PREPRINT

OAI/PMH Metadata Conformance to DLF/Aquifer MODS Guidelines

BY

COLIN KOTELES

B.A., University of Illinois at Urbana-Champaign, 1999
M.S.L.I.S., University of Illinois at Urbana-Champaign, 2002

April 13, 2008

Urbana, Illinois

Abstract

The DLF/Aquifer Implementation Guidelines for Shareable MODS Records were created to facilitate the creation of rich, sharable metadata for use in aggregated digital humanities collections. While guideline creators recognize most data providers do not meet criteria set forth in this document, this study attempts to quantify current levels of conformance to the base requirements set forth by DLF/Aquifer MODS guidelines. By analyzing collections for which MODS records are currently available to OAI-PMH service providers, predictions can be made as to both the nature and extent of future normalization processes required by service providers and the nature and extent of training and education required by data providers wishing to expose MODS records that are useful in a variety of contexts.

Table of Contents

Chapter 1 - Overview 1

Chapter 2 - Methodology 7

Chapter 3 - Results & Discussion 9

Chapter 4 - Conclusions 25

References 27

Appendix A - Repository Data 29

Appendix B - Table Structures 31

Appendix C - Database Queries 32

Appendix D - Result Summaries 53

Appendix E - Detailed Results 60

Chapter 1 - Overview

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) defines a framework for facilitating platform-independent application sharing of metadata between disparate digital collections and centralized aggregators [1]. In the OAI-PMH model, individual digital repositories, or "data providers", make structured metadata available for harvest by "service providers." Service providers aggregate these metadata records in an effort to develop "services that add value to the metadata" [2]. The records OAI-PMH data providers expose are structured XML documents that include administrative <header> and <about> sections as well as one or more descriptive <metadata> sections. While data providers can provide metadata in multiple formats in <metadata> sections, it is required "at a minimum" [3] each record include valid unqualified Dublin Core metadata [4]. The motivation for making unqualified Dublin Core the required metadata format in OAI-PMH-based systems was to reduce the burden on service providers to normalize heterogeneous metadata and to emphasize the focus on resource discovery as opposed to resource description in these systems [5].

While Dublin Core provides a "low barrier" method for data providers to describe resources and expose descriptive metadata to service providers, limitations have been identified in its usefulness in OAI-PMH records. First, unqualified Dublin Core may not be expressive enough to accurately and explicitly describe metadata [6]. This makes it difficult to map Dublin Core to more expressive schema, describe resources with fine granularity, and disambiguate content. Dublin Core also has no mechanisms for declaring the controlled vocabularies or schema used to generate metadata values [7]. Furthermore, all Dublin Core elements are optional and repeatable, which, in many cases allows data providers to use a small set of DC elements to describe resources. This may also introduce ambiguities into metadata. It also means the ways Dublin Core is applied varies greatly among data providers [8][23]. Dublin Core's limited and optional element set, ambiguities created by repeated elements and the inability to declare controlled vocabularies limit "quality" characteristics in OAI-PMH records.

Quality metadata is described by Bruce and Hillman as exhibiting seven important characteristics--completeness, provenance, accuracy, conformance to expectations, logical consistency and coherence, timeliness, and accessibility [9][10]. These characteristics may be measured by analyzing a number of quality indicators found in metadata records. The limitations found in Dublin Core encoded metadata most specifically diminish the completeness, conformance to expectations, and logical consistency and coherence of metadata. Dublin Core's limited element set and granularity (or lack thereof) limits its ability to "completely describe [...] objects." Optional elements may further reduce this completeness--data providers are left to decide for themselves how many Dublin Core elements are required to "completely" describe a

resource. The inability to qualify neither the role of elements nor controlled vocabularies used to generate element content further limits Dublin Core's ability to explicitly and predictably conform to expectations and be consistent within and across collections. When these quality criteria are diminished, the ability of users to find, identify, select and acquire resources is also diminished [10].

Besides being high-quality, aggregated metadata collections may exhibit additional properties that help make collections "shareable." The National Science Digital Library describes sharable metadata as quality metadata that also demonstrates proper context, content coherence, use of standard vocabularies, consistency, and technical conformance [7]. These characteristics make sharable metadata usable and interoperable in a wide variety of contexts beyond the context in which the metadata was created. The loosely defined semantic meaning and minimum usage requirements for Dublin Core elements makes it difficult to maintain coherence and consistency across contexts and collections; collections may implement varying subsets of the full Dublin Core element set or interpret the meaning of DC elements differently. Being unable to declare controlled vocabularies may also diminish metadata consistency and reduce overall use of standard vocabularies. As in the case of quality characteristics, the sometimes ambiguous requirements of unqualified Dublin Core metadata diminishes the shareability of OAI-PMH records. When metadata is not sharable, it demonstrates reduced interoperability and may be more difficult to find in aggregated collections [7]. This fact is at direct odds with the goals of the OAI-PMH.

The Digital Library Federation's Aquifer project--founded to "promote effective use of distributed digital library content for teaching, learning, and research in the area of American culture and life" [11]--has attempted to address Dublin Core's quality and shareability limitations in OAI-PMH systems. DLF/Aquifer promotes and develops digital library best practices, services, and tools while also acting as an OAI-PMH service provider through aggregated search portals such as the American Social History Online [12]. To advance their mission and--indirectly--address the identified quality and shareability limitations in Dublin Core, the DLF/Aquifer project created the *DLF/Aquifer Implementation Guidelines for Shareable MODS Records* [13]. These guidelines outline specific requirements and recommendations for creating high-quality, sharable MODS records describing digital objects intended for distributed use. Resulting metadata created using these guidelines should be "rich, sharable metadata that is coherent and consistent" [13] thus supporting a wider variety of anticipated and unanticipated aggregator services than possible using Dublin Core [7] .

The MODS metadata schema [14] defines an expressive element and attribute set that facilitates explicit, semantically rich metadata description and allows users to declare controlled

vocabularies used to generate element content. The DLF/Aquifer MODS guidelines outline specific rules for creating MODS records--requiring the use of nine specific elements and associated subelements and attribute/value pairs (Table 1.1). The guidelines also define requirements for generating element content and require the use of certain controlled vocabularies in conjunction with certain elements. The DLF/Aquifer's MODS guidelines and the more expressive MODS element and attribute set facilitate such actions as declaring elements conform to W3C Date and Time Format or Library of Congress Subject Heading values or making explicit differentiations between titles and subtitles. These declarations and distinctions are not required or possible with unqualified Dublin Core and having the ability to do so improves metadata shareability.

While the advantages of exposing MODS metadata to OAI-PMH service providers can be reasonably predicted, the DLF/Aquifer project concedes "the requirements and recommendations set forth [in the *DLF/Aquifer Implementation Guidelines for Shareable MODS Records*] are not currently met by most current and potential Aquifer participants" [13]. The goal of this study is to attempt to validate and quantify this statement by determining just how closely data providers currently adhere to DLF/Aquifer MODS guidelines. While the scope of this study necessitated a strict focus on only the nine explicitly required elements (and any additional required subelements, attribute/value pairs, and element content vocabularies), this work could be extended in the future to include recommended elements and attributes. Testing generally involved mechanical tests that quantitatively measured metadata record conformance to the MODS guidelines. And in some cases, additional heuristic analysis also offered qualitative insight.

Measuring current conformance to DLF/Aquifer MODS guidelines not only gives a snapshot of the current "state of the state" of MODS metadata creation, it also provides important insight into the issues service providers will confront when aggregating collections. When aggregating collections, service providers generally need to repurpose metadata to build value-added services. This repurposing first requires the analysis of harvested metadata to determine which augmentation and normalization processes are required to make harvested metadata most useful in the service provider's own context [15] [16]. In keeping with recommended practice [17], this study identified which required elements were present in records (and what percentage of records included required elements), the consistency of values within required elements, and what patterns were present within required element content. These observations were then used to make recommendations regarding required automated reprocessing during service provider ingest or recommend changes to data provider metadata creation practices before metadata is exposed for harvest. This study not only defines how closely data providers conform to

DLF/Aquifer MODS guidelines; it also gives an idea as to what measures can be implemented to bring records closer to conformance.

Beyond practical measurement of current MODS record conformance to DLF/Aquifer MODS guidelines, the results of this study may justify the approach used by the DLF/Aquifer project to enhance metadata shareability. By publishing and promoting a set of "best practices", DLF/Aquifer is hoping to raise awareness of MODS metadata creation practices that in turn will educate data providers to be better metadata producers and provide a roadmap for the creation of shareable metadata. This awareness may also cause service providers to demand and expect higher-quality shareable metadata that they can use to create richer value-added services. However, there are alternatives to this approach. Service providers can develop more powerful, algorithmic processing systems that automatically clean, normalize and augment metadata. These processing systems can manipulate existing metadata or infer metadata directly from an item (such as using file extensions to generate <internetMediaType> values). Another alternative may be to augment existing metadata with additional user generated metadata that can make metadata more useful in new contexts. Both of these alternate approaches can result in metadata that is more shareable. While this study doesn't provide a definitive answer as to which of these approaches is ultimately the "best", the results may provide insight into the viability or effectiveness of these different approaches to creating shareable metadata.

Finally, measuring conformance to DLF/Aquifer MODS guidelines had the unintended result of identifying possible inefficiencies in both the MODS guidelines and in OAI-PMH. Many of the guideline non-conformance factors identified pointed not necessarily to data provider errors, but to redundancies required by either the DLF/Aquifer MODS guidelines and/or OAI-PMH. Instances of non-conformance were often due to simple formatting errors (that could be easily and automatically repaired) and not due to missing, ambiguous or otherwise low-quality metadata. Addressing some of these inefficiencies would reduce administrative overhead on data providers and improve the overall ability of the DLF/Aquifer guidelines and OAI-PMH to facilitate the creation of shareable metadata. So while this study was designed to simply measure conformance of existing MODS records to the *DLF/Aquifer Implementation Guidelines for Shareable MODS Records*, the results can be also used to show the usefulness of different overall approaches to creating shareable metadata and to improve the guidelines themselves and even improve the overall effectiveness of OAI-PMH.

1.1 Element Requirements Summary	
Required Element	**Requirement Summary**

<titleInfo><title>	● Each record must include *at least* one <titleInfo> element with one <title> subelement. ● Instances of <titleInfo><title> are repeatable.
<typeOfResource>	● Each record must include *at least* one <typeOfResource> element. ● <typeOfResource> content must match one of the following values: ● text ● cartographic ● notated music ● sound recording ● sound recording-musical ● sound recording-nonmusical ● still image ● moving image ● three dimensional object ● software, multimedia ● mixed material ● <typeOfResource> elements are repeatable
<originInfo>	● Each record must include *at least* one <originInfo> element that includes a date-related subelement. ● Recommended date-related subelements include: ● <dateIssued> ● <dateCreated> ● <copyrightDate> ● <dateOther> ● One and only one date-related <originInfo> subelement per record must include a keydate="yes" attribute/value pair. ● The <originInfo> element is repeatable.
<language>	● Each record must include *at least* one <language> element when "language is primary to understanding the resource." ● Each <language> element must include the following pair of <languageTerm> subelements: ● One <languageTerm> subelement that includes a type="text" attribute/value pair. ● The content of this <languageTerm> subelement should appear in the MARC Code List for Languages. ● One <languageTerm> subelement that includes both type="code" and authority="iso629-2b" attribute/value pairs. ● The content of this <languageTerm> subelement should be valid ISO 639-2 content.
<physicalDescription>	Each record must include one and only one <physicalDescription> element. ● Each <physicalDescription> element must include one and only one <digitalOrigin> subelement. ● The content of the <digitalOrigin> subelement must match one of the following values: ● born digital

	● reformatted digital ● digitized microfilm ● digitized other analog ● Each <physicalDescription> element must include at least one <internetMediaType> subelement. ● The content of the <internetMediaType> subelement must be a value from the MIME Media Types list. ● This subelement is repeatable.
<subject>	● Each record requires the use of **at least** one <subject> element ● The <subject> element requires the use of *at least* one valid subelement: ● <topic> ● <geographic> ● <temporal> ● <titleInfo> ● <name> ● <genre> ● <hierarchicalGeographic> ● <cartographics> ● <geographicCode> ● <occupation>
<location>	● Each record requires **at least** one <location> element that includes **at least** one <url> subelement. ● One and only one <location><url> subelement must include a usage="primary display" attribute/value pair.
<accessCondition>	● Each record requires **at least** one <accessCondition> element that includes a type="useAndReproduction" attribute/value pair. ● This element is repeatable.
<recordInfo>	● Each record requires one and only one <recordInfo> element that includes one and only one <languageOfCataloging> subelement. ● Each <languageOfCataloging> subelement requires a pair of <languageTerm> subelements. ● One <languageTerm> subelement that includes a type="text" attribute/value pair. ● The content of this <languageTerm> subelement should appear in the MARC Code List for Languages. ● One <languageTerm> subelement that includes both type="code" and authority="iso629-2b" attribute/value pairs. ● The content of this <languageTerm> subelement should be valid ISO 639-2 content. (It should be noted the <recordInfo> example in the DLF/Aquifer Implementation Guidelines for Shareable MODS Records 1.0 does not

	meet this requirement.)

Chapter 2 - Methodology

Overview

A set of existing OAI-PMH records that include descriptive MODS records were harvested from ten different data providers. These records were then processed to extract and organize the individual MODS elements in each record. This processing resulted in a relational database that includes both a table of records and a table of elements in each record (a third, hand-generated table of data providers was also created). SQL queries were then applied to these tables to test conformance to *DLF/Aquifer Implementation Guidelines for Shareable MODS Records* for the nine explicitly required elements in the guidelines.

Test Set

The test set (Appendix A) consists of records harvested from ten different data providers at nine different institutions. Each of the nine data providers listed on the DLF MODS portal as of August 2007 [18] are represented with the addition of records harvested from the Digital Books from the University of Illinois at Urbana-Champaign and the Open Content Alliance collection. In all cases, records were harvested using the Reap OAI Command Line Harvesting Tool [19]. During these full harvests, Reap's metadataPrefix option was set to "mods" so only records containing MODS metadata were harvested. For most collections, the setSpec option was not set so any available MODS records were harvested regardless of the set. However, due to unresolved circumstances, using the setSpec option was required for successful harvest of the Indiana University Cushman collection and the University of Michigan DLPS collection. SetSpec values for these collections can be found in Appendix A. Records were harvested to separate directories for each collection. Harvests were completed between August 30, 2007 and October 27, 2007 and resulted in a set of 343,529 MODS records.

After harvesting, IndexReap (another VBScript-based tool developed by the UIUC OAI Metadata Harvesting Project) was used to extract metadata elements and element content and attribute/value pairs from harvested MODS records. IndexReap extracted this data into a Microsoft Access relational database. To run IndexReap, a table including identifying data for each repository in the test set was first generated by hand. This information was then used to process records on a repository-by-repository basis. Running IndexReap on harvested records resulted in a Records table that includes a row for each harvested record. Each row includes a column each for the corresponding repository ID, a unique autonumber record ID assigned as metadata is extracted (relevant only within the relational database), the OAI identifier, the OAI date stamp, and a column containing the entire XML content of the metadata record. Running

IndexReap also resulted in a Metadata table that contains a row for each element in each metadata record. For each row there is a column for the corresponding RecordID, an autonumber assigned unique ID for each metadata row (again, only relevant within the relational database), the property (or element) name, the parent property, the content of the property and any existing values of "type", "authority", "encoding", "href", "displayLabel", "keyDate", or "usage" attributes. Additionally, there are columns for property and parent property namespace values and the ID of the parent property's metadata row. The complete Metadata table includes rows for 13,770,392 MODS elements. Both the Records and Metadata tables became the basis for mechanical tests to measure record compliance to the DLF/Aquifer MODS Implementation Guidelines. Full structure details of the Repositories, Records, and Metadata tables can be found in Appendix B.

Tests

A series of SQL queries testing for conformance to each element requirement were applied to the Records and Metadata tables (Appendix C). Depending on the extent of a particular requirement, between four and seventeen SQL queries were created to, in aggregate, fully test a record's compliance to each DLF/Aquifer MODS guidelines requirement (and in some cases, additional SQL queries or brief heuristic analysis was applied to test some of the recommendations attached to explicit element requirements). The results of these queries could then point to deficiencies or adequacies in record conformance to DLF/Aquifer MODS guidelines. These SQL queries were applied to each repository separately and totaled so results could be analyzed on the test set as a whole and on a repository-by-repository basis. While intermediate results were stored in temporary, intermediate database, only certain, useful result categories were formatted for inclusion in this study (Appendix E). Formatted results generally display the presence (or lack thereof) of required elements and attribute/value pairs. And in certain instances they display element content values when useful for conformance analysis.

Chapter 3 - Results & Discussion

Results Summary

The results summary (Table 3.1) shows only two required elements--<titleInfo> and <typeOfResource>--are found in 90% or more of records the sample set. Two required elements--<subject> and <accessCondition>--are found in more than 50% of records. The remaining five required elements are found in less than 8% of records.

3.1 Summary of Number of Records with Required MODS Elements				
A	**B**	**C**	**D**	**E**
<titleInfo>	305386	**99.97%**	305386	**99.97%**
<typeOfResource>	281297	**92.09%**	281297	**92.09%**
<originInfo>	9170	**3.00%**	118825	**38.90%**
<language>	366	**0.12%**	165496	**54.18%**
<physicalDescription>	24356	**7.97%**	24356	**7.97%**
<subject>	201162	**65.85%**	201162	**65.85%**
<location>	1130	**0.37%**	281748	**92.23%**
<accessCondition>	179058	**58.62%**	305468	**100.00%**
<recordInfo>	308	**0.10%**	305468	**100.00%**

A – Element name

B – Number of records meeting element requirements

C – Percentage of records meeting element requirements

D – Number of records meeting element requirement after safe algorithmic normalization

E – Percentage of records meeting element requirement after safe algorithmic normalization

Although these numbers may at first seem discouraging, many of the problems can be fixed algorithmically with a high degree of accuracy. These "safe" fixes result in five element requirements being met by more than 90% of records, three elements requirements being met in 53% of records or more and only one element requirement that is met by less than 8% of records. In cases where automatic normalization and augmentation are overly costly or risk invalidating data, it may be required to alter data provider practices to more fully meet DLF/Aquifer MODS guidelines requirements. Detailed results summaries showing element requirement conformance by both repository and by element can be found in Appendix D. Discussion of results for each required element follow.

<titleInfo>

DLF/Aquifer Implementation Guidelines for Shareable MODS Records require each record to include at least one <titleInfo> element that includes a <title> subelement (Table 1.1).

Tests first returned all <titleInfo> elements that contain content and are children of <title>

elements. In Appendix E, Table 1 shows that almost every record includes at least one <title><titleInfo> element. In nine repositories, 100% of records satisfy this requirement while in the one remaining repository all but less than 1% of records satisfy the <titleInfo><title> requirement. Within the full test set, each record includes an average of 1.86 <titleInfo><title> elements per record. While two repositories only average one (one averages 1.01) <titleInfo><title> per record, six repositories average almost two to over two and a half <titleInfo><title> elements per record. Future testing could analyze the use <titleInfo> attributes and values to see if explicit differentiation is made between <titleInfo> elements when multiple titles present in a single record. Brief heuristic analysis of records show that multiple titles are in fact differentiated through the use of such techniques as applying type="alternative" attribute/value pairs in <titleInfo> elements.

