Notes on Operations

Name Authority Control in Local Digitization Projects and the Eastern North Carolina Postcard Collection

By Patricia M. Dragon

Authority control is a vitally important but frequently overlooked aspect of metadata creation for local digitization projects. The addition of digital projects metadata to the traditional cataloging environment creates a number of challenges for authority control, challenges arising in turn from the nature of the materials being digitized, choices made during the project, and the tools used for the project. By examining the authority control applied to named entities in the Eastern North Carolina Postcard Collection at East Carolina University, this paper describes these challenges in some detail, and also describes endeavors to overcome them.

M any libraries around the world are developing digitized collections emphasizing their regionally significant special collections holdings. By compiling these digitized collections, libraries are promoting access to the historical record of the local community while acting upon a belief that these resources inspire interest beyond the immediate community. An integral component of these digitized collections is the metadata that are applied to them, making the items in the collection findable through user searching. Authority control adds value to metadata by ensuring that the same heading is used to refer to all instances of a named entity, thus collocating related material.

The case studied in this paper is the Eastern North Carolina Postcard Collection, particularly the authority control of named entities and places used as subject headings. The case study illustrates how applying authority control to named entities in digital projects metadata creates several challenges for catalogers. These challenges include the complexity of work arising from the form and subject matter of the materials digitized, the volume of work created by a large number of new authorized headings per bibliographic description, and the inefficiency perpetuated by the lack of actual authority data in the repository database. The author proposes that, while some choices made by individual institutions may differ, the challenges encountered in this case study can be generalized to the many similar projects being undertaken in libraries throughout the world. Grappling with these challenges is essential to the achievement of goals with widespread appeal, such as the integration of digitized collections more fully into library collections and the improved usability of digital repositories through quality metadata. The paper commences with a literature review followed by a description of the Eastern North Carolina Postcard Collection. It then discusses in detail each of the challenges encountered when applying authority control to named entities in the subject analysis of this project as well as efforts to overcome these challenges. The paper concludes with a discussion of points needing future research and of the importance of meeting these challenges for the sake of future digitization projects.

Patricia M. Dragon (dragonp@ecu .edu) is Head of Special Collections Cataloging, Metadata and Authorities, Joyner Library, East Carolina University, Greenville, North Carolina.

Literature Review

Traditional application of authority control in the metadata context has not been explored extensively. Some even question whether the two concepts are compatible. Gorman disparages metadata standards such as Dublin Core (DC) because of their alleged lack of authority control.[1] He ignores the fact that many metadata standards, including DC in its qualified form, make possible, even if they do not require, the explicit use of controlled vocabularies for subject analysis. Granted, applying these qualifiers robs DC of much of its simplicity, one of its major advantages. Vellucci points out that many metadata schemes, including DC, offer the opportunity to apply authority control, but whether a particular project takes advantage of that opportunity is the result of local policy decisions.[2] She argues that when information specialists and catalogers create metadata for high-quality and long-lasting documents in a library context, "authority controlled data content should be the norm."[3] This position is supported by Baca, who argues for the importance of controlled vocabularies for the subject analysis of cultural heritage material.[4] Baca indicates that many institutions are finding that the best practice is to create a local thesaurus rather than using a single, established one, but she does not dwell on that decision process.[5] Instead, she outlines the benefits to users of a controlled vocabulary, regardless of its origin, including navigation helpers such as broader and narrower terms and cross references.[6] While both Vellucci and Baca argue for the importance of authority control in the metadata context, neither details the process of applying authority control to digitized materials, nor do they indicate the specific authority control challenges involved in working with these materials.

The literature describing digital projects often focuses on resource description. When authority control is mentioned, it is usually to say simply that names are entered according to the Library of Congress Name Authority File (LCNAF) or constructed according to *Anglo-American Cataloguing Rules*, 2nd ed. (*AACR2*).[7] Details about this process and experiences related to these types of materials are rarely given. An exception is Graham and Ross's article "Metadata and Authority Control in the *Civil Rights in Mississippi Digital Archive*," in which the authors explain the role of catalogers in creating new authorized subject headings and give philosophical underpinnings for having catalogers create authority records for personal and corporate names used as subjects: "The catalogers feel a sense of duty to establish headings of extramural reach in the Library of Congress (LC) authority file despite the increasing demands this activity requires."[8] This sense of duty, they write, arises out of a desire to "maintain global interoperability of the catalog" by using the same subject thesaurus for all library materials.[9] The authors, however, do not go into detail regarding these increasing demands.

The challenges involved in subject analysis for images of named entities and its concomitant authority control have been well noted. Cuccurullo offers an intense look at the difficulties inherent in applying the guidelines in the LC *Subject Cataloging Manual* to images of the built world, including the confusion regarding which set of rules to follow—those for the subject file or those for the name file—and the implications for the user when the cataloger follows one or the other set of rules.[10] She suggests several possibilities to alleviate the problem, including combining the authority files or adjusting or clarifying the rules. While she suggests that a quick resolution to the problem is needed because of the implications for digital projects metadata, she does not specifically address the challenge in the digital projects metadata context.

Cataloging Cultural Objects (*CCO*), a manual of *AACR2* application to works of art, is primarily concerned with dividing descriptive information into a linked hierarchical construct wherein some elements would apply to the object (or building) itself, some to the image of the object, and some to the digital reproduction of the image.[11] The introduction, however, contains a very useful examination of some of the challenges involved in the subject analysis of images, including the question of specificity and a rationale for making specificity decisions. The present paper addresses this issue below. Once again, however, specific challenges involved in establishing named entities as subjects are not directly addressed by *CCO*.

