

# Manual of Geospatial Data



OCHA

United Nations  
Office for the Coordination  
of Humanitarian Affairs

## PURPOSE AND SCOPE

This manual outlines the working practices of the collection, management and dissemination of geographic data at the United Nations Office for the Coordination of Humanitarian Affairs (OCHA). The purpose of this document is to set a systemized standard for data collection, storage and dissemination for efficient geographic information exchange and best practices in interoperability.

This document is consists of both strategic and technical practices covering multiple technologies to ensure the interoperability of data collected, managed and disseminated by OCHA with other relevant datasets including but not limited to geographic data, statistical data, etc.

This document is a working document and is subject to updates as needed. Please submit any feedback to [ocahvisual@un.org](mailto:ocahvisual@un.org).

# CONTENTS

- OVERVIEW OF PROCEDURES ..... 4
- DATA NEEDS ASSESSMENT ..... 5
- DATA INVENTORY ..... 7
- DATA PREPARATION ..... 7
  - Quality Check..... 7
  - Cleaning..... 7
  - Naming Convention ..... 7
  - Metadata Population ..... 8
- DATA STORAGE ..... 8
- DATA SHARING..... 9
  - Data formats ..... 9
  - Dissemination Channels ..... 9
- DATA PREPARATION MANUAL ..... 10
  - Quality Check..... 10
  - Cleaning..... 12
  - Naming Convention ..... 21
  - Metadata ..... 25
- TABLE 1: FEATURE DATASET CATEGORIES..... 27
- TABLE 2: FEATURE CLASS CATEGORIES ..... 29
- TABLE 3: METADATA CONTENT ..... 33
- DATA STORAGE MANUAL..... 35
  - Single-User Geodatabase..... 35
- DATA SHARING MANUAL..... 36
  - Data Formats and Conversion ..... 36

## OVERVIEW OF PROCEDURES

Data collection is a concurrent process, and an integral part of the daily work of all OCHA Geographic Information Systems (GIS) Officers and the vast majority of Information Management Officers (IMO). In order to be efficient in the data collection process (avoiding duplication of efforts and identifying gaps) as well as ensure that the highest quality data available are collected, five steps are identified in processing geospatial data: assess, collect, prepare, store and share.

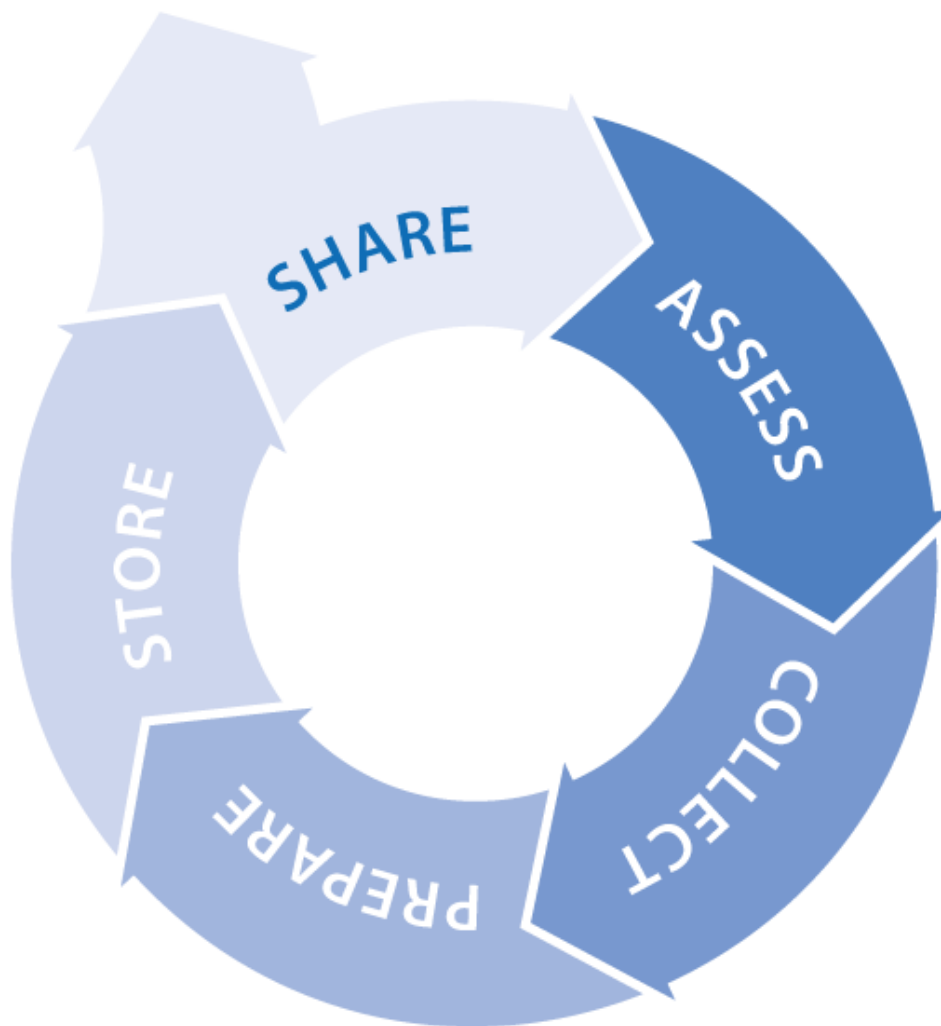


Figure 1 – The data cycle

## DATA NEEDS ASSESSMENT

The aim of the needs assessment is to identify the data thematic and the geographic scope of the data needed to support operations and decision-making for all actors in a humanitarian response. Ideally a needs assessment should take place for countries identified as vulnerable to a natural disaster or complex emergency. If the needs assessment did not take place prior to an emergency, the following methodology for data review is valid during the onset of a crisis.

### Thematic

The Inter-Agency Standing Committee (IASC) has outlined the Common Operational Datasets (COD) and Fundamental Operational Datasets<sup>1</sup> (FOD) needed in disaster preparedness and response. CODs are predictable, core sets of data needed to support operations and decision-making for all actors in a humanitarian response. FODs are datasets required to support multiple cluster/sector operations and complement the common operational datasets. These datasets are characterized by thematic areas such as education or health facilities.

It is of utmost importance that the “best available” source for the each of the Common Operational Datasets is identified. Multiple sources of the same data can create confusion and inhibit information flow between agencies.

OCHA is the “Guardian” of the CODs and will facilitate the distribution of the “best” available CODs/FODs in emergencies while managing forums for updates and distribution. Quality assurance for compliance with minimum format and data characteristics in CODs/FODs will be conducted by OCHA prior to distribution. These geographic data should be prepared in a way in which they are interoperable with non-spatial common data (such as demographic) via a common code (see [Data Preparation Manual](#) page 14). Additional datasets may be identified particular to a country and/or emergency.

The geographic common operational datasets are listed in the table below.

---

<sup>1</sup> IASC. *IASC Guidelines Common Operational Datasets (CODs) in Disaster Preparedness and Response*. Dec 2010.

Dataset	Recommended Governance	Key Data Included in Attributes
<b>Administrative Boundaries</b> (Geographic) admin level 1 admin level 2 admin level 3 admin level 4	<ul style="list-style-type: none"> <li>- Guardian: OCHA</li> <li>- Sponsor: OCHA (Other potential sponsors could include UNDP, Government agencies or INGOs)</li> <li>- Source: Government</li> </ul>	<ul style="list-style-type: none"> <li>- Unique identifier (P-Code)</li> <li>- Name</li> </ul>
<b>Populated Places</b> (Geographic)	<ul style="list-style-type: none"> <li>- Guardian: OCHA</li> <li>- Sponsor: OCHA, (Other potential sponsors could include UNDP, Government agencies or INGOs)</li> <li>- Source: Government</li> </ul>	<ul style="list-style-type: none"> <li>- Unique identifier (P-Code)</li> <li>- Names</li> <li>- Size classification</li> <li>- Population statistics</li> <li>- Status if capital of administrative division</li> <li>- Type (Village, spontaneous settlement, collective center, planned settlement)</li> </ul>
<b>Transportation Network</b> (Geographic)	<ul style="list-style-type: none"> <li>- Guardian: OCHA</li> <li>- Sponsor: Logistic Cluster</li> <li>- Source: Government</li> </ul>	<ul style="list-style-type: none"> <li>- Roads (Classified by size)</li> <li>- Railways</li> <li>- Airports/helipads</li> <li>- Seaports</li> </ul>
<b>Hydrology</b> (Geographic)	<ul style="list-style-type: none"> <li>- Guardian: OCHA</li> <li>- Sponsor: OCHA (Other potential sponsors could include UNDP, Government agencies or INGOs)</li> <li>- Source: Government</li> </ul>	<ul style="list-style-type: none"> <li>- Rivers (Classified by size)</li> <li>- Water bodies</li> </ul>
<b>Hypsography</b> (Geographic)	<ul style="list-style-type: none"> <li>- Guardian: OCHA</li> <li>- Sponsor: UNOSAT</li> <li>- Source: Remote sensing, Government</li> </ul>	<ul style="list-style-type: none"> <li>- Elevation</li> <li>- Resolution</li> </ul>
<b>Humanitarian Profile</b> (disaggregated by admin level and populated place)	<ul style="list-style-type: none"> <li>- Guardian: OCHA</li> <li>- Sponsor: OCHA</li> <li>- Source: Government, Assessments, UNHCR</li> </ul>	<ul style="list-style-type: none"> <li>- Internally Displaced<sup>2</sup></li> <li>- Non-displaced affected</li> <li>- Host family/resident community affected</li> <li>- Refugee<sup>3</sup></li> <li>- Dead</li> <li>- Injured</li> <li>- Missing</li> </ul>
<b>Population Statistics</b>	<ul style="list-style-type: none"> <li>- Guardian: OCHA</li> <li>- Sponsor: OCHA, UNFPA (Other potential sponsors could include UNDP, Government agencies or INGOs)</li> <li>- Source: Government</li> </ul>	<ul style="list-style-type: none"> <li>- Total population by admin level (Individuals)</li> <li>- Total population by admin level (Number of Households)</li> <li>- Age</li> <li>- Sex</li> <li>- Average family size by admin level</li> <li>- Unique identifier</li> </ul>

### Geographic Scope

Each office should house the identified thematic data for all countries and regions which fall under their area of work.

