



*New York State  
Archives*

# Managing and Preserving Geospatial Records

Jennifer O'Neill  
Scheduling and State Agency Services Supervisor  
October 2017



# Agenda

- Identify records retention period
- Organize and label files well
- Use widely adopted or open source formats and standards
- Populate metadata fields
- Store on centralized server with adequate backups
- Dispose of records per State Archives' retention schedules
- Identify appropriate preservation strategy
- Migrate files to newer formats as needed before access lost



# Archives' Retention Schedules

- Authorize destruction of records or transfer of records to appropriate archival repository
- Local governments:
  - MU-1 Schedule for municipalities
  - CO-2 Schedule for counties
  - ED-1 Schedule for school districts
  - MI-1 Schedule for miscellaneous governments
- State agencies:
  - State General Schedule
  - Agency-specific schedules



# Why is Retention Important?

- Meet a legal requirement
- Mitigate litigation issues
- Save resources
- Improve records retrieval
- Identify and preserve important records
- Eliminate ad hoc decisions



# Permanent/Archival GIS Records

- Orthoimagery
- Boundaries (e.g., municipal, school district, election district)
- Transportation (e.g., street centerlines, highways)
- Utilities and communication infrastructure (e.g., water, sewer, gas, phone)
- Land ownership (e.g., parcel, state-owned lands, survey data)
- Land use (e.g., zoning, conservation, agriculture)
- Hydrography (e.g., streams, lakes, water shed)
- Elevation



# Non-Archival GIS Records

- Municipal lighting
- Emergency dispatch
- Crime reports or studies
- Parcel data
- Highway or transportation structure maintenance and repair
- School bus routing and scheduling



# Retention of Related Records

- System security and usage monitoring records
  - System-wide: destroy 10 years after date of last entry
  - Individual user: destroy after 3 system backups
- GIS metadata and data dictionary records
  - Destroy 3 years after discontinuance of system but not before system data is destroyed or migrated to new system
- Backup files
  - Destroy after 3 system backups
- Application development records
  - Destroy 3 years after completion of project
- Data processing policies and procedures
  - Destroy 3 years after withdrawn, revised, updated, or superseded



# Challenges with Managing GIS Records

- Long-life
- Multiple types of data and formats
- Proprietary formats
- Large size
- Dynamic nature
- Extensive context and relational nature





# File Naming Convention

- Geographic-area-or-extent\_Dataset-name-or-theme\_Year\_Month
  - Wake\_Parcels\_2005\_09 or Shellfish\_Growing\_Areas\_2008\_08
  - MTLandCover2010 or MTCensusBlock2000
- Avoid blank spaces or special characters
  - Use an underscore or a dash instead
- Limit file name character and path length



# NYS Standards

- GPS Data Collection Guidelines
- Datum and Coordinate System Standard
  - NAD-83 – Horizontal; NAVD-88 – Vertical
  - Large scale, State Plane; Small scale, UTM zone 18 (extended)
- Metadata Standard
  - FGDC Content Standards for Digital Geospatial Metadata
  - Cadasteral Data Standard
- Four Character County Code Standard
  - First four letters of county, e.g., ESSE



# File Organization

- Define an organizational structure for datasets based on
  - data retention strategy
  - form of the datasets being retained, such as:
    - individual geospatial assets (e.g., individual shapefiles or a single layer in a geodatabase format)
    - collection of geospatial assets (e.g., multiple layers in a geodatabase, or an orthophoto comprised of numerous individual tiles)
- For archivists, base first level of organization on the general type of geospatial asset being stored:
  - Vector Digital Data
  - Raster Digital Data (assets such as aerial imagery, orthophotography, and elevations)
  - Projects
  - Digitized Maps



# Categorizing Vector Data

(based on ISO 19115 – Geographic Information Metadata topic categories or themes)

- biota
- boundaries
- climatologyMeteorologyAtmosphere
- economy
- elevation
- environment
- farming
- geoscientificInformation
- health
- imageryBaseMapsEarthCover
- inlandWaters
- intelligenceMilitary
- location
- oceans
- planningCadstre
- society
- structure
- transportation
- utilitiesCommunications