Because authority control is difficult, it is expensive. Tillett summarizes: "Since the 1970s people have claimed that authority work is the most expensive part of cataloging, and we still seek ways to automate and simplify the work to reduce costs."[12] She introduces the concept of the virtual international authority file, a means of sharing locally created authority data internationally without the unconstructive need to decide upon a single authorized heading for all. Borbinha interprets this need to share authority data created according to different standards in the context of digital libraries with the blunt statement, "Deal with heterogeneity."[13] Another practical approach to the challenges of authority control is offered by Younger, who argues for the selective application of authority control.[14] She refers to the concept as "utility" and explains that because of the time and cost involved, as well as because of the online catalog's superior access in comparison to the card catalog, authority control efforts in the new century should be directed to those areas in which it has the most potential effect on user retrieval in the online environment.[15] The concept of utility is one heavily invoked in the case study below.

While the aforementioned articles and books are largely practical in their approach, a philosophical document, *Functional Requirements for Authority Data: A Conceptual Model (FRAD)* serves as the basis for the application of authority control to the metadata environment.[16] Enumerating what authority data are, what place its creation has in the metadata assignment process, and what its functions are, *FRAD* is divorced from any specific schema or community of practice.[17] In the present case study, *FRAD* reveals problems with the implementation of authority control and suggests a path for future development.

Description of the Project

The Eastern North Carolina Postcard Collection currently consists of 404 picture postcards of Eastern North Carolina scenes. It was selected from several manuscript collections held by East Carolina University (ECU) Joyner Library. The postcards have been scanned by the library's Digital Collections Unit and ingested into a locally built digital object repository separate from the library catalog. This repository contains digitized versions of materials in the library's collections, and the postcard collection makes up only a small component of it. Materials enter the repository from a variety of work streams, including locally created digital exhibits and collections, grant projects, and user-initiated scanning requests. The public interface to the Joyner Library Digital Collections repository (http://digital .lib.ecu.edu) is searchable and browseable using a variety of access points, including subject headings. Metadata Encoding and Transmission Standard (METS) records, containing both Metadata Object Description Schema (MODS) and DC data, have been created for each image and uploaded to the repository to support this access. Users coming to the repository can

St. Peter's Episcopal Church, Washington, N.C.

Title	St. Peter's Episcopal Church, Washington, N.C.
Creator	Hardy, Ira M.
Description	Postcard of St. Peter's Episcopal Church, Washington, N.C. Addressed on verso to Mr. Martin Douglass, Chapel Hill, N.C. Date from postmark.
Date	1909
Subject(s)	St. Peter's Episcopal Church (Washington, N.C.) Anglican church buildings--North Carolina--Washington United States--North Carolina--Beaufort--Washington
Publisher(s) of Original	N.E. Paper & Stationery Co., Ayer, Mass.
Identifier	39.4.b.163
Type	still image
Medium	postcards
Physical Description	90mm x 140mm
Language	English
Collection(s)	Eastern North Carolina Postcard Collection Architecture
Finding Aid	Elmore, William E., Collection
Other Items	in this folder in this box in this collection
Rights	Rights to some materials may still be retained by the publisher and/or author (or their descendants) in accordance with U.S. copyright law.

Figure 1. Postcard with Its Associated Metadata

choose to limit their searches to the postcard collection, or they may search the postcards along with the entire contents of the repository. Figure 1 is an example of a postcard with its associated metadata.

Note also that although both sides of the postcard were scanned and are available for viewing, the front and back of the cards were treated as a unit for descriptive purposes. Because the library chose to view the postcards primarily as photographic resources rather than as items of correspondence, the photographer is considered the creator in the metadata, and the focus for subject analysis is the image portion, or the front of the card. Most postcards in the collection depict named structures of some sort, such as houses,

streets, buildings, bridges, cemeteries, and so on. For the subject vocabulary, the Library of Congress Subject Headings (LCSH) and LCNAF were chosen. The reasons for this choice are discussed below. The final subject heading (shown in figure 1) is not an LCSH, but rather a hierarchical geographic coverage element that could potentially be used to locate items on a map of Eastern North Carolina. These "subject headings" are not the concern of this paper.

Some of the metadata were entered by Digital Initiatives staff as a part of the scanning process, including identifier, type, medium, physical description, language, collection, finding aid, other items, and rights. Several of these elements are standard across

the collection, while others are dependent on handling the physical postcard (e.g., physical description) and noting its place in an archival folder or box (e.g., other items). Metadata entered by Digital Collections staff were later enriched by other descriptive information, including title, description, date, subjects, publisher of original, and creator, if there was one, by the author, a staff member in the cataloging department. Cataloging staff worked without the physical item in hand, and based descriptions on the digital image.

