KB

SCAPE
SCAlable Preservation Environments

PDF/A-3 for preservation

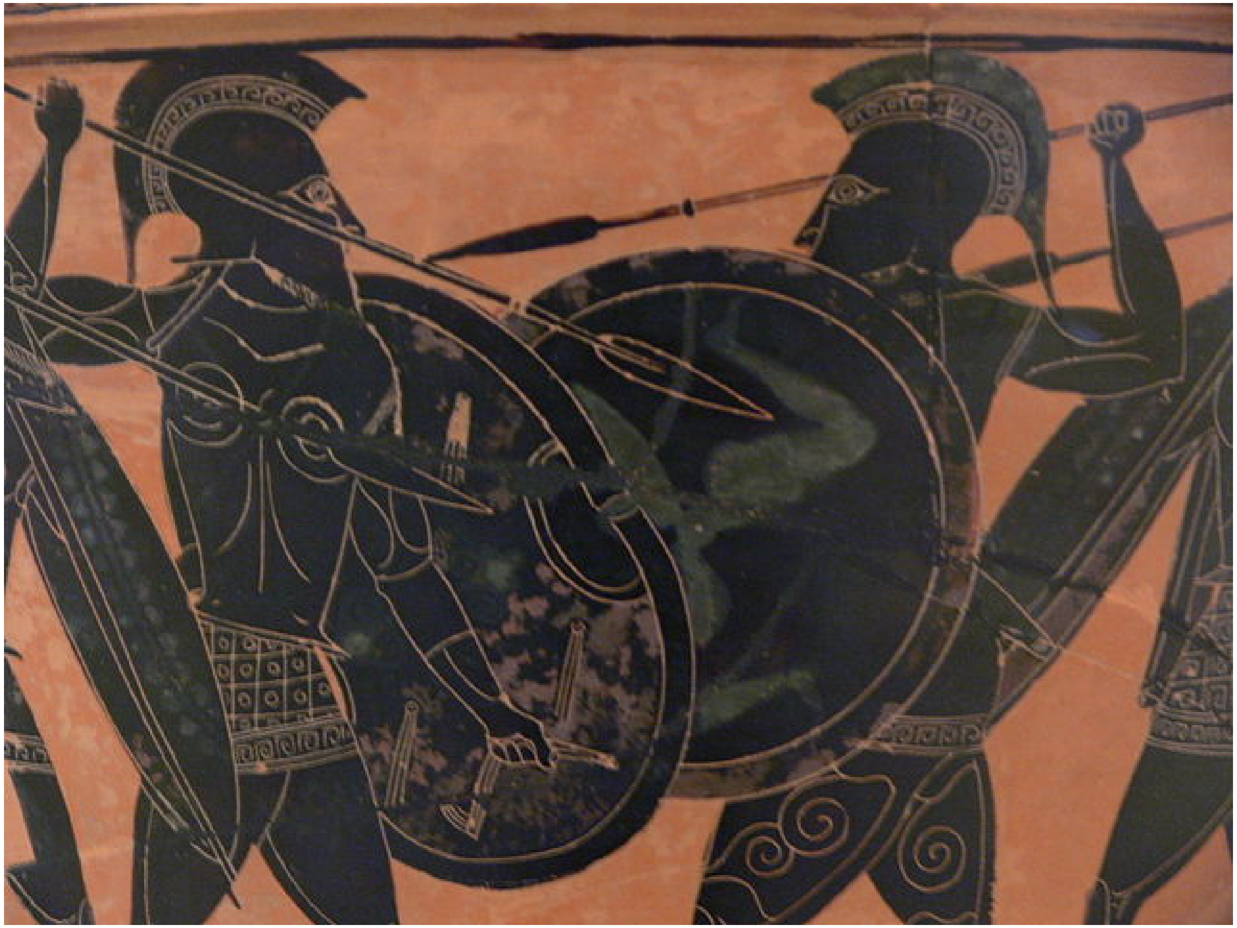
Notes on embedded files and JPEG 2000

Johan van der Knijff
Koninklijke Bibliotheek – National Library of the Netherlands

DPC, PDF/A-3 Briefing, Leeds, 13.3.2013

Part 1: Embedded files

PDF/A-3: embedding of *any* file (type)



Key point:

Use of “embedded files” really means “embedded file streams” = specific data structure in PDF!

File specification dictionary

```
31 0 obj
<</Type /Filespec /F (mysvg.svg) /EF <</F 32 0 R>> >>
endobj
```

File specification dictionary

```
31 0 obj
<</Type /Filespec /F (mysvg.svg) /EF <</F 32 0 R>> >>
endobj
```

