

Workshop 26. Mai 2004 in Bern

## Perspektiven für das E-Archiving für Lehre und Forschung in der Schweiz

Arlette Piguet

Dr. Matthias Töwe

- Geschäftsstelle Konsortium der Schweizer Hochschulbibliotheken
- Firma IBM

# Programm

---

09.30h – 09.45h	Begrüssung
09.45h – 11.00h	Projekt E-Archiving (M. Töwe / A. Piguet)
11.00h – 11.15h	<b>Pause</b>
11.15h – 12.00h	Global Technology Outlook (W. Hehl)
12.00h – 13.00h	<b>Stehlunch</b>
13.00h – 14.00h	Vorstellung Dokumentenspeichersysteme: MILESS und MyCoRe Projekt Univ. Jena (U. Reber)
14.00h – 14.30h	Diskussion
14.30h – 15.00h	<b>Kaffee</b>
15.00h – 16.00h	Vorstellung Archivierungssystem DIAS (Digital Information Archiving System) der Königlichen Bibliothek in Den Haag (S. Seybold)
16.00h – ca.16.30h	Fragen, Diskussion

# Konzeptstudie E-Archiving

---

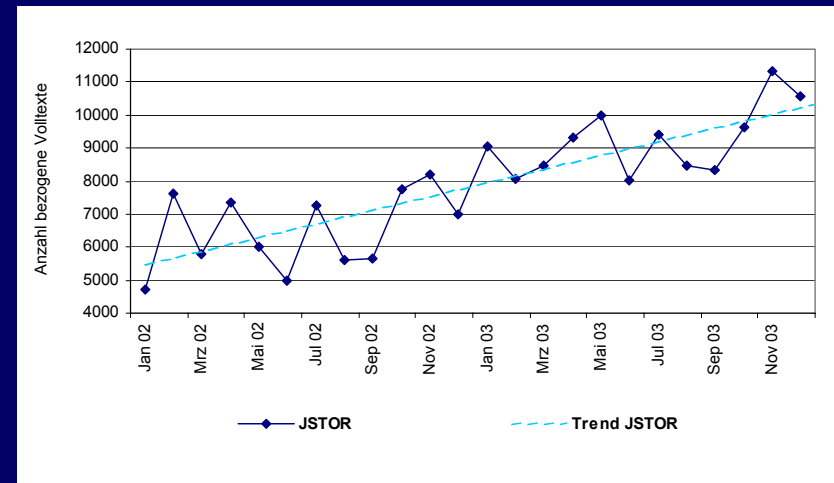
Generelles Ziel: Entwicklung einer Gesamtstrategie für das E-Archiving an den Schweizer Hochschulen

- E-Journals kommerzieller und nicht-kommerzieller Verlage, Fachdatenbanken
- Publikationen der Schweizer Hochschulen
- Noch nicht digitalisierte Zeitschriften

# E-Publikationen

---

- Zunahme der Verlagsangebote
- Zunahme der Angebote auf Hochschulservern („open access“)
- Zunehmende Nachfrage der Benutzer



# Verlagsangebote: E-Journals

---

- Bibliotheken oder Konsortien erwerben das Zugriffsrecht auf die Volltexte in Form von Lizenzen
- Zugriff via Verlagsserver (Suchfunktionen, Volltexte: HTML, PDF)
- Heute i.d.R. noch paralleler Erwerb der Printausgaben
- Immer häufiger Wechsel auf E-only

