GIS Data and File Management

Resource Note #10

Community Social Profile Project

January 2008

Preface

- The Resource Note is intended to increase our understanding of and broaden our knowledge base on key subject areas that are fundamental in building our capacity in numeric and geographic analysis.
- It is not an in-depth or comprehensive discussion of the subject matter.
- It highlights certain relevant and important areas that deserve our attention and consideration.
- It is intended to be informal and informative.

Introduction

- This is the 10th Resource Note, previous notes include:
  - # 1 – Census Geography
  - # 2 – Census Data
  - # 3 – PCensus Database
  - # 4 – Geocoding
  - # 5 – Cartographic Principles
  - # 6 – Thematic Mapping
  - # 7 – 2006 Census
  - # 8 – Population Pyramid
  - # 9 – PCensus DBX v8.5

Introduction

- This resource note addresses a number of issues related to GIS data and file organization. Some of us come across these issues on a regular basis in doing mapping and spatial analysis work.
- For example:
  - I cannot tell the exact content of a shapefile from the filename
  - I have created many files with similar file names over time
  - I cannot find a file that I have created some months ago
  - I have many files with this filename> Geocoding_Result.shp
  - Some of my mxd files have broken links
  - What is the minimal amount of information that should included in a file name?
  - I cannot open a shapefile sent to me by someone
  - What is the best way to organize my GIS data file?
  - Others…

Introduction

- Most of these issues arise as a result of the ways we create, document. organize and manage our GIS data and files
- A well-thought out and executed GIS data and file management system can help us and others to locate and retrieve the right file(s) and to avoid unnecessary confusion, frustration, stress and waste of resources
- The fact that a number of files are required to support a typical GIS mapping and analysis project and GIS files may be accessed by others further underscores the importance of GIS data and file organization and management
- Also, ArcMap.mxd file *does not contain* any data, only points to the data files which must be saved and available in addition to the mxd, one has to know the 'whereabouts' of the data sources in the event that the "links" are broken
- The areas covered in this note include:
  - General and project-based GIS data files
  - GIS file naming convention
  - Create and update Metadata
  - Referencing data – full and relative paths
  - Components of a shapefile

