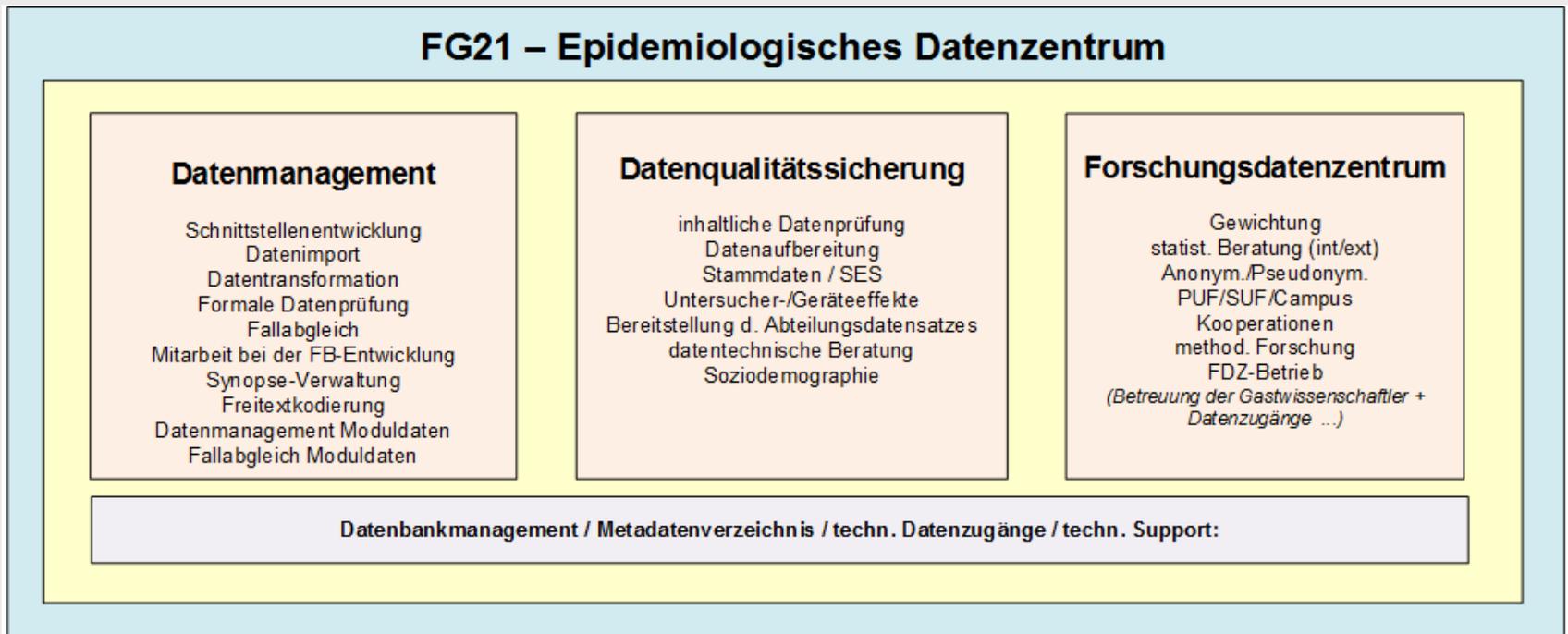


Datenqualitätssicherung bei der Erhebung von Primärdaten

Rüdiger Dölle
Robert Koch Institut

Die aktuelle FG-Struktur

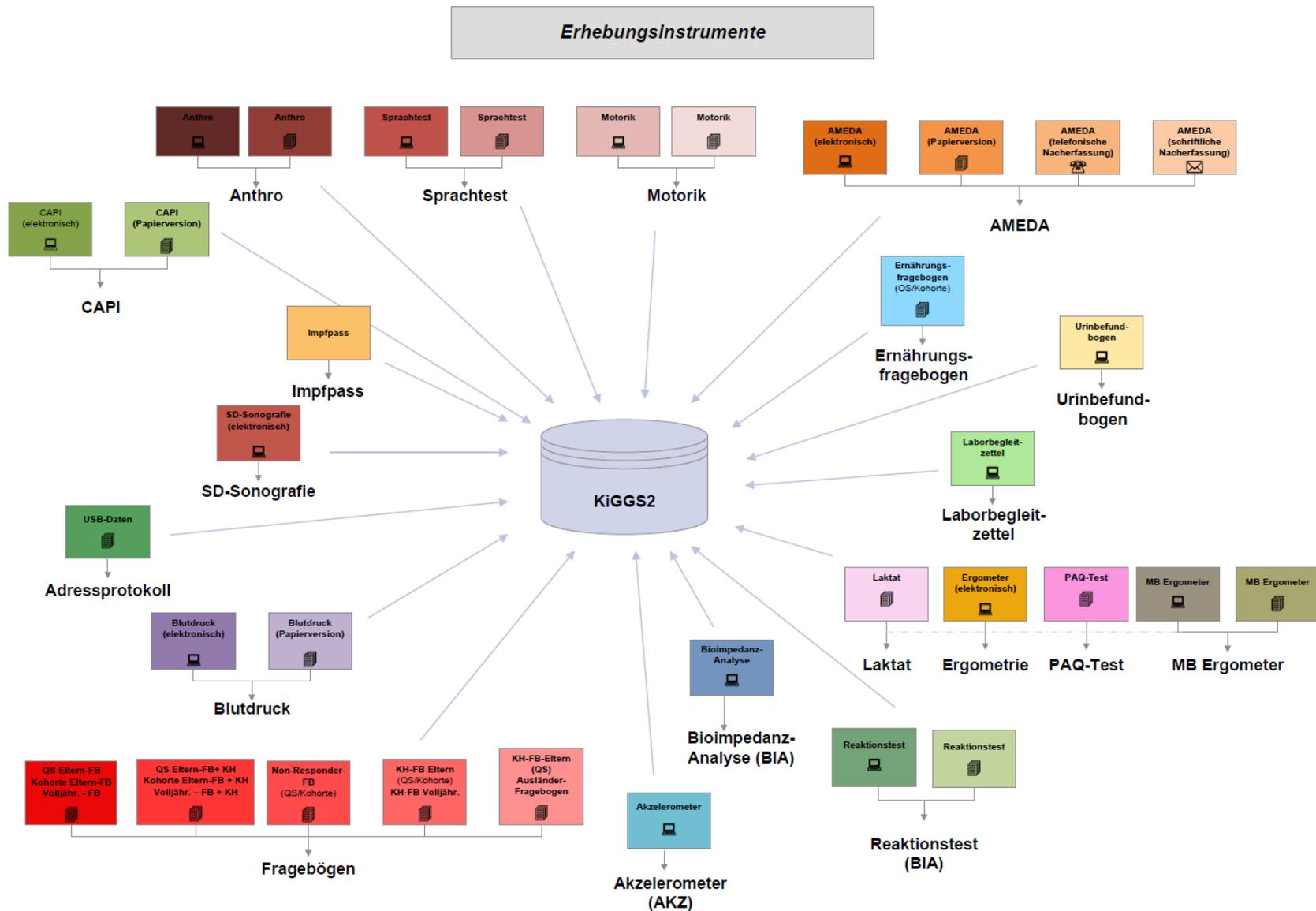


Aufgaben

- Aufbereitung, Prüfung, Standardisierung und Dokumentation der Surveydaten
- Bereitstellung von **validierten Datensätzen** für wissenschaftliche Auswertungen (intern/extern)
- Prüfung, Test, Anwendung und Vermittlung **statistischer Verfahren und Methoden**
- **statistische Beratung** der wissenschaftlichen Mitarbeiter
Unterstützung bei der Auswertung der Daten

Verfügbare Surveydaten der Abt.2

Name	Fallzahl	Verfügbar	Altersbereich	Samp.Points
NUST2 (1990-91)	5.311 (West)	Ja	25 - 69	100
Survey Ost (1992)	2.617 (Ost)	Ja	18 - 79	50
BGS98 (1998-99)	7.124	Ja	18 - 79	120
KiGGS (2003-06)	17.641	Ja	0 - 17	167
GSTel03 (2002-03)	8.318	Ja	18 +	-
GEDA09 (2008-09)	21.262	Ja	18 +	-
GEDA10 (2009-10)	22.050	Ja	18 +	-
GEDA12 (2012-13)	19.294	Ja	18 +	-
DEGS1 (2008-11)	8.152	Ja	18 - 79/91	180
KiGGS1 (2009-12)	16.452	Ja	0 - 23	167
GEDA14 (2014-15)	24016	In Vorbereitung	18 +	301



Stand: 09.12.2014

Erwartungshaltung an d. Datenqualitätssicherung

- Schnelligkeit
- Fehlerfreiheit
- Leicht verständliche Datensätze
- Vergleichbare Datensätze bzgl. der Kurationsmaßnahmen
- Umfassende Dokumentation

Qualität vs. Geschwindigkeit



Heterogene oder gar keine Dokumentation
Keine Standards in der Datenqualitätssicherung
Weniger Genauigkeit
Mehr Inkonsistenzen
Mglw. viele unkodierte Freitexte

