

An Operational Model for Library Metadata Maintenance

By Jim LeBlanc and Martin Kurth

Libraries pay considerable attention to the creation, preservation, and transformation of descriptive metadata in both MARC and non-MARC formats. Little evidence suggests that they devote as much time, energy, and financial resources to the ongoing maintenance of non-MARC metadata, especially with regard to updating and editing existing descriptive content, as they do to maintenance of such information in the MARC-based online public access catalog. In this paper, the authors introduce a model, derived loosely from J. A. Zachman's framework for information systems architecture, with which libraries can identify and inventory components of catalog or metadata maintenance and plan interdepartmental, even interinstitutional, workflows. The model draws on the notion that the expertise and skills that have long been the hallmark for the maintenance of libraries' catalog data can and should be parlayed towards metadata maintenance in a broader set of information delivery systems.

Librarians know how to maintain catalog data. Since the days of using industrial-grade erasers to correct and update information on catalog cards, they have made maintaining catalogs an important part of their business to ensure that the contents of the surrogate bibliographic records they present to users are complete and accurate. In spite of this history of catalog maintenance, librarians have not yet given the same kind of focus to the catalog data in newer resource discovery systems—that is, to information in databases other than the online public access catalog (OPAC) and in metadata formats other than MARC. This lack of attention to the integrity of these new catalogs is not necessarily intentional. Those responsible for the general upkeep of digital collections and the bibliographic metadata associated with these aggregates are often distributed throughout the library or even across multiple libraries, and they are not always the practitioners of traditional library technical services. These keepers of non-MARC metadata are as likely to be found in library systems offices or in metadata services departments as in cataloging, catalog maintenance, or database management units.

In a 2004 article on the redesign of database management (DBM) at Rutgers University Libraries, Bogan recounted the use of a core competency model to help identify those aptitudes and skills that most characterize traditional DBM staff.¹ She maintained that understanding these qualities and the values that underpin them allows technical services managers to reposition DBM staff to make useful contributions to the maintenance of a library's digital collections and the catalogs that describe them, and to work in the broader bibliographic infosphere for which libraries now create and maintain data in multiple resource discovery systems. At Rutgers, the DBM team defined its core competency and its role in the library as “fast and accurate maintenance and conversion of bibliographic and related metadata to support Rutgers University Libraries' resources.”² Further, Bogan noted that:

Jim LeBlanc (JDL8@cornell.edu) is Head, Database Management Services, and **Martin Kurth** (MK168@cornell.edu) is Director, Discovery Systems and Services, both at Cornell University Library, Ithaca, New York.

Submitted February 28, 2007; returned to authors April 14, 2007 for revision and second review; resubmitted July 7, 2007 and accepted for publication.

The expertise of DBM will be increasingly valuable as metadata proliferates across an increasing number of repositories. DBM's knowledge-building activities—shared problem solving, implementation of new processes, experimentation, and importing knowledge—strengthen the unit's ability to respond quickly to emerging opportunities. Such opportunities are not likely to be radical shifts requiring rebuilding of skills from the ground up; rather, they will be logical extensions of the expertise embodied in the unit.³

If Bogan is right, and DBM staff are poised to be redeployed as keepers of libraries' metadata in multiple formats and multiple resource discovery systems, perhaps in association with staff in metadata, cataloging, and systems units, how might libraries go about identifying catalog maintenance needs and priorities in this expanded DBM sphere? In the following pages, the authors propose an operational model with which to help answer this question.