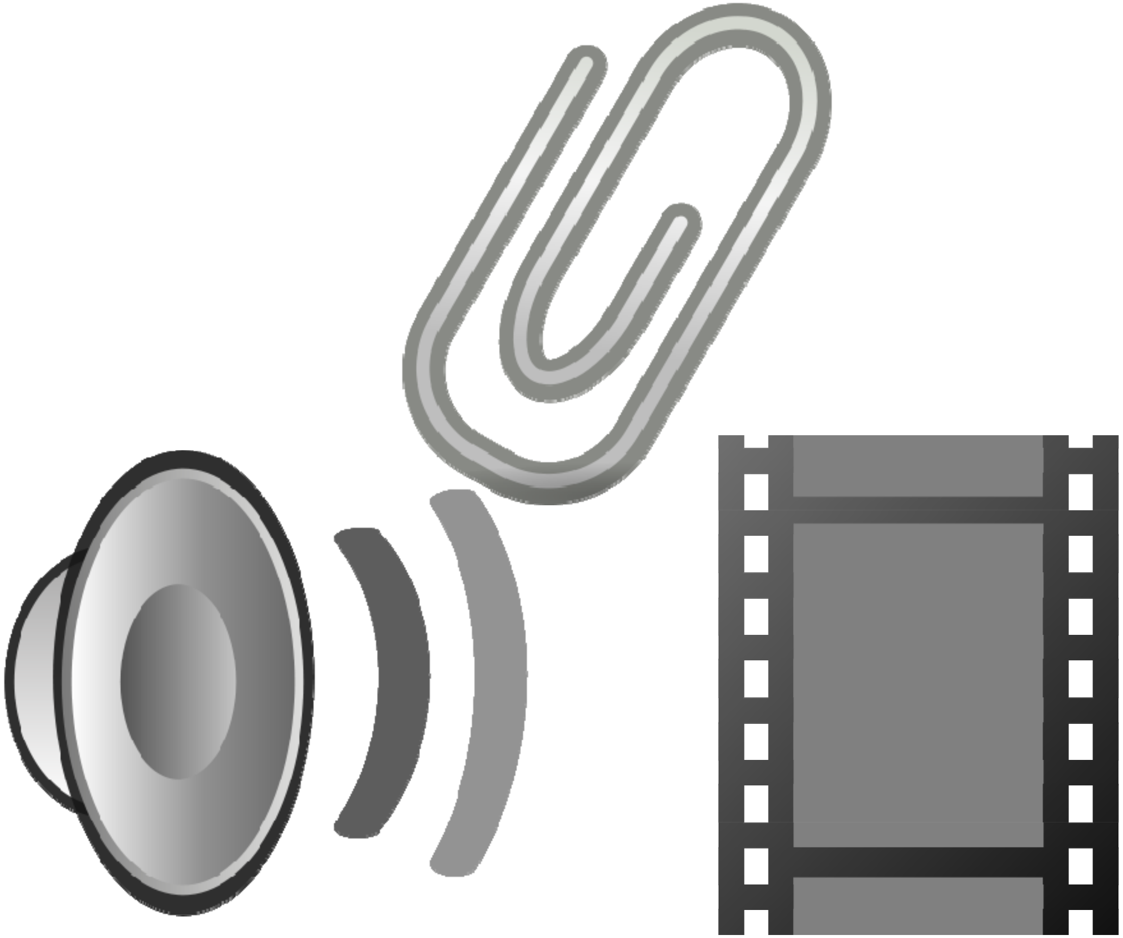
EF key

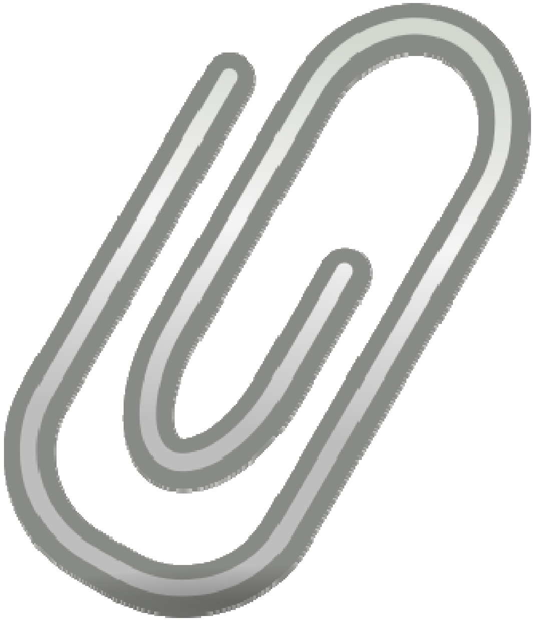
points to embedded file stream

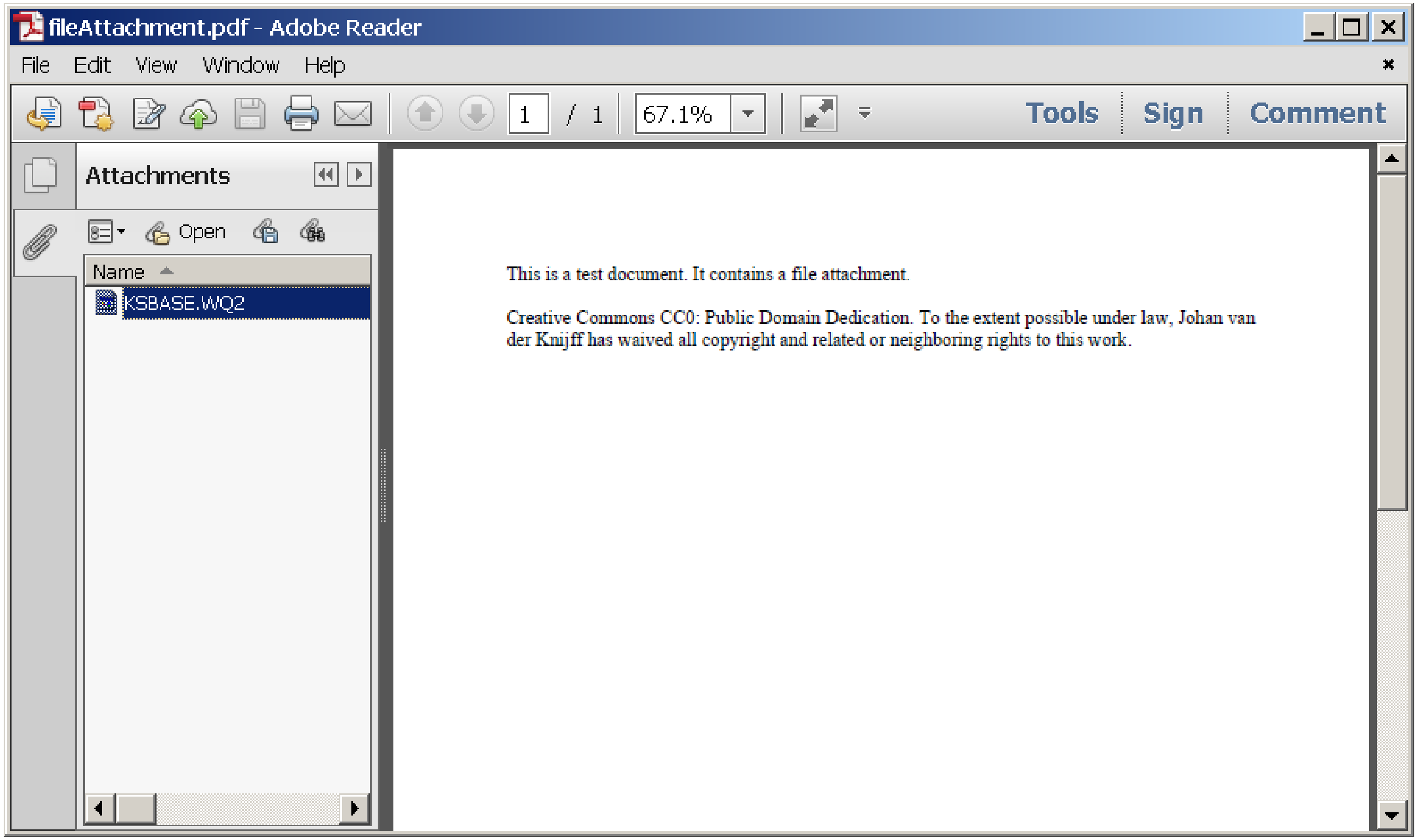
Embedded file stream

```
32 0 obj
<</Type /EmbeddedFile /Subtype /image#2Fsvg+xml /Length 72>>
stream
…SVG Data…
endstream
endobj
```

Uses of embedded file streams







fileAttachment.pdf - Adobe Reader
File
Edit
View
Window
Help
1
/ 1
67.1%
Tools
Sign
Comment
Attachments
Open
Name
KSBASE.WQ2

This is a test document. It contains a file attachment.

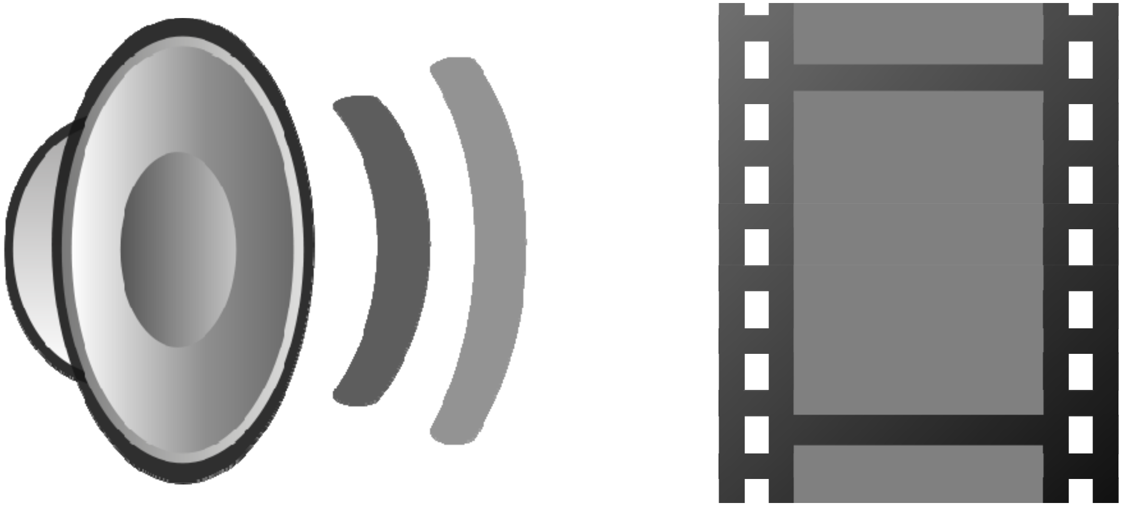
Creative Commons CC0: Public Domain Dedication. To the extent possible under law, Johan van der Knijff has waived all copyright and related or neighboring rights to this work.

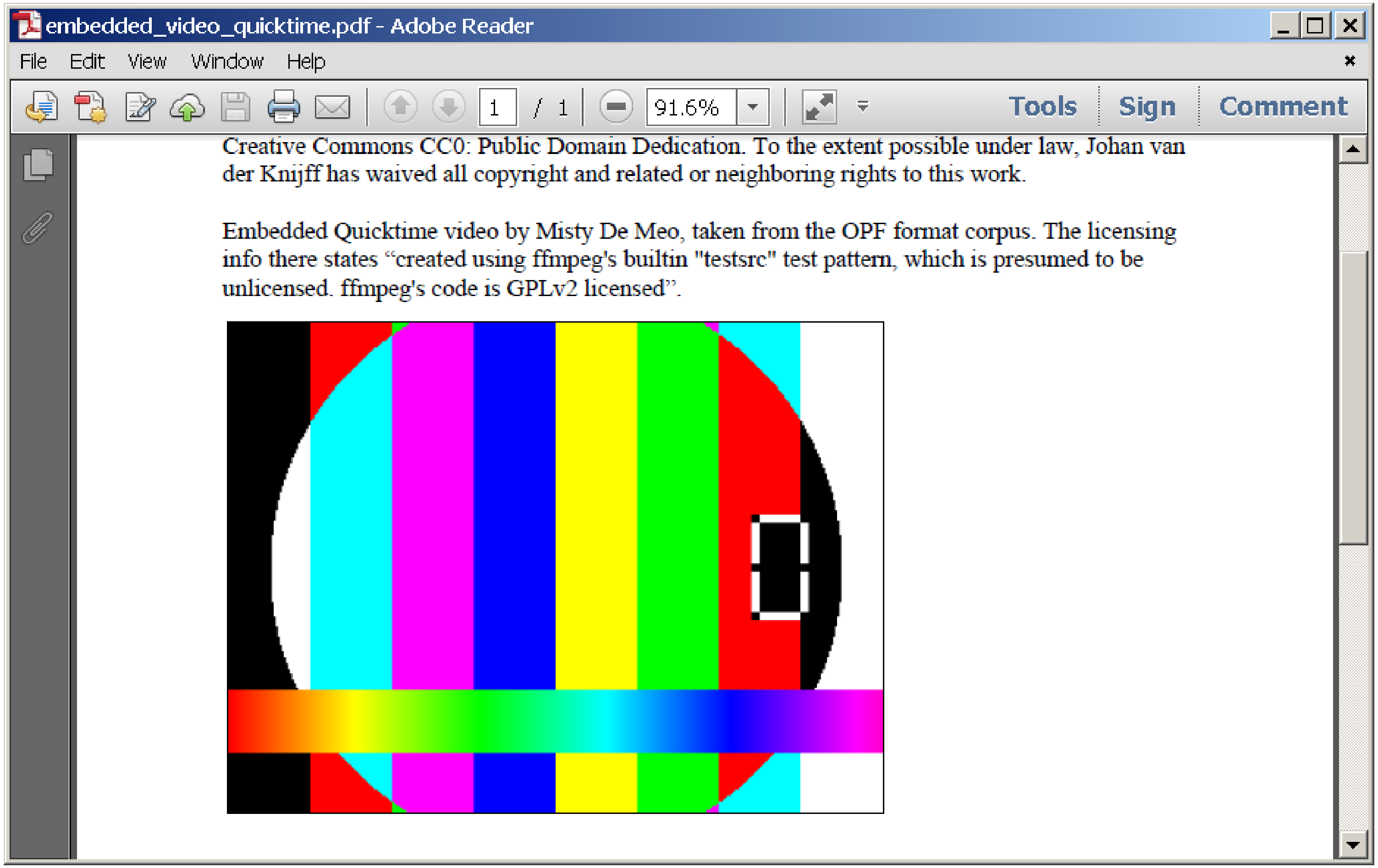
File attachments not meant to be rendered by viewer

