Data Access and Data Warehousing

Michael F. Goodchild
University of California
Santa Barbara

Outline

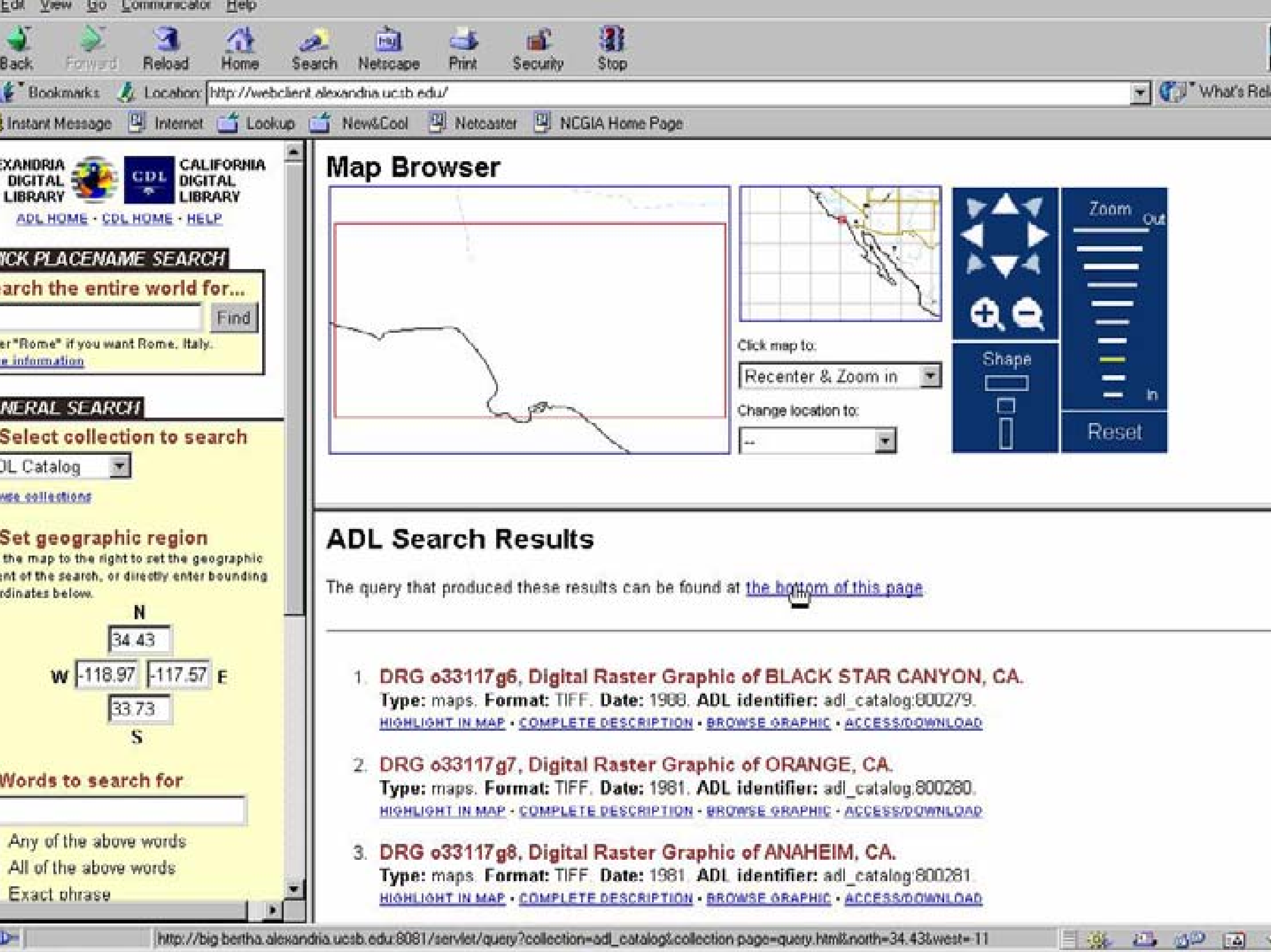
- Geolibraries
- Contrasting world views
- Object-oriented design
- Sharing information

The digital library

- The digital catalog
  - author, title, subject
  - ca 1985
- 10^6 books in a major research library
- 10^6 text characters per book
- 10^{12} total storage requirement
  - ca 1992
  - preWeb

The digital map library

- 10^6 maps and images
- 10^8 bytes per map or image
- 10^{14} bytes total
- Geographic location as the primary key
  - the *geolibrary*
- A physical geolibrary is infeasible
  - one-dimensional, discrete author/title/subject
  - multidimensional, continuous location


Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Location: http://webclient.alexandria.ucsb.edu/
What's Related
Instant Message Internet Lookup New&Cool Netcaster NCGIA Home Page
ALEXANDRIA DIGITAL LIBRARY
CDL
CALIFORNIA DIGITAL LIBRARY
ADL HOME · CDL HOME · HELP
QUICK PLACENAME SEARCH
Search the entire world for...
Find
Enter "Rome" if you want Rome, Italy.
More information
GENERAL SEARCH
Select collection to search
ADL Catalog
Browse collections
Set geographic region
Use the map to the right to set the geographic extent of the search, or directly enter bounding coordinates below.
N
34.43
W
-118.97
-117.57
E
33.73
S
Words to search for
Any of the above words
All of the above words
Exact phrase
Map Browser
Click map to:
Recenter & Zoom in
Change location to:
--
Zoom
Out
In
Shape
Reset
ADL Search Results
The query that produced these results can be found at the bottom of this page
1. DRG o33117g6, Digital Raster Graphic of BLACK STAR CANYON, CA.
Type: maps. Format: TIFF. Date: 1988. ADL identifier: adl_catalog:800279.
HIGHLIGHT IN MAP · COMPLETE DESCRIPTION · BROWSE GRAPHIC · ACCESS/DOWNLOAD
2. DRG o33117g7, Digital Raster Graphic of ORANGE, CA.
Type: maps. Format: TIFF. Date: 1981. ADL identifier: adl_catalog:800280.
HIGHLIGHT IN MAP · COMPLETE DESCRIPTION · BROWSE GRAPHIC · ACCESS/DOWNLOAD
3. DRG o33117g8, Digital Raster Graphic of ANAHEIM, CA.
Type: maps. Format: TIFF. Date: 1981. ADL identifier: adl_catalog:800281.
HIGHLIGHT IN MAP · COMPLETE DESCRIPTION · BROWSE GRAPHIC · ACCESS/DOWNLOAD
http://big.bertha.alexandria.ucsb.edu:8081/servlet/query?collection=adl_catalog&collection page=query.html&north=34.43&west=-11

The Alexandria Digital Library

- 10^7 items, 10^{13} bytes
- Three ways of specifying location
  - pointing to a map
  - latitude and longitude
  - gazetteer
- The library paradigm
  - putting data in the user's hands
  - a gatekeeper
- Beyond maps and images
  - information with a geographic footprint

Instances of geolibraries

- www.alexandria.ucsb.edu
- National Geospatial Data Clearinghouse
  - www.fgdc.gov
- www.geographynetwork.com

NRC report

- "Distributed Geolibraries: Spatial Information Resources", 1999

www.nap.edu

“Imagine, for example, a young child going to a Digital Earth exhibit at a local museum. After donning a head-mounted display, she sees Earth as it appears from space. Using a data glove, she zooms in, using higher and higher levels of resolution, to see continents, then regions, countries, cities, and finally individual houses, trees, and other natural and man-made objects. Having found an area of the planet she is interested in exploring, she takes the equivalent of a ‘magic carpet ride’ through a 3-D visualization of the terrain.”

Perspectives on Digital Earth

- High-end visualization
  - an immersive environment
  - specialized hardware
  - massive bandwidth requirements
- Spin, zoom, pan
  - "fly-by" technology
- 4 orders of magnitude zoom
  - 10km to 1m

Does DE scale?

- 500,000,000 sq km
  - 5 million at 10km resolution
  - 500,000,000,000,000 at 1m resolution

Transmitting Digital Earth

- 1m resolution at T1 (order 10 megabits/sec)
  - 69.4 working years
- 1m resolution at 56k
  - done in 12,400 years
- The Internet-killer

What resolution do we really need?

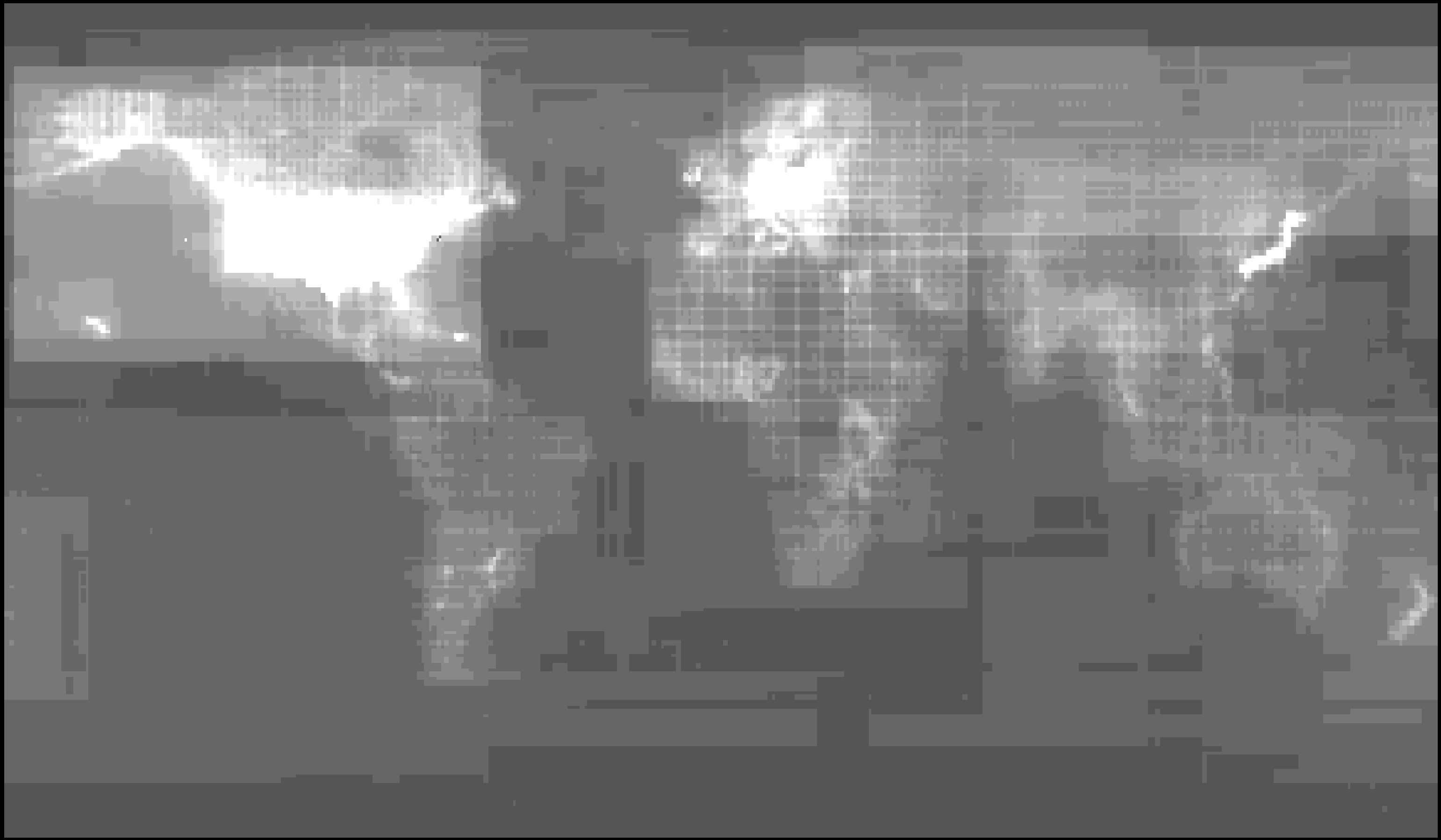
- Whole Earth at 10km
  - California at 1km
  - Santa Barbara County at 100m
- L/S (extent divided by resolution)
  - order 10^3 or 10^4
  - ratio for computer screen
  - ratio for human retina

