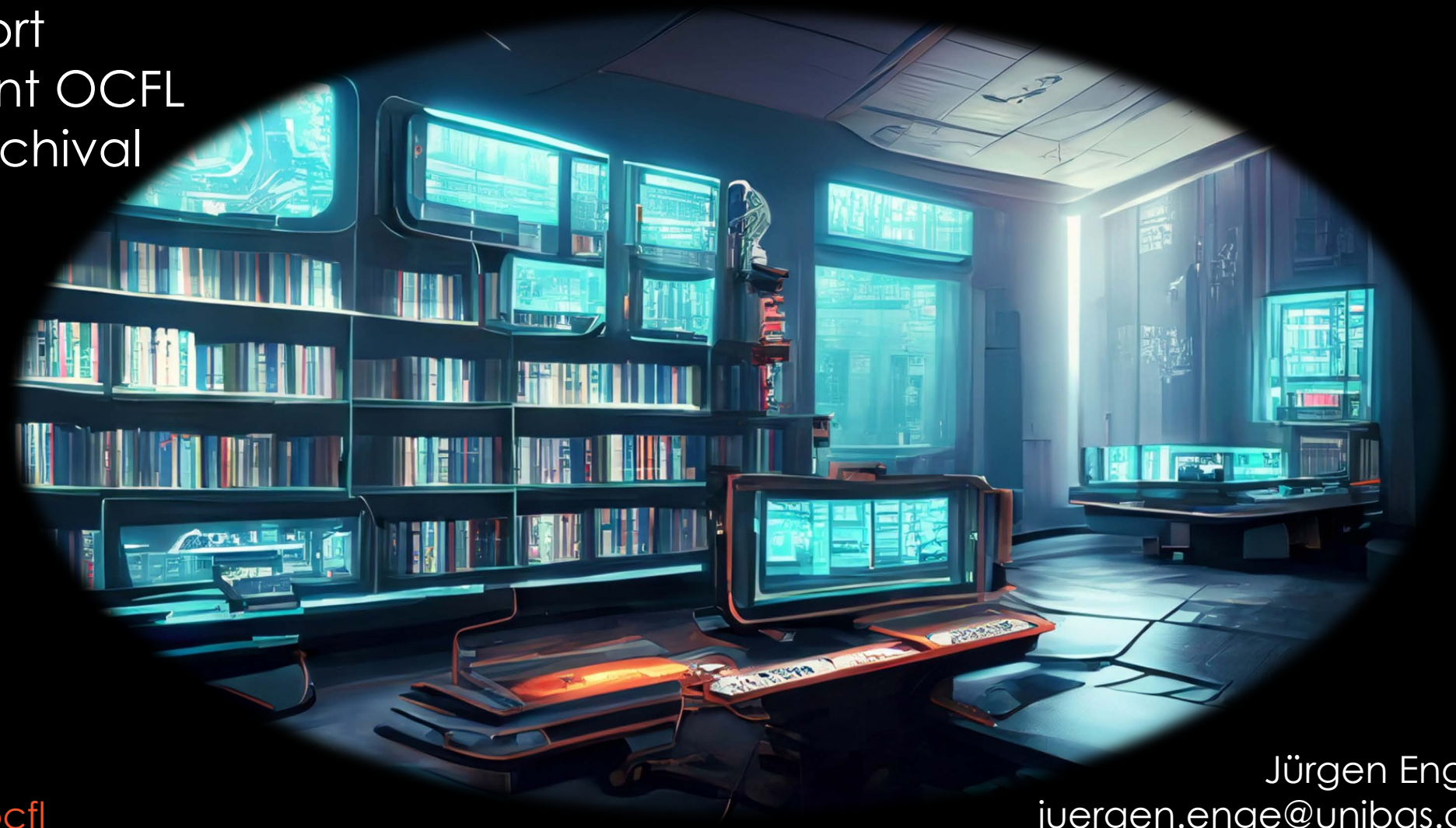
Universität Basel

GOCFL as a software approach to support standards-compliant OCFL extensions in the archival context



