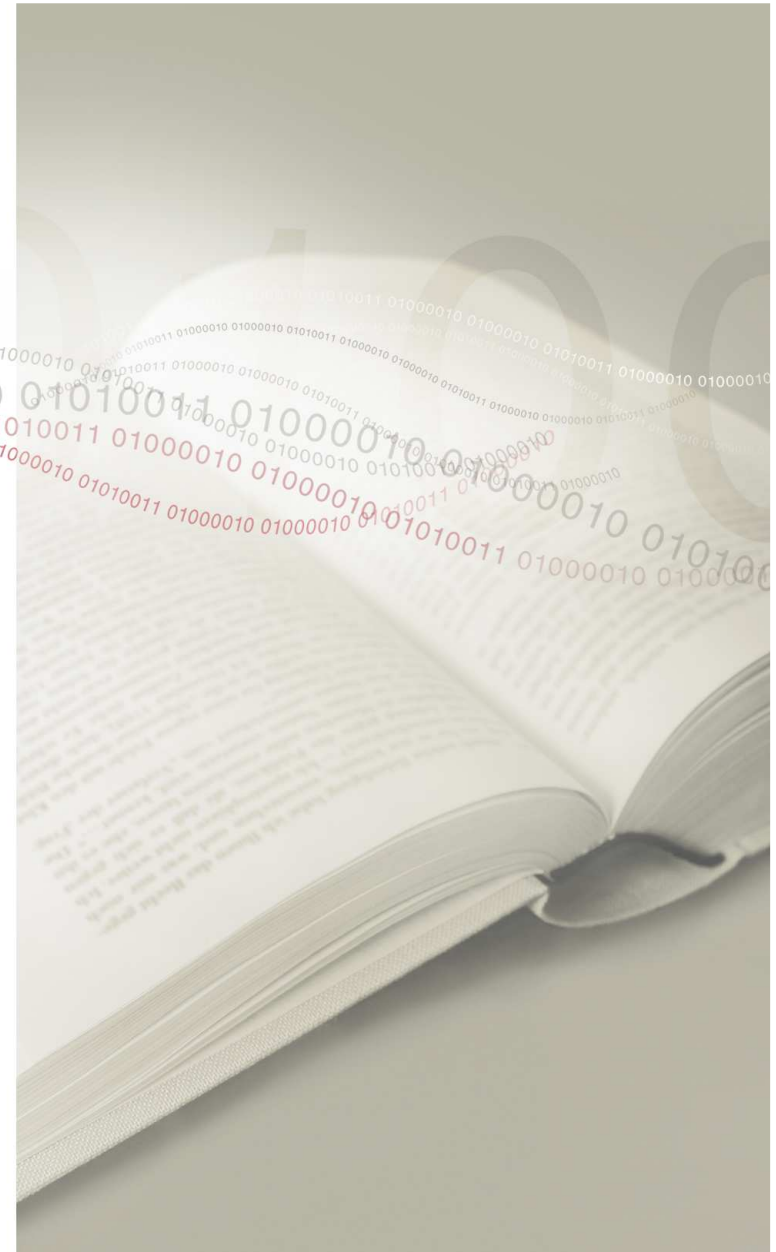


# **Verfahren und Werkzeuge aus nationalen und internationalen Projekten in der Praxis der digitalen Langzeitarchivierung an der Bayerischen Staatsbibliothek**



Tobias Beinert

4. Leipziger Kongress für Information und Bibliothek  
17. März 2010

# Agenda

- Kurzübersicht: digitale Langzeitarchivierung und das Projekt BABS2 an der BSB
- Die Anwendung des nestor-Kriterienkataloges in der Praxis
- Preservation Planning mit PLATO
- Web-Curator-Tool
- Technische Metadaten: JHOVE, DROID, FITS
- Ausblick: Nachnutzung der Projektergebnisse

