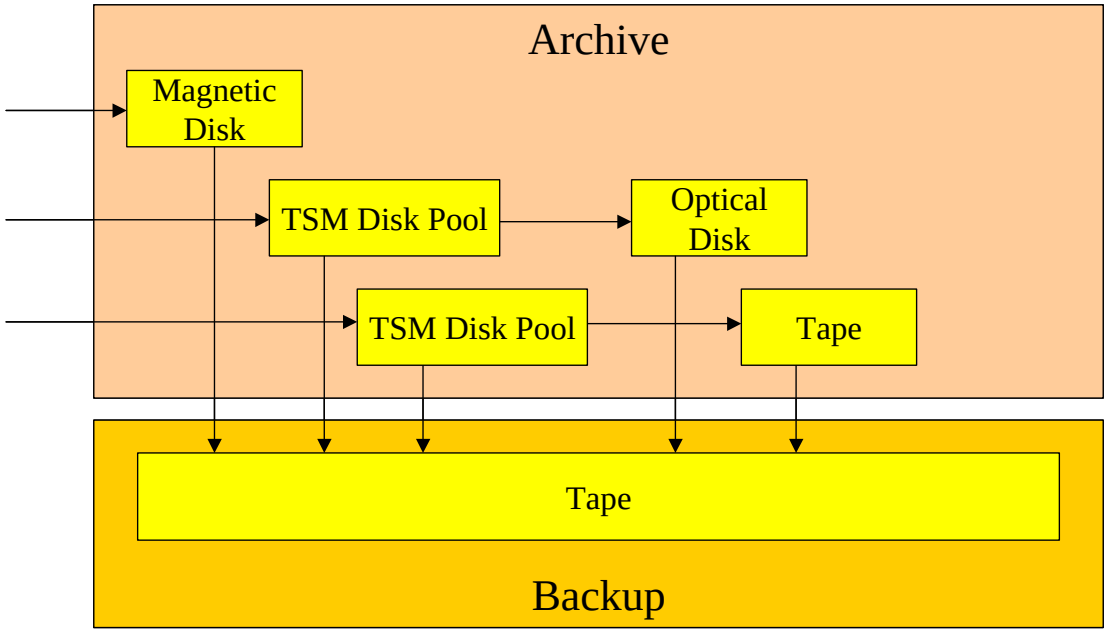
General Introduction: Technological Issues

Reinhard Altenhöner
Die Deutsche Bibliothek, Frankfurt am Main, Germany
altenhoener@dbf.ddb.de

Complex storage systems



Binary Code

1101110111001111001101001010101110101110010
1010111000110101010101000110101010101010101
0100010101010101010101010101010101000101010
1010101010100111000001011111100110111011100
1111001101001010101110101110010101011100011
0101010101000110101010101010101010001010101
0101010101010101010101000101010101010101010
0111000001011111100110111011100111100110100
1010101110101110010101011100011010101010100
0110101010101010101010001010101010101010101
01

Archiving the bit stream

- Digital information is stored as a bit stream on physical media
  - → Storage media types change quickly and are subject to obsolescence
  - → Storage media are unstable and can degrade quickly

Substantial preservation I

Starting point for all following activities, depends from

- Stability of the data-carrier
- Copying (refreshing)
- Migration of the data carrier
  e.g. magnetic --> optical
- Risk management

