

*Note: This material is taken directly from the deactivated OCHA IM Wiki as a part of the IM Toolbox migration process in 2014.*

## Purpose

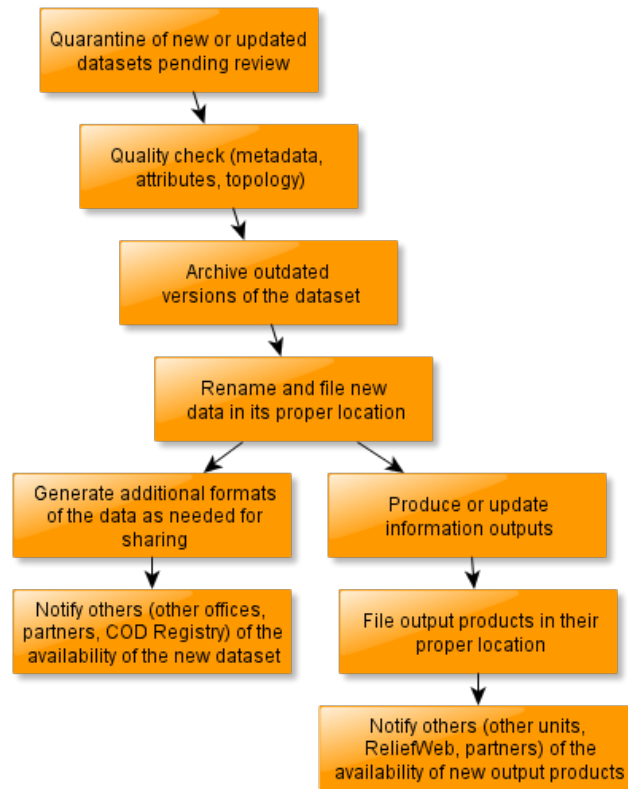
This manual outlines the recommended best practices in the management, storage and dissemination of datasets, information products, and associated files at the United Nations Office for the Coordination of Humanitarian Affairs (OCHA). These best practices were defined based on inputs from several country and regional offices as well as best practices from other UN agencies.

The purpose of this document is to provide a set of clear procedures and data structures that can support the varied uses of datasets within OCHA. In order to be considered for this document a given practice must help OCHA meet the following criteria:

- **Predictability of data filing:** new staff should be able to easily locate a given dataset
- **Archiving:** out of date data sets are retained, but do not clutter the file system
- **Ease of update to map products:** updating data sets should not break documents that link to them
- **Off-site duplication of data:** when 'official' backup systems are not in place, it must be easy to synchronize data to other locations
- **Flexibility to work quickly:** during emergency responses, data management must support the need to work quickly while maintaining sufficient order in the filing system to allow new staff to easily assume their duties
- **Interoperability:** the file formats and data structures must support data sharing with other members of the humanitarian community
- **Fulfill mandated roles:** OCHA has been mandated to be the guardian of the Common Operational Datasets. Any best practice must allow OCHA to fulfill that duty.
- **Respect data confidentiality and security:** data management strategies must ensure, to the extent practical, that sensitive data are kept secure.

## Idealized Data Management Workflow

The various components of this manual and the Geodata Preparation Manual are organized around a general data workflow.



## FILE Naming Convention

The data structure presented here represents a hybrid approach based on best practices identified in OCHA offices around the world. It includes standards both for where data is stored, in what format, and how it is named. Additionally, it provides structures for archiving outdated datasets and products.

### Folder Structure

The primary goals of the folder structure are to:

- Provide a predictable location for datasets, production files, and output products at country, regional, and global levels
- Make data and products easy to find
- Avoid breaking data links (such as within MXD files) when data is updated or copied from one computer to another

- Provide distinction between "finalized" datasets that are ready to be used in cartographic products and datasets that are still in production
- Provide a space for archiving outdated datasets and cartographic products
- Provide a space for working on datasets or products separate from the final location of the output files
- Handle data in multiple formats (Geodatabase, Shapefile, KML, GML, etc.) as well as production files (MXD, DOCX, AI, etc.) and output files (PDF, JPG, etc.)
- A note about existing folder structures

Most OCHA offices have evolved their own folder structure over time. This manual does not intend to impose this structure on any office. As offices review their folder structures or as new offices are opened, this structure is recommended as it satisfies the goals listed above. In particular, the highest level folders (the 1st and 2nd levels below the root level) are critical to enabling data synchronization between country, regional, and global offices.

