

Geoff Ross
Spring 2006

Independent Study Paper

Research Problem: how can the DCC project make the content in the IMLS National Leadership Grant collections more accessible to teachers of K-12 students?

I. Contexts

The intended users of the DCC aggregated collections are Educators of K-12 students. However, studies of this user group expose a mismatch between what the DCC collections offer, and what K-12 teachers need: most K-12 educators are looking for ready-to-use classroom materials (Fitzgerald et al. 26) like lesson plans and activities. The DCC currently has about thirty collections that already offer lesson plans, and these are probably ready to be mobilized, either through the DCC's own interface, or through an existing service like GEM. For the most part, however, the National Leadership Grant collections do not appear to have been packaged for use in K-12 classrooms, and yet they contain material that would likely be valuable in such settings. Can these collections support K-12 learning without the development of further educational apparatus? At the very least, teachers will have to develop some sort of lesson "plan" or assignment to direct student use of these collections. A collection should be an "information seeking context" (Lee 1106), but students will require guidance and direction as to what they are seeking. This problem could be framed in terms of both information retrieval, and information use: users of the DCC gateway want to find information, and use that information to compose lesson plans. Definitions of "digital libraries" increasingly emphasize the need to support use of content (Borgman 234). Metadata must be enhanced to facilitate both information retrieval and use.

The information retrieval problem posed by these collections begins as a cataloging problem, and is conditioned by the objects in those collections, most of which comprise "images of artifacts" (Palmer and Kuntson 460). Images have long been a special problem in librarianship (Anderson and Pérez-Carballo 141, and Svenonius 600). Corinne Jörgensen argues that "the major intellectual problem involved in access to images is the question of how to index them" (162). My own experiences working with and thinking about the objects in these collections certainly bear that argument out. Images are overloaded with meanings that resist codification, which might explain why so many attempts to describe images lexically have proven unsatisfactory (Jörgensen 162). In dealing with images, the DCC project must negotiate the collision of two distinct information cultures: museum informatics and librarianship. Museums tend to be document centered--the thing itself is the object of interest--whereas libraries tend to be information centered. Librarianship in general, and library cataloging in particular, take almost as a matter of faith Lubetzky's distinction between Book and Work, which has since been reproduced and expanded in the FRBR Group 1 entity-relationship model (Tillet 11). Lubetzky writes, "the book (i.e., the material record) and the work (i.e., the intellectual product embodied in it) are not coterminus" (Lubetzky 99). Pace Marshall McLuhan (6), Lubetzky argues that the medium is not the message at all.

Unfortunately, this information model fits only a certain class of information-bearing objects, what Jerome McGann calls scholarly or scientific texts, where the medium itself is only "vehicular": "for the scientist or scholar, the media of expression are primarily conceptual utilities, means rather than ends" (54). A scholarly Work can be realized in any number of Expressions--English or German--and a scholarly Expression

can be embodied in any number of Manifestations--the print edition of JASIS or a web-page produced by an e-content aggregator--without undermining information quality. These are the kinds of information objects that librarianship is especially well adapted to organizing. Librarianship is less well-suited to organizing objects that are artistic or artifactual, where the media are "incarnational" (McGann 54). With these kinds of information-bearing objects, one cannot easily separate the information from its material incarnation. Many of the objects in the DCC aggregated collection are of precisely this incarnational nature.

A further problem with images is that whatever meaning they can be said to possess does not inhere solely within the image. Rather an image is in part about the person who created it, and the place and time it was created: "an image showed how [. . .] the subject had once been seen by other people [. . .] the specific vision of the image-maker was also recognized as part of the record. An image became a record of how X had seen Y. This was the result of an increasing consciousness of individuality, accompanying an increasing awareness of history" (Berger 10). For example, the meaning of any object in the Charles W. Cushman collection is forever bound up with Charles W. Cushman, and his place and time in history. The farther any image moves from its originating context, the more difficult it is to recover its meanings: "An image is a sight which has been recreated or reproduced. It is an appearance or a set of appearances, which has been detached from the place and time in which it made its first appearance [. . .] Every image embodies a way of seeing" (Berger 9-10). Berger's comments seem especially relevant since the each image accessed through the DCC collections have undergone several steps of removed from its historical apparition, and as

it is reharvested and recombined with documents from other collections, this process of detachment becomes more and more pronounced. Reproduction "destroys the uniqueness" of an object, "its meaning multiplies and fragments into many meanings" (Berger 19).

