

# IT-Anforderungen der Nationalen Kohorte

Dr. Daniel Kraft  
Dipl.-Math. Stefan Ostrzinski

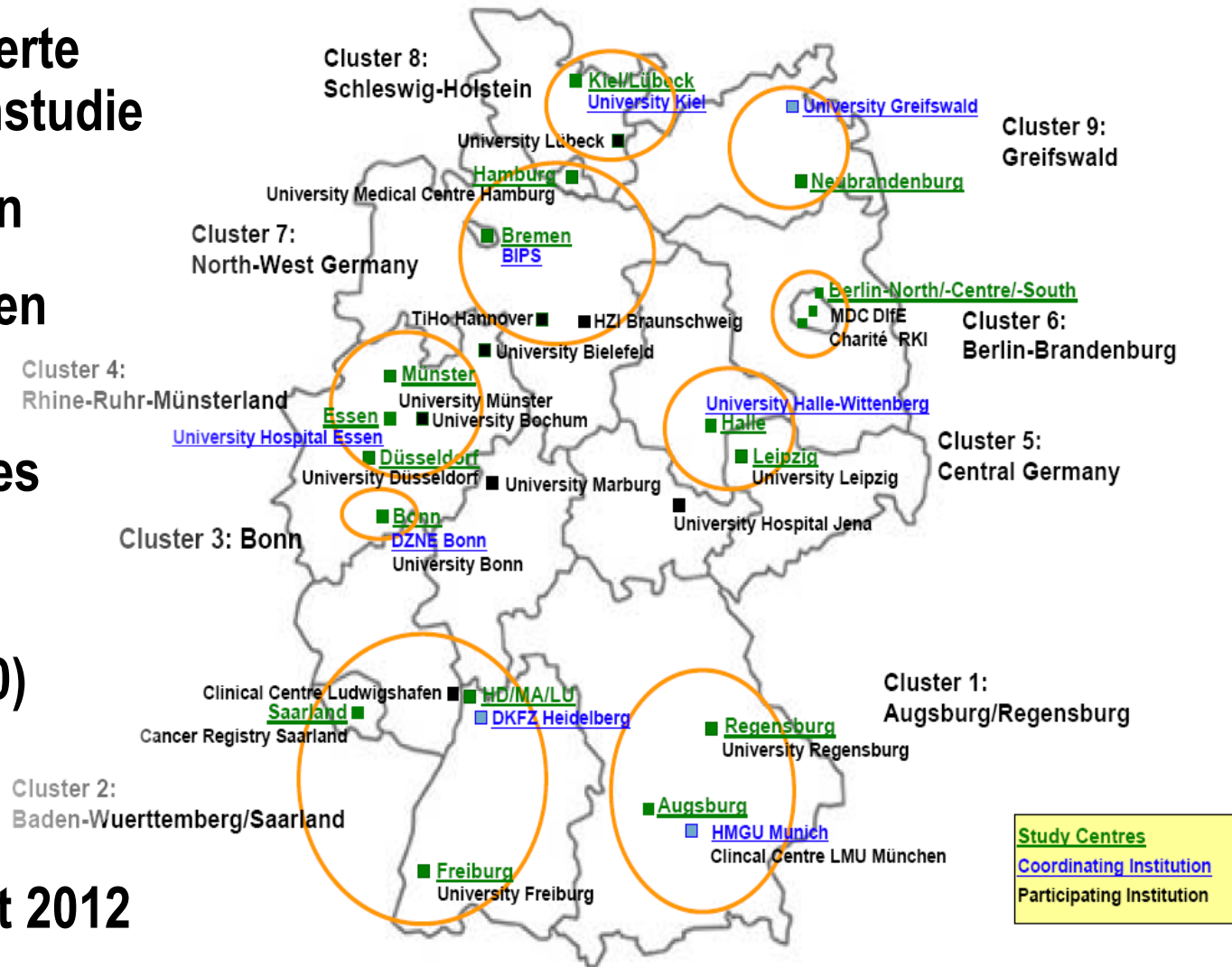
TMF-Workshop „Nachhaltige IT-Unterstützung von Kohorten und Registern“

