Web GIS data resources: US and global

Author Attribution

Major contributors to this curriculum include (alphabetical):

Maria Fernandez
Michael Hamel
Quentin Lewis
Maili Page
James Peters
Charlie Schweik
Alexander Stepanov

Module Licensing Information

Version 1.0. 

Reviewed by

Quentin Lewis

We live in a priviledged era in which internet access to digital spatial information is increasing everyday! This module is designed to provide a starting point to track down sources of digital geospatial data and attributes for USA and the WORLD. The goal of this module is to point you to a variety of Internet resources we know about.

While the links are all active on this page, note that we provide the full URL of each site so that the student can print this out and keep it as a future reference.

Before you go through this exercise, **think about a particular topic area and location you might want to search for data on**. While we've tried to include many international GIS data websites, some are examples of U.S. repositories. If you are particularly interested in locations outside of the U.S., you should certainly try searching for that location, but in the U.S. based data repositories you may want to try a U.S. location (e.g., Amherst, Massachusetts) to see what the website provides.

NOTE: This exercise references external websites out of our control. A few of the websites referenced appear to be having stability problems or server problems. We are keeping the references in the exercise regardless, with the hope that users of this exercise can successfully access them. But we are not responsible for sites that are not operational at the time that you do this exercise.

Virus Protection

An important note before we begin -- a warning that you should already be aware of. You should not download any data from the internet without some up-to-date virus protection running on your computer. If you do not already have virus protection there are several free products available, including AVG Anti-Virus for windows and Avast! antivirus, which is available for windows and linux. You can also purchase antivirus software such as Norton Antivirus or McAfee VirusScan® Plus.

General Research Using the Internet

First, let's talk generally about approaches to doing research on the Internet. The world wide web has provided the ability to share and find lots of GIS-related data, or at least find where they are stored. But what strategy would you take if you were going to look for GIS-data out on the web? When I ask this question in my face-to-face classes, the typical answer is: "I go to a web search engine, like Google www.google.com, and do a search." And indeed, Google is amazing and certainly works for many of our searching issues. But we sometimes forget that it is only one of many approaches to searching the net. Thanks to the librarians at the University of California at Berkeley, there is an excellent set of web pages that describe a much deeper internet searching strategy. This section will utilize their material.

- **Keyword searches on search engines**: Search engines are web databases that have been built by computer programs called "spiders" that "crawl" the web looking for web pages to inventory. Google has become the most popular of these, although there are others. See http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html for a description and comparison of search engines and to learn more about how search engines work.
    - Note the UC Berkeley librarian's warning: "Google's popularity ranking often makes pages worth looking at rise near the top of search results. Google alone is often not sufficient, however. Less than half the searchable web is fully searchable in Google. Overlap studies show that about half of the pages in any search engine database exist only in that database. Getting a second opinion is therefore often worth your time. For a second opinion, we recommend Teoma, now merged with Ask.com, or Yahoo! Search http://search.yahoo.com/"
- **Subject Directories**: What I find students don't always understand is that there is a distinction between keyword search engines that inventory web pages, and "subject directories" that use humans to organize various pages or sites. For a list of subject directories, see http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SubjDirectories.html, also graciously provided by the librarians at UC Berkeley.
    - One handy hint provided by the UCB librarians: Do a search in Google or other search engines above for a topic and add "subject directory" next to it. For example, if you are interested in Forestry, try "Forestry Search Directory" in the Google keyword line. You may find some specialized directories that are maintained that could be quite useful.
- **Meta Search Engines**: These are search engines that build their databases by searching other search engines. See http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch.html for a list. We tend to agree with the Berkeley librarians and don't recommend the use of these systems. In our view they "free ride" on the hard work of the other search engine organizations.
- **"Deep" or "invisible web" search engines**: There is a huge amount of information on the web that are stored in databases attached to the Internet that search engines like Google and others

cannot read. Deep web search engines are tools that actually provide information on these deeper databases. These searching facilities are less well known. See http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html for a list of search engines that do search these deeper Internet resources.

