Preventing Misuse of GIS Data and Products
Preventing Misuse of GIS Data and Products

Goals & Objectives

Review and discuss the following –

- History behind this effort
- Minimum metadata standards
- Standard disclaimers for data and products
- Guidelines for classifying data and products
- Case studies on implementing the above
- Cooperative outreach efforts
History

- Started in 2002 by the “Group of 12” consisting of 6 Surveyors and 6 GIS professionals.
  - Spurred by what was occurring in other states at that time in terms of licensing GIS professionals and implementing GIS data standards.
- The group expanded quickly over time under the umbrella of APLS and became the Geospatial Committee and eventually the Geospatial Chapter.
  - Focused on what differentiated the practice of surveying from non-survey mapping and how the two professions could work together to develop guidelines and potential policy for the governance of GIS data and products.
  - Evaluated the licensing of GIS professionals.
History (cont.)

- Arizona Spatial Data Accuracy and Georeferencing Standards.
  - First guideline published from this effort.
- Arizona Geographic Information Council (AGIC) involvement.
  - Expanded feedback from the GIS community.
- Arizona Board of Technical Registration (BTR).
  - Established an on-going dialog with the Legislative & Rules Committee.
- Dialog continued and feedback gathered between APLS, AGIC, and the BTR via committee/chapter meetings and both the APLS and AGIC annual conferences.
- The cross-education that has occurred during this time has been instrumental in achieving a better understanding of the professions and has driven the effort to tangible and accomplishable proposals.
Preventing Misuse of GIS Data and Products

History (cont.)

• The resulting conclusions on GIS professional licensure were –
  o It would be difficult given the breadth of the industry.
  o It would not necessarily alter data practices.
• The resulting recommended actions were –
  o Develop minimum metadata standards for GIS data and products that provide quality indicators to the end-user.
  o Develop a standard disclaimer for GIS data and products as an initial heads-up to the data quality.
  o Develop guidelines that assist in classifying the quality of GIS data and provide recommendations for preventing the misuse of GIS data and products.
Minimum Metadata Standards

• The adoption of standards for assessing and reporting quality indicators of GIS data would provide a tool that could effectively inform and guide data end-users in the proper use of a given dataset and/or its derivative products.
• Metadata standards for geospatial attribute data would be included.
• Key data quality indicators include:
  • Intended use of the data.
  • Data creation methodologies.
  • Known errors and qualifications.
Minimum Metadata Fields

- **Descriptive name/description/abstract**: General overview of what the data set encompasses.
- **Purpose of the data and intended uses**: What purpose and use was the data created for? (key data quality indicator)
- **Known errors and qualifications**: Indicator of usability and constraints. (key data quality indicator)
- **Dates**: Creation, update, maintenance frequency. Temporal information.
- **Contacts**: General and/or maintenance. References for acquiring additional information.
- **Attribute information**: Field metadata, for key fields at a minimum. Field metadata includes an explanation of each field’s content, and domain of values if applicable, especially for encoded values.
- **Limitations of use and/or distribution**: Augments the intended use with specific use and/or distribution restrictions.
Minimum Metadata Fields (cont.)

- **Feature type**: Geometry type, e.g. point, line, polygon, raster, etc.
- **File name**: Name of the digital file(s), as delivered.
- **Projection/coordinate system**: Georeferencing information.
- **Data creation methodologies, processing and oversight**: Including equipment used and whether or not an RLS was involved. How was the data created? Once the data was created, was it post-processed in any way that could affect its quality? (key data quality indicator)
- **Completeness**: Geometry AND attribute status in terms of what is missing.
Minimum Metadata Fields (cont.)

- Example – GIS parcel layer:
  - **Descriptive name/description/abstract**: GIS parcel boundaries – Pima County.
  - **Purpose of the data and intended uses**: This GIS layer was created for general reference purposes, along with providing a graphical interface into parcel data that is maintained by the Assessor’s Office.
  - **Known errors and qualifications**: This layer contains duplicate taxcodes for parcels, e.g. undeveloped subdivisions in which all lots are assigned the same taxcode, until the lots are developed. Further, a relatively small number of other parcels have more than one polygon per parcel resulting in duplicate taxcodes. Subdivision common areas are typical examples. Parcels with duplicate taxcodes in parcels resulting in duplicate parcel valuation data. Summing valuations for these duplicated taxcode parcels from regions will result in inflated totals. Users should normalize their analysis results based on the frequency of duplicate taxcodes.
Minimum Metadata Fields (cont.)