<sup>2</sup> As defined in the UN Guiding Principles on Internal Displacement UN Doc. E/CN.4/1998/53/Add.2

<sup>3</sup> As defined in Refugee: Article 1, The 1951 Convention Relating to the Status of Refugees

## DATA INVENTORY

The aim of the inventory is to compile a comprehensive list of the “best available” source of data for each thematic and geographic region identified in the needs assessment. This includes a review of the existing in-house data, updates to existing in-house data and other data available not previously collected or considered against the needs identified in the needs assessment.

The primary data source of data should always be the government or in-country institution supporting a country’s mapping activities. OCHA staff should be in contact with the National Mapping Agency in country in identifying these “best available” data, and keeping awareness of any updates.

## DATA PREPARATION

Data are only as good as their quality and usability. Before data are stored (see [Data Storage Manual](#) page 35) and shared (see [Data Sharing Manual](#) page 36), the standardized data preparation process outlined below should be followed to ensure that quality of the data is documented. All data stored and disseminated are quality checked, cleaned if necessary, named according to a standard naming convention and have a complete metadata record.

It is not always possible to have a perfect dataset, however, it is possible to determine the limitations of any dataset and document them to ensure that the data is used appropriately by the user community.

### Quality Check

All data stored and disseminated should have a known source, complete geographic extent, complete and accurate attribute information, known projection, correct topology and should be up-to-date and/or relevant to the current situation. If data do not meet these criteria, and/or the data cannot be cleaned to meet these criteria, the sources of the data should be reviewed. If no better data is readily available, these problems should be documented in the metadata record. See detailed guidance in the Data Preparation Manual under [Quality Check](#) (see page 10).

### Cleaning

Before data are used and shared, their geometry and attributes should be cleaned. Geometries should be not have any topological errors and attributes should include information pertinent to their use and not include irrelevant or undocumented information. In addition, subsequent datasets may be created from existing (e.g. polylines created from polygons) to enhance in cartographic representation. See detailed guidance in the Data Preparation Manual under [Verifying Attributes](#) (page 14) and [Verifying Geometry](#) (page 15).

### Naming Convention

The purpose of having a standardized naming convention is to provide an organized framework for the data, ensuring interoperability between users and platforms. OCHA has a defined naming convention which includes the most important information identifying a geographic dataset using recognizable coding. The convention is outlined in the Data Preparation Manual under [Naming Convention](#) (see page 21).

## Metadata Population

The population of metadata is vital to ensure that the copyright of the data are preserved and any specifics on the use of the data are maintained. OCHA uses the ISO 19115 metadata specification, as endorsed by the United Nations Geographic Information Working Group (UNGIWG). More specifics are outlined in the [Data Preparation Manual](#) under Metadata (see page 25).

## DATA STORAGE

Data may be stored in a variety of manners depending on the format of the data. OCHA stores data in ESRI geodatabases which, depending on the office and the licensing, may be single-user or multiuser. OCHA has a defined structure for the geodatabase which is outlined in the [Data Storage Manual](#) (see page 25). Geodatabases are recommended because they can store a number of feature datasets and feature classes in one package and advanced rules and relationships, relational models, editing, access and dissemination techniques may be applied. Data from a geodatabase may be exported in a variety of formats including shapefiles and kml/kmz ensuring data compiled by OCHA may be disseminated to partners who may not have ESRI software. The graphic below is an overview of the single-user and multiuser geodatabases. More detailed information can be found on ESRI's website.<sup>4</sup>

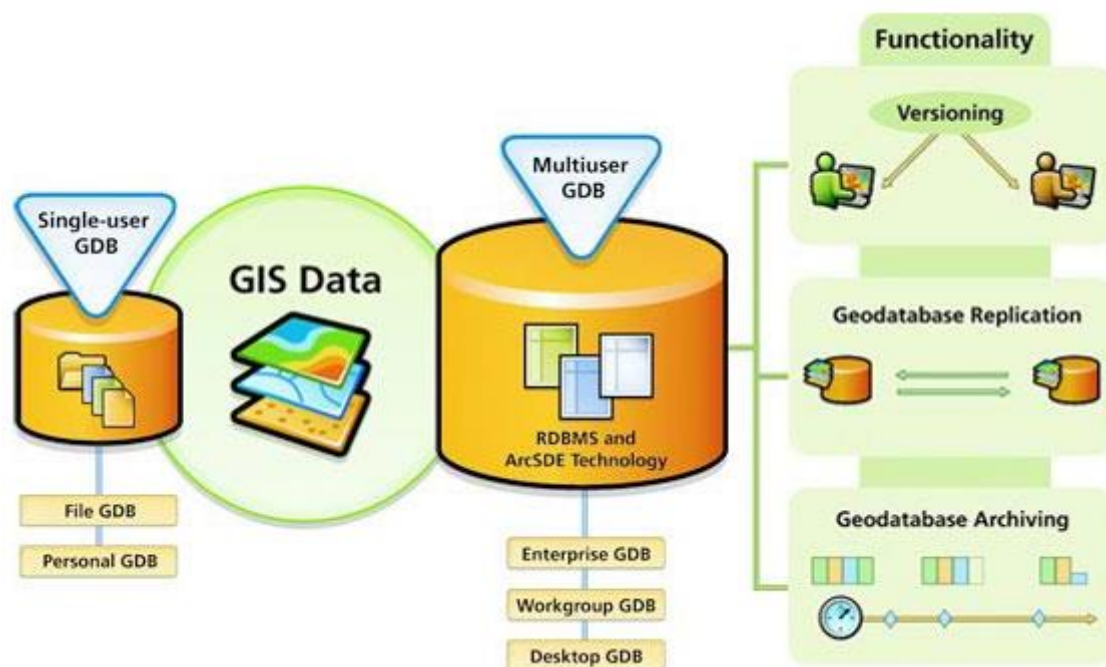


Figure 2 - Overview of ESRI geodatabases<sup>4</sup>

<sup>4</sup> <http://www.esri.com/software/arcgis/geodatabase/index.html>



## DATA SHARING

It is important that data are shared in formats which can be used by a user base ranging from technical to non-technical through channels where they can reach the widest audience with least effort on all parties involved.

### Data formats

#### **Static Data**

Static data include but are not limited to geodatabases, shapefiles and kml/kmzs. It is recommended that all data are disseminated as geodatabases, shapefiles and kml/kmz.

At minimum and dependant on resources, all data should be disseminated as a geodatabase, shapefile and kml/kmz through a website or ftp.

Technical guidance on data conversion of static datasets is outlined in the [Data Formats and Conversion](#) on page 36 of the [Data Dissemination Manual](#).

#### **Web Service**

Web services such as web feature services are a preferred option for data dissemination as they can be directly accessed through a variety of programs and allow for real-time updates, meaning the users do not have to check for updated versions of the data. The caveat dissemination through a web feature service is that the service provider must have a fixed server.

It is important to understand that online mapping portals such as a Web Map Service, Google Maps, etc. are a good option to disseminate the information in a dataset in a geographic display to users, who do not need to create their own maps, performed customized analysis or manipulate the data in any way. An online mapping portal does not allow access to the actual data, only for online visualization.

### Dissemination Channels

To ensure continuity, data prepared for a specific country, mission or emergency should be shared with OCHA headquarters and other interested parties preparing global datasets.

Depending on the office resources, data may be disseminated through a data portal, website, ftp or email. For each country and/or emergency, a file geodatabase of the common operational datasets should be disseminated to all interested partners via the appropriate channel of a particular office/situation. Technical guidance on the preparation, naming convention and database creation and population can be found in the [Data Preparation Manual](#) and [Data Storage Manual](#).

## DATA PREPARATION MANUAL

This manual provides technical guidelines on checking the quality of data and various editing techniques to ensure data are used appropriately and to their fullest capacity as well as outlines a standard naming convention and metadata population to ensure easy access by a wealth of users and good preservation.

