

Introducing veraPDF

A quick overview of the Open Source PDF/A Validator
Tallinn, October 2017



Industry-supported
PDF/A validation

The PREFORMA project

veraPDF development has been funded by the PREFORMA project

PREservation FORMAts for culture information/e-archives, is a Pre-Commercial Procurement (PCP) project co-funded by the European Commission under its FP7-ICT Programme. The project's main aim is to address the challenge of implementing standardised file formats for preserving digital objects in the long term, giving memory institutions full control over the acceptance and management of preservation files into digital repositories. <http://www.preforma-project.eu/>

The veraPDF consortium

Led by the

- Open Preservation Foundation
- PDF Association

With partners

- Dual Lab
- KEEP Solutions
- DPC



PDF/A validation

Latest release: 1.8 (Aug 17)

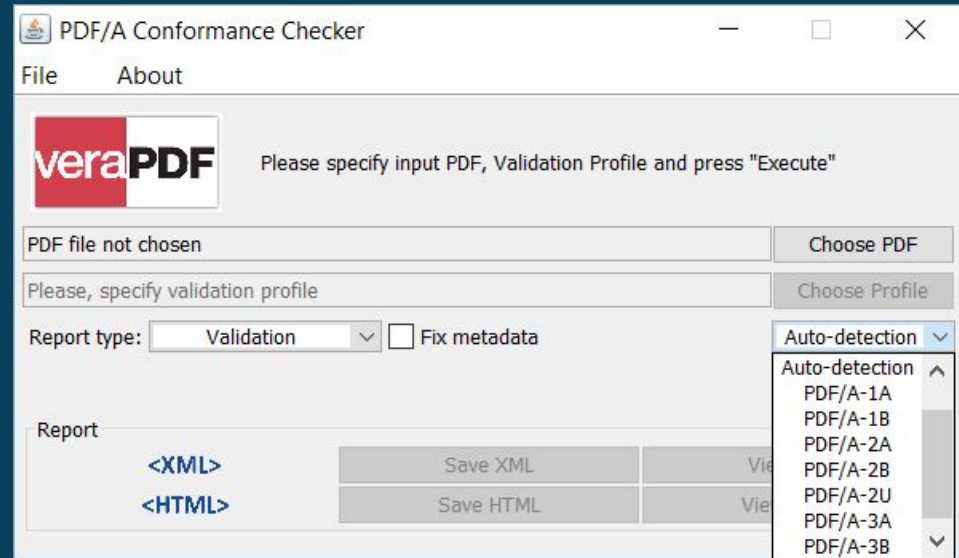
Conformance checker: full support for all PDF/A versions (1,2,3) and levels (A,B,U)

GUI - desktop version for single file evaluation

CLI - command line interface targeting large volume batch processing

Web - Online demo web site

Java library - Calling a Java API from custom Java-based applications



Open source for sustainability

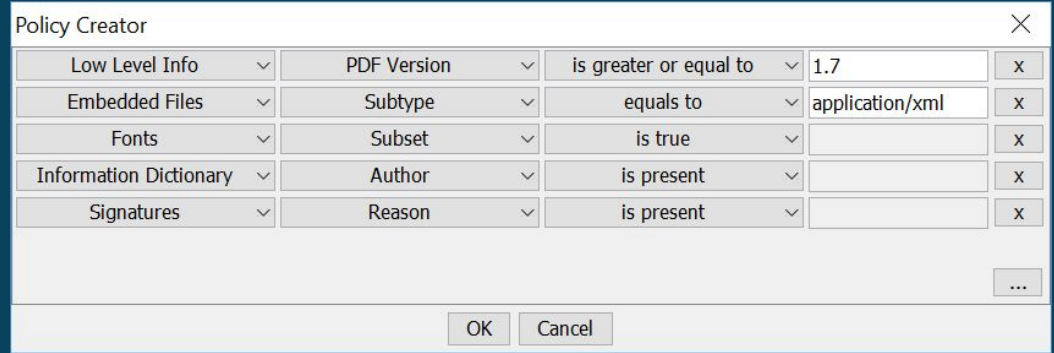
Software licenced under dual MPL v2+ / GPL v3+ allowing anyone to download or integrate the software free of charge

Test datasets and documentation licensed under CC-BY-4

Active open source community

Recent news and developments

- GUI for defining institutional **policy profiles**
- Support for recently published **PDF 2.0** in PDF feature extraction
- Optimizations for more **stable processing** of (very) large file collections and large individual documents
- Integration of veraPDF into a **web document crawler** and analyser (in partnership with Logius, the Hague, Netherlands)



The future

PREFORMA funding ended in July 2017. veraPDF is transitioning from a funded project to a stand alone open source project.

The OPF and Dual Lab will:

- continue to **address bugs** reported on GitHub;
- test and merge **small new features** submitted as pull requests submitted to GitHub; and
- **provide answers and support** for minor mailing list enquiries.

Future developments?

veraPDF is functionally complete for PDF/A but there is still work to be done, for example:

- **developing validators for other parts of the PDF standard**
- veraPDF supports a plug in mechanism that can be used to validate other formats embedded or attached to PDF/A documents
- integrating veraPDF with commercial preservation systems and other services
- the creation of more complex policy schematron that reflect collections of real institutional acceptance criteria

Coverage of existing PDF standards

✓ **PDF/A**
ISO 19005

PDF/UA
ISO 14289

PDF/X
ISO 15930

PDF/VT
ISO 16612-2

PDF
ISO 32000

PDF/A-4

PDF/E
ISO 24517

Partnership

Memory institutions alone cannot have expertise in-house for every format they collect and preserve.

The PDF family of formats is widely used in digital preservation. PDF/A is just one specification, there are no complete open source validators for the others.

OPF and PDF Association have bridged a gap between memory institutions and industry to create a successful product.

Find out more



<http://verapdf.org/>



users@lists.verapdf.org (Q&A, discussions)



<https://github.com/veraPDF>



<http://verapdf.org/subscribe/> (news)



[@_verapdf](https://twitter.com/_verapdf)



info@verapdf.org

This work was co-funded by the PREFORMA project and European Commission under the FP7-ICT Programme.

