

# UC Santa Barbara

## GIS Core Curriculum for Technical Programs (1997-1999)

### Title

Unit 1: Data Acquisition

### Permalink

<https://escholarship.org/uc/item/75f4d2cr>

### Authors

1, CCTP  
Willett, Karen B.

### Publication Date

1998

Peer reviewed

# UNIT 1: DATA ACQUISITION

Written by Karen Beardsley Willett, University of California, Davis

---

## Context

Acquiring data over the Internet has become common practice over the past few years. Numerous web sites provide free data layers on the web, and most provide adequate metadata (information about the data) for users to determine whether or not the data are appropriate for their needs. One must be cautious, however, when acquiring free data over the Internet. Many data formats, compression methods, and archive media are used for distributing and accessing GIS data sets.

This Unit will focus on methods for acquiring GIS data over the Internet, including understanding Internet data transfer methods and developing skills in manipulating a variety of data formats.

The following is an example of using a combination of ftp, tar, and gzip tools to retrieve and manipulate GIS data over the Internet.

---

## Example Application

**A small environmental consulting firm has hired a GIS technician to acquire basic data layers for their state. The company first requires data at a coarse scale (statewide), and later will want to obtain more detailed, finer scale data sets for particular regions in the state. Since the USGS makes numerous GIS data layers available on the Internet, several of these base layers can easily be retrieved by the GIS technician through use of a web browser. The technician proceeds to acquire data layers for Digital Elevation Model (DEM) data, transportation data, and Land Use/Land Cover data. In order to accomplish this task, the GIS technician follows several steps:**

- 1. From the Netscape browser, open the USGS page <http://www.usgs.gov/>**
- 2. Under Science Topics, click on Mapping**
- 3. Click on US GeoData under Mapping Products and Services**
- 4. Under 1:250,000-Scale Digital Elevation Model (DEM), review the Condensed User Guide**

5. Click on FTP via State
6. Using FTP, download one data set in Compressed format
7. Use GZIP utility (GUNZIP) to uncompress data
8. As indicated in the instructions, review and download the User Guide from  
*[outdated link removed]*
9. Convert the data from DEM format into the GIS software used by the company

**The technician obtained other base layers using a similar method. Some data sets are acquired from public agencies via tape or disk, and these are processed in a similar manner once extracted from the transfer media.**

---

## Learning Outcomes

The following list describes the expected skills which students should master for each level of training, i.e. Awareness/Competency/Mastery.

### Awareness:

The expected learning goals of this section are to achieve a general understanding of the Internet tools available for accessing and transferring GIS data files; gain a basic vocabulary of file formats, compression methods, and transfer protocols; and recognize and be able to utilize various media sources for data transfer.

### Competency:

The learning goals of this section are to develop the ability to connect to remote computers, locate desired data sets, and retrieve data of various formats over the Internet or from archive media.

### Mastery:

The learning goals of this section are to develop the ability to prepare data sets for transfer to other users and to understand interoperability issues relating to shared access of GIS data over the Internet.

---

## Preparatory Units

Recommended:

None

Complementary:

1. Unit 2 - Locating Demographic Data
2. Unit 3 - Locating Transportation Network Data
3. Unit 4 - Locating Land Records Data
4. Unit 5 - Locating Natural Resources Data

5. Unit 6 - Locating Terrain Data
  6. Unit 7 - Using and Interpreting Metadata
- 

## Awareness

### Learning Objectives:

1. Student can use web browser to find data.
2. Student can define basic vocabulary relating to the Internet and data transfer methods.
3. Student is able to retrieve data over the Internet using File Transfer Protocol (ftp).

