



Global Mapping International

World Language Mapping System

Version 16

In This Document

- [Welcome](#)
- [What's New in Version 16](#)
- [Features](#)
- [Installation](#)
- [Licensing](#)
- [Release History](#)
- [Sample Map Projects](#)
- [Data Organization](#)
- [Detailed Data Documentation](#)
 - [Source: GMI and NIMA VMap0, Edition 5](#)
 - [Country area features](#)
 - [Source: World Language Mapping System](#)
 - [Language area features](#)
 - [Mixed Language area features](#)
 - [Other Region Language area features](#)
 - [Language point features](#)
 - [Extinct Language point features](#)
 - [Widespread Language point features](#)
 - [Other Language point features](#)
 - [Unknown Location Language point features](#)
 - [Widespread Languages, Offset for Display point features](#)
 - [Source: SIL](#)
 - [Ethnologue Language attribute data](#)
 - [Source: GMI/Dr. Merritt Ruhlen](#)
 - [Conversion for Language Families](#)
 - [Source: Globe/GMI](#)
 - [Global Elevation/Bathymetry Colors compressed raster](#)
 - [Global Hillshades compressed raster](#)
 - [Global Elevation/Bathymetry Colors+Hillshades compressed raster](#)

Other Documents

- [End User License Agreement](#)

Welcome

The World Language Mapping System (WLMS) is the result of the Language Mapping Project, an eighteen-year collaboration between Global Mapping International (GMI) and the SIL International (SIL). The project seeks to map the languages described in SIL's *Ethnologue* book and web site (<http://www.ethnologue.com/>). The resulting GIS data is used by SIL to produce maps for the *Ethnologue* and other purposes, and has been published by GMI in GIS format for the Christian nonprofit community since 1995 as part of the *Global Ministry Mapping System*. This data set, updated with every edition of the *Ethnologue*, is available to the GIS community as the [World Language Mapping System](#).

What's New in Version 16

Changes between Version 3.2.1 (15th Edition *Ethnologue*) and Version 16 (16th Edition *Ethnologue*) include:

- Point, polygon, and attribute data have been updated to correspond to the 16th Edition of the *Ethnologue*.
- Language polygons for widespread languages have been added in nearly all populated areas where there is no minority language and a there is a single, dominant national language, e.g. Spanish in Latin America, Portuguese in Brazil, etc.

- Layers were added to display SIL language families, Ruhlen language families, and population of languages for points and polygons. Table was added to cross reference SIL language families with Ruhlen language families.
- When installed with the Seamless Digital Chart of the World, language features can be controlled by the base map wizard
- Major version number will henceforth correspond to the *Ethnologue* edition number

Features

- Polygons delineating the linguistic homelands of most of the language-in-country entries in the *Ethnologue, 16th edition*, with coastlines and international borders corresponding to the Digital Chart of the World (VMAP level 0, edition 5).
- Polygon overlays for mixed-language areas.
- Point locations of all *Ethnologue* languages.
- Additional polygons of language areas without corresponding *Ethnologue* listings (generally areas where a small part of a linguistic homeland described in the *Ethnologue* extends into an adjoining country).
- A joinable attribute table with fields nearly corresponding to the full data in the published *Ethnologue: Languages of the World, 16th Edition*, along with additional data developed by the Language Mapping Project and additional, unpublished *Ethnologue* data on immigrant languages.
- Layer files providing default symbology for ArcGIS.
- Sample map project (.mxd) files illustrating use in ArcGIS.
- Vector country boundaries and ECW-compressed raster elevation colors and hillshades to provide geographic context. Additional geographic context can be added using GMI's Seamless Digital Chart of the World or other forms of the Digital Chart of the World (VMap Level 0) base map.
- XML metadata compatible with ArcGIS.

Installation

The World Language Mapping System data is compressed in setup.exe on the CD-ROM, which must be executed on a Windows system. Setup.exe will prompt for a user name, company name, serial number, and unlocking code. The company name (even if misspelled or added to), serial number, and unlocking code must be entered exactly as provided by GMI, generally on a label inside the front cover of the CD-ROM case or in an email message. If the WLMS is being installed for use in conjunction with GMI's Seamless Digital Chart of the World data product, the installation directory chosen should be the same as that chosen for the DCW; the two products in either order. For normal first-time installation, choose the Complete installation option. If you need to restore a fresh copy of a part of the data at some later time, run setup.exe and select Custom installation. This allows you to select individual feature classes to be reinstalled.

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.

Release History

Notes:

- For changes in current release, see [What's New in Version 16](#)
- World Language Mapping System version numbers prior to version 16 correspond to the version numbers of the GMMS DCW Data Supplement from which they are derived. Changes in major and minor version levels did not reflect the extent of changes in the WLMS.

