Implementation of the DS 2.0 Data Model as the Digital Scriptorium Catalog
Outline of Principles, Practices, and Decisions for DS Authorities and Data Enrichment

Last updated: 2024-10-28

L.P. Coladangelo
DS Catalog Project and Data Manager

1. **DS 2.0 Record: Extraction, Reconciliation, and Enrichment of Metadata Values**
   1.1. **Century (and Date)**
      1.1.1. *From Metadata Record*
         1.1.1.1. Data values extracted from member records
            1.1.1.1.1. Semi-structured string value
               1.1.1.1.1.1. Production Date as Recorded
            1.1.1.1.2. Structured date information where available
               1.1.1.1.2.1. Earliest Date
               1.1.1.1.2.2. Latest Date
            1.1.1.1.3. Century values are automated at extraction based on range of date values
               1.1.1.1.3.1. Reconciled automatically against controlled list of AAT values at data extraction from structured date information
      1.1.2. *Reconciliation Process*
         1.1.2.1. Data values for century qualified/enriched with authority values
            1.1.2.1.1. Authorized Label
               1.1.2.1.1.1. Label from AAT record
            1.1.2.1.2. Structured Value
               1.1.2.1.2.1. URI from AAT record
         1.1.2.2. Automated reconciliation: Values matched to AAT values at data extraction using controlled list
            1.1.2.2.1. Rationale for Use of AAT
               1.1.2.2.1.1. Well-known, established controlled vocabulary used by many cultural heritage institutions
               1.1.2.2.1.2. Allows for an object-oriented approach to century values in support of faceted browsing and searching
               1.1.2.2.1.3. Offers additional discovery by time period beyond Wikibase Point in Time datatype
      1.1.3. *Wikibase Implementation*
         1.1.3.1. Production date string is represented in DS 2.0 Record as the value of a statement for a Production Date as Recorded (P23)
         1.1.3.2. Statement for Production Date as Recorded is qualified by
            1.1.3.2.1. Earliest Date (P37) value extracted from member data

1.1.3.2.2.   Latest Date (P36) value extracted from member data

1.1.3.2.3.   Century (P25) Point in Time value

1.1.3.2.4.   Production Century in Authority File (P24) where structured dates are reconciled to AAT

1.1.4.   Associated Wikibase Authority Record

1.1.4.1.   Type: Century (Authority File): Q13

1.1.4.2.   Authority Record acts as authority control for annotated/qualified Production Date as Recorded value

1.1.4.3.   Label is derived from authorized label from AAT

1.1.4.4.   External Identifier (P44): AAT URI

1.2.   Genre

1.2.1.   From Metadata Record

1.2.1.1.   Data values extracted from member records

1.2.1.1.1.   Semi-structured or structured string value

1.2.1.1.1.1.   Genre as Recorded

1.2.1.1.2.   Vocabulary Designator

1.2.1.1.3.   Authority Record Identifier/Link

1.2.2.   Reconciliation Process

1.2.2.1.   Data values for genres qualified through reconciliation

1.2.2.1.1.   Authorized Label

1.2.2.1.1.1.   Preferred Label from AAT record (default)

1.2.2.1.1.1.1.   For terms derived from AAT

1.2.2.1.1.1.2.   FAST labels are used for FAST/LCSH/LCGFT/TGM/RVMGF designated values

1.2.2.1.1.1.3.   RMBS labels are used for RBMS designated values

1.2.2.1.1.1.4.   LoBT labels are used for LoBT designated values

1.2.2.1.2.   Structured Value

1.2.2.1.2.1.   URI from AAT record (default)

1.2.2.1.2.1.1.   For terms derived from AAT

1.2.2.1.2.1.2.   FAST URIs are used for FAST/LCSH/LCGFT/TGM/RVMGF designated values

1.2.2.1.2.1.3.   RMBS URIs are used for RBMS designated values

1.2.2.1.2.1.4.   LoBT URIs are used for LoBT designated values

1.2.2.2.   Semi-automated reconciliation: Values matched to AAT through OpenRefine reconciliation service for Getty Vocabularies

