Validating Your Geospatial Data: Protect Your Investment and Yourself!

Author: Rodrick M. Erickson, Michael Baker Jr., Inc.

I. Abstract

The use of geospatial data in the decision making process continues to grow exponentially across government and private organizations. Large investments in technology and data development are made to achieve visual situational awareness for the decision-maker. But what are the implications of the types of decisions being made utilizing geospatial data? If challenged, will the data be legally sufficient? Is your investment worthy of the potential challenges it could face? This paper will present measures you can take to validate your data, set accuracy standards, and protect your investment and credibility in the process.

II. Introduction

Imagine the surprise for the City of Virginia Beach residents and Navy officials when reviewing a map of aircraft flight paths and discovering that two schools and a museum were in the crash zone for a newly proposed aircraft, the F/A 18 Super Hornet, at the city's neighboring Naval Air Station Oceana. The mayor stated that this is not the same map that the City has been using to make decisions about supporting the location of the proposed new aircraft. This was the result of one GIS layer depicting crash zones on two different base maps while evaluating potential impacts of the new aircraft on the surrounding community. As it turned out, the crash zones layer, also known as Accident Potential Zones (APZ), were spatially inaccurate. Initial maps showed the schools and museums within meters outside the APZ. In this case, the discrepancy was brought to the attention of decision makers during the early stages of impact analysis and was quickly resolved between the City and the Naval Air Station. While there may have been some embarrassment for Navy Officials and initial friction, envision this error being discovered during the aftermath of a crash investigation where lives were lost or put in harms way.

This all goes back to practices and standards we have all been made aware of in the GIS profession, but all to often take for granted or simply disregard in our haste to implement and run a GIS. The initial investments in GIS and data development are so great that there is often pressure to produce with immediate results of more efficient processes and powerful, visually enhanced, decision making tools. Whatever the case, there is always the focus on the product and output of a GIS, but not always enough on the content. Ensuring accuracy, completeness, and documentation of your geospatial data can provide an even greater return on investment as well as fortify confidence in GIS as a trusted information resource that will endure challenges and, if sustained properly, time.

For the purpose of this white paper accuracy of geospatial data is discussed in terms of spatial or positional accuracy for the features mapped and completeness or content accuracy of the attributes. In this paper, the case will be made for establishing or following existing standards, and existing National standards and validation procedures will be referenced.

A. Spatial Accuracy

Spatial accuracy is defined as the positional accuracy of features in the geospatial data as measured horizontally and vertically. Spatial accuracy tells us that when we see a feature on the map, there is known probable location within certain limits. For example, if stated accuracy for geospatial data is +/- 1 Foot, the true location of that feature on the ground will be within 1 foot of the location given on the map. The same applies to vertical accuracy. If elevation data were stated to be +/- 5 feet accuracy, then the elevation of a feature on the map will be within 5 feet of the true vertical position on the ground. National standards require stated accuracy to be within a 95% confidence level, meaning that of the points tested, 95% fell within the stated accuracy.

B. Content Accuracy

1. Attributes

Content accuracy can be defined and measured in a number of ways. In this case, we are concerned with the content or accuracy of attributes of the geospatial data and how the data was developed. For measuring content accuracy, evaluation of the attribute data is conducted and a simple percentage of error can be calculated. For example, if 5% of the records in your GIS attributes are found to contain errors, the data could be stated to have 95% accuracy. Depending on requirements or standards set, the process may be more complex. Requirements may call for individual assessment of each field in the attributes where accuracy will vary across fields depending on results. In this case, the average percentage of accuracy for all fields could be calculated.

2. Lineage

Knowing where the data came from is important. How was it developed? What processes were used to create the layer? How old is the data? What source scale was used? These factors should sound familiar to those of us familiar Federal Geographic Data Committee (FGDC) Metadata Standards. Placing a quantitative metric on lineage of GIS is not as easy. A scoring method could be devised weighing factors that will affect the decision to use or not use data based on accuracy requirements. This could be as simple as setting a scale of 1 to 10 in for scoring each criterion considered important and then calculating the average score.

