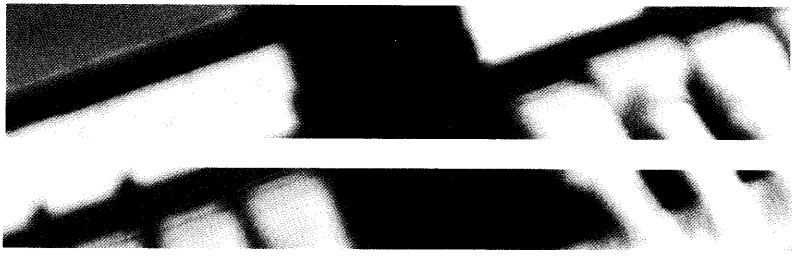
8-26-2004

"Introduction" to *Metadata in Practice*

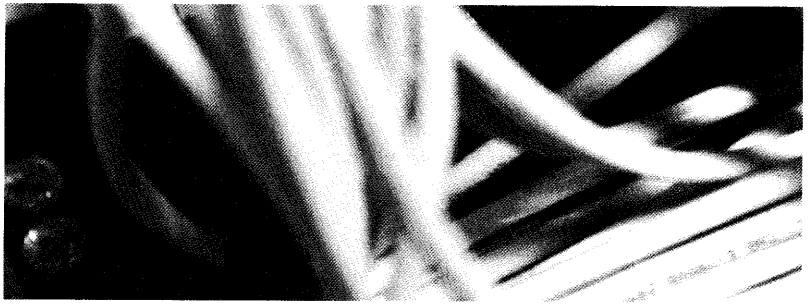
Diane I. Hillmann
Cornell University

Elaine L. Westbrooks
University of Nebraska-Lincoln, elainelw@email.unc.edu



Metadata *in* Practice

Diane I. Hillmann Elaine L. Westbrooks



American Library Association Chicago 2004

ACKNOWLEDGMENTS

THIS BOOK HAS been a group effort from the beginning, with much support offered to the coeditors every step of the way. Diane would like to thank in particular her colleagues at Cornell Information Science, whose assistance when she broke her arm in February 2003 was above and beyond the call of duty. Even when Diane was again possessed of two working arms, her colleagues' support for this project was generous and unconditional. Elaine would like to express her gratitude to her colleagues at Cornell University's Albert R. Mann Library, particularly Janet McCue and William Kara. A special thanks should be extended to Brian Lowe for his editing and formatting contributions.

INTRODUCTION

IT IS BOTH an exciting and frustrating time to be working in the world of metadata. Exciting because so many new communities are discovering the usefulness of metadata at the same time as librarians seriously consider the limitations of our traditional notions of the functions of libraries. New metadata formats seem to erupt like dandelions on a spring lawn, each seeking to bring together new communities with genuine needs to organize their important information.

For librarians or project managers who attempt to understand this world enough to plan a project implementation with a metadata component, the frustrations are also considerable. Although a library or cataloging background can be an asset when approaching metadata issues, to a traditional librarian the current metadata environment seems like the Wild West as seen from the point of view of a Boston Brahmin—very messy, and with armed cowboys behind every rock.

In such environments, prudent librarians review the literature. Unfortunately, information on the metadata context of relevant projects is sometimes difficult to find; and when relevant information is found, it rarely contains the detail that a planner desires. In addition, most of the research literature about digital libraries is not published in journals familiar to librarians; rather, it is scattered in digital library and computer science conference proceedings or journals. Consequently, taking advantage of the experience of others can be daunting. Those planners looking for the latest ideas in important areas of implementation have an even more difficult time. Developments are constantly in flux, and without active participation, it is a challenge to discover what is still relevant among the existing documentation.

Planning for metadata implementation is even more confusing, of course, for those without the benefit of a traditional cataloging background.

Determining what the options are for a new project, how to ensure the "interoperability" everyone seems to desire, and attempting to choose options with some promise of stability seem impossible. None of the emerging standards seems quite stable enough, there is little documentation that seems trustworthy, and each expert has a different opinion.

Even those metadata standards with the most promise have been slow to provide guidance directly to implementers. Element names, labels, and definitions may well be available, but what is often lacking is an experienced body of implementers to provide the documentation on what belongs inside the elements. Even creating simple metadata can be more difficult than new implementers might imagine, and richer metadata brings even more complications with its promise of improved discovery.

This situation is made even more difficult by the increasing requirements for metadata sharing, particularly since the emergence of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Although OAI implementation is touted as simple, requiring very little technical background, the technical skills necessary are beyond those of any but the most technically adept librarian. Since funding agencies in particular are determined to ensure that investments in data creation and maintenance will survive beyond the next grant, the project planner is on the horns of a dilemma. What's a conscientious planner to do to ensure that their project plan includes provisions for adequate technical support over time, in the face of such information gaps?

