

Repositorien und Langzeitarchivierung aus der Sicht der Bibliotheken

Repositorien – Workshop der Arbeitsgruppe Elektronisches
Publizieren der deutschen Akademien der Wissenschaften
Düsseldorf, 4. bis 6. Oktober 2010

Natascha Schumann (DNB/nestor)
Stefan Strathmann (SUB Göttingen/nestor)

Ziele und Aufgaben von nestor

- .. nestor bündelt das vorhandene Know-How und die Kompetenzen im Bereich der digitalen Langzeitarchivierung in Deutschland
- .. 2003-2009 in zwei Projektphasen vom BMBF gefördert
- .. Seit 2009: Kooperationsverbund
- .. 11 Partner aus den Bereichen Archiv, Museum, Bibliothek und Universitäten...



VERANSTALTUNGEN

Workshops
nestor-Vorträge



STANDARDISIERUNG

DIN-Normen

INTERNATIONALE KOOPERATION

Alliance for Permanent Access
DPC
NDDC
NDIIPP
IFLA
EU-Projekte

AG Media
AG Recht
AG Kooperation & Vernetzung
AG Digitale Bestandserhaltung
AG Emulation

nestor



VERNETZUNG

INFORMATIONEN- ANGEBOTE

News
Kalender
Personen
Projekte
Informations-Datenbank



QUALIFIZIERUNG

nestor-Schools
eTutorials
nestor-Handbuch

PUBLIKATIONEN

Kriterienkataloge
Handbuch
Ratgeber
Expertisen
Infosheets
nestor-edition
Newsletter

- Home
- Arbeitsgruppen
- Publikationen
- Informationsdienste
- Services
- Schwerpunkte

Willkommen bei nestor

Willkommen bei nestor, dem deutschen Kompetenznetzwerk zur digitalen Langzeitarchivierung. In nestor arbeiten Bibliotheken, Archive, Museen sowie führende Experten gemeinsam zum Thema Langzeitarchivierung und Langzeitverfügbarkeit digitaler Quellen. nestor ist ein Kooperationsverbund mit Partnern aus verschiedenen Bereichen, die alle mit dem Thema "Digitale Langzeitarchivierung" zu tun haben.

▶ nestor Partner



■ QUALIFIZIERUNG

nestor kooperiert mit Hochschulpartnern, die Aus-, Fort- und Weiterbildungsangebote im Bereich der digitalen Langzeitarchivierung in Deutschland entwickeln.



■ VERNETZUNG

nestor ist das nationale Kompetenznetzwerk für digitale Langzeitarchivierung, eingebunden in ein Netz europäischer Partner.



■ STANDARDISIERUNG

nestor bündelt Standardisierungsaktivitäten und vermittelt Standards in Anwender-Communities.



..■ Catalogue of Criteria for Trusted Digital Repositories, Version 2

Die englische Fassung des nestor Kriterienkatalogs Vertrauenswürdige digitale Langzeitarchive ist in der überarbeiteten Fassung erschienen....
[mehr]

..■ Aufruf zur Beteiligung am Normungsprojekt des Arbeitskreises "Rechtsichere Aufbewahrung von digital signierten Dokumenten"...
[mehr]

..■ Neuer DPC Technology Watch Report "File Formats for Preservation": Die Digital Preservation Coalition (DPC) hat einen neuen Technology Watch Report zum Thema: File Formats for Preservation veröffentlicht... [mehr]

Informationsangebote

- ..■ Informationsdatenbanken mit Artikeln, Konferenzankündigungen, Projektbeschreibungen, etc
- ..■ Personendatenbanken mit Kontaktmöglichkeiten
- ..■ News
- ..■ Newsletter
- ..■ Veranstaltungskalender

