

*Gemeinsam forschen für eine gesündere Zukunft*



# **IT-Infrastruktur der Nationalen Kohorte**

TMF-Registertage, Berlin, 22.-23.05.2014

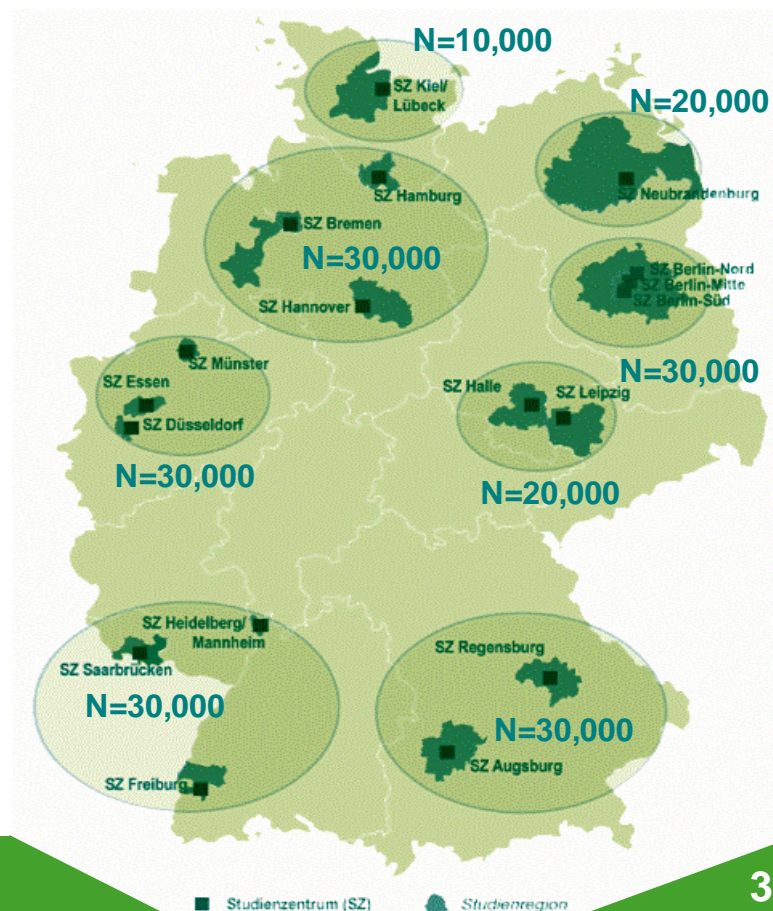
S. Ostrzinski, D. Kraft

# Überblick

- Einleitung
  - Wer ist / was macht die Nationale Kohorte?
  - IT in der Nationalen Kohorte
- Datenschutz
  - Anforderungen und Umsetzung
  - Management identifizierender Daten
- Datenerhebung Studiendaten
  - Data Dictionary als zentrales und verbindendes Element
  - Verschiedene Erhebungsmethoden / -instrumente
- Datennutzung

# Nationale Kohorte: kein Register

- Bevölkerungsbasierte prospektive Kohortenstudie
- Altersbereich 20-69 Jahre zur Baseline-Untersuchung (m/w)
- Repräsentativer Ansatz (Melderegisterstichprobe)
- 8 regionale Cluster, 18+ Studienzentren
- N = 200.000, Baseline 2014-2018
- Sekundär- und Registerdaten
- Nachbeobachtung, Follow-Up
- Ursachen / Risikofaktoren
- Forschungsinfrastruktur

