



Food and Agriculture
Organization of the
United Nations

How to Add Tabular Data

Hand-in-Hand Geospatial Platform Guide

Table of Contents

Introduction and Key Words	Page 3
System Architecture for Users and Developers	Page 4
Getting Started	Page 5
Step-by-Step Workflow	Page 6
Restrictions and Limitations	Page 6

Introduction

This document provides guidance on the entire tabular data workflow, including how to create datasets, update tables, and debug errors in the workflow process.

Key Words

ORGANIZATION_CODE: the code version of the name of the organization that manages the data. For example: “fao”.

WORKSPACE_CODE: the name of the team/project/department responsible for the data within the organization. It is a category lower than ORGANIZATION_CODE and higher than individual datasets. An example will be “countrystat”

Dimension tables: tables that contains data about different business entities across the dataset, including codes and descriptions. The dimension tables always start with the prefix **dim_**.

Fact tables: tables that contains the data. They can contain some or all code described in the dimension tables. Users can choose to avoid dimension tables if descriptions for each code are provided in the fact table. For example **variable_code** and **variable_name**. Fact tables always start with the prefix **fct_**.

System Architecture for Users

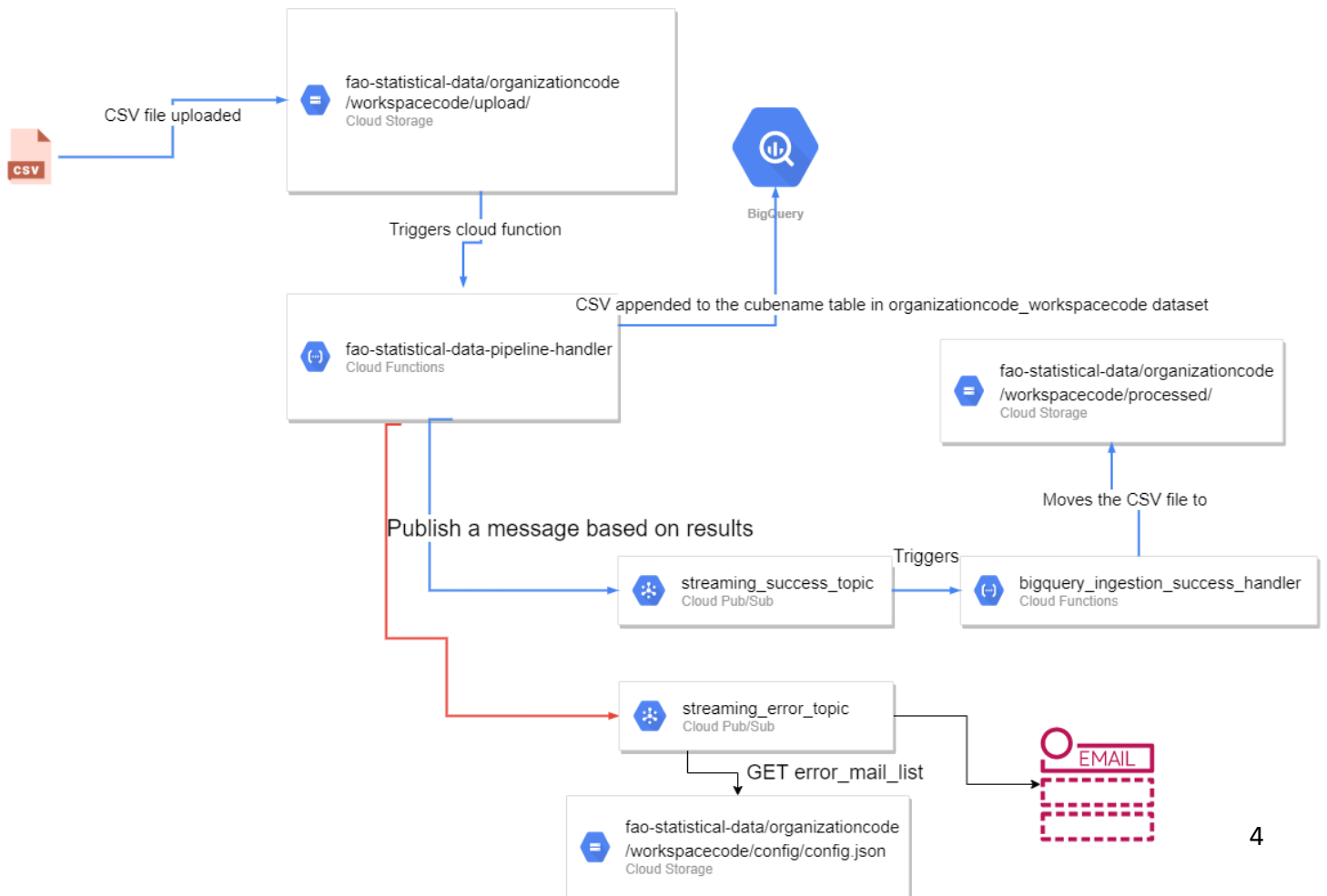
For each combination of ORGANIZATIONCODE and WORKSPACECODE, users will find:

1. One Google storage bucket named “ORGANIZATIONCODE-WORKSPACECODE-tabular” (i.e. fao-countrystat-tabular). There are three folders inside this bucket:
 - a. *Upload folder*: this is where new data will be added
 - b. *Processed folder*: successfully ingested data will be moved here
 - c. *Config folder*: this folder is for configurations regarding the data and error reporting
2. One BigQuery dataset named “ORGANIZATIONCODE_WORKSPACECODE” such as fao-countrystat

It is recommended to refer to the [Restrictions and Limitations section](#).

System Architecture for Developers

System Architecture on Google cloud:



Getting Started

1. Choose an ORGANIZATIONCODE and a WORKSPACECODE. It is encouraged to always refer to the [Restrictions and Limitations section](#).
2. Ensure the bucket is created with the associated three folders mentioned above
 - a. Upload folder
 - b. Processed folder
 - c. Config folder
3. Create a config.json file with the following key-value pair:



```
1 {  
2  
3   "error_mail_list": [  
4     "Guowei.Yin@fao.org",  
5     "Martin.SosaMelgarejo@fao.org",  
6     "mohamed.megahed@fao.org"  
7   ]  
8 }
```

4. Add the email address of the inbox that will receive the ingestion error messages
5. Upload config.json file to the Config folder in your bucket
6. Double check that there is a dataset in fao-maps project in BigQuery named "ORGANIZATIONCODE_WORKSPACECODE (example: fao_countrystat)

Step-by-Step Workflow

1. Upload one or more Fact tables (starts with **fct_**) or one or more Dimension tables (starts with **dim_**) to the *Upload* folder, then the ingestion process will begin.
2. If this is the first time a file with that name is added, a new table will be automatically created in your dataset. If a file with the **exact same name** was already ingested previously, the data will be appended to the table with the **same name**.
3. A new column named “__bigquery_upload_timestamp” will be added to any data that includes the ingestion timestamp.
4. In the case of a successful ingestion, the timestamp will be appended to the file name and the file will be moved to the processed folder.
5. In the case of a failed ingestion, the timestamp will be appended to the file name with an extra prefix (FILENAME_failed_TIMESTAMP.EXTENSION) and the file **will remain** in the upload folder. An email with the complete Callstack leading to the error will be sent the email address added to the config.json file.

HINT: Users can add as many Fact and Dimension tables as you want

Restrictions and Limitations

Data limitations:

1. The names of columns must only contain letters, numbers, and underscores, they must start with a letter or underscore, and must be maximum 128 characters.
2. The Column name “__bigquery_upload_timestamp” be cannot be usedColumns’ names **can’t** be repeated in the same file.
3. The table schema **must** be consistent.
4. Each column should contain only **one** type of data.
5. When you load data, values in TIMESTAMP columns must use a dash (-) separator for the date portion of the timestamp, and the date must be in the following format: YYYY-MM-DD (year-month-day). The hh:mm:ss (hour-minute-second) portion of the timestamp must use a colon (:) separator.
6. When you load CSV or JSON data, values in DATE columns must use the dash (-) separator and the date must be in the following format: YYYY-MM-DD (year-month-day).
7. The only supported separator is a comma [,].

Files limitations:

1. File names cannot contain hyphens (-) or spaces.
2. All tables start with either **fct_** or **dim_** and in the case of uncertainty on which to choose, users can simply choose to use **fct_**.
3. The supported file extensions are (.csv, .log).

- individual files can't be larger than 1.75 gigabytes. If you have large files split them into smaller files.

File name structure: (fct_|dim_)tablename(.csv|.log)

If ingestion is successful: (fct_|dim_)tablename_TIMESTAMP(.csv|.log)

If ingestion is unsuccessful: (fct_|dim_) tablename_failed_TIMESTAMP(.csv|.log)

GCP limitations:

The only limitation is that WORKSPACECODE and ORGANIZATIONCODE cannot contain hyphens and should be lowercase at all times.