### Quality Check

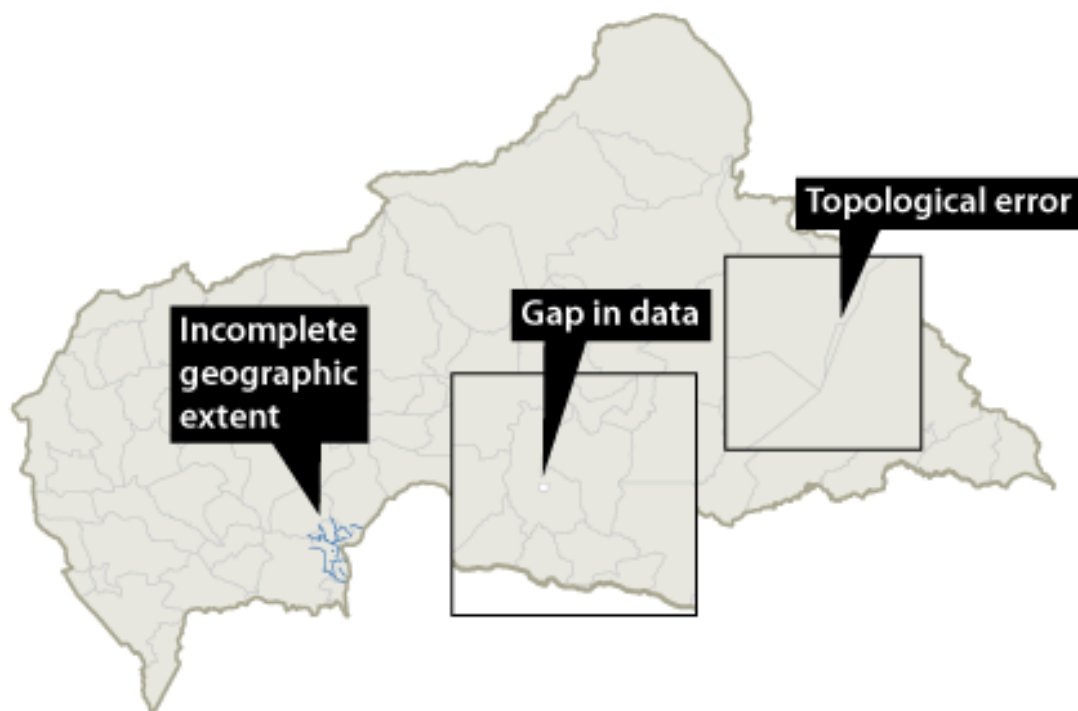


Figure 3 - Examples of data which are either incomplete in geographic scope or have topological errors.

The six themes outlined below should be considered before using or disseminating geographic data. If the data do not meet the criteria defined by these themes, and/or the data cannot be cleaned to meet these criteria, the sources for these data should be reviewed. If there is no other option to correct the problems, these issues should be documented in the metadata.

#### **1. They have a known source**

Data should not be used if the source is unknown because there is no guarantee of the verification of the data or the appropriate permission to use the data.

#### **2. They are complete in geographic scope**

Data need to span the entire country(s)/region(s) of interest. See example in [Figure 3](#). In this case more research may be done to see if data are available which span the entire country.

#### **3. They have complete and accurate attribute information**

If data do not have information about each geographic feature, there is an increased risk for the data to be used incorrectly. See more specific details on how much attribute information are needed under Editing > [Verifying Attributes](#) (see page 14).

#### 4. They have a known projection

Unknown or incorrect coordinate reference systems (including datum and projection) can prevent the data from being overlaid properly with other sources of geographic information and incorrect spatial analysis. If the data's coordinate reference system is unknown, refer to the source of the data to see if the original coordinate reference system can be determined.

If data are unprojected (that is, they use latitude and longitude) it is likely that the correct datum is WGS84, however this is not always true and applying the wrong datum can result in hundreds of meters of error.

#### 5. They are up-to-date and relevant to the current situation

The information associated with the data must be up-to-date OR useful to the situation for analysis. See example below of administrative boundaries not reflecting current situation. However, if updated data are not available, out of date data are better than none, but the problem should be documented in the metadata record.

#### 6. They have correct topology

The spatial properties of the data must be accurate for the data to be used correctly. See example of topological errors in a polygon file in [Figure 3](#). Topology is checked differently for polygon, arc and point files.

**Point Files** – All points are generally in the correct location. Two examples of files that do not pass the topology check are 1) a file where a type error was made in the latitude and/or longitude field(s) of the file and the point is not in the correct location or 2) the location for a populated place is obviously incorrect (e.g. located in the ocean or incorrect administrative unit).

**Polygon and Arc Files** – No gaps and/or overlaps between the lines that make up the arcs or polygons are present in the data. Guidance on checking polygon data for errors using ArcGIS is provided below:

- Create Empty Geodatabase
  - In ArcCatalog, go to the folder where you want to save the geodatabase.
  - Right click in open space and choose *New > Personal Geodatabase*
  - Name the geodatabase **ISO3\_update\_yymmdd.mdb** (e.g. usa\_update\_070801.mdb)
- Create Feature Dataset
  - Right click on the geodatabase and choose *New > Feature Dataset*
  - Name the dataset **ISO\_topology** (e.g. usa\_topology). Click *Next*.
  - Select the *Geographic Coordinate System > World > WGS 1984*. Click *Next*
  - For the vertical coordinate systems, leave as <None>. Click *Next*. Click *Finish*
- Import Layer to Feature Dataset
  - Right click on the feature dataset and choose *Import > Feature class (single)...*
  - Specify the following:
    - Input Features: The path to the data under check
    - Output Location: The empty feature dataset created in section 2 above
    - Output Feature Class: Name the feature class **ISO3\_work\_yymmdd**
- Create Topology
  - Right click on the feature dataset (**ISO3\_topology**) and choose *New > Topology...*
  - Click *Next*.
  - Accept the default name (**ISO3\_topology\_Topology**) by clicking *Next*

- Select the feature that will participate in the topology and Click *Next*
- Click *Next*
- Click *Add Rule...*
  - Select *Rule > Must Not Overlap* and Click OK
  - Leave *Show Errors* checked
- Click *Add Rule...*
  - Select *Rule > Must Not Have Gaps* and Click OK
  - Leave *Show Errors* checked
- Click *Next*
- Click *Finish*
- Click *Yes* to validate the topology
- Visualize Errors of the Topology
  - Click on **ISO3\_topology\_Topology** and the *Preview* tab
  - If the preview of **ISO3\_topology\_Topology** shows only one red line that follows the outline of the main polygon, the file is free of topological errors.
  - If the file has topological errors, it will have multiple lines; one associated with the main polygon, and the others associated with the locations of the gaps and/or overlaps in the data (see [Figure 4](#)).

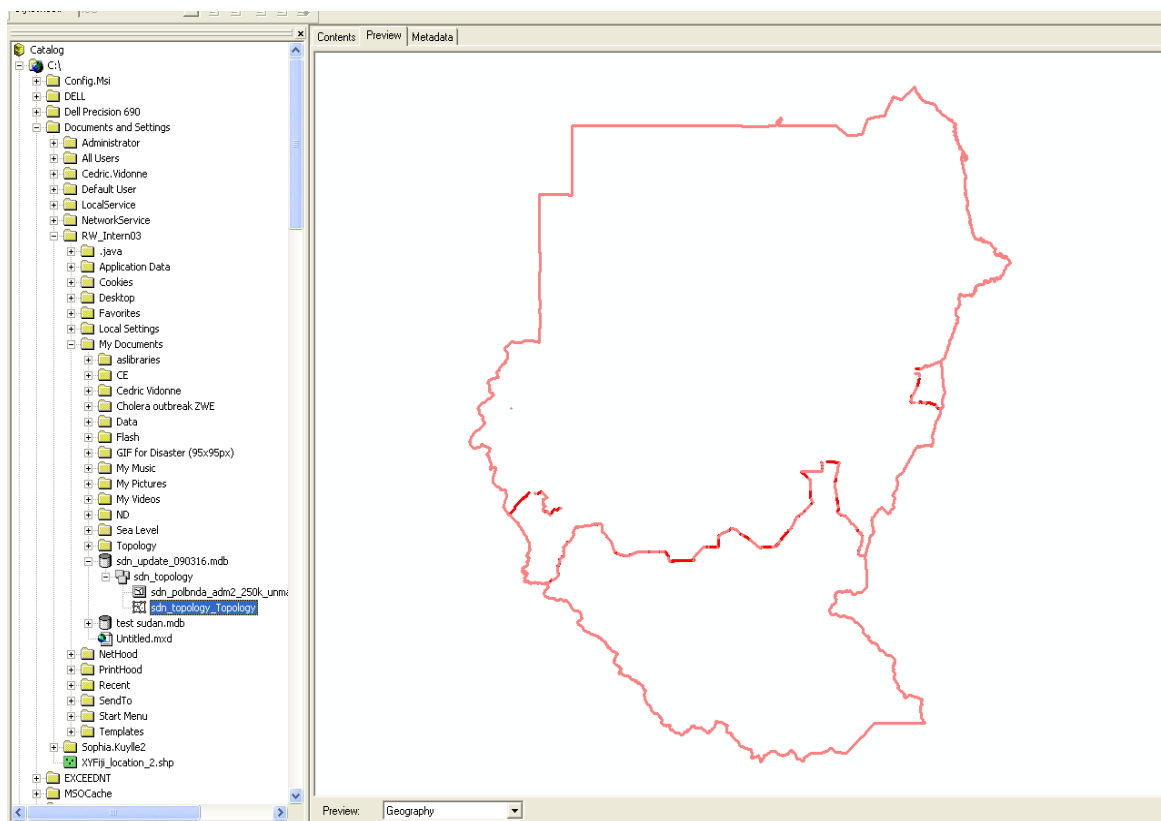


Figure 4 - The image shows an administrative unit polygon file, where a new topology was created to produce the layer shown in red outline. The map shows the extra lines in red inside the external boundary (in this case the international boundary), these represent the topological errors.

Detailed technical guidance for fixing errors in data polygon topology is covered in [Editing Geometry > Topological Errors](#) on page 15.