## E-Publikationen: „Open Access“, u.a.

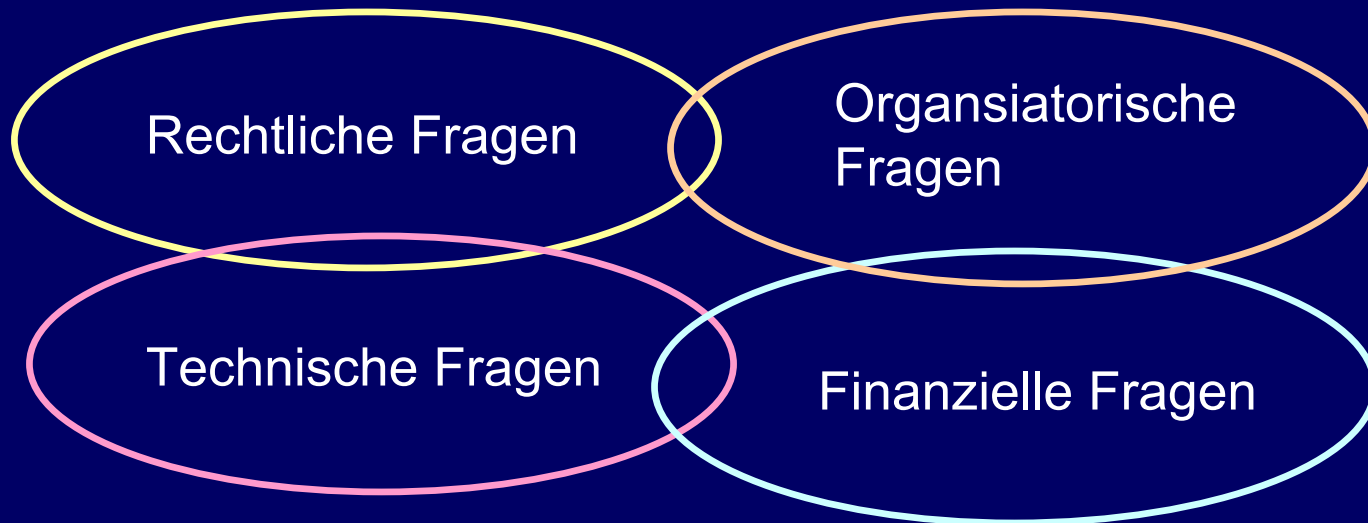
---

- Dokumentenserver: Hochschulschriften (Dissertationen, Preprints, Postprints, Lehr- und Lernmaterialien etc.)
- Server für digitalisierte Bilder, Karten, alte Drucke etc.
- Server für „Open Access Journals“

# E-Archiving


---

Definition: The maintenance of digital material over the long-term with a view of ensuring continued accessibility.

