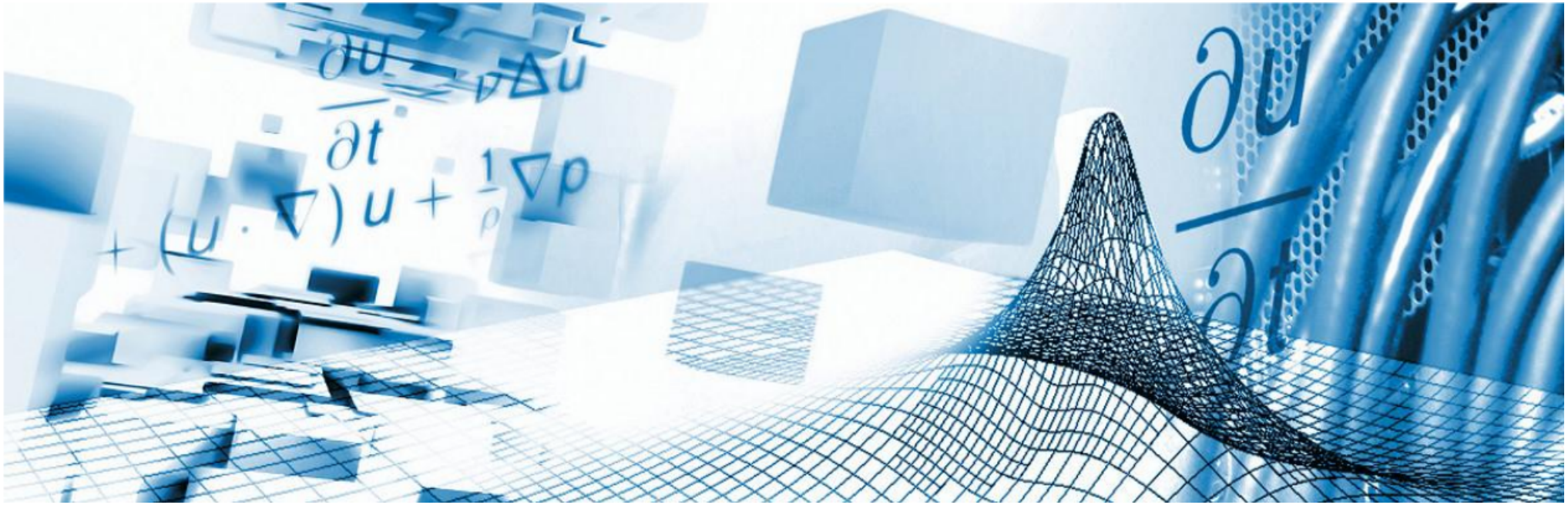
Juli 2018

UNIVERSITÄTS-
RECHENZENTRUM

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Langzeitarchivierung an der Universität Heidelberg: Status und Herausforderungen

Vincent Heuveline



Langzeitarchivierung in Heidelberg

Anforderungen / Bedarfe

- unterschiedlichste Daten sollen gespeichert werden
  (z.B. Bibliotheksdaten, Verwaltungsdaten, allg. Forschungsdaten)
- Daten sollen über (sehr) lange Zeiträume nutzbar erhalten bleiben
  (z.B. 30 oder mehr Jahre)
- Aufbewahrung besonders großer Kapazitäten an Forschungsdaten
  (z.B. Rohdaten im TB-Bereich für 10 Jahre, DFG)
- Orientierung am OAIS-Standard (Open Archival Information System)
  Referenzmodell und ISO-Standard 14721:2012
- Archivsystem soll keine öffentliche (Web-)Schnittstelle für den Datenzugriff haben („dark archive“)
  (zusätzliche Präsentation der Daten erforderlich, z.B. Repositorien)

Langzeitarchivierung in Heidelberg

UNIVERSITÄTS-
RECHENZENTRUM

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Anforderungen / Bedarfe

- hohes Maß an Sicherheit gegen Datenverlust
  → Replikation / Georedundanz / Medienbruch (Festplatten & Bänder)
- Standardisierung der Datenformate für zukünftige Nachnutzung
  → nur geeignete Formate, verlustfreie Konvertierung
- Rechtliche Aspekte
  (z.B. Lizenzrecht, Urheberrecht, ...)
- Einsatz standardisierter Metadaten
  → technische, strukturelle, administrative, rechtliche und den Inhalt der Daten beschreibende Metadaten (z.B. PREMIS)
- Persistente Identifier
  → interne vs. externe Identifier

Kompetenzzentrum Forschungsdaten (KFD)

