

Technische Aspekte der Langzeitarchivierung

Matthias Razum

Dr. Matthias Hahn

Fortbildungsveranstaltung der AjBD
Karlsruhe, 25.11.2011



Agenda

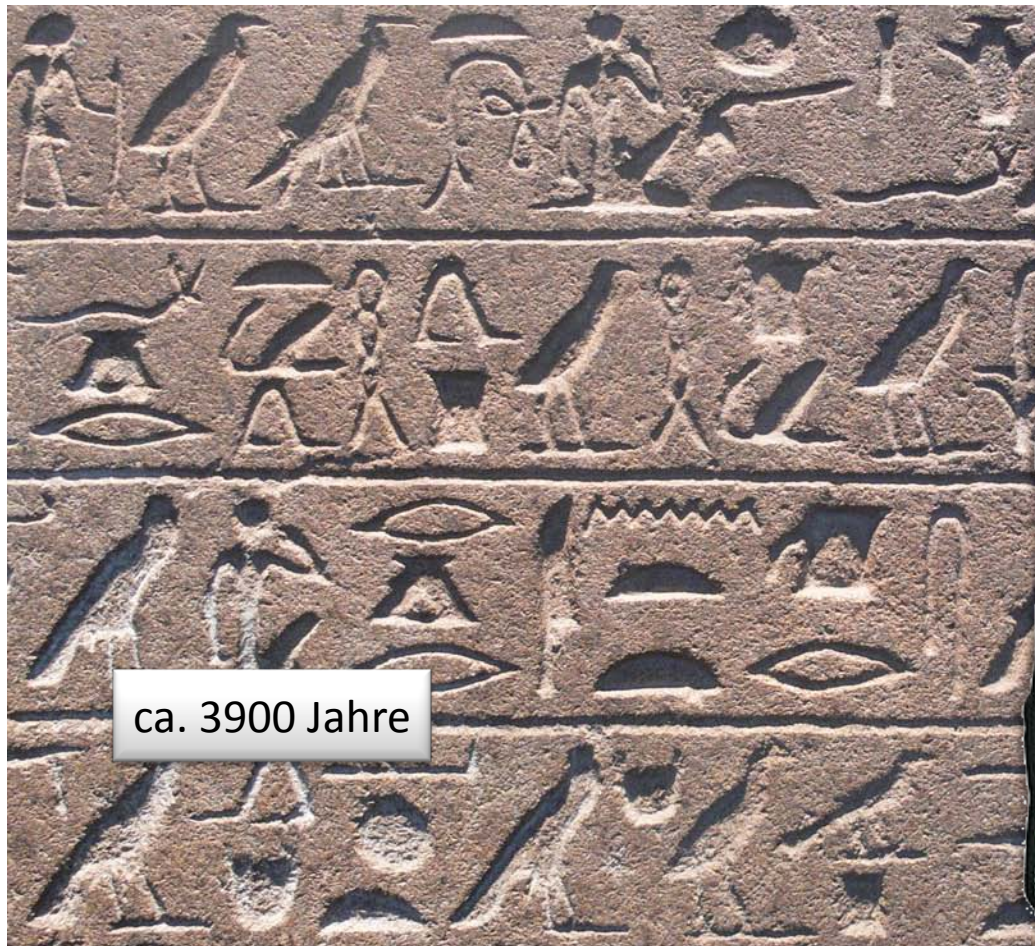
Was versteht man unter Langzeitarchivierung?

Wie betrifft das Bibliotheken?

Vorgehensweisen

Existierende Werkzeuge und Systeme

Was versteht man unter Langzeitarchivierung?



ca. 3900 Jahre



ca. 2200 Jahre

Stele aus Semna mit mittelägyptischer Aufschrift (12. Dynastie); Altes Museum, Berlin

Stein von Rosette; British Museum, London

Was versteht man unter *digitaler* Langzeitarchivierung?