## Cleaning

### Defining the Coordinate System

The coordinate reference system (CRS), often referred to as the “projection” of a given dataset, is simply a defined set of values that describe how to interpret the X and Y coordinates stored in a dataset. For example if we have a shapefile containing only one point and it’s coordinates are X=34.2 and Y=45.7, we have no idea if these are degrees of latitude and longitude, meters in UTM, or state plane feet. It is the coordinate reference system that specifies degrees, meters, or feet as well as other necessary parameters such as origin or various projection parameters.

The CRS has two primary parts: a projection (or possibly unprojected in the case of latitude-longitude coordinates), and a datum. The datum defines which mathematical model of the Earth’s shape is used in the CRS. Commonly this is WGS84, however in some regions other systems may be more commonly used.

If the CRS definition is missing from the dataset, many GIS softwares will assume that it is in geographic (lat/long) based on the WGS84 datum. If this assumption is wrong, the data will not align with other datasets have correct CRS definitions.

There are two operations that are used to change the CRS:

- *Defining* the CRS (or projection in ArcGIS terminology) does not change the coordinates in the data, it simply changes how the coordinate values are interpreted.
- *Reprojecting* the data actually changes the coordinate values based on a mathematical transformation from the old CRS to the new CRS.

In the data cleaning process, there is usually no need to *reproject* the data, only to make sure that the CRS definition is correct.

To view/edit the CRS, follow the guidance below:

- To view the coordinate reference system of a dataset
  - In ArcCatalog or ArcMap, right click on the dataset and choose properties
  - Choose the *XY Coordinate System* tab
  - The coordinate system is displayed
    - Note: if the coordinate system is *GCS\_Assumed\_Geographic* or something similar, it means that the CRS is undefined and ArcGIS is assuming that the coordinates are geographic (latitude-longitude) based on the WGS84 datum. If this assumption is accurate, the coordinate system should be redefined to *GCS\_WGS\_1984* as described below in order to eliminate any ambiguity.
- To change the coordinate reference system of a dataset
  - Ensure the dataset is not open in any ArcMap project
  - In ArcCatalog, right click on the dataset and choose properties
  - Choose the *XY Coordinate System* tab
  - Click *Select*
  - Locate the name of the CRS that the data set is actually in (remember, we are not reprojecting here, only defining)
    - For example, to define a data set as latitude-longitude based on the WGS84 datum, choose *Geographic Coordinate Systems>World>WGS 1984.prj*
- To define the same projection on many datasets at the same time
  - Ensure that none of the datasets are open in any ArcMap project
  - Open ArcToolbox
  - Choose *Samples>Data Management>Projections>Batch Define Coordinate System* and complete the wizard

To be included in future version of this document.

## Verifying Attributes and Features

### Essential vs. Non-Essential Information

A Dataset may have too much or not enough information in the attribute table. Too little information may cause the data to be unusable or misrepresented. Too much information may cause confusion to users of the data and also lead to misrepresentation. To guide in what is too much and too little information, attributes are broken into essential, marginal and external according to the following definitions:

It is fundamental to establish a link between spatial data and demographic or situational thematic data. If geospatial data are prepared with unique codes shared with those collecting demographic and situational information, geospatial analysis is possible.

#### Essential

Essential information is information pertinent to the use of the geometry and which provide linkages between the geometry and other essential datasets such as demographic or situational assessments. At very minimum, all data should include a unique ID and a standard name for each feature.

#### Marginal

Marginal information is information which aid in the use of the geometry. Data may be valid without this information, but this information enhances their use. For example, a polyline layer for rivers are usable with only the feature name (e.g. river name) and classification (e.g. river type), however marginal information such as drainage basin size may enhance the use.

#### External

External information is information which can be stored in a table and linked to the data using a unique ID. This information may be time sensitive (e.g. demographic data or a specific thematic such as % food insecure population, security risk, etc). The information should be stored in tables which share a unique ID to the geometry. Joins and related can be used to link these data when needed. Keeping these data external to the geodataset avoids the inclusion of obsolete data.

Essential	Marginal	External
Unique ID/Pcode	Feature Group 1	Demographic Information
Feature Name	Feature Group 2	Affected Population
Attribute differentiator	Feature Type 1	Risk
	Feature Type 2	...

Figure 5 - Essential and marginal information should be stored as attributes to geographic data. External information should be stored separately as tables, and linked by the unique ID using joins and relates.

### Attribute Names

Geodatabases allow for long attribute names, which is useful, but may become problematic when exporting to formats that do not support long filenames. Ideally, attribute names should be limited to 10 characters and not begin with numbers or contain spaces. The alias function in ArcCatalog can be used to assign a longer and more descriptive name to the attribute that will be visible in ArcGIS applications.

### Features

A geodatabase should include the minimum number of feature classes possible per feature dataset. Features of like geometry (e.g. points or polygons or lines) and content (e.g. populated places or land use or road) should be one feature class with the classification of each feature differentiated using an attribute field.

Example: Populated place gazetteer

In many cases there is a separate shapefile or feature class associated with the different types of populated places. Consider the following:

Shapefile/Feature Class 1: National Capital

Shapefile/Feature Class 2: Administrative Capitals

Shapefile/Feature Class 3: Cities with population greater than 100,000

Shapefile/Feature Class 4: Cities with population between 50,000 and 99,999

Shapefile/Feature Class 5: Cities with population less than 49,999

Ideally these separate layers are combined into one feature class for all populated places including small towns, large cities, administrative capitals, national capitals, etc. These files can easily be merged by creating an empty feature class with the following schema (essential attributes):

Attribute Field Name	Controlled Vocabulary	Data type
Unique ID/Pcode	No	Short Integer
Feature Name	No	Text
Feature Type	Yes	Text
Source	Maybe	Text

In this case, controlled vocabulary for the feature type defines the type of populated place (e.g. national capital, administrative capital, etc.)

OCHA does not have a defined schema for any particular dataset, but follows standards used by partners and the providers of the data.

### Verifying Geometry

In order to ensure data are used correctly and accurately in geospatial analysis and cartographic representation, topological errors should be fixed and features developed prior to use.

### Topological Errors

One of the most common problems in data quality is the presence of gaps or overlaps in polygons such as administrative boundaries. Detailed technical guidance for editing data topology in ArcGIS is below.

- Open ArcMap
  - Import into ArcMap **ISO3\_topology\_Topology** (created in [Quality Check](#) on page 11)
  - Click on **Yes** to add feature classes that participate in topology
- Set the Editing Parameters – In order to perform the editing procedures, it is important to set the editing parameters as follows:
  - Change map Units
    - Go to *View > Data Frame Properties*
    - Click on the General Tab
    - Change Units – Display: to Meters
    - Click *Apply > Okay*
  - Set up the parameters
    - If it's not already there, add the Editor toolbar to your workspace by going to *View > Toolbars > Editor*
    - On the toolbar, go to *Editor > Options*
    - Under the General Tab chose Snapping Tolerance: 50 map units

- Under the General Tab choose Sticky Move Tolerance: 1000 map pixels (This will prevent accidental moves of polygons during the editing process.)
    - Click *OK*
  - Start the Editing Process
    - Start Editing
      - Click on Editor > Start Editing
      - Select the Task: Modify Feature and the Target: **ISO3\_work\_yymmdd**
    - Set Snapping Parameter
      - Go to *Editor > Snapping*
      - For the **ISO3\_work\_yymmdd** layer, check the boxes for Vertex and Edge  
If toolbar not already opened, add Topology Tools
      - Go to *Editor > More Editing Tools > Topology* (another toolbar will be added)
  - Edit Overlap Errors

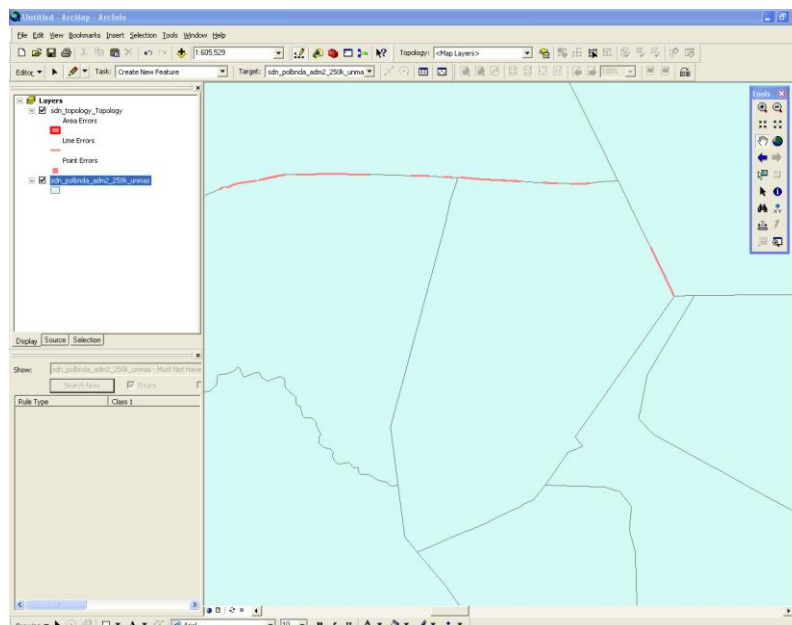

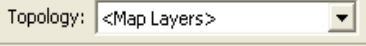





Figure 6 - Shows **ISO3\_topology\_topology** in red and **ISO3\_work\_yymmdd** in blue

