

Data formats, groups and sources in geospatial technology; main data portals of NOAA, NASA, USGS, GIS Clearinghouses

Below is a list of common and less common formats of data used in GIS analysis. Some data formats are being used solely in GIS (e.g. shapefiles, coverages) and some are used solely in remote sensing and earth science (e.g. NetCDF, HDF). Development of the software helps to integrate these data in easier way now.

Used abbreviations:

ESRI: Environmental Systems Research Institute, Redlands, CA

USGS: United States Geological Survey

NOAA: National Oceanographic and Atmospheric Administration

NASA: National Aeronautics and Space Administration

Related to ArcGIS software (ESRI Inc.):

1. Coverage (vector), original ESRI format, not much in use
2. Interchange Format (or Export Format, .e00), not much in use
3. Shapefile (vector), widely used by ESRI and other GIS software
4. ESRI GRID (raster), still in use in many ArcGIS applications
5. ESRI Geodatabases (file .gdb and personal, .mdb), standard compact formats

USGS

1. Digital Line Graph (DLG) (vector), USGS, still can be found in many USGS sites, not much in general use
2. Spatial Data Transfer Standard (SDTS), USGS, still can be found in many USGS sites, not much in general use
3. Digital Raster Graph (DRG) (raster), USGS, still can be found in many USGS sites, not much in general use

NOAA, NASA

1. NetCDF (.nc)
2. HDF (.hdf5)

Both formats are very common for many meteorological, oceanographic and other earth science data applications

General standard raster/image formats:

Various image formats: **.bil**, **.jpg**, **.tiff**, **.sid**, etc.

These and others are common to store raster data in ArcGIS and images from NASA and USGS (e.g. Landsat data).

ESRI FORMATS:

Coverage (vector format) and GRID (raster format):

COVERAGE is almost outdated but you might find them periodically. GRID is still in use.

Both data formats have **separate folders** for **spatial** data and **attribute tables**

Example: ESRI coverage named “towns” stores municipal boundaries of a township.

Its structure consists of two sub-folders: **/info** **/towns**

/info is a standard folder for storing **non-spatial information** (i.e. attribute tables or “what is there?”)

/towns is a standard folder containing **spatial information** (i.e. features; or “where is it?”)

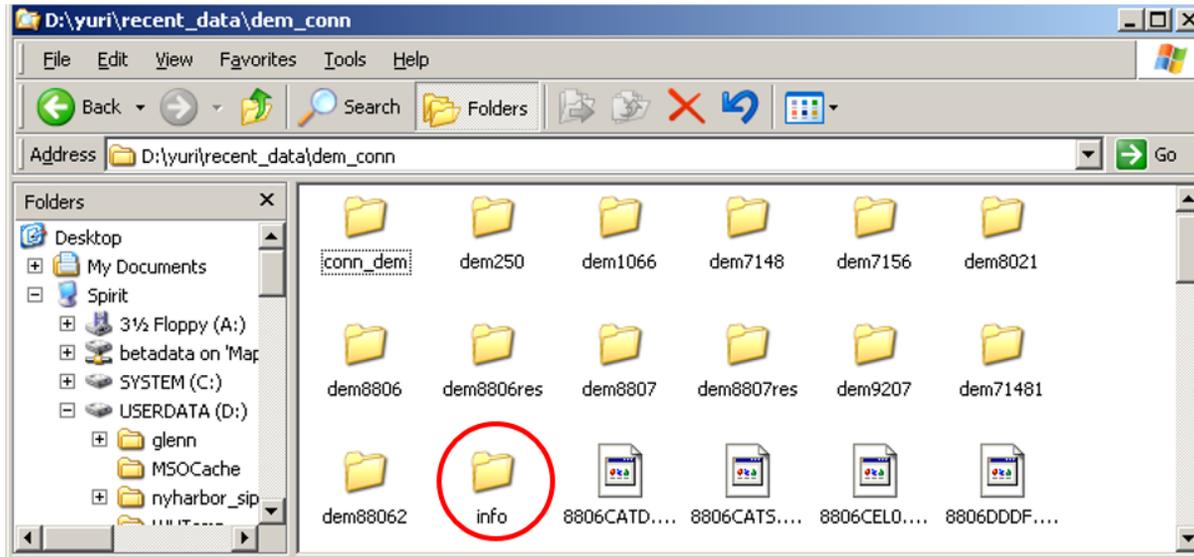
Directory with both **/info** and **/towns** is called “**Workspace**”

Workspace can have multiple coverages and grids; however **/info** directory will be common for all, it will store attribute data of all datasets.

Copying, renaming and deleting datasets can be done only through ArcInfo or ArcCatalog.

NO DRAGGING AND MOVING of these directories!!! You will lose internal links between data and software will not be able to work with them.

