



1. European i2b2 Academic User Meeting  
IDRT Hands-On Workshop am 26.03.2013

# TMF-Projekt "Integrated Data Repository Toolkit"

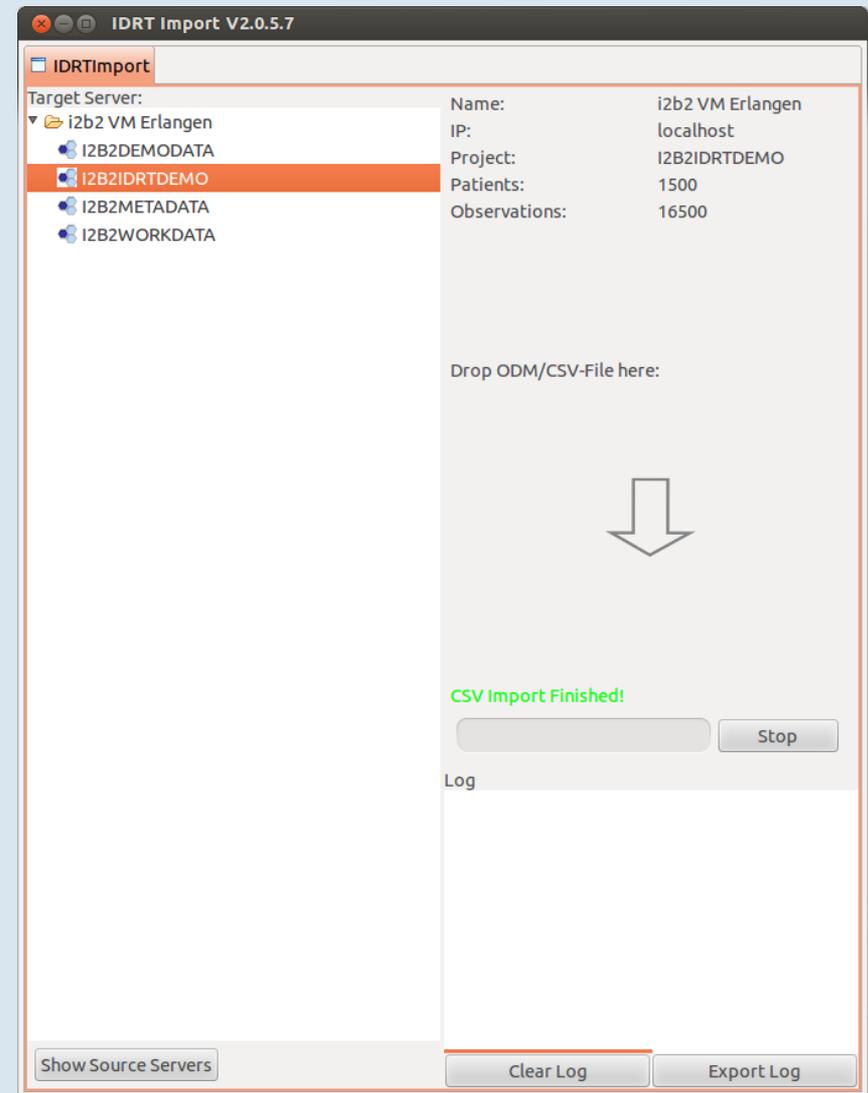
## **IDRT-Import-Tool**

- CSV-Import
- SQL-Import
- ODM-Import

Vortragender:

Benjamin Baum (Universitätsmedizin Göttingen)