- Select Map Topology (  ) from toolbar > Select **ISO3\_work\_yymmdd** > Click *OK*
    - Go to Topology toolbar > Click on <Map Layers> (  ) and Select **ISO3\_topology\_Topology**
    - Select all Overlap Errors using the Error Inspector tool (  ) > Show **ISO3\_work\_yymmdd** – Must Not Overlap > Click on *Search Now*
    - Select all the errors
    - Correct the selected Overlap Errors using the Fix Topology Error tool (  ) > Right click on the map and choose *Subtract*
    - Verify there is no more error with the Error Inspector tool (by repeating Step • )
    - Go to *Editor > Save Edits*
  - Edit Gap Errors (limited to one polygon at a time)
    - Zoom to error
    - Using the Fix Topology Error tool (  ), select all the topological errors associated with one of the polygons > Right Click > Select *Create Feature* (see [Figure 7](#))



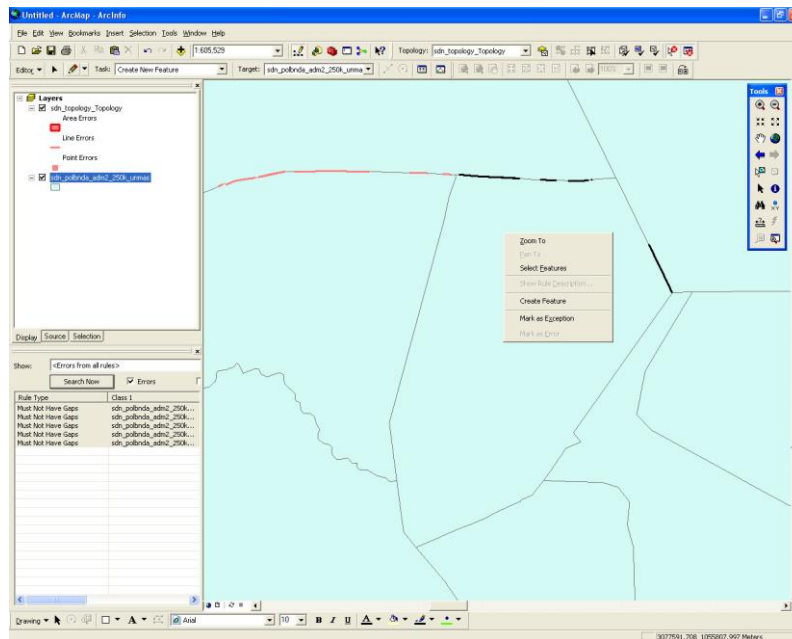



Figure 7 - Shows the topological errors selected from **ISO3\_topology\_topology** in black

- Select the polygon and the new polygons created using the Edit Tool (  ) or open the attribute table and highlight the main polygon and the newly created polygons (they will be at the bottom of the table and have null values).
- Go to Editor > *Merge* to merge the selected polygons. Select the main polygon as the feature with which polygon will be merged (to keep the Attributed Table information) > Click OK.
- Save Edits
- Continue for all errors

## Clip administrative data to the international boundary

- Create a working copy of the data for the country administrative boundaries
  - In ArcCatalog, right click the feature class and choose copy, then paste.
  - Name the feature class **ISO3\_adm2\_extended\_yymmdd** (e.g. usa\_adm2\_extended\_070801).
- Expand the boundary so that it is outside the international boundary
  - Start Editing.
  - Select and move one vertex at each end so they are beyond the international boundary and then delete all vertexes in between to obtain a polygon that extends beyond the international boundary, as illustrated below. To preserve topology don't move the vertex that connects the two adjacent administrative units but instead use the next one along, and ensure that the extended lines of adjacent administrative units are snapped together.
  - Save Edits.

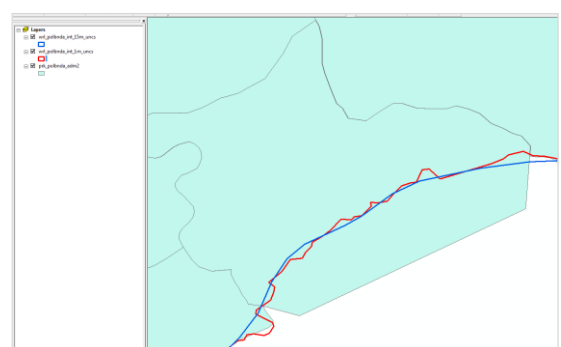
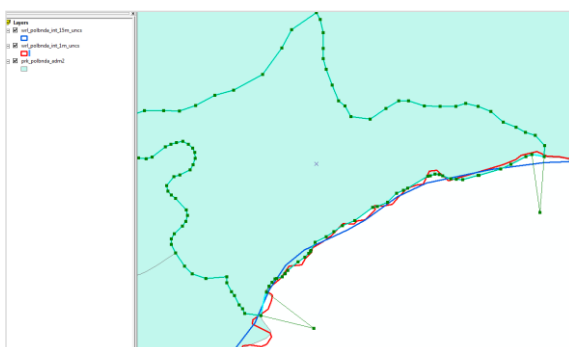


Figure X - These images shows how to extend the administrative unit boundaries beyond the international boundaries when preparing for clipping to the international boundaries.

Particular attention should be paid to vertices that form the intersection between country boundary and an administrative boundary. Note the following special cases and how they should be resolved.

*Special Case 1:* The boundary between 2 administrative units does not extend to the international boundary.

If this case is encountered the following procedure should be followed:

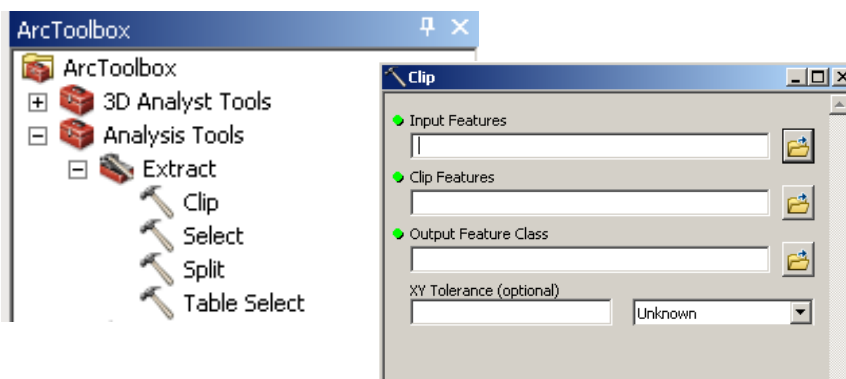
- 1) Measure the distance between the extremity of the line dividing the 2 administrative units and the international boundary. From that point we consider two cases:
  - a) this distance is smaller than 1 km. In this case we consider that the administrative boundary can be extended straight.
  - b) this distance is bigger than 1 km. In this case you will have to check with other layers such rivers, lakes, mountain ranges, etc. to see if the boundary is following a natural elements.

*Special Case 2:* Islands.

If this case is encountered the following procedure should be followed:

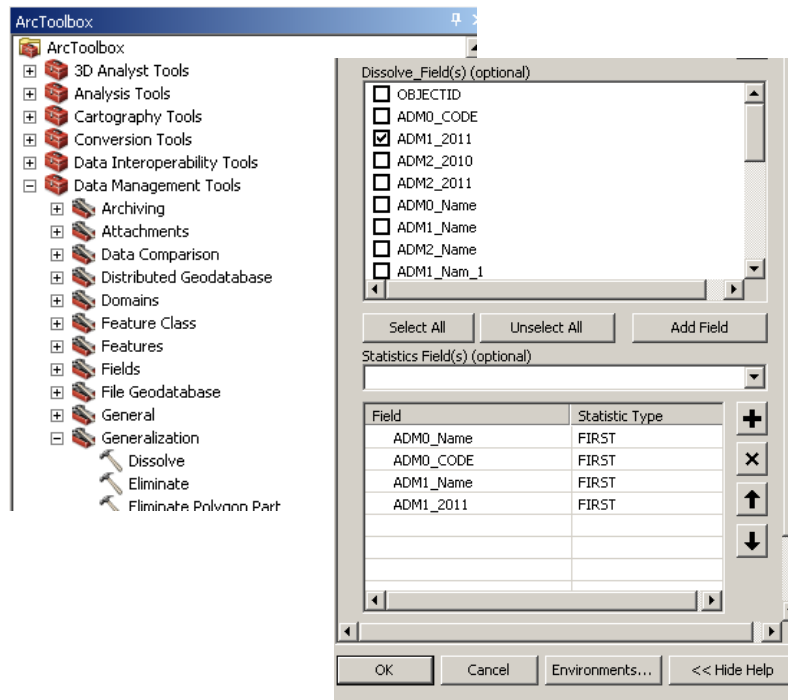
- 1) Identify which administrative unit the island belongs to.
- 2) Extend the administrative unit boundaries in a way that preserves the administrative unit correctly.

- Once all the administrative units boundaries have been extended do a topology check and fix any errors present in the feature class (refer to instructions for topology).
- Clip to the international boundary
  - In ArcToolbox open *Analysis Tools > Extract > Clip*.
  - Complete the tool parameters.
  - Name the output feature class according to the naming conventions.



- Review the output data
  - Check that the number of records/features is the same as it was at the start of the process.
  - Check the full extent of the boundary visually making sure everything has clipped as expected – paying particular attention to locations with islands or complex intersections between administrative units.
  - Do a topology check and fix any errors present in the feature class (refer to instructions for topology).

