Open Preservation Foundation
and
The Preservation Action Registry

Martin Wrigley, Executive Director, OPF
Martin Wrigley

30+ years experience delivering software and solutions - mostly in Mobile Telecoms

10+ years experience of managing a membership driven open source association

OPF Executive Director since September 2017

Expanding my knowledge of the finer points of Digital Preservation
Who is OPF?

• A not for profit, global membership association providing stewardship of open-source tools for the digital preservation community.

• Founded in 2010 to sustain the results of the EU PLANETS project

• The OPF reference toolset now includes veraPDF, JHOVE and more
What is OPF’s purpose?

OPF Vision
Open sustainable digital preservation

OPF Mission
Enabling shared solutions for effective and efficient digital preservation; the Open Preservation Foundation leads a collaborative effort to create, maintain and develop the reference set of sustainable, open source digital preservation tools.

This set of tools (including software and standards) enables organisations to evaluate, validate, document, mitigate risk, and process digital content to be preserved in line with desired policies and community best practice.

Values
• Open
• Member driven
• Collaborative & Inclusive
• Innovative
Who are OPF members?

Austrian Institute of Technology
British Library
Bibliotheque Nationale de France
Goportis
International Atomic Energy Archives
Jisc
Koninklijke Bibliotheek
Det Kgl. Bibliotek
Nationaal Archief
The National Archives UK
Nasjonalbiblioteket
Rigsarkivet
Ex Libris
Rahvusarhiiv

We welcome any organisation with a mandate to preserve digital information for the long term
What does OPF do?

- Community Knowledge
- Sharing knowledge
- Develop the OPF reference toolset
- Deliver to development roadmaps
- Community engagement
- Webinars and training
- Interest Groups and Tech Clinics
- OPF Software Maturity Model
- Hosting community services e.g. COPTR
- Website, blogs, events
OPF – Digital Preservation Knowledge and Tools

Knowledge – Member and Community input and sharing

Digital Preservation – Process, Policy & Best Practice

Practical Tools
- Open Source
- Reference Toolset
OPF Reference Toolset – generic process

Thing is (or is becoming) a Submission Information Package
OPF Tool Mapping

- Transform
  - Database archiving / Extraction tools
  - Recommended by OPF
- SIARD (SQL database to XML format)
- Derivative check tools
  - Maintenance through OPF
  - xcorssound WAV, MP3
- Information Packaging tools
- Quality check tools
  - E-ARK CEF SIP validator
- Cross Check tools
- TBA
- Quality check tools
  - SIARD (SQL database to XML format)
- Container explosion recursive
- Disk image explosion/analysis
  - Recommended by OPF
  - BitCurator

- Verification and Characterisation tools
  - Maintained through OPF
  - PDF/A
  - JHOVE PDF, JPEG, WAV, PNG, WARC, AIFF, UTF8 TEXT, XML, HTML, GZIP, ASCII TEXT, MP3, GIF, JPEG2000
  - TIFF
  - (DPF Manager) TIFF module TIFF
  - JPEG2000

- Validation and Characterisation tools
  - Maintenance through OPF

- Validation polices
- Fix/transform (migrate...)
- Cross Check polices
- Fix/transform (redact...)
- Package, Quality Assurance, Review, Cross Check
- Put into a Box (turn into an AIP)

- Put into a Box
- Periodic re-check

- Identification tools
  - Maintenance through OPF
  - fido
  - Format Sniff
    - Recommended by OPF
  - DROID
  - PRONOM
  - FILE
How do OPF projects work?

**FUNDING**
- OPF membership
- Donations
- Project income

**PLANNING (PRODUCT BOARD)**
- Prioritise fixes and features
- Define the release
- Manage the roadmap

**REQUIREMENTS & COMMUNITY FEEDBACK**
- Bug reports and new feature requests
- Hack day activities
- Code contributions
- Input from OPF interest groups
- Contribution of test files

**DEVELOPMENT & TESTING**
- GitHub for OS development
- Build a set of test data
- Continuous integration
- Quality Assurance

**FINAL TEST & RELEASE**
- Production release
- Freely available to community
- Patches (essential fixes)
Preservation Action Registry
PAR Background: The problem

• Users want the best advice, wherever it comes from
  o Identification, property extraction, validation, migration, rendering, tools

• Many sources for current ‘best practice’
  o Products such as Preservica & Archivematica
  o Practitioners
  o Academics
  o Specialists
  - but they don’t talk to each other effectively
Background: Motivation and Objectives

- To provide a mechanism to exchange good practice information between organisations and preservation system suppliers regardless of which system they use.

