Summer 7-10-2024

FAIRification of Geospatial Data

Attiksh Ansool Panda
Case Western Reserve University, aap184@case.edu

Olatunde D. Akanbi
Case Western Reserve University, oda10@case.edu

Vibha S. Mandayam
Case Western Reserve University, vsm21@case.edu

Erika I. Barcelos
Case Western Reserve University, eib14@case.edu

Roger H. French
Case Western Reserve University, rxf131@case.edu

FAIRification of Geospatial Data

Attiksh Ansool Panda[1,3,4], Olatunde Akanbi[1,2,4], Vibha Mandayam[1,3,4], Erika I. Barcelos[1,2,4], Roger H. French[1,2,3,4]

[1]SDLE, Department of Material Science and Engineering, Case Western Reserve University, Cleveland OH, USA
[2]Department of Material Science and Engineering, Case Western Reserve University, Cleveland OH, USA
[3]Department of Computer and Data Science, Case Western Reserve University, Cleveland OH, USA
[4]Center for Advancing Sustainable and Distributed Fertilizer Production (CASFER)

CASFER
An NSF Engineering Research Center

Introduction

Objectives:

- Extract, transform and load chemical and physical metadata for the state of Ohio from 2016-2023 from National Water Quality Monitoring Council which holds data from United States Geological Survey (USGS) and Environment Protection Agency (EPA) into the HPC setup at Case Western Reserve University.
- Create schemas, and ontologies to FAIRify the given dataset and get a subsequent .json and knowledge graph.
- FAIR stands for Findable, Accessible, Interoperable and Reusable.
- FAIR data is data that can be easily shared and explained between different groups. Since, the data is stored as .json and knowledge graphs, it is easy to find and access, the variables are standardized and helps for efficient analysis and modelling.

