

# Community Media Archive

## House Style Guide for Metadata

*The shared conventions every steward follows, so 26,000 items a month stay consistent and findable.*

Good metadata answers four questions about a video: **Who? What? Where? When?** The test for every field is simple — does this help someone tell *this* video apart from a similar one, and find it years from now? Place each video in geographic space, in time, and in subject.

**The two rules that prevent real damage:** a non-unique title can break the upload itself, and the first 180 characters of the description are the only part search engines index. Both are explained in full below — read those two first.

### The five required fields

Archive.org requires these five on every item. More than one subject is allowed and encouraged.

- **Title** — unique, descriptive (see card below).
- **URL / source** — where the video came from.
- **Description** — specifics first, boilerplate last.
- **Collection** — which collection the item belongs to.
- **Subject** — one or more tags placing the video in space, time, and topic.

### Field-by-field conventions

#### Title — make it unique

Always include series, episode, or date information so the title identifies *this specific instance*. Two items must never share a title.

- ✓ “City Council Regular Meeting — March 4, 2016”
- ✓ “Garden Club Workshop, Session 3 (2015) — Composting Basics”
- ✗ “City Council Meeting” (which one? collides with every other)

*Why: the YouTube→Archive pipeline uses the title as the identifier. If it sees a title it has seen before, only one details page is created — and which video uploads to it becomes unpredictable. Video and metadata can get out of sync, and duplicate metadata elements can appear. A unique title is what keeps the automated pipeline from corrupting the item.*

#### Description — specifics before boilerplate

Put the episode-specific information **first**, before any generic series boilerplate that repeats across many videos.

- ✓ Starts with what's unique: the episode topic, guests, date, what happened.
- ✗ Starts with the same three-paragraph series blurb on every item.

*Why: search engines index only the first 180 characters of the description (true as of early 2016). If those characters are boilerplate shared across hundreds of videos, none of them are distinguishable to search. Lead with the specifics that make this item findable.*

#### Subject / tags — place the video in space, time, and topic

Enter these as Subject fields or tags. Aim to capture, for each video:

- State, City, and Center Name — both the official long name and the short name or call letters.
- Category and Genre.
- Series information.

Collection-specific fields are encouraged too. For the People’s Media Collection, for example: “Free Library of Philadelphia,” the Philly neighborhood or branch library, the class/workshop session number and year, and the Knight Foundation program grant year.

*Why: keeping all grouping information at the item level (not buried in a website’s interface) means videos can be grouped easily, and an item can be removed or repurposed without losing what tied it to its group.*

## Acronyms — expand them, never reuse them

### Always include the long form

If locals would recognize an acronym, include the spelled-out version too, for searchers who aren’t local.

✓ “KCB (Keep Cincinnati Beautiful)” · “GTEI (Gateway to Economic Inclusion)” · “CCA (Citizen Complaint Authority)”

### Never reuse an acronym within a collection

The same letters meaning two things in one collection makes both unfindable.

✗ “CAC” used for both Cable Advisory Commission and Citizens Advisory Council (School Board) in the same collection.

*Why: archive.org search is literal. An unexpanded or ambiguous acronym either hides the item from non-locals or collides it with unrelated items.*

## Creator / Producer

Credit the individual producer of the video, even though the access-center collection aggregates many producers’ work.

*Why: it gives credit where due and lets someone search all of a producer’s videos across archive.org.*

## Language

Use “English” unless the audio is in another language. For other languages, look up and use the ISO 639 language code.

If any metadata uses accented characters, include an ASCII (unaccented) version of the term as well.

*Why: archive.org search is not multilingual and doesn’t fold accents — the ASCII version keeps the item findable by people typing plain characters.*

## Date — and why Year matters

The “Date Published” field on archive.org will be the date the item was uploaded to YouTube or Vimeo. “Year” is extracted from that Date.

### The backlog exception

This works fine except when a backlog of old videos from prior years is uploaded all at once — then the upload date is wrong for the content. In that case, look for a year fragment in the description or title and use it to populate the Year field correctly.

**Why this matters:** subcollections of the Community Media Archive are faceted by Year. A wrong Year puts the video in the wrong facet, where people browsing by year won't find it.

## Runtime

Express as HH:MM:SS — for example, 00:27:49. Nice to have, not required.

*Why:* viewers can see the length before the player loads and starts playing. If someone has 30 minutes, it helps to know a video runs over two hours before committing.

## License URL

Consider adding Creative Commons licensing information via the licenseurl field.

## Special handling & good practice

- **Opt-outs:** have a process for marking a media maker's video as “not for web distribution,” and a process for removing an item on request.
- **Everything is portable:** archive.org accepts any metadata fields you want to store, and anything you put in (and edit) can be exported back out as JSON or XML. Favor putting information in fields over leaving it implicit.

**Known issue — broadcast\_ready:** adding a collection tag of broadcast\_ready normally tells the Archive to derive an NTSC SD 4:3 version for playback servers. Note this has not worked correctly since a January 2019 Archive regression involving square-pixel vs. rectangular-pixel handling. Don't rely on it without checking the output.

## Before you submit — quick check

- Title is unique and includes series/episode/date.
- Description leads with this video's specifics in the first 180 characters.
- At least one Subject; ideally state, city, center (long + short), category, genre, series.
- Acronyms expanded; none reused within the collection.
- Creator/producer credited.
- Language set (ASCII alternate added if accented characters used).
- Year is correct — double-check if this was part of a backlog upload.
- Runtime as HH:MM:SS if you have it.

---

## References

Community Media Archive: [archive.org/details/community\\_media](https://archive.org/details/community_media)

CMA Wiki: [cma.accesshumboldt.net/wiki](https://cma.accesshumboldt.net/wiki)