



---

# Sicherheitsanforderungen bei der Langzeitarchivierung medizinischer Forschungsdaten – auf Grid-Basis

---

**Security-Workshop NGI-DE Jahrestreffen**

**19. Mai 2011, Garching**

**Fred Viezens**



- Nachvollziehbarkeit/Provenienz des Workflows bzw. der Entstehung und Weiterverarbeitung von Forschungsdaten sicherstellen
- Intellektuelle Nachvollziehbarkeit durch sinnvolle Metadaten (hier nicht die technischen Metadaten, wie Datenformat, Dateierzeugungsdatum etc.) ermöglichen
- Einhalten von Datenschutzvorschriften – auch im Hinblick auf bisher unbekannte Nutzungsmöglichkeiten von Daten
- Langfristigen Zugriff auf Forschungsdaten realisieren

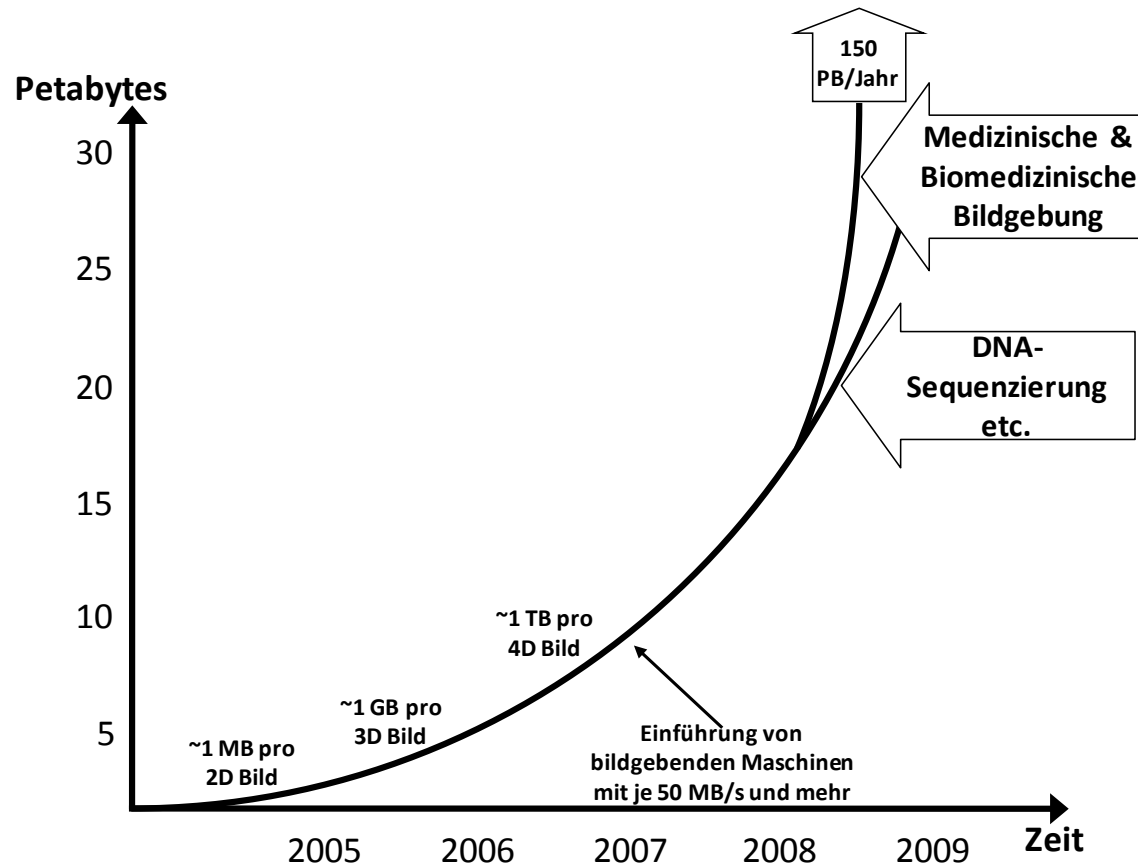
- Datenschutzerfordernngen werden nicht geringer im Zeitablauf
  - Genomische Daten betreffen auch die Folgegenerationen
- Datenmengen spielen eine wichtige Rolle
  - Nicht alle Daten sind für die Forschung relevant
  - Datenschutz empfiehlt nur notwendige Daten zu speichern, daher Auswahl von Forschungsdaten in den Arbeitsprozessen realisieren (Tendenz: lieber alles sichern; dies kann auf Dauer nicht gehalten werden!)
- Einfacher Zugriff
  - Mediziner bevorzugen keine Kommandozeile, aber einfache Workflows

