



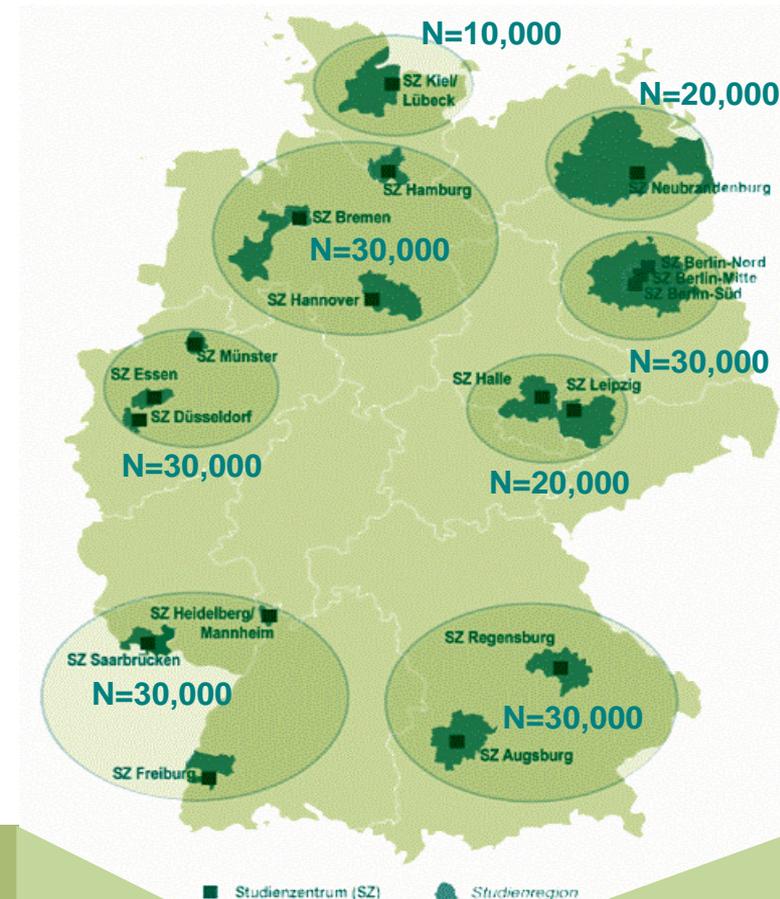
# Zentrales Datenmanagement in der Nationalen Kohorte

