



PACKED

Centre of Expertise in Digital Heritage



DIGITAL PRESERVATION: THREATS AND STRATEGIES

Bert Lemmens | PACKED vzw
29 October 2015 | iMAL Brussels



PACKED

Centre of Expertise in Digital Heritage

PACKED VZW



CULTUREEL
ERFGOED
STANDAARDEN
TOOLBOX



Toolbox & Richtlijnen voor
Archief- & Collectiezorg
in de Kunstensector



WEBSITE OVER
DE PRESERVERING VAN AUDIOVISUEEL ERFGOED
EEN INITIATIEF VAN PACKED

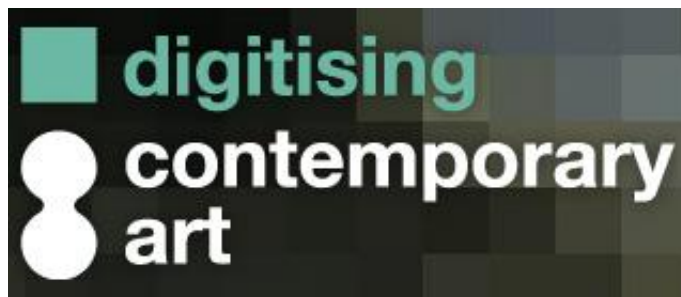
www.packed.be
www.projectcest.be
www.scart.be
www.projecttracks.be
www.scoremodel.org



PACKED

Centre of Expertise in Digital Heritage

PACKED VZW



CONTENTS

- digital preservation
- threats
- strategies



I. INTRO

- Introduce yourself !
- What type of 'digital objects' do you produce?
- How would you define 'digital preservation'? (in one sentence)



ANALOGUE PRESERVATION?



PACKED

Centre of Expertise in Digital Heritage

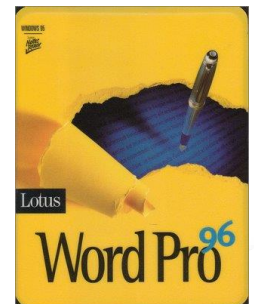
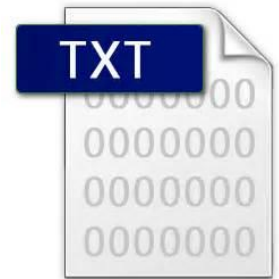
DIGITAL PRESERVATION?



PACKED

Centre of Expertise in Digital Heritage

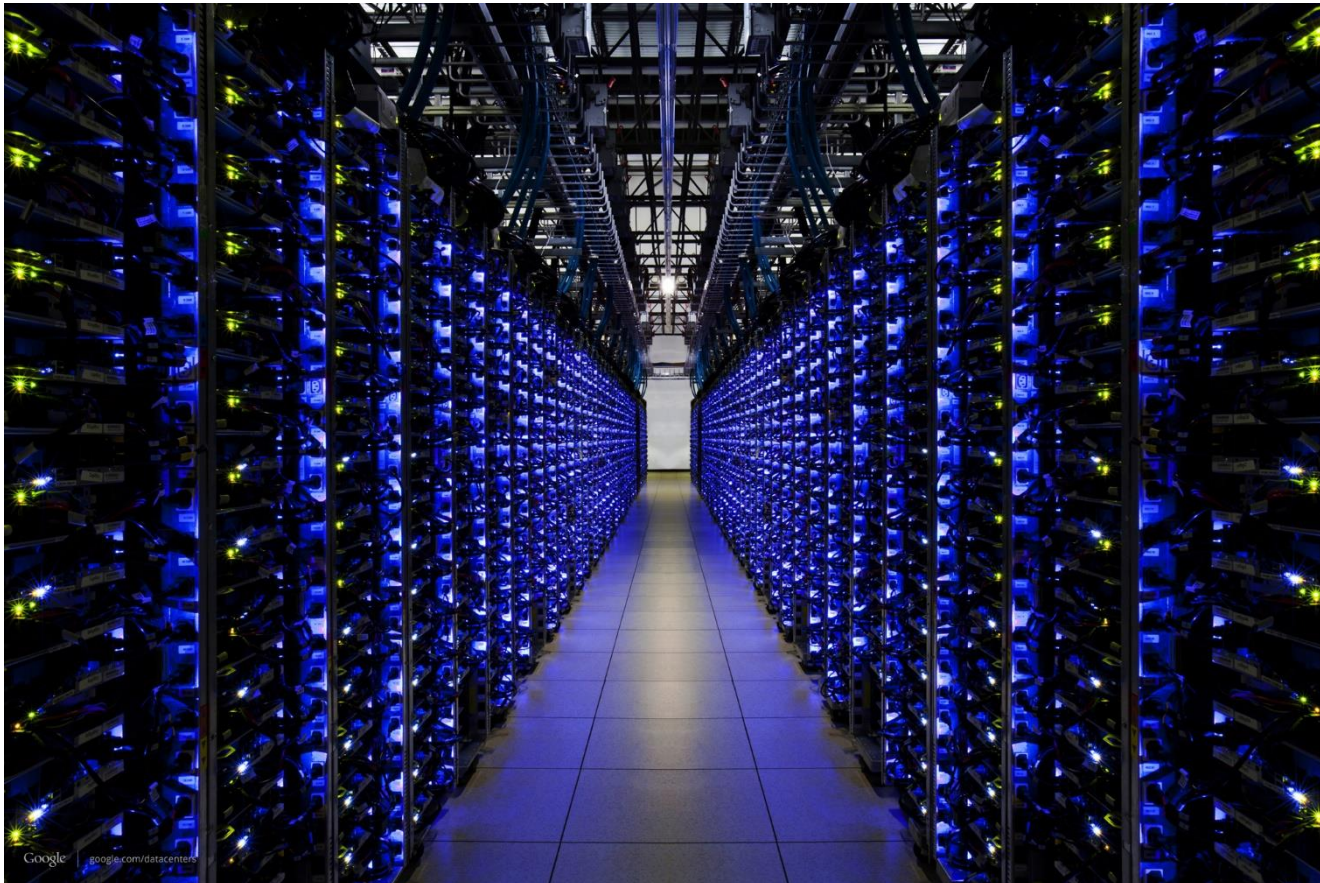
DIGITAL PRESERVATION?