- Es existieren unterschiedliche Aufbewahrungsfristen:
  - Z.B. RöntgenVO von 10 Jahre bis zu 30 Jahre
  - FDA-Compliance 15 Jahre (Trial Master File bei Studien mit Ziel einer Zulassung der Therapie in den USA)
  - Beweiskräftigkeit bei klinischen Studien ist notwendig
    - Der Umgang mit Zeitstempeln bei Datenbanken ist z.B. noch nicht gelöst (Zertifikatskette, dynamisches Schreiben, etc.)

- Die Menge medizinischer Daten wächst exponentiell
  - Dies ist eine direkte Folge der Weiterentwicklung von Möglichkeiten der Datenerhebung (z.B. Gen. Sequenzierung und MRT)
  - Dies gilt für Versorgung und Forschung, die in ihrer Ausrichtung aber klar abgegrenzt werden müssen
- Daten, die bislang erhoben worden sind, erhalten immer weitere Bedeutungen, z.B. aus Langzeitstudien und epidemiologische Erhebungen
  - Helmholtz-Kohorte, eine seit über 20 Jahren laufende Studie
  - Nachnutzung von in der Mayo-Clinic bis in die 1930-er Jahre zurückliegenden digitalisierten Patientendaten

- Medizinische Daten aus der klinischen Versorgung
  - Sind immer mit Personenbezug verbunden
  - Werden zur rechtlichen Absicherung erhoben
  - Dienen als Grundlage für die Abrechnung medizinischer Leistungen
  - Haben nicht den gleichen Qualitätsfokus wie die Forschung

→ Geplant ist der Secondary Use Case
- Forschungsdaten aus der präklinischen- bzw. Grundlagenforschung sowie aus Epidemiologie und Biostatistik haben i.d.R. keinen Personenbezug mehr



- weltweit geschätztes medizinische Datenaufkommen (medizinischen Bildgebung/ DNA-Sequenzierung und sonstige Daten)

# Datenschutzmethoden in der Medizin (1)



- Die medizinische Forschung setzt verstärkt IT-Plattformen für Forschungsnetze / vernetzte Forschung ein
  - Identifizierende Daten (IDAT) dürfen nur dem untersuchenden Arzt zur Verfügung stehen
  - Medizinische Daten (MDAT) sind für die Forschung im Verbund relevant und sollen gemeinsam genutzt werden
  - Neue Erkenntnisse / Verfahren führen zu möglichen Re-Identifikationen aufgrund vorliegender Daten
  - k-Anonymität ( $1/k$ ;  $k$  ist dabei die Anzahl an Individuen mit einem entsprechendem Identifikationsmerkmal)
  - Die Vertraulichkeit / ärztliche Schweigepflicht muss gewährleistet werden