- **Mechanisms to search listserv or discussion groups**: The Internet has a relatively long history of discussion groups, beginning with the "Usenet" groups that started in the early 1980s. The Google search engine provides a mechanism to search these groups. For example, if you are looking for GIS data relevant to Forestry, for example, you might be able to find a discussion group on this topic, or GIS using this search facility. See http://groups.google.com/. If you find a group that is of interest, you could then read their archives, or possibly send an email specifically to that group. But of course, rules of "netiquette" apply http://en.wikipedia.org/wiki/Netiquette.
    - Try going to http://groups.google.com and search for a group of your interest, such as "Forestry". You may discover discussion groups you were not aware of.
- **Direct e-mail**: This is another way of finding information, but is perhaps your last resort. Using this technique you find via search engines people with expertise in what you are looking for, and then email them directly with a question.
- **Author tracking**: Author tracking is a way of using your Internet browser to scroll through an author's web pages. If you find a web page of interest, you simply edit the active URL in the Browser address bar to "go up one level" and see if there are other related pages that aren't specifically linked to the one you are on.
- **Traditional methods**: on-line journals, library indices: Of course, there are a variety of traditional search approaches found at various libraries that shouldn't be forgotten as a resource!
- Lastly, the UC Berkeley library folks provide a nice document describing their recommendations for using all of these tools. See http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html for a five-step overview on Internet searching strategies.

Searching Specifically for GIS Data on the Internet

Because many organizations need similar datasets, there has been a special emphasis on the development of web based repositories for searching and finding geographic data. Crucial to this effort is the concept of "metadata."

What is metadata?

- Metadata is often defined as "data about data." Metadata is essential because it is information about a particular dataset: things like who created it, when was it created, why was it created, rules for its use, what time point(s) it represents, etc. Note that metadata is important for geographic and nongeographic datasets. It would be impossible to share data without metadata. Keep this in mind from now on. Documenting your data will be a key part of your GIS work.
- In the United States, the Federal Geographic Data Committee (FGDC) has made a major effort to develop standards for the development of GIS metadata.

- The GeoData Alliance also works on enabling the creation, flow, and beneficial use of geographic information.
- Other organizations and companies have developed tools that have become standards. Galdos Systems, Inc., for example, has become a leader in Geography Markup Language (GML), an XML application that is becoming the world standard for geographic information delivery over the Internet.
- In the USA, collection of metadata is often required for agencies or research programs that receive federal funding.

GIS Related Consortiums, Clearinghouses and Portals

There a numerous web sites that provide access to metadata (and often freely shared datasets themselves). The gradual existence of metadata is providing the capacity to search the web for GIS data much more easily. There is a collection of internet servers responsible for storing and sharing metadata listed below.

*** The Federal Geographic Data Commission (FGDC) Clearinghouse network**

- The Federal Geographic Data Committee coordinates the sharing of geographic data, maps, and online services through an online portal for domestic data (US): Geospatial One-Stop, that searches metadata held within the NSDI Clearinghouse Network. Take a few minutes to explore this website and search for data. For example, you might want to try searching for topographic information on Amherst, MA USA.
- The Clearinghouse Network is a community of distributed data providers who publish collections of metadata that describe their map and data resources within their areas of responsibility, documenting data quality, characteristics, and accessibility (US and International data). You can try to find some data, but **DO NOT** search all clearinghouse servers. Just choose a couple. For example, you might try to see what kinds of data are available for Africa, by selecting the "African Data Dissemination Service" and/or the "African Geospatial Data Sets" option.
- Global Spatial Data Infrastructure Gateway makes access to geospatial information possible as well as making it possible to publish your own geospatial data. NOTE: We had some periodic problems accessing this server]
- The Global Spatial Data Infrastructure Association "...is an inclusive organization of organizations, agencies, firms, and individuals from around the world. The purpose of the organization is to promote international cooperation and collaboration in support of local, national and international spatial data infrastructure developments that will allow nations to better address social, economic, and environmental issues of pressing importance." There are news letters and discussion lists, including ones focusing in on particular regions of the world, such as Africa.

*** Geography Network**

Go to Geography Network where you can find many types of geographic content including dynamic maps, downloadable data, and more advanced Web services. It is a service from the commercial GIS software company ESRI.

- You can enter a key word (Amherst, MA, for example) and search available data in a given category (e.g. environmental).
- Explore the ways in which you can access information clicking in the tabs: web services, data, maps, etc. Have a look at the option of getting "dynamic data and maps". This is one direction GIS is moving -- utilizing data that is maintained by the data owner and accessing it in real time over the Internet.

*** Other Portals and GIS resources sites**

The below provides a list of other portals and GIS resource sites that we are aware of.

