

Cataloging Directorate Strategic Plan

Goal 4, Group 2 : Processing Rule Analysis Group

Report

Executive Summary:

The Processing Rule Analysis Group was formed to provide recommendations about how the Cataloging Directorate can supply bibliographic control and access for that digital content for which it has bibliographic control responsibility. The group began its work by studying the various modes of bibliographic control/access that are currently in use at the Library of Congress for digital content as well as other possible modes of control/access, with the goal of identifying those modes that are both appropriate and implementable in the Library of Congress environment. Presentations by experts on these various modes allowed the group to gain information needed to identify the modes that are presented in the recommendations that follow.

It was clear to the Group that LC will not have the bibliographic control resources to create MARC/AACR records for all digital objects. The Group also made the assumption that, without further technical development, the ILS will continue to support only MARC records and therefore, some recommendations would be for bibliographic control outside of the ILS. After considering the key advantages and feasibility of various modes of control, the Group recommended three modes for use at LC: Web guides, MODS records, and "traditional" MARC/AACR records as well as new "access level" records which emphasize subject elements over descriptive ones. The intent behind this proposed new level of record is to save time and money by minimizing catalogers' time-consuming but often futile efforts to locate traditional descriptive elements on digital resources, while preserving the ability to perform subject and keyword retrieval on these resources.

A small number of modes was selected in order to limit the negative impact that a broad proliferation of modes might have on users, bibliographic control staff, and any federated searching/portal product that may be acquired by the Library. Because the group is recommending the employment of modes of control outside of the ILS, it is critical that any federated searching/portal product selected by the Library provide access to the non-ILS metadata-- the resource descriptions should be accessible to all users, even if access to the resources themselves is only available "on campus" at LC.

Based on criteria the group devised for determining which mode of control is appropriate for the variety of resources to be described, one (or more) of the modes of control would be applied. When descriptions are to be provided outside of the ILS, it is recommended that a MARC/AACR record for the aggregation be made in the ILS to point indirectly to the descriptions housed elsewhere. It is also recommended that regardless of the metadata scheme chosen (MODS or MARC/AACR), all records for digital content should be exposed via the Open Archives Initiative Protocol for Metadata Harvesting to support the use of LC-

created records in other metadata services. Additional recommendations related to each of the selected modes of control are provided in the report.

The scope of the deliberations of the group also includes the bibliographic control and access issues related to digital counterparts-- recommendations for the treatment of counterparts are included below.

It is important to note that, beyond the need for tools, the recommendations in this report will have major impacts on training and workflow. To ensure that the Cataloging Directorate can provide effective bibliographic control for digital content, it is important that the directorate be actively involved in planning for digitization activities and the acquisition of born-digital content and digital counterparts. This subgroup endorses the employment of the Digital Life Cycle Framework by Library Services to allow the Cataloging Directorate to manage its staff resources in a manner consistent with LC priorities.

Although the individual recommendations are found throughout the report in context of the discussion, they are also collected together in Appendix D.

Group Membership:

Caroline Arms (OSI)
Ardie Bausenbach (OPS/APLO)
Rebecca Guenther (OPS/NDMSO)
Allene Hayes (CAT/SMCD)
Gabe Horchler (CAT/SSCD)
Carolyn Larson (PSC/ST&B)
Tracy Meehleib (PSC/DRT)
Hien Nguyen (ACQ/SRD)
Carlos Olave (AREA/HISP)
Don Panzera (ACQ/ELAD)
David Reser (CAT/CPSO), leader
Regina Reynolds (ACQ/SRD)

Charge to Group

Task: Determine how digital content will be brought under bibliographic control and made accessible.

- A. Investigate the types of bibliographic control and access that are appropriate and implementable in LC's environment.
- B. Develop criteria for determining and applying appropriate types of bibliographic control and access to categories of digital content, and policies for ongoing maintenance of the resulting descriptions.
- C. Determine what standards and formats will be used to accomplish the types of bibliographic control that are identified (note that metadata for some resources may be made accessible in more than one standard/format).

D. Develop and promulgate guidelines for determining what level(s) of a Web site or other remote resource should be brought under bibliographic control and in what environment (LCDB, LC portal, other).

E. Determine the relationship of records for digital content and the LC catalog to any portal LC may implement.

Scope and Introduction

The scope this subgroup has assumed for our deliberations includes:

- remote access electronic resources¹
- both born digital and digitally reformatted materials
- materials for which the primary cataloging responsibility is the Cataloging Directorate
- descriptive metadata (i.e., not administrative, structural, preservation metadata)
- electronic resources requiring bibliographic control and access, regardless of whether the Library has truly “acquired” the resource, archived the resource in a repository, acquired licenced access to the resource, or is merely “pointing” to the resource
- bibliographic data; although the group was not specifically asked to address authority data, we do endorse the general principle of making authority files conveniently accessible in order to support application by metadata creators (using any mode of bibliographic control).

For many years libraries have been involved in discussions on the degree to which efforts should be made to bring digital content into bibliographic control, particularly whether libraries should invest their resources in describing “free” resources on the Web². At a 1995 conference held at the Library of Congress, a number of principles related to digital content were identified-- despite the vast technological advances since that time, it would seem that this list of principles has stood the test of time:

- “1. Libraries exist to provide value-added services to a wide variety of materials, including: selection, organization, access, location information, delivery, and preservation
2. Libraries will include a mix of traditional materials (print and non-print) and digital resources indefinitely

¹For the purpose of this discussion, “remote access” is used in the AACR2 context, that is to differentiate from “direct access,” the latter being tangible resources on physical carriers (e.g., CD-ROMs, zip disks). “Remote” is in no way intended to imply “distance” from LC, e.g., a remote access resource can be on local servers.

²There isn’t really a “whether” question for digital content actually brought under custody of the Library, it *must* be brought under control in some manner.

3. Library collections will continue to be only subsets of the universe of publications, resources, and information
4. Like traditional materials, digital resources will have more value and utility if they are organized, making resources known and available
5. Libraries should integrate access to digital resources with access to conventional materials
6. Genre is a more useful organizing principle than format
7. Information seekers benefit from self-indexing resources, producer-generated access, and librarian-generated access
8. Librarians will continue to use judgment in applying varying levels of description and access, as appropriate to each resource, in order to provide retrieval of relevant resources in a cost-efficient manner.”³

The work of this subgroup is necessarily centered on the last principle, and is in essence an opportunity for reflection and review rather than a “start from scratch” exercise, since LC staff have been providing bibliographic control and access to digital content for the last decade. Our work also assumes that valid selection criteria have been applied to the digital content prior to a decision on what mode of bibliographic control is appropriate. This report focuses on the “how” not the “whether,” and we make no attempt to identify reasons why libraries should provide access to digital content, though the remarks of Martin Dillon found in a recent collection of articles regarding metadata for digital content resonate:

“You may very well ask, why should we care? A second question provides a more than adequate counter: do libraries have the responsibility for providing access to patrons of knowledge objects on the Web? If your answer is, no, others are doing an adequate job, specifically Google, skip this article; indeed, skip the 21st century of librarianship.”⁴

A rather important caveat is worth noting: any conclusions that the modes of bibliographic control we recommend are effective modes for providing access to users are based purely on anecdotal evidence. That is to say, our current systems provide no “use statistics” to measure the degree to which any of the bibliographic products created are helping users get to the resources we describe or to determine which approach may accomplish the task “better.”

Method of Work

- Reviewed and revised the charge
- Studied the modes of bibliographic control/access currently in use at LC, with

³*Organizing the Global Digital Library Conference*, held at the Library of Congress, Dec. 11, 1995. See: <http://www.loc.gov/catdir/marvel>

⁴*Portal: Libraries and the Academy*, Vol. 3, No. 1 (2003), p. 138.

- presentations by experts in the various modes
- Developed criteria that would assist in decision making for choosing a mode of bibliographic control/access
- Identified the modes of bibliographic control that are implementable in LC's environment and the key advantages for each mode
- Made maintenance recommendations for each mode
- Established subgroup to make recommendations related to digital counterparts
- Identified needed tools and areas for further investigation.

Report Organization

For the ease of discussion and comprehension, the report has been divided into two sections, one for digital only/born digital content and a second for digital counterparts of original (print) materials.

Part 1: Digital only/Born Digital Content

Criteria for determining mode of control

The following is a compilation of criteria and/or resource characteristics that may lead one to choose one mode of bibliographic control over another. It is expected that these criteria would be applied⁵ in concert with other decisions made at the time of selection decision (e.g., cataloging priority, and perhaps someday, candidacy for local archiving). Although there may be some overlap in issues raised in these criteria with selection criteria, as presented here the criteria are intended only to drive a decision on the mode of bibliographic control.

