**Sally McCallum's** paper, 'Extending MARC for bibliographic control in the Web environment', was very interesting, but I could add nothing to the excellent appraisal of it given by Judith Hopkins. Jane Greenberg provided an unusual approach with 'A comparison of Web resource access experiments', and this also received thorough review. I am afraid I was defeated by Caroline Arms's 'Some observations on metadata and digital libraries'.

I have been reading Matthew Beacom's conference paper, 'Crossing a digital divide'. He discusses four major changes in how sources of information are used in the new environment and recommends twelve changes to AACR2 in the light of these. I agree with much of what he says, but find difficulty in accepting the idea that Part 1 of the rules should be reorganised according to ISBD areas of description. In this respect, I share the reservations of the Australian Committee on Cataloguing, which are mentioned by Ann Huthwaite in her paper for the conference, 'AACR2 and its place in the digital world'.

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

**Regina Romano Reynolds** in "Partnerships to Mine Unexploited Sources of Metadata" quotes the recommendation that "The Library [DLC] should actively encourage and participate in efforts to develop tools for automatically creating metadata. These tools should be integrated in the cataloging workflow." This is seen as necessary due to the high cost of traditional cataloguing, coupled with the growth of Web resources.

<SNIP>

She suggests that to facilitate this cooperative effort, there should be better conversion programs, interactive instructions, change in some catalogue rules (e.g. capitalization in titles), and change in descriptive practice (abandon paragraphs in favour of values entered for a list of elements).

That last one puzzles me. When I catalogue, I enter values after MARC tags; the arrangement in paragraphs in the OPAC display happens after the fact. I much prefer numeric tags to named fields; 130, 240 and 730 are much clearer to me than "uniform title", not to mention being language neutral.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

**Larson and Arret** report on a survey with 162 libraries responding out of 453 libraries' reference librarians contacted concerning the practices in their libraries, and their preferences, for Web control.

The findings are in accord with what we have been told. There is a mixture of OPAC listings and library gateway pages. Sixty-five reported access to subscription electronic resources through the OPAC, ninety-four through web lists alone; two through the OPAC alone; with two not answering. Sometimes reference librarians created the web lists.

Only a minority feel OPAC terminals should not be used for Web access. The minimum description required is felt to be in accord with the Dublin Core. The descriptive elements wished, in order of desirability, are:

- Title
- URL
- Date of last update\*
- Author
- Language

Subject keywords  
Time period covered  
Date created  
Genre  
Publisher  
Copyright/access rights  
Subject controlled vocabulary  
Relation to other works  
Link to index  
Format  
Geographic coverage  
Summary (by librarian)  
Other identity numbers  
Table of contents

This list would seem to be to differ rather remarkably from the order in which we cataloguers would list them, perhaps in part because we tend to think of them in MARC tag order (or is that just me?). I'm particularly struck by "Date of last update", item three. The date of last update at time of cataloguing would be no problem. But since many websites, like loose-leaf services, are updated on a continuing regular or irregular basis, I am not clear how this information could be kept up-to-date.

Finally, the surveyors report some variety in the level of communication between reference and cataloguing librarians.

It is interesting that DLC staffers are taking such a leading role in Web cataloguing and access, leap frogging over the other nonbook formats the rest of us have been struggling to catalogue. I read that DLC is purchasing music records, but most formats, such as videorecordings, are very under represented in DLC cataloguing. Based on the requests of our customers' reference people, there is greater need for good derived records for nonbook physical materials than for websites, since the later can be found through existing search engines, and the former are less will controlled. The Web is more glamorous, and easier to tout in funding requests, I suppose.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

**Matthew Beacom's** paper on the "Digital Divide" is required reading for everyone who is concerned with the future of library work, not just cataloging. And for anyone dealing or dabbling in metadata. His approach is profoundly perceptive and phrased in lucid language, a pleasure to read. The suggestion to put principles (purposes of the catalog) up front is overdue (Beacom's recommendation nr. 1). AACR don't even have a single paragraph citing the Paris Principles, but actually they belong into a preamble and printed in boldface on a page of their own.

Though there is next to nothing in his reasoning I'd want to argue with, there are two additions I'd want to suggest for his 12 recommended changes:

1. The treatment of multipart items looks back on a long history of neglect in AACR territory. Contents notes are just not good enough in many cases, and neither are long sequences of 700s containing subfields \$a and \$t. If there's talk of record linking between expression and manifestation, there should also be links between parts of a larger entity (the Part -> Whole relationship). IOW, there need to be records for parts in many cases where presently there are none. In the "REUSE" project, jointly undertaken by OCLC and our network based in Goettingen, we put forward new suggestions drawing on our different tradition of cataloging. The project summary (completed 1998) is at: <http://www.allegro-c.de/allegro/formate/reusep.htm> Actually, there is no need for a change in AACR! A shift to another option that has been lying dormant all the time, that's all.

