First European i2b2 Academic User Meeting

IDRT: Platform Architecture And Tools to Support The Re-use of Routine Clinical Data For Research

The IDRT Team (in alphabetical order):

Christian Bauer, Benjamin Baum, Jan Christoph, Igor Engel, Thomas Ganslandt, Matthias Löbe, **Sebastian Mate (presenter)**, Daniel Plog, Hans-Ulrich Prokosch, Matthias Quade, Ulrich Sax, Sebastian Stäubert, Lars Voitel, Alfred Winter

Introduction

Secondary use of routine clinical data provides a big chance for medical research:

To support clinical trials

- accurate determination and exploitation of the number of study patients through feasibility analysis and recruitment support
- optimize study design with retro perspective analysis
- re-use of routine data as a part of the study documentation

To support projects with third-party funding

- e.g. Harvard hospitals (2005)[1]: 40% of the running projects use clinical warehouse data, received funding >130 million USD

Competitive advances

- through exploitation of routine care data
- adequate infrastructure and access to data is necessary

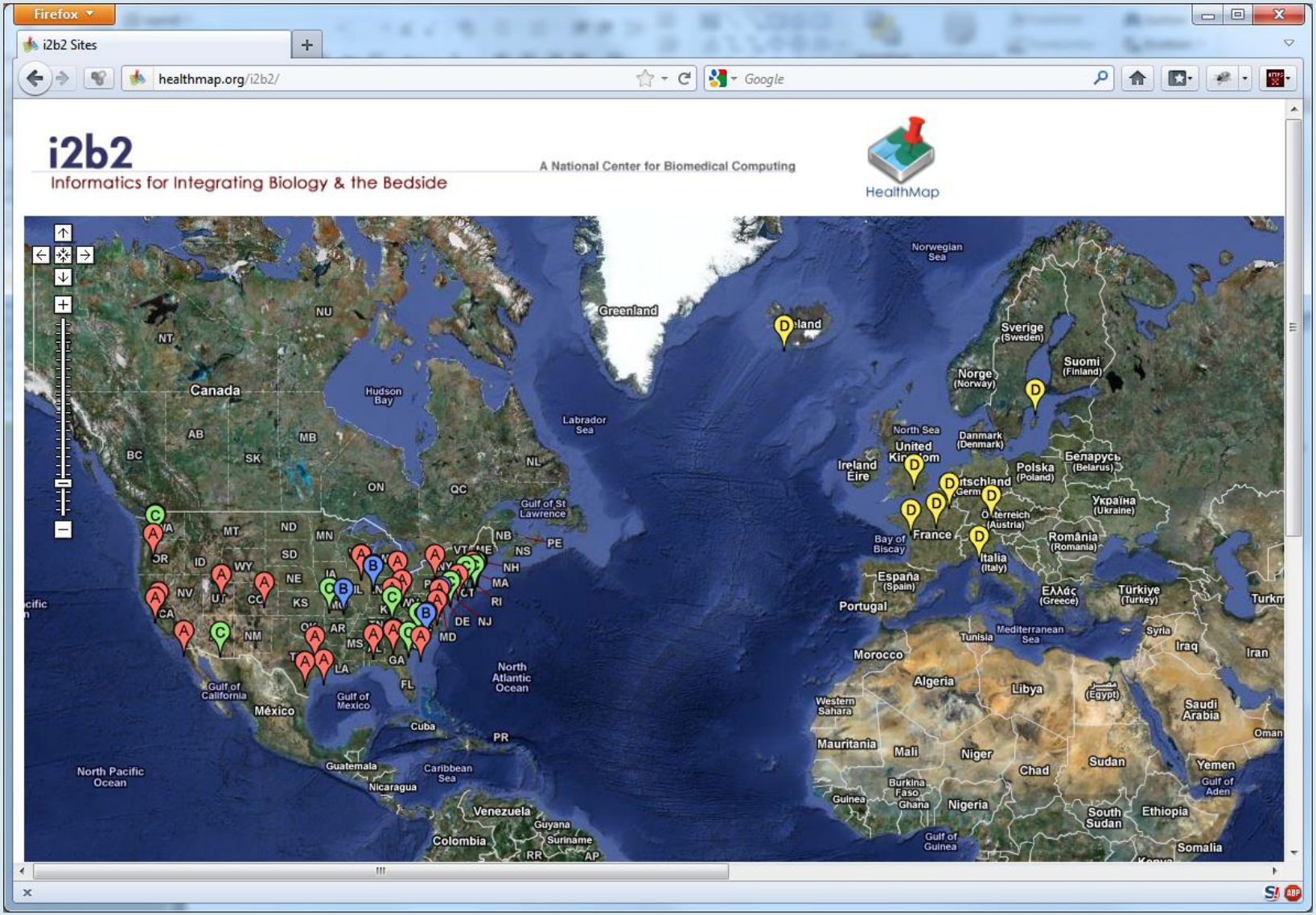
[1] Nalichowski R. Calculating the Benefits of a Research Patient Data Repository. Proc AMIA Symp. 2006;1044