# Random Access Metadata tool for Online National Assessment (RAMONA)

## Framework Data Category

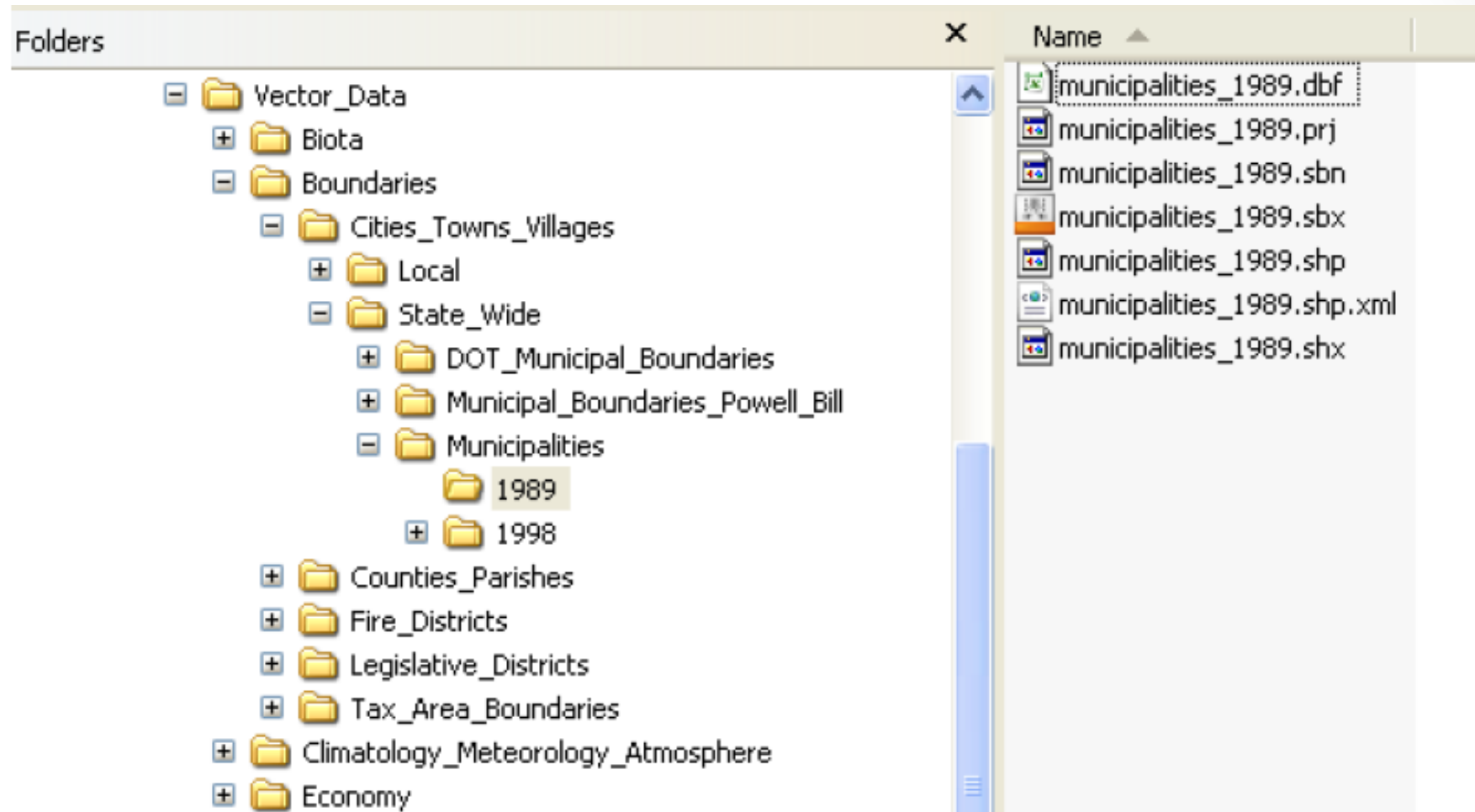
- Boundaries
- Imagery/Base Maps/Earth Cover
- Inland Waters
- Transportation