2. AACR has never had filing rules. What computers do is not filing but sorting, which is not the same. But worse, the absence of a filing code has led software manufacturers to try their luck with many different solutions. It may seem to be nuts and bolts stuff, but esp. for cross-system searching via Z39.50, these nuts and bolts have to be properly fitted or too many search results become unreliable. A few examples:

- o How are personal names indexed? Some actually do it on a word basis treating names like ordinary strings. What's to be done with hyphenated names or those containing various additions like articles and such?
- o How exactly are title strings to be broken up into words, considering all the special characters that can occur in titles (hyphens, brackets, slashes, quotes, punctuation marks ...)
- o Some systems index the German umlaut characters by just omitting the dots, some index a-umlaut as ae (which is what we do), some do both (like Melvyl). Even if your data is fully AACR-compliant, uninformed indexing can still wreak havoc. To date indexing, this last step on the way to the end-product (the usable database), is not covered by anything remotely resembling a guideline. (Correct me if I'm wrong)

Bernhard Eversberg, Universitaetsbibliothek

\* \* \* \*

RE: J. McRee (Mac) Elrod's comments below:

*>>It is interesting that DLC staffers are taking such a leading role in Web cataloguing and access, leapfrogging over the other nonbook formats the rest of us have been struggling to catalogue. I read that DLC is purchasing music records, but most formats, such as videorecordings, are very under represented in DLC cataloguing.<<*

I think the main reason LC catalogers, and many of the rest of us in the cataloging community, are keenly interested in creating catalog records for Web sites, is not that they are simply more glamorous, or easier to get funded for.

It is that the access created via Web search engines leaves libraries out, hardly anyone needs to get into a library to do a search engine search for Web sites, and this access does not in fact provide much in the way of bibliographic control for Web sites -- e.g. can you really call a hit list of 10s of thousands of un-ordered items bibliographic control?

The other reason is that there is simply far too much good information being published on the Web for us to ignore it, and this good stuff is all mixed up with an equally incredible amount of dreck. It is a \*tremendous\* opportunity for us as librarians to try to use our powers of selection and organization on the Web, and that's why so many of us want to.

This argument was already had at great length on this list the last two weeks, under the heading of alternative architecture, should we include catalog records for Web sites in our OPACS, so I don't believe I need to go on too much longer, since it is obvious by now that I side with the Martin Dillons and others who are promoting developing ways to catalog the good stuff on the Web, and integrate this material with traditional library materials, and not with the Barbara Baruths, who are waiting for the software developers to save us.

Debra Shapiro, Continuing Education Services, UW-Madison SLIS

\* \* \* \*

Bernhard said regarding filing rules:

*>>How are personal names indexed? Some actually do it on a word basis treating names like ordinary strings. ... <snip> Some systems index the German umlaut characters by just omitting the dots, some index a-umlaut as ae (which is what we do), some do both (like Melvyl).<<*

Yes to most of your points, but in terms of simplicity of filing, I have found strict word by word, nothing before

something, numbers before letters, ignoring diacritics, works best.

At a major university we used to follow the then filing rules. We kept a record of the common "mistakes" made by filers which had to be corrected by filing revisors. We found they were usually moving from filing "as is" to something else, e.g., "Mc" to "Mac", "u" umlaut to "ue", arabic numbers to as spelled, etc., etc.; in other words where a computer would have put them. We deduced that if that is where student filers filed them, that is where student patrons would expect to find them. Filing and filing revision became much easier after the refileing of the catalogued. Refiling in inverse chronological order under subject headings helped too.

I realize the expectations of our North American students might differ from European ones. Like the international authority file already mentioned here, there might have to be national differences in filing practices.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

Debra accurately said:

*>>The other reason is that there is simply far too much good information being published on the Web for us to ignore it, and this good stuff is all mixed up with an equally incredible amount of dreck. It is a \*tremendous\* opportunity for us as librarians to try to use our powers of selection and organization on the Web, and that's why so many of us want to.<<*

Absolutely. No argument. Selectivity is the prime argument for inclusion of Web material in the OPAC.

But in terms of getting \*my\* work done, with the help of MARCIt, I can create online a bibliographic record for a Website far more easily than I can sit and watch a videorecording, install a CD-ROM, figure out how to play a DVD, and then create a bibliographic record.

I'm simply commenting on our various national libraries having not created records for much of the wealth of nonbook material among which "there is simply far too much good information ... for us to ignore it".

Why the leapfrog is my question. Will all these other formats ever get the attention they deserve? Will we each continue doing our own, and deriving records of far poorer quality than those DLC and NLC do for print material? How did the Web escape the prejudice against nonbook material, if not due to the glamour of it?

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

*>>The other reason is that there is simply far too much good information being published on the Web for us to ignore it.<<*

Regarding volume (which is incredible), an item of interest regarding permanence is at:

[http://www.pfeifferreport.com/trends/ett\\_online.html](http://www.pfeifferreport.com/trends/ett_online.html)  
dated 20Sept2000  
Title: On-line publications: the beginning of the end?

[how's that for a citation format?]

Regarding Mac's point about the Web resources leapfrogging of other media in our priorities, I tend to agree with him,

but I wouldn't single out just the national libraries. This is a very large and well populated bandwagon. I believe I was on it for awhile too, but am now dragging one foot. I'm not questioning the value, just the prioritization of our focus.

Daniel CannCasciato, Head of Cataloging, Central Washington University Library

\* \* \* \*

I'm afraid that you misunderstand some of my points or perhaps one of Martin Dillon's. Neither one of us is advocating a future where catalog records for Web resources are located in the traditional OPAC. (At least, this part of Martin Dillon's paper leads me to that conclusion: "...the new task we are designing our systems for is fitting surrogate descriptive records into a universal catalog for Web knowledge resources, with the added need, at least for the foreseeable future of having this catalog work congruently and seamlessly with the bibliography of the paper world.")

Perhaps both M. Dillon and I envision a type of library system architecture that will include the capability to simultaneously search two or more disparate databases. My vision is that the databases will be both human-created (our OPACs of physically-held resources, consortial databases of mutually held e-books and e-journals, collections of digital library resources, etc.) and machine-created (pre-harvested and/or harvested on-the-fly), while apparently M. Dillon's vision is that humans will do all of the resource discovery, filtering and cataloging. In a nutshell, we differ on who ultimately will create most the metadata for billions of Web resources. I would put my money on Web mining via machine while OCLC has put its money on an army of librarians. I am a librarian myself ... I'm not talking about what I would personally like, but rather a realization that the task before the army is far too great ... it just won't scale.

Barbara Baruth

\* \* \* \*

I have been reading **Carl Lagoze's** conference paper, 'Business unusual : how "event-awareness" may breathe life into the catalog?'

The writer proposes that, in a world of constantly changing internet resources, where the continuing authenticity of a resource may be called into question and in which descriptions relating to the changes, or "events", in the life-cycle of a resource may come from a variety of sources, the role of the library catalogue should be to provide a history of the resource by mapping the different descriptions representing changes to it. He presents a model for accomplishing this task.

How to keep track of modifications to networked digital resources is a widely recognised problem. Whether 'libraries should promote the catalog as a mapping, or interoperability mechanism amongst distributed descriptions' is, however, another matter. The writer says, 'I have no doubt that this suggestion might meet some resistance from my library colleagues'. I agree with him in that. The model appears, however, to be a clever piece of work.

Philip Davis, Assistant Librarian, Redditch Library, England

\* \* \* \*

Umm, excuse me, I don't misunderstand you at all. At least two of the areas where you think the best course is to rely on software -- selecting and cataloging -- are the areas where human abilities are still superior, and may always be, to machine. Search engines are great for discovery; they are not great at evaluating resources, and they are not great at creating descriptions of resources based on an ability to sum up and characterize what a resource is about, in the way that a human cataloger can. I think Dillon (and others) advocate using machines for what they are good at and humans for what they are good at. And I am not sure that Dillon would not advocate including Web resources in the OPAC; there is a variety of ways to interpret "this catalog work congruently and seamlessly with the bibliography of the paper world."

We have to remember that the traditions of librarianship were based on the idea of information scarcity; we now have to recreate a librarianship based on information overload. This is what I mean by saying that the web presents us with tremendous opportunities -- there is far too much stuff on the web, and information seekers want us to tell them which is the good stuff, and one of the ways to do this (gracefully!) is by selecting quality Web resources, and cataloging them, therefore adding them to our library collections, thereby imbuing these selected resources with the trust that people give libraries.

