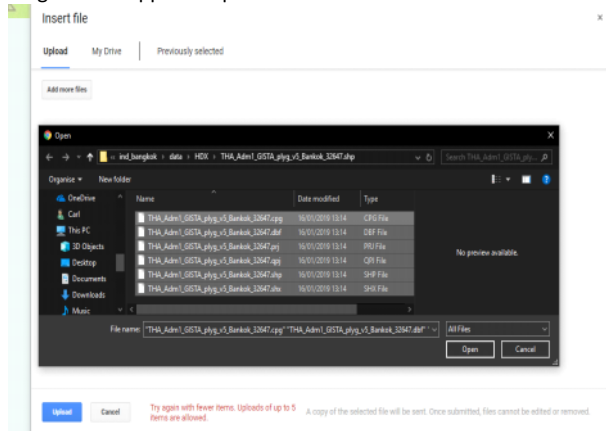


Feedback on spatial data survey for Lancet series

Tuesday, 29 January 2019 13:06

- Boundary file (and other spatial data types)
 - might want zipped shape file?



- I think it would be good to formally separate out the source metadata --- a field for custodian, and separate field for URL. This will help ensure these are stored consistently, and more reliably provided
 - May want to specify format for dates (eg yyyyymmdd) to ensure consistency; or have calendar clicker
 - May be good to formally get people to specify spatial reference of provided file(s), rather than rely on auto-detection
- GTFS
 - Should allow for multiple zips (ie. different agencies may have separate GTFS)
 - Should have a more clear opportunity to describe limitations in this data (e.g. in my brief search for Bangkok I could only find official GTFS for a particular coach company, which actually seems more geared towards inter-city rather than local transport --- but that notwithstanding, there are many other separate transport agencies - e.g. various train lines run by different companies -- operating in Bangkok; should come with comments about data completeness or caveats come under "restrictions"? I think of it more as limitation - but maybe an example could be provided for this item if that's where it should be entered --- or could rephrase as "restrictions, limitations or other important points to note on the use of this data in this project"?). I interpret 'restrictions' more from a licencing / usage perspective. (see below under 'Restrictions')
 - Opportunity to provide description of suggested alternative where GTFS is not available / adequate? Bangkok does not really have any appropriate GTFS; there are various route planning APIs which perhaps could be made use of unofficially. But there is no collated data source of transport schedules / routing that I could find, so I will have to leave the GTFS data page blank
- POIs
 - If users provide supplementary POI data, there should be a field containing a classification to identify what kind of POIs these relate to
 - Perhaps could prompt for the name of the classification field? (e.g. 'category')
 - Should there be options to provide multiple POI data sources, or is it assumed that the researchers have pre-combined their supplementary POI data sources? Or should there be some direction if multiple files are uploaded, e.g. for data sources ("Where multiple files have been provided, please separate responses using a semicolon" (although, it could be hard to post hoc match such lists to the corresponding file reliably)
 - As per above, shp should be zipped, ideally spatial reference is explicitly recorded, and option for additional notes / limitations should be possible
- Restrictions
 - Where a restriction is specified by selecting 'yes', there didn't seem to be scope for describing what is meant by this --- if yes is selected there should be option to provide detail necessary to determine if use is practicable
 - Alternately, as per above suggestion, do away with 'yes/no/maybe' and make this solely qualitative response: "Please detail any restrictions, limitations or other important points to note on the use of this data in this project". This will require some manual interpretation on our end, but at least the important information should be recorded
- Population
 - Rather than request 'Please ensure that it is clear which attribute column should be joined to the ID provided in the template above.', I think preferable would be to get the researcher to specify the join field for each data source uploaded -- ie. "Please specify the field name to be used for matching with geographical boundaries", and "Please

