Upload of the Data Portrait database to the GIS platform– Documentation

General description the data portrait database

The Data Portrait of Small Family Farms is a project developed by FAO with the objective to set the ground for a standardized definition of smallholders across countries. For this, the project used household surveys developed by national statistical offices in conjunction with the World Bank as part of its Living Standards Measurement Study (LSMS) for 19 countries. As a result, the smallholder farmers' dataportrait is a comprehensive, systematic and standardized data set on the profile of smallholder farmers across the world.

As mentioned, the Data portrait is made up of surveys from 19 countries, and for some of them, data is reported for more than one round, resulting in a total of 29 surveys, with a total of 233,523 farmers providing information related to agriculture. As all surveys are under the LSMS framework, they share similar characteristics. A two-stage stratified random sampling technique was followed, and all surveys (with exception of the Ethiopian Rural Household Survey), are nationally representative and cover urban and rural areas.

Objective

The process had as objective, to feed the Hand in Hand Geospatial (GIS) platform with farmers' household information from the Data Portrait survey, disaggregated at the lowest administrative level divisions as available. This process of cleaning, organizing, disaggregating, and uploading the data portrait survey into the GIS platform was carried out in the second semester of 2020.

To upload the household's surveys in the GIS platform, 73 variables were selected from the Data Portrait to be displayed in the platform. The 73 variables are divided in 11 categories: access to markets; capital; consumption; general; income; infrastructure; inputs; labor; production; social and technology adoption. This information is also disaggregated between total farmers and smallholders for each country.

This note documents the process that was carried out to find the different levels of administrative divisions available in each database that makes up the Data Portrait, and the results of matching the available divisions with the GIS platform.

Methodology

Each survey that makes the Data Portrait, represents a country and a specific wave. For each survey a process of searching for the available administrative divisions was carried out. The goal was to find names and codes available in the surveys of the different divisions until the lowest administrative level possible. In some cases, codes and names were found only for the first administrative level, in some cases until the second and some others until the third. Also, in several cases names or codes of some divisions were not available.

When names were not available but codes where, those codes were matched with information available from official registries and names were included. The level of accuracy in this process varies from country to country. Once each household had assigned codes and names for all administrative level divisions available, aggregated means of each variable were calculated at the different levels in order to obtain average information per each administrative division. This process was carried out for all producers and smallholders separately.

The available administrative information per country was matched with the information available in the GIS platform. The variable used for matching was the name of each administrative division and for every match, it was assigned the code of the division available in the GIS platform. For this process, two different databases of administrative divisions from the GIS platform where used, the GADM and the GAUL databases. The matching was done with the database where the level of matching was higher.

After the matching, once each division had assigned a platform code, the averages per region were uploaded in the GIS platform.

In cases were not all the available divisions in the survey were matching with the divisions from the GADM or GAUL databases, an average of the non-matched divisions was calculated. This average was then assigned to all the remaining regions from the platform (GAUL or GADM databases) that the surveys did not have data about, to complete blanks in the maps. On the other hand, when all the divisions from the surveys matched the ones in the platform, but remaining divisions in the platform were left without values, as surveys usually do not cover all regions inside a country, national averages for each variable were assigned.

Albania

Survey**:** Living Standards Measurement Survey, 2005

Information available: The LSMS survey covered 36 districts (second administrative level) and 211 municipalities (third administrative level) [1]. Names and codes per each administrative division were available in the survey.

Matching with the GIS platform: Matching of the first three administrative level divisions of Albania with the GIS platform was carried out with the GADM database. Second and third level that were available in the survey were matched and based on this matching, first level divisions from the GADM database were assigned (12 divisions). The matching was 100% for level 2, and 99.5% for level 3, with 1 municipality not matching. Information from the survey covered 100% of level 1, 97% of level 2 and 56% of level 3 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Bangladesh

Survey: Household Income and Expenditure Survey, 2005

Information available: The HIES 2005 survey covered 6 divisions (first administrative level), 64 districts (second administrative level) and 357 Upazilas or Thanas (third administrative level). Codes were available for all three levels; however, names were only available for the first two. To guarantee a matching with the GIS platform, codes for level 3 were matched online to obtain names. The matching was made with the information available at the Geo Code list of March 2013, from the National Statistical Office (NSO) available in the web portal of the Department of Disaster Management of Bangladesh[2]. The Thanas codes in the database matched with 96.6% of the codes found online. Therefore, 345 Thanas from the 357 were assigned a name.

