The Spatial Web

Michael F. Goodchild
University of California
Santa Barbara

Outline

- Is the Web organized spatially or functionally?
  - the locations of computing
- The future spatial Web
  - beyond data sharing
  - scaling properties
  - Digital Earth
  - the Grid

The Death of Distance

- Cairncross 1997
- Cost(distance) = 0
  - every point is as accessible as every other point
  - "there is no more there, everywhere is here" (Anna Paquin)
  - social networks are independent of distance
    - *p*(receiving email from any point on Earth) = constant
  - location on the Web is transparent
  - returns to Web searches are independent of location
    - *p*(hit anywhere on Earth) = constant

Consider the consequences

- Retailing
  - a network of central facilities serving a dispersed population
  - min(cost of traveling to stores, cost of constructing and maintaining stores)
  - threshold, range
  - each store of the same type contains approximately the same contents
    - general interest
    - specialized interest due to language, location, culture
- Distance is dead, range is infinity
  - only one bookstore needed
  - amazon.com conquers the world

The library

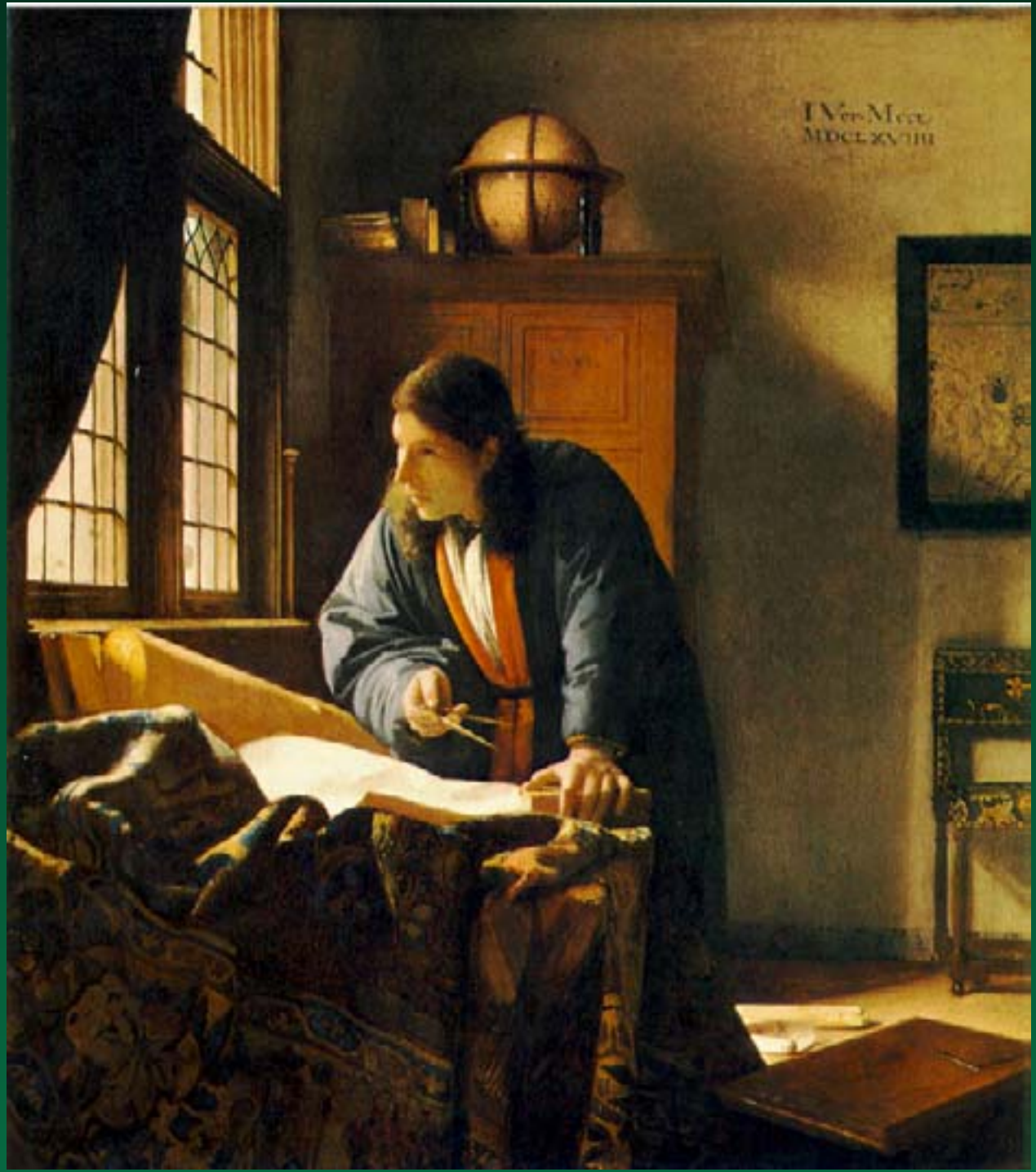
- A central service
  - each library tries to have everything
    - the one with the most wins
- An Internet service
  - one digital library for the world
    - one copy of all of the world's books
    - 100 TB ASCII
  - became technically feasible circa 1995

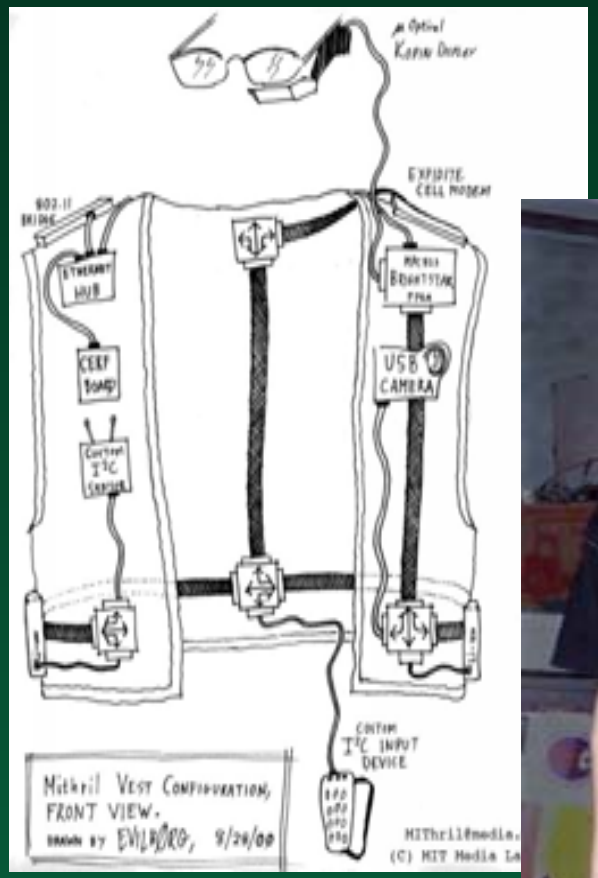
The locations of computing

- User location **u**
  - the user interface
- Processing location **p**
  - ||**u-p**||
    - 1960s < 10m
    - dedicated lines ca 1970 <10km
    - now no limit
- Data storage location **d**
  - independent of **u**, **p**
- Subject location **s**
  - independent of **u**, **p**, **d**

User and subject

- ||**u**-**s**|| = 0
  - augmented reality (AR)
    - information system augments the senses
    - information system replaces impaired senses
  - LBS
  - contextual awareness
  - field work
  - emergency management
- ||**u**-**s**||>>0
  - virtual reality (VR)
    - virtual tourist
    - desk-top GIS
    - Digital Earth




Optional
Kopin Driver
EXPRESS CELL MODEM
802.11 BRIDGE
ETHERNET HUB
MICRO BRIGHTSTAR PDA
CERF BOARD
USB CAMERA
Custom I²C Sensor
CUSTOM I²C INPUT DEVICE
Mithril VEST CONFIGURATION, FRONT VIEW.
DRAWN BY EVILROBO, 9/20/00
MIThril@media.
(C) MIT Media La


MIThril
Rich DeVaul
Steve Schwartz
Sandy Pentland
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

Text input

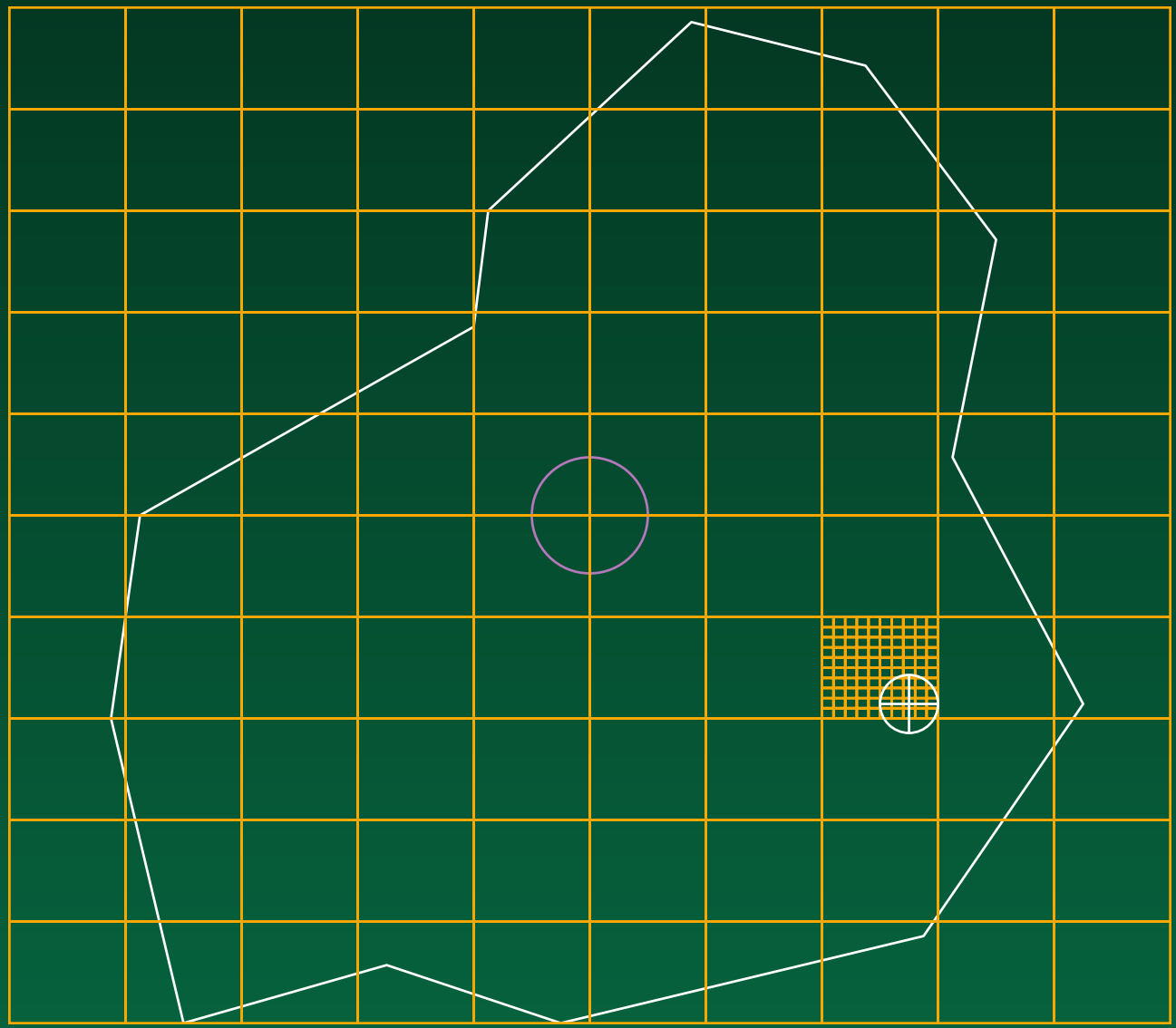
- The Twiddler2 chorded keyboard is designed for one-handed input with an array of 12 finger keys and six thumb keys.
- Frequent users can enter text at close to two-hand touch-typing speeds.

