

## 5.3 Data Mart Schema for Study

Each OpenClinica dataset extracted from an OpenClinica Study creates its own database schema within the Data Mart. The schema name consists of the Study Protocol Name, plus the name of the OpenClinica dataset.

### **IMPORTANT:**

#### *Potential Conflicts with Table and Column Name Maximum Lengths*

PostgreSQL has a maximum length of 63 characters for the names of columns and tables. The Datamart generates table and column names based a composite of the CRF, group, item, and (for multi-selects/checkboxes) the item's response options text. If you have long names for these objects within your CRF, Data Mart may have problems not creating the required tables/columns. For instance, before adding a column to a table, Postgres checks the column name length for validity. If the column name is too long, Postgres will truncate the proposed column name, and attempt to create the column using the truncated name. The problem is that if the truncated column name is found to already exist, instead of skipping adding the column gracefully, the SQL script may error out.

Essentially the combination of the CRF, group, item names and Response Options text can cause a column name to be long enough to be truncated. If two different columns get truncated to the same name the Datamart will not work as expected. We suggest using names that are as short as possible to avoid this conflict.

In OpenClinica 3.1.4 we have made some changes to the way that Data Mart handles very long column names for checkbox/multi select options this change only affects checkbox and multi-select options. To prevent the long column names in the SQL scripts from causing errors, we have added logic when building the column names which does some truncation if the column name would exceed 63 characters. The column names are built by concatenating the Item Name and the Response Options Text. If the Item Name exceeds 30 characters, then it will be truncated and a unique number will be added to it. If the Response Options Text exceeds 30 characters, then it will be truncated and a unique number will be added to it. For any columns that have been truncated, they will be added to a new mapping table. This mapping table has been built to make it easy to reference any items that have been truncated. The mapping table contains the Column name (as truncated), the original Item Name and original Response Options Text.

Additionally in 3.1.4, CRF table names will now be truncated to prevent long names from causing Data Mart errors. Currently table names can be comprised of the following elements concatenated together:

- CRF Name
- Item Group Name
- '\_resp\_opts'

CRF & Item Group Name concatenations exceeding 52 characters will be truncated with a unique number appended to prevent duplicate names. This limit will leave room in the table name for '\_resp\_opts', in case a separate table for multiselect items needs to be created.

From the [PostgreSQL Manual Section 4.1.1](#), Identifiers and Key Words:

"SQL identifiers and key words must begin with a letter (a-z, but also letters with diacritical marks and non-Latin letters) or an underscore (\_). Subsequent characters in an identifier or key word can be letters, underscores, digits (0-9), or dollar signs (\$). Note that dollar signs are not allowed in identifiers according to the letter of the SQL standard, so their use may render applications less portable. The SQL standard will not define a key word that contains digits or starts or ends with an underscore, so identifiers of this form are safe against possible conflict with future extensions of the standard.

