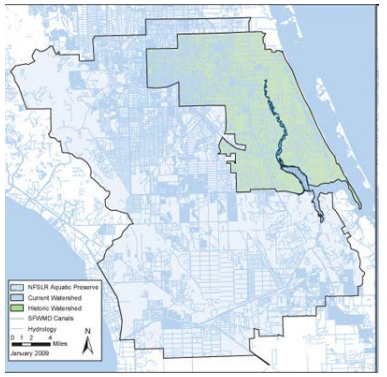
Geography 128 Winter Quarter 2017

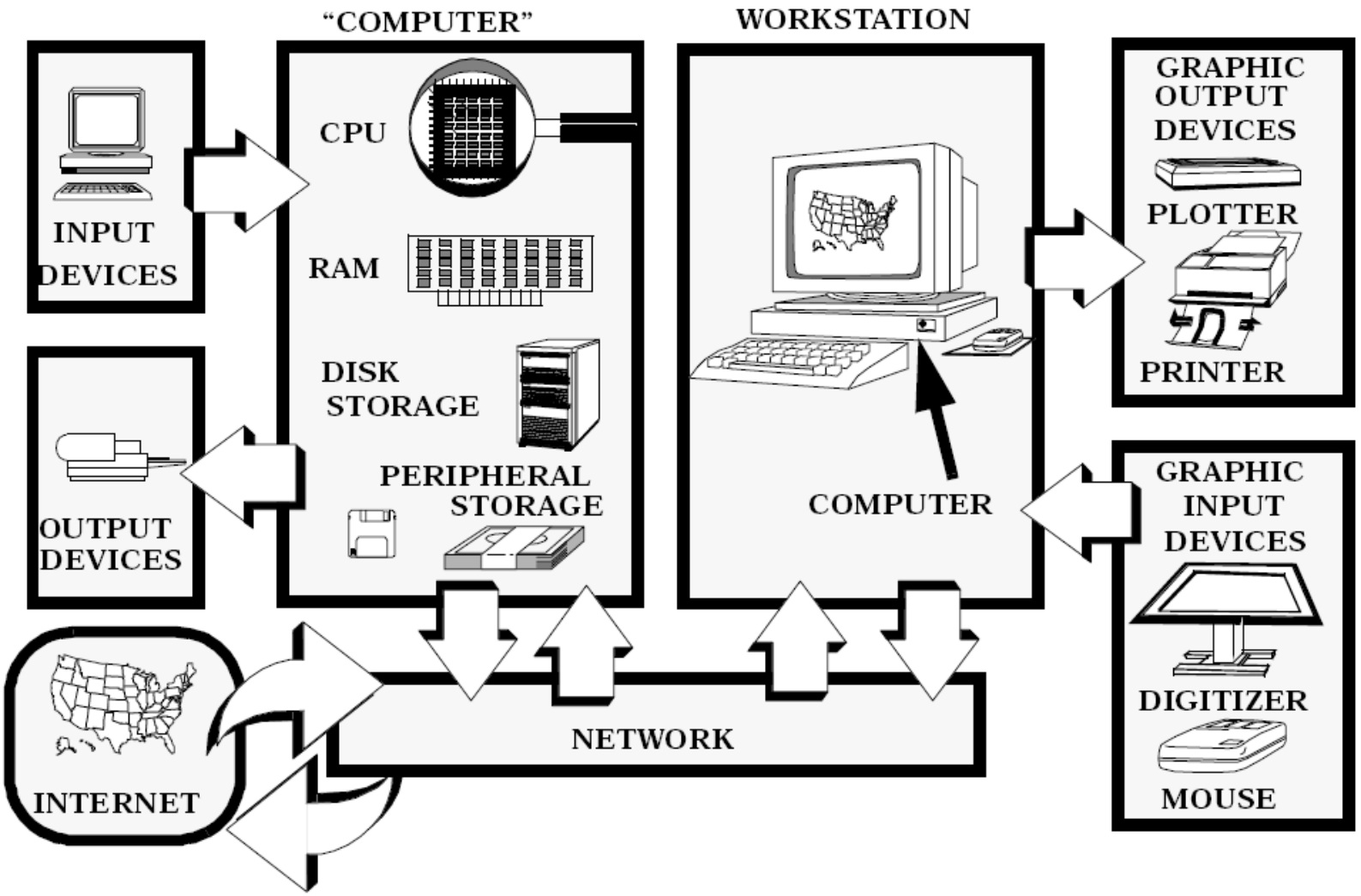
Lecture 6: Data Storage and Representation

Today's Themes

- How we represent the world makes a big difference in how we map it
- Technology impacts how the computer can store maps
- During the data capture and initial transformations errors abound: be wary!



Hardware for Computer Cartography

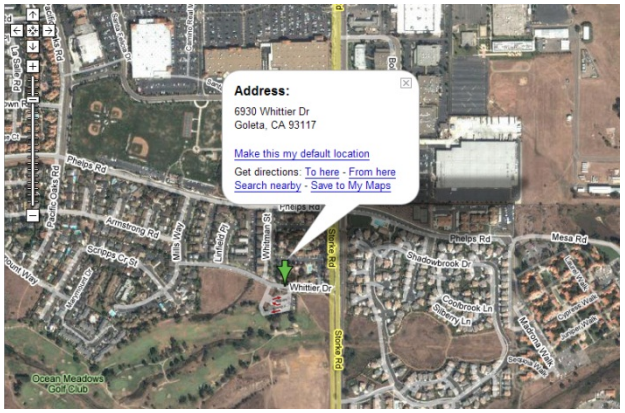


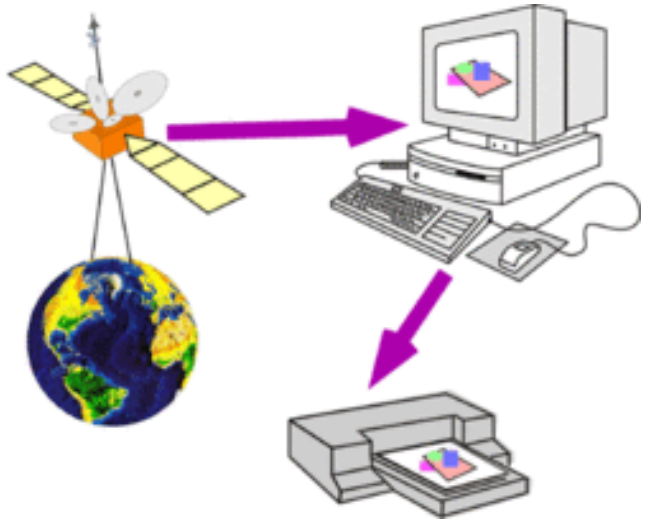
Clarke, K., 1995, Analytical and Computer Cartography

Input Devices for Computer Cartography

What is Geocoding?

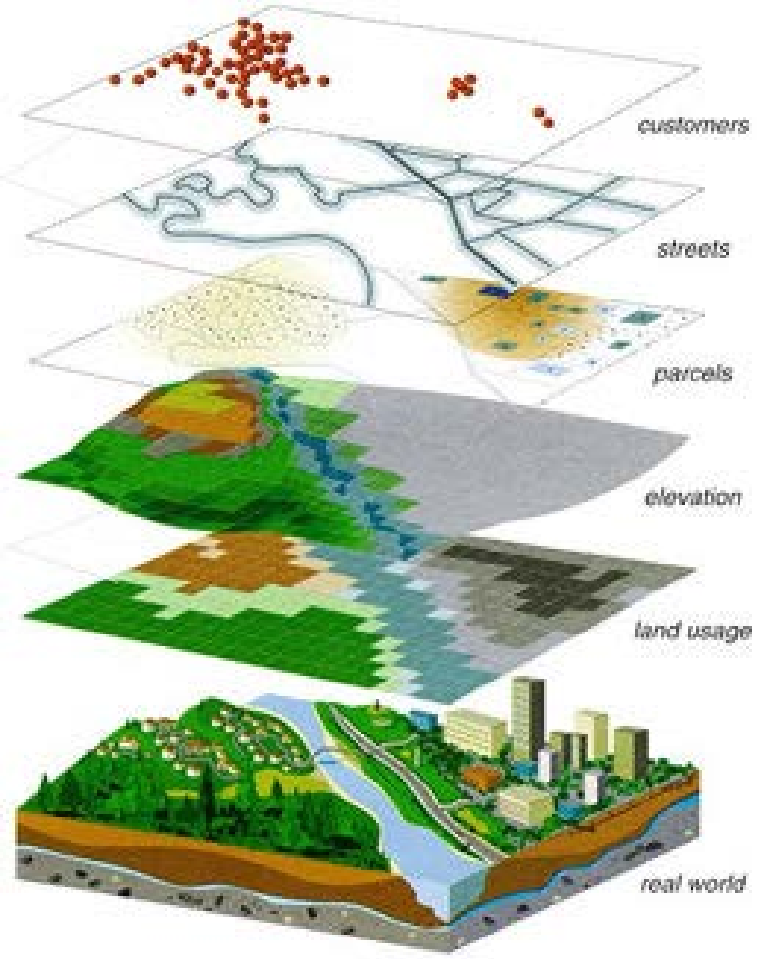
- *"Geocoding is the conversion of spatial information into computer-readable form. As such, Geocoding, both the process and the concepts involved, determines the type, scale, accuracy and precision of digital maps"* – K., Clarke, 1995
- Geocoding involves capturing the coordinates, and sometimes also capturing the attributes
- Often involves address matching





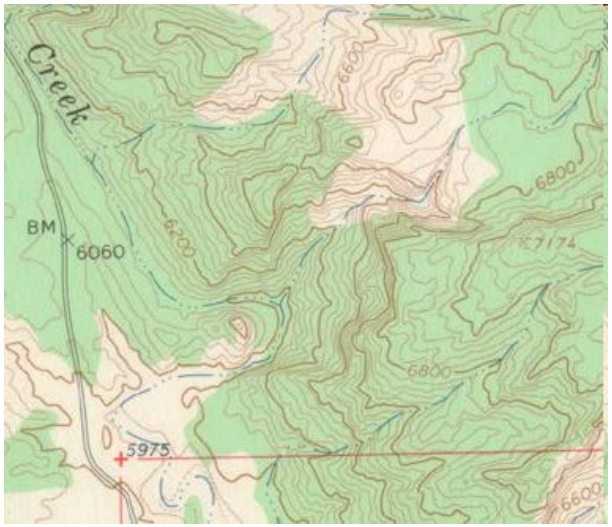
What is Geocoding? (ctd)

- Real world broken into phenomena, landscapes
- Phenomena can be broken down into cartographic entities
- Entities are geocoded to become cartographic objects
  - Geometry
  - Topology
  - Attribute

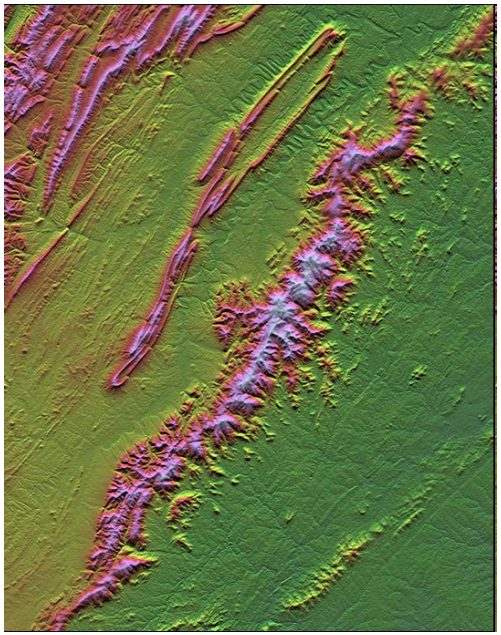


Two fundamentally different data models

- Features
  - Objects
  - Entity-by-entity
  - Point/Line/Area/(Volume)
  - Based on coordinates with given precision

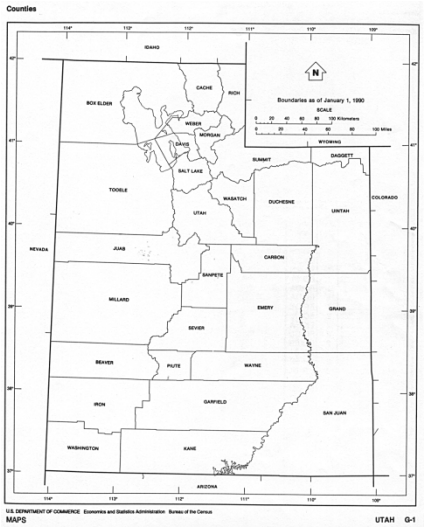


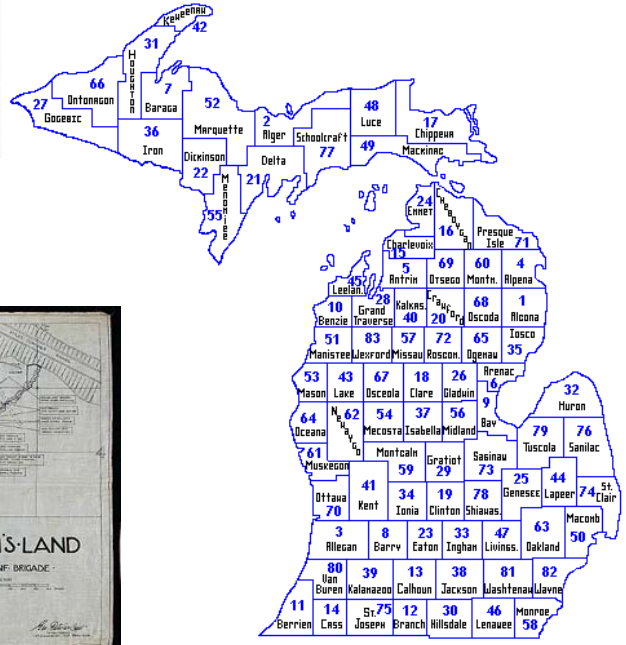
- Fields
  - Variable is continuous
  - Measurement can take place anywhere
  - Samples are used, sampling strategy matters

