

AGIC

ARIZONA GEOGRAPHIC
INFORMATION COUNCIL

GUIDELINES FOR GEOSPATIAL DATA SHARING

*FOR USE BY ARIZONA ORGANIZATIONS WHO CREATE, MAINTAIN, AND SHARE
GEOSPATIAL DATA*

DRAFT

TABLE OF CONTENTS

Acknowledgements	3
AGIC Geospatial Data Guidelines – Summary Sheet.....	4
Objectives	4
Methodology.....	4
Objectives.....	5
Reduced Costs.....	5
Improved Productivity.....	5
Enabling Business.....	5
Reduced Risk.....	6
Promoting Cross-Jurisdictional Information Exchange.....	6
Broader Benefits	6
Methodology.....	7
Examining Sensitivity	8
Determining Classification.....	8
Desensitizing Controlled Data.....	8
Suggested Technology	9
Control Your Data’s Attributes.....	9
Disclaimers.....	10
Metadata	12
Open Data for Categorical Sharing	13
Sharing with the General Public.....	13
Sharing Confidential Data Through AZGeo Groups.....	13
Appendices.....	14
A. Arizona Geographic Information Council (AGIC) and AZGeo.....	14
About AZGeo	14
How does AZGeo work?	14
B. A.R.S. § 37-177 & A.R.S §37-178.....	15
Open Data – Why, What, How?	17
Why Open Data? ¹	17
What is Open Data? ²	18

ACKNOWLEDGEMENTS

AGIC would like to thank the Data Committee's Data Sharing Workgroup comprised of the following persons and organizations for their time and expertise in the development of these guidelines.

Kevin Blake	Yavapai County
Tim Colman	Arizona State Land Department
Jenna Leveille	Arizona State Land Department
Phil Ponce	Engineering Mapping Solutions
Curtis Pulford	Arizona State Land Department
Jason Ramsey	Arizona Public Service
Susan Smith	Arizona Forestry and Fire Management
Gene Trobia	Arizona State University
Steve Whitney	Pima County

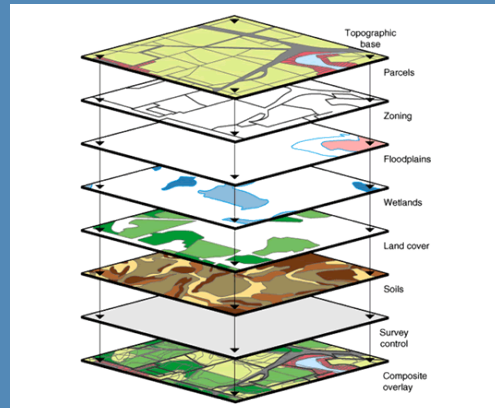
AGIC GEOSPATIAL DATA GUIDELINES – SUMMARY SHEET

Objectives

- Reduce the costs of data development associated with duplicative efforts.
- Improve productivity through one-stop, federated access to relevant data and information.
- Facilitate a self-service data download/access environment
- Eliminate bottlenecks and consumer wait times associated with order fulfillment; freeing up devoted staff time
- Enable entrepreneurial use by businesses and citizens.
- Reduce the use of unauthorized or out of date information.
- Facilitate sharing, integration and standardization of geospatial data and services.

Methodology

- Determine agency custodianship
- Categorize data sets with regard to sensitivity levels
- Encourage standardized disclaimer (AGIC or customized)
- Ensure that metadata is complete
- Share data with the broadest audience possible
- Utilize Arizona’s state geospatial clearinghouse, AZGeo



Geospatial data is data that pertains to geographic objects, features, locations, and boundaries on Earth, and includes information regarding their location, size, and shape. This data also includes additional attribute data that provides more information about the entity that is being represented. Geospatial data is consumed and analyzed by Geographic Information Systems (GIS), which in turn enables the data to be mapped, analyzed, and consumed by geospatial applications across government, the business community, educational institutions, and the general public.

Making geospatial data “open” leads to increased interoperability amongst diverse organizations and provides more opportunities for users to utilize this data. Increased interoperability can enhance the ability for organizations to work together, since no single organization develops all of the geospatial that they need.

This document serves as a best practice guide for Arizona organizations who wish to engage in the sharing of geospatial data.

