



IT-Infrastruktur von Kohorten und Registern - aus Sicht der Nationalen Kohorte -

Stefan Ostrzinski, Daniel Kraft

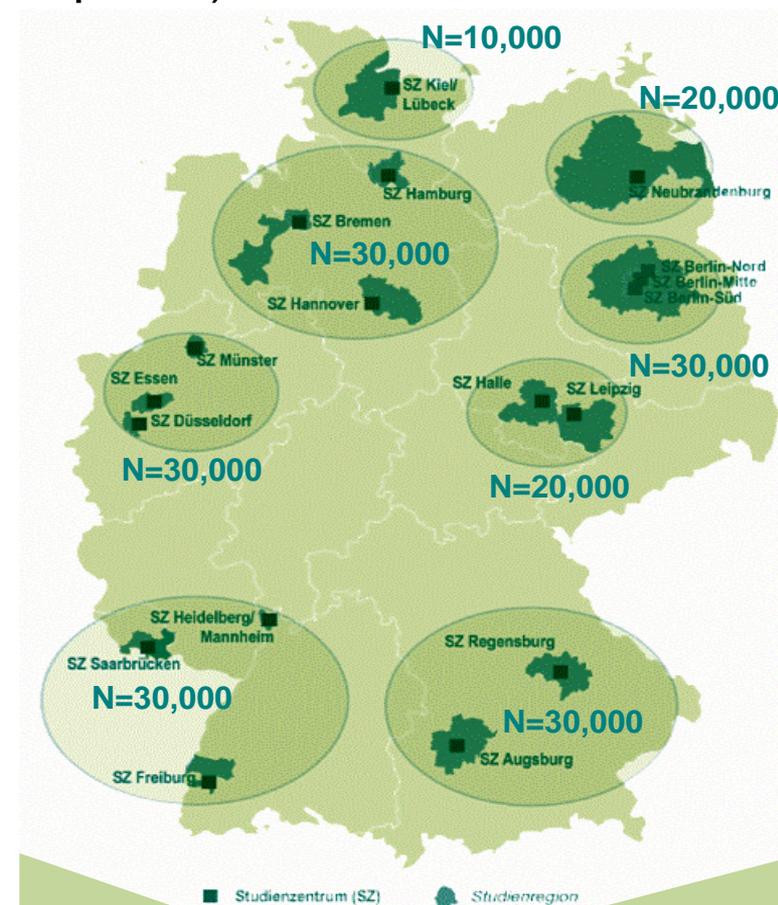
Workshop KoRegIT, Berlin, 08.04.2013

Nationale Kohorte

Studiendesign



- Bevölkerungsbasierte prospektive Kohortenstudie
- Altersbereich 20-69 Jahre zur Baseline-Untersuchung (m/w)
- Repräsentativer Ansatz (Melderegisterstichprobe)
- 8 regionale Cluster, 18+ Studienzentren
- Level 1 (Basisprogramm)
ca. 2,5 h, N = 200.000
- Level 2 (intensiviertes Programm)
ca. 4 h, N = 40.000
- Level 3 (zusätzliche Fragestellungen)
N variabel
- MRT-Programm, N≈30.000
- Baseline-Untersuchung 4 Jahre
Kalibrierung nach 1 Jahr
Nachbeobachtung alle 2-3 Jahre
Follow-Up-Untersuchung 4 Jahre



Nationale Kohorte

Zielsetzung



- Untersuchung der Entstehung der wichtigsten chronischen Krankheiten
 - Herz-/Kreislaufsystem
 - Lunge
 - Diabetes
 - Krebs
 - Neurodegenerative / -psychiatrische Erkrankungen
 - Infektionskrankheiten
- Infrastruktur für die epidemiologische Forschung

IT-Einsatzgebiete und Anforderungen



- IT-Unterstützung von Rekrutierung, Probandenmanagement, Terminverwaltung
- *Effiziente und standardisierte* Datenerfassung
- *Sichere langfristige* Speicherung und Archivierung der Daten, *schnelle Verfügbarkeit*
- *Automatisierung* von Prozeduren des Qualitätsmanagements
- Datenintegration und -bereitstellung

Übergeordnete Aspekte



- **Datenschutz**

- Konzeptionelle Umsetzung
- Schutz von Probanden- und Mitarbeiterdaten

- **IT-Sicherheit**

- Verfügbarkeit, Integrität, Vertraulichkeit
- Technische und organisatorische Maßnahmen

