

Peoples of South Asia

Version 10.0.1

In This Document

- [Introduction](#)
- [Features](#)
- [Installation](#)
- [Licensing](#)
- [Political Sensitivity](#)
- [Making Maps and Reports with the Omid data](#)
- [List of Maps with South Asia data](#)
- [Other Mapping Approaches](#)
- [Data Sources](#)
- [Tables in Omid.mdb](#)
- [Change History](#)

Introduction

The **Peoples of South Asia** dataset allows ArcGIS users to create maps and reports of peoples, languages, and religions in India, Pakistan, Afghanistan, Bangladesh, Myanmar, Nepal and Sri Lanka from an extensive (three-quarters of a million records) database at a district level (province in Afghanistan). Data is consistent with the most recent available census for which ethnographic and linguistic data is available for each country (2001 in India and Pakistan) but data from earlier censuses (pre-1948) has been used to estimate the ratios of various community (caste and tribal) groups not counted individually in the current census. The data was collected and extrapolated by a researcher who prefers to be known simply as OMID, word meaning “hope” in Hindi, Urdu, and Farsi. The name OMID will appear throughout the data product.

Features

- Population density (persons per district, persons per square mile, persons per square kilometer) can be mapped for 2,710 peoples (including caste groups), 850 languages, and 11 religions.
- Map titles and statistical summaries are automatically generated for each group queried
- Query can be re-run and map refreshed immediately to allow rapid mapping of a number of groups.
- Database includes generation of statistical reports in addition to maps.
- Advanced users may query and map the database in additional ways, such as religious or linguistic subgroups of peoples.

Installation

The Peoples of South Asia dataset is compressed in setup.exe in the downloadable zip file or CD-ROM, which must be executed on a Windows system. Setup.exe will prompt for an

installation password, which is provided by the WorldGeoDatasets.com store on the download page and confirmation email. It is recommended, but not required, that the dataset be installed in the same folder as other WorldGeoDataSets.com data.

Licensing

This data set is covered by an [End-User License](#), and is subject to both copyright and license restrictions. In particular, redistribution of the data is not allowed except as specifically permitted in the End User License Agreement or by explicit, prior, written permission of Global Mapping International.

Political Sensitivity

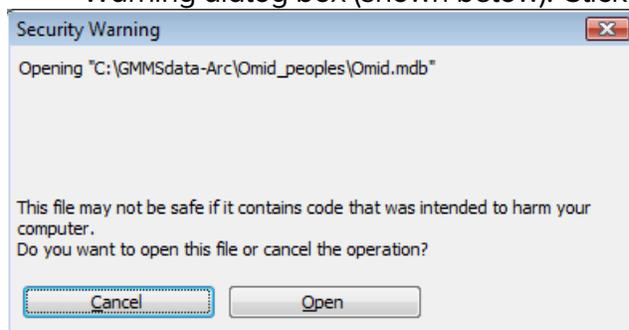
Note that the borders of Jammu and Kashmir are a very sensitive issue with the Indian and Pakistani governments. We have included ArcView project files (.mxd) to show the Indian view and international view of the border. The Indian view includes areas occupied by Pakistan and China but claimed by India as part of India. The international view includes these areas as part of Pakistan and China, but uses a dashed line to show Indian, Pakistani, and Chinese claim and has a note describing the disputed area. If your maps will be published or displayed publicly in India, it is **VERY important** to use only mxd's beginning with *IndView* when making India maps; our understanding is that it is a criminal offense there to publish maps showing borders for India other than the official claim lines. If you want to make maps of all the countries of South Asia but still show the Indian View, there are mxd's in the \S_Asia\IndianView folder which will show this view. To use these mxd's, copy them from the IndianView folder into the \S_Asia folder (where the other South Asia data mxd's are located). If the maps are for audiences in other countries, you can use the mxd's which begin with *IntlView*.

GMI is less familiar with sensitivities on these issues in Pakistan and China, but we believe it would be insensitive, if not dangerous, to use the *IndView* maps in these countries, and may be problematic to use even the international view. If you are a user who needs to distribute maps in Pakistan or China and need to adjust borders to reflect your country's sensitivities in these disputed areas, please contact mapsupport@gmi.org for assistance. We are willing to add other national views to the standard distribution of this data if there is substantial demand.

Making Maps and Reports with the Omid data

Using the Omid data requires you to have two programs running: a Microsoft Access database to select the statistical data used, and ArcMap to map the data. You can change the map to show a new group of the same statistical data by simply making a new selection and refreshing the map. Use the following steps:

1. In Windows Explorer (My Computer), navigate to the folder where you have installed the South Asia data (by default \GMI_GEODATA\S_Asia. Double click on Omid.mdb. If you are using the run-time Microsoft Access program you will receive a Security Warning dialog box (shown below). Click Open.



