DISEASE SPECIFIC DATA MARTS

Workflow and instructional documentation for installing the plug-in and creating disease specific data marts within the SCILHS Network

SCILHS Disease Specific Data Mart Script Generator Plug-in

Table of Contents

DOCUMENT MANAGEMENT

Revision Number	Date	Author	Comment / Modifications
1.01	12/01/14	Janice Donahoe	Document created
1.02	02/18/15	Janice Donahoe	Updated with instructions on how to use the plug-in and run the script.
1.03	03/05/15	Mike Mendis	Added installation information.
1.04	03/11/15	Janice Donahoe	Updated some minor changes.
1.05	03/13/15	Janice Donahoe	Added documentation on setting up a new project and updated db_lookup tables.

INTRODUCTION

The SCILHS Network is a Clinical Data Research Network (CDRN) that is part of a larger network known as PCORnet, the National Patient-Centered Clinical Research Network. One of the objectives of SCILHS is to build a scalable system that enables clinician and patient participation in research. As a result, sites participating in the SCILHS Network will be creating multiple disease specific data marts that are based on the data found in their i2b2 database. In order to assist with this process, the i2b2 team has developed a plug-in that sites can add to their local i2b2 Web Client.

Outlined in the document is the process of requesting a new data mart and how to use the *SCILHS Data Mart Request* plug-in to generate the appropriate database scripts for creating a disease specific data mart.

Assumptions:

The following assumptions are made in the writing of this document:

- The site is part of the SCILHS Network and has both a SHRINE and local instance of the i2b2 running in their environment.

- The *Principal Investigator* has the appropriate access to run a SCILHS network query in the SHRINE client.

- The *Site Administrator* has the appropriate level of access to do the following in their local i2b2 Web Client and Admin Module.

    o View SHRINE queries run by the Principal Investigator.
    o Generate a patient set in the i2b2 by running a local query.
    o Add new projects and users from within the i2b2 Admin.

- The *Site Administrator* has the ability to create database users and schemas in their local environment.

WORKFLOW:

The overall workflow for creating a new disease specific data mart is defined in the steps outlined within this section.

1. In the **SCILHS SHRINE Network**, the **Principal Investigator** ("PI") runs a query to find the appropriate number of cohorts needed for the study.

2. The **PI** flags this query so it can be easily identified in the local instance of the i2b2. The *Site Administrator* ("*Admin*") will use this query to generate the set of patients to be included in the new disease specific data mart.

> **ⓘ Note**
>
> Depending on your study, there may be specific case studies that you do not want included in your new disease specific data mart. In this instance you will need to run a **second query** with the **exclusion criteria** and flag it so it can be used later on in the data mart creation process.

3. The **PI** submits a request for a new project (data mart) to be created by the local *Admin*. The request needs to include the following:

    a. The name of the SHRINE query that will be used by the Admin to generate the list of patients to be **included** in the new data mart.

    b. The name of the SHRINE query that contains criteria for those cases to be **included** in the new data mart.

    c. The name of the SHRINE query that contains the criteria for those cases to be **excluded** in the new data mart. *Optional*

    d. The username of all i2b2 users who can have access to the new project. The list of users should also include the appropriate level of access for each user.

4. The **Admin** creates the new database schemas for the CRC Data, Ontology Metadata and Workplace tables.

> **✅ Tip**

> Do no create the tables for these schemas as they will be created when running the create data mart script. You do need to make note of the new schema names, as they will be needed when creating the new data mart scripts.

5. The **Admin** creates the database users for the new schemas.

> ⚠ **Important**
>
> Make sure the database user(s) for the new schemas have read / write access.

6. In the **local i2b2**, using the SHRINE query flagged by the *PI*, the **Admin** creates the list of patients to be included in the new data mart.

> ℹ **Note**
>
> To obtain the list of patients to be excluded from the new data mart, the Admin will need to run the second SHRINE query that was flagged and identified by PI for exclusion.

7. The **Admin** creates the list of patients to be excluded in the new data mart. This step is only necessary if the PI has flagged and identified a second SHRINE query for exclusion.