The Internet can support DE

- 1 refresh per second, 1 megapel images (L/S=10^3)
  - T1 rates without compression
  - 10+ refreshes per second with compression
  - sufficient for zoom, pan, flyby

Research challenges for geolibraries

- Defining footprints
  - fuzzy, vernacular
- Access for the child of ten
  - scale
- Search over a distributed archive
  - search engines
  - object-level metadata (OLM)
  - collection-level metadata (CLM)

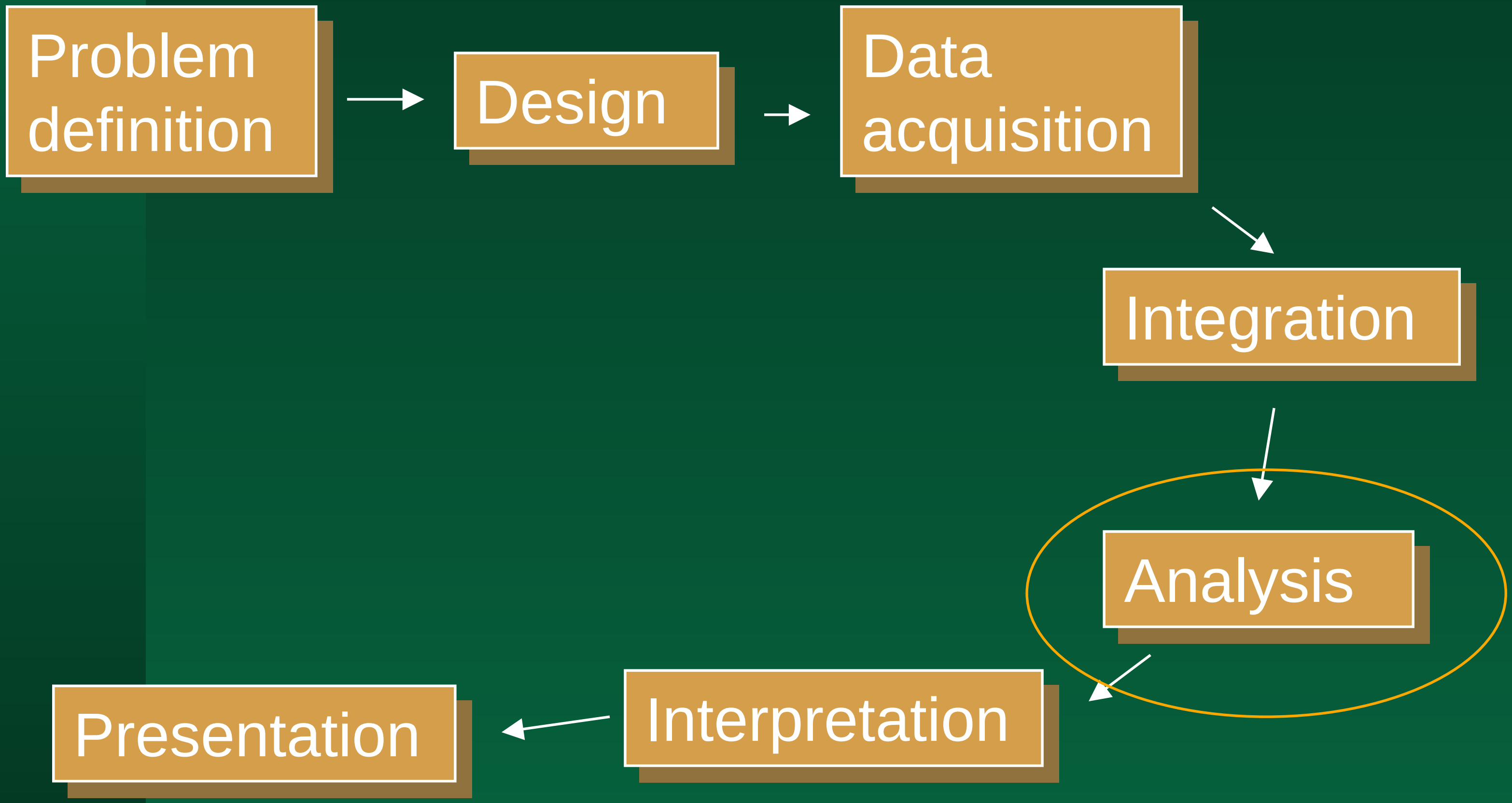


CLM of the Alexandria Digital Library

Research challenges (2)

- Approaches to CLM
  - by data type
    - ortho.mit.edu
  - by area of the globe
    - SRI's Digital Earth
  - the one stop shop
    - www.fgdc.gov
  - a new generation of search engines
    - identifying footprints

Stages of problem solving



Why does it take so long?

- Analysis at the speed of light
- Why can't we solve problems in real time?
- How can we make it faster?

Dear Waldo

View of the Outer Banks of North Carolina
from Apollo 9

This photograph was taken on March 12, 1969 at 10:00 a.m. EST, from an altitude of about 120 miles.

Posted at the old seafaring village of Hatteras, I know this card, with its complete and accurate address will get to you.

