

BWF and XMP and ID3 - Oh my!

Building practical software for digital audio metadata

Michael Andrec

Archivist, Ukrainian History and Education Center, Somerset, NJ

m.andrec@UkrHEC.org

NumerousHats

www.UkrHEC.org

Ukrainian-History

UkrHEC

Abstract When digitizing audio recordings in-house, it can be convenient to immediately embed preservation and descriptive metadata in the resulting audio files. Unfortunately, existing open-source tools, such as the de facto standard FADGI BWF MetaEdit, are cumbersome to use. This poster introduces new open-source software developed to simplify the interactive creation and management of embedded metadata in audio files. autoBWF is an alternative GUI for BWF MetaEdit that pre-fills standard Broadcast Wave metadata fields with values derived from file naming conventions, filesystem metadata, and a choice of customizable boilerplate texts. Unlike BWF MetaEdit and exiftool, autoBWF can write XMP tags in addition to BWF and RIFF fields, thereby greatly increasing the descriptive capabilities of the embedded metadata. autoBWF can also pre-populate fields based on existing metadata in another file, thereby reducing or even eliminating the need for manual data entry when creating derivative files. The autoLame tool (a wrapper for the Lame MP3 encoder) can be used to automatically migrate Wave metadata to XMP and ID3v2 tags in MP3 distribution files. This allows for an end-to-end audio metadata pipeline that includes the possibility of export to access and discovery systems like the Oral History Metadata Synchronizer (OHMS) or to PBCore.

The Problem

The Ukrainian History and Education Center's holdings of Ukrainian and Ukrainian American recorded sound are some of the most significant in the world. A substantial fraction are unique recordings on magnetic tape, and a good number date from the 1960s. This makes digitization of these materials a high priority, and metadata management for the resulting digital items is a critical part of the process.

Possible solutions

Digital repository system: overkill (for now)

- Ideal, but needs dedicated server, planning, and \$\$\$ (or IT staff time)
- "Perfect is the enemy of good" - need to start managing metadata *immediately*

PBCore XML sidecar files: good, but cumbersome

- Rich schema designed for sound and moving images
- Sidecars can become "detached"
- Creation and use can be time-consuming without custom software

