Improving Metadata Quality: Augmentation and Recombination

Diane Hillmann, Naomi Dushay, Jon Phipps
National Science Digital Library, Cornell University,
301 College Ave., Ithaca, NY 14850, USA
EMail: {hillmann, naomi, jphipps} @cs.cornell.edu

Abstract: Digital libraries have, in the main, adopted the traditional library notion of the metadata "record" as the basic unit of management and exchange. Although this simplifies the harvest and re-exposure of metadata, it limits the ability of metadata aggregators to improve the quality of metadata and to share specifics of those improvements with others. The National Science Digital Library (NSDL) is exploring options for augmenting harvested metadata and re-exposing the augmented metadata to downstream users with detailed information on how it was created and by whom. The key to this augmentation process involves changing the basic metadata unit from "record" to "statement."

1 Introduction

Metadata today is likely to be created by people without any metadata training, working largely in isolation and without adequate documentation. Metadata records are also created by automated means, often with poorly documented methodology and little or no indication of provenance. Unsurprisingly, the metadata resulting from these processes varies strikingly in quality and often does not play well together. Nevertheless, many metadata aggregators use this metadata to build services for end users, thus contributing to criticisms that metadata is of limited value, can't be trusted or that it's demonstrably so incomplete as to be worthless.

Crawling full text resources is largely believed to have taken the place of topic assignment for digitally available resources, but when resources can't be crawled, perhaps due to intellectual property issues or because the format isn't text, the resource's metadata becomes even more important. Moreover, resource crawling has its limitations. Even when a crawled resource can be automatically described effectively, optimal search and discovery metadata does not depend solely on information in the resource itself. One of the distinguishing characteristics of Google's popular search engine is its effective use of aggregated metadata to identify desirable matches to a search query. Google's PageRank relies heavily on information obtained by harvesting, following, and indexing the contents of hyperlinks [1]. In addition to analyzing link topology, Google applies a weighted trust metric to the source of those links, e.g. giving a higher ranking to resources referenced by links in the .edu and .org domains, as well as many other factors. Google's Page-Rank algorithm also appears to give greater weight to terms contained in the <title> metadata tag than to terms in the body of the resource itself [2].

2 The NSDL Environment

The NSDL is a wide-ranging program of the National Science Foundation, engaged in building library collections and services for all aspects of science education [3]. Now in its third year of operation, the NSDL is building upon the technical foundation already established (described more fully in [4], [5]). The NSDL Metadata Repository (MR) is fully operational, gathering and updating increasing amounts of metadata pertaining to resources in the fields of science, technology, engineering and mathematics. The MR, based on Qualified Dublin Core, uses a simple two-tier model comprised of "collections" and their "items." An item may be large or small, and it may itself contain parts or smaller units; a collection is defined as an organized arrangement of items. Associating every individual item with a collection, though fairly primitive as an organizing principle, allows some basic assertions of quality based on the reputation of the entity responsible for the collection. This simple collection/item principle has also facilitated the construction of an automated "ingestion" system, based on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [6], whereby metadata flows into the MR with a minimum of ongoing human intervention. The NSDL, from this perspective, functions essentially as a metadata aggregator.

While the MR and its related systems currently ingest metadata primarily from services providing Dublin Core (DC) metadata via the OAI-PMH, the system architecture has been

https://doi.org/10.23106/dcmi.952107705

designed to harvest, store, and redistribute a wide range of heterogeneous metadata from disparate sources. In order to leverage maturity, scalability and performance of the relational data model [7], [8], particularly in the context of the non-hierarchical structure of DC metadata, the ingest process shreds harvested XML metadata into a set of related tables: metadata providers serve *collections* of *records*, each of which is comprised of individual *elements* defined by XML *schemas*. Incoming records are split up into elements for storage in the MR, and are reassembled into records for MR output. This design allows us to perform efficient, detailed, element-level data analysis across many providers' collections to evaluate both aggregated metadata content and overall metadata quality, while allowing the MR to easily maintain and update metadata from each provider.

