

Digital Archive First Steps for Preservation of Legacy Collections



Library
Leabharlann

Boole Library, UCC,

Raelene Casey

3rd February 2025

Scope

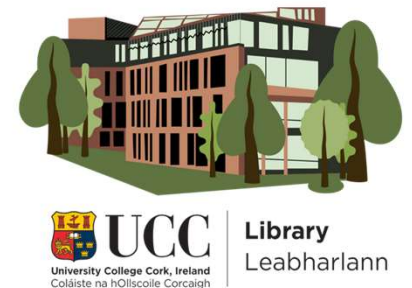
Special Collections and Archives Department

Legacy data accumulated in the Library over 15 years years.

- Digitised material (internal) (projects and reprographics)
- Digitised material (external)
- Born digital material

On-prem servers in the University. Some sharepoint storage

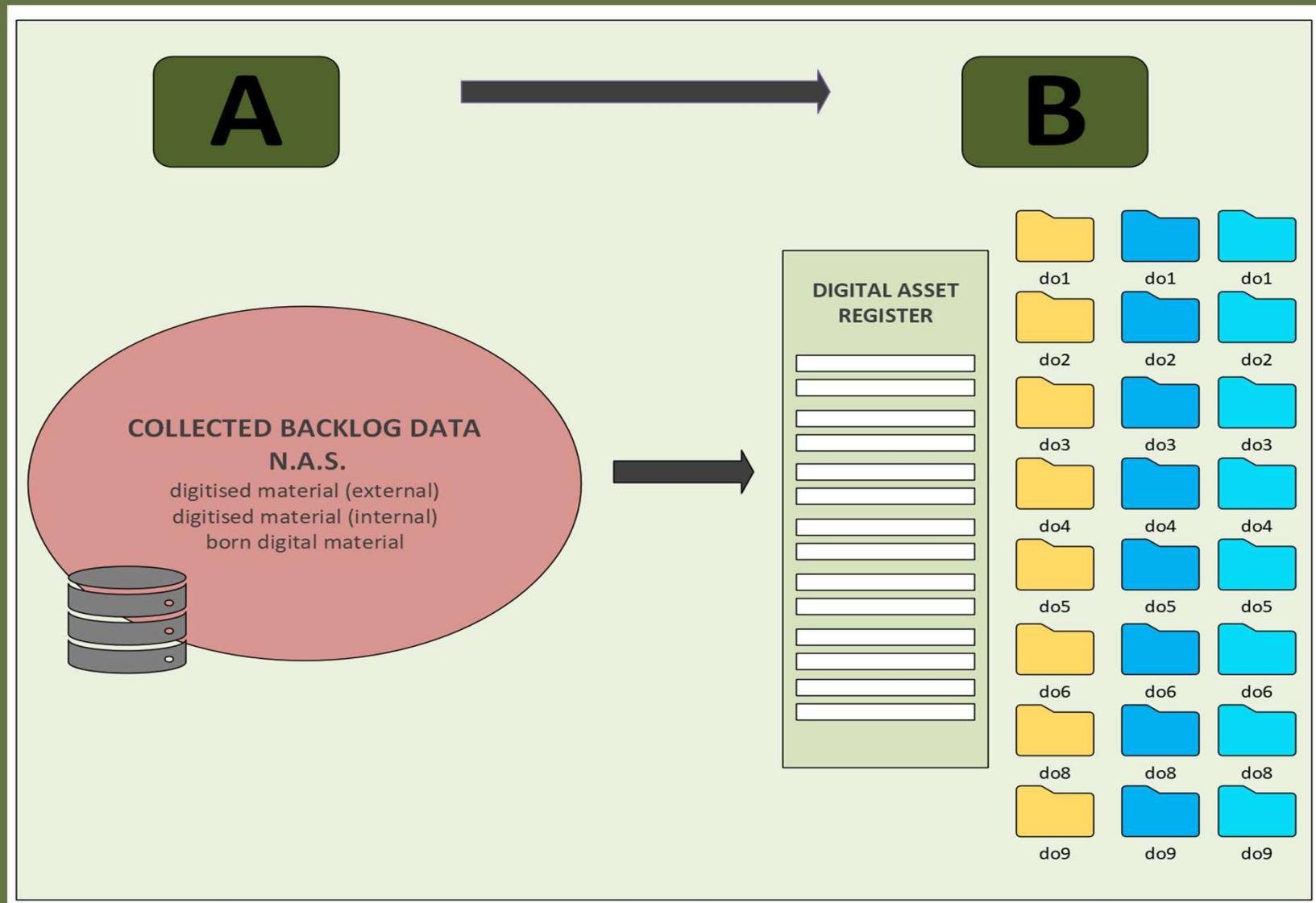
10TB



Purpose of Workflow

- **Collections Driven Approach to Decisions for Preservation and Management of Digital Objects**
- **Fuller understanding of the content of the collections to inform responsible decision making**
- **Buy Time to learn more about the collections before deciding on a long-term digital preservation solution.**
- **Ensure are safely and securely stored, backed-up with integrity and cared for in the short to medium-term.**





LEGACY DATA

ASSESS

PACKAGING

BACK-UP

REGISTRATION

NAS

UCC-Library / UCC-Library-DigiPres

<> Code

Issues

Pull requests

Discussions

Actions

Projects

Security

Insights

UCC-Library-DigiPres

Public

Watch 1

Fork 0

Star 1

main

Go to file

Code

About

Abhi21298

return to formats

9b4d3f7 · 4 months ago

bash_scripts	Updated the bash readme file	5 months ago
.DS_Store	UCC library scripts	6 months ago
.gitignore	Adding pdf2csv and updates pushed ...	5 months ago
LICENSE.md	Update and rename LICENSE to LICE...	5 months ago
README.md	Update README.md	5 months ago
av_format_mapper.csv	add formats to format mappers	4 months ago
folder_summary.py	Appending changes	5 months ago
image_format_mapper.csv	JHOVE added to ip_creator.py	5 months ago
ip_creator.py	added .pdf format to jhove option	4 months ago
logger.py	log updates	5 months ago
metadata_extractor.py	renamed other parameter to other_su...	5 months ago
other_format_mapper.csv	return to formats	4 months ago
pdf2csv.py	Adding pdf2csv and updates pushed ...	5 months ago
remove.py	Appending changes	5 months ago
search_duplicates.py	Adding pdf2csv and updates pushed ...	5 months ago

README

MIT license

No description or website provided.

digital-humanities

digital-preservation

oais

fixity

digipres

information-package

Readme

MIT license

Activity

1 star

1 watching

0 forks

Report repository

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

Contributors 3

Abhi21298

Abhijeet Madhus...

UCC-Library

UCC-Library

LEGACY DATA

ASSESS

PACKAGING

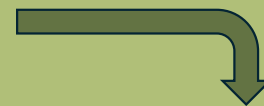
BACK-UP

REGISTRATION

>
fixity



NAS



- .EXTENSION
- SIZE
- COUNT



>
folder_
summary
.py

```
ucc_library — zsh — 125x41
Last login: Thu Jan 30 12:46:02 on ttys006
/Users/librarysystems/.zprofile:6: no such file or directory: /opt/homebrew/bin/brew