PACKED

Centre of Expertise in Digital Heritage

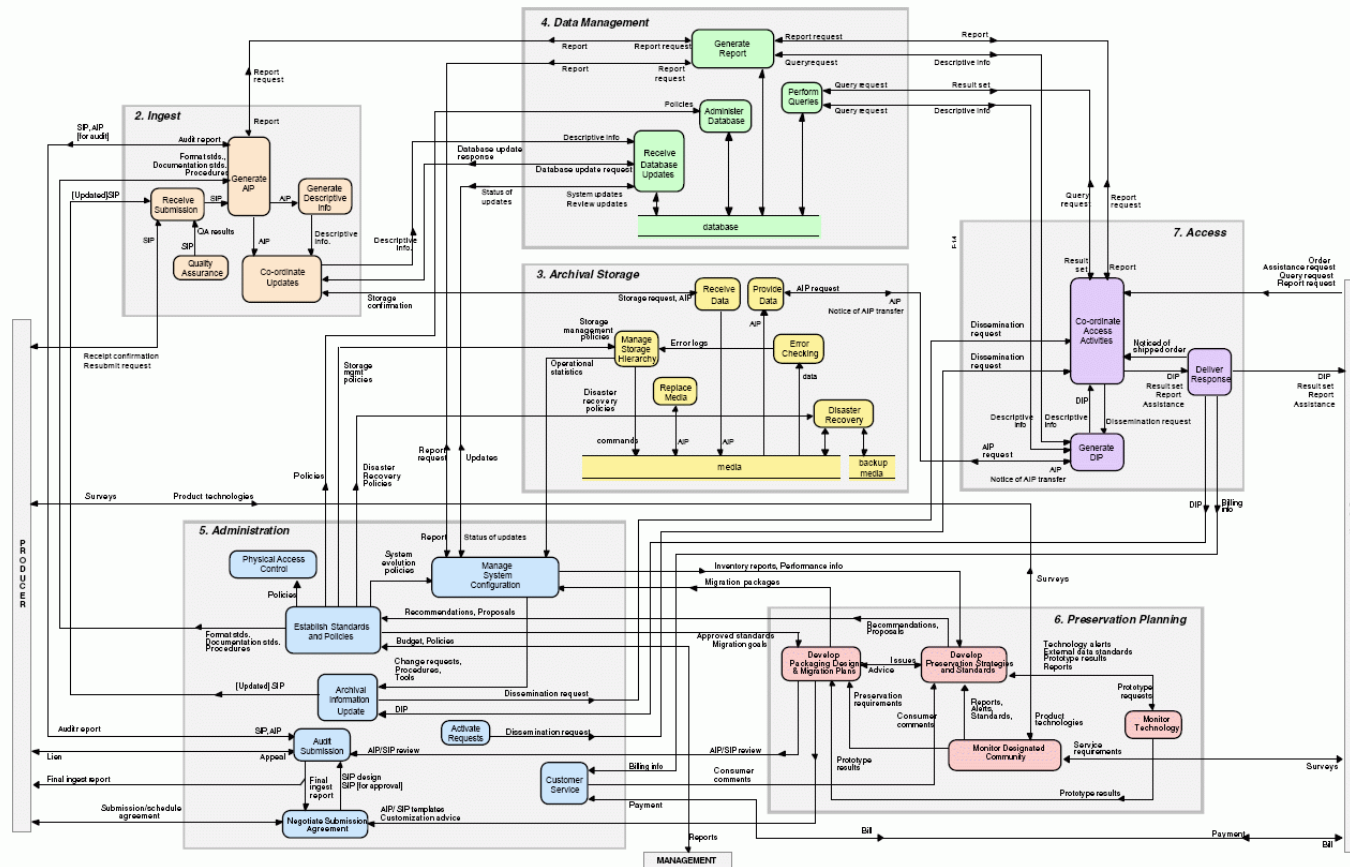
DIGITAL REPOSITORIES



PACKED

Centre of Expertise in Digital Heritage

DIGITAL REPOSITORIES



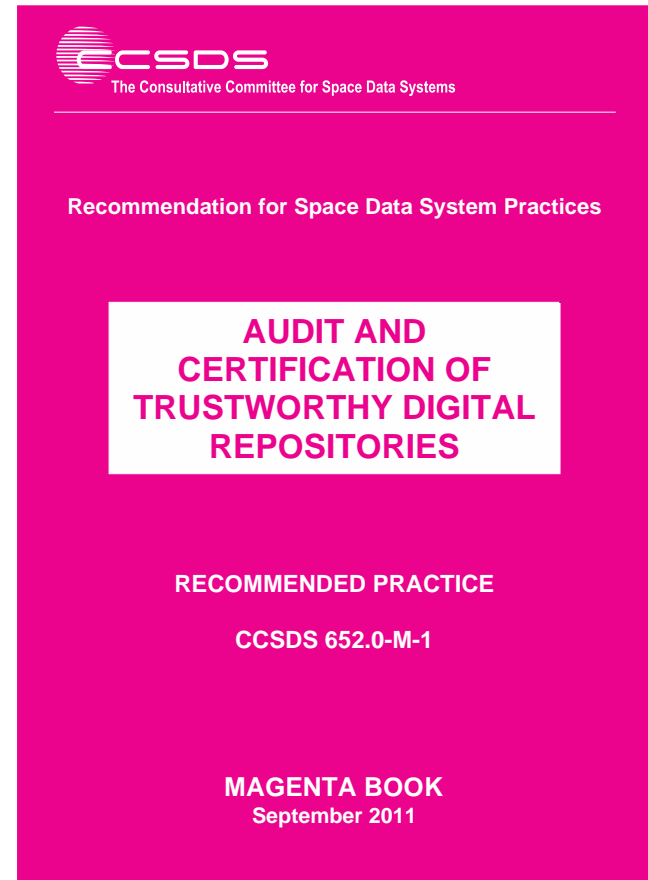
DIGITAL REPOSITORIES

OAIS MAGENTA BOOK

- ISO 14721:2003
- 77 pagina's

4.1.1 The repository shall identify the Content Information and the Information Properties that the repository will preserve.

Supporting Text. This is necessary in order to make it clear to funders, depositors, and users what responsibilities the repository is taking on and what aspects are excluded. It is also a necessary step in defining the information which is needed from the information producers or depositors.



