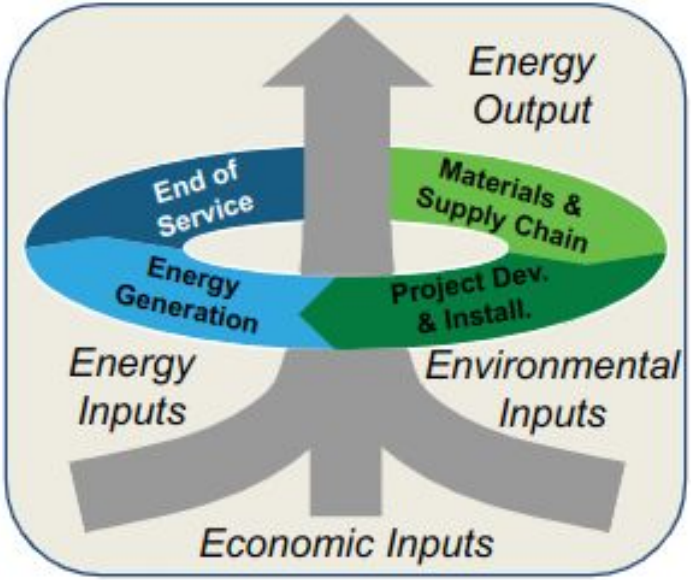


Using FAIRification and PV Ontologies to enable 2nd Life PV via provenance of modules, their energy production and O&M history

Roger H. French, Kyocera Professor, Case Western Reserve University

IEA PVPS Task 13 Webinar: Enabling 2nd Life Photovoltaics April 18th, 2024

Factors necessary for a Robust PV 2nd Life Market

How to Value 2nd Life PV Modules

- Need to know Module's Provenance
  - Where has it been, for how long?
  - How old is it?

PV Power Plants Sell Electricity

- Therefore there is timeseries Power Production Data

Managing PV Power Plant Assets

- Bidding on 1 hour, next day markets for electricity sales
- Analysing PV Performance Loss Rate

New Data Science, Deep Learning and AI technologies

- Are rapidly changing how PV assets are studied
- Data-enabled analyses can establish 2nd Life PV Market

AI/ML Opportunities: Accelerating Time to Science

To develop AI/ML for Science, such as Photovoltaic Science

We have High Performance Computing (HPC)

- "Scaled Up" Computing: Works for Physics Simulation Modeling
  - But doesn't handle massive datasets

Yet Big Tech uses Distributed Computing (DC)

- "Scaled Out" Computing: e.g. used by Google, Meta, etc.

AI/ML for Science needs D/HPC Computing

- Needs the integration of "Scaled Out & Scaled Up" Computing
- CRADLE^{tm}: Common Research Analytics & Data Lifecycle Environment[1]
  - Automated pipelines, FAIRification[2], Efficient Insights

Data Centric AI[3] presents humans with a grand opportunity

- "*Computational Inflection Point for Scientific Discovery*"[4]
  - Augmenting human reasoning; Working alongside human researchers
  - Scientific investigations restructured around the "salient human tasks"
    - With computers handling the routine and onerous tasks
    - Supplementing our human capabilities

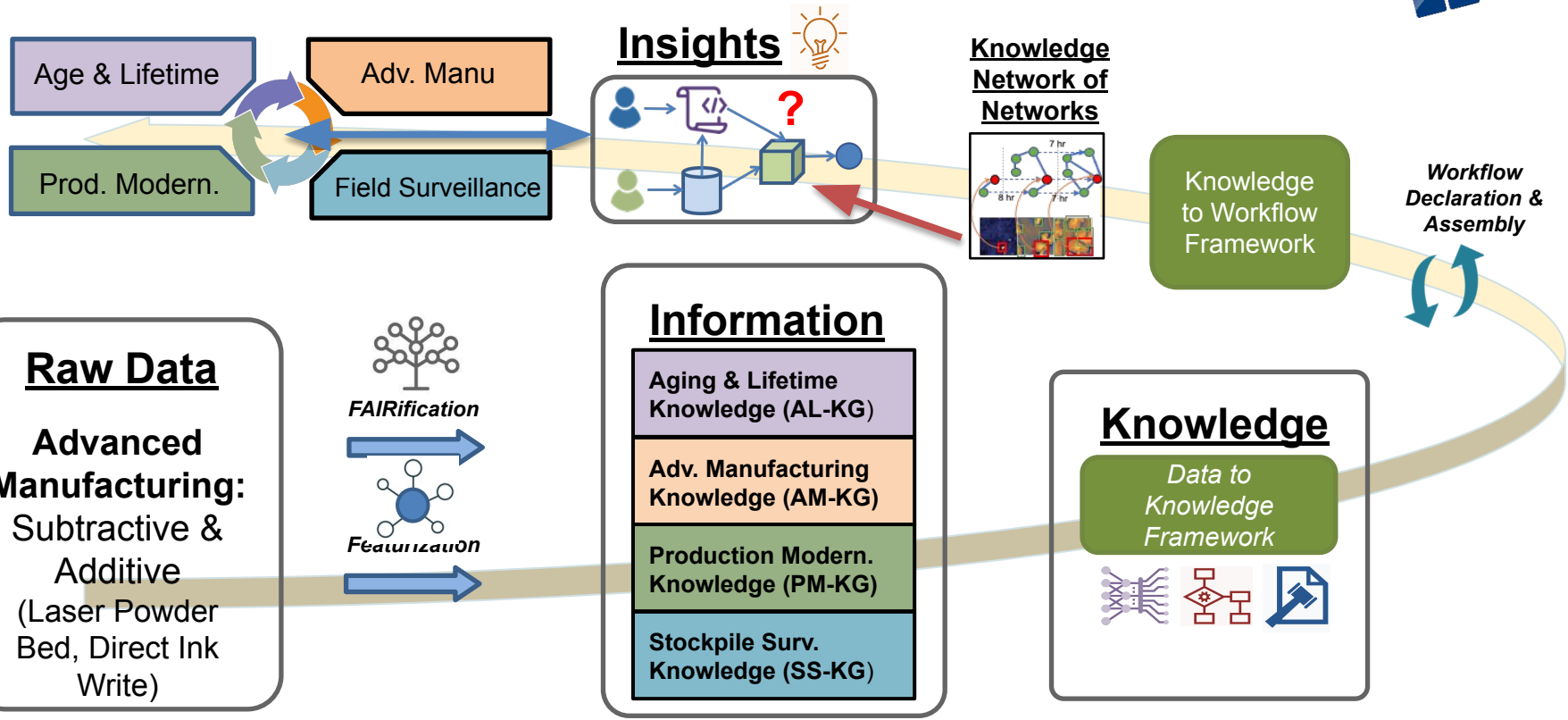
While decreasing reductionist approaches in scientific research

In SDLE Res. Cntr.