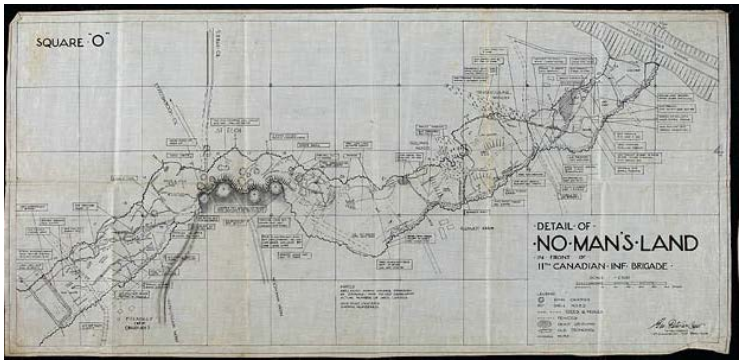


Common issues with Cartographic Data Models

- Holes
- Undefined areas
- Discontinuities
- Multi-values
- Uncertainty

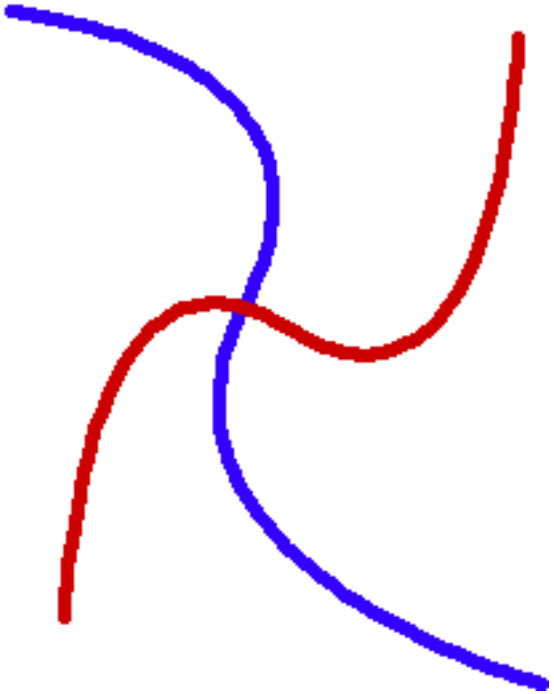




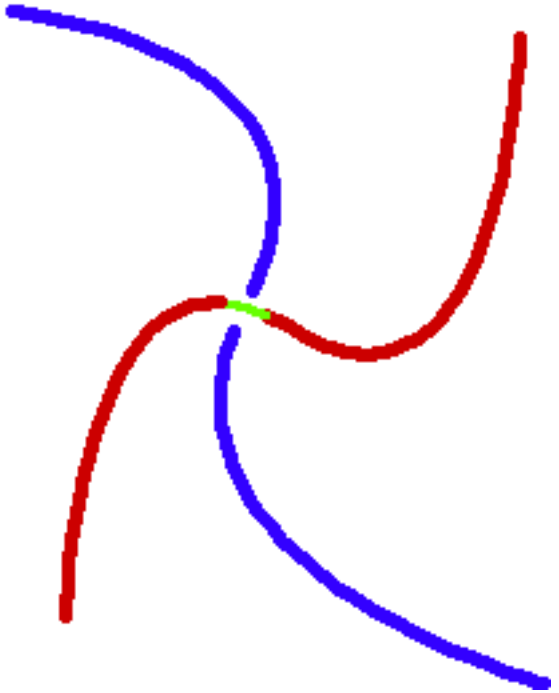


Topology

Topology - The spatial relationships between connecting or adjacent map features (e.g., points, lines, and polygons) that remain when geometry is removed



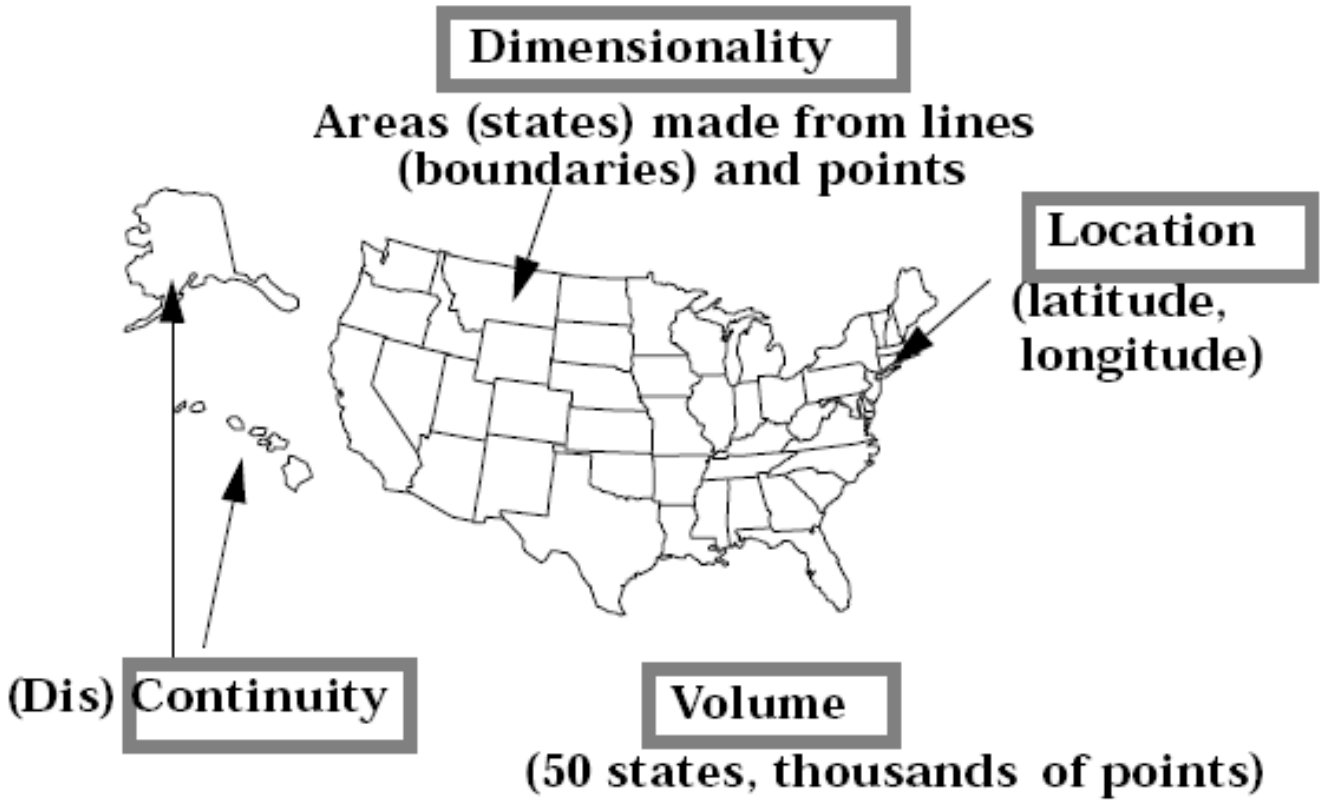
Non-topological Approach



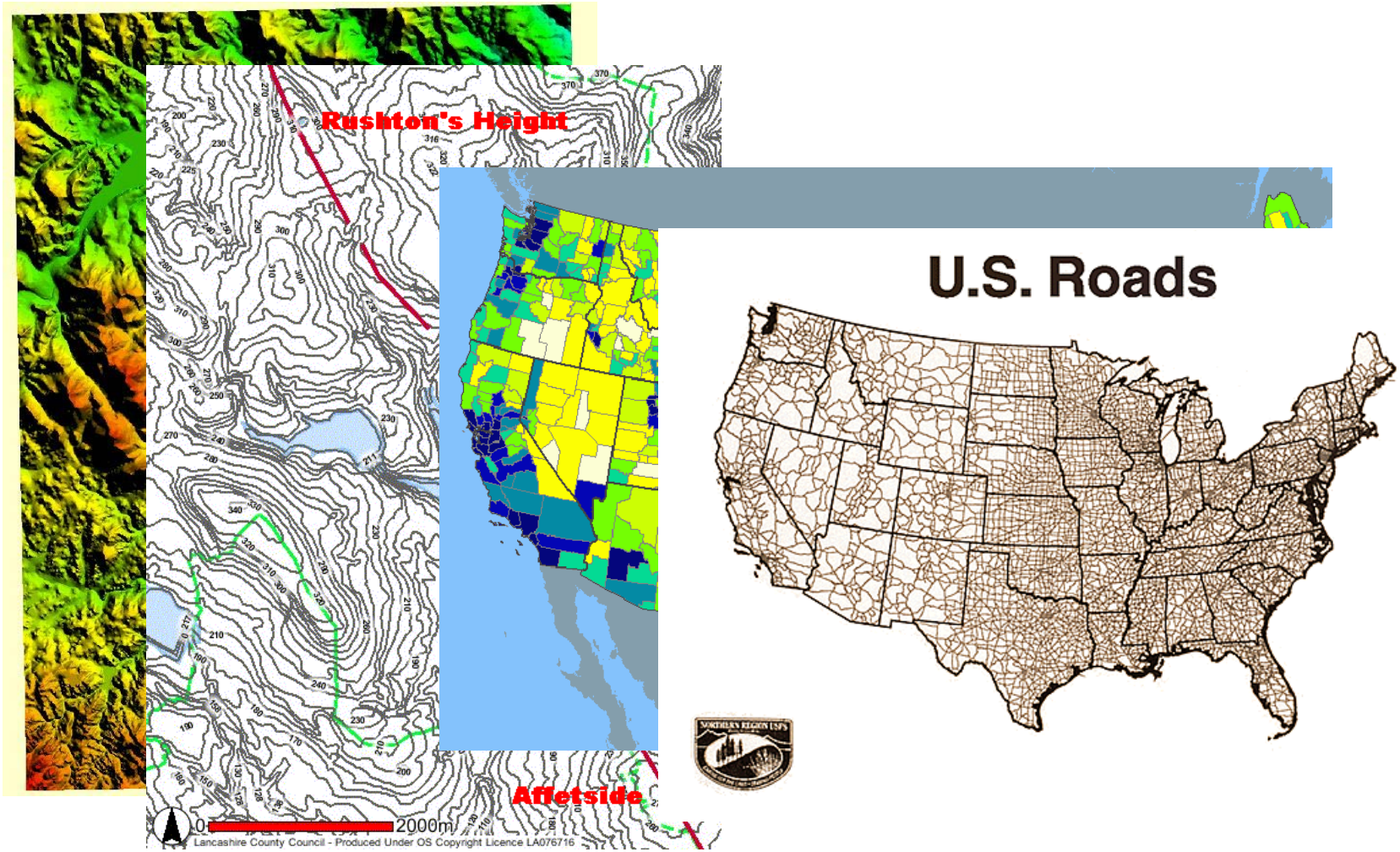
Topological Approach

Characteristics of Geographic Data

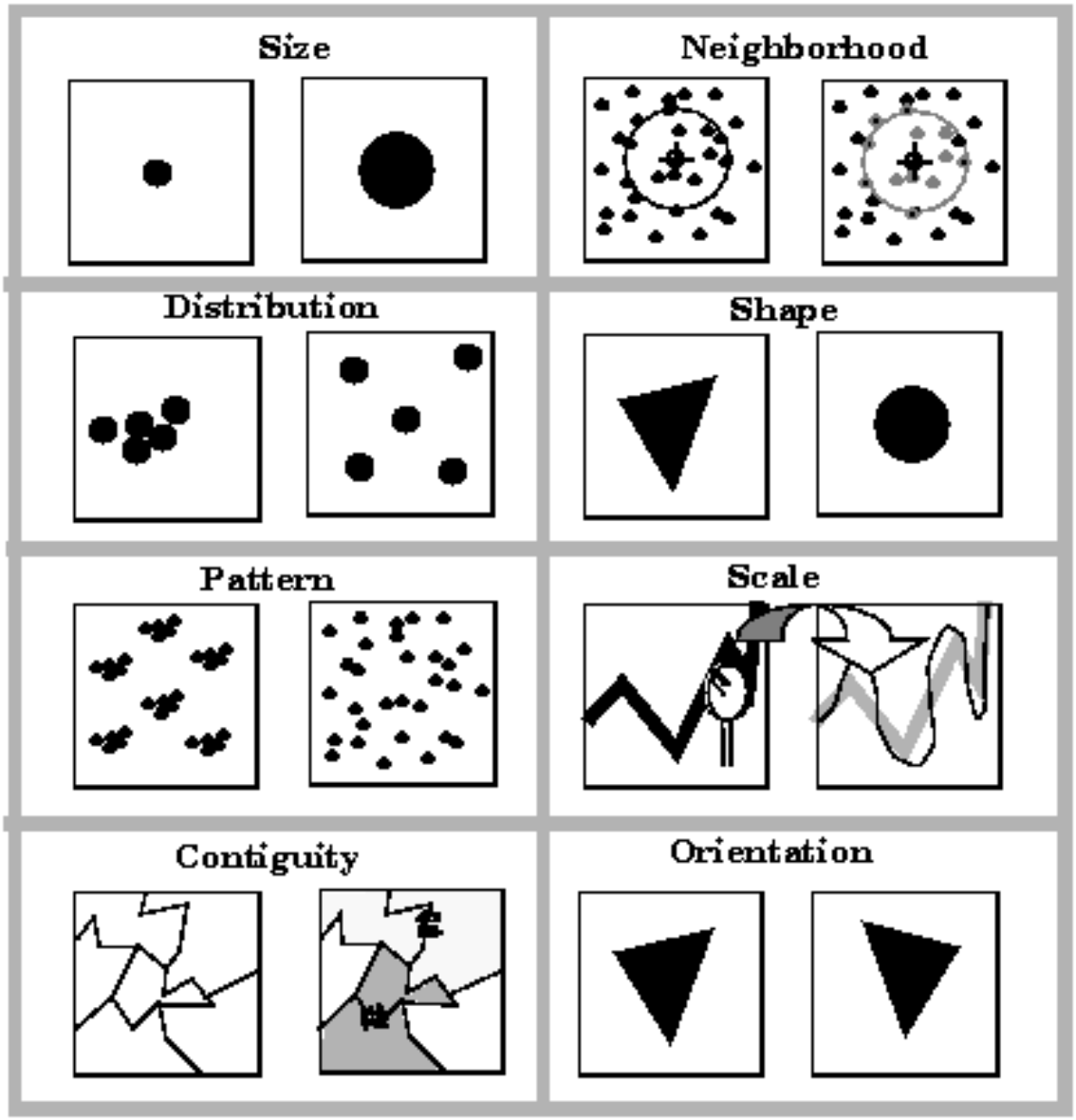
- Geocoding is the process of digital-encoding the fundamental characteristics of geographic data
  - Location
  - Volume
  - Dimension
  - Continuity