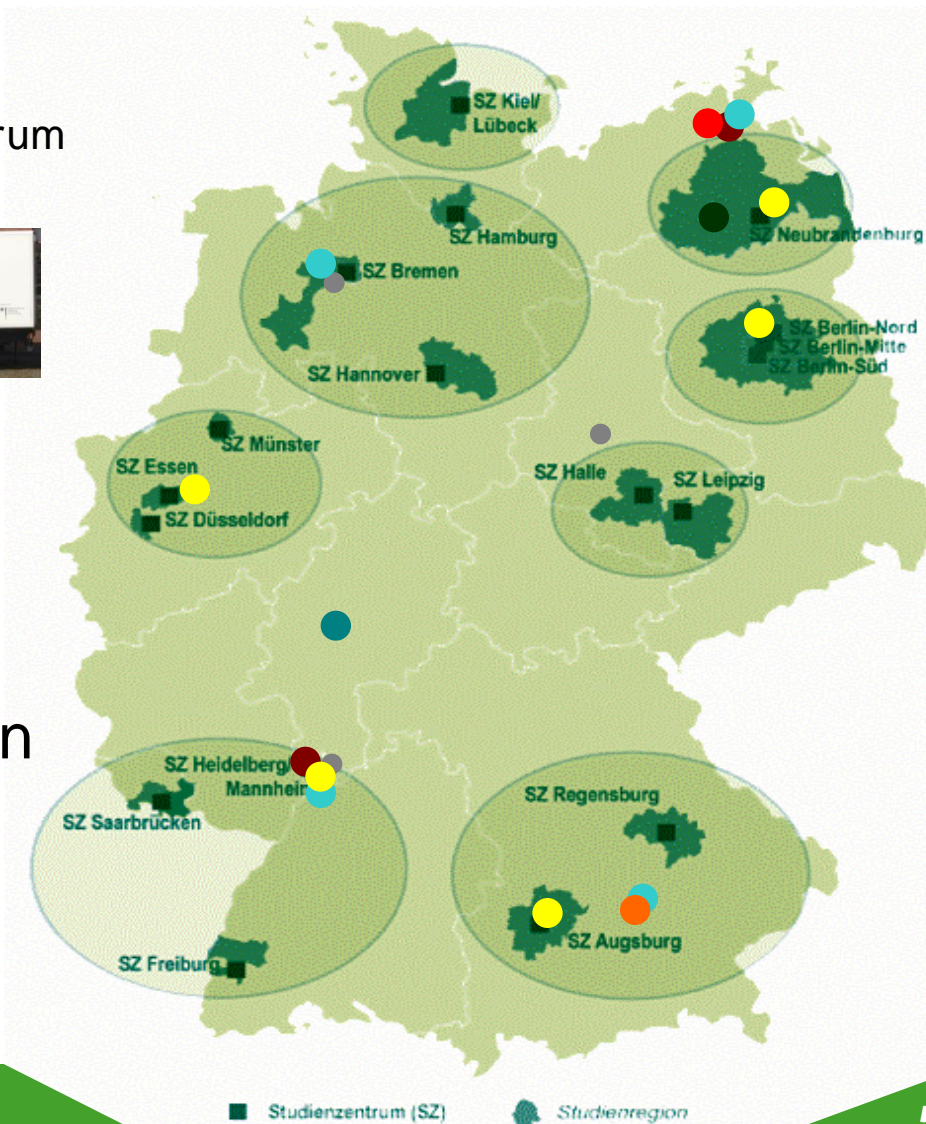


# Untersuchungsprogramm (Baseline)

- Level 1 (ca. 2,5 h, N=200.000)
  - Interview, Selbstausfüller-FB, Neurokognitive Tests
  - Anthropometrie, BIA, Blutdruck, Puls, Vascular Explorer
  - Handgreifkraft, Akzelerometrie, Spirometrie, Zahnstatus
  - Laborbasisprogramm, Biobank
- Level 2 (ca. 4 h, N=40.000)
  - Ultraschall Bauchfett, AGE-Reader, 3D-Echokardiographie
  - Ruhe- und Langzeit-EKG, Schlafcharakteristika
  - NO in Ausatemluft, Sinnesorgane
- Weitere Untersuchungen (N = 30.000 ... 70.000)
  - Ergometer, Muskel-Skelett-System, Zahnuntersuchung
  - Oraler Glukosetoleranztest (oGTT)
  - Ganzkörper-MRT (ca. 90 TB Rohdaten)
- Level 3 (lokal, zusätzlich, N variabel)

