

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™: Common Research Analytics & Data Lifecycle Environment¹
 - Automated pipelines, FAIRification², Efficient Insights

Data Centric AI³ presents humans with a grand opportunity

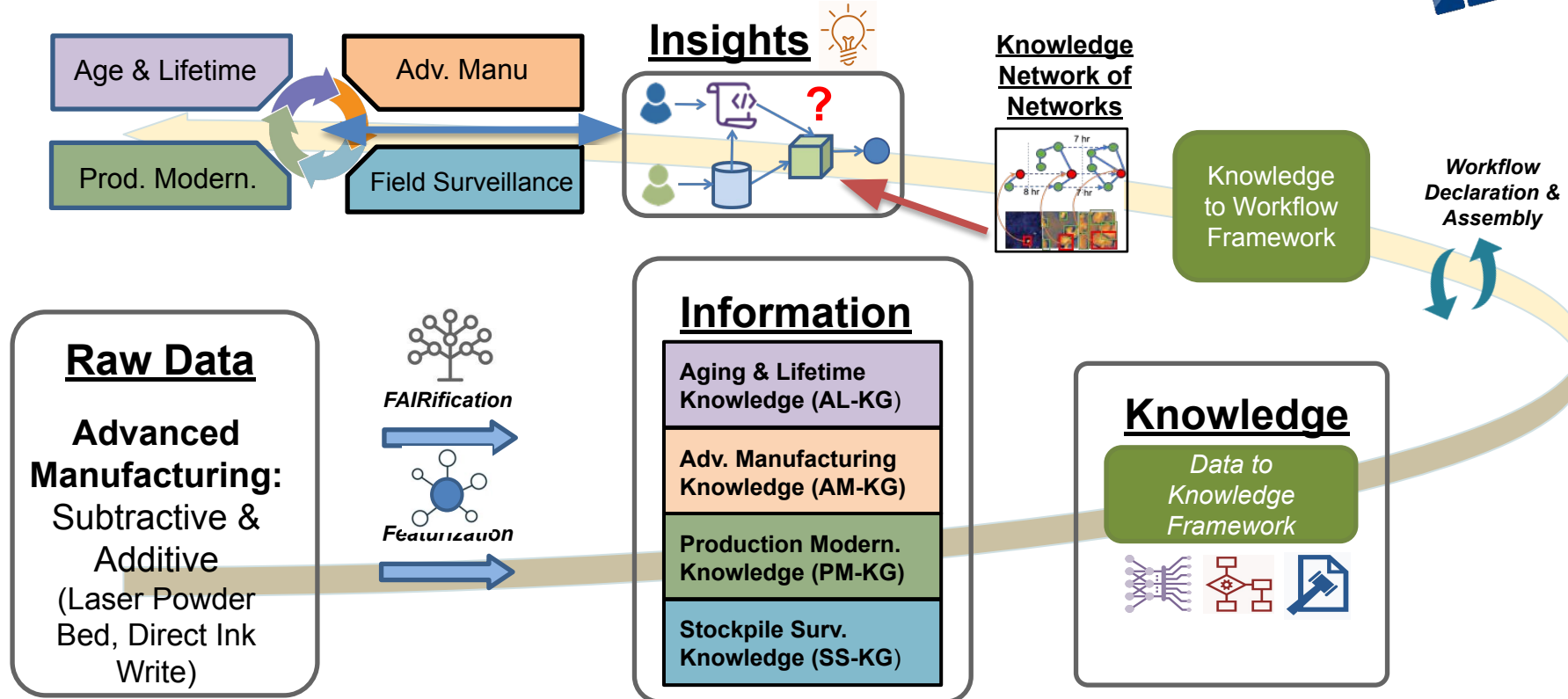
- “Computational Inflection Point for Scientific Discovery”⁴
 - Augmenting human reasoning; Working alongside human researchers
 - Scientific investigations restructured around the “salient human tasks”
 - With computers handling the routine and onerous tasks
 - Supplementing our human capabilities

While decreasing reductionist approaches in scientific research

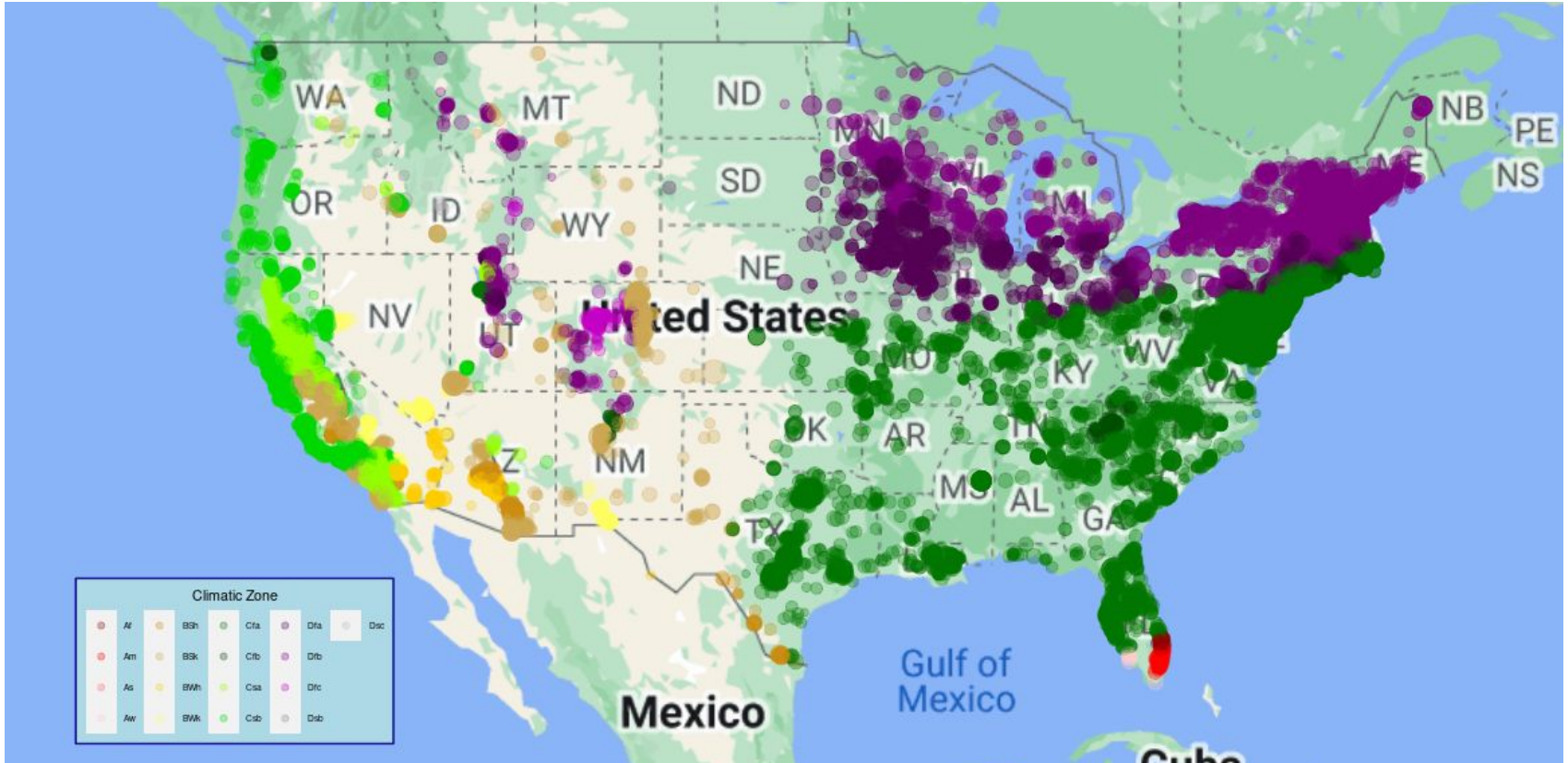
In SDLE Res. Cntr.

