

CARRF @CAS

The Ralph G. Connor Alcohol Research Reference Files is an archive of instruments (questionnaires, interview schedules, surveys, etc) that have been used in studies of substance use. The primary focus of this collection is on various aspects of drinking behavior and alcoholism.

The Society for Study of Social Problems founded this collection, which is now named in memory of Dr. Ralph G. Connor, in 1960. The Files were developed and maintained by Dr. Connor at Eastern Washington State College, and following his death they were transferred to the Center of Alcohol Studies at Rutgers University.

The Center has continued to add new instruments to the collection, and periodically issues new updates to the descriptive inventory, providing brief information on the materials.

Project Goal

The CAS Library is attempting to digitize the historical collection using a new content management system, which is in the testing process. The preliminary work of digitization has been implemented as part of the testing for the center's system and information migration.

Streamlining the digitization process is a great challenge. This project is an exhibition about what could go wrong and what lessons we have learned from this process.

Option 1:

Digitize one historical document, load, and index it in the new system. One at a time.

Option 2:

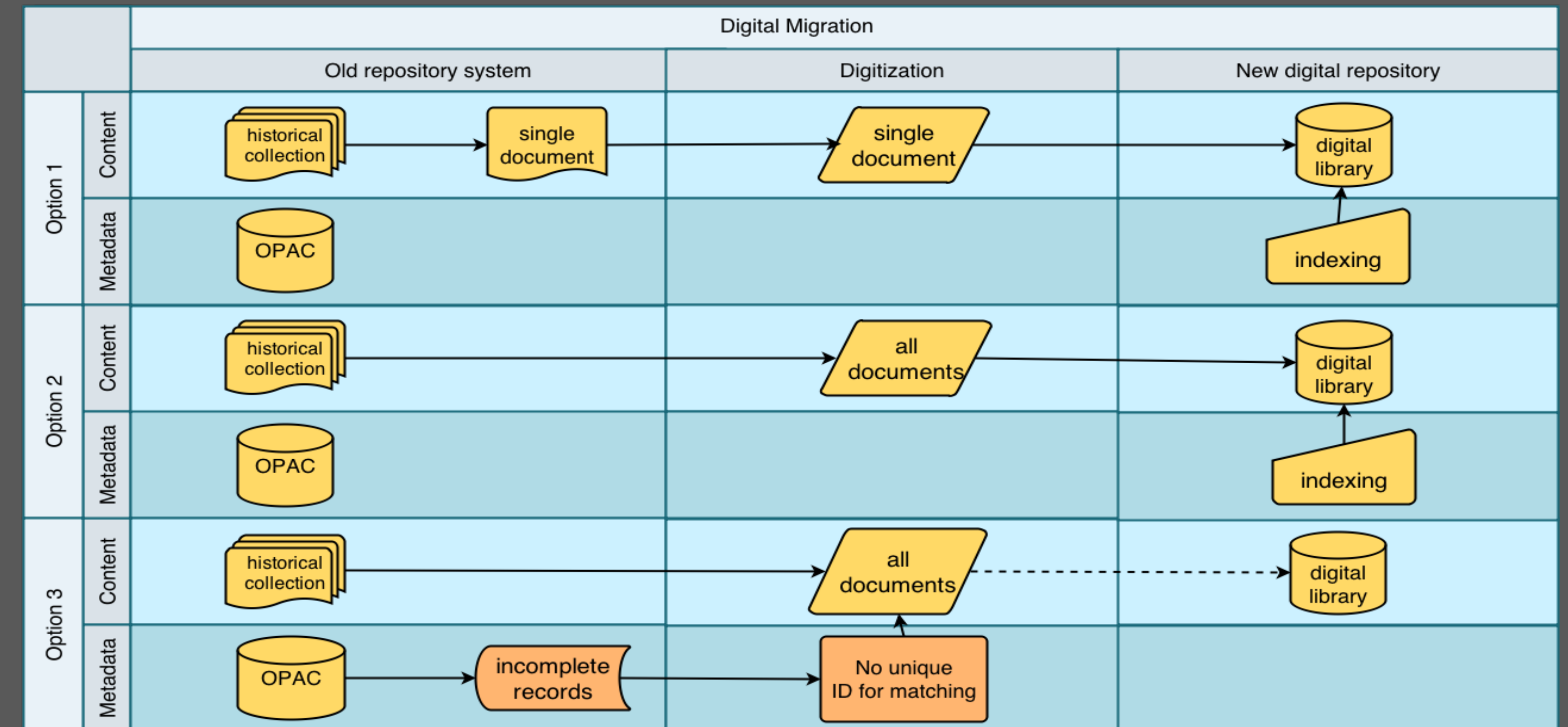
Digitize all the historical documents, load, and index them in the new system.