Continuity vs. Discontinuity

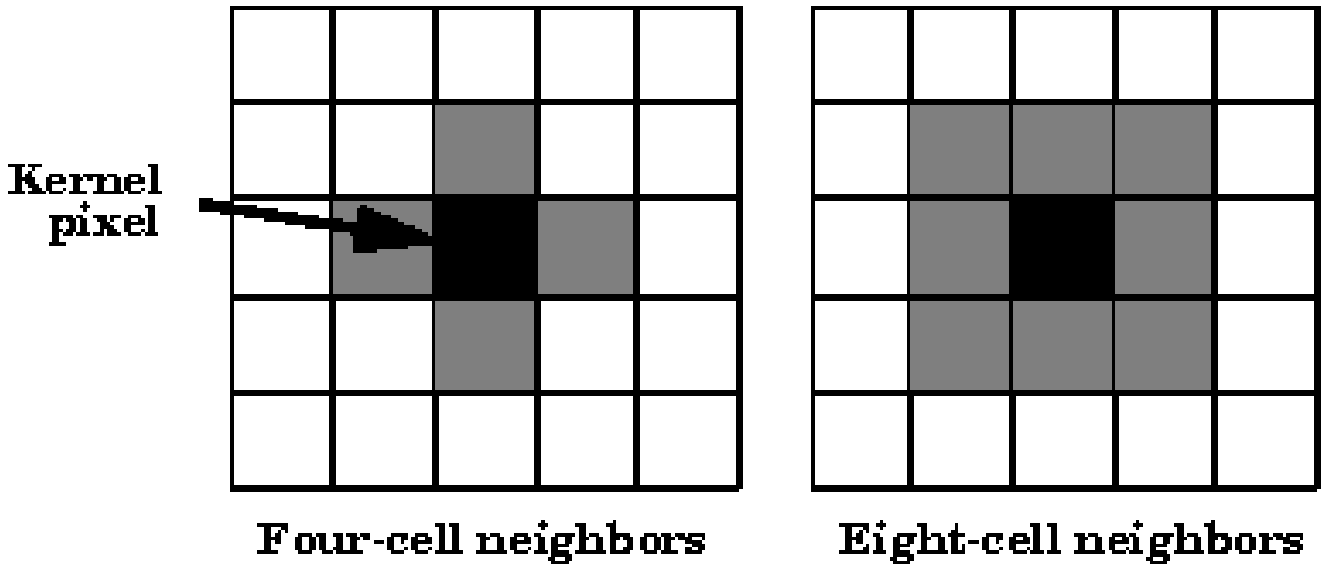


Fundamental Properties of Geographic Objects



Fundamental Properties of Geographic Objects (ctd)

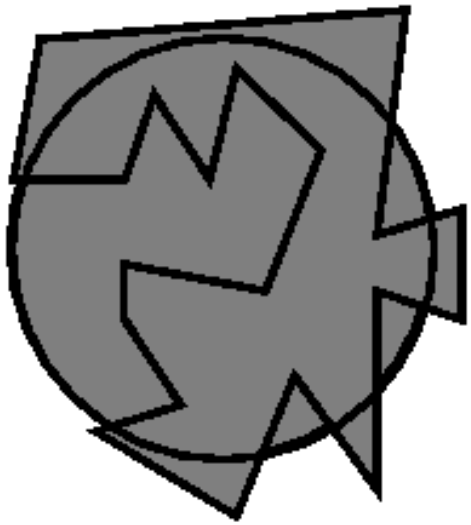
- Neighborhood



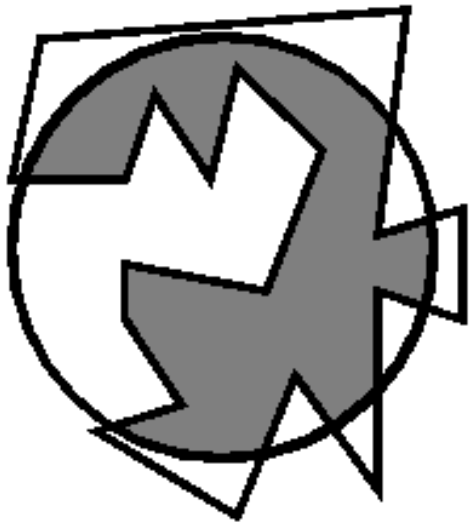
Fundamental Properties of Geographic Objects (cnt.)

- Shape
  - Lee and Sallee index

$$s = 1 - \frac{A \cap B}{A \cup B}$$



Union $A \cup B$



Intersection $A \cap B$

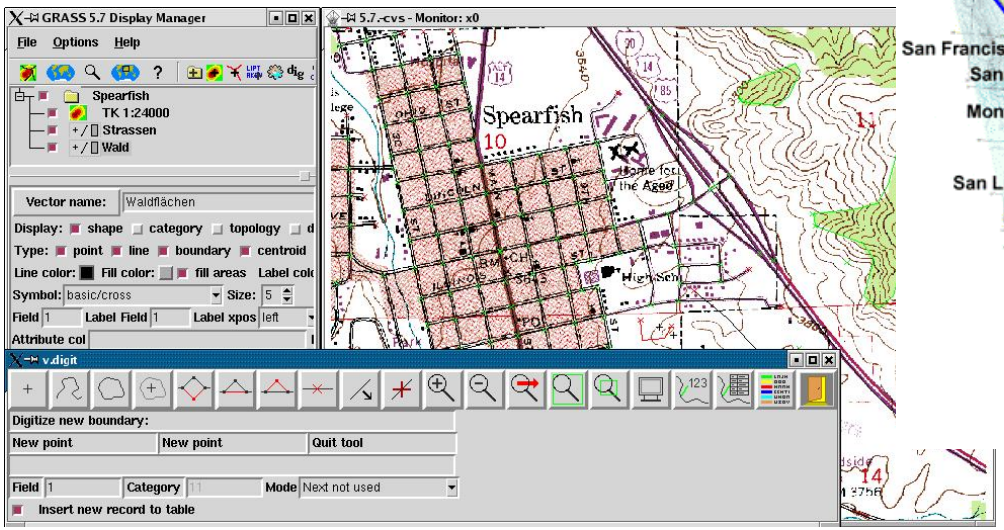
North Korea's Yongbyon Nuclear Scientific Research Center is pictured in January 16, 2017 handout satellite image

Measure the Properties of Geographic Objects

	POINT	LINE	AREA	VOLUME
SHAPE	Feature type	Curvature	Shape measure	Dimension Resemblance to figure (e.g. cone)
CONTIGUITY	Link	Intersection	Shared boundary	Shared face
ORIENTATION	Of cluster or pattern	Bearing Trend	Of axis Of pattern	Dip, drift, trend, aspect
SIZE	Number	Length	Area	Volume Surface area
SCALE	Range at which object is a point	Range at which object is a Line	Range at which object is an area	Range at which object is a volume
NEIGHBORHOOD	Set of nearby points	Connected lines. Lines within a range	Contiguous areas Area within a range Connected areas	Adjacent voxels Overlapping volumes Shared faces
PATTERN	Pattern matching Fourier analysis	Curve measures Fractal dimension	Shape distribution Description	Fourier power spectrum Trend surface
DISTRIBUTION	Standard distance Nearest neighbor number Autocorrelation	Line density Length, Intersection frequency	Coverage Autocorrelation	Variogram

Goals of Geocoding Methods

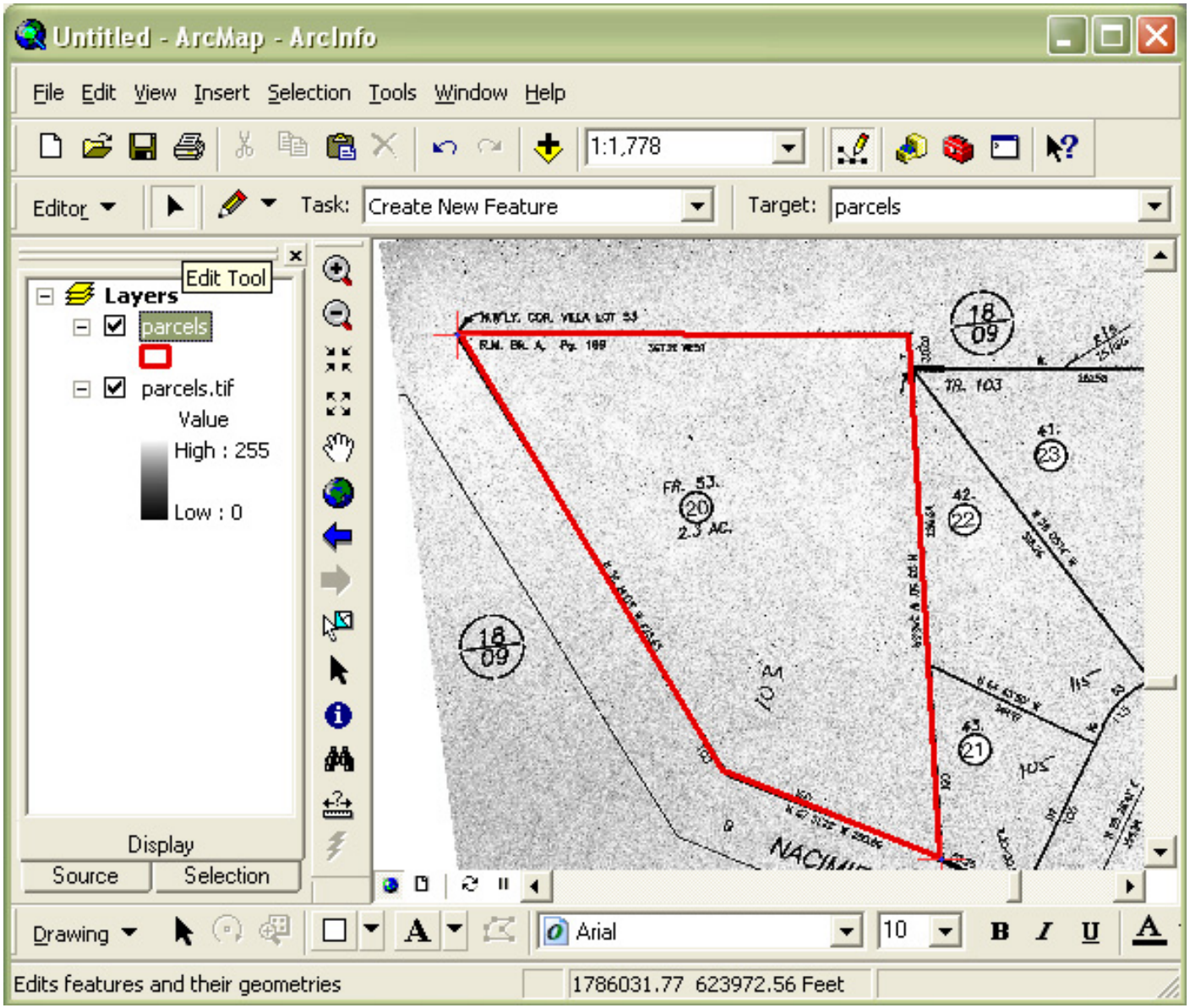
- Minimize Labor Input
- Detect and Eliminate Errors
- Optimize Storage Efficiency
- Maximize Flexibility



