Strategic Pathway 4

DATA

*This **strategic pathway** establishes a geospatial data framework and custodianship guidelines for best practice collection and management of integrated geospatial information that is appropriate to ensure cross sector and multidisciplinary collaboration.*

*The **objective** is to enable data custodians to meet their data management, sharing and reuse obligations to government and the user community through the execution of well-defined data supply chains for organizing, planning, acquiring, integrating, curating, publishing and archiving geospatial information.*

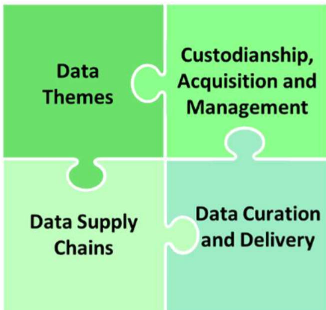
Summary

Geospatial data is the foundation on which governments base many decisions. It is used in policy development and in the provision of government services. Its use is growing exponentially across all sectors for e-commerce, business intelligence to make timely and accurate decisions, and to inform policy.

Having access to the right data and at the right time is crucial to good decision-making. It is data that provides new levels of insight into our past, present and future. For this reason, governments, businesses and the community need to know they are using the most accurate and authoritative data for planning, analysis, navigation and visualization – good data underpins good decisions.

As the amount, variability and availability of data rapidly increases, the requirements for 'organized' geospatial data holdings have never been more important. Geospatial data has grown in use across almost every market and institution. Every part of government creates and consumes geospatial data. It is a nation's 'digital currency', an asset that must be properly governed, designed, and managed to provide enduring consistency and completeness in quality, accuracy, security and use.

An ecosystem that fosters the proper collection, acquisition and management of geospatial data, leads to cutting-edge innovation and revolutionary methods across a range of sectors. Advances can already be seen in the health, financial, disaster management and transportation sectors; where geospatial data is enabling the discovery of new patterns and influences by combining geography, and social and cultural norms in a way that reveals new knowledge through enhanced visualizations, analysis and traceability.



Common to all government and business applications are four key elements associated with data coordination that need to be achieved to enable an environment where innovation, and pioneering research and development can thrive. These four elements are:

- **Data Themes** – the organization of priority national data themes, aligned to the globally endorsed fundamental geospatial data themes.
- **Custodianship, Acquisition and Management** – leading to responsible collection, management, maintenance and dissemination of fit-for-purpose geospatial information.
- **Data Supply Chains** – and interlinkages that support cooperative data sharing and integration.
- **Data Curation and Delivery** – enables enduring accessibility and value of data, and an information resource for broader usage across all sectors.

These elements are underpinned by principles that promote consistent data management, sharing and reuse so that data custodians may meet their obligations to government and the user community. These principles are put into practice through strategic actions that deliver and strengthen effective integrated geospatial information management. Tools, such as matrices, examples and checklists, are provided in the appendices to assist countries to work through concepts and processes to successfully complete each action. The overall structure for data is illustrated in and anchored by Figure 4.1 below.

When implemented the actions (and their interrelated actions) will enable the achievement of the four elements, which in turn will deliver significant and sustainable outcomes and benefits for a country. These outcomes include attaining:

- Increased range and scope of authoritative, integrated geospatial data available for decision-making and policy-setting to address economic, social and environmental challenges;
- Critical mass of centrally coordinated data discovery to support national development and innovation, leading to economic growth and improved quality of life for citizens;
- Cost reduction through productivity improvements achieved via well-defined data supply chains that eliminate duplication and ensure standardized data is accessible to end users for integration and reuse;
- Ability to monitor and measure progress towards achieving broad socio-economic benefits, including the sustainable development goals, through access to quality geospatial information.

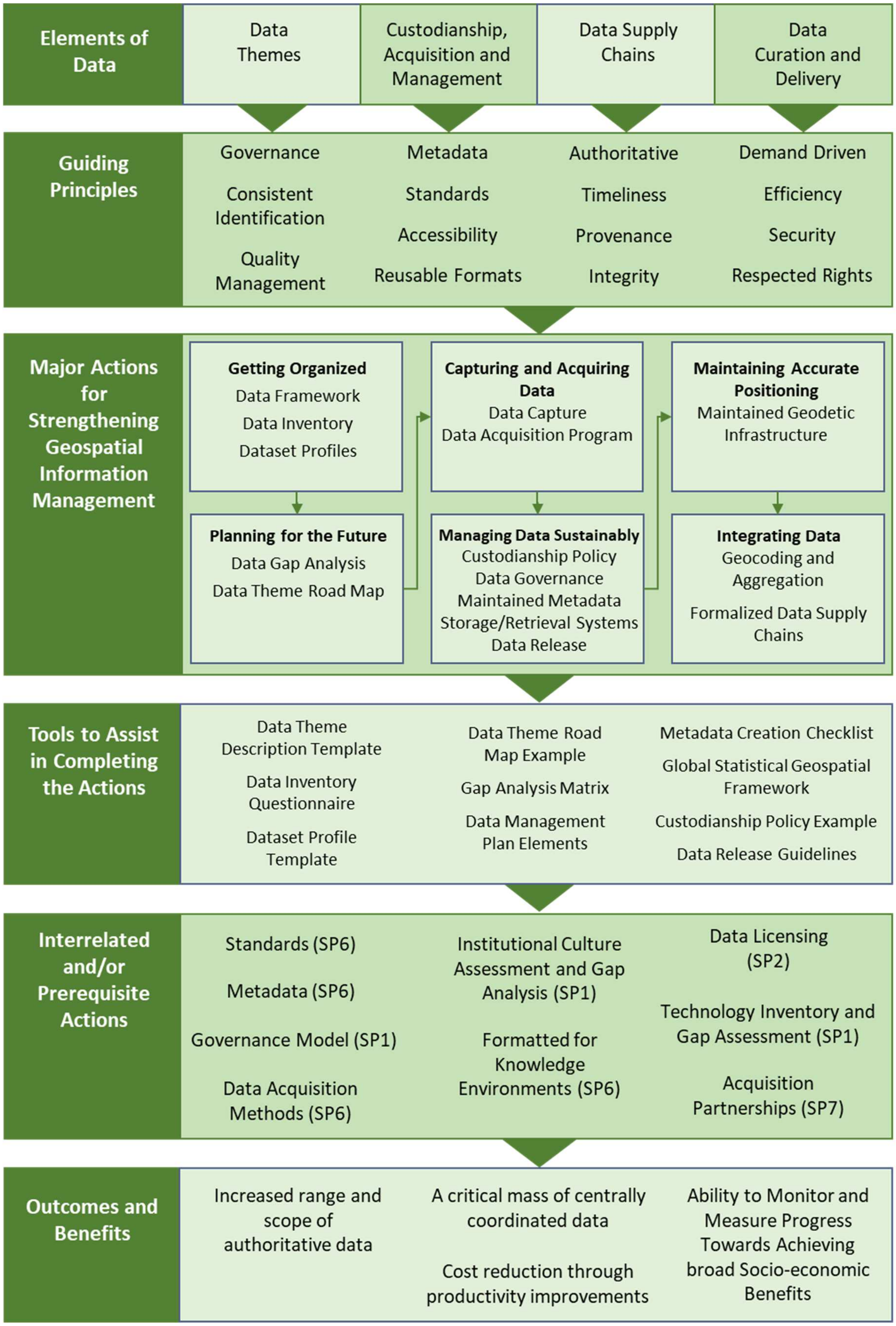


Figure 4.1 *The overall structure for data - showing the four key elements, guiding principles, actions and interrelated actions, and the tools provided in the appendices to support the achievement of outcomes.*

4.1 Introduction

With so much geospatial data becoming increasingly available, it is important to have guidelines and a framework to manage it consistently.

Geospatial information[1], reflects the physical world in which all human, economic and environmental activity takes place, and provides the digital version of our world - without which a digital economy is not possible. Geospatial information describes the physical location of geographic features and their relationship to other features and associated statistical information. Geospatial information is presented in many forms and mediums including maps, satellite imagery and aerial photography. describes the connection between a place, its people and their activities. It illustrates what is happening - where, how and why, and can be used to examine what has happened in the past, and to create likely future scenarios.

Geospatial information is collected and managed using a significant range of technologies and tools, such as Geographic Information Systems (GIS), photogrammetric software, satellites, mobile devices, sensors, etc. GIS in partcular, is typically used to manage geographic features in 2 and 3-dimensional space. It is also used to visualise the dynamics of the environment (4-dimensional space) as a series of data of the same area captured over time.

In many nations, there is a steady increase in the number of government departments, businesses and academic organisations that collect geospatial information, as well as end users that are continually adding value to data and redistributing as new products. Decision-makers are increasingly recognizing the opportunities afforded through the ability to harness this information for evidenced-based policy setting, and integrated and transparent government services.

However, geospatial information is often not well integrated across the broader government sector, nor in a format that can be easily used for analysis. This limits the full use and realization of geospatial information for decision-making.

Strategic Pathway 4 is a guide for data custodians (and producers) to manage their geospatial information using a whole-of-data lifecycle management approach (e.g., to organize, plan, acquire, harvest, curate, catalog, analyze, integrate, publish and archive this information) so that it can be easily integrated, discovered and used/reused for decision-making.

This objective is crucial. Every decision made, every event or activity in our daily lives, occurs at, or as a result of, a geographic location. Whether it is to determine the best site for a new hospital, choose the location of a new business

---

[1] *In this document the terms 'geospatial data' and 'geospatial information' are used interchangeably in general contexts. In specific contexts, 'Geospatial Data' refers to unprocessed facts and figures; 'Geospatial Information' refers to data that has been processed, organised, structured and presented in a meaningful way.*

venture, stage a community event or respond to an emergency – geospatial information is inherent in decision-making.

For this reason, geospatial information needs to be integrated across government processes and services, and managed according to guidelines within the broader data, goverance and technology frameworks of government.

Once integrated, geopatial information becomes a powerful tool for determining appropriate policy interventions and decisions for a wide variety of government services that contribute to economic growth, national security, sustainable social development, environmental sustainability and national prosperity.

4.2 Context and Rationale

Authoritative geospatial information is often managed in closed systems of government, resulting in a multitude of information silos. Because of this, geospatial information is often not discoverable nor interoperable, making data integration and exchange problematic. Over time, duplicated information can arise resulting in inconsistent data versions, gaps in data coverage, a waste of resources, and an inability to effectively and efficiently integrate geospatial information across a broad range of services. Sometimes information required is not collected at all. There needs to be more institutional collaboration, interoperability and integration across the various national data information systems and platforms that exist.

Geospatial information is often managed in silos, making data integration and exchange problematic. Cross-sector cooperation, standards and multidisciplinary collaboration can overcome these issues.

Cross-sector coordination, multidisciplinary collaboration, and standards are required to overcome problems associated with data integration and overly complicated data supply chains, particularly when organizing, planning, acquiring, curating, cataloging, analyzing/integrating, publishing and archiving geospatial information.

Multidisciplinary issues like climate change, disaster risk reduction, emergency response and urban planning require a coordinated effort across multiple government departments. Activities will often overlap when dealing with complex planning systems, such as implementing and operating an integrated flood risk management system or improving transportation infrastructure.

While departments use geospatial information in the course of their activities, this information is typically managed independently and not interoperable with other agency data. This means it is not possible to effectively bring together initiatives from the many departments into collective action plans.

Moreover, planning and risk management information will often be paper-based. In addition to being difficult to share, paper maps are time consuming to update and risk becoming so obsolete that they are detrimental to decision-making.

4.3 Approach

With so much geospatial data becoming increasingly available there is an essential requirement to deliver a well-coordinated program for the collection, management and distribution of geospatial information. Being able to leverage geospatial information effectively and efficiently is what will afford increased economic activity and growth over the longer term.

Considerations about how to coordinate the acquisition of geospatial data, and management of geospatial information, are an essential part of a country's geospatial strategy (SP1: Activity 1.7.4). These considerations are not only about what data to collect, but also about how to manage and maintain the data throughout its lifecycle.

In this strategic pathway, the approach for enabling data custodians to meet their data management, sharing and reuse obligations is through the execution of well-defined data supply chains and custodianship guidelines for organizing, planning, acquiring, integrating, curating, publishing and archiving geospatial information (Figure X).

The approach includes four key elements that are a guide for nations to achieve effective and efficient cross-sector and multidisciplinary geospatial data management, sharing and reuse. These elements include the implementation of priority fundamental data themes, guidelines for custodianship, acquisition and management, streamlined data supply chains and well-coordinated data curation and delivery mechanisms. These elements are explained in more detail below (See section 4.4).

The approach includes strategic pathway actions that are recommended as a means to achieve the four key elements. The actions are underpinned by guiding principles, and there are several interrelated actions detailed in other strategic pathways that may need to be achieved prior to, or in conjunction with, these strategic pathway actions. These interrelated actions are referenced in the text. Tools are available in the appendices. The Approach for Strategic Pathway 4: Data is illustrated in Figure 4.2 and explained in the following sections.

The actual implementation approach of each strategic pathway action will depend on country-specific needs, which may be influenced by country priorities, existing capabilities, resourcing potential, culture and other practicalities. Whatever the implementation approach, each action should reference the guiding principles below (section 4.5) as these describe what is important for effective and efficient geospatial information management.

Outcomes

- Increased range and scope of data available for decision-making and policy setting
- A critical mass of data for national development and innovation
- Cost reduction and productivity improvement
- Ability to monitor and measure progress towards achieving SDGs

Elements

- Data Themes
- Custodianship, Acquisition and Management
- Data Supply Chains
- Data Curation and Delivery

Tools

- Data Theme Description Template
- Data Inventory Questionnaire
- Dataset Profile Template
- Data Theme Road Map Example
- Gap Analysis Matrix
- Data Management Plan Elements
- Metadata Creation Checklist
- Global Statistical Geospatial Information Framework
- Custodianship Policy Example
- Data Release Guidelines

Principles

- Governance
- Consistent identification
- Quality Management
- Metadata
- Standards
- Accessibility
- Reusable Formats
- Authoritative
- Timeliness
- Provenance
- Integrity
- Demand Driven
- Supply Chain Efficiency
- Security
- Respected Rights

Actions

Getting Organized

- Data Framework
- Data Inventory
- Dataset Profiles

Planning for the Future

- Data Gap Analysis
- Data Theme Road Map

Capturing and Acquiring Data

- Data Capture
- Data Acquisition Program

Managing Data Sustainably

- Custodianship Policy and Guidelines
- Data Governance
- Maintained Metadata
- Storage and Retrieval Systems
- Data Release

Maintaining Accurate Positioning

- Maintained Geodetic Infrastructure

Integrating Data

- Geocoding and Aggregation
- Formalised Data Supply Chains

Interrelated Actions

- Geospatial Strategy (SP1)
- Country Action Plan (SP1)
- IGIF Governance Model (SP1)
- Geospatial Steering Committee (SP1)
- Geospatial Information (SP1) Coordination Unit
- Policy Framework (SP2)
- New Capture Methods (SP5)
- Global Geodetic Reference Frame (SP5)
- Crowdsourcing (SP5)
- Data Standards (SP6)
- Storage Solutions (SP5)
- Data Partnerships (SP7)
- Capacity Assessment (SP8)
- Identify Key Stakeholders (SP9)

Figure 4.2 *The Approach to data*

4.4 Elements

4.4.1 Data Themes

End users have a recurring need for fundamental data themes, as well as application-specific and socio-economic themes

National priority and fundamental data themes, and the datasets that fall within each of these themes, are relevant to a broad range of applications; hence, end-users have a recurring need for this information. As a guide, nations are encouraged to adopt the *Minimum List of Global Fundamental Data Themes*[2]. These 14 themes can be adapted to align with national strategic and statutory needs and mandates (See Page XX for a detailed explanation on UN-GGIM Global Fundamental Data Themes).

In addition, data themes can include application data themes that are captured for a specific purpose, such as health and utilities; and socio-economic themes that are used for demographic studies.

4.4.2 Custodianship, Acquisition and Management

Data custodianship mandates responsibilities for the acquisition, management, maintenance and quality of the information.

Data custodianship is usually assigned to a department. It mandates certain rights and responsibilities for the collection of geospatial information and the management of this information on behalf of the community. The rights and responsibilities may include the right to set conditions for data release and responsibilities for the acquisition, management, maintenance and quality of the information. This is important, as geospatial information is only useful if the end-user is assured that the data is correct and up-to-date or that is the best available at the time. Data custodianship usually includes the need to provide a recognised contact point for the distribution, transfer and sharing of the information. Mandating custodianship is one of the mechanisms used to avoid cross-government duplication in the acquisition and management of information. Data custodians are typically subject-matter experts and the data in their care is recognised as the authoritative source of truth.

4.4.3 Data Supply Chains

Data supply chains and their interlinkages need to be formalized and streamlined to improve the quality of information for end users

Data supply chains refers to the flow of geospatial information from one organization to another. In a supply chains, organizations are referred to as supply chain nodes, and information flows are referred to as the supply chain links. Each organization (node) will typically add value, such as updates, to the data before transferring the information on to the next organization (node). The types of interlinkages common to geospatial information flows are those between:

- One level of government and another, such as between national and district levels;

[2] *The Minimum List of Global Fundamental* Data Themes *endorsed by Member States at the Seventh Session of the United Nations Committee of Experts for Global Geospatial Information Management* can be accessed at *http*

- Organizations at the same level of government, such as between the geospatial organization and statistical organization;
- Departments and volunteer geographic information providers and various projects; and between
- Departments and private companies.

Many data supply chains, particularly long ones, or those that span multiple organizations and different levels of government, are not well-integrated and information becomes inconsistent and out-of-date very quickly. Consequently, information is not reliable. The resynchronization of datasets is a highly manual task, and therefore having formalized and streamlined, (i.e. standard data exchange) processes in place from the outset, and data from an identified authoritative source, will save considerable effort later, minimize the supply chain length, and improve the quality of information for users.

4.4.4 Data Curation and Delivery

Data curation and delivery refers to the art of maintaining the value of data and delivering it to end users in a way it can be visualized and used. The main purpose of data curation is to ensure that data is retrievable for future purposes or reuse. Data curation is usually assigned to a data curator (or aggregator) who is responsible for collecting data from many different sources and then aggregating and integrating the data into an information resource, such as map portal. Data curation has become increasingly important for data analytics, and the private sector has typically engaged in the curation of data to transform independently created government department data sources into unified datasets ready for analysis. Curated geospatial information is therefore often more valuable for decision-making than the individual geospatial datasets.

Data curators typically aggregate data sources into a unified dataset ready for analysis.

4.5 Principles

There are specific principles and elements for managing geospatial information to promote consistent data governance, management, discovery, sharing and reuse so that data organizations may meet their obligations to government and the user community. These principles also need to be embedded into acts, policies, and directives for update/compliance and integration into business practices. The guiding principles for data are:

By applying these principles, data custodians will be able to manage and share reusable geospatial data, and in doing so, meet their obligations to government and the user community.

- **Governance**: High quality authoritative datasets are delivered through designated data governance roles and responsibilities. Data governance roles are mandated for each dataset to ensure responsibility for the integrity and quality of data.
- **Consistent Identification**: A common data dictionary, vocabulary, ontology and persistent identifies are applied to the identification of data to enhance accessibility, manage effective use of data, and avoid duplicated collection or purchase.