Changes from 3.2 to 3.2.1

- Language areas have been added for English, Spanish, and French in the U.S. and Canada
- Languages polygons have been moved from Other Language Areas (rother.shp) to the main Language Areas (langa.shp) where the *Ethnologue database* contains a language record (usually a record for an immigrant group, designated by field IM = Y). Previously, languages were only included in langa.shp where the *Ethnologue book* (which excludes immigrant languages) included an entry for the language group.
- Country and language boundaries, with corresponding data attributes, have been adjusted to reflect the split of Serbia and Montenegro planned for January, 2007.
- A number of other minor corrections have been made to the data.

Changes from 3.1 to 3.2

- Point, polygon, and attribute data have been updated to correspond to the 15th Edition of the *Ethnologue*.
- Unified point shapefile allows the use of classification and/or definition queries to display various classes of points (polygon centroids, widespread language points, extinct language points, etc) formerly represented only as separate shapefile layers. The separate shapefile layers are retained for back compatibility.
- Additional attribute data in shapefiles requires less frequent joining of the *Ethnologue* attribute table for many applications.
- 44 additional fields of information in the *Ethnologue* data table reflect the additional scope of the published 15th Edition *Ethnologue*, an improved breakdown of data into separate fields, and improvements (described below) in geographic coding, language coding, name rendering, and population.
- The *Ethnologue* attribute table now has decimal-degree latitude and longitude fields added, allowing it to be mapped directly as an XY or event table.
- Languages are coded using both the [ISO DIS 639-3](#)-compliant codes used in the 15th Edition *Ethnologue* and a legacy code compatible with earlier versions of the *Ethnologue*. For clarity, legacy codes are given in the traditional upper case, while ISO 639-3 codes are given in lower case. Several codes for language-in-country are provided:
 - Legacy *Ethnologue* codes with upper-case ISO three-letter country codes (the previous WLMS standard for language-in-country coding).
 - ISO 639-3 codes with lower-case ISO two-letter country codes (the recommended new standard for ISO-compliant language-in-country coding).
 - ISO 639-3 codes with upper-case U.S. government FIPS 10-4 country codes (the recommended standard for those already using the widely-used FIPS standard for country coding, and for those doing historical research, since FIPS codes represent consistent geographic entities over time).

- Name rendering for map labels is substantially improved, with traditional upper-case ASCII (diacritic-stripped) names supplemented by mixed-case ASCII, mixed-case ANSI (ISO Latin 1) diacritical forms, and comma-free ANSI diacritical forms.
- Populations of most languages are now provided as a numeric field in addition to the earlier text discussion of population research.
- The Widespread Languages, Offset for Display point coverage, which allows ranged-size point symbols placed in ocean or sparsely populated land areas to substitute for polygon symbology, has been substantially enhanced.
- Additional and updated sample maps illustrate a number of possibilities for using new population, widespread language, and name data.

Changes from 3.01 to 3.1

- Adjustments of some language borders to align with DCW country borders (one or two languages of Russia, Oman, Saudi Arabia, Ethiopia, Djibouti, Sudan and some islands which were associated with wrong country).

Sample Map Projects

Sample map projects in ArcGIS project format (.mxd) are included in the `<installdir>\projects\Samples-lang` directory to illustrate use of the various language layers. Some of these projects use the ECW compressed topography included in the WLMS installation. Display of this data in ArcMap requires a free extension available from <http://www.ermapper.com/downloads>; for more information on the ECW format, see <http://www.ermapper.com/ecw>.

- Africa Language Families.mxd – Language families of Africa, widespread languages shown with symbols.
- Kenya and Tanz Language.mxd – All languages of Kenya and Tanzania classified by language family. Languages labeled.
- Languages in Southern Africa.mxd – Language family map zoomed in to southern Africa region.
- Languages of Papua New Guinea.mxd – Papua New Guinea languages randomly colored.
- Myanmar and Thailand.mxd – All languages of Myanmar and Thailand, includes mixed language areas.
- Myanmar-Thailand-1_1mil_langonly.mxd – Some Myanmar and Thailand languages. Languages labeled.
- SE Papua New Guinea.mxd – Dot density map showing number of speakers of particular language in southeast Papua New Guinea region.
- SE Papua New Guinea zoom to 1_5mill_langonly.mxd – Papua New Guinea map zoomed in to southeast.
- Widespread_Population.mxd – World map showing all languages, including widespread “offshore” languages.
- World Languages Point.mxd – Language points, classified by type.

Additional projects illustrate the use of the data in combination with GMI's Seamless Digital Chart of the World product are included in `<installdir>\projects\Samples-dcwstring`.

