



Leipziger Forschungszentrum
für Zivilisationserkrankungen

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

Die Infrastruktur zum umfassenden Forschungsdatenmanagement in LIFE

T. Kirsten, A. Kiel, M. Kleinert, R. Speer,
C. Engel, H. Binder, M. Löffler

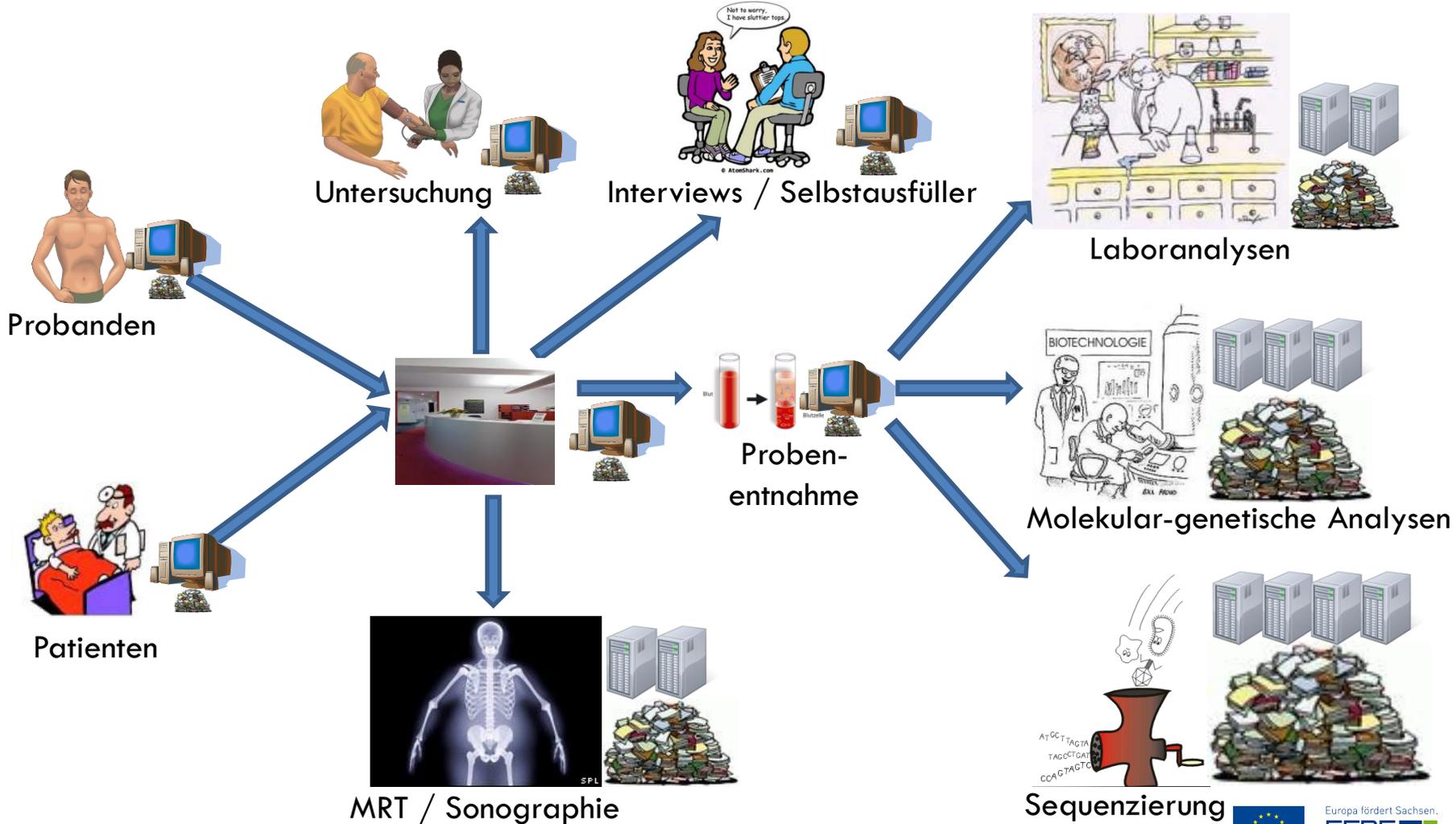
TMF-Workshop: IT in der epidemiologischen Forschung