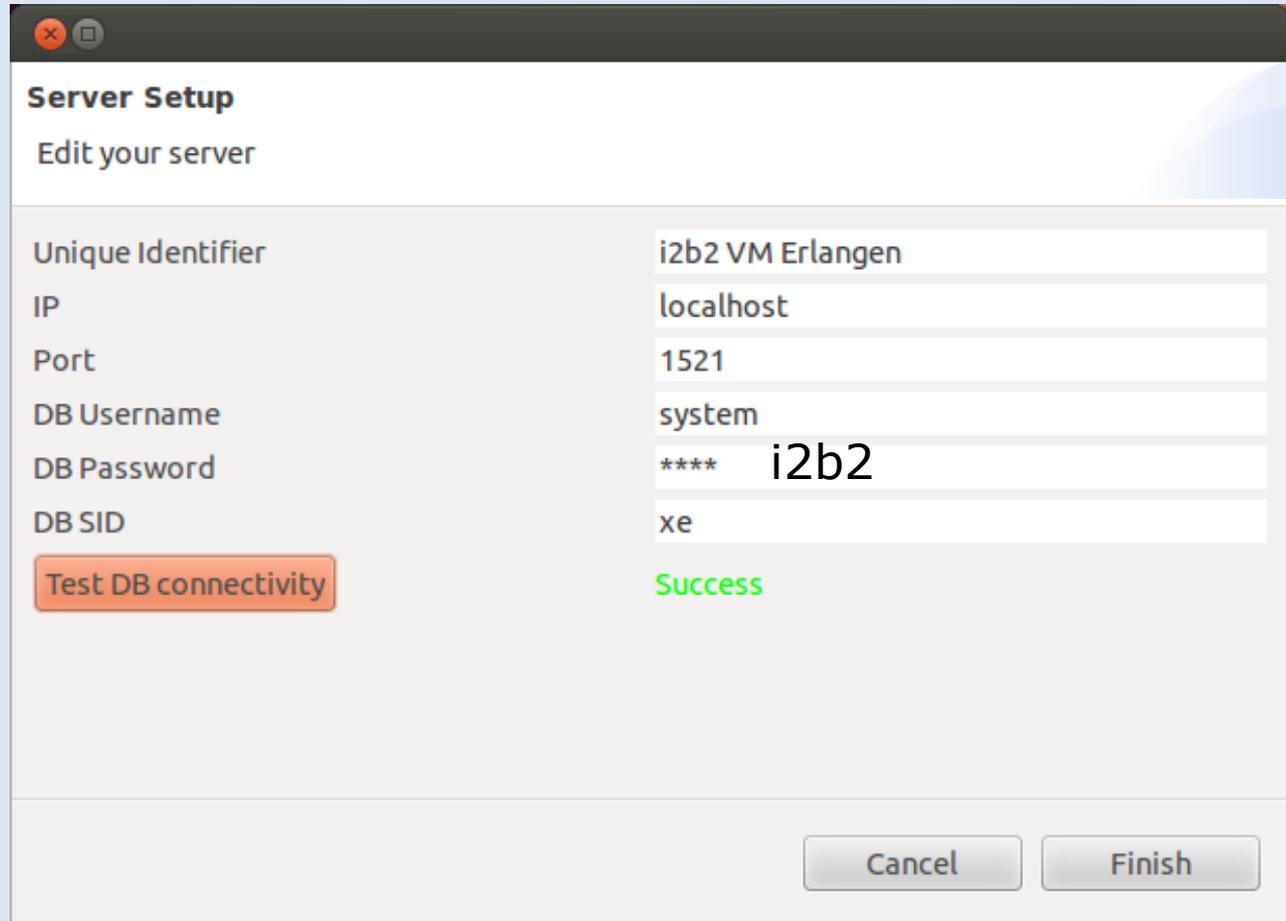
- Ziele
  1. CSV-Daten importieren
  2. Daten aus einer Datenbank (Oracle) importieren
  3. ODM-Daten importieren
  
- Methode
  - IDRT-Import-Tool
    - Java (Eclipse RCP)
    - Benutzt Talend Open Studio Jobs (TOS)
    - Windows/Linux (Mac geplant)



