2D TALENT TRACKS

AVINEON®

DATA INTEROPERABILITY

BEYOND THE FIRE

For those who want to do more

TABLE OF CONTENTS

SNO	CHAPTER	DURATION IN HRS
01	TRANSFORMATION OF DATA FROM ONE SOFTWARE TO ANOTHER SOFTWARE	42HRS
02	COORDINATE SYSTEMS AND REPROJECTION OF THE DATA FROM ONE PROJECTION SYSTEM TO ANOTHER PROJECTION SYSTEM	48HRS
03	IMPROVING DATA LOCATION ACCURACY OR SPATIAL CORRECTION OF DATA FROM ONE LOCATION TO ANOTHER LOCATION	8HRS
04	TRANSFORMATION OF VECTOR DATA TO RASTER DATA OR VICE VERSA	16HRS

Week – 1 to 10

TRANSFORMATION OF DATA FROM ONE SOFTWARE TO ANOTHER SOFTWARE

Week -1 4Hrs

In GIS, data interoperability refers to the ability of different geographic information systems (GIS) or spatial data systems to exchange, integrate, and use geospatial data seamlessly. The goal is to ensure that data can be shared, accessed, and used across different GIS platforms, software applications, and data formats without losing its integrity or meaning.

Data interoperability encompasses a wide range of concepts and technologies, including:

- **Data standards:** Standardized formats and definitions for data elements ensure that data can be interpreted and used consistently across different systems.
- **Data transformation:** Techniques for converting data from one format to another, enabling seamless data exchange and integration.
- **Metadata:** Descriptive information about data, providing context and meaning to facilitate discovery, understanding, and utilization.
- **Data quality:** Ensuring data is accurate, complete, consistent, and relevant to its intended use.
- **Data governance:** Establishing policies and procedures for managing data effectively, including data sharing, access control, and security.

A) Transformation of data from one software to another software

Data standards:

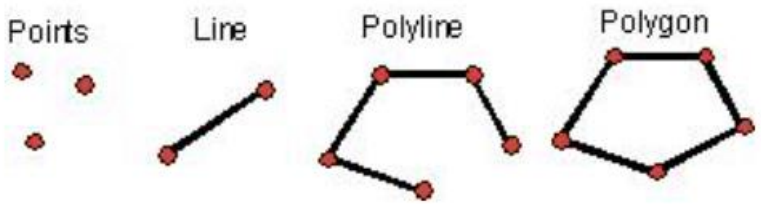
- **What are data standards?**

  Standardized formats and definitions for data elements ensure that data can be interpreted and used consistently across different systems.

  The three types of GIS Data are -spatial, –attribute, & —metadata

  Spatial:

  - Point Data — layers containing by points (or "events") described by x, y (latitude, longitude, easting, northing)



  - Line/Polyline Data — layers that are described by x, y points (nodes, events) and lines (arcs) between points (line segments and polylines).
  - Raster or grid data (matrices of numbers describing e.g., elevation, population, herbicide use, etc.
  - Images or pictures such as remote sensing data or scans of maps or other photos. This is special "grid" where the number in each cell describes what color to paint or the spectral character of the image in that cell. (To be used, the "picture" must be placed on a coordinate system, or "rectified" or "georeferenced")
  - TINs – Triangular Irregular Networks – used to discretize continuous data.

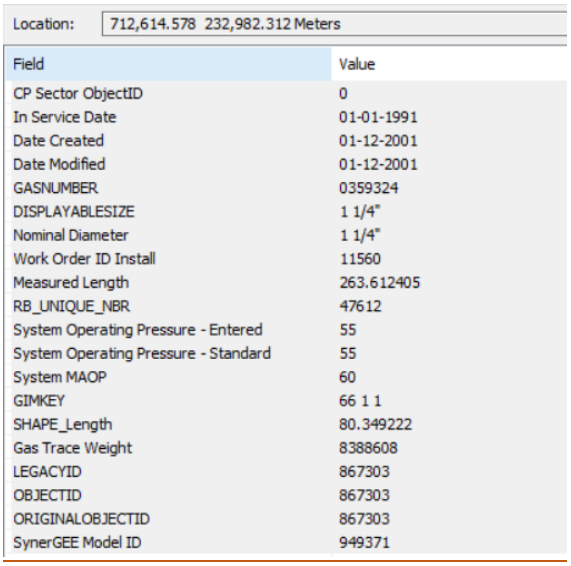
  Attribute:

  Attribute data are non-spatial characteristics that are connected by tables to points, lines, "events" on lines, and polygons (and in some cases GRID cells)

  1) A point, vector or raster geologic map might describe a "rock unit" on a map with a single number, letter or name, but the associated attribute table might have

  Age, lithology, percent quartz, etc., for each rock type on the map.