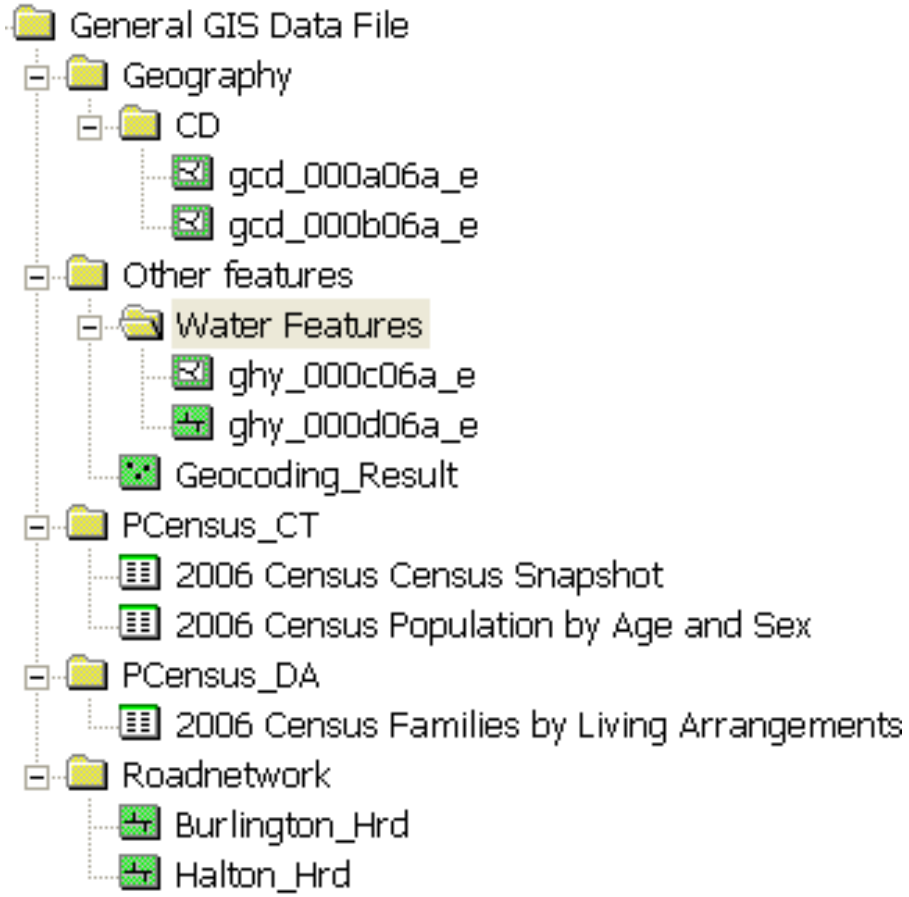
Organize GIS data

- One effective way to organize GIS data is to group data into file folders
- '…you know where you put them, you know where to find them…'
- Generally, there are two main groups of GIS data
  - General GIS Data : one you use on a regular basis organized by type of data (Census data) or geographic extent, e.g. City of Toronto by Census Tract or Town of Oakville by Dissemination Area
  - GIS Project Data: the other group is project based which contains all data, and ArcView project files for particular project and/or set of maps, e.g. Urban Poverty Project for the City of North Bay.

General GIS data

- Examples of General GIS data are:
  - Base map – holds shapefiles such as CSD boundaries, CT/DA boundaries, Electoral Districts, or any user defined areas
  - Road network – holds shapefiles of different levels of road network (provincial highways, regional roads, local streets)
  - Other geographic features – water bodies, parks, schools, shopping centres…
  - Database – holds data from sources such as Statistics Canada, user directories, school boards
    - For Census data, one can further group data by sources
      - From Statistics Canada (Beyond 20/20)
      - From Tetrad Computer Applications Inc (PCensus DBX)
  - Others

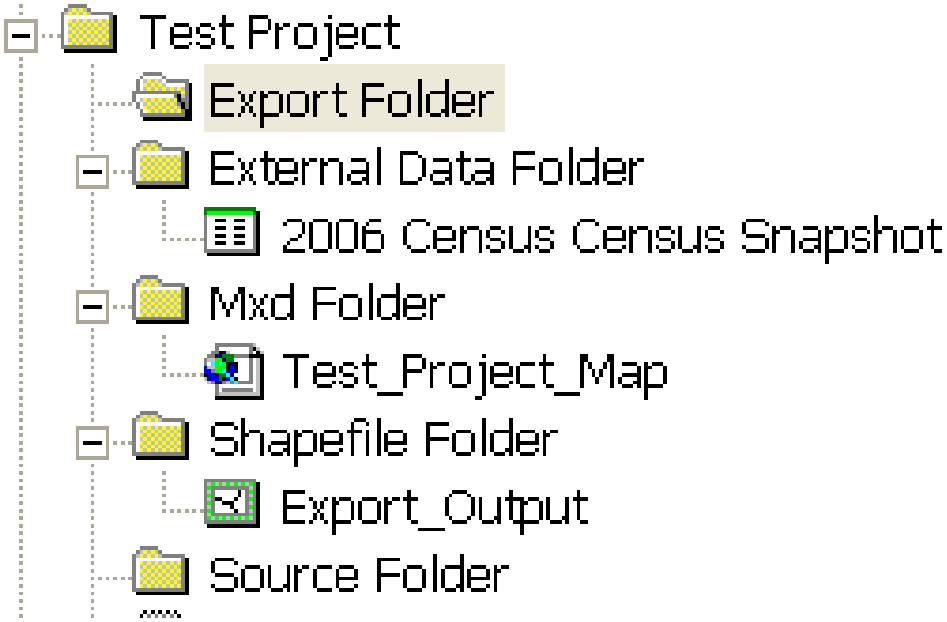
General GIS Data



GIS Project Data

- Create an overall Project Folder where you will keep all folders of data, tables and maps, etc
  - Source Folder: Word or Excel document for lists of sources (websites, etc) that you have used to download data, to keep track of where you data came from
  - Shapefile Folder: boundaries of your study area, roadnetworks, Geocoded results…..
  - External Data Folder: Census data, local data, survey data..
  - MXD Folder: all maps created
  - Exports Folder: map documents that have been exported as PDFs, JPGs, etc.