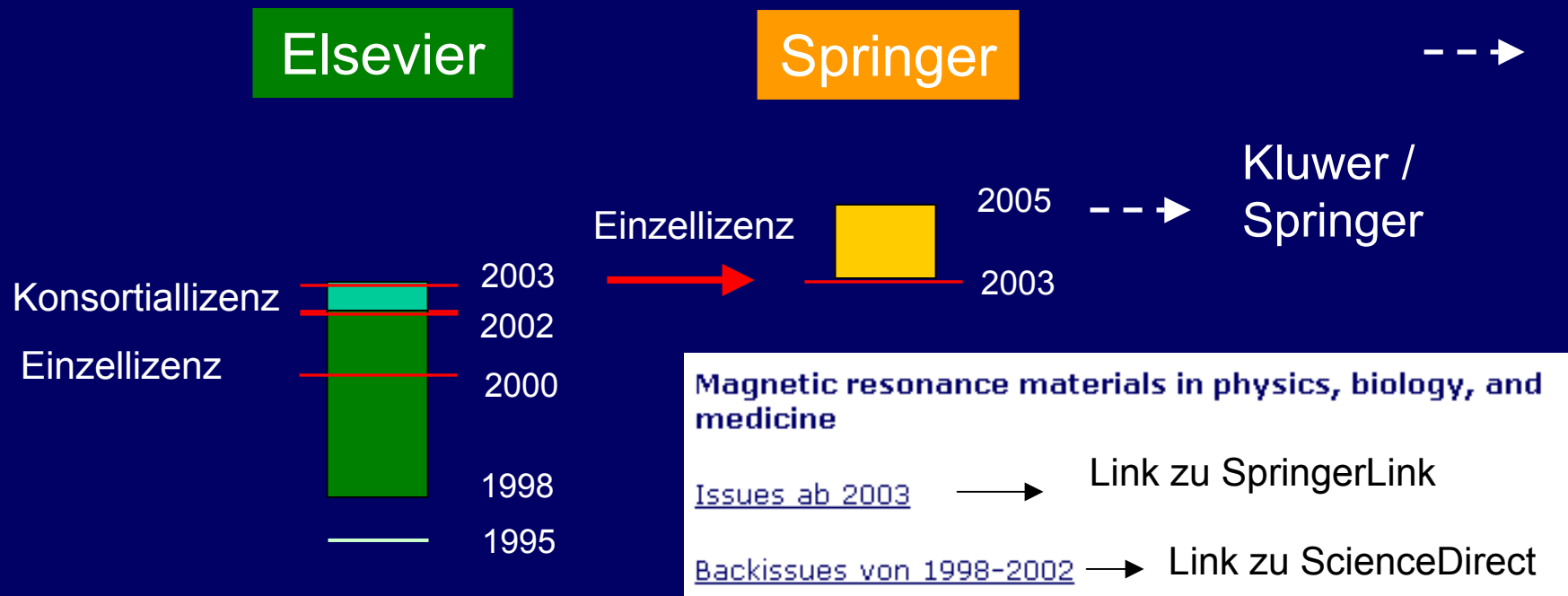


## Verlagsangebote: langfristiger Zugriff und langfristige Lesbarkeit der E-Journals?

---

- Die Bibliothek oder das Konsortium führt die Lizenz nicht mehr weiter
- Die Zeitschrift wird eingestellt
- Die Zeitschrift wechselt den Verlag 
- Technologie (Hard- und Software) ändert sich: neue Formate und Funktionalitäten
- ...

# Verlagsangebote: langfristiges Zugriffsrecht auf die E-Journals?



Titel der Zeitschrift: Magnetic Resonance Materials in Physics, Biology and Medicine

# Verlagsangebote: Lizenzvertrag für die E-Journals

---

- Lizenzvertrag zwischen Bibliothek oder Konsortium und Verlag
- „Archivklausel“ (Empfehlungen u.a. von ARL, ICOLC, IFLA, LIBER)
  - Recht für den Zugriff auf die einst erworbenen Daten („paid content“)
  - Kosten minimal
  - Festlegung der Verantwortlichkeiten (Verlag, Bibliothek oder Konsortium, Drittanbieter)
  - Technische Spezifikationen offen

## Archivierung von E-Journals / Zugriff

---

- Bedingungen für ein Dienstleistungsniveau, das zum heutigen vergleichbar ist:
  - Metadaten
  - Benutzerschnittstelle mit angemessene Such- und Blätterfunktionen
- Anbieter geben diese Daten und die eigene Software ungern preis

Es gibt viele offene technische Fragen zur  
Langzeiterhaltung elektronischer Daten.

Aber die Technik ist nicht das einzige Problem  
und wird auch nicht die alleinige Lösung bieten  
können.

# Technische Herausforderungen

---

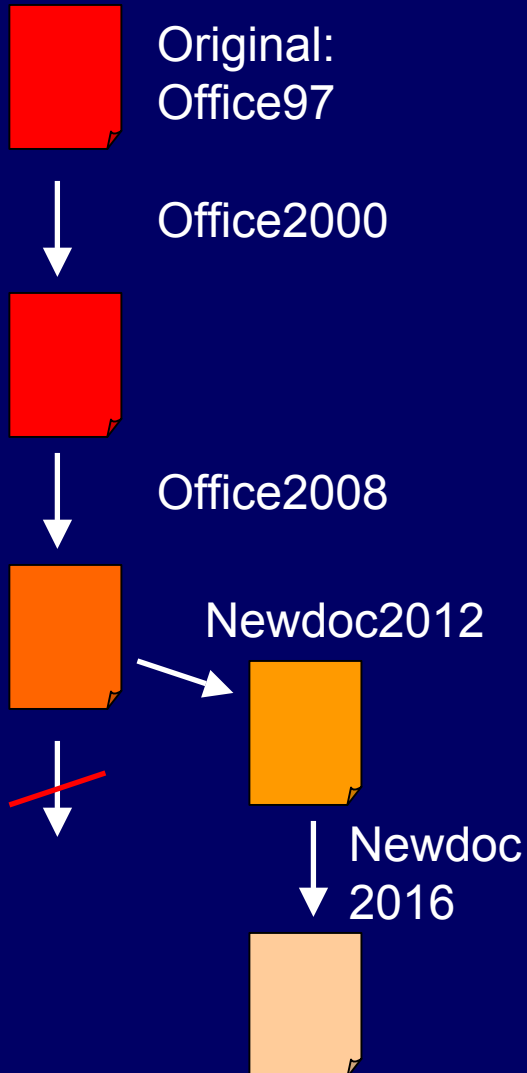
- Bits müssen physisch erhalten werden:
  - Über längere Zeit auf dem gleichen Träger („Refreshing“)
  - Über wechselnde Mediengenerationen („physische“ Migration)
- Die Wahl des Zeitpunktes für diese Massnahmen ist entscheidend für Erfolg und Kosten.

