

Repositorien – Zukunft der Bibliotheken oder Digitalmüll der Zukunft?

Prof. Dr. Gerhard Schneider
direktor@rz.uni-freiburg.de

Albert-Ludwigs-Universität Freiburg

**UNI
FREIBURG**

Repositorien

- Typische Repositorien der Vergangenheit:
 - Bibliotheken
 - Sammeln Bücher
 - Deutsches Rundfunkarchiv
 - Sammelt Schallplatten
 - Filmarchiv
 - Sammelt Filme
- Neuere Repositorien:
 - Wetteramt
 - Sammelt Wetterdaten
 - Nasa
 - Sammelt kosmische Daten
- Stets gilt: aufheben, vielleicht braucht es jemand

Repositorien

- Typische Anforderungen an Repositorien:
 - Seriös und langlebig
 - Schließlich will man auch in 100 Jahren etwas wiederfinden
 - saubere Datenhaltung und Auffindbarkeit der Inhalte
 - Nichts verändern, gut katalogisieren, Wahrnehmung steigern
 - Beratung
 - Wie gehe ich mit den Daten/Dingen/Inhalten um?
 - Wie steht es wirklich um die Langlebigkeit?
 - Finanzierung durch die öffentliche Hand
 - Oder Privatisierung?
 - Oder „Augen zu und durch“?
 - Kein Rezept hat bisher in jedem Fall funktioniert
 - Manche Repositorien sind auch einfach abgebrannt ☹

Neue Herausforderungen

- Alles wird digital
 - Auch die Bücher und (wissenschaftlichen) Zeitschriften
 - Schneller, ortsunabhängiger Zugriff
 - Es wird digital produziert –
 - Papier wird als Speichersystem zur Ausnahme
 - *Hilfe, mein Buch* – denken Sie an email: wann wurden die Zauderer bekehrt?
 - E-books – eine Gefahr für das Abendland?
 - Eher für die Wohnzimmermöbel-Industrie
- Vieles lässt sich gar nicht mehr auf Papier darstellen
 - Filme, Audio
 - Interpretationsspielraum beim Speichern von Musik auf Papier
 - Wörterbücher mit Suchfunktion, Datenbanken

Fehler

- Jeder weiss es am besten
 - Denn es gibt in Deutschland noch mehr IT-Experten als Fussball-Bundestrainer
 - Grandiose Entscheidungen: wir ernennen MS-Word zum Standard
 - Und vertrauen daher unsere Daten dem Programm einer Firma an – ohne zu wissen, wie es funktioniert
 - Man ignoriert Expertenmeinungen
- Machterhalt / Machtgewinn
 - „Sogenannte“ Standards durchsetzen, um an den Lizenzgebühren zu verdienen
 - Echte Standards kann jeder nutzen
 - Oder zahlen wir Lizenzgebühren dafür, dass die Bahnschienen 1,458m breit sind?
 - Warum wurde mp3 so erfolgreich?
 - Abhängigkeiten schaffen

Folgen

- Digitale Inhalte kann man nicht so einfach lesen – man braucht ein Darstellungsprogramm
 - Ok, einfache Dinge würde man noch hinbekommen:
ASCII-Code: 0100 1001 0100 0010 0100 1101 = IBM
 - Immer notwendig: die Art, wie Inhalte binär dargestellt werden, muss bekannt sein.
 - Sonst bedeutet: „out of print“ = Datenverlust
- „out of print“:
 - Klassisch: der Verleger verabschiedet sich, das Buch wird nicht mehr gedruckt – in der Bibliothek (dem Repository) liegt aber noch ein Exemplar
 - Digital: Die Softwarefirma verschwindet, und mit ihr das Darstellungsprogramm
 - Niemand kann/darf mehr die Inhalte lesen, die vielleicht in einer Bibliothek liegen

Folgen

- Das ist möglicherweise überraschend...
- erscheint doch die verschwindende Hardware vielen als das größere Problem 😊
 - Alte Floppies kann niemand mehr lesen..
 - Tonbandgeräte/Plattenspieler existieren nicht mehr
- Dieses Problem lässt sich jedoch lösen
 - Jeder digitale Speicher enthält „nur“ bits (0,1)
 - Diese kann man auslesen und auf einem neuen Speicher ohne Informationsverlust wieder abspeichern
 - Man kopiert automatisiert von einem Träger auf den nächsten
 - Und muss nur die Einträge im Katalog aktualisieren
 - Das können Roboter, wenn man recht bald den richtigen Träger wählt.
 - Achtung: das kann dauern...
 - Bei 50 Mbyte/s dauert es 20000 sec = ca 6h, um 1 TB zu kopieren.
 - Das Freiburger Uni-Bandarchiv hat 2500 TB 😊



