



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



Universitäts- und  
Landesbibliothek  
Darmstadt

# Archivematica an der ULB Darmstadt

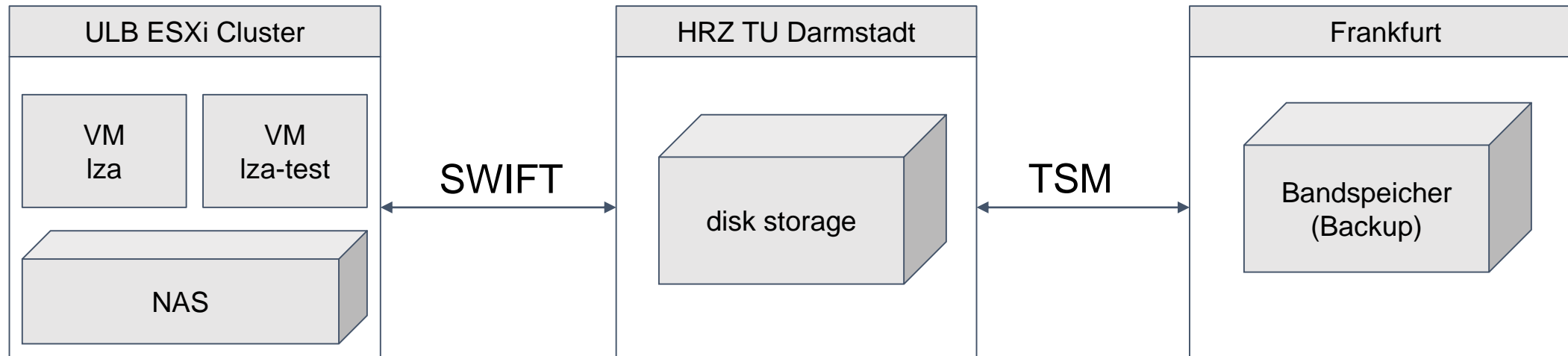
Aktueller Stand und Perspektiven



# Inhalt

- 1** Systemarchitektur
- 2** Anwendungsfälle, Herausforderungen und Lösungen
- 3** Perspektiven

# Architektur des Systems



- 2 Archivematica-Instanzen (Version 1.16) an der ULB für Produktiv- und Testbetrieb
- Jeweils 12 virtuelle CPU cores @2.6 GHz und 16 GB RAM
- Archivematica-Prozessierungsordner auf NAS

# Anwendungsfälle

LZA-System muss sehr verschiedene Anwendungsfälle abdecken

- Digitalisate aus ULB Digitalisierungszentrum
- Projekte:
  - DFG Workflow Digitale Medien
  - DFG Idovir
  - NFDI4Ing ing.est

# Digitalisate

- Retrodigitalisierung des historischen Bestands
- Digitalisierungsprojekte, z.B. Darmstädter Tagblatt ab 1740
- Veröffentlichung über Plattform TUDigit (Kitodo)
- Ablage der Digitalisate zur Langzeiterhaltung nach Durchlaufen des Workflows



<https://doi.org/10.25534/tudigit-14001>



# Digitalisate

- Selbstentwickelte Erweiterung von Kitodo für Langzeitarchivierung
  - bash Skripte zur Erzeugung von Bagit-Paketen
  - bash Skripte für Transfer nach Archivematica über NAS und Nutzung der Archivematica API
  - gesteuert über cronjobs
- Paketgrößen bis mehrere hundert GB
- Anzahl zu archivierender Scans ~2 Mio
- Archivierung im Produktivsystem



# Workflow Digitale Medien

- Gezieltes Sammeln und Bereitstellen digitaler Inhalte in einem einheitlichen Format
  - elektronische Zeitschriften
  - E-Books
  - damit verbundene Medienelemente
  - Digitalisate mit strukturierten Volltexten
- Archivematica zentrales System im Workflow
  - (1) Eingangskontrolle und Vorsortierung
  - (2) Langzeitarchivierung
  - (3) DIP Erstellung mittels Automation Tools
  - (4) Restliche Verarbeitung ausgehend von DIP
  - (5) Langzeitarchivierung von bearbeiteten Inhalten
- Aktuell in Entwicklung und Testbetrieb

# Workflow Digitale Medien

- Unterschiedliche Datentypen
  - PDF
  - XML
  - Bilder in unterschiedlichen Formaten
  - Office Dateien
  - LaTeX
  - ...
- Sehr viele Pakete (BagIt)
  - Ziel: >> 1 Million Pakete/Jahr
- Hohe Prozessierungsgeschwindigkeit notwendig
- Paketgröße variabel, aber eher klein (wenige MB)





