



10/9/2012

# Medizinische Forschungsdaten

Das Spannungsfeld zwischen Datenschutz

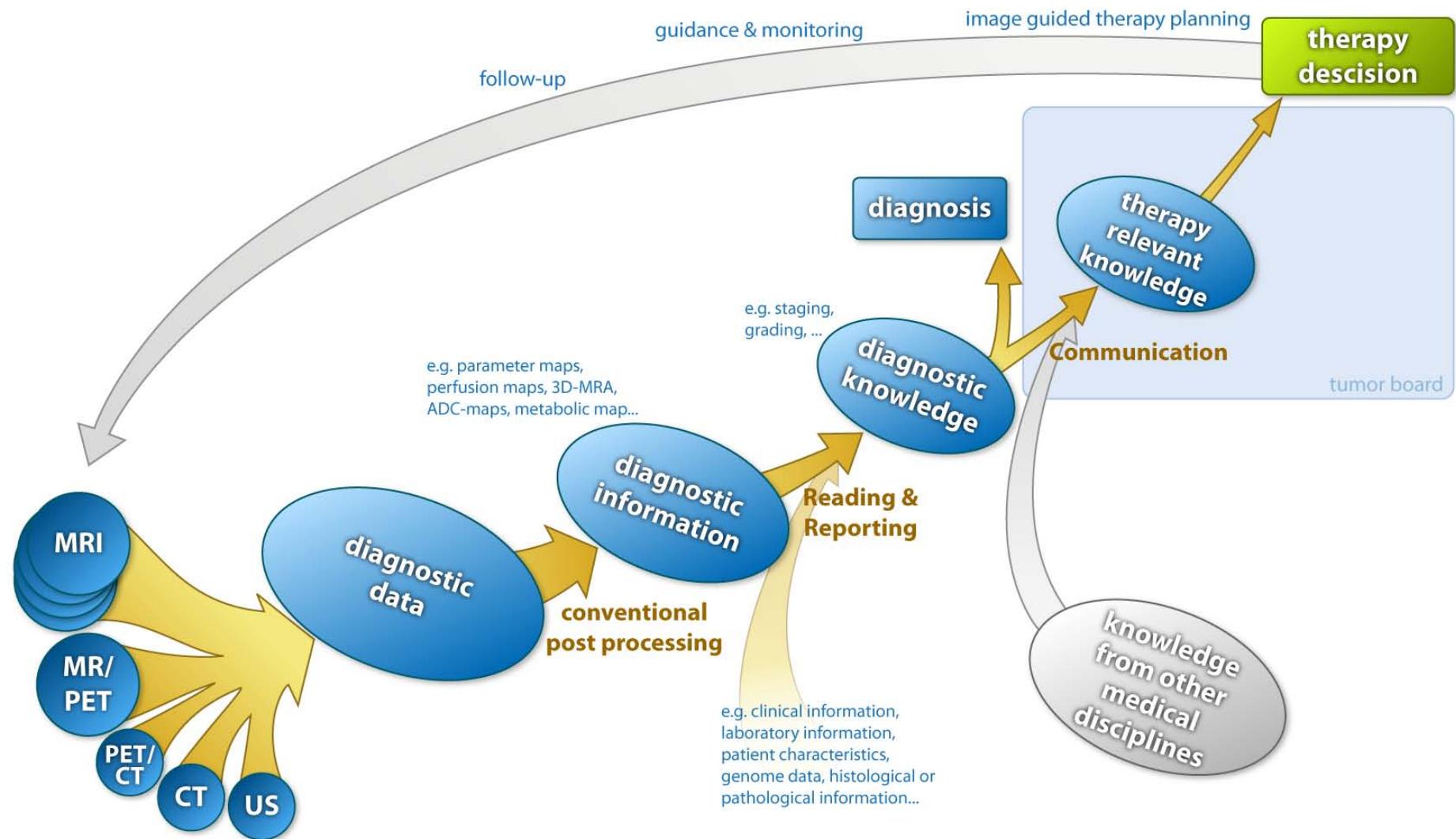
und moderner Medizinforschung am Beispiel der