BWF MetaEdit: limited and frustrating

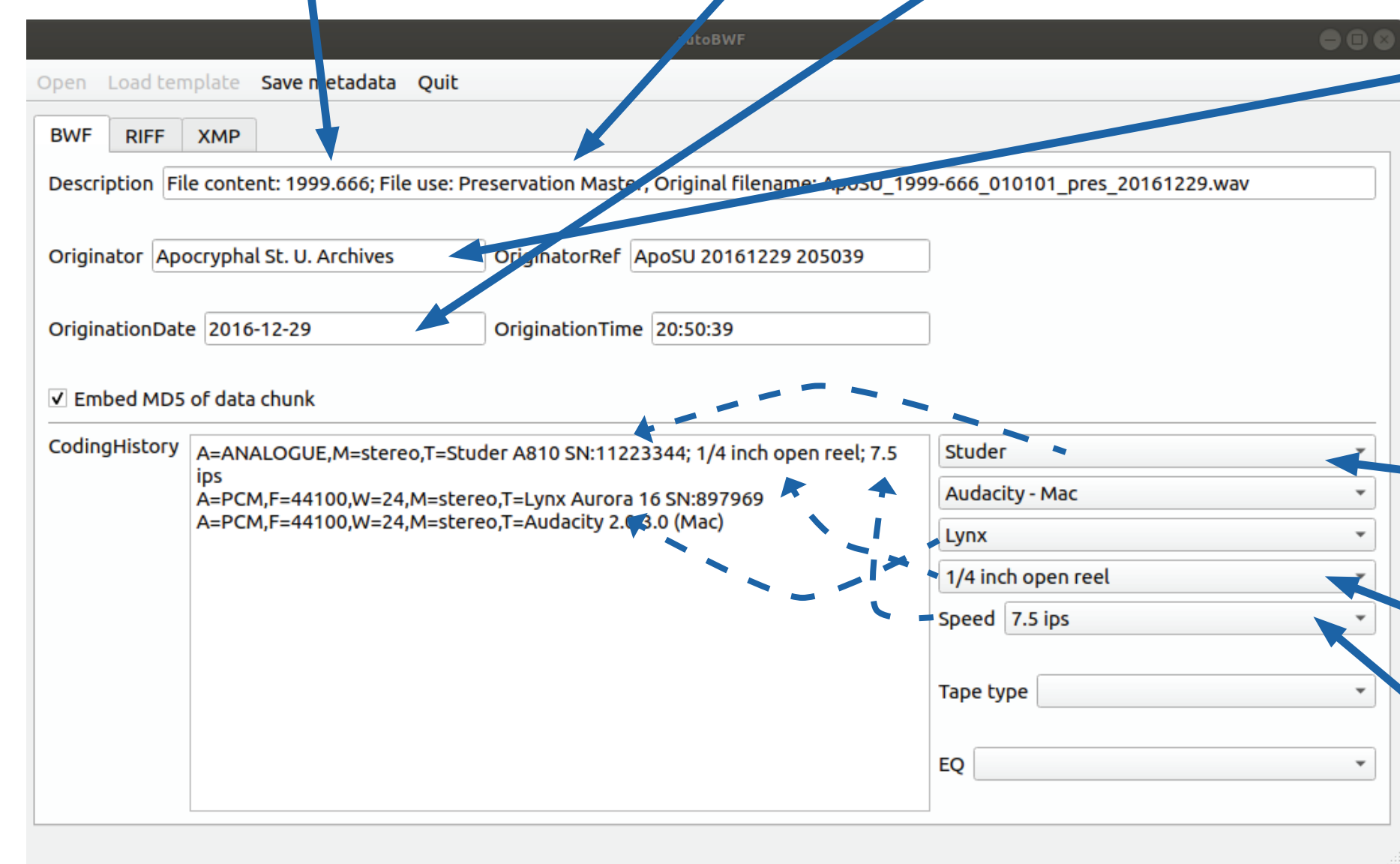
- The "industry standard"
- Lacks sufficient descriptive richness
- GUI can be annoying for interactive production work

Our solution: autoBWF Metadata embedding with a custom GUI tool

BWF metadata can be automatically generated from filename conventions.

The UHEC uses the Indiana University convention described in "Sound Directions" [1]