librarysystems@Librarys-iMac ucc_library % python3 folder_summary.py -i /Users/librarysystems/Desktop/legacy_server
Bad files removed - exiting process
*****
Timestamp - 2025-01-30T12:46:36

        Directory details of - /Users/librarysystems/Desktop/legacy_server

Subdirectories count - 20

File-type                File-Counts                Total-Size
.xml                      1                          0.003643035888671875 MB
.md5                      2                          0.0062389373779296875 MB
.jpg                      57                         419.18346405029297 MB
.csv                      25                         0.09747600555419922 MB
.txt                      20                         0.095245361328125 MB
.psd                      4                          88.70683097839355 MB
.tif                      68                         2922.1823863983154 MB
.xlsx                     2                          0.018718719482421875 MB
unknown                   43                         9.521719932556152 MB
.zip                      1                          6.098291397094727 MB
.html                     1                          0.005387306213378906 MB

librarysystems@Librarys-iMac ucc_library %
```



```

ucc_library — zsh — 125x41
Last login: Thu Jan 30 12:46:02 on ttys006
/Users/librarysystems/.zprofile:6: no such file or directory: /opt/homebrew/bin/brew
librarysystems@Librarys-iMac ucc_library % python3 folder_summary.py -i /Users/librarysystems/Desktop/legacy_server
Bad files removed - exiting process
*****
Timestamp - 2025-01-30T12:46:36

        Directory details of - /Users/librarysystems/Desktop/legacy_server

Subdirectories count - 20

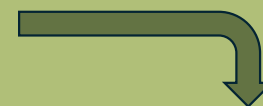
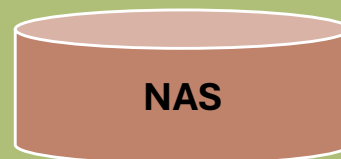
        File-type                File-Counts                Total-Size
        .xml                      1                          0.003643035888671875 MB
        .md5                      2                          0.0062389373779296875 MB
        .jpg                      57                         419.18346405029297 MB
        .csv                      25                         0.09747600555419922 MB
        .txt                      20                         0.095245361328125 MB
        .psd                      4                          88.70683097839355 MB
        .tif                      68                         2922.1823863983154 MB
        .xlsx                     2                          0.018718719482421875 MB
        unknown                   43                         9.521719932556152 MB
        .zip                      1                          6.098291397094727 MB
        .html                     1                          0.005387306213378906 MB

librarysystems@Librarys-iMac ucc_library %

