

Transformation von OpenClinica-Studiendaten in ein Data Warehouse (i2b2) mit Hilfe des IDRT-Importers

TMF-Workshop

OpenClinica

10. September 2013 | Berlin

Christian Bauer

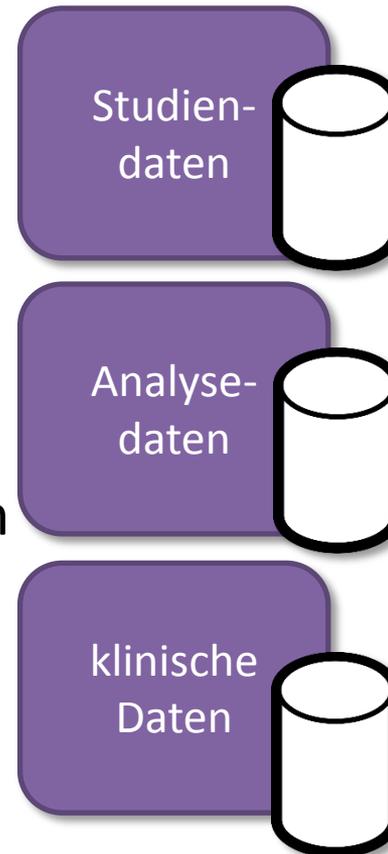
christian.bauer@med.uni-goettingen.de

www.mi.med.uni-goettingen.de



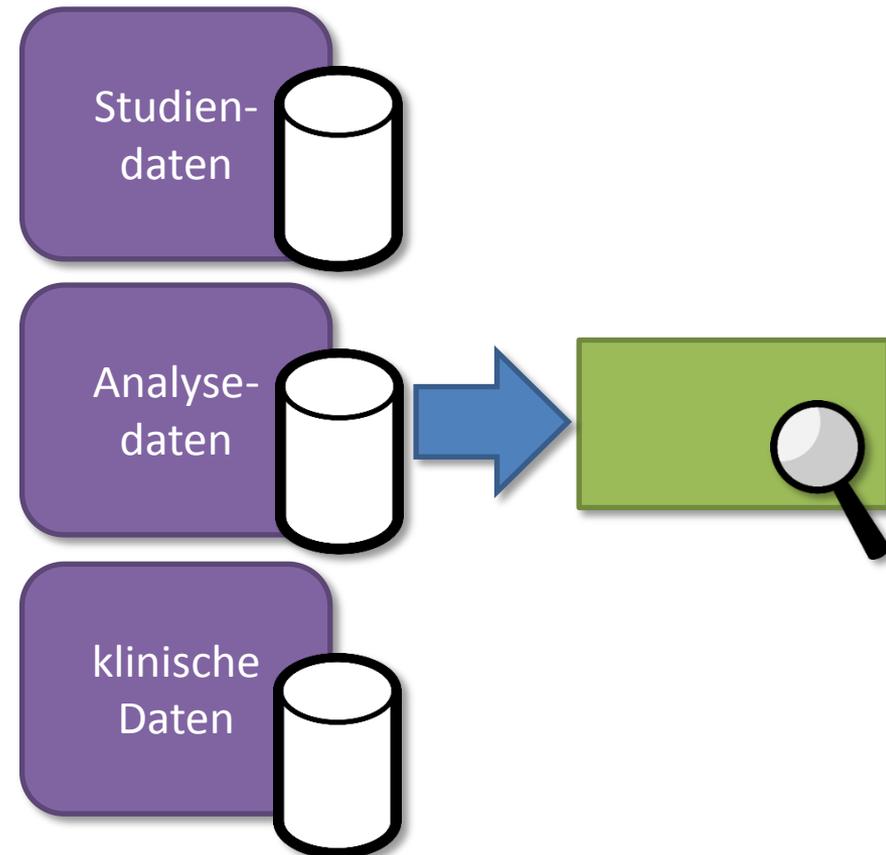
Hintergrund

- Menge an elektronisch verfügbaren medizinische Daten steigt
- neue Schwerpunkte des Datenmanagements
 - Wiederverwendung von Daten
 - Zusammenführung von heterogener Daten
 - gezielte Abfragen



Hintergrund Lösungen

- Data Warehouse
 - Ablage von heterogenen Daten
 - ETL-Tools
- Forschungsdatenbank
 - einfache Abfragetools
 - Analysewerkzeuge



Hintergrund

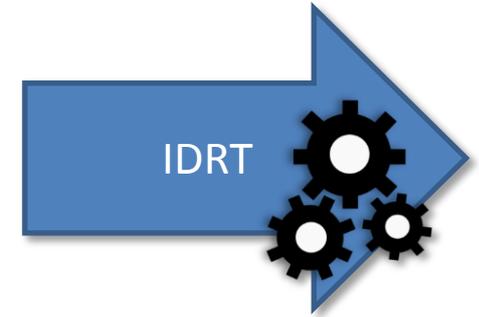
i2b2



Informatics for Integrating Biology and the Bedside (i2b2)
ist eine Open Source Forschungsdatenbank:

- Datawarehouse Schema zur Datenablage
- modularer Anwendungsserver
- Suchtools (Webanwendung und Java-Programm)

Hintergrund IDRT



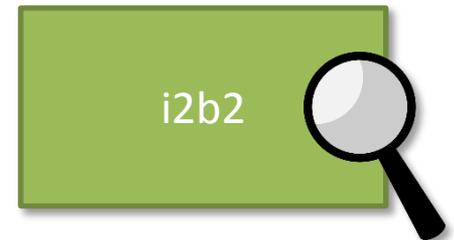
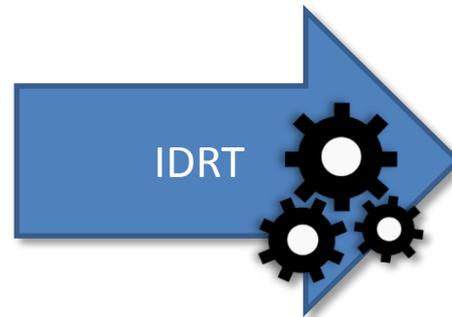
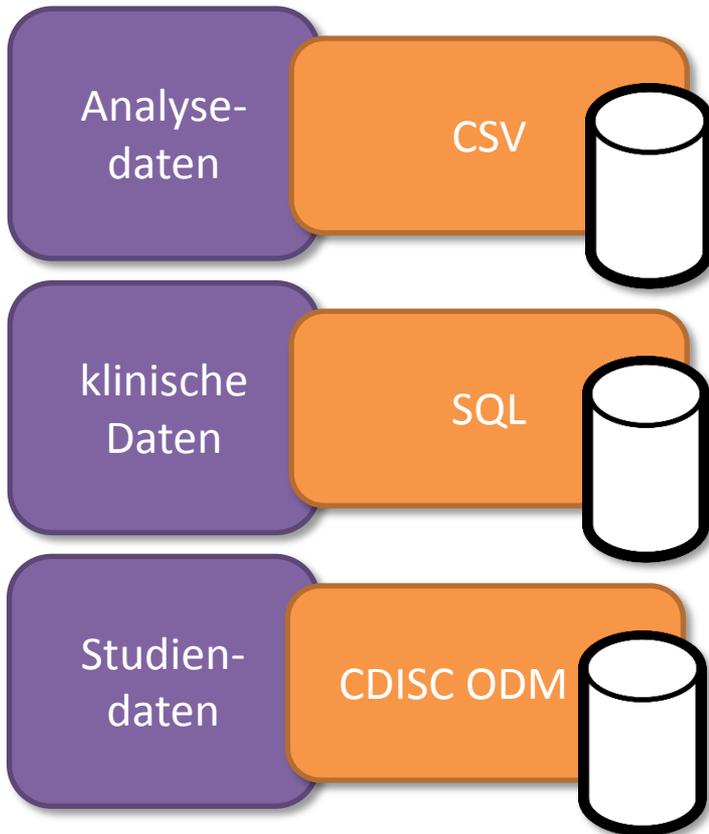
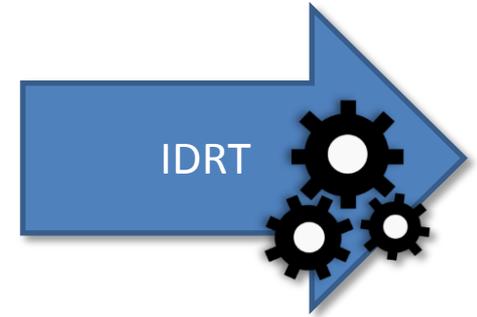
Integrated Data Repository Toolkit (IDRT¹) ist eine durch TMF²/BMBF geförderte Ergänzung von i2b2:

- Tool zur einfachen i2b2-Installation (i2b2Wizard)
- Import von Standardterminologien (z.B. ICD-10) und §21-Datensätzen (IDRT Import Tool)
- generischer Ad-Hoc-Import von medizinischen Daten für i2b2 (IDRT Import Tool)

¹ Ganslandt T, Sax U, Löbe M, Drepper J, Bauer C, Baum B, Christoph J, Mate S, Quade M, Stäubert S, Prokosch HU. Integrated Data Repository Toolkit: Werkzeuge zur Nachnutzung medizinischer Daten für die Forschung. In: Ursula Goltz, Marcus Magnor, Hans-Jürgen Appelrath, Herbert Matthies, Wolf-Tilo Balke, Lars Wolf (Hrsg.): INFORMATIK 2012. GI-Edition. Lecture Notes in Informatics. P-208 (2012), 1252-1259.

² TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (http://www.tmf-ev.de/Themen/Projekte/V091_IDRT.aspx)

Hintergrund generischer IDRT Ad-Hoc-Import



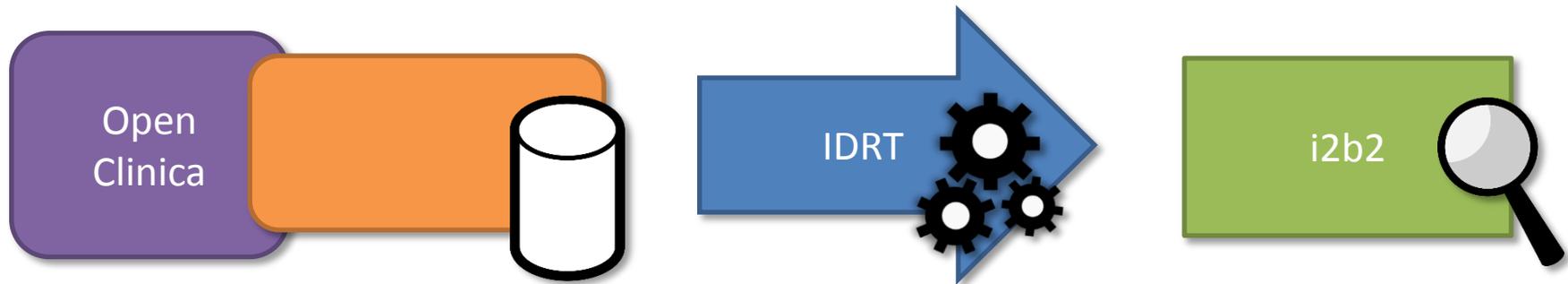
Hintergrund Open Clinica

- Clinical Data Management System



Ziel

Transformation von OpenClinica-Studiendaten in ein Data Warehouse (i2b2) mit Hilfe des IDRT-Importers.



Alerts & Messages

Your extract job completed successfully. The file is available for download [here](#).

Instructions

Other Info

Study: INSEP

Start Date: 01-Jan-2013

End Date: N/A

PI: Prof. Dr. med. Gernot Marx

Protocol Verification/IRB Approval Date:

Download Data: INSEP

Dataset Name:	INSEP
Dataset Description:	Alle Daten aus allen CRF, auch nichtabgeschlossene.
Item Status:	Data from all Available CRFs

To view or download data, select from the formats provided below. You may also select from the archived dataset files at the bottom of the page.

- CDISC ODM XML 1.3 Full with OpenClinica extensions [Run Now](#)
- CDISC ODM XML 1.3 Clinical Data with OpenClinica extensions [Run Now](#)
- CDISC ODM XML 1.3 Clinical Data [Run Now](#)
- CDISC ODM XML 1.2 Clinical Data with OpenClinica extensions [Run Now](#)
- CDISC ODM XML 1.2 Clinical Data [Run Now](#)

- View as HTML [Run Now](#)
- Excel Spreadsheet [Run Now](#)
- Tab-delimited Text [Run Now](#)
- SPSS data and syntax [Run Now](#)
- Annotated CRF (PDF) [Dann los!](#)
- Datamart (Oracle/Sqlplus) [Ab geht's!](#)
- Datamart (Postgres) [Ab geht's!](#)

Note: Large datasets may take a few minutes to generate in a pop-up window. Please disable pop-up blocking to allow the dataset to be displayed.

If you continue to have trouble, please see [Help](#) for additional tips on configuring your browser.

Internet Explorer users: If you are having trouble downloading files over an https connection, see [Help](#) for information on modifying your settings to enable https file downloads.

Archive of Exported Dataset Files:

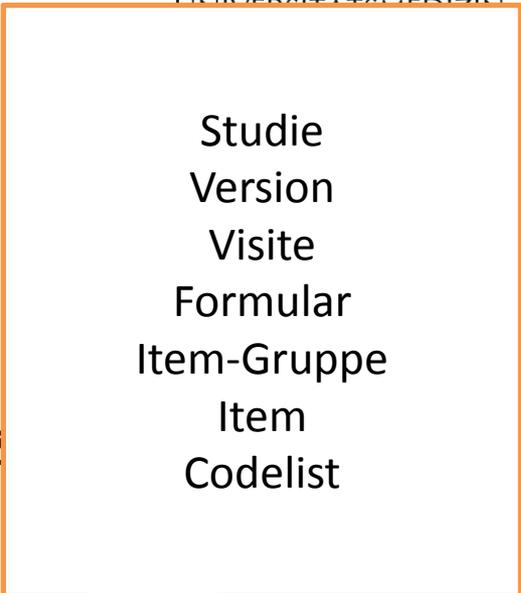
Page 1 of 1 [Find](#)

