



Datenmanagement und Archivierung im Rahmen der klinischen Forschung

Dr. Maximilian Fünfgeld

Heinrich-Lanz-Center for Digital Health, University Medicine Mannheim

03.12.2019 Bonner Archivtage

Daten sind das Öl des 21. Jahrhunderts



“Data is the new oil. It’s valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.”

— Clive Humby

Versorgungsdaten Daten

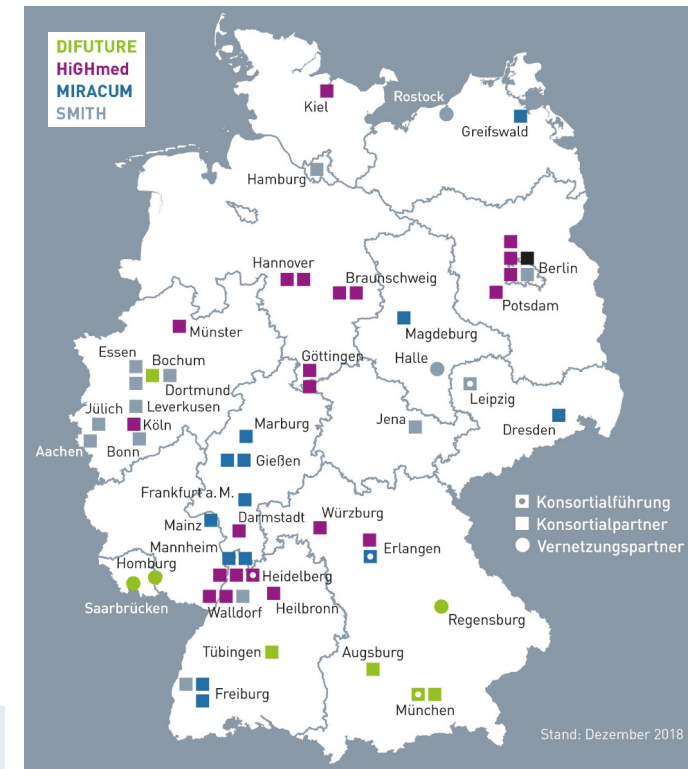
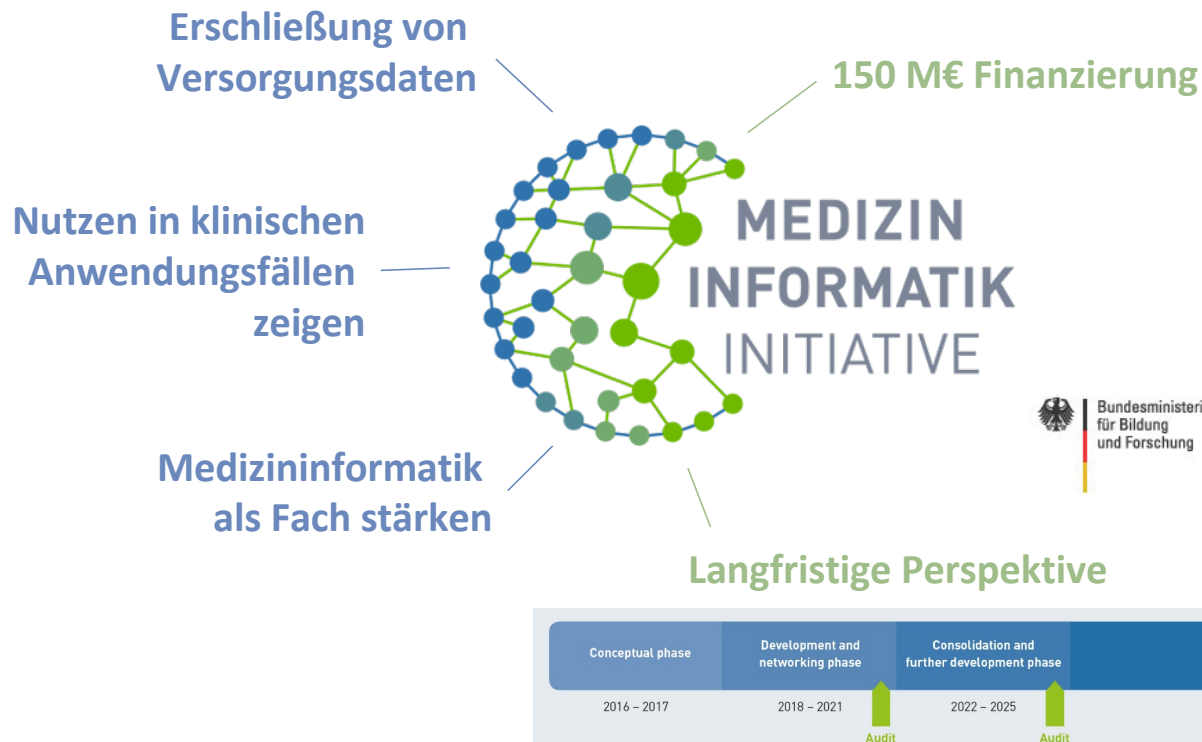
- Verlieren nicht an Wert
- International gültig
- Werden nicht durch Benutzung verbraucht