Substantial preservation II

Factors of disturbance

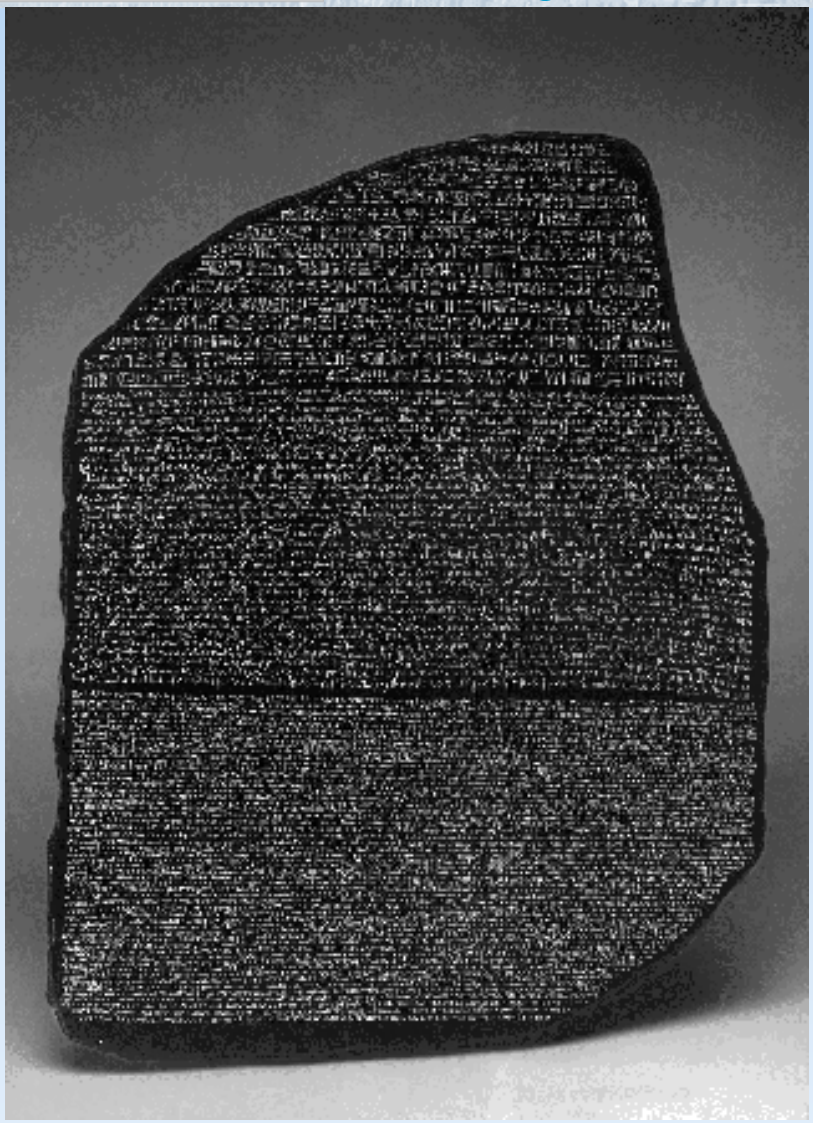
- the mass of data
  - ➲ automation
- unexpected material behavior
  - ➲ current material observation (in general: technology watch)
- failures in copy procedures
  - ➲ strictly control of the processes
- copy protection, DRM
  - ➲ no solution until today ...