- **Quality Management**: Quality management processes are used to manage the currency, completeness, accuracy and consistency of data for a specified purpose.
- **Metadata**: Appropriate metadata is applied according to standards and used to accurately define and describe geospatial data, including content, geographic extent, purpose, characteristics, currency and provenance etc., together with contact details for further information.
- **Standards**: Appropriate standards are adopted and enforced throughout the data lifecycle to enhance integration and interoperability of individual and disparate data sets.
- **Accessibility:** Easy, efficient and equitable access to spatial data through common geospatial platforms where technology, data formats, organizational arrangements, licensing, location, costs and conditions do not inhibit its use.
- **Reusable Formats**: Data is in a form suitable for further value-adding by internal and external users.
- **Authoritative:** Data is managed responsibly by the designated data custodian to eliminate the proliferation of duplicate data sets. The notion is to collect once and use many times.
- **Timeliness**: Data is managed according to priority, and where required, is maintained as close to real-time as possible.
- **Provenance**: The origin and quality of data is readily accessible to the user via metadata so that they can determine if it is 'Fit for Purpose"
- **Integrity**: The interrelationships between data themes collected by multiple agencies are managed with topological integrity.
- **Demand Driven:** Data acquisition and maintenance is aligned to user needs and requirements to achieve optimal resource allocation.
- **Efficiency:** Geospatial data products are differentiated as close as possible to the user to create more opportunities for reuse along the supply chain.
- **Security:** Data are held with adequate provision for long-term care including disaster recovery and backup procedures, are disposed or archived in accordance with government regulations, and considering technological advancements.
- **Respected Rights:** Confidentiality, privacy, intellectual property rights and the security of sensitive information are preserved, and the sharing of Indigenous knowledge is contingent upon consent of the knowledge holders in alignment with UNDRIP principles.

4.6 Actions

The following strategic pathway actions are typically used to address gaps in capability. They are a guide to best practice collection and management of integrated geospatial information.

The actions are recommended as a way to achieve the four key elements. Some actions have interrelated actions that need to be achieved prior to, or in conjunction with, the strategic pathway actions. These interrelated actions are referenced in the text and detailed under other strategic pathways.

Country-specific needs may be influenced by factors such as country priorities, existing capabilities, resources, culture and other practicalities. These will influence approaches for implementing each strategic pathway.

Whatever the implementation approach, each action should take into account the guiding principles above as these describe drivers for effective and efficient geospatial information management.

The strategic pathway Actions are divided into six categories that reflect the order in which the actions are typically completed. A road map illustrating this order and where the Actions typically occur is presented in Figure 4.3 and detailed more with interrelated actions in Figure 4.4. The categories of Actions are:

1. Getting Organized
2. Planning for the Future
3. Capturing and Acquiring Data
4. Managing Data Sustainably
5. Maintaining Accurate Positioning
6. Integrating Data

The following Actions are typically used to address gaps in capability. They serve as a guide to building the necessary capacity to strengthen integrated geospatial information management processes and systems.

Figure 4.3: *Data includes several actions and tools designed to assist countries to best practice collection and management of integrated geospatial information. The actions are divided into six categories and reflect the order with which these actions are typically completed.*

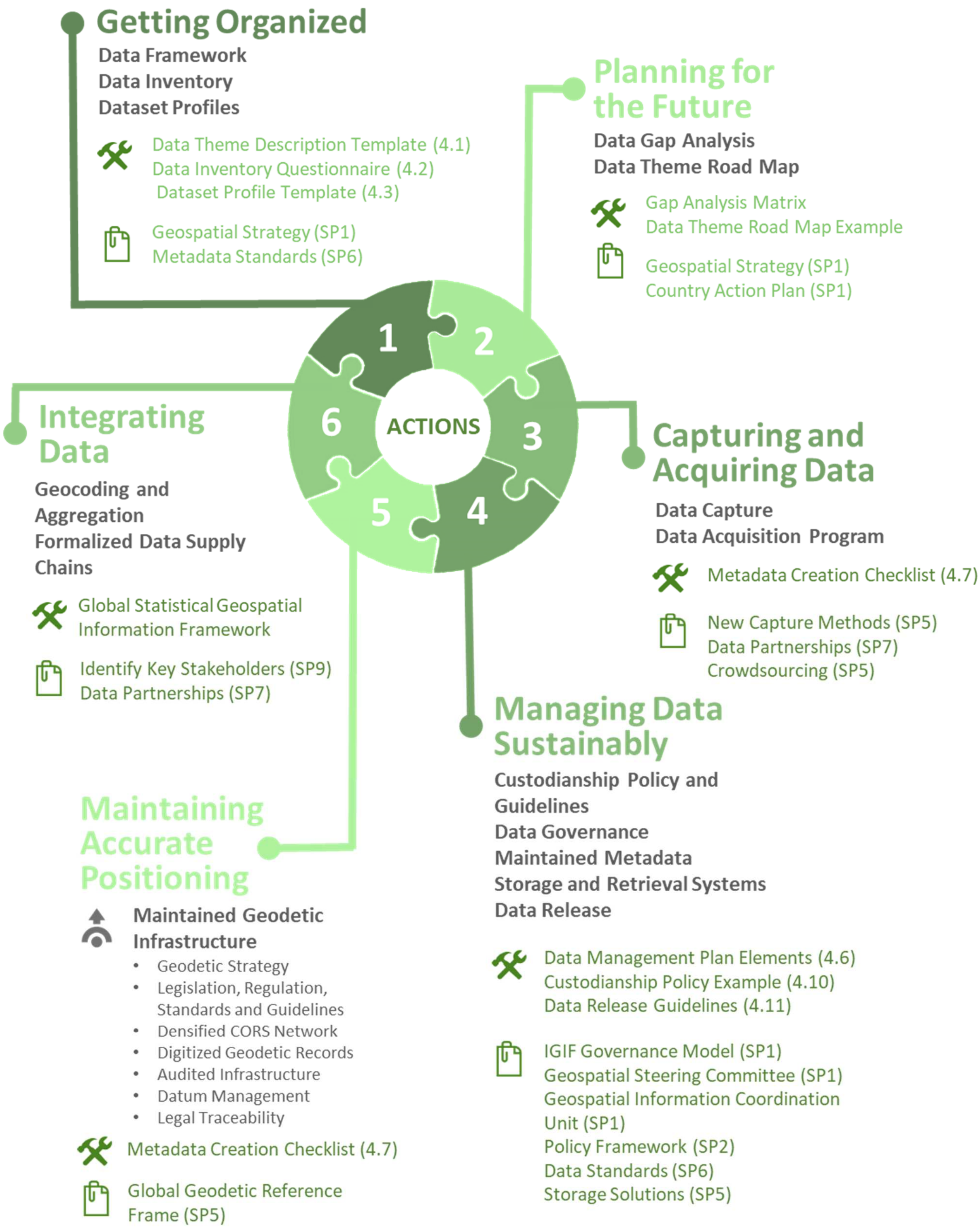


Figure 4.4: *Data includes several actions and tools designed to assist countries to achieve political endorsement and strengthened institutional mandates for building a cooperative data sharing environment. The interrelated actions provide key linkages to other strategic pathway actions.*

1 Getting Organized

4.6.1 Implement a Data Framework

The Data Framework provides a way to organize geospatial information so that it can be accessed easily and meaningfully.

The Data Framework is a methodology for organising a country's geospatial and statistical information so that is can be accessed and used. This is an important, as it provides a way to organize geospatial information so that it can be accessed easily and meaningfully. Being able to find information and understand its purpose is critical to good decision-making. There are typically three classification tiers to a Data Framework:

1. **Fundamental** data themes, such as transportation, that are required for a broad range of decision-making applications, and for which users have a recurring need (See inset page XX);
2. **Application** data themes, such as flood models, required for specific studies; and
3. **Socio-economic** data themes that provide demographic information, such as census data.

The data themes are essentially categories for grouping geospatial datasets in each classification category. The themes make it easy for people to locate information and are used to structure machine-readable data catalogs so that information can be located easily by search engines.

The Data Framework records Data Theme Descriptions in a way that makes it easy for people to understand what data is available within the theme, and the purpose for which it can be used. A Data Theme Description typically includes (a) access, pricing and licensing categories; (b) the data custodian; (c) data characteristics, such as data structure, accuracy, and coverage; (d) data standards and regulations; and (e) the purpose for which the dataset is best used.

An example of a Data Framework is included in Appendix 4.1

Additional guidance for defining data layers and models can be found in the SDI Manual for the Americas that was prepared for ECOSOC by the Permanent Committee for Geospatial Data Infrastructure of the Americas (p. 45).

An example of a Data Theme Description for the Administrative Boundary Theme is provided in Appendix 4.1.

Importantly, data themes, particularly fundamental data themes, should be considered when developing the nations *Geospatial Strategy* [SP1: Activity 1.6.4] as they provide the foundation (base) information for achieving national goals and international commitments, such as the SDGs.

Significant digitization of geospatial datasets may be required to collect data for each fundamental data theme. Nevertheless, new methods and emerging technologies can be adopted to fast track digitisation (SP5: Activity x.x.x).

4.6.2 Conduct a Data Inventory

A Data Inventory enables government to fully understand the extent of their data holdings.

An inventory of all geospatial data and information held by institutions is required to fully understand the extent of a Nation's data holdings. The data inventory should record, as a minimum, the spatial data format, currency, accuracy, ownership, and the purpose for which the data is primarily used. This information will be incorporated into the Data Framework document (SP4: Activity 4.6.1).

It is important to differentiate (a) what datasets are used by organisations as a reference only, as this highlights the benefits of data sharing; and (b) what data sets are collected and managed by organisations as this will identify if there any areas of duplication in data collection.

Not all government spatial information will currently be in digital format. Some of the country's wealth of knowledge is stored as paper-based documents/maps. This information will make a valuable contribution to decision-making capabilities when it is converted into digital form. It should therefore, be considered in the data inventory process.

An example of a comprehensive data inventory questionnaire is provided in Appendix 4.2.