If you are using the Windows Vista operating system with run-time Microsoft Access, you will next get a dialog box with a Security Warning about blocking unsafe expressions. Enter Yes. Click OK to restart the run-time Microsoft Access and repeat the above step from the main menu.

2. Go back to Windows Explorer (My Computer) and in the same folder as above double click on one of the mxd files to open a map in ArcMap. See [List of maps of South Asia Data](#) below for a full list and description of maps. Note that you can select a map which only shows data for India or data for all countries in South Asia. Some maps will show the Jammu and Kashmir situation as India sees it (*IndView*) or as the international community sees it (*IntlView*). Maps show the following types of data as selected in the database:
 - Peoples (known as Communities in the database, and as peoples, castes and tribes in common usage)
 - Language Speakers
 - Religious Adherents

Variants of these maps report density of the selected group in percent of population, persons per district, persons per square kilometer, or persons per square mile. When a map project is opened, the map title and thematic data will reflect the most recent selection of the type required for the map made in the database. For the example below open the *IndViewIndiaPeoples.mxd*.

Additional maps show several data items pre-calculated on a per-district basis; the database does not have to be opened or selections made:

- Population density (persons per square kilometer or mile)
- Max Language (the language with the greatest number of speakers per district)

The screenshot shows two panels of the 'South Asia Reporting and Mapping' interface. The top panel has a title bar and two options: 'Reporting for South Asia' (selected with a checked checkbox) and 'Mapping front-end for South Asia' (unchecked). The bottom panel also has a title bar and six options: 'Build mapping work tables for one community' (checked), 'Build mapping work tables for one language' (unchecked), 'Build mapping work tables for one religion' (unchecked), 'Miscellaneous mapping maintenance' (unchecked), 'Search community names' (unchecked), and 'Return to main menu' (unchecked).

When the mxd opens, it will display the thematic data and title corresponding to the most recent query of the type desired in the database.

3. Go back to the Microsoft Access or run-time MS-Access program. You will see the opening South Asia Reporting and Mapping screen (right). Select “*Mapping front-end for South Asia*”. Note that reports for peoples, language, religions for a country or sub-country level can be created by selecting the other option, “*Reporting for South Asia*”. We are thankful to Bill Morrison from Joshua Project for this MS–Access interface.
4. The next form to appear allows you to choose what type of map to make (see below). Click on the button matching the map you opened in ArcMap (i.e. select “*Build mapping work tables for one community*”) since the mxd which was opened shows communities (peoples). (**Note: it is common for the same or similar names to**

be used both for a community (people group) and the language spoken by a people. If a change in group selection does not change your map as expected, it is likely there is a mismatch between the type of data being selected and the map project being displayed.)

- The following discussion applies to the selection of "Build Mapping Work Tables for One Community". See screen shot below. Options for other query screens for other data types may be slightly different, but are mostly similar.

Note that to get data for a map for single country, use the top part of the form. For a map of all countries in South Asia, use the bottom part of the form. Use the pull down in *Step 1* to select only the Communities present in a particular country. Use the pull down in *Step 2*: to choose a Community within that country from the list which appears (list may take a few seconds to appear). A check box determines whether only data for the selected country, or data for the selected community in all countries, will be mapped and included in the title statistics. A second check box allows inclusion of data for disputed districts in India and Pakistan.

For our example, since *IndViewIndiaPeoples.mxd* was opened in ArcMap, select "India" in *Step 1*. Select any community from *Step 2* (you may want to select one with a somewhat large population so you could see it on the map). Click to remove the check from "Map this Community in all countries", and click to check "Include disputed districts..."

After you select a Community, click the "Create Mapping Table" button to extract the data for the selected Community. Note that the terms "Community" and "People" are somewhat confusingly used in this screen. The meaning of the two terms is identical. A message will appear that the mapping table is created.

- Using the Windows taskbar, return to ArcMap.
- Click on the refresh view icon  near the lower left of the layout frame. The map will redraw showing the name and summary statistics for the selected people.

8. Alternate between the query in the Omid database and refreshing the map to create maps for additional peoples.
9. The India mxd files show data for the country of India only. Thus the India mxd's should only be used if the country of India is chosen in the above form. If the India mxd's are used in other cases, the summary numbers under the title (number of districts and population for the community) will not match what is shown on the map and the data for other countries won't display. Thus only use the mxd files which have SouthAsia in their name to see data for countries other than India.

