Creating an online library of map and geospatial data: Challenges and Opportunities

Digital Library Federation Spring 2005

T. Wangyal Shawa

GIS Librarian

Princeton University Library shawatw@princeton.edu

April 14, 2005

Welcome to Princeton University

When the project was started

- The project was formally started on February 20, 2004 as a pilot map scanning project.
- The goals of the pilot project were to:
 - a) Virtually integrate Princeton University library's analog and digital maps.
 - b) Provide a flexible intuitive user interface that will enable users to search, browse, view, or download maps.
 - c) Design a system architecture and standards for digitizing maps, metadata, and compressing images.
 - d) Establish workflows.

Challenges of creating an online library of map and geospatial data

- Design a system to integrate various forms of geographic information and make them accessible online.
- Select software packages that will integrate complex systems.
- Design an interface that will not only allow you to search but also to browse, and download.
- Organize data if browsing option is allowed.
- Assign file name once map is scanned.

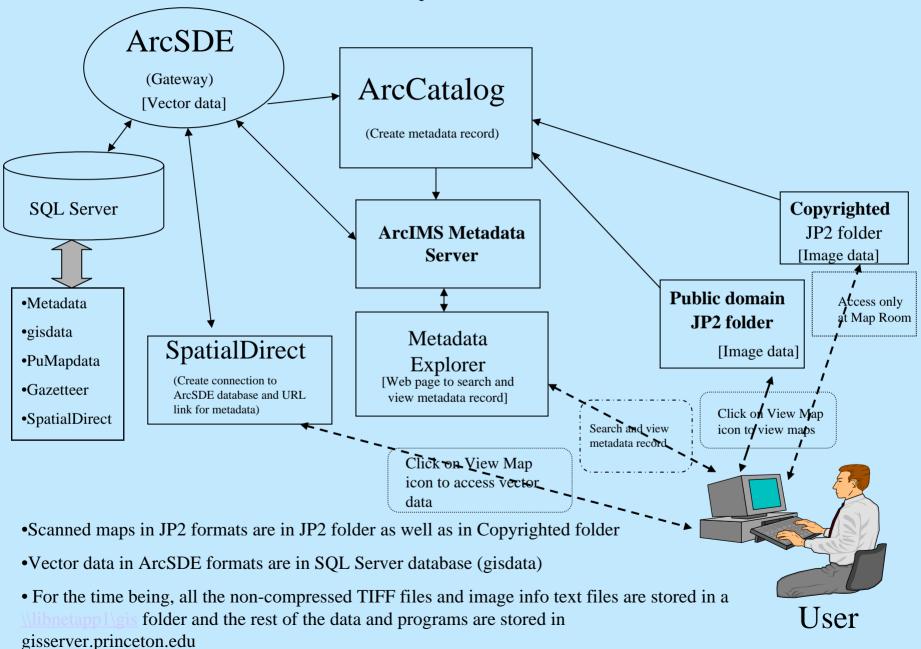
Challenges of creating an online library of map and geospatial data

- Design workflows that are easy to maintain and implement.
- Convince administrators to give necessary support to start the project.
- Determine type of server and disk space needed.
- Select and purchase scanner and computer for scanning maps.

Software packages for system design

- ESRI software packages (ArcCatalog, ArcIMS, and ArcSDE)
- Mapping Science's GeoJP2 Image Server, and Encoder and Decoder
- Microsoft's SQL Server database
- Safe Company's FME and SpatialDirect

Present System Architecture



Standards

• Scanning Standard: All the maps were scanned at 400 dpi with 256 colors and saved in a non-compressed TIFF file format. The reason for making this decision was based on in-house tests.

Scanning of a paper map (USGS 1:24,000 topographic map) at 400 dpi with 256 colors versus 500 dpi with 24-bit color shows very little difference. A minor quality improvement hardly justifies the larger file sizes (500 dpi with 24-bit color: file size 441MB, 400 dpi with 24-bit color: file size 278MB, 400 dpi with 256 color: file size 96.2MB) and extra time it takes in scanning and saving it on a computer.

The objective of our map scanning is to preserve map information and therefore, it is not important to capture all the subtle color differences or color "noise" generated by the condition of the paper and the printer.

Standards

- Compression Standards: The TIFF files are later compressed using Mapping Science's GeoJP2 software into JPEG2000 file with 10:1 compression ratio.
- Metadata Standards: All the scanned maps were individually cataloged using ESRI style ISO XML format. The ISO XML format was based on the DTD included with the draft ISO 19115 specifications.

Maps scanned

- 1951 Mercer county aerial photographs (112) and New Jersey Geological Survey Atlas Sheets (14)
- New Jersey USGS 1:62500 topographic maps (102)
- NJGS Topographic Name Sheets (45)
- USGS NJ 7.5 Minute Topographic series (161)
- A.M.S L902 series maps covering Japan (123)
- G.S.G.S 4416 series maps of Middle Danube 1:100,000 (70)
- G.S.G.S 4275 series maps covering Tunisia and Algeria (17)
- A.M.S L951 maps covering Korea (18)
- Russian Army's City maps of China (51)
- A.M.S. S901 series maps covering Philippine cities (47)
- NIMA's Iraq maps (5)
- NIMA's Honduras (4)
- England 1:10 000 published by German 1942 (118)

Workflows

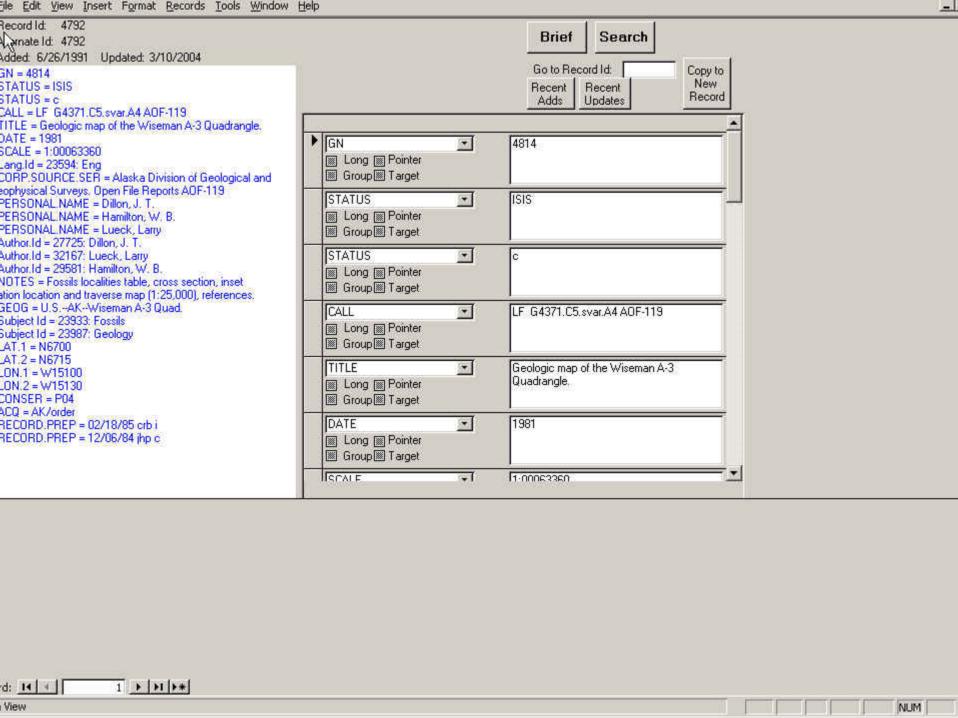
- I developed workflows for scanning maps and for disseminating vector data.
- Scanning maps: one workflow deals with regular scanning and another with scanning of maps brought in or requested by our patrons to be scanned for their class or research.
- At this stage, we are scanning only those maps that were cataloged in GeoMap database (a local map cataloging database).

