Research Data Workshop: infrastructural URBANISM

Jennifer Moore
GIS & Data Projects Manager
& Anthropology Librarian
j.moore@wustl.edu
314.935.5492
Background Research

- Reference – Encyclopedia of Urbanism
- Articles – Web of Science & G Scholar
- Urban Land Studies Development Case Studies
Finding and Assessing Data

- What data is...
- Assessment of metadata
  - Consider the source
  - Consider the age
  - Consider the lineage
  - Does it have a code book?
- Assessment in practice?
  - Does it surprise you; ask important questions of the data.
Part 1: Tabular Data

- Census Bureau
- ICPSR
- Bureau of Transport
- Re3data.org
Part 2: Geospatial Data

- The WUSTL campus GIS server: \`\`\maps.wustl.edu\gispublic\`
- MSDIS
- STL Open Data
- GeoFabrik
How GIS works

GIS stores information about the world as a collection of layers that can be linked together by a common locational component such as latitude and longitude, a postal zip code, census tract name, road name. These references allow you to identify features on the earth's surface for analysis of patterns, relationships & trends. Using GIS analysis you can explore evolution and/or make predictions.
Spatial Data Types

Vector Data

- Rivers, trees, buildings
- Have attributes
- Point, line or area format
- Values exist at discrete locations
Spatial Data Types

Features

Attributes

<table>
<thead>
<tr>
<th>UID</th>
<th>Name</th>
<th>Type</th>
<th>County</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>Oak Tree</td>
<td>Natural</td>
<td>St. Louis</td>
<td>MO</td>
</tr>
<tr>
<td>002</td>
<td>Bus Stop</td>
<td>Transport</td>
<td>St. Charles</td>
<td>MO</td>
</tr>
<tr>
<td>003</td>
<td>Fire Hydrant</td>
<td>Emergency</td>
<td>Jefferson</td>
<td>MO</td>
</tr>
</tbody>
</table>
Spatial Data Types

Vector Data cont.

More than one feature in a set

= Feature class

= Dataset

= Layer

= Shapefile
You can only have one geometry type per feature class; Geometries don’t mix in a dataset, but you can layer the sets in your project.

points  or  lines  or  polygons
Vector* File Formats

Shapefiles (.shp)
- Vector data only
- Contain one feature class
- Commonly shared
- Made up of several files
- Have to package them

Geodatabases (.gdb)
- Spatial database
- Container
- Good for analysis
- Proprietary
- Can contain
  - Multiple feature classes
  - Tables
  - Raster data
  - Relationships
Many kind of raster formats, very common are TIFF

- Tagged Image File Format – TIFF supports black-and-white, grayscale, pseudo color, and true color images, all of which can be stored in a compressed or decompressed format.

- Other supported raster formats can be found here
Shapefile Anatomy

- **shp** -- Main file (mandatory); a direct access, variable-record-length file in which each record describes a shape with a list of its vertices.
- **shx** -- Index file (mandatory). In the index file, each record contains the offset of the corresponding main file record from the beginning of the main file. The index file (.shx) contains a 100-byte header followed by 8-byte, fixed-length records.
- **dbf** -- dBASE Table file (mandatory); a constrained form DBF that contains feature attributes with one record per feature. The one-to-one relationship between geometry and attributes is based on record number. Attribute records in the dBase file must be in the same order as records in the main file.
- **sbn** -- Part 1 of spatial index for read-write instances of the Shapefile format. If present, essential for correct processing.
- **sbx** -- Part 2 of spatial index for read-write instances of the Shapefile format. If present, essential for correct processing.
- **prj** -- Projections Definition file; stores coordinate system information.
- **xml** -- contains metadata, as used by ArcGIS.
Spatial Data Types

Raster Data

- Continuous data such as temperature measurements or satellite imagery
- Values exist over an area
- Every location has a value
**Raster = Grid**

**Pixel**

Abbreviation for PICTURE ELEMENT, which is the smallest unit in an image. In raster based GIS systems, attribute information can be assigned to each pixel.

The **bounding box** defines the geographic extent of the grid in terms of its coordinates

\[ \text{[min}_x, \text{max}_x, \text{min}_y, \text{max}_y] \]

---

<table>
<thead>
<tr>
<th>2</th>
<th>4</th>
<th>5</th>
<th>16</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
<td>11</td>
<td>14</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>13</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>8</td>
<td>12</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>13</td>
</tr>
</tbody>
</table>

Matrix of Equal-Area Cells

The bounding box defines the geographic extent of the grid in terms of its coordinates

\[ \text{[min}_x, \text{max}_x, \text{min}_y, \text{max}_y] \]
Raster = Grid

Matrix of Equal-Area Cells

Low | High

The bounding box defines the geographic extent of the grid in terms of its coordinates

[min_x, max_x, min_y, max_y]

Pixel

Abbreviation for PICTURE ELEMENT, which is the smallest unit in an image. In raster based GIS systems, attribute information can be assigned to each pixel.
Many kind of raster formats, very common are TIFF

- **Tagged Image File Format** – TIFF supports black-and-white, grayscale, pseudo color, and true color images, all of which can be stored in a compressed or decompressed format.

- **Other supported raster formats** can be found [here](#).
ArcCatalog Basics

- Connect to Folder
- GIS Servers
- Campus Data: \maps.wustl.edu\gispublic
- Preview and Metadata tabs
- Types of Data (points, lines, polygons, raster)
- View data in Windows Explorer
Data Management

We will:

- Create a file hierarchy
  - Not too many layers deep
- Develop consistent naming conventions
  - File names don’t have spaces
  - File names don’t have special characters
  - File names should be meaningful, but brief
Find Data and other Resources
libguides.wusdl.edu\infraURB_dataWS
Questions?

Jennifer Moore
GIS & Data Projects Manager
& Anthropology Librarian
j.moore@wustl.edu
314.935.5492