

Week 2: Wednesday

Data Literacy

Data Definitions, Catalogs, and Glossaries

Data Catalog

- What is a data catalog?
- Why is it important?
- What are the key components?

Data Catalog

- A technical solution that helps organizations manage their data assets
 - Where is specific kinds of data stored?
 - Who is responsible for it?
 - How is it used?
 - What is the storage types for the data?
 - Data models.

What is a Data Model

Name	Date	location
string	string(YYYY-MM-DD)	point(longitude,latitude)

Data Model Continued

- GeoJSON Standard:

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "geometry": {
        "type": "Point",
        "coordinates": [125.6, 10.1]
      },
      "properties": {
        "name": "Dinagat Islands"
      }
    }
  ]
}
```

We are not focused too much on the technical details

- They shape how data is executed and deployed.
- We are focused on the data definitions and how they shape the data we use.

Data Glossary

- What separates a data catalog from a data glossary?
- What is a data glossary?
- Why is it critical?

Data Glossary

- A data glossary is a collection of terms and definitions that are used in a particular domain of knowledge or data.
- It is frequently included in a data catalog.
- It is a way to standardize the language used in a particular domain.

Data Glossary Example

- **Data Definition:** A data definition is a detailed description of the data that is stored in a data catalog.
- **Data Model:** A data model is a visual representation of the data that is stored in a data catalog.

Data Definitions

- Why are they important?
- How do they shape the data we use?

Data Definitions, examples

- What is a student? — Let's say this is for a university database system.
- What is middle class? — for an economic study.
- What is a diplomat? — for a political science study.

Data definitions

- How do these definitions shape our data?
- How do these definitions shape how our users understand the data?

Group Work

- Break into your domain groups.
- Discuss the three terms in your domain that are significantly ambiguous and need a definition.
- What are the different options for defining these terms?
- What are the implications behind those options?
 - How would the different definitions impact data collection/analysis?