Note if clipping needs to be done for multiple administrative unit levels the work above should be completed for the lowest level, and the data for the higher level units should then be generated from the lowest level data to ensure consistency in the geometry between the levels. The higher level units should be generated by 'dissolving' the units on the attribute code for the unit level concerned (*ArcToolbox > Data Management > Generalization > Dissolve*) and using the *statistics field* option to populate attributes that should be retained in the output.



### Homogeneisation of the attribute table for the administrative data

In order to obtain an homogeneous data set the final attribute tables of each shape file should only contain the following fields:

- Cntry\_name: Country name
- Cntry\_code: ISO3 code of the country
- Adm1\_name: names of the administrative units at the 1st level
- Adm1\_code: administrative unit code at the 1st level
- Adm2\_name: names of the administrative units at the 2nd level
- Adm2\_code: administrative unit code at the 2nd level

### Create polyline from polygon

It is recommended that all data repositories include polygons AND polylines of administrative and international boundaries. Boundaries should be represented with the polyline file and background landmass and labels should be represented with polygons. Preparing data repositories with both of these files in advance allows for quick map creation, and proper cartographic representation.



Figure 8 – The 4 maps above are meant to demonstrate the importance of appropriate data preparation of international and administrative boundaries in cartographic representation. The figure at top left displays the map polygons and each consecutive map show the polyline layers added in order from bottom (first administrative boundary) to top (coastline).

The polyline layer should always be created from same source and data that will be used for the polygon layer. A simple method for creating a line from polygon file in ArcGIS is outline below:

- Create the line file from the polygon in ArcCatalog
  - Open ArcCatalog
  - Right click on the polygon layer and select *Export > to coverage*
  - Maximize the coverage to see the arc, label, polygon, region and tic. Right click on the arc and select *Export > to Shapefile (single) or geodatabase*
- Remove the outer border in ArcMap\*
  - Open ArcMap
  - Add the arc shapefile created from the coverage
  - Go to Editor, and select *> Start Editing*
  - Go to Selection in the main panel, and select *Select by Attributes*
  - Set query "RIGHTPOLY" = 1

\* The outer boundary is removed so that it does not conflict with the boundary of the administrative unit at a higher level. For example, first administrative level boundaries will be encompassed by international boundaries, and international boundaries will be encompassed by coastlines. See [Figure 8](#) for reference

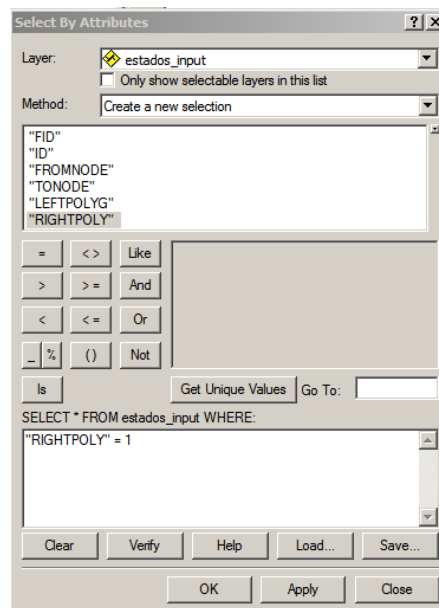


Figure 9 - Query to select outer line.

- Select *Apply* and *OK*
- You will see the outline highlighted. Now select *Delete* on your keyboard, and the outline is deleted.

## Naming Convention

The purpose of having a standardized naming convention is to provide an organized framework for the datasets, ensuring interoperability between users and platforms.

### **Base Vector Data**

As OCHA's preferred storage environment is the ESRI geodatabase, OCHA has defined naming conventions for baseline geographic data at the 3 levels of the geodatabase: geodatabase, feature dataset and feature class.

When distributing data in alternative formats, such as shapefiles or kml/kmz, the feature class naming convention may be used for the files. The files may then be organized in folders according to the feature dataset names.

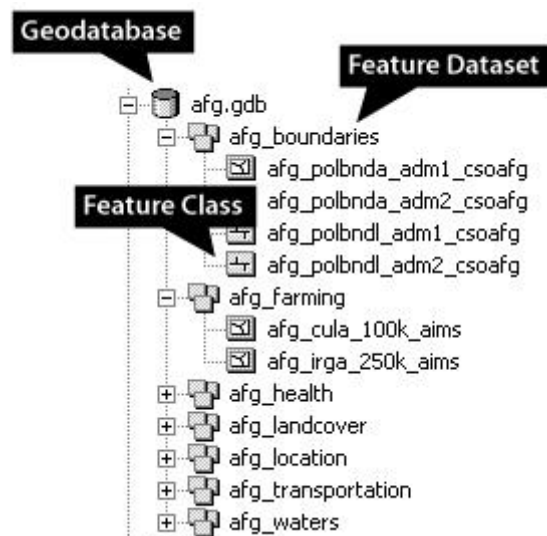


Figure 10 - Example of geodatabase for Afghanistan.

## Geodatabase

The name of the geodatabase is from the International Organization for Standardization (ISO) country code, ISO3 code<sup>5</sup> of the country/region of interest. For example: wrl, afg, alb, etc. See Table X for a list of ISO3 codes.

## Feature Dataset

Feature datasets are objects that are used to group together related feature classes. There are two parts to the feature dataset name naming convention, each separated by an underscore (\_). They are as follows:

**ISO3 Code** – As with the geodatabase name, the first part of the naming convention consists of the ISO3 code. For example: wrl, afg, alb, etc.

**ISO 19115 Topic Category** – The second part of the name is one of the 19 ISO19115 Topic Categories. For example: boundaries, farming, health, etc. See [Table 1: Feature Dataset Categories](#) on page 27 for a list of the categories and their associated definitions.

Example: Provinces in Afghanistan

ISO3 Code = afg

ISO 19115 Topic Category = boundaries

Feature Dataset Name = afg\_boundaries

## Feature Class

Within each feature datasets can be a number of different feature classes in a number of different formats (e.g. point, line, polygon, annotation). For example, in a boundaries feature dataset you

<sup>5</sup> FAO has published ISO3 codes as a resource free of charge here: <http://www.fao.org/countryprofiles/iso3list.asp>

may find first administrative level boundary polygons, first administrative level boundary lines, second administrative level polygons, etc. See [Table 2: Feature Class Categories](#) on page 29 for a list of feature classes. The table only covers data which may fall under the Common Operational Datasets<sup>1</sup> and are in the holdings of the OCHA offices in New York and Geneva. If data obtained by OCHA are not covered in this table, the person responsible for the input of the data should consult with the OCHA geospatial data focal point consider the addition of a new class.

There are 5 parts to the naming convention, each separated by an underscore (\_). They are as follows:

- ISO3 Code - As with the geodatabase name, the first part of the naming convention consists of the ISO3 code. For example: wrl, afg, alb, etc.
- Feature Class Code + Feature Class Type – The feature code as defined in [Table 2: Feature Class Categories](#) followed by the first letter of the feature class type where:

a = polygon  
l = arc  
p = point  
t = text  
r = raster

- Field Code (if applicable) – The field code (if applicable) as defined in [Table 2: Feature Class Categories](#). For example, for political boundaries field codes include: adm1, adm2, adm3, etc.
- Scale – The denominator for the scale of the dataset in the following form:

Example 1 – 1:1,000,000 = 1m

Example 2 – 1:250,000 = 250k

Example 3 – scale not known or of mixed scales (should be documented in metadata) = unk

Example 4 – scale not applicable for this dataset (such as utm zone boundaries) = na

Example 5 – for raster data, this parameter is the nominal pixel size in meters or cm = 30m, 130cm

- Source – The acronym or short version of the source of the data. See Table X for a list of acronyms and source name short versions.

Example 1 – United Nations Cartographic Section = uncs

Example 2 – Government of Guinea = govgin

Example: World International Boundaries

ISO3 Code = wrl  
 Feature Class Code = polbnd  
 Feature Class Type = polygon  
 Feature Sub-Class Code = int  
 Scale: 1:1,000,000  
 Source: United Nations Cartographic Section  
 0  
 Feature Class Name = wrl\_polbnda\_int\_1m\_uncs

### *Special Case 1: Two Datasets having the same naming convention*

In the case where two datasets have the same naming convention and there is insufficient time to clean the data to merge them to one dataset (see Editing > [Verifying Geometry](#) on page 15), numbers are used to differentiate between the two datasets and differences are specified in the metadata title and abstract until the data may be combined to one. The numbers run in descending order from the dataset at the lowest detail to the dataset at the highest detail. Consider the following:

#### Example of Feature Class Naming Convention – Special Case 1

*Two sets of population data for a particular country, one has the population for major cities and the other population data for small towns. The data for major cities are labeled with a “1” and the data for small towns are labeled with a “2”.*

Dataset 1: Major cities in Burundi from Government of Burundi at 1:1M scale

Dataset 2: Cities in Burundi from Government of Burundi at 1:M scale

Feature Class Names (interim solution)

Dataset 1: bdi\_pplp1\_1m\_gov

Dataset 2: bdi\_pplp2\_1m\_gov

Feature Class Name (long term solution)

Combine the two feature classes to 1 using guidance from [Verifying Geometry](#). The resulting label would be: bdi\_pplp\_1m\_gov.

