This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SAND2021-7224C

Toward Findable, Accessible, Interoperable and Reusable (FAIR) Photovoltaic System Time Series Data

Arafath Nihar*, Alan J. Curran*, Ahmad M. Karimi*†, Jennifer L. Braid‡, Laura S. Bruckman*, Mehmet Koyutürk*, Yinghui Wu*, and Roger H. French*
*SDLE Research Center, Department of Materials Science and Engineering, CWRU, Cleveland, OH, 44106, USA
†Oak Ridge National Laboratory, Oak Ridge, TN, 37830, USA
‡Sandia National Laboratories, Albuquerque, NM, 87123, USA

***Abstract*—We present the application of FAIR principles to photovoltaic time series data to increase their reusability within the photovoltaic research community. The main requirements for a "FAIRified" dataset is to have a clearly defined data format, and to make accessible all metadata for this dataset to humans and machines. To achieve FAIRification, we implement a data model that separates the photovoltaic data and its metadata. The metadata and their descriptions are registered on a data repository in a human and machine readable format, using JSON-LD. Also, secure APIs are developed to access photovoltaic data. This approach has long term scalability and maintainability.**

***Index Terms*—Photovoltaics, Time Series data, FAIR principles, Metadata, JSON-ld, RDF/XML, W3C, RDF**

I. Introduction

The accessibility and reusability of digital research objects [1] generated from scientific research plays a critical role in the resulting impact of this research for society and the world. Research objects that are Findable, Accessible, Interoperable, and Reusable (FAIR) by both humans and machines ensure transparent, reproducible, and reusable science. The publication of Wilkinson et al.'s paper [2], [3] introduced a set of principles that aim to enhance the ability of machines to find and process data while also improving the reusability of data in the research enterprise while also supporting efforts to scientific reproducibility [4], [5]. The FAIR principles differ from other studies exploring the reusability of data in that the FAIR principles place particular emphasis on improving the reusability and findability of data by and for machines.

We present the application of FAIR principles to photovoltaic (PV) time series data of a fleet of 316 commerical PV power plant systems distributed across the US and studied by Karimi et al. [6]. Karimi demonstrates a deep learning model (a geospatiotemporal graph neural network) which uses the time series datasets of 316 PV systems (referred to as the raw PV data) from PV power plants to forecast PV power data (referred to as forecasted PV data) of all the power plants in the fleet. To our knowledge, this study is the first to present the application of FAIR principles to PV time series data. However, the FAIR principles have been extensively applied in other fields.

We can observe a movement towards data-driven research in many scientific fields and consequently the wide adoption of FAIR principles. For example, [7] applies FAIR principles to a data repository for plant phenomics called 'GnpIS', in order to enhance its interoperability with other data repositories. [7] is an instance of big federal data centers implementing FAIR data principles; FAIRification of Atmospheric Radiation Measurement [8] data. In [9], FAIR principles are applied to the German Network for Bioinformatics Infrastructure (de.NBI), which is a large distributed bioinformatics infrastructure. Further, [10] discusses the application of FAIR principles to health research, with [11] presenting the FAIRification of the Open-Source Registry for Rare Diseases (OSSE). Although, to our knowledge there is no literature on the application of FAIR principles to PV data, there are studies exploring the data-driven approach to PV research [12], [13], [14]. Therefore, the current study introducing the FAIRification of PV data is relevant and timely.

II. PV Data Description

We gather photovoltaic (PV) time series data from a fleet of 316 power plants located in various climate zones distributed across the US. There are multiple DC to AC electricity inverters in a single power plant, with each inverter controlling multiple racks of PV modules. The inverters, which are the primary devices of data collection, come in different varieties. The specifications of a particular inverter can be looked up if its manufacturer and model number are available.