8. Using the *SCILHS Disease Data Mart Request* plug-in, the **Admin** generates the database scripts for creating the new disease specific data mart.

    Additional Resource / Instructions: Additional information on how to generate the database script can be found in the section titled *Generating Data Mart Scripts in i2b2*.

9. The **Admin** creates the new disease specific data marts by running the data mart scripts created in the previous step. The data mart script will create the required tables and populate them with the appropriate data from your i2b2 database.

Additional Resource / Instructions: Additional information on how to use the database script can be found in the section titled *How to use the Database Script*.

10. Update the CRC_DB_LOOKUP, ONT_DB_LOOKUP, and WORK_DB_LOOKUP tables with the new database schema names and project information.

11. Setup the new project and it's users in the i2b2 Admin module.

    Additional Resource / Instructions: The i2b2 Installation Guide on the i2b2 Community Wiki will provide additional information on how to setup and manage Projects and Users in the i2b2 Admin module.

12. The **Admin** notifies the PI the new project has been setup and is available for use.

INSTRUCTIONS:

The SCILHS Disease Data Mart Request plug-in is designed to assist administrators with the creation of a new disease specific data mart. The plug-in will use the information entered on the form to generate a database script. This database script will contain a stored procedure that will create a new data mart from an existing one.

The following sections outline instructions on installing the new plug-in as well as the various aspects of creating the new disease specific data marts.

Installing the Plug-in

The process for installing the new SCILHS Data Mart Request plug-in involves two basic steps; (1) Add the plug-in code to your existing local i2b2 Web Client files and (2) Add new plug-in information to the Web Client's loader file.

Step 1: Add plug-in to existing Web Client

The first step in this process is to copy the **/SCILHSDiseaseRequest** folder into the appropriate folder in your existing local i2b2 Web Client.

1. Save the zip file for the SCILHS Disease Request plug-in into a target area of your choice.

2. Extract the files.

3. Copy the **/SCILHSDiseaseRequest** folder to the following location in your existing local i2b2 Web Client:

> *{root http}/webclient/js-i2b2-cells-plugins/standard folder*

Second and Final Step: Add new plug-in information to the Web Client's loader file

In order for the new plug-in to appear in the list of available Analysis Tools in the Web Client you need to update the **i2b2_loader.js** file with the appropriate information.

1. Open the **i2b2_loader.js** file located in */webclient/js-i2b2/*

2. Scroll to the section in the file that contains the following line:

> *i2b2.hive.tempCellsList = [*

3. Add the following lines to this section (see others already listed in the file):

```
{ code: "SCILHSDiseaseRequest",
  forceLoading: true,
  forceConfigMsg: { params: [] },
  forceDir: "cells/plugins/standard"
},
```

4. Save the changes to the file.

5. The *SCILHS Disease Data Mart Request* plug-in will now appear in your list of Analysis Tools.

Identifying Patients in the i2b2

Once the Primary Investigator runs a query from within the SCILHS SHRINE Network it will appear in the local Web Client.

The local Admin will log into the local instance of the i2b2 Web Client and use the SHRINE query to run a new query. This new query will generate the Patient Set to be used when creating the disease specific data mart.

> **ⓘ Note**
>
> Make sure *Patient Set* is selected when you run the local query.

Generating Data Mart Scripts in i2b2

The following instructions explain how to use the *SCILHS Disease Data Mart Request* plug-in to generate the scripts for creating a disease specific data mart in your environment.

1. Verify the following information is entered on the *Specify Data* page.

Script Name:

This is an optional field that can be used to help organize and keep track of the various scripts you may be running. The name entered will be added to the script as a comment and will not be used when the script is run against the database.

Cases:

In simplest terms, cases are a set of existing i2b2 patients that will be **included** or **excluded** when the new data mart is created. In order to populate the new data mart with the appropriate patients and their observations, it is necessary you identify the cases to be included in the new data mart. Identifying the cases to be excluded is optional.

The patient sets to be included or excluded were generated in the current i2b2 database and can be added to the form by dragging the name of the patient set from Previous Queries or Workplace to the appropriate box.

