Geospatial Data Governance Plan
GIS Project

Prepared by: State Geospatial Data Governance Work Group
Department: OIT
Co-authors: Chris Brown, Division of Water Resources; Jon Gottsegen, OIT; William Johnson, Department of Transportation; Mike Rigirozzi, Department of Agriculture; Bob Sacco, OIT/Division of Wildlife; Mary Sullivan, Historical Society
Date: November 16, 2010

Executive Summary of Recommendation

This plan outlines a data governance process for geospatial data sets in the state. The governance process will integrate with the enterprise data architecture and governance being developed through the Government Data Advisory Board (GDAB) and under the state's Chief Data Officer.

Among the business drivers for this geospatial data governance effort are the Information Technology (IT) consolidation mandated by SB 08-155, recommendations for data stewardship mentioned in work by Applied Geographics[1] and CH2MHill[2], and the difficulty in discovering relevant geospatial data. Goals and objectives include defining an organizational structure and roles for geospatial data governance, identifying high priority geospatial data sets for stewardship and defining a stewardship process for them, promoting participation in data stewardship among local and federal stakeholders, monitoring the benefits of data stewardship, and identifying constraints to stewardship. The metrics for success are categorized into financial, process, customer and infrastructure measures. This plan identifies the desired direction of change in these measures if the data governance program is successful, and the time frame over which the metrics should be assessed.

Context for this plan has been set by IT consolidation and the GDAB. The plan also builds on the Colorado Geospatial Information Advisory Council, which represents the variety of Geographic Information Systems (GIS) stakeholders across all jurisdictional levels in the state. The plan has been developed by a working group of state agency staff from the Department's of Agriculture, Natural Resources, Transportation, the Historical Society and the Office of Information Technology (OIT). The working group began this effort by taking an inventory of the geospatial data sets maintained and used by state agencies. This inventory included over 300 data sets and approximately 25 state agencies. The working group also surveyed the agencies regarding standards they used for metadata, disclaimers, quality statements, distribution and data structures.

The plan provides definitions of several important concepts for data governance generally and the geospatial arena specifically. These definitions are: data governance, data stewardship, geographic data, primary data user, secondary data user, data steward, data provider, data owner, primary content provider and secondary content editor. These definitions will be aligned with definitions from the GDAB.

One of the most significant contributions of this plan comes in the section describing standard steps involved in data stewardship. These steps provide geospatial data users with a common

---

[1] Colorado GIS Coordination Strategic Plan, Applied Geographics, Inc., February 2008
[2] Colorado Statewide GIS Coordination Findings and Recommendations, CH2MHill, April 2007

understanding or set of expectations of how data will be managed, and therefore the quality they can expect when a state agency undertakes data stewardship. This plan does not imply that a state agency will do all of the steps, as more advanced steps may be beyond the resources of the agency, but agency stewards of geospatial data will identify the set of common steps they will pursue and as a result the level of stewardship they can support. The steps involved in stewardship are 1) assessing the needs for the data, 2) establishing and maintaining a stewardship team, 3) compiling and maintaining the data, 4) distributing the data, and 5) defining standards or policies for the data. While step three is commonly thought of in the GIS domain as integrating data from various sources into one normalized, consistent and seamless data set across the state, it may simply involve collecting data from local and other sources and making these separate data sets available from one location, or even simply coordinating metadata about the available data sets, depending on the resources.

The plan describes how this geospatial data governance process should integrate and align with the enterprise data governance. There should be representation of geospatial expertise and needs among the data stewards action council and a member of the working group should participate in or attend the GDAB meetings.

Last, this plan considers the priority of specific data sets for active stewardship. This prioritization depends on a number of factors including need as evidenced in the data inventory, available stewards, complexity of the data, and opportunities presented by federal grants or other programs. The working group also identified four data sets as pilots for stewardship. They are the National Hydrography Dataset, currently under a fairly mature stewardship program at the Division of Water Resources, local parcel (property ownership) data, “Community Anchor Institutions” data (i.e., education, health, public safety and other government facilities), and roads. Each of these data sets will have a stewardship plan for it in a standard format and structure. Plans for two of the data sets area appended to this plan.

Table of Contents

Section I: Introduction 1
I.1: What is this Plan Trying to Accomplish 1
I.2: Why Geospatial Data Governance 2
I.3: Business Drivers 2
Section II: Goals, Objectives, Benefits and Metrics 3
II.1: Goal 3
II.2: Objectives 3
II.3: Metrics for Success 5
Section III: State Context 6
III.1: Geospatial Data Governance Work Group 7
III.2: Colorado Geospatial Information Advisory Council 7
III.3: GIS Consolidation 8
III.4: Government Data Advisory Board and House Bill 1364 9
III.5: Statutes and Rules 10
III.6: Current Status of Spatial Data in the State of Colorado 11
Section IV: Definitions 13
IV.1: Data Governance 14
IV.2: Data Stewardship 14
IV.3: Geographic Data 14
IV.4: Primary Data User 15
IV.5: Secondary Data User 15
IV.6: Data Steward 15
IV.7: Data Provider 16
IV.8: Data Owner 16
IV.9: Primary Content Provider 16
IV.10: Secondary Content Editor 16
Section V: Governance Process and Structure 17
V.1: Stewardship Process 17
V.1.1: Needs Assessment 18
V.1.2: Establishing and Maintaining a Team 21
V.1.3: Data Compilation/Maintenance 23
V.1.4: Distribution 26
V.1.5: Data Governance and Standards Process 27
V.2: Governance Structure 32
V.3: Communications 36
V.4: Data Security 36
Section VI: Data Set Stewardship 37

VI.1: Prioritizing data sets for stewardship 37
VI.2: Data Sets To Be Considered Long Term 39
VI.3: Data sets covered in this plan 41
Section VII: Conclusion 41
Appendix A: National Hydrography Dataset Stewardship Plan 42
Appendix B: Community Anchor Institutions Stewardship Plan 46

Section I: Introduction

The Colorado Geospatial Data Governance Plan has been developed to provide a starting point and set a course for the effective management of statewide geospatial data assets. Geospatial data have been effectively developed, managed and utilized by many state agencies including the Colorado Department of Natural Resources, the Colorado Department of Transportation, Colorado Department of Public Health and Environment, and many other agencies in support of their business processes for close to three decades.

Recent direction in the organization and management of state information technologies has provided new motivation to address a longstanding need to formally manage geospatial assets. Legislation passed by the Colorado General Assembly and signed by the Governor in 2008 and 2009 has provided the context and impetus for Colorado state government to formally manage information as enterprise assets. These bills are described in the next section of the plan.

This plan specifically addresses the current status of spatial data within Colorado government as well as actions that should be taken in order to establish accurate, complete, timely, secure and authoritative data sources that can be utilized across state government and beyond. The plan includes definitions relevant to data governance and the geospatial industry. The plan reviews current standards in use among state agencies relative to the management and distribution of spatial data. It recommends refinement of those standards within a structure that aligns Colorado's geospatial assets to the broader context of Colorado government's information assets, and proceeds with a goal of continuous improvement.

I.1: What is this Plan Trying to Accomplish

This plan defines a governance structure and proposes stewardship mechanisms for priority geospatial data sets in the state. It also defines a process for prioritizing additional data sets for stewardship. The plan will describe a governance process that state agencies should follow to result in more consistent handling of data and authoritative data sets for the state.

Entities external to state government will benefit from authoritative data sets because the data quality will be known and documented. The stewardship processes in this plan will identify how local, federal, regional and tribal entities should participate in building the statewide data sets. The stewardship process outlined in this plan will also clarify expectations and constraints and resource requirements of state agencies compared to desired data structures or content among the geospatial community in the state.

I.2: Why Geospatial Data Governance

The reasons for pursuing geospatial data governance are the same as those for other data governance: protecting investment, securing data, improving decision making. A 2008 paper from the National Association of State Chief Information Officers (NASCIO) stated, "Data governance is essential to ensuring that data is accurate, appropriately shared, and protected. Data is rationalized to create information…The quality of data and information will certainly impact the quality of the decisions that consume it."[1] Geospatial data differs somewhat from other data in that:

- geospatial data is often widely shared and developed with the intent to be shared,
- comprehensive data often is built from multiple contributors, and
- several data sets that are used routinely don't have clear owners, so specific mechanisms for integration of data across the state and clearly defining owners versus stewards become key issues.

Data governance will improve data sharing by:

- making authoritative data known and discoverable,
- improving the reliability of data because its quality is known, and
- defining a consistent system for identifying access or security restrictions for data rather than individual ad hoc decisions.

I.3: Business Drivers

Senate Bill 08-155, passed by the Colorado legislature and signed into law by the Governor in 2008, establishes the Governor's Office of Information Technology (OIT) as the lead agency for statewide geographic information system coordination. In addition, the Geographic Information Systems (GIS) community, both within state government and among other stakeholder groups in the state, has identified access to authoritative data for various geospatial data sets as an important business driver. This plan will include consideration of effective and efficient governance or stewardship of geospatial data in the state, particularly among state agencies.

SB08-155 mandates that all duties and responsibilities for statewide geographic information system coordination be transferred to the Office of Information Technology, leading OIT to develop an implementation plan for appropriate consolidation of GIS activities among state government entities. This plan will include consideration of effective and efficient governance or stewardship of geospatial data in the state, particularly among state agencies.

State agencies, like other entities have trouble locating authoritative data for their business activities. In some cases, several state agencies are editing data for similar needs. Geospatial data

[1] Data Governance – Managing Information As An Enterprise Asset Part I – An Introduction. NASCIO. April 2008. http://www.nascio.org/publications/documents/NASCIO-DataGovernance-Part1.pdf

governance will lead to efficiencies in maintenance of priority spatial data sets by clarifying a maintenance process for spatial data sets. Geospatial data users often need expanded data structures and content for statewide data. Data governance and stewardship will clearly specify the content of data provided by the state thereby providing a foundation for potential users of the data to assess its fitness for particular uses. It will provide a basis for a gap analysis between state data content and the specific business needs of GIS users within the state, particularly those in local and federal entities, and allow for identification of necessary resources to overcome the identified gaps. For certain data sets, data governance will potentially increase opportunities for federal funding to support maintenance, development, discovery and accessibility of spatial data. Last, data governance will also allow the state to document and demonstrate the benefit of developing an accessible and widely-used state data clearinghouse or repository for storing and distributing useful data products resulting from a data governance effort.

Recent reports by CH2MHill[2] and Applied Geographics[3] have both identified data governance as an important component in furthering GIS coordination in the state. In addition, the GIS community, both within state government and among other stakeholder groups in the state, has identified access to authoritative data for various critical geospatial data sets as an important business driver.

Section II: Goals, Objectives, Benefits and Metrics

This plan presents the following short and long term goals regarding the effective management and usage of spatial data within the public sector.

II.1: Goals

The short term goal of this effort is to establish a plan for the governance of priority geospatial data sets in the state. It will establish a foundation to identify the priority data sets for which stewardship mechanisms should be defined, recommend governance approaches and stewards for those data sets as well as suggest an overall governance model for "framework" data in the state's spatial data infrastructure. In the long term, this plan will provide best practices for agencies within Colorado State Government to identify and document authoritative spatial data and further develop the content of these spatial data sets to support the business needs of state agencies and then of other stakeholders.

II.2: Objectives

To achieve these goals this plan seeks to achieve the following objectives:

- To define what is meant by data stewardship, ownership, governance and authoritative data.

[2] Colorado Statewide GIS Coordination Findings and Recommendations, CH2MHill, April 2007
[3] Colorado GIS Coordination Strategic Plan, Applied Geographics, Inc., February 2008

- To define a data governance procedure that establishes roles, responsibilities, and priorities for stewarded geospatial data sets within the state. This procedure will leverage and align with data governance efforts of the Government Data Advisory Board and the State's Enterprise Architecture Office.
- To identify priority data sets for the State of Colorado, and provide examples showing how the defined stewardship procedure could be applied.
- To include state agencies and then local, federal and other entities in the maintenance of priority data through the defined stewardship procedures.
- To identify and monitor necessary standards for data delivery, documentation and updates, where appropriate, and recommend data access procedures for the priority data sets. These standards can also include an approach to defining security for these data.
- To identify business needs and benefits for geospatial data governance and stewardship including obligations arising from statute or administrative rule.
- To identify constraints to effective data stewardship for the identified data sets.

II.3: Benefits

A defined data governance process for the state will yield the following benefits:

- Lead to efficiencies in maintenance of priority spatial data sets.
- Provide effective, definitive and standard approaches to data security for spatial data.
- Specify data content of the data provided by the state thereby providing a foundation for users of these data to assess their fitness for particular uses.
- Provide a basis for a gap analysis between state data content and specific business needs of GIS users, and allow for identification of necessary resources to overcome the identified gaps.
- Provide a primary, "authoritative" set of state data that GIS users can refer to when these critical data are required, and a documented process for identifying authoritative data sets.
- Increase opportunities for federal funding to support maintenance, development, discovery and accessibility of spatial data.
- Leverage successful data maintenance models in state agencies for other data sets.
- Highlight opportunities for utilizing enterprise approaches to data governance, maintenance, discovery and dissemination.
- Document and demonstrate the benefit of developing an accessible and widely-used state data clearinghouse or repository for storing and distributing useful data products resulting from a data governance effort.