The most common variables, and data, that are collected via inverters are listed in table I. In addition to the data collected from the power plants, weather and insolation data pertaining

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number DE-EE0009353.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Abridged Legal Disclaimer: The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

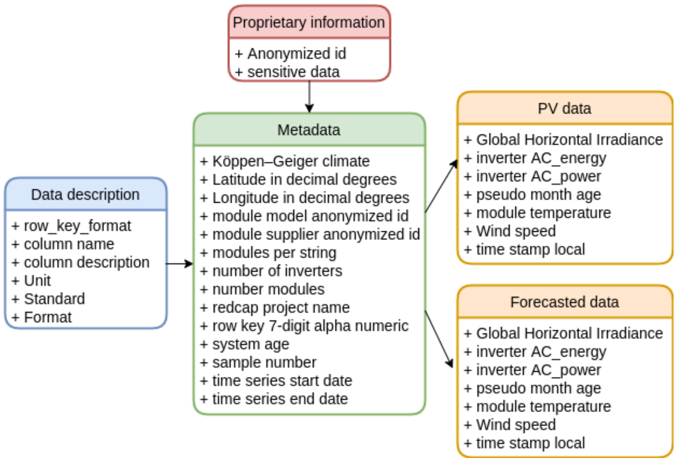

Fig. 1: The data model of the FAIRified PV data and metadata.The logical separations and relations between different entities are shown.

to the locations of all the power plants are collected via the SolarGIS API. SolarGIS data is based on their analysis of satellite images and development of a detailed atmospheric model [15], [16]. Table 2 lists the variables collected via solarGIS. The inverter data is typically 5 minute interval data, while the SolarGIS is intrinsically 30 minute interval due to the satellite's periodicity over any particular location, and is interpolated to 15 minute interval for ingestion.

The raw PV data used in this st-GNN deep learning study are a small segment of the full timeseries datasets of these systems, and consist of the power time series of the 316 PV power plants over a two year period. The PV data are stored in the SDLE Research Center's Apache Hadoop/Hbase/Spark cluster [17], [18], which we refer to as CRADLE (Common Research Analytics and Data Lifecycle Environment)), and is based on the Cloudera CDH distribution.

HBase, is a NoSQL database based on triples, where each cell has an associated rowkey and columnkey. These triples are similar to Resource Description Framework triples, so that HBase is often used as a RDF database store [19], [20]. This PV data is proprietary, so is not currently accessible, or findable by the general public PV research community, but the same FAIRification approaches can be used for open datasets that are share publically. The current study applies the FAIR principles to both raw and forecasted PV data.

III. Results

A. FAIR Data Model

The data model of the raw and forecasted PV data after FAIRification is shown in Fig. 1. The metadata and the PV time series data are clearly separated in this resultant data model. The data model includes detailed descriptions of the metadata. Additionally, the sensitive data, such as inverter or module supplier identification information, in the metadata are anonymized and stored separately in Research Electronic Data Capture (REDCap), which is a Health Insurance Portability and Accountability Act (HIPAA) compliant flat-file database and web application [21]. In the resultant data model the mapping between the metadata and the PV data is achieved via a unique alphanumeric row-key string, that serves as the anonymized PV system identifier. Although the raw and forecasted PV data are stored separately, they share the same row-key string.

B. FAIRified Implementation

The data are collected using the data transfer protocols sFTP, https, and REST APIs. The collected raw data are then moved to HDFS. After moving the data, Spark data ingestion jobs are scheduled. In the process of scheduling, the metadata are separated from the raw data and JSON-LD objects are generated from the metadata. Finally, these objects are published to CRADLE for our private cloud researchers or to the public on our CWRU-SDLE OSF site. Once the data are ingested to HBase, the data become easily accessible to the SDLE researchers for analysis, modeling, and prediction. The models and their results are ingested back into HBase and become available for future research work. Fig. 3 illustrates the pipeline developed for the FAIRification of the data.

Fig. 2 shows the infrastructure used to implement the FAIRified raw and forecasted PV data. The data are stored in HBase. Each entry in HBase is stored as RDF triples; a row key, a column key, and the input value. The physical representation of the FAIRified metadata and PV time series data consists of two data tables in HBase. The PV data are stored in the HBase table 'ecradle' and the metadata are stored in the HBase table 'meta'. A single row of the 'ecradle' table stores a single month's PV time series data. Further, each row is assigned a rowkey defined by the 7 character alphanumeric-yearmonth i.e bq03nu9-201803. The Scrypt algorithm with Salt is used to create the unique PV system identifier alphanumeric [22]. The

