



**Office of the Librarian
Office of Strategic Initiatives
Library Services**

*The Library is Taking the Right Steps
to Ensure that Its Web Search
Strategy Is an Enterprise-wide
Solution but Needs to Expand Its
Oversight of Other Projects*

Report No. 2013-IT-102
September 2013

FOR PUBLIC RELEASE



UNITED STATES GOVERNMENT

LIBRARY OF CONGRESS

Memorandum

Office of the Inspector General

TO: James H. Billington
Librarian of Congress

September 30, 2013

FROM: Karl W. Schornagel
Inspector General

A handwritten signature in black ink, appearing to read 'Karl Schornagel'.

SUBJECT: *The Library is Taking the Right Steps to Ensure that Its Web Search Strategy Is an Enterprise-wide Solution but Needs to Expand its Oversight of Other Projects*
Report No. 2013-IT-102

This transmits our final report summarizing the results of the Office of the Inspector General's review of the Library's enterprise search strategy. The executive summary begins on page *i*, and complete findings appear on pages 3 to 10.

Based on written comments to the draft report, we consider all of the recommendations resolved. Please provide, within 30 calendar days, an action plan addressing implementation of the recommendations, including an implementation date, in accordance with LCR 2023-9, *Rights and Responsibilities of Library Employees to the Inspector General*, §6.A.

We appreciate the cooperation and courtesies extended by the Office of the Librarian, Library Services, and the Office of Strategic Initiatives during this review.

cc: Deputy Librarian of Congress
Associate Librarian for Library Services
Director, Information Technology Services

▶▶ TABLE OF CONTENTS

▶▶ Executive Summary	<i>i</i>
▶▶ Introduction.....	1
▶▶ Objectives, Scope, and Methodology.....	2
▶▶ Findings and Recommendations.....	3
I. The Library is Taking the Right Steps to Ensure that its Web Search Strategy is an Enterprise-wide Solution.....	3
Recommendation.....	5
Management Response	5
II. Library Services Needs to Determine Whether any Data from its Terminated Project Can Be Migrated into the Library’s Enterprise-wide Strategy.....	5
Recommendation.....	8
Management Response.....	8
OIG Response.....	8
III. The Information Technology Steering Committee Should Expand Its Oversight of Information Technology Projects.....	8
Recommendation.....	10
Management Response	10
▶▶ Appendix A: Timeline of Events	11
▶▶ Appendix B: Management Response.....	13

▶▶ EXECUTIVE SUMMARY

Enterprise search involves the capability to perform a Library-wide search to locate information from all of the Library's websites. A long-standing challenge for the Library is making all of its digital collections easily accessible to the public, researchers, staff, and Congress. For years, the Library lacked policies and procedures for creating and maintaining digital collections in a collaborative environment. We performed this evaluation to determine whether the Library's enterprise search strategy has resulted in efficiencies and potential cost savings.

A summary of our evaluation is below.

The Library is Taking the Right Steps to Ensure that its Web Search Strategy is an Enterprise-wide Solution

The Library has created a Web Governance Board and developed a Web Strategy that is standardizing how the Library's digital collections are presented on the Web. The Web Governance Board's work is addressing an issue we identified in a 2009 report¹ by creating a comprehensive search tool and developing a comprehensive index of the Library's multiple digitized collections and Websites. We believe the Library's search strategy will result in a more effective Web presence.

Library Services Needs to Determine Whether any Data from its Terminated Project Can Be Migrated into the Library's Enterprise-wide Strategy

Library Services' decision to terminate the entire National Library Catalog project after 33 months of development and \$1.25 million did not include an assessment of whether data from the project could be salvaged and migrated to the Library's new Web search platform. We recommended that Library Services collaborate with the Web Governance Board and ITS to evaluate whether there is usable data from the terminated IT project and to develop a plan to migrate any potential data to the new platform.

¹ *Information Technology Strategic Planning: A Well-Developed Framework is Essential to Support the Library's Current and Future IT Needs*, Report No. 2008-PA-105 (March 2009), pgs. 10-13.