Krebsforschung

Ralf Floca

- Vorstellung Radioonkologie
- Situation/Rolle des Patienten
- Anforderung der medizinischen Forschung
- Wieso ist der Datenaustausch wichtig?
- Herausforderungen

# Diagnostische Kette in der Radioonkologie





# Radioonkologie in Zahlen

Kenndaten Allgemein:

Pro Jahr:

~11 TB Daten → in Zukunft ~40 TB Daten

~10.000 Untersuchungen/Jahr

Nur DKFZ; wird durch Kooperationsinitiativen wie das Deutsches Konsortium Translationale Krebsforschung deutlich ansteigen.

Vergleich:

Sequenzierung des Genoms (nur Exons): ~200 GB/Patient

Bis 2015: 1000 Patienten pro Jahr -> 195 TB/Jahr



# Besondere Rolle/Situation des Patienten

## Daten-Akquise

- Ereignis nicht planbar
- Langer Beobachtungszeitraum
- Zusatzbelastung
- Compliance

## Daten-Nutzung

- Zustimmung des Patienten
  - Zur Behandlung
  - Zur Studie
  - Zur Nutzung seiner Daten spezifisch, erweitert oder unspezifisch ?

# Zwischen Versorgung und Forschung

## Dualität der Zielsetzung:

- Versorgung des Patienten; klinische Arbeit
- Forschung (im Rahmen von Studien)

Abhängig von der Zielsetzung unterschiedliche Ansprüche an

- Aggregation
- Archivierung
- Retrieval
- Datenschutz



# Anforderungen in der Forschung

## Aggregation

- Hohe Ansprüche
- zusammen Führung ganzer Kollektive
- Multizentrische Studien

## Archivierung

- Gegeben durch gute wissenschaftliche Praxis
- Anspruch „für immer“
- „Günstiger“ als Neugewinnung



# Anforderungen in der Forschung

## Retrieval

- Wichtig
- Hoher Aufwand für Identifikation und „Bergung“
- Suche: Profil spezifisch
- Berücksichtigung der Patientenakte

## Datenschutz

- Geeignete Pseudonymisierung
- Zugriffskontrolle

## Bedeutung von Austausch und Zusammenführung

„[...] neither Google nor Facebook would make a change to an advertising algorithm with a sample set as small as that used in a Phase III clinical trial.“

John Wilbanks, Sage Bionetworks

Kotz, J.; *SciBX* 5(25); 2012



# Bedeutung von Austausch und Zusammenführung

## Datenlage

- Viele Einflussgrößen  
Patientenbiologie, Tumorbiolegie, Lebensweise, Umweltfaktoren, Therapie
- Sehr heterogen  
Gesundheitswesen, Dokumentation (Formate/Medien), Arbeitsabläufe
- Zu geringe Fallzahlen in einzelnen Institutionen

→ Essentiell für:

- Medizinischer Fortschritt  
Validierung der Nützlichkeit von Verfahren
- Personalisierte Therapie
- Data intensive science / eScience  
„The 4th Paradigm: Data-Intensive Scientific Discovery“, Tony Hey et al., 2009

(OpenAccess; als Gegenentwurf zu kommerziellen Sammlungen)

# Bedeutung von Austausch und Zusammenführung

Was bedeutet Daten verfügbar machen?

- Offen
- Klassiert/normiert
- Maschinen lesbar
- Standarisierter Format

„Radiological images contain a wealth of information, such as anatomy and pathology, which is often not explicit and computationally accessible“

Rubin, DL.; AMIA Annu Symp Proc 626-630; 2008

# Herausforderungen

Wieso ist es so problematisch?

- Beteiligte
- Beispiele
  - Regulatorisch
  - Organisational
  - Technisch

## Beteiligte

- Patient
- Klinik
- Forscher
- Forschungseinrichtung
- Ethikkommission
- Geldgeber
- Gesetzgeber
- Weitere: etwa Forschungsadministratoren, IT-Administratoren...

