

1. European i2b2 Academic User Group Meeting

# **IDRT: Integration and Maintenance of Medical Terminologies in i2b2**

---

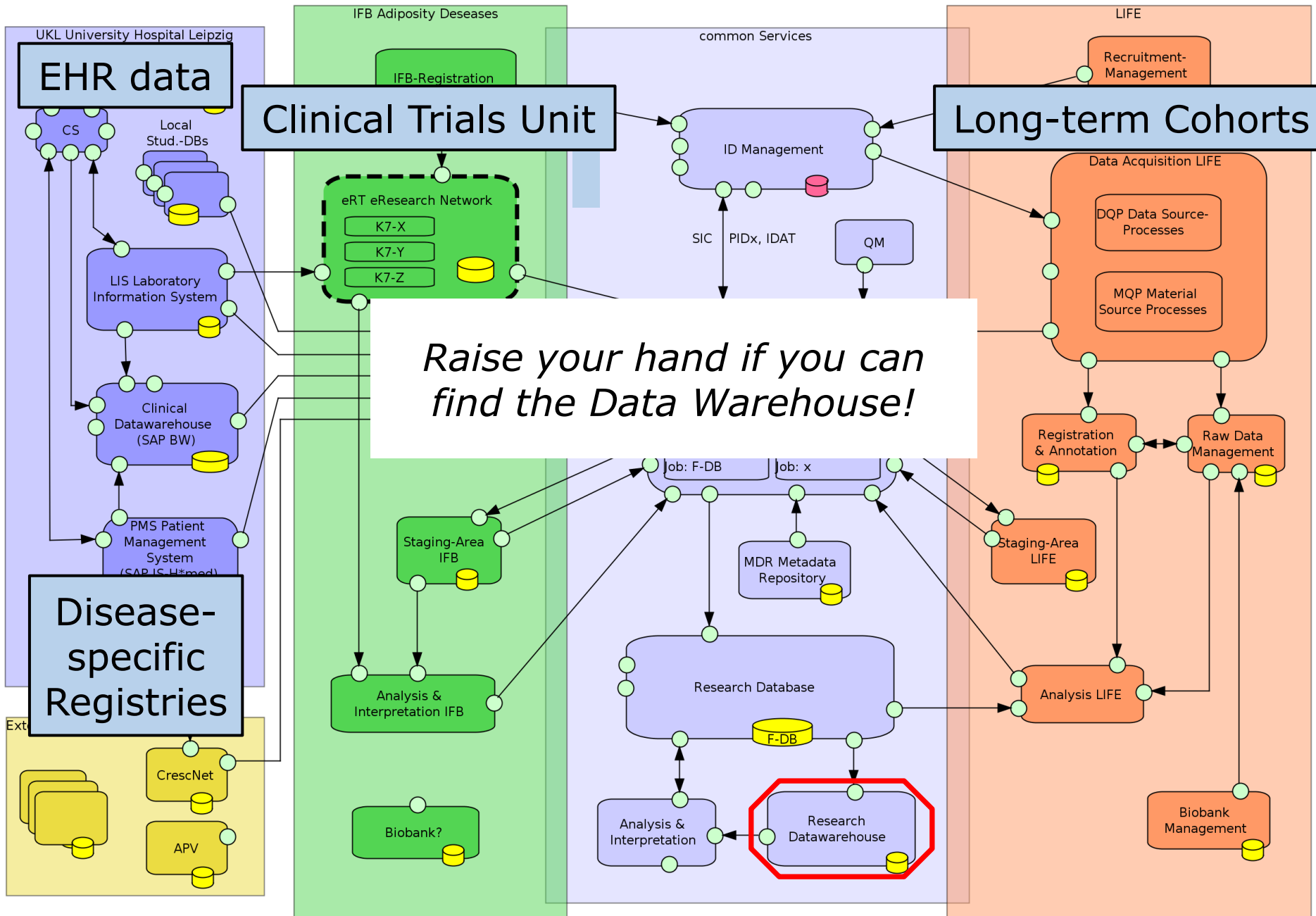
TMF-funded project "Integrated Data Repository Toolkit"

Matthias Löbe, Sebastian Stäubert, Alfred Winter  
Institute for Medical Informatics (IMISE), Universität Leipzig

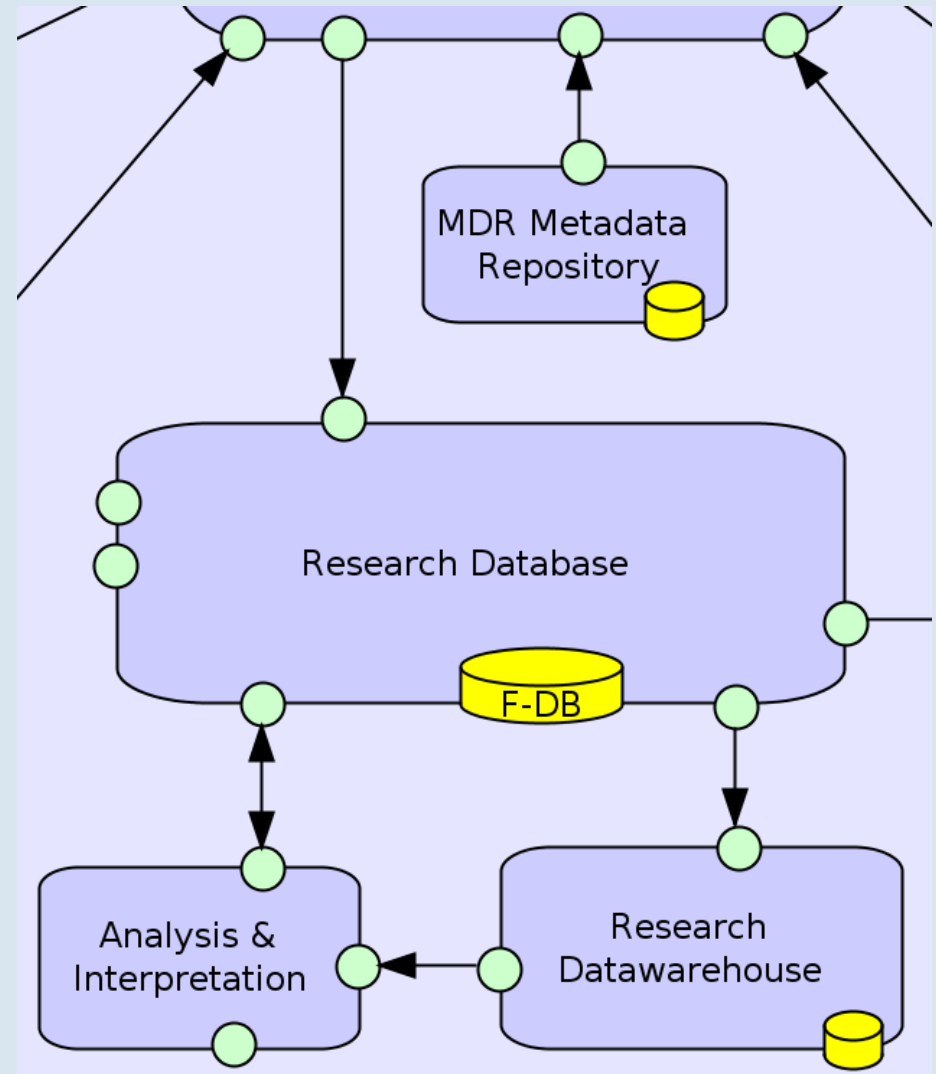
*„I've got a mission for the two of you ...“*

