



PAKED

Centre of Expertise in Digital Heritage



**(HOW TO PICK)
THE BEST
FILE FORMAT FOR
DIGITAL PRESERVATION**

Bert Lemmens | PACKED vzw
18 May 2015 | Lleida



PACKED

Centre of Expertise in Digital Heritage



PACKED

Expertisecentrum Digitaal Erfgoed

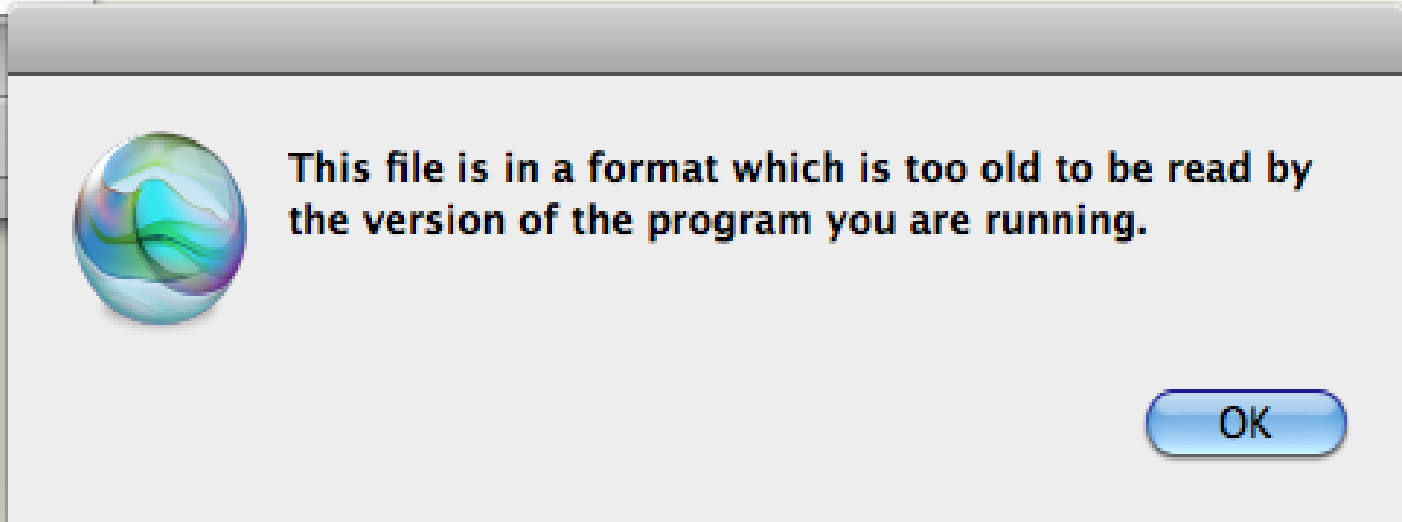
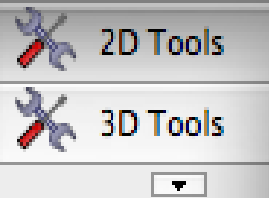
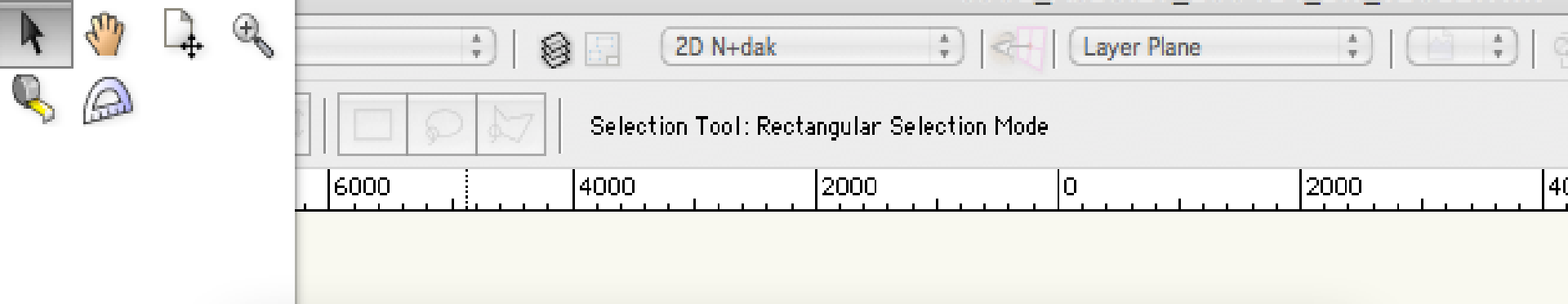
- not-for-profit organisation
- 2005: founded by four museums of contemporary art (Museum Dhondt-Dhaenens, S.M.A.K., M HKA & argos.)
- 2011: funded by the Flemish as Expertise Centre for Digital Heritage..
- gather and further develop knowledge
- disseminate among memory institutions
- initiate innovative projects.
- European projects: LinkedHeritage, ATHENAplus, europeana.space ,PREFORMA,
- packed.be | scart.be | projectcest.be | tracks.be