Database System:

Depending on your environment, you will need to select the Database Management System that is appropriate for your i2b2 database. The plug-in uses this information to generate a script that is specific to your type of database. Currently there are two options; SQL Server and Oracle.

Database Names:

The database names are the schema names for your **existing** i2b2 database. The script will use this information as the source for the data that will be inserted into the new data mart. Therefore, it is important that you correctly enter the names for your existing i2b2 database so the appropriate data is selected.

2. Click the **Submit** button to generate the database script.

⚠️ **Important**

The information entered on the *Specify Data* page will be used to define the variables that will be used by the script when it is run in your environment. Therefore it is important you have defined everything correctly before submitting the form.

---

3. The page will automatically change to the *View Results* tab once the SCILHS Data Mart plug-in finishes creating the database script.

4. Create a new file on a local machine in your environment.

5. Copy and paste the database script into the new file.

> ✅ **Tip**
>
> To make it easier for users the plug-in will highlight the entire script in one simple step. All you need to do it use your mouse to click on the text of the generated script. At this point the text for the entire script will be highlighted. To copy the text and paste it in your new file you can use one of the following keyboard commands:
>
> **Windows:**
>
>   Copy = Ctrl+C
>   Paste = Ctrl+V
>
> **Macintosh:**
>
>   Copy = Command+C
>   Paste = Command+V

6. Save the file containing the newly generated script.

How to use the Generated Database Script

The information in this section includes instructions on how to use the database script once it has been generated via the plug-in.

1. Log into the application you use to manage your databases.

2. Connect to your **NEW** database.

3. Open the local file that contains the database script your previously created.

4. Run the script from your **NEW** database.

⚠️ **Important**

The database script creates all the required i2b2 tables, indexes, and stored procedures when it is run against your new database schema. If a table with the same name already exists in the new database schema then the script will stop running.

If the tables do exist in your new database then you will need to delete them before you proceed with running the scripts. Make sure you also delete the indexes and stored procedures if they also exist in your new environment.

If you ran the script and it stopped running because of existing tables, then you will need to manually delete the existing tables, indexes and stored procedures; This includes any tables the script was able to create before it encountered the problem. Once you have deleted the table(s) you will need to run the script again.

---

5. "**Finished creation of datamart**" will display once the script finishes running. At this point your new CRC and Ontology tables have been created and the appropriate data inserted into them.

Setting up New Data Mart in i2b2

Once the new data mart has been created the next step is to set it up in the i2b2 so users can log into it and begin using it for their study.

The following steps need to be taken before users can begin using the new data mart.

1. Update the DB_LOOKUP tables

2. Create a new project in the i2b2 Admin Module

3. Add users to the project with the appropriate level of access

Update the DB Lookup Tables

The DB lookup tables contain the database information needed for the i2b2 cells to connect to the right database and database schema. These tables are part of the hivedata database schema.

The following three DB lookup tables that need to be updated with the data source information for the new data mart.

1. CRC_DB_LOOKUP

2. ONT_DB_LOOKUP

3. WORK_DB_LOOKUP

Each table contains the following six columns:

Column Name	Description / Usage	Example
C_DOMAIN_ID	Enter the ID for the domain in which your i2b2 hive is located. The value entered in this column needs to match the Domain Id defined in your PM_HIVE_DATA table.	i2b2demo
C_PROJECT_PATH	The project path will be used to link this data source to a project in the i2b2. The value entered in this column needs to match the Project Path entered when setting up your project in the i2b2 Admin Module.	/Demo/
C_OWNER_ID	Enter an id for the user who is responsible for the database you are setting up. If there is only one owner for all your databases you can enter an ampersand (@).	@
C_DB_FULLSCHEMA	Enter the database schema name for database you are setting up. Make sure you enter the full name of the schema. This means if it is a SQL Server database then you need to include .dbo The cells use this information to know which database schema in the database it should connect to.	*Oracle:* i2b2demodata *SQL Server:* i2b2demodata.dbo *PostgreSQL:* i2b2demodata
DB_DATASOURCE	Enter the JNDI name that will be used by the cells to connect to the data source	*CRC:* java:/QueryToolDemoDS