- Building a research DW is a “strategic aim”
  - ☑ “Strategic” like “long-range effort”
  - ✗ Not so “strategic” in defining milestones and success criteria
- Business data warehouse already exists (since 2006)
  - ↳ Used for financial controlling
  - ↳ Not accessible for researchers
  - ↳ Data islands
- **What do we want to do with a research warehouse?**
- Center for clinical trials (2010): facilitate patient recruitment!
  - ↳ Find eligible subjects by querying for inclusion criteria
  - ↳ Or at least support cohort estimation by retrospective queries
- 6-month evaluation using i2b2
  - ↳ Nice GUI, easy to understand, scales to large datasets
  - ↳ Problems with getting data into it => IDRT project

# University Hospital Leipzig Big Picture



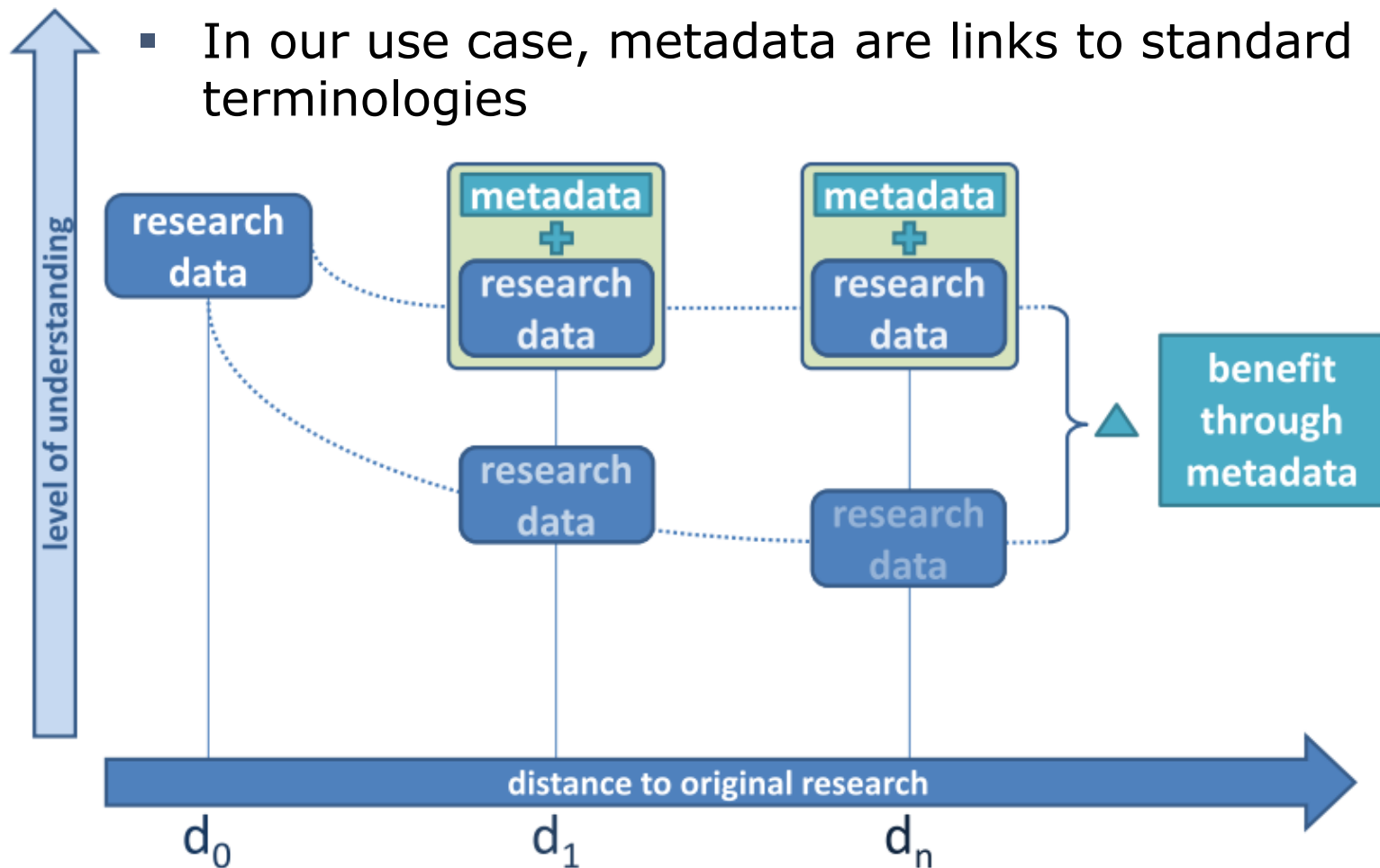
- Research database will store all research data untouched and versioned
- Data warehouse will provide an integrated, up-to-date view on research-specific hypotheses
- (To-be-developed) Metadata Repository will store complex semantic annotations on data elements



