

Langzeitarchivierung Bewahrung digitaler Kultur

Fakultätentag Informatik

19. November 2015

Langzeitarchivierung



Stadtarchiv Köln

Langzeitarchivierung

(engl.: digital preservation, long-term preservation)

'Langzeit' ist die Umschreibung eines nicht näher fixierten Zeitraumes, währenddessen wesentliche nicht vorhersehbare technologische und soziokulturelle Veränderungen eintreten, die sowohl die Gestalt als auch die Nutzungssituation digitaler Ressourcen in rasanten Entwicklungszyklen vollständig umwälzen werden.

(Zitat aus: Grundlagen der praktischen Information und Dokumentation. - München, 2004)

Langzeitarchivierung digitaler Objekte umfasst alle Maßnahmen, die dazu dienen, digitale Objekte für die Nachwelt dauerhaft zu erhalten.

NESTOR - Kompetenznetzwerk Langzeitarchivierung

Gedächtnis

Neuronales Gedächtnis (Embodied)

Soziales Gedächtnis (Embedded)

Kulturelles Gedächtnis (Externalisiert)

„Unter dem Begriff des kulturellen Gedächtnisses fassen wir den jeder Gesellschaft und jeder Epoche eigentümlichen Bestand an Wiedergebrauchs-Texten, -Bildern und -Riten zusammen, in deren »Pflege« sie ihr Selbstbild stabilisiert und vermittelt, ein kollektiv geteiltes Wissen vorzugsweise (aber nicht ausschließlich) über die Vergangenheit, auf das eine Gruppe ihr Bewußtsein von Einheit und Eigenart stützt.“
(Jan Assmann)



Ebstorfer Weltkarte



Deutsche Nationalbibliothek, Leipzig

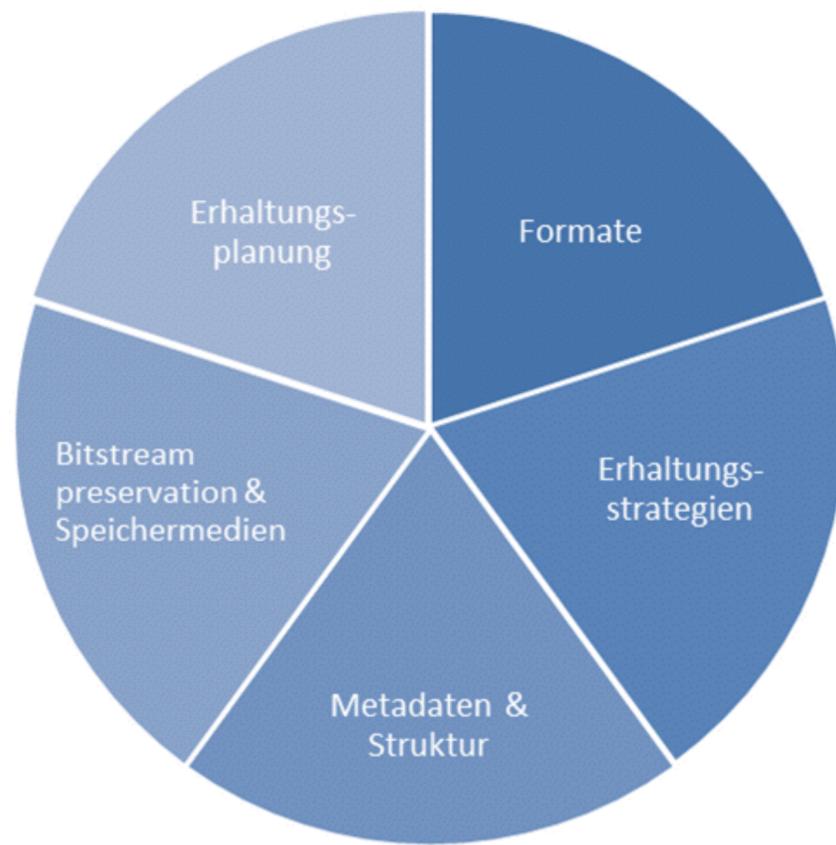
Gedächtnisorganisationen

- Bibliotheken
- Archive
- Museen

Aber auch

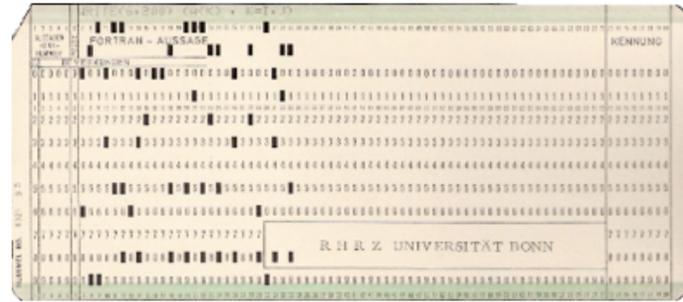
- Bildungseinrichtungen
- Forschungseinrichtungen
- Unternehmen
- Banken, Versicherungen

Dimensionen der LZA



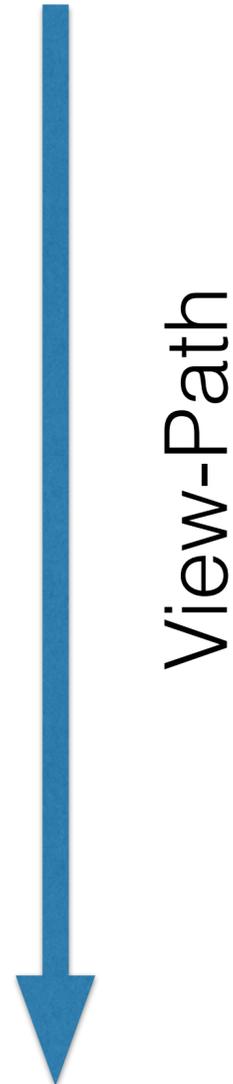
- Recht (UrhG, Berner Übereinkunft)
- Ökonomie (Kosten und Finanzierung)
- Organisation (Akquise, Ingest, Verantwortung, Archivinfrastruktur)
- Auswahlkriterien
- Medienspezifik (Film, Spiel, Soziale Netze)
- Technik: Bitstream Preservation, Speichermedien
- Pädagogik: Aus- und Fortbildung

Jürgen Keiper: *Überblick über die Expertendiskussion zur Nachhaltigkeit*



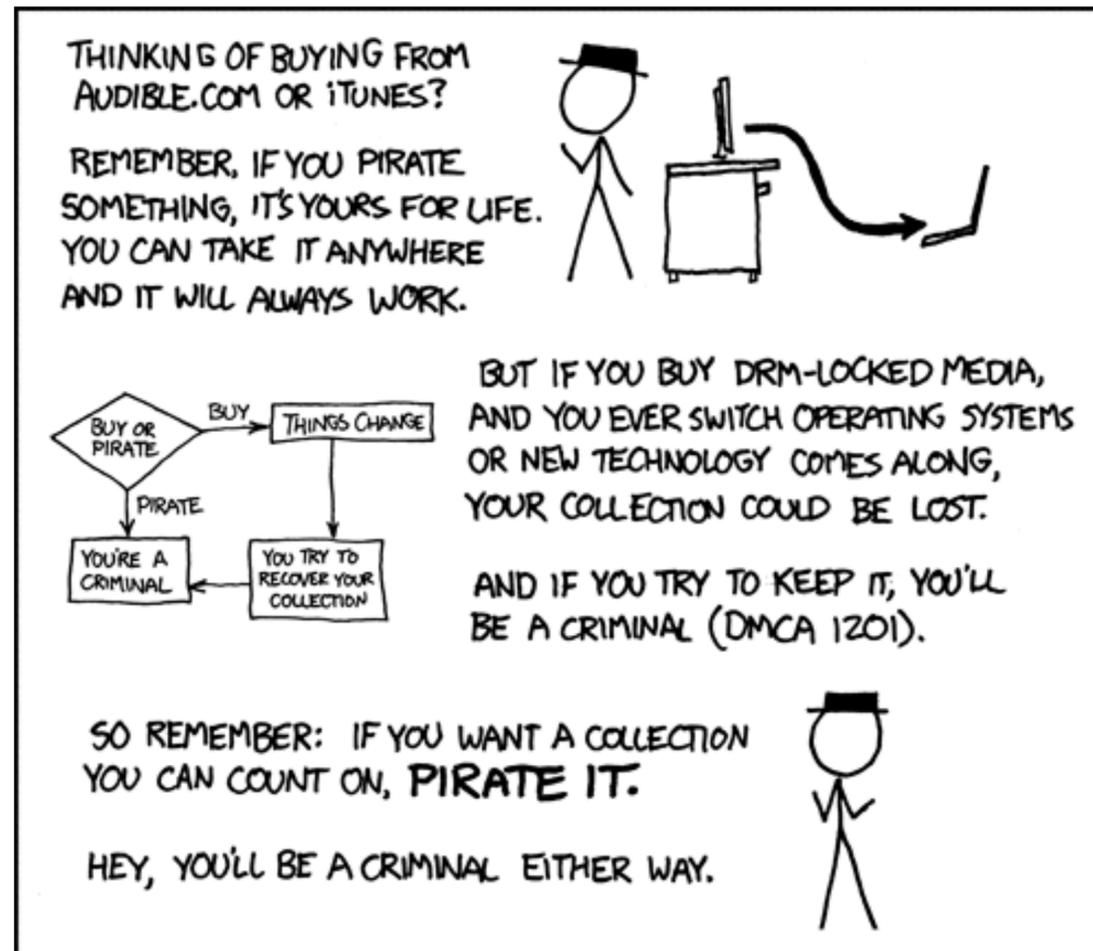
Preservation Layer

- Dateiformat
- Anwendung
- Betriebssystem
- Datenträger
- Hardware



Rechtliche Herausforderungen

http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_1.pdf

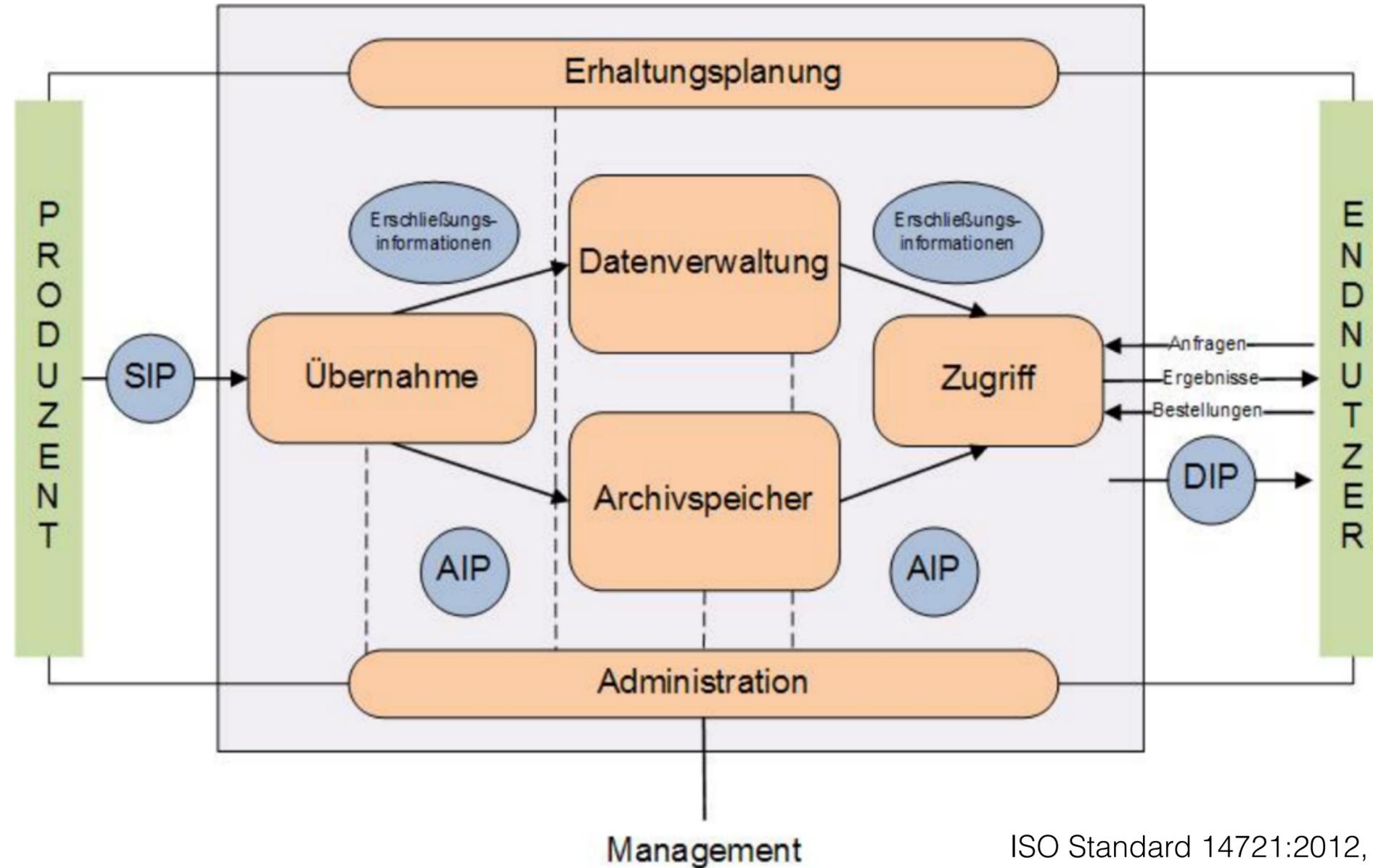


(IF YOU DON'T LIKE THIS, DEMAND DRM-FREE FILES)

<http://xkcd.com/488/>

- Rechte zur Durchführung notwendiger Eingriffe in die Gestalt der elektronischen Ressourcen im Interesse der Langzeiterhaltung (z.B. DNB)
- Einschränkungen durch Digital Rights Management Systeme (z. B. Kopierschutz)
- Konditionen des Zugriffs auf die archivierten Ressourcen und deren Nutzung (z.B. beim Cloud Computing)
- Haftung für die Inhalte archivierter Objekte (z.B. Äußerungsdelikte)

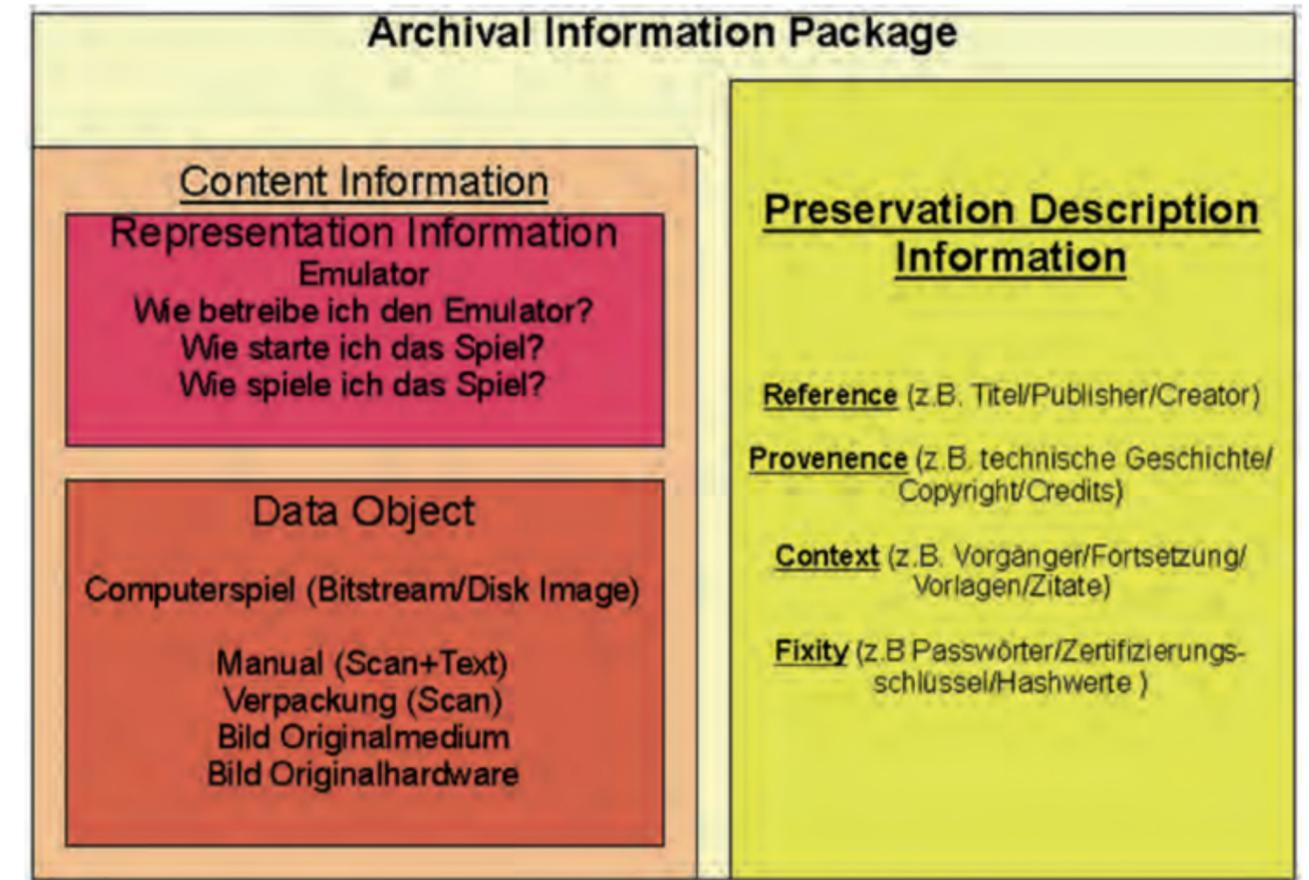
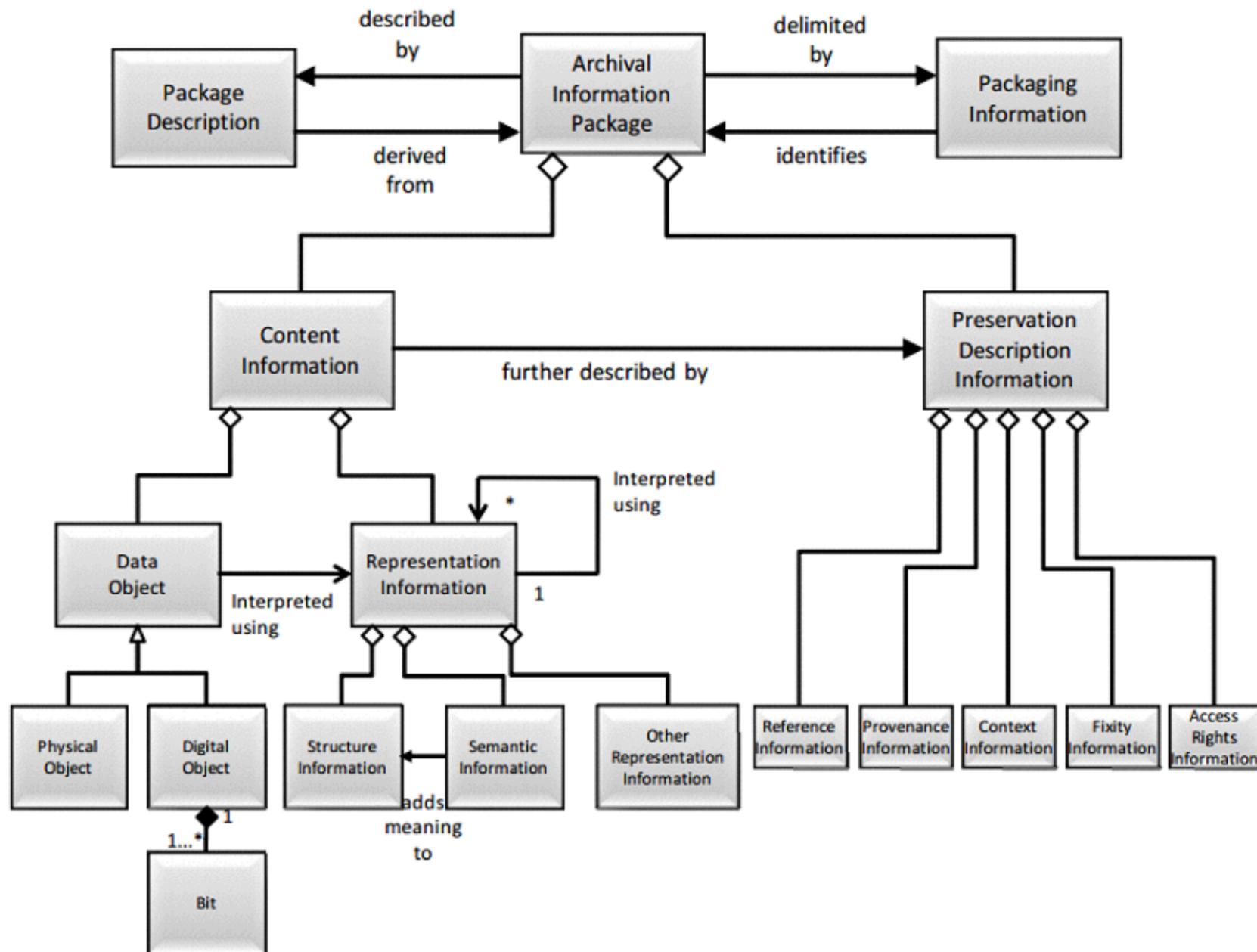
Archivmodell



ISO Standard 14721:2012, Referenzmodell

OAIIS (open archival Information system)

Archival Information Package



Huth; Lange: AIP für ein Computerspiel

Metadaten

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE mapping SYSTEM "C:\Documents and Settings\leirini\Desktop\crm_mapping.dtd">
<mapping>
  <map>
    <domain_map>
      <src_domain>DCCAP</src_domain>
      <src_domain_condition>If DC.Type = DCT collection</src_domain_condition>
      <target_domain>E78 Collection</target_domain>
      <target_domain_condition/>
    </domain_map>
    <combined_links joined_on="x2">
      <link_map>
        <range_map>
          <src_range>DC.Title</src_range>
          <src_range_condition/>
          <target_range>E35 Title</target_range>
          <target_range_condition/>
        </range_map>
        <path_map>
          <src_path_condition/>
          <src_path>has</src_path>
          <target_path_condition/>
          <target_path>
            <internal_link>P102 has title</internal_link>
          </target_path>
        </path_map>
      </link_map>
      <link_map>
        <range_map>
          <src_range>DC.TitleAlternative</src_range>
          <src_range_condition/>
          <target_range>E55 Type</target_range>
          <target_range_condition/>
        </range_map>
        <path_map>
          <src_path_condition/>
          <src_path/>
          <target_path_condition/>
          <target_path>
            <internal_link>P102 has title</internal_link>
            <internal_entity>E55 Title</internal_entity>
            <internal_link>P132 has alternative form</internal_link></target_path>
          </target_path>
        </path_map></link_map>
      </combined_links>
    </map>
  </mapping>
```

Administrativ: Name, Herkunft, Speicherort

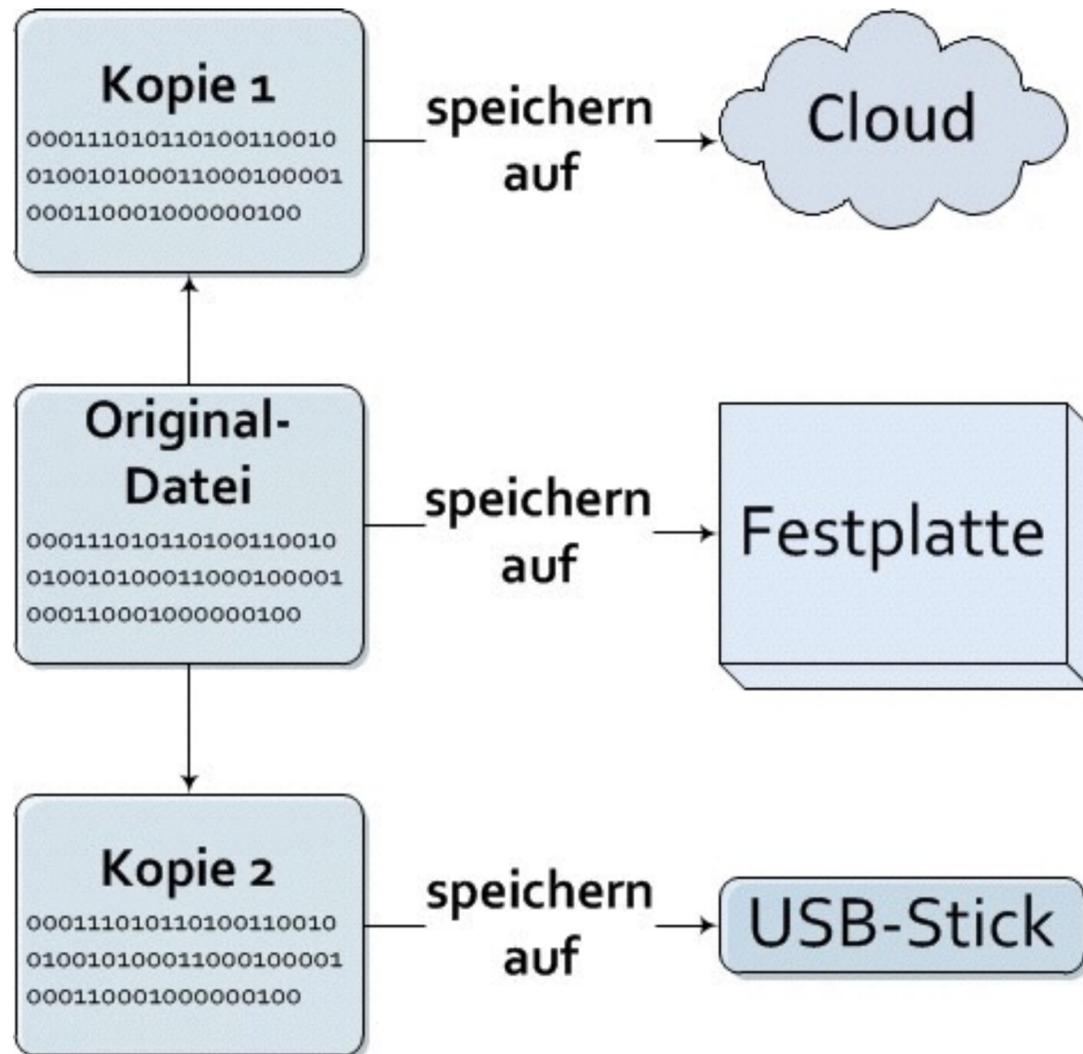
Deskriptiv: Inhalt

Technisch: Format, Umfang, Software

Strukturell: Beziehungen, Verknüpfungen

Formatspezifisch: Je nach Medientyp

Erhaltungsstrategien



- Analoge Sicherung
- Museumsansatz
- Migration
 - Umkopieren
 - Backup
- Emulation
- LOCKSS
- Privatarhive

Analoge Sicherung



Memory of Mankind

Weil wir keine haltbaren schriftlichen Aufzeichnungen mehr hervorbringen, werden Historiker in der Zukunft unser Zeitalter als das "finstere 21. Jahrhundert" bezeichnen.

- Computerausdruck
- Computer Output on Microfilm (COM)

Pro

lange Haltbarkeit, robust

Contra

aufwändig (einmalig), nur möglich für Text und Bilder, eingeschränkte Verfügbarkeit

Museumsansatz

- Keine Migration, Bewahrung der Inhalte in Ihrer originalen Umgebung
- Versuch der Aufbewahrung
 - Originalhardware, -software
 - Lesegeräte
 - Ein- und Ausgabegeräte

Pro

Höchstmaß an Authentizität

Contra

aufwändig, Zugang eingeschränkt
keine Ersatzteile

Migration

Auffrischung | Replikation | Umverpackung | Transformation

- Erhalt der Metadaten
- Umkodierung in aktuelle Speicherformate
 - verlustfreie vs. verlustbehaftete Konvertierung
 - sinnhafte Konvertierung digitale Artefakte

Pro

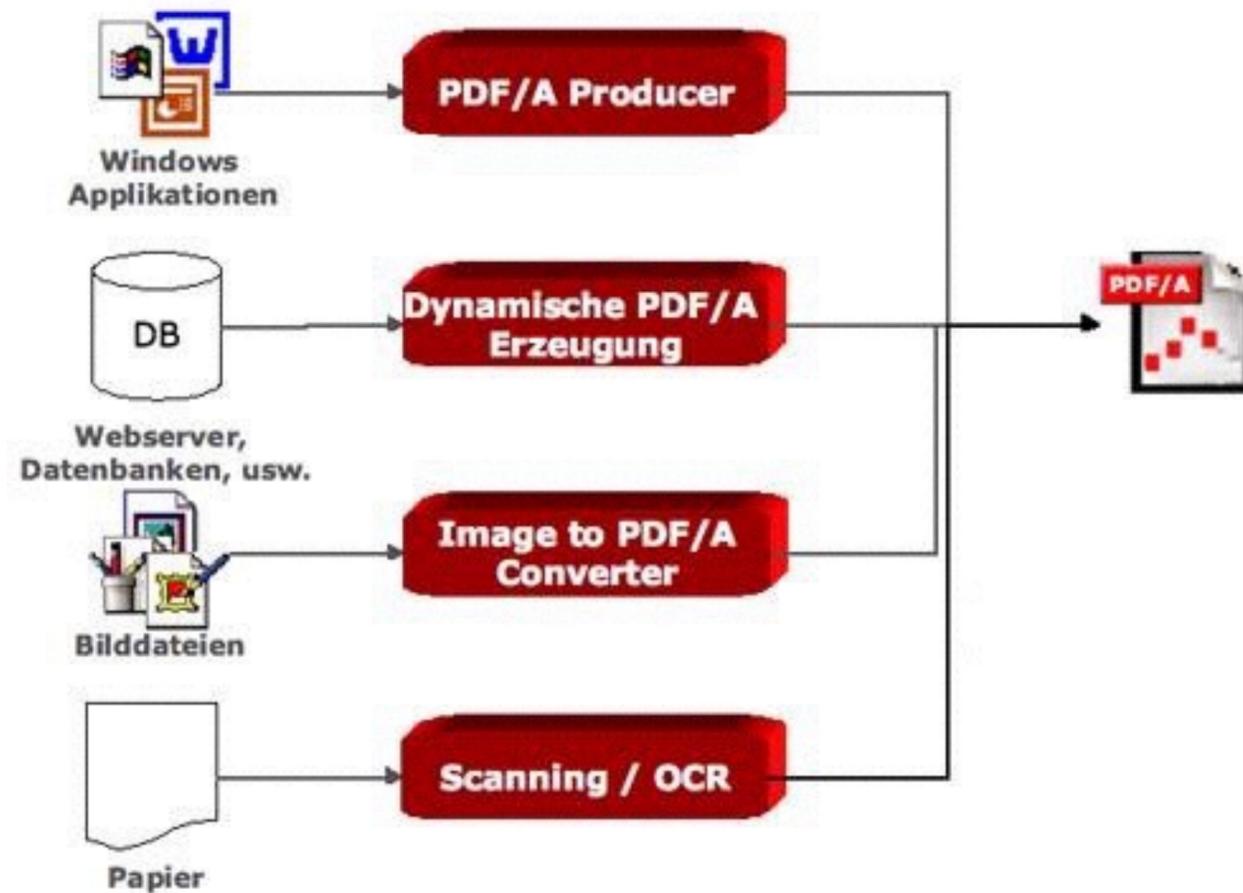
geringer Aufwand

Contra

Original wird verändert,
manuelle Nachbearbeitung sehr aufwändig
bei multimedialen Daten nicht sinnvoll anwendbar
nicht möglich bei proprietären Formaten

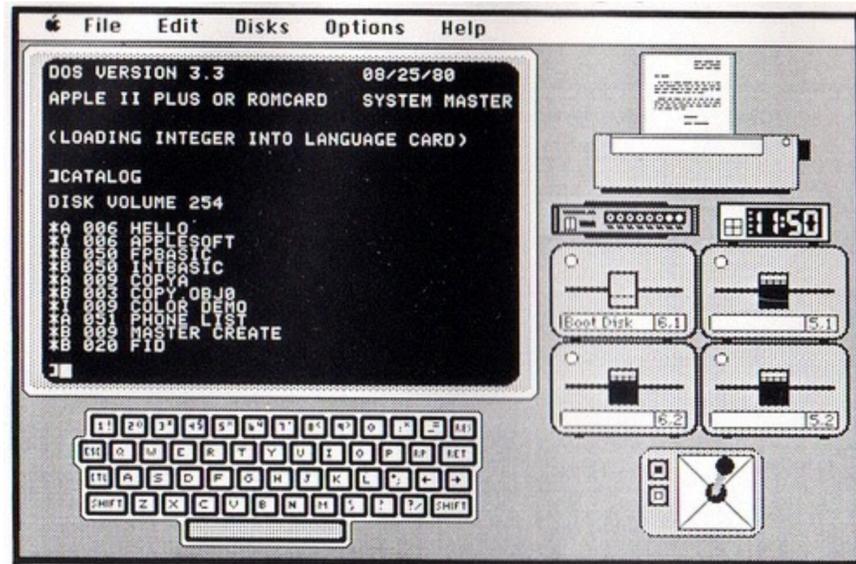
Archivstandards

- Wenige, bevorzugt offene Standards
- Archivfreundliche Formate
 - PDF/A (ISO 19005-1:2005)
 - XML
 - TIFF
 - AIFF



Emulation

- lat. *aemulari*: nachahmen
- Emulation der Hardware und/oder Software (API)
- zuerst 1962 bei IBM



Pro

originalgetreu, keine Konvertierung notwendig
erfolgreich im Spielebereich

Contra

sehr hoher Aufwand,
Emulator muss für neue Systeme angepasst werden



LOCKSS

- Lots Of Copies Keeps Stuff Save
- Peer-to-Peer Ansatz
- “Ein Mann-Bibliothek”
- OAI-kompatibel
- Migration verwendet

archive.org (Sue Me)



Internet Archive is a non-profit library of millions of free books, movies, software, music, and more.



8.4M



2.2M



2.6M



113K



1.1M



158K



178K



GO

[Advanced Search](#)

Announcements

Pro-Airbnb advertising dominated recent political TV ads in San Francisco

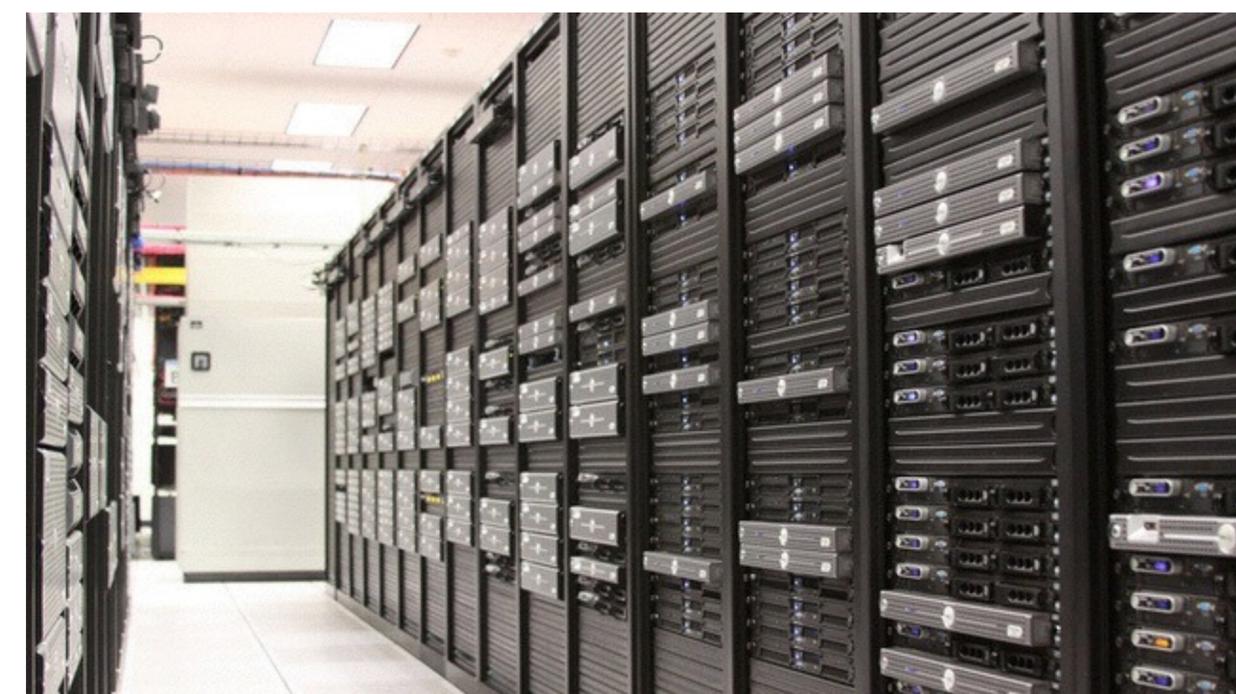
Aaron Swartz Day – Hackathon, Privacy-enabling conference and Reception

Grant to Develop the Next Generation Wayback Machine

[SEE MORE](#)

Kapazitäten

Wie viele Bits umfasst die Welt?



- 1 B
- 1 KB = 10^3 B
- 1 MB = 10^6 B
- 1 GB = 10^9 B
- 1 TB = 10^{12} B
- 1 PB = 10^{15} B
- 1 EB = 10^{18} B
- 1 ZB = 10^{21} B
- 1 YB = 10^{24} B
- 1 Buch: $10 \text{ MB} = 10 \cdot 10^6 \text{ B} = 10^7 \text{ B}$
- 1 CD: $100 \text{ MB} = 10^8 \text{ B}$
- 1 Film: $1 \text{ GB} = 10^9 \text{ B}$
- 1 Spiel: $500 \text{ MB} = 5 \cdot 10^6 \text{ B}$
- 1 Bild = 5 MB
- 1 Tag Radio: $1.440 \text{ MB} = 1.4 \text{ GB}$, 1 Jahr $\hat{=} 525 \text{ GB}$

130 Mio. Bücher (Google Books 2010)
 7,6 Mio CDs (Gracenote 2008) ~8 Mio.
 2 Mio Filme + TV (IMDB)
 75.000 Computerspiele (MobyGames 2013)
 100 Mio. Bilder, Pläne, Karten (Schätzung)

Buch: $130 \cdot 10^6 \cdot 10^7 \text{ B} = 1.3 \cdot 10^{15} \text{ B} = 1.3 \text{ PB}$
 CDs: $8 \cdot 10^6 \cdot 10^8 \text{ B} = 8 \cdot 10^{14} \text{ B} = 800 \text{ TB}$
 Film: $2 \cdot 10^6 \cdot 10^9 \text{ B} = 2 \cdot 10^{15} = 2 \text{ PB}$
 Spiel: $8 \cdot 10^4 \cdot 5 \cdot 10^8 \text{ B} = 40 \cdot 10^{12} = 4 \cdot 10^{13} = 40 \text{ TB}$
 Bild: $100 \cdot 10^6 \cdot 5 \cdot 10^6 \text{ B} = 10^8 \cdot 5 \cdot 10^6 = 5 \cdot 10^{14} = 500 \text{ TB}$

= 4,64 ~ 5 PB

156.250 €

Megaupload (19.01.2012): 25 PB



8 TB, ~250 €

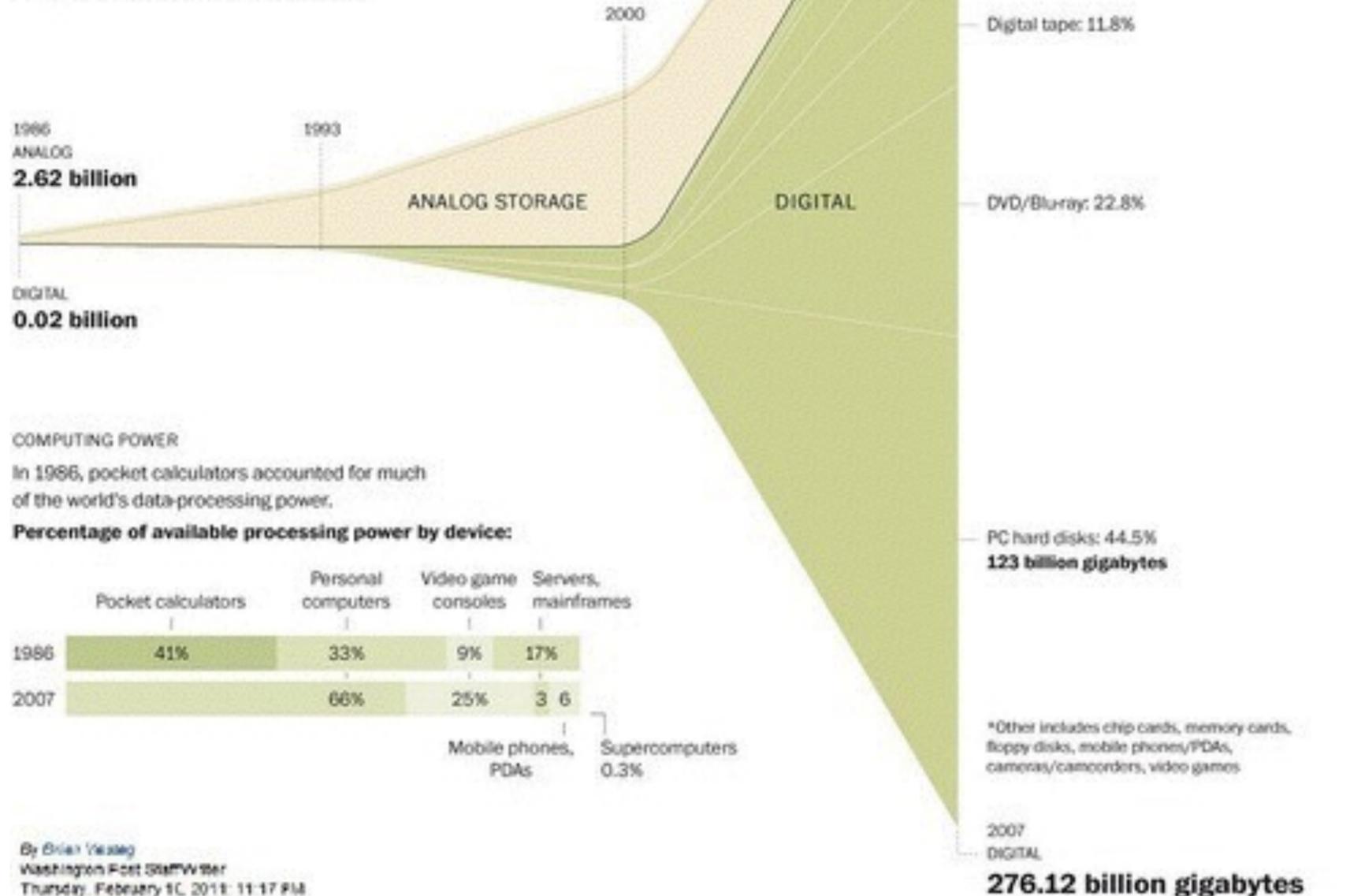
The Washington Post

Exabytes: Documenting the 'digital age' and huge growth in computing capacity

THE WORLD'S CAPACITY TO STORE INFORMATION

This chart shows the world's growth in storage capacity for both analog data (books, newspapers, videotapes, etc.) and digital (CDs, DVDs, computer hard drives, smartphone drives, etc.)

In gigabytes or estimated equivalent

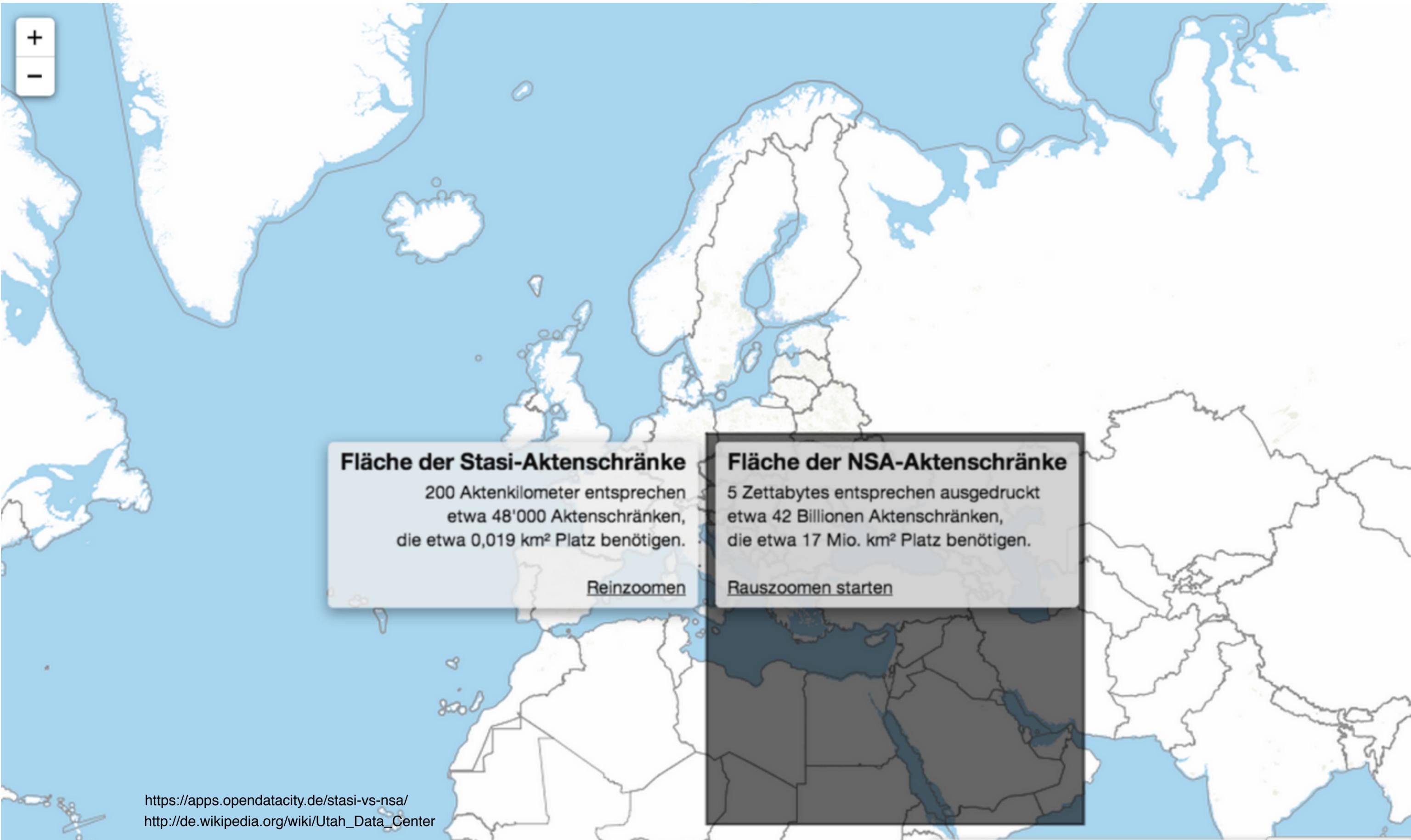


By Brian Yeasey
 Washington Post Staff Writer
 Thursday, February 10, 2011 11:17 PM

Weltinformation

In 2007, humankind was able to store some 300 exabytes of optimally compressed information in their technological devices.

<http://www.martinhilbert.net/WorldInfoCapacity.html>



Fläche der Stasi-Aktenschränke
200 Aktenkilometer entsprechen etwa 48'000 Aktenschränken, die etwa 0,019 km² Platz benötigen.
