

# 1 DLF AIG METADATA WORKING GROUP

---

This is the website for the [Digital Library Federation \(DLF\) Assessment Interest Group \(AIG\) Metadata Working Group](#), also known as the DLF Metadata Assessment Working Group.

The DLF Metadata Assessment Working Group aims to build guidelines, best practices, tools and workflows around the evaluation and assessment of (mostly, descriptive) metadata used by and for digital libraries and repositories. The foci of the work of the 2016 DLF Metadata Assessment Working Group are:

1. performing an environmental scan on the topic of metadata quality and assessment;
2. gathering use cases and definitions for metadata assessment needs and realities;
3. creating a preliminary framework and set of recommendations on metadata assessment.

## 2 TAKE PART

---

We want this website and the information contained therein to be a living document. Working Group members will continue to add resources, recommendations, and information to this website as well as to our working documents. If you want to make a comment, suggestion, correction or update, you can:

[submit a pull request to this website](#),  
[open a GitHub issue on this repository](#),  
[comment on our working docs in this Google Drive folder](#),  
[send a message to our Google Group](#),  
or [join one of our calls](#).

Anyone and everyone is welcomed to take part in the DLF AIG Metadata Working Group. Use the link below to find out more about the group or get involved.

[our DLF wiki page](#)

## 3 ENVIRONMENTAL SCAN

---

This is our first area of work for the 2016 DLF Metadata Assessment group. We performed a literature, tools, presentations, and organizations review on the topics of metadata assessment and metadata quality - with a focus on (but not limited to) digital repositories descriptive metadata.

### 3.1 ORGANIZATIONS & GROUPS

---

[Draft and Notes for this Section](#)

#### 3.1.0 SUMMARY

A wide range of groups are addressing issues related to metadata and issues related to assessment of library services, but relatively few are directly working on the assessment of metadata. Here are some organizations and resources of interest.

## 3.1.0 EUROPEANA

Europeana is actively working to develop quality standards for metadata. The [Data Quality Committee](#) is addressing many issues related to metadata, including required elements for ingest of EDM data and meaningful metadata values in the context of use. “This work includes measures for information value of statements (informativeness, degree of multilinguality...)” (p. 3). Of particular note is the committee’s statement on data quality: “The Committee considers that data quality is always relative to intended use and cannot be analysed or defined in isolation from it, as a theoretical effort” (p. 1).

[Europeana’s Report and Recommendations from the Task Force on Metadata Quality](#) is an essential read, outlining broad issues related to metadata quality as well as specific recommendations for the Europeana community. This report defines good metadata quality as “1. Resulting from a series of trusted processes 2. Findable 3. Readable 4. Standardised 5. Meaningful to audiences 6. Clear on re-use 7. Visible” (p. 3). In addition, the report explores hindrances to good metadata quality: lack of foresight for online discovery, treating metadata as an afterthought, lack of funding and resources, describing digitized items with little information, and not understanding the harvesting requirements.

The [Task Force on Enrichment and Evaluation’s Final Report](#) provides ten recommendations for successful enrichment strategies. See Valentine Charles and Juliane Stiller’s presentation [Evaluation of Metadata Enrichment Practices in Digital Libraries](#) provides additional background information for this report.

## 3.1.0 HYDRA GROUPS

The [Hydra Metadata Interest Group](#) has multiple subgroups that have developed best practices for [technical metadata](#), [rights metadata](#), [Segment of a File structural metadata](#), and [Applied Linked Data](#). The best practices and metadata application profiles developed by these groups can help in the assessment of metadata quality, but the work of these groups has not yet explicitly included metadata assessment. The [Hydra Metrics Interest Group](#) is involved in the use of scholarly and web metrics to assess the performance of various aspects of Hydra instances.

## 3.1.0 SAA

Although very little directly related to metadata assessment is available from the SAA, the 2010 presentation by Joyce Celeste Chapman, “[User Feedback and Cost/Value Analysis of Metadata Creation](#)” contains many findings that merit consideration. This project studied the information seeking behavior of researchers and regarded successful searches as indicative of the value of metadata. The fields used most often by researchers were identified and the time needed to create metadata for those fields was analyzed in order to determine if the time spent creating metadata was related to the frequency of researcher use.

