Fedora Program and Community Update

David Wilcox, Program Leader, LYRASIS

2020 Technical Roadmap

1. Release Fedora 6
2. Release updated migration tooling
3. Test and validate application and tooling with the community

Fedora™

1. Reduce effort required to migrate

2. Enhance long term digital preservation

3. Improve performance and scale

Fedora™

Steps to Achieve Goals

1. Replace ModeShape

2. Implement the Oxford Common File Layout

3. Retain alignment with specified Fedora API

4. Release with migration tooling and support
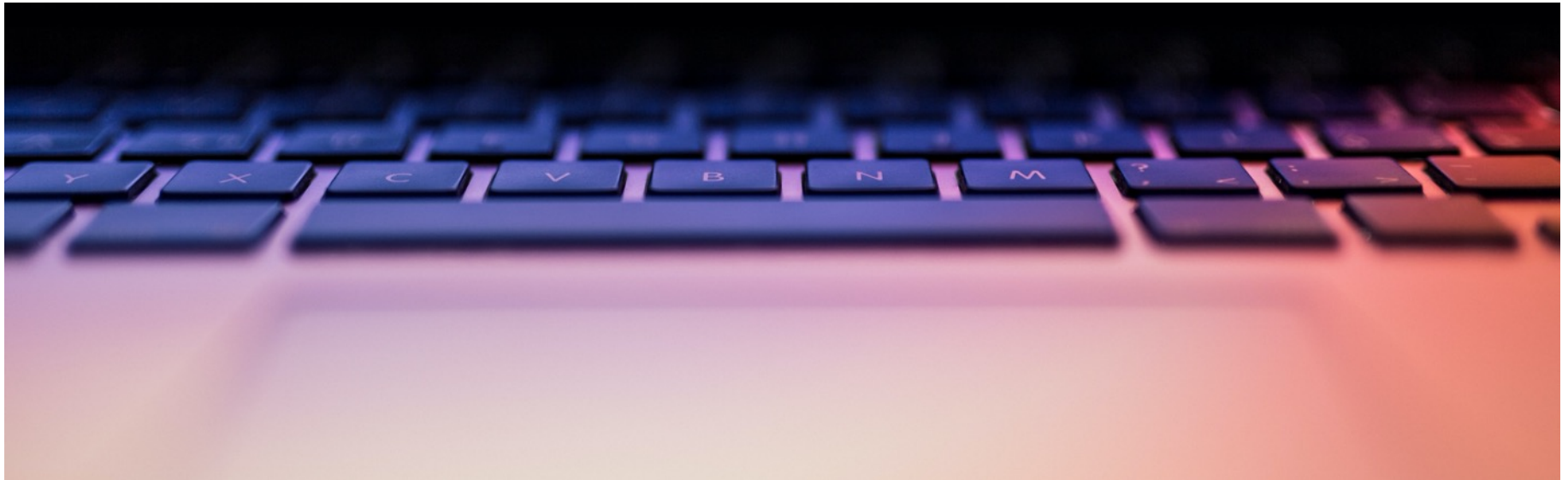
What is the Oxford Common File Layout (OCFL)?

Fedora™

A **simple, non-proprietary, specified, open-standards approach** to the layout of preservation persistence.

OCFL Offers…

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

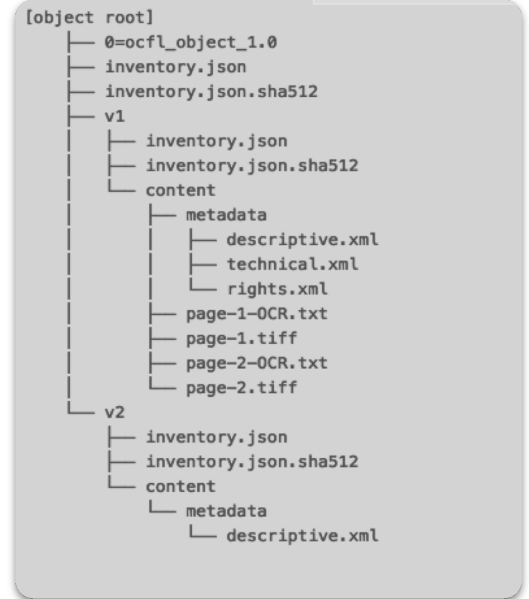**Versioning**, so repositories can make changes to objects allowing its history to persist

Storage diversity, to ensure content can be stored on diverse storage infrastructures including cloud object stores

Completeness, so that a repository can be rebuilt from the files it stores

Benefits of the OCFL: Parsability

both by humans and machines, to ensure content can be understood in the absence of original software

- In disaster recovery situations, humans should be able to understand the content.