- Kurds.mxd – Locations of four Kurdish groups.
- Myanmar-Thailand-1_1mil_dcw.mxd – Some Myanmar and Thailand languages including additional DCW data for geographic context.

- SE Papua New Guinea zoom to 1_5mill dcw.mxd – Papua New Guinea map zoomed in to southeast including additional DCW data for geographic context.

Data Organization

When uncompressed during installation, data will be placed in these subdirectories under a specified installation directory (<installdir>, by default c:\gmi_geodata). Click links below for detailed documentation of data in each subdirectory.

- <installdir>\bnd – Boundary: Country boundaries
- <installdir>\lang – Language points, polygons, and attributes
- <installdir>\doc – Documentation files (including this document)
- <installdir>\topogrid – Raster topography grids which display elevation and hillshading.
- <installdir>\projects -- Sample project files [documented above](#)

Detailed Data Documentation

Note: use Acrobat’s Bookmarks pane or Find (Ctrl-F) features to find specific data in this section. Titles indicate point, line, or area (polygon) features.

Source: GMI and NIMA VMap0, Edition 5

Boundary

Country area features

bnd\DCW_countries.shp

Abstract: This shapefile contains countries of the world and their names. It was derived by dissolving the updated first-level administrative divisions of GMI’s DCW base map into countries. **Notes:** Some data fields from the VPF form of polbnda have been changed to more intuitive names:

- NA2 changed to FIPS10_4
- NA2_DESCRI changed to COUNTRY
- NA3 changed to CONTCODE
- NA3_DESCRI changed to CONT

Attributes:

Field	Field Name	Type	Width	Description
1	ID	Character	16	Unique identifier
2	FIPS10_4	Character	2	FIPS 10-4 code
3	COUNTRY	Character	44	Country name
4	CONTCODE	Character	1	Continent/Region code
5	CONT	Character	15	Continent/Region name
6	F_CODE	Character	5	Feature code
7	F_CODE_DES	Character	19	Feature code description
8	OW_ABBREV	Character	4	Patrick Johnstone, Operation World, Country Code
9	ISO_A2	Character	4	ISO 3166 2 letter country code
10	ISO_A3	Character	3	ISO 3166 3 letter country code
11	VMAP_ID	Character	15	Unique ID
12	FACTBK_CTY	Character	44	Country name as given in World Factbook

Used in Projects: Africa Language Families.mxd, Kenya and Tanz Language.mxd, Kurds.mxd, Languages in Southern Africa.mxd, Languages of Papua New Guinea.mxd, Myanmar and Thailand.mxd, Myanmar-Thailand-1_1mil_langonly.mxd, SE Papua New Guinea.mxd, SE Papua New Guinea zoom to 1_5mill_langonly.mxd, Widespread_Population.mxd, World Languages Point.mxd

Table Joins:

Source: World Language Mapping System

Language Areas

Language area features

lang\langua.shp

Abstract: Languages areas of the world.

Attributes:

Field	Field Name	Type	Width	Description
1	ID	Character	7	Old style LMP identifier- Legacy Ethnologue with uppercase ISO A3 country code
2	ID_ISO_A3	Character	7	Unique identifier with ISO language code and ISO A3 country code
3	ID_ISO_A2	Character	7	Unique identifier with ISO language code and ISO A2 country code (preferred for most applications)
4	ID_FIPS	Character	7	Unique identifier with ISO language code and FIPS 10-4 country code (preferred for historical studies due to code consistency over time)
5	NAM_LABEL	Character	40	Name in mixed-case, ANSI (ISO Latin 1) character set, inverted to move modifiers to beginning and remove commas (preferred for most map labeling purposes)
6	NAME_PROP	Character	40	Name in mixed-case, diacritic-stripped ASCII
7	NAME2	Character	40	Name in upper-case ASCII
8	NAM_ANSI	Character	40	Name in mixed-case, ANSI (ISO Latin 1) character set, inverted to move modifiers to beginning and remove commas (preferred for most map labeling purposes)
9	CNT	Character	8	Continent
10	C1	Character	33	Country Name
11	POP	Character	86	Population (narrative text)
12	LMP_POP1	Numeric	11	Population (numeric value extracted from POP where available)
13	G	Character	225	Language genetic tree
14	LMP_CLASS	Character	5	Language mapping class: A=Point mapped, no polygon L=Polygon mapped; point location is polygon centroid U=Unknown location; point location based on REG field in Ethnologue or arbitrary W=Widespread language (national or regional language of wider communication) X=Extinct language blank or null: unclassified (generally immigrant languages not in Ethnologue)

15	FAMILYPROP	Character	30	Language family; first element of G field
16	FAMILY	Character	30	Language family; first element of G field, in upper case for compatibility with family classification in older maps
17	LMP_C1	Character	32	Country Name used by WLMS (created for cases like Kosovo where Ethnologue doesn't have entry for country)

Used in Projects: Africa Language Families.mxd, Kenya and Tanz Language.mxd, Kurds.mxd, Languages in Southern Africa.mxd, Languages of Papua New Guinea.mxd, Myanmar and Thailand.mxd, Myanmar-Thailand-1_1mil_langonly.mxd, SE Papua New Guinea.mxd, SE Papua New Guinea zoom to 1_5mill_langonly.mxd, Widespread_Population.mxd.