Locations in field computing

- Many workers in the field (many **u**'s at **s**)
  - interacting with eachother through the senses and through the technology
  - interacting with distributed services and data
- Possibly a hierarchical arrangement
  - a field manager also at **s**
  - a field base

Locations and the Web

- Domain names
  - .com, .edu, .org
  - .ca, .ee, .it
  - relays
  - spoofed return addresses
- Efforts to spatialize IP addresses
  - .geo proposal
  - Go2 coordinates (www.go2online.com)



Location = .81 .73 .XYZ

Options for **d**

- Where to store data in the SDI?
  - cost of dissemination goes to zero
  - close to **s**
    - access to ground truth
  - level of interest determined by ||**u-s**||
    - information of geographically determined interest
    - geographic information is IGDI
    - but other information is not
  - convergence of **u** and **s**

Implications for finding data

- Heuristics for the SAP
- Geographic information is most likely to be found on a server located within its footprint
  - convergence of **d** and **s**
  - but at what level in the hierarchy?
    - jurisdiction that most closely matches the footprint
    - max $||J \cap F|| / (||J||\ ||F||)^{1/2}$

Transitioning map libraries

- From central services to unique services
  - from general collections to special collections
  - from UCSB's Map and Imagery Laboratory to the Alexandria Digital Library
- There will always be more than one service
  - no amazon.com of geographic information
    - www.alexandria.ucsb.edu
    - www.geographynetwork.com
    - www.fgdc.gov
- Unique services must declare themselves
  - through collection-level metadata (CLM)
  - formalizing and publishing **d**

EXANDRIA DIGITAL LIBRARY

CDL CALIFORNIA DIGITAL LIBRARY

ADL HOME · CDL HOME · HELP

NICK PLACENAME SEARCH

earch the entire world for...

Find

er "Rome" if you want Rome, Italy.
e information

NERAL SEARCH

Select collection to search

DL Catalog

ose collections

Set geographic region

the map to the right to set the geographic
ent of the search, or directly enter bounding
rdinates below.

N 34.43

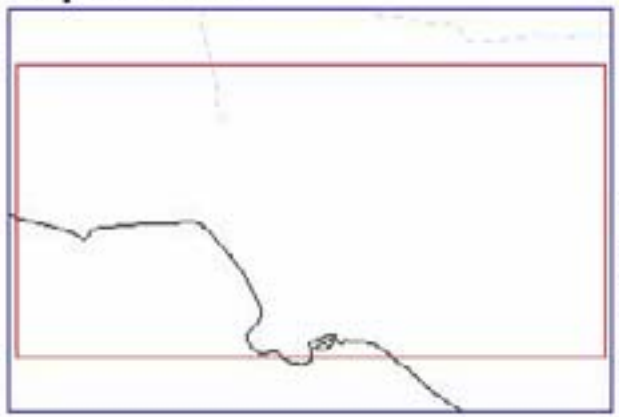
W -118.97 -117.57 E

33.73 S

Words to search for

Any of the above words
All of the above words
Exact phrase

Map Browser

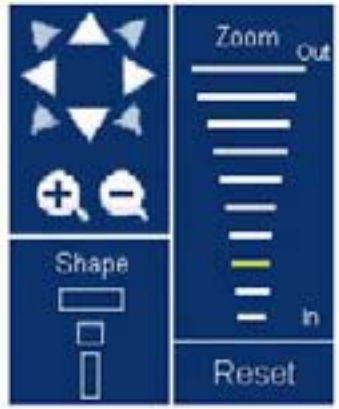


Click map to:

Recenter & Zoom in

Change location to:

--



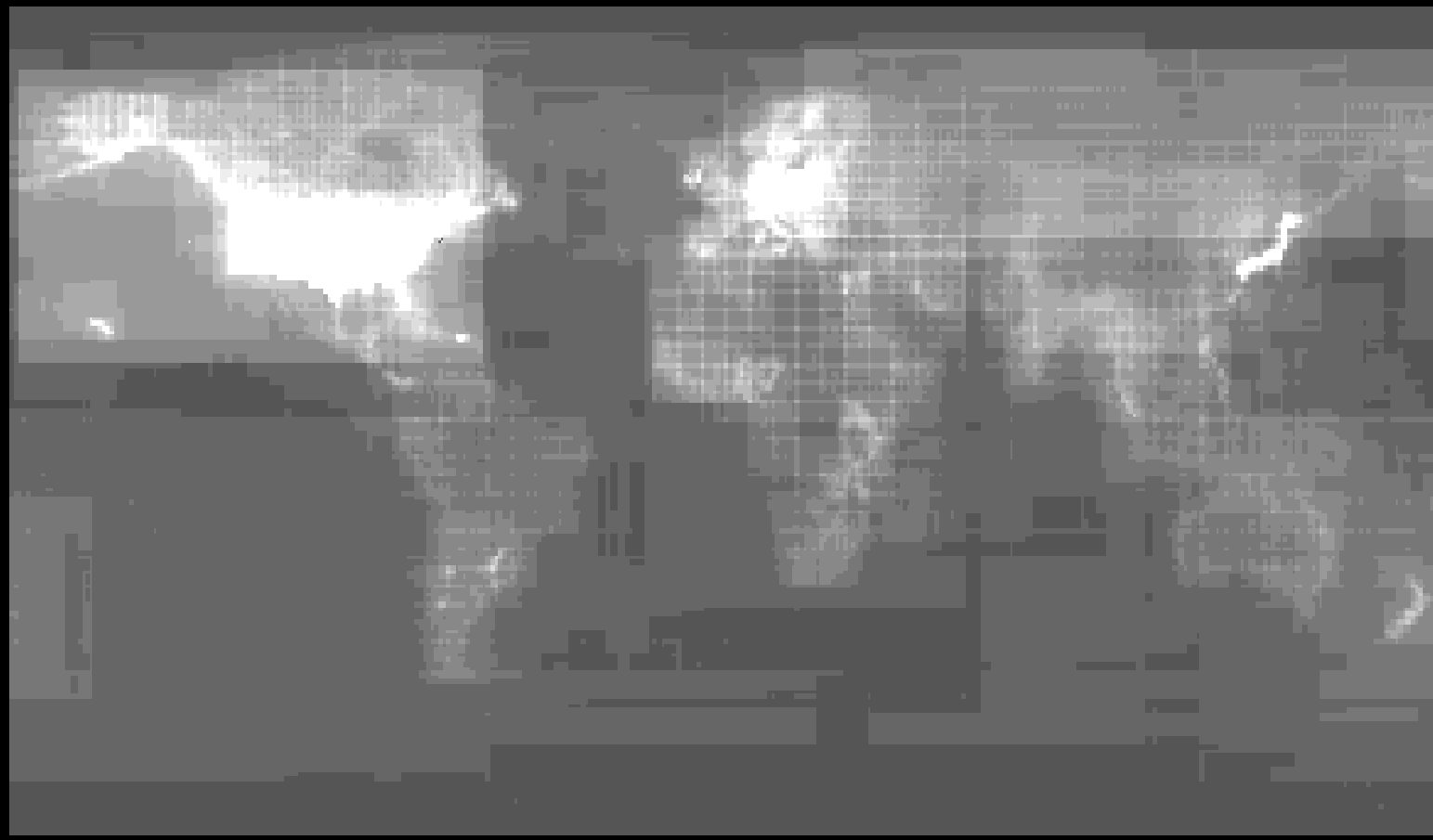
ADL Search Results

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

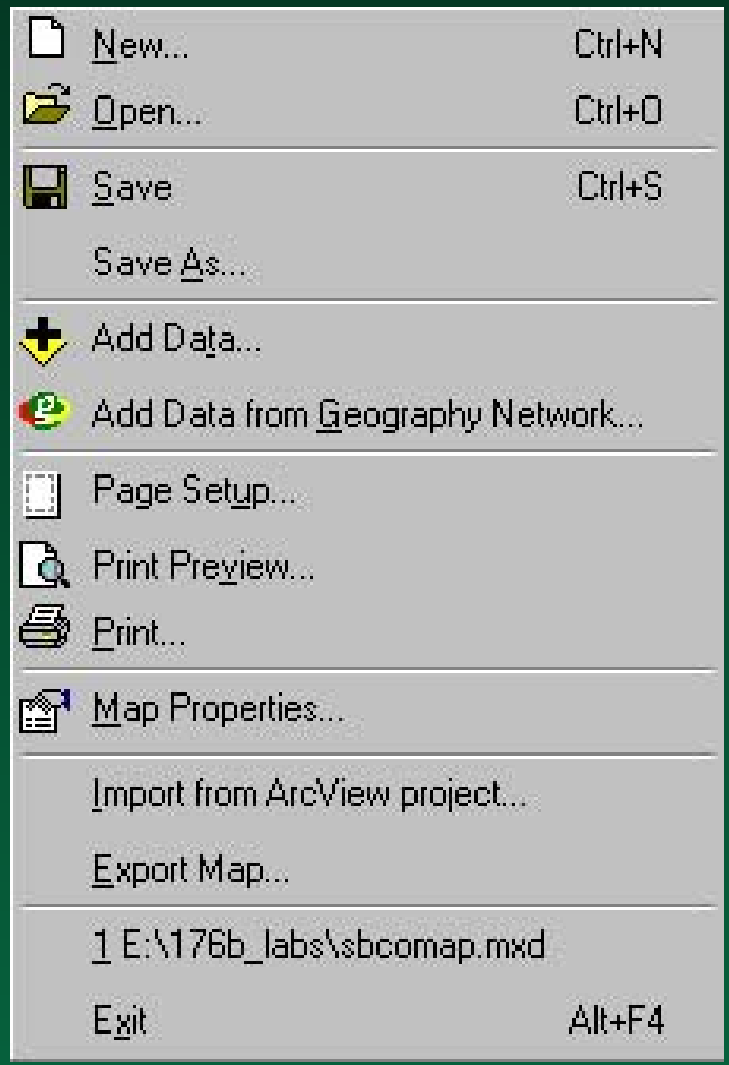
3. **DRG o33117g8, Digital Raster Graphic of ANAHEIM, CA.**
   Type: maps. **Format:** TIFF. **Date:** 1981. **ADL identifier:** adl_catalog:800281.
   HIGHLIGHT IN MAP · COMPLETE DESCRIPTION · BROWSE GRAPHIC · ACCESS/DOWNLOAD



CLM of the Alexandria Digital Library

Knowing where to look

- Approaches to CLM
  - by data type
    - ortho.mit.edu
  - by area of the globe
    - Arctic Data Directory
  - the one stop shop
    - www.fgdc.gov
  - a new generation of search engines
    - identifying footprints