# Langzeitarchivierung@BSB: 3 verknüpfte Handlungsfelder

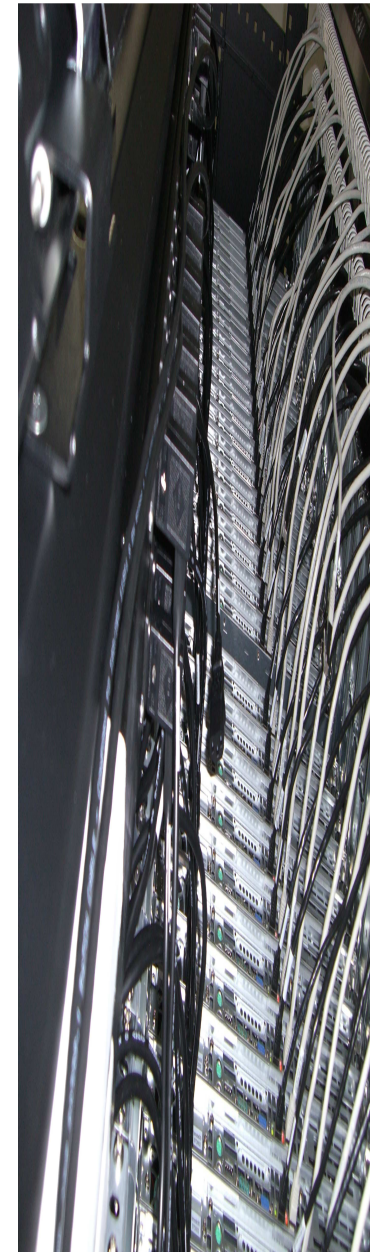
## Routinebetrieb im Rahmen der digitalen Produktion

- Bibliothekarisches Archivierungs- und Bereitstellungssystem (BABS):
  - technischer Betrieb und Weiterentwicklung
  - bibliothekarische Geschäftsgänge

## Ausbau und Konsolidierung der Strukturen

- Aufbau- und Ablauforganisation
- LZA-Profil und Konzepte
- Technologische Infrastruktur (Rosetta-Projekt)

## Forschung und Entwicklung



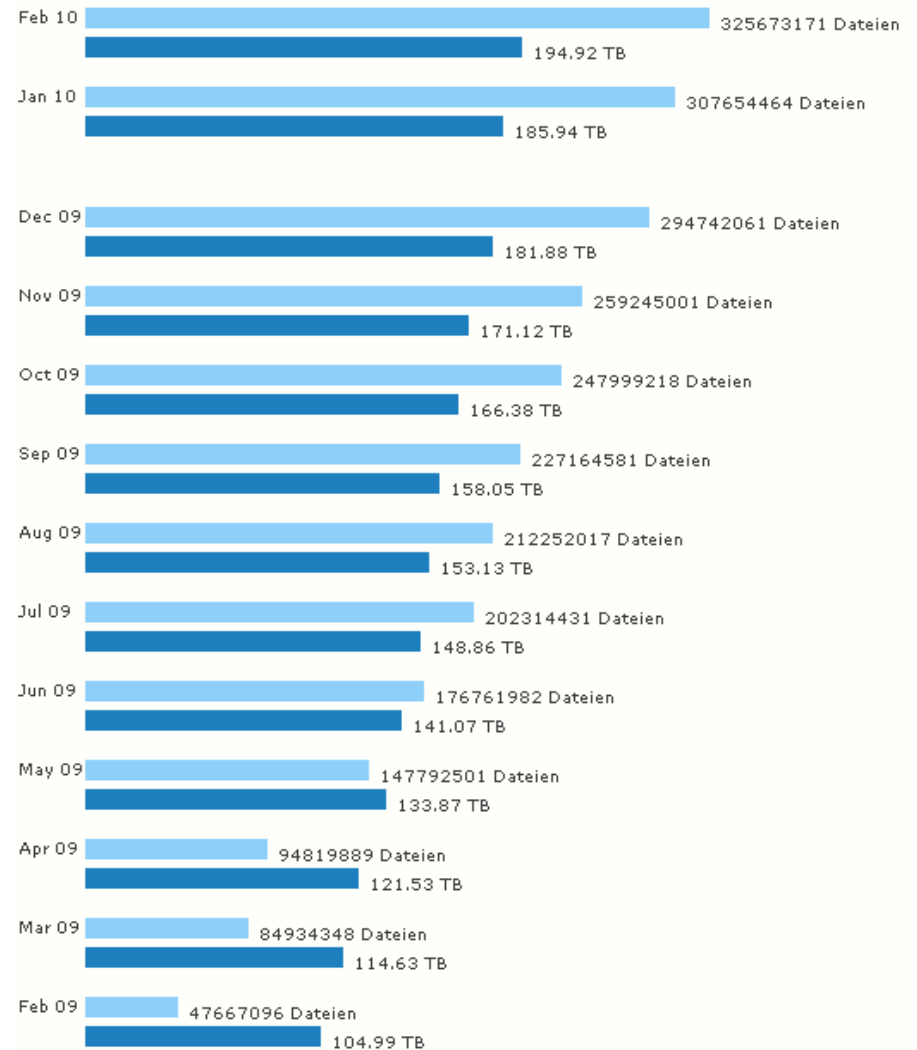
# Die Herausforderungen des Routinebetriebs in Zahlen