Special layer files: lang family poly.lyr, lang family - Ruhlen (poly).lyr, lang population (poly).lyr

Table Joins: eth_wlms.dbf, SIL2Ruhlen.dbf

Mixed Language area features

lang\mixlanga.shp

Abstract: Areas where languages overlap in langa.shp.

Notes: The Unique identifier field is created by two (if there are two language areas overlapping) or three (if there are three or more language areas overlapping) groups consisting of a three letter *Ethnologue* language code and a hyphen. The three letter ISO 3166 code for the country is after the final hyphen.

In the NAM_LABEL and NAME_PROP fields, the various language names are separated by a semi-colon (;). When necessary, some language names have been abbreviated (or, as a last resort, omitted) for polygons with several long language names.

Attributes:

Field	Field Name	Type	Width	Description
1	ID	Character	32	Unique identifier; hyphen-separated legacy Ethnologue language codes followed by ISO A3 country code.
2	ID_ISO_A3	Character	32	As above, with ISO language codes.
3	ID_ISO_A2	Character	31	As above, with ISO language codes and ISO A3 country code.
4	ID_FIPS	Character	31	As above, with FIPS 10-4 country code
5	NAM_LABEL	Character	254	Names of languages, semicolon separated, in mixed-case, ANSI (ISO Latin 1) character set, inverted to move modifiers to beginning and remove commas (preferred for most map labeling purposes)
6	NAME_PROP	Character	254	Names of languages, semicolon separated, in mixed-case, diacritic-stripped ASCII
7	CNT	Character	8	Continent
8	C1	Character	33	Country Name

Used in Projects: Myanmar and Thailand.mxd, Myanmar-Thailand-1_1mil_langonly.mxd

Table Joins:

Other Region Language area features

lang\rother.shp

Abstract: Language areas where information developed by the Language Mapping Project indicates a language has a homeland area in a particular country, but the language is not listed in the Ethnologue for that country. Many of these areas are extensions of a linguistic homeland for the same language in an adjoining country; these are usually listed in the Ethnologue. There are relatively few such areas in the world.

Notes: The Unique identifier field is created by the three letter *Ethnologue* language code, a hyphen, and then the three letter ISO 3166 code for the country.

Attributes:

Field	Field Name	Type	Width	Description
1	ID	Character	7	Old style LMP identifier- Legacy Ethnologue with uppercase ISO A3 country code
2	ID_ISO_A3	Character	7	Unique identifier with ISO language code and ISO A3 country code
3	ID_ISO_A2	Character	7	Unique identifier with ISO language code and ISO A2 country code (preferred for most applications)
4	ID_FIPS	Character	7	Unique identifier with ISO language code and FIPS 10-4 country code (preferred for historical studies due to code consistency over time)
5	NAM_LABEL	Character	40	Name in mixed-case, ANSI (ISO Latin 1) character set, inverted to move modifiers to beginning and remove commas (preferred for most map labeling purposes)
6	NAME_PROP	Character	40	Name in mixed-case, diacritic-stripped ASCII
7	NAM_ANSI	Character	40	Name in mixed-case, ANSI (ISO Latin 1) character set, inverted to move modifiers to beginning and remove commas (preferred for most map labeling purposes)
8	CNT	Character	8	Continent
9	C1	Character	33	Country Name
10	POP	Character	82	Population (narrative text)
11	LMP_POP1	Numeric	11	Population (numeric value extracted from POP where available)
12	G	Character	225	Language genetic tree
13	LMP_CLASS	Character	5	Language mapping class: A=Point mapped, no polygon L=Polygon mapped; point location is polygon centroid U=Unknown location; point location based on REG field in Ethnologue or arbitrary W=Widespread language (national or regional language of wider communication) X=Extinct language blank or null: unclassified (generally immigrant languages not in Ethnologue)

14	FAMILY	Character	33	Language family; first element of G field, in upper case for compatibility with family classification in older maps
15	FAMILYPROP	Character	30	Language family; first element of G field

Used in Projects: Myanmar-Thailand-1_1mil_langonly.mxd
Table Joins: eth_wlms.dbf

Language Points

Standard Language Point Attributes

Notes: Unless otherwise indicated below, all language point shapefiles have the following attributes:

Attributes:

Field	Field Name	Type	Width	Description
1	ID	Character	7	Old style LMP identifier- Legacy Ethnologue with uppercase ISO A3 country code
2	ID_ISO_A3	Character	7	Unique identifier with ISO language code and ISO A3 country code
3	ID_ISO_A2	Character	7	Unique identifier with ISO language code and ISO A2 country code (preferred for most applications)
4	ID_FIPS	Character	7	Unique identifier with ISO language code and FIPS 10-4 country code (preferred for historical studies due to code consistency over time)
5	NAM_LABEL	Character	40	Name in mixed-case, ANSI (ISO Latin 1) character set, inverted to move modifiers to beginning and remove commas (preferred for most map labeling purposes)
6	NAME_PROP	Character	40	Name in mixed-case, diacritic-stripped ASCII
7	NAME2	Character	40	Name in upper-case ASCII
8	NAM_ANSI	Character	40	Name in mixed-case, ANSI (ISO Latin 1) character set, inverted to move modifiers to beginning and remove commas (preferred for most map labeling purposes)
9	CNT	Character	8	Continent
10	C1	Character	33	Country Name
11	POP	Character	86	Population (narrative text)
12	LMP_POP1	Numeric	11	Population (numeric value extracted from POP where available)
13	G	Character	225	Language genetic tree
14	IM	Character	1	Immigrant language, not included in printed Ethnologue if Y
15	LMP_CLASS	Character	5	Language mapping class: A=Point mapped, no polygon L=Polygon mapped; point location is polygon centroid U=Unknown location; point location based on REG field in Ethnologue or arbitrary W=Widespread language (national or regional language of wider

ID	Field Name	Field Type	Field Length	Description
16	LMP_SUBCL	Character	6	Language mapping subclass, indicating point reliability; available for some class U, W, X, and A points: N=No reliable location; point placed arbitrarily in corresponding country R=Regional reliability; point placed in general region suggested by REG in Ethologue M=Most reliable location; sufficient data for point, but not for polygon
17	LMP_LON	Numeric	19.5	Longitude, decimal degrees
18	LMP_LAT	Numeric	19.5	Latitude, decimal degrees
19	LMP_C1	Character	32	Country Name used by WLMS (created for cases like Kosovo where Ethologue doesn't have entry for country)

Used in Projects: World Languages Point.mxd
Table Joins: eth_wlms.dbf

Ethologue table as XY event table

Note: Fields LMP_LON and LMP_LAT with decimal degree coordinates have been added to the [Ethologue Attributes table \(lang\eth_wlms.dbf\)](#), allowing it to be used as a language point layer in GIS systems capable of displaying tables containing coordinates. These are known as XY tables or Event Tables in ArcView.

All Language Point Features

lang\lang_all_pt.shp

Abstract: Point features for all published Ethologue languages, including widespread, extinct, and unknown location languages. Where necessary, points have been placed arbitrarily within the correct country to create a complete set.

Notes:

Attributes: See [Standard Language Point Attributes](#)

Used in Projects: World Languages - Points.mxd

Special layer files: lang family poly.lyr, lang family - Ruhlen (poly).lyr, lang population (poly).lyr

Table Joins: eth_wlms.dbf, SIL2Ruhlen.dbf

Language point features

lang\lang_points_pt.shp

Abstract: Points within the language areas (langa.shp).

Attributes: See [Standard Language Point Attributes](#)

Table Joins: eth_wlms.dbf

Extinct Language point features

lang\lang_extinct_pt.shp

Abstract: Points representing extinct languages. Located where possible at the center of the homeland of the last known population of speakers of the language in a country.

Notes: File included for back compatibility with older maps; for many new applications, [All Language Point Features](#) or [Ethnologue table as XY event table](#) with appropriate classification and definition queries will be easier to set up and produce more understandable legends

Attributes: See [Standard Language Point Attributes](#)

Table Joins: eth_wlms.dbf

Widespread Language point features

lang\lang_widesprd_pt.shp

Abstract: National or regional trade languages which cannot be properly mapped as polygons since they overlap a large number of other language areas are represented by these points, which are placed arbitrarily within the correct country or region of a country.