Further analysis (Appendix E, Table 2) shows that just under 25% of records include <titleInfo> subelements other than <title>. While seven repositories make use of <titleInfo> subelements other than title, three repositories use only <title>. Appendix E includes a summary of the use of other acceptable <titleInfo> subelements and shows <subTitle>, <partName>, and <nonSort> see the most significant use; although all four other <titleInfo> subelements see significant use in at least two or more collections. While the use or non-use of additional <titleInfo> subelements may point to "completeness" quality in these records [10], it is hard to make definitive judgments about the use and utility of these subelements without detailed and costly hand analysis of each record. It may be the case that title information is sufficiently encoded with a single <titleInfo><title> element in records that do not include additional subelements.

Appendix E also includes a summary of <nonSort> element content. In instances where <nonSort> elements are used, their content generally includes appropriate, recommended values. However, a brief heuristic scan of <title> content showed there are significant numbers of instances where non-sorting characters are present in <title> content and no <nonSort> elements are used. So when used, <nonSort> elements are generally used according to recommendation, but there are instances when they should be used and they are not. This heuristic analysis also showed there are instances where <title> content values are repeated in multiple records. While repeated <title> content may indicate duplicate or overlapping records [15], further analysis of <titleInfo> attribute/value pairs is required to confirm these redundant titles are exposing duplication or ambiguity within records.

A primary identifier for digital objects, title information is a primary mechanism for differentiating items within collections. Title information is also extremely important as a display mechanism for individual records and in results lists where the title is generally the most

prominent information displayed for each item. In these lists, users often rely solely on title information to differentiate objects and may use title information as a primary resource for deciding whether or not an item meets their search success criteria. While almost all records in the test set meet the base DLF/Aquifer MODS implementation guideline requirements for title information, there is still uncertainty as to the deeper quality of this information and how well records adhere to further <titleInfo><title> recommendations. Repeated <title> content and non-use of <nonSort> elements are the two most critical areas for further analysis. While it may be fairly easy for aggregators to extract and properly encapsulate non-sorting characters from <title> content, trying to disambiguate repeated <title> values is difficult for service providers and would be best remediated by data providers who have intimate familiarity with the items in question making them better equipped to supply this more subjective content. But quantifying the exact extent and severity of this issue requires deeper analysis beyond core DLF/Aquifer MODS requirements.

<typeOfResource>

The DLF/Aquifer Implementation Guidelines for Sharable MODS require the use of at least one <typeOfResource> element. The content of this element must be one of eleven values outlined in the MODS User Guidelines (Table 1.1).

Test results (Appendix E, Table 5) show over 92% of all records in the test set and almost 100% or records in nine repositories in the test set include at least one <typeOfResource> element. Results in Appendix E also reveal the content in these elements is always valid as prescribed by the MODS user guidelines. (It may also be interesting to note that over 97% of this content is one of three values from the list of eleven available values: still image, notated music and text). Further analysis also revealed that all but four records missing <typeOfResource> elements were from one repository--Repository 10.

<typeOfResource> content is used to determine a primary categorization of items--often used to create top-level browse indexes. This information can also be used as a primary limiter in user searches and may be used by service providers as a primary determining factor in what mechanisms are used to display a record. If missing, service providers need to implement costly (in terms of processing overhead) and possibly inaccurate normalization routines to try to generate <typeOfResource> information from elsewhere in the record (for instance, from file extension or <physicalDescription> information). But these results show missing or invalid <typeOfResource> elements are not widespread across collections--these elements are only missing from one collection. Therefore, instead of trying to algorithmically repair records, it may be a better solution to ask the data provider to add this information to their exposed OAI-PMH

records.

<originInfo>

DLF/Aquifer MODS Implementation Guidelines require the use of at least one <originInfo> element. Furthermore, one <originInfo> element must include a recommended date-related subelement that includes a keyDate="yes" attribute/value pair. The recommended date-related subelements include <dateIssued>, <dateCreated>, <copyrightDate>, and <dateOther> (Table 1.1).

Appendix E, Table 7 shows that at least 95% or records in every collection includes at least one <originInfo> element with a recommended date-related subelement. However, as indicated in Appendix E, Table 8, only four collections identify a key date using keyDate="yes" attribute/value pairs. Furthermore, in only three collections are there records with single keyDate="yes" attribute/value pairs (in one collection, there are always two keyDates, and in another, only 81% of records include a single keyDate="yes" attribute/value pair). For this reason, a total of only 3% of records conform fully to DLF/Aquifer <originInfo> requirements (Appendix E, Table 9). Further analysis summarized in Appendix E reveals that although <dateIssued> is generally the most frequently used date-related <originInfo> subelement, all other date-related subelements see significant use in at least two collections and <dateIssued> and <dateCreated> are generally the most common elements designated as key dates.

Further analysis of date-related subelement values found in Appendix E shows trends in encoding attribute values. Encoding schema are declared almost approximately 59% of the time for date-related subelements. However, when encoding is declared, the declared encoding schema is only allowable W3CDTF or ISO 8601 schema in six collections. MARC schema, which is not allowed by the DLF/Aquifer guidelines, is declared in three collections. Furthermore, a loose heuristic evaluation of date-related subelement content reveals a significant number of values are valid W3CDTF or ISO 8601 content but there are a significant number of invalid values. These invalid values include both incorrectly formatted single date values and incorrectly formatted date ranges and qualified dates. The DLF/Aquifer MODS guidelines recommend using "point" and "qualifier" attributes to accurately and explicitly encode this information as opposed to representing these values in a single date-related element. (It should be noted the metadata processor used in this study did not extract "point" or "qualifier" attribute values. It would be recommended to accommodate for these values for future processing and analysis.) This visual analysis of date content makes it apparent there is significant non-conformance and variation in date vales.

Date information included with <originInfo> content is generally a primary criteria used to judge relevance of search results. It is also an important category used by service providers to build, organize, and categorize collections. In instances when date content is not homogeneous across aggregated collections, service providers must apply costly processing to normalize a multitude of invalid date values. And in cases where multiple dates are provided without an identified key date, service providers must make difficult, subjective judgments as to which date values are primary to accessing and understanding resources. In both cases, fixing and making sense of uncertain date values is difficult for service providers and a better solution may be to educate data providers so they are better equipped to provide conforming date content. However, it was noted that of the records that include an <originInfo> date-related subelement, 118,825 of these records include only one date-related subelement. It could be inferred then, that this date is, by default, the key date. Service providers could apply a keyDate="yes" attribute/value pair to records with single date-related subelements, thus vastly improving conformance to this element requirement with little effort and a high degree of accuracy.

Future analysis of <originInfo> content may included analysis of granularity of date encoding by evaluating use of "point" and "qualifier" attributes. Also, analysis of publication information by evaluating use of <place>, <publisher>, and <edition> elements would be useful.

<language>

DLF/Aquifer MODS Implementation Guidelines require the use of at least one <language> element (Table 1.1). In one <language> element, a pair of <languageTerm> elements "representing the primary language of the text" [8] is also required. One <languageTerm> element must include a type="text" attribute/value pair. The content of this element should be a value from the MARC Code List for Languages. The other <languageTerm> element must include both a type="code" attribute/value pair and an authority="iso639-2b" attribute/value pair. The content of this element must be valid ISO 639-2b content. It is also important to note this requirement is only necessary for "resources in which language is primary to understanding the resource" [13].

Appendix E, Table 11 shows <language> elements are present in nearly 100% of records in seven collections and 56% of records in one collection "in which language is primary to understanding the resource" (there are almost no <language> elements in two collections that are comprised solely of images or musical scores). The collection in which <language> elements are only present in 56% of records identifies <typeOfResource> values of "still image" for over half of all items--giving an explanation for missing <language> elements. Tests also showed that for every <language> element there exists at least one corresponding <languageTerm> subelement

(Appendix E, Table 12). However, further testing revealed <languageTerm> elements including both required type="code" and authority="iso639-2b" attribute/value pairs are included in only 53.87% of records (Appendix E, Table 13) (one collection includes either no qualifying <languageTerm> attributes or uses only authority=”rfc3066” attribute/value pairs). It was also revealed less than 1% of records include the required <languageTerm type="text"> element (Appendix E, Table 14). Therefore, this means less than 1% of all records meet the full <language> element requirements outlined in the DLF/Aquifer MODS guidelines (Appendix E, Table 15). (It should also be noted that in the cases of both types of required <languageTerm> elements, valid values from prescribed controlled vocabularies are generally used. These values are included in Appendix E.)

<language> element content is often a primary limiter for user searches and allows users to eliminate results in languages they do not understand or return results that meet criteria of language-specific searches. For service providers, <language> content can be used to create language-specific collections or create subsets within collections. Service providers may also use <language> content to render items in specific ways. When <language> content is not explicitly available, user search capabilities and value-added services are limited.

Records in the test set that do not conform to DLF/ Aquifer MODS <language> guidelines can be normalized using relatively simple algorithms. In the majority of cases, records include valid, declared ISO 639-2b <languageTerm> content. They are only missing <languageTerm type="text"> elements. Therefore, the content of the "code" <languageTerm> content can be mapped to valid MARC Code List for Languages values and the missing <languageTerm> elements could be accurately added to deficient records during normalization. And in the case of the collection that provided RFC 3066 language content, this is at least explicitly stated by the "authority" attribute; thus facilitating automatic mapping and normalization of content.

The most difficult aspect of evaluating <language> content (other than whether or not the identified language truly is the item's primary language) is for records that do not include <language> elements, is it truly the case that language is not primary to the understanding the resource? Answering this question requires subjective analysis and careful heuristic testing that is costly to service providers. In the case of <language> content, the data provider needs to be trusted to correctly evaluate the need for <language> declarations.

<physicalDescription>

DLF/Aquifer Implementation Guidelines for Shareable MODS Records require the use of one and only one <physicalDescription> element (Table 1.1). The <physicalDescription> element must include one and only one <digitalOrigin> subelement and at least one <internetMediaType>

subelement. Content of the <digitalOrigin> subelement must be one of four allowable values as defined by the MODS schema [14]. Content of the <internetMediaType> element must be a value from the MIME Media Types list [21].

Appendix E, Table 21 shows that nearly all records in the test set include a <physicalDescription> element. Testing also showed in all but one collection there is only one <physicalDescription> element per record (one collection includes an average of 5.5 <physicalDescription> elements per record). But while almost every record contains a <physicalDescription> element, less than 8% of all records include a <digitalOrigin> element (Appendix E, Table 22). However, in each record where <digitalOrigin> is used, the element is not repeated--as required--and the content is valid (this is the case for nearly 100% of records in four collections). <internetMediaType> subelements are also missing from a significant number of records; appearing in just over 16% of records (Appendix E, Table 23). (<internetMediaType> is used in all records in six of the collections and not at all or in less than 1% of records in four collections--the overall percentage is skewed as one of the collections that includes almost no <internetMediaType> elements is Repository 9--a disproportionately large collection.) In three of the six collections in which <internetMediaType> has significant use, this element is frequently repeated within records. In the remaining three, the element is generally only used once per record. But like <digitalOrigin> content, <internetMediaType> content is generally an accepted value (only Repository 9 includes invalid content). In final analysis, only 7.97% of all records satisfy the full <physicalDescription> requirement (Appendix E, Table 24).

<digitalOrigin> information can be useful administratively for service providers and may be a useful limiter for end-users. It also may give insight to service providers and end users as to the version of the resource they are accessing (are they accessing the original item or a secondary, converted version). Even more important, <internetMediaType> information may be a primary search criteria for user searches and it is extremely important to service providers who will use this information to categorize collections and make a primary determination as to how to display and render an item (for instance, should the item be embedded in a video player, image viewer, or audio player).

Generating missing <digitalOrigin> content is very difficult for service providers. There is no precise way of inferring these values from elsewhere in records; although, these values are somewhat consistent across collections when present in the test set meaning a single value may be valid for an entire collection. While training data providers to consistently provide this information in exposed records, it may be the case that service providers can ask data providers for this data and add the same value to each record in a collection during normalization. And in cases of missing <internetMediaType> elements, it may be possible to generate these values

from elsewhere in metadata records such as file extensions in <location><url> content. However, these processes add substantial overhead and imprecision to normalization processes.

<subject>

DLF/Aquifer Implementation Guidelines for Shareable MODS Records require the use of at least one <subject> element within records (Table 1.1). This s only required "when applicable" [13]. Declaring and using controlled vocabularies to generate subject content is recommended. It is also recommended to use <subject> subelements to describe subject content with increased granularity.

Basic results of testing (Appendix E, Table 29) show that over 67% of all records in the test set include at least one <subject> element. While all repositories average at least 1.75 <subject> elements per record, most average more than two and two repositories average nine or more subjects per record. Among all repositories, there is an average of 2.69 subjects per record. Subject authorities are declared in nine of the ten repositories with nearly 49.15% of all <subject> elements including an authority attribute (Appendix E, Table 30). A total of 49.98% of all records include at least one <subject> element that includes an authority attribute. The most common subject authorities declared are "lcsh", "lctgm" and "local"--together accounting for the vast majority of total values. A small number of invalid authority attribute values and blank values were found in two repositories.

In six repositories, at least 96% of <subject> elements include a subelement (Appendix E, Table 31). In three of the remaining repositories, 66% or more of all <subject> elements include subelements. But in the final repository, less than 27% of <subject> elements include subelements. By far the most commonly used <subject> subelement is <topic> (accounting for more than 60% of all <subject> subelements). All allowable <subject> subelements see at least some sort of significant use except for <occupation> (and while most are used more than 4800 times, <genre> and <titleInfo> are each used less than 700 times).

An informal analysis of <subject> content revealed a large number of repeated <subject> values. However, unlike repeated <title> content that may indicate duplication or ambiguity, repeated subject content may confirm the use of controlled vocabularies [15]. A quick visual inspection of these values also confirms significant use of standardized controlled vocabularies to generate subject content (although 18,569 <subject> elements do declare the use of local vocabularies). Further informal heuristic analysis found in one repository instances of <subject> elements with no subject authorities declared and no subelements that include locally generated

subject content directly in the <subject> element.

<subject> element content describes an item's "aboutness." Or as the DLF/Aquifer MODS guidelines explain, subject values "typically answer such questions as *who*, *what*, *where*, and *when*" [13]. Subject content is a key mechanism for generating a collection's browse categories and this information is often included in both abbreviated and full item records and may even appear in results lists. In both brief and full-record displays <subject> content is crucially important because users often rely on this content to judge an item's relevance in relation to search criteria. This automated testing reveals <subject> requirements to have generally been met by records in the test set. Further heuristic testing is required to judge whether or not, in records that do not include <subject> elements, is subject content truly not applicable? Formalized heuristic testing is also required to determine whether or not chosen subject subelement values truly describe the content "represented by the work." In the case of missing authority declarations, it is possible processes could be designed that supply this missing content; but due to the size, extent, and shifting nature of controlled vocabularies it may be too costly to detect and add these values automatically. It again may be a case where data providers should be trained to supply these values before exposing records for aggregation.

<location>

DLF/Aquifer Implementation Guidelines for Shareable MODS Records require the use of at least one <location> element in conjunction with a <url> subelement (Table 1.1). One and only one of the <location><url> elements is required to have a usage attribute with the value of "primary display." While <location> and <url> elements are repeatable, it is very important only one <url> element include the usage="primary display" attribute/value pair.

Final results (Appendix E, Table 36)) show that less than 1% of all records in the test set included conforming <location><url usage="primary display"> elements and content. This content is present in only three repositories (although, within these repositories, at least 99% of records include valid <location> information). In the remaining seven repositories, no records conform to this guideline. However, all is not lost. Testing for the inclusion of <location><url> elements without usage="primary display" qualification (Appendix E, Table 35) reveal over 99% of all records in the test set include at least some kind of URL location pointer. It should also be noted that these records generally include a single <location><url> element (an average of 1.09 per record). So although these locations are not qualified, it may be accurate to infer that if only a single location URL is present in a record, this content should be used as the primary display URL. It should also be noted that over 99% of all <location><url> values in the test set are unique values. This would lead one to believe each record generally points to a unique item.

And finally, an unscientific heuristic analysis on <location><url> values reveal these values are generally well formed URLs.

<location> data is an extremely crucial piece of metadata for items in harvested collections. <location><url> content is generally the primary access point for digital items and is usually one of the primary pieces of information available in results lists. With missing or inconsistent primary display data it is difficult for service providers to make available an item's intended primary access point. This significantly reduces an item's accessibility. But because records in the test set generally only include a single <location><url> element (over 99% of records do so), it would be easy to add usage="primary display" attribute/values to the <url> element in these records during normalization processes. However, in the case of records that do not include or repeat <location><url> elements, it would most likely be very difficult to generate an accurate primary display URL. That means the accessibility of about 1% of records in the test set are severely limited by this missing and incomplete metadata.

Further analysis of <location> subelements and <url> access attributes would be a useful extension of this study. This analysis would be especially useful for determining the granularity of description of access to multiple digital representations of an item and physical access points.

<accessCondition>

All records adhering to DLF/Aquifer MODS Implementation Guideline requirements must contain at least one <accessCondition> element (Table 1.1). At least one of these <accessCondition> elements must include a type attribute with the value "useAndReproduction" and the content of this element should explain any restrictions on use and reproduction of the item described by the record. This content should also state any lack of restrictions on use and reuse if the item is in the public domain as well as contact information for obtaining permission to reuse the item. Importantly, this content should "be as free from legalese and technical jargon as possible" [13].

The results of testing (Appendix E, Table 39) show that five of the ten repositories in the test set include <accessCondition> elements and <accessCondition> content for all records (one additional collection includes this content in over 66% of records). Of the remaining four repositories, none include significant numbers of <accessCondition> elements (with or without type="useAndReproduction" attribute/value pairs). Of the six repositories that have a significant number of records with <accessCondition> elements, five include type="useAndReproduction" attribute/value pairs for almost every <accessCondition> element. Only Repository 10--where an <accessCondition> element is present in every record--omits type="useAndReproduction" attributes for each <accessCondition> element. A total of 58.62% of all records include valid

<accessCondition type="useAndReporduction"> elements and content.

It is very interesting to note repeated <accessCondition> content is used frequently within repositories (Appendix E, Table 41). In four of the six repositories that include conformant <accessCondition> elements, only a single content value is used for all records. In one of the remaining two repositories, three values are used (one is only used in less than 1% of records) and one uses over 20 values. It is possible this observed use of single <accessCondition> values throughout repositories means a single <accessCondition> statement can be used to accurately describe use restrictions across collections or at least across sets within collections. If this is the case, this would help service providers normalize or augment metadata globally within repositories.

When <accessCondition> content is missing from records, the use of an item is impacted greatly. Users do not know if the the item can be repurposed and reused and who to contact for copyright clearance. This significantly reduces the usefulness of the item. However, because this content is often demonstrated to be consistent across collections, it may be fairly easy to ask data providers to provide a blanket use statement that can be automatically added to records during any metadata normalization processes. And in the case of data providers like Repository 10, it was easy to determine the single <accessCondition> content instance used for the entire collection was intended to be a use and reproduction statement. The type="useAndReproduction" attribute/value pair could confidently and easily be added to <accessCondition> elements in this collection thus achieving conformance to this guideline.

