

**Tobias Steinke**

# **Erfahrungen mit kopal und digitaler Langzeitarchivierung**

# Inhaltsverzeichnis

## **1. Das Projekt kopal**

- 1. Ergebnisse**
- 2. Perspektiven**

## **2. Praxisprobleme**

- 1. Dateiformate**
- 2. Kopierschutz**
- 3. Webseiten**

## Das Projekt kopal

- Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen
- BMBF-gefördertes Projekt 2005 – Juli 2007
- Partner: Deutsche Nationalbibliothek, SUB Göttingen, IBM Deutschland, Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG)
- Gemeinsamer Archivkern DIAS von IBM, lokale Software koLibRI (Open Source) von DNB und SUB

## **kopal: Ergebnisse (1)**

- DIAS als OAIS-konformes Kernsystem
  - Flexibles Datenmodell für Langzeitarchivierungsmetadaten und Migrationsversionierung
  - Universelles Objektformat für Datenaustausch
  - Technische Schnittstelle für komplexe Migrationsabfragen
  - Mandantenfähigkeit
  - Stand: Finale Nachbesserungen durch IBM

## kopal: Ergebnisse (2)

- koLibRI als generische Softwarebibliothek:
  - Vielseitige Java-Tools für Workflowaufbau, Archivpaketerstellung, gekapselte Kommunikation mit Archivsystem, technische Metadatengenerierung mit JHOVE, Zugriff und Migrationsmanagement
  - Hohe Konfigurierbarkeit und Modularisierung
  - Grundlegendes Framework für Dateiformatmigration
  - Stand: Version 1.0 mit Dokumentation frei verfügbar bei <http://kopal.langzeitarchivierung.de>

## **kopal: Perspektiven**

- kopal-Einbettung in DNB: Integration in produktive Arbeitsabläufe der Netzpublikationen
- DPFuse: Zusammenarbeit von DNB, SUB und KB (Niederlande) zur Erarbeitung von gemeinsamen Anforderungen für Zeit bis 2010
- DP4Lib: DFG-Antrag für den Aufbau praktischer Konkretisierungen der kopal-Nachnutzung
- EU-Projekt SHAMAN: Einbindung in aktuelle Forschung

## Praxisprobleme: Dateiformate

- Vielfältige digitale Veröffentlichungspraxis ist technisch nicht immer geeignet für Archivierung
- Verlagskunde kann wählen (z. B. HTML oder PDF), intern anderes Format (eigenes XML): Was archivieren?
- Was bedeutet Inhaltsgleichheit bei digitalen Medien?
- ISO-Standard PDF/A: Von Verlag nicht geliefert („kein Kundenbedarf“)
- Nötig: Informationsaustausch und praktikable Lösungen

## Praxisprobleme: Kopierschutz

- Datenträger größtenteils mit Kopierschutz
- Vielfältige Ausprägung, u. a.:
  - Hardwaremanipulation („Defekte“)
  - Seriennummereingabe
  - Online-Freischtaltung
  - Individualisierung
- Langzeitarchivierung nicht möglich – Verlust droht
- Nötig: Schutzfreie Archivexemplare, Dokumentation

## Praxisprobleme: Webseiten

- Webseiten abliefern oder holen?
- Anbietersicht vs. Nutzersicht
- Dokumente als Teil des Web: Abgrenzung möglich?
- Authentizität: Änderung der Webseiten für Nutzung und Langzeitarchivierung erlaubt?
- Nötig: Praktikable und pragmatische Lösungen für Ablieferer und Archiv, internationale Zusammenarbeit (etwa im IIPC)

## Zusammenfassung

- Projekt kopal erfolgreich abgeschlossen
- Nachnutzung und Weiterentwicklung wird fortgeführt
- Konkrete Praxisprobleme werden immer deutlicher
- Digitale Langzeitarchivierung bleibt auch zukünftig große Herausforderung für bewahrende Institutionen