

MedInfoGRID



**Provider für Integrierte Medizinische Information:
 Bilddaten, Therapieoptionen, Dokumentation,
 Forschung**

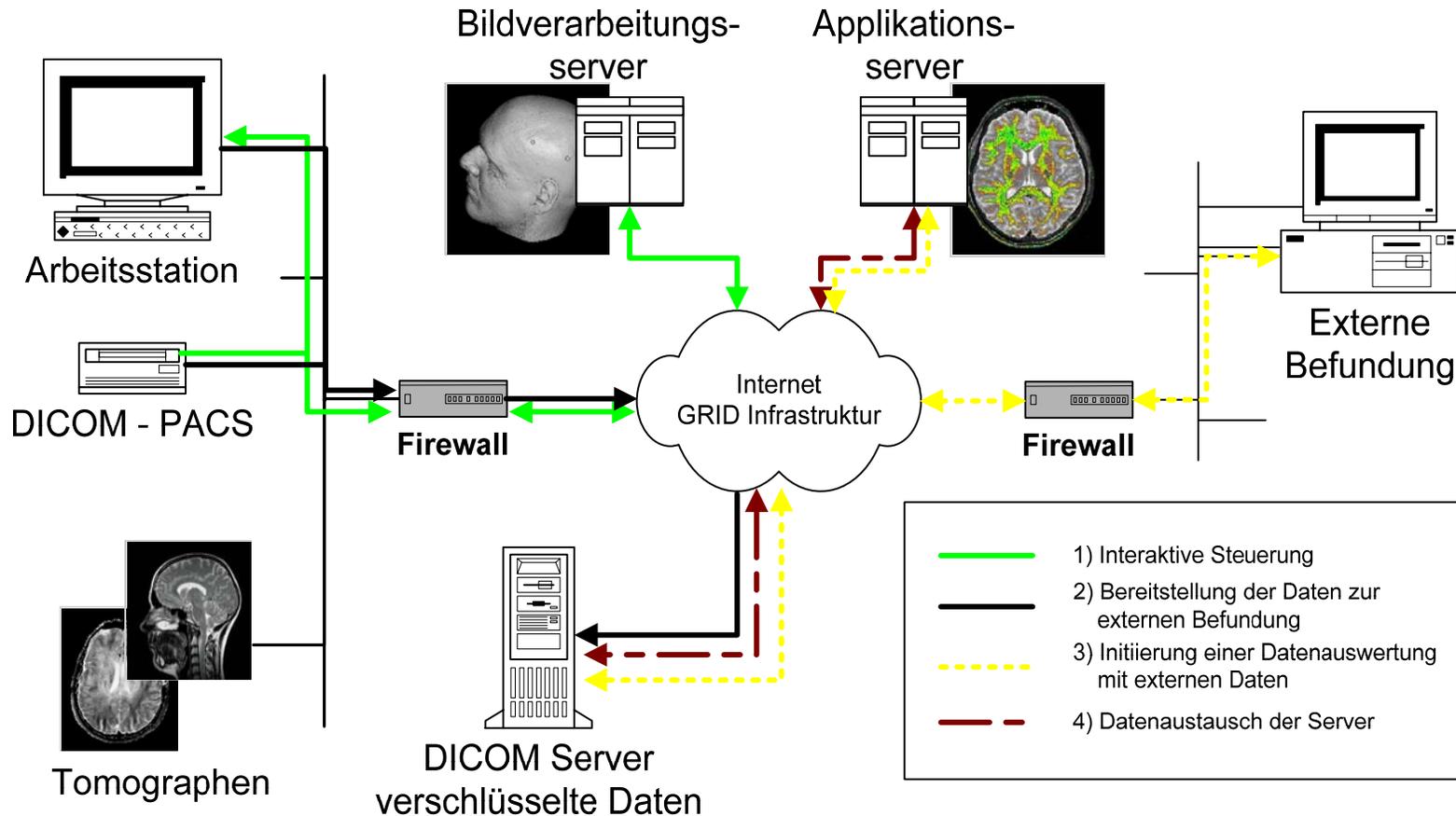
Vertikales ServiceGRID für die biomedizinische
 Verbundforschung

Andreas Thiel

Projektschwerpunkte

- Provider für Dokumentation von Tumorerkrankungen unter Einschluss digitalisierter pathologischer und radiologischer Bilddaten mit Kompatibilität zu anderen großen Initiativen wie caBIG (ServiceGRID)
- DataGRID aus hochkomplexen medizinischen Datenstrukturen durch Zusammenbindung von Befund- und Bilddaten (bis zu mehreren Gigabyte pro Fall) unter Beachtung medizinischer Standards (DICOM, IHE)
- Aufbau eines Bereiches zur Erprobung verschiedener Algorithmen (Computing und KnowledgeGRID) wie Kompression, Bildverarbeitung, Pseudonymisierung.