<recordInfo>

The DLF/Aquifer MODS Guidelines require the inclusion of one and only one <recordInfo> element that includes at least one <languageOfCataloging> subelement (Table 1.1). The <languageOfCataloging> element is required to include a pair of <language Term> elements. One of these <languageTerm> elements is required to include a type attribute with the value "code" along with an authority attribute with the value of "iso639-2". The value of this element should also be valid ISO 639-2b content. The second <languageTerm> element is required to include a type attribute with the value of "text" and the content of this element should be in the MARC Code List for Languages value that corresponds to the value of the <languageTerm type="code" authority="iso639-2b"> element.

Appendix E, Table 42 shows that 92.09% of all records in the test set include at least one <recordInfo> element. But of these records, only records from four repositories include required <languageOfCataloging> subelements. And of these <languageOfCataloging> elements, only

three repositories include required <languageTerm type="code" authority="iso639-2b"> subelements (two of the repositories have 100% inclusion, while one repository has less than 1% inclusion). Furthermore, only one repository includes required <languageTerm type="text"> subelements (this repository includes this subelement for all instances of <languageOfCataloging>). When <langaugeTerm> elements are used, their content is always valid and declares the English language (Appendix E, Table 43; Appendix E, Table 44). Due to missing <languageOfCataloging> and <languageTerm> subelements, less than 1% of all records in the test set conform fully to DLF/Aquifer MODS <recordInfo> requirements.

Missing <recordInfo> language content will add to the administrative overhead of service providers. Knowing the language used to create records can be used by service providers to automate record processing and facilitate future maintenance of collections. But while this data is missing from most records in the test set, it should be possible to accurately generate <recordInfo> language information during normalization processes. Because this information is usually consistent across collections or across sets, in most cases, <languageOfCataloging> values need to be determined only once per collection or set. It may be possible to use information from other parts of the record or have the data provider provide this information once. And in the case of records that are only missing <languageTerm type="text"> content, this information is easily generated using <languageTerm type="code" authority="iso639-2b"> element content present in the record. So while there is currently very little conformance to this requirement, it may be fairly easy automatically add this information during ingest.

Future analysis of this requirement may include looking at <languageTerm> values when no authority is declared to see if these values are consistent with certain controlled vocabularies that could be identified automatically.

Safe/Algorithmic Normalizations

In analyzing record conformance to the *DLF/Aquifer Implementation Guidelines for Sharable MODS Records*, it was noted that while many instances of non-conformity such as missing <subject> elements require altering data provider practices for remediation, many other instances of non-conformity could be remediated through fairly simple, "safe" and automatic metadata transformations and augmentations. These transformations would be safe in that they could bring records closer to conformance with "no risk of degradation" to the original metadata [16]. In the test set, possible safe automatic transformations (Table 3.2) are enabled by three scenarios: single values in a record requiring no differentiation (such as single <originInfo> dates or single <location> URLs), global statements that are true for all or a majority of records in a collection (such as <accessCondition> or <recordInfo> information), or missing natural language

content that can easily be generated by existing authority content (as in the case of <language> elements).

Upon completion of these safe transformations and augmentations (that require little administrative or computing overhead from data or service providers) the overall conformance of test set records to the DLF/MODS guidelines is much, much higher. Column E in the Results Summary (Table 3.1) shows that if these safe transformations are applied, there would be over 90% compliance for five required elements. There would also be between 38% and 66% compliance for three of the remaining four elements and only one element--<physicalDescription>--would still show very low compliance at 7.97%. These numbers are significantly better than before safe transformations are applied and may demonstrate records in the test set are not deficient in significant ways.

Table 3.2 Summary of Safe Transformations

Required Element	Transformations/Augmentations & Results
<originInfo>	● For records that only include a single date-related <originInfo> subelement, service providers can safely supply missing keydate="yes" attribute/value pairs. ● There are 118,825 records with single <originInfo> date-related subelements. Applying keyDate="yes" attribute/value pairs to these records improves conformance from 3% to just under 39%.
<language>	● Service providers can accurately and automatically generate <languageTerm type="text"> content from supplied <languageTerm type="code" authority="iso639-2b"> content. ● Over 53% of all records include at least <languageTerm type="code" authority="iso639-2b"> subelements for <language> elements. Therefore, 53% of all records could easily conform to this requirement as opposed to the much lower 0.12% conformance rate before these transformations are applied.
<location>	● For records that only include a single <location><url> element, service providers can safely supply missing usage="primary display" attribute/value pairs. ● There are 281,748 records with single <location><url> elements. Applying usage="primary display" attribute/value pairs to these records improves conformance from 0.37% to over 92%.
<accessCondition>	● Service providers could apply a global <accessCondition type="useAndReproduction"> value provided by data providers.

	This could be general information such as a link to a web page with contact information that is applied when a record is missing required <accessCondition> information. This would bring conformance to 100%.
<recordInfo>	• Using single values supplied by data providers, service providers could automatically apply global <languageOfCataloging> information within collections. This would bring conformance to 100%.

The ease with which record conformance problems can be fixed through safe automatic processes may point to some inefficiencies in the *DLF/Aquifer Implementation Guidelines for Shareable MODS Records* or OAI-PMH. In many instances (as in the case of <languageOfCataloging> and <accessCondition> elements), automatic fixes are possible because metadata values were demonstrated to be consistent across sets or collections. The nature of these fixes could point to a useful alteration of OAI-PMH that would allow data providers to supply special records that apply to whole sets or whole collections. These records would include global values for a set or collection that could be overridden by values in individual item records within the set or collection. For instance, a data provider could supply a global <accessCondition type="useAndReproduction"> value for a collection that is overridden, if necessary, by a value within an item record in the collection. <languageOfCataloging> values could also be provided one time in these globally applicable records. Requiring data providers to provide globally consistent values only once would not only reduce the administrative overhead of declaring redundant values over and over, it may also cause them to more carefully consider these values and improve metadata quality by spotlighting these one-time declarations.

Many of the other automatic fixes focus on inserting content or attribute values that can be inferred accurately from elsewhere in metadata records. For instance, <language> element non-conformity was largely due to missing <languageTerm type="text"> elements and content while valid <languageTerm type="code" authority="iso639-2b"> elements and content were generally present (when <language> elements were present) . So it wasn't really the case data providers were providing inaccurate, inconsistent content or providing incomplete content, they were simply not fully meeting somewhat redundant administrative requirements. Metadata guideline creators and OAI-PMH could work together to leverage service provider automatic processing capabilities to reduce some of this administrative overhead (like requiring data providers to provide language content in both text and code format). Once a code value is declared in a record, service providers can transform and reformat this value easily and safely. This is also similar to the case of missing usage="primary display" attribute/value pairs for

<location> elements. In many instances, only a single <location> element was present in a record meaning service providers could safely infer and denote this single location is intended for primary display. Therefore, data providers may really only need to include usage="primary display" attributes when more than one <location> element appears in a record. This is another instance when guidelines creators can work with OAI-PMH to leverage service provider processing capabilities and reduce administrative overhead for data providers.

By realigning OAI-PMH structures and future data provider requirements in a manner that takes advantage of service provider processing capabilities, the DLF/Aquifer Project and OAI-PMH could reduce the onus on data providers to consistently provide redundant data. This could free up data provider resources to focus on creating shareable, high-quality content for elements such as <titleInfo>, <subject>, and <originInfo> that rely on the data provider's intimate knowledge of an item for quality and shareability. Service providers are much better equipped to repeatedly and consistently produce information that can be inferred from elsewhere in the record or by one-time global statements. By logically distributing metadata generation responsibilities, OAI-PMH and DLF/Aquifer could provide a more efficient and effective means for creation of shareable metadata.

Post-DLF/Aquifer MODS Implementation Guidelines 1.0 Results

One way to gauge the usefulness and effect of the DLF/Aquifer Implementation Guidelines for Shareable MODS Records is to compare the conformance rates of records with OAI datestamps from before the release of the 1.0 version of the guidelines document to records datestamped after the release of the document. Any trends in the conformance of records in the pre- and post-release subsets could point to the usefulness of the document in guiding and encouraging the creation of shareable MODS-based metadata.

Only two repositories contain records datestamped both before and after the release of DLF/Aquifer 1.0 (I used the cut-off date of October 31, 2006 to reflect the November, 2006 version 1.0 release date). Repository 8 includes 23,301 pre-release records and 866 records post-release while Repository 9 includes 216,980 pre-release records and 37,090 post-release records. The remaining repositories contain records the were either all datestamped before or after the release of the 1.0 version of the document. Repositories 2 and 7 include records all datestamped before DLF/Aquifer MODS 1.0 and Repositories 1, 3, 4, 5, 6, and 10 include records that were all datestamped after DLF/Aquifer MODS 1.0. Therefore, conformance comparisons could only be made within two collections while the rest were made between collections.

When making comparisons within collections, there was very little indication of any major

effect made by the release of the MODS 1.0 guidelines (test results broken down by pre- and post-release of DLF/Aquifer MODS 1.0 for Repository 8 and 9 can be found in Appendix E). <accessCondition> conformance improved from 58.74% to 94.75% and <language type="code" authority="iso639-2b"> usage improved from 51.69% to 78.48% in Repository 9. But <subject> conformance drops from 69.97% to 48.92% in the same repository. For all other elements in both repositories, conformance generally stays at the same rate both before and after the cut-off date making it difficult to draw any conclusions regarding the effect of the guidelines document.

Comparisons can also be made of conformance rates between collections that include only records datestamped before or after the release date. However, these comparisons also reveal little in the way of trends. Repository 7--which includes only records datestamped before the release of the guidelines document--has nearly 100% conformance for five elements; this is one of the top 3 or 4 conformance rates in the entire test set. And while Repository 2 only has 100% conformance for three elements (making it one of the least conforming collections) there are two other repositories that include only post-guideline records that have as low or lower overall conformance rates. So while deeper analysis may unearth a measurable effect (and over time, this effect should become more apparent), the records in this test set don't point to any profound effect caused by the release of the DLF/Aquifer Implementation Guidelines for Shareable MODS Records 1.0.

Chapter 4 - Conclusions

The results of this study initially showed significant non-conformance to the *DLF/Aquifer Implementation Guidelines for Shareable MODS Records* in currently exposed MODS records. However, it may be possible to rectify many of these areas of non-conformance through relatively simple changes to data provider and/or service provider practices and processes. <originInfo>, <language>, <location>, <accessCondition>, and <recordInfo> requirement non-conformance is generally due to missing information that can inferred automatically from elsewhere in a record. In these cases, service providers could apply metadata normalization and augmentation processes that would safely remediate non-conformance to these requirements. On the other hand, remediation for certain element requirements (such as for <subject> and <physicalDescription>) that require subjective intimate familiarity with an item may be best achieved by altering data provider practices. And although many overall process are required to significantly improve conformance to DLF/Aquifer MODS requirements, no process is overly complex or unexpected. And after only applying safe, automatic normalizations and augmentations, this study shows record conformance would in fact improve significantly.

That overall conformance is significantly improved by relatively simple, safe and automatic processing points to possible inefficiencies in the DLF/Aquifer MODS guidelines and OAI-PMH. It is possible that both OAI-PMH and metadata creation guidelines could be realigned to better make use of service provider processing capabilities. This would allow data providers to focus less on administrative redundancies and more on creating metadata facets that require subjective and intimate knowledge of items and collections. Possible solutions include allowing data providers to provide one-time, global metadata values for entire sets or collections and streamlining guidelines to eliminate repetitive requirements (such as providing both code and text versions of a single value or declaring differentiations when only a single instance of an element exists). Service providers are well-equipped to generate repetitive, consistent content. Leveraging this fact and aligning data provider practices with service provider processes would make creation of shareable metadata easier and more likely.

This study's proposed remediations for requirement non-conformance make the argument that a combination of data provider practices and service provider processing will ultimately be the most thorough and effective way to generate shareable metadata. While service providers can efficiently process, repurpose and normalize many aspects of metadata; service providers will always have difficulty applying subjective metadata that can't be implied from an explicit characteristic of an item or an item record. So while service providers can apply a global value repeatedly or generate a text value from a code value, only data providers (or possibly end-users) can most efficiently provide subjective metadata such as <subject> values or content that

requires other intimate knowledge of an item such as <digitalOrigin> content. Data providers and service providers are each uniquely qualified to supply different aspects of shareable metadata and tighter and more efficient integration between the two will result in a system that better encourages and facilitates the creation and exposure of shareable metadata.

Although this study provides a good starting point for gauging how close current metadata creation practice is aligned with requirements and some recommendations in the DLF/Aquifer MODS guidelines, it is important to realize the test set is comprised of records from a limited number of data provider institutions. These institutions are generally leaders in digital library technologies and may generally exhibit higher quality metadata creation practices. Real world applications will most likely include much more heterogeneous metadata from a much wider variety of data providers. For instance, the Illinois OAI-PMH Cultural Heritage project estimated an aggregation of metadata from approximately 580 institutions [4]. In these broader applications, it can be expected records will be nonconformant to DLF/Aquifer MODS guidelines in more varied and severe ways. However, the issues discovered in this study should also apply to a wider sample set and still be useful to future applications of these guidelines.

The results of this study can be furthered by extending analysis beyond the base requirements of DLF/Aquifer MODS guidelines. While some element recommendations were analyzed during this study, delving deeper into current data provider conformance to more recommendations would give deeper insight into the current shareability of MODS records Measuring conformance to one particular recommendation--use of the <name> element--would be particularly insightful and relate closely to work being done to develop DLF/Aquifer MODS guidelines levels of adoption [22]. While conforming to all MODS guidelines requirements satisfy almost all requirements of the first two proposed levels of adoption, inclusion (if applicable) of <name> elements is required for the first level of adoption. So analysis of use of <name> elements would help bring the MODS guidelines in sync with the levels of adoption. This, in addition to deeper analysis of the current quality of required elements would give an even more comprehensive overview of current MODS creation practice and give further insight into how data providers and service providers can work to together to facilitate creation of shareable MODS records.

References

[1] Open Archives Initiative, "Open Archives Initiative Protocol for Metadata Harvesting," <http://www.openarchives.org/pmh/>.

[2] C. Lagoze and H. Van de Sompel, "The Making of the Open Archives Initiative Protocol for Metadata Harvesting.," Library Hi Tech, vol. 21, pp. 118-28, 2003.

[3] "Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0," <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

[4] "Dublin Core Metadata Element Set, Version 1.1," <http://dublincore.org/documents/dces/>.

[5] Lagoze and H. Van de Sompel, "The Open Archives Initiative: Building a Low Barrier Interoperability Framework," JCDL 2001: Proceedings of the First ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 54-62, 2001.

[6] R. Tennant, "Bitter Harvest: Problems & Suggested Solutions for OAI-PMH Data & Service Providers," <http://www.cdlib.org/inside/projects/harvesting/bitter_harvest.html>.

[7] Digital Library Federation, "Best Practices for Shareable Metadata," <http://webservices.itcs.umich.edu/mediawiki/oaibp/index.php/ShareableMetadataPublic>.

[8] J. Ward, "A Quantitative Analysis of Unqualified Dublin Core Metadata Element Set Usage within Data Providers Registered with the Open Archives Initiative," JCDL 2003: Proceedings of the 2003 Joint Conference on Digital Libraries, 2003.

[9] D.I. Hillman, "Metadata Quality: From Evaluation to Augmentation," <http://ecommons.library.cornell.edu/bitstream/1813/7899/1/Metadata_Quality_rev.pdf>.

[10] T.R. Bruce and D.I. Hillman, "The Continuum of Metadata Quality: Defining, Expressing, Exploiting," in Metadata in Practice, ALA Editions, 2004.

[11] "DLF Aquifer," <http://www.diglib.org/aquifer/>.

[12] "American Social History," <http://www.dlfaquifer.org/home/about>.

[13] Digital Library Federation Aquifer Metadata Working Group "Digital Library Federation / Aquifer Implementation Guidelines for Shareable MODS Records," 2006. <http://www.diglib.org/aquifer/dlfmodsimplementationguidelines_finalnov2006.pdf>.

[14] "MODS User Guidelines," <http://www.loc.gov/standards/mods/v3/mods-userguide.html>.

[15] M. Foulonneau and T.W. Cole, "Strategies For Reproceessing Aggregated Metadata," 2005. <http://cicharvest.grainger.uiuc.edu/documents/metadatareprocessing.pdf>.

[16] D. Hillmann, N. Dushay, J. Phipps, "Improving Metadata Quality: Augmentation and Recombination," 2004. <http://purl.org/metadataresearch/dcconf2004/papers/Paper_21.pdf>.

[17] R. Tennant, "Specifications for Metadata Harvesting Tools," <http://www.cdlib.org/inside/projects/harvesting/metadata_tools.htm>.

[18] Digital Library Federation, "DLF MODS Portal About Data Contributors," <http://quod.lib.umich.edu/m/mods/viewcolls.html>.

[19] "Reap: OAI Command Line Harvesting Tool," <http://sourceforge.net/docman/display_doc.php?docid=27545&group_id=47963>.

[20] Creative research Systems, "Sample Size Calculator," <http://www.surveysystem.com/sscalc.htm>.

[21] "IANA | MIME Media Types," <http://www.iana.org/assignments/media-types/>.

[22] Digital Library Federation Aquifer Project, "MODS Guidelines Levels of Adoption," <http://wiki.dlib.indiana.edu/confluence/display/DLFAquifer/MODS+Guidelines+Levels+of+Adoption>.

[23] S. Shreeves, J. Kaczmarek, and T. Cole, "Harvesting Cultural heritage metadata using the OAI protocol," Library Hi Tech, vol. 21, pp. 159-169, 2003.

Appendix A - Repository Data

Repository Names

A.1 Repository Names	
Repository ID	**Repository Name**
1	A Celebration of Women Writers
2	OCLC Research Publications
3	University of Tennessee
4	Southern Spaces
5	Digital Books from UIUC and the Open Content Alliance
6	University of Chicago Metadata Repository
7	Indiana University Library Cushman Collection
8	Deep Blue at the University of Michigan
9	Library of Congress Memory Collection
10	University of Michigan. University Library. Digital Library Production Service.

Repository URLs & Sets

A.2 Repository URLS & Sets		
Repository ID	**Base URL**	**Sets**
1	http://digital.library.upenn.edu/webbin/OAI-celebration	*
2	http://errol.oclc.org/orpubs.oclc.org.srw2oai	*
3	http://diglib.lib.utk.edu/cgi/b/broker20/broker20	*
4	http://oai.library.emory.edu/sspaces	*
5	http://ratri.grainger.uiuc.edu/oca-oaiprovider/oai.asp	*
6	http://oai.lib.uchicago.edu:8180/	*
7	http://oai.dlib.indiana.edu/phpoai/oai2.php	cushman
8	http://deepblue.lib.umich.edu/dspace-oai/request	*
9	http://memory.loc.gov/cgi-bin/oai2_0	*
10	http://quod.lib.umich.edu/cgi/b/broker20/broker20/	(note 2)

Repository Harvest & Record Data

A.3 Repository Harvest & Record Data							
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
1	09/23/07	293493	293796	304	304	304	0
2	09/04/07	294656	295507	852	852	0	852
3	09/04/07	293797	294655	859	859	859	0
4	10/17/07	1	62	62	62	62	0
5	08/30/07	10147	10912	766	767	766	0
6	09/04/07	295508	295879	372	372	372	0
7	09/23/07	254901	269325	14425	14425	0	14425
8	09/23/07	269326	293492	24167	24299	866	23301
9	09/02/07	63	254900	254072	292000	37090	216980
10	10/27/07	295880	305468	9589	9589	9589	0
			Totals:	**305468**	**343529**	**49908**	**255558**

A – Repository ID

B – Harvest Date

C – Start Record ID

D – End Record ID

E – Total Records Extracted

F – Total Records Harvested

G – Records with OAI Datestamp After DLF/Aquifer MODS Implementation Guidelines 1.0

H – Records with OAI Datestamp Before DLF/Aquifer MODS Implementation Guidelines 1.0

Notes:

1 - A technical anomaly inserted records from the Digital Books from UIUC and the Open Content Alliance into the ID range of records from the Library of Congress Memory Collection.

