

Werkzeuge zur Unterstützung der Studienspezifikation in OpenClinica

Dr. Frank Meineke
TMF Workshop OpenClinica, Berlin
10.9.2013

Integriertes Forschungs- und
Behandlungszentrum (IFB)
AdipositasErkrankungen

Integrated Research
and Treatment Center
AdiposityDiseases

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

- Unterstützung der klinischen Studien
 - Projektmanagement, Biometrie
 - Datenmanagement
- Aufbau einer Forschungsdatenbank (FDB)
 - Konzept und Aufbau einer sicheren Infrastruktur
 - Einbinden der Quelldaten, Aufbau der Datenbank
 - Bereitstellung von Retrieval Tools
 - Dokumentationskonzept,
 - Basis- und harmonisierter Datensatz
- IFB Adipositas Erkrankungen, Leipzig
 - Unterstützung der Versorgung
 - Erhöhung der Datenqualität, Prozessoptimierung
- IFB Sepsis & Sepsisfolgen, Jena
 - Aufbau Studien-Datenmanagement
 - Einsatz OpenClinica 12.2010-6.2013

IFB CORE UNIT
KLINISCHE STUDIEN /
DATA CENTRE

 Zentrum
für Klinische Studien
Leipzig · KKS

imise.



Unterstützung bei

1. der Erzeugung annotierter CRF
2. der Erzeugung von Validierungsregeln
3. dem Aufbau einer Forschungsdatenbank
 - in eine SQL-Datenbank („DataMart“)
 - nach i2b2 („DataWarehouse“)

1. ANNOTIERTE CRF AUS OPENCLINICA

MOTIVATION

- Der Biometriker benötigt eine Abbildung der CRF zu den Namen der Datenbankfelder und den verwendeten Codelisten

ALTER	AL_GESCHLECHT	AL_GROESSE	AL_AMBERWORBEN	AL_ITSAUFNAHMEART
49	2	164	1	1
68	1	176	0	1
79	1	178	0	3
70	2	162	0	1

- die OC internen OIDs werden einmal ausgelesen
- die Druckanzeige des eCRF im Web-Browser wird über javascript mit Annotationen grafisch angereichert
- Ergebnis kann als PDF gespeichert werden

AL-Baseline V1.0 **ALBAS**

Alter (Jahre)

Geschlecht null keine Angabe
 AL_Geschlecht=1 männlich
 ...=2 weiblich

Gewicht (kg)

Größe (cm)

Bestehende Infektion bei stationärer Aufnahme AL_AmbErworben=1 ja
 ...=0 nein

Aufnahmeart Klinikum null keine Angabe
 AL_ITSAufnahmeart=1 elektiv operativ
 ...=2 elektiv konservativ
 ...=3 Notfall operativ
 ...=4 Notfall konservativ

Unterbringung vor Aufnahme null keine Angabe
 AL_Wohnsituation=1 eigene Wohnung
 ...=2 Altersheim / Pflegeheim
 ...=3 anderes Krankenhaus / Reha
 ...=4 unbekannt