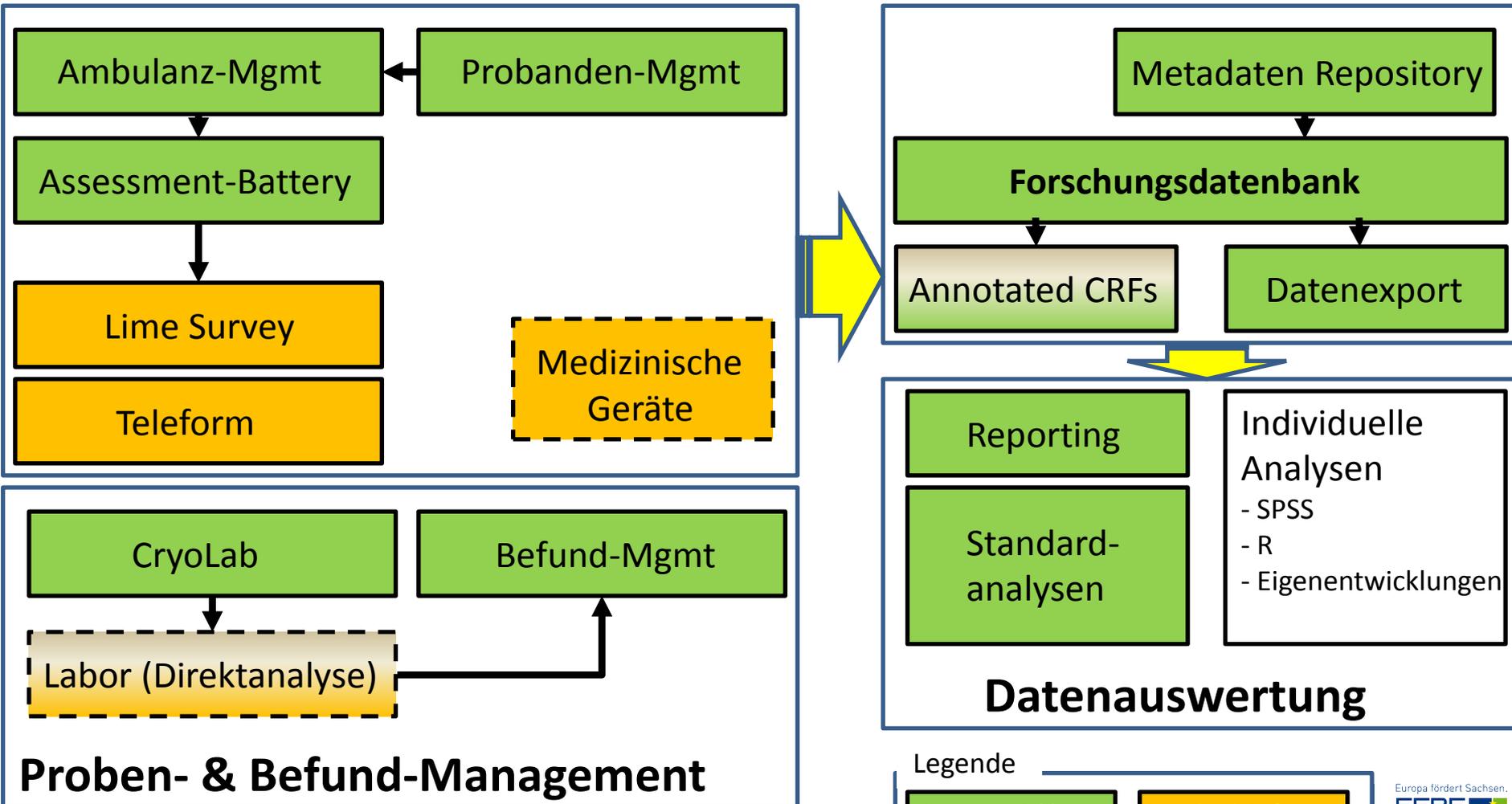
LIFE-Studie

- Leipziger Forschungszentrum für Zivilisationserkrankungen
- Start in 2009 für 5 Jahre
- Gefördert durch Sachsen und EU
- Ziel: Evaluierung klinischer und genetischer Faktoren für ausgewählte Zivilisationserkrankungen
 - Einbeziehung von Lebensstil, soziales & Wohnumfeld
 - Rekrutierung aus der Leipziger Bevölkerung
- Kooperation verschiedener Institutionen