Option 3:

Digitize all the historical documents, extract the file names, retrieve existing metadata records, match the file names and the records by common IDs, and load them all into the new system.

The last option was assumed to be the most efficient way of digitization.

Possible ways of digitizing a historical collection



The Issues

Issue 1:

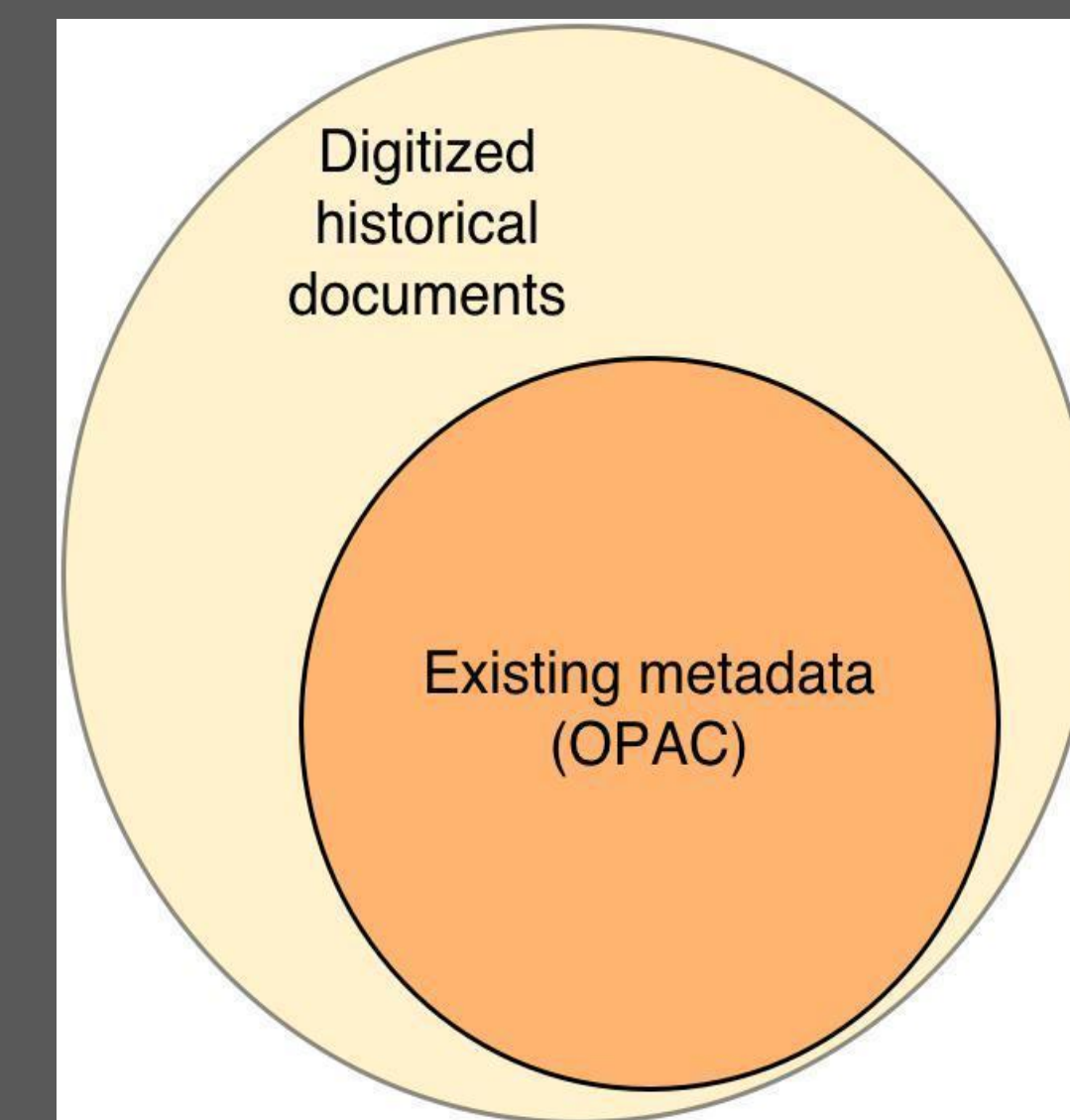
Testing found that the existing metadata records were not updated to cover all the print historical documents (see the Venn diagram at the right).

