(Web) Archiving
Online Media

Nicholas Taylor
Web Archiving Service Manager
Digital Library Systems and Services

Stanford Media Group
December 10, 2014
overview

• services for media archiving
• web archiving use case
• web archiving mechanics
• technical challenges
• approaches for online media archiving
Services for

MEDIA ARCHIVING
SUL media archiving

services

• analog and digital reformatting
• long-term preservation
• discovery and referenceability
• access interfaces and research tools

scope

• broad range of library collections
  – institutional legacy
  – creative works
  – oral histories, documentary
  – research outputs
  – lectures, events
  – broadcasts and podcasts
  – rare commercial works
SUL media archiving

services
- analog and digital reformatting
- long-term preservation
- discovery and referenceability
- access interfaces and research tools

scope
- broad range of library collections
  - institutional legacy
  - creative works
  - oral histories, documentary
  - research outputs
  - lectures, events
  - broadcasts and podcasts
  - rare commercial works
SUL web archiving

services

• one-time and repeating collection
• long-term preservation
• discovery and referenceability
• access interfaces and research tools

scope

• select SU sub-sites
  – institutional legacy
  – compliance
  – scholarly outputs
• third-party content
  – government information
  – scholarly inputs
SUL web archiving

services
• one-time and repeating collection
• long-term preservation
• discovery and referenceability
• access interfaces and research tools

scope
• select SU sub-sites
  – institutional legacy
  – compliance
  – scholarly outputs
• third-party content
  – government information
  – scholarly inputs
Web Archiving

USE CASE
whole greater than sum of parts
web archiving not always best approach
video or webpage?

PayPal

Check out faster for more holiday cheer.

Sign Up for Free

Own a business? Open a business account

Every purchase with PayPal enters you into our holiday sweepstakes with $100k in total prizes. See rules | Learn more

https://www.paypal.com/home#Pause
web page composed of files
collect and store w/ metadata
repeat over time

Web Page

Collect Data

Web Archives

Constituent Files
make accessible via SWAP

Web Page → Collect Data → Stanford Web Archive Portal → Index Data → Web Archives

Constituent Files
make accessible via SWAP
Web Archiving Online Media

TECHNICAL CHALLENGES
easy: linked video files
Steve Jobs' 2005 Stanford Commencement Address (with intro by President John Hennessy)
YouTube: “Steve Jobs’ 2005 Stanford Commencement Address (with intro by President John Hennessy)”
capture challenges

• obfuscated or short-lived links
• ties up crawler, harming crawl quality
• hard to delimit scope
• few objects, big data volume

“I have the light in my hands” by Ashley Campbell under CC BY 2.0
replay challenges

- relating capture context
- re-inserting embeddable players
- platform-specific engineering
Approaches for

ONLINE MEDIA ARCHIVING
mix-and-match approaches

media archiving

- none
- master
- derivatives

web archiving

- none
- exclude derivatives only
- low-resolution version only
- highest-resolution version
- best effort
archiving approach considerations

• technical **affordances**
• **cost** and budget
• **formats** of extant media
• **nature** of work
• what **else**?
thank you!

Nicholas Taylor
ntay@stanford.edu

“stanford dish at sunset” by Dan under CC BY-NC-SA 2.0
University Archives Use Cases

• Media Only
Philosophy Talk

Philosophy Talk is a talk radio program co-hosted by Stanford professors John Perry and Ken Taylor. The program deals both with fundamental problems of philosophy and with the works of famous philosophers, especially as these relate to our contemporary, day-to-day lives. Some of the wide range of topics of past programs include terrorism, Bush's doctrine of pre-emptive self-defense, Descartes, genetic engineering, and virtue. The program airs from the studios of KALW, 91.7 FM, Information Radio, San Francisco, California, and is produced by Ben Manilla Productions, Inc. on behalf of Stanford University, as part of its Humanities Outreach Initiative.
Knowledge and inspiration, one entrepreneur at a time.

Stanford University’s Entrepreneurship Corner offers 3000 free videos and podcasts, featuring entrepreneurship and innovation thought leaders. Start exploring now.