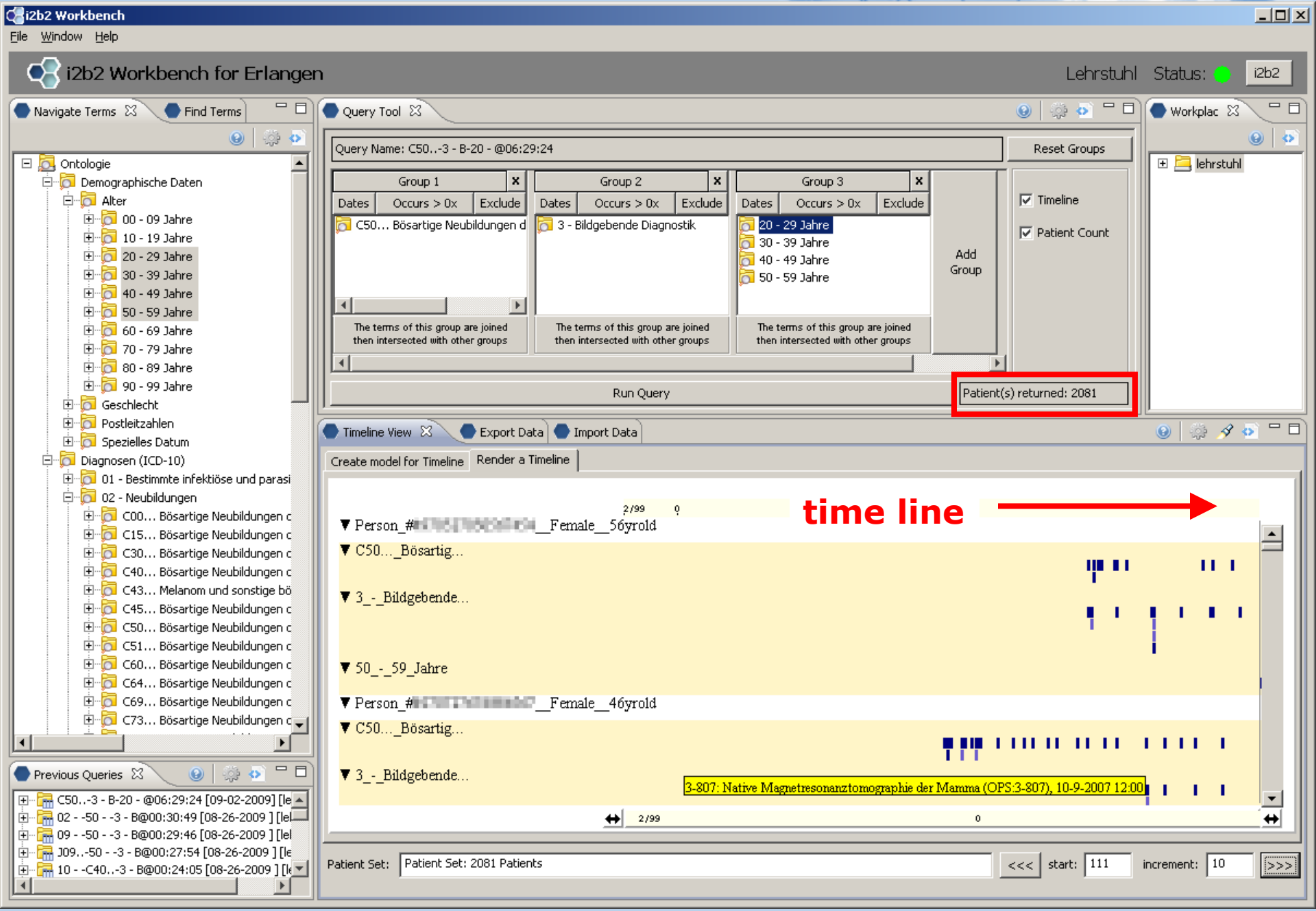
Introduction

Short history of i2b2

- **i2b2 is the "open source variant" of the RPDR (Patient Research Data Registry)**
  - developed since 1999 at the Massachusetts General Hospital (MGH) and Partners Healthcare, Inc., which is in productive use since 2002
  - 2005: already 1073 users
  - used in several hospitals: MGH, BWH, FH, SRH, NWH
- **2004: request for applications by the NIH as part of their "Roadmap for National Centers for Biomedical Computing (NCBC)"**
- **Birth of i2b2 (Informatics for Integrating Biology and the Bedside): Application of i2b2 by I. Kohane (Harvard Medical School) and J. Glaser (CIO of Partners Healthcare)**
  => received funding in 2005 as one of the NCBCs
- **Since 2008: i2b2 is available for download**
- **2010: second funding phase**