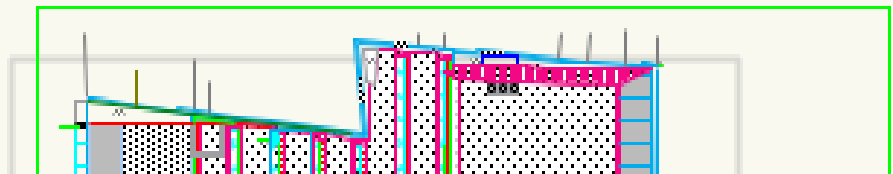
PREFORMA

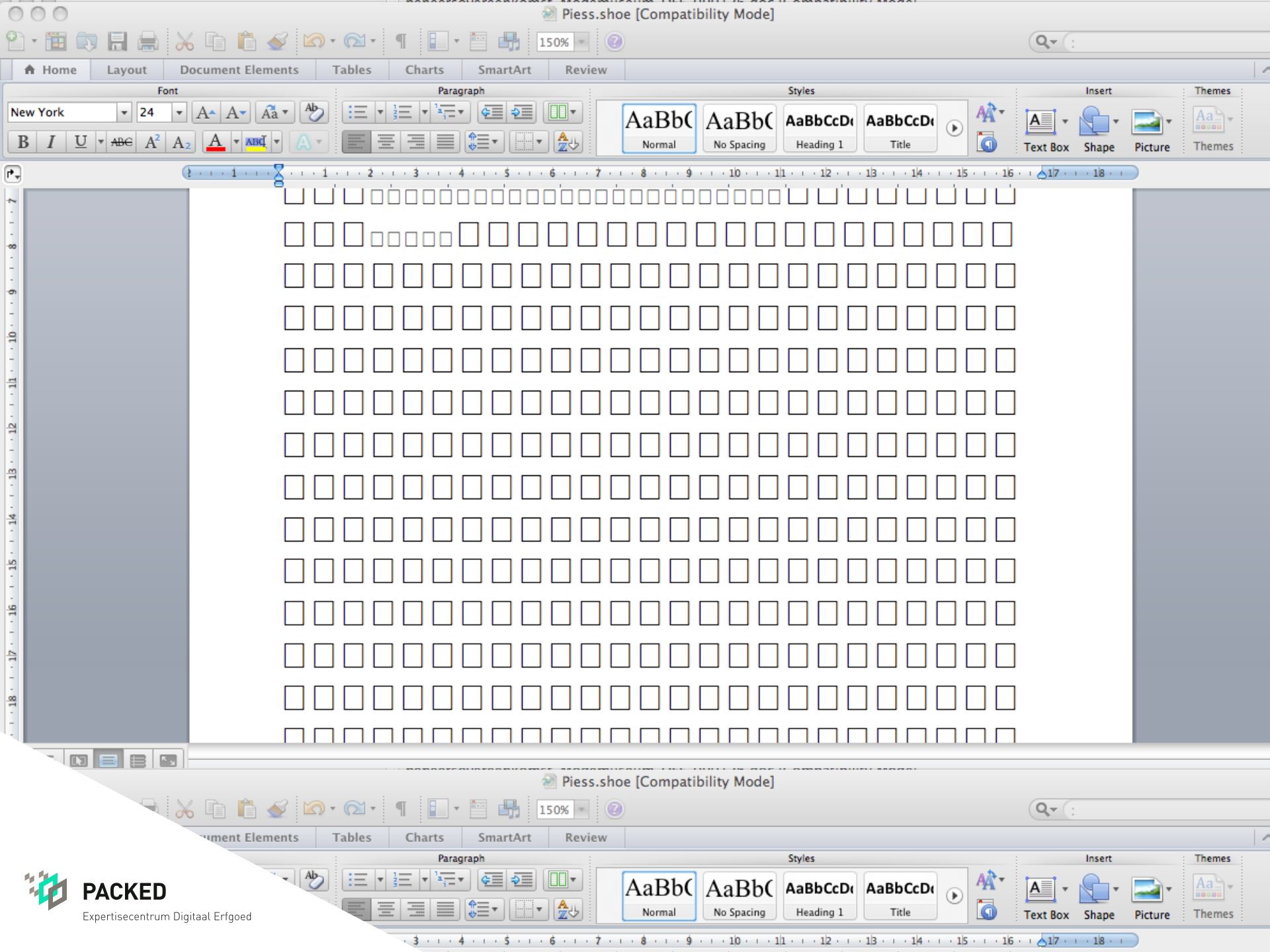
FUTURE MEMORY STANDARDS

- Pre-commercial Procurement project, co-funded by the European commission under the FP7-ICT Programme
- 3.500.00 euro R&D budget
- 48 months (1 Jan 2015 – 31 Dec 2017)
- Riksarkivet, Sweden (Coordinator)
- PROMOTOR, Italy (Technical Coordinator)
- PACKED, Belgium (WP2 Lead – R&D Tender)



4000







A problem has been detected and windows has been shut down to prevent damage to your computer.

UNMOUNTABLE_BOOT_VOLUME

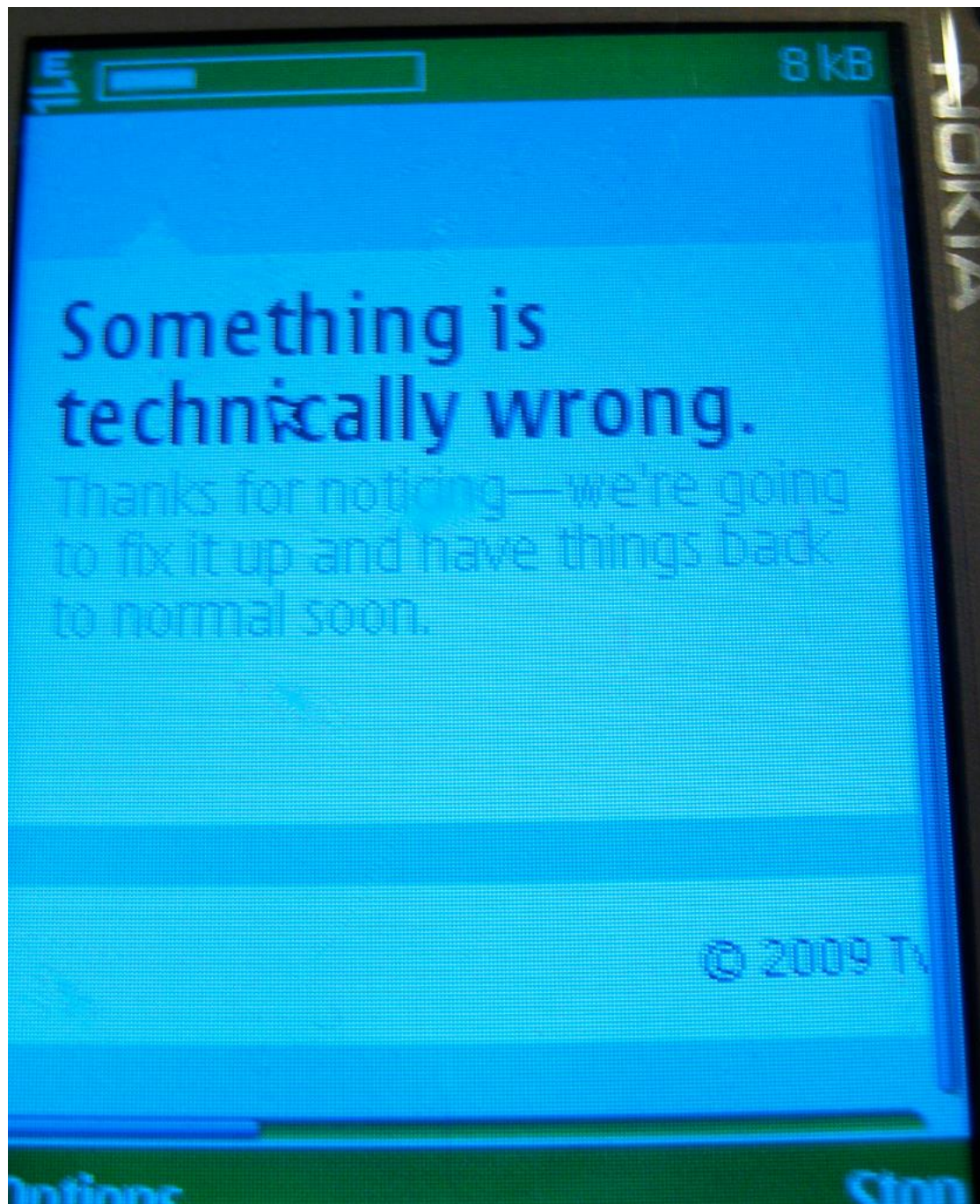
If this is the first time you've seen this stop error screen, restart your computer. If this screen appears again, follow these steps:

Check to make sure any new hardware or software is properly installed. If this is a new installation, ask your hardware or software manufacturer for any windows updates you might need.

If problems continue, disable or remove any newly installed hardware or software. Disable BIOS memory options such as caching or shadowing. If you need to use Safe Mode to remove or disable components, restart your computer, press F8 to select Advanced Startup Options, and then select Safe Mode.

Technical information:

*** STOP: 0x000000ED (0x89BE4168,0xC000009C,0x00000000,0x00000000)



Something is technically wrong.

Thanks for noticing—we're going to fix it up and have things back to normal soon.

© 2009 TM

PRESERVATION FILE FORMAT?

- has a life-cycle that is as long as possible
- capture a digital copy of the original (digital or analog) document in a quality that is high enough to substitute the original in case it is damaged or destroyed and that is
- intended for storing content in a trusted repository.



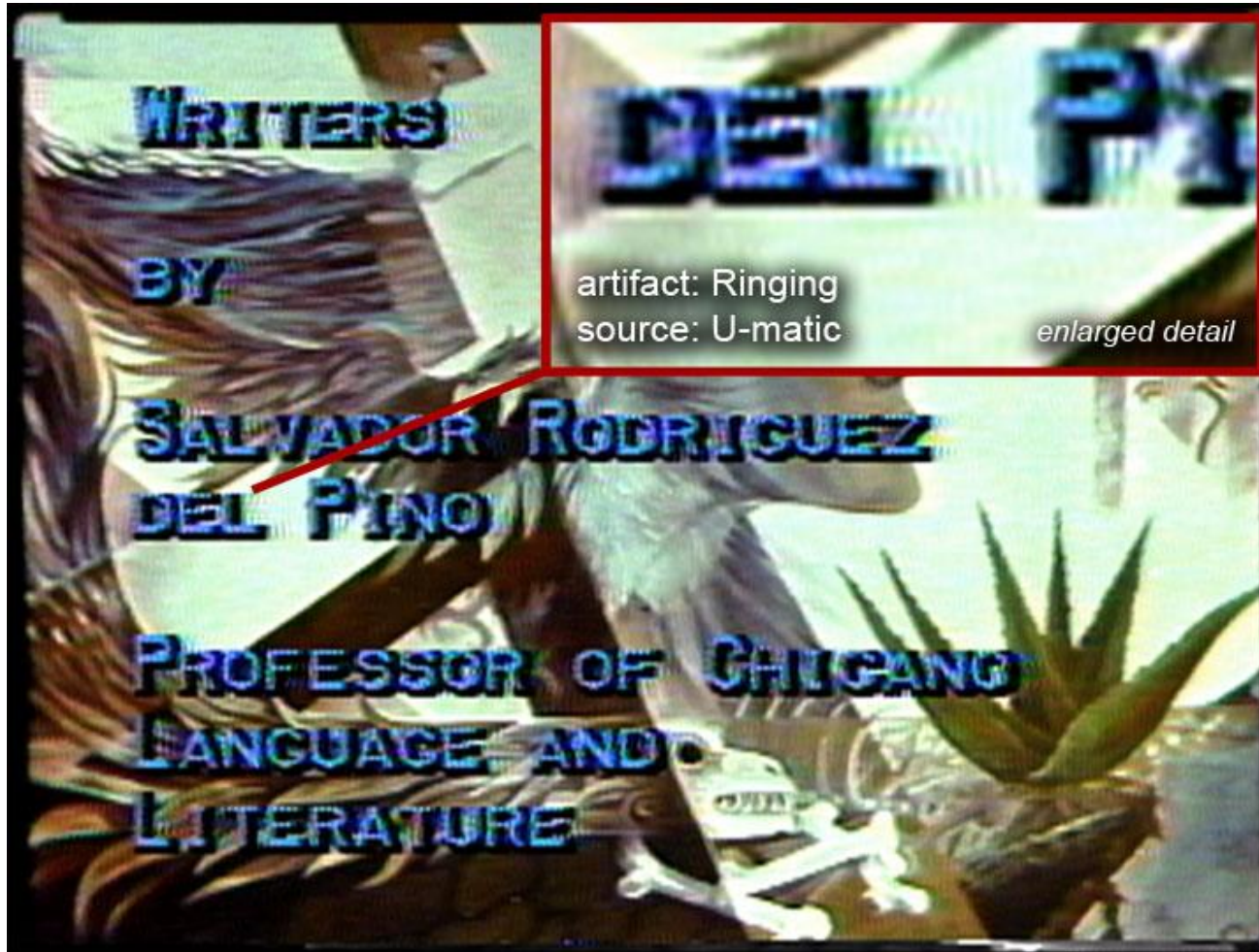
THREE TOPICS

- How do you know if a file format is fit for long term preservation?
- How do you make people use a preservation file format?
- case: PREFORMA pre-commercial procurement