# Standorte

- 18 Studienzentren  
Probandenmanagement / Untersuchungszentrum
- Temporäre Zentren
- Mobiles Zentrum
- Zentrales Biorepository
- 5 MRT-Zentren
- MRT Imaging Core
- Kompetenznetz Sekundärdaten
- Mortalitäts-Follow-up
- 2 Integrationszentren
- Unabhängige Treuhandstelle



# IT-Aufgaben in der Nationalen Kohorte

- IT-Unterstützung von Rekrutierung, Probandenmanagement, Terminverwaltung
- *Effiziente und standardisierte* Datenerfassung, Datenintegration (thematisch, örtlich, zeitlich)
- *Sichere, langfristige* Speicherung und Archivierung der Daten, *schnelle Verfügbarkeit*
- *Automatisierung* von Prozessen des Qualitätsmanagements
- Datenbereitstellung für Forschungsprojekte
- Technologischer Datenschutz, IT-Sicherheit

# IT-Einsatz: Herausforderungen

- **Dezentrale Organisation**  
Viele Beteiligte mit unterschiedlichen Voraussetzungen und Interessen  
Vergleichbarkeit über alle Studienzentren und über den Zeitverlauf
  - **Heterogenität der Daten / Erhebungsmethoden**  
Interview, SAQ, Untersuchungen, Gerätedaten, Proben, Sekundärdaten
  - **Lange Projektlaufzeit**  
Daten gewinnen im Lauf der Jahrzehnte an wissenschaftlichem Wert
  - **Ethik / Datenschutz**  
Einwilligung, Trennung IDAT – MDAT, Pseudonymisierung, DV im Auftrag
- ⇒ **Standardisierung, Zentralisierung, Automation**

# Besondere Herausforderung Datenschutz

- Gesetzliche Anforderungen:
  - Verantwortliche Stelle
  - Genaue Zweckbestimmung
  - Datensparsamkeit
  - Benennung der Empfänger
- Nationale Kohorte:
  - 18+ Studienzentren
  - Offene Zweckbestimmung
  - Breite Datensammlung
  - Empfänger stehen nicht fest
  - Biobank, Sekundärdaten, ...
- Nationale Kohorte e.V., Auftragsdatenverarbeitung
- Technische und organisatorische Maßnahmen
- Fortlaufende Information der Teilnehmer / Öffentlichkeit
- Modulare Einwilligung, verschiedene Widerrufsoptionen



# Management identifizierender Daten

- lokal*
  - Probandenmanagement-Einheiten der Studienzentren
    - Stichprobenziehung, Kontaktierung, Terminverwaltung
    - Getrennt von Untersuchungszentren
    - Kein Zugang zu Studiendaten
    - Software: MODYS (BIPS Bremen), VPN zur Treuhandstelle
- zentral*
  - ID- und Einwilligungsmanagement in der Treuhandstelle
    - Unabhängig von anderen Einheiten des Datenmanagements
    - Master Participant Index (Software: E-PIX)
    - Pseudonymisierungsdienst (Software: gPAS)
    - Consent Management (Software: gICS)
    - Datenaustausch mit anderen Institutionen



- Ziel

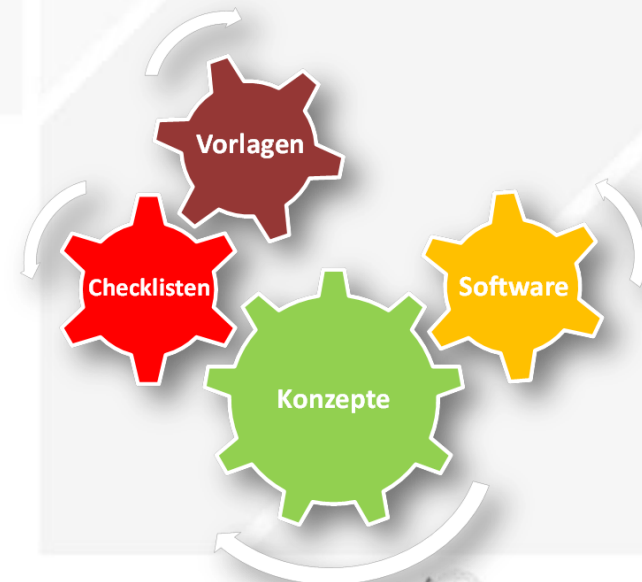
Realisierung des *zentralen Datenmanagements* für Kohorten und Register durch direkt nutzbare Werkzeuge und Lösungsstrategien *deutlich vereinfachen*