- The Center for Spatially Integrated Social Scienceshttp://www.csiss.org/
- The GIS Data Depot http://data.geocomm.com. Has data for US states and counties (some of this data is free, some must be purchased)
- The ESRI guide to GIS http://www.gis.com
- The GIS Portal http://www.gisportal.com/
- GIS and remote sensing educational materials and tutorials http://www.photogrammetry.ethz.ch/general/persons/jana/isprs/ed_material.html
- UMASS Open GIS curriculum project link page with quick access to the most important open source GIS programs.
- Massachusetts GIS information about GIS around the world http://www.mass.gov/mgis/giswww.htm
- University of Edimburgh GIS WWW resource list http://www.geo.ed.ac.uk/home/giswww.html
- University of Texas's useful Internet sites for GIS and water resources http://www.ce.utexas.edu/prof/maidment/gishydro/docs/websites/othr_web.htm
- USGS GIS data for water resources http://water.usgs.gov/maps.html
- WMS Sites, a catalog of public web mapping service resources online. http://wms-sites.com/

GIS-US Federal Agency websites

There are lots of GIS and map sources available using US Federal websites, although most data is now accesible through the Geospatial One-Stop that we visited earlier. Some of those sites keep data on other countries in addition to the US.

Take some time to take a look at these sites, and perhaps try searching for something of interest to you.

- http://seamless.usgs.gov/ Seamless Data Distribution System, Earth Resources Observation and Science (EROS). Here you can download seamless US and global data for a given geographical area that you select interactively. It is VERY useful to obtain elevation data (DEM) at different resolutions, as well as land cover, transportation, hydrography, orthoimagery, boundaries and national atlas data.
- http://redhook.gsfc.nasa.gov/%7Eimswww/pub/imswelcome/ The NASA Earth Observing System Data Gateway. A portal for U.S. NASA data.

- National Geophysical Data Center
- http://edcsns17.cr.usgs.gov/EarthExplorer/ The USGS EarthExplorer site. A tool to query and order satellite images, aerial photographs, and cartographic products through the USGS. As an example, this is how you would do a search on available Landsat satellite images for the Amherst, MA region:
    - Enter as GUEST
    - Click on define on map
    - Zoom in to your geographic area of interest.
    - Select the types of datasets you are interested in. Let's just select Landsat for this exercise.
    - Press continue
    - Select the acquistion date or temporal range you are interested in.
    - Press search
    - Click on the hyperlink that is returned to see the list. You can browse available images, look at their previews and footprints, and details about their quality.
- http://www.census.gov US Census Bureau website. Note the "Geography" option and the reference to "Tiger files". These are GIS layers related to census data (such as various census geographic boundaries).
- http://www.nationalatlas.gov/atlasftp.html The U.S. National Atlas. Provides abundant US data, for example:
    - Agriculture: Agriculture Census 2002 - Crops, Expenses, Farmland...
    - Biology: Bat Ranges, Butterflies, Forests, Invasive Species, Land Cover...
    - Boundaries: Congressional Districts, Counties, Federal lands, States...
        - This is one location where students in the US can get congressional boundaries in GIS format.
    - Climate: Precipitation, Hazard Events, Hurricanes, Sea Temperature...
    - Environment: Air Releases, Hazardous Waste, Toxics Release...
    - Geology: Earthquakes, Landslides, Shaded Relief, Volcanoes...
    - History: Presidential General Election 2000 County and State Results.
    - Map Reference: Cities and Towns, Urban Areas...
    - People: Census, Crimes, Energy Consumption, Mortality...
    - Transportation: Airports, Parkways and Scenic Rivers, Railroads, Roads...
    - Water: Aquifers, Dams, Watersheds, Streams and Waterbodies...

Nonspatial Data Clearinghouses

Non-GIS data can be linked as secondary data to any GIS layer, enhancing the power of GIS analysis. In other words, you may often have a need for some data that are not currently associated with a GIS layer, but could be. For example, suppose you were interested in summary statistics for a certain public health problem for each U.S. State. If you could find that summary statistical data in some kind of tabular or spreadsheet format and having for each row its associated state, you could "join" this data to a polygon layer of US states. Once done, you could do various queries looking at the geographic pattern of this issue across the U.S. at the state level of aggregation.

