

Democracy Now!

Automating Access,
Preservation Workflows for Digital Tape,
and Cataloguing File-based Metadata

David Rice 2007-11-06

Access to New Digital Programming

- Recording
- Encoding
- Delivery / Hosting

Access to New Digital Programming

- Recording

8:00:00 - 8:59:06 a Linux computer runs DVGrab to record the outgoing program as a DVstream.

Specifications: DV-NTSC codec, 720 x 480, stereo 48 kHz 16 bit audio

Access to New Digital Programming

- Encoding (via ffmpeg)

In:
DV media file

Out:

- MP3 (audio podcast)
- flac
- 256 kb MPEG4 (video podcast)
- 1 MB MPEG4 (torrentcast)
- MPEG2

Transcription: Closed Captioning text is captured, divided into segments, send on a rotational schedule to remote volunteer editors, returned, turned into HTML, and put online.

Access to New Digital Programming

- Delivery and Hosting (via FTP and Internet Archive)

| | |
|-------------------------------|------------------------|
| •MP3 (audio podcast) | Internet Archive |
| •flac (lossless audio codec) | Internet Archive |
| •256 kb MPEG4 (video podcast) | Internet Archive |
| •1 MB MPEG4 (torrentcast) | BitTorrent Server |
| •MPEG2 | Internet Archive & DVD |

We also make Ogg Theora and Ogg Vorbis.

Internet Archive makes MPEG I, Flash video, and 64 kb mpeg4

All files then written to LTO3 tape.

Preservation Strategies for Digital Media Tapes

Video: DV tape
DVCam / DVCPro / miniDV

Audio:
Digital Audio Tape (DAT)
Discontinued format used
from mid-1980s to ~2000/



Preservation Strategies for Digital Media Tapes

Goals for Migration of Content from digital tape
to a digital asset management system:

- Retain the media in original codec and specifications with no new encoding or digital generation loss
- Retain not only audio and video but also timecode and content metadata and recording information
- Log the migration of the content from tape, keep track of errors and specifications

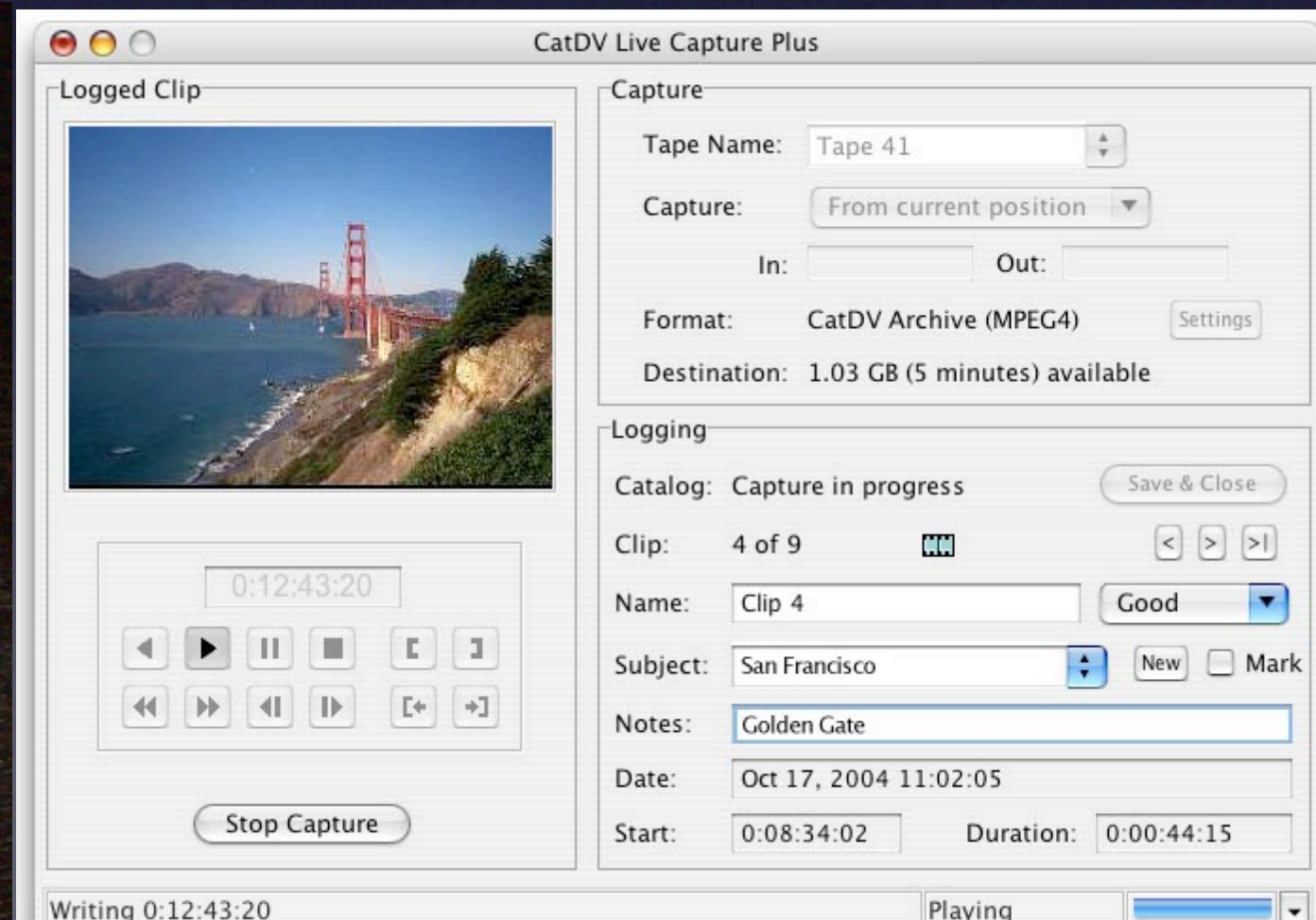
Preservation Strategies for Digital Media Tapes

Video: DV tape

Deck: Sony miniDV Tape Player GV-D1000 (\$800)