Notes: File included for back compatibility with older maps; for many new applications, [All Language Point Features](#) or [Ethnologue table as XY event table](#) with appropriate classification and definition queries will be easier to set up and produce more understandable legends

See also: [Widespread Languages, Offset for Display point features](#)

Attributes: See [Standard Language Point Attributes](#)

Used in Projects: World Languages - Points.mxd

Table Joins: eth_wlms.dbf

Widespread Languages, Offset for Display point features

lang\lang_widesprd_offset.shp

Abstract: This shapefile contains the same languages as lang_widesprd_pt.shp, but points have been moved either to ocean areas off the coast of the corresponding country or to land areas where a reasonable-sized point symbol will not obscure any languages. This allows the points to be displayed as relatively large symbols, which may be filled with colors or patterns equivalent to those used for language areas in thematic maps. Africa Language Families.mxd and lang_widesprd_population.lyr file illustrate use of these points to create symbols colored by one thematic variable (language family) and sized by another (population).

Attributes: See [Standard Language Point Attributes](#)

Table Joins: eth_wlms.dbf

Other Language point features

lang\lang_other_pt.shp

Abstract: Points of the [Other Region Language area features \(rother.shp\)](#)

Notes: File included for back compatibility with older maps; for many new applications, [All Language Point Features](#) or [Ethnologue table as XY event table](#) with appropriate classification and definition queries will be easier to set up and produce more understandable legends

Attributes: See [Standard Language Point Attributes](#)

Table Joins: eth_wlms.dbf

Point-only point features

lang\lang_only_pt.shp

Abstract: Points of languages lacking an associated polygon.

Notes: File included for back compatibility with older maps; for many new applications, [All Language Point Features](#) or [Ethnologue table as XY event table](#) with appropriate classification and definition queries will be easier to set up and produce more understandable legends

Attributes: See [Standard Language Point Attributes](#)

Table Joins: eth_wlms.dbf

Unknown Location point features

lang\lang_unkwn_loc_pt.shp

Abstract: These points represent Ethnologue languages for which insufficient data was available to create a language polygon in langa.shp. Where some textual description of the

language location was given in the Ethnologue, the point was located in the approximate center of the area described. Where the text description is absent or unintelligible, the point is located arbitrarily within the country.

Notes: File included for back compatibility with older maps; for many new applications, [All Language Point Features](#) or [Ethnologue table as XY event table](#) with appropriate classification and definition queries will be easier to set up and produce more understandable legends

Attributes: See [Standard Language Point Attributes](#)

Used in Projects: World Languages - Points.mxd

Table Joins: eth_wlms.dbf

Source: SIL

Ethnologue Language attribute data

Ethnologue Attributes

lang\eth_wlms.dbf

2005

Abstract: Language attributes from the Ethnologue database used to create *Ethnologue: Languages of the World, 15th Edition*.

Notes:

- The Ethnologue database is a text database maintained by the SIL International (SIL) for use in publishing the *Ethnologue* series of books. A portion of this database, converted to a .DBF attribute table by GMI, has been provided as part of the World Language Mapping System.
- The Ethnologue database includes entries for a substantial number of instances of language-in-country for immigrant languages; these are not included in the printed Ethnologue. These additional records have attribute IM = "Y". These records are not intended to be comprehensive, are not maintained to the same standards as other Ethnologue records, and are in most cases not mapped as points or polygons.
- In its original form, the Ethnologue database has no maximum size for text fields. \LANGA\ETH_WLMS.DBF was converted from the Ethnologue database using as field sizes either the length of the longest data element in the field or the 254-byte maximum allowed in ArcGIS-compatible .DBF files. In the latter case, some data has necessarily been truncated.
- Additional coded identifiers, alternate name forms, and derived fields (such as LMP_POP1 and FAMILY) have been added to facilitate mapping.
- For more information on the *Ethnologue*, see <http://www.ethnologue.com/>. To suggest updates or corrections to the *Ethnologue*, see: http://www.ethnologue.com/ethno_docs/feedback.asp

Attributes:

Field	Field Name	Type	Width	Description
1	ID	Character	7	Old style LMP identifier- Legacy Ethnologue with uppercase ISO A3 country code
2	ID_ISO_A3	Character	7	Unique identifier with ISO language code and ISO A3 country code
3	ID_ISO_A2	Character	7	Unique identifier with ISO language code and ISO A2 country code (preferred for most applications)
4	ID_FIPS	Character	7	Unique identifier with ISO language code and FIPS 10-4 country code (preferred for historical studies due to code consistency over time)