And as far as scaling goes, we never cataloged every book or print publication either. I worked for an Internet resource cataloging project where, using 5 FTE librarians, we cataloged about 7,000 resources into a stand alone database in about 2 years. That does not scale, you're right. OCLC's CORC, on the other hand, leveraging an army of librarians in over 100 libraries, has grown to over 350,000 resources in about 2 years. The point is not that the Web is too big for us; it is that if we do not make efforts to select and organize the quality information from the Web, and include it in library collection, libraries will be left out. I have already heard anecdotal reports from librarians wishing to add records for Internet resources to their OPACS, that they found already extant records for a very high percentage of the resources they wished to add in CORC.

And since this reply came RE: Beacom's paper, I would point out that Beacom is even more "radically conservative" than Dillon, since rather than talking about adapting a new technology to the old (creating a DC-MARC as Dillon advocates) Beacom writes about making a major overhaul to our cataloging rules, to make them work for cataloging Internet resources.

Debra Shapiro, Continuing Education Services, UW-Madison SLIS

\* \* \* \*

I apologize for getting into this discussion a bit late -- I have been out of email touch for two weeks. First I harken back to comments made by Barbara Baruth dealing with my paper posted on the conference site. She was kind enough to forward her comments to me and in them she expresses her unhappiness with how I characterize her position:

*>>While on the topic of Mr. Dillon's article, I'd like to correct a glaring misstatement he made about my position. He summarized my vision as: "The job of resource discovery will be accomplished primarily through software acting on Web resources without benefit of human intervention, particularly of the catalog sort.<<*

According to Ms. Baruth, her position is more accurately characterized by the following:

"The future of library systems architecture rests in the development of umbrella software that digests search results from rapid, coordinated searches of a variety of disparate databases--OPACs of locally-held print and audiovisual materials, union catalogs, consortial catalogs of e-books and journals, and specialized digital library collections."

Let me clarify. My subject was access to material on the Web. I was speaking, in the quoted section, of Ms Baruth's vision of how Web materials would be accessed -- primarily through advanced search engines. "Resource discovery" in the above quote is intended to mean "resource discovery on the Web." I hope that this restriction is clearer in context. I have no quarrel with Ms Baruth's insistence that patron retrieval results will often be formed from coordinating searches over many databases.

The key point that separates us, of course, is the advisability of using humans to create (or edit) resource descriptions for Web resources versus the dependence on search engines. More precisely, we disagree over the scalability of the first and the likelihood that the second can improve to a point where satisfactory results can be achieved.

The evolution of the quality of search engines is highly speculative and I won't address that issue here; I will address briefly the issue of scalability which is primarily a matter of definition and arithmetic.



First, the definition: what kinds of Web resources would we want human resource descriptions for and what quality of record are we after? Michael Gorman has argued for an approach that divides resources into 4 levels based on their merit: the lowest level, automated extraction of attributes; the next higher, quick and dirty brief records lightly edited; the next higher, carefully edited brief records; and finally, full bibliographic descriptions, reserved for material that is of lasting significance.

Now the arithmetic: do libraries command sufficient human resources to accomplish the job? If we restrict the task to items in the top two levels, this question reduces to estimating how many such items there are on a yearly basis. We don't know that yet but we do know that around 2 million original MARC cataloging records of the traditional sort are added to OCLC each year. Let us imagine that we can divert staff now doing 1 million of these to working on Web materials. My guess is that some 200,000 Web resources deserve the full level 1 treatment. We will do these and however many level 2 records that we can with the remaining resources -- cataloging staff for 800,000 level 1 records. If we can provide level 2 descriptions at the rate of 3 to 1 of the more complex records, that give us  $800,000 \times 3 = 2.4$  million of level 2 records yearly. If we had begun this 5 years ago, we would have a database of 13 million records (2.6 million yearly). Would this be sufficient? Your guess is as good as mine, but my guess is yes.

A great deal of over-simplification has gone into the above paragraph; I would argue that the more carefully the steps are carried out, the more reasonable the approach would become. That is, under sensible assumptions, the concept of human resource descriptions for the Web is scalable.

A final point before I conclude this overly long note. Ms Baruth, in writing about my ideas, states, >>*Understandably, the OCLC position is different. OCLC has already pinned its hopes and assigned much of its resources to the CORC initiative.*<<

The position I expressed is my own, not OCLC's; I would never presume to speak for OCLC. And even though I am no longer with OCLC, having recently retired, I like to think of OCLC as having no views other than the pursuit of what is best for libraries.

