

A cartoon image of Bart Simpson from The Simpsons, looking out a window with a wooden frame. He has a surprised or concerned expression. The background outside the window shows a blue sky and green hills.

**DAM'ED IF YOU DO,
DAM'ED IF YOU DON'T**

Using Generative AI to Support a Digital Asset Manager Migration

R.C. Miessler, Gettysburg College
DLF Forum 2025, November 18, 2025
CC-BY 4.0

IT'S A PERFECTLY CROMULENT PRESENTATION

This presentation isn't:

- A call to replace library workers with some sort of AI
- An attempt to convince you that AI solves all your library problems and should be used uncritically
- Anti-AI

A PRETTY BIG WHEEL DOWN AT THE CRACKER FACTORY

- Digital Initiatives Librarian
- Not a developer
- Not a cataloger/metadata librarian
- Not a digital asset manager



I TIED AN ONION TO MY BELT

- CONTENTdm shop for around 20 years
- Increasing dissatisfaction with CONTENTdm
- Staffing changes meant no dedicated digital collections manager
- Migration to [AM Quartex](#) with go-live of June 1, 2025
- No IT or developer support to automate

WE`VE TRIED NOTHING AND WE`RE ALL OUT OF IDEAS

- Assets on network storage are not organized or named for a migration
- Metadata is inconsistent and ugly
- [CONTENTdm API](#) not incredibly intuitive on its own
- [Lyrasis Cdmtools](#) provides tools to leverage the API, but designed for migrations to Islandora

(QUARTEX) CELLENT

- Images in compound objects (assets) need to be foldered together when uploaded to Quartex
- Compound object images and folders need to have their unique identifier as the filename
- Asset and image metadata are separate ingests
- Non-OCR transcripts have to be named with the unique identifier before uploading and foldered together
- PDFs can't be part of a compound object

GOOD LORD, WHAT IS HAPPENING IN THERE

- TSV files exported from CONTENTdm
- JSON files: 0.json, 1.json, 2.json, 3.json, 4.json, etc.
- Images: 0.jp2, 1.jp2, 2.jp2, 3.jp2, etc.
- PDFs
- CPD “files”

\$20 BUYS MANY PEANUTS

- Previous experience using ChatGPT to develop R scripts for sentiment analysis
- Build flowcharts to determine connections and workflows
- Provide example scenarios
- Visualize final results
- Upgrading to ChatGPT for \$20/month allowed for better results

Compound object JSON file

(4.json)

```
"migchilddata": {  
  "0": {  
    "title": "Page 1 - 28 September 1915",  
    "file": "1.jp2"  
  }, /0  
  "1": {  
    "title": "Page 2 - 28 September 1915",  
    "file": "2.jp2"  
  }, /1  
  "2": {  
    "title": "Page 3 - 28 September 1915",  
    "file": "3.jp2"  
  }, /2  
  "3": {  
    "title": "Page 4 - 28 September 1915",  
    "file": "4.jp2"  
  } /3  
}, /migchilddata  
"migcompobjtype": "Document",  
"migobjcategory": "compound",  
"migobjlevel": "top",  
"migfiletype": "cpd",
```

Image JSON file

(0.json)

```
"dmrecord": "0",  
"restrictionCode": "1",  
"cdmfilesize": "1104342",  
"cdmfilesizeformatted": "1.05 MB",  
"cdmprintpdf": "0",  
"cdmhasocr": "0",  
"cdmisnewspaper": "0",  
"migobjcategory": "Document",  
"migobjlevel": "child",  
"migparentptr": "4",  
"migtitle": "Page 1 - 28 September 1915",  
"migfile": "1.jp2",
```

Files on network drive downloaded from CDM

Ps 0.jp2

11/8/2024 1:56 PM

Ps 1.jp2

11/8/2024 2:06 PM

Ps 2.jp2

11/8/2024 2:10 PM

Ps 3.jp2

11/8/2024 2:24 PM

Compound object JSON file (4.json)

```
"migchilddata": {  
  "0": {  
    "title": "Page 1 - 28 September 1915",  
    "file": "1.jp2"  
  }  
}, /0
```

Image JSON file (0.json)

```
"dmrecord": "0",  
"restrictionCode": "1",  
"cdmfilesize": "1104342",  
"cdmfilesizeformatted": "1.05 MB",  
"cdmrelated": "0"
```