## 3.1.0 ALA ALCTS "BIG HEADS"

(i.e. “[ALA ALCTS Technical Services Directors of Large Research Libraries IG \(Big Heads\)](#)”)

The [Final Report of the Task Force on Cost/Value Assessment of Bibliographic Control](#) defines the value of metadata as:

1. Discovery success
2. Use
3. Display understanding
4. Ability of our data to operate on the open web and interoperate with vendors/ suppliers in the bibliographic supply chain
5. Ability to support the FRBR user tasks
6. Throughput/Timeliness
7. Ability to support the library’s administrative/management goals

The use of “ability to support the FRBR user tasks” as a criterion for assessment of metadata quality was cited in Chapman’s (2010) presentation as an indicator of metadata quality.

The report found that describing the cost of metadata is extremely difficult, especially when considering the various operations that support and enable the creation of metadata. The authors acknowledge that the nebulous definitions of value outlined in the report also create challenges for defining what is meant by “cost” in this context.

## 3.1.0 USETDA

(i.e. “[US Electronic Thesis and Dissertation Association](#)”)

The 2015 presentation “[Understanding User Discovery of ETD: Metadata or Full-Text, How Did They Get There?](#)” describes the use of web traffic for metadata analysis. The percentage of successful searches that included terms from an item’s descriptive metadata vs. the percentage of successful searches that included terms from the full text of an item was analyzed to determine how often descriptive metadata contributed to the discovery of an item.

## 3.1.0 ALCTS/LITA METADATA STANDARDS

This joint committee has recently drafted “[Principles for Evaluating Metadata Standards](#)”, which provides seven principles intended to apply to the “development, maintenance, governance, selection, use, and assessment of metadata standards” by LAM organizations. The principles recommend metadata standards that meet real-world needs, are open, flexible, and actively maintained, and that support network connections and interoperability. A recent committee blog post [summarizes and responds to public comments](#) made on the initial draft, with a subsequent draft expected later this spring.

## 3.2 PRESENTATIONS

---

[Draft and Notes for this Section](#)

Presentations collected are organized below from earliest to most current. If you have a presentation or description to add, please see the [Take Part](#) section - and thank you!

### 3.2.0 2003

#### 3.2.0.0 DLF FORUM

*Cushman Exposed! Exploiting Controlled Vocabularies to Enhance Browsing and Searching of an Online Photograph Collection*

Dalmau, Michelle; Riley, Jenn.

[Slides](#)

An interesting looking at early metadata quality control/assessment.

### 3.2.0 2015

#### 3.2.0.0 DPLAFEST

*Can Metadata be Quantified?*

Harper, Corey.

[Slides](#)

[Visual](#)

This presentation shares preliminary results of a study of converting data on metadata into numeric and visual representations, based on a case study using DPLA Providers’ metadata.

#### 3.2.0.0 ELAG

*Datamazed*

Koster, Lukas

[Slides](#)

[Notes on presentation](#)

A presentation about the blog post “[Analysing library data flows for efficient innovation.](#)”

#### 3.2.0.0 ALA ANNUAL

*We’ve Gone MAD: Launching a Metadata Analysis & Design Unit at the University of Virginia Library*

Glendon, Ivy.

[Slides](#)

A look at the background and results of a reorganization of metadata work at the UVA Library. This evaluation of university and library needs in relation to metadata services resulted in a new unit that focuses on a holistic approach that hopes to ensure consistency across systems, library units, and the university. An approach that includes metadata assessment as a part of overall plan.

### 3.2.0.0 DCM1

*Metadata Quality Control for Content Migration: The Metadata Migration Project at the University of Houston*

Weidner, Andrew; Wu, Annie.

[Slides & Paper](#)

This is a report on a migration project that resulted in the development of scripts to create reports on existing metadata used to identify problems and allow for cleanup and preparation for the metadata to be published as linked data.

### 3.2.0.0 TENNESSEE SHAREFEST

*Metadata Quality Analysis*

Harlow, Christina.

[GitHub Repository for Interactive Presentation](#)

This is an introduction to resources that can help with extracting metadata for reviewing for quality analysis.

Tools covered: MARCEdit, OpenRefine, Python Scripting, and Catmandu

### 3.2.0.0 DLF FORUM

*Statistical DPLA: Metadata Counting and Word Analysis*

Harper, Corey

[Session Notes](#)

This is a report on progress of a research project that focuses on word analysis in DPLA metadata to discover relationships in word pattern usage among DPLA providers versus search terms used versus social media language used in reference to DPLA collections. The research results will help inform best practices in metadata implementation.

*Automating Controlled Vocabulary Reconciliation*

Neatrou, Anna; Myntti, Jeremy

[Slides](#)

A case study at the University of Utah of metadata cleanup approaches, including OpenRefine, as applied to controlled vocabulary.

