Automatisierung des Datenreinigungsprozesses in einer internationalen multizentrischen Follow-up-Studie (I.Family-Studie)

Willempje Hummel-Bartenschlager
Leibniz-Institut für Präventionsforschung und Epidemiologie - BIPS

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 266044

Building on ideficsstudy

Überblick

- Eckpunkte der I.Family-Studie
- Aufgaben des BIPS Datenmanagements innerhalb der Studie
- Datenbereinigung in der Vorgänger-Studie IDEFICS
- Umstrukturierung der Eingabe-Datenbanken in I.Family
- Datenbereinigungprozess in I.Family
- Ergebnisse

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

I.Family

- Multizentrische Follow-Up-Studie in 8 europäischen Ländern
- Erforschung von Einflussfaktoren auf das Ernährungsverhalten europäischer Kinder, Jugendlicher und ihrer Eltern
- BIPS als Koordinator
- 2013 Start der Erhebung
  - über 2000 Variablen bei mehr als 17000 Probanden
- 10 Eingabe-Datenbanken
  - auf zentralem Server
  - Erst- und Zweiteingabe
  - diverse Interviews und Tests auf lokalen Computern

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Aufgaben des Datenmanagements im BIPS in I.Family

- Programmierung und Bereitstellung aller Datenbankapplikationen
- Aufbereitung der in Erst- und Zweiteingabe-DB eingegebenen Daten
  - Bereinigung von Dopplern
  - Löschung leerer Datensätze/falsch eingegebener Daten
  - erste Datenkorrekturen
  - Erstellung von Dateien als Grundlage für die Weiterverarbeitung
- Datenbereinigung
  - Vergleich Erst- und Zweiteingabe
  - Prüfung auf unplausible Werte
  - Bereitstellung von Korrektur-Datenbanken
- Bereitstellung Analyse-Datenbank

