

FORSCHUNGSDATENMANAGEMENT IN EINER INTERNATIONALEN MULTIZENTRISCHEN FOLLOW-UP-STUDIE (I.FAMILY/IDEFICS)

Dr. Achim Reineke

Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS
- on behalf of the I.Family consortium

5. Workshop AK-Forschungsdaten
Berlin, 28.06.2016



Beschreibung I.Family

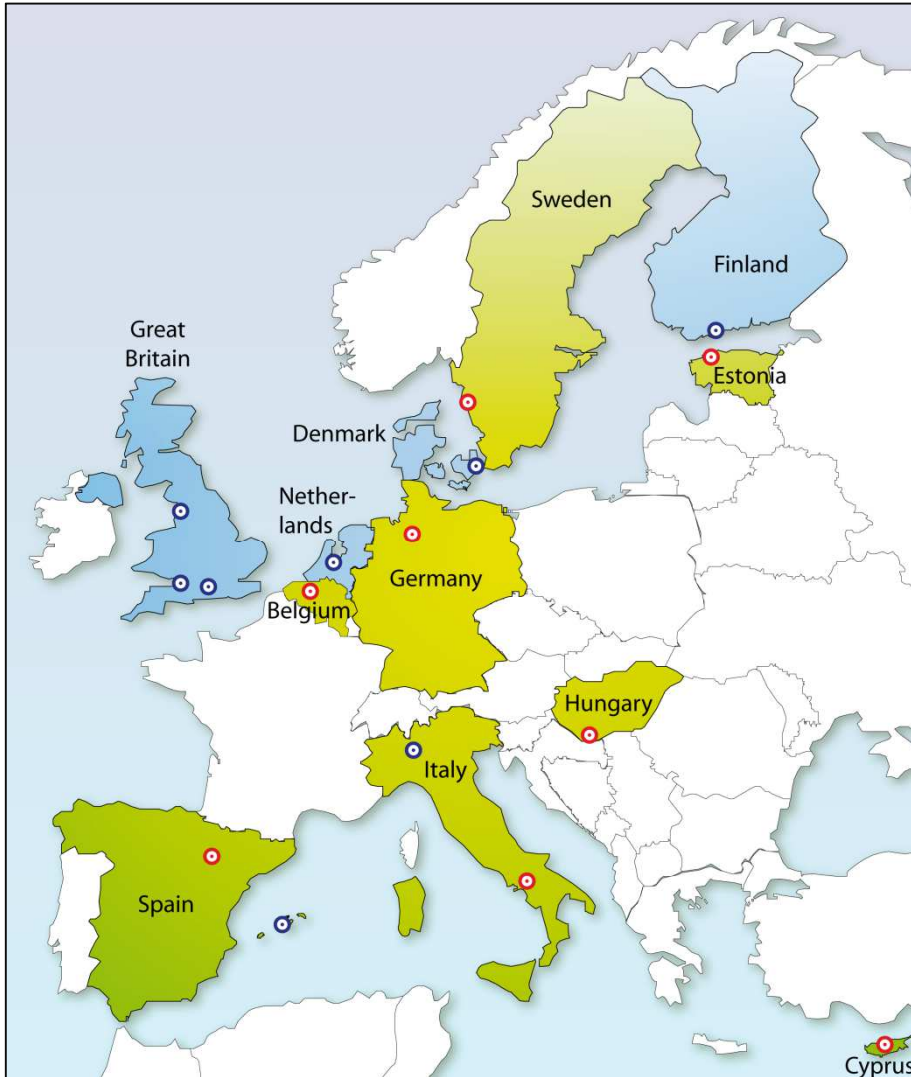


Einleitung

3

- Multizentrische Follow-Up-Studie
- Ziel: Erforschung von Einflussfaktoren auf das Ernährungsverhalten europäischer Kinder, Jugendlicher und ihrer Eltern
- 2013: Start der Erhebung
- Über 2000 Variablen bei mehr als 17.000 Probanden
- BIPS als Koordinator