GIS Project Data



PCensus database

- Since PCensus naming convention of the Export Pointfile (attribute table for joining with shapefile) does not distinguish different geographies (e.g. CSD, CT or DA), e.g.
  - 2006 Census Census Snapshot.dbf
  - 2006 Census Census Snapshot1.dbf

  you may wish to create separate files for each geography

Organizing data with ArcCatalog

- GIS data can be organized with ArcCatalog the same way we manage files with Windows Explorer
- Use all standard Windows shortcuts and drag and drop techniques
- When you copy, rename, or delete items with ArcCatalog, ancillary files associated with the item are also copied, renamed or deleted. For example, metadata, projection, and index files may all accompany a shapefile

Organizing data with ArcCatalog

- Creating a new folder
  - Click the folder or folder connection in which you want to create a new folder
  - Click the File menu.
  - Point to New and click Folder
  - Type a new name for the folder
  - Press Enter.

- Creating a new file or personal geodatabase
  - Click the folder or folder connection in which you want to create a new file or personal geodatabase
  - Click the File menu
  - Point to New and click File Geodatabase or click Personal Geodatabase
  - Type a new name for the geodatabase
  - Press Enter.

Organizing data with ArcCatalog

- Copying an item
  - Click the item you want to copy
  - Click the Copy button
  - Click the folder or geodatabase to which you want to copy the item.
  - Click the Paste button
- Renaming an item
  - Click the item you want to rename
  - Click File and click Rename
  - Type the new name
  - Press Enter.

GIS Data Naming Convention

- The purpose of a naming convention for GIS files is to promote consistency in naming, clearer understanding of the information of the files and to facilitate the identification of GIS data sets between data users
- A poorly named file not only causes confusion and frustration but also may lead to erroneous results and interpretation
- Many jurisdictions have made recommendations on GIS data naming convention in order to provide some level of consistency so that sharing of data becomes easier for the end user. We are not recommending any particular naming convention for your organization because you are in a better position to determine what meets your own conditions and requirements
- Here are some examples of naming convention used by government organizations and a mapping company

GIS Data Naming Convention

- Statistics Canada – Spatial file
  - gxxxmmma05x_e
    - g – projection of file (g- geographic (longitude/latitude)
    - xxx – primary geographic area (csd-census subdivision)
    - mmm- geographic code of coverage (035-Ontario)
    - a – file type (a-digital boundary file)
    - 05 – dissemination year (05- disseminated in 2005)
    - x – file format (a-ARC/INFO format)
    - _e – language (_e – English)

GIS Data Naming Convention

- The Indiana Geographic Information Council recommends:
  - keyword_steward_extent-date
  - roads_DNR_city_20070107
    - Roadnetwork
    - Created by the Department of Natural Resources
    - For the City or municipality
    - Created January 7, 2007

GIS Data Naming Convention

- Avenza Systems Inc's MAPDataUSA3.0 TIGER:
- cccccAAAAAxxx.yyy
  - ccccc – City or municipality name
  - AAAAA – 5 digit County FIPS code
  - xxx – layer abbreviation (e.g. rds – roads)
  - yyy – file extension (e.g. shp – shapefile)

GIS Data Naming Convention

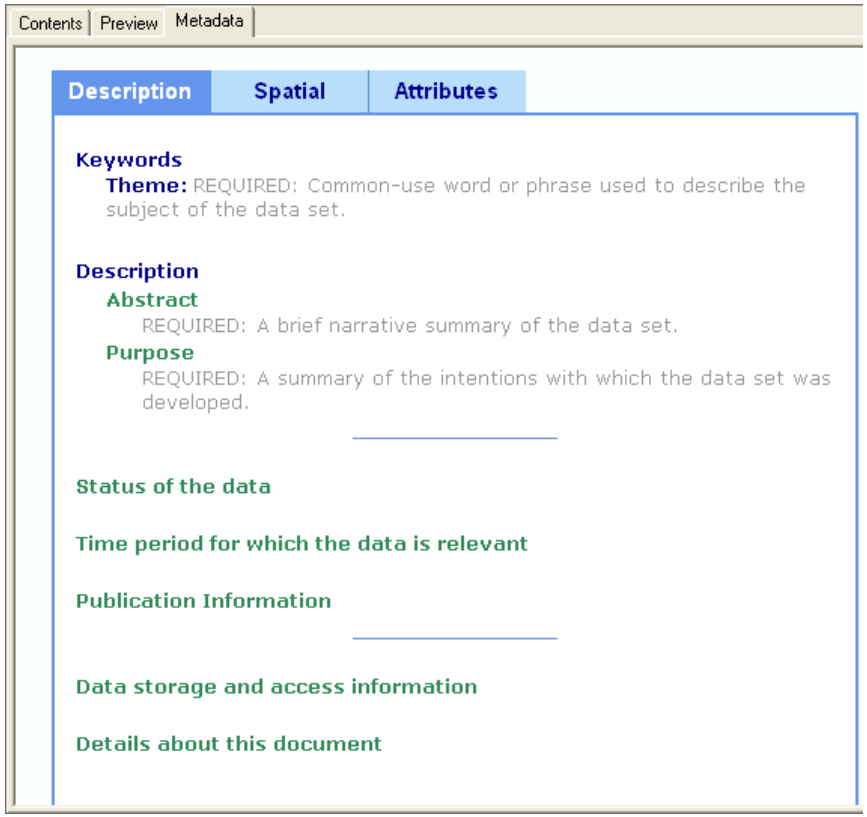
- Some useful guidelines
  - Name contains the key information about the file
  - Keep the name short (8-15 characters)
  - Do not use spaces
  - Do not use special characters (except the underscore (__))
  - Do not put a number as the first character (e.g. y2001 instead of 2001)
  - Keep the name simple (use abbreviation where practical)
  - Use terminology that are familiar within your organization and community