Over the past eight years, we have spent a great deal of time working in this new world of emerging metadata standards. Diane Hillmann's work with the National Science Digital Library, and as editor of the Using Dublin Core guidelines and administrator of the new AskDCMI service, has brought many questions from implementers her way. Many who ask for help are close to panic. "Where do I start?" is a common beginning. Others have made it through some initial research and decision-making but have gone aground on the details. Sometimes it is a question of determining where to find out what other, similar projects have done, perhaps regretted, and most likely redone. Perhaps the problem is determining what the current, standard manner of doing things is, when none of the documentation available uses the same terminology. Unlike the traditional library world, where there are well-trodden paths toward acknowledged competence, this new world has few maps. Many implementers find themselves working in isolation, feeling ill-prepared for the task they have

taken on. This book was conceived to assist these wanderers in the wilderness, to provide both background and signposts for the journey ahead. This book is not a guide to the options available to implementers; it does not provide definitions or advise on choices of metadata format, and not every metadata standard in use today is covered. For such a survey, Priscilla Caplan's *Metadata Fundamentals for All Librarians* is surely the best source.

Metadata in Practice is divided into two parts. Part 1, "Project-Based Implementations," brings together the work of a number of significant projects. Because so much of the interesting work being done in metadata implementations is focused in specialized communities, we have attempted to cover a broad range of communities and metadata formats in this book. But aside from orienting our contributors to their target audience, perhaps the most important question we have asked them is: "What would you have done differently, knowing what you know now?" Their answers provide much food for thought.

Two early projects begin part 1. Stuart Sutton describes the work of Gateway to Educational Materials, one of the first and still one of the most influential projects gathering educational materials for teachers. Another early project, Heritage Colorado, described by Liz Bishoff and Elizabeth Meagher, is particularly significant for its statewide collaboration between libraries and museums. Museums and archives have been major players in the metadata movement, with many issues quite distinct from libraries. Angela Spinazzè describes some of these issues and the efforts of the museum community.

Three projects centered on gathering materials together on specific campuses follow. First, Robin Wendler describes Harvard University's campuswide image database and the issues inherent in bringing together image descriptions from many sources. Moving westward to the University of Minnesota, Charles Thomas discusses a project with a similar campus focus dealing with more heterogeneous data. Karen Coyle, a recent retiree from the California Digital Library, describes a project to bring vendor-supplied journal article information into the University of California's library catalog.

Clearly, metadata is an international issue, involving researchers and practitioners from many countries. Norm Friesen's chapter on CanCore, a Canadian educational project using the IEEE Learning Object Metadata, highlights the importance of integrating project implementation with

standards efforts. Continuing with our survey of specialized communities and metadata formats, we examine two very different projects focused on geographic data: the Alexandria Digital Library (ADL) at the University of California-Santa Barbara and the Cornell University Geospatial Information Repository (CUGIR). Linda Hill and Greg Janée describe the groundbreaking work of the ADL, in particular their efforts to make georeferenced data more generally available in general digital library applications. The CUGIR project, presented by coeditor Elaine Westbrooks, focuses on distributing specialized geographic data through preexisting general channels in an automated fashion.

Part 1 ends with two chapters on the special problems of aggregation and sharing in the new world of metadata. The team of Rachael Bower, David Sleasman, and Edward Almasy, all of the well-respected Internet Scout Project, talk about the development of Scout as both an aggregator and disseminator of information on Internet resources. Their work has led to the creation of tools to assist others: the Scout Portal Toolkit and the Collection Workflow Information System. Last in this section is an excellent summation of the problems inherent in metadata aggregation by Timothy Cole and Sarah Shreeves, who discuss the Illinois Open Archives Initiative Metadata Harvesting Project.

A few lessons have emerged consistently from these metadata projects. They can perhaps be grouped into three major themes. The first is the most important: change happens, and it happens constantly. Get used to it, accept it, and plan for it. Waiting for emerging standards to settle down is a futile exercise; it will probably not happen in our lifetimes.

A second theme is more concrete: stick to standards as much as possible, but if and when you diverge, document what has been done and why it was done. Someone will be managing your project or using your data after your tenure, and will need to understand the context of your decisions. A third theme arises from the second: try to anticipate future uses of your data. This is, of course, why we have standards in the first place, but it cannot be too strongly emphasized.

Part 2 of this book, "The Future of Metadata Development and Practice," moves beyond the lessons learned from the recent past and looks to the future of metadata. Clearly, if change is to be a constant in our lives, we must cultivate the ability to anticipate the trends that will soon wash over us. We begin with two chapters describing how two communities are organizing the development and maintenance of metadata

standards. Harriette Hemmasi describes the work being done to define the needs and options open to the music community as it attempts to come to consensus on community-specific rules, logic, labels, and vocabularies. Steven Bird and Gary Simons describe the Open Language Archives Community and the model that group has developed to manage metadata standards and aggregation, as well as the creation of tools to support reuse.

Caroline and William Arms collaborate to convey their knowledge and experience regarding searching functionality and the relation of metadata to the provision of search and browse services, particularly in large, heterogeneous projects. As metadata is shared, questions of quality and reusability become more pressing. Thomas Bruce and Diane Hillmann begin to take the quality discussion beyond the traditional boundaries defined in library experience toward one more useful in a world of harvesting, reuse, and repurposing. Rachel Heery extends the discussion to the Semantic Web and the issues of identification that underlie the promise of future interoperable metadata. Understanding these issues at the planning stages will make the future much less messy than the present.

The contributions of knowledge and experience from these pioneers in the metadata Wild West will go a long way toward disarming the cowboys behind the rocks. Project managers and planners will find much to learn from their successes, failures, and restarts, and will gain their own experience and knowledge. We ask these new pioneers to follow the lead of the contributors to this book, consider those coming behind them, and pass on their knowledge freely, for the betterment of us all.