

# 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 primarily audio or video data, including speech, used for natural language processing and analysis. The recording may be accompanied by a transcript, which can be analyzed as a text dataset. Projected media is included here for when it is evaluated for its sound characteristics; when evaluating for visual characteristics, see the document on text/image datasets.

## GUIDELINES FOR FIXED FIELDS

*(Required fields are marked in bold. Instructions in orange are local Yale practices)*

- **Leader:**
  - Type of Record: m=Computer file
  - Bibliographic Level:
    - m=Monograph/item *(treat as monograph unless it is specifically noted to be a continuing resource)*
    - i=Integrating resource
    - s=Serial
  - Encoding level: \_=Full level
  - Cataloging Form: i=ISBD

Record Status	c : Corrected or revised
Type of Record	m : Computer file
Bibliographic Level	m : Monograph/item
Type of Control	_ : No specific type of control
Encoding Level	_ : Full level
Cataloging Form	i : ISBD punctuation included
Multipart resource record level	_ : Not specified or not applicable
Length of the length-of-field portion	4 : Number of characters in the length-of-field portion of a Directory entry
Length of the starting-character-position portion	5 : Number of characters in the starting-character-position portion of a Directory entry
Length of the implementation-defined portion	0 : Number of characters in the implementation-defined portion of a Directory entry
Undefined	0 : Undefined

- **006 Additional Material Characteristics:** *(add 006 for type(s) of resource)*
  - g=Projected medium:
    - Form of Item:
      - o=Online
      - q=Direct electronic *(for CD-ROMs, DVD-ROMs, etc.)*
    - Type of Material:
      - f=Filmstrip
      - m=Moving image
      - v=Videorecording i=Non-musical sound
  - i=Non-musical sound recording:
    - Form of Item:
      - o=Online
      - q=Direct electronic *(for CD-ROMs, DVD-ROMs, etc.)*

- j=Musical sound recording:
  - Composition *(for resources limited to one type of music)*
  - Form of Item:
    - o=Online
    - q=Direct electronic *(for CD-ROMs, DVD-ROMs, etc.)*
- s=Continuing resource:
  - Type of Continuing Resource: *(selected)*
    - n=Newspaper
    - p=Periodical
  - Form of Item:
    - o=online *(for online version)*
    - q=direct electronic *(for CD-ROMs, external hard drives, USB flash drives, etc.)*
  - Nature of Entire Work: *(selected)*
    - f=Handbooks
    - o=Reviews
- 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, DVD-ROMs, etc.)*
    - r=Remote *(for online format)*
  - Dimension:
    - g=4 ¾. or 12 cm *(for CD-ROMs, DVD-ROMs, etc.)*
    - n=Not applicable *(for online format)*
    - =No attempt to code *(for external hard drives, USB flash drives, etc.)*

○ Ex. 1: Online format

 007 - Physical Description (c - Computer File)

<input type="checkbox"/> Video Recording	<input type="checkbox"/> Remote Sensing Image	<input type="checkbox"/> Unspecified
<input type="checkbox"/> Kit	<input type="checkbox"/> Notated Music	<input type="checkbox"/> Sound Recording
<input type="checkbox"/> Projected Graphic	<input type="checkbox"/> Microform	<input type="checkbox"/> Nonprojected Graphic
<input type="checkbox"/> Map	<input checked="" type="checkbox"/> Computer File	<input type="checkbox"/> Globe
Specific Material Designation	r : Remote	
Original vs. Reproduction Aspect (OBSOLETE)	_ : (OBSOLETE) Undefined	
Color	<input type="checkbox"/> : No attempt to code	
Dimension	n : Not applicable	
Sound on Medium	<input type="checkbox"/> : No attempt to code	
Image Bit Depth	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> : No attempt to code	
File Format	<input type="checkbox"/> : No attempt to code	
Quality Assurance Target(s)	<input type="checkbox"/> : No attempt to code	
Antecedent/Source	<input type="checkbox"/> : No attempt to code	
Level of Compression	<input type="checkbox"/> : No attempt to code	
Reformatting Quality	<input type="checkbox"/> : No attempt to code	

o Ex. 2. CD-ROM or DVD-ROM

007 - Physical Description (c - Computer File)

Video Recording |  Remote Sensing |  Unspecified |  
 Kit |  Notated Music |  Sound Recording |  Text  
 Projected Graphic |  Microform |  Nonprojected Gra... |  Motion Picture  
 Map |  Computer File |  Globe |  Tactile Material