Archiv

- Medizinische Archive: Ölquellen des 21. Jahrhunderts
- Die Daten müssen raffiniert werden



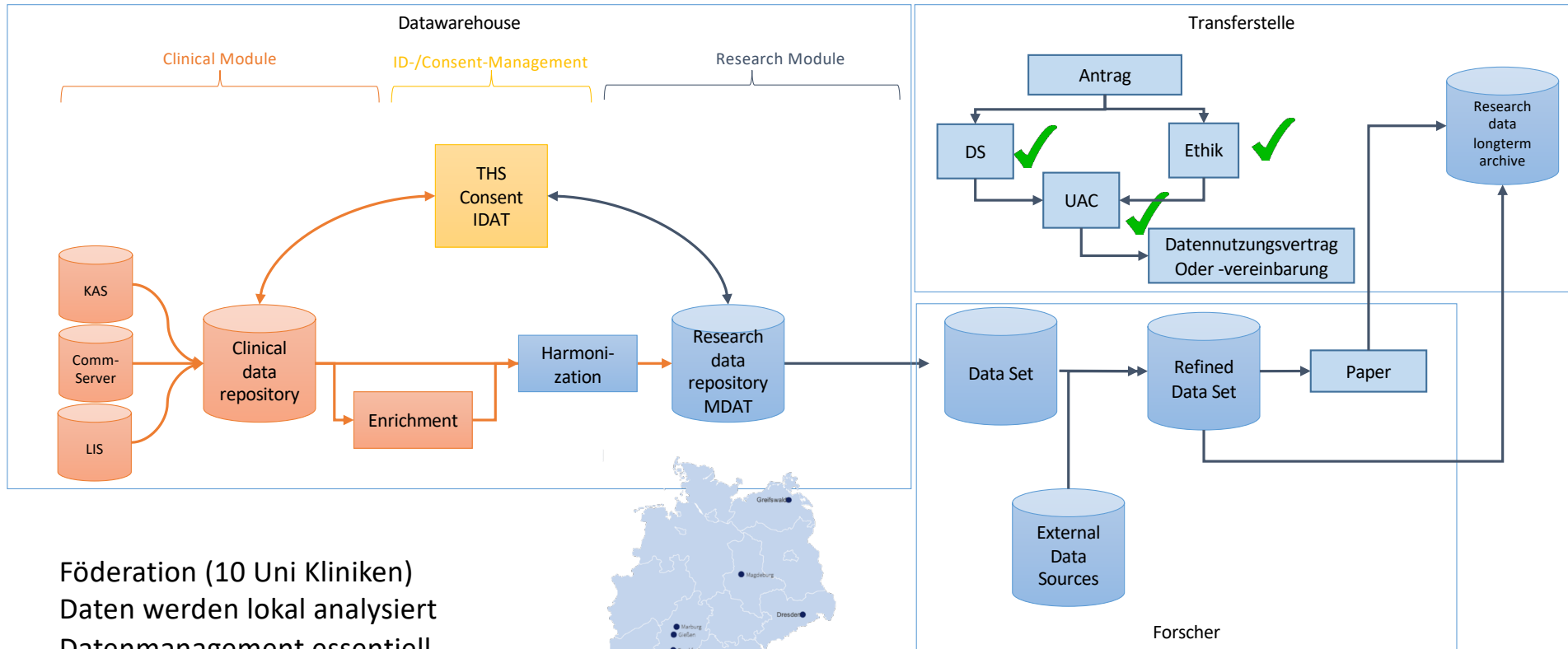
Medizininformatik Initiative: gezielte Erschließung des Versorgungsdatenschatzes



