

# Übersicht zu verschiedenen Perspektiven der Langzeitarchivierung

Offener Workshop zum nationalen Metadaten  
Repository

23.09.2011, TMF, Berlin

Frank Dickmann

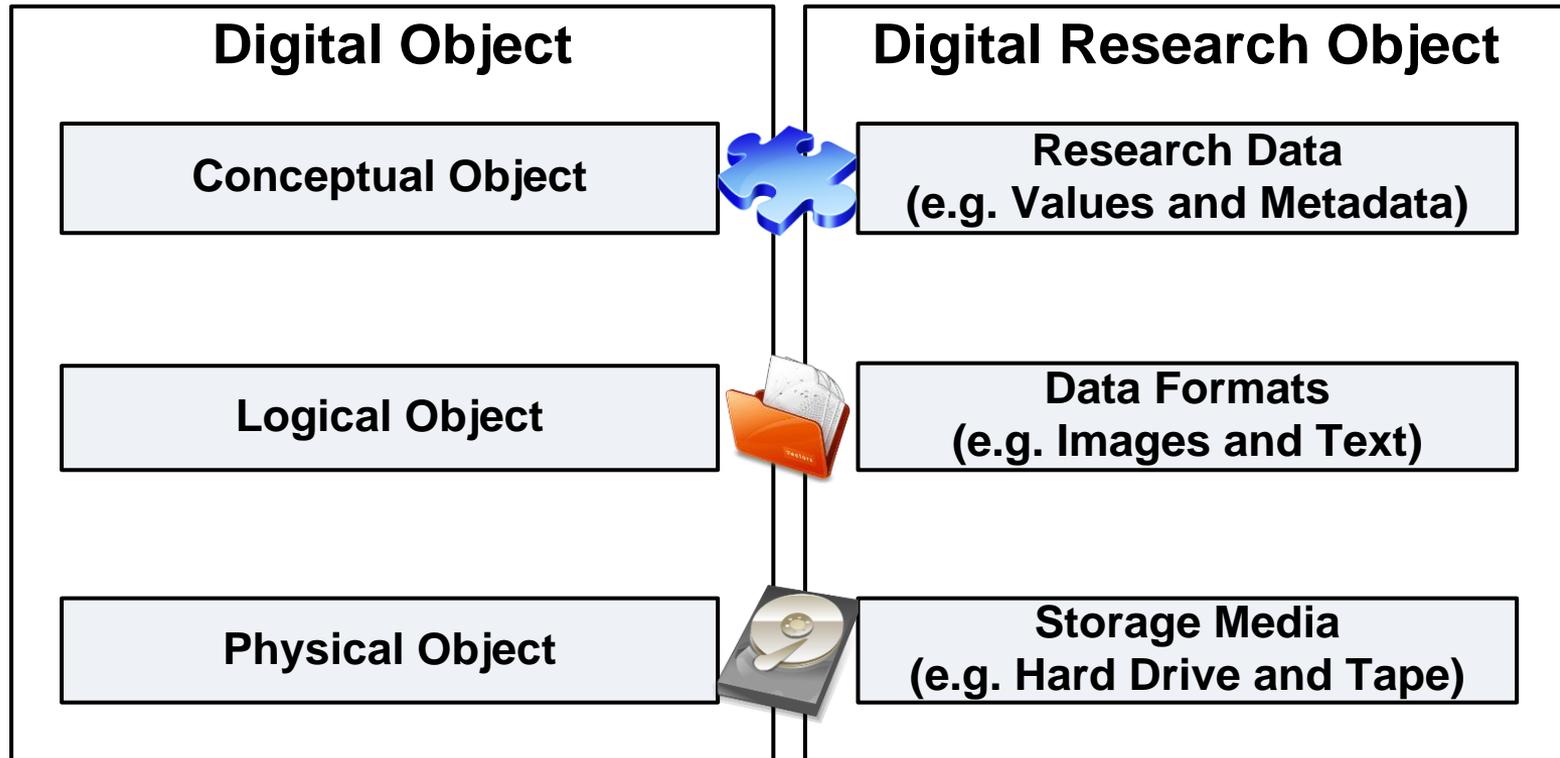


Bundesministerium  
für Bildung  
und Forschung

**DFG**

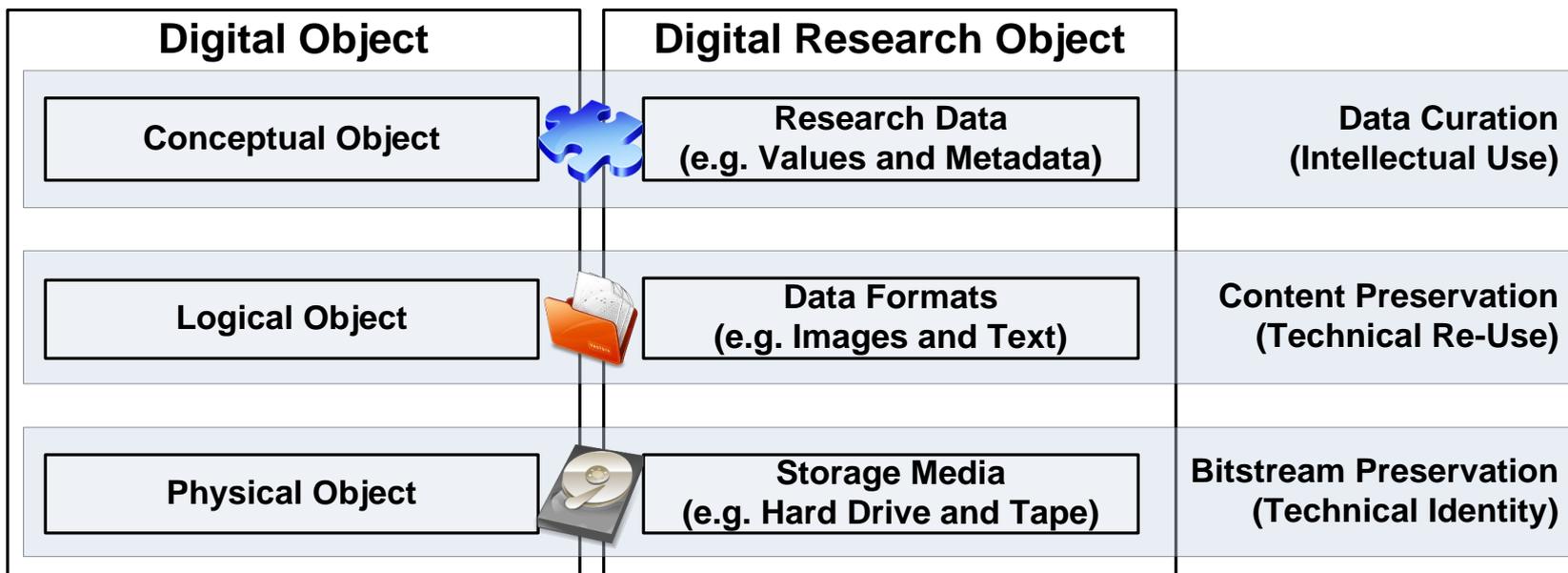
- Getrieben durch Bibliothekswesen
- Es ist kein „dark archive“ gemeint
- Der Fokus liegt auf Forschungsdatenmanagement
  - „Datenkuration“ / Metadaten(-pflege)
  - Sicherstellen logischer & intellektueller „Lesbarkeit“
  - Erhalten der physikalischen Speicherung
- Zentrale Aspekte sind
  - Zeithorizont: 5+ Jahre
  - Verfügbarkeit/Zugriffsmöglichkeit
  - Fachwissenschaftliche Anforderungen
  - Metadaten und Datenformate
  - Nachhaltige Finanzierung

