



OPF Advisory Group (OAG)

Jisc (London) and online

29 June 2022



Welcome!

Welcome Willkommen Bienvenido/a Bienvenue Välkommen 欢迎 Witamy أهلا بك
selamat datang bem-vinda ようこそ Tere tulemast Velkommen Welkom Croeso Chào
mừng bạn Benvenuta/o Vítejte Tervetuloa Сардэчна запрашаем ברוך הבא Laipni lūdzam
Fàilte vitajte Wèllkomm Dobrodošli მოგესალმებით Üdvözöljük i mirëpritur

Agenda and aims

- Welcome and introductions
- Agenda and aims
- Year in review: tools
- Year in review: member groups
- Year in review: projects
- Next steps
- AOB
- Wrap-up



A Year in Review

Tools

Overview

- 35 minutes: JHOVE review and planning
- 15 minutes: veraPDF review and 1.22 preview
- 5 minutes: FIDO review and planning
- <5 minutes: Jpylyzer

JHOVE v1.26 Community Release

- Released on Thursday 9th June
- Contributions gathered during previous OPF Hack events
- Other community contributions submitted since v1.24 release
- The release comprises:
 - More than 70 pull requests
 - 14 direct contributors
 - Around 50 closed issues following triage on 14th June
- Release notes:
<https://github.com/openpreserve/jhove/releases/tag/v1.26.0>

JHOVE v1.26 Highlights

- GUI Improvements:
 - Addition of keyboard shortcuts
 - Allow folder analysis
- PDF Module improvements:
 - Fixed some issues with parenthesis handling in PDF document dictionary
 - Retired faulty PDF/A validation, veraPDF should be used instead
 - Fixed several loop conditions causing PDF processing to hang
 - Performance improvements
- XML Module:
 - Fixed XML signature handling
 - Improvements to handling of external schemas
- EPUB Module now supports latest EPUB Check v4.2.6
- Language settings now configurable
- Fix JHOVE build compatibility for JDK 11 onwards

JHOVE v1.28 Planning

- UTF-8 module overhaul
- XML module fixes vs. JAX based OPF module
- Member issues
- Filter message reporting

JHOVE Estimation & Resource Allocation

- Fine-grained accurate estimation of issues/tasks is difficult
- Small errors accumulate
- An alternative approach:
 - Divide time between prioritised objectives, e.g. fix PDF loops
 - Prioritise issues within an objective
 - Work on an objective until out of time
 - Move on to the next objective

veraPDF v1.18 (Feb 2021)

- Support for PDF/UA-1 Matterhorn Protocol machine validation
- Minor Fixes:
 - STDIN support
 - Parsing of inline images
 - Edge case glyph parsing for TrueType and CFF fonts
 - Better parsing of private data in Type1 fonts
 - Infinite loop when parsing circular CMap dependencies
 - Handle zero page documents

veraPDF 1.20 (Jan 2022)

- Support for PDF/A-4 including:
 - Level F (File attachments)
 - Level E (Engineering)
- Java 11-16 support
- Fixed validation of the permissions dictionary in PDF/A-2 and PDF/A-3
- Improvements to PostScript and CFF font parsing
- More robust handling of malicious PDF documents
- Various XMP parsing/validation improvements:
 - Allow empty LangAlt arrays in XMP metadata
 - Better handling of missing standard identification XMP fields
 - Fixed validation of value types that are redefined in extension schema
 - Validate XMP URLs as text

veraPDF v1.22 (Summer 2022)

- Implemented PKCS7 encryption handler to replace deprecated JavaX version
- Support for evaluation of PDF functions
- Better handling of user config location for custom installs
- Fixed text output bug for custom policy reporting
- Fixed bug with handling spaces in output folder names on Windows
- Drag and Drop fixed on Mac OS

PDF/UA Validation and Viewer Demo

Boris to give a demonstration of the online PDF/UA viewer:

<https://verapdf.duallab.com/>

FIDO v1.6

- The previous production release was version 1.4
- Release candidate out in March, can be downloaded from the Python Package Index (PyPi) using pip, the Python package installer
- Draft release notes are available:
<https://github.com/openpreserve/fido/releases/tag/v1.6.0rc5>
- FIDO v1.6.0rc2 will be released on Tuesday 5th July

FIDO v1.6 Highlights

- PRONOM v104 supported for latest signature files
- Signature web service
- Fixed issue causing crash for XLS files
- Unicode support for Windows Python 2.7
- Generated signature file validated against the XML schema
- Addressed errors with signature generation
- Addressed technical debt, particularly code style warnings

Jpylyzer

Jpylyzer development schedules driven by Johan, the maintainer.

- OPF will liaise with Johan on the next release
- If Johan is present he can give a quick update



Break

10 mins



A Year in Review

Member Groups

Archives Interest Group

- Published [The Significant Properties of Spreadsheets \(OPF AIG Final Report\) | Zenodo](#) (1000+ views, 500+ downloads) and presented it at iPRES 2021
- Received feedback, updated version to be reviewed in OPF publication review process
- Spawned ICRFF Working Group, Errors and Warnings Working Group in preparation
- **Discontinued AIG on June 13, 2022**
- Looking back:
 - Not just the work, also ‘an inspiration group’
 - Fond memories of Copenhagen trip (Feb. 2020)
 - Expanded smaller working bubbles (pandemic)
 - Enjoyed learning and sharing what the member organisations were doing



Diversity and Inclusion Working Group

- Established with DDHN after member interest in D&I at the 2020 AGM
- Areas of interest:
 - Inclusive language and terminology in digital preservation
 - Bias in AI
 - Community-driven preservation (selection, acquisition, metadata)
 - Inclusive metadata
 - Diversifying the preservation community
 - The Designated Community

Reflections

- Regular meetings may not be required.
- OPF members groups should be run by an engaged chair from the community.

Moving forward, OPF will...

- maintain the group slack channel (email Charlotte to be added).
- facilitate D&I meetings, webinars and discussions at request of members.
- produce D&I resources for our website (e.g. translation guides and best practices).

D&I at OPF

In addition to the working group, OPF has been more conscious about improving D&I since the 2020 AGM:

- Reviewing contributor diversity at OPF events
- Paid webinar slots for independent speakers, small archives, charitable organisations, or minority-focused organisations.
- Inclusive hiring practises, leading to our first female ED and greater gender balance across our team.
- Taking steps to diversify our board of directors.
- Accessibility review of our content and platforms, working to address gaps (e.g. webinar captions).

ICRFF Working Group

International Comparison of Recommended File Formats Working Group

- Established 2022
- Discusses and collaborates on issues related to file formats
- Created and maintains the table “International Comparison of Recommended File Formats”
- Monthly meetings for entire group
- Ad-hoc meetings on particular “working items”



A Year in Review

Projects

E-ARK

E-ARK involvement took a lot of OPF time in 2020-21

- Probably over committed in retrospect
- Production of E-ARK specifications for information packages
 - CSIP: <https://earkcsip.dilcis.eu/>
 - SIP: <https://earksip.dilcis.eu/>
 - DIP: <https://earkdip.dilcis.eu/>
 - AIP: <https://earkaip.dilcis.eu/>
- Development of an information package validator:
<https://eark.openpreservation.org/>

E-ARK Planning

The next phase of E-ARK will be a procurement project:

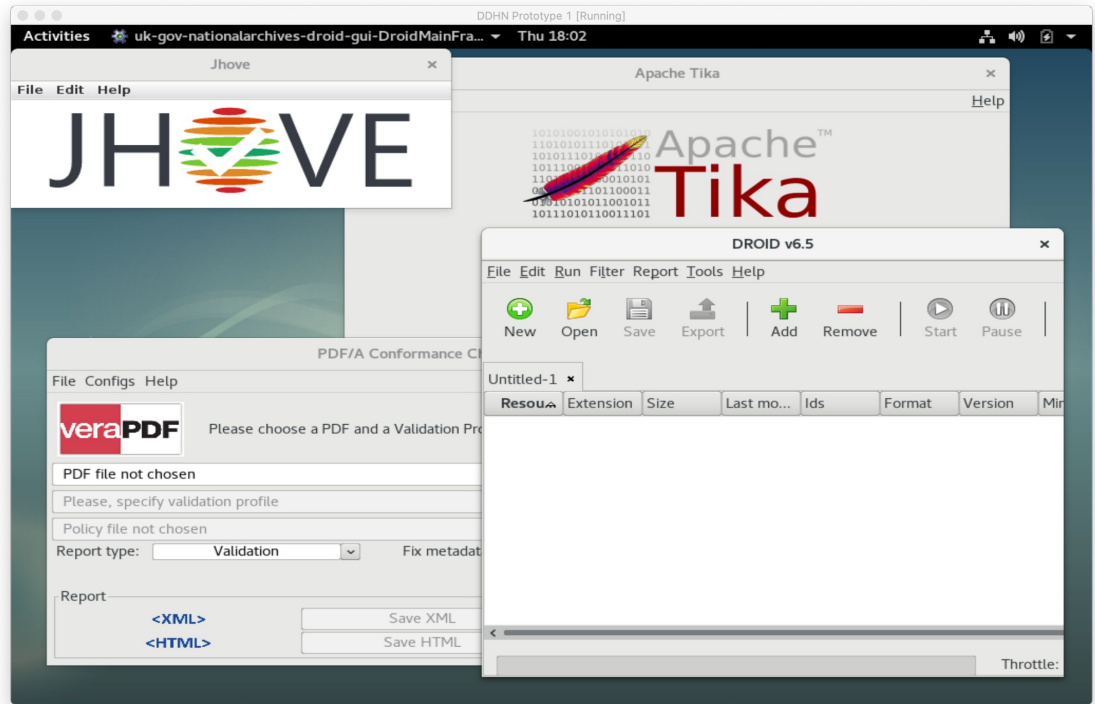
- Aimed at getting organisations up and running with E-ARK
- Production of specifications for specific content types, e.g. financial records, health records, etc.
- Development of validation software for the specifications
- OPF involved in bid, but with significantly reduced technical involvement
- If the bid is approved OPF will resume in 2023

Preservation Tools VRE

- Funded by and developed with the Dutch Digital Heritage Network (DDHN)
- Virtual machine environment with popular digital preservation tools pre-installed
- New Release Candidate prepared and will be released early July
- Mostly bug fixes and tool updates
- Another release with functionality/tool updates due for start 2023

Preservation Tools VRE

Providing an easy way for the community to test preservation tools without the necessity of installing them. We prioritized tools with a GUI in order to make VRE accessible to organizations who are starting to implement digital preservation tools.