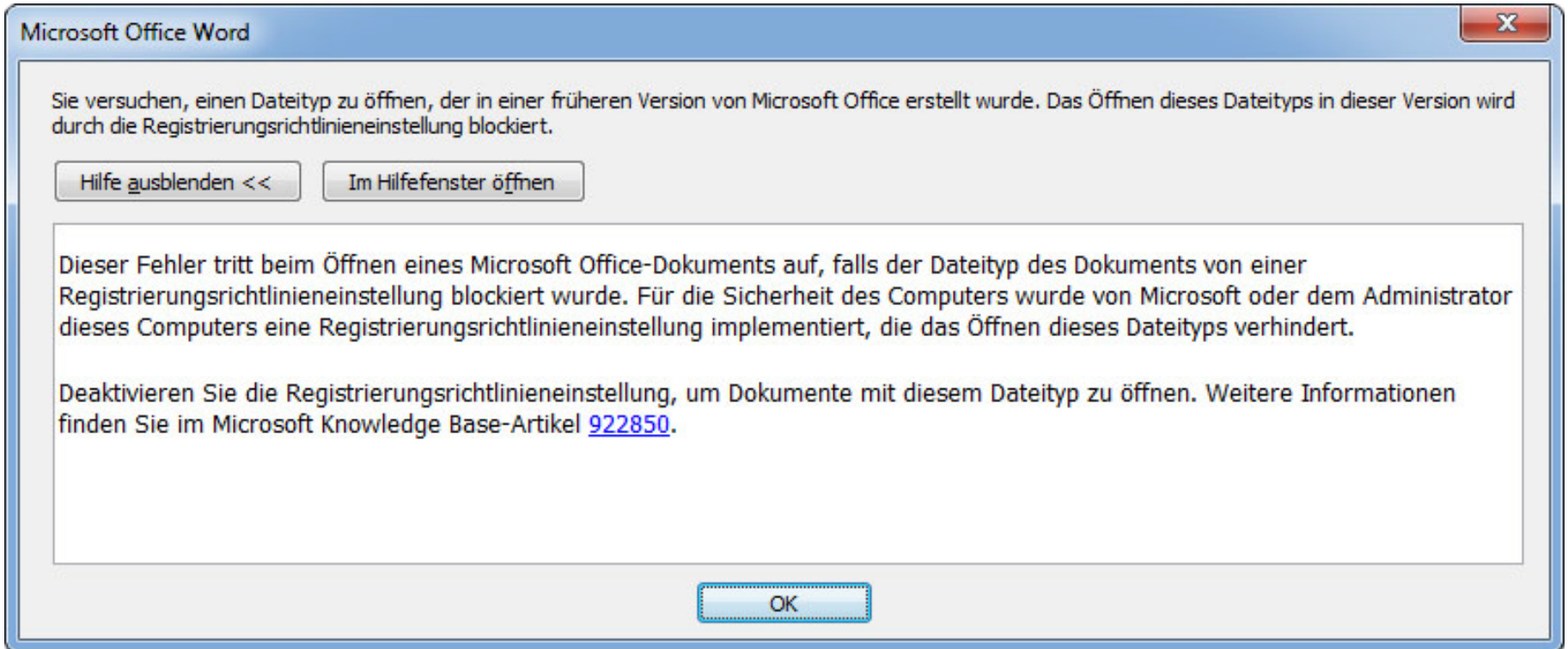
Digital information will last forever.

Or five years.

Whatever comes first...

Jeff Rothenberg

Die Probe auf's Exempel



Was versteht man unter Langzeitarchivierung?

- ❑ Wesentliche Inhalte sollen dauerhaft erhalten bleiben!
 - „Significant Properties“
- ❑ Fokus auf Nutzbarkeit, weniger auf Bewahrung der Unversehrtheit
 - „Designated Community“
- ❑ Entwicklung eines institutionsspezifischen Regelwerks
 - „Preservation Planning“
- ❑ Kontinuierliche Überwachung
- ❑ Integrität und Authentizität der Objekte gewährleisten
- ❑ Dokumentation (z.B. durch Provenance-Metadaten)

Agenda

Was versteht man unter Langzeitarchivierung?

Wie betrifft das Bibliotheken?