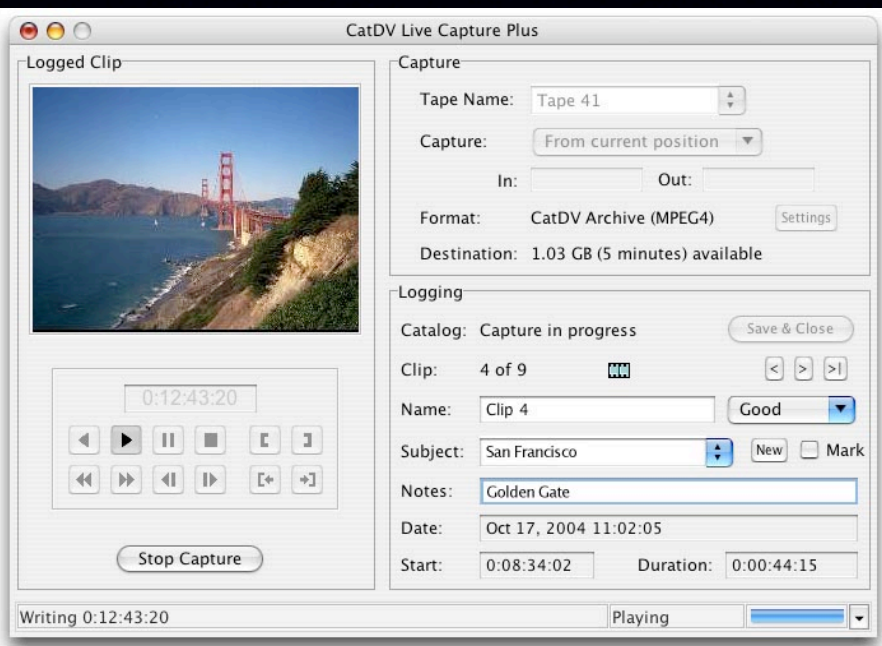
Computer: Mac Mini (\$600)

Software: Live Capture Plus by CatDV (\$49)



Preservation Strategies for Digital Media Tapes

Live Capture Plus



- Ingests miniDV over Firewire
- Retains original timecode and content metadata
- Automatically performs multiple passes over the tape when errors are detected
- Performs ingest as whole tape, in to out, and from current positions

- Automatically detects start/stop markers on the tape and can split resulting files accordingly
- Searches beyond blank portions of tape to capture all data
- Logs all audio and video glitches, dropped frames, and unexpected events
- Can automate post encoding if desired such as MPEG4 derivatives
- From the resulting files the original MiniDV tape could be recreated

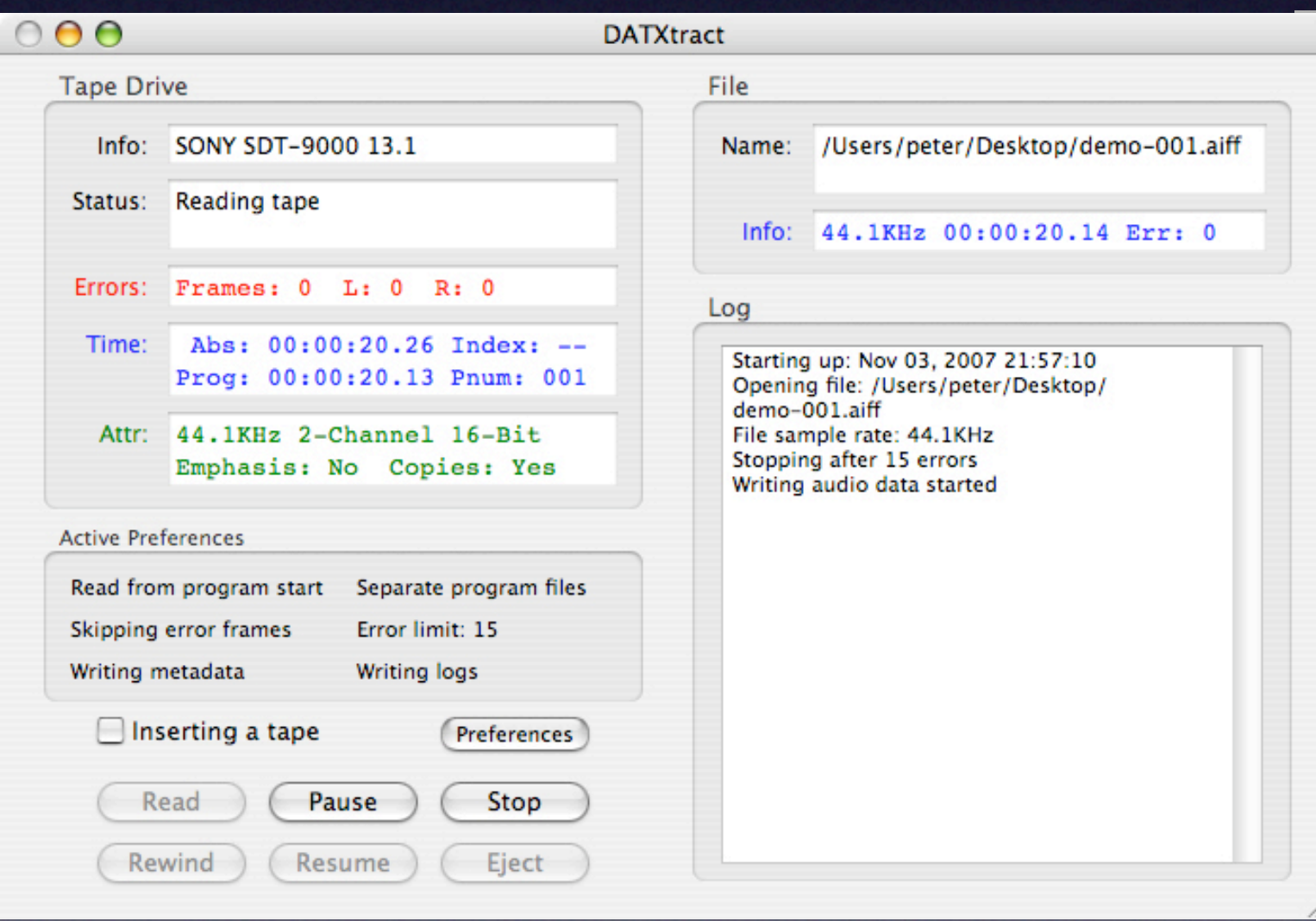
Preservation Strategies for Digital Media Tapes

