DETAILED INSTRUCTIONS FOR CATALOGING SOUND DATASETS

INTRODUCTION

This document outlines procedures for cataloging sound datasets, including remote and direct electronic formats. A sound dataset is a collection of audio data, including musical and nonmusical sound, such as speech, used for natural language processing and analysis.

GUIDELINES FOR FIXED FIELDS   *(Required fields are marked in bold. Instructions in orange are local Yale practices)*

- **Leader:**
  - Type of record (Leader/06):
    - i=Nonmusical sound recording
    - j=Musical sound recording
  - Bibliographic Level
  - Encoding level
  - Cataloging Form: i=ISBD
  - Ex. 1: Nonmusical sound recording

Leader	
Record Status	c : Corrected or revised
Type of Record	i : Nonmusical sound recording
Bibliographic Level	m : Monograph/item
Type of Control	_ : No specific type of control
Encoding Level	_ : Full level
Cataloging Form	i : ISBD punctuation included
Linked Record Requirement	_ : Not specified or not applicable

  - Ex. 2.: Musical sound recording

Leader	
Record Status	c : Corrected or revised
Type of Record	j : Musical sound recording
Bibliographic Level	m : Monograph/item
Type of Control	_ : No specific type of control
Encoding Level	_ : Full level
Cataloging Form	i : ISBD punctuation included
Linked Record Requirement	_ : Not specified or not applicable

- **006 Additional Material Characteristics:**

  m=Computer File
  - Form of Item:
    - o=Online
    - q=Direct electronic *(for CD-ROMs, etc.)*
  - Type of File:
    - h=Sound *(musical and nonmusical sound)*
  - Ex.: Online sound dataset

006 - Additional Material Characteristics (m - Computer File)			
☐ Projected Medium	☐ Nonmusical Sound	☐ Musical Sound	
☐ Notated Music	☐ Manuscript Notated Mu...	☐ Printed Map	☐ Manuscript Map
☐ 3-D Artifact	☐ Continuing Resource	☐ Manuscript Lang.	☐ Books
☐ 2-D Nonprojectable	☑ Computer File	☐ Kit	☐ Mixed Material
Target Audience	_ : Unknown or not specified		
Form of Item	o : Online		
Type Of File	h : Sound		
Govt. Publication	_ : Not a government publication		

- **007 Physical Description:**

  c=Electronic Resource

  - Specific Material Designation:
    - b=Chip cartridge *(for flash drives)*
    - e=Disc cartridge, type unspecified *(for external hard drives)*
    - o=Optical disc *(for CD-ROMs, etc.)*
    - r=Remote *(for online format)*
  - Dimension:
    - g=4 ¾. or 12 cm *(for CD-ROMs, etc.)*
    - n=Not applicable *(for online format)*
    - ☐=No attempt to code *(for external hard drives, USB flash drives, etc.)*
  - Sound on medium:
    - a=Sound on medium
  - Ex. 1: Online format

  - Ex. 2. CD-ROMs, etc., format

- **008 General Description:**
  - Publication Status
  - Date 1/Date 2
  - Place of publication
  - Composition:
    - Identify specific type of music if applicable or
    - nn=Not applicable *(for nonmusical sound)*
  - Format of music:
    - Identify specific type of music if applicable or
    - n=Not applicable *(for nonmusical sound)*

- o Form of Item:
  - ▪ o=Online
  - ▪ q=direct electronic *(for CD-ROMs, external hard drives, USB flash drives, etc.)*
- o Literary Text for Sound Recording:
  - ▪ _=Item is a musical sound recording
  - ▪ Code nonmusical sound recordings appropriately, e.g. t=Interviews
- o Language
- o Cataloging source: d=Other
- o Ex.: Online dataset of interviews

008 - General Description (Music)

Publication Status	s : Single known date/probable date
Date 1 (yyyy)	2018
Date 2 (yyyy)	___
Place of Publication	pau : Pennsylvania
Composition	nn : Not applicable
Format of Music	n : Not applicable
Music Parts	_ : No parts in hand or not specified
Audience	_ : Unknown or not specified
Form of Item	o : Online
Accompanying Matter 1	_ : No accompanying matter
Accompanying Matter 2	_ : No accompanying matter
Accompanying Matter 3	_ : No accompanying matter
Accompanying Matter 4	_ : No accompanying matter
Accompanying Matter 5	_ : No accompanying matter
Accompanying Matter 6	_ : No accompanying matter
Literary Text for Sound Recording 1	t : Interviews
Literary Text for Sound Recording 2	�□ : No attempt to code
Transposition and Arrangement	n : Not applicable
Language	eng : English

GUIDELINES FOR VARIABLE FIELDS   *(Required fields are marked in bold. Instructions in orange are local Yale practices)*