Involving the cataloging staff in this way seemed to be natural given the expertise catalogers possess in descriptive metadata creation. The project also fits the current work trends within the cataloging department, which is spending an increasing percentage of effort on local and special collections materials. This is a result of spending less time on widely owned materials for which it is more efficient and cost effective to buy cataloging services. The regional and special collections emphases of the digitization program and the cataloging department naturally encourage these groups to work together. Part of Joyner Library's in-house cataloging effort for traditional materials goes toward authority control. Though Joyner Library and Collections does not participate in the Name Authority Cooperative (NACO) or the Subject Authority Cooperative (SACO), catalogers create local authority records for named entities and places from Eastern North Carolina in the library catalog. While this commitment to authority work for local named entities was also applied to the postcard project in the repository database, it could only be partially applied because the repository does not include authority records *per se*. Rather than creating entire authority records, therefore, the cataloger merely assigned subject headings, which often involved creating new headings for named entities. The remainder of this paper will focus on the experience of applying authority control to named entities and places used as subject headings in the Eastern North Carolina Postcard Collection, and on the challenges faced by the cataloger performing this work.

Challenges

Complexities in the Subject Analysis of Images

Images are popular to digitize because they are attractive to users with varying levels of sophistication, and they may be manipulated in more ways in their digital form than in their print form, for instance by zooming in. They frequently belong to "hidden" collections of primary sources that were not easily accessible before digitization, and they give users valuable views of specific structures and places unencumbered by someone else's interpretation, as users would find in a book. In the words of Coyle and Hillman, "these collections [of primary sources] . . . are not the product of the scholarly enterprise, but instead the precursor."[18] To the cataloger performing subject analysis, however, these images can be much more challenging than monographs or other library materials, precisely because of their form. Free from the interpretation a book gives its topic, an image often lacks the context that tells the cataloger what is important about it, how users may wish to search for it, or under what circumstances they may wish to find it. The problem of ambiguity is made more acute by the fact that, for an image, the cataloger is responsible for providing the only words by which this item may be recalled, which would not be the case for a full-text searchable digitized text. While this problem is somewhat mitigated in the case of postcards, which typically include a caption that provides some information about the focal point of the image, postcards remain much closer to contextless images than to interpretive monographs.

This lack of context gives rise to the question of how specific (and how general) the subject analysis should be. For instance, for a postcard of St. Peter's Episcopal Church in Washington, N.C., should the cataloger assign a subject heading for the specific name of the church? LC practice is to assign subject headings at a level of specificity matching the content of the work being cataloged.[19] But would exclusively assigning a name heading impede access for those interested in Washington churches but not knowing the name of this particular church? Should more generic headings be assigned, such as "Anglican church buildings—North Carolina—Washington," even though such headings are normally assigned to works that discuss such churches collectively? What about users interested in Washington architecture? Should the heading "Buildings—North Carolina—Washington" also be assigned? In that case, where should one stop? On the question of specificity, *Cataloging Cultural Objects (CCO)* advises, "The greater the level of specificity . . . in catalog records, the more valuable the records will be for researchers."[20] In keeping with this advice, two headings were assigned: a specific name heading and a heading for a generic category of entity: "Anglican church buildings—North Carolina—Washington." This practice of dual (specific and generic) heading assignment is an attempt to make up for the lack of a syndetic structure of broader and narrower terms contained within LC subject authority records. That this deviation from LC practice creates a more useful repository is again supported by *CCO*: "If it is not possible to link to hierarchical authorities, it may be necessary for catalogers to enter both specific and generic terms in each record to allow access, which may differ from traditional bibliographic practice."[21]

It could be argued that it is unnecessary to create specific subject headings for named entities such as St.

Peter's Church, since keyword access to the titles, or captions, of the postcards should render the desired entity findable. Many similar projects rely on such keyword access and do not include subject analysis at all, or assign only generic subject headings. It is true that many users may discover the images in Joyner Library's repository through an external search engine such as Google, which operates exclusively on a keyword basis. The value of specific subject headings, however, is increased by their linking potential. If the metadata contains a specific heading, once users view an image of St. Peter's Episcopal Church within the repository, they are able to click on that heading to find all the pictures of that particular church in the repository, and not of any other church. If future links are to be made between the digital repository and the library catalog, the data will be consistent, and the user could find books about the church and pictures of the church in one search. Taylor emphasizes this superiority of access points in establishing explicit relationships between materials: "When relationships are merely described (e.g., mentioned in a note [or a title]) . . . the user is left with chance as the means for discerning related information packages that may be useful. Access points can make relationships explicit."[22] Any institution committing the time and effort to digitization projects such as this one should be concerned with providing quality metadata to make the resulting collection navigable by users. Minimization of the necessity of relying on chance for information discovery is the mark of quality metadata.

Complexities in Name Authority Work for Subjects of Local Images

While some complexities arise in determining the correct level of specificity to which to adhere in subject analysis, the real challenge comes with the name authority work involved in assigning specific headings for a project such as this one. These images are about a particular location, such as a building, street, beach, or waterfall. The local nature of these named entities means that few of them will be found already established in the LC authority files. This requires that the cataloger either determine the correct form of the heading according to complex rules, or else forgo assigning a specific heading for a named entity. Thus the very quality that makes these images valuable to the user—their unique representation of named entities not a focus of published works—increases their complexity from the perspective of the cataloger.