List of Maps of South Asia Data

The following are mxd files in the S_Asia folder. Mxd and layer files should work with any version of ArcView from 9.0 and up. An asterisk (*) indicates a layer file also exists.

- IndViewIndiaLanguages.mxd – Number of speakers of specific language of India by district, disputed political boundaries from India perspective
- IndViewIndiaPeoples.mxd – Number in specific people group (community) of India by district, disputed political boundaries from India perspective
- IndViewIndiaReligions.mxd – Number of adherents of specific religion of India by district, disputed political boundaries from India perspective
- IntlViewIndiaLanguages.mxd - Number of speakers of specific language of India by district, disputed political boundaries using line of control (international view of boundaries)
- IntlViewIndiaPeoples.mxd - Number in specific people group (community) of India by district, disputed political boundaries using line of control (international view of boundaries)
- IntlViewIndiaReligions.mxd - Number of adherents of specific religion of India by district, disputed political boundaries using line of control (international view of boundaries)
- IntlViewSouthAsiaLanguages (per sq km).mxd* - Number of speakers of specific language per square kilometer in each district/province of countries in South Asia
- IntlViewSouthAsiaLanguages (per sq mi).mxd* - Number of speakers of specific language per square mile in each district/province of countries in South Asia
- IntlViewSouthAsiaLanguages (percent).mxd* - Percent of speakers of specific language compared to speakers of all languages in each district/province of countries in South Asia
- IntlViewSouthAsiaLanguages (person).mxd* - Number of speakers of specific language in each district/province of countries in South Asia
- IntlViewSouthAsiaMaxLanguage.mxd – Name of language with most speakers in each district/province of countries in South Asia
- IntlViewSouthAsiaPeoples (per sq km).mxd* – Number in specific people group (community) per square kilometer in each district/province of countries in South Asia
- IntlViewSouthAsiaPeoples (per sq mi).mxd* - Number in specific people group (community) per square mile in each district/province of countries in South Asia

- IntlViewSouthAsiaPeoples (percent).mxd* - Percent in specific people group (community) compared to members of all people groups in each district/province of countries in South Asia
- IntlViewSouthAsiaPeoples (person).mxd* - Number in specific people group (community) in each district/province of countries in South Asia
- IntlViewSouthAsiaPopDensityKm.mxd* - Population density by square kilometer in each district/province of countries in South Asia
- IntlViewSouthAsiaPopDensityMile.mxd* - Population density by square mile in each district/province of countries in South Asia
- IntlViewSouthAsiaReligions (per sq km).mxd* - Number of adherents of specific religion per square kilometer in each district/province of countries in South Asia
- IntlViewSouthAsiaReligions (per sq mi).mxd* - Number of adherents of specific religion per square mile in each district/province of countries in South Asia
- IntlViewSouthAsiaReligions (percent).mxd* - Percent of adherents of specific religion compared to adherents of all religions in each district/province of countries in South Asia
- IntlViewSouthAsiaReligions (person).mxd* - Number of adherents of specific religion in each district/province of countries in South Asia

The following mxd's are located in the \S_Asia\IndianView folder. They show data for all countries in South Asia, but the Indian borders are drawn to show an India political view. To be used, the mxd's will need to be copied to the \S_Asia folder. Since these mxd's have only limited interest and use, they were put in a separate folder.

- SouthAsiaLanguages (percent).mxd – Percent of speakers of specific language compared to speakers of all languages in each district/province of countries in South Asia, disputed political boundaries from India perspective
- SouthAsiaLanguages (person).mxd – Number of speakers of specific language in each district/province of countries in South Asia, disputed political boundaries from India perspective
- SouthAsiaMaxLanguage.mxd – Number of speakers of specific language in each district/province of countries in South Asia, disputed political boundaries from India perspective
- SouthAsiaPeoples (percent).mxd – Percent in specific people group (community) compared to members of all people groups in each district/province of countries in South Asia, disputed political boundaries from India perspective
- SouthAsiaPeoples (person).mxd – Number in specific people group (community) in each district/province of countries in South Asia, disputed political boundaries from India perspective
- SouthAsiaPopDensityKm.mxd – Population density by square kilometer in each district/province of countries in South Asia, disputed political boundaries from India perspective
- SouthAsiaPopDensityMile.mxd – Population density by square mile in each district/province of countries in South Asia, disputed political boundaries from India perspective

- SouthAsiaReligions (percent).mxd - Percent of adherents of specific religion compared to adherents of all religions in each district/province of countries in South Asia, disputed political boundaries from India perspective
- SouthAsiaReligions (person).mxd - Number of adherents of specific religion in each district/province of countries in South Asia, disputed political boundaries from India perspective