Folgen

- Das funktioniert aber nur, wenn man umkopieren darf
 - Kopierschutz verhindert also eine langfristige Speicherung
 - Kopiergeschützte Materialien sind somit keine Kulturgüter
 - Hat da irgendjemand daran gedacht, als die Diskussion um Kopierschutz und dessen Verankerung im Urheberrecht brandete?
 - „Eigentümersicht“: mit jeder neuen Technologie kann der Inhalt wieder neu verkauft werden
 - Wie schon beim Wechsel von Schallplatte zu CD
 - Lizenz zum Gelddrucken
 - „Konsumentensicht“: wer moderne Technologie liebt, wird gemolken.
- Wer Daten besitzt, besitzt Macht:
 - ... kontrolliert den Zugang/Zugriff
 - ... erschließt sich neue Geldquellen
 - Nur noch reiche „xyz“ dürfen die Deutsche Bibliothek nutzen?
 - ... kann die Informationen verändern, die gespeichert sind

Folgen

- Welche Anforderungen sind an Repositorien zu richten?
 - Den Fachleuten sind diese schon klar:
 - Ungehinderter, nicht diskriminierender Zugriff
 - Manipulationen ausschließen
 - Verteilung auf viele Institutionen, welche unterschiedlichen Interessen gehorchen müssen
 - Aber wie verhindert man, dass Lobbygruppen und Politik auf die falsche Spur geraten?
 - und nach einem Terroranschlag umgehend den Zugriff auf Chemieinformationen verbieten
 - Was würde geschehen, wenn man in einer solchen Situation alle Bibliotheken anweisen würde, ihre Chemiebücher zu verbrennen?
 - Oder aus Sparsamkeit das „Zentralarchiv“ fordern?
- Sind diese Anforderungen neu?
 - Sie galten schon immer und wurden erst nach vielen Kämpfen von der Gesellschaft verstanden

Euphorie – und die Folgefehler

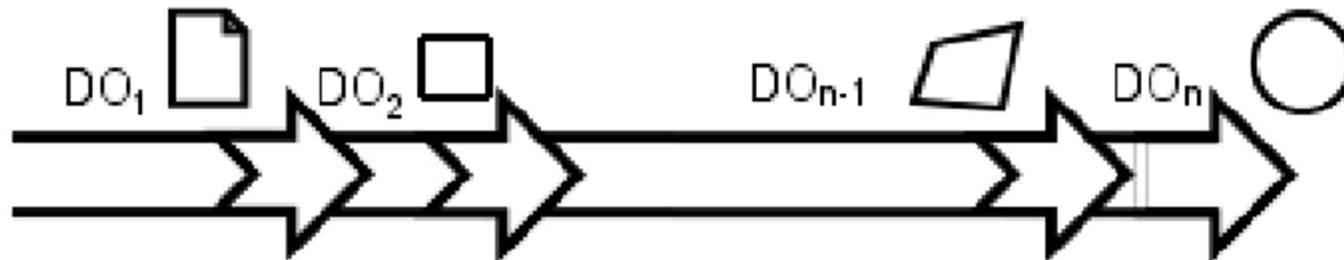
- Digitale Informationen stehen nun dank Internet weltweit zur Verfügung
- Immer mehr Wissenschaftler (u.a.) wollen nun wichtige Dinge digitalisieren
 - Weil die digitalen Quellen verloren gegangen sind – oder nie existierten
 - Es spart viele Reisen und erlaubt den direkten Zugriff auf forschungsrelevante Daten
 - Sie kommen mit dem Urheberrecht in Konflikt
 - Nur „alte“ Dinge darf man u.U. digitalisieren
 - Massive Probleme beim e-Learning!
 - Und andere setzen die Standards...
 - Stört es Sie, wenn Sie ein digitalisiertes Werk „downloaden“, dass es von der Library of Congress stammt
 - Und nicht etwa von der Deutschen Nationalbibliothek?
 - Verzerrte Wahrnehmung der Bestände führt zu verzerrten Aussagen

Euphorie – und die Folgefehler

- Weitere Fehlerquelle: der falsche Ansatz
 - das Rad wird neu erfunden
- Antrag auf Mittel zur Digitalisierung enthält typischerweise:
 - Geräte zur Digitalisierung
 - Entwicklungsgelder für ein Portal
 - Natürlich anders als die existierenden 400 Portale
 - Geld für Hardware zur Anschaffung eines Servers
 - Und was ist in fünf Jahren????
 - Ruf nach einem Datengrab = Repositorium
- Übergabe an ein Repositorium gelingt nur, wenn Standards eingehalten werden
 - Sonst wird es zu teuer / nicht wartbar.
 - Natürlich kann es kein Repositorium für alles geben
 - Spezialisierung ist unvermeidlich

Heile Welt?

- Angenommen, es ist alles im Griff:
 - Repositorien nehmen die Daten auf, erschließen sie und halten sie zum Abruf bereit – **für immer**
 - Repositorien sorgen automatisiert dafür, dass die Daten auf aktuelle Träger umkopiert werden
- Ist dann die Nutzbarkeit der Daten sichergestellt?
 - Kann word2040 noch die .doc-Formate aus dem Jahr 2003 lesen?
 - Nun – man könnte doch umspeichern: 2007 ruft man das Dokument mit Word2007 auf und speichert es im richtigen Format ab, das man 2010 mit Word2010 aufruft, um...



