

Jumping the DAM fence

from CONTENTdm to Quartex: gambling on a new digital repository

Tanya Hollis, Alex Cherian, Michael Herrick,
Matthew Martin, & Alex Post

J. Paul Leonard Library, San Francisco
State University

Feeling fenced in by CONTENTdm

We were looking for:

- Better discovery and usability
- Ability to configure user interface
- Accessibility, including closed captioning
- Flexibility and reuse of assets within the system
- Interoperability, moving content in and out of the system

Quartex offered:

- Ease of use on all levels
- Ability to edit, design, and customize our own pages
- Transcription of audio, video, and handwritten texts
- Exhibits and curated collections



San Francisco State University Special Collections & Archives

The J. Paul Leonard Library's Digital Collections include materials drawn from our collections.



Special Collections and Archives

The J. Paul Leonard Library's Special Collections & Archives contain a wealth of primary source materials on a wide variety of topics. The collections include manuscripts, organizational records, personal papers, photographs, rare books, periodicals, audio-visual media, artwork, artifacts and ephemera. These collections support academic and personal scholarship on campus, in the wider community, and internationally.

Can Quartex make our DAM grass greener?

Key advantages of closed captions in Quartex:

- Active role in software development process.
- Functional upgrades to caption editing software.

Systems and Metadata

Good fences make good neighbors

- 1 Assess and solve ContentDM-related issues
- 2 Get Implementation Team basic training and access
- 3 Agree to the basic structure of backend, to support 3 front e...
- 4 Define FieldTypes and other essential configurations
- 5 Test all types of data imports with a Test Collection
- 6 Review essential configurations carefully / Sign off on basic s...
- 7 Draft file ingest and metadata review, wrangling and uploadi...
- 8 Design decisions for the SpecColl and Strike Collection URLs
- 9 Implement data ingest workflows and Design features
- 10 Full QC of first Promotions / Launch
- 11 Launch / Promote, iteratively

Field definitions based on:

- 1) CONTENTdm legacy data
- 2) Envisioned Quartex needs
- 3) Eye toward eventual linked data enrichment



The timeline to meet instructional needs dictated that the draft metadata ingest became the final decisions

Fields can be configured at 3 levels: Document (i.e. Object), Image and Section. Controlled Vocabularies then have their own validation rules and decision points

Document fields

Set up and edit your document metadata fields. Choose the field name and define the field type to identify how metadata should be created for your assets.

Show help

Add new group

- Standard Fields [Rename](#) 29 fields
- Translations [Rename](#) 2 fields
- Test [Rename](#) 4 fields
- SCA Audio [Rename](#) 5 fields
- BATA [Rename](#) 4 fields

Field name	Field type	Field required	Promote	
Duration	Free Text	Yes	Yes	Advanced x
Current Media Master	Free Text	Yes	No	Advanced x
Opening Credits By	Free Text	No	Yes	Advanced x
Subject Tags	Controlled Vocabulary	Yes	Yes	Advanced x

Add Custom Field Add Field Cluster

Mapping CONTENTdm fields to a Quartex structure

This can be tweaked post-migration	DO NOT ALTER after MIGRATION			Can be changed after	This can be changed easily too and takes effect once running a Promotion				
Draft Quartex Field Name	Field Group	Q Field type	Make limited values	Q Field required	Q Promote	Q repeatable	Notes	New Req	SF Decision
Title	Standard	Free Text	N/A	Yes	Yes	N/A			
Date	Standard	Date	N/A	Yes	Yes	N/A			
Creator	Standard	Controlled Vocabulary	N/A	No	Yes	N/A	CV?		fff to CV
Description	Standard	Free Text	N/A	No	Yes	N/A			
Additional Date Information	Standard	Free Text	No	No	Yes	No			
Publisher	Standard	Free Text	N/A	No	Yes	N/A			
Physical Description	Standard	Free Text	N/A	No	Yes	N/A			
Contributor	Standard	Free Text	N/A	No	Yes	N/A			
Other Title	Standard	Free Text	N/A	No	Yes	N/A	Discuss G		Yes to No
Rights URI	Standard	Free Text	Yes	Yes	Yes	No	For control? Or	Yes	Change to FTF
Rights	Standard	Controlled Vocabulary	Yes	Yes	Yes	No		Yes	Change to field nar
Rights Statement	Standard	Free Text	N/A	No	Yes	N/A			Change to Rights S
Rights Holder	Standard	Free Text	N/A	No	Yes	N/A			
Subject (Name)	Standard	Controlled Vocabulary	Yes	No	Yes	Yes			
Subject (Place)	Standard	Controlled Vocabulary	Yes	No	Yes	Yes			
Subject (Topic)	Standard	Controlled Vocabulary	Yes	No	Yes	Yes			
Form/Genre	Standard	Controlled Vocabulary	Yes	No	Yes	Yes			
Country of Creation	Standard	Controlled Vocabulary	Yes	No	Yes	No			
Language	Standard	Controlled Vocabulary	Yes	Yes	Yes	Yes		Yes	Probably make Not
Digital Object Type	Standard	Controlled Vocabulary	Yes	Yes	Yes	Yes		Yes	
ID Number	Standard	Free Text	N/A	Yes	Yes	N/A			