Global Fundamental Data Themes

Data Theme	Description
Global Geodetic Reference Frame	The **Global Geodetic Reference Frame** is the framework which allows users to p recisely determine and express locations on the Earth, as well as to quantify changes of the Earth in space and time. It is a crucial prerequisite for the accurate collection, integration and use of all other geospatial data.
Geographical Names	**Geographical Names** provide orientation and identity to places. They are location identifiers for cultural and physical features of the real world, such as regions, settlements, or any feature of public or historical interest.
Addresses	**Addresses** are a structured label, usually containing a property number, a street name and a locality name. Addresses are often used as a proxy for other data themes such as Land Parcels and usually are linkable to geographic coordinates.
Functional Areas	**Functional Areas** are the geographical extent of administrative, legislative, regulatory, electoral, statistical, governance, service delivery and activity management areas.
Buildings and Settlements	A **Building** refers to any roofed structure permanently constructed or erected on its site, for the protection of humans, animals, things, or the production of economic goods. **Settlements** are collections of buildings and associated features where a community carries out socio-economic activities
Land Parcels	**Land Parcels** are areas of land, cadastral parcels, or areas of the Earth's surface (land and/or water) under common rights (such as ownership or easements), claims (such as minerals or indigenous land) or use.
Transport Networks	The **Transport Networks** theme are the suite of road, rail, air and water transport routes and their connectivity.
Elevation and Depth	The **Elevation and Depth** theme describes the surface of the earth both on land and under a body of water, relative to a vertical datum.
Population Distribution	The **Population Distribution** theme covers data for the spatial distribution of population and its characteristics, as well as how population impacts urbanisation, regional development or sustainability.
Land Cover and Land Use	**Land Cover** represents the physical and biological cover of the earth's surface. **Land Use** is the current and future planned management, and modification of the natural environment for different human purposes or economic activities.
Geology and Soils	**Geology** is the composition and properties of geologic materials (rocks and sediments) underground and outcropping on the Earth's surface. **Soil** is the upper part of the earth's crust, formed by mineral particles, organic matter, water, air and living organisms.
Physical Infrastructure	The **Physical Infrastructure** theme includes industrial & utility facilities, and the service delivery facilities associated with administrative & social governmental services such as public administrations, utilities, schools and hospitals etc.
Water	The **Water** theme covers the geographic extent and conditions of all water features including rivers, lakes and marine features.
Orthoimagery	**Orthoimagery** is geo-referenced rectified image data of the Earth's surface, from satellite or airborne sensors.

4.6.3 Create Dataset Profiles

Datasets Profiles are created from the information collected during the data inventory. Like Data Theme Descriptions, dataset profiles are often incorporated into a data catalogue so that end-users can determine if the dataset will suit their purpose. This improves the accessibility and use of geospatial information. Dataset Profiles include specific information about the content of the dataset and can include:

Dataset Profiles improve the accessibility and use of geospatial information.

- **Description:** General overview of the dataset theme and content.
- **Dataset Uses:** Lists the general applications the data is suitable for.
- **Current Status:** Data created and last modified.
- **Standards and specifications:** Data and metadata standards applied.
- **Access and Licensing:** Conditions under which the dataset can be used.
- **Data Theme:** The Data Framework theme to which the dataset belongs.
- **Update Strategy:** Provides an indication of currency and when updates will be made available.
- **Dataset Custodian:** Organisation responsible for acquiring and managing the dataset.
- **Data Curator (Aggregator):** Organisation responsible for data aggregation, integration.
- **Distributor:** Responsible for delivering datasets to end-users.

Note: It is plausible that the data custodian, curator and distributor is the same organisation.

An example of a Dataset Profile is provided in Appendix 4.3. Note: Dataset Profiles are in human-readable formats; whereas metadata is in machine-readable formats

See Interrelated Action on Metadata Standards ISO 19115 (SP6)

4.6.4 Undertake a Data Gap Analysis

A Data Gap Analysis is used to identify the strategies required to address the gaps in geospatial information capability.

The data inventory describes what datasets a nation has, but not the actual data needs of the nation, nor how to achieve the data outcomes required. A Gap Analysis Matrix is required to organize information so that it simpler to identify the strategies required to address the gaps in geospatial information capability.

Data gaps may include, incomplete data coverage in priority areas, inadequate accuracy, inconsistent geodetic referencing system/projection, limited integration between land parcels and addresses, restricted access to utilities data, difficulties maintaining adequate levels of currency for building footprints, missing data themes etc.

A Gap Analysis considers:

- **Current Situation** – This information is generally derived from conducting a SWOT and PEST Analysis in a workshop situation with a broad stakeholder group. In this way, the current situation is viewed from multiple perspectives.
- **Desired Future State (Strategic Goals) -** these are the goals a nation is endeavoring to achieve. Goals are generally documented in the Geospatial Strategy (See SP1: Activity 1.6.4).
- **Gaps in Capability (Challenge)** – These are the challenges that need to be overcome.
- **List of Actionable Strategies** – A list of strategies, such as the collection of new data or the implementation of a new geodetic datum, which are required to move nations from their current situation to a desired future state where geospatial data resources are concerned. These strategies provide input to the Data Theme Road Maps (SP4: Activity 4.6.5) and are often included as activities in the Country Action Plan for strengthening geospatial information management (See SP1: Activity 1.6.7).

An example of a Gap Analysis Matrix is provided in Appendix 4.4.

See Interrelated Actions on developing a Geospatial Strategy (SP1); Conducting a SWOT and PEST Analysis (SP9); Data Theme Road Map (SP4); Country Action Plan (SP1)

4.6.5 Prepare a Data Theme Road Map

A Data Theme Road Map is a powerful strategic tool for coordinating cross-government activities that will lead to strengthening integrated geospatial information management.

A Data Theme Road Map is a powerful strategic tool for coordinating cross-government activities that will lead to strengthening integrated geospatial information management.

The Data Theme Road Map documents the major steps and milestones required to reach the desired outcomes and close the capacity gaps determined during the Gap Analysis (SP4: Activity 4.6.4). This high-level document serves as a communication tool to help articulate the strategic thinking behind the goals and the plan for getting there. As a minimum, the road map typically includes (Figure 4.5):

- A **Vision** for what the datasets within the Data Theme will be in the future. The vision is usually driven by particular use cases i.e. what the data is required for as a matter of priority.
- Short term **Goals** to be achieved and associated milestone dates.
- **Activities** to be undertaken, such as improving data quality, data supply chains, standards compliance, delivery mechanisms, policies associated with data release, retention and sharing, archive management for reuse and stakeholder engagement.
- Assignment of **responsibility** for change management:
  - A Data Theme Sponsor with responsibility to engage with dataset custodians to guide initiatives to move each Data Theme towards the stated vision.
  - An activity owner (usually the custodian) with responsibility to implement initiatives to improve the dataset.
- The **outcomes** associated with each activity.
- **Funding** situation for each activity.

The Road Map is a key tool for achieving stakeholder buy-in. It must therefore, achieve some immediately measurable benefits whilst progressing towards the eventual desired future state.

Acquiring, enhancing and updating geospatial information is a costly and resource intensive process. The Road Map is typically designed to be completed over a three to five-year period and reviewed annually

.

Figure 4.5 *Elements of Data Theme Road Map*

Priorities will differ by country and should be based on the political, social and economic drivers. Typically, the Fundamental Data Themes are the most important because they have an impact across a broad range of applications. They are also more likely to have an impact on economic growth. With benefits from initial efforts realized, it may be possible to add more data themes over time.

The Imagery Theme is likely to be a priority for many nations, as it is used to create other Fundamental Data Themes, such as the digitisation of buildings and settlements. While imagery is costly, there is a growing range of free Earth Observation data that can be used by countries to start the transformation to digital geospatial information use and management (See SP:7 Activity 7.x.x).

The data custodian is integral to developing the Road Map for the data theme/dataset under their responsibility and therefore, custodians should be determined at the outset.

An example of a Data Theme Road Map is provided in Appendix 4.5.

Interrelated Activity: Establishing partnerships (SP7)

4.6.6 Capture Geospatial Data

Geospatial information is collected in various formats, typically - raster (images) and vector (lines, point and polygons). There are various methods for capturing new datasets, as well as enhancing and maintaining existing datasets. The method chosen will be dependent on the nature of the geographic feature, level of details required, ease of updating, available resources and skills capacity, and budget.

Determining the best data capture methods to implement is an evolving practice and each nation will have a unique starting point - what is a suitable method for some, will be impractical for others.

Determining the best data capture methods to implement is an evolving practice and each nation will have a unique starting point - what is a suitable method for some, will be impractical for others. Data capture methods are considered under Strategic Pathway 5: Innovation (See SP5: Activity 5.x.x). Methods are discussed in terms of the potential to fast-track data collection and updating, as well as opportunities to leapfrog traditional manual methods and move straight to automated processes and new levels of location intelligence.

As a guide, the following need to be considered when embarking on a data capture program.

- *Does the data already exist elsewhere in a usable format?* This may seem an obvious question, however data can be difficult to find within closed systems of government and therefore, a thorough investigation is required.
- *What information is necessary to meet internal organisation business needs and; what do end-users require and how will they use the information?* This is crucial. For example, while an organisation may require point locations of buildings, an end-user may need the actual building footprints. In this situation it would be logical to capture the building as polygons and derive the point locations from the building footprints.
- *Are there options to partner with another organisation?* See potential partnership options in Strategic Pathway 8 that could lead to potential savings and resource sharing.
- *Who will be the Data Custodian and how will the data be managed in the long term?* This is an important consideration, particularly for projects, which often result in isolated geospatial investments that have one-off usage because consideration was not given to the longevity and potential future value of the data.

See Interrelated Action: Modern Geospatial Data Capture Methods (SP5)

4.6.7 Implement a National Data Acquisition Program

A nationally coordinated Data Acquisition Program supports the most strategic investment in, and use of, geospatial data resources

Each year government agencies purchase or collect a wide range of geospatial information such as aerial photography, satellite imagery and topographic surveys. The procurement of this information is often not centralised and therefore, there is significant potential for the same information to be purchased, shared or digitized.