In addition, the resource identifier in each record, generally a URL in a DC Identifier element, is normalized and indexed to provide a simple resource equivalence service, allowing identification and lookup of records from multiple metadata providers that may be describing the same resource. "Duplicate" metadata statements about a particular resource, anathema in a traditional catalog, can be seen in our context as potentially different assertions about a resource. Provider A may see the resource as an earth science resource, because that reflects the area of interest of A's community; provider B may be more focused on astronomy, and describe the resource from a different point of view. If an equivalence relationship can be established between these two metadata descriptions, a broader view of the resource is then available to others.

3 Transforming Metadata "Safely"

The NSDL architecture is based on the recognition that, with a library of this complexity, it is impossible to impose detailed requirements for metadata standards that every collection must follow. Instead, the NSDL must accommodate a broad spectrum of metadata quality, anticipating a wide variety of errors or inconsistencies.

Reality has fully justified these minimal expectations. Dushay and Hillmann [9] identify four categories of problems encountered with the metadata harvested by the NSDL:

1. missing data – metadata elements not present in supplied metadata
2. incorrect data – metadata values not conforming to standard element use
3. confusing data – multiple values crammed into a single metadata element, embedded html tags, etc.
4. insufficient data – e.g., no indication of controlled vocabularies used

In attempting to provide a reasonable level of quality and predictability for the metadata served from the MR, in particular to provide services to end users at our public portal at NSDL.org, we wanted to correct as many problems with the harvested metadata as possible, in a scaleable fashion. Initially, we processed and corrected metadata on a collection-by-collection basis. Our corrections were individually tailored to each collection, with the expectation that a collection's metadata would always require the same transform. We quickly discovered that collections' metadata practices changed over time, implying an ongoing cost of tailoring transforms for individual harvests. However, with the experience we gained, we were able to address some of the common metadata quality problems with an automated technique we refer to as "safe transforms." We use the word "transform" both because it changes the metadata and because we use eXtensible Stylesheet Language: Transformations (XSLT) to modify the XML metadata as supplied by the provider in order to produce a normalized, "smartened-up"[1] version of the XML metadata for storage in the MR.

Safe transforms are designed to enhance the information present in the original metadata with no risk of degradation. The goal is to improve the utility of the metadata for the NSDL in the following ways:

1. *remove "noise"* – a partial solution to the "incorrect data" problem. For example, we remove metadata with no information value, such as empty metadata elements, metadata elements with values such as "unknown" or "n/a" or consisting entirely of dashes or other punctuation.

[1] Used colloquially by the NSDL to indicate the opposite of "dumbing-down."

https://doi.org/10.23106/dcmi.952107705

2. *detect and identify controlled vocabularies in use whenever possible* – a partial solution to the "insufficient data" problem. For example, the DCMIType encoding scheme is applied to DC "Type" elements when their value is one of the allowed DCMITypes [10]. This works well for small controlled vocabularies; however, it does not scale well to large vocabularies such as LCSH.
3. *normalize metadata presentation* – clean up the values: remove double XML encodings ("<" becomes "<"), extra whitespace (a tab followed by five spaces becomes a single space), etc.

Because our intent is to normalize the metadata exposed by the MR, we apply safe transforms to every metadata record harvested by and ingested into the Metadata Repository as a minimal means of quality assurance. While this does improve the quality of metadata served by the MR, it just scratches the surface of a more endemic problem.

Some missing data and confusing data problems cannot be fixed with safe transforms applied to every ingested metadata record, but can be tackled on a collection-by-collection basis. For example, it is not unusual for a collection to consist solely of resources of a single format or a single type and for the collection's metadata to lack any type or format information, as the information is assumed in the context of the originating collection [11]. We also encounter multiple values in a single metadata element, such as

<dc:contributor>Sanders, G.S., T.R. Brice, V.L. DeSantis, Jr., and C.C. Ryder.</dc:contributor>
<dc:creator>Van Gogh, Vincent, George Jackson, Humphrey Little and Stanley Black</dc:creator>

Unfortunately, the separator used, and the exact formatting of the names differs from collection to collection, and sometimes within the same collection. A comma may be an automatically detectable separator for the DC Creator fields in collection A, while collection B uses commas entirely differently in its Creator fields, and collection C uses a semicolon as a separator. Thus, the safe transform approach is necessarily conservative; otherwise, it would improve some collections' metadata at the cost of degrading that of other collection [12]. This has made it necessary to continue to specify collection-specific transforms as well as safe transforms, to catch these additional problems that occur only within particular collections and that cannot be generalized.