5	NAM	Character	40	Name without diacritics in mixed-case ASCII
6	NAM_ANSI	Character	40	Name with diacritics in mixed-case ANSI (ISO Latin 1) character set.
7	NAM_UC	Character	40	Name without diacritics upper-case ASCII
8	NAM_LABEL	Character	40	Name in mixed-case, ANSI (ISO Latin 1) character set, inverted to move modifiers to beginning and remove commas (preferred for most map labeling purposes)
9	XXX	Character	3	Legacy Ethnologue language code
10	ISO	Character	3	ISO language Code
11	C1	Character	32	Country Name
12	CNT	Character	8	Continent
13	Z	Character	25	Zone within country (for certain countries, esp Indonesia and Malaysia)
14	C	Character	254	Other countries where language is spoken
15	NAL	Character	253	Alternative names for language
16	IM	Character	1	Immigrant language, not included in printed Ethnologue if Y
17	D	Character	254	Dialect names (alternate names in parentheses)
18	FAMILYPROP	Character	30	Language family; first element of G field
19	FAMILY	Character	30	Language family; first element of G field, in upper case for compatibility with family classification in older maps
20	G	Character	225	Language genetic tree
21	HUB	Character	32	Country of origin or where most speakers are
22	REG	Character	254	Location of language in country
23	PTO	Character	22	Population total in all countries. ('or more' indicates some estimate(s) missing)
24	POP	Character	82	Number of speakers of language in country (narrative text, preferred numeric value at beginning). POPU in earlier Ethnologue database versions. See also numeric population in LMP_POP1.
25	PPE	Character	254	Population of the ethnic group
26	PPX	Character	254	Extra population data
27	BL	Character	150	Bilingual proficiency levels
28	DOM	Character	254	Domains of use in L1
29	USE	Character	254	L1 (Comments on vitality)
30	ATT	Character	254	Attitude. (MT or other languages)
31	BLR	Character	254	Language and identity Bilingual remarks. (Domains for L2). Functionally replaces PIN in earlier versions.
32	AGE	Character	254	Age range of speakers
33	BB	Character	10	Dates the Bible was published (First-Latest)
34	NT	Character	13	Dates the NT was published (First-Latest)

35	TCP	Character	1	Extinction Status: X=Extinct U=Nearly Extinct Z=Second language only
36	LXS	Character	254	Lexical Similarity remarks (Usually %)
37	INR	Character	254	Intelligibility dialects closeness remarks. Includes remarks about dialects and genetic relationships.
38	WR	Character	158	Script and orthography (May be more than one, or one is obsolete)
39	NAT	Character	38	Status of this language as national or official language in country
40	TRA	Character	189	Trade Language (Y)
41	WID	Character	1	Language of Wider Communication (LWC). (Y) (Without official recognition)
42	GOV	Character	84	Official languages (vs. national language) (Language entry) (Y)
43	ELE	Character	211	Elementary Education in L1.
44	SEC	Character	146	Secondary Education in L1 (Instruction or subject) (Y)
45	LIR	Character	117	Literacy rate in L2 (figures from Lit office)
46	LIV	Character	71	Literacy rate in L1
47	TG	Character	254	Grammar Typology (see questionnaire)
48	ADI	Character	254	Dictionaries (Y) Give ref. if available.
49	AGR	Character	254	Grammars available. "Grammar" Give reference if available.
50	SBT	Character	254	Subsistence type
51	ECT	Character	254	Ecosystem, type of land
52	GET	Character	71	Geologic type
53	ALR	Character	109	Altitude range. Printed in feet or meters.
54	REL	Character	121	Religions
55	MT	Character	133	Television in L1.
56	MD	Character	144	Radio broadcasts in L1.
57	MN	Character	56	Newspapers in L1.
58	MF	Character	66	Movies or films in L1.
59	MM	Character	89	Magazines in L1. (Y)
60	LMP_CLASS	Character	5	Language mapping class: A=Point mapped, no polygon L=Polygon mapped; point location is polygon centroid U=Unknown location; point location based on REG field in Ethnologue or arbitrary W=Widespread language (national or regional language of wider communication) X=Extinct language blank or null: unclassified (generally immigrant languages not in Ethnologue)

61	LMP_SUBCL	Character	6	Language mapping subclass, indicating point reliability; available for some class U, W, X, and A points: N=No reliable location; point placed arbitrarily in corresponding country R=Regional reliability; point placed in general region suggested by REG in Ethnologue M=Most reliable location; sufficient data for point, but not for polygon
62	LMP_LAT	Numeric	19.5	Latitude, decimal degrees
63	LMP_LON	Numeric	19.5	Longitude, decimal degrees
64	LMP_POP1	Numeric	11	Population (numeric value extracted from POP where available)
65	LMP_POP2	Numeric	11	Population preferring LMP_POP1 but adding IMB 1996 data where LMP_POP1 is not available

Additional cautions:

1. The original Ethnologue database is a structured-text database developed to support the publication of the Ethnologue book. Full proofreading and consistency checks on the data are only done just before the book is published, so releases of the data such as the one used here may have minor inconsistencies with the published book. The information in the Ethnologue is designed for human readers of a printed book, rather than for the kind of statistical, analytical study a structured table would seem to imply. In particular, sources and other annotations of the data tend to be mingled with numeric data, and numeric information tends to be formatted for human rather than machine reading.
2. The tagged database structure used for the Ethnologue database makes it very easy and non-costly in storage space to add tags (fields), which are only infrequently used. It is unwise to assume that because a tag has been assigned that data for that tag is available for all languages for which it might be relevant. The tags included here correspond to those used in publishing the Ethnologue book. The tagged structure also sets no fixed width for fields. In converting data to the fixed-field structure, we have attempted to set field widths which truncate the data for no more than 1% of entries, or to the maximum 255 character field width of a .dbf file.