1.2.2.2.1.   Rationale for Use of AAT (default)

1.2.2.2.1.1.   Genre values in metadata records often use a variety of vocabularies, with no emergent standard

- 1.2.2.2.1.2. AAT has robust coverage for multiple terms which relate to unique cultural heritage materials and objects, including manuscript culture
- 1.2.2.2.1.3. Makes important differentiations between concepts explicit through parenthetical qualifiers
- 1.2.2.2.1.4. Allows automatic post-coordinate faceting of terms since the vocabulary is not intended for the pre-coordinate indexing
- 1.2.2.2.2. Where genre strings are constructed, the designated vocabulary is used for the first term and an implied vocabulary is used for all remaining terms in the string (usually FAST) based on the terms used

- *1.2.3. Wikibase Implementation*
  - 1.2.3.1. Genre is represented in DS 2.0 Record as the value of a statement for Genre as Recorded (P18)
  - 1.2.3.2. Statement for Genre as Recorded is qualified by
    - 1.2.3.2.1. Term in Authority File (P20) where genres are reconciled to AAT (or FAST, RMBS, LoBT)
      - 1.2.3.2.1.1. Allows retention of string value from original record but semantically annotated to corresponding items in the Term Authority

- *1.2.4. Associated Wikibase Authority Record*
  - 1.2.4.1. Type: Term (Authority File): Q11
  - 1.2.4.2. Authority Record acts to annotate/qualify string values for the presence of genre information
    - 1.2.4.2.1. Database locally duplicates a skeletal AAT/FAST/RBMS/LoBT record in Wikibase by linking to the external AAT/FAST/RBMS/LoBT record for the same concept or entity represented in a DS 2.0 Record
    - 1.2.4.2.2. Label is derived from authorized label for AAT/FAST/RBMS/LoBT concept or entity obtained through reconciliation
    - 1.2.4.2.3. External identifier (P44): AAT/FAST/RMBS/LoBT URI

1.3. Language

- *1.3.1. From Metadata Record*
  - 1.3.1.1. Data values extracted from member records
    - 1.3.1.1.1. Semi-structured string value
      - 1.3.1.1.1.1. Language as Recorded
    - 1.3.1.1.2. Structured language information where available
      - 1.3.1.1.2.1. Language Code

- *1.3.2. Reconciliation Process*
  - 1.3.2.1. Data values for language qualified through reconciliation
    - 1.3.2.1.1. Authorized Label
      - 1.3.2.1.1.1. Preferred Label from Wikidata Item

1.3.2.1.2.　　Structured Value

1.3.2.1.2.1.　　QID from Wikidata Item

1.3.2.2.　　Semi-automated reconciliation: Values matched to Wikidata through OpenRefine reconciliation service for Wikidata

1.3.2.2.1.　　Matches made based on semi-structured language string and structured language code

1.3.2.2.1.1.　　If language string is available, those values are reconciled to Wikidata (as the authority)

1.3.2.2.1.2.　　Absent a string, language codes are reconciled to Wikidata

1.3.2.2.2.　　Rationale for Use of Wikidata

1.3.2.2.2.1.　　While ISO codes would be the normal standard, ISO codes are not automatically structured as linked data

1.3.2.2.2.2.　　Additionally, different ISO code systems exist for different types of uses at different levels of granularity

1.3.2.2.2.3.　　Wikidata is structured as linked data and contains links to not only ISO codes but other controlled value vocabularies

1.3.2.2.2.4.　　Linking to Wikidata allows queryable connections to many other different language code systems and controlled vocabularies beyond ISO codes

1.3.3.　　Wikibase Implementation

1.3.3.1.　　Language is represented in DS 2.0 Record as the value of a statement for Language as Recorded (P21)

1.3.3.2.　　Statement for Language as Recorded is qualified by

1.3.3.2.1.　　Language in Authority File (P22) where structured or semi-structured language values are reconciled to Wikidata