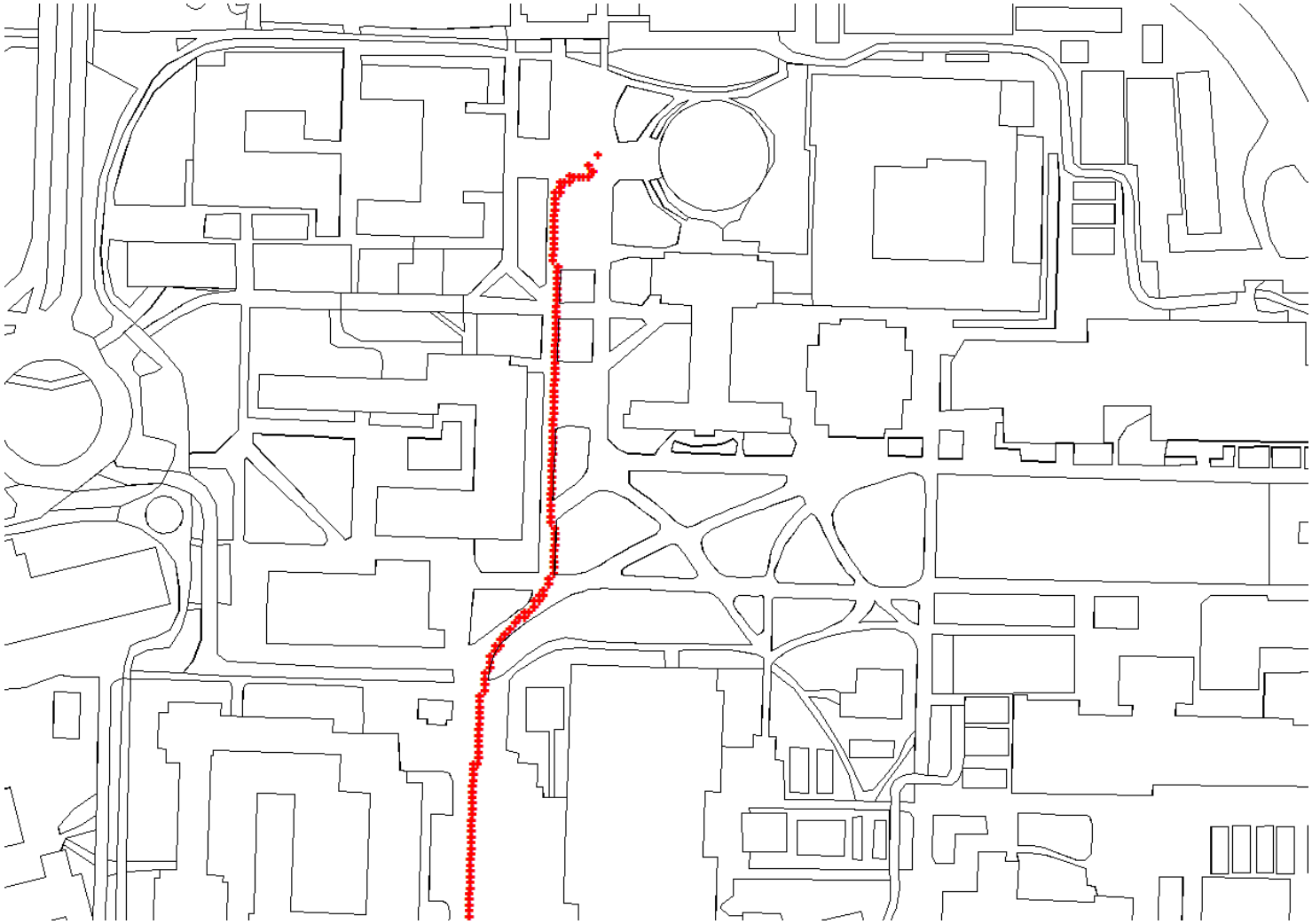
Geocoding methods for maps

- Digitizing
- Scanning
- Field data collection

On-screen digitizing

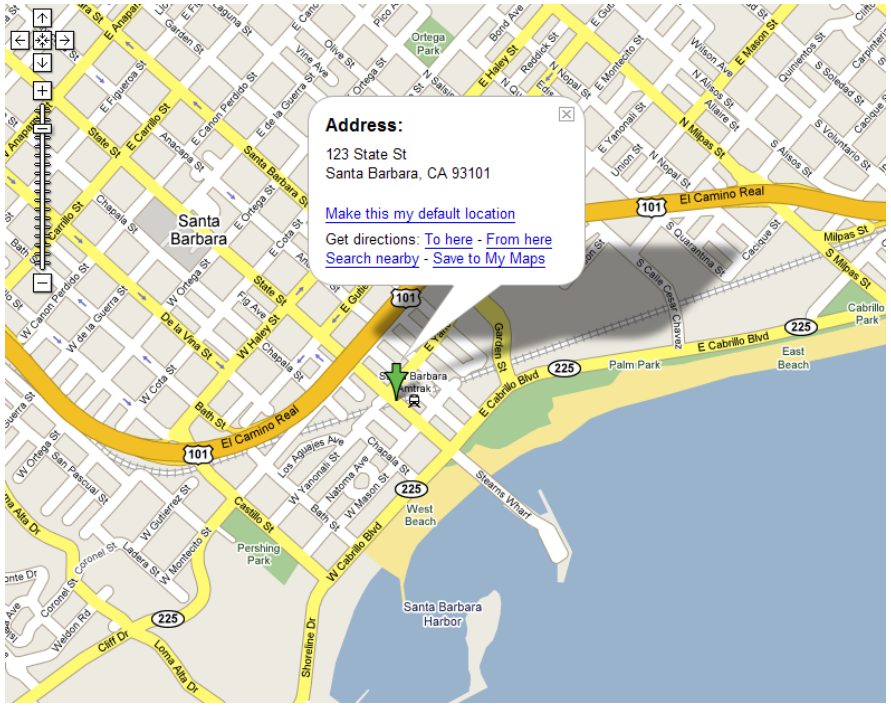


GPS navigation/tracks



Address Matching

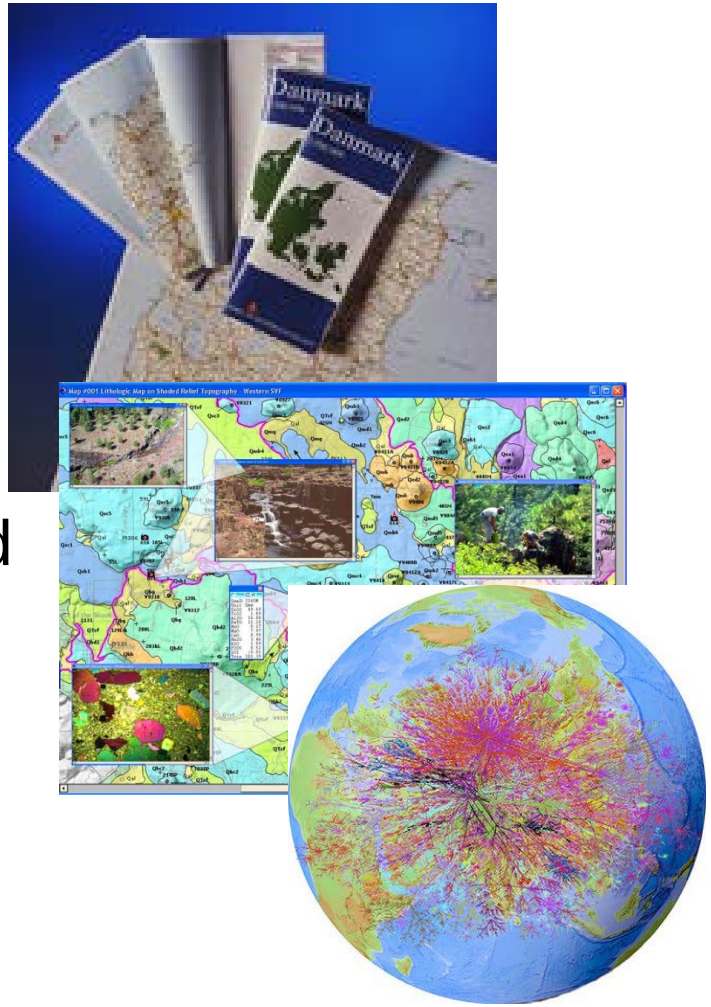
- Most GISs contain capability
- Start with 123 State St, Santa Barbara, CA 93101
- End with Coordinates



- May need to interpolate along blocks
- Street number range, left and right side e.g. 101-199

Storage Media

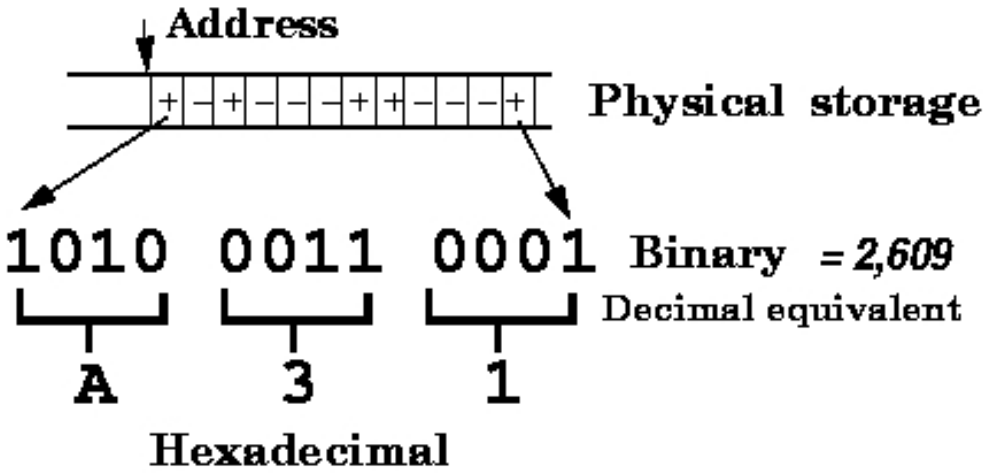
- Traditionally, the paper map has performed a storage function for spatial information
- Computer cartography requires information to be digital and stored explicitly
- Storage is increasingly distributed over networks
- Many mapping programs require local storage of data
- Cost and size restraints now less important

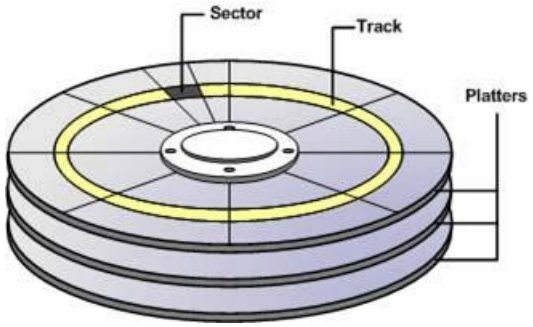


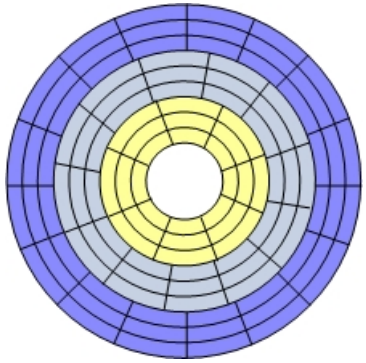
Evolution of Storage Media

Physical Storage

- Bit - the most basic information unit in a binary system (1 / 0)
- 1 Byte = 8 bits
- Binary (2-based), Decimal(10-based), and Hexadecimal (16-based) System
- Binary Operator – AND, OR, NOT
- Data on a disk -Sectors, Tracks, Platters
- File system – File, Directory







Maps as Numbers

- Map data is stored in the computer's memory in a physical data structure (i.e. files and directories).
- Files can be written in binary or as ASCII (American Standard Code for Information Interchange) text.
- Binary is faster to read and smaller, ASCII can be read by humans and edited but uses more space.

ASCII Table

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040		Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	(	(	72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051	)	)	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[	[	123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174			|
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135	]	]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177	&#127;	DEL

Source: www.LookupTables.com

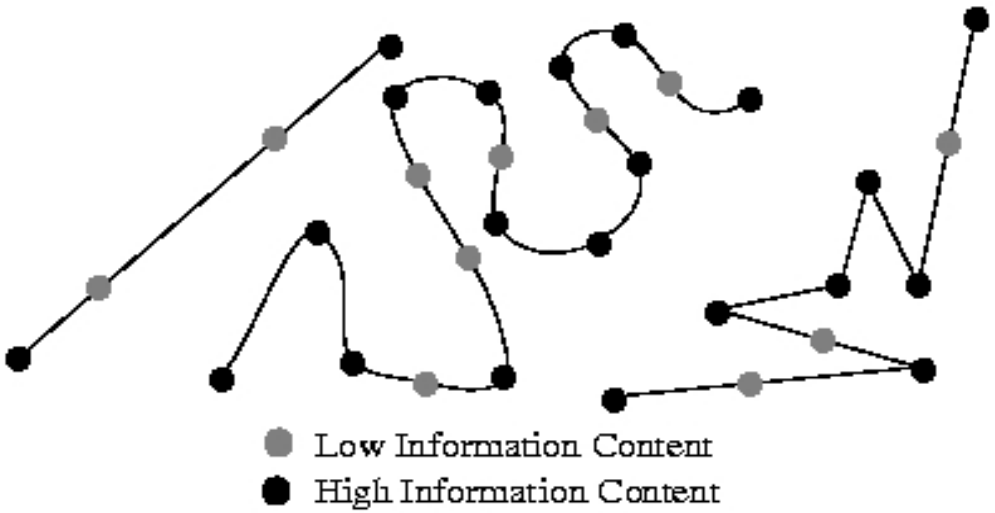
ASCII Table (extend)

128	Ç	144	É	161	í	177	▒	193	┴	209	╤	225	ß	241	±
129	ü	145	æ	162	ó	178	▓	194	┬	210	╥	226	Γ	242	≥
130	é	146	Æ	163	ú	179	│	195	├	211	╙	227	π	243	≤
131	â	147	ô	164	ñ	180	┤	196	─	212	╘	228	Σ	244	⌠
132	ä	148	ö	165	Ñ	181	╡	197	┼	213	╒	229	σ	245	⌡
133	à	149	ò	166	ª	182	╢	198	╞	214	╓	230	µ	246	÷
134	å	150	û	167	º	183	╖	199	╟	215	╫	231	τ	247	≈
135	ç	151	ù	168	¿	184	╕	200	╚	216	╪	232	Φ	248	°
136	ê	152	_	169	_	185	╣	201	╔	217	┘	233	Θ	249	·
137	ë	153	Ö	170	¬	186	║	202	╩	218	┌	234	Ω	250	·
138	è	154	Ü	171	½	187	╗	203	╦	219	▉	235	δ	251	√
139	ï	156	£	172	¼	188	╝	204	╠	220	▄	236	∞	252	_
140	î	157	¥	173	¡	189	╜	205	═	221	▌	237	ϕ	253	²
141	ì	158	_	174	«	190	╛	206	╬	222	▐	238	ε	254	■
142	Ä	159	ƒ	175	»	191	┐	207	╧	223	▀	239	∩	255	
143	Å	160	á	176	░	192	└	208	╨	224	α	240	≡		

Source: www.LookupTables.com

Storage Efficiency and Data Compression

- Cartographic data sets are typically large
- Need to reconfigure data formats, structures etc.
- Seek to retain information content, lose volume.
- Is redundancy necessary?