```

LEGACY DATA

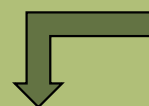
>_
fixity



ASSESS

>_
brunnhilde

>
metadata_
extractor
.py



- .EXTENSION
- SIZE
- COUNT



=

>_
folder_
summary
.py

PACKAGING

>_
brunnhilde
jhove



>_
exiftool
mediainfo
fixity

>_
ip_creator
.py

=

information
package

- IDENTIFICATION
- TECHNICAL
- CREATION



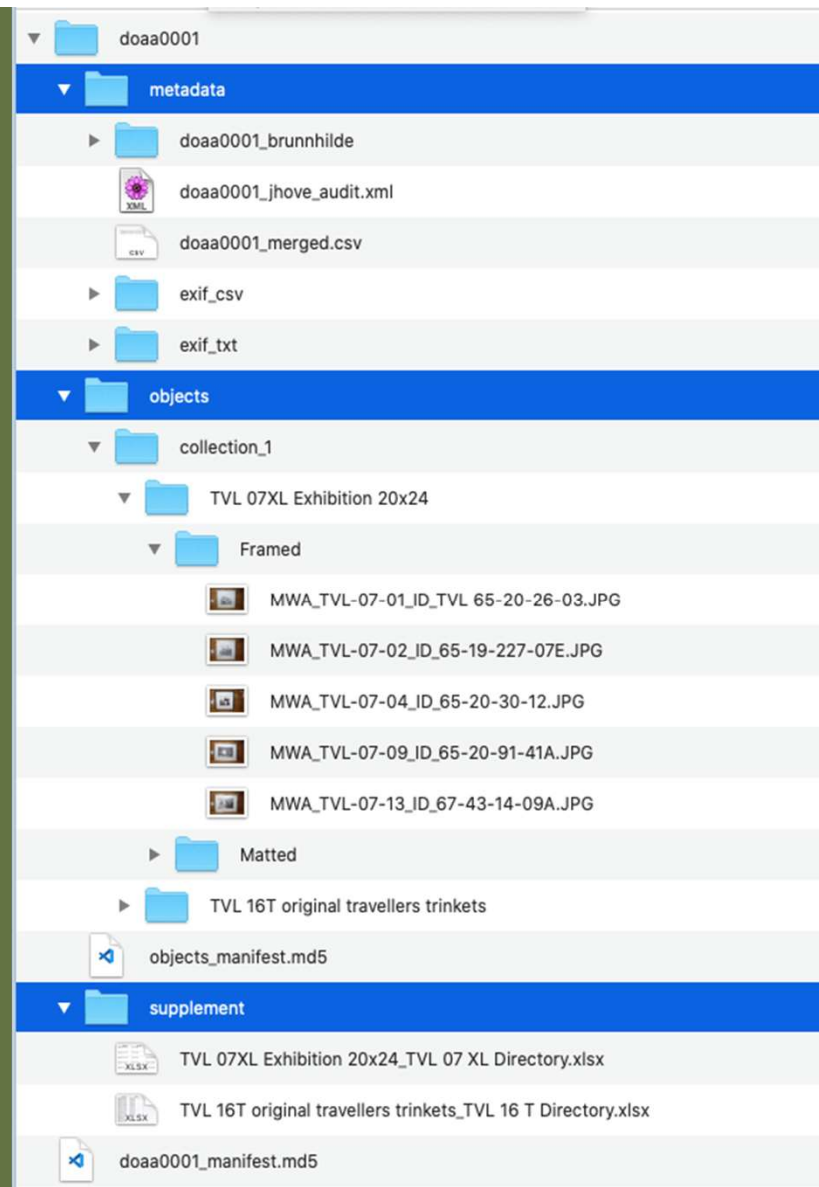
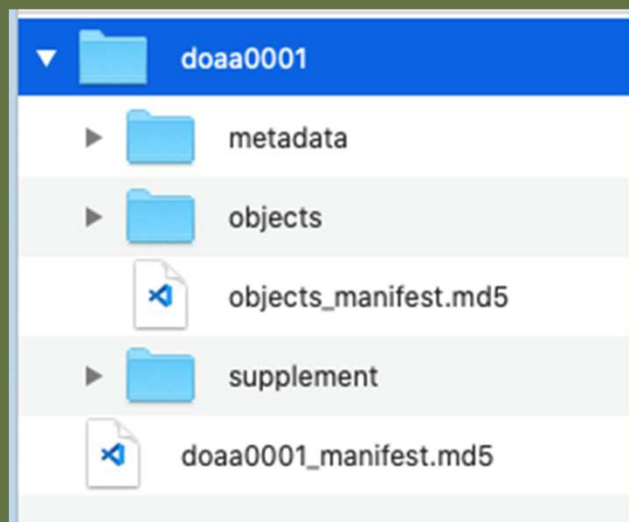
BACK-UP

