Digital Collections Metadata Quality Report: Baseline Elements

January 10, 2017

Compiled by Sara Rubinow and Shawn Averkamp, Metadata Services Unit

Overview

Metadata Services Unit (MSU) is conducting a large-scale audit of our Digital Collections metadata to identify the frequency and distribution of our most common metadata errors and opportunities for enhancement to help us strategize more targeted remediation projects.

This report shares the highlights from a data quality assessment of the required metadata elements necessary for the minimum viable description of digitized items. These elements directly impact core functions of the Digital Collections, such as search, display, and data sharing, and user comprehension of content. Selected results are presented with recommended remediation solutions to bring our item-level metadata up to our basic level of quality standards.

This report is intended to serve as an extensible template for future iterations. By initiating regular audits of our metadata, we'll create benchmarks to mark progress over time that will not only measure growth but also be used to identify potential correlation with Digital Collections analytics. We'll also be positioned to develop reproducible workflows for routine maintenance and remediation.

Methodology

Metadata quality dimensions

A metadata audit entails measuring metadata quality. The three dimensions of metadata quality we are measuring in this audit are:

1. Completeness: Is the element or attribute present?
2. Accessibility: Can it be understood by a human and/or a machine?
3. Conformance: Does it meet core needs of our user community by conforming to expected standards?

(See the glossary of metadata quality criteria at the end of this report for more detailed definitions of these and other additional metadata quality dimensions that we will explore in future audits.)

This audit evaluates 965,065 item-level records included in a database extract from Metadata Management System (MMS) on June 2, 2016. From the original data extract, we excluded two collections and four divisions that were not relevant to the audit evaluation.

Approach and tools

To assess the quality of our baseline metadata along the completeness, accessibility, and conformance dimensions, we created a set of fifteen quality "assertions" about our six mandatory elements, such as whether or not the element was present or if it matched an allowable data value. We used a Python script to score each assertion on a pass/fail basis for every item record and calculated total scores based on the composite scores of each mandatory element.[1] A record with a score of 6 passes all assertions and exceeds baseline criteria, while a record with a score of 0 reflects a failure of all assertions.

Results were output to a CSV file and analyzed with pandas, a Python library for data analysis. Charts were generated with pandas in conjunction with matplotlib, a Python plotting library.

Remediation methods

We suggest a broad range of remediation methods, defined below, to support a flexible approach and provide a foundation for the creation of tools and workflows to enhance efficiency and enable routine metadata maintenance.

Batch → Downloading data and cleaning in batch with another tool (e.g. Open Refine, Excel). Remediation via API.
Co-occurrence → Asserting or inferring data through existing metadata in other fields. Remediation via API or directly in MMS
Crowdsourcing → R&D-built tool collecting consensus-based corrections for larger remediation projects. May be staff-facing or public-facing depending on task. Remediation via API.
Data analysis → Review of data to identify patterns, inconsistencies. Remediation via API or directly in MMS.
Human → Data review that requires human eyes. Remediation directly in MMS.
Source record → Checking if missing data exists in source record or Registry. Remediation through API or directly in MMS

Elements

Title

Definition

A name given to a resource to identify or describe it.

[1] See https://github.com/NYPL/metadata-stats/blob/master/items-min-mandatory/getscores-items-min-mandatory.py for all assertions included in this audit.

Purpose

The primary function of the title is to help the user identify a resource or to briefly describe the content of a resource. Titles can help users quickly navigate or disambiguate within result sets. Titles marked as "primary" are indexed to serve as labels for thumbnails in display and search result sets, so their presence is required for proper functioning within MMS and Digital Collections. In addition, a title marked as "primary" ensures the preferred title is displayed in aggregators, such as DPLA.

Assertions tested:

We tested each item record for the presence of a title element and for the presence of one "primary" attribute.

Summary of results

Only five records did not include at least one title element. Three have since been deleted, and the other two records are part of a Brightcove General Audio Collection ingest and will be deleted in MMS once the captures can be removed from the repository *(item ids 5073444 and 5075118)*.

Of records with a title element, 1.4% (13,847) lacked exactly one title element marked as "primary." 95% of those records are in the Theatre Division, with the overwhelming majority in the Friedman-Abeles photograph collection (11,980), followed by the Martha Swope photographs collection (832).

Remediation actions and possible methods

Add missing titles: Identify the records that were missing a title and add a title element.

Human Due to the small number of records missing a title element, it makes sense to remediate these manually.

Add missing "primary" attributes: Identify the records missing the "primary" attribute by collection and determine whether the record has single or multiple title elements.

Batch For records with a single title, add the `usage="primary"` attribute to the title through the API.

Batch For records with multiple titles, all without "primary" attributes, add the `usage="primary"` attribute to the first title through the API.

Batch For records with multiple titles, all with "primary" attributes, remove the `usage="primary"` attribute from all titles except the first title through the API.

Batch For records with multiple titles marked "primary" that include one or more inherited titles, uninherit the inherited title element(s) through the API.

Potential for data enhancement

Completeness: Use title information to populate missing elements.

Co-occurrence For records in which information in the title could be used to populate other missing elements, such as date, genre, or name.

