Different approaches to digital preservation

Hilde van Wijngaarden

Digital Preservation Officer

Koninklijke Bibliotheek/ National Library of the Netherlands

www.kb.nl/e-depot

Digital preservation:

- Safe storage
- Preservation metadata
- Permanent access

Safe storage:

- Secure storage media
- Seperating storage from access
- Refreshment procedures
- Back-up procedures
- International standard: OAIS
- Trusted depositories

Preservation metadata:

- Content description
- Specific preservation information:
  - Provenance
  - Rights
  - Technical metadata
- File format information

Permanent access:

- Rendering may become impossible due to obsolescence of soft- and hardware
- Different strategies possible
- Goal and audience have to be determined

Permanent access policy:

- What kind of digital objects is the repository responsible for?
  - Fixed format texts, web resources, complex digital objects, datasets, programmes, ...
- What do you want to render in the future?
  - Keep the original?
  - What is the original?
  - Offer extended functionalities?
- How do you want to provide this access?
  - Options for the user?
  - Provide the software or give a recommendation?

Possible strategies:

Processing the original:

- Migration
- Normalisation
- Data-extraction

Keeping the original:

- Emulation
- Encapsulation
- Technology preservation (Hardware museum)
- Re-engineering/Data recovery/Digital archaeology

Migration

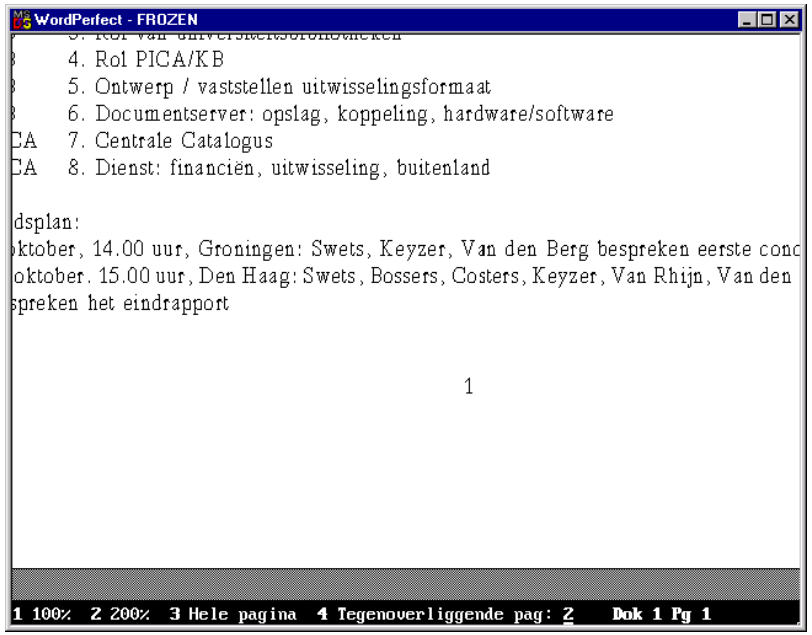
- Hardware migration: refreshment
  - Transferring data to new carriers
- Software migration:
  - Migrate to a new version of the same format
  - Migrate to another format
- Migration at point of access

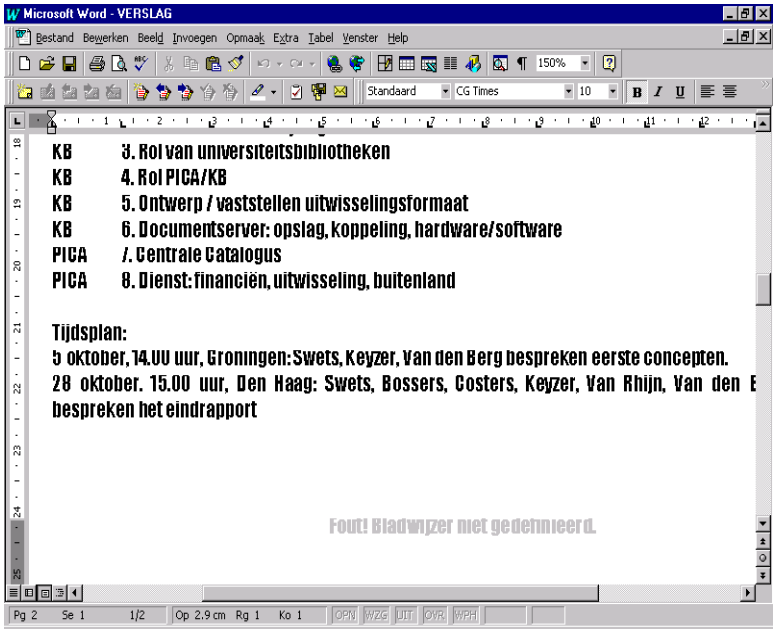
Examples:

- Dutch digital preservation testbed: Migration of wordprocessing documents
- Scientific data archives like EROS, NASA,SDSC
- Camileon: Migration-on-request

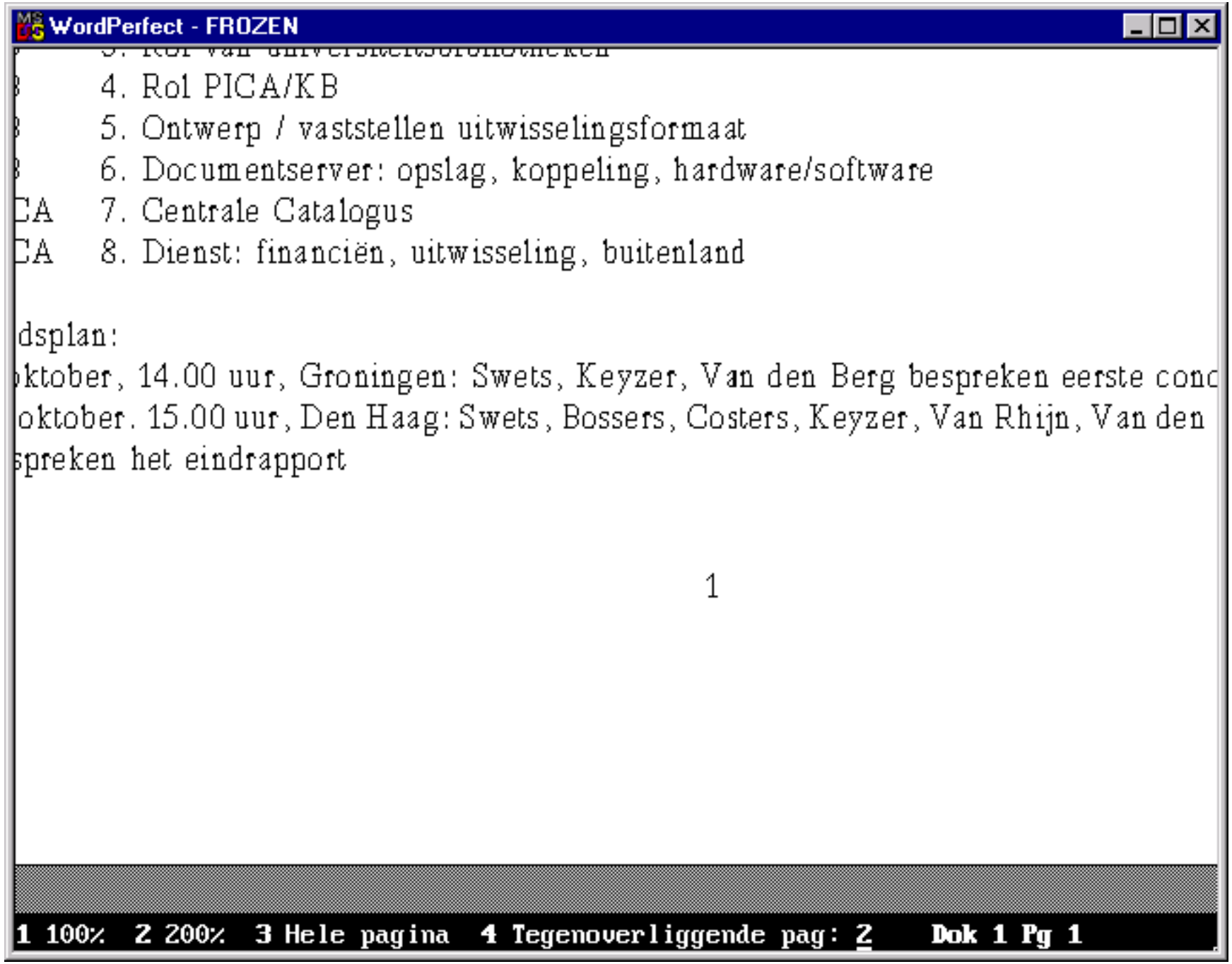