As an example of this fragmenting of meaning, take item P10757 from the Cushman Collection: "River steam boat Mark Twain Disneyland." Disneyland reproduces this Mississippi River steamboat and puts it in a California theme park; Cushman photographs this reproduction and puts it in a private collection; Indiana University reproduces the entire collection which is then harvested by the DCC. What finally, is that steamboat about? Is it about the Mississippi River of Mark Twain's youth, or about how that place and time has been recreated in the popular imagination? How does Cushman himself figure into these constructions of meaning? Because subject cataloging is in part about identifying the "aboutness" of an information-bearing object, and because images are potentially about so much, they present real problems to any system trying to support exploratory information discovery. Berger's characterization of the image dovetails nicely with Jörgensen's own observations on strategies of providing access to images: she notes that indexing of images might benefit from some of the strategies developed for cataloging fiction (168). Both Berger and Jörgensen zero in on a suppressed or obscured story that needs somehow to be elucidated before an image's meaning can be known. An understanding of the problems posed could prove useful in informing decisions about where the DCC should focus its enrichment efforts. For my own study, determining what age an object is appropriate for, or what school-subject it could be used with, is not more straightforward than determining what an image is about.

II. Metadata

During my independent study, I investigated how the metadata might be enhanced, within the organizational constraints of the project, to make it more useful to educators. To get an idea how others have provided access to educational content, I looked at similar services available to teachers: the Gateway to Educational Materials (GEM), United Streaming, and Marco Polo. GEM and Marco Polo are free services, while United Streaming is a commercial content provider, owned by the Discovery Channel. Two key access points are common to all these services: school subject and grade level. DCC already offers access by school subject, but access by grade level remains an elusive goal, and the reason is that unlike GEM, Marco Polo, and United Streaming, the DCC is not providing access to lesson plans, but to primary resources. Many of the lesson plans in GEM, Marco Polo, and United Streaming have been developed by teachers to work with state learning standards. Where a lesson plan conforms to a specific learning standard, that information can be made available in the metadata. In these systems, a teacher can literally plug in facts about his information need--school subject, grade level, and state-specific learning standard--and retrieve a set of highly relevant material. Not surprisingly, the commercial system, United Streaming, provides the most comprehensive level of access: every available state learning standard in the United States and Canada is mapped to content. The baseline functionality, however, was the ability to retrieve materials appropriate to a specific grade level.

I attempted to provide a similar grade level access point to the collection-level records in the DCC. I ruled out the item level records from the start, not just because of the sheer quantity of work that would involve, but because so many of the objects in

these collections are images, or photographic reproductions of art objects and artifacts, or realia: for the reasons similar to those described above, assigning grade levels to such material would be extremely difficult. As I quickly discovered, however, assigning grade levels to entire collections is not much easier. Only where lesson plans accompanied the collections was I able with confidence to assign a grade level value. Where no classroom-ready material was available, I found it difficult to judge. This problem was in part due to my own lack of background in K-12 education. In short: the problem I anticipated in the item-level repository reproduced itself in the collection-level catalog. The age-appropriateness of a set of photographs probably depends on how they are used, which brought me again back to the problem discussed earlier: I was trying to adapt strategies from systems that provide access to lesson plans, to a system that provides access largely to primary resources. However, the heterogeneous nature of the NLG collections meant that some of those collections did have lesson plans and classroom activities available, so as a starting point I decided to locate these collections and provide grade level access.