- **040 Cataloging Source:**
  - o   040 // |a CtY |b eng |e rda |c CtY *(original cataloging)*
  - o   040 // |d CtY *(copy cataloging)*
- 041 Language Code:
  - o 041 x/ |a [Language code] |h [Language code of original]
- 043 Geographic Area Code:
  - o 043 // |a [Geographic code]
- 050 LC Call Number:
  - o 050 /4 |a [LC call number]
- **090 Local Call Number:**
  - o 090 // |a yuldset *(add to all datasets)*
  - o 090 // |a yuldsetmediated *(add to all mediated datasets)*
  - o 090 // |a yuldsetsnd *(add to all sound datasets)*

090		‡a yuldset
090		‡a yuldsetsnd

- **1xx Main Entry** *(required when applicable)*
  - o 1xx xx |a [Author or uniform title].
- **245 Title Statement:**
  - o 245 xx |a [Title] dataset : |b [subtitle], |f [inclusive dates of contents] |g [bulk dates of contents if applicable].
- 246 Varying Form of Title:
  - o 246 3/ |a [Variant title] dataset

- **264 Publication Information:**
  - 264 /1 |a [Location of publisher] : |b [Publisher], |c [Date].

264		1 ‡a [Ann Arbor, Mich.] : ‡b [ProQuest Information and Learning Co.], ‡c [between 2002 and 2017?]

    (Note: When deriving a record for an unseen dataset from the database record, bracket the database publication information and assign date range between database publication date and date the dataset was received.)

- **300 Physical Description:**
  - 300 // |a 1 computer disc (5 audio files) ; |c 4 ¾ cm
  - 300 // |a 1 external hard drive (10 audio files)
  - 300 // |a 1 USB flash drive (15 audio files)
  - 300 // |a 1 online resource (approximately 1 million audio files)

    (Note: Include number of files for closed sets if known but NOT for continuing resources.)

- **336 Content Types:**
  - 336 // |a computer dataset |b cod |2 rdacontent *(for all datasets)*
  - 336 // |a performed music |b prm |2 rdacontent
  - 336 // |a sounds |b snd |2 rdacontent
  - 336 // |a spoken word |b spw |2 rdacontent

- **337 Media Types:**
  - 337 // |a computer |b c |2 rdamedia

- **338 Carrier Type:**
  - 338 // |a computer disc |b cd |2 rdacarrier *(for CD-ROMs, etc.)*
  - 338 // |a online resource |b cr |2 rdacarrier *(for online versions)*
  - 338 // |a other |b cz |2 rdacarrier *(for external hard drives, USB flash drives, etc.)*

- 347 Digital File Characteristics:
  - *(Note: Include each type of subfield (|a, |b, |c) in a separate field*
  - Ex.: Online text dataset:
  - 347 // |a audio file |2 rdaft
  - 347 // |b [Encoding format, e.g. MP3, etc.] |b [additional encoding format]
  - 347 // |3 [Materials specified if available, e.g. compressed [specify format if more than one is present], uncompressed [specify format if more than one is present] |c [File size]
  - *(Note: Subfield 3 is not repeatable; use separate fields if necessary; do not include for continuing resources)*
  - Ex.1: Speech dataset on CD-ROM:

300		‡a 1 computer disc ; ‡c 4 3/4 in.
336		‡a computer dataset ‡b cod ‡2 rdacontent
336		‡a spoken word ‡b spw ‡2 rdacontent
337		‡a computer ‡b c ‡2 rdamedia
338		‡a computer disc ‡b cd ‡2 rdacarrier
347		‡a audio file ‡2 rdaft

  - Ex.2: Online sound dataset:

300		‡a 1 online resource
336		‡a computer dataset ‡b cod ‡2 rdacontent
336		‡a sounds ‡b snd ‡2 rdacontent
337		‡a computer ‡b c ‡2 rdamedia
338		‡a online resource ‡b cr ‡2 rdacarrier
347		‡a audio file ‡2 rdaft
347		‡b PCM
347		‡3 Compressed ‡c 9.1 GB