# Ebenen der Langzeitarchivierung



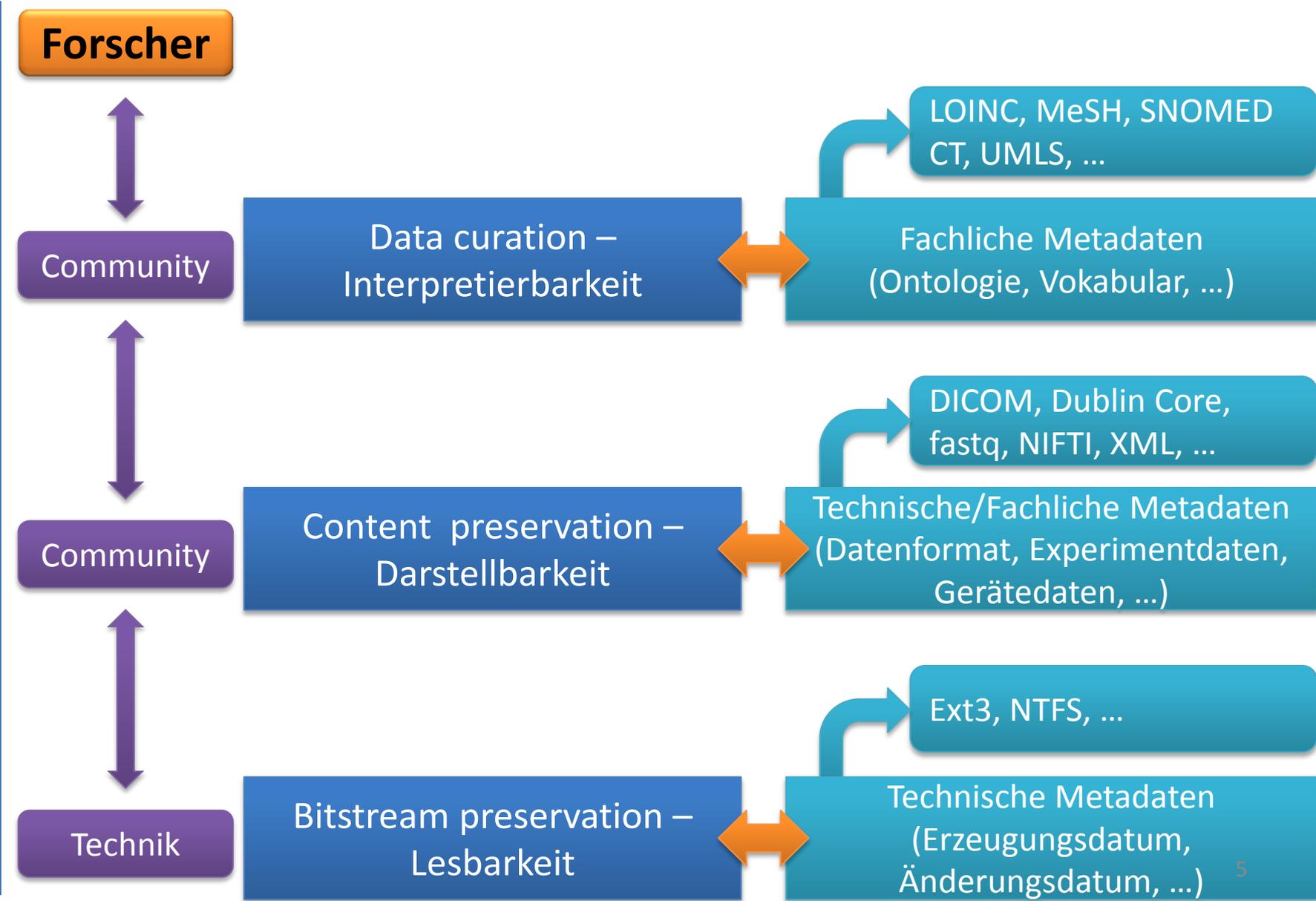
Angelehnt an Thibodeau: Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years, 2002.  
<http://www.clir.org/pubs/reports/pub107/thibodeau.html>.