OBJECTIVES

The benefits to open data sharing are numerous. Open data sharing can positively impact community relations, improve workflows, support decision making, and have long-lasting benefits to an organization and the customers it serves. In today's information age, it is imperative to demonstrate value and efficiency, provide on-demand services, and have more transparent operations.

Demonstrating this with geospatial data has previously been thought to be problematic due to inherent size, complexity, and content. However, technologies are in place that are fully capable of dealing with the worries of the past. Data formatting and complex file types, once the managed domain of only one or two large GIS software vendors, are no longer as daunting and can be transformed with ease. We can, and should, encourage data sharing between engineering, planning, science, academic, and various other entrepreneurial uses.

The objectives in sharing geospatial content can contribute to the following:

Reduced Costs

- Reducing administrative costs in processing data requests
- No need to duplicate the creation of data managed by others
- Partnership in shared acquisitions lower total cost of ownership to participants
- Service delivery, or one-stop data marts, reduce the need for local storage of external data, e.g. by consuming the data through web services

Improved Productivity

- Reduction of time spent fielding data requests from external consumers
- Generally, datasets are more easily discoverable through Open Data portals
- Improvement of data reliability through increased usage; increasing usage results in more feedback to the data custodians, helping improve and/or enhance the data
- Accessing reliable/authoritative sources reduces the need to quality check data

Enabling Business

- Public data access improves business relations with external businesses and organizations

- Encourages innovation with new and different uses of the data
- Enables value added services to be created from base-level data sets (e.g. tourism, planning, analyses, etc.)
 - Third-party applications are built on top of public data
- Economic development benefits from access to the latest data, utilizing spatial overlays to determine best siting, resources, etc.

Reduced Risk

- Reducing the number of duplicative datasets also reduces the use of old or inaccurate data
- Appropriate custodians can share the data for all to use versus subsets used to create ad-hoc derivatives
- Commercial routing and guidance applications would be able to take advantage of the most current/accurate data.

Promoting Cross-Jurisdictional Information Exchange

- Increasing data availability to a wider user-community, across a larger collection of organizations, also increases the ability to see the data within a more regional context
- Streamlining information processing across lines of government becomes easier to achieve

Broader Benefits

- Transparency and democratic control
- Increased participation
- Improved and/or new products and services
- Innovation
- Improved efficiency of government services
- Improved effectiveness of government services
- Increased understanding of government program impact on communities
- New knowledge from combined data sources and patterns

METHODOLOGY

Information is knowledge and timely, easily discoverable information is today's imperative. Both geospatial data custodians and their data consumers strive for accurate, current, and informative data.

The methodology for sharing data consists of determining the sensitivity of the data, providing appropriate metadata, and opening the data for sharing. When these concepts are followed, the sharing of data is safe, and efficiencies are improved within an organization.

Utilizing best practices for assessing sensitivity allows for more flexible data both internally and when shared externally. In most cases, the spatial component of data is typically not sensitive and can be provided without divulging sensitive attribute information.

Metadata is a key feature of spatial data and provides the end-user with information about currency and quality. The metadata helps to ensure that the data is properly used. Sharing data can be most effective when the data is distributed through a clearinghouse such as AZGeo. The clearinghouse will provide enhanced security, reporting, and tools to make the datasets easily discoverable.

By following these methods for sharing data, evaluating sensitivity, developing metadata, and open distribution through a clearinghouse, efficiencies will be recognized both internally and externally. Efficiencies are gained through reduced administrative overhead, and the distribution of vetted data.

Characteristics of well-maintained custodial data are current, accurate, and informative. However, when consumers (e.g. other agencies, businesses, and citizens) have difficulty obtaining data, it is often due to the administrative processes in place for gaining access to data. Data currency and/or data containing misinformation due to poor methodology in its collection and maintenance is problematic. When this data is used for mapping and analyses, the results can be misleading and detrimental.

With the speed of Arizona business increasing daily, even simple data-sharing agreement signings can take too long. As a result, data is sought on the internet or previously obtained out of date data may be utilized. Unfortunately, in the haste of needing to produce information, old or inaccurate data may not be checked until problems arise.

Better customer service can be provided by increasing accessibility to data, and

ensuring that the data has been vetted, is accompanied by metadata, and appropriate disclaimers are included. In turn, data providers avoid time-consuming inquiries regarding the data and are protected against damage, loss or liability arising from any misuse, and/or misinterpretation.

The following steps can be taken to open your geospatial data to multiple uses:

Examining Sensitivity

Determining Classification

Determining how to classify sensitivity and share that information within your organization will prevent distribution mistakes. It will also result in identifying how much of your data can and should be shared. A simple classification structure includes:

- Public data is the lowest level of sensitivity. There are no inherent risks to privacy or sensitive information contained in the data itself. Examples include public domain data such as street networks and city boundaries.
- Controlled data is data which may contain personal identifiable information or detailed locational or attribute information identifying sensitive infrastructure. Examples of controlled data include consumer transactions, personal and financial information, and archaeological sites. These data are most likely not appropriate for release as open data without processing to remove sensitive information or to aggregate the data to a level at which individual features are indistinguishable.
 - It is recommended that all data identified as controlled include a date for review of that status, along with documented reasons for the sensitivity. The date for review may vary depending on the nature of the control reason.

Desensitizing Controlled Data

Desensitizing, where possible, may be accomplished by an extract, transform, and load (ETL) process to remove the fields which contain sensitive information. The ETL process output could then be written to a 'Public Data' shareable file. There are other desensitizing options involving altering records by substituting random, yet validly formatted values, but this method can substantially alter data to the point that it may not retain its value.

The ETL process has two parts:

- Step one is to determine a minimum content standard, for each common thematic type
 - Initial layers on which to implement a minimum content standard would likely include streets, parcels, addresses, structures, geographic names, administrative boundaries and districts, and similar widely used spatial datasets
 - Keep the attributes that make features identifiable and valuable for mapping and analysis, remove any sensitive information fields, and add custodian contact information (e.g. website or email).

This process would create a single standardized data model that is useful across unique organizations.

- Step two is to improve operations through streamlining the process with automation. Standardized data model population with distribution through AZGeo provides multiple benefits to both the data custodian and consumers, including:
 - Fewer questions about how to get the data, where to find it, and what public information policies affect distribution.
 - Easier access for entrepreneurial use across jurisdictional boundaries
 - Associated mapping and analytical products based on authoritative data

Suggested Technology

Spatial ETL is utilized by many state geospatial clearinghouses. There are advantages to using ETL technology, including:

- Standardization: data is mapped to a consistent set of lookup values (e.g. AZ, Ariz., and Arizona all mapped to the standard state code)
- Validation: perform quality checks and either remove or modify sensitive data items

Control Your Data's Attributes

Attribute field mapping and field value transfer from the input dataset to the standardized output dataset can be accomplished through an ETL process. The recommended steps are:

- Create a dataset template with the standardized attribute fields.
- Map the attribute fields from the input dataset to the standardized attribute fields in the template dataset.
- As part of this step, the input fields/values that are transferred can be controlled.
- Transfer the input fields/values to the template dataset.
- Create a permanent relationship between the input dataset and the template dataset so that the data transfer process can be automated as updates to the input dataset are applied.

Disclaimers

The purpose of a disclaimer is to define the scope of rights and obligations between the data provider and the data user. Because no dataset is completely error-free, risk is implicit in using data even from a known source. Thus, a disclaimer serves two purposes. First, it serves to inform the data user that a level of uncertainty exists from the standpoint of the data provider regarding completeness and accuracy, regardless of quality control measures in place. Second, it serves to inform the data provider that any potential users of the data acknowledge and accept these risks.

AGIC has created a disclaimer for default use in the AZGeo Clearinghouse which addresses liability, accuracy, and re-use, as components that organizations can use independently or together, for their own use:

Disclaimer

We give no warranty, expressed or implied, as to the accuracy, reliability, or completeness of these data. This disclaimer applies both to the direct use of the data and any derivative products produced with the data.

Any type of boundary, linear, or point locations contained within this data or displayed within this product are approximate and should not 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. If legally-defensible boundaries or locations are required,

they should first be established by an appropriate state-registered professional.

Per A.R.S. §37-178: A public agency that shares geospatial data of which it is the custodian is not liable for errors, inaccuracies or omissions and shall be held harmless from and against all damage, loss or liability arising from any use of geospatial data that is shared. The information contained in these data is dynamic and may change over time. It is the responsibility of the data user to use the data appropriately and consistent with the intent stated in the metadata.