File Name	Run Time (Seconds)	File Size (Bytes)	Created Date	Created By	Action
TAB_INSEP_2013-09-06-144208495.tsv.zip	2.0	1501	Sep 6, 2013	mloebe	↓ ✗
odm1.2_clinical_INSEP_2013-09-06-144203010.xml.zip	1.0	6490	Sep 6, 2013	mloebe	↓ ✗
odm1.2_clinical_ext_INSEP_2013-09-06-144201901.xml.zip	1.0	7741	Sep 6, 2013	mloebe	↓ ✗
odm1.3_clinical_INSEP_2013-09-06-144201182.xml.zip	1.0	6525	Sep 6, 2013	mloebe	↓ ✗
odm1.3_clinical_ext_INSEP_2013-09-06-144200338.xml.zip	1.0	13572	Sep 6, 2013	mloebe	↓ ✗

Hintergrund CDISC-ODM

XML-Standard zum Transport und zur Speicherung von klinischen Studien:

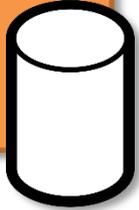
- Abbildung von Studienstrukturdaten
 - an Papier-Studien angelegte Hierarchie
- mit dieser Struktur verknüpfte Patientendaten
- Nutzung von Auszeichnungsstandards (CDISC STDM)
- AuditTrail
- Transactions



Studie
Version
Visite
Formular
Item-Gruppe
Item
Codelist

DM

von



Vorgehen Überblick

- Installation von i2b2 per i2b2Wizard des IDRT / download der IDRT-VM
- Export der Studienstruktur und Studiendaten aus OpenClinica in ODM 1.3
- Nutzung des ODM-Moduls des IDRTImportTools zum Import der OpenClinica-Daten in i2b2



Vorgehen

i2b2-Installation

- eigene Installation per i2b2Wizard
- download der IDRT-VM mit vorinstalliertem i2b2, i2b2Wizard und IDRT-ImportTool

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

```
sebmate — root@sebmate-laptop: /root/i2b2wizard-1.4.2.1 — ssh — 93x27
[JBoss/i2b2: STOPPED] [Tomcat/SHRINE: STOPPED] ::::::::::::::. i2b2 Wizard 1.4.2.1 (with IDR)

Main Menu
Move using [UP] [DOWN], [Enter] to select

System Setup      Install required software, prepare system
Projects & Users  Manage i2b2 projects and users
Boston Demodata   Load and remove the Boston Demodata
Build i2b2        Compile and deploy i2b2 webservices
JBoss Control     Start, stop and restart JBoss
Run ETL           Execute i2b2 staging, import data
SHRINE            SHRINE installation and configuration
Info              Show program and version info
Quit              Exit this program

< OK >           < Quit >
```

Alerts & Messages

Your extract job completed successfully. The file is available for download here .

Instructions

Other Info

Study: INSEP

Start Date: 01-Jan-2013

End Date: N/A

PI: Prof. Dr. med. Gernot Marx

Protocol Verification/IRB Approval Date:

Download Data: INSEP

Dataset Name:	INSEP
Dataset Description:	Alle Daten aus allen CRF, auch nichtabgeschlossene.
Item Status:	Data from all Available CRFs

To view or download data, select from the formats provided below. You may also select from the archived dataset files at the bottom of the page.

- CDISC ODM XML 1.3 Full with OpenClinica extensions [Run Now](#)
- CDISC ODM XML 1.3 Clinical Data with OpenClinica extensions [Run Now](#)
- CDISC ODM XML 1.3 Clinical Data [Run Now](#)
- CDISC ODM XML 1.2 Clinical Data with OpenClinica extensions [Run Now](#)
- CDISC ODM XML 1.2 Clinical Data [Run Now](#)
- View as HTML [Run Now](#)
- Excel Spreadsheet [Run Now](#)
- Tab-delimited Text [Run Now](#)
- SPSS data and syntax [Run Now](#)
- Annotated CRF (PDF) [Dann los!](#)
- Datamart (Oracle\Sqlplus) [Ab geht's!](#)
- Datamart (Postgres) [Ab geht's!](#)

Note: Large datasets may take a few minutes to generate in a pop-up window. Please disable pop-up blocking to allow the dataset to be displayed.

If you continue to have trouble, please see [Help](#) for additional tips on configuring your browser.