2. VALIDIERUNGSREGELN

MOTIVATION

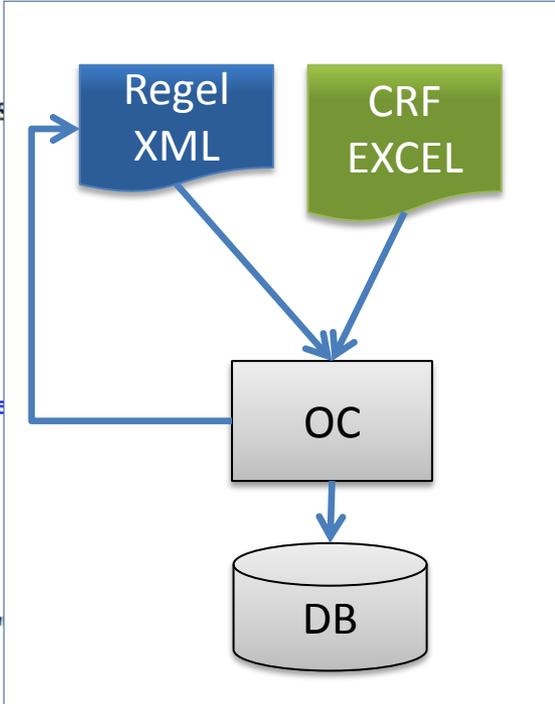
- Fehleingaben sind so früh wie möglich abzufangen
 - a. durch gutes Design (keine Fehlangebe möglich)
Methode: auch dynamische Anzeige (ab. 3.1)
 - b. durch Prüfung vor Speicherung
Abhängigkeiten intra-CRF, zum Teil inter-CRF
Methode: Online Überprüfung
 - c. durch Prüfung nach Speicherung
Prüfung auch komplexer Abhängigkeiten
Methode: nächtliche Batch-Läufe auf der Datenbank, Prüfung in Forschungsdatenbank
- OpenClinica
 - Erstellung komplexer Regeln, auch inter-CRF, ist möglich
 - Unterstützung nur bei Fehlerprüfung und Testumgebung*
 - Manuelle Erstellung problematisch →

* OC RuleDesigner ab OC 3.1. EnterpriseEdition?