- **Qualitätssicherung**

- Standardisierung, Prozessüberwachung, Historisierung, Versionsmanagement



- **Dezentrale Organisation**
Viele Beteiligte mit unterschiedlichen Voraussetzungen und Interessen
- **Lange Projektlaufzeit**
Daten gewinnen im Lauf der Jahrzehnte an wissenschaftlichem Wert (Follow-ups, Sekundär- und Registerdaten, ...)
- **Zusammenführung dezentraler Daten**
Effiziente Integration und detaillierte Regelung der Nutzung erforderlich
- **Dezentrale Erhebung – Qualitätssicherung**
Vergleichbarkeit über alle Studienzentren und über den Zeitverlauf
- **Ethik / Datenschutz**
Ethische und rechtliche Anforderungen „erschweren“ IT-Unterstützung

Anmerkungen aus NatKo-Sicht



- Generischer Anforderungskatalog
 - Aufgaben und Anforderungen in Entwicklung, Betrieb und Abschluss
 - Umfassend strukturierte Beschreibung der Use Cases
 - IT-Lastigkeit der Use Cases sehr unterschiedlich
 - Use Cases vs. Anforderungen vs. Spezifikation vs. Lösungen
- Evtl. noch nicht ausreichend abgedeckt:
 - Schnittstellen / Einbindung externer Datenquellen / “komplexe Daten”
 - Management personenbezogener Daten
 - Bioproben: Gewinnung, Verwaltung, Labore / Biorepository
- Datenschutz vs. IT-Sicherheit
 - Datenschutz: Gesetzliche Verpflichtung
 - IT-Sicherheit: “Eigeninteresse” und Anforderungen aus Datenschutzgesetzen

Anmerkungen aus NatKo-Sicht



- Softwareentwicklung:
 - UC-Liste Grundlage für SW-Entwicklung, ggf. nicht ausreichend für Sonderlösungen
 - Standardsoftware vs. Eigenentwicklung
 - zentrale vs. dezentrale Lösungen
 - Modularisierung
 - Wartungsaufwand (bewusst) nicht berücksichtigt
- Generelle Anforderungen an (Meta-)Datenstrukturen
- Prozessunterstützung:
Entwurf -> Durchführung -> Kontrolle -> Ergebnisse
- Umfassendes IT-Konzept notwendig
-> Begriff “Datenmanagement”

Anmerkungen zu Use Cases



- TL006/UC0195 Rekrutierungskonzept
 - Zusammenhänge mit wiss. Fragestellung/Protokoll
 - Repräsentativität (“Randgruppen”)
- TL006 Einwilligung
 - Modularisierung
 - Inhalte:
 - Bereitschaft zur Teilnahme
 - Datenschutzrechtliche Einwilligung
 - Bioproben: Eigentumsrechtliche Übertragung
 - Sekundärdaten etc.: Ermächtigung zur Datenabfrage

Anmerkungen zu Use Cases



- TL041/UC0204 Versicherungspflicht prüfen:
 - Abschluss von Pflichtversicherungen ersetzt keine Risikoanalyse
- TL009 Analyseplan
 - Für prospektive Kohortenstudien:
Nutzungsregeln (“Use and Access Policy”)
- TL010/UC0034 Variablen ... festlegen
 - Ablaufdokumentation und “Sprungregeln” nicht vergessen
- TL043 Schulung des Zentrumspersonals
vs. TL018/UC0066 Personal schulen
 - Zusammenhang?
 - Schulung/Zertifizierung sind auch laufende Aufgaben im Betrieb,
nicht nur Vorbereitung!

Anmerkungen zu Use Cases



- TL020/UC0080 Informationsmaterial an potentielle Teilnehmer versenden
 - Ggf. auch mehr entsprechend Rekrutierungsstrategie
- TL022/UC0100 Widerruf Einwilligung bearbeiten
 - Änderung statt Widerruf (modulare Einwilligung)
- TL023/UC0104 Zentrales Monitoring
 - Automatisierung von Prozessen
 - Bestimmung und Prüfung von Qualitätsindikatoren
 - “Beschleunigung” des Monitorings

Anmerkungen zu Use Cases



- TL026 Datenmanagement
 - “Organisation und Pflege der Daten”
-> Prozesse, Datenflüsse
- TL026/UC0122 Datensicherheit entsprechend Datenschutzkonzept umsetzen
 - Datensicherheit vs. Datenschutz
 - Eigener Punkt ID-Management
 - Verarbeitung identifizierender Daten
 - Pseudonymisierung, Anonymisierung
 - Einwilligungsmanagement
- TL033 Datenintegration, -zusammenführung
 - Nicht von TL026 Datenmanagement trennen