### Vocabulary:

- anonymous ftp
- client
- file transfer
- ftp (file transfer protocol)
- internet
- internet protocol
- JAZ disks
- magnetic tape (4mm, 8mm)
- protocol
- server
- TCP/IP
- web browser
- world-wide web (www)
- ZIP disks

### Topics:

1. Unit Concepts
  - The Internet is a federation of computer networks that speak the same protocols ("language"). The networks are connected to each other with high-speed telephone circuits. The protocols spoken are computer networking protocols, used to enable computers to communicate with each other (just as human protocols help humans communicate). Three roles played on the Internet: Information provider, User (or customer) provider, and Connection provider.
  - There are three basic methods for accessing the Internet: Modem, dial-up networking, and high-speed telephone circuits leased from the phone company.
  - There are a number of different ways to find and acquire existing digital data over the Internet. The basic requirement for doing so is an internet browser and a relatively fast connection. (For example, a 14400 speed modem from home is usually slower than users are willing to put up with.)

## 2. Types of Connectivity and Data Transfer

- FTP (File Transfer Protocol)

FTP is a utility which transfers files to and from TCP/IP networked computers. TCP/IP (Transmission Control Protocol/Internet Protocol) consists of a suite of standardized protocols for networking computers. FTP is one protocol in this suite. Other significant protocols within TCP/IP are telnet, SMTP (Simple Mail Transfer Protocol), and NFS (Network File System). The Client portion of Unix FTP lets users access file systems on a remote computer. The Server portion of Unix FTP lets users on remote computers access local computer files.

Using FTP, you can access files on a remote computer and perform common operations, such as list and change directories, transfer files, create directories, rename files and directories, and delete files and directories. Once you have launched the FTP utility, a connection is established to the remote computer. You are then free to transfer files to and from the remote system using FTP commands. The connection to the remote system remains in effect until terminated by the user. Unix FTP supports both local help for FTP commands via the man pages, and remote help, which displays FTP commands and options available on the remote system. A list of ftp commands will be displayed if one types first ftp, then "?" at the unix prompt:

```
% ftp
ftp> ?
```

You will see a [list of FTP Commands](#)

- Tape Storage Media

Numerous media types are commonly used today. Some examples include 4mm tapes, 8mm tapes, IOMEGA Zip/Jaz drives and disks, and many others. Zip disks hold 100MB, while Jaz disks hold 1 Gigabyte of information. Keep in mind that it is best not to compress data on any media since retrieval on another machine may be problematic.

Whenever you receive data from another users, you must ask the data provider how data were loaded and what command is required to retrieve information from medium. You should also make sure that you have the resources to read the particular media type before having the information send using that medium. Floppy disks (1.4 MB) are still used to transfer information, and often data are compressed, or "Zipped" using the winzip utility.

### Tasks:

- Use the Internet to search for GIS data for your state.
  - Hint: Use an Internet search engine, such as
    - Excite
    - Lycos

- Yahoo
- Infoseek

and look for help on the best searching strategies.

- Download several relevant data sets that you discover during the data search.

## Competency

Learning Objectives:

1. Student can use the Internet to connect to a remote computer.
2. Student can find and retrieve relevant data layers.
3. Student is able to process data from numerous formats, including GIS export files, tar files, and compressed files.

Vocabulary:

- download
- gzip, zip
- tar (tape archive and retrieval)
- telnet
- upload
- vt100

Topics:

1. Unit Concepts
  - Telnet

Telnet is a software program used for logging onto other computers on the Internet. It allows you to sit at a keyboard connected to one computer and log onto a remote computer across the network. The connection can be to a machine in the same room, on a local network, or on the other side of the world. When you are connected, it is as if your keyboard were connected directly to the remote computer. You can access whatever services the remote machine provides to its local terminals. You can run a normal interactive session or access special services provided only on that computer. When displaying GIS and other graphic information, you must be sure to set the display environment to be that of your local computer. Otherwise, your graphics window will open on the host computer, which could cause problems for those actually sitting at and using the host computer!

The telnet command (UNIX) launches the telnet application, and the argument

you must supply is the domain name of the computer to which you want to connect (remote computer).