PDF/A-3

File attachment annotation

EmbeddedFiles **entry in name dictionary**



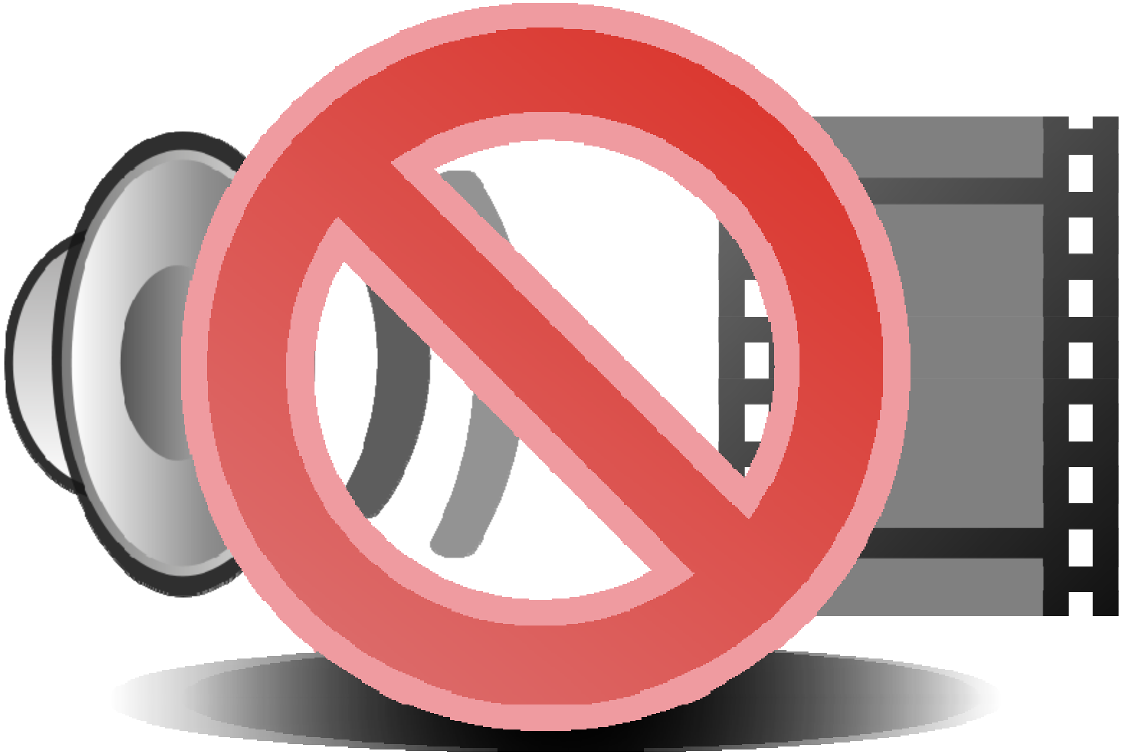

embedded_video_quicktime.pdf - Adobe Reader
File Edit View Window Help
1 / 1
91.6%
Tools
Sign
Comment

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 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”.

Rendered in/by PDF viewer

PDF/A-3



Rendition actions

Screen annotations

What about inline images?