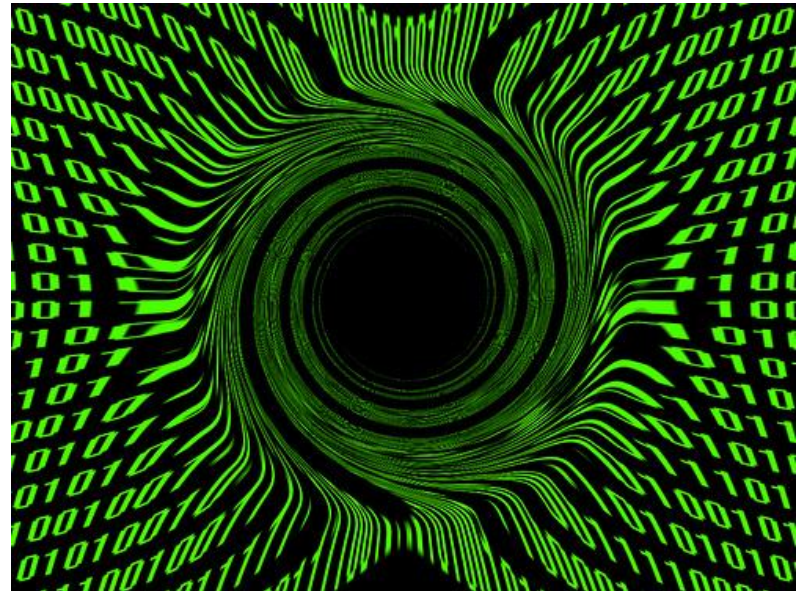
PACKED

Centre of Expertise in Digital Heritage

DIGITAL REPOSITORIES (IN PRACTICE)

- 'bit level preservation
- normalisation > only for specific file formats

=> 'digital black hole

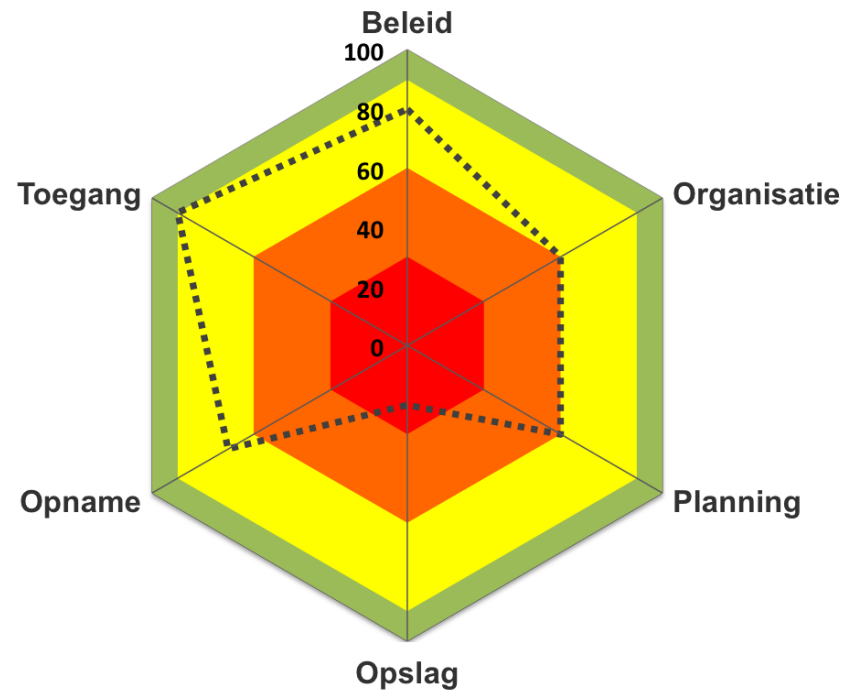


ALTERNATIVE APPROACH: MATURITY MODEL (CH. DOLLAR)

SEVEN ASPECTS:

- Policy
- Strategy
- Expertise & organisation
- Storage
- Planning & controle
- Ingest
- Access

★ SCOREMODEL



PACKED

Centre of Expertise in Digital Heritage

ACT PROACTIVELY!

digital preservation

=

intervene in the environment where you
create and store documents
to reduce the risk of damage to a minimum



PACKED

Centre of Expertise in Digital Heritage

ACT PROACTIVELY!

digital preservation

=

identify threats



apply a strategy to counter it



PACKED

Centre of Expertise in Digital Heritage

ACT PROACTIVELY!

digital preservation

=

technical solutions + proper arrangements
clever tools + 'getting things organized'



PACKED

Centre of Expertise in Digital Heritage

SUMMARY

- analogue vs digital preservation
- digital repositories
- act proactively!