For example, if you wish to connect to domain monet.ucdavis.edu from niko.dfg.ca.gov (using UNEX c shell) the commands would be:

(Note: % is used to indicate the UNIX prompt)

```
% telnet monet.ucdavis.edu
```

```
Trying 128.120.4.10...
Connected to monet.ucdavis.edu.
Escape character is '^'].
```

```
ULTRIX V4.3A (Rev. 146) (monet)
login: user1
Password: passwd$$
monet% setenv DISPLAY niko.dfg.ca.gov:0.0
monet% arc
```

Copyright (C) 1982-1997 Environmental Systems Research Institute, Inc.  
All rights reserved.  
ARC Version 7.1.2 (Wed Aug 13 07:45:00 PDT 1997)

This software is provided with RESTRICTED AND LIMITED RIGHTS. Use, duplication, and disclosure by the U.S. Government are subject to restrictions as set forth in FAR Section 52.227-14 Alternate III (g)(3) (JUN 1987), FAR Section 52.227-19 (JUN 1987), and/or FAR Section 12.211/12.212 [Commercial Technical Data/Computer Software] and DFARS Section 252.227-7015 (NOV 1995) [Technical Data] and/or DFARS Section 227.7202 [Computer Software], as applicable. Contractor/Manufacturer is Environmental Systems Research Institute, Inc., 380 New York Street, Redlands, CA 92373-8100, USA.

```
Arc: &term 9999
Arc: ae
```

Copyright (C) 1982-1997 Environmental Systems Research Institute, Inc.  
All rights reserved.  
ARCEDIT (COGO) Version 7.1.2 (Wed Aug 13 07:45:00 PDT 1997)

```
Arcedit: display 9999
```

This brings up the display windows on your local host, and you are ready to proceed using Arcedit. As long as your display environment is set properly, ArcView may also be run remotely from telnet.  
A number of different Telnet programs are available for Macs and PCs. Some of these are public domain, and can be obtained directly over the Internet.

>From a Unix system, a list of telnet commands can be obtained by entering the following commands:

```
% telnet
telnet> ?
```

You will see a [list the telnet commands](#)

- Compression Utilities and Export Formats
  - Tape Archive and Retrieval (tar)
  - ZIP, GZIP
  - Export files (.e00 for ARC/INFO)

Tasks:

- Using gzip, zip, tar, and other methods, process the data sets retrieved earlier into a form usable by your GIS program.

## Mastery

Learning Objectives:

1. Student understands basic concept of interoperability.
2. Student is able to work with Java, Active X, and other Internet mapping tools.

Vocabulary:

- Java
- Active X
- Interoperability
- Internet Map Servers

Unit Concepts:

1. Java

The Java programming language and environment is designed to solve a number of problems in modern programming practice. Java started as a part of a larger project to develop advanced software for consumer electronics. These devices are small, reliable, portable, distributed, real-time embedded systems. When they started the project they intended to use C++, but encountered a number of problems. Initially these were just compiler technology problems, but as time passed more problems emerged that were



best solved by changing the language.

It is commonly thought of as a way to make Web pages sexy -- incorporating stock tickers, sound or video into Web pages. It has evolved into much more. It is becoming known as a computing platform -- the base upon which software developers can build applications. Developers can build a variety of applications using Java -- traditional spreadsheets and word processors in addition to mission critical applications used by the biggest companies: accounting, asset management, databases, human resources and sales.

Java applications, or applets, are different from ordinary applications in that they reside on the network in centralized servers. The network delivers the applet to your system when you request them. For example, let's say that you want to check your personal financial portfolio. You'd dial in to your financial institution and use your Web browser to log into the bank's system. The portfolio data will be shipped to you along with the applet needed to view it. Let's assume that you're considering moving your money from one account to another. No need to perform a series of cut-and-paste exercises. The system will also send you an applet that will allow you to change the rate of interest and length of investment to perform a series of "what-if" scenarios.

>From the corporations' point-of-view, Java will simplify the creation and deployment of applications thus saving money. Applications created in Java can be deployed without modification to any computing platform, thus saving the costs associated with developing software for multiple platforms. And because the applications are stored on centralized servers, there is no longer a need to have people insert disks or ship CD's to update software.