aposu_1999-666_010101_pres_20161229.wav



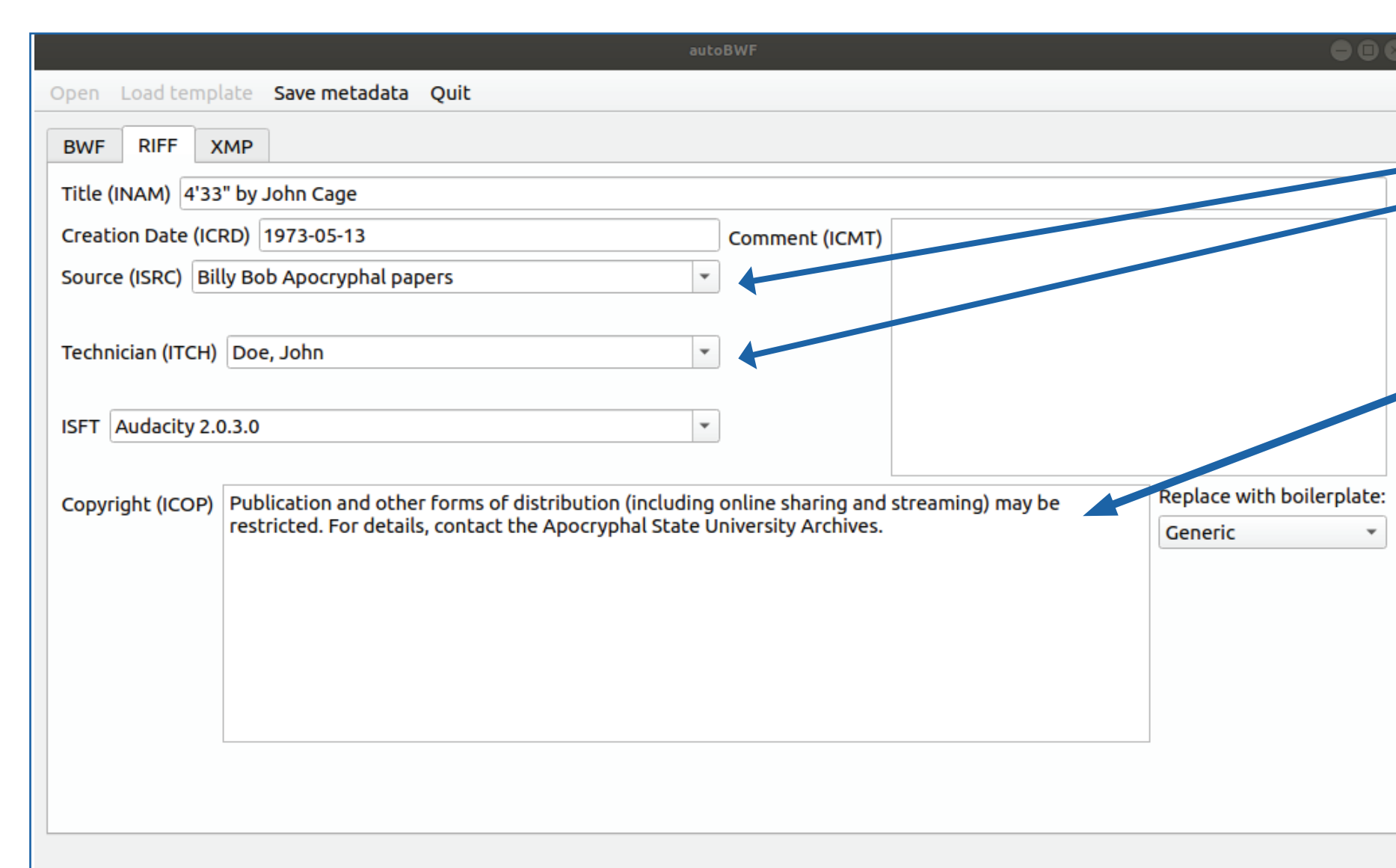
Other metadata is pre-filled from a configuration file.

"originator": "Apocryphal St. U. Archives",

Signal chain metadata is auto-generated using drop-down menus from hardware data in the same configuration file.

```
"deck": {
  "list": ["Studer", "Realistic"],
  "Studer": "Studer A810 SN:11223344",
  "Realistic": "Realistic 909A SN:1234321"
},
"media": [
  "1/4 inch open reel",
  "cassette"
],
"speed": [
  "7.5 ips",
  "3.25 ips"
],
```

