Mapping out the OPF Reference Toolset

Martin Wrigley, Executive Director
OPF Webinar, 20 September 2018
This is the story behind the OPF reference toolset. We’ll look at how we can help the community with shared knowledge and useful digital preservation tools.

But how does OPF choose the tools that we will support and adopt - what is the criteria?

And how do we then take those tools forward and deliver what the community needs?
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?

<table>
<thead>
<tr>
<th>Austrian Institute of Technology</th>
<th>Österreichische Nationalbibliothek</th>
</tr>
</thead>
<tbody>
<tr>
<td>British Library</td>
<td>Preservica</td>
</tr>
<tr>
<td>Bibliotheque Nationale de France</td>
<td>Yale University Library</td>
</tr>
<tr>
<td>Goportis</td>
<td>Albert-Ludwigs Universitat</td>
</tr>
<tr>
<td>International Atomic Energy Archives</td>
<td>University of North Carolina</td>
</tr>
<tr>
<td>Jisc</td>
<td>Portico</td>
</tr>
<tr>
<td>Koninklijke Bibliotheek</td>
<td>PSNC (Poznan Supercomputing &amp; Networking Centre)</td>
</tr>
<tr>
<td>Det Kgl. Bibliotek</td>
<td>Artefectual</td>
</tr>
<tr>
<td>Nationaal Archief</td>
<td>Biblioteca Nacional de Portugal</td>
</tr>
<tr>
<td>The National Archives UK</td>
<td>Arcsys Software</td>
</tr>
<tr>
<td>Nasjonalbiblioteket</td>
<td></td>
</tr>
<tr>
<td>Rigsarkivet</td>
<td></td>
</tr>
<tr>
<td>Ex Libris</td>
<td></td>
</tr>
<tr>
<td>Rahvusarhiiv</td>
<td></td>
</tr>
</tbody>
</table>
What does OPF do?

- Develop the OPF reference toolset
- Deliver to development roadmaps
- Community engagement and knowledge sharing
- Best practice development
- 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

- **Identify**
  - Thing
  - Meta Thing
  - Container explosion recursive

- **Validate**
  - Validation polices
  - Fix/transform (migrate...)
  - Fix/transform* (redact...)

- **Characterise**
  - Characterisation polices
  - Characterisation polices

- **Packaging**
  - Packaging polices
  - Fix/transform*

- **Put into a Box**
  - (turn into an AIP)

- **Periodic re-check**

*Quality check derivative

**Note:**
- Thing is (or is becoming) an SIP
OPF Tool Mapping

Derivative check tools
Maintained through OPF
xorsound WAV, MP3

Identification tools
Maintained through OPF
fido Format Sniff
Recommended by OPF
DROID
PRONOM
FILE

Transform
Database archiving / Extraction tools
Recommended by OPF
SIARD (SQL database to XML format)

Validation and Characterisation tools
Maintained through OPF
PDF/A
PDF, JPEG, WAV, PNG, WARC, AIFF, UTF8 TEXT, XML, HTML, GZIP, ASCII TEXT, MP3, GIF, JPEG2000

Validation polices
Fix/transform (migrate...)

Packaging polices
Fix/transform*

Quality & Cross Check polices

Put into a Box (turn into an AIP)

Periodic re-check

Information Packaging tools
TBA

Cross Check tools
TBA

Quality check tools
E-ARK CEF SIP validator

Characterisation polices

Characterise
T M T+

Validate
T

Package, Quality Assurance, Review, Cross Check
T M T+

Validated polices
Fix/transform* (redact...)

Fix/transform (migrate...)

Disk image explosion/analysis
Recommended by OPF
BitCurator

Container explosion recursive

T M T+

Meta Thing

Identify

Thing

*Quality check derivative

( haunted by BitCurator )

Transform
Database archiving / Extraction tools
Recommended by OPF
SIARD (SQL database to XML format)
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
Improvements to documentation

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)
Any questions?
Get in touch!

Send us your feedback...

✉️ martin@openpreservation.org

🏠 http://openpreservation.org/

🐱 https://github.com/openpreserve

🐦 @openpreserve

✉️ http://openpreservation.org/subscribe/