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

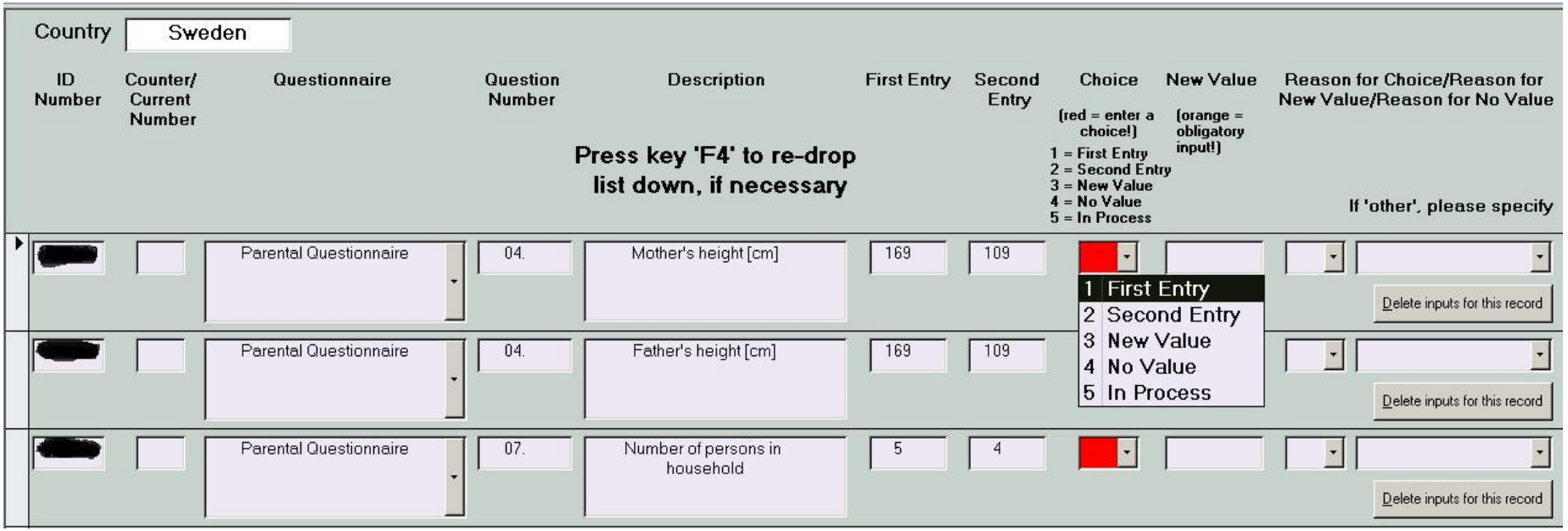
Datenbereinigung in der IDEFICS-Studie

- Vorgängerstudie von I.Family
  - Erforschung und Prävention von Übergewicht und Lebensstil-bedingten Erkrankungen bei Kindern in acht europäischen Ländern
  - Erhebung von 2007 bis 2010
- T0 (Studienphase 1)
  - SAS-Programme
    - Prüfung auf Unterschiede zwischen Erst- und Zweiteingabe
      - Nach Beendigung der gesamten Datenerhebung
    - Prüfung auf unplausible Werte
- Übermittlung der Korrekturlisten
- Manuelle Eingabe der Korrekturen durch die Zentren in die Ersteingabe-Datenbank

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Datenbereinigung in der IDEFICS-Studie

- T1 (Studienphase 2)
  - Erster Schritt zur Automatisierung
    - Einsatz von Access-Datenbanken für die Korrekturen
    - Automatischer Export der Korrekturen in die Ersteingabe-Datenbank



Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Vorteile durch Einsatz von Korrektur-Datenbanken

