

# Das Nationale Metadata Repository (MDR)

**Standardisierte Datenelemente zur Dokumentation in klinischen Studien und epidemiologischen Registern**

Matthias Löbe

Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE)

DGEpi-Jahrestagung 2013, Leipzig

25.9.2013

**imise.**

**Institut für Medizinische Informatik, Statistik und Epidemiologie**

# Motivation

- Zentrum für Klinische Studien Leipzig, IMISE, Integrierte Forschungs- und Behandlungszentren
  - 100 laufende klinische Studien, hauptsächlich akademische Forschung
  - jährlich 20 neue Studien mit 5-30 CRFs, ca. 200-1000 Datenelemente
- LIFE-Forschungszentrum für Zivilisationserkrankungen
  - Große epidemiologische Kohorte (26.000 Probanden + Patienten)
  - Befragungen, Selbstausfüller, Messungen mit Geräten, Laborbestimmungen, ca. 17.000 Datenelemente (und steigend)
- Großer Aufwand bei der Spezifikation und Wartung des Dokumentationskonzeptes/ der enthaltenen Datenelemente
- Immer wieder neue Mitarbeiter erschaffen Formulare unterschiedlicher Qualität

# Wie entstehen CRF-Bögen heute?

- In diesem Bild ist ein Verbesserungsbedarf versteckt!

Vorbekannte Tumorerkrankungen, ggfs mit ICD-10: Suchmöglichkeit ICD-10 C00-D48

Chemotherapie in Anamnese: J/N/weiß nicht

Strahlentherapie in Anamnese: J/N/weiß nicht

[Sonstige Tumorbehandlungen: J/N, bei J Freitext zum Spezifizieren **!zunächst rauslassen!**]

f Relevante Begleiterkrankungen: s. Datei „relevante Begleiterkrankungen“ (oder *relevantes Organ* → *Spezifität*)

[Tumor-Histologie n. ICD O.3: Freitext mit genauer Bezeichnung / ~~zunächst rauslassen!~~

Tumor-Grading: G 1-4

↘ Art des Primärtumors

Rezeptorstatis, ~~KBR-Statis~~  
HER2-Statis

[Patholog. Institut + Einsendenummer]???) **zunächst rauslassen**

Klin. Bzw. Pathologischer TNM-Status: Pulldown für T - N - M

# Wo ist das Problem?

- DE nicht detailliert spezifiziert
  - Fehlende Einheiten (Alter)
  - Unterschiedliche Maßeinheiten (Grad Celsius und Fahrenheit)
  - Fehlende Formatierung (Gewicht mit Nachkommastelle?)
  - Unklare Normbereiche („guter Allgemeinzustand“)
  - Verschiedene Messmethoden und -bedingungen (Blutdruck stehend/liegend mit/ohne Belastung)
  - Unterschiedlich geeichte Laborgeräte und Analyseverfahren
  - Unklare Konditionen (Not done)
  - Unterspezifizierung (Gründe für unerwünschte Ereignisse: Therapie)
  - Viel Freitext („Andere“)
  - Unvollständig ausgefüllte Felder (Bedeutet das „Nein“?)




## Wo stört das?

- Studienablauf
  - Häufige Rückfragen (Queries)
- Datenqualität
  - Probleme bei statistischer Auswertung (z.B. unterschiedliche Genauigkeiten von °C und °F oder cm und ft)
- Datenintegration
  - Problem bei Metaanalysen
- Kosten
  - Mehrfache Studien zum gleichen Problem sind teuer
- Ethik
  - Mehrfache Studien zur gleichen medizinischen Fragestellung sind unethisch

# Interoperabilität mit Kerndatensätzen

- Harmonisierte Ausprägung medizinischer Konzepte

## Gemeinsamer einheitlicher Onkologischer Basisdatensatz

ADT und GEKID 

Mit Hilfe des **einheitlichen onkologischen Basisdatensatzes** für alle Krebsarten, der im März 2008 durch den Kooperationsverbund Qualitätssicherung durch Klinische Krebsregister (KoQK) als politisches Gremium verabschiedet wurde, ist in Zusammenarbeit mit den epidemiologischen Krebsregistern, der Deutschen Krebsgesellschaft und Deutschen Krebshilfe, ebenso wie mit den Leistungsträgern und der Politik ein Instrument geschaffen worden, das einen einheitlichen onkologischen Standard vorgibt, Mehrfachdokumentationen verhindert und in allen Bundesländern und klinischen Strukturen eine vergleichbare Erfassung und Auswertung von Krebsbehandlungen ermöglicht.