CONTENTdm
Metadata

Title	Identifier	CONTENTdm number	CONTENTdm file name
Page 1 - 28 September 1915	Peirs_Le_1915-09-28_01	0	1.jp2
Page 2 - 28 September 1915	Peirs_Le_1915-09-28_02	1	2.jp2
Page 3 - 28 September 1915	Peirs_Le_1915-09-28_03	2	3.jp2
Page 4 - 28 September 1915	Peirs_Le_1915-09-28_04	3	4.jp2
H.J.C. Peirs to Father, 28 September 1915	Peirs_Le_1915-09-28	4	5.cpd
Page 1 - 30 September 1915	Peirs_Le_1915-09-30_01	5	6.jp2
Page 2 - 30 September 1915	Peirs_Le_1915-09-30_02	6	7.jp2

```
    "title": "Page 4 - 28 September 1915",  
    "file": "4.jp2"  
  } /3  
}, /migchilddata  
"migcompobjtype": "Document",  
"migobjcategory": "compound",  
"migobjlevel": "top",  
"migfiletype": "cpd",
```

Files on network drive downloaded from CDM

Ps 0.jp2	11/8/2024 1:56 PM
Ps 1.jp2	11/8/2024 2:06 PM
Ps 2.jp2	11/8/2024 2:10 PM
Ps 3.jp2	11/8/2024 2:24 PM

Compound object JSON file (4.json)

```
"migchilddata": {  
  "0": {  
    "title": "Page 1 - 28 September 1915",  
    "file": "1.jp2"  
  }, /0  
  "1": {  
    "title": "Page 2 - 28 September 1915",  
    "file": "2.jp2"  
  }, /1  
  "2": {  
    "title": "Page 3 - 28 September 1915",  
    "file": "3.jp2"  
  }, /2  
  "3": {  
    "title": "Page 4 - 28 September 1915",  
    "file": "4.jp2"  
  }, /3  
}, /migchilddata  
"migcompobjtype": "Document",  
"migobjcategory": "compound",  
"migobjlevel": "top",  
"migfiletype": "cpd",
```

Matched files with identifiers

Ps 0.jp2	Peirs_Le_1915-09-28_01
Ps 1.jp2	Peirs_Le_1915-09-28_02
Ps 2.jp2	Peirs_Le_1915-09-28_03
Ps 3.jp2	Peirs_Le_1915-09-28_04

Pull it all together

Name	
Peirs_Le_1915-09-28	1
Peirs_Le_1915-09-30	1
Peirs_Le_1915-10-02	1
Peirs_Le_1915-10-05	1
Peirs_Le_1915-10-08	1
Peirs_Le_1915-10-09	1
Peirs_Le_1915-10-10	1

Name	
Ps Peirs_Le_1915-09-28_01.jp2	1
Ps Peirs_Le_1915-09-28_02.jp2	1
Ps Peirs_Le_1915-09-28_03.jp2	1
Ps Peirs_Le_1915-09-28_04.jp2	1

Page 1 - 28 September 1915	Peirs_Le_1915-09-28_01
Page 2 - 28 September 1915	Peirs_Le_1915-09-28_02
Page 3 - 28 September 1915	Peirs_Le_1915-09-28_03
Page 4 - 28 September 1915	Peirs_Le_1915-09-28_04
H.J.C. Peirs to Father, 28 September 1915	Peirs_Le_1915-09-28