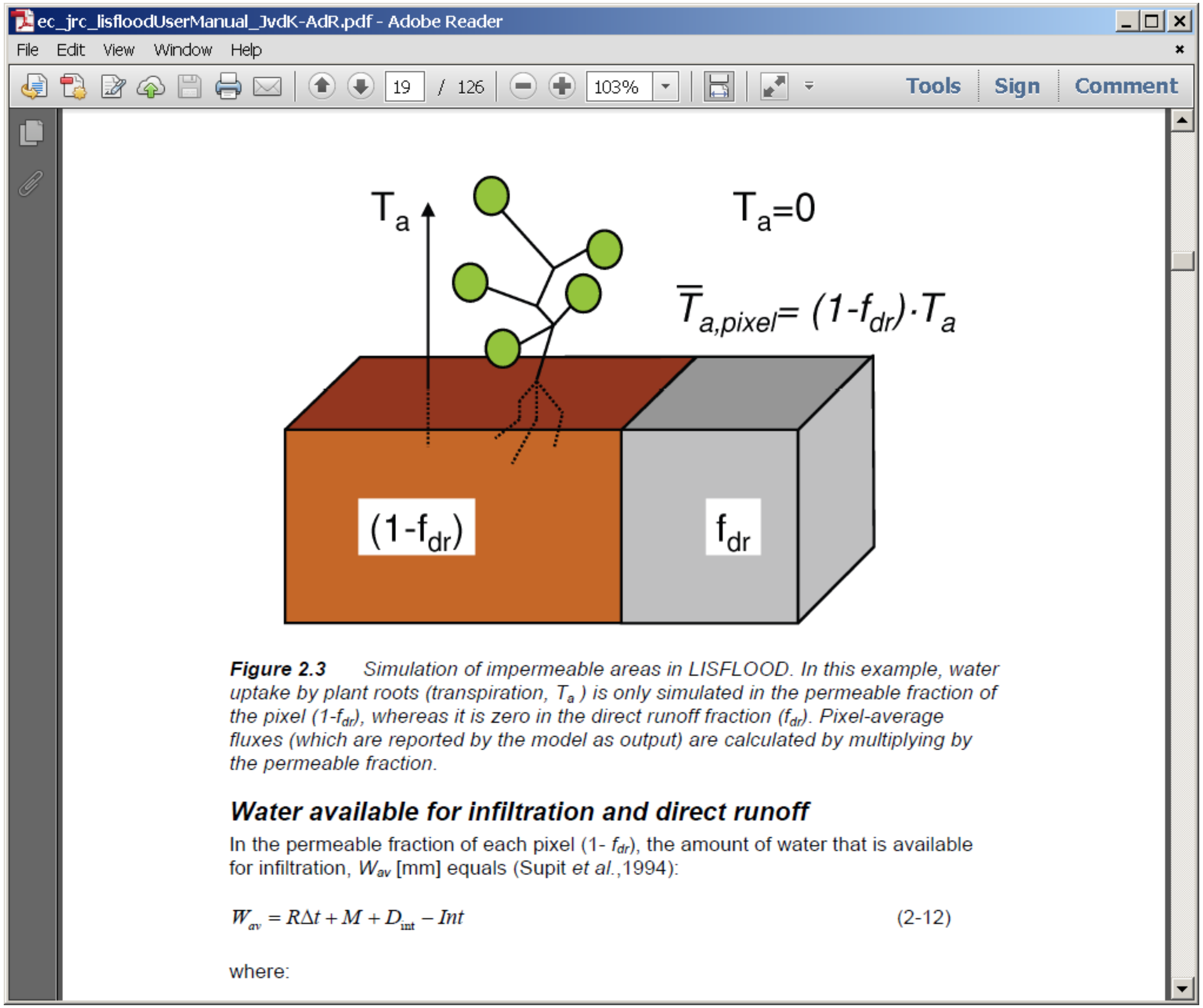

ec_jrc_lisfloodUserManual_JvdK-AdR.pdf - Adobe Reader
File
Edit
View
Window
Help
19
/ 126
103%
Tools
Sign
Comment
T_a
T_a=0
$\overline{T}_{a,pixel} = (1-f_{dr})\cdot T_a$
(1-f_{dr})
f_{dr}

Figure 2.3 *Simulation of impermeable areas in LISFLOOD. In this example, water uptake by plant roots (transpiration, T_a) is only simulated in the permeable fraction of the pixel (1-f_{dr}), whereas it is zero in the direct runoff fraction (f_{dr}). Pixel-average fluxes (which are reported by the model as output) are calculated by multiplying by the permeable fraction.*

Water available for infiltration and direct runoff

In the permeable fraction of each pixel (1- f_{dr}), the amount of water that is available for infiltration, W_{av} [mm] equals (Supit *et al.*,1994):

$$W_{av} = R\Delta t + M + D_{int} - Int \tag{2-12}$$

where:

What about inline images?

Not based on “embedded file stream”, but on “Image XObject” data structure (allows limited set of pre-defined formats)