Introduction




i2b2 Workbench
File Window Help
i2b2 Workbench for Erlangen
Lehrstuhl Status:
i2b2
Navigate Terms
Find Terms
Ontologie
Demographische Daten
Alter
00 - 09 Jahre
10 - 19 Jahre
20 - 29 Jahre
30 - 39 Jahre
40 - 49 Jahre
50 - 59 Jahre
60 - 69 Jahre
70 - 79 Jahre
80 - 89 Jahre
90 - 99 Jahre
Geschlecht
Postleitzahlen
Spezielles Datum
Diagnosen (ICD-10)
01 - Bestimmte infektiöse und parasi
02 - Neubildungen
C00... Bösartige Neubildungen c
C15... Bösartige Neubildungen c
C30... Bösartige Neubildungen c
C40... Bösartige Neubildungen c
C43... Melanom und sonstige bö
C45... Bösartige Neubildungen c
C50... Bösartige Neubildungen c
C51... Bösartige Neubildungen c
C60... Bösartige Neubildungen c
C64... Bösartige Neubildungen c
C69... Bösartige Neubildungen c
C73... Bösartige Neubildungen c
Previous Queries
C50..-3 - B-20 - @06:29:24 [09-02-2009] [le
02 - -50 - -3 - B@00:30:49 [08-26-2009] [le
09 - -50 - -3 - B@00:29:46 [08-26-2009] [le
J09..-50 - -3 - B@00:27:54 [08-26-2009] [le
10 - -C40..-3 - B@00:24:05 [08-26-2009] [le
Query Tool
Query Name: C50..-3 - B-20 - @06:29:24
Reset Groups
Group 1
Group 2
Group 3
Dates
Occurs > 0x
Exclude
C50... Bösartige Neubildungen d
3 - Bildgebende Diagnostik
20 - 29 Jahre
30 - 39 Jahre
40 - 49 Jahre
50 - 59 Jahre
The terms of this group are joined then intersected with other groups
Add Group
Timeline
Patient Count
Run Query
Patient(s) returned: 2081
Workplac
lehrstuhl
Timeline View
Export Data
Import Data
Create model for Timeline
Render a Timeline
time line
Person_#__Female__56yrold
C50..._Bösartig...
3_-_Bildgebende...
50_-_59_Jahre
Person_#__Female__46yrold
C50..._Bösartig...
3_-_Bildgebende...
3-807: Native Magnetresonanztomographie der Mamma (OPS:3-807), 10-9-2007 12:00
Patient Set:
Patient Set: 2081 Patients
<<<
start:
111
increment:
10
>>>

Introduction

i2b2's difficulties from a user's point of view:

Installation & configuration

- complex: although a comprehensive documentation exists, much can go wrong

Import of clinical data

- no integrated support (except for data in i2b2 format)
- especially not for common standard formats (ODM, CSV, §21)

Import of German standard terminologies and non-standard terms

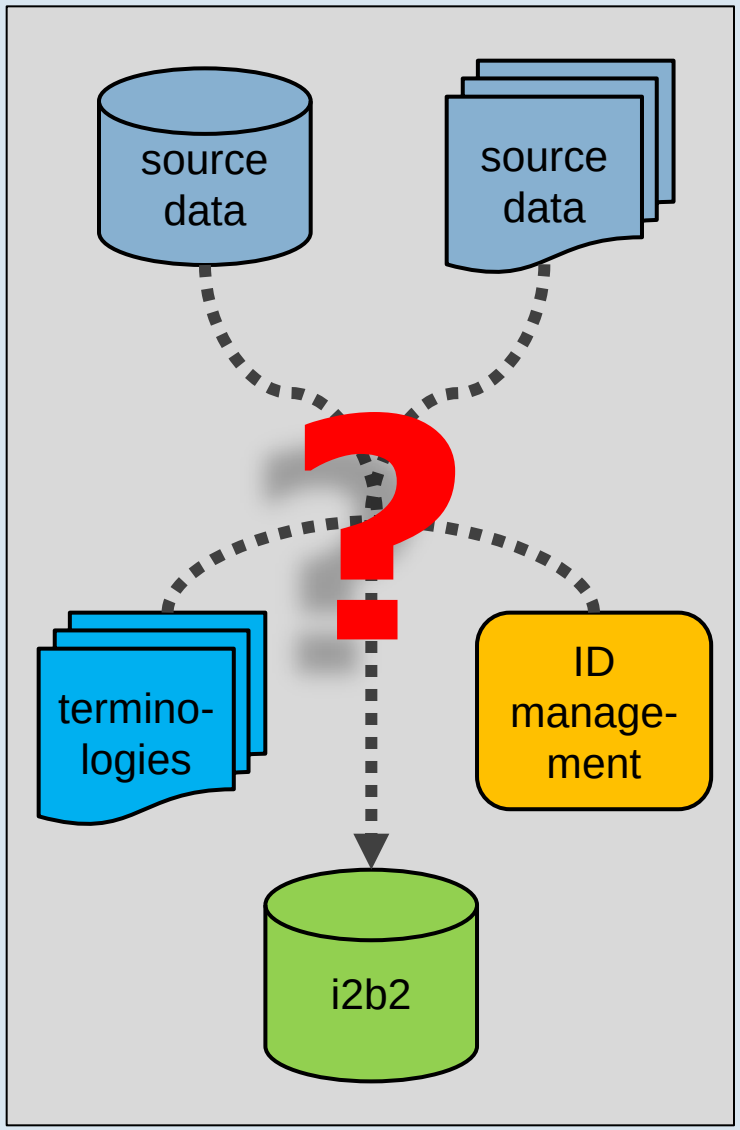
- i2b2 demo data only contains some US standard terminologies
- import to i2b2 has to be re-invented always again, although:
  - standard terminologies already exist in all hospitals (supplied by DIMI: ICD, OPS, …)
  - other standard formats (e.g. ODM, CSV) already include their metadata terms

  => why not develop a standard methodology to import those common terminologies?