1.3.4.　　Associated Wikibase Authority Record

1.3.4.1.　　Type: Language (Authority File): Q12

1.3.4.2.　　Authority Record acts as authority control for annotated/qualified Language as Recorded value

1.3.4.3.　　Label is derived from authorized label from Wikidata

1.3.4.4.　　External Identifier (P42): Wikidata QID

1.4.　　Material

1.4.1.　　From Metadata Record

1.4.1.1.　　Data values extracted from member records

1.4.1.1.1.　　Semi-structured string value

1.4.1.1.1.1.　　Material as Recorded

1.4.2.　　Reconciliation Process

1.4.2.1.　　Data values for material qualified through reconciliation

1.4.2.1.1.　　Authorized Label

        1.4.2.1.1.1.   Label from AAT record

     1.4.2.1.2.   Structured Value

        1.4.2.1.2.1.   URI from AAT record

   1.4.2.2.   Semi-automated reconciliation: Values matched to AAT in OpenRefine using controlled list after data wrangling/cleaning

     1.4.2.2.1.   Rationale for Use of AAT

        1.4.2.2.1.1.   Well-known, established controlled vocabulary used by many cultural heritage institutions

        1.4.2.2.1.2.   Range of materials for manuscripts is well-represented in AAT

1.4.3.   Wikibase Implementation

   1.4.3.1.   Material is represented in DS 2.0 Record as the value of a statement for Material as Recorded (P30)

   1.4.3.2.   Statement for Material is Qualified by

     1.4.3.2.1.   Material in Authority File (P22) where semi-structured material values are reconciled to AAT

1.4.4.   Associated Wikibase Authority Record

   1.4.4.1.   Type: Material (Authority File): Q17

   1.4.4.2.   Authority Record acts as authority control for annotated/qualified Material value

   1.4.4.3.   Label is derived from authorized label from AAT

   1.4.4.4.   External identifier (P44): AAT URI

1.5.   Name / Associated Name

1.5.1.   From Metadata Record

   1.5.1.1.   Data values extracted from member records

     1.5.1.1.1.   Semi-structured string values

        1.5.1.1.1.1.   Associated Name as Recorded

        1.5.1.1.1.2.   In Original Script (name value)

     1.5.1.1.2.   Structured contextual information

        1.5.1.1.2.1.   Source Authority Record Identifier/Link

        1.5.1.1.2.2.   Role

           1.5.1.1.2.2.1.   Author

           1.5.1.1.2.2.2.   Artist

           1.5.1.1.2.2.3.   Scribe

           1.5.1.1.2.2.4.   Former Owner

           1.5.1.1.2.2.5.   Associated Agent

1.5.2.   Reconciliation Process

   1.5.2.1.   Data values for names qualified and enriched through reconciliation

     1.5.2.1.1.   Instance of

        1.5.2.1.1.1.   Human

        1.5.2.1.1.2.   Organization

     1.5.2.1.2.   Authorized Label

        1.5.2.1.2.1.   Preferred Label from Wikidata Item

1.5.2.1.3.   Structured Value
    1.5.2.1.3.1.   QID from Wikidata Item
1.5.2.2.   Semi-automated reconciliation
    1.5.2.2.1.   Name values matched to Wikidata through OpenRefine reconciliation service for Wikidata
        1.5.2.2.1.1.   Rationale for Use of Wikidata
            1.5.2.2.1.1.1.   National/Traditional name authorities are limited in their scope for artists, scribes, and former owners
            1.5.2.2.1.1.2.   National/Traditional name authorities have institutional barriers to adding or editing authority records
            1.5.2.2.1.1.3.   Wikidata provides low barriers to entry to add items, so that our authority work can be done externally to DS and in collaboration with Wikimedians (sustainable)
            1.5.2.2.1.1.4.   Name authority work for DS 2.0 can be limited in scope to the needs of DS project, but linked to Wikidata to leverage other data not found in our authorities
            1.5.2.2.1.1.5.   Wikidata links to national/traditional name authorities, if such data is desired
    1.5.2.2.2.   Role values matched to AAT values at data extraction using controlled list
1.5.3.   Wikibase Implementation
1.5.3.1.   Name is represented in DS 2.0 Record as the value of a statement for Associated Name as Recorded (P14)
1.5.3.2.   Statement for Associated Name as Recorded is qualified by
    1.5.3.2.1.   In Original Script (P13) string value extracted from member metadata record
    1.5.3.2.2.   Role in Authority File (P15) data value extracted from member metadata record reconciled to AAT value
    1.5.3.2.3.   Name in Authority File (P17) where semi-structured name values are reconciled to Wikidata
        1.5.3.2.3.1.   Allows retention of string value from original record but semantically annotated to its corresponding item in the Name Authority
1.5.4.   Associated Wikibase Authority Records
1.5.4.1.   Types
    1.5.4.1.1.   Personal (Name Authority): Q8
    1.5.4.1.2.   Corporate (Name Authority): Q9
    1.5.4.1.3.   Role (Authority File): Q10
1.5.4.2.   Authority Record acts as aggregator for annotated/qualified string values for associated names

