

**GENERAL DISCUSSION  
FINAL SESSION**

**DAVID BEARMAN, FACILITATOR**

**DAVID BEARMAN:**

Fortunately, my most important role is to let you ask questions that the rest of the speakers can try to answer. I would just like to make a couple of preliminary observations that may be fairly obvious.

First, we are expressing a tremendous faith in what appears to be a very powerful, logical concept — that if a data base is more consistent, users will have a better chance of retrieving the information that they require from the data base. Not only is that an important proposition to examine, it's one that the research findings and the literature suggest may not be entirely true.

We don't really know very much about why users are satisfied with data bases. But from a research point of view, satisfaction doesn't seem to be as strongly correlated with consistency as we would like to believe. It is strongly correlated, however, with another factor also relevant to authority work that we haven't discussed yet today. A very, very crude finding emerges in virtually all of the research literature, that the more access points, values, or index terms provided, the more likely users will find what they want.

And one of the things we haven't really discussed today is that authority control provides a large number of access points to a particular item even when the cataloger has provided only a single access point. So that one of the powers of the authority file is to expand the number of ways in which one can get to a specific item.

This is also relevant to our earlier discussion of the number of different places that archivists could begin to implement authority control, including names, subjects, geography, form, function, and physical material types. All of these are candidates for authority control because all of them are access points.

But we're not simply talking about consistency in the way in which the library community uses authority control. Rather we are talking about potential richness: the potential richness of information that we create as curators, or information sources to which we add value, as being potential access points for users through which they can ripple to lots of other materials.

I think that as we think about the meaning of authority work for the archival profession, and as the museum community begins to develop the potential of authority work within particular disciplines in museum curatorship, we're going to become much more aware of the value of authority work to provide access to yet another kind of reference file through which yet another kind of access is provided to get another kind of user. And this is very clear, I think, from Jackie Dooley's example of the history of Rhodesia, Miasaland, and Zimbabwe.

The real promise from the museum point of view is not so much to control words in particular headings, which, as I said, depends on a great deal of faith in the power of the logical proposition that better words get us better access by users. The real promise lies in the possibility that we can multiply access, and by multiplying access, produce a better recall, at least, if not better precision.

With that, I'd like to open the floor to a general discussion.

**TED WEIR:**

One thing that wasn't mentioned at all today, but is very important at least in public archives, is the facility for authority files to support appraisal. We keep files because they contain information that results from a certain function; they represent a certain form of material. There are other judgements that affect appraisal considerations as well, and therefore, the use of authority files to support appraisal and acquisitions is very important. Presumably, once you establish authority control over basic information, you can implement some sort of expert system to support appraisal as well.

**DAVID BEARMAN:**

Is a different search strategy involved in using authority control for appraisal decisions than that used to find materials that are relevant for research? We haven't talked about the reasons why anybody wishes to acquire information about holdings, but it is very clear that we are our own largest user. And to the extent that we make these files more accessible to users, we make our own work easier because that work depends on knowledge of what those files have in them.

**RICH SZARY:**

We haven't touched on appraisal specifically because the Smithsonian Archives is the only SIBIS unit that does that type of work on a regular basis. I don't think anything that we've said restricts the implementation or the value of this approach for materials in our repository that have been saved for whatever reason.

**DAVID BEARMAN:**

I think Jackie Dooley said something earlier in her talk that's probably relevant to this question, and that is that library authority work has, in part, been pushed forward by the need to borrow material. Also Tom Garnett said that the Smithsonian Institution Library user is a user of holdings outside the Smithsonian Institution libraries.

I think that whether there is a system in place to support use of material outside the institution or not, one of the things that authority control does is to connect you to the universe of documentation. That universe of documentation consists not only of things that are collected, but things that have not been collected, and it might, it's interesting to note, consist of things that were once collected and have been disposed of, if we kept that sort of information.

**AUDIENCE SPEAKER:**

I have a question for Tom Garnett. You said today that you don't create authority records when there is only one bibliographic record to link them to. I wondered why you do that, since it seems this practice precludes the possibility of collocating subsequent bibliographic records with the same heading.

