## Introduction

The [GC-RDA maDMP Excel Workbook](https://docs.google.com/spreadsheets/d/1OfY5dKEfbvFhlhBjRb4UfdPKqQiB9mjZwe_60R7mu-A/edit?usp=drive_link) contains information such as fieldnames, property ids, descriptions, example values, user-friendly questions, data types, allowed values, cardinalities, requirements, and dependencies. The information is specifically present in the worksheet, GC maDMP Master sheet; which is commonly referred to as the “orange tab.” The information serves as an input to creating the GC-RDA maDMP Application Profile.

The orange tab is used by humans to collaborate and record their decisions.

The **GC-RDA maDMP Excel Workbook** contains all relevant information for the [**GC-RDA maDMP Application Profile**](https://docs.google.com/spreadsheets/d/1j40xvC_GKmGswCT-B-Eh08rxh5Ho8b37VQ_qK7lQkkw/edit?usp=drive_link); however, changes to formatting must be made to create the user-friendly Application Profile visible on the [web page](https://fairerdata.github.io/maDMP-Standard/). The following below documents the process on converting the **GC-RDA maDMP Excel Workbook** into the **GC-RDA maDMP Application Profile**. **The** **GC-RDA maDMP Application Profile** serves as the input for [Go code](https://007gc-my.sharepoint.com/personal/emilywenjia_chu_ec_gc_ca/Documents/Microsoft%20Teams%20Chat%20Files/rda_dmp_common_standard_doc_generator/src) which creates the README.md output in [the root directory of the maDMP-Standard Github repository](https://github.com/FAIRERdata/maDMP-Standard). The Readme.md file is then deployed as a web page on github.io.

## GC-RDA maDMP Application Profile Sheets

The GC-RDA maDMP Application Profile contains 5 different sheets.

* **data\_types:** This sheet defines the types of data expected in the application profile.
* **vocabularies:** describes the various controlled vocabularies used in the application profile (for reference).
* **properties:** Defines the properties or fields used in the application profile.
* **values:** Provides allowable values for specific properties.
* **entity\_descriptions:** Describes entities or sections of the application profile.

Only 3 sheets of the Application Profile, *properties*, *values* and *entity descriptions*, are used to generate the readme.md file and the associated maDMP web page. At the top of the readme.md, the entity descriptions are on the left and the list of properties (label human) is displayed on the right as a list of dropdown items organized using the parent property column.

Below, are tables which display all properties details: data type, cardinality, description, allowed values, example value.

## Guide for Updating the Application Profile:

Only two of the sheets need updating to reflect changes made to the orange tab: the *properties* sheet and the *values* sheet. The conversion code specifically focuses on these two sheets; all other sheets will be unchanged. Note that running the Python code requires a Google Service Account because the service account allows for Python to use Google Sheets API to update the Google Sheets.

1. Download the [maDMP repository](https://github.com/FAIRERdata/maDMP-Standard) and open the RDA-DMP-common standard doc generator/Conversion folder in VSCode (file > open folder), this will be the relevant working folder for the code
2. Ensure that **worksheet\_name** variable in the Python code is set to the correct sheet in the GC maDMP Excel Workbook (worksheet id is found in the url)
3. Set the relevant excel sheet to be plain text, particularly for column “Example Value.” This will prevent Python from automatically interpreting certain cells as Datetimes/Dates. You find Plain text in Format>number>plain text
4. Ensure that credentials.json is present in the main work folder. If credentials.json is not present, read **How to set up a Google Service Account** section below.
5. Install relevant libraries on your Python IDE of choice. On Terminal

**py -m pip install pandas numpy gspread oauth2client**

If it doesn’t work you can try running these 3 prompts:

**pip install --user pandas**

**pip install --user gspread**

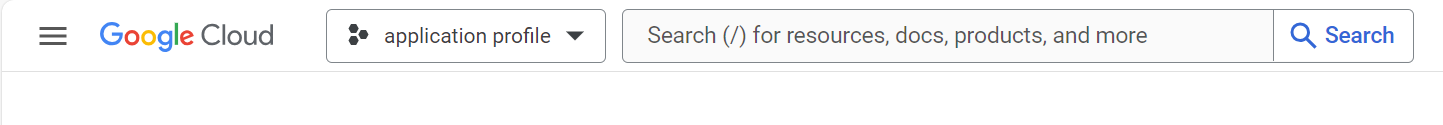
**pip install --user oauth2client**

1. Ensure that the relevant column header names (those in the kept\_columns variable) are the same in the Python script and input file (examples: Property ID, Description, etc.). If they are not, a print message will show up in the terminal and say that the “Column names are not the same.” When this happens, there is a possibility that the outputs will not be correct. Please change the Python code to have the updated names in the variable **kept\_columns**
2. Run the Python code. If ran successfully, the **properties** and **values** sheet in the Application Profile will have been updated

**NOTE:** If there is trouble with the Google Service Account not functioning, manual outputs of the **properties** and **values** sheet can be done with the last two lines of code.

