Koninklijke Bibliotheek
National Library of the Netherlands

Why PDF/A validation matters

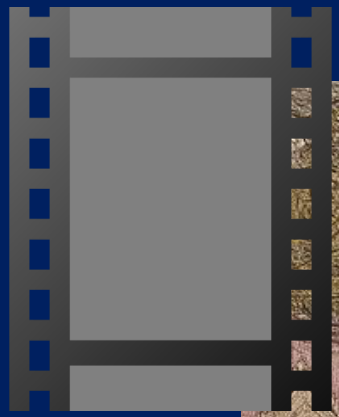
(even if you don't have PDF/A)

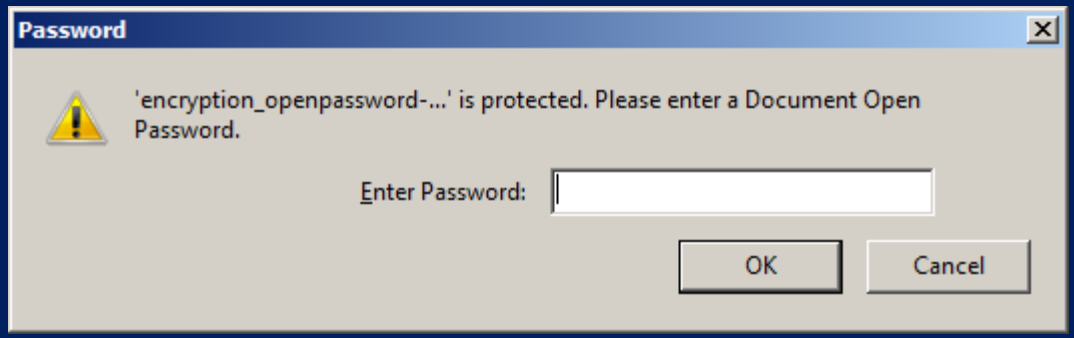
Johan van der Knijff

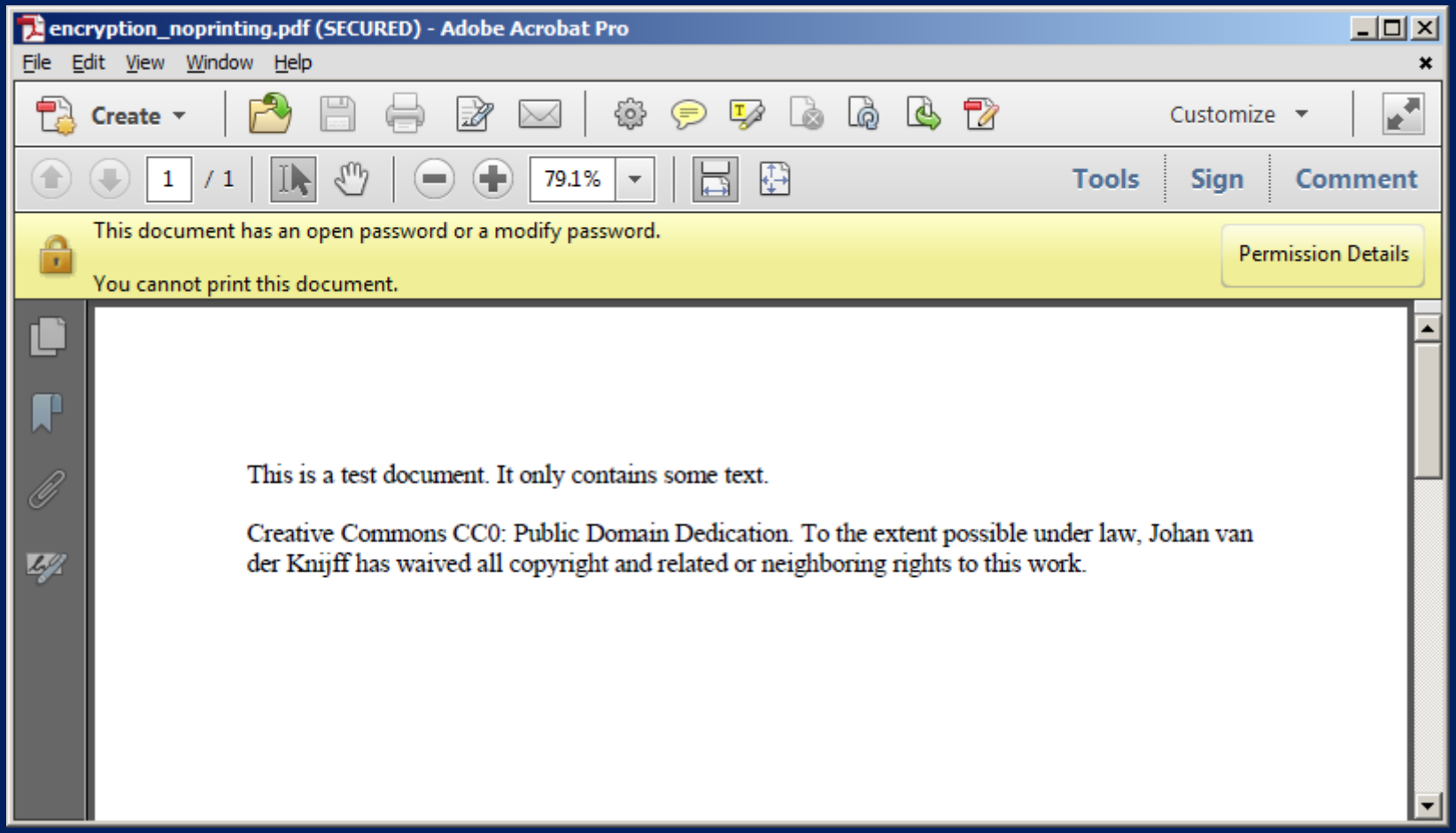
PANDORA opening her Box.

PANDORA opening her ~~Box~~ PDF

abcd



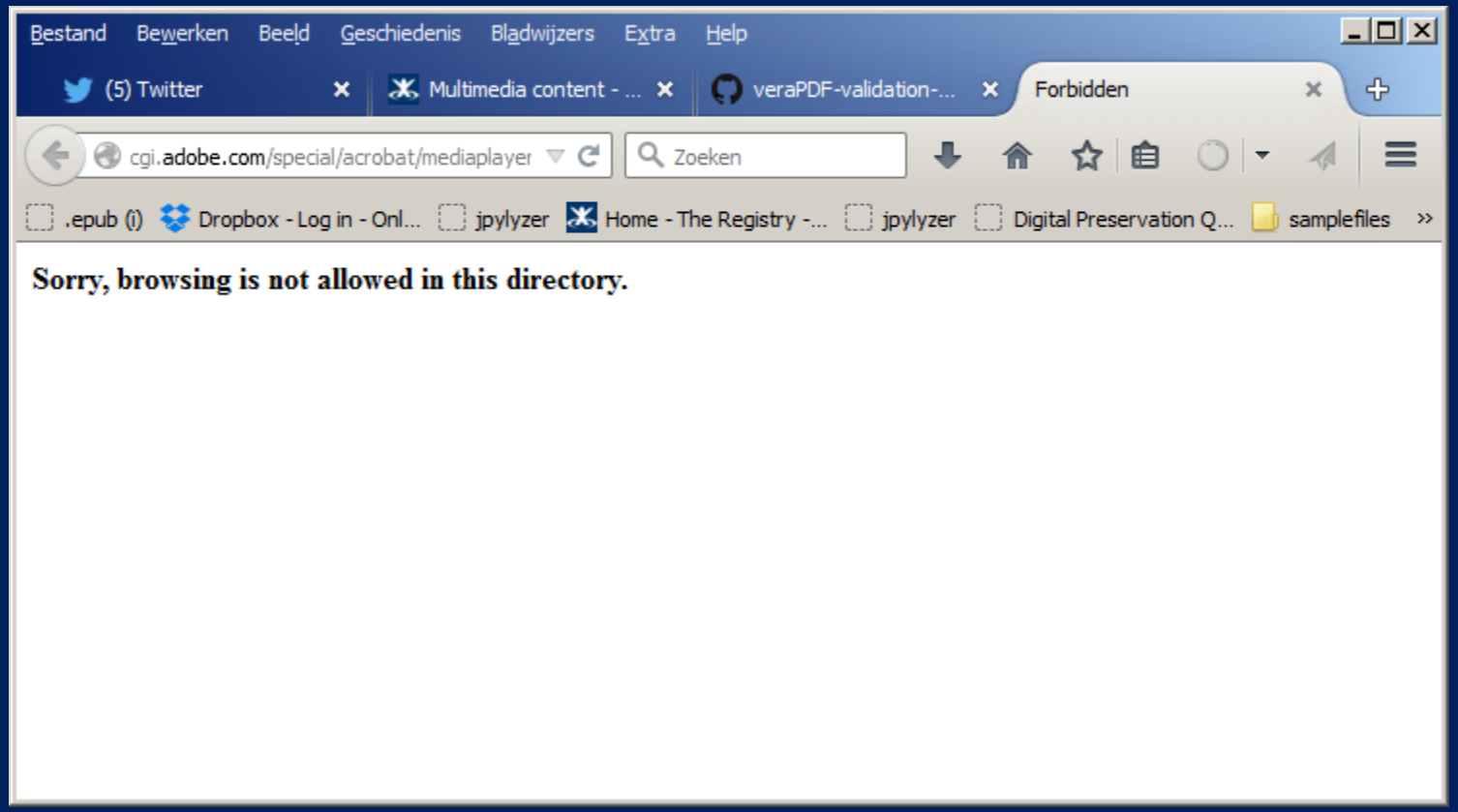

Password
'encryption_openpassword-...' is protected. Please enter a Document Open Password.
Enter Password:
OK
Cancel


encryption_noprinting.pdf (SECURED) - Adobe Acrobat Pro
File Edit View Window Help
Create
Customize
1 / 1
79.1%
Tools Sign Comment
This document has an open password or a modify password.
You cannot print this document.
Permission Details
This is a test document. It only contains some text.
Creative Commons CC0: Public Domain Dedication. To the extent possible under law, Johan van der Knijff has waived all copyright and related or neighboring rights to this work.

embedded_video_quicktime.pdf - Adobe Acrobat Pro
File Edit View Window Help
Create
Customize
1 / 1
111%
Tools Sign Comment

der Knijff has waived all copyright and related or neighboring rights to this work.

Embedded Quicktime video by Misty De Meo, taken from the OPF format corpus. The licensing info there states “created using ffmpeg's builtin "testsrc" test pattern, which is presumed to be unlicensed. ffmpeg's code is GPLv2 licensed”.


Bestand Bewerken Beeld Geschiedenis Bladwijzers Extra Help
(5) Twitter
Multimedia content - ...
veraPDF-validation-...
Forbidden
cgi.adobe.com/special/acrobat/mediaplayer
Zoeken
.epub (i)
Dropbox - Log in - Onl...
jpylyzer
Home - The Registry -...
jpylyzer
Digital Preservation Q...
samplefiles
Sorry, browsing is not allowed in this directory.