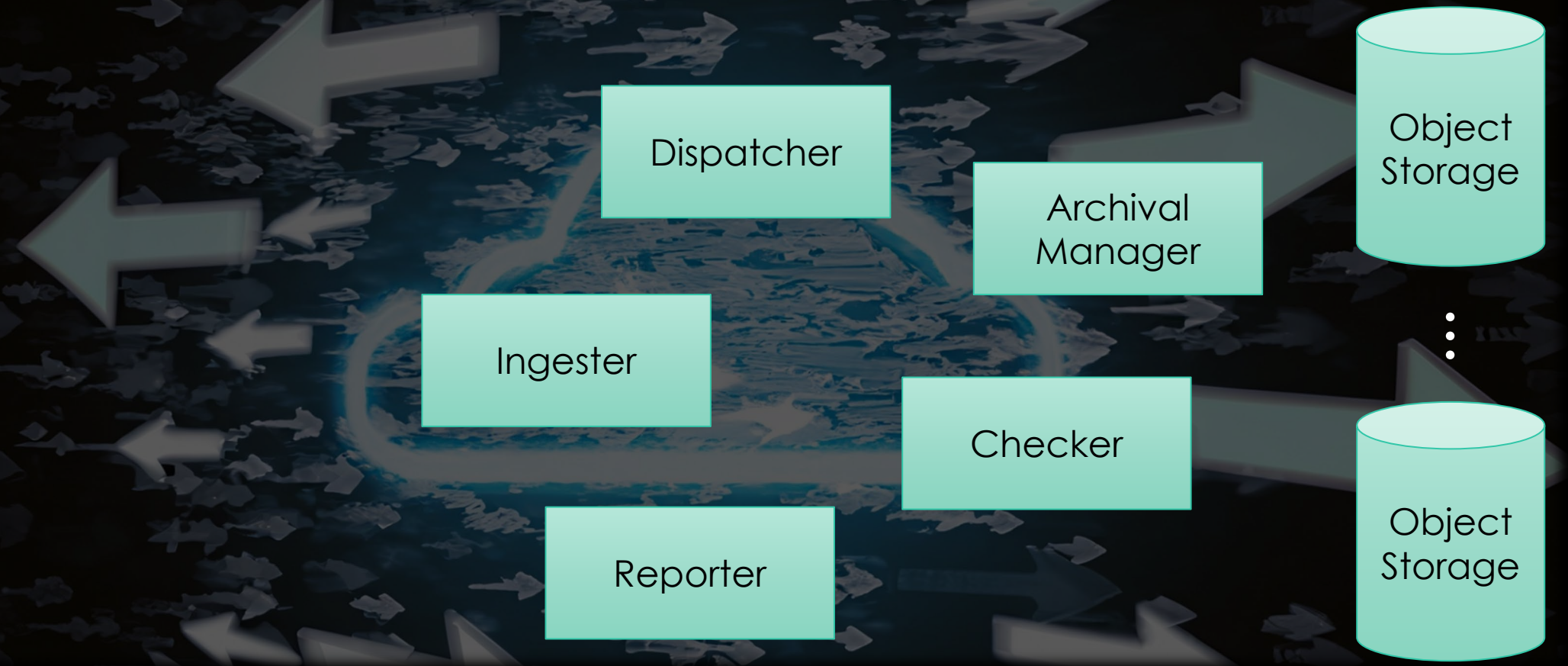
https://github.com/je4/gocfl

Jürgen Enge
juergen.enge@unibas.ch

Universität Basel

Starting point

The current generation of digital (long-term) archives is moving towards microservice architectures using cloud infrastructures.



Universität Basel

Starting point

Assuming the management and administration system is in place, the Archival Information Package (AIP) must also meet modern requirements.

Universität Basel

AIP – OCFL

Common AIP requirements face data, metadata, preservation information, ...
This presentation will focus on the benefits that are natively included in the OCFL standard, such as:

OCFL Object
AIP

- **Completeness**, so that a repository can be rebuilt from the files it stores
- **Parsability**, both by humans and machines, to ensure content can be understood in the absence of original software
- **Robustness** against errors, corruption, and migration between storage technologies
- **Versioning**, so repositories can make changes to objects allowing their history to persist
- **Storage** diversity, to ensure content can be stored on diverse storage infrastructures including conventional filesystems and cloud object stores

https://ocfl.io

Universität Basel

OCFL

Based on the OCFL standard, AIP features can be embedded in OCFL extensions. Following the OAIS standard, and based on experience of preservation needs and brain drain issues, our aim was to meet the following requirements:

OCFL Object
AIP

- Automatic recognition of technical metadata
- Minimal Semantic Metadata
- Safe path and filenames
- Migrations using versioning
- Self containment
- Self documentation
- ZIP container (with optional encryption)
- Reporting