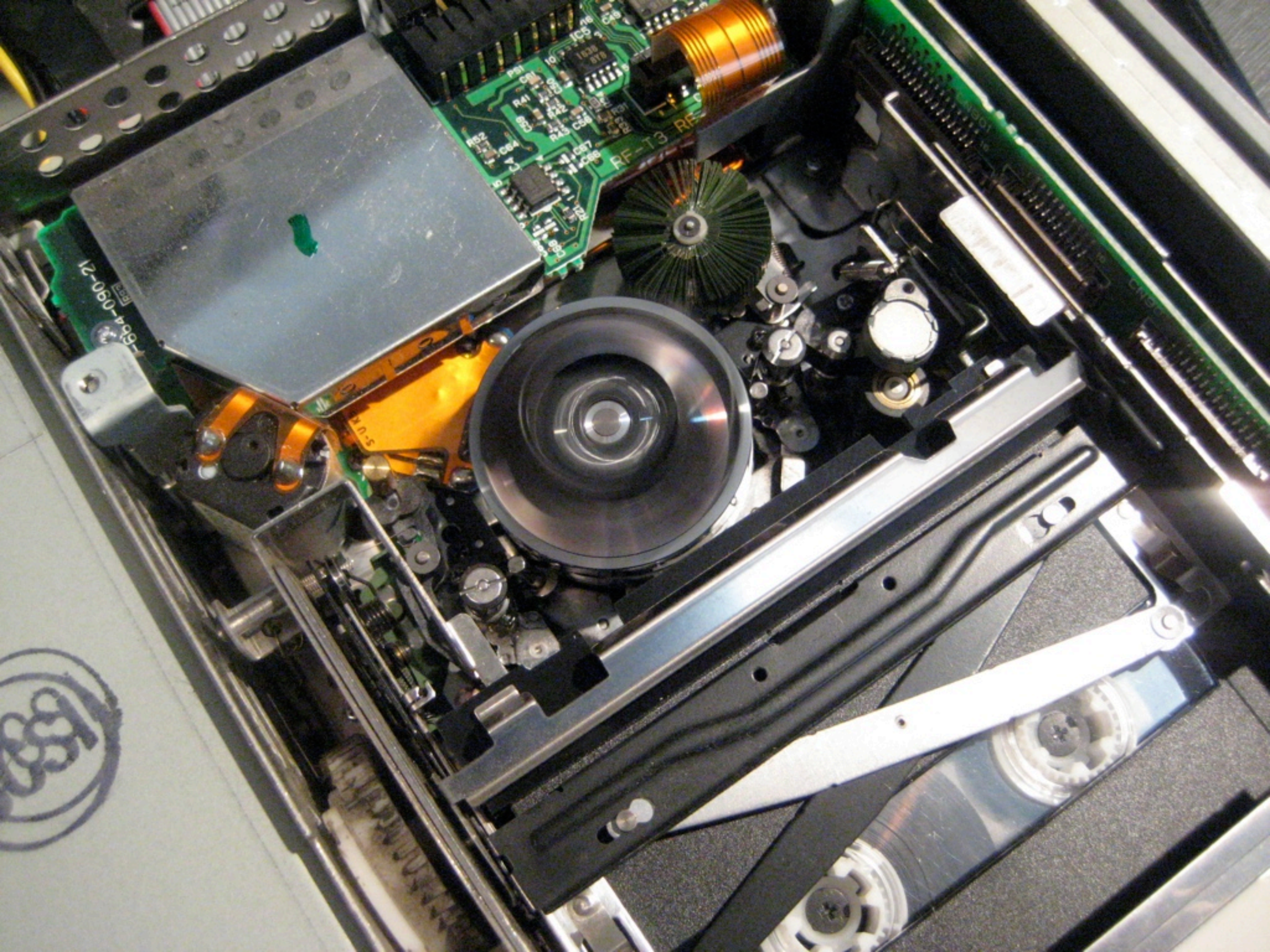
Audio: Digital Audio Tape (DAT)

Deck: Sony SDT-9000 DDS (digital data storage) drive (~\$100)

Computer: Mac G4 with SCSI port (~\$400)

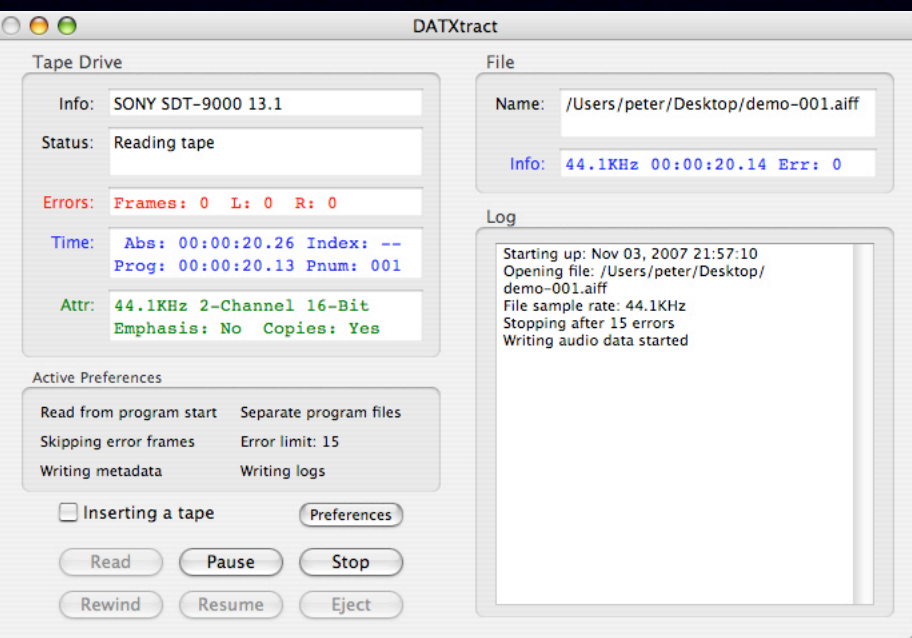
Software: DATXtract 1.3 (open source)





