

# Open Source Digitization Tools from NARA

Martin Jacobson, Kate Murray, Courtney Egan  
and Jeff Reed

Digitization Services Branch

# Presentation Overview

- Introduction to NARA and Digitization Services Branch (Martin Jacobson)
- Business needs for tool development (Kate Murray)
- Tool demos:
  - AVI MetaEdit (Kate Murray)
  - Media Info (Kate Murray)
  - Video Frame Analyzer (Courtney Egan)
  - File Analyzer (Jeff Reed)
- Wrap up and questions

# Digitization Services Branch (IDS)

- Digitization Services is NARA's primary source of in-house digitization and analog reformatting of motion picture film, audio, video, photographs, aerial film, textual documents, cartographic documents, microfilm, and microfiche.
- Mission statement: to provide agency-wide laboratory services and technical expertise in order to facilitate preservation and access to the content of Federal Records in the custody of NARA

# IDS Quick Look

- Located in College Park, Maryland (Archives 2)
- About 45 staff members
- Four reformatting labs (2 for still images, 1 for audio/video and 1 for motion picture film)
- Development team to support the work of the labs
- Mostly analog-to-digital reformatting (exceptions are film and aerial)

# Business Needs for Tools

- Standardization and consolidation
  - Products and Services project (P&S):  
<http://www.archives.gov/preservation/products/>
- Alignment with NARA's Open Government philosophy (<http://www.archives.gov/open/>)
  - Transparency
  - Participation
  - Collaboration
- If WE need it, maybe OTHERS do too

# Tool Info and Demos

<https://github.com/usnationalarchives>

# AVI MetaEdit

- Built off code and functionality of BWF MetaEdit (created by AVPS and FADGI)
- Batch embed metadata into AVI files in a structured way through existing LIST-INFO RIFF structure using an existing data source or input directly into tool
- Basic QC reports
- Can embed MD5 on video stream only
- Value add: adapted best Coding History for video in the IMIT element
- FADGI working on approving embed guidelines

# reVTMD

- Technical and process metadata schema for reformatted video
- Addresses the needs of cultural heritage communities
- Draws from PBCore and videoMD
- Adds parseable structures to record capture history data which would include critical information such as make and model of playback and capture devices
- Available:  
<http://www.archives.gov/preservation/products/reVTMD.xsd>

# MediaInfo Tool Augmentation

- Popular existing open source tool to view and organize audio and video technical metadata from a wide variety of codecs, formats and wrappers
- Now includes support for view as reVTMD
- Can export the data as a reVTMD schema compliant sidecar XML file

# Video Frame Analyzer

- SAMMA- System for Automated Migration of Media Assets
- SAMMA produces the media files, an XML and a PDF
  - In the XML, about 30 pieces of data are captured 30 times per second
- Tool enables automated QC by analyzing XML files

# File Analyzer

- Java tool
- Framework for actions to perform on files
- Ability to filter file names and types
- Ability to export results
- Ability to compare results

# Use Examples

- Match By Name – did every file transfer?
- Match By Base Name – do we have a JPEG for every TIFF?
- Get Checksum By Name – provide checksum value for file recipients
- Random Sampling MIL 105E – create list of file paths for inspection

# Contact Information

- Martin Jacobson: [martin.jacobson@nara.gov](mailto:martin.jacobson@nara.gov)
- Kate Murray: [kate.murray1@nara.gov](mailto:kate.murray1@nara.gov)
- Courtney Egan: [courtney.egan@nara.gov](mailto:courtney.egan@nara.gov)
- Jeff Reed: [jeffrey.reed@nara.gov](mailto:jeffrey.reed@nara.gov)