Martin Dillon

\* \* \* \*

>>*Bernhard said regarding filing rules: How are personal names indexed? Some actually do it on a word basis treating names like ordinary strings. ... <snip> Some systems index the German umlaut characters by just omitting the dots, some index a-umlaut as ae (which is what we do), some do both (like Melvyl).*

>*Yes to most of your points, but in terms of simplicity of filing, I have found strict word by word, nothing before something, numbers before letters, ignoring diacritics, works best.*<<

While I agree that filing rules should be simple, the point Bernhard is making is that computers don't file, and programmers aren't programming them to file. I just looked in OCLC. Smith Ayala, Emilie sorts several screens before Smith, A. (I did a scan author search using just "smith" in the authorities file). No one expects this. While I was clearly looking for an example, and few would really do such I search, I have encountered this phenomenon in other contexts doing real searches. I'm sure readers have, too, and unless one is very alert, one can miss the entry one is looking for.

Uniform titles are another area where computers just are not programmed to file things correctly, at least, none that I have seen.

Michael S. Borries, Cataloger, City University of New York

\* \* \* \*

I hesitate to add my two cents to the discussion but for what they are worth here are a few comments.

I'm very much in agreement with the general comments made at the beginning of the **Lagoze** paper. The profession is far too conservative in taking up new ideas and certainly the catalogue has been challenged in recent times as the library's primary finding tool. Having said that, I'm drawn to conclude not that we should adopt radical changes to the catalogue, its role and composition in order to survive as suggested by Lagoze, but that we ought to apply what is already developed to the new situation. It's a case of avoiding throwing the baby out with the bath water.

If one thinks in terms of libraries today there are not one but at least two primary finding aids. These are the traditional catalogue and the library's Web site. It appears that more and more importance is being placed on the WEB site rather than the catalogue. One has only to compare library sites to see that there is no systematic approach taken by the different institutions and one can contrast this to the very structured approach offered by a catalogue founded on bibliographic standards such as ISBD, AACR and yes MARC.

Rather than discard this entire structure it seems more pragmatic to see how it needs to be modified or edited to allow for the description of information rather than just the more limited notion of bibliographic items. In this regard it doesn't seem too far fetched as for example the descriptive manual developed by OCLC based on AACR2 chapter 9 for electronic formats.

The question then becomes how to integrate the library web site with the catalogue. This integration it would seem to me would consist of: (1) user interface, (2) the information filter (i.e., the catalogue), (3) electronic delivery of information, (i.e., centralized delivery to the user in various formats, physical, ftp, fax, etc.), and (4) a transactional database working behind the scene managing each interaction with the system to ensure that tasks are completed, and that statistics are recorded. This is a model that can be distributed to the client group, i.e., delivered to the desktop in the case of a special library or limited to service within the physical library itself whatever would seem the more appropriate.

Mac Nason, Algonquin College, Ottawa, Ontario

\* \* \* \*

Before commenting on **Michael Gorman's** paper just posted, "From Card Catalogues to WebPACS", perhaps I should admit my underlying prejudice.

My Cataloguing Mount Olympus has two living residents (among the several shades from the past such as Cutter, Osborn, and Lubetzky). One is Judith Hopkins, because of her grasp of detail and ability to express them in such a manner that anyone may understand. The other is Michael Gorman, because of his grasp of underlying principle and his ability to bring that to bear on the concern of the moment. His immunity to infection by the buzz words in fashion have resulted in clear headed pronouncements, particularly in reaction to the complex records bruited at the Toronto conference on the future of AACR.

Gorman in this paper sees the story of cataloguing in the 20th century as the story of two structures: the codes and rules from Cutter to AACR2, and the expression of those rules from cards to web-based catalogues. Views that codes expressed in catalogues are no longer relevant in this time of digital information are "noxious" in that they impede progress. Digital materials will not catalogue themselves. The important question is not "How should we catalogue electronic resources?" but rather "Which electronic resources should be catalogued, and how shall we preserve them?" Effective cataloguing involves controlled vocabularies and adherence to the standards that have evolved in the past 100 years. (He doesn't address class numbers as an alternate method of subject control.)

Other writers for this conference have mentioned the selection value of library cataloguing of websites, but Gorman also addresses the question of libraries as conservators in relation to this ephemeral mass of information.

There follows a history of the successive codes of this century, with the principles each developed, which I am printing out against the time this website may vanish. He mentions the change from LC practice being stated in the rules, to the rules being contradicted by LCRIs. He sees the abandonment of principles and citing of cases as progressing to the "red and green books", ignoring the position of Osborn, and being only reversed by the work of Lubetzky. "Alas, reality intervened in the unholy alliance of traditional cataloguers ... and library administrators ... that caused Seymour Lubetzky to be replaced as editor of the code ..." The possibility of a spare and coherent international code, in which every rule is consistent with the principles and is properly related to other rules, was lost through lack of nerve.