EXAMPLE OF FILE MANAGER FOLDER WITH WORKSPACE AND COVERAGES/GRIDS.



(Figure shows the content of Windows folder)

What is the workspace name?

What is the name of the first coverage or grid in the file manager window?

Why do we consider this subdirectory a workspace?

SHAPEFILE FORMAT: most common in general GIS use

Shapefile characteristics:

1. Vector format
2. Have **separate files** for **spatial** and **non-spatial** (attribute table) data
3. Does not follow or maintain topological rules
4. Does not automatically calculate arc lengths and polygon areas
5. Requires projection definition file “.prj” that stores information about spatial reference of your data; you can still view data without it but the software will issue a warning message and you will not be able to use them together with other data properly set with spatial reference.

Example: ESRI shapefile that stores municipal boundaries and is called “towns”.

towns.shp (spatial component, main file)

towns.shx (index file)

towns.dbf (non-spatial component, attribute dBase table, can open with EXCEL)

towns.prj (projection definition (spatial reference) file)

The minimum necessary number of shapefile components for viewing it in ArcGIS is three: .dbf, .shp, .shx.

Files can be ZIPed (WinZip), RARed (WinRAR), dragged and moved together, but **NOT SEPARATED!!!**

When compressing file make sure that ALL relevant files are inside your .zip or .rar file

Especially critical are .dbf, .shp and .shx

More Info on shapefiles: <http://en.wikipedia.org/wiki/Shapefile>

EXAMPLE OF FILE MANAGER FOLDER WITH SHAPEFILE

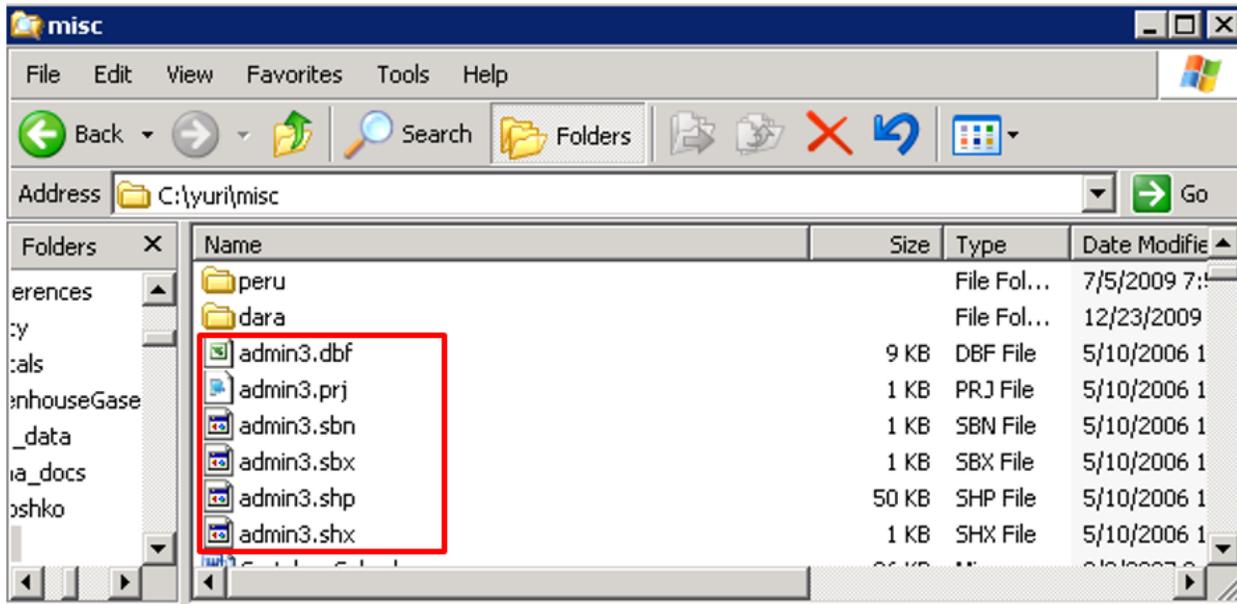


Figure shows the content of Windows folder

INTERCHANGE FILE (aka Export): almost outdated but you might encounter it occasionally on governmental data portals

1. Compressed version of ESRI Coverage or Grid, similar to TAR, ZIP, RAR, etc.
2. Combines in one file (.e00) **spatial** data folders and **attribute tables** from **/info** folder.



Figure shows conversion of coverage into interchange file format

3. Helps transfer COVERAGES or GRIDS easily over the internet
4. Contains only ESRI coverages or grids
5. Has to be IMPORTED to produce back coverage or grid. See below.