Medizinische Fakultät Mannheim der Universität Heidelberg
Universitätsklinikum Mannheim



Konzept eines MIRACUM Datenintegrationszentrums(DIZ)

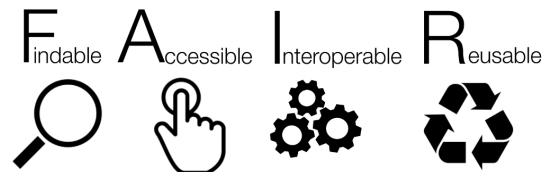
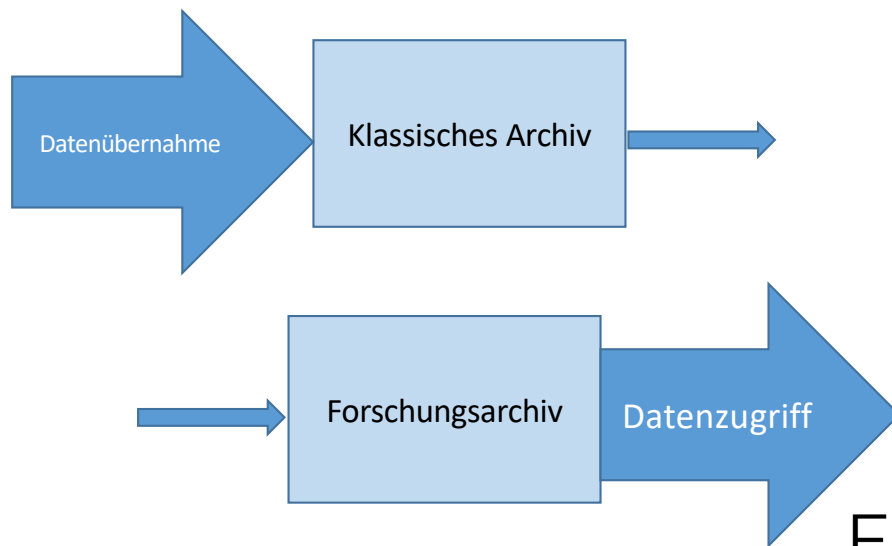


Föderation (10 Uni Kliniken)
 Daten werden lokal analysiert
 Datenmanagement essentiell



Was soll ein Forschungsarchiv leisten

Daten sollen wiederverwendet werden!