Some examples of useful websites for data sets that are not necessarily GIS layers:

- Social and economic statistics
    - Statistics and Social Science Data at the University of Virginia

http://fisher.lib.virginia.edu/collections/stats/
    - FEDSTATS: The U.S. Federal Government main statistics portal. More than 100 agencies contribute their data to this site. http://www.fedstats.gov
    - The US Census Bureau site http://www.census.gov
- Biodiversity
    - The Global Biodiversity Information Facility http://www.gbif.org
    - HerpNet http://herpnet.org
    - The Mammal Netowrked Information System at UC Berkeley http://manis.mvz.berkeley.edu/pres/PresentationServlet?action=home
    - The Ocean Biogeographic Information System http://www.iobis.org/Welcome.htm
    - The Ornithological Information System http://ornisnet.org/
    - The System-wide Information Network for Genetic Resources http://singer.grinfo.net/. A database of crop, vegetable and agroforestry information.
    - The United National Environment Program http://na.unep.net/datasets/datalist.php

Global Data

For students interested in global datasets or data about countries outside of the U.S., we have found a couple useful sites (although we have referenced others already, and certainly there are more we are not aware of).

- The Global Spatial Data Infrastructure Association (http://www.gsdi.org/) is an organization of organizations around the world that promotes the development of spatial data infrastructures. Browse their site to access some of their http://www.gsdi.org/Electroninc%20Gateways.asp "gateways" to global data.

There are also sources of GIS (and useful secondary global data) on various sustainability and environmental topics by the:

- World Resources Center (http://pubs.wri.org/datasets.cfm?SortBy=1)
- The United Nations Environment Programme (http://www.grida.no/db/gis/prod/html/toc.htm), and
- FAO's Geonetwork (http://www.fao.org/geonetwork/srv/en/main.search).

The rest of this section, presents information about some common categories of data available at global scales:

Elevation Data

- Browse through the variety of elevation products provided by the USGS at http://edc.usgs.gov/products/elevation.html.
- SRTM3 (Version 2) (ftp) is one of the primary sources of gridded elevation data with global coverage. The resolution is 3 arc-second or roughly 90 meter. These data were obtained by the Shuttle Radar Topography Mission (SRTM) which used Synthetic Aperture Radar (SAR) to measure all land surfaces between +/-60 degrees of latitude. They provide a globally consistent base dataset for standard analyses that are terrain-dependent. The SRTM data suffer from data

gaps in higher-relief areas, so one could patch and fill them with data from SRTM30 (30 arc-second or approximately 1km resolution), itself a blend of GTOPO30 and SRTM data. A complete assessment of the SRTM Topographic Products (pdf) has recently been released and provides important information on both absolute and relative height and locational errors in the SRTM3 data.

- However, recall that you can get seamless digital elevation data for the United States and some other locations as well at http://seamless.usgs.gov. This is very nice because you don't have to mosaic data together and the seams between datasets have been delt with.

Hydrography

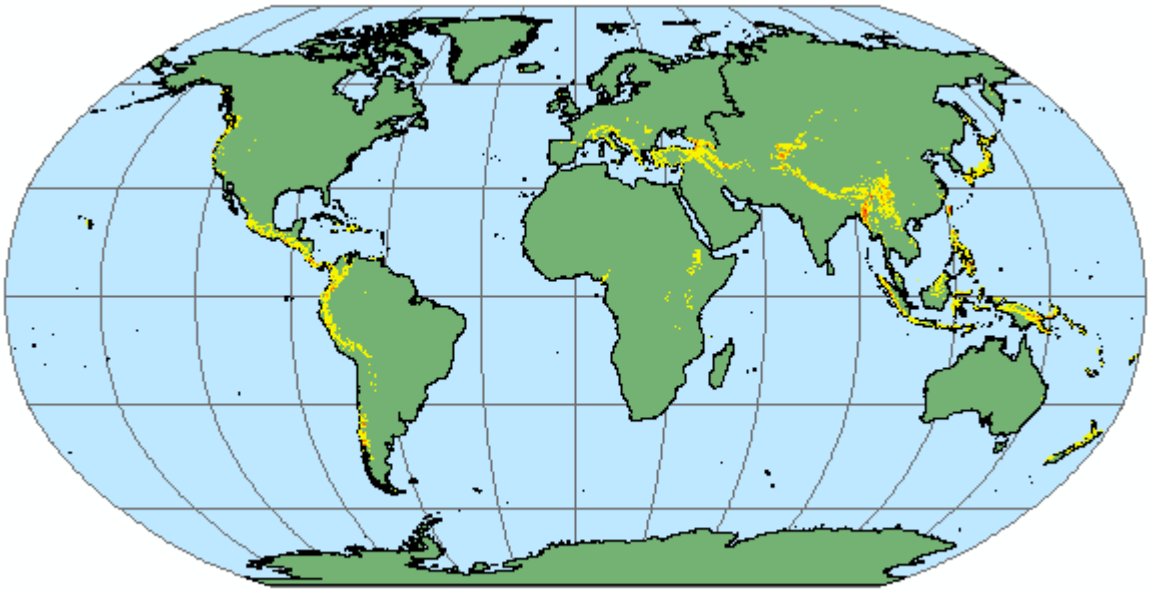
- The World Wildlife Fund's "Hydrosheds" project. http://www.worldwildlife.org/freshwater/hydrosheds.cfm. "Hydrosheds" stands for "Hydrological data and maps based on Shuttle Elevation Derivatives at multiple scales." These are data derived from the Space Shuttle topographic mission.
- The USGS watershed HYDRO1K project. http://edc.usgs.gov/products/elevation/hydro1k.html This dataset contains all watersheds on a nested basis down to the resolution obtainable with 1km digital elevation grids. One could query HYDRO1K to find the location that a specific point occupies, then use the special topological encoding of the HYDRO1K basins to additionally select all upstream basins, if any. These selected basin polygons then represent the entire watershed for the site.