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

- Gemeinsame Serviceeinrichtung von UB und URZ
- Beratung und Unterstützung in allen Projektphasen:

Project Planning

- Data Management Plans
- Courses & workshops
- Technical, organizational & legal information

Data processing

- heiBOX
- heiCLOUD
- SDS@hd
- High Performance Computing

Data Archiving & Publication

- heiDATA
- heidICON
- SDA@hd
- Research Data Catalougue

Pictures: © Universität Heidelberg, Kommunikation und Marketing

Langzeitarchivierung in Heidelberg
Konzept mit Service-Levels

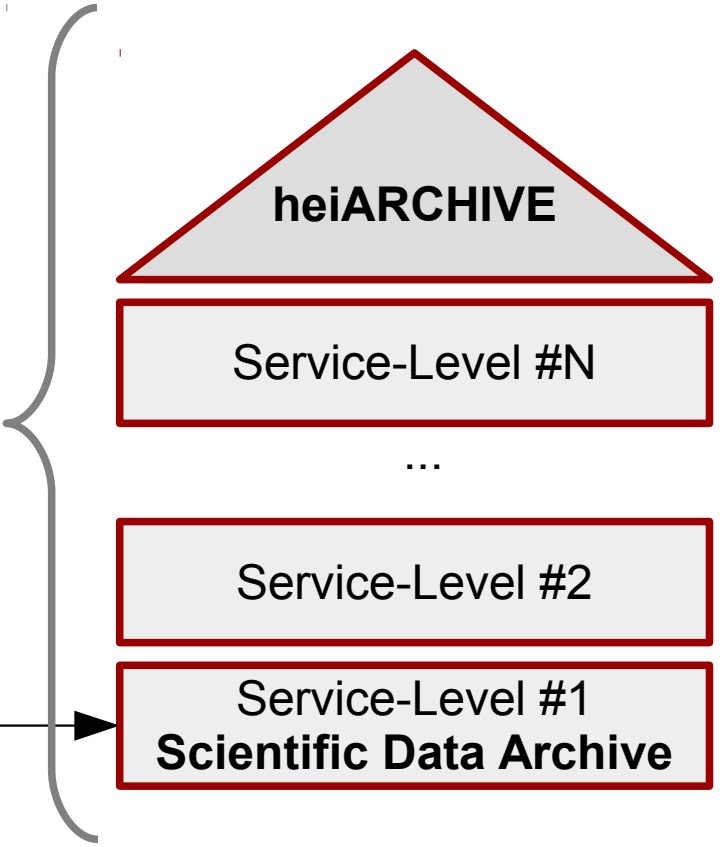
UNIVERSITÄTS-
RECHENZENTRUM

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Entwicklung eines Archivierungsdienstes an der Universität Heidelberg

- Digitales Langzeitarchiv **heiARCHIVE** mit unterschiedlichen Service-Levels
  → gemeinsames Projekt mit UB und UniArchiv