PACKED

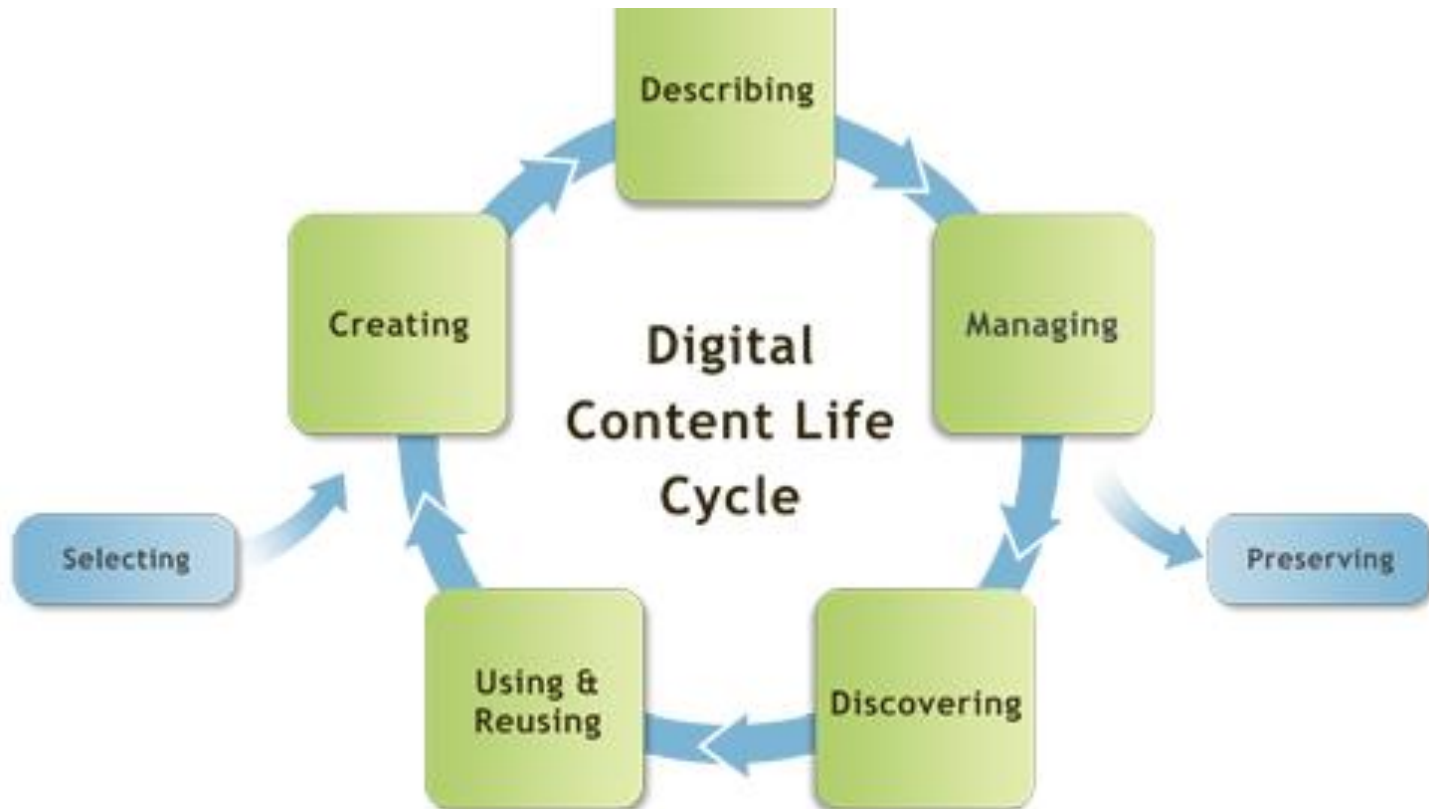
Expertisecentrum Digitaal Erfgoed

II. THREATS

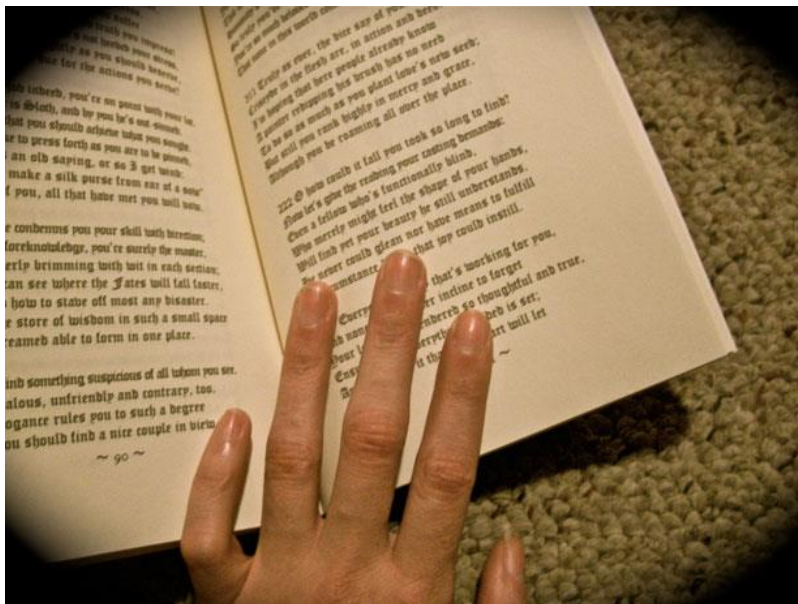
- obsolete technology
- unreliable carriers
- rights infringement
- managing extent

=> What applies to your content?