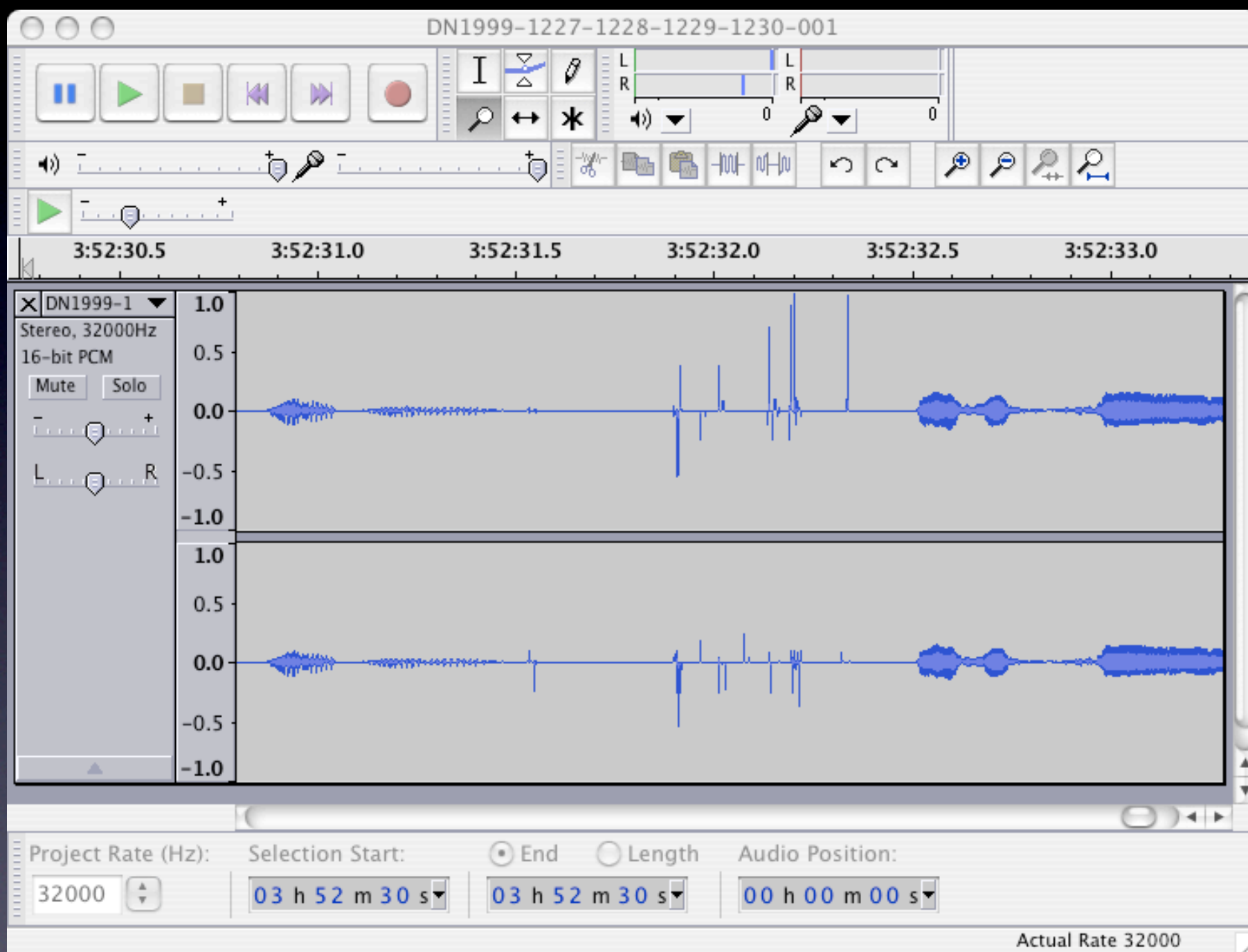
Preservation Strategies for Digital Media Tapes

DATXtract



- Ingest DAT audio over SCSI from Sony SDT-9000 DDS drive
- Retains original timecode and content metadata
- Automatically detects program markers on the tape and can split resulting files accordingly

- Supports digitally extracting audio from 'long-play' 32kHz 12 bit recordings (the majority of Democracy Now's DAT collection)
- Logs audio glitches, error frames, and unexpected events
- Faster than real time audio transfers



DN1999-1227-1228-1229-1230-001

Channels: L
Abs: 01:00:06.04 Prog: 00:00:00.08
File start: 01:58:42.00
File end: 01:58:42.02 (1920 samples)
Including error frame:

Channels: L
Abs: 01:57:01.11 Prog: 00:53:34.29
File start: 03:52:31.17
File end: 03:52:31.19 (1920 samples)
Including error frame:

Channels: L
Abs: 01:57:01.17 Prog: 00:53:35.08
File start: 03:52:31.29
File end: 03:52:31.31 (1920 samples)
Including error frame:

Channels: L
Abs: 01:57:01.18 Prog: 00:53:35.10
File start: 03:52:31.31
File end: 03:52:32.00 (1920 samples)
Including error frame:

Channels: L
Abs: 01:57:01.19 Prog: 00:53:35.12
File start: 03:52:32.00
File end: 03:52:32.02 (1920 samples)
Including error frame:

Channels: L
Abs: 01:57:01.20 Prog: 00:53:35.14
File start: 03:52:32.02
File end: 03:52:32.04 (1920 samples)
Including error frame:

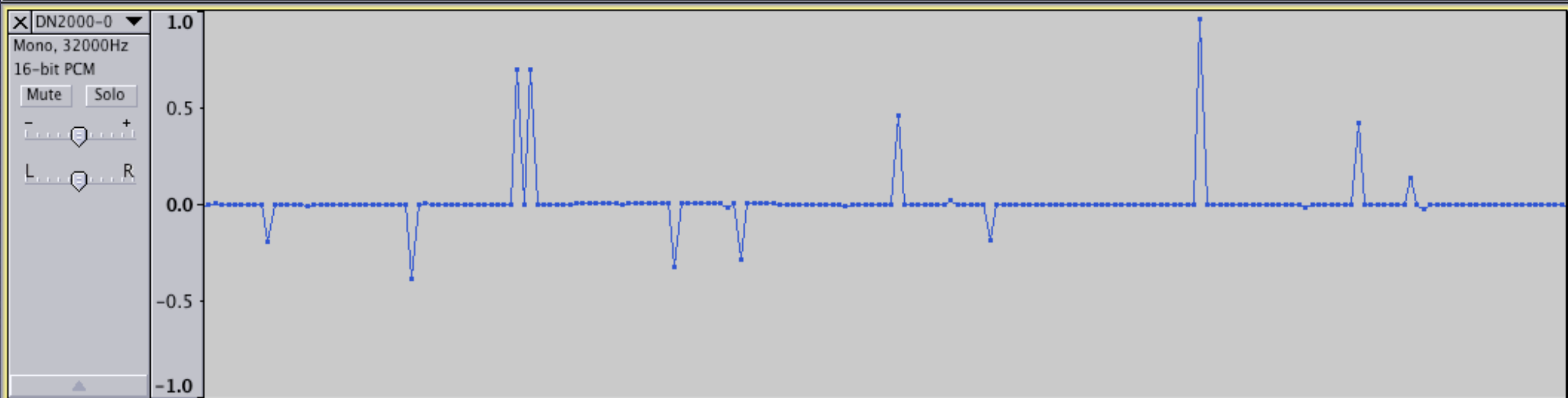
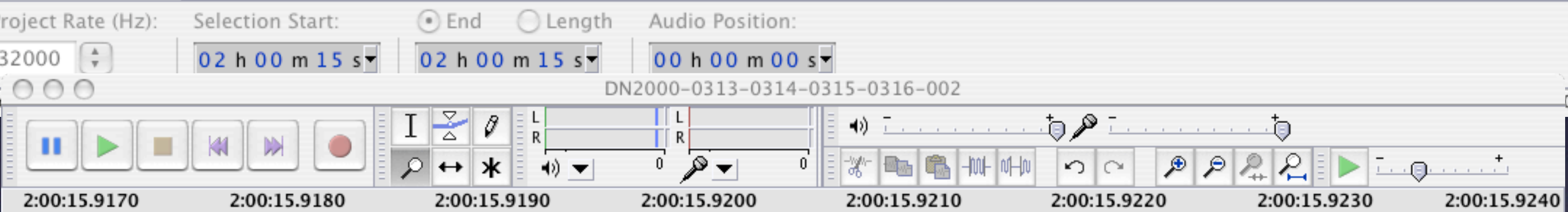
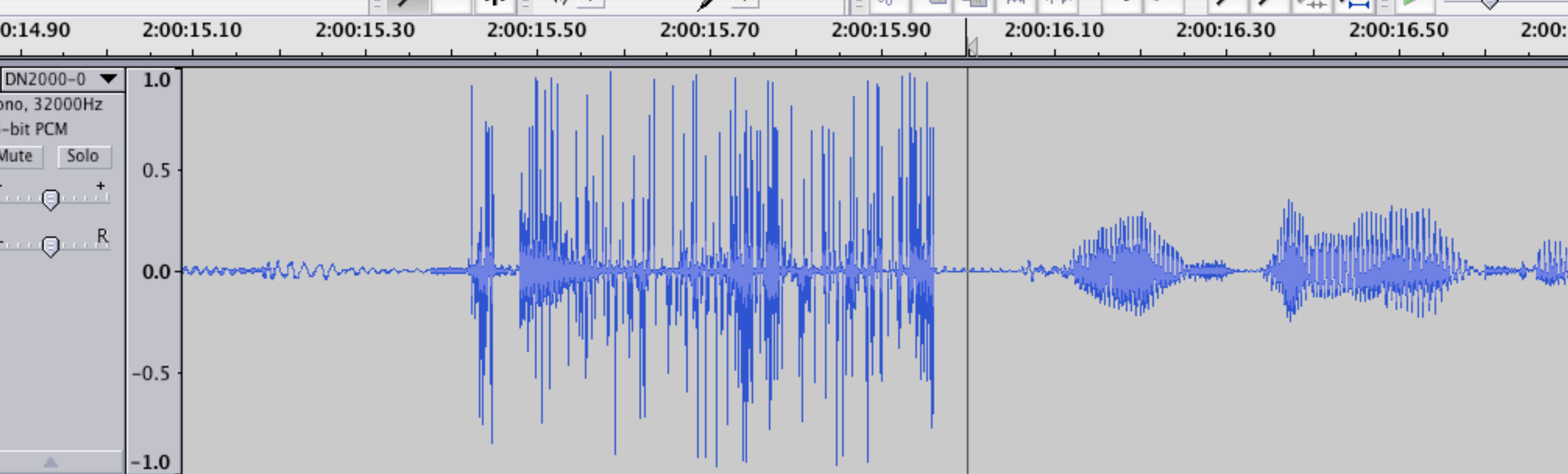
Channels: L
Abs: 01:57:01.21 Prog: 00:53:35.16
File start: 03:52:32.04
File end: 03:52:32.06 (1920 samples)
Including error frame:

Channels: L+R
Abs: 01:57:01.22 Prog: 00:53:35.18
File start: 03:52:32.06
File end: 03:52:32.08 (1920 samples)
Including error frame:

Channels: L
Abs: 01:57:01.24 Prog: 00:53:35.22
File start: 03:52:32.10
File end: 03:52:32.12 (1920 samples)
Including error frame:

Interpolate flags: 80
Abs: --:--:-- Prog: --:--:--
File start: 03:58:34.13
File end: 03:58:34.15 (1920 samples)
Including error frame:

Interpolate flags: 80
Abs: . . . Prog: . . .