## Data (sub)layer

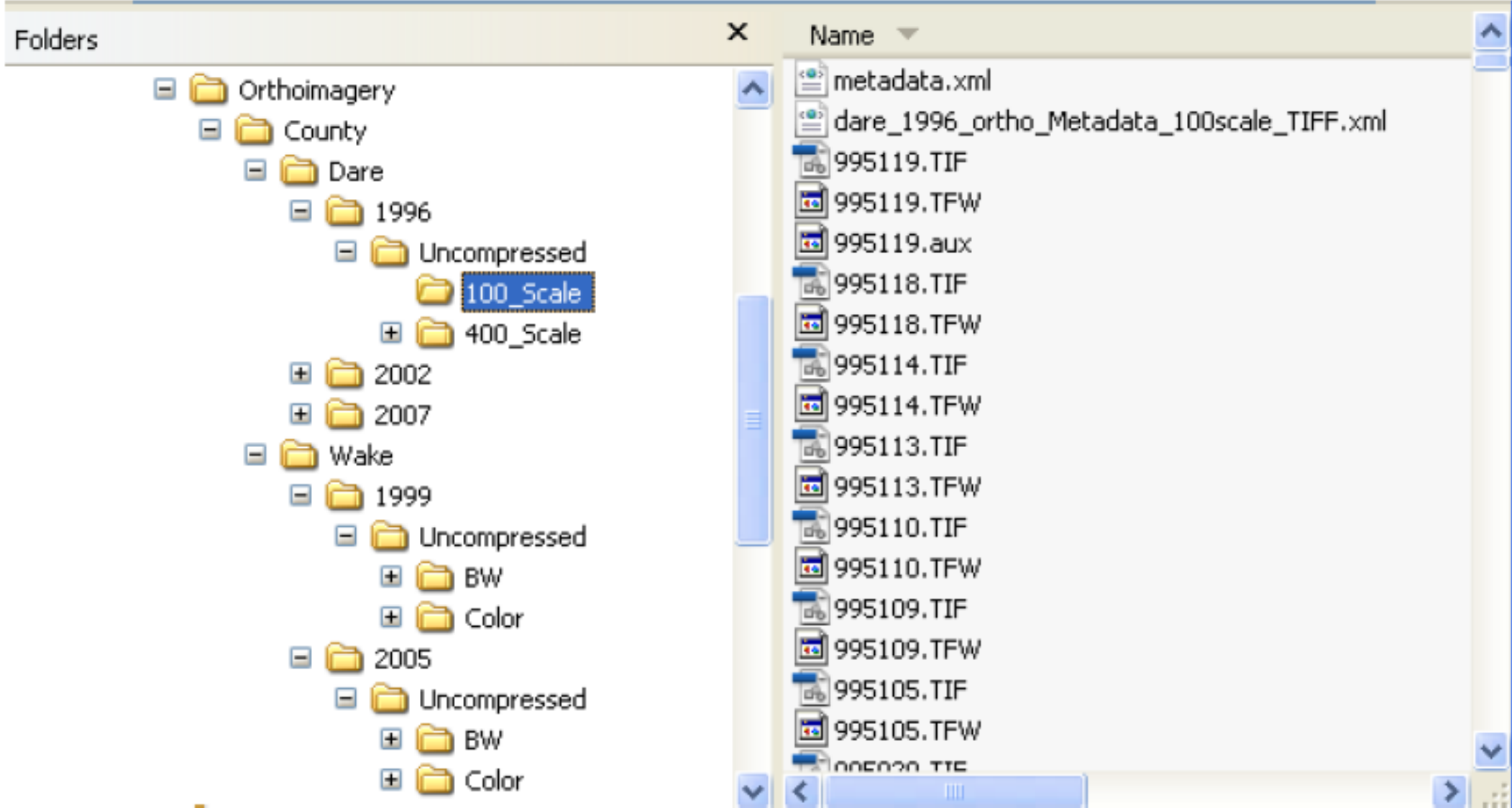
- American Indian Reservation
- Cities/Towns/Villages
- Civil Township or Equivalent
- Counties/Parishes
- State
- Digital Orthophotography/ Orthoimagery
- Land Cover
- Hydrography
- Watershed Boundaries
- Airports & Airfields
- Railroad Lines
- Roads/Street Centerlines



