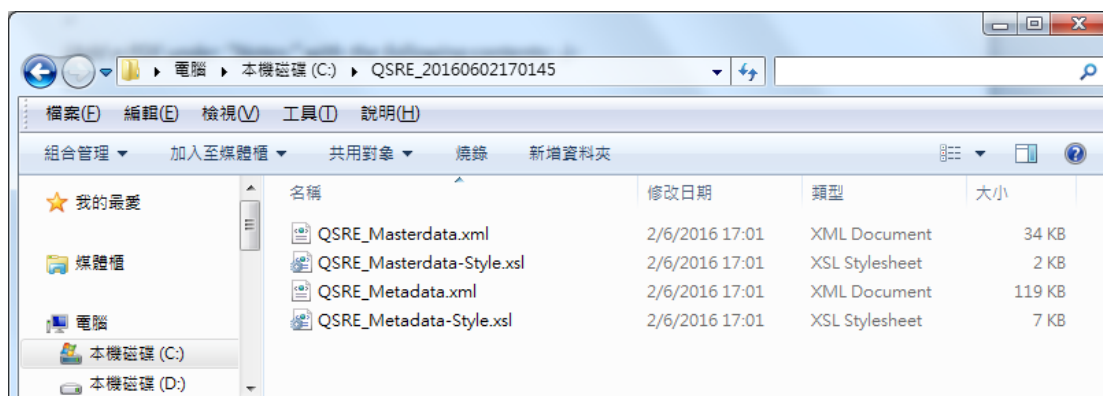


Usage Guide of Statistical Dataset (XML format)

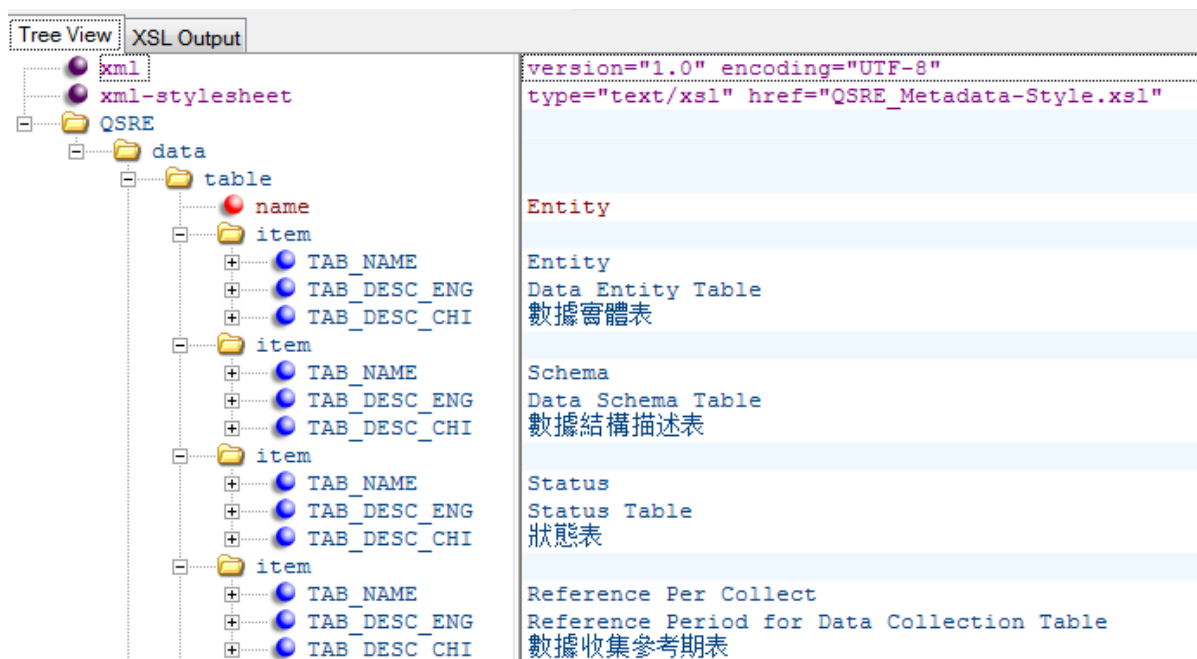
A. Download and Unzip the Statistical Dataset