## Schwieriger: *Logische* Erhaltung, um eine spätere Interpretation zu ermöglichen

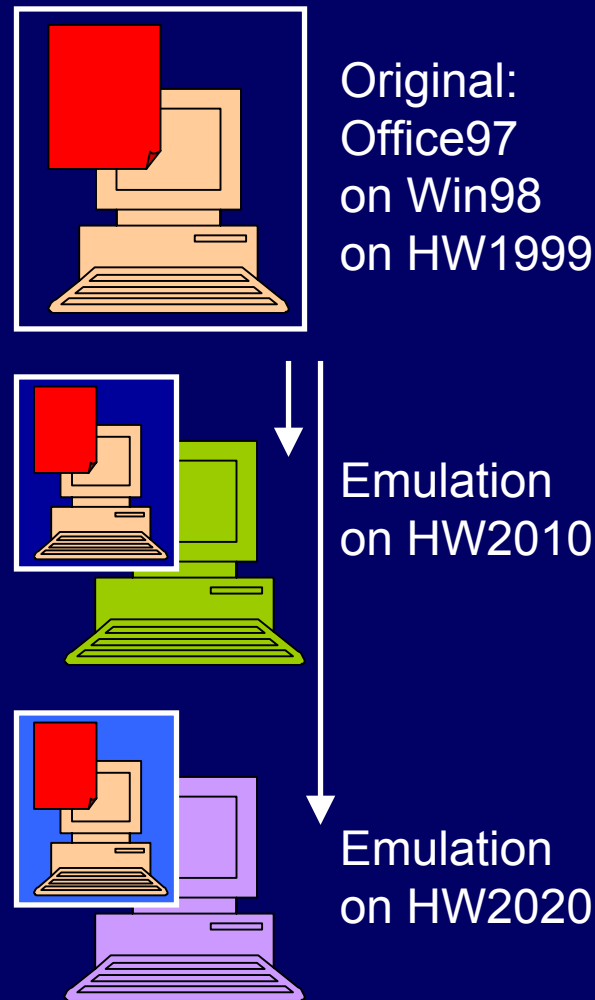
---

- Daten- und Dateiformate müssen bekannt sein
- Benötigte Software, Betriebssysteme und Hardware müssen entweder erhalten sein oder in zugänglicher Form erschöpfend beschrieben werden
- Authentizität muss gewährleistet sein

