

Project proposal at Philips Research, Eindhoven

TranSMART

Introduction

TranSMART is a knowledge management platform that enables scientists to develop and refine research hypotheses by investigating correlations between genetic and phenotypic data, and assessing their analytical results in the context of published literature and other work¹. A major feature of tranSMART is the Data set Explorer. Data Explorer is an i2b2²-based tool that represents the concepts in tree form and lets one compare two sets of study groups based on one or more points of comparison (e.g. age or gender). One can define both the criteria that populate the study groups and the points of comparison between the study groups.

Data model

TranSMART uses an i2b2 database. The i2b2 data model consists of facts and dimension. A fact is the piece of information being queried, and the dimensions are groups of hierarchies and descriptors that describe the facts (Figure 1). Facts in i2b2 are observations about a patient, including things like diagnoses, demographics or laboratory results.

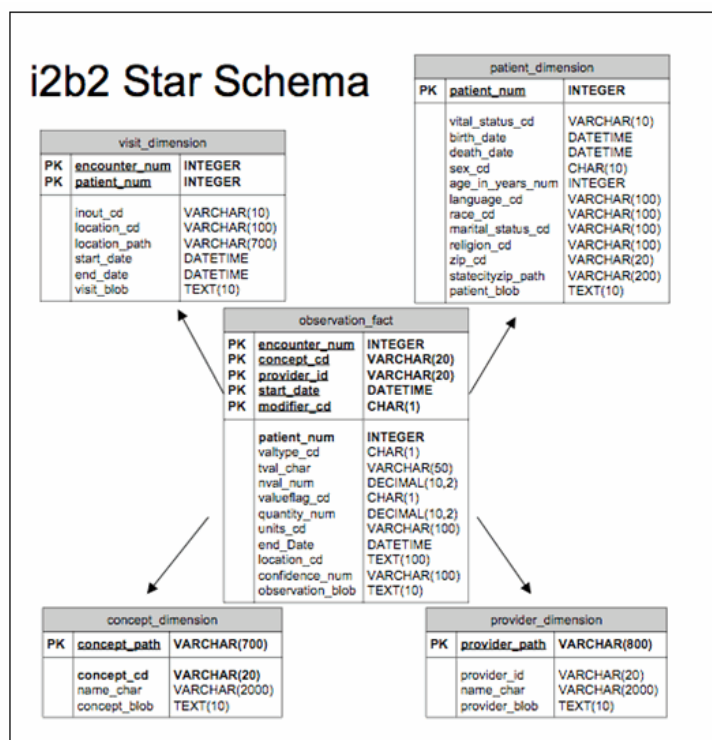


Figure 1 I2B2 star schema

¹ <http://www.transmartproject.org/>

² <https://www.i2b2.org/>

Common Information Model

To provide homogeneous access to different data sources, we, at Philips, use an infrastructure that is based on a Common Information Model (CIM) which will represent, manage and retrieve clinical information stored in our platform. The proposed model is comprised by two main components: the Core Dataset (CD) and the Common Data Model (CDM).

Core Dataset

The CD is the structure that contains a shared medical vocabulary (SNOMED CT³, LOINC⁴ and MedDRA⁵), relationships among concepts and tools to infer knowledge from these ontologies.

Common Data Model

The CDM is the structure that contains the data coming from different sources and follows a common schema including the vocabulary of the CD.

Project proposal

TransSMART does not enforce or uses a standard ontology. The ontology one defines affects how data appears in the Data set Explorer tree. Different instances of transSMART (e.g. running at a company, institute or university) can use different, existing ontologies. Also, it is possible to create a study-driven ontology.

A seminar project could be to investigate whether it is possible to provide a methodology for integrating transSMART data sources with the semantic CIM (Figure 2).

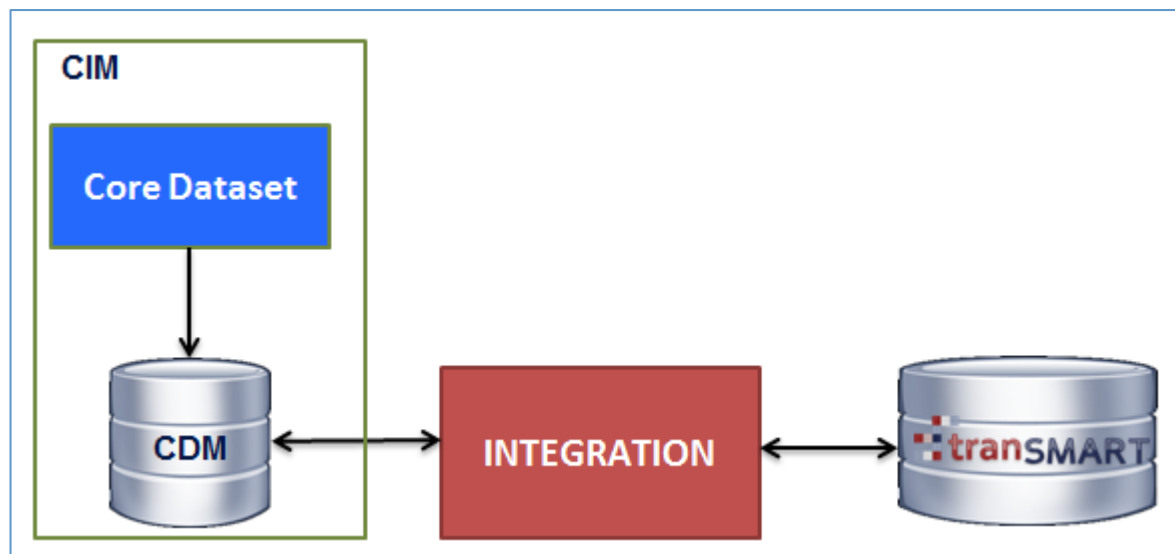


Figure 2 Overview integration CIM and transSMART

³ Systematized Nomenclature Of Medicine Clinical Terms (SNOMED CT), <http://www.ihtsdo.org/snomed-ct/>

⁴ Logical Observation Identifiers Names and Codes (LOINC®) - <http://www.loinc.org>

⁵ Medical Dictionary for Regulatory Activities Maintenance and Support Service Organization (MedDRA), <http://www.meddrasso.com/>