- Dist. Compute
  - 2.5 Pb Cluster
  - 7 TB Ram
  - 1164 CPU Cores
  - 30 GPUs
    - 480 GPU VRAM
    - 384k Cuda Cores
    - 1.2k Tensor Cores
- High Perf. Compute
  - 7152 CPU Cores
- Nvidia AISC 8 DGX
  - 2.5 Tb VRAM
  - 4 Tb RAM
  - 15 Tb nvme storage

PVPS

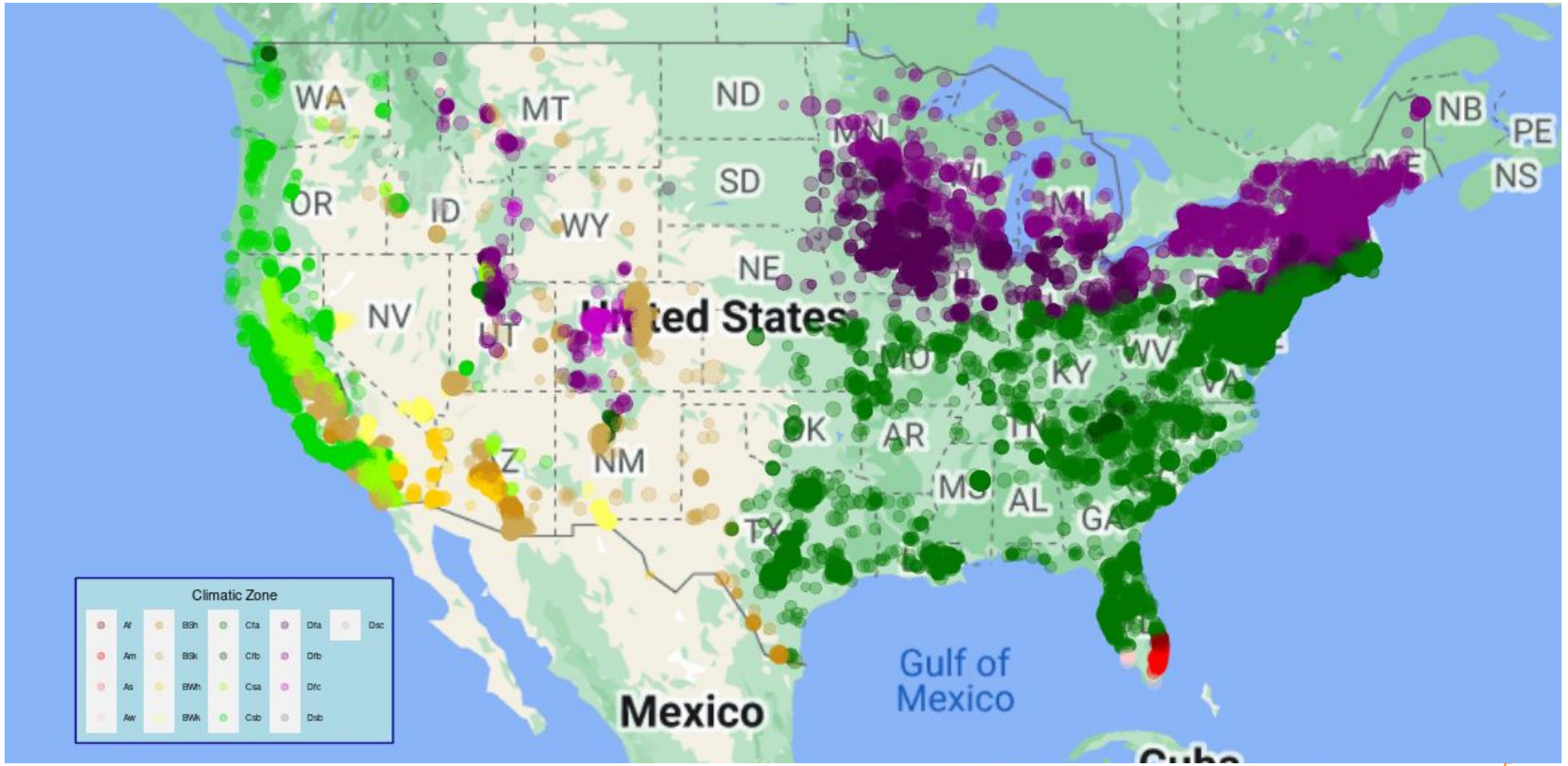
AI/ML Framework: Data to Knowledge, Knowledge to Workflow



PVPS

[1] Arafath Nihar *et al.*, "Accelerating Time to Science using CRADLE: A Framework for Materials Data Science," presented at the 30th IEEE International Conference On High Performance Computing, Data, & Analytics, Goa, India: IEEE, Dec. 2023. doi: 10.1109/HiPC58850.2023.00041.

Large Scale Photovoltaic Fleet Monitoring: 104,700 PV Systems



PVPS

104,700 PV Systems Ingested: EDA

Exploratory Data Analysis (EDA)