- Bleibt das Dokument dabei authentisch???

Heile Welt?

- Viele Daten brauchen die Laufzeitumgebung, in der sie entstanden sind
 - Auch nicht neu:
 - 3D-Bücher benötigen eine rot/grün-Brille
 - Der geschichtlich/soziale Kontext einer Aussage muss verstanden sein
Unschärfen durch Informationsverluste sind akzeptiert
 - Dokumente benötigen das Textbearbeitungsprogramm der damaligen Zeit
 - Konstruktionspläne benötigen die Darstellungssoftware aus der Vergangenheit
 - Wann kam der A320 auf den Markt? Wie sah die IT-Welt aus?
 - Digital gespeicherte Hologramme (3D-Bilder) benötigen eine ganze „Wiedererweckungs-“Infrastruktur.
- Manche modernen Dinge sind gar nicht klassisch speicherbar:
 - Interaktion: Computerspiele, Simulationsmodelle, Lernumgebungen
 - Sie sind mit ihrer Laufzeitumgebung „siamesisch“ verbunden

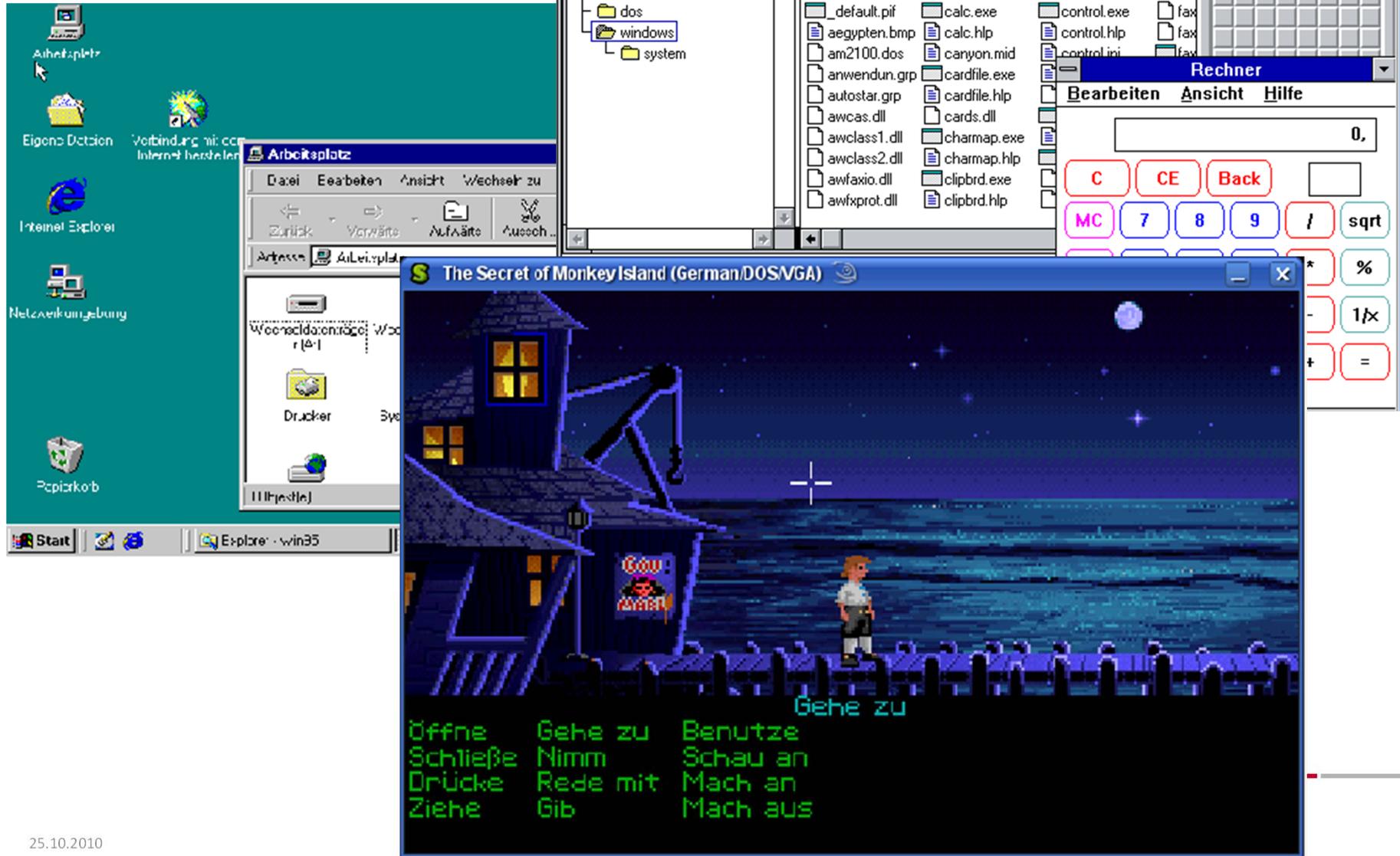
Heile Welt?