Temporal criteria

- fleeting nature of interest in the topic, i.e., resources that may not have long-term interest
- timeliness of required description (e.g., different modes of bibliographic control/access may be accomplished more quickly than others to support an immediate need; note that more than one type of control may be required to accommodate both speed and depth of description such that early control may take one form, but final control another)

⁵This subgroup does not make any recommendation on "who" applies the criteria, i.e., recommending officers or selection officers (a separate CPC work group has been charged to examine issues related to selection), but we hope that as for all other materials, cataloging staff have an avenue to consult with selection officers on questionable decisions.

- ephemeral nature of the resource

"Context" criteria (i.e., related to where users might predict the description to live)

- value of collocation using controlled vocabulary (authoritative descriptive and subject headings and/or classification)
- analogous treatment for different manifestations of the same work and/or similar categories of content to support consistent retrieval by users
- research value, e.g., reference works may require different modes of control than other resources
- value of individual description vs. collective nature of description (e.g., granularity of subject access desired)

User access criteria

- importance of accessibility to descriptions via federated searching
- available alternatives to traditional bibliographic search access (e.g., hidden Web vs. "Google"-accessible Web)
- citation value, i.e., support for finding resources that are likely to be quoted or cited in bibliographic references

Enterprise-wide system dependencies

- integrated system dependencies (e.g., purchase orders, invoice payment, descriptive metadata embedded in METS objects)
- impact of cooperative endeavors (e.g., shared cataloging, registries of digital content, metadata exposure and/or distribution)
- impact of decisions about how to manage items and groups of items as digital objects in repositories (e.g., granularity relationships of metadata to digital objects)
- impact of granularity of receipt and associated workflow
- impact of permanent storage decisions (e.g., archived at LC)

Technical infrastructure

- metadata transformation or extraction potential, where mappings and tools are available
- availability/usability of tools to support the modes of control

Resource issues (e.g., staffing, training)

- need to keep metadata in sync with resource over time (e.g., AACR2 requirements to re-describe based on changes between iterations)
- available metadata from publisher/suppliers
- efficient treatment for large aggregations of content (e.g., large collections of sites bulk harvested or purchased in aggregations)
- staffing availability (e.g., staff with the appropriate training and at the appropriate grade level)

Recommended Modes of Bibliographic Control

In its report of April 2002⁶, the Library of Congress Metadata Policy Group (LC MPG) made an observation that “Metadata is not required for its own sake, but to support desired functionality.” With this in mind, this Goal 4 subgroup discussed the various modes of control that were possible in light of the functions they are to accomplish. Although the Library of Congress has long been in the business of MARC/AACR cataloging, it should be understood that we will not be able to afford item-level cataloging in the ILS for all digital content and we will have to employ additional means of bibliographic control. The group also realized that we need to limit the number of modes of bibliographic control: a limited set of disparate data sets should make it easier to target the sets for federated searching and to configure a portal to supply public access (i.e., limit the impact of multiple modes on users); a proliferation of modes of control would make it difficult to effectively train bibliographic control staff if they are expected to master many modes of control and switch between them on a regular basis.

The group identified three basic modes of bibliographic control/access to recommend as “implementable” in LC’s environment: **Web guides**, **MODS records**, and **MARC/AACR catalog records**. Each will be discussed below, along with the advantages of that particular mode and any additional recommendations related to the mode. It is important to note that some resources being brought under bibliographic control will require only one mode of control to provide adequate access, while others may warrant more than one mode of control.

1. Web-guides⁷

For several years, LC reference staff have been creating Web-accessible lists using a variety of approaches that can range from sophisticated annotated bibliographies to what in reality are “shared bookmarks.” Pathfinders created by OCLC’s software (annotated bibliographies primarily derived from records in the OCLC Resource Catalog, sometimes supplemented with embedded links that perform dynamic searches) are examples of the former and are used for business-related resources and the “Portals to the world” sites created by Area Studies. Hard-coded HTML pages using a variety of styles (e.g., “Alcove 9” (reference websites identified by the Main Reading Room staff) and simple A-Z lists on various reading room pages) are illustrations of the latter. In many respects, listings such as these are an efficient means to identify and collect Internet sites useful for reference purposes. For researchers, the primary benefit of such lists is that the

⁶Report of the Metadata Policy Group, prepared for the Library of Congress by Members of the Metadata Policy Group, April 2002
http://www.loc.gov/staff/deog/resources/metadata_policy_group.pdf

⁷Examples include “Portals to the World” pathfinders (e.g., <http://www.loc.gov/rr/international/amed/afghanistan/afghanistan.html>), Alcove 9 (e.g., <http://www.loc.gov/rr/main/alcove9/library.html>), and American Memory “related resources” pages (e.g., <http://memory.loc.gov/ammem/aamhtml/aamrel.html>).

Library's subject-area specialists selected these resources as the "best" or most useful on the Internet, thus providing a critical "value added" sorting of the vast resources on the Web that users would face if merely searching the Web via search engines. The advantages of using Web guides as a mode of providing bibliographic control and access are that they:

- consolidate access to ready reference resources, especially those tools that incorporate ever-changing parameters to remain current, e.g., Web sites that provide calculators for interest rates, exchange rates, mortgage rates
- provide "hot-topic" access: reference staff are able to post information immediately in response to high demand, e.g., sources for information on frequent requests related to current news events
- provide access to resources that are judged to be reliable and useful but possibly unstable, e.g., no apparent institutional commitment to maintain the service; in the case of commercial sites, information made available to attract users that may be removed from the site with corporate refocus/reorganization (e.g., an accounting glossary on a business publisher's site, a business search engine on a bank Web site); in the university setting, possibly a personal page maintained by a professor in the field (e.g., prepublication drafts of articles, studies, guides of professors' personal pages on university sites)
- provide thematic access to relatively small resources that traditionally have not received item-level bibliographic control (e.g., resources that simulate resources known in the print world as pamphlets, brochures, news articles, individual encyclopedia articles as well as born-digital resources such as Listserv/discussion group messages)
- provide "table of contents" type access, such as lists of digitized pamphlets on particular topics or collections of digitized short stories
- share responsibility for bibliographic control/access with reference colleagues-- resources listed solely in Web lists (i.e., not also recommended for additional modes of control) are created and maintained outside of the Cataloging Directorate
- provide access to the Web lists, where possible, via search engines both within the Library (e.g., Inktomi, the LC website search engine and possibly any new Portal/metasearch application) or externally (e.g., Google), thus making these valuable resources accessible to a greater number of searchers

It should be understood that a subset of the individual resources added to these guides will also meet criteria for additional bibliographic control (e.g., MARC/AACR records in the ILS). Although Web guides have been successfully used at the Library, a number of recommendations for improvement were identified by the subgroup-- it is suggested that consultations with the library-wide Internet Operations Group be held to resolve several of the recommendations.

Web guide recommendation 1: Investigate how to assure that all such Web guides (including “on the fly” dynamic pages) created by LC reference staff are indexable by Internet search engines thus increasing access to the guides.

Web guide recommendation 2: Investigate the possibility of defining and embedding a core set of metadata tags in all LC-created Web guides (e.g., embed the XML tags from a MODS record, develop guidelines for title construction). Embedded metadata may facilitate access to the guides from a library portal/metasearch facility and/or LC website search engine (Inktomi).

Web guide recommendation 3: Investigate mechanisms to enable users who access a guide from a search engine to be alerted to the presence of other guides produced by the Library’s subject specialists.

Web guide recommendation 4: Encourage any group addressing the possible standardization of Web guide styles across the Library (e.g., by adopting a single middleware product for developing Web guides, or applying standardized content schemes) to consider the impact of such standardization on any portal/metasearch product⁸ selected by the Library.

Web guide recommendation 5: Because the Web guides created by LC reference staff represent a resource investment by the Library and are seen as authoritative resources in their own right, each Web guide or cluster of guides should be represented by a record in the ILS that supports subject/keyword access to lead ILS searchers to these valuable resources.⁹

Web guide recommendation 6: Determine if it is possible to generate use statistics to determine how effective Web guides are (e.g., how frequently are they consulted? How often do users link to sites described in the guide? What percentage of the described sites do users link to?).

2. MODS Records

The Metadata Object Description Schema (MODS)¹⁰ is a MARC-compatible schema, developed by LC’s Network Development and MARC Standards Office and interested experts, for a bibliographic element set (i.e., descriptive metadata) that

⁸See Appendix A for background information on resource discovery via library portals (also known as metasearching or federated searching).

⁹See LCCN 2002564514 for a prototype example of such a record.

