

PERmanent Visual ARchive

**PEVIAR - ein System zur migrationslosen
Archivierung digitaler Daten**

KOST

20. November 2007

Dr. Peter Fornaro
peter.fornaro@unibas.ch



Die digitale Revolution ist bereits Teil unserer Geschichte

Woher kommen die digitalen Daten?

- Daten „born digital“
 - Bsp: 99 von 100 verkauften Kameras sind digital
 - Bsp: Emailverkehr
 - Bsp: Medizinische Dokumente
- Daten aus Digitalisierungen
 - Analoge Größen unterliegen einem steten Alterungsprozess
 - Ausweg: **Digitalisieren**

Digitale Datenbestände eröffnen neue Möglichkeiten (**beträchtlicher Mehrwert**):

- Zugriff
- Suche
- Verbreitung

3

Digitale Archivierung

Die Folgen sind:

- Extremes Wachstum der zu erhaltenden digitalen Daten

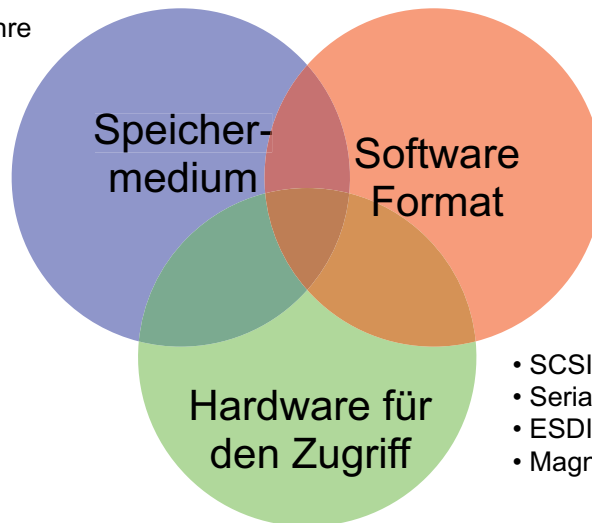
Realität in der IT:

- Der **Datenträger** altert
- Die **Softwareformate** ändern sich
- Die **Hardwaretechnologie** unterliegt einem Wandel

Drei Problembereiche im digitale Archiv

- CD-R Lebensdauer: 2-30 Jahre
- Magnetband: 30 Jahre

- WordStar (CP/M)
- DisplayWrite (IBM)
- MacDraw (Apple)
- PS16 (ProTracker)



- SCSI
- Serial ATA
- ESDI
- Magneto-Optical Drive

03/30/2007

5

Digitales Archiv? Nichts Neues!

- Digital bedeutet: „Mit **diskreten** Zeichen dargestellt“
 $S = \{s_1, s_2, s_3, \dots, s_n\}; n \geq 2$
- Jeder Computer arbeitet **digital**: Mit einem zweiwertigen „**binären**“ Zeichensatz
 $-S = \{0,1\}, S = \{\text{true}, \text{false}\}, S = \{+5V, -5V\}, S = \{\uparrow, \downarrow\}$
- **Unser Alphabet ist allerdings auch digital** => 26 diskrete Zeichen!

Wie kann digital gespeicherte Information lange erhalten bleiben?

03/30/2007

6

Die (kurze) Geschichte der digitalen Archivierung

03/30/2007

7

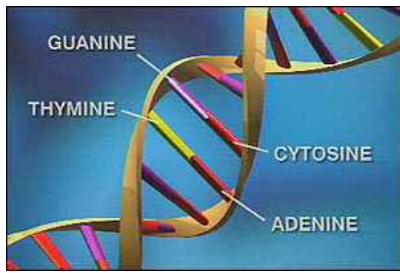
Weg 1

Durch **stete (automatische) Migration** eines relativ kurzlebigen Speichersystems