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

ABOUT

MAPS

DATA

GEOSERVICES

SOLUTIONS

COMMUNITY

Geography Network Explorer

Free Resources

Be a Publisher

Challenge

SEARCH & VIEW

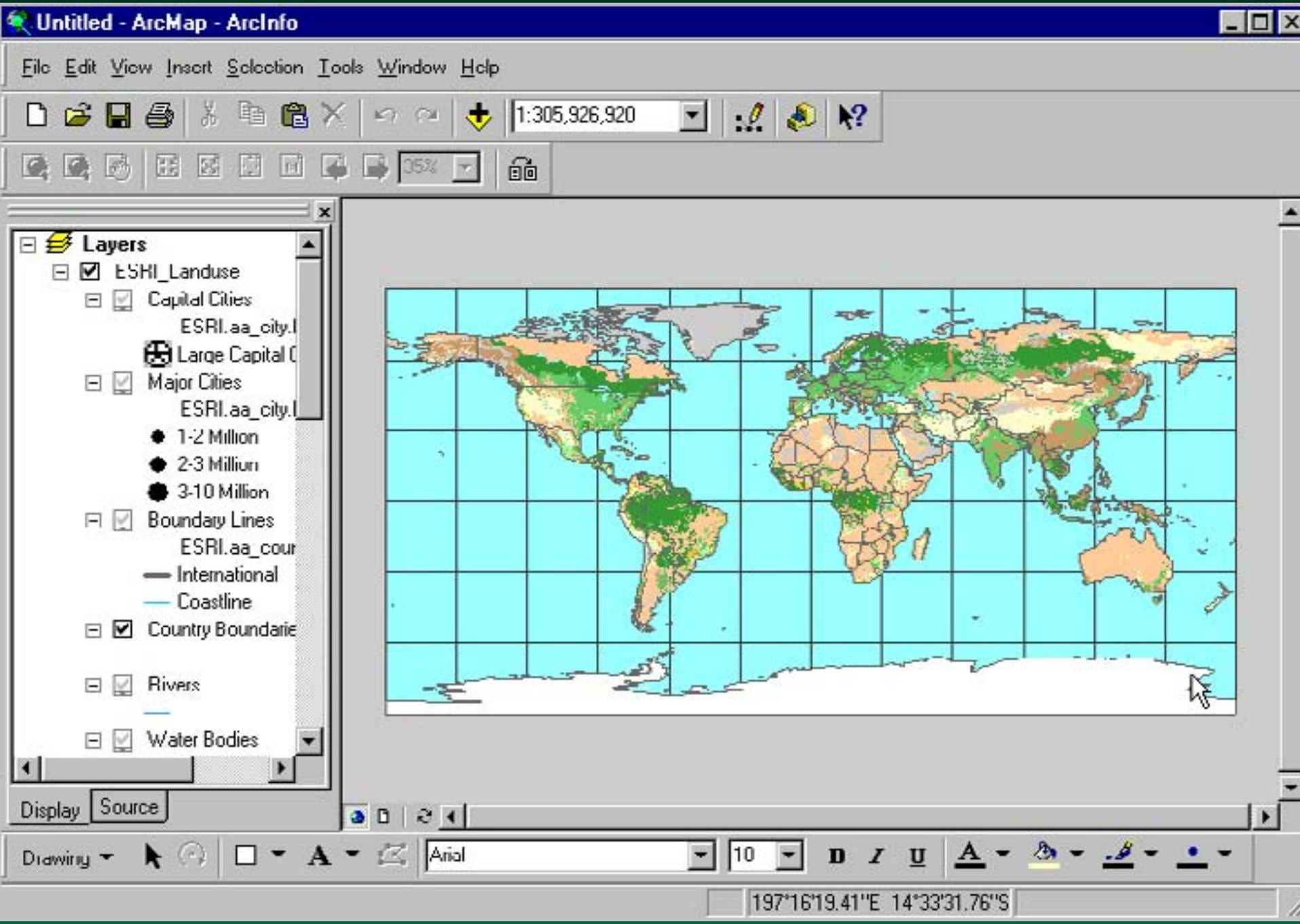
use the **Geography Network Explorer** to search and view maps and other geographic content over the Internet

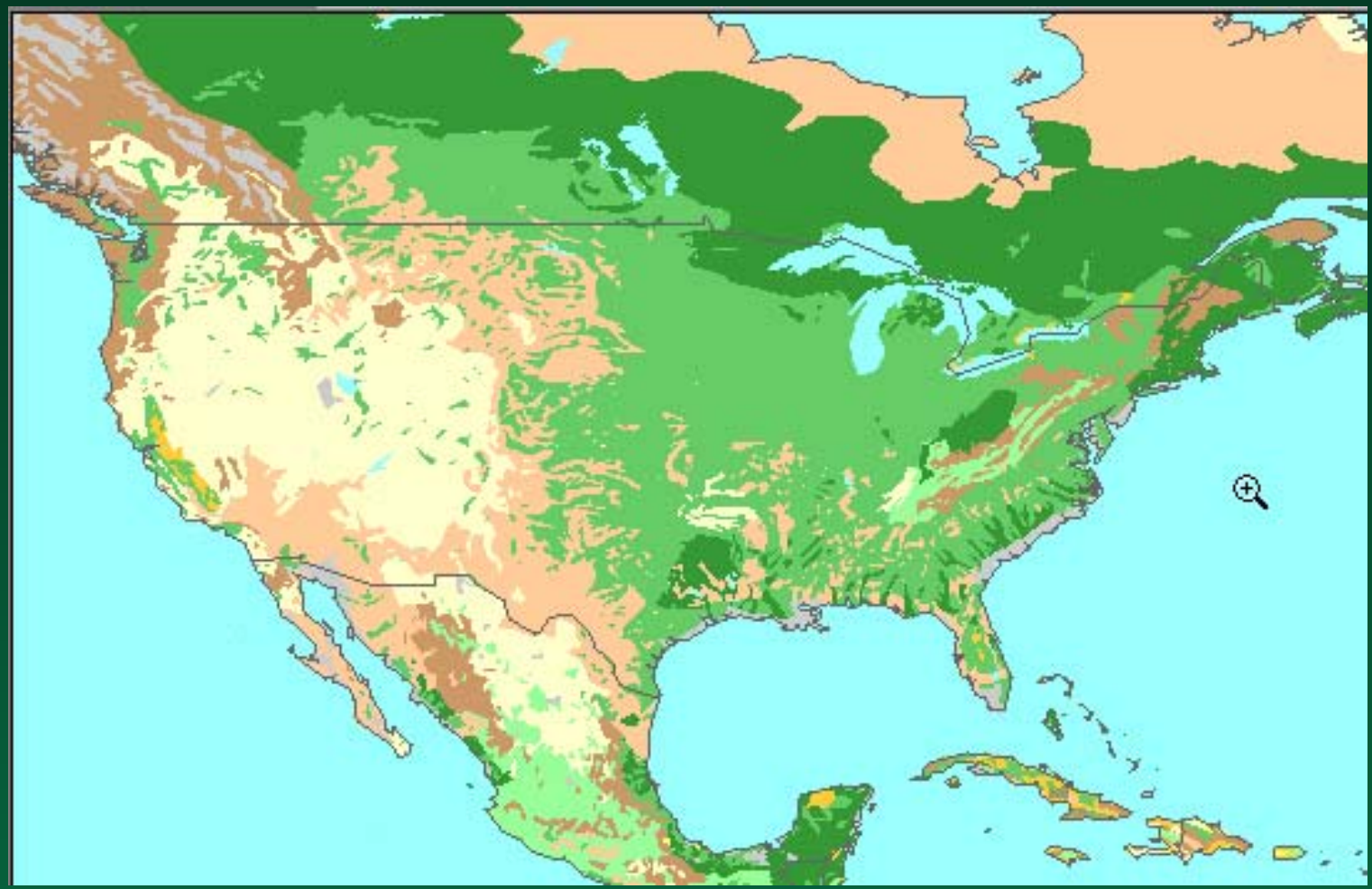
The Geography Network is a global community of data providers who are committed to making geographic content available. This content is published from many sites around the world, providing you immediate access to the latest maps, data, and related services. This portal to the Geography Network enables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