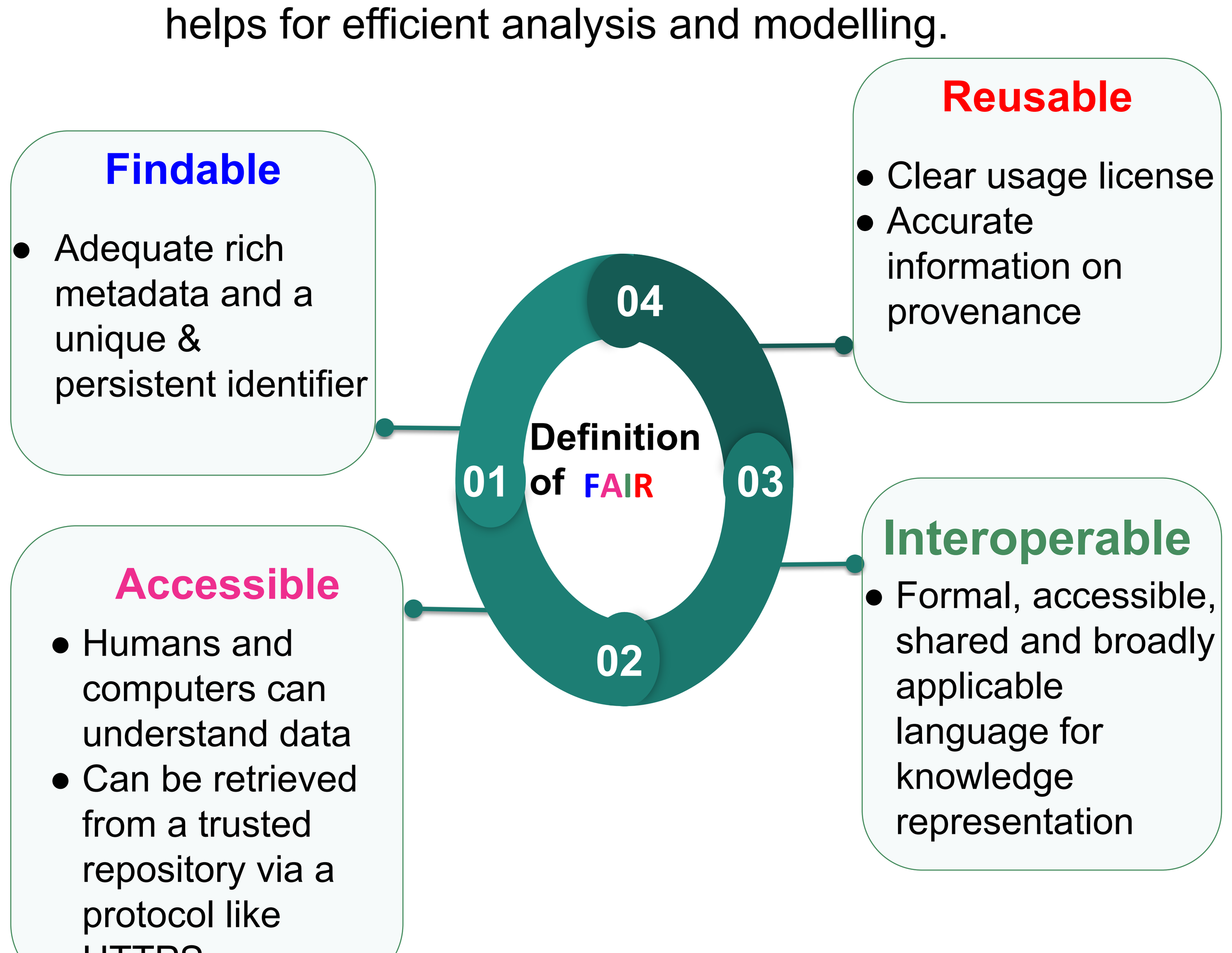

Fig. 1. FAIR Process

USGS Stream Water Ontology

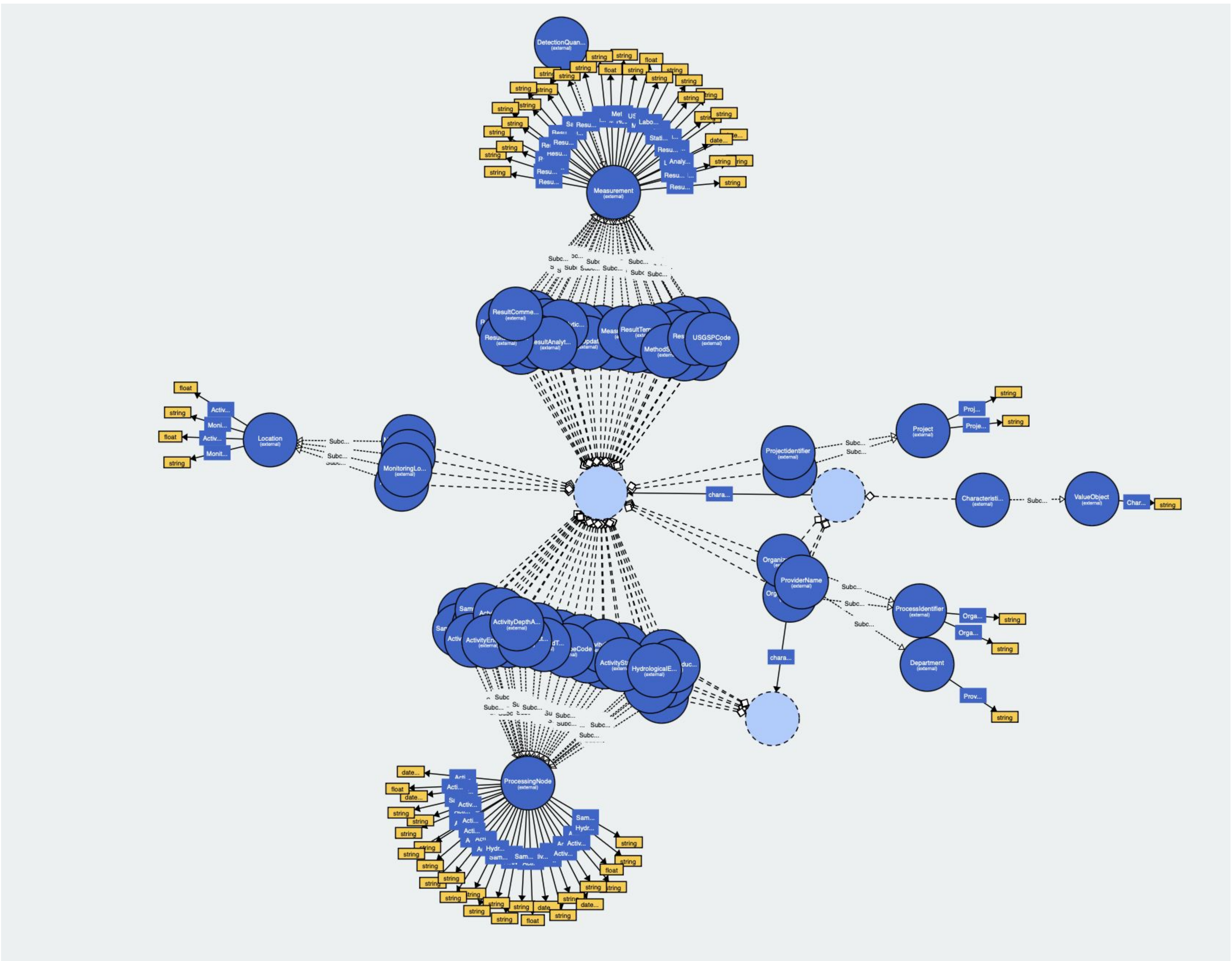

Fig. 3. Visualization of Ontology developed from .json

This is a knowledge graph that has been developed from the .json file obtained from running the FAIRmaterials R package. The details about FAIR, FAIRmaterials package documentation (both R and Python), directories and vignette for the package, and ontology visualization tools can be found at FAIR Materials FindTheDocs.

Methodology

- The data was extracted from National Water Quality Monitoring Council website and cleaned to remove data points that were either null or not required for our analysis.
- The dataset primarily divided data based on the provider of the data - either from the NWIS (National Water Information System) or STORET (STOrage and RETrieval) Data Warehouse.
- The data (from USGS and EPA) is from different water stations containing details about different chemical and physical features ('water temperature', 'nitrate + nitrite', 'ammonia'). The update period varies - some are static, some are updated yearly while some are updated every 15 minutes.
- A schema was created on draw.io from the metadata to map datasets variables to existing or newly-created ontology terms, which are based off of PMDco (Platform MaterialDigital core ontology).
- For my schema, I used the following structure: CharacteristicName → ProviderName → Organization → Activity. Activity → Location. Activity → Project. Activity → Measurements. All of the different subheadings of the dataset were sorted under one of these structure heads for my schema and they were paired with an ontology term from the Platform MaterialDigital core ontology (PMDco).
- The ontology terms from the schema were filled out in the OfficialPackageFAIRSheetInput sheet.
- This was extracted as CSVs which were then run through the FAIRMaterials package on R Studio (also available on Python) to get 3 files - knowledge graph, .json and .ttl sheet which can be used to generate graphs.