Choosing the correct form of name for these entities is made more complicated because many of the headings for buildings, structures, and various types of corporate bodies fall into an acknowledged group of ambiguous entities. According to the guidelines contained in the LC *Subject Cataloging Manual* instruction sheet H405, some of these named entities (e.g., Banks, Cemeteries, Churches) are established in the LCNAF following name heading conventions and others (e.g., Bridges, Courthouses, Dwellings) in the Subject Authority File (LCSAF) following subject heading conventions.[23] Following one set of conventions rather than the other could result in a different heading. It has been argued that "the separation of controlled names and terms into a Name Authority File and a Subject Authority File is artificial" because it adds no value for the user and is ignored by many library systems.[24] The policy of separation nevertheless remains codified, and debating the pros and cons of these rules is beyond the scope of this paper.

The process of creating headings was sometimes relatively straightforward. For example, the postcard with the caption "A. C. Monk Tobacco Company, Farmville, North Carolina" (figure 2) required the creation of a corporate name heading with no representation in the LCNAF. *AACR2* directs the cataloger to enter a corporate body under the name by which it is commonly identified in items issued by the body, or lacking those, from reference sources.[25] A search of OCLC's WorldCat revealed no items issued by the body, necessitating consultation of reference sources. Following the NACO practice of using the item itself as a reference source when no conflict is found in WorldCat, the cataloger created the name as found on the item itself: A. C. Monk Tobacco Company.

Sometimes significantly more work was necessary to determine the correct form of the name. Take, for example, the three postcards entitled St. Peter's Church, Washington, N.C.; St. Peter's Protestant Episcopal Church, Washington, N.C.; and St. Peter's Episcopal Church, Washington, N.C. No heading exists in the LCNAF for this church, so the cataloger needed to make a decision about the correct form of the name to use. A search of WorldCat revealed three books issued by the church, with varying usage: Saint Peters Parish, St. Peter's Episcopal Church, and St. Peter's Church, the last with a variant usage on the cover, St. Peter's Episcopal Church. Using the predominant form of the name found in works issued by the body, St. Peter's Episcopal Church (Washington, N.C.) was chosen.

Frequently, variant names are used for the same structure on different postcards in the collection. For example, a particular pavilion in Wrightsville Beach called variously Lumina, Greater Lumina, or Lumina Dancing Pavilion, is the subject of a number of postcards (figure 3). No appropriate heading exists in the authority file. According to the *Subject Cataloging Manual* H405, pavilions are established in the LCSAF following instruction sheet H 1334, "Buildings and Other Structures." This instruction sheet states, "Enter the heading for

190 Dragon

LRTS 53(3)

a particular building or structure directly under its own name, in uninverted form, and qualify it by the name of the geographic entity in which the structure is located."[26] But what is this structure's "own name?" The conflict of the various names used was settled by a book about Wrightsville Beach that contained a chapter on Lumina.[27] There the author states that Lumina was an entertainment center that opened in Wrightsville Beach in 1905. A new center containing a movie theater opened in 1909 and was called Greater Lumina.[28] Following the instruction in the *Subject Cataloging Manual* H1334, the cataloger created the heading "Lumina (Pavilion : Wrightsville Beach, N.C.)."

The preceding examples give a sense of the complexity involved in assigning specific subject headings for named entities in a project such as this one. It could be argued that the library brought this challenge on itself by the choice of LCSH and LCNAF as the controlled vocabulary for the repository, saddling itself with requirements to follow complex rules

Figure 2. A. C. Monk Tobacco Company

of questionable applicability outside the context of NACO contribution. The choice of LCSH and LCNAF for items in the repository was made prior to undertaking the Eastern North Carolina Postcard Collection project and had to do with the fact that many of the first materials in the repository were digitized books. Perpetuating this choice came from a desire to make all metadata added to the repository compatible with each other in terms of controlled vocabulary. The choice of LCSH and LCNAF also makes the repository more compatible with the library catalog, increasing consistency for users and furthering the library's goal to integrate digital collections more fully with the rest of the library collections. The use of such a widely used standard also increases potential for interoperability not only within the library's collections but with other collections from other institutions, and makes possible the future sharing of authority data should such a step be deemed viable or useful. In fact, it is probable that, apart from a simplification of the rules for heading construction, the only solution to the challenge of complexity is found in sharing as much data as possible. If more institutions continue to undertake the digitization of material with a local focus, it is reasonable to expect a proliferation of geographic-based NACO and SACO funnels mirroring the proliferation of locally focused digitization projects.

Figure 3. Greater Lumina

Volume of Authority Work Created

The large number of named entities pictured in this collection presents an additional challenge. As noted above, the local nature of the named entities indicates that few of them will be found already established in the authority file. Cuccurullo warns of the impending flood of names to be established as a result of cataloging digitized collections at the image level: "The number of such headings [for buildings and structures] is likely to increase exponentially as libraries focus attention on cataloging of digital collections."[29] The Eastern North Carolina Postcard Collection of 404 postcards depicts approximately 429 named entities, with some entities appearing on more than one postcard and some single postcards depicting more than one entity. Of these, 113 were associated with authority records in the LC authority file; the remaining 316 were not.

Most institutions have only a limited amount of personnel time dedicated to doing subject analysis and authority work for digital projects. As noted previously, the work for the Eastern North Carolina Postcard Collection was limited to a portion of one staff member's time. When time and personnel are limited, it is prudent to invoke the law of diminishing returns on that time investment, a concept Younger calls "utility" in authority control.[30] By "utility," Younger means that authority control efforts should be directed to those areas in which it has the most potential positive effect on user retrieval. "Categories," she writes, "can be defined for names requiring more or less control."[31] Though Younger's article is concerned mainly with personal names of authors, her thesis is applicable to named entities as subjects as well.