Stefan Ostrzinski & Daniel Kraft



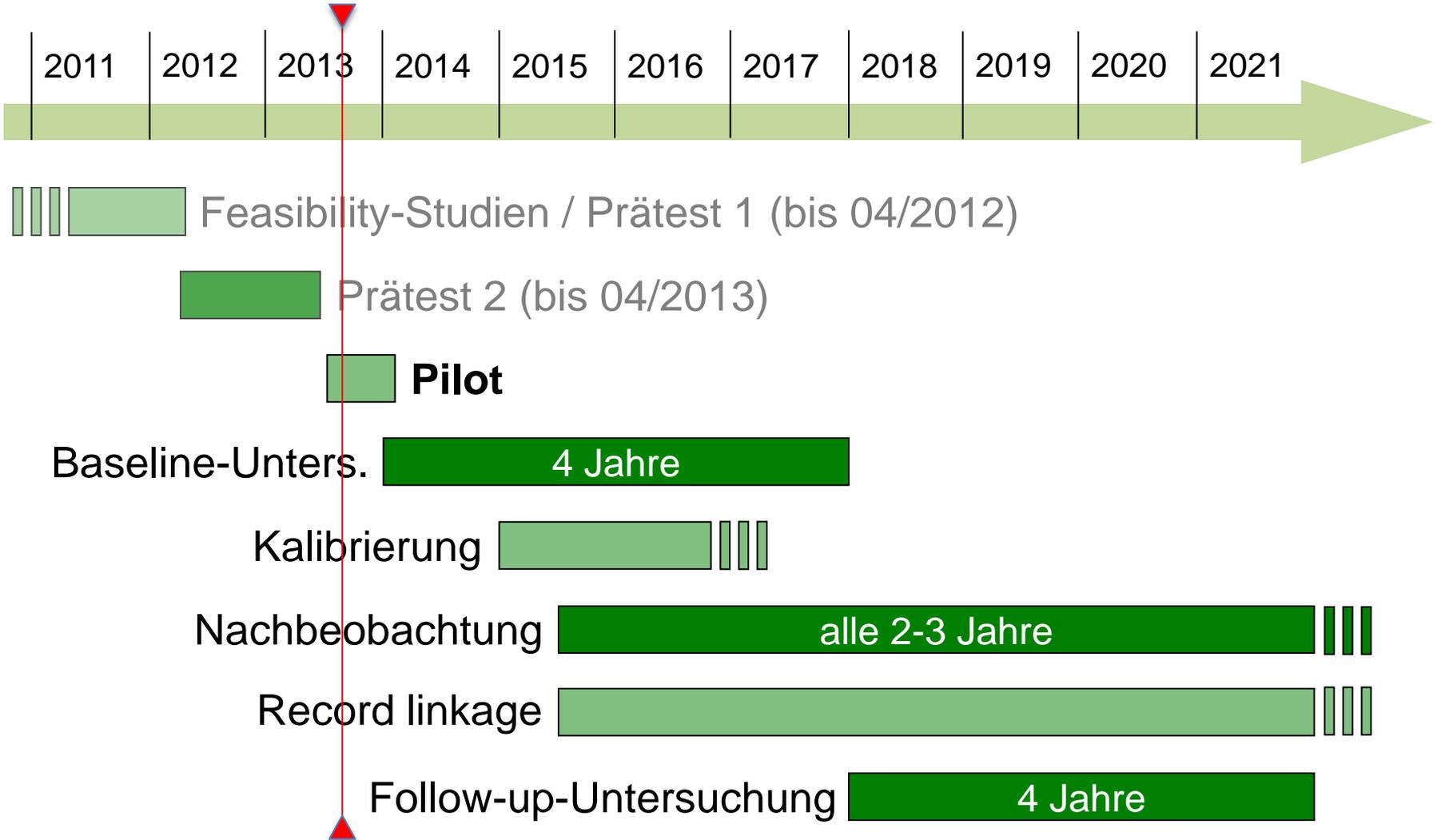
# Nationale Kohorte: Studiendesign

- Bevölkerungsbasierte prospektive Kohortenstudie
- Altersbereich 20-69 Jahre zur Baseline-Untersuchung (m/w)
- Repräsentativer Ansatz (Melderegisterstichprobe)
- 8 regionale Cluster, 18+ Studienzentren
- Level 1: ca. 2,5 h, N = 200.000
- Level 2: ca. 4 h, N = 40.000
- Level 3: lokal, zusätzlich, N variabel
- MRT-Programm: N $\approx$ 30.000





# Nationale Kohorte: Zeitplan





# Nationale Kohorte: Ziele

- **Ursachen von Volkskrankheiten** aufklären
- **Risikofaktoren** und Möglichkeiten der **Früherkennung** von Krankheiten identifizieren
- Wege einer wirksamen **Vorbeugung** aufzeigen
- **Infrastruktur** für die epidemiologische Forschung



[1] H.-E. Wichmann, · R. Kaaks, · W. Hoffmann, · K.-H. Jöckel, · K.H. Greiser, · J. Linseisen; Die Nationale Kohorte. Bundesgesundheitsblatt 2012 · 55:781–789; [2] [www.nationale-kohorte.de](http://www.nationale-kohorte.de)