1. Download the statistical dataset (in ZIP format) from the Census and Statistics Department's website.
2. Unzip the statistical dataset in a local folder which contains:
 - 1 Metadata dataset (XML format)
 - 1 or more Master Data dataset(s) (XML format)
 - 1 Stylesheet File for Metadata dataset (XSL format)
 - 1 Stylesheet File for Master Data dataset (XSL format)



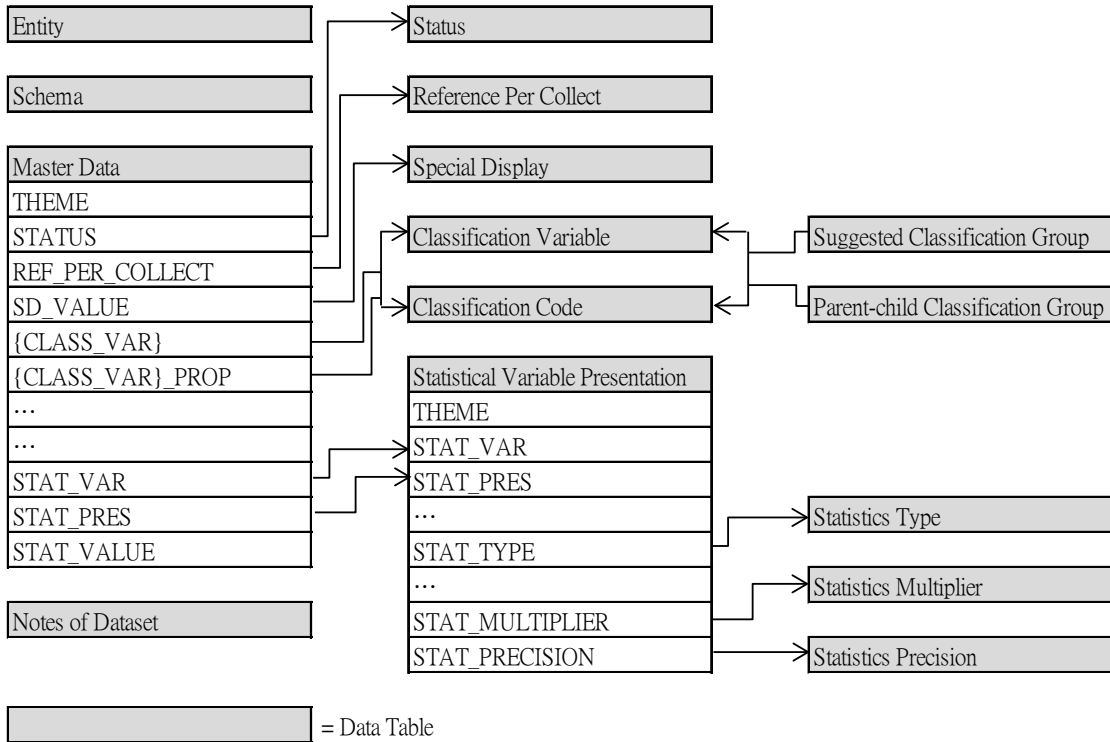
B. Metadata Dataset

3. The Metadata dataset is a data dictionary of the statistical dataset which provides specifications and descriptions of the statistical data in the Master Data dataset.



(Screenshot of opening Metadata dataset in XML editor software, such as Microsoft XML Notepad)