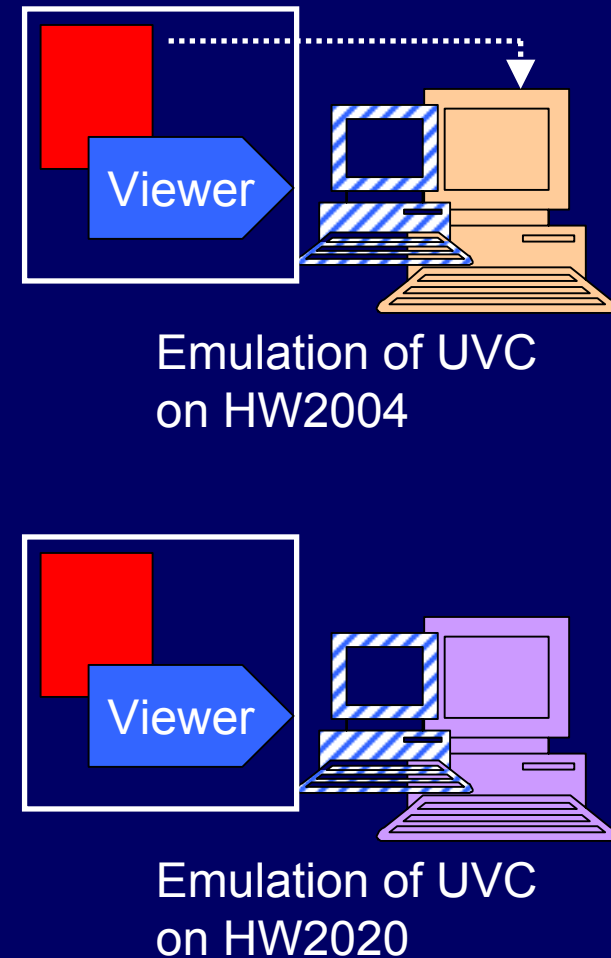
## Migration



## Emulation



## UVC



# Meist diskutierte Herangehensweisen I

---

- („Logische“) Migration: Umwandlung von Dateien aus einer Anwendung in neuere Versionen dieser Software oder in eine andere Anwendung
  - Muss periodisch mit jeder Einzeldatei durchgeführt werden
  - Fehleranfällig
  - Informationsverlust ist zu erwarten
  - Machbar eher für kürzere Zeiträume

## Meist diskutierte Herangehensweisen II

---

- Emulation: „Simulation“ überholter Hardware (seltener SW oder OS) auf aktueller Hardware. Dies ermöglicht es, erhaltene Software zu betreiben, die auf die alte Hardware angewiesen war.
  - „Look and feel“ werden erhalten
  - Gute Dokumentation wird benötigt
  - Kann sehr komplex werden
  - Muss für jede Hardwaregeneration wiederholt werden

## Meist diskutierte Herangehensweisen III

---

- Universal Virtual Computer (UVC): Heute werden Viewer für aktuelle Formate auf einem wohldefinierten, aber einfachen UVC programmiert.  
In der Zukunft muss dann nur der UVC emuliert werden, um die Viewer zu benutzen.
  - Viewers können heute getestet werden
  - Es wird eine heutige Sicht konserviert
  - Bisher nirgends produktiv

## Ergänzende Massnahmen I:

---

- Schaffung und Einhaltung offener, internationaler und gut dokumentierter Standards.
- Schaffung und Unterhalt von Verzeichnissen für Formate sowie Soft- und Hardware und ihre wechselseitigen Abhängigkeiten sowie ihre „Lebenserwartung“.
- Erfassung von Metadaten zum Erhaltungsprozess und zu technischen Eigenschaften.

## Ergänzende Massnahmen II:

---

- Koordinierte Archivierung von Printexemplaren – verteilt oder zentral. Diese könnte Teil einer insgesamt koordinierten Erwerbungspolitik für Zeitschriften sein.
- Klare Verantwortlichkeiten definieren in intensiver Zusammenarbeit.

## Internationale Projekte I:

---

- Staatsarchive
  - Versuchen die Dokumentenerzeugung zu beeinflussen, um einheitlichere Daten und notwendige Metadaten zu erhalten
  - Arbeiten an Format-Verzeichnissen (z.B. PRONOM, UK)
  - Erhaltung von Datenbanken ist eine wichtige Aufgabe (z.B. ARELDA/SIARD, CH)

## Internationale Projekte II:

---

- Nationalbibliotheken
  - Verantwortlich für die *nationale* Produktion einschliesslich Internetseiten (in Auswahl oder umfassend)
  - Installationen im grossen Massstab
  - Für kommerzielle Inhalte i.d.R. „semi-dark“: Zugänglich nur aus der Bibliothek heraus
  - Z.B. Koninklijke Bibliotheek, NL; SLB-e-Helvetica, CH ...

## Internationale Projekte III:

---

- Hochschulbibliotheken und Konsortien:
  - Lokales hosting von Zeitschriften: (OhioLINK, US; HeBIS, D)
  - Verschiebung zu e-only ? (BIBSAM, S)
  - Aufbau vernetzter Hochschuldokumentenserver (DARE, NL; DiVA, S)
  - Nutzung von Drittanbietern für Zugang zu und Archivierung von Zeitschrifteninhalten
  - Skeptisch gegenüber Inhalt von Lizenzen

## Internationale Projekte IV:

---

- Vertrauenswürdige Drittanbieter:
  - Wichtigste Beispiele: JSTOR and OCLC-ECO
  - Überzeugende Konzepte für die Archivierung
  - Aber:
    - Dienstleistung basiert auf Abonnement ohne weiteren Anspruch nach dessen Ende

## Internationale Projekte V:

---

- **Kompetenzzentren und Netzwerke:**
  - Üblicherweise nicht selbst als Archiv-  
einrichtung aktiv, aber sie koordinieren die  
Forschung, bieten fachliche Unterstützung  
und/oder fördern Weiterbildung und  
Informationsaustausch:  
Digital Curation Centre, UK; nestor, D; DPC,  
UK; NDIIPP, US; PADI, AUS...