Komplexe Workflows in LIFE



IT-Systeme im Überblick





Ambulanz-Management

- Komplexe Anforderungen
 - Unterschiedliche Kohorten
 - Unterschiedliche Rekrutierungsstrategien
- Einheitliche Pseudonymisierung (zweistufig)
- Einladungs- und Termin-Management
- Umfangreiche und konfigurierbare Untersuchungsprogramme → Laufzettel
- Generierung von Informationsschreiben an Teilnehmer
- Erstellung und Export von Abrechnungsdaten





Datenerhebung

- **Online** Datenerhebung (Lime Survey)
 - Direkte Dateneingabe für ausgewählte Assessments eines Laufzettels
 - Monitoring der Dateneingabe für alle Teilnehmer
 - Datennacherfassung (Laufzettel, Assessments)
- **Offline** Datenerhebung
 - Nutzung von TeleForm (Papier → Scan)
- Diverse Spreadsheets und Desktop-Datenbanken
- Unterschiedliche Geräte: Desktops, Tablets, ...

Datenerhebung mit LimeSurvey

Assessment Battery (T:0.1-SNAPSHOT)

Benutzer: tkirsten Kategorie: SELBSTAUSFÜLLER Logout

Integration der Lime Survey Eingabemaske

Lfd. Nr.	Name des Assessment	Geplante Zeit	Status
2	Befinden während der letzten Woche V1	12 min	
3	Soziales Netzwerk V3	6 min	
4	Ernährung und Getränke V1	8 min	

Liste aller Selbstausfüller

Soziales Netzwerk V3

0% 100%

Soziales Netzwerk 2

In den folgenden 3 Fragen geht es um Ihre Kontakte zu Freunden und Nachbarn.

	9 oder mehr	5 bis 8	3 oder 4	2	1	keine Antwort
Mit wie vielen Freunden oder Nachbarn treffen Sie sich mindestens einmal im Monat oder haben Sie mindestens einmal im Monat sonstigen Kontakt (z. B. telefonisch)?	<input type="radio"/>	<input checked="" type="radio"/>				
Mit wie vielen Freunden oder Nachbarn sind Sie so vertraut, dass Sie sie um Unterstützung bitten können?	<input type="radio"/>	<input checked="" type="radio"/>				
Mit wie vielen Freunden oder Nachbarn...	<input type="radio"/>					

Nächster Teilnehmer



Monitoring der Datenerhebung

Assessment-Battery-Monitor (P:0.2.6) Benutzer: **kirsten**

Aktuelle Laufzettel

Laufzettelfilter (SIC) :

Nr.	SIC	Druckzeit	Status
1	U00598392	02.01.2012 07:12:15	
2	U005987071	02.01.2012 07:15:37	
3	U00598393	02.01.2012 07:41:49	
4	U00598614	02.01.2012 09:22:58	
5	U00598691	02.01.2012 09:23:35	
6	U00594336	02.01.2012 07:07:06	
7	U00598412	02.01.2012 09:39:25	
8	U00612930	02.01.2012 07:36:20	
9	U00478214	02.01.2012 08:10:18	
10	U00453665	02.01.2012 08:11:26	
11	U00185176	02.01.2012 14:17:19	
12	U00597376	02.01.2012 07:27:07	
13	U00596118	02.01.2012 07:52:30	
14	U00598215	02.01.2012 07:54:35	
15	U00421551	02.01.2012 08:12:41	
16	U00598394	02.01.2012 08:20:20	
17	U00592436	02.01.2012 09:59:54	

Assessments für den Laufzettel vom 02.01.2012 09:23:35 des Teilnehmers LI00586591

Assessment-Filter (Nr., DQP, Name, Kategorie) :

Nr.	Name	Status
2	FB_PHQ_parents_self V3	
3	FB_BFI_parents_self V3	
4	FB_FV_parents_self V3	

Liste aller Assessments
pro Laufzettel inkl. Status

Details zum selektierten Laufzettel

SIC	
Teilnehmer	
Druckzeit	
Status	geschlossen

Liste aller Laufzettel

Details zum selektierten Assessment

Nr	3
Kategorie	SELBSTAUSFÜLLER
DQP	DQP_01_0222_20110713
Name	FB_BFI_parents_self V3
Zeit	0 h 5 min
Kommentar	N/A



CryoLab: Das LIFE-LIMS

- Dokumentation der Probenabgabe in den Ambulanzen und Klinken sowie –annahme im Labor
 - Profilgestützte Probenabnahme
 - Etikettierung
- Proben-Tracking
 - Wo befindet sich welche Probe?
 - Welcher Teilnehmer mit welchen Proben?
- Biobank-Management: Dokumentation und Unterstützung der Einlagerung und Auslagerung entsprechend den Lagerungsstrategien