- Digitalisierungsaktivitäten
- Ausweitung im Bereich Pflichtablieferung und lizenzierter bzw. Open Access-Materialien
- Webarchivierung

## → Wachstum des digitalen Archivs

- 325 Mio. Files
- 194 TB
- + ca. 10 TB pro Monat

Wachstum des Archivs



# Leibniz-Rechenzentrum: Strategischer Partner

## Hochleistungsrechenzentrum für die Bayerische Akademie der Wissenschaften

### Dienstleister für

- allgemeine **IT-Services** für die Universitäten in München (mehr als 100.000 Kunden) und die Akademie
- leistungsfähige **Kommunikationsinfrastruktur**
- umfangreiche Platten- und Bandspeicher zur langfristigen **Sicherung großer Datenvolumen**
- technisch-wissenschaftliches **Hochleistungsrechenzentrum** für dt. Universitäten



## Projektaktivitäten

- **BABS I:** Entwicklung des Bibliothekarischen Archivierungs- und Bereitstellungssystems (Pilot)
- **BABS II:** Evaluation und Verbesserung des BABS-Systems im Hinblick auf Vertrauenswürdigkeit und Skalierbarkeit
- **nestor:** Deutsches Kompetenznetzwerk zur dLZA
- **LOCKSS:** Teilnahme an einem weltweiten Pilotprojekt



# Das Konzept der Vertrauenswürdigkeit

- **Was ist Vertrauenswürdigkeit?**
  - Eigenschaft eines Systems, gemäß seinen Zielen und Spezifikationen zu operieren (d.h. es tut genau das, was es zu tun vorgibt).
  - Es kann dieses glaubhaft machen.  
(RFC 4949, Internet Security Glossary; Vers.2)
  
- **Warum Vertrauenswürdigkeit in der Langzeitarchivierung?**
  - Überprüfung und Nachweis von Qualität und Nachhaltigkeit
  
- **Vier Säulen des Konzeptes in Bezug auf digitale Objekte:**
  - Integrität
  - Authentizität
  - Vertraulichkeit
  - Verfügbarkeit

# Der nestor-Kriterienkatalog und seine Anwendung in der Praxis

- Orientierungsfunktion mittels Kriterien:
  - Leitfaden, Selbstevaluierung, Befragungen, Test-Audits
- Perspektivische Entwicklung:
  - Formaler Zertifizierungsprozess
    - › Basierend auf durch Normierungsprozesse geschaffenen Standards (ISO, DIN)
- Internationale Synergien: CRL, DCC, DPE (TRAC, Drambora, Data Seal of Approval)