- Machine readability allows for simple applications to be placed on top of an existing OCFL storage root.

```
[object root]
├── 0=ocfl_object_1.0
├── inventory.json
├── inventory.json.sha512
├── v1
│   ├── inventory.json
│   ├── inventory.json.sha512
│   └── content
│       ├── metadata
│       │   ├── descriptive.xml
│       │   ├── technical.xml
│       │   └── rights.xml
│       ├── page-1-OCR.txt
│       ├── page-1.tiff
│       ├── page-2-OCR.txt
│       └── page-2.tiff
└── v2
    ├── inventory.json
    ├── inventory.json.sha512
    └── content
        └── metadata
            └── descriptive.xml
```

Fedora™

Benefits of the OCFL: Robustness

against errors, corruption, and migration between storage technologies

- Strong fixity is built into OCFL.

- Content can easily be validated using the `inventory.json`.

- Objects can be completely self-contained.

Fedora™

Benefits of the OCFL: Versioning

so repositories can make changes to objects allowing its history to persist

- Changes to objects are tracked over time.
- Forward delta is employed to reduce the amount of content stored.
- Previous versions of objects can be reconstructed using the `inventory.json` file

```
},
"type": "https://ocfl.io/1.0/spec/#inventory",
"versions": {
  "v1": {
    "created": "2018-01-01T01:01:01Z",
    "message": "Initial import",
    "state": {
      "7dcc35...c31": [ "v1/content/metadata/descriptive.xml" ],
      "cf83e1...a3e": [ "v1/content/metadata/technical.xml"],
      "ffccf6...62e": [ "v1/content/metadata/rights.xml"],
      "ge72e1...d6e": [ "v1/content/page-1.tiff" ],
      "jdferd...56d": [ "v1/content/page-1-OCR.txt" ],
      "gk4er6...57d": [ "v1/content/page-2.tiff"],
      "7adjhe...4ad": [ "v1/content/page-2-OCR.txt"]
    },
    "user": {
      "address": "alice@example.com",
      "name": "Alice"
    }
  },
  "v2": {
    "created": "2018-02-02T02:02:02Z",
    "message": "Fix descriptive.xml",
    "state": {
      "4d27c8...b53": [ "v2/content/metadata/descriptive.xml" ],
    },
```

Benefits of the OCFL: Storage diversity

to ensure content can be stored on diverse storage infrastructures including cloud object stores

- Designed to work with various storage infrastructures including object stores prevalent in cloud offerings (e.g. Amazon S3).

- Supports conventional file system metaphor.

- Can be implemented to ensure deduplication of content, lowering overall storage costs.

Benefits of the OCFL: Completeness

so that a repository can be rebuilt from the files it stores

- The complete intellectual object is stored together with its metadata.

- Falls in line with Trusted Digital Repositories (TDR, ISO 16363), NDSA Levels of Preservation, and Open Archival Information Systems (OAIS).

- Allows ease of mapping from one system to another.

Many of these standards talk about what you should do, but not how. The OCFL provides the how.