Ähnlichkeiten eines Archiv mit einem Data Lake



Analogien

- Unveränderliche Daten (immutable data)
- Heterogene Daten (Strukturiert aber kein einheitliches Schema)
- Temporale Daten

Unterschiede

- Archiv steht am Ende eines Prozesses
- Data Lake steht am der Anfang

Konvergenzen

- Data Lakes nähern sich Archiven an (Meta-Daten und Annotationen werden wichtiger)
- Wir wollen unser MIRACUM Archiv an dem Data Lake Konzept orientieren



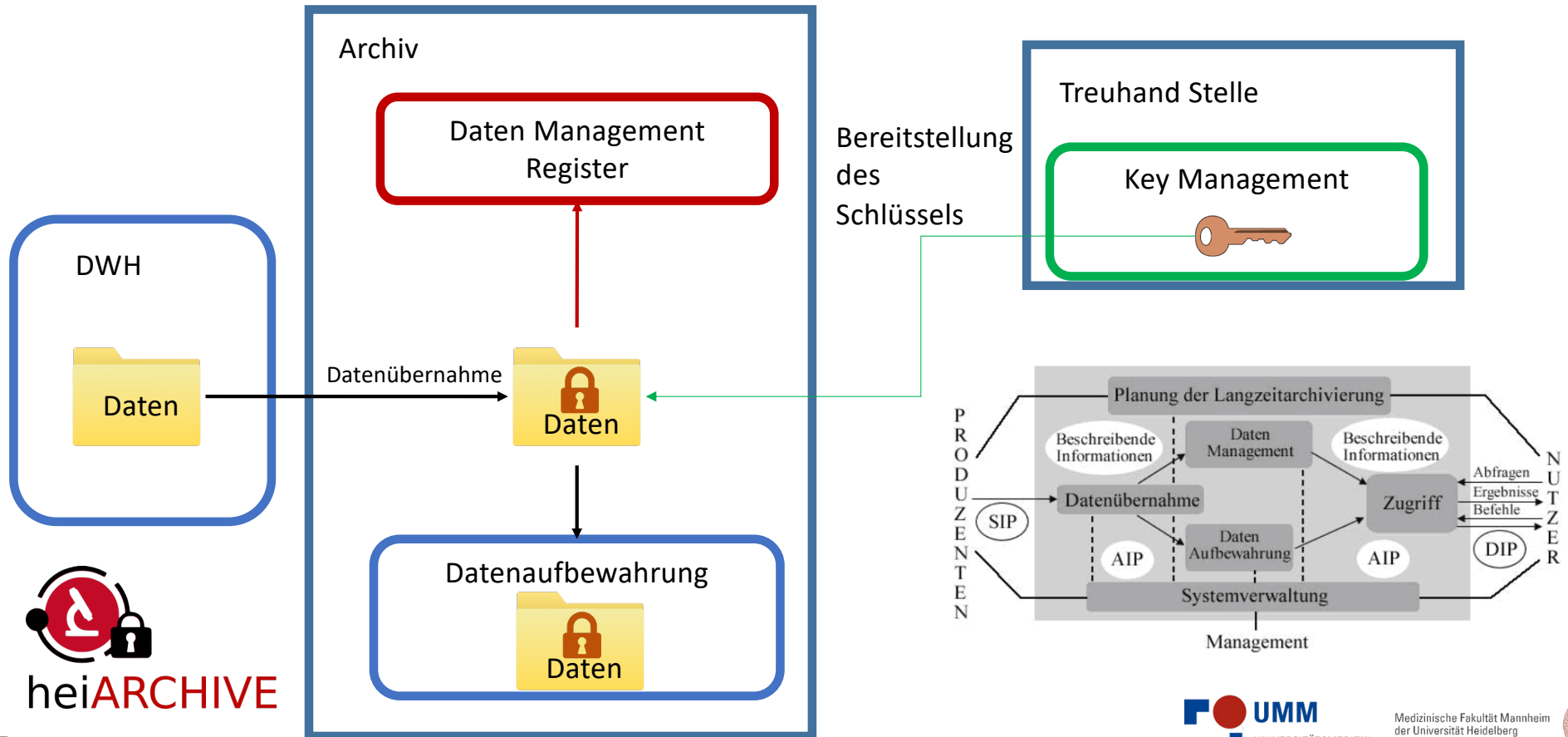
MIRACUM Archivierungskonzept



- **Föderiert - aber jedes DIZ soll auch isoliert funktionieren**
- **Daten müssen ohne Probleme lokal und/oder global (cloud) archiviert werden können**
- **Datenschutzkonform**
- **Verschlüsselte Datensätze**
- **Register globale Komponente - aber jedes Register soll auch isoliert funktionieren**
- **Datenmanagement Integraler Teil**
- **Wie sieht es in der Praxis aus?**