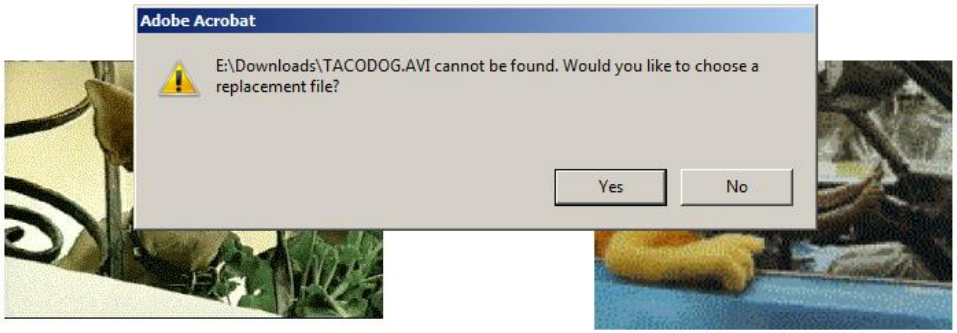
movie.pdf - Adobe Acrobat Pro

File Edit View Window Help

Create Customize

1 / 1 111% Tools Sign Comment

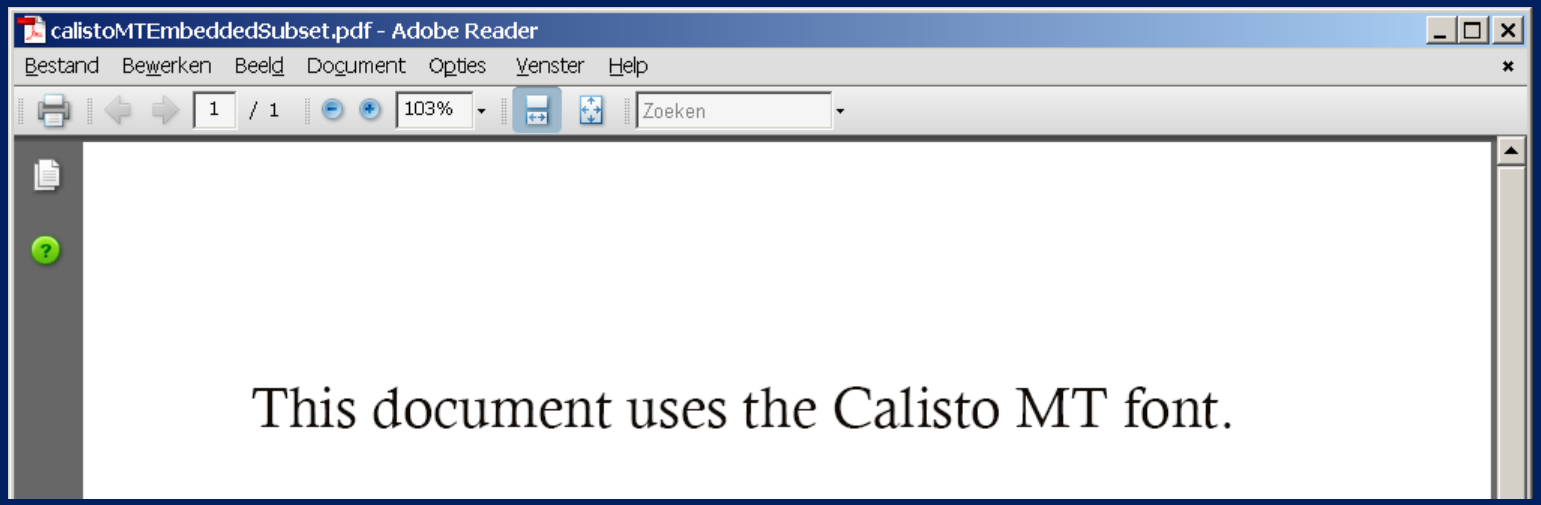
Going to the movies!



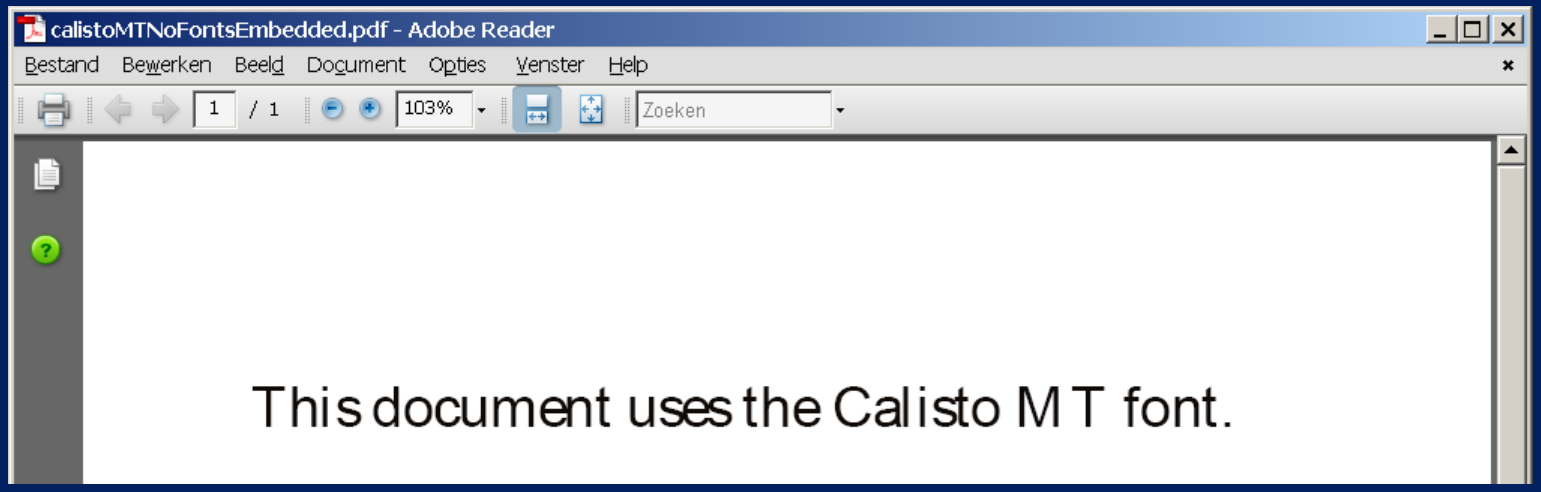
AVI file with the Controller

AVI file w/out the controller

On creator machine



On Linux machine

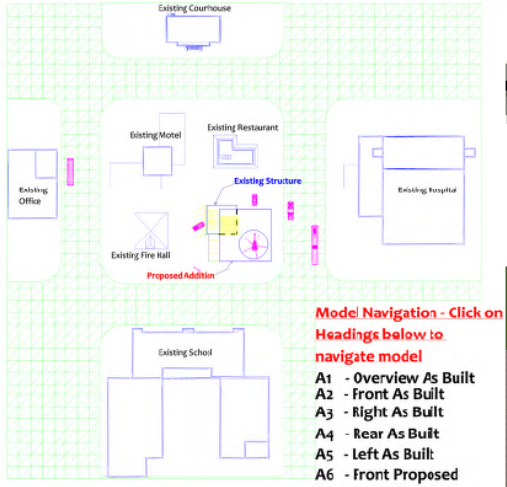


pdf 1.7 portfolio 3D multiple sheets.pdf - Adobe Acrobat Pro

File Edit View Window Help

Create Customize

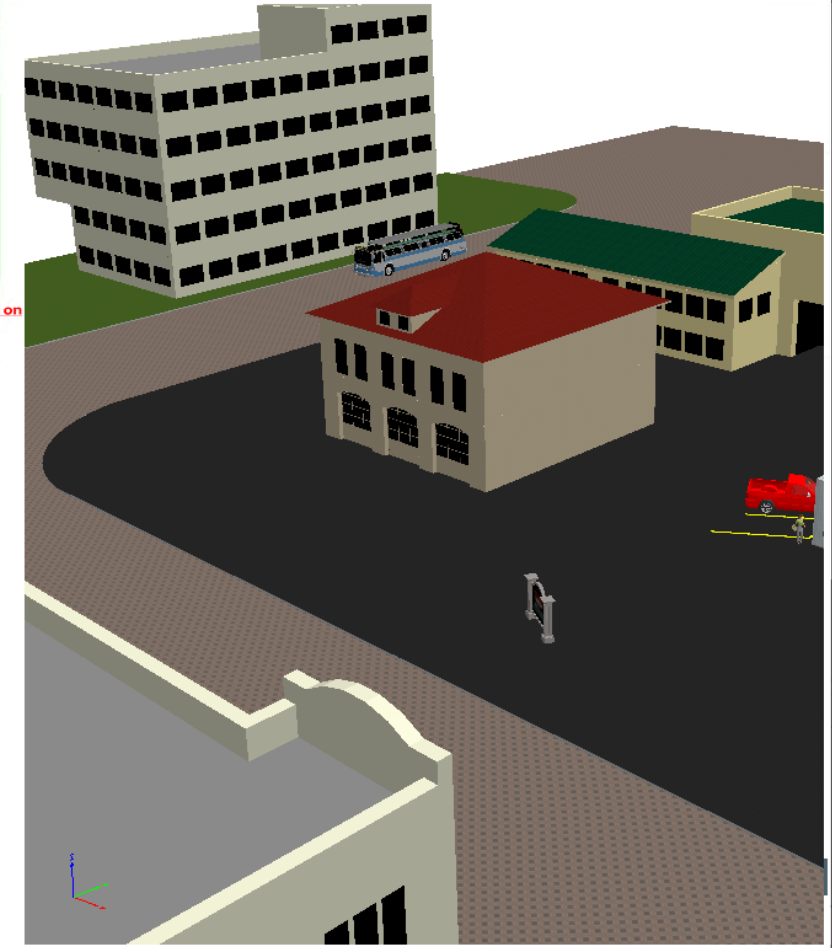
1 / 1 32.1% Tools Sign Comment



Model Navigation - Click on Headings below to navigate model

A1 - Overview As Built
A2 - Front As Built
A3 - Right As Built
A4 - Rear As Built
A5 - Left As Built
A6 - Front Proposed
A7 - Right Proposed
A8 - Rear Proposed
A9 - Left Proposed
A10 - Section 1
A11 - Section 2
A12 - Section 3
A13 - Section 4
A14 - Credits
A15 - Contact Us
A16 - Visit our Website
A17 - Back to the Gallery

LEGAL NOTICE:
THESE DRAWINGS ARE MADE AS A VISUALIZATION TOOL AND ARE INTENDED TO AUGMENT, NOT REPLACE REQUIRED ARCHITECTURAL OR ENGINEERING PLANS. VIRTUAL ARCHITEC LLC, IT'S OFFICERS, OWNERS & EMPLOYEES ARE NOT ARCHITECTS OR ENGINEERS. ANY ENGINEERING REQUIRED FOR THIS DESIGN WILL BE PROVIDED BY OTHERS. CONSTRUCTION FROM THESE PLANS SHOULD NOT BE UNDERTAKEN WITHOUT THE ASSISTANCE OF A CONSTRUCTION PROFESSIONAL. THE PURCHASER AND BUILDER OF THIS PLAN RELEASES VIRTUAL ARCHITEC LLC, IT'S OFFICERS, OWNERS & EMPLOYEES FROM ANY CLAIMS OR LAW SUITS THAT MAY ARISE DURING THE CONSTRUCTION OF THIS STRUCTURE OR ANY TIME THEREAFTER.
ADDITIONALLY
VIRTUAL ARCHITEC LLC, IT'S OFFICERS, OWNERS & EMPLOYEES ARE UNABLE TO ACCEPT ANY LIABILITY FOR THE ACCURACY OR OVERALL INTEGRITY OF THESE DRAWINGS SINCE ANY COMPUTER GENERATED PLANS CAN BE MODIFIED OR ALTERED WITHOUT OUR KNOWLEDGE. THEREFORE THE BUILDER MUST CAREFULLY INSPECT ALL DIMENSIONS AND DETAILS IN THESE DRAWINGS AND ASSUME RESPONSIBILITY FOR THE SAME.