### 3.2.0.0 LITA FORUM

*Data Remediation: A View from the Trenches*

Harlow, Christina; Wilson, Heather

[Slides & Resources](#)

This session was a sharing of the difficulties we still face in automating data cleanup processes and a look at tools that can complement each other and, when used together, can solve some of the challenges. Tools discussed include OpenRefine, MARCEdit, PyMARC, Python, Catmandu, Google Apps Scripts, SUSHI scripting, and API calls.

### 3.2.0.0 SWIB

*Evaluation of Metadata Enrichment Practices in Digital Libraries: Steps towards Better Data Enrichments*

Charles, Valentine; Stiller, Juliane.

[Slides](#)

[Video](#)

A look at semantic enrichment tools and their effectiveness within. Covers an overview and evaluation of the why and what of semantic enrichment, using the Europeana Cultural Heritage domain as an example. See the [Report on Enrichment and Evaluation](#).

## 3.2.0 2016

### 3.2.0.0 CODE4LIB

*Measuring Your Metadata Preconference at Code4Lib 2016*  
Averkamp, Shawn; Miller, Matt; Rubinow, Sara; & Hadro, Josh.  
[Information on workshop](#)

This was a hands-on workshop that explored visualization and analysis of metadata using Python and d3. The workshop notes point to [other helpful resources](#). An outline of the workshop can be found [here](#).

*Get Your Recon*  
Harlow, Christina  
[Slides](#)

This presentation discusses the possibility of more efficient methods of preparing library data for the linked data environment beyond the traditional manual cleanup workflows.

### 3.2.0.0 DPLAFEST

*Perspectives on Data and Quality.*  
Gueguen, Gretchen; Harper, Corey; & Stanton, Chris.  
[Session information](#)  
[Slides](#)

This presentation offers three perspectives on DPLA data: an overview of the data, usage, and language in the descriptions; the strategies involved in data quality control across the collection; and data quality in aggregation.

## 3.3 PUBLICATIONS

---

[Draft and Notes for this Section](#)

### 3.3.0 SUMMARY

The group surveyed more than 50 documents produced between 2002 to 2016, ranging from journal articles, white papers, and reports to blog posts and wikis.

Bruce and Hillmann's 2004 article, "The Continuum of Metadata Quality," which defines a framework with seven categories of metadata quality (completeness, accuracy, conformance to expectations, logical consistency, accessibility, timeliness, provenance), is particularly noteworthy for influencing the subject's subsequent exploration.

In 2013, Hillmann and Bruce revisited their original framework in the context of the linked open data environment, highlighting additional considerations such as licensing, correct/consistent data modeling, and the implications of linked data technology on definitions of metadata quality. Exploring what metadata quality means in large-scale aggregators, such as Europeana and DPLA, is another topic found in recent work.

Metadata assessment involves articulating conceptual criteria and frameworks as well as developing actionable methods to collect specific information about collections. The documents we surveyed tend to focus on the following themes:

- Development of conceptual frameworks/models/metrics for defining metadata quality
- Enrichment of existing datasets to meet quality metrics
- Changes to metadata over time
- Measurement of auditing quality
- Considerations for shared metadata

A common theme across the publications we reviewed is the subjective nature of "quality," since its definition is dependent upon local context and content as well as institutional goals. According to Hillmann and Bruce (2013), conceptual criteria are "the lenses that help us know quality when we see it." Through building a community of practice for assessing metadata quality, we will be better positioned to have a shared vision, one that provides for the sustainability of our resources and meets the needs of our users and systems.

### 3.3.0 PUBLICATIONS OF NOTE

The publications review group gathered citations from existing personal collections, library-science focused databases, and Google scholar. These are available in a Zotero group library ([https://www.zotero.org/groups/metadata\\_assessment](https://www.zotero.org/groups/metadata_assessment)), which serves as a collaborative shared repository of all the resources explored as part of this process. We found a subselection of the articles, listed below, to be especially influential—these articles are a good starting point for librarians interested in learning about metadata assessment.