The development and maintenance of authoritative data within the state will yield the following benefits:

- Clear data sets appropriate for specific purposes will be discoverable and available reducing time and cost involved in searching for these data.
- The lineage and therefore limitations and applicability of authoritative data sets will be immediately understandable, eliminating time required to research the data and its history.
- Stakeholders outside of state government will be able to discover, obtain and easily understand the quality of data within the state, thereby facilitating their processes and inter-jurisdictional collaboration.
- Documentation regarding the data's use by the GIS community can be accessed from the same location allowing decision-makers to see the return on investment for data sets easily.

II.3: Metrics for Success

This project will use several metrics to evaluate the success of the data governance program. These can be categorized into financial, process related, customer related and infrastructure metrics. The metrics will be assessed annually. The annual change will demonstrate the degree to which data governance is succeeding in improving the availability and quality of data and the business processes that rely on it.

Table 1 below lists the metrics that the geospatial data governance work group has identified. It also indicates the direction of change that would indicate successful geospatial data governance, as well as the time frame for evaluating each metric.

Table 1: Metrics for Success

Category	Measure	Desired Trend	Forecast Time Frame
Financial	$ saved through data stewardship	higher	long-term
	$ spent developing data stewardship plans	lower	long-term
	Person/consultant hours spent collecting data (as primary or secondary content providers)	lower	long-term
	Person/consultant hours spent processing data (as primary or secondary content providers)	lower	long-term
	Person/consultant hours spent publishing data (as primary or secondary content providers)	lower	long-term
	Cost of product delivery/receipt	lower	long-term
	Value of data asset	higher	long-term
	$ spent on storage infrastructure associated with each data set	lower	long-term
	$ spent on software for data maintenance	lower	long-term
Process	# of priority theme layers stewarded	higher	mid-term
	#/% of redundant priority theme layers	lower	mid-term
	#/% of priority theme layers with value added	higher	long-term

Category	Measure	Desired Trend	Forecast Time Frame
	#/% of priority theme layers with stewardship plan (state departments)	higher	mid-term
	#/% of priority theme layers with stewardship plan (others)	higher	mid-term
	# of state departments using stewarded data sets	higher	short-term
	# of other agencies using stewarded data sets	higher	long-term
	#/% of priority theme layers using statewide spatial reference/coordinate system standard	higher	short-term
	#/% of priority theme layers using statewide metadata standard	higher	short-term
	#/% of priority theme layers with process documentation (data quality, content, collection)	higher	short-term
	#/% of priority theme layers with standard decription of security constraints	higher	short-term
Customer	# of priority theme layers published	higher	mid-term
	#/% of geographic web services using stewarded data	higher	mid-term
	# of web site hits (clearinghouse and per web service using stewarded data)	higher	long-term
	#/% of data delivery through email	lower	mid-term
	#/% of data delivery through hard copy map/report	lower	mid-term
	#/% current data sets (updated within past 2 years)	higher	long-term
	Rating on statewide customer satisfaction survey	higher	long-term
Infrastructure	# of dedicated servers	lower	mid-term
	# of servers shared across agencies (includes virtual)	higher	mid-term
	Storage capacity used for priority data sets	lower	mid-term

Section III: State Context

The governance effort for geospatial data is proceeding among several other projects related to information technologies and particularly geospatial information technologies in the state. They either impact the geospatial data governance effort directly or provide a context to which the geospatial data governance program should align. This section these other efforts and how they will influence the geospatial data governance. The efforts include:

- Geospatial Data Governance Work Group
- Colorado Geospatial Information Advisory Council
- GIS Consolidation within the State of Colorado
- Government Data Advisory Board (GDAB) and House Bill 1364
- Colorado Statutes (namely the Colorado Open Records Act)

It also first describes the geospatial data governance work group that has been developing this plan and then concludes with a description the current environment of geospatial data among state agencies.

III.1: Geospatial Data Governance Work Group

This plan has been developed by a geospatial data governance working group composed of personnel from various state departments. The working group has met approximately biweekly since January 2009. Participation in this working group was strictly voluntary. The primary participants in the working group are:

Chris Brown, OIT/Department of Natural Resources (Div. of Water Resources)
Simon Chen, Department of Transportation
Jon Gottsegen, Office of Information Technology
Lou Henefeld, Department of Transportation
William Johnson, Department of Transportation
Marv Koleis (Chair), Department of Transportation
Aaron Rhodes, Department of Transportation
Mike Rigirozzi, Department of Agriculture
Bob Sacco, OIT/Department of Natural Resources (Div. of Wildlife)
Mary Sullivan, Colorado Historical Society

This group is related to the Geospatial Information Advisory Council, described below, in that one of the Council's responsibilities is proposing data governance for geospatial data. The data governance process is also a component effort of the consolidation of geospatial activities in the state. Several of the members of the geospatial data governance working group participate in the Geospatial Information Advisory Council and the GIS consolidation effort.

In addition to developing this plan, the working group has inventoried the geospatial data sets maintained and used by state agencies and surveyed the agencies regarding the standards or protocols in uses for distributing and maintaining data. This survey is described in Section III.6. The working group has also prioritized data for stewardship based on the survey.

This working group will continue to function after the plan is complete to review and recommend standards and policies related to geospatial data. It will be a standing group to provide guidance from the state agency GIS to the GDAB on deliberations that impact geospatial activities. The working group will also review progress on stewardship undertaken by state agencies and annually evaluate the metrics proposed in this document. As it moves into this role, it will meet less frequently.

III.2: Colorado Geospatial Information Advisory Council

The Colorado Geospatial Information Advisory Council (CGIAC) is a body of representatives from the state's geospatial community. Its purpose is to advise the State Chief Information Officer (CIO) on matters related to coordination of geospatial information and technologies and specifically on how the state can provide services that can assist the community of geospatial

information users. The CIO convened the CGIAC and tasked it with several specific responsibilities. They are:

a. Advise the State's Chief Information Officer (CIO) of the Governor's Office of Information Technology (OIT) on plans that involve statewide GIS activity.

b. Recommend appropriate data governance, including stewardship procedures and state agency stewardship roles, for (at a minimum) the geospatial framework data layers defined by the Federal Geographic Data Committee (FGDC).

c. Recommend metadata standards and data exchange standards as well as spatial information sharing guidelines among the state geospatial community.

d. Recommend procedures for defining priorities, methods to coordinate development and use of geospatial information, and forums for sharing technical expertise and ideas from all sectors in the geospatial community.

e. Recommend an effective strategy to share and integrate spatial data across all levels of government as well as with institutions of higher education, the private sector and non-profit organizations.

f. Promote cross agency cooperation and recommend cost sharing and collaborative arrangements resulting in more efficient government processes as well as more effective service delivery and decision-making.

g. Deliver an annual progress report to the State CIO, to include current status of GIS activities in the state, progress made throughout the year on priorities and activities, and objectives for the following year.

Responsibility (b) specifically deals with geospatial data governance. This data governance plan will be vetted through the CGIAC to provide for input and recommendations from the geospatial information community at large. Although this won't obligate state agencies to respond to the needs of the broader community, it will allow those needs to be documented.

The concrete, detailed work of the CGIAC will be accomplished through working groups. A working group for data governance may be formed to continue the development of data governance programs statewide.

III.3: GIS Consolidation

In 2008 the state legislature passed Senate Bill 08-155. This bill mandated consolidation of information technologies under the Governor's Office of Information Technology (OIT). This bill also included a GIS provision:

> **24-37.5-111. Geographic information system - coordinator - statewide plan.** On and after July 1, 2008, all duties and responsibilities for statewide geographic information system coordination shall be transferred from the department of local affairs to the office. The office shall develop a statewide geographic information system plan on or before July 1, 2010, and submit such plan to the Governor and to the State, Veterans, and Military Affairs committees of the Senate and the House of Representatives, or their successor committees.

OIT is pursuing consolidation of geospatial activities as part of its IT consolidation program. The GIS consolidation includes various components. One piece is data management. This reflects a recognition that it is important for the state to manage geospatial data, like any other data, across the state enterprise as an important asset for the state. Enterprise-wide data governance protects and optimizes the investments the state is making in data.

The state has convened a working group of state agency GIS personnel who are developing an implementation plan for GIS consolidation. Part of the plan is to identify activities that are appropriate to be managed or executed centrally versus as independent line of business functions. While data governance may be a centrally managed function, the stewardship of specific data sets for the state may be best executed within specific agencies. This data governance plan identifies these central versus line of business structures.

III.4: Government Data Advisory Board and House Bill 1364

Recognizing the increase in demand for data sharing and the need for more information based decision-making, balanced with increasing privacy issues related to personally identifiable information, the State Legislature passed House Bill 08-1364. HB 08-1364 directed OIT to convene a Data Protocol Development Council ("Council") to assist in designing and implementing an interdepartmental data protocol. The goal of the cross-departmental data protocol is to facilitate information sharing across agencies, and to assist in formulating and determining the effectiveness of state policies.

The Council was composed of state personnel from many different state agencies. Since the primary motivation for it was education policy, it focused on “unit records,” or data about individuals. It identified challenges to data sharing and enterprise wide data management and made various recommendations to overcome these challenges, including expanding data sharing in existing service delivery channels and standardizing approaches to data architectures, data security and other data management practices, which can be accomplished through accelerated IT consolidation. This Council identified more opportunities to share data in most business aspects of state government, including education, workforce development, and citizen services.

As a follow-up to this Council, the Legislature passed House Bill 09-1285 which created the Government Data Advisory Board. This board has several duties mandated by the legislation.

Among them are: advising the CIO regarding the ongoing development, maintenance, and implementation of the interdepartmental data protocol (from the HB 1364 Council); advising the CIO concerning best practices in sharing and protecting data in state government; recommending rules and procedures that a state agency shall follow in requesting, or responding to a request for, data from another state agency, including but not limited to strategies for enforcing said rules; and advising the CIO on rules and procedures for responding to data requests submitted by an entity outside of state government. The legislature also created an Education Subcommittee of this board to deal with education related data immediately. While geospatial data is not specified in this legislation, the GDAB will clearly be making recommendations that will impact geospatial data policies in the state.

III.5: Statutes and Rules

There are several state and federal statues or rules that determine how geospatial data, and data more generally, is developed, distributed and maintained in the state. This section will describe some of these statutes or rules. However, much of the data developed by state agencies is produced to meet specific business requirements mandated for those agencies by Colorado or federal law. These laws that drive the business of state agencies, and therefore the development of particular data sets will be described in the section pertaining to the stewardship of those individual data sets.

Colorado Open Records Act (CORA) 24-72 C.R.S.

CORA (citation) specifies what data must the obligations of state government and political subdivisions are regarding shared data with the public. Based on work done with the Attorney General's Office in previous years, CORA includes spatial data as a "public writing." It specifies that government entities must provide for inspection of their data on request.

Government Data Advisory Board 24-37.5-703 C.R.S. et seq.

As described above, this board is mandated to recommend data governance policies to the State Chief Information Officer. Geospatial data will also fall under these policies.

Survey Records 38-51 C.R.S. et seq.

This statute defines standards for survey records and plats. It also deals with GIS explicitly by including a provision for GIS records allowing less rigorous accuracy than property surveys. The statute also includes a statutory Colorado coordinate system.

Several other statutes are salient for various types of geospatial data. They tend to be specific to particular data sets, and part of the stewardship process for data sets will involve identifying these laws or rules that determine how the data are created or maintained.

III.6: Current Status of Spatial Data in the State of Colorado

The State of Colorado has several agencies that use geospatial technologies to varying degrees. The heaviest users of geospatial technologies among state agencies are Colorado Department of Transporation (CDOT), Colorado Department of Public Health and Environment (CDPHE), Division of Natural Resources (DNR), Colorado Department of Public Safety (CDPS) and Division of Local Affairs (DOLA). Agencies that have some geospatial technology users include the Departments of Agriculture, Regulatory Agencies, Corrections, Labor and Employment, Health Care Policy and Financing, the Legislative Council and the Governor's Office. Each of these agencies generates data for their own programs and consumes data from other sources. There are several data sets that are commonly used and provide a standard base map or set of reference layers for GIS efforts. Some of these data sets are included in the "framework" data sets for the National Spatial Data Infrastructure (NSDI) established by federal Executive Order 12906.

Up to this point, in Colorado, there is no single clearinghouse for geospatial data and no identification of "authoritative" data sets. Authoritative data is an important concept in geospatial data governance. It provides for data that is managed by an identified steward according to the best practices for that particular data given a defined level of accuracy and resolution. Authoritative data sets are those the state can specify as the preferred data of known quality.

As part of the data governance effort, the state has developed a comprehensive inventory of the geospatial data developed and used by state agencies. This comprehensive work identified over 300 data sets in use. It identified whether individual agencies required the data sets for their business or whether they used the data sets but the data were not actually critical to the business processes of the agencies. The inventory also identified which agencies are primary data creators or editors and whether an agency is a secondary or additional content editor for each data set. This detailed inventory is available at: http://bit.ly/9t3yjx, but a summary of this inventory is shown in Table 2 below.

Table 2 lists state agencies that responded to the inventory as columns. The rows list the following information:

- Primary Content Provider – This is the agency that is primarily responsible or has assumed the primary role of maintaining a given data set. An example of this is CDOT for roads.
- Additional Content Editor – This is an agency that takes data sets from the primary content provider and adds or edits data for its own purposes. An example of this is the Division of Wildlife adding roads to CDOT's road data in national forests.
- Required layer – This indicates that a data set is required for an agency to accomplish its business. If the data were no longer available, the agency would be forced to find an alternate source for the data.

- Data consumer – This indicates that an agency periodically uses the data for mapping or other purposes, but it is not a critical component of the agency's business.