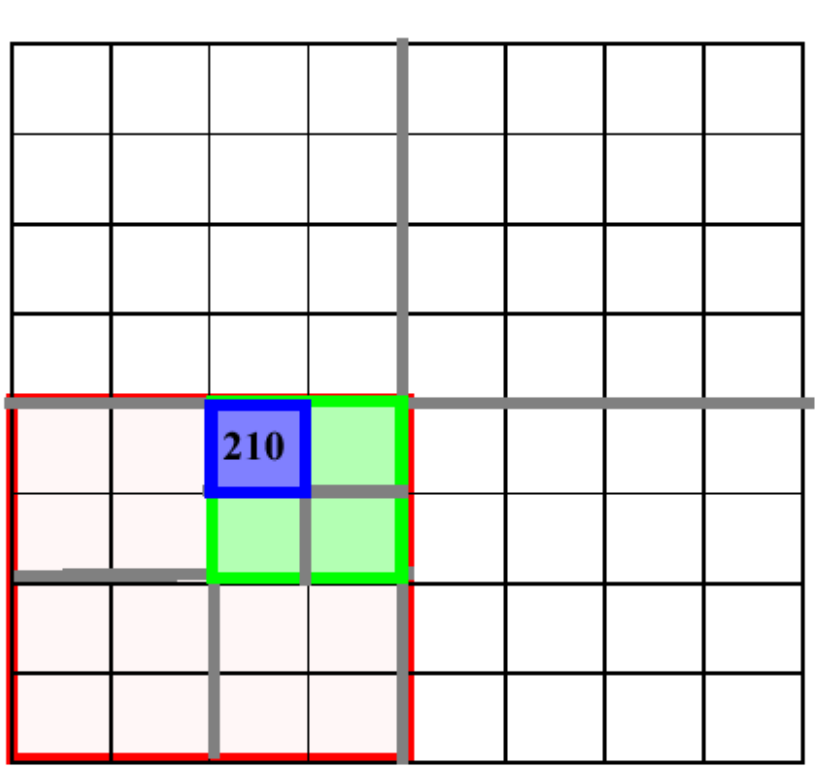


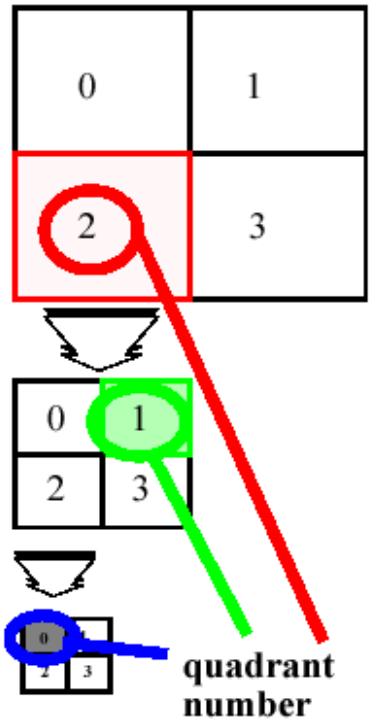
Storing Coordinates (Vector)

- Physical Compression
  - 4,513,410 m N;587,310 m E; Zone 18,N (32 characters, 15 digits)
  - 4513410 587310 (13 digits, one space) Need metadata
  - 98 96 7F 0F 42 3F (six bytes)
- Logical Compression
  - Drop last two digits (10 ASCII or 2 bytes per coordinate)

Raster data Compression

- Run-length encoding
- Quad-trees
- …

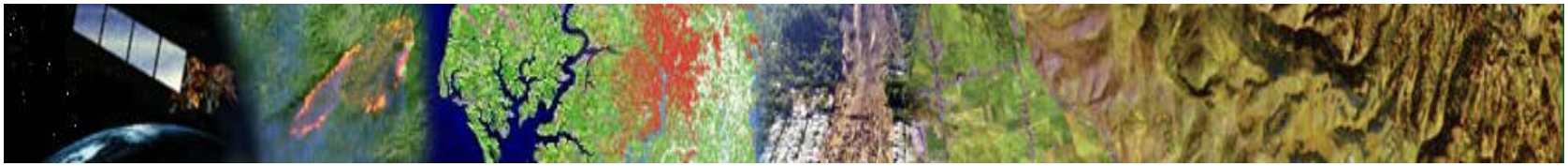




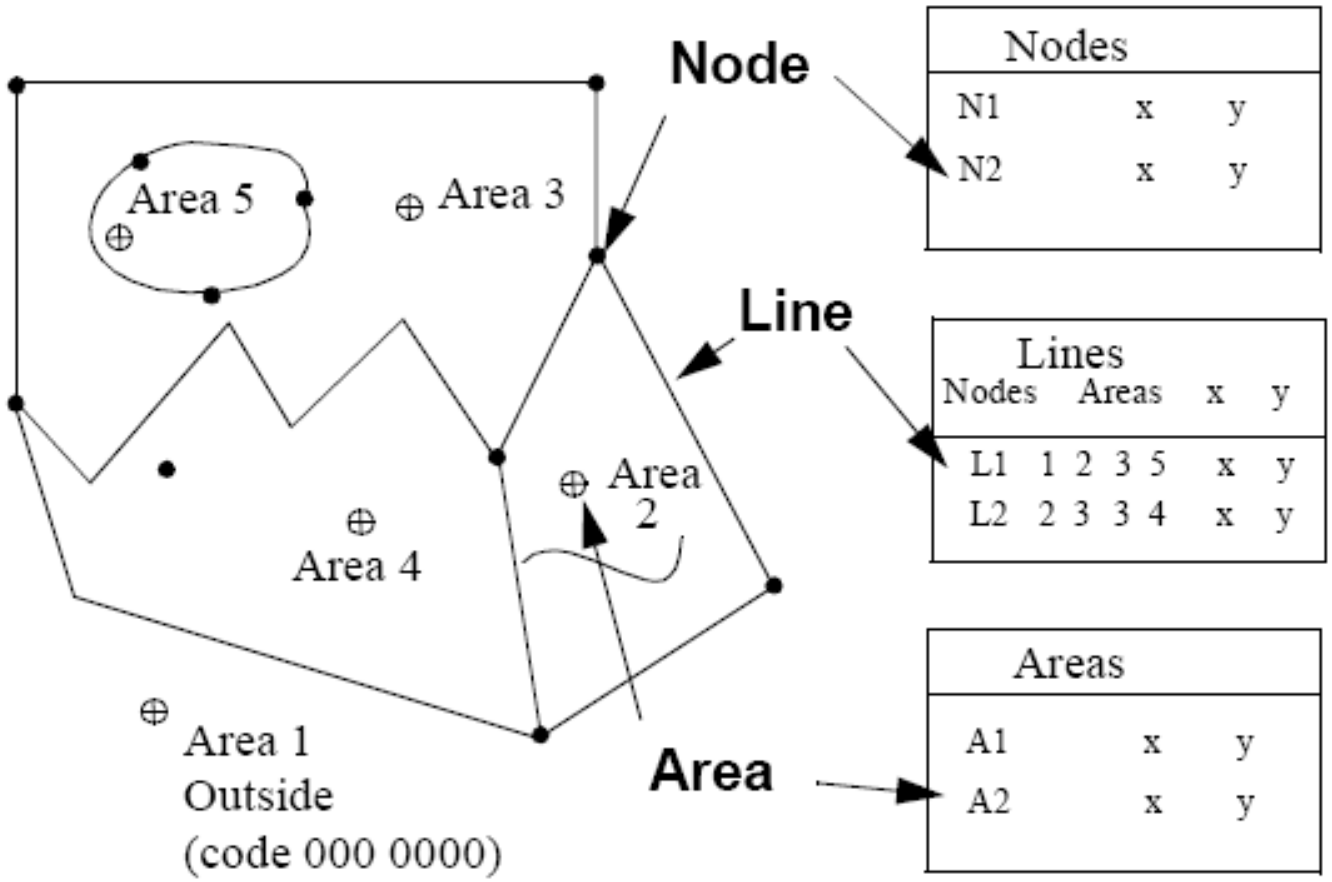
Data Storage Formats for Cartography - U.S. Geological Survey

- DLG – Digital Line Graphs (1:24,000; 1:100,000; 1:2,000,000)
- DEM – Digital Elevation Model (1:24,000; 1:250,000)
- GIRAS – Land-use and Land-cover Digital Data (1:100,000; 1:250,000)
- GNIS – Digital Cartographic Text

USGS
science for a changing world

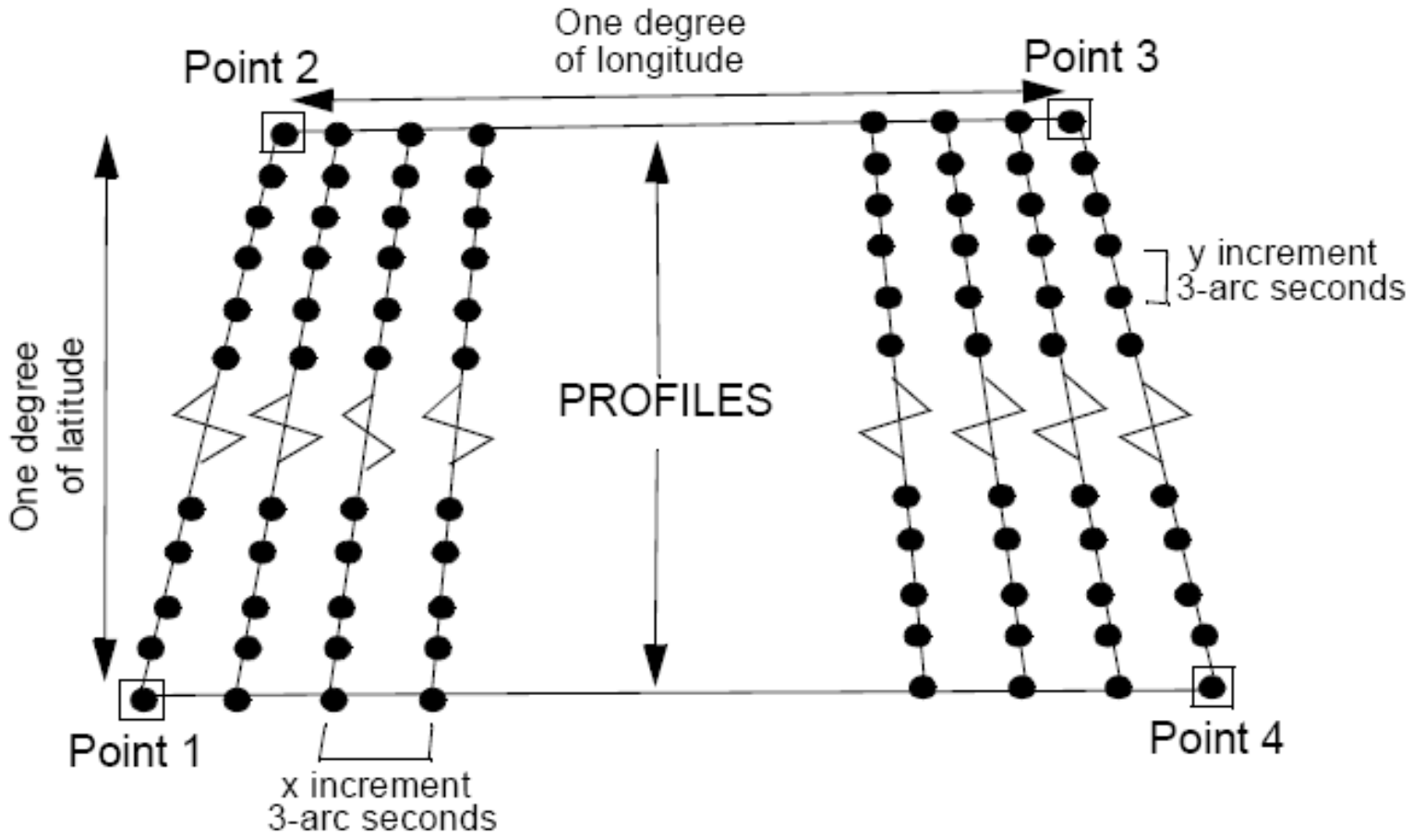


Data Storage Formats for Cartography - U.S. Geological Survey



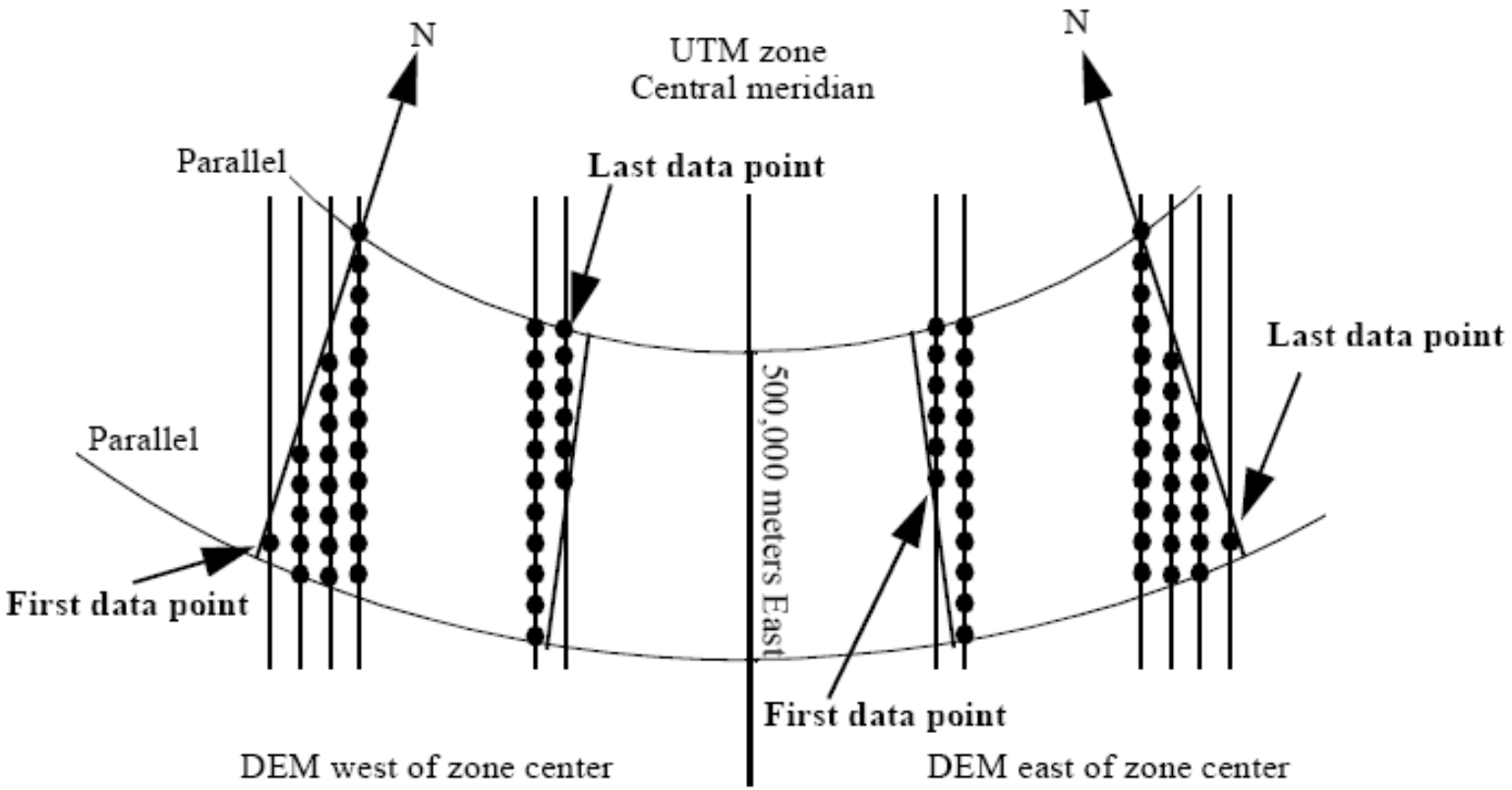
USGS DLG format

Data Storage Formats for Cartography - U.S. Geological Survey



USGS 1:250,000 3-arc second DEM format (1-degree block)