Berlin, 15. 03. 2011

# Nationale Kohorte: Das Projekt

## Bevölkerungsbasierte Langzeit-Kohortenstudie

- 18 Studienzentren
- 200.000 Probanden
- Subkohorten:
  - Intensiviertes Programm (40.000)
  - MRT (40.000)
  - ggf. weitere
- Rekrutierung geplant ab Herbst 2012



# Nationale Kohorte: Studienprotokoll

---

- **Basisprotokoll (Level 1, 100%):**  
Interviews, Fragebögen, Untersuchungen, Bioproben, ...
- **Intensiviertes Protokoll (Level 2, 20%):**  
OGTT, Carotis-Sono (IMT), 24h-EKG, Schlaflabor, ...
- **Bildgebung mit MRT (20%):**  
Ganzkörper, Herz, Gehirn
- Eventuell weitere Sub-Protokolle / lokale Erweiterungen
- **Follow-up:**
  - nach einem Jahr: Kalibrierungsuntersuchung (3%)
  - alle 2-3 Jahre: Nachbefragung (100%)
  - nach 5 Jahren: Nachuntersuchung (100%)
- **Sekundärdatenabgleich**

# Nationale Kohorte: Zeitplan

|             |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |       |    |
|-------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-------|----|
| Kalenderj.  | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | ..... | 37 |
| Studienjahr | -3 | -2 | -1 | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | ..... | 25 |

**Planung, Machbarkeitsstudien  
Pilotstudie**

**Erste Erhebung / Untersuchung  
(N=200,000 / 40,000)**

**Zweite Erhebung / Untersuchung  
(N=200,000 / 40,000)**

**Aktive Nachbeobachtung, Registerbeobachtung,  
Sekundärdatenerfassung**

**Nutzung für epidemiologische Studien**

# IT-Unterstützung: Einsatzgebiete

---

- Probandenmanagement
- Datenerhebung
- Datenaufbereitung
- Datenspeicherung
- Datennutzung

# IT-Unterstützung: Einsatzgebiete

---

- **Probandenmanagement**
  - Stichprobenziehung, Probandenliste
  - Kontaktierung, Termin-Management
  - Pseudonymisierung
- Datenerhebung
- Datenaufbereitung
- Datenspeicherung
- Datennutzung

# IT-Unterstützung: Einsatzgebiete

---

- Probandenmanagement
- **Datenerhebung**
  - Ablauf- und Ergebnisdokumentation
  - Datenaustausch mit diagnostischen Geräten
- Datenaufbereitung
- Datenspeicherung
- Datennutzung

# IT-Unterstützung: Einsatzgebiete

---

- Probandenmanagement
- Datenerhebung
- **Datenaufbereitung:**
  - Zentrale Integration
  - Gewinnung abgeleiteter Daten  
(z. B. Befundung von Messwerten/Bildern)
  - Qualitätssicherung
- Datenspeicherung
- Datennutzung



# IT-Unterstützung: Einsatzgebiete

---

- Probandenmanagement
- Datenerhebung
- Datenaufbereitung
- **Datenspeicherung**
  - Verfügbarkeit, Vertraulichkeit, Integrität
  - Nachvollziehbarkeit
  - Langzeitspeicherung
- Datennutzung

# IT-Unterstützung: Einsatzgebiete

---

- Probandenmanagement
- Datenerhebung
- Datenaufbereitung
- Datenspeicherung
- **Datennutzung**
  - „Idee - Antrag - Datenübergabe - Auswertung - Ergebnisse“

# IT-Unterstützung: Übergeordnete Aspekte

---

- **Datenschutz**
  - Konzeptionelle Umsetzung
  - Schutz von Probanden- und Mitarbeiterdaten
- **IT-Sicherheit**
  - Verfügbarkeit, Integrität, Vertraulichkeit
  - Technische und organisatorische Maßnahmen
- **Qualitätssicherung**
  - Standardisierung, Prozessüberwachung, Historisierung, Versionsmanagement

