

Defining and Providing Guidance for Evaluating Data on AZGeo, Arizona's State Geospatial Data Clearinghouse

Executive Summary

The Arizona Geographic Information Council (AGIC) AZGeo Advisory Committee tasked the AGIC Outreach Committee with developing guidance for AZGeo users and publishers on what constitutes authoritative data. Providing guidance to data creators and users aligns with AZGeo's long term goal of providing discoverable, authoritative data that is fit for use and appropriate for decision making. The intent is to guide users to think critically about the data they are consuming and developing, which is best supported by complete and comprehensive metadata. This document reviews provides guidance on proper data management including metadata standards, criteria for authoritative data tagging, community resources, and much more.

The Problem

Geospatial professionals across jurisdictions and disciplines rely on location based data for meeting organizational business requirements and supporting executive leadership in data-driven decision making. Discovering and identifying the best available source data is essential for analysis and visualization and is often a complex and nuanced process. For example, data can have dual ownership, creating complexity in determining data fitness for a specific purpose. Defining best practices for identifying authoritative data, including criteria for what constitutes authoritative, will ensure that the right data is used to solve problems. This document will provide guidance and best practices for AZGeo users seeking the best available data for their purpose.

Guidance for Data Users

For data to be flagged in AZGeo as 'Authoritative', data creators must request this designation. A request initiates a review process to ensure that all 'Authoritative' data have been thoroughly reviewed, meet industry standards and are evaluated for fitness. This designation is intended to guide data users to locate and distinguish authoritative data from other datasets.

However, because a dataset is considered authoritative does not mean it is the best or most appropriate for all purposes. This section provides guidance on what "authoritative" means and how to evaluate a dataset to determine if it will meet your project needs, whether it is authoritative or not.

What is Authoritative Data?

According to the Federal Geographic Data Committee (FGDC) Subcommittee for Cadastral Data, authoritative data are created by an authoritative source, which is an entity that is authorized by a legal authority to develop or manage data for a specific business purpose. The data from authoritative sources are typically the most current and accurate available and have been vetted according to official rules and policies (FGDC Subcommittee for Cadastral Data, 2008). An example of authoritative data are the Arizona Department of Transportation (ADOT) highways. ADOT is required by federal and state laws to manage and maintain records, including geographic information, for all interstates, U.S. routes, state highways, and other roads owned by the state. This data is rigorously reviewed and must meet standards for publication.

It is possible to have official data that are not authoritative. Data created by “trusted sources” are datasets that follow an official process to combine and engineer one or more authoritative datasets, which make it easier for end-users to consume the data. While these datasets may not be authoritative, they are typically recognized by the authoritative sources as an official publisher (FGDC Subcommittee for Cadastral Data, 2008). An example of trusted data are the ADOT publication of local roads. While ADOT is the authoritative source for all state-owned roads, local government agencies are the authoritative source for locally owned roads. With permission from the local agencies, ADOT does provide a dataset that combines all local roads with state-owned roads, creating a single data source for all Arizona roads and making it easier for end-users to locate and consume the data.

When evaluating a dataset on AZGeo, it is important to distinguish if the data are coming from an authoritative source, a trusted official source, or another source. Authoritative data will be the most accurate as they were created using an official process mandated by law. Trusted sources can be useful if they include the appropriate data for your project. It is advised to carefully read the metadata of trusted data sources to understand how current the data are and how the data were published.

Intended Use

Although a dataset is marked as authoritative, it does not always mean it is appropriate for each project. As stated above, trusted official sources may better meet user needs by combining multiple authoritative sources into a single dataset. However, it’s also important to understand that each dataset was created for an intended purpose. Understanding what that purpose is may influence whether the dataset will fit user requirements.

An example of evaluating intended use would be locating an appropriate ZIP code boundary file. While the USA ZIP Code Areas dataset published to the Esri Living Atlas is marked as authoritative and does provide a nationwide ZIP code file, it may not be the most current or accurate dataset needed for more local studies. Some counties may have ZIP code boundary data that are more appropriate for projects in their areas.

