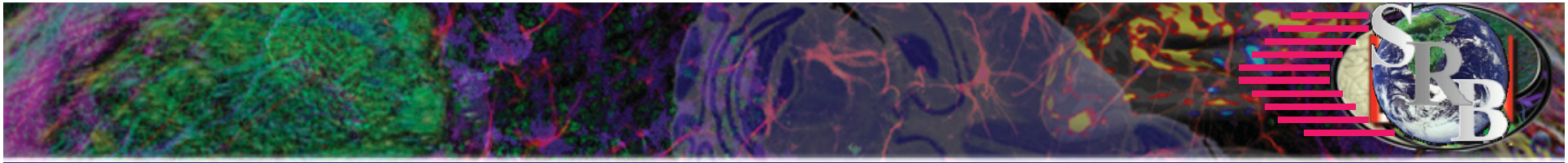


Preservation Environments for GIS Systems

Reagan Moore

Richard Marciano

Ilyz Zaslavsky

San Diego Supercomputer Center

UCSD

SDSC

Topics

- **Preservation concepts and infrastructure**
  - Authenticity
  - Integrity
  - Infrastructure independence
- **Preservation of GIS systems**
  - NARA
    - Vietnam war Herbicides data (Ilya Zaslavsky zaslavsk@sdsc.edu)
    - Home Owners Loan Corporation (Richard Marciano marciano@sdsc.edu)
  - NHPRC Persistent Archive Testbed
    - Michigan precinct voting records (Richard Marciano marciano@sdsc.edu)
    - Maine GeoArchives (Ilya Zaslavsky zaslavsk@sdsc.edu)
  - InterPARES
    - VanMap project

UCSD SDSC

Digital Preservation

- **Extract a digital record from the environment in which it was created**
  - Accession into the preservation environment
- **Preserve authenticity**
  - Maintain links from provenance metadata to the digital record, and chain of custody
- **Preserve integrity**
  - Maintain persistent naming, persistent access controls, checksums, replicas

UCSD

SDSC

Infrastructure Independence

- **Concept that the preserved records can be migrated from the current preservation environment into another choice of technology, while preserving authenticity and integrity.**
- **Challenge is that all components of the preservation environment will evolve:**
  - Hardware systems
  - Software systems
  - Encoding formats
  - Access mechanisms
  - Preservation attributes

UCSD

SDSC

Data Grids

- **Manage technology evolution for software and hardware systems**
  - Data virtualization - manage data collection properties independently of the storage systems
    - Assert fixity properties on the data collection while storing in an evolving storage system
  - Trust virtualization - manage access controls and authentication independently of the storage systems
    - Assert fixity properties on the name spaces while storing across administrative domains

UCSD

SDSC

Storage Resource Broker 3.3.1

Preservation Environment

C Library, Java	Unix Shell	Linux I/O C++	NT Browser, Kepler Actors, Cheshire, Transana	DLL / Python, Perl, Windows	HTTP, DSpace, Fedora, OpenDAP	OAI, WSDL, (WSRF), GridFTP

Federation Management

Consistency & Metadata Management / Authorization, Authentication, Audit

Logical Name Space	Latency Management	Data Transport	Metadata Transport

Database Abstraction	Storage Repository Abstraction			
Databases - DB2, Oracle, Sybase, Postgres, mySQL, Informix	Archives - Tape, Sam-QFS, DMF, HPSS, ADSM, UniTree, ADS	ORB	File Systems Unix, NT, Mac OSX	Databases - DB2, Oracle, Sybase, Postgres, mySQL, Informix

UCSD

SDSC

Storage Resource Broker Collections at SDSC (8/2/2005)	GBs of data stored	Number of files	Users with ACLs
Data Grid	Ź	Ź	Ź
NSF/ITR - National Virtual Observatory	53,862	9,536,751	100
NSF - National Partnership for Advanced Computational Infrastructure	36,149	7,539,180	380
Static collections Š Hayden planetarium	8,013	161,352	227
Pzone Š public collections	12,998	6,707,952	68
NSF/NPACI - Biology and Environmental collections	40,155	76,083	67
NSF/NPACI Š Joint Center for Structural Genomics	15,731	1,577,260	55
NSF - TeraGrid, ENZO Cosmology simulations	176,730	2,125,945	3,267
NIH - Biomedical Informatics Research Network	10,561	7,596,888	303
Digital Library	Ź	Ź	Ź
NSF/NPACI - Long Term Ecological Reserve	256	9,033	36
NSF/NPACI - Grid Portal	2,620	53,048	460
NIH - Alliance for Cell Signaling microarray data	741	84,594	21
NSF - National Science Digital Library SIO Explorer collection	2,733	1,083,998	27
NSF/ITR - Southern California Earthquake Center	131,010	2,702,421	73
Persistent Archive	Ź	Ź	Ź
NHPRC Persistent Archive Testbed (Kentucky, Ohio, Michigan, Minnesota)	100	382,186	28
UCSD Libraries archive	4,147	408,050	29
NARA- Research Prototype Persistent Archive	1,478	893,434	58
NSF - National Science Digital Library persistent archive	3,600	27,034,150	136
TOTAL	501 TB	68 million	5,335

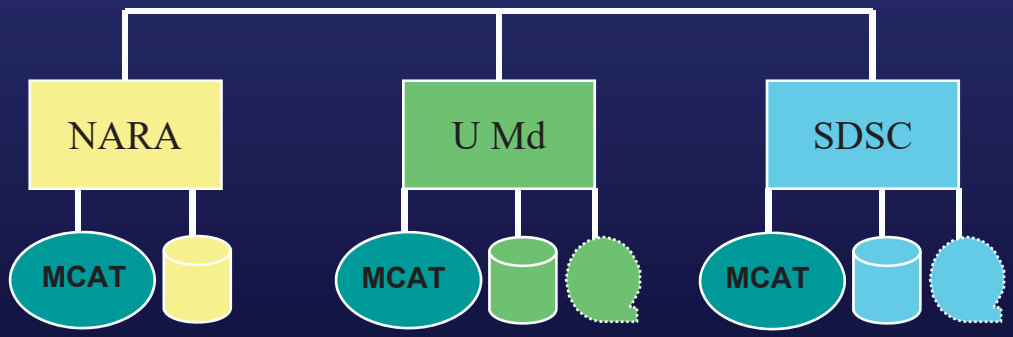
Storage Resource Broker

NARA Research Prototype Persistent Archive

Demonstrate preservation environment

- Authenticity
  - Producer-Archive submission pipeline
  - LCDRG metadata
- Integrity
  - Replication of data
  - Federation of catalogs
- Infrastructure independence
  - MCAT - Oracle, Informix
  - Storage - HPSS, Sun disk, commodity disk
- Scalability
  - EAP collection - 350,000 files

Federation of Three Independent Data Grids



Principle copy stored at NARA with complete metadata catalog

Replicated copy at U Md for improved access, load balancing and disaster recovery

Deep Archive at SDSC, no user access, but complete copy

UCSD

SDSC

Preserving GIS Systems

- **GIS system is a compound digital object**
  - Disassociate into components
- **Preserve the components**
  - Configuration file defining interaction between layers
  - Data layers which are geo-referenced
  - Data layer metadata defining layer properties
  - Data layer markup standard
  - Geo-reference display standard operations

UCSD

SDSC

Preservation Strategies

- **Emulation**
  - Migrate the display application onto new operating systems
- **Transformative migration**
  - Migrate the encoding format to a preservation standard
  - Formatting standard is expected to be good for 5-10 years
- **Persistent object**
  - Characterize the encoding format
  - Parse the original record based on the characterization, keeping the original record unchanged
  - Migrate the characterization forward in time

UCSD

SDSC

Graphics Standards

- Shape files
  - Definition for polygon shapes in GIS system
- SVG - Scalable Vector Graphics
  - Define end-points of vectors
- GML - Geographic markup language
  - XML vocabulary for capturing spatial content
  - Specify layers, formats, coordinate system
  - Annotation of names for shapes
  - GML viewers - Universal Viewers operate directly on GML
  - NARA accepts GML formatted graphics files

Generic Transformation Software

- Safe Software
  - Transforms between spatial formats
  - Transforms shape files (polygons) to GML
- Based on transformation workflows
  - Self documenting workflow description
  - Can be part of transformative migration pipeline
- Authenticity metadata still required
  - Need to document the workflow components
  - Identify features that do not get transformed

Preserving GIS Environments

- **Thought experiment**
  - Is it possible to import paper records into a GIS environment?
  - Are the required steps the same for preserving the components of a GIS environment?

UCSD

SDSC

Preservation of GIS

- **Creation of GIS system from paper records**
  - Homeowners Loan Corporation - Red-lining
- **Creation of GIS system from digital records**
  - Michigan voting records
- **Re-creation of GIS system from archived GIS data**
  - Vietnam war Herbicides data
- **Creation of GIS records from active GIS environment**
  - InterPARES VanMap project

UCSD

SDSC

Home Owners Loan Corporation

- Civilian records preserved on paper at NARA II

- Demonstration of the conversion of paper records into a GIS environment
  - Photocopy the paper records
  - Digitize the paper records
  - Convert maps from shape files to vector graphics
  - Convert vector graphics to GML
  - Create separate GIS layer for each time snapshot

Homeowners Loan Corporation (HOLC)