At Tim's suggestion, I looked at the DCMI Education Working Group to see how they recommended using Qualified Dublin Core to improve access to educational content. The Education Working Group has developed an Education Application Profile, which includes an EducationLevel element refinement of Audience. Following the Working Group's recommendations and the DC-Education Application Profile, I added <dcterms:educationLevel> elements to 44 collection level records. An xsi:type="imlsdcc:educationLevel" attribute on the <educationLevel> element ties the value of the field to a controlled vocabulary. As I could find no recommendations for

controlled vocabulary on the Education Working Group site, I used the GEM Level Element Controlled Vocabulary. This vocabulary might be more fine-grained than is really necessary. However, the IMLS DCC project has on its timeline a GEM Exchange for this spring. It should be easier to ingest records into the GEM test region if the DCC records use the same controlled vocabulary. Also, it would in theory be easier to translate the elements over to a blunter vocabulary, than it would be to go in the other direction. A potential problem with the GEM vocabulary is that, in its expressiveness, it has the potential to suggest a misleadingly precise level of subject analysis. A blunter vocabulary might make the improvisational nature of the subject cataloging more apparent to potential users. I modified the DCC's XML Schema to include the new controlled vocabulary:

<http://leep.lis.uiuc.edu/publish/gtross/Fellowship/dcterms.xsd>
<http://leep.lis.uiuc.edu/publish/gtross/Fellowship/dcprofile.xsd>
<http://leep.lis.uiuc.edu/publish/gtross/Fellowship/imlsdccprofile.xsd>
<http://leep.lis.uiuc.edu/publish/gtross/Fellowship/imlsdcctypes.xsd>
http://leep.lis.uiuc.edu/publish/gtross/Fellowship/IMLSDCCVocab_educationLevel.xsd

Tim and Tom then created for me a clone of the IMLS database, in which I entered the new field and its values. They also helped me to create a test interface, through which this new access point could be combined with the already existing browse and search structures.

While working with National Leadership Grant collections, I began to wonder what one might learn from them by conducting collections assessment and evaluation. Although such an evaluation was beyond the scope of my work, I did reflect on what insights might be gained from theorizing the NLG collections. I began by thinking that many resembled thematic research collections, but the more I worked with them, the less

they seemed to fit the characteristics of thematic research collections. For example, while many achieve an undeniable density of content, there is often little evidence of continuing and systematic collection development (Palmer). Many of the NLG collections are digitizations of extant special collections. There are other characterizations of digital collections, for example that developed by the NISO Framework Advisory Group. The NISO Framework highlights what has been a big obstacle in implementing this federated collection: "A good collection fits into the larger context of significant related national and international library initiatives. For example, collections of content useful to education in science, math, and/or engineering should be usable in the NSF-funded National Science Digital Library" (10). Palmer and Knutson found that the NLG projects tended to fall short of this goal, choosing instead to base decisions on "immediate, local production requirements" (461). It appears that many of the NLG projects are enacting a collections theory at odds with theories favored in the library community. How can these differences be reconciled?

A clear understanding of library collections in general, and digital collections in particular, might help frame an approach to the problem, especially if librarians are to make the most effective intervention in providing access to this content. Do we as librarians consider the National Leadership Grant collections to be (digital) library collections? Will our established methods of organizing and providing access to digital library collections work with the NLG collections? The answer to that question will partially determine the nature of our response to the challenge they pose. These questions are value-neutral with regard to each NLG collection itself: whether it fits our idea of a library collection or not has nothing to do with its success or failure. Librarians

must deal with each collection on its own terms. That said, librarians approach different classes of information differently: archives, special collections, periodicals, monographs, and monographic series all call for different methods of selection, organization, preservation, and access.