Limitations of use: (Select One)

- This data is not to be redistributed and/or used for commercial purposes
- This data is being provided with no restrictions for redistribution and/or commercial use

Metadata

Metadata is a very important component of any GIS data or GIS-derived product. Providers of GIS data and/or products should ensure that metadata is included, and recipients should ensure that metadata is included. The adoption of metadata guidelines for assessing and reporting quality indicators of GIS data provides a tool that can effectively inform and guide data end-users in the proper use of a given dataset and/or its derivative products.

To ensure metadata provides minimum essential data quality indicators, it should include:

1. Descriptive name/description/abstract
 - General overview of what the data set encompasses.
2. Purpose of the data and intended use
 - What purpose and use were the data created for?
3. Known errors and qualifications
 - Indicators of usability and constraints.
4. Dates
 - Creation, update, and maintenance frequency.
 - Temporal information.
5. Contacts
 - General and/or maintenance.
 - References for acquiring additional information.
6. 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 and especially for encoded values.
7. Limitations of use and/or distribution
 - Augments the intended use with specific use and/or distribution restrictions.

Open Data for Categorical Sharing

Sharing with the General Public

There are three main steps in making data open through AZGeo.

1. Choose your dataset(s)
2. Share the data through AZGeo with the General Public
 - See more detailed instructions at:
https://azgeo.az.gov/azgeo/sites/default/files/help/AZGEO_HelpDoc_0.pdf
3. Make it discoverable through metadata
 - Add Keywords, Organization, Categorization, etc.

Alternatively, an organization could also share data with the public through an Open Data portal; typically, available within the geospatial software tools commonly utilized. Access to these organization-specific portals can be linked through AZGeo to provide a one-stop shop for users to discover and use.

Sharing Confidential Data Through AZGeo Groups

Groups have been created in AZGeo to provide security for data contributors. It is recommended that Group Administrators are assigned to ensure that access to the group data is granted for appropriate users only. For example, only Arizona State Land Department (ASLD) employees are a part of the ASLD group. Group structure allows for data contributors to understand who they are sharing their data with, especially if it is data of a sensitive nature that cannot be shared with the general public.

APPENDICES

A. Arizona Geographic Information Council (AGIC) and AZGeo

Arizona Geographic Information Council's primary mission is to coordinate the development and management of geographic information in Arizona. The AGIC vision is to facilitate access to credible, timely, and accurate geographic information to be used by both decision makers and the citizens of Arizona. This type of sharing reduces duplication of data and personnel costs throughout all branches and agencies of government. Agencies may leverage this sharing model to fulfill public records requests.

The guidelines presented in this document reference current Arizona Revised Statutes (A.R.S. §37-178, reference below) and are intended for use by anyone who engages in the creation, maintenance, and/or publication of geospatial data. These guidelines do not override or change any national or local laws, rules, policy, and regulations which an individual or organization must follow.

The intent of these guidelines is to encourage those who create, maintain, and steward geospatial data to adopt an open data sharing policy while also providing justification for doing so.

About AZGeo

The AZGeo Clearinghouse is an initiative of the AGIC. AZGeo is designed to provide GIS users with links to internet map services, metadata, and geospatial data downloads. AZGeo is hosted and maintained by the Arizona State Land Department. Data on AZGeo includes GIS layers for administrative boundaries, demography, environmental factors, hydrology, imagery, indices, mining, natural features, transportation, and much more.

How does AZGeo work?

If an agency creates and manages GIS data that is frequently requested by others, becoming an AZGeo member will allow sharing the data in an efficient and effective manner. Data access is determined by the contributor and may be fully public and open, or targeted. AZGeo is organized in groups, making it possible to control access to sensitive data, thus ensuring security. Additionally, tracking features on the website make it possible to monitor downloads and identify who is downloading data.

The data and services available on AZGeo are contributed by a variety of public and

private organizations in Arizona. Different from other clearinghouses and a significant feature and strength of AZGeo is that all accepted data must come from the authoritative source for such information. Contributing members agree to provide data, metadata, and services in approved formats which are standardized for user convenience. This service cannot exist without the data and services content which is provided by the contributing members.