**I.
HOW DO I KNOW
IF A FILE FORMAT IS FIT
FOR LONG TERM PRESERVATION?
(4 CRITERIA)**



#1. CAPTURES UNCOMPRESSED OR LOSSLESS CONTENT

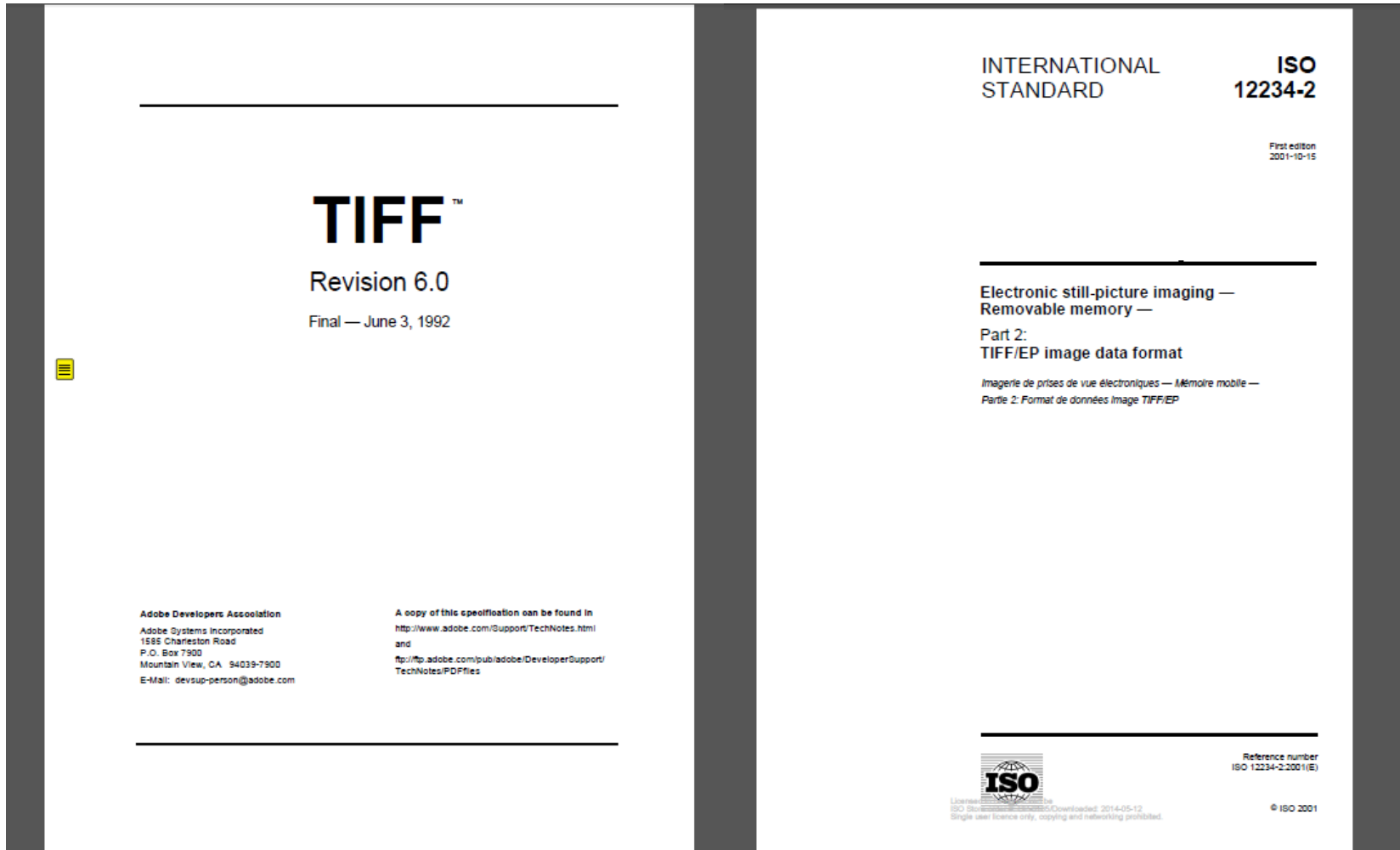


#2. FREE / LIBRE

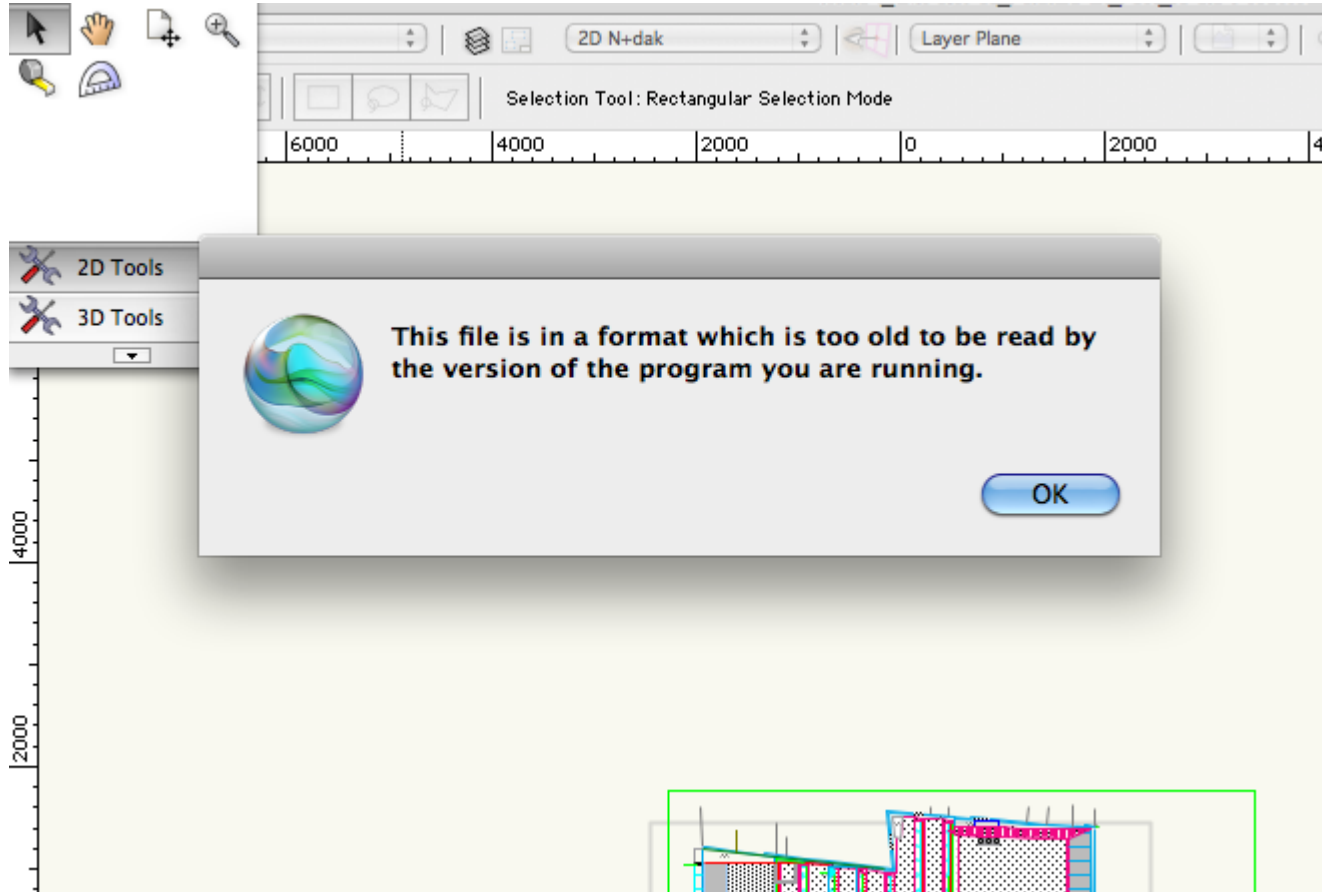
- The standard is adopted and will be **maintained by a not-for-profit organization**, and its ongoing development occurs on the basis of an open decision-making procedure available to all interested parties (consensus or majority decision etc.).
- The standard has been published and the standard specification document is **available either freely or at a nominal charge**. It must be permissible to all to copy, distribute and use it for no fee or at a nominal fee.
- The **intellectual property** - i.e. patents possibly present - of (parts of) the standard is made irrevocably **available on a royalty-free basis**.
- There are no **constraints on the re-use** of the standard

*OpEuropean Interoperability Framework for Pan-European eGovernment Service
(version 1.0 2004)*

#3. WELL DOCUMENTED / STANDARD



#4. ADOPTED BY USERS & SERVICE PROVIDERS:



**II.
HOW DO I
MAKE PEOPLE ADOPT
A PRESERVATION FILE FORMAT?
(3 THREE STRATEGIES)**



#1. PUT IT IN A LAW?

“Cultural heritage organizations that manage a collection should use appropriate, generally accepted standards and high quality, dynamic work methods.”

“Digitization projects must be done in accordance with generally accepted international standards and, if applicable, the standards imposed by the Flemish Government.”

Guide the cultural heritage decree, Flemish Government 2011

#1. PUT IT IN A LAW?

Electronic documents file are in line with the regulation when they are based on **standardized, open** and **non-proprietary** formats.

This policy will be updated according to the evolution of technology.

The preferred format is PDF, in accordance with the technical specifications of the **ISO-19005 PDF / A**.

Digitized documents, regardless of being scanned in or outside the organisation should have the following **technical requirements**:

- a) Apply OCR, whenever there is information of interest and according to assessment of SGDAP.
- b) minimum resolution of 300 dpi.
- c) image compression algorithm CCITT Group 4, should one bit images depth.