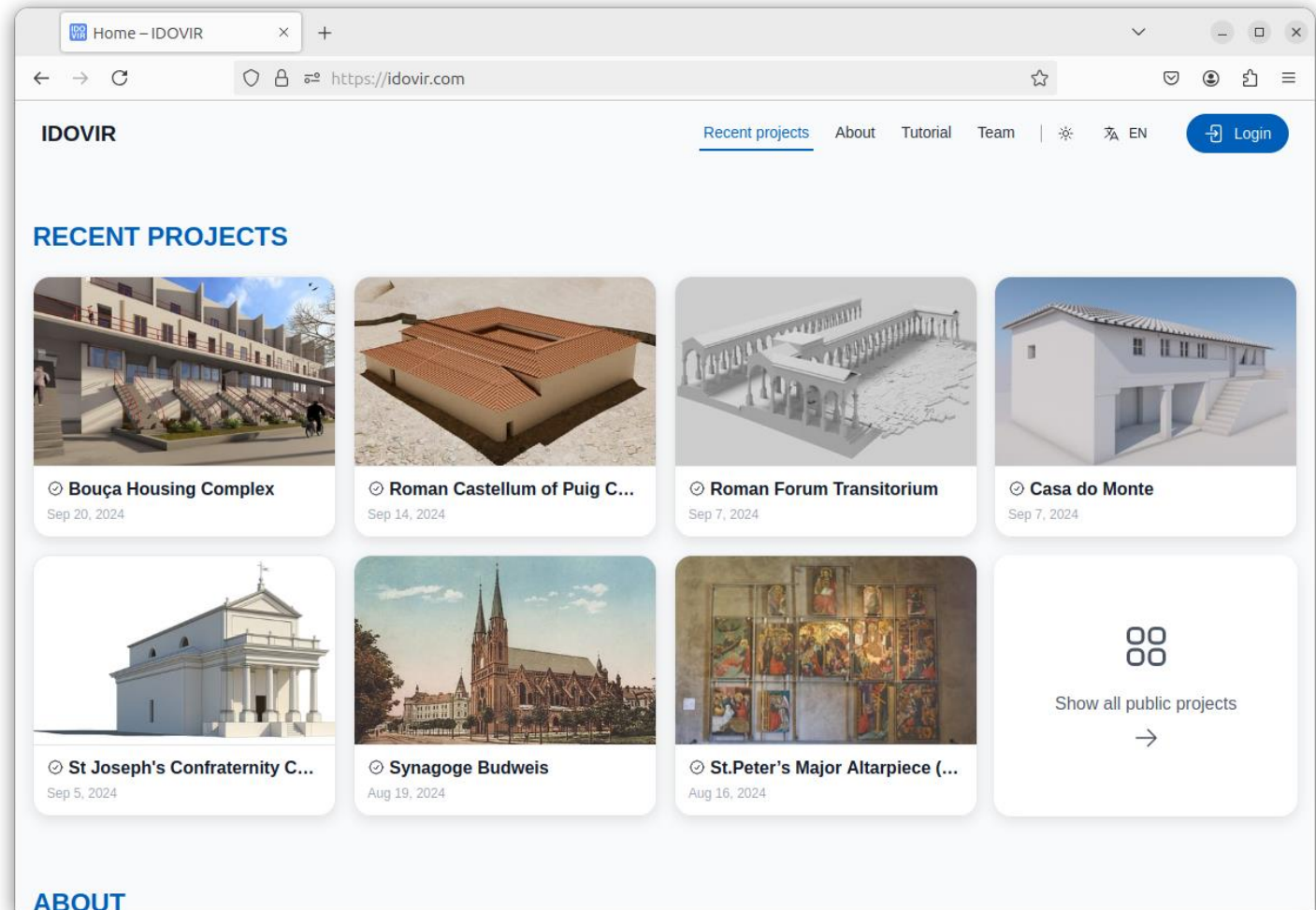
# Transfer/Ingest Optimierung

- Paralleler vs. sequentieller Ingest kleiner Pakete
  - 5 Pakete paralleler Ingest
    - Faktor ~2,5 schneller als sequentieller Ingest für die Pakete insgesamt
    - Jedes einzelne Paket dauert länger
- Anzahl MCP Clients
  - MCP Clients sind die Worker, die die Archivemata-Microservices ausführen
  - MCP Clients schrittweise von 1 auf 12 erhöht
  - Ab 8 MCP Clients bei uns kein Geschwindigkeitszuwachs mehr



# Idovir

- DFG gefördertes Projekt
- Web-Anwendung zur Dokumentation von Entscheidungsprozessen bei virtuellen Architekturekonstruktionen
- Betrieben in einem Kubernetes Cluster an der ULB



