GeoPack Format

- Overview
  - Example
- CSV Geometry & Aliases File(s)
  - Naming Conventions
    - Example
  - File Format
    - Example Level 1
    - Example Level 2
    - Example Level 3
- CSV Demographic File
  - Naming Convention
    - Example
  - File Format
    - Example
- Metadata File
  - Naming Convention
    - Example
  - File Format
    - Example
- Source & Disclaimer

Overview

top

This page will explore the files that are required to create GeoPacks. These include:

1. A CSV file for **each** level of the geographical hierarchy in the pack, containing geometry and aliases.
2. A CSV file containing demographic names and values for the lowest level of the geographical hierarchy. (You only need the lowest level, as Yellowfin will automatically aggregate these values up if higher levels are used instead)
3. A metadata file specifying how the CSVs link together.

Source & Disclaimer information is required for all GeoPacks. It's important to ensure any source data used has been acquired in an appropriate manner and is allowed to be redistributed in this form.

Example

For the purposes of the examples throughout the page, we will explore a scenario around creating a GeoPack for Australia, which has three levels of geographical hierarchy:

1. **State**
2. **Postcode**
3. **Suburb**

This pack will require five files:

1. A CSV file containing geometry and aliases for **States**.
2. A CSV file containing geometry and aliases for **Postcodes**.
3. A CSV file containing geometry and aliases for **Suburbs**.
4. A CSV file containing demographic names and values for **Suburbs**.
5. A metadata file specifying how the CSVs link together.

CSV Geometry & Aliases File(s)

top

Naming Conventions

The CSVs containing geometry and aliases must all share a common name, suffixed with an underscore and the order of the CSV in the pack hierarchy.

Example

In the case of the Australian GeoPack we're creating, the CSV Geometry & Aliases files will be called:

1. **auspack_1.csv** - this will contain the **State** information from the hierarhcy.

2. **auspack_2.csv** - this will contain the **Postcode** information from the hierarhcy.
3. **auspack_3.csv** - this will contain the **Suburb** information from the hierarhcy.

File Format

The CSVs will need to contain data in the following structure:

1. First row containing the label for the level.
2. Second row containing labels for each alias column, labels for the point and polygon columns should not be included.
3. Remaining rows containing:
    a. Columns containing aliases (there can be as many aliases for a particular level as you require. Some levels may have one or two different aliases, but others may have several).
    b. **If the CSV is not the top level of the hierarchy** it must now have a column containing the alias value of the level directly above it that the row belongs to in order to **link** it.
    Note: as each level can have more than one alias, the linking aliases are defined in the metadata file described later.
    c. One Column containing a centroid Point.
    d. One Column containing Polygons (this can be left completely empty if polygons are not required or available for the level, but if it's populated, **all** rows must contain a polygon, no blanks will be accepted).

Example Level 1

For this example, when creating the top **State** CSV, we've decided to use two aliases; **State Name**, and **State Code**. This will mean that the file will look something like this:

```
"State"
"State Name","State Code"
"New South Wales","NSW","<WKT Point>","<WKT Polygon>"
"Victoria","VIC","<WKT Point>","<WKT Polygon>"
...
etc ...
```

Note: the **<WKT Point>** and **<WKT Polygon>** text displayed in the examples above will have to actually contain Point and Polygon data, we've just displayed these as place-holders. Coordinates should be listed in the WKT as latitude longitude, not the other way around.

Example Level 2

The second level of the hierachy, the **Postcode** CSV, might contain a single alias, and we've chosen to link it to the **State** level on the **State Code** field. This will mean that the file will look something like this:

```
"Postcode"
"Postcode"
"2000","NSW","<WKT Point>","<WKT Polygon>"
"2009","NSW","<WKT Point>","<WKT Polygon>"
...
"3000","VIC","<WKT Point>","<WKT Polygon>"
"3008","VIC","<WKT Point>","<WKT Polygon>"
...
etc ...
```

Example Level 3

The third level of the hierarchy, the **Suburb** CSV, might again contain a single alias, and we link to the **Postcode** level on the only alias available, **Postcode** . This will mean that the file will look something like this:

```
"Suburb"
"Suburb Name"
"Sydney CBD","2000","<WKT Point>","<WKT Polygon>"
"Darling Harbour","2000","<WKT Point>","<WKT Polygon>"
"Haymarket","2000","<WKT Point>","<WKT Polygon>"
"Darling Island","2009","<WKT Point>","<WKT Polygon>"
...
"Melbourne CBD","3000","<WKT Point>","<WKT Polygon>"
"Docklands","3008","<WKT Point>","<WKT Polygon>"
...
etc ...
```

CSV Demographic File

top

Naming Convention

This CSV file, containing demographic values for the bottom level of the hierarchy, must use the same common name used by the Hierarchy CSVs, suffixed with an underscore and the word "demo".