Examples of quality issues not yet tested

Syntactical issues that impact display on DC:

- end of line brackets and punctuation or words that should be in the title field but are in the non-sorting characters field—a holdover from traditional cataloging. (*Accuracy)*

- since titles also frequently serve as labels for thumbnails, an abbreviated title could be added and marked as “primary” for those records identified as having lengthy titles. (*Conformance)*

Usage of recommended descriptive attributes for context and accessibility:

- title language and type (e.g., “uniform”, “translated”, “abbreviated”) attributes are infrequently used but could added to enhance access and discovery. (*Completeness)*

Semantic issues that impact discovery and access on DC

- collection-level titles that are repeated in error at the container or item level (*Accuracy)*
- items in a collection that cannot be disambiguated by title (Accuracy/*Consistency)*

Type of resource

Definition

Expresses the general type and characteristics of the original resource content.

Purpose

The primary function of the resource type is to categorize material in a general way using a defined set of values. This can help applications determine the appropriate viewing or playback system and allows users to facet results by general type.

Assertions tested

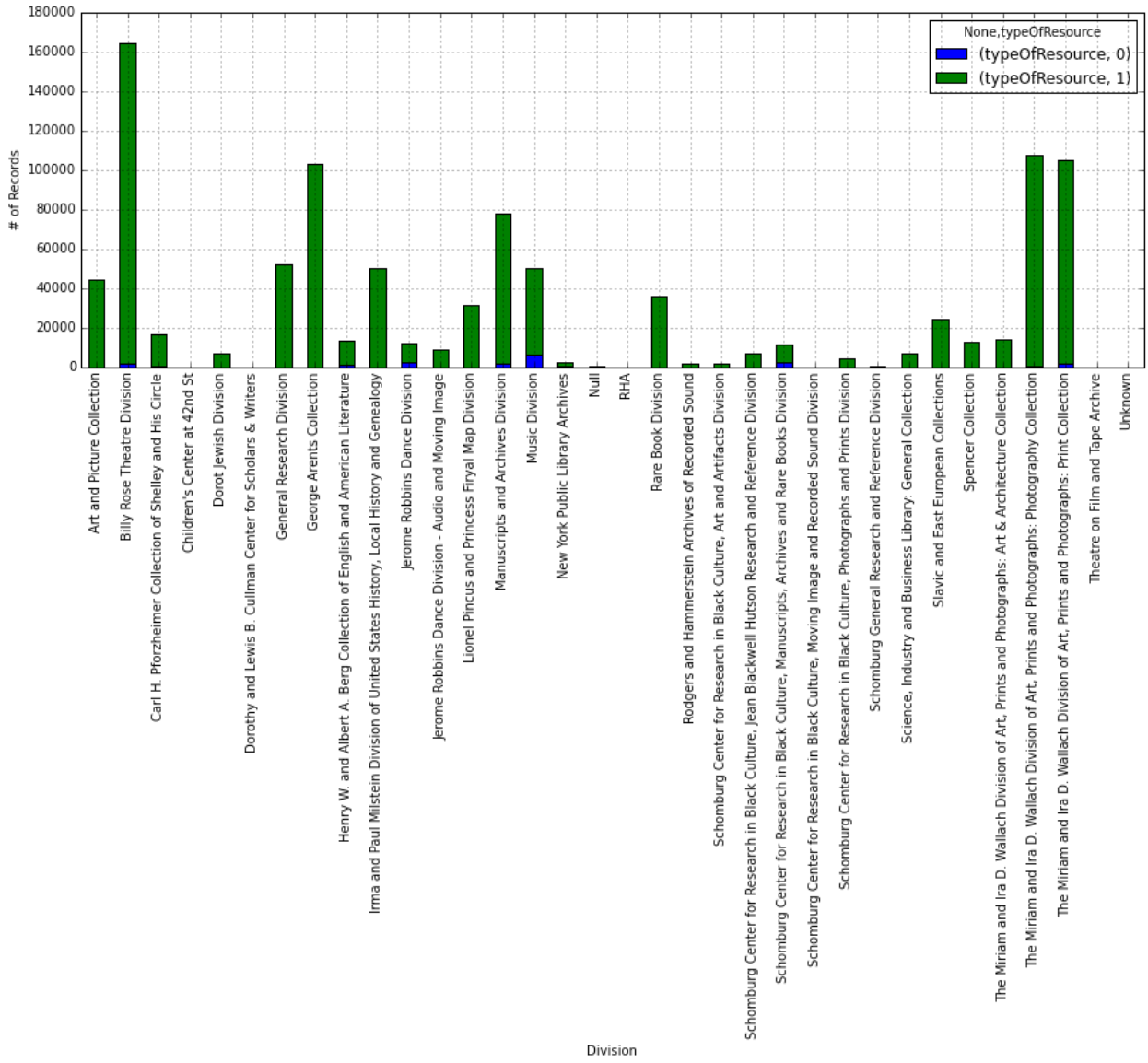
We tested each item record for the presence of at least one type of resource value from a controlled MODS list of ten values.

It is a known issue that many legacy records migrated from Hades erroneously have “still image” as the resource type for textual assets. The type of resource’s accuracy in describing the original resource content was not tested.

Summary of results

1.8% of records (18,068) did not have a type of resource.