DIGITALE LEVENSCYCLUS



#1 OBSOLETE TECHNOLOGY



PACKED

Centre of Expertise in Digital Heritage

#1 OBSOLETE TECHNOLOGY



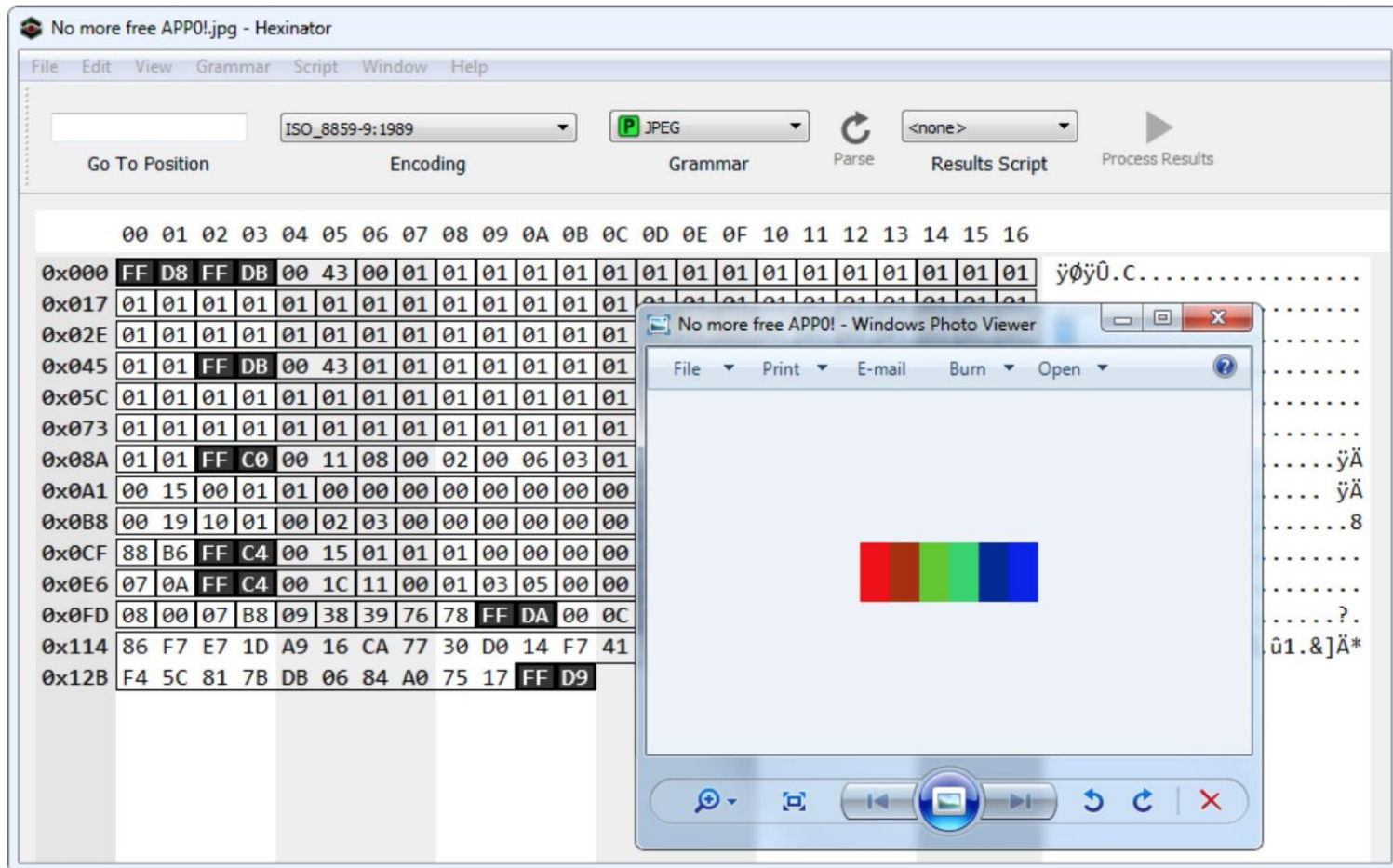
#1 OBSOLETE TECHNOLOGY



PACKED

Centre of Expertise in Digital Heritage

#1 OBSOLETE TECHNOLOGY



#1 OBSOLETE TECHNOLOGY

- Access depends on the availability of the corresponding technology
- Problems:
 - Algorithm (codec) to decode your content is lost
 - Software to decode your content is lost
 - Device to decode your content is lost



#2 UNRELIABLE CARRIERS



PACKED

Centre of Expertise in Digital Heritage

#2 UNRELIABLE CARRIERS



PACKED

Centre of Expertise in Digital Heritage

#2 UNRELIABLE CARRIERS

- Problem
 - Inherent physical deterioration (bitrot)
 - Physical damage by simple usage
 - Errors when (de-)coding (e.g. copying)



#3 RIGHTS INFRINGEMENT



Office



OS X



PACKED

Centre of Expertise in Digital Heritage

#3 RIGHTS INFRINGEMENT



Linux



PACKED

Centre of Expertise in Digital Heritage

#3 RIGHTS INFRINGEMENT



#3 RIGHTS INFRINGEMENT

Problems:

- intellectual property rights, patents, ... on:
- Format (wrapper)
- Codec (essence)
- Software (implementation)
- hardware (carrier, hardware codec)



PACKED

Centre of Expertise in Digital Heritage

#4. MANAGING EXTENT



PACKED

Centre of Expertise in Digital Heritage

#4. MANAGING EXTENT



PACKED

Centre of Expertise in Digital Heritage

#4. EXTENT & MANAGEMENT

- Problem:
 - Lacking metadata
 - endless copying
 - Ignorance
 - project-based work





PACKED

Expertisecentrum Digitaal Erfgoed

II. THREATS

- obsolete technology
- unreliable carriers
- rights infringement
- managing extent

=> What applies to your content?

III. STRATEGIES

- do nothing...
- conservation
- documentation
- copy & distribute
- regular checks
- migration & transcoding
- emulation

=> What strategy would counter your threats?

#1 DO NOTHING



PACKED

Centre of Expertise in Digital Heritage

#1 DO NOTHING

- because you don't fully understand the preservation threat/solution
- because you don't find a convenient solution

pro:

- avoid obvious mistakes

con:

- obsolescence > inevitable
- ignorance



PACKED

Centre of Expertise in Digital Heritage

#2 CONSERVATION



#2 CONSERVATION

- Store hardware, software and carriers in a safe, climatised environment

pro:

- relevant when the 'essence' is in both the digital and physical manifestation of the object

con:

- requires maintenance (e.g. batteries!)
- hardware obsolescence



PACKED

Centre of Expertise in Digital Heritage

#3 DOCUMENTATION



#3 DOCUMENTATION

- user manuals of hardware and software
- technical specifications of hardware, software and file formats
- documentation of the system environment (software libraries, programming languages, OS)

pro:

- helpful for reverse engineering/emulation

Con:

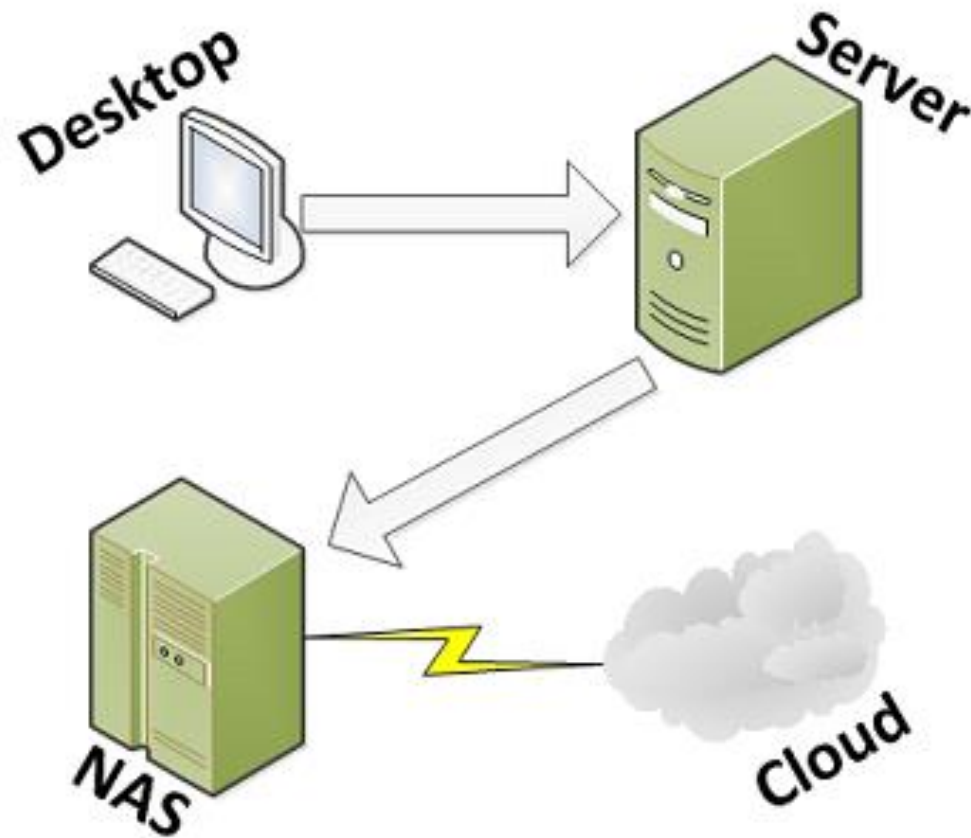
- Passive: depend on expertise somewhere in the future



