APPENDICES

STRATEGIC PATHWAY 5: INNOVATION

APPENDIX 5.1: IGIF Technology Maturity Index

Levels of Maturity Generational Bracket	Level 1: Analogue Mapping	Level 2: Digital Cartography	Level 3: Geographic Information Systems	Level 4: Spatial Data Infrastructure	Level 5: Integrated Geospatial Information Management
Focus	Map Production	Product-Based	Process-Based	User Centric	Knowledge On-Demand
Operational Level	National, Subnational, Private Sector	National	National, Subnational, Private Sector	Cross-Sector Integration	Global Network
Data Supply Patterns	Siloed Production	Siloed Production and Delivery	Informal Individualised Supply Chains	Formalised Hierarchical Supply Chains	Published Direct to the Web
Storage	Plan Press	Computer Hard Drives, Portable Hard Disk	Optical Disk/ Mainframe Computing	Cloud Storage	Cloud/Edge Computing
Acquisition	Photogrammetry	Digitisation/ Scanning	Digitisation/ Image Interpretation	Automated Image Interpretation, Social Media, Crowdsourcing	IoT sensors, Machine-learning, Artificial Intelligence
Access	Counter sales	FTP Sites	Web Portal (multiple agency portals)	Centralised National Web Portals	Global Web of Data
Data Formats	Paper Maps	CAD (2D)	GIS (2D)	GIS (Discontinuous, 2D, 3D, 4D)	Linked Data (Seamless 2D, 3D, 4D)
Users Services	NA	NA	NA	Data Catalogue/ Security Services	Brokering Services
Standards	Ad-hoc Technical Specifications	Organisation-Wide	National/ISO	ISO	OGC/W3C
Knowledge Representation	Map Legend and Production Notes	Analogue Metadata	Digital Metadata	Digital Metadata and Provenance	Metadata, Provenance, Domain and Process Ontologies
User Domain	Government	Government	Government, Private Sector, Academia	Government, Private Sector, Academia, Community, Citizens	Everyone
Analytics	NIL	Predominantly Analogue Analysis	Digital Analysis, Manually Executed Algorithms	Automated Algorithms	Real-time Query Responses
Reference Frame	Map Projection	Various Map Projections/ Datums	National Geodetic Datums	Global Reference Frame (Static)	Global Reference Frame (Dynamic)

APPENDIX 5.2: Capability Framework Matrix

Capabilities	Current State	Desired Situation	Gaps in Capability	Enablers	Priorities
Focus					
Operational Level					
Data Supply Patterns					
Storage					
Acquisition					
Access					
Data Formats					
Users Services					
Standards					
Knowledge Representation					
User Domain					
Analytics					
Reference Frame					

APPENDIX 5.3: Geospatial Drivers and Trends

The Geospatial Drivers and Trends determined in the Third Edition of the Future Trends Report, 2020.

Trends

- Relevance of ***data integration and interoperability*** increase
- Products and solutions produced from ***multiple data sources*** becoming the norm
- ***New opportunities for data gathering***, i.e. autonomous vehicles
- ***Crowdsourcing and VGI*** become established ways of data collection
- ***High-resolution high-revisit Earth Observation*** data become valid alternative to aerial imagery
- ***Big Data processing*** has become a normal path of geospatial data processing
- Integration of multiple data sources requires ***licensing harmonisation***
- ***Digital platforms*** provide access to data at scale
- ***Linked Data*** enables knowledge-on-demand

Drivers: Rise of new data sources & analytical methods

- ***Ubiquitous connectivity*** enables deployment of new tech
- ***Digital infrastructure*** through sensors and the Internet of Things
- Interconnecting modes of transport through ***intelligent mobility***
- ***Digital Twins*** for modelling, simulation and prediction
- Wide uptake of ***edge computing*** to enable intelligent mobility, the Internet of Things, and smart cities
- Visualisations and ***immersive technology*** widely used to enhance customer experience and decision making
- ***Machine learning, deep learning, and AI*** disrupt geospatial production
- ***Quantum computing*** enables intensive processing

Drivers: Technological advancements

- Rise of products and services specifically designed for the ***urban environment***
- Demand for ***real-time information*** provision
- ***Digital divide and exclusion*** continue to hold back universal digital transformation
- Seamless experience between ***outdoor and indoor mapping*** becomes an expectation
- Viable integrated ***Smart City*** solutions becoming wide spread

Drivers: Evolution of user requirements

- Increased ***diversity at work*** in technology, science, and innovation
- ***Talent and consumer shift*** - changing values and attitudes
- ***Incubator spaces*** enable innovation to enter markets swiftly
- ***Regeneration of business ecosystem*** through the rise of non-geospatial start-ups
- ***New collaboration agreements*** with industries outside of geospatial emerge

Drivers: Industry structural shift

- ***Digital ethics and privacy*** addressed by national and international initiatives
- ***Cybersecurity*** conversations increase in tandem with increase in digital devices
- ***Pace of digital and tech change*** puts pressure on national institutions to address policy and legislative shortcomings
- ***Pressure on government institutions*** to be more tech and digital savvy

Drivers: Legislative environment

APPENDIX 5.4: ICT Data Inventory

Hardware

Hardware (CPU) Monitor, Printer, Keyboard, Mouse etc)	Hardware Capcity (RAM and CPU)	Model	Serial Number	Date Purchased

Software

Software Product	Version	Serial Number	Source of Acquisition

APPENDIX 5.5: PEST and SWOT Analyses

PEST Analysis

The PEST Analysis considers the external environment and focusses on the **P**olitical, **E**conomic, **S**ocial and **T**echnology issues that may have a positive or negative impact on the implementation of integrated geospatial information management.

An example of issues that may be raised during a PEST Analysis are presented below.

POLITICAL	ECONOMIC	SOCIAL	TECHNOLOGICAL
• Safer Country • Policy and legislation • E-Government • Regional Needs • Sufficient government support and Funding • Copyright and Intellectual Property • Value & importance to the country	• Investment Opportunities for revenue growth • Savings • Modernization and maintenance • Professional Skills • Plant, equipment and personnel availability • Public-Private Partnerships	• Institutional Culture • Community needs • Intergenerational issues • Geographic and geospatial education capacity • Computer literacy • Community safety	• Data quality • Legislation • Technology level • Power (utilities) availability • Broadband capacity • Standards, Metadata etc. • Innovation

SWOT Analysis

The SWOT Analysis considers internal factors; namely the characteristics of government that are an advantage or disadvantage to geospatial information management, those aspects that can be exploited to advantage through IGIF implementation and those realities that are a threat to implementation. While threats are often considered to be out of people's control, they may still have an influence on outcomes and be able to make a valuable contribution. For example, while legislators are responsible for drafting Open Data Policy, geospatial professionals will have specialist expertise that shapes the policy to bring out the opportunities.