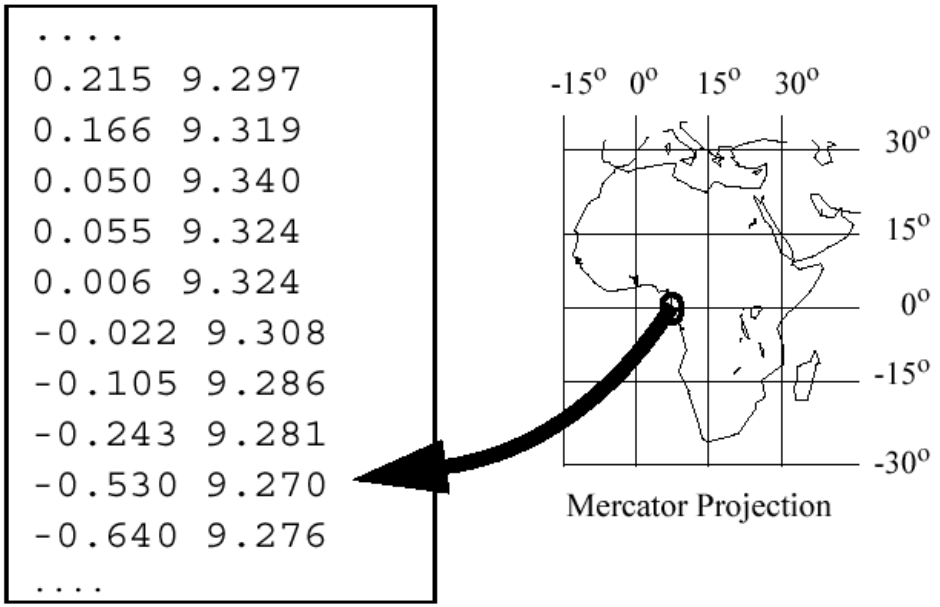
Data Storage Formats for Cartography - U.S. Geological Survey



USGS 1:24,000 30 meter DEM format (7.5-minute quadrangle)

Data Storage Formats for Cartography - CIA World Data Bank

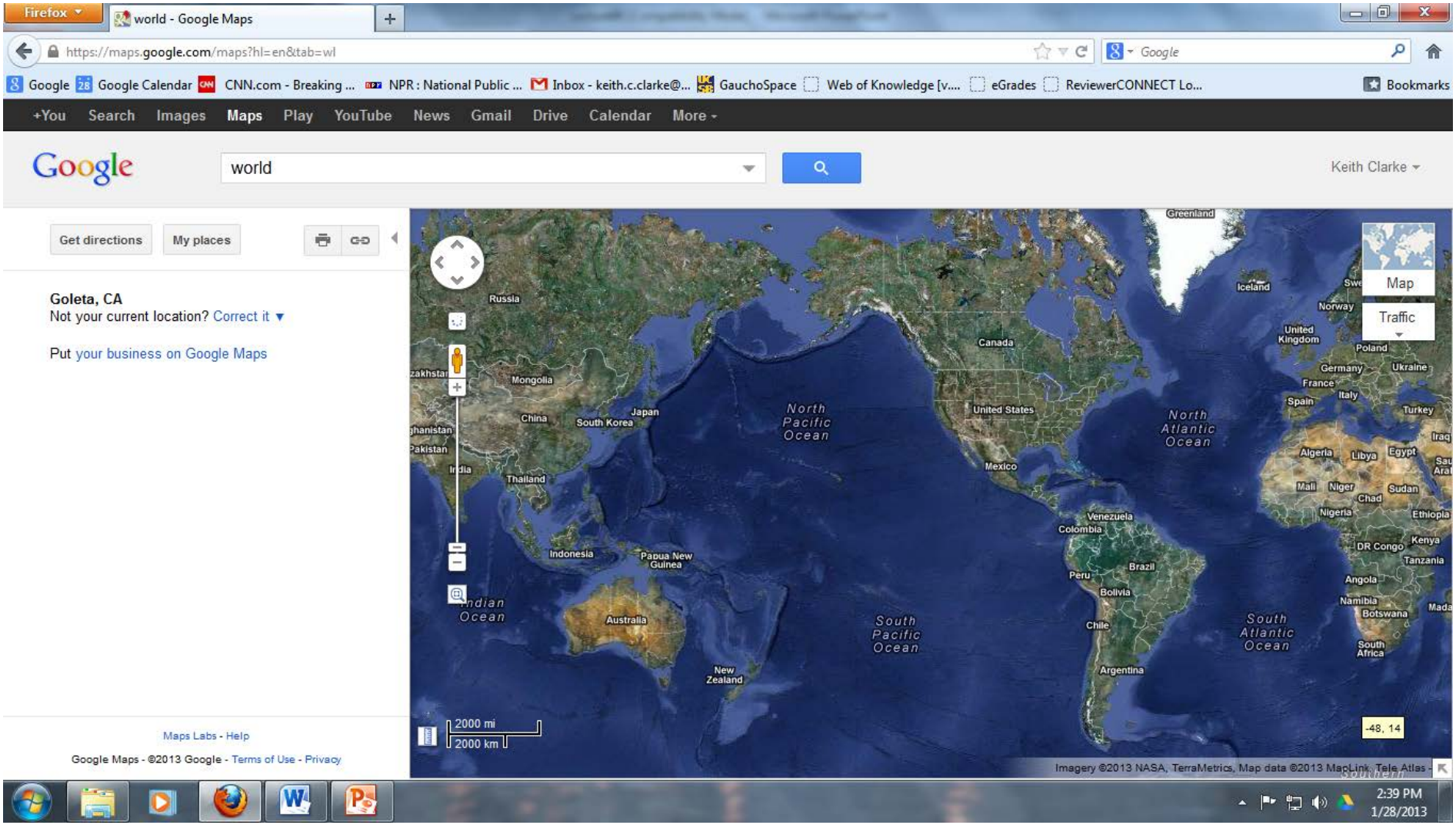
- WDB I (1:12M base, 100K points)
- WDBII (1:3M base, 6M Points)
- DCW 1:1M base- 4 CDs, 14 layers DMAs VPF



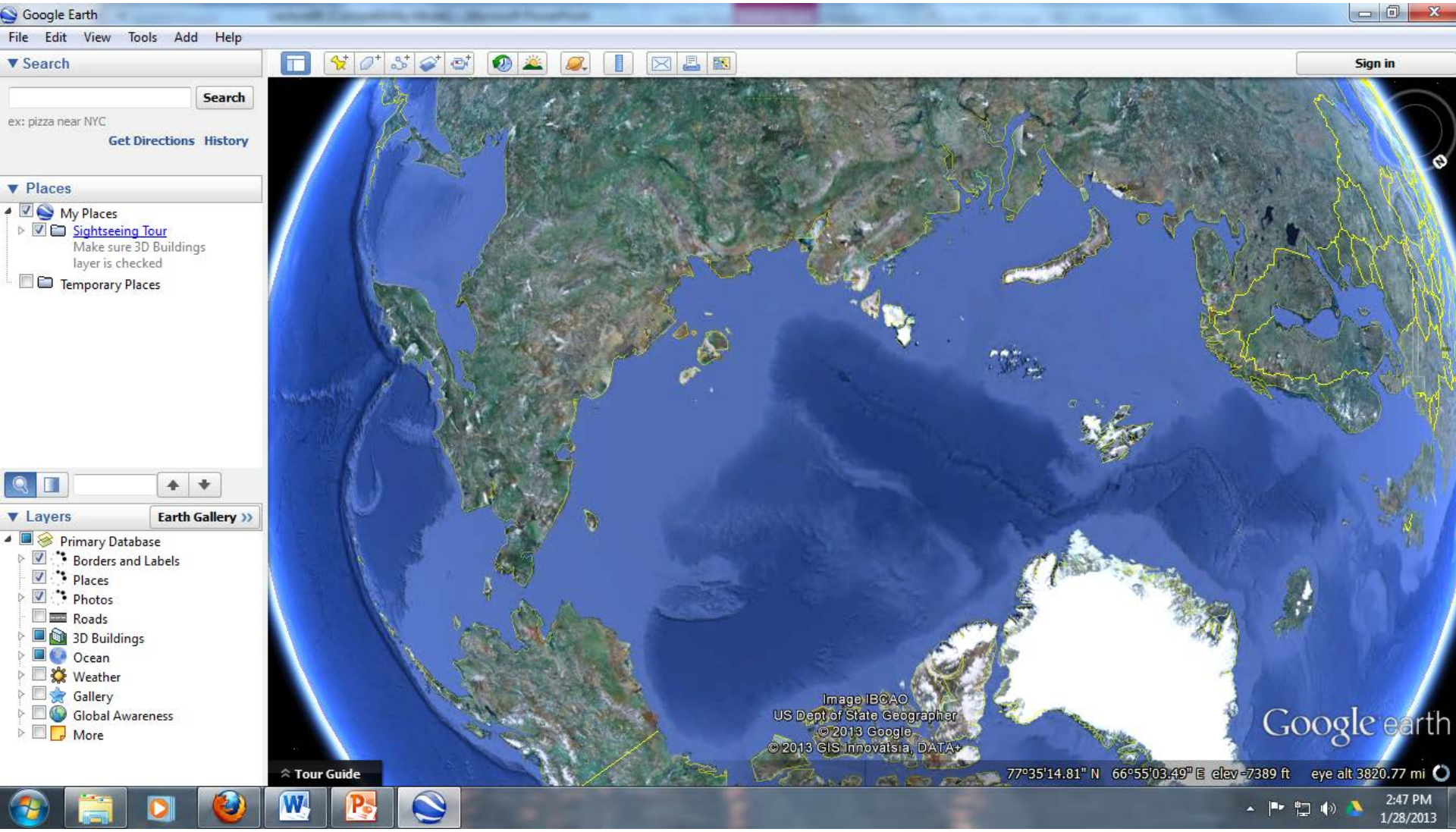
Storage Issues

- Compression: Lossy vs. Lossless
- Ease of Access (Time, # operations)
- Reliability
- Permanence
- Backup
- Redundancy
- Detail and Scale
- Tiling, Mosaicing and Joining “Seamless” Database
- Maintenance, management
- Metadata (embed vs. attach) e.g. GeoTIFF

