

PIM 2022-012 – Attachment 4
File Naming Standard for Control Document Index Records in GLORAS

Creating Import Batches for the Control Document Index (CDI) Document Management System

Prerequisite

If the image's document is in BLM's LR2000 database, this needs to be determined before the import batch is created. CDI uses the document's LR2000 case ID, when available, to populate its document information records with data from LR2000.

If an image in the import batch has no case ID specified, CDI assumes LR2000 does not include the document, and the CDI operators may have to manually enter information from the image itself before the image is available on the web site. So, to keep things efficient, every effort must be made to determine an LR2000 case ID for the page image's document before import.

Contents of an Import Batch

CDI Import is a process that attempts to import every page image in each folder and any subfolders.

1. There must be one Access 2007 **database file** (with .mdb extension) somewhere in the folder set
2. The Access file must contain a so-called **linking table**.
3. The linking table must contain a **record for each page image file**.
4. The **record should include the LR2000 case ID** for the document the image belongs to, along with certain other key information described below.
5. The **page image files** which follow the file naming convention.

Linking Table File

GLO staff will provide a CDI linking table template upon request.

The linking table's Access file can have any filename that ends in .mdb, but there must be only one .mdb file in the import folder or subfolders. If the import process finds multiple .mdb files, the process will stop with an error message.

The linking table name within the Access file must be or begin with **imagedatalocation**. Character case does not matter. Any other tables in the Access database file with other names are ignored; you are welcome to create and use additional tables for your own purposes.

Note: Make sure only one table in the Access file fits this naming criteria. If there is more than one such table the CDI import process will use the first one it comes to; which one that will be can't be predicted.

An empty linking table file, with the proper table and column names, is available.

Required Columns

The linking table is required to contain columns with the names in the table below. Some columns may be empty in certain cases, but all the columns need to be in the table. Columns with names other than the ones below are ignored. You are welcome to create and use additional columns for your own purposes.

Column name	Meaning	Required?	Note
imagenamecal	The full name of the page image file, including the .tif extension	Required	Do not include the path.
Caseid	The LR2000 case ID of the page image's document.	Optional	See discussion above. If empty, a stub record for the document is created during import containing only the information that can be gleaned from the other required columns in the linking table record and the image file name.
DocTypeCode	(see below)	Required	If empty, a warning is given during import, and image is copied to holding area without further processing.
State	Geographical state of the document.	Required	
Docid	The document number	Required	
Actiondate	The document date	Required	
Imagestatus	n/a	n/a	This column must exist, but CDI Import ignores any values in the column.

Page Image Files

The CDI Import process uses an image's filename to find its record in the linking table. A CDI filename convention (below) describes how to name an image file for CDI import. Image filenames consist of several fields of information separated by dashes, ending with the .tif extension.

The image file naming convention does a few things:

- It contains certain information used by CDI in lieu of the page image's document being in LR2000 (and therefore has no case ID in the linking table). If the image's record in the linking table does specify an LR2000 case ID, the rest of the data is taken from LR2000 instead of the linking table.
- It helps ensure that images within an import batch can all have unique names
- It helps keep CDI's internal holding area organized efficiently.

CDI scanning requirements

Each CDI image must be scanned at a minimum of 300 DPI in 8 bit grayscale or color. The images will be saved in an uncompressed TIF (a single "F" .TIF extension) file format.

CDI File Naming Convention

Each page image file submitted to the CDI Document Management System must have an alpha-numeric file name composed of the following elements. These approximately match the alpha-numeric text on the physical Control Document Index aperture card:

Example: OR330410S0430E-1969-05-21-PA36690059-01.tif

<i>Content</i>	<i>Length, type</i>	<i>Contains</i>	<i>Notes</i>
!!	2 exclamation points	Optional. Indicates image is illegible.	This is completely ignored by CDI Import. Illegibility is not an image attribute concept understood by CDI Document Management System. A scanning contractor developed this convention for their own use.
OR	2 characters	State code	
33	2 digits	Meridian code	Zero padded
041	3 digits	Township number	Zero padded
0	1 digit	Township fraction	2 represents ½, 0 for whole
S	1 character	Township direction	
043	3 digits	Range number	Zero padded
0	1 digit	Range fraction	2 represents ½, 0 for whole
E	1 character	Range direction	
-	<i>hyphen</i>		
1969	4 digits	Document year.	CDI import process will use the year to form the name of any needed folders for the image holding path.
-	<i>hyphen</i>		
05	2 digits	Optional. Document month	CDI does not use these values.
-	<i>hyphen</i>		
21	2 digits	Optional. Document day.	CDI does not use these values.
-	<i>hyphen</i>		
PA	2 digits	Document type code	See Appendix for examples
36690059	1 to 10 digits	Document number	
-	<i>hyphen</i>		
01	1 to 4 digits	Page number.	If the import process cannot interpret the characters as a number (i.e. if it has anything but digits in it), the page number is considered to be zero. No warning is given.
@00	<i>At sign, zero, zero</i>	Optional. Indicates image is the aperture card header.	
.tif	File extension	Always .tif	

About the State, Meridian, Township, and Range:

This is the information from the “CDI source” (e.g. the physical file drawer home of the aperture card). This is not necessarily the “key township”. The values in the field are used in two ways:

- To pre-populate the land description information in the CDI database’s document record in the absence of an LR2000 case ID
- To label thumbnails of the images for CDI operators during image approval

Header Image Files

If an aperture card scanner is used, it may support scanning the printing at the top of the aperture card (in addition to scanning the microfilm). We call the image files from these scans “header images”. The CDI import process will look for header image files that match up with page image files and import both files if possible. Such header images, while not required by CDI, can help the CDI operators index any missing document information.