# Why Standard Terminologies?

Research data are only interpretable by those who made them

- In our use case, metadata are links to standard terminologies

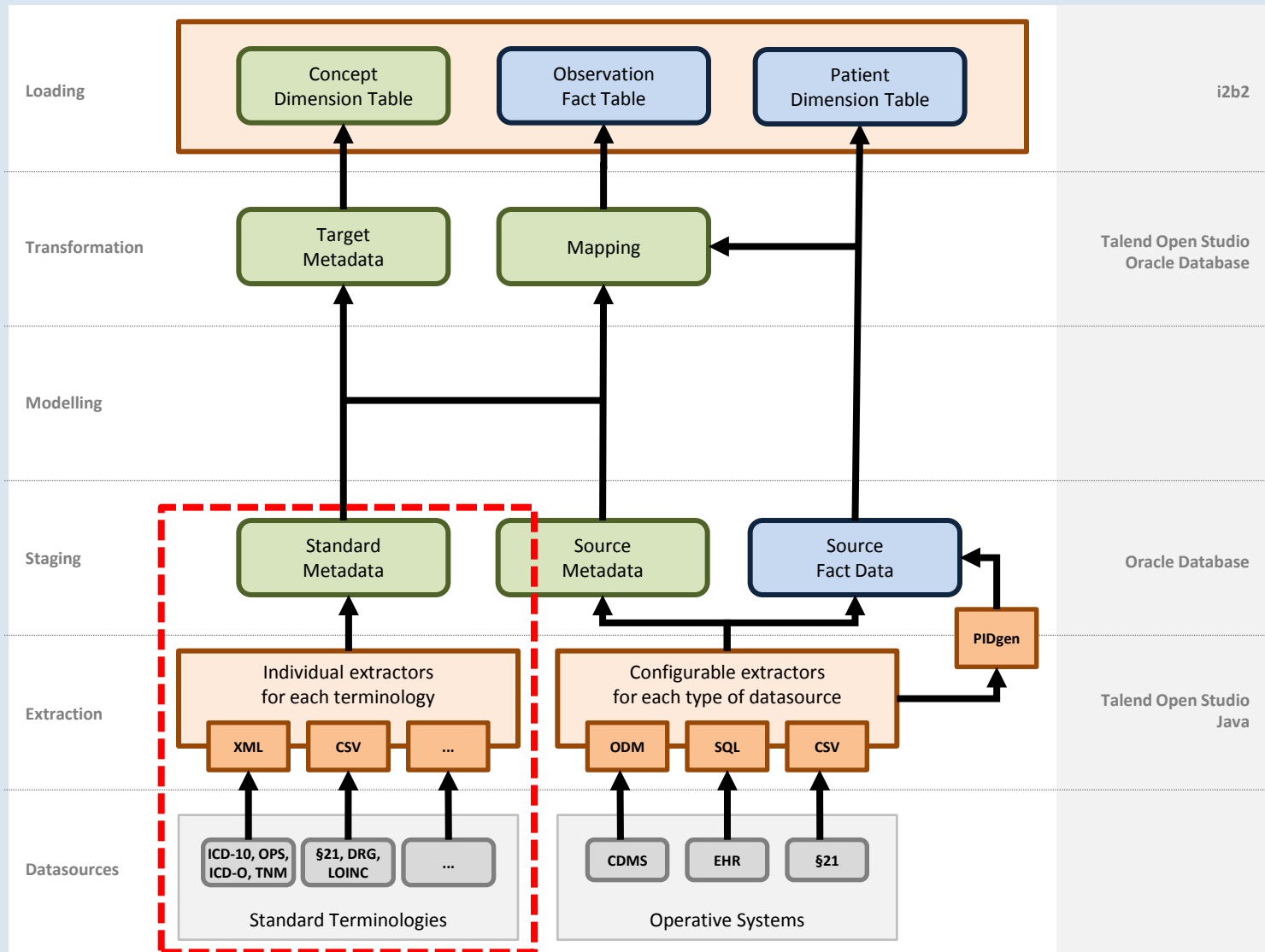


Quelle: Dickmann, Frank and Grütz, Romanus (2011): LABIMI/F - Digital Preservation of Biomedical Research Data, Knowledge Exchange: Workshop Research Data Management – Activities and Challenges, 14.-15. November 2011, Bonn, Poster, Access date: 2012.01.12, URL: [http://www.labimi-f.med.uni-goettingen.de/Publications/Poster\\_LABIMI-F\\_v6.pdf](http://www.labimi-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

- Patient demographics
- Diagnoses and observations
- Procedures and surgeries
- Laboratory
- Drugs and medication
- Biobank samples
- Encounters and admissions
- Billing and cost
- Patient reported data

AOD	FDA	MDR	OMIM
AOT	GDRG	MEDLINEPLUS	OPS
BioC	GO	MED	PDQ
CBO	HCPCS	MGED	PMA
CCS	HL7V3.0	MSH	PNDS
CDC	HUGO	MTHFDA	QMR
CDISC	ICD10AE	MTHHH	RADLEX
CDT	ICD10	MTHICD9	RAM
COH	ICD9CM	MTHICPC2ICD107B	RENI
COSTAR	ICDO	MTHICPC2ICD10AE	RXNORM
CRCH	ICH	MTHMST	SNOMEDCT
CSP	ICPC2ICD10ENG	MTHSPL	SPN
CST	ICPC	MTH	SRC
CTCAE	JAX	NCBI	TNM
CTEP	KEGG	NCI-GLOSS	UCUM
DCP	LNC	NCI-HL7	UMD
DICOM	LNC	NCIMTH	USPMG
DTP	MBD	NCISEER	UWDA
DXP	MCM	NCI	VANDF
ELC	MDBCAC	NDFRT	ZFIN

- Data elements in clinical research as well as in patient care often use codes from controlled vocabularies to define their permissible values
- i2b2-included terminologies are not used in Germany





# A Clinical Form in Real Life

## Data Entry on Case Report Form in OpenClinica

- Attribute: an abstract concept which describes a object's characteristic

↳ Can be part of a medical classification

- Value: a value that is being assigned to an attribute

↳ Example: *true*, 38, 176

**View Section Data Entry for PSKA\_Baseline 9.85 ?**

CRF Info ▾

Press the little flag icon beside an input to enter discrepancy notes, please note that you can only save the notes if

Exit

Informa... (0/4) Demogra... (0/4) Medical... (0/6)

**Title: Demographics**

Demographics

Gender  male  female

Age 38 \* (years)

Height 176 \* (cm)

Weight \* (kg)

-- Select to Jump --

- Cover
- Inclusion and Exclusion Criteria
- Informed Consent
- Demographics
- Medical History
- Premedication
- Qualifying ischemic stroke

Source: S. Mate





# i2b2 Web Client

## Hierarchy for navigation and content selection

i2b2 Query & Analysis Tool

Project: TERM

User: Terminus

Find Patients | Analysis Tools | Message Log | Help | Logout

Navigate Terms

Find Terms

- Ontology
  - Patientendaten
    - S\_PSK\_PRED
      - MetaDataVersion\_v1.0.0
        - 001 PSK\_Tag 1
          - PSKA\_Baseline - 0.85
            - Ungrouped
              - 001 Predictors of Sepsis - Patient-ID
              - 002 VISIT 1 Baseline - Visit Date
              - 003 Inclusion Criteria - Acute ischemic infarction
              - 004 Patient age >= 18
              - 005 NIH stroke score (NIHSS) >= 8
              - 006 Exclusion Criteria - Clinical or paraclinical s
              - 007 History of previous cerebrovascular event
              - 008 Cardiac arrhythmia
              - 009 Has the nature of the study been explained
              - 010 Informed Consent - Date of written informec
              - 011 If the patient is not able to give informed co
              - 012 Date of written informed consent
              - 013 Demographics - Gender
                - female
                - male
              - 014 Age
                - 014 Age (Numerical Query)
              - 015 Height
              - 016 Weight
              - 017 Are any pre-existing diseases/conditions ac
              - 018 Are any medications currently being taken p
              - 019 Qualifying ischemic stroke - Onset of CVE