- Übersicht IDRT-Import-Tool

## Server Optionen

- Hinzufügen
- Editieren
- Löschen
- Exportieren
- Importieren

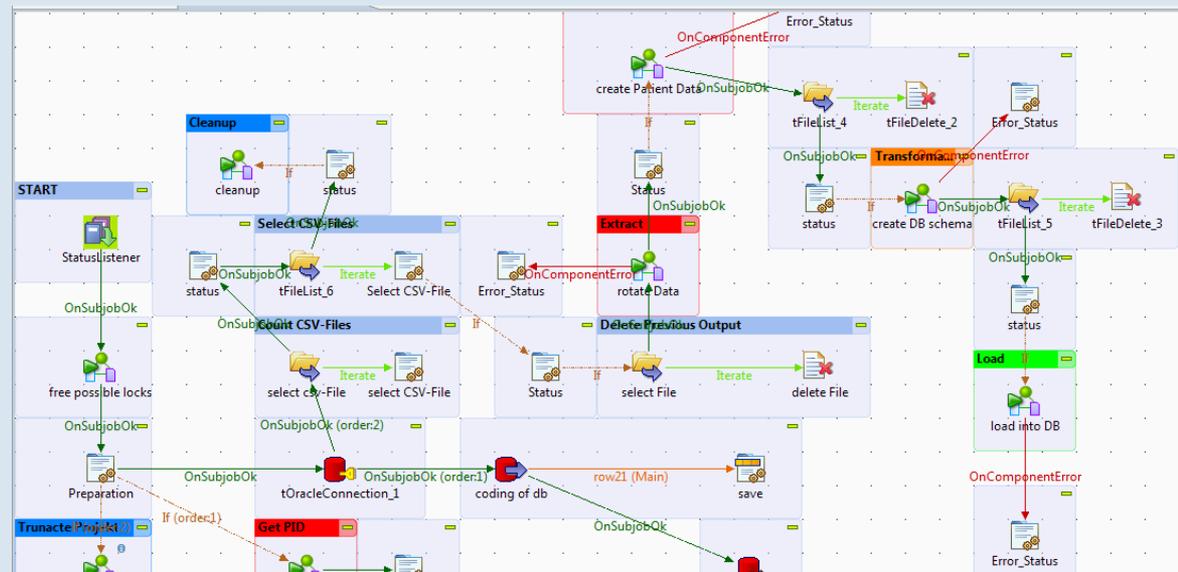


**Server Setup**  
Edit your server

Unique Identifier	i2b2 VM Erlangen
IP	localhost
Port	1521
DB Username	system
DB Password	**** i2b2
DB SID	xe

Success

- Generischer CSV-Import
- 1-n Dateien (sequenzielle Abarbeitung)
- Grundvoraussetzung
  - Komma/Tabulator getrennt
  - Eine Spalte mit Patienten ID (oder Fallnummer, wenn Patient bereits in der i2b2-Datenbank ist)
- Umsetzung in TOS und Java





## ■ IDRT-CSV-PatientenDaten.csv

PID	Geschlecht	Geburtsjahr	Alter	Größe	Gewicht	Downloaddatum	Importdatum	Startdatum	Enddatum
1	f	1959	54	172	124	2012-11-21	2013-03-21	1909-05-04	2011-08-10
2	m	1908	105	163	144	2012-11-21	2013-03-21	1909-05-04	2007-04-13
3	m	1957	56	156	56	2012-11-21	2013-03-21	1910-02-02	2007-04-13
4	f	1940	73	208	44	2012-11-21	2013-03-21	1908-02-05	2006-03-28

## ■ IDRT-CSV-MedizinischeDaten.csv

PID	BMI	BP sys	BP dia	LV Mass	RV Mass	Oxy Sat in %	Fallnummer	Downloaddatum	Importdatum	Startdatum	Enddatum
1	41	118	81	1.44	1.72	98	1001	2012-11-21	2013-03-21	1909-05-04	2011-08-10
2	54	154	81	1.04	1.01	96	1002	2012-11-21	2013-03-21	1909-05-04	2007-04-13
3	23	127	101	1.19	1.34	95	1003	2012-11-21	2013-03-21	1910-02-02	2007-04-13
4	10	147	67	2.00	1.56	90	1004	2012-11-21	2013-03-21	1908-02-05	2006-03-28

- CSV Import durchführen



## Ergebnis

**i2b2 Query & Analysis Tool**      Project: idrtdemo      User: i2b2 User      Find Patients | Analysis Tools | Message

**Navigate Terms**    Find Terms

- Ontology
  - Patientendaten
    - IDRT-CSV-MedizinischeDaten
      - BMI
        - BMI (Numerical Query)
        - BP dia
        - BP sys
        - LV Mass
        - Oxy Sat in %
        - RV Mass
      - IDRT-CSV-PatientenDaten
        - Alter
          - Alter (Numerical Query)
          - Geburtsjahr
          - Geschlecht
            - f
            - m
          - Gewicht
          - Groesse

**Query Tool**

Query Name: \_\_\_\_\_

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 ▾		
<div style="border: 1px solid black; padding: 5px; display: inline-block;">drop a term on here</div>								

Run Query    Clear    Print Query      0 Groups

**Query Status**

- Import baut auf CSV-Import
  - Generisch
  - Sequenziell
  - Patienten ID (oder Fallnummer) obligatorisch
- Beispieltabelle I2B2IDRTDEMO.IDRT\_DB\_IMPORT:

PATIENTID	NIKOTIN	ALKOHOL	ALLERGIEN	MEDIKAMENTE	OPERATIONEN	ERKRANKUNGEN	DOWNLOADDATE	IMPORTDATE	STARTDATE
238	Nein	wöchentlich	Nickel	Xusal	Wechsel von Herzklappe	Darmfunktionsstörung	2012-11-21	2013-26-03	1923-03-12
239	Ja	monatlich	Nein	Omniflora	Operationen an Gallenblase und Gallenwegen	Harnwegsinfektion	2012-11-21	2013-26-03	1910-01-06
240	Nein	wöchentlich	Laktose	Cetirizin	Operationen an Dünn- und Dickdarm	Harnwegsinfektion	2012-11-21	2013-26-03	1923-03-12
241	Ja	täglich	Heuschnupfen	Cotrim	Operationen an Dünn- und Dickdarm	Harnwegsinfektion	2012-11-21	2013-26-03	1911-03-07
242	Nein	monatlich	Nickel	Cotrim	Rekonstruktion des Darmes	Multiple Sklerose	2012-11-21	2013-26-03	1909-05-04
243	Ex-Raucher	monatlich	Laktose	Omniflora	Wechsel von Herzklappe	Nikotinsucht	2012-11-21	2013-26-03	1910-02-02
244	Ja	wöchentlich	Heuschnupfen	Ciprofloxacin	Rekonstruktion des Darmes	Multiple Sklerose	2012-11-21	2013-26-03	1911-03-07
245	Nein	täglich	Nickel	Ciprofloxacin	Operationen an Gallenblase und Gallenwegen	Nikotinsucht	2012-11-21	2013-26-03	1908-02-05
246	Ja	monatlich	Nein	Ciprofloxacin	Operationen an Gallenblase und Gallenwegen	Multiple Sklerose	2012-11-21	2013-26-03	1908-02-05
247	Ex-Raucher	seltener	Laktose	Cetirizin	Operationen an Gallenblase und Gallenwegen	Darmfunktionsstörung	2012-11-21	2013-26-03	1911-03-07
248	Ja	seltener	Heuschnupfen	Cotrim	Operationen an Dünn- und Dickdarm	Morbus Chron	2012-11-21	2013-26-03	1905-11-08
249	Ja	täglich	Heuschnupfen	Cetirizin	Wechsel von Herzklappe	Darmfunktionsstörung	2012-11-21	2013-26-03	1910-01-06
250	Ja	seltener	Nickel	Prednisolon	Wechsel von Herzklappe	Morbus Chron	2012-11-21	2013-26-03	1908-02-05
251	Ex-Raucher	täglich	Heuschnupfen	Azathioprin	Wechsel von Herzklappe	Nikotinsucht	2012-11-21	2013-26-03	1910-01-06
252	Ja	wöchentlich	Heuschnupfen	Prednisolon	Ersatz von Herzklappe	Morbus Chron	2012-11-21	2013-26-03	1911-03-07
253	Ex-Raucher	täglich	Heuschnupfen	Mirena	Ersatz von Herzklappe	Nikotinsucht	2012-11-21	2013-26-03	1911-04-03
254	Ex-Raucher	wöchentlich	Nickel	Ciprofloxacin	Operationen an Dünn- und Dickdarm	Nikotinsucht	2012-11-21	2013-26-03	1910-02-02
255	Ja	täglich	Heuschnupfen	Xusal	Wechsel von Herzklappe	Morbus Chron	2012-11-21	2013-26-03	1905-11-08
256	Ja	wöchentlich	Nein	Prednisolon	Rekonstruktion des Darmes	Multiple Sklerose	2012-11-21	2013-26-03	1900-10-09
257	Ja	täglich	Heuschnupfen	Mirena	Wechsel von Herzklappe	Darmfunktionsstörung	2012-11-21	2013-26-03	1910-02-02