Two divisions account for just under half of those records: the Music Division (6,155) and Schomburg Rare Books (2,465). The following chart, organized by division, shows records with at least one type of resource element in green, and records missing a type of resource element in blue.



Remediation actions and possible methods

Add missing type of resource elements: Identify the records that were missing a type of resource and add a resource type.

Co-occurrence For records with a genre element present. Items with the genre “maps” would be assigned the resource type “cartographic materials”, for example. This would take place after genre authority remediation.

Crowdsourcing Type of resource remediation requires a human eye to evaluate the digital asset as well as an understanding of the element’s definition. Because the types are from an enumerated list of ten values, a public-facing tool could provide guidance and enable users to select the type or types that best characterize the original resource content.

- This approach could be enhanced by grouping records by MMS collection ID, which would help identify issues of inheritance; there may be many instances in which the type of resource exists at the collection level and is appropriate to inherit down to child containers

and items. One user task could be verifying if a parent resource type is applicable to all children.

Potential for data enhancement

Accuracy: Remediate incorrect resource types.

Co-occurrence Facet by genre to remediate type of resource values that incorrectly characterize the content of the original resource. Items with type of resource “still image” and a genre of either “correspondence” or “maps”, for example, could be remediated to type of resource “text” and “cartographic”, respectively. This would take place after genre authority remediation.

- (722,745) records passed both resource type and genre completeness prior to genre authority remediation, indicating a robust foundation for pursuing this course of action.
- Caveat: this does not address existing issues of semantic accuracy. For example, Image ID: 3946315 has the genre “Photographs” and the resource type “still image,” but the object is a typescript document.

Examples of quality issues not yet tested

Semantic issues that impact discovery and viewing on DC and indexing in Fedora

- check that values match values in controlled MODS list *(Accuracy*)
- update instances of the “mixed material” value, which is deprecated locally (*Consistency*)
  - instances of “mixed material” have been identified, remediated, and removed from the MODS date types available in the MMS menu of date types available to metadata creators.
- verify that the values characterize original resource content *(Accuracy*)

Identifier

Definition

A unique standard number or code that distinctively identifies a resource within a given context.

Purpose

Identifiers serve to uniquely identify a resource locally, within a repository or globally, across a universe of resources. Identifiers can also serve to point the user to a description of the resource in another database, such as an original catalog record or a finding aid.

When applied consistently, identifiers can help enhance the online display of a resource by facilitating links to online source records or digital exhibits. In Digital Collections, identifiers of type “local_bnumber”, “local_mss”, and “local_tms” connect records and link users to their corresponding records in the catalog, Archives Portal, and the Wallach Prints and Photographs catalog, respectively. These connections are necessary to provide context to items in Digital Collections.

Assertions tested

We tested each item record for the presence of at least one identifier type that corresponds to the catalog (local_bnumber), the Archives Portal (local_mss), or the Wallach catalog (local_tms).

