Indexing Multilingual Content with OHMS

Teague Schneiter | Brendan Coates
Sr Manager OHP | Sr Archivist OHP
Academy of Motion Picture Arts & Sciences

About OHP

Founded in 2012

Produce, collect, preserve, make accessible stories of filmmakers, craftspeople, and technologists from Hollywood & worldwide

174 interviews produced in-house to date - 672 hours of raw footage

8 languages | 13 countries | 47 film crafts |

2050ish interviews in our collection - over 4000hrs of raw footage

MICD | MPOHDA

Agenda

1. Intro to OHMS
2. OHMS setup
3. OHMS metadata
4. Indexing
   a. as a big group
   b. in small groups
5. Extending OHMS

OHMS intro

Who?

-University of Kentucky Louie B. Nunn Center for Oral History

What problems is it trying to solve?

-Long-form video isn't very user friendly!

-Archival OH collections are really complicated!

OHMS in action - PST/LALA

Academy OHP PST site: pstlala.oscars.org

Pacific Standard Time LA/LA Cultural Festival

Context for OHMS

20% of our interviews have a non-English primary language

OH is not just the recording:

recording | transcript | subtitles | controlled vocabs | name authorities

OHMS brings it all together

Archival/ production workflows + OHMS

OHMS is sort of the end

Record/ accession interview -> cataloging -> transcription -> QC ->

editing -> edit approval -> PICTURE LOCK -> subtitling -> QC -> prep on OHMS

Your IT dept. + OHMS

viewer is on GitHub: https://github.com/uklibraries/ohms-viewer

All PHP

works on a variety of web publishing platforms

-PST is Wordpress wrapping an <iFrame>

Getting buy-in

OHMS setup

OHMS is 2 things:

OHMS Backend

-where you make the index and coordinate metadata

OHMS Viewer

-where you display your stuff to the public

OHMS overview

Interview Manager

Thesaurus Manager

Import Interviews

Users

Repository MGMT

OHMS Metadata

Let's add a video to our repo!

QUIRKS

Sync

Language

Media URL must contain https:/ prefix, no private videos

OHMS backend login isn't HTTPS

1 repository per user email

Maybe JSON?

1:|14(7)|26(1)|38(1)

{ syncBase:60,

    {

    syncPoint:{line:14,word:7},

    syncPoint:{line:26,word:1},

    syncPoint:{line:38,word:1}

    }

}

Pythonically:

```
sync = { base:60,
data:[[14,7],[26,1],[38,1]] }

>>> sync['base']
60
>>> for syncPoint in sync['data']:
...          line = syncPoint[0]
…          word = syncPoint[1]
```

Indexing

Philosophy - Levels of indexing

Best Practices

Let's make some index points!

Now you!

Let's pair off and make some index points!

Extending OHMS

Exporting the index

XML walk-through

ohmsindex2mkvchapters.py

re-importing interview metadata?