Code ...

```
   …À „;   _^¸   [ƒÄ Ã¡Øò   ƒøÿ „!   L$,j Q T$ j R L$ j Qhp " PÇD$,xV4 ÿ è°   …À „ð ‹D$ _%ÿÿ ^
   [ƒÄ Ã‹
P.   …É „Í ‹ ƒÆ VÿR$÷Ø À_÷Ø%ÿÿ ^Áà [ƒÄ Ãƒù ,¦ ‹
P.   …É „˜ ‹V ‹ RÿP(÷Ø À_÷Ø%ÿÿ ^Áà [ƒÄ Ã=ç ‡o „3 P¤ƒú ‡] ÿ$•˜: ‹
P.   …É „H ƒø<¸€² t ¸¬² Pj j ÿ Ì° ‹ð…ö „$ j j j j Vÿ Ä° ‹ø…ÿ „ ‹
P.   W‹ ÿP …À •ÁWˆL$0ÿ Ô° Vÿ ± ‹D$,_%ÿ ^Áà [ƒÄ Ãƒù ¸Ì ¡P. …À „¿ hÐ² ÿ |± ƒÀ
       Pè°â ‹V ƒÄ ‹ØRhÐ² h@õ Sÿ ² ‹F ƒÄ ƒø=¸€² t ‹ÃPj j ÿ Ì° S‹øè â ƒÄ …ÿ „a j j j
       j Wÿ Ä° ‹Ø…Ûu Wÿ ± _^3À[ƒÄ Ã‹V ‹
P.   R‹V ‹ SRÿP …À •ÀSˆD$0ÿ Ô° Wÿ ± ‹D$,_%ÿ ^Áà [ƒÄ Ãƒù ¸õ
       ƒøP¸€² t ¸D³ Pj j ÿ Ì° ‹ð…ö „Ñ j j j j Vÿ Ä° …À „³ ‹
L.   P‰ ÿ Ô° Vÿ ± _^¸ [ƒÄ Ã‹
P.   …É „Š ‹ D$,PÿR,…À „x ‹D$,_%ÿ ^
   [ƒÄ Ãƒù ¸Z ‹
P.   …É „L ŠF ‹ PÿR0÷Ø À_÷Ø%ÿÿ ^Áà [ƒÄ Ã‹
P.   …É „ ƒø^¸€² t ¸ø² Pj j ÿ Ì° ‹ð…ö „ü j j j j Vÿ Ä° ‹ø…ÿ „Ü ‹
P.   W‹ ÿR4…À •ÀWˆD$0ÿ Ô° Vÿ ± ‹D$,_%ÿ ^Áà [ƒÄ Ã‹
P.   …É „Ÿ ƒø_¸€² t ¸ ³ Pj j ÿ Ì° ‹ð…ö „{ j j j j Vÿ Ä° ‹ø…ÿ „[ ‹
P.   W‹ ÿR8…À •ÀWˆD$0ÿ Ô° Vÿ ± ‹D$,_%ÿ ^Áà [ƒÄ Ãƒù ¸# ‹
P.   …É „ ‹ D$0P‹F PÿR<‹T$0_…À •ÁˆL$(‹D$(%ÿ âÿÿ Áà ^
       Â[ƒÄ Ãƒù ¸Ö ‹
P.   …È „È ‹V ‹ R‹V RÿP@÷Ø À_÷Ø%ÿÿ ^Áà [ƒÄ Ã ù ¸š ‹
P.   …É „Œ – ‹ R V R‹V R‹V RÿPD÷Ø À_÷Ø%ÿÿ ^Áà [ƒÄ Ãƒù rZ‹
P.   ‹V R‹ ÿPP÷Ø À_÷Ø%ÿÿ ^Áà [ƒÄ Ãƒù r1‹
|.   ‹F j j h€ Q£ . ÿ T² _^¸ [ƒÄ ÃVÿ ± _^3À[ƒÄ Ã !/ 7/ g/ {/ 0 =0 ]0 ä0 Z1 Ö9

                                                    ‹ÿ 2 &2 s2 3 €5 6 x3 ª3 ß3 4 I4 §4 Ø6
       û4 Ö9

>7 s7 ¨7 )8 ª8 ÷8 09 w9 €5 6 ¨7 )8 Ø6
       S‹\$ Sÿ (² …Àu [Ã‹L$ ‹T$ D$ PQRè íÿÿƒÄ …Àu [Ã‹D$ €Áà ‹ˆhð …Ét 3À[Ã‹L$ ‹T$
     V‹t$ W‹|$ ‰°`ð ‰¸dð ‰ˆlð ‹
P.   ‰ pð ‰˜hð ‹€tð ‹ j PÿR ‹D$ ‹
P.   V‹ €‹ Åxð PÿR ‹D$ ‹
P.   W‹ €‹ Å|ð PÿR _^¸ [Ã ‹L$ ‹T$ D$ PQRè\ìÿÿƒÄ …Àu Ã‹D$ €Áà ‹ˆhð …Éu 3ÀÃ‹
```