- Andererseits ändert sich die IT-Infrastruktur und lässt Generationen von Hardware hinter sich
 - Wer kennt den Motorola 68000?
 - Wer erinnert sich, wie oft z.B. Apple die Hardwareplattform geändert hat, so dass manche Software nicht mehr ausgeführt werden konnte – „schade um die Inhalte – das Leben ist hart“
 - Verluste sind der IT-Industrie egal – wie den Verlagen das „out of print“
- „Lösungen“:
 - Ein Hard- und Softwaremuseum?
 - Kann man kaputte IT-Bausteine in 100 Jahren überhaupt nachbauen?
 - NASA zahlte vor einigen Jahren Höchstpreise für uralte PentiumIII-Prozessoren!
 - Kann man die Kombination dann auch noch bedienen?
 - Wer versteht heute wirklich eine Dampflokomotive? Eine Ritterrüstung?

Heile Welt?

- „Lösungen“:
 - Museen vermitteln nur einen Anblick, selten einen Einblick in die Funktion
 - Schon mal ein Schwert in Händen gehalten? Wie führt man es?
- Lösungsansatz: **Emulation**
 - Man kann Hardware mit ihrer ganzen Funktionalität in Software abbilden!
 - Grandios für Entwickler:
Sie brauchen nicht 10 Smartphones, sondern nur einen Rechner, auf dem 10 „virtuelle Handys“ laufen
 - Die virtuellen Smartphones funktionieren genauso wie die realen, kein technischer Unterschied
 - Lediglich die Geschwindigkeit der Ausführung
 - Also kann man „alte Umgebungen“ in Software wiedererwecken
 - wenn

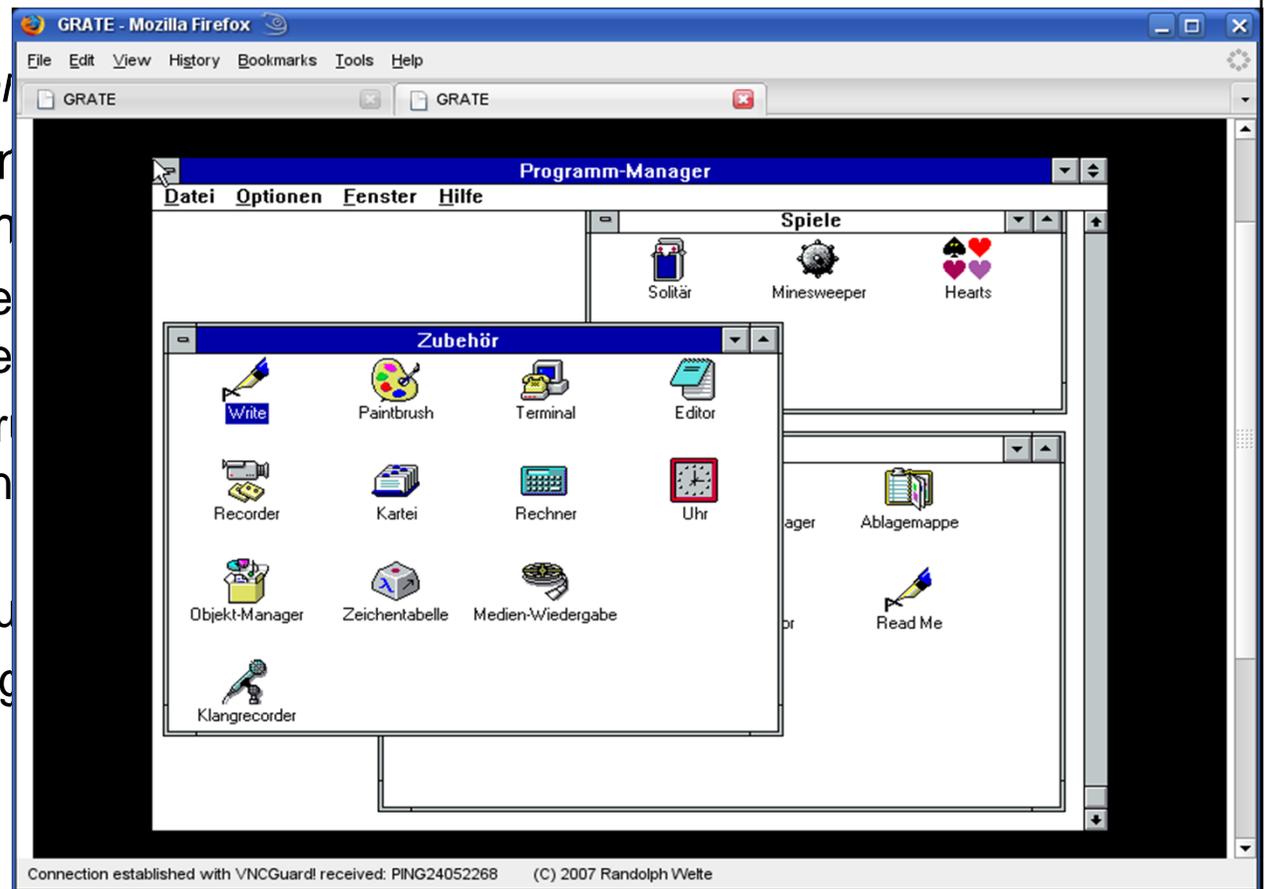
Beispiele:



Und wie geht es weiter?