Librarianship has evolved methods for acquiring, organizing, providing access to, and preserving highly synthesized discursive objects, the exemplars being monographs and serials. Both monographs and serials are typically larger chunks of information than one will find in the NLG collections, but in most cases they are much smaller than entire digital collections. The DCC has found that "[i]ntermediate levels of descriptive access [between item level and collection level] might enable or enhance other functionalities of federated registries and repositories" (Cole 2005, 2). Furthermore, as argued above, many of the objects in traditional library collections are "vehicular" (McGann 59) in relationship to the information they carry. In contrast, objects in the NLG collections are often incarnational or, posing an even bigger challenge, not really discursive at all: what, for example, is a maternity dress about? What argument is it making? In cataloging realia, one must be careful not to confuse what an object is with what an object is about. In short, can a maternity dress really be treated as a document? One might usefully invoke here Susan Briet's well-known argument that something is a document only when submitted as evidence in discourse (qtd. in Buckland 806). Alternatively, one might think of the distinction between information and knowledge (Gorman 23): while the objects in these collections are often information-bearing objects, the aim of librarianship is to organize the graphic records of human knowledge (Shera qtd. in Budd 94). These distinctions might seem like quibbling, and yet the project's work with the GEM initiative

seems to be an attempt to address them, to think of the items in these collections as discourse, knowledge, or Britian documents. The GEM subject headings situate the objects within a school-subject context. Similarly, many of the NLG collections create discourse by combining information objects with mini-essays, biographies, guided-browsing, and contextualizing facts. For example, "The American Missionary Association and the Promise of a Multi-Cultural America" self-consciously tells a story, makes an argument, in a way that library collection developers generally try to avoid.

The item-level metadata repository creates the opportunity for end users to tell these kinds of stories. Combined with the new DLF Aquifer Asset-Actions suite of tools, the item-level repository allows users, in this case teachers, to take a piece of information out of its native collection and use it in concert with other objects to convey or create knowledge. Interestingly, teachers are potentially both users and authors, or knowledge-makers, but only if there exists a mechanism for capturing their use of the objects in the repository: "users can play an important role in the aggregator's quality improvement activities. Adding user feedback/error reporting capabilities to the system can be an effective and efficient tool in increasing the collection's quality inexpensively" (Stvilia et al. 2004 p. 121). The project's grant proposal points towards this potential: "Teachers suggested the inclusion of an interface component for submitting commentary on objects" (Cole 2002). Such an interface creates the opportunity for a kind of feedback loop, where the object's rhetorical deployment returns to the repository as a publication in its own right. Furthermore, these kinds of "comments" come close to Jørgensen's idea of using narratives as access points. An invented example: I used this image of a maternity dress in a unit on *The House of Mirth* where we discussed articulations of feminity through

clothing. In this case, the maternity dress was submitted as evidence in an argument on early Twentieth Century femininity. A teacher might submit this information to the repository in the form of syllabi and related course materials, à la ERIC's Resources in Education subfile, or simply as a user comment. Such an approach borrows from the citation index, where each use of a digital object works like a citation, and might lay the groundwork for a system that could make semantic inferences on those citations.

Whatever the approach, the value is that the object's "meaning" is tied to its uses, and facts about those uses might create valuable access points for supporting exploratory discovery.

Another alternative to viewing the NLG collections as library collections might be to approach them as unified discursive objects. Many of these collections more closely resemble anthologies, albums, or books of collected essays than they do library collections. Some of them even seem argument-driven. Librarians might deal with the NLG collections as they deal with monographic or continuing publications. The challenge posed at the beginning of this paper remains, however: how do we make these complex discursive objects accessible to K-12 educators? If these collections are treated as publications, as discursive objects, then catalogers should be able to evaluate age-appropriateness. This is something librarians do every day in selecting and organizing materials for access. If these collections resemble publications, then it should be possible to assign grade levels to them. An indicator of age-appropriateness might be the reading level of collections' introductory and contextualizing material. The project could use Gunning Fog and Flesch Kincaid scores to gauge these reading levels (Renear). For example, the introductory material for Mark Twain's *Mississippi* has a Flesch Kincaid