- ⇒ Der gemeinsame **onkologische Basisdatensatz** der ADT / GEKID wird für jegliche onkologische Dokumentation empfohlen.
- ⇒ Die XML-Umsetzung des Basisdatensatzes zum Herunterladen: [XSD](#) | [PDF](#) | [XLS](#)
- ⇒ **Handbuch** zur Basisdokumentation
- ⇒ [Benutzerhinweise](#) zum Basisdatensatz

### Module zum Basisdatensatz

Der einheitliche onkologische Basisdatensatz wird um weitere Module ergänzt, die in der AG Daten des KoQK entwickelt werden. Die Module **Mamma-** und **Kolorektales Karzinom** wurden auf dem Lenkungsausschuss des KoQK am 30.11.2010 und die Module **Malignes Melanom** und **Lungenkarzinom** am 22.11.2011 verabschiedet.

In Abstimmung mit den Fachgesellschaften werden derzeit die Module zum Prostatakarzinom, zu Gliomen und Lymphomen erarbeitet.

# Bsp: ADT-Datensatz der Tumorzentren

**Version 6.0 Stand 02/08/Basis**  
**DIAGNOSEDATEN**

**Einwilligung zur Meldung an das Klinische und Epidemiologische Krebsregister**
Liegt vor 
Liegt nicht vor 
Abgelehnt

Patientenidentifikationsnummer: 
Krankenkasse:

Telefon: 
Name:

Meldende Institution: 
Geburtsname: 
Geschlecht: w/m

Geschlecht aus PDF-Formular


```

<xs:element name="Geschlecht">
  <xs:simpleType>
    <xs:restriction base="xs:token">
      <xs:pattern value=" [MWU] " />
    </xs:restriction>
  </xs:simpleType>
</xs:element>
<xs:element name="Geburtsdatum" type="Datum_Geburt_Typ"/>
<xs:element name="Strasse" type="xs:string"/>
    
```

Geschlecht aus XML-Schema

Bezeichnung	Codierung	Bedeutung / Erläuterung	Anwendung
<b>Stammdaten</b>		Einmal pro Patient vorzuhaltende Informationen	
Name, Vorname, Geburtsname, Adresse	Geschlecht aus Handbuch	Identifikation, aktuelle Adresse, für EKR eigentlich die Adresse zum Zeitpunkt der Erkrankung zur korrekten Einordnung der Inzidenz	Anschreiben, EKR-Meldung
Geschlecht	M=Männlich W=Weiblich X=Unbekannt		Stratifikation und Selektion von Daten, EKR-Meldung

# Idee: Zentrales Verzeichnis von Datenelementen (MDR)

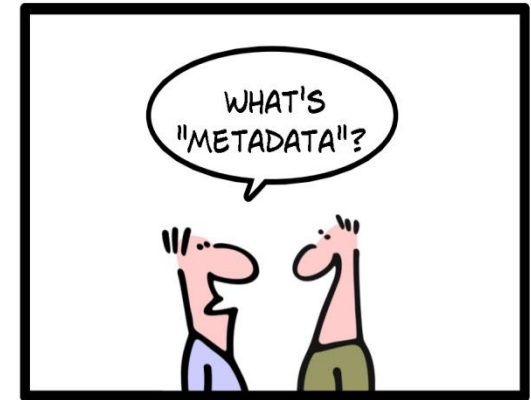
1. Kann man ein **Lexikon** wiederverwendbarer Datenelemente für klinische Studien, aber auch für Register oder Patientenakten anlegen („Biometrie-pedia“)?
2. Kann man so ein System in einem Studienzentrum **effektiv** nutzen?
3. Kann man Datenelemente in einer **Community** austauschen? Kann man „gute“ Datenelemente markieren?
4. Kann man **Versionen**, Varianten, Alternativen und Abbildungen verwalten?
5. Kann man  vorhandenes Wissen aus medizinischen **Terminologien** (ICD-10, OPS, LOINC, MedDRA) nutzen?  
imgBurmlink
6. Kann man damit genormte **Dokumentationskonzepte** für Forschungsprimärdatensammlungen erstellen (DRKD)?
7. Kann man Vergleiche oder **Metaanalysen** von Studien unterstützen?