- Weniger fehlerträchtig
- Erhebliche Zeitersparnis
- Elektronische Dokumentation der Korrekturen gemäß GEP

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Umstrukturierung der Eingabe-Datenbanken in I.Family

- Access-Backend-Datenbanken wurden durch MySQL-Datenbanken ersetzt
  - Erhöhte Datensicherheit
  - Unbegrenzte Dateigröße
  - Bessere Performance durch schnelleren Zugriff und Datentransfer
  - Änderungen an den Daten-Tabellen zentral
  - Einsatz von Trigger
    - Nachverfolgung von Änderungen an den Daten

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Einsatz eines zentralen Datenservers

- Bereitstellung der Eingabe- und Korrektur-Datenbanken auf zentralem Server
  - Frontend: Access-Datenbank
    - Programmierte erste Plausibilitätsprüfungen
  - Backend: MySQL-Datenbank
    - eine Datenbank für alle Zentren
    - Zusammenspielen der Daten entfällt

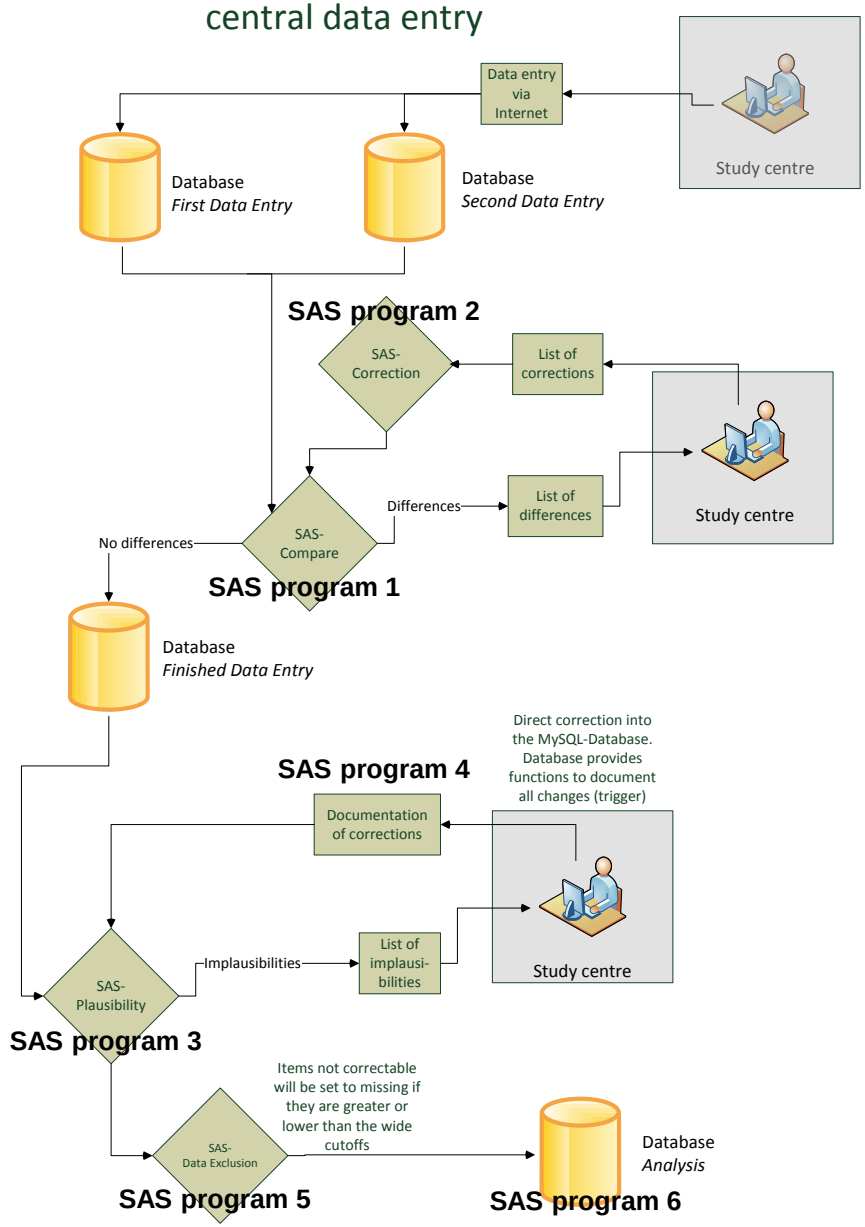
Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Vorteile eines zentralen Datenservers