Universität Basel

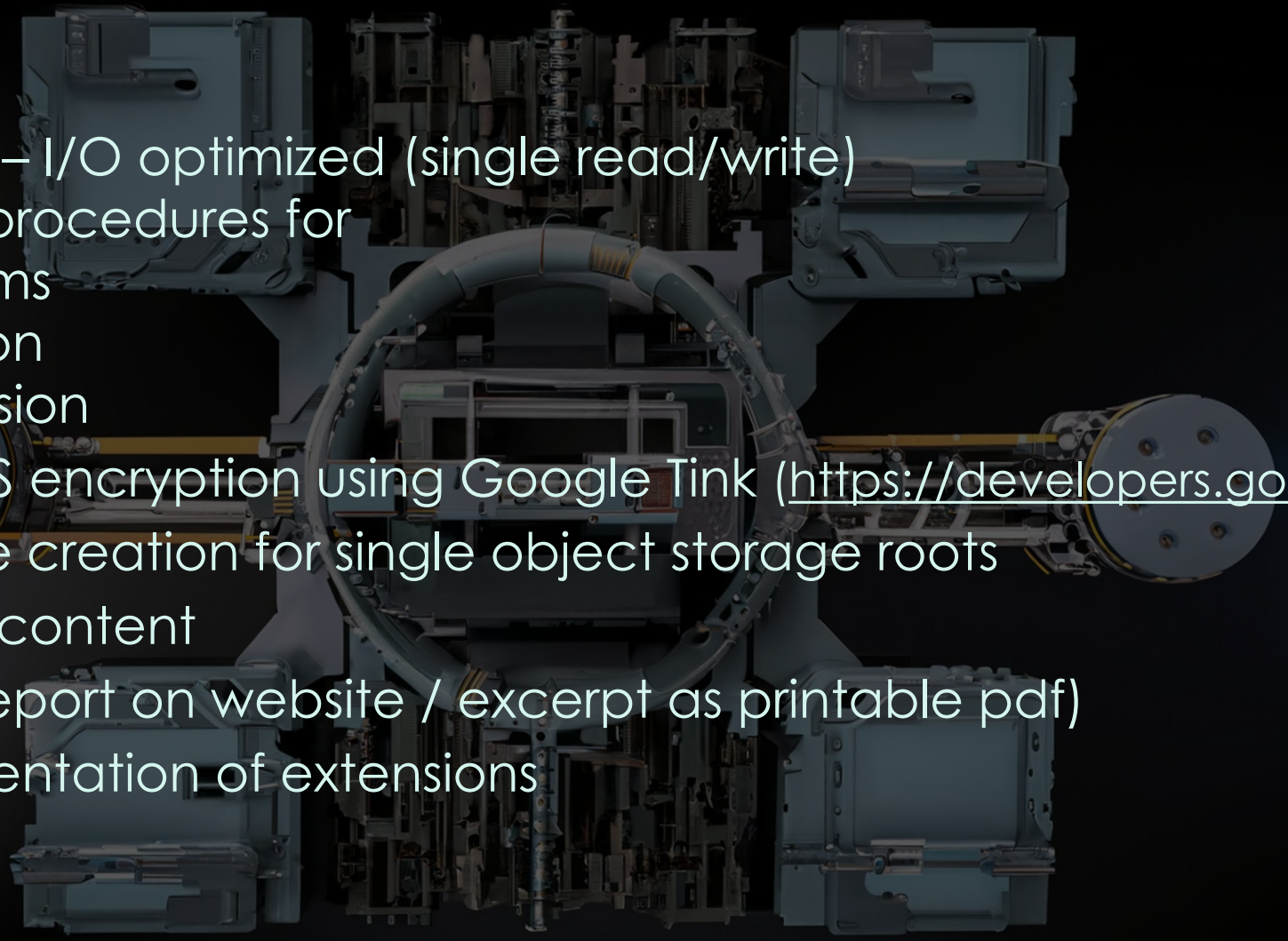
OCFL - gocfl

```
Jürgen Enge (University Library Basel, juergen@info-age.net)
Version v1.0-beta.8

Usage:
  gocfl [flags]
  gocfl [command]

Available Commands:
  add         adds new object to existing ocfl structure
  completion  Generate the autocompletion script for the specified shell
  create      creates a new ocfl structure with initial content of one object
  display     show content of ocfl object in webbrowser
  extract     extract version of ocfl content
  extractmeta extract metadata from ocfl structure
  help        Help about any command
  init        initializes an empty ocfl structure
  stat        statistics of an ocfl structure
  update      update object in existing ocfl structure
  validate    validates an ocfl structure

Flags:
      --config string                config file (default is $HOME/.gocfl.toml)
  -h, --help                         help for gocfl
      --log-file string              log output file (default is console)
      --log-level string             log level (CRITICAL|ERROR|WARNING|NOTICE|INFO|DEBUG) (default "ERROR")
      --s3-access-key-id string      Access Key ID for S3 Buckets
      --s3-endpoint string           Endpoint for S3 Buckets
      --s3-region string             Region for S3 Access
      --s3-secret-access-key string  Secret Access Key for S3 Buckets
      --with-indexer                 starts indexer as a local service

Use "gocfl [command] --help" for more information about a command.
```

gocfl is a cross-platform command line tool for creating OCFL standard compliant objects.