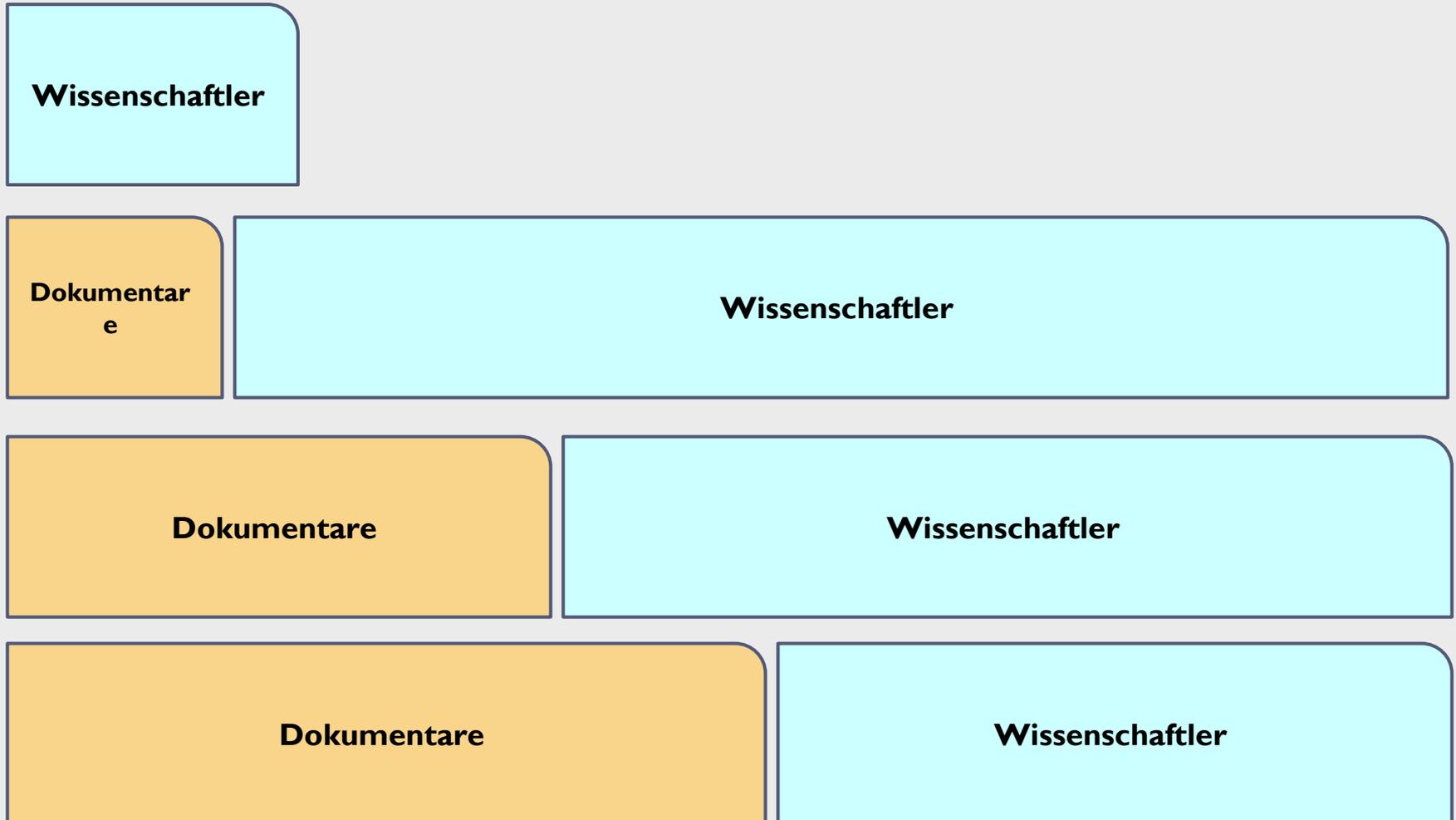
...