2) most GIS programs can either plot the polygon by the identifier or by one of the attributes.



Location: 712,614.578 232,982.312 Meters

Field	Value
CP Sector ObjectID	0
In Service Date	01-01-1991
Date Created	01-12-2001
Date Modified	01-12-2001
GASNUMBER	0359324
DISPLAYABLESIZE	1 1/4"
Nominal Diameter	1 1/4"
Work Order ID Install	11560
Measured Length	263.612405
RB_UNIQUE_NBR	47612
System Operating Pressure - Entered	55
System Operating Pressure - Standard	55
System MAOP	60
GIMKEY	66 1 1
SHAPE_Length	80.349222
Gas Trace Weight	8388608
LEGACYID	867303
OBJECTID	867303
ORIGINALOBJECTID	867303
SynerGEE Model ID	949371

Metadata:

metadata are the most forgotten data type and necessary if you are going to use data, or if someone is going to use your data later (or your derivative information)

contains information about

scale

accuracy

projection/datum

data source

manipulations

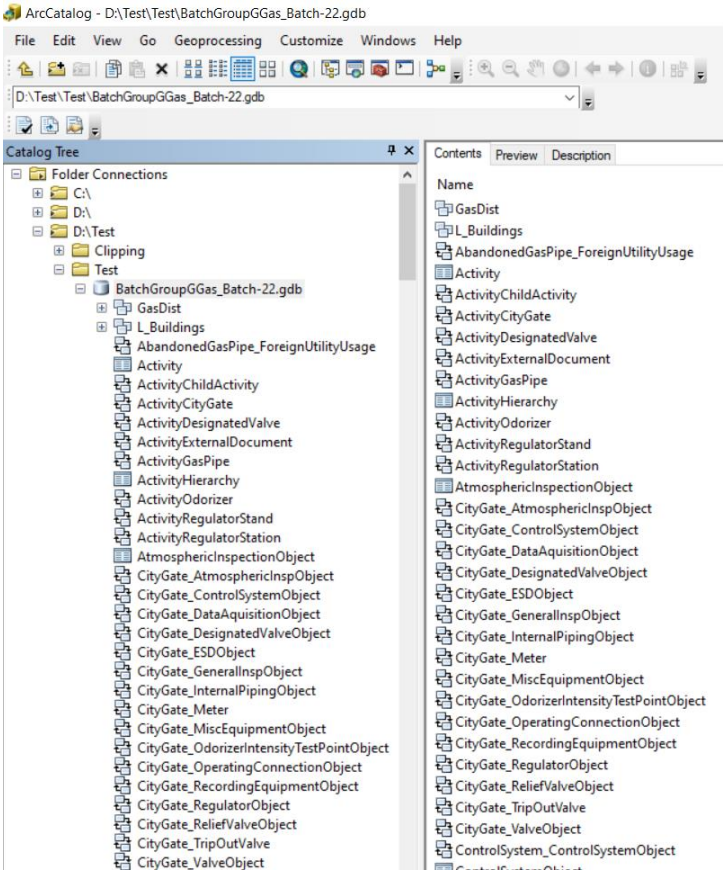
how to acquire data

many different “standards” for collection and presentation of metadata, such as FGDC used by US gov’t agencies.

What is data catalog?

A data catalog is an organized list of all the data assets which empower data teams throughout the company.

A Data Catalog is a collection of metadata, combined with data management and search tools, which helps analysts and other data users to find the data that they need, serves as an inventory of available data, and provides information to evaluate fitness of data for intended uses.



In case if you need further information, please browse in google for relavent information.