Metadata

Datasets published to AZGeo should include metadata, or data about the data. Information about the dataset, including when it was published, intended use, and any linkages to original source datasets if applicable, are commonly included. Referencing metadata will provide important information to help data users determine if the dataset is appropriate for their project.

Best Practices for Data Publishers

Best practices and considerations for sharing data in the state geospatial data clearinghouse, AZGeo.

- Metadata needs to comply with [AZGeo metadata recommendations](#). This includes proper data tagging and categorization for improved discoverability. Recommendations also include but are not limited to data latency, contact information, collection methods, known data limitations, and much more. The use of robust metadata cannot be understated, it helps to provide context around data. It also is a method to reduce risk related to the proper use of data and allows for team's to be proactive with customer input and questions that will be asked.
- Having clear roles and responsibilities such as data owner, data steward, and data custodian will help a data consumer to understand the level of complexity of who is responsible for evaluating the data, its intended use, and appropriateness for analysis.
 - Data users are responsible to do their own due diligence to determine the best dataset for their needs.
 - Data owner is responsible for the management of the data and for documentation.
 - Data stewards are responsible for transforming the data into information and dissemination.
 - Data custodians ensure that the data is properly stored, secured and backed up.
- Other considerations include, define: agency specific, dissemination vs. authority of data, owner vs. maintainer, shared authority (ARNOLD vs. City of Phoenix). Considerations such as data:
 - Intended Use & Limitations
 - Data Creator - should be documented with general contact information
 - Documented data development process (workflow)
 - Update cycle (i.e. real-time, monthly, annually, etc.)
 - Coordinate system
 - Metadata are voluntary and not enforceable but highly recommended
- AZGeo encourages data creators and custodians to include a disclaimer as part of the data's metadata. This best practice alleviates risks to a member agency if the published data is used for purposes outside of its documented intended uses.

The table below provides some examples of **authoritative data** and why they are **trusted sources**:

Dataset Name	Data Author	Authoritative or Trusted Source?	Comments
Arizona's All Roads Network	ADOT Local Agencies	Both	Due to the Federal reporting requirements, ADOT has a business requirement to collect and aggregate street centerlines annually.
County Boundaries	Census	Both	The Census Bureau coordinates with local, regional and state agencies to validate the data.
Incorporated City Boundaries	ASLD (via ALRIS DOR)	Trusted	ASLD is Arizona's State Certifying Official (SCO) for incorporated boundaries and annexations submitted to the Census Bureau. ASLD coordinates with the Department of Revenue and the state Demographer to ensure the representational datasets are accurately depicted.
Arizona's Address points	State Tribal Local Agencies	Both	Federal requirements governing Next Generation 911 data collection dictate the aggregation of local authoritative data to support emergency response.
Tax Parcels	Local Assessors	Authoritative	
Land Ownership (aka Surface Management Agency or SMA)	BLM ALRIS	Both	

AGIC Guidelines for Identifying Authoritative Data on AZGeo

The proposed guidelines for identifying and categorizing authoritative data on AZGeo are referenced throughout this document. Below are the requirements for requesting their data be identified on AZGeo as the authoritative source. A series of authoritative categories are listed in the glossary of terms and each data provider should specify the level of authority their data should hold.

Below are what is needed to identify data as authoritative and what the processes for identifying data sources as authoritative:

- Data owner must identify the data source as authoritative
- Metadata must be completed to define FGDC standards and community standards for data to be considered authoritative
- Provide documentation for on intended use and data restrictions
- Depends on the AZGeo data curator and validation from authority
- If the data comes from data owners - can be considered authoritative
- If the data doesn't come from data owners - work with AZGeo AC for review and approval
- A high level decision tree needed to support visualizing the process
- Intended use, data author, community standards
- Use AGIC to facilitate the determination of specific data authorities. Improve transparency and isn't one-sided for decision making.