4 Replacing Safe Transforms with Metadata Augmentation

As mentioned earlier, our implementation of safe transforms uses XSLT. The initial goal was to provide a low cost, scaleable way to improve the quality and predictability of metadata in the NSDL MR. XSLT kept programming costs down, but there were some trade offs. Because it is difficult and unwieldy to write XSL stylesheets that track exactly which metadata elements are changed during the safe transform process, it is necessary to perform the safe transforms every time data is harvested or updated, because any transformed data is over-written by each new harvest. This problem occurs with collection-specific transformations as well as with the safe transforms. In addition, while XML and XSL tools have allowed automation, the immaturity of these tools makes it difficult, if not impossible, to automatically detect errors in source metadata and to automatically supply useful explanations of errors to metadata providers.

Another issue is the opacity of our normalized data: there is no indication in metadata served from the MR of changes we have made. Although we certainly believe our changes are improvements, we provide no means for partners or downstream users to determine or evaluate whether or not they agree. Our normalized version of Qualified Dublin Core was not designed to expose detailed information on changes at that level of specificity. Similarly, OAI-PMH guidelines for provenance information [13] only allow a true/false assertion that changes have been made in a record. Our methodology therefore requires that we flag all normalized records "true" for change, even if our transformations didn't touch them, because we have no good way to track if and where a change has been made.

As we've continued to consider strategies for improving metadata quality, we've also reconsidered our view of portals or services as not just consumers of metadata but potential providers of metadata as well. In a simple case, a metadata provider may have supplied unqualified DC Format information describing a resource as an *image*. A service, such as the NSDL archive service, that routinely crawls

https://doi.org/10.23106/dcmi.952107705

resources might provide a more specific Format description of *image/png*, based on the Internet Mime Type (IMT) list recommended by DCMI [14]. Coming from a trusted service that uses known methods for detecting the Format of the resource, this assertion might be inherently more trustworthy than even the resource provider's own metadata.

Another example illustrates a slightly more complex scenario. The Eisenhower National Clearinghouse (ENC) was an early provider of metadata to the NSDL MR; this year it will begin harvesting aggregated metadata from the MR in order to create a portal targeted for middle school teachers and students. ENC is planning to enhance this MR-supplied metadata with audience and education level information so they can provide services to their users that build on the enhanced metadata. As part of this effort they will also provide information on a resource's relevance to national and state educational standards. ENC will then expose this new information for harvest by the NSDL MR via OAI-PMH.

We've also observed that collections providing metadata to the NSDL have been slow to use standard controlled vocabularies, and they often do not expose them in a way that promotes automated interoperability. To be sure, much of the difficulty is due to the lack of understanding and infrastructure available to expose vocabularies interoperably.[2]. The NSDL is currently working with a group from the INFOMINE Project [15] on methods to add topical information from controlled vocabularies and classifications to metadata already existing in the NSDL MR, using automated crawling and metadata generation technologies.

5 From Records to Elements

Several interesting conclusions began to emerge from these explorations into metadata enhancement. First, managing and exposing metadata to services solely as discrete records, though clearly necessary with the use of OAI-PMH, limits opportunities for automated metadata maintenance and enhancement. Tom Baker, in his article on "A Grammar of Dublin Core" [16] calls Dublin Core "… a small language for making a particular class of ***statements about resources***." Stephen Downes has developed a notion of a Resource Profile, which he describes as "… a multi-faceted, wide ranging description of a resource [that] conforms to no particular XML schema, nor is it authored by any particular author." [17].

If a metadata record can be seen as a series of statements about resources, then it should be possible to manage the metadata at the statement or *element* level, rather than the *record* level. Aggregating both complete and fragmentary metadata from many sources provides the opportunity to build a more complete profile of a resource. As we began considering the possibilities of an augmentation strategy for NSDL, Downes' ideas about the potential of Resource Profiles began to resonate.

