

MODS XML to CSV

Related documentation:

- 3-Projector design schematic

This document provides the specifications used to transform MODS XML records from the University of Wisconsin Digital Collections site to a CSV for use in the *Broadcasting A/V Data* project. This specification was used to develop [this code](#), which transforms the XML records and writes them to CSV. An example XML record can be found [here](#). (These are available at the bottom of the user-facing pages for program records on UWDC, under "Library Staff Details".)

CSV column name: Identifier

Element name(s): identifier

Path(s): identifier[@type="hdl"]

Example of one value, formatted for CSV: 1711.dl/HQ2PQVZZSVNYL8C

Example of values when repeated, formatted for CSV: should not repeat - unique identifier

Formatting specs: none

CSV column name: Title

Element name(s): nonSort; title; subTitle

Path(s): mods/titleInfo/nonSort; mods/titleInfo/title; mods/titleInfo/subTitle

Example of one value, formatted for CSV: Farm program OR Thanksgiving program 1950 : a chapel in the woods

Example of values when repeated, formatted for CSV: should not repeat

Formatting specs:

- When the <nonSort> element appears, its value should just be concatenated with the <title> value and any other values within the <titleInfo> element without a separator (the nonSort element just blocks characters and words that should be ignored when sorting in alpha order)
 - E.g., these elements: <nonSort>"</nonSort> and <title script="Latn">Cappy and Columbus" and "The Ball Game"</title> should just be concatenated to get "Cappy and Columbus" and "The Ball Game"
- The <subTitle> element value should be concatenated with the <title> element value with a space, colon, and another space
 - E.g., <title script="Latn">Thanksgiving program 1950</title> and <subTitle>a chapel in the woods</subTitle> should become Thanksgiving program 1950 : a chapel in the woods

CSV column name: CPF authorities

Element name(s): namePart; role; name[@valueURI]

Path(s): mods/name/namePart; mods/name/role/roleTerm[@type="text"]; mods/name[@valueURI]

Example of one value, formatted for CSV: "Moore, J.G. (James Garfield)####Speaker\$\$\$https://lccn.loc.gov/n90650257"

Example of values when repeated, formatted for CSV: "WHA (Radio station : Madison, Wis.)####Broadcaster", "Moore, J.G. (James Garfield)####Speaker\$\$\$https://lccn.loc.gov/n90650257"

Formatting specs:

- Separate namePart and role with '####' (if role is present)
- Separate role and valueURI with '\$\$\$' (if valueURI is present)
- Wrap full string in double quotes
- Separate double quote-wrapped strings with comma and space

CSV column name: Date created

Element name(s): dateCreated; dateCreated[@point]

Path(s): mods/originInfo/dateCreated; mods/originInfo/dateCreated[@point]

Example of one value, formatted for CSV: "1947-05-19" OR "1940####start"

Example of values when repeated, formatted for CSV: "1940####start", "1949####end"

Formatting specs:

- Separate dateCreated and dateCreated[@point] with '####' (if point attribute is present)
- Wrap full date string with point if present in double quotes
- Separate double quote-wrapped strings with comma and space

CSV column name: Abstract

Element name(s): abstract

Path(s): mods/abstract

Example of one value, formatted for CSV [example shortened]: Two discussions on the Farm Program...

Example of values when repeated, formatted for CSV: should not repeat

Formatting specs: none

CSV column name: Notes

Element name(s): note

Path(s): mods/note[@type="public"]

Example of one value, formatted for CSV [example shortened]: "[Label transcription] SERIES: Farm Program..."

Example of values when repeated, formatted for CSV [examples shortened]: "[Label transcription] SERIES: Farm Program...", "[Transfer notes] Content clips..."

Formatting specs:

- Wrap each individual note value in double quotes
- Separate double quote-wrapped notes with comma and space

CSV column name: Topical subjects

Element name(s): topic

Path(s): mods/subject/topic

Example of one value, formatted for CSV: "Cows" OR "Cattle—Feeding and feeds"

Example of values when repeated, formatted for CSV: "Cows", "Cattle—Feeding and feeds"

Formatting specs:

- For each subject, the <subject> element repeats, but within a <subject> element, the <topic> element can repeat, or the <namePart>, <geographic>, or <temporal> element may appear. When multiple topic, namePart, geographic, or temporal tags appear within a single <subject>, each value should be concatenated with two hyphens (--). Those strings are considered a single subject (aka coordinated subject)
- Wrap each subject string in double quotes (not the individual values for those that are concatenated)
- Separate double quote-wrapped subjects with comma and space

CSV column name: CPF subjects

Element name(s): namePart

Path(s): mods/subject/name/namePart

Example of one value, formatted for CSV: "Quisling, Vidkun, 1887-1945"

Example of values when repeated, formatted for CSV: "WHA (Radio station : Madison, Wis.)", "Quisling, Vidkun, 1887-1945"

Formatting specs:

- For each subject, the <subject> element repeats, but within a <subject> element, the <namePart> element can repeat, or the <topic>, <geographic>, or <temporal> element may appear. When multiple topic, namePart, geographic, or temporal tags appear within a single <subject>, each value should be concatenated with two hyphens (--). Those strings are considered a single subject (aka coordinated subject)
- Wrap each subject string in double quotes (not the individual values for those that are concatenated)
- Separate double quote-wrapped subjects with comma and space

CSV column name: Geographic subjects

Element name(s): geographic

Path(s): mods/subject/geographic

Example of one value, formatted for CSV: "Madison, Wisconsin"

Example of values when repeated, formatted for CSV: "Madison, Wisconsin", "St. Paul, Minnesota"

Formatting specs:

- For each subject, the <subject> element repeats, but within a <subject> element, the <geographic> element can repeat, or the <namePart>, <topic>, or <temporal> element may appear. When multiple topic, namePart, geographic, or temporal tags appear within a single <subject>, each value should be concatenated with two hyphens (--). Those strings are considered a single subject (aka coordinated subject)
- Wrap each subject string in double quotes (not the individual values for those that are concatenated)
- Separate double quote-wrapped subjects with comma and space

CSV column name: Series

Element name(s): title

Path(s): mods/relatedItem[@type="series"]/titleInfo/title

Example of one value, formatted for CSV: "Let's Find Out"

Example of values when repeated, formatted for CSV: "Let's Find Out", "Rhythm and Games"

Formatting specs:

- Wrap title string in double quotes
- Separate double quote-wrapped series titles with comma and space

Template:

CSV column name:

Element name(s):

Path(s):

Example of one value, formatted for CSV:

Example of values when repeated, formatted for CSV:

Formatting specs: