

Nachhaltigkeit

Quelle:

Borghoff, Uwe M., Rödiger, Peter, Scheffczyk, Jan, Schmitz,
Lothar:

Langzeitarchivierung.

In: Informatik-Spektrum 28, Heft 6. Dezember 2005. Seite
489 - 492

Das Problem

- Früher hat man von Texten Papierkopien oder Mikrofilme gemacht, heute
 - macht man oft Computerkopien.
- Oft ist das Material aber auch
 - nur auf einem Computer vorhanden („Digitale Dokumente“),
 - wird ausschließlich auf Computer archiviert („Digitales Archiv“),
 - ein genuin computerbasiertes Objekt (Spreadsheets, Animation, Programm, Spiel etc.)
- Digitale Dokumente sind gegenüber Papier nur indirekt zugänglich und altern schnell. Direkt zugängliche (Höhlenmalereien, Bücher) altern relativ langsam und werden oft nur zufällig zerstört.

Beispiele

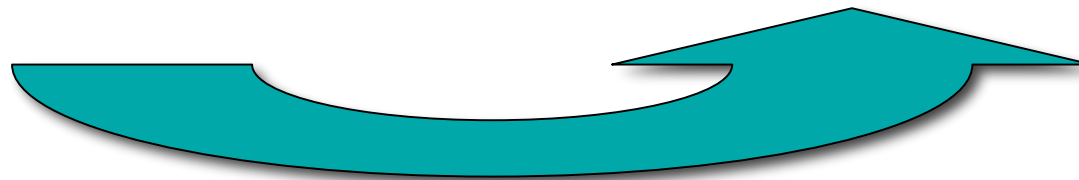
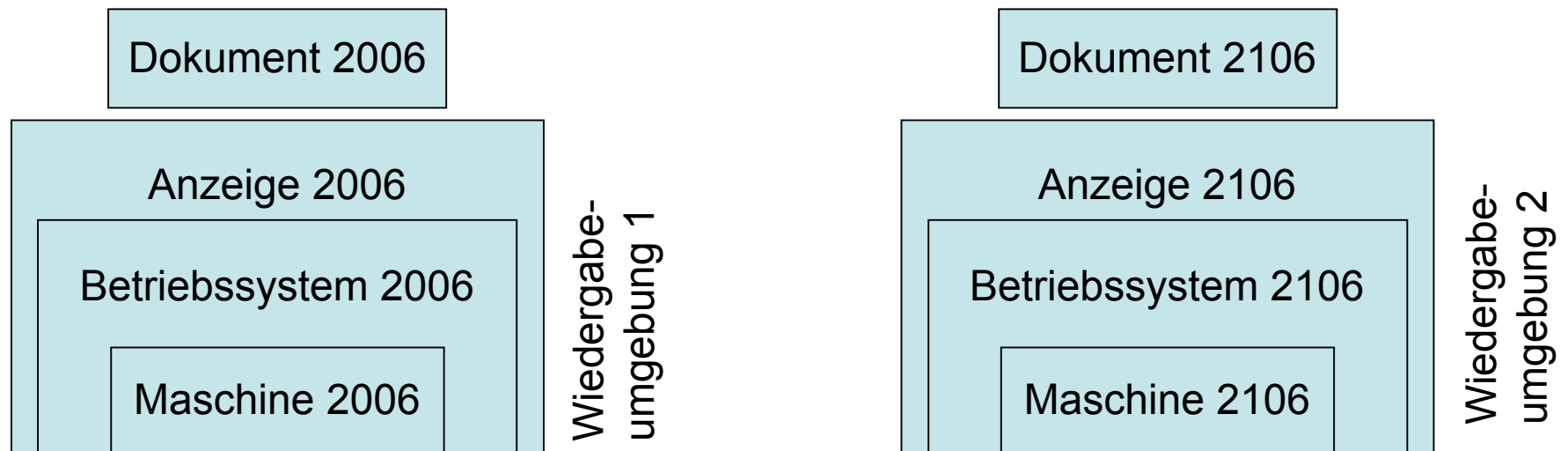
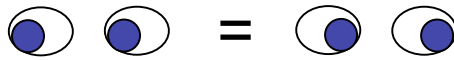
- Die digitalen Satelitenbilder der 70er Jahre über den Zustand des Regenwaldes sind heute nicht mehr lesbar und nicht mehr rekonstruierbar.
- Die Volkszählungsdaten von 1960 sind in den 70ern fast 10 Jahre lang rekonstruiert worden. 10.000 Datensätze sind für immer verloren.
- Eine CD kann Daten etwa 10 Jahre lang fehlerfrei speichern.

Erhalt eines digitalen Dokuments

Heißt

- Kopieren von Datenträger zu Datenträger zur Erhaltung der Bitfolge; ist technisch gelöst („Refreshing“)
- Erhaltung des Inhalts unabhängig von der Wahrnehmung („Inhaltmigration“)
- Erhaltung des Zugang im Sinne von identischer oder sehr ähnlicher Wahrnehmung.