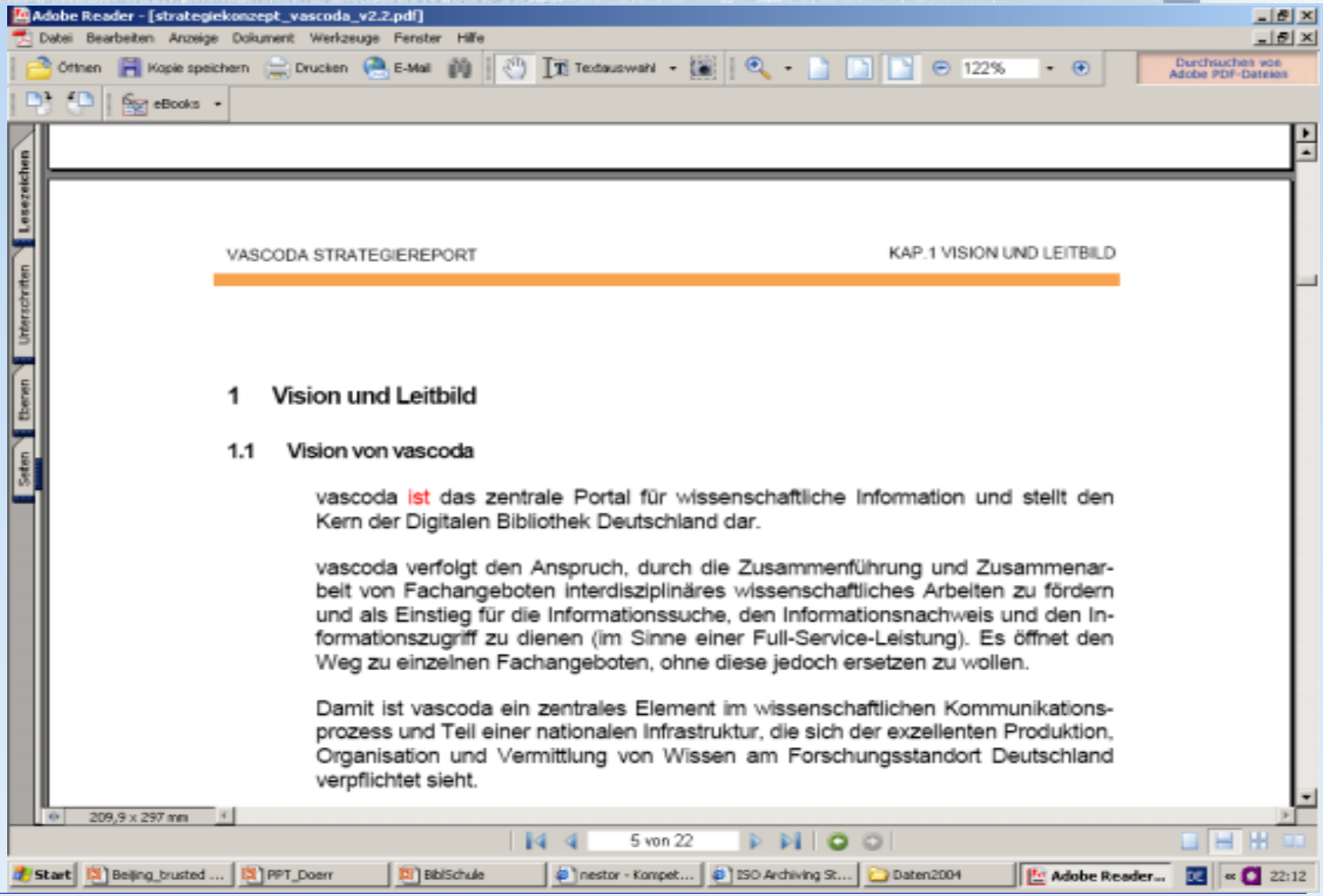
Of course code ...

```
</script>
	</head>

	<frameset cols="238,*" framespacing="0" border="0" frameborder="NO" onLoad="BrowserTest()">
		<frameset rows="58,*" border="0" framespacing="0" frameborder="NO">
			<frame src="home/streifen_home.htm" name="Sprachwahl" title="Bilder der drei
	Standorte" noresize scrolling="NO">
			<frame src="home/menu.htm" name="Menuepunkte" title="Links" noresize>
		</frameset>
		<frameset rows="58,*" border="0" framespacing="0" frameborder="NO">
			<frame src="home/signet.htm" name="Signetrahmen" title="Rahmen mit Signet"
	noresize scrolling="NO">
			<frame src="home/home.htm" name="Textrahmen" title="Textbereich" noresize>
		</frameset>
	</frameset>
	<noframes>
		<body bgcolor="#f3f0ef">
			<h2>Leider kann diese Seite nicht angezeigt werden, da Ihr Browser keine Frames
	unterst&uuml;tzt.</h2>
		</body>
	</noframes>
</html>
```



