Topic 6 - File Management Practice

Speaker notes

No notes on this slide.

What is this about?

- File naming
- Organizing files
- Describing files

Speaker notes

No notes on this slide.

File naming

Speaker notes

No notes on this slide.

File naming

- `Untitled1.mkv`
- `video.mkv`
- `my_great_recording_2019_day1.mkv`
- `My great recording 2019: Day 1.mkv`
- `April 2019 - Großartige Aufnahme Tag 1 [FFV1,PCM,1080p].mkv`
- `>>>GREAT<<< Recording!❤! feat. Péri Änders, Sœkůr T. in กรุงเทพมหานคร.mkv`
- `mgrd1.mkv`
- `VX-00815.mkv`
- `19530103/2.mkv`
- `de305d5475b4431badb2eb6b9e54601.mkv`
- …

Speaker notes

No notes on this slide.

Filenames Matter

A CATALOGUE OF

BAD FILE NAMING

16

BUILT IN VERSIONING & APPROVALS

ArticleContentEdited10-5-14JLK10-6-15ATKrevised10-17-15approved.docx

I don't even know.

Source: A catalogue for bad file naming (2015)

Speaker notes

No notes on this slide.

File naming

What to look out for (=avoid) when naming files?

- non-alphanumeric / non-ASCII
- Spaces!
- < > " / \ ? % : | * :
- (total) string length

Why?

Speaker notes

Modifiying filenames to make them work more smoothly for computers can be a very emotional thing: Invisible Defaults and Perceived Limitations: Processing the Juan Gelman Files (by Elvia Arroyo-Ramirez, 2016)

I find it good that she highlights that lanuguage and expression is a very personal thing and represents cultural codings - and that it's our duty as archivists to safeguard what is handed over to us as original as we can.

She also stumbles over incompatibilites between computer (operating) systems.

When reading the article I felt that she was putting blame on the Linux system for not treating certain encodings well, when in fact it's the other way around: Windows introduced its own codepages, and also they didn't configure their Linux system correctly so it would read these characters properly.

I want to add though, that actually it's open source systems that support the biggest number of languages and encodings that exist worldwide: Especially the ones that are not as "commercially interesting" for other systems to really put effort it.

Keep and guard the materials handed over to you as best as you can, yet be aware that computers that use filesystems have a hard time dealing with certain characters in filenames. Preserve the text of the original filenames, but maybe not necessarily in the filename, but rather as metadata somewhere where it doesn't cause machines to freak out.

A filename is not a catalogue. Information that goes beyond the usage as a human+machine filename handle, but causes problems is IMO better put somewhere where special characters are kept and treated properly - like a database, or inside the file as embedded metadata.

“Stabilizing” Filenames

A filename is not a catalogue. *Replacing (or removing) potentially “problematic” characters from file- and foldernames makes sense.*

…but document it!

Speaker notes

Any modification to file- or folder names is theoretically reversible, but:

Therefore: When applying any modifications to file- or folder-names, by replacing potentially “tricky” characters (spaces, umlauts, etc) it is good practice to document the original name.

This chapter was previously called “Detoxing Filenames”.

File naming your personal stuff?

- Sortable timestamp
  20190618_103918-awesome_friends.jpg
  (so alphabetical = chronological)

- Consistent naming per folder?
  IMG_0801.JPG, SL740465.jpg, DSCN7717.JPG, …

- Folders = groups, collections
- Hierarchical folders:
  2021/05-EQZE/slides (=year / event / record_source)

Speaker notes

No notes on this slide.

Organizing Files

Establish file and folder naming conventions, identifiers, and storage locations

Speaker notes

No notes on this slide.

Exercise

Organizing Files

- Copies: Original, Access
- Media types: AV, audio, images?
- Sidecar / Metadata files?
- Stabilize original names?
- Does it scale?
- Uniqueness?

Speaker notes

You will be creating a file and folder naming structure for your assets. Expect that you will be making low resolution access copies of all files. Ensure your directory/file naming structure clearly identifies preservation masters from access copies. This could be done using folders, or using file naming.

You may choose to rename the original files, or not. Be prepared to explain why you decided to keep a file name or not.

They key to file organization is consistency. Ensure that your structure can grow and scale to accommodate additional content beyond just these field recordings.

Document your structure and present to the group. After discussion and input from other participants, create your hierarchies and re-arrange your files.

Describing files

Which fields/data to gather for…?

- Collection level
- Item level
  - Preservation files
  - Access files
- Content level

Speaker notes

For preservation, it is common (and necessary) to distinguish between different levels of description. They are related to each other, and the boundaries may sometimes be fuzzy, but important to draw them.

Examples:

Collection level

- ID
- Title
- Description
- Genre
- Rights
- Which recordings exist
- Meta-Metadata:
  - Cataloger
  - Entry Date

Item level

Describing the physical item or digital file.

- ID
- Which Collection? (=Collection ID)
- Title
- Duration
- File name / location
- File format
- Hashcode
- Filesize
- Time in/out
- Recording details (date, engineer, location, etc)
- Meta-Metadata:
  - Cataloger
  - Entry Date

Content level

Describing what's inside the recording. The content. The subject focus and level of detail may vary greatly - depending on intention and capabilities.

Examples:

- Abstract
- Synopsis
- Naming people, objects, etc.
- Historical details
- …

Exercise

Creating a Catalog

Speaker notes

We will be creating a catalog for the collection using a simple data model. During this segment, you will be building out this model in a spreadsheet.

Create tabs for Collection and Item Set all worksheets to Text format (walk through how to do this together) Create an identifier scheme (hint: make it sequential and use pull down method) Create at least one controlled vocabulary and implement it (walk through how to do this together) Create a set of guidelines that document what information to enter in each field, and in what format. Document specifically your entry formats for: Dates Technical information (e.g., bitrate, file size)

Once you are finished, Swap your Catalog form and guidelines with another team. You will be using the other team's Catalog form for cataloging in the next activity. Review your new Catalog form and guidelines.

Do you have any questions for the other group?

Comments?

Questions?

Speaker notes

No notes on this slide.

Links:

- Detox: Filename cleaner
- Filename (Wikipedia)
- Filename (Wikipedia): Reserved Characters and Words
- Naming Files, Paths, and Namespaces (Microsoft)

Speaker notes

No notes on this slide.