Supplementary Material (2) for Xiaodong Qin et al.:

GPlates–Shapefile interoperability

Contact: Christian.Heine@sydney.edu.au

Revision: 1792 Author: christian Date: 2012-05-11

This document proposes the use of a common set of attributes to be included with shapefiles for plate tectonic applications. The agreement on a community standard of attributes attached to shapefiles containing plate tectonic data should facilitate the exchange of files and allow for a basic tracking of modifications without the need of a complex centralised versioning system.

Contents

Introduction

Metadata attached to geospatial information describes important information about the origins, quality and modifications. In the age of digital data distribution, these metadata become increasingly important to track the origins of data and their quality for both research and industry use. In the field of plate tectonics, the community has been using a file format which is largely based on the legacy PLATES data format. This format, however, was developed in the 1980's, uses simple ASCII text files and has severe limitations in terms of interoperability, metadata containment and metadata standardisation. With the advent of the open-source GPlates application and other, partly proprietary, tools for plate tectonic data visualisation, a large part of the community acknowledged the importance of metadata attached to plate tectonic models and data. Furthermore, ESRI shape files (*.shp) have evolved into a a de-facto standard for data exchange across a wide range of proprietary and open source tools used in industry and academic software packages — from seismic interpretation packages to general GIS. To facilitate the integration of data across various disciplines and tools, and for integrated geodynamic modelling a standardised set of plate tectonic attributes attached to shape file data is proposed.

Contact Christian Heine (`mailto:christian.heine@sydney.edu.au`) for questions, comments or change proposals.

Current deficiencies

The shapefile format has severe limitations due to its dBase format which does not allow attribute columns longer than 10 characters, no topological information. Additionally it is a proprietary format. Attribute names in shapefiles are *silently* truncated to 10 characters. For a futher discussion on the capabilities and limitations of the shapefile format, please refer to the ESRI Shapefile Technical Description. More information on the dBase format can be found at `www.inprise.com`. The OSGEO pages also contain some technical background info on shapefiles `http://www.gdal.org/ogr/drv_shapefile.html` and here `http://trac.osgeo.org/gdal/wiki/UserDocs/Shapefiles`.

Aims of this standardisation

This standardisation aims to unify the naming of shapefile attributes for plate tectonic applications in order to facilitate data exchange, preventing the loss of important metadata, and help to trace changes to features. At time of writing, most open and commercially available applications rely on the old PLATES data type format, which needs a PlateID, FromAge and ToAge associated with each geometry in order to reconstruct the data through time.

The legacy PLATES format as specified in the PLATES database manual is:

```
92 1    3 NORTH & CENTRAL ATLANTIC RIDGE AXIS
101 0.1 -999.0 RI    1 101    1    8
37.4118   37.1394 3
37.1760   37.0349 2
37.0643   36.9175 2
99.0000   99.0000 3
```

In this model, which has over the time being adopted to various other applications, the first line of the header contains the following information:

92 North Atlantic region

1 reference number for the data

3 String number

North... Geographic description

The second header line contains the following information:

101 Moving plate ID

0.1 Age of appearance (FromAge)

-999.0 Age of disappearance (ToAge)

RI Data type code

1 Data type code number

101 Reference plate id

1 Colouring

8 number of location points/nodes in string

Subsequently, shapefiles have largely replaced fixed-format plain text files in geospatial applications and plate tectonic software until recently, when the GPlates project invented a XML-based GML derivative dubbed the "GPlates Markup Language (GPML)" and an associated information model (GPGIM). The goal of this document is to lay out a community standard for attributes in shapefiles which facilitates the exchange of information without the loss of important metadata.

This paper lays out a suggestion for adopting the current Shapefile format, developed by ESRI as an option for representing geospatial data for plate tectonic applications. The reasoning behind this is that ESRI shapefiles are a current de-facto standard across a wide range of institutions and companies. However, it should be explicitly mentioned that Shapefiles do not represent an adequate container and are no replacement for the advanced data infrastructure used in the GPlates Information Model and associated GPML files. The intent of this specification is rather to present a workaround for people using Desktop GIS systems where Shapefiles are a de-facto standard for data exchange.

Currenty used file formats

At time of writing, at least three different file formats are being used in plate tectonic applications. These are the new, GML-based GPlates Markup Language (GPML) with the associated GPlates Geological Information Model (GPGIM). Additionally, ASCII-text based files in the legacy PLATES format and ESRI shapefiles. The latter are usually attributed with metadata from the PLATES legacy format.