Grade Level score of 22--that is to say, the 22d grade--and a Gunning Fog Index score of 30, where scores over 22 are considered to indicate a post-graduate reading level. The Themes Overview on the same site scored 15 for Flesch Kincaid Grade Level, and 24 on the Gunning Fog Index--better than the introductory material, but still far in advance of K-12 students. If many of the NLG collections score that high, then we gain nothing by viewing the collections as publications. NLG collections with Flesch-Kincaid scores above 12 are probably not ready for use in K-12 classrooms. Therefore, the DCC project might need to narrow down its intended user group even further: those K-12 educators who want or need to create lesson plans that use the DCC content in a supporting role. In narrowing down its intended user group, the project will actually end up broadening its scope of services, because the collection would need to support, even if only indirectly, the authoring of lesson plans.

What I learned while working with these collections is that they are big, difficult to describe, and challenging to use. It's tempting to view the NLG collections' lack of built-in support for K-12 education as a weakness, and perhaps it is. It might also, however, be a strength. From this perspective, the educational support already provided becomes a barrier to learning. Julia Flanders, who directs the Women Writers Project, has written about her experiences in trying to provide access to large, complicated collections of information resources. Her experiences provide a valuable counter-text to the "the academic masterplot" (Bloom 130), where students are successfully guided by expert readers to discover correct readings of textual objects. Flanders notes that scholarly editions tend to "encrust" texts with the insights and knowledge of expert readers (Flanders 50). She notes that critical apparatus serves to "elicit" the correct

reading from students, rather than encouraging students to make their own discoveries, to construct their own meanings.

These problems are in part systemic to the way education is practiced: "the American educational system, with its emphasis on syllabi, is designed to avoid [the challenge of confronting an unknown body of knowledge]" (Flanders 50). Flanders notes from studying users of her own collection that "[c]oming to the [Women Writers Project] collection through the medium of a syllabus, students encounter it only as a way of accessing a particular set of prescribed readings, not as a space that itself poses any kind of intellectual challenge [. . .] students navigate the collection as if it were fully known to them" (50). From this perspective, federated collections like the DCC create the opportunity for students to confront the unknown past, and the salient question is, how do teachers want to make it known? How do they want to facilitate that act of knowing? This is a fraught question, especially in the culture of standards-based teaching. Librarians have an obligation to serve our user communities, regardless of our theoretical stance towards those communities' practices. We might believe that standards-based teaching to be less-than-ideal, but our collections still need to serve teachers whose daily practice must conform to standards.

At the same time, standards-based learning need not be seen as the final word on education. The process of using and developing digital collections remains a in part an unexplored field. Flanders argues, "Anyone with prior knowledge of the existence and general location of a book can find it in even the most arcanelly organized library; the test of a collection's design is what happens when its user comes to it in a state of ignorance" (50). Within the item-level repository, the DCC creates the opportunity for students to

approach these objects "in a state of ignorance"--hardly an obvious advantage. Flanders's argument suggests that we don't fully know what kind of "information seeking context" (Lee) is best suited to supporting learning. It could be that my attempts to categorize these collections into appropriate grade levels has been counterproductive, has served to foreclose on the real learning potentials of these collections. On the other hand, studies of our users strongly suggest that pursuing that potential comes at a cost in time which most teachers can ill-afford to pay.

A library exploits economies of indirection and abstraction (Renear) in order to provide richer, more fine-grained access to the information in its collections. Users often do not begin a search by going straight to the book shelves: they use indexes, bibliographies, catalogs, and other reference sources to obtain guided, subject-specific access to the collections. In fact, research libraries have long been unable to provide unified access to the objects they make accessible: circulating collections in departmental libraries, stacks, and remote storage; rare book collections; special collections; archives; microform collections; periodical collections; shared collections; and reference collections are all distributed and hybrid. When one considers the heterogeneous and distributed nature of research libraries, the idea that "hybridity" is new and unique to electronic library collections (Manoff 857) begins to seem a little off target. The DCC fits into this tradition of providing indirect access to content: it brings together information about resources that might be relevant to a specific user group, in this case educators of K-12 teachers.