A pinhole shows you where we are—

Yours Geographically

Bill

Elizabeth City News Co. Elizabeth City, North Carolina

Plastichrome®

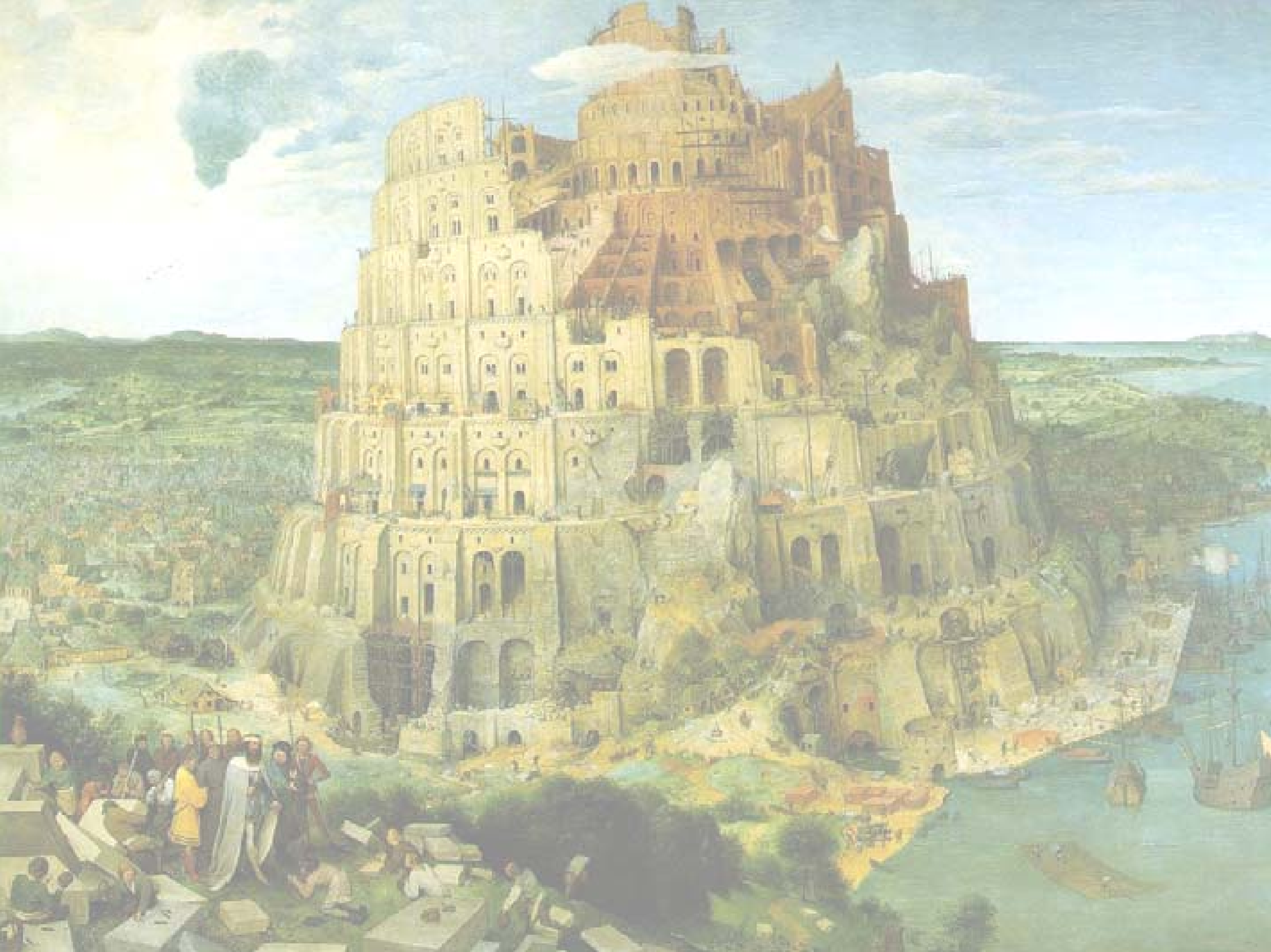
PRINTED IN IRELAND

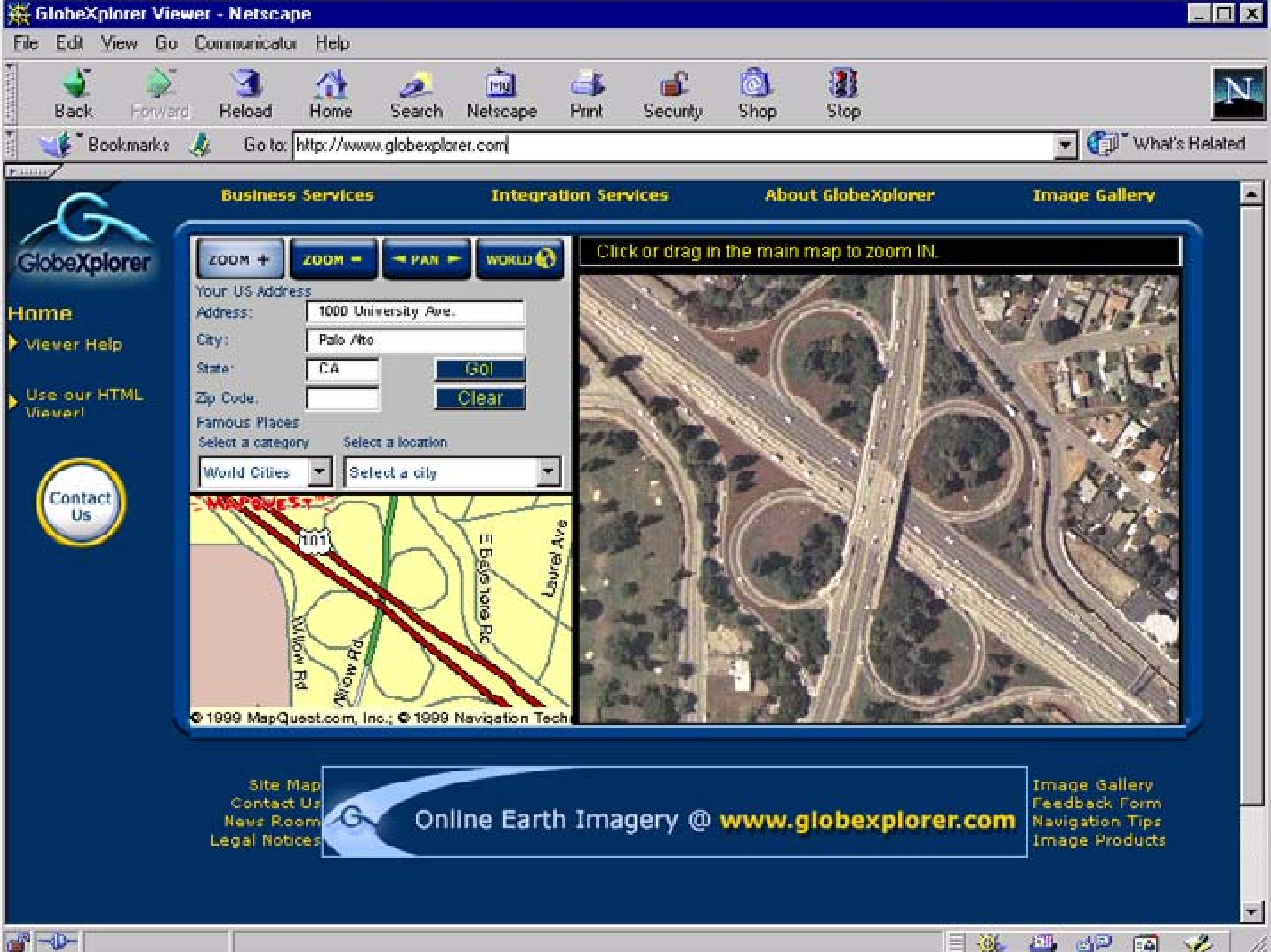
POST CARD

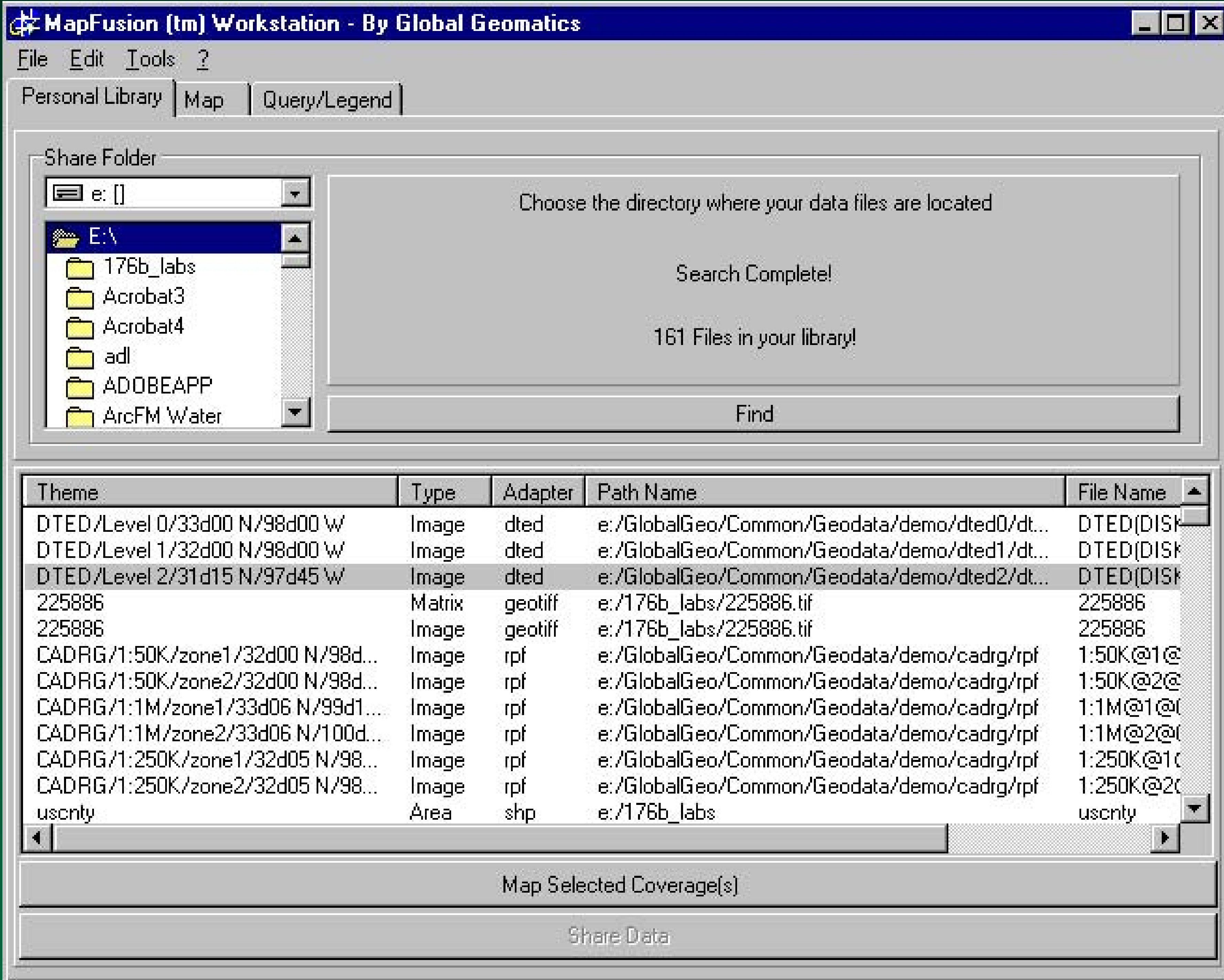
Professor Waldo Tobler

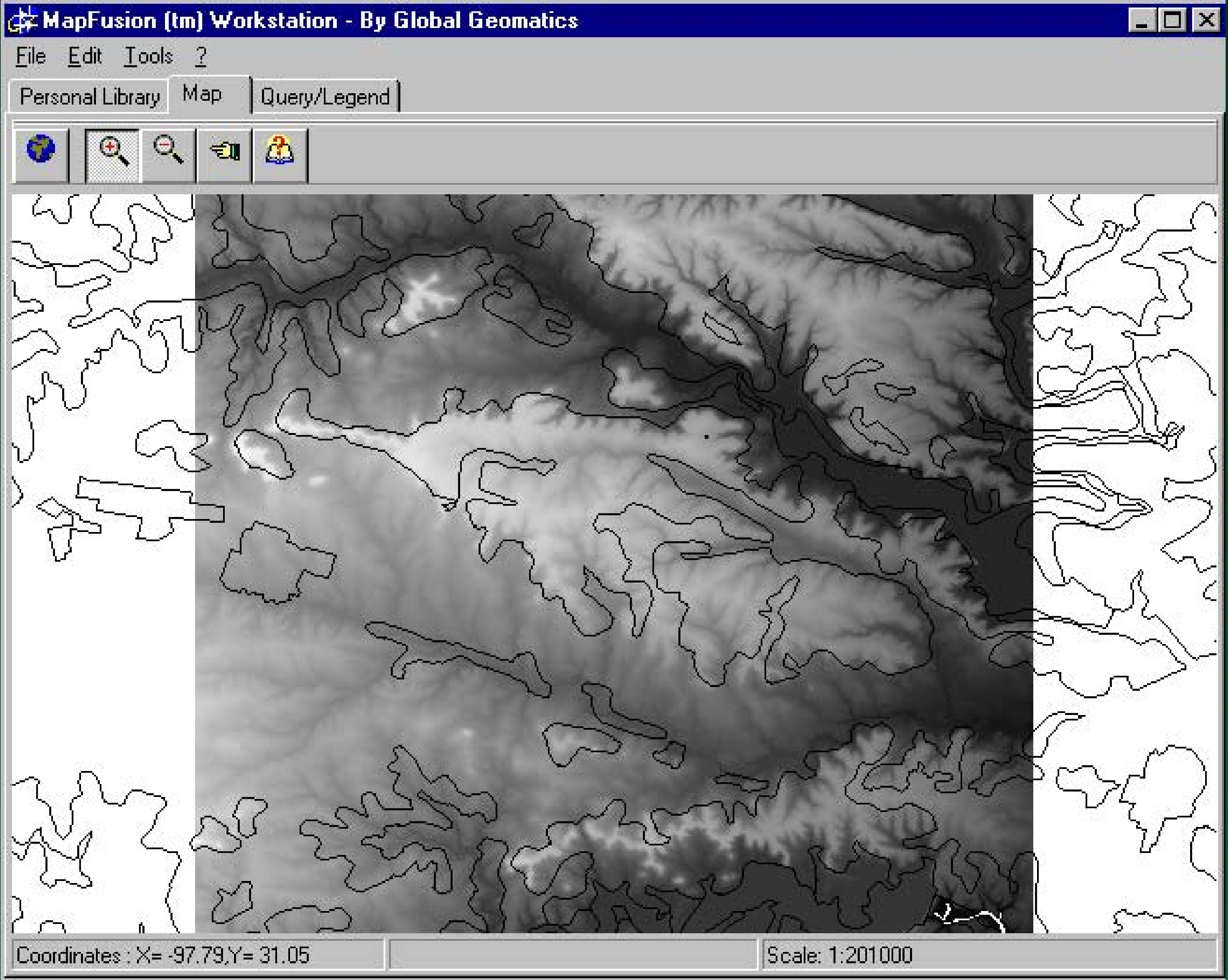
34° 26' 41" N

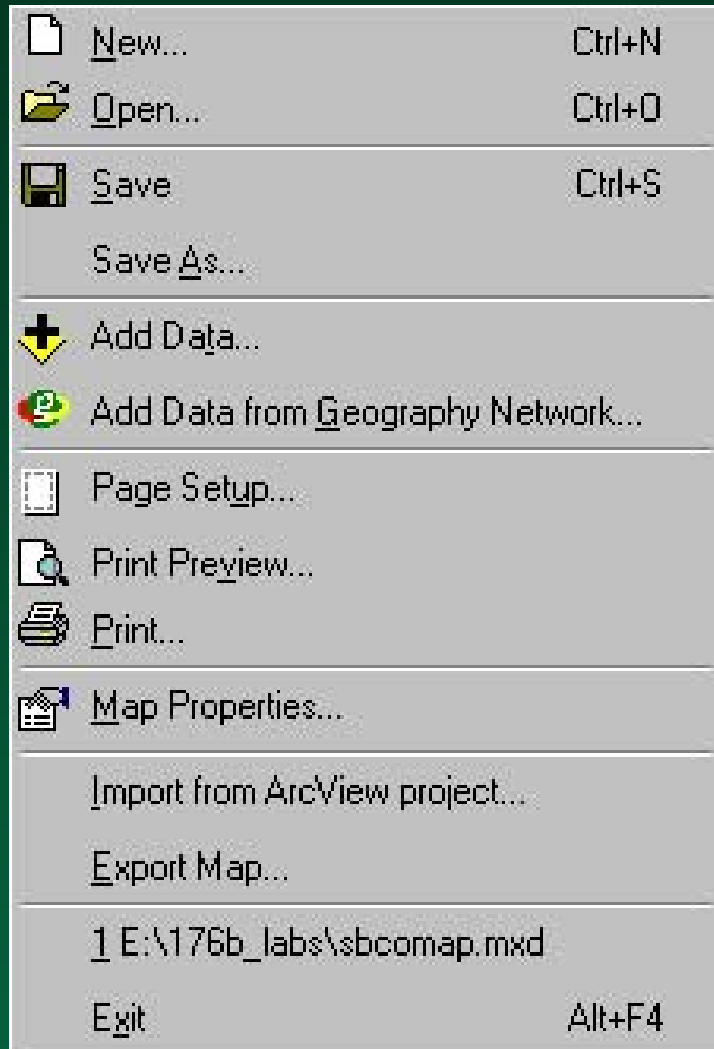
119° 48' 26" W




GlobeXplorer Viewer - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks
Go to: http://www.globexplorer.com
What's Related
Business Services
Integration Services
About GlobeXplorer
Image Gallery
GlobeXplorer
Home
Viewer Help
Use our HTML Viewer!
Contact Us
ZOOM +
ZOOM −
PAN
WORLD
Your US Address
Address: 1000 University Ave.
City: Palo Alto
State: CA
Zip Code:
Go!
Clear
Famous Places
Select a category
Select a location
World Cities
Select a city
Click or drag in the main map to zoom IN.
101
Willow Rd
E Bayshore Rd
Laurel Ave
© 1999 MapQuest.com, Inc.; © 1999 Navigation Tech
Site Map
Contact Us
News Room
Legal Notices
Online Earth Imagery @ www.globexplorer.com
Image Gallery
Feedback Form
Navigation Tips
Image Products


MapFusion (tm) Workstation - By Global Geomatics
File Edit Tools ?
Personal Library
Map
Query/Legend
Share Folder
e: []
E:\
176b_labs
Acrobat3
Acrobat4
adl
ADOBEAPP
ArcFM Water
Choose the directory where your data files are located
Search Complete!
161 Files in your library!
Find
Theme Type Adapter Path Name File Name
DTED/Level 0/33d00 N/98d00 W Image dted e:/GlobalGeo/Common/Geodata/demo/dted0/dt... DTED(DISK
DTED/Level 1/32d00 N/98d00 W Image dted e:/GlobalGeo/Common/Geodata/demo/dted1/dt... DTED(DISK
DTED/Level 2/31d15 N/97d45 W Image dted e:/GlobalGeo/Common/Geodata/demo/dted2/dt... DTED(DISK