# Ebenen der Langzeitarchivierung

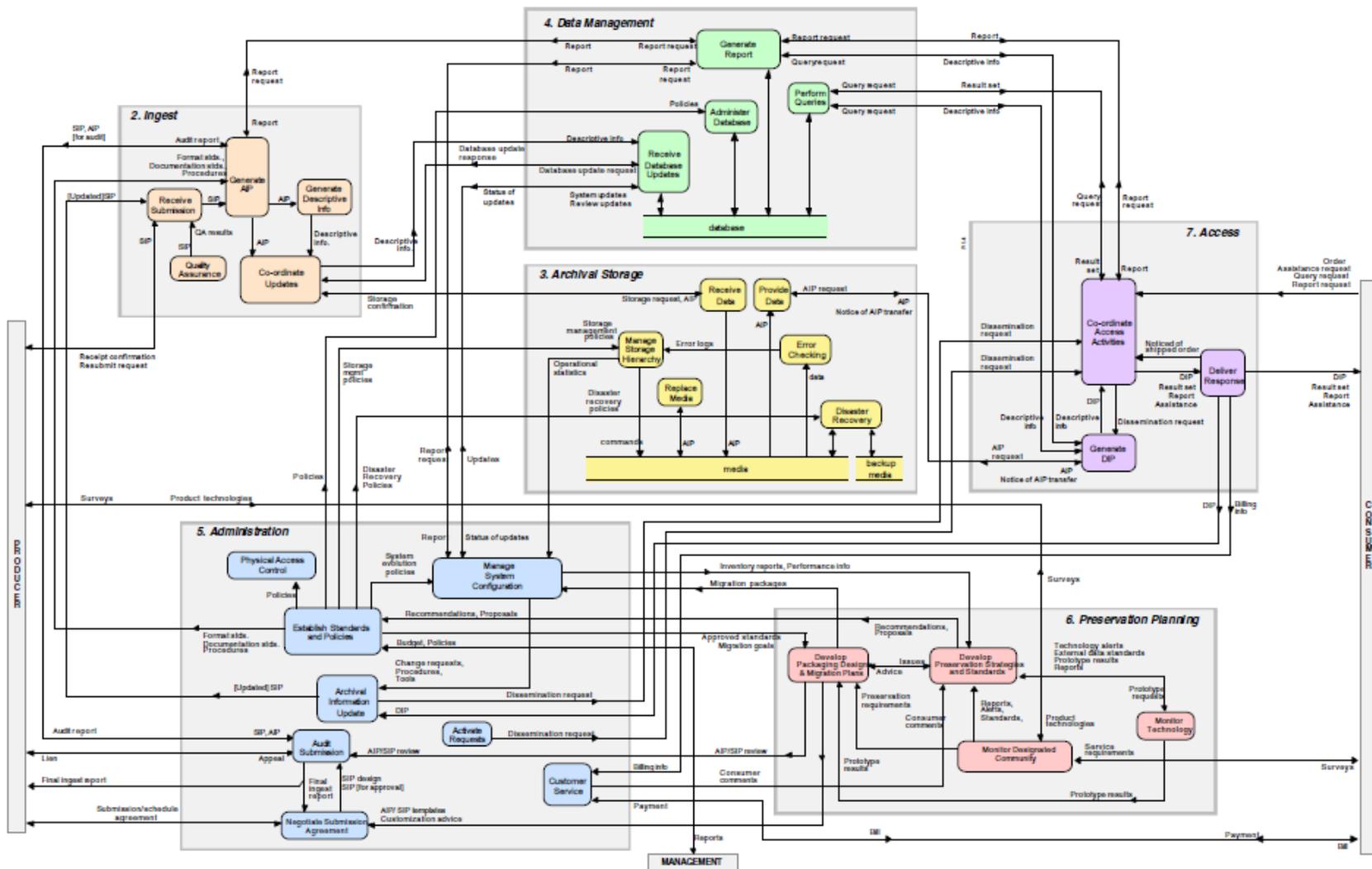


Modifiziert übernommen von Ludwig, J., Long-term Preservation of Digital Research Data. 2009, DESY: Hamburg, Germany.  
 URL: [http://www.desy.de/dvsem/WS0910/ludwig\\_talk.pdf](http://www.desy.de/dvsem/WS0910/ludwig_talk.pdf).