- Zielgruppe

Epidemiologische Forscher in „*kleinen*“ Kohorten und Registern, die mit minimaler IT-Unterstützung organisatorische und datenschutzrechtliche Herausforderungen meistern müssen

- Bereitstellung kostenfreier Werkzeuge

- plan.Tau – ein interaktives Referenzportal als zentrale Wissensdatenbank und Plattform für Erfahrungsaustausch
- Musterdokument zur Erstellung von Datenschutzkonzepten
- separate Software-Tools



# Unterstützung für Kohorten und Register



MOSAIC bietet direkte Unterstützung bei

- Einsatz einer flexiblen Speicherlösung
- Abbildung von Studienitems
- formularbasierte Datenerfassung
- Integration externer (strukturierter) Datenquellen
- Maßnahmen zur Sicherung des Forschungsdatenbestands
- Erstellung von QS-Reports
- Realisierung eines „Use and Access“-Verfahrens

Kontakt:

[mosaic@uni-greifswald.de](mailto:mosaic@uni-greifswald.de)

<http://mosaic-greifswald.de>

# Erhebung der Studiendaten

- Ziele:
  - Datenvalidierung (Plausibilität, Konsistenz, Vollständigkeit)
  - unmittelbare Speicherung, Zusammenführung
  - höchstmögliche Standardisierung der Erhebungsverfahren
  - Automatisierung, Vermeidung von Fehlerquellen ("Abschreiben")
- Randbedingungen:
  - Datenschutz / Ethik: Einwilligungsstatus, Pseudonymisierung
  - Heterogenität der Quellen (Interview, Geräte, Proben, ...)
  - Plattform für QS-Prozesse und Datennutzung
  - Effizienz: minimale Anforderungen an dezentrale IT-Infrastruktur

# Data Dictionary als zentrales Element

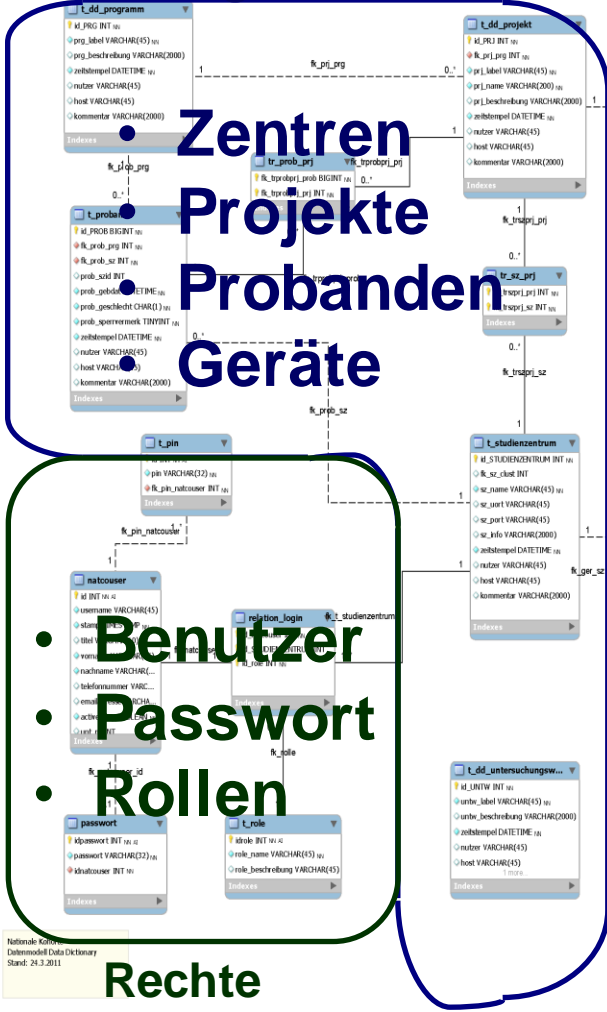
- Definition / Design der Studiendaten
- Erstellung der Erhebungswerkzeuge
  - Interview, Untersuchungen, Selbstausfüller-Fragebögen
  - Dokumentation Durchführung und Ergebnisse, Nachdokumentation
- Datenerhebung
- Datenzusammenführung und –integration
- Datenbereinigung, Qualitätssicherung
  - Plausibilitäts-,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