225886 Matrix geotiff e:/176b_labs/225886.tif 225886
225886 Image geotiff e:/176b_labs/225886.tif 225886
CADRG/1:50K/zone1/32d00 N/98d... Image rpf e:/GlobalGeo/Common/Geodata/demo/cadrg/rpf 1:50K@1@
CADRG/1:50K/zone2/32d00 N/98d... Image rpf e:/GlobalGeo/Common/Geodata/demo/cadrg/rpf 1:50K@2@
CADRG/1:1M/zone1/33d06 N/99d1... Image rpf e:/GlobalGeo/Common/Geodata/demo/cadrg/rpf 1:1M@1@
CADRG/1:1M/zone2/33d06 N/100d... Image rpf e:/GlobalGeo/Common/Geodata/demo/cadrg/rpf 1:1M@2@
CADRG/1:250K/zone1/32d05 N/98... Image rpf e:/GlobalGeo/Common/Geodata/demo/cadrg/rpf 1:250K@1
CADRG/1:250K/zone2/32d05 N/98... Image rpf e:/GlobalGeo/Common/Geodata/demo/cadrg/rpf 1:250K@2
uscnty Area shp e:/176b_labs uscnty
Map Selected Coverage(s)
Share Data


MapFusion (tm) Workstation - By Global Geomatics
File
Edit
Tools
?
Personal Library
Map
Query/Legend
Coordinates : X= -97.79,Y= 31.05
Scale: 1:201000


New... Ctrl+N
Open... Ctrl+O
Save Ctrl+S
Save As...
Add Data...
Add Data from Geography Network...
Page Setup...
Print Preview...
Print...
Map Properties...
Import from ArcView project...
Export Map...
1 E:\176b_labs\sbcomap.mxd
Exit Alt+F4

geography network

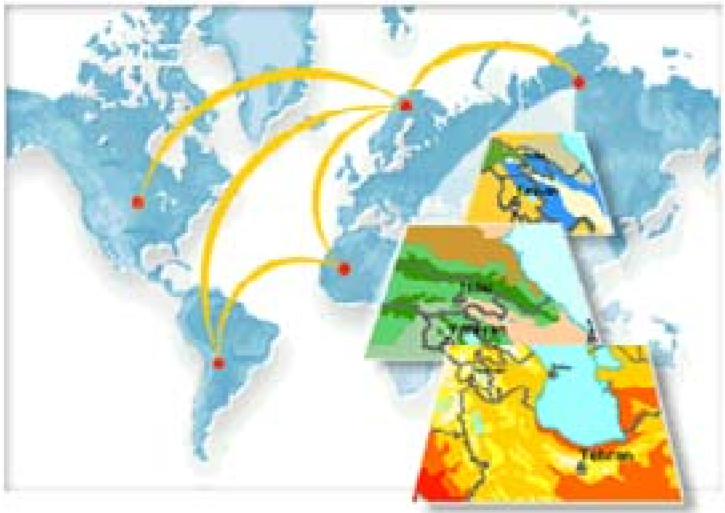
ACCESS A WORLD OF INFORMATION

GeographyNetwork.com Challenge

SEARCH & VIEW

use the **Geography Network Explorer** to search and view maps and other geographic content over the Internet

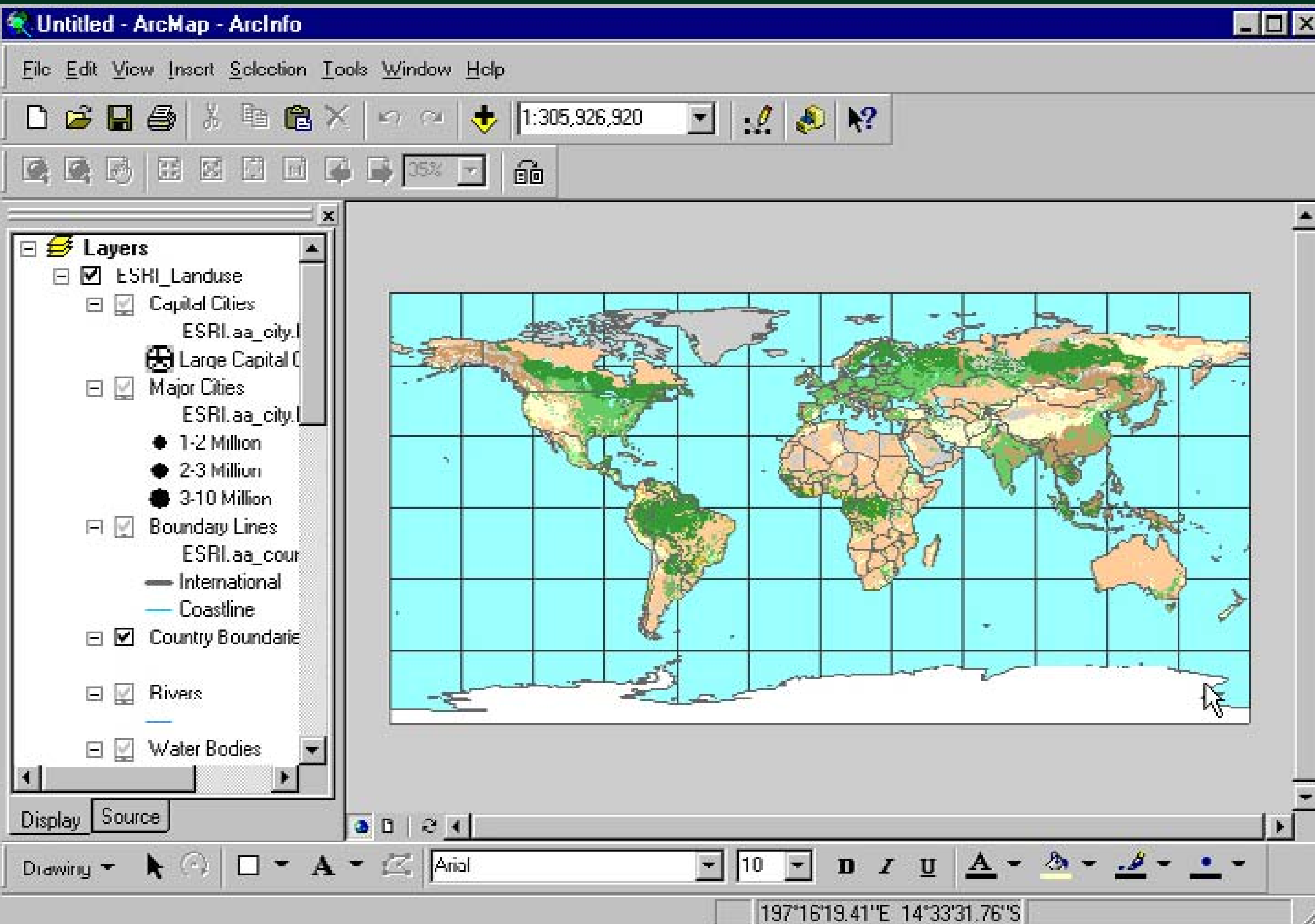
The Geography Network is a global community of data providers who are committed to making geographic content available. This content is published from many sites around the world, providing you immediate access to the latest maps, data, and related services. This portal to the Geography Network enables you to discover this content and share your own.