¹⁰See the official Web site of MODS for more information:
<http://www.loc.gov/standards/mods/>

may be used for a variety of purposes, particularly for library applications. As an XML (eXtensible Mark-up Language) schema, MODS can carry selected data from MARC 21 (i.e., converted from MARC 21 to MODS), as well as enable the creation of original resource description records. The schema was developed as a response to concerns that the Dublin Core metadata schema defined in NISO Z39.85-2001 is too "simple" for library-based applications¹¹, and that the full MARC 21 format is too complex and not user-friendly for use outside of integrated library systems. Advantages to using MODS as a mode of bibliographic control and access include:

- XML is the environment adopted by the Internet technologists who are developing the tools and services for the future
- as a MARC derivative, MODS records are highly compatible with traditional MARC records
- mnemonic XML "tags" allow for easy identification of elements by specialists and non-specialists alike
- flexible displays of MODS records can be generated and easily changed using style sheets
- XML structure facilitates data creation through the use of templates
- MODS supports the coding of hierarchical relationships in a single record rather than requiring the creation and association between multiple records
- XML structure makes MODS compatible with other XML-based standards (e.g., METS), allowing descriptive metadata to be packaged with other types of metadata (preservation, structural, administrative, etc.)
- information can be expressed clearly, (e.g., MODS allows for more types of dates, date elements and use of the ISO standards for encoding dates)
- use of an XML schema allows for extensibility (e.g., inclusion of elements from other metadata schemes and/or locally defined elements)
- data is expressed with less redundancy than MARC records (combines categories of "like" data that are derived from multiple elements in MARC; eliminates the need to code control fields that correspond to variable field data)
- MODS is not tied to any particular content description scheme (e.g., AACR2, ISBD punctuation), but can enforce business rules at the institution level to assure data quality and consistency, with an eye toward data compatibility. Given this flexibility, MODS records could range from minimal descriptions not requiring highly-trained staff to

¹¹See Appendix B for a more complete explanation of why this working group is not recommending Dublin Core as a mode of bibliographic control.

- rich records built by professional catalogers¹²
- character set is Unicode™ based
- language may be coded at the data element level as well as the record level (allows for bilingual records that may support cooperative metadata projects with other national libraries, e.g., Global Gateways)
- metadata may be tagged at a sufficiently granular level to support effective parsing of citation information used by linking technologies such as OpenURL
- MODS provides the data elements recognizable and useful in library contexts, as opposed to more general and/or simple metadata schemes.

Because MODS is a very new standard, the Cataloging Directorate has relatively little experience in creating MODS records. An ideal project that LC has used to experiment with MODS is the Web archiving project MINERVA (Mapping the INternet Electronic Resources Virtual Archive). For the Election 2002 Web Archive¹³, over 4,000 different Web sites were selectively identified and archived. Although a record for the thematic collection as a whole has been created in the ILS using traditional methods (i.e., MARC/AACR) to represent the archived collection, the costs of producing catalog records for the 4,000 constituent sites using traditional methods was prohibitive¹⁴. Yet some type of metadata was necessary to provide bibliographic control and access to the harvested sites. Working under the guidance of the MINERVA project team, contractors created MODS records for the Web sites that contained the metadata elements necessary for effective retrieval of each Web site (e.g., candidate names, political party, office sought, jurisdiction); some attempts were also made to extract metadata from the sites themselves (e.g., HTML titles). User-friendly displays were then generated from the records¹⁵. MODS seems ideal for such projects because:

- MINERVA record creation, using templates and without strict

¹² For descriptive metadata records, the LC Metadata Policy Group identified a minimal/essential set of elements as: a unique identifier/identifying number and a title (the number is often enough for a machine to retrieve the object if it includes the path to the resource and the title is required to provide basic identification to a human). Although it is not common that a MARC/AACR record could be this minimal, MODS is a suitable alternative when these minimal requirements are necessary to provide "control" in the context of a digital repository, presuming alternative methods for discovery and access are provided (e.g., full text search engine).

¹³ See <http://www.loc.gov/minerva/collect/elec2002/index.html> for more information on the collection.

¹⁴ A cautionary note: due to the flexibility of MODS, there can be no assumption that using MODS in and of itself is less-costly than traditional MARC/AACR records-- costs will depend on the desired richness of the records.

¹⁵ See <http://www.loc.gov/minerva/collect/elec2002/ulmer-record.html> for a display version of a MODS record, and <http://www.loc.gov/minerva/collect/elec2002/ulmer-record.xml> for the raw XML version of the same record.

- adherence to AACR2 and LCSH, can be done by contractor staff not familiar with traditional library cataloging standards
- although tools exist to convert the MINERVA MODS records to MARC 21 for loading to the ILS, the impact of adding large quantities of non-standard records to a rule-based catalog has not yet been thoroughly analyzed-- MODS records provide a valuable alternative resource discovery tool
 - because candidate Web pages are volatile by their very nature, the re-description for each iteration required by AACR2 is not feasible (some sites were archived dozens of times); the MODS descriptions are generic to all iterations and no attempt was made to adjust the MINERVA records to reflect new iterations of the site
 - although the archived sites can be accessed, in many cases the "active" site no longer exists-- other libraries are unlikely to expect traditional LC cataloging to represent such fleeting resources since they are unlikely to have archived the sites
 - the MINERVA records could be exposed via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for use in any OAI-compatible services.

Other important pilots evaluating the use of MODS at the Library of Congress include digital library projects such as: I Hear America Singing, the AV-Prototyping project, and the Veterans Oral History project.

MODS recommendation 1: The Cataloging Directorate should continue to experiment with MODS as the mode of bibliographic control for large collections of archived Web sites.

MODS recommendation 2: When resource discovery is provided via MODS records, a subject/keyword-rich MARC/AACR record for the thematic aggregation as a whole should be made in the ILS, with a link that leads the user to the system/interface where the MODS records can be searched directly.¹⁶

MODS recommendation 3: LC should consider exposing all MODS records that describe digital content via the Open Archives Initiative-Protocol for Metadata Harvesting¹⁷.

MODS recommendation 4: A prototype support infrastructure for MODS

¹⁶See LCCN 2003556434 for a sample record.

¹⁷A recent survey by the Program for Cooperative Cataloging (see http://www.loc.gov/catdir/pcc/tgsrvyeres_final.pdf) reveals that only 18% of responding libraries have cataloging staff involved in routine creation of non-MARC metadata, although the interest in LC-created MODS records may grow in the future.

should be developed to determine if expanding the use of MODS as a mode of bibliographic control for resources beyond archived Web sites is feasible and cost effective. Infrastructure components include (but are not limited to):

- tools for MODS record creation and maintenance including Unicode™ support (being developed in NDMSO)
- XML-compatible database software capable of storing and indexing MODS records created across the Library
- tools for searching MODS records, including:
 - the ability to accommodate Z39.50 searching and provide a means to target collections of MODS records via federated searching/metasearching/library portal application (being investigated by NDMSO as part of digital library projects), and
 - tools for generating and evaluating “usage” statistics to determine the effectiveness of MODS records

MODS recommendation 5: Establish Library-wide profiles and best practices for all instances of LC’s implementation of MODS records, that address such issues as: persistent identifiers for records (control numbers), mechanisms to represent rights statements or access rights information, mandatory elements.

MODS recommendation 6: Consider issues related to integrating access to MODS records with other metadata (e.g., analyze the impact on “the catalog” if selected MODS records are converted to MARC and added to the ILS).

3. MARC/AACR records

The MARC and AACR standards have undergone considerable revision in the last several years to accommodate the cataloging of digital resources. There are currently over 2,500 MARC/AACR records for monographic and integrating resource remote access electronic resources in the Library of Congress Database. These records primarily provide access to freely available resources, but also describe LC-created reference tools (e.g., Web guides, databases), LC and partner-developed digitized exhibits/collections, and subscription/licenced resources. The subgroup believes that MARC/AACR records continue to be an appropriate mode of bibliographic control in the LC ILS in that they support:

- Shared cataloging distribution via traditional channels, e.g., the Cataloging Distribution Service. This may be particularly valuable for resources that are likely to be represented in the catalogs of other institutions, e.g., subscription resources purchased by many research libraries, resources archived by a trusted repository assuring long-term access, document-like objects assigned standard numbers
- Cross-content discovery: supports a user’s ability to find all format types from multiple LC “silos” in a single search, e.g., all books, periodicals, moving

- images, photographs, maps, Web resources related to a given person, topic, place, etc.
- Depth of descriptive granularity and markup: supports more sophisticated searching than typical "simplified" metadata schemes without the data or granular markup, e.g., searching at field/subfield levels, limits
 - Expectations of users to find analogous materials together in a single source, e.g., "book-like" content in digital form with like analog content (ebooks, electronic conference proceedings, government documents, and more importantly, reference genres such as directories, dictionaries, encyclopedias). Genre categories outlined in the OpenURL standard¹⁸ may be used to help make decisions on the kinds of digital objects that would get analogous treatment.
 - Richly described collection level records can serve as pointers to item level descriptions that, for whatever reason, are maintained outside of the ILS (e.g., the item-level descriptions in Web guides, MODS records, other finding aids)
 - Business interdependencies required in an integrated system, e.g., purchase orders, paying of invoices
 - Controlled subject access supported by systems that have implemented thesaural controls using authority records that lead searchers from "used for" and "see also" terms, as well as assisting the user to navigate hierarchical relationships¹⁹
 - Mechanisms to indicate resources truly acquired by the Library, e.g., the Library stores the resource or has made archival arrangements with trusted partners
 - Resource discovery from federated searching/metasearch applications capable of targeting bibliographic databases using protocols such as Z39.50.