- 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

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



Large Scale Photovoltaic Fleet Monitoring: 104,700 PV Systems



104,700 PV Systems Ingested: EDA



Exploratory Data
Analysis (EDA)

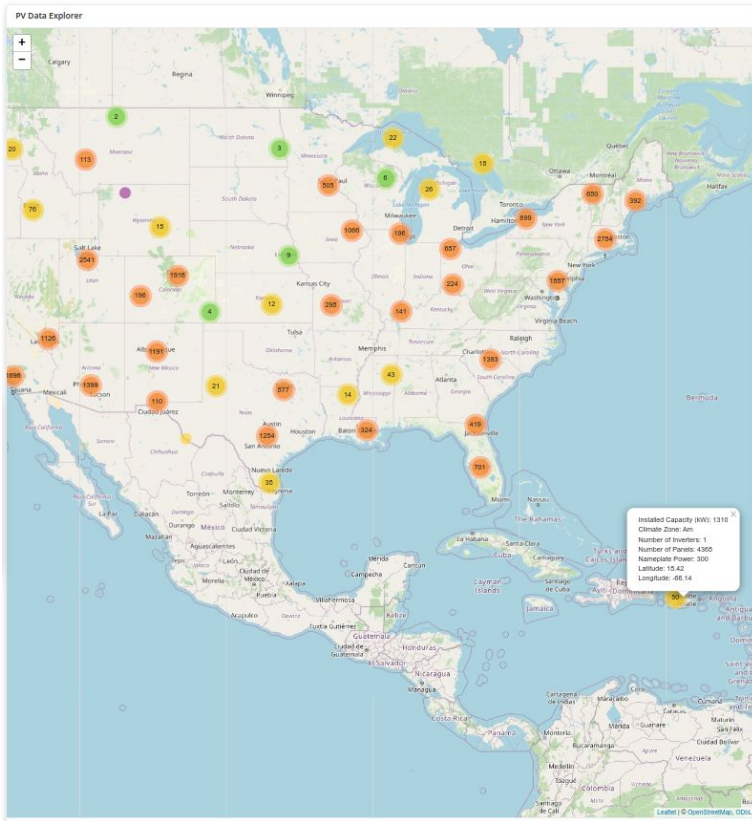
EDA of
SDLE PV Systems

Massive Data

Integrated with
Metadata



CRADLE Data Explorer



R Query to Fetch PV Metadata

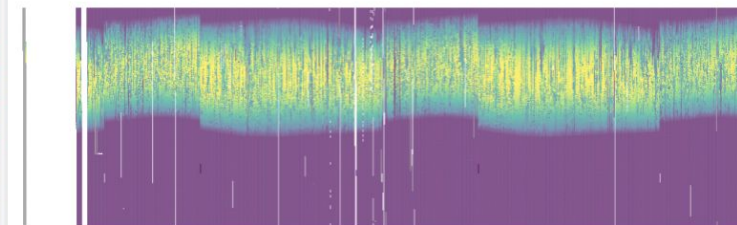
Metadata of PV Systems

1	pp	ss1	19.83	-155.79	b0580cn	Cfb	6be77d805385e0735c1b057a3ad18f6ddaef9a0	MLEET
2	pp	ss1	19.93	-155.79	cn78irs	Cfb	f7bc2d4a7ef6a9aee37b9beeafca9aaba45b1cc	MLEET
3	pp	ss1	19.93	-155.79	qwfu080	Cfb	f7bc2d4a7ef6a9aee37b9beeafca9aaba45b1cc	MLEET
4	pp	ss1	21.33	-157.9	wx8lr7g	As	f7bc2d4a7ef6a9aee37b9beeafca9aaba45b1cc	MLEET
5	pp	ss1	21.34	-157.9	a4mwbbm	As	850dbf76696f7dda659114892857d4c8b8e7a33c	MLEET
6	pp	ss1	21.36	-157.95	pimgpdv	As	f7bc2d4a7ef6a9aee37b9beeafca9aaba45b1cc	MLEET
7	pp	ss1	21.36	-157.95	pkupb0f	As	f7bc2d4a7ef6a9aee37b9beeafca9aaba45b1cc	MLEET
8	pp	ss1	27.19	-82.4	l2zq550	Cfa	f7bc2d4a7ef6a9aee37b9beeafca9aaba45b1cc	MLEET
9	pp	ss1	32.25	-111.26	pdku8hz	BSh	8a96ac37b7739f71b5753089d51783912531917	MLEET
10	pp	ss1	32.25	-111.26	uvvr8b0	BSh	8a96ac37b7739f71b5753089d51783912531917	MLEET

Showing 1 to 10 of 101,520 entries

Previous 1 2 3 4 5 ... 10152 Next

Heatmaps of Selected PV System



PVPS

CRADLE
Containerized Data
Science Tools

Photovoltaic Systems: CRADLE Data Explorer



CRADLE Data Explorer PV Sytems XRD AFM

PV Data Explorer

Installed Capacity (kW): 81
Climate Zone: Am
Number of Inverters: 3
Number of Panels: 336
Nameplate Power: 240
Latitude: 18.45
Longitude: -66.99

R Query to Fetch PV Metadata

```
1 select * from get_hbase_connection() ...  
2 tbl('pvsysmeta') %%%  
3 group_by (latd, lond)
```