Other Mapping Approaches

The standard projects described above are designed to quickly show a wide range of population sizes of peoples, languages, or religions on a standard map. Doing this required showing populations on an exponential scale. A wide variety of other approaches are possible, and will likely yield maps that communicate the distribution of a group more accurately and intuitively. Some ideas include:

- Adjusting classification to give a more linear scaling with endpoints appropriate to the particular group
- Using dot-density fills (1 dot = x persons). For groups with significantly different spatial distribution, it is possible to show multiple groups on the same map with different colored dots
- Investigate other ways of querying the data, using the Microsoft Access OLEDB connection in Database Connections in ArcCatalog or Add Layer to connect to the query. See [Building queries in Omid.mdb](#)

Nearly all of the development of the current system (both the map templates and the database design) has been done by users who wanted to find a new way of looking at the Omid data. If you develop an idea that others might benefit from, contact Bill Dickson (bill@gmi.org) to explore having your sample projects added to the Omid distribution.

Data Sources

- The Omid data was collected and extrapolated from pre-1948 Census data (particularly the “Great Census” of 1934) by a researcher who prefers to be known simply as Omid, a word meaning “hope” in Hindi, Urdu, and Farsi. After 1948, the Census of India did not allow full self-identification of caste membership. Data from these early censuses has been used to subdivide data from the latest available census into the smaller subgroups using the ratios of populations of various groups from earlier censuses.
- Shapefiles have been adapted from the Digital Chart of the World (VMap level 1) coastlines and international borders. This is a 1:1,000,000 basemap for world.
- **India_prov.shp** – states of India to 2001 census created by GMI. **ID** field contains ISO 3 letter country code – 3 letter GMI code (often first 3 letters or first 3 consonants of name). **HISCODE** field contains the FIPS10-4 code for state (U.S. government code - two letter country, two digit number or letters for state). The **ISOCODE** field contains a two letter ISO country code with a two letter province code (some of these are incomplete).
- **omid_dist.shp** – districts of India according to 2001 census from GMI – DCW data supplement. District borders of Pakistan, Nepal, Bhutan, Bangladesh, and Sri Lanka

from UNESCO (1987) through UNEP/GRID-Sioux Falls (these may not be updated to 2001). Afghanistan provinces (first level admin borders) from DCW data supplement (doesn't have one 2001 update). **ID** field for India is ISO 3 letter country code – state code – 3 letter GMI code for district (often first 3 letters or first 3 consonants of name). This is also true for Bangladesh. For other countries this is just province code (so will not be unique for districts). **Local code** field is the numeric India 2001 Census code. The **OMIDCODE** field is a two letter state/province code with a district code which joins to Omid data.

Tables in Omid.MDB (for advanced users, good knowledge of Access required)

Some fields contain very few data records especially latter fields in COMMUNITY DATA and LANGUAGE DATA. It would be best to work from the sample queries at the end of this documentation. If you do work with the tables to get accurate results it is very important to know how the data is structured.

COMMUNITY DATA

This main table contains people group data for the south Asian countries coded by district, province, or country. Note that very low one digit, non-zero values usually don't indicate actual values, but that there are at least some of the particular community in the district. "Some" could mean from a few to several thousand.

Following is a list of the fields along with descriptions.

Field Name	Type	Wid	Dec	Description
COMMUNITY	String	50		This field contains the names of people groups such as Brahmans. An entry of AALL indicates a summary of all the communities in the district. Each community is identified by a unique combination of community name and NCode. If EITHER changes as you pass thru the file you have come to a new community. There are a small number of communities that are distinct communities, same name, different NCODE
SUBGROUP	String	50		Subgroup of community. Normally records with non-nulls in this field should be bypassed, else you will double count the populations. This field is in a state of flux with a number of subgroups becoming identified as try people groups. Example Bania Agarwal used to be as COMMUNITY Bania SUBGROUP Agarwal. Where this has been done there are 2 extra fields, Group Community Name and Community Family, in this case Bania is entered in both, but not in the matching line against Bania which is the total for all Bania
SUBSUB	String	50		Sub-subgroup. Seldom used.