Bearing German data protection and privacy in mind:

- Standardized creation of pseudonyms and record linkage based on demographic data

Goals



Courtesy of Thomas Ganslandt

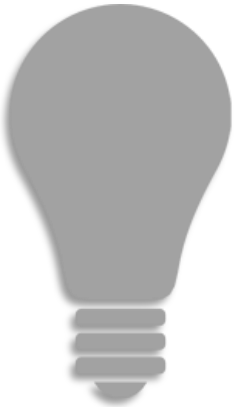
Basic Workplan

Installation & Administration

Improve i2b2 Wizard

ETL for Common Data Formats

ODM, §21, CSV, SQL
Easy IDRT Import Tool
Flexible & powerful
Talend Jobs



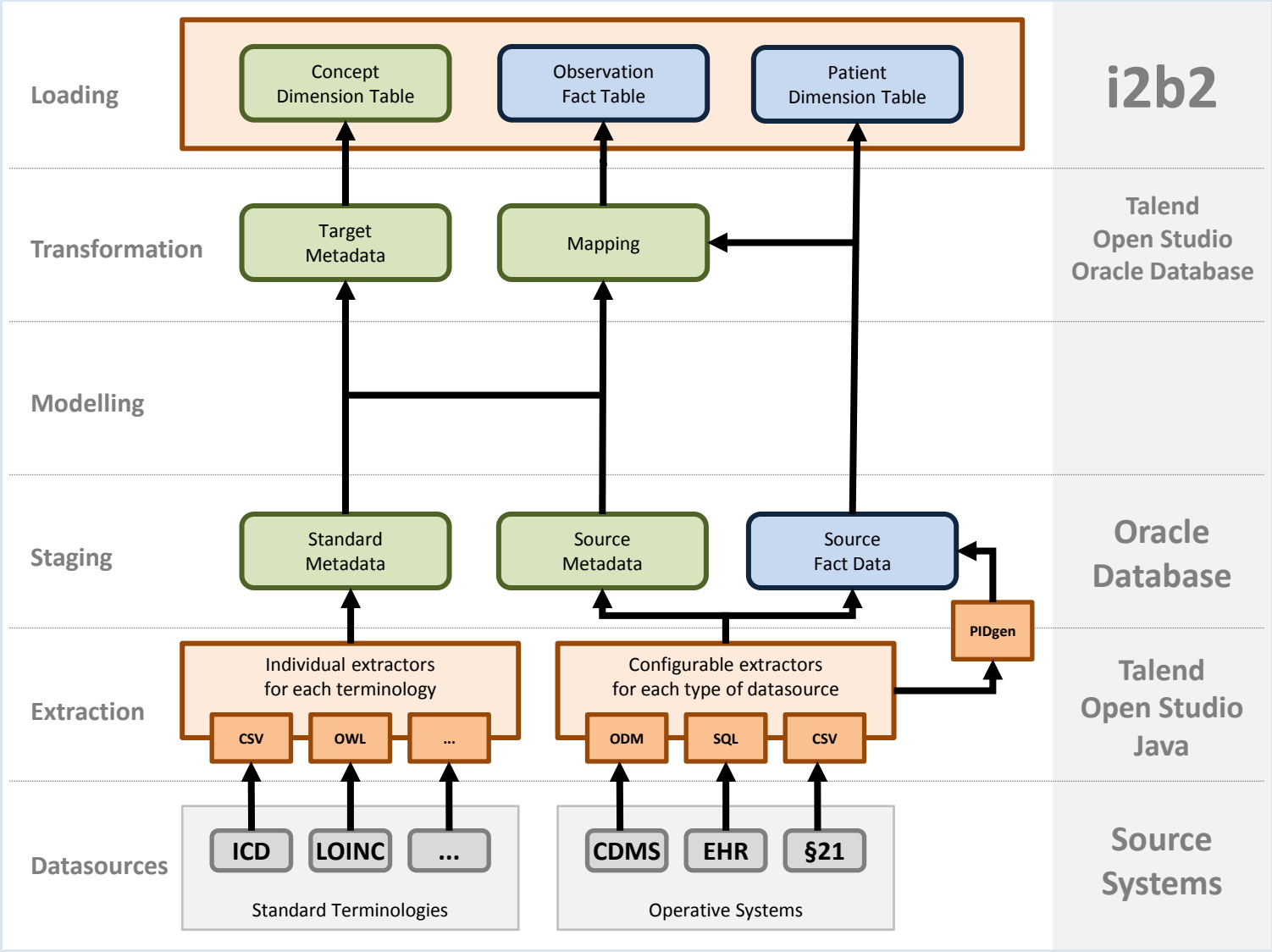
Semantic Integration

Provide standard terminologies for i2b2 and automatic mapping of data during import