MARC provided a vehicle for catalogue data, and ISBD provided an international standard for that data. AACR2 has a core of Lubetzkyan truth in its rules for entry, and ISBD gives a coherent structure to the rules for description. The Dublin Core and Metadata can be seen as subsets of MARC, but inferior subsets which lack the instructions on standardization of content necessary to create a bibliographic control system, and which ignore the central question, what should be catalogued?

Gorman denounces a settling for tenth best, a weird amalgam of free text searching and unstandardized uncontrolled ersatz cataloguing masquerading as a branch of information "science". He calls instead for us to incorporate all other forms of human communication into our standards and catalogues, and thus usher in another golden age of cataloguing that supports our unique task as librarians - the preservation and onward transmission of the human record.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

**Gorman's** call for us to catalogue and preserve the Web selectively in his Bibcontrol conference paper raises many questions I hope he will address.

Passing by copyright and permission questions, is each major library to attempt to store inhouse websites which relate to its mission? Or would major institutions agree to specialize, and access preserved websites in each others' electronic collections? What about websites which are in constant flux? Would only a recent manifestation be kept, continually being overwritten, or successive versions?

When SLC published a cumulating loose-leaf print version of Canadian Law Symposia Index, we were astounded to learn that the National Library of Canada kept successive superseded cumulations. The Index is now an electronic publication on QuickLaw. I'm fairly certain QL keeps only the current manifestation.

The more I think about Gorman's position, the more it seem congruent with the best in library tradition. The more I experience the ephemeral nature of some of the best electronic information sources, the more I see the value of his proposition. Even a decade ago, such a project could not have been considered. Now for about \$100 we can add more computer capacity than even the largest library could have afforded when we began cataloguing in 1979. We have every MARC record DLC has produced for books in English sitting in a little box by my foot. Before automation, this required a stack range. In automation's early days, it required a mainframe.

I think Gorman has enunciated a worthy and doable mission. I hope he is not ignored like Osborn nor shunted aside like Lubetzky.

J. McRee (Mac) Elrod, Special Libraries Cataloguing

\* \* \* \*

The conference paper by **Karen Calhoun** is quite good at articulating the current environment in which many libraries and technical services operate, and in pointing to the future directions necessary for us to follow with respect to managing Internet resources in libraries.

Among various studies described, her own study indicates a lack of full accountability with respect to access to Internet resources--not only in the catalogs of the libraries she examined, but also in the Web gateway listings. I would take her discussion further by saying that experience has shown myself and others that given title listings on the Web (created by public services staff), it is true that new titles will often appear sooner there than in the catalogs. However, the Web listings lack the authority control over the elements of titles, author names and subjects that will be necessary as changes alter those elements over the long term. So there is more than one question, not just how to increase the volume of Internet resources in catalogs through mechanisms such as streamlining of certain descriptive cataloging practices and standards, but also how to ensure long term access to a library's Web listings. Experience has shown that the background and conceptual training necessary to account for long term access, or even to account for thoughtful discussion of long term access, must be acquired by public services staff, systems staff and library administrators before workable long term solutions can be managed. While the issue of URL maintenance has begun to be widely understood, there is more that needs to be widely understood with respect to the needs of long term accessibility. As such, targeting a need for sustained communication to libraries regarding the necessity of reallocating time and existing resources, in order to successfully retrain at least part of a library's cataloging, public services, and systems staff to work together in building databases of Internet resources (be they as part of the library catalog or as part of a library Web listing) would be a highly beneficial result of this conference. In my experience at CORC meetings to date, it has been clear that there is still too much communication simply to individuals within the cataloging community and too little targeted and sustained communication to library administrators and other sections of library staff by external supporting institutions. As a result, only some of the libraries involved with CORC have effectively integrated workflows with respect to CORC. Therefore, a larger, widespread effort is necessary to facilitate needed changes in thought and direction. Leaving the proposal and implementation of necessary institutional changes simply to individual librarians, as has been the case to date, is not likely to expedite the process much. Instead, fostering institutional changes within libraries (such as workflows redesign mentioned by Karen Calhoun) in order to deal with Internet resources efficiently and effectively over the long term, will require new and redoubled commitments to cross-institutional communication and training from leading standards institutions to library administrators and to targeted groups of library staff alike.

Laurel Jizba, Head of Cataloging, Portland State University Library

\* \* \* \*

I have been reading the draft of **Regina Reynolds's** conference paper, "Partnerships to mine unexploited sources of metadata".