Legal Description
Put legal Desc. Here
Anytown

Designer:
Virtual Architec LLC

Contractor Name
Contractor/Builder:

Cover Page / Table of Contents
Sheet Label:

Scale: NTS

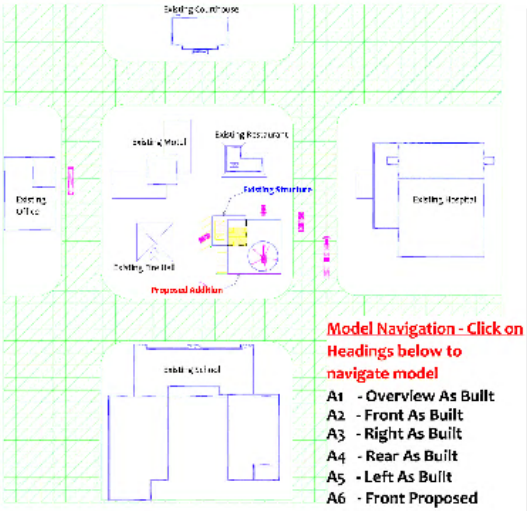
Client:
Sample Commercial Addition/ Remodel Proposal for City Hall

Date: Date

Sheet: A1

pdf 1.7 portfolio 3D multiple sheets.pdf - GSview

File Edit Options View Orientation Media Help



Model Navigation - Click on Headings below to navigate model

- A1 - Overview As Built
- A2 - Front As Built
- A3 - Right As Built
- A4 - Rear As Built
- A5 - Left As Built
- A6 - Front Proposed
- A7 - Right Proposed
- A8 - Rear Proposed
- A9 - Left Proposed
- A10 - Section 1
- A11 - Section 2
- A12 - Section 3
- A13 - Section 4
- A14 - Credits
- A15 - Contact Us
- A16 - Visit our Website
- A17 - Back to the Gallery

LEGAL NOTICE
THESE DRAWINGS ARE MADE AS A VISUALIZATION TOOL AND ARE INTENDED TO AUGMENT, NOT REPLACE REQUIRED ARCHITECTURAL OR ENGINEERING PLANS. VIRTUAL ARCHITECT LLC, ITS OFFICERS, OWNERS & EMPLOYEES ARE NOT ARCHITECTS OR ENGINEERS. ANY ENGINEERING REQUIRED FOR THIS DESIGN WILL BE PROVIDED BY OTHERS. CONSTRUCTION FROM THESE PLANS SHOULD NOT BE UNDERTAKEN WITHOUT THE ASSISTANCE OF A CONSTRUCTION PROFESSIONAL. THE PURCHASER AND BUILDER OF THIS PLAN RELEASES VIRTUAL ARCHITECT LLC, ITS OFFICERS, OWNERS & EMPLOYEES FROM ANY CLAIMS OR LAWSUITS THAT MAY ARISE DURING THE CONSTRUCTION OF THIS STRUCTURE OR ANY TIME THEREAFTER.
ADDITIONALLY
VIRTUAL ARCHITECT LLC, ITS OFFICERS, OWNERS & EMPLOYEES ARE UNABLE TO ACCEPT ANY LIABILITY FOR THE ACCURACY OR OVERALL INTEGRITY OF THESE DRAWINGS SINCE ANY COMPUTER GENERATED PLANS CAN BE MODIFIED OR ALTERED WITHOUT OUR KNOWLEDGE. THEREFORE THE BUILDER MUST CAREFULLY INSPECT ALL DIMENSIONS AND DETAILS IN THESE DRAWINGS AND

Virtual Architect LLC

Contractor Name

Cover Page / Table of Contents

Scale: NTS

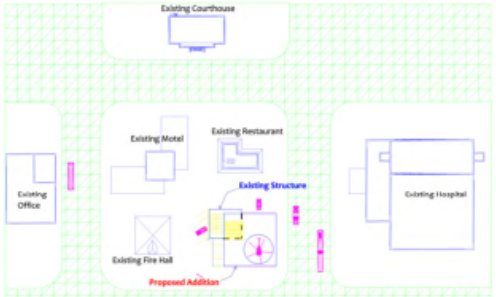
Sample Commercial Addition/ Remodel Proposal for City Hall

...ortfolio 3D multiple sheets.pdf 236, 1562pt Page: "1" 1 of 1

pdf 1.7 portfolio 3D multiple sheets.pdf - [Sample Commercial Addtion Proposal] - SumatraPDF

File View Go To Zoom Favorites Settings Help

Page: 1 / 1 Find:



Model Navigation - Click on Headings below to navigate model

A1 - Overview As Built
A2 - Front As Built
A3 - Right As Built
A4 - Rear As Built
A5 - Left As Built
A6 - Front Proposed
A7 - Right Proposed
A8 - Rear Proposed
A9 - Left Proposed
A10 - Section 1
A11 - Section 2
A12 - Section 3
A13 - Section 4
A14 - Credits
A15 - Contact Us
A16 - Visit our Website
A17 - Back to the Gallery

LEGAL NOTICE:
THESE DRAWINGS ARE MADE AS A VISUALIZATION TOOL AND ARE INTENDED TO AUGMENT, NOT REPLACE REQUIRED ARCHITECTURAL OR ENGINEERING PLANS. VIRTUAL ARCHITEC LLC, IT'S OFFICERS, OWNERS & EMPLOYEES ARE NOT ARCHITECTS OR ENGINEERS. ANY ENGINEERING REQUIRED FOR THIS DESIGN WILL BE PROVIDED BY OTHERS. CONSTRUCTION FROM THESE PLANS SHOULD NOT BE UNDERTAKEN WITHOUT THE ASSISTANCE OF A CONSTRUCTION PROFESSIONAL. THE PURCHASER AND BUILDER OF THIS PLAN RELEASES VIRTUAL ARCHITEC LLC, IT'S OFFICERS, OWNERS & EMPLOYEES FROM ANY CLAIMS OR LAW SUITS THAT MAY ARISE DURING THE CONSTRUCTION OF THIS STRUCTURE OR ANY TIME THEREAFTER.
ADDITIONALLY
VIRTUAL ARCHITEC LLC, IT'S OFFICERS, OWNERS & EMPLOYEES ARE UNABLE TO ACCEPT ANY LIABILITY FOR THE ACCURACY OR OVERALL INTEGRITY OF THESE DRAWINGS SINCE ANY COMPUTER GENERATED PLANS CAN BE MODIFIED OR ALTERED WITHOUT OUR KNOWLEDGE. THEREFORE THE BUILDER MUST CAREFULLY INSPECT ALL DIMENSIONS AND DETAILS IN THESE DRAWINGS AND ASSUME RESPONSIBILITY FOR THE SAME.

Legal Description
Put legal Desc. Here
Anytown

Designer:

Contractor/Builder:
Contractor Name

Sheet Label:
Cover Page / Table of Contents

Scale:
NTS

Client:
Sample Commercial Addition/ Remodel Proposal for City Hall

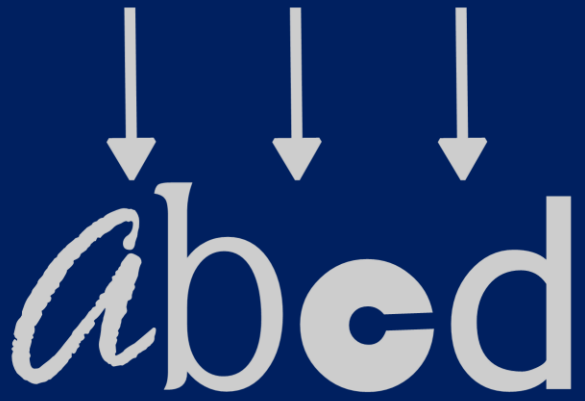
Date:
Date

Sheet:
A1

WARNING

Command Prompt

```
e:\Downloads>c:\qpdf\bin\qpdf --check encryption_nocopy.pdf
checking encryption_nocopy.pdf
PDF Version: 1.7
R = 4
P = -1044
User password =
extract for accessibility: allowed
extract for any purpose: not allowed
print low resolution: allowed
print high resolution: allowed
modify document assembly: not allowed
modify forms: allowed
modify annotations: allowed
modify other: allowed
modify anything: not allowed
stream encryption method: AESv2
string encryption method: AESv2
file encryption method: AESv2
File is linearized
ERROR: linearized file contains an uncompressed object after a compressed one in
 a cross-reference stream
WARNING: page 0 has shared identifier entries
WARNING: page 0: shared object 25: in hint table but not computed list
```

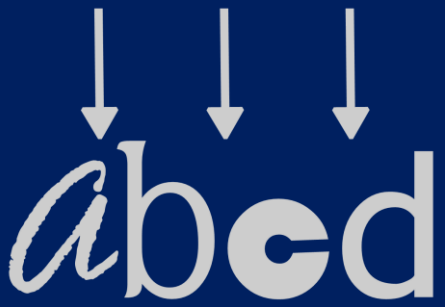


Command Prompt

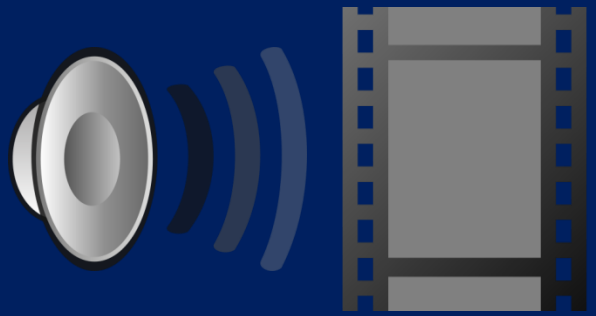
```
e:\preflightTests>pdffonts notembedded_pm65.pdf
name                                 type              emb sub uni object ID
------------------------------------ ----------------- --- --- --- ------ --
AdobeCorpID-MyriadRg                 Type 1            no  no  no      13  0
ComicSansMS                          TrueType          no  no  no       8  0
Tahoma                               TrueType          no  no  no      11  0
Haettenschweiler                     TrueType          no  no  no      16  0
TimesNewRoman                        TrueType          no  no  no      18  0
Helvetica                            Type 1            no  no  no      35  0

e:\preflightTests>_
```

Tool A

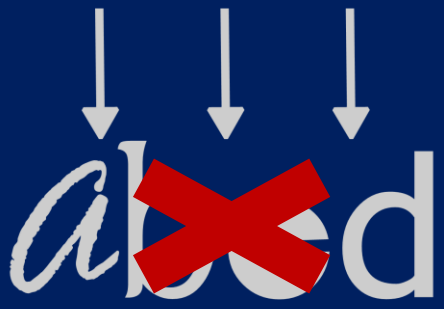
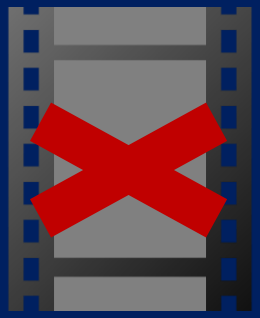