Datenmanagement in der Nationalen Kohorte

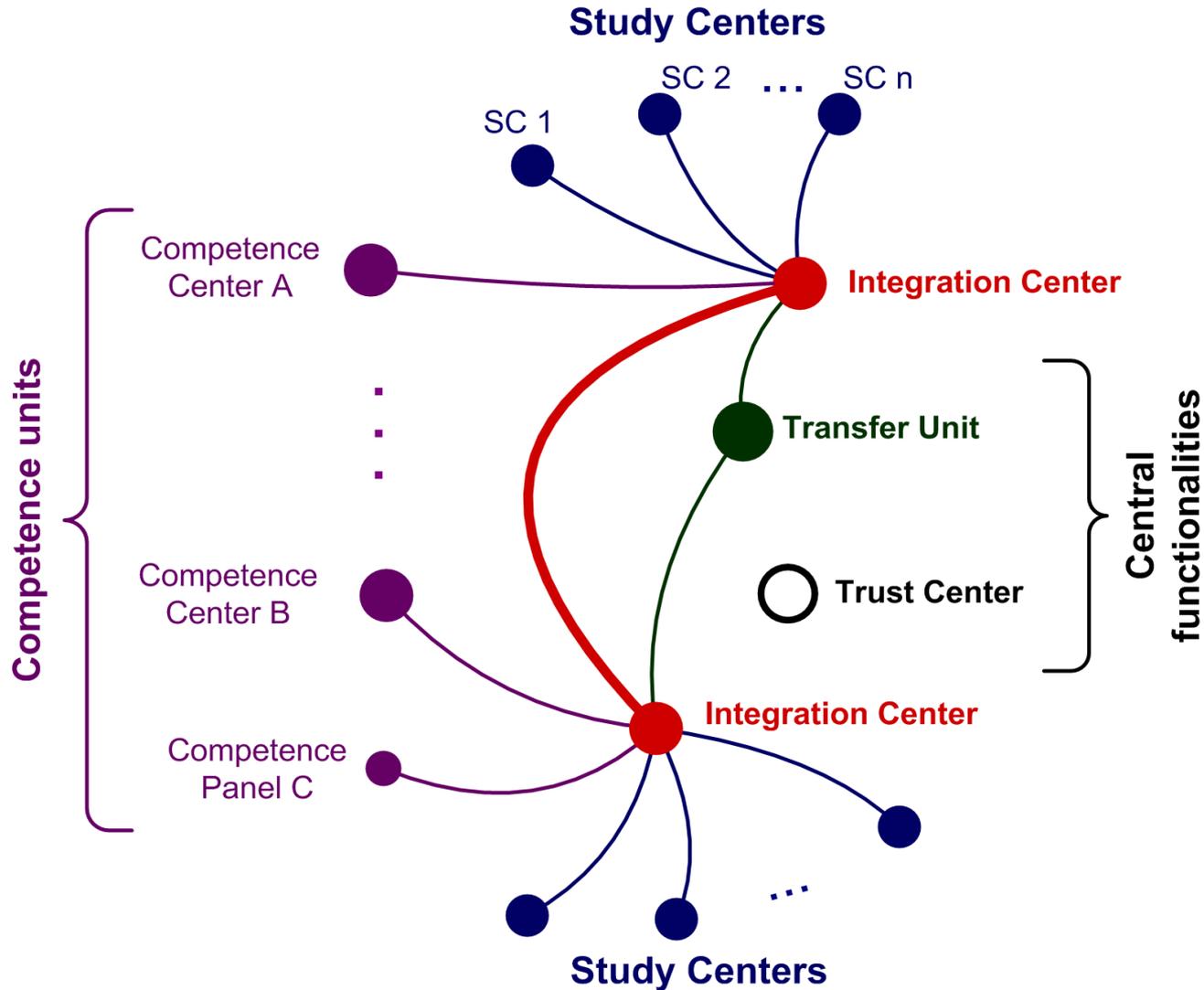
Lösungswege



- **Standardisierung - Zentralisierung - Automatisierung**
 - Zentrales „Data Dictionary“: formale Beschreibung zu erhebender Daten
 - Datenaufbereitung, Qualitätssicherung, Sekundärdatenerhebung
 - Frühzeitige automatisierte Plausibilitäts- und Konsistenzprüfungen
 - Datenweitergabe für wissenschaftliche Auswertung
- **Hohe Verfügbarkeit zentraler Komponenten**
 - Organisation / Technik / Personal
- **BSI- und TMF-Standards**