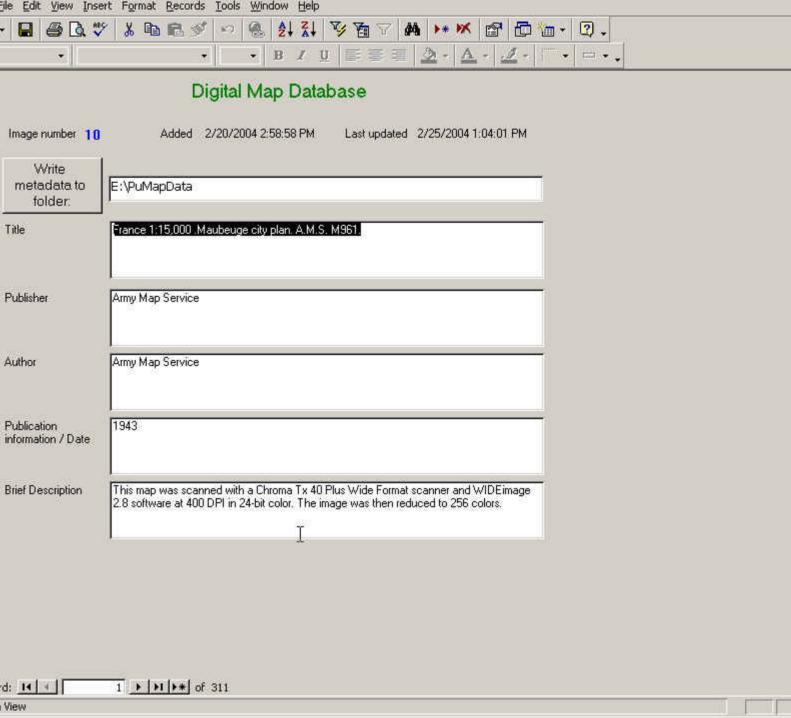
Regular scanning workflow

- a) First, find the GN number in the GeoMap database
- b) Open the PuMapdata database
- c) Enter basic information into the PUMapdata
- d) Create a text file for each entry

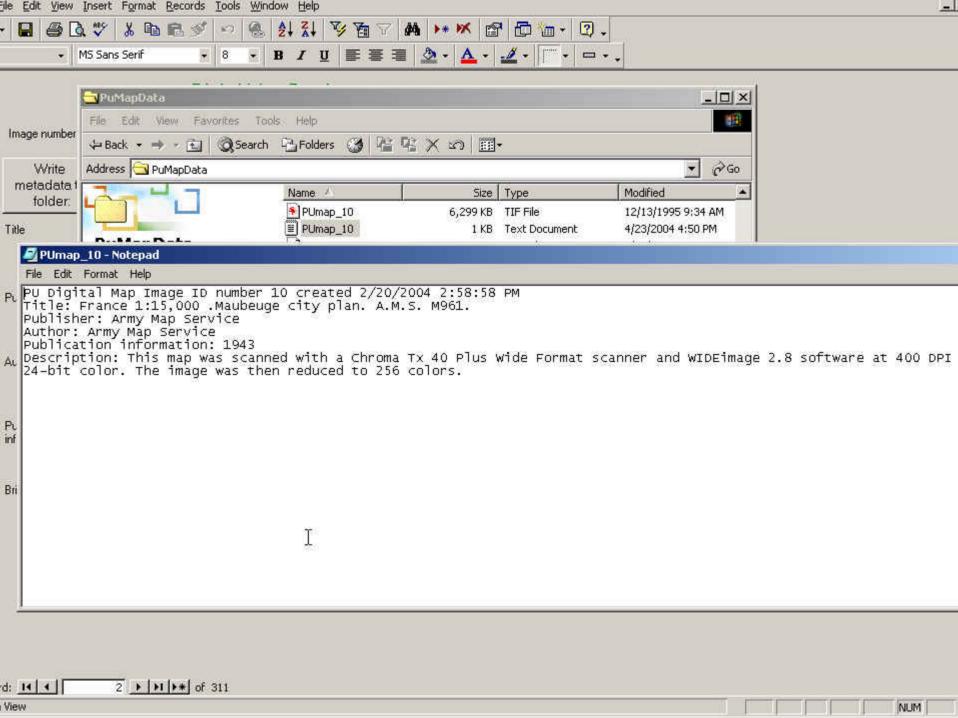
Regular scanning workflow

- e) Open WIDEimage software and scan the map
- f) Save the image in the same folder as the one in which you saved the database text file. The filename should be the same as the corresponding .txt file. For example, with E:\PuMap10.txt, your image will be "E:\PuMap10.tif"
- g) Encode all the TIFF images overnight into JP2 files
- h) Copy JP2 and TIFF images to server, and delete images from local computer
- i) Share JP2 files for creating a metadata
- j) Open ArcCatalog and select a JP2 image and create metadata
- k) Search JP2 image record in the GeoMap database by GN number
- 1) Enter necessary information in Metadata Editor from the GeoMap
- m) Publish metadata in ArcIMS Metadata Server. Then it is ready for search in the Metadata Explorer
- n) Create Excel spreadsheet to keep a record of images that have metadata records.