Metadata of PV Systems

Show 10 entries Search:

	dtyp	styp	latd	lond	row_key	kgcz	mods	npow	invm	loc
1	pp	ss1	19.83	-155.79	b0580cn	Cfb	6be77d805385e0735c1b057a3ad18f6daeffa90			
2	pp	ss1	19.93	-155.79	cn78irs	Cfb	f7bc2d4a7e6a9aee37b9beeaefca9aaba45b1cc			
3	pp	ss1	19.93	-155.79	qwfu080	Cfb	f7bc2d4a7e6a9aee37b9beeaefca9aaba45b1cc			
4	pp	ss1	21.33	-157.9	wx8lr7g	As	f7bc2d4a7e6a9aee37b9beeaefca9aaba45b1cc			

Heatmaps of Selected PV System

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

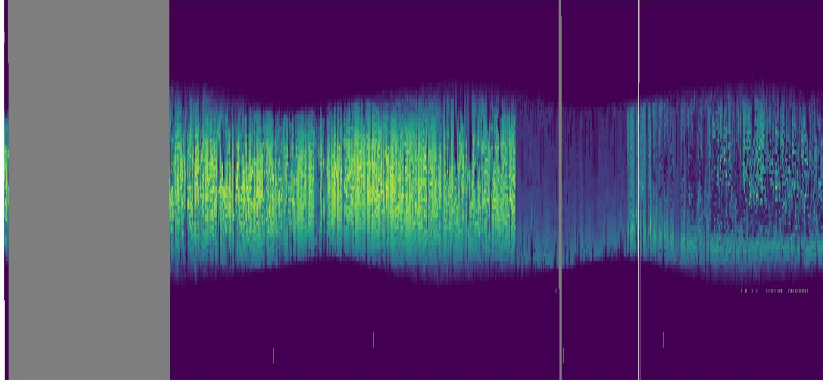
Interactive and quick assessment of data quality with missingness visualization