## Internationale Projekte VI:

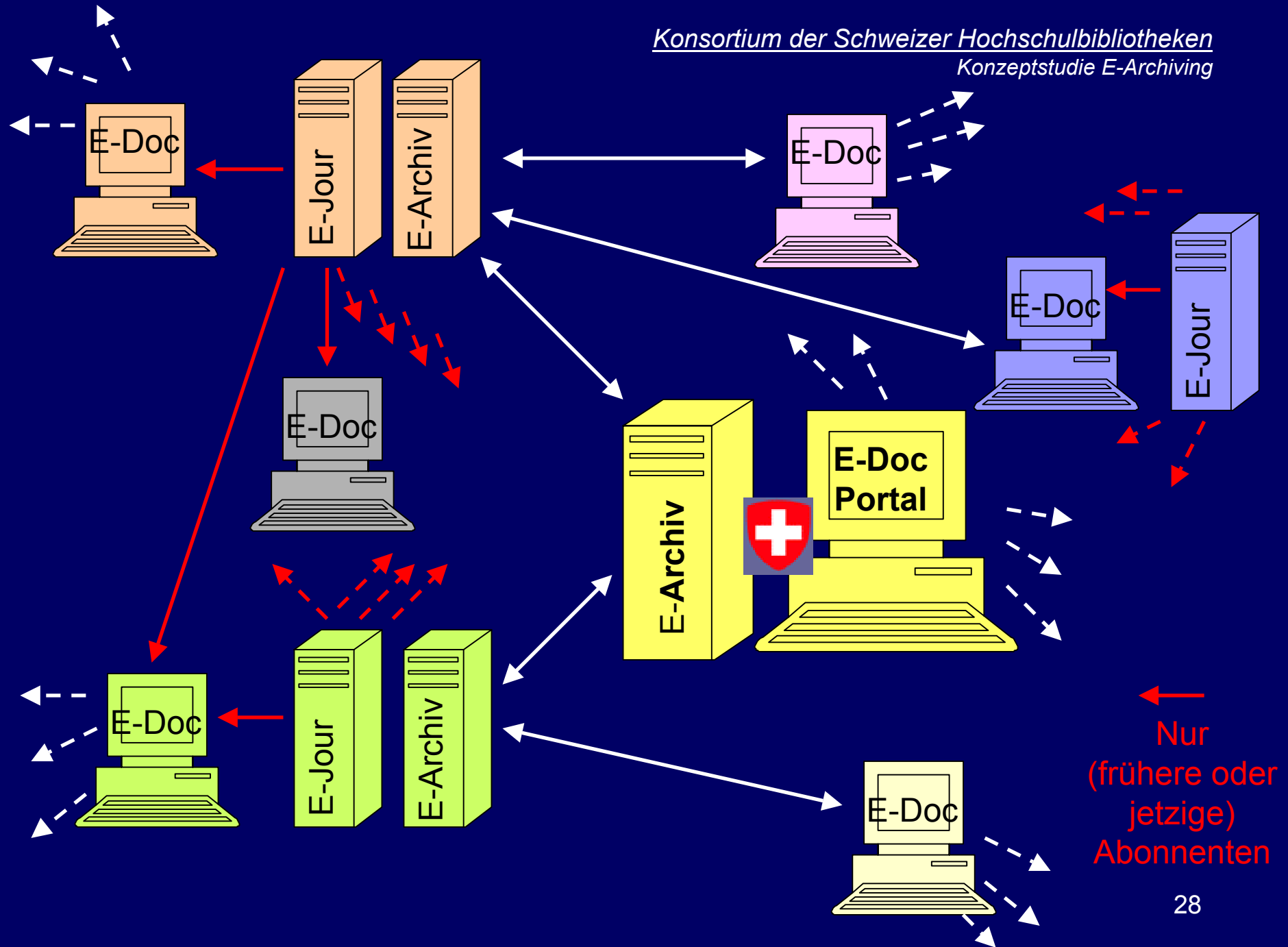
---

- Verlage:
  - Grosse Verlage arbeiten bereits mit Nationalbibliotheken zusammen, kleinere werden dies tun müssen, sobald die Pflichtexemplarregelungen erweitert sind (NL, D, CH, UK...). Die Inhalte bleiben „semi-dark“, so lange es den Verlag noch gibt.
  - Die Verlage selber werden vermutlich ihre Daten so lange archivieren und pflegen, wie dies für sie lohnend erscheint.