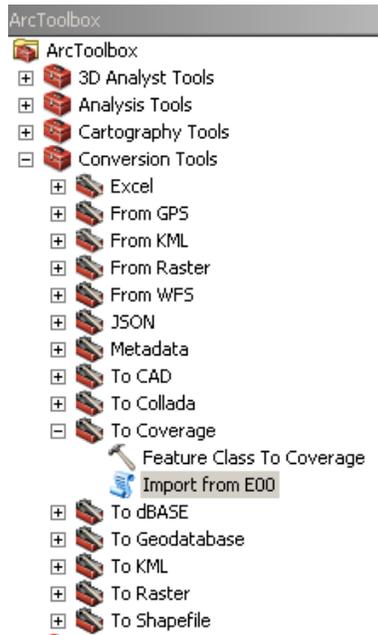


Figure shows ArcGIS ArcToolbox content

Example: ESRI export file that stores municipal boundaries and is called “towns”.

towns.e00

Can be moved, dragged, etc. outside of ESRI interfaces, however it has to be converted back to coverage or grid either via ArcCatalog or ArcMap or ArcTools or ArcInfo

ESRI GEODATABASES: very common data format in GIS community

1. Consist of data imported from variety of sources, such as shapefiles, coverages, grids, text files, images, etc. Like a closet.
2. Compact storage and handling of datasets, easy to transfer
3. Very efficient use for data processing by applying and devising set of rules and properties for all data within them.

File Geodatabases - *Stored as folders* in a file system (.gdb). Each dataset is held as a file that can scale *up to 1 TB* in size.

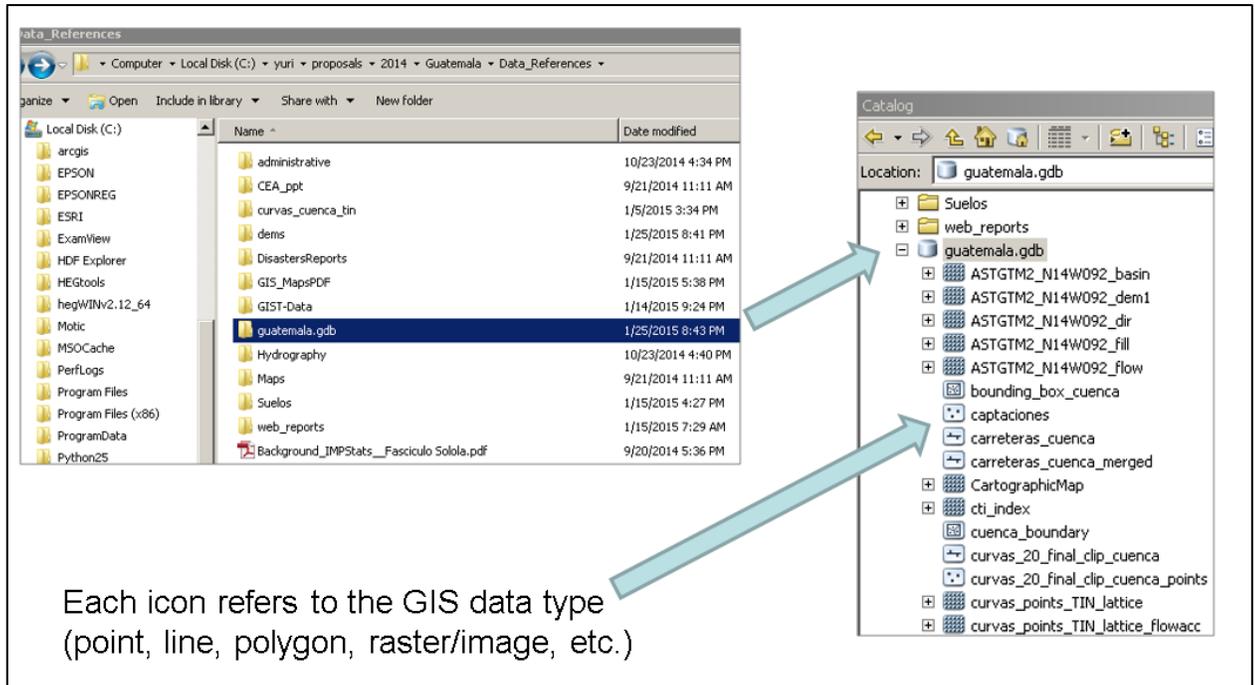
Personal Geodatabases - All datasets are *stored within a Microsoft Access* data file (.mdb), which is *limited in size to 2 GB*. More vulnerable to corruption since it depends on (and only supported by) the Microsoft windows file system management and security; however it is very compact and can be easily transferred as one file.