Field Name	Type	Wid	Dec	Description
DISTRICT	String	50		This field contains the code for each district (each province in Afghanistan). The first two letters are an abbreviation of the state/province while the remaining letters are an abbreviation of the district or in name. In Afghanistan the first two letters are an abbreviation (AF) for the country and the remaining three letters are an abbreviation for the province. Note that if the remaining 3 letters are AAA the record is a summary record for a whole province or state. The XrefDistricts table can be used to lookup the district names for the district codes.
STATUS	String	50		Housekeeping field used by Omid. Generally sources of the data; D1931 is 1931 Census, D1991 is 1991 Census, L1961 is Linguistic Census of 1961, D1891 is 1891 Census, SC91 is Scheduled Caste tables of 1991 Census, ST91 is Scheduled Tribes tables of 1991 Census. There are other data sources
ALTNAME	String	250		Alternate name(s) for the community
NCODE	String	50		This code distinguishes between different religious subgroups within a particular community. Oftentimes, it is necessary to include the NCODE in a query in order to get one record per district.
DISTRICT_1	Double	8		2001 population of the community
HX2001	Double	8		2001 Hindu population of the community
MX2001	Double	8		2001 Muslim population of the community
CX2001	Double	8		2001 Christian population of the community
BX2001	Double	8		2001 Buddhist population of the community
JX2001	Double	8		2001 Jain population of the community
SX2001	Double	8		2001 Sikh population of the community
IX2001	Double	8		2001 Indigenous/Other religion population of the community
STATE	String	50		Two letter code for the state. The XrefStateCodes table can be used to lookup the state names for the state codes
NATION	String	50		Country name
not_in_addup	String	50		Indicates records (generally sub community, sub sub community, and aggregate records) which should not be included when adding up national, state, or district populations
notes	String	50		Notes about community and data
DISTRICT_NAME	String	50		District name. This field is usually empty. It is better to use the XrefDistricts table to determine the district name. Being added as data is reviewed; always present for AALL (all communities) aggregate records
OCCUPATION1	String	50		Primary traditional occupation of caste group
SUBDISTRICT	String	50		Sub-district geographic identifier (seldom used)
OCCUPATION2	String	50		Additional traditional occupation of caste group
OCCUPATION3	String	50		Additional traditional occupation of caste group
OCCUPATION4	String	50		Additional traditional occupation of caste group
OCCUPATION5	String	50		Additional traditional occupation of caste group

Field Name	Type	Wid	Dec	Description
MXGEO	String	50		A term that identifies Muslims by Geography; example, Muslim of Bengal, Muslim of Khandesh; best left out of any addup queries. Of use primarily for those interested in Muslims specifically.
GEOGRAPHICAL_MUSLIM	String	50		{Construction field; to be deleted in final release}
SUBDISTRICT_NAME	String	50		Applicable for data at a sub-district level (tehsil, cd block, block, taluka, etc.)(seldom used)
TAB	String	50		{Construction field; to be deleted in final release}
KX2001	Integer	4		2001 Kirat population of the community
Ba2001	Integer	4		2001 Ba'hai population of the community
JEW2001	Integer	4		2001 Jewish population of the community
ZX2001	Integer	4		2001 Zoroastrian population of the community
HXGEO	String	50		{Construction field; to be deleted in final release}
group_community_name	String	50		Used for Bania, Jat, Brahman where subgroups are being reclassified as communities. All Bania, for instance, have Bania in this field. Does not affect addup queries as long as COMMUNITY FAMILY field is set
dalit	String	50		Is group dalit (untouchable)?
village	String	50		Rarely used; village level geographic descriptor
language_	String	50		Predominant language of group
ethcode	String	50		Ethnologue code for language
community_family	String	50		Used for Bania, Jat, Brahman, etc. where subgroups are becoming reclassified as distinct communities. In the main entry for Bania, Jat, Brahman and other communities this field is empty. It contains Bania, Jat, etc. against community names such as Bania Agarwal or Jat Sindhu. In an addup query this field MUST be set at "IS NULL"
not_significant	String	50		Flag to indicate groups with insignificant presence, generally omitted from queries. Added because of changes being made to Jat, Bania, and Brahman, soon Rajput, Baloch, Brahui, and other large communities
MIGRANT_FACTORED	String	50		Only important for S Asia migrants in Canada, US, NZ, Australia. Includes a scaling from language data to cover migrants who identified English instead of a South Asian language on census forms. It has only been done for S Asia, and Pathan/Pushtun/Pashto who overlap Afghanistan and rest of S Asia

wrkMapTable

Note that wrkMapTable is used to map data for a community. This table is built from the COMMUNITY DATA table each time a community is specified in the Switchboard. For mapping the DistrictCode field is joined with the OMIDCODE field in amid_dist.shp.