- **S**trengths: characteristics of government that are an advantage to the IGIF Project.
- **W**eaknesses: characteristics that place the IGIF project at a disadvantage
- **O**pportunities: elements that the IGIF could exploit to its advantage
- **T**hreats: elements in the environment that could cause trouble for the IGIF during and after implementation.

An example of issues that may be raised during a SWOT Analysis are presented below.

STRENGTHS	WEAKNESSES	OPPORTUNITIES	THREATS
• Leadership • Skills • Technology • R &D • Community demand •	• Missing, outdated, or sub-standard data themes • Policy • Cross agency collaboration • ROI and Pricing Models	• Expansion of data use • New applications • Community crowdsourcing • Government branding • Community trust	• Free data policy impacts on ROI • Change in policy • Consumer behaviour • Obsolete Technology • Insufficient resources

APPENDIX 5.6: Examples of Modernizing Data Assets

Today, governments can choose from an array of methods to capture and modernize (including transitioning from maps to digital data) geographic information. Methods include:

- **Flatbed Scanning:** Much like a photocopier, large format flatbed scanners are used to scan, geo-reference, geocode and uniquely identify large maps and historical aerial photographs (film or printed images) so that they can be incorporated into a GIS or preserved in a digital archive.
- **Line Scanning:** Line scanners are used to digitize the lines, points and text on a map – firstly by creating a raster image of the map and then vectorizing the adjacent pixels to form linear features. Background noise such as dirt or imperfections in the map are typically cleaned as part of the scanning process.
- **Digitizing (and Visual Image Interpretation):** Digitizing is the process of examining, identifying and tracing features, such as roads, vegetation and buildings), recorded on a map or scanned image to create vector data (coordinate data - points, lines and polygons). The vector data, which delineates the features identified, is able to be incorporated directly into a GIS. The manual digitizing process can be highly reliable, however the process is time consuming and typically requires a skilled analyst who has familiarity with the area.
- **Automated Feature Extraction (and machine-learning):** A number conventional digitizing and image interpretation techniques have now been successfully automated through supervised and unsupervised machine-learning techniques. Geographic features, such as can be automatically detected and extracted as vector data using specific algorithms that analyze aerial and remotely sensed imagery.
- **Paper-based Maps:** In many developed countries geospatial datasets were originally created from digitizing analog maps. This was achieved using a large format scanner and geo-referencing the images to known points in a GIS. While scanning paper maps is now an outdated data capture method, having scanned historical images is crucial for chronicling cadastral boundaries, land rights and land use records to verify ownership on land titles. Additionally, lines scanners still have a role to play in land boundary capture. Often cadastral paper plans are the only record of ownership and this information can be digitized using line scanners. However, depending on the quality and scale of the map, and when the map was last updated, the spatial accuracy of the boundary will be questionable and require manual verification against land registry documents or field validation. Paper maps are also important to preserve where they have been annotated by hand, such as the recording of flora and fauna sightings that have historical significance.
- **Limited or No Geospatial Data:** Often countries will have geospatial information but lack data on important themes. For example, understanding the location of buildings, their shape and structure is essential for emergency management and exposure risk to natural hazards and disasters, however these features have often not been collected. This is because the traditional digitizing methods (including photogrammetric interpretation methods) still used some agencies are prone to interpretation error and the features are difficult to update over time. This scenario often applies to the capture of water feature (dams, swimming pools and springs) critical for water resource management, and offshore rocks that are

important for maritime navigation and territorial waters etc. These features can be collected using high-resolution imagery combined with machine-learning techniques, which provide a rapid means of data collection as well as automated feature validation using training sets. Some features, such as points of interest (monuments, shops and amenities) can be collected via crowdsourcing or focused VGI groups.

- **Partially completed geospatial datasets:** Most countries have embarked on their geospatial information management journeys and many are keen to extend data coverage and implement regular updating programs, typically focusing on densely populated areas and priority data themes. The challenge for these countries is whether to continue with traditional methods or move to more contemporary methods such as automated feature extraction methods. The challenge lies in being able to conflate existing datasets with the new information. In some cases it may be is more efficient and effective to start certain datasets from scratch.
- **Comprehensive geospatial datasets**: There are many countries with comprehensive datasets covering a broad range of themes. However, the maintenance of these datasets is often problematic. Most organizations rely on digitizing methods to update databases; or business processes that result in a 'just in time' spatial transaction, such as the land development process that triggers new land boundaries, addresses, road names, and settlement boundaries. For some data themes, such as buildings, moving to automated feature extraction methods is the next innovative step; for others it the development of rule bases to automate the procedural processes, such as the approval of new road names.

The methods used to capture data will vary depending on the type of data. Data collection typically falls under 3 main categories: (1) information that is created through a business process (e.g. address data); (2) information that is derived from imagery and sensors (e.g. building footprints); and (3) information that is derived through precise measurement (e.g. land subdivisions). This is a challenge when formulating plans to escalate production as there is no single solution. For example, property street addresses (text files), building footprints (GIS Polygons) and 'As Constructed' engineering drawings (2/3D models) – have different data formats, purpose, change frequency and update triggers. Therefore, what is an optimum data collection method for one feature type, will not be suited to another.

APPENDIX 5.7: Examples of Modern Data Creation Methods

Some examples of modern data capture and creation can be achieved by acquiring Earth observation data for mapping and monitoring, GPS and GNSS for location-based services, targeted imagery via unmanned aerial systems, data streaming from autonomous vehicles and IoT Sensors, and accurate 3D city models using laser scanning and photo mesh modelling.

- **Earth Observation Data:** Satellite Imagery and artificial intelligence has come of age and is now being used automatically in extracting features and intelligence from imagery to support topographic mapping, cadastral mapping, environmental monitoring and disaster management and recovery. In the Earth observation domain, nanosatellite technology are now providing a low-cost solution to rapidly acquiring Earth observation satellite imagery for the entire globe on a daily basis - providing valuable evidence for climate change studies and resilience management. Now able to provide 50 centimeter resolution, satellite imagery is proving a more affordable high-resolution option than aerial imagery.

  In addition to commercial imagery, there are a number of data hubs worldwide that provide free access to Earth observation data, such as the Copernicus Open Access Hub[1], USGS Earth Explorer[2], Global Visualization Viewer (GloVis), ESA Earth Online[3], International Institute of Space Research (INPE)[4], Buhvan Indian Geoplatform[5], Japan Aeropsace Agency (JAXA) Global ALOS[6], NASA Earth Observation Data[7], NOAA Data Access Viewer[8], and VITO EO Data[9].

- **GPS and GNSS:** Global Positioning System (GPS) is now the most widely used Global Navigation Satellite Systems (GNSS) in the world (REF). It provides continuous precise positioning and timing information globally, and under any weather conditions. GPS/GNSS is used in a variety of fields, such as agriculture, transportation, machine control, marine navigation, vehicle navigation, mobile communication and athletics. Both GPS and GNSS consist of three major segments: the space segment (satellites), the ground segment (ground control stations), and the user segment (GNSS or GPS receivers).

  Satellites are continuously sending radio signals toward Earth, which are picked up by GNSS or GPS receivers. Countries that have limited cadastral data are using GNSS and GPS receivers to survey land boundaries. While ground positioning is not as precise as traditional surveying methods, in many cases the boundary data is fit for recording land ownership and land transactions.

---

[1] Copernicus Open Access Hub - https://scihub.copernicus.eu/dhus/#/home

[2] USGS Earth Explorer - https://earthexplorer.usgs.gov/

[3] ESA Earth Online - https://earth.esa.int/web/guest/home

[4] INIP - http://www.dgi.inpe.br/CDSR/

[5] Buhvan Indian Geoplatform - https://bhuvan-app3.nrsc.gov.in/data/download/index.php

[6] JAXA https://www.eorc.jaxa.jp/ALOS/en/aw3d30/

[7] NASA Earth Observation Data- https://neo.sci.gsfc.nasa.gov/

[8] NOAA - https://coast.noaa.gov/dataviewer/#/

[9] VITO - http://www.vito-eodata.be/PDF/portal/Application.html#Home

GPS devices can be used in a number of contexts including, car navigation, traffic congestion maps, and location-based searching. The GNSS network is continually being modernised. The Russian GLONASS network reached full operational capacity in 2017, providing instantaneous high-precision real-time position data. The European GNSS, Galileo, is under development and the Chinese Beidou system is expected to provide global coverage in 2020 (Figure 1).

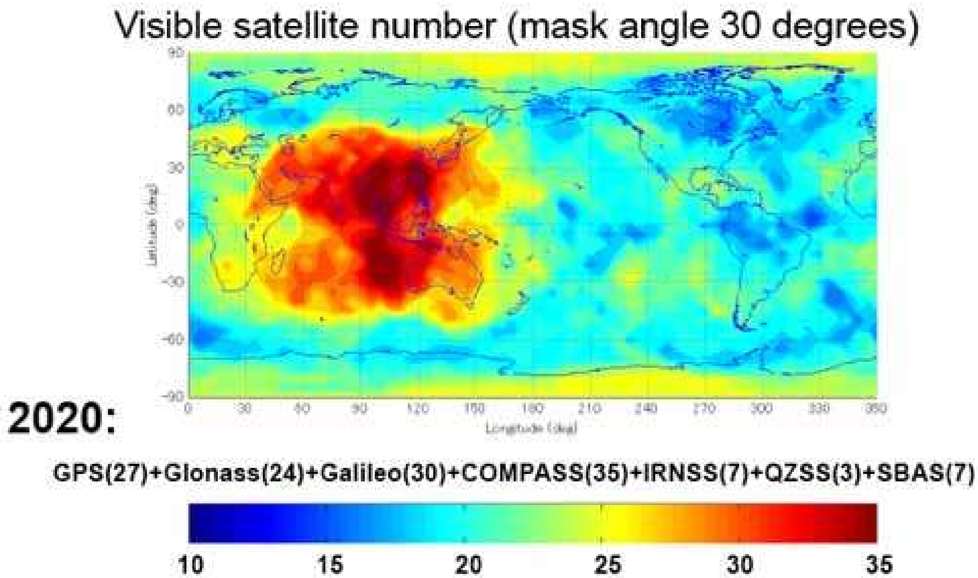


Figure 1: *Map of satellite availability from global and regional satellite navigation systems by 2020 (REF)*

- **Unmanned Aerial Systems (UAS):** Commonly known as drones, UAS have matured significantly in recent years. The aircraft has improved to the point where medium format cameras can be used to produce good quality and highly accurate imagery. The two most common UAS's used are the vertical take-off multi-propeller craft and the static take-off craft that requires pre take-off flight planning. UAS have the ability to change payloads, such as swapping to LiDAR sensors to acquire 3D point clouds over small project areas.

  The advantages of UAS include fast capture for areas less than $15km^2$, the ability to produce high-resolution data even on cloudy days, and cost effective small mapping projects such as narrow corridor mapping like transmission lines, pipelines, and roads.

- **Autonomous Vehicles:** Whilst not yet mainstream, autonomous vehicles are one of the most significant technological developments of the last 20 years, and the level of commercial investments worldwide make it hard to ignore. So too are the publicized benefits to consumers – increased mobility, reduced traffic congestion, lower emissions, reduced energy consumption and improved road safety, all make for a compelling argument for driverless technology.

  With GPS receiver's inbuilt, sourcing geospatial information from vehicle sensors will constitute a large source of data in the future. Already companies are using the perception capabilities of human-driven vehicles to update cloud-based high-definition maps. They use vehicle sensor data to detect semantic features and landmarks, such as road boundaries, lanes, signs, traffic lights etc. and these features are automatically compared with existing maps and updates performed where necessary.

- **IoT Sensors:** Sensors are increasingly being used to measure a physical phenomenon (like temperature, water levels, and air quality etc.) and to detect changes in the environment over time. By itself, a sensor is ineffective, but when it is integrated in a digital network, it can be used to transfer data via electronic signal without requiring human-to-human or human-to-computer interaction.

  When connected to the Internet, data from sensors can be integrated with geospatial information and used across a range of location-based applications. For example, optical sensors are used to detect vacant car bays helping motorists to find parking easily; water quality sensors detect changes in oxygen levels providing early warning of the presence microorganisms from sewage, urban or agriculture runoff or discharge from factories; and vibration sensors are used the monitor structural health of building bridges and historical monuments. These IoT sensors can transmit and process data locally (on the sensor itself or at the nearest network node) and are suited to real-time applications.

- **3D Laser Scanners:** Laser scanners have been available for quite some time. They provide a cost effective and fast approach to capturing 3-dimensional models of urban environments. The laser scanner produces millions of coordinate points (point clouds) of geographical features, such as buildings and road assets. Mobile laser scanning (MLS), particularly vehicle-mounted laser scanning systems, are in the process of becoming a newly established way in measuring road and urban environments. Further, with the development of algorithms that enable simultaneous localization and mapping, MLS has progressed to deliver 3D data from environments not covered by Global Navigation Satellite System (GNSS), such as tunnels and indoor environment. These digital replicas permit accurate measurements of any environment – indoors, outdoors, underwater, etc.

- **Sub-surface Surveying:** Until recently, the sub-surface infrastructure was hardly considered by those developing utility networks and urban underground transportation systems. The little data that does exist is often fragmented and outdated. More so, studies have shown that uncertainty in the location of underground assets can have serious knock on effects on the economy as missing or inaccurate data can cause construction delays.