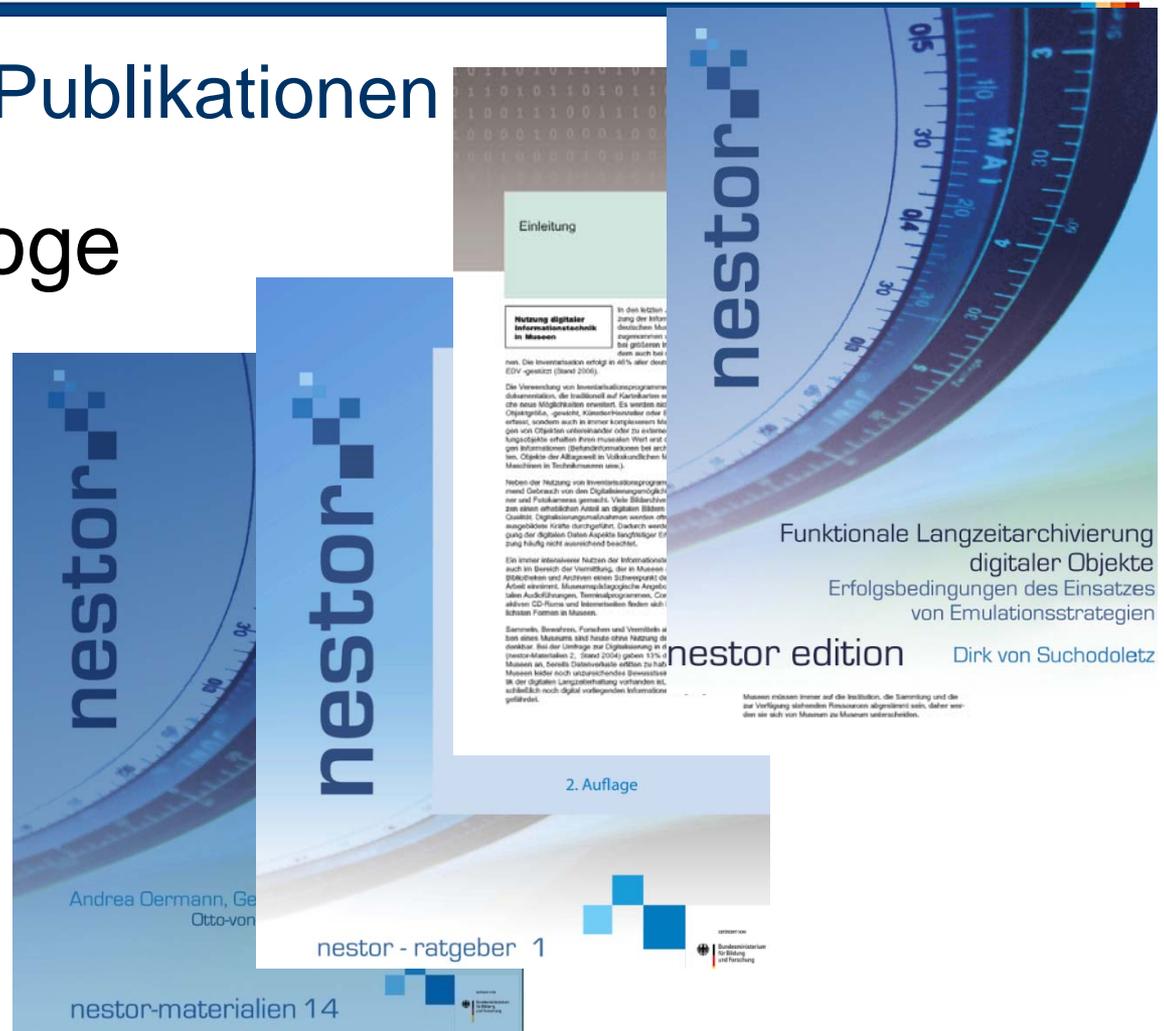
nestor Handbuch

- ..■ Alle relevanten Bereiche
- ..■ Online Version 2.3 frei zugänglich
- ..■ Version 2.0 als gedruckte Ausgabe
- ..■ Permanente Weiterentwicklung
- ..■ Autoren: Experten aus verschiedenen Bereichen



Materialien/Publikationen

- Kriterienkataloge
- Ratgeber
- Infosheets
- nestor edition



Vernetzung durch Arbeitsgruppen

- ..■ Ehemalige AGs sind in DIN- Arbeitskreisen aufgegangen und bereiten dort Normierungsverfahren vor
- ..■ Aus der Task Force Recht der AG Vernetzung & Kooperation ist die AG Recht geworden
- ..■ Neu-Strukturierung der AGs Media und Vernetzung & Kooperation
- ..■ Neu eingerichtet wurde die AG Digitale Bestandserhaltung
- ..■ Ganz neue eingerichtet wurde die AG Emulation