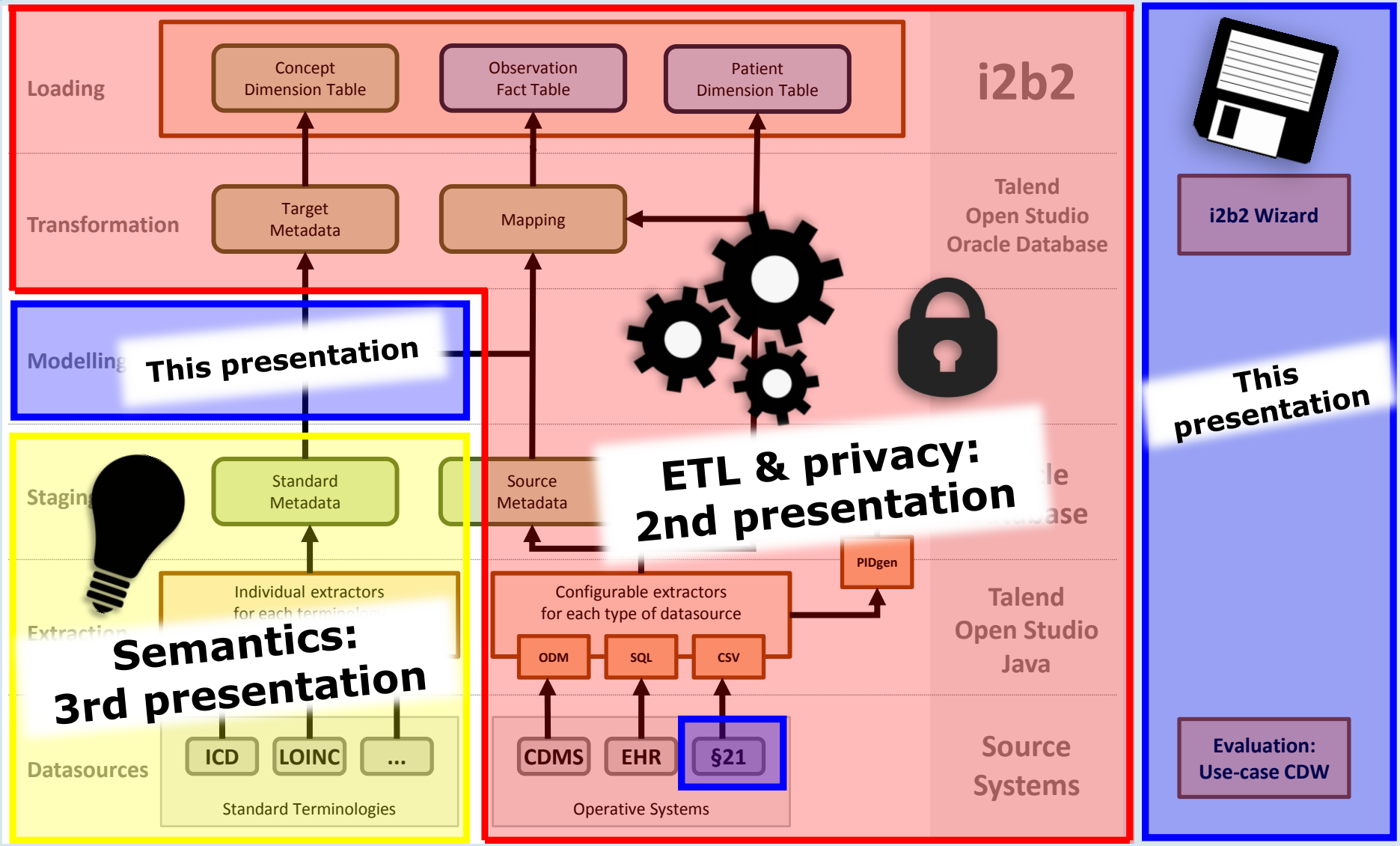
Data Protection & Privacy

Integration of TMF PID-Generator

IDRT Architecture

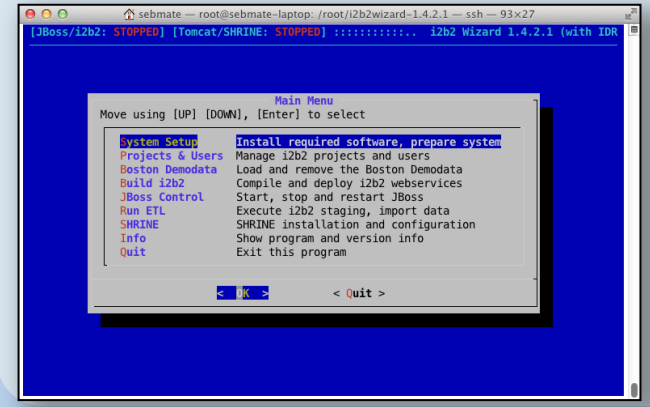


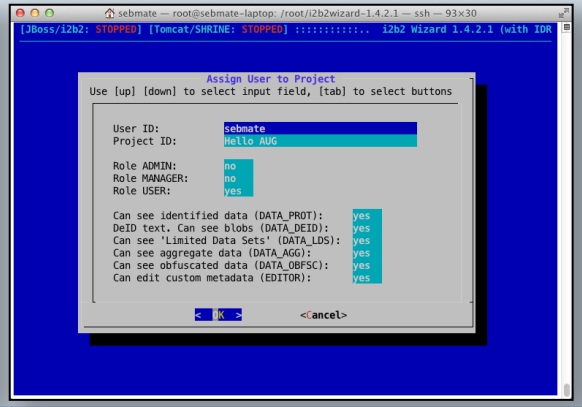
IDRT Architecture

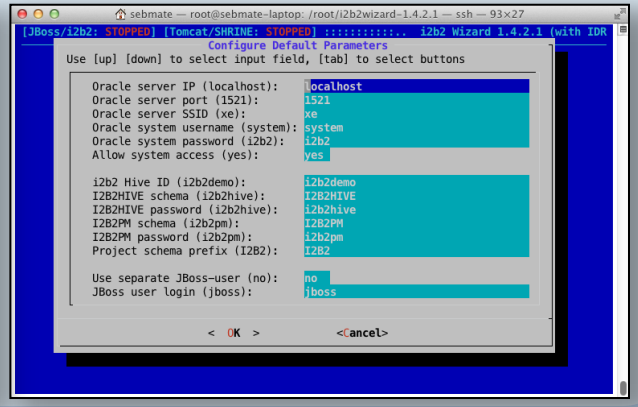


i2b2 Wizard

- **"Semi-graphical" program to install and configure i2b2**
- **Full i2b2 installation support**
  - Automatic download and extraction of required software packages
  - Automatic installation of Linux software packages
  - Automatic configuration and compilation of the i2b2 source code
  - Automatic handling of all database work (schema creation, etc.)
- **Full support for i2b2 administration**
  - Create and remove i2b2 projects and users
  - Load and remove the Boston Demodata
  - Change Hive ID and network interface and database configuration
  - Change various passwords (e. g. from I2B2HIVE and I2B2PM)
  - Recursive resolving of dependencies for setup targets / administration tasks
- **SHRINE installation (currently not up-to-date)**







