BRIDG HSCT Physical Model: A Model – Driven Architecture for Stem – Cell Data Exchange

Introduction

Hematopoietic Stem Cell Transplantation (HSCT) is the process of transplantation of stem cells extracted from bone marrow, peripheral blood or umbilical cord blood to cure certain types of blood related cancers.

The National Marrow Donor Program (NMDP) is an organization that operates the Be The Match Registry, which holds the samples of donors. About 50,000 transplants have been conducted since 1987. The Be The Match Registry is the central hub of a global network spanning across 41 countries.

The Center for International Blood and Marrow Transplant Research (CIBMTR) is a research collaboration between NMDP and the Medical College of Wisconsin (MCW). The CIBMTR conducts research on the transplant outcomes data to find a faster and easier match for the patients.

Existing Problem

Transplant centers, which use samples from the Be The Match Registry, are required to submit all data related to the transplant to CIBMTR for research purposes. Since each transplant center has its own data collection mechanism, there arises a problem when organizations try to exchange this information.

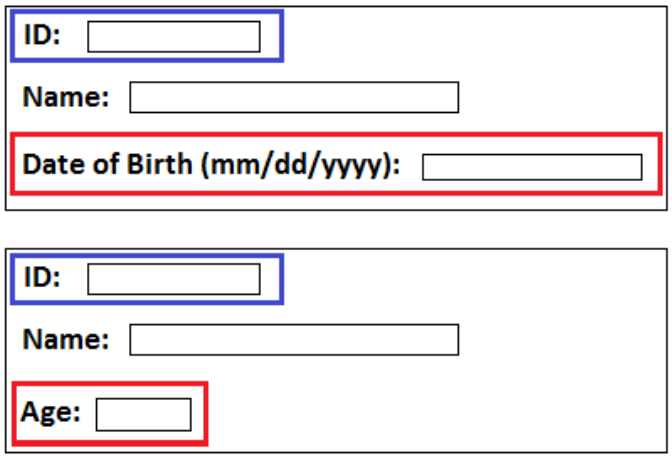

Fig. 1 Problems faced by organizations

The above diagram highlights a couple of problems faced by organizations trying to exchange information related to a bone marrow transplant.

1) **Identifiers:** Assume Person1 has ID:1 in organization A. Now organization A sends the person data to organization B. Now organization B has the data of Person1. But B might already be using ID:1 for another person.

2) **Different attributes:** Organization A might collect date of birth from the person and send it to organization B. But organization B might be collecting age and not date of birth. As a result, organization B needs to restructure the data it receives from Organization A before storing it.

Solution to the Problem

The solution to this problem is the Biomedical Research Integrated Domain Group (BRIDG) model. The BRIDG model is an information model consisting of shared view of concepts between its various stakeholders like Clinical Data Interchange Standards Consortium (CDISC), Health Level Seven International HL7), National Cancer Institute (NCI) and Food and Drug Administration (FDA). The BRIDG model is also intended to provide semantic interoperability and be the future data collection mechanism of the HSCT domain.

Project Goal

The goal of this project is to create a Physical Data Model (PDM) from the BRIDG model using a model driven architecture and distribute it so that it serves as a basis for organizations to develop their data systems and facilitate transplant information exchange. The physical model should also organize the process of collecting, managing, maintaining and analyzing data for CIBMTR.

Implementation

The implementation consists of the following steps:

1) A Common Data Element (CDE) represents a data point that depicts information collected through one of the CIBMTR mandated forms. The first step is to create instance diagrams for each CDE that provides a better visual representation and strengthens the understanding for biostatisticians and analysts who will be conducting research on this data.

CDE	Mapping Path
2866037	PerformedSubstanceAdministration.dateRange WHERE PerformedProcedure.nameCode = Administer Hematopoetic Stem Cell Product {Transplantation}

Fig. 2 Diagram showing a sample CDE

The above diagram shows a sample CDE that is trying to find the date of substance administration in the transplantation procedure.