Although at first glance, shifting the granularity of the managed metadata unit from *record* to *element* seems to add unnecessary complexity, in reality it enables several important new possibilities. First, it allows us to expose the source of each statement, and to reassemble these statements about resources in a variety of ways and for a variety of purposes, rather than expose a mélange of records and expect the downstream users to perform complex dissociations and recombinations. In addition, any information associated with a particular provider—including methodology used, reputation or rating of provider, and age of the metadata—could be linked to each individual statement. Extending this thinking, it became clear that we needed to consider any augmentation to a metadata record or its statements, including our own safe transforms, as separate statements that we can combine at certain points for distribution to downstream services but manage separately.

Figure 1 illustrates how these several providers might contribute to an augmented metadata record. The MR harvests a simple DC record from Provider A containing unqualified Title, Identifier, Creator and Type elements. The safe transforms done at the MR recognize the Identifier as a valid URI, so the URI encoding scheme is added to the Identifier element, and similarly, the Type value is a valid DCMIType, so the DCMIType encoding scheme is added to the Type element. ENC is able to provide audience and education level information for the resource, while the INFOMINE iVia service provides subject information in three different encoding

[2] The Dublin Core Metadata Initiative (DCMI) has been attempting for over a year to create a registry to identify subject vocabularies used in metadata, but the effort is now on hold due largely to a lack of resources.

https://doi.org/10.23106/dcmi.952107705

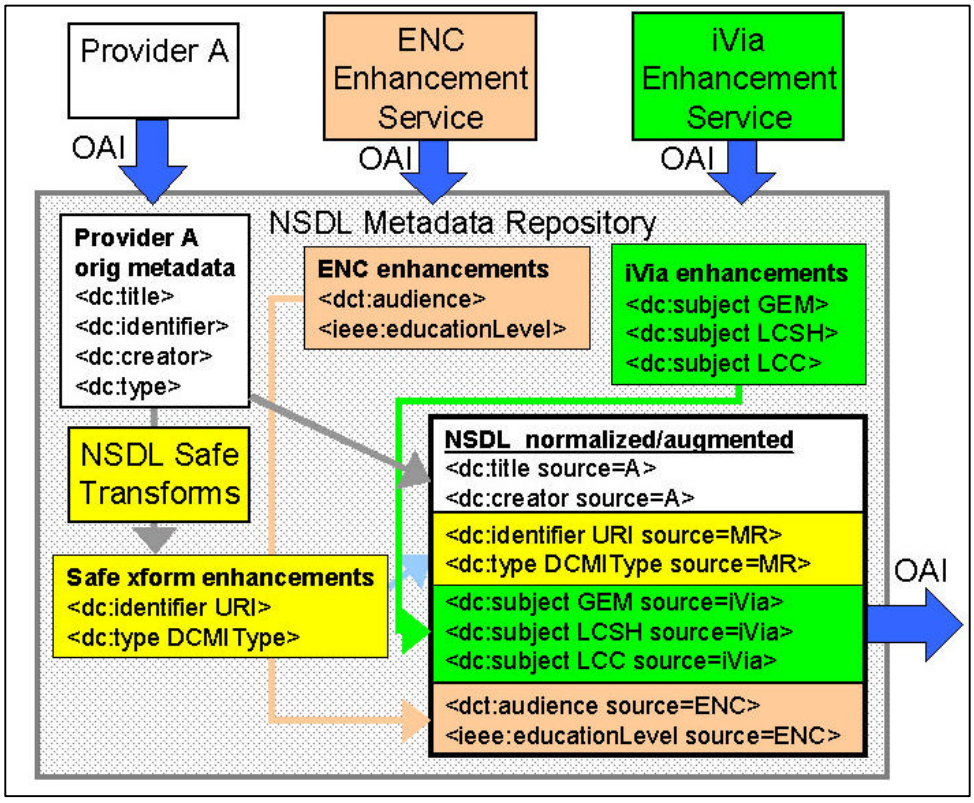

Figure 1– Sources, storage and redistribution of augmented metadata in the MR

schemes. These enhancements are exposed via OAI-PMH, and the MR harvests them. The MR associates the original metadata from provider A, the safe transform enhancements, the ENC enhancements and the iVia enhancements, and it aggregates this information into an augmented, normalized metadata record, which is then exposed by the MR's OAI server.