**TOM GARNETT:**

The reason is staff time. We have a large number of bibliographic records in our data base for which only one personal author heading appears in the entire data base. You only need authority control, I would argue, at least in the narrow library sense, and especially if you operate under tight staffing and budgetary constraints, when you have multiple bibliographic records because then you have something to coordinate. If you have only one bibliographic record that has the only occurrence of a name in the data base, it's not going to get out of sync with other bibliographic records.

Ideally we would like to have an authority record for every single heading. Presently, we don't see that we will be able to afford to do that. With more experience, maybe we will.

**DAVID BEARMAN:**

The argument is certainly compelling from the collocation point of view, but it's not very compelling from a vocabulary control point of view. From a collocation point of view, there aren't two things to connect. But from a vocabulary control point of view, we don't know that the form that we have entered is the one that a user will bring to the catalog. Therefore, we're in desperate shape because we only have one variant and we have essentially authorized it and don't know whether it will do.

**MARION MATTERS:**

I think that part of why you might take that point of view in working with a single name is because if you use authority files primarily to distinguish one form of name from another similar form of name, like the different John Gardners, then for that reason it may seem that it's not important to have a full authority record for a name that apparently has no other variants.

**DAVID BEARMAN:**

I think actually there is only one thing that each of the speakers said throughout the day,

and it may be significant that that was the one thing they all said: That what one controls and how one controls it relates to what the objectives of the catalog are.

And we've begged that issue, so perhaps starting with Rich Szary who has to implement this, I can put our speakers on the spot. What are the purposes of the archival catalog other than those mentioned for the library catalog?

**RICH SZARY:**

I think we've begged the issue not only as a panel or speakers but within our professions as well, including the archival profession, the museum profession, as well as related professions.

I have my own personal biases about what I see as the purpose and role of an archival catalog. My vision stems very heavily from the view that an archives is a repository of cultural memory, and that archives are custodians of not only papers and photographs but of information, as a number of people have said, that has been buried in finding aids or in the heads of reference archivists.

I think the network of authority reference files is the purpose of an archives catalog: to be able to capture whatever information we have, whether it's about collections, activities, organizations, or people, and to make it accessible to the public.

**DAVID BEARMAN:**

There are a couple of words that haven't come up much today that seem to underlie a lot of this discussion. One of them is *explicit* and the other is *structured*.

I think that nobody in this room wishes to see archivists, museum staff, or librarians responsible for creating information they previously haven't created, or managing information they previously haven't wanted to manage. But we're struggling for a framework in which we can make the effort we're already investing pay off. That effort seems to depend on finding a way to make explicit what we know, and to place it in structures that allow others to traverse what we know and utilize it.

I think it's very important that we recognize that the added effort that we all agree authority work represents is an added effort on two fronts. First, there is an added effort up front. That is to figure out why we are capturing all this information to begin with, and how it could be captured for better use. Second, there is the added effort to actually do what we want to do; that is, to enforce the very rules that we're constructing to make the information useful.

While those are both added efforts, it's hard for us as professionals to see what the alternative is. To do things stupidly and badly doesn't seem to be a really good promotional campaign.

**AUDIENCE SPEAKER:**

I would like to address a question to Marion Matters or Lisa Weber. I recall one speaker mentioned that the overwhelming bulk of inquiries seemed to be with geographic terms at her repository.

Presumably that impression results from observations of research or inquiries. Has anyone attempted a survey or study of reference inquiries on a scientific basis to determine exactly what types of queries users pose?

**MARION MATTERS:**

Not at our repository to my knowledge.

**LISA WEBER:**

In terms of a study, no, not that I know about. I think most people at this point are using *Library of Congress Subject Headings*. There are, however, no archival guidelines for use of LCSH.

There are also local fields within the MARC format; therefore, if you've developed a local subject heading list, you can use that with the format as well.

It's hotly debated whether LCSH works with archival materials. My personal feeling is that we need guidelines for applying LCSH but that by and large they do work relatively well.