# Langzeitarchivierung und Metadaten

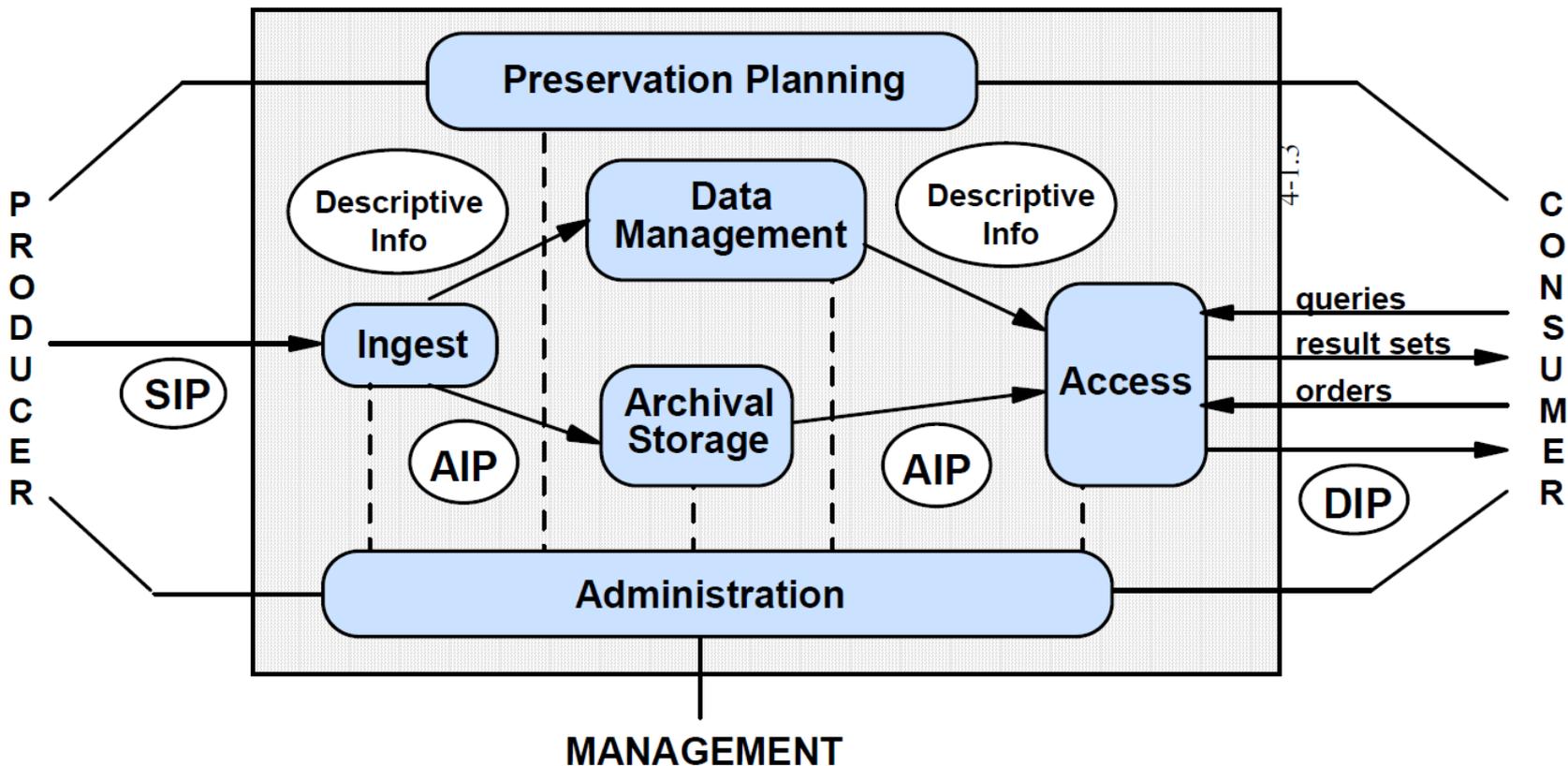


# Generisches Langzeitarchivierungskonzept - OAIS



Quelle: CCSDS (Consultative Committee for Space Data Systems) (2002): Reference Model for an Open Archival Information System (OAIS). Blue Book 1, Washington, DC, CCSDS Secretariat, 2008.08.25, URL: <http://public.ccsds.org/publications/archive/650x0b1.pdf>

# Generisches Langzeitarchivierungskonzept - OAIS



- Dublin Core Metadata Initiative
- Ein Projekt der Online Computer Library Center, Inc. (OCLC) und des National Center for Supercomputing Applications (NCSA)
- Gegründet 1994
- Zielsetzungen
  - Beliebige Objekte im WWW wiederauffindbar zu machen
  - Interdisziplinärer und internationaler Kern für Metadaten
- Als Standard definiert u.a. als ISO Standard 15836:2009

## 15 zentrale Bestandteile / „element set“ (erweiterbar):

1	Title	given name
2	Creator	responsible entity
3	Subject	topic of resource
4	Description	abstract or similar
5	Publisher	responsible entity for publication
6	Contributor	contributing entity
7	Date	point in time of the lifecycle
8	Type	nature or genre
9	Format	file format, physical medium
10	Identifier	unique reference to the resource
11	Source	related sources of derivation
12	Language	resource language
13	Relation	related resource
14	Coverage	spatial / temporal applicability
15	Rights	rights in and over the resource

Quelle: American National Standards Institute (ANSI) und National Information Standards Organisation (NISO) (2007): The Dublin Core Metadata Element Set. Bethesda, MD, USA, NISO Press, Z39.85-2007, Letzter Zugriff: 2011.09.20, URL: [http://www.niso.org/apps/group\\_public/download.php/6576/The%20Dublin%20Core%20Metadata%20Element%20Set.pdf](http://www.niso.org/apps/group_public/download.php/6576/The%20Dublin%20Core%20Metadata%20Element%20Set.pdf).