## The Folder Structure

The primary goals of the folder structure are to

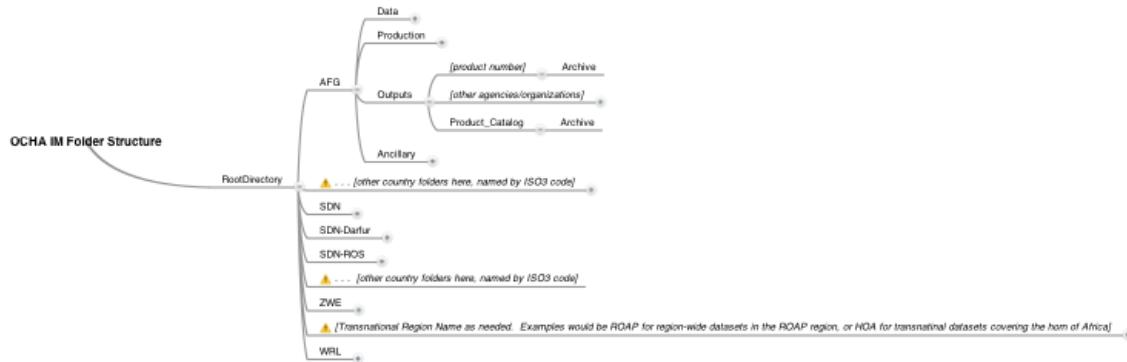
- Provide a predictable location for datasets, production files, and output products at country, regional, and global levels
- Make data and products easy to find
- Avoid breaking data links (such as within MXD files) when data is updated or copied from one computer to another
- Provide distinction between "finalized" datasets that are ready to be used in cartographic products and datasets that are still in production
- Provide a space for archiving outdated datasets and cartographic products
- Provide a space for working on datasets or products separate from the final location of the output files
- Handle data in multiple formats (Geodatabase, Shapefile, KML, GML, etc.) as well as production files (MXD, DOCX, AI, etc.) and output files (PDF, JPG, etc.)

A note about existing folder structures

Most OCHA offices have evolved their own folder structure over time. This manual does not intend to impose this structure on any office. As offices review their folder structures or as new offices are opened, this structure is recommended as it satisfies the goals listed above. In particular, the highest level folders (the 1st and 2nd levels below the root level) are critical to enabling data synchronization between country, regional, and global offices.

It is difficult to present a data folder structure in a static document. Two tools have been developed to help explain the system:

File Structure Mind Map:



A zipped empty folder structure using Afghanistan as an example is found in the example section of the IM Toolbox.

## Archiving

Notice there is no date on dataset names. This allows us to update data files without breaking links to the MXD files. To archive a file using this approach: Give the outdated file a date stamp at the end of the file name. Use the date it was removed at the end of the file name in `yyyymmdd` format. Place the outdated file in the co-located 'archive' folder. The new dataset should have the same name and location as the previous (now archived) file. Any project files referring to this data layer will then automatically use the new layer.

For example:

`AFG/Data/Boundaries/afg_polbnda_adm1_1m_salb.shp` is archived on March 13, 2011 as

`AFG/Data/Boundaries/archive/afg_polbnda_adm1_1m_salb_20110313.shp`

The updated file will be named

`AFG/Data/Boundaries/afg_polbnda_adm1_1m_salb.shp`

## Working Directories

Any folder can have a working directory. It's just a place to make a mess when you are working on something. When finished, move the file (whether it's data or a product) to its appropriate place and clean out the working folder. If there are working files you want to keep, put them into a well-named folder and store them in the appropriate Archive folder.

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

## Naming Convention for Datasets ([Naming convention for information products](#))

The naming convention for datasets (which are used to generate information products but are not usually information products in and of themselves) is different from the naming convention for information products. This section describes the naming convention for datasets, including geodata and other datasets.