---

[1] In 2014, Albania changed its administrative divisions.
[2]Link:http://ddm.portal.gov.bd/sites/default/files/files/ddm.portal.gov.bd/notification_circular/f8d45c66_c 3ee_4cf9_9584_e784aed88924/geocodeweb.pdf

Matching with the GIS platform: Matching of the first three administrative divisions of Bangladesh with the GIS platform was carried out using the GADM database. Matching for level 1 was 100%, while for level 2 and 3 was 98.4% and 99% respectively. This as 1 out of 64 districts did not match, two Thanas did not have name and one Thana did not match. Information from the survey covered 100% of level 1, 98% of level 2 and 63% of level 3 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Bolivia

Survey: Household Survey, 2005

Information available: the HS 2005 recollected information from the sample households until administrative level 3. However, the survey files only had codes and names available at the department (first) level. Therefore, the survey only provided names and codes for the 9 departments of the country.

Matching with the GIS platform: Matching of the first administrative division of Bolivia with the GIS platform was carried out using the GADM database. All the 9 departments matched and information for them is available in the GIS platform.

Cambodia

Survey: Household Socio-Economic Survey (CSES) 2003/04

Information available: The CSES covered 24 provinces (first administrative level), 165 districts (second administrative level) and 693 communes (third administrative level). In the databases information of codes were available for all three levels of divisions, however names were only available for the first level. To guarantee a matching with the GIS platform, names of the two remaining levels were matched with available official information. The information was taken from The Humanitarian Data Exchange Platform developed by OCHA. Where OCHA ROAP publishes Cambodia's sub national administrative boundaries obtained from the Department of Geography of the Ministry of Land Management, Urbanization and Construction in 2008 and unofficially updated in 2014 according to sub-decrees on administrative modifications[3].

The matching with the official information for the second administrative level was of 95%, where names were not matched for 7 of the 165 districts. Level 3 had a matching of 88%, where the codes of 83 from the 692 communes did not match with the official information. The differences in codes can be explained as the survey was carried out between 2003 and 2005, and since then several administrative regions names and codes have suffered changes, and the official information was updated in 2014.

Matching with the GIS platform: Matching of the first three administrative divisions of Cambodia with the GIS platform was carried out using the GADM database. Matching for level 1 was 96%, as one province did not match. From the 158 available districts, 95% matched with the GIS platform as 150 names matched. Finally, 91% of level 3 divisions matched with the GIS information, leaving 53 communes without a GIS code. Information from the survey covered 96% of level 1, 84% of level 2 and 39% of level 3 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

---

[3] Link: https://data.humdata.org/dataset/cambodia-admin-level-0-international-boundaries

Ethiopia

Survey**:** Ethiopian Rural Socioeconomic Survey - ERSS 2011/12

Available information**:** the ERSS covered 10 regions (first administrative level), 69 zones (second administrative level), and 263 districts (third administrative level). The databases provided codes and names for all the first three levels of divisions. However, nine out of 69 zone names and 74 out of 263 districts names were missing. Since the matching with the GIS platform is made by name, the missing names for zones and districts were found in official data and matched by code. The information used was taken from the Information Technology Outreach Services (ITOS) with funding from USAID from 2019 available in The Humanitarian Data Exchange Platform developed by OCHA[4]. With this information, name was assigned to 6 zones and 65 districts. The imperfect match can be attributed to the date of availability of information. The survey contains information from 2011-12 while the information from OCHA is from 2019.

Matching with the GIS platform**:** Matching of the first three administrative divisions of Ethiopia with the GIS platform was made using the GADM database. Matching for level 1 was 100%, for level 2 was 97% with 2 zones not marching, and level 3 was 97% with 9 districts not matching. Information from the survey covered 91% of level 1, 80% of level 2 and 37% of level 3 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Ghana

Surveys**:** Ghana Living Standards Survey - LSMS, 2005/06 (round 5) and 2012/13 (round 6)

Available information**:** the LSMS survey on round 5 covered 10 regions (first administrative levels) and 110 districts (second administrative level). The round 6 covered 10 regions and 170 districts. Codes for all administrative levels, and region names were available in both waves. However, district names were not available for round 5. Therefore, a matching between rounds 5 and 6 was carried out using district codes, and names were assigned for round 5.