Latest Videos

Popular Videos
Popular Speakers

Browse Topics:
Creativity & Innovation
Opportunity Recognition
Product Development
Marketing & Sales
Finance & Venture Capital
Leadership & Adversity
Team & Culture
Globalization

Nailing the Hard Things [Entire Talk]
Ben Horowitz | Andreessen Horowitz
Nov. 19, 2014 | Views: 5138 | Duration: 52:53

Consumers and Brands in the Digital Age [Entire Talk]
Tina Wells | Buzz Marketing Group
Nov. 12, 2014 | Views: 1088 | Duration: 55:33

Seeking the Full Potential of Education [Entire Talk]
Jennifer Carolan | New Schools Seed Fund
Nov. 05, 2014 | Views: 1713 | Duration: 55:57
Deposit specs

For Digital AV Content

Preparing to deposit digital audio and video in the Stanford Digital Repository (SDR)? Read these guidelines and use these tools to facilitate the process. Have questions? Write us at sul-media-lab@lists.stanford.edu.

Describing Digital Audio and Video for the SDR

Includes general guidelines and information about required and optional metadata fields as well as available templates.

Preparing Digital Audio and Video Content for the SDR

How to select, name, encode, organize and deliver media files to the SDR.

Documenting Your Digital Video Production

For media producers to provide detailed technical information that will be useful for optimal content preservation and delivery.
Describing Digital Audio and Video Content for the SDR

STANFORD UNIVERSITY LIBRARIES

Describing Digital Audio and Video Content for the Stanford Digital Repository:
General Guidelines and Recommendations

This document provides guidelines for creating information (metadata) about digital audio and video content to be archived in the Stanford Digital Repository (SDR) for preservation and access. It is intended for use by members of the Stanford University community.

By following these guidelines, you help to ensure that your digital audio and video content can be reliably discovered by users searching for relevant content on the web.

For guidelines about preparing audio and video content files for deposit in the SDR, please see:

General Guidelines

1. The goal is to provide description for a discrete unit (or "object") that may be archived, which is not necessarily each and every file that is part of that unit.
2. Be consistent in your metadata: develop patterns and conventions for expressing the description and apply them as uniformly as possible. Consistency makes for better searching and simplifies making changes if necessary.
3. Avoid abbreviations, especially for departments, programs, buildings, etc.
4. Aim to record, at minimum, the "Required Descriptive Elements" listed below. These represent the most commonly used fields on campus, and are considered the minimal required descriptive information to ensure future access to your materials.
5. There may be cases where additional descriptive information not listed below would also be appropriate to include, or where some of this information is not known. Please consult with the University Archives archivist@stanford.edu or the SDR Service Team sdr-contact@library.stanford.edu if you have any questions.

September 2014
Preparing Digital Audio and Video Content
for the Stanford Digital Repository:
General Guidelines and Recommended Technical Specifications

This document provides guidelines for digital audio and video content to be deposited in the Stanford Digital Repository (SDR) for preservation and access. The intended audiences for this information:

- Donors of collection material to Stanford University Archives, the Stanford Archive of Recorded Sound, and Stanford University Libraries Department of Special Collections,
- Members of the Stanford community who use SDR services, including SDR Online Deposit.

The information is relevant whether you are planning to record new content or preparing existing digital content. It is organized into a set of general recommendations followed by a series of technical specifications for both audio and video.

By following these guidelines, you help to ensure to the long-term interoperability and usability of your digital audio and video content preserved in the SDR.

For guidelines about describing the content (creating metadata), please see http://www.stanford.edu/group/tdl/SDR/MetadataGuidelines.pdf.

General Recommendations:
1. At a minimum, plan to archive the final edited production master of your project. Deposit of the raw material - unedited camera master tapes or files - is also recommended.
2. When preparing new content to be archived, choose the highest quality settings appropriate for your project and your recording device. Avoid high compression whenever possible.
3. If your device does not record directly to a file, physical formats - whether tape (MiniDV, DigiBeta, audio cassette, etc.) or disc-based (XDCAM Professional Disc, DVD, CD, etc.) - are acceptable but not preferred.
4. Content on SD cards or internal camera memory should be delivered as is: do not alter the directory structure or file names.

September 2014
Contact the University Archives

Daniel Hartwig
dhartwig@stanford.edu

Josh Schneider
josh.schneider@stanford.edu