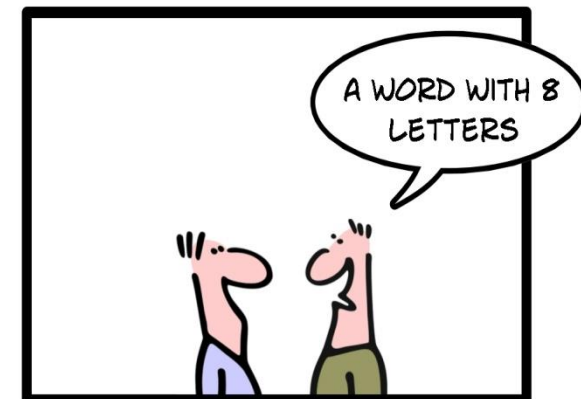
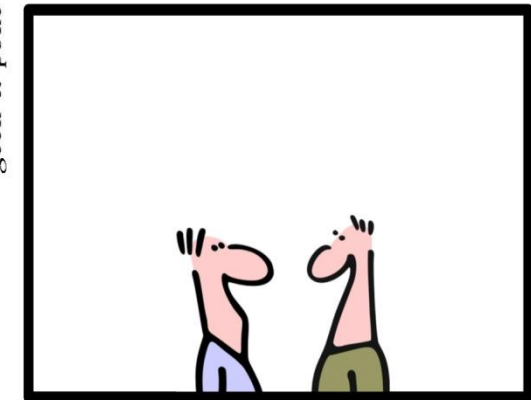
# Simply explained: Metadata Repository (MDR)

- Metadaten:
  - explizite Definitionen von Attributen von Datenelementen (nicht von med. Fakten!)
- Datenelemente:
  - klinische Variablen oder „Items“ wie sie auf CRF-Bögen, in Studiendatenbanken oder statistischen Tools verwendet werden
- Repository:
  - Datenbank für detailliert spezifizierte, qualitativ hochwertige, häufig genutzte, Datenelemente