Vorgehensweisen

Existierende Werkzeuge und Systeme

Wie betrifft das Bibliotheken und Archive?

- ❑ Auftrag, das kulturelle Erbe zu bewahren
- ❑ Immer mehr Quellen stehen digital zur Verfügung:
 - Zeitschriften
 - Bücher
 - Datenbanken
- ❑ „Born-digital“-Materialien
- ❑ Neue Subskriptionsmodelle
 - „post-cancellation access“
- ❑ Auch digitale Daten altern (Speichermedien, Dateiformate, Hardware)

Agenda

Was versteht man unter Langzeitarchivierung?

Wie betrifft das Bibliotheken?

Vorgehensweisen

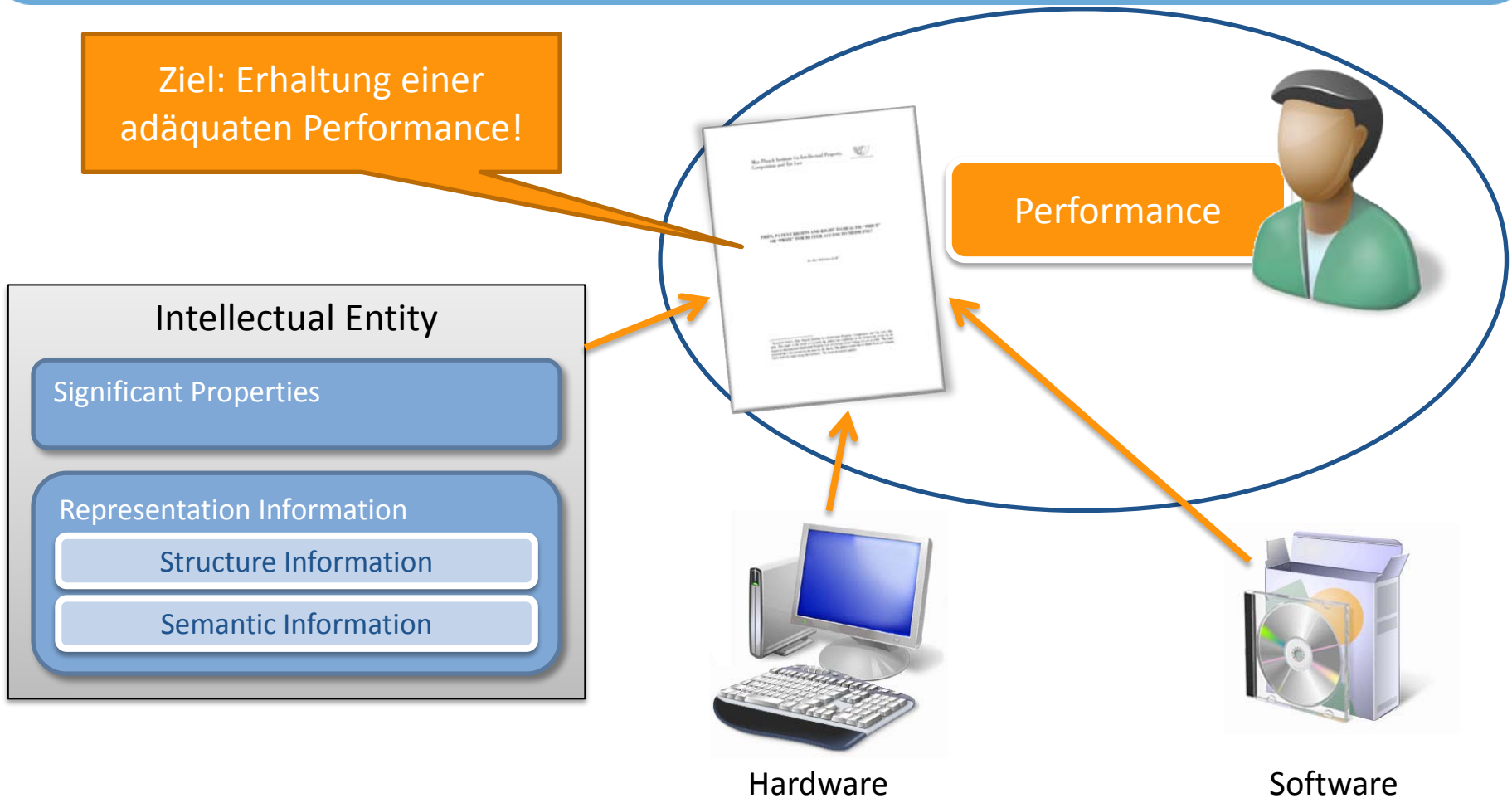
Existierende Werkzeuge und Systeme

Was sind die Herausforderungen?