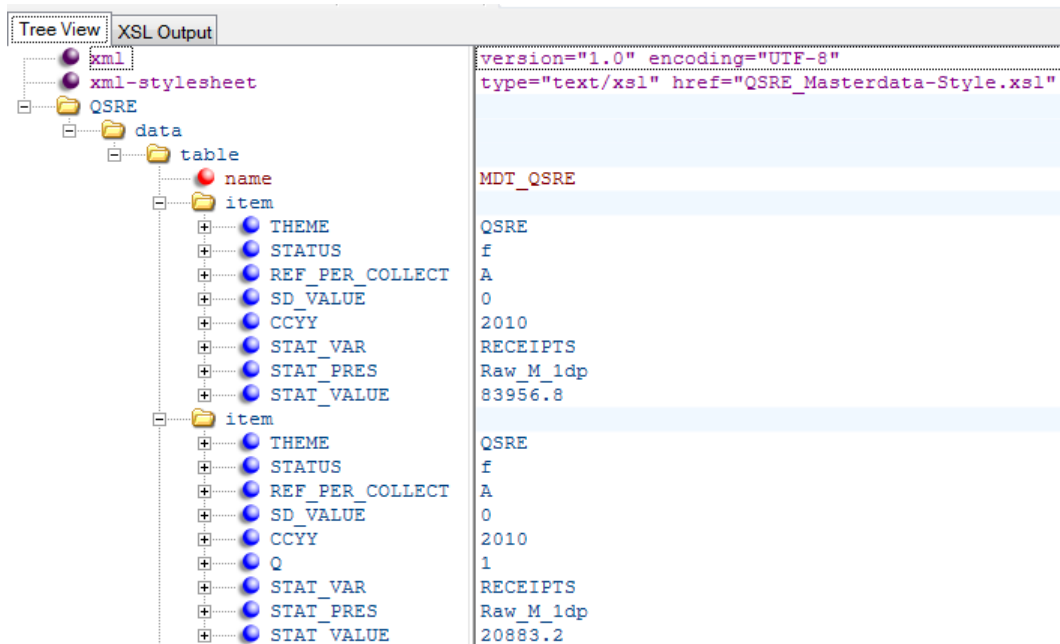
4. A diagram showing the relationship between different data tables is shown below:



5. The Entity table and the Schema table specify the table name, column name and column type of each data table (including the Master Data dataset). Also, the Notes of Dataset table provides points to note (if any) about the statistical dataset.

C. Master Data Dataset

6. The Master Data dataset provides the values of individual statistical data.



(Screenshot of opening Master Data dataset in XML editor software, such as Microsoft XML Notepad)

7. The attributes of each statistical data value (STAT_VALUE) are specified by:
 - Statistical Variable (STAT_VAR), together with a particular Statistical Variable Presentation Code (STAT_PRES) such as raw figure, proportion figure or rate of change which are defined in the Statistical Variable Presentation table and the Statistics Type table;
 - Classification Variable(s) {CLASS_VAR} and their corresponding Classification Code(s) {CLASS_CODE}, including the reference time of statistical data, which are defined in the Classification Variable table and the Classification Code table;
 - Special Display Code (SD_VALUE) indicating the special remark of statistical data (if any), which is defined in the Special Display table; and
 - Other codes like Status Code (STATUS) and Reference Period for Data Collection Code (REF_PER_COLLECT), which are defined in the Status table and the Reference Per Collect table respectively.
8. Also, the order of magnitude and the level of rounding of the statistical data are specified by the Multiplier Code (STAT_MULTIPLIER) and the Precision Code (STAT_PRECISION) in the Statistical Variable Presentation table respectively. Such codes are defined in the Statistics Multiplier table and the Statistics Precision table respectively.
9. With regard to the classification variables and classification codes, the Suggested Classification Group table and the Parent-child Classification Group table provide reference information about the common groupings of classification codes and the parent-child relationship of classification codes respectively.
10. User should take note of the attributes of the statistical data and exercise caution before performing further computation on them.

D. Stylesheet Files

11. To facilitate the reading of the datasets by general users, the stylesheet files in XSL format enable the display of the XML format datasets in a structured way in Excel, XML editor software or web browsers (for XML datasets less than 20MB). Normally, when the XML format datasets are placed together with the corresponding XSL format stylesheet files in the same file folder, the datasets will be displayed in list view automatically when they are opened in web browser or XML editor software.

TAB_NAME	TAB_DESC_ENG	TAB_DESC_CHI
Entity	Data Entity Table	數據實體表
Schema	Data Schema Table	數據結構描述表
Status	Status Table	狀態表
Reference Per Collect	Reference Period for Data Collection Table	數據收集參考期表
Special Display	Special Display Table	特殊顯示表
Statistics Multiplier	Statistics Multiplier Table	統計數字倍數表
Statistics Precision	Statistics Precision Table	統計數字精確度表
Statistics Type	Statistics Type Table	統計數字類型表
Statistical Variable Presentation	Statistical Variable Presentation Table	統計變數展示表
Classification Variable	Classification Variable Table	分類變數表
Classification Code	Classification Code Table	分類編碼表
Suggested Classification Group	Suggested Classification Group Table	建議分類組別表
Parent-child Classification Group	Parent-child Classification Group Table	主次分類編碼組別表
Notes of Dataset	Notes of Dataset Table	數據集註釋表
Master Data	Master Statistics Table	主要統計數字表