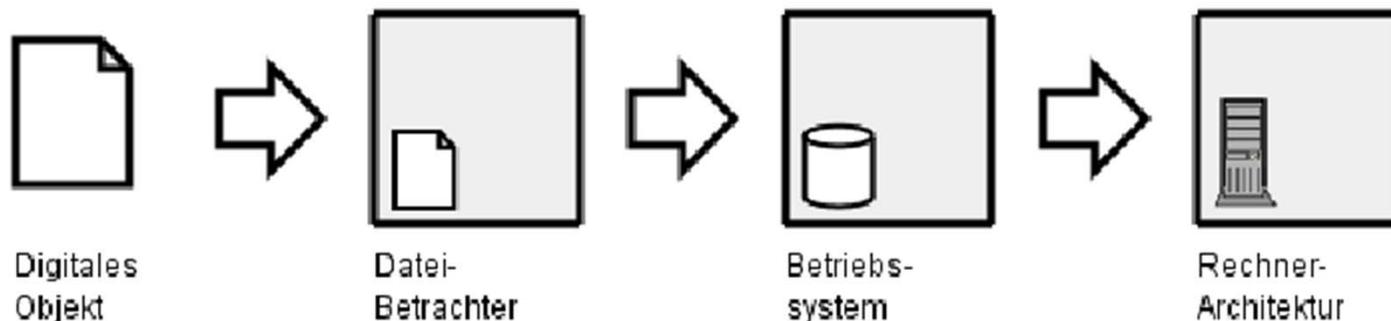
- Forschung: EU-Projekt „Planets“
Preservation and Long-term Access through Networked Services
<http://www.planets-project.org>
- Ziel (u.a.): Entwicklung von Mechanismen der Entdeckung und Nutzung von digitalen Ressourcen
 - D.h. Nutzer findet ein „lebendiges digitales Objekt“
 - Automatischer Aufruf (OPAC), Ausführung des Nutzers
 - Einfacher Zugriff auf die Ressourcen
 - Ausführen des Programms



Die Aufgaben:

View-Path

- *View-Path* als Modell zur Ermittlung notwendiger Sekundärobjekte
 - Primärobjekt kann Applikation erfordern
 - Applikation benötigt Ablaufumgebung in Form eines Betriebssystems (typischerweise)
 - Betriebssystem erfordert Hardwareumgebung
 - Emulator läuft in bestimmter Referenzumgebung