Specific Material Designation	o : Optical disc
Original vs. Reproduction Aspect (OBS)	_ : (OBSOLETE) Undefined
Color	<input type="checkbox"/> : No attempt to code
Dimension	g : 4 3/4 in. or 12 cm.
Sound on Medium	<input type="checkbox"/> : No attempt to code
Image Bit Depth	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> : No attempt to code
File Format	<input type="checkbox"/> : No attempt to code
Quality Assurance Target(s)	<input type="checkbox"/> : No attempt to code
Antecedent/Source	<input type="checkbox"/> : No attempt to code
Level of Compression	<input type="checkbox"/> : No attempt to code
Reformatting Quality	<input type="checkbox"/> : No attempt to code

• 008 General Description:

- o Publication Status
  - o Date 1/Date 2
  - o Place of publication
  - o Form of Item:
    - o=Online
    - q=direct electronic (for CD-ROMs, DVD-ROMs, external hard drives, USB flash drives, etc.)
  - o Type of File:
    - c=Representational (for moving images)
    - h=Sound
    - (do not use m=Combination)
  - o Language
  - o Cataloging source: d=Other
- o Ex. 1: Online sound dataset

008 - General Description (Computer)

Publication Status	m : Multiple dates
Date 1 (yyyy)	2010
Date 2 (yyyy)	2018
Place of Publication	miu : Michigan
Frequency (OBSOLETE)	_ : No determinable frequency (OBSOLETE)
Regularity (OBSOLETE)	_ : Not applicable (OBSOLETE)
Audience	_ : Unknown or not specified
Form of item	o : Online
Type of File	h : Sound
Govt. Publication	_ : Not a government publication
Language	eng : English
Modified Record	_ : Not modified
Cataloging Source	d : Other

o Ex. 2: Direct electronic resource moving image dataset

008 - General Description (Computer)	
Publication Status	m : Multiple dates
Date 1 (yyyy)	2013
Date 2 (yyyy)	2018
Place of Publication	miu : Michigan
Frequency (OBSOLETE)	_ : (OBSOLETE) No determinable frequency
Regularity (OBSOLETE)	_ : (OBSOLETE) Not applicable
Audience	_ : Unknown or not specified
Form of item	q : Direct electronic
Type of File	c : Representational
Govt. Publication	_ : Not a government publication
Language	eng : English
Modified Record	_ : Not modified
Cataloging Source	d : Other

**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(s)]

• 043 Geographic Area Code:

- o 043 // |a [Geographic code(s)]

• 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

• **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

245	0	0	#a Psychological warfare and propaganda in World War II dataset : #b air dropped and shelled leaflets and periodicals, #f 1939-1945 #g 1942-1945.
246	3		#a Psychological warfare and propaganda in World War Two dataset
246	3		#a Air dropped and shelled leaflets and periodicals dataset

• **264 Publication Information:**

- o *(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.)*
- o 264 /1 |a [Location of publisher] : |b [Publisher], |c [Date].

- **300 Physical Description:** \*\*\**(include number of files for closed sets if known but NOT for continuing resources)*\*\*\*
  - 300 // |a 1 computer disc (5 audio files) ; |c 4 ¾ cm
  - 300 // |a 1 external hard drive (10 video files)
  - 300 // |a 1 USB flash drive (15 video files)
  - 300 // |a 1 online resource (1 million audio files)
  - 300 // |a 1 online resource *(if number of files is unknown or resource is incomplete)*
- **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
  - 336 // |a text |b txt |2 rdacontent *(for transcripts)*
  - 336 // |a three-dimensional moving image |b tdm |2 rdacontent *(for 3D film)*
  - 336 // |a two-dimensional moving image |b tdi |2 rdacontent *(for standard film)*
- **337 Media Types:**
  - 337 // |a computer |b c |2 rdamedia
- **338 Carrier Type:**
  - 338 // |a computer disc |b cd |2 rdacarrier *(for CD-ROMs, DVD-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:** *(include each type of subfield (|a, |b, |c) in a separate field)*
  - 347 // |a audio file |2 rdaft
  - 347 // |a video file |2 rdaft
  - 347 // |a text file |2 rdaft *(for transcripts)*
  - 347 // |b [Encoding format, e.g. CD audio, DVD video, MP3, MP4 etc.] |b [additional encoding format]
  - 347 // |3 [Materials specified if available, e.g. compressed [format], uncompressed [format]] |c [File size]  
*(do not include for continuing resources)*