- 500 General Note:
  - 500 // |a Accompanied by ….
  - 500 // |a Documentation in README file.
  - 500 // |a Title devised by cataloger.
- 505 Formatted Contents Note:
  - 505 00 |t [Title 1] dataset, (inclusive dates of contents for Title 1) -- |t [Title 2] dataset, (inclusive dates of contents for Title 2) |t [Title 3] dataset, (inclusive dates of contents for Title 3).
  - *(Note: Add enhanced contents note when the dataset includes data from specific titles, using catalogers' judgement to decide whether the information would be useful.)*
- **506 Restrictions on Access Note:** *(if applicable)*
  - 506 // |a Access restricted by licensing agreement and agreement to terms of use. *(for mediated resources)*
  - 506 // |a Access restricted by licensing agreement. *(for all other licensed datasets)*
- 520 Summary Note:
  - 520 // |a [Summary including keywords (including "sound data mining") and information on geographic, topical, and chronological coverage].
- 538 System Details Note:
  - 538 // |a [System requirements if special software is required; do not make note of standard information or formats].
- **588 Source of Description Note:**
  - 588 0/ |a Description based on [source record, e.g. database, print] record.
  - 588 0/ |a Description based on online resource (viewed [Date]).
  - 588 0/ |a Title from [Location, e.g. file header, README file, etc.] (viewed [Date]).
- **590 Local Note:**
  - 590 // |a Access is available to the Yale community.
- **6xx Subject Access Headings and Genre Terms:**
  - *(Note: when deriving a dataset record from a database record, retain the original subject headings but delete "|v ~~Databases.")~~*
  - Subject headings for Speech Datasets:
    - 600 x0 |a [Individual person] |x Language.
    - 610 x0 |a [Corporate body] |x Language.
    - 650 /0 |a [Individual war] |x Language.
    - 650 /0 |a [Class of person] |z [Geographic location] |x Language.
    - 650 /0 |a [Discipline] |x Language.
    - 650 /0 |a [Ethnic group] |x Language.
    - 650 00 |a [Language] |x Spoken [language, e.g. "Spoken English"] |z [Location].
    - 650 /0 |a [Language, e.g. English language] |x Government jargon.
    - 651 /0 |a [Geographic location, e.g. country, city, etc.] |x Language.
  - Genre/form headings for Speech Datasets:
    - 655 /7 |a Speech corpora. |2 lcgft *(for speech datasets)*
    - 655 /7 |a Data sets. |2 lcgft *(add to all datasets)*
    - 655 /7 |a [Additional genre heading(s) to reflect type of speech, e.g. Interviews, Speeches] |2 lcgft
  - Subject headings for Non-Speech Sound Datasets:
    - 650 /0 |a [Animated films, Motion pictures, Radio broadcasting, Television broadcasting, Theaters, and Video games] ǂx Sound effects.
  - Genre/form headings for Non-Speech Sound Datasets:
    - 655 /7 |a Music. |2 lcgft

- - - 655 /7 |a Sound corpora. |2 lcgft *(for sound datasets)*
    - 655 /7 |a Data sets. |2 lcgft *(add to all datasets)*
    - 655 /7 |a [Additional genre heading(s) to reflect type of sound, e.g. *Animal sounds, Human sounds, Nature sounds, or genre heading for specific type of music*] |2 lcgft
- 7xx Added Entries:
  - 7xx x/ |a [Corporate or personal name].
- **740 Dataset Collection Name** *(required if applicable)*
  - 740 x/ |a [Cataloger supplied title, usually based on the collection title of the data source].
- 76x-78x Linking Entry Fields:
  - 776 Additional Physical Form Entry:
    - 776 08 |i [Relationship information, e.g. Also issued as, Online version, etc.]: |t [Title]. |h [Physical description, e.g. 1 computer optical disc] |w [Record control number]
  - 786 Data Source Entry:
    - 786 08 |i Based on (work): |s [Uniform title]. |t [Title]. |k [Series] |w [Record control number]
  - 787 Other Relationship Entry:
    - 787 08 |i Related work: |s [Uniform title]. |t [Title]. |k [Series] |w [Record control number]

- **856 Electronic Locations and Access:**
  - For unmediated resources:
    - 856 40 |y Online dataset |u [URL]
  - For resources that require staff mediation:
    - 856 40 |y For data access contact researchdata@yale.edu |u mailto:researchdata@yale.edu?subject=data%20access%20inquiry%20from%20Quicksearch
  - Additional link for related resource:
    - 856 42 |3 [Link text of related resource, e.g. Documentation, Manifest, etc.] |u [URL]

    ```
    856  4  0  ‡y For data access contact researchdata@yale.edu ‡u
               mailto:researchdata@yale.edu?subject=data%20access%20inquiry%20from%20Quicksearch
    ```
- **946: Local Data:**
  - 946 // |a DO NOT EDIT. DO NOT EXPORT.

MFHD

- 852: Location:
  - 852 80 |b yulint |h None |z Online resource *(for all online datasets)*

    ```
    852  8  0  ‡b yulint ‡h None ‡z Online resource
    ```
  - 852 80 |b yulintx |h None |z Online resource *(for online format hosted locally)*

    ```
    852  8  0  ‡b yulintx ‡h None ‡z Online resource
    ```
  - 852 00 |b [Location code, e.g. Marx] |h [Classification part] |i [Item part] *(for physical items)*

    ```
    852  0  0  ‡b sml ‡h PJ6715 ‡i .B658 2018 DVD (LC)
    ```
- **583 Action Note** *(Use for materials with digital copies added to Preservica):*
  - 583 0/ |a [Action, e.g. transformed digitally, transferred to optimal storage] |c [Time/date of action] |k [Action agent, e.g. Preservica] |2 [Source of term, e.g. pda] |5 [Institution to which field applies, e.g. CtY]

    ```
    583  0      ‡a transformed digitally ‡c 2019 ‡k Preservica ‡2 pda ‡5 CtY
    ```