- Bruce, Thomas R. & Hillmann, Diane I. (2004). The Continuum of Metadata Quality
- Bruce, Thomas R. & Hillmann, Diane I. (2013). Metadata Quality in a Linked Data Context (blog post).
- Europeana Tech. Evaluation and Enrichments Task Report Outcomes. URL: <http://pro.europeana.eu/get-involved/europeana-tech/europeanatech-task-forces/evaluation-and-enrichments>
- Gavrilis, Dimitris, et al. (2015). “Measuring Quality in Metadata Repositories.” In S. Kapidakis, C. Mazurek, & M. Werla (Eds.), Research and Advanced Technology for Digital Libraries: 19th International Conference on Theory and Practice of Digital Libraries, TPDL 2015, Poznań, Poland, September 14-18, 2015, Proceedings.
- Park, Jung-ran, and Tosaka Yuji. “Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms.” *Cataloging & Classification Quarterly* 48, no. 8 (2010): 696–715.
- Stvilia, B., Gasser, L. (2008). Value based metadata quality assessment. *Library & Information Science Research*, 30(1), 67-74. URL: <http://dx.doi.org/10.1016/j.lisr.2007.06.006> (Full paper: [http://myweb.fsu.edu/bstvilia/papers/stvilia\\_value\\_based\\_metadata\\_p.pdf](http://myweb.fsu.edu/bstvilia/papers/stvilia_value_based_metadata_p.pdf))
- Zavalina, Oksana; Kizhakkethil, Priya; Alemneh, Daniel Gelaw; Phillips, Mark Edward; & Tarver, Hannah. (2015). Building a Framework of Metadata Change to Support Knowledge Management. URL: <http://digital.library.unt.edu/ark:/67531/metadc505014>

### 3.3.0 FULL CITATIONS LIST

We will continue to collect citations of interest in [the Zotero Group](#). We welcome any additions or updates you want to offer to that list.

## 3.4 TOOLS

[Draft and Notes for this Section](#)

### 3.4.0 SUMMARY

The DLF Metadata Assessment group surveyed and analyzed:

- general data tools,
- cultural heritage institution metadata-specific tools,
- programming languages/libraries that support metadata-specific activities, and
- datasets and dataset aggregators.

The 2016 environmental scan captures information about the use, status, and application of 20 tools.

#### 3.4.0.0 HOW TO READ OUR TOOLS DOCUMENT

The [Tools Documentation](#) is intended to aid the evaluation of tools for potential use in metadata assessment.

The documentation presents general information about each tool, such as its purpose and type along with a descriptive summary and URL. The documentation also provides details that may influence adoption, such as technical requirements, support, and budgetary considerations. Links to source code and documentation are included for further research.

#### 3.4.0.0.0 TOOLS OVERVIEW SHEET

Columns	Definitions	Values
Lit Review ID	Identifier to track Tool description across multiple tabs	MA-### (abbreviation for Metadata Assessment with incrementing number)
Assessment Grouping	Description of emerging trends identified in the group’s literature	Free text

	review which tool supports/could support	
Tool Name	Name of the tool assessed	Free text
Designed For	Description of intended use based on documentation or user feedback	efficiency and assessment across large datasets, assessing metadata, statistical computing, graphics, integrated development environment (IDE), data visualization, business intelligence, sharing and testing [datasets]
Type	Type of tool assessed	programming language or library, stand-alone script, tool, tools package, dataset, computing framework
URL	General URL for tool or tool information	URL
Abstract	Brief summary of the tool, its significant characteristics and relevant considerations	Free text
Other	Additional notes field.	Free text
Tool Creator/Maintainer	Individual or organization responsible for tool creation and/or maintenance	Free text
Source code / download URL	Destination for source code or download	URL
Documentation	Destination for tool documentation	URL
GUI	Designates if tool has a graphical user interface	y,n
CLI	Designates if tool is available for the command line	y,n
Free?	Designates if tool is freely available	y,n
OSS or proprietary	Designates if tool is open source or proprietary	OSS, proprietary
Written in...	Programming language tool is written in	Free text

### 3.4.0 LIST OF TOOLS & SAMPLE DATASETS TO BE ASSESSED

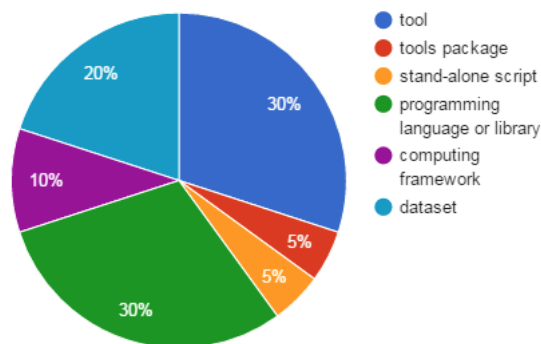
- Anaconda distribution of Python
- Apache Spark
- Completeness Rating in Europeana
- D3
- Digital Public Library of America: Bulk Metadata Download Feb 2015
- Google Analytics
- Hadoop
- Internet Archive Dataset Collection
- LODrefine
- Mark Phillips’ Metadata Breakers