            1.5.4.2.1.   Traditional name authority work is then largely outsourced to Wikidata and other linked data authorities to which Wikidata is linked

            1.5.4.2.2.   Allows for streamlined DS 2.0 name records that act as nodes to connect string data values in metadata records to other more robust linked data knowledge bases and name authorities (Wikidata and beyond)

        1.5.4.3.   Label is derived from authorized label of Wikidata item obtained through reconciliation

            1.5.4.3.1.   No alternate labels are included in an Authority Record, because

                1.5.4.3.1.1.   Alternative labels, that is, string values as recorded in metadata records, can be (reverse) queried using SPARQL

                1.5.4.3.1.2.   Where an original string value is "messy" in unstructured legacy data, these "labels" would be inappropriate or confusing to automatically include

                1.5.4.3.1.3.   A potentially infinite number of alternative labels may exist, so it is not feasible to include them for display purposes as if representing an exhaustive list

                1.5.4.3.1.4.   A DS 2.0 Authority Record is not intended to capture exhaustive information about a named entity, but to point to other linked data resources where more information can be found

                1.5.4.3.1.5.   Collecting labels in an DS 2.0 Authority Record would use resources to duplicate the authority work already accomplished by other organizations maintaining authority records, which we can leverage more easily through linked data

        1.5.4.4.   Personal or Corporate identity based on enrichment of data

        1.5.4.5.   Role identity based on extracted data related to association of name to manuscript object

        1.5.4.6.   External Identifiers: (P42) Wikidata QID for names; (P44) AAT URI for roles

1.6.   **Place**

    1.6.1.   From Metadata Record

        1.6.1.1.   Data values extracted from member records

            1.6.1.1.1.   Semi-structured string value

                1.6.1.1.1.1.   Place as Recorded

    1.6.2.   Reconciliation Process

        1.6.2.1.   Data values for places qualified through reconciliation

            1.6.2.1.1.   Authorized Label

                1.6.2.1.1.1.   Label from TGN record

1.6.2.1.2.   Structured Value
1.6.2.1.2.1.   URI from TGN record
1.6.2.2.   Semi-automated reconciliation: Values matched to TGN through OpenRefine reconciliation service for Getty Vocabularies
1.6.2.2.1.   Rationale for Use of TGN
1.6.2.2.1.1.   Well-known, established controlled vocabulary for geographic locations used by many cultural heritage institutions
1.6.2.2.1.2.   Hierarchical structure of TGN allows linking and querying from most granular location known to larger geographic entities
1.6.2.2.1.3.   TGN includes granular representation of geographic and geopolitical entities, including specific locations or buildings, which occur with some regularity in DS data (abbeys, monasteries, estates, etc.)
1.6.2.2.1.4.   TGN includes historical locations which are no longer extant, but contextualized with contemporary location information for the historical site
1.6.2.2.1.5.   TGN differentiates between geographic and political entities as well as contemporary and historical entities for the same coordinates or coordinate boundaries
1.6.3.   *Wikibase Implementation*
1.6.3.1.   Place is represented in DS 2.0 Record as the value of a statement for Production Place as Recorded (P27)
1.6.3.2.   Statement for Production Place as Recorded is qualified by
1.6.3.2.1.   Place in Authority File (P28) where semi-structured place values are reconciled to TGN
1.6.3.2.1.1.   Allows retention of string value from original record but semantically annotated to corresponding items in the Place Authority
1.6.3.2.1.2.   String value may be qualified by multiple Place values if string represents multiple named locations (e.g., the string "Dijon, Burgundy, France" is qualified by links to Authority Records for all three Places individually: Dijon; Burgundy; and France)
1.6.4.   *Associated Wikibase Authority Record*
1.6.4.1.   Type: Place (Authority File): Q16
1.6.4.2.   Authority Record acts to annotate/qualify string values for the presence of production place information
1.6.4.2.1.   Database locally duplicates a skeletal TGN record in Wikibase by linking to the external TGN record for the same geographic entity represented in a DS 2.0 Record

