**Keywords;**

Spreadsheet  
Multidimensional

Dimension

Ingestion

Transformation  
Cube

Spreadsheets form an important means of displaying data and making data public. For the government, this succeeds in getting important information into the public domain, however barriers and challenges exist in accessing and using the data produced.  
Data in spreadsheets is human readable and not machine readable thus constituting a barrier and challenge, as mentioned above, to accessing the information present in the data.

A key step towards making data machine readable is to follow best practices when publishing statistics. This includes producing spreadsheets in such a way that they are structured efectively. In order for data to become machine readable, it needs initially to be converted to CSV (comma separated values) and ultimately be incorporated into a data cube.

A data cube is a multidimensional structure that stores and analyses information. The word cube conjures a vision of 3 dimension, but the word multidimensional means that there can be more than three dimensions, or sides. Each dimension/side represents a different category, that helps classify your data. Think octagonal rubik cube it is can be multisided.

A [dimension](https://gss-cogs.github.io/csvcubed-docs/external/glossary/?h=dimension" \l "dimension) is a category. It enables you to be able slice and dice your data. Examples of dimension includes time, location, product type, customer segment, etc. With this in mind designing the spreadsheet you need to create the spreadsheet in a [standard shape](https://gss-cogs.github.io/csvcubed-docs/external/quick-start/designing-csv/" \l "standard-shape).

Key to this design is

* Di*[mension](https://gss-cogs.github.io/csvcubed-docs/external/guides/configuration/qube-config/columns/dimensions/?h=dimension+column" \l "what-is-a-dimension-column) columns* identify the sub-set of the population that has been observed in a given row.
* Value columns contain the value which has been observed or measured; there is only ever one observed value per row in the [standard shape](https://gss-cogs.github.io/csvcubed-docs/external/guides/shape-data/" \l "standard-shape).
* Measure columns describe what has been observed or measured; note that the measure should not include any information about the units of measure.
* Unit column describes the unit of measure in which the Value has been recorded.

Ingestion

Ingestion is transforming the spreadsheet into csv format and forms the extraction process of the ELT activity. The extraction and convertion of the spreadsheet to csv format requires the [columns headers](https://github.com/GSS-Cogs/application-profile/blob/draft/csv.md" \l "in-brief) of the spreadsheet to be set up appropriately.

Columns should be ordered as follows:

* [Dimension columns](https://gss-cogs.github.io/csvcubed-docs/external/guides/configuration/qube-config/columns/dimensions/?h=dimension+column" \l "what-is-a-dimension-column), order first by time period, then by geography, then in descending order by volume of options in each column (i.e. a column with 17 values would come before a column with 3 values).
* [Observation column](https://gss-cogs.github.io/csvcubed-docs/external/guides/configuration/qube-config/columns/observations/?h=observation+column" \l "what-is-an-observations-column).
* [Measure column(s)](https://gss-cogs.github.io/csvcubed-docs/external/guides/configuration/qube-config/columns/measures/?h=measure+column" \l "what-is-a-measures-column). (e.g. count and spending per capita)
* [Unit column](https://gss-cogs.github.io/csvcubed-docs/external/guides/configuration/qube-config/columns/units/?h=unit+column" \l "what-is-a-units-column). (e.g. number, GBP, miles per hour, portion, percent)
* [Literal attribute](https://gss-cogs.github.io/csvcubed-docs/external/guides/configuration/qube-config/columns/attributes/attribute-literals/?h=literal+attribute+column) columns providing model output (e.g. upper confidence level, sample\_size, standard deviation).
  1. [Observation status](https://gss-cogs.github.io/csvcubed-docs/external/guides/missing-observed-values/?h=observation+status) column (if necessary).
  2. All other [attribute columns](https://gss-cogs.github.io/csvcubed-docs/external/guides/configuration/qube-config/columns/attributes/?h=attribute+columns" \l "what-is-an-attribute-column).