i2b2 Wizard

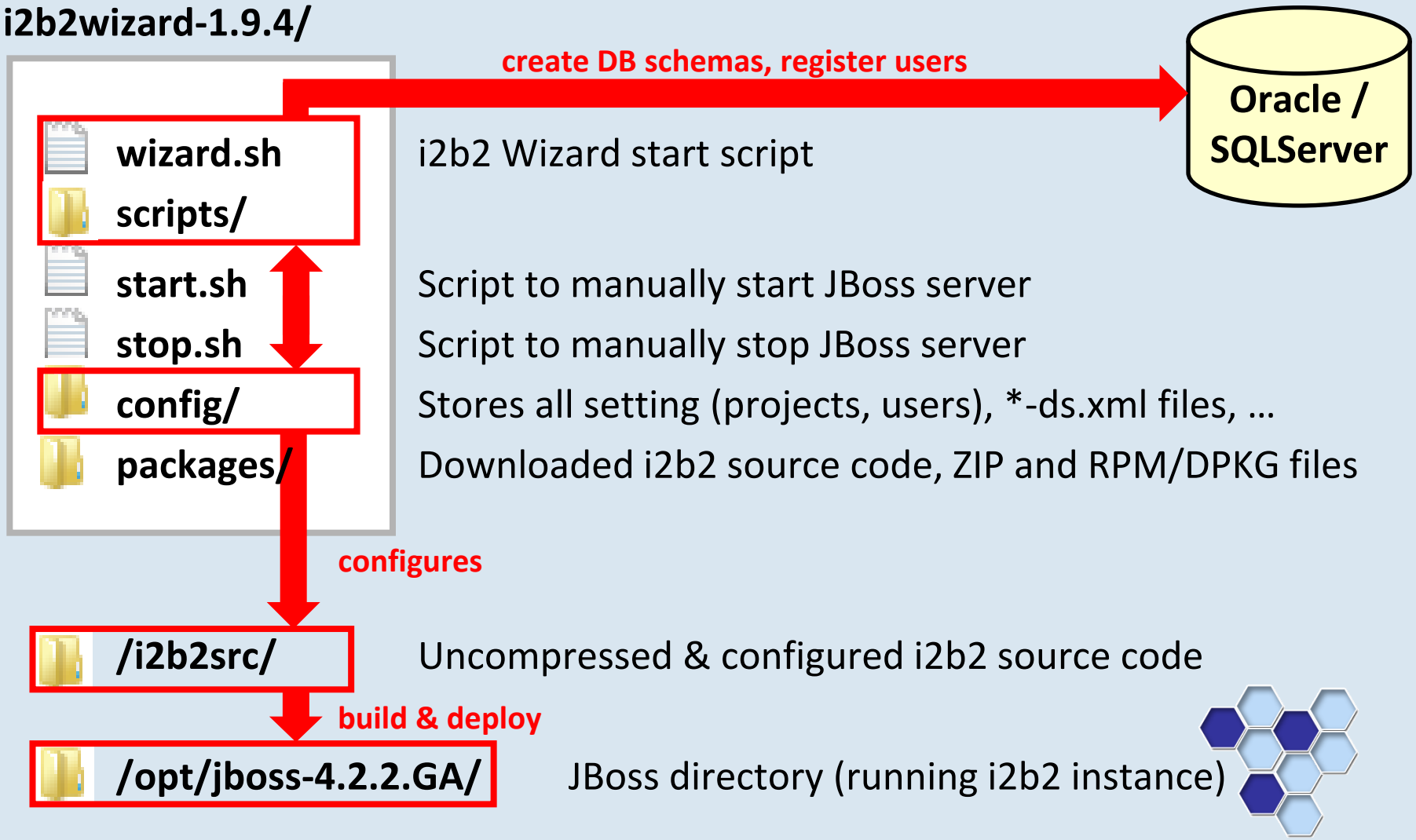
Achieved enhancements in the IDRT project

- Always updated to latest i2b2 versions
- JBoss log monitoring: automatic detection of installation errors
- Automatic JDBC tests to ensure correct database settings
- Custom schema names for I2B2PM and I2B2HIVE
- ***Many*** minor changes and improvements