- Import durchführen

## Ergebnis

**i2b2 Query & Analysis Tool** Project: idrtdemo User: i2b2 User Find Patients | Analysis

**Navigate Terms** Find Terms

- Ontology
  - Patientendaten
    - IDRT-CSV-MedizinischeDaten
      - BMI
      - BP dia
      - BP sys
      - LV Mass
      - Oxy Sat in %
      - RV Mass
    - IDRT-CSV-PatientenDaten
      - Alter
      - Geburtsjahr
      - Geschlecht
      - Gewicht
      - Groesse
    - i2b2\_DB\_Import\_I2B2IDRTDEMO\_IDRT\_DB\_IMPORT
      - ALKOHOL
      - ALLERGIEN
        - Heuschnupfen
        - Laktose
        - Nein
        - Nickel
      - ERKRANKUNGEN
      - MEDIKAMENTE
      - NIKOTIN
        - Ex-Raucher
        - Ja
        - Nein
      - OPERATIONEN

**Query Tool**

Query Name:

Temporal Constraint:

Group 1			Group 2		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently		
<div style="border: 1px solid black; padding: 5px; display: inline-block;">drop a term on here</div>					

Run Query Clear Print Query 0 Groups

**Query Status**

- Beispiel ODM-Datei vom CDISC Connect-A-Thon, erweitert auf 1500 Patienten
- Strukturdaten:

```

IDRT-ODM-Sample.xml x
1 <?xml version="1.0" encoding="windows-1252"?>
2 <ODM xmlns="http://www.cdisc.org/ns/odm/v1.2" xmlns:ds="http://www.w3.org/2000/09/xmldsig#" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" ODMVersion="1.2.1" xmlns:cdisc="http://www.cdisc.org/ns/odm/v1.2" xmlns:ds="http://www.w3.org/2000/09/xmldsig#" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" ODMVersion="1.2.1" xmlns:cdisc="http://www.cdisc.org/ns/odm/v1.2" ...
3 <Study OID="IDRT-ODM-Sample-Study">
4 <MetaDataVersion OID="v1.1.0" Name="Version 1.1.0">
5 <Protocol>
6 <StudyEventRef StudyEventOID="SE.VISIT1" OrderNumber="1" Mandatory="Yes"/>
7 </Protocol>
8 <StudyEventDef OID="SE.VISIT1" Name="Visit 1" Repeating="Yes" Type="Common">
9 <FormDef OID="FORM.AE" Name="Form 1" Repeating="Yes">
10 <ItemGroupDef OID="IG.AE" Repeating="Yes" SASDatasetName="AE" Name="Adverse Events" Domain="AE" Comment="Some adverse events from this trial">
11 <ItemDef OID="ID.TAREA" SASFieldName="TAREA" Name="Therapeutic Area" DataType="text" Length="4">
12 <ItemDef OID="ID.PNO" SASFieldName="PNO" Name="Protocol Number" DataType="text" Length="15"/>