Example

In the case of the Australian GeoPack we're creating, the CSV Demographic file will be called:

1. **auspack_demo.csv** - this will contain the **Suburb** demographic values.

File Format

The demographic data included in packs explore various values associated with census demographics available in the geographic area. The CSV will need to contain data in the following structure:

1. First row containing a list of demographic labels.
2. Remaining rows containing values for each demographic, as well as an additional column to specify which value in the bottom level of the hierarchy they belong to.

Example

In this example, we will need to link our demographic information to the bottom of our hierarchy, the **Suburb** level. As our bottom level only has one alias, we'll use that to identify each row, if there were multiple available we would have to select one to use throughout. Either way we will need to specify this link in the metadata file, described later.

```
"Population","Median Income"
14308,75000,"Sydney CBD"
28371,60000,"Melbourne CBD"
...
etc ...
```

Metadata File

top

Naming Convention

This file, containing linking information about the CSVs, must use the same common name used by the CSV files, with no suffix **or** file extension.

Example

In the case of the Australian GeoPack we're creating, the CSV Metadata file will be called:

1. **auspack** - this will contain the information on how each of the CSV files are linked together.

File Format

The metadata included in this file must define how each of the hierarchy and demographic CSVs are linked so that the pack is correctly generated.

1. First rows contain links between the hierarchy CSVs. There will be one row for each child level. The way this is done is a pair of comma separated numbers:
    a. The first number tells you which column of the parent level file the linking field is.
    b. The second number tells you which column of the child level file the linking field is displayed in.
2. Second last row containing links between the bottom level of the hierarchy and the demographic CSV.
3. Last row containing how each of the demographic metrics should be aggregated. The values available are "AVG", "SUM", "MAX", and "MIN". These should be listed in the same order the fields are listed in the demographic CSV.

Note: if your file only contains one hierarchy level, or it doesn't contain demographic information, simply do not include metadata rows to link up this data.

Example

In this example, we have a three level hierarchy, as well as demographic information. This will mean that we need the following in the metadata file:

1. The first row will need to pair the **State** and **Postcode** levels.
   The **Postcode** file uses the **State Code** field as the pairing field. The **State Code** is the **2nd** column in the **State** file and the linking field is the **2nd** column of the **Postcode** file. So the row will be `2,2`.
2. The second row will need to pair the **Postcode** and **Suburb** levels.
   The **Suburb** file uses the **Postcode** field as the pairing field. The **Postcode** is the **1st** column in the **Postcode** file and the linking field is the **2nd** column of the **Suburb** file. So the row will be `1,2`.
3. The third row will need to pair the **Suburb** level with the **Demographic** data.
   The **Demographic** file uses the **Suburb Name** field as the pairing field. The **Suburb Name** is in the **1st** column in the **Suburb** file and the linking field is in the **3rd** column of the **Demographic** file. So the row will be `1,3`.
4. The fourth row will need to nominate how each metric in the **Demographic** file should be aggregated.
   The **Population** value should be summed up, and the **Median Income** should be averaged, so the row will be `SUM,AVG`.

```
2,2
1,2
1,3
SUM,AVG
```

Source & Disclaimer

top
All GeoPacks come with both source and disclaimer information, which is presented to users when they attach a pack to a field. A sample of this information is displayed here:

This GeoPack provides geometry for every country on Earth. It includes a metric for country population estimates as of 2005.
Please note that several of the borders in this pack are disputed.

Source:

- All data is from the World Borders Dataset

World Borders Dataset

Disclaimer:

Although the data in this GeoPack has been produced and processed from sources believed to be reliable, no warranty, expressed or implied, is made regarding accuracy, adequacy, completeness, legality, reliability or usefulness of any information. This disclaimer applies to both isolated and aggregate uses of the information. The information is provided on an "as is" basis. All warranties of any kind, express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, freedom from contamination by computer viruses and non-infringement of proprietary rights are disclaimed. Changes may be periodically made to the information herein; these changes may or may not be incorporated in any new version of the publication.

Data can also quickly become out-of-date. It is recommended that careful attention be paid to the contents of any data associated with a file, and that the originator of the data or information be contacted with any questions regarding appropriate use. If you find any errors or omissions, please report them to the provider.

As you can see, this pack has provided:

1. Information on the contents of the pack and any possible data disputes.
2. Information on the source of the data, including a link to the website itself.
3. Disclaimer information.

top