Paper index of NARA holdings

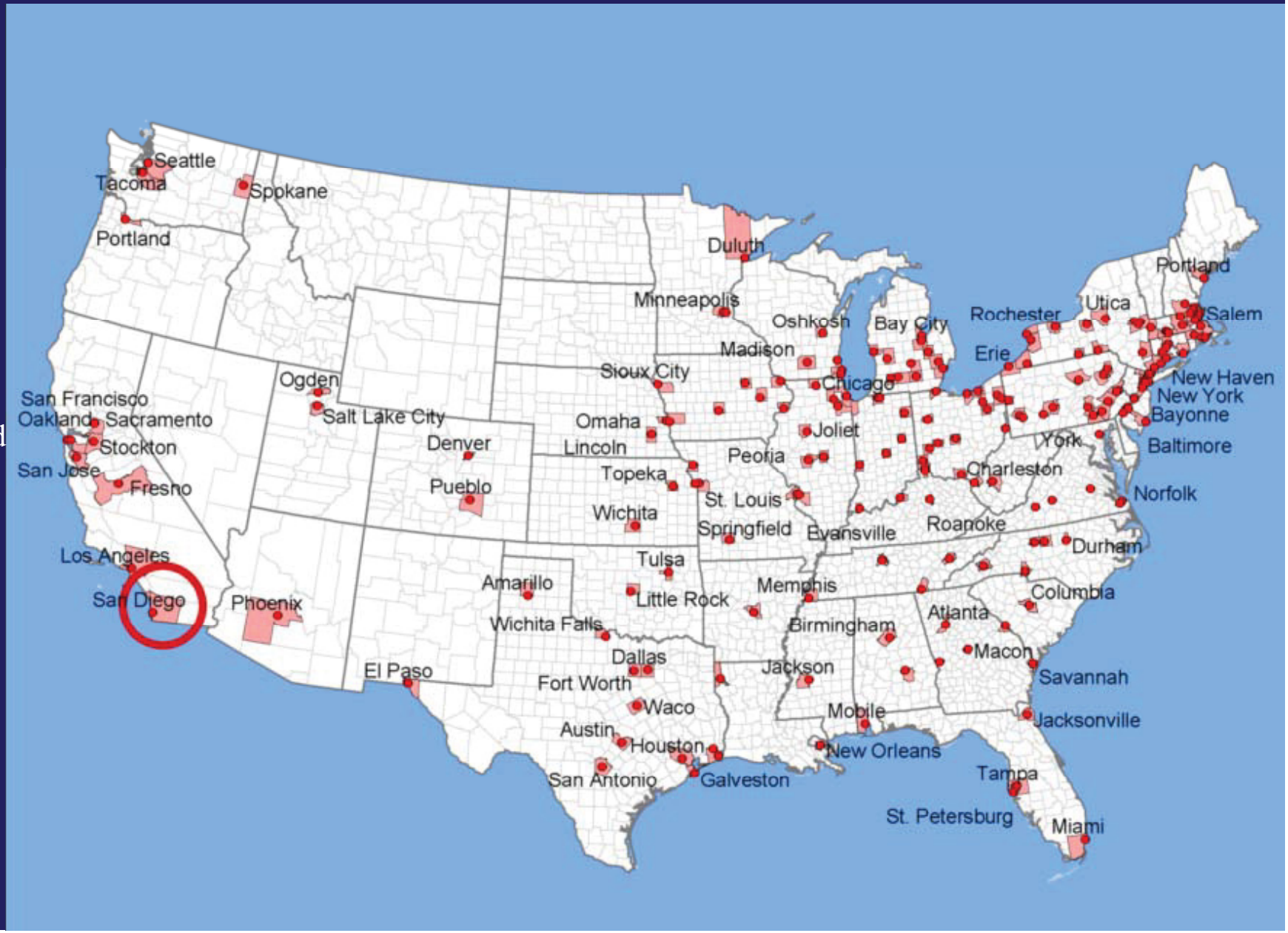
Digitized index

Spatial index for towns and cities

Geo-reference

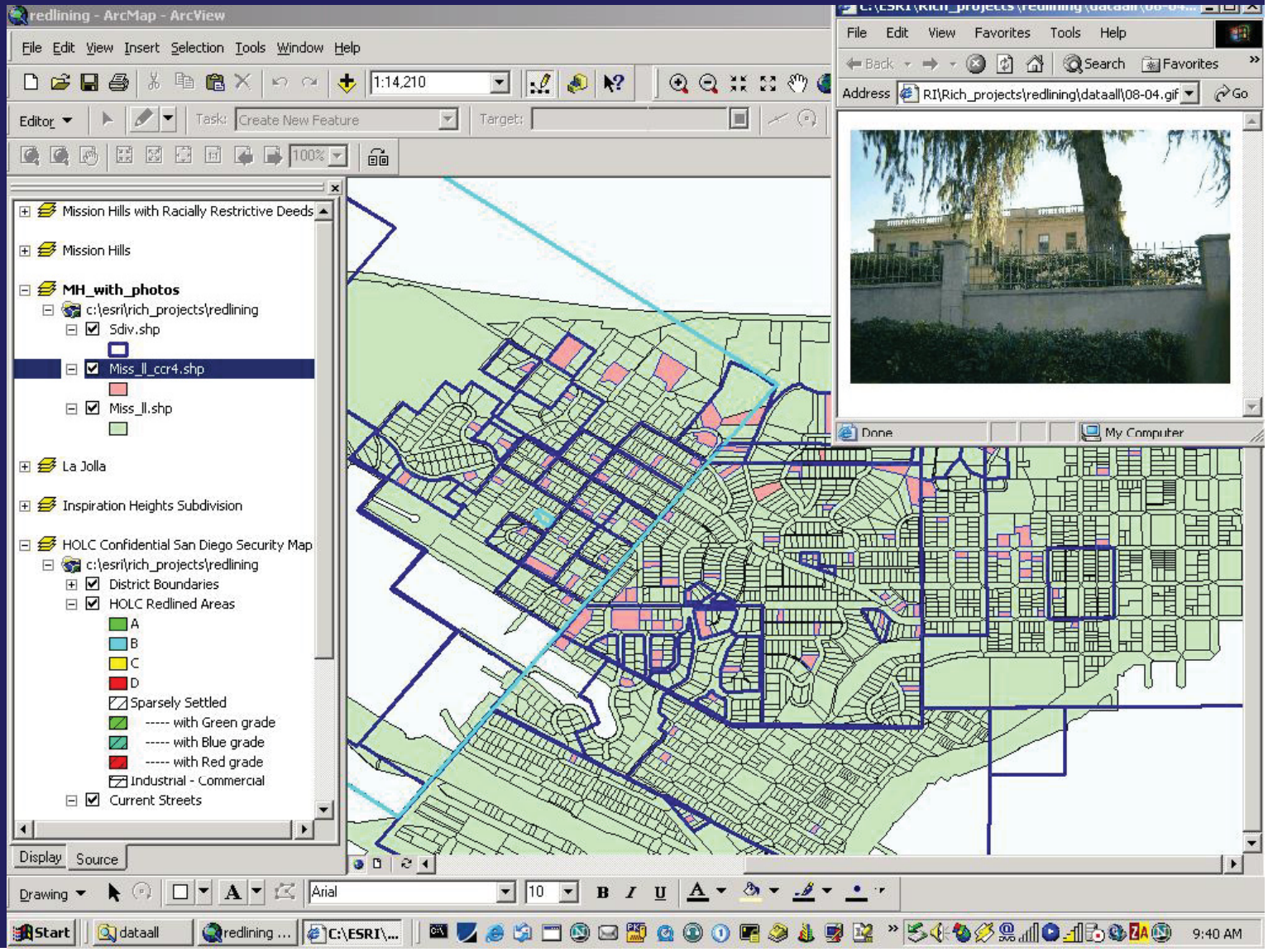
Acquire base map

Highlight cities and counties



Historical Subdivisions Based on Parcel Maps for Mission Hills, (Old Town) San Diego Shown in ESRI ArcView

- Digitize 50 historical maps
- Spatially reference subdivision maps
- Vectorize the maps (shape to GML)
- Create 50 layers
- Import into ESRI ArcView
- Query temporal changes



Viewing of NARA maps & documents (civilian records)

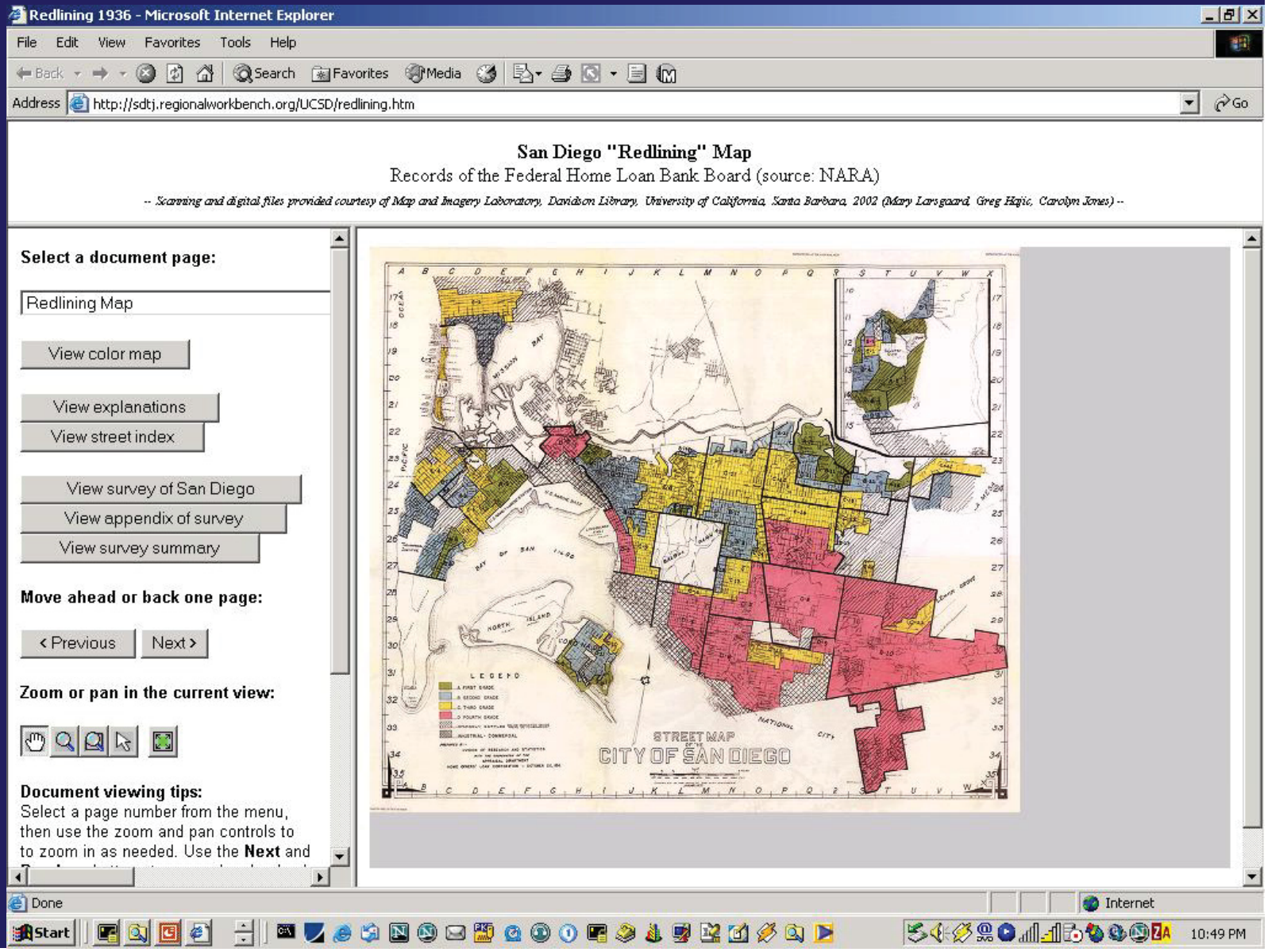
Digitize base map as Tiff

Digitize documents

View associated documents as a function of location

Used to assess impact of redlining - assessing neighborhood risk for mortgage insurance

Tiff viewer



Indexing HOLC Collection

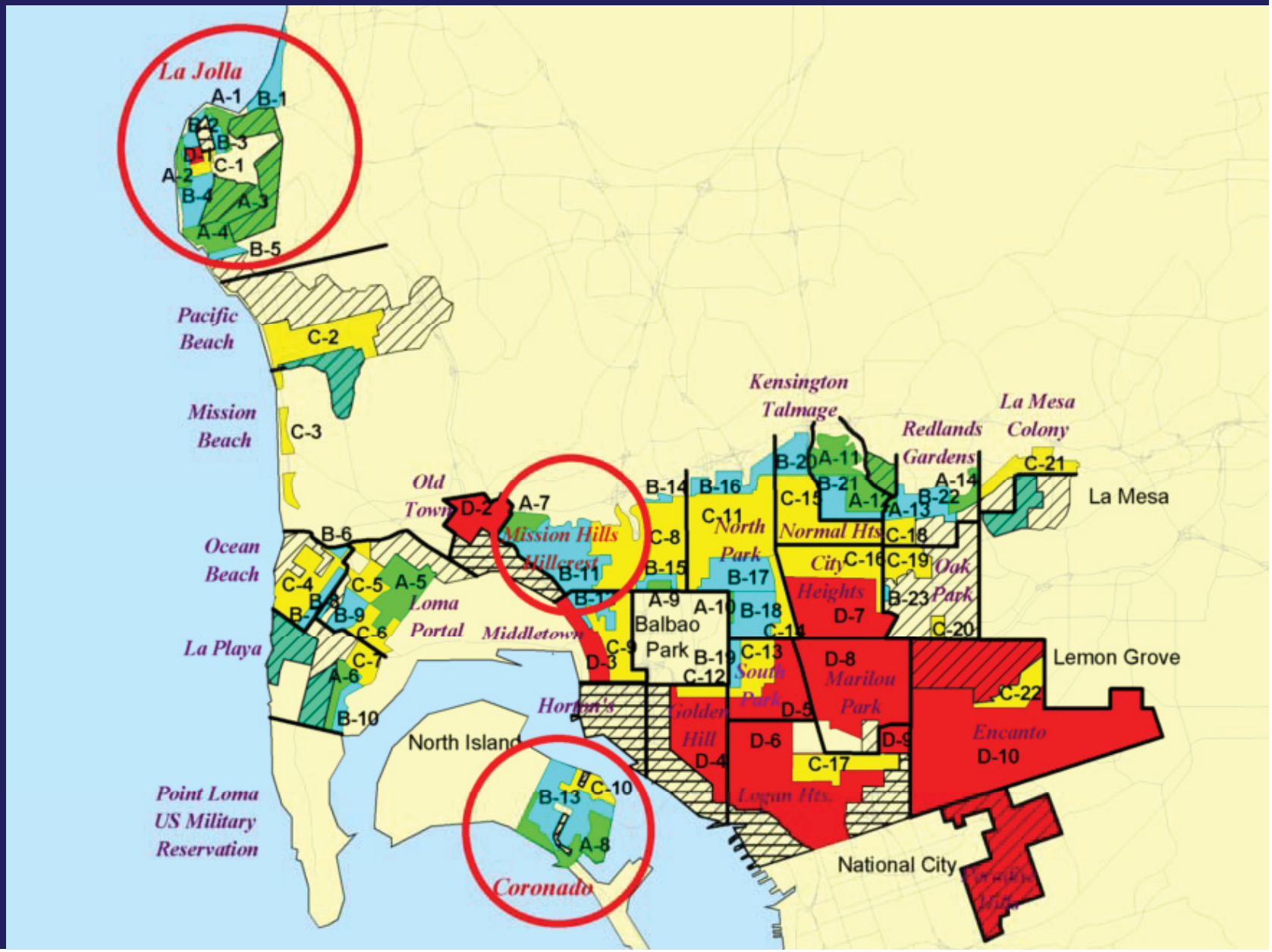
Digitize map and create 69 areas

Convert from raster to vector

Vectorize each red-line parcel

Each parcel is now clickable

ESRI shape file for each area

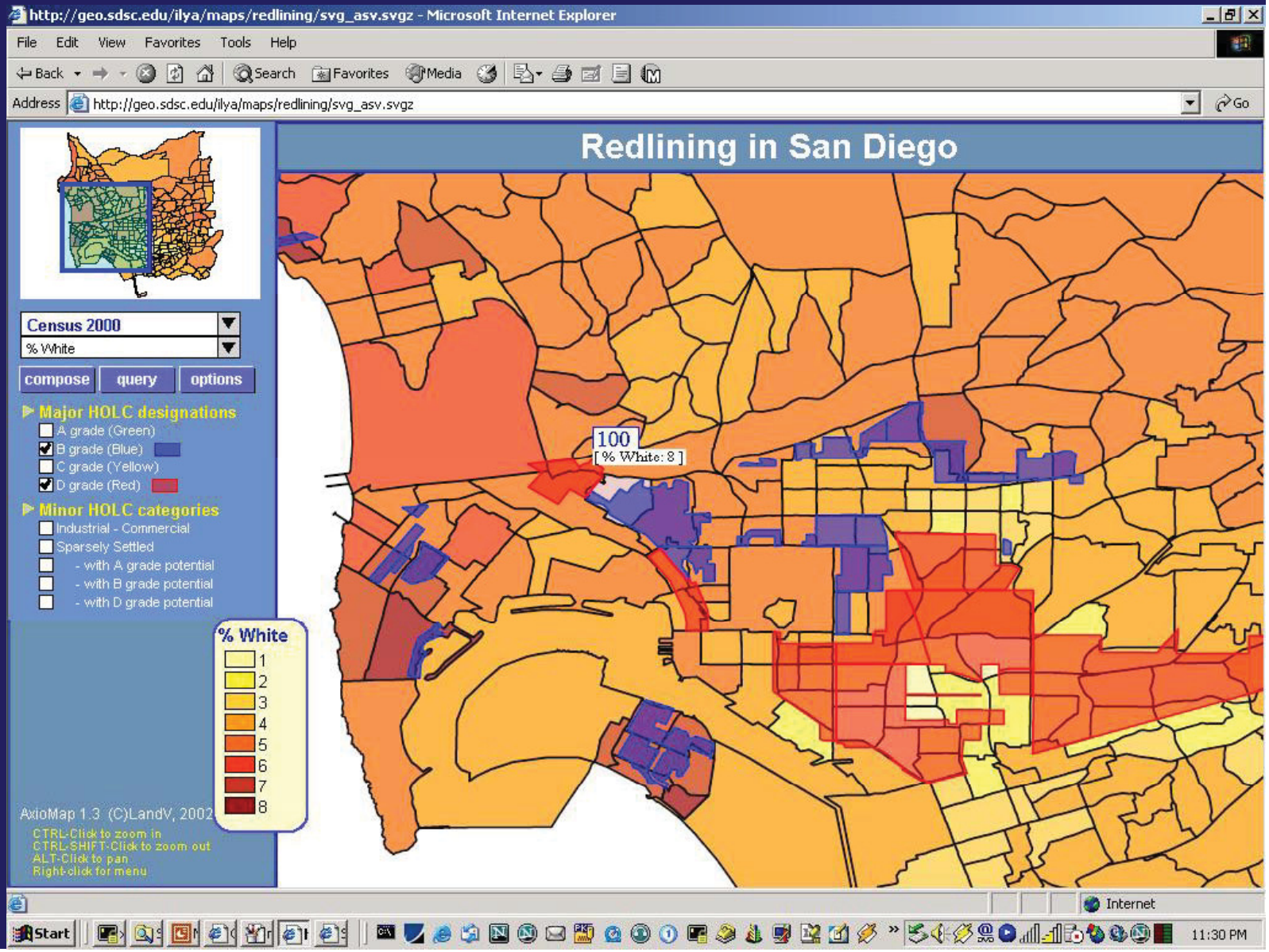


Color Separation of Redlining Zones (SVG) (1/3)