Tool B



Tool C

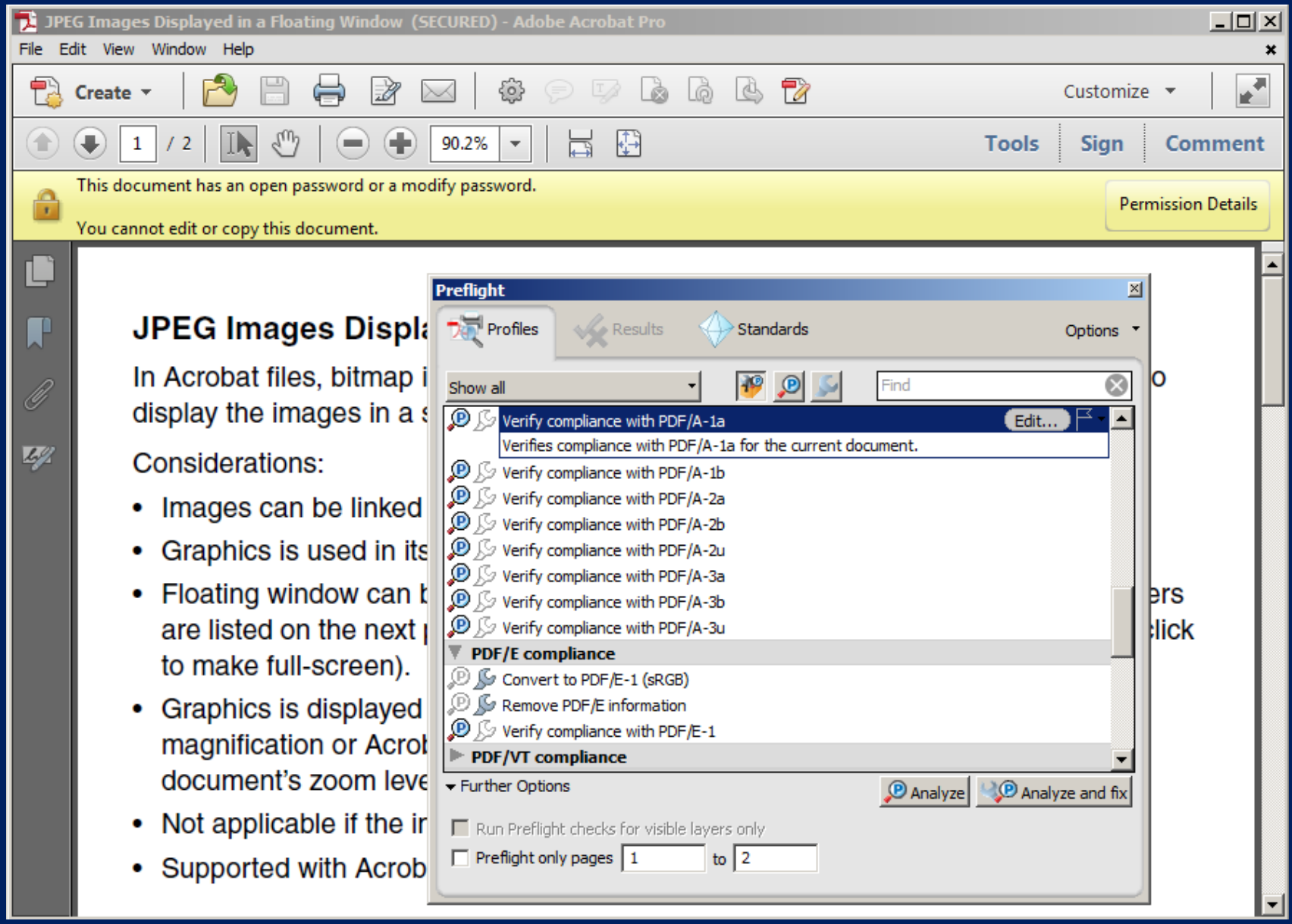
PDF/A

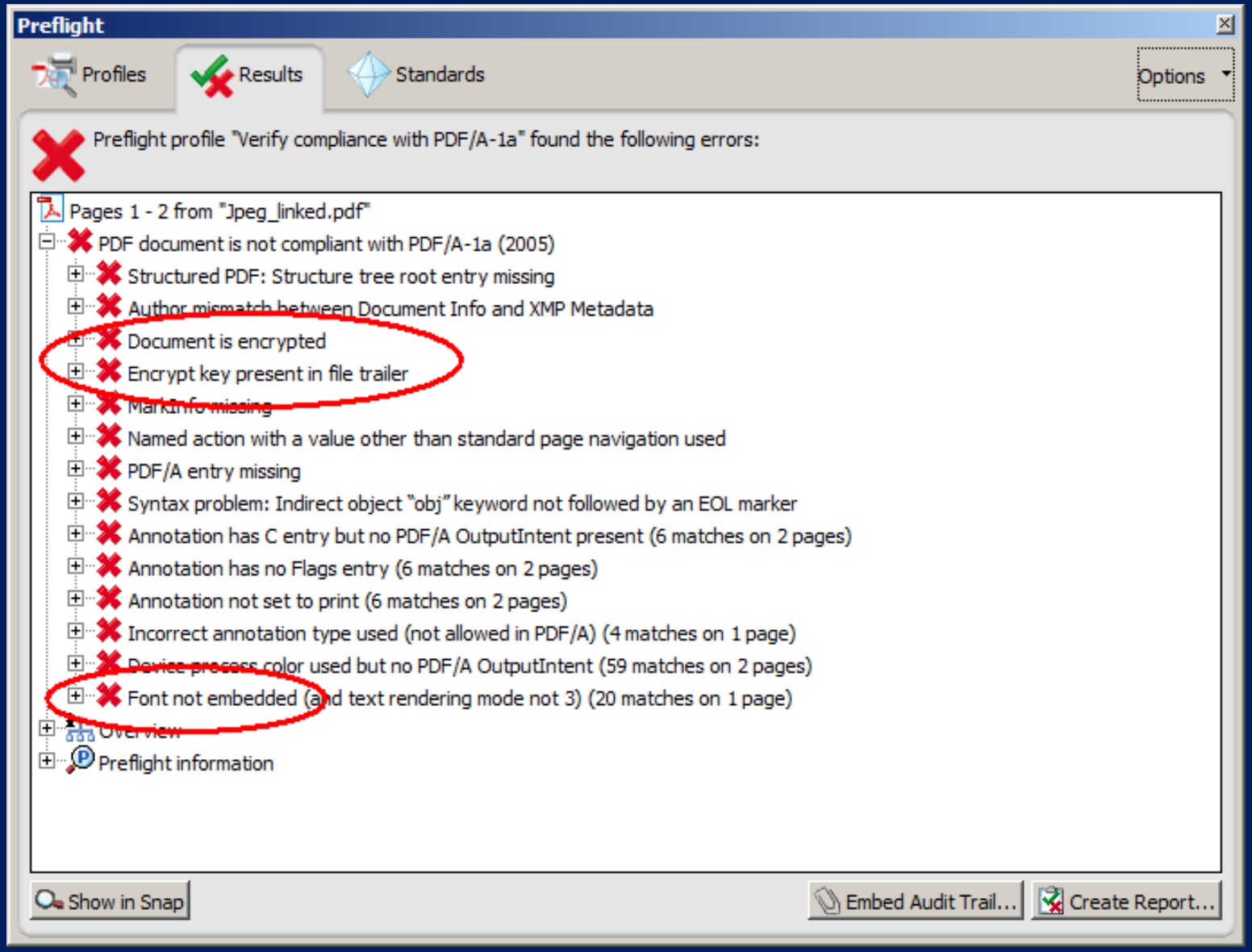

abcd


Identify potential preservation risks of a PDF –*any PDF!*- by assessing against PDF/A standard.

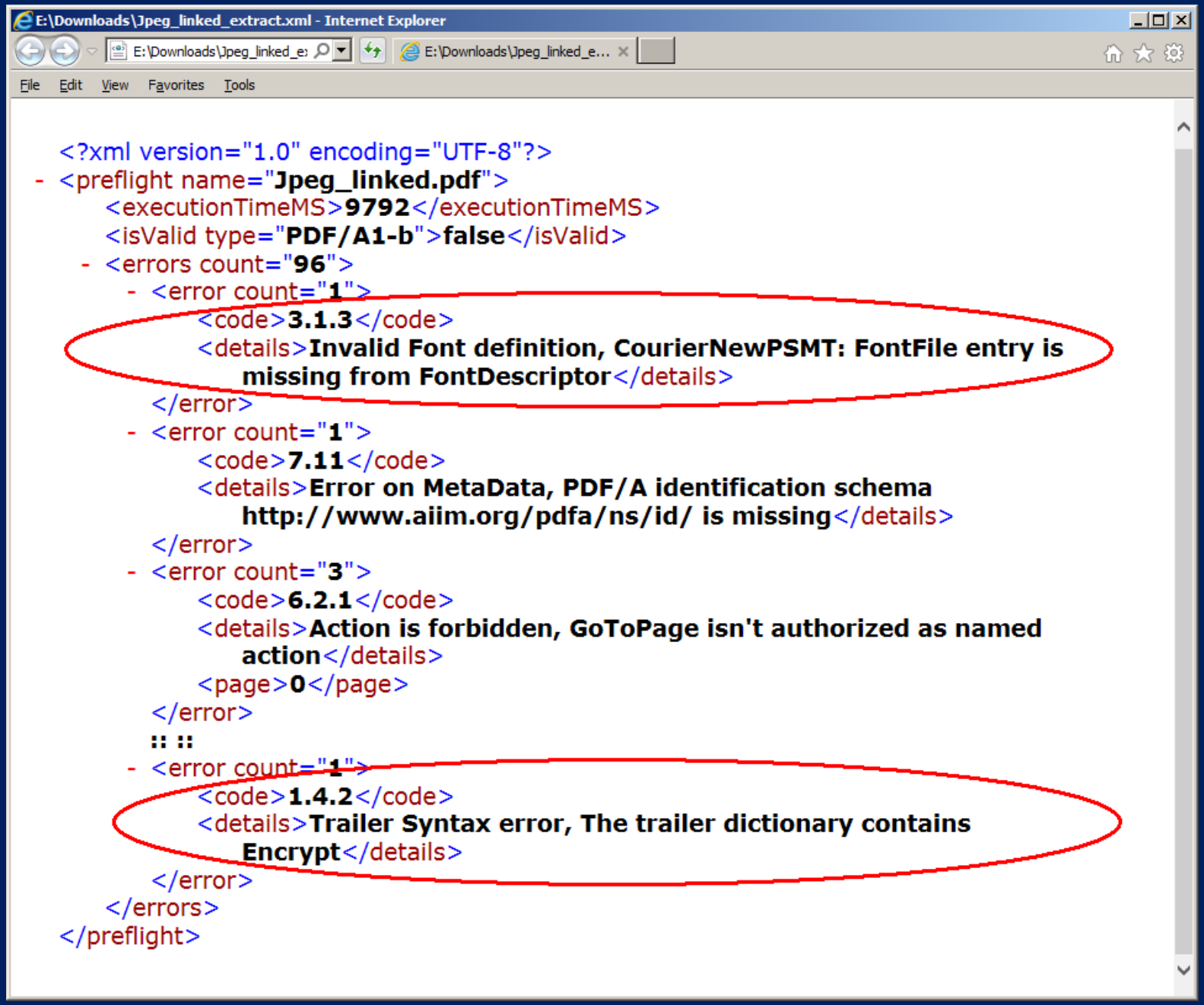
How?

Use PDF/A validator!