  Economic development is a key driver for the application of geospatial in this field. The lack of information regarding the location of buried assets increases the cost of its management and maintenance and, therefore, has a direct impact on a nation’s economy. Quantum sensing is said to revolutionize how surveys are performed, by using sensitive sensors that can detect even the slightest fluctuation in gravity speeding up survey times and increasing accuracy. Nonetheless, it should be noted that, whilst quantum sensors exist, they only do so within a laboratory environment and the resolution required to meet the expectations of underground assets surveys are yet to be developed.

- **Photo mesh mapping:** An alternative to laser scanning for mapping 3D urban environments is photo mesh mapping. Mesh mapping is used to rapidly capture large areas. Often captured from helicopter using a hand held camera, photo mesh mapping generates a dense triangular mesh of a scene from several digital still images and a sparse point cloud. Because these mesh maps can be generated quickly, they are often used for coastal monitoring (Figure 2).

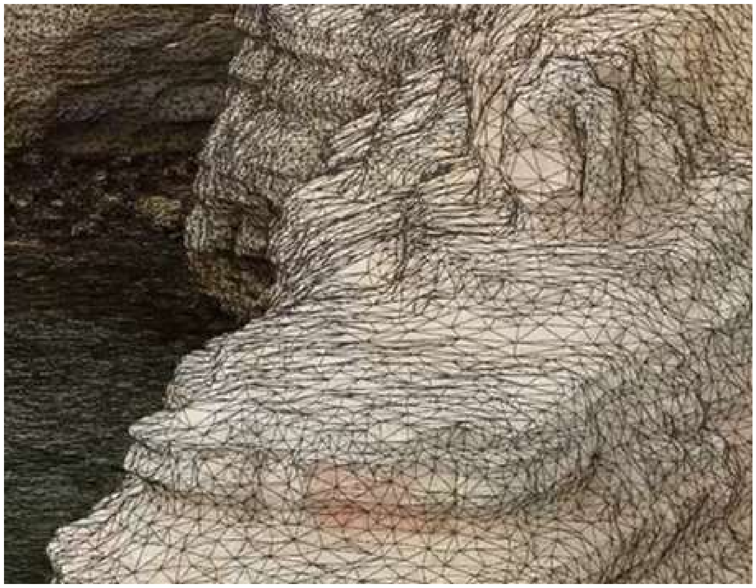


Figure 2: *Photo Mesh Mapping of Onkaparinga cliffs, South Australia (Photograph courtesy AEROmetrex).*

- **Volunteered Geographic Information:** Community mapping partnerships, commonly referred to as Volunteered Geographic Information (VGI), and more generally crowdsourcing, enable governments to tap into the collective intelligence of the community to assist in the collection of geospatial information that organizations would otherwise not collect. This is an important option, as the collection of geospatial information is time consuming, and resourcing this task is often a challenge for governments. Technology today makes it possible to collect information quickly by enlisting a large number of non-technical people from the broader community to do mapping tasks. This has significantly enhanced the availability of geospatial data globally, but at the same time, blurred the lines between authoritative and non-authoritative data.

  VGI has proven effective for continuous (almost real-time in some situations) updating of specific geospatial datasets, such as road networks, buildings and addresses etc. VGI has the potential to reduce government costs associated with mapping as well as creating better data for the benefit of citizens. There are several excellent open-source solutions that demonstrate the value of engaging with the community, such as OpenStreetMap, created by volunteers using local knowledge. In some cases, crowdsourced maps may be more current and accurate than the official 'authoritative' data.

  The notion of this type of community input is not new, and in some developing countries, crowdsourced geospatial data is used for social and economic development, particularly in areas where no, or only limited, data is available. As well as generating data, VGI is a powerful mechanism for encouraging public participation and empowering citizens in the use of geospatial information. However, questions around quality continue to hold back the wider uptake of crowdsourced information by public bodies of developed countries.

There are two key challenges that organizations will face:

- Being able to effectively integrate and validate crowdsourced data into authoritative data systems of government. Manual data verification processes are time consuming and workflows will need to be critically analyzed to achieve process improvements.
- The need to build awareness within the community so that VGI is adopted and supported by a large number of citizens. This will require a range of communication strategies.

- **Data Harvesting:** One of the best ways to find geographic information is to search the Web. Rather than collect some information themselves, governments are now accessing and consuming authoritative data as web services, which have been published to the Web by other agencies.

  Data harvesting is often used to develop Regional Data Hubs based on data services. For instance, the Arctic Spatial Data Infrastructure (SDI)[10] is a regional model of coordinated data flows for serving and harvesting data. The Artic SDI leverages machine-to-machine data access via a web service API to:

  - create and maintain a "service of services" that harvests data from each participating National Mapping Authority (NMA) to create the Arctic SDI base map and gazetteer of place names; and
  - provide catalogues services for other authoritative external data providers (e.g. academic, NGOs) that are integrated into a data viewer (GeoPortal).

  The collaborative expertise of multiple NMAs provides the additional benefit of improving the accuracy of the data, such as increasing the resolution of the Arctic Digital Elevation Model.

  A further example is the Government of Canada's Federal Geospatial Platform (FGP). This platform is an integrated model for visualizing, analyzing and disseminating federal government data. The FGP provides a base layer (or framework data) of commonly used fundamental data themes It utilizes services to harvest data from other federal departments – thus giving end users access to other thematic layers from a central location.

  Data harvesting offers significant opportunities to enrich existing and future geospatial databases. As such it will remain increasingly important for governments to facilitate coordination between all sources of information, finding new ways to join information into nationally recognized datasets. Key to the successful delivery of integrated data will be the development of standards and methods to assure the quality and fit of different information sources.

  Global and regional web portals often provide web services and thus data harvesting opportunities to global, regional, national and local content. The type of data, content and areas of coverage will differ between portals, and information is not available for all countries. Nonetheless, these portals are worth

[10] Arctic SDI [Online] available at https://arctic-sdi.org/

investigating as they provide a pool of rich data resources that are available for use, and associated metadata is available to enable an assessment of fitness-for-purpose.

Examples of freely available global datasets include Esri Open Data ArcGIS Hub[11], Natural Earth Data[12], Socioeconomic Data and Applications Centre (SEDAC)[13], Open Topography[14], UNEP Environmental Data Explorer[15], FAO) GeoNetwork[16], Global Map Data Archives[17] and Terra Populus[18], which integrates census (from 160 countries) with environmental data - land cover, land use and climate.

