
Subject: Survey identifiers

Posted by [harry](#) on Tue, 28 Apr 2015 16:27:49 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi,

I'd like to be able to reliably and automatically match between DHS survey data (from any survey) and the corresponding regional boundaries downloaded (where available) from the spatial data repository.

To do this I'm looking for some clarification on survey identifiers that are used in the DHS surveys. As far as I can tell there are (at least) three different systems and I can't quite see how I can reliably match between them.

Downloaded survey data files have names such as BDIR61 and this page shows that this means it's a survey for Bangladesh, Phase 6, first survey done in that country on that phase.

The boundary data downloaded from the new spatial data site here contain a column "SVYID" which is a three digit identifier, as well as country and year; in the case of the above Bangladesh survey it has the value 349. (Unfortunately this isn't in the GPS cluster data when available).

The DHS API when used to retrieve information on available surveys with a request like this returns a field SurveyID which has a different value again; in the case of the above survey it has the value BD2011DHS.

I can see that for a given survey file I could extract the country code and year of interview from HV000 and HV007 and use that to match the boundary polygons. But I don't know if this is always going to match the year recorded in the corresponding boundary file, e.g. for surveys that spanned more than one year? It feels like there ought to be a cleaner way of doing this - a single survey identifier that is common to the survey data and the boundaries (and the GPS data when available) and the DHS API. But I can't find much information on the different survey identifiers. Can anyone explain / clarify these? Is there a single, published mapping between (in this case) BDIR61, 349, and BD2011DHS?

Thanks,
Harry

Subject: Re: Survey identifiers

Posted by [Trevor-DHS](#) on Tue, 28 Apr 2015 20:13:11 GMT

[View Forum Message](#) <> [Reply to Message](#)

You are right that there isn't a published mapping between these various IDs. We will look into updating the API to provide a full mapping of these. I'm attaching a file here that provides the current mapping.

You will find included here the survey ID from the API, as well as the numeric ID, plus the range for the datasets (as version numbers change). For the dataset range, the first two characters are

the country code, the next two are the dataset type (represented in this file as ??), and the next two are the phase and sequence/version number.

Let us know if you find any problems using this file.

File Attachments

1) [SurveyIDs.csv](#), downloaded 990 times

Subject: Re: Survey identifiers

Posted by [harry](#) on Wed, 29 Apr 2015 09:19:48 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hello Trevor,

Many thanks for the quick reply. That CSV is exactly what I need! Thank you.

Yes, if the API could be updated to include this information going forward that would be a great idea.

Harry

Subject: Re: Survey identifiers

Posted by [cippi](#) on Mon, 30 Nov 2020 11:02:09 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hello, I am also looking for a way to link the REG_ID from the Spatial Repository with variables in the Survey data. Has anything changed since this old post? My understanding is that there isn't a unique way to identify provinces in both files, is this correct? What would be the variables in the survey data that would allow constructing the same REG_ID as in the Spatial repository files? Thanks!

EDIT: Just found this file which seems useful:

https://spatialdata.dhsprogram.com/references/subnational_data_schema.pdf

However, I don't understand what CHAR_CAT_ID & CHAR_ID are. Any help?

Thanks