Remote-sensing Imagery

- The Global Land Cover Facility (GLCF)] at the University of Maryland (http://glcf.umiacs.umd.edu/index.shtml) maintains multi-decadal global coverage of Landsat (http://landsat.gsfc.nasa.gov/) data. A path/row pair is needed to retrieve each image. Landsat 7 offers recent (1999 and later) coverage of all land surfaces.
- Google Earth http://earth.google.com/ and some other sources on the internet provide high-resolution (approximately 1-meter) satellite imagery of some urban areas, a great resource for visual inspection.
- Some of the earlier sites such as the USGS Earth Explorer provide mechanisms to find remotely sensed images as well.

Hazards Data

- The Center for Hazards and Risk Research (http://www.ldeo.columbia.edu/chrr/research/hotspots/coredata.html at Columbia University provides interesting global data sets of cyclone, drought, earthquake, flood, landslide and volcano hazard frequency and distribution, and also cyclone, drought, earthquake, flood, landslide and volcano associated mortality risks. Here is an example of global landslide data obtained from their website:



Climate data

- The Global Historical Climatology Network (GHCN) version 2 (http://www.ncdc.noaa.gov/oa/climate/climatedata.html#monthly provides station-based monthly precipitation totals for about 10,000 stations worldwide among other climate data.
- Willmott et al. (http://climate.geog.udel.edu/%7Eclimate/html_pages/archive.html have used these and other climate information to develop a global terrestrial optimally-interpolated half-degree resolution monthly timeseries grid. They provide global data in the form of monthly air temperature, total precipitation, terrestrial water budgets and moisture indices. They also have additional climate datasets for the Arctic, Tropical and South American regions.
- The World Meteorological Organization (WMO) maintains a database of selected long-term daily precipitation records at http://www.wmo.ch/web/gcos/gcoshome.html
- The U.S. National Climate Data Center is another resource http://www.ncdc.noaa.gov/oa/ncdc.html
- Finally, WorldClim (http://www.worldclim.org/ is a set of global climate grids with a spatial resolution of a square kilometer. The data are described in: Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978. A pdf of this paper can be found at http://www.worldclim.org/worldclim_IJC.pdf.

Population data

The World Data Center for Human Interactions in the Environment at CIESIN (http://www.gateway.ciesin.org/wdc/) provides population data and administrative boundaries. Some

examples of data sets are a 5X5 minute (long/lat) gridded population of the world and World Bank data collection of 170+ countries, targeting the social effects of economic development.

Country data

Last, but not least, the dataserver diva-gis (http://biogeo.berkeley.edu/bgm/gdata.php) provides a search mechanism for data by country around the world.

Conclusions

The intention of this module is to give you an overview of many of the GIS data repositories out on the the Internet. These are constantly changing, the information in the module is current as of October 2006. If we are missing an important GIS data resource, we welcome you to contact us to let us know what it is.

Note that this module only provided an opportunity to review and explore these sites. Many of the datasets that are available in these sites are in various data formats.