# Data Dictionary: Allgemeine Anforderungen

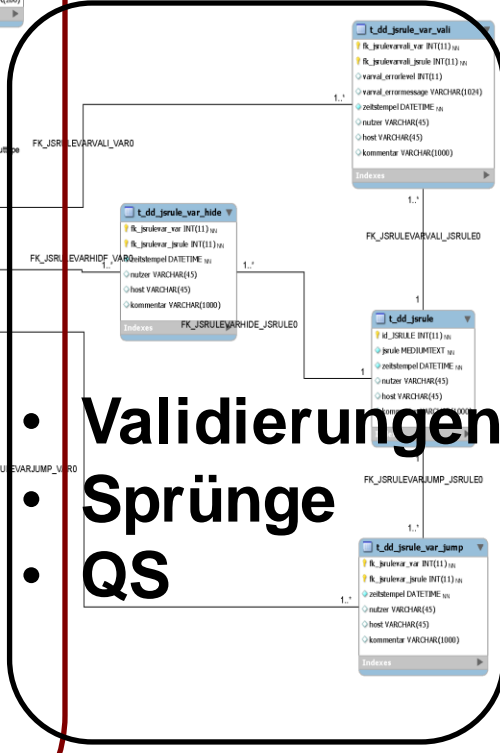
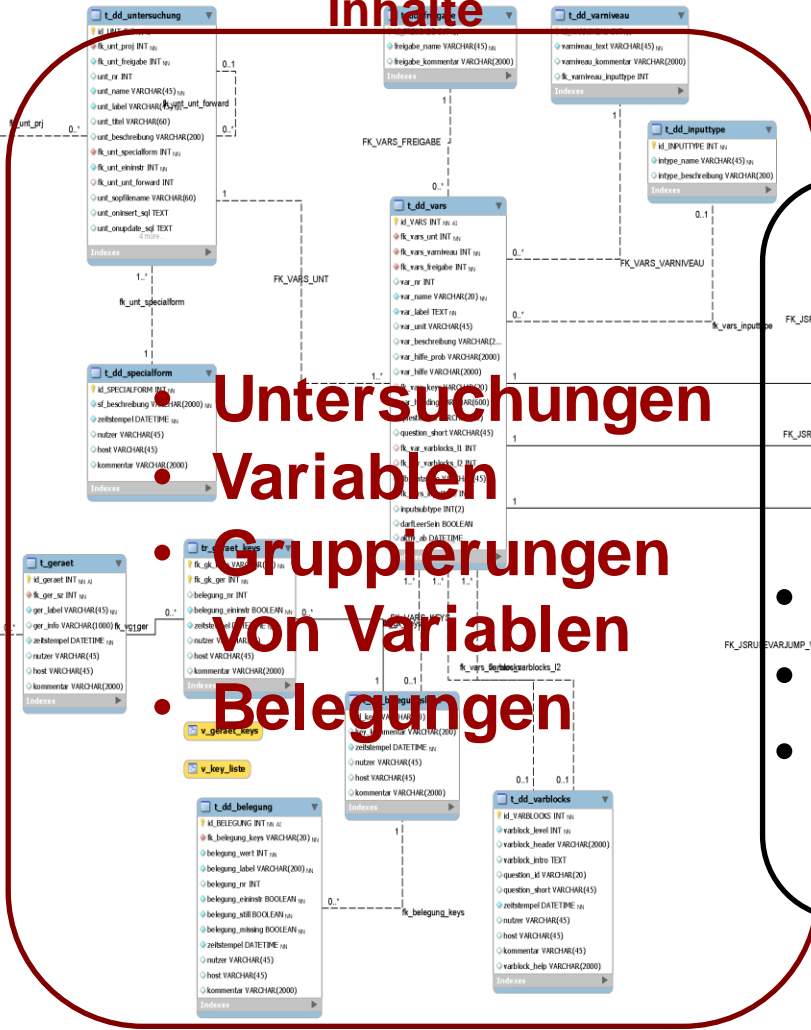
- Multizentrisch, Quer- und Längsschnitt
- Vollständigkeit
  - Alle benötigten Informationen abgebildet
- „Normalisierung“
  - Abhängigkeiten in der Struktur, nicht in den Daten (keine impliziten Constraints)
- Versionierung
  - Zuordnung zwischen gespeicherten Studiendaten und zugehörigem Versionsstand des Data Dictionary
- Nutzung von „Vorarbeiten“
  - TMF MDR, SHIP, SFB-TRR19, Transferstelle FVCM, GANI\_MED, ...