1. ID (identifizier)
2. Technische Metadaten (format, type, language)
3. Inhaltliche Beschreibung (title, subject, coverage, description)
4. Rechte, Rollen und Personen (creator, publisher, contributor, rights)
5. Vernetzung von Objekten (source, relation)
6. Lebenszyklus (date)

- LABIMI/F
- nestor
- KoLaWiss
- WissGrid

- „Langzeitarchivierung biomedizinischer Forschungsdaten“
- DFG-Förderung seit 06/2011 bis 06/2013
- Zielsetzungen
  - Präzisieren von Anforderungen an die Langzeitarchivierung in der Biomedizin
  - Entwickeln einer Laborimplementierung für Langzeitarchivierungsdienste
  - Definieren von einheitlichen Metadaten und Datenformaten für zwei Use Cases
  - Entwickeln eines Betriebsmodells
  - Entwickeln eines Finanzierungsmodells

- Use Cases
  - Medizinische Bilddaten, Magdeburg
  - Transkriptomdaten, Kiel
- Workshops für beide Use Cases Anfang 2012
  - Anforderungen weiter verfeinern
  - Einbinden von „Endnutzern“ → Forschern
- Aufbau von Langzeitarchivierungsdiensten, Göttingen
  - Ebene der „data curation“ / Forschungsdaten
  - Verwenden von Grid-Technologie → verteilte Infrastruktur
  - Gegenwärtig werden vorhandene Lösungen evaluiert

- “Network of Expertise in long-term Storage and availability of digital Resources in Germany”
- Förderung durch das BMBF von 2003 bis 2009
- Seit 2009 durch Eigenleistung weitergeführt
- Zielsetzungen / Schwerpunkte
  - **Standardisierung** → Persistent Identifier, Datei-Formate, Metadaten, Referenzmodelle (z.B. OAIS)
  - **Qualifizierung** → Kooperation mit Hochschulen für Weiterbildung und Publikationen
  - **Vernetzung** → Kooperation von Arbeitsgruppen und Projekten

- Arbeitsgruppen
  - Media (nicht-textuelle Medien)
  - Recht (rechtliche Rahmenbedingungen)
  - Digitale Bestandserhaltung
  - Vertrauenswürdige Archive – Zertifizierung
  - Langzeitarchivierungsstandards
  - Grid / eScience und Langzeitarchivierung
- Wichtige Publikationen
  - nestor Handbuch
  - In Vorbereitung:  
„Langzeitarchivierung von Forschungsdaten“