Matching with the GIS platform**:** Matching of the first two administrative divisions of both surveys in Ghana with the GIS platform was carried out using the GADM database. For round 5, matching for level 1 was 100%, while for level 2 was 66% with 37 districts not marching. For wave 6, matching for level 1 was 100%, while matching for level 2 was 82% with 29 districts not marching. Information from the 2005/06 survey covered 100% of level 1 and 58% of level 2 divisions available in the platform (GADM database). While information from the 2012/13 survey covered 100% of level 1 and level 2 divisions from the platform. The remaining divisions in the platform were assigned values as described in the methodology.

Guatemala

Survey: National Living Standards Measurement Survey - LSMS, 2006

Available information**:** The LSMS survey covered 22 departments (first administrative level) and 304 municipalities (second administrative level). All names of departments where available in the survey, however, only 293 names of municipalities were available.

---

[4] Link: https://data.humdata.org/dataset/ethiopia-cod-ab

Matching with the GIS platform: Matching of the first two administrative divisions of Guatemala with the GIS platform was carried out using the GADM database. For the first administrative level 100% of names matched, while for the second administrative level the matching was of 70%, with 204 names matching. Information from the survey covered 100% of level 1 and 82% of level 2 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Indonesia

Survey: Indonesia Family Life Survey 2000

Available information: The survey covered 19 provinces (first administrative level), 37 regencies (second administrative level) and 1,141 districts (third administrative level). All names of divisions were where available in the survey.

Matching with the GIS platform: Matching of the first three administrative divisions of Indonesia with the GIS platform was carried out using the GADM database. For the first administrative level 100% of names matched. For the second and third administrative level the matching was of 99%, with 2 regencies and 5 districts not matching. Information from the survey covered 58% of level 1, 7% of level 2, and 17% of level 3 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Kenya

Survey: Integrated Household Budget Survey – KIHBS, 2004/05

Available information: The KIHBS survey covered 8 provinces (first administrative level) and 69 districts (second level). The survey had available codes and names for both level divisions[5].

Matching with the GIS platform: Matching of the first two administrative divisions of Kenya with the GIS platform was carried out using the GAUL database. The matching for the first level was of 100%, while for the second level was 99%, with one district not matching. Information from the survey covered 100% of level 1 and 91% of level 2 divisions available in the platform (GAUL database). The remaining divisions in the platform were assigned values as described in the methodology.

Malawi

Survey: Integrated Household Survey, 2004/04 (round 2) and 2010/11 (round 3)

Available information: The IHS2 covered 26 districts in Malawi (first administrative level), and 221 Traditional authorities (second administrative level). The IHS3 covered 31 districts and 281 Traditional Authorities. Codes and names for all regions were available in the databases.

Matching with the GIS platform: Matching of the first two administrative divisions of Malawi with the GIS platform was carried out using the GADM database. For the IHS2, the matching of the first levels divisions was of 100%, and the matching of the second level was 77%, with 52 Traditional authorities not matching. For the IHS3, the matching of the first admin levels divisions was of 97%, with one district not matching, and the matching for the second level was 75%, with 70 Traditional authorities not matching. Information from IHS2 survey covered 93% of level 1 and 64% of level 2 divisions available in the platform (GADM database). While information the IHS3 covered 100% of

[5] Since 2010 Kenya implemented a new constitution and the administrative divisions were changed.

level 1 and 82% of level 2 divisions available in the platform. The remaining divisions in the platform were assigned values as described in the methodology.

Nepal

Survey: Nepal Living Standards Survey II, 2002/03

Available information**:** The NLSS II covered 72 districts (third administrative level). The survey contained names and not official codes for the districts. In order to get information of the upper level divisions, names of districts were matched with official information available at The Humanitarian Data Exchange Platform developed by OCHA, last updated on 2017[6]. Two districts did not match; therefore, no provinces were assigned to them. After this process, 7 provinces and 72 districts were available in the survey.

Matching with the GIS platform: Matching of the first three administrative divisions of Nepal with the GIS platform was carried out using the GADM database. District level information was matched with administrative level 3 divisions from the GADM database, while provinces were matched with level 2. First level divisions available in the GADM database where also assigned to the available regions. The matching with the GADM database for all the administrative divisions was of 100%. Information from the survey covered 100% of level 1 and level 2, and 96% of level 3 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Nicaragua

Survey**:** National Living Standards Household Survey (EMNV), 1998 (round 2) and 2005 (round 4)

Available information**:** The EMNV 1998 survey covered 17 departments (first administrative level) and 125 municipalities (second administrative level), while the 2005 covered the same 17 departments and 136 municipalities. Both surveys contained codes for departments and municipalities, and names for departments. However only the 2005 survey contained municipality names. Therefore, municipalities names from 2005 were matched to the 1998 surveys using each survey codes.