- Explicitly: To provide compatibility/interoperability between JISC RDSS project systems.

However:
- It is not a single ‘Best Practice’
- It is not ‘one registry to rule them all’
Background: Jisc RDSS Project

Development of a multi-vendor shared services platform led to discussions of interoperability of format policies (i.e. “preservation actions”) between preservation systems.
A JISC funded project to initiate the process to deliver benefits to RDSS users

Arkivum, Preservica and Artefactual as RDSS product suppliers

Open Preservation Foundation as respected independent shared DP technology supplier
Preservation is not just about file formats, it’s about making sense of data.

The specific action depends on the context, and the policies.
– what action is being taken and why? What is the business rule?

Today - preservation actions are not portable across systems
(e.g. Archivematica, Preservica, others)
Current Registry (In)compatibility

Preservica Registry

Archivematica FPR
## What have we produced and why?

| Conceptual Model | ● Common framework for everyone  
|                  | ● Language between preservation systems  
|                  | ● Still under definition…  
| Json Schemas     | ● Formal definition of the conceptual model  
|                  | ● Machine readable, used in API payloads  
|                  | ● Used to test and validate interoperability  
| API              | ● Common interface for preservation systems  
|                  | ● Well defined way to exchange information  
| Executable Digital Preservation Actions | ● Cross-platform way to deploy/run tools  
|                  | ● Unambiguous and vendor independent  
| Proof of Concept | ● Reference implementation to share  
|                  | ● Make the idea really work between Preservica and Archivematica  

---

[Open Preservation Foundation](https://openpreservationfoundation.org)
PAR Conceptual Model
JSON schemas

- Tool
- Action
- Action Type
- Format
- Property
- Business Rule
APIs

Preservation Actions

Retrieve all preservation actions
Allow to retrieve the details of all the preservation action.

QUERY PARAMETERS

- limit
  Integer to limit the number of preservation action returned. Default value is zero, which will not filter the result.

- offset
  Integer to specify the offset of the first element of the list of preservation action returned. Default value is zero, which will not filter the result.

- modifiedAfter
  Filter the preservation action to return only the ones having a local modified date value GREATER than the one passed in. Allowed date time formats are YYYY-MM-DD/T11h:mm:ssZ and YYYY-MM-DD. In the second scenario, when the time is not provided, it will be assumed to be 00:00:00 UTC time.

- modifiedBefore
  Filter the preservation action to return only the ones having a local modified date value LOWER than the one passed in. Allowed date time formats are YYYY-MM-DD/T11h:mm:ssZ and YYYY-MM-DD. In the second scenario, when the time is not provided, it will be assumed to be 00:00:00 UTC time.

HEADER PARAMETERS

- tool
  "Filter the preservation actions by the Tool they use, provided its ID. Multiple IDs can be passed in within the same string, linked together using an ampersand symbol as follow: ID1; ID2; ID3. When more than one ID is provided, the endpoint will return all the preservation actions that use one of the IDs in the list. When used together with the preservation action type header parameters, the endpoint will return only those preservation actions satisfying both filters' criteria. A list of Tool IDs cannot be combined with another list of Preservation Action Type IDs."

https://github.com/JiscRDSS/rdss-par/tree/master/api
Executable Tool Definitions

- Machine readable spec for running a tool
  - Tool command line
  - Parameters and flags
  - Inputs and outputs
  - Pre and post processing

```json
[job mediaInfo2.cwl] completed success
{
  "width": "1280",
  "bitrate": "748253",
  "height": "720"
}
Final process status is success
```

Property extraction

```
[job md5check2.cwl] completed success
{
  "fixity_report": "PASS"
}
Final process status is success
```

Fixity check
Next steps

• OPF coordination
  o Define project deliverables and stages in more detail
• More use cases demonstrating real benefits
• Looking for more organisations to be involved

• Extend the conceptual model to more practical cases that involve more organisations

Make PAR useful to communicate good practice between systems and organisations
Join OPF today!

For more information get in touch...

✉️ martin.wrigley@openpreservation.org
🏡 http://openpreservation.org/
)}> https://github.com/openpreserve
🐦 @openpreserve

Newsletter: www.openpreservation.org/subscribe/

For more info on PAR go to

www.openpreservation.org/about/projects/par