MARC/AACR cataloging is a resource intensive proposition requiring trained staff-- the Cataloging Directorate is fortunate to have a large, highly-trained cadre of cataloging staff conversant in the application of AACR and MARC, although most have not yet applied their skills to digital content (efforts are currently underway to expand the staff pool capable of cataloging electronic resources). For true monographs (i.e., not integrating resources coded as monographs), records have been built at both "full" and "core" levels. Because there is currently no core level standard for integrating resources, all records for these resources represent "full" level cataloging. The subgroup concludes that not all resources require full-level cataloging treatment in order to provide the required bibliographic control and access. Following on an objective found in the Report of the Metadata Policy Group

¹⁸OpenURL genres such as: Article, Conference Proceeding, Book, Report, Document, Dissertation; a full list is available at:
http://library.caltech.edu/openurl/StandardDocuments/KEV_Guidelines-20031112.pdf

¹⁹For more detail, see the article "Why LC subject Headings are More Important than Ever" by Thomas Mann (American Libraries, October 2003, Vol. 34, Issue 9, p. 52).

that "existing models, standards, and practices should be adopted, adapted, and extended rather than inventing new ones,"²⁰ the subgroup proposes that a new level of cataloging ("access" level) be developed, within the MARC/AACR context (i.e., "adapting" the current model), to provide a more appropriate level of control to some digital resources at a lower cost than full cataloging treatment. This is one of several recommendations related to MARC/AACR cataloging we make below to help reduce the costs of cataloging.

In connection with its charge, the subgroup also reviewed information on INFOMINE²¹, a cooperatively built virtual library of both free and subscription Internet resources useful for research at the university level, created and hosted by the University of California at Riverside. As elaborated on below, this project was of particular interest to the group for several reasons:

- its focus on scholarly Internet resources
- its cooperative development by librarians at multiple institutions
- the designation of "editor" librarians to create "expert" records as well as to oversee the development and maintenance of each of the broad subject areas
- its ability to import MARC records into its database and map from them to a simplified record display, but which uses subject terms based on LCSH
- the use of sophisticated crawling software to identify additional records of scholarly interest beyond those identified by librarians, an identification based in part on the number of times the resource is cited in "expert created" resources as well as on the number of times the resource is used by INFOMINE users.
- the availability of INFOMINE records for OAI harvesting
- its potential as a "collection development" tool for identifying scholarly Internet resources which already have been brought under some bibliographic control.

Presently there are approximately 109,826 resources described in the INFOMINE database, including databases, electronic journals, electronic books, bulletin boards, mailing lists, online library catalogs, articles, and directories. Approximately one-quarter of these have been selected and described by subject editors, the remaining resources have been identified through the use of an open

²⁰Report of the Metadata Policy Group, p. 35.

²¹ See <http://infomine.ucr.edu> for more information; cooperating partners are: University of California, Wake Forest University, California State University, the University of Detroit - Mercy. The Library of Congress Business Reference Section in conjunction with the Bibliographic Enrichment Advisory Team (BEAT), is currently involved in a pilot project with INFOMINE to supply MARC records for resources cataloged for the BEOnline project to INFOMINE, and to review existing INFOMINE business and economic records for addition to the BEOnline project.

source software product, iVia²². Frequently cited and/or often consulted resources are flagged for attention by editors for possible upgrading to full expert-created records. Users can choose whether to search both expert and robot created records or focus only on expert created records. Recommendations specifically related to INFOMINE are included below.

MARC/AACR Recommendation 1: In addition to the use of “full” cataloging for selected resources (e.g., resources of high research value), devise a new level of cataloging within the MARC/AACR context for a subset of digital content that de-emphasizes certain descriptive cataloging fields, and emphasizes subject/content-oriented fields. This new level would increase the chances for users to find the record through subject and keyword searching, and thus access the resource. For a rationale and more complete description, see Appendix C.

MARC/AACR Recommendation 2: For resources recommended by reference staff for MARC/AACR treatment, require the presence of a summary note constructed by the recommender who has already analyzed the site and, as a subject specialist, may be in a better position than the cataloger to describe the resource²³. (Note that this has implications for adjustments to the TrackER system (Goal 4, Workgroup 4, Workflow) and for training recommending and selecting staff (Goal 5, Initiative 1, Implementation.)

MARC/AACR Recommendation 3: For resources (primarily textual) that need to be cataloged yet lack a summary (e.g., from Acquisitions or other sources), consider whether text analysis tools might be available that are capable of analyzing the content of a Web resource and suggesting possible summary statements and keywords that the cataloger can incorporate in the record. (Implication for Goal 4, Workgroup 3, Tools analysis.)

MARC/AACR Recommendation 4: For systematic identification of records for digital content (e.g., to associate with a project, to export records to collaborative partners), define project identification strings in MARC field 985 (local record history) and use them when appropriate. (Implication for TrackER system, Goal 4, Workgroup 4.)

MARC/AACR Recommendation 5: The Cataloging Directorate should investigate making a commitment to collaborate with selected projects

²² see <http://infomine.ucr.edu/iVia/>

²³It should be noted that many recommenders, primarily those creating/editing bibliographic records in OCLC for use in OCLC Pathfinders, already construct such summary notes. This not only saves the cataloger from constructing the notes themselves, but greatly aids the cataloger in the selection of subject terms.

involved in the identification and control of Internet resources with high research value (e.g., could assist the INFOMINE project by identifying and exporting applicable MARC/AACR records (retrospective and future) for inclusion in the INFOMINE database of scholarly Internet resources, could make LC records available via OAI in the MARCXML format).

MARC/AACR Recommendation 6: Explore how LC recommending and selecting officers could exploit existing research tools (e.g., the INFOMINE database) as a collection development source for identifying valuable research-oriented Internet resources to receive MARC/AACR treatment.

MARC/AACR Recommendation 7: Investigate how LC may be able to exploit the resource descriptions developed by other projects (e.g., see if the OAI-exposed records from the INFOMINE database might be converted to MARC or MODS). (Implication for Goal 4, Subgroup 3, Tools analysis.)

MARC/AACR Recommendation 8: Investigate the feasibility of using the iVia open source software provided by INFOMINE in the LC environment as a tool to generate resource descriptions. (Implication for Goal 4, Subgroup 3, Tools analysis.)

MARC/AACR Recommendation 9: Provide indirect access through MARC/AACR records in the ILS to all instances where individual descriptions are handled by another mode of bibliographic control. These records will guide users from the ILS to the more complete resource descriptions (e.g., Web guides, collections of MODS records)²⁴.

MARC/AACR Recommendation 10: When the Library is in a position to process metadata received as part of the digital objects themselves, from copyright registrations, or through available metadata packages (in XML, RDF, METS, MPEG-7, ONIX or other "known" formats), analysis should be done to determine the effectiveness of converting the supplied metadata to MARC (using tools developed by NDMSO) for use as the basis of a record in the ILS.

MARC/AACR Recommendation 11: Examine the feasibility of generating "use statistics" that would indicate the extent to which users access electronic resources from MARC/AACR records in the catalog.

²⁴Report of the Metadata Policy Group, p. 23 "At a minimum, access will be made available through the Library of Congress online catalog to the reference tools or finding aids or collection guides or general collection-level records to reach all digital and non-digital resources."

Recommended Maintenance Models

Web guides

The responsibility for maintaining LC-generated Web guides is primarily in the hands of the reference staff that create them. For Web guides that reside in LC Web space, webmasters routinely receive link-checker error reports when URLs embedded in guides are problematic. Webmasters also receive error reports for those Web guides that reside elsewhere (e.g., OCLC Pathfinders). Reviewing and acting upon the error reports is a necessary but resource intensive activity.

In addition to automated link checking, Web guides must also be routinely reviewed for content: do the cited resources still present the content appropriate to the guide for which they were selected? Are there newer, better resources for addition to a guide, etc.? Web guide creators/editors must also be aware that if they remove resources from their guides that have counterpart records in the ILS, or if the focus of the guides shifts enough to warrant changes in the cataloging record for the guide, notification must be made to Cataloging.