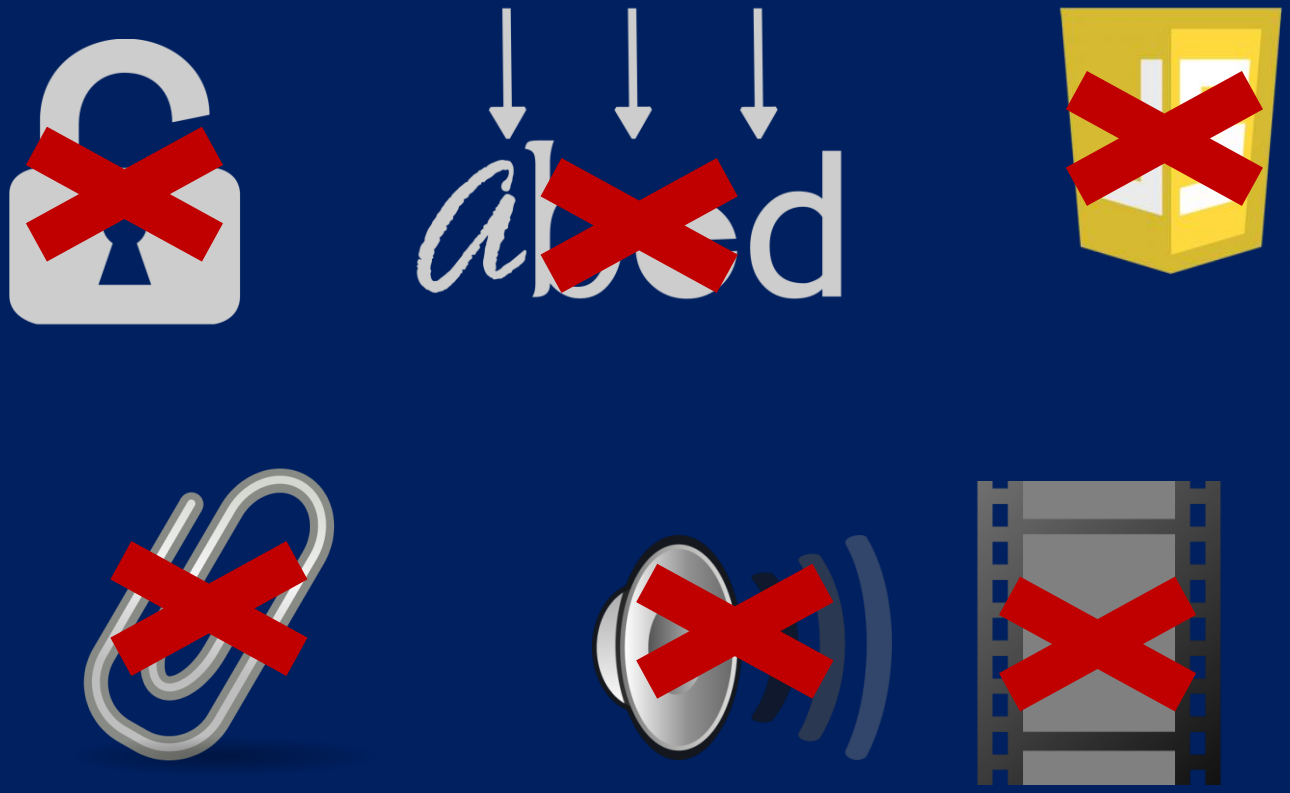

JPEG Images Displayed in a Floating Window (SECURED) - Adobe Acrobat Pro
File Edit View Window Help
Create
Customize
1 / 2
90.2%
Tools
Sign
Comment
This document has an open password or a modify password.
You cannot edit or copy this document.
Permission Details
JPEG Images Displa
In Acrobat files, bitmap i
display the images in a s
Considerations:
• Images can be linked
• Graphics is used in its
• Floating window can b
are listed on the next p
to make full-screen).
• Graphics is displayed
magnification or Acrob
document's zoom leve
• Not applicable if the ir
• Supported with Acrob
Preflight
Profiles
Results
Standards
Options
Show all
Find
Verify compliance with PDF/A-1a
Edit...
Verifies compliance with PDF/A-1a for the current document.
Verify compliance with PDF/A-1b
Verify compliance with PDF/A-2a
Verify compliance with PDF/A-2b
Verify compliance with PDF/A-2u
Verify compliance with PDF/A-3a
Verify compliance with PDF/A-3b
Verify compliance with PDF/A-3u
PDF/E compliance
Convert to PDF/E-1 (sRGB)
Remove PDF/E information
Verify compliance with PDF/E-1
PDF/VT compliance
Further Options
Analyze
Analyze and fix
Run Preflight checks for visible layers only
Preflight only pages 1 to 2

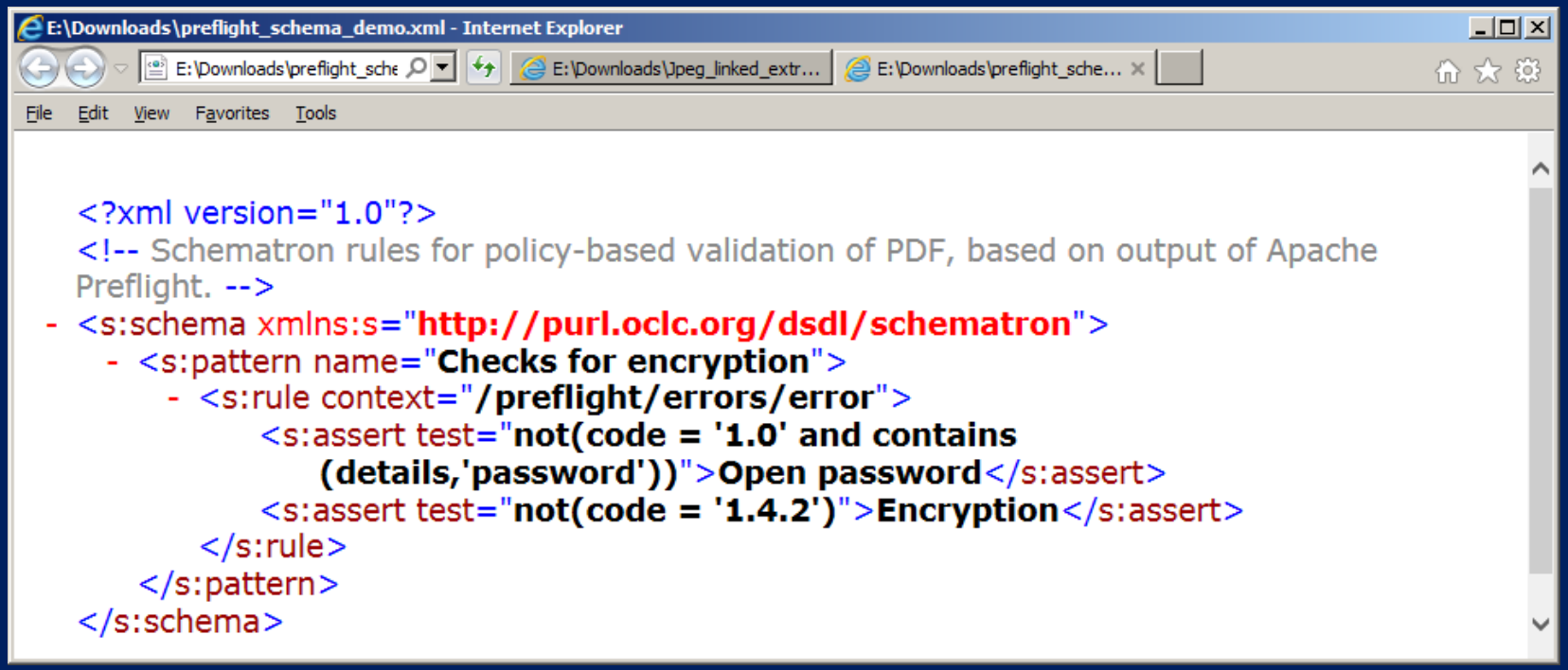

Preflight
Profiles
Results
Standards
Options
Preflight profile "Verify compliance with PDF/A-1a" found the following errors:
Pages 1 - 2 from "Jpeg_linked.pdf"
PDF document is not compliant with PDF/A-1a (2005)
Structured PDF: Structure tree root entry missing
Author mismatch between Document Info and XMP Metadata
Document is encrypted
Encrypt key present in file trailer
MarkInfo missing
Named action with a value other than standard page navigation used
PDF/A entry missing
Syntax problem: Indirect object "obj" keyword not followed by an EOL marker
Annotation has C entry but no PDF/A OutputIntent present (6 matches on 2 pages)
Annotation has no Flags entry (6 matches on 2 pages)
Annotation not set to print (6 matches on 2 pages)
Incorrect annotation type used (not allowed in PDF/A) (4 matches on 1 page)
Device process color used but no PDF/A OutputIntent (59 matches on 2 pages)
Font not embedded (and text rendering mode not 3) (20 matches on 1 page)
Overview
Preflight information
Show in Snap
Embed Audit Trail...
Create Report...