	The value in this column needs to match the jndi-name in the datasource configuration file for the associated cell. CRC_DB_LOOKUP matches crc-ds.xml ONT_DB_LOOKUP matches ont-ds.xml WORK_DB_LOOKUP matches work-ds.xml	*ONT:* java:/OntologyDemoDS *WORK:* java:/WorkplaceDemoDS Additional information about the ds.xml files can be found in the Installation Guide located on the i2b2 Community Wiki.
C_DBSERVERTYPE	Enter the type of database.	ORACLE, SQLSERVER, or POSTGRESSQL

STEPS TO UPDATE THE DB_LOOKUP TABLES

1. Log into the database management system you use to maintain your i2b2 database.

2. Connect to your local i2b2 database.

3. Find the table called CRC_DB_LOOKUP

> ℹ **Note**
>
> The specific location of this table will depend upon how you have setup the i2b2 in your environment. The DB_LOOKUP tables are part of the i2b2 hive.

4. Edit / Alter the table

5. Add a new row that contains the appropriate information for the new data mart that you just created.

> ⚠ **Important**
>
> Make note of the C_PROJECT_PATH as you will need this information when setting up the project in the i2b2 Admin Module.

6. Save / commit the changes to the CRC_DB_LOOKUP table.

7. Edit / Alter the table called ONT_DB_LOOKUP table.

8. Add a new row that contains the appropriate information for the new data mart that you just created.

9. Save / commit the changes to the ONT_DB_LOOKUP table.

10. Edit / Alter the table called WORK_DB_LOOKUP table.

11. Add a new row that contains the appropriate information for the new data mart that you just created.

12. Save / commit the changes to the WORK_DB_LOOKUP table.

Once you have finished updating the DB_LOOKUP tables you are ready to setup a new project in the i2b2 Admin Module.

Setting up a new project

The next step in this process is to enter a new project in the i2b2 Admin module. The i2b2 Installation guide contains detailed information about setting up and managing your projects in the i2b2 Admin Module.

The following is quick guide on setting up a new project in relation to your new data mart. For detailed steps on setting up a new project see the Add New Project section of the i2b2 Installation Guide.

1. Log into the i2b2 Admin Module that is associated to your local instance of the i2b2.

2. In the Navigation Tree click on **Manage Projects** to display the Project Summary page.

3. Click on the **Add New Project** button located at the top of the page.

4. The Project Data page will display.

5. Enter the appropriate information for this new project.

> ⚠ **Important**
>
> Make sure the Project Path matches what you entered in the DB_LOOKUP tables (C_PROJECT_PATH column).

6. Click on the **Save** button to save the project.

Add Users to the Project

The last step in set up process is to add the users to project and define the level of access they will have in the project. The i2b2 Installation Guide contains detailed information on how to Add User to a Project and Project User Roles.

> ⚠️ **Important**
>
> In order to run a query and perform other functions in the i2b2 you have to add the **AGG_SERVICE_ACCOUNT** to the new project. This user has to be set up with the *Admin Path* of User and the *Data Paths* of Aggregated and Obfuscated.

1. Click on **Manage Projects** to refresh the list of projects.

2. Expand the Project you just created to display a list of options

3. Click on **Users** to display the *Project Users* page.

4. Click on **Add User to Project** to associate a user to the project.

5. On the *Add Project User* page enter the name of the user you want to add to the project.

6. Click on **Add User to Project** to add the user.

7. In the Navigation Panel click on **Users** to refresh the tree and display the user you just added.

8. In the Navigation Panel, expand the name of the user and click on the **Roles** option.

9. On the *Roles Management* page check of the roles that appropriate for that user in the new project.

10. Click on **Save** to save the changes.

ABOUT THE DATABASE SCRIPT

What the Database Script Does

The following is an outline of what the database script will do when you run it against your new database.

1. Declares and sets the variables to be used by the stored procedure. The values for these variables are taken from the information you entered when submitting the form.