# Datenschutzmethoden in der Medizin (2)

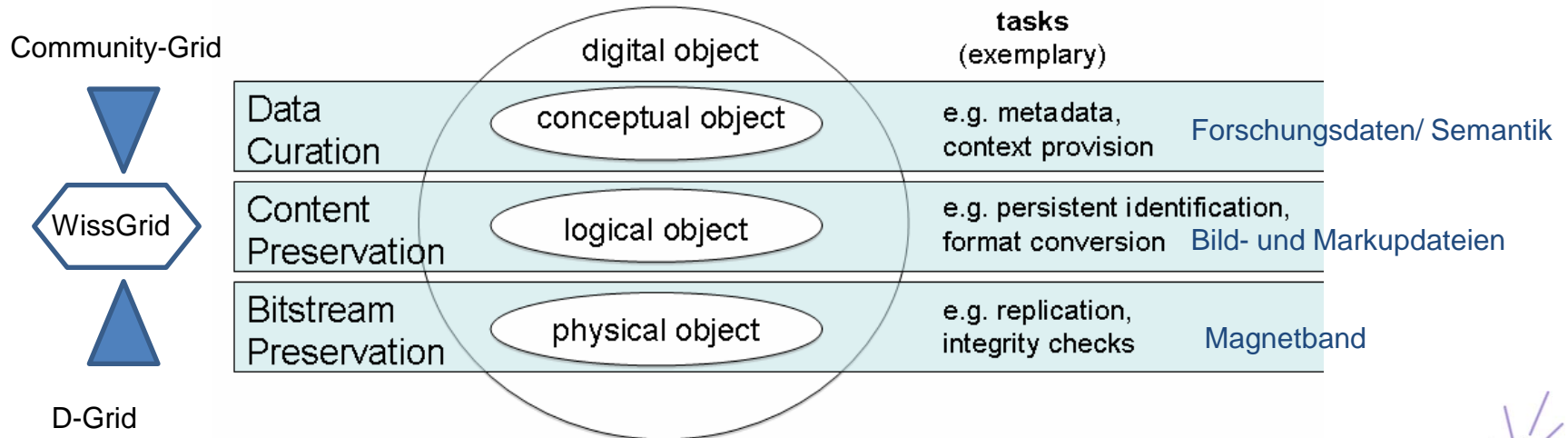


<b>Merkmal</b>	<b>Privatsphäre</b>	<b>Transparenz</b>	<b>Umkehrbar</b>	<b>Bemerkung</b>
<b>Methode</b>				
Anonymisierung	+	-	Nein	IDAT werden entfernt
Verschlüsselung	+	-	Ja	Daten werden komplett verschlüsselt
Pseudo-nymisierung	+	+	Unter definierten Bedingungen	IDAT werden durch Pseudonym ersetzt

T. Neubauer et al. (2010): Pseudonymisierung für die datenschutzkonforme Speicherung medizinischer Daten. In: Elektrotechnik & Informationstechnik, 5. 2010 (127), S. 135-142.



- Lösungen auf Basis von Grid-Technologie sollen erarbeitet werden innerhalb von
  - WissGrid
    - entwickelt/ konzeptioniert generische Lösungen zur LZA, getrieben aus den Communities anhand von Checklisten (Blaupausen)



## – LABIMI/F (DFG)

- präzise Anforderungen an die Langzeitarchivierung biomedizinischer Forschungsdaten
- Laborimplementierung einer Grid-basierten Lösung für zwei Use Case:
  - Bild- und
  - Genomdaten

- Benutzbarkeit und Akzeptanz sicherstellen
  - Möglichst optimierte Eingabe/Ingest in dauerhafte Speicherlösungen
  - Einfachen Zugriff bei hohem Sicherheitsbedarf dauerhaft gewährleisten
- Fachlich relevante Forschungsdaten selektieren
  - Überblick bewahren
  - Wirtschaftlichkeit sicherstellen



## Universitätsmedizin Göttingen

Medizinische Informatik

<http://www.mi.med.uni-goettingen.de/>

**Fred Viezens**

Computational Medicine und Grid-Computing

[Fred.viezens@med.uni-goettingen.de](mailto:Fred.viezens@med.uni-goettingen.de)

Tel.: (0551) 39 - 13125