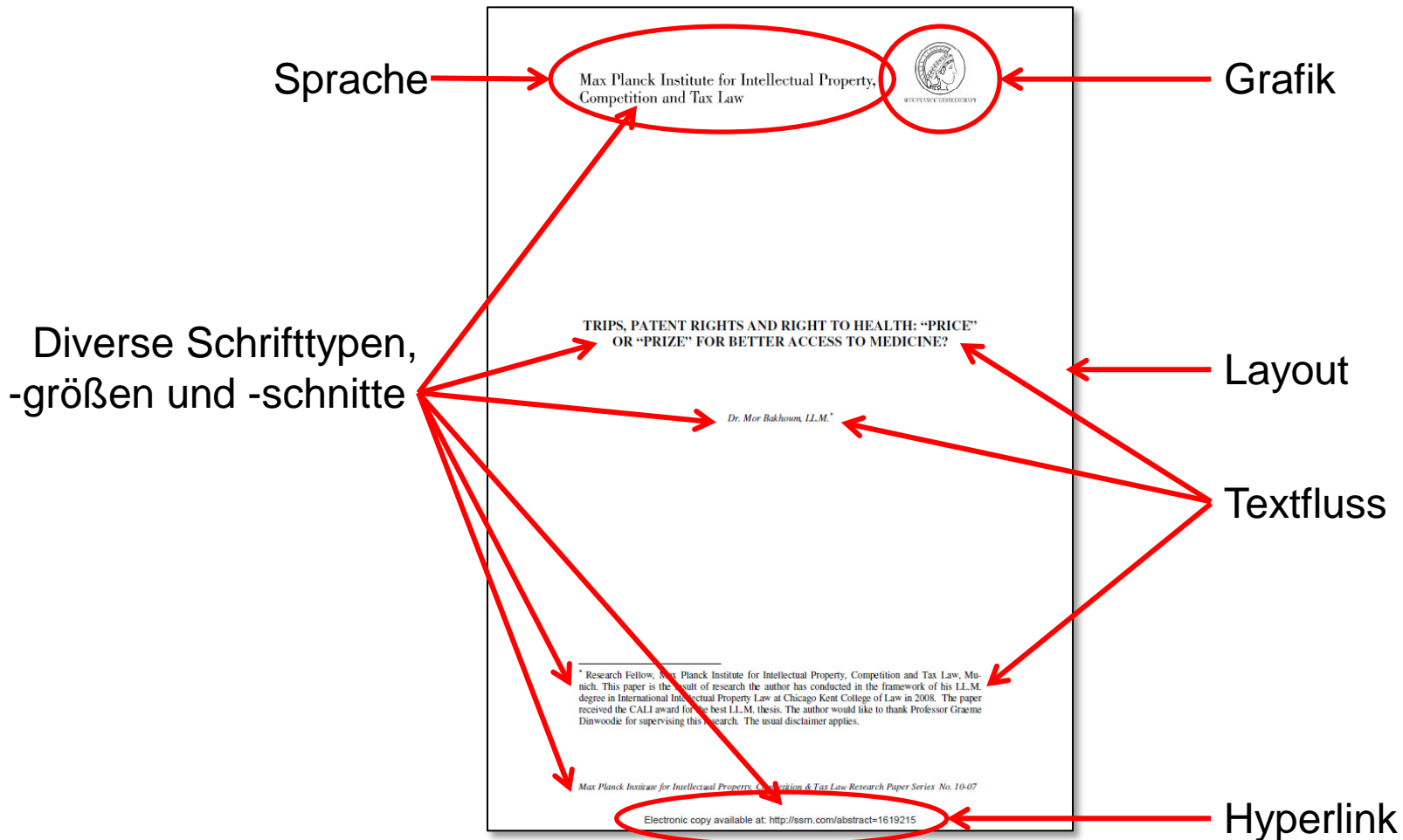
- ❑ Organisatorische Fragestellungen
 - Bewertung der Archivwürdigkeit
 - Preservation Planning
 - Qualitätssicherung
- ❑ Kosten
- ❑ Vertrauenswürdigkeit gewährleisten
- ❑ Formate (Vielzahl, Normierung)
 - Archivfähigkeit herstellen
 - Obsolete Formate erkennen
- ❑ Automatisierung
- ❑ Rechte berücksichtigen