- Header image files have the same name as page image files but end with the three characters “@00” (just before the . extension).
- The microfilm on an aperture card may include exposures of more than one page. In this case, the same header image file can be duplicated once for each page image (and named based on the file names of the page image files to which it applies).
- Header images do not have their own records in the linking table. They will only be imported if there is a matching page image file exists in the import set and the page image has a record in the linking table.

Duplicate page image filenames in same batch

If the CDI import process finds the same page image filename in multiple folders in the same CDI import batch, none of the instances of the file are imported, but it will warn about the duplication and output a log file.

Preventing multiple imports of the same image file across batches

CDI remembers the file name, file size, and creation date of every page image file it successfully imports (e.g. did not just copy to the holding area). During import, CDI compares the name, size, and date of every candidate page image file in an import batch against the files from previous imports. If it does find a match, a warning is given and the candidate page image file is simply ignored (because it’s been imported before). This mechanism prevents CDI from creating duplicate image files if the import process needs to be restarted after a failure, while allowing updated or replacement page image files to be imported.

Records in the linking table that do not match to any files in the import folder

The linking table can have records for other files than are included in the folder. They will be ignored.

Files in the import folder that are not in the linking table

The import process will copy all image files (*.tif) in the main and subfolders (except for files with the same names, see above) that do not have corresponding records in the linking table to a holding area. There is no automated process for further processing the image files in the holding area.

Importing does not change a document record after its first image

CDI document records are created the first time a page image is imported for a document. At that point, the document record is initialized with information taken from LR2000, the linking table, and/or the filename as appropriate. As subsequent page images are imported for the document, during the same or different import batch, the page images are simply “attached” to the already existing document record in CDI. The document record is not changed even if LR2000, the linking table, or the file name contains different information.

Note: The information in CDI document records can be edited and updated manually by the CDI application operators.

Example of an Import Batch

This is a basic example consisting of one page image to be imported. The import folder contains one Access file and one image file. Note that the example values are representative and may not be correct for the document used.

\\My Import Folder\

- LinkingDB.mdb
- OR330410S0430E-1969-05-21-PA36690059-01.tif

LinkingDB.mdb is an Access 2007 file that contains a table named imagedatalocation. The table contains one record with columns and values like so:

imagedatalocation

<i>imagenamcalc</i>	<i>caseid</i>	<i>doctypecode</i>	<i>state</i>	<i>docid</i>	<i>actiondate</i>	<i>imagestatus</i>
OR330410S0430E-1969-05-21-PA36690059-01.tif	6167191	PA	OR	36690059	5/21/1969	[empty]

Appendix: Document Type Codes

CDI doc type codes follow similar BLM data element standards as LR2000. For example only, here's the list of doc type codes as of August 2022.

<i>Code</i>	<i>Description</i>
86	W/D from all forms
91	Restrictive Covenant
98	Improve USA title
AC	Act of congress
AG	Agreement
AN	Air Navigation Order
AP	Appeal
BO	Blm order
BS	Bargain and sale deed
CA	Cooperative agreement
CD	Condemnation deed
CL	Clear list
CO	Class decision/order
CP	Certificate Patent
CT	Certificate of transfer
DD	Decision document
DT	Declaration of taking
ED	Easement deed
EO	Executive order
FA	Ferc application
FC	Final Certificate
FO	Fpc/ferc order

<i>Code</i>	<i>Description</i>
GD	General deed
GL	Glo order
GR	Grant
GS	Usgs order
IB	Indem lieu base
IC	Interim conveyance
ID	Interchange Deed
IL	Indem lieu sel
KC	Known classified area
LC	License
LI	Lease
LS	Lieu selection
MD	Missing document
NA	Native allotment
ND	Not determined
NR	Nora
NT	Notice
OO	Other order
OS	Order solicitors
PA	Patent
PC	Power Site CI Order
PI	Permit
PL	Public law
PO	Public land order
PP	Pres proclamation
PS	Power Site Reserve Order
PT	Petition to Restore
PW	Power Projects
QC	Quit claim deed
RD	Recordable disclaimer
RG	Rr Grant Patent
RW	Right of way
SD	State dir order
SG	State grant
SI	Succession of Interest
SL	State law
SO	Sec order
SR	Shore Space Restoration
SS	State selection
ST	Small Tract Classification
TA	Tentative Approval
TV	Title Vested
TY	Treaty
UD	Unspecified deed
UN	Unknown
WD	Warranty deed
XD	Exchange deed