- Dwellings
 - If the dwellings geography is same as previous population geography that was provided, there is no need to re-upload these boundaries. So perhaps make clear "if not already uploaded"
- Journey to work
 - Need to be explicit about units used (ie. require this to be in units of metres, else allow for the researcher to specify alternate units --- former is preferable, I think)
 - The CloudStor csv template appears to have some odd formatting - might be disconcerting for people:

```
,Area_ID_place_of_work,,,,,,,,
Area_ID_usual_residence,1000001,1000002,1000003,1000004,1000005,1000006,1000007,1000008,🔒,🔒
1000001,280,220,954,370,230,819,125,879,🔒,🔒
1000002,389,182,58,666,406,849,613,582,🔒,🔒
1000003,171,226,565,881,691,584,485,420,🔒,🔒
1000004,48,619,561,155,188,758,427,168,🔒,🔒
1000005,433,16,612,591,828,622,884,407,🔒,🔒
1000006,592,650,663,904,52,823,25,347,🔒,🔒
1000007,203,53,996,674,254,785,196,792,🔒,🔒
1000008,425,875,43,115,513,850,165,71,🔒,🔒
🔒,🔒,🔒,🔒,🔒,🔒,🔒,🔒,🔒,🔒,🔒
🔒,🔒,🔒,🔒,🔒,🔒,🔒,🔒,🔒,🔒,🔒
```

↓ Download Global indicators project template - JTW.csv (524 B)

Direct link <https://cloudstor.aarnet.edu.au/plus/s/wStxdR7SfbhFtm4/download>

- I suggest using an alternate template with special characters removed, and slight restructure to heading (I will attach separately). Existing version is on the left; new version (with plain ellipsis, rather than special character) is on the right

The screenshot displays two Excel spreadsheets side-by-side, both titled 'Global indicators project template - JTW.csv - Excel'.

The left spreadsheet shows a table with columns A through N. The data is organized into rows, with the first row (row 1) containing headers: 'Area_ID', 'Area_ID_place_of_work', and then columns D through N. The data rows (rows 2 through 11) contain numerical values. The value in cell G11 (row 11, column G) is highlighted with a green border.

The right spreadsheet shows a similar table with columns A through K. The data is organized into rows, with the first row (row 1) containing headers: 'Area_ID', 'Area_ID_place_of_work', and then columns D through K. The data rows (rows 2 through 11) contain numerical values. The value in cell G11 (row 11, column G) is highlighted with a green border.

- Should just provide the direct download link (<https://cloudstor.aarnet.edu.au/plus/s/wStxdR7SfbhFtm4/download>); then no preview is displayed (for less data literate people, text version of csv could be confusing)

- Residential addresses
 - The example might be a bit obscure for non-data people, ie. I can read this as being a text representation of a csv, but I'm not sure if all will be able to (might be better off linking to a csv template, as done in Journey to work section):

Residential addresses

NOTE: This data is optional. Please provide it only if it is available from an open data source and therefore available for use in this project. An mock example of the format of the data is as follows:

```
ID, prefix, number, suffix, street_name, street_type, suburb, state, lat, lon
123454321, "", "124", "", "La Trobe", "St", "Melbourne", "Victoria", 37.808077, 144.9644363
```

- For the sake of consistency between study regions, as well as simplicity, perhaps it would be preferable to not ask for individual addresses
 - I don't imagine most cities will have this, at least as open data; even for Australia, we don't have 'residential addresses' per se

- Instead we could as a rule take sampling points along the network at regular intervals to ensure consistent coverage
 - ◆ address points can yield quite different sampling in outer compared with inner areas
 - ◆ the high rate of sampling of inner city addresses really is excessive;
 - ◆ arguably constant rate is more fair
 - ◆ Still can weight to dwellings and/or population in areas
 - ◆ No worry about distance of snapping to network; sample points would be pre-located on network segments
 - ◆ Could choose to only sample network segments with particular attributes (e.g. exclude certain road types if thought inappropriate; e.g. excluding footpaths might sidestep the mal-connected representation of footpaths we have identified in OSM)
- So I think the easiest option might be to just not include this section?
- Green space
 - Should it be specified that these should be polygons not centroid points? (to illustrate the kind of file we might receive, I've uploaded a rar of a park centroid shapefile for Bangkok)
 - Do we want to have some comment around what sort of green space this actually represents? (e.g. are they 'parks', only public, include private sports facilities?). If we don't ask, we might not have sense of how different provided green space layers might be.