GPlates standard shapefile attributes

A technical description of the shapefile capabilities is listed in the ESRI whitepaper on shapefiles. However, this slightly arcane format limits the length of attribute names to a maximum length of 10 characters and the maximum length of characters in a given cell to 254 characters.

A minimum set of attributes for each feature carrying the crucial information to reconstruct geometries back through time is proposed here. One goal is to make the loading of such “standardised shapefiles” in GPlates automatic if they carry attributes as shown in Fig. 1. Out of the attributes listed in Fig. 1, only the *FeatureID* and *Conjugate PlateID* are optional.

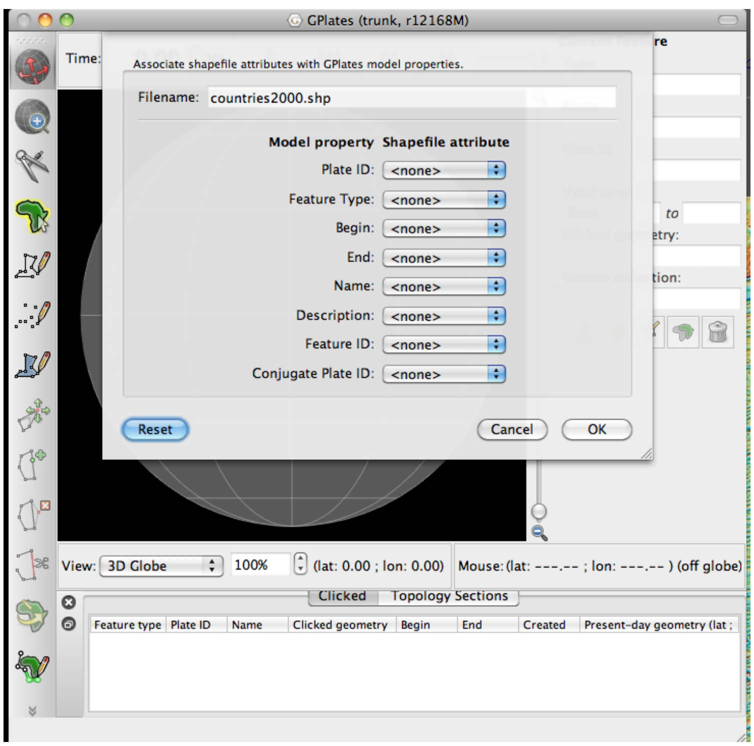

Figure 1: Shapefile loading dialogue in GPlates.

Standardised minimum set of attributes

Explanation of indvidual attribute names

`PLATEID` *Core attribute* – PlateID column in shapefile. If the column is not filled with values, the GPlates cookie-cutting routing will populate it.

Attribute name 1234567890	Type;length	Description/Comment
PLATEID1..	long	Unique plate identifier
FROMAGE...	float;	Age of appearance
TOAGE.....	float;	Age of disappearance
FEAT_TYPE	text; 50	Follow GPlates GPGIM or PLATES data type codes (See)
FEAT_NAME	text; 255	Name of the feature
FEAT_DESCR	text; 254	Verbose description
DOI_REF...	text; 254	Unique DOI or verbose reference
MOD_AUTHOR.	text;254	Modification details
MOD_DATE..	date	Modification date
COMMENT...	text; 254	Comments on feature
STDVERSION	text; 20	Version of the GPlates tectonic shapefile attribute standard

Table 1: Basic set of shapefile attributes which are automatically recognised in GPlates

`FROMAGE` *Core attribute* – Age of appearance of the feature in geological history.

`TOAGE` *Core attribute* – Age of disappearance of the feature in geological history.Commonly, using the PLATES legacy format, this is -999.0 for the distant future and any number for a numerical geological age.

`FEAT_TYPE` The feature type should adhere to the GPlates Geological Information Model (GPGIM), or the PLATES project data types. The current set of hardcoded FeatureTypes in GPlates is listed in Table 3. The `FEATURETYP` list will likely undergo changes which will be incorporated in this standard in subsequent updates. Example: `SM: Seamount`

`FEAT_NAME` Column to hold the name of a given feature. It should be populated using common names from literature or as used in the community. Example: `Ninety East Ridge Seamounts`

`FEAT_DESCR` Column for a verbose description of individual features. The `FeatureDescr` attribute should capture an understandable description using geological and geophysical terminology, if possible referring to datasets where the interpretation has been based upon. Example: `500 m depth contour line along Australian North West Shelf based on SRTMplus bathymetry model, 2009`