Convert from shape file to SVG

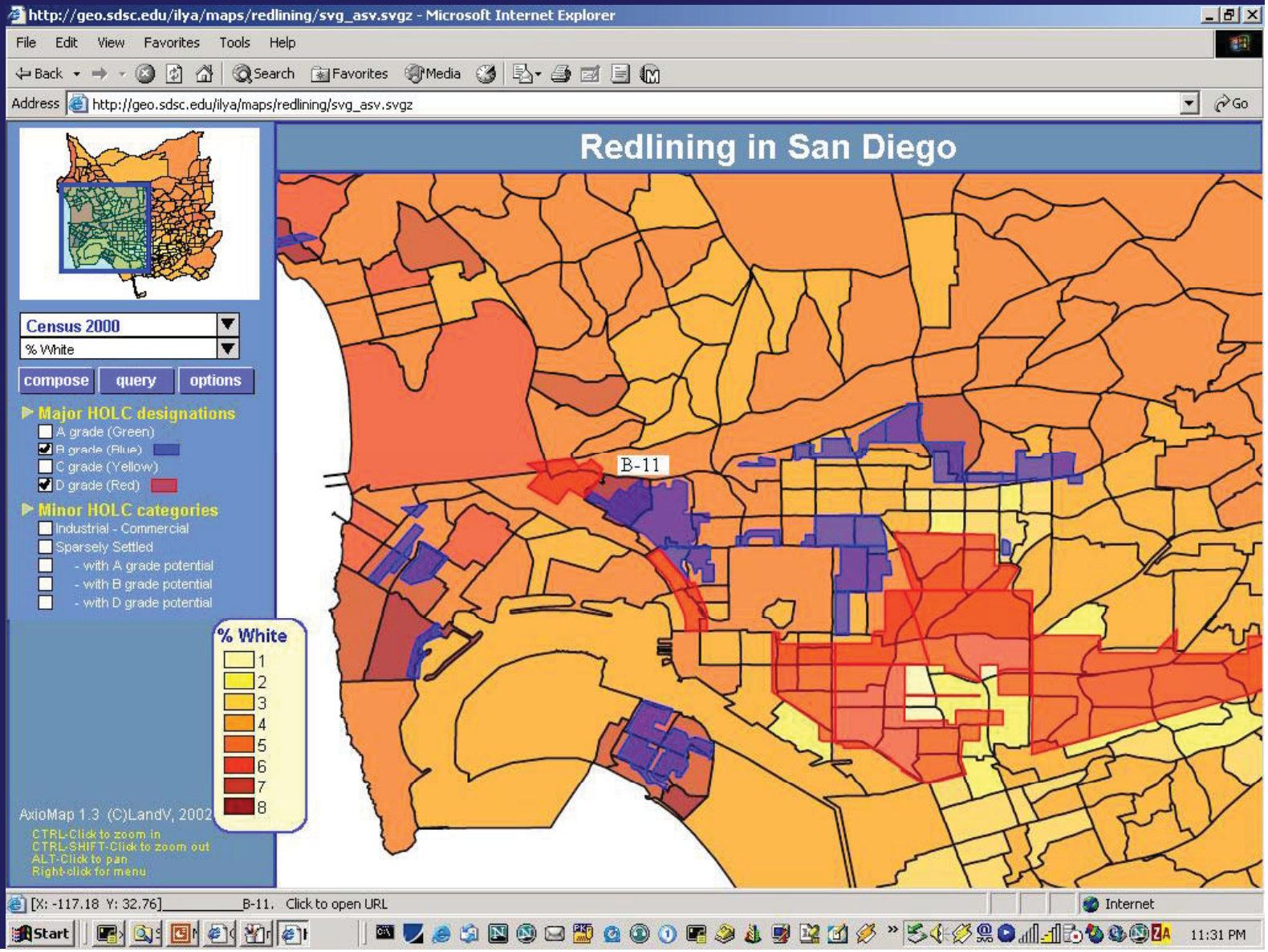
Group the 69 layers into four SVG layers by HOLC grade

Select HOLC layer and overlay on Census 2000 data for the % white population per Census tract



Creating GIS Environment (2/3)

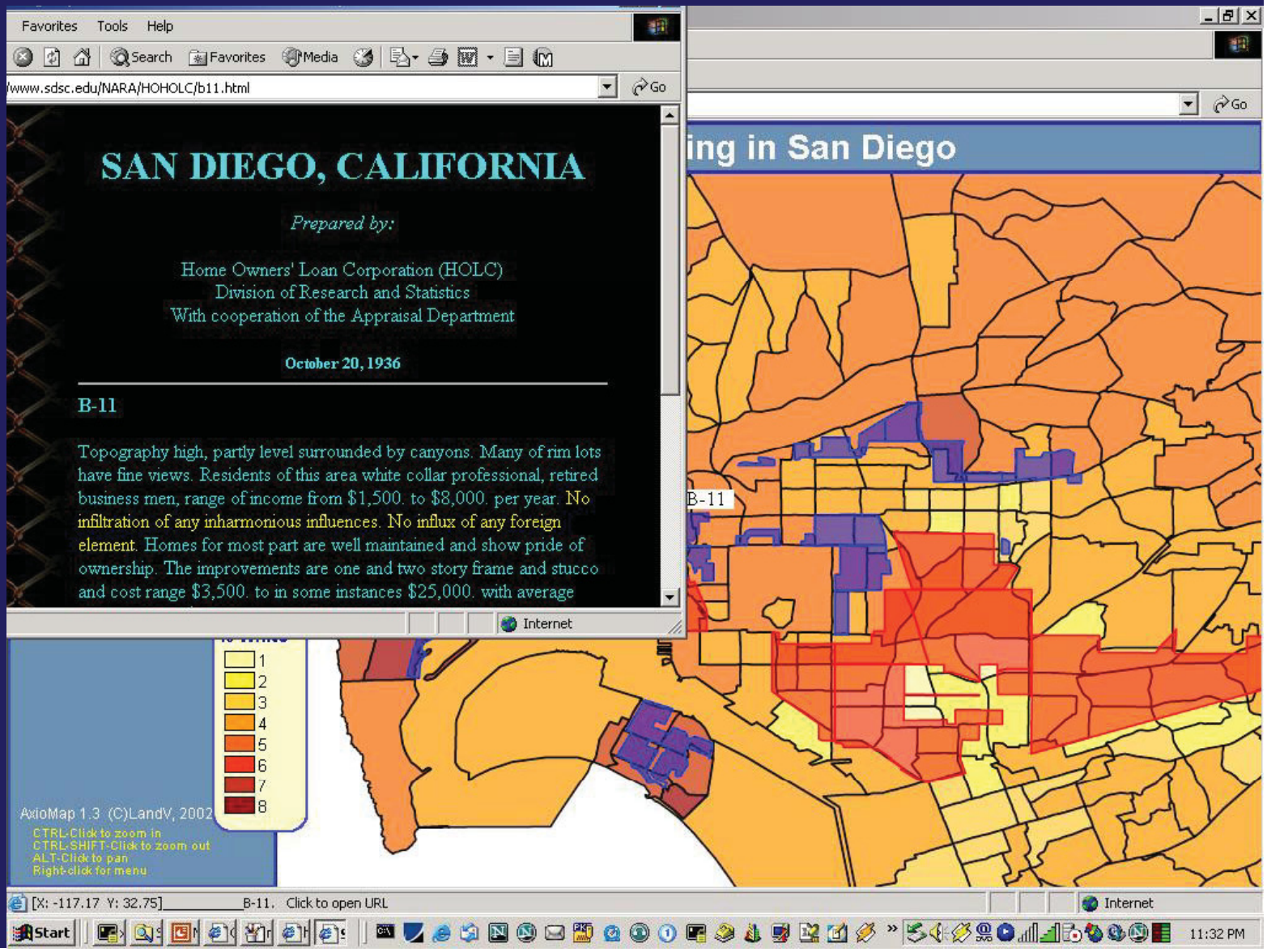
Add dynamic display of historic label as move cursor



Creating GIS Environment (3/3)

Click on the historic label and display the historic document

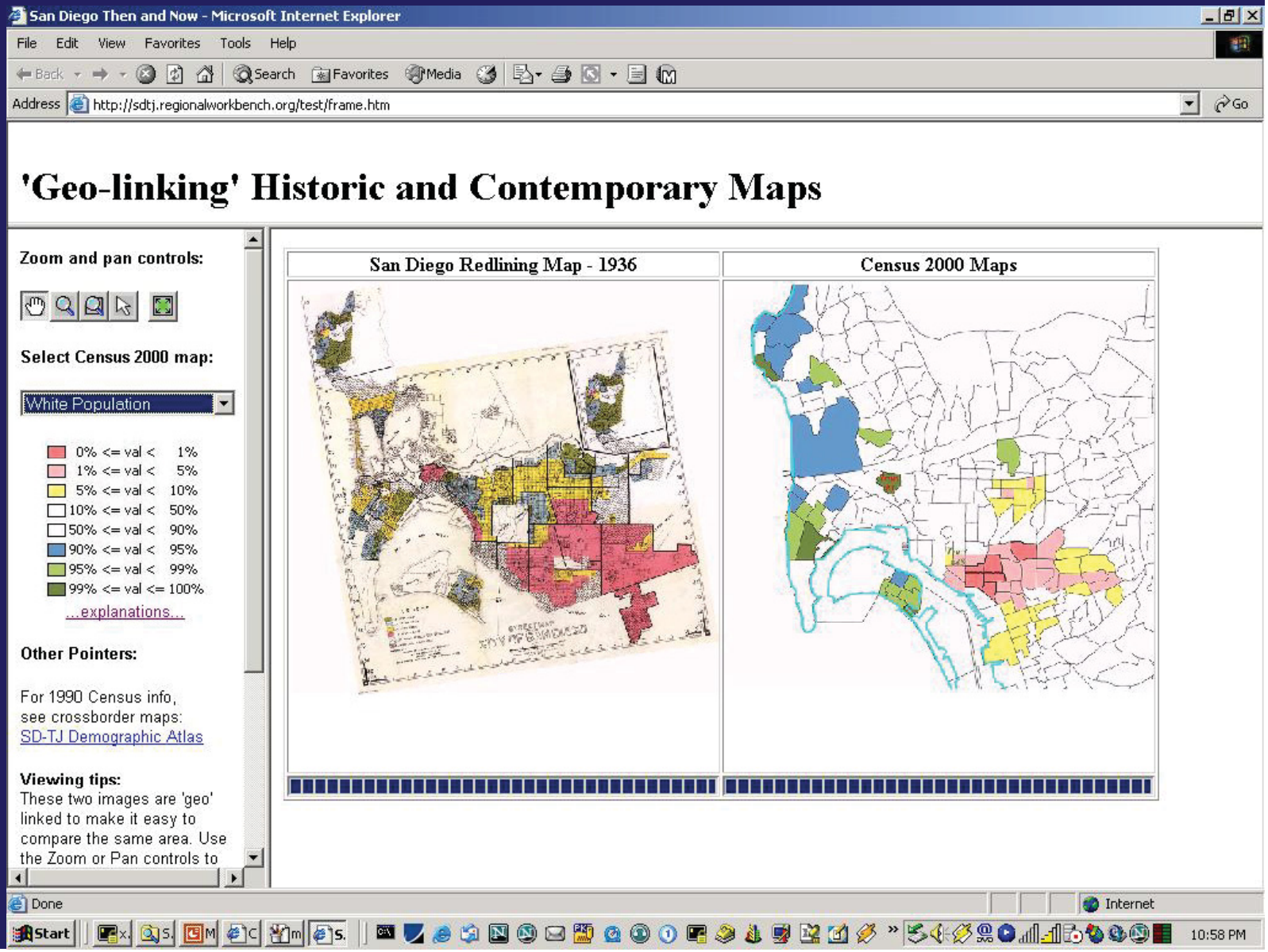
Geo-linking of documents for Mission Hills



Geo-Linking Past & Present (1/3)

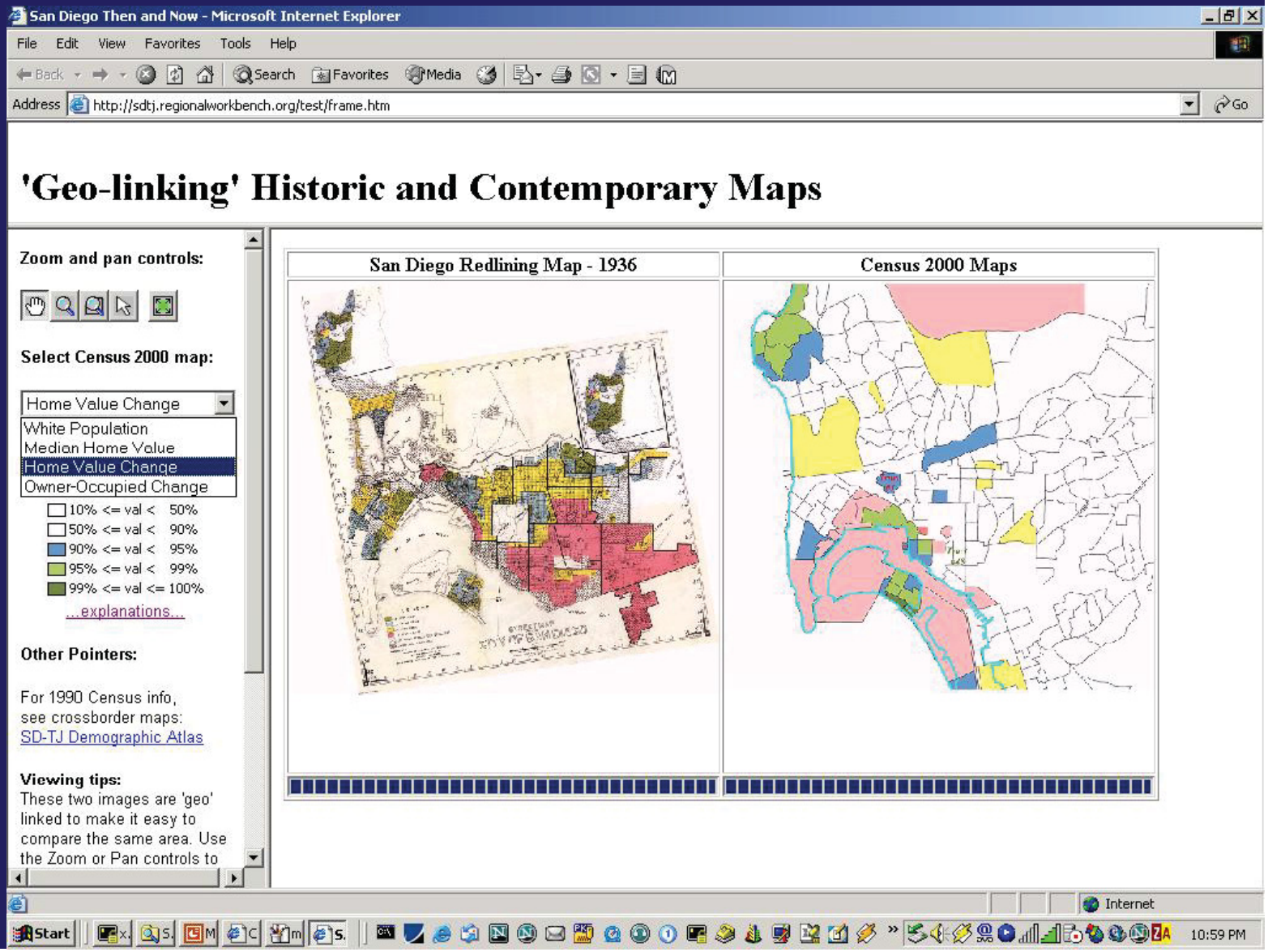
Geo-link the historic and present Census maps