She says, "My premise is that metadata created for other purposes--particularly metadata created in association with existing and emerging identifiers--ISSN, ISBN, DOI--or captured from registrations such as Copyright or Cataloging in Publication, are potential sources of bibliographic data, and that non-traditional solutions such as this are the only way libraries will be able [to] gain some bibliographic control over the explosion of Web-based resources. Partnerships with agencies which collect this metadata can provide opportunities to share libraries' experience using metadata so as to make it readily adaptable for library cataloging purposes".

She lists an amazing range of potential sources of data. Her idea is a very useful one, provided, of course, that a cataloguer is responsible for the final product.

Philip Davis, Assistant Librarian, Redditch Library,

\* \* \* \*

Another possible source which she does not mention, maybe because it is too new, is ONYX. It is described thus:

ONIX International is the international standard for representing and communicating book industry product information in electronic form, incorporating the core content which has been specified in national initiatives such as BIC Basic and AAP's ONIX Version 1.

More information is available at: <http://www.editeur.org/onix.html>

This is one we should watch and work with closely.

David Bigwood, Lunar & Planetary Institute

\* \* \* \*

Yes, I could not agree more, librarians need to look at ONYX as a potential source of commodity cataloging data. I address ONYX and the larger EPICS data dictionary in my conference paper:

Priscilla Caplan

\* \* \* \*

I agree with Priscilla Caplan and others. In one of my first postings on this list, I referred to Dick Miller's LJ Netconnect article: "XML: Libraries' Strategic Opportunity", in which he notes that libraries are on the wrong side of the digital divide and that we should consider moving to XML (XMLMARC??) He goes on to note that XML is supplanting EDI (Electronic Data Interchange) standards. I'm certainly no expert in this area, but it seems that as part of a conversion to XML, librarians will need to consider developing a new schema (DTD--Document Type Definition), adapting AACR?? to DTD-hood, or working with others to influence further development of an existing DTD. Perhaps ONYX, the new standard for representing book industry product information in electronic form, is a strong possibility for the third alternative. I'm not familiar with EPICS and am looking forward to reading Caplan's article.

In case you are confused, no, I have not changed my mind. I don't advocate that we catalog the millions/billions of free sites on the Internet. But I do think that our existing catalogs for physical materials should be migrated to XML and I also think we should use XML and DTDs (a revised ONYX?, EAD, etc.) to create catalogs (consortial when at all possible) of purchased or licensed electronic books, journals, digitized special collections, etc. We need the architecture to search these disparate databases (including CORC IF it scales ... or else incorporating a strong Internet search engine), summarize, and report results to the user. To quote Miller: "Disparate data sources are easier to integrate and process when they share XML's syntax." Let's get on the right side of this digital divide.

Barbara Baruth

\* \* \* \*

B. Baruth writes:

>>... *I do think that our existing catalogs for physical materials should be migrated to XML and I also think we should use XML and DTDs (a revised ONYX?, EAD, etc.) to create catalogs...*<<

XML alone helps nobody, even with a DTD. Application software still needs to be written to digest XML data. To use XML as storage format in database systems is entirely the wrong idea. It would be a gross waste of space and make everything more complex (and thus slower) internally. What we need is the ability of exporting and importing data with XML tagging. But a DTD will be needed first that maps all the data elements currently in use. There exists one created by LC:

<http://lcweb.loc.gov/marc/marcsgml.html>

but it will be no use before it is not accepted as standard, and it will have to prove its usefulness first. As a communication vehicle, not as a storage format.

A rush into XML is not possible - no one would know, at present, just how to code their data, in the absence of a

standard. And once full MARC data can be represented in XML, everybody will find it only slightly easier to use the data than they find it now. The inherent complexity would not be removed. Most people talking about XML as a data communication format have relational database structures in mind. MARC is notoriously ill at ease with relational concepts, and this wouldn't change with XML.

Bernhard Eversberg, Universitaetsbibliothek

\* \* \* \*

I'm still at sea trying to understand the relationships and conversions possible/necessary between XML, MARC, AACR2 and various DTDs. However, I can steer you to Dick Miller's article, which I'm sure you will find interesting: <http://xmlmarc.stanford.edu/LJ/> He very briefly notes the LC literal mapping project, MARC SGML, that you also reference.

Just a couple of highlights from the article:

"The best argument for the feasibility of using XML in conjunction with database management is the example of the National Library of Medicine." XML will be the only distribution format for MEDLINE beginning in 2001. NLM plans to produce an XML version of MeSH (Medical Subject Headings) and eventually of its MARC cataloging.

"The NASA Astrophysics Data System also chose XML for reformatting all its bibliographic records."

"Despite its simple hierarchical structure, XML reveals a remarkable accommodation for complex bibliographic data. Librarians must take a very serious look at MARC and AACR2 in view of the many advantages afforded by XML."