6 Exposing Quality Information

One of the more significant challenges presented by this model is that of establishing the relative and intrinsic value of each metadata statement—a value that might vary based on intended use. How much can a particular DC Identifier be trusted? What is known about the provider of a particular DC Description? In an environment where the norm is complete records from one source being passed through an aggregation service without change, the issue rarely arises; when "recombinant" metadata is served out, the provenance of the metadata is very much an issue.

One reason these questions arise is because of the inherent subjectivity of some metadata assertions. Metadata such as IEEE-LOM:interactivity adheres to no universally accepted objective metric or ontology. Each metadata provider may be using different criteria to assign levels of interactivity. The publisher of the resource may, for marketing reasons, assert that all of its resources are 'high' interactivity regardless of the actual level of interactivity. A fifth-grade teacher who has used the resource might assert that its interactivity level is 'low'. Without access to the criteria each uses to assign an interactivity level, the usefulness of the assignments is limited.

Quality issues for automatically generated metadata are oftentimes different than those for human generated metadata. As Bruce and Hillmann point out [18], "… a computer program that extracts metadata will produce absolutely consistent results over an indefinite period of time, where a churning pool of student employees assigned to a markup project will not." Bruce and Hillmann maintain that, given this

https://doi.org/10.23106/dcmi.952107705

understanding, exposing "… provenance information at a more detailed level, including (in addition to source, date and identifier) information on the methodology used in the creation of the metadata …" is not only a way to identify quality information *about* metadata, but comprises an indication of metadata quality by its very presence.

7 Technical and Schema Issues

To date, metadata schemas have rarely provided a facility for indicating provenance of metadata, particularly not at the metadata statement level. In order to accomplish this goal, particularly in the NSDL context, we must have automated mechanisms to indicate candidate metadata records available to augmentation services

- to ingest metadata augmentations from the augmentation services into the MR
- to match augmentations to individual metadata records and/or the resource URLs to which the augmentation statements apply
- to store metadata so that the source of the assertion for each metadata element, and the last update date for the element (a measure of freshness) can be easily determined