The Information Technology Steering Committee Should Expand Its Oversight of Information Technology Projects

The Information Technology Steering Committee (ITSC) missed an opportunity to provide oversight on the National Library Catalog project. The ITSC was created in March 2010, nine months after Library Services purchased software to develop the National Library Catalog. However, the ITSC chose to consider only new (“proposed”) IT projects rather than including current IT projects that would otherwise meet the criteria for oversight. We recommended that the Library determine whether there are any projects that are in development but did not otherwise meet the ITSC’s criteria of being a new project, and include them under the ITSC’s purview.

Management concurred with our recommendations.

▶▶ INTRODUCTION

Enterprise search involves the capability to perform a Library-wide search to locate information from all of the Library's websites and databases. Search results depend on whether the content is publicly accessible (i.e., the Library's homepage), limited to on-site researchers, or available only to staff (i.e., Intranet).

Over the years, the Library has accumulated a vast collection of digital content and data on disparate systems. As a result, users of the Library's main Internet homepage (www.loc.gov) had to navigate through multiple websites in order to find requested information. A long-standing challenge for the Library is making all of its digital collections easily accessible to the public, researchers, staff, and Congress. For years, the Library lacked policies and procedures for creating and maintaining digital collections in a collaborative environment.

In January 2010, the Librarian of Congress created the Web Governance Board "to serve as the coordinating authority for all of the Library's web presence." The Web Governance Board formulated a Web Strategy with three facets (One Library, Core Areas, and Network Intelligence) to take full advantage of the diversity of the Library's diverse collections while identifying opportunities for strategic alignment.

This report includes our evaluation of the Library-wide search strategy.

►► OBJECTIVES, SCOPE, AND METHODOLOGY

The objective of our evaluation was to determine whether the Library's enterprise search strategy resulted in efficiencies and potential cost savings.

We reviewed planning documentation relating to the Library's chosen enterprise search strategy and a project initiated by Library Services that terminated because of the enterprise search platform. We also reviewed relevant reports by the Government Accountability Office (GAO). GAO reports included 1) a 2008 report² issued internally to the Library that identified challenges the Library faced in collecting, preserving, and making available print and digital materials, and 2) the 2013 Annual Report³ on overlap and duplication throughout the federal government that stressed the importance of management oversight when issues span multiple organizations or multiple entities within an organization. We also reviewed our 2009 report⁴ on the Library's information technology strategic planning to see whether the Library has addressed the issue of having multiple digitized collections, spanning multiple Library websites, with no common search and access tools. We also interviewed staff from Library Services, the Office of Strategic Initiatives, and the Web Governance Board.

We conducted this evaluation in accordance with the Council of the Inspectors General on Integrity and Efficiency's *Quality Standards for Inspection and Evaluation*. We planned and performed the evaluation to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our evaluation objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our objectives.

² GAO, *Library of Congress Collections Management-Opportunities to Improve Effectiveness through Digitization-Options to Reduce Costs of Overseas Offices* (September 2008)

³ GAO, *2013 Annual Report: Actions Needed to Reduce Fragmentation, Overlap, and Duplication and Achieve Other Financial Benefits*, GAO-13-279SP (April 2013)

⁴ *Information Technology Strategic Planning: A Well-Developed Framework is Essential to Support the Library's Current and Future IT Needs*, Report No. 2008-PA-105 (March 2009)

►► FINDINGS AND RECOMMENDATIONS

This report presents the results of our evaluation of the Library's enterprise search strategy. Details on the results of our review follow.

I. The Library is Taking the Right Steps to Ensure that its Web Search Strategy is an Enterprise-wide Solution

In 2009, we issued a report⁵ that found the Library had disjointed and disparate information technology (IT) solutions. This resulted in multiple digitized collections, spanning multiple Library websites, with no common search and access tools and no comprehensive index or inventory. We recommended that the Library should implement its IT solutions as a total institution with one voice, and have a plan of execution with implemental details and buy-in from the service and support units.

The Library is taking steps to ensure that its Web search strategy is enterprise-wide. It created a Web Governance Board and stopped development of search solutions that were not the enterprise-wide solution.