Additional Resources on ArcCatalog and Managing Databases:

http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/A_quick_tour_of_ArcCatalog/006m000000100000/

GEODATABASE EXAMPLE

Figure shows content of Windows folder and ArcGIS Catalog



Main data formats used by NASA:

Hierarchical Data Format (HDF, HDF4 (.he4), or HDF5 (.he5)) is the name of a set of file formats and libraries designed to store and organize large amounts of numerical data. (**COMPLICATED**)

NetCDF (Network Common Data Form) (.nc) is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. (**EASY**)

GeoTIFF (Geostationary Earth Orbit Tagged Image File Format) (.tif) is a public domain metadata standard which allows georeferencing information to be embedded within a TIFF file. The potential additional information includes map projection, coordinate systems, ellipsoids, datums, and everything else necessary to establish the exact spatial reference for the file. (**EASIEST**)

How .hdf data formats can be viewed and analyzed in ArcGIS?

.hdf can be opened in ArcMap by using **Add Data** button and then selecting a parameter to view. Below is an example of viewing Land Surface Temperature (LST) parameter from the MODIS file from Subdataset Selection menu.

Problems: georeferencing. Often ArcGIS does not use spatial reference correctly. See following slides with solutions for .he4 and .he5.

Solution: Get HEGTool from NASA (free) and convert .hdf file into .tif

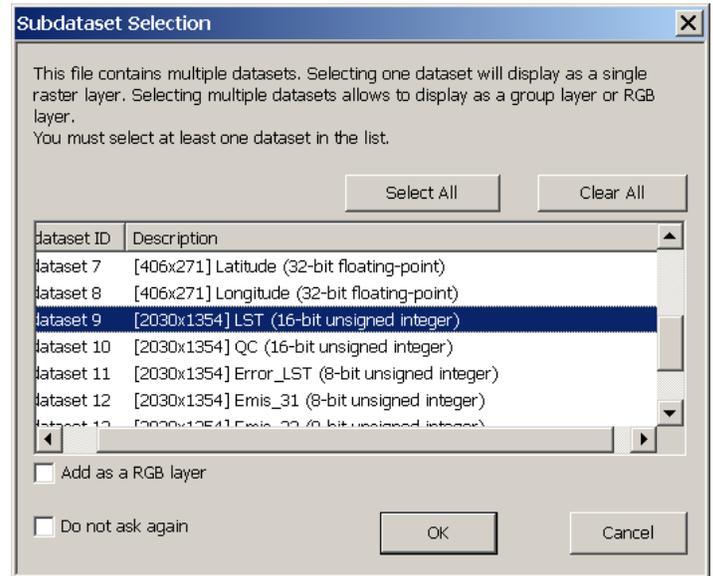


Figure shows Subdataset selection window

NASA offers this method:

<https://disc.gsfc.nasa.gov/information/howto?title=How%20to%20Import%20MERRA%20Surface%20Product%20ata%20into%20ArcGIS>

How .he4 or .he5 data formats can be viewed and analyzed in ArcGIS?

(.he4), or (.he5) files can be opened in ArcMap by changing (rename) these extensions to **.hdf** and then using ArcGIS **Add Data** button and then selecting a parameter to view (see previous slide).

Problems: georeferencing. Often ArcGIS does not use spatial reference correctly.

Solution: But here are few options:

1. Sometimes HEGTool works and you can convert your **.hdf** into **GeoTIFF**.
2. You can use HDFView (HDFExplorer) software (but an old version, older than 2.9) to view data fields and export Lon, Lat and your parameter (e.g. SO₂) into three separate text files. New version (HDFView 2.9) does not allow you to create a “one field” text file. I use HDFExplorer 1.4. Then you can combine all three files together in Excel and finally add this text file to ArcGIS as X,Y Data, convert into shapefile and create a surface using Interpolation (e.g. Spline, IDW, Kriging, etc.)
3. See next slide explaining use of “Warp” tool in ArcGIS

Error: spatial references do not match

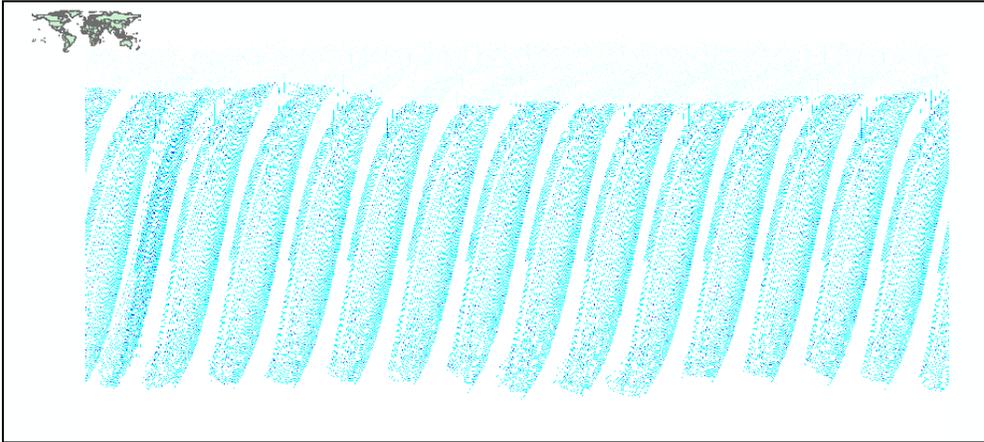


Figure shows spatial mismatch between geographic and satellite data

Spatial reference in .he5 data (in ArcGIS)