# Vector Filing Example



# Orthoimagery Filing Example



# Selecting a Preservation Format

- Format should be:
  - Publicly and openly documented
  - Non-proprietary
  - Widely used
  - Self-documenting
  - Able to be opened, read, and accessed using readily-available tools





# Archival Geospatial Formats

- Preservation formats
  - GeoTIFF
  - Geographic Markup Language (GML)
  - Keyhole Markup Language (KML)
  - ESRI Shapefile (ver. 1997 to current)
- Access formats
  - MrSID or JPEG tiles
  - Geospatial PDF
  - ESRI Shapefile



# Other Types of File Formats 1

- Textual
  - ASCII
  - Unicode
  - PDF/A
  - OpenDocument Text Format (ODF)
- Digital images
  - TIFF
  - PDF/A



# Other Types of File Formats 2

- Video
  - Audio-Video Interleave format (AVI)
  - Material Exchange Format (MXF)
  - Quicktime format (MOV)
  - Windows Media Video format (WMV)
- Computer Aided Design (CAD)
  - Extensible 3D (X3D)
  - Standard for the Exchange of Product Model Data (STEP)
  - PDF/E



# Populate Metadata Fields

- Define for current staff and future researchers
- GIS software auto-populates some metadata
- Follow Federal Geographic Data Committee (FGDC) Content Standards for Digital Geospatial Metadata (CSDGM) guidelines



# Metadata Example

- County orthoimagery md provided by NYS Orthoimagery Program:
  - Identification\_Information
  - Data\_Quality\_Information
  - Spatial\_Data\_Organization\_Information
  - Spatial\_Reference\_Information
  - Entity\_and\_Attribute\_Information
  - Distribution\_Information
  - Metadata\_Reference\_Information



# Storing GIS Records

- Store on centralized server with adequate backups
  - Identify sensitive infrastructure or other information
  - Can use internal, vendor, or cloud server
  - Apply appropriate access controls
  - Ensure backups are multiple copies stored in separate locations



# Disposing of GIS Records

- Destroy any non-archival records per retention schedule
- Transfer archival records to appropriate repository
  - State agencies: State Archives
  - Local governments: designated unit (e.g., clerk, creating unit, IT)



# Preserving Long-term GIS Records

- Determine preservation strategy
  - Capture one copy of static maps/orthophotos
  - Take snapshots of GIS data (individual or bulk)
  - Migrate data to new software version or format
- Determine frequency of capture
  - Based on significance of record and its alterability
  - May be different for various layers
- Determine what to capture
  - Geometry, attributes, and metadata





# Conclusion

- Design system, select file formats, and organize records with final disposition in mind
- Consider future research uses
  - Develop meaningful metadata
  - Identify copyright or other legal access restrictions
- Develop a preservation plan
  - Plan to migrate records regularly
  - Transfer records to “archival” repository



# Resources

- NYS GIS Clearinghouse <https://gis.ny.gov/>
- GeoMAPP <http://www.geomapp.net/>
- NYS Archives' Digital Imaging Guidelines  
[http://www.archives.nysed.gov/common/archives/files/mr\\_erecords\\_imgguides.pdf](http://www.archives.nysed.gov/common/archives/files/mr_erecords_imgguides.pdf)
- National Archives Table of File Formats  
<https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>
- Library of Congress Recommended Formats Statement  
<http://www.loc.gov/preservation/resources/rfs/>