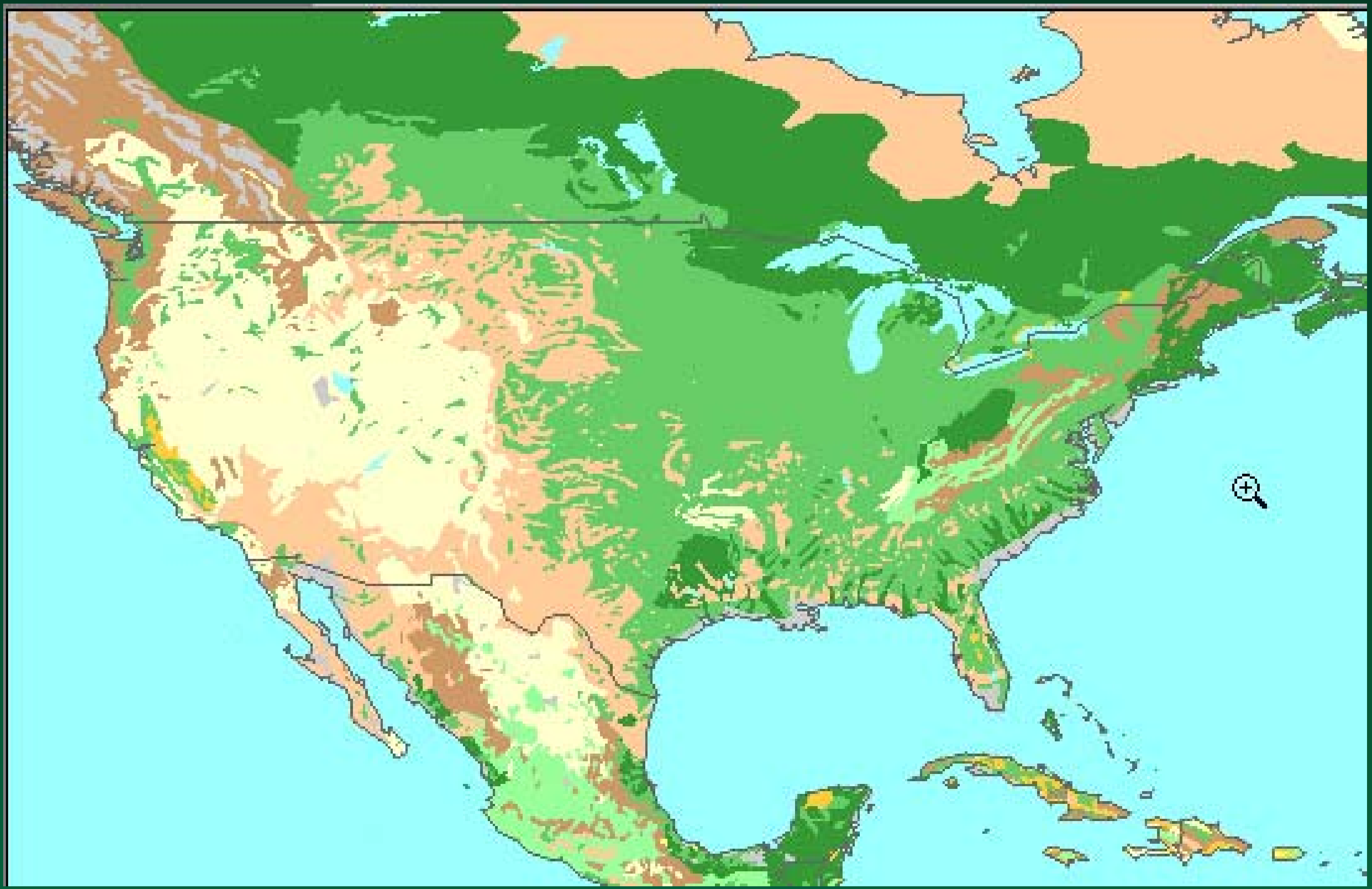


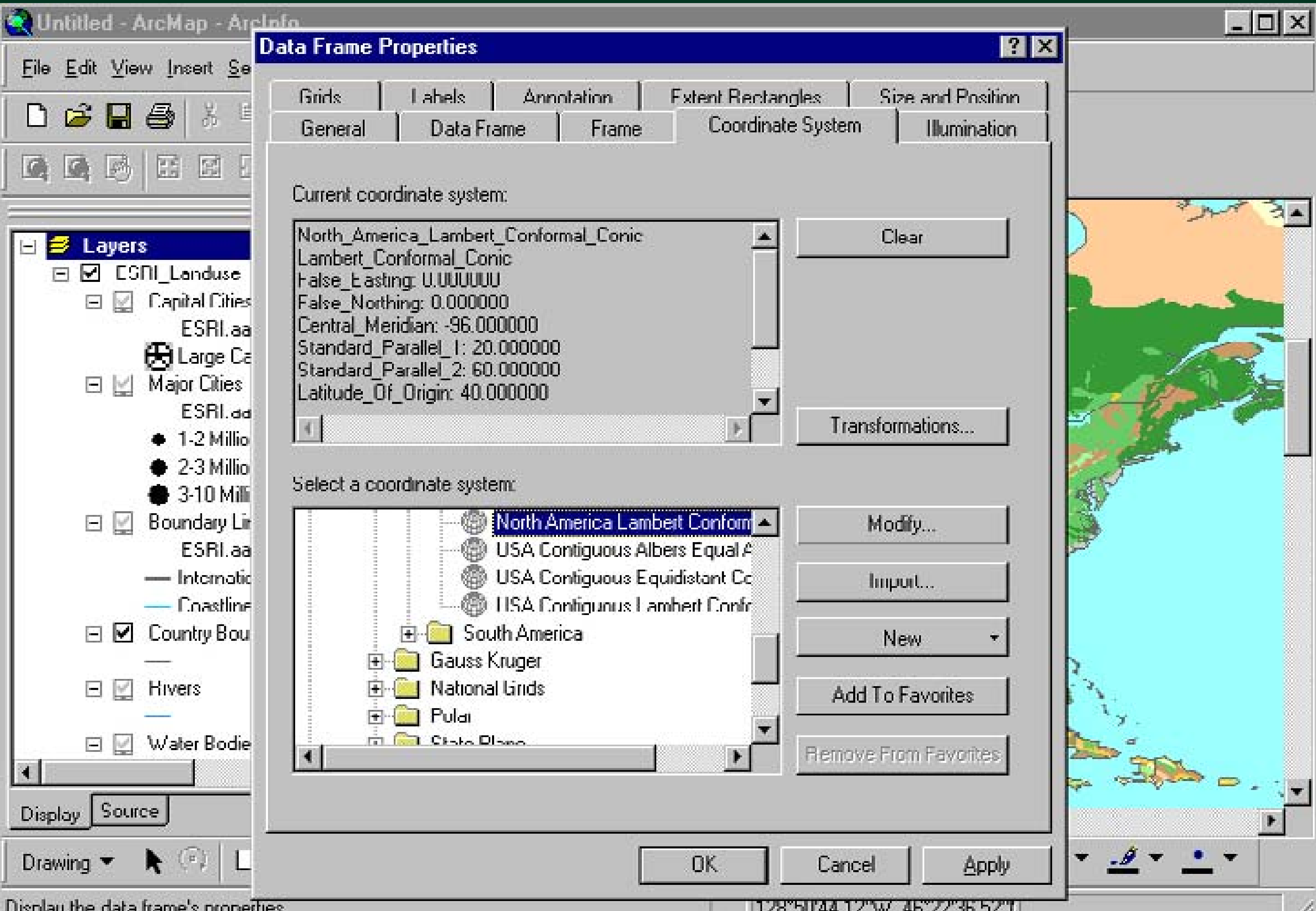
Featured Content

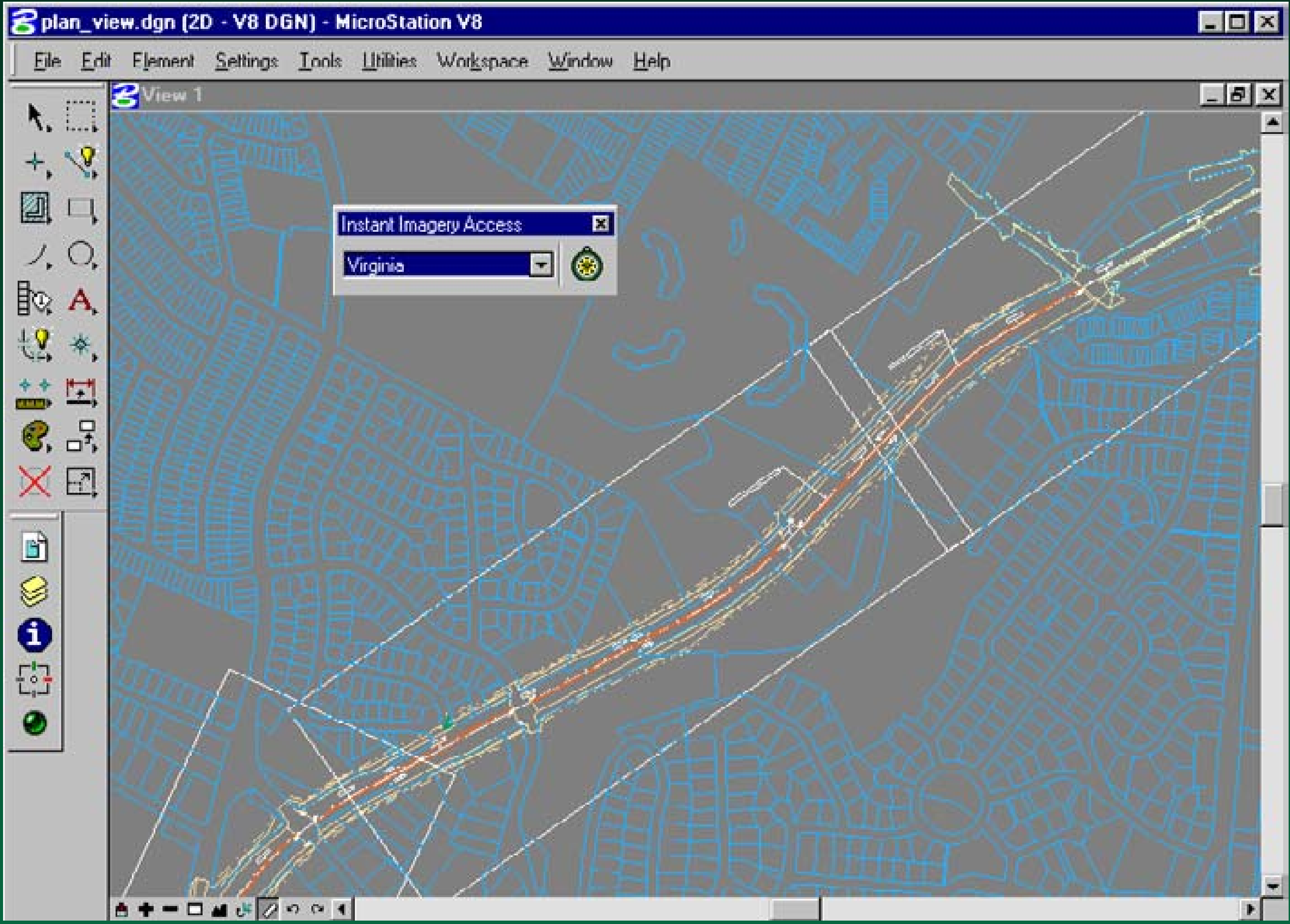
U.S. Census TIGER 2000

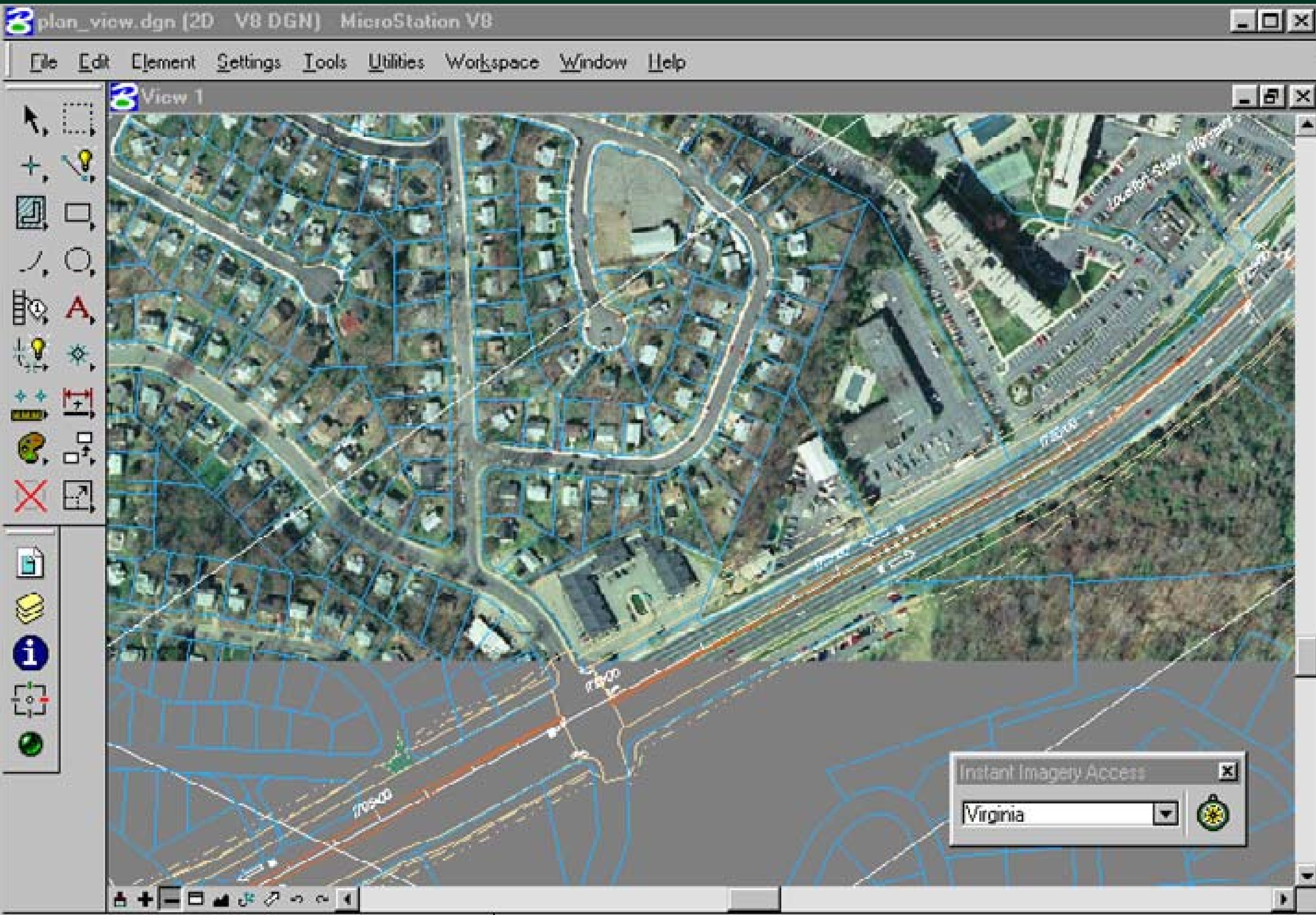
- View Live Maps
- Download Data Sets
- Build Custom Apps
- Publish Your Content
- Find Useful Tools
- Share Your Ideas


Untitled - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
1:305,926,920
Layers
ESRI_Landuse
Capital Cities
ESRI.aa_city.l
Large Capital C
Major Cities
ESRI.aa_city.l
1-2 Million
2-3 Million
3-10 Million
Boundary Lines
ESRI.aa_cour
International
Coastline
Country Boundarie
Rivers
Water Bodies
Display
Source
Drawing
Arial
10
197°16'19.41"E 14°33'31.76"S




Untitled - ArcMap - ArcInfo
File Edit View Insert
Data Frame Properties
Grids
Labels
Annotation
Extent Rectangles
Size and Position
General
Data Frame
Frame
Coordinate System
Illumination
Current coordinate system:
North_America_Lambert_Conformal_Conic
Lambert_Conformal_Conic
False_Easting: 0.000000
False_Northing: 0.000000
Central_Meridian: -96.000000
Standard_Parallel_1: 20.000000
Standard_Parallel_2: 60.000000
Latitude_Of_Origin: 40.000000
Clear
Transformations...
Select a coordinate system:
North America Lambert Conform
USA Contiguous Albers Equal A
USA Contiguous Equidistant Co
USA Contiguous Lambert Confo
South America
Gauss Kruger
National Grids
Polar
Modify...
Import...
New
Add To Favorites
Remove From Favorites
OK
Cancel
Apply
Layers
ESRI_Landuse
Capital Cities
Large Ca
Major Cities
1-2 Millio
2-3 Millio
3-10 Milli
Boundary Li
Internati
Coastline
Country Bou
Rivers
Water Bodie
Display
Source
Drawing
Display the data frame's properties
128°50'44.12"W 46°22'36.52"N


plan_view.dgn (2D - V8 DGN) - MicroStation V8
File Edit Element Settings Tools Utilities Workspace Window Help
View 1
Instant Imagery Access
Virginia


plan_view.dgn (2D - V8 DGN) - MicroStation V8
File
Edit
Element
Settings
Tools
Utilities
Workspace
Window
Help
View 1
Instant Imagery Access
Virginia

