**3.5** **Obtaining data from external sources**

**Reading material**

|  |
| --- |
| Now that you know more about data capture, we will dive into data conversion and transfer.  For many projects, it would be nearly impossible to gather all the necessary data on your own. That is where external data sources come in. Regardless of where the data comes from, GIS software can overlay all the information into a single, layered map. |
| Now let us start with looking at Spatial data conversion and Spatial data transfer Standards (SDTS) as data management processes for obtaining external data.   1. **Spatial data conversion**   One of the biggest problems with data obtained from external sources is that they can be encoded in many different formats. Data conversion is the process of translating data from one format to another. It is a critical process in the migration of information from existing information databases to new ones that often require changes in data formats. This step enables the data to be read, altered, and executed in an application or database other than that in which it was created, when the systems undergo replacement or updates. It is also of great importance in the insurance sector. Companies can make use of different strategies for converting data to ensure that the data is compatible with their systems. |
| GIS Data Conversion  Source : <https://www.igismap.com/gis-data-conversion/> |
| There are many ways to convert GIS data and the conversion may use special conversion programs. Simply, it may involve complex data exporting or importing procedures. Following are the most used spatial data format.   * The commonly used formats for raster data model are**: GeoTIFF (TIFF), ERDAS Imagine, Grid, ECW** * The commonly used formats for the vector data model are**: .shp ,( shapefile ), \*.gpkg(Geopackage) ,** SpatiaLite, **GeoJSON (a lightweight format based on JSON, used by many open source GIS packages), KML (Keyhole Markup Language a XML based), etc.** Vector data can also be stored in database formats as opposed to file-based formats. |
| **Right data conversion should ensure the following:**   * Data is converted into an appropriate format and is transferred correctly. * Data works in the new destination database. * Data retains its quality and data consistency is maintained at all times across all the systems.   As said above one of the biggest challenges with spatial data obtained from external sources is that they can be encoded in many different formats. For that, many tools have been developed to move data between systems and to reuse data through open application programming interfaces (APIs).  GIS Consortium has the capability to convert hard copy into a wide range of electronic formats. GISC’s team of experienced and multi – skilled specialists have a vast knowledge of converting various types of geospatial data. This helps to deal with the present-day challenges such as complexity of data, project timelines and effect on the quality and accessibility of the data, ensuring a smooth and successful data conversion. |
| **What data conversion is NOT**  Data conversion is often confused with processes known as ***data migration***, ***data transformation***, and ***data cleansing.*** Let’s take a look to clarify these processes.   * **Data migration:** Where data conversion translates individual computer objects and data types from one format to another, data migration transfers entire databases or programs from one location to another. Data migration often entails data conversion, data transformation, and/or data cleansing. * ***Data transformation:*** Data conversion translates one format to another. An example would be converting an .***shp file*** to ***raster***. Data transformation changes the data presentation. A common data transformation process is to condense the data. Note: the format itself does not change. * ***Data cleansing:*** Data cleansing finds and corrects inaccurate, repeated, and incomplete data. This procedure often occurs after a data conversion, data transformation, or data migration process. |
| 1. **Spatial data transfer Standards (SDTS)**   Data transfer is the exchange of data and is essential to foster sharing of geographic data. The ***Spatial Data Transfer Standard (SDTS)*** is a robust way of transferring earth-referenced spatial data between different computer systems with the potential of no information loss. It is a general mechanism for the transfer of geographically referenced spatial data and its supporting metadata (i.e., attributes, data quality reports, coordinate reference systems, security information).  Data transfer standards are operating rules that accomplish the transfer of data between incompatible systems and provide a common format for exchanging data from various software systems. Data transfer standards provide a road map for interpreting various geographic data models. Since each GIS contains its own unique, conceptual, logical, and physical data model, standards can provide a means for exchanging data between two GIS systems. Resolving characteristics of various spatial data models used by the different systems is a fundamental requirement for data transfer standards. |
| The full SDTS specification creates a framework for spatial data transfer by defining different “levels,” from the real world to the physical encoding of the data. The **conceptual level** describes a way to represent real world entities, including their geometric and topological characteristics and relationships. The **logical level** presents a data model for identifying and encoding information for an SDTS transfer. SDTS also defines the **physical level** with rules and specific formats for encoding data on a medium of choice (e.g., magnetic tape).    Source: “Design of a Spatial Data Transfer Processor.” Cartography and Geographic Information Systems, Vol. 19, by Altheide, Phyllis |
| As usual, at an abstract level, every GIS packages supports at least one of the two spatial data models - ***the raster model and the vector model.*** In the raster model, the information is provided in the form of images (or pixels). In the vector model, the information is provided in the form of topology comprising of points, lines, and polygons. A data format should adhere to any one of the two data models. |