Universität Basel

OCFL - gocfl

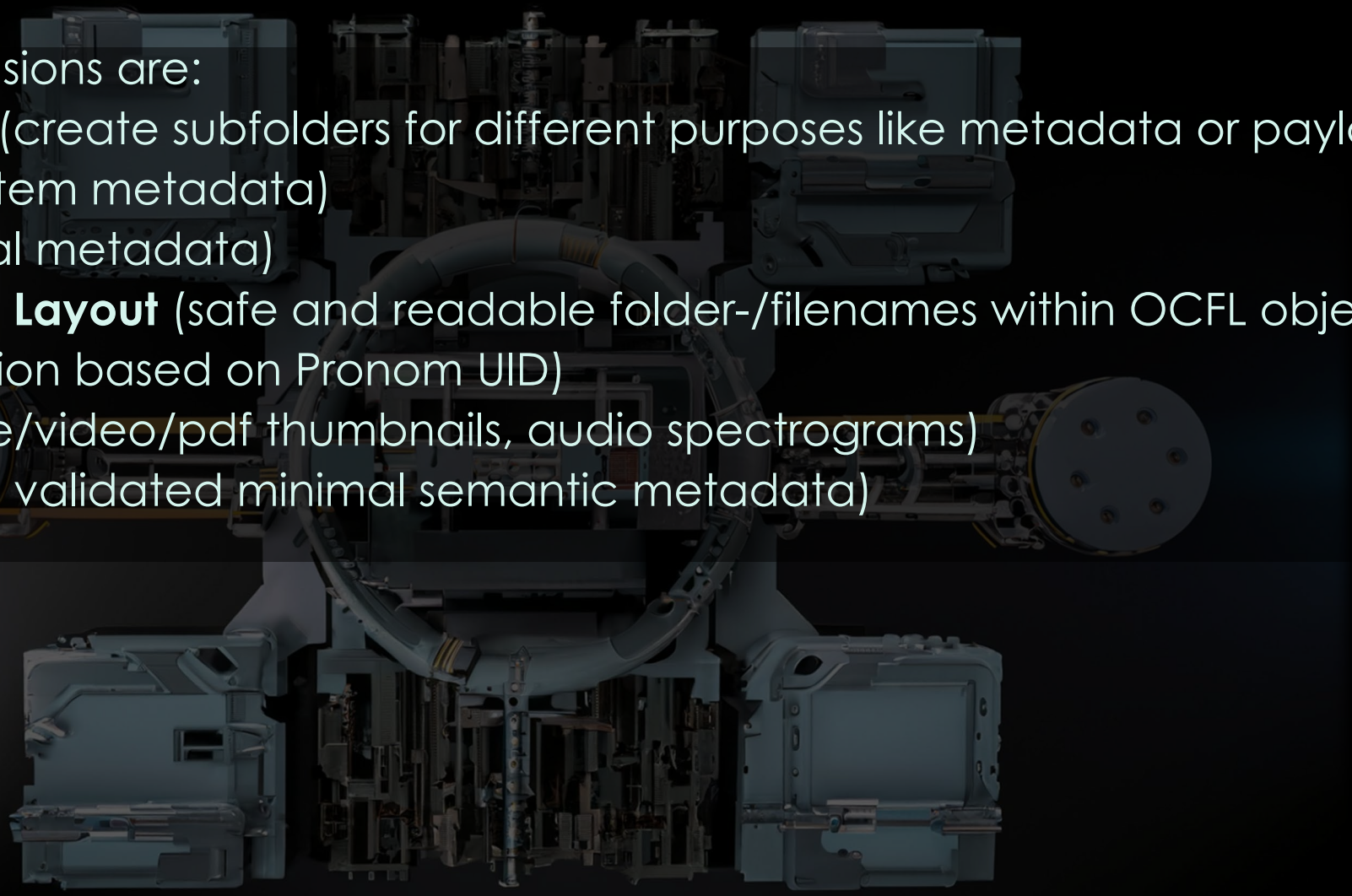


gocfl allows:

- Multi-Threading – I/O optimized (single read/write)
  - concurrent procedures for
    - checksums
    - encryption
    - compression
- Encryption – AES encryption using Google Tink (https://developers.google.com/tink)
- One step ZIP-File creation for single object storage roots
- Web Access to content
- Reporting (full report on website / excerpt as printable pdf)
- Internal documentation of extensions

Universität Basel

OCFL – gocfl - extensions



Current gocfl extensions are:

- **Content subpath** (create subfolders for different purposes like metadata or payload)
- **Filesystem** (filesystem metadata)
- **Indexer** (technical metadata)
- **Direct Clean Path Layout** (safe and readable folder-/filenames within OCFL object)
- **Migration** (migration based on Pronom UID)
- **Thumbnail** (image/video/pdf thumbnails, audio spectrograms)
- **Metafile** (schema validated minimal semantic metadata)