Datenmanagement in der Nationalen Kohorte

Organisationsmodell



Data Dictionary als zentrales Element



- Definition / Design der Studiendaten
- Erstellung der Erhebungswerkzeuge
 - Interview, Untersuchungen, Selbstausfüller-Fragebögen
 - Dokumentation Durchführung und Ergebnisse, Nachdokumentation
- Datenerhebung
 - Plausibilitäts-, Konsistenz- und Vollständigkeitsprüfungen
 - Warnhinweise, Fehlermeldungen
- Datenbereinigung, prozessbegleitende Qualitätssicherung
- Datenweitergabe für wissenschaftliche Auswertung
 - Beantragung, Übergabe und Rückführung
- Wissenschaftliche Analysen/Datenauswertung

Nationale Kohorte: Datenerhebung

Erhebungsinstrumente



Abmelden Benutzer: fitzerk@Neubrandenburg UntZ Proband: 2810009992@Pretest 2

Passwort ändern

Startseite

- Intro
- Blutdruck ^{SOP}
- Anthropometrie ^{SOP}
- Handkraft ^{SOP}
- Start Interview ^{SOP}
- SES
- MAN_HKE
- MAN_KRE
- MAN_SWE
- MAN_SDE
- MAN_RME
- MAN_LUN
- MAN_ALL
- MAN_MDE
- MAN_GLF
- MAN_HTE
- MAN_NBE
- MAN_NME**
- MAN_AOE
- MAN_INF
- MAN_UNF
- MAN_OP
- Frauen
- Ende
- Blutentnahme ^{SOP}
- Blutverarbeitung Basis ^{SOP}
- NeuroKog Config
- Zufriedenheit
- Abschlussgespräch
- Geruchstestung ^{SOP}
- Touchscreen Intro ^{SOP}
- Knochenbrüche
- Männer
- Touchscreen: NeuroKog

MAN_NME MNME_AN_G1

Neurologische und psychische Erkrankungen

Wie würden Sie Ihr Gedächtnis zurzeit beurteilen, würden Sie sagen, es ist...

Weiter Zurück Zur nächsten

Bitte auswählen...
Bitte auswählen...
Sehr gut
Gut
Mittelmäßig
Schlecht
Sehr schlecht
[Weiß nicht]
[Keine Angabe]

Notiz anlegen Sprung

- Dynamische Formularerstellung (Datenbank)
- Zentrale Studiendatenbank

NatCoEdc r3153 · Nutzerhandbuch · Zentrales Datenmanagement

- Webbasierte Erhebungsanwendung
- Touchscreen für Selbstausfüller-Fragebögen, Tests



- Aufklärung / Einwilligung
- Aktualisierung identifizierender Daten
- Ablaufdokumentation