13 <ItemDef OID="ID.SCTRY" SASFieldName="SCTRY" Name="Country" DataType="text" Length="4">
14 <ItemDef OID="ID.F STATUS" SASFieldName="F STATUS" Name="Record status, 5 levels, internal use" DataType="text" Length="1">
15 <ItemDef OID="ID.LINE_NO" SASFieldName="LINE_NO" Name="Line Number" DataType="integer" Length="2"/>
16 <ItemDef OID="ID.AETERM" SASFieldName="AETERM" Name="Conmed Indication" DataType="text" Length="100"/>
17 <ItemDef OID="ID.AESTMON" SASFieldName="AESTMON" Name="Start Month - Enter Two Digits 01-12" DataType="integer" Length="2"/>
18 <ItemDef OID="ID.AESTDAY" SASFieldName="AESTDAY" Name="Start Day - Enter Two Digits 01-31" DataType="integer" Length="2"/>
19 <ItemDef OID="ID.AESTYR" SASFieldName="AESTYR" Name="Start Year - Enter Four Digit Year" DataType="integer" Length="4"/>
20 <ItemDef OID="ID.AESTDT" SASFieldName="AESTDT" Name="Derived Start Date" DataType="date"/>
21 <ItemDef OID="ID.AEENMON" SASFieldName="AEENMON" Name="Stop Month - Enter Two Digits 01-12" DataType="integer" Length="2"/>
22 <ItemDef OID="ID.AEENDAY" SASFieldName="AEENDAY" Name="Stop Day - Enter Two Digits 01-31" DataType="integer" Length="2"/>
23 <ItemDef OID="ID.AEENYR" SASFieldName="AEENYR" Name="Stop Year - Enter Four Digit Year" DataType="integer" Length="4"/>
24 <ItemDef OID="ID.AEENDT" SASFieldName="AEENDT" Name="Derived Stop Date" DataType="date"/>
25 <ItemDef OID="ID.AESEV" SASFieldName="AESEV" Name="Severity" DataType="text" Length="1">
26 <ItemDef OID="ID.AEREL" SASFieldName="AEREL" Name="Relationship to study drug" DataType="text" Length="1">
27 <ItemDef OID="ID.AEOUT" SASFieldName="AEOUT" Name="Outcome" DataType="text" Length="1">
28 <CodeListRef CodeListOID="CL.$AEOUT"/>
29 </ItemDef>
30 <ItemDef OID="ID.AEACTTRT" SASFieldName="AEACTTRT" Name="Actions taken re study drug" DataType="text" Length="1">
31 <CodeListRef CodeListOID="CL.$AEACTTR"/>
32 </ItemDef>
33 <ItemDef OID="ID.AECONTRT" SASFieldName="AECONTRT" Name="Actions taken, other" DataType="text" Length="1">
34 <CodeList OID="CL.$STAREAF" SASFormatName="$STAREAF" Name="$STAREAF" DataType="text">
35 <CodeListItem CodedValue="ONC">
36 <Decode>
37 <TranslatedText xml:lang="en">Oncology</TranslatedText>
38 </Decode>
39 </CodeList>
40 </ItemDef>
41 </FormDef>
42 </StudyEventDef>
43 </StudyEventRef>
44 </Protocol>
45 </Study>
46 </ODM>