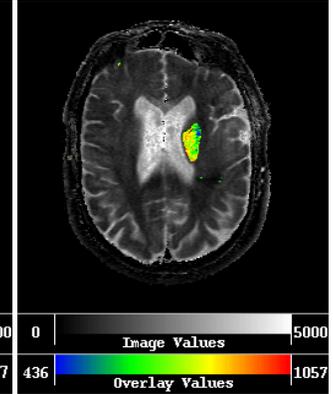
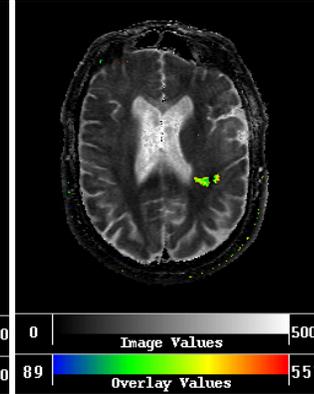
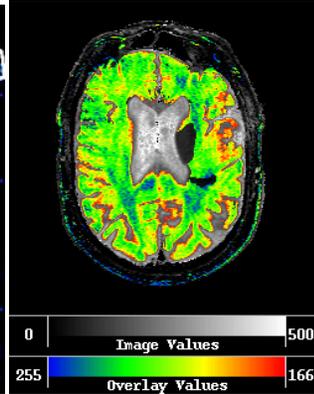
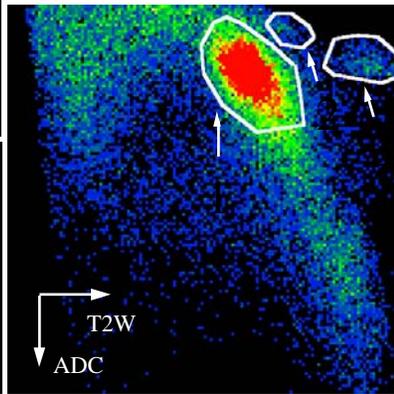
Szenarium



Forschung: Infarktverlaufskontrolle

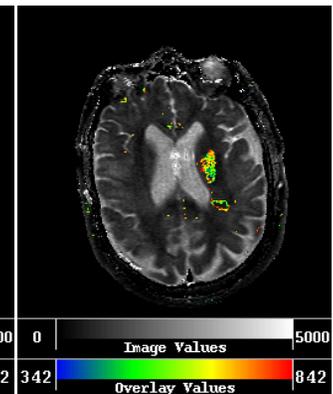
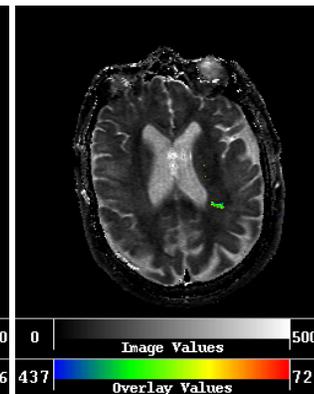
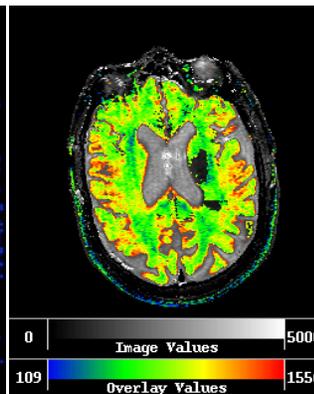
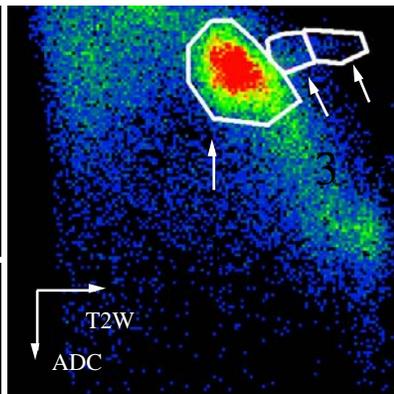
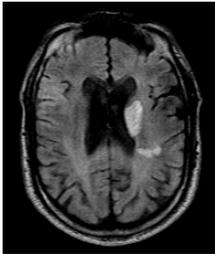
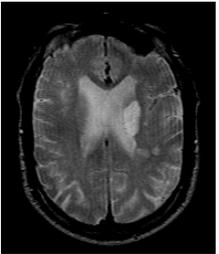
Tag 6

