

Agenda

- 1. Kulturdatenarchiv Thüringen**
- 2. Verallgemeinerung**
- 3. Umsetzung**
- 4. Weiterentwicklung**

Projektskizze

Kulturdatenarchiv Thüringen
Projekt der
Thüringer Landes- und Forschungsbibliothek
und der
Universitäten Erfurt, Jena und Ilmenau

Grundlage ist das
Digitales Archiv Nordrhein-Westfalen (DA NRW)
des
Landschaftsverband Rheinland
und die
Universal Multimedia Electronic Library (UrMEL)
der
Thüringer Universitäts- und Landesbibliothek Jena (ThULB)

Ziel des Projektes

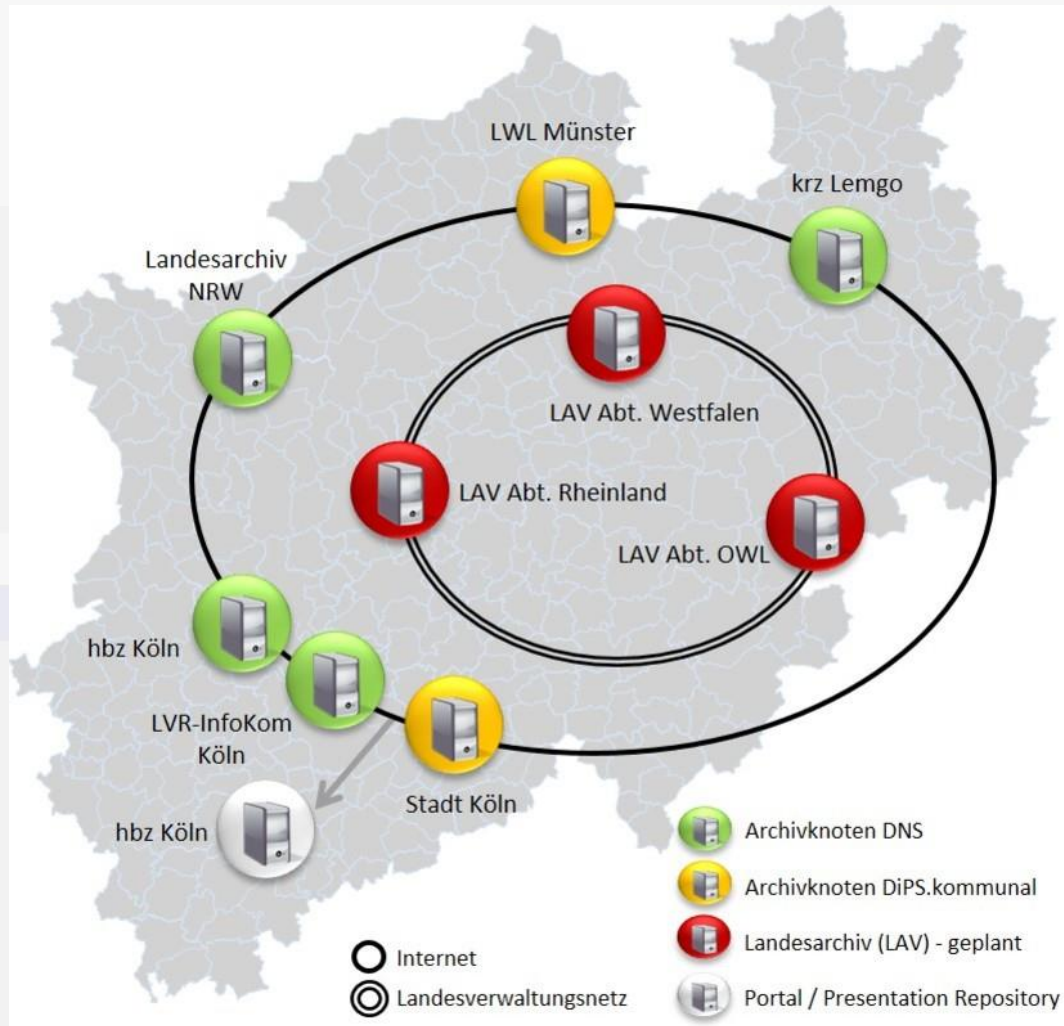
- Etablierung eines dezentralen Langzeitarchives
- Schaffung einer flexiblen ausbaufähigen Architektur
- Anbindung vorhandener Software (UrMEL/MyCoRe, DA DNS-Core)
- Anbindung vorhandener Storage Hardware
- Nutzung von OpenSource
- wenig bzw. keine Softwareentwicklung (Nutzung vorhandener Projekte)
- kostengünstige Umsetzung