2. VALIDIERUNGSREGELN XML-FORMAT

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>  
<RuleImport>  
  <RuleAssignment>  
    <Target Context="OC_RULES_V1">SE_PS_BASELINE.F_PSFM_BASELIN.IG_PSFM_UNGROUPED.I_PS  
    <RuleRef OID="PSFM_INFOPATIENT_D1">  
      <DiscrepancyNoteAction If  
        <Message>"Date of writ  
      </DiscrepancyNoteAction>  
    </RuleRef>  
    <RuleRef OID="PSFM_INFOPATIE  
      <DiscrepancyNoteAction If  
        <Message>"Date of writ  
      </DiscrepancyNoteAction>  
    </RuleRef>  
  </RuleAssignment>  
  <RuleDef Name="PSFM_INFOPATIENT  
    <Description>"Date of writte  
    <Expression>I_PSFM_PSFM_INFOPATIENT_D eq "" and I_PSFM_PSFM_INFOPATIENT_YN eq "1"  
  </RuleDef>  
  <RuleDef Name="PSFM_INFOPATIENT_D2" OID="PSFM_INFOPATIENT_D2">  
    <Description>"Date of written informed consent Patient" should be empty.</Description>  
    <Expression>I_PSFM_PSFM_INFOPATIENT_D ne "" and I_PSFM_PSFM_INFOPATIENT_YN eq "0"</Expression>  
  </RuleDef>  
</RuleImport>
```

ALERTS-Studie des CSCC
> 300 Regeldefinition
> 220 Targets
> 3500 XML-Zeilen



2. VALIDIERUNGSREGELN XML-FORMAT

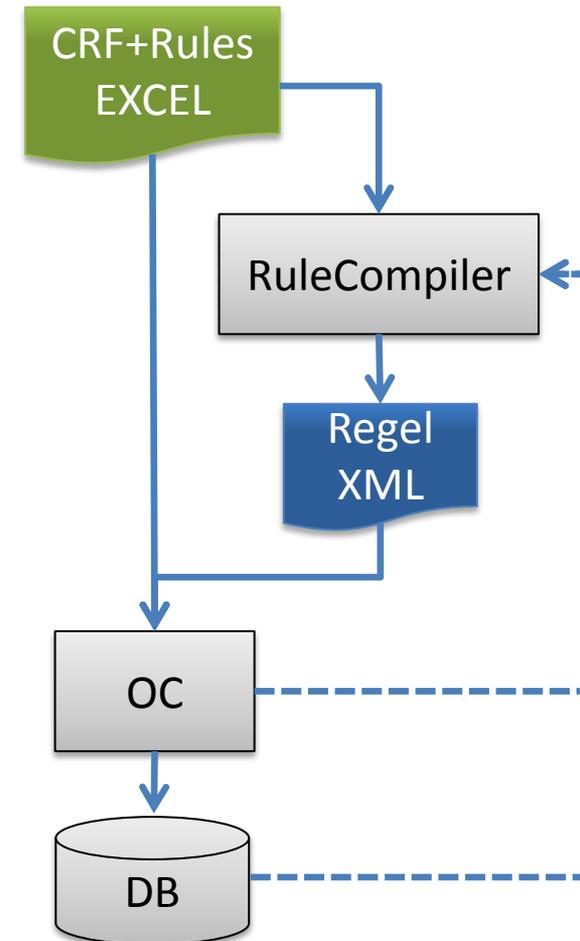
```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<RuleImport>
  <RuleAssignment>
    <Target Context="OC_RULES_V1">SE_PS_BASELINE.F_PSFM_BASELIN.IG_PSFM_UNGROUPED.I_PSFM_PSFM_INFOPATIENT_D</Target>
    <RuleRef OID="PSFM_INFOPATIENT_D1">
      <DiscrepancyNoteAction IfExpressionEvaluates="true">
        <Message>"Date of written informed consent Patient" is required.</Message>
      </DiscrepancyNoteAction>
    </RuleRef>
    <RuleRef OID="PSFM_INFOPATIENT_D2">
      <DiscrepancyNoteAction IfExpressionEvaluates="true">
        <Message>"Date of written informed consent Patient" should be empty.</Message>
      </DiscrepancyNoteAction>
    </RuleRef>
  </RuleAssignment>
  <RuleDef Name="PSFM_INFOPATIENT_D1" OID="PSFM_INFOPATIENT_D1">
    <Description>"Date of written informed consent Patient" is required.</Description>
    <Expression>I_PSFM_PSFM_INFOPATIENT_D eq "" and I_PSFM_PSFM_INFOPATIENT_YN eq "1"</Expression>
  </RuleDef>
  <RuleDef Name="PSFM_INFOPATIENT_D2" OID="PSFM_INFOPATIENT_D2">
    <Description>"Date of written informed consent Patient" should be empty.</Description>
    <Expression>I_PSFM_PSFM_INFOPATIENT_D ne "" and I_PSFM_PSFM_INFOPATIENT_YN eq "0"</Expression>
  </RuleDef>
</RuleImport>
```

2. VALIDIERUNGSREGELN XML-REGELN GENERIEREN

- Spezifikation direkt im Excel



- häufig auftretende Regeltypen definieren