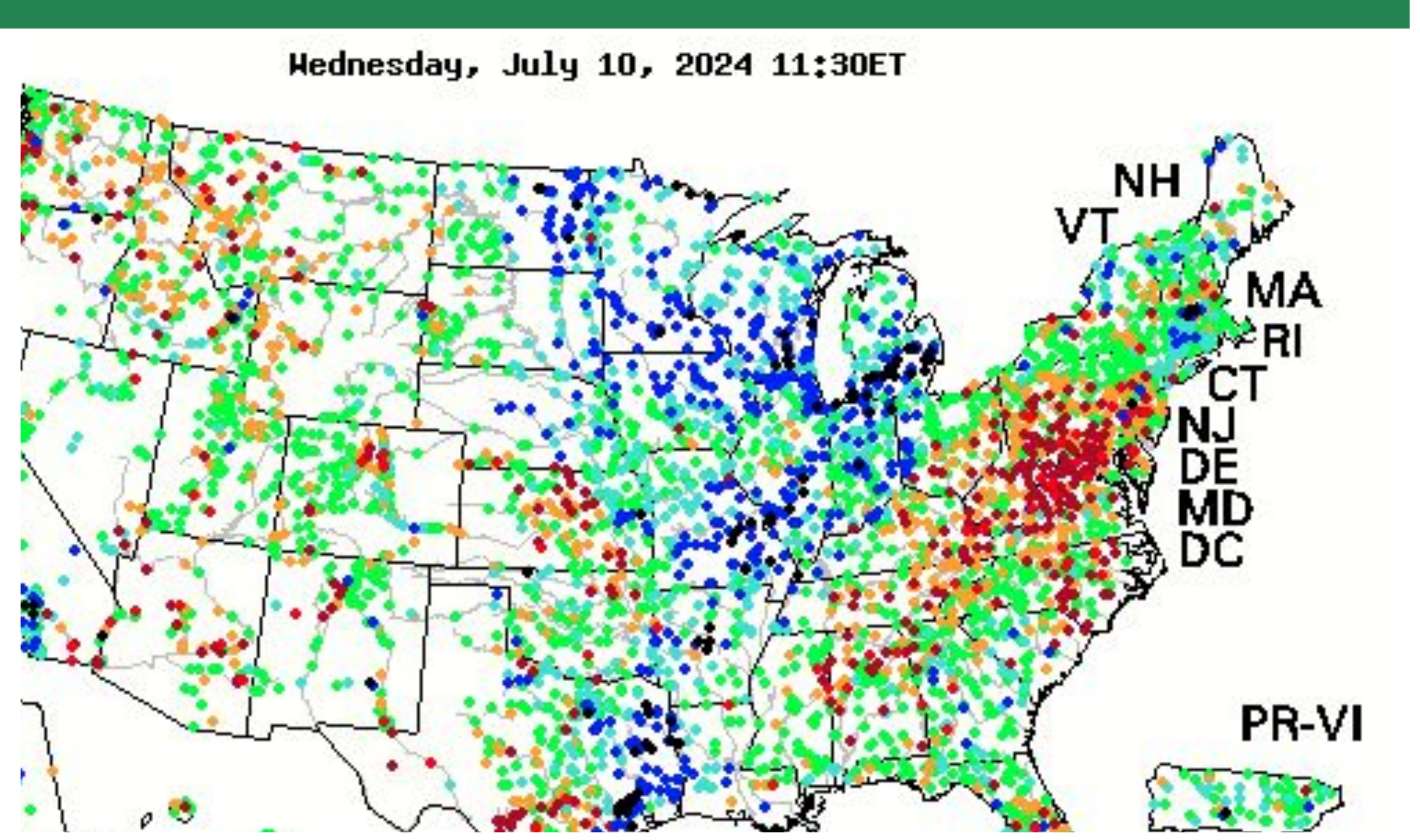

Fig. 2. Map of USA's water stations

Conclusion

Present method of data handling lead to lost metadata, a difficult integration process because of variable terms defined by the original user which are not known to everyone and the need for multiple sets of scripts to access and standarize different interlinked datasets. Developing ontologies and FAIR principles help us tackle these problems by making us use schemas on metadata to map variables to existing or newly-created ontologies which are then put through a script to develop .json files which help us visualize our ontologies into knowledge graphs. This helps us convert unstructured data into structured graphs. the ontological framework developed is helpful for understanding the datasets.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 2133576 This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at CaseWestern Reserve University.

References

1. Draw.io, "Flowchart Maker & Online Diagram Software," *app.diagrams.net*, 2024. https://app.diagrams.net/
2. "FAIRMaterials FindTheDocs," https://cwrusdle.bitbucket.io/
3. Jonathan E. Gordon, Alexander Harding Bradley, et al, "FAIRmaterials 0.4.2.4," PyPI. https://pypi.org/project/FAIRmaterials/
4. Hernandez, Kristen & Barcelos, Erika et al (2024). A data integration framework of additive manufacturing based on FAIR principles. MRS Advances. 10.1557/s43580-024-00874-5.

Thrust Interactions