Despite generally espousing the principle of specificity, cataloging staff did not create all the specific headings possible for the postcard project.

Figure 4. Pitt County Court House

In an attempt to meet the challenge of high volume, a selection process was applied to pare down the number of specific headings created. This selection process was informed by Younger's concept of utility. Those images not receiving a specific subject heading received only a generic category subject heading. In deciding whether to create a specific heading for a named entity, the cataloger tried to be as consistent as possible across a generic subject category. The decision about how to treat subject categories was guided mainly by considerations of the effect on user retrieval. User retrieval is greatly improved by specific headings when that heading adds unique information to the subject file, when there are many similar names to keep organized, or when there are many potential links to other materials about the entity through specific subject analysis. When these conditions were met, the cataloger was more likely to create specific headings for that category of entity. Other considerations included practical concerns regarding the prohibitive amount of research involved in specific heading creation. By the end of the project,

126 unique new specific headings for proper named entities had been created. The table of generic category treatment (see appendix) shows the breakdown of how generic subject categories were treated.

To make a decision about how to treat each subject category, it was necessary to create a checklist of three questions. The first question on the checklist was whether assigning the specific names in the subject category added unique information to the subject file. A generic term plus geographic qualifier has a high likelihood of resulting in exactly the same words as a generic heading with geographic subdivision. For instance, specific headings were created for hotels because most of the hotels had nongeneric names (e.g., Hotel Kennon adds the unique word "Kennon" to the subject file). Courthouses, which have generic names like Pitt County Court House (figure 4), did not receive specific headings, rather only generic heading with the nearly identical words "Courthouses—North Carolina—Pitt County." Assigning exclusively generic headings only for courthouses actually results in more consistency because

then "courthouse" is always entered as a single compound word in the subject file rather than as it appears on the postcard, where it is sometimes "courthouse" and sometimes "court house." If it was judged that there was a positive effect for user searching through unique and more consistent information in the subject file, the cataloger was more likely to assign specific headings to a given subject category.

The second question on the checklist was whether this category was a focus of the postcard collection and of the larger library collection. That is, are there a significant number of images in this category, perhaps several images of each entity in the category, to keep organized? If so, this was a point in favor of assigning specific subject headings in that category. For example, the largest category of named entities is street names. Thirty-seven distinct named streets are depicted in the postcards, several streets featuring in as many as eleven different postcards. Specific street name headings were assigned because of the benefit perceived in linking directly to other images of the same street. Having only one image of a pier, however, the collection does not gain much in navigability by a specific name heading in this subject category. If the library owned other materials by or about the entities in that category, it was also a point in favor of specific heading assignment. For instance, because the library collection includes many histories of local churches, it was useful to assign specific headings for churches to integrate the postcard collection with the existing library collection.

Finally, were the names in this subject category actual proper names? Some names are clearly proper names (for example, St. Peter's Episcopal Church), but what about Old Tar River Bridge? Is that the name of a bridge, or simply a description of a bridge that goes over the Tar River? The "name" is capitalized on the postcard, but that does not necessarily signify a proper name. *Subject Cataloging Manual* H1334 instructs catalogers, "Do not formulate a heading for a named structure that consists solely of a generic term with a geographic qualifier unless there is evidence that this is also the *proper* name of the structure."[32] Being certain it was a proper name would probably require prohibitively extensive research into city records, and even then it may be difficult to be positive. *CCO* warns, "Catalogers should never use a specific term unless they have the research, documentation, or expertise to support that use."[33] This last item on the checklist shows how, in addition to considerations of effect on user retrieval, decisions about the treatment of particular subject categories must be guided by practical concerns regarding the amount of research involved in specific heading creation as well as a desire to avoid introducing misleading data into the repository.

When in doubt, the cataloger tended not to create specific headings because, should circumstances change (for instance, if the library should digitize a whole collection of images of university buildings), specific subject analysis could then be done on categories initially left generic. A vital complement to Younger's concept of utility is her assertion that cataloging of monographs should be viewed as more of a dynamic, iterative process than it traditionally has been. "Authority control," Younger writes, "requires continual evaluation of how a name fits into the larger context of the catalog."[34] This iterative process is perhaps more compatible with digital repository metadata creation, in which standards are more flexible and policies more subject to continual revision than they have been in traditional monograph cataloging.

As an attempt to meet the challenge of the high volume of authority work associated with this digital project, a checklist of considerations was used to select various subject categories for which to create specific name subject headings. Nevertheless, the volume of authority work produced by such a relatively small project was significant. The exact approach would probably not be scalable to larger projects, at least not without substantial increases in committed personnel. While the use made of the checklist may differ from project to project, the process of separating the images into generic subject categories and the checklist itself remain valuable tools in meeting the challenge of the high ratio of new headings to image descriptions that is typical for these types of projects.

Inefficiency Caused by Lack of Authority Data in the Repository Database

Though many new headings were created for named entities, no actual authority records were created because there was no structure to accommodate them in the repository database. While initially a time saver, this omission ultimately results in inefficiencies that constitute an additional challenge for the cataloger. A lack of structure for authority data is not atypical for metadata projects, where most of the emphasis is on resource description. In the non–MARC metadata context, "doing authority control" is usually used to mean merely "using controlled vocabulary." However, this is akin to doing authority work without the authority record. In traditional bibliographic databases with authority control, the data in an authority record includes not only the authorized form of the heading but also cross-references to variants of the name—in the case of corporate bodies to earlier and later names and in the case of headings established under subject conventions—to broader or narrower terms. It also includes additional information about the entity represented by the heading, useful when determining

the appropriateness of a heading to a given bibliographic record, and references to the sources of information used when establishing the heading. Lacking authority records, the library's repository lacks these data.