2. Drops and creates the required CRC, Ontology, and Workplace tables. (see list of tables created in the section titled "*Tables Created During This Process*")

3. Populates the QT_QUERY_RESULT_TYPE, QT_QUERY_STATUS_TYPE, and QT_PRIVILEGE tables with data from your existing i2b2 database.

4. Creates the set of patients to use when selecting the data to insert into the patient oriented tables. This set of patients is based on the "*include*" and "*exclude*" patient lists you entered when submitting the form.

5. Populates the Ontology tables with data from your existing i2b2 database.

6. Populates the WORKPLACE_ACCESS table with the required data (Shared folder).

7. Populates the OBSERVATION_FACT, PATIENT_DIMENSION, and VISIT_DIMENSION tables with selected data. All observations, patients, and visit data for the patients listed in your patient set will be inserted into the appropriate tables.

8. Populates the remaining dimension tables (CONCEPT_DIMENSION, MODIFIER_DIMENSION, and PROVIDER_DIMENSION) with data from your existing i2b2 database.

9. Populates the mapping tables with self-mapping data.

10. Creates the required stored procedures for the CRC data mart.

Tables Created During This Process

The following tables are created when you run the database script. If the tables already exist then the script will drop them before creating them.

CRC Dimension Tables

- CONCEPT_DIMENSION

- MODIFIER_DIMENSION

- OBSERVATION_FACT

- PATIENT_DIMENSION

- PROVIDER_DIMENSION

- VISIT_DIMENSION

CRC Mapping Tables

- PATIENT_MAPPING

- ENCOUNTER_MAPPING

CRC QT Tables

- QT_ANALYSIS_PLUGIN

- QT_ANALYSIS_PLUGIN_RESULT_TYPE

- QT_BREAKDOWN_PATH

- QT_PATIENT_ENC_COLLECTION

- QT_PATIENT_SET_COLLECTION

- QT_PDO_QUERY_MASTER

- QT_PRIVILEGE

- QT_QUERY_INSTANCE

- QT_QUERY_MASTER

- QT_QUERY_RESULT_INSTANCE

- QT_QUERY_RESULT_TYPE

- QT_QUERY_STATUS_TYPE

- QT_XML_RESULT

Ontology Tables

- CODE_LOOKUP

- PCORNET_DEMO

- PCORNET_DIAG

- PCORNET_ENC

- PCORNET_PROC

- PCORNET_VITAL

- PCORNET_ENROLL

- SCHEMES

- TABLE_ACCESS

Workplace Tables

- WORKPLACE

- WORKPLACE_ACCESS

GLOSSARY

General Terms

The following table contains terms that are used throughout this document.

Term	Definition
Local i2b2	This is the instance of i2b2 that is used in your environment. It is not the SHRINE or Disease Specific instances of i2b2. In relation to the process of creating disease specific data marts, the local i2b2 is used to find the set of patients to be used when creating the disease specific data marts. The data contained in the local i2b2 database is not specific to any one disease or group of patients.
i2b2 Web Client	The i2b2 Web Client will be used when querying your local i2b2 to find the cases (patients) that are to be included in the new data mart.
Project	
SHRINE Web Client	The SHRINE Web Client is the used when querying other sites in the SCILHS network. Aggregated results will be returned from each site. No patient specific information is available.
SCILHS Network	

Database Terms (Oracle vs. SQL Server)

To avoid confusion the following table contains database terms that are used within this document and may have different meanings depending upon the database management system you use.

Term	Oracle Definition [1]	SQL Server Definition [2]
Database	An Oracle database is a collection of data treated as a unit. The purpose of a database is to store and retrieve related information.	A collection of information, tables, and other objects organized and presented to serve a specific purpose, such as searching, sorting, and recombining data.
Instance	A system global area (SGA) and the Oracle background processes constitute an Oracle database instance. Every time a database is started, a system global area is allocated and Oracle background processes are started. The SGA is deallocated when the instance shuts	A copy of SQL Server that is running on a computer.

	down.	
Schema	Collection of database objects, including logical structures such as tables, views, sequences, stored procedures, synonyms, indexes, clusters, and database links. A schema has the name of the user who controls it.	The names of tables, fields, data types, and primary and foreign keys of a database. Also known as the database structure. * Microsoft refers to this as the "database schema".