**WARUM GELINGT ES UNS NICHT STANDARDS UND
NACHHALTIGE STRUKTUREN FÜR DIE
DATENQUALITÄTSSICHERUNG
ZU SCHAFFEN UND AUSZUTAUSCHEN?**

STRUKTURMAßNAHME:

AUFBAU UND VERSTETIGUNG VON KNOW-HOW

Personalstruktur im Wandel 2003 - 2018



STRUKTURMAßNAHME:

**AUFBAU UND BETRIEB EINER TECHNISCHEN
INFRASTRUKTUR**

Arbeitsaufgaben im Wandel 2003 - 2018

DatenQS
PUF Verwaltung

Softwareentwicklung / Datenbankentwicklung

Verwaltungsdatenbank - Labordatenbank - Datenschnittstelleneentwicklung - Software für die Feldarbeit - diverse Workflowmanagementsysteme - Koop.DB- ErlassDB, PUFDB

DatenQS
PUF Verwaltung

Softwareentwicklung / Datenbankentwicklung

Verwaltungsdatenbank - Labordatenbank -
Datenschnittstelleneentwicklung - Software für die Feldarbeit - diverse
Workflowmanagementsysteme - Koop.DB- ErlassDB, PUFDB

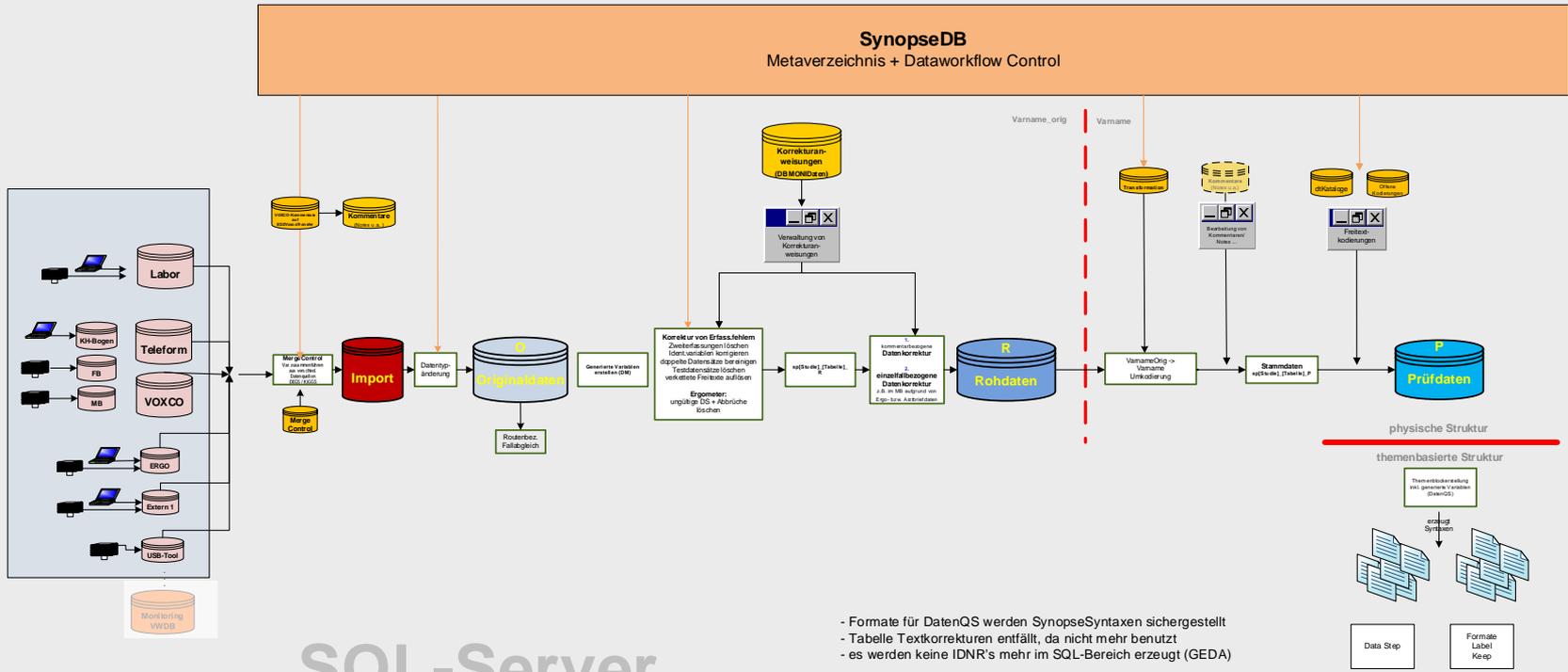
DatenQS
PUF Verwaltung

Datenbankentwicklung
Datenbankmanagement

DatenQS
PUF Verwaltung

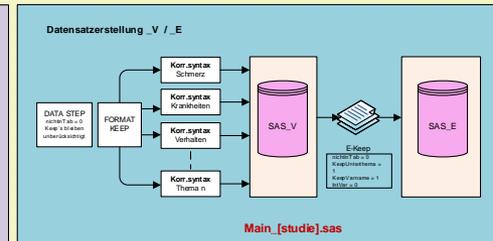
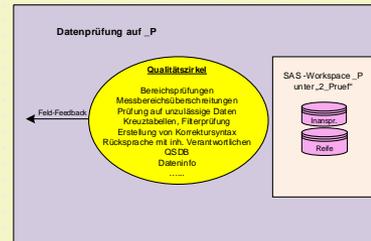
... mit einem Maximalaufwand für die DatenQS (teilautomatisiert)

Survey – Dataworkflow (RDO 23.01.2015)



SQL-Server

SAS



Warum gelingt es uns nicht Verfahren, Tools und Regelwerke zu standardisieren und auszutauschen ?