- Bereitstellung der Anwendungsprogramme
- Problemlose und schnelle Bereitstellung von Versionsänderungen
- Unkomplizierter Datentransfer ohne Zwischenschritte
  - Verfügbarkeit der Daten für Datenmanagement direkt nach Eingabe
  - Vermeidung unterschiedlicher Versionen

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Data entry and plausibility checks in I.Family central data entry



Data entry

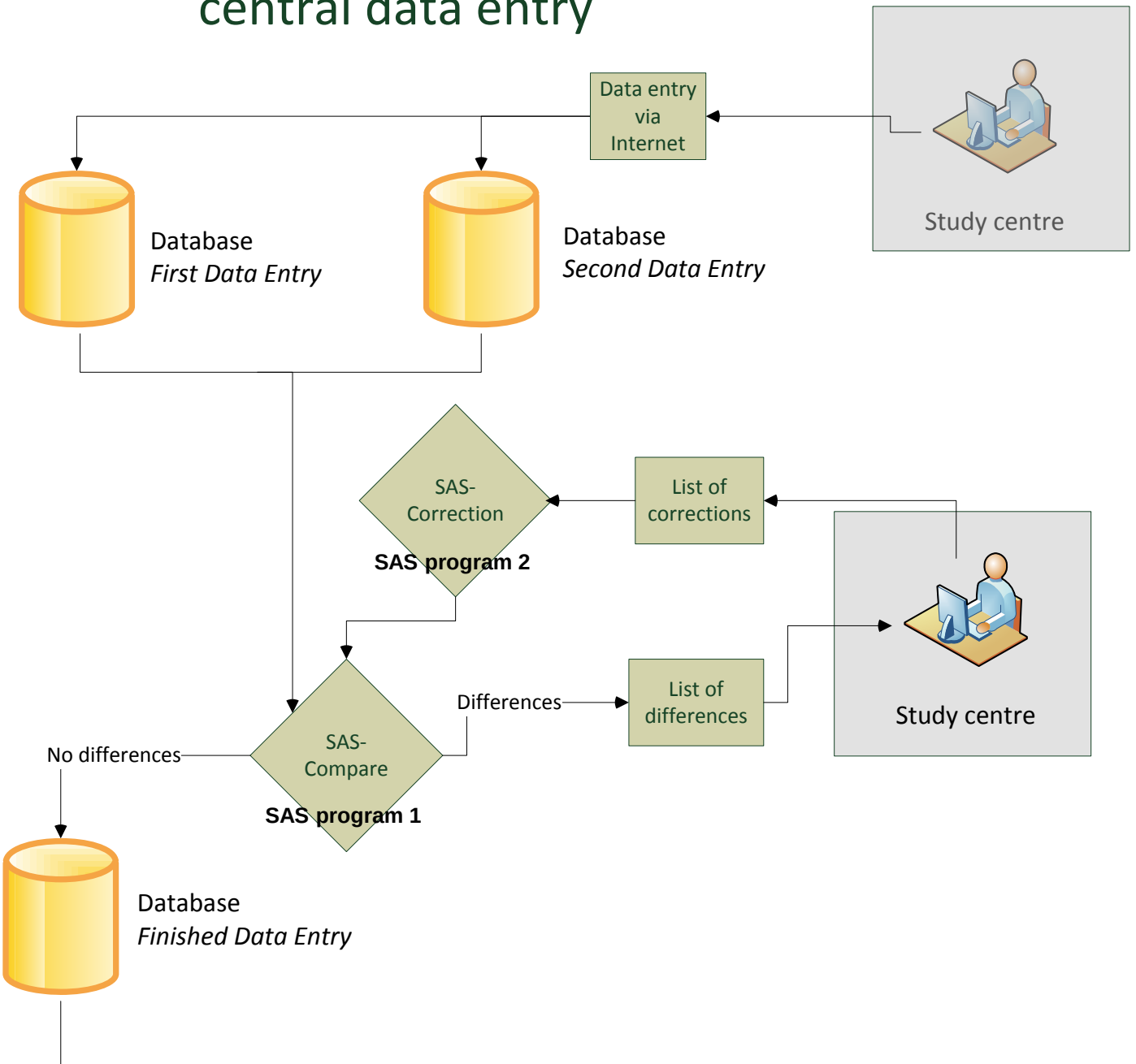
Comparison of 1st and 2nd entry and corrections