Web guide maintenance recommendation 1: Develop a communication/notification mechanism for Web guide creators to report to Cataloging when ILS records are affected by Web guide changes and to define these conditions for reference staff. (Implication for Goal 4, Subgroup 4, Workflow.)

Web guide maintenance recommendation 2: Consider adding a button to Web guides for users to report to the webmasters any problems they encounter with sites listed in the guides.

MODS records

Because the MODS records created for the individual websites archived by the MINERVA project are item-level records (in most cases, the description covers multiple iterations of the websites), the concept of maintenance currently has less applicability than with records created for dynamic resources where descriptions need to be kept up to date as resources change. (Note: the same may not be true of other MODS pilot implementations in the Library.) As the MINERVA pilot progresses, the impact of changes to authoritative name or subject headings used on MODS records should be assessed (i.e., if a subject heading changes, will an attempt be made to ripple the change through MODS records?).

MARC/AACR records

Like serials (for which a decision to catalog often results in a long-term commitment to maintain records as the resource changes), it is clear that many

integrating resources that receive MARC/AACR cataloging change at some point after original resource description activities have been completed-- some, obviously, more frequently than others. Although staff in the Cataloging Directorate have been creating MARC/AACR records for digital content for many years, there has been no systematic effort to review these records to determine whether the original description still accurately reflects the resource²⁵. Problems with some records have been reported on an "as encountered" basis by reference staff, catalog users, consumers of distributed records, etc., but we have no accurate measures for determining whether records should routinely be reviewed for accuracy given the potential volatility of the sites they describe. Neither do we have any way to predict the costs associated with such a review. Routine checking of the URI links can be done in the short term, but until LC actually implements systematic URI checking, we again are unable to estimate the resources that would be required to analyze the link checking reports, adjust records, etc.

MARC/AACR maintenance recommendation 1: Establish a mechanism to routinely check the links on records in the catalog²⁶.

MARC/AACR maintenance recommendation 2: In order to provide basic information regarding record maintenance needs, prepare a study of a sample of existing records to determine the degree to which the records no longer accurately reflect the resources they describe. A well-designed study should be able to generate the information needed to determine the necessary policies for ongoing record maintenance. For example, the study may help determine whether a more systematic approach to maintenance is warranted, whether certain "errors" are acceptable without change, whether automated tools may be required to monitor record content, or whether some categories of resources are too volatile to be described effectively (or economically maintained) by MARC/AACR records.

Part 2: Digital Counterparts to be brought under bibliographic control by the Cataloging Directorate

This discussion of digital counterparts is divided into two parts: first, LC²⁷ digitized books or other printed materials from the LC collections; second, digitization done by other organizations.

²⁵Business-related resources are more likely to have been reviewed post-cataloging, due to efforts to identify the subset of records to be exported to INFOMINE; it should be noted that the 2003 survey by the PCC indicates that 72% of libraries do not routinely review records to see if they still reflect the item cataloged.

²⁶A similar proposal has already been made regarding URI links on serial bibliographic records.

²⁷"LC" here is loosely defined to include digitization done by LC staff or by contractors acting on LC's behalf.

LC Digitization of materials in LC collections

The Library of Congress has digitized several thousand books, primarily from the custodial areas of the General Collections, Law Library, Music, and the Rare Book and Special Collections Division (the major categories, with examples, are identified in Table 1 below). For the vast majority of these books, there is already an extant bibliographic description in the ILS for the book that was digitized. Since 1997, the policy for bringing the digital manifestation under bibliographic control has been to add details for the electronic manifestation to the record for the analog material²⁸.

A group of interested stakeholders from a variety of Library Services units, the Law Library, and the Office for Strategic Initiatives met in 2002 to identify the goals for bringing LC-digitized content under bibliographic control:

- LC ILS OPAC users who encounter bibliographic records for original manifestations (print) will discover that a digital manifestation exists (and can access the resource directly via an 856 field link); in all likelihood, service of the digital manifestation will help protect the analog materials-- fewer call-slip or ILL requests, reduced photocopying, etc. (supports user task of "obtain")
- Bibliographic records for original manifestations from the ILS, enhanced to include information for digital manifestations, can be used as metadata to drive the American Memory application whenever possible (i.e., exported MARC records used in the InQuery/Aurora system); this eliminates the need for duplicate metadata generation and maintenance (supports user tasks of find, identify, select)
- LC staff will be aware of the presence of a digital manifestation, eliminating the risk, expense, and "wear and tear" of digitizing the same work more than once
- Bibliographic records can be distributed via CDS to inform other institutions that LC has digitized the work, allowing other institutions to avoid the expense of digitization and focus limited preservation resources on other titles. LC is participating in the development of guidelines for a DLF-sponsored activity: "Digital Library Federation Registry of Digital Masters"²⁹

²⁸For specific guidelines, see section B19.5 "Electronic manifestation--Original in non-electronic form" of the *Descriptive Cataloging Manual (DCM B19)*.

²⁹See <http://www.diglib.org/collections/reg/reg.htm> for more information.

- Metadata sets (as MARC, MODS, or Dublin Core records) for American Memory collections, etc., could be “exposed” to harvesting partners (OAI, DLF, RLG, etc.)

Unfortunately, due to the fact that the materials being digitized are generally “old” (e.g., of historical significance for American Memory, in the public domain), so are the extant bibliographic records-- for books this means that most of the records are from the old MUMS PreMARC file. The quality of these records is notoriously bad: inaccurate data, inaccurate content designation, abbreviated descriptions (most notes and many other fields were not transcribed), invalid headings that are not synchronized with authority records and headings in the rest of the catalog.

Merely adding links to the digital content on the records may suffice to provide access to the resources from the ILS (presuming the record itself can be retrieved), but this does not provide for the other goals identified above (e.g., cannot be used to supply metadata to InQuery/Aurora (American Memory, Global Gateways, etc.), cannot be distributed by CDS, cannot be exposed for harvesting or reporting to a registry). Re-cataloging from scratch under current cataloging rules is not feasible; however, “upgrading” the PreMARC records from the printed source cards and modernizing the headings can be accomplished efficiently and is the model currently in place (at least for resources that were cataloged and represented in PreMARC). In order to formulate an effective workflow for record upgrade and adding of links, cataloging staff must have access either to the analog materials that have been/will be digitized, or access to the images post-digitization-- the approach may vary depending on deadlines, etc.

Digital counterpart recommendation 1: See specific recommended methods for bibliographic control in Table 1. below.

Digital counterpart recommendation 2: Investigate need for “mode of expression” identifier (a high-level marker used to indicate categories of resources) in LC ILS records pulled for use in American Memory application to facilitate record handling, indexing, etc. in American Memory.

Digital counterpart recommendation 3: For categories of digitization below that have not yet been candidates for OAI harvesting, investigate the mechanisms for retrospectively adding these records to LC’s OAI offerings.

Digital counterpart recommendation 4: As plans for the DLF-sponsored registry for digital masters evolve, commit to LC participation by adjusting and sending records to the registry when appropriate. Participation will require planning for implementation that will affect workflow, and record content (e.g., this may involve expanding use of the 583 field (Action note) and other massaging of records in order to fit the DLF-developed guidelines, either at time of record upgrade or via conversion programs after record extraction from the ILS).

Digital counterpart recommendation 5: Assure that the Cataloging Directorate participates appropriately in the “plan” process of the Digital Life Cycle Framework to provide accurate estimates of bibliographic control costs during the approval process for proposed digital conversion projects.

Table 1. Digital Counterparts made by LC

<p align="center">Categories of Digital Counterparts --Digitization by LC--</p>	<p align="center">Recommended Method of Bibliographic Control/Access</p>
<p>LC digitization of book materials featured in LC digital collections (e.g., American Memory, Global Gateway, and other cooperative endeavors)</p> <p><i>Examples:</i> 07035019 (<i>Slaves and the courts</i>) 02001704 (<i>Travel narratives</i>) ca 30000742 (<i>The Capital and the Bay</i>) 05003696 (<i>Meeting of the Frontiers</i>)</p>	<p>Add details of electronic manifestation to ILS record for original book (DCM B19); often involves upgrade of PreMARC records</p> <hr/> <p>Export ILS records for InQuery/Aurora search interface and selectively expose the metadata via the Open Archives Initiative* (http://memory.loc.gov/ammem/oamh/)</p> <p>*not a CATDIR responsibility</p>
<p>LC digitization of book materials for digital delivery of ILL</p> <p><i>Examples:</i> 33008506 ca 22000067</p>	<p>Add details of electronic manifestation to ILS record for original book (DCM B19); often involves upgrade of PreMARC records</p> <hr/> <p>A-Z Web list (maintained by CALM) (http://lcweb.loc.gov/rr/loan/illscanhome.html)</p>
<p>LC digitization of book materials for preservation/surrogate purposes</p> <p><i>Examples:</i> 08020989 (<i>PRD</i>) 87631686 (<i>microform guide digitized by PRD</i>)</p>	<p>Add details of electronic manifestation to ILS record for original book (DCM B19); often involves upgrade of PreMARC records</p>