WHAT THE HELL WAS THAT

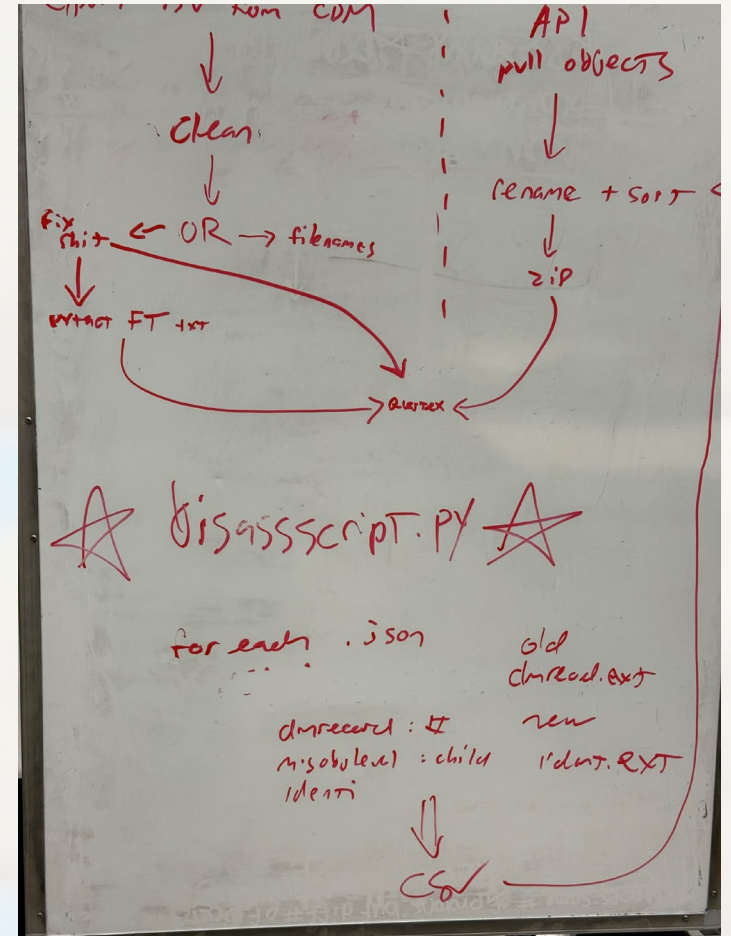
- Renaming files to their identifier
- Moving images in compound objects to appropriate folder
- Splitting metadata spreadsheets into asset and image
- Creating transcript files and foldering
- ImageMagick-ing and transcoding

JUST DO A HALF- ASSED JOB

- Between ChatGPT and Python scripts, renaming and organizing files went ok
- Still massive amounts of metadata cleanup ... enter, OpenRefine
- ChatGPT sort of sucks at OpenRefine GREL, and likes to gaslight you
- A lot of the cleanup was done by hand via OpenRefine facets and filters

I AM SO SMART, S-M-R-T

- [SMARTER prompt engineering](#)
- Be specific ... GIGO
- Explain the problem and the intended solution
- Provide examples of input and expected output
- Don't trust anything at first
- Iterate as needed and provide feedback



Preliminary whiteboard brainstorming

I have two folders - one `_migrecords` that has json files (0.json, 1.json, etc.) and one `_objects` that has various image and document files (2.pdf, 3.jp2, 4.jpg, etc). These are files that have been exported from CONTENTdm and the file name corresponds to the pointer value in CONTENTdm. The json files have relevant metadata about the files.

I also have a csv file with three relevant columns - `newIdentifier` (this will be the final file name (without a file extension)) of the file, `CONTENTdm` number (which are pointers that correspond to the files in `_migrecords` and `_objects` (without file extensions)) and `CONTENTdm` file path, which the only relevant piece I am looking for is to see if that value ends in `cpd` which indicates it is a compound object in CONTENTdm.

The JSON files I am concerned about will have the following:

```
"migobjcategory": "compound",  
"migobjlevel": "top",  
"migfiletype": "cpd",
```

as well as a list of child pointers

```
"migchildptrs": [  
    "1",  
    "2",  
    "3",  
],
```

I want to accomplish the following:

1. Create folders for each compound object
2. Move the child files for each compound object into their relevant folders
3. If it is not a compound object, leave it at the top level
4. Rename each folder and file based on the value in `newIdentifier` in the CSV

is this possible with python

need a script that will do the following:

look in a folder of txt files (transcripts) (make interactive)

look in a folder of assets (make interactive)

each asset could be a folder or an individual file

in the folder of txt files, those files need to be sorted into the same folder structure as the assets files

example

transcript\InterviewA_01.txt

transcript\InterviewA_02.txt

transcript\InterviewA_03.txt

transcript\InterviewB_01.txt

transcript\InterviewB_02.txt

transcript\InterviewC.txt

these correspond to

\collection\InterviewA\InterviewA_02.jpg

\collection\InterviewA\InterviewA_02.jpg

\collection\InterviewA\InterviewA_02.jpg

\collection\InterviewB\InterviewB_01.jpg

\collection\InterviewB\InterviewB_02.jpg

\collection\InterviewC.jpg

the final result should be

\transcript\InterviewA\InterviewA_01.txt

\transcript\InterviewA\InterviewA_03.txt

\transcript\InterviewA\InterviewA_03.txt

\transcript\InterviewB\InterviewB_01.txt

\transcript\InterviewB\InterviewB_02.txt

\transcript\InterviewC.txt

I need to create item level metadata CSVs for digital collections that I am migrating. I already have asset level metadata CSVs.