---

[11] ArcGIS Hub - https://hub.arcgis.com/search

[12] Natural Earth Data - http://www.naturalearthdata.com/downloads/

[13] SEDAC - https://sedac.ciesin.columbia.edu/

[14] Open Topography - http://opentopo.sdsc.edu/datasets

[15] UNEP - http://geodata.grid.unep.ch/

[16] FAO GeoNetwork - ,http://www.fao.org/geonetwork/srv/en/main.home

[17] Global Map Data Archives - https://globalmaps.github.io/

[18] Terra Populous - https://terra.ipums.org/

APPENDIX 5.8: Examples of Data Integration Approaches

Data integration is about methodically blending data from different sources to make it more useful and valuable for analysis. It is not about moving data into a single repository as such, although warehousing is often referred to as data integration - it is about making the data more comprehensive and usable.

Data integration in many countries is problematic, because there are too few linkages across key government datasets making the relationships between geographic features difficult to analyze. When data are not integrated evidence-based decision-making is difficult for end users, particularly non- geospatial experts. For example, it is not easy for a land developer to query a land parcel to understand what can be done on land (land use rights); what cannot be done (restrictions due to landslide risk), and what must be done (reforestation of cleared areas). Nonetheless, accessible data can be integrated – the problem is that this integration process has to be done by each end user.

Data Integration means different things to different people and there are multiple ways to integrate data - data federation, data conflation, update propagation and Linked Data (Figure 1).

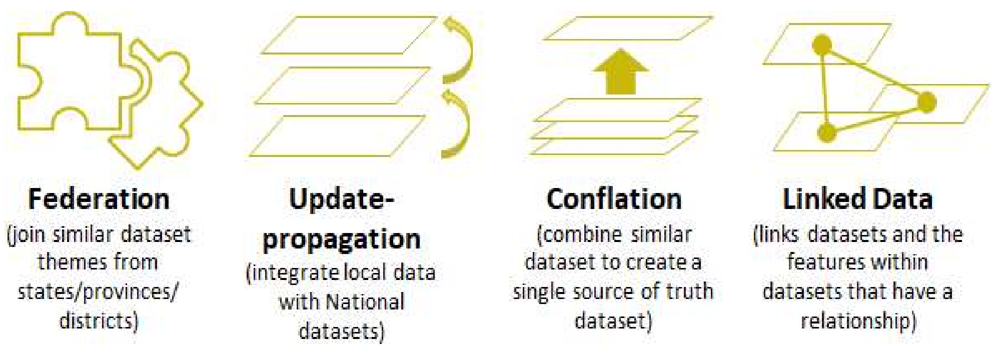


Figure 1: *Data Integration Methods.*

From an innovation and process improvement perspective, these integration processes enable geospatial information to be (a) viewed seamlessly across administrative boundaries (i.e. through data federation), (b) managed as a single source of truth data set and eliminate duplication (i.e. through data conflation); (c) synchronize updates between the various levels of government (i.e. update propagation; and (d) interlink datasets and expose relationships between data at the feature (i.e. through Linked Data). The four data integration concepts are explained below:

- **Data Federation:** Federated data is a unified view of heterogeneous data from different systems, created using a virtual database and a common (or harmonized) data model. Data federation is commonly used to bring State/Provincial/Municipal dataset together to form a seamless country-wide dataset – just like joining the pieces of a puzzle. Once federated, data can be analyzed from a national perspective through applications.
- **Data Conflation:** Data conflation is the compilation of two (or more) different geospatial datasets covering overlapping regions. Data conflation combines the best quality elements of both datasets to create a composite dataset that is better than either of the original datasets.

This integration technique is used when similar datasets have been created by different departments, such as geographic points of interest, in the course of their business, and there is a need to have a single authoritative dataset to reduce government duplication. Conflation is commonly used when each organization's dataset has valuable but slightly different information. With conflation processes is possible to retain the accurate data, remove redundancy, and reconcile data conflicts.

- **Update-propagation:** Is an event-driven data integration process used to transfer (propagate) updates performed in a local database (municipal/district/state dataset) to a national dataset. The process can be done either synchronously (at the same time) or asynchronously (at specified time intervals). Some synchronous data propagation software also supports a two-way data exchange between the source and the target, meaning that updates at a national level can be propagated to the local level. However, a prerequisite to update-propagation between autonomous databases is to develop equivalence between the objects that represent the same real world phenomena. This calls for: (a) the schema of one database to be linked to the other; and (b) the matching of corresponding object instances so that explicit relationships can be defined. Updates can then be propagated using a rule base developed as a result of the schema integration and data matching process.
- **Linked Data**: Linked data is the ultimate standard for creating interoperable datasets. The term is used to describe a method of exposing and connecting data from different sources on the Web. Linked data is structured information that is in a machine-readable format. With Linked data it is possible to access more complete information across "data silos" of information on the Web. In contrast, much of government data is structured in tables or graphs in PDF or Word documents or within HTML web pages. For this data to be reused by others, it typically needs to be re-entered by hand, or copied and manipulated into a useable format. This adds significant time and effort to reusing the information.

Data integration using the Linked Data approach requires geospatial information to be structured according to a standard-like data format. The notion is that once all data is published the same way (i.e. as Linked Data), it is possible for machines to search for, interrogate and analyze huge volumes of data. To do this, data needs to be published according to the Linked Data standard - the Resource Description framework (RDF) format, and have a Uniform Resource Identifier (URI) to identify it as a unique data resource as well as a Uniform Resource Name (URN) for identifying the namespace.

With this structure in place it is possible to add structural relationships between the datasets (right down to the feature level) to enable knowledge inferencing and complex semantic data querying possible. Search engines can then find the most appropriate data to answer a user's query and quickly infer the answers they need.

No matter how data is stored, it is possible to transform it into Linked Data and publish it as part of a giant global graph on the Web. There are several workflows for publishing Linked Data and the choice will depend on the data volume, rate of change, and how the data needs to be managed for legacy systems. Releasing this data in an open and machine-readable manner makes the data much quicker and easier to reuse and can be as simple as exporting raw geospatial data from the source in an appropriate format (such as Comma Separated Value (CSV)).

APPENDIX 5.9: Data Storage Processes

The increasing volumes of geospatial data being generated today creates challenges on how this data is stored, maintained, and used. While data archiving, backup and retrieval systems have advanced considerably, for many developing countries the primary form of data storage, is still the desktop computer. Using desktop storage is problematic – data is often inaccessible, as information can only be accessed by the computer owner/user, and in some cases this is the only copy of the information. Therefore, risk of information loss is extremely high, as hard drive failure is common and data can easily become corrupted. The lack of data storage is typically a financial problem. Although decreasing in cost, data storage is relatively expensive to buy and costly to manage, which is why desktop computers and portable hard disk storage remain the main storage systems in some organizations. Urgent reform is required.