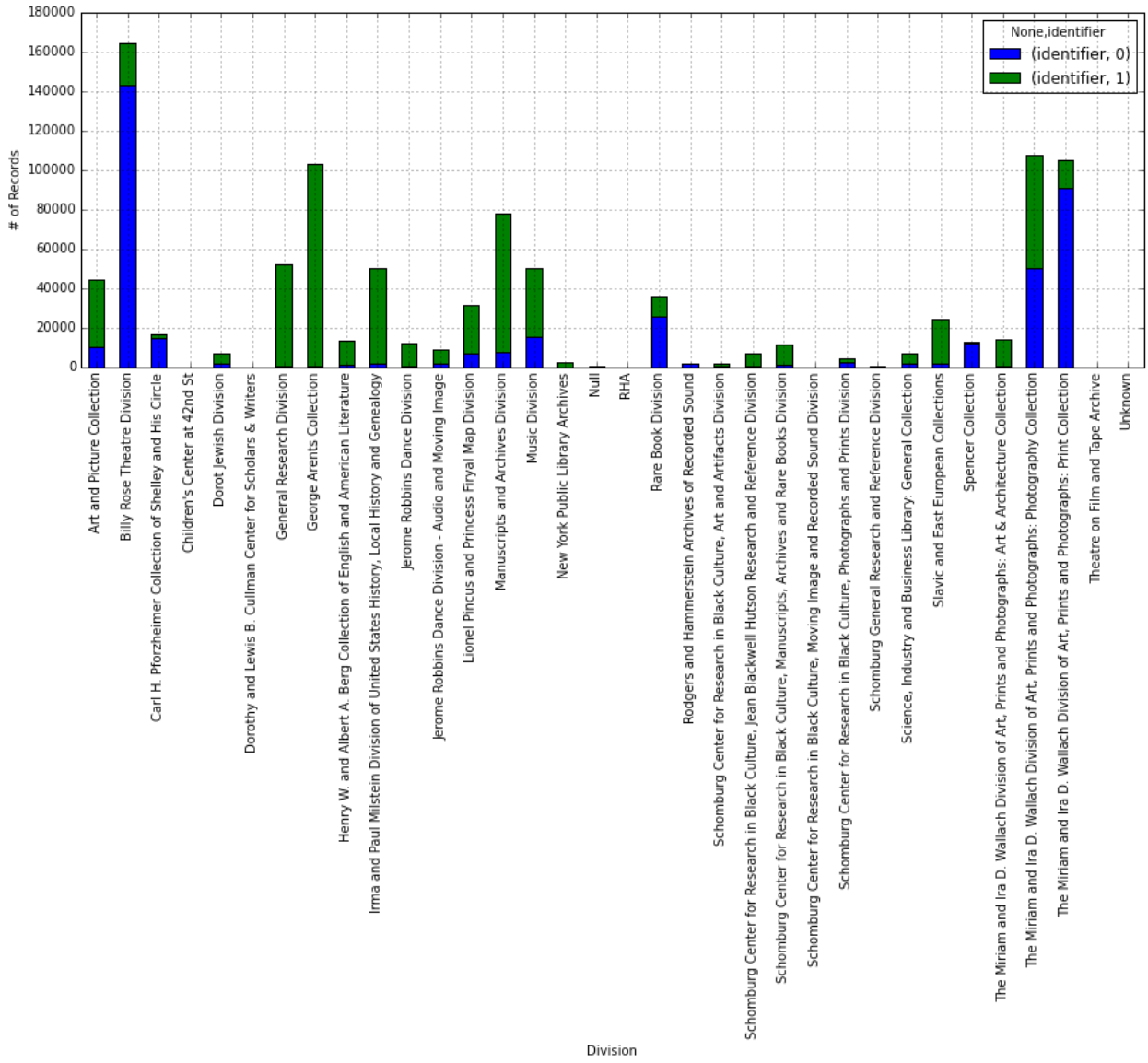
Summary of results

Of all the elements tested, the identifier element had the highest number of records that failed to meet the minimum requirement: 41% of records (393,443) did not have at least one of the three requisite identifiers.

Over a third of those records were in the Theatre Division (143,316), representing about 87% of that division's total MMS records. This can likely be attributed to be due to a substantial amount of uncataloged Theatre Division material represented in MMS. Over half (80,499) of Theatre Division records without a requisite identifier are from an uncataloged collection, the "Billy Rose Theatre Collection photograph file."

Another third of the records without the required identifiers came from the Wallach Print Collection (90,705) and Wallach Photography Collection (49,887). The majority of Print Collection records missing required identifiers (72,659) are from another uncataloged collection, the Print Collection Portrait file. Apart from the presence of uncataloged material, the lack of identifiers in the Wallach Collections may be due to the fact that scoring was limited to the presence of a "local_tms" identifier and not a TMS Object identifier.

Other divisions with a high number of records without requisite identifiers include Rare Books (25,536), Music (15,414), and Pforzheimer (14,695). The following chart, organized by division, shows records with at least one required identifier element in green, and records missing a required identifier element in blue.



Remediation actions and possible methods

Add missing identifier elements: Identify the records that were missing a required identifier and add an identifier element.

Source record Use shelf locator data to match identifiers for catalog and archives portal material.

Source record Use title and shelf locator data to match identifiers for “local_tms” identifiers.

Human The absence of identifiers in the Pforzheimer collection was previously known and is currently in the process of remediation.

Data analysis There may be collections for which child containers and items are missing an identifier that is present at the collection level and is applicable to child components. Group records by division and collection ID to identify opportunities for identifier inheritance.

Examples of quality issues not yet tested

Source record context and online display issues:

- frequency of recommended identifiers, like Archives EAD ID (*Completeness)*

- frequency of all identifier types *(Consistency)*

Syntactical issues that impact discovery and connection to source record:

- frequency of invalid identifier values due to typographical errors *(Accuracy)*
- identifier type corresponds to identifier value (e.g., an instance in which an “Archives EAD ID” type is assigned an “MSS Unit ID” value) *(Accuracy)*
- frequency of deprecated types that could be converted to updated types *(Consistency)*

Genre

Definition

Genre describes the nature of the content or function of the resource at a greater level of specificity than Type of Resource.

Purpose

The primary function of the genre element is to categorize material in a more specific way using a defined set of values drawn from controlled vocabularies. This allows users to facet results and will help to automate resource navigation and display based on the nature of the content.

Assertions tested

We tested each item record for the presence of a genre element and for the presence of an “authority” attribute that would serve as an indicator that the genre term was from an approved controlled vocabulary.

Summary of results

25% of records (237,724) did not have a genre element populated.

Four divisions account for about half of those records: General Research Division (37,070), Art and Picture Collection (33,263), Manuscripts and Archives (27,429), and the Wallach Photography Collection (17,364).

Of the 727,342 records with a genre element present, 3% (22,874) lacked an “authority” attribute. Over a third of the records were from the Wallach Print Collection (8,881).

Many genre terms migrated incorrectly from form terms in Hades, which may account for a significant number of unauthorized terms. In addition, the actual number of genre terms not from a controlled vocabulary is likely much greater since it is possible for a genre element to have an “authority” attribute with any value in the genre field. This issue is addressed in the second remediation method listed below.

Remediation actions and possible methods

Reconcile unauthorized genre terms: This step would yield the most immediate results and provide groundwork for subsequent remediation.

> Batch Identify the unauthorized genre terms and remediate through a clustering and evaluation of terms followed by batch data cleaning and reconciliation against preferred controlled vocabularies with OpenRefine.