III. Making the Case for Accuracy & Validation

GIS is an influential tool that serves as a valuable information resource to decision makers across multiple disciplines. Private businesses use it to measure economic and demographic conditions that will influence financial and investment decisions. Local governments use it to regulate land uses. Federal decision makers use it to implement policy that affects private property owners. In all instances, the stakes are high and the geospatial data in use can and will be challenged. Any discrepancy found that is not already stated in accuracy statements or disclaimers will erode the integrity of the data and lose the confidence of its intended users and audience. The following examples are just a few of the uses for geospatial data in the decision-making environment where accuracy is a crucial element.

A. Zoning

GIS is a common tool found in all levels of government involved with land use regulation. The District of Columbia Office of Zoning has been utilizing GIS for almost four years now. The organization has successfully migrated all of its original hard copy zoning maps into GIS layers. This enables their staff to research zoning for properties more efficiently and better serve a never ending flow of inquiries about zoning for DC residents and property owners. The office has even implemented a map publishing tool that generates maps in the portable document format for printing and web distribution.

Still, with all this automation and success happening as a result of migrating to GIS there are two things they cannot do. One, they cannot refer to the updated zoning maps published from the GIS data as the Official and Legal Zoning designation map for the District of Columbia. Two, the Zoning Commission cannot use the maps generated by the GIS to make decisions in zoning cases. The reasons for these limitations stem from the fact that if that map published by the GIS were stamped with the “Official Map” designation or if the Zoning Commission were to make a decision in a zoning case based on one of those maps, the data itself becomes subject to challenge. The legal implications of using this data that has not yet been validated for accuracy are risks that the Office of Zoning is not yet ready to take on.

At no fault of the Office of Zoning, the zoning boundaries in the GIS do not have current, real property boundaries to snap zoning boundary lines to and will not for at least another year. In addition, errors and inconsistencies are occasionally found in the GIS layers after extensive research on a case-by-case basis. As a result, they are forced to refer back to the 1996, official version of the hard copy map and then go through a research process of archived zoning orders from 1996 to present to identify changes that have occurred since that version of the map. Fortunately, The GIS is currently used to automate this research process and has improved efficiency in that aspect.

B. Measurement & Analysis

The case discussed in the introduction of this paper involving Virginia Beach and Naval Air

Station Oceana is an instance where geospatial data was used to measure the impacts of an action that could affect the general health, safety, and welfare of the public. The potential Crash Zones or Accident Potential Zones (APZ) for the new aircraft were mapped because, under Federal Law, the military must retain control of property within the Clear Zones and any APZs that extend beyond base boundaries must be communicated to the adjacent property owners and public.

To map APZs for aircraft, a formula has been developed. All that is needed to map these zones accurately are end of runway ground control points with and compass heading or angle of runway direction. From there the width and length of the APZs are developed following a standard model backed by substantial study and research.

Given the definition of APZs and how they are developed, now consider the fact that if the ground control point at the end of the runway is off, the APZ itself will be off. The simple error of +/- one meter can place a school or museum in or out of the APZ and lead to a potentially hazardous situation. Such an error can be eliminated or brought within an acceptable level by testing and validating the accuracy of the APZ data before releasing it to the public.

C. Natural & Cultural Resources Management

In the field of natural and cultural resource management, GIS serves as an effective tool for tracking and managing sensitive resources. At a military installation in Florida, GIS was used to track the location and population of Red Cockaded Woodpeckers. This helped the installation to comply with Federal DoD requirements to manage and protect Threatened and Endangered (T & E) Species as well as register and track the existence of T & E species with State authorities. In fact, the installation actually received build credits for the large number of species maintained, permitting them an allowance to actually build in such critical habitat areas so long as a certain population was maintained.

The Red Cockaded Woodpeckers were tracked by nesting tree sites that had been surveyed, tagged, and mapped as a point feature in GIS. In a field validation study of this point layer, the sites that were mapped were found to have an average horizontal error of 250 feet! In addition, several of the sites were found to be inactive or no longer used by the Red Cockaded Woodpeckers. As a result this data was not good for identifying the location of T & E species that may be impacted by a project or physical action nor should it be used to report populations of the Red Cockaded Woodpeckers to the state.

Further research on the lineage of this data turned up the fact that this data was surveyed in a crude manner before modern GPS resources were available for more accurate recording of the sites. The data was developed by finding the site, finding the nearest street intersection or known mapped landmark, taking a compass heading, and pacing off the distance to the site. This information was then used to develop a point layer in the GIS. Many of the sites evaluated in the test were found to be inactive because the department tasked with tracking them was not required to actually map the sites and lacked funding or resources to maintain and update the geospatial layer.