Data Reconstruction: Block Outages & Anomalous Measurements

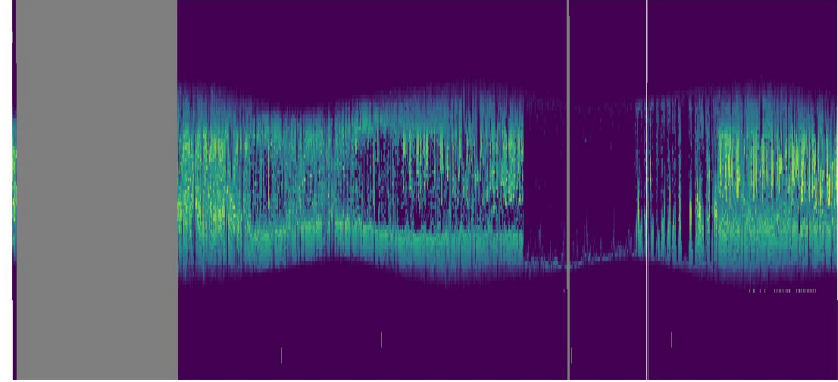


RAW

s2025_inv1_17.3

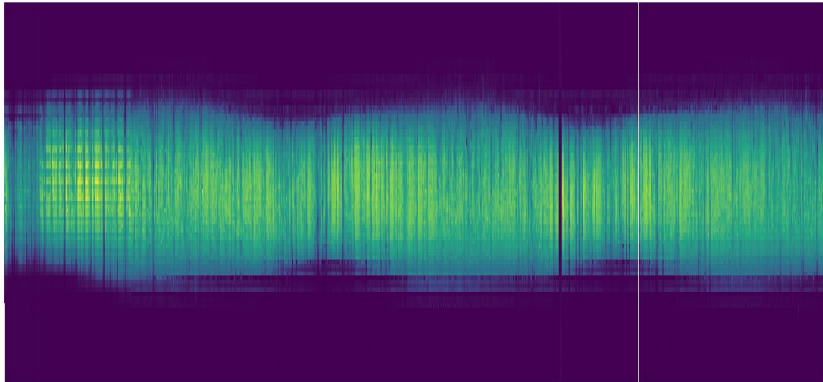


s2025_inv2_18

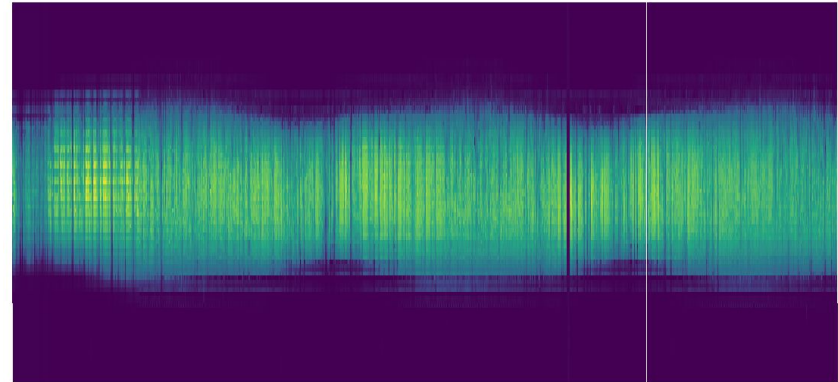


Reconstruction

s2025_inv1_17.3

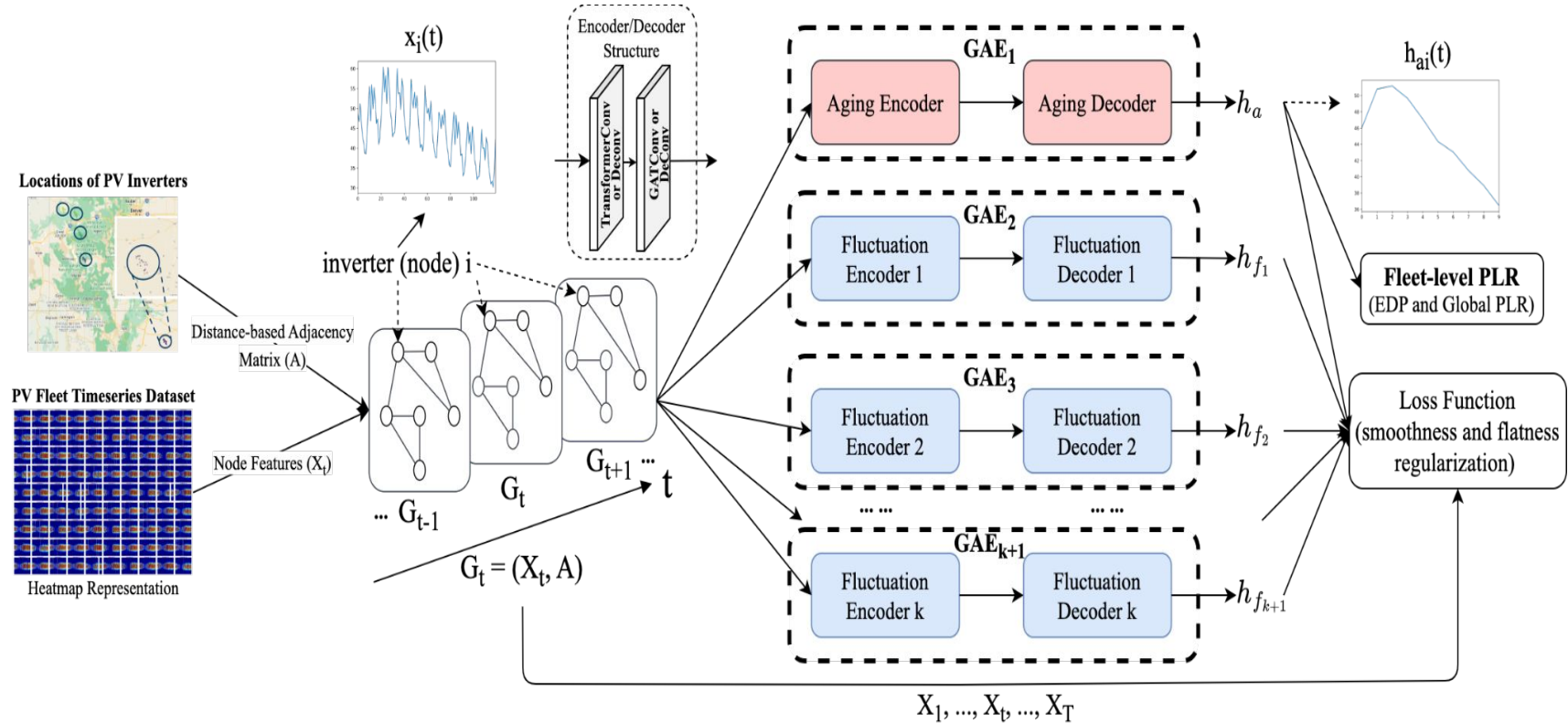


s2025_inv2_18

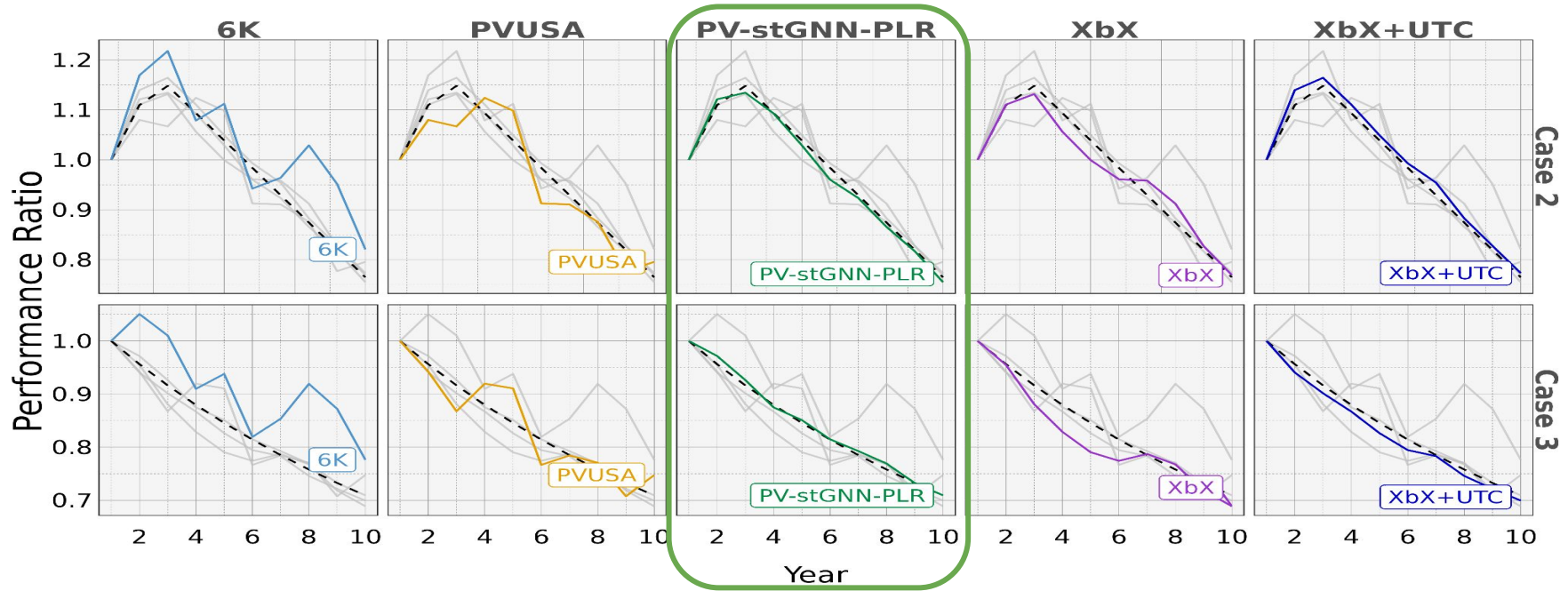


PVPS

Performance Loss Rate (PLR) Determination



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

Making Data & Models FAIR



What is FAIR DATA?



Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

FINDABLE



Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

ACCESSIBLE



Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

INTEROPERABLE



Data and collections have a clear usage licenses and provide accurate information on provenance.

REUSABLE

Unstructured real data



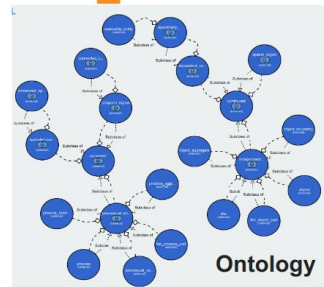
Structured FAIRified data



```

<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "LocalBusiness",
  "name": "Lucy's Cake Shop",
  "url": "https://www.lucyscakeshopkent.co.uk",
  "telephone": "01622664969",
  "address": {
    "@type": "PostalAddress",
    "streetAddress": "83-84 Bank Street",
    "addressLocality": "Kent",
    "postalCode": "ME141SD",
    "addressCountry": "GB"
  },
  "openingHoursSpecification": [ {
    "@type": "OpeningHoursSpecification",
    "dayOfWeek": [
      "Monday",
      "Tuesday",
      "Wednesday",
      "Thursday",
      "Friday"
    ],
    "opens": "09:00",
    "closes": "18:00"
  } ],
  "openingHoursSpecification": [ {
    "@type": "OpeningHoursSpecification",
    "dayOfWeek": [
      "Saturday",
      "Sunday"
    ],
    "opens": "11:00",
    "closes": "16:00"
  } ]
}
</script>
    
```

Domain knowledge



Ontology

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by F1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

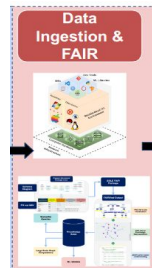
- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. (meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards



[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

OSFHOME Search Support Donate Sign Up Sign In

CWRU SDLE Research Center Metadata Files Wiki Analytics Registrations

CASE WESTERN RESERVE UNIVERSITY CWRU SDLE Research Center

Contributors: Roger H. French, Arafath Nihar, Erika Barcelos, Raymond Wieser, Alan Curran, Ahmad Maroof Karimi, Jennifer L. Bradi, Devin Gordon, JIQI Liu, Menghong Wang, Laura Bruckman, Arash Khalilnejad, Kunal Rath, Sacha Lindig, Xuanji Yu, Sameera Nalin Venkat, Will Ojien, Olatunde Akanbi

Affiliated institutions: Case Western Reserve University

Date created: 2019-03-25 10:34 AM | Last Updated: 2023-06-04 06:41 PM

Identifier: DOI:10.17605/OSF.IO/WN35J

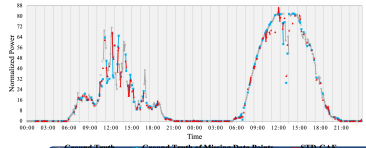
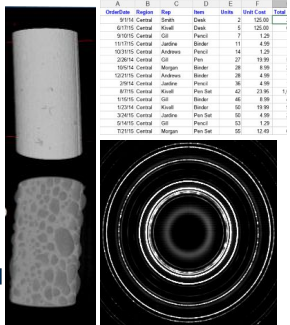
Category: Project

Description: Open Data and Open Codes shared from CWRU SDLE Research Center

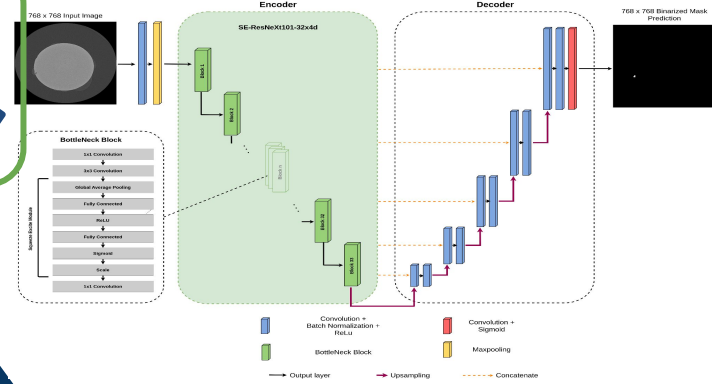
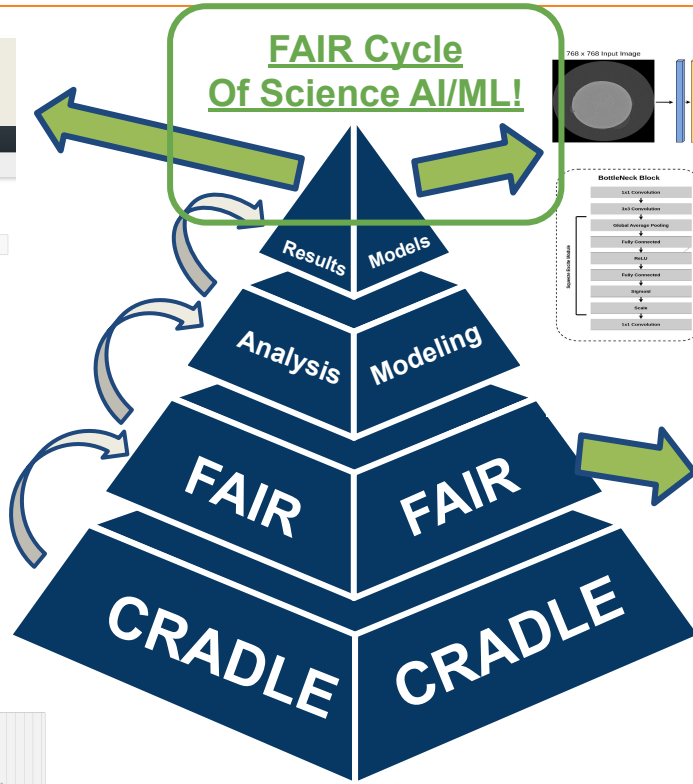
License: CC-BY Attribution 4.0 International



PVPS



FAIR Cycle Of Science AI/ML!



Open source tools to community

fairmaterials 0.4.0

pip install fairmaterials

fairMaterials: Make Materials Data FAIR

We provide here tools used by the Solar Diversity and Lifetime Extension Center (SDLEC) for FAIRifying data from materials science projects. Functions have been created for accessing tools common in the field in order to make the materials more Findable, Accessible, Interoperable, and Reproducible.

Version: 0.4.1
 Depends: R (≥ 2.10)
 Imports: astropy, pandas, sklearn, numpy, matplotlib, seaborn, plotly, pytorch, tensorflow
 Suggests: kdtree, matplotlib, sklearn (≥ 2.10)
 Published: 2023-08-09

Author: Will Ojien, Mengjiao Lu, Xuanji Yu, Liangliang Huang, Arafath Nihar, Sunny Claret, Erika Barcelos, Pooja Tropic, Ashish Dandekar, Deepa Braungott, Sage Chrookor, Hsin-Tzu Ang, Kishor Kumar, Kim-Su Kim, Sameera Nalin Venkat, Arash Khalilnejad, Kunal Rath, Sacha Lindig, Xuanji Yu, Sameera Nalin Venkat, Will Ojien, Woei Vei, Yaqun Fan, Rounak Chandra, Lucas Li, Jinchuan Zhang, Zhen Li, Jonathan Gordon, Jeffrey Yano, Mengjie Li, Christopher O'Donnell, Laura Bruckman, Yiqing Wu, Roger French, Salar Diversity and Lifetime Extension Center (epf, hd)

Pipfile: Performance Loss Rate Analysis Pipeline

The pipeline contained in this package provides tools used in the Solar I feature correction, power predictive modeling, PLR determination, and usage. This material is based upon work supported by the U.S. Department of the High Performance Computing Resource in the Core Facility for Advanced Computing Resources at Case Western Reserve University.

License: BSD 3-Clause - No LicensE

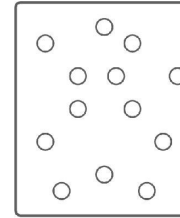
Data ingested at CRADLE/HPC

What is an Ontology?