# Data Dictionary: Datenmodell

## Organisation



## Inhalte



## Regeln

Nationale Kohorte  
 Datenmodell Data Dictionary  
 Stand: 24.3.2011

# Datenerhebung: Webanwendung

Abmelden Benutzer: osi@Schulung UntZ Proband: 2030000014@Pretest-2

Passwort ändern  
Startseite

Intro  
NeuroKog Config  
Kontaktmöglichkeiten  
Blutdruck  
Anthropometrie  
Handkraft  
Start Interview

**SES**

MAN\_HKE  
MAN\_KRE  
MAN\_SWE  
MAN\_SDE  
MAN\_RME  
MAN\_LUN  
MAN\_ALL  
MAN\_MDE  
MAN\_GLE  
MAN\_HTE  
MAN\_NBE  
MAN\_NME  
MAN\_ADE  
MAN\_INF  
MAN\_UNF  
MAN\_OP  
Frauen  
Ende

Blutentnahme  
Blutverarbeitung Basis  
Laborwerte  
Zeiterfassung TS-Core  
Touchscreen-Intro  
Tabak  
Alkohol  
Drogen  
Männer  
Knochenbrüche  
Familienanamnese  
Frühkindliche Faktoren  
Gewichtsgeschichte  
Phänotyp  
Screening  
Gesundh bez. Lebensqual.  
WHOQOL BREF

Neurokognitive Tests  
Neuropsychologische Tests  
FeNO-Messung  
Zufriedenheit  
Abschlussgespräch  
Spirometrie  
Zahnstatus

**SES**

**Bildung/Ausbildung**

Im nächsten Fragenkomplex werden wir uns mit Ihrer Schul- und Ausbildung beschäftigen.

**Welchen höchsten allgemein bildenden Schulabschluss haben Sie?  
Sagen Sie es mir bitte anhand dieser Liste**

**INT: Liste 2a vorlegen**  
EOS: Erweiterte Oberschule  
Wenn Schulabschluss im Ausland erworben, bitte wenn möglich, ein bereits in Deutschland erworbenes, äquivalentes  
Schulabschluss" angeben.  
Formulierung der Antwortmöglichkeiten an das Geschlecht des Probanden anpassen.  
Erläuterung der Antwortoption laut Liste

Bitte auswählen...

Schülerin, besuche eine allgemein bildende Vollzeitschule  
Von der Schulzeit bis zum Ende der Schulzeit kein Schulabschluss  
Hauptschulabschluss (Volksschulabschluss)  
Realschulabschluss (Mittlere Reife)  
Polytechnische Oberschule der DDR mit Abschluss der 8. oder 9. Klasse  
Polytechnische Oberschule der DDR ohne Abschluss  
Fachhochschulreife, Abschluss einer Fachoberschule  
Allgemein- oder fachgebundene Hochschulreife/Abitur

auch EOS mit Lehre)

Weiter Zurück Zur näch

- Dynamische Formularerstellung (Datenbank)