And finally: "A fully XML-based integrated library system is feasible within three to five years. ILS vendors, notably Endeavor, are beginning to incorporate XML into existing systems."

Mr. Miller understands and presents the case much better than I could.

Barbara Baruth

\* \* \* \*

On the matter of volume of electronic resources:

Karen Coyle wrote on the USMARC list, April 12, 2000, about having "processed something close to 50 million article records ... ". At a time when computing power is huge, the catalog's purported death (or limitations) due to file size seems -- at least possibly -- to be mistaken.

**Carl Lagoze** ... has a very interesting take on some aspects of cataloging, yet in his development of the decentralized aspect of building something akin to a catalog record he lacks a specific focus on the patron. I believe adoption of his focus would first require a substantial reworking of what our catalog's objectives are. While much of the information he suggests capturing could be of interest, I don't see that information as being an inherent cataloging responsibility. As a practical approach, the computing needed to capture this information, and the costs associated with it, should be clearly within our mission before those costs are justifiable.

I must say that this line in the first paragraph -- "The conservative "business as usual" perspective of libraries ... " seems way off the mark. I have not, I admit, visited all libraries, but what I don't see is a conservative approach. A brisk walk through the main hallways of almost any library will demonstrate this. The continual call for Internet filters also indicates a lack of complacency in libraries.

Overall, I found the lack of explicit focus on a patron and a relation to the catalog to be lacking. Some of the ideas were fascinating, though. I support his view of the library as adding a form of trust to the information in the networked environment (p. 6 on my printout, second paragraph). Thomas Mann's article suggests a strength in this regard as well, as does Sarah Thomas.

**Michael Gorman and Ann Huthwaite** ... In general, I agree across the board. I especially like Gorman's challenge to the assertions that cataloging standards cannot be used for networked resources. Ann Huthwaite was very strong in this regard as well -- no need to reinvent the wheel. My only hope is that the flood of networked resources starts a very strong move towards either explicit multiple version records, or that ILS's make post-hoc ones a reality. These issues are not the ones that prove that current standards are inadequate. As Lagoze mentions in his paper, many metadata initiatives provide less than does the catalog record. In a time of flooded information, our patrons don't need cruder mechanisms, they need better ones. As Gorman concludes, we are in the process of providing less to our patrons. Or perhaps not abandoning ship, but instead continuing to improve what we do.

**Sarah E. Thomas** ... I react strongly not in support of the idea of reducing the amount of time spent on books (or videos, or CDs-Audio, etc.). Especially in conjunction with the idea of accepting copy and having something cataloged only once. Again, I think we need better records in a world of more items. In fact, many moves the past few years have been to increasing the comprehensiveness of records (TOCs and increased subject access). To accept core records, rather than advocate their enhancements by others in a process of upgrade to full records is not something I would like to see. **Thomas Mann** gives a good account of the strengths of the catalog and the value of comprehensive records. While I understand that there is a practicality issue especially for LC (increased funding? from congress?) that has an impact, I believe we should at least first fail at getting the increases necessary for such an easily documentable increase in workload. [Perhaps that's already happened.]

Also, the patrons we favor in this shift of internal staff time should be acknowledged. It's complimentary issues, those patrons we ... disfavor? slight? ... should also be acknowledged. Unfortunately for LC, it's internal policy has a large impact on almost everyone else's (at least in the US) and the information age is not yet comprehensively distributed.

I see the idea of the catalog as portal as true. I don't see that as a change, however. We've (always?) provided catalog access to the "home page" so to speak of materials that weren't analyzed in the catalog: Proceedings from conferences, short story, drama, and poetry collections, pretty much all journals and newspapers, and citation indexes. The Web has made this more manifest, it seems to me, but not fundamentally changed the nature of the catalog, but only its growth rate.

**Michael Kaplan** ... I enjoyed his assertion that an effort to catalog the Internet is equivalent to cataloging an electronic rubbish heap. It coincides well with Gorman's assertion that selection is vital in this discussion, as well as other's comments (Hal Cain, for example) on this matter.

While I'm intrigued by the idea behind the reconciliation of metadata structures into some sort of unified record, I find it's applicability to the library's mission and the catalog's mission to be at least debatable. Also, one thing I disagree with, and that is in no way unique to Mr. Kaplan, is the continued emphasis on providing access to dust jacket illustrations. Is anyone right now saving dust-jackets for future scanning and access? If not, why not? If not, why do we want to clog our OPACs with information we right now trash -- not just ignore, but trash?

Daniel CannCasciato, Head of Cataloging, Central Washington University Library