### *Special Case 2: Data do not span an entire country or region*

In the case where the dataset only covers part of a country, administrative names are used to differentiate between administrations and city names are used to differential between urban areas. See example below:



### Example of Feature Class Naming Convention – Special Case 2

Dataset 1: IDP Camps in Aceh, Indonesia

Dataset 2: IDP Camps Afgooye Cooridor, Somalia

#### Feature Class Names

Dataset 1: idn\_aceh\_cmpp\_idp\_1m\_unhcr

Dataset 2: som\_afgooye\_cmpp\_idp\_1m\_unhcr

### Situational/Emergency Specific Vector Data

To date OCHA does not have a standard framework for situational vector data such as image derived vectors for flooding or damage. This will be addressed in future versions of this manual. However, it is desirable to try and adapt the above described naming convention for these data layers to the extent possible.

### Raster Data

To date OCHA does not have a standard framework for raster data. This will be addressed in future versions of this manual.

## Metadata

The population of metadata is vital to ensure that the copyright of the data are preserved and any specifics on the use of the data are maintained. OCHA uses the ISO 19115 metadata specification, as endorsed by the United Nations Geographic Information Working Group (UNGIWG). The specifics for this standard can be found under Interoperable Services on the UNGIWG website.<sup>6</sup>

### ArcCatalog Environment

Metadata are created and/or updated in ArcCatalog using the ISO metadata editor. To set the metadata editor in ArcCatalog:

- ▣ Open ArcCatalog
- ▣ Go to *Tools > Options*
- ▣ Go to the *Metadata* tab
- ▣ Choose the *Default Style sheet* as *ISO*
- ▣ Chose the *Metadata Editor* as *ISO Wizard*

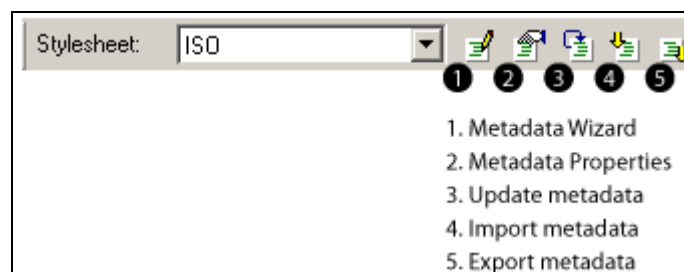


Figure 11 - Overview of the metadata toolbar in ArcCatalog.

Metadata templates may be exported as an xml for use in other data catalogs or reloaded in ArcCatalog to populate metadata for data with like information (e.g. 5 different layers from the same datasets/source). To export the metadata as an xml, use the Export metadata tool in the

<sup>6</sup> <http://www.ungiwg.org/interoperable.htm>

metadata toolbar and to import metadata from an xml use the Import metadata tool (see [Figure 11](#)).

**Metadata Content**

All available information on the dataset should be stored in the metadata and at minimum the fields outlined in [Table 3: Metadata Content](#) on page 33.

TABLE 1: FEATURE DATASET CATEGORIES

ISO Code	Feature Dataset	Definition
001	Farming	Rearing of animals and/or cultivation of plants Examples: agriculture, irrigation, aquaculture, plantations, herding, pests, and diseases affecting crops and livestock
002	Biota	Flora and/or fauna in a natural environment Examples: wildlife, vegetation, biological sciences, ecology, wilderness, sea life, wetlands, habitat
003	Boundaries	Legal land descriptions Examples: political and administrative boundaries
004	Climate	Processes and phenomena of the atmosphere Examples: cloud cover, weather, climate, atmospheric conditions, climate change, precipitation
005	Economy	Economic activities, conditions, employment Examples: production, labor, revenue, commerce, industry, tourism and ecotourism, forestry, fisheries, commercial or subsistence hunting, exploration and exploitation of resources such as minerals, oil and gas
006	Elevation	Height above or below sea level Examples: altitude, bathymetry, digital elevation models, slope, derived products
007	Environment	Environmental resources, protection and conservation Examples: environmental pollution, waste storage and treatment, environmental impact assessment, monitoring environmental risk, nature reserves, landscape
008	Geoscientific	Information pertaining to earth sciences Examples: geophysical features and processes, geology, minerals, sciences dealing with the composition, structure and origin of the earth's rocks, risks of earthquakes, volcanic activity, landslides, gravity information, soils, permafrost, hydrogeology, erosion
009	Health	Health, health services, human ecology and safety Examples: disease and illness, factors affecting health, hygiene, substance abuse, mental and physical health, health services
010	LandCover	Base maps Examples: land cover, topographic maps, imagery, unclassified images, annotations

ISO Code	Feature Dataset	Definition
011	Intelligence	Military bases, structures, activities Examples: barracks, training grounds, military transportation, information collection
012	Waters	Inland water features Examples: rivers and glaciers, salt lakes, water utilization plans, dams, currents, floods, water quality, hydrographic charts
013	Location	Positional information and services Examples: addresses, geodetic networks, control points, postal zones and services, place names
014	Oceans	Features and characteristics of salt water bodies (excluding inland waters) Examples: tides, tidal waves, coastal information, reefs
015	Cadastre	Information used for appropriate actions for future use of the land Examples: land use maps, zoning maps, cadastral surveys, land ownership
016	Society	Characteristics of society and cultures Examples: settlements, anthropology, archaeology, education, traditional beliefs, manners and customs, demographic data, recreational areas and activities, impact assessments, crime and justices, census information
017	Structure	Man-made construction Examples: buildings, museums, churches, factories, housing, monuments, shops, towers
018	Transportation	Means and aids for conveying persons and/or goods Examples: roads, airports/airstrips, shipping routes, tunnels, nautical charts, vehicle or vessel location, aeronautical charts, railways
019	Utilities	Energy, water and waste systems and communications infrastructure and services Examples: hydroelectricity, geothermal, solar and nuclear sources of energy, water purification and distribution, sewage collection and disposal, electricity and gas distribution, data communication, telecommunication, radio, communication networks

TABLE 2: FEATURE CLASS CATEGORIES

Feature Dataset Name	Feature Class Alias	Feature Class Code	Feature Sub-class Name	Feature Sub-class Code
Biota	Multiple	biota		
	Cleared Way / Firebreak*	firebrk		
	Grassland*	grass		
	Hedgerow*	hedge		
	Marsh / Swamp*	swamp		
	Oasis*	oasis		
	Trees*	trees		
	Tundra*	tundra		
Boundaries	Barrier	barrier	Multiple	
			Fence	fence
			Wall	wall
	Geographical	geobnd	Continent	cnt
			Region	reg
	Grid	grd	Multiple	
			Latitude*	lat
			Longitude*	lon
			Time Zone	
	Markers	markers	Multiple	
			Control Point*	control
	Political	polbnd	International	int
Admin Level			admX	
Cadastre	Multiple	cadastre		
Climate	Multiple	climate		
	Precipitation	precip	Isohyet	isohyt
Economy	Livelihood	live		
	Market	mkt		
Elevation	Multiple	contour		
	Elevation Contour	elev		
	Depth Contour	depth		
Environment	Multiple	environment		
	Natural / National Park*	landmrk		
Farming	Multiple	landuse		
	Agricultural Area*	agri		
	Agricultural Storage Site	agristr		
	Irrigated Area*	irg		
	Orchard / Vineyard	orchard		

\*Those which should ideally be part of one more general complete dataset, where the features are differentiated in the attribute table and not through a separate feature dataset or feature class. See Data Preparation Manual under [Verifying Attributes](#) on page 14. These classes are outlined here as use for a temporary solution for incomplete datasets and/or datasets/gazetteers under consideration and/or development.

Feature Dataset Name	Feature Class Alias	Feature Class Code	Feature Sub-class Name	Feature Sub-class Code
Geoscientific	Multiple	geoscientific		
	Caves, Mountains & Volcanoes	mtn	Multiple	
			Cave*	cave
			Mountain Pass*	pass
			Volcano*	vol
	Cuts & Embankments / Fills	embank		
	Glaciers & Snow / Ice Fills	landice		
	Hydrogeology	hydro		
	Ice Shelf, Polar Ice, Pack Ice	seaice		
	Land Forms	Indfrm	Multiple	
Fault			fault	
Natural Resource	resource	Multiple		
		Oil / Gas Field*	oil	
Health	Multiple	hlt		
	Facility	hltfac		
LandCover	Multiple	Indcvr		
Location	Multiple	ppl		
	Built Up Area	cuiltup		
	Camp	cmp	IDP Camp*	idp
Refugee Camp*			ref	
Oceans	Multiple	oceans		
	Coastline	coast	Multiple	
Islands*			isl	
Society	Linguistics	ling		
	Religion	rel		
Structure	Multiple	structure		
	Bank*	bnk		
	Building General*	build		
	Factory*	fct		
	Hotel*	htl		
	Industry Feature*	ind		
	Landmark Site / Park*	landmrk		
	Plaza / City Square*	plaza		
	Ruins*	ruins		
	Shelter* **	shl		
	Sport Field	sport		
	Tower (non-communication)*	tower		
	Warehouse*	whs		
Water Facility*	wtfac			

\*Those which should ideally be part of one more general complete dataset, where the features are differentiated in the attribute table and not through a separate feature dataset or feature class. See Data Preparation Manual under [Verifying Attributes](#) on page 14. These classes are outlined here as use for a temporary solution for incomplete datasets and/or datasets/gazetteers under consideration and/or development.

\*\*May be a combination of other structures, but used as a shelter. If these data are time sensitive, should be stored as a table and linked to the structure data via a unique ID. Data Preparation Manual under [Verifying Attributes](#) on page 14.