NUM





=

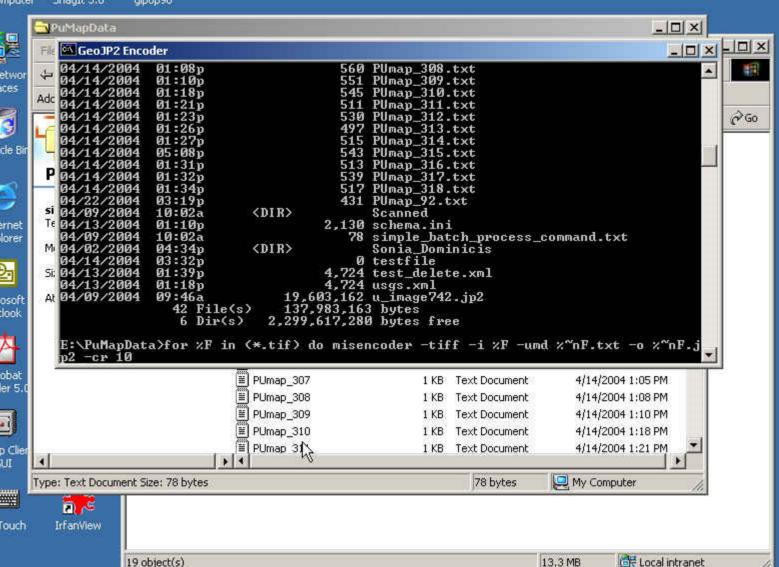
ices

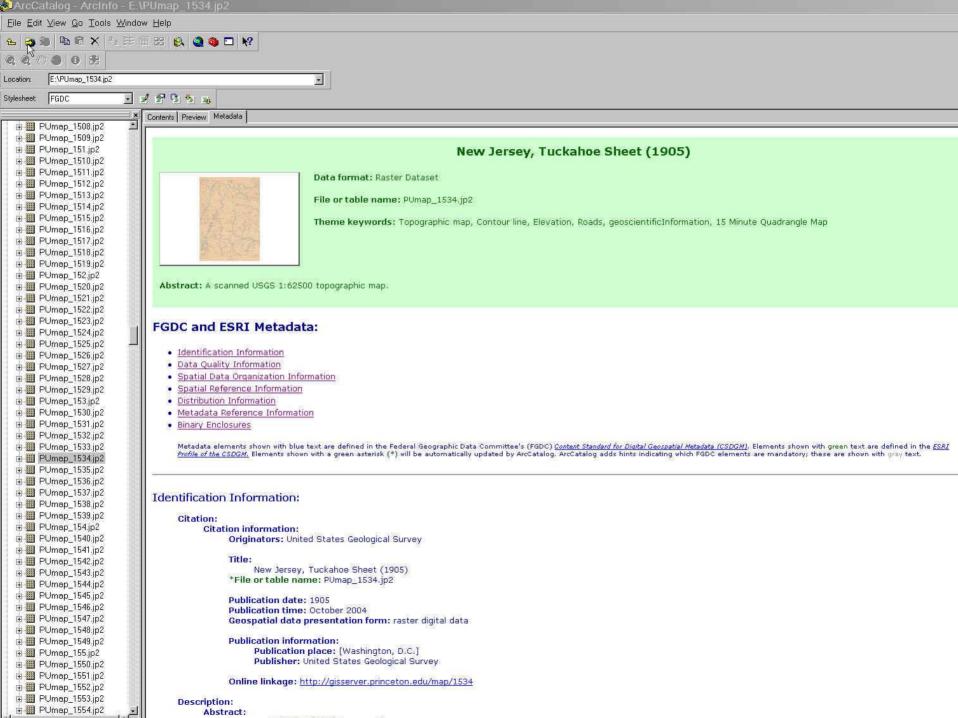
5

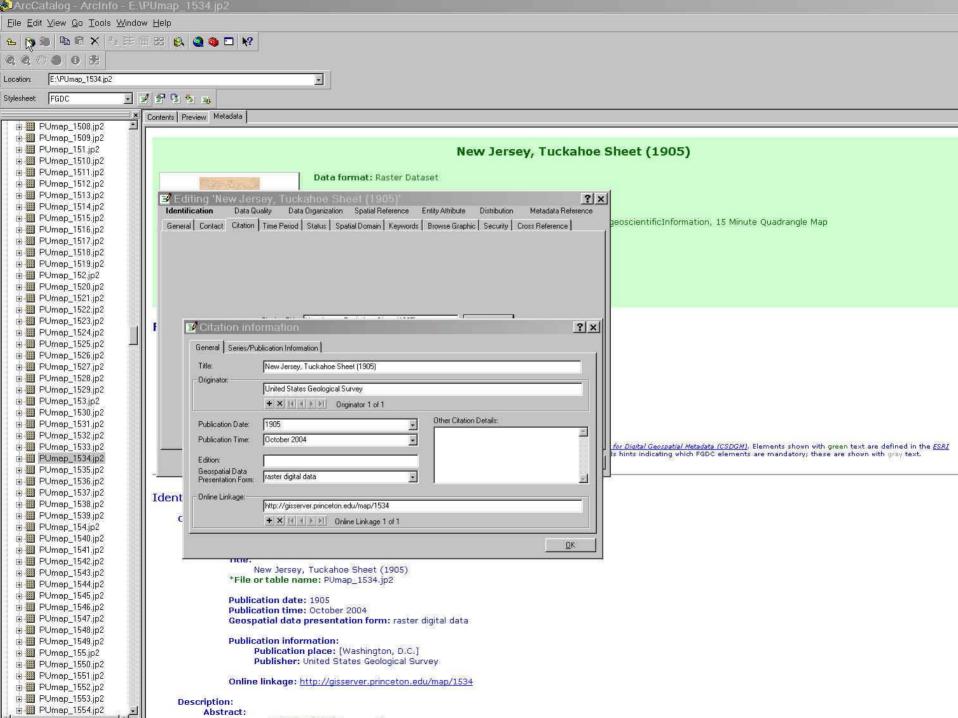
look

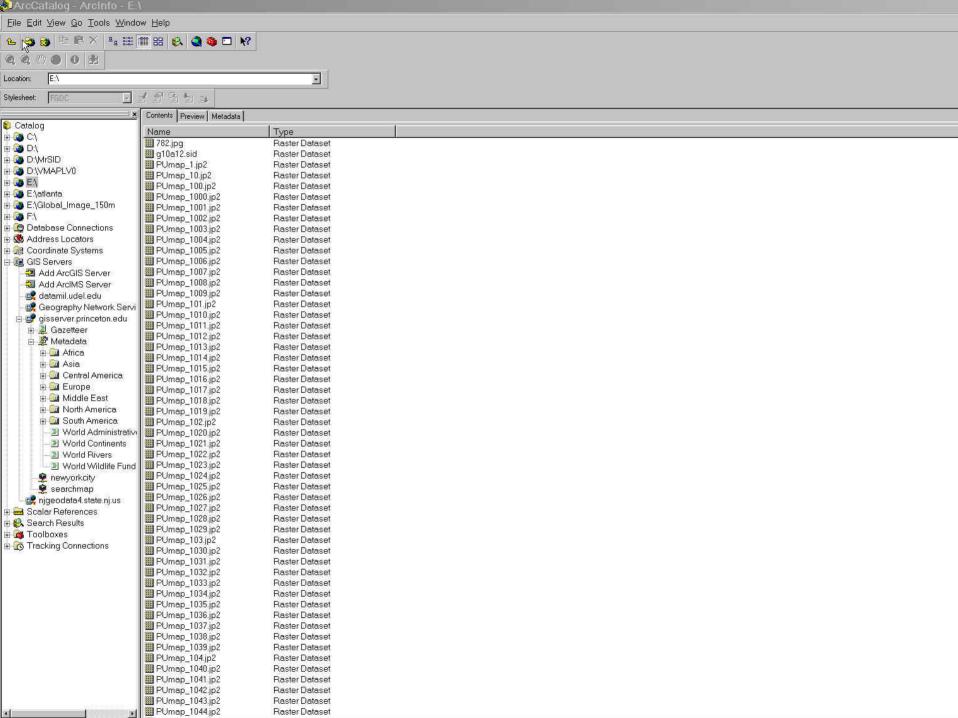
obat

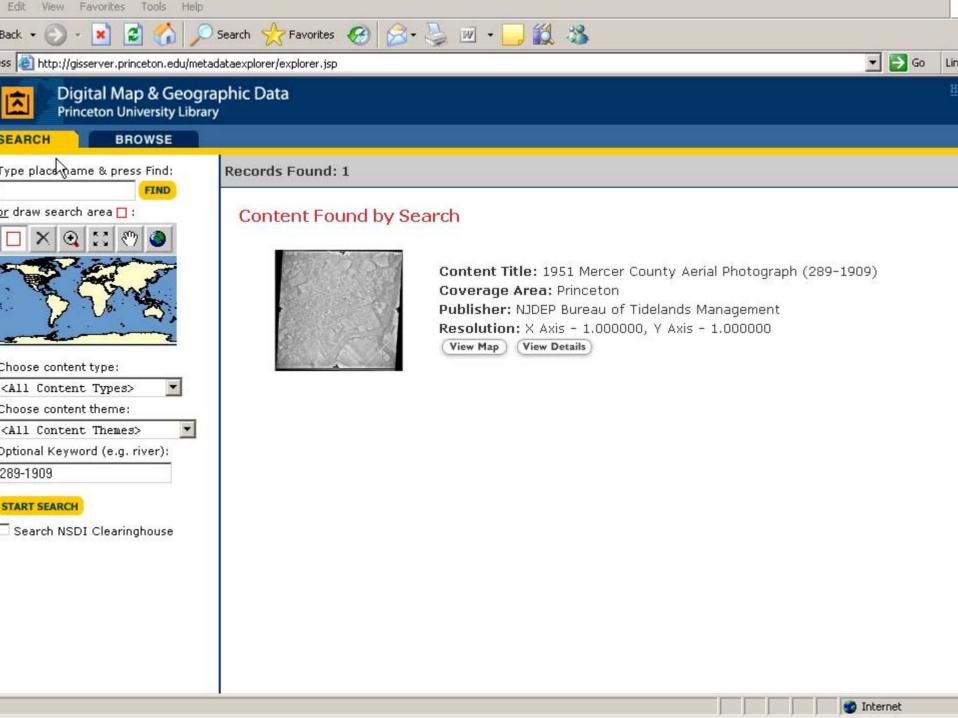
UI