REGISTRATION

```
ucc_library — zsh — 88x33
Last login: Mon Feb  3 09:56:50 on ttys058
/Users/librarysystems/.zprofile:6: no such file or directory: /opt/homebrew/bin/brew
librarysystems@Librarys-iMac ucc_library % python3 ip_creator.py -h
usage: ip_creator.py [-h] -i I -format FORMAT [-uid UID] -o O
                    [-supplement SUPPLEMENT] -kfs {y,n} [-jhove {y,n}]
                    [-brunnhilde {y,n}] [-other_sup OTHER_SUP]

this script is an information package creator. It packages all files belonging
to a specific format of interest into a folder and creates manifests for the
packaged file. Additionally, metadata is extracted and stored in csv and txt
files for reference. Optionally, supplementary files could be stored as well.
Written by, Abhijeet Rao, UCC 2023-2024

options:
  -h, --help            show this help message and exit
  -i I                  Full path of input directory
  -format FORMAT        Enter the format you would like to package
  -uid UID              Enter the destination uid name you would like to
                        assign
  -o O                  Full path of output directory to place the uid package
  -supplement SUPPLEMENT
                        Enter the supplementary formats you would like to
                        preserve
  -kfs {y,n}            (KFS - Keep folder structure) : Enter your choice on
                        preserving directory structure for the objects in the
                        destination
  -jhove {y,n}          Enter your choice on using 'jhove' utility IF
                        available
  -brunnhilde {y,n}     Enter your choice on using 'brunnhilde-ClamAV' utility
                        IF available
  -other_sup OTHER_SUP Enter the additional directory/file to be copied from
                        a different source to the destination
librarysystems@Librarys-iMac ucc_library %
```