Feature Dataset Name	Feature Class Alias	Feature Class Code	Feature Sub-class Name	Feature Sub-class Code
Transportation	Multiple	transportation		
	Airdromes	airdrm	Multiple	
			Unspecified*	unsp
			Airports*	airp
			Airfields*	airf
			Airstrips*	airs
			Fixed HLZs*	fhlz
			Improved HLZs*	lhlz
	Bridges	bdge		
	Obstacles	obstacle		
	Ports	prt	Multiple	
			Unspecified*	unsp
			Seaports*	sea
			Riverports*	rvr
			Lakeports*	lake
			Beachings*	beach
			Banks*	bank
			Piers*	pier
			Ramps*	ramp
			Achorages*	anch
Railways	rlw			
Roads	rds			
Stations	stn			
Waterways	wtw			
Utilities	Multiple	utilities		
	Communication Buildings / Towers	comm.		
	Disposal Site / Wreck Yard	extract	Multiple	
			Mine	mine
	Materials Treatment Plant*	treat		
	Obstructions*	obstr		
	Pipelines*	pipe		
	Powerlines & Power Plants*	power		
	Pumping Stations*	pumping		
	Rigs & Wells*	rigwell		
	Substations*	substat		
	Telephone / Telegraph Lines*	tele		

\*Those which should ideally be part of one more general complete dataset, where the features are differentiated in the attribute table and not through a separate feature dataset or feature class. See Data Preparation Manual under [Verifying Attributes](#) on page 14. These classes are outlined here as use for a temporary solution for incomplete datasets and/or datasets/gazetteers under consideration and/or development.

Feature Dataset Name	Feature Class Alias	Feature Class Code	Feature Sub-class Name	Feature Sub-class Code	
Waters	Multiple	inlandwaters			
	Aqueduct*	aqueduct			
	Cistern*	cistern			
	Dam	dam			
	Land Subject to Inundation	inund			
	Lakes & Reservoirs	lakeres			
	Rapids & Waterfalls	rapids			
	Water Course	watcrs	Multiple		
			Canal*	canal	
			Ditch*	ditch	
			Fluvial Islands*	isl	
			Rivers / Streams*	rvr	
	Watersheds*	watshed			
Water Point	watr				
Wells & Springs	wellspr				

\*Those which should ideally be part of one more general complete dataset, where the features are differentiated in the attribute table and not through a separate feature dataset or feature class. See Data Preparation Manual under [Verifying Attributes](#) on page 14. These classes are outlined here as use for a temporary solution for incomplete datasets and/or datasets/gazetteers under consideration and/or development.



TABLE 3: METADATA CONTENT

Information	ArcCatalog ISO Metadata Wizard Tab > Field Name	Description / Instruction
Data Date (if available)	General Information > Title	At minimum, record the year of the current version of the data.
Original Creation Date (if available)	General Information > Creation Date and Language	If available, at minimum record the year of the original version of the data.
Data Description	General Information > Abstract	<p>This may include:</p> <ul style="list-style-type: none"> <li>■ Primary use for which the data set was created</li> <li>■ Completeness of the dataset</li> <li>■ Modifications made from the original dataset</li> <li>■ Sharing restrictions - If the data MAY NOT be shared publically, use of the following statements: “Restricted to use by UN staff”, “Restricted to use by OCHA staff” or “For internal reference use only” and mark the reason for the restriction under <a href="#">Legal Restrictions</a></li> <li>■ Sharing – If the data may be shared freely, mention it here.</li> </ul>
Data Source	General Information > Point of Contact 1	Fill in the contact information, and assign the function to “originator”.
Data Point of Contact	General Information > Point of Contact 2	Whom to contact to acquire further knowledge or updates to the data. This may or may not be the same as the data source. Fill in the contact information, and assign the function to “point of contact”.

Information	ArcCatalog ISO Metadata Wizard Tab > Field Name	Description / Instruction
Topic Category	Dataset identification > Themes or Categories	ISO topic category as outlined in <a href="#">Table 1: Feature Dataset Categories</a> (see page 27)
Coordinate System	Automatically stored	
Use Restrictions (if any)	Dataset identification > Restrictions overview > Use restrictions	Includes any information that needs to be taken into account in the use of the data.
Legal Restrictions	Dataset identification > Restrictions overview > Legal restrictions	Copyright License Restricted

# DATA STORAGE MANUAL

This manual provides technical guidance on database creation and population in a single-user (e.g. personal or file) geodatabase environment. Guidance for a multi-user (e.g. enterprise) geodatabase environment are available upon request.

## Single-User Geodatabase

Single-user geodatabases include personal and file geodatabases. A personal geodatabase is a database using Microsoft Access as the storage system. It is limited 2GB (however performance is decreased between 250 and 500 MB), and may only be used on a Windows Operating System. A file geodatabase is a collection of GIS data held in a file system folder. File geodatabases are preferred over personal geodatabases as they are not as limited in size (up to 1TB) and may be used on a number of different platforms.<sup>7</sup>

This section covers file geodatabase creation and population. It should be noted that the process for personal geodatabase creation and population is almost identical.

### Database Creation

- Open ArcCatalog and right click on the location where the file geodatabase is to be stored (e.g. \data\vector)
- Select *New > File Geodatabase*
- Name the geodatabase as outlined in [Naming Convention](#) (see page 21)

### Database Population

- Create a feature datasets
  - Right click on the geodatabase and select *New > Feature Dataset*
  - Name the feature dataset as outlined in [Naming Convention](#)
  - Chose the Geographic Coordinate System WGS 1984
  - Leave the Vertical Coordinate System as None
  - Accept the default resolution and domain extent
- Load feature classes
  - Right click on the feature dataset and select *Load > Feature Class (single) OR Feature Class (multiple)*
  - Name the feature class as outlined in [Naming Convention](#)

#### **Trouble Shooting Tip – Projections and geodatabases**

To be included in future version of this document.

<sup>7</sup> <http://www.esri.com/news/arcuser/0309/files/9reasons.pdf>

# DATA SHARING MANUAL

## Data Formats and Conversion

At minimum and depending on resources, all data should be disseminated as a geodatabase and kml/kmz through a website or ftp.

### **Shapefiles and Geodatabase feature classes**

Guidance on geodatabase creation can be found under [Database Creation](#) on page 35.

### **Geodatabase to shapefile**

- Open ArcCatalog and navigate to the feature class which you would like to convert to a shapefile
- Right click on the feature class and select *Export > To shapefile (single)...*
- Name the shapefile according to the guidance in [Naming Convention](#) (see page 21)

### **KML/KMZ to shapefile/feature class**

There are many different methods for converting KML/KMZ to shapefiles, some work better than others and/or work for specific KML/KMZ and not others. To this date not one preferred method has been discovered. Please submit suggested alternative methods as feedback to this document.

- **ArcGIS Interoperability Extension**  
The ArcGIS Interoperability extension allows for easy conversion of various geographic data to a shapefile or feature class. If a license is available, the extension may be accessed through ArcToolbox under Data Interoperability Tools.
- **Global Mapper**  
Global Mapper is a Mapping Software which can be used to convert from KML/KMZ to shapefiles and works great with most KML/KMZ, however is not free. It can be downloaded here <http://www.globalmapper.com/>.
- **Kmltoshp**  
Kmltoshp is a free program which can be added and used in ArcToolbox however it does not work for all KML/KMZ, especially those with extensive attribute information and/or symbology. It can be downloaded here <http://arcscripts.esri.com/details.asp?dbid=15603>. An online version can be accessed here <http://www.zonums.com/online/kml2shp.php>

## **KML/KMZ**

### **Feature class or shapefile to KML/KMZ**

#### In ArcGIS

This method is preferred if you would like to preserve the symbology used in ArcGIS

- Open ArcToolbox
- Go to *Conversion Tools > to KML > Layer to KML*

What is the difference between a KML and KMZ?

A KMZ is a zipped KML file which may include images, icons, overlays, etc. used in the original creation of the KML.

- Select the layer to be converted to KML and set the output scale<sup>8</sup> to 1

### In Google Earth

This method is preferred if you would like to preserve the attribute information with the data.

- Load the shapefile to GoogleEarth
  - Open GoogleEarth
  - Go to *File > Export*
  - Select .shp as file type
  - The shapefile will appear in the places panel
- Export the file as KML/KMZ
  - Right click on the layer and select *Save Place As*
  - Save the file as KML or KMZ and name according to the guidance in Naming Convention (see page 21)

### **Tabular data to KML/KMZ**

- If not already, convert the data to a csv
- Open GoogleEarth and go to *File > Import*
- Select .csv as file type
- When prompted *Do you want to apply a style template to the features you ingested?* select 'yes'
- Choose from the options to select the style you want
  - Under the Name tab, make sure to set the *name field* to field which will be used as the label
  - The icon style, color and height can be modified in the associated tabs
- Select OK
- Save the style in a convenient location as it can be used to reproduce the same style with updating the data or creating a new dataset
- Export the file as a KML/KMZ
  - Right click on the layer and select *Save Place As*
  - Save the file as a KML or KMZ
  - To update the KML, simply make the correct changes to the csv file and create the kml following the same steps above using the style created above.

---

<sup>8</sup> Layer Output Scale is the scale at which to export the layer. Any scale-dependent rendering will be observed, so if the layer is not visible at the export scale, it will not be included in the created KML file. The symbology for the layer will be driven by this scale. Only numeric characters should be entered. For example, enter the scale as 20000, not 1:20000 or 20,000.