# IT-Unterstützung: Übergeordnete Aspekte

---

- **Datenschutz**

- Konzeptionelle Umsetzung

- Trennung personenidentifizierender Daten von Forschungsdaten

- Pseudonymisierung

- Erfassung von Aufklärung und Einverständnis

- Schutz von Probanden- und Mitarbeiterdaten

- **IT-Sicherheit**

- Verfügbarkeit, Integrität, Vertraulichkeit

- Technische und organisatorische Maßnahmen

- **Qualitätssicherung**

- Standardisierung, Prozessüberwachung, Historisierung, Versionsmanagement

# IT-Unterstützung: Übergeordnete Aspekte

---

- **Datenschutz**
- **IT-Sicherheit**
  - Verfügbarkeit und Integrität der Systeme
    - Möglichst geringe Beeinträchtigung durch Ausfälle von Komponenten oder Personen
    - Untersuchungsbetrieb muss laufen
  - Verfügbarkeit und Integrität der Daten
    - Sofortige dauerhafte Speicherung und Sicherung
    - Historisierung bei nachträglichen Änderungen
  - Vertraulichkeit der Daten
    - Kein Zugriff durch Unbefugte
  - Technische und organisatorische Maßnahmen
- **Qualitätssicherung**

# IT-Unterstützung: Übergeordnete Aspekte

---

- **Datenschutz**
  - Konzeptionelle Umsetzung
  - Schutz von Probanden- und Mitarbeiterdaten
- **IT-Sicherheit**
  - Verfügbarkeit, Integrität, Vertraulichkeit
  - Technische und organisatorische Maßnahmen
- **Qualitätssicherung**
  - Standardisierung der Erhebungsinstrumente
  - Prüfung auf Vollständigkeit, Plausibilität, Konsistenz
  - Prozessautomatisierung und -überwachung, Historisierung, Versionsmanagement
  - Elektronische Schnittstellen für Gerätedaten
  - Scannen von Papierformularen

# IT-Unterstützung: Herausforderungen

---

- **Dezentrale Organisation**

*Viele Beteiligte mit unterschiedlichen Voraussetzungen und Interessen*

- **Lange Projektlaufzeit**

*Daten gewinnen im Lauf der Jahrzehnte an wissenschaftlichem Wert*

- **Dezentrale Erhebung - Zentrale Verarbeitung**

*Effiziente Integration und detaillierte Regelung der Nutzung erforderlich*

- **Dezentrale Erhebung – Qualitätssicherung**

*Vergleichbarkeit muss über viele Studienzentren und über den Zeitverlauf hinweg gesichert sein*

- **Ethik / Datenschutz**

*Ethische und rechtliche Anforderungen „erschweren“ IT-Unterstützung*

# Datenmanagement: Anforderungen

---

- IT-Unterstützung von Rekrutierung, Probandenmanagement, Terminverwaltung
- Effiziente und standardisierte Datenerfassung
- Sichere Speicherung und Archivierung der Daten
- Automatisierung von Prozeduren des Qualitätsmanagements
- Datenintegration und -bereitstellung zur Nutzung



# Datenmanagement: Lösungsansätze

---

- Standardisierung - Zentralisierung - Automatisierung
  - „Data Dictionary“: formale Beschreibung der zu erhebenden Daten
  - Automatisierte Plausibilitäts- und Konsistenzprüfungen so früh wie möglich
- Transparente Historisierung, Versionsmanagement
- Hohe Verfügbarkeit der zentralen Komponenten (Integrationszentren)
- Organisatorische Zentralisierung der Weiterverarbeitung (Kompetenzzentren und -panels)
- BSI- und TMF-Standards

# Danke für Ihre Aufmerksamkeit.

Für Nachfragen:



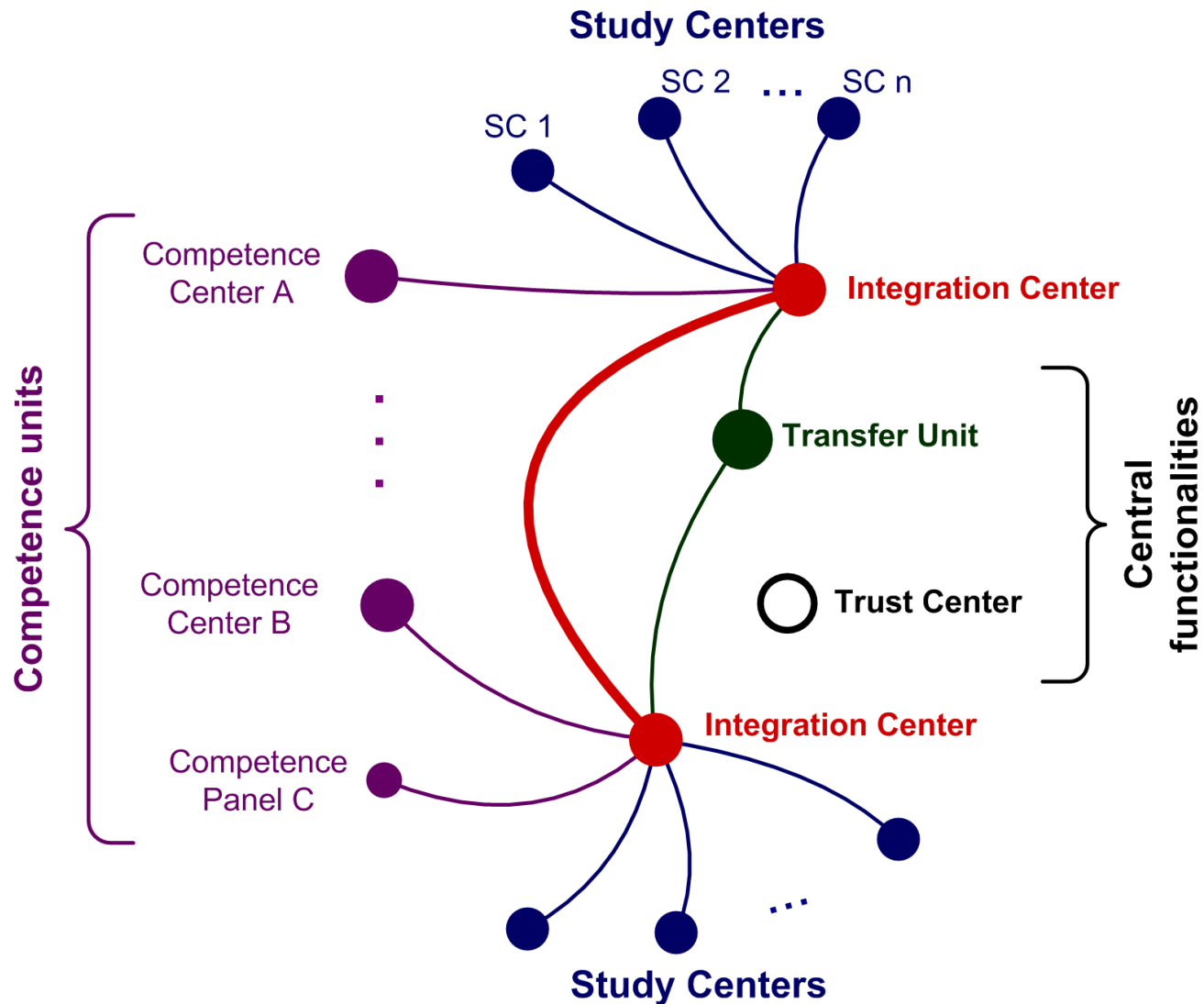
Dr. Daniel Kraft  
DKFZ Heidelberg, Abt. Krebs-Epidemiologie  
Datenmanagement für die Nationale Kohorte  
E-Mail: [d.kraft@dkfz-heidelberg.de](mailto:d.kraft@dkfz-heidelberg.de)  
Tel.: +49 6221 42 3153