        1.6.4.2.2.   Label is derived from authorized label for TGN entity obtained through reconciliation

        1.6.4.2.3.   External identifier (P44): TGN URI

1.7.   **Subject**

  1.7.1.   From Metadata Record

      1.7.1.1.   Data values extracted from member records

        1.7.1.1.1.   Semi-structured or structured string value

          1.7.1.1.1.1.   Subject as Recorded

            1.7.1.1.1.1.1.   Where possible, named subjects and topical subjects are extracted separately

        1.7.1.1.2.   Vocabulary Designator

        1.7.1.1.3.   Subfield Codes for subject string construction, if applicable

        1.7.1.1.4.   Authority Record Identifier/Link

  1.7.2.   Reconciliation Process

      1.7.2.1.   Data values for subjects qualified through reconciliation

        1.7.2.1.1.   Authorized Label

          1.7.2.1.1.1.   Preferred Label from OCLC FAST record (default)

            1.7.2.1.1.1.1.   For terms derived from FAST, LCSH/LCNAF, MeSH, Répertoire de vedettes-matière

            1.7.2.1.1.1.2.   AAT labels are used for AAT designated values

            1.7.2.1.1.1.3.   RMBS labels are used for RBMS designated values

        1.7.2.1.2.   Structured Value

          1.7.2.1.2.1.   URI from OCLC FAST record (default)

            1.7.2.1.2.1.1.   For terms derived from FAST, LCSH/LCNAF, MeSH, Répertoire de vedettes-matière

            1.7.2.1.2.1.2.   AAT URIs are used for AAT designated values

             1.7.2.1.2.1.3.   RMBS URIs are used for RBMS designated values

        1.7.2.1.3.   Named subjects and topical subjects are reconciled separately

      1.7.2.2.   Reconciliation Service Used: FAST through OpenRefine (https://github.com/cmharlow/fast-reconcile)

        1.7.2.2.1.   Rationale for Use of FAST (default)

          1.7.2.2.1.1.   Subject values in metadata records are often constructed as LC Subject Heading strings

            1.7.2.2.1.1.1.   While LC is the standard used among nearly all institutions for subjects, linked data representation of subjects in the LC Linked Data Service is complicated by

        hierarchical relationships between Subject Headings and Subheadings, which are dependent on rules for constructing valid subject headings

1.7.2.2.1.1.2.   While some subject strings are represented as a single entity in the LC Linked Data Service, not all valid subject strings are present as linked data

1.7.2.2.1.1.3.   Pre-coordinated subject strings are not useful in a dynamic, user-driven faceted search and browse system in which individual subjects can combined as components at the moment of retrieval (post-coordinate)

1.7.2.2.1.1.4.   A faceted or post-coordinate system with equal weight given to subject components found in LC Subject Heading strings would be ideal for representation as linked data

1.7.2.2.1.1.5.   FAST, which is faceted, derived from LC Subject Headings, and structured as linked data, meets such requirements

1.7.2.2.1.2.   FAST links to other relevant authorities and linked databases like LC and Wikidata

1.7.3.   *Wikibase Implementation*

1.7.3.1.   Subject is represented in DS 2.0 Record as the value of a statement for Subject as Recorded (P19)

1.7.3.1.1.   Named subjects and topical subjects expressed in a DS 2.0 Record are Subject as Recorded when uploaded to Wikibase

1.7.3.2.   Statement for Subject as Recorded is qualified by

1.7.3.2.1.   Term in Authority File (P20) where subject strings are reconciled to FAST (or AAT, RMBS)

1.7.3.2.1.1.   Allows retention of string value from original record but semantically annotated to corresponding items in the Term Authority

1.7.3.2.1.2.   String value may be qualified by multiple Subject values if string represents multiple subjects (e.g., the string "Catholic Church--Liturgy--Calendar--Early works to 1800" is qualified by links to Authority Records for all four subject facets/components individually: Catholic Church; Liturgics; Calendar; and Early works)