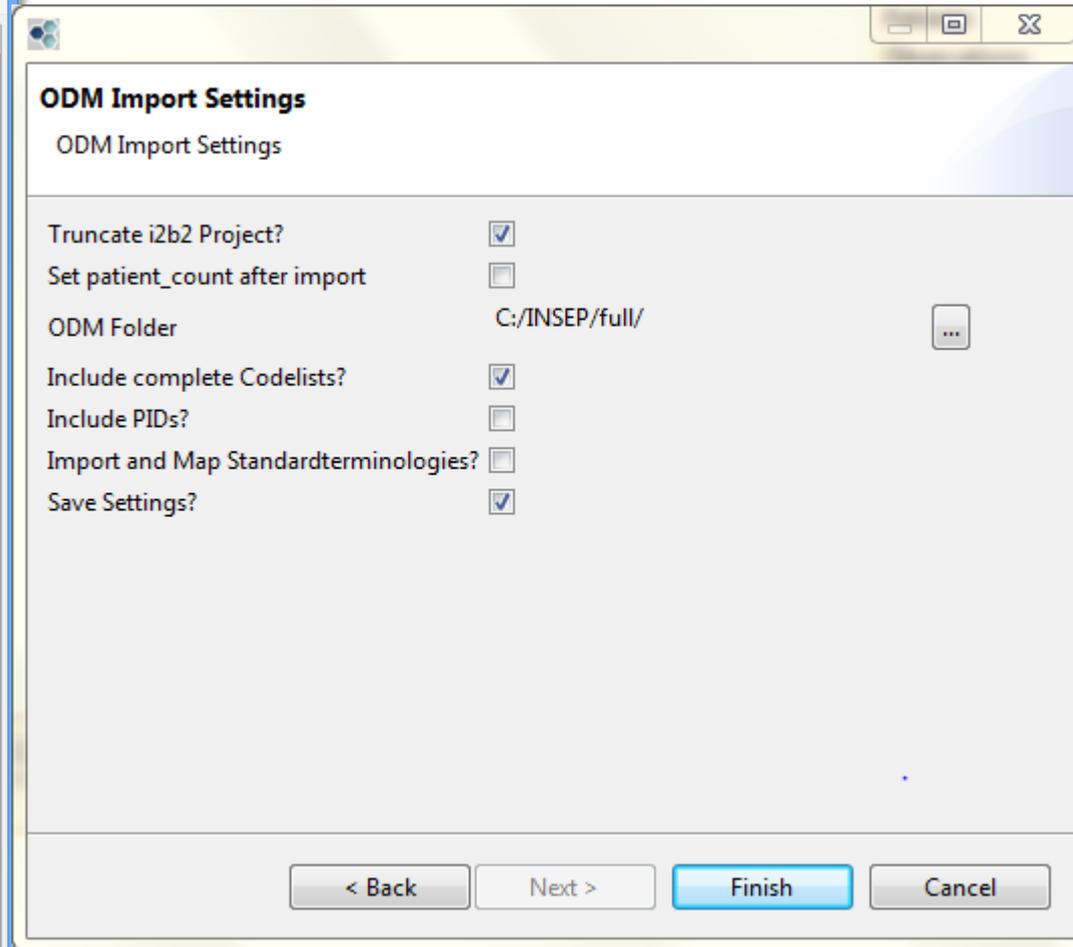
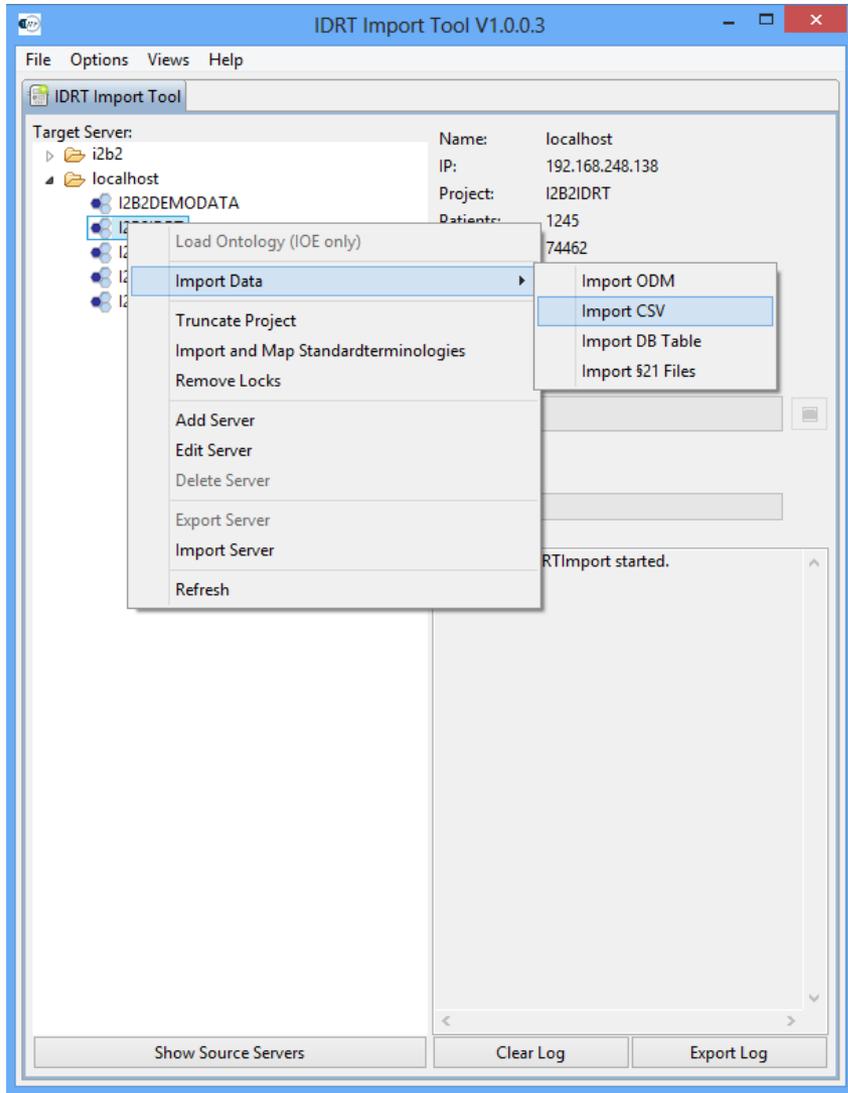
Internet Explorer users: If you are having trouble downloading files over an https connection, see [Help](#) for information on modifying your settings to enable https file downloads.