Probenabnahme und -annahme

LIFE CryoLab - Lagerverwaltungs- und Probenverfolgungssystem v0.8 Prod hürsten Abmelden

Navigation

Mandanten

Stammdaten

Ambulanz

Labor

Lager

Blutentnahmen

Suche: Entnahme starten Urinbecher Zeitpunkt 0 Zeitpunkt 1 Zeitpunkt 2 Zeitpunkt 3 Zeitpunkt 4

Ankunft	SIC	Name	Profil	Entnahmenummer	Nüchtern	Urinbecher Rückgabe	Zeitpunkt 0	Zeitpunkt 1	Zeitpunkt 2	Zeitpunkt 3	Zeitpunkt 4
10:02	0080024	Herr Uwe Eitz	A1-oGTT								
09:40	0080003	Herr Florian Jacobi	B4e-Fal-KJP								
08:10	0080029	Frau Max Drumm	B1-O-oGTT								
08:19	0080070	Frau Nicole Dutschke	B1-O-oGTT	B1036F	Ja	11:12	08:34	09:10	09:40	10:10	10:40
09:48	0080080	Frau Nadja Olschew	A2-Mutter-oGTT	A20BRT	Ja	12:44	10:40	11:09	11:40	12:12	12:39
08:37	0080029	Herr Sven Schöler	A2-A-1J	A20SPZ	Ja	10:21	09:28				
09:14	0080080	Frau Carina Schulz	A2-Mutter-oGTT	A20BQV	Ja	12:07	09:55	10:24	10:55	11:25	11:55
08:11	0080070	Frau Silke Treusch	A2-Mutter-oGTT	A20BPX	Ja	11:25	09:50	09:23	09:53	10:20	10:52
08:21	0080080	Frau Lisa Probst	B1-A2-oGTT	B102HY	Ja	11:29	09:05	09:35	10:05	10:34	11:05
10:07	0080004	Herr Hans Paschel	A1-oGTT	A1156T	Ja	10:46	10:47		1		
09:40	0080080	Frau Kathi Schöler	A2-Etern	A20V9K	Ja	10:21	09:36				
09:03	0080004	Herr Stefan Krause	A1-oGTT	A1157R	Ja	09:45	09:46				
10:15	0080040	Herr Frank Hübner	Initial	B707R7	Nein	10:05	10:00				
08:56	0080080	Frau Julia Lucht	A1-oGTT	A115AL	Ja	09:35	09:37				
09:30	0080080	Herr Volker Jahn	Initial	B707G9	Nein	09:00	08:50				
09:24	0080080	Herr Andreas Jahn	Initial	B707FB	Nein	09:05	09:00				
08:58	0080080	Herr Roger Probst	Initial	B707NF	Ja	08:30	08:45				
07:44	0080080	Herr Tom Rüger	A1-oGTT	A1155V	Ja	08:24	08:27				
07:08	0080080	Herr Sebastian Zehn	A1-oGTT	A1153Z	Ja	07:57	07:59				
07:24	0080080	Frau Franziska Schöler	A1-oGTT	A1154X	Ja	08:03	08:06				
07:18	0080080	Herr Bast Janaszek	A1-oGTT	A11514	Ja	08:06	07:49				

Dokumentation verschiedener
Probenabnahmezeitpunkte

■ abgeschlossen
 ■ Belastungstest starten
 ■ Belastungstest überfällig
 ■ Belastungstest unmöglich



Proben-Tracking

Navigation

- Mandanten
- Stammdaten
- Ambulanz
- Labor
- Lager

Entnahmen

Datum	SIC	Entnahmnr.	Profil
02.01.2012		A1-oGTT	
22.12.2011		A2-A-6M	
22.12.2011		A2-A-6M	
22.12.2011		A1-oGTT	
22.12.2011		A2-Mutter	
22.12.2011		A2-Mutter	
22.12.2011		A2-Etern	
22.12.2011		A1-oGTT	
22.12.2011		A1-oGTT	
22.12.2011		A2-Neo	
22.12.2011		A1-oGTT	
22.12.2011		A2-Etern	
22.12.2011		A2-A-6M	
22.12.2011		A2-Etern	
22.12.2011		A1-oGTT	
22.12.2011		A1-oGTT	
22.12.2011		A1-oGTT	
22.12.2011		A2-Mutter	
22.12.2011		A2-A-6M	
22.12.2011		A2-Etern	
22.12.2011		A2-Etern	
22.12.2011		A2-A-1J	
22.12.2011		A1-oGTT	
22.12.2011		A1-oGTT	
22.12.2011		A2-Etern	
22.12.2011		A2-A-1J	
22.12.2011		A2-AZS-8J	