EDA of SDLE PV Systems

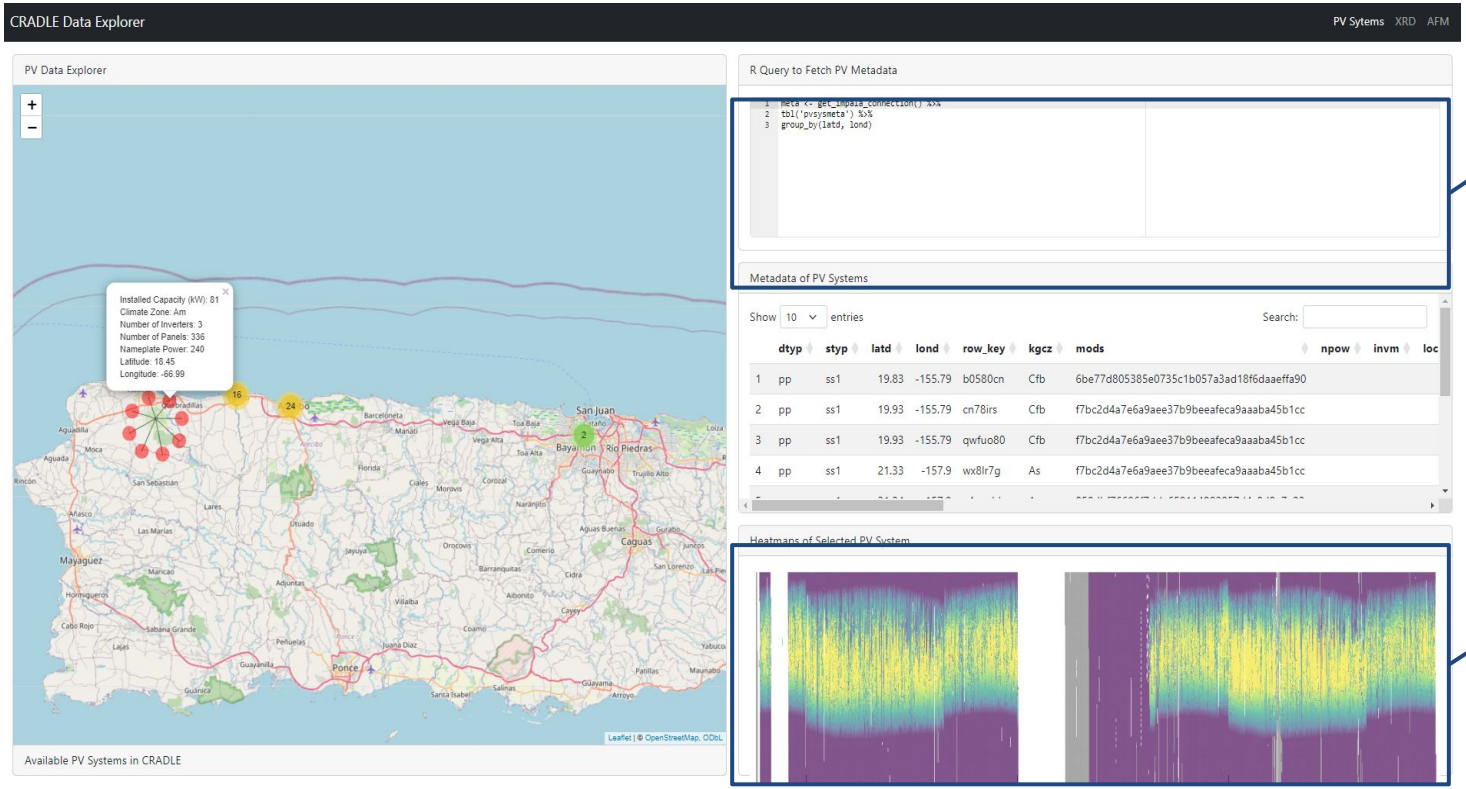
Massive Data

Integrated with Metadata

CRADLE
Containerized Data Science Tools

PVPS

Photovoltaic Systems: CRADLE Data Explorer



Query to extract selected metadata from HDFS/ HBase into a usable R dataframe

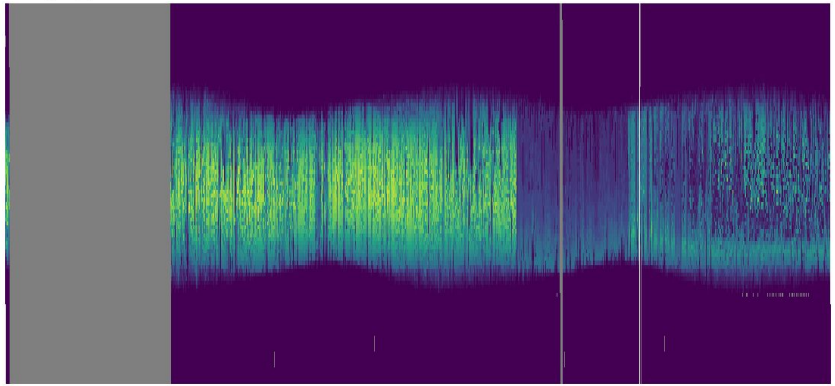
Interactive and quick assessment of data quality with missingness visualization

PVPS

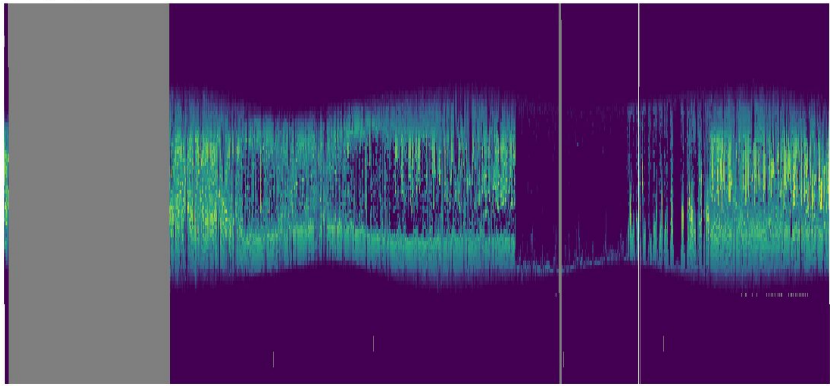
Data Reconstruction: Block Outages & Anomalous Measurements