- Zentrale Studiendatenbank

- Stets aktuelle Statusanzeige

- Touchscreen für Tests und Selbstausfüller-Fragebögen

- Gerätedatentransfer

- Aufklärung / Einwilligung
- Aktualisierung identifizierender Daten
- Ablaufdokumentation



# Datenerhebung: Touchscreen

- „Unbeobachtete“ Nutzung durch Probanden
  - Nutzerführung (Anmeldung, Anleitung, Validierung der Eingaben)
  - Sicherheitsaspekte
- Eingliederung in bestehende Struktur
  - Zugriff auf bereits erhobene Daten
  - Wiederverwendung vorhandener Anwendungen
- Hardware-Bedienbarkeit
  - Auch mit Prothesen oder Handschuhen
  - Mögliche Einschränkungen der Beweglichkeit oder Sehkraft des Probanden
  - Robustheit (mechanisch, chemisch), wartungsarm



# Datenerhebung: Diagnostische Geräte

## Standardisierte, gesicherte Datenübertragung und Datenintegration

Ausgabedateien  
enthalten Teilnehmer-  
und Geräteerkennung

Server in Integrationszentren:

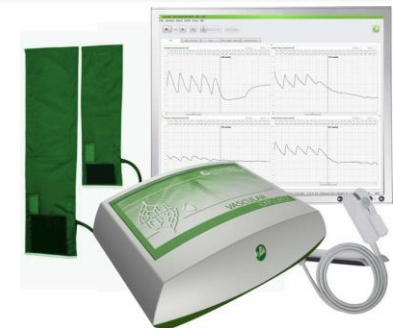
- ① Initialisierungsdatenerzeugung
- ⑦ Integration in Studiendatenbank
- ⑧ Statusanzeige in Web-Anwendung
- ⑨ Archivierung

Gerät

- ③ initialisiert sich aus Datei
- ④ speichert Daten lokal

Lokale Software (Java-basiert)

- ② legt Initialisierungsdaten lokal ab
- ⑤ überwacht Ausgabe-Verzeichnis
- ⑥ lädt neue Dateien gesichert hoch



# Datenerhebung: weitere Schnittstellen

- Labordaten
  - Bioprobengewinnung:  
Blut, Urin, Speichel, Nasenabstriche, Stuhl
  - Laborbasisprogramm („Präsenzlabor“)
  - Biorepository (LIMS)
- MRT-Substudie
  - Entkopplung Ablauf - Bilddatenverarbeitung – Reading
- Sekundär- und Registerdaten
- Qualitätssicherung
  - Validierung, Prozesssteuerung
  - spezielle QS-Sichten und automatisierte Reports