Kooperationen

- ..■ Die meisten nationalen Projekte sind mit nestor auf die eine oder andere Weise verbunden
- ..■ nestor-Partner sind in verschiedenen Projekten aktiv
- ..■ Zusammenarbit mit Initiativen auf europäischer und internationaler Ebene:
 - ..■ Austausch
 - ..■ Gemeinsame Veranstaltungen

Qualifikation

- ..■ nestor Schools
- ..■ nestor Handbuch
- ..■ Zusammenschluss mehrerer deutscher und deutschsprachiger Hochschulen auf Grundlage eines MoU mit dem Ziel, dLZA in der Ausbildung stärker zu verankern
 - ..■ Entwicklung von e-Tutorials
 - ..■ Kooperative Entwicklung eines gemeinsamen Curriculums

Standardisierung

- ..■ DIN Arbeitskreis Normenausschuss Bibliotheks- und Dokumentationswesen (NABD), Schriftgutverwaltung und Langzeitverfügbarkeit digitaler Informationsobjekte
- ..■ Vorbereitung von nestor-Kriterienkatalogen zur Normung:
 - ..■ Kriterienkatalog vertrauenswürdige digitale Langzeitarchive
 - ..■ Ingest-Leitfaden: Wege ins Archiv
 - ..■ Kriterienkatalog Vertrauenswürdige Persistent Identifier Systeme

Norm-Entwürfe

- ..■ DIN 31644 - Kriterien für vertrauenswürdige digitale Langzeitarchive
- ..■ DIN 31645 - Leitfaden zur Informationsübernahme in digitale Langzeitarchive
- ..■ DIN 31646 - Anforderungen an die langfristige Handhabung persistenter Identifikatoren

Vertrauenswürdige Archive

- ..■ Verschiedene Initiativen neben nestor/DIN
 - ..■ CCSDS/ISO Repository Audit and Certification Working Group (RAC),
 - ..■ Data Seal of Approval DSA
- ..■ Bestrebungen der EU diese zu harmonisieren
 -  Memorandum of Understanding (Juli 2010)

Memorandum of Understanding

- ..■ 3-Stufige Zertifizierung:
 - ..■ Basic Certification: Self-Audit DSA,
 - ..■ Extended Certification: Zusätzlich zum DSA Self-Audit nach ISO 16363 oder DIN 31644,
 - ..■ Formal Certification: Externes Audit nach ISO 16363 oder DIN 31644
- ..■ Testphase Extended Certification bis Ende des Jahres

Repositorien und Langzeitarchivierung aus der Sicht
der Bibliotheken

-

Teil II: Die Rolle Bibliothek im Forschungsdatenmanagement

Datengetriebene Wissenschaft und Nachnutzung

In den Wissenschaften werden riesige Datenmengen produziert:
oft teuer (mehrere Milliarden €/Jahr allein in D), oft aufwändig und
oft nicht reproduzierbar (z.B. Wetterdaten)

Datenmenge nicht (nur) Problem, sondern auch Chance: Neue
Fragestellungen werden möglich, z.B. quantitative Analysen großer
Textkorpora

Nachnutzung für z.T. unvorhergesehene Zwecke: Verknüpfung von
Google Earth und epidemiologischen Daten, Re-Evaluierung „alter“
Daten mit neuen Erkenntnissen