[-] Extent	
Top	0.5
Left	-0.5
Right	2879.5
Bottom	-1439.5
[-] Spatial Reference	<Undefined>
Linear Unit	

Solution (provided by Yang Wenli, NASA GES DISC):

Use the “Warp” tool (ArcToolBox → Data Management → Projections and Transformations → Raster → Warp, using the following parameters:

Source control points X/Y = 0,0 => Target control Points X/Y = -180,-90

Source control points X/Y = 2879,0 => Target control Points X/Y = +180,-90

Source control points X/Y = 0,-1439 => Target control Points X/Y = -180,+90

Source control points X/Y = 2879,-1439 => Target control Points X/Y = +180,+90

NASA offers this method:

<https://disc.gsfc.nasa.gov/information/howto?title=How%20to%20Import%20MERRA%20Surface%20Product%20Data%20into%20ArcGIS>

How .nc data formats can be viewed and analyzed in ArcGIS?

Use: ArcToolBox → Multidimension Tools → Make NetCDF Raster Layer.

AURA Data Import in ArcMap

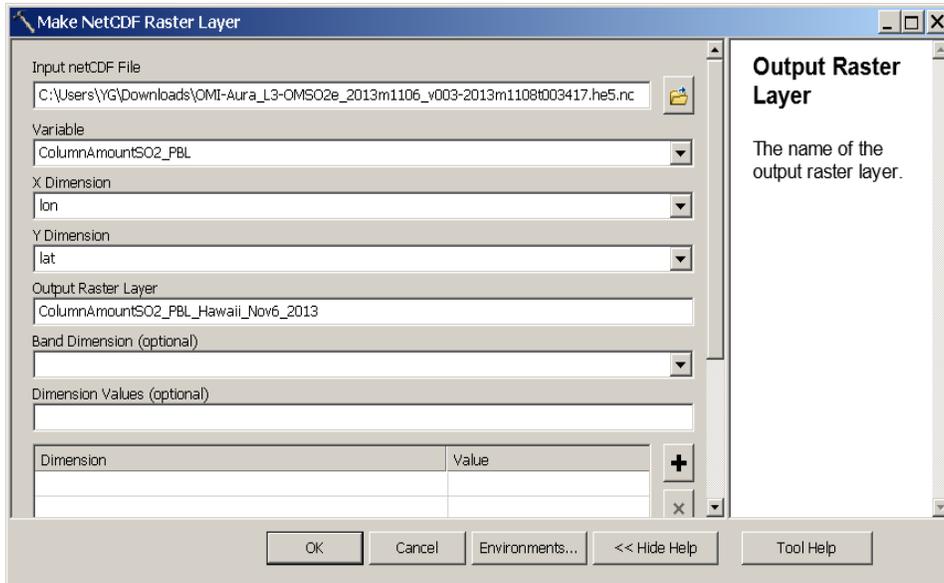


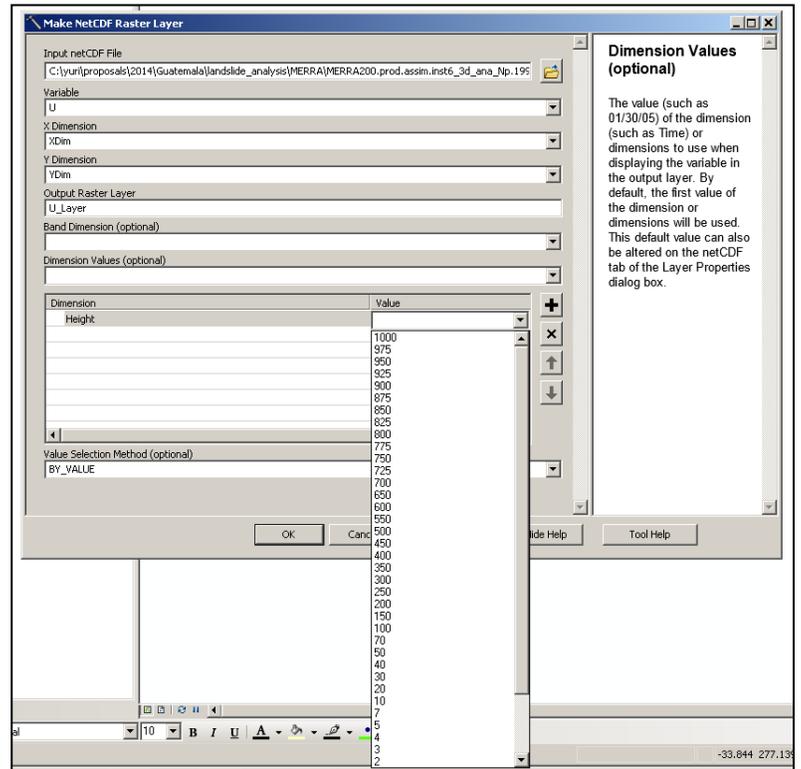
Figure shows Make NetCDF Raster Layer menu