LEGACY DATA

>_
fixity

NAS

ASSESS

>_
brunnhilde

>
metadata_
extractor
.py

- .EXTENSION
- SIZE
- COUNT

>_
folder_
summary
.py

PACKAGING

>_
brunnhilde
jhove



>_
exiftool
mediainfo
fixity

>_
ip_creator
.py



information
package

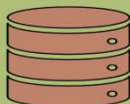
- IDENTIFICATION
- TECHNICAL
- CREATION

BACK-UP

3

2

1



>_
copy
fixity

REGISTRATION

DAR		

Description ▾	dor uid ▾	Ref numb... ▾	Collection ▾	Quantity ▾	Distinct Files ▾	file format ▾	Media Type ▾	size GB ▾	source location ▾	library copy locat
Walter Pfeiffer Colle...	doaa0005		Walter Pfeiffer	63		tiff	image	4.2	oc0003	oc0001
John Minihan Collec...	doaa0006	IE/BL/P/JM	John Minihan	88	85	tiff	image	2	FILE-SHARES/LIBRARY/LIB1	oc0001
John Minihan Collec...	doaa0007	IE/BL/P/JM	John Minihan	181		jpeg	image	0.006	FILE-SHARES/LIBRARY/LIB1	oc0001
John Minihan Collec...	doaa0008	IE/BL/P/JM	John Minihan	1		mp4	audio, video	0.003	FILE-SHARES/LIBRARY/LIB1	oc0001
Bantry TIFF Master I...	doaa0009	IE/BL/EP/B	Bantry Estate	184	184	tiff	image	10	FILE-SHARES/LIBRARY/LIB1/MASTER FILES	oc0001
BANTRY ESTATE CO...	doaa0010	IE/BL/EP/B	Bantry Estate	184		jpeg	image	0.04	FILE-SHARES/LIBRARY/LIB1/MASTER FILES	oc0001
James Barry Commo...	doaa0011	U.333	Strongroom	442		jpeg	image	0.36	FILE-SHARES/LIBRARY/LIB1/MASTER FILES	oc0001

0.015

FI

2

FI

8

FI

0.56

FI

0.17

FI

0.07

FI

1

FI

0.39

FI

0.5

FI

0.413

FI

0.024

FI

0.776

FI

1.2

FI

10

FI

606

FI

0.0855

FI

☑ back-up copy 3 location

oc0015

📅 date entered *

November 08, 2024

☑ entered by *

Casey, Raelene (2023-)

≡ additional information


updated ip 20241108

≡ questions

Enter value here

📎 Attachments

doaa0006_report.html



Be the first one to add a comment

Brunnhilde HTML report

[Provenance](#) | [Statistics](#) | [Virus report](#) | [File formats](#) | [Versions](#) | [MIME types](#) | [Dates](#) | [Unidentified](#) | [Errors](#) | [Duplicates](#)

Provenance

Input source (directory or disk image): /Users/librarysystems/Desktop/legacy_server/georgeboole/The Laws of Thought

Accession/identifier: brunoputput2

Brunnhilde version: brunnhilde 1.9.6

Siegfried version: siegfried 1.11.1 /usr/local/Cellar/siegfried/1.11.1/share/siegfried/default.sig (2024-06-21T16:01:11+10:00)
identifiers: - pronom: DROID_SignatureFile_V118.xml; container-signature-20240501.xml

Siegfried command: sf -csv -hash md5 "/Users/librarysystems/Desktop/legacy_server/georgeboole/The Laws of Thought" > "/Users/librarysystems/Desktop/legacy_server/georgeboole/brunoputput2/siegfried.csv"

Scan started: 2025-02-03 10:31:27.250946

Statistics

Total files: 10

Total size: 378 MB

Distinct files: 8

Duplicates: 2 (of 2 distinct files) ([see list](#))

Empty (zero byte) files: 0

Years (last modified): 2015 - 2015

Earliest date: 2015-06-02T11:27:05+01:00

Latest date: 2015-06-02T11:29:35+01:00

Identified file formats: 1

Unidentified files: 0

Siegfried errors: 0

Virus report

----- SCAN SUMMARY -----

Known viruses: 8704106

Engine version: 1.4.1

Scanned directories: 1

Brunnhilde HTML report

[Provenance](#) | [Statistics](#) | [Virus report](#) | [File formats](#) | [Versions](#) | [MIME types](#) | [Dates](#) | [Unidentified](#) | [Errors](#) | [Duplicates](#)

File formats

Format	ID	Count
Tagged Image File Format	fmt/353	10

File format versions

Format	ID	Version	Count
Tagged Image File Format	fmt/353		10

MIME types

MIME type	Count
image/tiff	10

Last modified dates by year

Year Last Modified	Count
2015	10

Unidentified

None found.

Errors

None found.

Duplicates

Duplicates are grouped by hash value.