David Bearman argues that providing access through form of material and function for archival records would provide better access than subject headings.

**DAVID BEARMAN:**

I think the question needs to be divided as it was asked. The first question is: Have we studied users? And we haven't studied them enough. I would point to Paul Conway's recent work and suggest that you look at it for evidence of what little we do know about how people pose reference questions and the kind of literature there is about how people ask questions. It's clear that we need to do a lot more research.

We also need to know better what reference archivists do with questions when they are posed. Again, some of Paul's work applies here, as does the study done at NARA for the artificial intelligence project. This project is very interesting in that it tells us a good deal about the translation that takes place when a reference archivist is confronted with a user question. It tells us how that question is translated into terms that are access points to the holdings, and how we actually negotiate that question until it gives us a vehicle that can be used to navigate to an access point. So, that work is also very useful.

The second part of the question has to do with what can work. And I think that the first observation is that no one access point will provide adequate recall consistently across collections held by archives for a large number of reasons that have to do with the varieties of material we hold. All one needs to do is list the names of the archival repositories that are participants in SIBIS to realize that there isn't a single access point that will recall what a user might want from all of those repositories.

While the creator heading comes closest to being one that everybody uses, it's not very powerful from a recall point of view because you have to know who a creator is. Therefore, no single access point is likely to work. My arguments haven't been that form and function will work by themselves, but rather that we haven't done what we should be doing, especially in the government records area, with creating access by form and function. They're much more powerful than we've given them credit for being, but we've buried that stuff pretty deep inside our finding tools. We've not done a good job of making them explicit or structuring them so that they can be used. But I don't for a moment believe that they alone are going to answer all the questions either.

**MARION MATTERS:**

I want to make a brief comment on what I said initially.

We do know a little bit about what our users are doing from the forms they fill out once a year designating who they are with a brief description of their research. If a person is doing family history or local history, I think you can infer that the primary access points are going to be personal names and geographic locations.

There is, however, another type of information that we provide in our finding aids that could help those types of users. We currently don't provide it as an access point. In one case, it's arrangement. Potentially any collection in which the components of the collection are arranged alphabetically by the surname of the person to whom the material relates is of potential interest to a family historian, or anyone else looking for name access. Our records need not provide the actual names but we could give an indication that the arrangement makes it easy for them to look for names.

The same thing applies to collections that contain materials that are classified and then arranged by geographic name. Right now, there is no way for the user to identify collections arranged in a manner best suited to their research objectives. I wish we could.

**DAVID BEARMAN:**

Yes, I think one of the things we need to know more about is the way our finding tools

are structured. ~~Twenty years ago on the way I did work at a reference library.~~ I put together my very first reference tool. It was an index to the way in which indexes to scientific material were accessible. We don't really know this.

Peter Hirtle and I were having a discussion over coffee about how we don't provide access to archival collections by social roles. In fact, we think a lot about social roles, especially since we know that so many of the individuals who are the creators of our material are, in effect, anonymous. But when we think about social roles at all, we think with this terrible bias, that the only relevant social role is occupations, even though we know perfectly well that our users are interested in homemakers on the frontier, or pregnant women in large cities, or labor union organizers in the 19th century, or community figures in religious organizations, or whatever. And these things are not people's occupational roles at all.

One of the things we do need to do in conjunction with understanding the specific uses of our collections better is to take advantage of existing files that contain vocabularies of these sorts and use them in association with our materials.

The fact that the occupations of persons in the United States have been collected since 1810 should tell us something about the possibility of not only accessing archival material by occupation, or museum material by occupation, but a much more exciting possibility. That is, of being able to link the materials that we find in archives with all kinds of data.

We have biological species in the museum. We also have records of people who collected biological species that are now extinct in our archives. How are we going to move between these different data sources?

The very first question you have to ask is: What are the access points across which we could move, and what could we get out of controlling them? Are there, for instance, existing vocabularies that could be imported?

When Rich Szary says that a vocabulary can be imported into SIBIS and points out programming problems, the programming problems are immensely trivial compared with the intellectual problems of how we make the connection.