*Metadata**

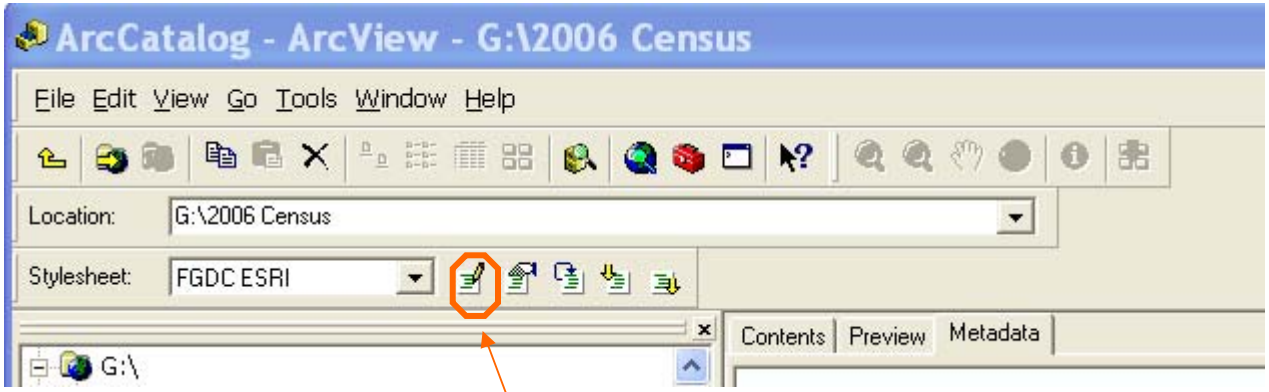
- What is Metadata? It is data on data
- As in a library, GIS file name is the title of a book and metadata serves as a card in the card catalogue describing the book
- It contains information on the file such as:
  - Date
  - Creator
  - Geographic extent
  - Coordinate system
  - Attribute domain
- Metadata is critical for sharing tools, data, and maps and for searching to see if the resources you need already exist
- Metadata is created, edited and updated with ArcCatlog
- Once created, metadata is copied, moved, and deleted along with the item when it is managed with ArcCatalog or ArcInfo Workstation

