Authority and Authoritative Data: A Clarification of Terms and Concepts

by David Stage

The statements made or opinions expressed by authors in **Fair & Equitable** *do not necessarily represent a policy position of the International Association of Assessing Officers.*

This article is intended to clarify the issues surrounding the terms *authority* and *authoritative* as they relate to sources of parcel data and the sharing of those data. Claims are frequently made, implied, or omitted about the appropriateness of applications and data sources that purportedly represent the extent of rights in land, the characterization of those lands, and the status of ownership. These claims have caused a great deal of confusion for the public as well as for professionals in two ways: first in how the source is chosen and the data are used and second in the discourse among cadastral data stakeholders. These issues are interrelated with the terminology used to describe the production and publication of cadastral data, but that terminology is treated only briefly here, since it should be the subject for another article.

With few exceptions, assessment data are developed under the auspices of an assessor's office.

The proliferation of parcel data sets on the Internet is pushing this issue into the public arena, but not because the data looks bad. As a matter of fact, the data that are available look quite good and seem to work well in the applications in which they are being used. In these images, property boundaries and those in the neighborhood appear to be exactly where they are supposed to be and the value does appear to be accurate, so why can't these data be used to address all land ownership needs? Answering this question requires an understanding of a number of issues, ranging from technical and legal to business processes and semantics.

In making assumptions about data, users should recognize the inverse relationship between the distance of the leap-in-faith needed to use the data and the likelihood the user will be disappointed, publicly embarrassed, or even worse. The first question users should ask is whether they can trust the data. Are the data sufficiently accurate to answer the questions the user wants answered? If the user is simply satisfying curiosity, it is fairly easy to give the data the benefit of the doubt as long as there are no glaring errors. On the other hand, if the users are Small Business Administration managers providing disaster victims with low interest loans, then they want the data to be of the highest *quality* to ensure these loans are being awarded to the owners of the property and the value of that property is what the applicant claims it to be.

Data Quality

What is data quality? One definition states that, "data are of high quality if they correctly represent the real-world construct to which they refer." A more technical definition declares quality to be "the state of completeness, validity, consistency, timeliness and accuracy that makes the data appropriate for a specific use" (Wikipedia contributors 2009). Determining an appropriate level of quality requires the user to first define their *real-world construct*, otherwise known as business requirements, and then match those requirements with the source data to determine whether the quality (completeness, validity, consistency, timeliness, and accuracy) of the data meets their specifications. This process leads to determining the best source for the needed data.

Data Sources

What are the different data sources of assessment data for a user? With few exceptions, assessment data are developed under the auspices of an assessor's office. These data can be acquired and packaged in a variety of ways. It is important to characterize and understand the subtle differences in the following terms when data sources are being considered.

- **Authoritative Data. These are o**fficially recognized data that can be certified and provided by an authoritative source.
- **Authoritative Data Source (ADS)**. This is not a source of data, but it is mistakenly used as one when data sources are discussed. ADS is an information technology (IT) term system designers use to identify a system process that ensures the veracity of data sources when a database is created.
- **Authoritative Source**. This is an entity authorized by a legal authority to develop or manage data for a specific business purpose. The data this entity creates are authoritative data.

- **Authority.** In the context of public agencies, authority is the legal responsibility provided by a legislative body to conduct business for the public good.
- **Authorization**. This is the result of an act by a legislative or executive body that declares or identifies an agency or organization as an authoritative source.
- **Certified Data. These d**ata are created for a specific purpose and have been vetted by an authorized entity.
- **Uncertified Data.** Uncertified data are generally a more current copy of the data that has yet to be certified and may also include information used for creating certified data but is not part of the certified data.

The sentence "The integrity and validity of any data set is open to question when there is no traceable connection with a known System of Record" deserves repeating.

- **Data Steward**. This is the organization within an authoritative source charged with the collection and maintenance of authoritative data. The term data steward is often confused with the term authoritative source.
- **Shadow Data.** These are copies of the assessor's data that are packaged and often enhanced or value-added for distribution and resale. The enhancements may consist of reorganizing the structure of the data, putting the data in a reseller's format, and/or adding data from sources to make a new data set. The level of trust that can be placed in the data depends upon how much the user knows about the data and the degree to which the enhancements improve the data or "muddy the waters."
- **Trusted Source and Trusted Data**. This is a service provider or agency that publishes data from a number of authoritative sources. These publications are often compilations and subsets of the data from more than one authoritative source. These data are trusted because there is an official process for compiling the data from authoritative sources and the limitations, currency, and attributes are known and documented.