According to *FRAD*, the conceptual model developed to accompany the *Functional Requirements for Bibliographic Records* (*FRBR*), authority data have five functions, some aiding the cataloger and some the user.[35] Authority data should

1. document decisions for the cataloger;
2. serve as a reference tool for the cataloger creating more descriptions;
3. control the form of access points so that a user may find all relevant items under a certain access point;
4. support access to the bibliographic file by the user, e.g., through cross-references and additional information; and
5. link bibliographic and authority records for automatic bibliographic file maintenance.[36]

In the metadata implementation of the Eastern North Carolina Postcard Collection, only the third function listed above is fully executed. Though some workarounds partially fulfill the other four functions and mitigate the effect of the lack of an authority file, inefficiencies remain that could be eliminated by the introduction of an actual authority file. These inefficiencies are a major source of challenges for the user as well as the cataloger.

One example of a workaround addresses the functions of documenting cataloging decisions and serving as a reference tool for future cataloging. In the absence of MARC authority field 670, records in the repository include an additional note field that cites outside sources consulted in creating the description and access points for each image. For instance, in

the case of the image of the Lumina Pavilion, this note field contains a citation to the history of Wrightsville Beach consulted to confirm the name of the pavilion. This note is only used for citations of outside reference sources, not simply to justify the form of the name, as would be the case for MARC field 670. Though not visible to users, these data, which could be useful to a future cataloger adding records to the repository, will not be lost. Tying this information to the image description rather than to the named entity itself, as it would be in an authority record, is not ideal. The future cataloger would have to locate that particular image and its description to use the information in the note unless the note were repeated in each description with an access point for that named entity.

FRAD states that authority data also support access to the bibliographic file by the user through cross-references and additional information.[37] The lack of these data in digital repositories is a clear disadvantage to the user, who may not know the exact name of an entity for which to search. A portion of this problem has been addressed by adding headings for generic subject categories in the subject analysis for each image, hopefully helping more users find what they are looking for. The inability to navigate through cross-reference structures and broader and narrower trees is a detriment, as is the inability to view additional identifying information to help the user determine whether a named entity is actually the intent of their search. These shortcomings, however, are frequently found even in databases with authority files if they are implemented imperfectly.

Automatic maintenance of headings in the bibliographic file, the final *FRAD* function of authority data, is impossible without linked authority records, of course. Without the ability to effect a global change by changing one authority record, the cataloger

must manually update each affected bibliographic record if it is determined that a heading needs to change. There is no workaround for this. The change process can be made as easy as possible with the inclusion of a good search function that identifies all the records that need to be changed and with the ability to cut and paste the updated headings into each record, but nevertheless the process is inefficient and prone to human error.

There is no doubt that workarounds, when they are even a possibility, are responses to the limitations of the current system. From the cataloger's and presumably the user's perspectives, the addition of authority file functionality to the repository would be a welcome development, aiding information retrieval and making maintenance more efficient. Such functionality could be accomplished in a variety of ways, whether by the addition of an actual authority file to the database or through links to an external authority file. The latter may be preferable, since it would eliminate the need to maintain a separate authority file just for the repository. Regardless of how it is done, programming complexities would ensue. Whether the investment of time and effort would result in overall savings depends on projections for the future growth of digitization efforts at a particular institution. For a small project such as the Eastern North Carolina Postcard Collection, success in maintaining a certain level of consistency in subject analysis despite the lack of an authority file was due in major part to the concentration of subject heading assignment responsibility in one person. Success may not be scalable to larger projects, however. If digitization of local materials is a growth area for an individual library, new solutions will need to be found.

Fortunately, progress toward these new solutions is being made. While authority control has frequently been an afterthought to metadata

creation, some evidence exists that this is changing. One sees increasing interest in authority data in the metadata environment. Metadata Authority Description Schema (MADS) was developed as a counterpart to MODS by the LC.[38] Derived from the MARC 21 authority format, MADS records could satisfy the five functions of authority data listed in *FRAD*.[39] Although it has been available since 2005, MADS has not been widely implemented nor experience with it broadly addressed in the professional literature. Practical discussions of MADS have started to appear in blogs, however.[40] Other metadata communities besides the library cataloging community also have begun to recognize the importance of authority data. The archival community, for example, has developed Encoded Archival Context (EAC) to complement the more established Encoded Archival Description (EAD) by housing authoritative data about the creators of archival collections separately from but linked to the descriptions of the collections.[41] The implementation of EAC, according to Pitti, would not only enable easier collocation of materials with a particular provenance, but also facilitate the sharing of archival authority data across institutions, enable the expression of relationships between different creators, and take advantage of the unique source material held in archives to make this contextual information available.[42] Whether use of schemas such as MADS and EAC will become widespread remains to be seen, although it seems clear that their use has the potential to address the challenge of inefficiencies created by the lack of authority data in digitization project repositories.