The numbers in each cell represent the number of data sets for which each agency assumes the roles identified in the rows. For example, CDPHE has said that it is the Primary Content Provider for 27 out of the 300+ data sets included in the inventory.

Table 2: Geospatial Data Inventory

	Public Health & Environment (CDPHE)	Military and Veteran Affairs (DMVA)	Department of Public Safety (DPS)	Colorado Information Analysis Center (CIAC)	Critical Infrastructure Protection Mission Assurance (CP-MAA)	DNR - Colorado Water Conservation Board (CWCB)	DNR - Division of Water Resources (DWR)	DNR - Colorado Geological Survey (CGS)	DNR - Division of Wildlife (DOW)	DNR - Division of Parks and Outdoor Recreation (DPOR)	DNR - Division of Reclamation Mining and Safety (DRMS)	DNR - State Land Board (SLB)	DNR - Oil & Gas Conservation Commission (OGCC)	Department of Agriculture (DOA)	Department of Corrections (DOC)	DOLA - Division of Local Government	DOLA - Division of Emergency Management	Department of Transportation (DOT)
Primary Content Provider	27	4	11	7	1	2	13	9	44	1	2	2	4	14	2	4	0	51
Additional Content Editor	4	0	5	0	0	2	1	0	6	0	0	2	0	1	0	0	0	0
Required GIS Layer Count	0	5	7	4	25	21	13	2	24	0	0	9	51	9	2	16	83	21
Data Consumer	11	6	31	4	3	5	7	2	23	0	0	15	6	92	0	35	86	9
Total Count	42	15	54	15	29	30	34	13	97	1	2	28	61	116	4	55	169	81

The inventory provides some very rich information. Some notable conclusions are:

- There are few redundant efforts at maintaining data sets, although there are some instances of several agencies creating similar derivative products from a data set or adding content to a data set to support its own business needs.
- There are several data sets that are clearly higher priority for data stewardship based on the number of agencies that consider these data sets critical for their work.
- Some collaborative successes exist in the state: National Hydrography Dataset, Denver Regional Council of Governments (DRCOG) Aerial Photography Program, and the National Agricultural Imagery Program (statewide aerial photography)

In addition to this inventory, the working group produced a survey of high level data documentation and maintenance approaches and standards in use. The questions in this survey related to:

- Data Delivery
- Metadata
- Coordinate Systems
- Data Disclaimers
- Data Disclosure/Quality
- Data Constraints/Security
- Data Content
- Data Collection

Seven of 11 state agencies responded to the survey although two indicated that they did not create any data. Some notable results of the survey include:

- A significant number of agencies deliver digital spatial data through email, DVDs/CDs or web services.
- Metadata exists for an overwhelming majority of the data sets covered in the survey, although there is not a strong use of a particular standard. The most commonly used metadata standard is the Federal Geographic Data Committee metadata content standard.
- Data sets are almost entirely maintained on one coordinate system (UTM zone 13, NAD 83)
- The large majority of data sets are accompanied by a disclaimer statement, but the data sets were approximately evenly split in terms of whether they include a statement of data quality.
- Most of the data sets did not have use or security constraints.
- Most of the data sets did not follow a data content standard for the data itself, nor did they have a documented data collection methodology.

While authoritative data sets with accompanying stewardship programs generally do not yet exist across all of the priority data sets, there are some that are developed on a comprehensive statewide basis through agency programs and one data set with a maturing statewide stewardship process. These data sets include transportation and hydrography. The current custodians of these data are likely candidates for stewardship of an authoritative version of the data sets.

Section IV: Definitions

It is critical to have a common understanding of what is meant by data governance and stewardship and the potential roles in data stewardship to arrive at standard data stewardship

processes. These definitions provide a framework and guidance to agencies in embarking on data stewardship efforts. The associated stewardship plans should address the roles of data steward, data provider, and data owner in the context of these definitions and in a manner appropriate to their line-of business and customer needs and subsequently the needs of GIS stakeholders in the state. These definitions reflect the experience of the Geospatial Data Governance Working Group as well information gleaned from similar work in the State of Oregon[4] and publications of the Data Governance Institute[5].

It is understood that the enterprise data governance effort is developing definitions for many of the same terms, so these definitions provide suggestions for the definitions resulting from the broader enterprise effort.

IV.1: Data Governance

Data Governance refers to the operating discipline for managing data and information as a key enterprise asset. This operating discipline includes organization, processes and tools for establishing and exercising decision rights regarding valuation and management of data. Key aspects of data governance include decision making authority, compliance monitoring, policies and standards, data inventories, full lifecycle management, content management, records management, preservation, data quality, data classification, data security and access, data risk management, and data valuation.

IV.2: Data Stewardship

Data stewardship is the practice of managing data and providing users access to that data. Processes supporting geographic data stewardship will be based on clear, inclusive, and well-documented data architecture. Geographic data should be shared widely among the primary and secondary user community with proper consideration to sensitivity, legal, and policy concerns that may restrict access and distribution.

IV.3: Geographic Data

Geographic data includes any data describing features that have a location. Geographic data is a valuable asset that supports the business needs of users. Its value should be maintained over time through an effective and efficient update program and should be maintained at a level that is well documented. Metadata providing information about the content, format, quality, authority, and availability of geographic data is vital, and this metadata should be updated along with the data itself.

---

[4] Oregon Gis Utility Project—Phase 1, Requirements Assessment And Business Case, Geographic Data Stewardship Best Practices, Deliverable 3E, PlanGraphics, Inc.

[5] Governance and Stewardship. Data Governance Institute, 2008. (http://www.datagovernance.com/adg_data_governance_governance_and_stewardship.html)

IV.4: Primary Data User

Primary data users are the agencies, persons, or processes that tie directly to line-of-business functions. Primary users are the primary reason the geographic data is produced and are the high priority customers for the data. There is often a formal agreement, contract, or mandate between the agency or group providing the data and the data users. For example, the Federal Highways Administration requires pavement data from the Colorado Department of Transportation. In this case, they can be considered primary data users. CDOT engineers or planners that use the data may also be primary data users. Changes of data format, type, quality, or products are typically driven by primary data users.

IV.5: Secondary Data User

Secondary data users are the agencies, persons, or processes that use the data, but are not in the data provider's line-of-business. Secondary users are the entities that rely on production of geographic data to support their processes, but are outside of the data provider's line-of-business. There is no formal agreement, contract, or mandate that requires the data provider to produce data for this customer. There is no requirement on the data provider to change the data format, type, quality, or products based on the needs of secondary users.

IV.6: Data Steward

A data steward is a person or organization delegated the responsibility for managing a specific set of data resources entrusted to them by data providers and/or data owners.

Best practices for geographic data stewardship will be based on the following principles:

1. Geographic data is a valuable asset that supports the business needs of users. Its value should be maintained over time through an effective and efficient update program.
2. Organizations with missions encompassing geographic data collection should have a lead role in updating and providing that geographic data in an environment and format that can be accessed and used by a larger audience.
3. The processes supporting geographic data stewardship will be based on a clear, inclusive, and well-documented data architecture.
4. Policies, procedures, and technical processes for data update should be well documented and widely communicated.
5. Metadata providing information about the content, format, quality, authority, and availability of geographic data is vital, and this metadata should be updated along with the data itself.
6. Maintaining a high-level of geographic data quality is critical. Data should be maintained at a specified quality level that is well documented. This quality level should be met unless there is a good reason to deviate from it, and any deviations should be documented.

7. Geographic data should be shared widely among the entire user community in Colorado with proper consideration to legal and policy concerns that may restrict access and distribution.

IV.7: Data Provider

A data provider is a person or organization that functions as the primary custodian and/or owner of a data source made accessible to a wide audience of users. This includes organizations or persons with missions encompassing or requiring geographic data collection, management, or publication. Data providers should have a lead role in updating and providing geographic data in an environment and format that can be accessed and used by a larger audience, with proper consideration to sensitivity, legal, and policy concerns that may restrict access and distribution. A geographic data provider should agree upon a specific level of data quality for the data that is going to be shared, and this level should be maintained unless there is a good reason to deviate from it, and any deviations should also be well documented.

IV.8: Data Owner

A data owner is a person or organization having the responsibility and authority for an entrusted data resource. Entrusted data is data that is owned by an entity that can authorize or deny access to this data, and is responsible for its accuracy, integrity, and timeliness, and maintenance/production of record level metadata. Policies, procedures, and technical processes for the accuracy, integrity, timeliness, and satisfying standards of metadata should be documented and widely communicated.

IV.9: Primary Content Provider

A primary content provider is the entity that is primarily responsible for maintaining data available to users. In the inventory of data holdings and requirements described above, it was evident that most of the Primary Content Providers arose organically based on agency business needs rather than being appointed or mandated to act in this role. Nonetheless as data governance proceeds, the agencies should be explicitly identified as Primary Content Providers.

IV.10: Secondary Content Editor

A secondary content editor is an entity that adds content, either geometry, attributes or topology or other relationships, to the data provided by the primary content provider. This is often done for the entities' individual use or needs. However, this function may be included the stewardship process for a data set.

Section V: Governance Process and Structure

V.1: Stewardship Process

A structure for governance of geospatial data includes a standard set of steps for stewarding data sets. Some of these steps may have utility in broader governance processes, like prioritizing specific data sets for stewardship, but the intent of listing them here is to provide a well-defined enumeration of what is expected in the stewardship of data. This does not imply that a data steward should be responsible for all of these steps, but it does mean that if an agency accepts the responsibility of acting as a steward of a particular data set, it will accomplish at least a subset of these tasks. When this agency accepts the stewardship role, it will identify the level of stewardship it can accomplish by identifying which of the tasks it can complete. In this way, the stakeholders and users of this data set will have a clear idea of how the data is being stewarded, and this idea may be consistent across data sets.

The steps below are categorized into logical groupings. Again, an agency stewarding data should not necessarily pursue all of the steps in each group, but such groups do clarify the process. Some of these steps are administrative in nature while others are technical. The technical steps introduce additional infrastructure resource requirements as well as personnel requirements. The steps may be thought of as increasing levels of maturity in stewarding data from basic coordination types of activities to full data integration. The steps in data stewardship are:

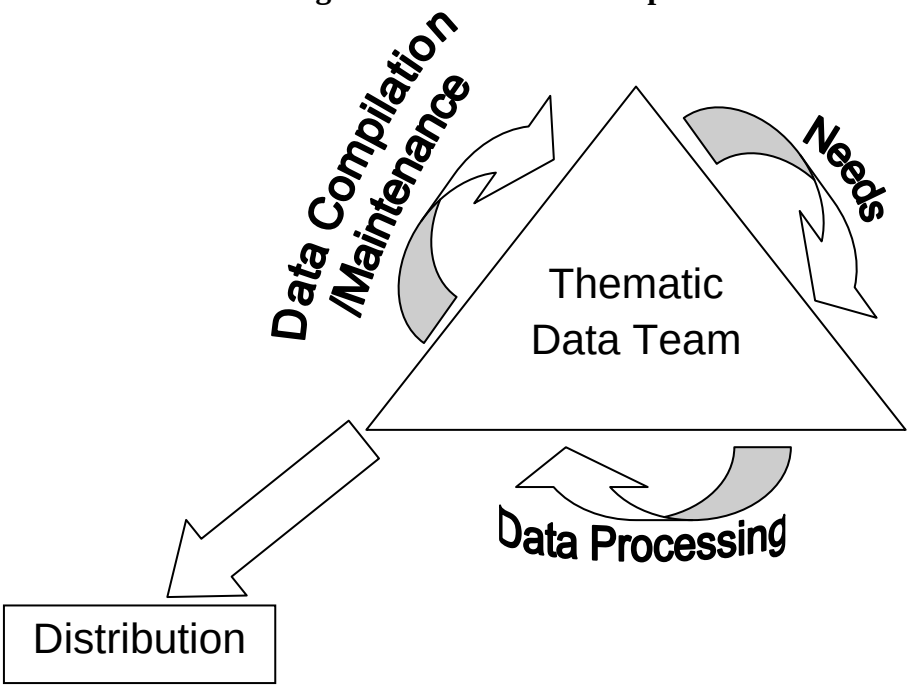
- Perform a needs assessment including identifying needs, data sources, return on investment and clarifying the value of data assets and data
- Establish and maintain a team to define processes for change management, standards enforcement, establishing decision rights, conflict management and pool funding and acquire grants.
- Perform data compilation and maintenance potentially including collecting and centralizing data, assessing the quality of data, compiling data into single data sets and integrating edits to data on an ongoing basis.
- Support data distribution potentially by publishing data to a clearinghouse but also by providing adequate descriptions of the data and its currentness and managing risk related to the data.
- Identify data governance policies and standards appropriate for data sets.

These steps do not necessarily have to be pursued sequentially. That is, the last two steps in the list may be performed at any time. However, the first two steps listed should be the first tasks undertaken when engaging in stewardship of a data set.

In actuality, the ongoing process of stewarding data will involve all of these steps in a continuous cycle as illustrated in Figure 1 below. That is, the needs for a data set are identified followed by

data collection and integration and then ongoing compilation and maintenance of the data. Throughout the process, needs should continually be evaluated and compared to the data processing and maintenance. The data can be made available at any time during this cycle depending on the requirements identified through the data governance process.

Figure 1: Data Stewardship



V.1.1: Needs Assessment

One of the first steps in data stewardship is identifying the need for and status of a data set. This aspect of the stewardship is meant to identify the potential return on investment from stewarding a data set based on the need for it and the work required to address that need. This will be critical in determining the level of stewardship the state assumes for the data sets. The needs assessment step should include the following tasks.

V.1.1.1 Identify Stakeholders

This should be done at the start of a stewardship effort to ensure that the right stakeholders are involved in future processes. The most common way to identify stakeholders is to look at who is providing inputs to data products and then look at who is receiving the outputs of these products. It is recommended that stakeholders be categorized during this step. This information will be important when prioritizing user needs, quantifying return-on-investment, and clarifying the value of the data assets. Also, this step should begin to determine the stakeholder roles and responsibilities.

When categorizing the stakeholders, it is important to recognize that some stakeholders may fit into more than one classification. Stakeholders can be generally classified into the following groups (please see Section IV: Definitions):

- Data Owner
- Data Provider
- Data Steward
- Primary Data User
- Secondary Data User
- Primary Content Provider
- Additional Content Editor

V.1.1.2 Identify needs

Surveys or focus groups or other techniques can be used to identify the business need and business objectives for different stakeholder groups. The matrix in Figure 2 below identifies some techniques that can be used to identify needs of stakeholder groups.

How to Identify Needs

Technique	On Paper	Tacit Knowledge	No Knowledge	
Stakeholder Analysis	X	X	X	People
Interviews		X		People
Brainstorming		X	X	People
Focus Groups		X	X	People
Questionnaires	X	X		People
Observation		X		Process
Gap Analysis		X	X	Process
Document Current Processes	X	X		Process
Document Future Processes		X	X	Process
Document Review	X			Documents
Product Review	X			Documents

Figure 2: Needs Assessment Techniques

This matrix categorizes the techniques into those that elicit information from people or processes or written artifacts. In addition, columns in the matrix indicate the type of knowledge that each technique may expose. This knowledge may be explicit or recorded on paper, or it may be tacit knowledge (i.e., knowledge that has not been explicitly recorded). Alternatively, there may be no existing knowledge about specific needs and the assessment techniques may actually create such knowledge. For example, by utilizing Focus Groups,

which is a technique focused on people, you identify the tacit knowledge or in fact create knowledge among the stakeholders about their needs.

Different stakeholders have different backgrounds, interests, and views. It is important to focus on the business needs, and filter out what is just wanted. The following questions can help to identify the needs:

- What data does the business function require?
- How does the stakeholder use the data?
- Where is the data used in the business?
- Who uses the product of the data?
- When does the data need to be provided to the stakeholder?
- Why does the stakeholder group need the data?

V.1.1.3 Identify data sources

Statewide data products typically represent the aggregation of many data sources. Identifying the sources of data should start by analyzing the inputs into the data product. That is, who provided the pieces to actually create the product? A data model and data dictionary should be developed as a result of this step. Additionally, a correlation between the data, the data functions, and the line-of-business should be defined at this step. An example of this would be if state highways serve as a base layer in your GIS. The Colorado Department of Transportation would be the source of data for state highways.

If it is necessary to drill down to specific sources of attributes, an analysis should be performed that traces these data elements back to the point of initial collection. An example of this would be Russian Knapweed information. The data would have been collected by a field technician that is employed by the Department of Agriculture. In this situation, the Department of Agriculture is the data source.

V.1.1.4 Calculate investment in data assets and data

Full costing for the dataset should be understood across all stakeholder groups to the extent possible. This should be a dollar figure that expresses the total cost for collecting, managing, maintaining, and using the data.

Investment should be a summation of the following:

- Person hours spent collecting, processing, and publishing the original data set.
- Consultant costs to collect, process, and publish the original data set.
- Person hours spent collecting, processing, and publishing secondary content for the data set.
- Consultant costs to collect, process, and publish secondary content for the data set.
- Average hourly wage of staff

- Costs associated with the delivery and receipt of the data set.

In actuality, it may be very difficult to obtain accurate information about these cost components. In many cases the cost of collecting data is subsumed in consultant contracts or other project costs. In these cases, the ongoing stewardship effort should progress toward making these costs explicit and measurable.

V.1.2: Establishing and Maintaining a Team

Data stewardship programs are generally composed of a several organizational bodies. They include the primary Data Steward, a stakeholder group and a data governance office or group. At the least, a successful stewardship program must be composed of a primary steward and a stakeholder group. However, best practices in data governance also call for a data governance group that can provide consensus on high level data standards and application of those standards.

V.1.2.1 Governance Groups – Stakeholders and Data Governance Group

A Stakeholder Group should consist of those who use, create and have sets of rules and requirements for the data to be stewarded. Since stakeholders are affected by data-related decisions, they will have expectations that must be addressed by the stewardship program. Also, stakeholders will have insight on issues related to data accuracy, content and use of the data within the data community. Decisions as to the direction of stewardship activities should be vetted through the stakeholder group to provide guidance for those activities.

A Data Governance Group exists at a higher level to provide direction in the application of data standards related to metadata, spatial accuracy and publication of the data. The Data Governance Group should suggest the application of these standards to all datasets while providing direction in the correct application of these standards to individual datasets. In Colorado, the work of the Government Data Advisory Board and the Colorado Enterprise Architecture Framework will provide the management and oversight functions of a data governance group. The teams described here will fit into these structures. In addition, the Colorado Geospatial Information Advisory Council will provide governance oversight from the GIS specific perspective.

V.1.2.2 Change Management

Change management applies to changes in the stakeholders group and the data. It also relates to how the stakeholders participate in decisions about changes to the data.

- Stakeholders have a vested interest in the change management process of a stewarded dataset. First and foremost is the maintenance of the integrity of the data itself. Decisions on content of and changes to the data should be vetted through the stakeholder group for feedback. The reporting structure for the decisions can be decided upon by the group. That is, in some cases stakeholder groups will be expected to be consulted before decisions are made, while others will be satisfied to

be simply informed of the changes. Change management, as it relates to stakeholders, should also describe how new members of the stakeholder group will be solicited and involved in the stewardship process.

- The data itself should have a robust change management procedure. This change management should apply to both changes to processes used to steward the data and the actual changes to the data structure or elements themselves. The change management procedures should be defined for all data stewardship efforts at the data governance level. The changes to process or the data should then be documented and published for review by the stakeholder group emphasizing where the processes and associated technologies are applied in the stewardship program for the data.

V.1.2.3 Standards enforcement

In most cases standards are enforced from the top down. A Data Steward should have complete knowledge of the various standards that apply to the dataset being stewarded. Depending on the maturity of the dataset, some standards may be self enforced while others may not necessarily apply. Regardless, it is the primary Data Steward's responsibility to see that the stewardship program is in alignment with any standards applicable to the data. Each state stewarded data set should comply with the standards described in this document as well as other potential standards defined through the GDAB or the Colorado Enterprise Architecture Framework. In addition, the stakeholder group or the data steward may identify content or other standards with which the data set must comply.

V.1.2.4 Establishing Decision Rights

Decision rights relate to potential changes to the data, its structure or its required handling or access. Establishing decision rights for a data governance program should involve the data steward and the data owner with input from the stakeholder group. The data steward and owner should present such decision rights to the stakeholder group outlining decisions that need to be vetted by the group in a data stewardship plan.

V.1.2.5 Conflict Management

A data stewardship program will always have conflicts arise regarding the authoritative content of the data and as to who should provide this content. Conflicts should be managed by the Data Steward, stakeholder group and the Data Owner, and a process for resolving conflicts should be developed early in a data governance program. In regards to spatial data, conflicts may arise between governmental agencies regarding overlaps of jurisdictional boundaries for example. Attempting to identify, plan for and resolve conflicts before they arise should be considered a best management practice for a data governance program.

V.1.2.6 Pooled Funding and Grants

Securing funding is a critical determinant of the success of any data stewardship program. Unless there is a direct business need to steward a dataset by an agency, grants and pooled

funding are generally the most common forms of funding for data stewardship. In most cases, numerous agencies benefit from a single dataset, and the idea of pooling funds to support a data governance program should be investigated. Grants are another source of funding for data governance programs. The data governance group should keep abreast of potential grants from federal, state and private entities and provide assistance to data stewardship programs in securing these types of funds. Likewise, the data governance group should provide a mechanism for agencies to pool funds for a data governance program where necessary.

V.1.3: Data Compilation/Maintenance

This set of steps deals with the technical aspects of stewarding data. Rather than the coordination related steps involving collecting, analyzing and disseminating assessments of the status of a particular data set within the state, this grouping of steps involves the process of centralizing a data set for access and use by the data stakeholders. This centralized data set may be a seamless,integrated version of local or other data, or it may simply be a central collection of disparate data sets without actually physically compiling them into one database.

V.1.3.1 Collecting and centralizing storage

This is the simplest step in centralizing distribution and maintenance of data and does not include any modifications to the geometry of the data. Nonetheless the simple process of making a large amount of data from varied sources available in one location, and the resulting reduction in overhead in obtaining that data, can be highly valuable. Clearly it is an initial step required for the subsequent steps listed below. This step may involve the following issues:

- Identifying data sources – This is generally done through the needs assessment step, but it is an ongoing process and is more rigorously documented in this effort.
- Data sharing agreements – Data sharing agreements may be necessary to make data available to multiple agencies. Various local governments have developed their own data sharing agreements, but the state has developed a template agreement with consultation from the Attorney General's Office. Ideally, this template agreement should be used to reduce the amount of negotiation of individual agreements.
- Access/distribution – As with the clearinghouse description, the purpose of collecting data in a central location is to provide access to these multiple data sets from a single location. The technology to accomplish this should follow the clearinghouse technology or utilize that infrastructure.

This step does not have to involve normalization of data attributes, but that may be a follow up to the centralization. If normalization is desired, data may be translated to a standard model through an extract, transform and load process or some other process.

V.1.3.2 Quality Assurance/Quality Control (QA/QC) of Data

This step begins the process of being more proactive in the stewardship of data sets. It may or may not include actual corrections to data, but most importantly it involves assessing and documenting the consistency, accuracy or timeliness of the data. For data sets that have multiple uses or a variety of stakeholders, it will be difficult, if not impossible, to anticipate one standard measure of data quality that would describe the fitness of the data for the varied uses. From a stewardship perspective then, it is most important that the quality is expressed in a way that allows as many users as possible to assess the fitness of the data for their use. As will be described in the standards section below, it is desirable to have a standard way of representing data quality The QA/QC step includes two tasks:

- Monitor data quality – This involves routinely assessing the quality of data according to the metrics decided on by the stakeholder group and the data steward.
- Report data quality status – The assessment of data quality should be communicated to the stakeholders and the data governance group. In particular trends or changes in the quality of the data should be highlighted.

V.1.3.3 Compiling data into single data set

While a repository of local or "sub" data sets related to a theme or layer is useful, integrating them into a single data set begins to approach a maintained authoritative data set. This step involves combining the geometric features of the smaller data sets, but it does not have to involve editing these features to correct differences in the geometry or topology. In its simplest form, it requires standardizing and combining the attributes. The tasks involved in this compilation step are:

- Combine sub data sets into a seamless geometric data set with standardized attributes.
- Assess topological, semantic and geometric differences in the data, such as overlaps or gaps in adjacent data, different definitions of features and different levels of generalization in data features.
- Edit the data to correct the differences identified in the previous point to create a topologically seamless data set.

V.1.3.4 Integrating edits to data

The step builds an ongoing maintenance process to update the data potentially from several sources. A good example of this level of stewardship is the data stewardship model for the National Hydrography Dataset (NHD), which relies on edits or updates from various "substewards." These edits are integrated into a statewide set by the State. While the process for this level of stewardship may be well defined (i.e., the logistics of how data is submitted for integration, responsibilities and accountabilities for data updates, etc.), it is possible to pursue two different approaches from the data steward's perspective.

The first collects data from substewards in a variety of formats or structures. The data steward then uses either automated or manual methods for ingesting these data into the stewarded data set. Automated methods may include extract, translate and load tools or automated manipulations to geometric objects. This is a loose data stewardship.

In this case, the data provider should disclose the characteristics and content of their data in a standard fashion. Ideally, the provider should submit a sample of actual data so that technical staff can review and verify that characteristics are in accordance with information provided in their disclosure. Based on this information, technical staff will be better able to quantify the level of effort involved both on the part of the data provider and the potential steward to incorporate substewards' data into the clearinghouse. Advanced knowledge of these characteristics can also prevent extra time and cost associated with preparing the data and processing it. At the completion of this review, if technical staff find some anomalies that will present difficulties for the integration process, a discussion of options for addressing them with the data provider will be necessary.

The second approach requires substewards to provide data in a standard format and data structure. These subdata sets can also be ingested using either automated or manual approaches. However, this stricter approach to stewardship puts more burden on the substewards for preparing the data for incorporation into the stewarded data set.

In either of these approaches, the maintenance process for a stewarded data set may either load changes or reload entire data sets. The former is the more sophisticated process and could support a temporal data set. It also will allow for tying features in the stewarded data set directly to their source in the data provided from substewards. In this case, the process for loading data into the statewide data set should enter into a change detection process where the new data is compared directly against the version of the data that the provider sent previously. Only changed features and attribution would be promoted into the staging database. The data steward would create dates and perpetual unique identifiers for each feature a provider sends, which are maintained over time. When a feature changes the original record is retired (a retirement date is set for the record) and a new perpetual unique identifier is applied to the changed feature. Even with this temporal maintenance, when a major schema change takes place it is likely all features will be retired and all new ones will be loaded. This may necessitate revisiting many of the initial processes performed with a new data provider. A change detection report returned to the provider would contain information such as: the number of records, number of changed, number of added records, and the number missing records. It will be beneficial to revisit the process periodically and to recalibrate priorities based on what has been achieved and new technological and political developments.