## Visionen davon, wie es sein *könnte*...

---

- Ein Netzwerk von Hochschulservern
  - Mit Dokumentspeichern für qualitätsgeprüfte graue Literatur, die über ein Schweizer Portal oder über ein lokales konsultiert werden kann.
  - Mit Zeitschriften, die für alle Schweizer Hochschulen gehostet werden, die je für den betreffenden Inhalt bezahlt haben.
  - Mit zentralem und/oder verteiltem Archiv für Material, das nicht anderswo archiviert wird.



## ...mögliche Wege zum Ziel: Graue Literatur...

---

- Lokal betriebene Dokumentenserver, deren Inhalt *zusätzlich* zumindest über die Metadaten unter einer landesweiten Oberfläche zusammengeführt wird. Sogar die Software kann lokal gewählt werden, wenn bestimmte Bedingungen erfüllt sind (OAI-Fähigkeit; vgl. DARE).
- Koordinierte Archivierung dieser Inhalte auf Servern an mehr als einem Standort.

## ...mögliche Wege zum Ziel: Zeitschriften...

---

- Wenige lokal betriebene Server, die jeweils bestimmte Zeitschrifteninhalte nur den berechtigten Nutzern in der ganzen Schweiz bereitstellen.  
Zusätzlich möglich: Digitalisierte Zeitschriften.
- Koordinierte Archivierung dieser Inhalte auf Servern an mehr als einem Standort.
- Die Datenmenge wird ein Problem, wann immer das *ganze* Volumen bearbeitet werden muss.

## ... Arbeiten im Projekt...

---

- Beobachtung von Entwicklungen weltweit.
- Sichtung der Lizenzvereinbarungen
- Test der Dokumentenserversoftware „DSpace“ (abgeschlossen)
- Auswertung von Bestandesdaten gedruckter Titel für eine verteilte Archivierung
- Pilotdigitalisierung in Arbeit
- ...und mehr
- Konsortium: Beschäftigt sich vor allem mit den grossen Verlagen – Gefahr einer Lücke bei kleinen Anbietern.

## ...und Fragen aus der Wirklichkeit

---

- Was wollen „die“ Hochschulbibliotheken?
  - Gar keine eigenen Archivanstrengungen, sondern Verlass auf externe Institutionen?
  - Bereitstellung von Metadaten oder Inhalten für ein gemeinsames Portal für graue Literatur?
  - Zentrale oder lokal kontrollierte Lösungen?
- Benötigte und vorhandene Ressourcen?
  - Es wird teuer – niemand wagt konkrete Prognosen.
  - Mitfinanzierung aus Bundesmittel?

## Nächste Schritte:

---

- Fragebogen für die Konsortialbibliotheken
  - Recht detailliert – bitte versuchen Sie zu antworten!
- Der Bericht über die Konzeptstudie ist vorgesehen für Ende 2004.
- Umsetzung??

# Verbindung zu den heutigen Präsentationen

---

- **W. Hehl: Global Technology Outlook**  
Aktuelle und zukünftige Trends als allgemeine Hintergrundinformation
- **U. Reber: MILESS/MyCoRe**  
Eine produktive Beispiellösung für Hochschulserver
- **S. Seybold: DIAS**  
Eine produktive Lösung für ein digitales Archiv im grossen Massstab

## Links

---

- DARE: <http://www.darenet.nl/en/toon>
- DiVA: <http://www.diva-portal.se/about.xsql>
- eDepot (KB Den Haag):  
<http://www.kb.nl/kb/menu/ken-arch-en.html>
- e-Helvetica: <http://www.e-helvetica.admin.ch/index.htm>
- JSTOR: [www.jstor.org](http://www.jstor.org)
- OCLC-ECO:  
<http://www.oclc.org/electroniccollections/>
- PRONOM:  
<http://www.records.pro.gov.uk/pronom/>