Folder  
Leaf

Query Tool

Query Name:

Temporal Constraint:

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
drop a term on here								

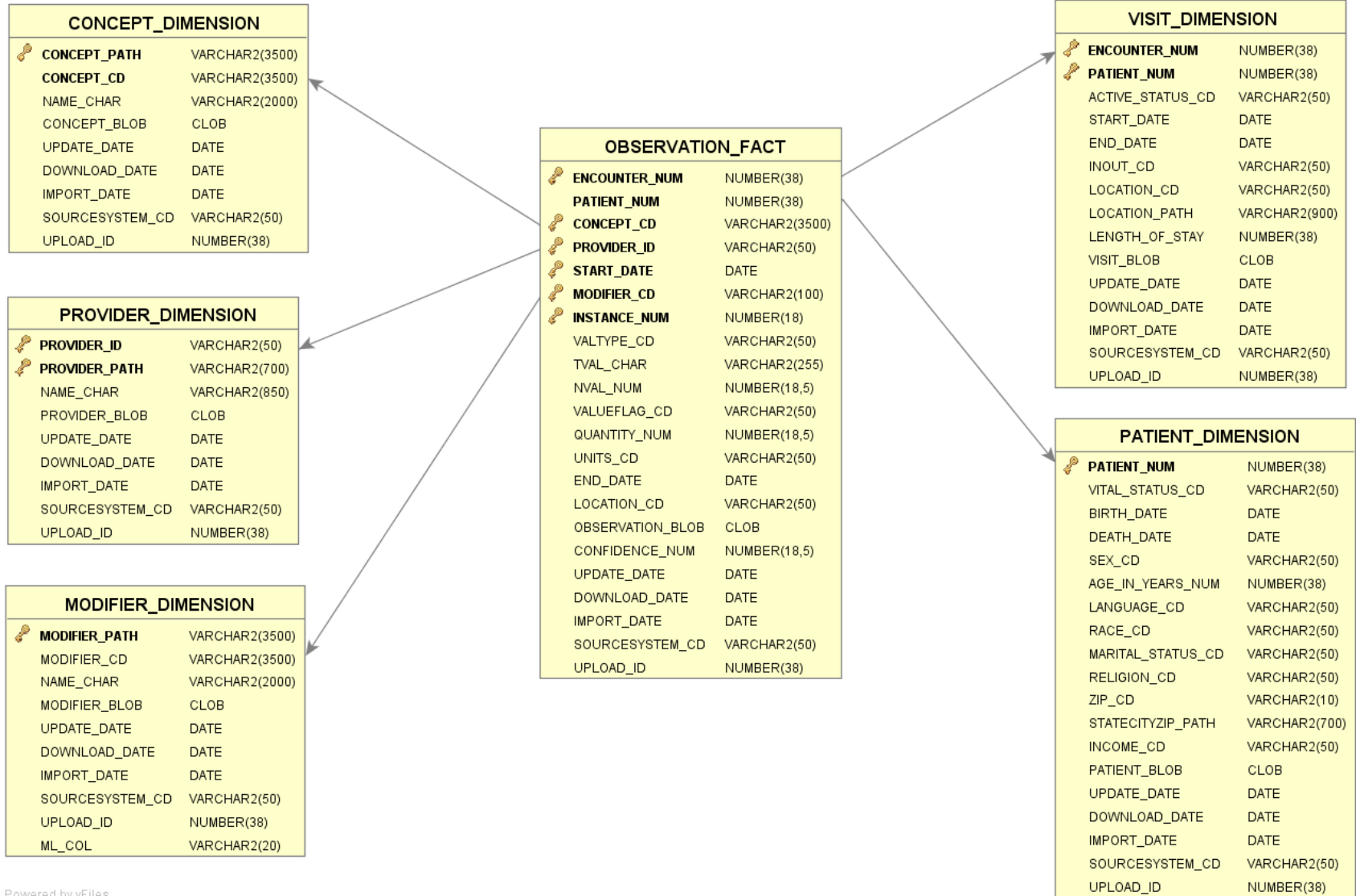
Run Query Clear Print Query 0 Groups New Group

Query Status



# i2b2 Star Schema

## Entity-Attribute-Value



Powered by yFiles



# i2b2 Ontology

## Hierarchy for navigation and content selection

Table: I2B2

TERM@localhost/2B2TERM/TABLE/I2B2

Info Columns Data Row Count Primary Key Indexes Grants Row Id

#	C_HLEVEL	C_FULLNAME	C_NAME
64	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	001 Predictors of Sepsis - Patie...
65	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	002 VISIT 1 Baseline - Visit Date
66	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	003 Inclusion Criteria - Acute is...
67	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	004 Patient age >= 18
68	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	005 NIH stroke score (NIHSS) >...
69	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	006 Exclusion Criteria - Clinical ...
70	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	007 History of previous cerebro...
71	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	008 Cardiac arrhythmia
72	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	009 Has the nature of the study...
73	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	010 Informed Consent - Date of...
74	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	011 If the patient is not able to...
75	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	012 Date of written informed co...
76	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	013 Demographics - Gender
77	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	014 Age
78	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	015 Height
79	7	\i2b2\PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1...	016 Weight

Table: OBSERVATION\_FACT

TERM@localhost/2B2TERM/TABLE/OBSERVATION\_FACT

Info Columns Data Row Count Primary Key Indexes Grants Row Id References

#	PATIENT_NUM	CONCEPT_CD	NVAL_NUM
10	5	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_ACUTEIN...	1
11	6	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_ACUTEIN...	1
12	7	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_ACUTEIN...	0
13	8	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_ACUTEIN...	1
14	9	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_ACUTEIN...	1
15	10	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_ACUTEIN...	1
16	11	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_ACUTEIN...	1
17	14	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_ACUTEIN...	1
18	1	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml	105
19	3	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml	38
20	11	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml	100
21	12	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml	100
22	4	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml	1
23	5	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml	80
24	6	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml	200
25	7	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml	80
26	9	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml	25