2 - Set limiters for Repository 10 include:

lincolnmods;oaimods:micountymods;oaimods:moamods;

oaimods:ppotpusmods;oaimods:railroadmods;oaimods:umhistmathmods

Appendix B - Table Structures

Repositories Table (dbo_Repositories)

B.1 Repositories Table (dbo_Repositories)							
Field Name	**Data Type**	**Field Size**	**Required**	**Indexed**	**New Values**	**Decimal Places**	**Allow Zero Length**
repoid	AutoNumber	Long Integer		Yes (No duplicates)	Increment		
baseURL	Text	255	Yes	No			Yes
setSpec	Text	50	No	No			Yes
whenHarvested	Date/Time		No	No			
repositoryName	Text	255	No	No			Yes

Records Table (dbo_Records)

B.2 Records Table (dbo_Records)							
Field Name	**Data Type**	**Field Size**	**Required**	**Indexed**	**New Values**	**Decimal Places**	**Allow Zero Length**
repoid	Number	Long Integer	Yes	No		Auto	
recordID	AutoNumber	Long Integer		Yes (No duplicates)	Increment		
OAIIdentifier	Text	255	No	No			Yes
OAIDateStamp	Date/Time		No	No			
XMLMetadata	Memo		No	No			Yes

Metadata Table (dbo_Metadata)

B.3 Metadata Table (dbo_Metadata)							
Field Name	**Data Type**	**Field Size**	**Required**	**Indexed**	**New Values**	**Decimal Places**	**Allow Zero Length**
recordID	Number	Long Integer	Yes	No		Auto	
metaRowID	AutoNumber	Long Integer		Yes (No Duplicates)	Increment		
propertyName	Text	50	No	No			Yes
propertyNS	Text	255	No	No			Yes
parent_propName	Text	50	No	No			Yes
parent_propNS	Text	255	No	No			Yes
parent_metaRowID	Number	Long Integer	No	No		Auto	
propText	Text	255	No	No			Yes
propTextOverflow	Memo		No	No			Yes
a_type	Text	100	No	No			Yes
a_authority	Text	100	No	No			Yes
a_encoding	Text	100	No	No			Yes
a_href	Text	100	No	No			Yes
a_displayLabel	Text	100	No	No			Yes
a_keyDate	Text	50	No	No			Yes
a_usage	Text	100	No	No			Yes

Appendix C - Database Queries

<titleInfo><title>

Query 1

Returns instances of <title> elements where <title> content is not empty. Results are written into a temporary table.

```
SELECT dbo_Records.recordID, dbo_Records.repoID, propText,
propTextOverflow, parent_propName INTO title_exists_temp_prelim
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE (((dbo_Metadata.propertyName="title") And
((dbo_Metadata.propText) Is Not Null)));
```

Query 2

Returns instances of <title> from the results of Query 1 that are children of <titleInfo> elements.

```
SELECT * INTO title_exists_temp
FROM title_exists_temp_prelim
WHERE (parent_propName='titleInfo');
```

Query 3

Count number of occurrences of <title> elements for each record in temporary table from Query 2.

```
SELECT recordID, COUNT(recordID) AS numOccurences
FROM title_exists_temp
GROUP BY recordID;
```

Query 4

Group and count result categories from Query 3.

```
SELECT numOccurences, COUNT(numOccurences) AS totalTitleCount
FROM titleInfo2_title_exists_count2
GROUP BY numOccurences;
```

Query 5

Return all instances of child elements of <titleInfo> that are not <title> elements. Write results into a temporary table.

```
SELECT dbo_Records.recordID, dbo_Records.repoID, propertyName,
propText, propTextOverflow INTO title_Subelements_temp
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE (((dbo_Metadata.propertyName<>"title") AND
(dbo_Metadata.parent_propName="titleInfo") AND ((dbo_Metadata.propText)
IS NOT null)));
```

Query 6

Group and count subelement occurrences from resulting temporary table from Query 5.

```
SELECT propertyName, COUNT(propertyName) AS numOccurences
FROM title_Subelements_temp
GROUP BY propertyName;
```

Query 7

Count number of occurrences of non-<title> <titleInfo> children for each record by querying table from Query 5.

```
SELECT recordID, Count(recordID) AS numOccurences
FROM title_subelements_temp
GROUP BY recordID;
```

Query 8

Group and count results of Query 7.

```
SELECT numOccurences, COUNT(numOccurences) AS totalOccurencesCount
FROM titleInfo55_count_group
GROUP BY numOccurences;
```

Query 9

Return occurrences of <nonSort> elements in result of Query 5.

```
SELECT * INTO title_nonSort_temp
FROM title_Subelements_temp
WHERE (propertyName="nonsort");
```

Query 10

Group and count <nonSort> content returned by Query 9.

```
SELECT propText, Count(propText) AS numOccurences
FROM title_nonsort_temp
GROUP BY propText;
```

<typeOfResource>

Query 1

Return all instances of <typeOfresource> elements that are children of <mods> elements. Results are written to a temporary table.

```
SELECT dbo_Records.recordID, dbo_records.repoID, propText,
propTextOverflow INTO typeOfResource_exists_temp
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE ((dbo_Metadata.propertyName="typeOfResource") And
(dbo_Metadata.parent_propName="mods"));
```

Query 2

Return all instances of <typeOfresource> elements that are children of <mods> elements that have valid content. Results are written to a temporary table.

```
SELECT dbo_Records.recordID, dbo_Records.repoID, propText,
propTextOverflow INTO typeOfResource_exists_valid_temp
FROM typeOfResource_exists_temp INNER JOIN dbo_Records ON
dbo_Records.recordID=typeOfResource_exists_temp.recordID
WHERE ((propText) In ('text','cartographic','notated music','sound
recording','sound recording-musical','sound recording-
nonmusical','still image','moving image','three dimensional
object','software, multimedia','mixed material'));
```

Query 3

Group and count <typeOfResource> content values from query 2.

```
SELECT propText, Count(propText) AS numOccurrences
FROM typeOfResource_exists_temp
GROUP BY propText;
```

Query 4

Return distinct records from query 2.

```
SELECT DISTINCT typeOfResource_exists_valid_temp.recordID
FROM typeOfResource_exists_valid_temp;
```

<originInfo>

Query 1

Return instances of recommended date-related elements and do not have empty content. Results are written into a temporary table.

```
SELECT dbo_Records.recordID, dbo_Records.repoID, propertyName,
propText, propTextOverflow, parent_propName, a_encoding INTO
originInfo_validSubElement_notNull_temp_prelim
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE (((dbo_Metadata.propertyName="dateIssued") OR
(dbo_Metadata.propertyName="dateCreated") OR
(dbo_Metadata.propertyName="copyrightDate") OR
(dbo_Metadata.propertyName="dateOther")) And
(dbo_Metadata.parent_propName="originInfo") AND (dbo_Metadata.propText
IS NOT null));
```

Query 2

Select results of Query 1 that are children of <originInfo> elements.

```
SELECT * INTO originInfo_validSubElement_notNull_temp
FROM originInfo_validSubElement_notNull_temp_prelim
```

```
WHERE (parent_propName='originInfo');
```

Query 3

Get distinct records from Query 2.

```
SELECT DISTINCT recordID
FROM originInfo_validSubElement_notNull_temp;
```

Query 4

Return all rows in result of Query 2 where the "encoding" attribute is present.

```
SELECT *
FROM originInfo_validSubElement_notNull_temp
WHERE a_encoding IS not null;
```

Query 5

Group and count "encoding" attribute values returned in Query 2.

```
SELECT a_encoding, count(a_encoding) AS occurrences
FROM originInfo_validSubelement_notNull_temp
GROUP BY a_encoding;
```

Query 6

Get distinct element values returned by Query 2.

```
SELECT DISTINCT propText
FROM originInfo_validSubElement_notNull_temp;
```

Query 7

Group and count values returned by Query 2.

```
SELECT propText, count(propText) AS occurrences
FROM originInfo_validSubElement_notNull_temp
GROUP BY propText;
```

Query 8

Count occurrences of date subelements in each record returned by Query 2.

```
SELECT recordID, COUNT(recordID) AS totalDateSubelements
FROM originInfo_validSubElement_notNull_temp
GROUP BY recordID;
```

Query 9

Group and count element values returned by Query 2.

```
SELECT originInfo_validSubelement_notNull_temp.propertyName,
Count(originInfo_validSubelement_notNull_temp.propertyName) AS
numOccurences
FROM originInfo_validSubelement_notNull_temp
GROUP BY originInfo_validSubelement_notNull_temp.propertyName;
```

Query 10

Count and group number of occurrences of date-related subelements returned by Query 8.

```
SELECT originInfo3_noKeyDate_totals.totalDateSubelements,
Count(originInfo3_noKeyDate_totals.totalDateSubelements) AS total
FROM originInfo3_noKeyDate_totals
GROUP BY originInfo3_noKeyDate_totals.totalDateSubelements;
```

Query 11

Return instances of valid date-related elements that are children of <originInfo> elements and do not have empty content and include keyDate="yes" attribute/value pairs. Results are written into a temporary table.

```
SELECT dbo_Records.recordID, dbo_Records.repoID, propertyName,
propText, parent_propName, propTextOverflow INTO
originInfo_validSubElement_validAttribute_notNull_temp_prelim
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE (((dbo_Metadata.propertyName="dateIssued") OR
(dbo_Metadata.propertyName="dateCreated") OR
```

```
(dbo_Metadata.propertyName="copyrightDate") OR
(dbo_Metadata.propertyName="dateOther")) And
(dbo_Metadata.parent_propName="originInfo") And
(dbo_Metadata.a_keyDate='yes') AND (dbo_Metadata.propText IS NOT null));
```

Query 12

Select results of Query 11 that are children of <originInfo> elements.

```
SELECT * INTO originInfo_validSubElement_validAttribute_notNull_temp
FROM originInfo_validSubElement_validAttribute_notNull_temp_prelim
WHERE (parent_propName='originInfo');
```

Query 13

Return distinct records from Query 11.

```
SELECT DISTINCT
originInfo_validSubElement_validAttribute_notNull_temp.recordID
FROM originInfo_validSubElement_validAttribute_notNull_temp;
```

Query 14

From results of Query 11 group and count number of "keyDate" elements in each record.

```
SELECT recordID, COUNT(recordID) AS totalKeyDates
FROM originInfo_validSubElement_validAttribute_notNull_temp
GROUP BY recordID;
```

Query 15

From results of Query 11, group and count "keyDate" element values.

```
SELECT
originInfo_validSubElement_validAttribute_notNull_temp.propertyName,
Count(originInfo_validSubElement_validAttribute_notNull_temp.propertyNa
me) AS numOccurences
FROM originInfo_validSubElement_validAttribute_notNull_temp
GROUP BY
originInfo_validSubElement_validAttribute_notNull_temp.propertyName;
```

Query 16

From result of Query 14, group and count total occurrences of "keyDate" elements in records.

```
SELECT originInfo5_keyDate_totals.totalKeyDates,
Count(originInfo5_keyDate_totals.totalKeyDates) AS total
FROM originInfo5_keyDate_totals
GROUP BY originInfo5_keyDate_totals.totalKeyDates;
```

Query 17

Return distinct records from result of Query 11.

```
SELECT DISTINCT recordID
FROM originInfo_validSubElement_validAttribute_notNull_temp;
```

<language>

Query 1

Return instances of <language> elements that are children of <mods> elements. Write results into a temporary table.

```
SELECT dbo_Records.recordID, dbo_Records.repoID,
CLng([dbo_Metadata.metaRowID]) AS metaRowID INTO language_temp
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE ((dbo_Metadata.propertyName="language") And
(dbo_Metadata.parent_propName="mods"));
```

Query 2

Return distinct records from results of query 1.

```
SELECT DISTINCT language_temp.recordID
FROM language_temp;
```

Query 3

Return all instances of <languageTerm> elements that are children of <language> elements.

write all results to a temporary table.

```
SELECT dbo_Records.recordID, dbo_Records.repoID,
CLng([dbo_Metadata.metaRowID]) AS metaRowID, dbo_Metadata.propText,
dbo_Metadata.a_type, dbo_Metadata.parent_MetaRowID,
dbo_Metadata.a_authority INTO language_languageTerm_temp
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE (((propertyName="languageTerm") And (parent_propName="language"))
And (propText Is Not Null));
```

Query 4

Return distinct records from query 3.

```
SELECT DISTINCT language_languageTerm_temp.recordID
FROM language_languageTerm_temp;
```

Query 5

Return all <languageTerm> elements that include both type="code" and authority="iso639-2b" attribute/value pairs. Write results to temporary table.

```
SELECT * INTO language_languageTerm_code_temp
FROM language_languageTerm_temp
WHERE a_type='code' AND a_authority='iso639-2b';
```

Query 6

Return distinct records from query 5.

```
SELECT DISTINCT language_languageTerm_code_temp.recordID
FROM language_languageTerm_code_temp;
```

Query 7

Return rows from query 3 where <languageTerm> element includes type="text" attribute/value pairs. Write results to temporary table.

```
SELECT * INTO language_languageTerm_text_temp
FROM language_languageTerm_temp
WHERE a_type='text';
```

Query 8

Return distinct records from query 7.

```
SELECT DISTINCT language_languageTerm_text_temp.recordID
FROM language_languageTerm_text_temp;
```

Query 9

Group and count element content values from result of query 5.

```
SELECT propText, count(propText) AS occurrrences
FROM language_languageTerm_code_temp
GROUP BY propText;
```

Query 10

Group and count element content values from result of query 7.

```
SELECT propText, count(propText) AS occurrrences
FROM language_languageTerm_text_temp
GROUP BY propText;
```

Query 11

Return rows from query 3 where type attribute is missing or has no value pair.

```
SELECT *
FROM language_languageTerm_temp
WHERE (((language_languageTerm_temp.a_type) Is Null));
```

Query 12

Count number of occurrences of <title> elements for each record in temporary database from query 1.

```
SELECT recordID, COUNT(recordID) AS numOccurences
FROM title_exists_temp
GROUP BY recordID;
```

Query 13

Count number of occurrences of <title> elements for each record in temporary database from query 1.

```
SELECT recordID, COUNT(recordID) AS numOccurences
FROM title_exists_temp
GROUP BY recordID;
```

Query 14

Count number of occurrences of <title> elements for each record in temporary database from query 1.

```
SELECT recordID, COUNT(recordID) AS numOccurences
FROM title_exists_temp
GROUP BY recordID;
```

Query 15

Count number of occurrences of <title> elements for each record in temporary database from query 1.

```
SELECT recordID, COUNT(recordID) AS numOccurences
FROM title_exists_temp
GROUP BY recordID;
```

<physicalDescription>

Query 1

Return instances of <physicalDescription> elements that are children of <mods> elements. Write results into a temporary table.

```
SELECT dbo_Records.recordID, dbo_records.RepoID,
CLng([dbo_Metadata.metaRowID]) AS metaRowID INTO
```

```
physicalDescription_exists_temp
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE ((dbo_Metadata.propertyName="physicalDescription") AND
(dbo_Metadata.parent_propName="mods"));
```

Query 2

Count occurrences of each record ID in results of Query 1.

```
SELECT recordID, Count(recordID) AS numOccurences
FROM physicalDescription_exists_temp
GROUP BY recordID;
```

Query 3

Group and total occurrence values from results of Query 2. This will show totals of <phyicalDescription> occurrences.

```
SELECT numOccurences, COUNT(numOccurences) AS totalOccurencesCount
FROM physicalDescription2_exists_count
GROUP BY numOccurences;
```

Query 4

Get all instances of <digitalOrigin> that are children of <physicalDescription>.

```
SELECT dbo_Records.recordID, dbo_Records.repoID, propertyName,
propText, propTextOverflow, parent_metaRowID INTO
physicalDescription_digitalOrigin_notNull_temp
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE ((dbo_Metadata.propertyName="digitalOrigin") And
(dbo_Metadata.parent_propName="physicalDescription") AND
(dbo_Metadata.propText IS NOT null));
```

Query 5

Count occurrences of each record ID in results of Query 4.

```
SELECT recordID, Count(recordID) AS numOccurences
FROM physicalDescription_digitalOrigin_notNull_temp
```

```
GROUP BY recordID;
```

Query 6

Group and total <digitalOrigin> values from results of Query 4.

```
SELECT physicalDescription_digitalOrigin_notNull_temp.propText,
Count(physicalDescription_digitalOrigin_notNull_temp.propText) AS
numOccurences
FROM physicalDescription_digitalOrigin_notNull_temp
GROUP BY physicalDescription_digitalOrigin_notNull_temp.propText;
```

Query 7

Get all instances of <internetMediaType> that are children of <physicalDescription>.

```
SELECT dbo_Records.recordID, dbo_Records.repoID, propertyName,
propText, propTextOverflow, parent_metaRowID INTO
physicalDescription_internetMediaType_notNull_temp
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE ((dbo_Metadata.propertyName="internetMediaType") And
(dbo_Metadata.parent_propName="physicalDescription") AND
(dbo_Metadata.propText IS NOT null));
```

Query 8

Group and total record occurrences from results of query 7. Used to get unique records from query 7.

```
SELECT recordID, Count(recordID) AS numOccurences
FROM physicalDescription_internetMediaType_notNull_temp
GROUP BY recordID;
```

Query 9

Group and total <internetMediaType> values from results of query 7.

```
SELECT propText, Count(propText) AS numOccurences
FROM physicalDescription_internetMediaType_notNull_temp
GROUP BY propText;
```

Query 10

Select all records that satisfy the requirements of both queries 4 and 7. These results satisfy all element requirements (as all results of query 4 included valid propText values).

```
SELECT DISTINCT physicalDescription_exists_temp.recordID
FROM physicalDescription_exists_temp,
physicalDescription_digitalOrigin_notNull_temp,
physicalDescription_internetMediaType_notNull_temp
WHERE
(((physicalDescription_exists_temp.metaRowID)=physicalDescription_digit
alOrigin_notNull_temp.parent_MetaRowID And
(physicalDescription_exists_temp.metaRowID)=physicalDescription_interne
tMediaType_notNull_temp.parent_MetaRowID));
```

<subject>

Query 1

Return instances of <subject> elements that are children of <mods> elements. Write results into a temporary table.

```
SELECT dbo_Metadata.recordID, dbo_Metadata.metaRowID,
dbo_Metadata.propertyName, dbo_Metadata.parent_propName,
dbo_Metadata.parent_metaRowID, dbo_Metadata.propText,
dbo_Metadata.propTextOverflow, dbo_Metadata.a_authority INTO
subject_exists_temp
FROM dbo_Metadata INNER JOIN dbo_Records ON
dbo_Metadata.recordID=dbo_Records.recordID
WHERE (((dbo_Metadata.propertyName)='subject') AND
(dbo_Metadata.parent_propName="mods"));
```