CULDA
archive

archive for cultural data

Strukturbild DA-NRW

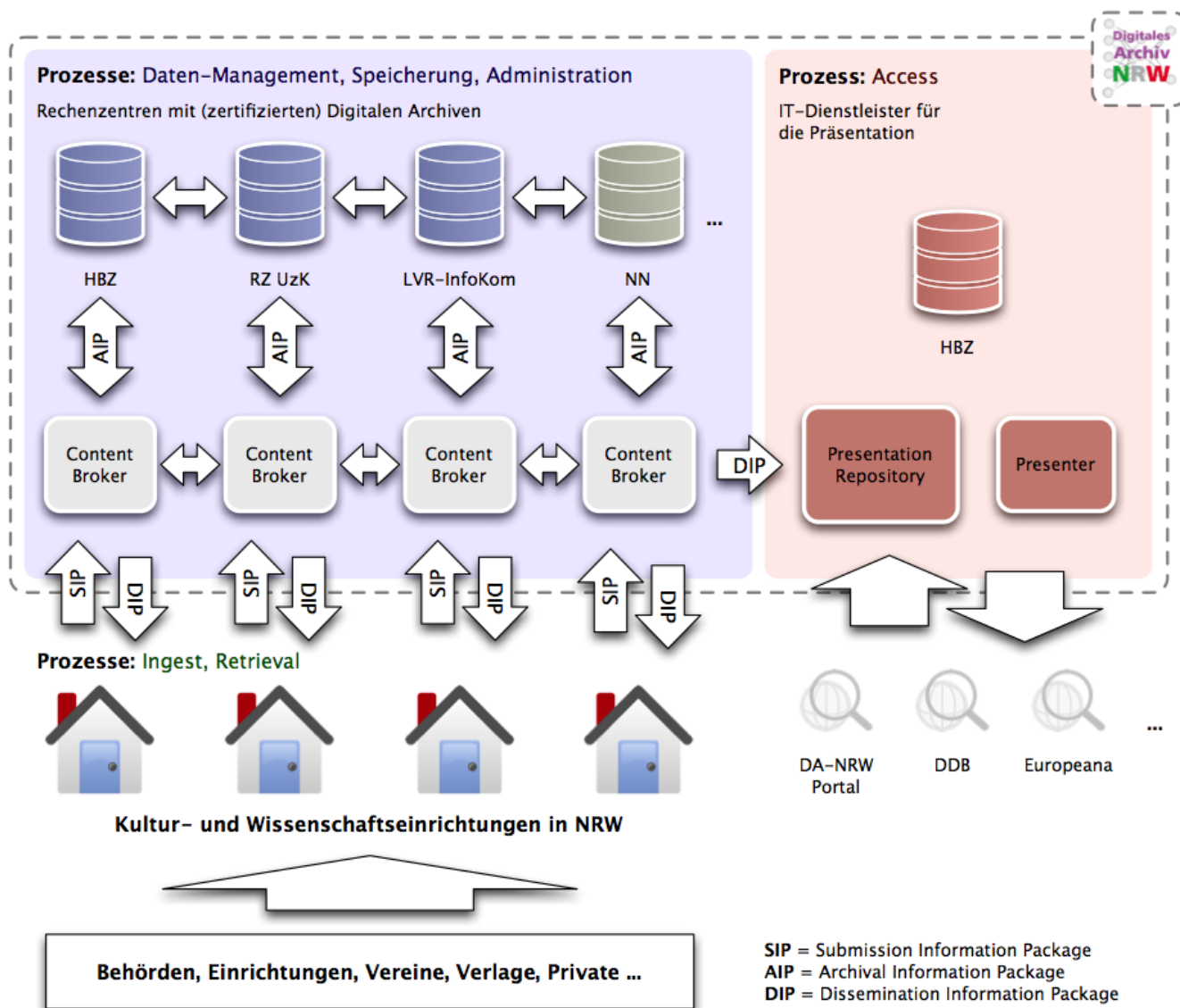




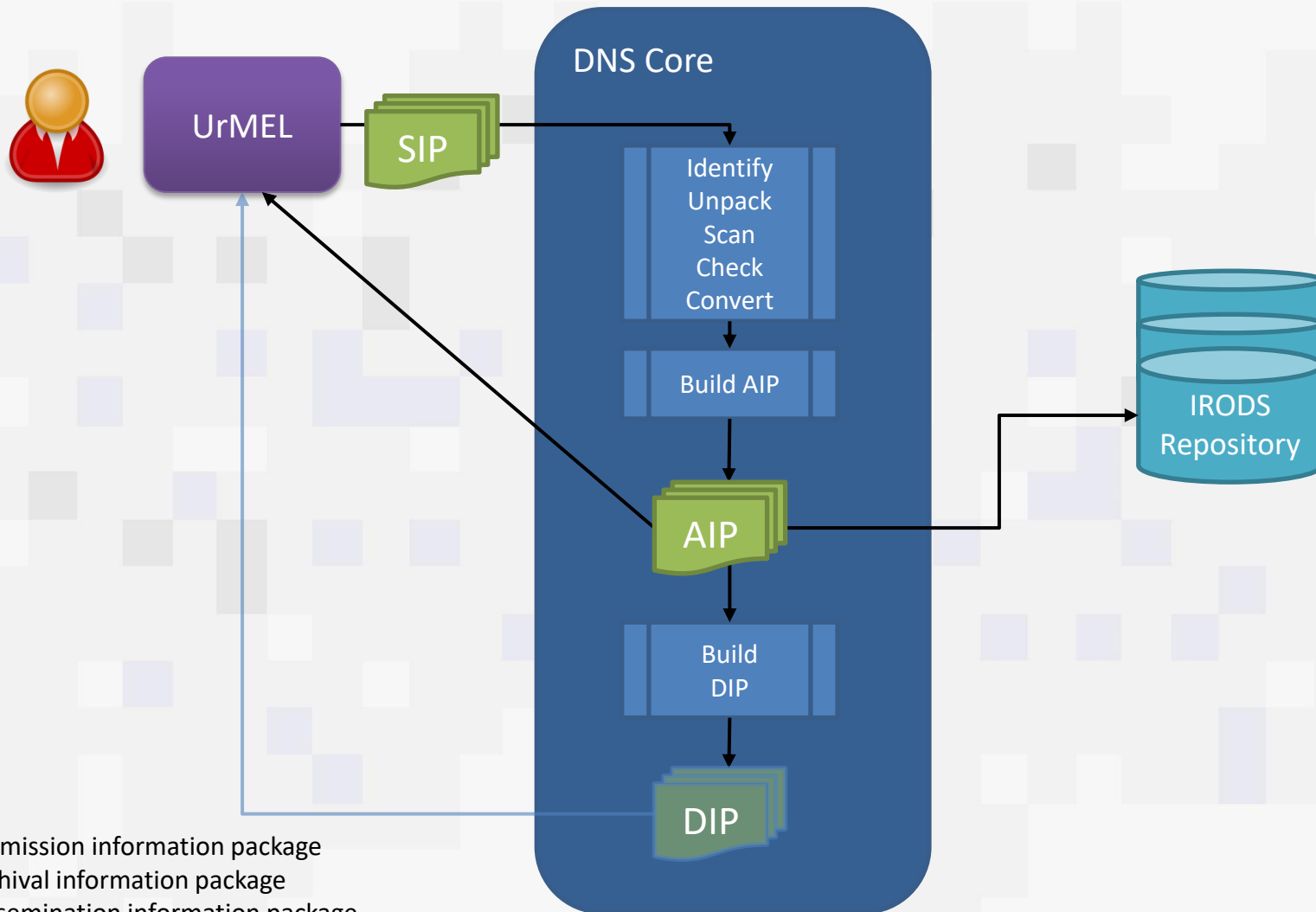
CULDA
archive

archive for cultural data