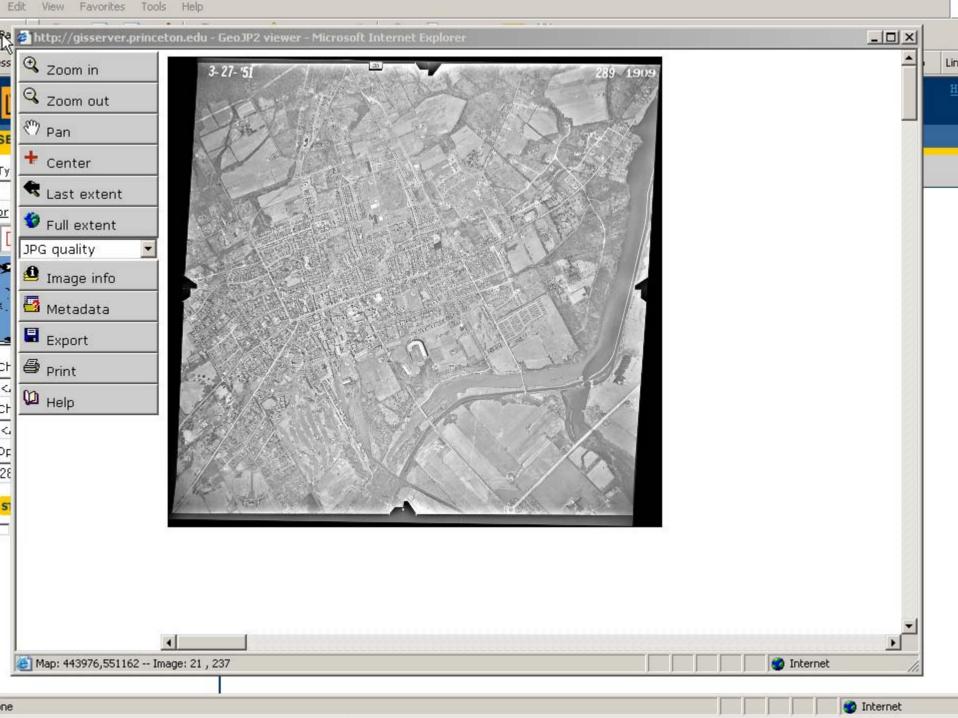


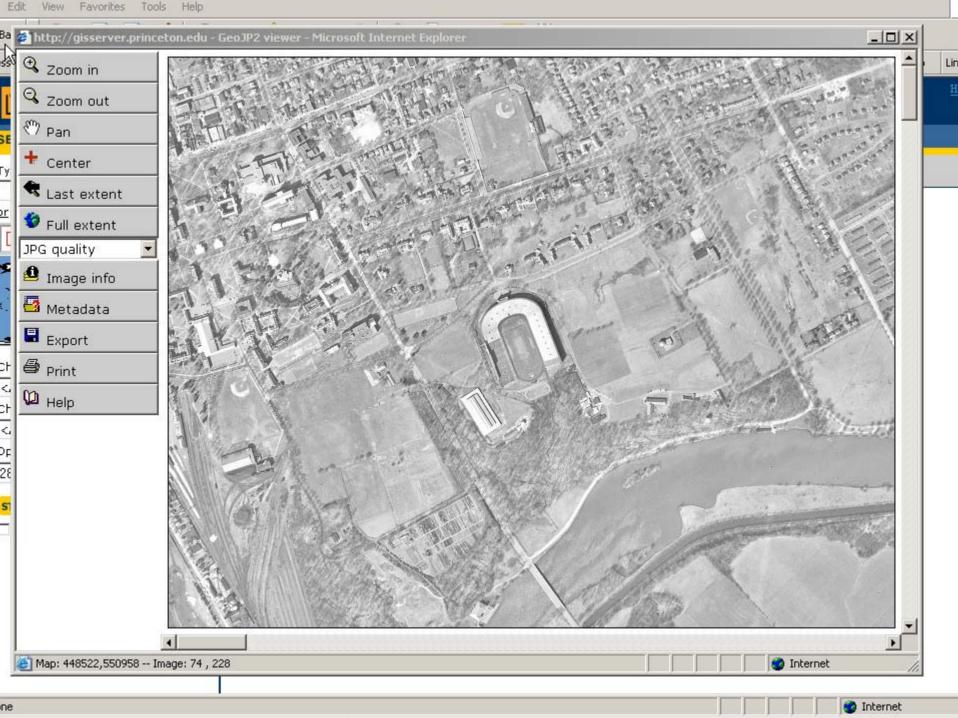


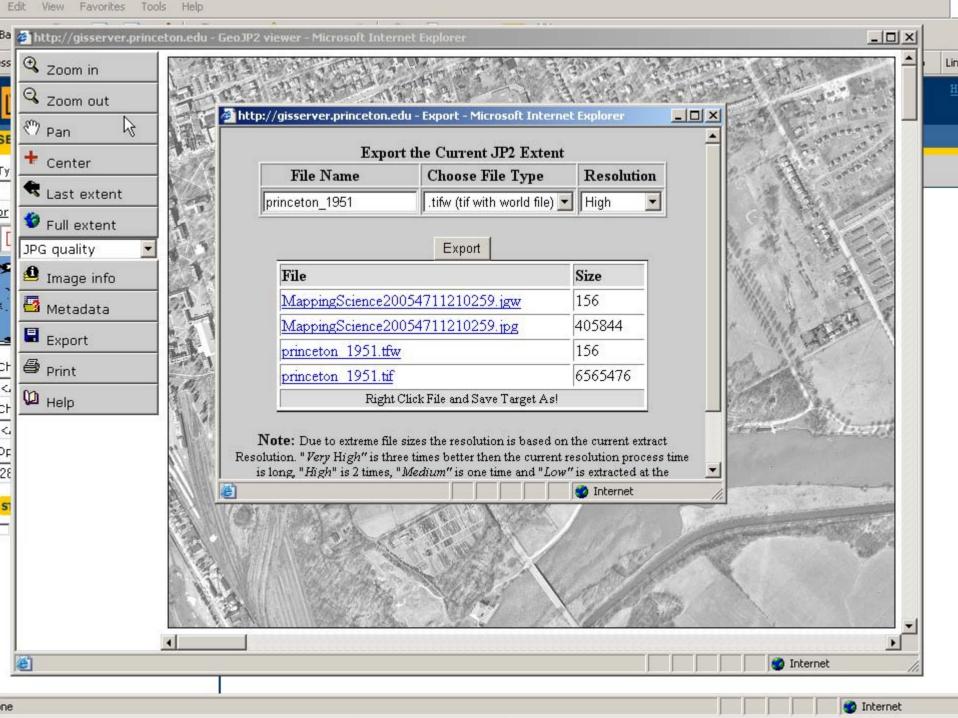






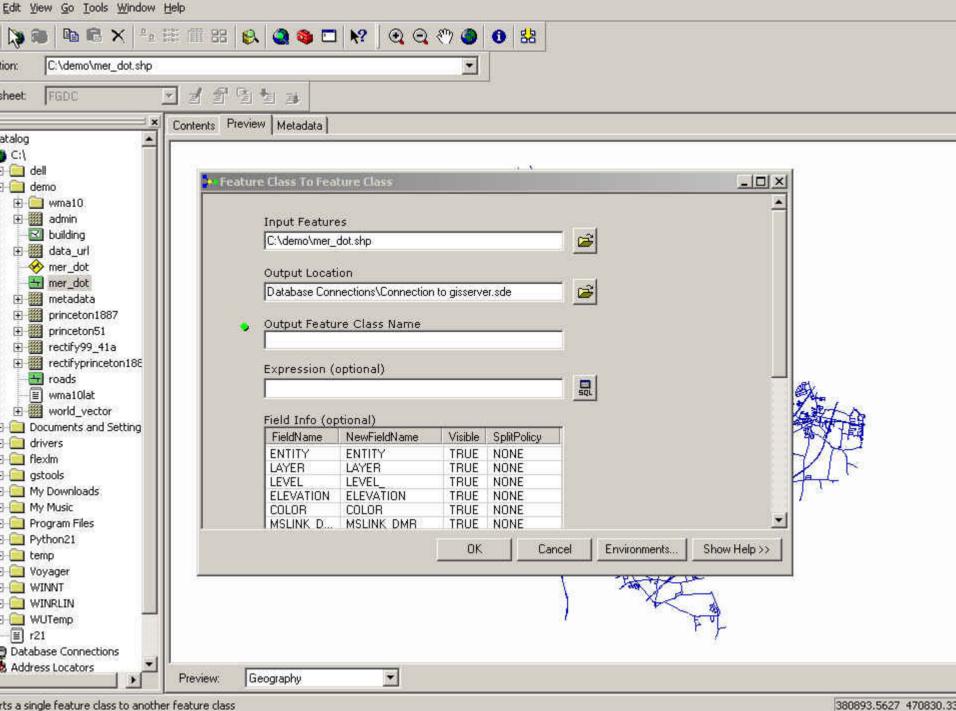


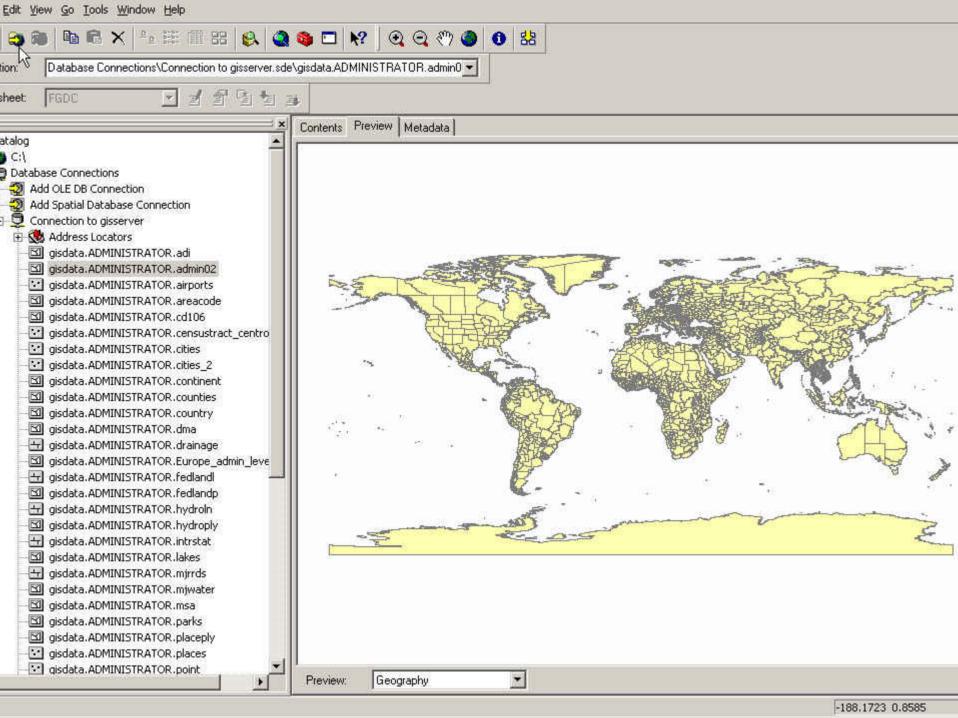


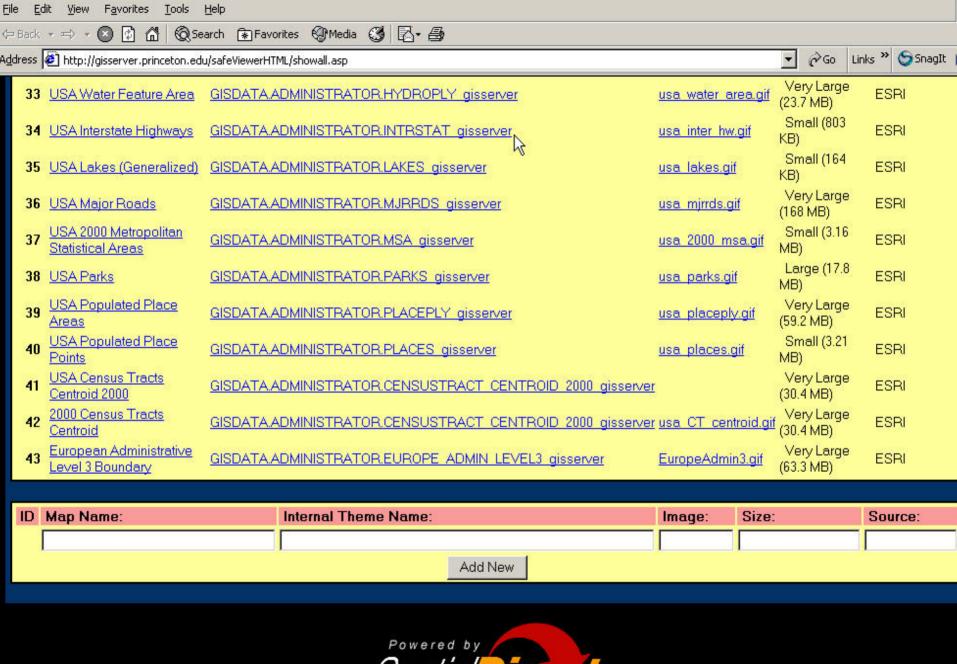


Workflow for vector data

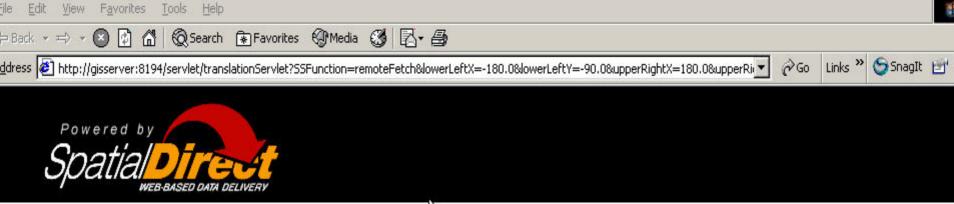
- Upload datasets into ArcSDE using ArcCatalog
- Use SpatialDirect's Spatial Assistant to connect to ArcSDE tables
- Select the data and set the right permission
- Open Administration interface web page and enter necessary information
- Create the map image and URL
- Create metadata, copy URL in metadata, and publish it in ArcIMS metadata server
- Data is ready to search, view, browse, and download in Metadata Explorer