As zoom into one map, the other map synchronizes to the same view



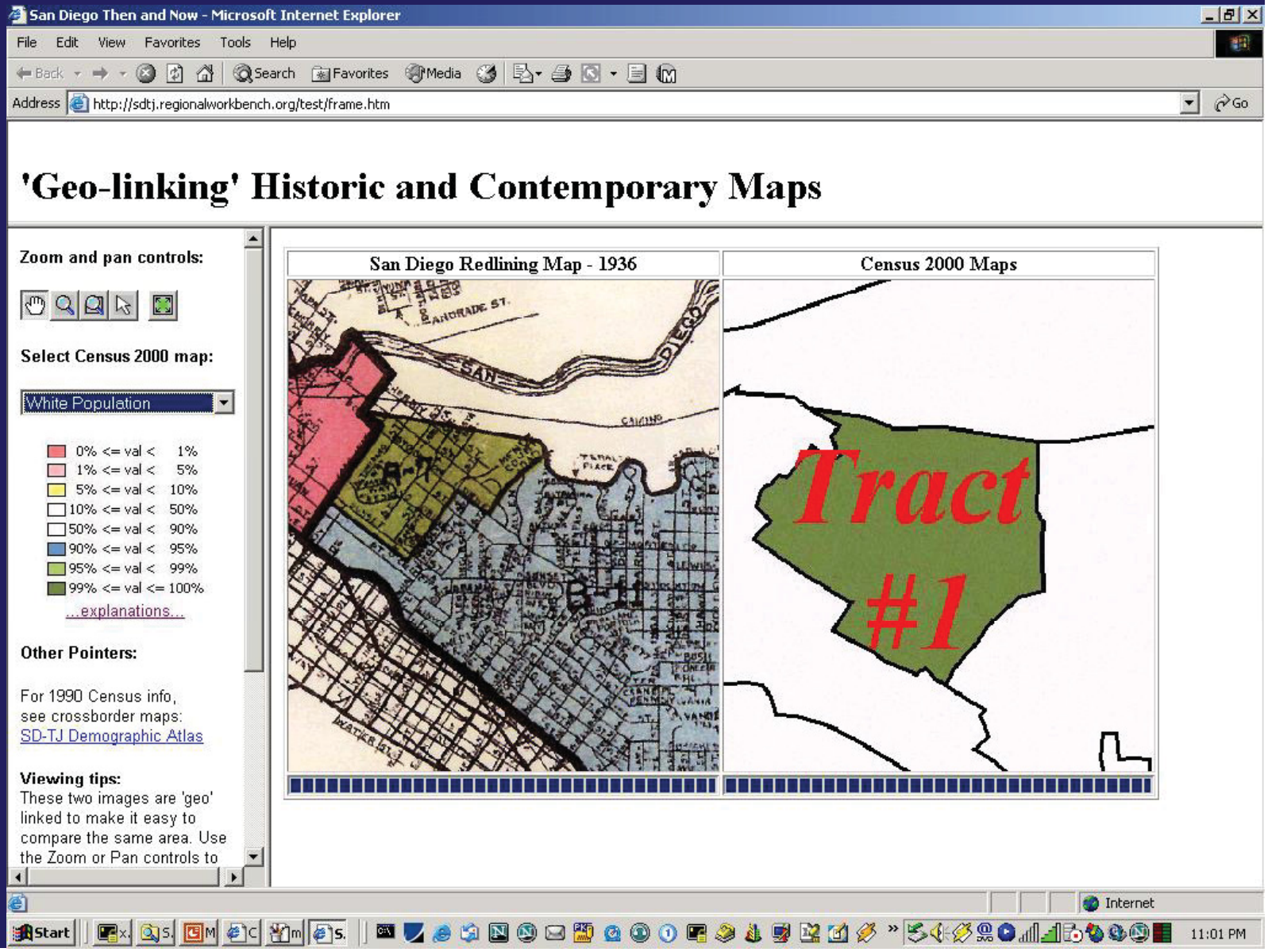
GEO-Linking Maps (2/3)

Change to home value census variable



Geo-Linking Maps (3/3)

Zoomed into the view



Persistent Archive Testbed Project
Michigan Precinct Level Election Results

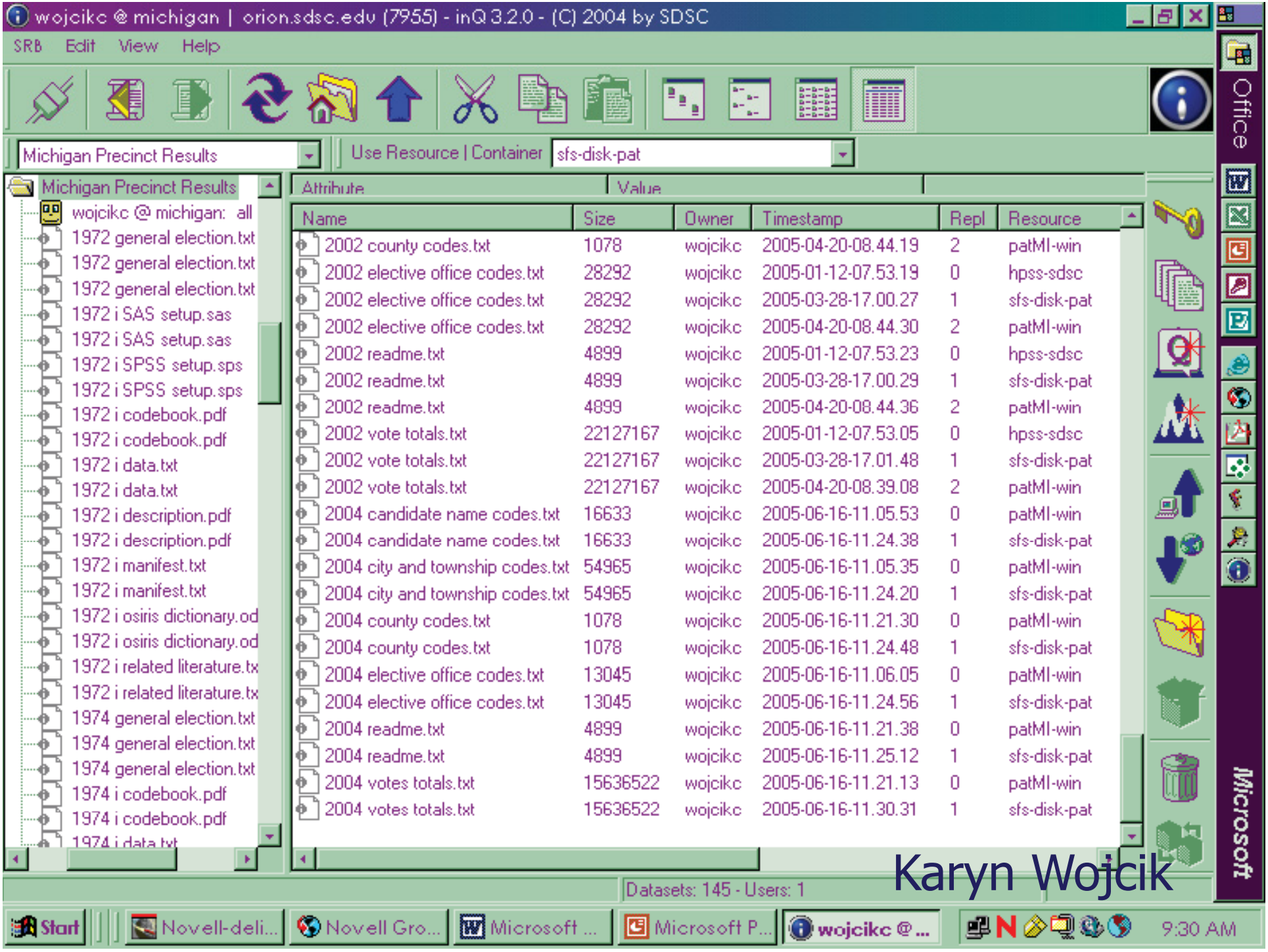
Caryn Wojcik
Michigan Department of History, Arts and Libraries
Michigan Historical Center
Records Management Services
(517) 335-8222
Wojcikc@michigan.gov
http://www.michigan.gov/recordsmanagement
http://www.michigan.gov/statearchives

Karyn Wojcik

Step #1: Data Recovery

- Michigan Department of Community Health magnetic tape machine → broke
- Muller Media ~ $50 per tape
- Load data into the Storage Resource Broker (SRB)
- Replicate data on Michigan grid brick and 2 SDSC grid bricks