Field Name	Type	Wid	Dec	Description
Nation	String	50		Country name
Community	String	50		Name of Community (people group)
NCode	String	50		This code distinguishes between different religious subgroups within a particular

Field Name	Type	Wid	Dec	Description
				community.
State	String	50		Two letter code for the state.
DistrictCode	String	50		This field contains the code for each district (each province in Afghanistan). The first two letters are an abbreviation of the state/province while the remaining letters are an abbreviation of the district or in name. In Afghanistan the first two letters are an abbreviation (AF) for the country and the remaining three letters are an abbreviation for the province.
TotalDistrictPopl	Integer	4		Total population in district
Population	Double	8		2001 population of the community (people group) in district
Hindus	Double	8		2001 Hindu population of the community
Muslim	Double	8		2001 Muslim population of the community
Christians	Double	8		2001 Christian population of the community
Buddhists	Double	8		2001 Buddhist population of the community
Jains	Double	8		2001 Jain population of the community
Sikhs	Double	8		2001 Sikh population of the community
Jewish	Double	8		2001 Jewish population of the community
Zoroastrian	Double	8		2001 Zoroastrian population of the community
Other	Double	8		2001 Indigenous/Other religion population of the community
Kirat	Double	8		2001 Kirat population of the community
Bahai	Double	8		2001 Ba'hai population of the community

LANGUAGE DATA

This main table contains language data for the south Asian countries coded by district, province, or country. Note that very low one digit, non-zero values usually don't indicate actual values, but that there are at least some speakers of the particular language in the district. "Some" could mean from a few to several thousand.

Following is a list of the fields along with descriptions.

Field Name	Type	Wid	Dec	Description
COMMUNITY	String	50		This field contains the names of people groups such as Brahmans. An entry of AALL indicates a summary of the speakers of all languages in the district
SUBGROUP	String	50		Subgroup of community
SUBSUB	String	50		Sub-subgroup of community
DISTRICT	String	50		This field contains the code for each district (each province in Afghanistan). The first two letters are an abbreviation of the state/province while the remaining letters are an abbreviation of the district or in name. In Afghanistan the first two letters are an abbreviation (AF) for the country and the remaining three letters are an abbreviation for the province. Note that if the remaining 3 letters are AAA the record is a summary record for a whole province or state.
DISTRICT_1	Double	8		Number of speakers in that district
LANGUAGE_	String	50		Language name
STATUS	String	50		Social status of group, e.g. Scheduled Caste, Scheduled Tribe, Forward Caste, etc.
ALTNAME	String	250		Alternate name(s) for the community.
NCODE	String	50		This code distinguishes between different religious subgroups within a particular community. Oftentimes, it is necessary to include the NCODE in a query in order to get one record per district.

Field Name	Type	Wid	Dec	Description
TAB	String	50		Not used
STATE	String	50		Two letter code for the state.
NATION	String	50		Country name
NOT_IN_ADDUP	String	50		Group is a subgroup of another and should not be included in statistics at a district level
HXGEO	String	50		{Construction field; to be deleted in final release}
MXGEO	String	50		A term that identifies Muslims by Geography; example, Muslim of Bengal, Muslim of Khandesh; best left out of any addup queries. Of use primarily for those interested in Muslims specifically.
NOTES	String	50		Notes about community and data
TOTPOP	Double	8		Total population
SUBDISTRICT	String	50		Applicable for data at a sub-district level (tehsil, cd block, block, taluka, etc.)(seldom used)
GEOGRAPHICAL_MUSLIM	String	50		{Construction field; to be deleted in final release}
ETHCODE	String	50		Three letter Ethnologue code of language
LANGORDER	String	50		{Construction field; to be deleted in final release}
district_name	String	50		District name. Frequently empty. It is better to use the XrefDistricts table to determine the district name. Being added as data is reviewed.
group_community_name	String	50		Used for Bania, Jat, Brahman where subgroups are being reclassified as communities. All Bania, for instance, have Bania in this field. Does not affect addup queries as long as COMMUNITY FAMILY field is set
county	String	50		{Construction field; to be deleted in final release}
subdistrict_name	String	50		Subdistrict name (seldom used)
migrant_factored	String	50		Only important for S Asia migrants in Canada, US, NZ, Australia. Includes a scaling from language data to cover migrants who identified English instead of a South Asian language on census forms. It has only been done for S Asia, and Pathan/Pushtun/Pashto who overlap Afghanistan and rest of S Asia
community_family	String	50		Used for Bania, Jat, Brahman, etc. where subgroups are becoming reclassified as distinct communities. In the main entry for Bania, Jat, Brahman and other communities this field is empty. It contains Bania, Jat, etc. against community names such as Bania Agarwal or Jat Sindhu. In an addup query this field MUST be set at "IS NULL"
NOT_SIGNIFICANT	String	50		Flag to indicate groups with insignificant presence, generally omitted from queries. This have been added because of changes being made to Jat, Bania, and Brahman, soon Rajput, Baloch, Brahui, and other large communities
ROL3	String	5		Ethnologue code for language (As supplemented or corrected by Joshua Project; generally more reliable than ETHCODE)
ROL4	String	5		Code for dialect