| **Purpose of SDTS**  The purpose of the SDTS is to promote and facilitate the transfer of digital spatial data between dissimilar computer systems, while preserving information meaning and minimizing the need for information external to the transfer. Implementation of SDTS is of significant interest to users and producers of digital spatial data because of the potential for increased access to and sharing of spatial data, the reduction of information loss in data exchange, the elimination of the duplication of data acquisition, and the increase in the quality and integrity of spatial data. SDTS is neutral, modular, growth-oriented, extensible, and flexible--all characteristics of an "open systems" standard. |
| **Interoperability and Spatial Data Standards**  The International Organization for Standardization defines interoperability as “the capacity to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units” (ISO 1993). In the context of geographic information, the functional units are Geographic information (GI) systems or Web services. The communication between these units comprises transfer of spatial data as well as querying and execution of remote services.  In this respect, two types of interoperability can be distinguished: ***syntactic interoperability*** refers to format of the transferred data that is the compliance with spatial data standards; ***semantic interoperability*** builds on syntactic interoperability and refers to the accurate preservation and interpretation of the meaning of the transferred information. |
| Interoperability is a key requirement for the seamless exchange of spatial data between users and organizations employing different Geographic information systems and Web services. On a broader scale, the standards enabling interoperability are the cornerstones of Spatial Data Infrastructure. Without agreements on the formats of the transferred spatial data and the interfaces for accessing the corresponding Web services, mutual data exchange between different geographic information systems would at least require manual transformation of the data or even be impossible.  Source: Encyclopedia of Geography by Wright, John Kirtland (1891–1969), <http://dx.doi.org/10.4135/9781412939591.n1264> , ©2010 SAGE Publications |
| **Geospatial information**, a variant of spatial information, is generally collected on thematic basis, where individual organizations are involved on any particular theme. Geospatial thematic data is being collected from decades and huge amount of data is available in different organizations. Information communities find it difficult to locate and retrieve required geospatial information from other geospatial sources in reliable and acceptable form.  The problem that has been incurred is the lack of standards in geospatial data formats and storage/access mechanism. Heterogeneity in geospatial data formats and access methods poses a major challenge for geospatial information sharing among a larger user community. |
| With the growing need of geospatial information and widespread use of Internet has fostered the requirement of geospatial information sharing over the Web. The Geo-Web is one of the reliable distributed networks of interconnected geographic information sources and processing services that are:   * Globally accessible, that is, they live on the internet and are accessed through standard Open Geospatial Consortium (OGC) and W3C interfaces * Globally integrated data sources that make use of standard data representation for sharing and transporting geospatial data.   Without successful interoperability approaches, the realization of Geo-Web is not possible. |
| Geospatial Standards, Interoperability, Metadata Semantics and Spatial Data  Infrastructure | Semantic Scholar  Source: Geospatial Standards, Interoperability, Metadata Semantics and Spatial Data Infrastructure by R. Longhorn (2005) |
| **Traits of successful interoperability:**   * Data producers must ensure that their data are readily accessible and understandable to potential data consumers. * Users must be able to identify and locate relevant information and know whether a given set of data is germane to their work. * Queries to dispersed sites must be formulated and processed in a manner meaningful to both data server and application client. * Geodata from one source must be capable of being integrated with data from another, in terms of both structure and semantics. * Display and analytical functions must be associated with particular data models and made available to the requester.   geospatial interoperability | GIS&T Body of Knowledge  Source: University of consortium of Geographic Information science |
| Source: Muvunyi Germain, INES Ruhengeri, Rwanda (Lecture note) |

**Exercise materials and tasks**

**Quiz questions**

As a recap of this session’s contents, please answer the following questions to test your understanding:

1. The biggest problems with data obtained from external sources is that they are in many different formats. Data transformation is the process of translating data from one format to another.
   1. True
   2. **False**
2. Link these two types of interoperability with their definition by matching the following terms with the correct definition:

syntactic interoperability, semantic interoperability

|  |  |
| --- | --- |
| refers to the accurate preservation and interpretation of the meaning of the transferred information |  |
| refers to format of the transferred data that is the compliance with spatial data standards |  |

*Answer:*

|  |  |
| --- | --- |
| refers to the accurate preservation and interpretation of the meaning of the transferred information | semantic interoperability |
| refers to format of the transferred data that is the compliance with spatial data standards | **syntactic interoperability** |

1. Pick the right traits of successful interoperability
   1. **Data producer must ensure that the data is accessible**
   2. Geodata cannot be integrated with other's data
   3. **Users must be able to locate relevant information**
   4. Data does not have to be understandable