The term ADS deserves to be elaborated upon for two reasons, first the term is often confused with *authoritative data* and *authoritative source* and second because ADS methodology, which is a best practice of the data warehousing community, should be kept in mind when comparing authoritative data and shadow data.

From a technology perspective, information managers create processes in which an information storage system is the *authoritative data source* and is defined for each data element or piece of information. These systems can be quite complicated, because data sets are being built by outputting data from multiple sources and then re-presenting the results for new business purposes. The problems faced in compiling data from multiple systems are as follows:

> *In these cases, multiple information systems (multiple jurisdictions) may disagree about the same piece of information. These disagreements may stem from semantic differences; differences in the timing of the extraction, transformation and loading (ETL) extracts that create the data they report against; or may simply be the result of bugs.*
>
> *Where the integrity of the data is vital, a data element must either be linked to, or extracted directly from its system of record. The integrity and validity of any data set is open to question when there is no traceable connection with a known System of Record.* (Wikipedia contributors 2009)

The sentence "The integrity and validity of any data set is open to question when there is no traceable connection with a known System of Record" deserves repeating. This statement is true, but there are also degrees of integrity and validity, for example, the degree to which the completeness and accuracy of the data have degraded since the last update, or the degree to which the "waters are muddied" by normalization of the data. These issues are particularly problematic for copies of the original source data, which are also known as *shadow data.*

In summary the term ADS, even though it is a best practice of the IT community, implies more than it able to deliver. ADS is an IT term that describes an IT process in an information storage system (data warehousing) that ensures the integrity and validity of the data being stored, but it does not necessarily mean that it is authoritative data coming from an authoritative source. (IT terms are labels used in *computer-world* constructs, not *real-world* constructs. IT terms often suggest similar real-world functionality but are metaphors at best. These terms should be qualified with the prefix "e-" for *electronic.* Consider the terms *desktop, window,* and *trash.* Their meaning would be more clear if they were *e-desktop, e-window,* and *e-trash.* One term that did manage to hang on to its prefix is *e-mail.*)

Real-World Operations

The primary use of assessment data is the valuation of property for tax purposes. The state legislature usually provides assessors with this authority alone. The expenses incurred in collecting data are justifiable only if they enhance the quality of the information used to assess properties and conduct the businesses of the assessor's office. It is not a mandate of the assessor's office to provide high-quality data for other business processes.

Nevertheless, the utility of assessment data is of great value to many organizations. These data provide the greatest degree of granularity for describing the characteristics of property and, with the advent of geographic information systems (GIS), a good visual description of the parcel boundary. The data in most jurisdictions are continually being updated, so they accurately portray the current status of all properties. The data are often validated by two sources: the assessment office through its own quality control and a state oversight agency.

In short, the product is of a very high quality for its intended purpose, and a subset of information coincides with the business needs of organizations for an accurate portrayal of how the land is being used, who is using it, and what is on

the land. The information requirements of these organizations can be described as highly coincident, and many organizations are looking for ways to acquire regular access to these data.

The primary use of assessment data is the valuation of property for tax purposes.... It is not a mandate of the assessor's office to provide high-quality data for other business processes.

Government-to-Government Data Sharing

Government agencies need high-quality data because they are making decisions that often affect life, property, and the economic viability of communities. Given the option, government agencies would prefer to go to the assessor's office to acquire high-quality authoritative data to base their decisions on. The issues of sharing data and providing benefits directly to the assessor's office, as opposed to benefits to the community as a whole, are complex and should be the subject of another article.

The purpose here is to describe a process, define the components of a data-sharing model, and differentiate *authoritative data* from *shadow data*. The following is the vision of the Federal Geographic Data Committee for the National Spatial Data Infrastructure (NSDI) and the National Cadastre within the NSDI.

> *The Cadastral NDSI will have a single source of authoritative cadastral data within a single geographic extent that is controlled and managed by designated data stewards. Access to this data is facilitated by compiling and integrating the data into trusted data sources at state or regional levels. This will reduce duplication of effort and assure that the best available information is used in decision making.* (FGDC Accessed 1-19-2009)

Figure 1 is an illustration of this model showing the relationship between the data sources, their legal authority for creating the data, and the different types of data—production, publication, and project.