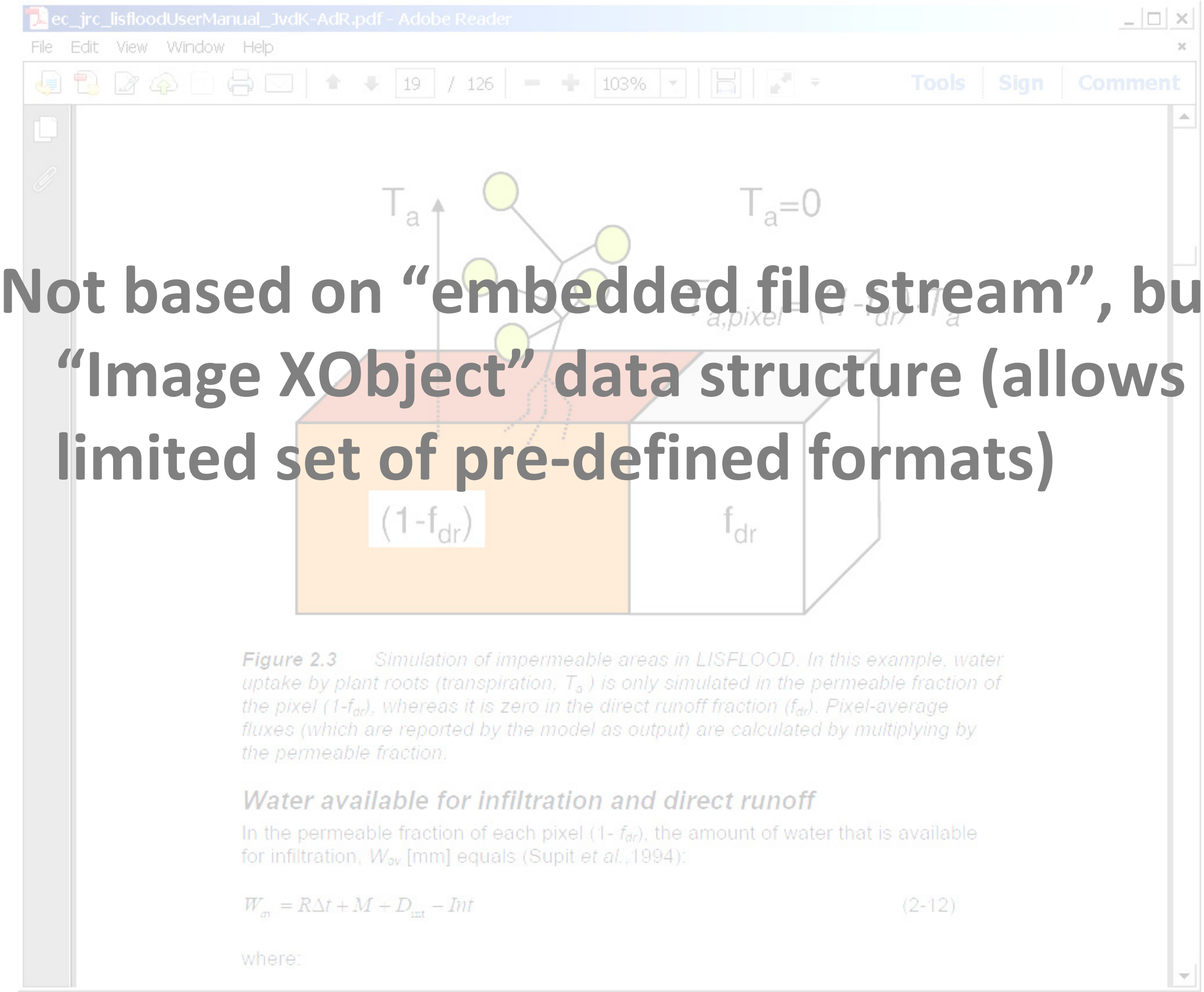

ec_jrc_lisfloodUserManual_JvdK-AdR.pdf - Adobe Reader
File Edit View Window Help
19 / 126
103%
Tools
Sign
Comment
T_a
$T_a=0$
$T_{a,pixel} = (1-f_{dr}) T_a$
$(1-f_{dr})$
f_{dr}

Figure 2.3 Simulation of impermeable areas in LISFLOOD. In this example, water uptake by plant roots (transpiration, T_a) is only simulated in the permeable fraction of the pixel ($1-f_{dr}$), whereas it is zero in the direct runoff fraction (f_{dr}). Pixel-average fluxes (which are reported by the model as output) are calculated by multiplying by the permeable fraction.

Water available for infiltration and direct runoff

In the permeable fraction of each pixel ($1- f_{dr}$), the amount of water that is available for infiltration, W_{av} [mm] equals (Supit *et al.*,1994):

$$W_{av} = R\Delta t + M + D_{int} - Int \qquad (2\text{-}12)$$

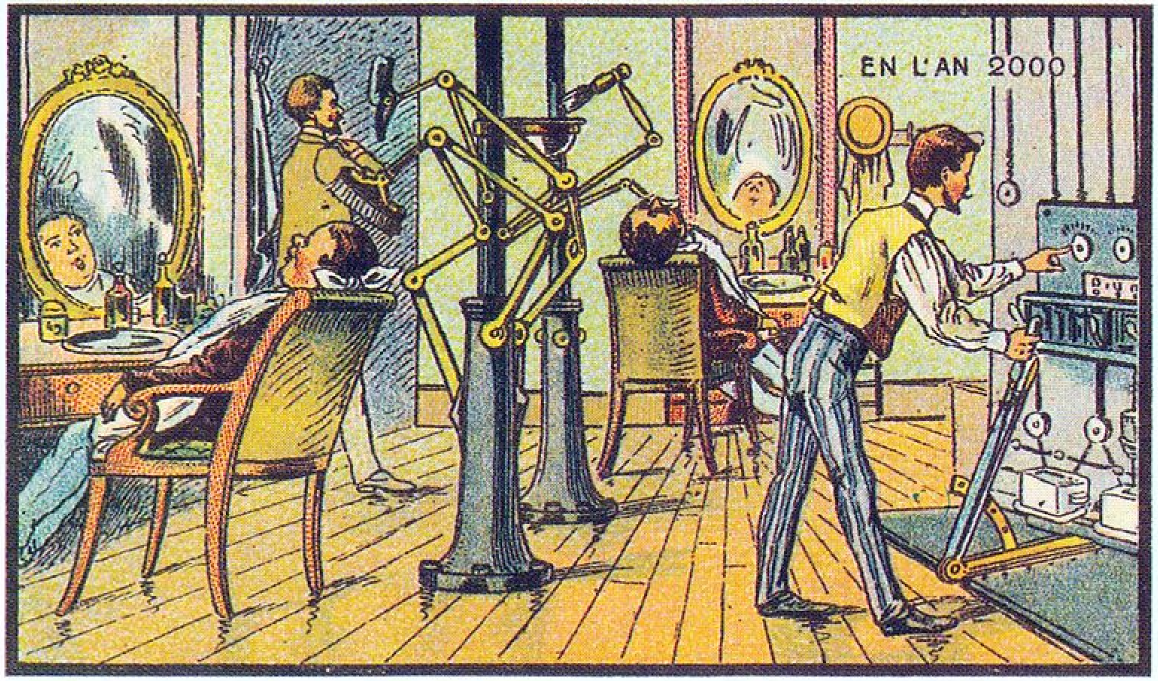
where:

Embedded files wrap-up:

No impact on content that is meant to be rendered by PDF viewer

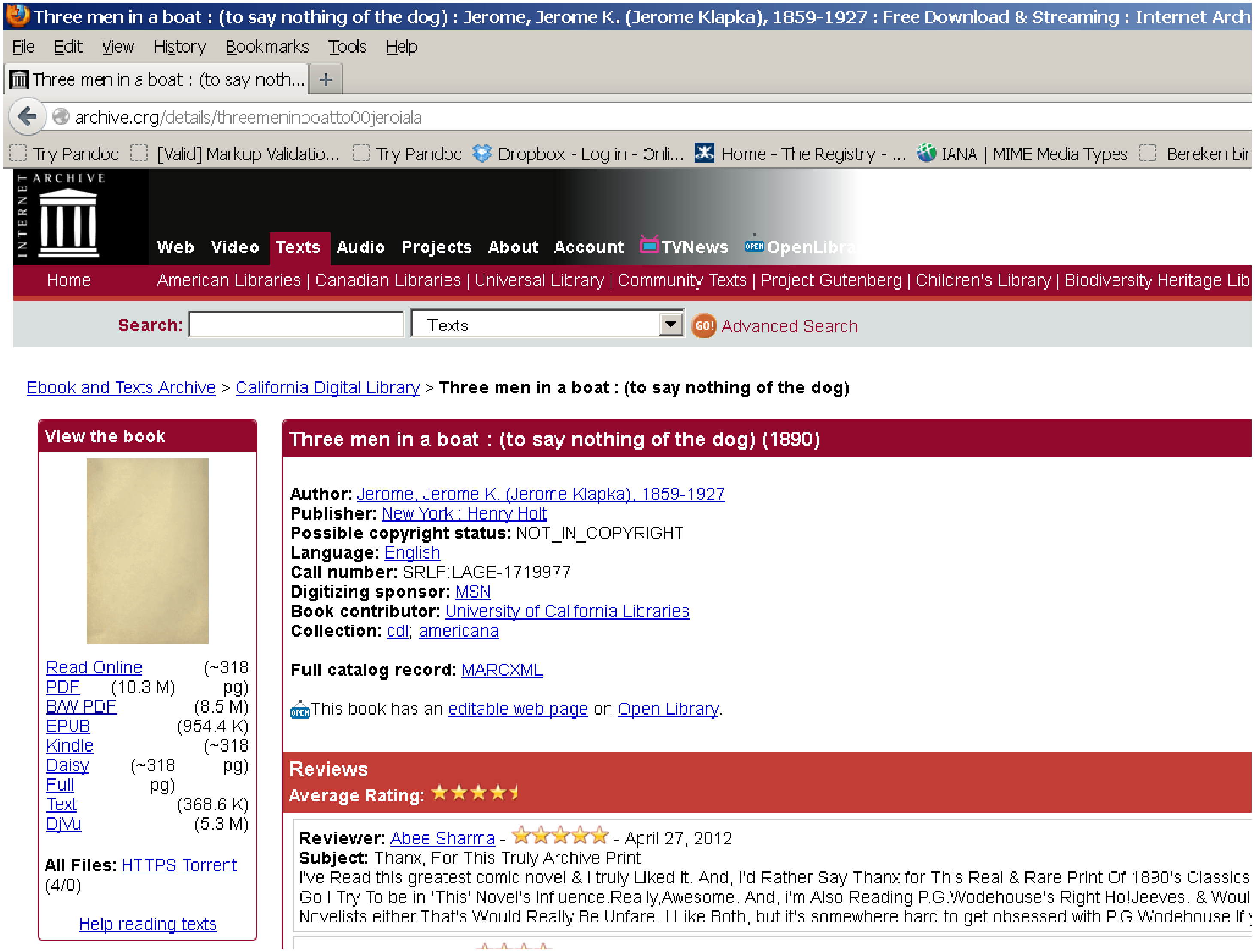
But PDF/A-3’s may contain file of *any* possible format as an attachment

Part 2: JPEG 2000



The New-Fangled Barber

Supported since PDF/A-2


Three men in a boat : (to say nothing of the dog) : Jerome, Jerome K. (Jerome Klapka), 1859-1927 : Free Download & Streaming : Internet Arch
File Edit View History Bookmarks Tools Help
Three men in a boat : (to say noth...
archive.org/details/threemeninboatto00jeroiala
Try Pandoc [Valid] Markup Validatio... Try Pandoc Dropbox - Log in - Onli... Home - The Registry - ... IANA | MIME Media Types Bereken bir
INTERNET ARCHIVE
Web Video Texts Audio Projects About Account TVNews OpenLibra
Home American Libraries | Canadian Libraries | Universal Library | Community Texts | Project Gutenberg | Children's Library | Biodiversity Heritage Lib
Search: Texts GO! Advanced Search

Ebook and Texts Archive > California Digital Library > **Three men in a boat : (to say nothing of the dog)**

View the book

- Read Online (~318 pg)
- PDF (10.3 M)
- B/W PDF (8.5 M)
- EPUB (954.4 K)
- Kindle (~318 pg)
- Daisy (~318 pg)
- Full Text (368.6 K)
- DjVu (5.3 M)

All Files: HTTPS Torrent (4/0)

Help reading texts

Three men in a boat : (to say nothing of the dog) (1890)

Author: Jerome, Jerome K. (Jerome Klapka), 1859-1927
Publisher: New York : Henry Holt
Possible copyright status: NOT_IN_COPYRIGHT
Language: English
Call number: SRLF:LAGE-1719977
Digitizing sponsor: MSN
Book contributor: University of California Libraries
Collection: cdl; americana

Full catalog record: MARCXML

This book has an editable web page on Open Library.

Reviews
Average Rating: ★★★★½

Reviewer: Abee Sharma - ★★★★★ - April 27, 2012
Subject: Thanx, For This Truly Archive Print.
I've Read this greatest comic novel & I truly Liked it. And, I'd Rather Say Thanx for This Real & Rare Print Of 1890's Classics
Go I Try To be in 'This' Novel's Influence.Really,Awesome. And, i'm Also Reading P.G.Wodehouse's Right Ho!Jeeves. & Woul
Novelists either.That's Would Really Be Unfare. I Like Both, but it's somewhere hard to get obsessed with P.G.Wodehouse If

Image XObject

```
1614 0 obj
<</Subtype/Image/Width 615/Height 978/ColorSpace/DeviceRGB
/BitsPerComponent 8/Interpolate true/Length 5278
/Filter/JPXDecode>>
stream
… Image data …
: :
: :
endstream
endobj
```