Universität Basel

OCFL – gocfl - files

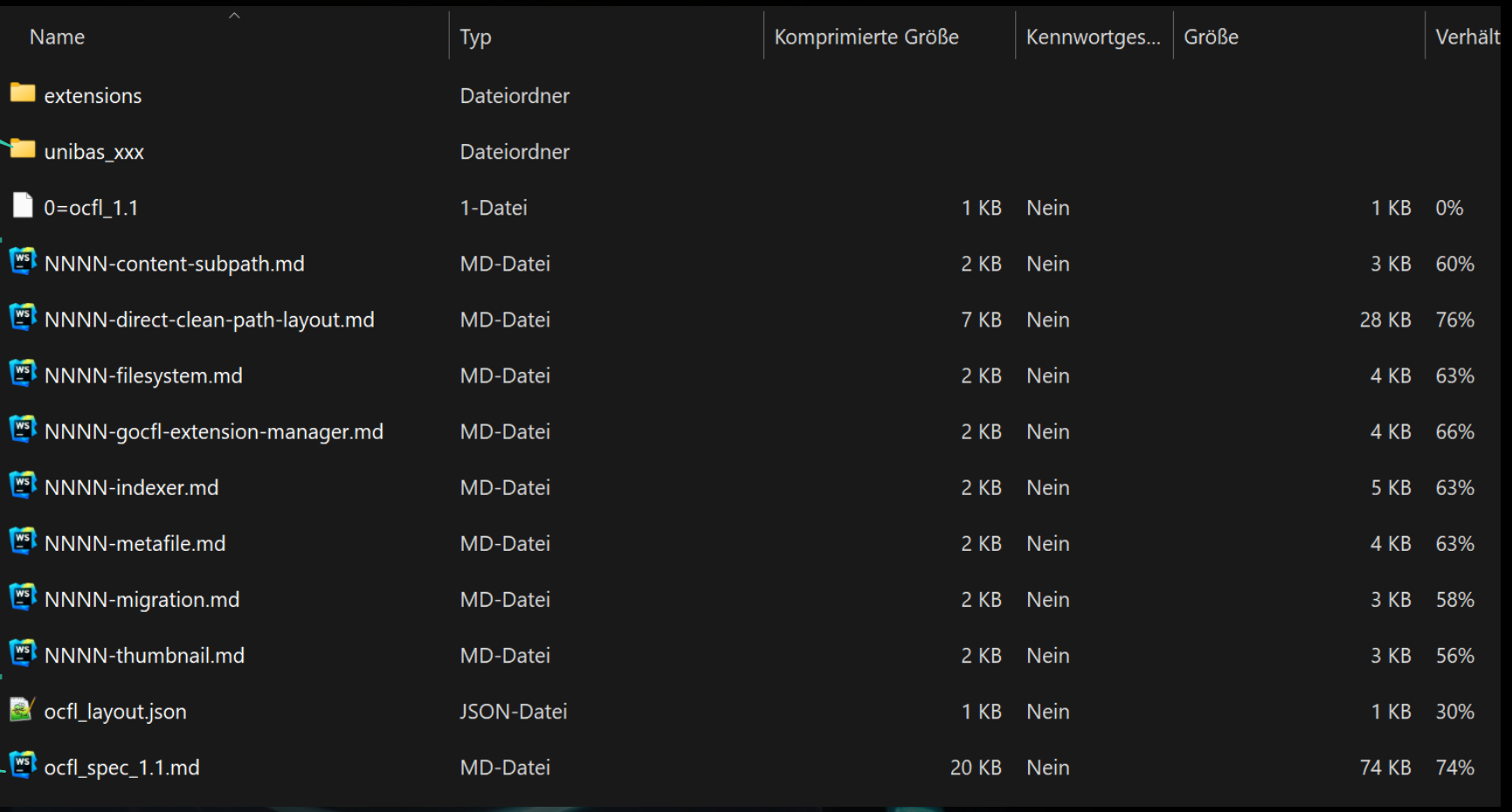
Universität Basel

OCFL – gocfl - files

Object Folder

Documentation of Extensions

OCFL Specification



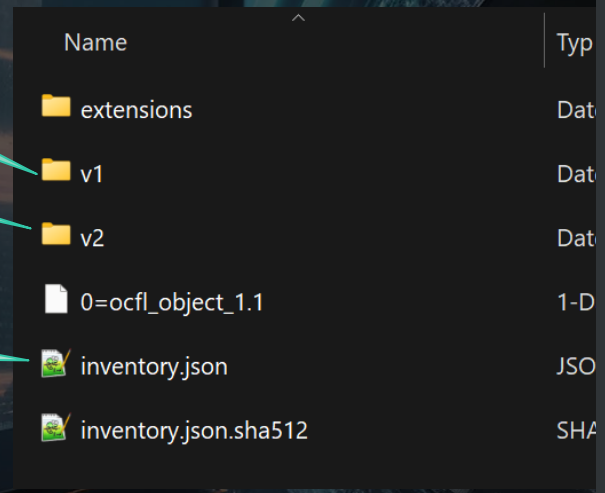
Name	Typ	Komprimierte Größe	Kennwortges...	Größe	Verhält
extensions	Dateiordner				
unibas_xxx	Dateiordner				
0=ocfl_1.1	1-Datei	1 KB	Nein	1 KB	0%
NNNN-content-subpath.md	MD-Datei	2 KB	Nein	3 KB	60%
NNNN-direct-clean-path-layout.md	MD-Datei	7 KB	Nein	28 KB	76%
NNNN-filesystem.md	MD-Datei	2 KB	Nein	4 KB	63%
NNNN-gocfl-extension-manager.md	MD-Datei	2 KB	Nein	4 KB	66%
NNNN-indexer.md	MD-Datei	2 KB	Nein	5 KB	63%
NNNN-metafile.md	MD-Datei	2 KB	Nein	4 KB	63%
NNNN-migration.md	MD-Datei	2 KB	Nein	3 KB	58%
NNNN-thumbnail.md	MD-Datei	2 KB	Nein	3 KB	56%
ocfl_layout.json	JSON-Datei	1 KB	Nein	1 KB	30%
ocfl_spec_1.1.md	MD-Datei	20 KB	Nein	74 KB	74%

Universität Basel

OCFL – gocfl - files

Versions

Inventory




```
{
  "id": "unibas:xxx",
  "type": "https://ocfl.io/1.1/spec/#inventory",
  "digestAlgorithm": "sha512",
  "head": "v2",
  "manifest": {
    "0316da5a5413ea3e1b581ba0c585407842bbb5fdfbcd0650f27f5ba3032d94a2f6d01eaf7c1578ec24ad0c19561d
      "v1/content/metadata/filesystem_v1.jsonl.gz"
    ],
    "0414c58c82fa6b5110e576070a30da75dca436738f4762da8a8e2e44c05bb6c88c43cbd1c30542a859789df9c794
      "v1/content/data/images/sample1.pef"
    ],
    "0dc7a8414ffe8ff4c0189091207e839d337412545401959cd0aa640043112941af00b79bcfa6b98f2b5f02368627
      "v1/content/data/together.png"
    ],
    "124682b5f508f169fbc1258758299ed14f62ef3c047fda94740b1d239a0d020a22f562f4b31dc047325aa2ed669e
      "v2/content/data/images/mig_sample1.cr2.png"
    ],
    "1bc4c902eb49a17dec7179f9ec1f3887d96de0f79b58e4cfe77203ea8b3cebac0debee7422334f1d9cb5a314b175
      "v1/content/metadata/indexer_v1.jsonl.gz"
    ],
    "1d497009f3aabd230bcd055ef09fd180e63e330c47f4c1afdfc36172e5421a220b78e3e3ce30c9ae533c516accfa
      "v2/content/metadata/filesystem_v2.jsonl.gz"
    ],
    "1fc86d9b9f9f4ab0a14b2032c7eb59c1e8ff8e9e93557f5f05f8933bfdf584fa8a64c7e175b731ef7a75a5532bee
      "v2/content/data/=u007Eblä=u0020blubb=u005Bin=u005D/mig_Modulhandbuch_MA_Gestaltung.pdf"
    ],
```

Universität Basel

gocfl - browse

Version 1 Information

Version 2 Information

gocfl

- Object
- Manifest
- Version v1
- Version v2
- Report

unibas:xxx

Versions

v1: Enge, Jürgen	v2: gocfl
2023-06-25 18:06:04 +0200 CEST	2023-06-25 18:07:01 +0200 CEST
mailto:juergen.enge@unibas.ch	https://github.com/je4/gocfl
Initial commit	automated version