Archive of Exported Dataset Files:

Page 1 of 1 [Find](#)

File Name	Run Time (Seconds)	File Size (Bytes)	Created Date	Created By	Action
TAB_INSEP_2013-09-06-144208495.tsv.zip	2.0	1501	Sep 6, 2013	mloebe	↓ ✖
odm1.2_clinical_INSEP_2013-09-06-144203010.xml.zip	1.0	6490	Sep 6, 2013	mloebe	↓ ✖
odm1.2_clinical_ext_INSEP_2013-09-06-144201901.xml.zip	1.0	7741	Sep 6, 2013	mloebe	↓ ✖
odm1.3_clinical_INSEP_2013-09-06-144201182.xml.zip	1.0	6525	Sep 6, 2013	mloebe	↓ ✖
odm1.3_clinical_ext_INSEP_2013-09-06-144200338.xml.zip	1.0	13572	Sep 6, 2013	mloebe	↓ ✖

Vorgehen



The screenshot displays the i2b2 Workbench interface. On the left, the 'Ontology' tree shows a path from 'Patientendaten' to 'INSEP' to '001 Tag 1' to 'IN_Basisdaten - V0.1' to 'IG_IN_BA_UNGROUPED' to '002 Geschlecht' to 'Weiblich'. A red arrow points from this 'Weiblich' term to the 'Query Tool' window. The 'Query Tool' window shows a query named '(+) Weibl-date-008 A@06:47:41' with three groups. Group 1 contains 'Weiblich', Group 2 contains 'date', and Group 3 contains '008 Art der ITS-Aufnahme'. A second red arrow points from the 'Run Query Above' button to the 'Timeline View' window. The 'Timeline View' window shows a table with 3 patients returned.

Select	Patient	View Image	002 Geschlecht	008 Art der ITS...	003 Krankenhau...
<input type="checkbox"/>	SS_006	SS_006	Weiblich	Medizinisch	2013-09-03
<input type="checkbox"/>	SS_005	SS_005	Weiblich	Andere	2013-07-02
<input type="checkbox"/>	SS_003	SS_003	Weiblich	Chirurgisch unge...	2013-08-12

Probleme

- „OpenClinica“-ODM benutzt eigene Erweiterungen
- alle *unrepeated* Item Groups werden in einer Gruppe „Ungrouped“ zusammengefasst
- MultiSelectLists werden als durch Komma getrennte Strings in ODM abgelegt

```
<FormDef OID="F_IN_BASISDATE_V01" Name="IN_Basisdaten - V0.1" Repeating="No">
  <ItemGroupRef ItemGroupOID="IG_IN_BA_UNGROUPED" Mandatory="Yes"/>
  <OpenClinica:FormDetails FormOID="F_IN_BASISDATE_V01" ParentFormOID="F_IN_BASISDATE">
    <OpenClinica:VersionDescription>Initiale Erstellung</OpenClinica:VersionDescription>
    <OpenClinica:RevisionNotes>03.01.2013 09:45 Uhr Matthias L&#246;be</OpenClinica:RevisionNotes>
    <OpenClinica:PresentInEventDefinition StudyEventOID="SE_TAG1" IsDefaultVersion="Yes" PasswordRequired="No" DoubleDataEntry="No" HideCR
SourceDataVerification="not_required"/>
  </OpenClinica:FormDetails>
</FormDef>
```