KB

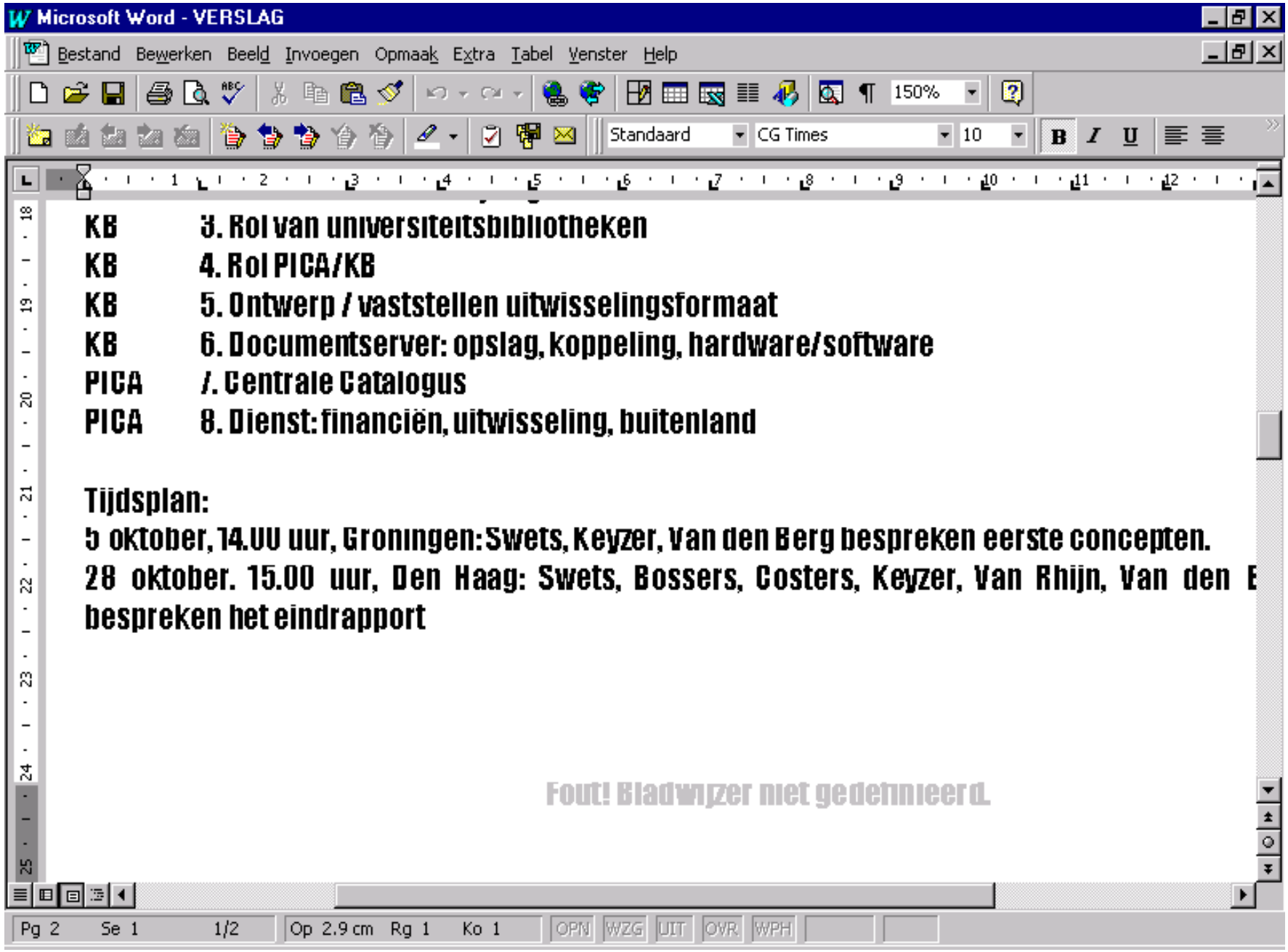
Different Approaches to Digital Preservation





Even a ‘simple’ conversion from WordPerfect to Word 97, shows how many differences can appear...