- North Carolina Digital Heritage Center DPLA Aggregation tools
- OpenRefine
- Plot.ly
- Python pandas
- R
- R Studio
- SPSS
- Tableau
- UNT Libraries Metadata Edit Dataset

### 3.4.0 TOOLS OVERVIEW VISUALIZATION

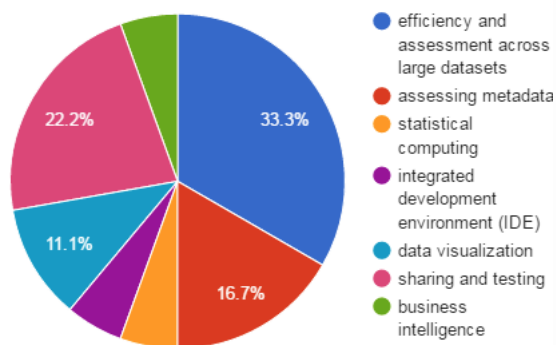
The following chart provides a quick overview of the types of tools selected for review. Many are standalone tools or programming languages; others are tools packages, standalone scripts, or computing frameworks.

Count of Type (as of 08-25-2016)



The tools we reviewed also reflect the variety of work associated with metadata assessment. Many are designed to help with assessment across large datasets, while others reflect the work of sharing and testing, statistical computing, or data visualization.

Count of Tool Design Purpose (as of 08-25-2016)



## 3.5 CITATIONS

This is a list of citations for all the resources, tools, publications, presentations, and other mentioned in the above environmental scan.

ALCTS/LITA Metadata Standards Committee. Principles for Evaluating Metadata Standards (draft). 2015-10-27. <http://metaware.buzz/2015/10/27/draft-principles-for-evaluating-metadata-standards/>



ALCTS/LITA Metadata Standards Committee. Summary of Comments Received on MSC Principles for Evaluating Metadata Standards (blog post). 2016-04-18. <http://metaware.buzz/2016/04/18/summary-of-comments-received-on-msc-principles-for-evaluating-metadata-standards/>

Alemneh, Daniel Gelaw. Understanding User Discovery of ETD: Metadata or Full-Text, How Did They Get There? 2015-09-30. <http://digital.library.unt.edu/ark:/67531/metadc725793/>

Anaconda, <https://www.python.org/>

Apache Spark, <http://spark.apache.org/>

Averkamp, Shawn; Miller, Matt; Rubinow, Sara; Hadro, Josh. Measuring Your Metadata Preconference at Code4Lib 2016 (workshop information). <http://2016.code4lib.org/workshops/#Measuring-Your-Metadata>

Chapman, Joyce Celeste. "User Feedback and Cost/Value Analysis of Metadata Creation". 2010-08-13. [http://www2.archivists.org/sites/all/files/saa\\_description\\_presentation\\_2010\\_chapman.pdf](http://www2.archivists.org/sites/all/files/saa_description_presentation_2010_chapman.pdf)

Charles, Valentine and Stiller, Juliane. Evaluation of Metadata Enrichment Practices in Digital Libraries: Steps towards Better Data Enrichments (slides and video from SWIB 2015). 2015.

Charles, Valentine and Stiller, Juliane. Evaluation of Metadata Enrichment Practices in Digital Libraries. 2015-12-18. <https://www.youtube.com/watch?v=U90Ajpgk6ic>  
<https://docs.google.com/document/d/1Henbc0lQ3gerNoWUd5DcPnNq4YxOxDW5SQ7g4f26Py0/edit#heading=h.l2fg46yn5tej>

D3, <https://d3js.org/>

Dalmau, Michelle and Riley, Jenn. Cushman Exposed! Exploiting Controlled Vocabularies to Enhance Browsing and Searching of an Online Photograph Collection. <http://www.slideshare.net/jenlrile/cushman-brownbag>

Digital Public Library of America: Bulk Metadata Download Feb 2015, <http://digital.library.unt.edu/ark:/67531/metadc502991/>

DLF/NSDL Working Group on OAI PMH Best Practices. (2007). Best Practices for OAI PMH DataProvider Implementations and Shareable Metadata. Washington, D.C.: Digital Library Federation. <https://old.diglib.org/pubs/df108.pdf>