MERRA Data:

MODERN-ERA RETROSPECTIVE ANALYSIS FOR RESEARCH AND APPLICATIONS

This example shows wind data at certain elevation above the land surface. Elevation is expressed in pressure units (mbar).

Figure shows Make NetCDF Raster Layer menu



EXAMPLES OF USEFUL GIS DATA PORTALS:

Karen Payne's famous spreadsheet (very extensive):

<https://docs.google.com/spreadsheets/d/1utQRlrX3IjNiBjWE3rNjLZeTRsbjH-zdixNmXhhvO9Q/edit#gid=63>

The National Map Viewer (US only): <https://www.usgs.gov/core-science-systems/ngp/tnm-delivery/>

National Atlas Data (US only): http://nationalmap.gov/small_scale/atlasftp.html

US Data by state and also International Data:

<https://web.archive.org/web/20180425051725/https://libraries.uark.edu/gis/datalinks.asp>

Census Bureau Cartographic Data: <https://www.census.gov/geo/>

General International and US data: <http://www.mapcruzin.com/>

PORTALS FOR ENVIRONMENTAL DATA ACCESS:

Precipitation:

1. For continuous precipitation data in equatorial and sub-equatorial zones use TRMM (only available until 2015): <http://trmm.gsfc.nasa.gov/>
2. For global precipitation measurements (GPM) use GPM (http://www.nasa.gov/mission_pages/GPM/main/index.html) data (go to <http://disc.sci.gsfc.nasa.gov/>, choose Precipitation and then follow with Data Holding, etc.)
GPM Global Precipitation Measurement: http://www.nasa.gov/mission_pages/GPM/main/
IMERG (Integrated Multi-satellite Retrievals for GPM): <https://pmm.nasa.gov/data-access/downloads/gpm>

Temperature:

1. For continuous global data use Landsat, (Band 10, 11 in Landsat 8):
<https://glovis.usgs.gov/app?fullscreen=1>
2. NASA Surface Temperature Analysis: <https://data.giss.nasa.gov/gistemp/>

Global Topography:

1. Shuttle Radar Topographic Mission (SRTM):
https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-void?qt-science_center_objects=0#qt-science_center_objects

EarthExplorer: <https://earthexplorer.usgs.gov/>

Excellent interface for 90 m SRTM data is: <http://srtm.csi.cgiar.org/>
3. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER):
<https://glovis.usgs.gov/app?fullscreen=1>

Soils:

1. HARMONIZED WORLD SOIL DATABASE:
<http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/>
2. US, NRCS, Web Soil Survey: <http://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx>

Land Cover (many datasets can be found at Earth Explorer or GloVis):

1. Global Land Cover from USGS (30 m): <https://earthexplorer.usgs.gov/>
2. Global Forest Change 2000 – 2015: http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.3.html
3. GlobCover (from ESA, 300 m): http://due.esrin.esa.int/page_globcover.php
4. PALSAR Forest/Non-Forest map (25 m):
http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/fnf_index.htm
5. GlobLand30: <http://www.globeland30.org/GLC30Download/index.aspx>

Climatic Data:

National Climatic Data Center: <http://www.ncdc.noaa.gov/cdo-web/search>

National Weather Service: <http://water.weather.gov/precip/>

New Jersey Climate Data: http://climate.rutgers.edu/stateclim_v1/njclimdata.html

Population Data:

1. The WorldPop project: <http://www.worldpop.org.uk/>
2. Gridded Population of the World: <http://sedac.ciesin.columbia.edu/data/collection/gpw-v3>
3. Global Rural-Urban Mapping Project (GRUMP), v. 1:
<http://sedac.ciesin.columbia.edu/data/collection/grump-v1>
4. Oak Ridge National Lab, Landscan: <http://web.ornl.gov/sci/landscan/>

Stream Data:

USGS, National Hydrography Dataset: <http://nhd.usgs.gov/>

NASA DAACs and GPM Data Sources and their Formats:

DAAC

Distributed Active Archive Centers:

<https://earthdata.nasa.gov/about/daacs>

EOS

Earth Observing System:

<http://eospsso.nasa.gov/>

EOSDIS

Earth Observing System Data Information System:

<https://worldview.earthdata.nasa.gov/>

<https://earthdata.nasa.gov/about>

HDF (data format)

Hierarchical Data Format:

<https://eosweb.larc.nasa.gov/HBDOCS/hdf.html>

NetCDF (data format)

Network Common Data Format:

<https://en.wikipedia.org/wiki/NetCDF>

Where do we go from here?