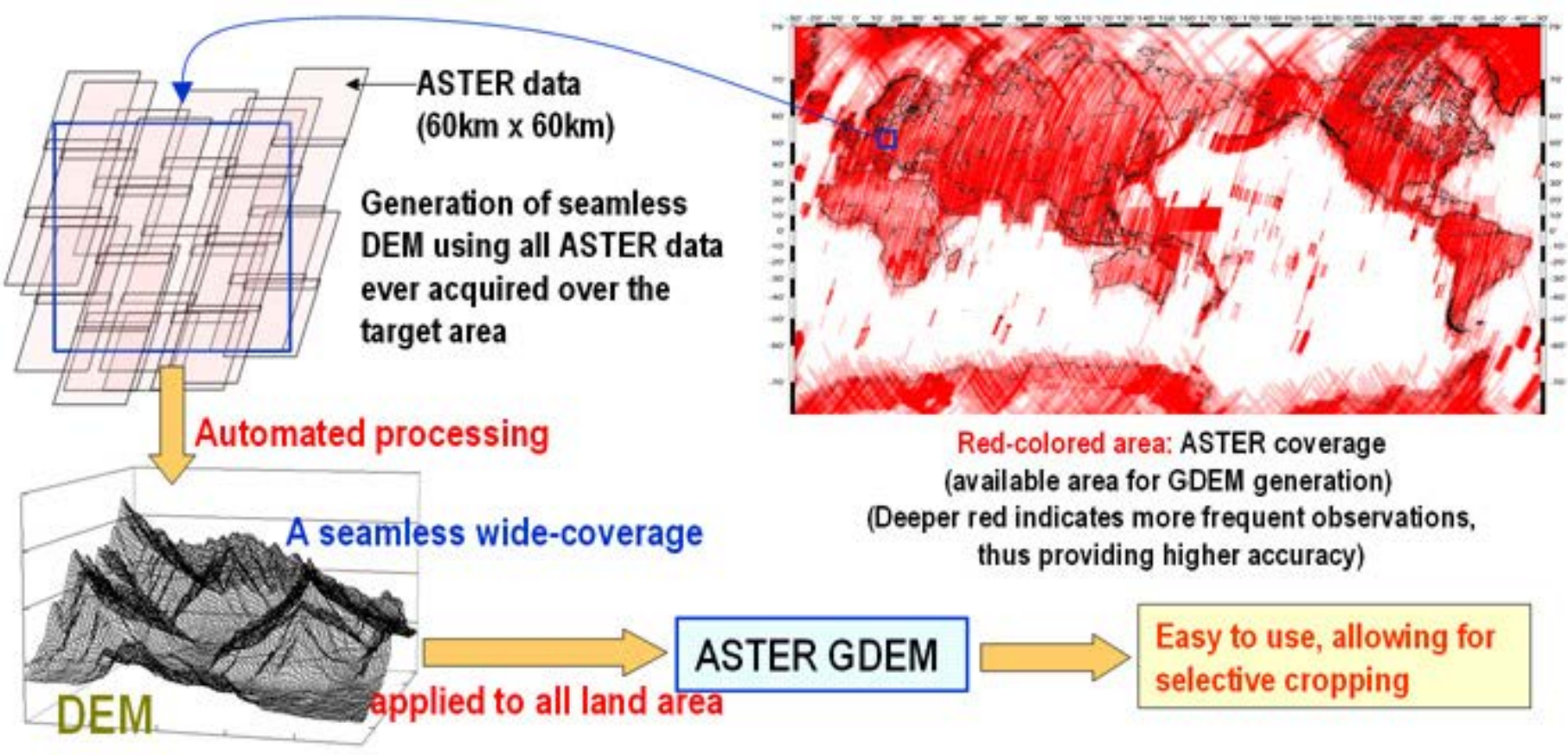
Web-Mercator



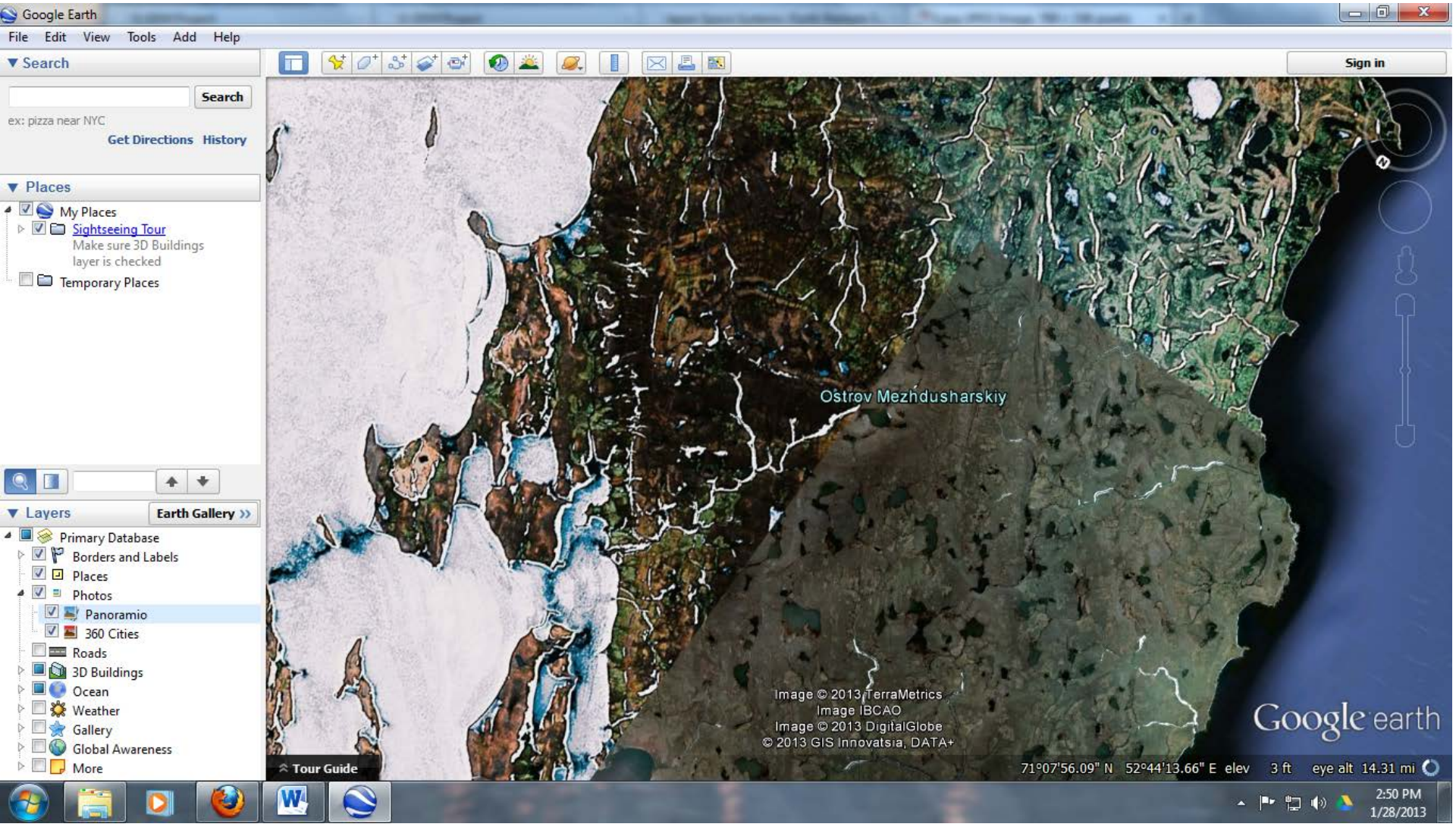
Tiles and Seams



Tiling: ASTER GDEM



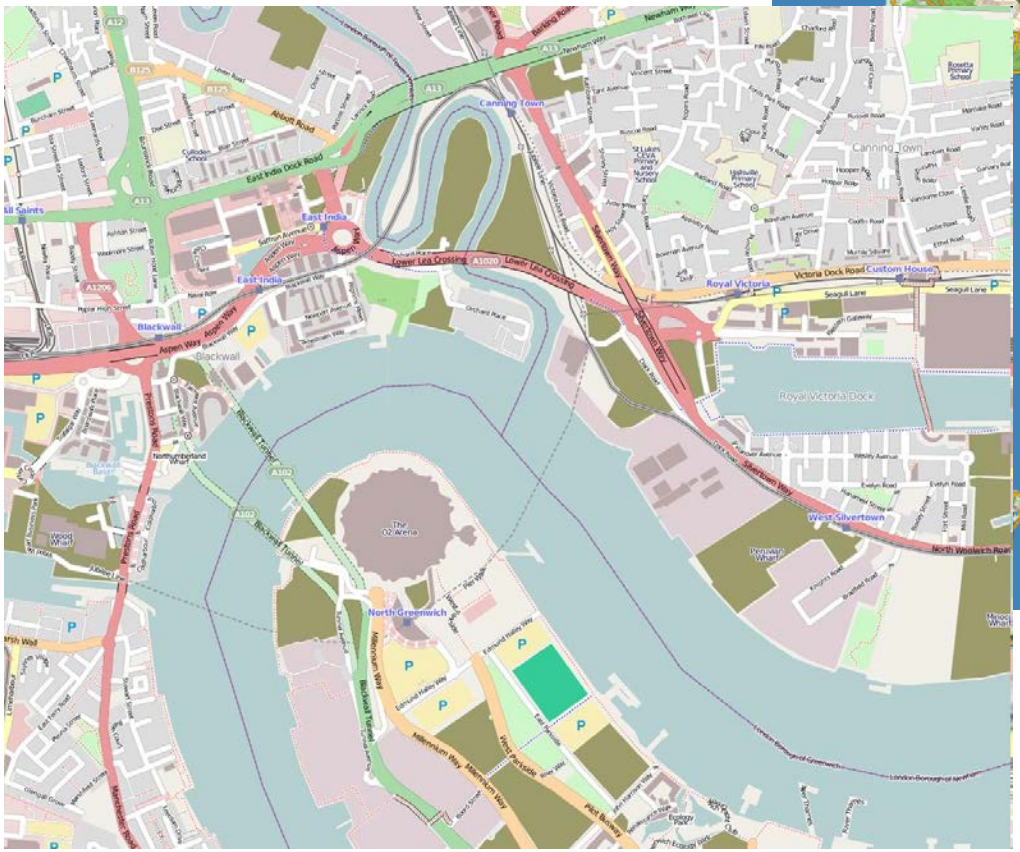
Discontinuity in Time and Space

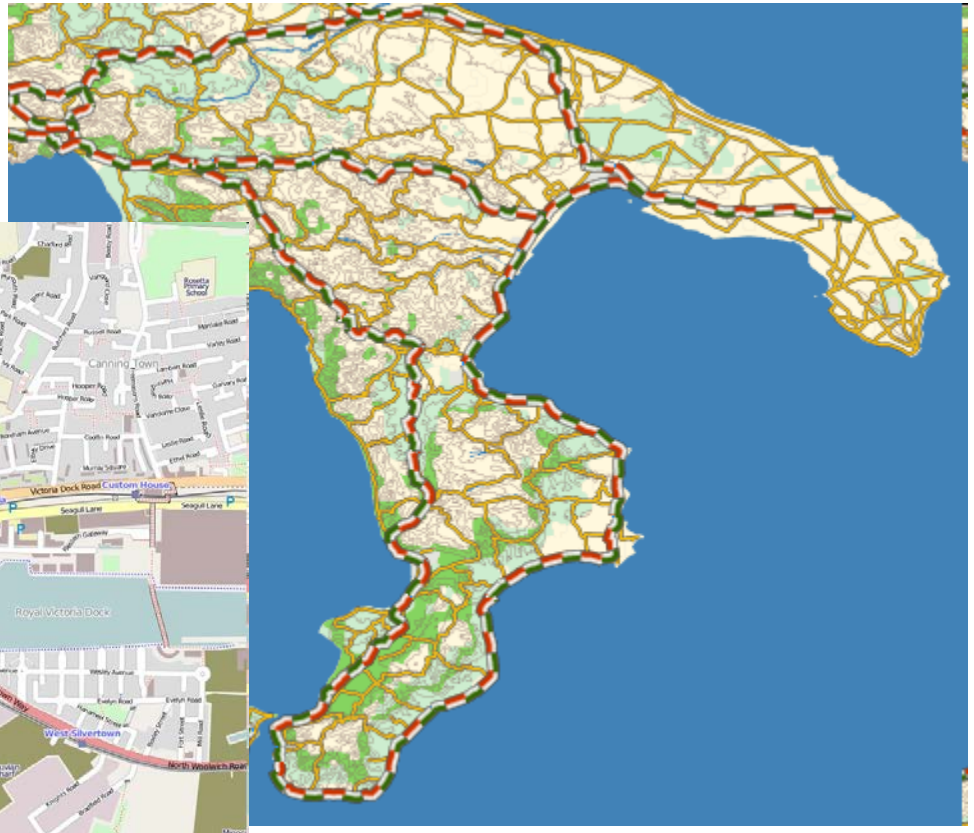


Data unlimited

- Proprietary vs. Open Source
- Accessible (i.e. Discoverable) vs. Isolated
- Protected, e.g. Private, Sensitive, Classified, Denied, Watermarked, Steganography
- Web-accessible, web-enabled, clearing house
- NSDI, GSDI, Digital Earth