DPLA Aggregation tools, <https://github.com/ncdhc/dpla-aggregation-tools>

Dublin Core Metadata Initiative. (2014). DCMI Task Group RDF Application Profiles. [http://wiki.dublincore.org/index.php/RDF\\_Application\\_Profiles](http://wiki.dublincore.org/index.php/RDF_Application_Profiles)

Dushay, N., & Hillmann, D. I. (2003). Analyzing Metadata for Effective Use and Re-Use. Presented at the DCMI International Conference on Dublin Core and Metadata Applications, Seattle, Washington, USA. <http://dcpapers.dublincore.org/pubs/article/view/744>

eCommons Metadata, <https://github.com/cmh2166/eCommonsMetadata>

Europeana Pro Data Quality Committee. <http://pro.europeana.eu/page/data-quality-committee>

Europeana Pro. [pro.europeana.eu](http://pro.europeana.eu)

Europeana. Report and Recommendations from the Task Force on Metadata Quality. 2015-05. [http://pro.europeana.eu/files/Europeana\\_Professional/Publications/Metadata%20Quality%20Report.pdf](http://pro.europeana.eu/files/Europeana_Professional/Publications/Metadata%20Quality%20Report.pdf)

Europeana. Task Force on Enrichment and Evaluation's Final Report. 2015-10-29. [http://pro.europeana.eu/files/Europeana\\_Professional/EuropeanaTech/EuropeanaTech\\_taskforces/Enrichment\\_Evaluation/FinalReport\\_Enrichment](http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation/FinalReport_Enrichment)

Fischer, K. S. (2005). Critical Views of LCSH, 1990–2001: The Third Bibliographic Essay. *Cataloging & Classification Quarterly*, 41(1), 63–109. [https://doi.org/10.1300/J104v41n01\\_05](https://doi.org/10.1300/J104v41n01_05)

Glendon, Ivy. We've Gone MAD: Launching a Metadata Analysis & Design Unit at the University of Virginia Library. (slides presented at the ALCTS Metadata Interest Group Meeting at 2015 ALA Annual Conference) 2015. <http://connect.ala.org/node/243993>

Google Analytics, <https://analytics.google.com/>

Gueguen, Gretchen; Harper, Corey; Stanton, Chris. Perspectives on Data and Quality (slides from DPLAFest 2016) 2016. [http://sched.ws/hosted\\_files/dplafest2016/69/DPLAFest2016PerspectivesonDataandQuality.pdf](http://sched.ws/hosted_files/dplafest2016/69/DPLAFest2016PerspectivesonDataandQuality.pdf)

Guinchard, C. (2006). Dublin Core use in libraries: a survey. OCLC Systems & Services: International Digital Library Perspectives, 18(1), 11. <https://doi.org/http://dx.doi.org/10.1108/10650750210418190>

Hadoop, <https://hadoop.apache.org/>

Harlow, Christina and Wilson, Heather. Data Remediation: A View from the Trenches. (slides and resources from the 2015 LITA forum) 2015. <https://drive.google.com/drive/folders/0ByxEB0pyAt5WOHZrOVJCVXc2X1k>

Harlow, Christina. Get Your Recon (slides from Code4Lib 2016). 2016. <http://2016.code4lib.org/Get-Your-Recon>

Harlow, Christina. Metadata Quality Analysis. (GitHub Repository for Interactive Presentation at Tennessee Sharefest 2015) 2015. <https://github.com/cmh2166/ShareFest15MetadataQA>

Harper, Corey. Can Metadata be Quantified? (slides presented at 2015 DPLAFest) 2015-04-18. [http://sched.ws/hosted\\_files/dplafest2015/c1/CanMetadataBeQuantifiedSlides.pdf](http://sched.ws/hosted_files/dplafest2015/c1/CanMetadataBeQuantifiedSlides.pdf)

Harper, Corey. Statistical DPLA: Metadata Counting and Word Analysis (session notes from DLF Forum 2015) 2015-10-28. [https://docs.google.com/document/d/1egAKg\\_Nw2kUvYJbuKOcpOGTTrQ4kIz4v5KYzGVLUEYw/edit#heading=h.c6q1qq3h66in](https://docs.google.com/document/d/1egAKg_Nw2kUvYJbuKOcpOGTTrQ4kIz4v5KYzGVLUEYw/edit#heading=h.c6q1qq3h66in)

Haslhofer, B., & Klas, W. (2010). A survey of techniques for achieving metadata interoperability. ACM Computing Surveys. <https://doi.org/10.1145/1667062.1667064>