## 2. Active X

ActiveX is Microsoft technology similar to Java but integrated fully for Microsoft products only.

For more information, read about the Microsoft Component Object Model (COM).

## 3. Internet Map Servers

### ◦ MapObjects Internet Map Server

Designed for Windows developers, MapObjects Internet Map Server extends the power of MapObjects to serve maps over the Internet. Applications built with the MapObjects Internet Map Server extension can access spatial data formats supported by MapObjects such as shapefiles, coverages, SDE layers, and many graphic images. In addition, MapObjects Internet Map Server includes a Web server extension that works with Netscape Server and Microsoft Internet Information Server. The Web server extension provides a unique framework for request management and load balancing that provides fast, efficient, and scaleable map serving capability.

For more information, see the [demos](#) of MapObjects Internet Map Server.

### ◦ ArcView Internet Map Server

With the ArcView Internet Map Server extension, you can use ArcView GIS out-of-the-box to put mapping and GIS applications on the Internet. ArcView Internet Map Server makes publishing maps on the Web almost as easy as printing a map. It includes a built-in setup wizard and a ready-to-use Java applet to help you publish your data quickly. Interactive maps can be created from a number of different types of spatial data including shapefiles, coverages, SDE layers, DWG, DXF, DGN, and a variety of graphic images. ArcView Internet Map Server works with Netscape Server, Microsoft Internet Information Server and other server products that support ISAPI/NSAPI web server extensions. ArcView Internet Map Server is scheduled to start shipping in the third quarter of 1997 and requires ArcView GIS Version 3.0 running under Windows 95, Windows NT, or UNIX.

For more information, see the demos of ArcView Internet Map Server.

- Implications for GIS users

Numerous java based mapping applications have been built. Interactive mapping over the Internet is possible by use of other methods, including Perl scripting and HTML programming. ESRI Internet Map Server for ArcView has a built in Java Applet (called Map Cafe) that allows Internet manipulation of GIS data layers in a manner that mimics ArcView. Map Objects Internet Map Server may be implemented using either Java, Active X, or neither (simple html coding).

Tasks:

- From the demo sites listed above, test out several Internet Map Server applications.

---

## Follow-up Units

1. Unit 9 - Converting digital spatial data between formats, systems and software
- 

## Resources

*[outdated links have been removed]*

- *Internet for Beginning Users*, Bureau of Land Management, National Training Center, December 1996.
- *Managing Internet Information Services*, Cricket Liu, Jerry Peek, Russ Jones, Bryan Buus, and Adran Hye, O'Reilly & Associates, 1994.
- Java in a Nutshell
- The Java Language Environment, A White Paper from May of 1996 by James Gosling

and Henry McGilton.

- So what exactly is Java?by Jason English
- Microsoft Component Object Model (COM)

---

*Created: May 14, 1997. Last updated: October 5, 1998.*

**FTP Commands**

!	debug	mget	pwd	status
\$	dir	mkdir	quit	struct
account	disconnect	mls	quote	system
append	form	mode	recv	sunique
ascii	get	modtime	reget	tenex
bell	glob	mput	rstatus	trace
binary	hash	newer	rhelp	type
bye	help	nmap	rename	user
case	idle	nlist	reset	umask
cd	image	ntrans	restart	verbose
cdup	lcd	open	rmdir	?
chmod	ls	prompt	runique	
close	macdef	proxy	send	
cr	mdelete	sendport	site	
delete	mdir	put	size	

## TELNET Commands

close	close current connection
display	display operating parameters
mode	try to enter line or character mode ('mode ?' for more)
open	connect to a site
quit	exit telnet
send	transmit special characters ('send ?' for more)
set	set operating parameters ('set ?' for more)
unset	unset operating parameters ('unset ?' for more)
status	print status information
toggle	toggle operating parameters ('toggle ?' for more)
slc	change state of special characters ('slc ?' for more)
z	suspend telnet
!	invoke a subshell
environ	change environment variables ('environ ?' for more)
?	print help information