B. A.R.S. § 37-177 & A.R.S. § 37-178

- A.R.S. §37-177 defines the Arizona Geographic Information Council. It covers the makeup, appointments and duties. Relative to this paper, those duties include:

D. The council shall:

1. Advise the Arizona state cartographer's office on matters related to geospatial data sharing and appropriate strategies to support a geospatial clearinghouse and statewide geographic information system.
 2. Facilitate interagency coordination for the purpose of geospatial data sharing and supporting a geospatial clearinghouse, a statewide geographic information system and location-based services that enhance and support federal, state and local government business systems, including emergency preparedness and other disaster mitigation response systems.
- A.R.S. §37-178 explains geospatial data sharing. The complete text reads as follows:
 - A. A public agency that shares geospatial data with another public agency may:
 1. Share geospatial data without entering into a written agreement with the other public agency.
 2. Share geospatial data of which it is the custodian.
 3. Retain custodial ownership of any geospatial data provided to other public agencies.
 4. Prohibit shared data from being redistributed by recipient public agencies if notification of the prohibition is given.
 - B. A public agency that shares geospatial data may exempt the data from commercial use fees prescribed in section §39-121.03, subsection A, paragraph 3.

- C. A public agency that shares geospatial data of which it is the custodian is not liable for errors, inaccuracies or omissions and shall be held harmless from and against all damage, loss or liability arising from any use of geospatial data that is shared.
- D. A public agency that shares geospatial data or receives shared geospatial data may withhold the shared data from public disclosure if the data consist of critical infrastructure information as defined in section §41-1801.

Open Data – Why, What, How?

Why Open Data?¹

Open data, especially open government data, is a tremendous resource that is as yet largely untapped. Many individuals and organizations collect a broad range of different types of data in order to perform their tasks. Government is particularly significant in this respect, both because of the quantity and centrality of the data it collects, but also because most of that government data is public data by law, and therefore could be made open and made available for others to use. Why is that of interest?

There are many areas where we can expect open data to be of value, and where examples of how it has been used already exist. There are also many different groups of people and organizations who can benefit from the availability of open data, including government itself. At the same time it is impossible to predict precisely how and where value will be created in the future. The nature of innovation is that developments often comes from unlikely places.

It is already possible to point to many areas where open government data is creating value. Some of these areas include:

- Transparency and democratic control
- Participation
- Self-empowerment
- Improved or new private products and services Innovation
- Improved efficiency of government services Improved effectiveness of government services Impact measurement of policies
- New knowledge from combined data sources and patterns in large data volumes

While there are numerous instances of the ways in which open data is already creating both social and economic value, we don't yet know what new things will become possible. New combinations of data can create new knowledge and insights, which can lead to whole new fields of application. We have seen this in the past, for example when Dr. Snow discovered the relationship between drinking water pollution and cholera in London in the 19th century by combining data about cholera deaths with the location of water wells. This led to the building of London's sewage systems, and hugely improved the general health of the population. We are likely to

see such developments happening again as unexpected insights flow from the combination of different open data sets.

This untapped potential can be unleashed if we turn public government data into open data. This will only happen, however, if it is really open, i.e. if there are no restrictions (legal, financial, or technological) to its re-use by others. Every restriction will exclude people from re-using the public data, and make it harder to find valuable ways of doing that. For the potential to be realized, public data needs to be open data.

What is Open Data?²

For our purposes, open data is as defined by the Open Definition:

Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share-alike.

The full Open Definition gives precise details as to what this means. To summarize the most important:

- **Availability and Access**: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Reuse and Redistribution**: the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets.
- **Universal Participation**: everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavor or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use or restrictions of use for certain purposes (e.g. only in education), are not allowed.

If you're wondering why it is so important to be clear about what open means and why this definition is used, there's a simple answer: interoperability.

Interoperability denotes the ability of diverse systems and organizations to work together (inter-operate). In this case, it is the ability to interoperate - or intermix - different datasets.

Interoperability is important because it allows for different components to work together. This ability to componentize and to 'plug together' components is essential to building large, complex systems. Without interoperability this becomes near impossible.

Providing a clear definition of openness ensures that when you get two open datasets from two different sources, you will be able to combine them together, and it ensures that we avoid our own ‘tower of babel’: lots of datasets but little or no ability to combine them together into the larger systems where the real value lies.

¹ <http://opendatahandbook.org/guide/en/why-open-data/>

² <http://opendatahandbook.org/guide/en/what-is-open-data/>