Kontakt

Christian Bauer

Universitätsmedizin Göttingen
Institut für Medizinische Informatik
Robert-Koch-Str. 40
37075 Göttingen

Tel.: +49 551 39-8227

Fax: +49 551 39-22493

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

E-Mail: christian.bauer@med.uni-goettingen.de

TMF – Technologie- und Methodenplattform
für die vernetzte medizinische Forschung e.V.
Projektnummer V091-01M



 Bundesministerium
für Bildung
und Forschung

Download der IDRT-Tools: <http://idrt.imise.uni-leipzig.de>

Kontakt

Christian Bauer

Universitätsmedizin Göttingen
Institut für Medizinische Informatik
Robert-Koch-Str. 40
37075 Göttingen

Tel.: +49 551 39-8227

Fax: +49 551 39-22493

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

E-Mail: christian.bauer@med.uni-goettingen.de

TMF – Technologie- und Methodenplattform
für die vernetzte medizinische Forschung e.V.
Projektnummer V091-01M



Bundesministerium
für Bildung
und Forschung

Download der IDRT-Tools: <http://idrt.imise.uni-leipzig.de>

i2b2 Workbench for Spielewiese

AHF Admin Status

Navigate Terms Find Terms Edit Terms

Query Tool

Ontology

Patientendaten

INSEP

MetaDataVersion_v1.0.0

001 Tag 1

IN_Basisdaten - V0.1

IG_IN_BA_UNGROUPED

001 Alter

002 Geschlecht

Männlich

Weiblich

003 Krankenhausaufnahme

date

004 Übernahme aus anderem KH

005 ITS-Aufnahme

006 SAPS II-Score

007 SAPS II-Score

008 Art der ITS-Aufnahme

Andere

Chirurgisch geplant

Chirurgisch ungeplant (Notfall)

Medizinisch

Trauma

009 Entlassung Krankenhaus

010 Entlassung ITS

011 Status bei ITS-Entlassung

012 Status bei KH-Entlassung

Query Name: Weibl-008 A-003 K@06:40:43

Group 1	Group 2	Group 3
<input type="checkbox"/> Dates <input type="checkbox"/> Occurs > 0x <input type="checkbox"/> Exclude Treat Independently	<input type="checkbox"/> Dates <input type="checkbox"/> Occurs > 0x <input type="checkbox"/> Exclude Treat Independently	<input type="checkbox"/> Dates <input type="checkbox"/> Occurs > 0x <input type="checkbox"/> Exclude Treat Independently
Weiblich	008 Art der ITS-Aufnahme	003 Krankenhausaufnahme
The terms of this group are joined then intersected with other groups	The terms of this group are joined then intersected with other groups	The terms of this group are joined then intersected with other groups

 Get Everyone

Run Query Above

Patient(s) returned

Timeline View Analysis View Biomaterial (beta) Chili/Web Viewer Additional Data V1.1.0

Query: Weibl-008 A-003 K@06:40:43 (5 patients)

Select	Patient	View Image	002 Geschlecht	008 Art der ITS-Aufnahme	003 Krankenhausaufnahme
<input type="checkbox"/>	SS_006		Weiblich	Medizinisch	2013-09-03
<input type="checkbox"/>	SS_005		Weiblich	Andere	2013-07-02
<input type="checkbox"/>	SS_003		Weiblich	Chirurgisch ungeplant (Notfall)	2013-08-12
<input type="checkbox"/>	SS_002		Weiblich	Medizinisch	2013-02-18
<input type="checkbox"/>	SS_001		Weiblich	Chirurgisch ungeplant (Notfall)	2013-02-03

Previous Queries

Weibl-008 A-003 K@06:40:43 [09-09-2013] [ahfadmin]

Navigate Terms Find Terms Edit Terms

Ontology

Patientendaten

INSEP

MetaDataVersion_v1.0.0

001 Tag 1

IN_Basisdaten - V0.1

IG_IN_BA_UNGROUPED

001 Alter

002 Geschlecht

Männlich

Weiblich

003 Krankenhausaufnahme

date

004 Übernahme aus anderem KH

005 ITS-Aufnahme

006 SAPS II-Score

006 SAPS II-Score (Numerical)

007 SAPS II-Score

008 Art der ITS-Aufnahme

Andere

Chirurgisch geplant

Chirurgisch ungeplant (Notfa)

Medizinisch

Trauma

009 Entlassung Krankenhaus

010 Entlassung ITS

011 Status bei ITS-Entlassung

Query Tool

Query Name: (+) Weibl-date-008 A@06:47:41

Group 1	Group 2	Group 3
Dates Occurs > 0x Exclude	Dates Occurs > 0x Exclude	Dates Occurs > 0x Exclude
Treat Independently	Treat Independently	Treat Independently
Weiblich	date	008 Art der ITS-Aufnahme
Drag terms from Navigate, Find and Workplace into this group	Drag terms from Navigate, Find and Workplace into this group	The terms of this group are joined then intersected with other groups
<input type="checkbox"/> Get Everyone <input type="button" value="Run Query Above"/>		