The following describes the components of the data-sharing model.

Application service providers require official sources of data if their applications are to be credible. Thus the source is authorized by a public body such as a federal agency, state legislature, or local government statute or rule. The two types of official data are *authoritative* and *trusted data,* and correspondingly there are *authoritative* and *trusted sources* for these two types of data.

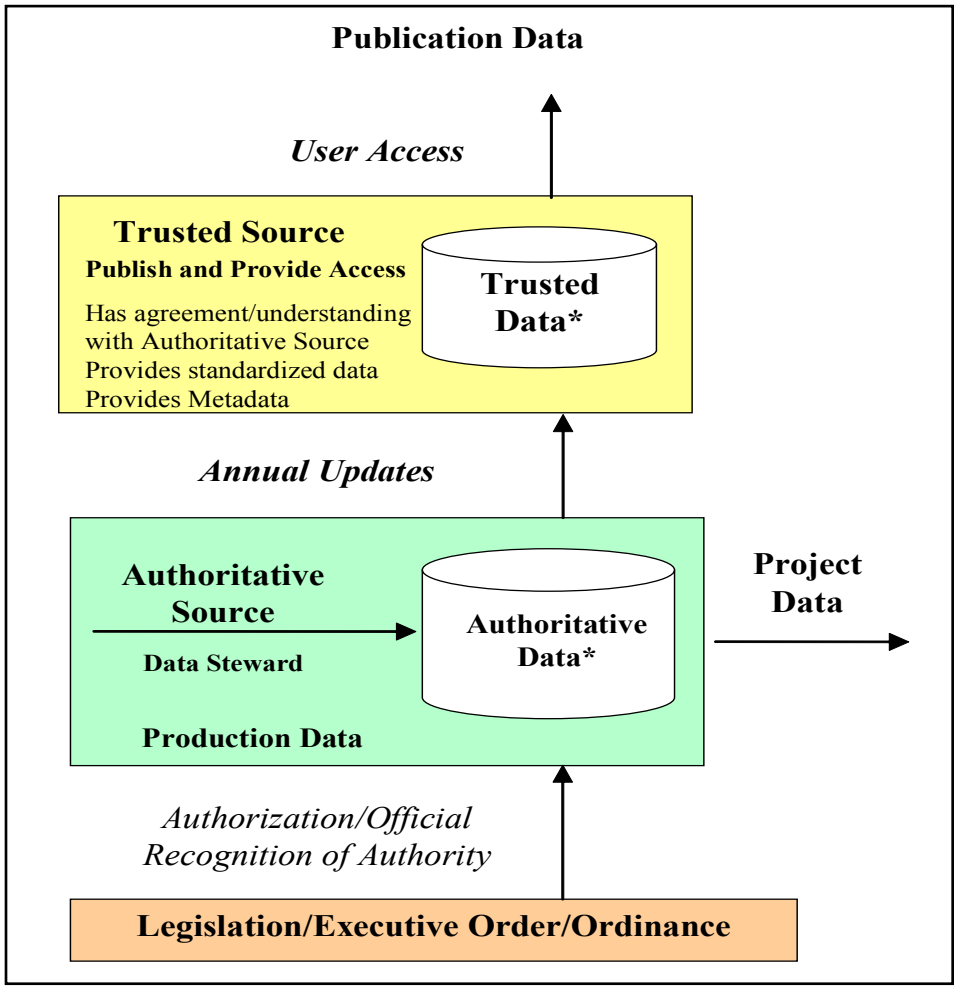
Authoritative Data

Authoritative data come directly from the creator or authoritative source. These data are the most current and accurate and have been vetted according to official rules and policy. The data have a known accuracy and lineage and can be verified and certified by data stewards in the authoritative source.

Trusted Data and Trusted Data Sources

Trusted data and t*rusted data sources* describe a situation in which data sets are published by someone other than the authoritative source and are often the compilation of multiple sources of authoritative data. The data are *trusted* because there is an *official process* for compiling the data from authoritative sources and the limitations, currency, and attributes are known. Metadata are provided; the data are often formatted into a standardized form; and linkages to the originating source are provided with the data. This trusted source is recognized by the authoritative source as an *official* publisher of this subset. Typically a trusted source is established to integrate data from multiple jurisdictions and to compile them into a standard format. These trusted data are adequate, convenient, and cost-effective for users that need a regional view

Figure 1. Illustration of the relationships among the defined terms.



and have to deal with multiple sources of data. However, there is an understanding of the necessity when final decisions are being made, particularly about rights and interests of specific properties, that users must go to an authoritative source and acquire authoritative data directly to ensure that they have the most current and accurate data.