Objectives of interoperability

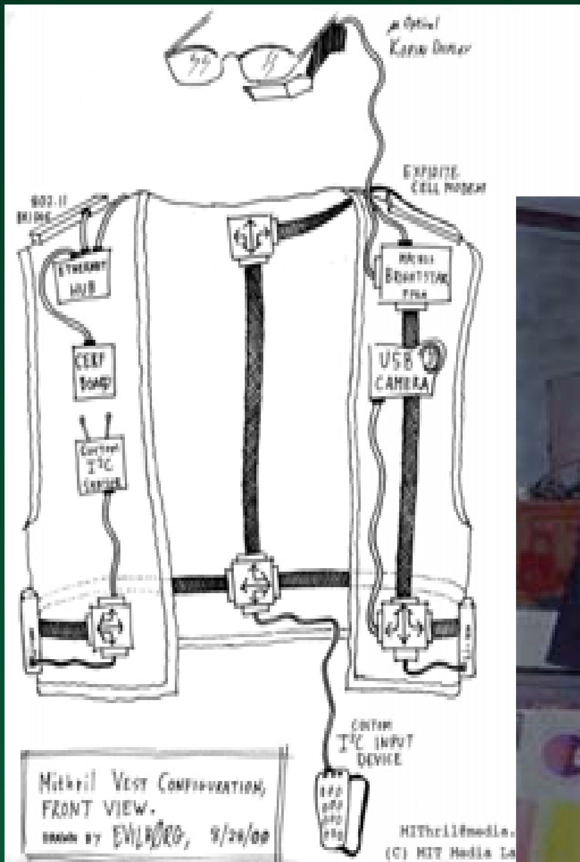
- Using technology to overcome differences
  - rather than imposing uniformity
  - enabling rather than intrusive
  - specifications not standards
- Bridging information communities
- Speeding and easing access to data

Major forces in spatial data interoperability

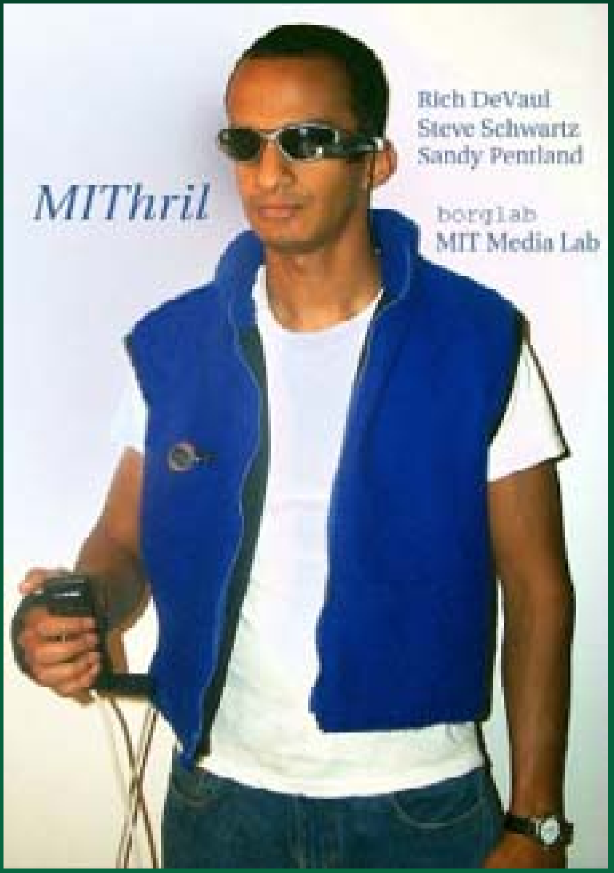
- National Spatial Data Infrastructure
  - Federal Geographic Data Committee
- Open GIS Consortium
  - industry, government, academic
- National, regional, and international standards organizations

GIS in the field

- Mobile, traveling with the user
- Ubiquitous, operating anywhere
- Augmenting the senses with information from digital representations
  - of the past
  - of what is beyond the senses
  - of the future



802.11 BRIDGE
ETHERNET HUB
CERF BOARD
USB CAMERA
CUSTOM I²C INPUT DEVICE
MiThril Vest Configuration, FRONT VIEW.
DRAWN BY EVILBORG, 9/20/00



Rich DeVaul
Steve Schwartz
Sandy Pentland
MIThril
borglab
MIT Media Lab

CharmIT™ Developer's Kit

- CharmIT™ is built on the PC/104 specification, which has been an industry standard for embedded computing for nearly ten years
- hundreds of companies manufacture a wide variety of PC/104 hardware
- majority of components are low power and ruggedized
- CharmIT™ Developer's Kit is lower cost (approximately $2000), low power (approximately 7 watts with Jumptec 266) and offers enough computing power for most everyday wearable tasks

Head-mounted displays

Integrated Eyeglassisplays ($5000)

ClipOn Display ($2500)
-evaluation kit comes with a belt-worn, VGA interface box connected to the display by a 4' cable
Display format: 640x480, 24-Bit color, 60 Hz refresh rate
Field of View: Approximately 16 degrees horizontal

MicroOptical

Text input

- The Twiddler2 chorded keyboard is designed for one-handed input with an array of 12 finger keys and six thumb keys.
- Frequent users can enter text at close to two-hand touch-typing speeds.