Replace inconsistent genre terms:

Data analysis Analysis of the frequency of remediated genre values—not assessed in this test—will inform additional clean up of genre terms with an eye to facilitating user discovery.

Add missing genre terms: Identify the records that were missing a genre and add a genre element. The newly-cleaned up list of genre terms will provide a foundation for this next step.

Source record Of the 237,724 records that are missing a genre, 186,485 have a qualifying identifier. Batch remediation may be possible for those records with genre information in the source record. For example, identifiers could be iterated through Registry/ShadowCat to retrieve 650X $v information where available, or through the portal database to retrieve “material type” information as available.

Human The absence of genre terms in the Pforzheimer collection (14,525 records or about 87% of the collection) was previously known and is currently in the process of remediation.

Human Remediation will require human eyes for the remaining records.

or

Crowdsourcing A staff-facing crowdsourcing tool could be rolled out with a limited list of terms from which to select. The terms would be based on the list of genre terms vetted in the first remediation pass plus additional terms specified by each division/domain—validated by MSU—for their frequency of use, clarity, and relevance to specific collections or material.

- If the crowdsourcing tool were public-facing, the list would need to be significantly restricted to omit esoteric, technical, or potentially ambiguous terminology.

Date

Definition

Dates related to the creation or issuance of a resource.

Purpose

In addition to informing users about the origin of a resource, dates and date ranges provide access points in NYPL Digital Collections and partner aggregators, and are necessary for making copyright determinations. Values in the date field intended to be used as machine-readable, indexed data, so a minimum of one structured (i.e. encoded) date or date range is required.

Date types (such as, “date created” and “date issued”) provide context for the value in the date field. When multiple date elements are present, the coexistence of the same date type or the coexistence of a date range and a single date serve as flags for potential issues regarding the accuracy of item-level date description that may misrepresent the material and adversely impact discovery, access, and copyright determination.

Assertions tested

We tested four date element assertions: two related to completeness and two related to accuracy.

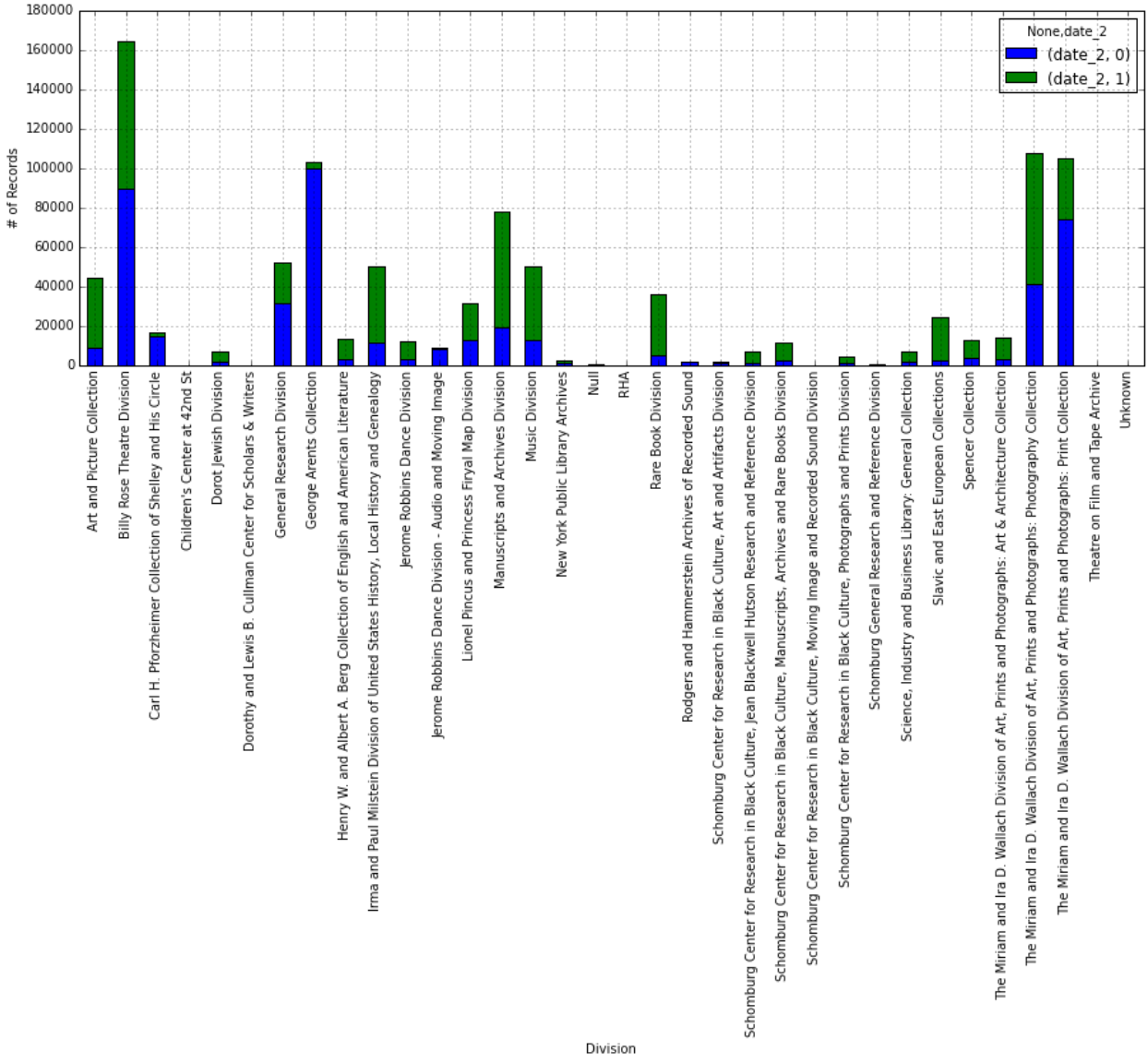
1. The presence of a date element *(Completeness)*
2. The presence of a value in the date field with an encoding attribute *(Completeness)*
3. The presence of different date types for multiple single date values *(Accuracy)*
4. The absence of date range values if multiple date elements were present. *(Accuracy)*