Technical annex to the Regulations on the management of electronic records,
Girona City Council, 2007

#2. PUT IT IN A GUIDELINE?

The screenshot shows the homepage of the Project_CES website. The browser address bar displays 'www.projectcest.be/index.php/Hoofdpagina'. The page features a blue pixelated background and a navigation menu with tabs for 'Hoofdpagina', 'CEST Richtlijnen', 'Software', 'Standaardenregister', 'Handleidingen', and 'Glossarium'. A central content area is divided into several sections: 'CEST?' (Cultuurel erfgoed standaarden toolbox), 'Richtlijnen', 'Standaarden', 'Software', 'Handleidingen', and 'Glossarium'. Each section includes a brief description and a 'Lees meer...' link. A sidebar on the right contains 'Instellingen', 'Colofon', and 'Zoek' (search) options. At the bottom, there is a footer with contact information, logos for PACKED and the Flemish Government, and a Creative Commons license.

Hoofdpagina - Project_CES x

www.projectcest.be/index.php/Hoofdpagina

Hoofdpagina CEST Richtlijnen Software Standaardenregister Handleidingen Glossarium

CEST CULTUREEL ERFGOED STANDAARDEN TOOLBOX

Laatste wijzigingen op CEST

- [Opensourcesoftware voor audiovisueel materiaal](#)
- [Nieuwe lay-out voor CEST](#)
- [Nieuwe CEST handleiding Rechten Klaren](#)

Stel je vraag aan PACKED vzw!

CEST? [bewerken]

De [Cultuurel erfgoed standaarden toolbox](#) maakt je wegwijs in het gebruik van standaarden bij het creëren, beheren en toegankelijk maken van digitale erfgoedcollecties.

[Lees meer...](#)

Richtlijnen [bewerken]

De [CEST-richtlijnen](#) helpen je de geschikte standaarden te kiezen om de kwaliteit van je digitaliseringsproject te garanderen. Er zijn minimumrichtlijnen en aanbevolen richtlijnen.

[Lees meer...](#)

Standaarden [bewerken]

Hier vind je een overzicht van de [belangrijkste standaarden](#) voor het digitaliseren van cultureel erfgoed.

[Lees meer...](#)

Software [bewerken]

Hier vind je een overzicht van [softwarepakketten](#) die het leven van een digitalecollectiebeheerder wat makkelijker maken.

[Lees meer...](#)

Handleidingen [bewerken]

Hier vind je [handleidingen](#) van CEST of andere organisaties die je stap voor stap door een bepaald digitaliseringstraject loodsen.

