

Selecting the Data Source

Stored Workplace data is distributed to projects through the existence of independent databases (in SQL Server) or schemas (in Oracle). These will be referred to in the rest of the document as the "**persistent storage location**" or **PSL**. These PSL's are organized so that the data from two metadata representations can be merged to a "Super" data set. While a person is working on a specific project, they will be directed to data in a PSL associated with that project.

In order to support the i2b2 project distribution strategy, the user is enrolled in numerous projects recorded within the i2b2 project management cell. The projects available to the user are returned in the web service call to the Project Management cell. The logic of selecting the correct PSL for the project is embodied in the following table.

DB_LOOKUP		
PK	C_DOMAIN_ID	VARCHAR(255)
PK	C_PROJECT_PATH	VARCHAR(255)
PK	C_OWNER_ID	VARCHAR(255)
	C_DB_FULLSCHEMA	VARCHAR(255)
	C_DB_DATASOURCE	VARCHAR(255)
	C_DB_SERVERTYPE	VARCHAR(255)
	C_DB_NICENAME	VARCHAR(255)
	C_DB_TOOLTIP	VARCHAR(255)
	C_COMMENT	CLOB
	C_ENTRY_DATE	DATE
	C_CHANGE_DATE	DATE
	C_STATUS_CD	CHAR(1)

The logic for selecting the PSL is as follows:

1. There are two methods to select the correct PSL, an implicit one, and an explicit one. Both rely only on information available within the i2b2 header.
 - a. The implicit one relies upon the data within the <domain> tag, the <username> tag, and the <project_id> tag.
 - b. The explicit one relies upon the data only within the <project_id> tag. It has the format represented as the following string:

"DOMAIN"	"PROJECT"	"USER_ID"
----------	-----------	-----------

- *These may not actually match the domain and username that is actually being used (since it is being built by the client), and must be checked when the PM cell is accessed.*

1. The table is meant to provide a series of default locations if ones are not specifically listed. If a project is listed in the _C_PROJECT_PATH_ column, then that PSL may be used, otherwise a domain source will be used.
2. If a username is listed in the _C_OWNER_ID_ column, and the project also matches the *PROJECT_ID*, the PSL in that row may be used otherwise a project PSL will be used. If the project PSL does not exist, the domain PSL will be used.

For example, only if the *domain | project | user_id* is an EXACT match to the entries in the database will that PSL be used.

1. The project may not have an entry in the table and in that case any project would be designated the PSL of the domain.
2. If a general domain PSL is not available in the table and only a specific project is associated with the domain in the table, then any incoming messages not associated with that project will return an error.
3. In the table, the "@" character is used to represent the absence of an entry (rather than a blank or a null).
4. In the explicit string and in the <project_id> an "@" can be used to optionally represent a blank column.

Other columns are specified as follows:

1. The column `_C_DB_FULLSCHEMA_` is used to contain the path to a table when the data source is used. Software is written so that the absence of the delimiter (usually a ".") does not need to be explicitly stated.
2. The column `_C_DB_DATASOURCE_` is used to contain a short string that represents a data source configured in some other location.
3. The column `_C_DB_SERVERTYPE_` can be "ORACLE" or "SQLSERVER".
4. The column `_C_DB_NICENAME_` is a string that can be used in the client software to describe a data source.
5. The column `_C_DB_TOOLTIP_` contains a longer (hierarchical) representation of the nicename.

To restate, many cells need to access some kind of persistent storage, and these cells will organize their persistent storage so that it is self-contained and can be apportioned in a way consistent with the project-based requirements of i2b2 that are described above. To that end, a table exists in many cells to make the decision of what persistent storage location to which a specific user will be directed, depending on the project and domain to which they are associated.