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/">
  <responseDate>2004-04-27T20:30:07Z</responseDate>
  <request identifier="oai:nsdl.org:providerA:sample1" metadataPrefix="nsdl_augmented"
verb="GetRecord">http://nsdl.baseurl/</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:nsdl.org:providerA:sample1</identifier>
        <datestamp>2004-04-08T15:19:15Z</datestamp>
      </header>
      <metadata>
        <nsdl_augmented xmlns="http://ns.nsdl.org/nsdl_augmented/"
            xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dct="http://purl.org/dc/terms/">
          <dc:title source="providerA" datestamp="2003-09-01T01:01:01Z">Recombined Metadata Sample Record</dc:title>
          <dc:creator source="providerA" datestamp="2003-09-01T01:01:01Z">Naomi Dushay</dc:creator>
          <dc:identifier xsi:type="dct:URI" source="NSDL" datestamp="2003-09-
13T13:13:13Z">http://some.url.com/sample1</dc:identifier>
          <dc:type xsi:type="dct:DCMIType" source="NSDL" datestamp="2003-09-13T13:13:13Z">Text</dc:type>
          <dc:subject xsi:type="GEM" source="iVia" datestamp="2003-09-14T22:22:22Z">Computer science</dc:subject>
          <dc:subject xsi:type="dct:LCSH" source="iVia" datestamp="2003-09-14T22:22:22Z">Computer simulation</dc:subject>
          <dc:subject xsi:type="dct:LCC" source="iVia" datestamp="2003-09-14T22:22:22Z">QA76.9.C65</dc:subject>
          <dct:audience source="ENC" datestamp="2004-03-13T04:04:04Z">undergraduate</dct:audience>
          <ieee:interactivityLevel source="ENC" datestamp="2004-03-13T04:04:04Z">high</ieee:interactivityLevel>
        </nsdl_augmented>
      </metadata>
      <about>
        <provenance xmlns="http://nsdl.provenance/">
          <originDescription harvestDate="2003-08-22T11:12:13Z" altered="true">
            <baseURL>http://providerA.baseURL/</baseURL>
            <origOAI_identifier>sample1</origOAI_identifier>
            <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
          </originDescription>
          <sourceDescription source="NSDL" moreInfo="http://nsdl.org/safetransforms/"/>
          <sourceDescription source="iVia" moreInfo="http://nsdl.org/iVia_enhancements/"/>
          <sourceDescription source="ENC" moreInfo="http://nsdl.org/ENC_enhancements/"/>
        </provenance>
      </about>
    </record>
  </GetRecord>
</OAI-PMH>
```

Figure 2-- An augmented metadata record in a preliminary format

https://doi.org/10.23106/dcmi.952107705

- to indicate for downstream users the source of the assertion of each metadata element in a metadata record exposed by the MR, and where additional information about the source can be found.

The NSDL is in the process of designing this infrastructure, which will include XML formats expressed as XML schemas, among other pieces. Initially we will be using the OAI-PMH for both harvest and exposure of metadata, at the MR and at augmentation services, although we recognize that at some stage we will need a wider range of publish/subscribe methodologies.

The sample record (which corresponds to the diagram shown earlier) illustrates our preliminary work in this area. Of particular note is the extension to the OAI About container including a link to additional information about the service providers represented within the record itself.

8 Conclusions

The utility of metadata can best be evaluated in the context of services provided to end-users. Different services require different kinds of metadata, perhaps tailored for different purposes, or with different confidence ratings. In our view, metadata tailoring, recombination and repurposing will require metadata aggregators and others to think of elements, rather than records, as the basic metadata unit. Aggregators, in the middle between metadata creators and service providers, will need to track information such as source, date, and creation methodology for metadata statements in order to enable quality assessments by downstream consumers. These assessments, in the context of real services, may well provide the best answer to the recurring questions of metadata utility.

Clearly, as the NSDL integrates human- and machine-generated metadata using scalable methods, quality issues will need to be addressed more directly. We are already approaching the point where most of the relevant metadata stores exposed via OAI servers are already included in the MR, and to continue to grow, other methods of gathering materials will need to be incorporated. As we develop our augmentation ingest processes and work with partners on the specific projects described earlier, we expect machine-generated metadata to increase in importance to the NSDL, and our augmentation strategies must reflect this reality.

The effort reported here to re-orient attention from metadata records to metadata statements is paralleled in some respects by the movement away from building and using specific metadata schemas towards the use of application profiles. Described by Heery and Patel in their seminal article in 2000 [19] as the “mixing and matching of metadata schemas,” application profiles allow communities to document use of metadata for particular domains at the element level, rather than the schema level. The CEN Workshop Agreement CWA14855 [20], created in cooperation with the DCMI, provides specific guidelines on creating application profiles, including the important documentation of the use constraints imposed by domains who may be using metadata elements in ways that downstream users must know about to correctly interpret and use the information.