Researching the process involved in developing this data layer first would have been enough to determine the potential inaccuracies and possibly prevented further field validation study. Unfortunately this knowledge did not exist in metadata and only came out from one of the original creators after being asked why the data had such poor accuracy. If at all possible, try to evaluate the data on its lineage prior to conducting extensive field research.

IV. Spatial Accuracy Standards

A. National Map Accuracy Standards (NMAS)

Geospatial data, like maps are a representation or model of the Earth's surface. As an abstraction of reality, some level of distortion will occur. Long before the first digital mapping data, cartographic database, or land information system was ever developed; the U.S. Bureau of Budget addressed the issue of distortion or map accuracy as early as 1941. It was then that the National Map Accuracy Standards (NMAS) were established and later revised in 1947 to set standards of accuracy for published maps. The standards were set for horizontal and vertical accuracy based on published map scales. Maps published at scales greater than 1:20,000 required that not more than 10 percent of points tested shall be in error by more than 1/30 of an inch on the map, and 1/50 of an inch on the map for scales of 1:20,000 or smaller. All published mapping that failed to comply with those accuracy standards were not allowed to print any mention of standard accuracy on the map.

In order to state that the map complies with NMAS, the accuracy of the map should be tested but is not required. Unfortunately the standards limit guidance in validation techniques to a single sentence that reads as follows:

" Compare the position of points whose locations or elevations are shown upon it with corresponding positions as determined by surveys of a higher accuracy."

However, at least the need for standards and validation of accuracy for mapping was recognized and addressed. This is important to keep in mind as a significant proportion of geospatial data was developed using published, hard copy maps. If that source map followed the NMAS, at least one can begin to assess the potential accuracy of the resulting geospatial data.

B. American Society for Photogrammetry and Remote Sensing

(ASPRS)

The American Society for Photogrammetry and Remote Sensing formed a Specifications and Standards committee in 1990 to develop ASPRS Accuracy Standards for Large Scale Maps. This formed the basis for a much-needed update of the NMAS, which was last revised in 1947. Using this standard, accuracy is reported at the 95% confidence level meaning that 95% of points tested will be equal to or smaller than the reported accuracy value. For example, a stated accuracy of +/- 4 Meters means that 95% of all well defined test points on the map were found to be within +/- 4 Meters of the true location or elevation of the test point on the ground.

ASPRS standards brought accuracy standards to a higher level for large scale mapping and further clarified the ambiguous definition under NMAS. Still, these standards revolved around published maps. Geospatial data was hardly an entity to be reckoned with at this point.

C. National Standard for Spatial Data Accuracy (NSSDA)

With exponential growth in use and popularity of GIS and geospatial data during the mid to late 1990s, NMAS and ASPRS standards were helpful but not specifically designed for digital geospatial data. Due to the fact that geospatial data can be easily manipulated and reproduced at multiple scales and output formats, a more comprehensive standard was in order. Developed as a committee of the Federal Geographic Data Committee (FGDC) in 1998, the National Standard for Spatial Data Accuracy (NSSDA) was formed to develop a statistical and testing methodology for estimating the positional accuracy of points on maps and in digital geospatial data. NSSDA uses Root Mean Square Error (RMSE) as a metric for positional accuracy of geospatial data. NSSDA defines RMSE as the square root of the average of the set of squared differences between dataset coordinate values and coordinate values from an independent source of higher accuracy for identical points. For full explanation of the NSSDA methodology and formulas to calculate positional Accuracy, see document FGDC-STD-007.3-1998, Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy located at http://www.fgdc.gov/standards/status/sub1_3.html.

NSSDA recommend that at least 20 well defined test points be tested. Well defined test points are points in the geospatial data that can be easily identified on the map and the ground. Using 20 points, a 95% confidence level allows one point out of the 20 to fail the parameters of the set accuracy specifications. This technique also calls for use of an independent source of higher accuracy for the geospatial data to be tested by. NSSDA states that the independent sources should be of highest accuracy feasible and practicable to evaluate the accuracy of the dataset. This independent source can vary depending on what is available and what level of accuracy is required.