```



## ■ Patientendaten:

```

332  →→→→→<SubjectData SubjectKey="ODM3">
333  →→→→→<StudyEventData StudyEventOID="SE.VISIT1" StudyEventRepeatKey="1">
334  →→→→→<FormData FormOID="FORM.AE" FormRepeatKey="1">
335  →→→→→<ItemGroupData ItemGroupOID="IG.AE" ItemGroupRepeatKey="1">
336  →→→→→<ItemData ItemOID="ID.TAREA" Value="ONC"/>
337  →→→→→<ItemData ItemOID="ID.PNO" Value="143-02"/>
338  →→→→→<ItemData ItemOID="ID.SCTRY" Value="USA"/>
339  →→→→→<ItemData ItemOID="ID.F_STATUS" Value="1"/>
340  →→→→→<ItemData ItemOID="ID.LINE_NO" Value="1"/>
341  →→→→→<ItemData ItemOID="ID.AETERM" Value="Chest Pain"/>
342  →→→→→<ItemData ItemOID="ID.AESTMON" Value="04"/>
343  →→→→→<ItemData ItemOID="ID.AESTDAY" Value="13"/>
344  →→→→→<ItemData ItemOID="ID.AESTYR" Value="2007"/>
345  →→→→→<ItemData ItemOID="ID.AESTDT" Value="2007-04-13"/>
346  →→→→→<ItemData ItemOID="ID.AEENMON" Value="11"/>
347  →→→→→<ItemData ItemOID="ID.AEENDAY" Value="24"/>
348  →→→→→<ItemData ItemOID="ID.AEENYR" Value="2004"/>
349  →→→→→<ItemData ItemOID="ID.AEENDT" Value="2004-11-24"/>
350  →→→→→<ItemData ItemOID="ID.AESEV" Value="1"/>
351  →→→→→<ItemData ItemOID="ID.AEREL" Value="0"/>
352  →→→→→<ItemData ItemOID="ID.AEOUT" Value="2"/>
353  →→→→→<ItemData ItemOID="ID.AEACTRT" Value="3"/>
354  →→→→→<ItemData ItemOID="ID.AECONTRT" Value="0"/>
355  →→→→→</ItemGroupData>
356  →→→→→</FormData>
357  →→→→→</StudyEventData>
358  →→→→→</SubjectData>
359  →→→→→<SubjectData SubjectKey="ODM4">
360  →→→→→<StudyEventData StudyEventOID="SE.VISIT1" StudyEventRepeatKey="1">
361  →→→→→<FormData FormOID="FORM.AE" FormRepeatKey="1">
362  →→→→→<ItemGroupData ItemGroupOID="IG.AE" ItemGroupRepeatKey="1">
363  →→→→→<ItemData ItemOID="ID.TAREA" Value="ONC"/>
364  →→→→→<ItemData ItemOID="ID.PNO" Value="143-02"/>
365  →→→→→<ItemData ItemOID="ID.SCTRY" Value="EU"/>
366  →→→→→<ItemData ItemOID="ID.F_STATUS" Value="2"/>
367  →→→→→<ItemData ItemOID="ID.LINE_NO" Value="1"/>
368  →→→→→<ItemData ItemOID="ID.AETERM" Value="Chest Pain"/>
369  →→→→→<ItemData ItemOID="ID.AESTMON" Value="03"/>
370  →→→→→<ItemData ItemOID="ID.AESTDAY" Value="28"/>

```



- ODM-Import durchführen



## Ergebnis

**i2b2 Query & Analysis Tool**
Project: idrtdemo
User: i2b2 User

Navigate Terms
Find Terms

**Ontology**

- [-] Patientendaten
  - [-] IDRT ODM Sample Study
    - [-] Version 1.1.0
      - [-] 001 Visit 1
        - [-] 001 Form 1
          - [-] Adverse Events
            - [-] 001 Therapeutic Area
              - Oncology
              - [+] 002 Protocol Number
              - [-] 003 Country
                - Asia
                - Europe
                - United States
              - [+] 004 Record status, 5 levels, internal use
              - [+] 005 Line Number
              - [+] 006 Conmed Indication
              - [+] 007 Start Month - Enter Two Digits 01-12
              - [+] 008 Start Day - Enter Two Digits 01-31
              - [+] 009 Start Year - Enter Four Digit Year
              - [+] 010 Derived Start Date
              - [+] 011 Stop Month - Enter Two Digits 01-12
              - [+] 012 Stop Day - Enter Two Digits 01-31
              - [+] 013 Stop Year - Enter Four Digit Year

**Query Tool**

Query Name:

Temporal Constraint:  Treat all g

Group 1			Group 2	
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x
Treat Independently ▾			Treat Independently ▾	
<div style="background-color: yellow; padding: 5px; border: 1px solid black; display: inline-block;">               drop a term on here             </div>				

Run Query
Clear
Print Query
0 Gr

TMF - Technologie und Methodenplattform für die vernetzte medizinische Forschung e. V.  
 1. European i2b2 Academic User Meeting - IDRT Hands-On Workshop am 26.03.2013

Seite 15



## **Universitätsmedizin Göttingen**

Abteilung Medizinische Informatik

<http://www.mi.med.uni-goettingen.de/>

## **IDRT-Projekt**

<http://www.tmf-ev.de/idrt>

<http://idrt.imise.uni-leipzig.de>

## **Benjamin Baum**

benjamin.baum@med.uni-goettingen.de

Tel.: +49 551 39 – 22 99 6