<p align="center">Categories of Digital Counterparts --Digitization by LC--</p>	<p align="center">Recommended Method of Bibliographic Control/Access</p>
<p>LC digitization of book materials for research or reference needs (RBSCD, PRD), e.g., individual requests by researchers, exhibits, special requests for Library officials for donor relations, educational presentations, and media requests</p> <p><i>Examples:</i> 65059243 (<i>Rosenwald</i>) 90210592 (<i>7 leaves of a mss.</i>) 01001768 (<i>MRR exhibit</i>)</p>	<p>Add details of electronic manifestation to ILS record for original book (DCM B19); generally involves upgrade of PreMARC records</p> <hr/> <p>A-Z Web lists (maintained by RBSCD) (http://lcweb.loc.gov/rr/rarebook/digitalcoll.html)</p>

Digitization of materials by other organizations

Appropriate bibliographic control for digital counterparts when the digital manifestations are made/held by organizations other than the Library of Congress may, in some cases, require different models than LC digitization. For example, the first item in Table 2. below is a BEAT project that, via automated means, adds details of publicly available electronic manifestations made by other organizations to LC book records. Although most of the target records would be candidates for the PreMARC upgrade process described above, this task becomes less critical for these materials-- upgraded records are not required for the purpose of supplying metadata to American Memory, digital registries (that would be the responsibility of the digitizing agency), etc. When staff resources are not available to take the extra steps of PreMARC upgrade for such categories, the upgrade should not be required.

Purchased/licensed access to electronic manifestations of resources held in LC's collections pose additional challenges. If we buy an aggregation of electronic content should we add links to that content from our analog records³⁰? If we don't have analog records for all titles in the aggregation should we make them (or acquire them, if possible, from the vendor)? Should we make distinctions between purchases of "permanent" access and "short term" access (e.g., access allowed only if we continue to pay subscription costs)? Should bibliographic control be considered for "trial" subscriptions or wait until permanent purchase decisions are made? Because the bibliographic control/access issues and decisions are likely to vary from aggregation to aggregation based on varying factors, it is difficult to recommend a "single" approach that would apply across the board.

³⁰Anecdotal evidence supplied by Abbie Grotke based on the trial subscription of netLibrary e-book content suggests that the resources are used more if pointers to the content are in the OPAC.

Digital counterpart recommendation 6: Decisions to purchase aggregated content should be made in the context of the digital life cycle framework³¹-- the planning process should include identification of the appropriate "describe" process. The goal should be to maximize access while balancing costs. Factors that might influence the degrees of bibliographic control may include (but are not limited to):

- quantity of the aggregated content
- availability of readily adaptable metadata (from existing LC or vendor-supplied records)
- alternative methods of access supplied by content providers (e.g., can a vendor-supplied search/retrieval system be a target of an LC-purchased federated searching tool?).

Digital counterpart recommendation 7: Develop mechanisms by which reference staff (on an ad hoc basis) can notify Cataloging to add links to existing ILS records for print materials when they become aware of freely available digital manifestations.

Table 2. Digital counterparts made by other organizations

Categories of Digital Counterparts --Digitization by institution other than LC--	Recommended Method of Bibliographic Control/Access
<p>Non-LC digitization of books also in LC collections (e.g., books digitized by other libraries)</p> <p><i>Examples:</i> 06002464 (<i>digital copy of microfilm held by Indiana Univ.</i>) 12014838 (<i>digitized by both Univ. of Michigan and Indiana Univ.</i>)</p>	<p>BEAT project adds link to LC ILS record for original book (DCM B19 guidelines followed for delineation techniques, but PreMARC records not upgraded)</p>

³¹This recommendation echoes a statement in the Metadata Policy Group Report (p. 36): "when selecting materials, the associated metadata costs need to be considered, since gathering and maintaining metadata can be a significant expenditure."

<p align="center">Categories of Digital Counterparts --Digitization by institution other than LC--</p>	<p align="center">Recommended Method of Bibliographic Control/Access</p>
<p>LC-purchased access to aggregations of digitized books</p> <p><i>Examples:</i> 70255868 (<i>digitized book available as one of 292 titles in "African American biographical database"</i>) 87880137 (<i>digitized rare pamphlet available as one of 36,000 titles in Evans digital edition of early American imprints</i>)</p>	<p>No action to date; link added to holdings for one title at request of reference librarian because all LC copies are missing in inventory</p> <p>Possible to "target" provider's database through federated searching?</p>
<p>Trial subscription ebooks in proprietary formats (e.g., netLibrary where print book held by LC)</p>	<p>No action to date, will need to address in future if "trial" status changed to permanent subscription</p> <p>Possible to "target" provider's database through federated searching?</p>
<p>Simultaneous (more or less) publication of publicly available digital counterparts to print manifestations collected by LC (e.g., GPO, other publishers), made known at time of cataloging print or subsequent notification</p> <p><i>Examples:</i> 2002727923 (<i>GPO PURL</i>) 2003431497 (<i>CIA document</i>) 00024341 (<i>print book with expanded footnotes only available on publisher's Web site</i>) 99058555 (<i>print book with full text available on author's Web site</i>)</p>	<p>Add details of electronic manifestation to ILS record for book (guidelines on when to add links to be prepared by CPSO)</p>

Appendix A: Portal Implications

Library portals come in many flavors and the variations appear to defy a single definition that is not so abstract to be unhelpful³². The one common feature is the intent to provide a form of one-stop shopping for library patrons. At one extreme are portals that allow users (or groups of users) to build their own customized web pages that provide convenient access to the library resources and services used most frequently. At the other extreme are portals offering federated search of a fixed set of bibliographic resources. Most portal applications are somewhere in between, supporting some federated search capability and some degree of personalization or customization by users.

It is for the federated search capability of any future portal developed by the Library of Congress that the deliberations and recommendations of this working group might have impact. Federated search, while a useful tool for exploring the broad information universe, is unlikely to be the panacea many hope for. In practice, many expert researchers find federated search services somewhat frustrating for four main reasons:

- variation in descriptive practice,
- variation in index configuration and query processing of the different search systems;
- slow performance as a federated system waits for results from many sources and merges them into a single results set or display;
- and the availability of only a limited set of query and navigation capabilities in the portal compared to those found in powerful native interfaces.

The Library of Congress Portal Applications Interest Group (LCPAIG³³) has explored portal products designed primarily to provide integrated, cross-database searching of a local catalog, other library catalogs, licensed or locally-mounted full-text and abstracting/indexing databases, and public domain or publicly accessible abstracting and indexing services. In the Library of Congress context, such a portal would provide service to library and congressional staff and onsite patrons. Another form of portal for the Library of Congress that has been discussed would focus on remote public access to all of the resources that the Library provides, including the Library's bibliographic services, searchable full-content resources such as American Memory and THOMAS, exhibits, online publications, and all the rich resources presented on the Library's Web site. There will be benefit to LC users of either of these portal possibilities from some degree of consistency of bibliographic description whether the records are in MARC, MODS, or some other format.

³²Mary Jackson quotes some definitions and one dream for library portals in "The Advent of Portals." *Library Journal*. 9/15/2002.
<http://libraryjournal.reviewsnews.com/index.asp?layout=articleArchive&articleid=CA242296%20&publication=libraryjournal>

³³see <http://www.loc.gov/catdir/lcpaig/> for more information

However, inconsistency of indexing and retrieval capabilities among the ILS and other search systems in operation at the Library, will also detract from a satisfactory experience when searching across resources. And, because any portal will offer a small number of query types, there is no point in expensive enforcement of a degree of consistency from which users will reap no benefit.

Potentially independent of any particular portal service is the increasingly popular function of context-sensitive linking among independently managed licensed resources. This functionality is supported by the proposed OpenURL standard (ANSI/NISO Z39.88-2003. Practices for describing monographic publications, articles in journals and proceedings, technical reports, etc., should take into account what is necessary for generating a "good" OpenURL.

Appendix B. Thoughts on Dublin Core³⁴

The Dublin Core Metadata Element Set³⁵ standardized through NISO (Z39.85-2001) is a very simple "language" for describing resources; it has fifteen elements, all optional, all repeatable, and no guidance as to content for an element beyond its definition. The concept behind it was to be able to support basic discovery across heterogeneous content from different domains. Indeed, it has often been described by those closest to its development as a pidgin.