Karyn Wojcik

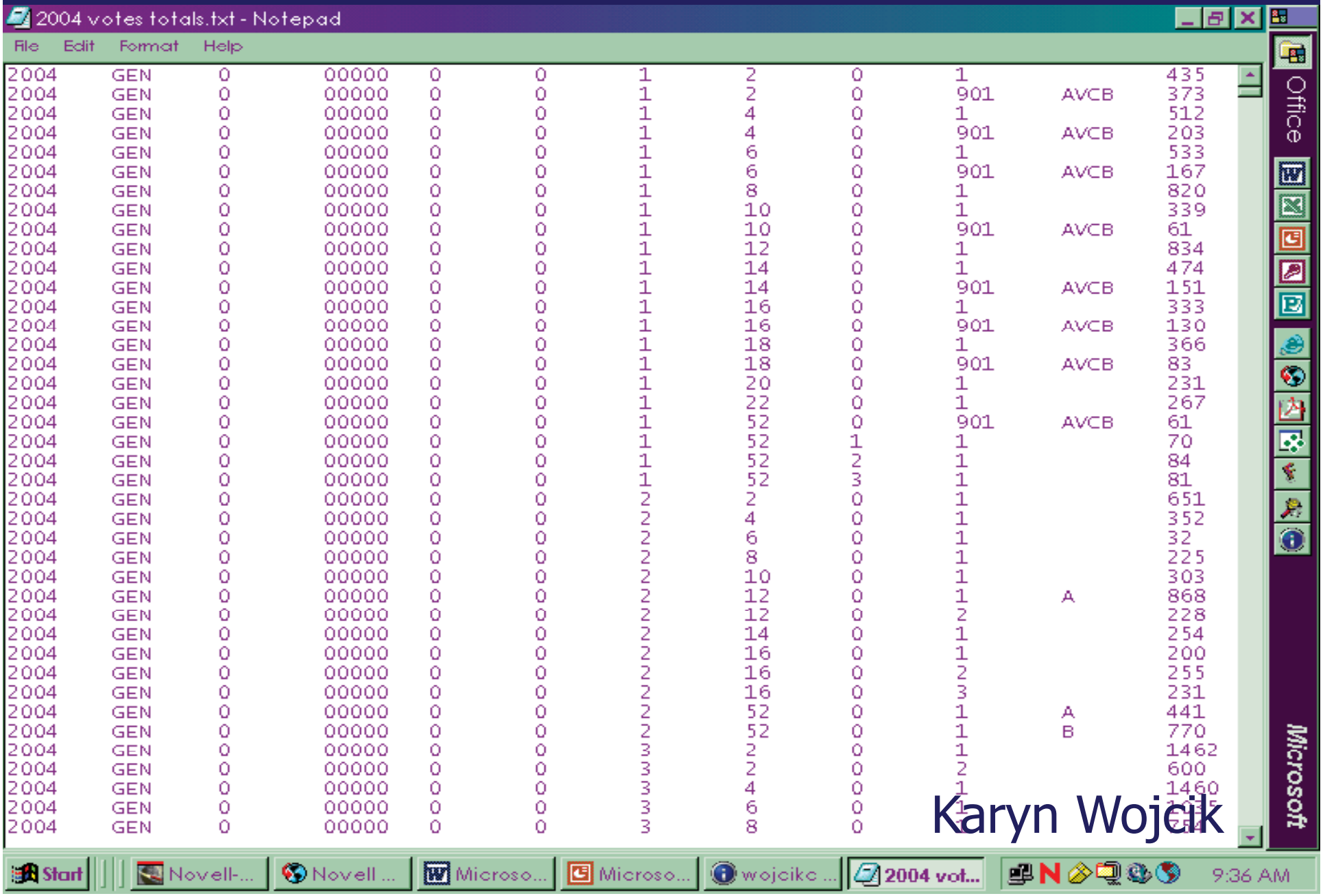

wojcikc @ michigan | orion.sdsc.edu (7955) - inQ 3.2.0 - (C) 2004 by SDSC
SRB Edit View Help
Michigan Precinct Results
Use Resource | Container
sfs-disk-pat
Michigan Precinct Results
wojcikc @ michigan: all
1972 general election.txt
1972 general election.txt
1972 general election.txt
1972 i SAS setup.sas
1972 i SAS setup.sas
1972 i SPSS setup.sps
1972 i SPSS setup.sps
1972 i codebook.pdf
1972 i codebook.pdf
1972 i data.txt
1972 i data.txt
1972 i description.pdf
1972 i description.pdf
1972 i manifest.txt
1972 i manifest.txt
1972 i osiris dictionary.od
1972 i osiris dictionary.od
1972 i related literature.tx
1972 i related literature.tx
1974 general election.txt
1974 general election.txt
1974 general election.txt
1974 i codebook.pdf
1974 i codebook.pdf
1974 i data.txt
Attribute
Value
Name Size Owner Timestamp Repl Resource
2002 county codes.txt 1078 wojcikc 2005-04-20-08.44.19 2 patMI-win
2002 elective office codes.txt 28292 wojcikc 2005-01-12-07.53.19 0 hpss-sdsc
2002 elective office codes.txt 28292 wojcikc 2005-03-28-17.00.27 1 sfs-disk-pat
2002 elective office codes.txt 28292 wojcikc 2005-04-20-08.44.30 2 patMI-win
2002 readme.txt 4899 wojcikc 2005-01-12-07.53.23 0 hpss-sdsc
2002 readme.txt 4899 wojcikc 2005-03-28-17.00.29 1 sfs-disk-pat
2002 readme.txt 4899 wojcikc 2005-04-20-08.44.36 2 patMI-win
2002 vote totals.txt 22127167 wojcikc 2005-01-12-07.53.05 0 hpss-sdsc
2002 vote totals.txt 22127167 wojcikc 2005-03-28-17.01.48 1 sfs-disk-pat
2002 vote totals.txt 22127167 wojcikc 2005-04-20-08.39.08 2 patMI-win
2004 candidate name codes.txt 16633 wojcikc 2005-06-16-11.05.53 0 patMI-win
2004 candidate name codes.txt 16633 wojcikc 2005-06-16-11.24.38 1 sfs-disk-pat
2004 city and township codes.txt 54965 wojcikc 2005-06-16-11.05.35 0 patMI-win
2004 city and township codes.txt 54965 wojcikc 2005-06-16-11.24.20 1 sfs-disk-pat
2004 county codes.txt 1078 wojcikc 2005-06-16-11.21.30 0 patMI-win
2004 county codes.txt 1078 wojcikc 2005-06-16-11.24.48 1 sfs-disk-pat
2004 elective office codes.txt 13045 wojcikc 2005-06-16-11.06.05 0 patMI-win
2004 elective office codes.txt 13045 wojcikc 2005-06-16-11.24.56 1 sfs-disk-pat
2004 readme.txt 4899 wojcikc 2005-06-16-11.21.38 0 patMI-win
2004 readme.txt 4899 wojcikc 2005-06-16-11.25.12 1 sfs-disk-pat
2004 votes totals.txt 15636522 wojcikc 2005-06-16-11.21.13 0 patMI-win
2004 votes totals.txt 15636522 wojcikc 2005-06-16-11.30.31 1 sfs-disk-pat
Datasets: 145 - Users: 1
Karyn Wojcik
Start
Novell-deli...
Novell Gro...
Microsoft ...
Microsoft P...
wojcikc @ ...
9:30 AM
Office
Microsoft

Step #2: Creating Usable Data

- Work backwards: start with the most recent data first
- SDSC mapped the data to the codes in the metadata files
- SDSC validated the data and recorded anomalies
- Data was cleaned of errors by reviewing the paper printouts
- SDSC copied the web interface used by the Bureau of Elections to display the data and to allow data queries.
- SDSC joined the data with geographic polygons supplied by CGI

