2.5.7 SAS Data and Syntax

OpenClinica version 3.11 introduced the **SAS Data and Syntax** extract format, which were tested using SAS Studio. This extract format functions as follows:

- The output includes three files:
    - **SAS_DATA.xml** - The extracted data.
    - **SAS_Format.sas** - For items defined as single-select or radio button, OpenClinica creates the library and maps response values to the appropriate response text.

        Note: *Because multi-select and checkbox items include multiple values in a string format in OpenClinica (e.g., 1,2,7), these cannot be mapped to individual response text options.*

    - **SAS_MAP.xml** - A mapping file that maps the data to the appropriate structures (e.g., LIBNAME, Table, Column) OpenClinica forces appropriate object names as required by SAS. For example, all Studies start with "S" and all Table names and Column names start with an underscore.
- Once the extract files are downloaded, upload the **SAS_DATA** and **SAS_MAP** files into SAS Studio.
- Open the **SAS_Format.sas** file, copy the text, and paste it into SAS Studio.
- Click the **Run** icon.
    - This generates all the data tables based on Item Groups.
    - OpenClinica Items become SAS Column Names.
    - Tables include the master set of items (i.e., Item Groups span CRF Versions, though the SAS file does not indicate which version of the CRF was the source for the item.)
    - There are two resulting data types: Numeric or Char. All OpenClinica items that are Integer or Real are classified as Numeric. All other OpenClinica data types are classified as Char.
    - The SAS datasets/tables are generated from the OpenClinica metadata. Tables are created for all Item Groups in the extract. If no data was entered for a specific item group, the SAS table is still created, but is empty.

The following apply due to SAS name limitations:

- OpenClinica and DataMart allow 3,999 single-byte characters in a text field. When this size string is extracted to SAS, the full string is in the SAS_DATA.xml file.
- SAS data set names must not exceed 32 characters and must start with either a letter (A-Z) or underscore. As a result, Openclinica uses a modified Item Group OID for the data set name as follows:
    - If group is Ungrouped use the CRF Name, otherwise:
        - To reduce the number of characters the pre-pended IG is removed (This means Group labels start with "_" + 5CHAR (of CRF Name) + _GROUPLABEL)
        - If the resulting value exceeds 35 characters, OpenClinica appends the dataset name with the three- or four-digit number appended to the IG_OID
- SAS column names must not exceed 32 characters and must start with a letter (A-Z) or underscore. As a result, OpenClinica uses a modified Item OID for the column names as follows:

- Truncate from the left to remove the I_5CHAR prefix to each Item Name.
- Use the portion of the OID starting with _ (underscore) followed by ITEMNAME (this ensures no Column Names start with a number.)
- Retain appended three- or four-digit numbers to ensure item/column name uniqueness.