PAR: OPF Role

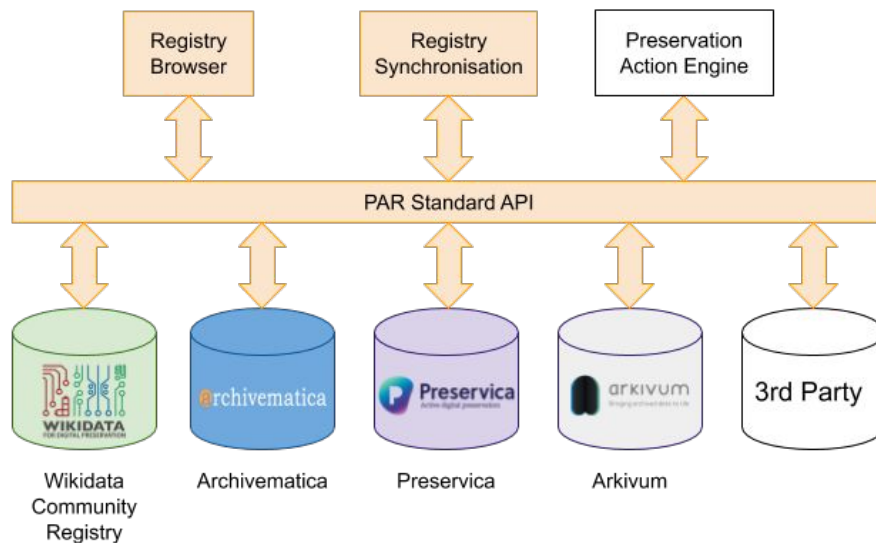
Under the stewardship of Jisc, PAR has moved forward from a standing start to its current position that includes a feature-rich API, an endpoint browser with a prototype API implementation spanning Preservica, Archivematica and Arkivum

With release 1.0 now on a firm footing the ongoing stewardship of PAR is to be transferred to the OPF

This change is aimed at:

- providing a more independent technical oversight and overall task coordination of PAR activities, and
- To expand the community knowledge and involvement with PAR via:
 - regular webinars and workshops
 - encouraging involvement via hack events, and
 - in some cases, as the providers of best practice trusted sources

PAR Overview



By way of introduction the ultimate purpose of PAR is to:

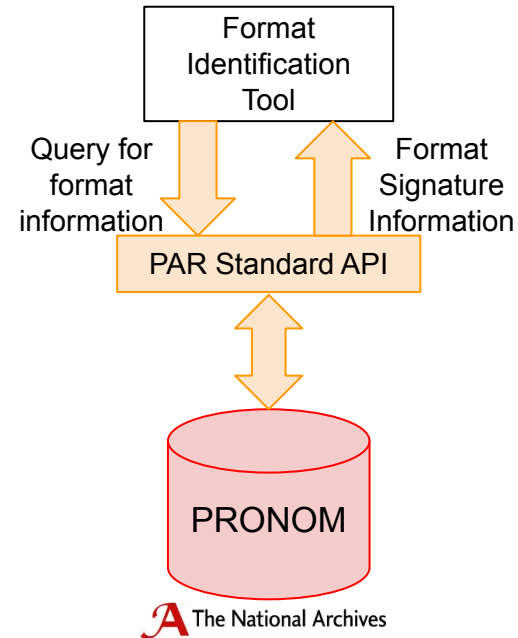
- facilitate sharing preservation actions between disparate DP systems
- allow well-formed preservation workflows, such as a set of compounded preservation actions, to be imported from an external trusted source for use within your local environment



PAR Use Cases: Updating format signatures

DROID and FIDO are file format identification tools that rely on format signature data that is downloaded from the UK National Archives PRONOM service. While this data is of high quality there can be long gaps between signature file releases meaning it can take a significant time for new format signatures to make it into DROID or FIDO. A PAR based solution would involve:

- Implementing a PAR endpoint for PRONOM allowing PAR clients to query and obtain format data.
- Developing a new signature update tool for DROID and FIDO to use the PAR API to query for and download new format signatures.
- Users could choose alternative PAR registries as sources of format information.





Next Steps

Member Champions and Chairs

Tool//project/group	Champion/Chair	Staff Contact
JHOVE	Micky Lindlar, Thomas Ledoux, David Russo	Carl
Fido	Sarah Romkey	Darren
veraPDF	Boris Dubrov, David Russo	Carl
Jpylyzer	Johan van der Knijff	Darren
PAR	Paul Stokes, Sarah Romkey	Carl
ICRFF	Asbjørn Skødt	Georgia
E-ARK		Carl
Carrus		Darren
VRE	Lotte Wijsman	Charlotte



AOB



Close

Thank you!