[Lees meer...](#)

Glossarium [bewerken]

Hier vind je een overzicht van [termen](#) met betrekking tot de creëren en beheren van digitaal cultureel erfgoed.

[Lees meer...](#)

INSTELLINGEN

- [Voortuuren](#)
- [Afdrukken](#)
- [Speciale pagina's](#)
- [Print](#)

COLOFON

- [Colofon](#)
- [Over CEST](#)
- [About CEST](#)
- [Hulp](#)
- [Wijzigingen](#)

ZOEK

OK Zoeken

Pagina Overleg Bewerken Geschiedenis Verwijderen Hernoemen Beveiliging opheffen Volgen

Sinds januari 2011 maakt het beheer en de verdere ontwikkeling van CEST deel uit van de beheersverantwoordelijkheid van PACKED vzw en de Vlaamse Gemeenschap. De eerste fase van het project (september tot september 2010) werd uitgevoerd door PACKED vzw en de Vrije Universiteit Brussel.

Met steun van de Vlaamse overheid

Powered by MediaWiki

CC BY-NC-SA

Deze pagina is het laatst bewerkt op 21 okt 2011 om 10:54. Deze pagina is 6.242 maal bekeken. [Privacybeleid](#) [Verkeer](#)

#2. PUT IT IN A GUIDELINE?

The screenshot shows a web browser window with the address bar displaying `digitalpreservation.gov/formats/fdd/fdd000167.shtml`. The page title is "JP2 File Format with JPEG 2000 Core Coding, Lossless". The page content includes a navigation menu, a search box, and several sections: "Table of Contents", "Format Description Properties", "Identification and description", and "Local use".

Table of Contents

- [Identification and description](#)
- [Local use](#)
- [Sustainability factors](#)
- [Quality and functionality factors](#)
- [File type signifiers](#)
- [Notes](#)
- [Format specifications](#)
- [Useful references](#)

Format Description Properties

- ID: fdd000167
- Short name: JP2_J2K_C_LL
- Content categories: still-image
- Format Category: file-format, encoding
- Other facets: unitary, binary, sampled
- Last significant FDD update: 2012-01-09
- Draft status: Full

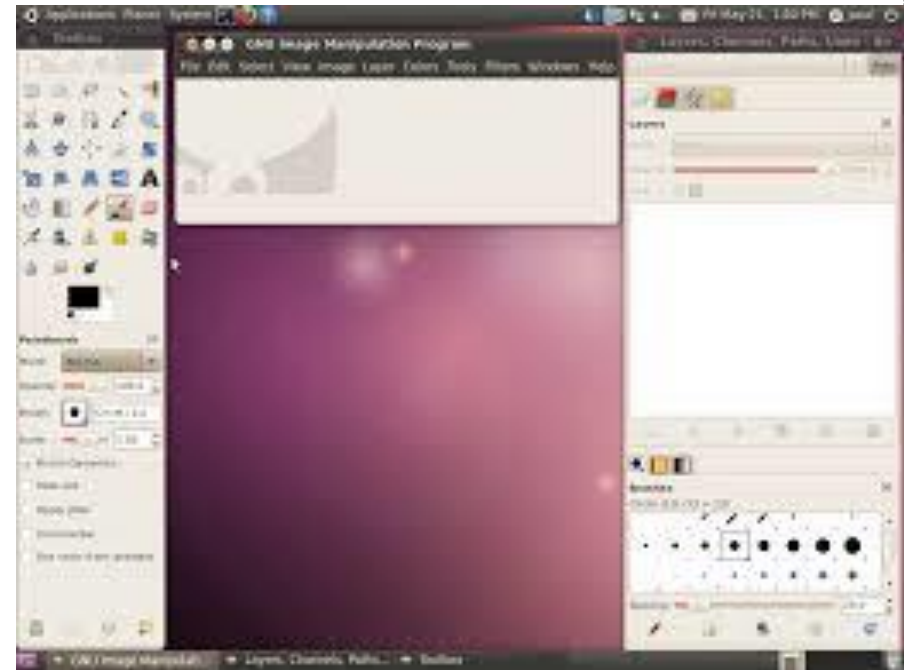
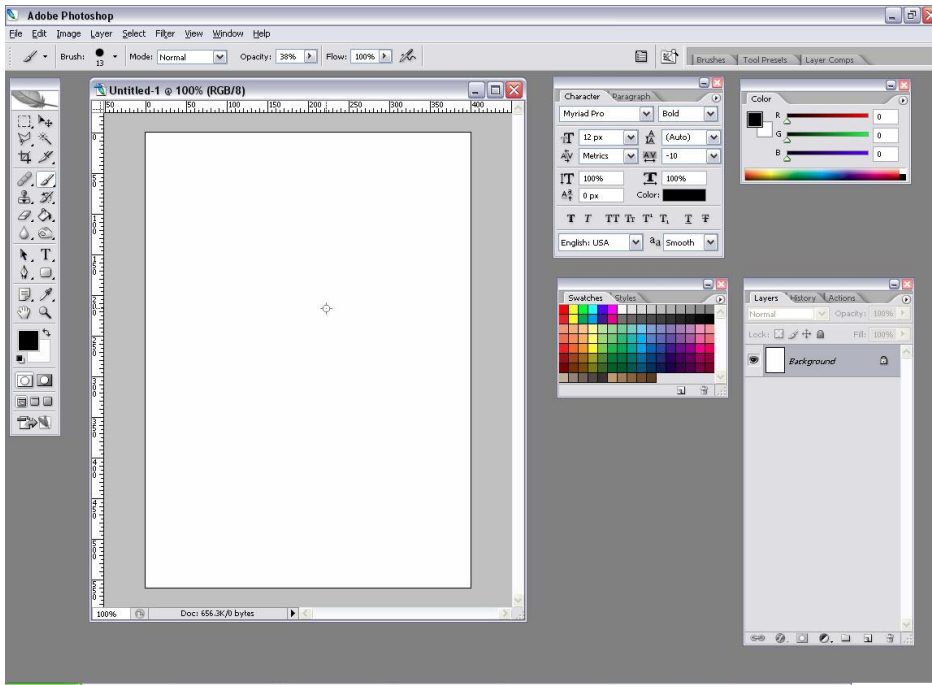
Identification and description

Full name	ISO/IEC 15444-1:2004. Information technology -- JPEG 2000 image coding system -- Part 1: Core coding system, file format as specified in Annex I, with lossless encoded bitstream (formal name); JPEG 2000 jp2 file format with lossless data (common name)
Description	JPEG 2000 core (part 1) wrapper containing lossless encoded data.
Production phase	May be used for content in initial, middle, and final states.
<i>Relationship to other formats</i>	
Subtype of	JP2_FE , JPEG 2000 Part 1 (Core) jp2 File Format
Contains	J2K_C_LL , JPEG 2000 Part 1, Core Coding, Lossless

Local use

LC experience or existing holdings	Video reformatting in the Library's Motion Picture, Broadcasting, and Recorded Sound Division employs individual frame images with lossless compression;
------------------------------------	--

#3. PROVIDE TOOLS THAT WORK FROM-THE-SHELF?