Statistics

Quantities

Number of Files: 47

Number of different Files: 45

Size of Data: 653 MB

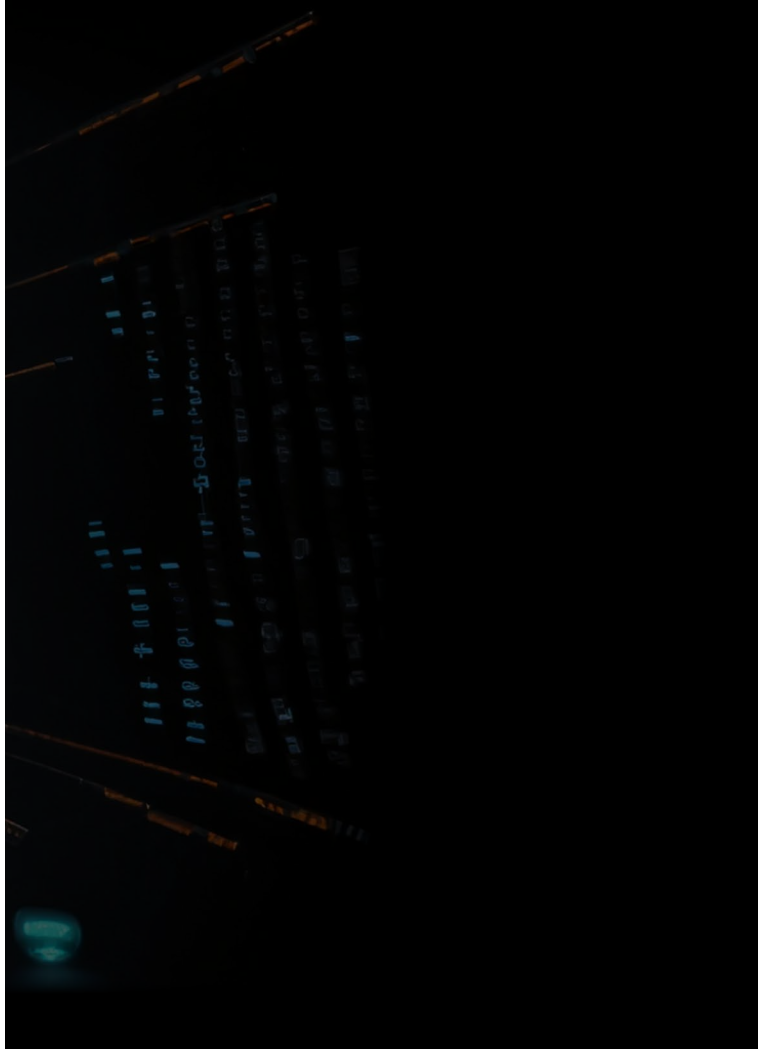
Files without size data: 7

Pronoms

- 5 fmt/11
- 1 fmt/1101
- 1 fmt/119
- 2 fmt/12

Mimetypes

- 2 application/pdf
- 1 application/vnd.oasis.opendocument.text
- 1 application/vnd.openxmlformats-officedocument.presentationml.presentation



Universität Basel

gocfl - browse

Version

- Object
- Manifest
- Version v1
- Version v2
- Report

unibas:xxx

data/2017-07-25_20-15-18_980.jpeg

669 kB | 2023-05-09T17:37:51+02:00 | windows | Archive | fmt/645 | image/jpeg | 669 kB

data/DSC_0111.JPG

759 kB | 2023-05-09T17:37:51+02:00 | windows | Archive | fmt/645 | image/jpeg | 759 kB

data/Kopie von bangbang_26_3_gemacht_V2.xlsx

1.7 MB | 2023-05-09T17:37:51+02:00 | windows | Archive | fmt/214 | application/vnd.openxmlformats-officedocument.spreadsheetml.sheet | 1.7 MB

data/README.md

data/bangbang_0044_00244_2021-0701_kakophonie-bild-def.mp4--web_master.mp4

33 MB | 2023-05-09T17:37:51+02:00 | windows | Archive | fmt/199 | video/mp4 | 33 MB

data/images/sample1.cr2

67 MB | 2023-05-09T17:40:19+02:00 | windows | Archive | fmt/592 | image/x-canon-cr2 | 67 MB

data/images/sample1.heif

2.5 MB | 2023-05-09T17:40:19+02:00 | windows | Archive | fmt/1101 | image/heic | 2.5 MB

data/images/sample1.nef

19 MB | 2023-05-09T17:40:19+02:00 | windows | Archive | fmt/202 | image/x-nikon-nef | 19 MB

Universität Basel

gocfl - browse

Checksums	
sha512	6c9fec902ab8ab76c4e8a7c1943d9b235f6d7f1232bb3ae5dfbf2e6c68e93054315647f1c257d401486b421c3eb7eeaecf27e0c4ec675cfe3f35a30bf82374e2
blake2b-384	0f6d67e3f79113dfe1d73aa58a1e144165a4c97cc5af7aee3942ce0a676e779172ea1cf2c61d9e4c0553f1dc2795dfc0
md5	21c04277358d6c9946fbc65568087356
sha256	7652c7e041b9a41b2f60ea5fbadfa2ab5299433c37d82c891924b360e90c0eec

Internal Names
v1/content/data/=u007Eblä=u0020blubb=u005Bin=u005D/Modulhandbuch_MA_Gestaltung.pdf

Direct clean path layout extension

External Names

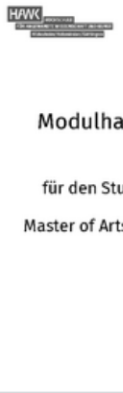
Version	Name	Time	Size	Attr	OS	Sys
v1	data/~blä blubb[in]/Modulhandbuch_MA_Gestaltung.pdf	C: 2023-05-09 17:37:51 A: 2023-06-25 18:06:34 M: 2022-05-04 13:24:51	669 kB	Archive	windows	▶CreationTime:{ HighDateTime: 31031948, LowDateTime: 925677207 } FileAttributes:32 FileSizeHigh:0 FileSizeLow:669241 ▶LastAccessTime:{ HighDateTime: 31041407, LowDateTime: 57686177 } ▶LastWriteTime:{ HighDateTime: 30957481, LowDateTime: 2454233071 }
v2	data/~blä blubb[in]/Modulhandbuch_MA_Gestaltung.pdf					