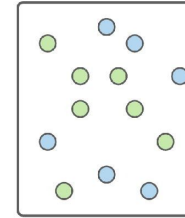


An ontology is a formal dictionary

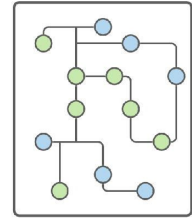
- of terms for a given industry or domain that shows how the terms are related
- Terms are stored as object-relationship pairs
Subject, predicate, and object



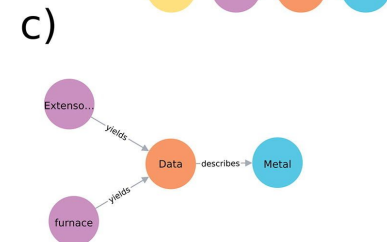
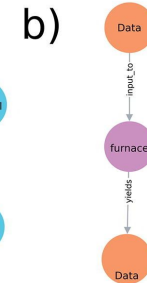
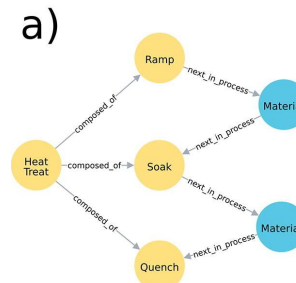
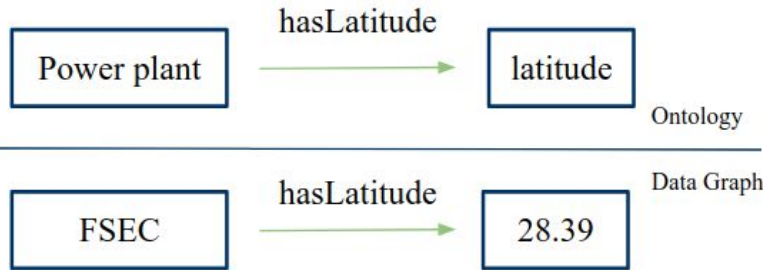
Data



Information



Knowledge



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

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

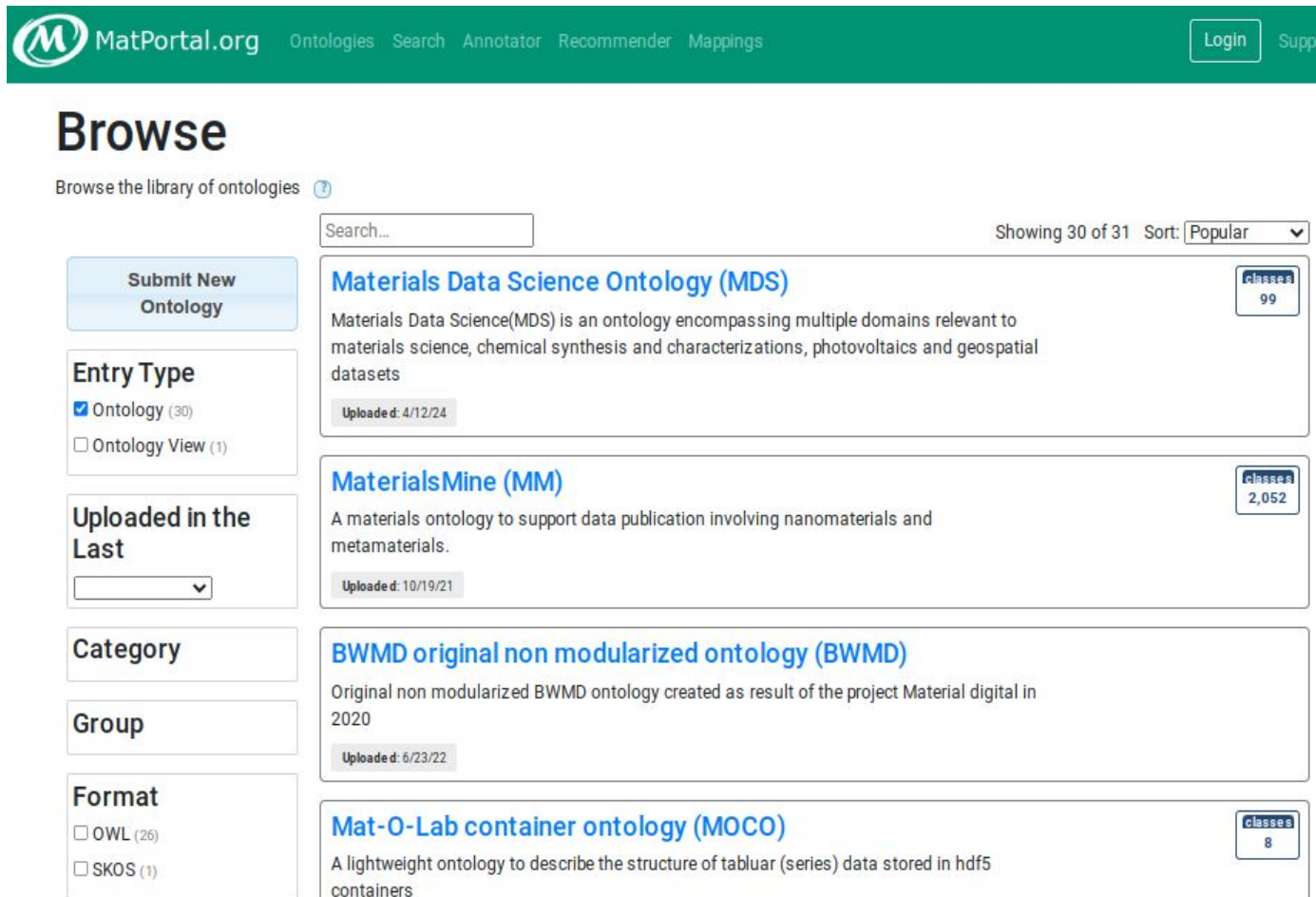
An Open Source Initiative

To standardize terminology

Across Materials Domains

While not requiring people
To change variable names

Major breakthrough in
Knowledge Management



The screenshot shows the MatPortal.org website interface. At the top, there is a green navigation bar with the MatPortal.org logo and menu items: Ontologies, Search, Annotator, Recommender, Mappings, Login, and Support. The main content area is titled 'Browse' and includes a search bar and a 'Submit New Ontology' button. On the left, there are several filter panels: 'Entry Type' (with 'Ontology (30)' selected), 'Uploaded in the Last' (with a dropdown menu), 'Category', 'Group', and 'Format' (with 'OWL (26)' and 'SKOS (1)' options). The main list displays three ontologies: 'Materials Data Science Ontology (MDS)' with 99 classes, 'MaterialsMine (MM)' with 2,052 classes, and 'BWMD original non modularized ontology (BWMD)' with 8 classes. Each entry includes a description and an upload date.

MatPortal.org Ontologies Search Annotator Recommender Mappings Login Support

Browse

Browse the library of ontologies ?

Search... Showing 30 of 31 Sort: Popular

Submit New Ontology

Entry Type

- Ontology (30)
- Ontology View (1)

Uploaded in the Last

Category

Group

Format

- OWL (26)
- SKOS (1)

Materials Data Science Ontology (MDS) classes 99

Materials Data Science(MDS) is an ontology encompassing multiple domains relevant to materials science, chemical synthesis and characterizations, photovoltaics and geospatial datasets

Uploaded: 4/12/24

MaterialsMine (MM) classes 2,052

A materials ontology to support data publication involving nanomaterials and metamaterials.