* ArcGIS Help_Data management with ArcCatalog

Metadata Summary - ArcCatalog



Creating and editing Metadata

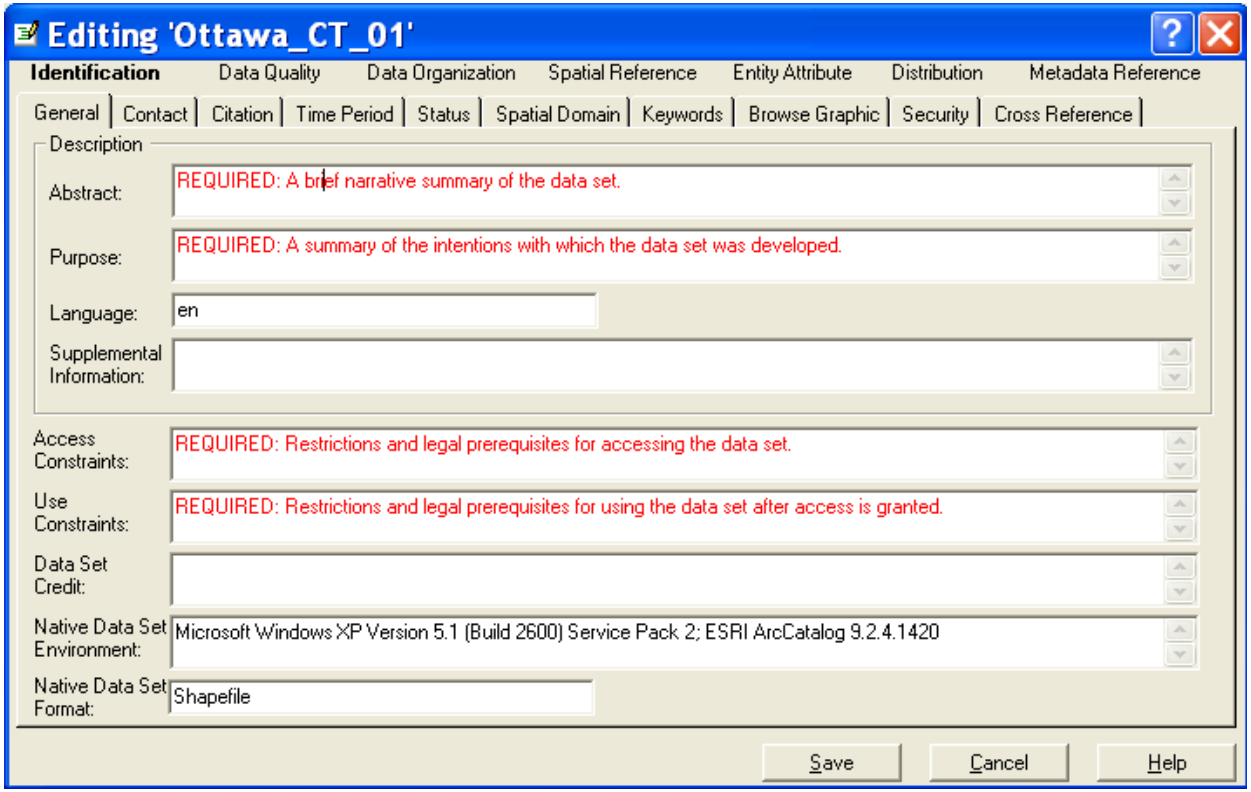


Edit Metadata

Sections and subsections

- All the information is captured in seven sections and subsections. The seven sections are;
  - Identification
  - Data Quality
  - Data Organization
  - Spatial Reference
  - Entity Attribute
  - Distribution
  - Metadata Reference

Metadata Template



Metadata - Help

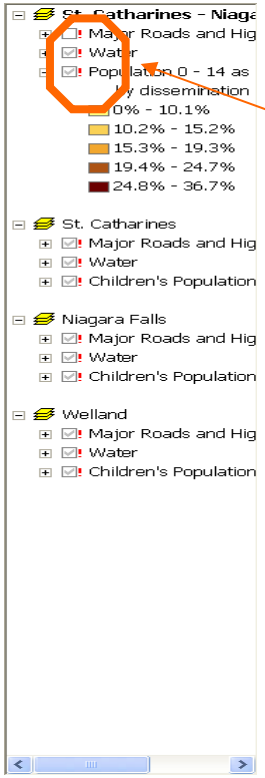
- ArcGIS Desktop Help provides information on how to create, edit, store and publish Metadata
- A useful step-by-step tutorial can be found at:

www.pasda.psu.edu/tutorials/metadata.asp

Referencing Data

- As mentioned earlier in this note, ArcMap .mxd file does not store the data files, instead it stores references to the locations of the data sources
- When a map document is opened, ArcMap reads the file and looks for all the path names to reconstruct the layers
- When data sources are moved, map documents can lose track of the source data for its layers. When this occur, the layer name appears in the ArcMap Table of Contents with a red exclamation mark, and nothing appears for that layer in the display area

Repair Broken Links



Broken links

The following steps will update the path to the data so that the layer can be drawn:

- Right-click the layer with the red exclamation mark in the Table of Content
- Click Data
- Click Repair Data Source
- Navigate to the correct folder where your data is stored
- Highlight the file name
- Click Add