# 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



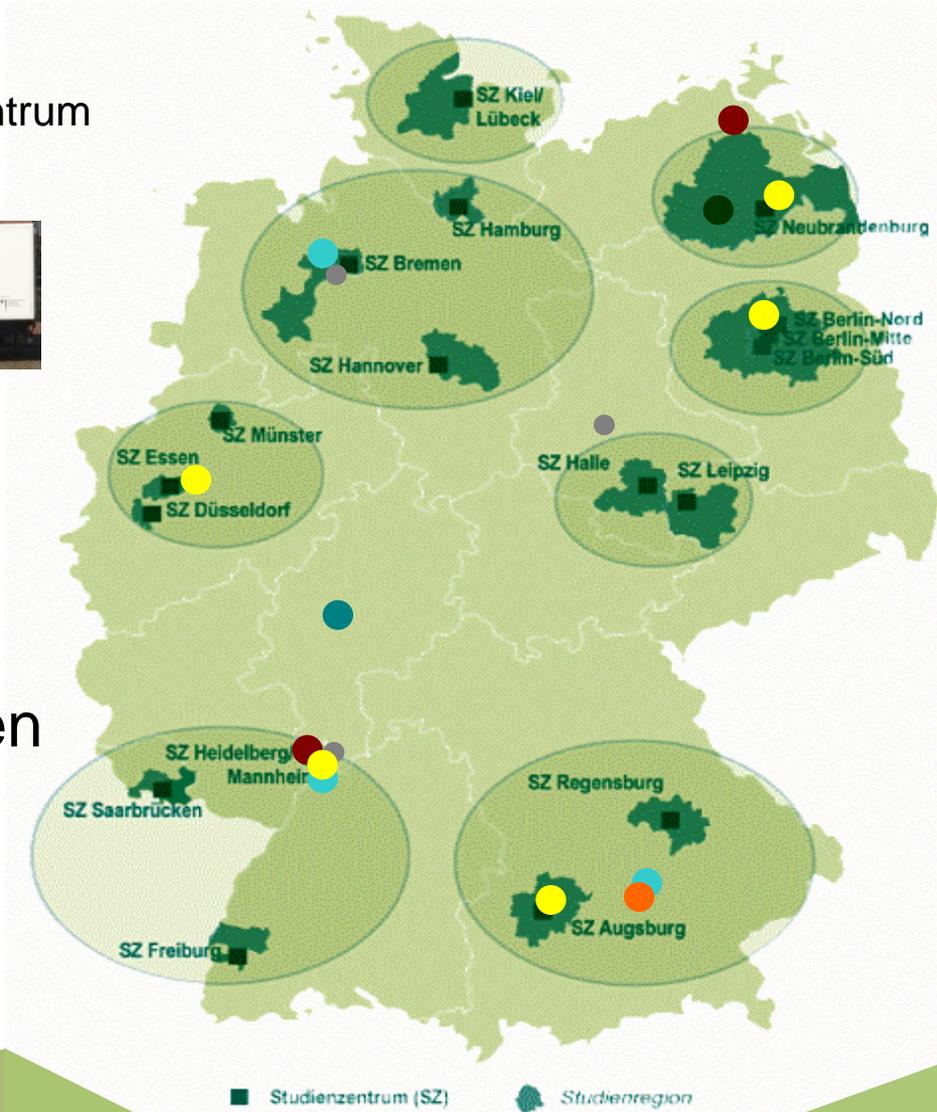
## Untersuchungsprogramm (Baseline)

- Level 1 (N=200.000)
  - Interview, Selbstausfüller-FB, Neurokognitive Tests
  - Anthropometrie, BIA, Blutdruck, Puls, Vascular Explorer
  - Handgreifkraft, Akzelerometrie, Spirometrie, Zahnstatus
  - Laborbasisprogramm, Biobank
- Level 2 (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)
  - MRT (ca. 53 TB Rohdaten)



# Standorte

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





# Datenmanagement: Herausforderungen

- **Dezentrale Organisation**  
Viele Beteiligte mit unterschiedlichen Voraussetzungen und Interessen
  - **Lange Projektlaufzeit**  
Daten gewinnen im Lauf der Jahrzehnte an wissenschaftlichem Wert
  - **Dezentrale Erhebung vs. zentrale Verarbeitung**  
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-Nutzung
- ⇒ **Standardisierung, Zentralisierung, Automation**



# Übergeordnete Aspekte (1)

- **Ethik und Datenschutz**
- **IT-Sicherheit**
- **Qualitätssicherung**



## Übergeordnete Aspekte (2)

- **Ethik und Datenschutz**
  - Informierte Einwilligung („Informed Consent“)
  - Trennung von identifizierenden und Studiendaten:  
lokales Probandenmanagement – unabhängige  
Treuhandstelle – Datenerhebung
  - Pseudonymisierte Datenerhebung, verschiedene  
Pseudonyme für unterschiedliche Teildatenbestände,  
Zweitpseudonymisierung vor Datenweitergabe
  - Unabhängige Treuhandstelle
  - Schutz von Probanden- *und* Mitarbeiterdaten
  - Votum des BfDI und der zuständigen Ethikkommissionen,  
Ethikbeirat, Kontakt mit AK Wissenschaft und lokalen DSB