Dipl.-Math. Stefan Ostrzinski  
Universitätsmedizin Greifswald  
ICM-VC - Zentrales Datenmanagement  
E-Mail: [stefan.ostrzinski@uni-greifswald.de](mailto:stefan.ostrzinski@uni-greifswald.de)  
Tel.: +49 3834 86 7545



# Organisationsmodell

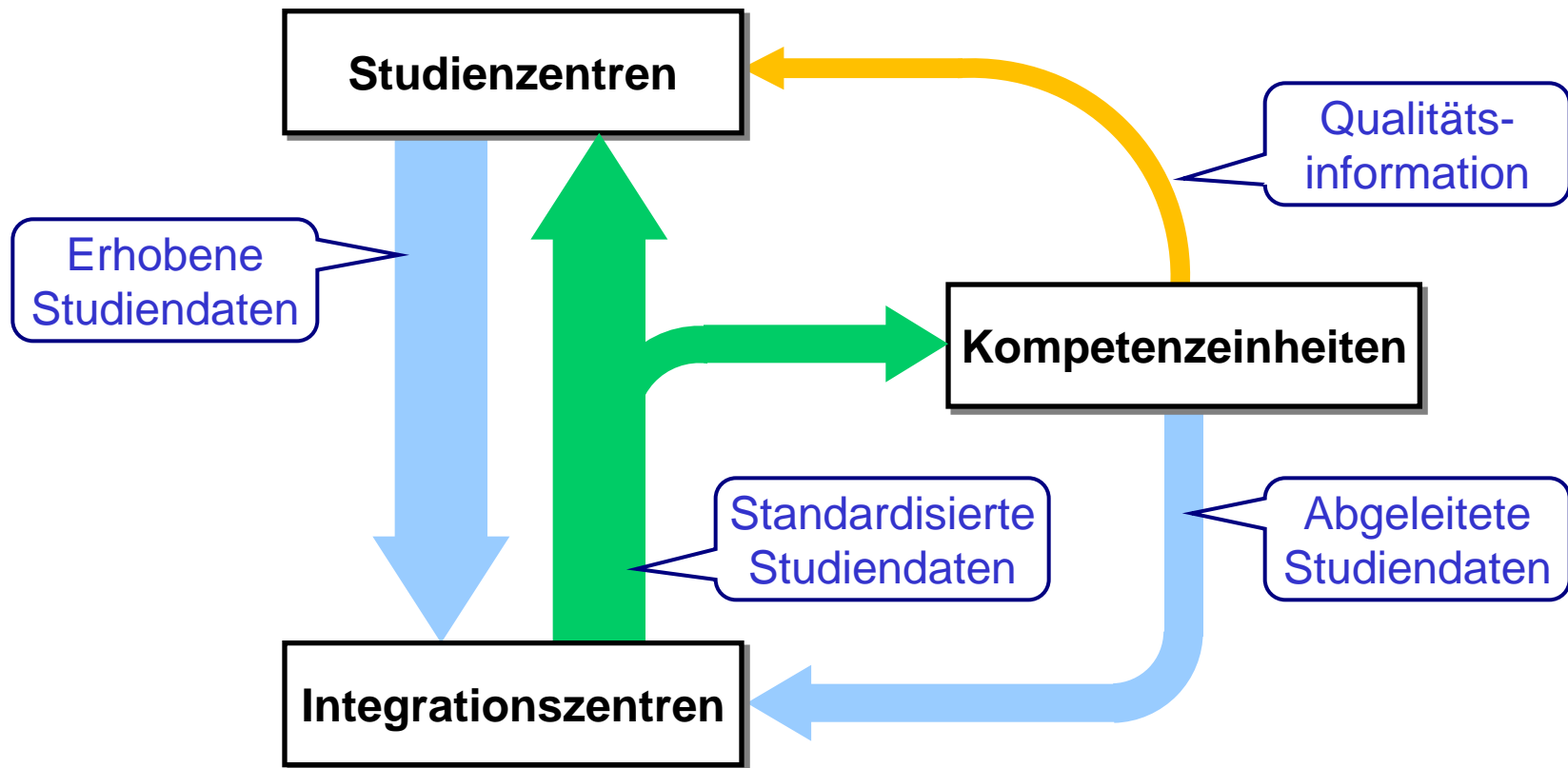


# Organisationsmodell

---

- **Studienzentren: Rekrutierung und Erhebung**
  - Rekrutierung und Verwaltung von Probanden
  - Erst- und Folgeuntersuchungen, Nachbeobachtung
  - Qualitätssicherung
- **Kompetenzzentren: Thematische Weiterverarbeitung**
  - Weiterverarbeitung komplexer Daten (MRT, EKG, ...)
  - Qualitätssicherung bestimmter Erhebungsinstrumente
  - Zentrale Erhebung und Abgleich mit externen Daten
- **Integrationszentren: Nationale Integration**
  - Integration aller Studiendaten aus Studien- und Kompetenzzentren
  - Qualitätssicherung auf Datenebene
  - Verfügbarmachung, Auswertungsunterstützung

# Organisationsmodell – Logische Datenflüsse



# Weitere organisatorische Einheiten

---

- **Kompetenzpanels:** Themenspezifische Qualitätskontrolle und Nachbeobachtung ohne eigene Infrastruktur
- **Treuhandstelle:**
  - Speicherung personenidentifizierender Daten
  - Erzeugung studienweit eindeutiger Pseudonyme
  - Unterstützung bei Abfrage von Sekundärdatenquellen
- **Transferstelle:** Verwaltung des Zugangs zu Studiendaten zu Verwertungszwecken

# Erhebung mittels elektronischer Formulare

---

- Identische, zentral gepflegte Erhebungsinstrumente für alle Studienzentren
- Sofortige Integration der Daten
  - Eingaben sofort überall sichtbar
  - Sofortige Konsistenzprüfung möglich
  - Zentrale Datensicherung
- Einfache Infrastruktur
  - Arbeitsstationen benötigen nur Web-Browser, damit kaum Wartungsbedarf