Plausibility checks and corrections

Standardised corrections / exclusion of non-correctable values

Data entry and plausibility checks in I.Family

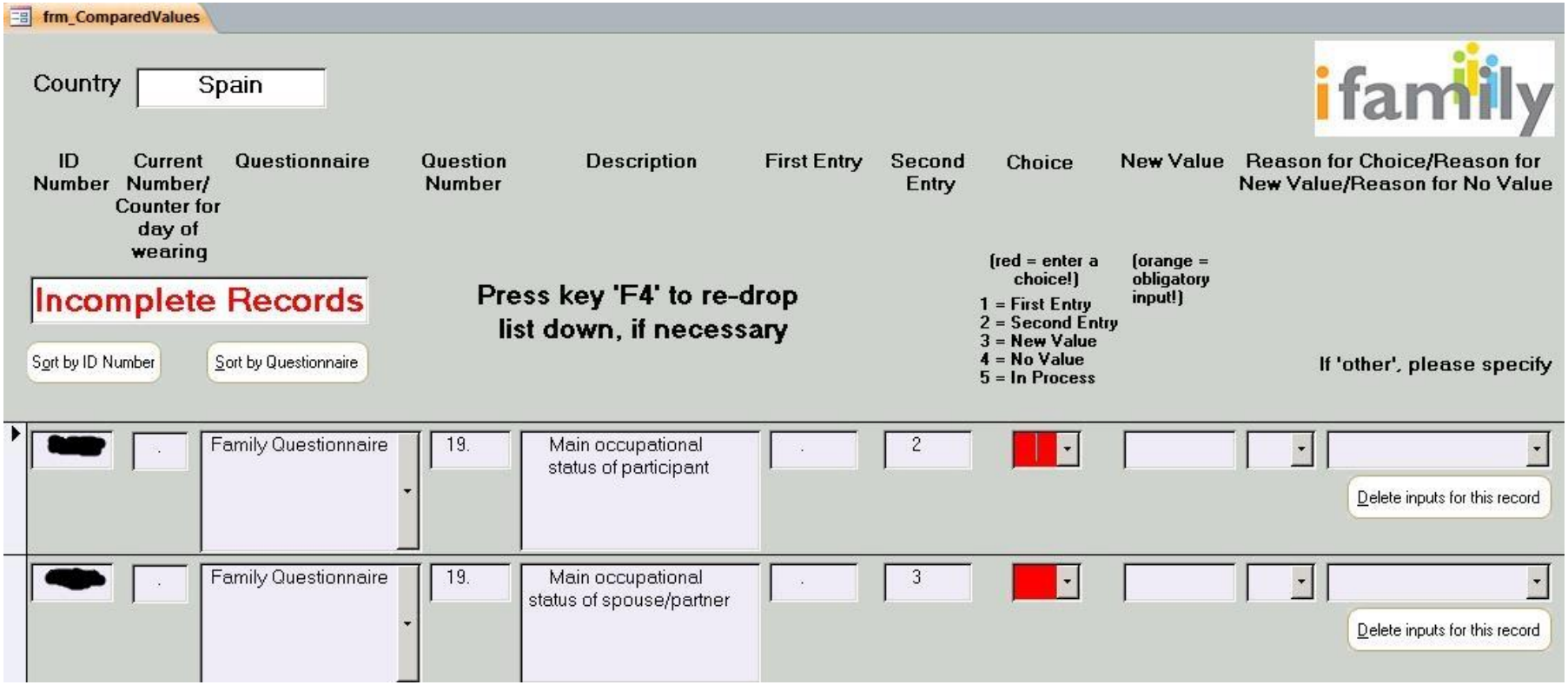
central data entry



Data entry

Comparison of 1st and 2nd entry and corrections

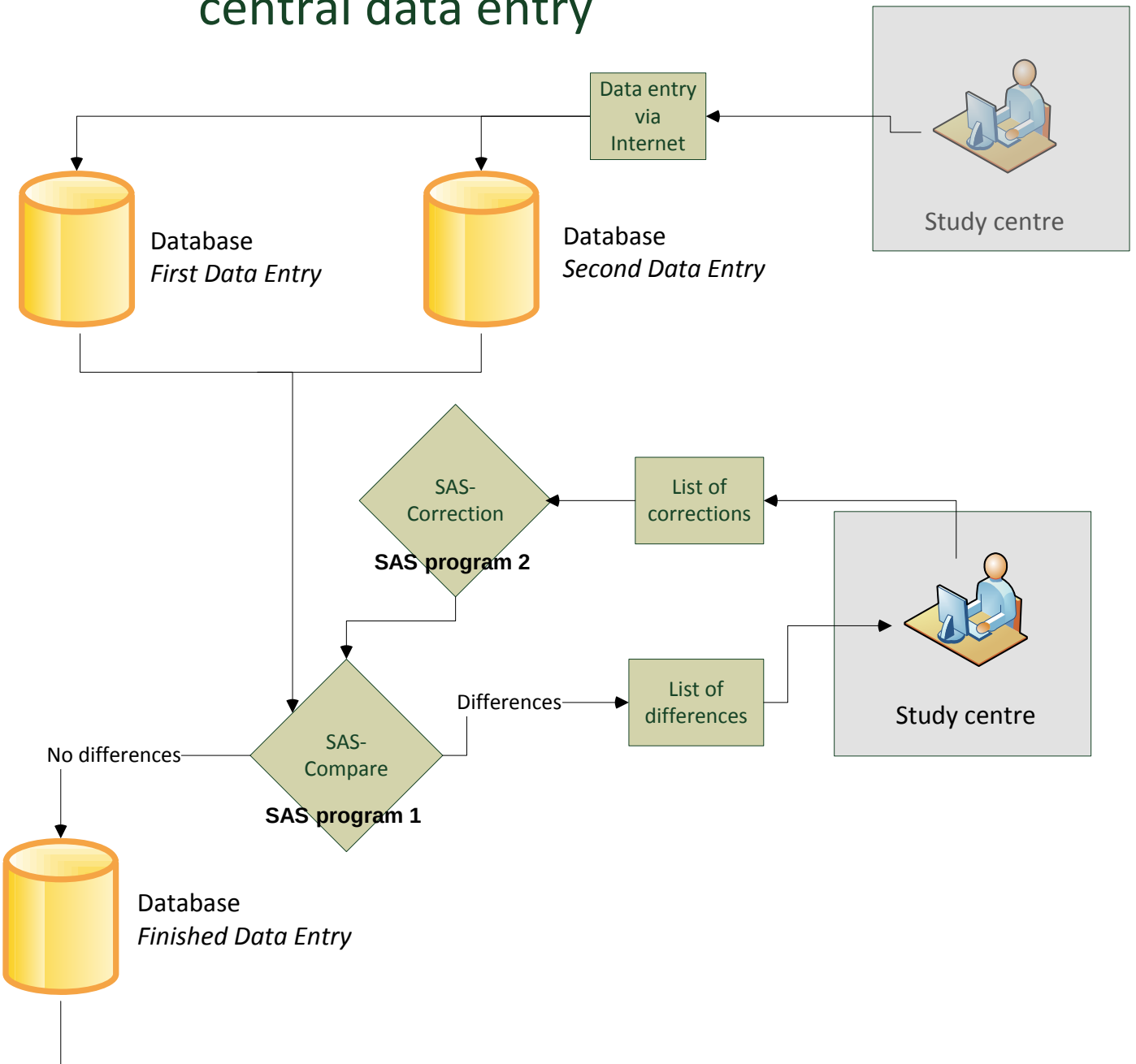
Korrektur-Datenbank für die Korrektur von Unterschieden zwischen Erst- und Zweiteingabe



Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

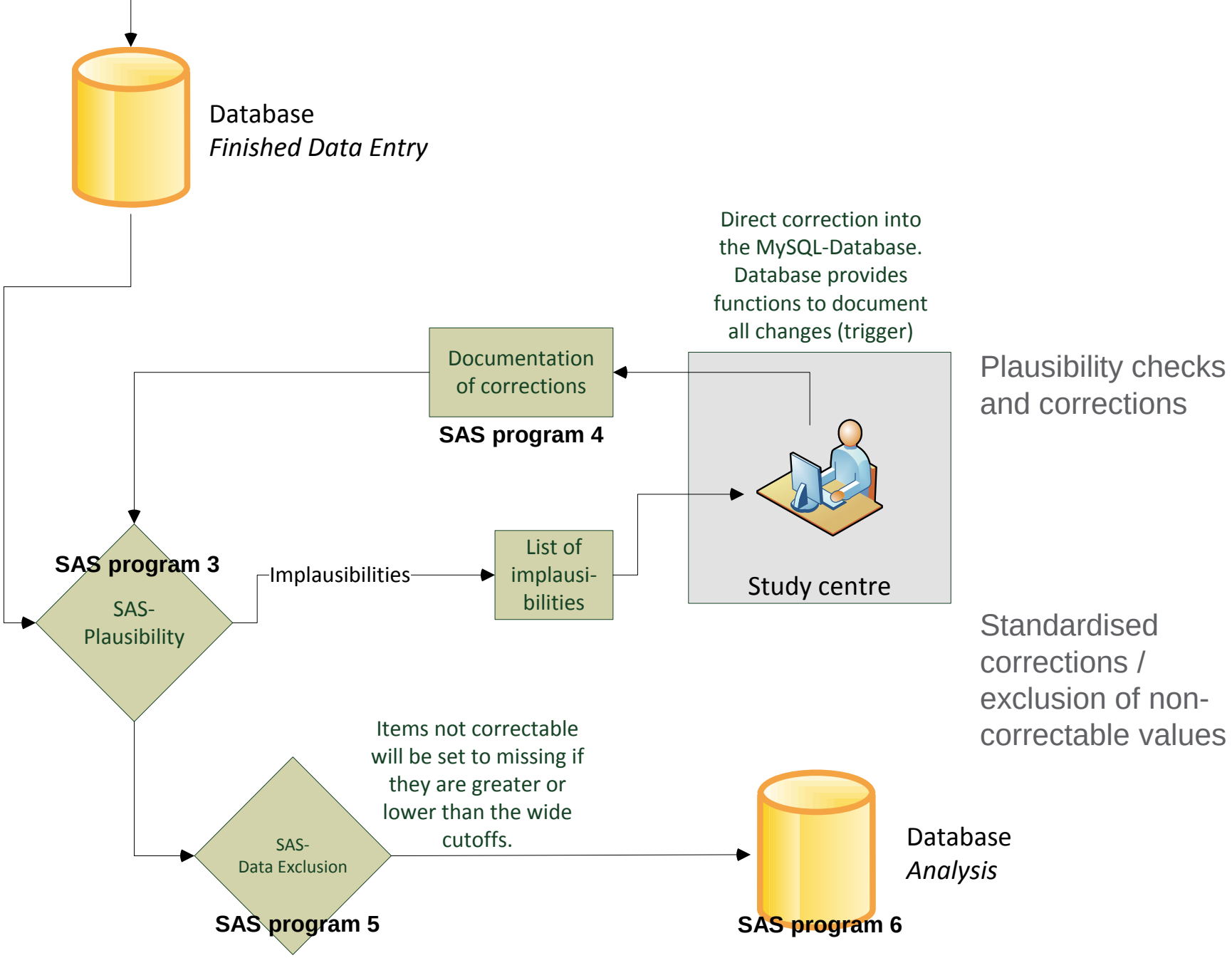
Data entry and plausibility checks in I.Family