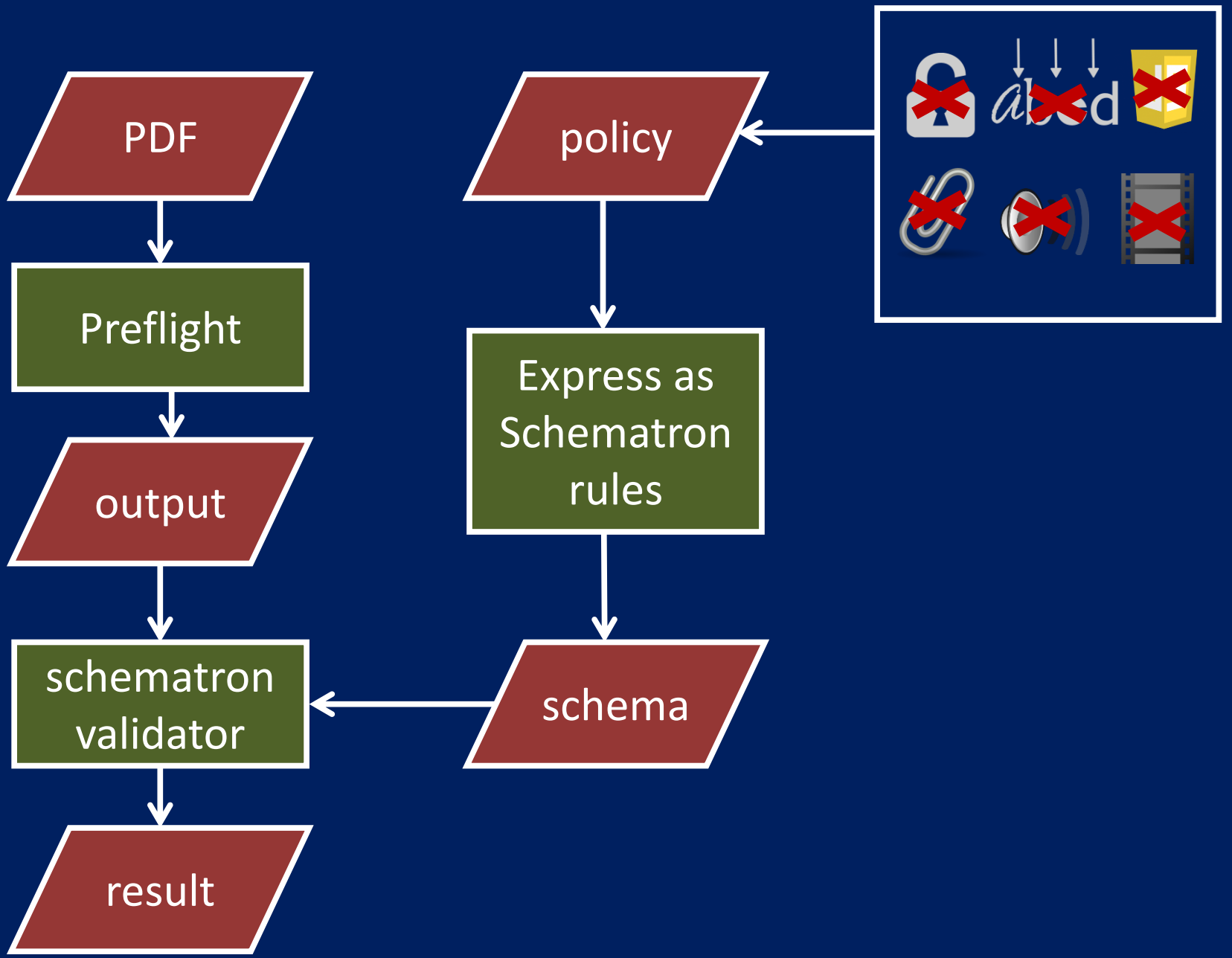
Preflight


E:\Downloads\Jpeg_linked_extract.xml - Internet Explorer
File Edit View Favorites Tools
<?xml version="1.0" encoding="UTF-8"?>
<preflight name="Jpeg_linked.pdf">
<executionTimeMS>9792</executionTimeMS>
<isValid type="PDF/A1-b">false</isValid>
<errors count="96">
<error count="1">
<code>3.1.3</code>
<details>Invalid Font definition, CourierNewPSMT: FontFile entry is missing from FontDescriptor</details>
</error>
<error count="1">
<code>7.11</code>
<details>Error on MetaData, PDF/A identification schema http://www.aiim.org/pdfa/ns/id/ is missing</details>
</error>
<error count="3">
<code>6.2.1</code>
<details>Action is forbidden, GoToPage isn't authorized as named action</details>
<page>0</page>
</error>
:: ::
<error count="1">
<code>1.4.2</code>
<details>Trailer Syntax error, The trailer dictionary contains Encrypt</details>
</error>
</errors>
</preflight>

Policy assessment



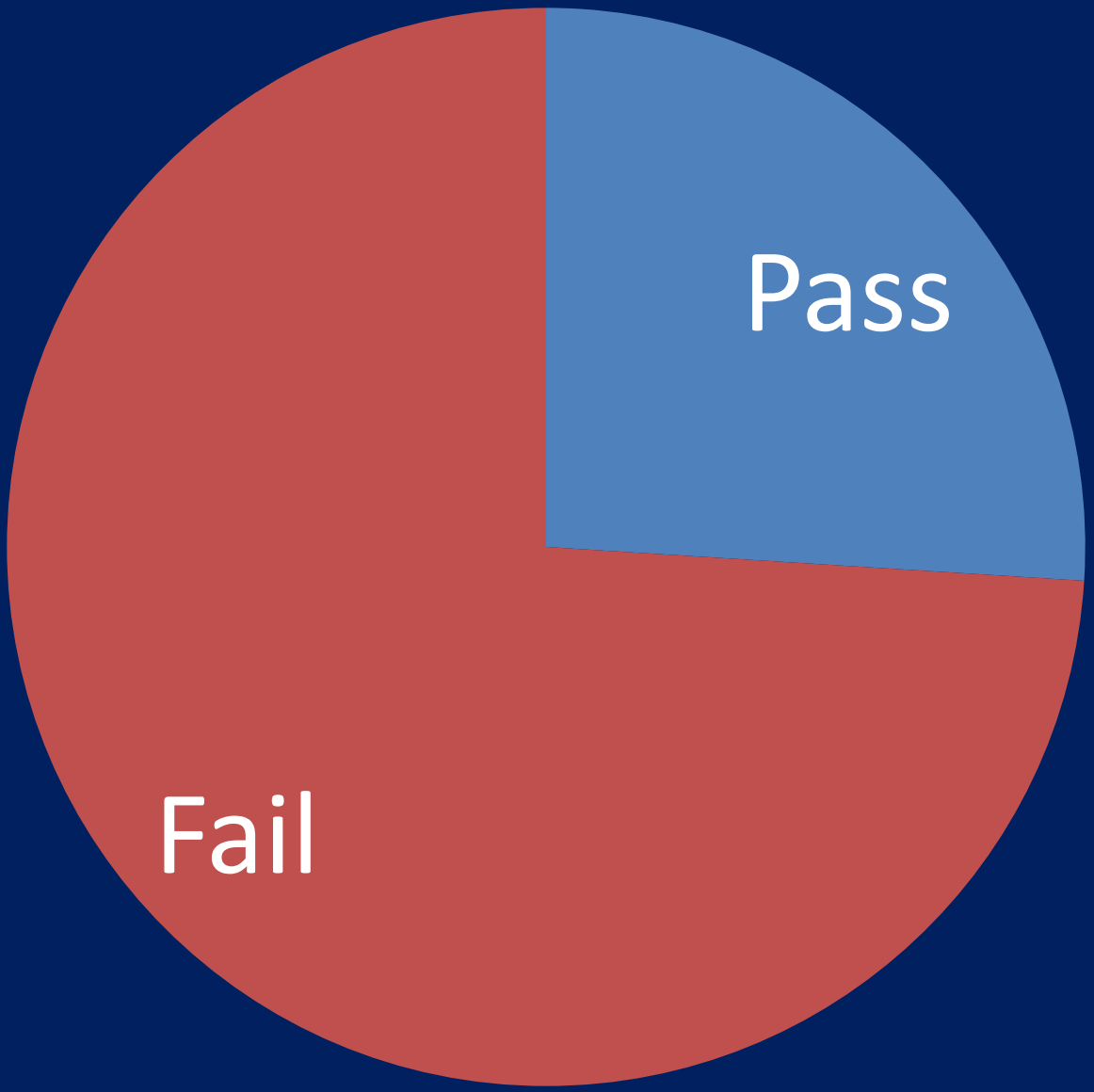

E:\Downloads\preflight_schema_demo.xml - Internet Explorer
E:\Downloads\preflight_sche
E:\Downloads\Jpeg_linked_extr...
E:\Downloads\preflight_sche...
File Edit View Favorites Tools
<?xml version="1.0"?>
<!-- Schematron rules for policy-based validation of PDF, based on output of Apache Preflight. -->
- <s:schema xmlns:s="http://purl.oclc.org/dsdl/schematron">
- <s:pattern name="Checks for encryption">
- <s:rule context="/preflight/errors/error">
<s:assert test="not(code = '1.0' and contains (details,'password'))">Open password</s:assert>
<s:assert test="not(code = '1.4.2')">Encryption</s:assert>
</s:rule>
</s:pattern>
</s:schema>


PDF
Preflight
output
schematron validator
result
policy
Express as Schematron rules
schema