Typically, organizations need to consider three data storage processes – data archiving, backup and warehousing:

- Data archives are used to store historical data (typically original data) that is not actively used, but needs to be retained in the long-term as a record, and/or for future reference. Sometimes archived data is stored offsite and on cheaper storage devices, such as tapes.
- Data backups are copies of current data that can be restored if the current data is ever damaged or lost. Data backups are often incrementally overwritten after a specified period to free up storage.
- Data warehouses are repositories of structured, filtered information that has already been processed for a specific purpose. The data stored in the warehouse is searchable and comprises catalog and metadata indexes for data discovery. Data warehouses typically need to store a large amount of data that is amassed from a wide range of sources from within and across organizations.

Computer Network-attached storage (NAS) are used by many organizations as their dedicated file storage systems. NAS drives enable multiple users to retrieve data on a local area network (LAN) or via a standard Ethernet connection. They are built more robustly than a desktop computer hard drive and are often the preferred choice for developing nations. However, the cost of NAS drives often makes this prohibitive.

Cloud-based storage systems offer a modern alternative and provide scalable, cost-effective data storage and computing resources that can be used as and when required. These computing resources are anchored by third-party providers that offer a combination of physical hardware, networks, storage, services and interfaces that are needed to deliver computing as a service. Cloud computing and the internet have transformed the way in which organisations are able to manage data. An increasing amount of data is now being stored in cloud services. There are several advantages, including the opportunity to access, adjust and share information more efficiently, thus increasing economies in data storage.

However, despite the increasing use of cloud computing, the infrastructure required to access and use these systems is not always available to developing countries whom are unable to meet the needs of basic electricity and high-speed internet connectivity for accessing, sharing, and processing large quantities of data. Additional barriers include a lack of adequate legal and regulatory frameworks for electronic commerce and cybersecurity, concerns about data privacy and security, the location of the data and reliability of services, and lack of skills to make effective use of ICT.

To take advantage of a cloud computing model, governments will need to create an enabling policy framework so that organization's wishing to migrate data and services to the cloud can do so easily and safely. This will require consideration of different business models and services available including public, private or hybrid clouds, at national, regional and global levels.

Moving forward, policy makers in developing countries may need to research and analyze the level of cloud readiness in their country, and the potential implications of cloud adoption. Then, prepare a cloud strategy, which addresses infrastructure, legal and regulatory issues, and the supply side of the cloud economy ecosystem - human resources, whole-of-government cloud use (for potential economies of scale) and financial implications.

APPENDIX 5.10: Pillars of an Innovation Program

Innovation programs typically include five pillars – innovation management, innovation infrastructure, the internal innovation community, open innovation community and monitoring and evaluation:

- **Innovation Management:** This includes the strategic intent and focus of the Innovation Program, the leadership and governance mechanisms, as well as the mechanisms for enacting change, such as the executive leadership, innovation program management and change management. The Innovation Management Team will need to be empowered to:
  - Endow grants for any partnerships, alliances or reciprocal arrangements that will result in the development of innovative solutions;
  - Provide funding to assist staff members in the pursuit of innovative solutions;
  - Enter into any contractual arrangement, for the purposes of developing innovative solutions;
  - Procure services, solutions, hardware and software;
  - In consultation with Managers take staff out of their substantive positions to temporarily work on an innovation project; and
  - Fund, execute and monitor projects, and apply appropriate project management controls.
- **Innovation Infrastructure:** This is the physical infrastructure of the innovation program that may include centers of excellence and innovation labs as well as tools such as innovation portals that support communication and collaboration among the various members of the innovation community. The infrastructure may also include an innovation database (repository) that provides procedural information for conducting innovation, such as procedures for:
  - Assessment processes, including feedback mechanisms on 'innovative ideas' generated by staff
  - Intellectual property rights management
  - Risk profiling
  - Legal advice and In-house Counsel
  - Commercial Confidence
  - Project Approval and Funding
  - Roles and Responsibilities
  - Project Reporting and Performance
- **Internal Innovation Community:** This is the staff-facing part of the innovation program that may include establishing innovation communities of practice within an organization and setting up award and recognition schemes. Internal communication is an extremely important part of the innovation program and consideration needs to be given to establishing staff feedback mechanisms. This is vital to achieving corporate goals of institutionalizing innovation within the organization (Evans, 2015).

- **Open Innovation Community:** This is the external-facing portion of an innovation program, and it is likely to a central part of a geospatial innovation strategy. It includes grant programs and crowd-sourcing for bringing in ideas from the outside, together with collaboration in various innovation consortia for best practice sharing. Alliance partners programs are often effective for stimulating research and development partnerships and the creation of new products and services.
- **Monitoring and Evaluation:** The success of the innovation program is typically measured by: (a) the number of innovation ideas submitted, funded and implemented; (b) the flow-rate through the innovation pipeline; (c) Staff participation rates; and (d) revenue generated or productivity improvement realized as a consequence of innovation.

APPENDIX 5.11: Critical Path Analysis

Geospatial information is about enabling business intelligence, and therefore innovation and process improvement is often about working towards the end game – having the right information, in the right format, at the right time for decision-making.

Critical path innovation is one way to progressively work towards common end goals. The process is to critically analyze existing, and often complex, geospatial information management workflows and procedures. This analysis is important for uncovering inefficiencies and barriers. For example, the ability to share information often hinges on a few critical resources or constraints (human or computing) that may not be obvious at first glance. And yet, without these critical resources and the removal of barriers, it is not possible to make the data sharing task more efficient.

Critical path innovation is not necessarily about being radical. In many cases, small incremental process improvement steps can be taken early to streamline data access. When connected, these steps will lead to improved business intelligence capabilities down the track. Small innovative process improvements are often more effective than one large project, particularly given financial constraints and difficulties associated with overcoming cultural barriers to change.

The critical pathway analysis focusses on both administrative processes as well as observing the flow of data. Templates for undertaking the critical path analysis steps are provided below. The process involves identifying:

- all key tasks e.g. required to deliver a product, service or knowledge;
- the sequence of activities required to complete a task;
- the time it takes to perform each activity;
- dependencies between activities, particularly those that have potential to cause delays; and
- the handover points within and between organisations, and the time lapse between the handover and the start of the next activity.

From the critical path analysis, it is possible to determine the points:

- that may lead to a high-risk of failure or delays;
- where the process can be fast-tracked;
- where lag time between activities is lengthening turnaround time delays;
- where additional resources are needed; and
- where process improvements need to be applied to ensure the critical path can be sustained.