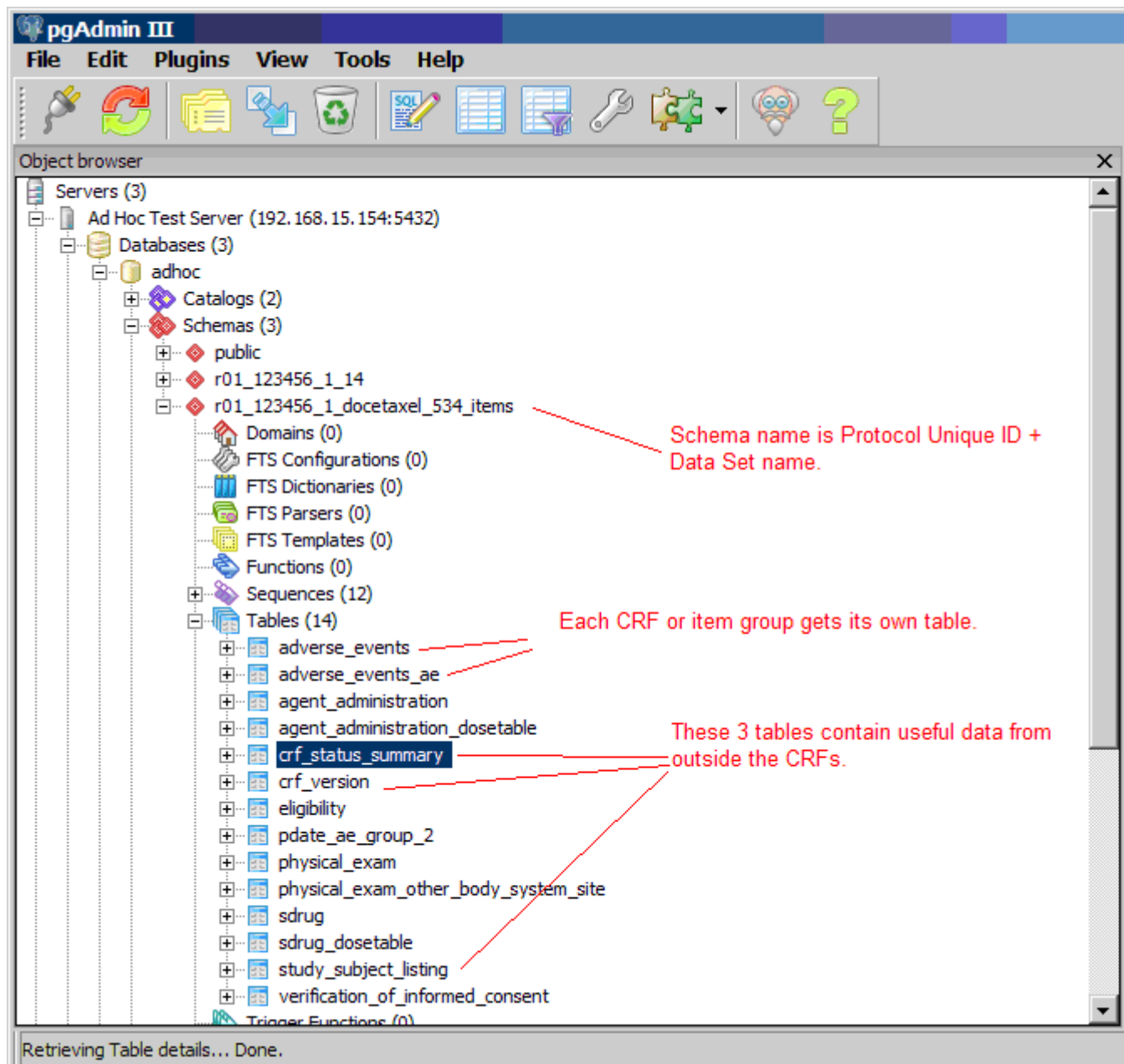
The system uses no more than NAMEDATALEN-1 characters of an identifier; longer names can be written in commands, but they will be truncated. By default, NAMEDATALEN is 64 so the maximum identifier length is 63. If this limit is problematic, it can be raised by changing the NAMEDATALEN constant in src/include/postgres\_ext.h."

### *Reserved Words*

PostgreSQL reserves certain words for creating objects within its database schema, which means that it will restrict you from creating tables named with these reserved words. If you have objects within a CRF (such as item\_name, item\_group) with these reserved words, Data Mart will not create the required tables/columns. Therefore, we advise that you refrain from using these words as your identifiers of an item during the CRF design. If you wish to use these reserved words for your items in the CRF, you may need to add additional characters in order to differentiate it from a reserved word (e.g. name the data item INTEGER\_123 instead of INTEGER). For a complete list of PostgreSQL reserved words, please refer to the following link:

<http://www.postgresql.org/docs/8.4/static/sql-keywords-appendix.html>

*Data Mart Study Schema and CRF Tables (Example from pgAdmin):*



To grant access to specific users for a subset of Study data (for example, only Subjects from a certain Site, or only certain CRFs and data Items) create a special dataset, and create a user Role in the database with access to only that specific schema.

When extracting data in Datamart format, Datamart creates one table for every non-grouped CRF and an additional table for grouped CRF. The maximum non-grouped CRF items count limit is 125. If non-grouped items count in the CRF exceeds the defined maximum limit; single-select labels and multi-select Booleans will be stored in another table.

This page is not approved for publication.