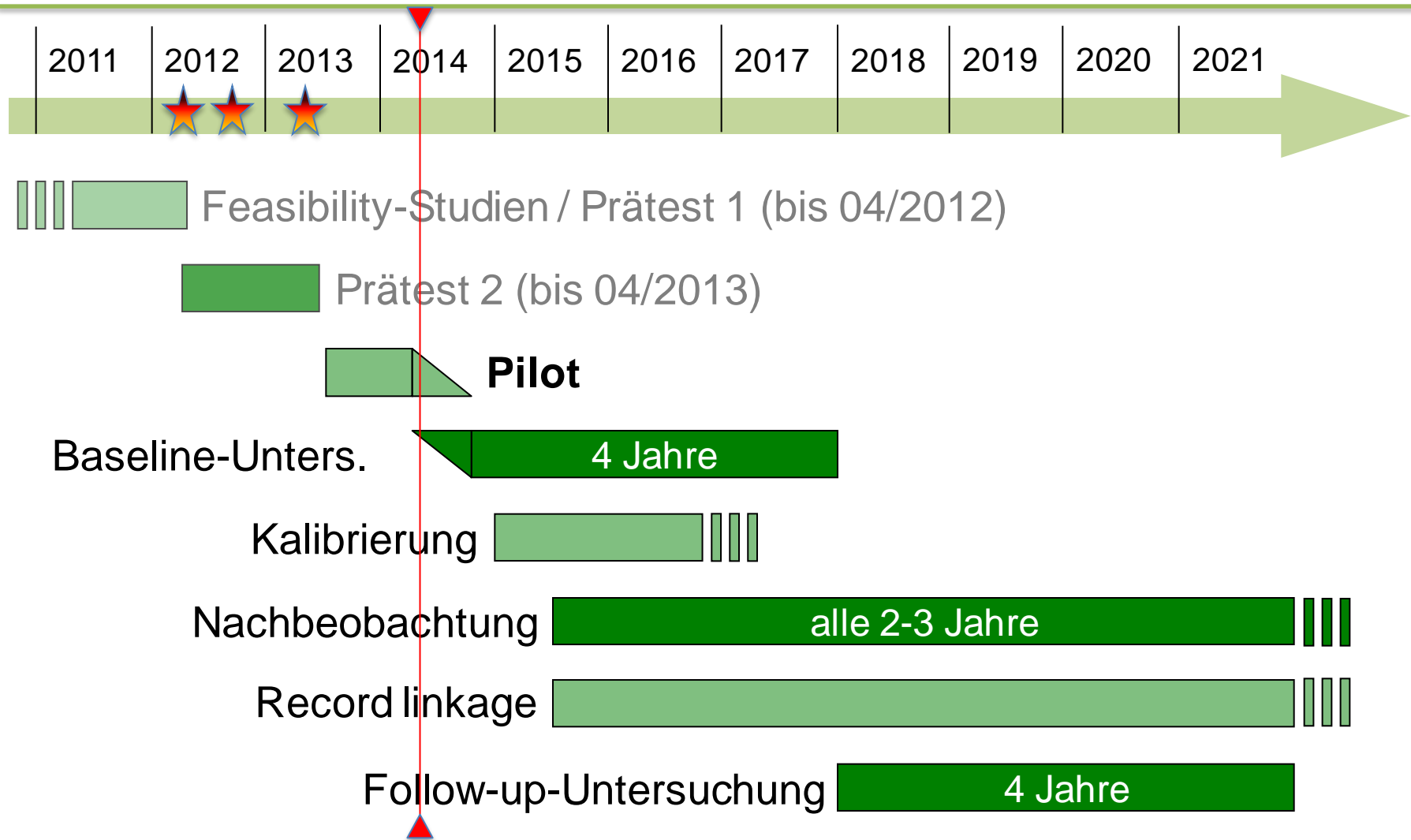
## Datenerhebung: Fazit

- Einfache und sichere Infrastruktur
  - Arbeitsstationen: nur Web-Browser und JavaScript
  - Touchscreen und Gerätedatentransfer integriert
  - Software-Updates und Änderungen am Datenmodell betreffen nur zentrale Server
  - Zwei identische Integrationszentren mit Online-Datenbankreplikation zur Ausfallsicherheit
  - Papierformulare als Backup erforderlich
  - Dezentralisierte Nutzerverwaltung, bereichs- und rollenbasierte Zugriffsrechte
- Spezielle Lösung zur Gerätedatenintegration

## Datenzugang und -nutzung

- Vollständig QS-gesicherte Daten verfügbar voraussichtlich erst ab 2018, **aber:**
- Art und Weise sowie Verfahren für den Zugang zu Daten und Bioproben *frühzeitig, vor Beginn der Hauptstudie* vereinbaren! (Gutachter, Datenschutz)
- Datenschutz: Re-Identifikationsrisiko minimieren
- Viele häufig wiederkehrende, streng geregelte Arbeitsschritte  $\Rightarrow$  Umsetzung in IT-Verfahren
- Nutzungsordnung (2013), Transferstelle (2015)

# Nationale Kohorte: Zeitplan



# Danke für Ihre Aufmerksamkeit.

## Nationale Kohorte – Zentrales Datenmanagement



Dipl.-Math. Stefan Ostrzinski  
Universitätsmedizin Greifswald  
ICM-VC - Zentrales Datenmanagement  
E-Mail: [stefan.ostorzinski@uni-greifswald.de](mailto:stefan.ostorzinski@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](mailto:d.kraft@dkfz-heidelberg.de)  
Tel.: +49 6221 42 3153