[Reinzoomen](#)

Fläche der NSA-Aktenschränke
5 Zettabytes entsprechen ausgedruckt etwa 42 Billionen Aktenschränken, die etwa 17 Mio. km² Platz benötigen.
[Rauszoomen starten](#)



Grand Challenge

Erhalt des digitalen Kulturerbes:
Sicherstellung der langfristigen
Bewahrung und des Zugangs zu den
digitalen und digitalisierten Artefakten
unserer Gesellschaft

Digitales Kulturerbe

Die Digitalisierung hat unsere Kultur tief durchdrungen. Musik, Videos, Fotos werden inzwischen überwiegend digital hergestellt und verbreitet, Bücher werden digitalisiert, Kommunikation findet über Handys, E-Mails oder Chats statt. Manches wie z. B. Computerspiele hat nicht einmal mehr eine Entsprechung in der analogen Welt. Wenn digitale Informationen nun nahezu von überall her und für jeden zugänglich sind, birgt dies zwar Chancen, stellt uns aber auch vor technische Herausforderungen: Wie lässt sich unsere digitale Kultur dauerhaft bewahren? Wie können virtuelle Objekte angemessen (re)präsentiert und zugänglich gemacht werden?

Wir benötigen Konzepte, um Kulturgüter auch für künftige Generationen begreifbar und erlebbar zu machen. Und ohne nachhaltige Langzeitbewahrung ist unsere digitale Kultur unwiederbringlich verloren. Nur mit Strategien zur Langzeitarchivierung können wir unser digitales Kulturerbe erhalten und ein „Zeitalter ohne Gedächtnis“ vermeiden.

GI-Fachgruppe Langzeitarchivierung

Gesellschaft
für Informatik



nestor 

www.digitale-bewahrung.de

Kontakt: Jens-Martin Loebel
loebel@uni-bayreuth.de

jochen.koubek@uni-bayreuth.de