In January 2010, the Librarian of Congress created the Web Governance Board "to serve as the coordinating authority for all of the Library's web presence." The Web Governance Board, chaired by the then Chief of Staff (now Deputy Librarian), created the Library's Web Strategy. The Web Strategy includes three facets:

- One Library—The Library will offer Congress and the American people seamless access to its online and onsite collections and services;
- Core Areas—The Library will be the "Go To" place on the Web for three core areas (Congress, National Library, and Copyright); and

⁵ *Information Technology Strategic Planning: A Well-Developed Framework is Essential to Support the Library's Current and Future IT Needs*, Report No. 2008-PA-105 (March 2009), pgs. 10-13.

- Network Intelligence—The Library will complement its established strengths in traditional cataloging and curation with adaptive and algorithmic⁶ approaches to knowledge organization that help people to explore and learn.

In order to achieve the Web strategy, the Web Governance Board created Project ONE to “build and manage a world-class, user-centered web presence that truly reflects the breadth, quality, expertise, and authority of the Library.” Project ONE is partially funded from requested appropriations for cyclical investments in technical infrastructure that increase the Library's base funding by \$15 million for fiscal years 2010 to 2014. Content delivery is one of the three components (others are core technology and content management) for building web-based search, discovery, and presentation functionality to enable users to find and use content with ease and flexibility.

In March 2011, the Web Governance Board hired an expert consultant group to review the Library's overall search strategy and assist with defining a plan and future strategy. The consultant recommended the Library implement a federated search strategy, “whereby a single technology crawls the multitude of repositories or their indexes, builds a single index, and presents an aggregated set of results back to the end user.” The expert consultant group recommended a Solr-based⁷ platform as the technology for the Library's federated search. The consultant's report stated that a federated approach “allows key indexes and technologies that also fulfill non-search-related functions (such as...the MarkLogic XML datastore initiative) to remain in place.” In May 2011, the Web Governance Board Search Task Group formally endorsed the consultant's recommendation and stated, “[a]doption of this recommendation means that the Library will not be using...MarkLogic⁸...as its enterprise search solution.”

⁶ Algorithmic is defined as a sequence of instructions embedded within software used to organize information and rank search results (e.g., by relevance or popularity).

⁷ Solr is an open source enterprise search platform from Apache Lucene.

⁸ A software used for the search function that was in development by Library Services.

The Web Governance Board is now standardizing the Library's digital collections presentations on the Web. We believe the Web Governance Board's work is addressing the problem identified in our 2009 report by creating a common search tool and developing a comprehensive index of the Library's multiple digitized collections and Websites. Because of the Web Governance Board's work, all Web projects in various Library service units have to conform to the adopted enterprise-wide search strategy. We believe that the enterprise-wide search strategy will result in a more effective Web presence.

Recommendation

None

Management Response

Management agreed with our finding.

II. Library Services Needs to Determine Whether any Data from its Terminated Project Can Be Migrated into the Library's Enterprise-wide Strategy

The strategic changes made because of the Web Governance Board's decision to use a different Web-interface solution led Library Services to terminate the entire National Library Catalog project. However, at the time of termination, Library Services ideally should have determined whether the XML datastore contained information that could have been salvaged and migrated into the Solr platform.

In 2006, Library Services issued its draft fiscal years 2008-2013 Strategic Plan with a performance goal to begin development of the next generation of information access utilities that bring together all its information resources and allow seamless and unified access to them. After more than two years of planning, Library Services purchased MarkLogic software in 2009 to provide a single access point for its collections by aggregating all metadata into a central repository. This allowed Library Services to gain intellectual control over its

holding, improve metadata⁹ stewardship and preservation, and allowed for a period of remediation of its metadata and, at the time, deemed as the appropriate solution by management.

In June 2009, Library Services began developing an XML datastore and a public Web-interface, all of which are later termed the National Library Catalog. Library Services was using the National Library Catalog to replace the then current version of the Online Public Access Catalog, an online database of materials held by the Library. The XML datastore was meant to be a metadata and resource repository. The National Library Catalog also included a search, discovery, and display interface to data stored in the XML datastore. Development of the XML datastore and search functionality began before the Library created a Web Governance Board.