`DOI_REF` This attribute should provide storage for references. These

CHHEI_ShapefileAttributeStandard.tex 1792 2012-05-11 14:47:48Z christian

can adhere to a user-defined system of citations or the global DOI (Digital Object Identifier) system `http://www.doi.org`. The field can also hold a list of individual citation keys which in turn point to a separate database. An example for a DOI is: `doi:10.1016/j.pepi.2008.05.008`. The string can be wrapped in an URL using the DOI referrer scheme at `http://dx.doi.org` like this: `http://dx.doi.org/10.1016/j.pepi.2008.05.008`. The example for a verbose, old fashioned citation would be adhering to the Author-Year citation, carrying at least the standard Harvard-like citation data: Heine, C., Müller, R.D., Steinberger, B, Torsvik, T. H. (2009): Anomalous subsidence in intracontinental basins. Phys. Earth Planet. Int. 71:23-39 would result into `Heine et al.,PEPI, 2009,71:23-39`.

`MOD_AUTHOR` The `MOD_AUTHOR` attribute is a container which should carry information about the author who initially digitised the feature or modified it subsequently. It should consist of an Authorname, institution name, and/or email address. If possible, the authorname can be replaced with a clear name email address. Example: `StanDard_EarthByte_USYD` or `Stan Dard; std@whatev.er` or `Stan.Dard@what.ev.er`

`MOD_DATE` The date attribute should be used to capture the date of modification of any given feature in the dataset as ISO-Date (YYYY-MM-DD, optional + THH:MM:SS). This can be altered automatically through the use of the field calculators in GIS.

`STDVERSION` A version number assigned to the version of the metadata standard, as layed out in this document. Starting with version 1.0 as according to the release of this document.

Handling of optional attribute fields

Optional attributes can be appended based on the requirements of the user, organisation or software. There are no restrictions on the amount of additional attributes beyond the standard set proposed in this document.

Standardised attribute names from old PLATES header

To ensure compatibility with the old PLATES standard and metadata preservation the fields of the old PLATES header, it is suggested to unify the attribute names for the remaining header infos which are not covered in the "standard" set (see above, Section **??**)

`PLATEID2` Conjugate plate identification code, specifying the plate.

CHHEI_ShapefileAttributeStandard.tex 1792 2012-05-11 14:47:48Z christian

Attribute name 1234567890	Type;length	Description/Comment
PLATEID2..	short; 4	Unique PlateID for conjugate plate
MEAS_VALUE	numeric	Measured value
MEAS_UNIT.	text; 50	Unit of measure

Table 2: Proposed optional set of shapefile attributes for PLATES format compatibility and metadata preservation

MEAS_VALUE he equivalent of the *PLATES Data type code number* - this has been used as "measured value" for e.g. bathymetry contours. One example would be `BA 2000` for the 2000 m isobath contour line in a PLATES header.

MEAS_UNIT The SI measurement unit of the original field.

CHHEI_ShapefileAttributeStandard.tex 1792 2012-05-11 14:47:48Z christian

Data type codes

These data type codes are used by GPlates when reading a properly attributed shapefile. The data type codes will automatically be translated to the appropriate GPGIM element.

Data Type Code	FeatureType (GPlates)
AR	AseismicRidge
BA	Bathymetry
BS	Basin
CB	PassiveContinentalBoundary
CF	ContinentalFragment
CM	PassiveConinentalBoundary
CO	PassiveContinentalBoundary
CR	Craton
CS	Coastline
EC	ExtendedContinentalCrust
FA	PassiveConinentalBoundary
FT	Fault
FZ	FractureZone
GR	OldPlatesGridMark
GV	Gravimetry
HF	HeatFlow
HS	HotSpot
HT	HotSpotTrail
IA	IslandArc
IC	Isochron
IM	Isochron
IP	SedimentThickness
IR	IslandArc
IS	UnclassifiedFeature
LI	GeologicalLineation
MA	Magnetics
NF	gpmlFault
N1	NavdatSampleMafic
N2	NavdatSampleIntermediate
N3	NavdatSampleFelsicLow
N4	NavdatSampleFelsicHigh
OB	OrogenicBelt
OP	BasicRockUnit
OR	OrogenicBelt
PA	MagneticAnomalyIdentification
PB	InferredPaleoBoundary
PC	MagneticAnomalyIdentification

CHHEI_ShapefileAttributeStandard.tex 1792 2012-05-11 14:47:48Z christian

PL	Pluton
PM	MagneticAnomalyIdentification
PO	PoliticalBoundary
RA	IslandArc
RF	Fault
RI	MidOceanRidge
SM	Seamount
SS	Fault
SU	Suture
TB	TerraneBoundary
TC	TransitionalCrust
TF	Transform
TH	Fault
TO	Topography
TR	SubductionZone
UN	UnclassifiedFeature
VO	Volcano
VP	LargeIgneousProvince
XR	MidOceanRidge
XT	SubductionZone

Table 3: Data type codes recognised by GPlates when reading properly attributed shapefiles.