Quote from standard: "The simplicity of Dublin Core can be both a strength and a weakness. Simplicity lowers the cost of creating metadata and promotes interoperability. On the other hand, simplicity does not accommodate the semantic and functional richness supported by complex metadata schemes. In effect, the Dublin Core element set trades richness for wide visibility. The design of Dublin Core mitigates this loss by encouraging the use of richer metadata schemes in combination with Dublin Core. Richer schemes can also be mapped to Dublin Core for export or for cross-system searching."

In practice, "simple" Dublin Core, as the version that was standardized is often called, is inadequate for almost any specific application or context apart from cross-domain discovery. Although the LC/Ameritech project guidelines allowed Dublin Core for bibliographic description, only one awardee institution found simple Dublin Core adequate, and when awardees mapped richer metadata to simple Dublin Core, the mapping would lose distinctions that are important in American Memory (such as explicitly tagged roles for Creators other than Authors). Rather, the American Memory team requested the richer descriptive records and mapped them directly to a form appropriate for use in American Memory. Requirements to support the level of description found valuable in American Memory (but using a local record format when MARC records were not available) were part of the input to the design of MODS.

There is one context in which LC is using simple Dublin Core. Records exposed for harvesting using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), primarily records from American Memory or the Prints & Photograph Division are exposed in simple Dublin Core. Since cross-domain discovery is one objective of that initiative, simple Dublin Core is the mandatory default schema for records. LC also makes records available for harvesting in MARC(XML) or MODS.

³⁴Through Rebecca Guenther (as member of the DCMI Usage Board), and to a lesser extent, Caroline Arms (OSI), the Library of Congress has been actively involved in the Dublin Core Metadata Initiative since 1995.

³⁵For more information on Dublin Core, see <http://dublincore.org>

As a response to expressed needs for richer metadata, the Dublin Core Metadata Initiative approved what was called "interoperability qualifiers," also known as "qualified Dublin Core." This effort provided the ability to make the elements more specific, by defining "element refinements" and "encoding schemes." Qualified Dublin Core has not been standardized by a format standards process, and the DCMI has been slow in providing tools for implementation; guidelines and examples are generally lacking, and a draft XML schema was only recently made available. Qualified Dublin Core layers complexity onto something quite simple, which has led to problems. Whether using simple or qualified Dublin Core almost every institution has to make adjustments to suit local needs, limiting its potential value for interoperating and exchanging records. The result of that and the lack of guidelines is "the reinvention at every new use, and in every adopting community, of a local set of instructions, a sorry waste of time, as well as a guarantee of inconsistent use in the application."³⁶ In addition, it is a flat list of elements which does not allow for any parsing of data within an element or relating one element to the next (e.g., a place, publisher and date cannot be associated with one another; if there are more than one of any, they are expressed as flat elements and not related to the appropriate one).

The Dublin Core Metadata Initiative (DCMI) has the concept of "application profiles." An application profile can refine or extend Dublin Core in ways consistent with the underlying Dublin Core data model, which allows for qualifiers and encoding schemes. The elements used in application profiles are subject to ratification by the Dublin Core Usage Board. Communities that develop application profiles are free to establish stricter guidelines for the content of elements. A DC-Libraries working group, under the chairmanship of Rebecca Guenther (LC/NDMSO) developed, over the course of two or three years, a proposed application profile for qualified DC for use by libraries. However, key features of the application profile, namely subelements and roles for Creator and Contributor, were rejected by the Dublin Core Usage Board. The DC-Libraries profile is inadequate for use in American Memory. In addition, the LC AV Prototyping Project and the MusicAustralia project (of the National Library of Australia and ScreenSound Australia) attempted to use the DC-Libraries profile, but found that it did not suit its needs (both decided to use MODS).

The use of Dublin Core in the OCLC Connexion (formerly CORC) system uses a non-standard set of qualified elements in order to provide a round-trip mapping between MARC and Dublin Core. The goal of a round-trip mapping to MARC was incompatible with the designed simplicity of Dublin Core. The resulting records are often unsatisfactory in either format.

Working group 2 does not recommend that LC use Dublin Core for describing

³⁶Martin Dillon, Director of Acquisitions, Libraries Unlimited and Adjunct Faculty, OCLC Institute, from *Portal: Libraries and the Academy*, Vol. 3, No. 2 (2003), p. 138.

the resources it manages. For simple records, when the expense of creating a MARC/AACR2 record in the ILS is not justified, the group recommends the use of MODS because of its higher compatibility with traditional bibliographic description and ability for hierarchical descriptions. The group does encourage the active use of one-way mappings from MARC and MODS to simple Dublin Core, in order to make records available to others for building cross-domain discovery services. Similarly, the Library of Congress should expect to receive digital content files with simple Dublin Core elements embedded in the files. One of the most successful aspects of the Dublin Core Metadata Initiative has been its visibility and adoption as a simple set of descriptive elements suitable for creators to embed in their own documents. For example, the latest version of PDF provides for embedded Dublin Core metadata. Automatic extraction of such metadata into simple, perhaps preliminary MODS or MARC records will be an important way to reduce the cost of bibliographic control, and NDMSO has already developed tools to convert Dublin Core descriptions.

Appendix C. Rationale for a new level of MARC/AACR catalog records-- "Access" level-- for selected digital content

Background

The Cataloging Directorate has traditionally followed a model of reducing record content (and/or the authoritativeness of some elements) in order to reduce cataloging costs. For example, "minimal level" cataloging has been used extensively to provide lower cost cataloging. Minimal level cataloging has traditionally meant that descriptive cataloging elements are generally applied as for higher levels of cataloging in order to support user tasks identified in the Functional Requirements for Bibliographic Records (FRBR) as identifying, selecting, and obtaining. However, record elements that support the FRBR user task of "finding," such as subject/content-oriented fields related to classification, subject headings, and summaries of content have traditionally been greatly reduced and usually eliminated altogether. Although such a model *could* also be applied to digital resources in order to lower the costs of cataloging, in many respects a completely opposite model would be more appropriate for these resources:

- Selective inclusion of Internet resources deemed worthy by LC recommending officers in the ILS with adequate content descriptions will be encountered by searchers serendipitously searching on keywords and subject terms and will be made aware that the Internet resources exist. The key will be to include rich content descriptions (e.g., subject headings, keywords, and summaries³⁷).
- Researchers intent on finding Web based resources (whether "known item" or not) are unlikely to consult a library catalog, preferring to search directly on the Internet, thus "skimming" on some descriptive cataloging elements will result in lower cataloging costs without a loss in discovery and access. Typical descriptive data elements that might be omitted include some variant titles (those that don't provide additional keyword access or critical identification information); certain notes; place, publisher, and date elements (these are often difficult for catalogers' to determine since Web pages generally do not present publication information as consistently as book title pages). Emphasis should be placed on "cataloger judgement" with regard to which critical elements are included.

Although the group considered coding such records as "minimal level," in order to unambiguously distinguish "access" records from a wide variety of past history with minimal level and avoid possible conflicts with the MARC 21 National

³⁷These elements have traditionally been considered a category of "descriptive metadata"; however the draft report *Guidelines on the Structure, Content, and Application of Metadata Records for Digital Resources and Collections* by the IFLA Cataloging Section Working Group on the Use of Metadata Schemas defines them as a separate type "analytical metadata" defined as "information analysing and enhancing access to the resource's contents."

Level Requirements for minimal level records, we recommend creating these "access" records at an "abbreviated" level (MARC 21 Leader/17 = 3), and continue to code them as AACR2 records (MARC 21 Leader/18) with the understanding that the descriptive elements that are provided follow that standard, thus enabling LC to continue to use these widespread standards and other institutions to incorporate the records without difficulty.

Implementation of this recommendation would require:

- the development of record content guidelines that are principle-based and easy for cataloging staff to distinguish from other levels of cataloging
- discussions with recommending/selecting officers to provide a clear understanding of when each level of record would be appropriate (e.g., access records for resources unlikely to be candidates of known-item searching, those where the value of the content would not warrant the expense of full cataloging, those whose content is so volatile that frequent re-description would be necessary; and perhaps fuller records for purchased subscriptions, resources of high research and reference value, those OpenURL genres likely to be resolved by that standard, etc.)
- consultation/notification with the Library's traditional bibliographic control partners and consumers
- adjustments to TrackER (the workflow management tool employed to track electronic resource work requests) so that recommending/selecting officers can indicate to cataloging which level of cataloging is desired
- training for cataloging staff who would create the records
- a pilot project to determine effectiveness of the proposed level (e.g., does it reduce costs, does it provide desired access, can the mode be mainstreamed with regular cataloging or assigned only to certain staff).

Appendix D. Recommendations

Web guide recommendation 1: Investigate how to assure that all such Web guides (including “on the fly” dynamic pages) created by LC reference staff are indexable by Internet search engines thus increasing access to the guides.