Hydra Metadata Interest Group. <https://wiki.duraspace.org/display/hydra/Hydra+Metadata+Interest+Group>

Hydra Metrics Interest Group. <https://wiki.duraspace.org/display/hydra/Hydra+Metrics+Interest+Group>

Internet Archive Dataset Collection, <https://archive.org/details/datasets>

Jackson, A., Han, M.-J., Groetsch, K., Mustafoff, M., & Cole, T. W. (2008). Dublin Core Metadata Harvested Through OAI-PMH. Journal of Library Metadata, 8(1), 5–21. <https://doi.org/10.1080/10911360802076682>

Király, P. (2015, September). A Metadata Quality Assurance Framework. Retrieved from <http://pkiraly.github.io/metadata-quality-project-plan.pdf>

Koster, L. (2014, November 27). Analysing library data flows for efficient innovation. Retrieved from <http://commonplace.net/2014/11/library-data-flows>

Koster, Lukas. Datamazed: Analysing library dataflows, data manipulations and data redundancies. (slides presented at ELAG 2015) 2015. <http://www.slideshare.net/lukask/datamazed-with-notes>

LODrefine, <https://github.com/sparkica/LODRefine>

Loshin, D. (2013). Building a Data Quality Scorecard for Operational Data Governance. SAS Institute Inc. Retrieved from [http://www.sas.com/content/dam/SAS/en\\_us/doc/whitepaper1/building-data-quality-scorecard-for-operational-data-governance-106025.pdf](http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/building-data-quality-scorecard-for-operational-data-governance-106025.pdf)

Ma, S., Lu, C., Lin, X., & Galloway, M. (2009). Evaluating the metadata quality of the IPL. Proceedings of the American Society for Information Science and Technology, 46(1), 1–17. <https://doi.org/10.1002/meet.2009.1450460249>

Margaritopoulos, T., Margaritopoulos, M., Mavridis, I., & Manitsaris, A. (2008). A Conceptual Framework for Metadata Quality

Assessment. Presented at the DCMI International Conference on Dublin Core and Metadata Applications, Berlin, Germany. Retrieved from <http://dcpapers.dublincore.org/pubs/article/view/923>

Metadata Breakers, [https://github.com/vphill/metadata\\_breakers](https://github.com/vphill/metadata_breakers)

Najjar, J., & Duval, E. (2006). Actual Use of Learning Objects and Metadata: An Empirical Analysis. *TCDL Bulletin*, 2(2). Retrieved from <http://www.ieee-tcdl.org/Bulletin/v2n2/najjar/najjar.html>

Neatrou, Anna and Myntti, Jeremy. Automating Controlled Vocabulary Reconciliation. (slides presented at DLF Forum 2015) 2015-10-26. <http://www.slideshare.net/aneatrou/automating-controlled-vocabulary-reconciliation>

Noh, Y. (2011). A study on metadata elements for web-based reference resources system developed through usability testing. *Library Hi Tech*, 29(2), 24. <https://doi.org/http://dx.doi.org/10.1108/07378831111138161>

Ochoa, X., & Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories. *International Journal on Digital Libraries*, 10(67). <https://doi.org/10.1007/s00799-009-0054-4>

Olson, J. E. (2003). *Data Quality: The Accuracy Dimension*. Morgan Kaufmann. Retrieved from [https://books.google.com/books/about/Data\\_Quality.html?id=x8ahL57VOtcC](https://books.google.com/books/about/Data_Quality.html?id=x8ahL57VOtcC)

OpenRefine, <http://openrefine.org>

Park, E. G. (2007). Building interoperable Canadian architecture collections: initial metadata assessment. *The Electronic Library*, 25(2), 18. <https://doi.org/http://dx.doi.org/10.1108/02640470710741331>

Pirmann, C. (2009, Spring). *Alternative Subject Languages for Cataloging*. Retrieved March 24, 2016, from <http://courseweb.lis.illinois.edu/~pirmann2/LIS577/toolbox/langhead.html>

Plot.ly, <https://plot.ly>

Python pandas, <http://pandas.pydata.org>

R Studio, <https://www.rstudio.com>

R, <https://www.r-project.org>

Sicilia, M. A., Garcia, E., Pages, C., Martinez, J. J., & Gutierrez, J. M. (2005). Complete metadata records in learning object repositories: some evidence and requirements. *ACM Digital Library*, 1(4), 14. <https://doi.org/10.1504/IJLT.2005.007152>