In early 2012, after creation of the Web Governance Board, Library Services terminated the entire National Library Catalog project. We determined that prior directions from the then Chief of Staff to stop the development and implementation of the search functionality of the National Library Catalog project led Library Services to terminate the entire project.

In March 2011, the beta version of the National Library Catalog was still in development and near public deployment when the consultant hired to assess the Library's search strategy issued its final report. On April 25, 2011, the former Associate Librarian of Library Services issued a memorandum to the then Chief of Staff stating, "[w]e accept the proposal that the Library use SOLR for federated searching." On April 27, 2011, the Associate Librarian of Library Services sent another memorandum to the Chief of Staff stating that Library Services would continue work on the National Library Catalog using the capabilities provided by MarkLogic to bring disparate data sets together—essentially continuing with the project, including the search functionality.

On May 31, 2011, Library Services requested the Web Governance Board's permission for public release of the beta version of the National Library Catalog, which still included

⁹ Metadata is information about data or other information. A metadata record is a file of information, usually presented as an XML document, which captures the basic characteristics of a data or information resource.

the web-based search functionalities using MarkLogic software and not the Web Governance Board's adopted solution, a Solr-based platform. The Web Governance Board subsequently denied Library Services public release (in June 2011) because it had significant implications for user experience, information architecture, and content management, all of which the Web Governance Board was in the process of defining. However, the Library Services' members sitting on the Web Governance Board did not ask the whole Board, at any time, for clarification of whether the entire project should be stopped or voice concerns as to why stopping the web-interface element would necessitate the scrapping of the more substantial element—that is, the XML datastore.

In March 2012, Library Services terminated the National Library Catalog project, but it is unclear why it terminated the XML datastore element, which was not part of the Web Governance Board's direction. With the National Library Catalog, there was much confusion and misinterpretation of direction on what to stop developing. The Web Governance Board wanted the web-interface and search element of the project to stop in its current form and be replaced with the enterprise-wide solution. In an era of rapid change, many software projects are properly started, well managed, and terminated before completion because their original assumptions have changed.¹⁰ It may have been appropriate to terminate the National Library Catalog project; ideally, however, at the time that decision was made, a disposition plan to migrate data from the XML datastore should have been prepared to support the metadata remediation efforts of Project ONE¹¹.

Library Services' decision to terminate the entire National Library Catalog project occurred after 33 months of development and the expenditure of approximately \$1.25 million to develop the XML datastore and other smaller projects using MarkLogic software. Estimated overhead and decommission/replacement costs brings the total expenditure to \$2.2 million. A project disposition plan that addressed the

¹⁰ Software Management Magazine, *Project Termination Doesn't Equal Project Failure*, by Barry Boehm, University of Southern California

¹¹ Project ONE is a multi-year work plan for executing the Library's Web strategy (published in June 2010).

costs and information technology alternatives was prepared two months after the entire project terminated. Library Services representatives stated that the datasets used to remediate metadata into the XML datastore are now outdated and obsolete and therefore would not be able to be migrated to Project ONE. It is not clear whether there may be other useful data; for example, there are over a million digital book covers stored in the XML datastore that may be useful to Project ONE. Library Services should work with ITS and the Web Governance Board to determine whether any data can be salvaged and migrated into the new platform.

Recommendation

We recommend Library Services collaborate with the Web Governance Board and ITS to evaluate whether there is usable data in the XML database and, if so, develop a plan to migrate the data stored to the new Web search platform.

Management Response

Management agreed with our recommendation. However, management disagreed with the narrative related to Library Services' decision to discontinue the National Library Catalog project. Specifically, it disagreed with our statement that "there was much confusion and misinterpretation of direction on what to stop developing," and stated that there was a clear management decision with the new Associate Librarian for Library Services.

OIG Response

We are not questioning the clarity of management's decision in March 2012 to terminate the National Library Catalog project. At the very least, there was confusion and misinterpretation by Library Services of the Chief of Staff's April 2011 direction—in the intervening months after this direction Library Services continued with development, including exercising an option to incur \$132,000 in annual maintenance costs. Eleven months later the project stopped, and there was no analysis of how the data (and monies spent to date) could be salvaged. The Library also stated that several factors were not included in the report that contributed to the decision to terminate use of MarkLogic, such as the need to redeploy

Library Services staff, a belief that MarkLogic did not provide any unique advantages or cost savings to the Library, and a parallel effort to develop a national library catalog. We saw no evidence of any cost analysis, and the other factors noted were not evident in April 2011 but likely evident 11 months later due to the time that had transpired.