Catalog Maintenance versus Metadata Maintenance

In a short piece published in 1986 in the *RTSD Newsletter*, Reid and Fiste outlined the new challenges for technical services managers in maintaining library catalogs in an online environment.⁴ For the purposes of their argument, they defined *catalog maintenance* as “the total work involved in maintaining a card file of bibliographic records for public use—including addition, correction, and deletion of records, as well as production of a syndetic structure connecting individual records.”⁵ According to Reid and Fiste, *database maintenance*, on the other hand, is “the comparable work done in maintaining a computerized file of bibliographic records,” but which includes an expanded, more complex set of tasks:

Sample database maintenance projects include elimination of duplicate records; deletion of alternate call numbers not used locally; addition of alternative title access for titles with abbreviations, initialisms, special characters, symbols, and numbers; checking of filing indicators; removal of initial articles in fields lacking filing indicators; updating fixed field data (e.g. imprint dates) previously ignored for card production; verification of holdings for given collections; determining if cancelled records were dropped during the tape load; and creation of authority files for verification purposes and cross-reference generation.⁶

Reid and Fiste concluded that the transition from a card to an online environment required a re-examination of catalog maintenance procedures and workflows, as well as new methods for compiling statistics and other management information. Though Reid and Fiste anticipated procedural changes, they did not talk directly about the impact of this evolution on DBM staff. As is now clear, although this transition did require additional training for some DBM practitioners, the core competencies of these personnel were, for the most part, adequate to the task. In other words, the work of maintaining catalog data was not normally reallocated to other library staff just because of the change in medium for delivering the data.

The twenty-first-century shift to *metadata maintenance* is no less complex and potentially disruptive, for with the expansion of standard library metadata formats to include non-MARC metadata, metadata experts (including traditional DBM practitioners) must contend with new variables—including converting data from one metadata scheme to another, establishing and maintaining semantic equivalents between metadata elements and values in different schemes, and exposing metadata for harvesting. Westbrook commented on this complexity in the broader context of metadata management:

Metadata management is the sum of activities designed to create, preserve, describe, maintain access to, and manipulate metadata, MARC and otherwise, that may be owned, aggregated, or distributed by the managing institution. These organizational and intellectual activities require the physical resources (web services, scripts and cross-walks), financial commitment (much like that already invested into OPACs), and policy planning that codifies the guiding framework within which metadata exists.⁷

While libraries do pay considerable attention to the creation, preservation, and transformation of descriptive metadata, little evidence exists that they devote as much time, energy, and financial resources to the ongoing maintenance of non-MARC metadata (especially with regard to updating and editing existing descriptive content) as they do to maintenance of such information in the MARC-based OPAC.

From a historical perspective, the number of maintenance functions associated with ensuring the ongoing integrity of library catalogs has been increasing as the world of the library catalog has evolved. As a first step in modeling metadata maintenance operations, the authors offer the following preliminary lists of typical maintenance functions for catalog records. These lists are not meant to be authoritatively defined taxonomies, but to help illustrate how an operational model for metadata maintenance would work.

In card catalogs, the range of maintenance tasks was relatively small:

- Accrual—Filing new catalog cards.
- Deletion—Removing existing cards.
- Modification—Manually revising information on the cards (or producing revised versions of the cards).
- Reporting—Compiling information regarding the cards.
- Export—Photocopying the cards for printed catalogs (such as NUC).

With the development of online catalogs, the number of data maintenance functions in which libraries could and did engage increased somewhat, as machine-readable records, unlike cards, could be moved around in cyberspace. They also could be turned on and off within the resource discovery system in which they were stored in order to make them available or unavailable to the public or other user groups. Thus, the scope of general maintenance tasks in an online environment widened to include the following:

- Accrual—Adding new records.
- Deletion—Removing existing records.
- Modification—Revising data within records.
- Reporting—Generating information regarding records.
- Export—Copying selected records for other uses.
- Migration—Transferring records from one integrated library system (ILS) to another.
- Activation/deactivation—Making records available or unavailable to selected user groups.

With scripting, construction of cross-walks, and other services to which Westbrook referred, the number of maintenance functions required to keep surrogate information in both MARC and non-MARC metadata catalogs clean and accurate has increased still further. The ten most obvious of these maintenance functions are:

- Accrual—Adding new records.
- Deletion—Removing existing records.
- Modification—Revising data within records.
- Transformation—Converting data from one metadata scheme to another.
- Reporting—Generating information regarding records.
- Export—Copying selected records for other uses.
- Mapping—Establishing semantic equivalents between metadata elements or values in different schemes.
- Migration—Transferring records from one system architecture to another.