**AVRA MICHELSON:**

We have covered a tremendous amount of information on authority control, and when we leave this room it's my job to provide assistance to SIBIS users. I'd like to ask all the speakers what would be the concrete steps to take towards implementing authority control?

**DAVID BEARMAN:**

Take it in the order of the day.

**JACKIE DOOLEY:**

I base what I'll say partly on the results of the survey that went out to the SIBIS-Archives users asking various questions about standards, and access points in use, and such.

The responses indicated that it was very important to offer training for staff in descriptive standards, to let you know what the concepts were all about, to write cataloging rules and standards, so that you had determined the role of your catalog record and what you wanted it to look like before you decided what kind of authorities were needed.

That is absolutely critical. If you haven't decided what you want your records to look like, particular vocabularies are difficult to address. But once you have defined your record structure and know how you're going to select and formulate the data to put into the records, it becomes easier. When you ask yourself some very hard questions about what your collections consist of and where their strengths lie, you can begin to look around at what vocabularies and existing systems are available and select one or two or three, as the case may be, that will really work for you.

**TOM GARNETT:**

I want to return to the basic question of what you want out of a catalog.

I think the questions of forms of access should follow from what the catalog is for, how it's going to be used, what you want it to do. Then I think you could consider issues of access. And then from there the issues can get more technical, but intellectually less demanding.

**MARION MATTERS:**

One of the things I would include is the importance of making conscious decisions about the relationship between the catalog record and other finding aids because they all need to work together as a system.

Deciding on the shape of the catalog record will affect how users get access to the finding aids that stand behind or don't stand behind them, depending on what your descriptive practices are like. To illustrate, I'll mention the process that we're going through right now. Among our state archives collections, our finding aid system consists of a series summary, or collection summary, followed by container lists. That is it. That is the finding aid.

When we pull that data into the automated data base, that series summary becomes the catalog record. So the summary and the container list is the entire finding aid system for the archives.

With our manuscripts collections, we have a slightly different problem in that we have for a long time created inventories that are fairly long narrative, descriptive, and biographic descriptions, followed by container lists.

The catalog record is yet another summary. The relationship between the catalog record and the finding aid is a little bit different in that circumstance, therefore, than it is for the archives. Implementation problems arise in deciding where the access points that are going into your automated catalog will come from. Will they come only from the abstract, or will they come from the abstract and the finding aid that backs it up? Deciding on the relationship between the catalog record and the rest of the finding aids is fairly crucial.

**LISA WEBER:**

I agree with everything everybody has said so far in terms of asking why you want to build this catalog and looking at what needs to be under authority control.

To say a little bit more about what Marion Matters is saying, I think in some ways the MARC AMC format adds to the confusion in determining what you're describing in the AMC record. One of the reasons is that it's very flexible. It doesn't prescribe this level or that level. I think that for various and sundry political reasons, the flexibility of the format was emphasized.

But my sense, after having led 12 AMC workshops at this point, is that people really want standards. Archivists really want to be told, essentially, this is how you use it.

The AMC record is a pointer to the finding aid. It is not the finding aid. There is no place to put an inventory or register within the format at this point, although I think that's one area we need to work towards because I think we do need to link terms in machine-readable form with the finding aid — but that's a whole different issue.

One thing that I find interesting about the library community, which is a good number of years ahead of the archival community in dealing with online catalogs, is that I sense that the library community has been building online catalogs for librarians. And now they're starting to think about the user. But they had to go through this transition. I think we're doing the same thing. I think we're building catalogs basically for archivists at this point. And we haven't really taken the leap of defining what a user catalog would look like.

Maybe we really do need to go through this process. I came from a repository that used and was a big proponent of SPINDEX. I think we needed to go through SPINDEX to get where we are today.

I think it's one long process, and we're never really going to get there. We're just going to

**RICH SZARY:**

I heartily endorse what everybody has said. One of the major problems that we have with SIBIS-Archives at the moment is, to some extent, a public relations problem. And it goes back to what the role of the catalog is. It goes back to are we, in fact, committed to sharing information across repositories within the institution? If we are committed to that, and if that is an objective for repositories in SIBIS-Archives, where are the resources to support it?