Filesystem extension

Universität Basel

gocfl - browse

Thumbnail

Thumbnail Process	PDF#01
Image	

Thumbnail extension

Indexer extension

Indexer

Mimetypes	• application/pdf
Pronom PUIDs	• fmt/18
Dimension	595 x 842px
Full Metadata	▶identify:{ frames: [...], magick: { ... } } ▼siegfried: ▶0:{ Basis: [...], Class: "Page Descriptio...", ID: "fmt/18", ... } ▶tika:[{ ... }]

▼identify:
▶frames:[{ ... }, { ... }, { ... }, ...]
▼magick:
▶image:{ backgroundColor: "#FFFFFFFFFFFF", baseDepth: 16, baseName: "gocfl_180170805...", ... }
version:"1.0"
▼siegfried:
▼0:
▶Basis:["extension match...", "byte match at [..."]
Class:"Page Description"
ID:"fmt/18"
MIME:"application/pdf"
Name:"Acrobat PDF 1.4 - Portable Document Format"
Namespace:"pronom"
Version:"1.4"
Warning:""
▼tika:
▼0:
Content-Type:"application/pdf"
PDFVersion:"1.4"
▶X-TIKA:Parsed-By:["org.apache.tika...", "org.apache.tika..."]
▶X-TIKA:Parsed-By-Full-Set:["org.apache.tika...", "org.apache.tika..."]
access_permission:assemble_document:"true"
access_permission:can_modify:"true"
access_permission:can_print:"true"
access_permission:can_print_degraded:"true"
access_permission:extract_content:"true"
access_permission:extract_for_accessibility:"true"
access_permission:fill_in_form:"true"
access_permission:modify_annotations:"true"
dc:format:"application/pdf; version=1.4"
dc:language:"x-unknown"
dcterms:created:"2022-05-03T12:47:33Z"
dcterms:modified:"2022-05-04T11:24:51Z"
language:"de"

Universität Basel

gocfl - browse

Internal Names

v2/content/data/=u007Eblä=u0020blubb=u005Bin=u005D/mig_Modulhandbuch_MA_Gestaltung.pdf

External Names

Version	Name	Time	Size	Attr	OS	Sys
v2	data/~blä blubb[in]/mig_Modulhandbuch_MA_Gestaltung.pdf					

Indexer

Mimetypes	• application/pdf
Pronom PUIDs	• fmt/477
Dimension	595 x 842px
Full Metadata	▶identify:{ frames: [...], magick: { ... } } ▶siegfried:[{ ... }] ▶tika:[{ ... }]

Migration

Migration Process	PDFA#01
Source	6c9fec902ab8ab76c4e8a7c1943d9b235f6d7f1232bb3ae5dfbf2e6c68e93054315647

Name	Acrobat PDF/A - Portable Document Format
Version	2b
Other names	
Identifiers	MIME: application/pdf PUID: fmt/477

Migration extension

Universität Basel

gocfl - report

Report

mediathek:FHNW-HGK-SNF_185436-b

Der vorliegende Report basiert auf der OCFL-Dokumentation des Datenpakets subFS(zipFS(fsFile(osFSRW(C:/temp/test)/ExpJourney.zip))/mediathek_FHNW-HGK-SNF_185436-b). Er enthält einen Auszug der Metadaten, die im Zuge der Archivierung für jede der im Datenpaket enthaltenen Dateien generiert wurden.

This report is based on the OCFL documentation of the data package subFS(zipFS(fsFile(osFSRW(C:/temp/test)/ExpJourney.zip))/mediathek_FHNW-HGK-SNF_185436-b). It contains an excerpt of the metadata generated in the course of archiving for each of the files contained in the data package.

OCFL Object Version v1

Created	2023-06-20 16:28:58
Name	Lurk, Tabea
Address	mailto:tabea.lurk@fhnw.ch
Message	Initial commit

Technology – Human – Design: Paradigms of Ubiquitous Computing - Journey

Hochschule für Gestaltung und Kunst FHNW Basel
IXDM Research Projects

Torpus, Jan
mailto:info.ixdm.hgk@fhnw.ch
2023-06-20

The research setting Journey focuses on human behavioral patterns and appropriation processes in unknown techno-social hybrid environments. We investigate and compare both perspectives: user experience and machine recognition of human behavior. To do so, we composed an interactive research setting of six separate spaces to stage paradigms of ubiComp such as: ubiquity, immediacy, invisibility, seamlessness/seamfullness and interconnectedness. By applying strategies and aesthetics from the media arts and by implementing technological sensor-actor systems, we created a measurable and debatable environment. It offers snippets of narratives that can be mentally combined and is carefully designed to enable test persons to fully engage and willingly suspend their disbelief. The bodily interference with the physical environment is intensified by an interactive garment that test persons put on before their journey. By doing so, they leave their familiar comfort zone and take on a new role in the narrative.