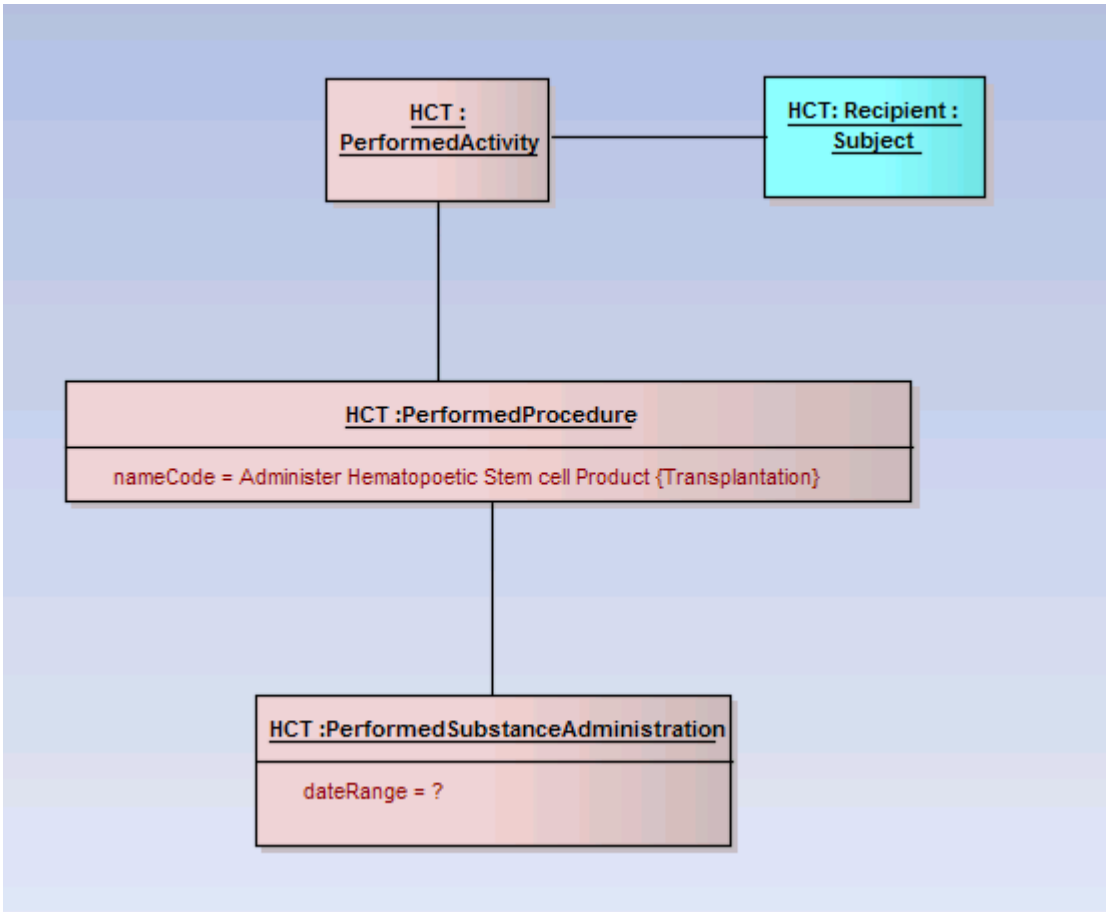

Fig. 3 Instance diagram for CDE

Fig. 3 shows the instance diagram for the above CDE. The instance diagram is created by using classes present in the BRIDG model and it provides a better visual representation when compared to text. It also provides additional information which may not be present in the mapping path. For instance the "Subject" class denotes the patient is recipient , and the "PerformedActivity" represents that an activity is being performed. The "PerformedProcedure" class denotes that the activity being performed is a procedure and the "PerformedSubstanceAdministration" class denotes that the procedure being performed is a substance administration.

2) Once an instance diagram is created for all CDEs (there are approx. 2000 CDEs) all classes used to represent the instance diagrams are extracted to be included in the Conceptual Data Model (CDM).

3) A CDM provides a high level description of a database system but abstracts the underlying details. It consists only of entities and their relationships and is not

sufficient to build a database system. However it allows the Subject Matter Experts (SME) and Business Analysts (BA) to quickly understand the model. The CDM was created using Sybase PowerDesigner tool.