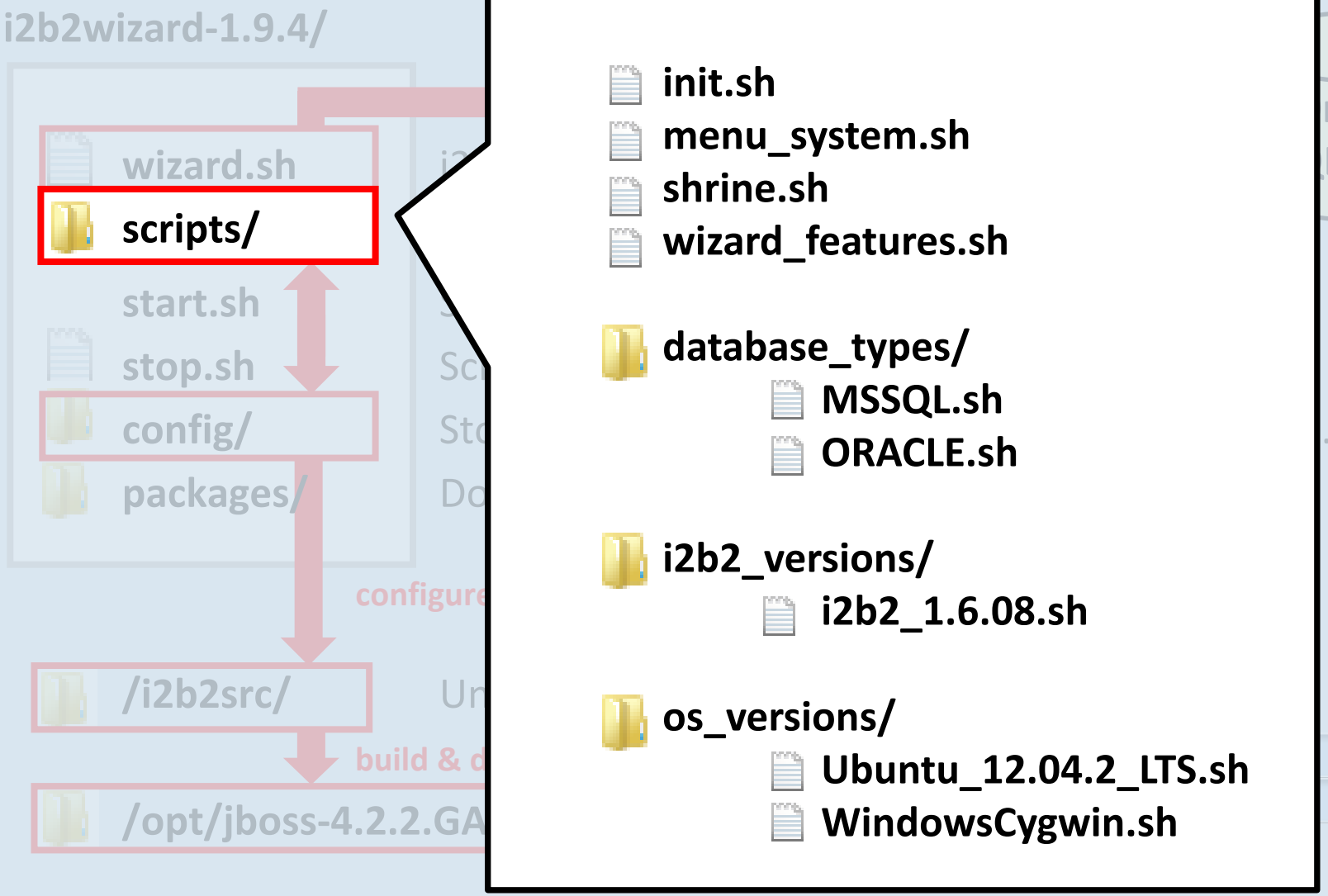
Current development: modular Version 2.0 (current beta is 1.9.4)

- Will support different i2b2 versions
- Will support different operating systems (Linux & Windows/Cygwin)
- Will support different database systems (Oracle & Microsoft SQLServer)
- Great backward-compatibility
- Increase community involvement

i2b2 Wizard: Architecture



i2b2 Wizard: Architecture



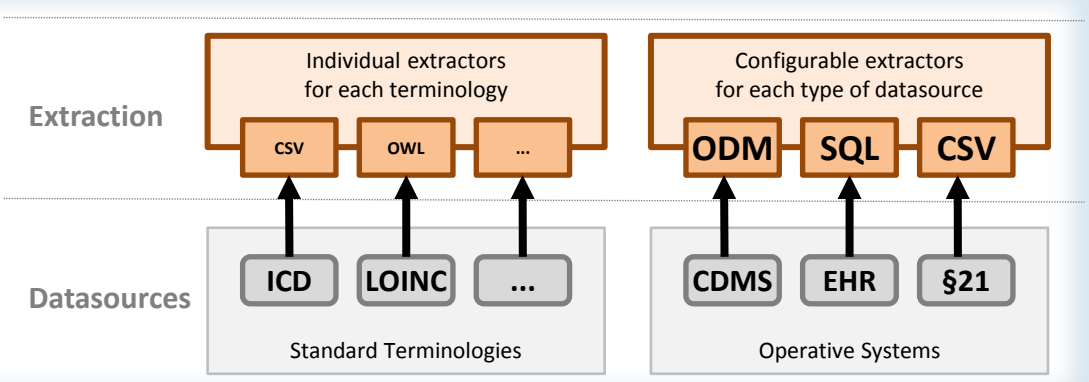
§21 Importer

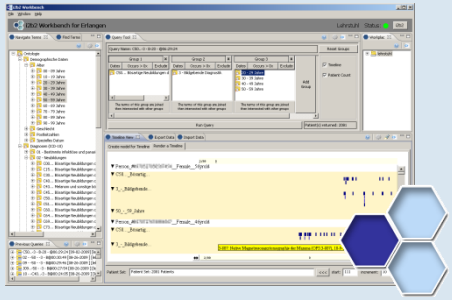
German §21 KHEntG data set

- used by InEK GmbH to create next DRGs and payment rates
- obligatory and therefore used by ***all*** German hospitals
- contains ICD10, OPS (procedures), DRGs, encounter and special hospital data => set of simple CSV files