ADC

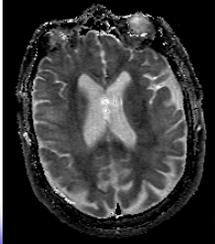


T2w

DWI

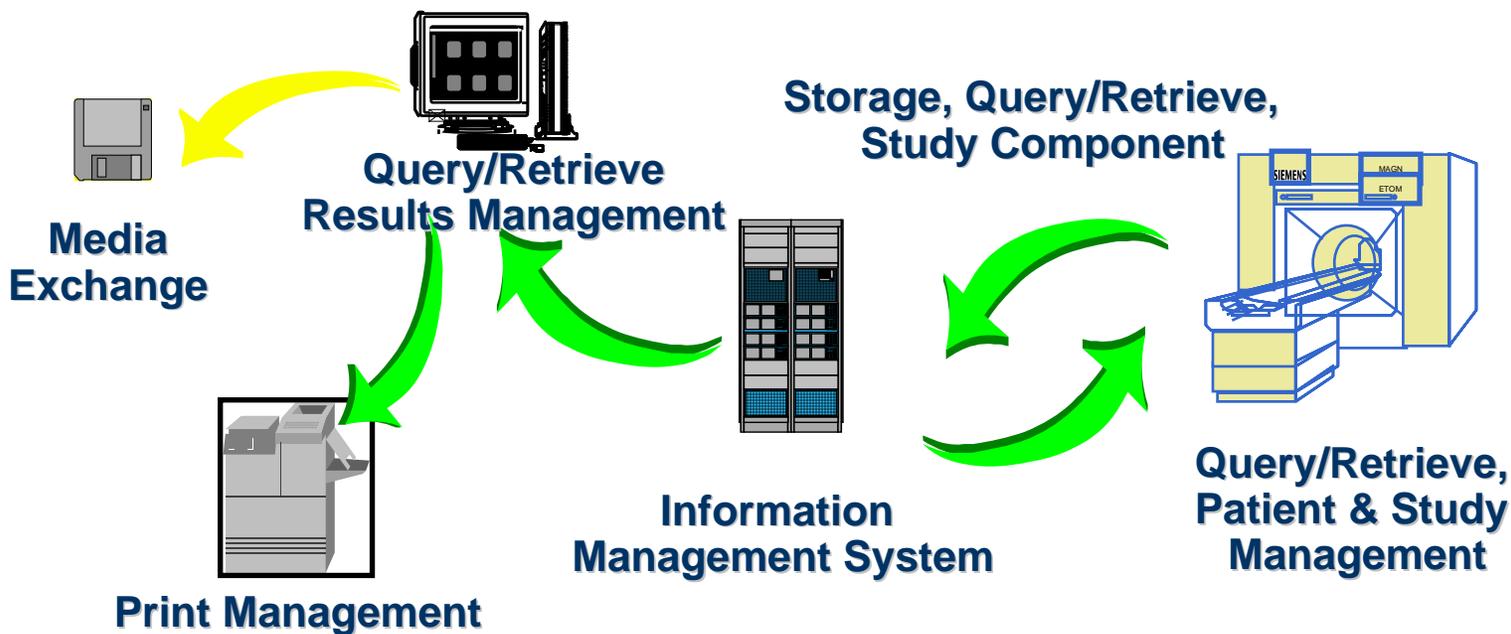


ADC

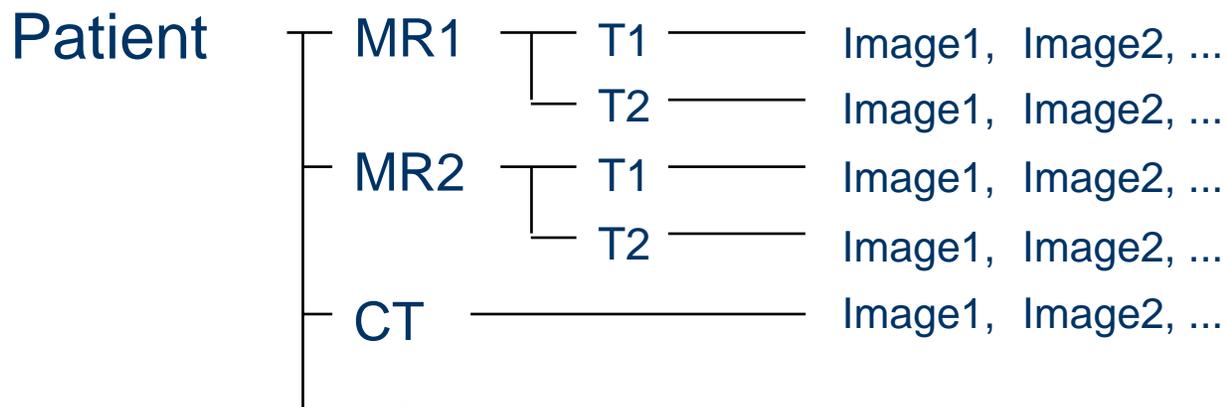


Tag 13

Digital Imaging and COmmunications in MEdicine



Datenstrukturen in DICOM



Eindeutigkeiten werden durch DICOM UID sichergestellt

DICOM UID

DICOM UID besteht aus:

Prefix:

1.3.12.2.1107. = Siemens

1.3.46.670589. = Phillips

und vom Anwender eindeutig zur haltendes Postfix

xxx.**11.0.4.1996051510410006**

Gerät / Seriennummer

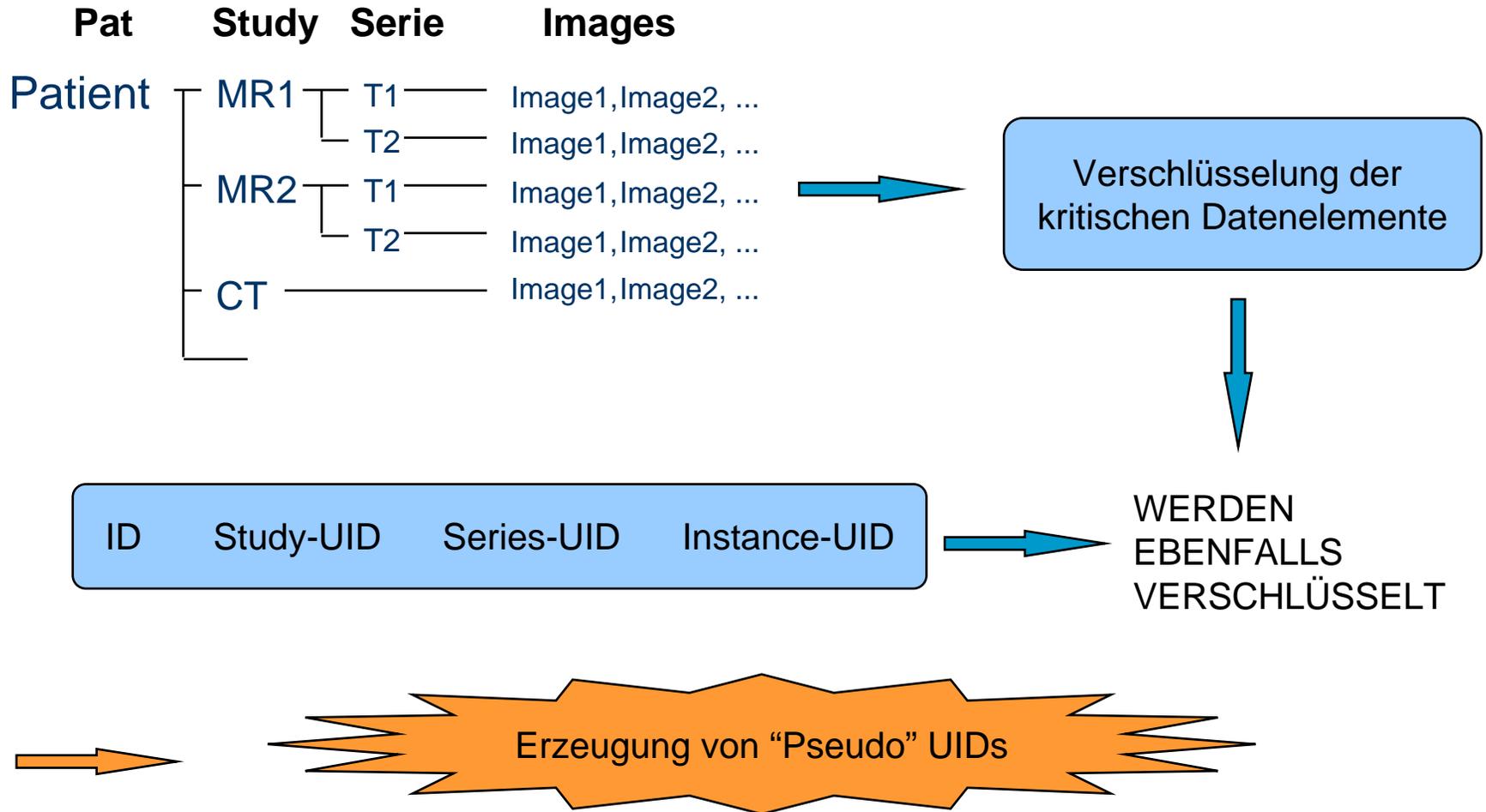


**DICOM UID lassen die Herkunft der Bilder erkennen.
Auch die UIDs müssen pseudonymisiert werden!**

Anforderungen am Workflow

- **Daten die extern bearbeitet werden müssen pseudonymisiert / verschlüsselt werden**
- **Applikationen haben keine Befugnis die Daten zu entschlüsseln**
- **Aufwändige Tabellen, die die Zuordnung Pseudonym – Originaldaten bereithalten sind inakzeptabel**
 - Sender – Empfänger können unterschiedliche Standorte haben
 - Die Tabelle müsste über Jahrzehnte gepflegt werden