central data entry

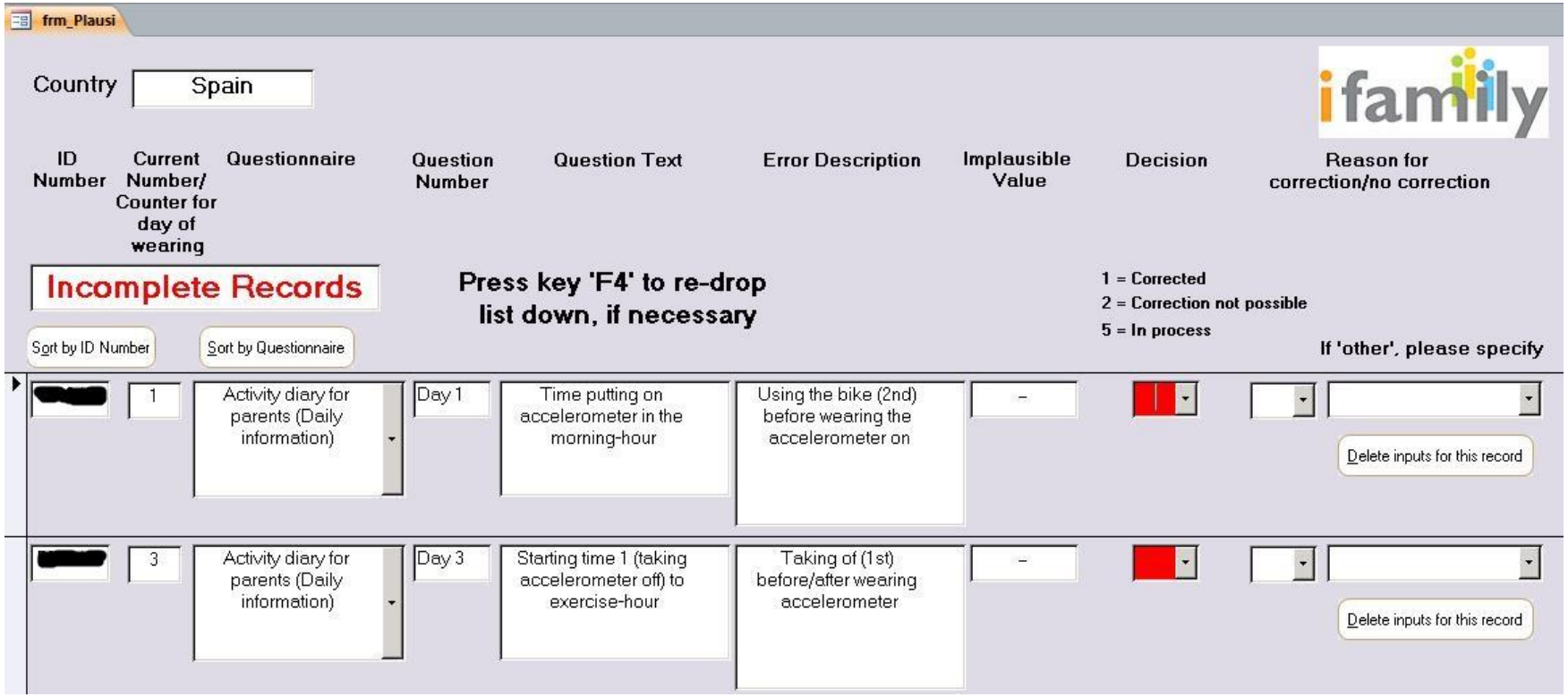


Data entry

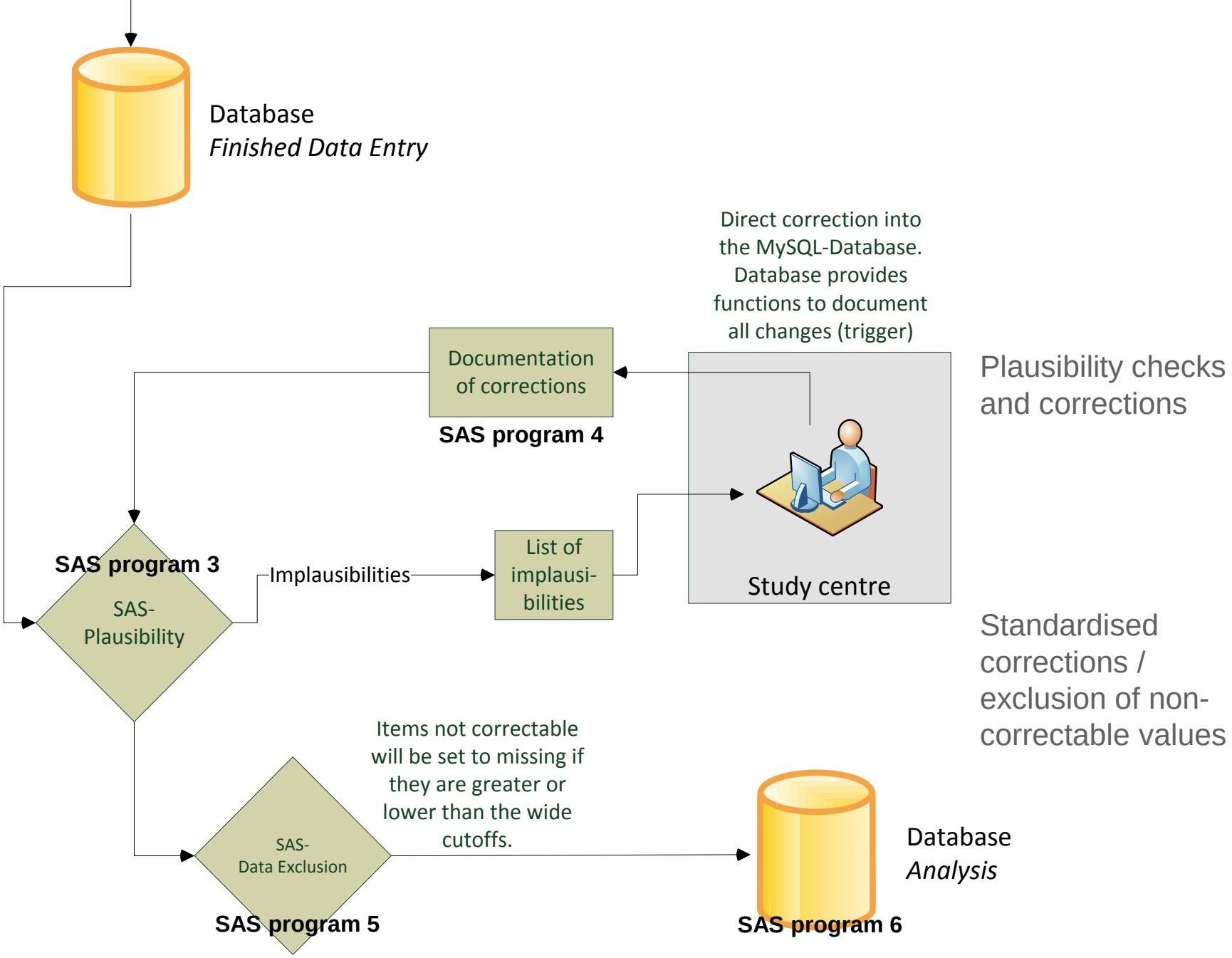
Comparison of 1st and 2nd entry and corrections


Database
Finished Data Entry
Direct correction into the MySQL-Database. Database provides functions to document all changes (trigger)
Documentation of corrections
SAS program 4
Plausibility checks and corrections
SAS program 3
SAS-Plausibility
Implausibilities
List of implausi-bilities
Study centre
Standardised corrections / exclusion of non-correctable values
Items not correctable will be set to missing if they are greater or lower than the wide cutoffs.
SAS-Data Exclusion
SAS program 5
Database
Analysis
SAS program 6

CorrectionDocuDB: Dokumentations-Datenbank für die Behandlung implausibler Werte



Funded by the EC, FP 7, Project No. 266044 - Building on idefics study


Database
Finished Data Entry
Direct correction into the MySQL-Database. Database provides functions to document all changes (trigger)
Documentation of corrections
SAS program 4
Plausibility checks and corrections
SAS program 3
SAS-Plausibility
Implausibilities
List of implausi-bilities
Study centre
Standardised corrections / exclusion of non-correctable values
Items not correctable will be set to missing if they are greater or lower than the wide cutoffs.