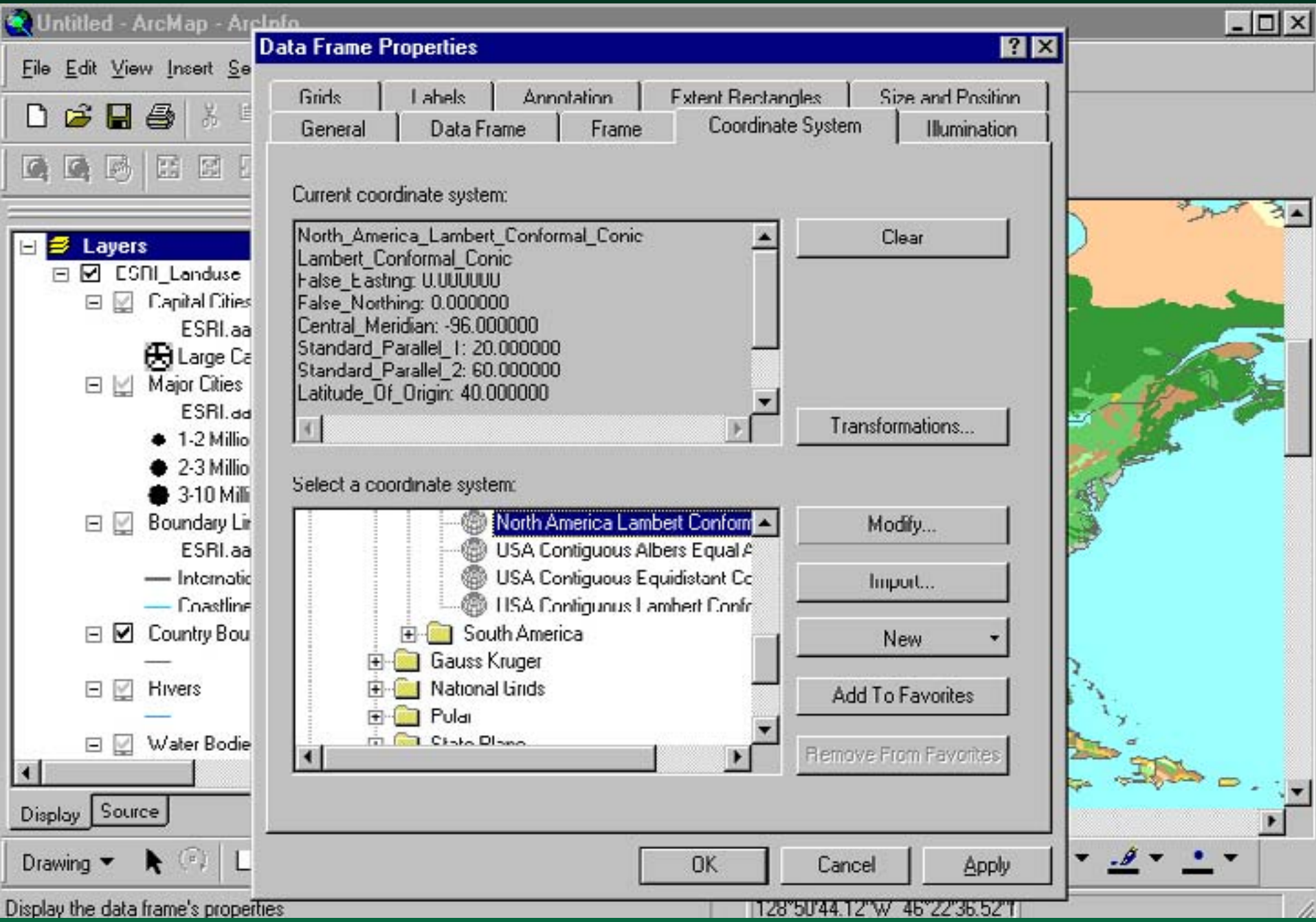
Featured Content

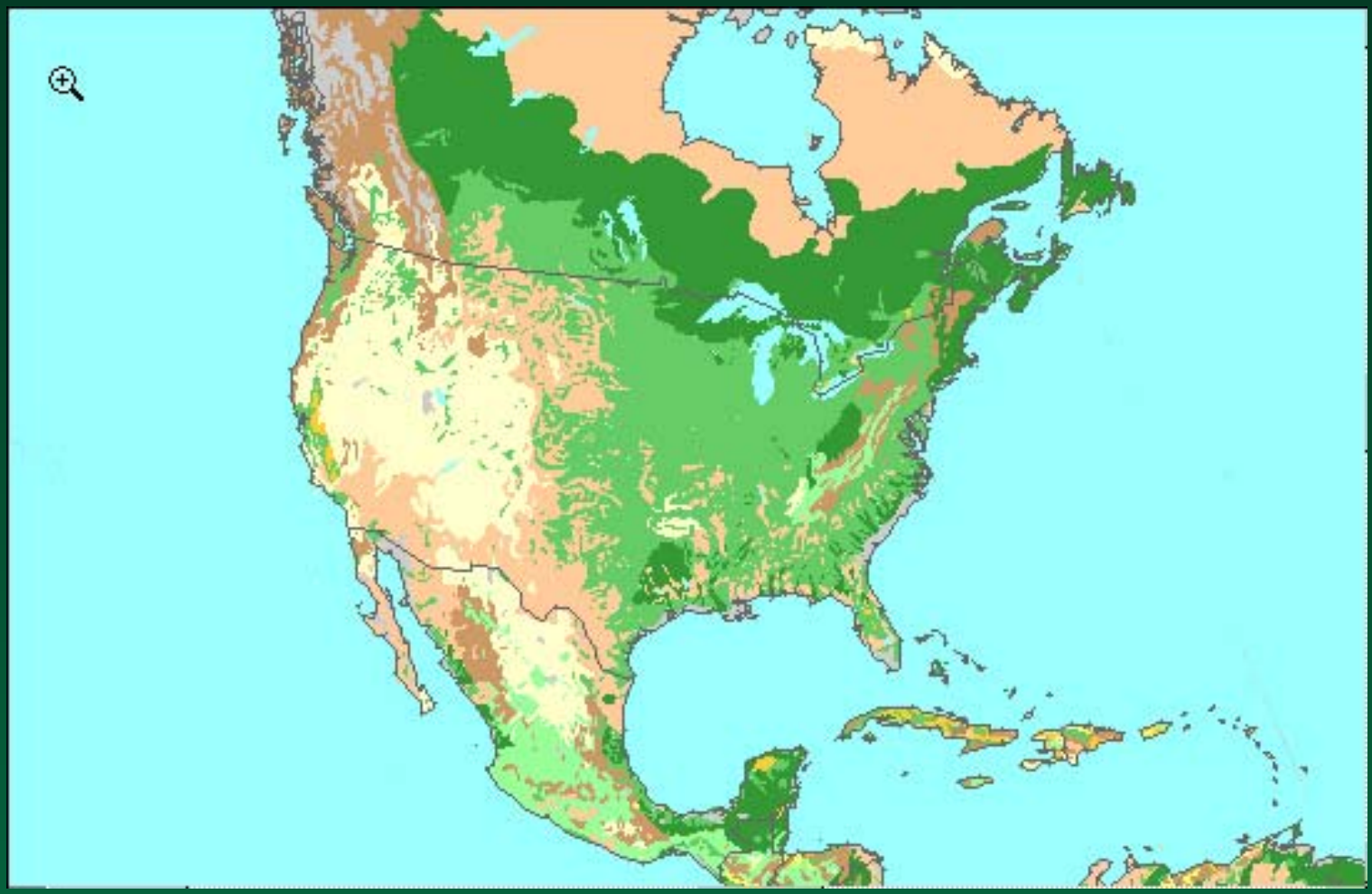
U.S. Census TIGER 2000

- 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.
Large Capital
Major Cities
ESRI.aa_city.
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 Se
Layers
ESRI_Landuse
Capital Cities
ESRI.aa
Large Ca
Major Cities
ESRI.aa
1-2 Millio
2-3 Millio
3-10 Milli
Boundary Li
ESRI.aa
Internatic
Coastline
Country Bou
Rivers
Water Bodie
Display
Source
Drawing
Display the data frame's properties
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 Cc
USA Contiguous Lambert Confr
South America
Gauss Kruger
National Grids
Polar
State Plane
Modify...
Import...
New
Add To Favorites
Remove From Favorites
OK
Cancel
Apply
128°50'44.12"W 46°22'36.52"N



Options for **p**

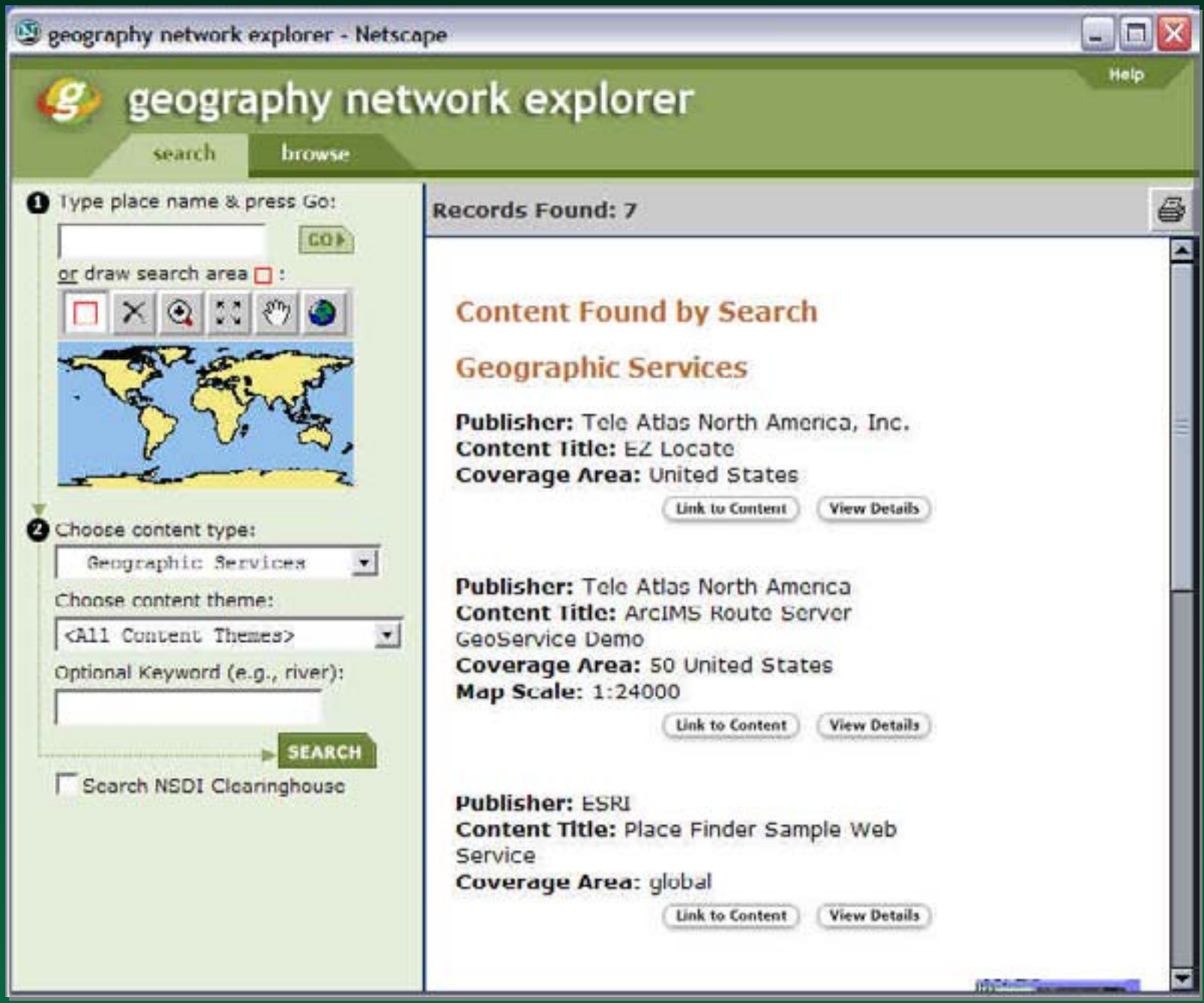
- Where to process?
  - server or client, which server?
  - published services
    - directories
    - www.geographynetwork.com
    - evolving g.net
  - description standards
    - UDDI: Universal Description, Discovery and Integration
    - WSDL: Web Service Definition Language

p and u

- $||\mathbf{p}-\mathbf{u}|| = 0$
  - computing in the client
    - using local data, $||\mathbf{u}-\mathbf{d}|| = 0$
    - using remote data
- $||\mathbf{p}-\mathbf{u}|| > 0$
  - send data to the service from the client
  - link a remote service to a remote data source, $\mathbf{p} \neq \mathbf{u}$, $\mathbf{d} \neq \mathbf{u}$