Alternative Titles	*Ubicombs Journey*
Signature	*FHNW-HGK-SNF_185436-b*
Organisation ID	*FHNW-HGK*
Collection ID	*IXDM-Research*
Sets	*IXDM, HGK FHNW, SNF, Forschung*
Keywords	*Arts, Medienepistemologie, Sensor-Aktor-Mapping, Internet der Dinge, Responsive Environment, Sensor-Aktor-Netzwerk, Ubiquitous Computing, Kunstforschung, Data Mapping*
Identifiers	*zotero2:2641719.S95EUAJS* *https://ubicombs.ch/journey/*
References	*FHNW-HGK-SNF_185436-a*
Deprecates	
Ingest Workflow	*WF01*

generated by gocfl / https://github.com/je4/gocfl

Metafile extension

Universität Basel

gocfl - report

Content Overview

Number of Files	526
Total Size of all files	50 GB
Total Length of A/V Media	40:06:00

Mimetype	Count	Size
application/octet-stream	4	122 MB
application/pdf	73	70 MB
application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	5	551 kB
application/zip	2	132 kB
audio/mp4	30	970 MB
audio/x-aiff	2	84 kB
audio/x-wav	9	3.0 GB
image/jpeg	4	3.3 MB
image/png	80	18 MB
image/svg+xml	6	527 kB
text/csv	7	661 MB
text/html	2	2.9 kB
text/markdown	21	126 kB
text/plain	77	2.2 MB
text/plain; charset=ISO-8859-1	73	13 MB
text/plain; charset=UTF-8	12	395 kB
text/x-csrc; charset=ISO-8859-1	5	20 kB
video/mp4	51	45 GB

File Listing

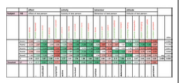
- data
  - plots
    - sense-making
      - sequence_139.png
      - sequence_182.png
      - sequence_166.png
      - sequence_184.png
      - sequence_160.png
      - sequence_174.png
      - sequence_134.png
      - sequence_175.png
      - sequence_150.png
      - sequence_106.png
      - sequence_148.png
      - sequence_146.png
      - sequence_143.png
      - sequence_102.png
      - sequence_178.png
      - sequence_111.png
      - sequence_144.png
      - sequence_151.png
      - sequence_186.png
      - sequence_108.png
      - sequence_189.png
      - sequence_133.png
      - sequence_109.png
      - sequence_177.png
      - sequence_142.png
      - sequence_115.png
      - sequence_101.png
      - sequence_167.png
      - sequence_162.png
      - sequence_171.png
      - sequence_149.png
      - sequence_112.png
      - sequence_156.png
      - sequence_152.png
      - sequence_181.png
      - sequence_104.png
      - sequence_173.png
      - sequence_107.png
      - sequence_140.png
      - sequence_188.png
      - sequence_113.png
      - sequence_176.png
      - sequence_155.png
    - EDA
      - EDA-collected-plots.pdf

Universität Basel

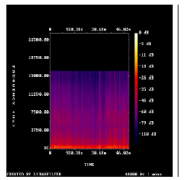
gocfl - report

Filename	data/code/src/evaluation-interface/config.default.json
Filesize	353 B
Created	2023-06-20 15:25:45
Modified	2022-02-10 20:13:24
Attributes	Normal
OS	windows

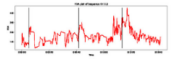
Pronom UID	fmt/13
Content-Type	image/png
Dimension	2686x658px
Filename	data/code/doc/visualisation/assets/grafana_plot_pos-xy.png
Filesize	654 kB
Created	2023-06-20 15:25:45
Modified	2021-06-22 23:07:54
Attributes	Normal
OS	windows



Pronom UID	fmt/12
Content-Type	image/png
Dimension	2654x1065px
Filename	data/ratings/tab3.png
Filesize	283 kB
Created	2023-06-20 15:26:06
Modified	2023-06-01 20:05:46
Attributes	Normal
OS	windows

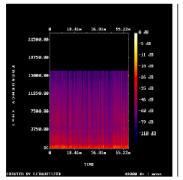


Pronom UID	fmt/199
Content-Type	audio/mp4
Duration	2945sec
Filename	data/interviews/AudioRecordings/112.m4a
Filesize	25 MB
Created	2023-06-20 15:25:51
Modified	2022-02-12 14:37:28
Attributes	Normal
OS	windows



Pronom UID	fmt/18
Content-Type	application/pdf
Dimension	864x288px
Filename	data/plots/EDA/sequence_113.pdf
Filesize	35 kB
Created	2023-06-20 15:26:06
Modified	2023-06-12 17:38:54
Attributes	Normal
OS	windows

Pronom UID	x-fmt/111
Content-Type	text/plain
Filename	data/code/src/ubicomp-max-package/package/snippets/latency-meter.maxsnip
Filesize	8.9 kB
Created	2023-06-20 15:25:46
Modified	2022-01-31 18:46:32
Attributes	Normal
OS	windows



Pronom UID	fmt/199
Content-Type	audio/mp4
Duration	3534sec
Filename	data/interviews/AudioRecordings/109_1.m4a
Filesize	30 MB
Created	2023-06-20 15:25:50
Modified	2022-02-12 10:13:46
Attributes	Normal
OS	windows
Filename	data/interviews/AudioRecordings/109_2.m4a
Filesize	30 MB
Created	2023-06-20 15:25:51
Modified	2022-03-02 07:19:58
Attributes	Normal
OS	windows

Universität Basel

gocfl

To achieve a system-independent, transportable Archival Information Package (AIP), the approach presented embeds semantic, technical and preservation metadata at the extension level of OCFL, rather than superimposing this type of data on top of OCFL.

Universität Basel

Lessons learned

The use of cloud technologies such as object storage, microservices, etc. not only enables new technological solutions. They also greatly simplify collaboration by eliminating the need for centralized, monolithic development. The portable OCFL structure is an important milestone in this respect.

The core OCFL standards community has not yet seen the impact of the extensions in this form. As a result, the field is still very much uncharted. But the community is very interested.

https://github.com/je4/gocfl

Contact: juergen.enge@unibas.ch