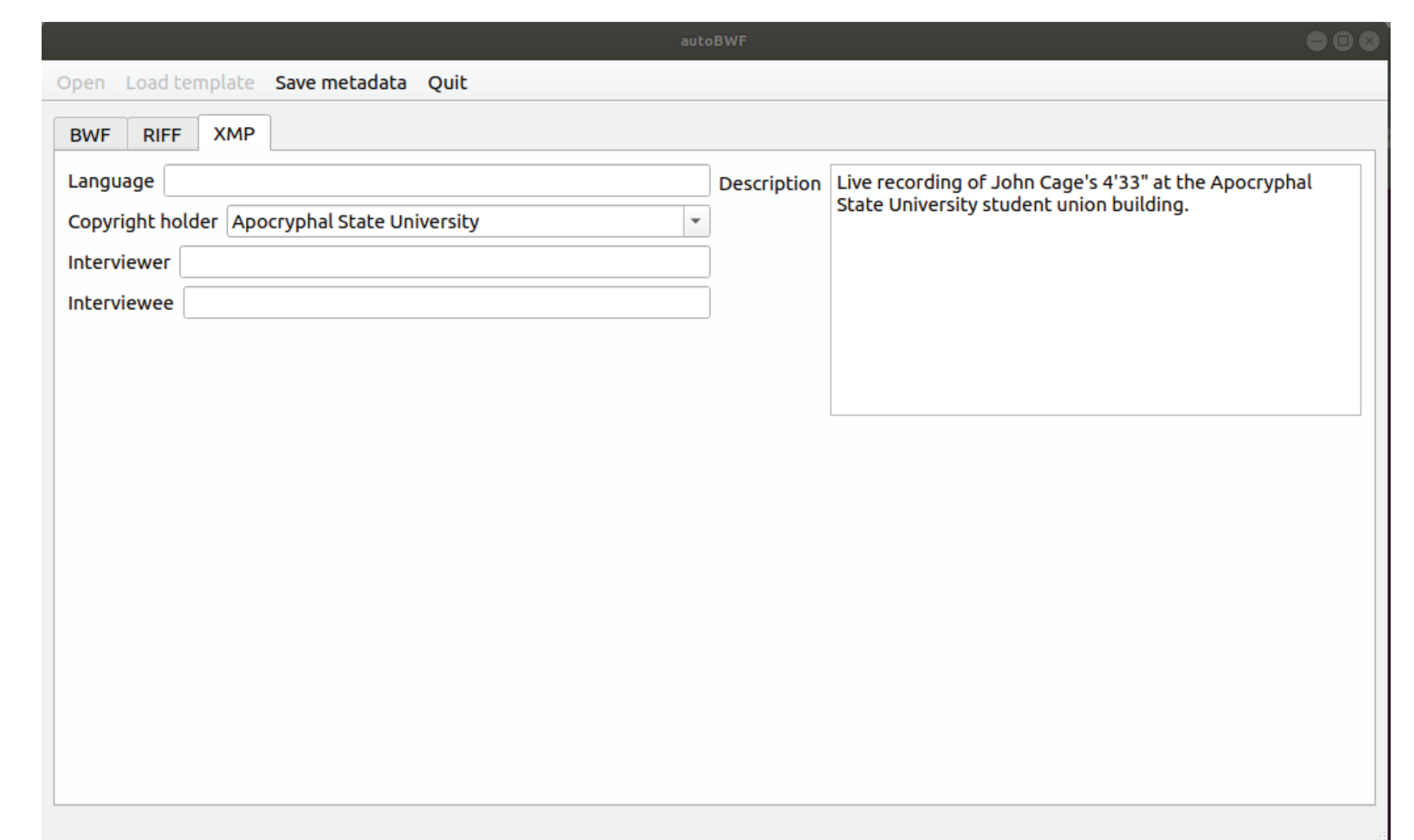
The configuration file also contains default values for Wave RIFF INFO fields.



Boilerplate texts can be pre-filled based on menu selections.

```
"copyright": {
  "list": ["Generic", "CC-BY-SA"],
  "Generic": "Publication and other forms of distribution (including online sharing and streaming) may be restricted. For details, contact the Apocryphal State University Archives.",
  "CC-BY-SA": "This content is copyright by the Apocryphal State University, and is licensed under Creative Commons BY-SA. See https://creativecommons.org/licenses/by-sa/4.0/ for details."
}
```

All menu elements and associated metadata defaults can be customized without having to edit Python code.



BWF and RIFF is supplemented with XMP

XMP ("eXtensible Metadata Platform") was developed by Adobe for their commercial products, but has been publicly released and codified as an ISO standard [2]. Its extensibility could in principle allow the embedding of a significant subset of PBCore. XMP can be embedded in both Wave and MP3 files.