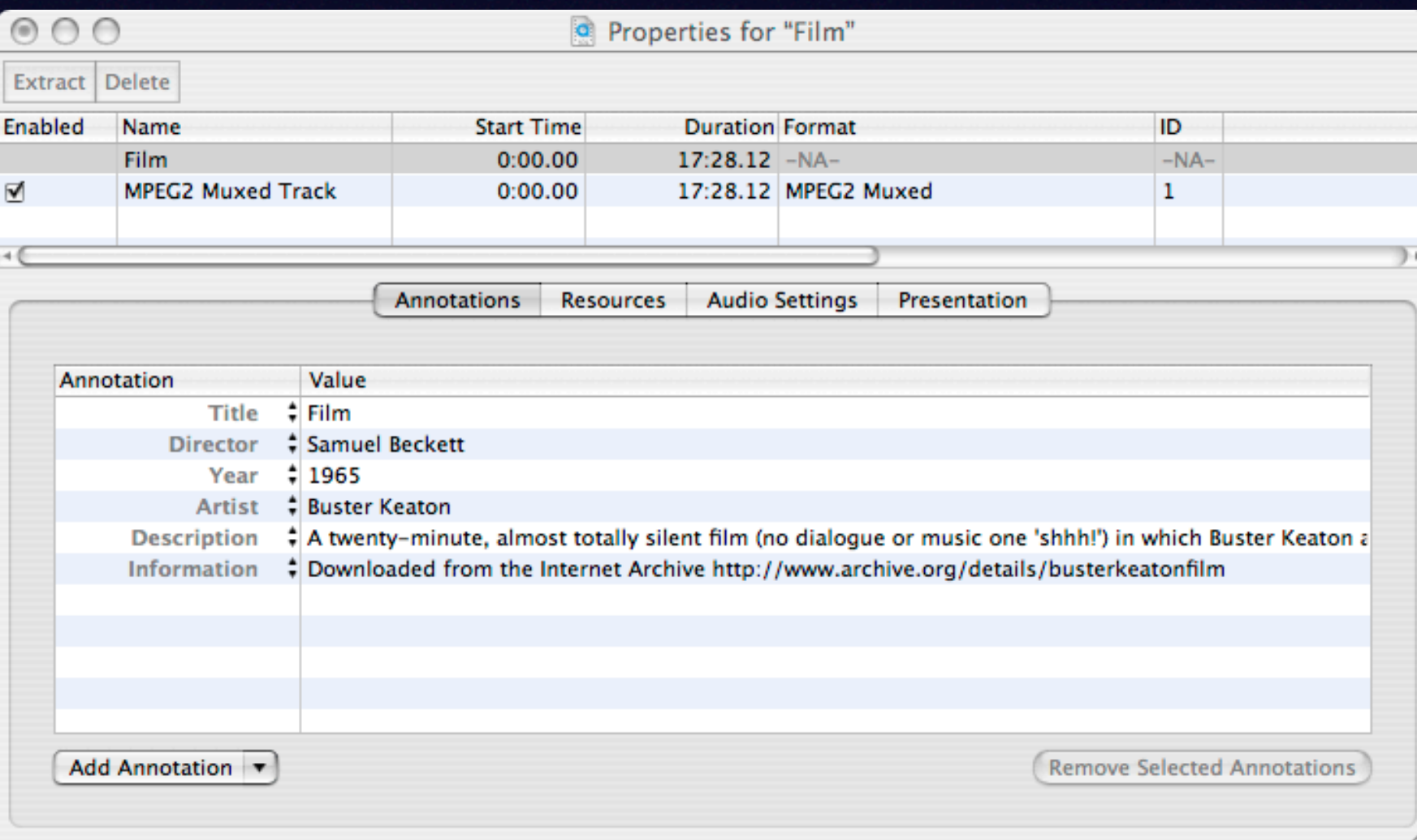
Cataloguing Media While in Production

How can an archive provide cataloguing and digital asset management for files that can be moved, deleted, or renamed by another department?



Cataloguing Media While in Production

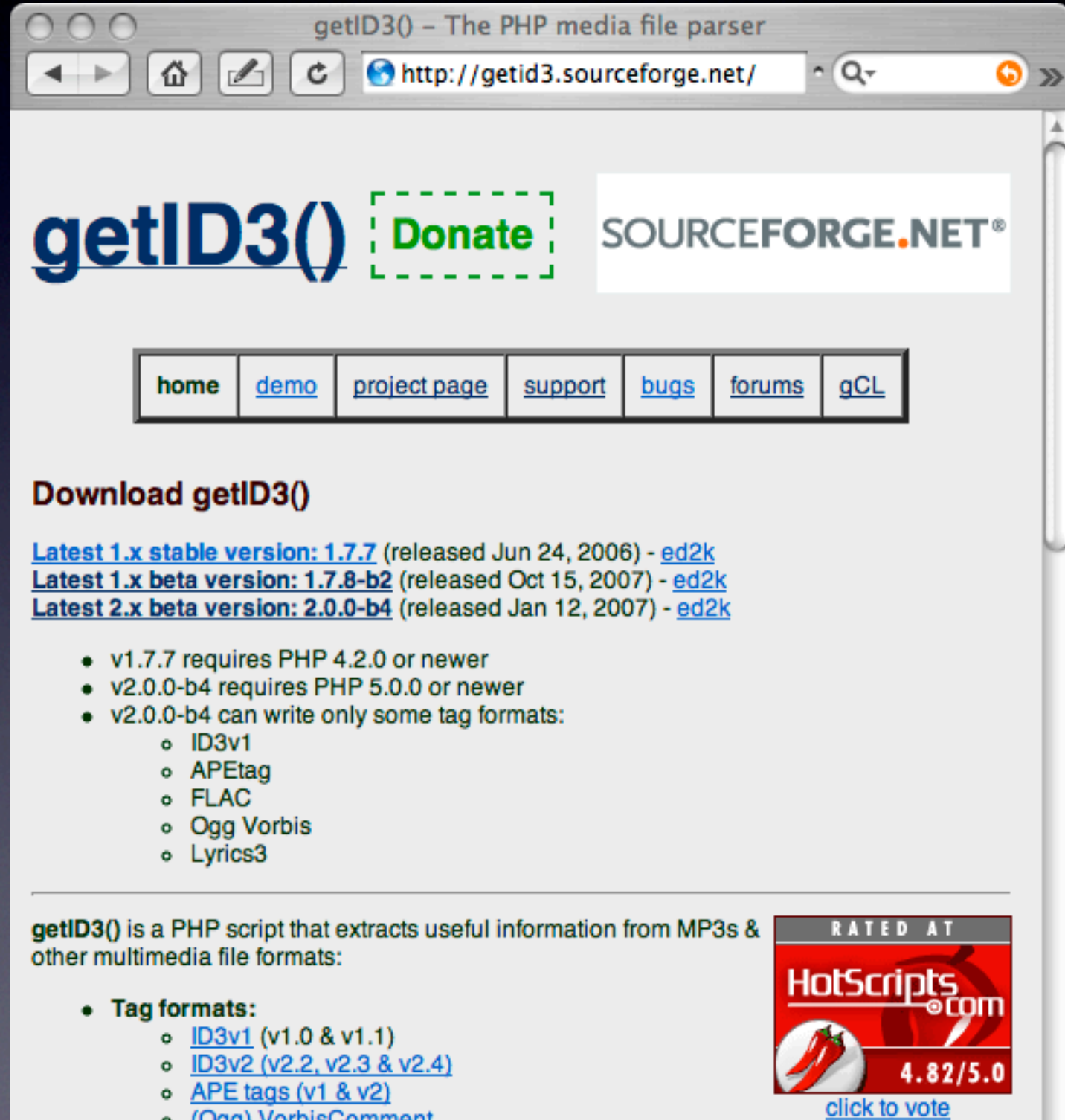
Cataloguing records existing within the media file.



Both production and archive can edit the Quicktime annotation tags of the file, using Quicktime Pro or other programs like Metadata Hootenanny.

Cataloguing Media While in Production

Extracting Quicktime
annotation and other
metadata



The screenshot shows a web browser window titled "getID3() - The PHP media file parser". The address bar displays "http://getid3.sourceforge.net/". The page features the "getID3()" logo, a green "Donate" button, and the "SOURCEFORGE.NET" logo. Below these are navigation links: "home", "demo", "project page", "support", "bugs", "forums", and "gCL".