Matching with the GIS platform**:** Matching for the first two administrative divisions of both surveys of Nicaragua with the GIS platform was carried out using the GADM database. For the EMNV 2005 matching at the first level was of 100%, while for second level it was 90% where 13 divisions did not match. For the EMNV 1998 matching at the first level was of 100%, while for the second level was of 94% where 7 municipalities did not match. Information from EMNV 2005 survey covered 94% of level 1 and 88% of level 2 divisions available in the platform (GADM database). While the EMNV 1998 covered 94% of level 1 and 84% of level 2 divisions available in the platform. The remaining divisions in the platform were assigned values as described in the methodology.

Niger

Survey**:** National Survey on Household Living Conditions and Agriculture (ECVM/A), 2010/11

Available information: The ECVM/A survey covered 8 regions (first administrative level), 37 departments (second administrative level), and 104 communes (third administrative level). On the

---

[6] https://data.humdata.org/dataset/administrative-bounadries-of-nepal

databases codes were available for all levels, and names were available in the survey documentation.

Matching with the GIS platform: Matching of the first three administrative divisions of Niger with the GIS platform was carried out using the GADM database. Matching at the first and second level was of 100%, while at the third level it was of 75% as 26 communes did not match. Information from the survey covered 100% of level 1 and level 2, and 60% of level 3 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Nigeria

Survey: General Household Survey, 2010/11 and 2012/13.

Available information: The surveys in 2010/11 and 2012/13 covered 37 states (first administrative level). The survey in 2010-11 covered 416 local government areas (second administrative level), while the 2012/13 covered 436. In both surveys, names and codes of each administrative level were available in the database.

 Matching with the GIS platform: Matching of the first two administrative divisions in both surveys of Nigeria with the GIS platform was carried out using the GADM database. In both surveys, matching for the first administrative level was of 100%. For the second administrative level, matching for both surveys was of 99% where 2 local government areas did not match. Information from the 2010/11 survey covered 100% of level 1 and 54% of level 2 divisions available in the platform (GADM database). While the 2012/13 survey covered 100% of level 1 and 56% of level 2 divisions available. The remaining divisions in the platform were assigned values as described in the methodology.

Panama

Survey: Encuesta Niveles de Vida, 2003

Available information: The survey covered 12 provinces (first administrative level), 75 districts (second administrative level) and 324 townships (third administrative level). The database only contained codes for all the divisions. Therefore, to obtain names of the divisions, codes were matched with the official numeric codification of the administrative divisions of the Republic of Panama[7].

Matching with the GIS platform: Matching of the first three administrative divisions of Panama with the GIS platform was carried out using the GADM database. For the first and second level, matching was of 100%, while for the third level matching was of 99% as 2 townships did not match. Information from the survey covered 92% of level 1, 95% of level 2, and 54% of level 3 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Tajikistan

Survey: Tajikistan Living Standards Survey, 2003 and 2007.

---

[7] Source: Direction of Statistics and Census of the General Comptroller of the Republic of Panama: https://www.inec.gob.pa/archivos/P6221DIV-DISTRITOS2013.pdf

Available information: The 2007 survey covered 5 country regions and 64 districts. The 2003 survey covered 5 regions and did not contain available information of the districts it covered. Names and codes were available in the surveys.

Matching with the GIS platform**:** Matching of the first two administrative divisions of Tajikistan with the GIS platform was carried out using the GADM database. In both surveys, matching for the first level was of 100%, while matching for the second level available in the 2007 survey was of 70%, where 19 districts did not match. Information from the 2003 and 2007 survey covered 100% of level 1 divisions and 2007 survey covered 76% of level 2 divisions available in the platform (GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Tanzania

Survey: National Panel Survey, 2008/09 (wave 1), 2010/11 (wave 2), and 2012/13 (wave 3)

Available information: The first wave covered 26 regions (first administrative level), 126 districts (second administrative level), and 360 divisions (third administrative level). The second wave covered 26 regions, 129 districts, and 732 divisions. Finally, the third wave covered 26 regions, 132 districts and 1018 divisions. Codes were available for all divisions in all waves. However, names were only available for the first two administrative divisions in the first two waves, while the third wave contained incomplete information on divisions names.