  - Änderungen an Software- und Datenmodell betreffen nur zentrale Server
- Bereichs- und rollenbasierte Zugriffsrechteverwaltung



# Aufgaben der Integrationszentren

---

- Integration und Qualitätssicherung:
  - Zentrale Speicherung aller Studiendaten mit Historisierung, Datensicherung und Spiegelung
  - Betrieb zentraler Erhebungswerkzeuge
  - Bereitstellung aller Daten für Studienzentren
  - Betrieb automatischer Qualitätssicherungsmaßnahmen, Berichterstellung
  - Bereitstellung von Möglichkeiten zur Anreicherung der Studiendaten durch Kompetenzeinheiten
  - Bereitstellung von Auswertungsmöglichkeiten (Verwaltung durch Transferstelle)
- Entwicklung von Software und Datenmodellen (Koord.)
- Systemadministration (Rechteverwaltung, Fernwartung)

# Kompetenzeinheiten

---

- Weiterverarbeitung, Interpretation und Qualitätskontrolle komplexer oder themenspezifischer Studiendaten
- Datenflüsse zu/von Kompetenzzentren:
  - Benötigte Daten liefern i.d.R. die Integrationszentren (Spezialfall: Direktübertragung von Studienzentren)
  - Erfassung von Ergebnissen wie in Studienzentren
- Vorgeschlagene Kompetenzzentren:
  - Bildgebung
  - Physische Aktivität und EKG
  - Sekundärdaten
  - Biobank
- Zusätzlich: Kompetenzpanels

# Datennutzung

---

- Nutzung von Studiendaten für wissenschaftliche Zwecke durch interne und externe Partner
- Datenzugang - Möglichkeiten:
  - Analyse auf gestellter Rechnerplattform, Zugriff auf projektspezifische Untermenge der Studiendaten
  - Export einer projektspezifische Untermenge nach zusätzlicher Pseudonymisierung
- Beteiligte Instanzen:
  - Datenfreigabe: Komitee für Datennutzung
  - Technische Datenbereitstellung: Integrationszentren
  - Administration: Transferstelle