# Erste Ergebnisse der Selbstevaluierung im BABS2-Projekt

## Identifizierte Problembereiche

- Eingeschränkte Transparenz hinsichtlich Aufgabenumfang und –wahrnehmung
- Integritätsnachweise für die digitalen Objekte
- Eingeschränkte Ablieferungswege für externe Produzenten
- Skalierbarkeit des TSM/HSM-Archivsystems

## Ergriffene Verbesserungsmaßnahmen

- Umfangreiche Dokumentation und Prozessbeschreibung
- Formulierung einer Policy
- Zielgruppenanalyse und Vorbereitung einer Bestandsaufnahme
- Konzeption von Integritätschecks und erweiterten tech. Metadaten
- Spezifizierung und Konfiguration eines Deposit-Moduls für Produzenten
- Durchführung eines exemplarischen Preservation Planning
- Skalierbarkeit durch Migration auf NAS-Laufwerk verbessert

# Feedback für den nestor-Kriterienkatalog

## Erkenntnisgewinn für die weitere Arbeit am Kriterienkatalog

- Für die grundlegende Orientierung gut geeignet und hilfreich
- Für eine Selbstbewertung und die Vergleichbarkeit von Evaluierungsergebnissen fehlen noch die geeigneten Maßstäbe und werden in Zukunft weiter zu entwickeln sein
- Weitere Rückmeldungen über praktische Anwendbarkeit werden benötigt