- Exposure—Making records available for harvesting.
- Activation/deactivation—Making records available or unavailable to selected user groups.

In metadata maintenance, all of these functions may conceivably come into play as the nature and content of digital objects or collections change. Librarians and library programmers already know how to perform this work for given targets, though how the various practitioners of this work must interact in the broader sphere of interrelated objects and collections is not always clear. The elements and values that represent these interactive relationships must be identified, defined, and codified in order to ensure the efficient functioning of the information system and the ongoing accuracy and integrity of the system's data.

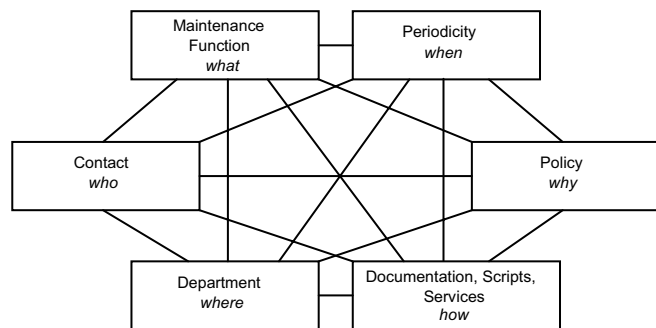
The Model

Although somewhat unknown within the library world, J. A. Zachman's descriptive framework for information systems architecture (ISA) has been widely adopted by systems analysts and database designers for use in businesses and institutions in which technology and effort are distributed.⁸ The ISA or Zachman framework examines entities and relationships within a given system in terms of six generic interrogatives: what, where, who, when, why, and how. Underlying this description is an understanding that individual pieces of the overall framework must be tailored to specific stakeholder perspectives. Zachman uses an architectural example to illustrate how the values inherent in the owner's, the designer's, and the builder's points of view may differ with regard to a structure. Elements of the Zachman framework may thus vary in nature, terminology, and level of detail, depending on the stakeholders at whom the particular elements are aimed. By identifying and associating elements in the system in this way, Zachman is able to construct a multidimensional description of interrelationships among work teams and the tasks or products (or both) they deliver.

Although the original ISA framework, and its later iterations, were intended as a basis on which to construct computer systems (that is, platform and software) architecture, Zachman's model also can be used to design workflows, both automated and manual, and to define variables and processes that can be used to inform strategic thinking, including planning and decision-making.⁹ What follows represents the adoption of a single point of view from the ISA framework, one that is roughly equivalent to a combination of what Zachman would call the designer's and builder's views. The resulting model allows for the examination of metadata maintenance workflows in a distributed environment. Properly speaking, the structure laid out below is far enough removed from the original aims and definitions of

the ISA framework that one should probably not refer to it as such at all, but rather as a Zachman-type or Zachman-inspired model.

The hexagonal diagram in figure 1 represents a simplified view of how metadata maintenance work can be seen in terms of Zachman's interrogatives. The six rectangular boxes depict the way in which these attributes can be understood as applied to metadata maintenance for a MARC or non-MARC catalog. In addition to the maintenance function itself, these attributes include those questions that must be answered in relation to each maintenance function: *periodicity*, or the frequency at which administrators should perform the function; *policy*, or the institutional decision or guidelines for performing the function; *documentation, scripts, and services* that describe the manual workflows and implement the automated processes that execute the function; the administrative *department* responsible for performing the function; and *contact*, the individual or group designated to receive communications regarding the function. Although department and contact may be redundant for maintenance functions carried out in a small operation, these entities may differ in a larger, more distributed environment. For example, in the latter case, the contact may be a collection's administrator, while the department may be the work unit where the metadata maintenance is actually performed. The linearly defined facets of the diagram reveal the interrelationships among attributes for each maintenance function. The real-world context for this description is much more complicated, and one must imagine ten levels (representing the ten metadata maintenance functions proposed above), with interrelational linkage in three dimensions among the individual boxes and hexagons at all levels, to visualize the complete framework on which data maintenance for a given collection should ideally be managed.