Source: GMI/Dr. Merritt Ruhlen

Conversion for Language Families

Conversion between SIL and Ruhlen Families

lang\SIL2Ruhlen.dbf

2010

Abstract: This is a simple translation table developed by GMI providing an approximate equivalency between the highest level of the *Ethnologue* genetic tree and the smaller number of top-level classifications in Dr. Merritt Ruhlen's *A Guide to the World's Languages*. There are a small number of cases where this simple translation between top-level categories fails to produce the correct Ruhlen classification, most notably Korean and Japanese.

Attributes:

Field	Field Name	Type	Width	Description
1	FAMILY	Character	21	SIL main language family (upper case)

Source: GLOBE/GMI**Topography Grids***Global Elevation/Bathymetry Colors compressed raster***TopoGrid\elelevation.ecw****2003**

Abstract: Elevation.ecw is a highly-compressed, high-resolution raster image applying conventional elevation colors (shades of blue/cyan for ocean depths, dark green at low elevations, browns at mid elevations, and gray values lightening to white at high elevations).

Notes:

A. Two Digital Elevation Models were used:

1. Land elevations were produced from GMI's WorldDTM90 DEM. This was created as follows:
 - a. We started with the GTOPO30 30 arc-second DEM (<http://edcdaac.usgs.gov/gtopo30/gtopo30.html>) as assembled for latitudes above 60 degrees south and converted to ESRI GRID format by Chris Deckert of Campus Crusade's Jesus Film project.
 - b. Additional elevation data for latitudes south of 60S were added from the GLOBE Project DEM (<http://www.ngdc.noaa.gov/seg/topo/globe.shtml>).
 - c. The Caspian Sea was converted to an ocean area by converting all cells at -28 meters elevation to NoData values. This is known to have added a few spurious NoData cells in land areas elsewhere having elevations of -28 meters, notably in areas north of the Caspian, near the Dead Sea, and in Death Valley, California.
2. Ocean depths (bathymetry) were produced with data from the ETOPO2 2-minute DEM (<http://www.ngdc.noaa.gov/mgg/fliers/01mgg04.html>).

Both DEMs were converted to ERMapper binary format, and colors from the ERMapper "elevation" color ramp were applied separately to land elevation and bathymetric DEMs using an ERMapper algorithm file. The color ramps are using several linear segments to increase color distinctions at low land elevations and ocean depths, and reduce distinctions at very high elevations and extreme depths. The resulting ERMapper algorithm and binary data were compressed at full resolution and 20:1 target compression using ERMapper 6.0 by Dawn Wise at San Diego State University.

Attributes: RGB color components

Used in Projects: Kurds.mxd

*Global Hillshades compressed raster***TopoGrid\hillshades.ecw****2003**

Abstract: hillshades.ecw is a highly compressed, high-resolution grayscale hillshade raster image. The image was calculated using the hillshade algorithm of ESRI ArcView spatial analyst applied to GMI's worldldtm90 grid dataset (derivation described under elevation.ecw above).

Notes:

- A. Hillshades were calculated at a sun elevation of 60 degrees and vertical scale factor of .00005, chosen to avoid black saturation in all but the steepest mountainsides. The resulting grayscale grid was converted to ERMapper binary format and compressed using ERMapper 6.0 by Dawn Wise at San Diego State University.
- B. The hillshade settings chosen for this image give good terrain definition in mountainous areas, but do not clearly distinguish terrain in areas of lower relief. This may be improved somewhat with the existing image by adjusting the maximum and minimum values of the

color ramp. Greatly improved rendering of low relief requires recalculation with lower sun angles and higher vertical scale factors.

Attributes: RGB color components

Used in Projects: Kenya and Tanz Language.mxd, Kurds.mxd, Languages in South Africa.mxd, Languages of Papua New Guinea.mxd, SE Papua New Guinea.mxd, SE Papua New Guinea zoom to 1_5mill.mxd

Global Elevation/Bathymetry Colors+Hillshades compressed raster

TopoGrid\lev-hs.ecw

2003

Abstract: This file is the hillshades.ecw file overlaying the elevation.ecw file at 70% transparency for use as a background in ArcView 3.x or other systems not supporting layer transparency.