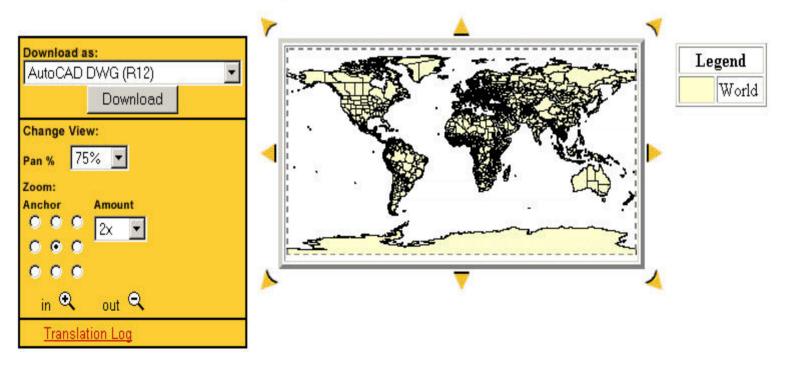
Main Menu

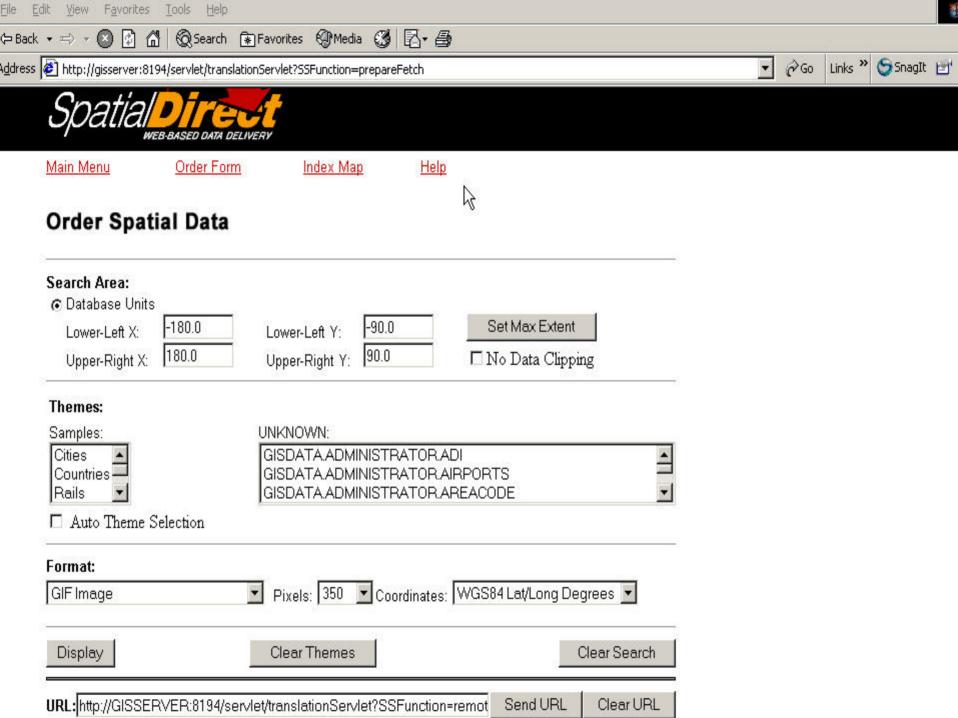
Order Form

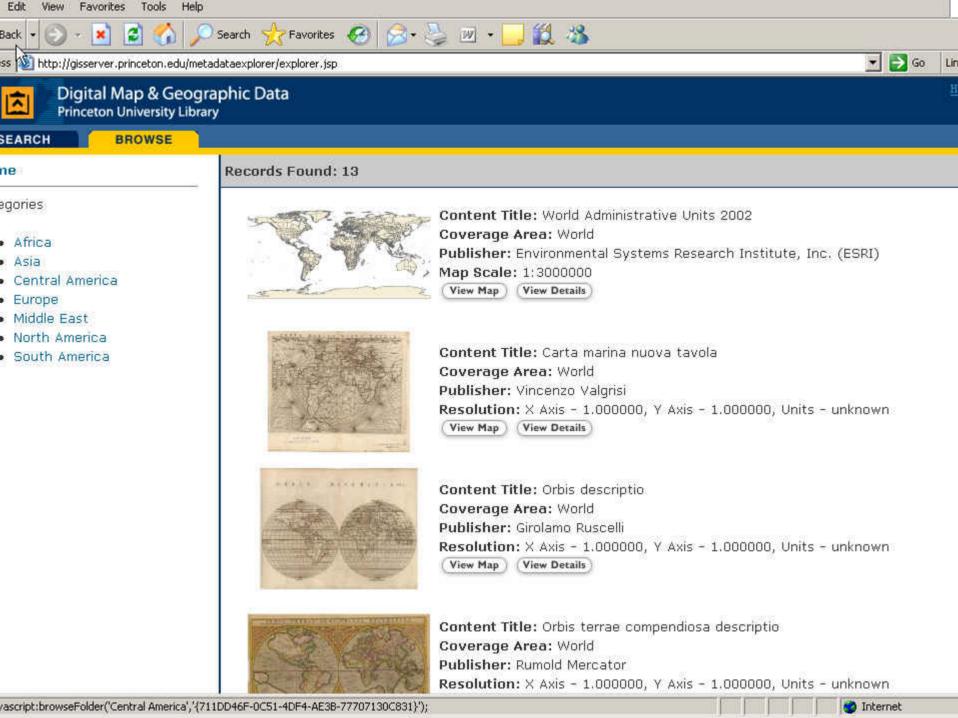
Index Map



Map Results









Opening page http://gisserver.princeton.edu/safeViewerHTML/frontpage.asp?id=9...