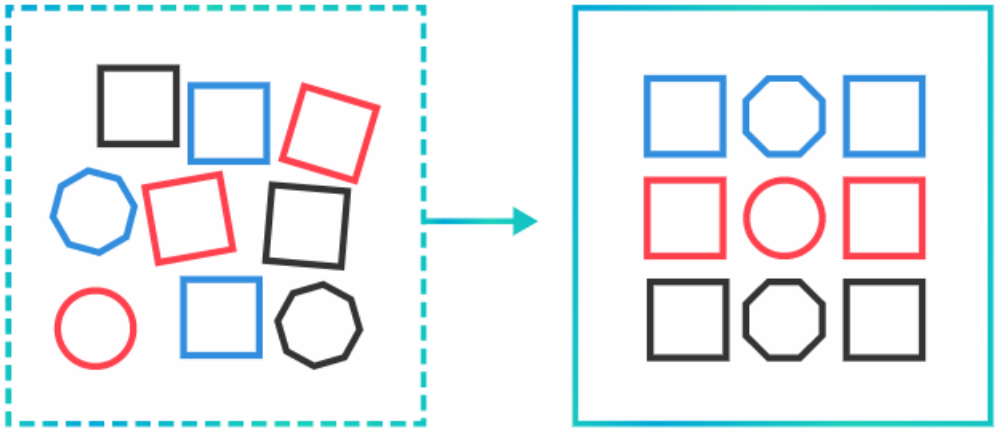
Week -2 4Hrs

Hands on experience on week1 content

Week -3 4Hrs

Data transformation:

It is the process of converting, cleansing, and structuring data into a usable format that can be analyzed to support decision making processes, and to propel the growth of an organization.



Data transformation is used when data needs to be converted to match that of the destination system. This can occur at two places of the data pipeline. First, organizations with on-site data storage use an extract, transform, load, with the data transformation taking place during the middle 'transform' step.

Organizations today mostly use cloud-based data warehouses because they can scale their computing and storage resources in seconds. Cloud based organizations, with this huge scalability available, can skip the ETL process. Instead, they use a transformation process that converts the data as the raw data is uploaded, a process called extract, load, and transform. The process of data transformation can be handled manually, automated or a combination of both.

Types (4) of data transformation:

Transformation is an essential step in many processes, such as data integration, migration, warehousing, and disagreeing. The process of data transformation can be:

- **Constructive**, where data is added, copied or replicated.
- **Destructive**, where records and fields are deleted.
- **Aesthetic**, where certain values are standardized, or
- **Structural**, which includes columns being renamed, moved and combined.

On a basic level, the data transformation process converts raw data into a usable format by removing duplicates, converting data types and enriching the dataset.

This data transformation process involves defining the structure, mapping the data, extracting the data from the source system, performing the transformations, and then storing the transformed data in the appropriate dataset. Data then becomes accessible, secure and more usable, allowing for use in a multitude of ways. Organizations perform data transformation to ensure the compatibility of data with other types while combining it with other information or migrating it into a dataset. Through data transformations, organizations can gain valuable insights into the operational and informational functions.

Given the massive amounts of data from disparate sources that businesses must deal with daily, data transformation has become an essential tool. It facilitates the conversion of data, irrespective of its format, to be integrated, stored, analyzed and mined for **business intelligence**.

How is Data Transformation Used?

Data transformation works on the simple objective of extracting data from a source, converting it into a usable format and then delivering the converted data to the destination system. The extraction phase involves data being pulled into a central repository from different sources or locations, therefore it is usually in its raw original form which is not usable. To ensure the usability of the extracted data it must be transformed into the desired format by taking it through a number of steps. In certain cases, the data also needs to be cleaned before the transformation takes place. This step resolves the issues of missing values and inconsistencies that exist in the dataset.

The data transformation process is carried out in five stages.

1. Discovery

The first step is to identify and understand data in its original source format with the help of data profiling tools. Finding all the sources and data types that need to be transformed. This step helps in understanding how the data needs to be transformed to fit into the desired format.

2. Mapping

The transformation is planned during the data mapping phase. This includes determining the current structure, and the consequent transformation that is required, then mapping the data to understand at a basic level, the way individual fields would be modified, joined or aggregated.

3. Code Generation

The code, which is required to run the transformation process, is created in this step using a data transformation platform or tool.

4. Execution

The data is finally converted into the selected format with the help of the code. The data is extracted from the source(s), which can vary from structured to

streaming, telemetry to log files. Next, transformations are carried out on data, such as aggregation, format conversion or merging, as planned in the mapping stage. The transformed data is then sent to the destination system which could be a dataset or a data warehouse.

Some of the transformation types, depending on the data involved, include:

Filtering which helps in selecting certain columns that require transformation.
Enriching which fills out the basic gaps in the data set.
Splitting where a single column is split into multiple or vice versa.
Removal of duplicate data, and Joining data from different sources

5. Review

The transformed data is evaluated to ensure the conversion has had the desired results in terms of the format of the data.

It must also be noted that not all data will need transformation, at times it can be used as is.

In case if you need further information, please browse in google for relevant information.

Week -4 4Hrs

Hands on experience on week3 content