
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