Metadata wrangling

←

Outline

In Excel

- Step 1: Add Units/Institution...
- Step 2: Use Priority 1/Pull L...

In Open Refine

- OpenRefine Step 1: Draft st...
- Notes from OpenRefine, f...
- OpenRefine Step 2: Try out ...
- OpenRefine Step 3: Implem...

After OpenRefine:

- Post-migration/in Quartex r...
- Project Note
- Metacollection

OpenRefine Step 2: Try out all steps on the Priority 2 Assets only

1. Opened the data source from Box: Master_Final_CdmEraData_Combined.xlsx
2. Reviewed OR Step 1 in comparison to the Excel steps and make sure can commence
 - a. Decided to try filtering, repeatedly, in OR rather than to break down sets in Excel
 - b. If this works, need to follow the steps, but include the reverse, when working with image-level metadata
 - c. Proceeding with one data set but each priority will treat CDM Blank first and then the "Collections", verifying the export set against the Quartex Asset Ingest Spreadsheet each time.
 - d. Look at Dates and Rights, very carefully in Priority 2 to ascertain how much can be automated across the full set. Consider doing for all of Priority 3-6 at once, if laborious, rather than repeating with each Batch
 - e. Need to add Pull Lot values, in OR, before export/releasing the filter
3. Copied all OR Step 1 tasks to this area, highlighting in yellow when there was an addition, or divergence from the Priority 1 steps

[For succeeding sets, need to first filter on Priority. Then work with Simple only but selecting CDM_LVL=blank. After exporting/setting Pull Lot for the Simple = Letter1 then select the Compound objects by Faceting on Priority/Pull Lot=blank/CDM_LVL_NAME=Collection. After wrangling, assign Pull Lot = Letter2, export the Document level data and go back in to assign

Facet / Filter Undo / Redo 814 / 814

6407 rows

Extensions Wikidata

Refresh Reset all Remove all

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

« first « previous 1 next » last »


Pull lot change

18 choices Sort by: name count Cluster

- A1 7
- A2 11
- A2_Image 409
- B1 62
- B2 32
- B2_Image 666
- C1 108
- C2 29
- C2_Image 135
- D1 1706
- D2 67
- D2_Image 1577

	All	Pull lot	Priority	Straggler	CDM_LVL	CDM_LVL_NAME	Quartex Name	ID Number	Title	Date	Date Statement	Date Published	Copyright Date	Date Searchable	Rights
		798.	B1	2	N			UARC_0333	UARC_0333	Summerskill. We won't support you until you support us	1968				Other
		799.	B2_Image	2	N	0	Side 1	SCRB_0001	Side [1]						
		800.	B2_Image	2	N	0	Side 2	SCRB_0001	Side [2]						
		801.	B2	2	N	0	Collection	SCRB_0001	SCRB_0001	[Missile Lomovcense Castrease : fragment]	1401-1499	circa 1440		1401; 1402; 1403; 1404; 1405; 1406; 1407; 1408; 1409; 1410; 1411; 1412; 1413; 1414; 1415; 1416; 1417; 1418; 1419; 1420; 1421; 1422; 1423; 1424; 1425; 1426; 1427; 1428; 1429; 1430; 1431; 1432; 1433; 1434; 1435; 1436; 1437; 1438; 1439; 1440; 1441; 1442; 1443; 1444; 1445; 1446; 1447; 1448; 1449; 1450; 1451; 1452; 1453; 1454; 1455; 1456; 1457; 1458; 1459; 1460; 1461; 1462; 1463; 1464; 1465; 1466; 1467; 1468; 1469; 1470; 1471; 1472; 1473; 1474; 1475; 1476; 1477; 1478; 1479; 1480; 1481; 1482; 1483; 1484; 1485; 1486; 1487; 1488; 1489; 1490; 1491; 1492; 1493; 1494; 1495; 1496; 1497; 1498; 1499	No Copyright
		802.	B2_Image	2	N	0	Side 1	SCRB_0002	Side [1]						

Like a good neighbor, don't throw a trash bag filled with files into your neighbor's DAM yard

A solid green diagonal shape that starts from the bottom-left corner and extends towards the top-right corner, occupying the lower half of the slide.