Interpretation of Code

- The Rosetta Stone
- * 196 a. Chr.
- founded 1799
- Today: British Museum, London


Adobe Reader - [strategiekonzept_vascoda_v2.2.pdf]
Datei Bearbeiten Anzeige Dokument Werkzeuge Fenster Hilfe
Öffnen Kopie speichern Drucken E-Mail Textauswahl 122%
eBooks
Lesezeichen
Unterschriften
Ebenen
Seiten
VASCODA STRATEGIEREPORT
KAP.1 VISION UND LEITBILD
1 Vision und Leitbild
1.1 Vision von vascoda
vascoda ist das zentrale Portal für wissenschaftliche Information und stellt den Kern der Digitalen Bibliothek Deutschland dar.
vascoda verfolgt den Anspruch, durch die Zusammenführung und Zusammenarbeit von Fachangeboten interdisziplinäres wissenschaftliches Arbeiten zu fördern und als Einstieg für die Informationssuche, den Informationsnachweis und den Informationszugriff zu dienen (im Sinne einer Full-Service-Leistung). Es öffnet den Weg zu einzelnen Fachangeboten, ohne diese jedoch ersetzen zu wollen.
Damit ist vascoda ein zentrales Element im wissenschaftlichen Kommunikationsprozess und Teil einer nationalen Infrastruktur, die sich der exzellenten Produktion, Organisation und Vermittlung von Wissen am Forschungsstandort Deutschland verpflichtet sieht.
209,9 x 297 mm
5 von 22
Start
Beijing_trusted ...
PPT_Doerr
BibliSchule
nestor - Kompet...
ISO Archiving St...
Daten2004
Adobe Reader...
22:12

Technical key issues

→ Migration

→ Emulation

→ Standardization

Migration (1)


Diese Präsentation wird aus einer früheren PowerPoint-Version aktualisiert.
Datei: C:\TEMP\MMB-Präsentation.ppt
Folien fertig konvertiert: 7 von 15
Abbrechen

Migration (2)