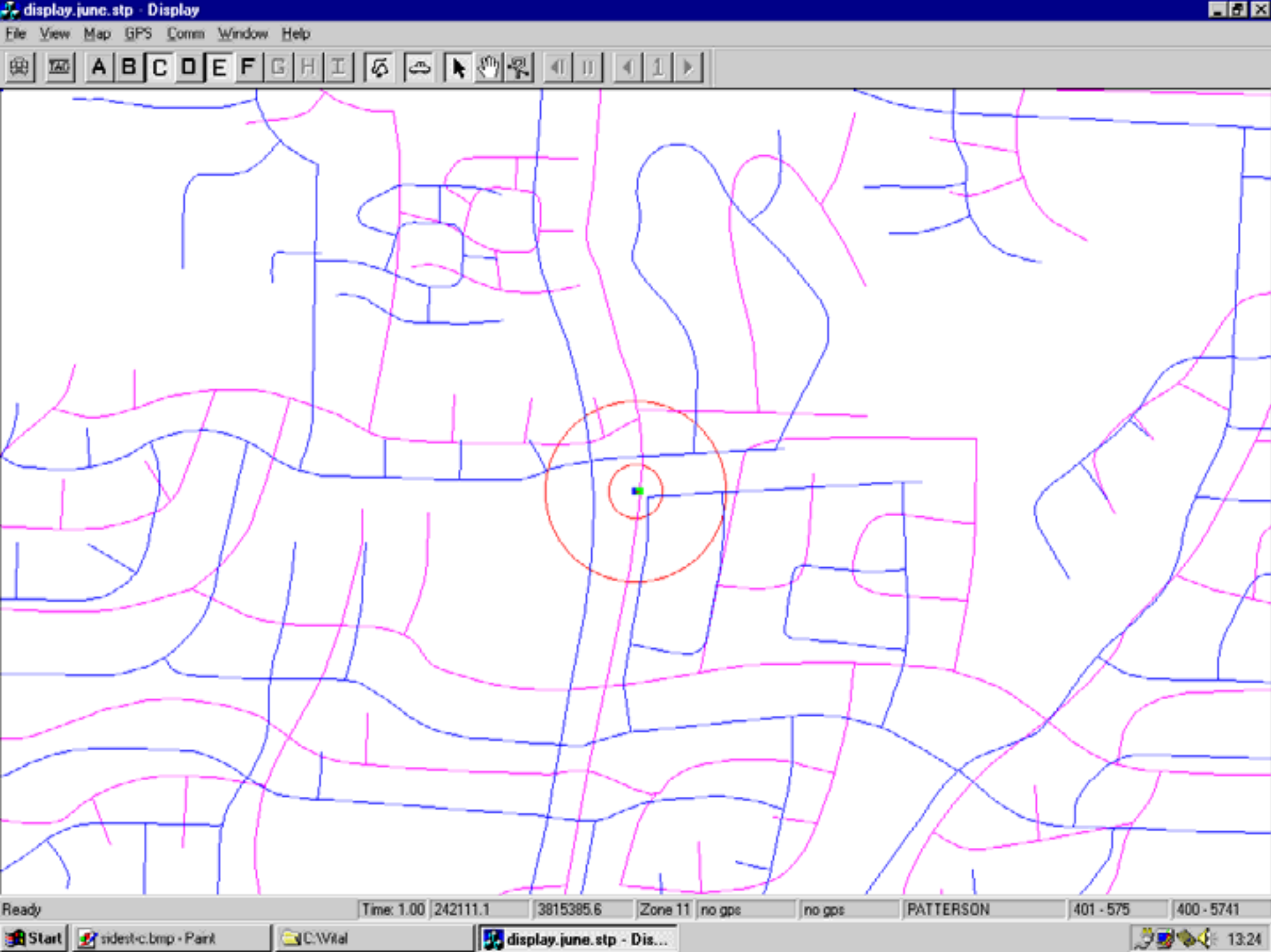
Costs and benefits

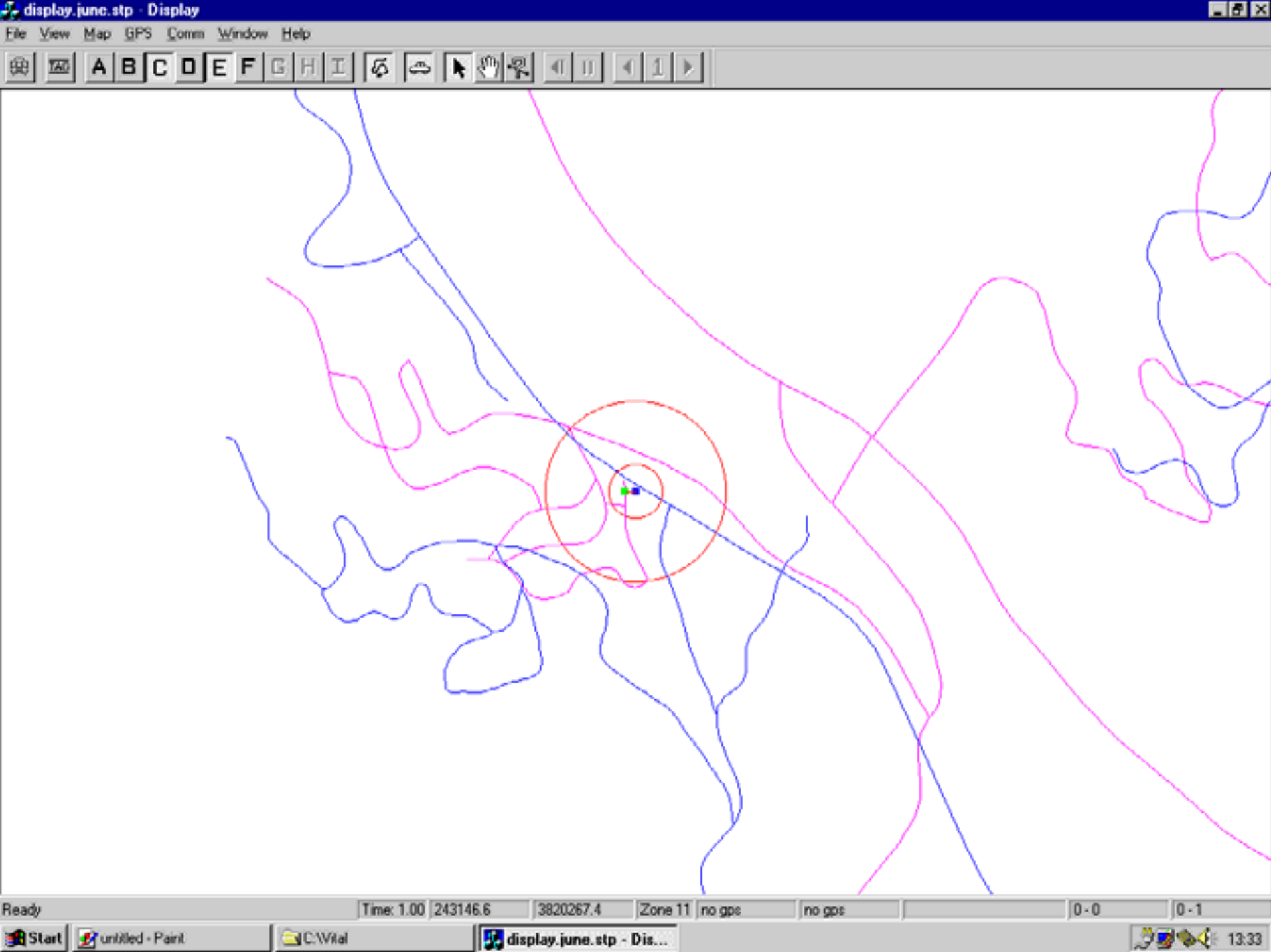
- More cycles available remotely
  - integrating and exploiting waste cycles
  - the Grid
  - SETI
- Intellectual property issues
  - intellectual value of service
  - risk of dissemination
  - commercial value
- Update, versioning issues
  - distributed service has versioning problems
- Process coupled to data, well defined


geography network explorer - Netscape
geography network explorer
Help
search
browse
1 Type place name & press Go:
GO
or draw search area :
2 Choose content type:
Geographic Services
Choose content theme:
<All Content Themes>
Optional Keyword (e.g., river):
SEARCH
Search NSDI Clearinghouse
Records Found: 7
Content Found by Search
Geographic Services
Publisher: Tele Atlas North America, Inc.
Content Title: EZ Locate
Coverage Area: United States
Link to Content
View Details
Publisher: Tele Atlas North America
Content Title: ArcIMS Route Server GeoService Demo
Coverage Area: 50 United States
Map Scale: 1:24000
Link to Content
View Details
Publisher: ESRI
Content Title: Place Finder Sample Web Service
Coverage Area: global
Link to Content
View Details

High-priority geoservices

- Geocoding
  - tied to data, update issue
- Gazetteer
  - conversion between general or domain-specific placename and coordinates
  - geoparsing
    - identification and decoding of placename references in text
    - mapping and associating news stories
  - queries based on placenames
    - how far is the capital of Belgium from the capital of France?
- What else, is there a general model?

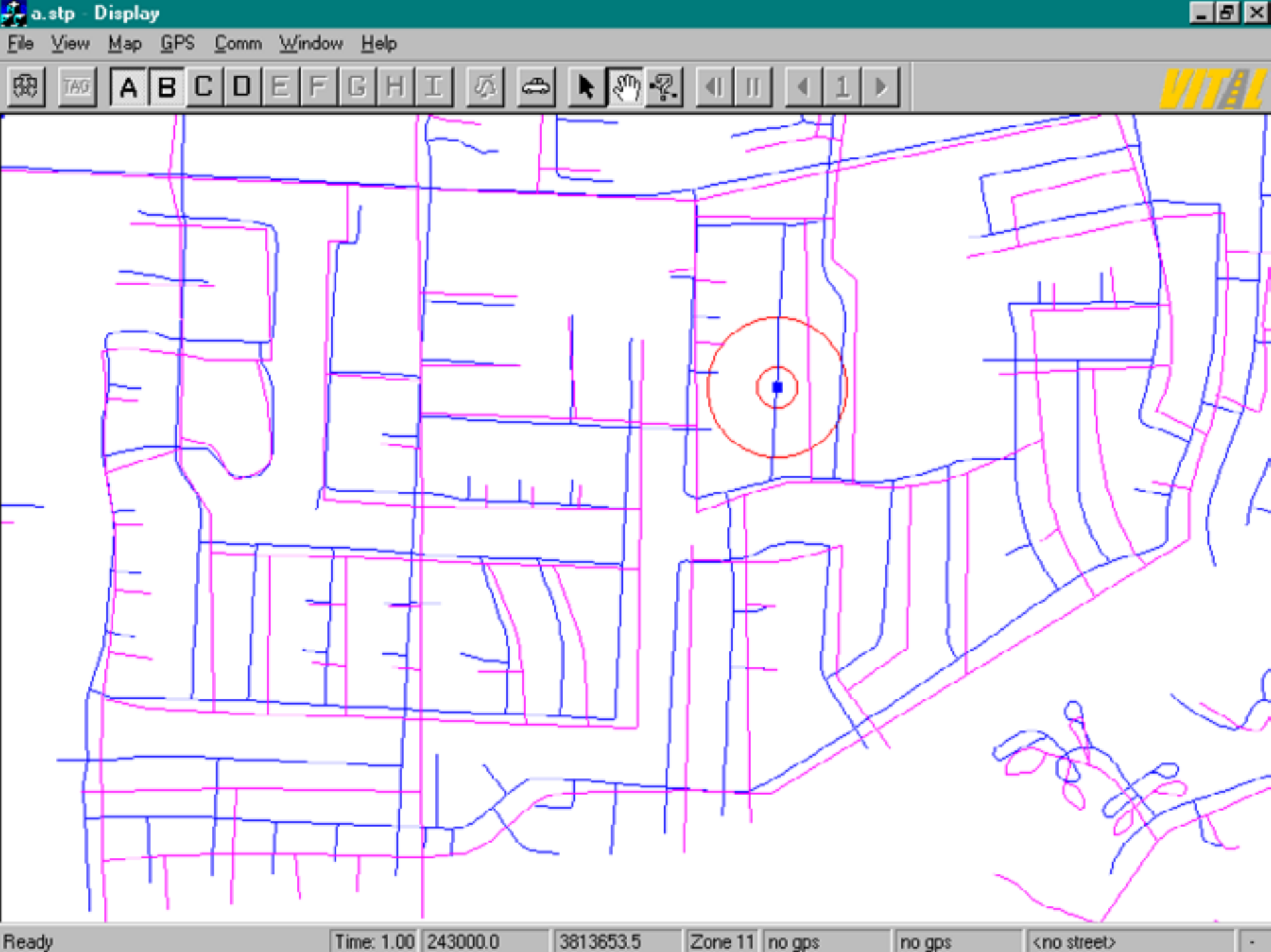

display.june.stp - Display
File View Map GPS Comm Window Help
A B C D E F G H I
Ready
Time: 1.00
242111.1
3815385.6
Zone 11
no gps
no gps
PATTERSON
401 - 575
400 - 5741
Start
sidest-c.bmp - Paint
C:\Vital
display.june.stp - Dis...
13:24

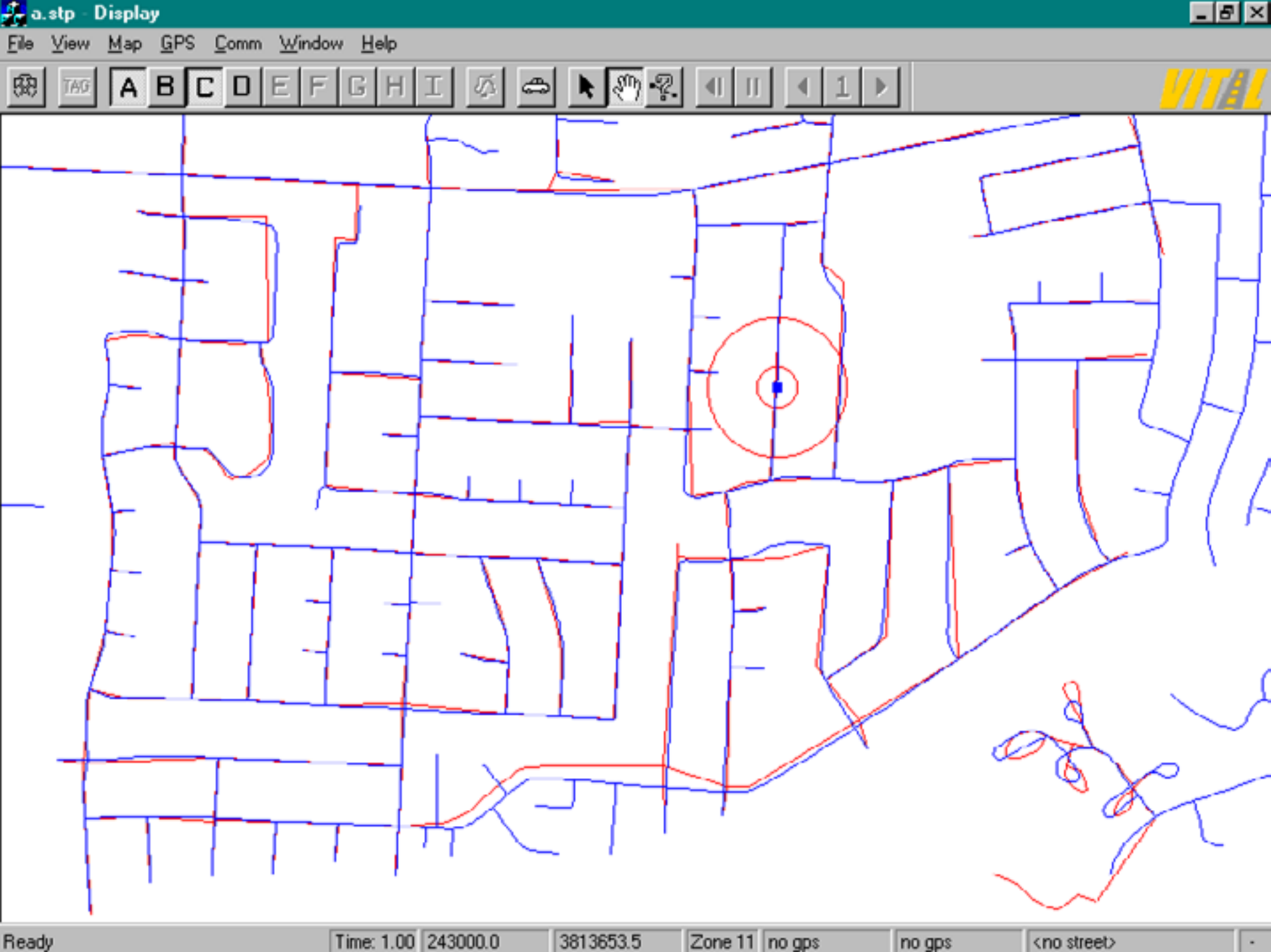

display.june.stp - Display
File View Map GPS Comm Window Help
A B C D E F G H I
Ready
Time: 1.00
243146.6
3820267.4
Zone 11
no gps
no gps
0 - 0
0 - 1
Start
untitled - Paint
C:\Vital
display.june.stp - Dis...
13:33


display.june.stp - Display
File View Map GPS Comm Window Help
A B C D E F G H I
Ready
Time: 1.00
241697.0
3814757.1
Zone 11
no gps
no gps
CLARENCE WARD MEMC
0 - 0
0 - 0
Start
fwy-ce.bmp - Paint
C:\Vital
display.june.stp - Dis...
13:19