Source: This diagram is adapted from the hexagonal model found in J. F. Sowa and J. A. Zachman, "Extending and Formalizing the Framework for Information Systems Architecture," *IBM Systems Journal* 31, no. 3 (1992): 611.

Figure 1. The operational model for library metadata maintenance

Can this model be used to describe catalog, database, and metadata maintenance across the ages, and thus show the increased sophistication and demands of maintaining library information over time? Can it thereby support Bogan's contention that the traditional core competency of DBM staff makes these individuals prime candidates for metadata maintenance assignments? A few examples are in order.

In a catalog card environment, accrual is defined as the filing of new catalog cards. Placed in the context of the hexagonal model outlined above, an operational scenario might look something similar to figure 2. The diagram in figure 2 depicts accrual in terms of the six Zachman interrogatives: what, when, who, why, where, and how. In this hypothetical workflow (which mirrors typical workflows for card filing that veterans of library technical services may remember from the days before OPACs), cards are received and filed weekly. The cards, produced by catalogers or in batch by a vendor, are sent to a contact person or address, from which they are distributed to the staff who will ultimately file them. Institutional policy supports and provides guidelines for this activity, and the catalog maintenance department will perform the task, according to local and national instructions (for example, the ALA filing rules). The institutional policy node in the diagram may seem a bit vague, but it is key to the operational model: whether in writing or simply assumed, the institution has made a decision to create cards (or have them created) and to file them according to a particular filing system. This operational element may be so obvious it seems not worth mentioning, but if one imagines the point at which the library decides to close or freeze the card catalog (such as at the point it decides to move to an online catalog), institutional support for this catalog maintenance function is withdrawn and the workflow becomes obsolete. The model also allows for a decision to take place even before a maintenance function is first implemented.

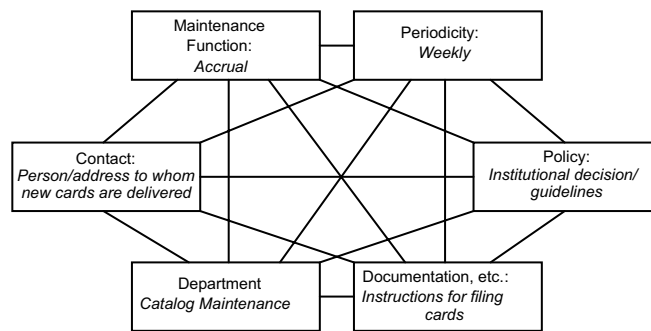


Figure 2. An application of the model to the accrual function in a catalog maintenance context

In the online environment, or from the point of view of what Reid and Fiste term database maintenance, deletion is defined as the removal of existing records from the online system. Using the hexagonal model, figure 3 depicts a hypothetical operational approach that addresses this maintenance function. As in the previous diagram, figure 3 outlines the database maintenance functions in terms of the six interrogatives. In this scenario, requests to delete records from the system are delivered as needed to a contact person or address from which the work will be distributed. Institutional policy supports and provides guidelines for this activity, and the database management department will perform the task according to established instructions, which may include the use of a computer program or script to perform the function.

Figure 4 illustrates a third example of the use of the operational model. In this case, the context is a multiformat metadata environment, and the function described is transformation, defined as the conversion of data from one metadata scheme to another. In this scenario, requests for transformation are sent to the contact person or address (which in this case might be a computer address) as the need arises to convert the data. Note that in a multiformat metadata environment, the contact is that person or computer address responsible for a particular catalog or collection—not, as in the previous examples, for *The Catalog*. The department that will perform the work, in this hypothetical case, will be either a metadata services or database management unit, depending on the catalog and metadata format in question. The transformation will be carried out using a program or information service, with or without a certain amount of manual intervention, again depending on the target catalog or metadata format. Once again, institutional policy will support and provide guidelines for this activity. If the library decides that it will not support transformation of data for a given catalog or collection, because of limited resources or low prioritization of the activity, this particular piece of the overall metadata maintenance model will not be

implemented. Nonetheless, the model signals the potential need for the work and how it can be accomplished, and prompts the policy question.