Referencing Data

- Broken links can be avoided by referencing the data differently
- ArcMap has two options for referencing file-based data:
  - Full paths – everyone who uses the map must either do so on the same computer or have the data on their computer in exactly the same folder structure
  - Relative paths – enable the map and its associated data to be moved to any disk drive without the map having to be repaired. As long as the same directory structure is used at the new location, the map will still be able to find its data by traversing the relative paths

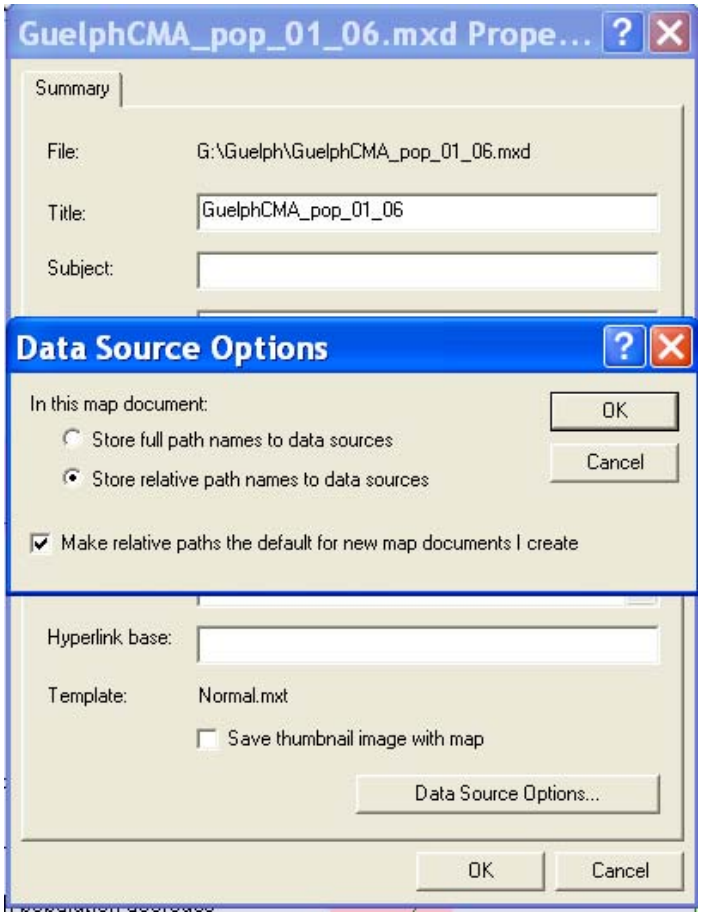
Referencing Data

- When you save your map always use relative path names
- This will allow you to move the whole project folder between computers and you do not lose any data (as long as all data displayed on your map is in the project folder)

Referencing Data

- To save map using relative pathnames:
  - Go to File> **Document Properties**
  - Click **Data Source Options** button
  - Select **Store Relative Path names to data sources**
  - Click **Ok**

Referencing Data



Shapefile

- An issue related to GIS file organization is the handling of shapefiles
- A common problem is the inability to open a shapefile in ArcMap although the file has an .shp file extension
- A shapefile is comprised of a number of files with same prefix for their filename but different extensions
- For example: community.shp has the following component files
  - Community.shp file contains the shape
  - Community.dbf file contains the attribute table
  - Community.shx file contains an index which connects the table to the shape
  - Community.prj file contains the coordinate system definition for the shapefile
  - Community.shn and Community.sbx are generated by the software

Shapefile

- Therefore, when copying, moving, renaming or deleting a shapefile you have to make sure that you perform the operation on all the component files. The preferred method for doing such an operation is to us ArcCatalog not Windows Explorer
- When viewing shapefiles in ArcCatalog (or any ArcGIS program), you will only see one file representing the shapefile; however, you can use Windows Explorer to view all the files associated with a shapefile
- However, if you do copy a shapefile outside ArcGIS, be sure to copy all the files that make up the shapefile

What Have We Learned?

- The importance of GIS data and file management
- GIS data and file organization
  - General and Project-based GIS data
  - Naming convention
  - Create and update Metadata
  - Referencing data – full paths and Relative paths
  - Shapefile and its component files

Any questions or comments?

Please contact me at:
Richard Lau
GIS Project Coordinator
Community Social Profile Project
Email: richard.lau@spno.ca
Phone: 905-632-1975, 878-0955
Fax: 905-632-0778

Thank You