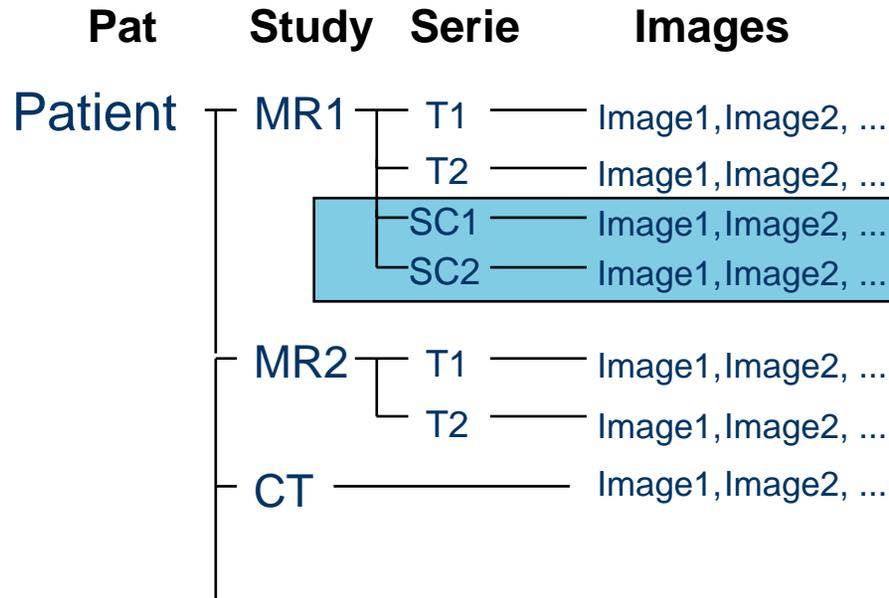
Bearbeitung der Daten



Nachbearbeitung der Daten

Entschlüsselung und Einsortierung
der Bilder in die Datenstruktur

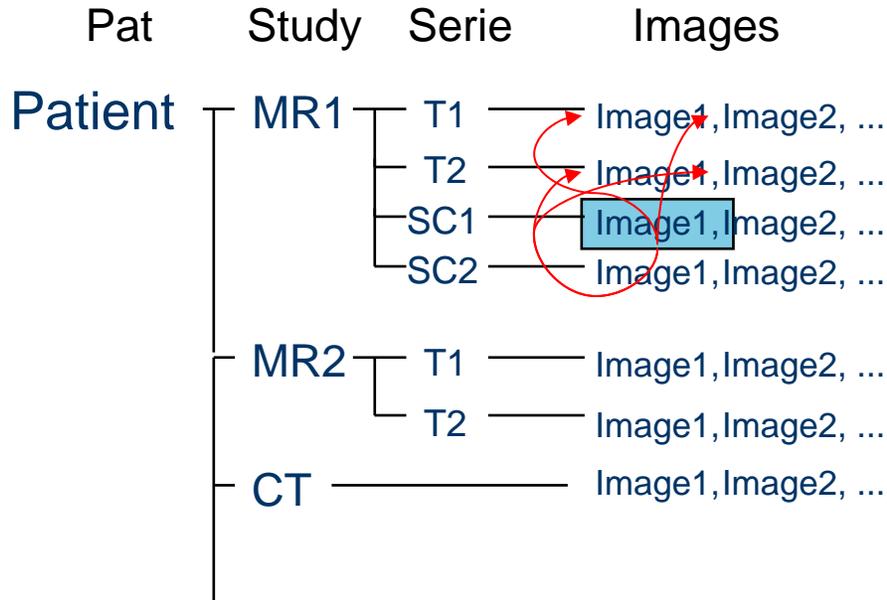
**Bild/Daten-
erzeugung
im
GRID**



Daten müssen richtig
einsortiert, obwohl die
erzeugende Applikation
keine Strukturinformationen
hatte

Nachbearbeitung der Daten

Interne Abhängigkeiten
Innerhalb der Datenstruktur



Die Bilder verweisen auf die Original Bilder von denen sie abgeleitet wurden

Herausforderungen

- **Die Daten müssen Standardkonform sein (DICOM, HL7...)**
- **Daten die auf externen Rechner erzeugt werden, verweisen auf die Pseudonyme**
 - Neu erzeugte Bilder müssen in die Patient – Studie – Serie Hierarchie integriert werden
 - Es müssen die Verweise auf die Original DICOM UID rekonstruierbar sein
- **Bei der Integration in eine medizinischen Datenbank muss die depseudonymisierende Stelle Wissen über die datenerzeugende Stelle und deren Vorgehensweise haben, um die Zuordnungen zu korrigieren**

Lösungsansatz / Projektziele

- **Portionsweise die Dateninhalte verschlüsseln, wobei die semantischen Elemente weiterhin im Klartext lesbar sind**
- **Integrieren dieser Daten innerhalb des Datensatzes**
- **Löschung/Dummy Werte treten an die original Position**

- **Verschlüsselte Daten können in andere Formate transferiert werden**
 - Z.B. DICOM SR->HL7 CDA / DICOM->XML
- **Repseudonymisierung ohne Vorhaltung von Transfertabellen ermöglichen**

Anforderung: Patientenidentifikation

Ziel

- Suche nach Organisationen/Datenquellen, die eine notwendige medizinische Information für einen spezifischen Patienten haben

Naive Lösung

- Versenden der die den Patienten identifizierenden Daten an alle Datenquellen (Krankenhäuser ...) und warte auf antworten

Hemmnisse

- Gesetzliche Regelungen
 - EU Empfehlungen und Direktiven / Datenschutz
 - Informationelle Selbstbestimmungsrecht
- Unterschiedliche identifizierende Daten sind in den Systemen vorhanden
- Es gibt (noch) keinen einheitlichen Patientenummer /ID