Datenübernahme



Nestor Handbuch, 7 Das Referenzmodell OAIS, Neuroth et al.,



Föderation der Archive

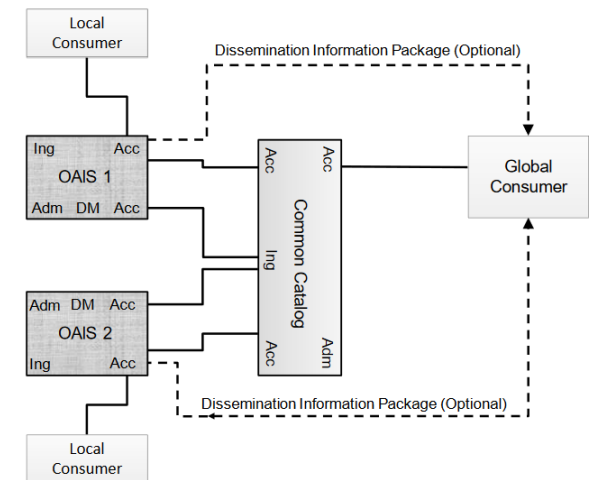
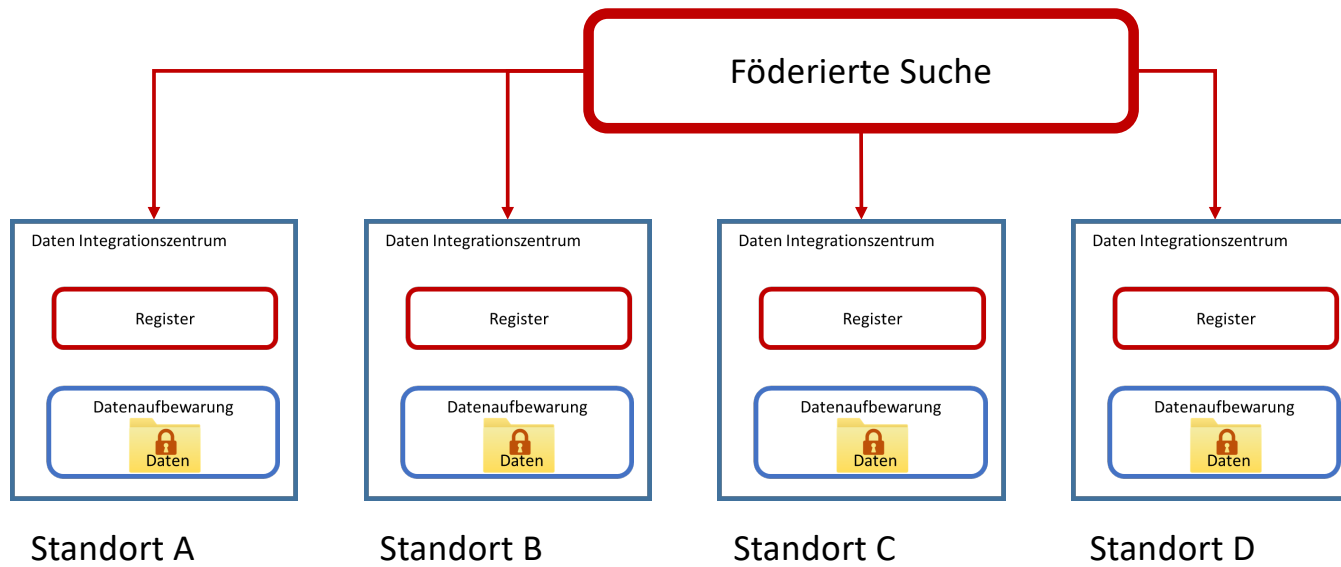
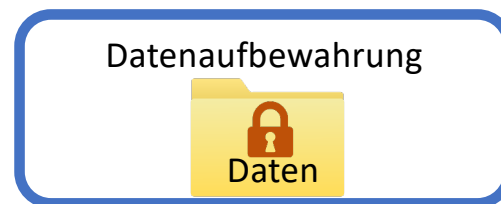
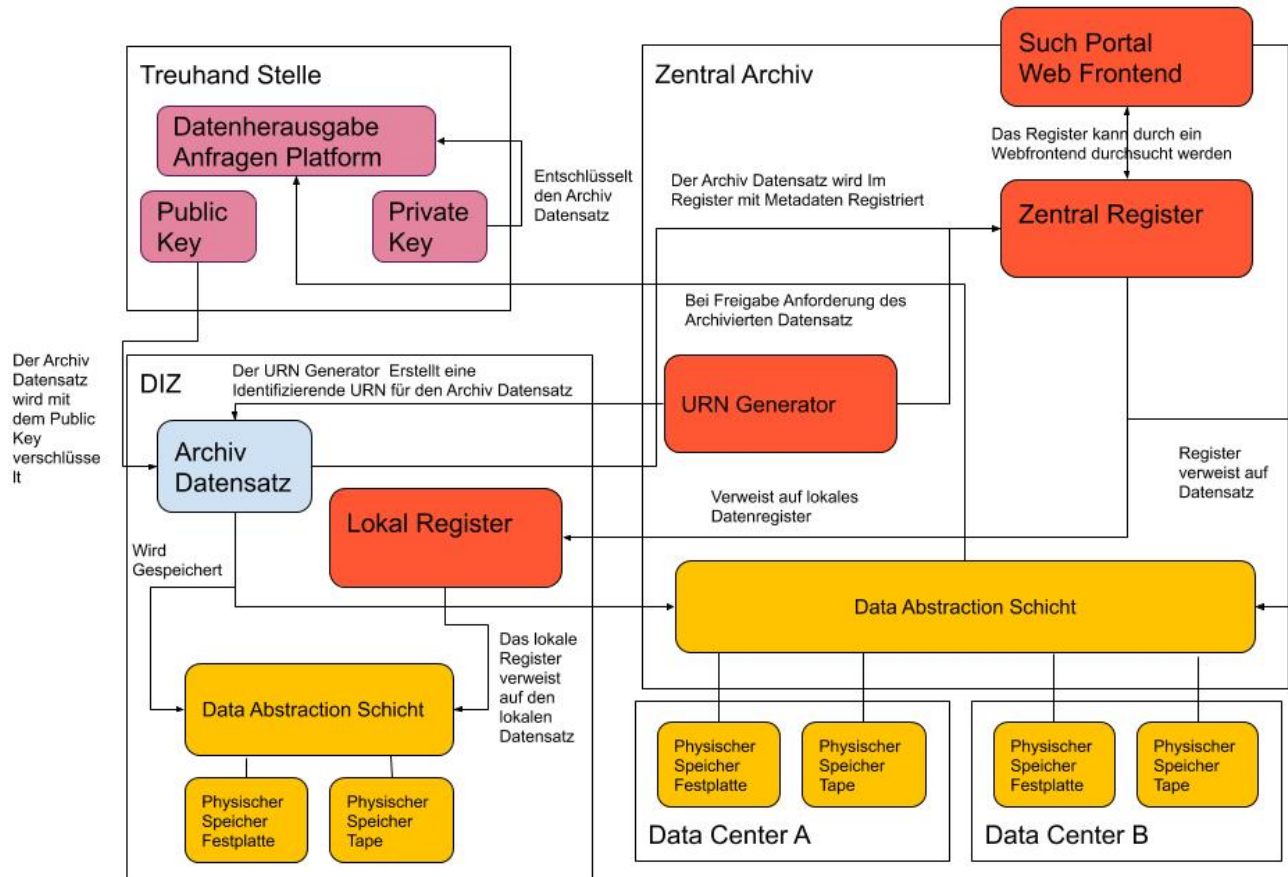