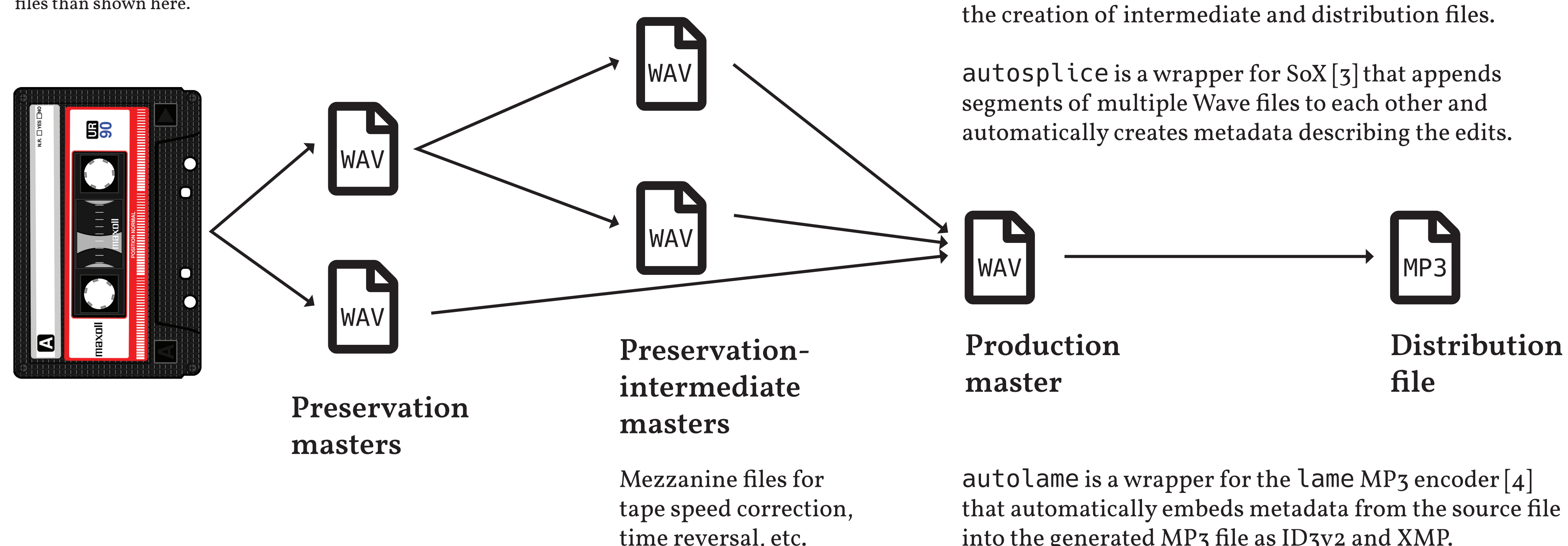
All pre-filled or auto-generated metadata can be manually edited if necessary before saving.

Embedded metadata can be "copied" from file to file, and existing metadata can be edited.

The goal is to be able to export metadata into CSV or XML for migration to discovery systems or a digital repository. A "work in progress" crosswalk between BWF, RIFF, XMP, ID3v2, PBCore, and OHMS is available at www.UkrHEC.org/audio-crosswalk.

The metadata pipeline

Unproblematic transfers typically require fewer files than shown here.



autoBWF is written in Python and PyQt. The source code is available under GPL at the UHEC Github repository (github.com/Ukrainian-History/autoBWF).

References

[1] Mike Casey and Bruce Gordon, *Sound Directions: Best Practices For Audio Preservation* (2007), http://www.dlib.indiana.edu/projects/sounddirections/papersPresent/sd_bp_07.pdf

[2] "File management, metadata integration | Adobe Extensible Metadata Platform (XMP)", Adobe, <https://www.adobe.com/products/xmp.html>

[3] "SoX - Sound eXchange | HomePage", <http://sox.sourceforge.net/>

[4] "LAME MP3 Encoder", <http://lame.sourceforge.net/>