Uploaded: 10/19/21

BWMD original non modularized ontology (BWMD)

Original non modularized BWMD ontology created as result of the project Material digital in 2020

Uploaded: 6/23/22

Mat-O-Lab container ontology (MOCO) classes 8

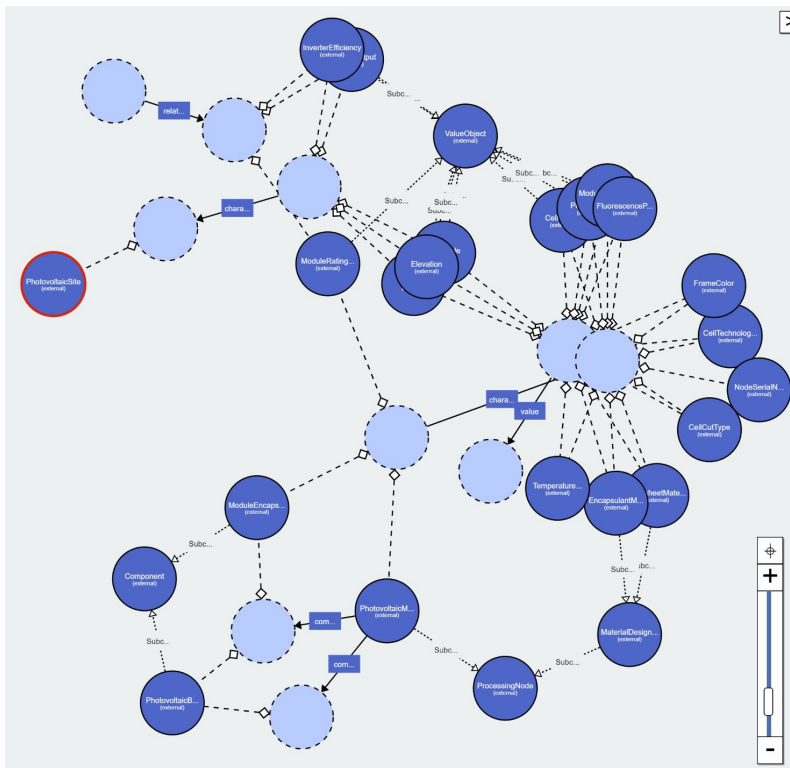
A lightweight ontology to describe the structure of tabluar (series) data stored in hdf5 containers

mds-PV Ontology v0.2.2

Graph of the terms in mds-PV Onto

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

JSON-LD



MDS-PV-Ontology

<https://cwrusdle.bitbucket.io/MDSPVOntology/Ontology>

Version: 0.2.2

Author(s): Balashanmuga Priyan Rajamohan Roger H. French, Jiana Kanbo, Kiefer Lin, Alexander C. Harding, Bradley Hayden Caldwell, Benjamin Pierce, Kris Davis, Mengjie Li, Jared Kellert, Raymond Wieser, Arafath Nihar, Erika I. Barcelos, Yinghui Wu, Laura S. Bruckman