As advances in technology have enabled many useful tools and services for discovery of resources, public perception may continue to doubt the utility of bibliographic metadata. What we think of as “metadata” will no doubt evolve as well. Outside of libraries, non-traditional sources of information such as usage statistics are increasingly mined for information discovery purposes (amazon.com: “others who bought this title also bought …”). Non-text-based resources such as images can now provide new types of metadata such as thumbnail images, XMP, Exif, and IPTC metadata. Nevertheless, as long as resources exist in a context with no one source knowing everything about them, bibliographic metadata is likely to remain an important means for discovering and using these resources.

9 Acknowledgements

This work was funded by the National Science Foundation under grant 0227648. The ideas in this paper are those of the authors and not of the National Science Foundation.

The authors would like to thank Anat Nidar-Levi for her help in preparing this paper for publication, the NSDL Cornell team (especially Tim Cornwell and Elly Cramer) and all the people and organizations who have provided metadata to or harvested metadata from the NSDL.

https://doi.org/10.23106/dcmi.952107705

References

1. Sergey Brin and L. Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. 1998. http://www-db.stanford.edu/pub/papers/google.pdf
2. *Google Top Level Domains*. May 12, 2004. http://www.searchengineworld.com/google/domains.htm
3. Lee Zia, et al. *The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program*, in *D-Lib*. 2001. http://www.dlib.org/dlib/march04/zia/03zia.html
4. Carl Lagoze, et al. *Core services in the architecture of the national science digital library (NSDL).* in *JCDL 2002*. 2002. Portland, OR. http://arxiv.org/ftp/cs/papers/0201/0201025.pdf
5. William Y. Arms and et al, *A Spectrum of Interoperability The Site for Science Prototype for the NSDL*. D-Lib Magazine, 2002. **8**(1). http://www.dlib.org/dlib/january02/arms/01arms.html
6. *The Open Archives Initiative Protocol for Metadata Harvesting*. May 12, 2004. http://www.openarchives.org/OAI/openarchivesprotocol.html
7. Jayavel Shanmugasundaram, et al. *Relational Databases for Querying XML Documents: Limitations and Opportunities*. in *Proceedings of the 25th VLDB Conference*. 1999. Edinburgh, Scotland. http://www.vldb.org/conf/1999/P31.pdf
8. Latifur Khan and Y. Rao. *A performance Evaluation of Storing XML data in relational database management systems*. in *Workshop On Web Information And Data Management - Proceeding of the third international workshop on Web information and data management*. 2001. Atlanta, GA. http://portal.acm.org/citation.cfm?id=502939&dl=ACM&coll=portal
9. Naomi Dushay and D. Hillmann. *Analyzing metadata for effective use and re-use*. in *Dublin Core 2003*. 2003. Seattle, Washington. http://dc2003.ischool.washington.edu/Archive-03/03dushay.pdf
10. *The DCMI Type Vocabulary*. May 12, 2004. http://dublincore.org/documents/dcmi-terms/#H5
11. Wendler, R., "The Eye of the Beholder." In Bruce et al., *Metadata In Practice*, 64. D.H.a.E. Westbrooks, Editors. 2004. Robin Wendler describes this as the " … 'on a horse' problem, in honor of a local Teddy Roosevelt database in which all subject headings are presumed to be about the great man. If you encounter the subject 'on a horse' in this context, you can safely assume that the Bull Moose himself is riding." http://content.nsdl.org/metadata_practice/hillmann_bruce_final.html
12. Coyle, K., "Crosswalking Citation Metadata." In Bruce et al., *Metadata in Practice*, 92-94. D.H.a.E. Westbrooks, Editors. 2004. p.http://content.nsdl.org/metadata_practice/hillmann_bruce_final.html
13. *Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting: XML schema to hold provenance information in the "about" part of a record*. May 12, 2004. http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm
14. *DCMI Metadata Terms: Encoding Schemes: IMT*. May 12, 2004. http://dublincore.org/documents/dcmi-terms/#IMT
15. Steve Mitchell, M.M., Julie Mason, Gordon W. Paynter, Johannes Ruscheinski, Artur Kedzierski, Keith Humphreys, *iVia Open Source Virtual Library System.* D-Lib Magazine, 2003. **9**(1). http://www.dlib.org/dlib/january03/mitchell/01mitchell.html
16. Baker, T., *A Grammar of Dublin Core.* D-Lib Magazine, 2000. **6**(10). http://www.dlib.org/dlib/october00/baker/10baker.html
17. Downes, S., *Resource Profiles*. November, 2003. http://www.downes.ca/files/resource_profiles.htm
18. Thomas R. Bruce, Diane Hillmann, "The Continuum of Metadata Quality: Defining, Expressing, Exploiting." In *Metadata In Practice*, 248-249. D.H.a.E. Westbrooks, Editors. 2004. http://content.nsdl.org/metadata_practice/hillmann_bruce_final.html
19. Rachel Heery and M. Patel, *Application profiles: mixing and matching metadata schemas.* http://www.ariadne.ac.uk/issue25/app-profiles/
20. *CEN Workshop Agreement CWA 14855: Dublin Core Application Profile Guidelines*, CEN. ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14855-00-2003-Nov.pdf

https://doi.org/10.23106/dcmi.952107705