RAW

s2025_inv1_17.3

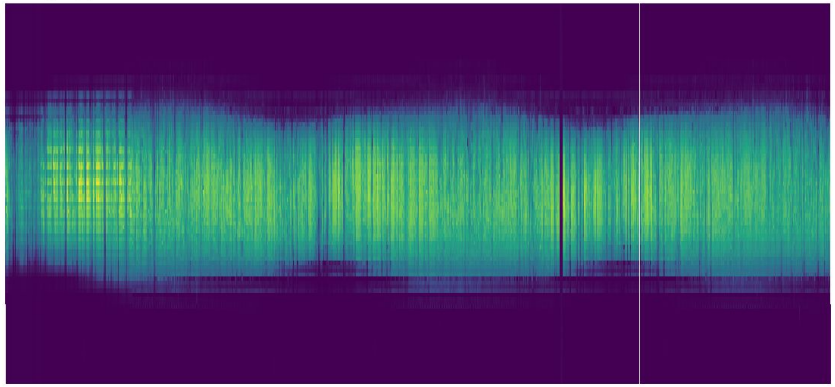


s2025_inv2_18

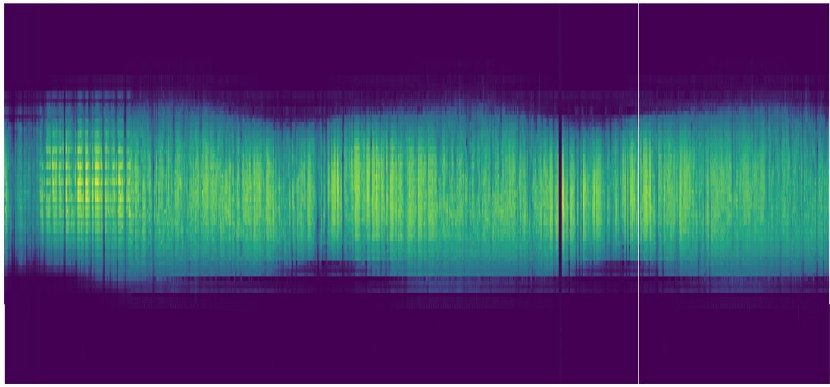


Reconstruction

s2025_inv1_17.3

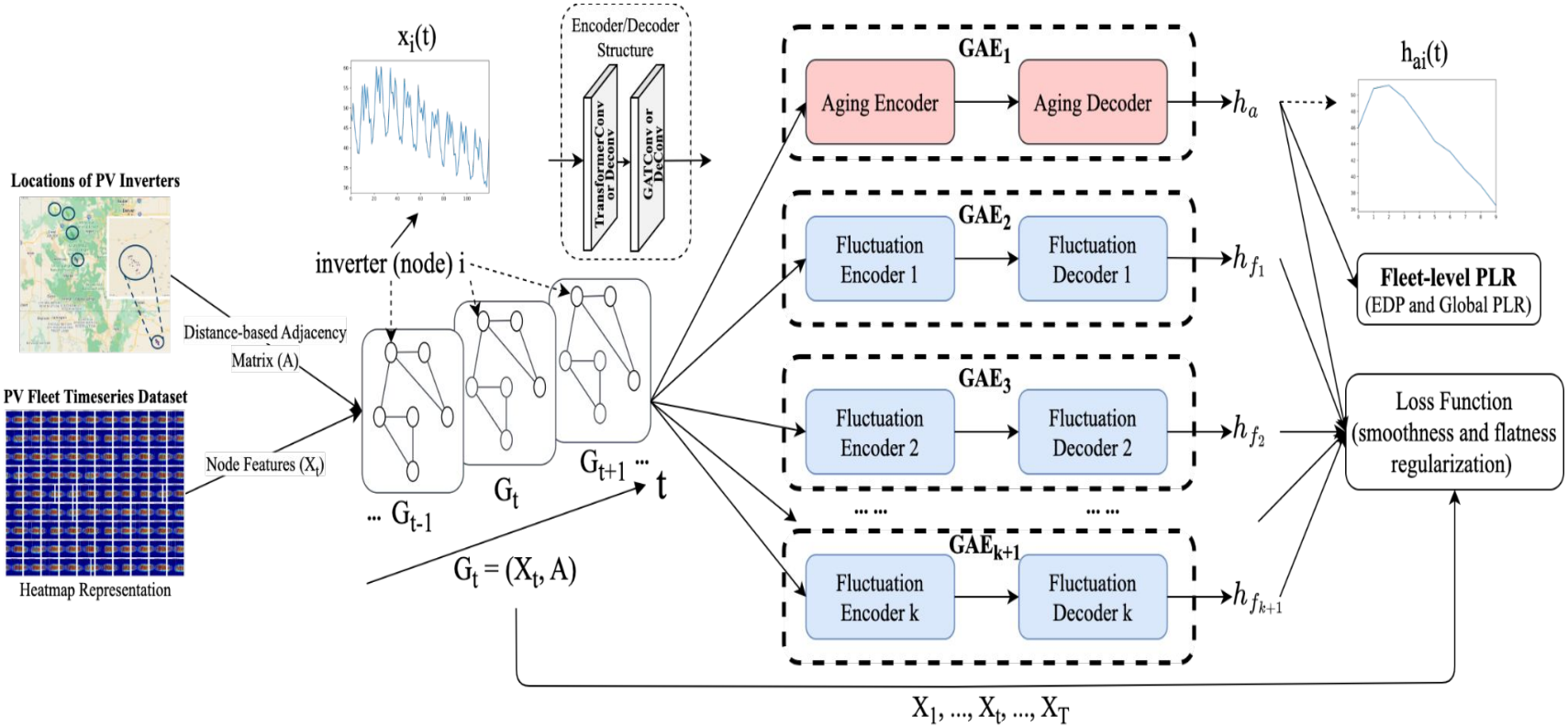


s2025_inv2_18



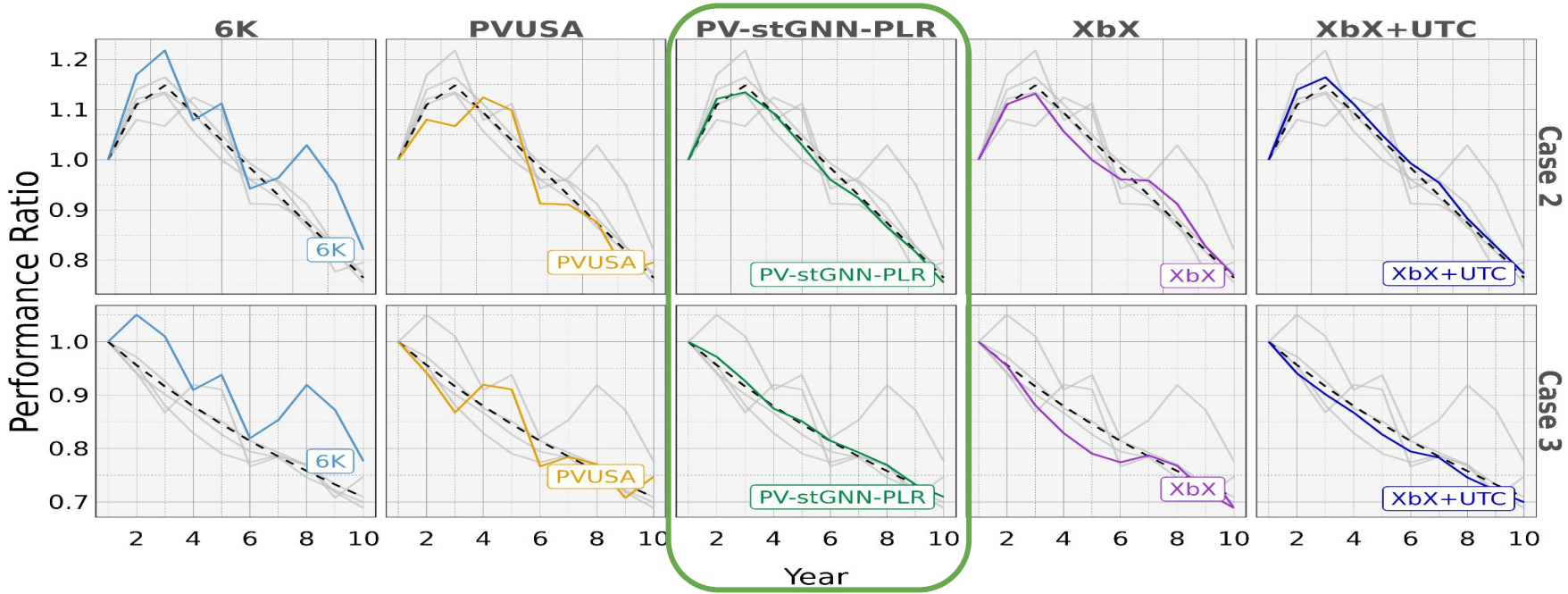
PVPS

Performance Loss Rate (PLR) Determination



PVPS

Best PLR Determination: Using st-Graph Modeling



Fleet wide performance analysis, performed using st-GNN Models

- Spatiotemporal Graph Neural Network Models

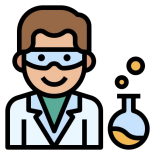
[1] Y. Fan, R. Wieser, X. Yu, Y. Wu, L. S. Bruckman, and R. H. French, "Using Spatio-Temporal Graph Neural Networks to Estimate Fleet-Wide Photovoltaic Performance Degradation Patterns," PLOS ONE, vol. 19, no. 2, p. e0297445, Feb. 2024, doi: 10.1371/journal.pone.0297445.
[2] Y. Fan et al., "Spatio-Temporal Denoising Graph Autoencoders with Data Augmentation for Photovoltaic Data Imputation," in Proceedings of the ACM on Management of Data, May 2023, pp. 1–19. doi: 10.1145/3588730.

The Problem with Data

More time is often spent locating, querying, and assessing if data is fit to use Than actually analyzing and learning from the data itself

- Data is hinged on extensive institutional knowledge
- Large datasets are difficult to get a high level sense of
- Querying data can be complex if stored in multiple formats and databases

Do you remember where John's old EBSD results are?



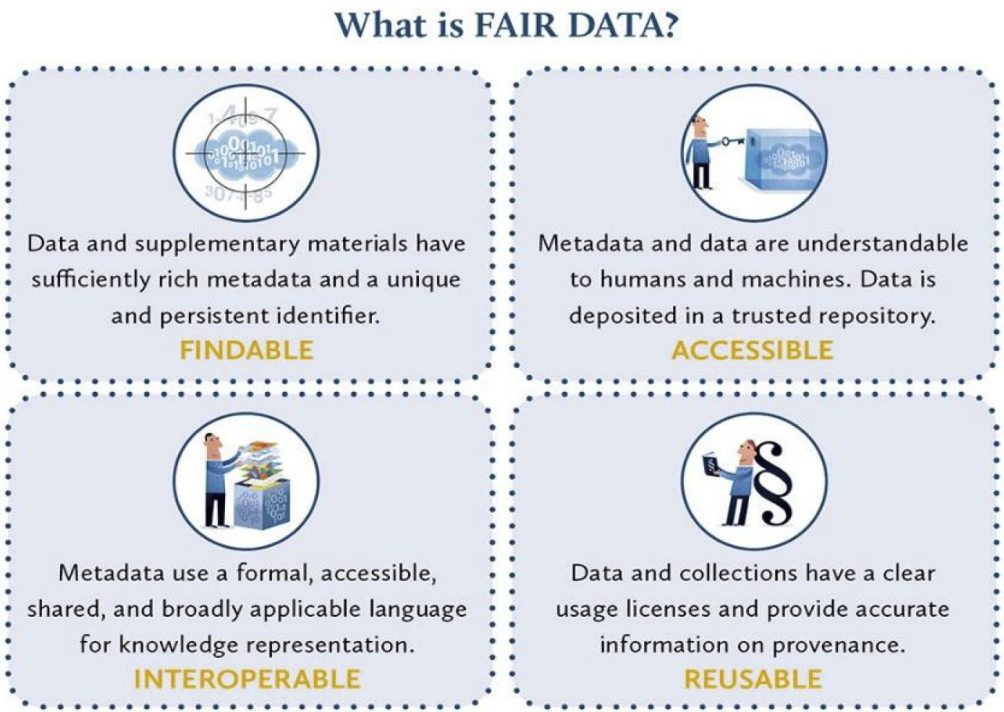
I think on CSV files on that old lab computer?

