Preservation Activities at the Library of Congress

Project Summary

Prepared by Tom Delsey

Revised
May 31, 2002

Project Summary

Background

The volume, range, and complexity of preservation activities carried out by the Library of Congress have increased significantly over the past several years. The Preservation Directorate currently handles approximately 400,000 items a year. The scope of activities now extends beyond the traditional binding, conservation, and microfilming processes to encompass new preservation technologies such as mass deacidification and digitization. Dealings with external service providers are increasing in scope and complexity, and preservation activities overall have become increasingly complex to plan, manage, and monitor.

Within the Library there is increasing recognition of the need for effective and efficient information systems to support the preservation function, and of the importance of addressing preservation systems requirements in a more comprehensive and structured way. The current systems environment for preservation activities within the Library comprises a mix of vendor-supplied proprietary systems (e.g., LARS, MAVIS), applications developed by the Library's automation support staff (e.g., spine label production), and databases developed using consumer software products (e.g., Microsoft Access). The Preservation Directorate and other divisions involved in preservation activities also make use of Voyager, but that system currently provides only minimal support for preservation-related functions. Virtually all of the systems currently supporting preservation activities operate in a stand-alone mode without any significant interfacing capability.

In the past two years, the need to address the Library's preservation systems requirements has become more urgent due to the fact that the system the Library relies on for communicating order information to its binding contractors (LARS) is due for replacement. In addition, the developers of Voyager have announced plans to support binding in the next major release. By defining the broader framework of preservation activities at the outset, there is greater potential for implementing binding support in a way that would facilitate the incorporation of additional preservation capabilities into the system in the future.

There is clearly a need for the Library, at this stage, to analyze and define functional requirements for the support of its preservation activities. It is in the Library's interest to define those requirements as broadly as possible, regardless of whether it embarks on a full-scale systems development project to address its needs in a comprehensive way, or

whether it continues its current practice of acquiring or developing systems supporting single applications. Understanding the relationships among the different functions and requirements is necessary to minimize redundancy and to support integration.

Within the Library there is also a recognition of the potential benefits to be gained through the definition of functional requirements for support of preservation activities in a form that would be relevant not only to the Library itself but to other libraries and preservation service providers as well. Libraries have a pragmatic interest both in communicating information relating to the preservation of materials in their own collections and in accessing information relating to the preservation of materials in other library collections. The transfer of information to service providers is also critical for the support of binding, conservation, and microfilming, and digitization activities, not only for the Library of Congress but for other libraries as well. However, the benefits achievable through the communication and transfer of preservation-related information are currently constrained by the fact that there has been relatively little standardization of data and communications protocols supporting the preservation function. The benefits to be gained by the Library of Congress through promoting the definition of functional requirements and the subsequent development of data standards and protocols are considerable, both in terms of the incentive that standardization would provide to systems developers and the potential for reducing development costs for the Library, and in terms of the recognition that the Library would gain within the larger preservation community by providing leadership in this area. By taking a leadership role in this area, the Library is also able to serve the larger preservation community.

Project Objective

The primary objective of this project is to provide an analytical framework both for the development of information systems supporting the preservation function at the Library of Congress, and for the development of standards to support cooperative effort among libraries engaged in preservation activities and communications between libraries and preservation service providers.

The documentation produced for the project is designed to serve as a reference model delineating and defining the scope of the processes and activities involved in the preservation function, their interrelationships, and the nature of the information that is required to support those processes. It is not meant to serve as a fully articulated, detailed statement of requirements for systems supporting the preservation function.

The models developed for the project function largely at a conceptual and logical level. Although they have been derived from an in-depth analysis of the processes carried out within the Library and the data that is used and generated in those processes, they are not exhaustive in incorporating the details of processing workflows and

individual data elements. They do, however, provide a comprehensive overview of the processes and activities carried out within the Library and the type of information required to support those processes.

The models are designed specifically to reflect functional relationships between processes and between the various classes of information that are used to support individual processes. Although the models do not define a systems architecture as such, they do provide sufficient definition at an information architecture level to highlight the interfaces required to support data transfer either between stand-alone systems developed for specific processes or functions or between discrete applications modules within an integrated system.

Scope

The models developed for the project are centred on the processes and activities carried out at the Library of Congress and on the information used to support those processes and activities. The models cover the Preservation Directorate as well as custodial divisions that have significant responsibilities relating to preservation. They are intended to reflect the full range of preservation processes and activities and to provide a comprehensive inventory of the classes of information used to manage preservation processes and activities within the Library.

Process Model

The process model developed for the project is a structured representation of the processes and activities carried out within the Library. It serves as a means of clarifying the extent of those processes and activities, their interrelationships, and the information transactions associated with them.