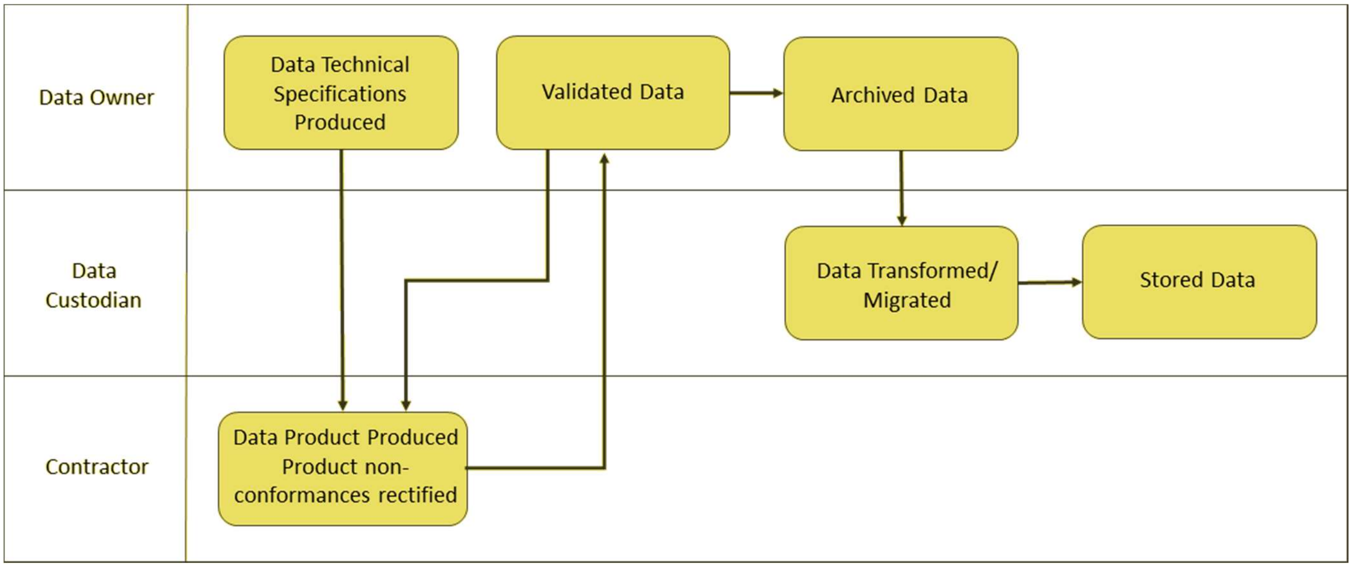
The critical path analysis is likely to raise several process-related issues that can be improved, particularly if the emphasis of the critical path is aimed at servicing the business intelligence and decision-making capabilities of government. This is because the critical pathway for information flows may not have been considered beyond organizational boundaries or between teams within the same organization. When embarking on the IGIF journey, there will be numerous processes identified as requiring improvement, and these will require multifaceted solutions, e.g. new technologies, polices and standards. This can be overwhelming. Priority should be given to mapping the critical path for geospatial information where existing workflows have:

- Known impediments as these workflows are already impacted along the critical path;
- Unsatisfactory turnaround times so that the lag time can be formally addressed;
- Duplicate data handling as the current situation is likely to be costly; and
- Potential bottlenecks that can be rectified using technology, standards and policies that can be scaled up as a whole-of-government approach, and thus have a significant positive impact.

Step 1: List of tasks currently undertaken

Task ID	Task Name	Task Description	Task Predecessors	Task Duration (hours)	Lapse time (Duration between Activities)
1	Specification	Data Technical Specifications Produced		65	
2	Send Specification to Contractor		1	1	
3	Draft Product	Data Product Produced	2		
4	Send Product to Data Owner		3		
5	Validate Product	Validate Product	4		
6	Send amendments to Contractor		5		
7	Revised Product	Amend Product	6		
8	Send Product to Data Owner		7		
9	Revalidate Product		8		
10	Archive Product		9		
11	Send Product to Data Custodian		10		
12	Transform Product		11		
13	Migrate Product		11		
13	Store Product		12		