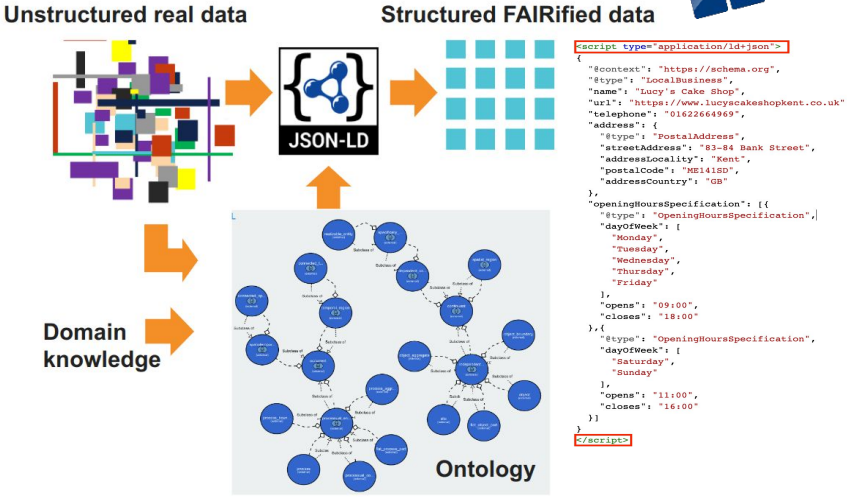
One week, 14 emails, and five people later…

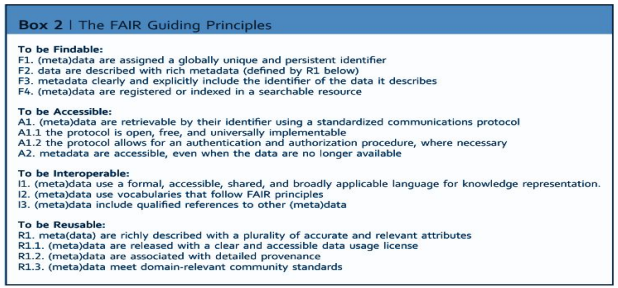
An outdated SQL query is found in a Word document retrieves only half of each experiment

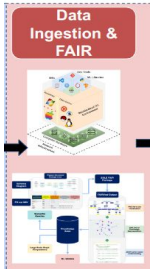
PVPS

Making Data & Models FAIR







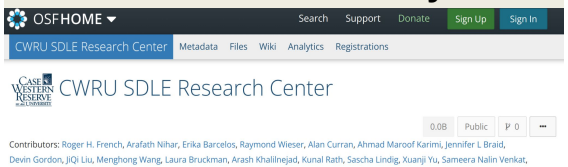


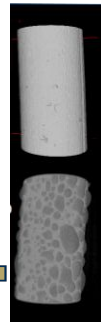
PVPS

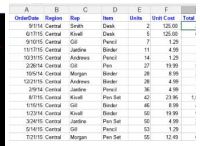
[1] W. C. Oltjen et al., “FAIRification, Quality Assessment, and Missingness Pattern Discovery for Spatiotemporal Photovoltaic Data,” in 2022 IEEE 49th Photovoltaics Specialists Conference (PVSC), Jun. 2022, pp. 0796–0801. doi: 10.1109/PVSC48317.2022.9938523.
[2] A. Nihar et al., “Toward Findable, Accessible, Interoperable and Reusable (FAIR) Photovoltaic System Time Series Data,” in 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC), Jun. 2021, pp. 1701–1706. doi: 10.1109/PVSC43889.2021.9518782.

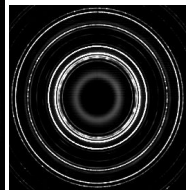
FAIRification & Data Science Workflow Pipelines

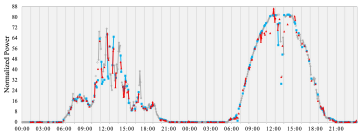
Datasets shared on osf.io with the community











Lawrence Livermore National Laboratory

Los Alamos NATIONAL LABORATORY EST. 1943

Sandia National Laboratories

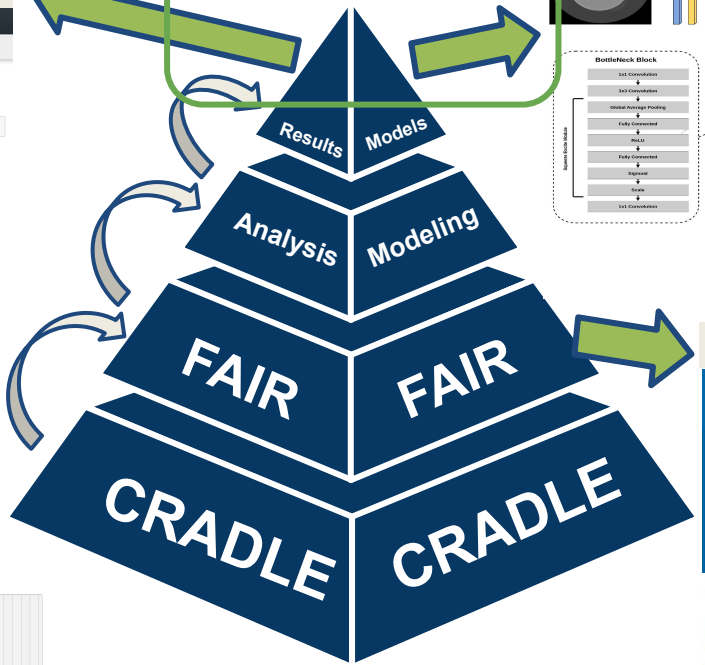
FAMU Florida Agricultural and Mechanical University

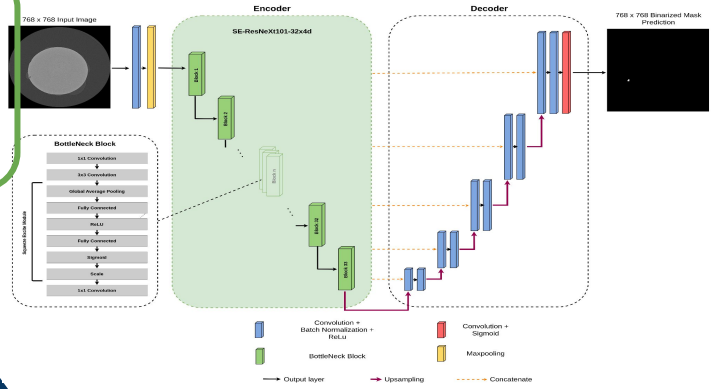
solaredge

Brookfield Renewable

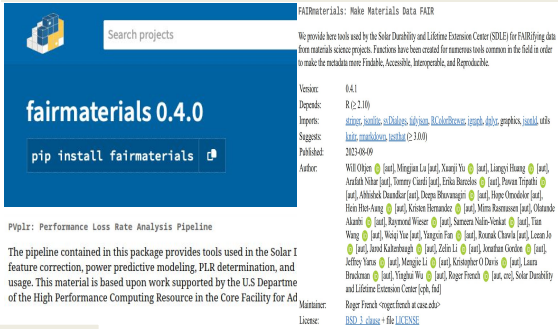
PVPS

FAIR Cycle Of Science AI/ML!





Open source tools to community



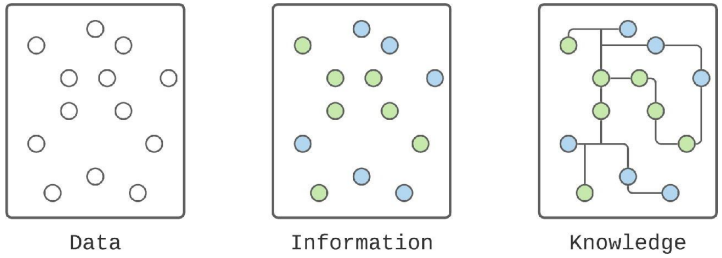
DATA

Data ingested at CRADLE/HPC

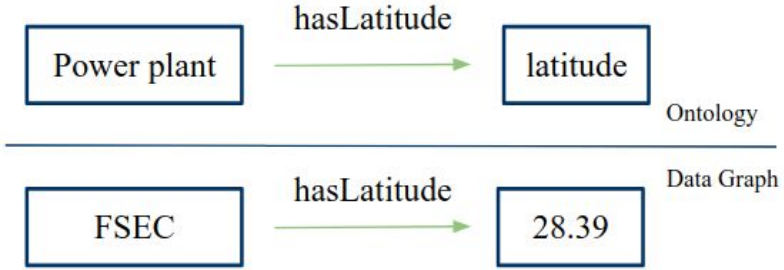
What is an Ontology?