III. The Information Technology Steering Committee Should Expand Its Oversight of Information Technology Projects

The Information Technology Steering Committee (ITSC) missed an opportunity to oversee the National Library Catalog project. In March 2010, the Library created the ITSC to provide project oversight of significant technology investments. The Library's Executive Committee appoints the ITSC and consists of Library leaders who possess information technology knowledge and represent each of the service units with the Chief Information Officer, who is also the Associate Librarian for Strategic Initiatives, serving as its Chair.

The ITSC is responsible for reviewing and analyzing proposed information technology investments that meet any of the following thresholds:

- a. The estimated three year cost for acquisition, development, operation, and maintenance is \$1,000,000 or more;
- b. The software enhancements or expansions in functionality are significant and require complex implementation;
- c. The investment impacts another service unit, either in the initial investment year or in subsequent years;
- d. The investment provides new common technology infrastructure (i.e., systems or services used throughout the Library, such as the telecommunications system, Internet backbone, or the Library's equipment or computer network);
- e. The effort has high visibility (internally or externally); or

- f. The investment is otherwise considered a high risk

In hindsight, the National Library Catalog project would have met most, if not all, of the criteria required for review.

However, the ITSC determined that it would only consider new (“proposed”) projects, not those that were already in development. The Library could have benefited from the oversight. Other projects in development may exist that are not under the purview of ITSC’s threshold policy because they were not considered as new projects. We believe their threshold should be expanded to include projects in development at that time, if they are still in development and not fully implemented.

Recommendation

We recommend the Library determine whether there are any projects that are in development but did not otherwise meet the ITSC’s criteria of being a new project, and include them under the ITSC’s purview.

Management Response

Management agreed with our recommendation.

Major Contributors to This Report:

Kurt Hyde, Assistant Inspector General for Audits

John Mech, Senior Auditor

Walter Obando, Lead Auditor

►► APPENDIX A: TIMELINE OF EVENTS

The sequence of events below summarizes the acquisition, development and communication relating to the Library's decision to terminate the entire MarkLogic project.

- **November 2006**—Library Services issues draft Strategic Plan for fiscal year 2008-2013, which includes the performance goal to “[b]egin development of the next generation of information access utilities that bring together all our information resources and allow seamless and unified access to them.”
- **July 2007**— Library Services performs four-month analysis recommending establishing an XML datastore to explore the aggregation of metadata records.
- **October 2008**—XML datastore project team conducts seven-month pilot project and issues report recommending further development of MarkLogic.
- **March 2009**—Consultant presents OSI with Information Architecture analysis recommending a federated search solution and using XML and other web standards.
- **June 2009**—Library Services purchases MarkLogic software for \$797,435.
- **January 2010**—Librarian creates Web Governance Board.
- **May 2010**—Library Services exercises option year 1 of MarkLogic annual maintenance for \$132,348.
- **March 2011**—Consultants issue report that recommends a federated search approach and endorses use of a Solr-based open source platform.
- **April 7, 2011**—Memo from Associate Librarian of Library Services to Chief of Staff addressing Library Services concerns of March 2011 consultant's report. Library Services' understanding is that the Solr-based federated search would be used when searching across the Library's content and that searches across the National Library's content would be driven by the XML datastore.
- **April 20, 2011**—Memo from Chief of Staff to the Associate Librarian of Library Services stating that the Library cannot pursue a siloed search strategy and that advanced searches of specialized collections would be made available whenever appropriate.

