North Carolina Geospatial Data Archiving Project (NCGDAP)

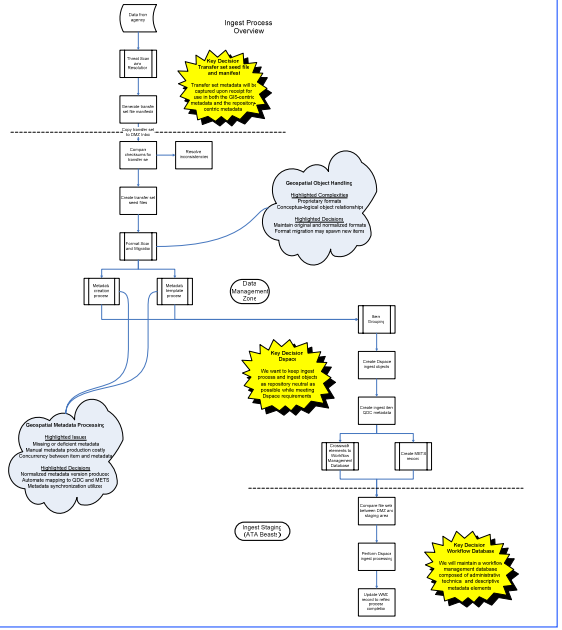
Digital Preservation Partners Meeting -- January 2006

Preparing Spatial Objects for Repository Ingest

Geospatial content will be received from a wide range of data producers including state, county, and city agencies. Each agency employs its own data dissemination practice and substantial processing is required in order to prepare content for the archive. The process needs to be flexible enough to handle numerous spatial formats, sophisticated enough to discriminate any and all metadata, yet normative enough to effectively transfer all these spatial data objects from data producer into a single spatial data repository. Several guiding principles govern the ingest process, including: the need to automate as much of the process as possible; the refusal to sacrifice accuracy; and the importance of designing a process flexible enough to create submission information packages for multiple repositories.

Geospatial content arrives in numerous formats and with metadata of significantly varying completeness. Overall ingest objectives include retention of the spatial data object and metadata (when included) in their original condition as well as archiving of more normative content formats. In general, these objectives dictate the creation of shapefiles for vector data, TIFFs for raster data, and FGDC-structured XML for metadata, when the data is not received in these forms. Geography Markup Language (GML) will be explored as a non-proprietary alternative to the shapefile.

Key Decisions Affecting the Ingest Process



Engaging Spatial Data Infrastructure in Preservation

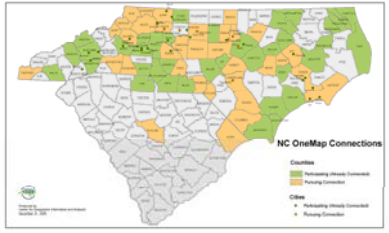
What is Spatial Data Infrastructure?

The **National Spatial Data Infrastructure (NSDI)** brings together federal, state, and local agencies, as well as commercial entities in the development of a framework in which to find, use, and contribute digital geospatial data. Spatial data infrastructure has addressed such issues as data and metadata standards, discovery mechanisms, and inventory processes; however, data preservation has not been addressed. One core objective of NCGDAP is to engage existing spatial data infrastructure in the problem of preservation. **NC OneMap** is a next generation mechanism to coordinate and disseminate geographic information in North Carolina. While the NC OneMap network continues to expand within the state, collaboration at the national level also deepens through interactions with the NSDI and the National States Geographic Information Council.

NC OneMap Objectives:

- Build a **common understanding** of North Carolina data resources
- Develop ongoing **data inventory** for all geospatial data holdings
- Develop **content standards** for key data themes
- Enable widespread **access and distribution** of geospatial data

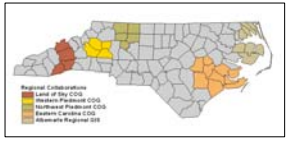
One of the defined characteristics of NC OneMap is that "*Historic and temporal data will be maintained and available*".



Regional Partnerships Form Building Blocks

Emerging regional collaborations focused on development of shared infrastructure for cultivating access to data, have lowered the barrier to participation in networks. These regional groups become test beds for innovation in the area of data sharing and data management, including archiving.

NCGDAP has embarked upon a series of regional site visits to tap into local learning experiences and to explore opportunities for deeper engagement with local agencies.



Exploiting Next-Generation Geospatial Technologies

The **NC One Map Viewer** and the **National Map** are based on Open Geospatial Consortium (OGC) specifications for data interoperability, drawing content from a variety of different local technical environments. One objective of NCGDAP is to explore the use of emerging geospatial web services technologies in the process of archive development and to introduce preservation issues into the discussions of the OGC.

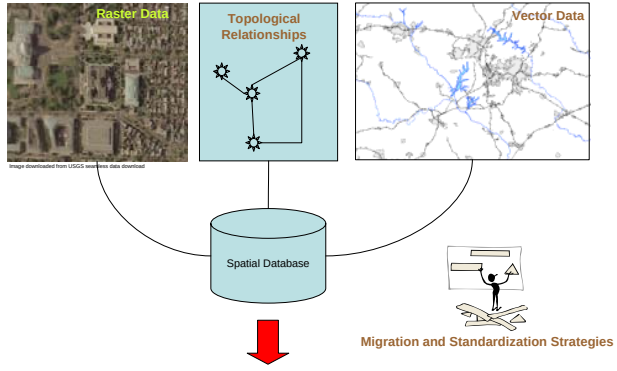
NCGDAP partnered with EDINA (at University of Edinburgh in the UK) to present on preservation issues to the **OGC Technical Committee** at the November 2005 Meeting in Bonn, Germany. In response, the Architecture Working Group of the OGC has added preservation to its list of working topics.

Geospatial Data: Going from Complex Objects to Ingest Items

An Example of Conceptual-Logical Object Relationships

The Spatial Database

- The spatial database is an increasingly common structure for managing spatial information
- A single spatial database will spawn multiple ingest object formats, including TIFF, XML, and shapefile
- TIFF and shapefile formats are complex objects composed of multiple required and optional logical components
- Other complex spatial object formats in common usage can have shared files and directories



Notable Ingest Item Creation Issues

- Rasters can only be exported from Geodatabases as ERDAS Imagine format, requiring further conversion of the images to TIFF
- Time versioning—different versions of the same object are certainly individual logical items, but should they be separate repository items?
- The coverage format is especially unique—a single index file may manage multiple otherwise independent data coverages, which also reside outside of the shared directory containing the index file

Presented by: Rob Farrell, Steve Morris, Zsolt Nagy, and Jim Tuttle