PACKED

Centre of Expertise in Digital Heritage

#4 COPY & DISTRIBUTE



#4 COPY & DISTRIBUTE

- ≠ copy types: archive file vs. reproduction file vs. access file
- back-up strategie: full/incremental - frequency - locations

pro:

- risk distribution

con:

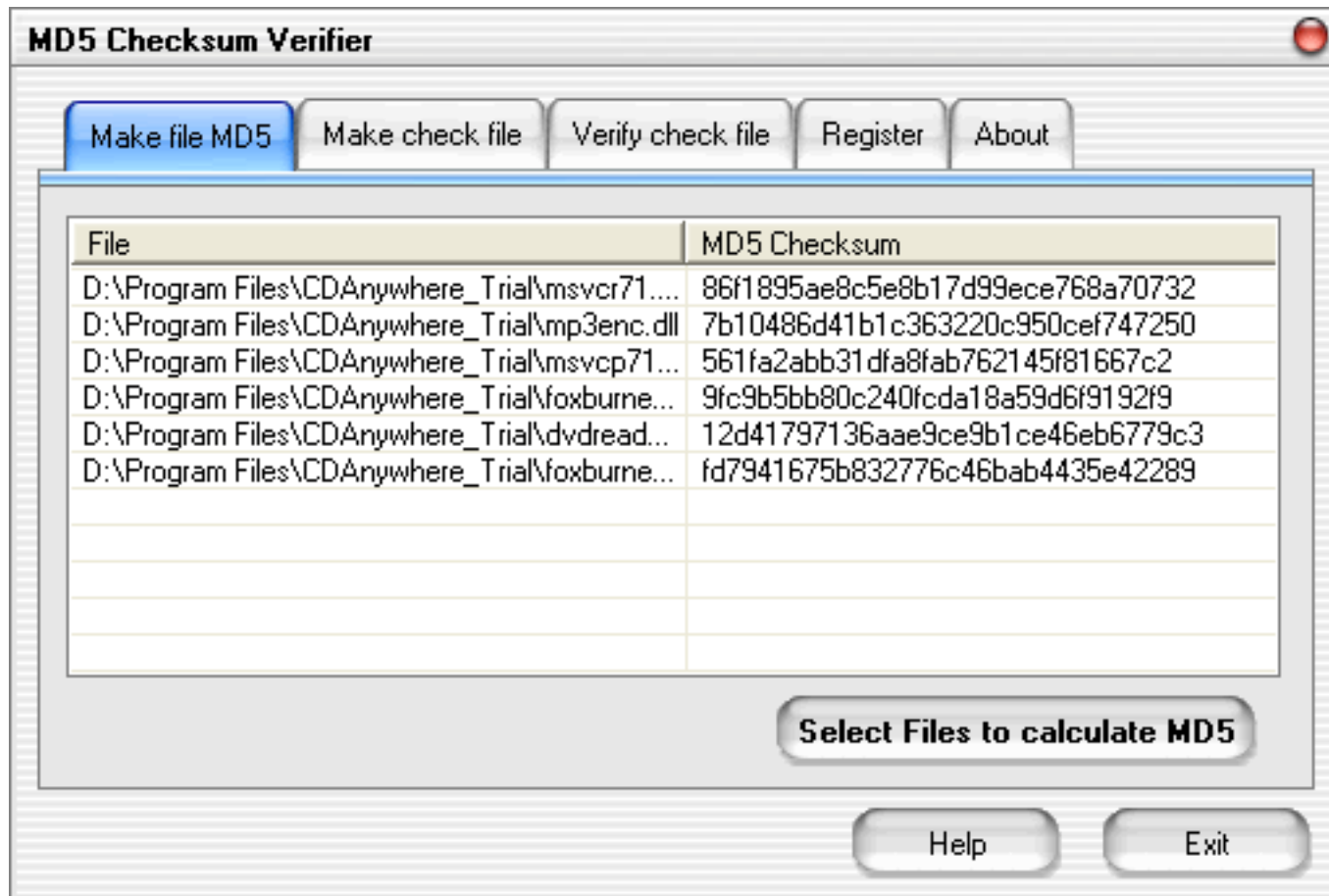
- risk of loosing track of copies



PACKED

Centre of Expertise in Digital Heritage

#5 REGULAR CHECKS



#5 REGULAR CHECKS

- operational hardware en software
- virus control
- completeness of your archive
- integrity of your archive

Pro:

- identify preservation issues at an early stage
- something you can automate

● Con:

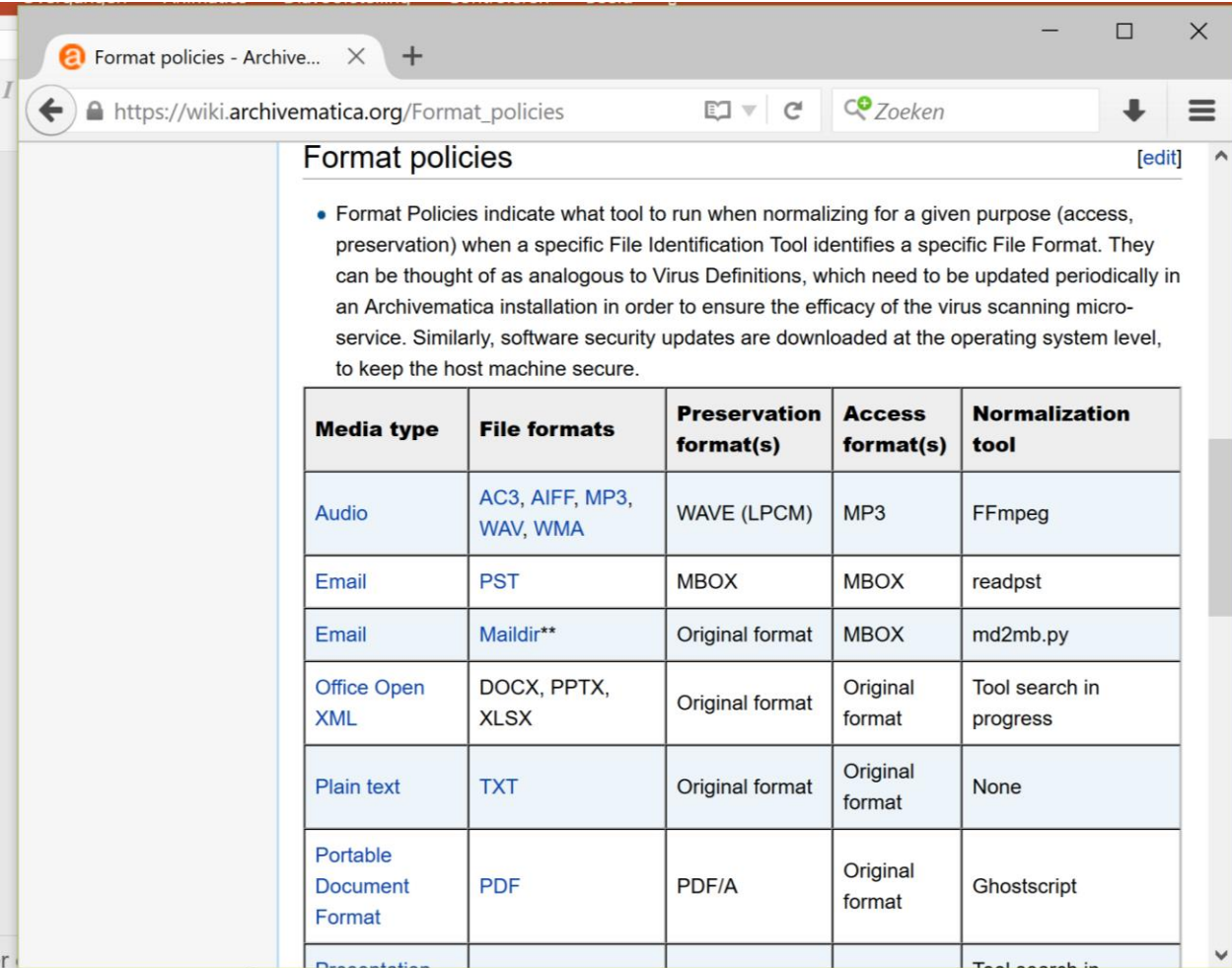
- allocate responsibilities
- discipline! Considerable IT-expertise!



PACKED

Centre of Expertise in Digital Heritage

#6 MIGRATION & TRANSCODING



The screenshot shows a web browser window with the address bar displaying https://wiki.archivematica.org/Format_policies. The page title is "Format policies" with an "[edit]" link. Below the title, a bullet point explains that Format Policies indicate the tool to run for normalization based on the File Identification Tool's output. A table follows, detailing the normalization tools for various media types and file formats.

Media type	File formats	Preservation format(s)	Access format(s)	Normalization tool
Audio	AC3, AIFF, MP3, WAV, WMA	WAVE (LPCM)	MP3	FFmpeg
Email	PST	MBOX	MBOX	readpst
Email	Maildir**	Original format	MBOX	md2mb.py
Office Open XML	DOCX, PPTX, XLSX	Original format	Original format	Tool search in progress
Plain text	TXT	Original format	Original format	None
Portable Document Format	PDF	PDF/A	Original format	Ghostscript
Presentation				Tool search in progress

#6 MIGRATION & TRANSCODING



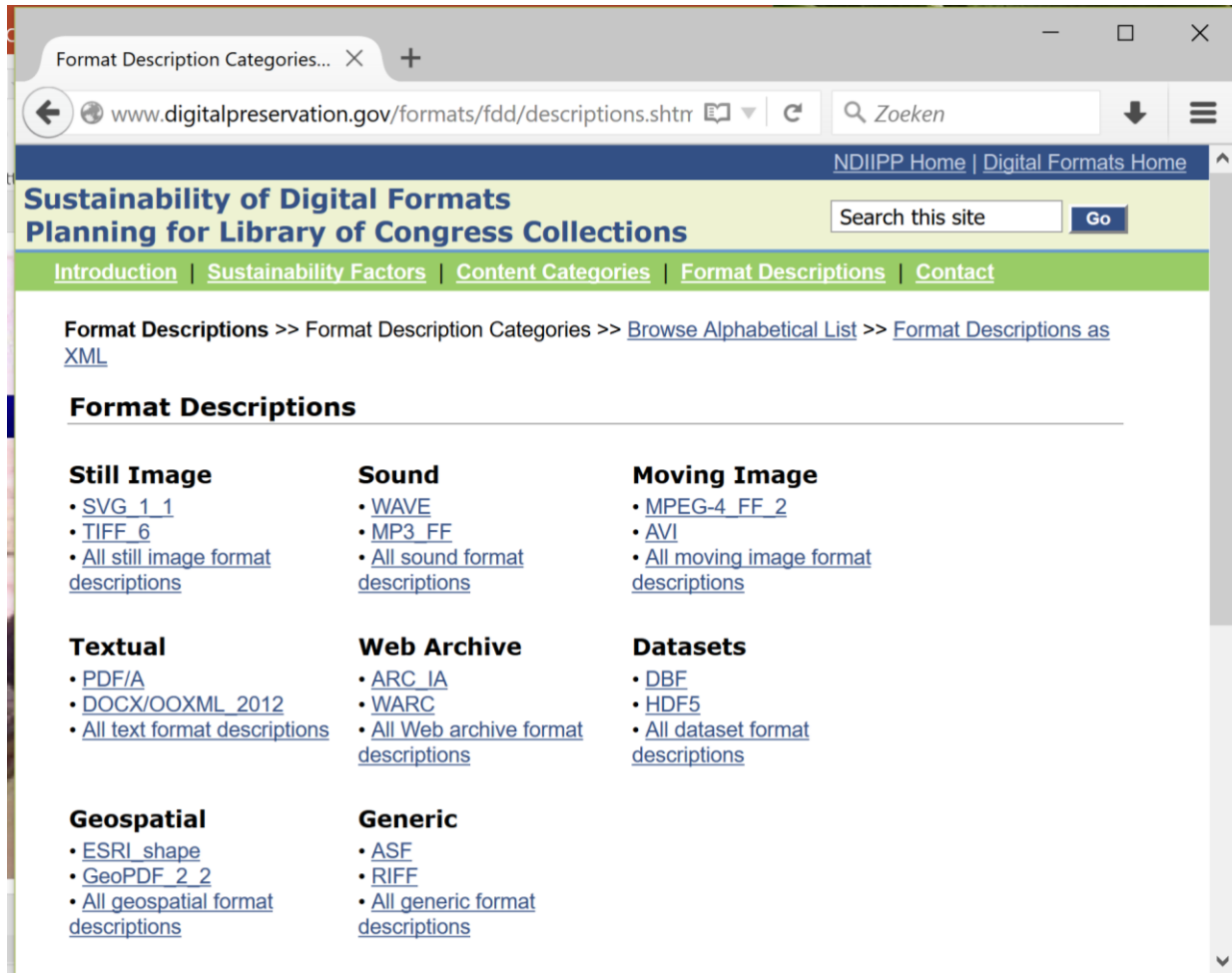
Preservation format	encoding	wrapper
TEKST	Utf-8	XML
IMAGE	TIFF v6.0 uncompressed baseline	TIFF v6.0 uncompressed baseline
	Lossless JPEG2000 pt1	Jp2
MOVING IMAGE	JPEG2000	MXF
	FFV1	MKV
SOUND	LPCM	WAV
		AIFF
	FLAC	FLAC