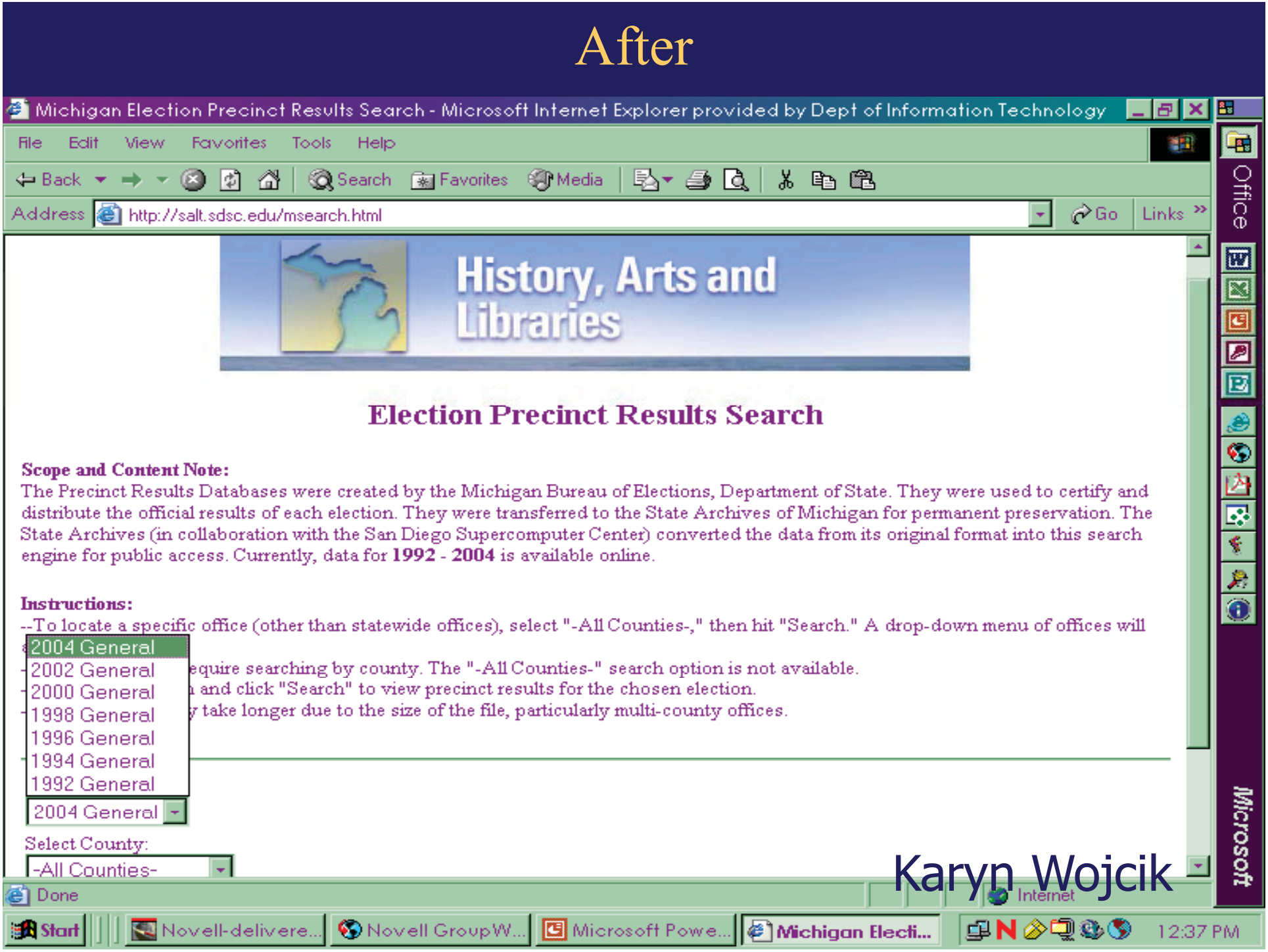
Karyn Wojcik

Before

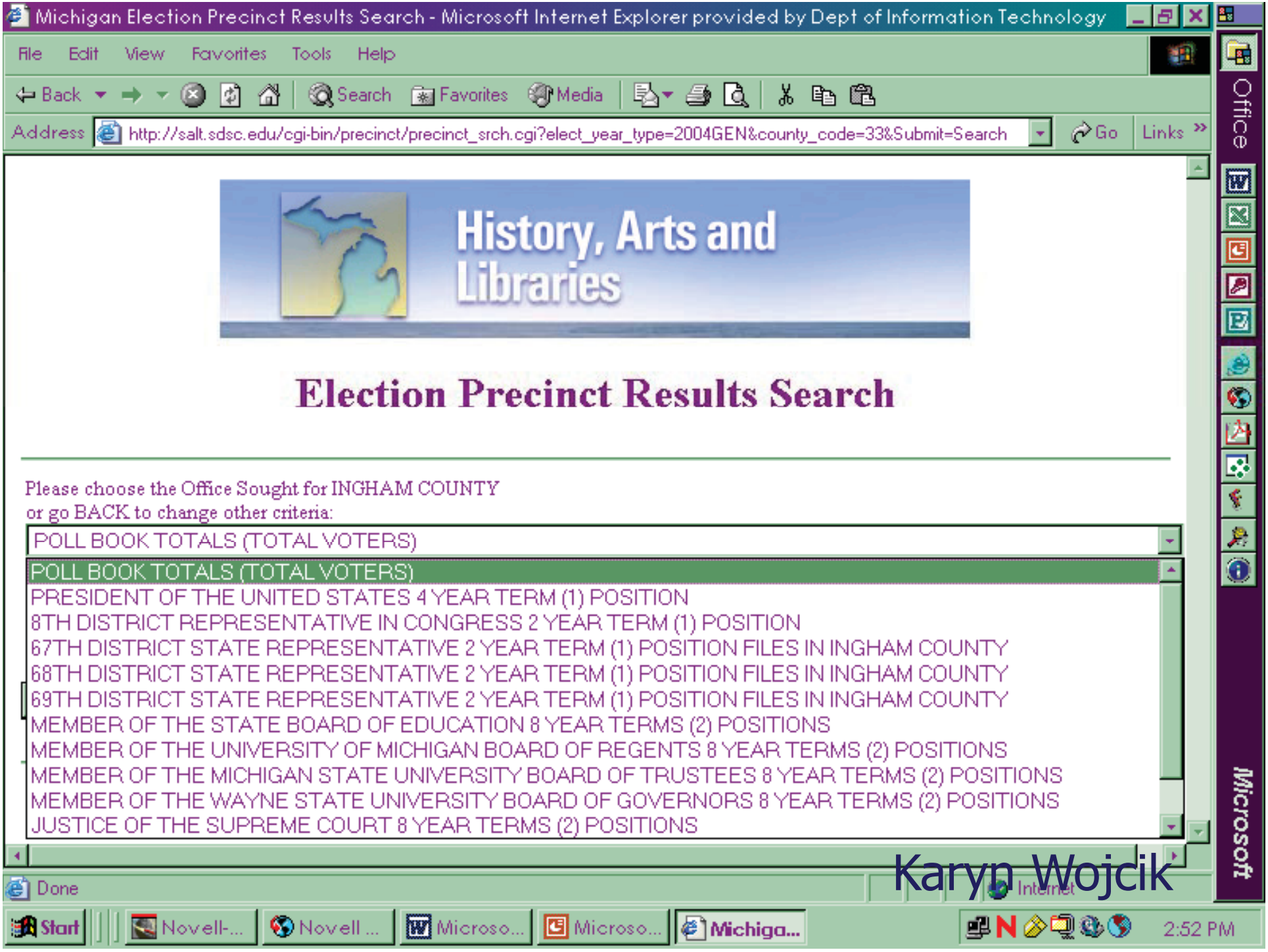


Karyn Wojcik

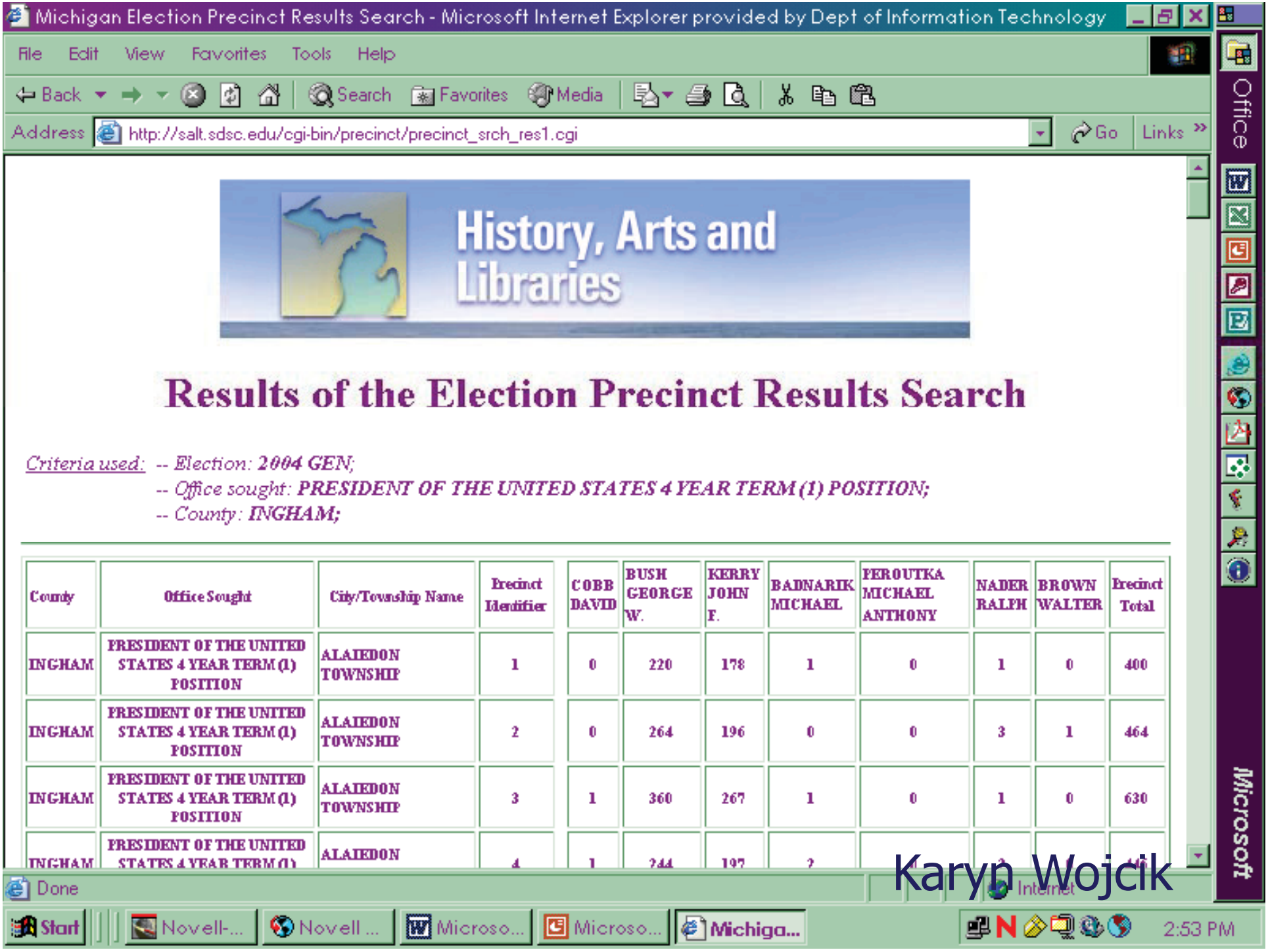
After



Karyn Wojcik



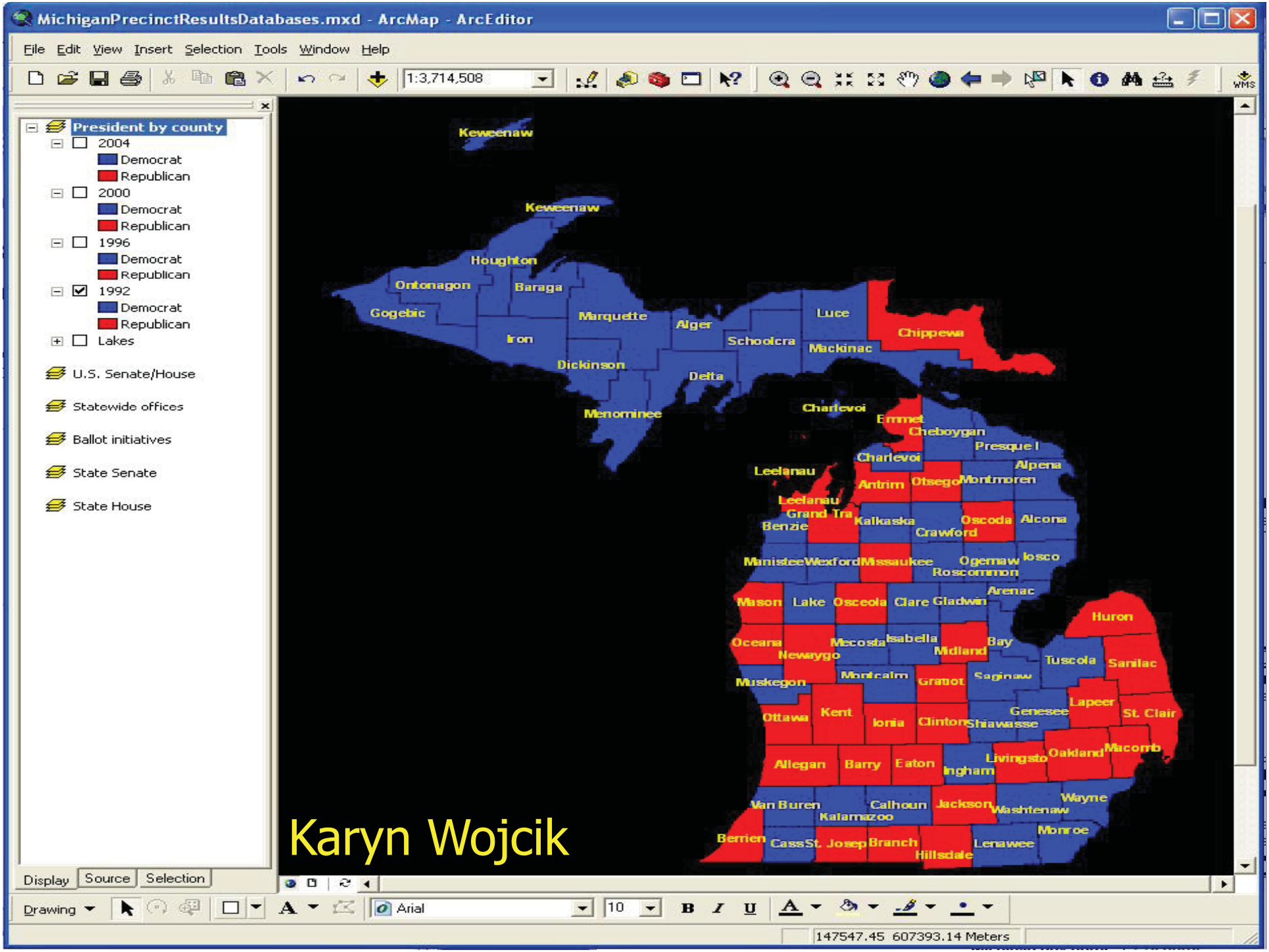
Karyn Wojcik

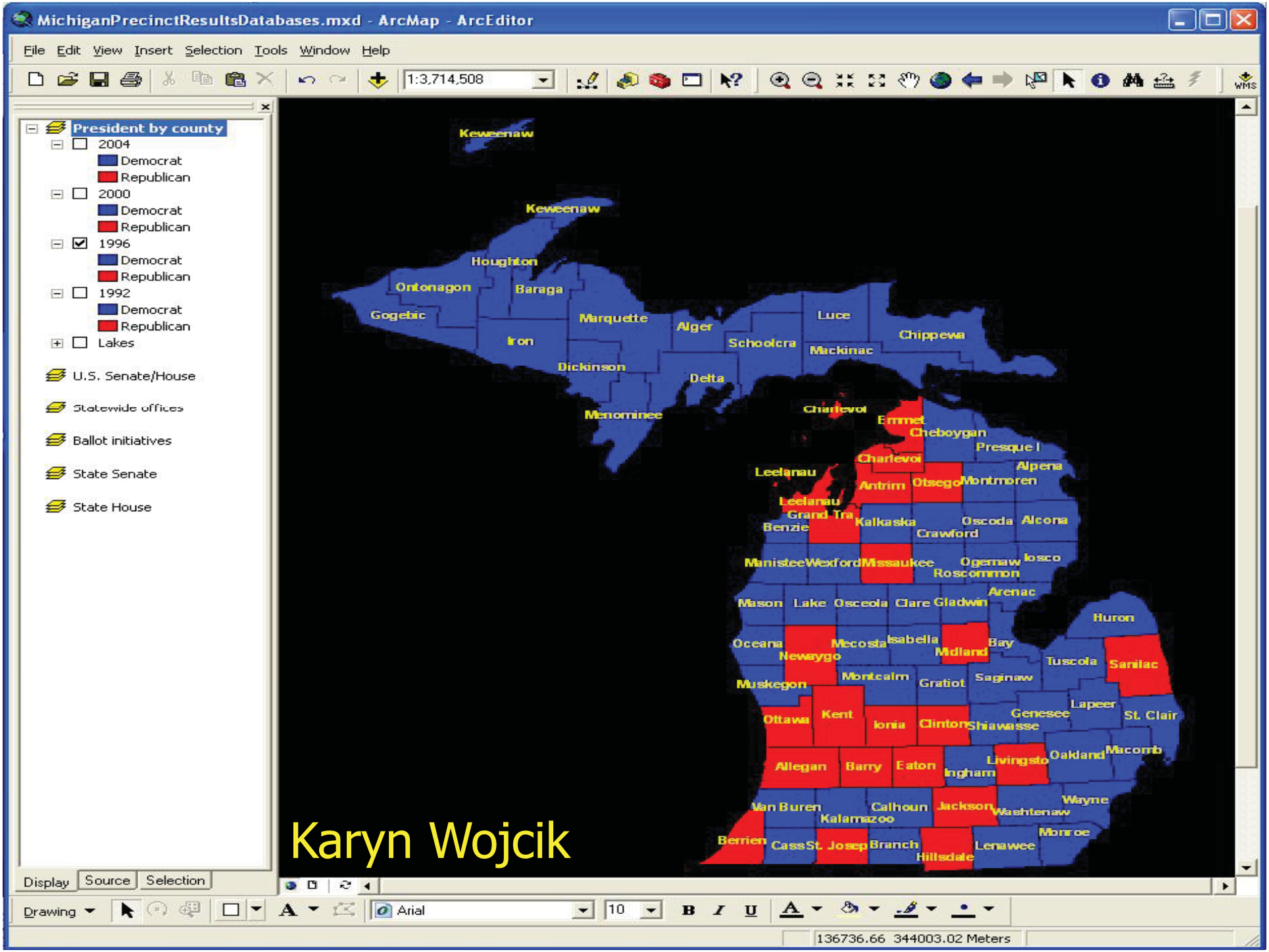

Michigan Election Precinct Results Search - Microsoft Internet Explorer provided by Dept of Information Technology
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://salt.sdsc.edu/cgi-bin/precinct/precinct_srch_res1.cgi
Go Links
History, Arts and Libraries
Results of the Election Precinct Results Search
Criteria used: -- Election: 2004 GEN;
-- Office sought: PRESIDENT OF THE UNITED STATES 4 YEAR TERM (1) POSITION;
-- County: INGHAM;
County | Office Sought | City/Township Name | Precinct Identifier | COBB DAVID | BUSH GEORGE W. | KERRY JOHN F. | BADNARIK MICHAEL | PEROUTKA MICHAEL ANTHONY | NADER RALPH | BROWN WALTER | Precinct Total
INGHAM | PRESIDENT OF THE UNITED STATES 4 YEAR TERM (1) POSITION | ALAIEDON TOWNSHIP | 1 | 0 | 220 | 178 | 1 | 0 | 1 | 0 | 400
INGHAM | PRESIDENT OF THE UNITED STATES 4 YEAR TERM (1) POSITION | ALAIEDON TOWNSHIP | 2 | 0 | 264 | 196 | 0 | 0 | 3 | 1 | 464
INGHAM | PRESIDENT OF THE UNITED STATES 4 YEAR TERM (1) POSITION | ALAIEDON TOWNSHIP | 3 | 1 | 360 | 267 | 1 | 0 | 1 | 0 | 630
INGHAM | PRESIDENT OF THE UNITED STATES 4 YEAR TERM (1) | ALAIEDON | 4 | 1 | 244 | 197 | 2
Done
Karyn Wojcik
Start Novell-... Novell ... Microso... Microso... Michiga... 2:53 PM