- Example – GIS parcel layer:
  - **Dates:**
    - Creation: 01/07/1997
    - Maintenance: Updated daily
  - **Contacts:** Steve Whitney, steve.whitney@pima.gov, (520) 724-6729.
  - **Attribute information:** Field metadata -

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Description</th>
<th>Domain</th>
<th>Enumerated Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACRES</td>
<td>Area in acres</td>
<td>Calculated</td>
<td>N/A</td>
</tr>
<tr>
<td>PARCEL</td>
<td>Parcel tax code</td>
<td>Range</td>
<td>N/A</td>
</tr>
<tr>
<td>PARCEL_USE</td>
<td>ADOR parcel use code</td>
<td>Enumerated</td>
<td>ADOR Property Use Code Manual</td>
</tr>
<tr>
<td>TOTALFCV</td>
<td>Assessed total full cash value</td>
<td>Freeform</td>
<td>N/A</td>
</tr>
<tr>
<td>LANDMEAS</td>
<td>Total number of land units</td>
<td>Freeform</td>
<td>N/A</td>
</tr>
</tbody>
</table>
| LANDUNIT      | Type of land unit            | Enumerated  | A=acres  
                             |                            |                          | F= sq.ft.  
                             |                            |                          | S=site                  |

The attribute data for the layer primarily is supplied from the Assessor’s parcel data, along with some GIS overlays for other fields.
Minimum Metadata Fields (cont.)

Example – GIS parcel layer:
- **Limitations of use and/or distribution:** This layer should not be used for determining legal boundary locations, surveying, engineering, or uses other than general reference. This layer is not to be distributed for commercial use.
- **Feature type:** polygon
- **File name:** paregion
- **Projection/coordinate system:** NAD83-92 (HARN), State Plane, Arizona Central Zone (FIPS Zone 0202), International Feet.
- **Data creation methodologies, processing and oversight:** The majority of the parcel boundaries were tablet digitized from Subdivision Plats and Assessor Record Maps. In 2007 the parcel boundaries in the metropolitan area were rubber sheeted to digital orthophotography. Over time as new subdivision data has been acquired in digital format from development firms, this higher accuracy data has been incorporated into this layer and existing boundaries have been adjusted to this new data. Those involved in the creation of this GIS layer included GIS Managers, Analysts, and Technicians. GIS Technicians maintain the data.
Minimum Metadata Fields (cont.)

- Example – GIS parcel layer:
  - **Completeness**: The GIS parcel layer is always incomplete in that there is lag time between the time that parcel data is recorded and the time that it takes to make its way through the recordation and assessment processes then into the GIS layer. The attribute data is complete, although it may not be totally up-to-date due to lag times the parceling process. The parcel geometry and Assessor’s data is synchronized once a year at the close of the tax roll, producing a one-to-one match between parcel geometry and Assessor attributes.
Preventing Misuse of GIS Data and Products

Standard Disclaimer

- The goal of adopting a standard disclaimer for non-survey GIS data and products is to give end-users an initial heads-up that they are using/viewing data that is not survey quality and that it is not intended to be used for the determination of authoritative and/or legal locations.
- Regardless of whether or not a disclaimer will stand up in the courts, this is an act of due diligence to notify the end-user that they should further research the quality of the data/product that they are using.
- This disclaimer is not meant to be an all-encompassing statement regarding the quality of the data/product, rather it is a simplified statement that can be printed on a map, provided with data, an intro to a web map, etc.
Preventing Misuse of GIS Data and Products

**Standard Disclaimer**

- Any type of boundary, linear or point locations contained within this data or displayed within this product are approximate, and cannot be used for authoritative or legal location purposes. Users should independently research, investigate and verify all information to determine if the quality is appropriate for their intended purpose. Legally-defensible property boundaries can only be established by a state-registered professional land surveyor. A list of Arizona Registered Land Surveyors is available at www.azbtr.gov.
Preventing Misuse of GIS Data and Products

Classifying Data and Products

- GIS data and products are in most cases different from survey data and products. Although GIS can be used to store and process survey data, GIS predominately uses non-survey data and therefore a distinction should be conveyed to end-users.
- Preface GIS data and/or products with “GIS”, e.g. GIS parcel boundaries, GIS street network, etc.
- "GIS" data should automatically be considered as representative and not positionally authoritative or legally enforceable.
- GIS data is not considered to be an authoritative survey product for positions or legal boundaries except when prepared by or under the supervision of a professional surveyor and sealed by a registrant.
Classifying Data and Products (cont.)