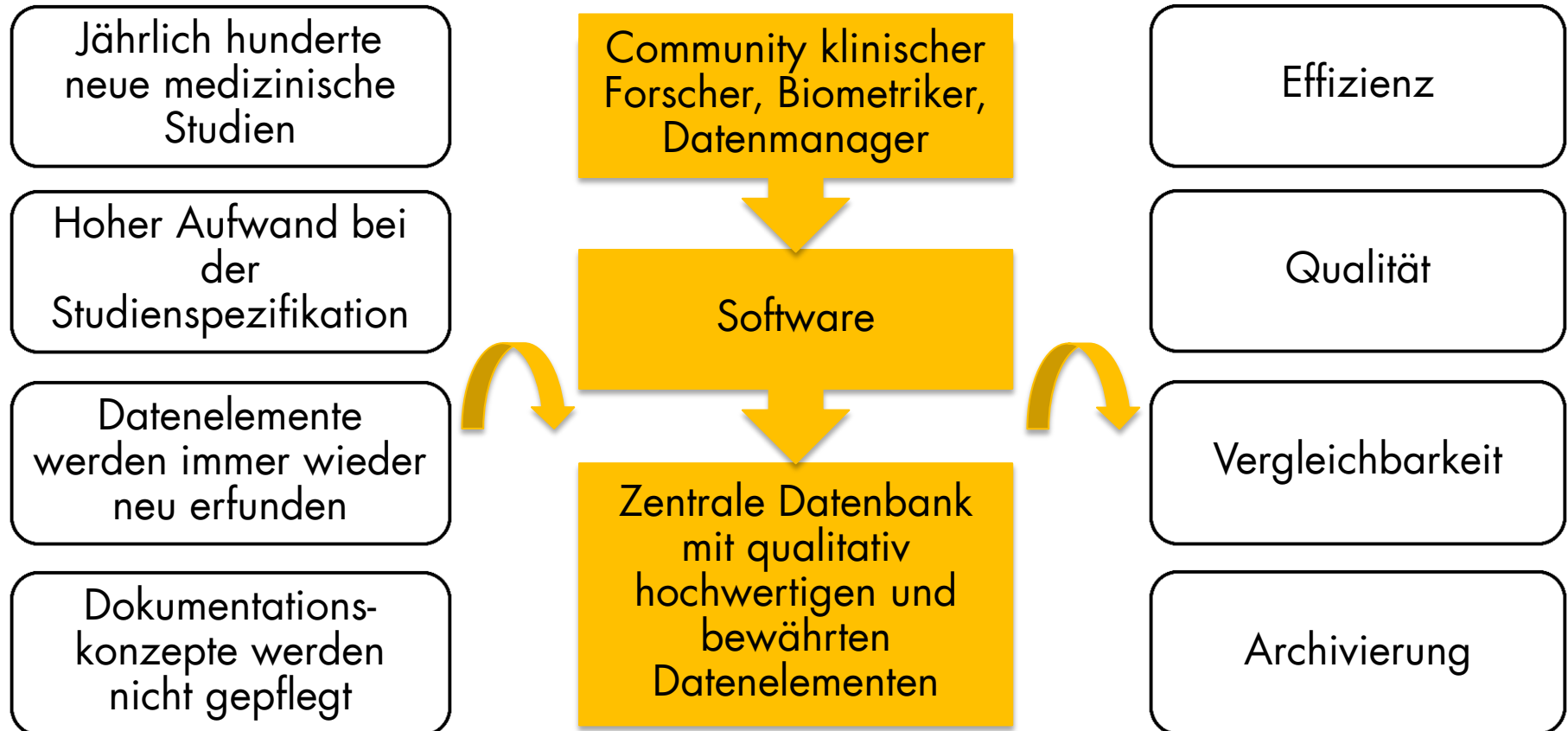
SIMPLY EXPLAINED:  
METADATA



geek & poke

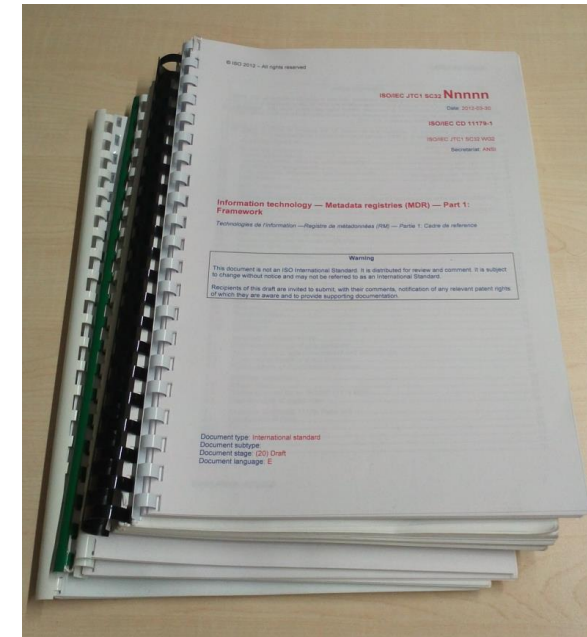


# Ziele des MDR-Projekts

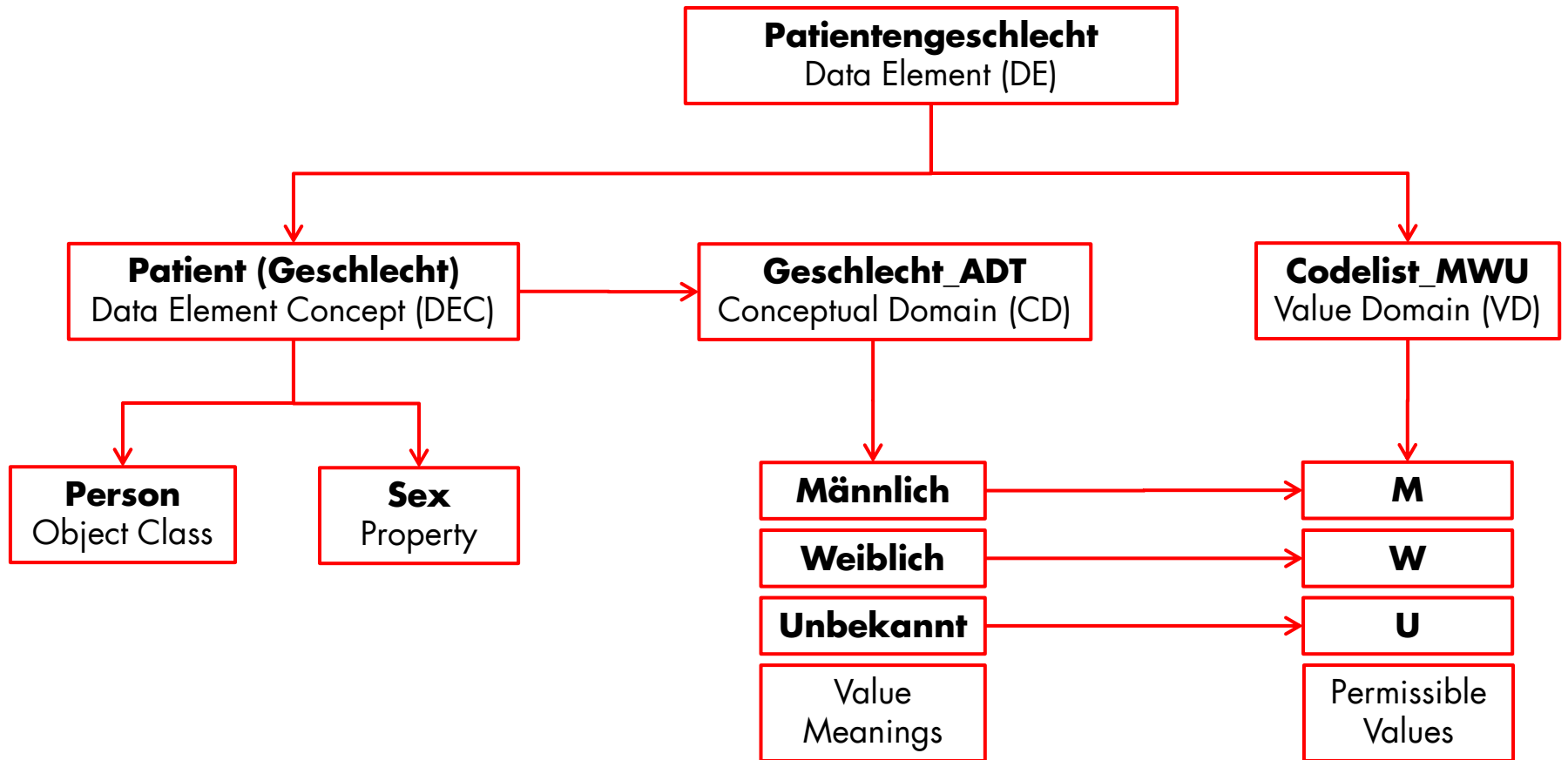


# ISO/IEC 11179 – Metadata Registries (MDR)

- Internationaler Standard zur Modellierung von Metadaten (Datenelementen), 6 Teile
- Unabhängig von der Art, Organisation und Zweck
  - „Beschreibt die Semantik“ der Daten, deren Repräsentation und Registrierung in einem zentralen Verzeichnis
- Sehr komplexer, aber auch recht verbreiteter Standard
  - Edition 2 (2004): etwas angestaubt
  - Edition 3 (2012): in Vorbereitung
- Fokus auf sauberer terminologischer Fundierung



# Beispiel für ein kategoriales Datenelement nach ISO11179



Konzeption  
(Was wird gemessen)

Repräsentation  
(Wie wird gemessen)

{ Object Class + Property = Data Element Concept } + Value Domain = Data Element

# Stand Softwareprototyp

## Web-Applikation mit GWT

The screenshot displays the MDR Design Studio web application interface. The browser address bar shows the URL `mdr.imise.uni-leipzig.de/Workshop/`. The application has a menu bar with 'New' and 'Edit' options, and a toolbar with buttons for 'Save', 'Home', 'Copy', 'Paste', 'Remove', 'Simple', 'ISO11179', 'Browser', 'Preview', 'Source', and 'Version'. The 'Preview' button is currently selected.