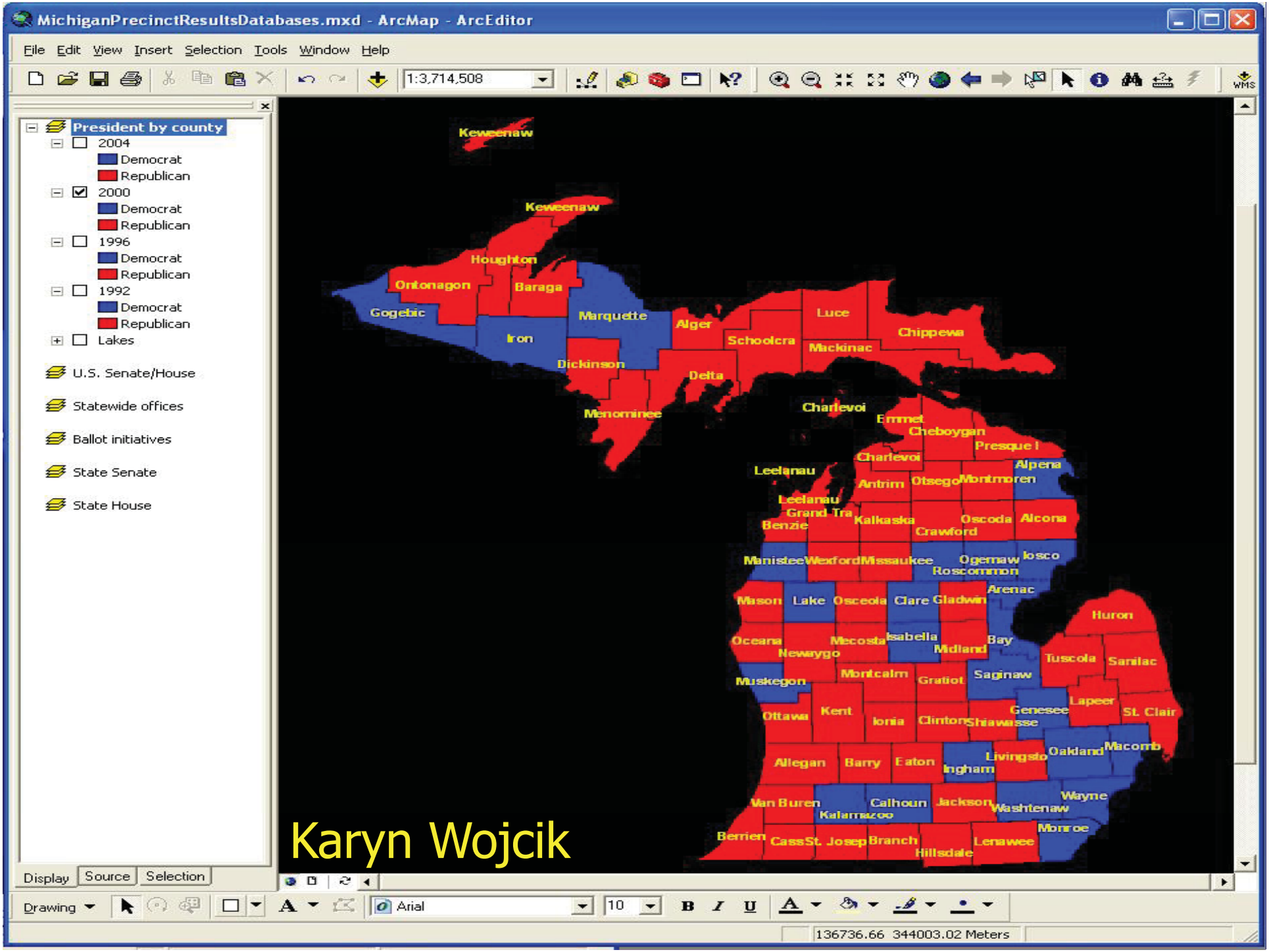
Step #3: GIS Interface

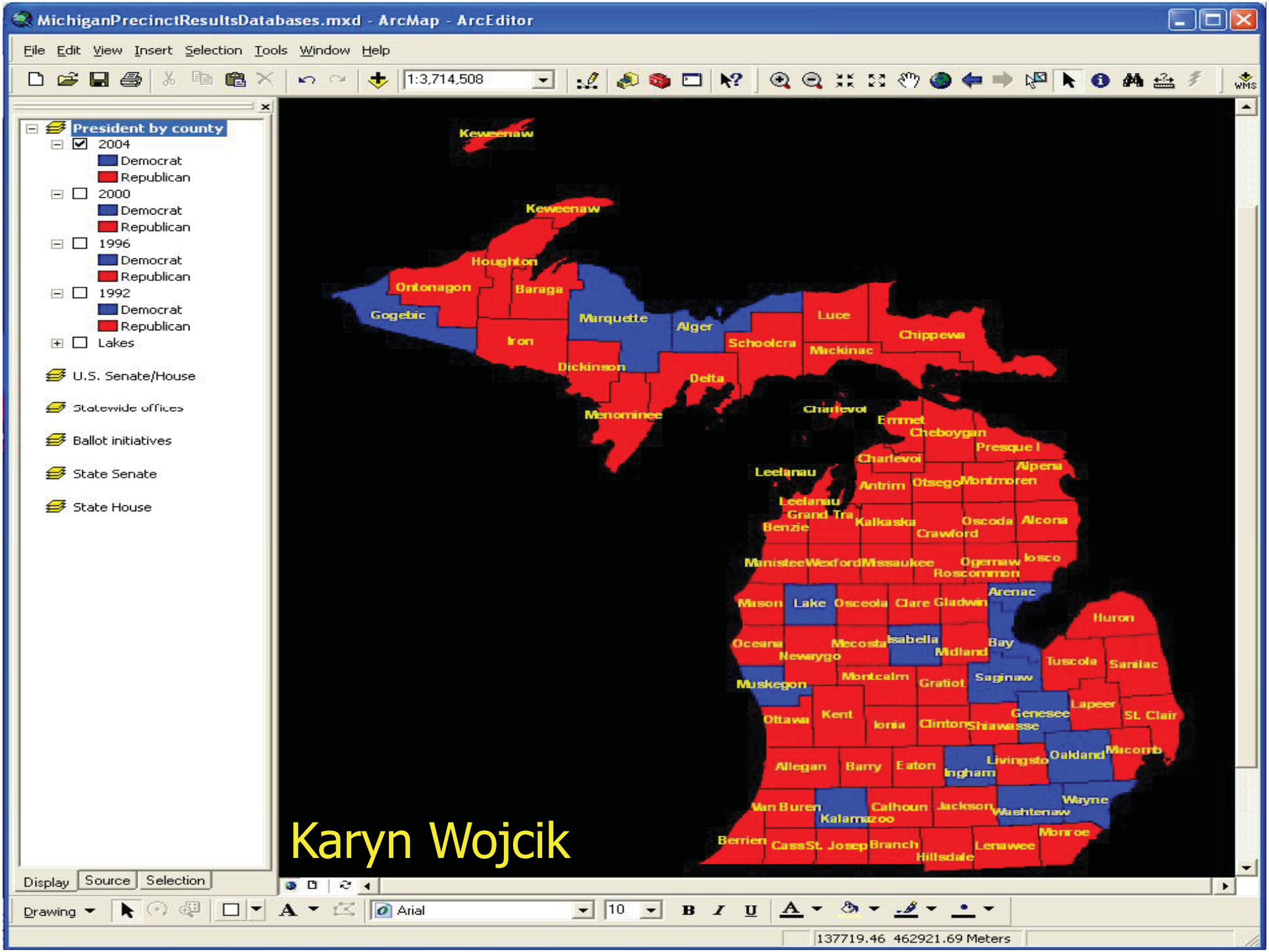
- Create statewide maps displaying the county level election results for each general election
- Identify blue and red counties for each race
- Identify voting trends

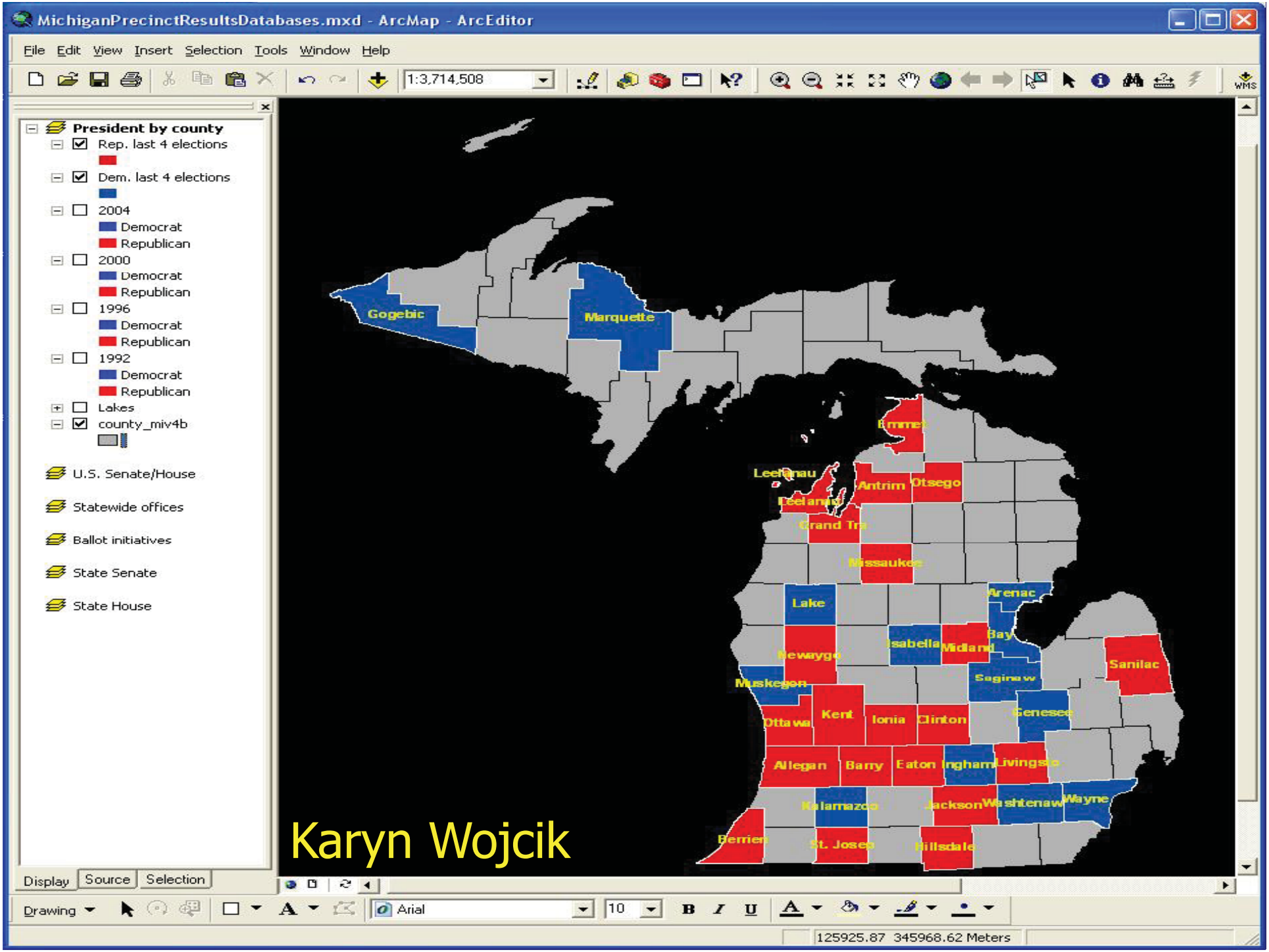
Karyn Wojcik


MichiganPrecinctResultsDatabases.mxd - ArcMap - ArcEditor
File Edit View Insert Selection Tools Window Help
1:3,714,508
President by county
2004
Democrat
Republican
2000
Democrat
Republican
1996
Democrat
Republican
1992
Democrat
Republican
Lakes
U.S. Senate/House
Statewide offices
Ballot initiatives
State Senate
State House
Display Source Selection
Drawing
Arial
10
147547.45 607393.14 Meters
Keweenaw
Houghton
Ontonagon
Baraga
Gogebic
Marquette
Alger
Luce
Chippewa
Iron
Schoolcra
Mackinac
Dickinson
Delta
Menominee
Charlevoi
Emmet
Cheboygan
Presque I
Leelanau
Antrim
Otsego
Montmoren
Alpena
Grand Tra
Kalkaska
Oscoda
Alcona
Benzie
Crawford
Manistee
Wexford
Missaukee
Ogemaw
Iosco
Roscommon
Arenac
Mason
Lake
Osceola
Clare
Gladwin
Huron
Oceana
Mecosta
Isabella
Midland
Bay
Newaygo
Tuscola
Sanilac
Muskegon
Montcalm
Gratiot
Saginaw
Lapeer
Ottawa
Kent
Ionia
Clinton
Genesee
Shiawasse
St. Clair
Allegan
Barry
Eaton
Ingham
Livingsto
Oakland
Macomb
Van Buren
Calhoun
Jackson
Kalamazoo
Washtenaw
Wayne
Monroe
Berrien
Cass
St. Josep
Branch
Hillsdale
Lenawee
Karyn Wojcik


MichiganPrecinctResultsDatabases.mxd - ArcMap - ArcEditor
File Edit View Insert Selection Tools Window Help
1:3,714,508
President by county
2004
Democrat
Republican
2000
Democrat
Republican
1996
Democrat
Republican
1992
Democrat
Republican
Lakes
U.S. Senate/House
Statewide offices
Ballot initiatives
State Senate
State House
Display Source Selection
Drawing
Arial
10
136736.66 344003.02 Meters
Keweenaw
Houghton
Ontonagon
Baraga
Gogebic
Marquette
Alger
Luce
Chippewa
Iron
Schoolcra
Mackinac
Dickinson
Delta
Menominee
Charlevoi
Emmet
Cheboygan
Presque I
Leelanau
Antrim
Otsego
Montmoren
Alpena
Grand Tra
Kalkaska
Oscoda
Alcona
Benzie
Crawford
Manistee
Wexford
Missaukee
Ogemaw
Iosco
Roscommon
Arenac
Mason
Lake
Osceola
Clare
Gladwin
Huron
Oceana
Mecosta
Isabella
Midland
Bay
Newaygo
Tuscola
Sanilac
Muskegon
Montcalm
Gratiot
Saginaw
Lapeer
Ottawa
Kent
Ionia
Clinton
Genesee
Shiawasse
St. Clair
Allegan
Barry
Eaton
Ingham
Livingsto
Oakland
Macomb
Van Buren
Calhoun
Jackson
Kalamazoo
Washtenaw
Wayne
Berrien
Cass
St. Josep
Branch
Hillsdale
Lenawee
Monroe
Karyn Wojcik

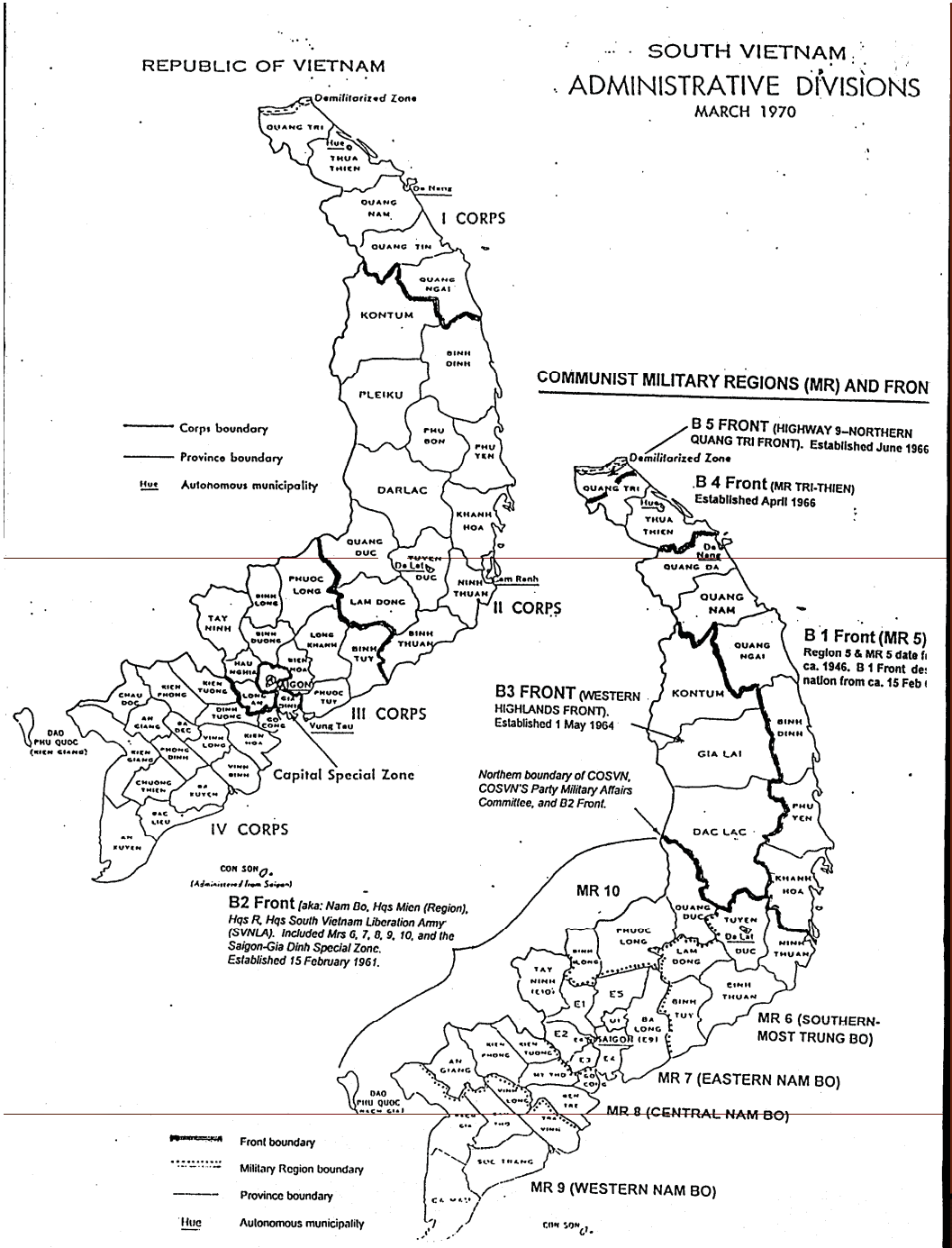

MichiganPrecinctResultsDatabases.mxd - ArcMap - ArcEditor
File Edit View Insert Selection Tools Window Help
1:3,714,508
President by county
2004
Democrat
Republican
2000
Democrat
Republican
1996
Democrat
Republican
1992
Democrat
Republican
Lakes
U.S. Senate/House
Statewide offices
Ballot initiatives
State Senate
State House
Display Source Selection
Drawing
Arial
10
136736.66 344003.02 Meters
Keweenaw
Houghton
Ontonagon
Baraga
Gogebic
Iron
Marquette
Alger
Schoolcra
Luce
Chippewa
Mackinac
Dickinson
Delta
Menominee
Charlevoi
Emmet
Cheboygan
Presque I
Alpena
Leelanau
Antrim
Otsego
Montmoren
Grand Tra
Kalkaska
Oscoda
Alcona
Benzie
Crawford
Manistee
Wexford
Missaukee
Ogemaw
Iosco
Roscommon
Arenac
Mason
Lake
Osceola
Clare
Gladwin
Huron
Oceana
Mecosta
Isabella
Midland
Bay
Newaygo
Tuscola
Sanilac
Muskegon
Montcalm
Gratiot
Saginaw
Ottawa
Kent
Ionia
Clinton
Shiawasse
Genesee
Lapeer
St. Clair
Allegan
Barry
Eaton
Ingham
Livingsto
Oakland
Macomb
Van Buren
Calhoun
Jackson
Kalamazoo
Washtenaw
Wayne
Monroe
Berrien
Cass
St. Josep
Branch
Hillsdale
Lenawee
Karyn Wojcik


