



Metadata Discovery Tools: R7.4 Implementation Specification

Using exiftool and mediainfo to automatically determine file format and codec/container information

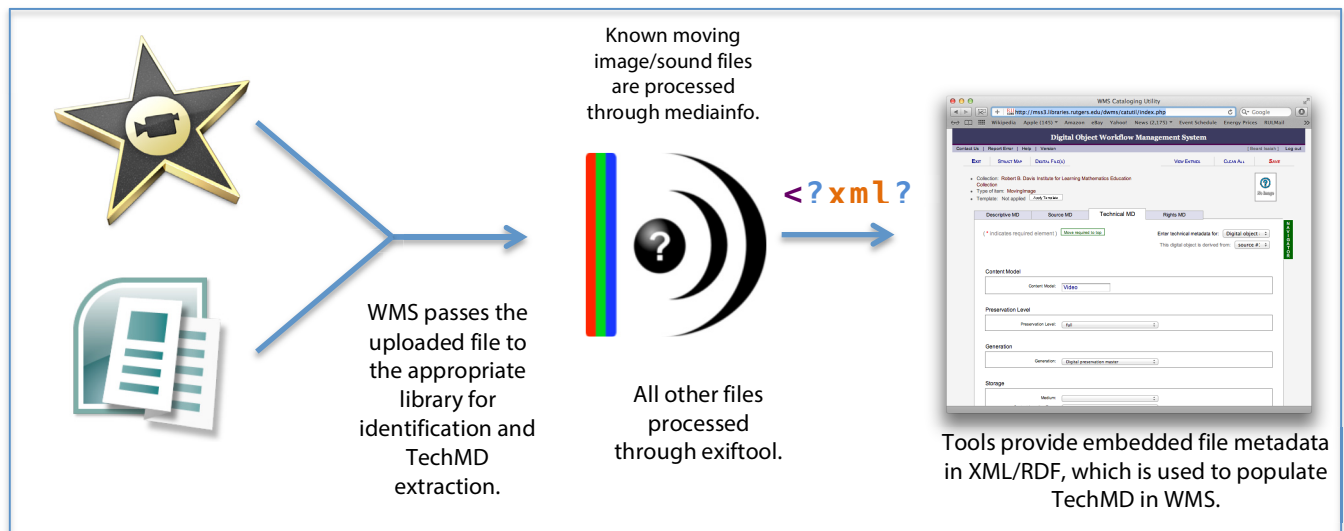
Overview

The use of automated software-based metadata discovery tools for populating technical metadata fields in the RUcore Workflow Management System (WMS) was discussed at the Metadata Working Group meeting on March 17, 2014. This document describes specifications for proposed implementation of these tools in release 7.4.

Technical requirements

A necessary requirement for this implementation is the installation of the most current stable release versions of the exiftool and mediainfo software libraries. These libraries are available as installation packages for SuSE Enterprise Linux¹, or are available as source code and direct installs from the authors' homepages.²

Proposed Operation in WMS



¹ Exiftool for OpenSUSE: <http://software.opensuse.org/package/exiftool>. A identically-named package should exist for SUSE Linux Enterprise. Mediainfo for OpenSUSE: <http://software.opensuse.org/package/exiftool>. Again, an identical package should exist for SUSE Linux Enterprise. <http://mediaarea.net/en/MediaInfo/Download/SLE>. libzen0, libmediainfo0 and the CLI interface will need to be installed; GUI implementation is optional.

² Exiftool: <http://www.sno.phy.queensu.ca/~phil/exiftool/index.html>. Mediainfo: <http://mediaarea.net/en/MediaInfo/Download/Source>. Also for mediainfo: <http://mediaarea.net/en/MediaInfo/Download/SLE>. libzen0, libmediainfo0 and the CLI interface will need to be installed; GUI implementation is optional.

The appropriate discovery tool will be invoked after a successful upload of a digital file into a WMS record:

- Mediainfo: To be used for sound and moving image objects (e.g. AAC, AVI, DV, FLV, M4A, M4V, MP3, MP4, MOV, M2TS, MTS, WAV and similar supported file types).
- Exiftool: To be used for documents, transcripts and still image objects (e.g. DNG, DOC/DOCX, JPEG, JP2, PDF, TIFF, XLS/XLSX).

Post upload, the relevant files will be passed to the appropriate program (exiftool or mediainfo) for processing. The output would then be populated into the appropriate TechMD fields in WMS, using provided metadata mappings.

The aforementioned file types are the proposed supported formats for release 7.4. It is highly probable that we may encounter additional file types going forward. As such, a method to configure exiftool and mediainfo to handle file types with other extensions will be necessary, in this or a future release.

Although this would be an automated process intended to streamline cataloging tasks and assist accuracy and consistency, verification by a cataloger is still a requirement to ensure that the TechMD being received is accurate. Additionally, a metadata manager may choose to add in their own additional information, or modify the fields that were input by the software. For this reason, it is still advisable to keep TechMD as an editable section in WMS.

Discovery Tool Output Display in RUCore Object Record View

During experimentation with mediainfo and exiftool, it was discovered that these tools output metadata information that goes significantly beyond the TechMD elements captured by WMS. Some of these elements may prove useful for RUCore purposes and may be proposed as added fields in a future RUCore release. There may also be fields we choose never to capture for our metadata record, but may still be of interest to some end users.

In light of this, we propose adding a “view exiftool/mediainfo output” option in the View Object interface of RUCore, under the “Complete Record” link. This would invoke a pop-up window which displays the output of the discovery tool used to populate the TechMD fields for the archival datastreams of each object, including fields not captured in the RUCore record.

Additional Implementation Steps

A discussion will need to occur, likely in the Software Architecture Working Group, as to whether and how these software tools have a role in certain RUCore administrative interfaces, such as dlr/EDIT. Additionally, some decisions will need to be made as to whether and how the output of mediainfo and exiftool for each archival datastream would be recorded in FOXML.