Currently, all data sources with the Authoritative label have been grandfathered in from the previous version of the AZGeo platform. The workflow moving forward for identifying the authoritative data on AZGeo will require data authors to submit requests for the authoritative data tag. The AZGeo Data Curator will review the request and if it meets minimum criteria, will bring it to the AZGeo Advisory committee to determine action on them. ALRS

DecisionTree:

https://drive.google.com/file/d/1Mvo2girYrEXg4Prx4RdvFu3JBbxg_opM/view?usp=sharing

Identifying what levels of authority a data source is has many nuances and is important for the AZ community that there is transparency in the decision making process. The data author (or requestor) will be involved in this discussion.

If the committee determines the authoritative data request does not meet the Authoritative threshold, it may be advised that a different category be added to identify it as a trusted source, data steward, or commonly used. Once the data source has been approved and designated in one of the categories, the data author will be added to the Authoritative AZGeo group where their data can be easily referenced on the open data hub.

It is recommended that the review and implementation of these guidelines will be the responsibility of the AZGeo Advisory Committee with the support of the Council. The Committee will ensure compliance of defining, hosting, and use of authoritative data. This will include:

- Guidelines
- Process for approval of 'authoritative' category
- Criteria for defining authority

At the end of the day, as a data user, it is your responsibility to think critically about which data source to use for each respective piece of analysis. Understanding this will help to reduce risk of using the improper data source for decision making.

Resources - Glossary of Terms

The following section on resources is designed to support data owners and users within the AZGeo community. The use of a business glossary of terms is helpful to ensure that all users are speaking the same language.

- **Authoritative Data** -Officially recognized data that can be certified and is provided by an authoritative source.
- **Authoritative Data Source** – An information technology (IT) term used by system designers to identify a system process that assures the veracity of data sources. These IT processes should be followed by all geospatial data providers. The data may be original, or it may come from one or more external sources all of which are validated for quality and accuracy.
Authoritative Source – An entity that is authorized by a legal authority to develop or manage data for a specific business purpose. The data this entity creates is authoritative data.
- **Authority** – In the context of public agencies it is the legal responsibility provided by a legislative body to conduct business for the public good.
- **Authorization** – The result of an act by a legislative or executive body that declares or identifies an agency or organization as an authoritative source.
- **Data Steward** – An organization within an authoritative source that is charged with the collection and maintenance of authoritative data. The term data steward is often confused with the term authoritative source.
- **Trusted Source and Trusted Data** – 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. It is “trusted” because there is an “official process” for compiling the data from authoritative sources and the limitations, currency and attributes are known and documented.

FGDC Subcommittee for Cadastral Data. (2008, August). *Authority and Authoritative Sources: Clarification of Terms and Concepts for Cadastral Data Version 1.1*. (Summary of Terms: [reference FGDC](#))

Supplemental Documents

- FGDC Standards Document: <https://drive.google.com/file/d/1TxcctHoYsxxTcyZTM6HrRCnhq2qnJuKw/view>
- What is authoritative geospatial data?: <https://geoawesomeness.com/what-is-authoritative-geospatial-data/>
- Open Geospatial Consortium (OGC): <https://www.ogc.org/>
- Authority and Authoritative Sources: Clarification of Terms and Concepts for Cadastral Data: <https://drive.google.com/file/d/1TxcctHoYsxxTcyZTM6HrRCnhq2qnJuKw/view>

Contacts

If you have questions on the content of this document or comments to improve upon it, please reach out to Jenna Leveille at jleveille@azland.gov or Patrick Whiteford at pwhiteford@azdot.gov.

References

FGDC Subcommittee for Cadastral Data. (2008, August 1). *Authority and Authoritative Sources: Clarification of Terms and Concepts for Cadastral Data*. Retrieved June 1, 2023, from https://glorerecords.blm.gov/reference/nationalCad_stagnate_old_reports/Authority_and_Authoritative_Sources_Final.pdf