Workflow DA-NRW



Workflow



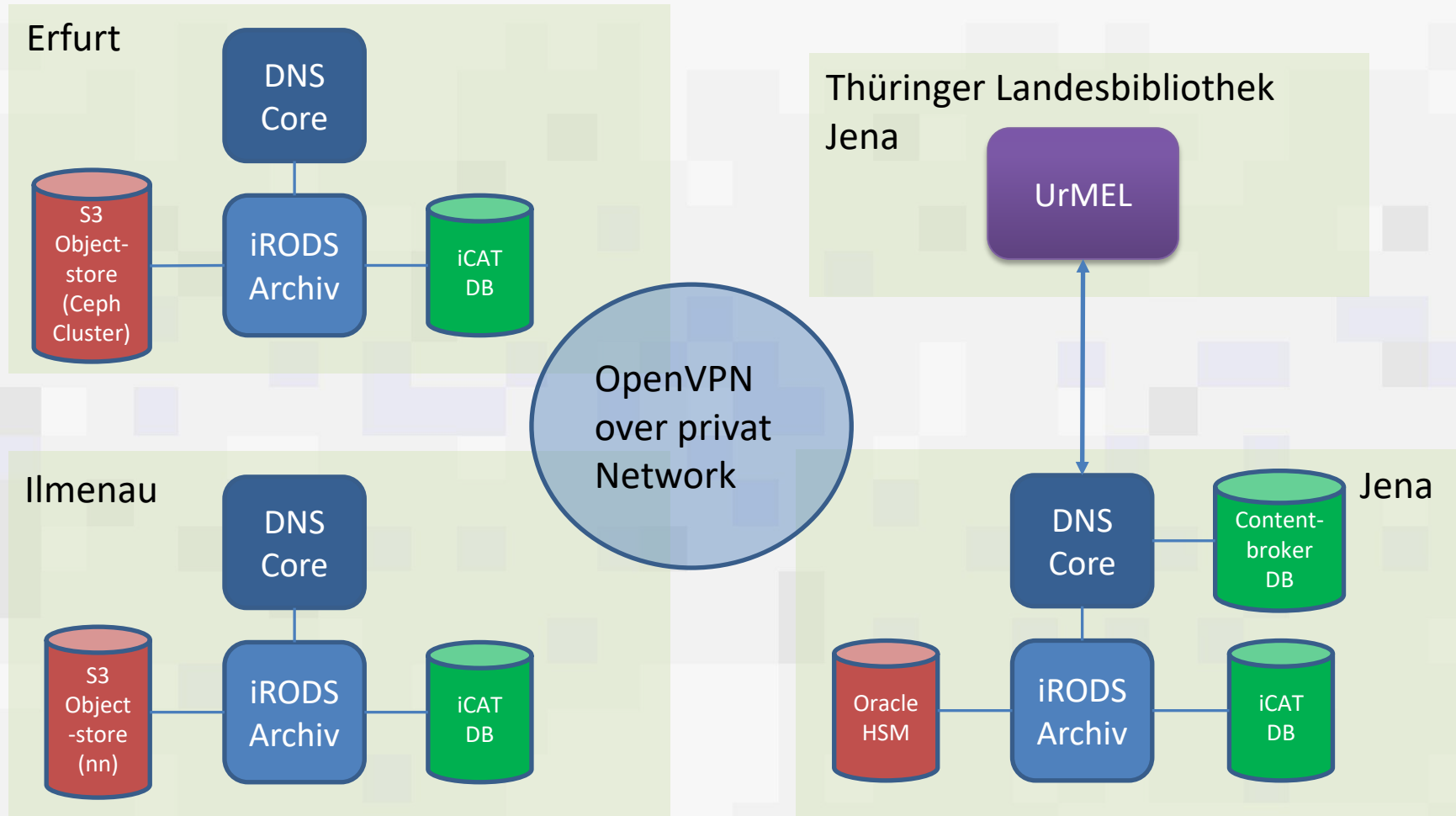
SIP: submission information package
AIP: archival information package
DIP: dissemination information package