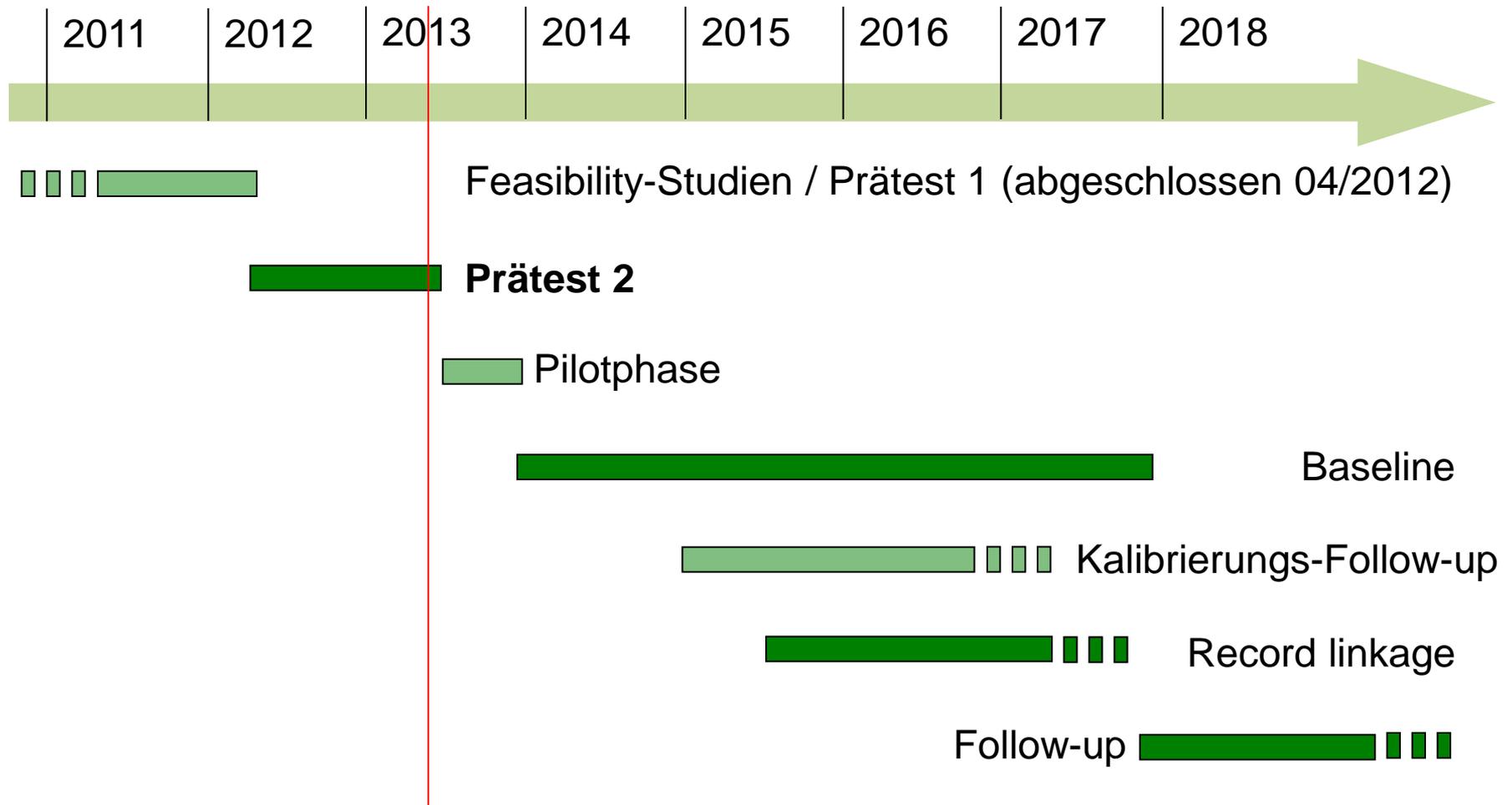
- Labor- und Gerätedaten
 - Ca. 10 diagnostische Geräte
 - Externe Web-Services und SW-Module für die Erhebung
 - Labordaten
- Biorepository (LIMS)
- MRT-Substudie (Ablauf, Bilddaten, Reading)
- Sekundär- und Registerdaten
 - Diagnose- und Abrechnungsdaten ambulante Versorgung
 - Verordnete Arzneimittel
 - Berufshistorie: Versichertengruppe, Reha-Maßnahmen, Arbeitslosigkeit, Kindererziehungszeiten, Rentenstatus
 - Aktuelle und frühere Krebserkrankungen
 - Tätigkeitsmerkmale und Betriebscharakteristika sowie Arbeitslosigkeit in der Berufshistorie

Regelungen für den Zugang zu Daten



- Daten verfügbar voraussichtlich erst ab 2017
aber:
- Art und Weise sowie Verfahren für den Zugang zu Daten und Bioproben *frühzeitig, vor Beginn der Hauptstudie* vereinbaren!
- Datenschutz: Re-Identifikationsrisiko minimieren
- Viele häufig wiederkehrende, streng geregelte Arbeitsschritte \Rightarrow Umsetzung in IT-Verfahren
- Nutzungsordnung (27.02.2013),
Konzept Transferstelle

Nationale Kohorte Zeitplan





- Projekt in der DFG-Ausschreibung „Informationsinfrastrukturen für Forschungsdaten“
- "Modular systematic approach to implement a Central Data Management"
- Laufzeit: 3 Jahre, Projektstart: Nov. 2012
- Zentrales Datenmanagement, Fokus auf Kohortenstudien
- Ziele:
 - Prozesse, Abläufe und Techniken generalisieren
 - Aufbau eines modularen Unterstützungssystems
 - Anleitungen, Werkzeuge, Musterlösungen begleiten sämtliche Phasen des Datenlebenszyklus

Danke für Ihre Aufmerksamkeit.



Nationale Kohorte – Zentrales Datenmanagement



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



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