## How to set up a Google Service Account:

1. Sign into your Google account
2. Go to https://console.cloud.google.com/
3. Create a new project, click the red circled icon to find your projects. Then click NEW PROJECT.



A blue line on a white background

Description automatically generated

1. Go to <https://console.cloud.google.com/apis/library> to enable two relevant APIs needed
   1. Google Drive API
   2. Google Sheets API
2. After enabling Google Sheets API, you will need to create credentials

A screenshot of a computer

Description automatically generated

* 1. Answer “Application Data” to create a service account
  2. A screenshot of a computer

     Description automatically generatedEnter the Service Account name and ID

1. Create a key for your Service Account as a JSON and store it safely. Rename the file to credentials.json for convenience. For safety reasons, do not upload the key onto GitHub
2. Go to the Google Sheets you wish to edit using the Service Account and click share. Enter the Service Account email address and set the permissions to “Editor”

## Changes between GC-RDA maDMP Application Profile and RDA Application Profile

The GC-RDA maDMP Application Profile is the extension of the RDA Application Profile. Therefore, there are new fields introduced and other changes. This section details the differences between the Application Profiles. Reference to the GC-RDA maDMP Application Profile will be abbreviated to GC AP and RDA Application Profile will be abbreviated to RDA AP.

1. Vocabularies (in properties and values sheet)
   1. The RDA AP details specific vocabularies such as dct, ali, cerif, foaf, pid\_system\_type, yes\_no\_unknown, etc. Currently, the GC AP only has two vocabularies: gc\_rda\_dmp\_extension and rda\_dmp\_common. The first meaning that it is an addition to the original RDA and the latter meaning that it was a part of the original RDA standard.
2. Properties sheet label\_human and label\_machine
   1. The label\_human and label\_machine is based off of the child of a fieldname. For example dmp/approval/status, the label\_human is Status with corresponding label\_machine status
3. Values Sheet
   1. The “id” column is created through the concatenation of “property” + \_ + “label” values in the GC AP. In the RDA AP, the “id” does not always follow this same rule.

Examples:

* + - “id” involving property contact\_id\_type is contact\_“label” instead of contact\_id\_type\_“label”
      * This occurs similarly for other properties with “id\_type.” Examples include: contributor\_id\_type, dataset\_id\_type, dmp\_id\_type, funder\_id\_type, grant\_id\_type, metadata\_id\_type
    - Other situations where the “id” differs would be in the RDA AP where the parent is omitted from the “id” name.
      * dataset\_personal\_data property has ids as personal\_data\_“label”
      * dataset\_sensitive\_data property has ids as sensitive\_data\_“label”
      * host\_certified\_with property has ids as certification\_“label”, in this situation the “with” has also been trimmed
      * host\_pid\_system property has ids as pid\_system\_“label”
      * host\_supports\_versioning property has ids as supports\_versioning\_“label”

1. Cardinality changes
   1. distribution\_format in RDA AP is 0..n, in GC AP it is changed to 1..n
   2. funding in RDA AP is 0..n, in GC AP it is changed to 1..n
   3. funding\_status in RDA AP is 0..1, in GC AP it is changed to 1
   4. grant\_id\_id in RDA AP is 1, in GC AP it is changed to 0..1
   5. grant\_id\_type in RDA AP is 1, in GC AP it is changed to 0..1
   6. license in RDA AP is 0..n, in GC AP it is changed to 1..n
   7. metadata in RDA AP is 0..n, in GC AP it is changed to 1..n
   8. security\_and\_privacy in RDA AP is 0..n, in GC AP it is changed to 1..n
2. Added or changed descriptions, example values, allowed values, new entity descriptions, vocabularies in vocabularies sheet
3. Added **question, dependency\_type, dependency\_reason** columns in the GC AP

## Changes to the Go code

1. Added user-friendly questions, dependency\_type, dependency\_reason
2. Added id to first table: table; added nav enclosing table1
3. Allowed values are now separated by commas instead of bullet points. In addition to ALLOWED VALUES, also added REQUIREMENT, and DEPENDENCY. Requirement determines if the field is optional, required, or conditional based on its dependency.
4. <ul> and <li> structure is changed from the previous to new structure below

|  |  |
| --- | --- |
| Previous | New |
| <ul>  <li>List item one</li>  <li>List item 2 with subitems:</li>  <ul>  <li>Subitem 1</li>  <li>Subitem 2</li>  </ul>  <li>Last list item</li>  </ul> | <ul>  <li>List item one</li>  <li>List item two with subitems:  <ul>  <li>Subitem 1</li>  <li>Subitem 2</li>  </ul>  </li>  <li>Last list item</li>  </ul> |