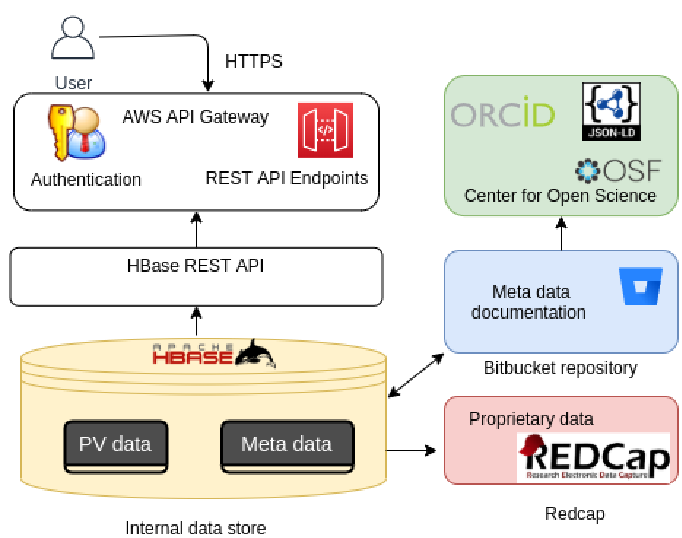

Fig. 2: The software architecture of the FAIRified PV data and metadata. The interconnections between the hardware and software components of the implementation are shown.

Variable	UN/CEFACT Code	schema.org/DataType	Data Source
Inverter AC power	KWT	schema.org/Number	SunFarm
Inverter DC power	KWT	schema.org/Number	SunFarm
Inverter AC energy	KWH	schema.org/Number	SunFarm
Module temperature	CEL	schema.org/Number	SunFarm
Date of measurement	DAY	schema.org/Date	SolarGIS
Time of measurement	MIN	schema.org/DateTime	SolarGIS
Global horizontal irradiance	D54	schema.org/Number	SolarGIS
Direct normal irradiance	D54	schema.org/Number	SolarGIS
Diffuse horizontal irradiance	D54	schema.org/Number	SolarGIS
Sun altitude angle	DD	schema.org/Number	SolarGIS
Sun azimuth angle	DD	schema.org/Number	SolarGIS
Air temperature at 2m	CEL	schema.org/Number	SolarGIS
Atmospheric pressure	HPA	schema.org/Number	SolarGIS
Relative humidity	-	schema.org/Number	SolarGIS
Wind speed at 10m	MTS	schema.org/Number	SolarGIS
Wind direction at 10m	DD	schema.org/Number	SolarGIS
Precipitable water	28	schema.org/Number	SolarGIS

TABLE I: Common variables collected via inverters and SolarGIS and there units in UN/CEFACT common code and data types reference for schema.org

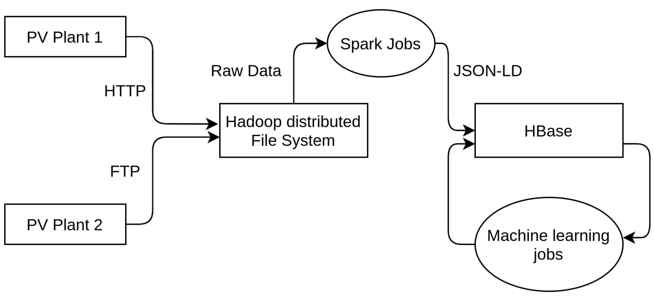

Fig. 3: The software architecture of the FAIRified PV data, metadata and modeling pipelines are shown.

'ecradle' table is a collection of packed cells, with each cell holding a single month's worth of data as a comma separated list. The proprietary, or sensitive, metadata information are stored separately in REDCap.

The version controlled documentation and metadata descriptions are maintained in a Git source versioning repository, hosted on Bitbucket. The final versions of the documentation and descriptions for FAIRified Open datasets are published to the CWRU-SDLE Open Science Framework (OSF) [23] site of our SDLE Research Center [24]. OSF is an open source web application supporting research collaborations.

OSF indexes the uploaded documentation and metadata descriptions so that they are human and machine searchable. Although open data are freely available on OSF, for proprietary data, a researcher may identify the required data from the descriptions of metadata on OSF and submit a direct request for access to the SDLE research center. The infrastructure of the FAIRified PV data are built to support such secure access to proprietary data through enabling REST API access to HBase via a Amazon Web Services (AWS) or Google Cloud Platform API gateway.