CULDA
archive

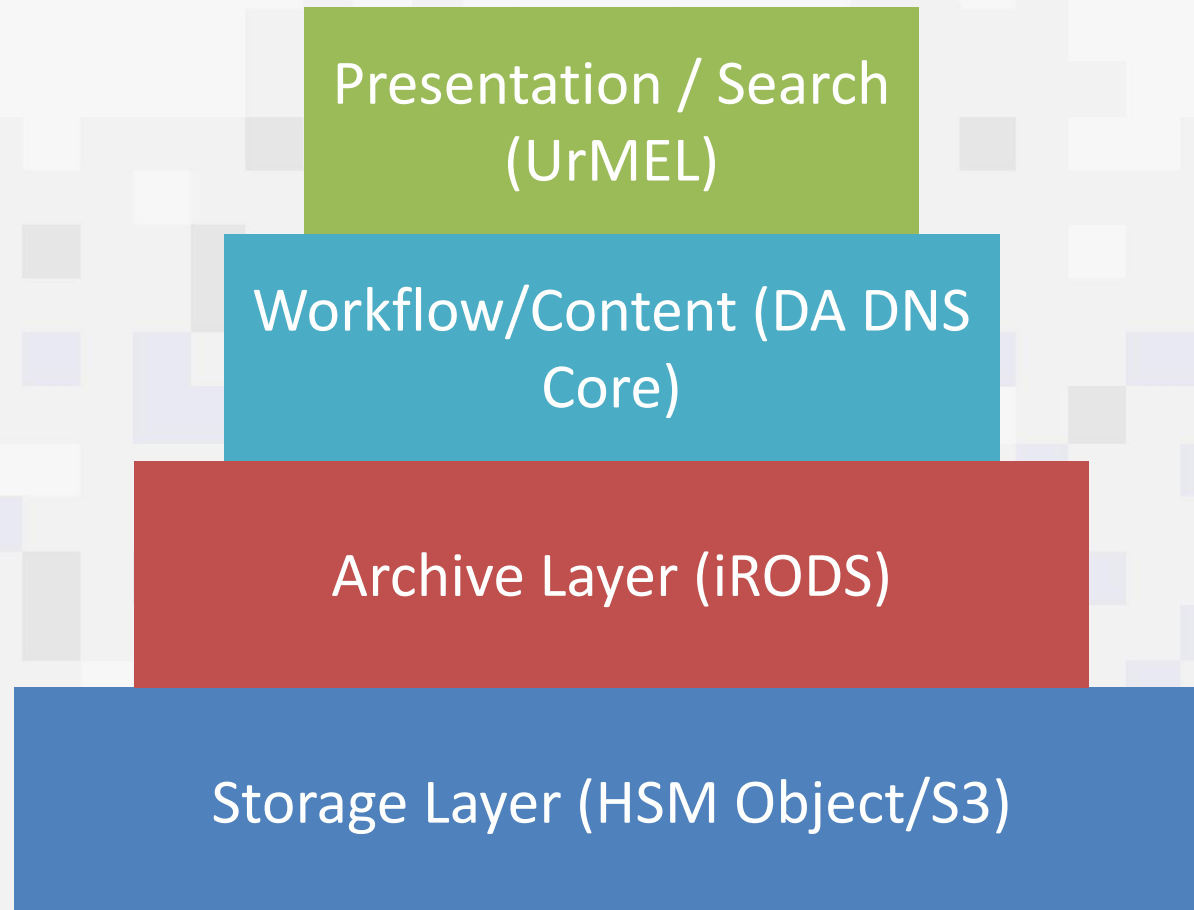
archive for cultural data

Strukturbild Culda



UrMEL: Universal Multimedia Electronic Library, basierend auf MyCoRe
DNS Core: DA-NRW Software Suite
iRODS: integrated Rule-Oriented Data System