Details der Entnahme A11CBU

Röhrchen (15)

Materialnummer	Röhrchen	Füllstand
A11CBU-CT	8,0 ml CPT-Vacutainer	0 %
A11CBU-CT	8,0 ml CPT-Vacutainer	0 %
A11CBU-CT	8,0 ml CPT-Vacutainer	0 %
A11CBU-CT	8,0 ml CPT-Vacutainer	0 %
A11CBU-CV	8,2 ml S-Monovette Gerinnung	100 %
A11CBU-ED	9,0 ml S-Monovette Hämatologie DNA	100 %
A11CBU-ET	2,7 ml S-Monovette Hämatologie	100 %
A11CBU-ET	9,0 ml S-Monovette Hämatologie	100 %

Aliquots (40)

Aliquotnummer	Typ	Material	Kücher	Lagerplatz
A11CBU-CP-01	Palette 0,5 ml	Citrat-Plasma	<input type="checkbox"/> KW-04NYE	/KL1/F2/KK-A1/A1-CP-
A11CBU-CP-02	Palette 0,5 ml	Citrat-Plasma	<input type="checkbox"/> KW-04M9V	/KL1/F2/KK-A1/B-CP-
A11CBU-CP-03	Palette 0,5 ml	Citrat-Plasma	<input type="checkbox"/> KB-01ZC5	
A11CBU-CP-04	Palette 0,5 ml	Citrat-Plasma	<input type="checkbox"/> KR-01ZG3	
A11CBU-CP-05	Palette 0,5 ml	Citrat-Plasma	<input type="checkbox"/> KB-01ZC5	
A11CBU-CP-06	Palette 0,5 ml	Citrat-Plasma	<input type="checkbox"/> KR-01ZG3	
A11CBU-EP-01	Palette 0,5 ml	EDTA-Plasma	<input type="checkbox"/> KB-01ZC5	
A11CBU-EP-02	Palette 0,5 ml	EDTA-Plasma	<input type="checkbox"/> KR-01ZG3	

Kücher (10)

Küchernummer	#Stra...	Lagerplatz
KR-01X4X	7	/KL1/F2/KK-A1/A1-UZ-1/K8
KG-01XFW	7	/KL1/F2/KK-A1/B-UZ-1/K8
KB-01ZC5	16	
KR-01ZF5	11	/KL1/F2/KK-A1/B-S0-1/K8
KR-01ZG3	16	
KB-02049	12	/KL1/F2/KK-A1/A1-UZ-2/K8
KB-024SH	12	/KL1/F2/KK-A1/A1-S0-1/K8
KW-04M9V	18	/KL1/F2/KK-A1/B-CP-1/K8

Neueste Kommentare

- 0,6ml. 22.12.2011 14:05:28 Löschen
- 0,7ml. 22.12.2011 14:05:03 Löschen
- Entnahme ca. 80µl, danach Material leer. 22.12.2011 11:50:54 Löschen
- Serum 9,5 ml 22.12.2011 11:14:06 Löschen
- Serum 700 µl, EDTA-Plasma 700 µl 22.12.2011 11:32:28 Löschen
- EDTA-Monovette zu ca. 400µl gefüllt, EDTA-Plasma ca. 150µl abgefüllt in Fluid X als A212H3EP01 21.12.2011 13:32:30 Löschen
- EDTA-Plasma hämolytisch 20.12.2011 11:35:02 Löschen
- Straw A213VAEP08 bitte löschen. 20.12.2011 10:44:05 Löschen
- EDTA-Plasma ca. 300µl 20.12.2011 09:03:34 Löschen
- hämolytisch 20.12.2011 09:03:09 Löschen
- EDTA-Plasma ca. 650µl 20.12.2011 09:02:41 Löschen
- EDTA-Plasma ca. 650µl hämolytisch+ml 15.12.2011 15:23:21 Löschen
- EDTA-Plasma ca. 650µl hämolytisch+ml 15.12.2011 15:23:21 Löschen