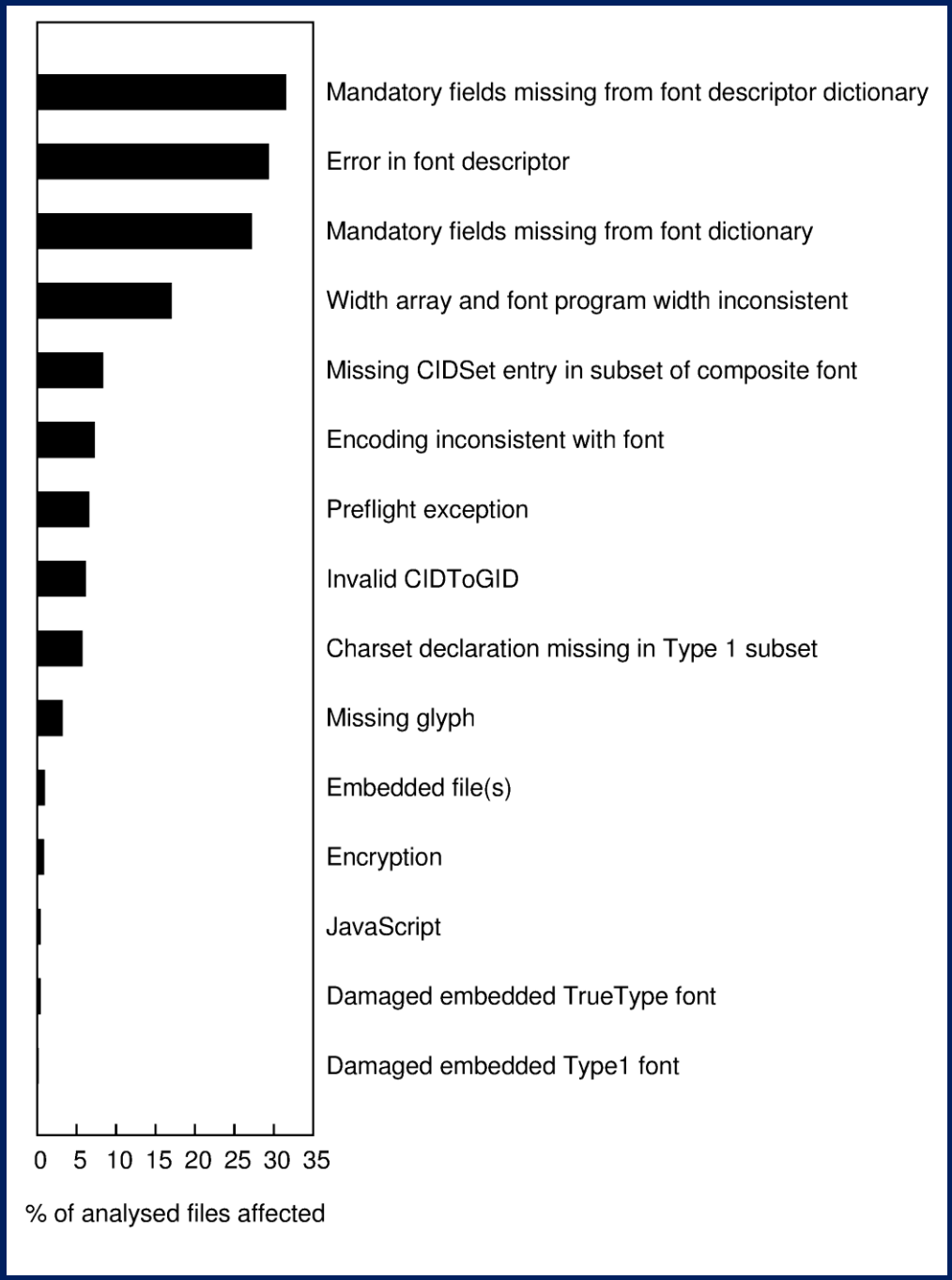
Test corpus

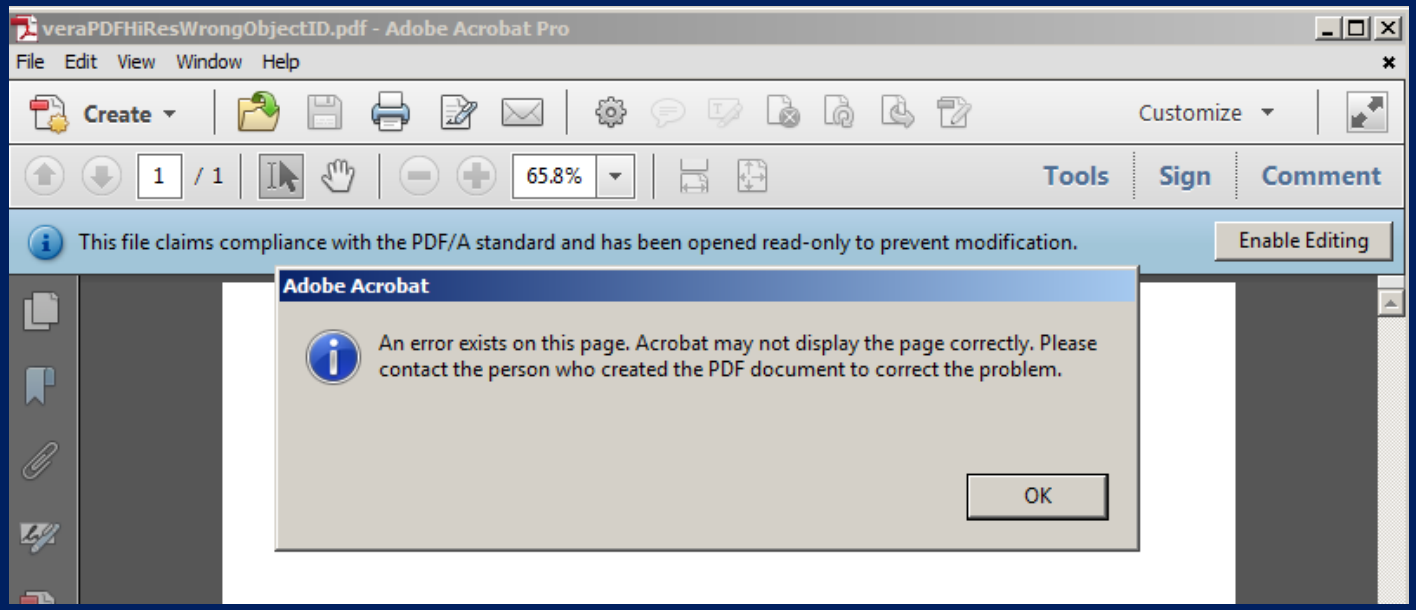
~ 15,000 PDFs from Govdocs1 dataset

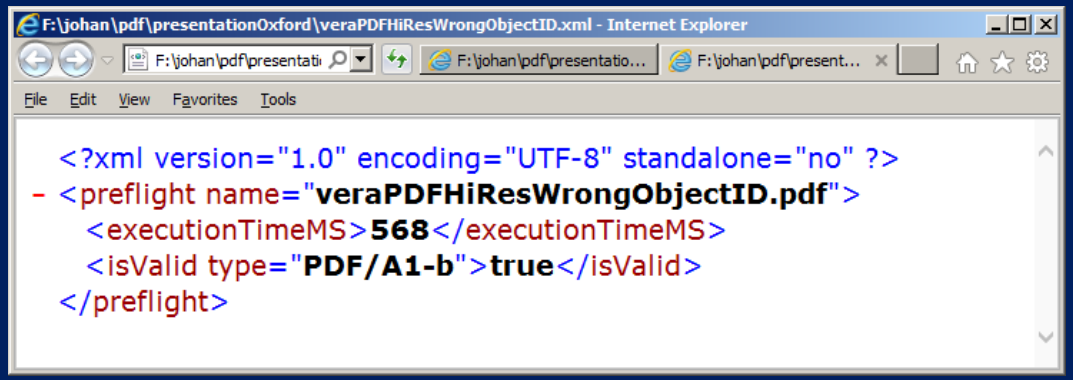
http://digitalcorpora.org/corpora/govdocs

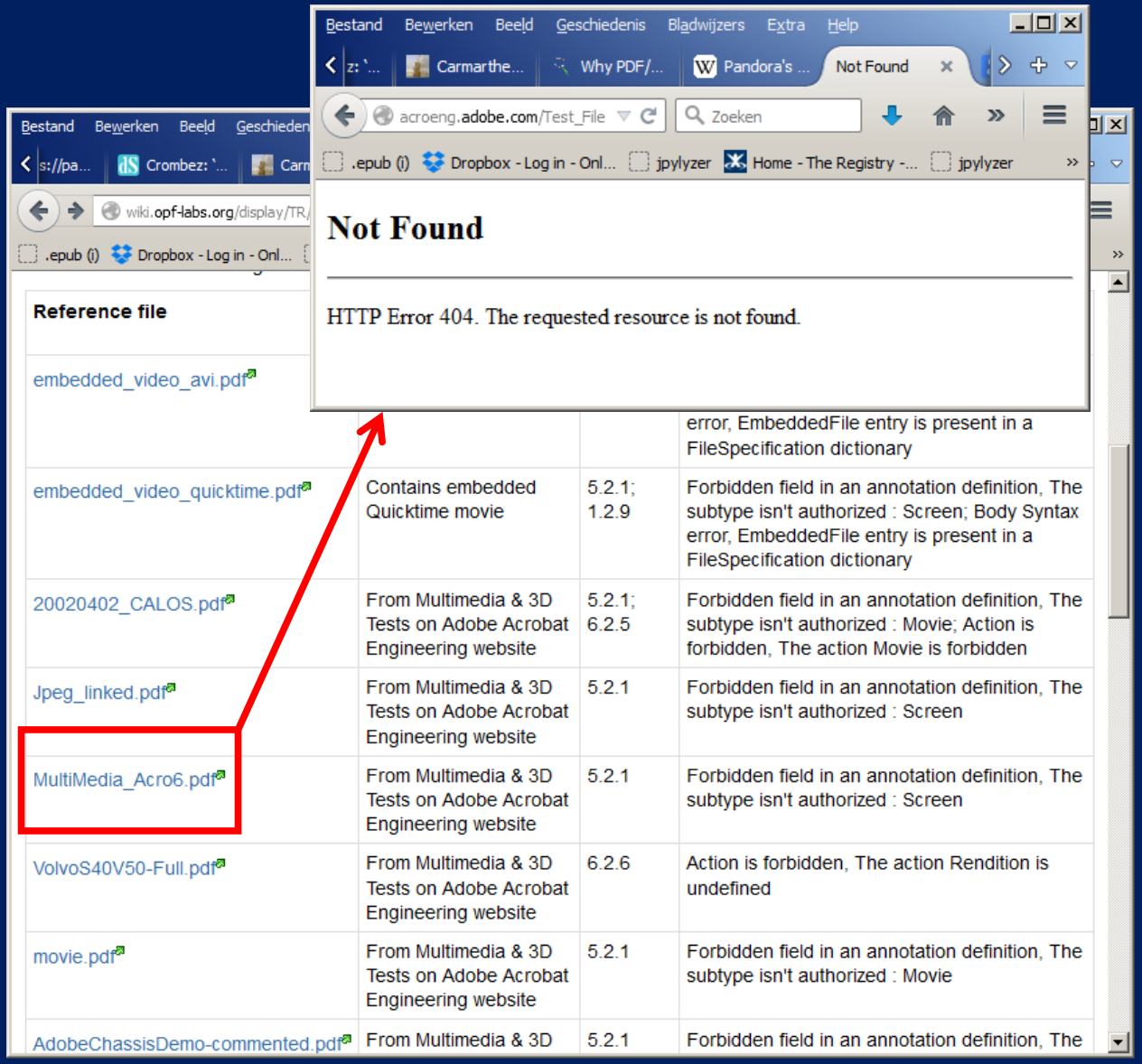

Pass
Fail

Failed tests assessment




veraPDFHiResWrongObjectID.pdf - Adobe Acrobat Pro
File Edit View Window Help
Create
Customize
1 / 1
65.8%
Tools
Sign
Comment
This file claims compliance with the PDF/A standard and has been opened read-only to prevent modification.
Enable Editing
Adobe Acrobat
An error exists on this page. Acrobat may not display the page correctly. Please contact the person who created the PDF document to correct the problem.
OK


F:\johan\pdf\presentationOxford\veraPDFHiResWrongObjectID.xml - Internet Explorer
File Edit View Favorites Tools
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<preflight name="veraPDFHiResWrongObjectID.pdf">
<executionTimeMS>568</executionTimeMS>
<isValid type="PDF/A1-b">true</isValid>
</preflight>


Bestand Bewerken Beeld Geschiedenis Bladwijzers Extra Help
Not Found
acroeng.adobe.com/Test_File
Zoeken
Dropbox - Log in - Onl...
jpylyzer
Home - The Registry -...
Not Found
HTTP Error 404. The requested resource is not found.
wiki.opf-labs.org/display/TR/
Reference file
embedded_video_avi.pdf
embedded_video_quicktime.pdf
Contains embedded Quicktime movie
5.2.1; 1.2.9
Forbidden field in an annotation definition, The subtype isn't authorized : Screen; Body Syntax error, EmbeddedFile entry is present in a FileSpecification dictionary
20020402_CALOS.pdf
From Multimedia & 3D Tests on Adobe Acrobat Engineering website
5.2.1; 6.2.5
Forbidden field in an annotation definition, The subtype isn't authorized : Movie; Action is forbidden, The action Movie is forbidden
Jpeg_linked.pdf
5.2.1
Forbidden field in an annotation definition, The subtype isn't authorized : Screen
MultiMedia_Acro6.pdf
VolvoS40V50-Full.pdf
6.2.6
Action is forbidden, The action Rendition is undefined
movie.pdf
AdobeChassisDemo-commented.pdf

Main challenges

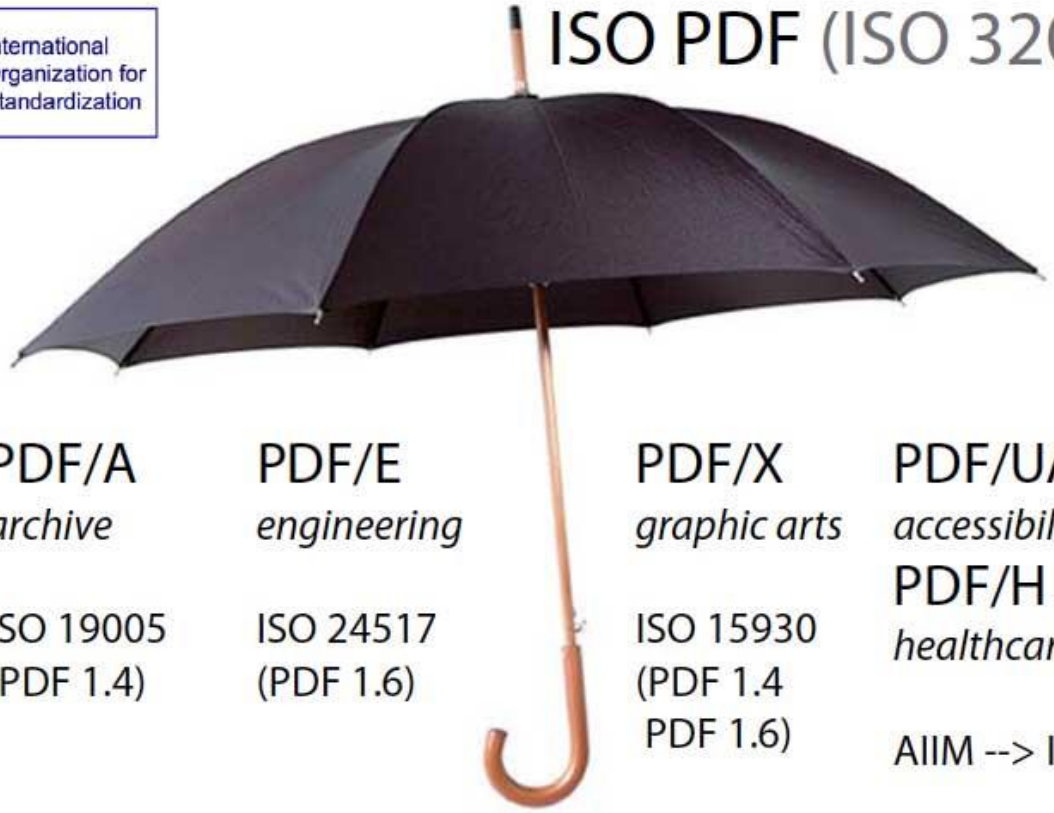
1. Font issues
2. Conformance to ISO 32000
3. Ground truth