There are 5 elements to the naming convention, each separated by an underscore: `_`. Optional elements are denoted by brackets: `[]`. They are as follows:

`ISO3_Code+DataType_SubCode_[Scale]_Source_[Additional Description]`

where:

- ISO3: The first part of the naming convention consists of the ISO3 code. For example: wrl, afg, alb, etc. Additional codes can be created for transnational datasets and are not limited to 3 characters. Example: hoa for horn of Africa.
- Code + Data type: The feature code as defined in the [Dataset Naming Table](#) followed by the first letter of the data type where:
  - a = polygon
  - l = arc
  - p = point
  - t = text
  - r = raster (can be omitted for data where Code = `image`)
- Sub-Code (if applicable): The sub-code (if applicable) as defined in [Dataset Naming Table](#). For example, for political boundaries sub-codes include: adm1, adm2, adm3, etc.
- Scale (optional parameter, omitted for tabular data): 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 or tabular data) = na (or omitted)
  - Example 5 – for raster data, this parameter is the nominal pixel size in kilometers, meters or cm = 30m, 130cm
- Source: The acronym or short version of the source of the data.
  - Example 1 – United Nations Cartographic Section = uncs
  - Example 2 – Government of Guinea = govgin
- Additional Description (optional parameter): This is a place holder for additional metadata that may make sense for a given type of dataset, such as:

a grid designator that may be used with datasets such as scanned toposheets or image datasets where the data is split into different files  
a date stamp for data where the specific date of publication is critical (such as humanitarian profile or other frequently published datasets)  
other metadata as needed

**IMPORTANT NOTE:** if the datasets are referenced by filename in other files (such as is common with MXD files) adding the date to the file name will often break the referring file when the date (and therefore the filename) is changed

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

In the case where two datasets have the same name and there is insufficient time to clean the data to merge them to one dataset (see [Data Cleaning](#) in the Geodata Preparation Manual), 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:

*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*

### **Dataset Names (interim solution):**

1. *Dataset 1: bdi\_pplp1\_1m\_gov*
2. *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 differentiate between urban areas. See example below:

### **Datasets not covering an entire country:**

- *Dataset 1: IDP Camps in Aceh, Indonesia*
- *Dataset 2: IDP Camps Afgooye Cooridor, Somalia*

### **Resulting Dataset Names:**

- *Dataset 1: idn\_aceh\_cmpp\_idp\_1m\_unhcr*
- *Dataset 2: som\_afgooye\_cmpp\_idp\_1m\_unhcr*

File naming within geodatabases

See also: [Folder Structure and Archiving](#)

The naming of datasets (feature classes) within a Geodatabase is identical to the scheme defined above. A geodatabase feature class, shapefile, and KML representation of the same dataset would have the same name (exclusive of the file extension). However, for geodatabases, the file naming convention must also define the names of the geodatabase and feature datasets which contain the feature classes. An example geodatabase can be found in the [folder structure](#).

**Geodatabase name:** The name of the geodatabase is from the International Organization for Standardization (ISO) country code, ISO3 code of the country/region of interest. For example: wrl, afg, alb, etc.

**Feature dataset name:** 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.
- Topic – The topic corresponds to the folder in the data structure where the data would reside in flat file formats (shapefile, kml, xls, etc.). These topics can be found in the [dataset naming codes table](#) and in the [folder structure](#).

Feature class name

As described above, the feature classes are named using the naming standard outlined above as if they were shapefiles.

Naming Convention for Information Products

OCHA Field Map names are made of four parts separated by an underscore:

1. The catalogue number, if in use (a good practice for catalog numbers is to have a three letter code for the country office and a sequential number)
2. A short map name (e.g. somalia\_3w)
3. The paper size (A4, A3, A0, etc)
4. The date of publication in YYYYMMDD format.

Examples:

- SUM001\_aceh\_reference\_map\_a4\_20050128
- LBN001\_Lebanon\_reference\_map\_20081029
- template\_sample\_a4\_20080917

## Dataset Naming Convention

This table provides some of the codes and sub-codes to be used for naming datasets as described in the [File Naming Convention](#). This list is not exhaustive and tries to address datasets commonly held by OCHA. Additional codes will almost certainly be needed by a country office to handle datasets particular to the local situation. If you would like advice on generating codes for

other datasets, or if you have identified codes that you think will be useful for other offices, please Contact ISS in Geneva.

\* Sub-codes denoted by an asterisk (\*) should ideally be part of one more general dataset, where the features are differentiated in the attribute table and not through a separate feature dataset or feature class. See the [Geodata Preparation Manual section on data cleaning](#) for details. These classes are outlined here as use for a temporary solution for incomplete datasets and/or datasets/gazetteers under consideration and/or development.