Patientenidentifikation

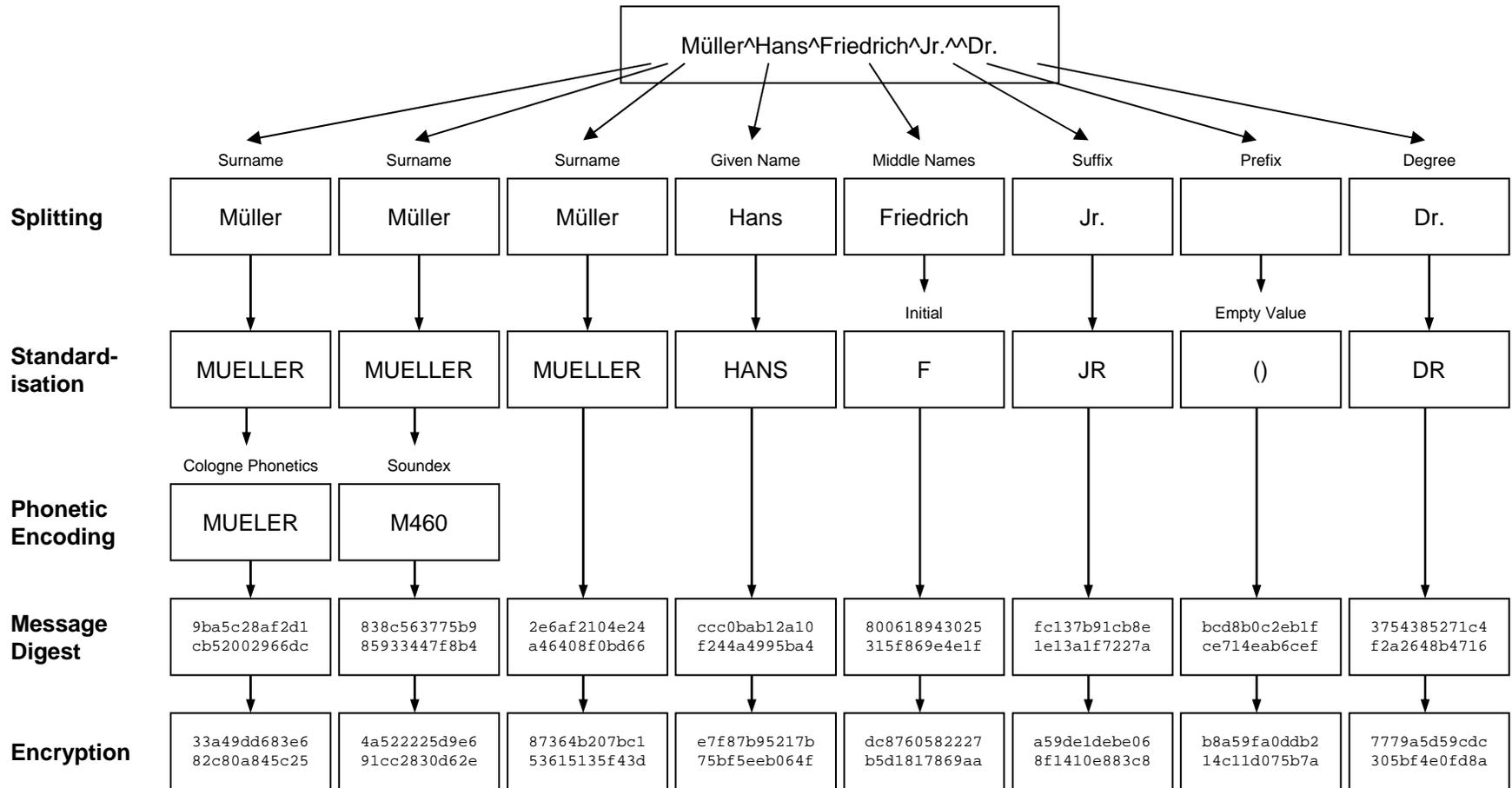
Anforderungen

- Dynamisch: Eigenheiten sind oft unterschiedlich in den Systemen und müssen berücksichtigt werden
- Skalierbar: Viele Datenknoten müssen befragt werden und deren antworten müssen verarbeitet werden
- Sicherheit: Kein Dritter darf die Patientendaten einsehen, auch nicht die Datenquellen nicht (In einer Anfrage steckt schon Information)
- Fehlertoleranz: Trotz Tippfehler und fehlende Werte bzgl. der Patienten identifizierenden Daten sollen die medizinischen Daten gefunden werden

Lösungskonzept basiert auf

- Kontrollnummern
 - Verschlüsselte Hashes der Patienten identifizierenden Daten
- Rekord-Linkage

Kontrollnummer Generierung



Kontrollnummer Vergleich

id	weight	given name	surname	...	address	Pl. of birth	Valid through	Link
1	0,99			...			31.07.2004 15:30	Hyperlink to service
2	0,83			...			01.08.2004 00:12	Hyperlink to service
			
			
			