V. Content Standards

There is a profound imbalance between existing policy and standards regarding spatial accuracy standards and those regarding content standards for geospatial data. Issues and concerns over spatial accuracy migrated and evolved to meet the needs for geospatial data. However, with GIS, content is different, consisting of the geospatial data or attributes associated with the mapping features. Standards and policy guidance for the content of geospatial data have come along more slowly. Fortunately, the Federal Geographic Data Committee (FGDC) is addressing the issue of content standards. More widely known in the GIS discipline for metadata standards, FGDC is leading the way on several fronts involving geospatial data quality.

A. FGDC & Spatial Data Transfer Standards (SDTS)

The FGDC is an interagency committee, organized in 1990 under OMB Circular A-16 that promotes the coordinated use, sharing, and dissemination of geospatial data on a national basis. FGDC has established or is currently in the process of establishing geospatial content standards for various types of geospatial data including cadastral data, soils and vegetation data, and digital orthorectified imagery.

With the objectives of interoperability and open sharing of geospatial data, FGDC developed the Spatial Data Transfer Standards (SDTS) to address specific data transfer issues so that geospatial data could be shared across organizations. SDTS includes five measures of geospatial data quality.

1. Positional Accuracy-The degree of horizontal and vertical control in the coordinate system.
2. Attribute Accuracy-The degree of error associated with the way thematic data is categorized.
3. Completeness- The degree to which data is missing and the method of handling missing data.
4. Logical Consistency- The degree to which there may be contradictory relations in the underlying database.
5. Lineage- The degree to which there is a chronological set of similar data developed using modeling and processing methods.

The FGDC SDTS indeed touches on content accuracy but leaves much to interpretation. Evaluating and maintaining content accuracy is common sense. If the time is taken to validate spatial accuracy, content accuracy should be evaluated as well. The feature may be highly accurate in terms of spatial location, but is that same feature what the attributes classify it to be? That also needs to be tested, as users need to be assured that there is some level of accuracy in reporting the “What” as well as the “Where”. Just as spatial distortion in mapping as a representation is inevitable; the potential for errors in geospatial data attributes is also unavoidable. To err is human and humans are the ones entering and maintaining the geospatial attributes.

Another important component of content accuracy is age. Depending on the intended use of the data, age will play a factor in determining its usefulness in meeting the desired objective. How old is the data? When was it last updated? Is there anything newer available? Perhaps it is something that was well maintained for a certain time frame but then is no longer current. This is where metadata, if available comes in to play. Complete, FGDC compliant metadata can answer a lot of these questions as well as tell about the five measures of geospatial data quality listed by SDTS.

VI. Validation as a Proactive Measure

A. Setting and Enforcing Standards

Be proactive and set accuracy standards and validation procedures within your organization. Do not reinvent the wheel. Follow the examples already out there. The references section of this very paper is a great starting point.

Before data is developed, acquired, or even provided by a vendor or contractor, accuracy standards should be set and enforced. If you already have data, reevaluate current practices and resolve accuracy issues with existing geospatial data. Or, at the very least, establish standards from this point forward to at least discontinue the problem and move forward in the right direction.

Setting and enforcing standards can be viewed as a proactive measure because you are preventing potential issues down the road, establishing trust and integrity of geospatial data, and investing in sustainable practices that will help your geospatial data endure the test of time and challenges. Many local government and Federal organizations have begun to implement policies or standards for geospatial data development. Whether it is done internally or externally by contractors, the primary lesson learned here is that it must be implemented as policy that requires compliance. People will not always comply with mere recommendations that have no teeth to enforce.

Speak softly, but carry a big stick. Enforce the policy! If working with a contractor or developing the data in house, incorporate geospatial data validation into the scope of work or project workflow. Require that a report of validation results is included with the deliverable. Also be sure to incorporate steps in the workflow for error resolution. Do it now as opposed to later. Don't accept the stated accuracy without evidence only to find out later, at a time when it is really critical to be accurate, that the data does not meet accuracy requirements.

VII. Auditing Geospatial Data?

Use of GIS and geospatial data has advanced at such a pace, that quality assurance practices to regulate accuracy have been slow to catch up. However, the time is drawing near that several Federal agencies will adopt the practice of auditing geospatial data. The standards and guidelines included in this paper will serve as their guide. Some may balk at the thought of an actual audit being performed on geospatial data, but is it really all that too far fetched?