Informationsobjekt und Performance

Ziel: Erhaltung einer adäquaten Performance!

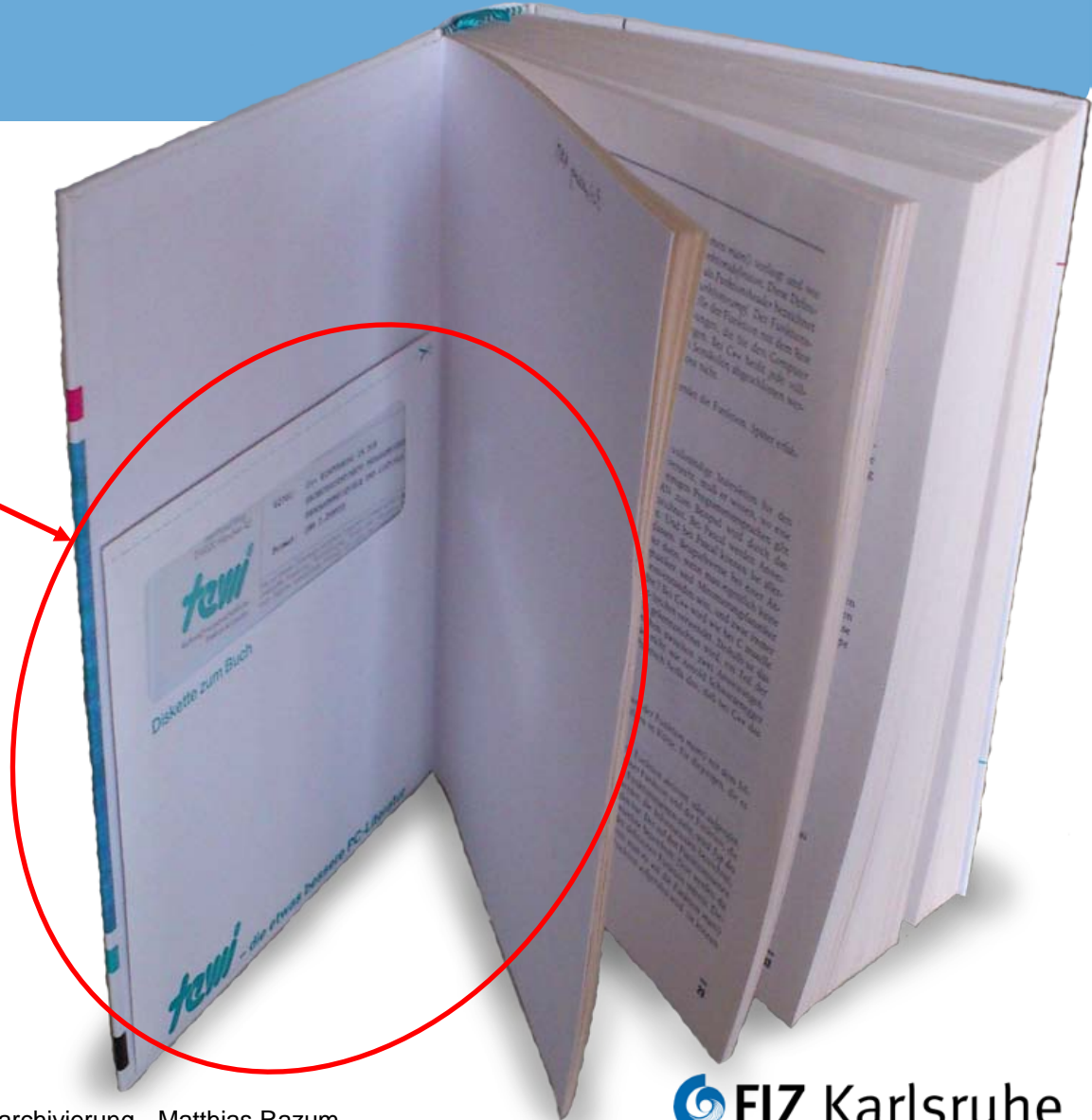


Significant Properties

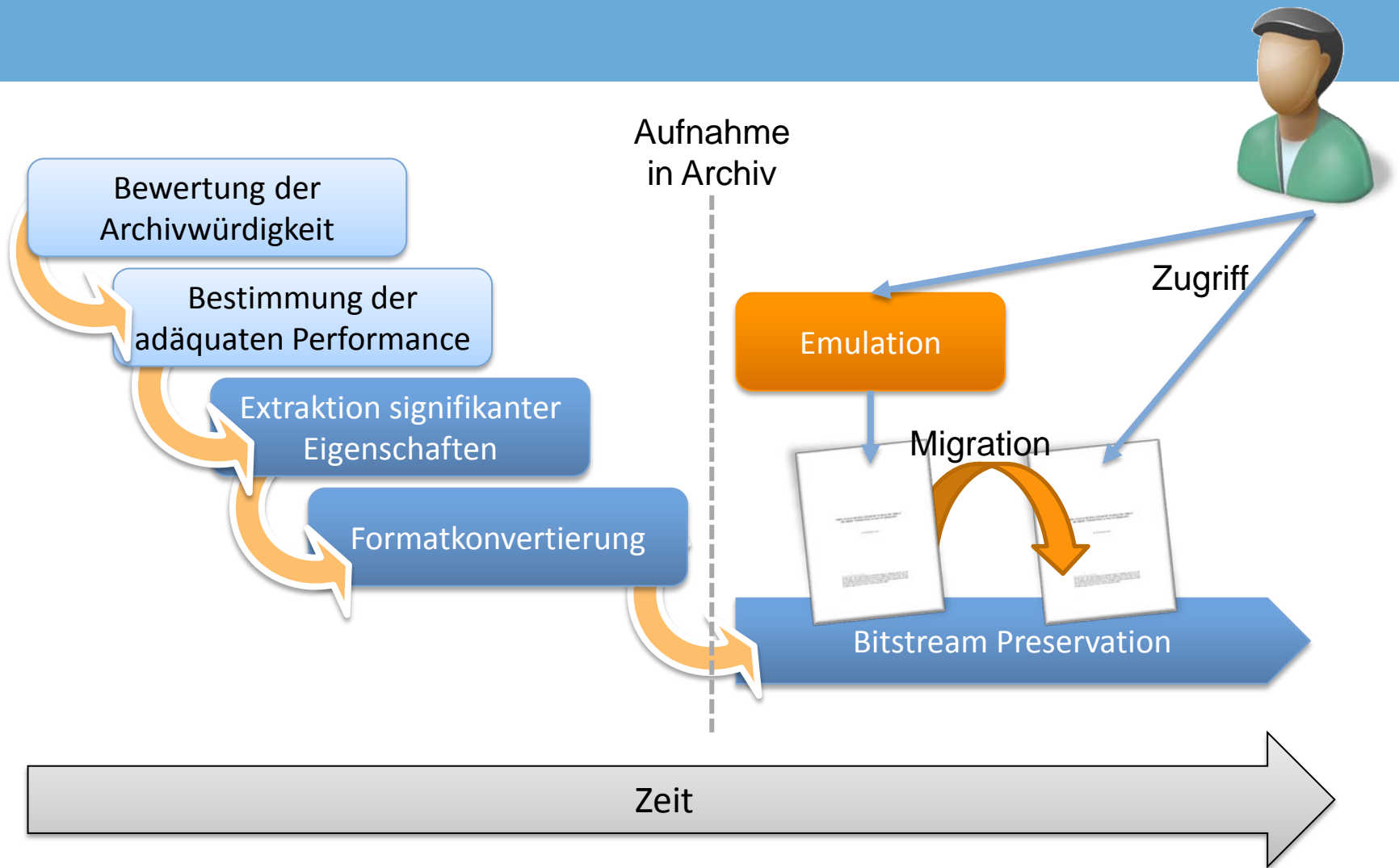


Wie erreicht man „adäquate Performance“?