Emulatoren

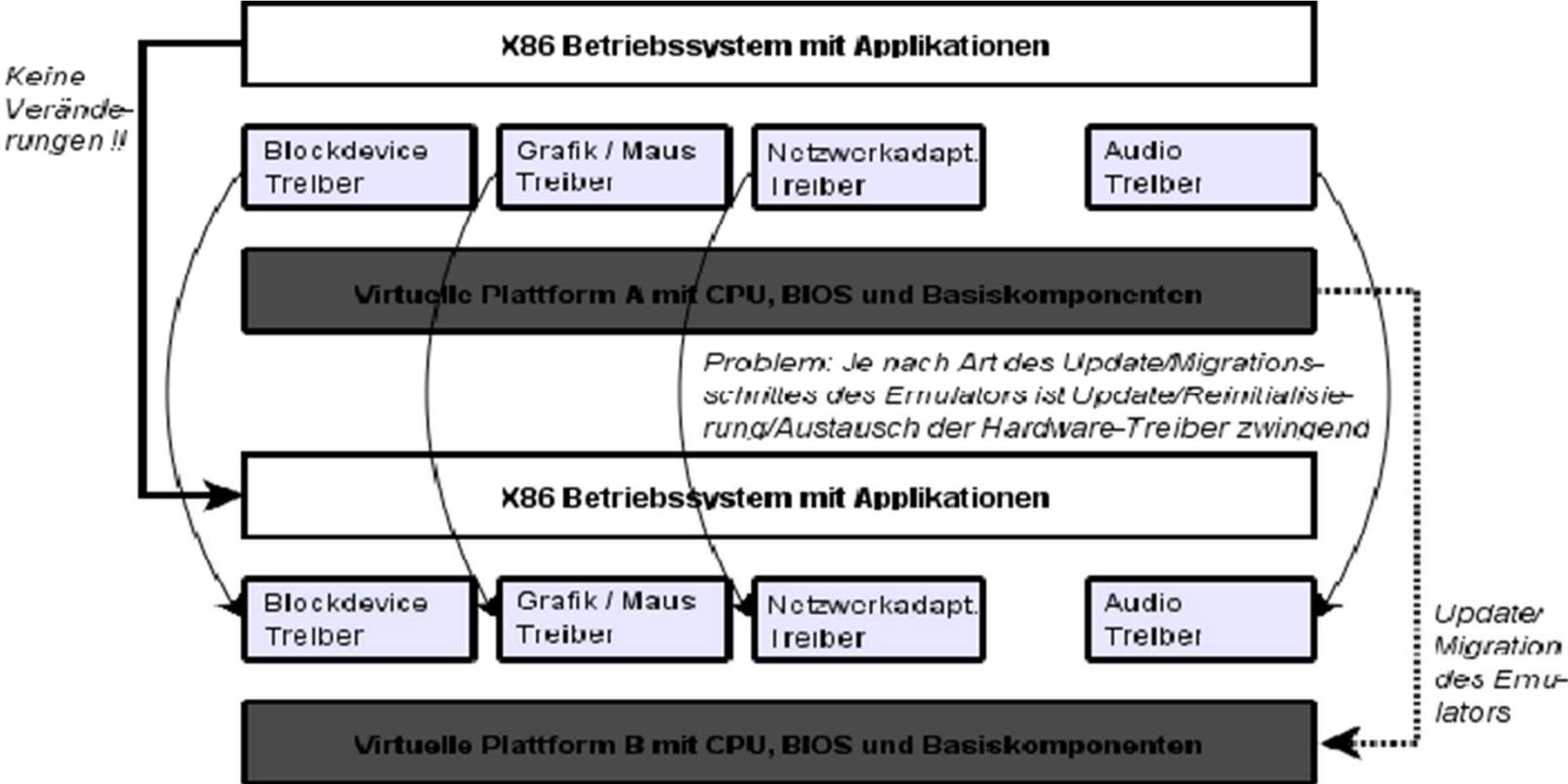
- Sie übernehmen die Brückenfunktion zwischen wiederherzustellender und aktueller Umgebung
- Sie bestehen selbst wieder aus Software
- Und stellen damit Anforderungen an ihre Umgebung als Applikation auf der aktuellen Umgebung (OS/Hardware)
- Übersetzung von Ein- und Ausgaben
- Datenaustausch zwischen “innen” und “außen”

- Konsequenz:
 - Regelmäßige Aktualisierung unumgänglich
 - Migration, modulare Emulation,
 - Möglicherweise sogar Schachtelung:
 - Wenn es für die aktuelle Umgebung keinen Emulator für die uralte Umgebung mehr gibt