Issue 2:

For the matched, the metadata records are not granular enough to denote the related documents.

Issue 3:

The IDs for file matching were made part of the file names of the digitized historical documents. They need to be read out to be the keys for matching (see the table at the right).



Wrong
The ID was a part of name of digitized document

File Name	CAARF-ID	CAARF-ID_1
1_10ASurvey.pdf	1.10	A
1_10B1Strassburger65.pdf	1.10	B1
1_10B2Bibliography.pdf	1.10	B2
1_11ASurvey.pdf	1.11	A
1_11B1Bibliography.pdf	1.11	B
1_12ABBailey69.pdf	1.12	A
1_12B2Bibliography.pdf	1.12	B
1_13ASurvey.pdf	1.13	A
1_13B1Bibliography.pdf	1.13	B
1_14ASurvey.pdf	1.14	A
1_14B1Bibliography.pdf	1.14	B
1_15ASurvey.pdf	1.15	A
1_15B1Greene72.pdf	1.15	B1
1_15B2Bibliography.pdf	1.15	B2
1_16ASurvey.pdf	1.16	A
1_16B1Huebner76.pdf	1.16	B1
1_16B2Bibliography.pdf	1.16	B2
1_17ASurvey.pdf	1.17	A
1_17B1TableVocationalRehab.pdf	1.17	B
1_17CTableCounselorsAttitudes.pdf	1.17	C
1_18A1PilotProject.pdf	1.18	A1

Correct
Column B & C were created to read the IDs for matching the existing metadata records

File Name	CAARF-ID	CAARF-ID_1
1_10ASurvey.pdf	1.10	A
1_10B1Strassburger65.pdf	1.10	B1
1_10B2Bibliography.pdf	1.10	B2
1_11ASurvey.pdf	1.11	A
1_11B1Bibliography.pdf	1.11	B
1_12ABBailey69.pdf	1.12	A
1_12B2Bibliography.pdf	1.12	B
1_13ASurvey.pdf	1.13	A
1_13B1Bibliography.pdf	1.13	B
1_14ASurvey.pdf	1.14	A
1_14B1Bibliography.pdf	1.14	B
1_15ASurvey.pdf	1.15	A
1_15B1Greene72.pdf	1.15	B1
1_15B2Bibliography.pdf	1.15	B2
1_16ASurvey.pdf	1.16	A
1_16B1Huebner76.pdf	1.16	B1
1_16B2Bibliography.pdf	1.16	B2
1_17ASurvey.pdf	1.17	A
1_17B1TableVocationalRehab.pdf	1.17	B
1_17CTableCounselorsAttitudes.pdf	1.17	C
1_18A1PilotProject.pdf	1.18	A1



Lessons Learned

1. If the metadata of an historical collection exists in an online public access catalog (OPAC) database, the digitization of each document should be accompanied by checking whether or not the corresponding metadata are included in the OPAC database.
2. When digitization and indexing / cataloging are not undertaken at the same time, a tentative ID used for the digitized document needs to match the ID used for the OPAC database.
3. The issue of compatibility is a challenge for metadata migration. Most of the time, only a part of the existing metadata elements and values are selected for the new repository system. How to link the metadata records with the digitized objects and then add in more metadata elements needs careful advance planning for peak efficiency.