Web guide recommendation 2: Investigate the possibility of defining and embedding a core set of metadata tags in all LC-created Web guides (e.g., embed the XML tags from a MODS record, develop guidelines for title construction). Embedded metadata may facilitate access to the guides from a library portal/metasearch facility and/or LC website search engine (Inktomi).

Web guide recommendation 3: Investigate mechanisms to enable users who access a guide from a search engine to be alerted to the presence of other guides produced by the Library’s subject specialists.

Web guide recommendation 4: Encourage any group addressing the possible standardization of Web guide styles across the Library (e.g., by adopting a single middleware product for developing Web guides, or applying standardized content schemes) to consider the impact of such standardization on any portal/metasearch product selected by the Library.

Web guide recommendation 5: Because the Web guides created by LC reference staff represent a resource investment by the Library and are seen as authoritative resources in their own right, each Web guide or cluster of guides should be represented by a record in the ILS that supports subject/keyword access to lead ILS searchers to these valuable resources.

Web guide recommendation 6: Determine if it is possible to generate use statistics to determine how effective Web guides are (e.g., how frequently are they consulted? How often do users link to sites described in the guide? What percentage of the described sites do users link to?).

MODS recommendation 1: The Cataloging Directorate should continue to experiment with MODS as the mode of bibliographic control for large collections of archived Web sites.

MODS recommendation 2: When resource discovery is provided via MODS records, a subject/keyword-rich MARC/AACR record for the thematic aggregation as a whole should be made in the ILS, with a link that leads the user to the system/interface where the MODS records can be searched directly.

MODS recommendation 3: LC should consider exposing all MODS records that describe digital content via the Open Archives Initiative-Protocol for Metadata Harvesting.

MODS recommendation 4: A prototype support infrastructure for MODS should be developed to determine if expanding the use of MODS as a mode of bibliographic control for resources beyond archived Web sites is feasible and cost effective.

Infrastructure components include (but are not limited to):

- tools for MODS record creation and maintenance including Unicode™ support (being developed in NDMSO)
- XML-compatible database software capable of storing and indexing MODS records created across the Library
- tools for searching MODS records, including:
 - the ability to accommodate Z39.50 searching and provide a means to target collections of MODS records via federated searching/metasearching/library portal application (being investigated by NDMSO as part of digital library projects), and
 - tools for generating and evaluating “usage” statistics to determine the effectiveness of MODS records

MODS recommendation 5: Establish Library-wide profiles and best practices for all instances of LC’s implementation of MODS records, that address such issues as: persistent identifiers for records (control numbers), mechanisms to represent rights statements or access rights information, mandatory elements.

MODS recommendation 6: Consider issues related to integrating access to MODS records with other metadata (e.g., analyze the impact on “the catalog” if selected MODS records are converted to MARC and added to the ILS).

MARC/AACR Recommendation 1: In addition to the use of “full” cataloging for selected resources (e.g., resources of high research value), devise a new level of cataloging within the MARC/AACR context for a subset of digital content that de-emphasizes certain descriptive cataloging fields, and emphasizes subject/content-oriented fields. This new level would increase the chances for users to find the record through subject and keyword searching, and thus access the resource. For a rationale and more complete description, see Appendix C.

MARC/AACR Recommendation 2: For resources recommended by reference staff for MARC/AACR treatment, require the presence of a summary note constructed by the recommender who has already analyzed the site and, as a subject specialist, may be in a better position than the cataloger to describe the resource. (Note that this has implications for adjustments to the TrackER system (Goal 4, Workgroup 4, Workflow) and for training recommending and selecting staff (Goal 5, Initiative 1, Implementation.)

MARC/AACR Recommendation 3: For resources (primarily textual) that need to be cataloged yet lack a summary (e.g., from Acquisitions or other sources), consider whether text analysis tools might be available that are capable of analyzing the content of a Web resource and suggesting possible summary statements and keywords that the cataloger can incorporate in the record. (Implication for Goal 4,

Workgroup 3, Tools analysis.)

MARC/AACR Recommendation 4: For systematic identification of records for digital content (e.g., to associate with a project, to export records to collaborative partners), define project identification strings in MARC field 985 (local record history) and use them when appropriate. (Implication for Tracker system, Goal 4, Workgroup 4.)

MARC/AACR Recommendation 5: The Cataloging Directorate should investigate making a commitment to collaborate with selected projects involved in the identification and control of Internet resources with high research value (e.g., could assist the INFOMINE project by identifying and exporting applicable MARC/AACR records (retrospective and future) for inclusion in the INFOMINE database of scholarly Internet resources, could make LC records available via OAI in the MARCXML format).

MARC/AACR Recommendation 6: Explore how LC recommending and selecting officers could exploit existing research tools (e.g., the INFOMINE database) as a collection development source for identifying valuable research-oriented Internet resources to receive MARC/AACR treatment.

MARC/AACR Recommendation 7: Investigate how LC may be able to exploit the resource descriptions developed by other projects (e.g., see if the OAI-exposed records from the INFOMINE database might be converted to MARC or MODS). (Implication for Goal 4, Subgroup 3, Tools analysis.)

MARC/AACR Recommendation 8: Investigate the feasibility of using the iVia open source software provided by INFOMINE in the LC environment as a tool to generate resource descriptions. (Implication for Goal 4, Subgroup 3, Tools analysis.)

MARC/AACR Recommendation 9: Provide indirect access through MARC/AACR records in the ILS to all instances where individual descriptions are handled by another mode of bibliographic control. These records will guide users from the ILS to the more complete resource descriptions (e.g., Web guides, collections of MODS records).

MARC/AACR Recommendation 10: When the Library is in a position to process metadata received as part of the digital objects themselves, from copyright registrations, or through available metadata packages (in XML, RDF, METS, MPEG-7, ONIX or other "known" formats), analysis should be done to determine the effectiveness of converting the supplied metadata to MARC (using tools developed by NDMSO) for use as the basis of a record in the ILS.

MARC/AACR Recommendation 11: Examine the feasibility of generating "use statistics" that would indicate the extent to which users access electronic resources

from MARC/AACR records in the catalog.

Web guide maintenance recommendation 1: Develop a communication/notification mechanism for Web guide creators to report to Cataloging when ILS records are affected by Web guide changes and to define these conditions for reference staff. (Implication for Goal 4, Subgroup 4, Workflow.)

Web guide maintenance recommendation 2: Consider adding a button to Web guides for users to report to the webmasters any problems they encounter with sites listed in the guides.

MARC/AACR maintenance recommendation 1: Establish a mechanism to routinely check the links on records in the catalog.

MARC/AACR maintenance recommendation 2: In order to provide basic information regarding record maintenance needs, prepare a study of a sample of existing records to determine the degree to which the records no longer accurately reflect the resources they describe. A well-designed study should be able to generate the information needed to determine the necessary policies for ongoing record maintenance. For example, the study may help determine whether a more systematic approach to maintenance is warranted, whether certain "errors" are acceptable without change, whether automated tools may be required to monitor record content, or whether some categories of resources are too volatile to be described effectively (or economically maintained) by MARC/AACR records.

Digital counterpart recommendation 1: See specific recommended methods for bibliographic control in Table 1.

Digital counterpart recommendation 2: Investigate need for "mode of expression" identifier (a high-level marker used to indicate categories of resources) in LC ILS records pulled for use in American Memory application to facilitate record handling, indexing, etc. in American Memory.

Digital counterpart recommendation 3: For categories of digitization below that have not yet been candidates for OAI harvesting, investigate the mechanisms for retrospectively adding these records to LC's OAI offerings.

Digital counterpart recommendation 4: As plans for the DLF-sponsored registry for digital masters evolve, commit to LC participation by adjusting and sending records to the registry when appropriate. Participation will require planning for implementation that will affect workflow, and record content (e.g., this may involve expanding use of the 583 field (Action note) and other massaging of records in order to fit the DLF-developed guidelines, either at time of record upgrade or via conversion programs after record extraction from the ILS).

Digital counterpart recommendation 5: Assure that the Cataloging Directorate

participates appropriately in the "plan" process of the Digital Life Cycle Framework to provide accurate estimates of bibliographic control costs during the approval process for proposed digital conversion projects.

Digital counterpart recommendation 6: Decisions to purchase aggregated content should be made in the context of the digital life cycle framework-- the planning process should include identification of the appropriate "describe" process. The goal should be to maximize access while balancing costs. Factors that might influence the degrees of bibliographic control may include (but are not limited to):

- quantity of the aggregated content
- availability of readily adaptable metadata (from existing LC or vendor-supplied records)
- alternative methods of access supplied by content providers (e.g., can a vendor-supplied search/retrieval system be a target of an LC-purchased federated searching tool?).

Digital counterpart recommendation 7: Develop mechanisms by which reference staff (on an ad hoc basis) can notify Cataloging to add links to existing ILS records for print materials when they become aware of freely available digital manifestations.