MichiganPrecinctResultsDatabases.mxd - ArcMap - ArcEditor
File Edit View Insert Selection Tools Window Help
1:3,714,508
President by county
2004
Democrat
Republican
2000
Democrat
Republican
1996
Democrat
Republican
1992
Democrat
Republican
Lakes
U.S. Senate/House
Statewide offices
Ballot initiatives
State Senate
State House
Display Source Selection
Drawing
Arial
10
137719.46 462921.69 Meters
Keweenaw
Houghton
Ontonagon
Baraga
Gogebic
Marquette
Alger
Luce
Chippewa
Schoolcra
Mackinac
Iron
Dickinson
Delta
Menominee
Charlevoi
Emmet
Cheboygan
Presque I
Leelanau
Antrim
Otsego
Montmoren
Alpena
Grand Tra
Benzie
Kalkaska
Crawford
Oscoda
Alcona
Manistee
Wexford
Missaukee
Ogemaw
Roscommon
Iosco
Mason
Lake
Osceola
Clare
Gladwin
Arenac
Huron
Oceana
Mecosta
Isabella
Midland
Bay
Newaygo
Tuscola
Sanilac
Muskegon
Montcalm
Gratiot
Saginaw
Ottawa
Kent
Ionia
Clinton
Shiawasse
Genesee
Lapeer
St. Clair
Allegan
Barry
Eaton
Ingham
Livingsto
Oakland
Macomb
Van Buren
Kalamazoo
Calhoun
Jackson
Washtenaw
Wayne
Monroe
Berrien
Cass
St. Josep
Branch
Hillsdale
Lenawee
Karyn Wojcik


MichiganPrecinctResultsDatabases.mxd - ArcMap - ArcEditor
File Edit View Insert Selection Tools Window Help
1:3,714,508
President by county
Rep. last 4 elections
Dem. last 4 elections
2004
Democrat
Republican
2000
Democrat
Republican
1996
Democrat
Republican
1992
Democrat
Republican
Lakes
county_miv4b
U.S. Senate/House
Statewide offices
Ballot initiatives
State Senate
State House
Display Source Selection
Drawing
Arial
10
125925.87 345968.62 Meters
Gogebic
Marquette
Emmet
Antrim
Otsego
Grand Tra
Missaukee
Lake
Arenac
Bay
Isabella
Midland
Newaygo
Sanilac
Muskegon
Saginaw
Ottawa
Kent
Ionia
Clinton
Genesee
Allegan
Barry
Eaton
Ingham
Livingsto
Kalamazoo
Jackson
Washtenaw
Wayne
Berrien
St. Joseph
Hillsdale
Karyn Wojcik

Summary of GIS Creation Steps

- Accession - data recovery from obsolete media
- Data encoding format migration - creating usable data
  - Digitize documents
    - Geo-reference documents
  - Digitize maps
    - Convert to Shape files
    - Convert to Scalable Vector Graphics encoding (SVG)
    - Convert to Graphics Markup Language encoding (GML)
  - Import into a generic GIS environment
    - XML-GML GIS viewer
    - Link layers
    - Map descriptive metadata to layers
    - Map documents to layers



NARA Herbicides Collection

DOD records of the locations where agents and pamphlets were dropped as a function of time (airplane flight paths)

Original Military Record (from 40 years ago) Data on EBCDIC Tapes

```
6507213207565     260404040 040000{0000D0000000{048{ {0000000{0000000{0000000{0000000{
6507243207565     260606060 060000{0000D0000000{072{ {0000000{0000000{0000000{0000000{
6507253207565     260606060 060000{0000D0000000{072{ {0000000{0000000{0000000{0000000{
6507263207565     260606060 060000{0000D0000000{072{ {0000000{0000000{0000000{0000000{
6507273207565     260606060 060000{0000D0000000{072{ {0000000{0000000{0000000{0000000{
6507283207565     260505050 050000{0000D0000000{060{ {0000000{0000000{0000000{0000000{
6507293207565     260404040 040000{0000D0000000{048{ {0000000{0000000{0000000{0000000{
6508022022365     060202020 010000{0000C0000000{012{ {0000000{0000000{0000000{0000000{1A         AS890255
     000{000{
6508022022365                                                                        1B         AS940140
     000{000{
6508042022365     060202020 006000{0000C0000000{007B {0000000{0000000{0000000{0000000{1A         AS925205
     000{000{
6508042022365                                                                        1B         AS970065
     000{000{
6508062022365     060202020 004000{0000C0000000{004H {0000000{0000000{0000000{0000000{1A         BS290320
     000{000{
6508062022365                                                                        1B         BS275298
     000{000{
6508073207565     260202020 020000{0000D0000000{024{ {0000000{0000000{0000000{0000000{1A         YT080110
     000{000{
6508073207565                                                                        1B         YT110060
     000{000{
6508113207565     260202020 020000{0000D0000000{024{ {0000000{0000000{0000000{0000000{
6508123207565     260202020 020000{0000D0000000{024{ {0000000{0000000{0000000{0000000{
6508151022465     020202020 008000{0000C0000000{009F {0000000{0000000{0000000{0000000{1A         YD350155
     000{000{
6508151022465                                                                        1B         YD450150
```

Steps Applied to Herbicide Collection

- Transformative migration of binary data file
  - Identify legacy format - NIPS EBCIDC
  - Read data file and convert to ASCI
  - Interpret paper record description of fields in binary file
  - Apply XML labels to identified fields
  - Convert coordinate system from 4 military geoids to UTM
- At this point have the individual airplane tracks for each flight mission
  - Organize missions into GIS layers by date or agent type
- Visualize using XML-SVG browser
  - Enabled interactive manipulation
  - Used to create resource for political science course

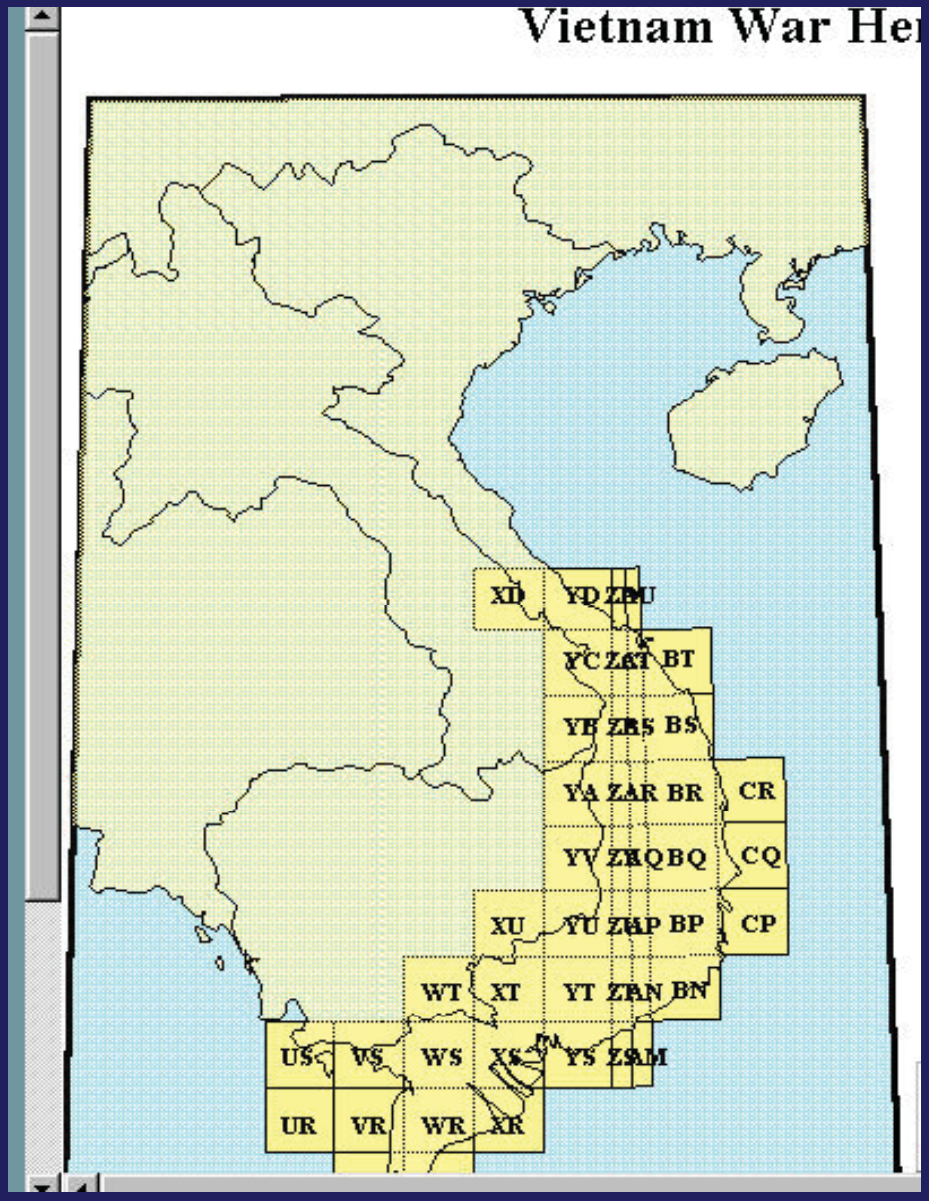
Herbicides Collection - 2

Converted to XML:

```
<YEAR><yearnum>66</yearnum>
<MONTH><monthnum>01</monthnum>
<DATE><datenum>01</datenum>
<MISSION><num>206866</num>
        <RUN><code>A</code>
        <ctz>3</ctz><multi></multi><prov>27</prov>
                <aircrafts>
                        <scheduled>02</scheduled><airborne>02</airborne><productive>02</productive>
                </aircrafts>
                <agent>O</agent><gal>02000</gal><hits>0</hits>
                <aborts>
                <maintenance>0</maintenance><weather>0</weather><battle_damage>0</battle_damage><other>0</other>
                </aborts>
                <type>D</type><area>024</area><rsult></rsult>
                <UTM>
                        <utmid>1A</utmid>
                        <utm_coor>YS240780</utm_coor>
                </UTM>
                <UTM>
                        <utmid>1B</utmid>
                        <utm_coor>YS290630</utm_coor>
                </UTM></RUN>
        <RUN><code>B</code>
        <ctz>3</ctz><multi></multi><prov>27</prov>
                <aircrafts>
                        <scheduled>02</scheduled><airborne>02</airborne><productive>02</productive>
                </aircrafts>
                <agent>O</agent><gal>02000</gal><hits>0A</hits>
                <aborts>
                <maintenance>0</maintenance><weather>0</weather><battle_damage>0</battle_damage><other>0</other>
                </aborts>
                <type>D</type><area>024</area><rsult></rsult>
```

Herbicides

The Complexity of Context



- Defining context
  - Not always in metadata
  - How was the data captured?
  - What are the constraints on the data?
  - How accurate is the data?
  - How to transform the original data into a conceptually realizable form?
- What standards should be used to define GIS context?

InterPARES VanMap Project

- Preserve an existing GIS environment
  - Manage accession of the GIS information and data in a dynamically changing environment

- Implications
- Identify relationships between the GIS environment and the information resources on which legislative records were based
- Preserve mapping between records within a preservation environment

VanMap

- VanMap: federates information from multiple Vancouver city departments, displays interactive maps (Oracle Spatial, Autodesk map viewer)
- Requirements:
  - Non-proprietary system for managing GIS data
  - Ability to import GIS data from a proprietary system into the preservation environment
  - Ability to query and display the GIS data in a similar fashion
  - Ability to archive web pages, “preserve look and feel”
  - Appraisal mechanism (“information value”?)
- Challenges:
  - Different update frequencies
  - Interactive browsing and mapping archived data
  - Preservation Model: snapshots, recording changes, a hybrid
  - Retaining relationships between spatial and related report data
  - What to do with hyperlinks to web pages

VanMap Project

1) Preserve content from a GIS environment.
   - Extract data for each GIS component
   - Data layers come from different storage resources
   - Demonstrate can assemble components into a viable GIS system
2) Preserve information about changes to the GIS environment.
   - Test whether a database can track changes to the GIS environment
   - From database, re-create GIS configuration file for requested time
   - Demonstrate application in a GIS system
3) Preserve snapshots of the VanMap system.
   - Demonstrate migration of GIS system into alternate technology

GIS Preservation Summary

- GIS system is a compound document
- The archivists control the preservation process and the identification of:
  - Preservation graphics formats
  - Components of compound record representing GIS environment
  - Relationships between components
  - Accession of components into the preservation environment
- View as migration process
  - Document the transformation / migration
  - Reassembly of disassociated GIS components is same as assembly of GIS environment from legacy data
  - Have to characterize the relationships between the layers in terms of the GIS operations

Appraisal and Accessioning of Electronic Records Produced with GIS (Mark Conrad, Richard Marciano)

- Potential research issues:
  - Can we develop a framework so that archivists can identify the essential characteristics of a record that must be retained after a transformative migration?
  - Can we develop a framework for validating that those essential characteristics have been retained after the transformative migration has been executed?
  - Some of the variables involved in the transformation include:
    - What software was used to create the records in the first place.
    - What software is used to transform the records from one format to another
    - What settings of the software are used for reading in the records in their native formats?
    - What settings of the software are used for writing out the records in a standards-based format?

Preservation of Types of GIS queries

- When did Feature X exist or cease to exist?
- What existed at Location A at Time T?
- What happened to a given feature or location between Time T1 and T2?
- Did Event A exist before or after Condition X (or Event B)?
- What patterns exist between Events A-B-C and Features X-Y-Z?
- Given data for Feature Y at Time T1 & T3, what was the likely state of this Feature at Time T2?
- What will be the likely state of Feature X at Time T?
- What is the predicted outcome following Event A after Time T?

For More Information

Richard Marciano
<marciano@sdsc.edu>
Ilya Zaslavsky
zaslavsk@sdsc.edu
Reagan W. Moore
<moore@sdsc.edu>