• Ex.1: Sound 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
347	#b computer disc

• 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.
- 505 Formatted Contents Note:
  - *(Note: add enhanced contents note when the dataset includes data from specific titles)*
  - 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).
- **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:
  - E.g.: “Dataset of materials for sound data mining, etc.”
  - *(Note: include information on geographic, topical, and chronological coverage.)*
- 538 System Details Note:
  - *(Note: use only if special software is required; do not make note of standard information.)*

```
538      |a System requirements: DjView software.
```
- **583 Action Note:**
  - *(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
```
- **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]); title from [Location]
  - 588 0/ |a Description based on [source of description]; title devised by cataloger.
- **590 Local Note:**
  - 590 // |a Access is available to the Yale community.
- **6XX Subject Access Headings and Genre Terms:**
  - Speech Datasets:
    - Provide subject access, e.g.:
      - 650 00 |a Automatic speech recognition.
      - 650 00 |a [Language] |x Spoken [Language].
      - 650 00 |a [Language] |x Phonetic transcriptions *(for transcripts)*
    - Genre/form headings:
      - 655 /7 |a Sound recordings. |2 lcgft *(for recordings)*
      - 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. Human sounds, Interviews, Oral histories] |2 lcgft
  - Other sound Datasets:
    - Provide subject access, e.g.:
      - 650 /0 |a [Organs and regions of the body and individual parts] |x Sounds.
      - 650 /0 |a [Chemicals, materials, and types of materials] |x Acoustic properties.
      - 650 /0 |a Sound effects recordings.

- 650 /0 |a Video games ðx Sound effects.
  - Genre/form headings:
    - 655 /7 |a Sound recordings. |2 lcgft (*for recordings*)
    - 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* ] |2 lcgft
- 710 Corporate Name:
  - 710 X/ |a [Corporate body, including author or publisher. Do NOT include field for a specific Yale Library *location*].
- **740 Dataset Collection Name:** (*required if applicable*)
  - 740 X/ |a [Dataset collection title based on vendor's database collection not YUL location].
 

740	0		a ProQuest historical newspapers dataset collection.
-----	---	--	---
- 76x-78x Linking Entry Fields: (*use catalogers' judgement to add when applicable and useful*)
  - 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]
 

786	0	8	i Based on (work): s Newsday (Nassau edition). t Newsday. b Nassau edition w (OCoLC)ocm05371847
-----	---	---	--
  - 787 Other Relationship Entry:
    - 787 08 |i Related work: |s [Uniform title]. |t [Title]. |k [Series] |w [Record control number]
    - Ex. 1: Dataset record points to database it is related to:
 

245	0	4	a The Louisville courier journal dataset, 1923-2000.
-----	---	---	---

787	0	8	i Related work: t Proquest historical newspapers, Louisville courier journal w (OCoLC)ocm08784449
-----	---	---	--
    - Ex. 2: One newspaper title is split into more than one part:
 

245	0	4	a The Indianapolis star dataset, 1903-1922.
-----	---	---	--

787	0	8	i Related work: t Indianapolis star dataset, 1923-2004 w 14423775
-----	---	---	--
- **856 Electronic Locations and Access:**
  - For unmediated resources with one link:
    - 856 40 |y Online dataset |u [URL]
 

856	4	0	y Online dataset u http://ssrs.yale.edu/data/SSDA/LDC/LDC9356B/
-----	---	---	---
  - For unmediated resources with multiple links:
    - 856 40 |y Online dataset |u [URL]
    - 856 42 |3 [Link text of related resource, e.g. Documentation, Manifest, etc.] |u [URL]
 

856	4	0	y Online Dataset u http://ssrs.yale.edu/data/SSDA/LDC/LDC94513B/
-----	---	---	--

856	4	2	3 Documentation u https://catalog ldc.upenn.edu/LDC94513B
-----	---	---	---
  - For resources that require staff mediation:
    - 856 40 |y For data access contact [researchdata@yale.edu](mailto:researchdata@yale.edu) |u <mailto:researchdata@yale.edu?subject=data%20access%20inquiry%20from%20Quicksearch>

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. (*required for licensed materials*)

## MFHD

- 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)