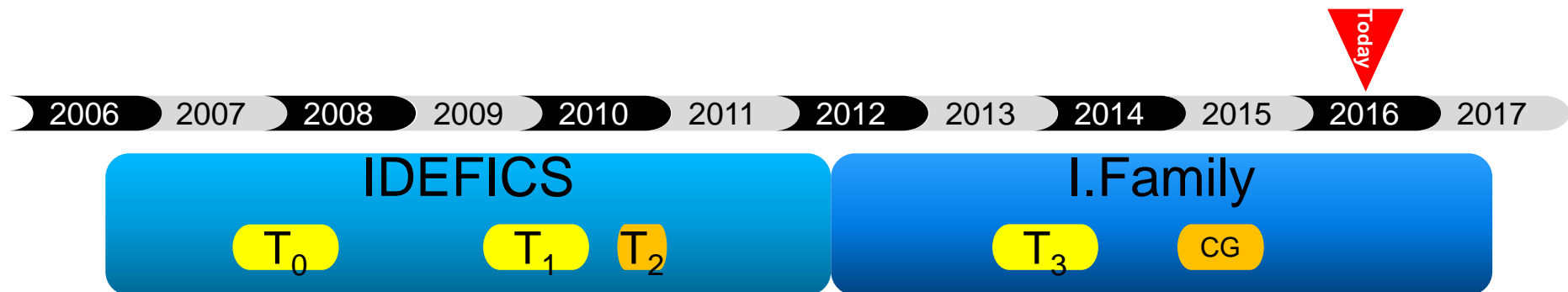
Partner



Partner:

- Strovolos, Cyprus
- Ghent, Belgium
- Copenhagen, Denmark
- Tallin, Estonia
- Helsinki, Finland
- Bremen, Germany
- Pécs, Hungary
- Avellino, Italy
- Milan, Italy
- Utrecht, Netherlands
- Palma de Mallorca, Spain
- Zaragoza, Spain
- Gothenburg, Sweden
- Bristol, United Kingdom
- Lancaster, United Kingdom
- Monxton, United Kingdom

Timeline



- T_3 : Follow-up der Index-Kinder (mit Geschwistern und Eltern)
- CG: Zusätzliche Untersuchungen in den **C**ontrasting **G**roups/ und Untergruppen

Instrumente

- Fragebögen (Eltern + Kind)
 - Medienkonsum, soziales Umfeld, Lebensstil
 - Medizinische Vorgeschichte, Medikamente
 - Ernährungsgewohnheiten, Ernährungsprotokoll
 - Körperliche Aktivität
- 24-Std Ernährungsprotokoll
 - SACANA: web-basierter „24-h dietary recall“
- Bewegungsverhalten
 - Accelerometer: 7 Tage

Please note the time and the date when you start completing this questionnaire

STARTING TIME: at [] [] [] h.

DATE: [] Day [] Month [] Year

GENERAL INFORMATION ABOUT YOUR CHILD

Please give some information about the child you are filling in

What is the date of birth of your child?
[] Day [] Month [] Year

What sex is your child?
Male Female

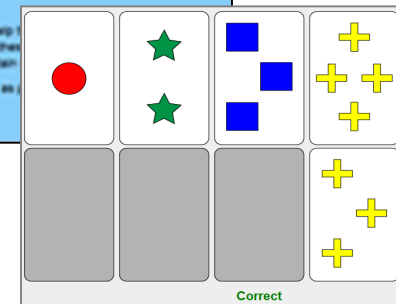
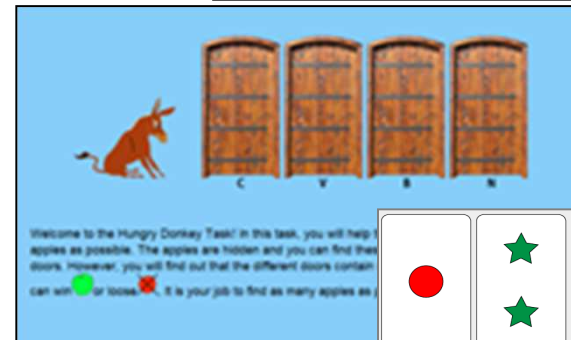
Please tick wherever the child ate here

Food Item	Rating (Smiley Faces)
Lasagne	☹ ☹ ☹ ☹ ☹
Braised Sprouts	
Sprouts	
Applesauce	
Old bread	



Untersuchungen und Messungen

- Ärztl. Untersuchung
 - Anthropometrie + Blutdruck
 - Muskuloskelettale Störungen
- Bioproben
 - Blut, Speichel, Urin
- Add-ons, z.B.
 - Körperliche Fitness
 - Geschmackssensorik
 - Bebaute Umgebung: GIS + GPS
 - Auswirkung von Lebensmittelwerbung
 - Neuropsychologische Tests