N	0,55			...			15.09.2004 12:17	Hyperlink to service

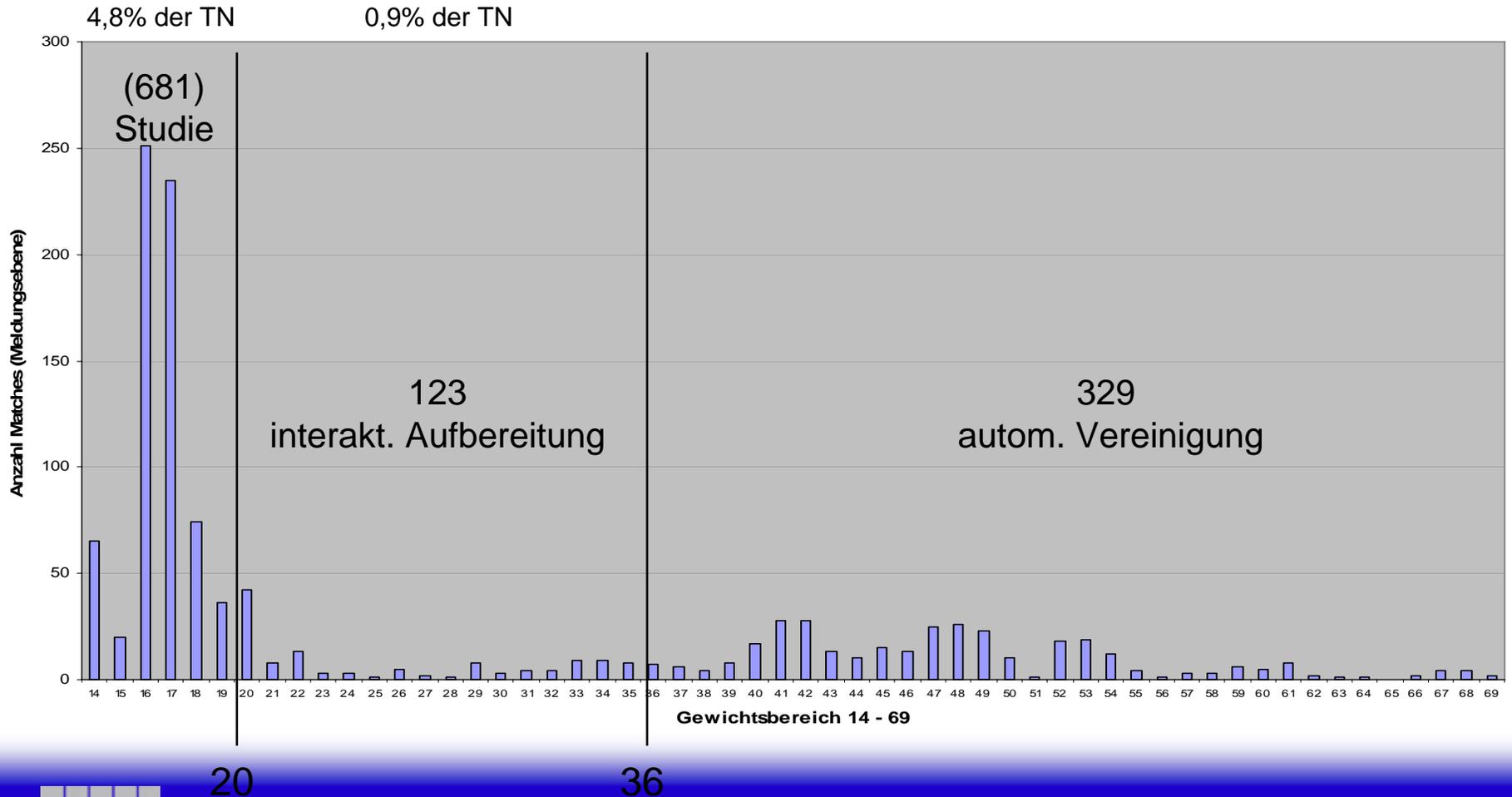
Probabilistisches Record-Linkage-Verfahren

Vergleich der Kontrollnummern zweier Meldungen

- **Gewichtsberechnung der Attributwerte**
- **Maß für die Wahrscheinlichkeit der Zugehörigkeit der Datensätze zu einer Person**
- **Es wird ein Gewicht als Maß für die Wahrscheinlichkeit der Zusammengehörigkeit zweier Datensätze vergeben.**
 - Wert < 19 keine Personenübereinstimmung
 - Wert > 36 automatische Zusammenführung
 - Dazwischenliegende sind unklare Matches

Beispiel Datenabgleich – Probabilistisches Record Linkage

Gewichtsverteilung Datenabgleich MSWE - EKN, 1. Screening-Runde
(n = 14.135 Teilnehmerinnen, 1.133 Matches ICD-10 C50+D05, EKN-Stand Juli 2005)



Zusammenfassung

- **Wir benötigen Verfahren zur sicheren (pseudonymen) Verarbeitung med. Daten, die:**
 - Strukturerhalten sind
 - Sich in bestehende Standards integrieren lassen
 - Feingranulare Sichtweisen auf die Daten zulässt
- **Gewünscht sind in Zukunft**
 - Eine Zusammenführung von Patientendaten ohne Verletzung der informationellen Selbstbestimmung
 - Studien über einen flächendeckenden Datenpool
- **Und GRID?**
 - Infrastrukturen via VO
 - Integration der Gesundheitskarte/ Heilberuferausweis