Works Cited

- Anderson, James D. and José Pérez-Carballo. *Information Retrieval Design: Principles and Options for Information Description, Organization, Display, and Access in Information Retrieval Databases, Digital Libraries, Catalogs, and Indexes*. East Brunswick, N.J.: University Publishing, 2005.
- Berger, John. *Ways of Seeing*. New York: Penguin, 1973.
- Bloom, Lynn Z. "Subverting the Academic Masterplot." *Composition Studies as a Creative Art*. Logan, Utah: Utah State University Press, 1998. 130-42.
- Borgman, Christine. "What Are Digital Libraries? Competing Visions." *Information Processing and Management*. 35.3 (1999): 227-43.
- Buckland, Michael. "What is a Document?" *Journal of the American Society for Information Science*. 48.9(1997): 804-809.
- Budd, John M. "Jesse Shera, Social Epistemology and Praxis." *Social Epistemology*. 16.1 (2002): 93-98.
- Cole, Timothy W., Principal Investigator. "Proposal for an IMLS Collection Registry and Metadata Repository." Champaign: UIUC, 2002.
- . "Proposal to Extend IMLS Collection Registry and Metadata Repository Project." Champaign: UIUC, 2005.
- Fitzgerald, Mary Ann, Robert Maribe Branch, Gita Williams, and Vicki Lovin. *The Gateway to Educational Materials: An Evaluation Study*. GEM: 2000.
- Flanders, Julia. "Learning, Reading, and the Problem of Scale: Using Women Writers Online." *Pedagogy*. 2.1 (2002): 49-59.
- Gorman, Michael. *The Enduring Library: Technology, Tradition, and the Quest for*

- Balance*. Chicago: American Library Association, 2003.
- Jørgensen, Corinne. "Attributes of Images in Describing Tasks." *Information Processing and Management*. 34.2-3 (1998): 161-74.
- Lee, Hur-Li. "What is a Collection?" *Journal of the American Society for Information Science*. 51.12 (2000): 1106-13.
- Lubetzky, Seymour. *Principles of Cataloging*. Los Angeles: University of California, 1969.
- Manoff, Marlene. "Hybridity, Mutability, Multiplicity: Theorizing Electronic Library Collections." *Library Trends*. 48.4 (2000): 857-76.
- McGann, Jerome. *Radiant Textuality: Literature after the World Wide Web*. New York: Palgrave MacMillan, 2001.
- McLuhan, Marshall. *Understanding Media: The Extensions of Man*. New York: Signet Books, 1966.
- NISO Framework Advisory Group. *A Framework of Guidance for Building Good Digital Collections*. 2nd ed. Bethesda, MD.: NISO, 2004.
- Palmer, Carole. "Thematic Research Collections." *A Companion to Digital Humanities*. Eds. Susan Schreibman, Ray Siemens, and John Unsworth. Malden, Mass.: Blackwell, 2004.
- Palmer, Carole L. and Ellen M. Knutson. "Metadata Practices and Implications for Federated Collections." *Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology*. Eds. Linda Schamber and Carol L. Barry. Medford, N.J.: Information Today, 2004. 456-62.
- Renear, Allen. Personal interview. 15 April, 2006. Champaign: GSLIS.

Stvilia, Besiki, Les Gasser, Michael Twidale, Sarah L. Shreeves, and Timothy W. Cole.

2004. "Metadata Quality for Federated Collections." *Proceedings of ICIQ04 - 9th International Conference on Information Quality*. Cambridge, MA : 111-125.

Svenonius, Elaine. "Access to Nonbook Materials: The Limits of Subject Indexing for Visual and Aural Languages." *Journal of the American Society for Information Science*. 45.8 (1994): 600-06.

Tillet, Barbara B. "FRBR (Functional Requirements for Bibliographic Records)." *Technicalities*. 23.5 (2003): 1, 11-13.