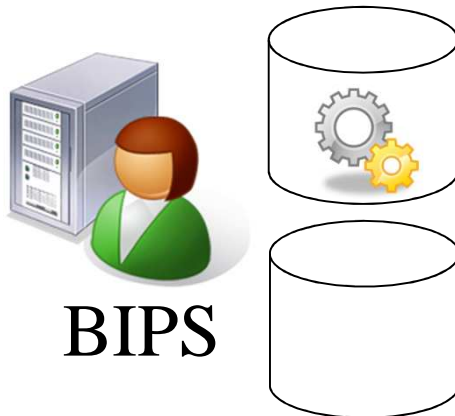
Erhebung der Daten



- Erhebung in acht Zentren
 - In verschiedenen Ländern
 - In unterschiedlichen Sprachen
- Konzeption der Instrumente erfolgte durch die jeweiligen Experten im Konsortium
- Umsetzung der Instrumente in Datenbanken (Datenbankapplikationen) und Eingabewerkzeuge erfolgte zentral durch das BIPS
- Datenmanagement erfolgt zentral

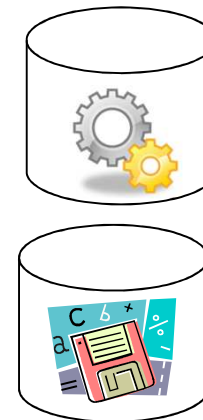
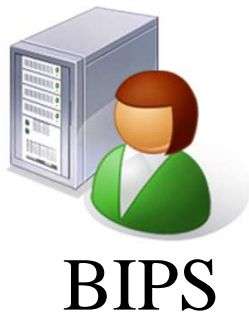
Datentransfer IDEFICS I

Interne Projekt-Web-Plattform



Datentransfer IDEFICS II

Interne Projekt-Web-Plattform



**Studien-
zentrum**



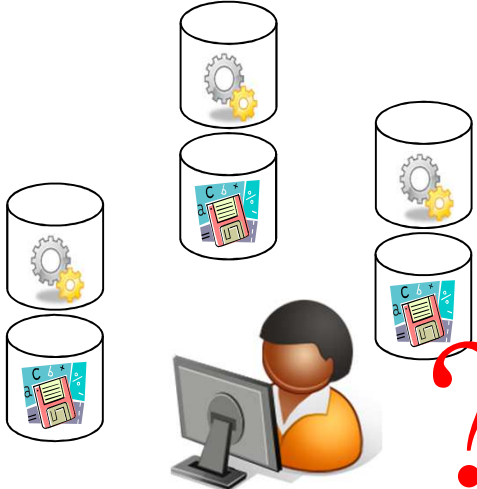
Datentransfer IDEFICS III



Interne Projekt-Web-Plattform



BIPS



Studienzentrum

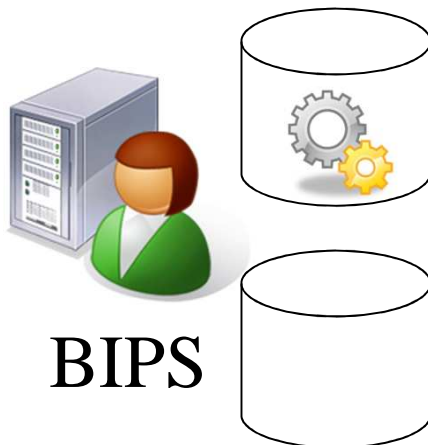
Probleme in IDEFICS

- Große Arbeitsbelastung durch Austausch der Datenbanken und der Daten
 - Doppelte Datensätze (mehrfach erfasst / Kopie der Eingabedatenbank) bzw. Testdatensätze mit ungültigen IDs
 - Verwendung alter Versionen der Eingabedatenbanken
- Probleme durch verzögerte Entwicklung einiger Instrumente
 - Verfügbare Zeit für die Programmierung und Verteilung aller Datenbanken reichte nicht aus
 - Dateneingabe konnte für einige Instrumente erst mit einigen Wochen Verzögerung gestartet werden
- Verzögerung nach dem Ende der Feldphase bis zur Verfügbarkeit des ersten bereinigten Datensatzes