GIS implementation and workflow example

Starting out new

The empty shapefiles can be loaded into any GIS and used for digitizng new features. I am using QGIS here to demonstrate the workflow.

Applying the standard to existing shapefiles

The empty files can be merged with existing ones to provide the standard class of attributes even if the user already has shapefiles containing attributed features. This can be done through a *merge* in all currently available GIS packages. As the empty shapefile does not contain any data apart from empty attribute columns there should be limited possibilities for conflicts.

The following section illustrates such a merge using ArcGIS on Windows using the Data Management → General → Merge tool.

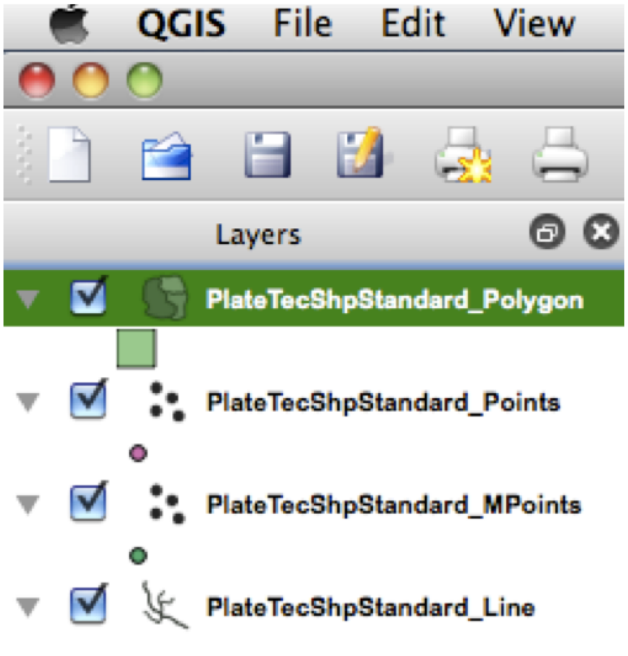

Figure 2: Loading the empty shapefiles into QGIS.

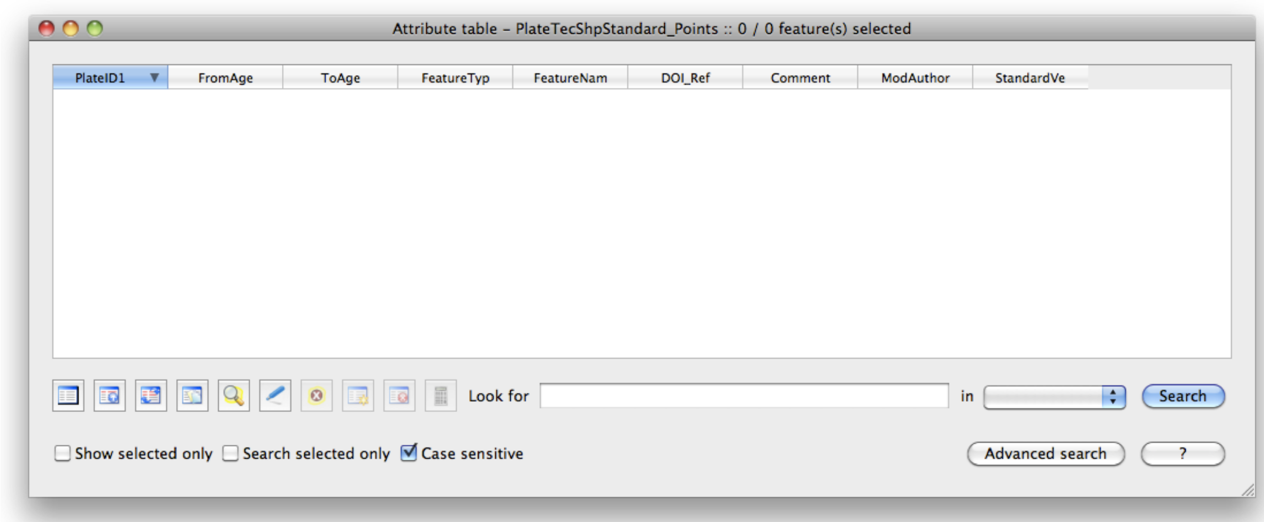

Figure 3: The emtpy attribute table in QGIS with GPlates shapefile standard columns.