WrkLangMapTable

Note that wrkLangMapTable is used to map data for a language. This table is built from the LANGUAGE DATA table each time a language is specified in the Switchboard. For mapping the DistrictCode field is joined with the OMIDCODE field in omid_dist.shp.

Field Name	Type	Wid	Dec	Description
Nation	String	50		Country name
LangName	String	50		Name of language
DistrictCode	String	50		This field contains the code for each district (each province

Field Name	Type	Wid	Dec	Description
				in Afghanistan). The first two letters are an abbreviation of the state/province while the remaining letters are an abbreviation of the district or in name. In Afghanistan the first two letters are an abbreviation (AF) for the country and the remaining three letters are an abbreviation for the province.
TotalDistrictLanguagePopl	Integer	4		Total population of all languages in district
Speakers	Double	8		Total population selected language in district

RELIGION DATA

This table is built from the COMMUNITY DATA and LANGUAGE DATA tables. It contains religion data for the Omid countries coded by district, province, or country. Note that very low one digit, non-zero values usually don't indicate actual values, but that there are at least some of the particular religion in the district. "Some" could mean from a few to several thousand.

Following is a list of the fields along with descriptions.

Field Name	Type	Wid	Dec	Description
Nation	String	50		Country name
DistrictCode	String	50		This field contains the code for each district (each province in Afghanistan). The first two letters are an abbreviation of the state/province while the remaining letters are an abbreviation of the district or in name. In Afghanistan the first two letters are an abbreviation (AF) for the country and the remaining three letters are an abbreviation for the province. Note that if the remaining 3 letters are AAA the record is a summary record for a whole province or state.
Religion	String	255		Name of religion
Adherents	Double	8		Practitioners of religion
TotalDistrictPopl	Integer	4		Total population of district

WrkReligionMapTable

Note that wrkReligionMapTable is used to map data for a religion. This table is built from the RELIGION DATA table each time a religion is specified in the Switchboard. For mapping the DistrictCode field is joined with the OMIDCODE field in *omid_dist.shp*.

Field Name	Type	Wid	Dec	Description
Nation	String	50		Country name
Religion	String	255		Religion name
DistrictCode	String	50		This field contains the code for each district (each province in Afghanistan). The first two letters are an abbreviation of the state/province while the remaining letters are an abbreviation of the district or in name. In Afghanistan the first two letters are an abbreviation (AF) for the country and the remaining three letters are an abbreviation for the province.
SumOfAdherents	Double	8		Practitioners of religion

wrkMaxLanguageByDistrict

This table is build by using Miscellaneous Maintenance from the switchboard and then the Build Maximum Language Table button. Unless new Omid data is imported, the table will exist already, so it does not have to be rebuilt each time the database is opened. The table tells which language has the most speakers in a particular district. For mapping the District field is joined with the OMIDCODE field in *omid_dist.shp*.

Field Name	Type	Wid	Dec	Description
Nation	String	50		Country name
District	String	50		This field contains the code for each district (each province in Afghanistan). The first two letters are an abbreviation of the state/province while the remaining letters are an abbreviation of the district or in name. In Afghanistan the first two letters are an abbreviation (AF) for the country and the remaining three letters are an abbreviation for the province.
Lang	String	50		Language name
Speakers	Double	8		Number of speakers of language

wrkMapTitle

This table passes information for the title for a community map. This would include the people group name (Joshua Project name), number of districts that have members of the community, and total number of community.

Field Name	Type	Wid	Dec	Description
Title	String	255		Title (not used)
Nation	String	255		Country Name (not used)
Community	String	50		Community (people group) name
NCode	String	50		This code distinguishes between different religious subgroups within a particular community.
CountOfDistrictCode	Integer	4		Number of districts where Community (people group) live
PoplTotal	Double	8		Total population of Community (people group) in all districts
JoshuaProjectName	String	50		Joshua Project name
Disputed	String	255		Are disputed districts of India included?

wrkLangMapTitle

This table passes information for the title for a language map. This would include the language name, number of districts that have speakers of the language, and total number of speakers.