03/30/2007

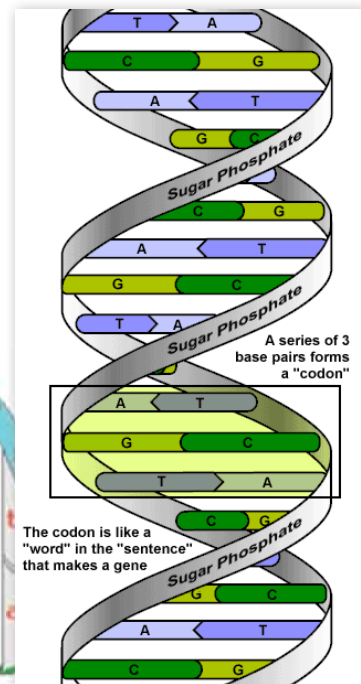
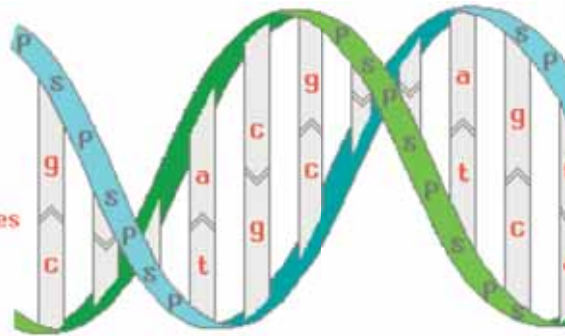
8

DNA (Deoxyribonucleic Acid) in Paaren = **2-bit Code**



s = sugar
p = phosphate

g = guanine
c = cytosine
a = adenine
t = thymine] bases



A series of 3 base pairs forms a "codon"

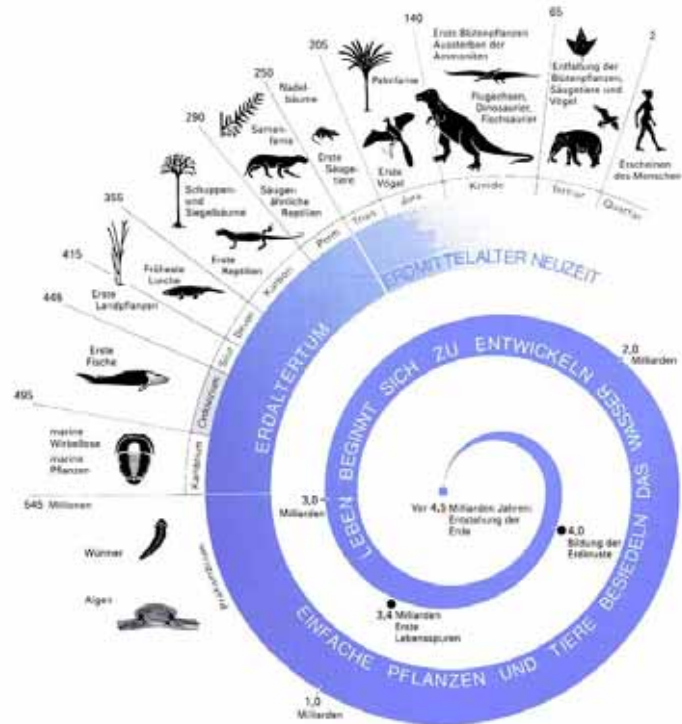
The codon is like a "word" in the "sentence" that makes a gene

Beispiel Quastenflosser - „lebende Fossilien“



Quastenflosser unverändert
über **> 70 Mio. Jahre**

mehrere Millionen
fehlerfreie Replizierungen!



Weg 2

Durch einen möglichst **langlebigen Speichersystem**, das mit wenigen oder gar keinen Migrationen auskommt

Keilschrift

- Permanentes Medium: gebrannter Ton
- "sichtbar"
- Syntax & Semantik entschlüsselt (William Henry Fox Talbot)
- Erfolg !!



MS 2020/08
Quantities of beer for offering.
Sumer, 2034 BC

13

Migration

- **Methode:**
 - Nicht das Medium erhalten, sondern die Information, indem die Daten kopiert werden, bevor das Medium obsolet oder unlesbar wird.
- **Vorteil:**
 - **Es funktioniert** (es wurde schon gemacht)
 - Mehrere Kopien (=Redundanz) - eine Notwendigkeit für eine saubere Migrationsprozedur - vermindert das Risiko eines Totalverlustes
(**LOCKSS** = **L**ots of **C**opies **K**eep **S**tuff **S**afe)

14

Migration

- Risiken/Probleme:
 - Eine korrekte Migration digitaler Daten ist ein **mühsamer** und **schwieriger** Prozess
 - **Konsistenz** der Datenbestände!
 - Man darf den **richtigen Zeitpunkt** nicht verpassen
 - **Wiederkehrende Kosten ohne unmittelbaren Mehrwert**
 - Hängt von einer funktionierenden, technisch orientierten Gesellschaft ab, die sich um das kulturelle Erbe kümmert.
 - Es ist keine „**fire and forget**“-Methode!

15

Welchen Weg wählen?

Aus heutiger Sicht:

- Beides macht Sinn und sollte komplementär eingesetzt werden
- Hängt von der **Anwendung** ab
- Beides sind (noch) Forschungsgebiete
 - Automatische Migration
www.distarnet.ch
 - Migrationslose Archivierung
www.peviar.ch

16

Anwendungen Migrationsbehaftet

- Hohe Lesewahrscheinlichkeit
 - Serverlösung
 - Verteilte zentral organisierte Institution bietet Erhalt der Daten als Dienstleistung an
 - Evaluation, Anschaffung und Migration werden von der zentralen Stelle organisiert
 - Economy of Scale
- Hohes Datenaufkommen
 - Film, Video, Audio (Bsp. BBC: A small migration is about **Petabyte**), Grosse Bildarchive
 - Zentral organisierte Institutionen

17

Anwendungen Migrationslos

- Geringe Lesewahrscheinlichkeit
 - Jegliche Kosten für allfällige Migrationsschritte sind zu vermeiden
 - Oft werden weniger als 1% der Daten wieder gelesen, aber 100% müssen für die ganze Pflichtdauer erhalten werden
- Geringes Datenaufkommen
 - Textdaten (Records, Amtl. Dokumente), Vektordaten (Baupläne, med. Daten), Bitmaps

18

Konzept der migrationslosen Speicherung

- Verfasst in einem **bekanntem** und verbreiteten **Format** (Sprache)!
- Gespeichert auf einem **stabilen Träger**
- **Einfacher Zugriff** für das Lesen
Durch visuellen Charakter der Information
=>human readable!
- „LOCKSS“



03/30/2007

19

Anwendung auf die IT

	Klassisch	Computertechnik
Datenträger	Pergamin, Papyrus, Papier	Lochstreifen, Fotomaterial, Rosetta-Storage
Format	Latein Griechisch Englisch	TIFF, AIFF RTF, PDF/A ASCII, XML
Interface	Auge (visuell)	Kamera Scanner

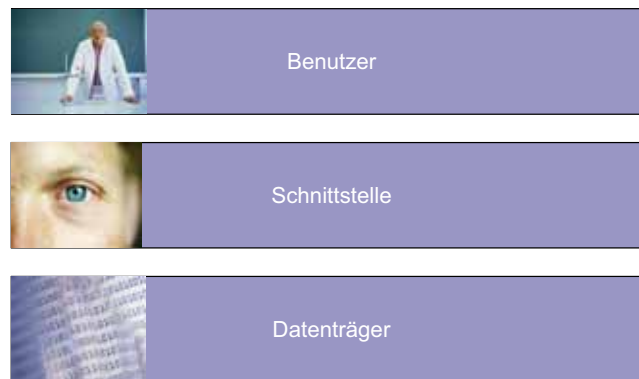
03/30/2007

20

Schnittstelle Mensch-Daten

Problem der **Schnittstelle**:

In der IT sind wir bis heute von **hoch technischen Geräten abhängig**. Ein Buch können wir lesen, da visuell!



21

Anforderungen an das "visuelle" Medium

- Sehr stabil und langlebig
 - Hohe Datendichte
 - Einfache Verarbeitung auch grosser Mengen
 - Akzeptiert bei den Archivaren (!)
 - Leicht in grossen Mengen erhältlich
 - Preis
- => **Photographischer Mikrofilm**