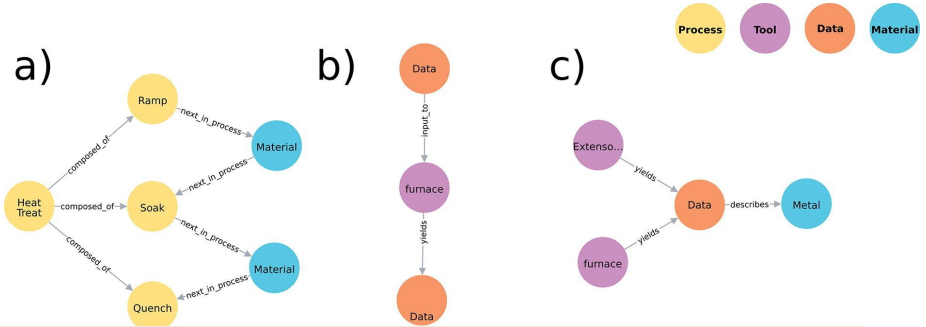
An ontology is a formal dictionary

- of terms for a given industry or domain that shows how the terms are related



- Terms are stored as object-relationship pairs ***Subject, predicate, and object***





By relating terms together, an ontology **can create a densely interconnected web**

PVPS

MDS-Onto is registered, Connects to Schema.org & W3C

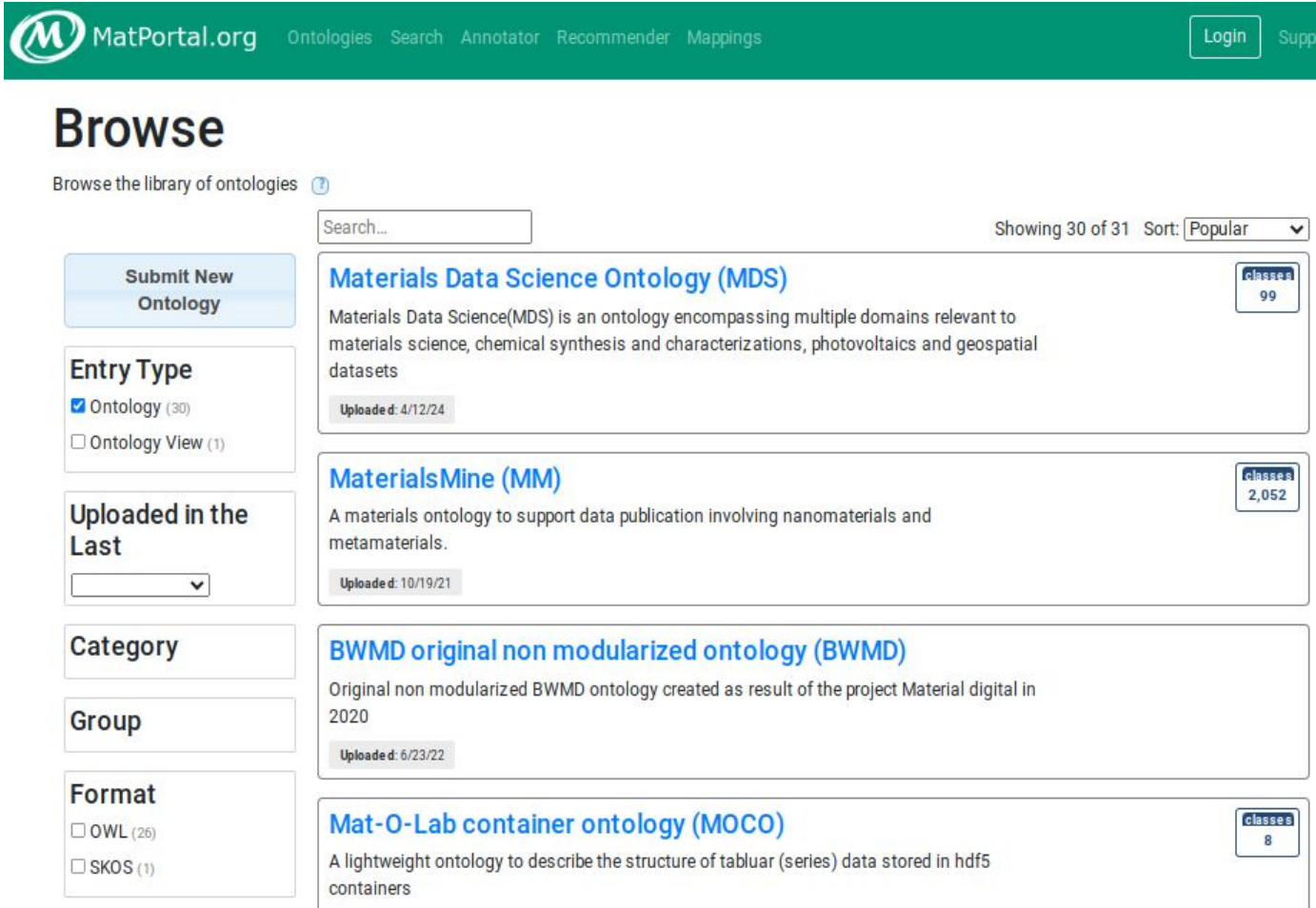
An Open Source Initiative

To standardize terminology

Across Materials Domains

While not requiring people
To change variable names

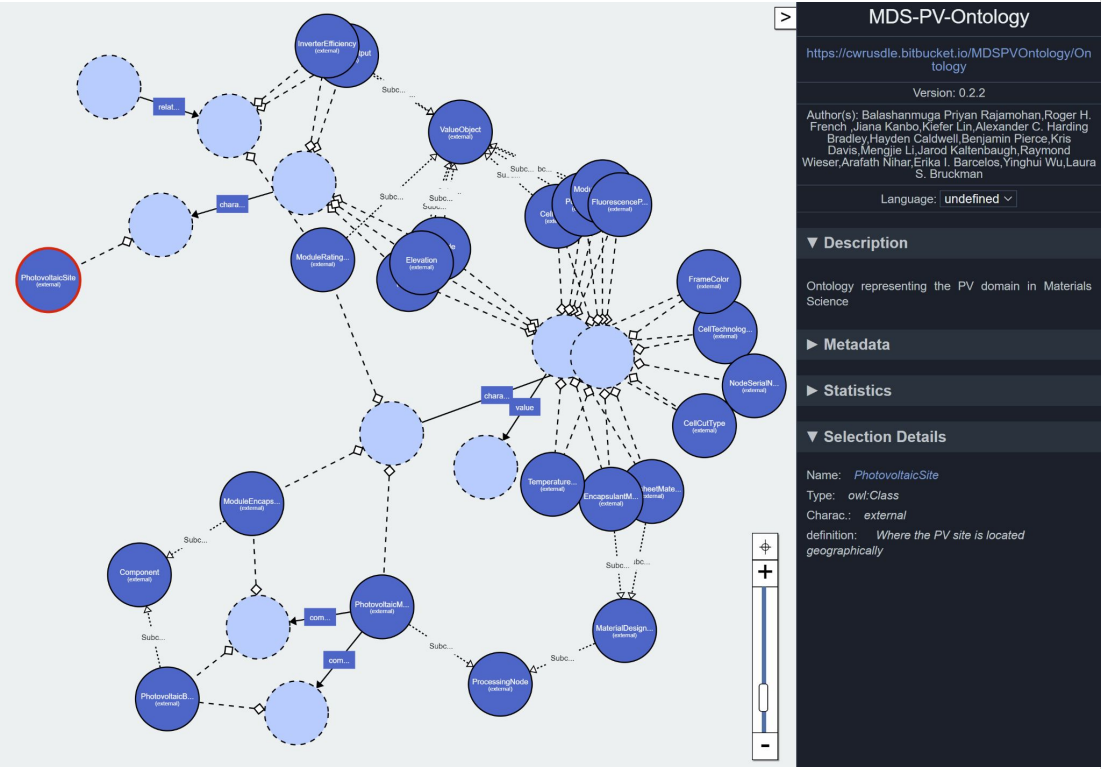
Major breakthrough in
Knowledge Management