Topics for Further Research

As is the case with much of the professional literature on metadata and their application to digitized collections during this time of intense and widespread development, this paper identifies more problems than it solves. These problems are in serious need of attention by the metadata community. They include the question of whether there is a continued need in LC practice to separate the Name Authority File and the Subject Authority File in an increasingly digital, metadata context. As noted above, this separation into two files with separate sets of rules for heading creation gives rise to much complexity in the creation of headings for named entities such as buildings and other structures, and now is the time to ask whether such complexity is worthwhile. Also worthy of investigation is the problem of how to apply authority control on a selective basis to larger and larger bodies of digitized materials. The applicability of Younger's utility principle to this problem remains to be tested. Also at issue is the best way for institutions to share the results of the time and effort that goes into the subject analysis of digital projects such as the Eastern North Carolina Postcard Collection. Institutions undertaking the digitization of local materials have widely varying resources, and one method of sharing (e.g., NACO) may not fit all. Nevertheless, the sharing of data is the only way to free it from local silos and thus increase the potential benefit in return for work expenses incurred. Lastly, but in the author's opinion most importantly, the problem of including authority data in digital repositories must be tackled. Implementations of schema such as MADS and EAC must be undertaken and reported upon, and pressure exerted on database vendors to incorporate such functionality, much as ILS vendors have been gradually pressured to include authority control functionality in their systems.

Conclusion

Authority control is a part of metadata creation for local digitization projects that has received insufficient attention. By examining the case of the Eastern North Carolina Postcard Collection (a small image collection), this paper discusses the particular challenges involved in the authority control of named entities used as subject headings for such projects. These challenges include (1) the complexity of work arising from the form and subject matter of the materials digitized, (2) the volume of work created by a high ratio of new authorized headings per bibliographic description, and (3) the inefficiency perpetuated by the lack of actual authority data in the repository database. How these challenges were addressed should be of interest to the many institutions undertaking similar projects if they are concerned with the visibility and usability of their digitized collections. The use of widely applied vocabularies and their rules such as LCSH and LNCAF in digital collections metadata enables the closer integration of digitized collections into the traditional collections of the library, collections whose metadata are privileged to reside in the library catalog. By ensuring name consistency, the cataloger is creating the potential for heading links across discovery tools and setting the stage for the implementation of a federated search function that would enable users to discover traditional library materials as well as digital projects in the same search. Authority control is a large part of what makes the difference between low- and high-quality metadata, and high-quality metadata improve the usability of digital repositories. By taking the time to determine which structure a particular image depicts, and by differentiating it from other similar structures, libraries avoid pushing that challenge off onto the user, as would be the case if they were to rely on keyword access or a generic subject heading only. The added value that authority control brings to traditional bibliographic databases should also, and perhaps even more urgently given the uniqueness of the subject

matter, be applied in local digitization projects. Establishing best practices for doing so is a major problem facing catalogers and digitization librarians in the near term.