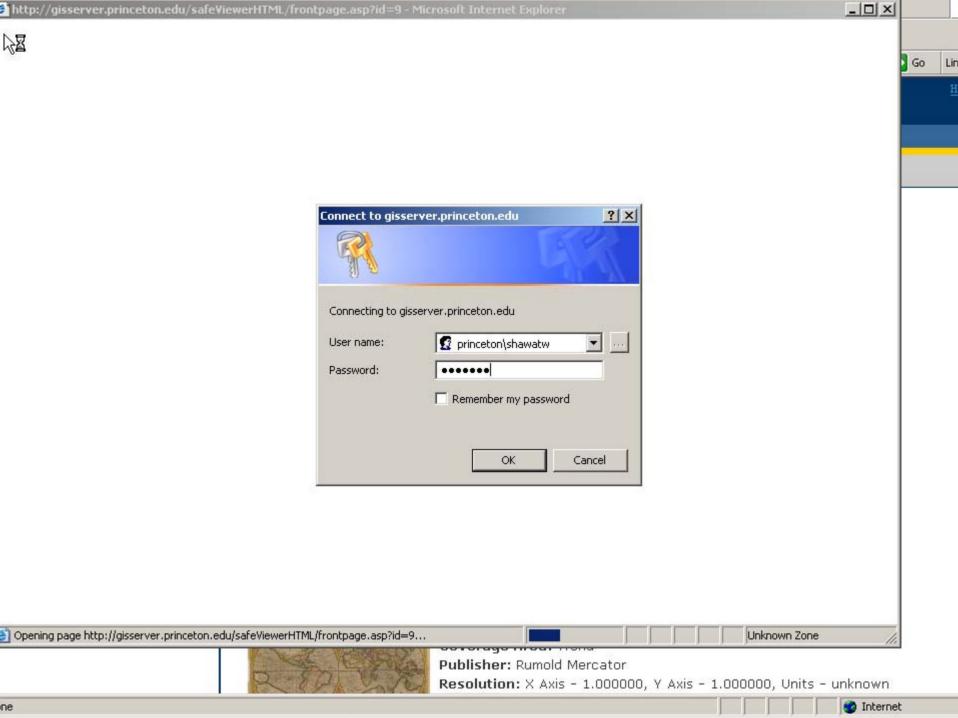
Unknown Zone

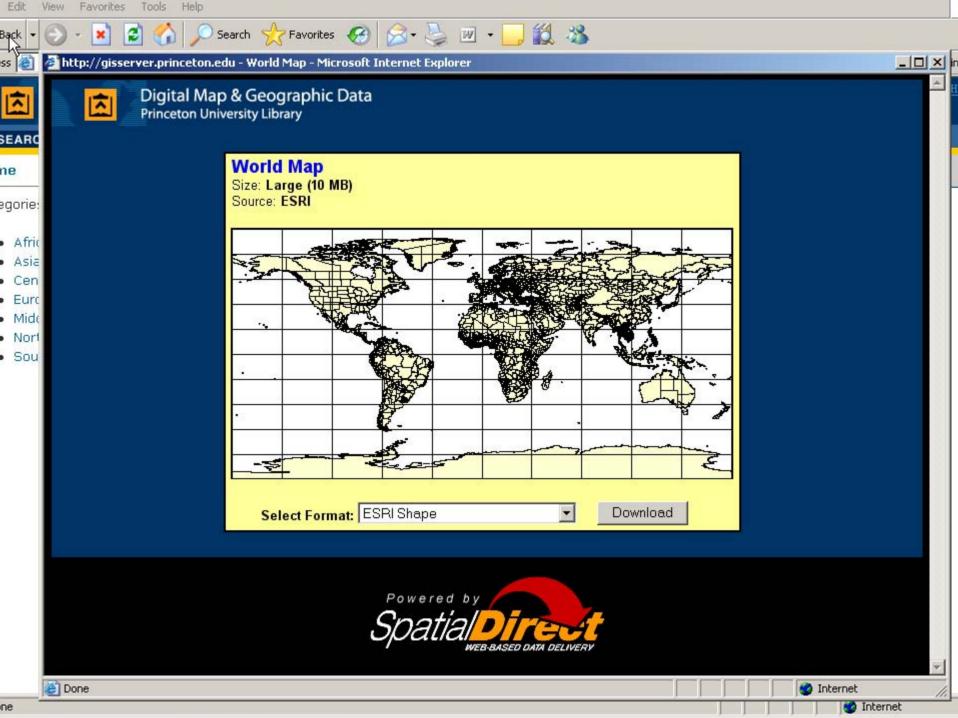
Internet

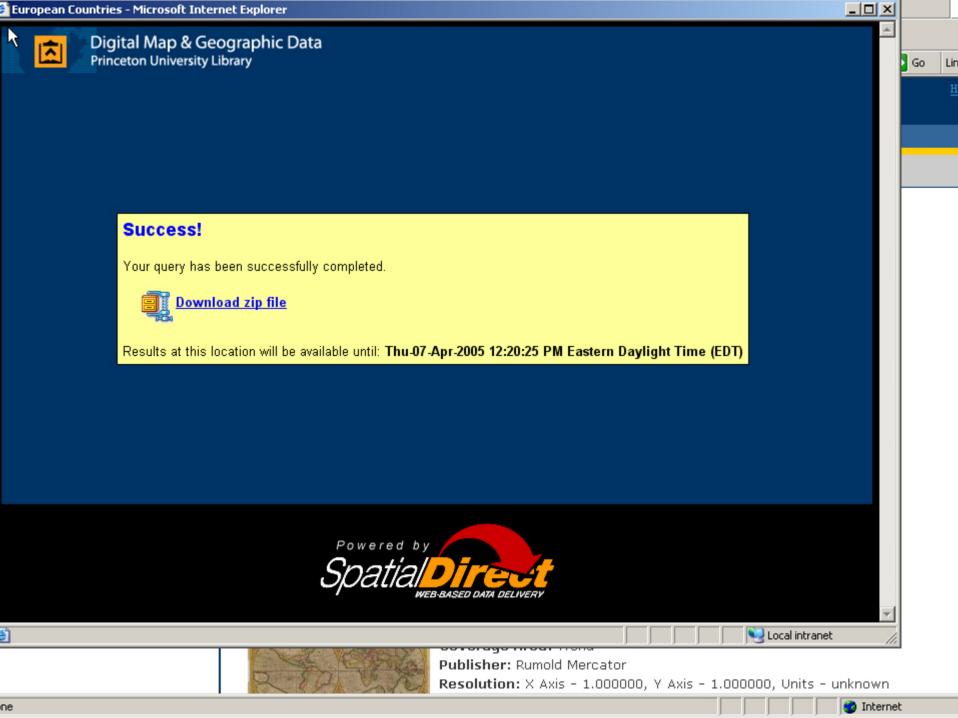
_ | U X

Lin

Go



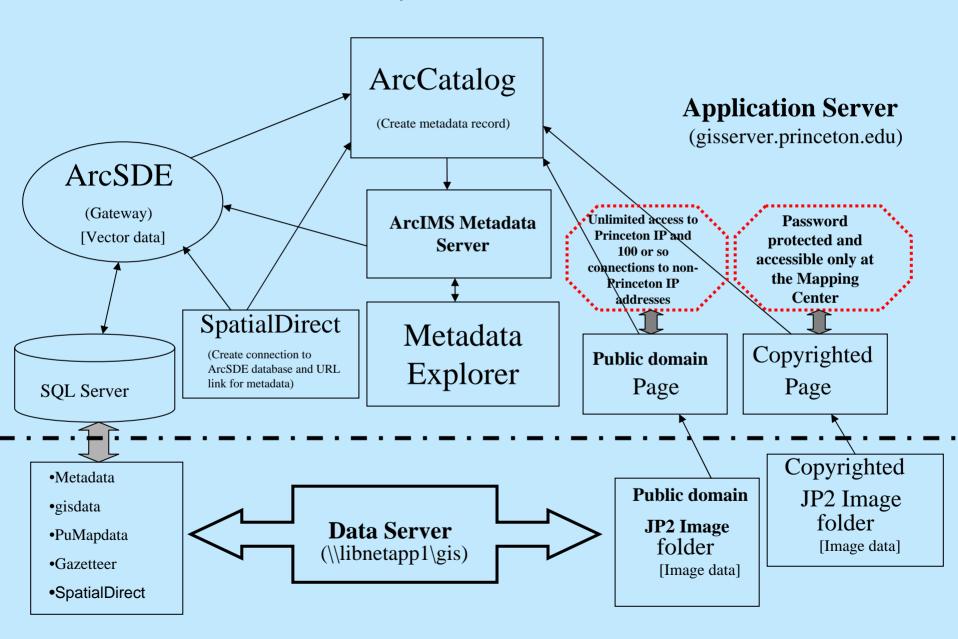




Lessons Learned

- Staff time: building a complex digital data infrastructure required help from Systems, database specialists, and programmers. To continue with a scanning project, you need a dedicated support staff. Depending on student workers is not very productive. However, using computer science students to design custom interfaces and applications can be very productive because they have different computer language skills.
- Hardware issues: to scan maps you need a lot of memory on a computer. We have a computer with 1GB of RAM and on a few occasions I found that it was not enough, especially when scanning a large map. I also found that putting scanned maps online required a large disk space. For example, 1300 maps (different sizes) that were scanned at 400 dpi with 256 colors needed 140 GB non-compressed TIFF file and 50GB of JP2 file compressed at 1:10 ratio.

Future System Architecture



Example of data visualization

• Click here to see an example of data visualization of Princeton University campus using georeferenced 1887 topographic map, 1951 and 1999 aerial photographs, campus building footprints, and Mercer county road dataset. All of these datasets were downloaded from our Digital Map and Geographic Data web site http://gisserver.princeton.edu/metadataexplorer/ex plorer.jsp

Thanks for coming to the presentation

Any Questions?

Tsering Wangyal Shawa, GIS Librarian Princeton University Library,