mds-PV Ontology v0.2.2

Graph of the terms in mds-PV Onto

- PV Site, System, Inverter, Module
- Backsheet, Cell



JSON-LD


```
{
  "@context": {
    "PMDCo": "https://w3id.org/pmd/co/",
    "PVBacksheet": "https://cwrusdle.bitbucket.io/PVBacksheetOntology/",
    "PVCellOntology": "https://cwrusdle.bitbucket.io/PVCellOntology/",
    "PVInverterOntology": "https://cwrusdle.bitbucket.io/PVInverterOntology/",
    "PVModuleOntology": "https://cwrusdle.bitbucket.io/PVModuleOntology/",
    "PVSiteOntology": "https://cwrusdle.bitbucket.io/PVSiteOntology/",
    "QUDT": "http://qudt.org/2.1/vocab/unit#",
    "brick": "https://brickschema.org/schema/Brick#",
    "csvw": "http://www.w3.org/ns/csvw#",
    "dc": "http://purl.org/dc/elements/1.1/",
    "dcam": "http://purl.org/dc/dcam/",
    "dcat": "http://www.w3.org/ns/dcat#",
    "dcmitype": "http://purl.org/dc/dcmitype/",
    "dcterms": "http://purl.org/dc/terms/",
    "doap": "http://usefulinc.com/ns/doap#",
    "foaf": "http://xmlns.com/foaf/0.1/",
    "geo": "http://www.opengis.net/ont/geosparql#",
    "odrl": "http://www.w3.org/ns/odrl/2/",
    "org": "http://www.w3.org/ns/org#",
    "owl": "http://www.w3.org/2002/07/owl#",
    "prof": "http://www.w3.org/ns/dx/prof/",
    "prov": "http://www.w3.org/ns/prov#",
    "qb": "http://purl.org/linked-data/cube#",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "schema": "https://schema.org/",
    "sh": "http://www.w3.org/ns/shacl#",
    "skos": "http://www.w3.org/2004/02/skos/core#",
    "sosa": "http://www.w3.org/ns/sosa/",
    "ssn": "http://www.w3.org/ns/ssn/",
    "time": "http://www.w3.org/2006/time#",
    "vann": "http://purl.org/vocab/vann/",
    "void": "http://rdfs.org/ns/void#",
    "wgs": "https://www.w3.org/2003/01/geo/wgs84_pos#",
    "xml": "http://www.w3.org/XML/1998/namespace",
    "xsd": "http://www.w3.org/2001/XMLSchema#"
  },
```

We check Terms

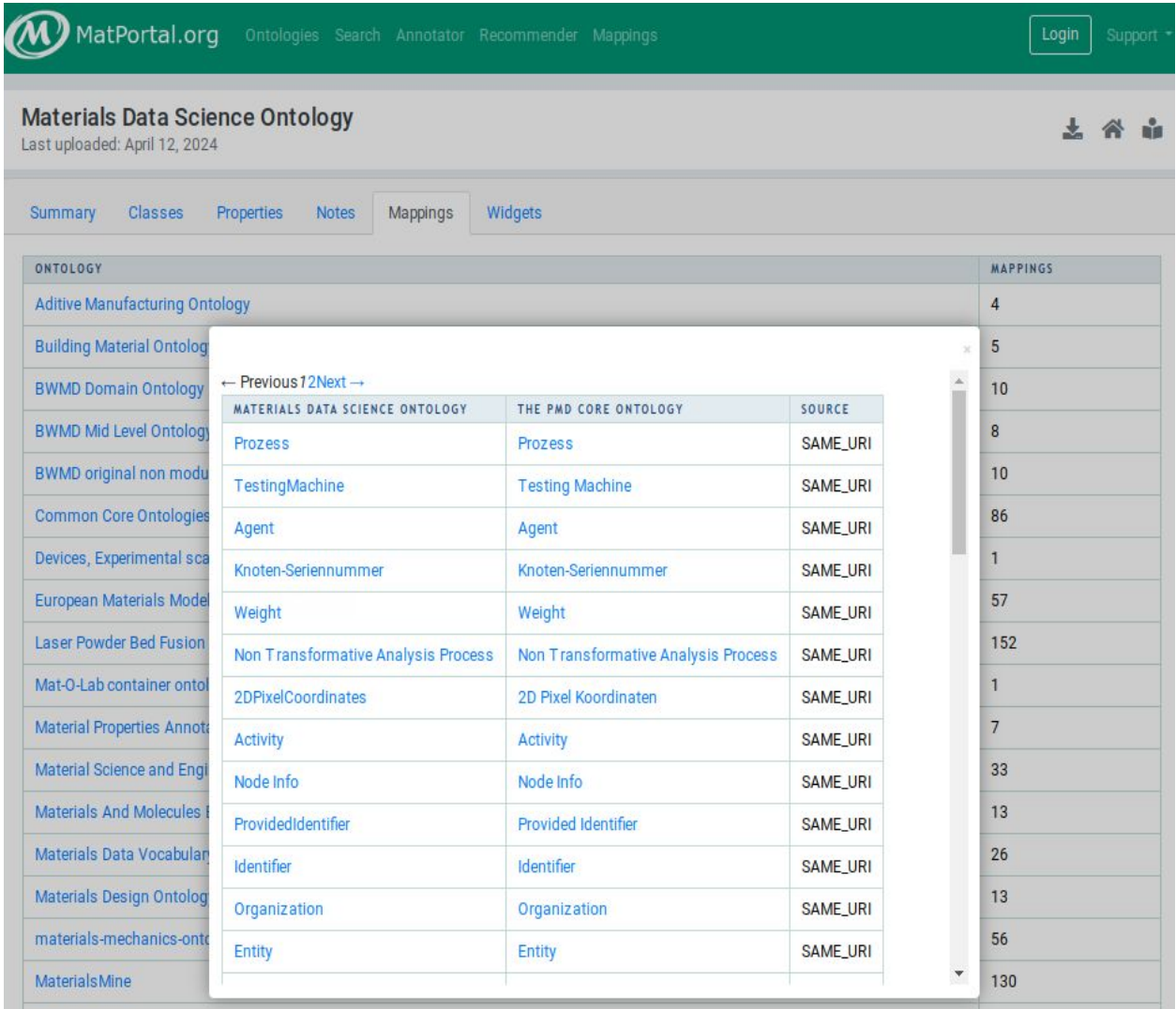
Across Materials Communities

A common terminology

- Is being established

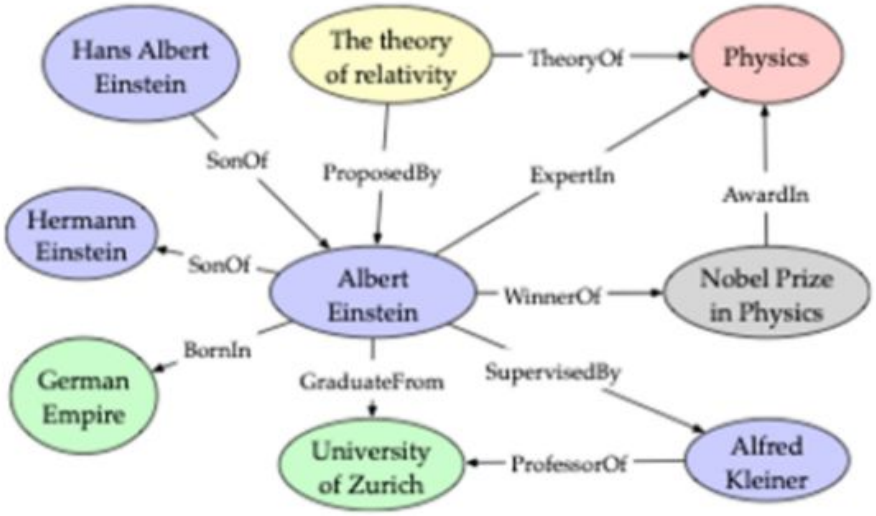
To Unify PV Datasets

- Enabling PV Histories
- For Valuation
- Of 2nd Life PV



Data Representations: RDF vs. LPG (Labeled Property Graph)