Geospatial information can be expensive, and given the potential for overlap, it is imperative that governments adopt a more coordinated whole-of-government approach to ensure a more strategic investment and use of geospatial resources.

One mechanism to achieve this is to establish a National Data Acquisition Program where organisations (including project teams) can register their requirements annually. In this way the planning, recording and acquisition of geospatial information can be funnelled through one governance channel enabling clear oversight of data acquisitions nationally.

The main objective of the National Data Acquisition Program is to reduce costs associated with the capture, storage and management of information by:

- Procuring data once and in a way that the data can be used many times.
- Planning, recording and acquiring geospatial information through one channel.
- Clear oversight of nationally/internationally funded Projects that acquire geospatial information.
- Acquiring multi-user licenses for imagery to enable reuse by many organisations.
- Maintaining a single version of key government geospatial information.
- Leveraging greater value through economies of scale and consolidated spend.
- Establishing an annual program of work where organisations can register their requirements annually.
- Fosters collaboration where organisations have similar requirements, such as for multi-user licensing for imagery, in order to leverage greater value through economies of scale and consolidated spend.

A National Data Acquisition Program is managed with consideration for cross-government needs and in the National interest. The program may be administered by the Geospatial Information Coordination Unit (SP1: Activity 1.6.2) and will include representation and input from across key government stakeholders to decide on acquisition priorities and expenditure.

The Geospatial Information Coordination Unit will typically take responsibility for the business plan, management of finances and achievement of outcomes. Essentially, Geospatial Information Coordination Unit acts as a Community of Practice for Geospatial Data Providers and conducts the following tasks:

- Calls for geospatial data requests annually from organisations across the government sector
- Maintains a register of all government geospatial data requests so as to avoid duplication and facilitate sharing and reuse.

The actual acquisition of geospatial data and imagery is typically done by the Survey Department (or equivalent) that is responsible for preparing the acquisition plan based on stakeholder submissions (needs and priorities), designing procurement specifications, and validating the quality of the information as it is received.

A new procurement/funding model may be required to administer the National Data Acquisition Program. Options are:

1. Agencies fund and manage their own spatial data capture program but register their data capture program with the National Data Acquisition Program to avoid potential duplication;
2. National Data Acquisition Program receives an allocation of funds from central government to procure data on behalf of all organisations;
3. Organisations contribute funds to the annual program in proportion to their data needs;
4. Commercial revenues from value-added data are redirected back into the National Data Acquisition Program.

There may also be opportunities to develop data acquisition partnerships (See SP7: Activity x.x.x) and negotiate more effective and consistent data licensing and copyright conditions for imagery, which may be more cost effective and result in reducing barriers to sharing and re-use of data within government.

4.6.8 Implement Data Custodianship Policy and Guidelines

A Custodianship Policy and Guidelines provides advice on the principles of custodianship, the responsibilities of custodians for the management of information products and can include responsibilities for end-users when acquiring and using information. Guidelines are generally based on experience and simply reflect good practice, convention and convenience.

A Custodianship Policy provides a means of accountability for the production and delivery of reliable geospatial information.

There are significant advantages to be gained through having a custodianship policy. Custodianship policy provides a means of accountability for the production and delivery of reliable information. For government departments, this means being recognised as the authoritative source of information. For end-users this means increased confidence in the accuracy, completeness and currency of government information.

Custodianship policy also eliminates unnecessary duplication when capturing and maintaining geospatial information, as the policy typically mandates responsibilities to a single agency, or to custodians working in partnership. This means funds previously spent on duplicated activities can be reallocated to other priorities.

An example of a Custodianship Policy and Guidelines is provided in Appendix 4.9.

4.6.9 Formalize Data Governance

Data Governance results in improved productivity and efficiency for an organisation.

Organisations responsible for geospatial information management have limited budgets and resources, and often struggle to collect and update the information in their care. Implementing Data Governance can result in improved productivity and efficiency for an organization. Data Governance[3] is a logical structure for classifying, organizing and communicating activities associated with making decisions and taking action on the management of geospatial information.

The model for data governance is different to the IGIF Governance Model proposed in SP1: Activity 1.6.x. The IGIF Governance Model is designed to institute cooperation, and visibility and alignment of decisions about data that have an impact across the broader government sector; whereas the Data Governance Model is implemented at the individual organization level to govern how individual datasets are managed.

Most organisations will have some form of data governance – informal or formal. Organisations that formalize data governance do so when data volumes are increasing and from far more sources, creating an environment for data inconsistencies that need to be identified and corrected.

[3] *For more information refer to the Data Governance Framework from the Data Governance Institute (2013 at. http://www.datagovernance.com/the-dgi-framework/*

Figure 4.6: Data Governance Framework

Data governance considers the people, processes and technologies that are needed to manage, protect, enhance and deliver reliable information. Data Governance typically includes a governance model and governing body, a set of procedures and a data management plan for executing those procedures (Figure 4.6).

4.6.10 Implement Data Management Plan

Enterprise data management plans promote consistency in geospatial data management practices within departments and across government. They are a strategic tool used to provide oversight for data assets as well as other ICT assets so that maximum value is derived through appropriate governance, maintenance and protections.

Data management plans are a useful tool for overcoming data silos. The guidelines in the plan facilitate data discoverability, accessibility and interoperability through uniform standards, processes and methods. They streamline processes through the implementation of standards and best practices, resulting in geospatial information that is suitable for frontline service delivery, analytics and decision-making.

The major elements considered in the Data Management Plan are data governance, accessibility, quality, security and knowledge management (Figure 4.7).

The elements of the Data Management Plan are detailed in Appendix 4.6.

Figure 4.7: Elements of a Data Management Plan

4.6.11 Maintain Metadata

Metadata makes it possible to access geospatial data and determine its usefulness for mapping and analysis.

Metadata describes the origin of geographic data and tracks the changes to a dataset over time. Metadata refers to the 'who', 'what', 'why', 'where', 'when' and 'how' about individual geospatial datasets and how they have been collected (Nebert, 2004). Without metadata it is difficult to access geospatial data and, when located, impossible to determine its usefulness for mapping and analysis.

The concept of metadata can be likened to a library catalogue which includes metadata records that help a person discover, use and manage a collection of books, documents and other information. Metadata for digital geospatial data is no different; except that the emphasis is on the geospatial component and includes producer, extent, accuracy, datum and other characteristics.

Metadata is managed at different levels including 'discovery', 'exploration' and 'exploitation':

- Discovery metadata is used to describe what data sets are available and provides a mechanism by which an agency can publish their data holdings.
- Exploration metadata allows users to explore data so that they can determine if the currency and accuracy of data are sufficient for their data analysis.

- Exploitation metadata includes those properties required to access, transfer, load, interpret and apply data in end applications where it can be exploited by the user.

Metadata management is done by the agency that collects and maintains the data, whereas metadata catalogs are managed by a central team.

The minimum amount of metadata recommended is the information contained in the Dataset Profiles (see 4.6.3 above). The form of metadata creation and its maintenance will depend on the size of an organizations data holdings and the patterns of data management within the organization. Traditionally, organisations with modest data holdings store metadata in conventional discrete word processing documents. This methodology meets minimum requirements. However, for large data holdings consideration should be given to investigating more advanced aspects of GIS to extract aspects of the metadata from the data itself.

Metadata should be created with a view to supporting the implementation format within a database or software system; an export (encoding) format for transfer between computers and; presentation formats for viewing. In this way, the system will support business and operational requirements, standard encoding for data exchange and permit a number of 'report' views of the metadata that will satisfy the needs and experience of the different users.

Consideration should be given to using International metadata standards that use a structured exchange format and for flexible presentation and software support for encoding.

Metadata validation is an important business process. However, there are few tools available to automatically verify the accuracy of the metadata other than those tools for checking its syntactical structure. Human review is required and should be incorporated into an organisations business process.

The management of the metadata is best left with the agency that collects and maintains the data. However, metadata catalogs are more appropriately managed by a central coordinating team, which can ensure consistent metadata management across organisations and thus ensure a consistent means of accessing data via a national data catalogue.

There is a risk that data producers will consider the task of creating metadata for their data sets too hard or they may have insufficient time to complete the task. This is particularly the case for organisations that have hundreds of legacy data sets. Managers of organisations may need to commit additional human resources to the task and be conscious of the fact that a day or two spent documenting each data set is small expense when compared to the months taken to produce the dataset in the first place.

Metadata based on International standards will have far reaching 'long term' benefits for both data producers and the geospatial information user community See Strategic Pathway 6 for more information regarding metadata standards.

4.6.12 Empower Data Release

Data Release Guidelines provide custodians with guidance on what can be lawfully released in the national interest.

Data custodians are accountable for the release of information, and this responsibility is often mandated under a Custodianship Policy. However, custodians often need additional guidance on what can be lawfully released in the national interest. This is usually achieved by assigning an access category to each dataset. Access categories are typically:

- **Open:** Open Access - Open access to spatial data takes precedence over restricted access unless there are specific, compelling reasons to restrict access.
- **Restricted:** Government Use Only – Approval of an authorized officer is required. Decisions on restricting access are based on privacy, commercial sensitivity, national security, environmental sensitivity or legislative requirements.
- **Confidential: Internal Use Only:** Access to authorised users only, where access is necessary to the function or activity of the collector/user.
- **Not to be Released:** A default position for all types of data and information until assigned a category.

When assigning access levels, there needs to be a balance between making information openly available for public good and economic growth; with the need to restrict access for the protection of individual rights and national security.

There are times when certain classes of geospatial data need to be withheld from public access and use. For example, the location of gem deposits, defense establishments, culturally sensitive sites, detailed bathymetry of harbor approaches, and locations of endangered species. There are also times when withholding data can degrade decision-making processes, such as emergency planning and response, and environmental management.