Download getID3()

[Latest 1.x stable version: 1.7.7](#) (released Jun 24, 2006) - [ed2k](#)
[Latest 1.x beta version: 1.7.8-b2](#) (released Oct 15, 2007) - [ed2k](#)
[Latest 2.x beta version: 2.0.0-b4](#) (released Jan 12, 2007) - [ed2k](#)

- v1.7.7 requires PHP 4.2.0 or newer
- v2.0.0-b4 requires PHP 5.0.0 or newer
- v2.0.0-b4 can write only some tag formats:
 - ID3v1
 - APEtag
 - FLAC
 - Ogg Vorbis
 - Lyrics3

getID3() is a PHP script that extracts useful information from MP3s & other multimedia file formats:

- **Tag formats:**
 - [ID3v1](#) (v1.0 & v1.1)
 - [ID3v2](#) (v2.2, v2.3 & v2.4)
 - [APE tags](#) (v1 & v2)
 - [\(Ogg\) VorbisComment](#)

RATED AT
HotScripts.com
4.82/5.0
[click to vote](#)

Cataloguing Media While in Production

Extracting Quicktime annotation

```
- <data key="5">
- <dict>
  <data key="hierarchy">moov udta</data>
  <data key="name">udta</data>
  <data key="size">867</data>
  <data key="offset">7960</data>
- <data key="subatoms">
- <dict>
  - <data key="0">
  - <dict>
    <data key="hierarchy">moov udta des</data>
    <data key="name">des</data>
    <data key="size">773</data>
    <data key="offset">7968</data>
  - <data key="data">
    Reverend Al Sharpton in Jena, LA: This is not about politics. It s not about black against white.
    protection under the law and justice. We spent about twenty, twenty-five minutes just now with
    heart to see him handcuffed and in leg shackles, but his spirit is high. He has said that he is very
    thousands of people have come into this little town to stand up for him and his five friends. He a
    that he is praying, and, with the grace of God, he will rejoin us soon. But he does not want anyt
    way tomorrow to disparage his name with violence or even a word that is negative. This is about
    This is not about being against anyone.
  </data>
  </dict>
</dict>
</data>
- <data key="1">
- <dict>
  <data key="hierarchy">moov udta key</data>
  <data key="name">key</data>
  <data key="size">52</data>
  <data key="offset">8741</data>
  <data key="data"> ( Al Sharpton, Jena Six, racial inequality</data>
</dict>
</data>
- <data key="2">
- <dict>
  <data key="hierarchy">moov udta WLOC</data>
  <data key="name">WLOC</data>
  <data key="size">12</data>
  <data key="offset">8793</data>
  <data key="location_x">179</data>
  <data key="location_y">272</data>
</dict>
</data>
- <data key="3">
```


Cataloguing Media While in Production

Extracting Quicktime
annotation and other
metadata

```
<data key="fileformat">quicktime</data>
<data key="playtime_seconds">50.1167834501</data>
<data key="video">
- <dict>
  <data key="resolution_x">720</data>
  <data key="resolution_y">480</data>
  <data key="codec">DV/DVCPRO - NTSC</data>
  <data key="bits_per_sample">24</data>
  <data key="lossless"/>
  <data key="pixel_aspect_ratio">1</data>
  <data key="dataformat">quicktime</data>
</dict>
</data>
<data key="comments"></data>
<data key="audio">
- <dict>
  <data key="codec">signed/two's complement (Little Endian)</data>
  <data key="sample_rate">48000</data>
  <data key="channels">2</data>
  <data key="bits_per_sample">16</data>
  <data key="lossless"/>
  <data key="dataformat">quicktime</data>
  <data key="channelmode">stereo</data>
+ <data key="streams"></data>
</dict>
</data>
<data key="bitrate">30316922.1846</data>
<data key="playtime_string">0:50</data>
<data key="warning">
```


Cataloguing Media While in Production

Mapping data from GetID3 to PBCore

| <u>GetID3</u> | <u>PBCORE</u> |
|-------------------------------|-------------------------------|
| creation_time | dateCreated |
| resolution_x resolution_y | formatFrameSize |
| [audio][channels/audio/codec] | formatChannelConfiguration |
| playtime_seconds | formatDuration |
| filesize | formatFileSize |
| codec | formatEncoding |
| mime_type | formatDigital |
| [audio][sample_rate] | formatSamplingRate |
| [audio][bit_depth] | formatBitDepth |
| director, producer, etc | contributor / contributorRole |
| title | title |
| description | description |
| keywords | subject |