PACKED

Centre of Expertise in Digital Heritage

#6 MIGRATION & TRANSCODING



The screenshot shows a web browser window with the URL www.digitalpreservation.gov/formats/fdd/descriptions.shtml. The page title is "Format Description Categories...". The header includes "NDIIPP Home | Digital Formats Home" and a search bar with the text "Zoeken". Below the header is a green navigation bar with links: [Introduction](#), [Sustainability Factors](#), [Content Categories](#), [Format Descriptions](#), and [Contact](#). The main content area is titled "Format Descriptions" and contains a breadcrumb trail: "Format Descriptions >> Format Description Categories >> [Browse Alphabetical List](#) >> [Format Descriptions as XML](#)". The content is organized into a grid of categories, each with a list of links to specific format descriptions.

Format Descriptions >> Format Description Categories >> [Browse Alphabetical List](#) >> [Format Descriptions as XML](#)

Format Descriptions

Still Image <ul style="list-style-type: none">• SVG 1.1• TIFF 6• All still image format descriptions	Sound <ul style="list-style-type: none">• WAVE• MP3 FF• All sound format descriptions	Moving Image <ul style="list-style-type: none">• MPEG-4 FF 2• AVI• All moving image format descriptions
Textual <ul style="list-style-type: none">• PDF/A• DOCX/OOXML 2012• All text format descriptions	Web Archive <ul style="list-style-type: none">• ARC IA• WARC• All Web archive format descriptions	Datasets <ul style="list-style-type: none">• DBF• HDF5• All dataset format descriptions
Geospatial <ul style="list-style-type: none">• ESRI shape• GeoPDF 2.2• All geospatial format descriptions	Generic <ul style="list-style-type: none">• ASF• RIFF• All generic format descriptions	



PACKED

Centre of Expertise in Digital Heritage

#6 MIGRATION & TRANSCODING

- Transcode to open and sustainable archive formats

Pro:

- (by far the most) efficient way of extending life cycle of a file

Con:

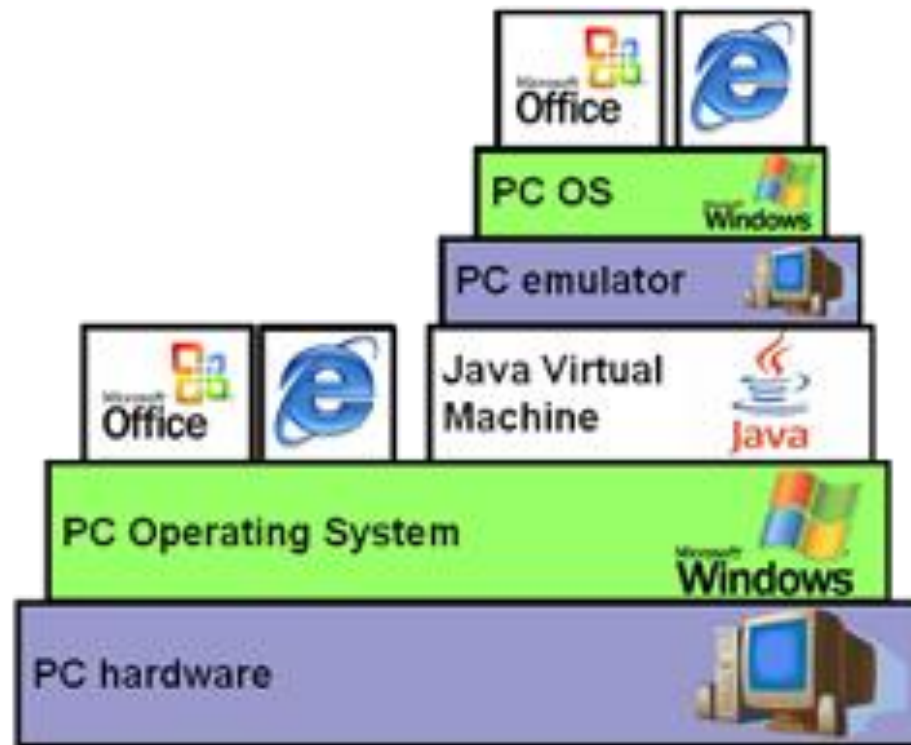
- Risk information loss
- Risk functionality loss
- Expertise!



PACKED

Centre of Expertise in Digital Heritage

#7 EMULATION



#7 EMULATION

- Mimic the original environment in which a file was used

Pro:

- Last resort for obsolete content...

Con:

- Specialist work
- Available for specific platforms
- Requires reverse engineering (legal?)



PACKED

Centre of Expertise in Digital Heritage



PACKED

Expertisecentrum Digitaal Erfgoed

III. STRATEGIES

- do nothing...
- conservation
- documentation
- copy & distribute
- regular checks
- migration & transcoding
- emulation

=> What strategy would counter your threats?

FINAL THOUGHTS:

- One-fits-all solution for long-term preservation DOES NOT exist
- long-term preservation = chain of short-term solutions based on a long term vision
- technology evolves >>> update your strategies!



GOEDE STRATEGIE = COMBINATIE

- verschillende strategie combineren
- alle aspecten moeten beschreven worden tot echte, volledige strategie
- meestal geen kant-en-klare oplossingen
- in bepaalde gevallen nog geen oplossingen





PACKED

Centre of Expertise in Digital Heritage