veraPDF

E:\Downloads\verapdf-profile-6-2-4-t01.xml - Internet Explorer
File Edit View Favorites Tools
<?xml version="1.0" encoding="UTF-8"?>
<profile model="org.verapdf.model.PDFA1a"
xmlns="http://www.verapdf.org/ValidationProfile">
<name>ISO 19005-1:2005 - 6.2.4 Images - Alternates</name>
<description/>
<creator>veraPDF Consortium</creator>
<created>2015-06-16T22:22:45Z</created>
<hash>sha-1 hash code</hash>
<rules>
<rule object="PDXImage" id="6-2-4-t01">
<description>An Image dictionary shall not contain the Alternates key</description>
<test>Alternates_size == 0</test>
<error>
<message>Alternates key is present in the Image dictionary (</message>
</error>
<reference>
<specification>ISO19005-1</specification>
<clause>6.2.4</clause>
</reference>
</rule>
</rules>
</profile>

International Organization for Standardization

ISO PDF (ISO 32000)



PDF/A
archive

ISO 19005
(PDF 1.4)

PDF/E
engineering

ISO 24517
(PDF 1.6)

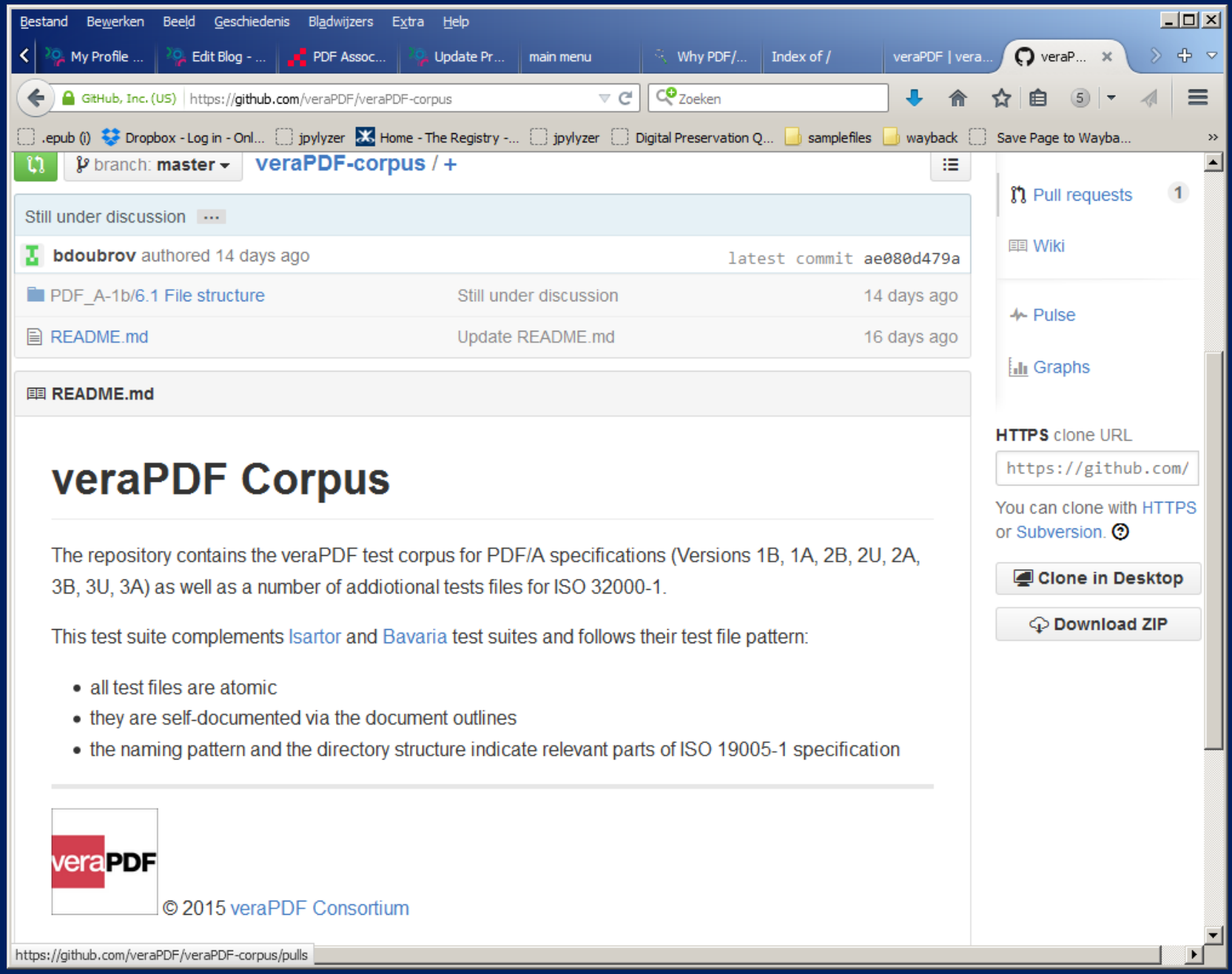
PDF/X
graphic arts

ISO 15930
(PDF 1.4
PDF 1.6)

PDF/UA
accessibility

PDF/H
healthcare

AIIM --> ISO


Bestand Bewerken Beeld Geschiedenis Bladwijzers Extra Help
My Profile ...
Edit Blog - ...
PDF Assoc...
Update Pr...
main menu
Why PDF/...
Index of /
veraPDF | vera...
veraP...
GitHub, Inc. (US)
https://github.com/veraPDF/veraPDF-corpus
Zoeken
.epub (i)
Dropbox - Log in - Onl...
jpylyzer
Home - The Registry -...
jpylyzer
Digital Preservation Q...
samplefiles
wayback
Save Page to Wayba...
branch: master
veraPDF-corpus / +
Still under discussion ...
bdoubrov authored 14 days ago
latest commit ae080d479a
PDF_A-1b/6.1 File structure
Still under discussion
14 days ago
README.md
Update README.md
16 days ago
README.md
veraPDF Corpus
The repository contains the veraPDF test corpus for PDF/A specifications (Versions 1B, 1A, 2B, 2U, 2A, 3B, 3U, 3A) as well as a number of addiotional tests files for ISO 32000-1.
This test suite complements Isartor and Bavaria test suites and follows their test file pattern:
all test files are atomic
they are self-documented via the document outlines
the naming pattern and the directory structure indicate relevant parts of ISO 19005-1 specification
veraPDF
© 2015 veraPDF Consortium
Pull requests 1
Wiki
Pulse
Graphs
HTTPS clone URL
https://github.com/
You can clone with HTTPS or Subversion.
Clone in Desktop
Download ZIP
https://github.com/veraPDF/veraPDF-corpus/pulls

blog.kbresearch.nl/2015/07/07/why-pdfa-validation-matters-even-if-you-dont-have-pdfa

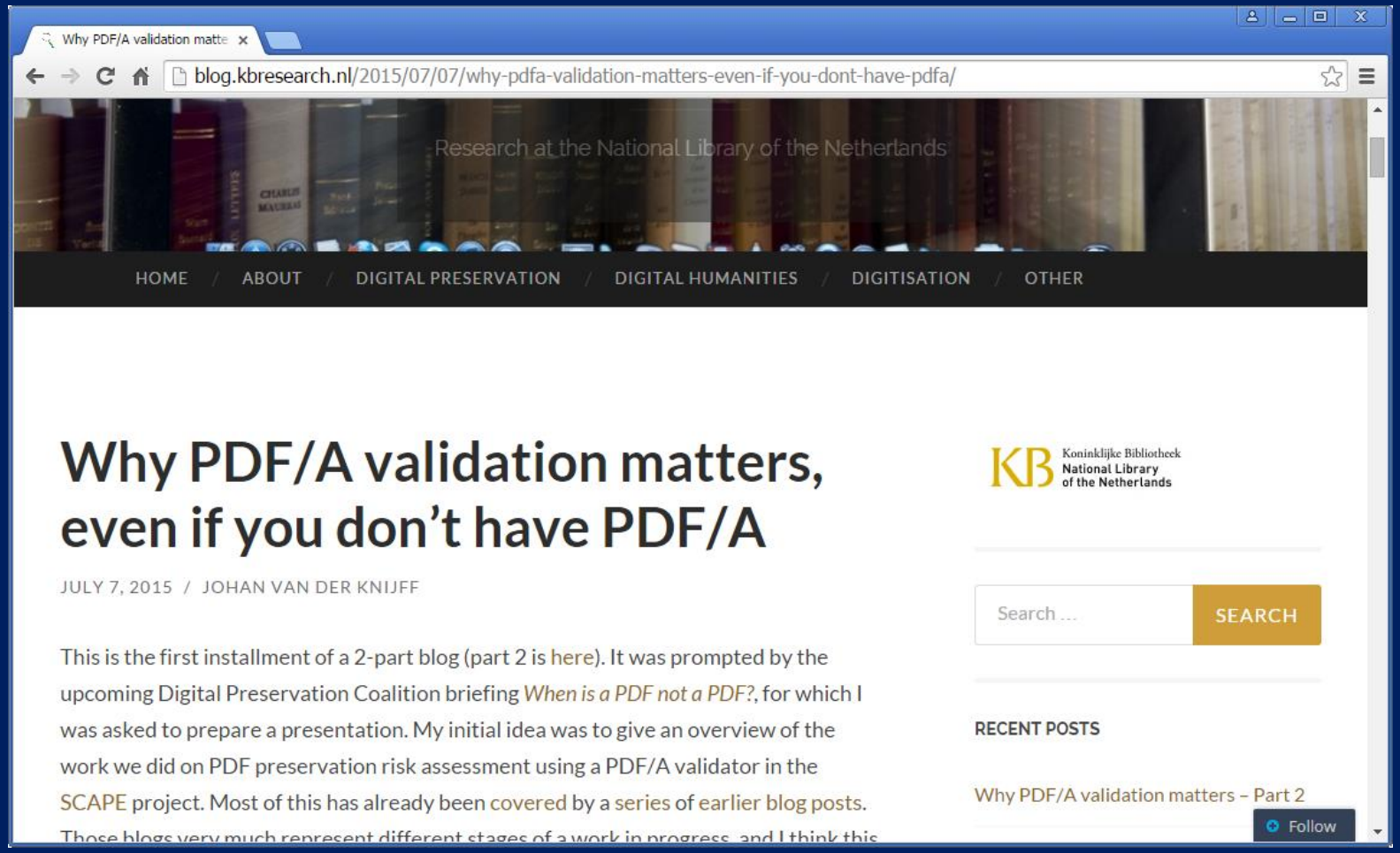


Image attribution

- Unencrypted by Brennan Novak from the Noun Project

* http://thenounproject.com/