1.7.4.   *Associated Wikibase Authority Record*

1.7.4.1.   Type: Term (Authority File): Q11

1.7.4.2.   Authority Record acts to annotate/qualify string values for the presence of subject information

1.7.4.2.1.   Database locally duplicates a skeletal FAST/AAT/RBMS record in Wikibase by linking to the external FAST/AAT/RBMS record for the same concept or entity represented in a DS 2.0 Record

1.7.4.2.2.   Label is derived from authorized label for FAST/AAT/RMBS concept or entity obtained through reconciliation

1.7.4.2.3.   External identifier (P44): FAST/AAT/RMBS URI

1.8.   Title

1.8.1.   From Metadata Record

1.8.1.1.   Data values extracted from member records

1.8.1.1.1.   Semi-structured string values

1.8.1.1.1.1.   Title as Recorded

1.8.1.1.1.2.   Uniform Title as Recorded

1.8.1.1.1.3.   In Original Script (both titles value)

1.8.2.   Reconciliation Process

1.8.2.1.   Data values for titles qualified through enrichment

1.8.2.1.1.   Authorized Label

1.8.2.1.1.1.   Standard Title assigned

1.8.2.2.   Manual enrichment: Values standardized through assignment of Standard Titles

1.8.2.2.1.   Rationale for Use of Local Authority

1.8.2.2.1.1.   Not seeking to implement or reproduce a bibliocentric model

1.8.2.2.1.2.   Contributed data does not always support bibliographic representation

1.8.2.2.1.3.   Offers the opportunity to enrich string values with our own authority values to better power faceted browsing, search, and retrieval

1.8.3.   Wikibase Implementation

1.8.3.1.   Title is represented in DS 2.0 Record as the value of a statement for

1.8.3.1.1.   Title as Recorded (P10)

1.8.3.1.2.   Uniform Title as Recorded (P12)

1.8.3.2.   Statement for Title as Recorded is qualified by

1.8.3.2.1.   In Original Script (P13) string value extracted from member metadata record

1.8.3.2.2.   Standard Title (P11) where title strings are reconciled to DS Authority values

1.8.4.   Associated Wikibase Authority Records

1.8.4.1.   Type: Standard Title (Q6)

1.8.4.1.1.   Authority Record acts to annotate/qualify string values for the presence of title information

1.8.4.1.1.1.   Label is derived from DS authority management standards and practices, including deference to existing vocabularies and authorities (Wikidata, FAST, LCNAF)

2. **Manuscript Object and Institutional Holding: Extraction and Publication of Metadata Values**
   2.1. **DS ID**
      2.1.1.   Assigned at ingest when record is uploaded
      2.1.2.   Contains DS identifying information
      2.1.3.   Represents the object in the real world
      2.1.4.   Unique and persistent
      2.1.5.   Linked to Holding
      2.1.6.   Linked from DS 2.0 Record
   2.2. **Holding**
      2.2.1.   Created upon ingest
      2.2.2.   Contains institutionally identifying information
      2.2.3.   Data values extracted from member records or supplied in supplemental documentation
         2.2.3.1.   Shelfmark / Call number
         2.2.3.2.   Institutional ID
         2.2.3.3.   Link to Institutional Record
         2.2.3.4.   Holding status
            2.2.3.4.1.   Current / Non-current
         2.2.3.5.   Start time and End time
         2.2.3.6.   Holding Institution
            2.2.3.6.1.   Qualified/enriched through the same process as Names (Section 1.5)

3. **Name Authority Management**
   3.1. **Distinct names that match items or are verified to exist in Wikidata**
      3.1.1.   Name as Recorded value reconciled to Wikidata item as described in Section 1.5.2.
   3.2. **Distinct names not found/represented in Wikidata**
      3.2.1.   Names not present
         3.2.1.1.   Create Wikidata item
            3.2.1.1.1.   Minimum data / statements needed*
               3.2.1.1.1.1.   QID* (supplied by Wikidata)
               3.2.1.1.1.2.   Label*
                  3.2.1.1.1.2.1.   If more than one label (name variant) exists, choose one label as main label and designate others as "also known as" / aliases
               3.2.1.1.1.3.   Description*