It may be decided that this temporal approach is too complicated for the data stewardship objectives for a particular data set, and a current snapshot of data across the state is adequate.

V.1.4: Distribution

Distribution of spatial data is typically accomplished in other states through a clearinghouse. In actuality a clearinghouse can provide for distribution of data or only discovery of data. Discovery of data is of primary importance, as users clearly need to know what data is available before considering gaining access to it. A federal example of a clearinghouse for data discovery is the Geospatial One Stop or Data.gov. Most important, the clearinghouse should manage geographic data in a way that facilitates data sharing and use by other agencies and the general public.

While the State of Colorado will be developing its own data.gov portal, this plan includes this section to identify those aspects of data distribution that are particularly relevant to geospatial data. Among these are:

- Data description
- Currentness of data
- Risk management and data security

V.1.4.1 Data Description

Metadata based on the FGDC standard described below support and present information about the data holdings in the clearinghouse including quality assessments, guidance and aids for locating and obtaining the data. If the data clearinghouse is primarily a data discovery portal, the metadata will indicate where the data may be accessed.

It is most important that the metadata allow for complete discovery of data in the most efficient manner possible. In addition, metadata content should be presented to the user in a simple manner to facilitate the user's evaluation of the appropriateness of the data for his or her use.

V.1.4.2 Currentness of Data

The temporal accuracy of a state-sponsored data clearinghouse is one of its greatest business values over purchased data. The frequency of updates to the clearinghouse depends on the likelihood of change to the data and the targets for temporal accuracy, determined by the data steward in consultation with the data stakeholder group. The update cycle for each data set should be clearly communicated in the clearinghouse.

V.1.4.3 Risk management and data security

Based on the sensitivity of data being provided, it will be necessary to implement some level of control to data assets. Data security is described in more detail below, but is mentioned here because of the importance to the clearinghouse of supporting data security and

mitigating and tracking risk to data assets. Of course, the simplest approach to this issue is separating public access to data from internal access. The state's approach to securing data of various levels of sensitivity is to store them on separate servers. However, if the clearinghouse is to provide the right to use data, with varying levels of access constraints, that are not statutorily required to be separated, role-based access and information classifications may provide adequate controls. Some examples would include: restricting the rights a user has to access information and restricting the rights a user has to perform certain functions (e.g., read/write).

When considering data security, there is sometimes a reaction to lock down entire data sets. The FGDC decision tree and other work have shown that it is possible to have a more nuanced approach to data security. For example, perhaps only certain data elements or attributes are sensitive. It will be important to approach data in a less monolithic manner than may be typical.

Classifying data involves identifying the risk to various state and other processes if the data is used or distributed inappropriately. This is one component of the FIPS199 standard. The risk may involve monetary damages, systems damage, and damage to government processes or infringement on privacy or rights of citizens. Therefore, the classification of data and proper security constraints is the first step in managing risk around the data clearinghouse. A plan should be developed for the clearinghouse that presents the possible risks and their mitigation strategies. These might include failover strategies in case the clearinghouse goes down or needs to be stopped, tracking activities to provide an audit trail if needed, and periodic reporting of other activities.

V.1.5: Data Governance and Standards Process

This step of data stewardship involves maintaining and structuring the process for ongoing stewardship. It deals with the way standards and policies are selected and monitored for their utility. A well defined set of policies and standards that establishes expectations for data collection assists decision makers in the review and acceptance of new data. Policy should also serve to establish rules of sharing data and contain a mechanism to periodically review how data is being acquired and shared to the larger community and whether user needs are being accommodated.

While standards and policies are important for ongoing maintenance of authoritative data, it is important to recognize that successful development and maintenance of data depends on solid, mutually beneficial relationships between the data stewardship effort and data providers. The long-term goals of data governance for the state should include creating agreements, processes and relationships which together will lead to the long-term improvement of data both for the providers and the distributors of data. The process to acquire and steward data should be designed to facilitate critical relationship building while providing consistency and predictability

throughout the ongoing maintenance of that layer. The objective should be to reduce the overhead in providing data as much as is feasible, both in the short-term and the long-term. This is particularly the case when a provider intends to make a significant modification in their data schema and/or format.

Several components are necessary to accomplish the efficient, predictable and collaborative process described above. They include:

- Aligning goals and benefits
- Collect, choose, review, monitor standards
- Align policies and standards (bottom up)
- Implement policies and standards
- Review, approve, and monitor policies

Each of these components are described below.

V.1.5.1 Aligning goals and benefits

The goals and potential benefits for providers of data to a state authoritative or stewarded data set should be explored so the objectives and benefits of the data governance effort can be aligned with those of the data providers.

V.1.5.2 Collect, choose, review, monitor standards

Data standards may be relevant at several levels. As described in section V.1: Stewardship Process, specific content standards for particular data sets are most salient when a data steward is active in integrating or compiling data. Such standards may define data structures, mapping or data compilation procedures, definitions of features and formats for transfer of data among others.

The FGDC has developed the Geographic Information Framework Data Standard that establishes common requirements for data exchange for seven “framework” themes of geospatial data. While the framework data standards do not specify a single structure for the interchange of data, the Geographic Information Framework Data Standard specifies a minimal level of data content that data producers, consumers, and vendors are expected to use for the interchange of framework data. Each thematic part of the Framework Data Content Standard should include a data dictionary based on a conceptual schema and contain, as appropriate, documentation of all features, attributes, and relationships as well as their definitions. This is an exchange standard for data needs and does not preclude the need for data stewards to endorse detailed data standards appropriate for a given theme.

High level data standards, such as those discussed in the survey undertaken by the working group, or procedures among state agencies are necessary for assembling data especially for a clearinghouse. The survey of standards and practices led to the following recommended possible standards:

- Metadata – All datasets should contain metadata that conform to some minimal set of mandatory elements for the FGDC Content Standard for Digital Geospatial Metadata. This standard is currently incorporated within many of the state agencies surveyed. Metadata content details various characteristics of the data description, the processing steps, attribute content and the quality of the data. Metadata is also critical for making the data set discoverable and informs users how to access the data. Metadata must also provide sufficient information to allow the user to determine if the geographic data set will meet the user's intended purpose. Specific guidance for populating the core free text fields should also be defined. In addition, the State of Colorado will be developing a data.gov portal that will rely on a limited set of metadata. The spatial components of this metadata should be consistent with the FGDC metadata standard so metadata for geospatial data can be compliant with both.
- Disclaimers – Several state agencies include disclaimer statements on their data regarding possible uses or limitations on data quality. These disclaimers tend to be at the discretion of the individual agency or individual who distributes the data. A standard disclaimer for data should be developed and vetted protecting agencies from the risk of misuse of data in a standard matter.
- Data Quality Description – A description of data quality is a required field in the FGDC metadata standards. However, the means for describing data quality is not prescribed. This leads to a variety of possible statements on data quality. Standard approaches to describing data quality should be developed so data across all agencies can be characterized and compared with regard to data quality. Such approaches may include the traditional map accuracy statements or it can be based simply on a detailed description of compilation methods or both. This description can be included in the metadata. Data quality statements may include:
    - Horizontal accuracy
    - Minimum mapping unit or data resolution
    - Compilation steps
    - Appropriate or inappropriate uses of the data
- Data Security/Access Constraints – Data security or access constraints again depend on individual agency discretion. Standard approaches to describing and classifying security constraints for data should be developed consistent with data classification methods from the Chief Information Security Officer and approaches specified by the GDAB or the state's data strategy. The FGDC has developed a rigorous approach to analyzing the potential security risks presented by geospatial data based on research

done by the Rand Corporation[6]. This "decision tree" should be incorporated into the thinking about data classification by the GDAB and OIT.

- Data Sharing Agreements – A formal data sharing agreement outlining the specific data shared, how it is shared, who may have access to it, how often data is shared, and responsibilities for updating the data are often requested by data providers. A common agreement should be used as much as possible to reduce the overhead in executing such an agreement and to ensure all parties are protected in a standard manner. In addition, clarification of who may execute these agreements is necessary to protect all involved.
- In addition to the metadata file a separate profile of data stewards and providers of data to a stewarded data set or clearinghouse should be created. This will include many of the elements expected in an FGDC compliant metadata file, but it can be used for vetting and recording providers who are approved to contribute information. The Data Provider Profile should include:
    - Name of the Organization or Jurisdiction,
    - Description of the Organization or Jurisdiction,
    - Organization Address,
    - Contact persons Full Name,
    - Contact persons address and all pertinent contact information (phone, FAX, Email etc),
    - What kind of data will be submitted,
    - What Theme (is this Road data, Hydro data, list of all possible approved data themes),
    - Types of files which will be submitted (e.g. Shape, Coverage, CAD),
    - Accuracy of data,
    - Projection (e.g. either UTM, State Plane, etc.),
    - Datum (e.g. NAD1983/HARN),
    - Frequency data will be provided

  The Data Provider Profile should be in place prior to any data being provided by that provider, and will allow the data coordinator to supplement data into missing elements of an incomplete metadata, begin the QA/QC processes, and allow the state coordinator to more effectively work with the provider to help improve their submitted data and metadata quality.
- Make data costs explicit in contracts – One of the difficulties of measuring the financial benefits of data stewardship, as called for in this plan, is accounting for current costs of data. This accounting can be extremely complex with many variables. There has been some work on Return on Investment in the GIS arena, but there are no easy answers. One of the confounding factors in this area is that data development or

[6] Mapping the Risks: Assessing the Homeland Security Implications of Publicly Available Geospatial Information. Rand Corporation. 2004.

update activities are often implicit in broader project contracts. Consequently it is impossible to isolate these costs in the contract. Data development costs, especially those related to stewarded data, should be explicit items in a contractor statement of work. Standard processes or language for accomplishing this may be addressed by the working group.

- Integrating data stewardship in business practices – A well-stewarded, authoritative data set is only beneficial if it is routinely referenced and used in state agency business practices. This requires sufficient communication efforts from the data steward and assessment of authoritative data by potential data consumers. Stewardship plans for each data set should include a communication plan for the identified stakeholders. However, potential users of the data should, as a common practice, review available authoritative data related to their projects under consideration and use these data, or at least assess the suitability of these data for their use, rather than engaging in separate data development activities. Data development through such projects should advance the stewardship of the authoritative data set rather than creating a separate maintenance activity.

V.1.5.3 Align policies and standards

Aligning policies and standards with methods for data development and management currently in use should be a regular part of the development process for new standards. However, it is important enough that it is included as a separate step here. This step analyzes the impact of standards on users or providers of the data and the potential resource requirements to adapting to the new standards. In addition, it determines whether standards should be amended based on the data providers' or users' needs.

V.1.5.4 Implement policies and standards

The implementation of these standards has a technical component and an administrative component. In fact many of the high level standards are administrative in nature. In this case the process of implementing these standards is one of determining how the standards are formalized, adopted and promulgated. This will likely be done through the Colorado Enterprise Architecture Framework or the GDAB. Of course as a first step, more specificity has to be developed for the standards identified above. The Geospatial Data Governance Work Group will undertake this specificity.

Technical standards for individual data sets will be the purview of stakeholder groups created for each data set. As described in Section V.1.2, these stakeholder groups may be composed of representatives from the variety of jurisdictions involved in the data. For example, the stakeholder group for the National Hydrography Dataset includes local, regional, state and federal representatives. To advance their standards to a statewide level, the NHD data stewards will likely introduce these technical standards through the Data Stewards Action Council of the GDAB. The implementation of such standards may start by establishing a

well-defined data model for a state stewarded or authoritative data set. As described above, such a standard may then expand to requirements for data providers.

If standards are promulgated, it will be necessary to communicate them and train users of the data in the most effective way possible. It will be incumbent on the state to provide this outreach and training.

V.1.5.5 Review, approve and monitor policies

Standards and policies developed through this program must be continually monitored and reviewed both for compliance among data users and for their effectiveness, benefits and adverse impacts. The providers and users of data should be surveyed annually regarding the work they undertake to comply with the standards, whether additional workload is created to use the data based on the standard and the magnitude of that work load. This must be compared against the documented benefit of maintaining the standard and stewarding a single data set based on the standards and policies. In addition, the level of compliance with the standards and policies or collaboration with regard to the standards and policies should be documented. All of this documentation should be provided to the State's Enterprise Architecture Office and the GDAB annually.

V.2: Governance Structure

As indicated in Section III: State Context, this plan is being developed in a complex governance and operational environment in the state. It is clear that geospatial information and technologies cut across state programs and departments, but that geospatial data does warrant some specific treatment.

The Data Stewards Action Council (DSAC) being established within the GDAB will organize data stewardship along broad subject areas and sub subject areas. These subject areas are identified in a business value matrix. A very preliminary suggestion for such a matrix is shown in Figure 3 below.