Die Anforderungen kommen vordergründig aus den eigenen Reihen

- Der QS-Mitarbeiter: Was kann ich dem Datennutzer überhaupt zutrauen, ich muss für ihn mitdenken, Ziel ist eine maximale Fehlerfreiheit des Datensatzes - > 80/20 Regel
- Der QS-Mitarbeiter: Ich bin verantwortlich für die Qualität des Datensatzes und meine Reputation soll nicht leiden?
- Der Fachwissenschaftler: Unsere Reputation könnte leiden.
- Der Fachwissenschaftler: Ich muss mit der Fachwelt vergleichbar sein
- Der Datenmanager: Never change a running system

Interne Probleme

- Standardisierung der Studienkerninhalte -> weniger Komplexität für mehr Schnelligkeit
Prüf-und Korrektursyntaxen können wiederverwendet werden
- Reduzierung der Freitextangaben
- Einbeziehung der QS-Info's der Vorstudien im Design-prozess neuer Studien
- komplexe Filterstrukturen vermeiden oder vereinfachen
- Vermeidung von Änderungen in den Variablenausprägungen

Warum gelingt es uns nicht Verfahren, Tools und Regelwerke zu standardisieren und auszutauschen ?

Die Anforderungen kommen vordergründig aus den eigenen Reihen

- Der QS-Mitarbeiter: Was kann ich dem Datennutzer überhaupt zutrauen, ich muss für ihn mitdenken, Ziel ist eine maximale Fehlerfreiheit des Datensatzes - > 80/20 Regel
- Der QS-Mitarbeiter: Ich bin verantwortlich für die Qualität des Datensatzes und meine Reputation soll nicht leiden?
- Der Fachwissenschaftler: Unsere Reputation könnte leiden.
- Der Fachwissenschaftler: Ich muss mit der Fachwelt vergleichbar sein
- Der Datenmanager: Never change a running system

Es besteht Unsicherheit darüber

- was das notwendige Maß für die DatenQS ist - GEP
- wie die akzeptierten (minimalen/maximalen) Standards in der Community sind
- welches Maß an Verantwortung liegt bei dem Datenmanager und welches bei dem Wissenschaftler
- ob man einem externen Anbieter dies bzgl. vertrauen kann
- wie man mit Änderungen von QS-Maßnahmen über die Zeit umgeht

Weitere Hürden

- Entwicklung und Betrieb von proprietären techn. Tools für spezifische Datenformate
- Viele Drittmittelprojekte sind nicht nachhaltig angelegt → Standards spielen kaum eine Rolle, personelle Stabilität ist nicht gegeben, Basiswissen der Mitarbeiter fehlt
- Proprietäre Systeme sind schwer zu übertragen (veraltet, benutzt andere Softwarestandards, bedingen Fachpersonal, müssen teilweise umfassend überarbeitet werden ...)
- Fehlende Netzwerke für den Austausch - Kompetenzzentren
- Hoher Zeitdruck in den Institutionen, kaum Zeit für die kreative Überarbeitung der bestehenden techn. Lösungen

Fragen und Handlungsoptionen

- Braucht jede Forschungseinrichtung seine eigene technische Infrastruktur?
- Muss jede Einrichtung ein Spezialist für die DatenQS sein oder könnten gemeinsam genutzte Strukturen (Kompetenzzentren) einen Mehrwert bieten?
- Welche Aufgaben können vollautomatisch verarbeitet werden? (Eurostat)
- Wir brauchen mehr technische Lösungsangebote
- Wir brauchen mehr Vernetzung
- Der allgemeine Unsicherheit im Umfeld der DatenQS kann nur mit Standards, SOP's und Regelwerken begegnet werden.

Kontakt:

Ruediger Doelle

DoelleR@rki.de

030 / 18 754 3194

Abt.

Methodenforschung und Forschungsinfrastruktur

Fachgebiet

Forschungsdatenmanagement