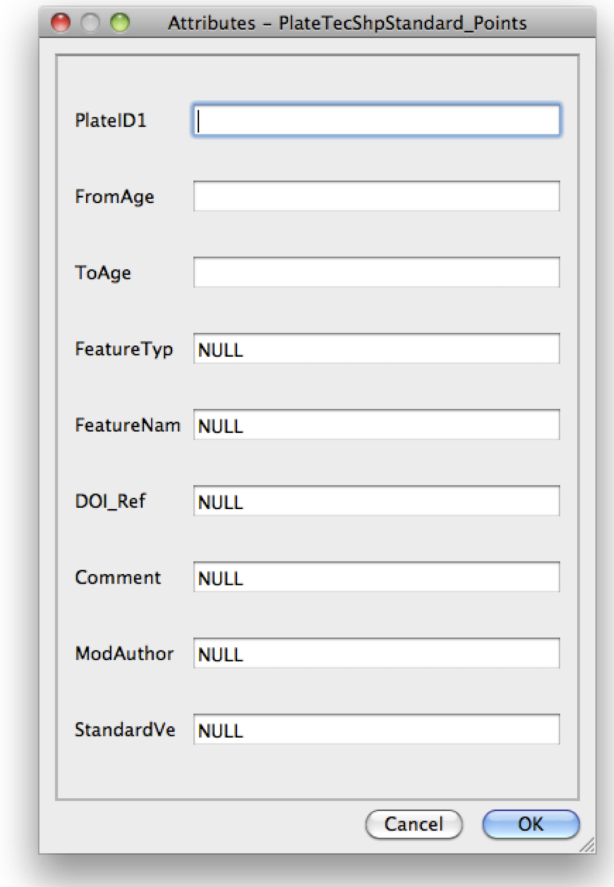

Figure 4: Dialogue to populate the feature attributes in QGIS, requires editing mode to be set to "on" for the respective layer.

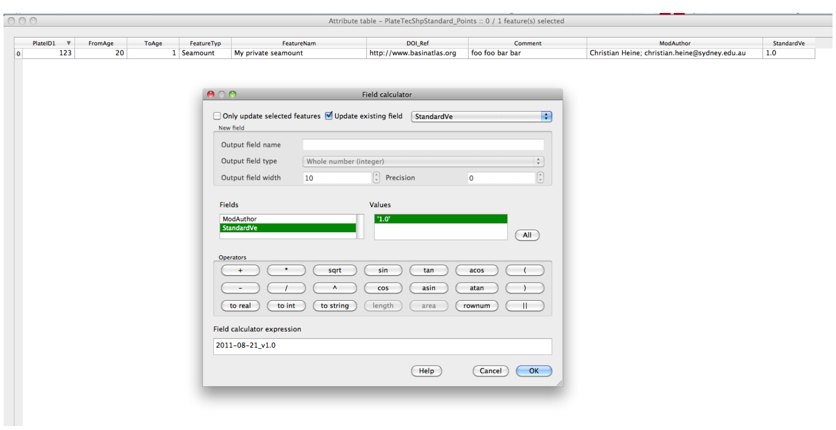

Figure 5: Example of the field calculator in QGIS which can be used to batch-populate attributes in a shapefile table. For example the user can modify the `Mod_Author` column to set one value to all existing features in the shapefile.