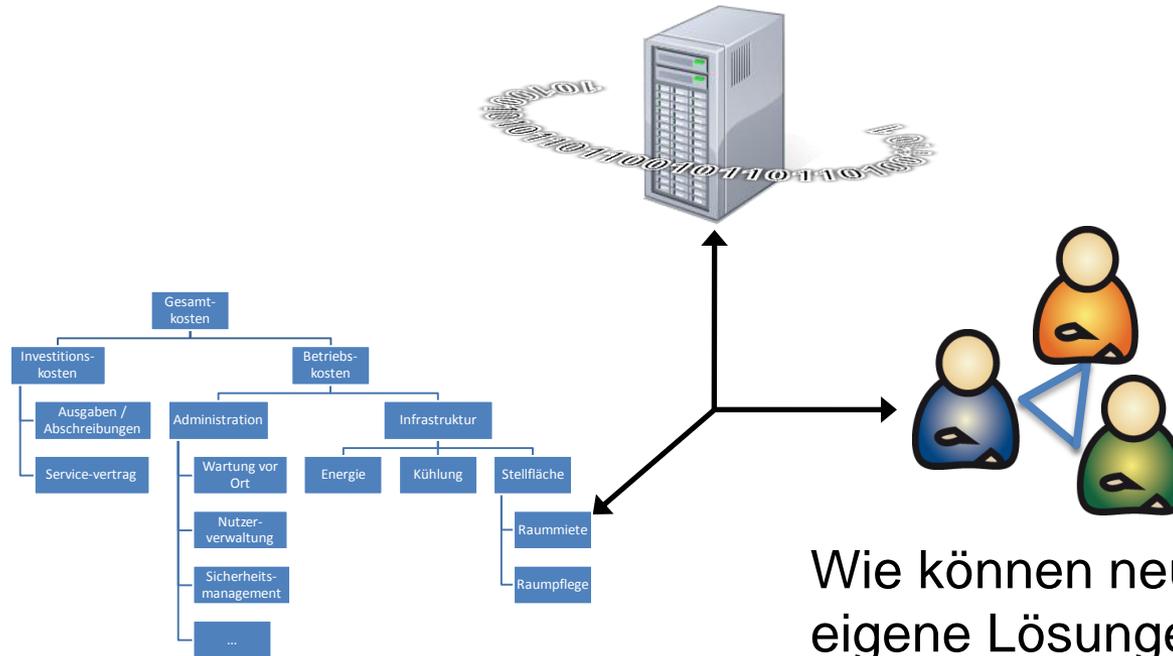


- „Kooperative Langzeitarchivierung für Wissenschaftsstandorte“
- DFG-Förderung von 02/2008 bis 02/2009
- Zielsetzungen
  - Entwickeln eines **Organisations- und Geschäftsmodells** für eine kooperative Langzeitarchivierung → am Beispiel Göttingen
  - Erarbeiten bundesweiter **Förderempfehlungen** für eine Umsetzung der Organisations- und Geschäftsmodelle von Langzeitarchivierungsknoten für die Forschung
- Arbeitspakete
  - Technik / Recht / Kosten / Organisation und Fördermaßnahmen

## Spezifische Inhalte „Biomedizinische Forschungsdaten“

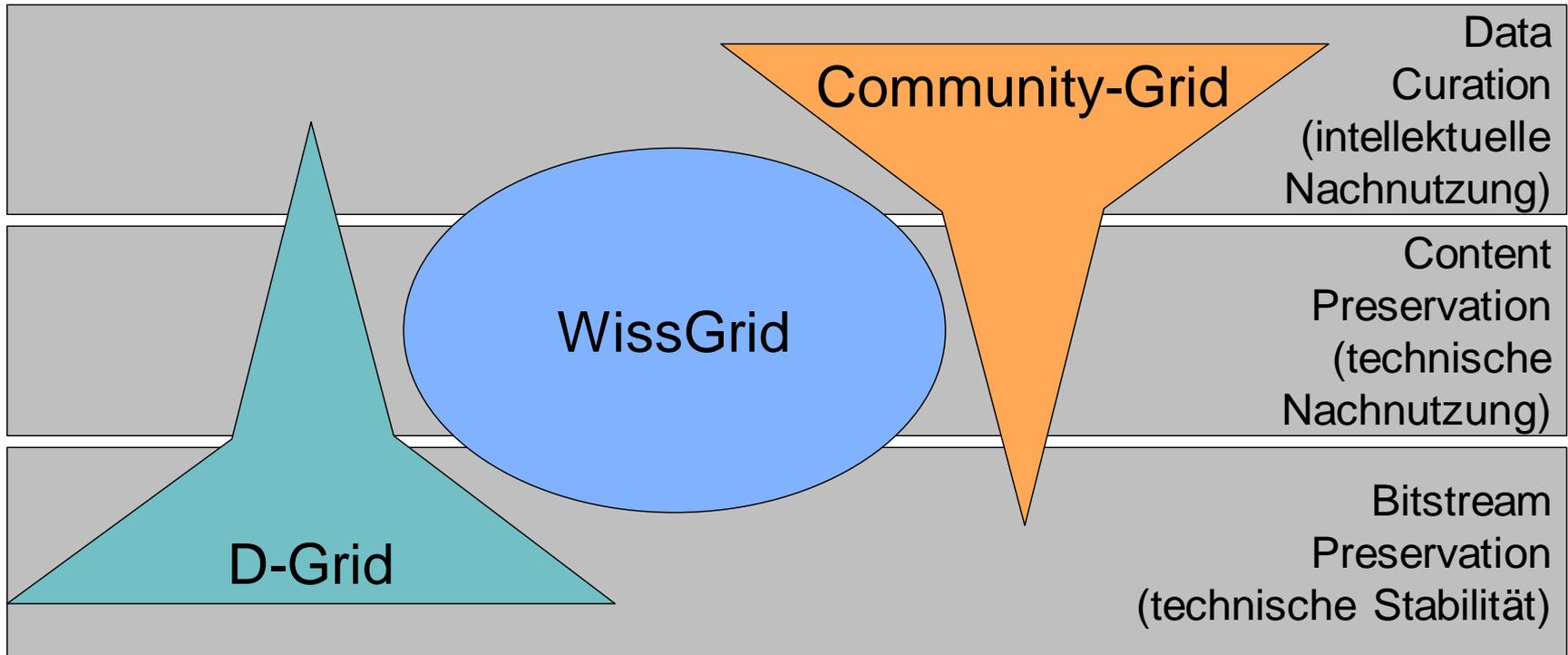
- Analyse von Anforderungen an Langzeitarchivierung durch die biomedizinische Forschung
- Behandlung von Fragestellungen zu Anforderungen der biomedizinischen Forschung in einem Rechtsgutachten
- Forschungsprimärdaten sind nicht durch Urheberrecht geschützt → eine „gewisse geistige Schöpfungshöhe“ ist notwendig
  - CSV-Daten (z.B. reine Messwerte) sind nicht geschützt
  - Durch Anwenden wissenschaftlicher Verfahren modifizierte Datenbestände sind geschützt
  - Methodisch abgeleitete Erkenntnisse sind geschützt