The main workspace is divided into three panels:

- Left Panel (Geschlecht):** Shows the configuration for the 'Geschlecht' data element. Fields include:
  - Parent: BGS
  - Language: de-DE
  - Designation: Geschlecht
  - Definition: Geschlecht einer Person
  - Value Domain: Geschlecht<sub>urn:mdr:ValueDomain@65544</sub>
- Middle Panel (Geschlecht):** Shows the configuration for the 'Geschlecht' data element, including:
  - Parent: Value Domains
  - Language: de-DE
  - Designation: Geschlecht
  - Definition: Geschlechtsausprägungen
  - Data Type: Zahl<sub>urn:mdr:DataType@32774</sub>
  - Format: (empty field)
  - Length: 1
  - Unit of Measure: (no value)
  - Members table:

Code	Text
1	Männlich
2	Weiblich
- Right Panel (Data Element Preview):** Shows the preview of the 'Geschlecht' data element. It includes a legend with radio buttons:
  - 1: männlich
  - 2: weiblich

On the far right, there is a sidebar with a search bar containing 'gesc' and a list of related data elements:

- Geschieden<sub>urn:mdr:ValueMeaning@6573</sub>
- Angestellter mit umfassenden Führu...
- Geschlecht<sub>urn:mdr:ValueDomain@65544</sub>
- Geschlecht<sub>urn:mdr:DataElement@65868</sub>
- Geschwister<sub>urn:mdr:DataElement@32844</sub>
- Geschlecht<sub>urn:mdr:DataElement@32839</sub>

# Datenelemente aus CRF

## View Section Data Entry for RE-T0 V1.1 ?

CRF Info

Press the little flag icon beside an input to enter discrepancy notes, please note that you can only save the notes if CRF data entry has already started.

Exit

Fokus (0/16)

Patient...(0/13)

Nebendi...(0/22)

Title: Patientendaten

Page: 4

Alter  \* (Jahre)

Geschlecht  \*

Gewicht  (kg)

Größe  (cm)

Ethnische Zugehörigkeit

Aufnahme im Klinikum (Datum)  \*

Aufnahme auf ITS (Datum)  \*

Aufnahme auf ITS (Uhrzeit)  \* (hh:mm)

Art der Zuweisung

Aufnahme auf ITS von

OpenClinica View Data Entry - Mozilla Firefox

uni-leipzig.de https://www.cscs.imise.uni-leipzig.de/OpenClinica/ViewSectionDataEntry?crfVersionId=81&...

OpenClinica View Data Entry

## View Section Data Entry for RE-T0 V1.1 ?

CRF Info

Press the little flag icon beside an input to enter discrepancy notes, please note that you can only save the notes if CRF data entry has already started.

Exit

Nebendi...(0/22)

Klinisc...(0/37)

Glasgow...(0/4)

-- Select to Jump --

Title: Klinische Daten

Page: 6

Werte des 24h-Zeitintervalls nach dem Beginn schwere Sepsis/septischer Schock

Falls nur ein Wert erhoben wurde, diesen bitte im Feld "min." eintragen.

Körpertemperatur  min.  (°C)

Mittlerer art. Blutdruck  min.  (mmHg)

Herzfrequenz  min.  (1/min)

Atemfrequenz (spontan oder beatmet)  min.  (1/min)

Patientendaten			
Name, Vorname	M107093587_Detlef		
GebDat	10.10.2009	Alter	2
Pat-ID	4545680	Fall	12861759
		Geschl	M

erw. Falldaten	
pfl. OE	KAID1
Fachl. OE	KAID
geplante Entlassung	

Dokumentdaten	
erstellt	Z1WASCH
erstellt am	18.01.2012
zuletzt geände	Z1WASCH
zuletzt geändert am	18.01.2012

- Patientenorganizer
- Laborbefunde
- Fallübersicht

# Datenelemente aus KIS

- aktuelle Anamnese
- aktueller Besuch
- Eigenanamnese
- Familienanamnese
- Medikamente
- Anamnesen

Datum	18.01.2012
Anamnese erfragt durch:	rsthbsyrhb
Anamnese nach Angaben von:	szhrbsrhb

Aktuelle Anamnese:

zshjnsrthrhb  
rtthg

rtah  
bare  
htgb  
rhb

HINWEIS: bei stat. Aufnahme evtl. Infektionen in der Umgebung des Kindes/ Inkubationen

Aktuelle Therapie	
Nahrung	
Medikamente	
Physio-, Ergother. etc.	
Unverträglichkeiten/ Intoleranzen	

Aktuelle Messwerte	
Körpergewicht (kg)	0,00
Körpergröße (cm)	0
Kopfumfang (cm)	0
Sitzhöhe (cm)	0
Halsumfang (cm)	0
Taille (cm)	0
Hüfte (cm)	0
Blutdruck (mmHg)	0 / 0
Herzfrequenz (b/min)	0
Knochenalter (Jahre)	0





# ... werden in das MDR importiert, geglättet und harmonisiert

The screenshot displays the MDR Design Studio interface in Mozilla Firefox. The browser address bar shows <http://www.test1.imise.uni-leipzig.de/>. The interface contains several panels for data elements and a Data Item panel.