Integratives Metadaten Management

- Zentrales Metadaten Repository
- Sammlung und Beschreibung aller Items aller Assessments über alle Eingabesysteme
 - Fragen, Kommentare und Fragentypen
 - Assoziierte Code-Listen
- Problem: Versionen, Varianten, Eingabesysteme wirken einer Datenintegration entgegen
- Harmonisierung und Überwindung der Quellspezifika notwendig

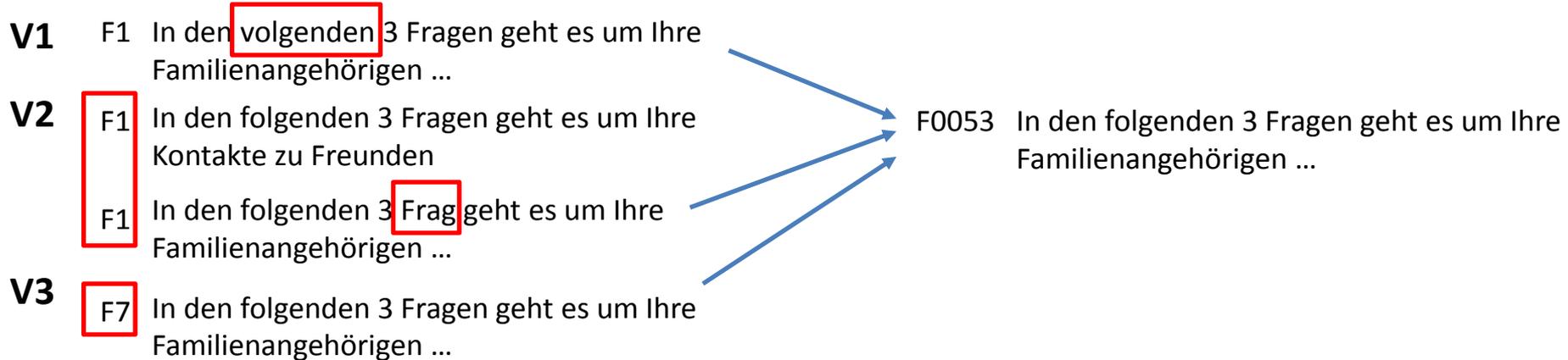


Mapping-basierte Datenintegration

- Manuelle Erstellung des Zielschemas und Herstellung der semantischen Integration zu aufwändig
- Instrumentenspezifische Harmonisierung
- Automatische Ableitung des Zielschemas aus Metadaten der Datenquellen
 - Generische Benennung von Tabellen und Spalten (& Aliasing)
 - Anpassung von Datentypen (Transformationsfunktionen)
 - Übernahme von Code-Listen (Modifikation mgl.)
- Generierung von Mappings zwischen Quellen und Ziel als Basis der täglich ablaufenden ETL-Prozesse

Mappings in LIFE

- Abbildung zwischen den Items



- Abbildung von Code-Listen (Umkodierungen)