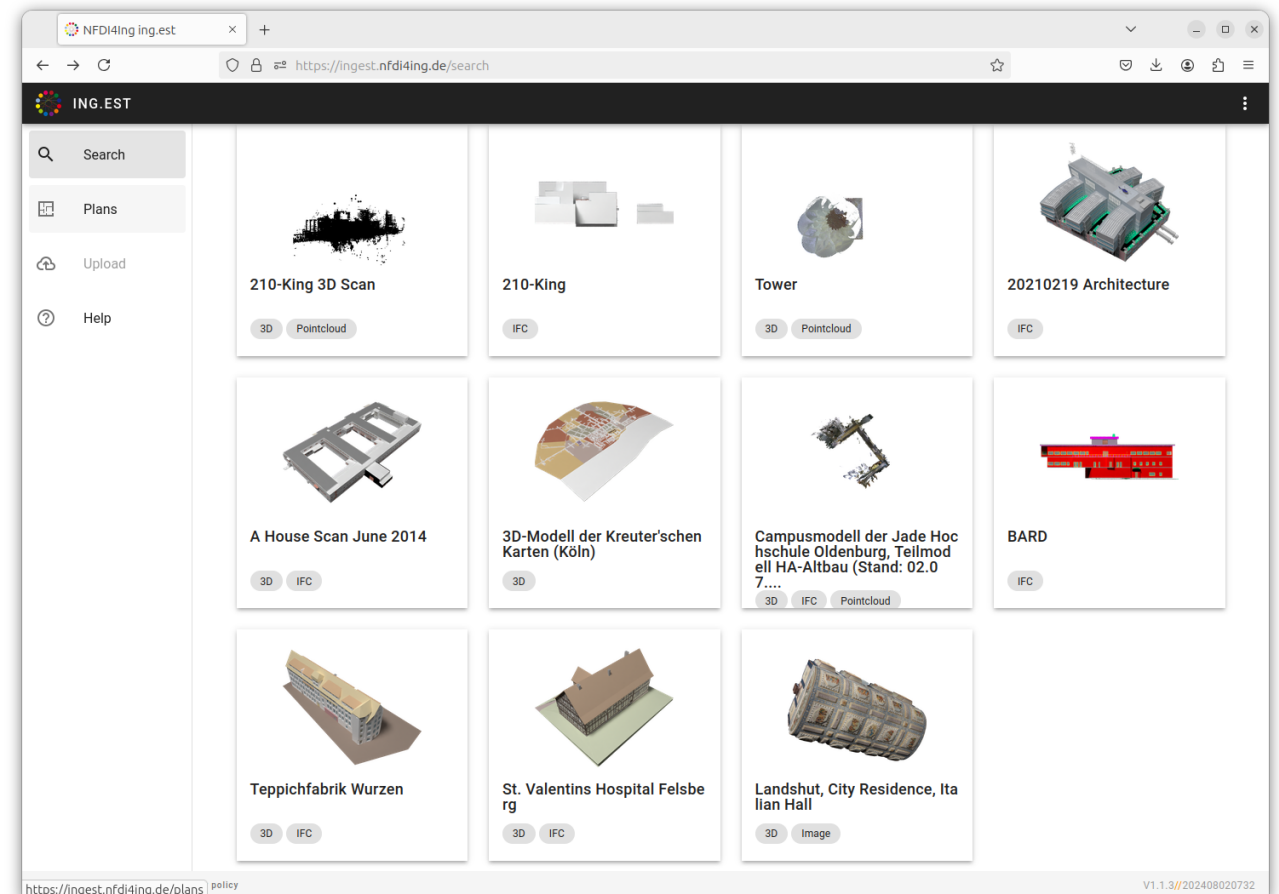
# Idovir

- Langzeitarchivierung im Projekt realisiert durch Archivematica der ULB
- Idovir-Applikation erzeugt selbst Bagit-Pakete mit:
  - 3D Modellen
  - Rekonstruktionsprozessen als XML
- Kopieren der Bagit-Pakete über SSH auf NAS der ULB
- Start des Transfers der Bagit-Pakete über Archivematica API
- Kontrolle des Ingest-Status über polling der Archivematica API
- Momentan Gesamtvolumen von 55 GB
  - Paketgröße ~1 GB



# NFDI4Ing ing.est

- Service zur Publikation von Forschungsdaten aus Architektur, Bauingenieurwesen und Urbanistik
- Extraktion von Metadaten aus domänenspezifischen Dateiformaten
- Vergabe von DOIs zu Datensätzen
- Langzeitarchivierung der Datensätze
- Betrieb im ULB K8S-Cluster
- LZA wie bei IDOVIR direkt über Archivematica API



# Perspektiven

Entwicklung einer Pre-Ingest application:

- Minimierung des Aufwands auf Seiten des Anwendungsfalls
- Sicherstellung einer einheitlichen Verarbeitung der Daten
- Optimale Auslastung der Systemressourcen ohne Überlastung

Weitere Optimierungen der Betriebsparameter

- Anzahl MCP Clients
- Ressourcen der virtuellen Maschine (CPU, RAM)
- Zugriffsgeschwindigkeit des Prozessierungsordners (NAS vs SSD)

Überführen der Projekte in den Produktivbetrieb