**PREFORMA:
“WHAT ARE THE THRESHOLDS FOR
ADOPTING A ‘GOOD’ PRESERVATION
FILE FORMAT IN A DIGITAL
REPOSITORY WORKFLOW?”**



PACKED

Centre of Expertise in Digital Heritage

A 'VALIDATOR'!

**SOMETHING THAT TELLS YOU IF A
FILE IS WHAT IT CLAIMS TO BE**



PACKED

Centre of Expertise in Digital Heritage

III. CASE: PREFORMA A PRE-COMMERCIAL PROCUREMENT (A BIT MORE ABOUT THE LAST STRATEGY)



PRE-COMMERCIAL PROCUREMENT

- procuring R&D services
- developing innovative solutions that address specific public sector challenges and needs
- competition-like procurement method
- phased procurement of development contracts to reduce risk.
- end result: multiple solutions for the same problem that almost ready to bring on the market

FOR WHICH FILES FORMAT?

MXF | MPEG | IMX | XDCAM
HD422 | DPX | DCP | JPEG2000 |
MOV | MPEG2 | AVI | MPEG4 |
AVC | PDF 1.4 | PDF/A1 | TIFF 6.0 |
JPEG | RAW | AS-07 | MPEG|AF |
PDF | MKV | FFV1 | OGG | Dirac |
PNG | WebM | VP8 | OGG | Theora
| PDF/A2 | PDF/A3 | LPCM



WHAT DID THE 'EXPERTS' SAY?

	AUDIO VISUAL			TEXT	IMAGE
	broadcast	film	Consumer		
PREFORMA stakeholders	MPEG-IMX (MXF/MPEG2) XDCAM HD422 (MXF/MPEG4)	DPX DCP (MXF/JPEG2000)	MOV/MPEG2 AVI/MPEG2 MPEG/MPEG2 MPEG/MPEG4 -AVC	PDF 1.4 PDF/A1	TIFF 6.0 JPEG JPEG2000 RAW
Industry standards	AS-07 (MXF/MPEG2) (MXF/JPEG2000) FIMS (MXF/MPEG2)	DCDM (TIFF 6.0) DCP (MXF/JPEG2000) IMF (MXF/MPEG4)	MPEG-AF	PDF	JPEG2000 TIFF
Developers	MKV/FFV1 OGG/Dirac	PNG	WebM/VP8 OGG/Theora	PDF/A1 PDF/A3 PDF/A3	PNG

WHAT DID THE GUIDELINES SAY?

uncompressed/ lossless?	AUDIO VISUAL			TEXT	IMAGE
	broadcast	film	Consumer		
PREFORMA stakeholders	MPEG-IMX (MXF/MPEG2) XDCAM HD422 (MXF/MPEG4)	DPX DCP (MXF/JPEG2000)	MOV/MPEG2 AVI/MPEG2 MPEG/MPEG2 MPEG/MPEG4 -AVC	PDF 1.4 PDF/A1	TIFF 6.0 JPEG JPEG2000 RAW
Industry standards	AS 07 (MXF/MPEG2) (MXF/JPEG2000) FIMS (MXF/MPEG2)	DCDM (TIFF 6.0) DCP (MXF/JPEG2000) IMF (MXF/MPEG4)	MPEG-AF	PDF	JPEG2000 TIFF
Open standards	MKV/FFV1 OGG/Dirac	PNG	WebM/VP8 OGG/Theora	PDF/A1 PDF/A3 PDF/A3	PNG

WHAT DID THE (SWEDISH) LEGAL EXPERT SAY?

Open standard?	AUDIO VISUAL			TEXT	IMAGE
	broadcast	film	Consumer		
PREFORMA stakeholders	MPEG-IMX (MXF/MPEG2) XDCAM HD422 (MXF/MPEG4)	DPX DCP (MXF/JPEG2000)	MOV/MPEG2 AVI/MPEG2 MPEG/MPEG2 MPEG/MPEG4 -AVC	PDF 1.4 PDF/A1	TIFF 6.0 JPEG JPEG2000 RAW
Industry standards	AS 07 (MXF/MPEG2) (MXF/JPEG2000) FIMS (MXF/MPEG2)	DCDM (TIFF 6.0) DCP (MXF/JPEG2000) IMF (MXF/MPEG4)	MPEG-AF	PDF	JPEG2000 TIFF
Open standards	MKV/FFV1 OGG/Dirac	PNG	WebM/VP8 OGG/Theora	PDF/A1 PDF/A3 PDF/A3	PNG

WHAT DID OTHER COLLECTIONS SAY?

adoption?

	AUDIO VISUAL			TEXT	IMAGE
	broadcast	film	Consumer		
PREFORMA stakeholders	MPEG-IMX (MXF/MPEG2) XDCAM HD422 (MXF/MPEG4)	DPX DCP (MXF/JPEG2000)	MOV/MPEG2 AVI/MPEG2 MPEG/MPEG2 MPEG/MPEG4-AVC	PDF 1.4 PDF/A1	TIFF 6.0 JPEG JPEG2000 RAW
Industry standards	AS 07 (MXF/MPEG2) (MXF/JPEG2000) FIMS (MXF/MPEG2)	DCDM (TIFF 6.0) DCP (MXF/JPEG2000) IMF (MXF/MPEG4)	MPEG-AF	PDF	JPEG2000 TIFF
Open standards	MKV/FFV1 OGG/Dirac	PNG	WebM/VP8 OGG/Theora	PDF/A1 PDF/A3 PDF/A3	PNG

WHAT WE BECAME AWARE OF...