Table: CONCEPT\_DIMENSION

TERM@localhost/2B2TERM/TABLE/CONCEPT\_DIMENSION

Info Columns Data Row Count Primary Key Indexes Grants Row Id References

#	CONCEPT_PATH	CONCEPT_CD
7	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_HISTORY_YN:xml
8	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_CARDIACARR_YN:xml
9	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_INFOPATIENT_YN:xml
10	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_INFOPATIENT_D:756909918
11	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_INFOPATIENT_YN:xml
12	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_AGE:xml
13	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_HEIGHT:xml
14	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_WEIGHT:xml
15	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_EXDISEASE_YN:xml
16	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_MEDHISTORY\I_PSKA_PSKA_DIAGN_DISEASE:492514880
17	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_DIAGN_YEAR:423973328
18	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_RECovYEAR:305047613
19	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_MEDHISTORY\I_PSKA_PSKA_DIAGN_YEARON:1814433547
20	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_VISIT_D:569489665
21	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_INFOPATIENT_D:756939672
22	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_INFOPATIENT_D:1077281734
23	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_PREMEDIC_YN:xml
24	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_CVESYMP_D:427540981
25	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_CVESYMP_T:1266670838
26	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_CRANIALCT_YN:xml
27	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_INSULARCX_YN:xml
28	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_CRANIALMRI_YN:xml
29	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_RINSULARCX_YN:xml
30	\i2b2\PD\S_PSK_PR...	PD\S_PSK_PRED\w1.0.0\SE_PSK_TAG1\F_PSKA_BASELIN_085\IG_PSKA_UNGROUPED\I_PSKA_PSKA_INFARCTGR13:xml

Table: PATIENT\_DIMENSION

TERM@localhost/2B2TERM/TABLE/PATIENT\_DIMENSION

Info Columns Data Row Count Primary Key Indexes Grants

#	PATIENT_NUM	VITAL_STATUS_CD	BIRTH_DATE	DEATH_DATE
1	1	N	(null)	(null)
2	3	N	(null)	(null)
3	4	N	(null)	(null)
4	5	N	(null)	(null)
5	6	N	(null)	(null)
6	7	N	(null)	(null)
7	8	N	(null)	(null)
8	9	N	(null)	(null)
9	10	N	(null)	(null)
10	11	N	(null)	(null)
11	12	N	(null)	(null)
12	13	N	(null)	(null)
13	14	N	(null)	(null)

Table: PATIENT\_DIMENSION

TERM@localhost/2B2TERM/TABLE/PATIENT\_DIMENSION

Info Columns Data Row Count Primary Key Indexes Grants

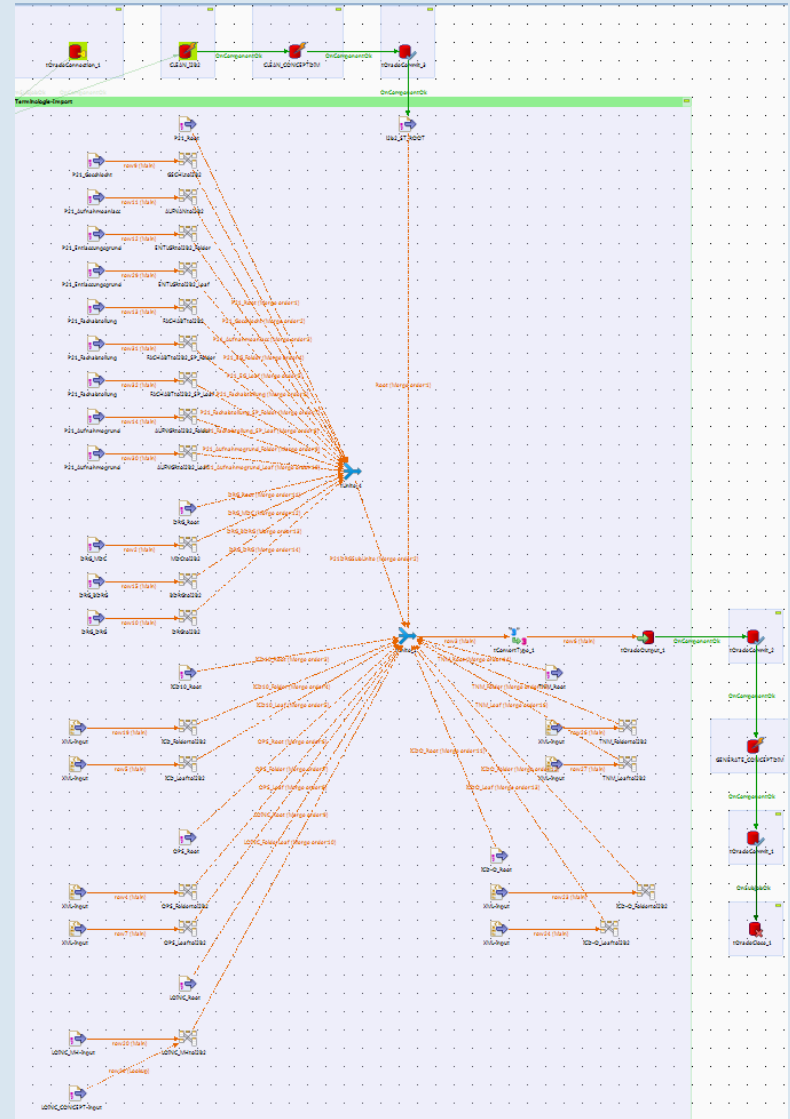
#	PATIENT_NUM	VITAL_STATUS_CD	BIRTH_DATE	DEATH_DATE
1	1	N	(null)	(null)
2	3	N	(null)	(null)
3	4	N	(null)	(null)
4	5	N	(null)	(null)
5	6	N	(null)	(null)
6	7	N	(null)	(null)
7	8	N	(null)	(null)
8	9	N	(null)	(null)
9	10	N	(null)	(null)
10	11	N	(null)	(null)
11	12	N	(null)	(null)
12	13	N	(null)	(null)
13	14	N	(null)	(null)



# IDRT: Realization of integration

Job in Talend Open Studio

- Mission: Import wide-spread medical terminologies for diagnosis, procedures, laboratory observations and administrative code systems like diagnosis-related groups
- Results: Job in Talend Open Studio generating i2b2 ontologies:
  - ↪ §21 code lists
  - ↪ German DRG
  - ↪ ICD-10-GM
  - ↪ OPS
  - ↪ LOINC
  - ↪ TNM
  - ↪ ICD-O