Figure 6-3: An OAIS Federation Employing a Common Catalog

CCSDS, Magenta Book, S. 6-5



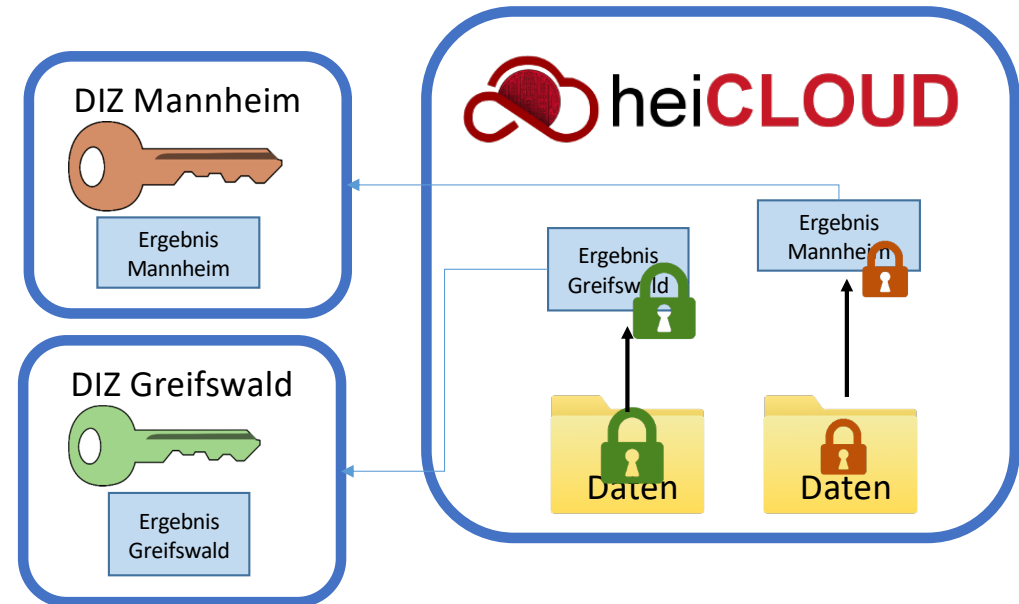
Architektur



Medizinische Fakultät Mannheim
der Universität Heidelberg
Universitätsklinikum Mannheim



- Was fehlt noch?
- Zugang zu Daten muss einfach und schnell sein
- Daten (Freigaben)
- Datenmanagement
- Für jedes Datenelement muss klar sein, wie die Datenverfügbarkeit geregelt ist
- **Verschlüsselung**
- Homomorphische Verschlüsselung
- Analyse verschlüsselter Daten die verschlüsselte Ergebnisse Produziert



Take-Home Message



- **Versorgungsdaten Daten sind wertvolle Datenschätze**
- **Funktion eines Forschungsarchiv ist die Sekundärnutzung**
- **Um die Daten nutzen zu können, müssen sie aufgereinigt werden**
- **Datenmanagement**
- **Das MIRACUM Forschungsarchiv orientiert sich an einem Data Lake**
- **Vorstellung eines Förderierten teilweise Cloudbasierten Archivkonzept**

- **Datennutzung wird in Zukunft in den Vordergrund treten**
- **Menge der Daten wird steigen (Cloud)**



Danke Für ihre Aufmerksamkeit



Bei weiteren Fragen:

maximilian.fuenfgeld@medma.uni-heidelberg.de

[linkedin.com/in/fuenfgeld](https://www.linkedin.com/in/fuenfgeld)



Medizinische Fakultät Mannheim
der Universität Heidelberg
Universitätsklinikum Mannheim