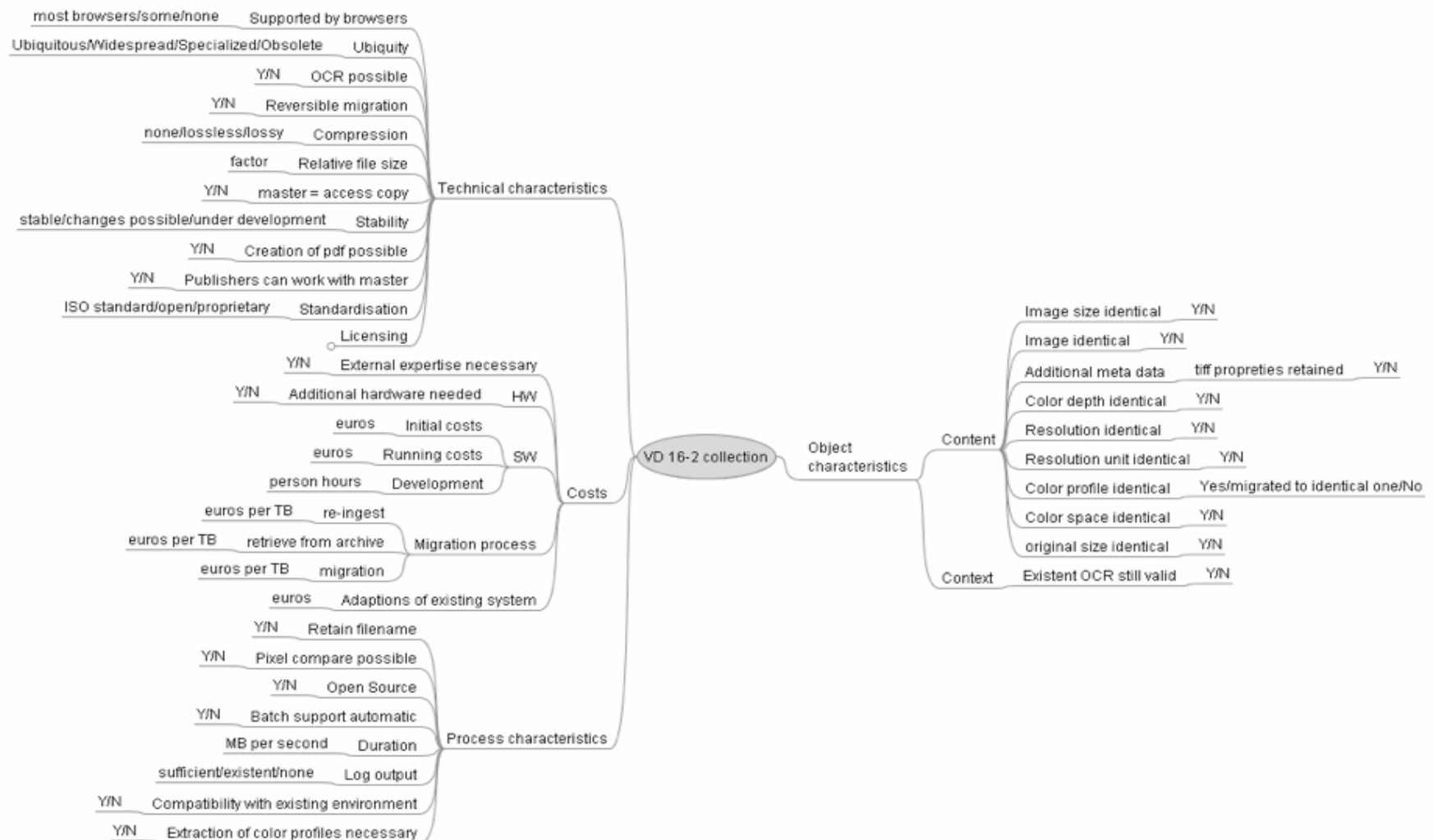


# Preservation Planning mit Plato

- **Plato:** Tool zur Unterstützung der dokumentierbaren, nachvollziehbaren und systematischen Entscheidungsfindung für das Feld des Preservation Planning mit integrierten Services für die Durchführung von Tests und zur Evaluierung der Ergebnisse
  - <http://www.ifs.tuwien.ac.at/dp/plato/intro.html>
- Gemeinsamer 2-tägiger Workshop mit Mitarbeitern der TU Wien
- **Ziele aus Sicht der BSB**
  - Gewinnung von Methodenkompetenz
  - Reflektion des eigenen Vorgehens in der Praxis
  - Entwicklung von Migrationsszenarien
- Durchführung von Tests an Beispielen aus zwei ausgewählten Kollektionen



# Anforderungen als Baumdiagramm





# PLATO: Ergebnisse



- Migration zu JPEG 2000 ist zum jetzigen Zeitpunkt noch keine Alternative für die BSB -> Wiederholung des Vorgehens nötig
- Individuelle Anforderungen -> individueller Preservation Plan
- PLATO erweist sich als geeignetes Tool für das Preservation Planning