Buch von 1992
mit 5 ¼“-Diskette
für MS-DOS 3.3



Prozessübersicht



Agenda

Was versteht man unter Langzeitarchivierung?

Wie betrifft das Bibliotheken?

Vorgehensweisen

Existierende Werkzeuge und Systeme

Extraktion signifikanter Eigenschaften

- ❑ Werkzeuge
 - Merkmalsextraktion (JHOVE, Tika, FITS)
 - Formatdatenbanken (PRONOM/DROID, GDFR)

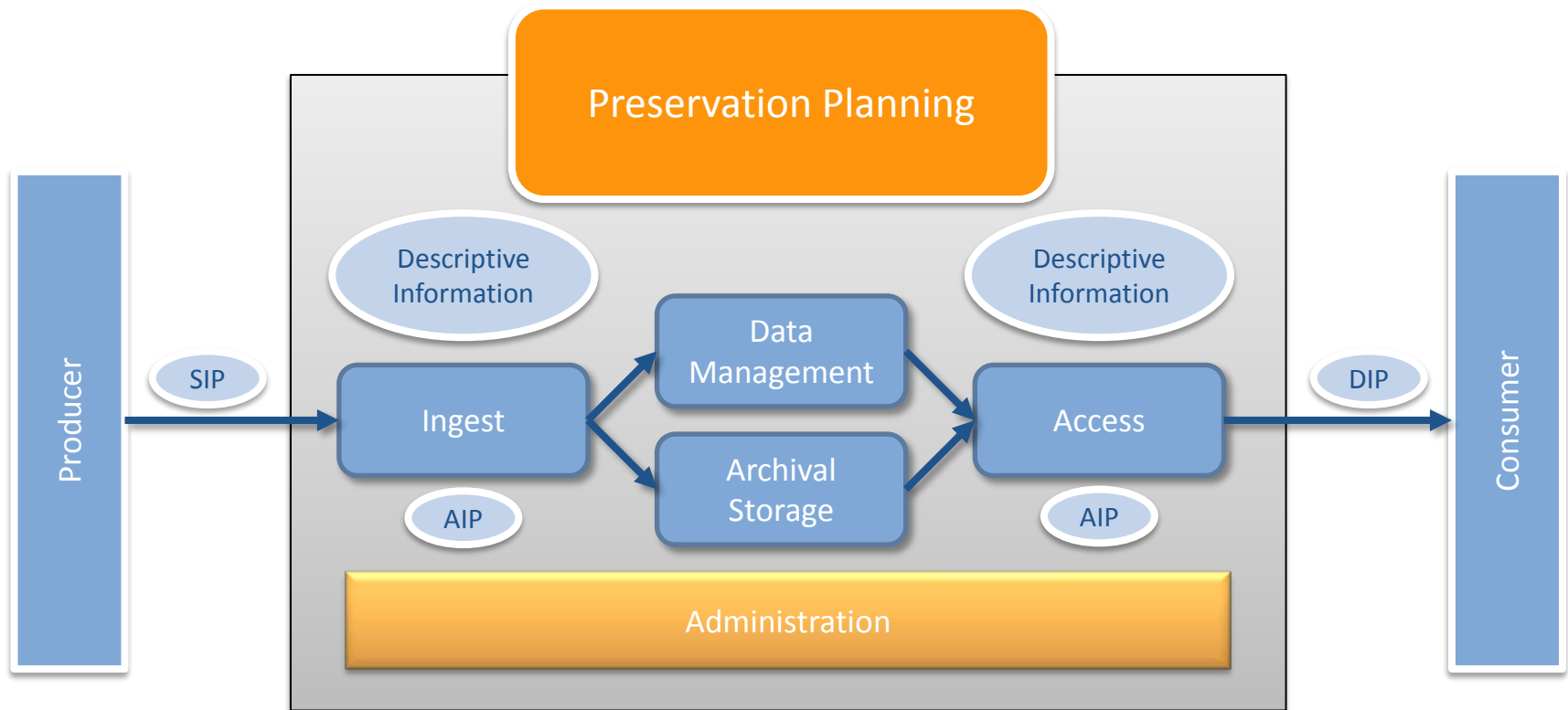
- ❑ Speicherung relevanter Eigenschaften
 - PREMIS-Metadaten
 - Deskriptive/technische Metadaten