• If horizontal and/or vertical accuracies are to be stated, or if the data is to be used as an authoritative record, then the data must be certified by a registrant and sufficient substantiation of certification provided.
• Non-certified boundaries should not be displayed above a certain scale based on a hierarchy of scale levels, or the linework is fattened to prevent fine-resolution location determination.
• Non-certified boundaries should not be displayed with imagery based on similar criteria as above.
• Set minimum data quality requirements for regulatory use.
Classifying Data and Products (cont.)

• Provide more than a disclaimer to the general public, e.g. couple it with metadata.
• Internal use vs. external use – Think about how the business community and/or general public will use the data. Is making it externally available really providing a benefit to the end-user?
• Turning off parcel boundaries or imagery at a specific scale, while continuing to provide access to needed data.  
  o Is this dependent on whether or not imagery is displayed?
  o Without imagery displayed, is misuse still possible? (e.g. inappropriate measuring)
  o A scale of 1:2400 may be a good threshold.
Case Studies

- GIS boundary misinterpretation.
- Inappropriate measuring.
- Mixing GIS data sources.
- GIS attribute data.
- Cartographic factors.
- GIS data for land development regulation.
- Analysis results.
- Projections and coordinate systems.
- GIS data visibility control.
Preventing Misuse of GIS Data and Products

Boundary misinterpretation

- Especially when displayed on top of imagery

- How good is the data?
- How good is the imagery?
- Mixing scales?
- Different projections/coordinate systems?

- Assessor Record Map
- Subdivision Plat Map
Preventing Misuse of GIS Data and Products

Inappropriate Measuring

- Inappropriate data quality.
- Scale is too small for the intended purpose.
- Interpreted view, e.g. oblique, integrated into applications, street-level, etc.
- Understand the quality of the source data to know if you are in the “ball-park” or better.

- Oblique image
- ARAN van image
Preventing Misuse of GIS Data and Products

Mixing Data Sources

- Census Blocks on Parcels
- Parcels on NAIP imagery (+/- 15-feet)
- Miss-registered layers
Preventing Misuse of GIS Data and Products

Mixing Data Sources

- Building roof lines, as captured from aerial imagery
- Planned features

- Spatial location differences between imagery
- Different image specifications
- Different image years
- Why so much imagery?
Preventing Misuse of GIS Data and Products

Mixing Data Sources

- Boundaries over orthogonal vs. oblique imagery
Preventing Misuse of GIS Data and Products

Attribute Data Misinterpretation

- Data Normalization

<table>
<thead>
<tr>
<th>AREA</th>
<th>PARCEL</th>
<th>LANDFCV</th>
<th>IMPFCV</th>
<th>TOTALFCV</th>
</tr>
</thead>
<tbody>
<tr>
<td>245555</td>
<td>20518018C</td>
<td>32580</td>
<td>17582</td>
<td>50162</td>
</tr>
<tr>
<td>721363</td>
<td>20518018A</td>
<td>33663</td>
<td>1551</td>
<td>35214</td>
</tr>
<tr>
<td>1281323</td>
<td>20518011A</td>
<td>56620</td>
<td>65101</td>
<td>121721</td>
</tr>
<tr>
<td>64349</td>
<td>20518011A</td>
<td>56620</td>
<td>65101</td>
<td>121721</td>
</tr>
<tr>
<td>33446</td>
<td>20518011A</td>
<td>56620</td>
<td>65101</td>
<td>121721</td>
</tr>
<tr>
<td>654160</td>
<td>20518018B</td>
<td>33012</td>
<td>2703</td>
<td>35715</td>
</tr>
<tr>
<td>6546707</td>
<td>20516001A</td>
<td>56000</td>
<td>0</td>
<td>56000</td>
</tr>
</tbody>
</table>

- Divide all aggregated values by the frequency

Know your table relationships:
- One-to-one
- One-to-many
- Many-to-many
- Many-to-one