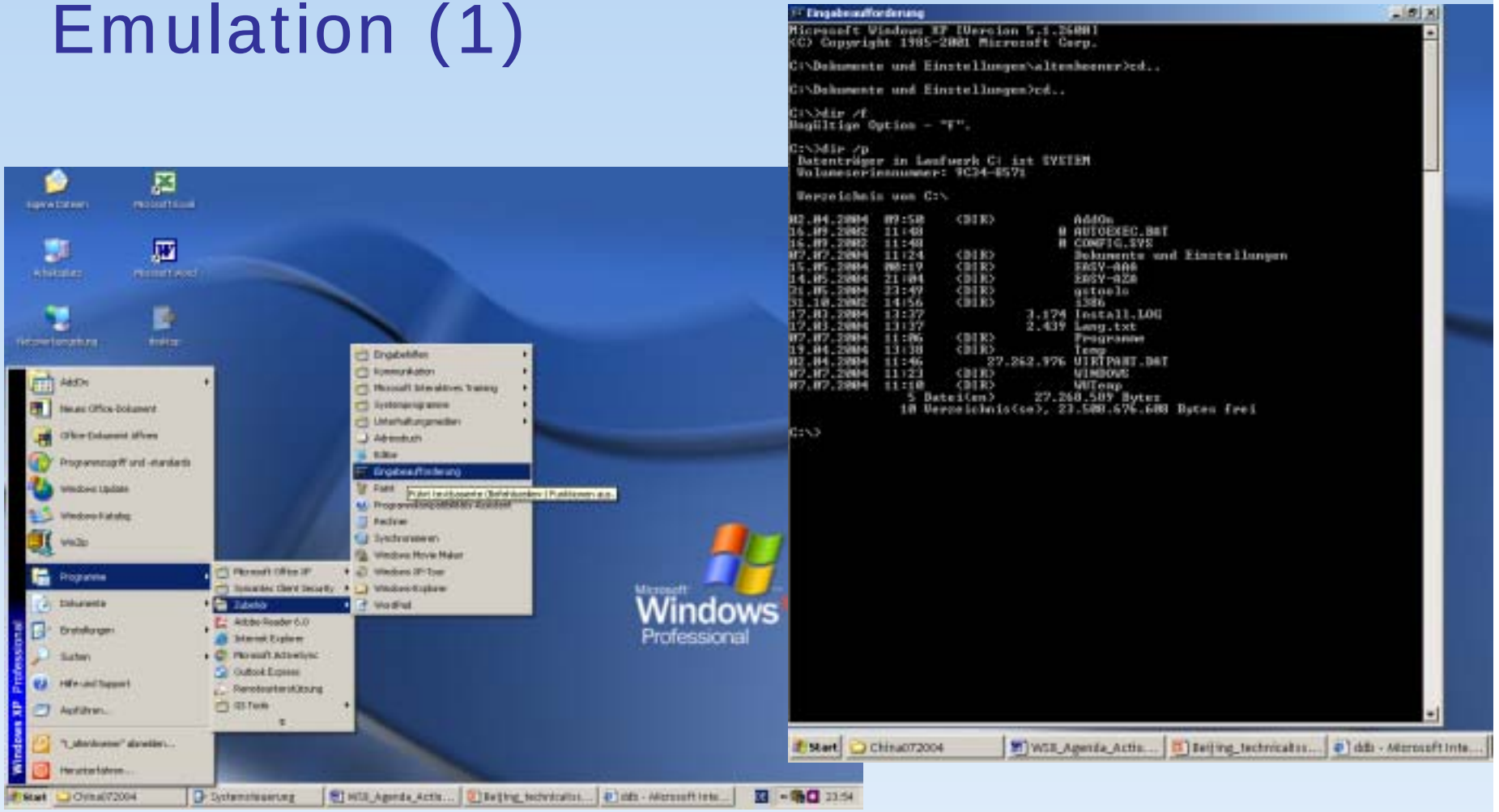
Advantages:

- Well-tried strategy
- just in case ??!!

Disadvantages

- No end
- Barriers
- Risk to fake the data
- Risk to lost information

Emulation (1)


Eingabeaufforderung
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.
Windows XP Professional
Microsoft
Windows
Professional

Emulation (2)

Advantages:

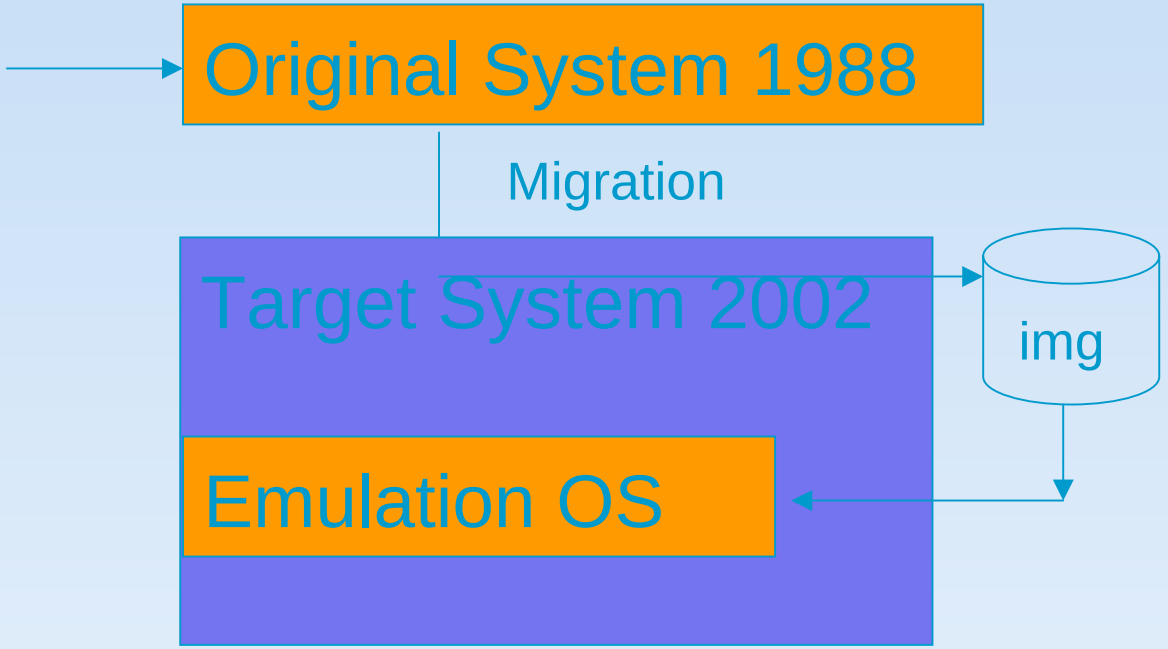
- Objects are preserved without change
- Calculable amounts
- For all digital objects