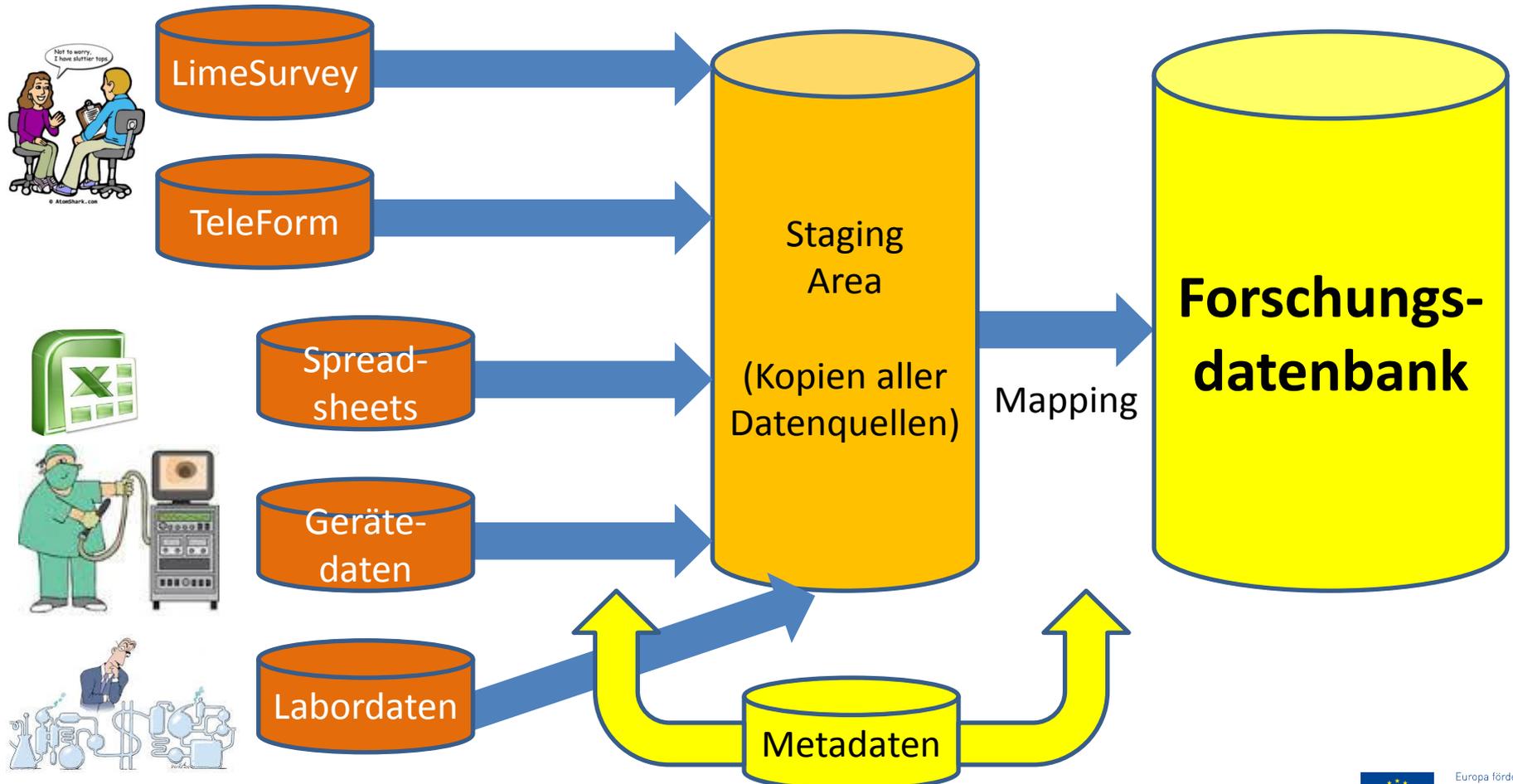
Code-Liste für Frage 4711 (F1, V1)

Code	Value
j	ja
n	nein
98	weiß nicht
99	Antwortverweigerung

Code-Liste für Frage F0053

Code	Value
1	ja
2	nein
98	weiß nicht
99	Antwortverweigerung

Integrationsansatz im Überblick





Aktuelle Mapping-Statistik

- # LimeSurvey: 798 (Ø 76 Items / Instr.)
- # Teleform: 107 (Ø 65 Items / Instr.)
- # Gerätedaten: 9 (Ø 67 Items / Instr.)
- # Labordaten: 79 (Ø 12 Items / Instr.)
- # Forschungsdatenbank: 485 Instrumente
- # Mapping (Version → harm. Instrument): 993
- Ø 2.3 Versionen pro Instrument (max. 9, min. 1)



Kuration von Forschungsdaten

- Qualitätssicherung notwendig für profunde Analysen
- Differenzierung
 - **Syntaktische Fehler:** Mapping-Fehler, ETL-Fehler, ...
 - Automatische Detektion im ETL-Prozess
 - **Semantische Fehler:** Ausreißer, Inkonsistenzen, Duplikate, ...
 - Detektion mit instrumentspezifischen und –übergreifenden Algorithmen
- Manuelle Korrektur der Fehler im Curation-IS
- Keine Kuration in den Quellsysteme sondern in Kopien