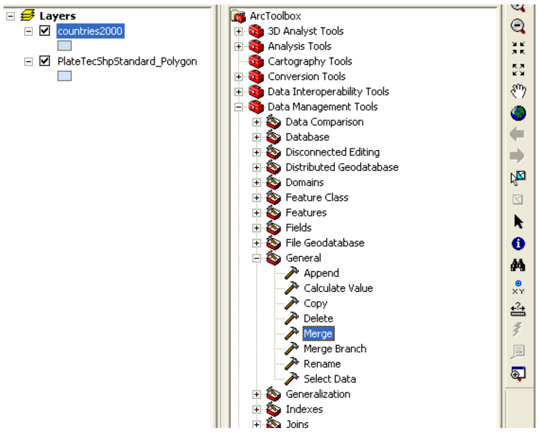

Figure 6: The Merge tool in ArcGIS's Toolbox

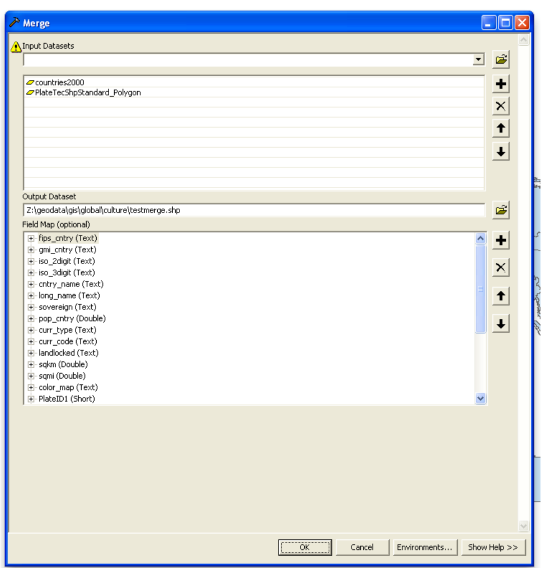

Figure 7: The Merge dialogue window in ArcGIS

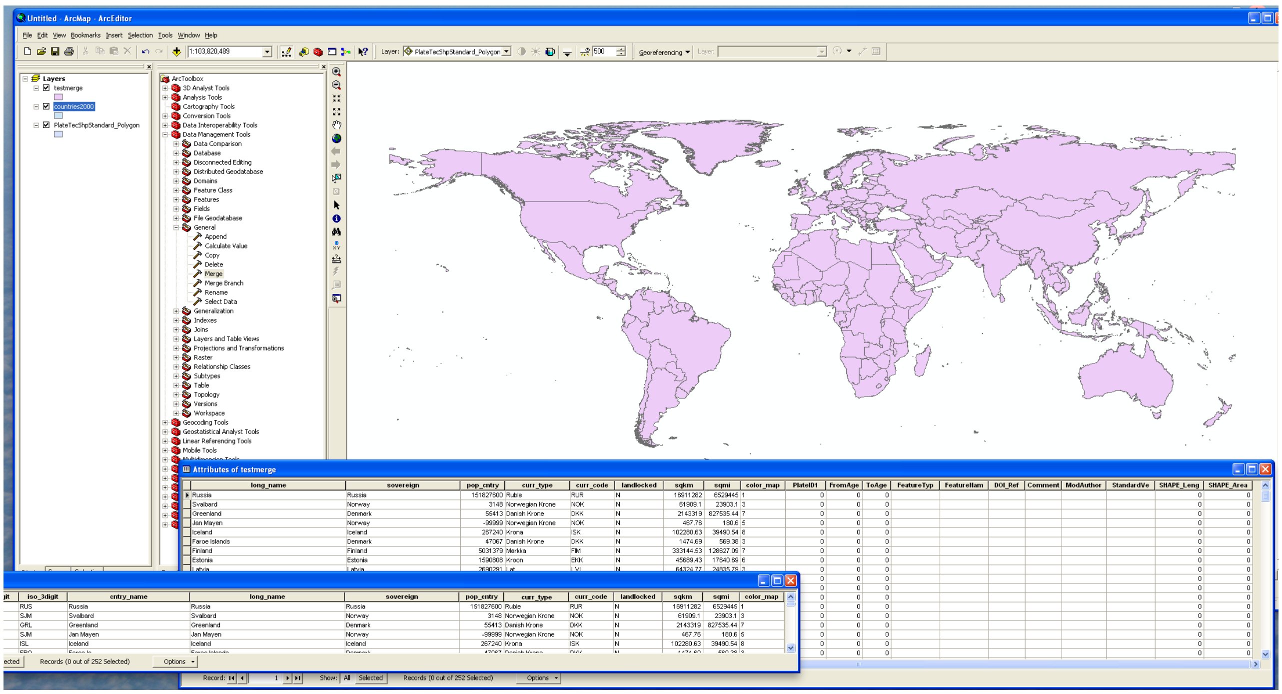

Figure 8: New shapfile generated by merging the provided Shapefile standard data set with a global country polygon dataset. The resulting new "testdata" layer has attributes of both shapefiles (countries2000 and the PlateTecShpStandard_Polygon) merged. These can now be polulated using ArcGIS' field calculator.