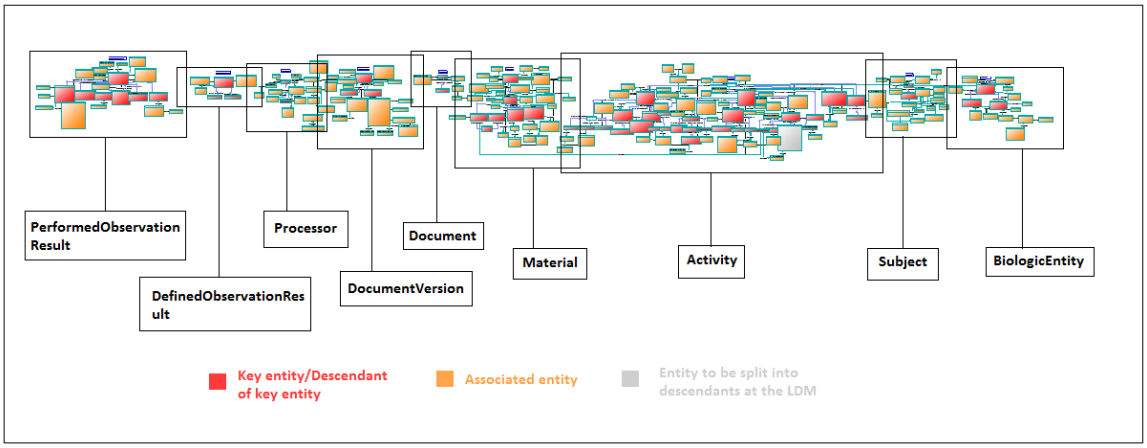

Fig. 4 Conceptual Data Model

4) From the CDM, all entities that need to be carried over to the Logical Data Model (LDM) are identified.

5) A LDM is created using the entities from step 4. The LDM focuses on the business requirements of the organization and provides detailed description about entities, attributes, their relationships, primary keys, foreign keys and normalization. The LDM was also created using PowerDesigner tool.

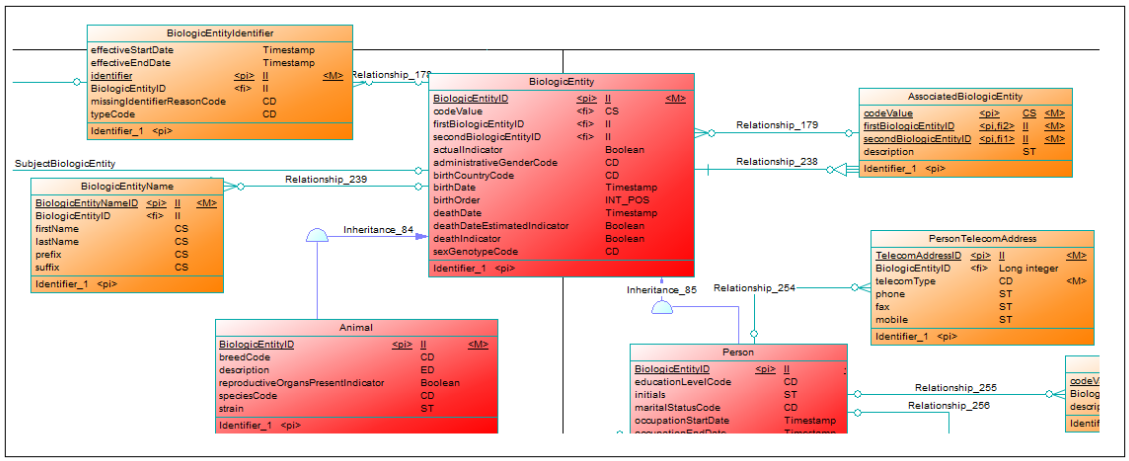

Fig. 5 Sample screenshot of one part of the Logical Data Model

6) From the LDM, the Physical Data Model is generated using the PowerDesigner tool. The PDM consists of tables, columns and relationships and allows for automatic derivation of database schema. The column data types are abbreviated to make it ANSI compliant (30 characters or less).