Kuration von Forschungsdaten

LIFE Curation DB v0.1.3 Test jwagner Abmelden

Navigation

- Datenkuration
- Transformationsfehler**

Transformationsfehlerkuration

Suchen: Refresh Kurierte anzeigen

Instrument	Fragen-Code	Frage	Wert
Anamnese_Eigen_Säuglin...	6_F1_1_1	1.2 Angaben zu Schwangerschaftswoche, Geburtsgewicht- und Länge - Schwangerschaftswoche -	40+1
Anamnese_Eigen_Säuglin...	6_F1_1_1	1.2 Angaben zu Schwangerschaftswoche, Geburtsgewicht- und Länge - Schwangerschaftswoche -	40+1
Anamnese_Eigen_Säuglin...	6_F1_1_1	1.2 Angaben zu Schwangerschaftswoche, Geburtsgewicht- und Länge - Schwangerschaftswoche -	38+1
Anamnese_Eigen_Säuglin...	6_F1_1_1	1.2 Angaben zu Schwangerschaftswoche, Geburtsgewicht- und Länge - Schwangerschaftswoche -	38+1
Anamnese_Eigen_Säuglin...	6_F1_1_1	1.2 Angaben zu Schwangerschaftswoche, Geburtsgewicht- und Länge - Schwangerschaftswoche -	39+2
Anamnese_Eigen_Säuglin...	6_F1_1_1	1.2 Angaben zu Schwangerschaftswoche, Geburtsgewicht- und Länge - Schwangerschaftswoche -	40+1
Anamnese_Eigen_Säuglin...	6_F1_1_1	1.2 Angaben zu Schwangerschaftswoche, Geburtsgewicht- und Länge - Schwangerschaftswoche -	39+4
Anamnese_Eigen_Säuglin...	6_F1_1_1	1.2 Angaben zu Schwangerschaftswoche, Geburtsgewicht- und Länge - Schwangerschaftswoche -	38+1
Anamnese_Eigen_ü1 V3	12_F5_2_1	1.5 Wenn das gelbe Heft nicht vorliegt anamnestische Erfragung von: - Geburtsmodus -	Sectio
Anamnese_Eigen_ü1 V3	12_F5_2_1	1.5 Wenn das gelbe Heft nicht vorliegt anamnestische Erfragung von: - Geburtsmodus -	Sectio
Anamnese_Eigen_ü1 V3	541_F127_02_4	11.2 Wie viele Zigaretten pro Tag, wie oft und wo wurde geraucht? - In der Umgebung des Kindes, z.B. auf dem Balkon (Inhaltsstoffe an Kleidung) - täglich	25/d
Anamnese_Eigen_ü1 V3	12_F5_2_1	1.5 Wenn das gelbe Heft nicht vorliegt anamnestische Erfragung von: - Geburtsmodus -	spontan
Anamnese_Eigen_ü1 V3	537_F127_01_4	11.2 Wie viele Zigaretten pro Tag, wie oft und wo wurde geraucht? - In der Wohnung des Kindes - täglich	10-15
Anamnese_Eigen...			HHL
Anamnese_Eigen...			spontan, Vakuum

Transformationsregelfehler

Fehlermeldung

Nicht zu Zahl konvertierbar.

Erzeugt durch Transformationsregel:
Text -> Zahl

Fehlerbezogene Korrektur der erfassten Werte

Alter Wert:

Jeder Wert

Neuer Wert:

Auf missing setzen

Problem:

Lösung:

Kurationshistorie

neuer Wert	alter Wert	erstellt am	Ersteller
------------	------------	-------------	-----------

Liste der aufgetretenen Transformationsfehler

Kurationshistorie

Für die Technikbegeisterten ...

- Server
 - SOFS-Cluster
 - TSM-Server für Backup der Daten
 - Domain-Controller
 - Terminal-Server
 - Datenbank-Server
 - ESX-Cluster (Virtualisierungslösung)
- Speicher
 - 100 TB Plattenspeicher (IBM DS5300)
 - 150 TB Bandspeicher (IBM TS3500 Library)





Zusammenfassung

- Hoher Unterstützungsgrad der Ambulanzen beim
 - Probanden- und Durchlauf-Management
 - Datenerfassungsprozessen (online, offline)
 - Proben-Management
- Hohe Anzahl an Erfassungsinstrumenten
- Integratives Forschungsdaten-Management
 - Mapping-basierte Integration relevanter Daten aus Erfassungssystemen in übergreifende Forschungsdatenbank
 - Umfassendes Konzept zur Detektion und Kuration von Forschungsdaten



Besuchen Sie unsere Poster ...

1st International LIFE Symposium

- The comprehensive IT-Infrastructure for epidemiological Research  425
- Mapping-based Integration of heterogeneous scientific Data  424
- CurationDB: Curating Scientific Data in LIFE  428
- Retrieval and Export of Scientific Data using the LIFE Investigation Ontology (LIO)  420
- CryoLab: The LIMS of LIFE  429
- Daten-Management in LIFE  419

**Vielen
Dank**