Disadvantages:

- High investment necessary for R&D
- Business value??!!
- feasibility

Migration / Emulation in production in DDB

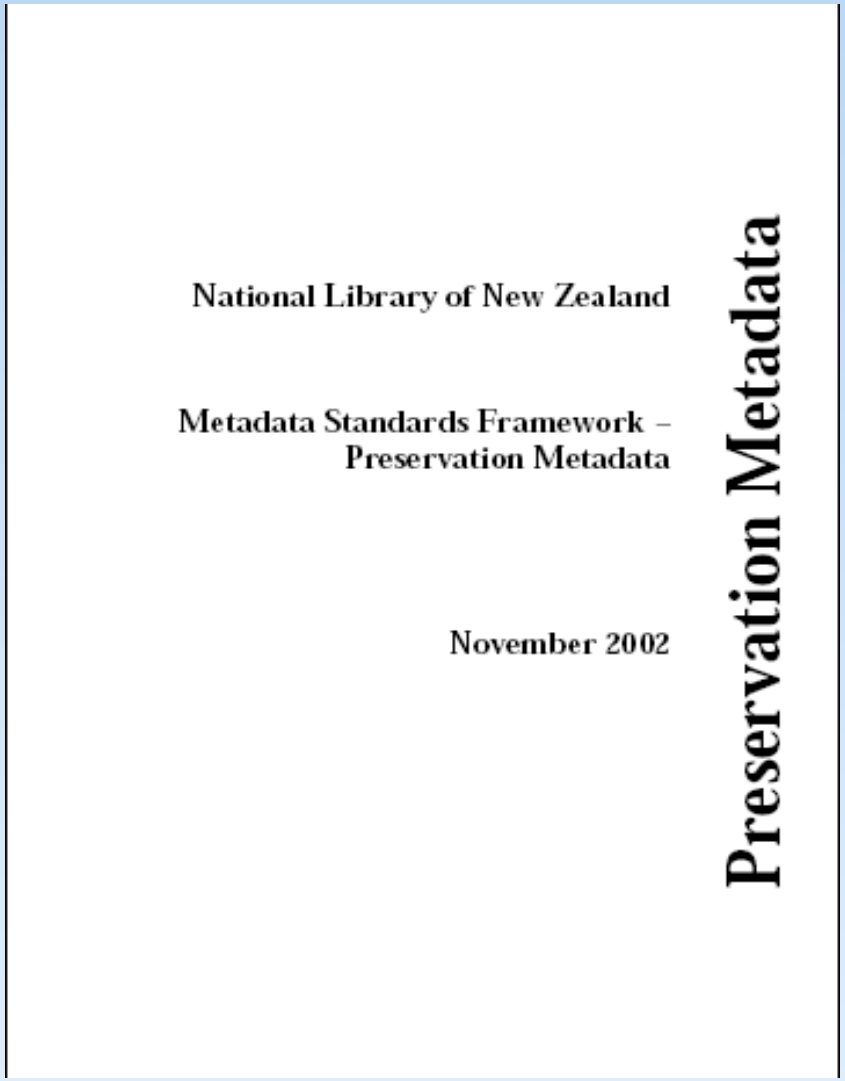
eg. C64 / AMIGA



Rendering objects

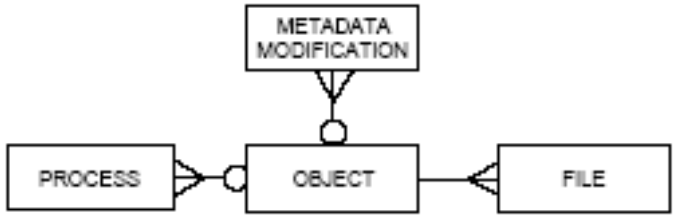
We need information about the production environment of the given object:

- Software
- Technical context
- Embedded objects
- Metadata (they also have to be considered in the long-term preservation)



Standardization

Appendix 1 – Preservation Metadata Model



ENTITY 1 - OBJECT

1.1 Name of object
1.2 Reference number
1.3 Identifier - Object IID
1.4 Persistent identifier - PID
1.5 UNIX location
1.6 Date of creation of preservation master
1.7 Technical composition
1.8 Structural type
1.9 Hardware environment
1.10 Software environment
1.11 Installation requirements
1.12 Access inhibitors
1.13 Access facilitators
1.14 Quirks
1.15 Authentication
1.16 Metadata record creator
1.17 Date of metadata record creation
1.18 Comments

ENTITY 2 - PROCESS

2.1 Object IID
2.2 Process
2.3 Purpose
2.4 Business unit or agency
2.5 Permission
2.6 Date of permission
2.7 Hardware used
2.8 Software used
2.9 Steps
2.10 Result
2.11 Guidelines
2.12 Completion date and time

ENTITY 3 - FILE

Common elements:
3.1 Object IID
3.2 File IID
3.3 Structural context
3.4 Filename and extension
3.5 File size
3.6 File date and time
3.7 MIME type/format
3.8 Version
3.9 Target indicator

3.10 Image:
3.10.1 Resolution
3.10.2 Dimensions
3.10.3 Tonal resolution
3.10.4 Colour space
3.10.5 Colour management
3.10.6 Colour lookup table
3.10.7 Orientation
3.10.8 Compression

3.11 Audio:
3.11.1 Resolution
3.11.2 Duration
3.11.3 Bit rate
3.11.4 Compression
3.11.5 Encapsulation
3.11.6 Track number and type

ENTITY 3 – FILE (cont.)

3.12 Video:
3.12.1 Frame dimensions
3.12.2 Duration
3.12.3 Frame rate
3.12.4 Compression
3.12.5 Encoding structure
3.12.6 Sound

3.13 Text:
3.13.1 Compression
3.13.2 Character set
3.13.3 Associated DTD
3.13.4 Structural divisions

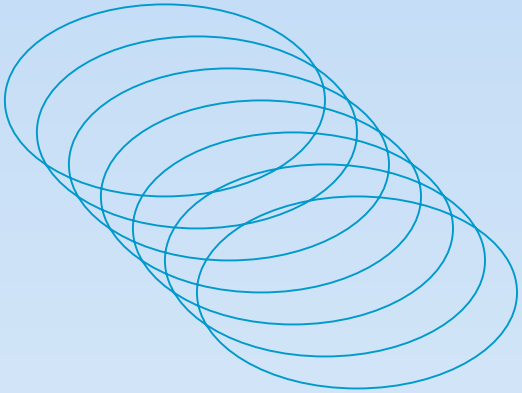
3.14 Datasets:
Uses common elements only

3.15 System Files:
Uses common elements only

ENTITY 4 - METADATA MODIFICATION

4.1 Object IID
4.2 Modifier
4.3 Date and time
4.4 Field modified
4.5 Data modified

Archiving of net-publications



DEPOSIT-

SYSTEM

Object-specific

metadata

& Data

Metadata

- Specific Preservation Metadata are necessary to ensure that information can be rendered in the future
  → Metadata about:
- Provenance
- Structure
- File Format(s)
- Technical Environment
- Rights

How much metadata can be extracted automatically during ingest process?

→ Tools needed!

Additional remarks

- Migration of the metadata
- Digital information can be easily changed

  → Threats to the authenticity and integrity of the resources
- Long-term preservation of metadata