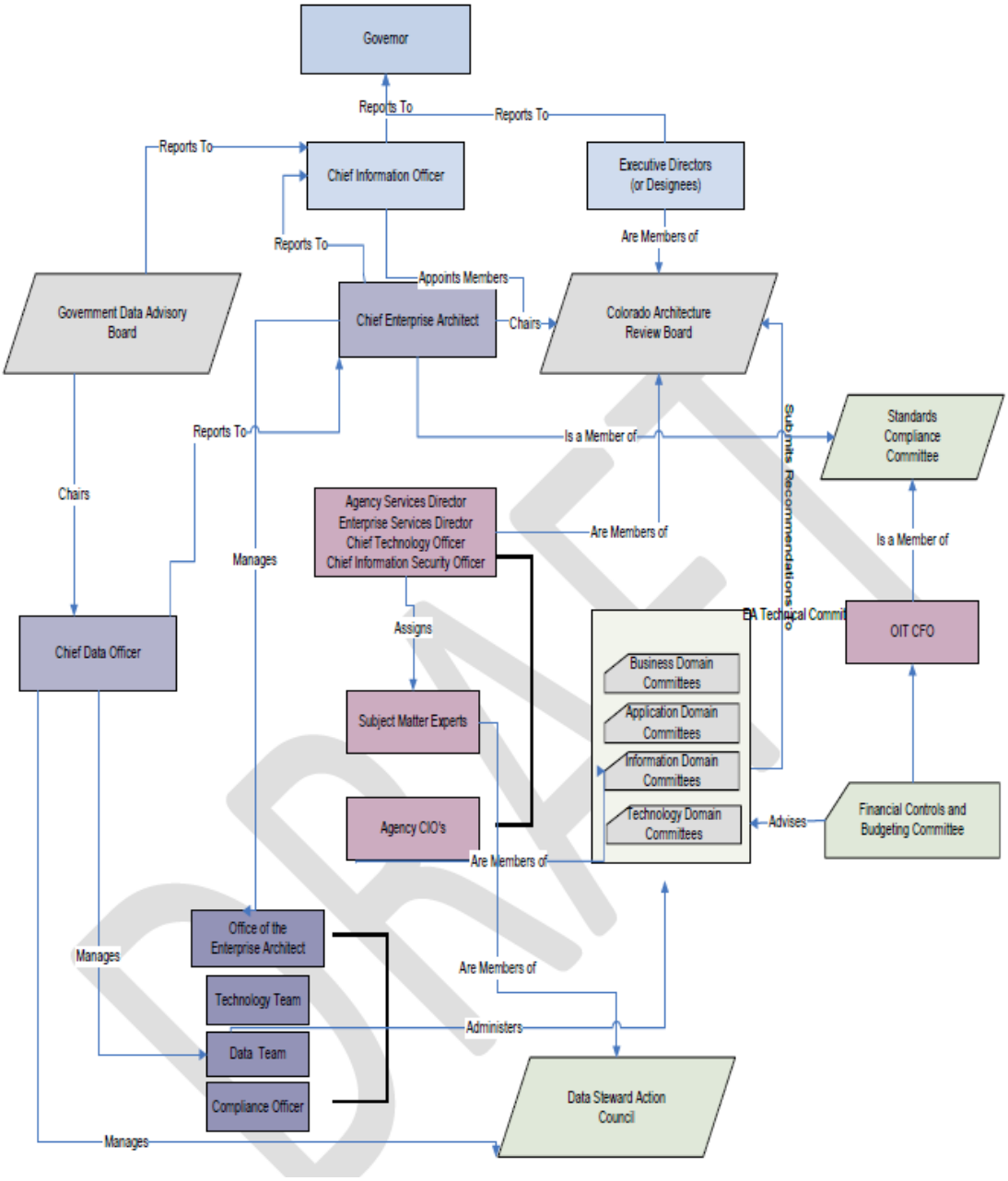
Figure 3: Preliminary Business Value Matrix

Subject Areas	Subject Area Major Domain sub-area	CDA	CDOC	HCPF	DHE	CDHS	CDLE	DOLA	DMVA	DNR	CDPS	CDPHE	DORA	CDOT	CDE	DPA	SOS	OIT	DOR
Case			X				X	X											
Filing		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Financial		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
License													X	X					X
Location																			
	Geographical	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
	Address	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Organization (Internal, External)		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Person (Employee, Client)		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Name	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Demographics	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Program/ Service		X					X			X									
	Education				X										X				
	Corrections			X				X	X										
	Health			X								X							
	Public Safety								X		X								
	Social					X							X						
	Transportation		X											X					
Project																			
Resources (depletable, non-depletable)		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Each of these subject areas may deal with data with a geographic component and their stewardship groups should therefore include somebody with knowledge of geospatial data.

The state's data strategy also defines a reporting structure within the Colorado Enterprise Architecture Framework. This structure depicts where the DSAC fits into an overall governance structure and how staff input is routed into architecture and data governance decisions. The structure is shown in Figure 4 below. Geospatial experts should be included as Subject Matter Experts (SMEs) that participate in the DSAC and should provide feedback to the various domain committees. It is not clear at this point who will comprise these committees.

Figure 4: Reporting Structure for Data Governance

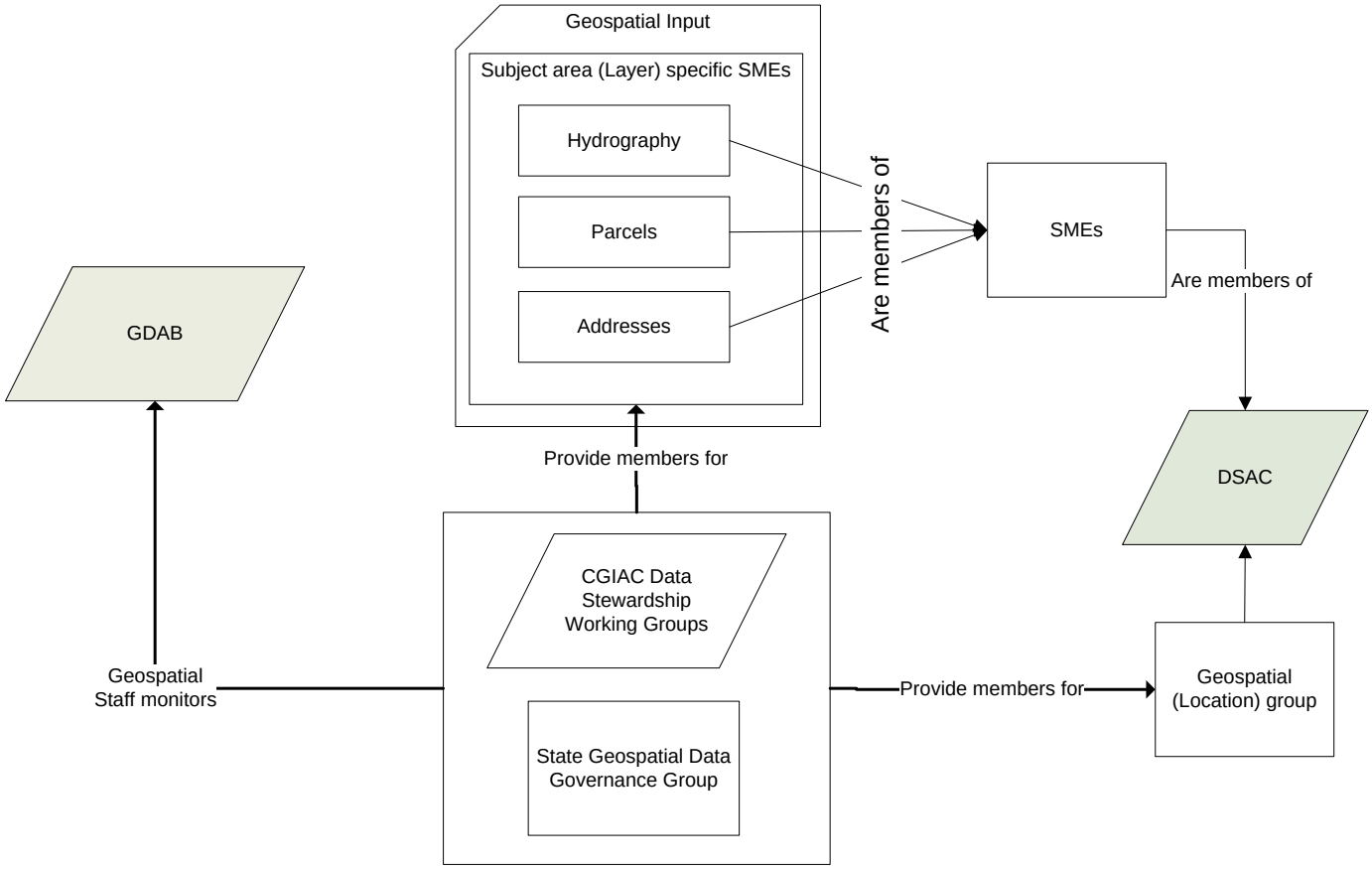


While geospatial specialists should be part of business committees or subject area committees in the DSAC, there is a need for a stewardship group for specific geospatial data sets as well. The CGIAC and the Geospatial Data Governance Working Group will address the geospatial-specific data stewardship. The state group will focus on state needs and policies from the state perspective, while the CGIAC will create specific working groups for data sets to incorporate the needs of other stakeholders. These groups will inform the Data Stewards Action Council as well.

In addition to Subject Matter Experts on the DSAC, the State GIS Coordinator will participate in the State's Enterprise Architecture team to continue to involve geospatial data in the enterprise architecture developments.

Figure 5 below depicts the recommended input into the salient groups represented in the Enterprise Architecture Reporting Structure. It shows that SMEs for individual layers or subject areas should compose overall geospatial input into the group of SMEs that are members of the DSAC. This geospatial input will inform and advise the group of subject area SMEs that make up the DSAC and that consider data stewardship that may relate to geospatial data. In addition, a SME group specific to geospatial data should be part of the DSAC. The personnel for roles will come from the CGIAC data stewardship groups and the state agency Geospatial Data Governance Group. Last, the State GIS Coordinator or other geospatial personnel from the CGIAC and the state agency group should attend and monitor the GDAB meetings.

Figure 5: Geospatial Reporting Structure to GDAB/DSAC



V.3: Communications

As mentioned earlier with regard to data standards, the data governance and stewardship process should include a plan for communicating the work undertaken and the processes and standards that are adopted. This plan can make use of a variety of modes and should reach out to all levels of stakeholders that are relevant for the geospatial data covered through these processes. In particular, this outreach effort should be proactive in communicating when the direction of the data governance plan is changed. Several bodies are already in place that should be utilized for this plan. They include:

- CGIAC
- GDAB
- GISColorado
- OIT communications, including OIT Plaza
- Other

Standard communications should include:

- New standards being considered
- Changes to existing standards
- Priority data sets for stewardship
- Methods for prioritizing data sets
- Progress on data stewardship, including availability of data
- Progress on data standards definition
- Updates to inventories or other base information

V.4: Data Security

Security of geospatial data should be approached in a manner similar to security for other data. The state's cybersecurity office is promulgating policies for classifying and protecting data. However, recent experience with geospatial data has led to a more nuanced approach to security than a binary restricting or allowing public access to data. The Federal Geographic Data Committee has published a decision tree to promote a standard approach to deciding whether data should be restricted or not. In addition, it is possible to discuss restricting access to specific attributes within the data rather than a data set in its entirety.

Suggestions about data security should be provided to the state Chief Information Security Officer and the Chief Enterprise Architect.

Section VI: Data Set Stewardship

VI.1: Prioritizing data sets for stewardship

The GIS Data Governance working group first identified various factors that should be considered when prioritizing data sets for stewardship. These factors are:

1. Need (The inventory of state data sets can be used to identify the need. Until such an inventory or assessment is done for local and federal data needs, we assume the state parallels local and federal needs.)
2. Funding/resources
3. Availability of data/# contributors
4. Complexity of data
5. Frequency of update
6. Identifiable steward

These criteria are guidelines, but they should not be applied inflexibly and in strict order as they are listed above. For example, there may be instances where stewardship for a particular data set is initiated before other data sets of higher need due to funding availability.

The working group considered the individual data sets in the data inventory and applied a priority classified into high, medium and low priority, to each data set as shown in Table 3. The prioritization generally followed the need ranking, but also considered if the layers were specifically the domain of a particular agency and therefore stewarded already, derivative products, or as mentioned, important to a grant or other funding agency.

Table 3: Priority Ranking of Data Sets

GIS Layer	# Depts Maintaining Data	# Depts Additonally Editing Content	# Data Consumers	# Depts Requiring Data	Total # Requiring & Consuming	Priority
Local Roads	1	1	10	4	14	H
Major Roads	1	1	9	5	14	H
Highways	1	1	9	4	13	H
National Hydrography Dataset	1	3	12	0	12	H
Orthophotographs (6 - inch color)	0	0	10	2	12	M
NAIP	0	0	3	9	12	H
County Boundaries	2	0	4	7	11	H
Census Blocks	1	0	5	5	10	H
Streams and Rivers	1	3	6	3	9	L
Municipal Boundaries	2	0	2	7	9	H

GIS Layer	# Depts Maintaining Data	# Depts Additonally Editing Content	# Data Consumers	# Depts Requiring Data	Total # Requiring & Consuming	Priority
Colorado State Boundary	1	0	1	8	9	L
Rail	1	1	7	1	8	L
Elevation	0	0	4	4	8	M
Populated Places (Derived)	1	0	2	6	8	M
Ortho-imagery	0	0	6	1	7	M
Topography	0	0	3	4	7	M
County Parcels	0	0	1	6	7	H
PLSS	0	0	1	6	7	H
Flood Zones	1	0	5	1	6	M
Lakes and Ponds	1	1	4	2	6	L
Fast Traks (proposed)	0	0	4	1	5	L
Fire Boundary Areas (historic or existing?)	0	0	4	1	5	L
Drainage Basins	1	0	3	2	5	L
Government Areas/Units	2	1	3	2	5	M
Water Quality data (CDPHE, CDA, other)	0	0	3	2	5	L
Oil and Gas Well Locations	1	0	3	2	5	L
Aquifers	1	0	2	3	5	L
Communication Towers	0	0	2	3	5	L
Water Division Boundaries	1	0	4	0	4	L
DOW Administrative Boundaries (3 layers)	1	0	3	1	4	L
Land Use	0	0	3	1	4	M
Wetlands	0	0	2	2	4	M
Public Airports and Air strips	1	1	2	2	4	M
LIDAR	0	0	2	2	4	L
Soils	1	0	2	2	4	M
Tribal Lands	0	0	1	3	4	L
US Congressional Districts	0	0	1	3	4	L
Drinking Water	1	0	1	3	4	L

GIS Layer	# Depts Maintaining Data	# Depts Additonally Editing Content	# Data Consumers	# Depts Requiring Data	Total # Requiring & Consuming	Priority
facilities						
Land Fills - current and historic - points and polygons	1	0	1	3	4	L
Police Stations	0	0	1	3	4	M

VI.2: Data Sets To Be Considered Long Term

The data sets with high and medium priorities in Table 3 can be considered the data that should be addressed over the next three years. These data sets are shown in Table 4.