References

1. Michael Gorman, "Authority Control in the Context of Bibliographic Control in the Electronic Environment," *Cataloging & Classification Quarterly* 38, no. 3/4 (2004): 11–21.
2. Sherry L. Vellucci, "Metadata and Authority Control," *Library Resources & Technical Services* 44, no. 1 (2000): 33–43.
3. Ibid., 40.
4. Murtha Baca, "Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information," *Cataloging & Classification Quarterly* 36, no. 3/4 (2003): 47–55.
5. Ibid., 52.
6. Ibid., 52-53.
7. *Anglo-American Cataloguing Rules,* 2nd ed., 2002 revision, 2005 update (Chicago: ALA; Ottawa: Canadian Library Association; London: Chartered Institute of Library and Information Professionals, 2005): 24.1A.
8. Suzanne R. Graham and Diane DeCesare Ross, "Metadata and Authority Control in the *Civil Rights in Mississippi Digital Archive*," *Journal of Internet Cataloging* 6, no. 1 (2003): 38.
9. Ibid., 39.
10. Linda Cuccurullo, "Kicked a Heading Lately? The Challenge of Establishing Headings for Buildings and Other Structures," *Art Documentation: Bulletin of the Art Libraries Society of North America* 25, no. 2 (2006): 56–60.
11. Murtha Baca et al., *Cataloging Cultural Objects: A Guide to Describing Cultural Works and their Images* (Chicago: ALA, 2006).
12. Barbara B. Tillett, "Authority Control: State of the Art and New Perspectives," *Cataloging & Classification Quarterly* 38, no. 3/4 (2004): 24.
13. José Borbinha, "Authority Control in the World of Metadata," *Cataloging & Classification Quarterly* 38, no. 3/4 (2004): 114.
14. Jennifer Younger, "After Cutter: Authority Control in the Twenty-First Century," *Library Resources & Technical Services* 132, no. 2 (1995): 133–41.
15. Ibid., 137.
16. IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), *Functional Requirements for Authority Data: A Conceptual Model* (The Hague: International Federation of Library Associations, 2007), www.ifla.org/VII/d4/FRANAR-ConceptualModel-2ndReview.pdf (accessed Sept. 11, 2008).
17. Ibid., 56–59.
18. Karen Coyle and Diane Hillman, "Resource Description and Access (RDA): Cataloging Rules for the 20th Century," *D-Lib Magazine* 13, no. 1 (2007), www.dlib.org/dlib/january07/coyle/01coyle.html (accessed Sept. 11, 2008).
19. Library of Congress Cataloging Policy and Support Office, "H180: Assigning and Constructing Subject Headings," in *Subject Cataloging Manual. Subject Headings*, 5th ed., 4 vols., loose-leaf (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1996–).
20. Baca et al., *Cataloging Cultural Objects*, 8.
21. Ibid., 9.
22. Arlene G. Taylor, "Metadata: Access and Authority Control," in *Organization of Information,* 2nd ed. (Westport, Conn.: Libraries Unlimited, 2004): 204.
23. Library of Congress Cataloging Policy and Support Office, "H405: Establishing Certain Entities in the Name or Subject Authority File," in *Subject Cataloging Manual. Subject Headings*.
24. Sherman Clarke, "FRBR and Buildings" (report to Subject Analysis Committee Task Force on Named Buildings, 2004), http://artcataloging.net/ala/mw04/frbrbldg.html (accessed Nov. 14, 2008).
25. *Anglo-American Cataloguing Rules*, 24.1A.
26. Library of Congress Cataloging Policy and Support Office, "H1334: Buildings and Other Structures," in *Subject Cataloging Manual. Subject Headings*.
27. Ray McAllister, *Wrightsville Beach: The Luminous Island* (Winston-Salem: John F. Blair, 2007).
28. Ibid., 49.
29. Cuccurullo, "Kicked a Heading Lately?" 58.
30. Younger, "After Cutter," 133.
31. Ibid., 137.
32. Library of Congress Cataloging Policy and Support Office, "H1334: Buildings and Other Structures."
33. Baca et al., *Cataloging Cultural Objects*, 9.
34. Younger, "After Cutter," 138.
35. IFLA Study Group on Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report* (Munich: G. K. Sauer, 1998), www.ifla.org/VII/s13/frbr/frbr.htm (accessed Sept. 12, 2008).
36. IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), *Functional Requirements for Authority Data*, 58–59.
37. Ibid., 59.
38. Library of Congress, MADS: Metadata Authority Description Schema Official Web Site, www.loc.gov/standards/mads (accessed Nov. 18, 2008).
39. Rebecca Guenther, "MADS," *Computers in Libraries* 27, no. 4 (2007): 14.
40. Winona Salesky, "Authority Control: MODS & MADS," online posting, the DIL, Oct. 30, 2007, http://thedil.wordpress.com/2007/10/30/authority-control-mods-mads (accessed Sept. 12, 2008).
41. Jean Dryden, "From Authority Control to Context Control," *Journal of Archival Organization* 5, no. 1 (2007): 1–13.
42. Daniel V. Pitti, "Creator Description: Encoded Archival Context," *Cataloging & Classification Quarterly* 38, no. 3/4 (2004): 201–26.

Appendix. Table of Generic Category Treatment

Generic Category	Establishment Conventions	Number of Occurrences	Treatment Decision	Reasoning
Banks	Name	9	Create specific	Have proper names.
[University]—Buildings	Subject	5	Do not create specific	Many not sure if proper names. Have very few of these.
Bridges	Subject	20	Do not create specific	Have generic names.
Boats	Subject	2	Do not create specific	Have very few of these and no other materials about.
Bodies of water (rivers, lakes, etc.)	Subject	26	Create specific	Have proper names.
Cemeteries	Name	3	Create specific	Own other materials by/about
Churches	Name	36	Create specific	Own other materials by/about.
Corporations	Name	17	Create specific	Have proper names.
Country clubs	Name	2	Create specific	Have proper names.
Courthouses	Subject	26	Do not create specific	Have generic names. Creating adds little unique information to subject file.
Grade schools	Name	18	Do not create specific	Have generic names (school plus location). Do not own other materials by/about.
High schools	Name	4	Create specific	Own other materials by/about (e.g., yearbooks).
Historic homes	Subject	13	Create specific	Have proper names.
Hospitals and institutions	Name	12	Create specific	Own other things by/about.
Hotels and motels	Name	34	Create specific	Have proper names. Own other things by/about.
Islands	Subject	1	Create specific	Have proper names.
Libraries	Name	1	Create specific	Have proper names.
Lighthouses	Subject	3	Create specific	Have proper names. Have other things by/about.
Military bases	Name	10	Create specific	All were already established in LCNAF.
Monuments	Subject	18	Do not create specific	Have generic names. Creating adds little unique information to subject file.
Parks	Subject	2	Create specific	Unclear whether proper names. Do not own other materials by/about.
Pavilions	Subject	5	Create specific	If it has a proper name.
Personal and family names	Name	24	Create specific	Have proper names.
Piers and docks	Subject	1	Do not create specific	Unclear whether proper names. Do not own other materials by/about.
Plantations	Subject	1	Create specific	Have proper names. Own other materials by/about.
Public buildings	Subject	4	Do not create specific	Have generic names. Creating adds little unique information to subject file.
Railroad stations	Subject	8	Do not create specific	Have generic names (name of railroad company + depot)
Restaurants	Name	4	Create specific	Have proper names.
Ships	Subject	2	Create specific	Were established already.
Singing groups	Name	1	Create specific	Have proper names.
Stores (retail)	Name	8	Do not create specific	Difficult to determine actual proper name.
Streets	Subject	75	Create specific	Have proper names.
Theaters	Subject	2	Create specific	Have proper names.
Universities and colleges	Name	25	Create specific	Most were already established. Have proper names. Own other materials by/about.