The data element panels are:

- RE\_Alter**: ID: RE\_Alterum.mdr.DataElement@33246, Domain: (no value), Meaning: (no value), Precision: 0. Designations: RE\_Alterum.mdr.Designation@490, Alter um.mdr.Designation@491.
- RE\_Gewicht**: ID: RE\_Gewichtum.mdr.DataElement@332, Domain: (no value), Meaning: (no value), Precision: 0. Designations: RE\_Gewichtum.mdr.Designation@494, Gewicht um.mdr.Designation@495.
- RE\_Groesse**: ID: RE\_Groesseum.mdr.DataElement@332, Domain: (no value), Meaning: (no value), Precision: 0. Designations: RE\_Groesseum.mdr.Designation@497, Größe um.mdr.Designation@498.
- RE\_HerzfrequenzMin**: ID: RE\_HerzfrequenzMinum.mdr.DataElement@332, Domain: (no value), Meaning: (no value), Precision: 0. Designations: RE\_HerzfrequenzMinum.mdr.Designation@572, min. um.mdr.Designation@573.
- RE\_HaematokritMin**: ID: RE\_HaematokritMinum.mdr.DataElement@332, Domain: (no value), Meaning: (no value), Precision: 0. Designations: RE\_HaematokritMinum.mdr.Designation@614, min. um.mdr.Designation@615.

The **Data Item** panel shows:

- Parent: CED-Register
- Language: en-US
- Designation: Knochenalter
- Definition:
- Value Domain: VD\_Jahreum.mdr.ValueDomain@...

Colored arrows from the top text point to various fields: red arrows point to the ID and Domain fields of RE\_Alter; purple arrows point to the ID and Domain fields of RE\_Gewicht; orange arrows point to the ID and Domain fields of RE\_Groesse; green arrows point to the ID and Domain fields of RE\_HerzfrequenzMin; blue arrows point to the ID and Domain fields of RE\_HaematokritMin; and a blue arrow points to the Designation field of the Data Item panel.

# Vergleich mit existierenden Clinical Metadata Repositories

Tool	11179	Content	Tools	API	Community	Open Source
caDSR (USA)	V2	✓	✓	✓	🥤	🥤
UK Cancergrid (GB)	V2	?	✓	✓	🚫	✓
METeOR (Australien)	V2	✓	🥤	✓	🚫	🚫
CIHI (Kanada)	V2	✓	Currently not available			
USHIK (USA)	V2	✓	🥤	🚫	🥤	🚫
MDR (Deutschland)	<b>V3</b>	✓★	✓★	✓★	✓★	✓★

Web-based Repository - Software as a Service - Community Approach - Bottom-up-Harmonization