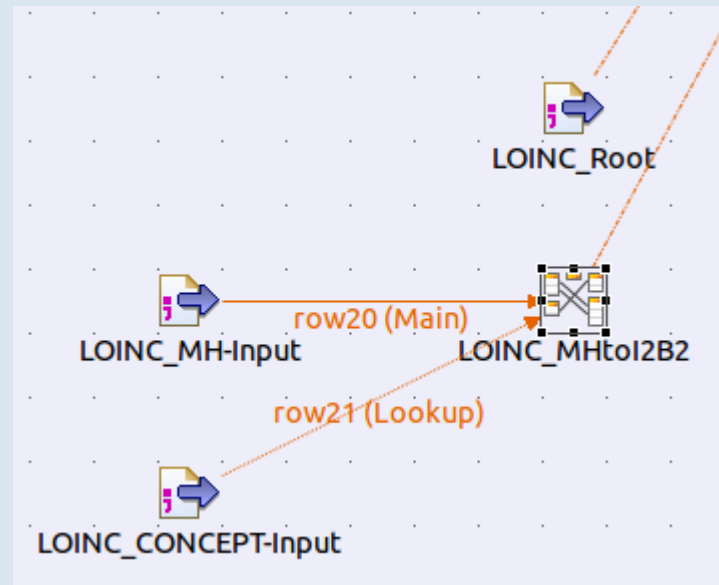




# Example 1: LOINC import

Step (1): Using the normative files from Regenstrief as input

- Only three prerequisites:
  1. LOINC\_Root: a manual-created file containing the root node for the LOINC hierarchy
  2. LOINC\_MH-Input: a CSV file containing the LOINC multi-axial hierarchy
  3. LOINC\_CONCEPT-Input: a CSV file containing the actual LOINC concepts
- The only Talend component used is the mighty tMap





# Example 1: LOINC import

## Step (2): Inside the tMap – joining concepts and hierarchy

The screenshot displays the tMap interface with three main panels:

- Left Panel (Schema editor):** Shows the schema for **row20** with columns: PATH\_TO\_ROOT, SEQUENCE, IMMEDIATE\_PARENT, CODE, and CODE\_TEXT. Below it, the **row21** properties are listed: Lookup Model (Load once), Match Model (Unique match), Join Model (Left Outer Join), and Store temp data (false). The **Expr. key** section shows **row20.CODE** mapped to **LOINC\_NUM**.
- Middle Panel (Var):** A table of variables:
 

Expression	Type	N	Variable
row20.PATH_TO_ROOT.equals	Integer	<input checked="" type="checkbox"/>	i2b2Level
row20.SEQUENCE.length()==1	String	<input checked="" type="checkbox"/>	sequence
row20.PATH_TO_ROOT.equals	String	<input type="checkbox"/>	i2b2pathToRoot
row20.CODE.startsWith("LP")	String	<input type="checkbox"/>	i2b2VisualAttribu
- Right Panel (Expression editor):** Shows the schema for **LOINC\_FolderLeaf** with columns: C\_HLEVEL, C\_FULLNAME, C\_NAME, C\_SYNONYM\_CD, C\_VISUALATTRIBUTES, C\_TOTALNUM, C\_BASECODE, C\_METADATAXML, C\_FACTTABLECOLUMN, C\_TABLENAME, C\_COLUMNNAME, C\_COLUMNDATATYPE, C\_OPERATOR, C\_DIMCODE, C\_COMMENT, C\_TOOL TIP, M\_APPLIED\_PATH, UPDATE\_DATE, DOWNLOAD\_DATE, IMPORT\_DATE, SOURCESYSTEM\_CD, VALUETYPE\_CD, M\_EXCLUSION\_CD, and C\_PATH. The **Expression** column contains various SQL-like expressions such as `Var.i2b2Level`, `"\\i2b2\\ST\\LOINC" + "\\| + Var.i2b2pathToRo`, and `row21.LONG_COMMON_NAME`.

At the bottom, two data tables are visible:

Column	Key	Type	Nullable	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
PATH_TO_ROOT	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		49	0		
SEQUENCE	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		2	0		
IMMEDIATE_PARENT	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		0	0		
CODE	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>		9	0		
CODE_TEXT	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		36	0		

Column	Key	Type	Nullable	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
C_HLEVEL	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		1	0		
C_FULLNAME	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		1000	0		
C_NAME	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		500	0		
C_SYNONYM_CD	<input type="checkbox"/>	Character	<input checked="" type="checkbox"/>		1	0		
C_VISUALATTRIBUTES	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		3	0		
C_TOTALNUM	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		0	0		
C_BASECODE	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		200	0		



# Example 2: ICD-O-3 import

## Step (1): Transform from ClaML into a simple XML



```

44 <xsl:template match="/">
45
46 <!-- xsl:text instructions provide nice formatting for generated code -->
47 <xsl:text>{#xA}</xsl:text><icdo><xsl:text>{#xA}</xsl:text>
48
49 <!-- Only <Class> elements that have SubClass elements are parsed into contents -->
50 <xsl:for-each select="//Class[SubClass]">
51
52 <!-- Add surrounding <class> element -->
53 <xsl:text>{#9}</xsl:text><class><xsl:text>{#xA}</xsl:text>
54
55 <!-- Add <code> element-->
56 <xsl:text>{#9}</xsl:text><code><xsl:value-of select="@code"/>
57 </code><xsl:text>{#xA}</xsl:text>
58
59 <!-- Add <label> element-->
60 <xsl:text>{#9}</xsl:text><label><xsl:value-of select=
61 "Rubric[@kind='preferred']/Label"/></label><xsl:text>{#xA}</xsl:text>
62
63 <!-- Add <path> element-->
64 <xsl:text>{#9}</xsl:text><path>
65
66 <!-- Write path prefix and call template "i2b2path"-->
67 <xsl:text>\i2b2\ST\ICD-O</xsl:text>
68 <xsl:call-template name="i2b2path">
69 <xsl:with-param name="code" select="@code"/>
70 </xsl:call-template>
71 <xsl:text>\</xsl:text>
72 </path><xsl:text>{#xA}</xsl:text>

```

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE ClaML SYSTEM "ClaML.dtd">
3 <ClaML version="2.0.0">
4
25 <Class code="T" kind="chapter">
43 <Class code="C00-C14" kind="block">
44 <SuperClass code="T"/>
45 <SubClass code="C00"/>
46 <SubClass code="C01"/>
47 <SubClass code="C02"/>
48 <SubClass code="C03"/>
60 <Rubric kind="preferred">
61 <Label xml:lang="de">Lippe, Mundhöhle und Pharynx</Label>
62 </Rubric>
63 </Class>
64 <Class code="C00" kind="category">
82 <Class code="C00.0" kind="category">
94 <Class code="C00.1" kind="category">
106 <Class code="C00.2" kind="category">
112 <Class code="C00.3" kind="category">

```

- Normative ClaML (Classification Markup Language) input file

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <icd>
3 <class>
4 <code>T</code>
5 <label>Topographie</label>
6 <path>\i2b2\ST\ICD-O\T</path>
7 <level>4</level>
8 </class>
9 <class>
10 <code>C00-C14</code>
11 <label>Lippe, Mundhöhle und Pharynx</label>
12 <path>\i2b2\ST\ICD-O\T\C00-C14</path>
13 <level>5</level>
14 </class>
15 <class>
16 <code>C00</code>
17 <label>Lippe</label>
18 <path>\i2b2\ST\ICD-O\T\C00-C14\C00</path>
19 <level>6</level>

