346. Import from MDR

Summary

A common problem while integrating data sets from heterogeneous sources in a Data Warehouse is to provide a mapping of data not only on a syntactical level, but also on a **semantical level**. In i2b2, every instance of an observation_fact has an association to an instance of a concept in concept_dimension. i2b2 concepts are rather straightforward. From a terminological perspective, they mainly consist of an identifier and a label. Furthermore, concepts are arranged hierarchically by a broader/narrower relation. In particular, there is no attribute defining the meaning of a concept. So, if there exists a concept *gender* in a clinical trial and a second concept *gender* in a registry, there is no way of recognizing whether they are equivalent. In fact, from i2b2's perspective, they are as much different as any other concept is.

Medical terminologies as human-curated ontologies play an important part in providing a common understanding but in reality, most data elements imported from sources like the clinical data management system, epidemiological registries, survey tools from longitudinal cohorts or the local EHR won't have an explicit annotation referencing terminologies.

A **Metadata Repository** (MDR) is a central registry of such data elements. Various MDRs exist, most popular the NCI caDSR, but also Australia's METeOR .

The IMT provides some support for a prototypical MDR software developed to establish a German National Metadata Repository.

Example

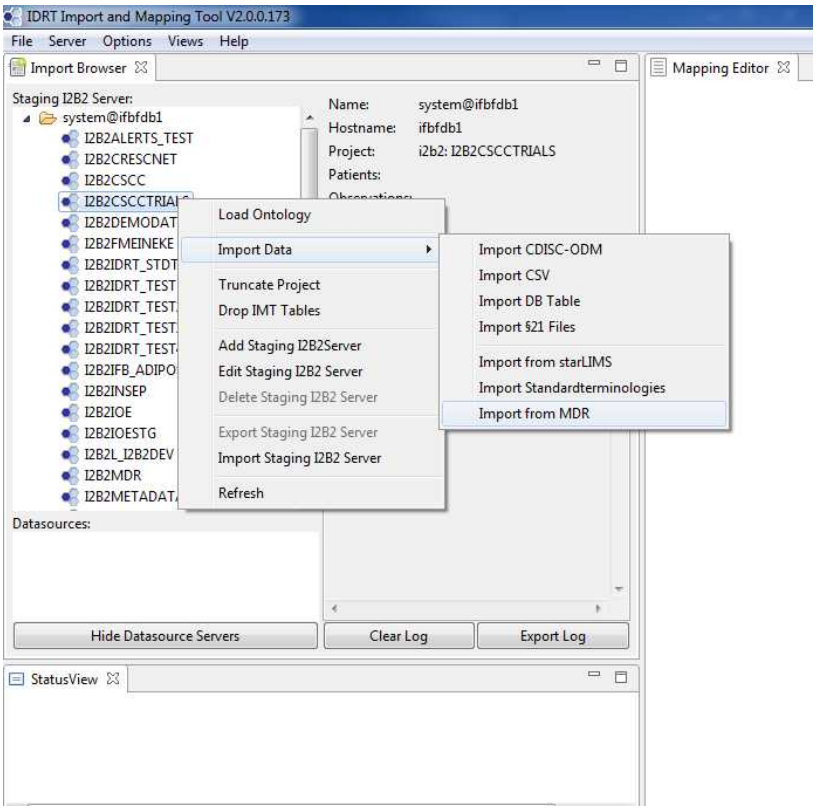
The MDR prototype is currently under evaluation in different projects running different instances. Some contain thousands of data elements from dozens of trials.

The screenshot below shows a graphical representation of a collection of data elements commonly used to measure the severity of a disease for patients admitted to Intensive care units, the *Simplified Acute Physiology Score* (SAPS-II). It is organized in certain submoduls from different domains, some being scores of their own (*Glasgow Coma Scale*).

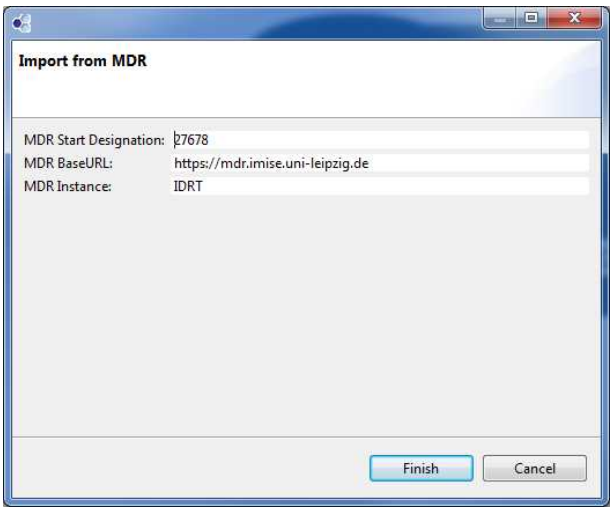
In most cases, a Metadata Repository contains much more details than medical terminologies, e.g. measurement units, lengths and format constraints, datatypes, codes for value sets and more.

Using the MDR with the IMT

The IMT treats the Metadata Repository just like any other data import.



If you choose *Import from MDR*, the following dialog is displayed:



Name	Description
MDR Start Designation	This is the ID of the label that designates the root container you want to import, e.g. a clinical trial specification.
MDR Base URL	The host the MDR is running on.
MDR Instance	In case multiple instance of the MDR are running on the host in different contexts, the right one is specified here.

Outlook

It is planned for the near future, when research grant resources bubble up, to implement a more convenient way of browsing through the MDR by providing access similar to Datasource Servers.