Special Collections Migration Initial Plan

- Constraints
 - 2 weeks original deadline
 - No access files
- Acquired access files from CONTENTdm
- Planned to upload those JP2 files directly to AM Quartex

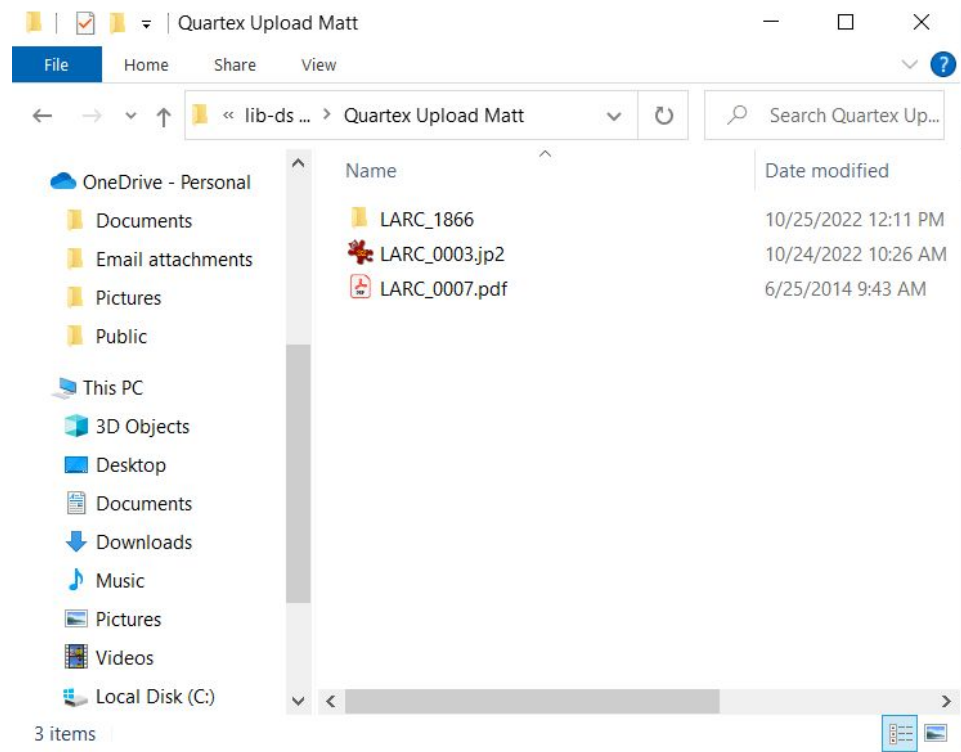
CDM_LVL	CDM_LVL	Identifier	Title	Creator	Date Created	CONTENTdm number	CONTENTdm file name	CONTENTdm file path
		DBEL0005	Frank V. de Bellis with Ferruccio Tagliavini and Federico Ghisi		circa 1953	17	6.jp2	/p16737coll4/image/6.jp2
		DBEL0014	Frank V. de Bellis playing wax cylinder recording		circa 1950	18	19.jp2	/p16737coll4/image/19.jp2
		DBEL0014	Frank V. de Bellis playing Pitt, Don Alex		1952	19	20.jp2	/p16737coll4/image/20.jp2
		DBEL0014	Frank V. de Bellis hosting radio show at KSMO in San Mateo		circa 1950	20	21.jp2	/p16737coll4/image/21.jp2
		DBEL0014	Frank V. de Bellis hosting "Music of the Italian Masters"		circa 1950	21	22.jp2	/p16737coll4/image/22.jp2
		DBEL0014	Frank V. de Bellis with Italo Tajo, Ferruccio Tagliavini, and Tito Go		1948	22	23.jp2	/p16737coll4/image/23.jp2
		DBEL0014	Cloe Elmo as Aida	Melançon, Louis	circa 1948	23	24.jp2	/p16737coll4/image/24.jp2
		DBEL0014	Guiseppe Valdengo performing Il Cenacolo (San Francisco, Calif.)		14-Oct-51	24	25.jp2	/p16737coll4/image/25.jp2
		DBEL0014	Frank V. de Bellis receiving Il Cenacolo (San Francisco, Calif.)		Oct. 14, 1951	25	26.jp2	/p16737coll4/image/26.jp2
		DBEL0014	Gaitano Merola greets soprano Brunetta Mazzolini at Stern Grove		30-Aug-53	26	27.jp2	/p16737coll4/image/27.jp2
		DBEL0014	Italo Tajo and American-born soprano Ina Souez at the opening of		circa 1955	27	28.jp2	/p16737coll4/image/28.jp2
		DBEL0014	Frank V. de Bellis with J. Paul Leonard, San Francisco Consul General		circa 1951-195	28	29.jp2	/p16737coll4/image/29.jp2
		DBEL0014	Frank V. de Bellis with Mayor George Christopher and Rev. John F		7-Mar-56	29	30.jp2	/p16737coll4/image/30.jp2
		DBEL0014	Frank V. de Bellis with Glenn S. Dumke, George Glenn Kendall, Mayor		circa 1964-196	30	31.jp2	/p16737coll4/image/31.jp2
		DBEL0014	Frank V. de Bellis with Maestro Vincenzo Giannini		circa 1966	31	32.jp2	/p16737coll4/image/32.jp2
		DBEL0014	Frank V. de Bellis in Verdi's apartment in Milano, Italy		circa 1965	32	33.jp2	/p16737coll4/image/33.jp2
		DBEL0014	Frank V. de Bellis with Governor Edmund Gerald "Pat" Brown		circa 1963-196	33	34.jp2	/p16737coll4/image/34.jp2