- (technically) open \neq (free/libre) open format
- standard formats \leftrightarrow multiple implementations
- technically appropriate formats $>$ little support/adoption

- too often we copy format guidelines/policies without checking if they are correct/appropriate



WHAT WE HAVE CHOSEN (EVENTUALLY...)

TEXT (strengthen the consensus)

- ISO 32000-1:2008 (PDF 1.7)
- ISO 19005-1:2005 (PDF/A-1)
- ISO 19005-2:2011 (PDF/A-2)
- ISO 19005-3:2012 (PDF/A-3)

IMAGE (improve the consensus)

- ISO 12234-2:2001 (TIFF/EP)
- ISO 12369:2004 (TIFF/IT)

MOVING IMAGE (virgin path...)

- OGG / MKV
- FFV1 / Dirac / ISO 15444-1 (JPEG2000 core coding system)
- LPCM



WHERE ARE WE NOW?

- Open Preservation Foundation – VeraPDFa (PDF/A)
- EasyInnova – DPF Manager (TIFF)
- MediaInfo – MediaConch (MKV|FFV1|LPCM)

- first public release: July 2015 (Quarterly)
- Open Source Workshop, Stockholm: April 2016





Search GitHub

[Explore](#) [Gist](#) [Blog](#) [Help](#)

bert-packed

**veraPDF**

Definitive PDF/A Validation

<http://verapdf.org>

Filters ▾

🔍 Find a repository...

People

0 >

This organization has no public members.
You must be a member to see who's a part
of this organization.

veraPDF-model

★ 0 📄 2

PDF model described with veraPDF syntax

Updated 12 minutes ago

veraPDF-corpus-PDFA-1b

★ 2 📄 1

Test corpus for PDF/A Part 1 Level B

Updated 12 minutes ago

pdfbox

Java ★ 0 📄 132

🍴 forked from [apache/pdfbox](#)

Mirror of Apache PDFBox

Updated 20 hours ago

veraPDF-library-GUI

★ 0 📄 0

Updated 21 hours ago

DPF Manager: Digital Preservation Formats Manager (Image files)

2 commits 2 branches 0 releases 1 contributor

branch: master DPFManager / +

Update README.md		
easyinnova	authored on Dec 22, 2014	latest commit 002ca1db42
LICENSE	Initial commit	5 months ago
README.md	Update README.md	5 months ago

README.md

DPFManager

DPF Manager: Digital Preservation Formats Manager (Image files)

Easy Innova proposes the development of a unique open platform specially designed to be a reference on conformance validation and file preparation of multiple standard formats for long-term preservation. One of its main features will be the capability to be integrated with other open source projects and proprietary applications, providing in a simple way a great range of tools to gain full control over the validation and preservation of digital content. The platform will be strategically modularized and prepared to easily incorporate conformance checkers for other formats and other multiple value-added functionalities (viewers, conversion, repair, optimization, ...).

Code

Issues 0

Pull requests 0

Wiki

Pulse

Graphs

HTTPS clone URL

<https://github.com/>

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

Clone in Desktop

Download ZIP



This repository Search

[Explore](#)
[Gist](#)
[Blog](#)
[Help](#)

bert-packed

MediaArea / **MediaConch**

Watch 10

Star 5

Fork 0

MediaConch (funded by PREFORMA)

713 commits

3 branches

3 releases

6 contributors



branch: master

MediaConch / +



adds link to release and demo



ablwr authored 8 hours ago

latest commit e4d2abf38c



ConformityChecks

removes old schematron

2 days ago



Demo

adds example of schematron within range

2 days ago



Design

draft reporter XML output

a month ago



Documentation

removes old schematron

2 days ago



Interviews

namechange across the board

a month ago



LettersOfSupport

added Pip LoS

3 months ago



Outreach

namechange across the board

a month ago



Press

namechange across the board

a month ago



ProjectReports

namechange except for report assembler

a month ago



Questionnaire

namechange across the board

a month ago



RelatedConformanceReports

JHOVE preference GUI

3 months ago



SampleTestFiles

small Matroska test file

4 days ago



SourceCode

Help.cpp: prefer lowercase mediaconch in help doc

2 days ago



.gitignore

adds .DS_Store to global ignore for us Mac users browsing test files ...

3 days ago



ArchitectureIntroduction.md

stress sectional headers

3 months ago



Code



Issues

6



Pull requests

0



Wiki



Pulse



Graphs

HTTP \$ clone URL

<https://github.com/>

 You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

Clone in Desktop

Download ZIP



This project has received funding from the European Union's Seventh Framework Programme under grant agreement no 619568

[HOME](#) [PROJECT](#) [PARTNERS](#) [TENDER](#) [EVENTS](#) [OPEN SOURCE PORTAL](#) [COMMUNITY](#) [DOWNLOAD](#) [CONTACTS](#)

COMMUNITY

PREFORMA is open to collaboration with experts, organisations, institutions and other projects in order to find synergies and discuss opportunities for cooperation.

The network of common interest of PREFORMA will be set up with representatives from memory institutions, researchers and developers who are interested in contributing to the definition of requirements and taking part in the assessment and exploitation of the results produced by the suppliers.

We are very much interested in engaging with:

- ▶ **Memory institutions** and cultural heritage organisations coordinating or representing memory institutions, that are involved in (or planning) digital culture initiatives, to use the tools developed by PREFORMA.
- ▶ **Developers** contributing code for the PREFORMA tools as well as developers implementing the reference implementations in production software.
- ▶ **Research organisations** providing technical advice and expertise to cultural stakeholders.

RELATED PROJECTS

[APARSEN - Alliance for Permanent Access](#)

[DAVID - Digital AV Media Damage Prevention and Repair](#)

[DCH-RP - Digital Cultural Heritage - Roadmap for Preservation](#)

[Europeana Creative](#)

[Europeana Tech](#)

[ForgetIT](#)

[NCDD - Netherlands Coalition for Digital Preservation](#)

[OSSCOM - Open Source Software Communities](#)

[PERICLES - Promoting and Enhancing Reuse of Information throughout the](#)



THANK YOU!

questions?