# Herausforderungen - Organisationen

- Kulturelle Mechanismen fördern nicht den Austausch
- Setzen richtiger Anreize
  - Status Quo: Publikation Studienergebnissen zählt  
→ Daten sind „privatisiert“ für „Wettbewerbsvorteil“
  - Publikation von offenen Daten als Leistungsmerkmal  
Reputation durch viel „zitierte“ Daten
- Setzen richtiger Rahmen
  - Forderung der Veröffentlichung durch Förderprogramme  
z.B. Multiple Myeloma Research Foundation; Forderungen an die National Institutes of Health (USA); Empfehlung des National Health and Medical Research Council (Australien)



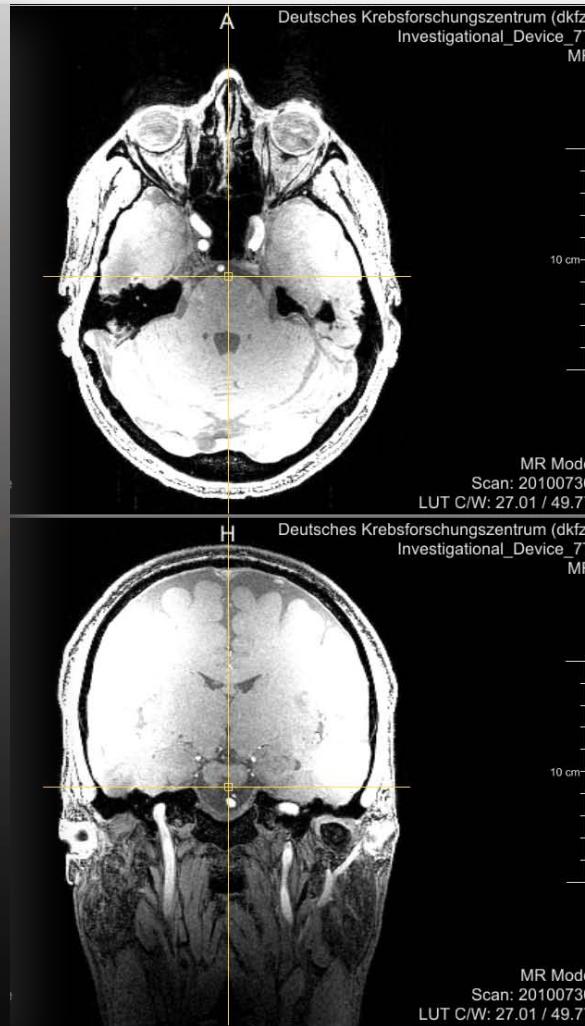
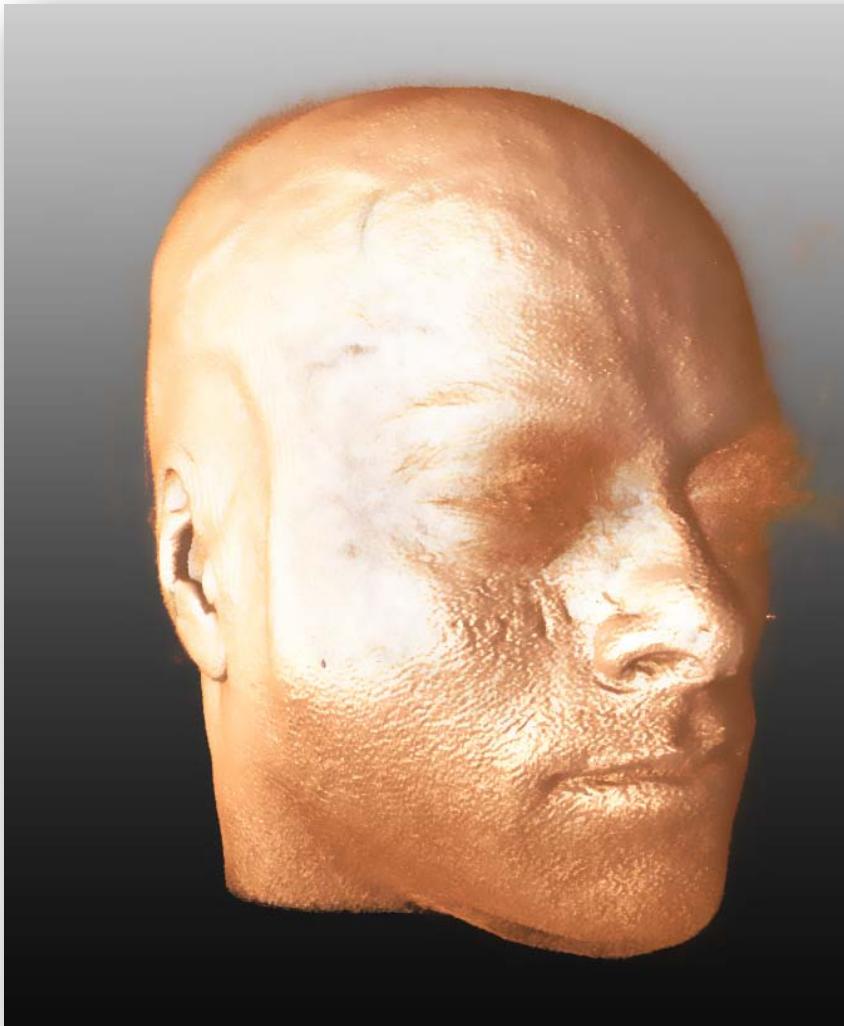