Datengetriebene Wissenschaft und Nachnutzung

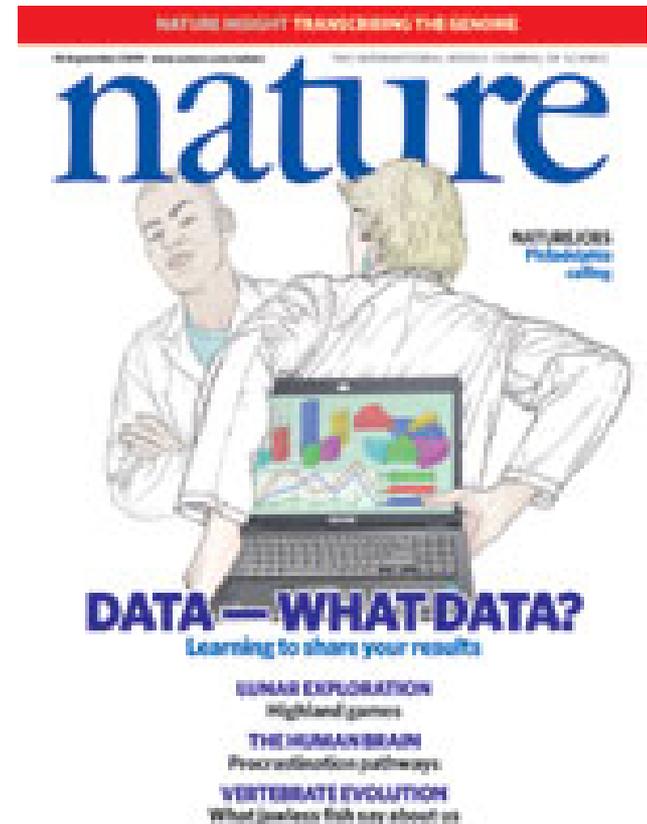
„Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly.“

Teaser zum Editorial:

Nature 461,

145 (09. September 2009),

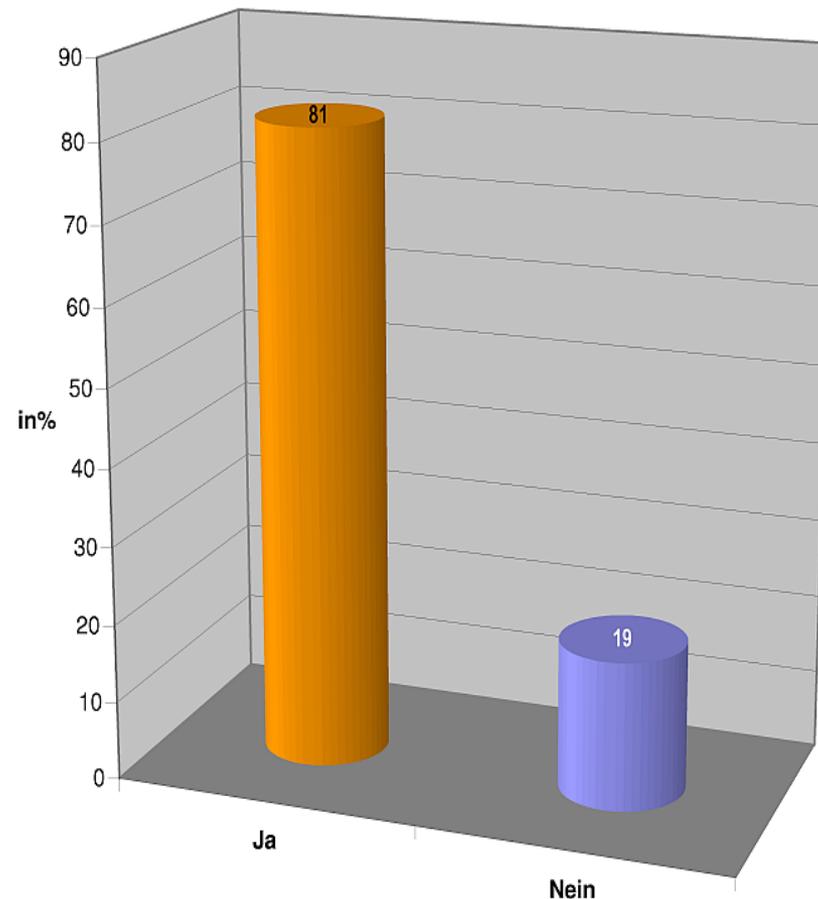
doi:10.1038/461145a



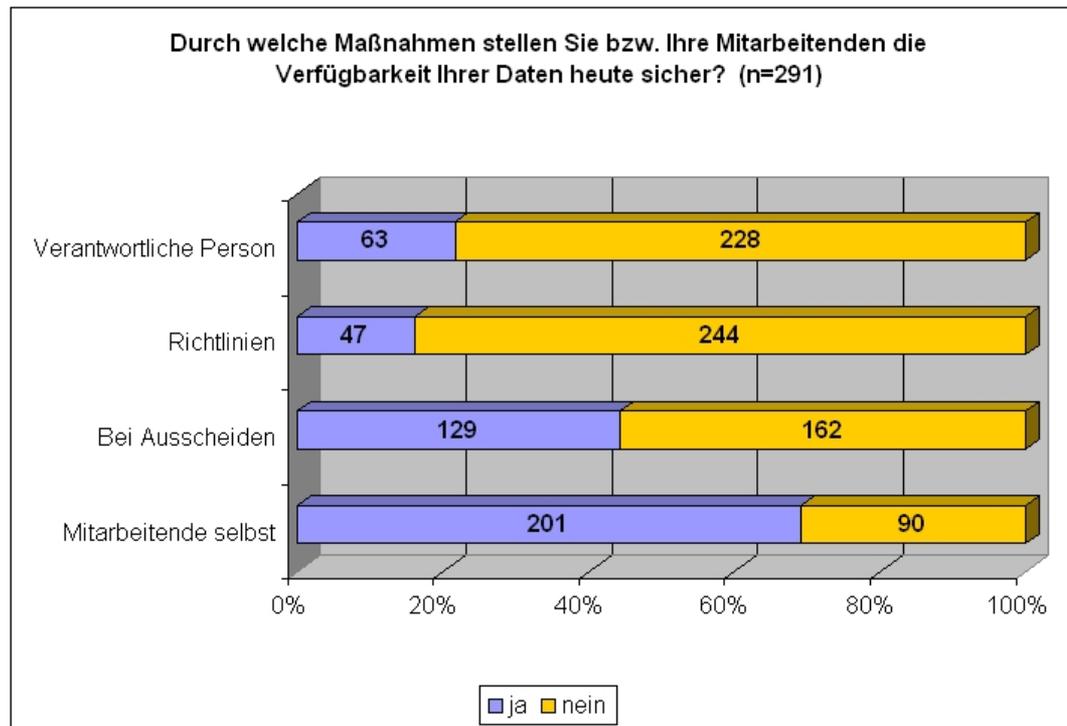
nestor Online-Umfrage

- ..■ Bedarfsabklärung zur Langzeitarchivierung (LZA) digitaler Daten an der Universität Göttingen
- ..■ Evaluierung des Ist-Standes
- ..■ In fast identischer Form an der **Universität Göttingen** und an der **ETH Zürich** durchgeführt
- ..■ 291 für die Auswertung nutzbare Antworten an der Universität Göttingen