Summary of results

66% (640,681) of records require some kind of date remediation.

40% (387,430) of records did not have at least one date element present. About a third of the records without a date element are from the Cigarette cards collection in the Arents Collection (). The next two largest sets of records without a date element are the Billy Rose Theatre Collection photograph file in the Theatre Division (75,143), and the Print Collection portrait file in the Wallach Print Collection (59,181).

7% (68,994) of the records with a date element did not include an encoding attribute. About 20% of the unencoded records were from the Wallach Photography Collection (15,107), with the Morris Huberland collection (9,255) accounting for the majority of records therein. The following chart, organized by division, shows records with at least one date element with an encoding attribute in green, and records missing that attribute in blue.



17% (163,937) of the records had multiple date elements with single date values and the same date type used more than once. 20% (33,049) of these records were from the Manuscripts and Archives division.

4% (37,789) of records had multiple date elements that included at least one date range. About a quarter of these records (8,981) were from the Theatre Division.

Remediation actions and possible methods

Add missing date elements: Identify the records without a date element and add a date element.

Source record 51% (197,679 -- *was 183,687*) of the records without a date element have one of the three required identifiers. Wait until identifier remediation has taken place for source record remediation.

Source record 26% (100,920) of the records without a date element or required identifier type have a shelf locator. Wait until identifier remediation has taken place for source record remediation.

Human and Batch Inherit date from nearest parent level with a date element present (e.g., from a container or collection).
Human Remediation will require human eyes for the remaining records.

Convert date strings to encoded values: Identify the collections and records that were missing and encoding attribute and convert those date strings to encoded formats.

Source record For records with one of the three required identifiers. Iterate through and replace date element with an encoded date.
Batch Use regular expressions to identify strings that contain values that are likely dates (e.g., "[1841]") and qualifiers (e.g., the presence of bookend brackets would correspond to the "inferred" date qualifier attribute).
Human Remediation will require human eyes for the remaining records.

Remove date ranges that are present in error: Identify the collections and records with multiple date elements that include at least one date range and determine if they accurately describe the resource.

Human For instances in which a date range has been inherited down in error from a higher level, uninherit the date at the collection/container-level or override the date element at the item level.

Determine if instances of the coexistence of the same date type are valid: Identify the collections and records that have multiple single dates with the same date type and review for semantic accuracy.

Human For instances in which a date element has been inherited down in error from a higher level, uninherit the date at the collection/container-level or override the date element at the item level. For instances in which the date element was added redundantly at the item level, delete the date element. For instances in which a date type was selected in error, change the date type.

Other quality issues not addressed by this test

Syntactical issues that impact discovery in DC and indexing:

- Date is encoded as an acceptable string: yyyy, yyyy-mm, or yyyy-mm-dd *(Completeness, Accessibility)*
- When date is encoded as a range, the start date is earlier than the end date (null end date values are acceptable) *(Accuracy)*
- Year(s) in the date field(s) is/are earlier than 2017 *(Accuracy)*
- Encoded date with type "dateCreated" is earlier or equivalent to an encoded date with type "dateIssued" *(Accuracy)*
- When more than one date instance is present, "key date" is checked only once *(Completeness)*
- When date is encoded as a range, "key date" is checked for the start of the range *(Completeness)*

Use of locally-preferred schema date types:

- Identify MODS date types not used in our profile: copyrightDate, dateCaptured or dateValid, or dateModified *(Consistency*)
    - instances of dateValid have been identified, remediated, and removed from the MODS date types available in the MMS menu of date types available to metadata creators.

Physical location

Definition
Information related to the location of and library unit responsible for the original physical resource.

Purpose
The primary function is to help users and curators find the original physical object. Location information also helps users and staff direct questions about the object to the appropriate division. Division location may be required in the future for applying access restrictions (i.e. access by IP range) to specific locations at NYPL. The division location also enables us to produce more in depth reports from usage analytics and production statistics for NYPL centers, divisions, and collections. The shelf locator field provides the relative intellectual location of the material within a collection and division. Sometimes, in the absence of a catalog or finding aid record, the shelf locator doubles as a unique identifier for the physical object within the division's collections.

Assertions tested
We tested five assertions for the physical location element.