22

- Mikrofilm ist bekannt und sehr stabil
- Preiswert
- Infrastruktur für

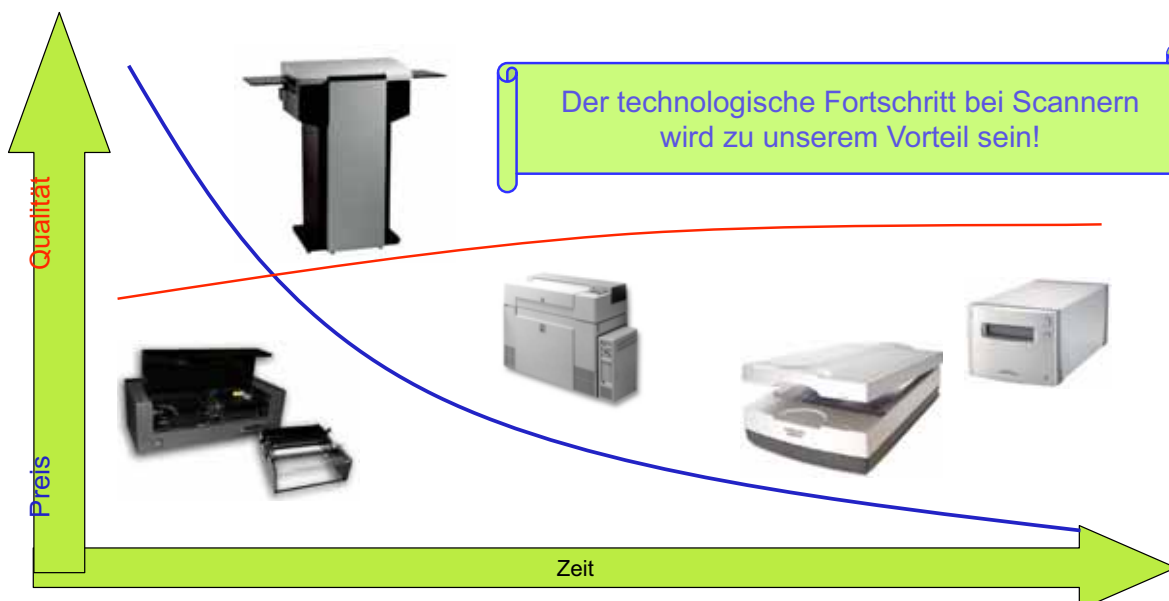
PEVIAR

- **Visuelles speichern** digitaler Daten
- **Daten plus Erklärung (z. B. in mehreren Sprachen) und Illustrationen = Information => Hybridspeicher**
- **Vorteile:**
 - Unabhängig von spezieller Hardware:
Jeder Scanner und jede Kamera kann visuelle Bilder aufnehmen
 - Technische Innovation wird den Leseprozess immer billiger und einfacher machen (neue, bessere Scanner etc.)
 - Bilder (analoge Information) und Text (symbolischer Code) kann **auf dem selben Medium** integriert werden um die Bedeutung der Daten zu erklären
 - „Decoding Manual“, sogar Source-Code auf dem **selben Medium**

03/30/2007

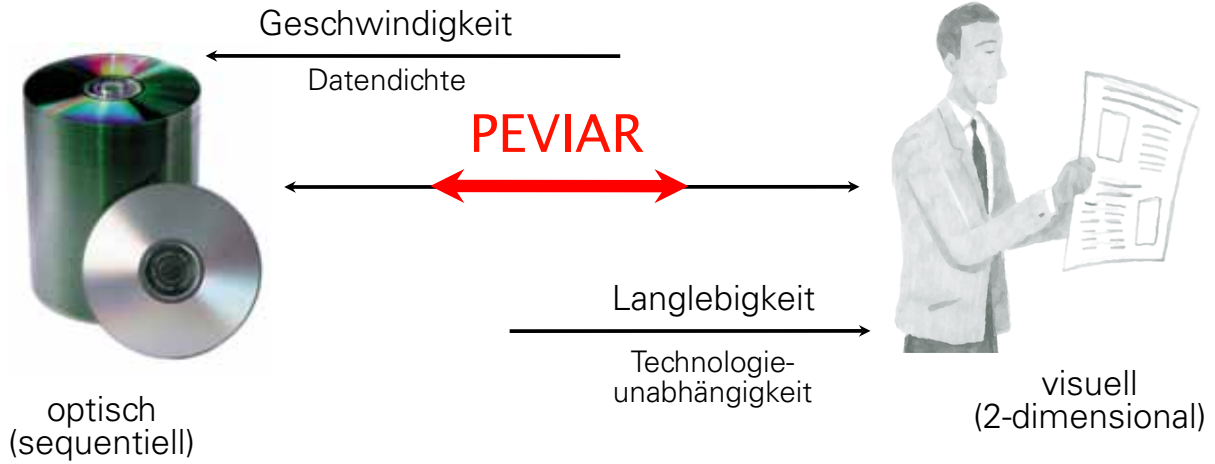
23

Entwicklung der Scanner-Technologie: Das Scannen von Bildern gehört zu den digitalen Basistechnologien, die sich in den letzten Jahren nur im Detail verändert haben.