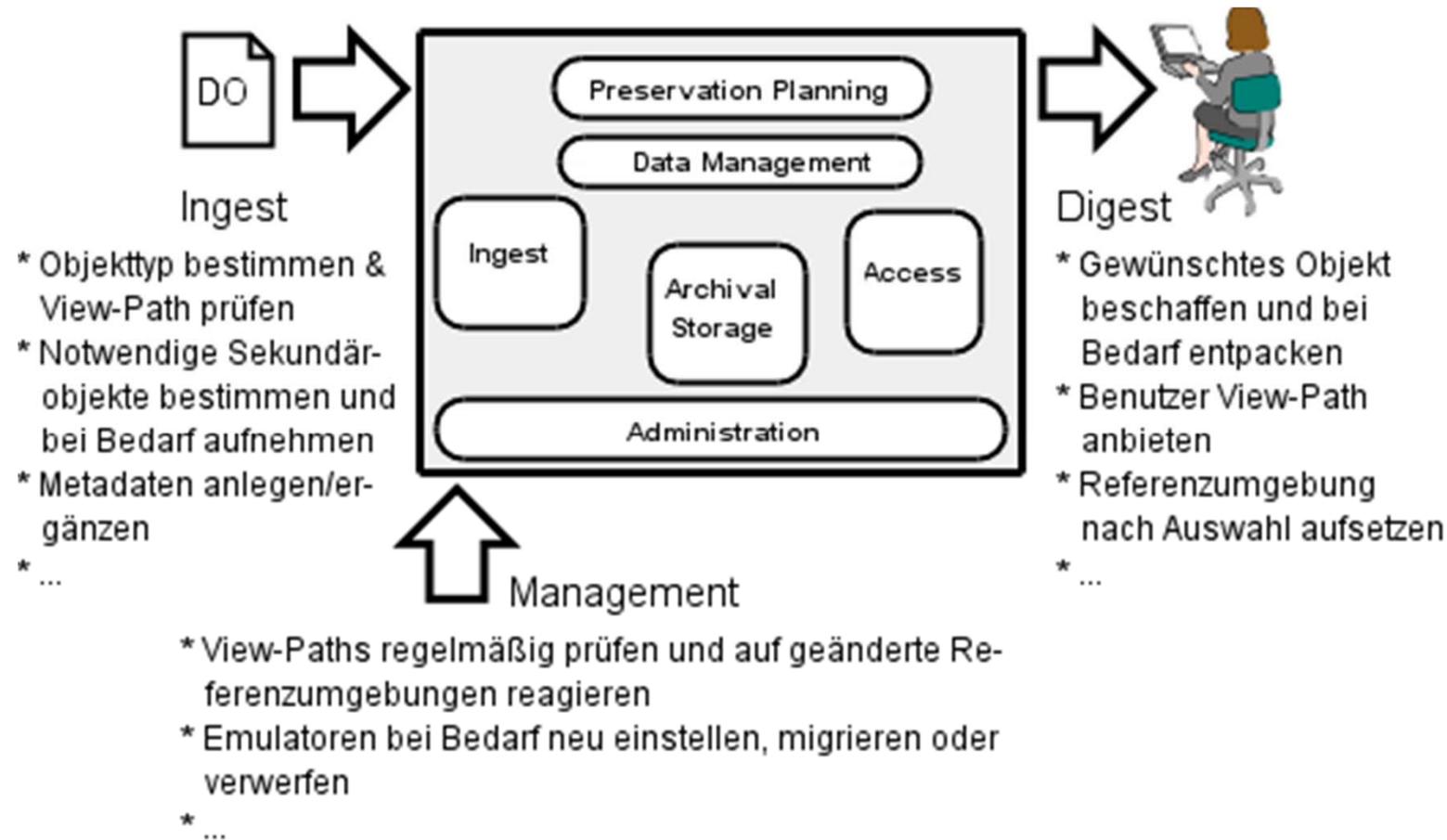
Emulatoren

- Ein planvolles Vorgehen wird erforderlich:
 - was benötigt eine Softwareumgebung, damit sie in Emulatoren ablaufen kann?
 - genaue Trennung der jeweiligen Anforderungen
 - Genaue Dokumentation der Installation und aller Aktionen bis hin zur Lauffähigkeit
- Dies muss in den Metadaten erfasst werden
 - Dann kann später ein Abbild der Softwareumgebung in einem neuen Emulator für die aktuelle Umgebung generiert werden
 - Oder man entscheidet sich, den alten Emulator zu portieren
 - Oder man läßt den alten Emulator in einer Emulation der alten Umgebung ablaufen
 - Diese Entscheidungen können automatisiert ablaufen und ggfls immer neu „ab initio“ entschieden werden.
 - Letztlich muss das jeweils nur eine Institution in der Welt durchführen