Query 2

Count number of occurences of <subject> elements for each record in results of Query 1. The number of rows in the result will be the number of distinct records containing <subject> elements.

```
SELECT recordID, Count(recordID) AS numOccurences
FROM subject_exists_temp
```

```
GROUP BY recordID;
```

Query 3

Group and total occurrence values from the results of Query 2.

```
SELECT numOccurences, COUNT(numOccurences) AS totalOccurencesCount
FROM subject2_exists_count
GROUP BY numOccurences;
```

Query 4

Check for that all <subject> instances in results of Query 1 have null propText values. Any non-null values indicate invalid use of <subject> elements.

```
SELECT RecordID, propText, propTextOverflow
FROM subject_exists_temp
WHERE propText IS NOT null;
```

Query 5

Returns all records from result in Query 1 where the authority attribute is not null.

```
SELECT subject_exists_temp.recordID, metaRowID, propertyName,
parent_propName, parent_metaRowID, propText, propTextOverflow,
a_authority INTO subject_authority_exists_temp
FROM subject_exists_temp INNER JOIN dbo_Records_FullDB ON
subject_exists_temp.recordID=dbo_Records_FullDB.recordID
WHERE (((subject_exists_temp.a_authority) Is Not Null) AND
(dbo_Records_FullDB.repoID=1));
```

Query 6

Group and total occurrence values from the results of Query 6.

```
SELECT recordID, Count(recordID) AS numOccurences
FROM subject_authority_exists_temp
GROUP BY recordID;
```

Query 7

Group and total authority values from the results of Query 7.

```
SELECT a_authority, Count(a_authority) AS numOccurences
FROM subject_authority_exists_temp
GROUP BY a_authority;
```

Query 8

Group and total authority values from the results of Query 7.

```
SELECT a_authority, Count(a_authority) AS numOccurences
FROM subject_authority_exists_temp
GROUP BY a_authority;
```

Query 9

Get all <subject> subelements by finding all metadata elements where the parentID matches the ID of a <subject> element from query 1.

```
SELECT dbo_Metadata.recordID, dbo_Metadata.metaRowID,
dbo_Metadata.propertyName, dbo_Metadata.propText,
dbo_Metadata.propTextOverflow INTO subject_subelements_temp
FROM dbo_Metadata, subject_exists_temp
WHERE subject_exists_temp.metaRowID=dbo_Metadata.parent_metaRowID;
```

Query 10

Group and count <subject> subelement propText values from results of Query 9.

```
SELECT propertyName, Count(propertyName) AS numOccurences
FROM subject_subelements_temp
GROUP BY propertyName;
```

Query 11

Group and count <subject> subelement occurrence values from results of Query 9.

```
SELECT numOccurences, COUNT(numOccurences) AS totalOccurencesCount
FROM subject66_subelements_record_count
GROUP BY numOccurences;
```

<location>

Query 1

Return instances of <url> elements. Write results into a temporary table.

```
SELECT dbo_Records.recordID, dbo_Metadata.propText,
dbo_Metadata.parent_propName, dbo_Metadata.a_usage INTO
location_url_temp_prelim
FROM dbo_Records INNER JOIN dbo_Metadata ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE ((dbo_Metadata.propertyName="url"));
```

Query 2

Return instances of <url> from Query 1 that have <location> elements as parents. Write results into a temporary table.

```
SELECT * INTO location_url_temp
FROM location_url_temp_prelim
WHERE (parent_propName='location');
```

Query 3

Count number of occurrences of record IDs in results of Query 2. The number of rows in the result will be the number of distinct records containing <url> elements.

```
SELECT recordID, count(recordID) AS occurrences
FROM location_url_temp
GROUP BY recordID;
```

Query 4

Group and total occurrence values from the results of Query 3.

```
SELECT occurrences, COUNT(occurrences) AS totalOccurrencesCount
FROM location_url_count
GROUP BY occurrences;
```

Query 5

Select unique <url> propText values. There should be the same number of unique number of values as number of rows in the result of Query 3.

```
SELECT DISTINCT propText
FROM location_url_temp;
```

Query 6

From the results of Query 1, return all instances of <location><url> that also required usage="primary display" attribute/value pair.

```
SELECT * INTO location_url_valid_temp
FROM location_url_temp
WHERE (a_usage='primary display');
```

Query 7

Count number of occurrences of records IDs for each record in results of Query 6. The number of rows in the result will be the number of distinct records containing <url> elements with required usage="primary display" attribute/value pairs.

```
SELECT recordID, count(recordID) AS occurrences
FROM location_url_valid_temp
GROUP BY recordID;
```

Query 8

Group and total occurrence values from the results of Query 7.

```
SELECT occurrences, COUNT(occurrences) AS totalOccurrencesCount
FROM location_url_valid_count
GROUP BY occurrences;
```

<accessCondition>

Query 1

Return instances of <accessCondition> elements that are children of <mods> elements and the content of the <accessCondition> element is not null. Write results into a temporary table.

```
SELECT dbo_metadata.RecordID, dbo_metadata.propText,
```

```
dbo_metadata.propertyName INTO accessCondition_valid_temp
FROM dbo_Metadata INNER JOIN dbo_Records ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE ((propertyName='accessCondition') AND (parent_PropName="mods')
AND ('propText' Is Not Null));
```

Query 2

Return instances of <accessCondition> elements that are children of <mods> elements and the content of the <accessCondition> element is not null. Also, these instance must also include the required type="useAndReproduction" attribute/value pair. Write results into a temporary table.

```
SELECT dbo_metadata.RecordID, dbo_metadata.propText,
dbo_metadata.propertyName INTO accessCondition_valid_temp
FROM dbo_Metadata INNER JOIN dbo_Records ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE ((propertyName='accessCondition') AND(parent_PropName="mods') AND
(a_type="useAndReproduction") AND ('propText' Is Not Null));
```

Query 3

Select the distinct records from the results of Query 2.

```
SELECT DISTINCT recordID
FROM accessCondition_valid_temp;
```

Query 4

Group and count <accessCondition> content values from the results of Query 2.

```
SELECT propText, COUNT(propText) AS numOccurences
FROM accessCondition_valid_temp
GROUP BY propText;
```

(In the case of Repository 10, this query was modified to select the content on <accessCondition>elements that were missing type="useAndReproduction" attribute/value pairs.)

<recordInfo>

Query 1

Return instances of <recordInfo> elements. Write results into a temporary table.

```
SELECT dbo_metadata.RecordID, dbo_metadata.metaRowID,
dbo_metadata.propertyName INTO recordInfo_temp
FROM dbo_Metadata INNER JOIN dbo_Records ON
dbo_Records.recordID=dbo_Metadata.recordID
WHERE (propertyName='recordInfo');
```

Query 2

Return instances of <languageOfCataloging> elements that have the same parent metadata row ID as an ID of an instance of <recordInfo> from the results of Query 1. Write results into a temporary table.

```
SELECT dbo_metadata.RecordID, dbo_metadata.metaRowID,
dbo_metadata.propertyName INTO recordInfo_languageOfCataloging_temp
FROM dbo_metadata INNER JOIN recordInfo_temp ON
dbo_metadata.parent_metaRowID=recordInfo_temp.metaRowID
WHERE ((dbo_metadata.propertyName='languageOfCataloging'));
```

Query 3

Return instances of <languageOfCataloging type="code" authority="iso639-2b"> elements that have the same parent metadata row ID as an ID of an instance of <recordInfo> from the results of Query 1. Write results into a temporary table.

```
SELECT DISTINCT dbo_metadata.RecordID, dbo_metadata.metaRowID,
dbo_metadata.propertyName, dbo_metadata.propText,
dbo_metadata.propTextOverflow, dbo_metadata.a_authority INTO
recordInfo_languageTerm_validAtt_notNull_temp
FROM dbo_metadata INNER JOIN recordInfo_languageOfCataloging_temp ON
dbo_metadata.parent_metaRowID=recordInfo_languageOfCataloging_temp.meta
RowID WHERE ((dbo_metadata.propertyName='languageTerm') And
(dbo_metadata.a_authority='iso639-2b') And (dbo_metadata.a_type='code')
And (dbo_metadata.propText Is Not Null));
```

Query 4

Return instances of <languageOfCataloging type=Text"> elements that have the same parent metadata row ID as an ID of an instance of <recordInfo> from the results of Query 1. Write results into a temporary table.

```
SELECT DISTINCT dbo_metadata.RecordID, dbo_metadata.metaRowID,
dbo_metadata.propertyName, dbo_metadata.propText,
dbo_metadata.propTextOverflow, dbo_metadata.a_authority INTO
recordInfo_languageTerm_validAtt_notNull_temp2
FROM dbo_metadata INNER JOIN recordInfo_languageOfCataloging_temp ON
dbo_metadata.parent_metaRowID=recordInfo_languageOfCataloging_temp.meta
RowID
WHERE ((dbo_metadata.propertyName='languageTerm') And
(dbo_metadata.a_type='text') And (dbo_metadata.propText Is Not Null));
```

Query 5

Select and count <languageTerm type="code" authority="iso639-2b"> element content from the results of Query 3.

```
SELECT propText, COUNT(propText) AS numOccurences
FROM recordInfo_languageTerm_validAtt_notNull_temp
GROUP BY propText;
```

Query 6

Select and count <languageTerm type="text"> element content from the results of Query 4.

```
SELECT propText, COUNT(propText) AS numOccurences
FROM recordInfo_languageTerm_validAtt_notNull_temp2
GROUP BY propText;
```

Additional Query Modifications

Repository Limiting

Simple where clauses were added to each query to limit results to individual repository. These where clauses would test whether or not the value of the repoID field in the records table matched a particular repository ID. An example is: `WHERE dbo_records.repoID = 1`.

OAI Timestamp Limiting

For repositories 8 and 9 that include records create both before and after the release of

the DLF/Aquifer Implementation Guidelines for Shareable MODS Records 1.0, an additional limiter was necessary to generate results for both before and after the release of the aforementioned guidelines. In these instances, a limiter that checks whether or not a record's OAI datestamp is before or after the November, 2006 release date of the 1.0 version of the DLF/Aquifer MODS guidelines was applied.

- For records before the document release:
  `(dbo_Records.OAIDateStamp)<#10/31/2006#)`
- For records after the document release:
  `(dbo_Records.OAIDateStamp)>#10/31/2006#)`

Appendix D - Result Summaries

Overview

3.1 Summary of Number of Records with Required MODS Elements				
A	B	C	D	E
<titleInfo>	305386	**99.97%**	305386	**99.97%**
<typeOfResource>	281297	**92.09%**	281297	**92.09%**
<originInfo>	9170	**3.00%**	118825	**38.90%**
<language>	366	**0.12%**	165496	**54.18%**
<physicalDescription>	24356	**7.97%**	24356	**7.97%**
<subject>	201162	**65.85%**	201162	**65.85%**
<location>	1130	**0.37%**	281748	**92.23%**
<accessCondition>	179058	**58.62%**	305468	**100.00%**
<recordInfo>	304	**0.10%**	305468	**100.00%**

A – Element name

B – Number of records meeting element requirements

C – Percentage of records meeting element requirements

D – Number of records meeting element requirement after safe algorithmic normalization

E – Percentage of records meeting element requirement after safe algorithmic normalization

Result Summary by Repository

D.2 Summary of Number of Records with Required MODS Elements	
Repository 1 (304 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	100.00%
<typeOfResource>	100.00%
<originInfo>	0.00%
<language>	100.00%
<physicalDescription>	100.00%
<subject>	86.84%
<location>	100.00%
<accessCondition>	100.00%
<recordInfo>	100.00%
Repository 2 (852 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	100.00%
<typeOfResource>	100.00%
<originInfo>	0.00%
<language>	0.00%
<physicalDescription>	0.00%
<subject>	100.00%
<location>	0.00%
<accessCondition>	0.23%
<recordInfo>	0.00%

Repository 3 (859 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	100.00%
<typeOfResource>	100.00%
<originInfo>	99.77%
<language>	0.00%
<physicalDescription>	0.81%
<subject>	100.00%
<location>	0.00%
<accessCondition>	100.00%
<recordInfo>	0.00%
Repository 4 (62 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	100.00%
<typeOfResource>	98.39%
<originInfo>	0.00%
<language>	100.00%
<physicalDescription>	100.00%
<subject>	100.00%
<location>	100.00%
<accessCondition>	100.00%
<recordInfo>	0.00%
Repository 5 (766 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	100.00%
<typeOfResource>	100.00%
<originInfo>	71.02%
<language>	0.00%
<physicalDescription>	0.00%
<subject>	96.87%
<location>	99.74%
<accessCondition>	0.00%
<recordInfo>	0.00%
Repository 6 (372 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	100.00%
<typeOfResource>	100.00%
<originInfo>	0.00%
<language>	0.00%
<physicalDescription>	0.00%
<subject>	100.00%
<location>	0.00%
<accessCondition>	0.00%
<recordInfo>	0.00%

Repository 7 (14425 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	99.45%
<typeOfResource>	100.00%
<originInfo>	0.00%
<language>	0.00%
<physicalDescription>	100.00%
<subject>	99.98%
<location>	0.00%
<accessCondition>	100.00%
<recordInfo>	0.00%
Repository 8 (24167 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	100.00%
<typeOfResource>	0.00%
<originInfo>	0.00%
<language>	0.00%
<physicalDescription>	0.00%
<subject>	26.64%
<location>	0.00%
<accessCondition>	0.00%
<recordInfo>	0.00%
Repository 9 (254072 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	100.00%
<typeOfResource>	100.00%
<originInfo>	0.00%
<language>	0.00%
<physicalDescription>	0.00%
<subject>	66.89%
<location>	0.00%
<accessCondition>	64.31%
<recordInfo>	0.00%
Repository 10 (9589 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	100.00%
<typeOfResource>	100.00%
<originInfo>	81.02%
<language>	0.00%
<physicalDescription>	99.68%
<subject>	75.23%
<location>	0.00%
<accessCondition>	0.00%
<recordInfo>	0.00%

Total (305468 Records)	
Element	**% of Records Meeting element Requirements**
<titleInfo>	99.97%
<typeOfResource>	92.09%
<originInfo>	3.00%
<language>	0.12%
<physicalDescription>	7.97%
<subject>	65.85%
<location>	0.37%
<accessCondition>	58.62%
<recordInfo>	0.10%

Result Summary by Element

D.3 Summary of Number of Records with Required MODS Elements			
<titleInfo>			
Repo ID	**Total Records**	**Valid Records**	**%**
1	304	304	100.00%
2	852	852	100.00%
3	859	859	100.00%
4	62	62	100.00%
5	766	766	100.00%
6	372	372	100.00%
7	14425	14346	99.45%
8	24167	24166	100.00%
9	254072	254070	100.00%
10	9589	9589	100.00%
Total:	**305468**	**305386**	**99.97%**
<typeOfResource>			
Repo ID	**Total Records**	**Valid Records**	**%**
1	304	304	100.00%
2	852	852	100.00%
3	859	859	100.00%
4	62	61	98.39%
5	766	766	100.00%
6	372	372	100.00%
7	14425	14425	100.00%
8	24167	0	0.00%
9	254072	254069	100.00%
10	9589	9589	100.00%
Total:	**305468**	**281297**	**92.09%**

<originInfo>			
Repo ID	**Total Records**	**Valid Records**	**%**
1	304	0	0.00%
2	852	0	0.00%
3	859	857	99.77%
4	62	0	0.00%
5	766	544	71.02%
6	372	0	0.00%
7	14425	0	0.00%
8	24167	0	0.00%
9	254072	0	0.00%
10	9589	7769	81.02%
Total:	**305468**	**9170**	**3.00%**
<language>			
Repo ID	**Total Records**	**Valid Records**	**%**
1	304	304	100.00%
2	852	0	0.00%
3	859	0	0.00%
4	62	62	100.00%
5	766	0	0.00%
6	372	0	0.00%
7	14425	0	0.00%
8	24167	0	0.00%
9	254072	0	0.00%
10	9589	0	0.00%
Total:	**305468**	**366**	**0.12%**
<physicalDescription>			
Repo ID	**Total Records**	**Valid Records**	**%**
1	304	304	100.00%
2	852	0	0.00%
3	859	7	0.81%
4	62	62	100.00%
5	766	0	0.00%
6	372	0	0.00%
7	14425	14425	100.00%
8	24167	0	0.00%
9	254072	0	0.00%
10	9589	9558	99.68%
Total:	**305468**	**24356**	**7.97%**

<subject>			
Repo ID	**Total Records**	**Valid Records**	**%**
1	304	264	86.84%
2	852	852	100.00%
3	859	859	100.00%
4	62	62	100.00%
5	766	742	96.87%
6	372	372	100.00%
7	14425	14422	99.98%
8	24167	6438	26.64%
9	254072	169937	66.89%
10	9589	7214	75.23%
Total:	**305468**	**201162**	**65.85%**
<location>			
Repo ID	**Total Records**	**Valid Records**	**%**
1	304	304	100.00%
2	852	0	0.00%
3	859	0	0.00%
4	62	62	100.00%
5	766	764	99.74%
6	372	0	0.00%
7	14425	0	0.00%
8	24167	0	0.00%
9	254072	0	0.00%
10	9589	0	0.00%
Total:	**305468**	**1130**	**0.37%**
<accessCondition>			
Repo ID	**Total Records**	**Valid Records**	**%**
1	304	304	100.00%
2	852	2	0.23%
3	859	859	100.00%
4	62	62	100.00%
5	766	0	0.00%
6	372	0	0.00%
7	14425	14425	100.00%
8	24167	0	0.00%
9	254072	163406	64.31%
10	9589	0	0.00%
Total:	**305468**	**179058**	**58.62%**

<recordInfo>			
Repo ID	**Total Records**	**Valid Records**	**%**
1	304	304	100.00%
2	852	0	0.00%
3	859	0	0.00%
4	62	0	0.00%
5	766	0	0.00%
6	372	0	0.00%
7	14425	0	0.00%
8	24167	0	0.00%
9	254072	0	0.00%
10	9589	4	0.04%
Total:	**305468**	**308**	**0.10%**

Appendix E - Detailed Results

<titleInfo><title>

<titleInfo><title> Basic Results

E.1 <titleInfo><title> Basic Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
1	304	304	306	100.00%	1.01	1	1	1
2	852	852	1724	100.00%	2.02	2	2	3
3	859	859	2044	100.00%	2.38	2	2	1
4	62	62	161	100.00%	2.6	2	2	8
5	766	766	1100	100.00%	1.44	1	1	9
6	372	372	980	100.00%	2.63	2	2	11
7	14425	14346	29172	99.45%	2.03	2	2	3
8	24167	24166	24167	100.00%	1	1	1	1
9	254072	254070	492691	100.00%	1.94	2	2	8
10	9589	9589	14767	100.00%	1.54	1	1	32
Total:	**305468**	**305386**	**567112**	**99.97%**	**1.86**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <titleInfo><title> element

D – Total <titleInfo><title> elements

E – Percentage of records with at least one <titleInfo><title> element

F – Mean <titleInfo><title> elements per-record

G – Median

H – Mode

I – Range

<titleInfo><title> Subelement Results

E.2 <titleInfo><title> Other Subelements Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
1	304	0	0	0.00%	-	0	0	0
2	852	344	402	40.38%	1.17	1	1	3
3	859	320	320	37.25%	1	1	1	0
4	62	15	17	24.19%	1.13	1	1	1
5	766	565	700	73.76%	1.24	1	1	2
6	372	367	632	98.66%	1.72	1	1	18
7	14425	0	0	0.00%	-	0	0	0
8	24167	0	0	0.00%	-	0	0	0
9	254072	68319	69124	26.89%	1.01	1	1	3
10	9589	6297	8285	65.67%	1.32	1	1	7
Total:	**305468**	**76227**	**79480**	**24.95%**	**1.04**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <titleInfo><title> subelement