A counter-example

- Fusing or conflating independent but overlapping sources
  - neither source will admit error
  - there is no truth
  - problem is binary not unary
- Service must be performed at the client
  - even by mobile clients


a.stp - Display
File View Map GPS Comm Window Help
A B C D E F G H I
1
VITAL
Ready
Time: 1.00
243000.0
3813653.5
Zone 11
no gps
no gps
<no street>


a.stp - Display
File View Map GPS Comm Window Help
TAG
A B C D E F G H I
1
VITAL
Ready
Time: 1.00
243000.0
3813653.5
Zone 11
no gps
no gps
<no street>

OPINION
WORLD
BUSINESS
FINANCE & ECONOMICS
SCIENCE & TECHNOLOGY
PEOPLE
BOOKS & ARTS
MARKETS & DATA
DIVERSIONS

CITIES GUIDE
COUNTRY BRIEFINGS

Economist Intelligence Unit
onlinestore

RESEARCH TOOLS
Articles by subject
Backgrounders
Surveys
Economics A-Z
Style guide
Business encyclopedia

GLOBAL EXECUTIVE
Executive Thinking
Business Education

PRINT EDITION
The Economist
Extinction of the

THE INTERNET

Printable page
E-mail this
Discuss this

The revenge of geography

Mar 13th 2003
From The Economist print edition

It was naive to imagine that the global reach of the internet would make geography irrelevant. Wireline and wireless technologies have bound the virtual and physical worlds closer than ever

IN THE early days of the internet boom, there was much talk of the "death of distance". The emergence of a global digital network, it seemed, would put an end to mundane physical or geographical constraints. There was some truth in this. E-mail made it cheap and easy to stay in constant touch with people, whether they lived around the corner or on the other side of the globe. Companies could communicate with customers and employees no matter where they were. And like-minded individuals who shared a common interest could get together online from all round the world.

Actually, geography is far from dead. Although it is often helpful to think of the internet as a parallel digital universe, or an omnipresent "cloud", its users live in the real world where limitations of geography still apply. And these limitations extend online. Finding information relevant to a particular place, or the location associated with a specific piece of information, is not always easy. This has caused a surge of innovation, as new technologies have developed to link places on the internet with places in the real world—stitching together the supposedly separate virtual and physical worlds

The Revenge of Geography

- *Economist* 3/15/03
  - the virtual and physical worlds are increasingly correlated
- Physical distance important in the virtual world
  - physical location allows determination of physical distance
  - hence physical location is important
- Physical location allows integration

Why is distance important?

- Virtual and physical interaction are complementary
  - telecommuters must visit the office
  - email contacts may originate in physical contacts
  - online orders must be filled physically
  - virtual contacts transition to physical contacts

Determining location

- Most computers do not know where they are
  - time zone defines a range of longitude
- Direct measurement
  - GPS
  - cellphone location measurement
  - WiFi, Bluetooth, …
- Input by user or system builder
  - coordinates
  - placename plus gazetteer

Inference about IP

- IP registration record
- Mining text for addresses
- Commercial incentive
  - targeted advertising, spam
  - biased search engines
- Military/intelligence incentive
  - email to a polygon
  - sourcing intelligence

The business of geolocation

- Quova: http://www.quova.com/services/services.html
- Digital Envoy: http://www.digitalenvoy.net/
- NetGeo: http://www.caida.org/tools/utilities/netgeo/
- InfoSplit: http://www.infosplit.com/

QUOVA®

Services | Solutions | Company | Partnerships | Press Room

Home > Services > Geo-location

Geo-location

Quova offers the industry's most comprehensive and sophisticated approach to providing geographic location information.

Our flagship service, GeoPoint, is an enterprise-class solution, offering superior data, accuracy and coverage.

Our GeoTraffic and GeoProfile services add a geographic dimension to other traffic analysis services and complement your own analyses.

Services

- GeoPoint
- GeoTraffic
- GeoProfile

Call Me! Let's Talk >

Why Location Matters >


InfoSplit
we know where your customers are
about infosplit
products
press room
career center
contact us
demo
Country:United States
State:CALIFORNIA
Metro area:SANTA BARBARA-S.
Infosplit's objective is to offer an accurate geographic profiling solution.
Our patent-pending technology consists in mapping the Internet as precisely as possible.
By growing our database, we improve the accuracy of our data day after day.
Design by
Legal Disclaimer
music On

And…

- Bandwidth is not infinite
  - latency is not zero
  - sites are often mirrored
  - www.google.com redirected for non-US IP
    - also for reasons of language
    - Italian-language site
- Interest in sites is likely centered on **u**
  - e.g., query for restaurants

And finally…

- Scalability
  - on a Web of *n* users/sites interaction rises as n^2
  - but on a Web partitioned spatially into *m* zones with *n*/*m* users per zone the interaction is only $m\ (n/m)^2 = n^2/m$
- Conclusion: the Web is in part spatially organized *and growing more so*

The future spatial Web

- Computing embedded everywhere
  - ground-based sensor networks
    - nano-scale dust
  - networks of space-borne sensors, UAVs
  - mobile, ubiquitous computing
    - GIS capabilities everywhere
  - infinite wireless bandwidth
    - living in a soup of radiation

Associated technologies

- Grid computing
  - harvested cycles
  - distributed services and data
    - transparent access
    - search engines, metadata
- Interoperability
  - semantic tools
  - ontologies

But…

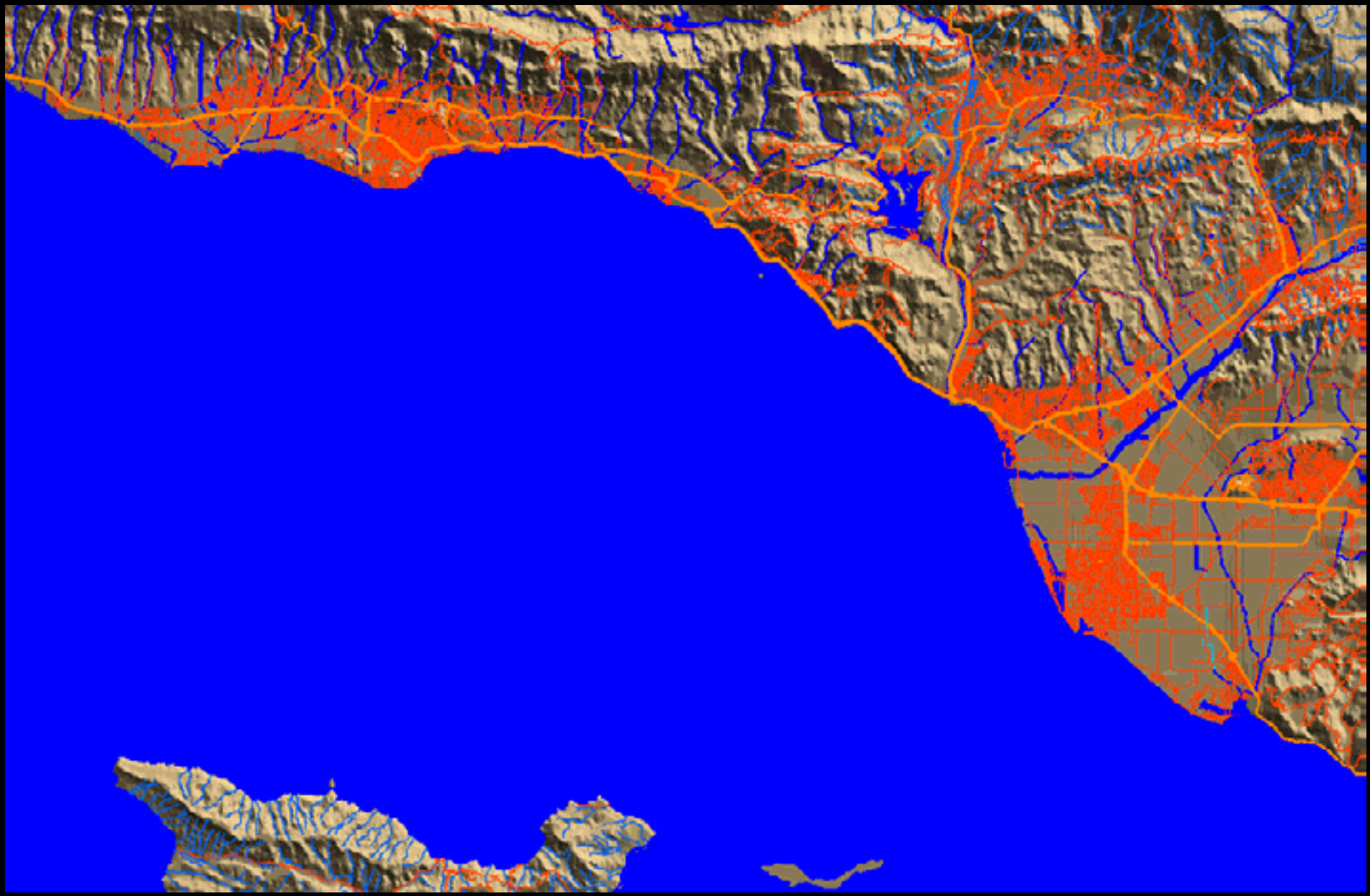
- Overlapping metadata domains
  - FGDC for geospatial
  - EML for ecological data
  - DDI for social data
    - but EML, DDI have spatial components
    - impossible to define non-overlapping domains
  - metadata light as the umbrella ontology
    - Dublin Core