Herausforderungen: Teilupdates

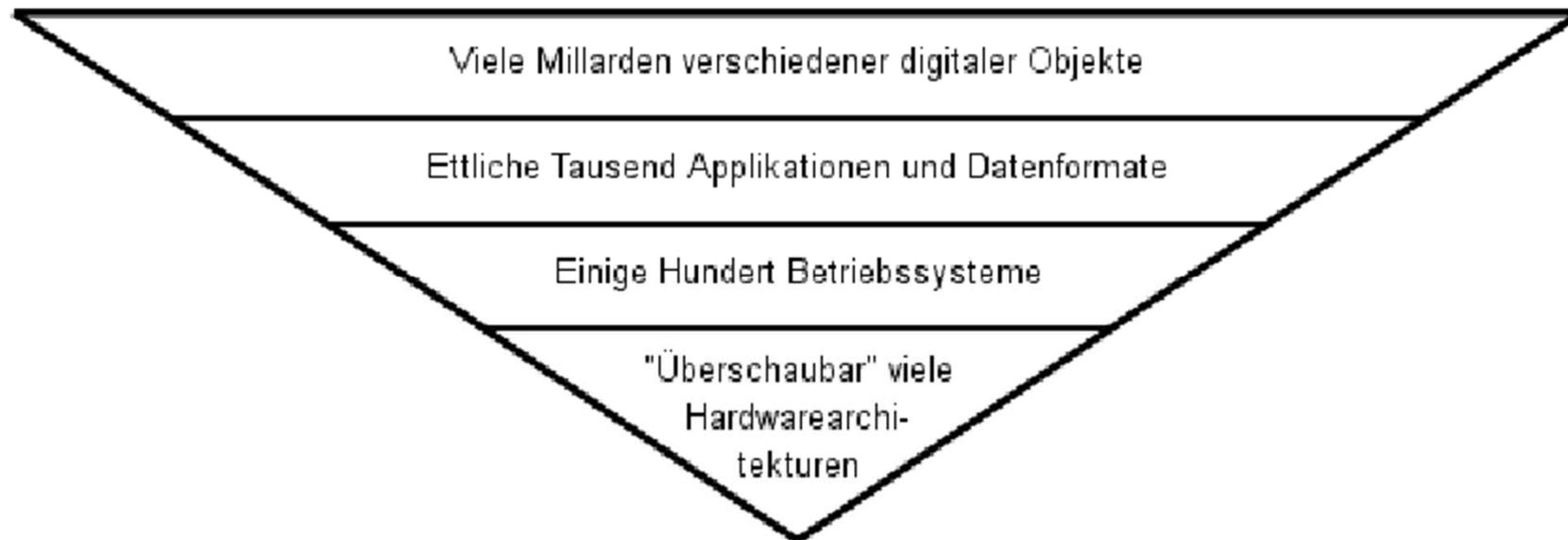


Repositorien-Management



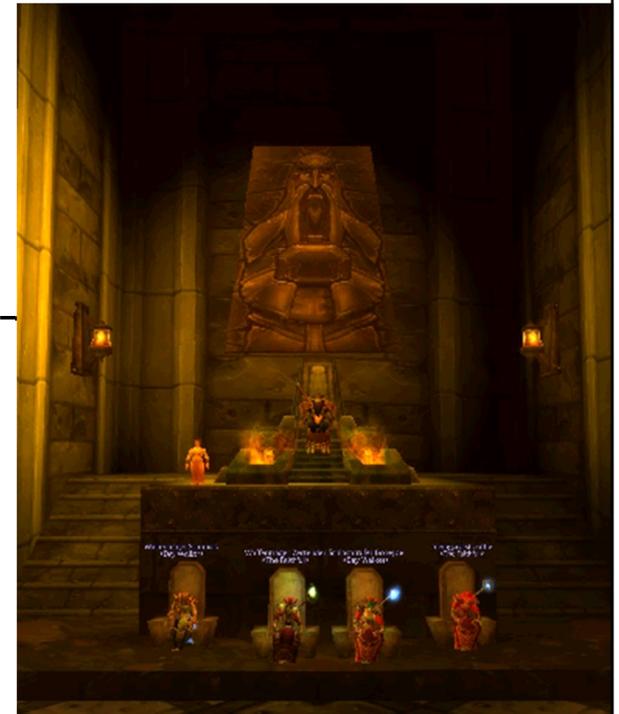
Alternativen?

- Zweifellos kein einfacher Prozess
 - Dennoch ein sinnvoller Weg, vermutlich der einzig gangbare
 - Angesichts der komplexen Welt.... ☹
 - Jeder Tag, der vergeht, ohne dass man damit beschäftigt, vergrößert den Abstand – und erhöht die Chance des Datenverlusts



Grenzen

- Zunehmender Abstand zwischen Ursprungs- und Referenzumgebung
- Digital Rights Management
- Neue Typen web-basierter Applikationen
- Massive Multiplayer Online Games
 - Hier kommt noch das Verhalten der Nutzer hinzu – man braucht viele Nutzer, um den Ablauf richtig verstehen zu können
- Ergänzende Herausforderung: Mitschneiden exemplarischen Nutzerverhaltens
 - Folge: noch weitere Herausforderungen an die Langzeitarchivierung
- Selbst wenn man „Ballerspiele“ nicht als aufbewahrungswürdig betrachtet – sie sind immerhin ein Beispiel für das Gesamtproblem



Zukunft

- Ohne Standardisierungen und ein geplantes Vorgehen werden viele Digitalisate untergehen
 - Denn sie brauchen eine Umgebung, in der sie ausgeführt werden
 - „tote Daten“ sind nicht so kritisch
 - JPG, MP3, ...
- Repositorien, die ihre Daten in Zukunft nicht mehr ablauffähig anbieten können, werden verschwinden.
- Einrichtungen ohne „ablauffähige“ digitale Daten werden zu Museen verkommen
- Eine renommierte Einrichtung hingegen wird noch mehr Renommé sammeln
 - Auch wenn es weit weg vom alten Berufsbild ist
 - Wer hat, dem wird gegeben
 - Und es ist klar, was zu tun ist! Aber man muss es tun.

Zukunft

- Die erfolgreichen Repositorien der Zukunft müssen Standards gesetzt haben
 - Nur dann ist die auf sie zukommende Arbeit überhaupt bewältigbar
 - Übrigens: auch der Erhalt von Büchern kostet Geld – prüfen Sie mal den Bibliotheksetat
 - Und alle sog. „Experten“ müssen sich an die Standards halten – sonst verschwinden die Daten wieder
 - Digitalisierungsprojekte in Zukunft wohl nur mit dem Repositoryum des eigenen Vertrauens
- Wer soll es machen?
 - Die Einrichtungen, die in der Vergangenheit schon einen entsprechenden Ruf aufgebaut haben
- Wer die Arbeit nicht übernimmt, verschwindet

Jeff Rothenberg: *“Digital information lasts forever, or five years, whichever comes first.”*

Zeigen wir, dass er Unrecht hat – yes, we can 😊

Mitwirkende = think tank:

Dr. Dirk v. Suchodoletz

Dr. Randolph Welte

Dipl-Inf. Klaus Rechert

u.a.