TAB_NAME	COL_NAME	COL_DESC_ENG	COL_DESC_CHI	COL_TYPE	PK
Entity	TAB_NAME	Table name	資料表名稱	Varchar(30)	Y
Entity	TAB_DESC_ENG	Table description in English	資料表的英文描述	Varchar(250)	N
Entity	TAB_DESC_CHI	Table description in Chinese	資料表的中文描述	NVarchar(250)	N

(Screenshot of opening Metadata dataset in web browser)

E. Proportion Type Statistical Data in Master Data Dataset

12. To specify the attributes of proportion type (i.e. $STAT_PRES = PROP_{\{xxx\}}$) statistical data like percentage share, it is required to specify the coverage of the denominator of such data additionally. In the Master Data dataset, accompanying with each classification variable $\{CLASS_VAR\}$, a corresponding column named “ $\{CLASS_VAR\}_PROP$ ” is used as a denominator indicator to specify the coverage of the denominator of proportion type statistical data.
13. If a classification code exists in the denominator indicator ($\{CLASS_VAR\}_PROP$), it means that the classification code is the coverage of denominator of the proportion type data; otherwise, if the denominator indicator has a null value, it means that the denominator covers all codes in the classification variable concerned.
14. An example is given in **Annex 1** for illustration purpose.

(2018.12)

Example of Proportion Type Statistical Data in Master Data Dataset

The table below shows hypothetical numbers of employed persons by occupation and sex for illustration purpose:

Number of Employed Persons by Occupation and Sex

Occupation (OCC)	Sex (SEX)		Total
	Male (M)	Female (F)	
Manager	22 (a)	28 (b)	50 (c)
Professional	65 (d)	35 (e)	100 (f)
...
...
Total	400 (g)	600 (h)	1000 (i)

For the proportion (STAT_PRES=Prop_1dp) of employed persons (STAT_VAR= EM(Employed Persons)) analyzed by the classification variables Occupation (OCC) and Sex (SEX) in the Master Data dataset, there are 2 columns (i.e. OCC_PROP and SEX_PROP) serving as the denominator indicator. The following table illustrates the attributes of the value (STAT_VALUE) of the proportions of employed persons.

Proportion Type Statistical Data in Master Data dataset											
THEME	STATUS	...	CCYY	OCC	OCC_PROP	SEX	SEX_PROP	STAT_VAR	STAT_PRES	STAT_VALUE
XXX	f	...	2015	...	1 (Manager)		F (Female)	F (Female)	EM (Employed Persons)	Prop_1 dp	4.7
XXX	f	...	2015	...	1 (Manager)		M(Male)	M(Male)	EM (Employed Persons)	Prop_1 dp	5.5
XXX	f	...	2015	...	2 (Professional)	2 (Professional)	M(Male)		EM (Employed Persons)	Prop_1 dp	65
XXX	f	...	2015	...	2 (Professional)	2 (Professional)	F(Female)		EM (Employed Persons)	Prop_1 dp	35
XXX	f	...	2015	...	1 (Manager)		F(Female)		EM (Employed Persons)	Prop_1 dp	2.8

Classification variables and corresponding denominator indicators

The values (STAT_VALUE) of the proportion were compiled based on the following formula

"4.7" = female manager as a percentage of females in all occupations $(=b)/(h) \times 100\%$
i.e. the denominator is the employed females (SEX_PROP="F") in all occupations (OCC_PROP="")

"5.5" = male manager as a percentage of males in all occupations $(=a)/(g) \times 100\%$
i.e. the denominator is the employed males (SEX_PROP="M") in all occupations (OCC_PROP="")

"65" = male professional as a percentage of professionals of all sexes $(=d)/(f) \times 100\%$
i.e. the denominator is the professionals (OCC_PROP="2") of all sexes (SEX_PROP="")

"35" = female professional as a percentage of professionals of all sexes $(=e)/(f) \times 100\%$
i.e. the denominator is the professionals (OCC_PROP="2") of all sexes (SEX_PROP="")

"2.8" = female manager as a percentage of all employed persons (all occupations and all sexes) $(=b)/(i) \times 100\%$
i.e. the denominator is the employed persons in all occupations (OCC_PROP="") of all sexes (SEX_PROP="")