„Gibt es Daten, die über den aktuellen Gebrauch hinweg erhalten bleiben sollen?“

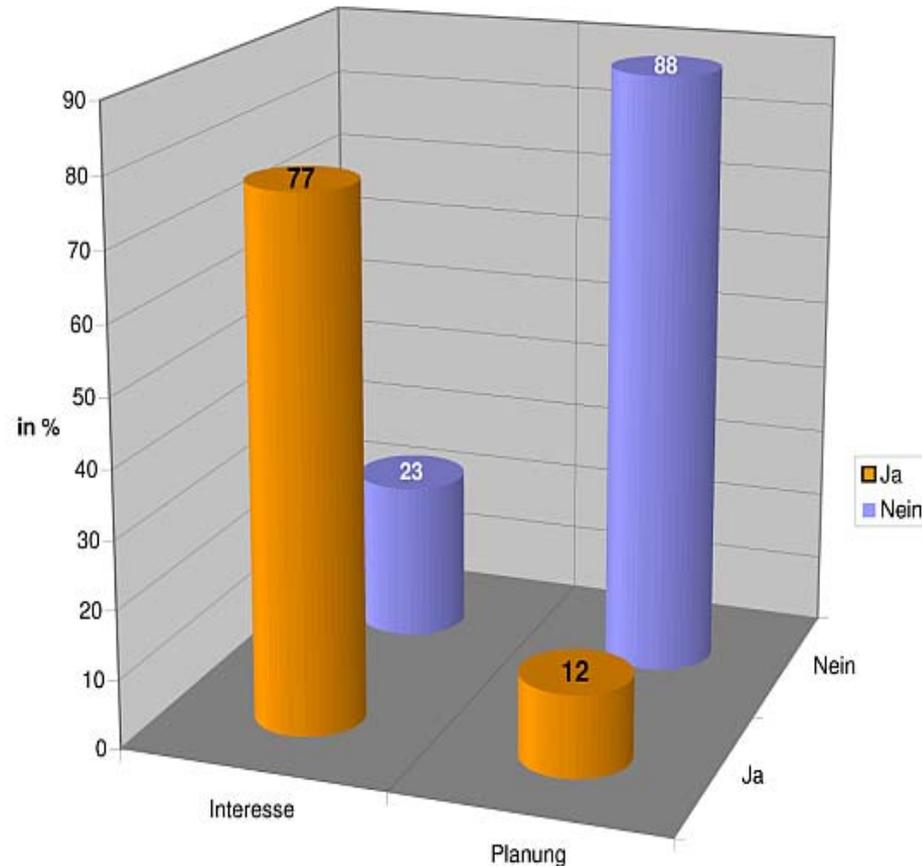


Verantwortlichkeit für Forschungsdaten



- In über 60% der Fälle sind die Wissenschaftler selbst verantwortlich
- Richtlinien und zentral verantwortliche Personen existieren nur bei ca. 20% der Befragten

Interesse an & Planung von Maßnahmen zur digitalen Langzeitarchivierung



Ergebnisse

- ..■ Beachtliches Interesse an Langzeitarchivierung
- ..■ Kaum konkrete Planungen
- ..■ Verantwortung für Daten liegt beim Wissenschaftler
- ..■ Geringe Policy-Kenntnisse
- ..■ Stark heterogene, oft auch proprietäre Formate
- ..■ Geringe Metadaten-Implementierung (v.a. technische Metadaten und Kontext-Metadaten)
- ..■ Starke Heterogenität bei Datenumfang und -typ von Forschungsdaten

Warum ist dies so?

- Forscher sehen sich als nicht verantwortlich für Datenpflege wegen...
 - ...fehlender Zeit
 - ...“Datenpflege“ keine Forschungstätigkeit
 - ...fehlende Anreize, fehlendes Bewußtsein
- Publikation von Forschungsdaten wird bisher nur in Ausnahmefällen honoriert
- **Notwendig: Forschungsdaten-Infrastruktur (am Standort Göttingen) unter Einbeziehung der Forschergruppen mit verbindlichen Richtlinien, Workflows und technologischem Rahmenkonzept**

Konzeptionelle Bausteine

Data Management Guides

Kriterienkataloge für vertrauenswürdige Archive
(insbesondere organisatorisch)

Metadatenstandards

Zur Verwaltung und Nachnutzung von Daten (technisch,
semantisch, Provenienz, ...)

Referenzmodelle, Befragungen, Studien, Roadmaps, ...

Technologische Bausteine

Archivsysteme

meist an Dokumenten orientiert, für Forschungsdaten
eher in der Entwicklung...

Persistent Identifier

Werkzeuge zur technischen Qualitätssicherung und
Metadatenerzeugung

Projekte zur sicheren verteilten Speicherung

Organisatorische Bausteine

Netzwerke

hps. national, z.B. im Bereich der Bibliotheken, Archive, Museen (nestor) und in einigen Fachdisziplinen