(Albert Einstein, **BornIn**, German Empire)
(Albert Einstein, **SonOf**, Hermann Einstein)
(Albert Einstein, **GraduateFrom**, University of Zurich)
(Albert Einstein, **WinnerOf**, Nobel Prize in Physics)
(Albert Einstein, **ExpertIn**, Physics)
(Nobel Prize in Physics, **AwardIn**, Physics)
(The theory of relativity, **TheoryOf**, Physics)
(Albert Einstein, **SupervisedBy**, Alfred Kleiner)
(Alfred Kleiner, **ProfessorOf**, University of Zurich)
(The theory of relativity, **ProposedBy**, Albert Einstein)
(Hans Albert Einstein, **SonOf**, Albert Einstein)

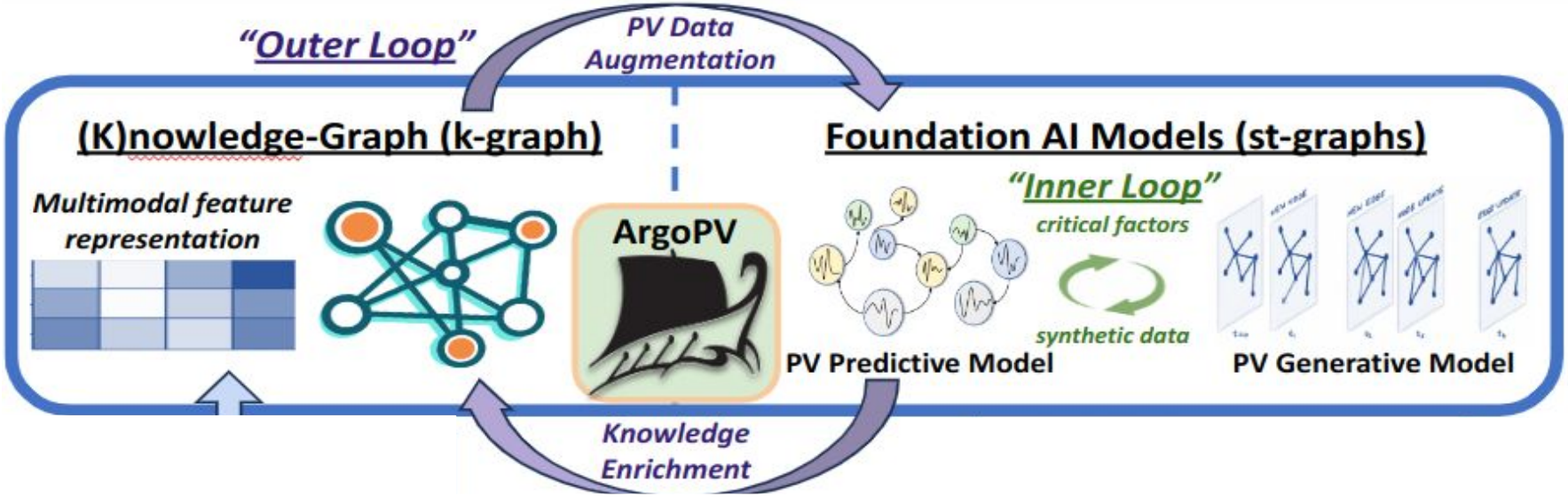


(a) Factual triples in knowledge base.

(b) Entities and relations in knowledge graph.

Using same "graph" approaches as Power Forecasting & PLR Determination

AI Inflection Point: FAIR data assists 2nd Life PV valuation



Knowledge Graphs **GNN ddDT/Foundation Models**

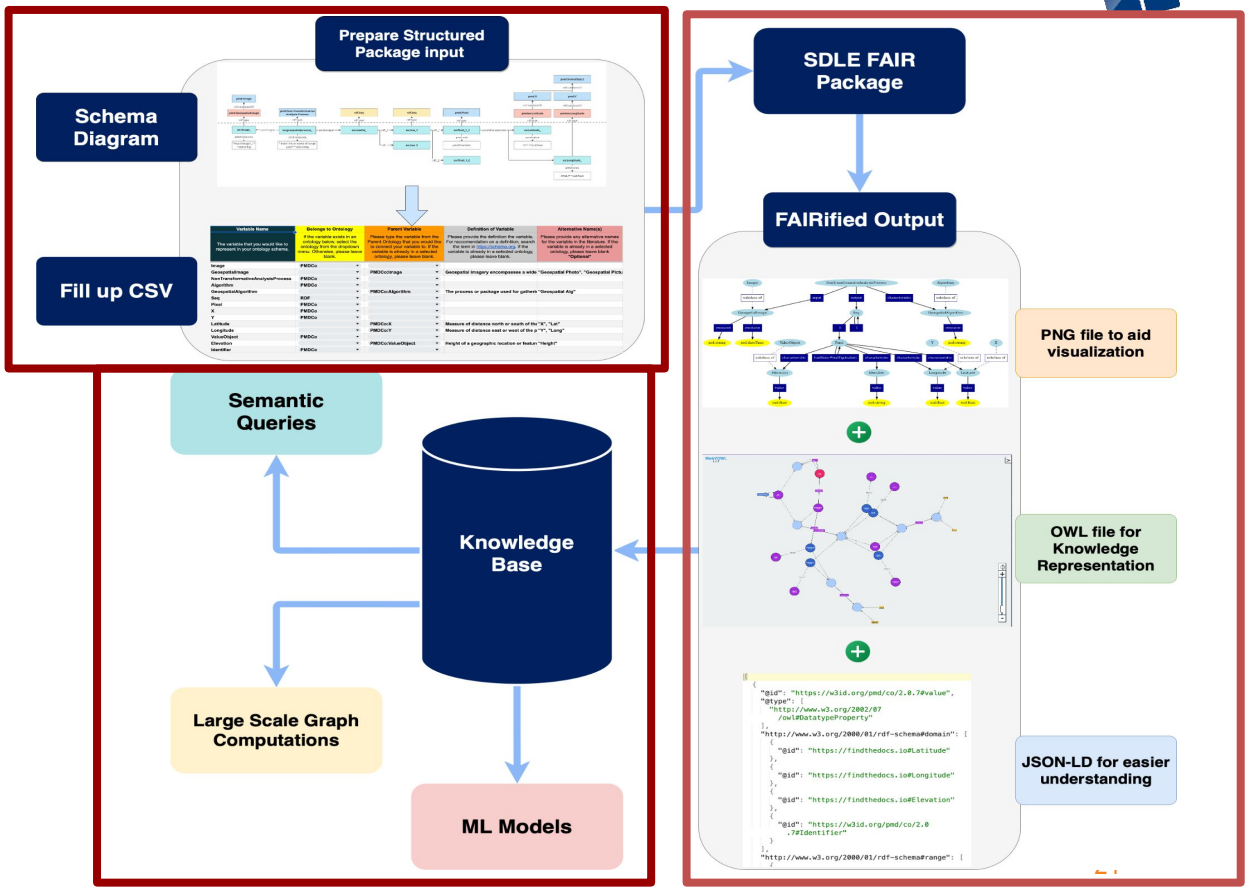
Combined st-Graph/ddDT Models & Knowledge Graph Repositories

PVPS

Questions

rxf131@case.edu

FAIR data model for the PV Lifecycle



PVPS