While it is recognised that some data cannot be made public because of its sensitivity, data should still form part of the Data Framework and be managed by nominated custodians. Authorized users can be given access for business purposes and mechanisms can be introduced to ensure that privacy, national security and other sensitivities are not compromised.

Importantly, the issue is not about whether geospatial data should be collected and made accessible, but rather what restrictions will be applied to its usage and how this should be decided. This can be achieved by specifying a License Type.

Examples of Licence types are provided in Strategic Pathway 2 - Appendix XX.

4.6.17 Implement Secure Data Storage and Retrieval Systems

Information management has become more difficult in recent times with larger volumes of data and limited ICT budgets for storage and retrieval.

In developing nations, geospatial information is often stored as hardcopy records of maps and plans. These maps are often the only records of land boundaries, urban development schemes and construction plans, etc. This paper-based information takes up a lot of physical storage space, tends to be of low quality, is prone to damage, and vulnerable to natural disasters.

Digitizing documents has many benefits - reduced physical storage requirements, lower document management costs and timeliness of processes, increased document security and reduced risk of damage, better knowledge management and search capabilities, opportunity to build in regulatory compliance, and opportunities for teamwork through digital collaboration technologies.

However, moving to a digital environment brings its own risks and the need for secure storage and retrieval processes. The digital storage capacity within government organisations of many developing nations is under capacity, increasing the risk of data becoming lost. Some organisations store data on computer hard drives and this information is only accessible to the computer owner/user. More concerning is that this information sometimes represents the only copy and as such, the risk of information loss is extremely high. In addition, digital imagery is often stored on transportable hard drives, while historical imagery may remain archived on film. Both of these situations present risks due to disk corruption and the continual deterioration of film.

Information management has become more difficult in recent times. Governments typically need to retain larger volumes of data for longer periods of time to meet governance and compliance requirements, and they are faced with limited information technology budgets. Cloud storage provides much needed storage capacity for government. However, few organisations are connected, mainly due to a lack of awareness of the storage facility, and concerns regarding the security and sovereignty of data that is not kept on home soil. Cost may also be a limiting factor. For a discussion on the difference types of storage solutions and innovative methods, see SP6: Activity 6.x.x.

Interrelated Activity – Data Management Plan (SP4) and Storage Solution (SP5)

When implementing storage and retrieval systems, the following guiding principles apply:

- **Security:** Data are maintained in a secure environment and transmitted through secure methods.
- **Maintain:** Data are held with adequate provision for long-term care including disaster recovery and backup procedures.

- **Disposal:** Data that are not required are disposed of, or archived in accordance with Government record keeping legislation.
- **Retrieval:** Data can be easily retrieved in a form that is needed by authorized persons
- **Compatible:** Information is stored in a way that is compatible with the hardware and software systems that will be used to access it.

Storage and retrieval systems are included in the Data Governance Framework and processes and procedures documented in an organization's Data Management Plan (SP4: Activity 4.6.7)

5 Maintain Accurate Positioning

4.6.13 Maintain the Geodetic Infrastructure

The geodetic infrastructure (also referred to as the positioning infrastructure) allows users to precisely determine and express locations/coordinates on the Earth, and to measure changes to the Earth's shape over time.

The geodetic infrastructure is a prerequisite for the accurate collection, integration and utilization of all other geospatial data. Without the existence of this predefined geodetic framework almost every endeavor that depends upon positioning information would become daunting, from both a practical and financial view point.

The geodetic infrastructure is a prerequisite for the accurate collection, integration and utilization of all other geospatial data.

The geodetic infrastructure underpins all geospatial data and provides the means for relating all cultural, administrative and geographic features to the one horizontal and vertical coordinate system (geodetic datum) that is used by a nation. Importantly this includes providing the positioning framework for the cadastral system.

At the global scale, the geodetic infrastructure includes the network of high fidelity geodetic observing stations, many with multiple collocated observing techniques – Very Long Baseline Interferometry (VLBI), Satellite Laser Ranging (SLR), Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS), GNSS Continuously Operating Reference Stations (CORS) and Gravimetry (Quantum, Superconducting, and/or Absolute). These stations, their data archive and the data processing/analysis centers form the basis of the Global Geodetic Reference Framework (GGRF) and are owned and operated in a collaborative effort by many contributing organisations across the globe.

At the National level, the geodetic infrastructure includes:

- Legislation, Regulations Standards and Guidelines that provide the authority and requirements for geodetic infrastructure practices and processes.
- A geodetic datum that allows a point on the Earth's surface to be uniquely described or identified by its latitude and longitude (and height). A geometric datum is based on a mathematical model of an ellipse and has reference points that are used to define the datum. A vertical datum will be based on a reference sea-level or on a defined gravimetric potential value.
- A practical realization of which enables access to the fundamental spatial reference system for all subsequent surveying requirements for cadastral, mining, mapping, engineering and related work.
- Geodetic data that includes reference point horizontal coordinates that support the positioning and mapping within a datum, height information, and the means of transforming different coordinate systems and relating different height systems, along with supporting metadata.
- Geodetic services are provided by many nations which transform client data into precise position information in global or national reference frames with error estimates and other ancillary information.
- Physical geodetic infrastructure, which includes:
    - Passive survey marks, of varying degrees of quality, ranging from low precision ground monuments to high precision GNSS pillars capable of measuring minute geodynamic motions.
    - Active CORS - A permanently configured, network of GNSS observing stations that support a range of positioning infrastructure purposes and applications.
    - Information storage, maintenance, delivery and access systems including the transfer of data files, data downloads from Web portals, integrated web services and mobile applications.
    - Suitably qualified and experienced personnel to operate and maintain the geodetic infrastructure.

Geodetic infrastructures differ between countries, often due to the nature of the geographic features (e.g. flat verses mountainous terrains) and economies, and because there are unique problems that need to be overcome to establish adequate geodetic infrastructure.

In developing nations, the high cost of establishing and maintaining geodetic infrastructure has meant that some countries have relied on temporary survey marks for project work. This has led to a lack of standardization and uniformity, and as a consequence, the geodetic infrastructure has become fragmented, unconnected to the national or regional framework, and inadequate for accurate positioning.

In this situation, the question is 'What should countries do to arrive at an adequate geodetic infrastructure?' The following are provided as a guide to improving existing geodetic infrastructures.

- **Develop a Geodetic Strategy:** To inform government and industry stakeholders as well as raising the profile to decision makers of the importance and value of a modern geodetic system, now more than ever, to support the new paradigm of positioning technology and the economic benefits it will bring.
- **Review and Update Legislation, Regulations, Standards and Guidelines**: To be reviewed in collaboration with industry and government stakeholders to support a modern geodetic and cadastral framework.
- **Establish and/or Densify and Maintain a CORS Network:** To support a higher level of positioning accuracy for both post processing and real time positioning. Such networks provide an improved understanding of earth dynamics, crustal deformation and tectonic motion and provides the vital link to the GGRF. Earth dynamics enable precise survey information to be shifted in time to account for movement of the earth with respect to the frame, a core requirement for legal surveys in some regions. Suitable CORS may potentially be included in either regional reference frames e.g. Asia Pacific Reference Frame (APREF), African Geodetic Reference Frame (AFREF) and/or the GGRF. It is important to consider Public/Private Partnership opportunities, such as commercial real time CORS networks where the user-pays for subscription-based services (See SP7 Act 7.6.x).
- **Digitize All Hardcopy Geodetic Records:** Protects historical records from loss and enables accessibility to a wider user base. Publish geodetic mark information online to ensure all users have access to this information.
- **Undertake an Audit of Current Geodetic Infrastructure:** Includes physical infrastructure, associated data and capacity (human resources) to fully understand current and future requirements. Analyse data by considering if it is correct, complete and current and has appropriate metadata. E.g. a coordinate or height without an accuracy statement and associated datum is of little value. Consider using crowd sourcing to gain information on the location and condition of geodetic marks – a form of geocaching, successfully used in the Australia, UK and USA. Understanding capacity in terms of

suitably qualified and experienced personnel is critical to supporting a modern geodetic framework and future training needs. Crowd-sourcing may also be used to facilitate rapid positioning through sharing of local atmospheric information with others.

- **Improve Datum Infrastructure Management** – This requires a program of work to relate all marks at the local level to the highest accuracy network of CORS sites. The program involves a structured process of GNSS field campaigns, rigorous processing of observations and rigorous network adjustment. The hierarchy system of relating local level marks directly to CORS sites is represented in figure that illustrates an Australian Case Study.
- **Achieve Legal Traceability for commercial GNSS service provision**: The advent of commercial GNSS real-time services in some nations has raised concerns about traceability of the resulting position information to the national datum. In Canada, a voluntary compliance program has been introduced to validate commercial network coordinates and to monitor station stability. Results of this monitoring are shown on a publically accessible government website. Some legislative actions have been implemented to require use of compliant services for land surveys.
- **Achieve Legal Traceability**: This requires the support of a legislative framework for legal metrology and a delegated authority to administer the Act, Regulations, Standards and Guidelines. It can be applied to measurements of length by electronic distance measurement instruments (EDMI)/total stations and for position by GNSS CORS. In some nations there is a well-established methodology for attaining legal traceability for length by establishing physical calibration baselines that are measured with EDMI that are certified against an established and legally traceable standard for length. Calibration of an EDMI is undertaken to prescribed standards and processes in order to achieve certification for legal traceability. The current trend to use GNSS for all measurements is making the establishment of EDMI calibration baselines obsolete, however it also complicates the legal traceability process as there are significant components of the GNSS measurement process that cannot be easily standardized. Legal traceability of position for GNSS CORS can be achieved by complying with a rigorous methodology to establish and maintain a reference standard for position with oversight and regular auditing by the delegated authority. This reference standard of position of GNSS CORS can then be used as the reference stations to define a datum.