Implementation of the Model

From a purely conceptual point of view, this ISA-inspired model offers a framework on which to support a method for addressing potential metadata maintenance needs beyond those of simply keeping up the MARC-based OPAC. Many digital library collections have been built as one-shot enterprises backed by grant money. After the digital images are created and the catalog metadata to describe those images has been loaded into the delivery system that will serve the collection, further editing of the metadata is often forsaken. That descriptive metadata might need to be corrected or enhanced over time is simply not part of most collection developers' or collection managers' mindset. Because the metadata for digital library collections is often derived from pre-existing MARC metadata, this oversight might initially seem a bit strange until one remembers that the managers of digital collections are not always technical services staff steeped in database maintenance practice and tradition.¹⁰

Using the operational model described above as a planning and documentation tool, digital collection managers could work cooperatively across library service divisions to pose questions, assign responsibility, and develop policy for ongoing maintenance of the collections they oversee. Knowing which questions to ask, managers and administrators also could give themselves the option of deciding not to pursue a given maintenance function for certain collections. For instance, if the descriptive metadata for a given digital collection has been derived from MARC records, and headings on the MARC records are subject to authority control updates (for example, when death dates are added to personal name headings or when Library of Congress subject headings are changed or split), collection managers

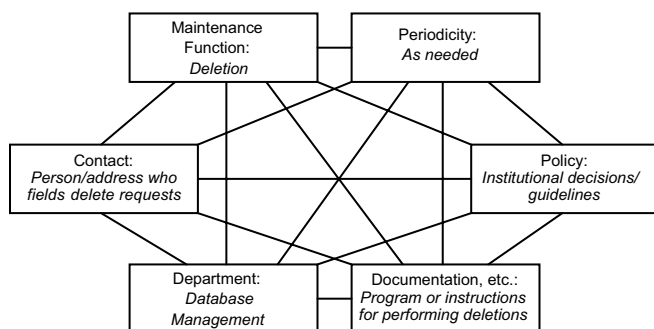


Figure 3. An application of the model to the deletion function in a database maintenance context

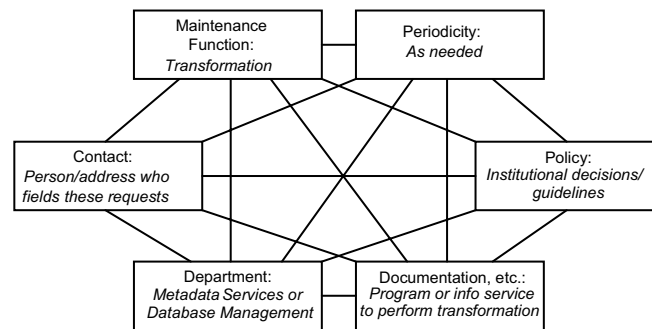


Figure 4. An application of the model to the transformation function in a metadata maintenance context

may choose to update the descriptive metadata for the corresponding records in the target digital collection as well. This work may be done manually, triggered perhaps by a routine report sent to the contact person or address in the model, or through the use of an automated script.

The model also provides a conceptual framework for a Dublin Core metadata maintenance application profile to help manage automated and manual processes and augment collection/service registries to support maintenance within a local or shared system.¹¹ Plugging such a model into what are so far are chiefly object-centered digital registries could provide the basis for communication, operation, and policy protocols, even in a distributed environment.¹² For instance, after reviewing required and available resources for the metadata maintenance of a given digital collection, the institution(s) involved could approve those functions that it chooses to fund and hand the implementation of operational details off to a metadata services or database management group, which would in turn develop the workflows and record its decisions regarding what, why, who, when, where, and how, along with the entry for the target collection in the digital registry. Reports and scripts could then be developed to manage both the manual and automated aspects of the ongoing descriptive metadata maintenance.