When considering that GIS and geospatial data is rapidly becoming integrated with other Information Technology and becoming just another form of data or ways of viewing data, the answer is no! We already audit other data resources, so what makes geospatial data so different? A couple of years ago, one might argue that it would be cost prohibitive and require extensive field research using survey grade GPS equipment. How many auditors do you know of that also know how to utilize differential GPS? Technology and availability of GPS equipment has radically improved in recent years. Improvements to the point that it is intuitive enough for even a novice to walk into a local sporting goods store and purchase and begin using GPS.

How real is this? Currently the United States Air Force GeoBase program is working with the Air Force Audit Agency to perform validation studies of geospatial data layers that will be used for decision making in the next round of Base Realignment and Closure. Their proactive measure to reduce the potential for error and prepare for challenges of the geospatial data used has gained the attention of the other services. As the Air Force Audit Agency develops and implements their geospatial auditing process, they are educating the General Accounting Office (GAO) as well as audit organizations within the other branches of the service.

As early as next year, the Air Force Audit Agency will be in the field testing geospatial data and interviewing data stewards about the content of their data. They will be applying the same standards of FGDC, NSSDA, and SDTS covered in this paper to their process. As initiatives such as this tend to gain momentum within the Federal government, the practice is likely to spread beyond DoD. Of course, the Federal government is not the only one out there validating and auditing their geospatial data, the best examples of applying NSSDA techniques found for this paper come from the State of Minnesota Planning and Land Management Information Center.

Their Positional Accuracy Handbook of using NSSDA to Measure and Report Geographic Data Quality provides 5 case studies where NSSDA techniques were actually applied. Geospatial data types involved include planimetric data, elevation data, real property boundaries, and watershed boundaries. In all case studies different independent geospatial data sets with known accuracy were used to compare test data against. In one case, no reliable independent data existed and survey grade GPS equipment was used to collect ground control points. The documentation of the processes and exploitation of lessons learned is invaluable to the GIS user community and has been well received.

Copies of the Positional Accuracy Handbook can be downloaded at: http://www.mnplan.state.mn.

us/ press/accurate.html. For additional information or printed copies of this handbook, contact the Land Management Information Center, 651-297-2488; e-mail gc@mnplan.state.mn.us.

VIII. Conclusion

The term validation is a polite way of saying, “Prove it.” As geospatial data continues to blend into our decision making process and enhancing our results with a visual outcome, the burden of proof weighs heavily. Especially if those decisions have legal implications, affect human health and welfare, or impact a local economy. Testing and documenting the spatial and content accuracy of geospatial data used in the process will only serve to protect the end user and all affected by decisions based on the data. It will also prevent use of geospatial data whose accuracy limitations make it unsuitable for use in decisions that require precision. The standards and references highlighted in this paper are offered as a guide to help you be proactive in protecting yourself and the user from future, costly embarrassments and promote more trust in the use of GIS and geospatial data.

IX. References

1. Accuracy Evaluation of Ikonos Orthorectified and Stereo Imagery Products, National Imagery and Mapping Agency (NIMA), November 2001.

2. Air Installations Compatible Use Zones, Department of Defense (DoD) Instruction 4165.57, November 8, 1977.

3. Beyond Maps, GIS Decision Making in Local Government, John O’Looney, PhD, Environmental Systems Research Institute, Redlands CA, June 2002.

4. CADD/GIS Technology Center Guidelines for Installation Mapping and Geospatial Data, US Army Corps of Engineers, Waterways and Experiment Station, Information Technology Laboratory, July 1999.

5. Geospatial Positional Accuracy Standards Part 3: National Standard for Spatial Data Accuracy (FCDC-STD-007.3-1998), National Spatial Data Infrastructure, Federal Geographic Data Committee, Reston, VA, 1998.

6. Map Use, Reading, Analysis, Interpretation, Third Edition, Phillip and Juliana Muehrcke, JP Publications, Madison WI, 1992.

7. Positional Accuracy Handbook, Minnesota Planning Land Management Information Center, St. Paul MN, October 1999.

8. Schools and Museums in New Crash Zone for NAS Oceana, Bonnita Harris, WVEC.Com-Hampton Roads, http:///www.wvec.com, August 16, 2002.

X. Author Information

Rodrick M. Erickson
GIS Project Manager
Michael Baker Jr., Inc.
711 Mount Holly Place N.E.
Leesburg, VA 20176
Phone: (703)-317-3060
Fax: (703)-960-9125
E-mail: rerickson@mbakercorp.com