3.2.1.1.1.3.1.   Standard format: [role] of [manuscript name] [ID], [holding institution]

3.2.1.1.1.4.   Instance of statement*

3.2.1.1.1.4.1.   Human or organization (or organization type)

3.2.1.1.1.5.   Occupation statement (human only)

3.2.1.1.1.5.1.   Author, artist, scribe, collector

3.2.1.1.1.5.1.1.   Qualified by assertion "stated in" "Digital Scriptorium Catalog"

3.2.1.1.1.6.   Industry statement (organization only)

3.2.1.1.1.6.1.   Valid value for industry property

3.2.1.1.1.7.   Date statements

3.2.1.1.1.7.1.   Start date statement

3.2.1.1.1.7.1.1.   Date of birth for human only

3.2.1.1.1.7.1.2.   Start date for organization only

3.2.1.1.1.7.2.   End date statement

3.2.1.1.1.7.2.1.   Date of death for human only

3.2.1.1.1.7.2.2.   End date for organization only

3.2.1.1.1.7.3.   Floruit statement

3.2.1.1.1.7.3.1.   Valid for either human or organization

3.2.1.1.1.7.3.2.   Use only if start and end dates unavailable

3.2.1.1.1.8.   External IDs

3.2.1.1.1.8.1.   SDBM*

3.2.1.1.1.8.2.   VIAF

3.2.1.1.1.8.3.   LC

3.2.1.1.1.8.4.   HHML*

3.2.1.1.1.8.5.   WorldCat Entities

3.2.1.1.1.8.6.   ISNI

3.2.1.1.1.8.7.   FnB

3.2.1.1.1.8.8.   Others

3.2.2.   *Names that cannot be disambiguated from existing names*

3.2.2.1.   Create Wikidata item as described in Section 3.2.1.

3.2.2.1.1.   Apply other disambiguating or distinguishing property-value statements to item as necessary

3.2.2.1.1.1.   May need to qualify with references

3.2.2.1.1.1.1.   "Stated in" property takes resource as represented in Wikidata as value (i.e., QID for database or catalog with that information)

3.2.2.1.1.1.2.   "Reference URL" property takes URL for web resource as value (i.e., URL pointing to a webpage with that information)

3.3.   "Messy names" in semi-structured or unstructured data
 3.3.1.   Manual parsing of names
  3.3.1.1.   Human labor to extract names from data string
   3.3.1.1.1.   Apply processes described in this documentation
    3.3.1.1.1.1.   Section 1.5.2. for existing names in Wikidata
    3.3.1.1.1.2.   Section 3.2.1. for names not present in Wikidata
 3.3.2.   Automated parsing of names
  3.3.2.1.   Entity Extraction
   3.3.2.1.1.   Process TBD

4.   Term Authority Management
 4.1.   Currently, entities and concepts occurring which are equivalent or near equivalent in different vocabularies are included in the Term authorities
  4.1.1.   Treated as separate entities/concepts
 4.2.   A plan will be formulated to map equivalents across vocabularies to consolidate Term authorities
  4.2.1.   As a result, a single Term may have multiple External URIs
  4.2.2.   Process TBD

5.   Wikidata Manuscript Item Creation and Management
 5.1.   Mapping process across three different levels of data
  5.1.1.   Data model to data model
   5.1.1.1.   Development of a crosswalk
    5.1.1.1.1.   Source schema: DS data model
    5.1.1.1.2.   Target schema: WikiProjects Manuscripts data model
  5.1.2.   Value vocabulary to value vocabulary
   5.1.2.1.   Data dictionaries mapping equivalents
    5.1.2.1.1.   Source vocabulary: DS authorities
    5.1.2.1.2.   Target vocabulary: Wikidata
  5.1.3.   Existing record to new record
   5.1.3.1.   SPARQL querying DS records
    5.1.3.1.1.   Generate CSV to load into OpenRefine
   5.1.3.2.   Use Wikibase schema in OpenRefine
    5.1.3.2.1.   Match SPARQL-generated CSV to Wikidata items and properties
   5.1.3.3.   Create new Wikidata items and upload statements for manuscript objects based on DS data