- **Upgrade the Datum through Modernisation:** (Also see SP5 Act 5.6.x) - With the advent of GNSS there is now a greater impetus to modernize geodetic datums and in this process there is an increasing trend to use a global reference frame and a geocentric ellipse that aligns with GNSS. This negates the need for complex transformations between localized datums and is in anticipation of centimeter accuracy real time GNSS positioning being available to the mass market on mobile devices. Given that existing geospatial data holdings often are in older datums, a thorough understanding, and an accurate transformation methodology are required to relate localized datums to the GGRF. Ultimately, to maximize the benefits of accurate positioning afforded by GNSS, a modern geodetic datum is required, but it is understood that transformation of existing data holdings is costly.
- **Height Modernisation:** Datum Modernisation also includes height Modernisation and whilst the GGRF can deliver three-dimensional coordinates, for practical management of water the height component needs to incorporate effects related to the Earth's gravity field rather than a mathematical ellipse. To achieve this, a thorough understanding of the current height datums and their origin is required. If there are multiple height datums (e.g. terrestrial and or hydrographic/chart datums) then a means of correlating them is highly desirable. Relating the height datum to chart datums and to a reference ellipse is facilitated by use of a geoid model. Geoid accuracy benefits from this relationship with sea-level and other information. Geoid modelling typically incorporates airborne and terrestrial gravity data, GNSS observations on ground marks with local height values, CORS stations collocated with tide gauges and other information. Ultimately, the GGRF will include the development of an International Height system at a defined geopotential value which will support transformations between local height datums and the IHRF.

6 Integrating Data

4.6.14 Geocoding and Aggregation

Integrated geospatial and statistical data enables good policy decision-making.

Geography and statistics are implicitly interconnected in a conceptual sense. Demographics (education, health, labour), economic (trade, tourism, sales, salaries, etc.) and environmental information (water quality, weather, traffic, etc.) are all associated with a geographic location, municipality, country and region.

However, geography and statistics are typically disconnected at logical and physical database implementation levels. This is partly because geospatial technologies are used to a limited extent by Statistics Agencies and generally only for census-related work, but also because geospatial data and statistics are managed in organisational silos.

With the 2030 Agenda for Sustainable Development, there is growing appreciation for the value of integrating geospatial and statistical information for making good policy decisions that will lead to the achievement of the SDGs, and for monitoring and measuring progress towards success.

There is also pressure on governments to release location-enabled socioeconomic data for broader business opportunities innovation, and research and development. This is because geospatial data adds significantly to the value of statistical representations – increasing the level of understanding of the interrelationships between statistics, geographic phenomena and human endeavor through enhanced static and dynamic multi-dimensional visualisations.

The linkage between a geographical point location/boundary and statistical records is achieved through a computational process referred to as *geocoding* and *aggregation*. Geocoding is used to link statistical data at micro level, such as linking household data to a building, whereas aggregation is used at the macro level to represent a generalized interpretation of statistics by area, such as the demographics of village/district/suburb. Geocoding is used to spatially locate statistical unit records to a high degree of accuracy. This is achieved by transforming a description of a location—such as a pair of coordinates, an address, or a name of a place—to a location on the Earth's surface. The land parcel or property is often used as the geographic feature with which statistics are associated, and the address and/or coordinate location of the parcel or property is typically used to link the two elements – statistics and location. Geocoding can be done by entering one location description at a time but is mostly achieved by processing a table of records at one time.

There are several methods that belong to the geocoding family of transformations (Figure 4.8). The methodology is dependent on the type and format of the data, the geographic location geometry (point/polygon), and amount of information to be geocoded. Geocoding need not be an expensive task, and there are several commercial and open source services available.

Correspondence tables	• Correspondences tables are a mathematical method of converting data from one geographic region to another.
Allocation tables	• A table, or classification code list, listing codes for one set of geographic regions and their nested relationship to a second set of larger regions.
Point-in-Polygon	• A process that uses a Geographic Information System (GIS) to allocate records to a geographic region based on the location of a point geocode.
Address Geocoding	• A process of linking a textual addresses to a geospatially referenced object, either a coordinate or a small geographic region.
Locality Coding	• A process when limited location information, such as only village or local area name, is available to geocode records to a geographic region.

Figure 4.8: *Geocoding methods*

The UN Expert Group on the Integration of Statistical and Geospatial Information (under both UN-GGIM and UNSC) is developing the Global Statistical Geospatial Framework[4] (GSGF), which is designed to provide a common method for geospatially enabling statistical and administrative data, as well as ensuring that this data can be integrated with geospatial information. The 5 principles of the framework have been endorsed by UN-GGIM and UNSC, and the Expert Group continues work on elaborating guidance that supports the implementation of these principles.

For more Information on the on the Global Statistical Geospatial Information Framework, see Appendix 4.8)

The steps typically undertaken to integrate geospatial and statistical data are as follows:

1. **Develop a Strategy:** Bring focus and provide clear direction to the Geospatial and Statistical Data Integration Initiative through a common vision, mission, and goals. This will guide the development and achievement of targets as a whole.
2. **Establish a Working Group:** To direct, communicate and oversee the strategy and plan for integrating geospatial information and statistics data, as well as to formalize the relationships between participating agencies

4 GSGF is available at http://ggim.un.org/meetings/GGIM-committee/documents/GGIM6/Background-Paper-Proposal-for-a-global-statistical-geospatial-framework.pdf through agreements, and foster an institutional environment for collaboration and cooperation through the implementation of policy, guidelines, standards and common processes. The working group will typically report to the Geospatial Steering Committee (See SP1: Activity 1.6.1) and interact closely with the Geospatial Coordination Unit (See SP1: Activity 1.6.3).

3. **Identify the Key Stakeholders**: Data producers and users, decision makers etc. (See SP9: Activity 9.x.x).
4. **Identify the Available Resources:** Statistical and administrative data is typically available from National Statistics Organisations (statistical units, and social, economic, demographic, agricultural, environmental and census statistics etc.). Geospatial information is typical available from National Mapping Organisations (administrative boundaries, addresses, transport, water networks, elevation data, satellite imagery and topographic data, etc.). (See Section 4.6.2 for how to conduct a data inventory).
5. **Specify Policy, Standards, Guidelines and Norms:** Policies, standards, guidelines and norms are required to support the utilization, access, analysis and visualisation of integrated geospatial and statistical information including terms of use agreements, data release policies (See Section 4.6.15), privacy policies, data standards to enable interoperability, and guidelines for the security of information. In addition, it is necessary to elaborate the technical norms and regulatory agreements related to the standardization of the geographical elements to be represented. These policies, standards, guidelines and norms are best addressed and implemented at a national level and where possible, are to be compliant with international laws and common practices.
6. **Develop the methodology:** Develop the methodologies and procedures that standardize the collection, generation and maintenance of data, both geographic and tabular:
    - **Address**: Use a standard for the collection and assignment of addresses nationwide to effectively capture the address label and physical location.
    - **Coding**: Establish coding classifications for disaggregation into spatial entities, such as government administrative boundaries or census areas.
    - **Verification:** Implement address verification at point of address data entry (for computer-based capture/mobile applications) to better manage and maintain data quality and reduce time spent on data cleansing.
    - **Georeferencing**: Align geographic data to a known coordinate system so it can be analyzed, viewed and queried with other data.
    - **Geocoding**: Adopt common geocoding practices at the national level.

- **Quality**: Specify a quality model and measures for both geospatial and statistical data.
- **Retrieval**: Establish an effective method for storing and managing integrated geospatial and statistical data files so that they can be automatically updated and retrieved without the need for reprocessing.
- **Linking**: Evaluate open source, bespoke and COTS services for connecting data, with consideration to the OGC Table Joining Services standard[5].

7. **Make data accessible -** Use standard services (WMS and WFS) to serve data and distribute results over the Web.

Figure 4.9: *Steps for integrating geospatial and statistical data*

4.6.15 Formalize Data Supply Chains

A data supply chain strategy is used to formalize data sharing, invoke supply chain partnerships, and establish service level agreements between organisations.

The exchange of information between multiple organisations participating in a national data supply chain is often not well documented. This leads to potential data duplication and a doubling up of human resources, software and storage overheads.

To mitigate this situation, a national data supply chain strategy is used to formalize data supply chains. It encompasses four domains (Figure 4.10) (Arnold, 2017):

- **Business Domain:** The supply chain strategy delivers on the national business (or economic) outcomes required such as return on

---

[5] A simple way to describe and exchange tabular data that contains information about geographic objects. See https://www.opengeospatial.org/standards/tjs

investments and incentives for businesses. These are generally specified in the strategy along with the collective vision, mission and goals of the supply chain partners. The supply chain strategy also considers the value proposition to the end-user. This value stems from the collective efforts and activities of the supply chain partners.

- **End User Domain:** The supply chain strategy considers the end-user requirements, the factors influencing product usage behavior and the design criteria that will create the most value for the end-user, such as timeliness, content, coverage, semantics and accuracy.
- **Production Domain:** The supply chain strategy coordinates the external forces that impact on demand planning, data sourcing complexity and the types of transformation processes required to make a data product. A compliance framework is required to support interoperability including data and technology standards, quality measures and metrics, and custodian roles and responsibilities. Collaboration with supply chain partners is a key component to sustaining production in the longer term.
- **Service Domain**: The supply chain strategy focuses on connecting people to products and services. It considers the integration of component products from multiple sources to create standard offerings as well as tailored solutions. A policy framework is required to manage open access to data products balanced with individual privacy, copyright and intellectual property considerations (See SP2: Activity 2.6.x). These aspects are more complex in extended supply chain networks. Communicating product suitability will require a rating system that is meaningful in the end-user's context.