Simon, A., Vila Suero, D., Hyvönen, E., Guggenheim, E., Svensson, L. G., Freire, N., ... Alexiev, V. (2014). *EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy: final report (Task Force Report)* (p. 44). Europeana. Retrieved from <http://pro.europeana.eu/get-involved/europeana-tech/europeanatech-task-forces/multilingual-and-semantic-enrichment-strategy>

SPSS, <http://www-01.ibm.com/software/analytics/spss>

Tableau, <http://www.tableau.com>

Tani, A., Candela, L., & Castelli, D. (2013). Dealing with metadata quality: The legacy of digital library efforts. *Information Processing & Management*, 49(6), 1194–1205. <https://doi.org/10.1016/j.ipm.2013.05.003>

Tarver, H., Phillips, M., Zavalina, O., & Kizhakkethil, P. (2015). An Exploratory Analysis of Subject Metadata in the Digital Public Library of America. In *Proceedings from the International Conference on Dublin Core and Metadata Applications 2015*. Sao Paulo, Brazil.

Task Force on Cost/Value Assessment of Bibliographic Control. Final Report. 2010-06-18. [http://connect.ala.org/files/7981/costvaluetafreport2010\\_06\\_18\\_pdf\\_77542.pdf](http://connect.ala.org/files/7981/costvaluetafreport2010_06_18_pdf_77542.pdf)

UNT Libraries Metadata Edit Dataset, <http://digital.library.unt.edu/ark:/67531/metadc304852>

Ward, J. H. (2002, November). A Quantitative Analysis of Dublin Core Metadata Element Set (DCMES) Usage in Data Providers Registered with the Open Archives Initiative (OAI) (Master's paper). School of Information and Library Science of the University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. Retrieved from <http://ils.unc.edu/MSpapers/2816.pdf>

Weidner, Andrew and Wu, Annie. Metadata Quality Control for Content Migration: The Metadata Migration Project at the University of Houston. (presentation from DCMI Global Meetings and Conferences, DC-2015) 2015. <http://dcevents.dublincore.org/IntConf/dc-2015/paper/view/339>

Zavalina, O. L. (2014). Complementarity in Subject Metadata in Large-Scale Digital Libraries: A Comparative Analysis. *Cataloging & Classification Quarterly*, 52(1), 77–89.

Zavalina, O. L., Kizhakkethil, P., Alemneh, D. G., Phillips, M. E., & Tarver, H. (2015). Building a Framework of Metadata Change to Support Knowledge Management. *Journal of Information & Knowledge Management*, 14(01). <https://doi.org/10.1142/S0219649215500057>

## 4 2016 CONTRIBUTORS

---

- Janet Ahrberg
- Shaun Akhtar «««« HEAD

### 4 – SHAWN AVERKAMP

---

master

- Filipe Bento
- Molly Bragg
- Anne Caldwell
- Joyce Chapman
- Tracy Chui
- Kevin Clair
- Robin Desmeules
- Maggie Dickson
- Laura Drake Davis
- Jennifer Eustis
- Arcadia Falcone
- Sharon Farnel
- Ethan Fenichel
- Kate Flynn
- Patrick Galligan
- Jennifer Gilbert
- Ivey Glendon
- Anna Goslen
- Peggy Griesinger
- Kathryn Gronsbell
- Wendy Hagenmaier
- Christina Harlow
- Violeta Ilik
- Dana Jemison
- Lukas Koster
- Liz Kupke
- Andrea Leonard
- Karen Majewicz
- Bill McMillin
- Timothy Ryan Mendenhall
- Amelia Mowry
- Jeremy Myntti

- Anna Neatrou
- Kayla Ondracek
- Bria Parker
- Sam Popowich
- Sarah Potvin
- Erik Radio
- Hilary Robbeloth
- Wendy Robertson
- Domenic Rosati
- Jason Roy
- Sara Rubinow
- Melissa Rucker
- Sibyl Schaefer
- Matt Schultz
- Sarah Beth Seymore
- Debra Shapiro
- Amber Sherman
- Laura Smart
- Ayla Stein
- Kathryn Stine
- Hannah Tarver
- Rachel Trent
- Friday Valentine
- Liz Woolcott
- Jennifer Young
- Angelina Zaytsev

## 5 RESOURCES & CONTACT

---

[Download this resource as a PDF](#)

This resource is a work in progress by the [DLF AIG Metadata Working Group](#). Get in touch with us via our [Google Group](#).