Files matching hash **38b79b001bab1e8cbc9ff76034fbc874**:

- **Size:** 39332348 bytes (38 MB)
- **ID:** [fmt/353](#)
- **Format:** Tagged Image File Format
- **MIME type:** image/tiff
- **Basis for ID:** extension match tif; byte match at 0, 4 (signature 1/2)

Filename	Date modified
/Users/librarysystems/Desktop/legacy_server/georgeboole/The Laws of Thought/00010.TIF	2015-06-02T11:27:51+01:00
/Users/librarysystems/Desktop/legacy_server/georgeboole/The Laws of Thought/0005.TIF	2015-06-02T11:27:51+01:00

Files matching hash **bfe3659d50265199698b54682f391b71**:

</

Description ▾	dor uid ▾	Ref numb... ▾	Collection ▾	Quantity ▾	Distinct Files ▾	file format ▾	Media Type ▾	size GB ▾	source location ▾	library copy locat
Walter Pfeiffer Colle...	doaa0005		Walter Pfeiffer	63		tiff	image	4.2	oc0003	oc0001
John Minihan Collec...	doaa0006	IE/BL/P/JM	John Minihan	88	85	tiff	image	2	FILE-SHARES/LIBRARY/LIB1	oc0001
John Minihan Collec...	doaa0007	IE/BL/P/JM	John Minihan	181		jpeg	image	0.006	FILE-SHARES/LIBRARY/LIB1	oc0001
John Minihan Collec...	doaa0008	IE/BL/P/JM	John Minihan	1		mp4	audio, video	0.003	FILE-SHARES/LIBRARY/LIB1	oc0001
Bantry TIFF Master I...	doaa0009	IE/BL/EP/B	Bantry Estate	184	184	tiff	image	10	FILE-SHARES/LIBRARY/LIB1/MASTER FILES	oc0001
BANTRY ESTATE CO...	doaa0010	IE/BL/EP/B	Bantry Estate	184		jpeg	image	0.04	FILE-SHARES/LIBRARY/LIB1/MASTER FILES	oc0001
James Barry Commo...	doaa0011	U.333	Strongroom	442		jpeg	image	0.36	FILE-SHARES/LIBRARY/LIB1/MASTER FILES	oc0001

0.015 FI
2 FI
8 FI
0.56 FI
0.17 FI
0.07 FI
1 FI
0.39 FI
0.5 FI
0.413 FI
0.024 FI
0.776 FI
1.2 FI
10 FI
606 FI
0.0855 FI

☒ back-up copy 3 location

oc0015

☐ date entered *

November 08, 2024

☒ entered by *

Casey, Raelene (2023-)

☐ additional information


updated ip 20241108

☐ questions

Enter value here

☐ Attachments

doaa0006_report.html



Be the first one to add a comment

LEGACY DATA

>_
fixity

NAS

ASSESS

>_
brunnhilde

>
metadata_
extractor
.py

- .EXTENSION
- SIZE
- COUNT

>_
folder_
summary
.py

PACKAGING

>_
brunnhilde
jhove



>_
exiftool
mediainfo
fixity

>_
ip_creator
.py

information
package

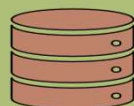
- IDENTIFICATION
- TECHNICAL
- CREATION

BACK-UP

3

2

1



>_
copy
fixity

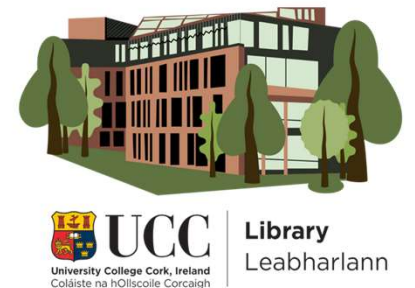
REGISTRATION

DAR		

Digital Archive

Next Steps

- **Appraisal**
- **Assess Risk**
- **Advocacy**
- **Born Digital submissions workflow**
- **Informed Requirements Gathering for Digital Preservation Solution**





Library
Leabharlann

Information

rcasey@ucc.ie

github.com/UCC-Library/UCC-Library-DigiPres

dpconline.org/blog/wdpd/blog-niamh-murphy-wdpd2024

ifiscripts.readthedocs.io/en/latest/index.html

Thank you to Abhijeet Rao, M.Sc. Data Science and Analytics, UCC 2023-2024