WordPerfect - FROZEN
4. Rol PICA/KB
5. Ontwerp / vaststellen uitwisselingsformaat
6. Documentserver: opslag, koppeling, hardware/software
7. Centrale Catalogus
8. Dienst: financiën, uitwisseling, buitenland
dsplan:
oktober, 14.00 uur, Groningen: Swets, Keyzer, Van den Berg bespreken eerste conc
oktober. 15.00 uur, Den Haag: Swets, Bossers, Costers, Keyzer, Van Rhijn, Van den
spreken het eindrapport
1
1 100% 2 200% 3 Hele pagina 4 Tegenoverliggende pag: 2 Dok 1 Pg 1


Microsoft Word - VERSLAG
Bestand Bewerken Beeld Invoegen Opmaak Extra Tabel Venster Help
150%
Standaard
CG Times
10
KB 3. Rol van universiteitsbibliotheken
KB 4. Rol PICA/KB
KB 5. Ontwerp / vaststellen uitwisselingsformaat
KB 6. Documentserver: opslag, koppeling, hardware/software
PICA /. Centrale Catalogus
PICA 8. Dienst: financiën, uitwisseling, buitenland
Tijdsplan:
5 oktober, 14.00 uur, Groningen: Swets, Keyzer, Van den Berg bespreken eerste concepten.
28 oktober. 15.00 uur, Den Haag: Swets, Bossers, Costers, Keyzer, Van Rhijn, Van den E
bespreken het eindrapport
Fout! Bladwijzer niet gedefinieerd.
Pg 2 Se 1 1/2
Op 2.9 cm Rg 1 Ko 1
OPN WZG UIT OVR WPH

Migration:

Advantages

- Conversion functionality supplied with software
- Result has a format that is familiar to the user
- New functionalities possible

Disadvantages

- Appearance changes
- Errors occur
- Meaning can be changed
- If applied at point of expected obsolescence, everything has to be migrated, usually repeatedly
- Migration at point of access may not be possible anymore at that time

Normalisation:

Converting all objects into

- One or more preferred formats
- A chosen preservation format, for instance XML
- A more generic format

Normalisation is also used to describe data-extraction:
Creating a logical description of the data, with tags

Examples:

- National Archives of Australia: Storing everything in XML
- Universal Virtual Computer
- Public Record Office Victoria: VERS

Normalisation:

Advantages

- A limited number of formats to maintain
- Formats chosen have a higher chance of surviving longer
- Using a logical description enhances the chances of future comprehension

Disadvantages

- (See migration)
- Not flexible
- Possible wrong choice of formats

Emulation:

Recreating the behaviour of one computer on another

Possibilities:

- Hardware emulation
- Software emulation
- Emulation of an operating system
- Emulation using an intermediate layer or virtual machine

Examples:

- Emulators for game computers
- Universal Virtual Machine
- Emulation Virtual Computer (Jeff Rothenberg)

Emulation:

Advantages

- Original file is kept accessible
- Applicable to every sort of digital object, including programmes
- One-time effort for large groups of digital objects

Disadvantages

- Never operationalised for digital preservation
- Technological challenging
- Result may not be what user wants

Encapsulation:

'Wrapping' the content in a description

Possibilities:

- Including the original file in an XML document
- Including links to software with the file in the description
- Including the software itself

Examples:

- Archival Information Packages (AIP) that contain metadata and content files
- VERS

Encapsulation:

Advantages

- Keeping options open through extensive descriptions

Disadvantages

- Updating metadata difficult
- In fact: nothing has really been done yet, strategy still has to be chosen
- Including (links to) software does not offer any guarantees

Technology preservation

- Often referred to as a hardware museum
- Saving everything: files, software and hardware and keep them alive
- Maintanance almost impossible
- Unworkable for larger quantities

Re-engineering

- Also called data recovery or digital archaeology
- Saving the bits and restore their readability/usability
- Labour intensive and technical challenging
- The original is not available so no way to know how it should look like

Current choices:

- Most repositories keep their options open
- Migration usually preferred
- Choices depend on sort of digital objects
  - Normalisation applied if content is considered the first priority
  - Encapsulation if context is important
  - Emulation (all thought not operational yet) for complex digital objects
- Choices depend on state of R&D
  - Large scale migration not necessary yet because digital archiving is new
  - Hesitation about emulation because there is not a working example available

Strategies are not enough: we need tools that...

- Make a strategy possible (emulators, virtual machines)
- Help choose a strategy
- Help perform the strategy
- Maintain the link between originals and conversions
- Enable interoperability and co-operation between different repositories

Tools have to be implemented

- In the digital archiving system
- In the digital archiving workflow

Any questions?