D – Total <titleInfo><title> subelements

E – Percentage of records with at least one <titleInfo><title> subelement

F – Mean <titleInfo><title> subelements per-record (that include a title subelement)

G – Median

H – Mode

I – Range

<titleInfo><title> Basic Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.3 <titleInfo><title> Basic Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Pre-DLF/Aquifer MODS 1.0								
8	23301	23301	23302	100.00%	1	1	1	1
9	216980	216980	427901	100.00%	1.97	2	2	7
Total	**240281**	**240281**	**451203**	**100.00%**	**1.88**			
Post-DLF/Aquifer MODS 1.0								
8	866	865	865	99.88%	1	1	1	0
9	37090	37090	64790	100.00%	1.75	2	2	8
Total	**37956**	**37955**	**451203**	**100.00%**	**11.89**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <titleInfo><title> element

D – Total <titleInfo><title> elements

E – Percentage of records with at least one <titleInfo><title> element

F – Mean <titleInfo><title> elements per-record

G – Median

H – Mode

I – Range

<titleInfo><title> Subelement Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.4 <titleInfo><title> Other Subelements Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Pre-DLF/Aquifer MODS 1.0								
8	23301	0	0	0.00%	-	0	0	0
9	216980	65149	65900	30.03%	1.01	1	1	3
Total	**240281**	**65149**	**65900**	**27.11%**	**1.01**			
Post-DLF/Aquifer MODS 1.0								
8	866	0	0	0.00%	-	0	0	0
9	37090	3170	3224	8.55%	1.02	1	1	2
Total	**37956**	**3170**	**3224**	**8.35%**	**1.02**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <titleInfo><title> subelement

D – Total <titleInfo><title> subelements

E – Percentage of records with at least one <titleInfo><title> subelement

F – Mean <titleInfo><title> subelements per-record (that include a title subelement)

G – Median

H – Mode

I – Range

<titleInfo><title> Subelement Data

Repository 1	
<title> Subelement Usage	
Subelement	**Occurrences**
subTitle	0
nonSort	0
partNumber	0
partName	0
Total:	0

Repository 2	
<title> Subelement Usage	
Subelement	**Occurrences**
subTitle	272
nonSort	109
partNumber	19
partName	2
Total:	402

Repository 3	
<title> Subelement Usage	
Subelement	**Occurrences**
subTitle	0
nonSort	0
partNumber	320
partName	0
Total:	320

Repository 4	
<title> Subelement Usage	
Subelement	**Occurrences**
subTitle	0
nonSort	17
partNumber	0
partName	0
Total:	17

Repository 5	
<title> Subelement Usage	
Subelement	**Occurrences**
subTitle	339
nonSort	201
partNumber	150
partName	10
Total:	700

Repository 6	
<title> Subelement Usage	
Subelement	**Occurrences**
subTitle	112
nonSort	13
partNumber	480
partName	27
Total:	632

Repository 7	
<title> Subelement Usage	
Subelement	Occurrences
subTitle	0
nonSort	0
partNumber	0
partName	0
Total:	0

Repository 8	
<title> Subelement Usage	
Subelement	Occurrences
subTitle	0
nonSort	0
partNumber	0
partName	0
Total:	0

Repository 9	
<title> Subelement Usage	
Subelement	Occurrences
subTitle	7446
nonSort	18233
partNumber	47
partName	43398
Total:	69124

Repository 10	
<title> Subelement Usage	
Subelement	Occurrences
subTitle	4622
nonSort	3382
partNumber	115
partName	166
Total:	8285

Total	
<title> Subelement Usage	
Subelement	Occurrences
subTitle	12791
nonSort	21955
partNumber	1131
partName	43603
Total:	79480

<titleInfo><title> <nonSort> Values

Repository 1	
<nonSort> Values	
Value	Occurrences
Total:	0

Repository 2	
<nonSort> Values	
Value	Occurrences
The	78
A	25
An	6

Repository 3	
<nonSort> Values	
Value	Occurrences
Total:	0

Repository 4	
<nonSort> Values	
Value	Occurrences
The	14
An	1
A	1
"	1

Repository 5	
<nonSort> Values	
Value	Occurrences
The	161
A	35
An	3
Eine	1
Das	1

Repository 6	
<nonSort> Values	
Value	Occurrences
Les	4
Le	2
La	2
"Les	2
"La	2
"Il	1

Repository 7	
<nonSort> Values	
Value	Occurrences
Total:	0

Repository 8	
<nonSort> Values	
Value	Occurrences
Total:	0

Repository 9	
<nonSort> Values	
Value	Occurrences
The	10896
A	2286
[The	1798
La	779
[	558
An	455
[A	237
Le	184
...	151
... An	135
Les	117
"The	86
"	51
L'	37
Il	36
De	32
(	31
El	26
... The	25
[La	22
[An	21
Die	19
The "	19
"A	18
Une	17
Der	16
[Le	14

Repository 10	
<nonSort> Values	
Value	Occurrences
The	2367
A	721
An	175
Die	44
Les	15
Das	10
La	7
Der	6
Le	6
"A	4
De	4
L'	4
"The	3
I	2
Il	2
[The	1
C	1
(The	1
Eine	1
I. The	1
Zur	1
L'é	1
Life	1
Ü	1
Un	1
Wing	1
El	1

... A	13
Das	12
Un	6
In the	6
*	4
At	3
["	3
Lo	3
[Les	3
[L'	2
In	2
Des	2
I	2
[In	2
...An a	2
A "	2
D'	2

On the	2
M	2
La	2
Una	2
Le	2
[Cap	1
[Cwe	1
[Das	1
[Dos	1
[Vi	1
[Il	1
[Unc	1
[Lla	1
[Saw	1
[N	1
[San	1
Team	1
[Pano	1
T[he?]	1
[Old	1
Rus	1
U	1
-	1
Yale	1
West	1
"La	1
"Le	1
"Lo	1
"Un	1
&#x2026	1
.	1
The [	1
Um	1
[Cad	1
The n	1
The C	1
...The	1
? ? ? a	1
[View	1
["The	1
The (	1
The '	1
[And	1
..	1
Ir	1

Farm	1
Fight	1
Follo	1
For	1
Fuji	1
Gran	1
Have	1
Het	1
Ils	1
Roya	1
In o	1
Do y	1
Iro	1
On	1
Ol	1
Las	1
New	1
Le K	1
Mrs.	1
Los	1
Ma	1
Mine	1
In a	1
An a	1
+ J	1
1	1
500	1
Dees	1
A b	1
Moon	1
A d	1
A g	1
A. B	1
A. J. B	1
Pa	1
Refl	1
Dr.	1
And	1
Anot	1
Ci	1
'	1
Rail	1
Zum	1
R	1
Popu	1
Dire	1
[Wom	1
A. L	1

<typeOfResource>

<typeOfResource> Basic Results

E.5 <typeOfResource> Basic Results					
A	**B**	**C**	**D**	**E**	**F**
1	304	304	304	304	100.00%
2	852	852	852	852	100.00%
3	859	859	861	861	100.00%
4	62	61	144	144	98.39%
5	766	766	766	766	100.00%
6	372	372	372	372	100.00%
7	14425	14425	14425	14425	100.00%
8	24167	0	0	0	0.00%
9	254072	254069	254069	254070	100.00%
10	9589	9589	9589	9589	100.00%
Total:	**305468**	**281297**	**281382**	**281383**	**92.09%**

A – Repository ID

B – Records in test set

C – Distinct records containing valid <typeOfResource> elements

D – Total valid <typeOfResource> elements

E – Total <typeOfResource> elements

F – Percentage of records with at least one valid <typeOfResource> element

<typeOfResource> Basic Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.6 <typeOfResource> Basic Results					
A	**B**	**C**	**D**	**E**	**F**
Pre-DLF/Aquifer MODS 1.0					
8	23301	0	0	0	0.00%
9	216980	216979	216979	216980	100.00%
Total:	**240281**	**216979**	**216979**	**216980**	**90.30%**
Pre-DLF/Aquifer MODS 1.0					
8	866	0	0	0	0.00%
9	37090	37090	37090	37090	100.00%
Total:	**37956**	**37090**	**37090**	**37090**	**97.72%**

A – Repository ID

B – Records in test set

C – Distinct records containing valid <typeOfResource> elements

D – Total valid <typeOfResource> elements

E – Total <typeOfResource> elements

F – Percentage of records with at least one valid <typeOfResource> element

<typeOfResource> Values

Repository 1	
<typeOfResource> Value	**Occurrences**
text	304
Total:	304

Repository 2	
<typeOfResource> Value	**Occurrences**
text	849
moving image	3
Total:	852

Repository 3	
<typeOfResource> Value	**Occurrences**
still image	467
text	394
Total:	861

Repository 4	
<typeOfResource> Value	**Occurrences**
text	62
moving image	28
still image	25
cartographic	17
sound recording-nonmusical	4
sound recording-musical	4
sound recording	1
software, multimedia	1
notated music	1
mixed material	1
Total:	144

Repository 5	
<typeOfResource> Value	**Occurrences**
text	766
Total:	766

Repository 6	
<typeOfResource> Value	**Occurrences**
notated music	371
mixed material	1
Total:	372

Repository 7	
<typeOfResource> Value	**Occurrences**
still image	14425
Total:	14425

Repository 8	
No values	

Repository 9	
<typeOfResource> Value	**Occurrences**
still image	172190
notated music	62977
text	11330
cartographic	6956
moving image	616
Total:	254069

Repository 10	
<typeOfResource> Value	**Occurrences**
text	9436
cartographic	141
software, multimedia	8
notated music	4
Total:	9589

Total	
<typeOfResource> Value	**Occurrences**
still image	187107
notated music	63353
text	23141
cartographic	7114
moving image	647
software, multimedia	9
sound recording-nonmusical	4
sound recording-musical	4
mixed material	2
sound recording	1
Total:	281382

<originInfo>

<originInfo> Basic Results

E.7 <originInfo> Basic Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
1	304	304	304	100.00%	1	1	1	1
2	852	852	1236	100.00%	1.45	1	1	4
3	859	857	878	99.77%	1.02	1	1	3
4	62	60	60	96.77%	1	1	1	1
5	766	733	1478	95.69%	2.02	2	2	4
6	372	370	370	99.46%	1	1	1	1
7	14425	14425	43239	100.00%	3	3	3	2
8	24167	24167	24429	100.00%	1.01	1	1	1
9	254072	253468	566084	99.76%	2.23	2	2	4
10	9589	9588	12167	99.99%	1.27	1	1	4
Total:	**305468**	**304824**	**650245**	**99.79%**	**2.13**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one recommended <originInfo> date-related subelement

D – Total <originInfo> recommended date subelements

E – Percentage of records with at least one <originInfo> date-related subelement

F – Mean <originInfo> date-related subelements

G – Median

H – Mode

I – Range

<originInfo> keyDate Results

E.8 <originInfo> keyDate Results (subset of E.7)				
A	**B**	**C**	**D**	**E**
1	304	0	0	0.00%
2	852	0	0	0.00%
3	859	857	857	99.77%
4	62	0	0	0.00%
5	766	544	544	71.02%
6	372	0	0	0.00%
7	14425	14407	28814	99.88%
8	24167	0	0	0.00%
9	254072	0	0	0.00%
10	9589	9588	12167	99.99%
Total:	**305468**	**25396**	**42382**	**8.31%**

A – Repository ID

B – Records in test set

C – Distinct records with <originInfo> keyDate

D – Total <originInfo> keyDate instances

E – Percentage of records with at least one <originInfo> keyDate

Note: In Repository 17 each record includes two keyDates

Note: In repository 10, approximately 1,819 records have more than one keyDate

<originInfo> Single keyDate Results

E.9 <originInfo> Single keyDate Results (subset of E.8)				
A	**B**	**C**	**D**	**E**
1	304	0	0	0.00%
2	852	0	0	0.00%
3	859	857	857	99.77%
4	62	0	0	0.00%
5	766	544	544	71.02%
6	372	0	0	0.00%
7	14425	0	14407	0.00%
8	24167	0	0	0.00%
9	254072	0	0	0.00%
10	9589	7769	9588	81.02%
Total:	**305468**	**9170**	**25396**	**3.00%**

A – Repository ID

B – Records in test set

C – Distinct records with one and only one <originInfo> keyDate

D – Total <originInfo> keyDate instances

E – Percentage of records with one and only one <originInfo> keyDate

<originInfo> Basic Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.10 <originInfo> Basic Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Pre-DLF/Aquifer MODS 1.0								
8	23301	23301	23320	100.00%	1	1	1	1
9	216980	216378	469738	99.72%	2.17	2	2	4
Total:	**240281**	**239679**	**493058**	**99.75%**	**2.06**			
Post-DLF/Aquifer MODS 1.0								
8	866	866	909	100.00%	1.05	1	1	1
9	37090	37090	96346	100.00%	2.6	2	2	4
Total:	**37956**	**517988**	**1084045**	**1364.71%**	**2.09**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one recommended <originInfo> date-related subelement

D – Total <originInfo> recommended date subelements

E – Percentage of records with at least one <originInfo> date-related subelement

F – Mean <originInfo> date-related subelements

G – Median

H – Mode

I – Range

<originInfo> Subelement Occurrences and keyDate Elements

Repo 1: Recommended <originInfo> subelements	
Element name	**Number**
dateIssued	192
dateCreated	0
copyrightDate	24
dateOther	88
Total:	304

Repo 1: keyDate elements	
Element name	**Number**
dateIssued	0
dateCreated	0
copyrightDate	0
dateOther	0
Total:	0

Repo 2: Recommended <originInfo> subelements	
Element name	**Number**
dateIssued	1236
dateCreated	0
copyrightDate	0
dateOther	0
Total:	1236

Repo 2: keyDate elements	
Element name	**Number**
dateIssued	0
dateCreated	0
copyrightDate	0
dateOther	0
Total:	0

Repo 3: Recommended <originInfo> subelements	
Element name	**Number**
dateIssued	0
dateCreated	544
copyrightDate	0
dateOther	334
Total:	878

Repo 3: keyDate elements	
Element name	**Number**
dateIssued	320
dateCreated	537
copyrightDate	0
dateOther	0
Total:	857

Repo 4: Recommended <originInfo> subelements	
Element name	**Number**
dateIssued	60
dateCreated	0
copyrightDate	0
dateOther	0
Total:	60

Repo 4: keyDate elements	
Element name	**Number**
dateIssued	0
dateCreated	0
copyrightDate	0
dateOther	0
Total:	0

Repo 5: Recommended <originInfo> subelements	
Element name	**Number**
dateIssued	1477
dateCreated	0
copyrightDate	1
dateOther	0
Total:	1478

Repo 5: keyDate elements	
Element name	**Number**
dateIssued	544
dateCreated	0
copyrightDate	0
dateOther	0
Total:	544

Repo 6: Recommended <originInfo> subelements	
Element name	Number
dateIssued	370
dateCreated	0
copyrightDate	0
dateOther	0
Total:	370

Repo 6: keyDate elements	
Element name	Number
dateIssued	544
dateCreated	0
copyrightDate	0
dateOther	0
Total:	544

Repo 7: Recommended <originInfo> subelements	
Element name	Number
dateIssued	0
dateCreated	28814
copyrightDate	14425
dateOther	0
Total:	43239

Repo 7: keyDate elements	
Element name	Number
dateIssued	0
dateCreated	28814
copyrightDate	0
dateOther	0
Total:	28814

Repo 8: Recommended <originInfo> subelements	
Element name	Number
dateIssued	24214
dateCreated	0
copyrightDate	0
dateOther	15
Total:	24229

Repo 8: keyDate elements	
Element name	Number
dateIssued	0
dateCreated	0
copyrightDate	0
dateOther	0
Total:	0

Repo 9: Recommended <originInfo> subelements	
Element name	Number
dateIssued	565979
dateCreated	15
copyrightDate	90
dateOther	0
Total:	566084

Repo 9: keyDate elements	
Element name	Number
dateIssued	0
dateCreated	0
copyrightDate	0
dateOther	0
Total:	0

Repo 10: Recommended <originInfo> subelements	
Element name	Number
dateIssued	12167
dateCreated	0
copyrightDate	0
dateOther	0
Total:	12167

Repo 10: keyDate elements	
Element name	Number
dateIssued	12167
dateCreated	0
copyrightDate	0
dateOther	0
Total:	12167

<originInfo> Declared Encoding Schema

Repository 1	
Encoding attribute values	**Occurrences**
w3cdtf	304
Total:	304

Repository 2	
Encoding attribute values	**Occurrences**
marc	386
Total:	386

Repository 3	
Encoding attribute values	**Occurrences**
w3cdtf	771
Total:	771

Repository 4	
Encoding attribute values	**Occurrences**
w3cdtf	60
Total:	60

Repository 5	
Encoding attribute values	**Occurrences**
w3cdtf	791
Total:	791

Repository 6	
Encoding attribute values	**Occurrences**
Total:	0

Repository 7	
Encoding attribute values	**Occurrences**
w3cdtf	43239
Total:	43239

Repository 8	
Encoding attribute values	**Occurrences**
iso8601	24229
Total:	24229

Repository 9	
Encoding attribute values	**Occurrences**
marc	312946
Total:	312946

Repository 10	
Encoding attribute values	**Occurrences**
marc	2631
Total:	2631

<language>

<language> Basic Results

E.11 <language> Basic Results				
A	**B**	**C**	**D**	**E**
1	304	304	304	100.00%
2	852	852	853	100.00%
3	859	859	859	100.00%
4	62	62	62	100.00%
5	766	766	770	100.00%
6	372	14	21	3.76%
7	14425	0	0	0.00%
8	24167	24167	24926	100.00%
9	254072	152108	152736	59.87%
10	9589	9577	9891	99.87%
Total:	**305468**	**188709**	**190422**	**61.78%**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <language> element

D – Total <language> elements

E – Percentage of records with at least one <language> element

<language><languageTerm> Basic Results

E.12 <language><languageTerm> Basic Results				
A	**B**	**C**	**D**	**E**
1	304	304	608	100.00%
2	852	852	853	100.00%
3	859	859	859	100.00%
4	62	62	124	100.00%
5	766	766	770	100.00%
6	372	14	21	3.76%
7	14425	0	0	0.00%
8	24167	24167	24926	100.00%
9	254072	152108	152736	59.87%
10	9589	9577	9891	99.87%
Total:	**305468**	**188709**	**190788**	**61.78%**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <language><languageTerm> element

D – Total <language><languageTerm> elements

E – Percentage of records with at least one <language><languageTerm> element

<languageTerm type="code" authority="iso639-2b"> Basic Results

E.13 <languageTerm type="code" authority="iso639-2b"> Basic Results				
A	**B**	**C**	**D**	**E**
1	304	304	304	100.00%
2	852	852	853	100.00%
3	859	859	859	100.00%
4	62	62	62	100.00%
5	766	766	770	100.00%
6	372	14	21	3.76%
7	14425	0	0	0.00%
8	24167	0	0	0.00%
9	254072	152108	152736	59.87%
10	9589	9577	9891	99.87%
Total:	**305468**	**164542**	**165496**	**53.87%**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <languageTerm type="code" authority="iso639-2b"> element