24

Digitale Information Mensch vs. Computer



Peviar verbindet beide Welten!

Das will der Archivar



Wahrnehmung durch den Nutzer mit Hilfe von Soft- und Hardware

Konzeptuelles Objekt

Das Objekt "zum Begreifen"

```
001001000111001
1001000011001101
001100110001100
1001001001010101
1001000100001111
100110000011010
```

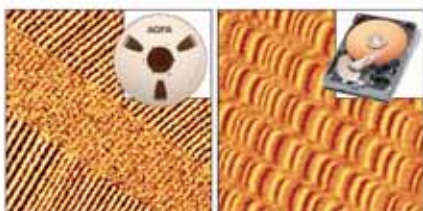


Der Bitstrom wird durch Software als Dateiformat erkannt

Logisches Objekt

Eine Folge von Bits als Einheit

Und damit beginnt alles

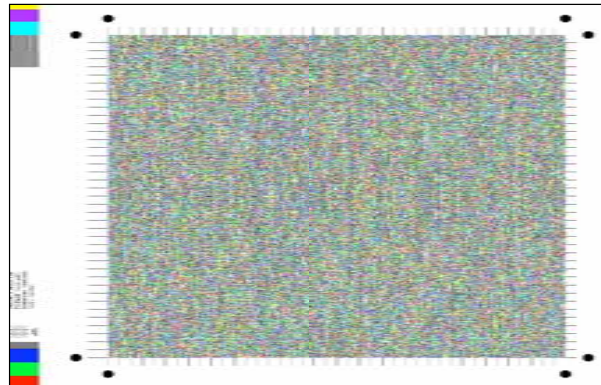
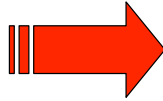


physische Datenstruktur

Physisches Objekt

Daten auf einem Speichermedium

Digital Encoding



1. Ein beliebiger Bit-Code (hier: eine TIFF Datei) wird in Form eines 2D Barcode wiedergegeben
2. Als Datenträger wird ein langzeitstabiles, technologie-unabhängiges Medium verwendet

03/30/2007

27

Visueller binäre Code Beispiele Barcode



Level=10



1234567890

Level=50

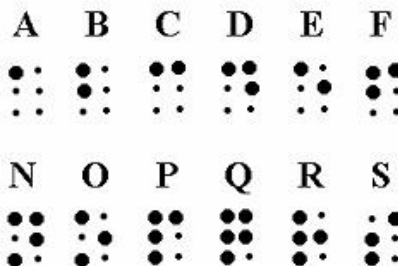


1234567890

Level=95



1234567890



28

Vorteile der Archivierung von digitalen Daten auf Mikrofilm

Stabilität

Beispiel:

Durch das **visuelle Interface** verbunden mit einem extrem stabilen Datenträger und selbst-erklärendem Code wird eine **sehr hohe Lebensdauer und Technologieunabhängigkeit** der Information erreicht.

03/30/2007

Migrationslosigkeit

Beispiel:

Datenbestände mit **Aufbewahrungspflicht** können mit Peviar so abgelegt werden, dass keine weiteren Migrationsschritte mehr notwendig sind und die Lesbarkeit garantiert bleibt.

Unveränderbarkeit

Beispiel:

Wichtige Dokumente wie Verträge oder Pläne können auf Mikrofilm **unveränderbar** in digitaler Form abgelegt werden. Eine Signatur bestätigt die Echtheit (Notariell beglaubigt)

31

Besten Dank...

Kapuzinerpater
Paolino unter dem
Friedhofstor von Sur,
1909, Rudolf
Zinggeler, EAD Bern

Wir danken dem
Schweizerischen
Nationalfonds für
die finanzielle
Unterstützung

<http://www.abmt.unibas.ch>
<http://www.iml.unibas.ch>
<http://www.distarnet.ch>
<http://www.peviar.ch>

peter.fornaro@unibas.ch