- Meldungen parametrisieren
- Binding zu OC-OIDs zentral
- interne Bezeichner generieren
- Konsistenz prüfen
 - Item in „Items“ definiert?
 - Item innerhalb „Sections“?
 - inter-CRF-Pfade



2. VALIDIERUNGSREGELN REGELDEFINITION IM EXCEL

- Pflicht / Wahl-Regelfelder sind abhängig vom Regeltyp

ITEM_NAME	RULE_TYPE	COMP	VAL	MIN	MAX	RULE_ERROR_MESSAGE
+	Required	o	o	-	-	+
+	Empty	o	o	-	-	+
+	Range	o	o	+	+	+
+	Before	+	-	o	o	+
+	After	+	-	o	o	+
+	Expression	+	-	-	-	+

- Required: Fehler, falls Feld leer ist
- Empty: Fehler, falls Feld nicht leer ist
- Range: Fehler, falls Wert nicht in Range (und COMP=VAL)
- After/Before: Fehler, falls vor/nach (MIN/MAX Tage) COMP liegt
- Expression: Fehler, falls freie Regeldefinition (OC Syntax) zutrifft.

Expression AL_GrundCFSU eq "" and (AL_AWCDC eq "J
FSU" or AL_AWCDC eq "C FSU")

ITEM_NAME	RULE_TYPE	COMP	VAL	MIN	MAX	RULE_ERROR_MESSAGE
PSFM_infoPatient_D	Required	PSFM_infoPatient_YN	eq 1			\$n is required.
	Empty	PSFM_infoPatient_YN	eq 0			\$n should be empty.

2. VALIDIERUNGSREGELN REGELDEFINITION IM EXCEL II

ITEM_NAME	RULE_TYPE	COMP	VAL	MIN	MAX	RULE_ERROR_MESSAGE
RE_Urogenital	Required	RE_Lokalisation	eq 1			Die Lokalisation der Infektion ist bekannt, aber \$n wurde nicht bear
	Empty	RE_AndereLokalisation_YN	ne 1			Wenn andere Lokalisation mit "nein" beantwortet ist, muss das Fe
RE_Alter	Range			18	100	Das erwartete Alter liegt zwischen 18 und 100 Jahren.
RE_Gewicht	Range			40	180	Das erwartete Gewicht liegt zwischen 40 und 180 kg.
	Required	RE_Gewicht_N	ne 1			Bitte Wert für \$n angeben!
	Empty	RE_Gewicht_N	eq 1			Bei "unbekannt" kein Wert für \$n erwartet.
RE_Groesse	Range			140	220	Die erwartete Größe liegt zwischen 140 und 220 cm.
	Required	RE_Groesse_N	ne 1			Bitte Wert für \$n angeben!
	Empty	RE_Groesse_N	eq 1			Bei "unbekannt" kein Wert für \$n erwartet.
RE_Ethnie	Required					Bitte eine Option für \$n auswählen!
RE_ITSAufnahme_D	After	RE_KKHAufnahme_D				Die ITS-Aufnahme kann nicht vor der Klinikum-Aufnahme liegen.
RE_ITSAufnahmeart	Required					Bitte eine Option für \$n auswählen!
RE_ITSAufnahmeort	Required					Bitte eine Option für \$n auswählen!

- RULE_ERROR_MESSAGE
 - lassen sich parametrisieren durch DESCRIPTION_LABEL von ITEM_NAME / COMP
- E-MAIL
 - falls eine E-Mail Adresse eingetragen, wird **zusätzlich** zur DiscrepancyNoteAction auch ein EMailAction erzeugt

2. VALIDIERUNGSREGELN OCRULECOMPILER

- java Programm
 - jdbc (PostgreSQL+Oracle) / apache POI, single jar
 - java -jar ocRuleCompiler.jar <NAME>.xls