Schichtenmodell



Storage Layer - HSM

Hierarchical Storage Management (HSM)

- + mehrere Kopien auf unterschiedlichen Medien
- + räumliche Verteilung der Kopien inkl. Tapes
- + erprobte Systeme (10+ Jahre Laufzeit)
- + Medienmigration als integraler Bestandteil
- + Datentransparenz bis zum Storage Device
- + integrierte Prüfsummen und WORM
- Komplexität
- Skalierung „nach unten“
- Kopien effektiv nur im „SAN“ Entfernungsbereich

Storage Layer - SDS

Software Defined Storage (SDS)

- + Verteilung der Blöcke auf mehrere Ort und Geräte (Erasure-Coding)
- + Austausch kompletter Server inkl. Reorganisation
- + Zugriff über S3, Block oder Filesystem
- Praktisch nur HDD (keine Tapes) als Storage Device
- begrenzte Performance oder hoher Netzwerkaufwand
- aufwändige Replikation über größere Entfernungen
- skaliert nicht für kleine Datenmengen

Storage Layer – Storage Array

Disk Array

- + preislich attraktiv
- + hohe Performance
- praktisch nur HDD (keine Tapes) als Storage Device
- räumliche Aufteilung aufwändig und inflexibel
- nur begrenzt nach oben skalierbar

Archive Layer

Zwischenschicht zur Abstraktion des Storage

Funktionen:

- Georedundanz, Replikation zwischen Standorten und Organisationen
- Migration zwischen Architekturen
- Nutzung vorhandener Storage Infrastruktur
- Zusammenarbeit verschiedener Organisationen
- Datensicherheit und Integrität
- Authentifizierung, Accounting, Compliance

Archive Layer – iRODS

iRODS, Integrated Rule Oriented Data System

Open Source

Distributed

Data Centric

Metadata Driven

Entwicklung erfolgt im iRODS Consortium

RENCI (North Carolina State University)

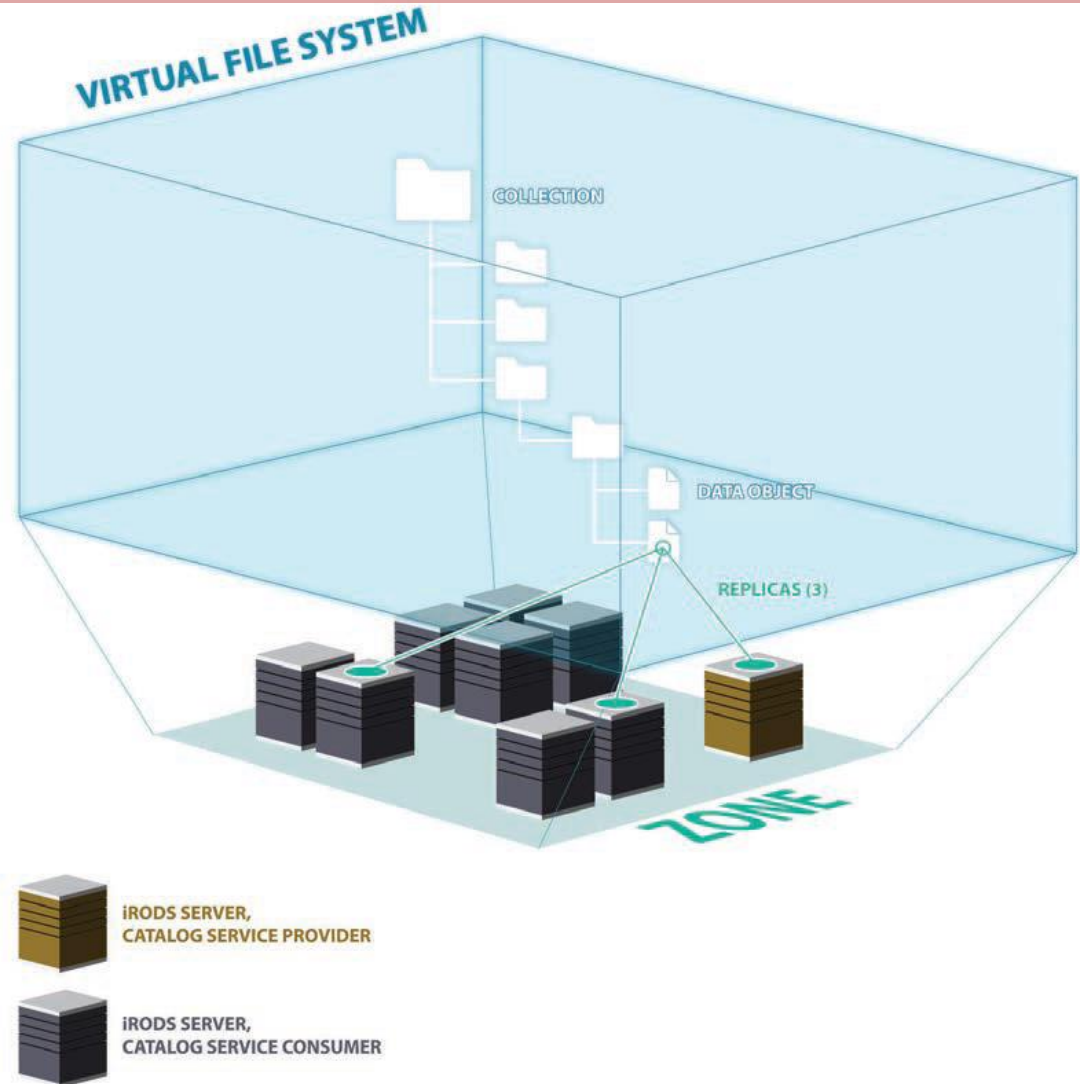
.....