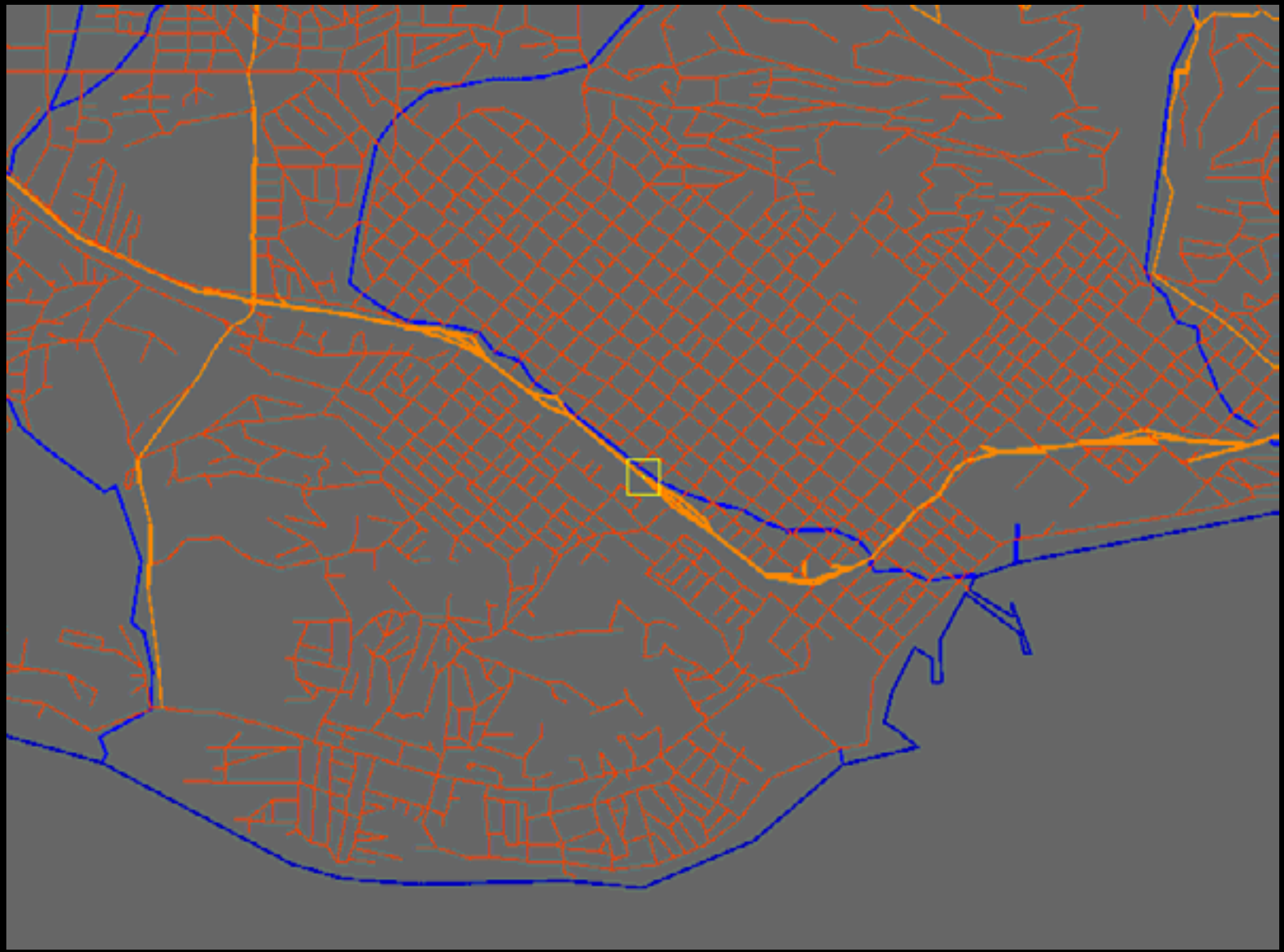
and…

- Mapping between ontologies, semantics
  - successful if 1:1, 1:*n*, *n*:1
  - but not if partial, fuzzy, uncertain
  - if it was simple enough to be automated it would have been done years ago
- The CLM issue
- The cost/benefit ratio for metadata
- IP address space
  - 2 billion assigned of 4 billion possibilities
    - 32 bit
  - IPv6 is on the way
    - 64 bit

A grand challenge of GIS

- To create useful, comprehensive digital representations of the enormous complexity of the Earth's surface in the limited space of a digital store, using a binary alphabet
- An integrated, coherent organization of geographic information





A virtual Earth

- A representation of form
  - distributed, seamless, vertically integrated
- Representations of process
  - dynamic simulation models
  - integrated with the data
- Integrated with visualization, analysis clients

“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.”

Is Digital Earth feasible?

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

[illegible]

The LS ratio

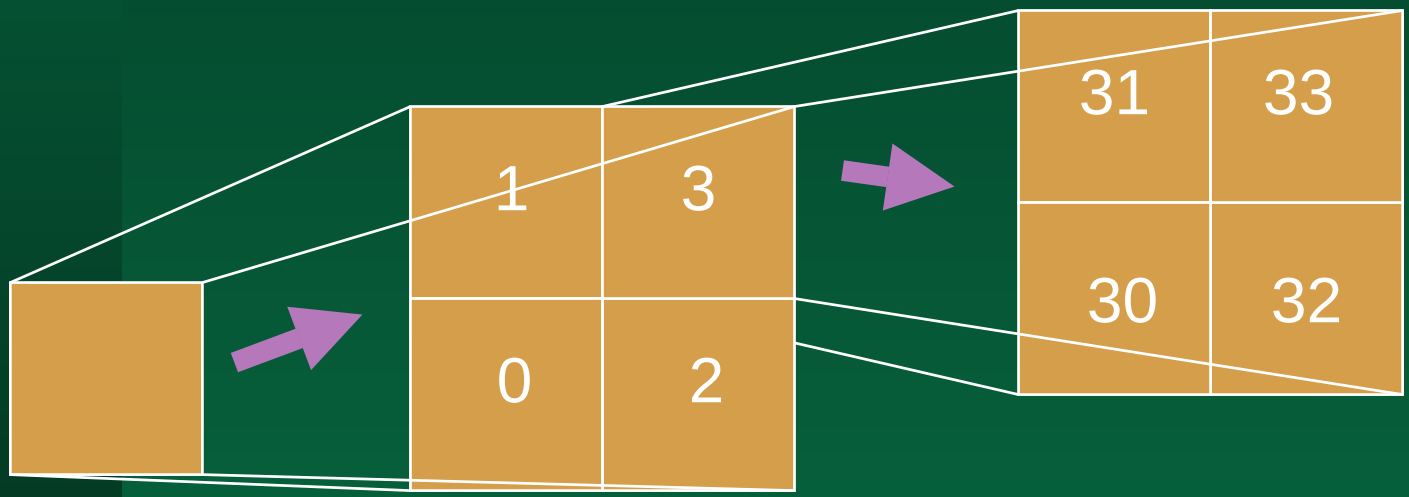
- Computer screen - 1000
- Digital camera - 1500
- Remotely sensed scene - 3000
- Paper map - 5000
- Dimensionless
- $Log_{10}L/S$ in range 3-4
- Human eye - 10,000

A data structure for DE

- To support smooth zooming over 4 orders of magnitude resolution
  - from 10km to 1m
  - maintaining LS ratio
- Vertically integrated
  - multiple layers

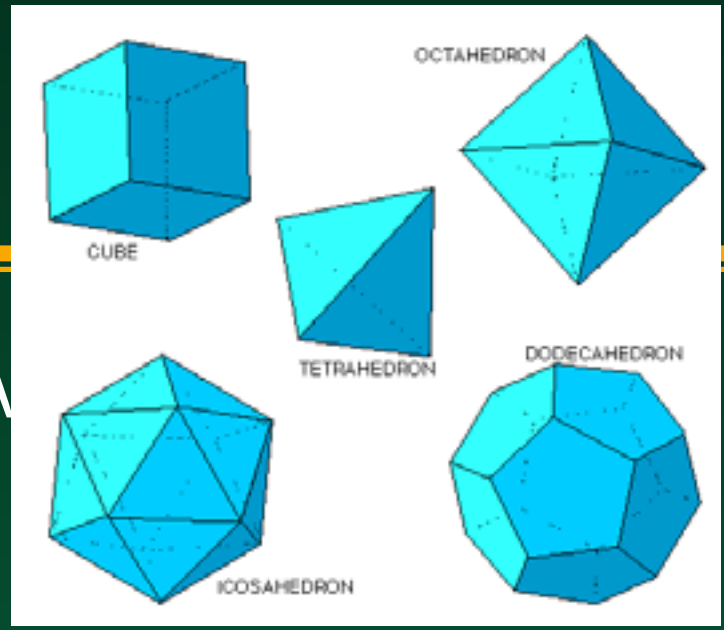
The quadtree

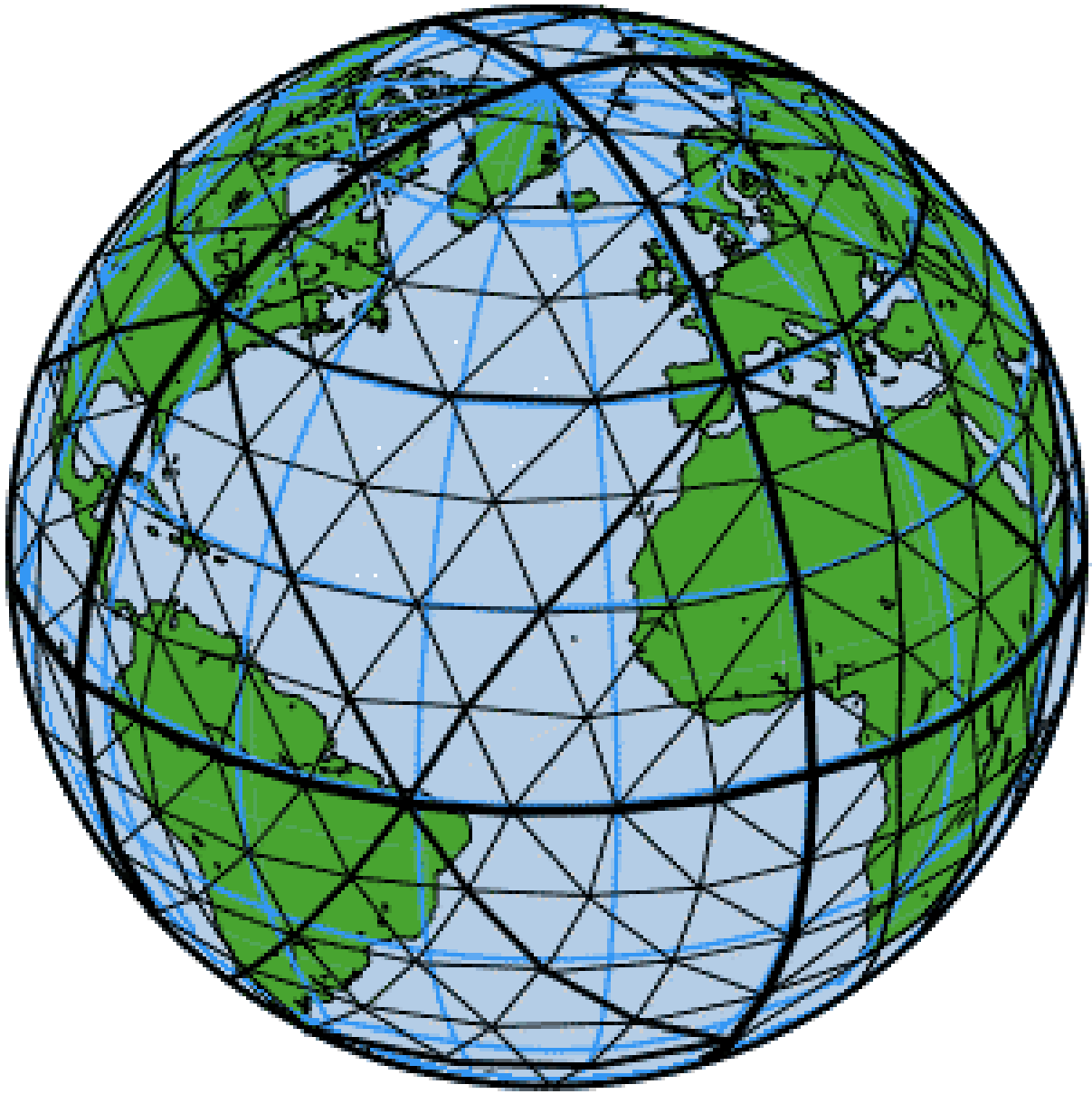
- Recursive subdivision
  - variable depth depending on local detail