```
ocRuleCompiler version 2013-02-05 (C) F. Meineke
usage: ocread
[-utf!latin]      : output xml encoding, default is UTF-8
[-driver xl]      : db jdbc driver, default is oracle.jdbc.driver.OracleDriver
[-user xl]        : db user/scheme
[-password xl]    : db password
[-url xl]         : db connection url, eg. jdbc:oracle:thin:@stargate:1521:xe
[-o <name>.xml]   : output, default is <name>-rules.xml, '-' is stdout
[-update]        : if set, xls is rewritten with OID
[-items]         : only create item.js for annotated CRF
<name>.xls       : OpenClinica Excel Sheet
```

- Liest „Rules“ Tabelle aus OC Excel-Mappe
- Gibt Warnungen / Fehlermeldungen aus
- Bindet optional ITEM OIDs direkt aus OC-DB ein
- Erzeugt „<NAME>-rules.xml“

2. VALIDIERUNGSREGELN OCRULECOMPILER

EVENT_OID	CRF_OID	UNGROUPED				
SE_PS_BASELINE	F_PSFM_BASELIN	IG_PSFM_UNGROUPED				

ITEM_NAME	RULE_TYPE	COMP	VAL	MIN	MAX	RULE_ERROR_MESSAGE
PSFM_infoPatient_D	Required	PSFM_infoPatient_YN	eq 1			\$n is required.
	Empty	PSFM_infoPatient_YN	eq 0			\$n should be empty.