- PCs, Laptops und Untersuchungsgeräte in den einzelnen Studienzentren
- Zentraler Datenaufbereitungs- und Auswertungsserver (Central Data Server, CDS) für:
 - Dateneingabe (mit Prüfeingabe)
 - Upload (automatisch) der gesammelten (Mess-)Daten
 - Plausibilitätsprüfungen
 - Datenaufbereitung
 - Datenanalyse
- ❖ Lokale Verarbeitung der personenidentifizierenden Daten (Name, Adresse, ...) in den Zentren

Datentransfer I.Family I

CDS

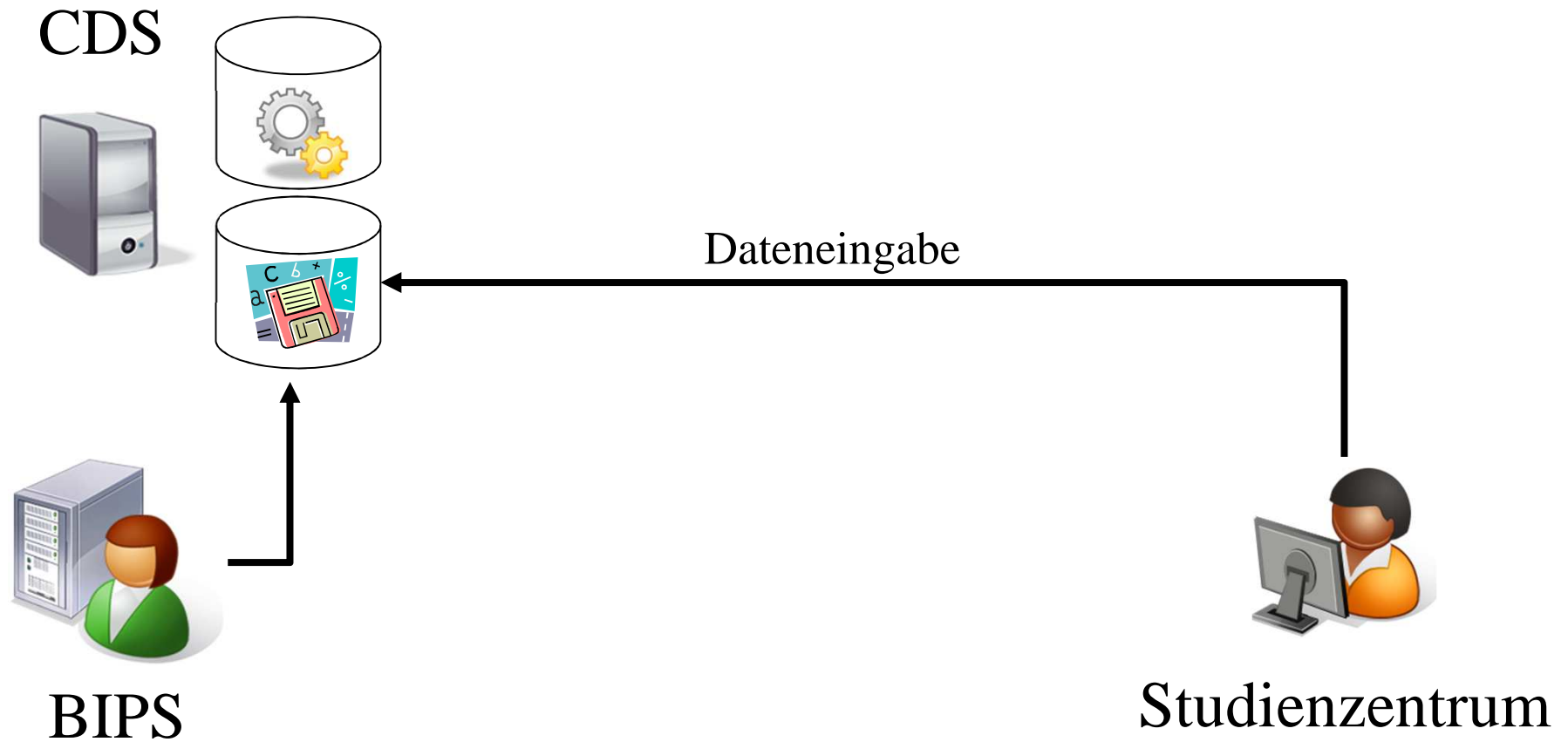


BIPS



Studienzentrum

Datentransfer I.Family II

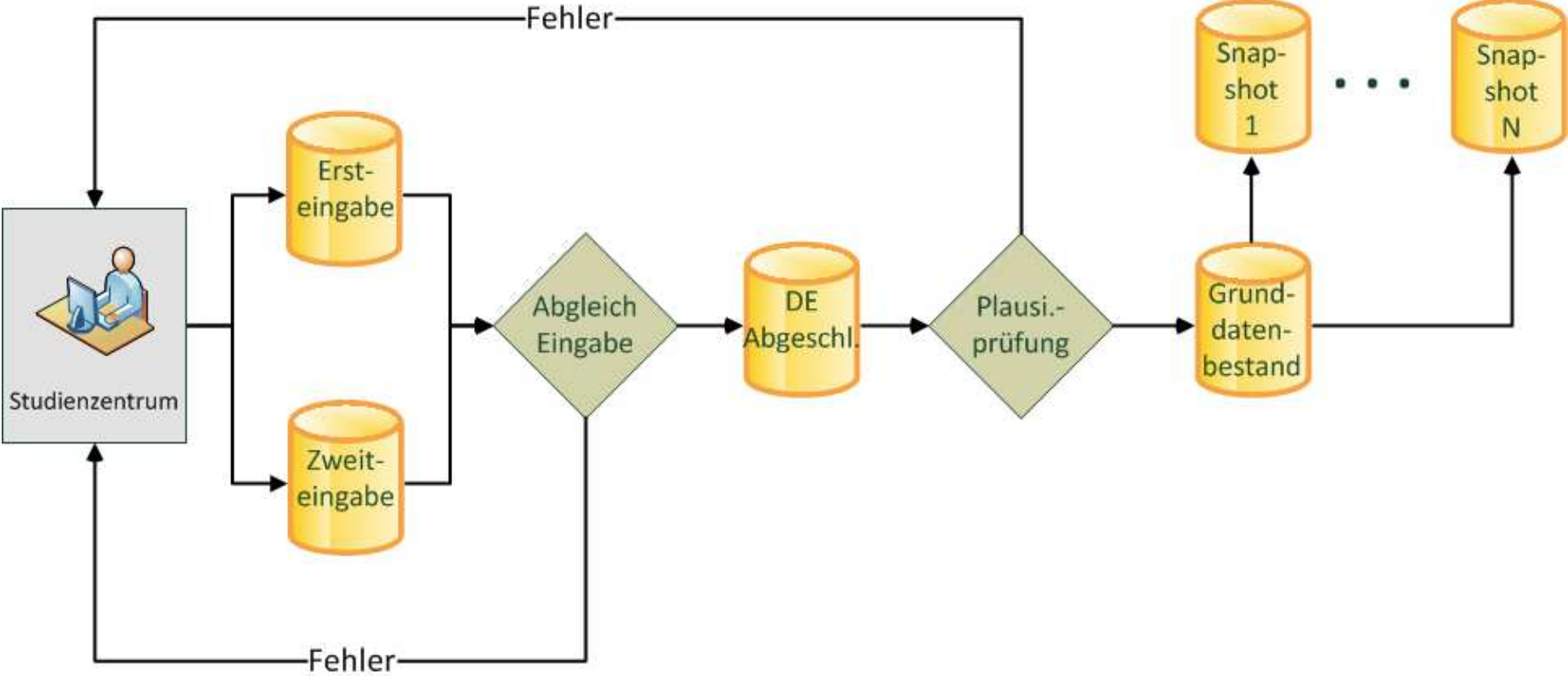


- Eingabe, Prüfung und Korrektur erfolgt direkt auf dem CDS
- Daten von Messgeräten wurden automatisch auf den CDS hochgeladen
 - Vorbereitete Kopier- und Archivierungsprogramme

➤ Vorteile

- Sämtliche Daten stehen kurzfristig auf dem CDS zur Verfügung
→ keine gesammelte Lieferung am Ende der Studie
- Konsistenz und Plausibilitätsprüfungen können sowohl die eingegeben Daten als auch die Messdaten umfassen (Crosschecks)
- Zeitnahe Rückmeldungen an die Partner möglich

Datenverarbeitung



- Heterogener Datenkörper:
 - Befragungsdaten
 - Daten aus ärztl. Untersuchung
 - Messdaten (Accelerometer, GPS, fMRT)
 - Bioproben
- Volumen:
 - 500 MB Befragungsdaten (pro Snapshot)
 - 250 GB Accelerometerdaten
- Zusammenfassung zu einen Core-Dataset
- Regelmäßige Bereitstellung von bereinigten Auswertungsdatensätzen (Snapshots)