Fedora™

Application-independent persistence

Ability to rebuild repository from contents on disk

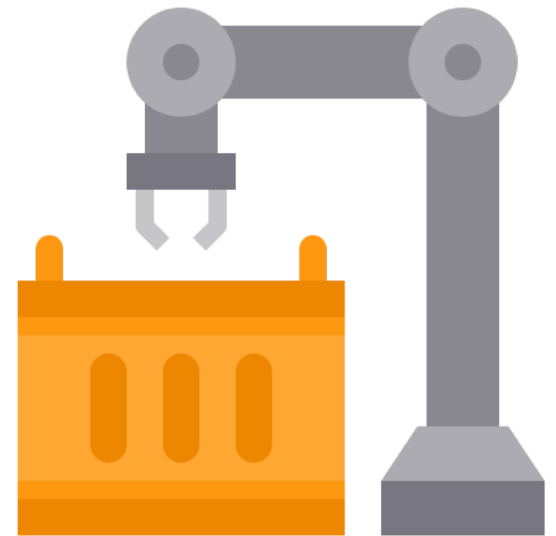Easier migrations to Fedora 6

Fewer migrations in the future

Fedora 6 Development Process

Four planned code sprints

Two in 2019, two in 2020

Working with pilot partners

Community reporting and feedback

Icon made by itim2101 from Flaticon

Fedora™

Fedora Members

Pilot Partners

Docuteam

National Library of Medicine

University of Wisconsin-Madison

Icon made by pixel-buddha from Flaticon

Fedora™

Implemented basic resource management

Upgraded migration-utils to support OCFL

Summary and demo now available

Icon made by Freepik from Flaticon

Fedora™

migration-utils

- Java command line tool

- Converts Fedora 3 data to OCFL on disk

- Simple and performant


```
dwilcox@Protagoras migration-utils % java -jar target/migration-utils-4.4.1-SNAPSHOT-driver.jar
Missing required options [--source-type=<f3SourceType>, --target-dir=<targetDir>]
Usage: migration-utils [-hrV] [--debug] -a=<targetDir> [-d=<f3DatastreamsDir>]
                       [-e=<f3ExportedDir>] [-l=<objectLimit>]
                       [-o=<f3ObjectsDir>] [-p=<pidFile>] -t=<f3SourceType>
                       [-y=<ocflLayout>]
  -h, --help                   Show this help message and exit.
  -V, --version                Print version information and exit.
  -t, --source-type=<f3SourceType>
                               Fedora 3 source type. Choices: AKUBRA | LEGACY |
                                 EXPORTED
  -d, --datastreams-dir=<f3DatastreamsDir>
                               Directory containing Fedora 3 datastreams (used
                                 with --source-type AKUBRA or LEGACY)
  -o, --objects-dir=<f3ObjectsDir>
                               Directory containing Fedora 3 objects (used with
                                 --source-type AKUBRA or LEGACY)
  -e, --exported-dir=<f3ExportedDir>
                               Directory containing Fedora 3 export (used with
                                 --source-type EXPORTED)
  -a, --target-dir=<targetDir>
                               Directory where OCFL storage root and supporting
                                 state will be written
  -y, --layout=<ocflLayout>    OCFL layout of storage root. Choices: FLAT |
                                 PAIRTREE | TRUNCATED
                                 Default: FLAT
  -l, --limit=<objectLimit>    Limit number of objects to be processed.
```

Fedora™

migration-utils

- Simple, folder-based structure

- Valid OCFL and compatible with Fedora 6

- Drop a Fedora 6 application on top to read and write
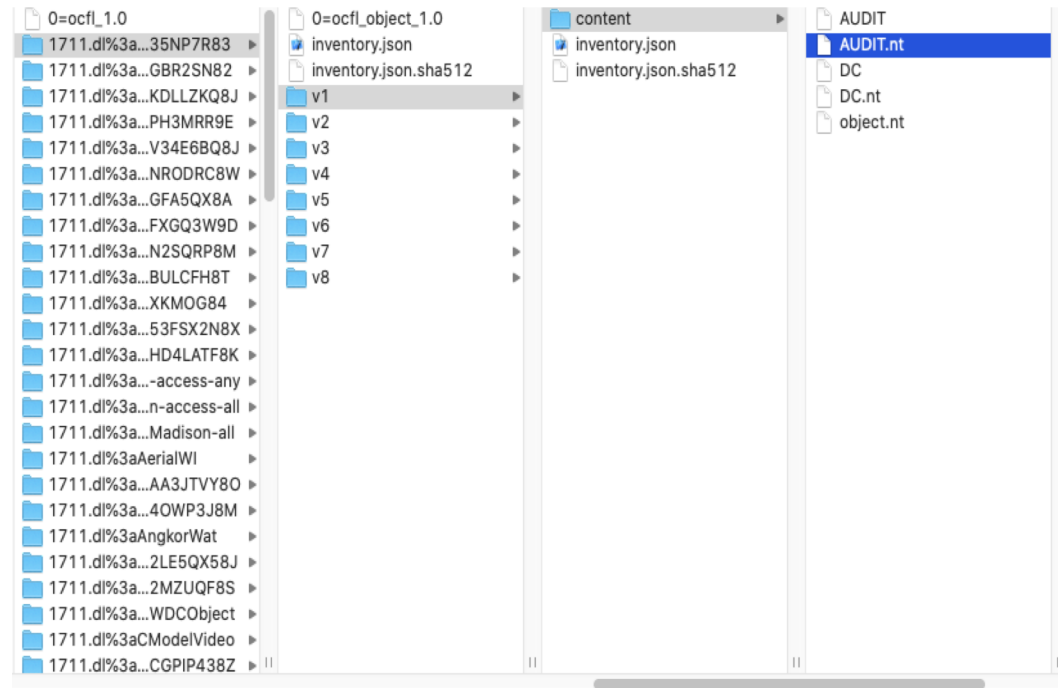
Resource Management

Create resources through the API and see OCFL on disk

```
📁 ~/GitHub/fcrepo4/fcrepo-webapp/target/ocfl-root
.
├── 0=ocfl_1.0
├── 14
│   └── 83
│       └── 65
│           └── _fedora_repository_root
│               ├── 0=ocfl_object_1.0
│               ├── extensions
│               │   └── mutable-head
│               │       ├── HEAD
│               │       │   ├── content
│               │       │   │   └── r1
│               │       │   │       └── info:fedora.nt
│               │       │   ├── inventory.json
│               │       │   └── inventory.json.sha512
│               │       ├── revisions
│               │       │   └── r1
│               │       └── root-inventory.json.sha512
│               ├── inventory.json
│               ├── inventory.json.sha512
│               └── v1
│                   ├── content
│                   ├── inventory.json
│                   └── inventory.json.sha512
├── extension-layout-n-tuple.json
├── ocfl_1.0.txt
└── ocfl_layout.json

12 directories, 14 files
dwilcox@Protagoras ocfl-root % 
```

Fedora™

Anticipated Timelines

Early testable code: Available now!

Next sprint: Week of January 27

Initial release: mid-to-late 2020

Icon made by Smashicons from
Flaticon

Fedora™

Test migration-utils and provide feedback

Participate in a code sprint

Join the conversation in Slack

Support us by becoming a member

Fedora™

please contact us for more info.

Email **david.wilcox@lyrasis.org**

David Wilcox
Fedora Program Leader