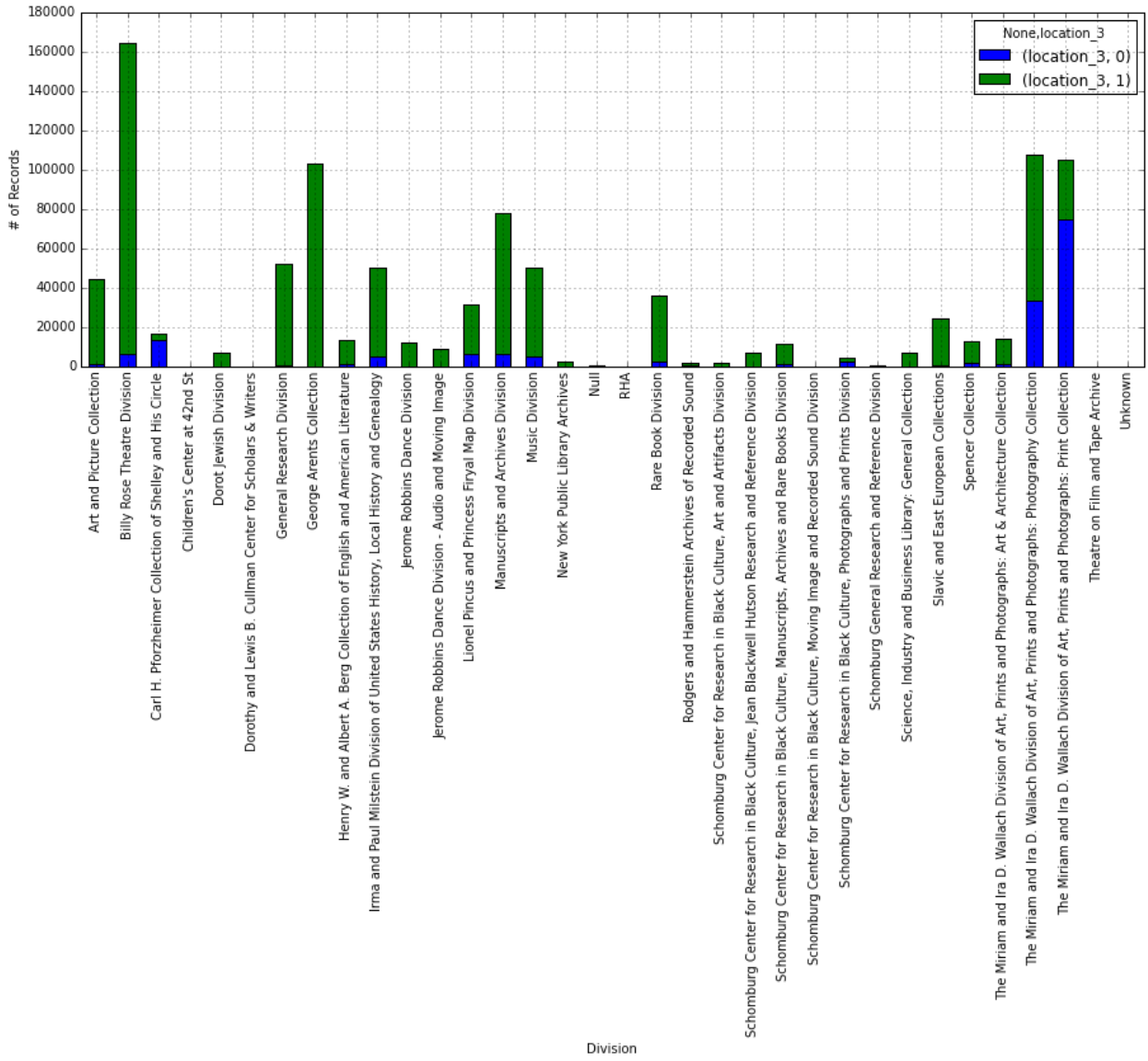
1. The presence of a repository location
2. The presence of a division location
3. The presence of a shelf locator location
4. If multiple division locations are present in a record they do not conflict
5. Division location information is expressed in the three division location types (long name, short name, code).

Summary of results
75% (729,856) of records have a repository, division, and shelf locator location..

8% (79,769) of records lack a repository location. Over half the records (42,222) without a repository location are from the Wallach Photography Collection, with 90% of those records (38,546) from the Robert N. Dennis collection of stereoscopic views.

Just 0.02% (275*)* of records lack a division location. 30% (83) of these records are from a collection that no longer exists in MMS; of the remaining 0.01% (182) of records that lack a division location*,* 25% (45) are from the John Bigelow papers, a Manuscripts and Archives collection*.*

17% (161,537) of records lack a shelf locator location. 67% (108,910) of records without a shelf locator are from the Wallach Division, with two-thirds (72,687) from a single collection, the Print Collection portrait file. The following chart, organized by division, shows records with at least one shelf locator in green, and records missing a shelf locator in blue.



3% (26,769) of records have multiple location elements with inconsistent division values. By collection: 34% (9,125) of the records with inconsistent division values are from the Martha Swope photographs collection and another 21% (5,566) of those records are from the Benjamin K. Miller collection of United States stamps. A large number of collections with inconsistent division values are due to the Dance Division's known practice of adding dual divisions ("Jerome Robbins Dance Division" and "Jerome Robbins Dance Division - Audio and Moving Image") to its MMS records. The Dance Division's dual division issue has since been resolved.