- Informationen zu einzelnen Objekten (Codebook, annotierter Fragebogen, Datensatzbeschreibung, ...)
 - Beschreibung
 - Quelle (Instrument, Zeitpunkt)
 - Wertebereiche
 - Algorithmus / Verfahren der Berechnung
- Änderungshistorie zu einzelnen Objekten
- Änderungshistorie zu den Analysedatensätzen

- Studienzentren
 - Kontaktierung der Probanden (personenidentifizierende Daten)
 - Datenerhebung
 - Dateneingabe / Upload
 - Datenkorrekturen / Abgleich mit Originaldaten (Fragebogen)
- Koordinierendes Zentrum (BIPS)
 - Bereitstellung der gemeinsam genutzten Infrastruktur
 - Definition und Koordination der Datenverarbeitungsprozesse
 - Plausibilitäts- und Konsistenzprüfungen
 - Speicherung (und später Archivierung) der Daten
 - Bereitstellung der Daten für Analysen

- Studienzentren
 - Wissenschaftliche Team
 - Erhebungskräfte
 - Lokale IT

- Koordinierendes Zentrum (BIPS)
 - Medizinische Dokumentare (Datenbanken, Datenmanagement)
 - Mitarbeiter aus den statistischen Fachgruppen (Plausibilitätsprüfungen)
 - IT (Betrieb des CDS, Upload-Programme, ...)

Anforderungen an den CDS



22

- Für alle Partner erreichbar
- Verfügbar
- Ausreichend dimensioniert (Speicher / Leistung)
- Bereitstellung benötigter Funktionalitäten (Software)
- Gesicherte Kommunikation / Zugriffskontrolle
 - Zugang über VPN
 - Vergabe von Benutzerkennungen und Passwörtern
 - Definition von Verfahren zur Vergabe der Zugriffsberechtigungen
 - Anleitungen und Vorgaben für die Nutzung
- ❖ Sonderregelungen für HPC-Cluster für Genomanalysen

- Standardsoftware auf dem CDS (englische Versionen)
 - MS Office / MS Access
 - SAS / SPSS / STATA / R
 - MySQL
 - Spezialsoftware (ArcGis, Nährwertberechnung, ...)
 - Datenaufbereitungs- und Datenmanagementprogramme
- Konvention zur Nutzung von UTF-8 als Kodierung für Dateien

Vergabe der IDs



24

- Vergabe mehrerer IDs pro Individuum (IDEFICS-ID, I.Fam-ID, Familien-ID, Bioproben-IDs, ...)
- Vorgabe der Struktur und der gültigen Bereiche (pro Zentrum) durch das BIPS
- Aufbau der IDs
 - Stelle 1 - Studie
 - Stelle 2 - Land
 - Stelle 3 - Region
 - Stelle 4 bis n - Proband
- Zuordnung der IDs zu realen Probanden verbleibt im jeweiligen Studienzentrum

Nutzung und Bereitstellung der Daten



25

- Kontrolle der Datennutzung durch projektübergreifendes Gremium
- Austausch ist begrenzt auf pseudonymisierte Daten
- Alle Partner kommen aus Ländern der EU
- Bereitstellung zur Nutzung für die nächsten X Jahre
- Anschließend Archivierung für 10 Jahre gem. GEP

- Gesamtprojekt: 13 Millionen Euro zuzüglich weiterer Eigenanteile der Partner
 1. Erhebung der Daten in der Feldphase
 2. Erstellung der notwendigen Datenbanken etc.
 3. Datenaufbereitung / Datenkorrekturen
 4. Betrieb des CDS (Hardware, Lizenzen und Personal)
- Zusätzliche Kosten für Internationalisierung
 - Berücksichtigung der lokalen Gegebenheiten und Vorgaben
 - Mehrsprachigkeit der Werkzeuge
 - Aufwand für Erstellung der Dokumentationen
 - Kommunikation

Vielen Dank für Ihre Aufmerksamkeit

www.bips.eu

Kontakt

Achim Reineke

Leibniz-Institut für Präventionsforschung
und Epidemiologie – BIPS
Achterstraße 30
28359 Bremen

areineke@leibniz-bips.de