Language: undefined

Description

Ontology representing the PV domain in Materials Science

Metadata

Statistics

Selection Details

Name: PhotovoltaicSite
Type: owl:Class
Charac: external
definition: Where the PV site is located geographically

```
{
  "@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/",
    "dcterm": "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

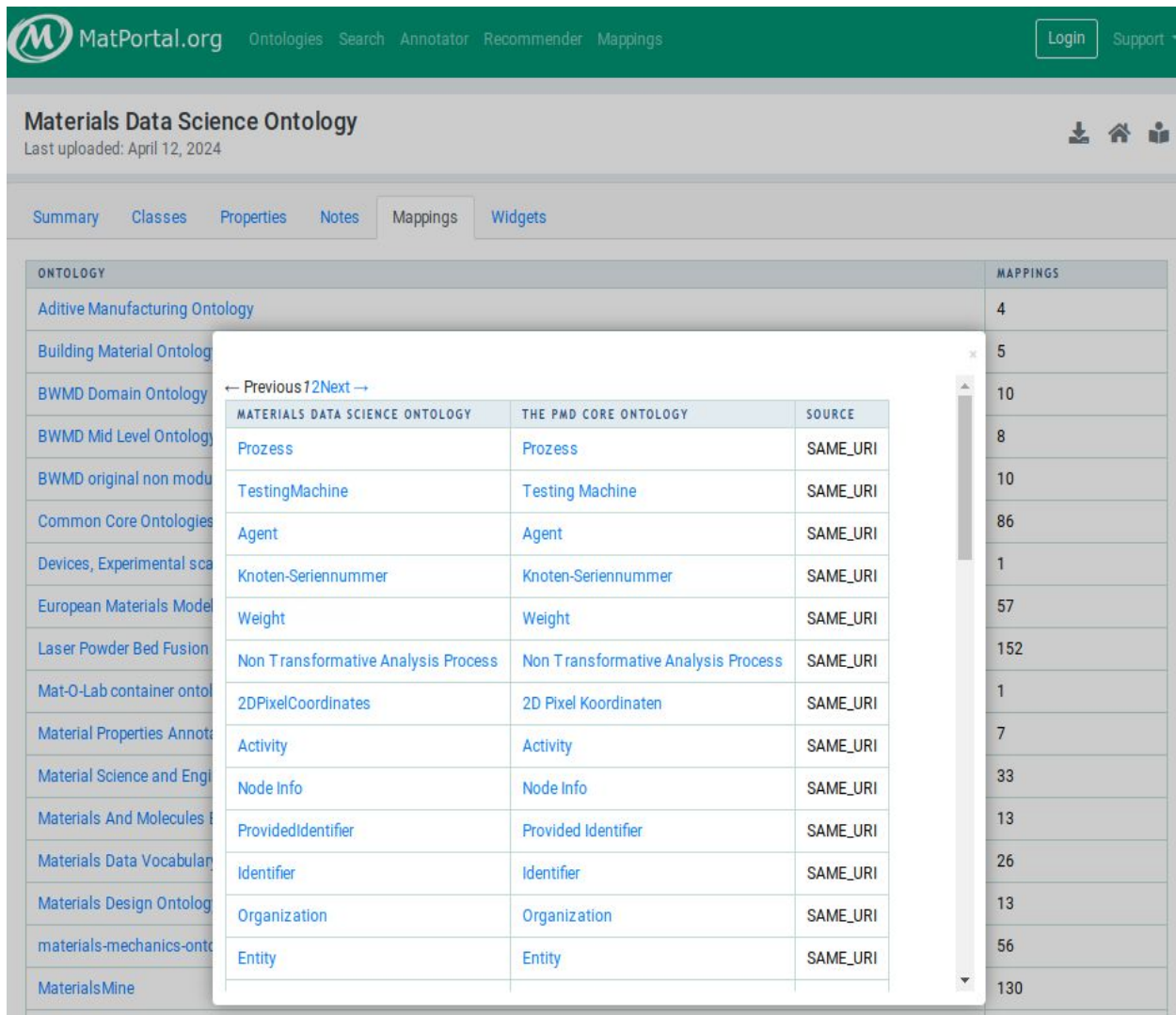
Across Materials Communities

A common terminology

- Is being established

To Unify PV Datasets

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



MatPortal.org Ontologies Search Annotator Recommender Mappings Login Support

Materials Data Science Ontology

Last updated: April 12, 2024

Summary Classes Properties Notes Mappings Widgets

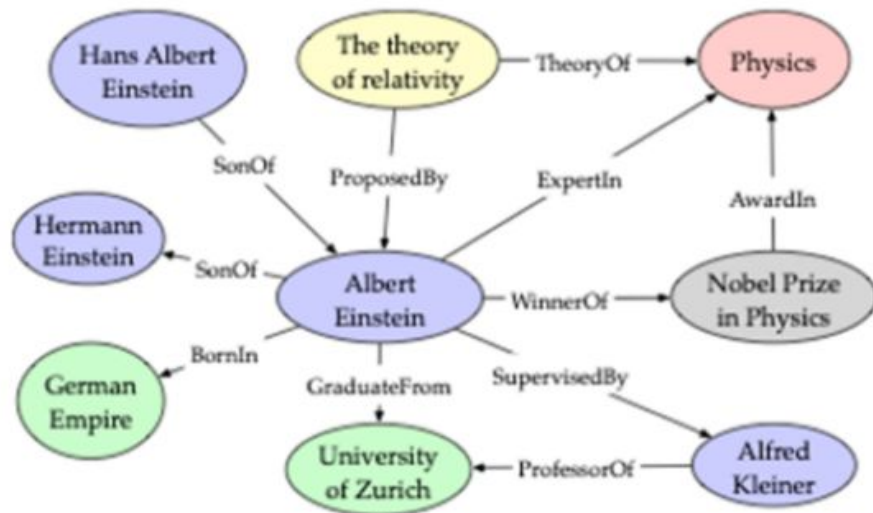
ONTOLOGY	MAPPINGS
Additive Manufacturing Ontology	4
Building Material Ontology	5
BWMD Domain Ontology	10
BWMD Mid Level Ontology	8
BWMD original non modular	10
Common Core Ontologies	86
Devices, Experimental sca	1
European Materials Model	57
Laser Powder Bed Fusion	152
Mat-O-Lab container ontol	1
Material Properties Annot	7
Material Science and Engi	33
Materials And Molecules F	13
Materials Data Vocabular	26
Materials Design Ontolog	13
materials-mechanics-ont	56
MaterialsMine	130

← Previous 12 Next →

MATERIALS DATA SCIENCE ONTOLOGY	THE PMD CORE ONTOLOGY	SOURCE
Prozess	Prozess	SAME_URI
TestingMachine	Testing Machine	SAME_URI
Agent	Agent	SAME_URI
Knoten-Seriennummer	Knoten-Seriennummer	SAME_URI
Weight	Weight	SAME_URI
Non T transformative Analysis Process	Non T transformative Analysis Process	SAME_URI
2DPixelCoordinates	2D Pixel Koordinaten	SAME_URI
Activity	Activity	SAME_URI
Node Info	Node Info	SAME_URI
ProvidedIdentifier	Provided Identifier	SAME_URI
Identifier	Identifier	SAME_URI
Organization	Organization	SAME_URI
Entity	Entity	SAME_URI

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

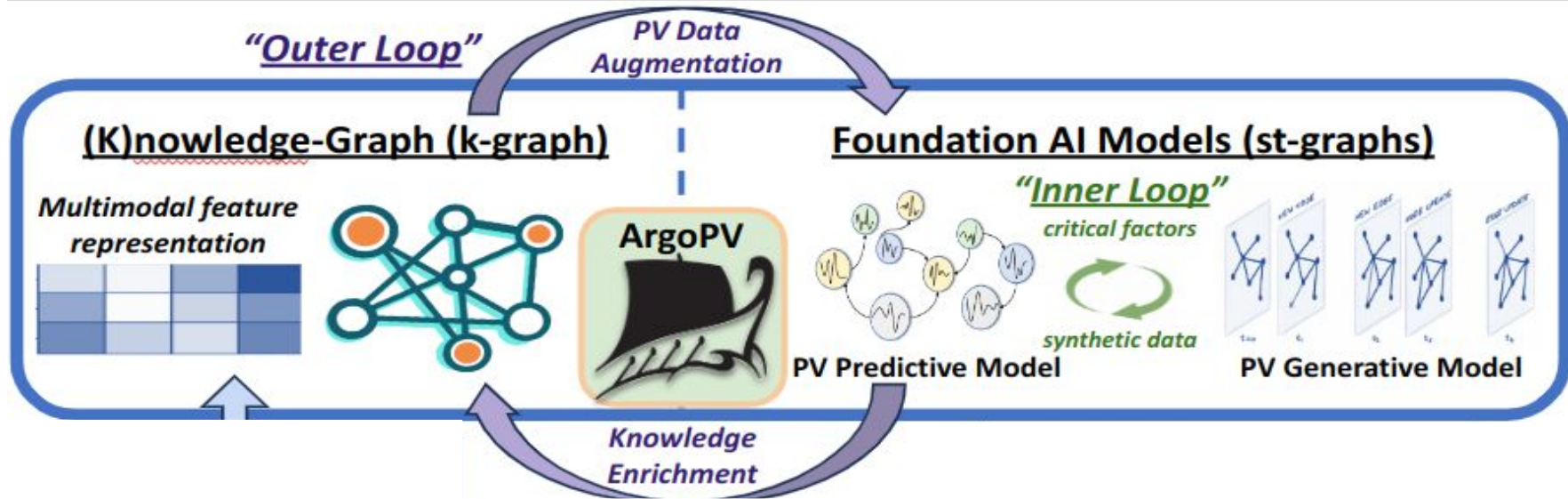
(Albert Einstein, **BornIn**, German Empire)
(Albert Einstein, **SonOf**, Hermann Einstein)
(Albert Einstein, **GraduateFrom**, University of Zurich)
(Albert Einstein, **WinnerOf**, Nobel Prize in Physics)
(Albert Einstein, **ExpertIn**, Physics)
(Nobel Prize in Physics, **AwardIn**, Physics)
(The theory 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

Questions

rxf131@case.edu



FAIR data model for the PV Lifecycle



PVPS