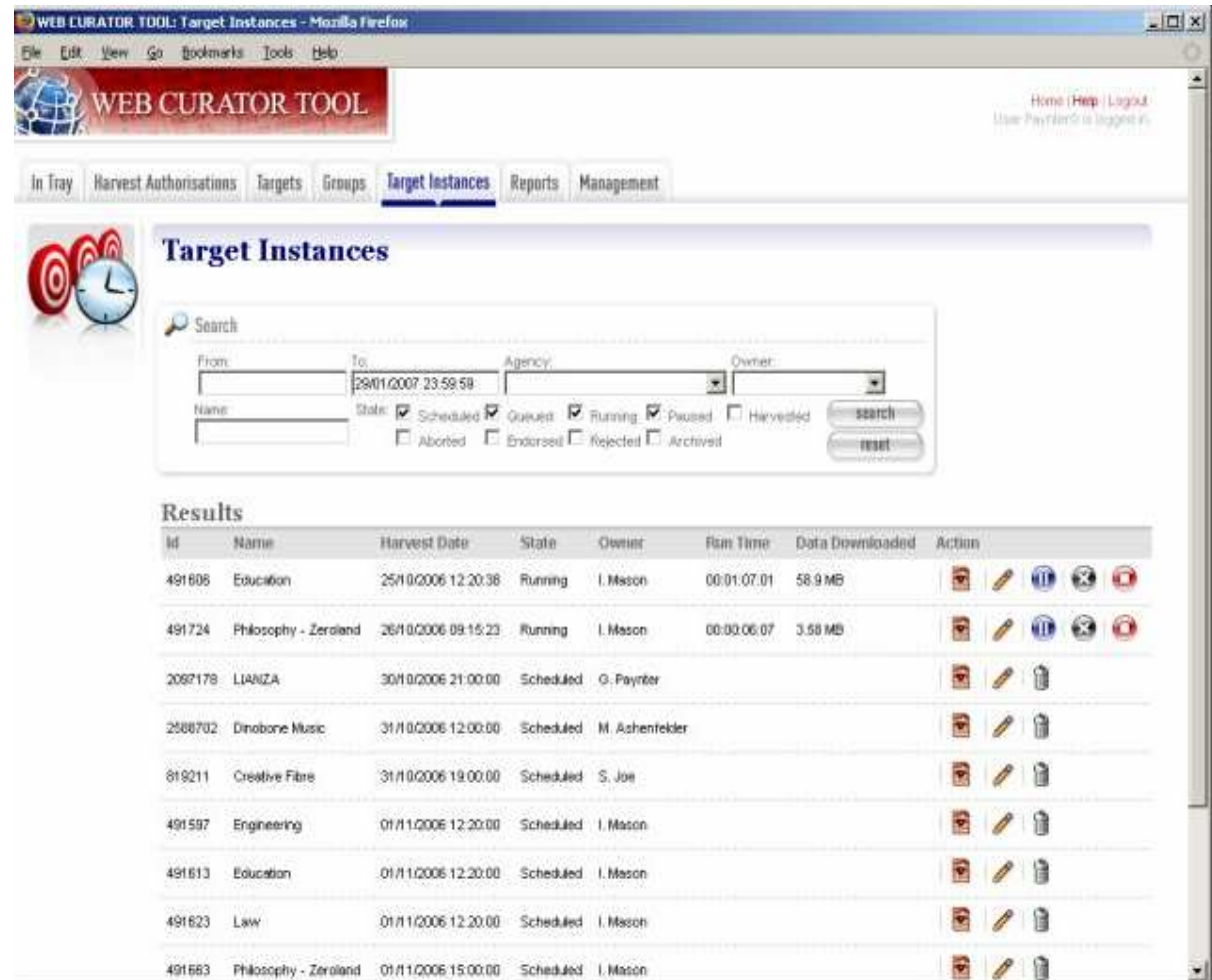
# Trainings-Events als Möglichkeiten der Aus- und Weiterbildung (und Vernetzung)

- **Planets-Training Events**
  - 19.-21. April 2010  
in Rom
- **Nestor Schools**
  - 14.-18. Juni 2010  
in Staufen



# Website-Archivierung mit dem Webcurator-Tool

- 2-monatige Implementierungsphase
- Konfiguration für den Regelbetrieb wird derzeit getestet
- Umfangreiche Funktionalitäten (end-to-end)



# JHOVE, DROID, FITS

- **Ziel: Integrität der digitalen Dokumente sichern**  
=> Dateiformat-Charakterisierung: Validierung, Extraktion und Speicherung technischer Metadaten der digitalen Dokumente
- OpenSourceTools im Test:
  - Jhove
  - DROID
  - Metadata Extraction Tool
  - **FITS**
  - XCL
- Perspektive: Risk-Management mit Drambora



## Synergieeffekte: Output

- Publikationen (dlib-Magazin, nestor-Materialien)
  - <http://www.dlib.org/dlib/november09/kulovits/11kulovits.html>
  - <http://www.langzeitarchivierung.de/publikationen/index.htm>
- Workshops
  - BABS2: LZA von Digitalisaten (ca. 12/2010)
  - nestor-Workshops (t.b.a.)
- Community-Aktivitäten / Vernetzung (nestor, DIN, ISO) / Informationsaustausch
- Beratung
- Beteiligung an Community-Aktivitäten im Rahmen von Rosetta

**Vielen Dank! Fragen?**

**Kontakt: [beinert@bsb-muenchen.de](mailto:beinert@bsb-muenchen.de)**