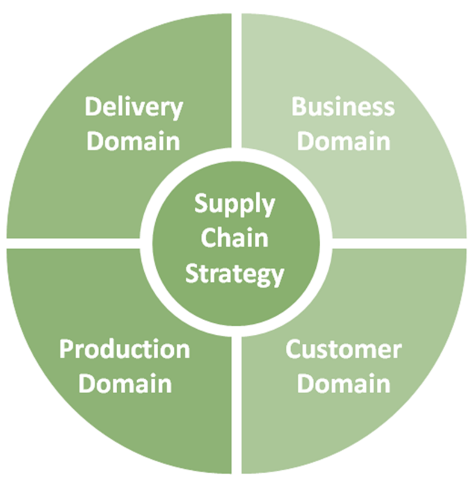


Figure 4.10: *The Four domains to be considered when developing a Supply Chain Strategy*

The National Supply Chain Strategy invokes supply chain partnerships and establishes service level agreements for updating and sharing data between organisations to ensure the ongoing currency of information. This is important. As soon as data is physically transferred from one agency to another, the data becomes stand-alone and either evolves into an obsolete dataset because it is not updated, or conversely, if it is updated the data becomes inconsistent in terms of currency with the parent data source. The national supply chain strategy formalizes consistent and timely data exchange.

Geospatial data supply chains in many countries have evolved to become a complex network that is difficult to visualise and manage. The relationships between suppliers, producers and consumers in national/cross-agency supply chains are difficult to enforce and understanding the origin of a piece of data is often challenging because of the ease with which data can be copied and transformed by individual supply chain participants.

Understanding the flow of geospatial data updates is crucial to maintaining, synchronizing and sustaining datasets over the longer term. However, creating supply chain models is not a simple task. Some datasets are stewarded from the bottom up, with local governments producing data that can be aggregated at the national level. Other datasets are managed through national programs with information transferred to the local level at specified intervals. In other cases, government agencies may acquire and use commercial data, project or collaborate with NGO's, citizens or volunteers.

There are often inherent complexities in existing data supply chains and therefore, developing a supply chain strategy provides an opportunity to review current capabilities and make an assessment on areas that could be done more effectively and efficiently. It is also an opportunity to consider new digital trends, appraise new methods (SP5 activity 5.x.x), and consider engaging with new supply chain partners, such as an external data aggregator or community volunteer geographic information providers.

There are three supply chain strategy levels that dovetail together to deliver a higher-level 'specifically' integrated supply chain strategy that is a plan for creating, delivering, governing and measuring spatial data content that is relevant to end users at national, jurisdiction and business levels. The approach recognizes that different organisations have their own strategic drivers (operational mandates), as well as meeting data sharing directives in the national interest. The three strategic levels are shown in Figure 4.11 (Arnold, 2019):

1 A **National Supply Chain Strategy** that is collectively developed, managed, adopted and monitored by channel partners, such as the digital cadastre supplier/producer representatives from each jurisdiction. It is generally an aspirational strategy.

2 A **Jurisdiction** (or sub-national or local government level) Supply Chain Strategy that encompasses national requirements as well as immediate

jurisdiction requirements, and formalizes local supply chain partner relationships

3 A **Business Strategy** (or Data Governance model – See SP: Activity 4.6.7) that is developed at an organizational level to deliver discretionary (or mandated) data products and services, as well as obligatory data products and services for jurisdiction and national end users.

The integrated strategic approach addresses the challenges of linking communications across supply chain channel participants of different orders (their level within the supply chain), and across broader policy initiatives and country-level development programs.

The Data Supply Chain Strategy is also an opportunity for organisations to evaluate their spatial data holdings and strengthen their data management processes. Supply chain strategies often bring clarity to data custodianship and responsibility for specific upstream and downstream activities.

Figure 4.11: *Integrated Supply Chain Strategy - three levels showing data flows and drivers.*

Also, given that data duplication may have occurred in the past, the supply chain strategy affords the opportunity to:

- consider which datasets represent an official dataset - fully endorsed by government for specific purposes, such as for the purpose of reporting on the Sustainable Development Goals; or
- put in motion procedures to have datasets recognised as an official source.

The Data Supply Chain Strategy is also opportunity for project sponsors to consider operating within a more sustainable supply chain framework rather than within an isolated project scope. Data supply chains are often a challenge for project teams because it is not always clear how their project-level data acquisitions, including community provided information, should be incorporated into mainstream data supply chains.

As government data supply chains can span several organisations, where to introduce project-level data can be problematic, particularly as each organisation in the supply chain contributes to its value in some way. Adding to this complexity, information flows can be bi-directional across organisations at the same level of government as well as between organisations at different levels of government. Because different data vocabularies, standards and schemas can occur, exchanging and reusing project-level data is not necessarily straightforward.

There are several methods available to organisations and project teams to streamline supply chains linkages. The most commonly used include:

- **Incremental Updating**, where project data updates are forwarded to the data custodian for the inclusion/integration into the authoritative database.
- **Database Versioning**, where updates are performed by the project team on a copy of the custodian's database, which is then integrated into the authoritative source by the custodian.
- **Direct Editing**, where updates are performed directly in the custodian's authoritative database.

With the exception of the direct editing approach, the integration of project-level data with the primary dataset is essentially a manual task. However, research is examining more automated methods to conflate project-level data acquisitions with authoritative sources (CRCSI, 2013). Current methods rely on the use of common standards so that data elements can be combined across datasets more readily.

In summary, the Supply Chain Strategy should:

- Assess existing supply chain capabilities relative to good practice.
- Consider the current and future business requirements of supply chain participants (contributors) as well as the needs of end-users.

- Evaluate new technologies and methods (SP5: Activity 5.x.x), and whether or not new core competencies are required (SP8: Activity 8.x.x)
- Consider new supply chain partnerships and community mapping programs and discontinue ineffective networks that lead to duplication of effort.
- Include a Supply Chain Road Map (multi-year) to improve supply chain interlinkages one step at a time. The Supply Chain Road Map should dovetail with the Data Theme Road Map.
- Gain buy-in and agreement for the Supply Chain Strategy from all participants.
- Consider how quality will be managed and coordinated along the supply chain including what standards need to be complied with (SP6: Activity 6.x.x), which organisation will be responsible for compliance, how quality control will be managed across the supply chain, and what procedures are required for quality assurance purposes.

4.6.11 Achieve Data Interoperability

Data interoperability is crucial to achieving integrated data supply chains. Having data that is interoperable means that systems and services that create, exchange and consume data have clear, shared expectations for the contents and contexts and meaning of the data [DISC, 2018]. In addition to promoting standardization for data sharing and reuse, interoperable data supports multidisciplinary knowledge integration, discovery, innovation and productivity improvements.

Data interoperability is crucial to achieving integrated data supply chains.

To be interoperable the data will need to use community agreed formats, language and vocabularies. The metadata will also need to use agreed standards and vocabularies and contain links to related information using identifiers (ANDS, 2019). For more information on common vocabularies, metadata and data standards for maximized Integration, see Strategic Pathway 6.

4.7 Deliverables

The following are products derived through the integrated geospatial information management development process under Strategic Pathway 4.

- Data Framework
- Data Inventory
- Data Profiles and Metadata
- Data Gap Analysis
- Custodianship Guidelines
- Data Release Policy
- Data Theme Strategic Plan and individual Dataset Road Maps
- National Data Acquisition Program and Plan
- A Fit for Purpose Geodetic Infrastructure
- Common Vocabularies/Data dictionary (that permit interoperability)
- Geocoded Datasets (that permit geostatistical data integration)
- Geospatial Data Supply Chain Strategy
- Robust Storage and Retrieval Systems

4.8 Outcomes

The following outcomes result from the execution of well-defined data supply chains and custodianship guidelines throughout the geospatial data life cycle:

- An increased range and scope of authoritative, integrated geospatial data available for decision-making and policy-setting to address economic, social and environmental challenges
- A critical mass of centrally coordinated data discovery to support national development and innovation leading to economic growth and improved quality of life for citizens
- Cost reduction through productivity improvements achieved via well-defined supply chains that eliminate duplication and make standardized data accessible to end users for integration and reuse
- Ability to monitor and measure progress towards achieving broad socio-economic benefits, including sustainable development goals through access to quality geospatial information.

4.10 Additional Resources

4.11 References

ArcticSDI (2019) ArcticSDI Strategic Documents [Online] Available at https://arctic-sdi.org/index.php/documents/strategic-documents/

Arnold (2019)

ANDS website. Accessed 2019 https://www.ands.org.au/working-with-data/the-fair-data-principlesDISC (2017) Data Interoperability Standards Consortium and is available at http://datainteroperability.org/

Government of Canada (2019) Open Maps [Online] Available at https://open.canada.ca/en/open-maps

Natural Resources Canada (2017) Spatial Data Infrastructure (SDI) Manual for the Arctic [Online] Available at https://arctic-sdi.org/index.php/documents/strategic-documents/

Natural Resources Canada (2019) Canadian Geospatial Data Infrastructure Resource Centre [Online] Available at https://www.nrcan.gc.ca/earth-sciences/geomatics/canadas-spatial-data-infrastructure/8904Open Geospatial Consortium (2019) Best Practices for Using and/or Implementing OGC Standards [Online] Available at ttp://www.opengeospatial.org/docs/bp

Permanent Committee for Geospatial Data Infrastructure of the Americas (PC-IDEA) (2013) Spatial Data Infrastructure (SDI) Manual for the Americas [Online] Available at https://unstats.un.org/unsd/geoinfo/RCC/docs/rcca10/E_Conf_103_14_PCIDEA_SDI%20Manual_ING_Final.pdf

UN-GGIM (2017) Report on Global Geodetic Reference Frame {Online] Available at http://ggim.un.org/knowledgebase/KnowledgebaseArticle51654.aspx, Accessed January 2019.

UN-GGIM Actions Forward – From a UN mandate to a roadmap for global geodesy [Online] Available at , Accessed January

UN-GGIM (2014) Unleashing the Power of Where

(Nebert, 2004). SDI Cookbook