Table 4 : High/Medium Priority Data Sets

GIS Layer	# Depts Maintaining Data	# Depts Additonally Editing Content	# Data Consumers	# Depts Requiring Data	Total # Requiring & Consuming	Priority
Local Roads	1	1	10	4	14	H
Major Roads	1	1	9	5	14	H
Highways	1	1	9	4	13	H
National Hydrography Dataset	1	3	12	0	12	H
County Parcels	0	0	1	6	7	H
PLSS	0	0	1	6	7	H
NAIP	0	0	3	9	12	H
County Boundaries	2	0	4	7	11	H
Census Blocks	1	0	5	5	10	H
Municipal Boundaries	2	0	2	7	9	H
Elevation	0	0	4	4	8	M
Populated Places (Derived)	1	0	2	6	8	M
Ortho-imagery	0	0	6	1	7	M
Topography	0	0	3	4	7	M
Flood Zones	1	0	5	1	6	M
Government Areas/Units	2	1	3	2	5	M
Land Use	0	0	3	1	4	M
Wetlands	0	0	2	2	4	M
Public Airports and Air strips	1	1	2	2	4	M

Soils	1	0	2	2	4	M
Police Stations	0	0	1	3	4	M

Of these data sets, some have actual or *de facto* stewards, while there are likely stewardship candidates for others. These agencies currently acting as stewards may not have developed a formal stewardship plan as defined by this governance plan though. The data sets and their stewards are listed in Table 5 below. Notice that stewards have not been identified for some of the data listed in Table 5.

Table 5: Data Set Stewards

GIS Layer	Steward
Local Roads	CDOT
Major Roads	CDOT
Highways	CDOT
National Hydrography Dataset	DNR
County Parcels	DNR (DNR is taking on the role of coordinating this data for state agencies. There has not been official agreement as to their stewardship role. They will be assisted by OIT)
PLSS	
NAIP	OIT
County Boundaries	CDOT (CDOT is a *de facto* stewards of this data set for the state as they update it annually, and most GIS users in the state use CDOT's data. DOLA is a likely candidate to steward the data given their statutory responsibility to maintain a record of changes to county and municipal boundaries, but they do not have the resources to dedicate to this stewardship. CDOT may consider their role as steward with coordination assistance from OIT).
Municipal Boundaries	CDOT (CDOT is a *de facto* stewards of this data set for the state as they update it annually, and most GIS users in the state use CDOT's data. DOLA is a likely candidate to steward the data given their statutory responsibility to maintain a record of changes to county and municipal boundaries, but they do not have the resources to dedicate to this stewardship. CDOT may consider their role as steward with coordination assistance from OIT).
Census Blocks	DOLA
Police Stations	CDPS (along with fire stations and other public safety facilities)
Elevation	
Populated Places (Derived)	
Ortho-imagery	
Topography	

Flood Zones	
Government Areas/Units	
Land Use	
Wetlands	
Public Airports and Air strips	
Soils	

VI.3: Data sets covered in this plan

Of the data sets listed in Table 5, stewardship programs for two data sets are under development and are fairly mature. They are:

- National Hydrography Dataset (NHD)
- "Community Anchor Institutions" (i.e., police stateions, fire stations, emergency medical services, health care, education and government facilities)

This plan includes stewardship plans for them as attachments and follow the rubric described in this plan. Stewardship for two other types of data, statewide parcels and roads, are just beginning, and plans will be developed for them and inserted into the framework of this plan in the next six months. The NHD and Community Anchor Institutions have been chosen as the first stewardship efforts because of a combination of their importance to the state and the availability of resources to support their stewardship.

Section VII: Conclusion

Geospatial data is a valuable asset to the State of Colorado and as such the creation and maintenance of accurate and complete state wide data sets is a business imperative. This document covers the reasons for governing state wide geospatial data, outlines a standard process for governance, identifies constraints as well as defining measures of success, and takes two state wide data sets through the process. However, it is important to understand that this is not static. It will evolve as additional data sets are included in stewardship programs, our understanding of the value data governance and effective mechanisms for data stewardship mature and the State's enterprise data governance process grows. This plan outlines a combination of technical and administrative steps that constitute a complete stewardship plan for an accurate, complete, timely, secure and authoritative source of data that can be utilized across state government and beyond.

This Geospatial Data Governance Working Group will continue to assess data stewardship activities in state government and compare the progress in data stewardship to the metrics outlined in this document to assess the overall success of these stewardship efforts. It will also identify resources that may be necessary to promote further data stewardship.

Appendix A

Stewardship Plan
National Hydrography Dataset (NHD)

Section I: Introduction

In the spring of 2007, the United States Geological Survey (USGS) held a conference in Lakewood, CO. detailing Data Stewardship and the National Hydrography Dataset (NHD). As a result of this conference, the Colorado Division of Water Resources (CDWR) received a grant to perform a pilot study to detail the level of effort for the State of Colorado to maintain/steward the NHD. The final report of this study also details recommendations and findings that would be applicable to a Stewardship program. As a result of this report, CDWR received a second grant from the USGS to initialize Stewardship of the NHD in Colorado over a three year period beginning July 1st, 2009 and ending June 30th 2012. CDWR intends to continue stewardship activities after the grant period ends and promote the use of the NHD in Colorado.

I.1: What is the NHD?

The National Hydrography Dataset (NHD) is the surface water component of *The National Map*. The NHD is a comprehensive set of digital spatial data representing the surface water of the United States using common features such as lakes, ponds, streams, rivers, canals, and oceans. These data are designed to be used in general mapping and in the analysis of surface-water systems using geographic information systems (GIS). In mapping, the NHD is used with other data themes such as elevation, boundaries, and transportation to produce general reference maps. Customized maps can be made to meet specific needs of the user by emphasizing certain aspects of the data. A map emphasizing hydrography can be produced by displaying more of the content embedded in hydrography.

The NHD often is used by scientists, specifically in surface-water analysis using GIS technology. This takes advantage of a rich set of embedded attributes that can be processed by a computer system to generate specialized information. This information can then be portrayed in specialized maps to better understand the results. These analyses of hydrography are possible largely because the NHD contains a flow direction network that traces the water downstream or upstream. It also uses an addressing system to link specific information about the water such as water discharge, water quality, and fish population. Using the basic water features, flow network, linked information, and other characteristics, it is possible to study cause and affect relations, such as how a source of poor water quality upstream might affect a fish population downstream.

I.2: Why Steward the NHD?

The most important benefit of stewarding the NHD is that it will provide a single and all-encompassing source of hydrography. Historically, individuals and agencies used various forms of hydrographic data. These datasets were often produced at different scales and from different sources. Some agencies chose hydrography best suited for mapping, and others used hydrography with reach codes or stream order information for modeling. Many agencies edited the data to meet their own needs and projects, and those changes were not returned to the source data unless there was a special funding to do so. Even then, it was difficult to determine what the "best" hydrography was since most of it was tailored for a specific use.

Section II: Need for and Status of the NHD

NHD stewardship encourages all NHD users to make official edits and submit those to the USGS distributed NHD so all can benefit from that work. As long as stewardship is managed properly, duplication in edit efforts is removed. By funneling all edits into a single dataset, the process is more efficient. The following programs and applications would be improved by using the NHD.

- CDWR's Water Rights Administration
- Colorado Water Conservation Board's (CWCB) Instream Flow and Lake Level Program
- Flood protection and Stormwater Management
- CDWR's Dam Safety Program
- Surface Water Quality Assessments
- Community Land Use and Transportation Planning
- Aquatic Habitat Monitoring

II.1: Who is Maintaining the NHD

The NHD is not a static dataset and is changing rapidly in Colorado and the nation. These edits are usually performed by either the USGS or a State Steward that has a defined stewardship agreement with the USGS. In Colorado's case, CDWR is acting as the primary content provider for the NHD and is soliciting data content from stakeholder groups statewide as well as adding State of Colorado data to the NHD.

Section III: Stakeholders

USGS
United States Forest Service (USFS)
Colorado Division of Water Resources
Colorado Water Conservation Board
Colorado Division of Wildlife
Colorado Department Public Health and Environment

Governor's Office of Information Technology
Northern Colorado Water Conservancy District
Aurora Water
Denver Water
Colorado River Water Conservancy District
Central Colorado Water Conservancy District

Section IV: Stewardship Team and Roles

CDWR is the primary steward of the NHD in Colorado and works under a cooperative agreement with the USGS to act as the Steward. CDWR has formed a steering committee, composed of the above stakeholders, to help provide direction for stewardship activities. These activities are currently funded through a partnership grant from the USGS and are outlined in a scope of work attached to the grant.

Section V: Stewardship Process

V.1: How the NHD is Maintained at the State Level

Currently, CDWR checks out the data from the USGS NHD Stewardship website and defines specific stewardship activities that are going to be performed on the data during this maintenance session. This website notifies all parties associated the data has been checked out and locks down the data. CDWR is currently on Phase 1 of its maintenance plan for Colorado and will begin Phase 2 in the fall of 2010. After the maintenance session is completed and QC has been run by the State, the data is uploaded to the USGS and additional QC is performed. If the data passes USGS QC checks, it is uploaded to the national database and is available for download by the public.

V.2: Update Frequency

The NHD is under continuous update.

V.3: Distribution

Currently distributed by the USGS
http://nhd.usgs.gov/

Section VI: Data Standards

Data standards for the NHD are defined by the USGS.
http://rockyweb.cr.usgs.gov/nmpstds/nhdstds.html

VI.1: Data Model

Data model is defined by the USGS.
http://nhd.usgs.gov/NHDv2.0_poster_6_2_2010.pdf

Section VII: Communication Process

CDWR currently holds biannual NHD steering committee meetings to update the stakeholders and public on the current maintenance activities and plans for upcoming activities. This is currently done through email and phone.

Appendix B

Stewardship Plan
Community Anchor Institutions

Section I: Introduction

This plan outlines stewardship procedures and expectations for a data set of Community Anchor Institutions (CAIs). These institution locations are critical for a variety of uses in the state and are of interest to local, state and federal entities as well as private citizens and academia. This work has been motivated by federal grants for mapping broadband service, but will benefit multiple applications. Some of the stewardship planning and testing has been funded by the USGS as well for maintenance of "structures" data, in USGS language, for homeland security.

I.1: What are CAIs?

The National Telecommunications and Information Administration defines CAIs to include health care, public safety, education and higher education facilities, as well as government buildings. The last group of facilities is open ended and should be refined. Consequently, the CAI work group is defining CAIs as the following types of features to start:

- Health care facilities
- Police stations
- Fire stations
- Emergency medical services locations
- Public schools
- Private schools
- Universities, colleges and community colleges
- Correctional institutions
- County courthouses
- City halls
- Emergency operations centers
- Public safety answering points
- Regional and local dispatch centers

Other facility types may be added in the future as this data is developed and stewardship proceeds.

I.2: Why Collect CAI Data

CAI information is used for a variety of activities. These activities include:

- Emergency service dispatch

- Broadband service identification
- Emergency management and response
- Managing, planning and optimizing government activities
- Managing government facilities and resources
- General information to citizens and government personnel
- Others....

Several of these uses require this information at a state or regional level, so a statewide stewardship program for these data is necessary to support these efforts that do not fall within a single jurisdiction.

Section II: Need for and Status of CAI Data

Community Anchor Institutions data is used for multiple purposes. This is due in part because it is such a broad collection of facilities, and various lines of business clearly need data about their particular facilities (e.g., Education requires data about schools, Public Safety requires data about law enforcement facilities). In addition, some business lines require information about the entire set of facilities, for example emergency response and dispatch.

II.1: Who is Maintaining CAI Data

Within State government, two entities maintain data about corresponding to the primary line of business. The Department of Public Health and Environment maintains data about health care facilities. The Department of Education maintains data about schools and libraries. In addition, the Department of Public Safety is involved in maintains some portions of data for multiple groups of these facilities. Like many other state agencies, the Department of Public Safety is interested in most of the facilities and collects information on them from various sources for their dispatch operations.