Ausbildung, Qualifizierung

„zarte“ Pflänzchen

Finanzierung und Organisation

einige Studien und Modelle

Kooperation

auf Projektebene ausgeprägt, einzelne Verstetigungen

Lokale Kompetenzen

GWDG: Rechenzentrum der MPG und der Universität

UMG: Universitätsmedizin mit eigenem Rechenzentrum

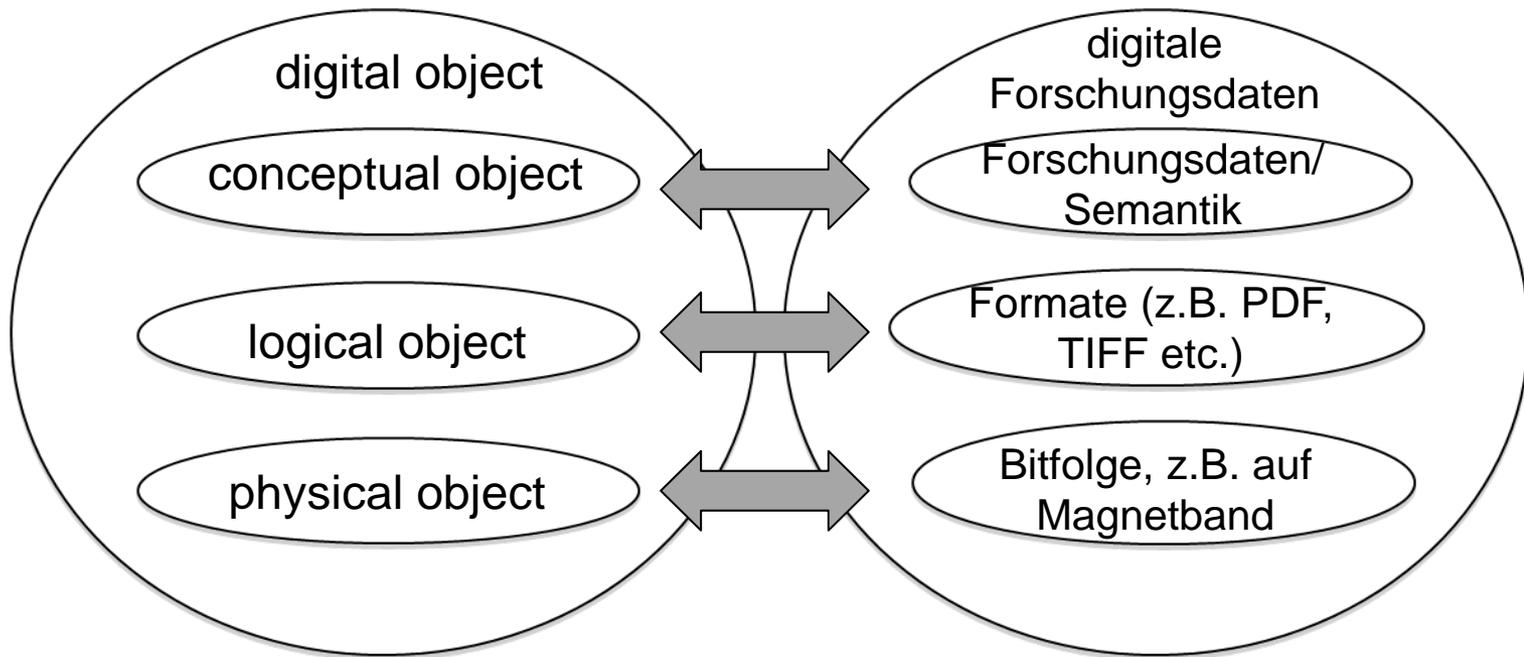
SUB: Universitätsbibliothek

Vielzahl von Projekten zu Langzeitarchivierung,
Forschungsdaten, Virtuelle Forschungsumgebungen und
Grid (Basis-Technologie)

Insbesondere:

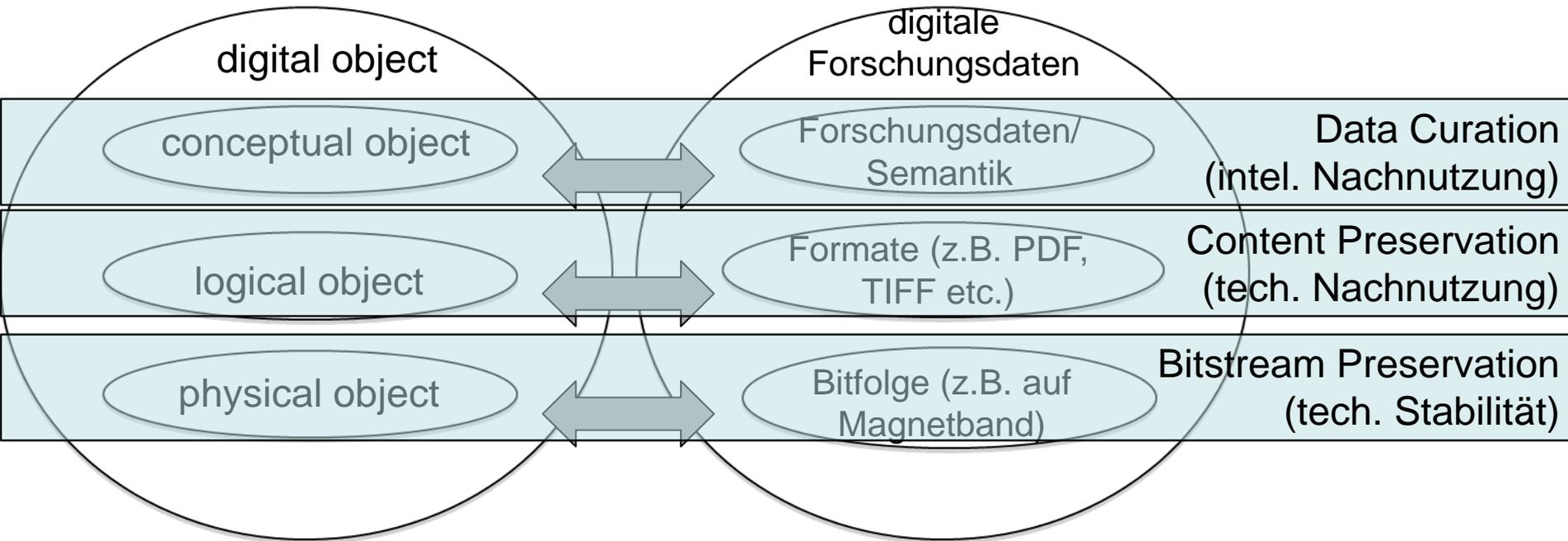
- PARSE.Insight (EC): SUB
- WissGrid (BMBF): UMG und SUB
- koLaWis (DFG): GWDG, SUB, UMG

Ausgangsmodell



Angelehnt an Thibodeau: Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years, 2002.

<http://www.clir.org/pubs/reports/pub107/thibodeau.html>



- ..■ Data Curation: intellektuelle Nachnutzbarkeit
 - ..■ Kontextinformationen, Objektmodelle, Versionierungen, ...
- ..■ Content Preservation: technische Nachnutzbarkeit
 - ..■ technische Qualitätskontrollen, Konvertierungen, ...
- ..■ Bitstream Preservation: technische Stabilität
 - ..■ genug unabhängige Kopien, Integritätsprüfung, ...

Kompetenzen für Langzeitarchivierung

- Benötigt Disziplin-Know-How, versteht wissenschaftliche Workflows und Daten, wissenschaftsnahe Institution oder Teil der Community
 - Experten für LZA-/Data Management Technik. Vermittelt zwischen Akteuren in Data Curation und Bitstream Preservation, versteht Authentizitäts und Integritätsanforderungen
 - Rechenzentren, technische Expertise und Massenbetrieb
- Data Curation (intellektuelle Nachnutzung)
- Content Preservation (technische Nachnutzung)
- Bitstream Preservation (technische Stabilität)

Vorgehen und Ausblick

- ..■ Ausbau der Kontakte zu Fakultäten, SFBs und Fachwissenschaftlern
- ..■ Gründung einer AG mit Vertretern der verschiedenen „Player“ (Präsidium, Infrastruktureinrichtungen, Wissenschaftler ...)
- ..■ Beratung der Fachwissenschaftler
- ..■ Unterstützung der Wissenschaftler bei der Beantragung von Drittmitteln für Forschungsdatenmanagement
- ..■ Gemeinsame Projektanträge der Infrastruktureinrichtungen mit den Fachwissenschaftlern
- ..■ Einrichtung einer Stelle zur Mitarbeit bei „ der Entwicklung eines fachübergreifenden Konzeptes zur Archivierung und zum Datenmanagement von Forschungsdaten für die Universität Göttingen“

- ..■ Herzlichen Dank für Ihre Aufmerksamkeit!
- ..■ Fragen?
 - ..■ Natascha Schumann: n.schumann@d-nb.de
 - ..■ Stefan Strathmann: strathmann@sub.uni-goettingen.de