D – Total <languageTerm type="code" authority="iso639-2b"> elements

E – Percentage of records with at least one <languageTerm type="code" authority="iso639-2b"> element

<languageTerm type="text"> Basic Results

E.14 <languageTerm type="text"> Basic Results				
A	**B**	**C**	**D**	**E**
1	304	304	304	100.00%
2	852	0	0	0.00%
3	859	0	0	0.00%
4	62	62	62	100.00%
5	766	0	0	0.00%
6	372	0	0	0.00%
7	14425	0	0	0.00%
8	24167	0	0	0.00%
9	254072	0	0	0.00%
10	9589	0	0	0.00%
Total:	**305468**	**366**	**366**	**0.12%**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <languageTerm type="text"> element

D – Total <languageTerm type="text"> elements

E – Percentage of records with at least one <languageTerm type="text"> element

Valid <language> elements with All Required Subelements Results

E.15 Records with valid <language> elements and all required subelements			
A	**B**	**C**	**D**
1	304	304	100.00%
2	852	0	0.00%
3	859	0	0.00%
4	62	62	100.00%
5	766	0	0.00%
6	372	0	0.00%
7	14425	0	0.00%
8	24167	0	0.00%
9	254072	0	0.00%
10	9589	0	0.00%
Total:	**305468**	**366**	**0.12%**

A – Repository ID

B – Records in test set

C – Records with all required <language> elements and subelements

D – Percentage of records with all required <language> elements and subelements

<language> Basic Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.16 <language> Basic Results				
A	**B**	**C**	**D**	**E**
Pre-DLF/Aquifer MODS 1.0				
8	23301	23301	24037	100.00%
9	216980	123000	123603	56.69%
Total:	**240281**	**146301**	**147640**	**60.89%**
Post-DLF/Aquifer MODS 1.0				
8	866	866	889	100.00%
9	37090	29108	29133	78.48%
Total:	**37956**	**29974**	**30022**	**78.97%**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <language> element

D – Total <language> elements

E – Percentage of records with at least one <language> element

<language><languageTerm> Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.17 <language><languageTerm> Basic Results				
A	**B**	**C**	**D**	**E**
Pre-DLF/Aquifer MODS 1.0				
8	23301	23301	24037	100.00%
9	216980	123000	123603	56.69%
Total:	**240281**	**146301**	**147640**	**60.89%**
Post-DLF/Aquifer MODS 1.0				
8	866	866	889	100.00%
9	37090	29108	29133	78.48%
Total:	**37956**	**29974**	**30022**	**78.97%**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <language><languageTerm> element

D – Total <language><languageTerm> elements

E – Percentage of records with at least one <language><languageTerm> element

<languageTerm type=”code” authority=”iso639-2b”> Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.18 <languageTerm type=”code” authority=”iso639-2b”> Basic Results				
A	**B**	**C**	**D**	**E**
Pre-DLF/Aquifer MODS 1.0				
8	23301	0	0	0.00%
9	216980	123000	123603	56.69%
Total:	**240281**	**123000**	**123603**	**51.19%**
Post-DLF/Aquifer MODS 1.0				
8	866	0	0	0.00%
9	37090	29108	29133	78.48%
Total:	**37956**	**29108**	**29133**	**76.69%**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <languageTerm type=”code” authority=”iso639-2b”> element

D – Total <languageTerm type=”code” authority=”iso639-2b”> elements

E – Percentage of records with at least one <languageTerm type=”code” authority=”iso639-2b”> element

<languageTerm type="text"> Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.19 <languageTerm type="text"> Basic Results				
A	**B**	**C**	**D**	**E**
Pre-DLF/Aquifer MODS 1.0				
8	23301	0	0	0.00%
9	216980	0	0	0.00%
Total:	**240281**	**0**	**0**	**0.00%**
Post-DLF/Aquifer MODS 1.0				
8	866	0	0	0.00%
9	37090	0	0	0.00%
Total:	**37956**	**0**	**0**	**0.00%**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <languageTerm type="text"> element

D – Total <languageTerm type="text"> elements

E – Percentage of records with at least one <languageTerm type="text"> element

Valid <language> elements with All Required Subelements Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.20 Records with valid <language> elements and all required subelements			
A	**B**	**C**	**D**
Pre-DLF/Aquifer MODS 1.0			
8	23301	0	0.00%
9	216980	0	0.00%
Total:	**240281**	**0**	**0.00%**
Post-DLF/Aquifer MODS 1.0			
8	866	0	0.00%
9	37090	0	0.00%
Total:	**37956**	**0**	**0.00%**

A – Repository ID

B – Records in test set

C – Records with all required <language> elements and subelements

D – Percentage of records with all required <language> elements and subelements

<language><languageTerm> Values

Repository 1			
<langaugeTerm> Type=code		**<languageTerm> Type=code**	
Value	**Occurrences**	**Value**	**Occurrences**
English	304	eng	304
Total:	304	**Total:**	304

Repository 2			
<langaugeTerm> Type=code		<languageTerm> Type=code	
Value	Occurrences	Value	Occurrences
		eng	852
		fre	1
Total:		Total:	853

Repository 3			
<langaugeTerm> Type=code		<languageTerm> Type=code	
Value	Occurrences	Value	Occurrences
	0	eng	859
Total:	0	Total:	859

Repository 4			
<langaugeTerm> Type=code		<languageTerm> Type=code	
Value	Occurrences	Value	Occurrences
English	62	eng	62
Total:	62	Total:	62

Repository 5			
<langaugeTerm> Type=code		<languageTerm> Type=code	
Value	Occurrences	Value	Occurrences
		eng	717
		und	33
		ger	7
		fre	5
		swe	2
		mul	2
		san	1
		ita	1
		dut	1
		cze	1
Total:	0	Total:	770

Repository 6			
<langaugeTerm> Type=code		<languageTerm> Type=code	
Value	Occurrences	Value	Occurrences
		ger	15
		fre	5
		eng	1
Total:	0	Total:	21

Repository 7			
<langaugeTerm> Type=code		<languageTerm> Type=code	
Value	Occurrences	Value	Occurrences
Total:	0	**Total:**	0

Repository 8
Note: All results from repository 8 were either <language><languag or <language><languageTerm>eng</languageTerm></language>.

Repository 9			
<langaugeTerm> Type=code		<languageTerm> Type=code	
Value	Occurrences	Value	Occurrences
		eng	114195
		und	16161
		|kn	13499
		rus	2491
		mul	1790
		ota	1328
		fre	1221
		ger	529
		spa	378
		lat	332
		ita	284
		jpn	283
		dut	52
		chi	33
		por	22
		swe	9
		wel	9
		5	7
		1.A	7
		vo	7
		l.	7
		00M	7
		0	7
		e C	6
		las	6
		12V	6
		Cas	6
		###	4
		yid	4
		kor	3
		dan	2
		--	2
		58	2
		cze	2
		grc	2

		haw	2
		53	1
		48	1
		12l	1
		ara	1
		?	1
		1 C	1
		[en	1
		61	1
		83	1
		afr	1
		42	1
		ind	1
		ul	1
		tur	1
		san	1
		pol	1
		oto	1
		nah	1
		mar	1
		er	1
		iro	1
		arm	1
		ice	1
		gre	1
		fro	1
		frm	1
		Cl	1
		at	1
		ass	1
		ase	1
		k	1
		Total:	152736

Repository 10			
<langaugeTerm> Type=code		**<languageTerm> Type=code**	
Value	**Occurrences**	**Value**	**Occurrences**
		eng	9068
		ger	371
		fre	236
		ita	56
		und	52
		lat	30
		dut	16
		rus	12
		spa	9
		grc	9

		gre	6
		dan	5
		swe	3
		hun	2
		gem	1
		chi	1
		chn	1
		syr	1
		sio	1
		san	1
		iri	1
		por	1
		jpn	1
		pol	1
		nor	1
		nai	1
		heb	1
		.	1
		mul	1
		ern	1
		Total:	9891

<physicalDescription>

<physicalDescription> Basic Results

E.21 <physicalDescription> Basic Results				
A	**B**	**C**	**D**	**E**
1	304	304	304	100.00%
2	852	852	852	100.00%
3	859	859	859	100.00%
4	62	62	62	100.00%
5	766	766	766	100.00%
6	372	372	372	100.00%
7	14425	14425	14425	100.00%
8	24167	24167	133979	100.00%
9	254072	254068	254068	100.00%
10	9589	9558	9558	99.68%
Total:	**305468**	**305433**	**415245**	**99.99%**

A – Repository ID

B – Records in test set

C – Distinct records with <physicalDescription>

D – Total <physicalDescription> elements

E – Percentage of records with one <physicalDescription> element

<physicalDescription><digitalOrigin> Results

E.22 <physicalDescription><digitalOrigin> Results				
A	**B**	**C**	**D**	**E**
1	304	304	304	100.00%
2	852	0	0	0.00%
3	859	7	7	0.81%
4	62	62	62	100.00%
5	766	0	0	0.00%
6	372	0	0	0.00%
7	14425	14425	14425	100.00%
8	24167	0	0	0.00%
9	254072	0	0	0.00%
10	9589	9559	9560	99.69%
Total:	**305468**	**24357**	**24358**	**7.97%**

A – Repository ID

B – Records in test set

C – Distinct records with <physicalDescription><digitalOrigin>

D – Total <physicalDescription><digitalOrigin> elements

E – Percentage of records with one <physicalDescription><digitalOrigin> element

<physicalDescription><internetMediaType> Results

E.23 <physicalDescription><internetMediaType> Results					
A	**B**	**C**	**D**	**E**	**F**
1	304	304	336	100.00%	1.11
2	852	2	2	0.23%	0
3	859	859	859	100.00%	1
4	62	62	196	100.00%	3.16
5	766	0	0	0.00%	0
6	372	0	0	0.00%	0
7	14425	14425	14425	100.00%	1
8	24167	24167	66160	100.00%	2.74
9	254072	40	40	0.02%	0
10	9589	9559	19120	99.69%	1.99
Total:	**305468**	**49418**	**101138**	**16.18%**	**0.33**

A – Repository ID

B – Records in test set

C – Distinct records with <physicalDescription><internetMediaType>

D – Total <physicalDescription><internetMediaType> elements

E – Percentage of records with <physicalDescription><internetMediaType> elements

F – Average number of <internetMediaType> elements

<physicalDescription> Valid & Complete Results

E.24 Valid/Complete <physicalDescription> Results			
A	**B**	**C**	**D**
1	304	304	100.00%
2	852	0	0.00%
3	859	7	0.81%
4	62	62	100.00%
5	766	0	0.00%
6	372	0	0.00%
7	14425	14425	100.00%
8	24167	0	0.00%
9	254072	0	0.00%
10	9589	9558	99.68%
Total:	**305468**	**24356**	**7.97%**

A – Repository ID

B – Records in test set

C – Records with valid and complete <physicalDescription> elements

D – Percentage of records with valid and complete <physicalDescription> elements

<physicalDescription> Basic Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.25 <physicalDescription> Basic Results				
A	**B**	**C**	**D**	**E**
Pre-DLF/Aquifer MODS 1.0				
8	23301	23301	125848	100.00%
9	216980	216979	216979	100.00%
Total:	**240281**	**240280**	**342827**	**100.00%**
Post-DLF/Aquifer MODS 1.0				
8	866	866	8131	100.00%
9	37090	37089	37089	100.00%
Total:	**37956**	**37955**	**45220**	**100.00%**

A – Repository ID

B – Records in test set

C – Distinct records with <physicalDescription>

D – Total <physicalDescription> elements

E – Percentage of records with one <physicalDescription> element

<physicalDescription><digitalOrigin> Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.26 <physicalDescription><digitalOrigin> Results				
A	**B**	**C**	**D**	**E**
Pre-DLF/Aquifer MODS 1.0				
8	23301	0	0	0.00%
9	216980	0	0	0.00%
Total:	**240281**	**0**	**0**	**0.00%**
Post-DLF/Aquifer MODS 1.0				
8	866	0	0	0.00%
9	37090	0	0	0.00%
Total:	**37956**	**0**	**0**	**0.00%**

A – Repository ID

B – Records in test set

C – Distinct records with <physicalDescription><digitalOrigin>

D – Total <physicalDescription><digitalOrigin> elements

E – Percentage of records with one <physicalDescription><digitalOrigin> element

<physicalDescription><internetMediaType> Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.27 <physicalDescription><internetMediaType> Results					
A	**B**	**C**	**D**	**E**	**F**
Pre-DLF/Aquifer MODS 1.0					
8	23301	23301	62096	100.00%	2.66
9	216980	40	40	0.02%	0
Total:	**240585**	**23645**	**62472**	**9.83%**	**0.26**
Post-DLF/Aquifer MODS 1.0					
8	866	866	4064	100.00%	4.69
9	37090	0	0	0.00%	0
Total:	**37956**	**866**	**4064**	**2.28%**	**0.11**

A – Repository ID

B – Records in test set

C – Distinct records with <physicalDescription><internetMediaType>

D – Total <physicalDescription><internetMediaType> elements

E – Percentage of records with <physicalDescription><internetMediaType> elements

F – Average number of <internetMediaType> elements

<physicalDescription> Valid & Complete Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.28 Valid/Complete <physicalDescription> Results			
A	**B**	**C**	**D**
Pre-DLF/Aquifer MODS 1.0			
8	23301	0	0.00%
9	216980	0	0.00%
Total:	**240281**	**0**	**0.00%**
Post-DLF/Aquifer MODS 1.0			
8	866	0	0.00%
9	37090	0	0.00%
Total:	**37956**	**0**	**0.00%**

A – Repository ID

B – Records in test set

C – Records with valid and complete <physicalDescription> elements

D – Percentage of records with valid and complete <physicalDescription> elements

<physicalDescription><digitalOrigin> Content

Repo 1: digitalOrigin Values	
Value	**Number**
born digital	91
reformatted digital	213
digitized microfilm	0
digitized other analog	0
Total:	304

Repo 2: digitalOrigin Values	
Value	**Number**
born digital	0
reformatted digital	0
digitized microfilm	0
digitized other analog	0
Total:	0

Repo 3: digitalOrigin Values	
Value	**Number**
born digital	0
reformatted digital	7
digitized microfilm	0
digitized other analog	0
Total:	7

Repo 4: digitalOrigin Values	
Value	**Number**
born digital	62
reformatted digital	0
digitized microfilm	0
digitized other analog	0
Total:	62

Repo 5: digitalOrigin Values	
Value	**Number**
born digital	0
reformatted digital	0
digitized microfilm	0
digitized other analog	0
Total:	0

Repo 6: digitalOrigin Values	
Value	**Number**
born digital	0
reformatted digital	0
digitized microfilm	0
digitized other analog	0
Total:	0

Repo 7: digitalOrigin Values	
Value	**Number**
born digital	0
reformatted digital	14425
digitized microfilm	0
digitized other analog	0
Total:	14425

Repo 8: digitalOrigin Values	
Value	**Number**
born digital	0
reformatted digital	0
digitized microfilm	0
digitized other analog	0
Total:	0

Repo 9: digitalOrigin Values	
Value	**Number**
born digital	0
reformatted digital	0
digitized microfilm	0
digitized other analog	0
Total:	0

Repo 10: digitalOrigin Values	
Value	**Number**
born digital	0
reformatted digital	9560
digitized microfilm	0
digitized other analog	0
Total:	9560

<physicalDescription><internetMediaType> Content

Repository 1	
<internetMediaType> Content	**Occurrences**
text/html	302
text/plain	31
application/pdf	3
Total:	**336**

Repository 2	
<internetMediaType> Content	**Occurrences**
text/html; charset=iso-8859-1	1
text/html	1
Total:	**2**

Repository 3	
<internetMediaType> Content	**Occurrences**
image/jpeg	472
Te[illegible]	[illegible]

Repository 4	
<internetMediaType> Content	**Occurrences**
text/html	59
video/quicktime	31
image/jpeg	31
application/vnd.rn-realmedia	30
video/x-ms-wmv	27
application/http	7
image/gif	4
audio/x-realaudio	3
application/x-shockwave-flash	2
audio/x-ms-wma	1
application/pdf	1
Total:	**196**

Repository 5	
<internetMediaType> Content	**Occurrences**
Total:	**0**

Repository 6	
<internetMediaType> Content	**Occurrences**
Total:	**0**

Repository 7	
<internetMediaType> Content	**Occurrences**
image/jpeg	14425
Total:	**14425**

Repository 8	
<internetMediaType> Content	**Occurrences**
text/plain	41980
application/pdf	24154
application/octet-stream	13
image/jpeg	11
text/html	1
application/vnd.ms-excel	1
Total:	**66160**

Repository 9	
<internetMediaType> Content	**Occurrences**
p coftsign})	34
pp	2
p999aX	1
p})	1
p sia)	1
9878 p	1
Total:	**40**

Total	
<internetMediaType> Content	**Occurences**
text/plain	42011
application/pdf	24158
image/jpeg	14939
image/tiff	9560
text/xml	9560
Text/html	750
p coftsign})	34
video/quicktime	31
application/vnd.rn-realmedia	30
video/x-ms-wmv	27
application/octet-stream	13
application/http	7
image/gif	4
audio/x-realaudio	3
application/x-shockwave-flash	2
pp	2
text/html; charset=iso-8859-1	1
audio/x-ms-wma	1
p sia)	1
application/vnd.ms-excel	1
p})	1
p999aX	1
9878 p	1
Total:	**101138**

Repository 8 - <physicalDescription> Occurrences

Repository 8	
Number of <physicalDescription> elements	**Records**
2	75
3	1
4	15269
5	1640
6	1818
8	6
10	5357
24	1

Note: In all other repositories, <physicalDescription> only occurs once if it is used.