Step 2: Document Workflow and responsibility handover points



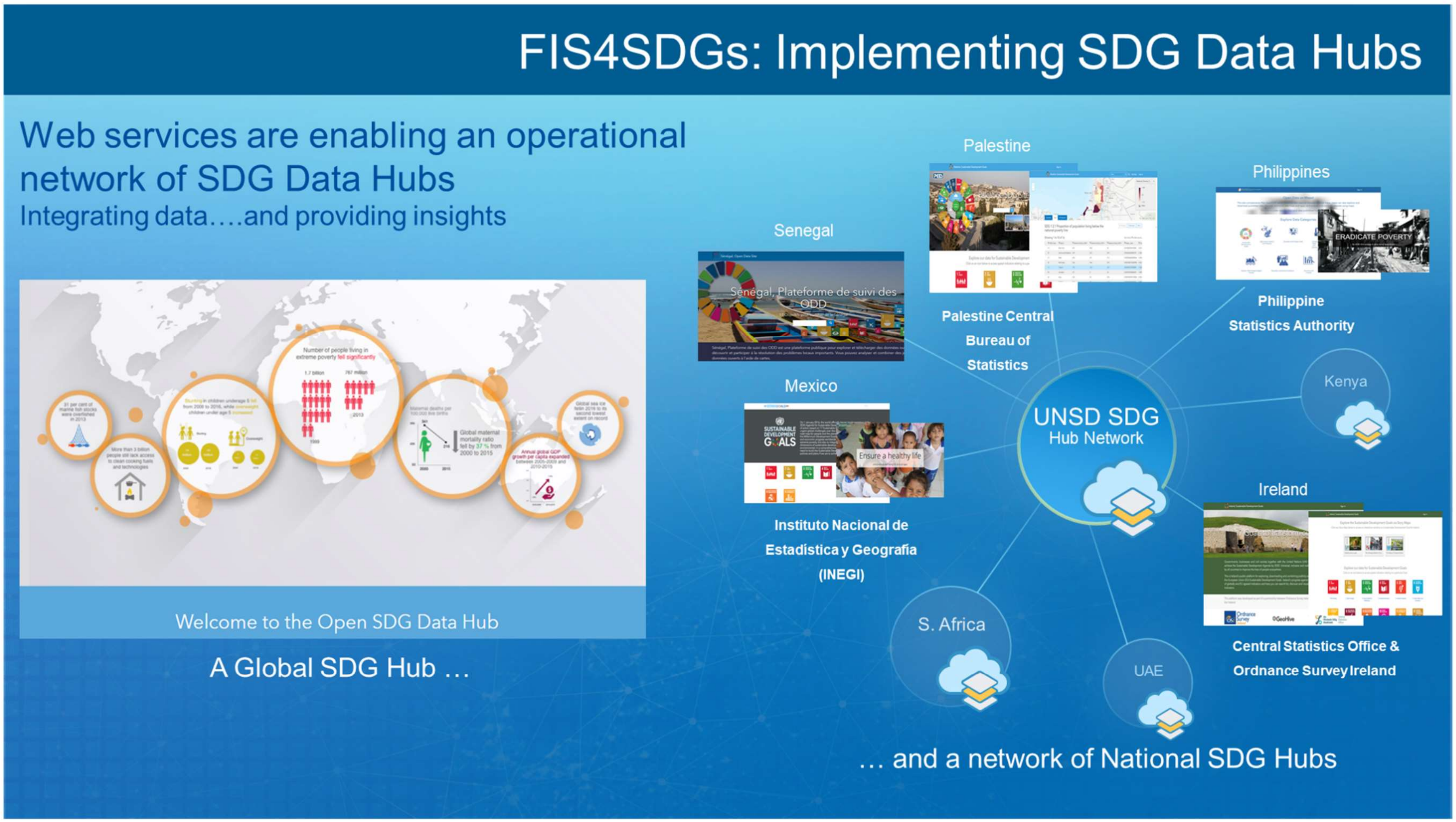
Step 3: Critical Path Analysis

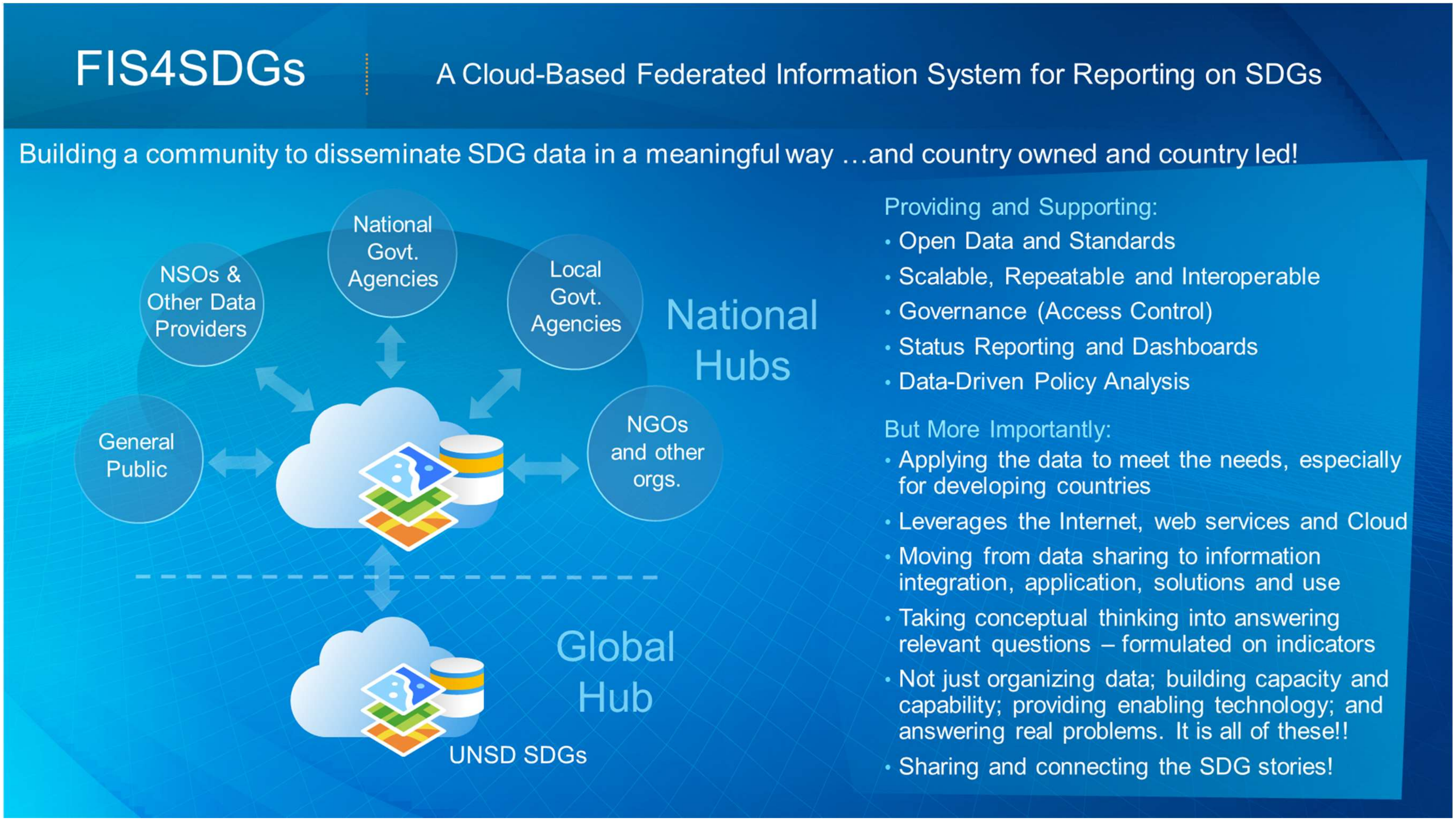
Some example questions that can be asked:

1. Highlight the critical path?
2. Determine the minimum and maximum path duration?
3. Identify what variables make a difference between min and max duration?
4. Are some task not critical?
5. Can lengthy tasks and handover points be automated?
6. Are their additional tasks that contractors can do?
7. If repeat errors are a factor, what can be done to resolve the issue?
8. What activities will delay the critical path?
9. Can the critical path be shortened in anyway?
10. Is lack of policy, technology, leadership and/or skills contributing to delays?

APPENDIX 5.12: Open SDG Data Hubs

To fully implement and monitor progress on the SDGs, decision-makers need data and statistics that are accurate, timely, sufficiently disaggregated, relevant, accessible and easy to use. The Open SDG Data Hub promotes the exploration, analysis, and use of authoritative SDG data sources for evidence-based decision-making and advocacy. Its goal is to enable data providers, managers and users to discover, understand, and communicate patterns and interrelationships in the wealth of SDG data and statistics that are now available.




FIS4SDGs
A Cloud-Based Federated Information System for Reporting on SDGs
Building a community to disseminate SDG data in a meaningful way …and country owned and country led!
NSOs & Other Data Providers
National Govt. Agencies
Local Govt. Agencies
National Hubs
General Public
NGOs and other orgs.
Global Hub
UNSD SDGs
Providing and Supporting:
• Open Data and Standards
• Scalable, Repeatable and Interoperable
• Governance (Access Control)
• Status Reporting and Dashboards
• Data-Driven Policy Analysis
But More Importantly:
• Applying the data to meet the needs, especially for developing countries
• Leverages the Internet, web services and Cloud
• Moving from data sharing to information integration, application, solutions and use
• Taking conceptual thinking into answering relevant questions – formulated on indicators
• Not just organizing data; building capacity and capability; providing enabling technology; and answering real problems. It is all of these!!
• Sharing and connecting the SDG stories!