The process model identifies five processes that might be considered to be at the centre of the preservation activity: *housing*, *binding*, *treatment*, *reformatting*, and *replacement*. Those central processes, however, are bracketed, as it were, by another thirteen that might be considered as preliminary, post-processing, or ancillary activities in the preservation workflow: *monitoring*, *assessing*, *marking*, *scheduling*, *preparation*, *shipping*, *receiving*, *quality review*, *storing*, *disposition*, *delivery*, *routing*, and *tracking*. In addition, the model identifies five activities that function outside the processing workflow *per se*: *planning*, *maintaining equipment and supplies*, *research*, *training*, and *outreach*.

An analysis of tasks and information transactions associated with each of the processes that was conducted on a division-by-division basis indicates that virtually all the processes identified in the model are carried out in more than one division, and that in many instances a particular task or type of information transaction is common not only to all five divisions within the Preservation Directorate, but to a number of custodial divisions as well.

The analysis of tasks and information transactions also serves to highlight the fact that the most information intensive aspects of the preservation activity as a whole are not the "technical" processes *per se* (*housing*, *binding*, *treatment*, and *reformatting*), but the preliminary and post-processing stages of the workflow (particularly *assessing*, *preparation*, and *quality review*).

Details of the process model and the task analysis are documented in Part 1.

Information Model

The information model developed for the project is a structured representation of the information required to manage the preservation processes and activities carried out within the Library. It serves as a means of clarifying the extent of the Library's information requirements for support of the preservation function and the interrelationships between and among those requirements. It also serves as a means of identifying areas of interface between the information required to manage the preservation activity *per se* and information that is generated and/or used in other library functions.

The model comprises seventeen entity-relationship diagrams depicting the various objects, agents, activities, events, etc. about which information is required in order to support the preservation activities of the Library.

The model builds on the work done for the IFLA study on the Functional Requirements for Bibliographic Records (FRBR), which serves as a point of reference both for the *item* entity and for its aggregate and component entities (*item aggregate*, *item part*, and *element*). Those entities, along with two other key entity groups are central to the model developed for this project. The second group comprises the "tools of the trade" for the preservation activity: *procedure*, *material*, *equipment*, *system*, and *facility*. The third group comprises the *employee* and the related entities *skill time* and *unit*.

Using those three key entity groups as their primary point of reference, the remaining diagrams in the information model depict the entities associated with each of the processes and activities identified in the process model. Each process or group of processes represented in the model has associated with it several entities in addition to the twelve represented in the three central entity groups noted above; in total, there are seventy-two entities defined in the model. They represent a broad range of actions and transactions, events, objects, and other "things" about which information is used or generated in the course of the preservation activity. Included in the model are entities associated with the *monitoring* process (*environment*, *survey*, *incident*, *response*, etc.), entities associated with the *assessing* and *scheduling* processes (*assessment*, *recommended action*, *action request*, *scheduled action*, etc.), entities associated with the *preparation* and *quality review* processes (*order item*, *vendor*, *contract*, *commitment*,

delivered product, etc.), and entities associated with the planning activity (*proposal*, *project*, *budget*, *allocation*, *audit*, etc.), in addition to many others.

The information model serves not only to illustrate the diversity of information required to support preservation activities, but to highlight as well the information interdependencies between the various processes involved. Not only do the central entity groups (*item*, *item aggregate*, *item part*, *element*; *procedure*, *material*, *equipment*, *system*, *facility*; *employee*, *skill time*, *unit*) come into play in virtually every process and activity, but other entities as well (e.g., *project*, *scheduled action*, *order*, *delivered product*) recur at several key points throughout the processing workflow. The model makes it clear that information flow is an essential part of the processing workflow, and that the design of systems supporting the preservation function must take into account the complex interdependencies and interfaces between and among the various information sources that the processes draw on.

Details of the information model are documented in Part 2.

Appendix A includes a comprehensive listing of attributes associated with each of the entities identified in the information model and the relationships that operate between each of the entities.

Functional Requirements

The statement of functional requirements developed for the project is designed to encapsulate the range of information transactions relating to preservation that occur across organizational units within the Library and across discrete processes, and to relate them to a structured set of information management functions.

The requirements statement is structured around broad information management functions. For each function, requirements are identified relative to specific classes of information derived from the entity analysis in the information model. For each class of information relevant to a given function, requirements are specified according to four categories of transactions derived from the information transaction analysis in the process model (create/update, search, format, and compile).

The ten information management functions around which the statement of functional requirements is structured are the following:

- Manage item description and action information
- Manage environmental and collections monitoring information
- Manage supplies and equipment inventory information

- Manage routing and tracking information
- Manage processing information
- Manage order, shipping, and receiving information
- Manage employee and unit performance information
- Manage policy and procedures information
- Manage planning information
- Manage research, training, and outreach information

In effect, the statement of functional requirements takes a higher-level view of the preservation activities carried out within the Library than what is depicted in the process model. Both the process model and the statement of functional requirements provide an overview that cuts across the organizational structure of the Library, but the statement of functional requirements goes further in that it also cuts across the work processes defined in the process model. Extrapolating from the information transactions associated with each of those work processes, the statement of functional requirements defines a set of information processes representing functions that are relevant in most cases to more than one work process.