		DBEL0014	Frank V. de Bellis with tenor Giacinto Prandelli and bass Salvatore		circa 1965-196	34	35.jp2	/p16737coll4/image/35.jp2
		DBEL0014	Frank V. de Bellis with the Manager of the San Francisco Symphony		circa 1960-196	35	36.jp2	/p16737coll4/image/36.jp2
		DBEL0015	Headquarters for America Vano, Tom		circa 1955	36	37.jp2	/p16737coll4/image/37.jp2
		DBEL0016	Fortepiano (1808-1809) manufactured by Muzio Clementi & Co.,		circa 1963	37	38.jp2	/p16737coll4/image/38.jp2
0	Original w	DBEL0001	Napoleonic era Woodcut		circa 1798-180	38	47.jp2	/p16737coll4/image/47.jp2
0	Impressio	DBEL0001	Napoleonic era Woodcut Impression		1966	39	48.jp2	/p16737coll4/image/48.jp2
0	Collection	DBEL0001	Napoleonic era Woodcut		circa 1798-180	40	49.cpd	/p16737coll4/image/49.cpd
0	Original w	DBEL0002	Renaissance Woodcut of Youth and Old Age		circa 1470-155	41	50.jp2	/p16737coll4/image/50.jp2
0	Print from	DBEL0002	Renaissance Woodcut of Youth and Old Age		1966	42	51.jp2	/p16737coll4/image/51.jp2
0	Collection	DBEL0002	Renaissance Woodcut of Youth and Old Age		circa 1470-155	43	52.cpd	/p16737coll4/image/52.cpd
0	Recto	DBEL0002	Recto			44	53.jp2	/p16737coll4/image/53.jp2
0	Verso	DBEL0003	Verso			45	54.jp2	/p16737coll4/image/54.jp2

Revised Plan

- Deadline to lose CONTENTdm access extended to 2 months
- Decided to create our own access files:
 - Inventory our assets
 - Normalize identifiers (e.g., LARC0001 becomes LARC_0001)
 - Convert to JPEG2000
- Spreadsheets to divide test assets, high priority assets, and repositories (e.g., Labor Archives, University Archives) to track file ingest, QC, metadata ingest, etc.

Ingest

- Identifier = Document ID
 - Links asset with metadata
 - Does not require page-level metadata to create a compound object
- Upload a combination of simple or compound objects and PDFs
- Some issues uploading compound assets



Asset details

Collection(s)

Special Collections

Standard Fields

ID Number

Title

Date



Date Statement

Rights

Rights URI

Language

Edit asset details

(12 out of 52 images)  



DBEL_0048_001



DBEL_0048_002



DBEL_0048_003



DBEL_0048_004



DBEL_0048_005



DBEL_0048_006



 Back to list of digital assets

What would we do differently? Implications for future migrations

- Don't perform a migration in 2 weeks.
- JPEG for access
 - Easy to convert from TIFF
 - Smaller
 - Universal
- Own your access copies. Have them clean and well-organized.

SFSU Quartex Discovery Layer

[insert funny DAM metaphor]