SAS-Data Exclusion
SAS program 5
Database
Analysis
SAS program 6

Ablauf der SAS-Programme

- Alle SAS-Programme laufen nach Start automatisch nach einander ab
- Speicherung von Zwischenergebnissen in temporären SAS-Dateien
  - Weiterverarbeitung vom nächsten SAS-Programm
- Import der Textdateien in Access-Korrektur-Datenbanken über programmierte Importfunktion

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Ergebnisse der Automatisierung des Datenbereinigungsprozesses

- Effektive Nutzung von Ressourcen
- Zeitnahe Bereinigung der erhobenen Daten von Anfang an
- Frühzeitige Erkennung von Problemen bei der Datenerhebung bzw. der Dateneingabe
- Schnelle Bereitstellung des ersten bereinigten Datensatzes für die Auswertung

- Ziel: Automatischer Import der Outputs in die entsprechenden Datenbanken

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Noch Fragen?

Noch Fragen?

Noch Fragen?

Noch

Fragen?

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study

Danke

für Ihre

Aufmerksamkeit

Kontakt

Willempje Hummel-Bartenschlager

hummel@bips.uni-bremen.de

Leibniz-Institut
für Präventionsforschung und
Epidemiologie – BIPS GmbH

Funded by the EC, FP 7, Project No. 266044 - Building on idefics study