- As a spatial analyst you need to understand spatial problem and create a:
 - list of themes (e.g. vegetation, geology, infrastructure, historical, etc.)
 - list of data needed for your analysis
 - list of data formats that you can use in analytical software (in our case ArcGIS)
- Data needs can be accommodated for the most part by browsing data portals; this is long and time consuming process that teaches you how to find appropriate data by:
 - Learning which agencies are responsible for the thematic content that you need (e.g. Census Bureau, USGS, NRCS, NOAA, NASA, etc.)
 - Reading documentation (it can be hard but very effective)
 - Using contact information available on data portals (e-mail, phone numbers)
 - Using various listservs, blogs, data user groups, etc.
 - Reading special articles and journals (Google Scholar, Scopus, etc.)
 - Learning how to manipulate online software (hidden menus, broken links, small fonts, small buttons, right-click/left-click, etc.)

Case study: landslides in Guatemala and their relation to local topography and atmospheric circulation

Landslides in Guatemala are a persistent problem during hurricane season and heavy rains, affecting economics, people and infrastructure. One of the problems is low preparedness and risk assessment due to inadequate knowledge about landslide phenomenon specific to the country. Because Guatemala is a mountainous country in the middle between Pacific and Atlantic oceans, landslides have strong dependency on topography and atmospheric circulation. The study conducted by Gorokhovich et al (see **Gorokhovich, Y., E.A. Machado, L.I. Giron Melgar, M. Gahremani. 2016. Improving landslide hazard and risk mapping in Guatemala using terrain aspect. *Natural Hazards*. 81(2), 869-886**) showed necessity to restrict development and use of slopes with south-eastern-southern and south-western aspect.

To conduct the study we need to compile spatial data!

Data sets:

Following is a list of potential useful datasets that we might need to accommodate spatial analyst who is working on any aspect of landslide hazards assessment in Guatemala:

1. Topography (continuous, DEM)
2. Precipitation (continuous, grid)
3. Atmospheric circulation (wind direction) (vector data or continuous)
4. Imagery: aerial photography and satellite data (images)
5. Historical data on landslides, images that can be used to digitize landslides (vector data, images)

Aerial Photography:

Table 1 Description of the data used in this study

Data	Source and description	Description	Date	Spatial resolution
Orthophoto imagery	Produced by the Guatemalan agency INSIVUMEH (Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología) and obtained from the NGO Asociación Vivamos Mejor	Collected in response to hurricane Stan (2005)	2006	1 m

Atmospheric Circulation (wind direction):

Data	Source and description	Description	Date	Spatial resolution
Modern-era retrospective analysis for research and applications (MERRA)	NASA Global Modeling and Assimilation Office web portal: http://disc.sci.gsfc.nasa.gov/daac-bin/DataHoldings.pl	Data represent “NASA reanalysis for the satellite era using a major new version of the Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5)” (http://gmao.gsfc.nasa.gov/GEOS/). We used the dataset “MAI6N-PANA” described as “MERRA 3D analyzed state, meteorology, instantaneous, on pressure levels, at native resolution” containing wind direction components U (eastward) and V (northward) (see NASA OPenDAP portal: http://goldsmr3.sci.gsfc.nasa.gov/opendap/MERRA/MAI6NPANA.5.2.0/)	1979–2015	$0.25^\circ \times 0.25^\circ$

Precipitation (rainfall):

Data	Source and description	Description	Date	Spatial resolution
Tropical Rainfall Measurement Mission (TRMM)	NASA Goddard Earth Sciences Data and Information Services http://mirador.gsfc.nasa.gov/cgi-bin/mirador/homepageAlt.pl?keyword=TRMM_3B42	Daily rainfall estimates of precipitation rates (mm/day)	1997–2014	0.25° × 0.25°

Topography:

Data	Source and description	Description	Date	Spatial resolution
Digital Elevation Model (DEM)	Original DEM created by INSIVUMEH from contour line vector data derived by photogrammetry prior to hurricanes Stan and Mitch. Dataset obtained from the NGO Asociación Vivamos Mejor	Floating point data format. No information about accuracy and precision available. Similar data derived by photogrammetry by US Federal Emergency Management Agency (FEMA, 2005) show a mean error of 0.34 feet and 1.2 feet vertical accuracy	NA	15 × 15 m

Historic Data on Landslides:

Data	Source and description	Description	Date	Spatial resolution
Digitized landslides, derived from orthophoto imagery	Digitized manually, using ArcGIS software and orthophoto imagery collected in 2006	Landslides in Sololá and Lake Atitlán watersheds (Hurricane Stan, 2005). Accuracy close to the orthophoto, i.e., 1–2 m. Total number of landslides in Lake Atitlán watershed: 11,167; total number of landslides in Sololá: 692	2015	1–2 m
Landslides available from USGS study (Bucknam et al. (2001))	USGS web site: http://pubs.usgs.gov/of/2001/ofr-01-0443/	ArcInfo Export format; was converted into shapefile format. According to the USGS report this data were digitized from 1:40,000-scale black-and-white aerial photographs taken between January 14 and March 6, 2000 with an accuracy 15 m (minimum size of the digitized landslide). Number of USGS landslides: 9626.	2000	15 m

Case study: Integrating Coastal Vulnerability and Community-Based Subsistence Resource Mapping in Northwest Alaska

The research sponsored by NOAA included study of the coastal vulnerability of the Northwest Arctic Borough (NAB) to sea-level rise (SLR) and coastal erosion in relation to subsistence resources of selected coastal villages.

Publications:

Gorokhovich Y., A. Leiserowitz, D. Dugan. 2014. Integrating Coastal Vulnerability and Community-Based Subsistence Resource Mapping in Northwest Alaska. *Journal of Coastal Research*, 30 (1), 158–169

Gorokhovich Y., A. Leiserowitz. 2012. Historical and Future Coastal Changes in Northwest Alaska. *Journal of Coastal Research*, 28 (1A), pp. 174-186)

Data sets:

- Coastal underwater slope was obtained from the **National Ocean Service (NOS)** data by **NOAA** at 1:160,000 scale. In 1986, the NOS produced digital maps at 1:250,000 scale in **PDF format**. Using ArcScan, the maps were converted into GIS format (**shapefiles**) and then converted in continuous bathymetric slope surfaces using a GIS interpolation technique.
- Topography (digital elevation models (DEM) is available from **National Elevation Dataset (NED)** by **USGS**.
- Coastal geomorphology at the scale of 1:250,000 (**NOAA**, 2011a); obtained from **Environmental Sensitivity Index (ESI)** data set (**shapefiles**).
- Coastline changes were derived from aerial photos from the 1950s, 1980s, and 2003, which were recently released by **National Park Service (NPS)**:
 - **Gray scale aerial photography** at scale 1 : 43,000 acquired by the U.S. Air Force between 1949 and 1956.
 - **Color infrared** data at scale 1 : 64000 were acquired by the USGS and NOAA as part of the Alaska High Altitude Photography program, between 1978 and 1985.
 - **Natural color orthophoto** data set was acquired and produced by Aero-Metric Inc. in July and September 2003 at scale 1 : 24,000.
- Subsistence resources: was obtained in the field using Community-Based Participatory GIS Mapping



Figure 4. Process of community-based participatory mapping using digital pens.

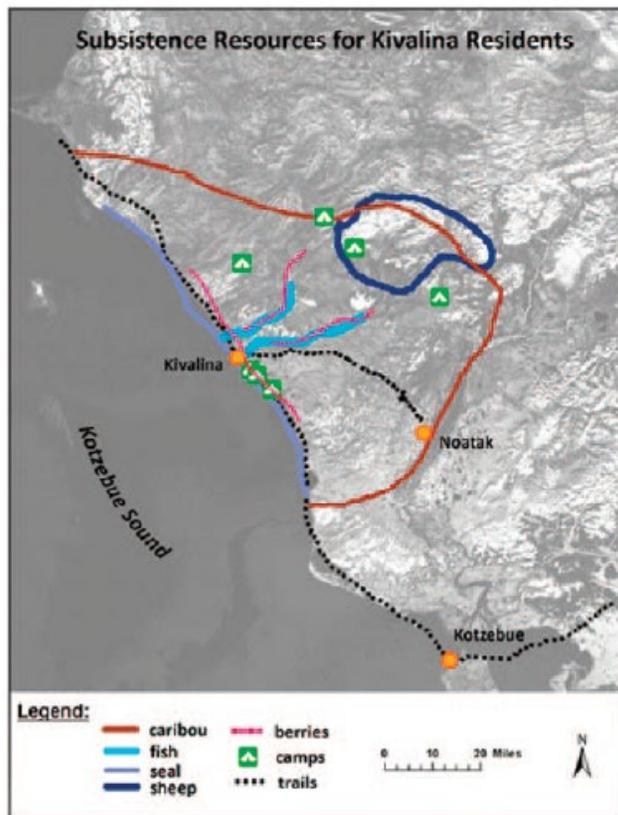


Figure 5. An example of the subsistence resources map for Kivalina village obtained from participatory GIS mapping.