User interface for augmented vision

Augmented: see-through map plus locator

Viewed reality

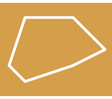
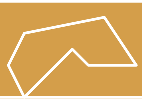
View options

Field-work applications

- Finding oneself in the field
  - recovering past sample locations
- Accessing previously collected data
  - the previous census
- Analyzing data continuously
  - progressive formation of geographic knowledge

Location as attribute

- The data table
  - Census summary table
  - county-level health statistics
- What value is location as an explanatory variable?
- Linking the table to a boundary file
  - enabling maps of summary data

Tract	Pop	Location	Shape
1	3786	*x,y*	
2	2966	*x,y*	
3	5001	*x,y*	
4	4983	*x,y*	
5	4130	*x,y*	
6	3229	*x,y*	
7	4086	*x,y*	
8	3979	*x,y*	

Abstraction of geographic space

- Cartograms

- Invariance under rotation, displacement
- Reconstruction from a distance matrix
- Reconstruction from ranked distances
  - ordered metric data (Coombs)

Space as a matrix

- ***W*** where w_{ij} is some measure of interaction
  - adjacency
  - decreasing function of distance
  - invariant under rotation, displacement
  - readily obtained from a GIS

Applications of the *W* matrix

- Spatial regression
  - add spatially lagged terms weighted by ***W***
  - Anselin's SPACESTAT
- Moran and Geary indices of spatial dependence

$$c = \frac{(n-1)\sum_i \sum_j w_{ij}\left(x_i - x_j\right)^2}{2\sum_i \sum_j w_{ij} \sum_i \left(x_i - a\right)^2}$$

The location-as-attribute world view

- Objective: scientific explanation, understanding of social processes
  - is location an explanatory factor?
- Relative location as expressed in the ***W*** matrix
  - a surrogate for spatial interaction
  - reflecting costs of transport, probability of interaction and acquaintance, probability of migration or travel, probability of seed dispersal

The Data Documentation Initiative

- A multinational multidisciplinary effort
- A standard for description of dataset contents
- Social science focus
- Geography working group
  - joint meeting August 02

Two world views

- Location as continuum
  - the FGDC metadata standard
  - ISO 19115
  - attribute tables as part of GIS
- Location as attribute
  - the DDI standard
- Reinforced by technical GIS design
  - the hybrid model
  - attributes in an RDBMS
  - geometry in a specialized file structure
  - ARC/INFO

Object-oriented design

- Objects as instances of classes
- Classes inherit properties of more generalized classes (*inheritance*)
- Methods associated with classes (*encapsulation*)

Specialized GIS data models

- The basic elements built into the GIS
  - points, lines, areas
  - the GIS mainstream
- How these elements are specialized in application domains (vertical markets)
  - railroad track as a class of transportation link
  - transportation link as a class of line

Unified Modeling Language

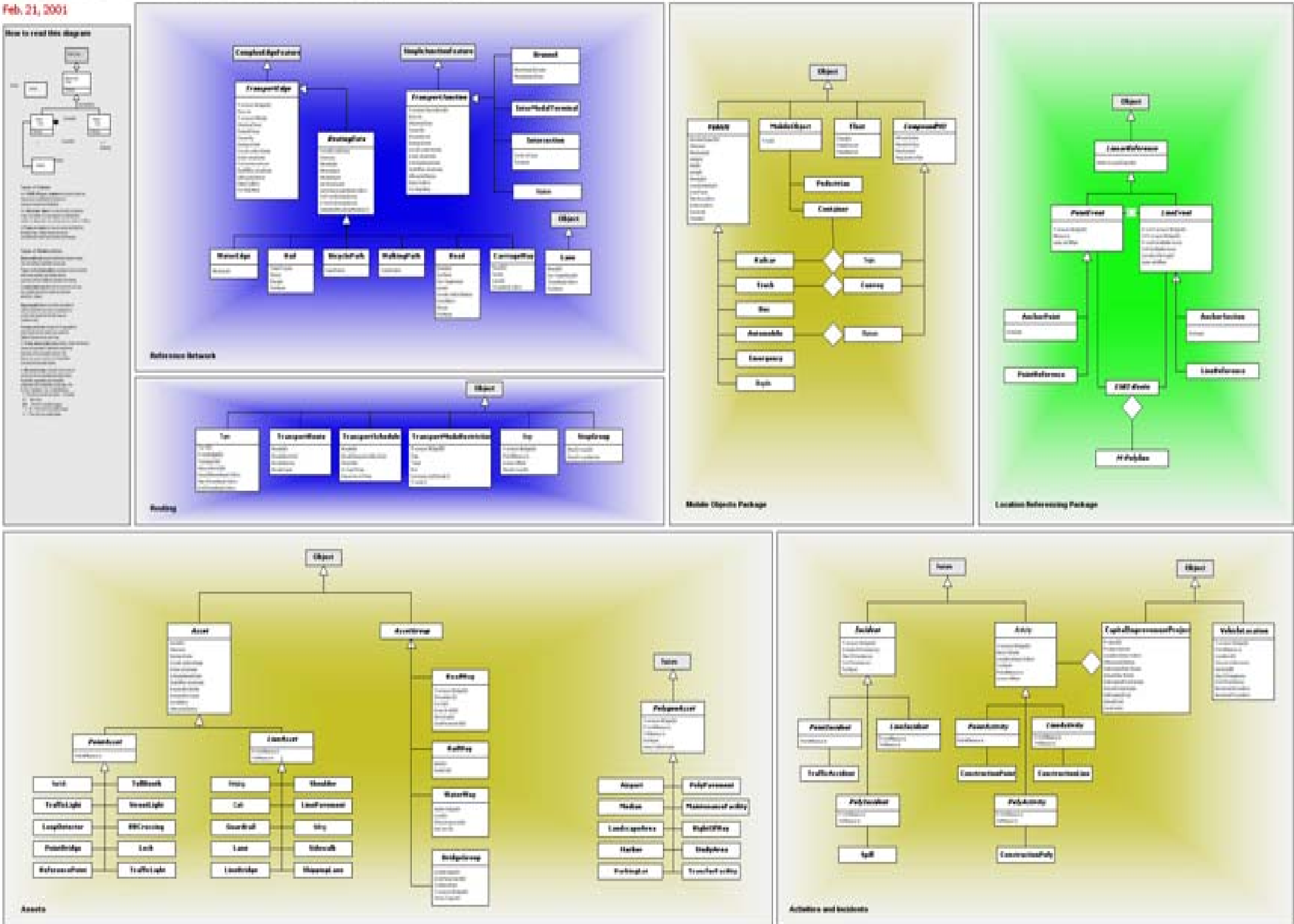
- Visual representation of a data model
  - conventional symbols
  - implemented in Visio
- Creation of database layout
  - use CASE tools
  - build tables
  - populate tables with data