> ℹ️ **Note**
>
> The behavior of schemas changed in SQL Server 2005. Schemas are no longer equivalent to database users; each schema is now a distinct namespace that exists independently of the database user who created it. In other words, a schema is simply a container of objects. Any user can own a schema, and its ownership is transferable. [3]

Oracle versus SQL Server Databases

The following list contains some additional things you may want to consider when comparing the database terminology for Oracle and SQL Server.

- Oracle instances cannot have multiple databases defined; each database in oracle is one instance. Where as in SQL Server a single instance can have multiple databases defined. Therefore when comparing Oracle and SQL Server you can consider an "Oracle Database" to be equivalent to a "SQL Server Instance"

- For the most part a schema in oracle is the same as a schema in SQL Server in that they are both a collection of database objects. The main difference is in Oracle, schemas are still equivalent to the database users.

  - When setting up a new database user in Oracle, the associated schema will be created automatically using the name of the new database user.

REFERENCES

[1] Oracle Help Center, (n.d.). Database Concepts, Glossary. Retrieved from
http://docs.oracle.com/cd/B19306_01/server.102/b14220/glossary.htm

[2] MSDN Library, (n.d.). Glossary (SQL Server 2008). Retrieved from
http://msdn.microsoft.com/en-us/library/ms165911(v=sql.100).aspx

[3] Microsoft Technet Library, (n.d.). User-Schema Separation. Retrieved from
http://technet.microsoft.com/en-us/library/ms190387(v=SQL.105).aspx

Appendix A: DB_LOOKUP Table Examples

The tables shown below include examples for both Oracle and SQL Server databases. Both types of databases have two projects. One is the main i2b2 demo project (Path=/Demo/) and the other is the new disease specific data mart (Path=/Demo2/)

CRC_DB_LOOKUP:

C_DOMAIN_ID	C_PROJECT_PATH	C_OWNER_ID	C_DB_FULLSCHEMA	DB_DATASOURCE	C_DBSERVERTYPE
i2b2demo	/Demo/	@	i2b2demodata	java:/QueryToolDemoDS	ORACLE
i2b2demo	/Demo2/	@	i2b2demodata	java:/QueryToolDemo2DS	ORACLE
i2b2demo	/Demo_SQL/	@	i2b2demodata.dbo	java:/QueryToolDemoSQLDS	SQLSERVER
i2b2demo	/Demo2_SQL/	@	i2b2demodata.dbo	java:/QueryToolDemo2SQLDS	SQLSERVER

ONT_DB_LOOKUP:

C_DOMAIN_ID	C_PROJECT_PATH	C_OWNER_ID	C_DB_FULLSCHEMA	DB_DATASOURCE	C_DBSERVERTYPE
i2b2demo	/Demo/	@	i2b2metadata	java:/OntologyDemoDS	ORACLE
i2b2demo	/Demo2/	@	i2b2metadata	java:/OntologyDemo2DS	ORACLE
i2b2demo	/Demo_SQL/	@	i2b2metadata.dbo	java:/OntologyDemoSQLDS	SQLSERVER
i2b2demo	/Demo2_SQL/	@	i2b2metadata.dbo	java:/OntologyDemo2SQLDS	SQLSERVER

WORK_DB_LOOKUP:

C_DOMAIN_ID	C_PROJECT_PATH	C_OWNER_ID	C_DB_FULLSCHEMA	DB_DATASOURCE	C_DBSERVERTYPE
i2b2demo	/Demo/	@	i2b2workdata	java:/WorkplaceDemoDS	ORACLE

i2b2demo	/Demo2/	@	i2b2workdata	java:/WorkplaceDemo2DS	ORACLE
i2b2demo	/Demo_SQL/	@	i2b2workdata.dbo	java:/WorkplaceDemoSQLDS	SQLSERVER
i2b2demo	/Demo2_SQL/	@	i2b2workdata.dbo	java:/WorkplaceDemo2SQLDS	SQLSERVER