```

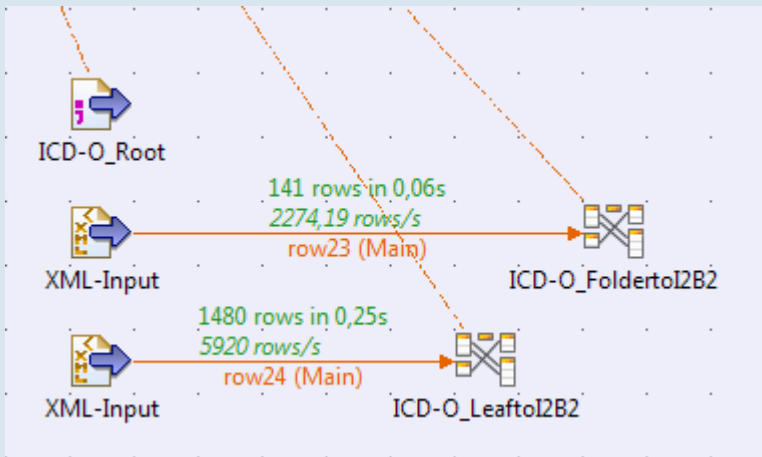
- XSL-Stylesheets for generating i2b2 folder and leafs

- XML intermediate format



# Example 2: ICD-O-3 import

## Step (2): Map simple XML to i2b2 database columns



```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <icd>
3   <class>
4     <code>T</code>
5     <label>Topographie</label>
6     <path>\i2b2\ST\ICD-O\T</path>
7     <level>4</level>
8   </class>
9   <class>
10    <code>C00-C14</code>
11    <label>Lippe, Mundhöhle und Pharynx</label>
12    <path>\i2b2\ST\ICD-O\T\C00-C14</path>
13    <level>5</level>
14  </class>
15  <class>
16    <code>C00</code>
17    <label>Lippe</label>
18    <path>\i2b2\ST\ICD-O\T\C00-C14\C00</path>
19    <level>6</level>

```



Talend Open Studio for Data Integration - tMap - tMap\_18

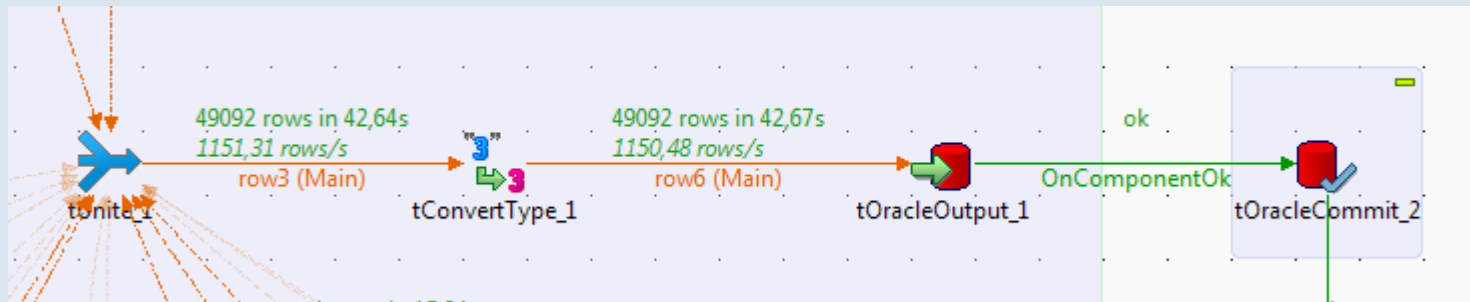
Auto map!

Column	Column
code	C_HLEVEL
label	C_FULLNAME
path	C_NAME
level	C_SYNONYM_CD
	C_VISUALATTRIBU...
	C_TOTALNUM
	C_BASECODE
	C_METADATAXML
"concept_cd"	C_FACTTABLECOL...
"concept_dimension"	C_TABLENAME
"concept_path"	C_COLUMNNAME
'T'	C_COLUMNDATAT...
"LIKE"	C_OPERATOR
row23.path	C_DIMCODE
	C_COMMENT
row23.label	C_TOOLTIP
'@'	M_APPLIED_PATH
TalendDate.getCurrentDate()	UPDATE_DATE
	DOWNLOAD_DATE
TalendDate.getCurrentDate()	IMPORT_DATE
"ICD-O"	SOURCESYSTEM_CD
	VALUETYPE_CD
	M_EXCLUSION_CD
	C_PATH
	C_SYMBOL

- XML intermediate format (continued)