C. FAIR Metadata Schema

The semantic models were defined after the data collection process. Solar plants, inverters, modules, PV data, and weather

```
{
  "@context": "https://www.schema.org",
  "@type": "Product",
  "brand": "Enphase",
  "name": "Enphase IQ 7A",
  "category": "Microinverter",
  "model": "IQ 7A ",
  "additionalProperty": [
      {
      "@type": "PropertyValue",
      "name": "Inverter Curtailment",
      "unitCode": "KWT",
      "value": "0.3"
      },
      {
      "@type": "PropertyValue",
      "name": "Number of modules Connected",
      "value": "1"
      }
    ],
  "description": "The high-powered smart grid-ready
      Enphase IQ 7A MicroTM achieving the highest
      system efficiency for systems with 60-cell and
       72-cell modules."
}
```

Fig. 4: JSON-LD code snippet for the RDF representation of the inverter 'Enphase IQ 7A'. Details of the inverter such as its type, brand, model etc. are included in the RDF representation so as to remove any ambiguity regarding the identity of the inverter.

data were all identified as model entities. We analyzed the attributes and metadata for each of the identified model entities, and came up with a linkable semantic model. Fig. 1 depicts the final semantic model. The implementation of this logical model was realized with the use of linked data technologies and the Semantic Web. Fig. 4, 6 and 7 shows the resultant JSON-LD/XML-RDF schema files.

The machine readable metadata schema of the FAIRified raw and forecasted PV data is in XML-RDF and JSON-LD [25] formats. The files are generated according to W3C RDF 1.1 metadata schema documentation [26]. The preferred

```
{
  "@context": "https://www.schema.org",
  "@type": "Product",
  "brand": "Canadian Solar",
  "name": "Canadian Solar CS6K-300MS",
  "category": "Solar Panel",
  "model": "  CS6K-300MS-T4",
  "additionalProperty": [
      {
      "@type": "PropertyValue",
      "name": "Technology",
      "value": "PERC"
      },
      {
      "@type": "PropertyValue",
      "name": "Cell",
      "value": "Mono-Si"
      },
      {
      "@type": "PropertyValue",
      "name": "Serial Number",
      "value": "11711481451496"
      },
      {
      "@type": "PropertyValue",
      "name": "Module Wattage",
      "unitCode": "KWT",
      "value": "0.3"
      }
    ],
  "description": "The 300 watt CS6K-300MS solar
      panel features efficient PERC solar cells to
      significantly improve power performance in
      morning, evening and other low light
      conditions."
}
```

Fig. 5: Example JSON-LD code snippet for the RDF representation of a solar module.

```
{
  "@context": "https://www.schema.org",
  "@type": "Dataset",
  "measurementTechnique": "5 min interval time
      series",
  "variableMeasured": [
      {
      "@type": "PropertyValue",
      "name": "Inverter AC power",
      "unitCode": "KWT"
      },
      {
      "@type": "PropertyValue",
      "name": "Inverter DC power",
      "unitCode": "KWT"
      },
      {
      "@type": "PropertyValue",
      "name": "Module temperature",
      "unitCode": "C"
      },
      {
      "@type": "PropertyValue",
      "name": "Inverter AC energy",
      "unitCode": "KWH"
      }
    ],
  "description": "Time Series data collected from
      Inverter"
}
```

Fig. 6: Example JSON-LD code snippet for the RDF representation of the data collected via inverters. The data format is as per the standards in schema.org/Dataset.

format is the lighter weight JSON-ld. It is well designed for use with REST API services, and has also been demonstrated as serving well for graph type datasets [27], [28].

Fig. 7 show the JSON-LD schema and the RDF/XML schema for the SDLE SunFarm power plant.

IV. Analysis

The conditions for each FAIR principle is satisfied in this study as detailed below.

Findability of the dataset is enabled through registering the metadata on OSF, with a global unique persistent identifier assigned to the metadata.

The researchers associated with the research likewise have global unique persistent identifiers through their Open Researcher and Contributor ID (ORCID). Additionally, the rich metadata of the PV data are descriptive for the benefit of human readers. Further, machine readability of the metadata is achieved through the use of W3C RDF 1.1 schema.