Archive Layer – iRODS

Nutzersicht ist ein
virtuelles Filesystem

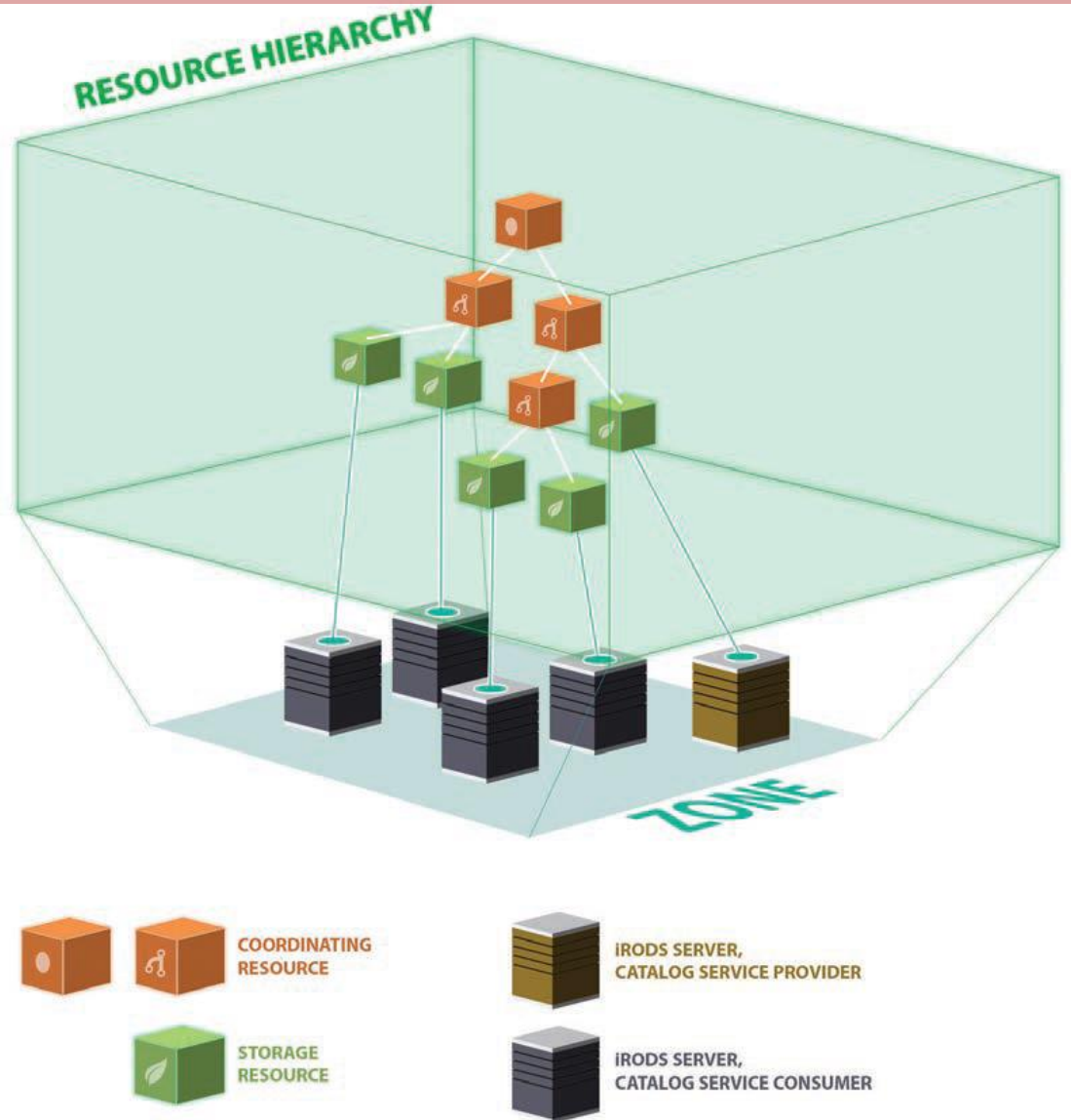
Wichtig:

In iRODS abgelegte Daten
finden sich 1:1 im Storage
Layer wieder



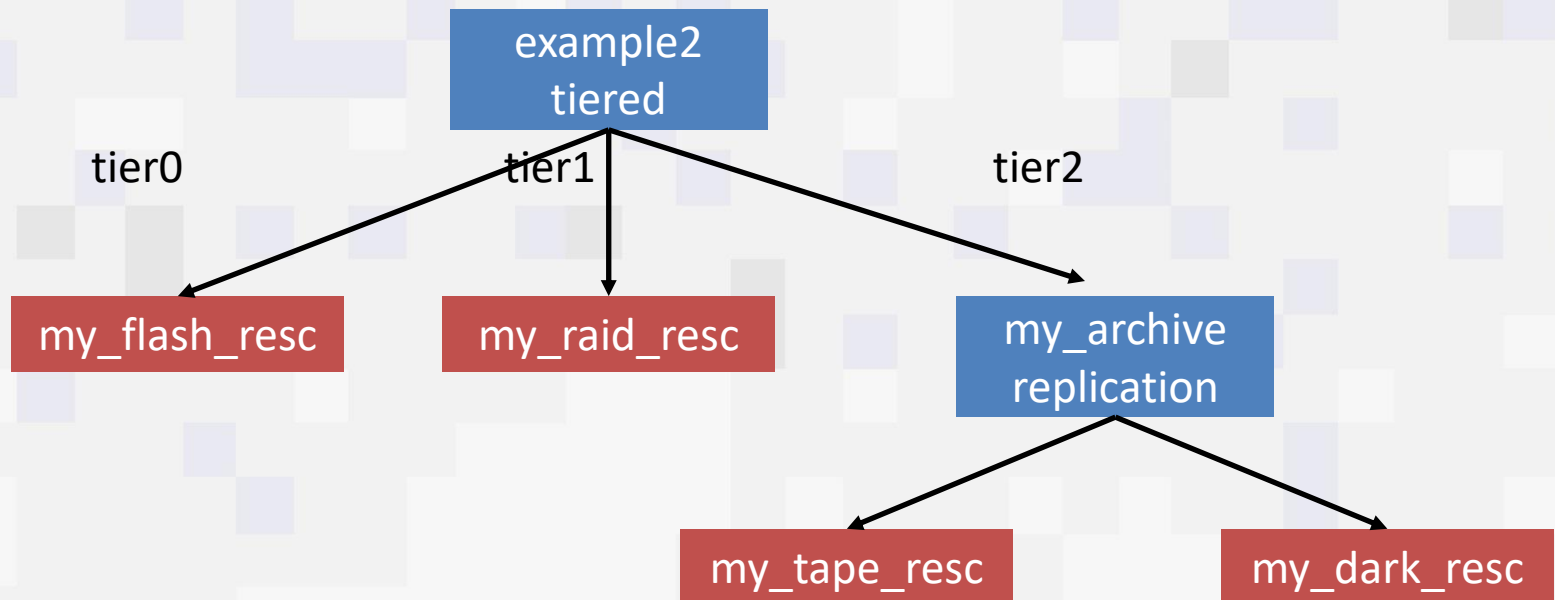
Archive Layer – iRODS

Datenverteilung über Ressourcen



Archive Layer – iRODS

```
$ iadmin mkresc example2 tiered
$ iadmin mkresc my_flash_resc "unix file system" fastest.example.org:/Vault
$ iadmin mkresc my_raid_array "unix file system" raid5.example.org:/Vault
$ iadmin mkresc my_archive replication
$ iadmin mkresc my_tape_resc "unix file system" tape.example.org:/Vault
$ iadmin mkresc my_dark_resc "unix file system" dark.example.org:/Vault
$ iadmin addchildtoresc example2 my_flash_resc "tier0"
$ iadmin addchildtoresc example2 my_raid_array "tier1"
$ iadmin addchildtoresc example2 my_archive "tier2"
$ iadmin addchildtoresc my_archive my_tape_resc
$ iadmin addchildtoresc my_archive my_dark_resc
```



Quelle: E-iRODS Composable Resources

Umsetzung

1. On-Premises (auf eigener Hardware)
2. Private Cloud
3. Public Cloud
4. Multicloud Ansatz / Kooperative Installationen

Unveränderlichkeit / Nachweisbarkeit

1. Realisierung im Storage Layer

- HSM Worm, Tape WORM, S3 Vault, NAS Vault
- Prüfsummen im Filesystem (Bsp. Oracle HSM)