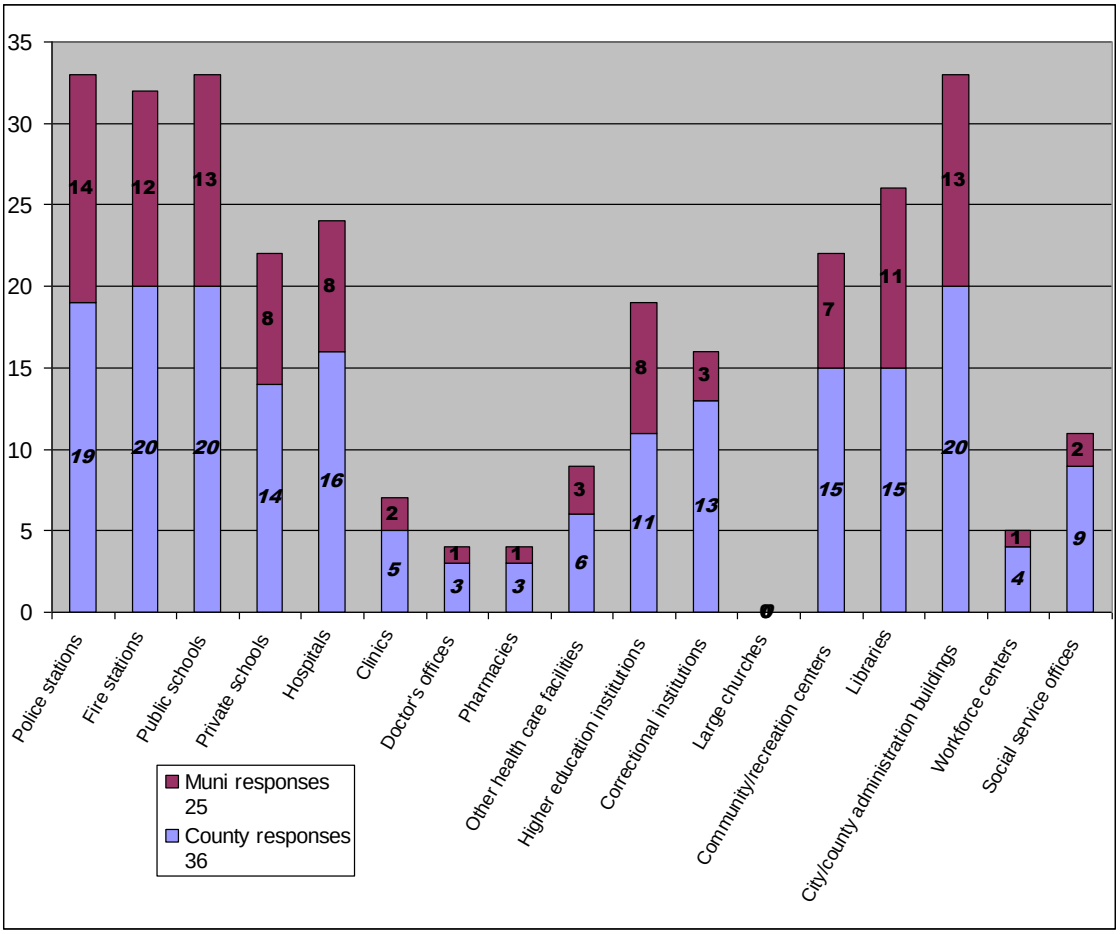
At the local level, many counties and municipalities maintain these data. A recent survey by the state received responses from 25 municipalities and 36 counties. Most of these respondents indicate that they maintain information about at least one category of CAIs.
The CGIAC is continuing to pursue more responses to the survey.

Of the 36 counties that answered the survey, 20 of them maintain data about CAIs. The most common institution types for which counties maintain data are public schools, fire stations and county administration buildings. Police stations are very close, and none of the counties indicated that they would not share the data.

16 of the 25 municipalities collect some sort of CAI data. Again, fire stations, police stations, public schools and administration buildings are the most commonly maintained data sets. The

bar chart in Figure 1 shows the number of counties and municipalities that collect the various types of CAI data. The list of counties and municipalities and the data that they collect is included as an attachment to this stewardship plan. Figure 2 maps the county responses for the most common types of CAI data.

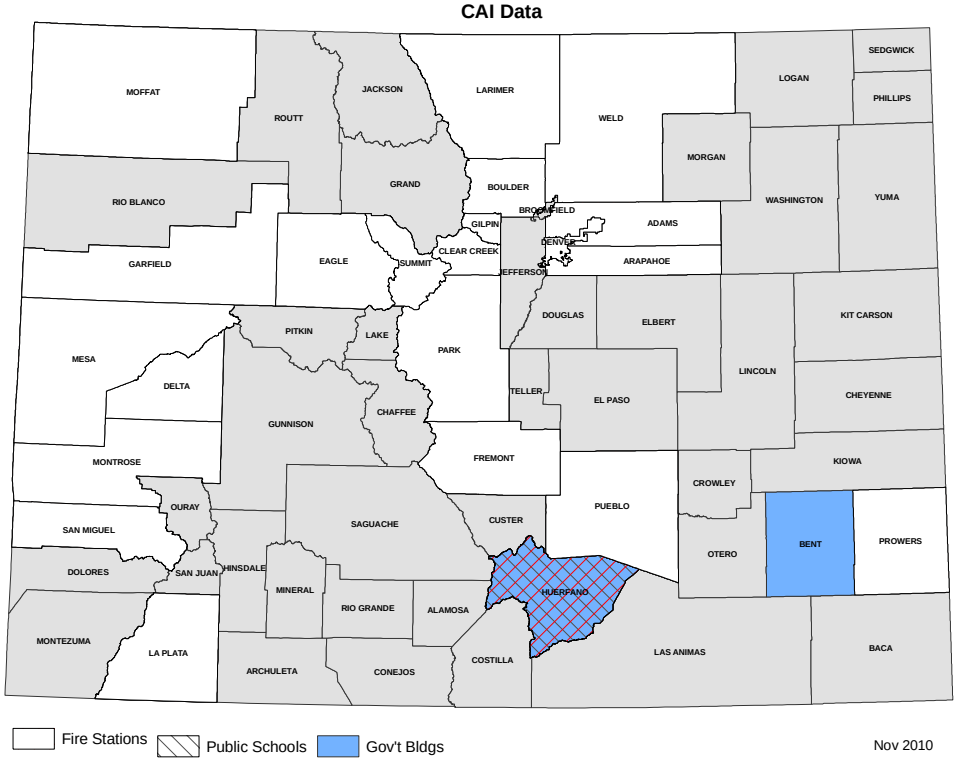
Figure 6: County and Municipal CAI Data Collected



In addition to local and state maintenance of these data, several federal agencies maintain versions of the data. Most notable among these agencies are FEMA for use in the HAZUS software, the Department of Homeland Security through their HSIP program and the USGS through the National Map and the Geographic Names Information System.

Figure 7

County Responses to CGIAC Data Survey
CAI Data



II.2: How CAI Data is Used

The earlier section of this plan identified some of the uses of these data. To understand the specific needs and therefore how the data should be structured, further detail related to the use of the data is warranted.

- Emergency dispatch – The locations of these institutions are used to direct emergency responders to an incident. That is, responders may be dispatched to a specific facility (e.g., school). Additionally, the location of an incident may be reported through an emergency call at one of these locations. In the case of dispatch, it is possible that the location is provided to a responder as a particular ingress/egress point for the institution. In addition to the location, contact information for the facility is very important for this use.
- Broadband mapping – In the case of broadband mapping, the location of the institutions is also reported with the broadband service to that location. This may be compared to general geographic information about broadband service across the state. This use of the data requires more generalized location than the emergency dispatch use.

- Emergency management and response – Emergency management and response uses of these data is concerned with identifying what facilities, and potentially special needs populations, may be at risk from a anticipated or actual hazard. The location data required for this use must be reasonably accurate, but again does not require the precision required for dispatch purposes. Contact information and potentially information about the populations or capacities of the facilities is important for this use.
- Managing, planning and optimizing government activities – As the State of Colorado consolidates information technologies and looks to optimize its IT operations, location of government facilities and the various services and IT resources (i.e., internet access points, desktop support, etc.) provided to these facilities is very useful. These locations and the source of the services may be compared to determine if there are redundant services to these locations or if there are ways to improve the efficiency of delivering services to the locations.
- Managing government facilities and resources – Knowing the locations of these facilities provides basic information for managing them efficiently from the perspective of facilities management. Again, this is particularly relevant at the state level for government buildings.
- General information to citizens and government personnel – Citizens are often interested in what types of institutions are close to them or within a certain distance of a location. Locations of CAIs allow for this comparison to such a spatial search. In some cases, ancillary information about the institutions may be of interest. For example, a family might want to know about school performance (e.g., CSAP scores) at schools near their home.

Section III: Stakeholders

The CAI stewardship team has identified the following agencies/groups as having an interest in CAIs. They are color coded according to their interest.

CIAC
Emergency Managers/EOCs
911 dispatch centers
Denver Regional Data Consortium (Planimetric subcommittee)
National Map
HAZUS/FEMA region 8
FCC/NTIA
DPS
CDPHE
OIT
DOLA Regional Managers, DEM

Dept. of Ed.
Fire Districts (Community Wildfire Protection Plans), Fed agency in Bent County?
Tourism boards
Economic Development councils
DHS
Utilities
School districts
Transit companies
Realtor/real estate companies

..... = data consumers
= data producers
= data consumer & editors

Section IV: Stewardship Team and Roles

The stewardship team for CAI data will build on the CAI working group in the CGIAC. Participants in this working group are:

Craig Barraclough – Park County
Susan Dellinger – City of Louisville
Carol Giffin -- USGS
Matt Goetch – City of Montrose
Jon Gottsegen – Colorado Office of Information Technology
Cindy Jones – Bent County
Bill Lucatuorto – Xcel Energy
Pete Magee – San Luis Valley
Mike Nath – CO Office of Information Technology/CO Department of Public Safety

This group represents state personnel and CGIAC members as well as likely regional substewards of the data. This group will initially decide on a strategy for stewardship and standards for maintenance process and data structures.

Mike Nath, in his role within the dispatch unit at the Department of Public Safety will serve as the data steward for CAIs. This role will involve integrating data into a standard data model and coordinating future stakeholder participation and deliberation on data models and protocols. He will be assisted by Jon Gottsegen.

Section V: Stewardship Process

V.1: How CAI Data is Maintained at the State Level

CAI data will be maintained through a local or regional integration approach. Data from the local level will be passed to the state and ingested into the state database. Ideally, these data will be assembled regionally before being transmitted to the state. There are already several regional initiatives in the regard under way. They include the Southwestern part of the state, the San Luis Valley and southeastern counties in the state and the Denver region.

The following steps will be followed for maintenance of CAI data:

- Data among the regions will be developed in the manner that best suits their needs. The regions may convert the data to the state format and data structure.
- The state will maintain point data for CAIs and follow a standard data model.
- The state CAI data steward will receive data through a request or regular "push" of data from a regional "substeward" to the state and translate and load the data into the state data set.
- In the future, an automated process for receiving the data and performing the translate and load process may be pursued.
- The state data steward will report to stakeholders including the CGIAC about updates to the data, changes to the data model and status of data collection.

V.2: Substewards for CAI Data

To this point, the State has identified the following regional substewards for the state:

Organization	Counties/Area Covered	Name
Montrose Regional E911 Authority	Montrose, San Miguel, Ouray, Hinsdale, portions of Gunnison	Matt Goetsch
Sangre de Cristo Resource Conservation and Development District	Park, Lake, Chaffee, Fremont, Custer, Pueblo, Huerfano, Las Animas	Craig Baraclough
San Luis Valley	Saguache, Rio Grande, Mineral, Conejos, Alamosa, Costilla	Pete Magee
Southeast Colorado	Bent, Prowers, Otero, Baca	Cindy Jones
Denver Regional Council of Governments	Denver, Boulder, Broomfield, Adams, Arapahoe, Douglas, Jefferson	Matt Krusemark
Mesa County	Mesa	Tristan Nelson

V.3: Update Frequency

The State will request updates from local stewards on a semi-annual basis.

V.4: Distribution

The basic CAI locations and core attributes will be considered Tier 1 in the State's information security classification and made available to the public through data download and a web service. It will also specifically be distributed to the following entities specifically:

- Substewards
- USGS
- US Department of Homeland Security

Section VI: Data Standards

The data will be developed as an ESRI point geodatabase. The precise location of the points at the state level (e.g., whether the points are place on the building, entrance to the property, building entrance, etc.) is not critical. The coordinate system of the data will be UTM zone 13, NAD83.

VI.1: Data Model

The attributes for the CAI data will include:

Field	Type	Values
ID	String	Still being determined
Address	String	
City	String	
Zip	String	
Type	String	Health & medical Education Gov't administration Public landmark Transportation Utilities Private Industry Public Safety/Emergency Services
Source Organization	String	
Source Name	String	
Source Phone	String	
Creation Date	Date	

Upload Date	Date	
Data Compilation/Quality	String	Professional grade GPS Consumer grade GPS NAIP aerial photography Local/higher resolution photography Point address geocoding Centerline geocoding Parcels Digitize from quad maps Other Don't know

VI.2: Quality

The quality of each feature in the data will be described in a field as shown above. The metadata for the data set will describe the range of quality among all of the features in the data based on the enumeration of values listed above.

Section VII: Communication Process

Information about the development of CAI data will be communicated at least semi-annually with the substewards and specific stakeholders. In addition, the GIS community will be informed at meetings of GIS Colorado, through the GIS Colorado listserve and through the OIT web page.