In order to complete the names for the third administrative divisions in all waves, official sources were used, however the official codes and the available in the databases did not match. Therefore, the available names in wave 3 were used to complete the first and second wave. In the third wave, from the 1018 divisions available, only 664 contained names. These names were matched with the second wave and 511 divisions were assigned with a name. Unfortunately, codes from the first wave where different from the other two waves, which made it impossible to assign ay name for the first wave. This means, no information for the third administrative divisions was uploaded for the first wave on Tanzania.

Matching with the GIS platform**:** Matching of the first three administrative divisions (when available) of Tanzania with the GIS platform was carried out using the GADM database. For the first level, matching was 100% in the three waves. For the second level it was 100% in wave 1, and 99% in wave 2 and 3 where one district did not match. Finally, for third level, in wave 2 matching was of 93% where 31 districts did not match, while in wave 3 it was 92% with 49 districts not matching. Information from wave 1 covered 87% of level 1 and 69% of level 2 divisions available in the platform (GADM database). While wave 2 covered 87% of level 1, 70% of level 2, and 19% of level 3 divisions available. Finally wave 3 covered 87% of level 1, 72% of level 2 and 17% of level 3 divisions. The remaining divisions in the platform were assigned values as described in the methodology.

Uganda

Survey: The Uganda National Panel Survey, 2005/06 and 2011/12

Available information**:** The 2005/06 survey covered 56 districts (first administrative level), 135 counties (second administrative level), and 345 sub counties (third administrative level). The 2011/2012 survey covered 108 districts, 185 counties, and 427 sub counties. These divisions were available in the databases by codes and names for the first level and for names for the second and third level.

Matching with the GIS platform: Matching of the first two administrative divisions with the GIS platform was carried out using the GAUL database, and matching of the third level was made with the GADM database. For the first level, matching was 100% for 2005/06 wave and 99% for the 2011/12 wave, where one district did not match. For the second level, matching was 100% for 2005/06 wave and 75% for the 2011/12 wave with 47 counties not matching. Finally, for the third level, matching was 100% for 2005/06 wave and 82% for the 2011/12 wave with 79 sub counties not matching. Information from the 2005/06 survey covered 50% of level 1, 80% of level 2, and 36% level 3 divisions available in the platform (GAUL and GADM database). While the 2011/12 survey covered 97% of level 1, 81% of level 2, and 36% of level 3 divisions available. The remaining divisions in the platform were assigned values as described in the methodology.

Viet Nam

Survey: Viet Nam Living Standards Survey, 1992

Available information: The survey covers 51 provinces (first administrative level), 147 districts (second administrative level) and 150 communes (third administrative level). Names were available for all divisions in the survey.

Matching with the GIS platform: Matching of the first two administrative divisions with the GIS platform was carried out using the GAUL database, and matching for the third level was using the GADM database. For the first level, matching was of 86% as 7 provinces did not match, while for the second level matching was of 88% where 17 districts did not match. Finally, for the third level matching was of 91% as 13 communes did not match. Information from the survey covered 69% of level 1, 20% of level 2, and 1% of level 3 divisions available in the platform (GAUL and GADM database). The remaining divisions in the platform were assigned values as described in the methodology.

Survey: Household Living Standard Survey, 2002 and 2008.

Available information: The 2002 survey covered 61 provinces (first administrative level), 607 districts (second administrative level) and 2,901 communes (third administrative level). Similarly, the 2008 survey covered 64 provinces, 643 districts, and 3,052 communes. Both surveys only had codes available and no names. Therefore, official administrative divisions for Vietnam were obtained from The Humanitarian Data Exchange Platform developed by OCHA [8]. Unfortunately, information was only available for the first two administrative divisions, meaning the third level was not matched with the GIS platform.

Matching with the GIS platform: Matching for the first two administrative levels of Viet Nam, for both 2002 and 2008, was carried out using the GAUL database. Matching at the first level was 100% for both years. Meanwhile, for 2002 matching was of 94% with 35 districts not matching and for 2008 matching was of 93% with 47 districts not matching. Information from the 2002 survey covered 95% of level 1 and 87% of level 2 divisions available in the platform (GAUL database). While the 2008 survey covered 100% of level 1 and 90% of level 2 divisions available. The remaining divisions in the platform were assigned values as described in the methodology.

---

[8] https://data.humdata.org/dataset/viet-nam-administrative-boundaries-polygon-polyline