## Herausforderungen - Technisch

- Fehlzuordnung des Patienten
- Heterogenität bei Datenumfang und Struktur ⇒ Vergleichbarkeit
- Pseudonymisierung
  - Zunehmend schwierig
    - Gestiegenes Risiko durch Data-Mining
    - Hohe Kosten und Aufwände ⇒
    - Gezielte Abwägung von Nützlichkeit und Sicherheit
    - Generalisierung oder Weglassen der Daten wo möglich
  - Teilweise nicht möglich

# Beispiel für Problematik von Pseudonymisierung von Bilddaten

Erkennbarer Personenbezug?



Vielleicht...

Ja!

Vielen Dank für Ihre Aufmerksamkeit!

Fragen?

# Anhang

- Beispiele für Re-Identifizierung
- OA-Erfolgsstories



# Beispiele für Re-Identifizierung

## 1. AOL search data

AOL put anonymized Internet search data (including health-related searches) on its web site. New York Times reporters were able to re-identify an individual from her search records within a few days.

## 2. Chicago homicide database

Students were able to re-identify a 35% of individuals in the Chicago homicide database by linking with the social security death index.

## 3. Netflix movie recommendations

Individuals in an anonymized publicly available database of customer movie recommendations from Netflix are re-identified by linking their ratings with ratings in a publicly available Internet movie rating web site.

## 4. Re-identification of the medical record of the governor of Massachusetts

Data from the Group Insurance Commission, which purchases health insurance for state employees, was matched against the voter list for Cambridge, re-identifying the governor's health insurance records.

## 5. Canadian Adverse Drug Event Database

A national broadcaster aired a report on the death of a 26 year-old student taking a particular drug who was re-identified from the adverse drug reaction database released by Health Canada. The national broadcaster matched the information from the report to the publicly available obituaries for that area of Ontario.

# Open Access – Erfolgsstories

- „Science as an open enterprise“

Juni 2012 hat die Royal Society die Ergebnisse der Studie „Science as an open enterprise“ veröffentlicht (<http://royalsociety.org/policy/projects/science-public-enterprise/report/>). Der Bericht enthält eine Vielzahl von Informationen und Beispielen zu diesem Thema.

- Open access success stories

Die Webseite <http://www.oastories.org/> sammelt Geschichten rund um Open Access / Open Science, die die Vorteile für den Einzelnen und die Gesellschaft veranschaulichen sollen. Die Seite wird bereitgestellt durch die Initiative knowledge-exchange.info welche unterstützt wird durch die [Denmark's Electronic Research Library](#) (DEFF, Dänemark), die Deutsche Forschungsgemeinschaft (DFG), das [Joint Information Systems Committee](#) (JISC; UK) und [SURF](#) (Niederlande)