How things are currently configured:

the files (jpg, mp3, mp4) for each collection are in
Y:\amq\Quartex\COLLECTIONID\assets

in the \assets directory, any of the following may be true:

1. there is a directory of files (this is a asset - it's a compound object of several files) - these files could be jpg, mp3, or mp4
2. there are individual jpg, mp3, or mp4 files (these are assets as well, just not compound objects)
3. there are pdfs (these are assets as well, just not compound objects)

The asset-level metadata is in
Y:\amq\Quartex\COLLECTIONID\metadata and has a filename ending in "-assets.csv"

essentially, I need a COLLECTIONID-items.csv ONLY if this is a compound object. so I am only looking at the folders and their files in Y:\amq\Quartex\COLLECTIONID\assets

I do not need any rows that are only for assets in the final COLLECTIONID-items.csv

the first column should be the folder name (Asset)
the second column should be the file name (Item) without the file extension
the rest should be the columns from
Y:\amq\Quartex\COLLECTIONID\metadata\COLLECTIONID-assets.csv. the values in the Identifier column in the COLLECTIONID-assets.csv are the same as the values in the Item column

so in the COLLECTIONID-items.csv, there will always be an Asset column, an Item column, an Identifier column, and the Title column from COLLECTIONID-assets.csv

as for the rest of the columns that will eventually be in COLLECTIONID-items.csv, I do not want metadata values in each column that duplicate the value at the COLLECTIONID-assets.csv level (because that would be duplicative in the final display)

for example, in Y:\amq\Quartex\TESTING\ AA_123 there are 3 files, AA_123_01.jpg, AA_123_02.mp4, AA_123_03.mp3

The asset metadata is:

Identifier: AA_123
Title: Some guy
Subject: Gettysburg
Type: Image; Video; Audio

the item metadata is

Identifier: AA_123_01, AA_123_02, AA_123_03
Title: Picture of man, Interview, Song
Subject: Gettysburg, Gettysburg, Gettysburg
Type: Image, Video, Audio

This would result in rows in the items.csv:

Asset: AA_123, AA_123, AA_123
Item: AA_123_01, AA_123_02, AA_123_03
Identifier: AA_123_01, AA_123_02, AA_123_03
Title: Picture of man, Interview, Song
Subject: [blank because it is duplicative of the asset csv]
Type: Image, Video, Audio

can a python script (or 2) be created to accomplish this?

Ok so there's one complication that has come up. Sometimes in the CSV the column "CONTENTdm file path" ends with a .pdf extension but it is not a PDF but rather a compound object. In this case, the column "CONTENTdm file name" usually ends in .cpd. Can the script be modified so that if CONTENTdm file path ends with .pdf, check to see if CONTENTdm file name ends in .cpd, and if it does, then use that as the trigger for a compound object?

modify this script further so that if CONTENTdm file name ends in .cpd it also renames the folder name from oldFilename to newIdentifier

I need the CSV to be a bit different. there should be a column for each separate file type (column header the file type) and then a final total files column that has the total number of files in /assets and its subdirectories

so this did something weird where it created the expected number of folders, sorted the files into them, but then recreated all the files in the files to amq directory as well.

great, that worked

still the same error please try a different approach

S-U-C-C-E-E-S

- Beat our go-live date – <https://gettdigital.gettysburg.edu>
- Modern digital asset manager and public interface
- Cleaner metadata (but still sort of messy)
- Pain points involved human input and decision-making
- I know a lot more about prompt engineering and code

IM LEARNING

- Saves time and allows for more nimble approach
- Possible solution for institutions without access to developers
- Still have to understand code and development environments
- AI may not be able to deal with more complex problems

SO IN CONCLUSION: GEN MEANS GENERATIVE AND AI MEANS AI

- “Technology is the active human interface with the material world.” - Ursula K. Le Guin, [A Rant About Technology](#)
- AI changes how we work and think
- Where is it getting this code from?
- How many gallons of water did I burn through?
- WTF do I do with the code?
- Is this moment with AI and libraries akin to a “moral panic”, a fundamental shift, or just another bump in a long line of changes?



Thanks! Questions?

Simpsons images: <https://frinkiac.com>