## Übergeordnete Aspekte (3)

- **Ethik und Datenschutz**
- **IT-Sicherheit**
  - Verfügbarkeit von Daten und Anwendungen
  - Integrität der Daten
  - Vertraulichkeit
  - Technische und organisatorische Maßnahmen

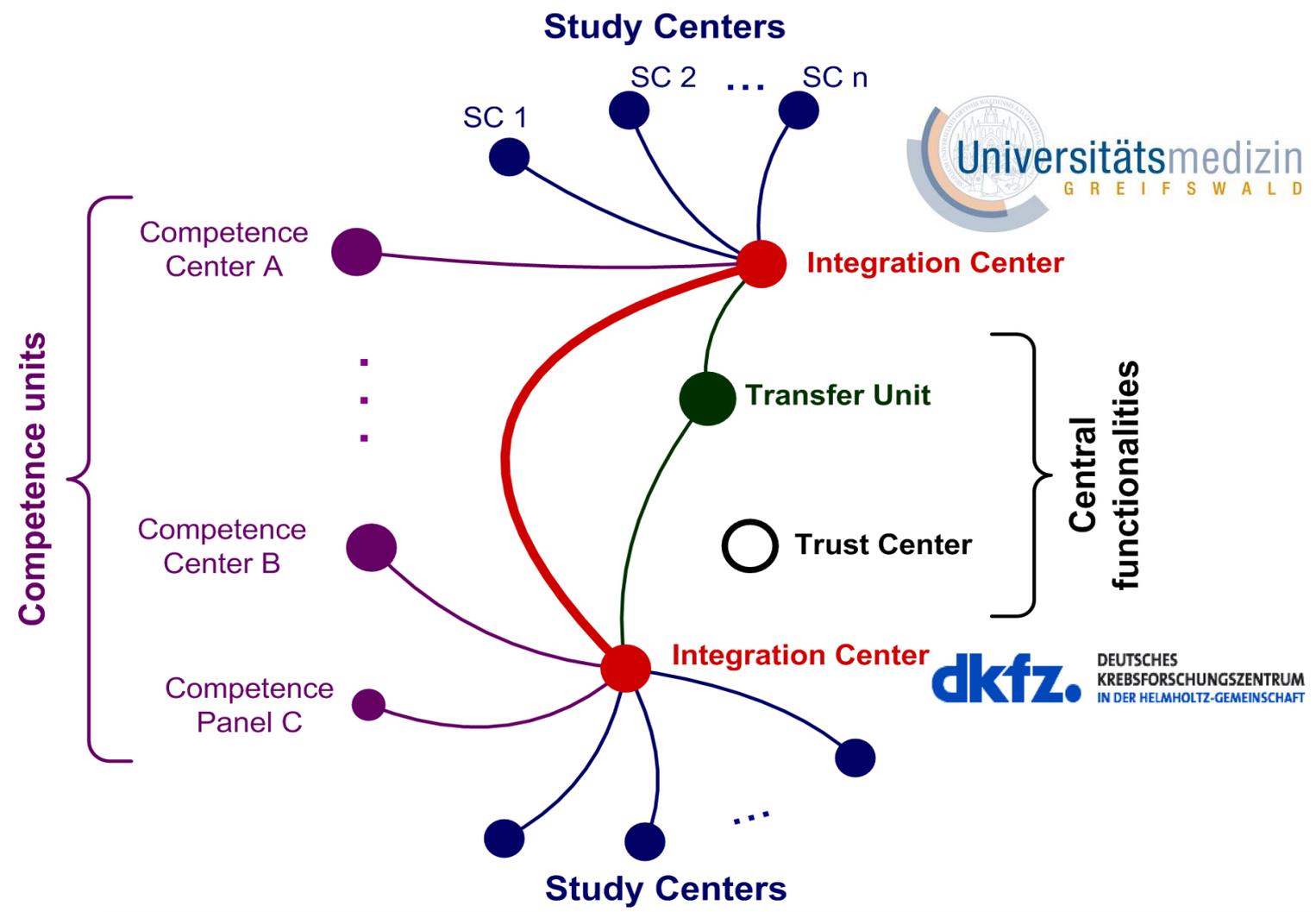


# Übergeordnete Aspekte (4)

- **Ethik und Datenschutz**
- **IT-Sicherheit**
- **Qualitätssicherung**
  - Standardisierung aller Prozesse  
(zentrumsübergreifend und über die Zeit)
  - Prozessüberwachung  
(Ablaufkontrolle, Validierung von Daten)
  - „Historisierung“ aller Datenänderungen
  - Versionsmanagement  
für Anwendungen und Data Dictionary