- **April 25, 2011**—Memo from the Associate Librarian of Library Services to Chief of Staff withdrawing Library Service’s concerns and accepting the proposal that the Library use Solr for federated searching. States that the rationale for moving bibliographic data into an XML database is to make it usable with any federated search system and to sustain the bibliographic record for the future.
- **April 27, 2011**—Memo from the Associate Librarian of Library Services to the Chief of Staff submitting Library Services’ official response to the consultant’s report. States that Library Services will continue its work on the National Library Catalog using the capabilities provided by MarkLogic to bring disparate data sets together.
- **May 2011**—Web Governance Board Search Task Group endorses March 2011 consultant’s recommendations.
- **May 2011**—ITS launches new metasearch application using Solr. Development began in 2007.
- **May 31, 2011**—Library Services requests Web Governance Board’s permission to publicly release National Library Catalog.
- **June 15, 2011**—Memo from the Chief of Staff to Library Services stating that the National Library Catalog should not be brought before the Web Governance Board since it appears to be a new platform with more than the capabilities of an open public access catalog. States that it is not advisable for any service unit to develop a separate solution for Library content.
- **July 2011**—Library Services exercised option year 2 of MarkLogic annual maintenance for \$132,348.
- **March 2012**—Newly appointed Associate Librarian of Library Services informs Library Services of discontinuance of the entire MarkLogic project.
- **May 2012**—Library Services prepares detailed cost estimate for replacing MarkLogic.
- **June 2012**—Option year 2 support for MarkLogic ends.

►► APPENDIX B: MANAGEMENT RESPONSE

 LIBRARY OF CONGRESS
OFFICE OF THE LIBRARIAN

MEMORANDUM

DATE September 26, 2013

TO Karl W. Schornagel
Inspector General

FROM Robert Dizard Jr. *RD*
Deputy Librarian of Congress

SUBJECT Comments on draft OIG Report No. 2013-IT-102: *The Library is Taking the Right Steps to Ensure that Its Web Search Strategy Is an Enterprise-wide Solution but Needs to Expand Its Oversight of Other Projects*

This memorandum provides the Library's comments on the findings and recommendations of the subject report. The report primarily addresses the work undertaken to develop a Library-wide search strategy and supporting technology.

Findings:

I. The Library is Taking the Right Steps to Ensure that its Web Search Strategy is an Enterprise-wide Solution (pg. 3)

Recommendation: None

Agree.

The Library's web strategy calls for an enterprise-wide search solution to make the Library's content more accessible. We are seeing the benefits of that strategy for users of our website and are continuing to execute the web strategy for this and its other objectives.

II. Library Services Needs to Determine Whether any Data from its Terminated Project Can Be Migrated into the Library's Enterprise-wide Strategy (pg.5)

Recommendation: We recommend Library Services collaborate with the Web Governance Board and ITS to evaluate whether there is usable data in the XML database and, if so, develop a plan to migrate the data stored to the new Web search platform.

Agree.

Page 1 of 2

Representatives of Library Services will confer with their colleagues on the Web Governance Board, and others as needed, to determine whether there is data or other content in the XML database that is not already available for the Library's website that can be cost-effectively migrated to the new platform and, if so, to develop a plan to do so.

We disagree, however, with the narrative related to Library Services' decision to discontinue the National Library Catalog project. As written, the report indicates that "there was much confusion and misinterpretation of direction on what to stop developing" (pg. 7). In fact, this was a clear management decision after several discussions between the Deputy Librarian and the new Associate Librarian for Library Services, Roberta Shaffer, and Library Services has cooperated.

Several factors, not noted in the report, contributed to the decision to terminate use of MarkLogic within Library Services. An inability to obtain buy-in from other service units, for example, was a major factor in support of termination as the original plan had been to distribute the substantial cost of supporting MarkLogic. Other factors included the need to redeploy Library Services staff from MarkLogic projects to other priorities; a belief that MarkLogic did not provide any unique advantages or cost savings to the Library in comparison with other open source options; and parallel efforts underway to develop a national library catalog.

III. The Information Technology Steering Committee Should Expand Its Oversight of Information Technology Projects (pg. 8)

Recommendation: We recommend the Library determine whether there are any projects that are in development but that did not otherwise meet the ITSC's criteria of being a new project, and include them under the ITSC's purview.

Agree.

The ITSC will consider establishing a process to determine those projects that are in development that the committee should review.

Thank you for the opportunity to provide comments. Please let me know if you have questions.

CC: Roberta Shaffer
Al Banks