```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<RuleImport>
  <RuleAssignment>
    <Target Context="OC_RULES_V1">SE_PS_BASELINE.F_PSFM_BASELIN.IG_PSFM_UNGROUPED.I_PSFM_PSFM_INFOPATIENT_D</Target>
    <RuleRef OID="PSFM_INFOPATIENT_D1">
      <DiscrepancyNoteAction IfExpressionEvaluates="true">
        <Message>"Date of written informed consent Patient" i
      </DiscrepancyNoteAction>
    </RuleRef>
    <RuleRef OID="PSFM_INFOPATIENT_D2">
      <DiscrepancyNoteAction IfExpressionEvaluates="true">
        <Message>"Date of written informed consent Patient" s
      </DiscrepancyNoteAction>
    </RuleRef>
  </RuleAssignment>
  <RuleDef Name="PSFM_INFOPATIENT_D1" OID="PSFM_INFOPATIENT_D1">
    <Description>"Date of written informed consent Patient" is
    <Expression>I_PSFM_PSFM_INFOPATIENT_D eq "" and I_PSFM_PS
  </RuleDef>
  <RuleDef Name="PSFM_INFOPATIENT_D2" OID="PSFM_INFOPATIENT_D2">
    <Description>"Date of written informed consent Patient" should be empty.</Description>
    <Expression>I_PSFM_PSFM_INFOPATIENT_D ne "" and I_PSFM_PSFM_INFOPATIENT_YN eq "0"</Expression>
  </RuleDef>
</RuleImport>

```

ALERTS-Studie des CSCC

- > 300 Regeldefinition
- > 220 Targets
- > 3500 XML-Zeilen
- > **300 Zeilen Excel**

GRENZEN?

- Allgemeine Beschränkungen durch OC
 - Keine Regeln auf Meta-/Kopfdaten
 - Regeln wie „minimal 2 aus 7“ werden sehr komplex
 - CRF übergreifende Regeln zwischen wiederholbaren CRFs / Events sind kaum möglich
 - Keine Regexp nicht in XML (aber EXCEL) möglich
 - Keine Regelwiederverwendung
- Grenzen ocRuleCompiler
 - Änderungen an OIDs müssen im Excel nachgetragen werden
 - Regel-enthaltendes Excel muss separat versioniert werden

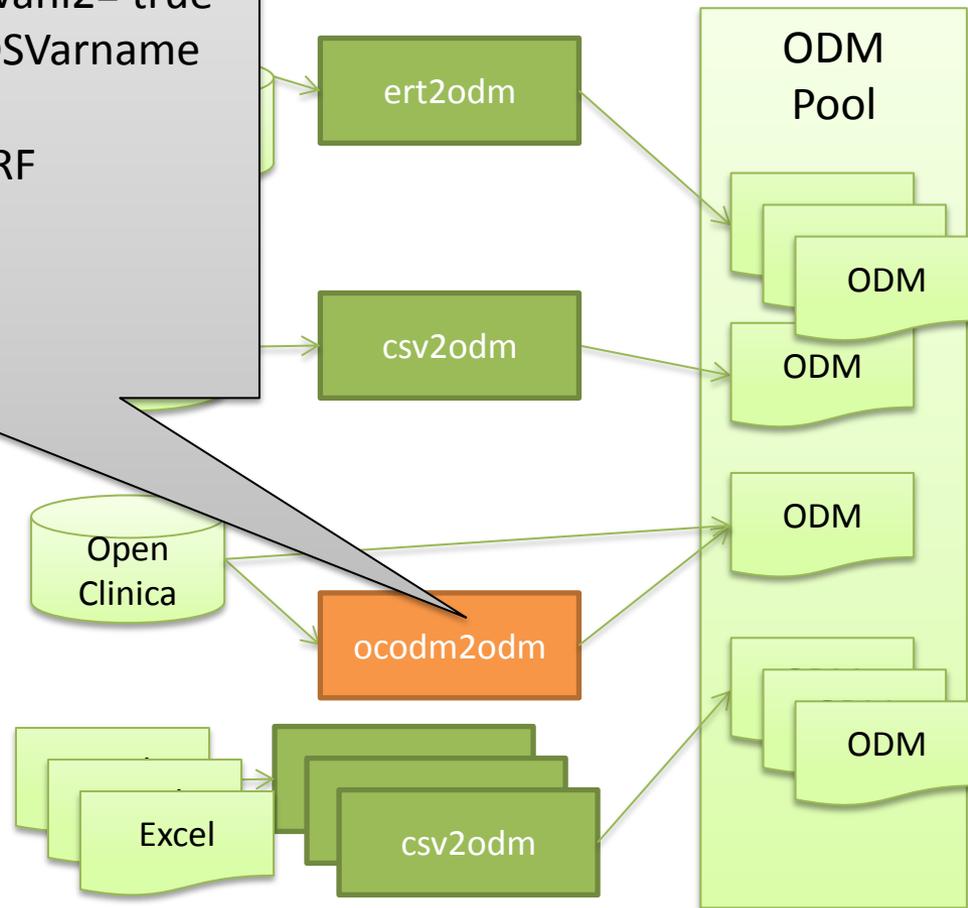
3. FORSCHUNG ETL ÜBER

1. Erzeuge ODM

- langfristige Standards
- Metadaten
- Bereit für die Erhebung verbunden, CSV werden mit MetaDaten angereichert
- Überschaubare ETL-Jobs (Talend Open Studio)

OC ODM Beautifier

- OpenClinica:MultiSelectList in boolesche Einzelitems aufbrechen
Auswahl = „1,4“
→ Auswahl1=true, Auswahl2= true“
- Metadaten zu Items mit SDSVarname taggen
- Ungrouped-Gruppe gem. CRF aufbrechen



Navigate Terms

Find Terms



- [-] Kerndaten
- [-] Klinische Studien
- [-] Patienten
 - [-] FAB
 - [-] FALL
 - [-] ICD - 12783
 - [-] OPS
 - [-] PMD-Adipositas-Erwachsene
 - [-] Adipositas
 - [-] Allgemeine Fragen
 - [-] Anthropometrie
 - [-] BMI (float)
 - [-] Körpergewicht (integer)
 - [-] Körpergröße (integer)
 - [-] Ausbildung und Beruf
 - [-] BIA-Messung
 - [-] Baros Score
 - [-] Dexa-Messung
 - [-] Eigenanamnese
 - [-] Familiäres Umfeld
 - [-] MRT-Abdomen
 - [-] Metabolische Erkrankungen
 - [-] Multimodales Therapieprogramm
 - [-] Patienteneinwilligung
 - [-] Psyche
 - [-] SAP-Kopfdaten

Query Tool



Query Name: PMD-Adi-ICD@18:01:09

Temporal Constraint: Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
PMD-Adipositas-Erwachsen			ICD - 12783					
one or more of these			AND			one or more of these		
			AND			drop a term on here		

Run Query Clear Print Query 2 Groups New Group

Query Status

Finished Query: "PMD-Adi-ICD@18:01:09" [33.9 secs]
 Compute Time: 31 secs

Patient Set for "PMD-Adi-ICD@18:01:09"

Number of patients for "PMD-Adi-ICD@18:01:09"
 patient_count: 28

- annotiertes CRF
 - Vision: nativ im OC (War schon auf Roadmap...)
- ocRuleCompiler
 - wurde zum Standardwerkzeug im CSCC
 - RuleDesigner? Konsequente Hidden-Items?
 - Vision: native Unterstützung im OC?
- Forschungsdatenbank
 - Schnittstellen zum i2b2 und klassischer SQL-Ansicht
 - „besseres“ ODM, z.T. bereits integrierbar
 - funktioniert auch in Oracle

Vielen Dank!

Kontakt: frank.meineke@imise.uni-leipzig.de