Timeline View Analysis View Biomaterial (beta) Chili/Web Viewer Additional Data V1.1

Query: (+) Weibl-date-008 A@06:47:41 (3 patients)

Select	Patient	View Image	002 Geschlecht	008 Art der ITS-...	003 Krankenhau...
<input type="checkbox"/>	SS_006	SS_006	Weiblich	Medizinisch	2013-09-03
<input type="checkbox"/>	SS_005	SS_005	Weiblich	Andere	2013-07-02
<input type="checkbox"/>	SS_003	SS_003	Weiblich	Chirurgisch unge...	2013-08-12

Previous Queries

- (+) Weibl-date-008 A@06:47:41 [09-09-2013] [ahfadmin]
- (+) Weiblich-date@06:47:25 [09-09-2013] [ahfadmin]
- (+) Weiblich-date@06:47:04 [09-09-2013] [ahfadmin]

Select All

View Matching Bi...

View selected

View in external bro...

i2b2 Workbench for Spielwiese

Find Terms Edit Terms

Query Tool

patientendaten
INSEP
MetaDataVersion_v1.0.0
001 Tag 1
IN_Basisdaten - V0.1
IG_IN_BA_UNGROUPED
001 Alter
002 Geschlecht
Männlich
Weiblich
003 Krankenhausaufnahme
date
004 Übernahme aus anderem KH?
005 ITS-Aufnahme
006 SAPS II-Score
007 SAPS II-Score
008 Art der ITS-Aufnahme
Andere
Chirurgisch geplant
Chirurgisch ungeplant (Notfall)
Medizinisch
Trauma
009 Entlassung Krankenhaus
010 Entlassung ITS
011 Status bei ITS-Entlassung
012 Status bei KH-Entlassung
013 Patient mit schwerer Sepsis / septischem Schock? - Falls "Nein", ist die Erhebung hier
IN_SOFASubscore - V0.1
IN_Sepsisdaten - V0.1
002 Tag 7
IN_SOFASubscore - V0.1
IG_IN_SO_UNGROUPED
001 Kreislauf
001 Kreislauf (Numerical Query)
002 Pulmo
003 Gerinnung
004 Niere
004 Niere (Numerical Query)
005 Leber
006 ZNS
003 Entlassung ITS

Previous Queries

Weibl-008 A-003 K@06:40:43 [09-09-2013] [ahfadmin]
Weibl-008 A-003 K@06:36:34 [09-09-2013] [ahfadmin]
Weiblic-008 Art@06:36:03 [09-09-2013] [ahfadmin]
Weibl-008 A-005 L@06:35:35 [09-09-2013] [ahfadmin]
Weibl-008 A-005 L@06:32:01 [09-09-2013] [ahfadmin]
Results of Weibl-008 A-005 L@06:32:01

Query Name: Weibl-008 A-003 K@06:40:43

Group 1	Group 2	Group 3	Group 4
<p>Dates Occurs > 0x Exclude</p> <p>Treat Independently</p> <p>Weiblich</p>	<p>Dates Occurs > 0x Exclude</p> <p>Treat Independently</p> <p>008 Art der ITS-Aufnahme</p>	<p>Dates Occurs > 0x Exclude</p> <p>Treat Independently</p> <p>003 Krankenhausaufnahme</p>	<p>Dates Occurs > 0x Exclude</p> <p>Treat Independently</p>
The terms of this group are joined then intersected with other groups	The terms of this group are joined then intersected with other groups	The terms of this group are joined then intersected with other groups	Drag terms from Navigate, Find and Workplace into this group

Get Everyone Run Query Above

Timeline View Analysis View Biomaterial (beta) Chili/Web Viewer Additional Data V1.1.0

Query: Weibl-008 A-003 K@06:40:43 (5 patients)

Select	Patient	View Image	002 Geschlecht	008 Art der ITS-Aufnahme	003 Krankenhausaufnahme
<input type="checkbox"/>	SS_006	SS_006	Weiblich	Medizinisch	2013-09-03
<input type="checkbox"/>	SS_005	SS_005	Weiblich	Andere	2013-07-02
<input type="checkbox"/>	SS_003	SS_003	Weiblich	Chirurgisch ungeplant (Notfall)	2013-08-12
<input type="checkbox"/>	SS_002	SS_002	Weiblich	Medizinisch	2013-02-18
<input type="checkbox"/>	SS_001	SS_001	Weiblich	Chirurgisch ungeplant (Notfall)	2013-02-03

Select All View Matching Biosamples View selected View in external browser