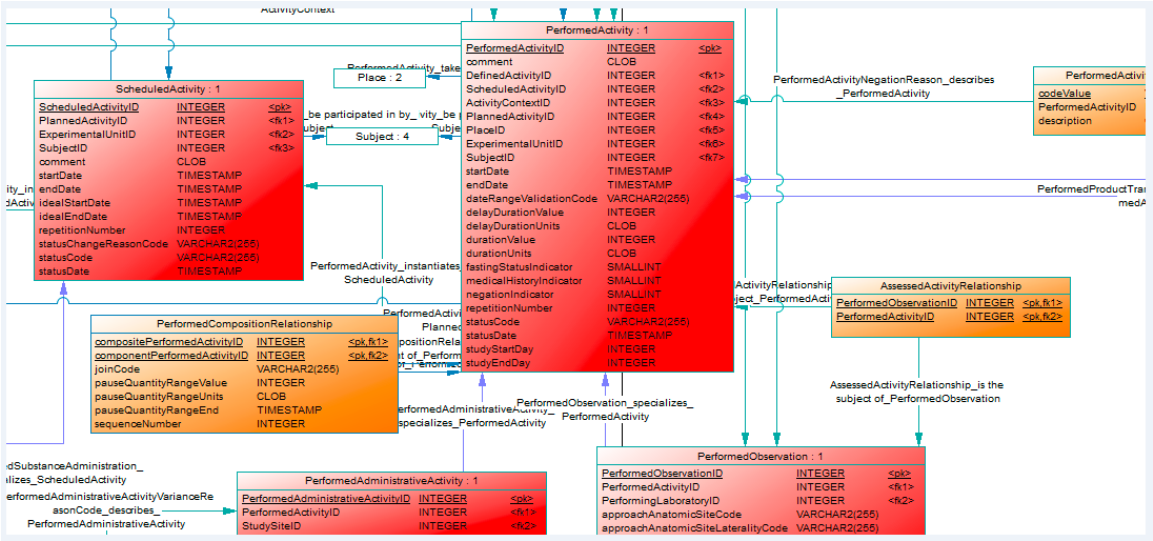

Fig. 6 Sample screenshot of one part of the Physical Data Model

7) Derive the Data Definition Language (DDL) from the Physical Model. Distribute the physical model to associated organizations.

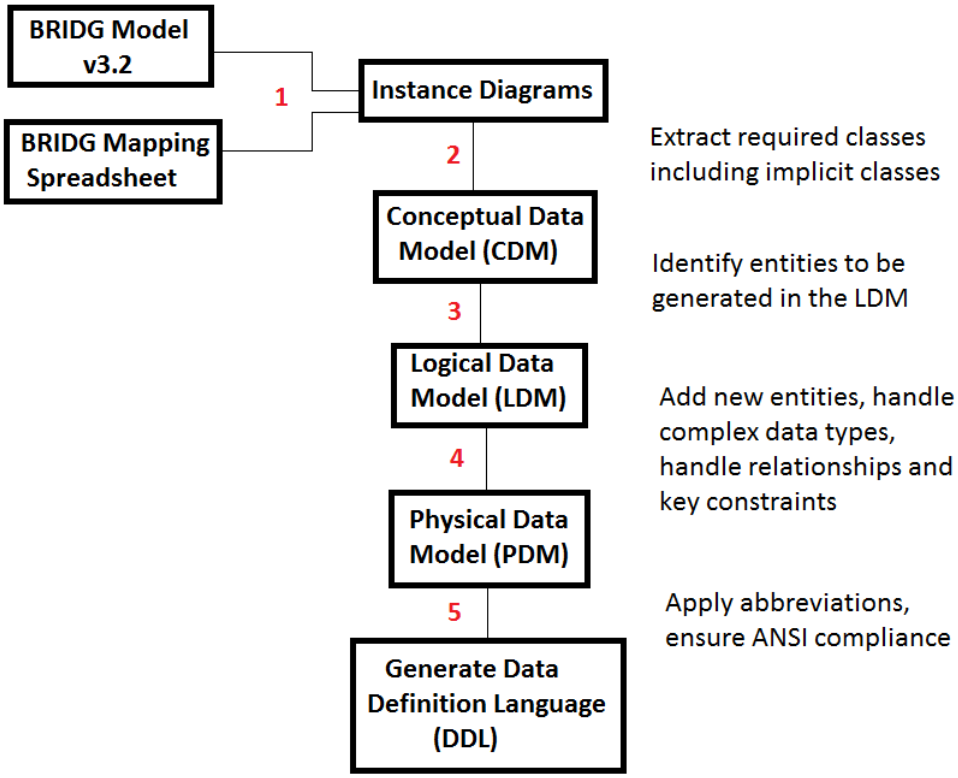

Fig.7 Diagram showing the steps involved in the implementation