# Datenmanagement: Organisationsmodell





# Datenerhebung: Grundsätze

- Identische, zentral gepflegte Erhebungsinstrumente
- Data Dictionary als Grundlage
  - Erstellung in Koordination mit SOP-Erstellung
  - Dynamische Generierung standard. Erhebungsinstrumente
  - Identische Beschreibung für Datenerhebung, -bereinigung -bereitstellung und -auswertung
- Sofortige Integration der Daten
  - Informationen sofort überall nutzbar (Formular/Zentrum)
  - Sofortige Konsistenz- und Plausibilitätsprüfung
  - Zentrale Datensicherung
  - Ansatzpunkt für automatisierte QS-Routinen



# Datenerhebung: Webanwendung

[Abmelden](#)

[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](#)

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

## 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, einen deutschen Abschluss mit den deutschen Entsprechung angeben.*  
*Formulierung der Antwortmöglichkeiten an das Geschlecht des Probanden anpassen.*  
*Erhöhe runter die Antwortoption laut Liste*

Bitte auswählen...

- Schüler/-in, besuche eine allgemeinbildende Vollzeitschule
- Von der Schule abgegangen ohne Hauptschulabschluss (Volksschulabschluss)
- Hauptschulabschluss (Volksschulabschluss)
- Realschulabschluss (Mittlere Reife)
- Polytechnische Oberschule der DDR mit Hochschulreife 8. oder 9. Klasse
- Polytechnische Oberschule der DDR mit Hochschulreife 10. Klasse
- Fachhochschulreife, Abschluss einer Fachoberschule
- Allgemeine Hochschulreife (Abitur)

[Weiter](#)   [Zurück](#)   [Zur näch](#)

- Webbasierte Erhebungsanwendung
- Touchscreen für Tests und Selbstausfüller-Fragebögen

• Dynamische Formularerstellung (Datenbank)

• Zentrale Studiendatenbank

- Aufklärung / Einwilligung
- Aktualisierung identifizierender Daten
- Ablaufdokumentation



# Datenerhebung: Touchscreen

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





# Datenerhebung: Formularerstellung

- ⇒ Papierformular als Vorlage
- ⇒ “Formale Beschreibung” (Data Dictionary)
- ⇒ Datenbank
- ⇒ Webanwendung
- ⇒ Webformular



# Datenerhebung: Diagnostische Geräte

Geräte unterstützen meist keine gesicherte Übertragung zu entferntem Ziel, deshalb ...

Ausgabedateien  
enthalten Teilnehmer-  
und Geräteerkennung

Gerät

- initialisiert sich aus Datei
- speichert Daten lokal

Einheitl. lokale Anbindungs-Software

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

Server in Integrationszentren:

- Initialisierungsdatenerzeugung
- Integration in Studiendatenbank
- Archivierung
- Statusanzeige in Web-Anwendung



# Datenerhebung: Schnittstellen ...

- 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
  - Inkl. Mortalitäts-Follow-Up



## Datenerhebung: Fazit

- Einfache Infrastruktur
  - Arbeitsstationen benötigen nur Web-Browser und JavaScript, damit kaum Wartungsbedarf
  - Touchscreen-Anwendung integriert
  - Software-Updates und Änderungen am Datenmodell betreffen nur zentrale Server
  - Papierformulare als Backup erforderlich
- Bereichs- und rollenbasierte Zugriffsrechte
  - Dezentralisierte Nutzerverwaltung
- Spezielle Lösung zur Gerätedatenübermittlung



# Datennutzung

- 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, Konzept Transferstelle



# 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](mailto: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](mailto:d.kraft@dkfz-heidelberg.de)  
Tel.: +49 6221 42 3153