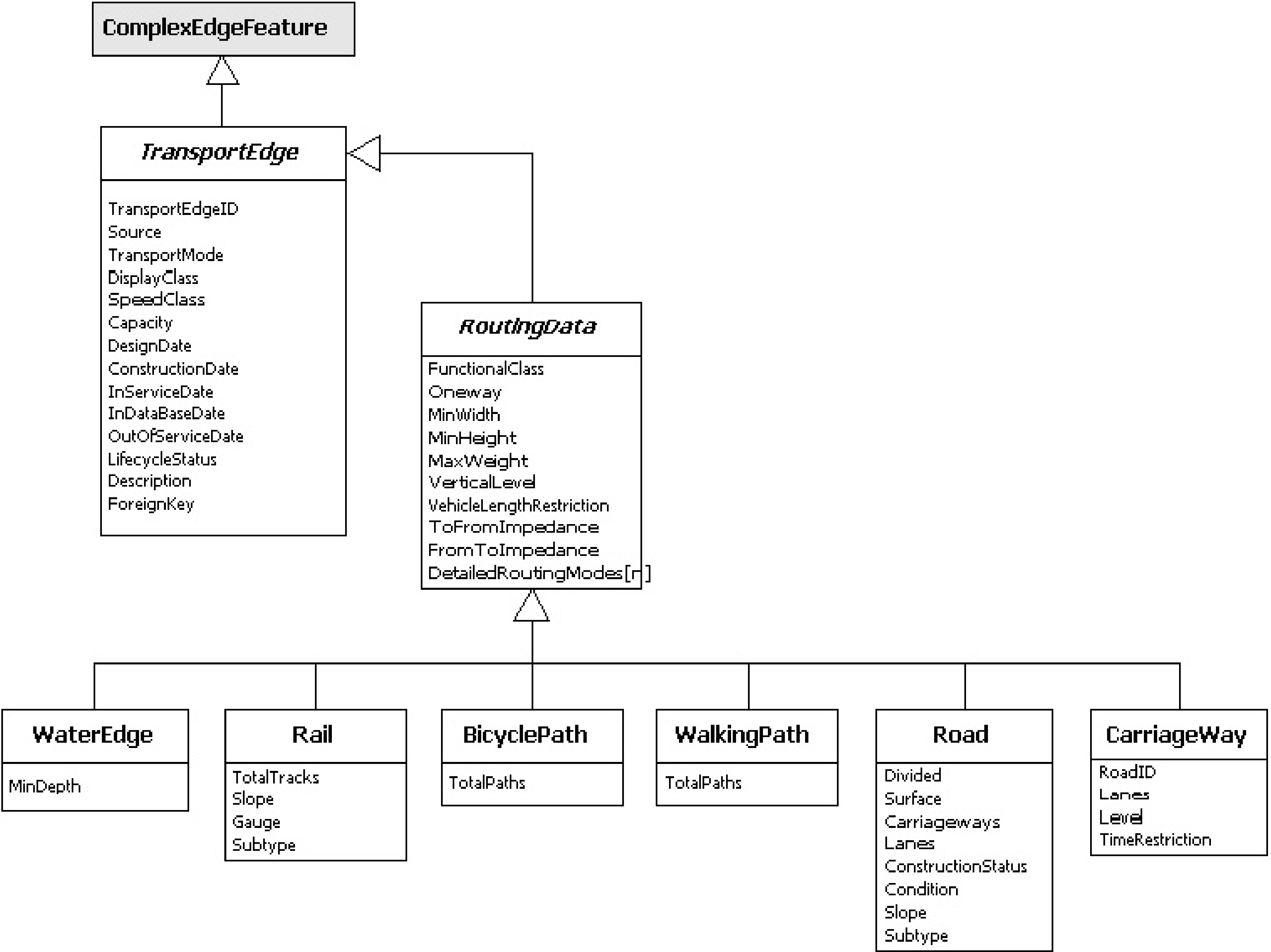
UNETRANS

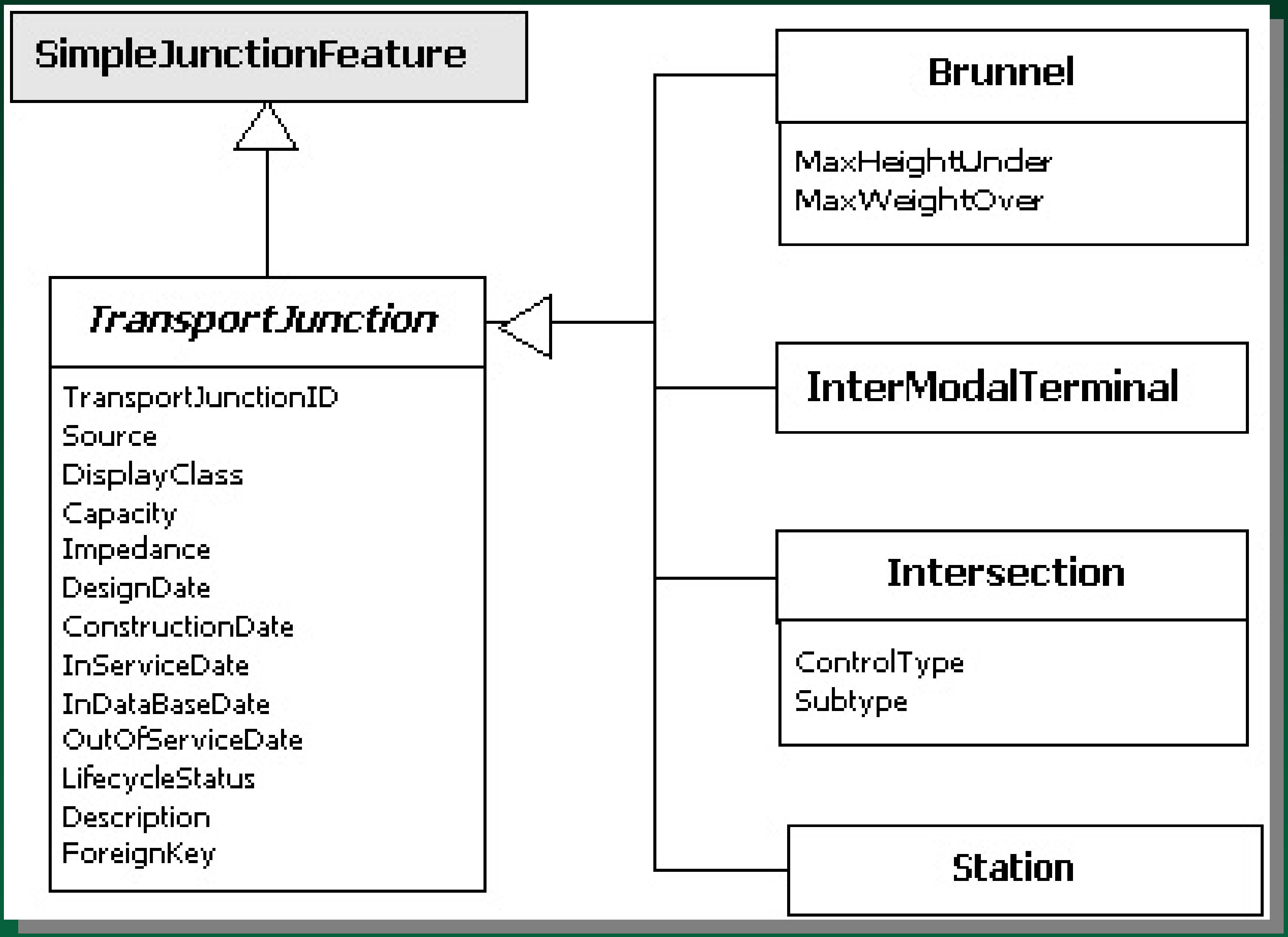
- Helping transportation users of ArcGIS by providing a database framework that includes familiar elements
  – contains the core items
  – is easy to extend and specialize
  – add new attributes
  – add specialized classes

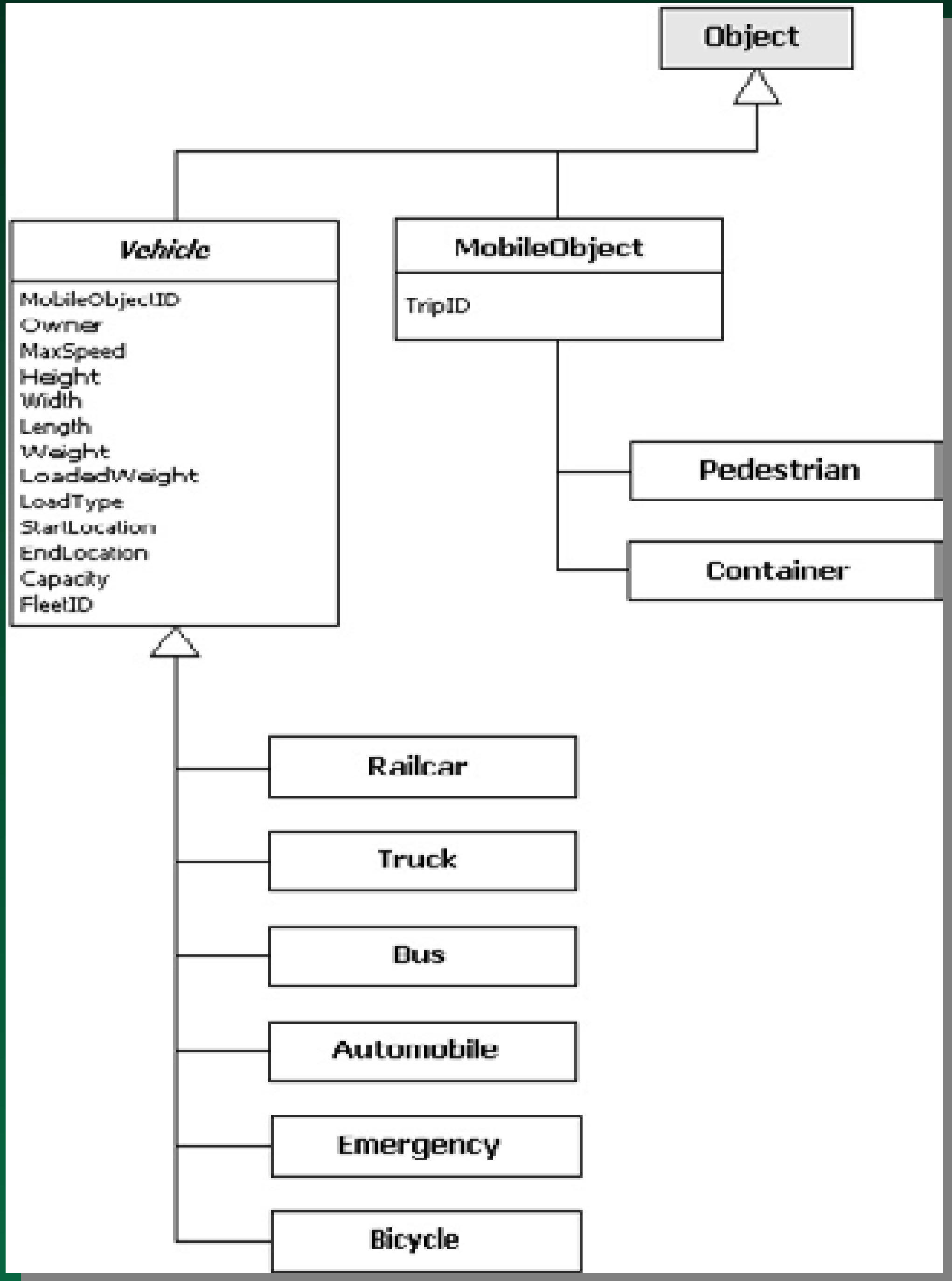
ArcGIS Transportation Data Model

Feb. 21, 2001




ComplexEdgeFeature
TransportEdge
TransportEdgeID
Source
TransportMode
DisplayClass
SpeedClass
Capacity
DesignDate
ConstructionDate
InServiceDate
InDataBaseDate
OutOfServiceDate
LifecycleStatus
Description
ForeignKey
RoutingData
FunctionalClass
Oneway
MinWidth
MinHeight
MaxWeight
VerticalLevel
VehicleLengthRestriction
ToFromImpedance
FromToImpedance
DetailedRoutingModes[n]
WaterEdge
MinDepth
Rail
TotalTracks
Slope
Gauge
Subtype
BicyclePath
TotalPaths
WalkingPath
TotalPaths
Road
Divided
Surface
Carriageways
Lanes
ConstructionStatus
Condition
Slope
Subtype
CarriageWay
RoadID
Lanes
Level
TimeRestriction


SimpleJunctionFeature
TransportJunction
TransportJunctionID
Source
DisplayClass
Capacity
Impedance
DesignDate
ConstructionDate
InServiceDate
InDataBaseDate
OutOfServiceDate
LifecycleStatus
Description
ForeignKey
Brunnel
MaxHeightUnder
MaxWeightOver
InterModalTerminal
Intersection
ControlType
Subtype
Station



Object
Vehicle
MobileObjectID
Owner
MaxSpeed
Height
Width
Length
Weight
LoadedWeight
LoadType
StartLocation
EndLocation
Capacity
FleetID
MobileObject
TripID
Pedestrian
Container
Railcar
Truck
Bus
Automobile
Emergency
Bicycle

But what about metadata?

- At the class (table) level?
- At the database level?
- Granularity
  - an unresolved issue

A comprehensive view of information

- To be sharable, information must be digital
- Many types of information
  - maps, images
  - tables
  - text
  - methods
  - simulation models
- What is the relative value per bit?
  - Windows XP >> a cloudy Ikonos image
  - academic paper >> survey data used

The infrastructure of information sharing

- Metadata standard
- Archives
- Interoperability
- To date, almost exclusively about data
  - much less about methods

Towards an infrastructure for dynamic models

- Infrastructure for sharing
  - search
  - discovery
  - evaluation of fitness for use
  - acquisition
  - execution
- Server-side or client-side execution

Falling through the cracks

- Text-sharing infrastructure
  – libraries, bookstores, books, journals, WWW, search engines
- Data-sharing infrastructure
  – metadata schema, archives, clearinghouses, data centers
- Model-sharing infrastructure
  – models are the highest form of sharable knowledge of the Earth system

Current status

- Some archives
  - some pre-WWW
- No standards
- No clearinghouses
- www.ncgia.ucsb.edu/~scott

Research in Metadata for Computer Models

Models available over the Web
Model research and articles
Metadata & Cataloging: Examples, Ideas & Articles

Meetings
Interviews
Readings

Reporting Model 'Fitness of Use' or 'Validation' in Metadata
Comparison chart for Model Metadata
An Easier Method for Metadata Collection
Creating a Computer Model Metadata Standard

A Special thanks to the people at The National Science Foundation for their support of the Research Education for Undergraduates program

link to Geography 5, Fall 2000

Building a metadata standard for describing models

- A model is a transformation
  - characterized by metadata for inputs and outputs
- Write down the key elements
  - compare FGDC CSDGM
- How do humans do it?
  - we’ve been doing it for decades
- A first-draft standard

Conclusions

- Instant geospatial data now feasible
  - access to distributed archives
- CLM is a problem
  - only experts know where to look
  - the one-stop shop will not happen
- Geohealth embodies two distinct world views
  - location as continuum
  - location as attribute
- Effective sharing of data must somehow embrace both
  - perhaps through new data-modeling technologies

Conclusions (2)

- Data may not be the most valuable type of information
  - other types may justify greater investment in sharing infrastructure
  - a holistic approach to information sharing is needed in geohealth