Field Name	Type	Wid	Dec	Description
Title	String	255		Title (not used)
Nation	String	255		Country name (not used)
LangName	String	50		Language name
CountOfDistrictCode	Integer	4		Number of districts where speakers of language live
TotalSpeakers	Double	8		Total number of speakers in all districts
Disputed	String	255		Are disputed districts of India included?
ROL3	String	5		Ethnologue code for language (As supplemented or corrected by Joshua Project; generally more reliable then ETHCODE)

wrkReligionMapTitle

This table passes information for the title for a religion map. This would include the religion name, number of districts that have practitioners of the religion, and total number of practitioners.

Field Name	Type	Wid	Dec	Description
Title	String	255		Title (not used)
Nation	String	255		Country Name (not used)
Religion	String	255		Name of Religion
CountOfDistrictCode	Integer	4		Number of districts where adherents of religion live
Adherents	Double	8		Number of adherents in all districts
Disputed	String	255		Are disputed districts of India included?

DISTRICT LANGUAGE SUMS

This table contains the sum of all languages in the districts (provinces for Afghanistan).

DISTRICT SUMS

This table contains the total number of people in each district (province in Afghanistan).

wrkStateCodesXRef

Cross reference table of state codes to state names used in current wrk tables

XrefDistricts

Cross reference table of district codes to district names

Field Name	Type	Wid	Dec	Description
ROG3	String	255		Registry of Geography 2 letter country code (usually FIP 10-4 code)
ROG4	String	4		Register of Geography 4 letter province code (usually FIPS 10-4 province code) . ROG3 2 letter country with two letter province
Nation	String	255		Name of country
StateCode	String	255		2 letter state code (first two letters of District or DistrictCode used in other files)
StateName	String	255		State name
DistrictCode	String	255		Code for district (last three letters of District or DistrictCode used in other files)
DistrictName	String	255		District name (provinces in Afghanistan)
Comment	String	255		Comment
Disputed	String	255		Is this a disputed district of India?

XrefStateCodes

Cross reference table of state codes to state names

wrkJPNames

Cross listing of community names, NCODEs, ROP3s, and Joshua Project names.

Field Name	Type	Wid	Dec	Description
Community	String	50		Community (people group) name
NCode	String	50		This code distinguishes between different religious subgroups within a particular community.
ROP3	Integer	4		Standard code for people group (Joshua Project)
JoshuaProjectName	String	75		Joshua Project name

Building queries in Omid.mdb

If you are building your own queries using the COMMUNITY DATA and LANGUAGE DATA tables, the following three rules will help avoid using duplicate records or sidestepping work in progress:

1. Use records where SUBGROUP is null (SUBGROUP is Null) unless you are specifically querying for a subgroup or sub-subgroup.
2. Use records where NCODE is not null (NCODE is not Null)
3. Use records where NOTINADDUP is null (NOTINADDUP is Null)

The two tables have only district codes, so relating the table to a xref table will probably be necessary.

Besides queries that are used by items in the switchboard, several sample queries have been built which can create new tables. These queries may be helpful to see what relationships are needed between the various tables. By changing the name of the community (e.g. in qryAssameseSpeakersByCommunity, choose a different community name than Assamese) these queries could be easily modified to a different community. The query names are fairly self-explanatory.

- QryAssameseSpeakersByCommunity
- QryAssameseSpeakersByDistricts
- QryBengaliSpeakersByCommunity
- QryBengaliSpeakersByDistrict

Change History

Changes refer to the Omid data and geography in Omid_Peoples unless otherwise noted.

August 2010

- Updated Omid.mdb database
- Omid.mdb language data has not been subdivided for the relatively recently created provinces of Afghanistan Khowst, Nurestan, and Sar-e Pol. We have provided separate shapefile geometry for languages with these provinces merged back into their former provinces (e.g. Khowst was merged into Paktia, Nurestan into Konar and

Laghman, and Sar-e Pol into Jowzjan). Peoples and religion maps have data for the new provinces and use a separate geometry with the new provinces defined.

January 2010

- First release not as part of the Global Ministry Mapping System.
- Some corrections/updates to data.
- Joins are now used instead of a database connection to associate the OMID data with geography. Thus the main folder can be placed any where in data structure.
- Mxd's created to show density of peoples and language speakers, i.e. speakers per square kilometer, and percent of people group or language as compared to total population of district or province. Also mxd's have been created to show population density. Layer files for many mxd's have also been added.
- Three provinces of Afghanistan, Sar-e Pol, Nurestan, and Kowst were reverted back to 1988 state, since no statistical data had been created for these provinces. It is hoped this will be reversed in the future.