Open Source

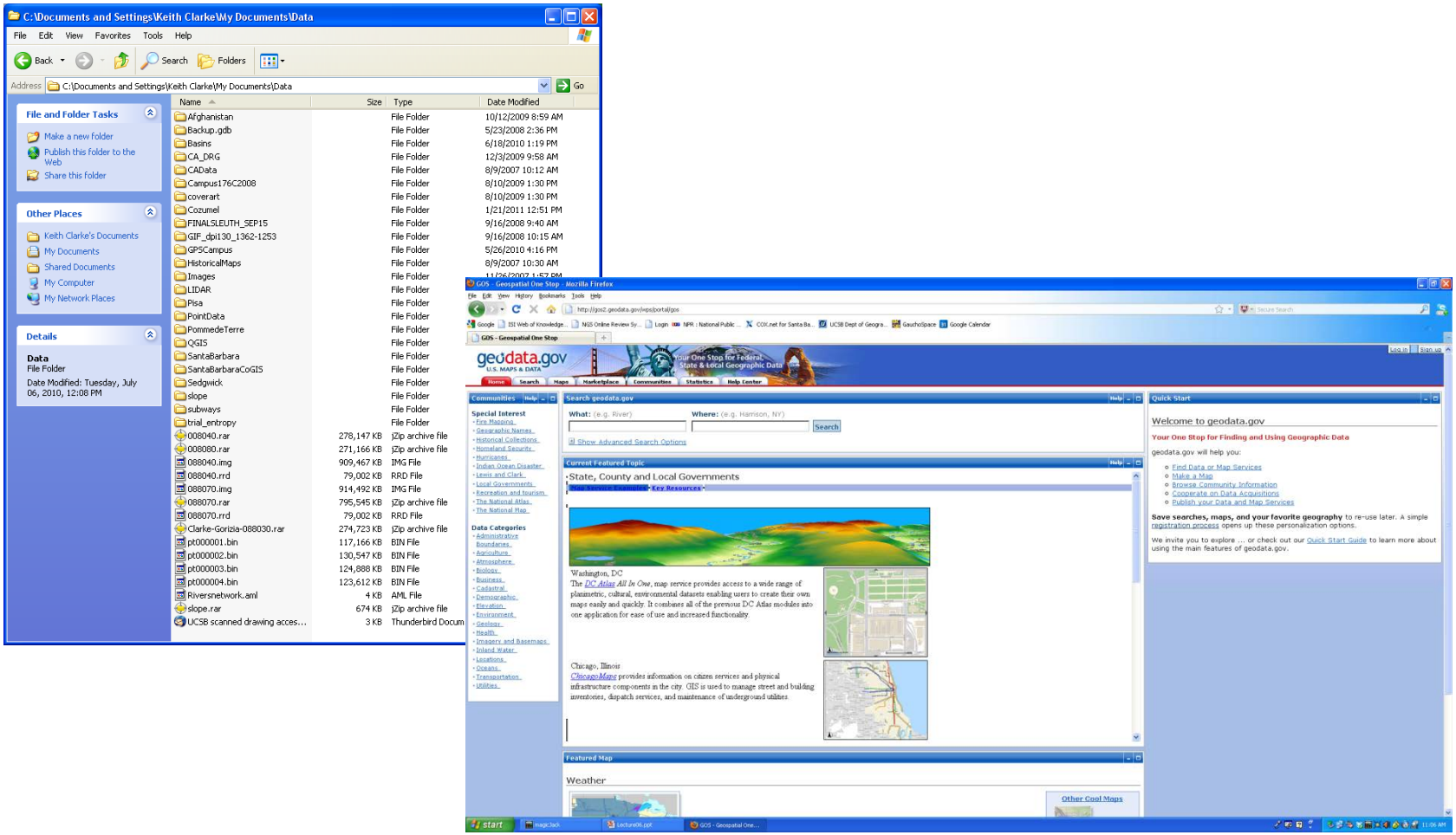




VMAP 0

Open Street Map

Isolated


C:\Documents and Settings\Keith Clarke\My Documents\Data
File Edit View Favorites Tools Help
Back Search Folders
Address C:\Documents and Settings\Keith Clarke\My Documents\Data
Go
File and Folder Tasks
Make a new folder
Publish this folder to the Web
Share this folder
Other Places
Keith Clarke's Documents
My Documents
Shared Documents
My Computer
My Network Places
Details
Data
File Folder
Date Modified: Tuesday, July 06, 2010, 12:08 PM
Name Size Type Date Modified
Afghanistan File Folder 10/12/2009 8:59 AM
Backup.gdb File Folder 5/23/2008 2:36 PM
Basins File Folder 6/18/2010 1:19 PM
CA_DRG File Folder 12/3/2009 9:58 AM
CAData File Folder 8/9/2007 10:12 AM
Campus176C2008 File Folder 8/10/2009 1:30 PM
coverart File Folder 8/10/2009 1:30 PM
Cozumel File Folder 1/21/2011 12:51 PM
FINALSLEUTH_SEP15 File Folder 9/16/2008 9:40 AM
GIF_dpi130_1362-1253 File Folder 9/16/2008 10:15 AM
GPSCampus File Folder 5/26/2010 4:16 PM
HistoricalMaps File Folder 8/9/2007 10:30 AM
Images File Folder
LIDAR File Folder
Pisa File Folder
PointData File Folder
PommedeTerre File Folder
QGIS File Folder
SantaBarbara File Folder
SantaBarbaraCoGIS File Folder
Sedgwick File Folder
slope File Folder
subways File Folder
trial_entropy File Folder
008040.rar 278,147 KB zIp archive file
008080.rar 271,166 KB zIp archive file
088040.img 909,467 KB IMG File
088040.rrd 79,002 KB RRD File
088070.img 914,492 KB IMG File
088070.rar 795,545 KB zIp archive file
088070.rrd 79,002 KB RRD File
Clarke-Gorizia-088030.rar 274,723 KB zIp archive file
pt000001.bin 117,166 KB BIN File
pt000002.bin 130,547 KB BIN File
pt000003.bin 124,888 KB BIN File
pt000004.bin 123,612 KB BIN File
Riversnetwork.aml 4 KB AML File
slope.rar 674 KB zIp archive file
UCSB scanned drawing acces... 3 KB Thunderbird Docum
GOS - Geospatial One Stop - Mozilla Firefox
geodata.gov
U.S. MAPS & DATA
Your One Stop for Federal, State & Local Geographic Data
Home Search Maps Marketplace Communities Statistics Help Center
Communities
Special Interest
Data Categories
Search geodata.gov
What: (e.g. River)
Where: (e.g. Harrison, NY)
Search
Show Advanced Search Options
Current Featured Topic
State, County and Local Governments
Key Resources
Washington, DC
The DC Atlas All In One, map service provides access to a wide range of planimetric, cultural, environmental datasets enabling users to create their own maps easily and quickly. It combines all of the previous DC Atlas modules into one application for ease of use and increased functionality.
Chicago, Illinois
ChicagoMaps provides information on citizen services and physical infrastructure components in the city. GIS is used to manage street and building inventories, dispatch services, and maintenance of underground utilities.
Quick Start
Welcome to geodata.gov
Your One Stop for Finding and Using Geographic Data
geodata.gov will help you:
Find Data or Map Services
Make a Map
Browse Community Information
Cooperate on Data Acquisitions
Publish your Data and Map Services
Save searches, maps, and your favorite geography to re-use later. A simple registration process opens up these personalization options.
We invite you to explore ... or check out our Quick Start Guide to learn more about using the main features of geodata.gov.
Featured Map
Weather
Other Cool Maps
start

Denied Data

The Attorney General
Washington, [redacted]

[redacted], 2[redacted]7

CLASSIFIED
TOP SECRET

President George W. Bush
[redacted]
Washington, D.C.

Dear [redacted]

It has always been my honor to serve at your pleasure. After much [redacted] and consideration, [redacted] believe this is the right time [redacted] to begin [redacted] our lives. Please accept my [redacted] as Attorney General of the United States, effective [redacted].

I am proud to have served with the career investigators, administrative and support staff, litigators, and prosecutors in the Department of [redacted]. It is because of their hard work and professionalism that our country today is safer from acts of terrorism, our neighborhoods suffer less from violent crime, and our [redacted] are better protected from [redacted].

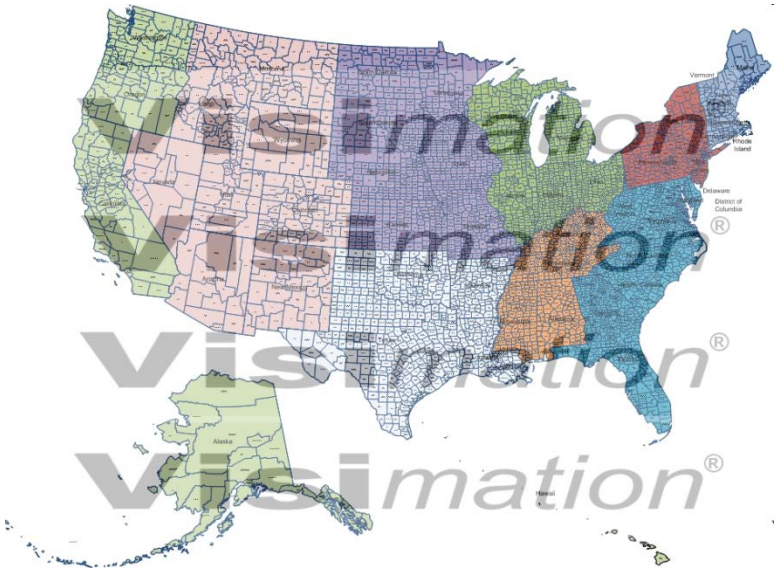
You have my deepest respect and admiration for the courage you have shown in leading our country through this extraordinary time[redacted]. Every day as the Attorney General has been a privilege and I am honored by your trust in me. Thank you for this opportunity to serve the [redacted].

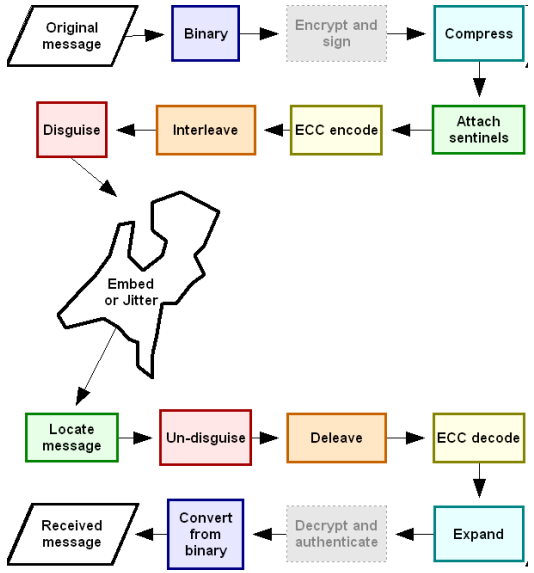
[redacted] are grateful for your friendship. We will continue to keep you and [redacted] in our prayers and ask [redacted] to provide both of you the [redacted], strength, and faith you need to lead [redacted] forward.

I remain by your [redacted],

[redacted]

Alberto [redacted] Gonzales





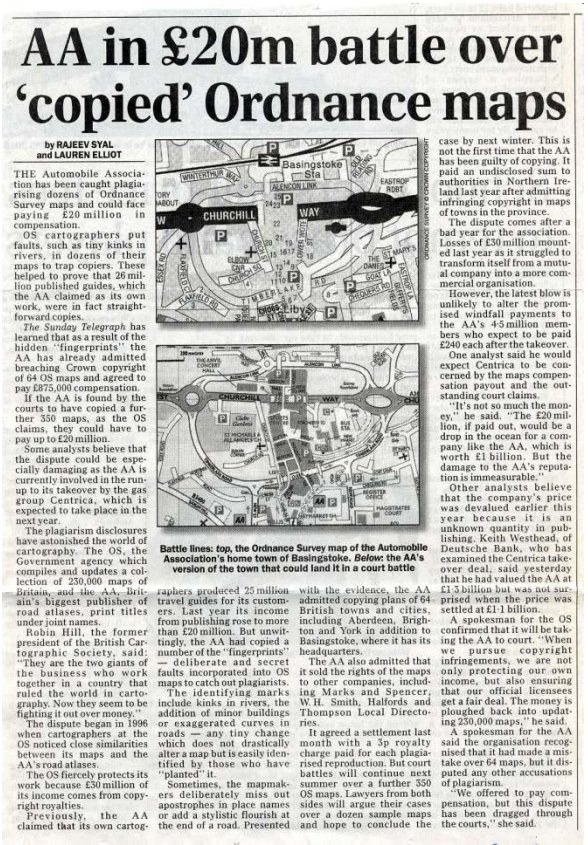
Hidden Data

Image of a tree. Removing all but the two least significant bits of each color component produces an almost completely black image. Making that image 85 times brighter produces the image below.

Image of a cat extracted from above image.

Cartographic traps

- Kerbela Street, Shrewsbury, England. Just off Meadow Farm Drive in Shrewsbury Google Maps shows a Kerbela Street, which does not physically exist at this location or even exist at all in Shrewsbury, according to Royal Mail.



AA in £20m battle over 'copied' Ordnance maps

by RAJEEV SYAL and LAUREN ELLIOT

THE Automobile Association has been caught plagiarising dozens of Ordnance Survey maps and could face paying £20 million in compensation.

OS cartographers put faults, such as tiny kinks in rivers, in dozens of their maps to trap copiers. These helped to prove that 26 million published guides, which the AA claimed as its own work, were in fact straightforward copies.

The Sunday Telegraph has learned that as a result of the hidden "fingerprints" the AA has already admitted breaching Crown copyright of 64 OS maps and agreed to pay £875,000 compensation.

If the AA is found by the courts to have copied a further 350 maps, as the OS claims, they could have to pay up to £20 million.

Some analysts believe that the dispute could be especially damaging as the AA is currently involved in the run-up to its takeover by the gas group Centrica, which is expected to take place in the next year.

The plagiarism disclosures have astonished the world of cartography. The OS, the Government agency which compiles and updates a collection of 230,000 maps of Britain, and the AA, Britain's biggest publisher of road atlases, print titles under joint names.

Robin Hill, the former president of the British Cartographic Society, said: "They are the two giants of the business who work together in a country that ruled the world in cartography. Now they seem to be fighting it out over money."

The dispute began in 1996 when cartographers at the OS noticed close similarities between its maps and the AA's road atlases.

The OS fiercely protects its work because £30 million of its income comes from copyright royalties.

Previously, the AA claimed that its own cartographers produced 25 million travel guides for its customers. Last year its income from publishing rose to more than £20 million. But unwittingly, the AA had copied a number of the "fingerprints" — deliberate and secret faults incorporated into OS maps to catch out plagiarists.

The identifying marks include kinks in rivers, the addition of minor buildings or exaggerated curves in roads — any tiny change which does not drastically alter a map but is easily identified by those who have "planted" it.

Sometimes, the mapmakers deliberately miss out apostrophes in place names or add a stylistic flourish at the end of a road. Presented with the evidence, the AA admitted copying plans of 64 British towns and cities, including Aberdeen, Brighton and York in addition to Basingstoke, where it has its headquarters.

The AA also admitted that it sold the rights of the maps to other companies, including Marks and Spencer, W.H. Smith, Halfords and Thompson Local Directories.

It agreed a settlement last month with a 3p royalty charge paid for each plagiarised reproduction. But court battles will continue next summer over a further 350 OS maps. Lawyers from both sides will argue their cases over a dozen sample maps and hope to conclude the case by next winter. This is not the first time that the AA has been guilty of copying. It paid an undisclosed sum to authorities in Northern Ireland last year after admitting infringing copyright in maps of towns in the province.

The dispute comes after a bad year for the association. Losses of £30 million mounted last year as it struggled to transform itself from a mutual company into a more commercial organisation.

However, the latest blow is unlikely to alter the promised windfall payments to the AA's 4·5 million members who expect to be paid £240 each after the takeover.

One analyst said he would expect Centrica to be concerned by the maps compensation payout and the outstanding court claims.

"It's not so much the money," he said. "The £20 million, if paid out, would be a drop in the ocean for a company like the AA, which is worth £1 billion. But the damage to the AA's reputation is immeasurable."

Other analysts believe that the company's price was devalued earlier this year because it is an unknown quantity in publishing. Keith Westhead, of Deutsche Bank, who has examined the Centrica takeover deal, said yesterday that he had valued the AA at £1·5 billion but was not surprised when the price was settled at £1·1 billion.

A spokesman for the OS confirmed that it will be taking the AA to court. "When we pursue copyright infringements, we are not only protecting our own income, but also ensuring that our official licensees get a fair deal. The money is ploughed back into updating 230,000 maps," he said.

A spokesman for the AA said the organisation recognised that it had made a mistake over 64 maps, but it disputed any other accusations of plagiarism.

"We offered to pay compensation, but this dispute has been dragged through the courts," she said.

Battle lines: *top*, the Ordnance Survey map of the Automobile Association's home town of Basingstoke. *Below*: the AA's version of the town that could land it in a court battle

Future Memory: Efficiency vs. Access

Whole internet on a disk?

Search systems: Information value vs. content

Dynamic and Real Time GIS

Summary

- Maps can be converted to data by digitizing, scanning or be native digital
- Digital conversion is geocoding: adding coordinates, “geotags”, metadata
- Data formats need logical data models
- Different models retain different geographical measures and properties
- Ultimately, geography is encoded into bits and bytes on storage media
- Many factors can restrict the efficiency of map encoding
- Map data storage must balance storage efficiency with ease of use