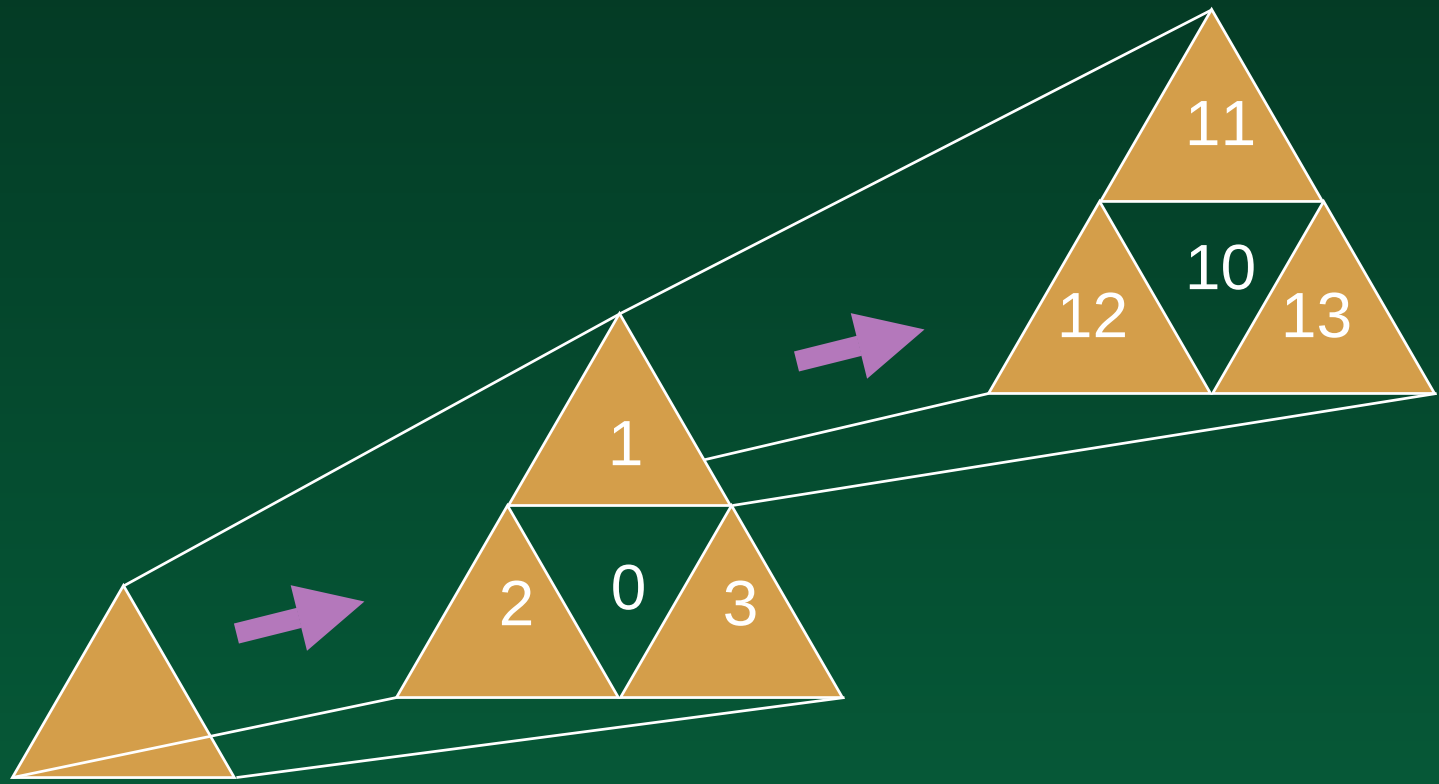


Grids on the globe

- Impossible to tile a curv
  squares
- Five Platonic solids
  - tetrahedron: 4 triangles
  - cube: 6 squares
  - octahedron: 8 triangles
  - dodecahedron: 12 pentagons
  - icosahedron: 20 triangles







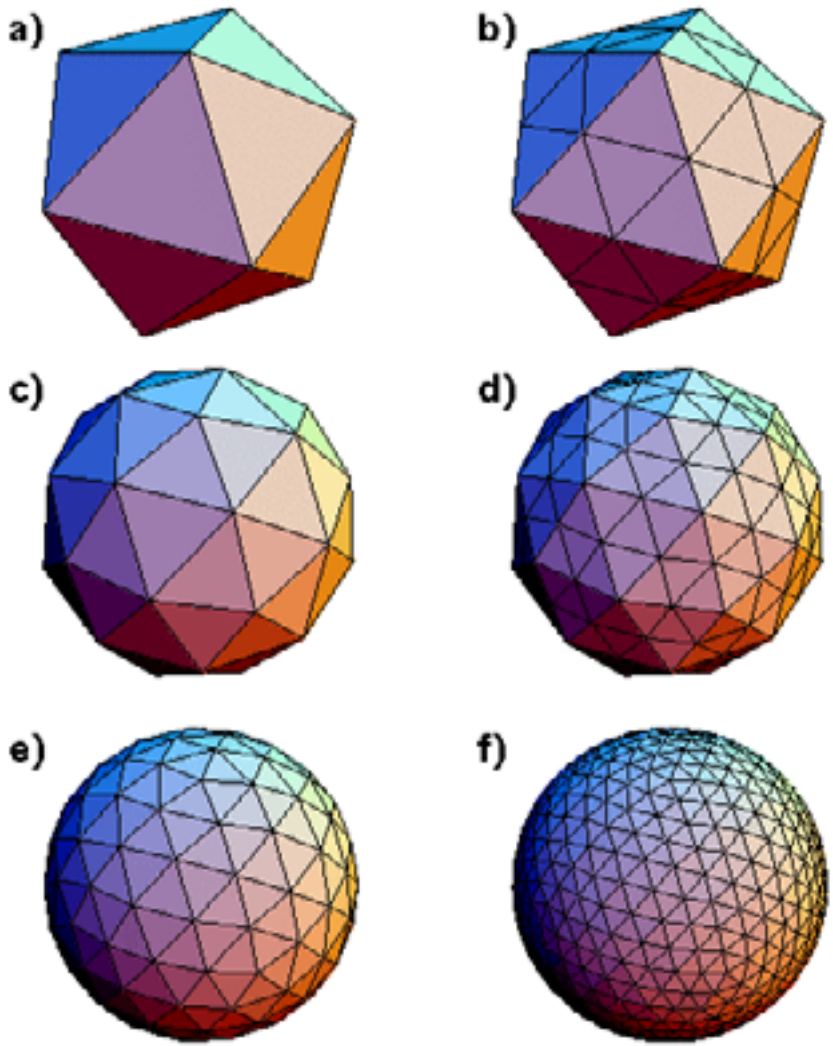
Octahedron: 1 base 8 digit plus unlimited base 4 digits

Discrete global grid based on the Icosahedron (20 triangles, 1:4 recursive subdivision)

Ross Heikes and David Randall, Colorado State University

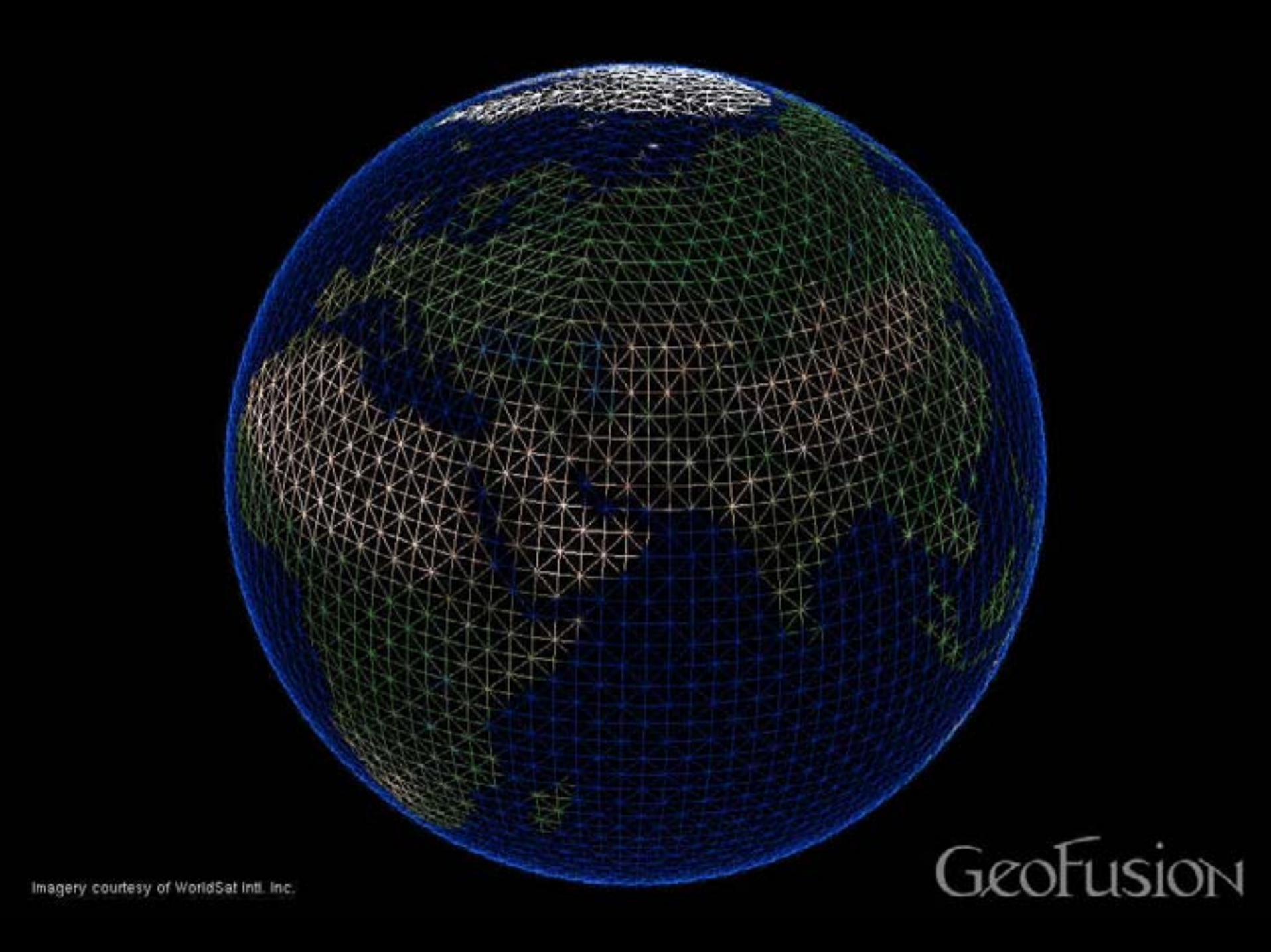
Construction of a simple Icosahedral grid

a) Suppose we have an icosahedron inscribed inside of a unit sphere.
b) Bisecting each edge forms 30 new vertices, and partitions each equilateral face into four pieces.
c) Project the new vertices onto the unit sphere.
d) Bisect and partition again.
e) Project again.
f) And so on.... The result is a sequence of polyhedrons that increasingly approximate the sphere.



Comparison of Criteria for the Assessment of Global Grids

Criteria in Goodchild (1994)	*Criteria in Kimerling et al. (1999) (Goodchild's Numbers given in parentheses)*
1. Each area contains one point	Areal cells constitute a complete tiling of the globe, exhaustively covering the globe without overlapping. (3,7)
2. Areas are equal in size	Areal cells have equal areas. This minimizes the confounding effects of area variation in analysis, and provides equal probabilities for sampling designs. (2)
3. Areas exhaustively cover the domain	Areal cells have the same topology (same number of edges and vertices). (9, 14)
4. Areas are equal in shape	Areal cells have the same shape. ideally a regular spherical polygon with edges that are great circles. (4)
5. Points form a hierarchy preserving some property for m < n points	Areal cells are compact. (10)
6. Areas form a hierarchy preserving some property for m < n areas	Edges of cells are straight in a projection. (8)
7. The domain is the globe (sphere, spheroid)	The midpoint of an arc connecting two adjacent cells coincides with the midpoint of the edge between the two cells.
8. Edges of areas are straight on some projection	The points and areal cells of the various resolution grids which constitute the grid system form a hierarchy which displays a high degree of regularity. (5,6)
9. Areas have the same number of edges	A single areal cell contains only one grid reference point.(1)
10. Areas are compact	Grid reference points are maximally central within areal cells. (11)
11. Points are maximally central within areas	Grid reference points are equidistant from their neighbors. (12)
12. Points are equidistant	Grid reference points and areal cells display regularities and other properties which allow them to be addressed in an efficient manner.
13. Edges are areas of equal length	The grid system has a simple relationship to latitude and longitude.
14. Addresses of points and areas are regular and reflect other properties	The grid system contains grids of any arbitrary defined spatial resolution. (5,6)


Imagery courtesy of WorldSat Intl. Inc.
GeoFusion

Imagery courtesy of WorldSat Intl. Inc.

GeoFusion

Some take-home messages

- Geography is having its revenge
  - cyberspace is spatial after all, and maps to real space
- There are four well-defined locations in GIS
  - and they interact in interesting ways, defining different application domains
- $\log_{10}$ L/S is an important design parameter
  - limiting the amount of data needed by an application at any time

GIScience and the Grid

- The Grid is becoming shorthand for fully integrated, distributed, interoperable services and data
- The GI community is ideally poised for a major new initiative to exploit Grid computing and collaborative technologies
  - how to prioritize services
  - how to achieve interoperability, transparency over the Grid
  - how to integrate at **s**
  - how to integrate geoservices, geodata into application domains that exploit the Grid