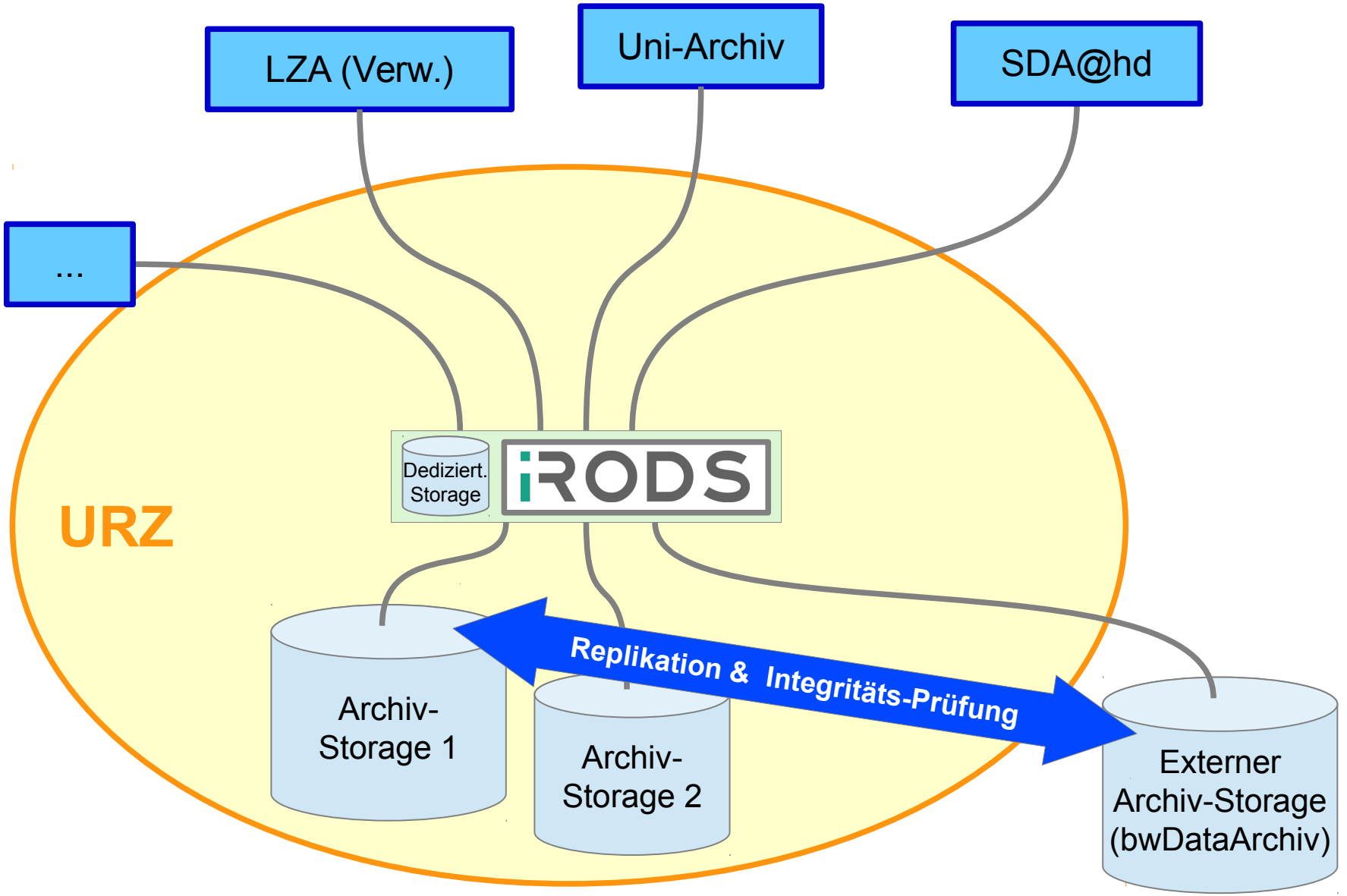
- Wissenschaftliches Datenarchiv **SDA@hd – Scientific Data Archive**
  → Projekt im Rahmen des Kompetenzzentrums Forschungsdaten



Langzeitarchivierung in Heidelberg
Konzept mit Service-Levels

- Mehrere Service-Levels mit folgenden Unterscheidungsmerkmalen
  - Dauer der Aufbewahrung
  - Qualität der (Geo-)Replikation
  - Datenkuration (inkl. Formatkonversion)
  - Regelung für Löschen und Datenherausgabe/Zugriff
  - Umfang der Metadaten
  - ...
- Service-Level „wissenschaftliche Daten“ (SDA@hd)
  - ersten Stufe (andere Service-Levels bauen darauf auf)
  - Speicherkapazität stellt Herausforderung dar (> 10 TB)

Technisches Speicherkonzept



Service Level #1:
SDA@hd - Scientific Data Archive

- Aufbewahrung von wissenschatlichen (Roh-)
  - Große Kapazitäten möglich
  - 10 Jahre (gem. DFG Richtlinie)
- Minimalsatz an Metadaten (gem. Standards)
- Dokumente (Archivierungsvertrag, z.B. für Beschreibung der Urheber- und Lizenzthemen)
- Paketierung der Daten gemäß Standard
- Geeignete Ingest-Prozesse und -Technologien für große Daten (z.B. Datenübernahme aus SDS@hd – Scientific Data Storage für „hot data“)

  => Geplanter Dienststart SDA@hd Ende 2018 (zunächst als Pilotdienst)

Tape-Library am URZ

UNIVERSITÄTS-
RECHENZENTRUM

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

- **Quantum Scalar i6** Tape Library:
  - 700 Slots
  - 6 Laufwerke
  - 1 Roboter
- Redundanzen:
  - Stromversorgung (UPS)
  - Datentransfer
  - (geplant: Geographische Redundanz)
- Tapes:
  - LTO7: 9 TB
  - LTO8: 12 TB
  - Voller Ausbau mit LTO8:

    ~ 8 PB

E-Science Tage in Heidelberg

Vielen Dank für die Aufmerksamkeit!

E-Science-Tage 2019: „Data to Knowledge“

27. bis 29. März 2019 in Heidelberg