Wahrnehmungserhaltung



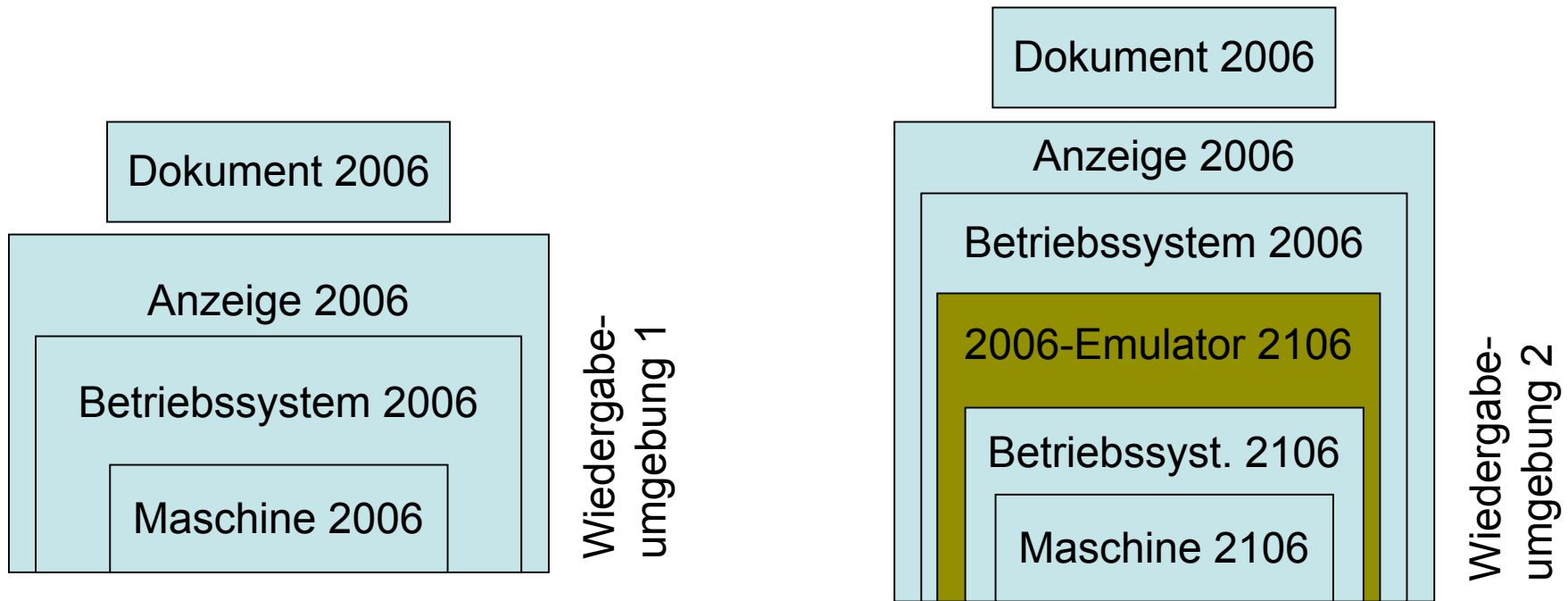
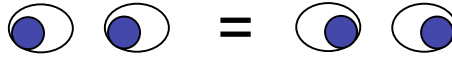
Randbedingungen

- Erhaltung (Erzeugung) von Metadaten,
- Möglichst automatische Migration,
- Authentisch, d.h. keine fehleranfälligen menschlichen Arbeitsschritte (Scannen, OCR, (überwachtes) Einlesen in neue Datenformate)
- Erzeugung von Preservation Metadata

Lösungsansätze

- Weiterhin *Papierkopien* einlagern. Deckt nur den deklarativen Anteil ab.
- *Museumsansatz*: Eine Maschine mit Betriebssystem, Programmiersprachen, Programmen und Daten wird betriebsbereit gehalten. Kaum durchführbar (Ersatzteile, Bedienungserhalt) und außerordentlich aufwändig im Betrieb
- *Emulationsansatz*: Software-Nachbau einer Maschine und eines Betriebssystems
- *Universal Virtual Machine-Ansatz*: Bau einer virtuellen Maschine, die alle existierenden Maschinen und Betriebssysteme simulieren kann.
- „*Stein von Rosetta*“-Ansatz: Man erhält die Bitfolge und erstellt eine exakte Spezifikation der Dokumentformate, Betriebssysteme und Maschinen. Erhaltungsaufwand gering, Rekonstruktionsaufwand riesig.

Emulationsansatz



Folgerungen im Kleinen

- Daten regelmäßig umkopieren, (Externe) Platten sind viel langlebiger als CDs,
- Herstellerunabhängige definierte Standards benutzen: ASCII, Unicode, SVG, XML, XSL, allenfalls Quasi-Standards wie PDF oder TIFF.
- Regelmäßig bei neuen Software-Versionen Umcodierung in kleinen Schritten, da diese eher formerhaltend sind.
- Metadaten in ASCII erzeugen über
 - Datenentstehung mit Autor, Datum und Datenentstehungsprozess,
 - Hardware mit Ausbau
 - Betriebssystem mit Version,
 - Softwareumgebung mit Version und dessen Handhabung,
 - Codierung,
 - Bericht über Erhaltungsmaßnahmen mit Datum, Aktion und Bearbeitern