- Data View Settings

<table>
<thead>
<tr>
<th>TOTAL_POP</th>
<th>POP_DENS</th>
<th>AREA_SQFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2823</td>
<td>0.000020560103</td>
<td>13730475.77900000000</td>
</tr>
<tr>
<td>5700</td>
<td>0.00003385918</td>
<td>168344309.65200000000</td>
</tr>
<tr>
<td>4733</td>
<td>0.00003665918</td>
<td>129108194.50100000000</td>
</tr>
<tr>
<td>1752</td>
<td>0.00005217327</td>
<td>33580412.85220000000</td>
</tr>
<tr>
<td>3361</td>
<td>0.00003769203</td>
<td>89170054.78820000000</td>
</tr>
<tr>
<td>5098</td>
<td>0.00008569848</td>
<td>72498368.61260000000</td>
</tr>
<tr>
<td>3728</td>
<td>0.00009063870</td>
<td>36993028.01510000000</td>
</tr>
<tr>
<td>3353</td>
<td>0.00003769203</td>
<td>56000000.00000000000</td>
</tr>
<tr>
<td>3400</td>
<td>0.00009063870</td>
<td>358585121.99100000000</td>
</tr>
<tr>
<td>3599</td>
<td>0.00009063870</td>
<td>27826360419.30000000000</td>
</tr>
<tr>
<td>3728</td>
<td>0.00009063870</td>
<td>39832048.43930000000</td>
</tr>
</tbody>
</table>

- Data value is stored correctly, but the display column width is too small to show the entire value
Preventing Misuse of GIS Data and Products

Cartographic Factors

- Colors that are too alike (and patterns)
- Too many layers displayed together
- Colorblindness
- Grayscale
- Black & white
Preventing Misuse of GIS Data and Products

Inappropriate Use for Regulatory Applications

- Floodplains?
- Development/building restrictions
  - Zoning
  - Overlay zones, e.g. hillside, riparian, etc.
  - Building setbacks
  - Density
- In/out/maybe zones of confidence

- FEMA floodplain changes

- Zones of confidence
Preventing Misuse of GIS Data and Products

Bad Analysis Practices

• Overstating the accuracy of the results
  • Data accuracy/quality
    • Vector
    • Raster
    • Elevation
  • Area proportion overlays
  • Mixed scales
  • Attribute accuracy
• Ignoring topological errors

Internal Use Only Data, That Gets Loose

• Critical Infrastructure and other DHS classified data
• Developmental/incomplete data sets
• Non-public data, e.g. health, public safety, etc.
Preventing Misuse of GIS Data and Products

Projections and Coordinate Systems

- Metadata
- Projecting data
  - Use correct Geographic Transformations
  - Datum changes add complexity
  - Projections/coordinate systems have a sweet spot or zone of “best” accuracy

Three Map Projections Centered at 39 N and 96 W

- Mercator
- Lambert Conformal Conic
- Un-Projected Latitude and Longitude

Sinusoidal 8pt Projection Map
Preventing Misuse of GIS Data and Products

Projections and Coordinate Systems

State Plane (AZ Zone C) “linear distortion” (parts per million)

1 ft/mile
= 189 ppm

1 ppm
= 0.005 ft/mi
= 1 mm/km
Preventing Misuse of GIS Data and Products

Layer Visibility Based on Scale

- Turning off parcel boundaries or imagery at a specific scale, while continuing to provide access to needed data.
Outreach and Education

• Education could be considered the most important component of this effort.
• The relationship that has formed between the survey and other geospatial professions has proved extremely beneficial to all by sharing valuable knowledge and best practices.
• Expand these relationships to other professionals and their groups, such as government public counter personnel.
• This knowledge gets shared further amongst professionals and the public.
• Focus on college students as well.
Conclusions

- Adopt minimum metadata standards.
- Classify GIS data and/or products as “GIS” based.
- Adopt a standard disclaimer for GIS data and/or products.
- Adopt guidelines for preventing misuse of GIS data and/or products.
- Provide educational guidance and materials to both data/product creators and end-users, including students.
Questions?

Contact Info
Steve Whitney
GIS Manager – Geographic Data Management and Mapping Applications
Pima County Information Technology Department
steve.whitney@pima.gov
(520) 724-6729