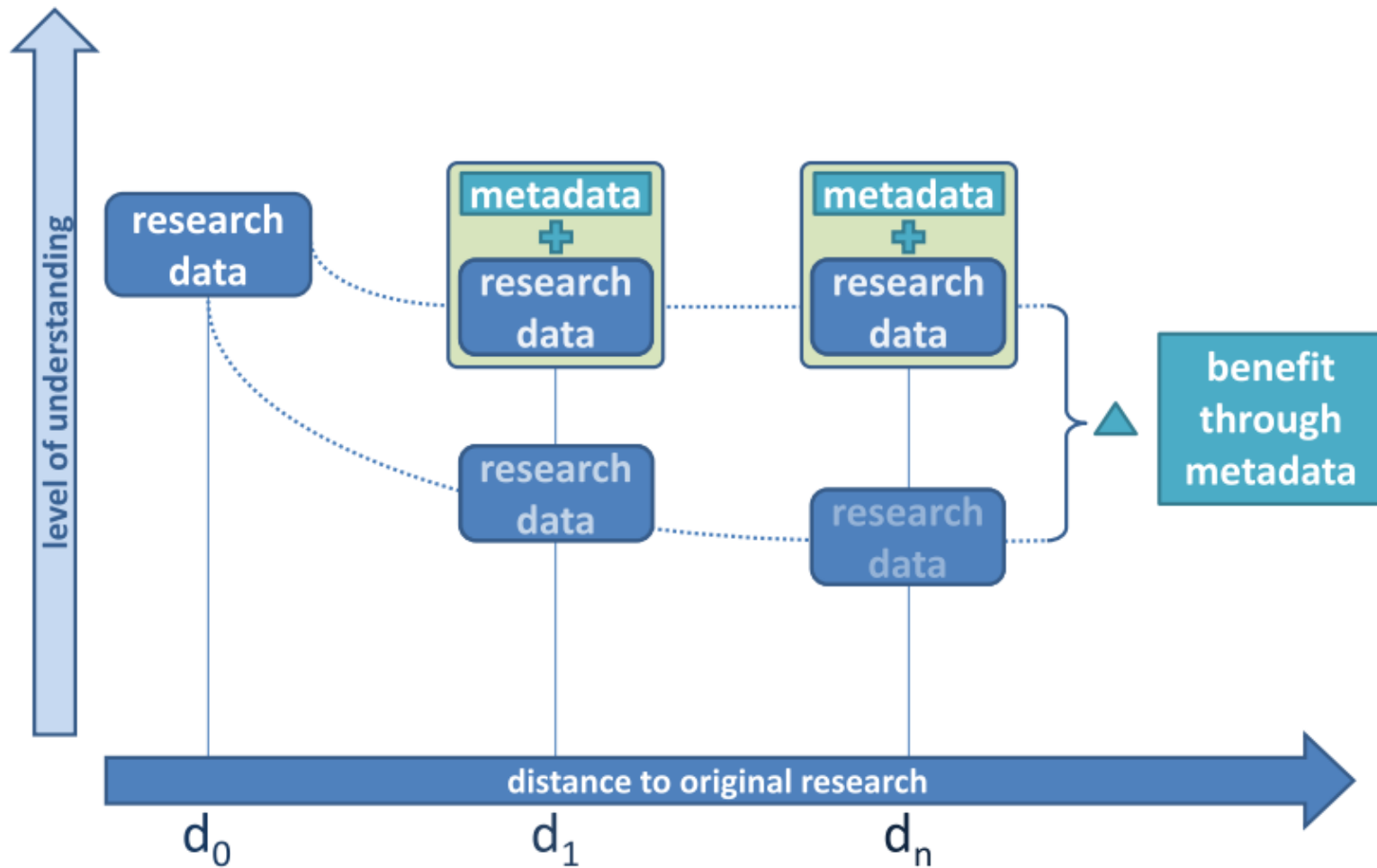
★ Geplant 😊

# Aktuelle Forschungsschwerpunkte

- Nutzer müssen beliebige Sätze von Datenelementen hochladen dürfen, sonst keine Akzeptanz
  - Was passiert mit Dubletten? Welche Optionen zur **Kuration** bestehen bei unter- oder überspezifizierten DE gibt es?
  - Was bedeutet **Ähnlichkeit** von DE?
  - Wie sollen **Änderungen** gehandhabt werden? Welche Auswirkungen hat das Löschen bzw. Kopieren von DE?
  - Wie kann die **Standardisierung und Harmonisierung** unterstützt werden (Kerndatensätze etc.)?
  - Welche Kriterien bestimmen die **Güte** von DEs?
  - Wie sieht eine gute **Benutzeroberfläche** aus?
  - Wie sieht eine gute **Suche nach Varianten** aus? Semantische Suche, Personalisierung, Visualisierungen, Facetten
  - Unter welchen **Lizenzen** sollen Datensätze angeboten werden?
  - Welche **Rechte, Rollen und Sichtbarkeiten** soll es geben?

# Why Metadata

Research data are only interpretable by those who made them



Quelle: Dickmann, Frank and Grütz, Romanus (2011): LABIMI/F - Digital Preservation of Biomedical Research Data, Knowledge Exchange: Workshop Research Data Management – Activities and Challenges, 14.-15. November 2011, Bonn, Poster, Access date: 2012.01.12, URL: [http://www.labimi-f.med.uni-goettingen.de/Publications/Poster\\_LABIMI-F\\_v6.pdf..](http://www.labimi-f.med.uni-goettingen.de/Publications/Poster_LABIMI-F_v6.pdf..)

3