One of the lessons that emerged today is a sense of the immense amount of resources needed to create and maintain these files. I think the SIBIS-Archives users need to take a hard look at what they want their catalog to do.

How much of it can the system do for them? How much are they willing to invest in the effort? It may be as stated before, that there are going to be compromises that need to be made. LCSH isn't perfect, LC names are not perfect. There are problems with them, but if that's all we can afford to buy into at this point in terms of human resources, does that buy us enough?

Once we establish that baseline of what we are doing, why we are doing it, and what we are willing to invest to get there, then we can start dealing with some of the conceptual issues as well.

**JACKIE DOOLEY:**

One more thing about user studies and studies of use of collections that comes from libraryland.

Only 15 or 20 years ago, librarians started doing serious systematic user studies. Probably a little farther back than that they were doing studies of manual card catalogs, and then maybe 5 to 10 years ago, they started trying to do systematic studies of the use of online catalogs. The main results of those earlier manual catalog studies, which was that the known item is supreme, suddenly started being shot all to hell by user studies on online catalogs.

Suddenly subject searches reigned supreme; the vast majority of searches were seen to be subject searches. People came to the library and wanted to know something about a subject. And, we, as librarians, had been putting most of our effort into describing the known item. Various lessons started to emerge from that.

The other thing they discovered was that people not only did a lot of subject searches, they were happy with their subject searches despite the fact that there were no online thesauri; that there was no online subject authority control; that usually all they had was something like keyword searching of titles. Again, this was exactly what is not supposed to work. And it starts telling you that keyword searching has its place.

The moral of the story is that there are many different things going on in the catalog, and the big problem is trying to figure out how to put them together into a whole set of search techniques that work together well. Keyword access has a wonderful place as an entry point to authority control. Keyword searching of the LC names file, keyword searching of a subject thesaurus file, gets you into the file, and then you have a higher level of consistency at that point. Your users do not go away unhappy after conducting online searches that resulted in nothing.

**DAVID BEARMAN:**

I think there are really two issues. There are tactical issues, which are addressed largely in my report summarizing the survey of SIBIS users. The tactics have to do with how best to improve the quality of the existing finding tools that are being created within the archival units, including the finding tool called "SIBIS-Archives."

It's not surprising that tactically it appears that some fairly basic things should be done first: We should decide what those tools look like and what the rules for building them are. We

should teach everybody on the staff who is building them to build them in the same way, then check to see what the existing backlog of records that we have already created looks like and clean up those records. Then only at some point later down the road, is one really prepared to move into thinking about how to implement authority control and where to implement authority control.

Now, that's not surprising. That came directly out of the survey as the analysis of what you told yourselves, in effect. It's also a path that is possible to pursue.

The strategic issues need to be addressed, however, at the same time. And they have to do with the question that keeps coming up here: What do we want the catalog to do for us?

But even larger issues than that face the Smithsonian archival units, and probably the archival profession, and almost certainly the museum profession. They have to do with how we are going to make the cultural resource, which we invest our energies in creating, — that value-added product that represents the holdings that we spend so much time curating and managing — how are we going to make that cultural resource valued by the culture?

That turns out to be a political question about how we are going to make people in the society value a product that we are creating: How are we going to make it relevant to them? How are we going to deliver it to them? How are we going to make it exciting to them? How are we going to package it? How are we going to attract people to its use?

Ultimately, those are the issues that are at the heart of the survival of the cultural repositories that we serve in. If the society finds that they're a luxury, doesn't wallow in them, then the finding aids we create will soon be starved. I think that issue has got to be addressed by all of us.

One of the reasons we've talked about authority control at all, is that there is a sense that authority control is part of the answer to that strategic question. It might make things more accessible to more people from different perspectives and different levels of education and background. It might bring those holdings that are relevant to their attention in ways that we have not been able to do in creating finding tools that serve only our internal purposes. I think that is why authority control is on the agenda.