| Identifier | QuicktimeAnnotation | Date In Filename | Production Folder | Production Type | FileSize | Duration | Date File Crea |
|--------------------------------|---|---------------------|----------------------|--------------------|----------|----------|----------------|
| 0_Singh_SOT20071031.mov | Keywords: % Manmohan Singh, India, Nuclear | 10/31/07 | DN20071031 | SOT | 156 MB | 0:00:43 | 10/30/07 18: |
| 0_Abbas_SOT20071031.mov | Keywords: 1 Mahmoud Abbas, Palestine, Gaza, | 10/31/07 | DN20071031 | SOT | 71.8 MB | 0:00:19 | 10/30/07 20: |
| 0_RoqueSOT20071031.mov | Keywords: ! Felipe Perez Roque, Cuba, Embargo | 10/31/07 | DN20071031 | SOT | 136 MB | 0:00:37 | 10/30/07 20: |
| 0_VO_Abbas20071031.mov | Keywords: 2 Mahmoud Abbas, Palestine, West | 10/31/07 | DN20071031 | BILLBOARD | 114.6 MB | 0:00:31 | 10/30/07 20: |
| 0_VO_GAZA.mov | Keywords: Gaza, Palestine, Israel | | DN20071031 | VO | 435.7 MB | 0:02:00 | 6/11/07 7:5 |
| 1_Nader_SOT20070709.mov | Keywords: Corprate Power, Ralph Nader | 7/9/07 | DN20071031 | SOT | 1.9 GB | 0:10:33 | 7/8/07 23:41 |
| 1_VO_DemDebate20070604.mov | Keywords: \$ Wolf Blitzer, Democratic Debate, | 6/4/07 | DN20071031 | VO | 309.8 MB | 0:01:25 | 6/4/07 6:4 |
| 0_VO_SCHIP20070724.mov | Keywords: Children, Babies, kids | 7/24/07 | DN20071031 | VO | 221.3 MB | 0:01:01 | 7/23/07 19:0 |
| 0_Tanner_SOT20071015.mov | Keywords: ' John Tanner, Department of Civil | 10/15/07 | DN20071031 | SOT | 131.4 MB | 0:00:36 | 10/14/07 18: |
| 0_VO_DeathPenalty20071030.mov | Keywords: C Capital Punishment, death penalty, | 10/30/07 | DN20071031 | VO | 161.9 MB | 0:00:44 | 10/29/07 17: |
| 1_Break_Animal_Cruelty.mov | Keywords: # PETA, animal rights, animal cruelty | | DN20071031 | BREAK | 591.1 MB | 0:02:43 | 10/30/07 19: |
| 0_VO_Cars20070625.mov | Keywords: Traffic, cars | 6/25/07 | DN20071101 | VO | 283.2 MB | 0:01:18 | 6/25/07 7:1 |
| 0_Mccormack_SOT20071101.mov | Keywords: > Sean McCormack, State Department, | 11/1/07 | DN20071101 | SOT | 87.2 MB | 0:00:24 | 10/31/07 18: |
| 0_Haiti_SOT20071101.mov | Keywords: * Tropical Storm Noel, Port-au-Prince, | 11/1/07 | DN20071101 | SOT | 84.7 MB | 0:00:23 | 10/31/07 18: |
| 0_Carlos_SOT20071101.mov | Keywords: 5 Madrid, Bombings, Juan Carlos Vives | 11/1/07 | DN20071101 | SOT | 71 MB | 0:00:19 | 10/31/07 18: |
| 1_carbonBREAK20071101.mov | Comment: "Nature Song" Sweet Honey in the | 11/1/07 | DN20071101 | BREAK | 934.1 MB | 0:04:18 | 10/31/07 20: |
| B1813Tele.mov | Information: NSA, Telecom, phones, MCI, ATT, | | DN20071101 | BROLL | 988.7 MB | 0:04:35 | 9/27/07 18:1 |
| 0_VO_Cars20070625.mov | Keywords: Traffic, cars | 6/25/07 | DN20071101 | VO | 283.2 MB | 0:01:18 | 6/25/07 7:1 |
| 0_BurmaHRWSOT20071101.mov | Keywords: , David Scott Mathieson, Burma, | 11/1/07 | DN20071101 | SOT | 98.3 MB | 0:00:27 | 10/31/07 17: |
| 0_Maung_SOT20071101.mov | Keywords: Muang Maung, Burma, protest | 11/1/07 | DN20071101 | SOT | 107.7 MB | 0:00:29 | 11/1/07 6:3 |
| 0_Church_SOT20071101.mov | Keywords:) Homophobia, Topeka, Church, Albert | 11/1/07 | DN20071101 | SOT | 61.4 MB | 0:00:16 | 11/1/07 6:5 |
| Broll_VenezuelaClashes.mov | Description: Tear Gas being used by police at a | | DN20071102 | BROLL | 756.9 MB | 0:03:29 | 11/1/07 22:4 |
| 0_VO_SCHIP20070724.mov | Keywords: Children, Babies, kids | 7/24/07 | DN20071102 | VO | 221.3 MB | 0:01:01 | 7/23/07 19:0 |
| B1813Tele.mov | Information: NSA, Telecom, phones, MCI, ATT, | | DN20071102 | BROLL | 988.7 MB | 0:04:35 | 9/27/07 18:1 |
| 2_Break-MoneySOT20070912.mov | Description: Song Theme: I want Money! Video | 9/12/07 | DN20071102 | BREAK | 545.1 MB | 0:02:30 | 9/11/07 23:5 |
| 0_Bhutto_SOT20071105.mov | Keywords: % Benazir Bhutto, Pakistan, marshal | 11/5/07 | DN20071105 | SOT | 56.9 MB | 0:00:15 | 11/4/07 20:1 |
| 0_Granier_SOT20071105.mov | Keywords: . Tobasco, flood, global warming, | 11/5/07 | DN20071105 | SOT | 77.5 MB | 0:00:21 | 11/4/07 19:4 |
| 0_Bush_SOT20071105.mov | Keywords: Bush, Gallup Poll | 11/5/07 | DN20071105 | SOT | 81.4 MB | 0:00:22 | 11/4/07 18:4 |
| 0_VO_FtJack20071105.mov | Keywords: + Troops, Fort Jackson, applause, | 11/5/07 | DN20071105 | VO | 128.3 MB | 0:00:35 | 11/4/07 18:4 |
| 0_Ramirez_SOT20071105.mov | Keywords: ' Rafael Ramirez, Chavez, referendum | 11/5/07 | DN20071105 | SOT | 75.5 MB | 0:00:20 | 11/4/07 20:0 |
| 0_VO_Floods20071105.mov | Keywords: K Mexico southern gulf coast, | 11/5/07 | DN20071105 | VO | 293.4 MB | 0:01:21 | 11/4/07 18:4 |
| 0_Perez_SOT20071105.mov | Keywords: = Oscar Perez, Chavez, Caracas, | 11/5/07 | DN20071105 | SOT | 64 MB | 0:00:17 | 11/4/07 19:5 |
| 0_Cheney_SOT20071105.mov | Keywords: + Dick Cheney, Venezuela, Peru, | 11/5/07 | DN20071105 | SOT | 121.7 MB | 0:00:33 | 11/4/07 18:4 |
| 0_Displaced_SOT20071105.mov | Keywords: E Faduma Hassan Geelle, Mogadishu, | 11/5/07 | DN20071105 | SOT | 95.4 MB | 0:00:26 | 11/4/07 20:0 |
| 0_VO_PakistanBlast20071019.mov | Keywords: ' Prime Minister Benazir Bhutto, | 10/19/07 | DN20071105 | VO | 387.6 MB | 0:01:47 | 10/19/07 6: |
| 0_VO_Bhutto20071022.mov | Keywords: ' Prime Minister Benazir Bhutto, | 10/22/07 | DN20071105 | VO | 270 MB | 0:01:14 | 10/21/07 17: |
| 3_Algiers_SOT20071105.mov | Keywords: Torture, "The Battle of Algiers" | 11/5/07 | DN20071105 | SOT | 862.7 MB | 0:03:58 | 11/5/07 7:4 |
| 1_VO_LahoreArrests.mov | Keywords: F Lahore demonstrations, marshal law, | | DN20071105 | VO | 152.6 MB | 0:00:42 | 11/5/07 8:2 |
| 1_Riceok_SOT20071105.mov | Keywords: & Condoleezza Rice, Pakistan, Mashal | 11/5/07 | DN20071105 | SOT | 149.2 MB | 0:00:41 | 11/5/07 8:2 |
| 105-GUATEMALA-CLOSINGVOTE.mov | Keywords: Alvaro Colom, Guatemala election | | DN20071105 | VO | 448 MB | 0:02:03 | 11/5/07 8:2 |
| 1_PakistanSecurity20071105.mov | Keywords: * Pakistan barricades, Mussharaf | 11/5/07 | DN20071105 | | 612.5 MB | 0:02:49 | 11/5/07 8:3 |
| B0707IraqResistance.mov | Keywords: Iraq resistance, occupation, war | | DN20071105 | BROLL | 1.5 GB | 0:08:31 | 5/29/07 23:0 |
| 0_VO_SriLanka20070807.mov | Keywords: Sri Lanka, civil war, L.T.T.E. | 8/7/07 | DN20071105 | VO | 168.7 MB | 0:00:46 | 8/7/07 6:13 |

Thanks

GetID3

<http://getid3.sourceforge.net/>

Live Capture Plus

<http://pdicamillo.org/~peter/datxtract/>

DATXtract

<http://www.squarebox.co.uk/lcplus.html>