0.05% (549) of records with a division location lack all three division location types (long name, short name, code). 90% (498) of records without all three division types are from the Schomburg General Research and Reference Division. This is a known issue that is currently in the Digital Repository development backlog.

Remediation actions and possible methods

Add missing repository values: Identify the records without a repository value and add a repository value.

Source record 75% (59,938) of the records without a repository value have one of the three required identifiers. Wait until identifier remediation has taken place for source record remediation.
Co-occurrence If the division is an NYPL division, add NYPL as the repository.

Add missing division values: Identify the records without a division value and add a division value.

Human Due to the small number of records with easily-identified collections, remediation could be expedited without need for API ingestion or repo team bandwidth. The largest single collection, for example, could be remediated by inheriting down the collection element.
Source record The majority of the records without a division value have one of the three required identifiers. For any records unable to be reconciled via human action, wait until identifier remediation has taken place for source record remediation.

Add missing shelf locator values: Identify the records without a shelf locator value and add a shelf locator value when available.

Source record Just 4% (6,463) of the records without a shelf locator value have one of the three required identifiers. Wait until identifier remediation has taken place for source record remediation.

Resolve conflicting division values: Identify the records with conflicting division values and resolve to a single division value.

Human Items in 181 collections have conflicting division values. Remediation will be manual on a collection-by-collection basis.
Source record 79% (21,273) of the records with conflicting division values have a shelf locator location.
Human Remediation will require human eyes for the remaining records.

Add missing division location types : Identify the records with a division location that are missing all three division location types and add missing location types

Batch Use division type(s) present in the record to generate the missing type(s)

Address outstanding missing values: Identify the records that are still missing values that were not remediated with other methods.

Human Remediation will require human eyes for the remaining records.

Other quality issues not addressed by this test

Duplication/redundancy of field values within a record:

- Identify records with multiple location elements that share identical division values *(Conformance)*

Source record context and online display issues:

- Identify records that have multiple location elements with inconsistent shelf locator values *(Accuracy)*

Syntactical issues that impact discovery and connection to source record:

- Identify records with shelf locator locations that do not match a shelf locator value *(Accuracy)*

Addenda

The Pforzheimer records missing a shelf locator have been remediated.

Glossary of metadata quality criteria

Accessibility: Values are either machine-readable and/or human-readable. ”Metadata that cannot be read or understood by users has no value.”[2]

Accuracy: “[T]he semantic distance between the information a user could extract from the metadata instance and the information the same user could obtain from the resource itself…”[3] plus the additional specificity that may be added as encoded data (e.g., an encoded longitude/latitude is more precise than a street address). Note: Only syntactic accuracy was measured in this audit because it is straightforward to automate.

Completeness: “[A] metric indicating the percentage of completeness of the elements of a schema… [This] can be analyzed into three partial measures: (i) completeness of the mandatory set of elements, (ii) completeness of the ‘recommended’ element set and (iii) completeness of optional elements.”[4]

Conformance: How values adhere to the expectations of a user community. “The degree to which metadata fulfills the requirements of a given community of users for a given task…”[5]

Consistency: When metadata values are consistent within the domain and are applied in a consistent manner across records.

> *Semantic consistency* is the "extent to which the collections use the same values (vocabulary control) and elements for conveying the same concepts and meanings throughout."[6]
> *Structural consistency* is the extent to which similar elements of a metadata record are represented with the same structure and format.[7]

---

[2] Bruce, Thomas R., and Diane I. Hillmann. “The continuum of metadata quality: defining, expressing, exploiting.” in *Metadata in Practice*, edited by Diane I. Hillmann and Elaine L. Westbrooks, 238-256. Chicago: ALA Editions, 2004.

[3] Ochoa, Xavier, and Erik Duval. “Automatic Evaluation of Metadata Quality in Digital Repositories,” [preprint], revised version published in *International Journal on Digital Libraries* 10 (August 2009): 61-97. doi: 10.1007/s00799-009-0054-4.

[4] Gavrilis, Dimitris, et al. (2015). “Measuring Quality in Metadata Repositories.” In S. Kapidakis, C. Mazurek, & M. Werla (Eds.), *Proceedings from the 19th International Conference on Theory and Practice of Digital Libraries*. Paper presented at TPDL 2015, Poznań, Poland, September 14-18, (pp. 56-67). Cham: Springer. doi: 10.1007/978-3-319-24592-8_5.

[5] Ochoa and Duval, 2009.

[6] Shreeves, Sarah L., Knutson, Ellen M., Stvilia, Besiki, Palmer, Carole L., Twidale, Michael B., and Cole, Timothy W., "Is “Quality” Metadata “Shareable” Metadata? The Implications of Local Metadata Practices for Federated Collections," ACRL Twelfth National Conference, April 7–10, 2005, Minneapolis, Minnesota, https://www.ideals.illinois.edu/bitstream/handle/2142/145/shreeves05.pdf

[7] Westbrook, N. R., Johnson, D., Carter, K., & Lockwood, A. (2012). Metadata Clean Sweep: A Digital Library Audit Project. D-Lib Magazine, 18(5/6), doi:10.1045/may2012-westbrook