All cadastral data collections that do not come from authoritative or trusted sources are* unofficial *copies whose value degrades over time relative to the rate of updates to the authoritative data.

Data Stewards

Data stewards have the responsibility for organizing, collecting, maintaining, and providing data. Data stewards are those closest to the data creation—they have recognized expertise in the field and follow professional standards. In the realm of cadastral data, the land surveyor is the data steward charged with collecting and interpreting information on the location and geometry of a parcel, and the county tax assessor is the data steward charged with providing the valuation and related property information for the same parcel.

Authoritative Data Source

ADS is a term used by IT systems designers to define a process for ensuring the data integrity in an application. When data are created or edited, data integrity is preserved by providing a single object through which updates can occur. When data are imported, which may be from multiple sources, there is a strict vetting process to ensure the integrity of the imported data. The process of managing the updates of records, from the systems designer's perspective, results in authoritative data. This is authoritative in a database sense but not necessarily in a legal sense.

Metadata

Metadata are an essential part of the process; they allow the user to determine the usability of the data as well as serve as an audit trail on the authority of the data. Trusted data sources and authoritative data sources must provide accurate metadata that describe the lineage, quality, and currency for cadastral data. In addition, the trusted data sources must provide linkages to the ADSs.

Conclusion

Unlike other spatial data (hydrography, topography, orthoimagery, and the like), cadastral data define rights and interests in land, and how they are described and presented to the public can cause considerable confusion about value, ownership, and so on. Cadastral data are also unique because they are created and maintained by more than 4,000 separate entities across the country. Although the original business purpose of cadastral data is at the local government level for assessment purposes, they have become critical to the efficient operations of many state and national business operations because of their level of detail about the land and the currency of that information. Uses vary widely, ranging from emergency response and recovery, to environmental management, to health and safety, to fleet management and more. As a result, there is a high demand for compiled and standardized cadastral information that can be incorporated into a wide variety of applications.

All cadastral data collections that do not come from authoritative or trusted sources are *unofficial* copies or shadow data whose value degrades over time relative to the rate of updates to the authoritative data. Unofficial data are often duplicative and create redundancy because they are re-published data that are already available from a trusted source. Unofficial sources create confusion among the general public by providing un-maintained duplicative data and *unofficial* parcel-like data sets that can unexpectedly harm or damage property rights with inaccurate, out-of-date information.

Recognizing the importance of authoritative sources for authoritative cadastral data provided through a trusted source is essential to (1) protecting individual land rights, (2) supporting local governments and other parcel producers in their authorized role of data stewards, and (3) ensuring that the user community has the best available and most current cadastral information.

References

Wikipedia contributors, "Data quality," *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/w/index.php?title=Data_quality&oldid=255796765 (accessed January 8, 2009).

Wikipedia contributors, "System of record," *Wikipedia, The Free Encyclopedia,* http://en.wikipedia.org/w/index.php?title=System_of_record&oldid=237742899 (accessed January 8, 2009).

FGDC Subcommittee for Cadastral Data, *Cadastral NSDI Reference Document*, p12, http://www.nationalcad.org/showdocs.asp?docid=158&navsrc=Report&navsrc2=

David Stage is a private consultant specializing in information management and institutional and technical infrastructure issues that have to do with the use and sharing of cadastral data He has been the Eastern Cadastral Coordinator for the FGDC Subcommittee for Cadastral Data (Subcommittee) since 2002 working on projects to build a national spatial data infrastructure for cadastral data. Projects and publications cover a wide variety of subjects that include the best practices for state parcel management programs, the acquisition of parcel data in support of emergency response for hurricanes and wildland fire; addressing issues for government to government data sharing of cadastral data between federal, state, local governments and the private sector.

Prior to working for the Subcommittee he had been the State GIS Coordinator for Florida from 1990 to 2001 and the Florida Governor's Liaison for the 2000 Census.

David Stage, dstage10@comcast.net
850-668-2604