IRAP Digital Commons Frequently Asked Questions (FAQ)

1. What datasets should make it into the IRAP_DB schema?

Only datasets that have been tested, vetted and certified by more than a single pair of eyes should make it into the IRAP_DB schema. All data pending peer review must remain in personal schemas. Users with datasets that need to be migrated to IRAP_DB for sharing must load the logic used in the creation of the data set into the IRAP code library and conduct peer review prior to loading the data into the IRAP_DB schema.

2. How do I load data into the IRAP Digital Commons?

Data can be loaded into the IRAP Digital Commons using the following methodologies:

I. SQL
II. SAS
III. Python
IV. ETL

3. How many loaded tables can I have in my personal schema at the same time?

This depends on the number of records within each table, how many columns there are in the table and what the column data types are. Use discretion when creating tables in your personal schema. Due to limited storage in this environment, if there are tables that are no longer in use or have been replaced with a different dataset, take the extra steps to drop unwanted/un-needed tables. Try not to have multiple tables created that contain pretty much the same dataset.

4. How do we ensure data quality within the IRAP Digital Commons?

It is only by publishing datasets that have been tested and certified (where the logic that created the dataset has been tested/ certified by more than a single individual and placed in IRAP Code Library), that data quality and consistency between the data within the IRAP Commons and data contained in the UCDW production environment can be ensured.

5. How do we ensure data integrity within the IRAP Digital Commons?

If the datasets being created in the IRAP Commons is based on existing UCDW production data and queries on which the datasets are based have been tested and certified, there should be no issue with regards to data integrity. For situations where a dataset is a result of the combination of UCDW production data and external enrichment data, the creator of the dataset is responsible for ensuring that integrity is maintained within the merged data. This can be accomplished via a proper peer-review process.

6. What are some of the operational rules around using the IRAP Digital Commons?

a. Users will be encouraged not to create multiple views of the same data with different table names.
b. Truncate data from tables that is no longer useful. You can simply reload with updated data if table structure is the same.
c. Drop tables that are no longer needed or were created for one-time use.
d. If you have the ability to create tables in your personal schema, try to limit the number of tables that you create by being aware of the limited storage space available.
e. Check the code library before you create a brand new data set. The dataset you are attempting to create could very well be a duplicate of an existing dataset.
f. If any dataset is to be published to the IRAP_DB schema for sharing with IRAP team members, you must ensure that both the query and data has been thoroughly tested and verified as accurate via peer review.
g. Each code submitted to the code library as dataset logic should be clearly described to help prevent the duplication of datasets. This will include the name of the resulting table to be created in the IRAP Commons.
h. All code submitted to the IRAP Code Library as part of the creation of tables within the IRAP Commons is subject to the appropriate and agreed IRAP formatting rules.