Conclusion

The operational model for library metadata maintenance described here offers a simple scheme for organizing the resources involved in metadata catalog maintenance operations, such as documentation, scripts, and contacts. The authors have sought simplicity in the model in order to ensure maximum flexibility for its use—whether merely to organize concepts and planning, to implement interdepartmental or interinstitutional workflows, or to develop automated scripts with which to identify and perform maintenance tasks. Refinement of this model will involve clearer articulation of its potential use and identification of business cases for its implementation.

It is in this regard that library technical services managers may be able to leverage both the traditional skills of their catalog maintenance workforce and the potential applications of the metadata maintenance model described in this paper to address the fundamental operational questions posed by Zachman for any complex or distributed workforce. Even when faced with limited resources for ongoing metadata upkeep, the key elements of this operational model can provide a framework for developing and discussing workflow options and, by extension, can furnish an

inventory of tasks for use in determining data maintenance priorities at an institutional or multi-institutional level.

References and Notes

1. Ruth A. Bogan, "Redesign of Database Management at Rutgers University Libraries," in *Innovative Redesign and Reorganization of Library Technical Services: Paths for the Future and Case Studies*, ed. Bradford Lee Eden, 161–77 (Westport, Conn.: Libraries Unlimited, 2004).
2. *Ibid.*, 176.
3. *Ibid.*, 176.
4. Marion T. Reid and David Fiste, "Catalog Maintenance: Manual to Machine," *RTSD Newsletter* 11, no. 1 (1986): 4–6.
5. *Ibid.*, 4–5.
6. *Ibid.*, 5.
7. Elaine L. Westbrooks, "Remarks on Metadata Management," *OCLC Systems & Services* 21, no. 1 (2005): 6.
8. J. A. Zachman, "A Framework for Information Systems Architecture," *IBM Systems Journal* 26, no. 3 (1987): 276–92. For examples of adaptations of Zachman's framework, see R. Evernden, "The Information Framework," *IBM Systems Journal* 35, no. 1 (1996): 37–68; Edmund F. Vail III, "Causal Architecture: Bringing the Zachman Framework to Life," *Information Systems Management* 19, no. 3 (2002): 8–18; and Chun-Che Huang and Chia-Ming Kuo, "The Transformation and Search of Semi-Structured Knowledge in Organizations," *Journal of Knowledge Management* 7, no. 4 (2003): 106–23.
9. See, for example, the hypothetical case involving the "Oz Car Registration Authority (OCRA)" in J. F. Sowa and J. A. Zachman, "Extending and Formalizing the Framework for Information Systems Architecture," *IBM Systems Journal* 31, no. 3 (1992): 590–616.
10. For more on metadata maintenance issues surrounding the repurposing of MARC metadata for describing digital collections, see Martin Kurth, David Ruddy, and Nathan Rupp, "Repurposing MARC Metadata: Using Digital Project Experience to Develop a Metadata Management Design," *Library Hi Tech* 22, no. 2 (2004): 153–65.
11. The authors proposed a data model for a Dublin Core metadata maintenance application profile in Martin Kurth and Jim LeBlanc, "Toward a Collection-Based Metadata Maintenance Model," in *Metadata for Knowledge and Learning: DC-2006, Proceedings of the International Conference on Dublin Core and Metadata Applications*, ed. Myriam Cruz Calvario, 31–41 (Colima: Universidad de Colima, 2006); also available in *arXiv* <http://arxiv.org/abs/cs/0605022v1> (accessed Feb. 16, 2007).
12. *Ibid.*, 38–40. Commentary on relevant object-centered registries include Christophe Blanchi and Jason Petrone, "Distributed Interoperable Metadata Registry," *D-Lib Magazine* 7, no. 12 (2001), www.dlib.org/dlib/december01/blanchi/12blanchi.html (accessed Feb. 16, 2007); Stephen L. Abrams, "Establishing a Global Digital Format Registry," *Library Trends* 54, no. 1 (2005): 125–43.