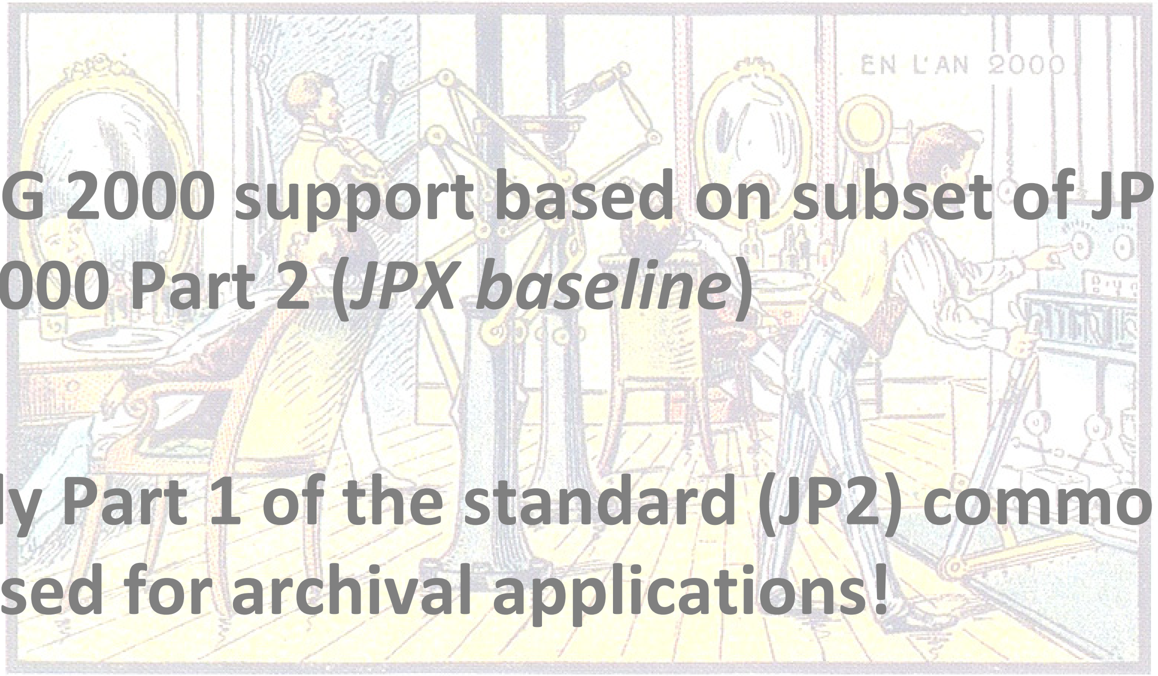
Image XObject

```
1614 0 obj
<</Subtype/Image/Width 615/Height 978/ColorSpace/DeviceRGB
/BitsPerComponent 8/Interpolate true/Length 5278
/Filter/JPXDecode>>
stream
… Image data …
::
::
endstream
endobj
```

Identifies object as

JPEG 2000 image

ISO 19005-2 (PDF/A-2):



JPEG 2000 support based on subset of JPEG 2000 Part 2 (*JPX baseline*)

Only Part 1 of the standard (JP2) commonly used for archival applications!

The New-Fangled Barber

JP2 vs JPX

JP2

JPEG 2000 Part 1:

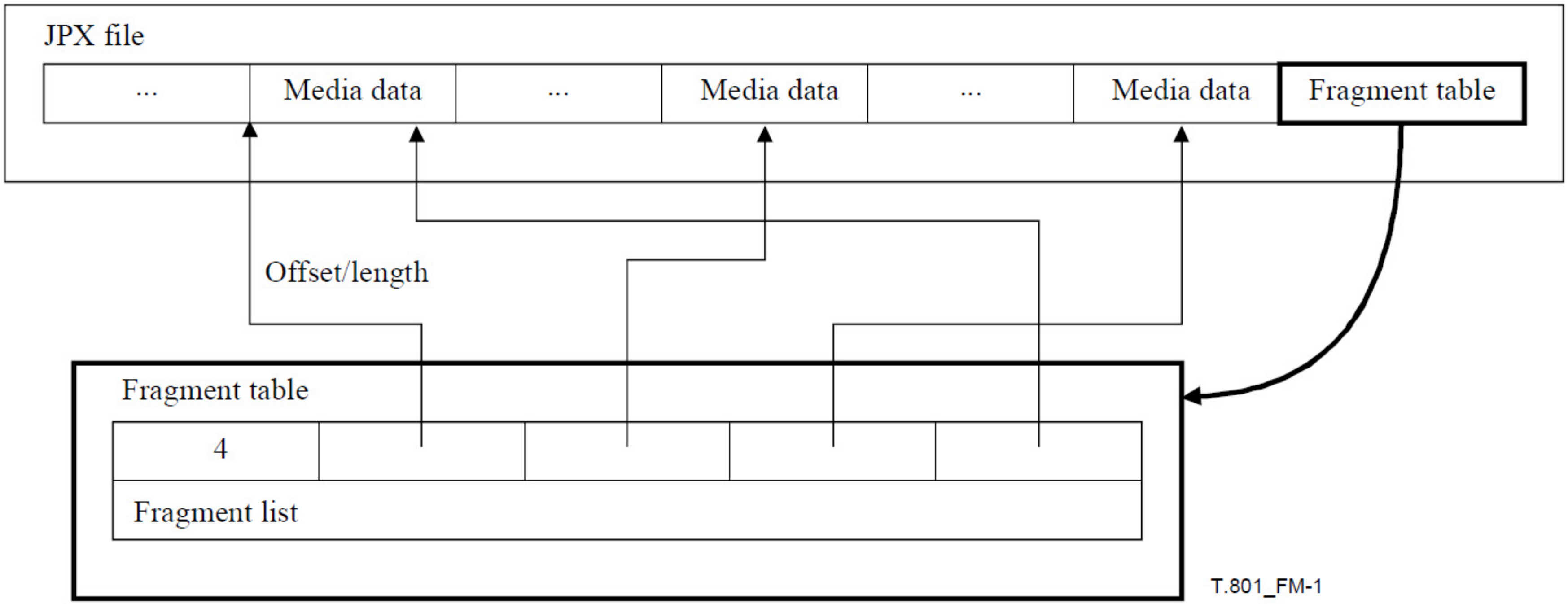
Basic still image format

JPX

JPEG 2000 Part 2:

= JP2 + assorted
advanced stuff …

Fragmented codestreams


JPX file
...
Media data
...
Media data
...
Media data
Fragment table
Offset/length
Fragment table
4
Fragment list
T.801_FM-1

Allowed in JPX Baseline!

OS PDF viewers – JPEG 2000 libraries

Ghostscript: OpenJPEG or JasPer

Evince: OpenJPEG

Mupdf: OpenJPEG

Firefox PDF viewer: built-in decoder

→ None of these libraries support fragmented codestreams!

Is it really a problem?

Fragmented codestreams extremely rare

But why is this feature even allowed in a long-term archival format?

OS support of JPEG 2000 in general remains problematic

Funding

This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

http://www.scape-project.eu

twitter **#SCAPEProject**