The structuring of the requirements statement around these broader information management functions serves two purposes. First, it provides a structure to facilitate a senior management assessment and strategic prioritization of the Library's needs for systems development. Second, it provides a definition of requirements in a form designed to facilitate discussion with systems developers on the scope of the Library's needs and the kinds of applications that would be best suited to address those needs. In addition, the statement of functional requirements provides sufficient definition at an information architecture level to highlight the interfaces required to support data transfer between existing applications (e.g., the module of Voyager that supports the management of item description) and any new applications developed to support specific processes or functions.

Details of the information model are documented in Part 3.

External Review

In addition to being reviewed by managers and staff within the Library of Congress, the models developed for the project were reviewed by an external working group of representatives from a number of major research libraries in the U.S. The objectives of the external review were as follows:

a) to determine whether the processes reflected in the draft process model for the Library of Congress are consistent with the processes that comprise the preservation function in other libraries;

b) to determine whether the range of information reflected in the draft information model is consistent with information requirements in other libraries; and
c) to identify a subset of information requirements that libraries would want to share or have access to across institutions (i.e., via MARC records).

Overall, input from the external working group indicates that the processes and activities identified and defined in the process model are consistent with the processes that comprise the preservation function in the libraries represented by the working group participants and that the range of information reflected in the information model is consistent with the information requirements of those libraries.

The external working group also provided helpful input in re-assessing the use of the "Action Note" field in the MARC 21 format as a vehicle for communicating information on action taken to preserve an item to external organizations.

Details of the external review are documented in Appendix B.

Recommendations on the use of the "Action Note" in MARC 21 are documented in Appendix C.

Next Steps

This project was designed as the first stage of a longer-term initiative to address the preservation systems needs of the Library. The models produced for the project serve to document the extent of Library's preservation activities and its information requirements in a structured form that is designed to serve as an overall frame of reference for subsequent, more detailed analysis.

In order to move the systems development initiative forward from this stage, efforts need to be focussed next on strategic planning, standardization, a review of current systems initiatives within the Library.

Strategic Planning

It is clear from the models developed for this project that the scope of work involved in developing automated systems to support the preservation function will be substantial. The range of processes involved, the extent of the Library's information requirements, and the diversity of applications and interfaces that will have to be put in place all indicate that addressing the Library's systems needs in a meaningful way will require a significant investment in time and money. Given the scope of the effort involved and the probability that addressing the full range of requirements will

take a considerable length of time, it is essential for senior management to undertake a strategic assessment of its needs and to establish a set of priorities to guide longer-term planning.

The models and statement of functional requirements developed for this project provide a framework that senior management can use to focus its strategic planning. It is recommended that the planning process begin with a review of the ten information management functions identified in the statement of functional requirements, in the context of program priorities for the Library as well as the operational priorities of the divisions responsible for carrying out preservation function. For each of the information management functions identified, senior management should assess the strategic benefits to be achieved through systems support, and estimate the investment required for each in terms of approximate order of magnitude.

Once senior management has completed its strategic assessment of benefits and costs, and established its systems priorities, a more detailed analysis of requirements will need to be done for the function(s) and/or application(s) determined to have the highest priority.

Standardization

The external review served to confirm the applicability of the models developed for the project to preservation activities in other libraries. It can be assumed, therefore, that systems development undertaken to address requirements identified by the Library of Congress has the potential to serve the needs of a wider community of users as well. From a systems developer's perspective, the market potential is likely to have a significant bearing both on the priority assigned to the development effort and on the calculation of costs to the client. However, in assessing the market potential for applications supporting the preservation function, systems developers will also give significant weight to the use of standards for data and communications in the area of the application. Standardization is also a critical factor for those applications that involve communications between the library and its service providers.

The models developed for the project provide a well structured starting point for assessing the needs and potential for standardizing preservation data and communications. In order to move the standardization process forward, however, further analysis is needed to evaluate the suitability and adequacy of existing standards (e.g. the ANSI/NISO Z39.76 standard on data elements for binding library materials), as well as to assess needs in areas where there are currently no widely implemented standards. The assessment and analysis will require consultation with other libraries, service providers, and systems developers.

Current Systems Initiatives

There are a number of systems initiatives currently underway within the Preservation Directorate as well as in a number of custodial divisions to address specific operational needs. For the most part, they are being planned and implemented at the divisional level. However, it is evident from the process analysis conducted in this project that in several cases the initiatives have the potential to serve parallel needs in other divisions as well, both within the Preservation Directorate and in Public Service Collections and Area Studies. It is recommended that senior management review preservation systems activities across the Library (including projects at the planning stage as well as those for which implementations are currently in place) to assess the potential for wider application and increased benefits to the Library as a whole.

It is also recommended that the Library examine the feasibility of enhancing the data structures currently defined in Voyager at the holdings and item level to accommodate some of the additional requirements for item description identified in the information model. Part of that discussion should focus on metadata requirements relating to the digital reformatting of collection materials.