IDRT §21 Importer: "one mouse click importer"

- reduces difficulties in getting started with i2b2 and IDRT
- makes use of IDRT CSV importer and terminology mapping

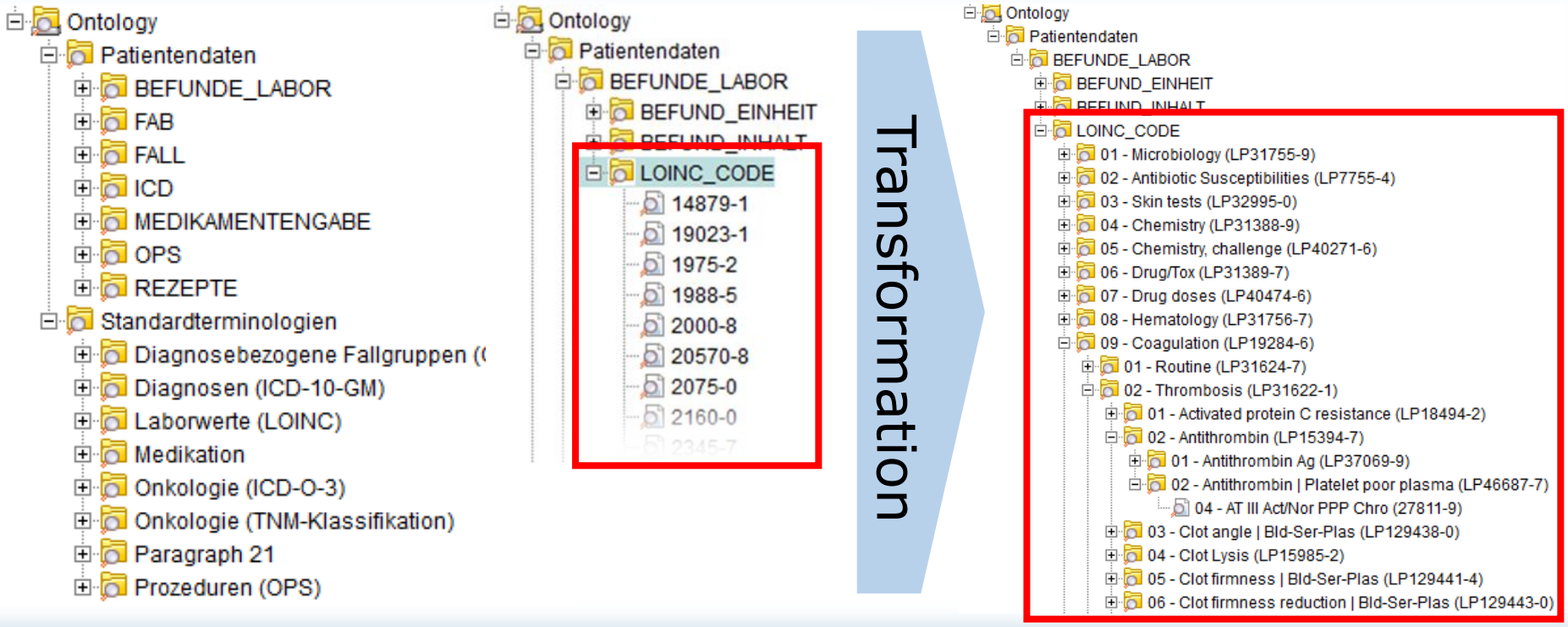




Use-case CDW

Idea: Show that it is possible (and efficient) to combine the simple importers (e.g. §21 importer) with other data imports. Prototypical setup in Erlangen:

1. Import §21 (IDRT §21 importer), lab values and medications (IDRT SQL Importer)
2. Load the corresponding terminologies (by using the IDRT jobs)
3. Run a transformation step to enrich imported data with the terminologies



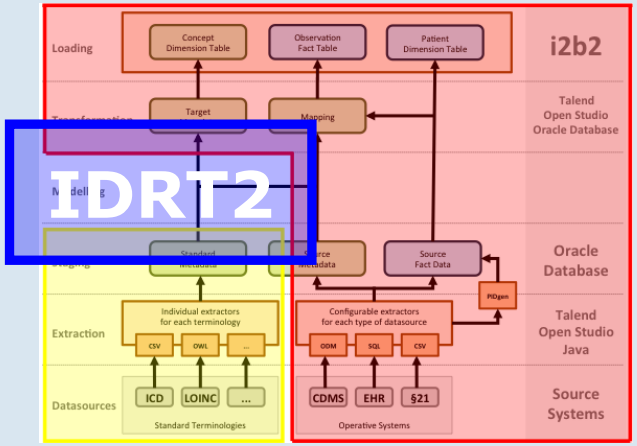
Discussion & Outlook

IDRT adds tremendous value to the standard i2b2 package

- IDRT delivers an easy-to-use package to install i2b2 and to populate its database with common German data sets
- Flexible approach: “power users” can use the Talend jobs (encapsulated into the IDRT tool) directly in Talend
- Take just what you need: i2b2 Wizard and ETL tools can be used independently

The i2b2 software is evolving: IDRT2

- Add support for i2b2 ontology modifiers (required to group complex data elements such as laboratory values)
- IDRT editor to customize i2b2 ontologies
- Interface to the German MDR project
- Update SHRINE support



Thanks for your attention!

sebastian.mate@imi.med.uni-erlangen.de

For more information about the IRDT project and downloads please visit:

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