<subject>

<subject> Basic Results

E.29 <subject> Basic Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
1	304	264	462	86.84%	1.75	1	1	5
2	852	852	2371	100.00%	2.78	2	2	13
3	859	859	4596	100.00%	5.35	5	5	22
4	62	62	619	100.00%	9.98	9.5	7	22
5	766	742	1719	96.87%	2.32	2	1	16
6	372	372	958	100.00%	2.58	3	3	27
7	14425	14422	135182	99.98%	9.37	9	8	22
8	24167	10585	28037	43.80%	2.65	2	2	19
9	254072	169937	364313	66.89%	2.14	2	2	62
10	9589	7214	14177	75.23%	1.97	2	1	85
Total:	**305468**	**205309**	**552434**	**67.21%**	**2.69**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <subject> element

D – Total <subject> elements

E – Percentage of records with at least one <subject> element

F – Mean <subject> elements per-record

G – Median

H – Mode

I – Range

Note: 4147 of the records in Repository 8 include propText content for <subject> meaning these are invalid uses of <subject>

<subject> Authority Results

E.30 <subject> Authority Results					
A	**B**	**C**	**D**	**E**	**F**
1	304	264	462	86.84%	100.00%
2	852	852	2286	100.00%	96.42%
3	859	859	4589	100.00%	99.85%
4	62	62	619	100.00%	100.00%
5	766	733	1493	95.69%	86.85%
6	372	372	612	100.00%	63.88%
7	14425	0	0	0.00%	0.00%
8	24167	4147	14025	17.16%	50.02%
9	254072	135143	234787	53.19%	64.45%
10	9589	7199	12652	75.08%	89.24%
Total:	**305468**	**149631**	**271525**	**48.98%**	49.15%

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <subject> element with authority attribute

D – Total <subject> elements with authority attributes

E – Percentage of records with at least one <subject> element with authority attribute

F – Mean <subject> elements with authority attributes per-record

<subject> Subelements Results

E.31 <subject> Subelements Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
1	304	264	462	86.84%	1.75	1	1	5
2	852	852	3262	100.00%	3.83	3	2	34
3	859	859	4703	100.00%	5.47	5	5	40
4	62	62	1070	100.00%	17.26	15.5	13	40
5	766	742	3265	96.87%	4.4	3	2	56
6	372	372	1134	100.00%	3.05	3	3	31
7	14425	14422	135182	99.98%	9.37	9	8	22
8	24167	6480	14012	26.81%	2.16	2	2	8
9	254072	169937	509371	66.89%	3	2	2	112
10	9589	7214	24689	75.23%	3.42	2	1	170
Total:	**305468**	**201204**	**697150**	**65.87%**	**3.46**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <subject> subelement

D – Total <subject> subelements

E – Percentage of records with at least one <subject> subelement

F – Mean <subject> subelements per-record

G – Median

H – Mode

I – Range

<subject> Basic Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.32 <subject> Basic Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Pre-DLF/Aquifer MODS 1.0								
8	23301	9770	26306	41.93%	2.69	2	2	19
9	216980	151827	323138	69.97%	2.13	2	2	26
Total:	**240281**	**161597**	**349444**	**67.25%**	**2.16**			
Post-DLF/Aquifer MODS 1.0								
8	866	815	1731	94.11%	2.12	2	2	19
9	37090	18146	41174	48.92%	2.27	1	1	62
Total:	**37956**	**18961**	**42905**	**49.96%**	**2.26**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <subject> element

D – Total <subject> elements

E – Percentage of records with at least one <subject> element

F – Mean <subject> elements per-record

G – Median

H – Mode

I – Range

<subject> Authority Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.33 <subject> Authority Results					
A	**B**	**C**	**D**	**E**	**F**
Pre-DLF/Aquifer MODS 1.0					
8	23301	4116	13958	17.66%	53.06%
9	216980	125913	214258	58.03%	66.31%
Total:	**240281**	**130029**	**228216**	**54.12%**	65.31%
Post-DLF/Aquifer MODS 1.0					
8	866	31	67	3.58%	3.87%
9	37090	9230	20529	24.89%	49.86%
Total:	**37956**	**9261**	**20596**	**24.40%**	48.00%

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <subject> element with authority attribute

D – Total <subject> elements with authority attributes

E – Percentage of records with at least one <subject> element with authority attribute

F – Mean <subject> elements with authority attributes per-record

<subject> Subelements Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.34 <subject> Subelements Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Pre-DLF/Aquifer MODS 1.0								
8	23301	5696	12348	24.45%	2.17	2	2	8
9	216980	151827	448327	69.97%	2.95	2	2	71
Total:	**240281**	**157523**	**460675**	**65.56%**	**2.92**			
Post-DLF/Aquifer MODS 1.0								
8	866	784	1664	90.53%	2.12	2	2	5
9	37090	18146	61044	48.92%	3.36	1	1	112
Total:	**37956**	**18930**	**62708**	**49.87%**	**3.31**			

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <subject> subelement

D – Total <subject> subelements

E – Percentage of records with at least one <subject> subelement

F – Mean <subject> subelements per-record

G – Median

H – Mode

I – Range

<subject> Subelements

Repository 1	
<subject> Subelement	**Occurrences**
topic	462
Total:	462

Repository 2	
<subject> Subelement	**Occurrences**
topic	2822
geographic	193
name	156
geographicCode	75
titleInfo	15
temporal	1
Total:	3262

Repository 3	
<subject> Subelement	**Occurrences**
genre	25
geographic	446
name	376
temporal	343
topic	3513
Total:	4703

Repository 4	
<subject> Subelement	**Occurrences**
geographic	281
name	74
temporal	89
titleInfo	13
topic	613
Total:	1070

Repository 5	
<subject> Subelement	**Occurrences**
genre	292
geographic	953
geographicCode	175
name	483
temporal	104
titleInfo	5
topic	1253
Total:	2973

Repository 6	
<subject> Subelement	**Occurrences**
geographic	2
hierarchicalGeographic	5
name	20
temporal	446
topic	661
Total:	1134

Repository 7	
<subject> Subelement	**Occurrences**
topic	109208
hierarchicalGeographic	15039
name	6536
geographic	4399
Total:	135182

Repository 8	
<subject> Subelement	**Occurrences**
topic	14012
Total:	14012

Repository 9	
<subject> Subelement	**Occurrences**
topic	275486
hierarchicalGeographic	90969
geographic	69650
temporal	47450
name	13973
geographicCode	7410
cartographics	4411
titleInfo	22
Total:	509371

Repository 10	
<subject> Subelement	**Occurrences**
topic	12963
geographic	6597
name	1853
geographicCode	1382
temporal	976
cartographics	391
genre	374
titleInfo	151
hierarchicalGeographic	2
Total:	24689

Total	
<subject> Subelement	**Occurrences**
topic	420993
geographic	82521
name	23471
temporal	49409
hierarchicalGeographic	106015
geographicCode	9042
genre	691
cartographics	4802
titleInfo	206
Total:	697150

<subject> Authorities

Repository 1	
<subject> Authority	**Occurrences**
lcsh	462
Total:	462

Repository 2	
<subject> Authority	**Occurrences**
lcsh	2208
rvm	60
mesh	18
Total:	2286

Repository 3	
<subject> Authority	**Occurrences**
local	4544
lcsh	45
Total:	6893

Repository 4	
<subject> Authority	**Occurrences**
	74
GNIS	1
LCSH	349
local	195
Total:	619

Repository 5	
<subject> Authority	**Occurrences**
lcsh	1457
lcshac	1
mesh	22
rvm	13
Total:	1493

Repository 6	
<subject> Authority	**Occurrences**
lcsh	612
Total:	2140

Repository 7	
<subject> Authority	**Occurrences**
Total:	0

Repository 8	
<subject> Authority	**Occurrences**
local	14025
Total:	16777

Repository 9	
<subject> Authority	**Occurrences**
lctgm	123810
lcsh	110888
gmgpc	22
lctgm.	16
lctgmlctgm	8
lcgtm	6
lcshac	5
lctgm	3
lctg	3
lctm	3
.lctgm	2
;ctgm	2
csh	2
lctmg	2
lxtgm	1
1930-1940	1
ctgm	1
disgm	1
mesh	1
lctg0	1
lcdtgm	1
lctgma	1
lcltgm	1
lcttgm	1
rvm	1
lctgmou	1
lctgm,	1
lctgmm	1
gmgpclctgm	1
Total:	234787

Repository 10	
<subject> Authority	**Occurrences**
lcsh	12572
mesh	80
Total:	247441

<location>

<location><url> Basic Results

E.35 <location><url> Basic Results										
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**
1	304	304	304	100.00%	1	1	1	0	304	304
2	852	328	437	38.50%	0.51	1	1	1	321	219
3	859	859	866	100.00%	1.01	1	1	1	866	852
4	62	62	62	100.00%	1	1	1	0	61	62
5	766	764	1536	99.74%	2.01	2	2	2	1529	764
6	372	372	372	100.00%	1	1	1	0	372	372
7	14425	14425	14425	100.00%	1	1	1	0	14425	14425
8	24167	24167	36671	100.00%	1.52	2	2	1	36659	11663
9	254072	254069	267819	100.00%	1.05	1	1	11	267781	243891
10	9589	9500	9500	99.07%	0.99	1	1	0	9500	9500
Total:	**305468**	**304850**	**331992**	**99.80%**	**1.09**					**281748**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <location><url> element

D – Total <location><url> elements

E – Percentage of records with at least one <location><url> element

F – Mean <location><url> elements per-record

G – Median

H – Mode

I – Range

J – Number of distinct <url> values

K – Records with a single <location><url>

<location><url> Valid Results

E.36 <location><url> Valid Results				
A	**B**	**C**	**D**	**E**
1	304	304	304	100.00%
2	852	0	0	0.00%
3	859	0	0	0.00%
4	62	62	62	100.00%
5	766	764	764	99.74%
6	372	0	0	0.00%
7	14425	0	0	0.00%
8	24167	0	0	0.00%
9	254072	0	0	0.00%
10	9589	0	0	0.00%
Total	**305468**	**1130**	**1130**	**0.37%**

A – Repository ID

B – Records in test set

C – Distinct records with valid <location><url>

D – Total valid <url> elements

E – Percentage of records with valid <location><url>

Note: When records include a valid <location><url usage="primary display"> element, there are one and only one per record

<location><url> Basic Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines

E.37 <location><url> Basic Results										
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**
Pre-DLF/Aquifer MODS 1.0										
8	23301	23301	35788	100.00%	1.54	2	2	1	35776	10814
9	216980	216979	228824	100.00%	1.05	1	1	11	228792	208398
Total:	**240281**	**240280**	**264612**	**100.00%**	**1.1**					**219212**
Post-DLF/Aquifer MODS 1.0										
8	866	866	883	100.00%	1.02	1	1	1	883	849
9	37090	37090	38995	100.00%	1.05	1	1	11	38989	35493
Total:	**37956**	**37956**	**39878**	**100.00%**	**1.05**					**36342**

A – Repository ID

B – Records in test set

C – Distinct records containing at least one <location><url> element

D – Total <location><url> elements

E – Percentage of records with at least one <location><url> element

F – Mean <location><url> elements per-record

G – Median

H – Mode

I – Range

J – Number of distinct <url> values

K – Records with a single <location><url>

<location><url> Valid Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.38 <location><url> Valid Results				
A	**B**	**C**	**D**	**E**
Pre-DLF/Aquifer MODS 1.0				
8	23301	0	0	0.00%
9	216980	0	0	0.00%
Total:	**240281**	**0**	**0**	0.00%
Post-DLF/Aquifer MODS 1.0				
8	866	0	0	0.00%
9	37090	0	0	0.00%
Total:	**37956**	**0**	**0**	**0.00%**

A – Repository ID

B – Records in test set

C – Distinct records with valid <location><url>

D – Total valid <url> elements

E – Percentage of records with valid <location><url>

Note: When records include a valid <location><url usage="primary display"> element, there are one and only one per record

<accessCondition>

<accessCondition> Basic Results

E.39 <accessCondition> Basic Results					
A	**B**	**C**	**D**	**E**	**F**
1	304	304	304	304	100.00%
2	852	2	2	4	0.23%
3	859	859	859	859	100.00%
4	62	62	62	62	100.00%
5	766	0	0	0	0.00%
6	372	0	0	1	0.00%
7	14425	14425	14425	14425	100.00%
8	24167	0	0	0	0.00%
9	254072	163406	163501	169316	64.31%
10	9589	0	0	9589	0.00%
Total:	**305468**	**179058**	**179153**	**194560**	**58.62%**

A – Repository ID

B – Records in test set

C – Distinct <accessCondition> elements with type=”useandreproduction”

D – Total <accessCondition> elements with type=”useandreproduction”

E – Total <accessCondition> elements

F – Percentage of records with valid <accessCondition> element

Note: Repository 10 is only missing the required attribute/value pair

<accessCondition> Basic Results - Pre & Post DLF/Aquifer MODS Implementation Guidelines 1.0

E.40 <accessCondition> Basic Results					
A	**B**	**C**	**D**	**E**	**F**
1					
8	23301	0	0	0	0.00%
9	216980	127444	127537	132048	58.74%
Total	**240281**	**127444**	**127537**	**132048**	**53.04%**
6					
8	866	0	0	0	0.00%
9	37090	35962	35964	37268	96.96%
Total	**37956**	**35962**	**35964**	**37268**	**94.75%**

A – Repository ID

B – Records in test set

C – Distinct <accessCondition> elements with type=”useandreproduction”

D – Total <accessCondition> elements with type=”useandreproduction”

E – Total <accessCondition> elements

F – Percentage of records with valid <accessCondition> element

Note: Repository 10 is only missing the required attribute/value pair

<accessCondition> Content

Repo ID	E.41 <accessCondition> Content	Occurrences
1	Personal, noncommercial use of this item is permitted in the Unted States of America. Please see http://digital.library.upenn.edu/women/ for other rights and restrictions that may apply to this resource.	303
	The editor of The Online Books Page believes that free access to this item for personal, noncommercial use is permitted in the United States of America, and in the country of the site providing this item. Other use and reproduction rights may also apply,	1
2	Includes bibliographical references (p. 45).	1
	Includes bibliographical references (p. 24).	1
3	For rights relating to this resource, visit: http://idserver.utk.edu/?id=200500000001941	465
	Digital Image Copyright (c) 2005. The University of Tennessee Libraries. All Rights Reserved. For permission to use, contact: Great Smoky Mountains Regional Project, The University of Tennessee Libraries, Knoxville, TN. For current rights information, vis	387
	For current rights information, please visit: http://idserver.utk.edu/?id=200600000001198	7
4	Copyright for contributions published in the Southern Spaces forum is retained by the authors, with publication rights granted to the forum. Content is free to users. Any reproduction of original content from Southern Spaces must a) seek copyright fr	62
7	Copyright and reproduction rights for all Charles W. Cushman photographs are held by Indiana University and administered by the University Archives, Indiana University, Bloomington, IN 47405.	14425
9	No known restrictions on publication.	139211
	Publication may be restricted. For information see "Horydczak Collection" (http://lcweb.loc.gov/rr/print/res/100_hory.html)	14322
	No known restrictions on reproduction.	4217
	For publication information see "Carl Van Vechten Photographs (Lots 12735 and 12736)"	1397
	No known restrictions on publication. For information see "World War I Posters" (http://lcweb.loc.gov/rr/print/res/243_wwipos.html)	998
	No known restrictions on publication. No renewal found in Copyright Office.	679

	Publication may be restricted. For information see "Brumfield Photograph Collection" (http://lcweb.loc.gov/rr/print/res/273_brum.html)	618
	No known restrictions on publication. No renewal in Copyright office.	538
	Publication may be restricted. For information see "Brumfield Photograph Collection" (http://lcweb.loc.gov/rr/print/res/273_brum.html).	518
	Rights status not evaluated. For general information see "Copyright and Other Restrictions..." (http://lcweb.loc.gov/rr/print/195_copr.html).	380
	No known restrictions.	303
	No known restrictions on publication. For information see "World War I Posters,"	109
	No known restrictions on publication. No copyright renewal found.	39
	Rights status not evaluated. For general information see "Copyright and Other Restrictions ...,"	37
	Rights status not evaluated. For general information, see "Copyright and Other Restrictions ..." (http://lcweb.loc.gov/rr/print/195_copr.html)	33
	Publication may be restricted. For information see "G. Eric and Edith Matson Photograph Collection,"	27
	Rights status not evaluated. For general information see "Copyright and Other Restrictions..." (http://lcweb.loc.gov/rr/print/195_copr.html)	9
	No known restrictions on publication. No copyright registration found.	5
	Publication may be restricted. For information see "Civil War Photographs ..." (http://lcweb.loc.gov/rr/print/res/120_cwar.html)	5
	No known restrictions on publication. No renewal in copyright office.	5
	No known restrictions on publication in the U.S. Use elsewhere may be restricted by other countries' laws. For general information see "Copyright and Other Restrictions..." (http://lcweb.loc.gov/rr/print/195_copr.html)	4
	No known restrictions on publication. DLC	4
	No known restrictions on publication	2
	No copyright renewal.	2
	Rights status not evaluated. For general information see "Copyright and Other Restrictions...,"	2
	Publication may be restricted. For information see "National Photo	2

	Company Collection" (http://lcweb.loc.gov/rr/print/res/275_npco.html)	
	No known restrictions on publicaton.	2
	Publication may be restricted. For information see "Edward Weston rights and restrictions information,"	2
	Publication may be restricted. For information see "Laura Voldkerding rights and restrictions information,"	2
	Publication may be restricted. For information see "G. Eric and Edith Matson Collection,"	2
	This record contains unverified, old data from caption card.	2
	Publication may be restricted. For general information, see "Copyright and Other Restrictions ...," (Reproduced in American Memory with permission from Oscar Bailey, 1997).	2
	Publication may be restricted. For general information, see "Copyright and Other Restrictions ...," (Reproduced in American Memory with permission from James Ivey, 1997).	2
	No known restrictions on publication in the U.S. Use elsewhere may be restricted by other countries' laws.	1
	"If used for publication, credit must be given to the Smithsonian Institution, Bureau of American Ethnology."	1
	RNo known restrictions on publication.	1
	George Grantham Bain Collection.	1
	Horse Capture is great-grandfather of museum director and historian George P. Horse Capture.	1
	No copyright renewal per LC Photoduplication Service.	1
	No known restrctions on publication.	1
	No known restriction on publication.	1
	Publication may be restricted. For information see "G. Eric Matson Photograph Collection,"	1
	Reproduced with the permission of the Braun Research Library, Southwest Museum, Los Angeles, California.	1
	No known restrictions on publication. No renewal in Copyright Office	1
	No known restrictions on publiction.	1
	No restrictions on publication.	1

	Restricted: Information on reproduction rights available at Reference Counter.	1
	Rights status not evaluated. For general information, see "Copyright and Other Restrictions ..."	1
	Rights status not evaluated. For general information see "Copyright and Other Restrictions . . . "	1
	Publication may be restricted. For general information, see "Copyright and Other Restrictions ...," (Reproduced in American Memory with permission from Barbara Crane, 1997).	1
	Publication may be restricted. For general information, see "Copyright and Other Restrictions ...," (Reproduced in American Memory with permission from Michael Smith, 1997).	1
	No known; restrictions on publication.	1
	No known restrictions on publicat	1
10	Where applicable, subject to copyright. Other restrictions on distribution may apply. Please go to http://www.umdl.umich.edu/ for more information.	9589

<recordInfo>

<recordInfo> Full Results

E.42 <recordInfo><languageOfCataloging><languageTerm> Basic Results								
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
1	304	304	304	304	304	304	100.00%	100.00%
2	852	0	0	0	0	852	100.00%	0.00%
3	859	0	859	0	859	859	100.00%	0.00%
4	62	0	0	0	62	62	100.00%	0.00%
5	766	0	0	0	0	766	100.00%	0.00%
6	372	0	0	0	0	372	100.00%	0.00%
7	14425	0	0	0	0	14425	100.00%	0.00%
8	24167	0	0	0	0	0	0.00%	0.00%
9	254072	0	0	0	0	254070	100.00%	0.00%
10	9589	0	50	0	9639	9589	100.00%	0.00%
Total	**305468**	**304**	**1213**	**304**	**10864**	**281299**	**92.09%**	**0.10%**

A – Repository ID

B – Records in test set

C – Distinct records containing required <recordInfo> element and required subelements and attribute/value pairs

D – Records with required <languageTerm type="code" authority="iso639-2b"> elements

E – Records with reguired <languageTerm type="text"> elements

F – Total <languageOfCataloging> elements

G – Total <recordInfo> elements

H – Percentage of records with at least one <recordInfo> element

I – Percentage of records that meet all <recordInfo> Requirements

<recordInfo><laguageTerm type="text"> Values

E.43 <laguageTerm type="text"> values		
Repo ID	**Value**	**Occurrences**
1	English	304

<recordInfo><laguageTerm type="code" authority="iso639-2b"> Values

E.44 <laguageTerm type="code" authority="iso639-2b"> values		
Repo ID	**Value**	**Occurrences**
1	eng	304
3	eng	859
10	eng	50