Wie können die digitalen Forschungsdaten langfristig genutzt werden? → **AP3**



Wie können die projektgeförderten  
Virtuellen Forschungsumgebungen (VRE)  
**nachhaltig** betrieben werden? → **AP1**

- Grid für die Wissenschaft
- Gefördert vom BMBF seit 04/2009 bis 04/2012
- Zielsetzungen bezüglich Langzeitarchivierung
  - **Grid als Basistechnologie** für Langzeitarchivierung
  - Entwickeln **modularer Langzeitarchivierungsdienste**
  - Bereitstellen von Dokumentation zur Unterstützung von Nutzern und Beratung
  - Entwickeln einer **generischen Grid-Architektur für Langzeitarchivierung von Forschungsdaten**, die durch wissenschaftliche Communities adaptiert werden kann





## **Universitätsmedizin Göttingen**

Medizinische Informatik

<http://www.mi.med.uni-goettingen.de/>

**Frank Dickmann**

CIOffice Forschungsnetze

[fdickmann@med.uni-goettingen.de](mailto:fdickmann@med.uni-goettingen.de)

Tel.: (0551) 39 - 14355

## Partner

- Universitätsmedizin Göttingen, Abteilung Medizinische Informatik (UMG)
- Universitätsklinikum Magdeburg, Institut für Biometrie und Medizinische Informatik
- Universitätsklinikum Schleswig-Holstein, Institut für Medizinische Informatik und Statistik (UKSH)
- Universitätsklinikum Schleswig-Holstein, AG Genomische Gastroenterologie, Klinik für Innere Medizin I (UKSH)
- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaft e.V. (AWMF)
- Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF)
- Zuse-Institut Berlin (ZIB)

## Partner

- Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
- Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB)
- Universitätsmedizin der Georg-August-Universität Göttingen Geschäftsbereich Informationstechnologie (UMG)
- Medizinische Informatik, Universität Göttingen (UMG)

## Partner (Stand 09/2011)

- Bayerische Staatsbibliothek
- Deutsche Nationalbibliothek
- FernUniversität Hagen
- Niedersächsische Staats- und Universitätsbibliothek Göttingen
- Humboldt-Universität zu Berlin
- Landesarchiv Baden-Württemberg
- Institut für Deutsche Sprache
- Computerspiele Museum Berlin
- Goportis (Leibniz-Bibliotheksverbund Forschungsinformation, inkl. ZB MED)
- PDF/A Competence Center

## Partner (ohne Unteraufträge)

- Universität Göttingen
- Astrophysikalisches Institut Potsdam
- Alfred-Wegener-Institut Bremerhaven
- Deutsches Elektronen Synchrotron
- Deutsches Klimarechenzentrum GmbH
- Konrad-Zuse-Zentrum für Informationstechnik Berlin
- Niedersächsische Staats- und Universitätsbibliothek
- Technische Universität Dortmund
- Universitätsmedizin Göttingen
- Universität Heidelberg
- Universität Trier
- Universität Wuppertal