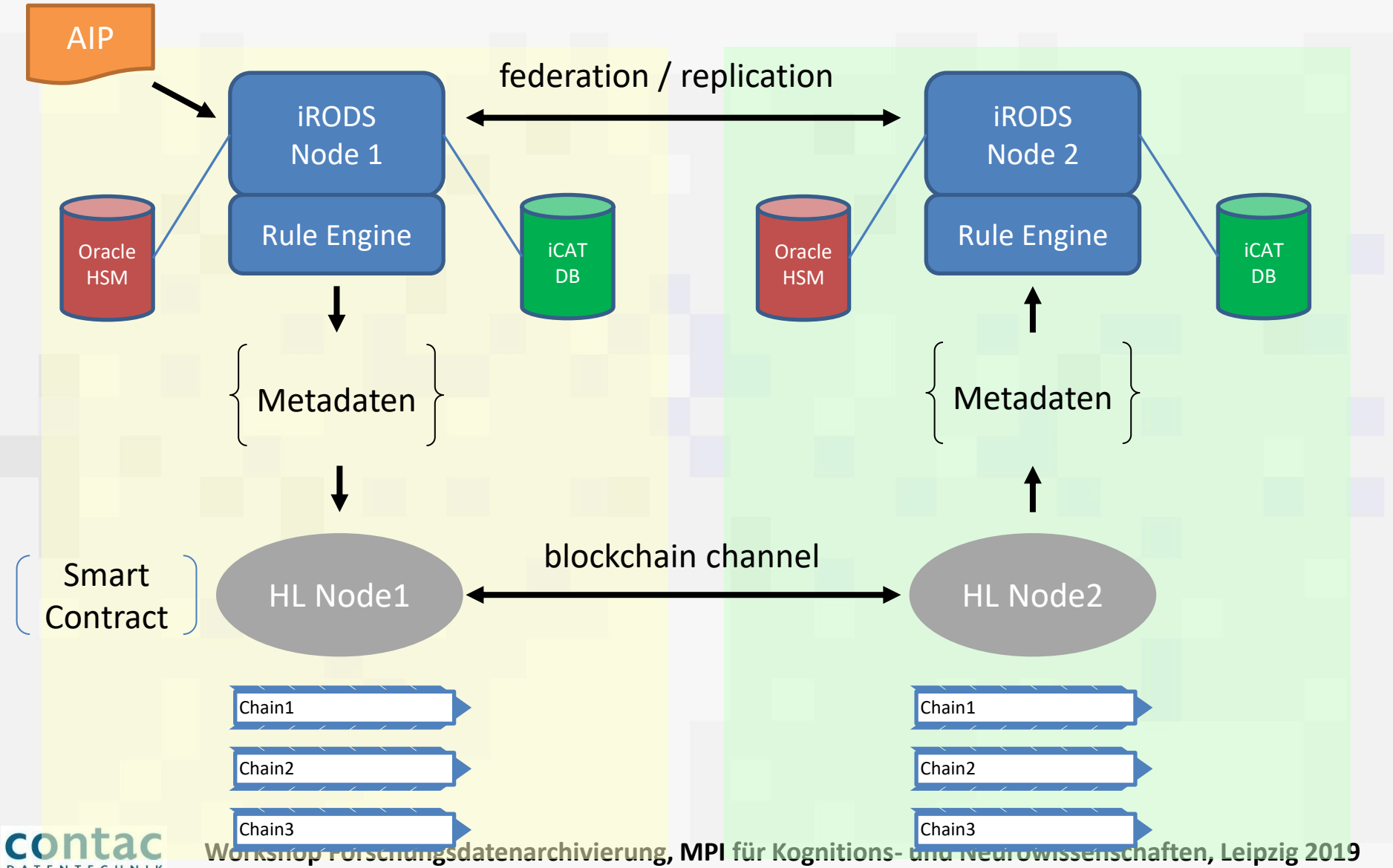
2. Realisierung im Archive Layer

- Speicherung der Metadaten inkl. Prüfsummen in einer Datenbank (iRODS-iCAT)
- Speicherung der Metadaten inkl. Prüfsummen in einem kommerziellen Archivsystem (Artec)
- Speicherung der Metadaten inkl. Prüfsummen in einer verteilten Blockchain (Hyperledger Fabric)

Blockchain als Speicher für Metadaten

- Jeder Knoten in einer Blockchain verifiziert und verarbeitet jede Transaktion.
- Blockchain-basierte Transaktionen und Daten sind aufgrund ihrer integrierten Redundanz extrem fehlertolerant.
- Jeder Record wird an eine Liste und unveränderlichen Datensätzen angefügt.
- Der Algorithmus der Transaktion / des Objektes wird versioniert in der Chain selbst unveränderlich gespeichert (-> Smart Contract)

Blockchain als Speicher für Metadaten



Unsere Leistungen

- Projektmanagement
- Qualitätskontrolle
- Roll Out der Komponenten
- Rule
- Anpassung
- Dokumentation

Ihr Vorteil

- Nutzung vorhandener Hardware
- keine Softwareentwicklung
- Georedundanz
- Open Source