- Schema mapping to i2b2





Schema vontOracleOutput\_1

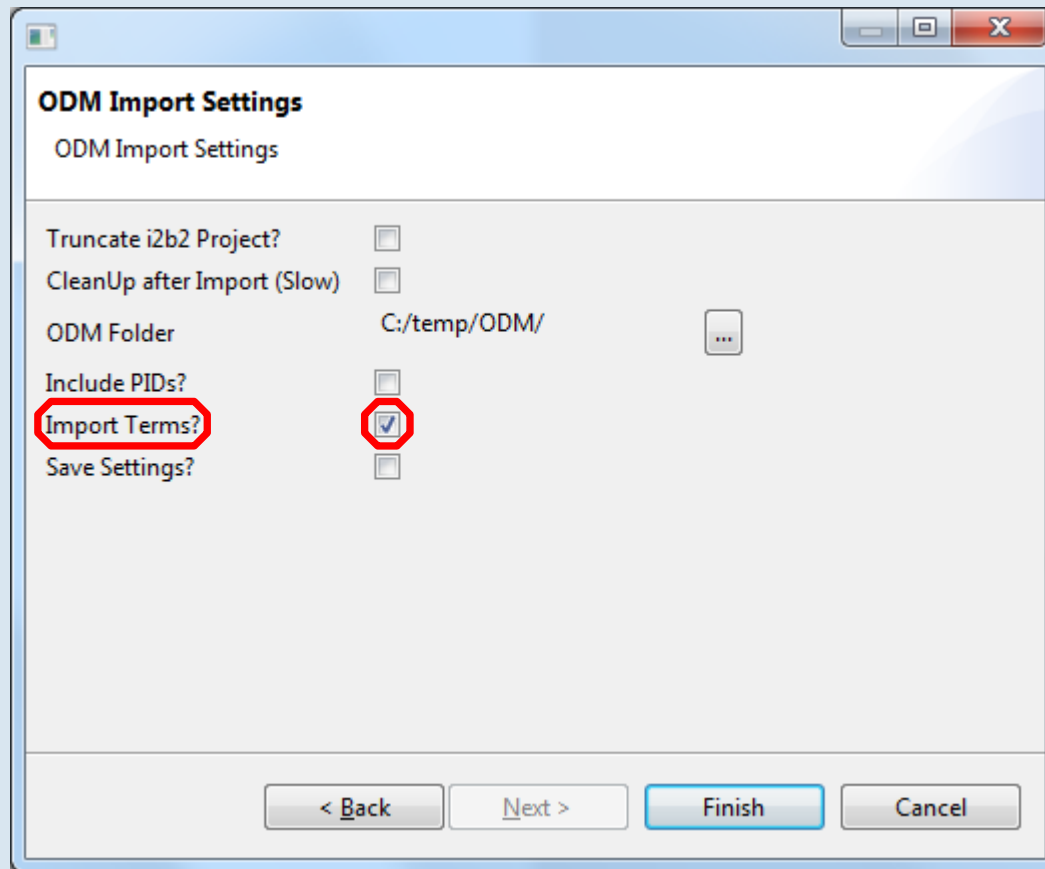
tConvertType_1 (Input - Main)							tOracleOutput_1 (Output)														
Spalte	Sc...	Typ	<input checked="" type="checkbox"/>	N..	Datumsfor...	Länge	Gen...	Def...	Com...	Spalte	Db Column	Sc...	Typ	Db Type	<input checked="" type="checkbox"/>	N..	Datums...	Lä...	Ge...	D...	Co
C_HLEVEL		Integer	<input checked="" type="checkbox"/>			1	0			C_HLEVEL	C_HLEVEL	<input checked="" type="checkbox"/>	Int...	INT	<input checked="" type="checkbox"/>			1	0		
C_FULLNAME		String	<input checked="" type="checkbox"/>			1000	0			C_FULLNA...	C_FULLNAME	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			1000	0		
C_NAME		String	<input checked="" type="checkbox"/>			500	0			C_NAME	C_NAME	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			500	0		
C_SYNONYM_CD		Char...	<input checked="" type="checkbox"/>			1	0			C_SYNON...	C_SYNONYM...	<input checked="" type="checkbox"/>	Ch...	CHAR	<input checked="" type="checkbox"/>			1	0		
C_VISUALATTRI...		String	<input checked="" type="checkbox"/>			3	0			C_VISUAL...	C_VISUALATT...	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			3	0		
C_TOTALNUM		String	<input checked="" type="checkbox"/>				0			C_TOTAL...	C_TOTALNUM	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>				0		
C_BASECODE		String	<input checked="" type="checkbox"/>			200	0			C_BASECO...	C_BASECODE	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			200	0		
C_METADATAX...		String	<input checked="" type="checkbox"/>			4000	0			C_METAD...	C_METADAT...	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			4000	0		
C_FACTTABLEC...		String	<input checked="" type="checkbox"/>			100	0			C_FACTTA...	C_FACTTABL...	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			100	0		
C_TABLENAME		String	<input checked="" type="checkbox"/>			100	0			C_TABLEN...	C_TABLENAME	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			100	0		
C_COLUMNNA...		String	<input checked="" type="checkbox"/>			100	0			C_COLUM...	C_COLUMNN...	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			100	0		
C_COLUMNNDAT...		Char...	<input checked="" type="checkbox"/>			100	0			C_COLUM...	C_COLUMNND...	<input checked="" type="checkbox"/>	Ch...	CHAR	<input checked="" type="checkbox"/>			100	0		
C_OPERATOR		String	<input checked="" type="checkbox"/>			10	0			C_OPERAT...	C_OPERATOR	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			10	0		
C_DIMCODE		String	<input checked="" type="checkbox"/>			1000	0			C_DIMCODE	C_DIMCODE	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			1000	0		
C_COMMENT		String	<input checked="" type="checkbox"/>				0			C_COMM...	C_COMMENT	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>				0		
C_TOOLTIP		String	<input checked="" type="checkbox"/>			500	0			C_TOOLTIP	C_TOOLTIP	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			500	0		
M_APPLIED_PATH		Char...	<input checked="" type="checkbox"/>			1000	0			M_APPLIE...	M_APPLIED_P...	<input checked="" type="checkbox"/>	Ch...	CHAR	<input checked="" type="checkbox"/>			1000	0		
UPDATE_DATE		Date	<input checked="" type="checkbox"/>		"yyyy-MM...	19	0			UPDATE_D...	UPDATE_DATE	<input checked="" type="checkbox"/>	Date	DATE	<input checked="" type="checkbox"/>		"yyyy-...	19	0		
DOWNLOAD_D...		Date	<input checked="" type="checkbox"/>		"yyyy-MM...	19	0			DOWNLO...	DOWNLOAD_...	<input checked="" type="checkbox"/>	Date	DATE	<input checked="" type="checkbox"/>		"yyyy-...	19	0		
IMPORT_DATE		Date	<input checked="" type="checkbox"/>		"yyyy-MM...	19	0			IMPORT_D...	IMPORT_DATE	<input checked="" type="checkbox"/>	Date	DATE	<input checked="" type="checkbox"/>		"yyyy-...	19	0		
SOURCESYSTEM...		String	<input checked="" type="checkbox"/>			100	0			SOURCESY...	SOURCESYST...	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			100	0		
VALUETYPE_CD		String	<input checked="" type="checkbox"/>			100	0			VALUETYP...	VALUETYPE_CD	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			100	0		
M_EXCLUSION_...		String	<input checked="" type="checkbox"/>			100	0			M_EXCLUS...	M_EXCLUSIO...	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			100	0		
C_PATH		String	<input checked="" type="checkbox"/>			100	0			C_PATH	C_PATH	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			100	0		
C_SYMBOL		String	<input checked="" type="checkbox"/>			100	0			C_SYMBOL	C_SYMBOL	<input checked="" type="checkbox"/>	Str...	VARC...	<input checked="" type="checkbox"/>			100	0		



# But users won't see much of this

Job is encapsulated in an dedicated import tool

- Import is done with just one click
- Only terminologies with corresponding factual data are imported



- All terminologies were imported successfully
  - ↪ i2b2 has only simple needs (Concept: label + code, Concepts relate to each other <broader> or <narrower>)
  - ↪ Talend Open Studio is fast and flexible enough to do the job
  - ↪ Import (125.000 concepts) takes one minute on our server and about 20 seconds on a subnotebook
  
- Problems
  - ↪ Modifier: Primary/secondary diagnosis is not distinguished
  - ↪ Versions: Most terminologies are updated once or twice a year, codes retire and reincarnate
  - ↪ Format: Some terminologies come as colored Excel files with textual annotations, requiring a manual cleaning

- More terminologies will be imported at request
- Planning for IDRT2:
  - ↳ Fine-grained mapping capabilities (Göttingen i2b2 ontology editor)
  - ↳ Connection to the National Metadata Repository (reading i2b2 ontologies from harmonized datasets)