Archivierung

- ❑ Softwaresysteme
 - Repositories (DSpace, Eprints, Fedora)
 - Archivsysteme

- ❑ Relevante Standards und Normen
 - Open Archival Information System (OAIS)
 - „Vertrauenswürdige digitale Langzeitarchive“ (DIN 31644)

Open Archival Information System (OAIS)



Quelle: nach nestor Leitfaden zur digitalen Bestandserhaltung, nestor Kompetenznetzwerk Langzeitarchivierung, 2011

Erhaltung der digitalen Objekte (Archival Storage)

- ❑ Bitstream Preservation
 - Redundanz
 - Replikation
 - Regelmäßiges Umkopieren auf aktuelle Datenträger

- ❑ Zugriffsmethode
 - Festplattenspeicher
 - Bandarchive
 - Mischlösungen (Hierarchical Storage Solutions)

- ❑ Validierung
 - Checksummen

Zugriff (Access)

- ❑ Migration
 - Immer mit Informationsverlust verbunden
 - Rechtzeitiges Erkennen obsoleter Formate
 - Abwägung von Kosten und Nutzen

- ❑ Emulation
 - Sehr aufwändig
 - Zukünftige Nutzung der Objekte eventuell unklar
 - Informationsverlust kann durch verbesserte Emulation verringert werden

- ❑ Validierung von Ergebnissen notwendig

Systeme

Gerade für die Archivierung von lizenzierten E-Journals und E-Books gibt es Alternativen (greifen im Fall von „trigger events“):

❑ LOCKSS (Lots Of Copies Keep Stuff Safe)

- Open-Source-System zur verteilten Langzeitarchivierung digitaler Objekte, entwickelt von David Rosenthal (Stanford University), arbeitet wie ein Proxy-Server

❑ Portico

- Semi-kommerzielles System für Inhalte wissenschaftlicher Verlage, gibt bei Eintreten eines „trigger events“ Inhalte für Portico-Mitglieder frei

„Trigger Events“ von Portico

- When a publisher ceases operations and titles are no longer available from any other source
- When a publisher ceases to publish and offer a title and it is not offered by another publisher or entity
- When back issues are removed from a publisher's offering and are not available elsewhere
- Upon catastrophic failure by a publisher's delivery platform for a sustained period of time

Quelle: <http://www.portico.org/digital-preservation/services/reliable-access/trigger-events/>

Wer kümmert sich schon darum?

Wichtige Projekte und Initiativen:

- Nestor
 - <http://www.langzeitarchivierung.de/>
- Alliance for Permanent Access
 - <http://www.alliancepermanentaccess.org/>
- Open Planets Foundation
 - <http://www.openplanetsfoundation.org/>
- Digital Preservation Coalition
 - <http://www.dpconline.org/>
- National Digital Information Infrastructure and Preservation Program
 - <http://www.digitalpreservation.gov/>

Weiterführende Informationen

- ❑ nestor Handbuch – Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
- ❑ nestor Leitfaden: Wege ins Archiv – Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv
- ❑ nestor Leitfaden zur digitalen Bestandserhaltung

Danke !