The metadata registered on OSF are accessible to researchers via HTTPS protocol, which is open, free, and universally implemented. The proprietary PV data are accessible via the same HTTPS protocol using the SDLE AWS API gateway. Authentication and authorization is also implemented in SDLE AWS gateway to ensure secure access to proprietary data. In addition, registering on OSF also guarantees long term accessibility to the PV metadata.

In order to enhance interoperability of PV data, metadata descriptions utilize the PV vocabulary from the Department of Energy [29]. References to other metadata in the metadata descriptions are cited and linked via persistent identifiers.

Also, the metadata are registered on OSF with the Open Database License (ODbL) [30] usage license to ensure the reusability of PV data. Reusability is further improved by the inclusion of detailed provenance in metadata.

V. Discussion

This study introduces the application of FAIR principles to PV time series data. The gathering and management of PV data is a resource intensive process. However, by applying FAIR principles, the gathered PV data become reusable across multiple studies. FAIRification also ensures the preservation of the gathered PV data.

The PV data subjected to FAIRification in this study is from [6] which introduces a state-of-the art machine learning model for PV data forecasting. The FAIRification of raw and forecasted data from [6] makes it very convenient for continuous research into the proposed machine learning model and also comparison with other models of forecasting.

According to FAIR principles, community standard vocabulary, community specific data repository, and community

```json
{
  "@context": {
    "name": "http://schema.org/name",
    "description": "http://schema.org/description",
    "image": {
      "@id": "http://schema.org/image",
      "@type": "@id"
    },
    "geo": "http://schema.org/geo",
    "latitude": {
      "@id": "http://schema.org/latitude",
      "@type": "xsd:float"
    },
    "longitude": {
      "@id": "http://schema.org/longitude",
      "@type": "xsd:float"
    },
    "xsd": "http://www.w3.org/2001/XMLSchema"
  },
  "name": "SDLE SunFarm",
  "description": "The SDLE SunFarm provides extensive
      outdoor exposure capabilities, including
     fourteen Opel SF-20 dual axis trackers for
     samples and modules, with a capacity of more
     than 15,000 samples at 1-5x concentration,
     along with racking for fixed -mount modules.",
  "image": "https://engineering.case.edu/centers/sdle
     /node/84",
  "geo": {
    "latitude": "41.510032",
    "longitude": "-81.616433"
  }
}
```


(a) JSON-LD schema


```xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
    syntax-ns#" xmlns:schema="http://schema.org/">
  <rdf:Description rdf:nodeID="
      Ne3c9a06c9f5c499390be77d346fb6691">
    <schema:description>The SDLE SunFarm provides
        extensive outdoor exposure capabilities,
        including fourteen Opel SF-20 dual axis
        trackers for samples and modules, with a
        capacity of more than 15,000 samples at 1-5
        x concentration, along with racking for
        fixed -mount modules.</schema:description>
    <schema:name>SDLE SunFarm</schema:name>
    <schema:geo rdf:nodeID="
        Nd3b118f42524412ea4bb185cc695b726" />
    <schema:image rdf:resource="https://engineering
        .case.edu/centers/sdle/node/84" />
  </rdf:Description>
  <rdf:Description rdf:nodeID="
      Nd3b118f42524412ea4bb185cc695b726">
    <schema:longitude rdf:datatype="http://www.w3.
        org/2001/XMLSchemafloat">-81.616433</schema
        :longitude>
    <schema:latitude rdf:datatype="http://www.w3.
        org/2001/XMLSchemafloat">41.510032</schema:
        latitude>
  </rdf:Description>
</rdf:RDF>
```


(b) RDF/XML schema

Fig. 7: The JSON-LD and XML RDF schema for SDLE SunFarm power plant. The JSON-LD is 876 characters while the XML is 1024 characters in length, for the same metadata information.

standards for metadata schema must be employed. This emphasizes the need for standardized vocabulary, data repository, and metadata schema in the PV research community, an effort we have initiated. The lack of such agreed upon standardization is an obstacle to reaping the maximum benefit from the application of FAIR principles.

VI. Conclusion

We present the application of FAIR principles to PV time series data which enhances the findability, accessibility, interoperability, and reusability of the data within the broader PV research community. The FAIRification of the PV data results in a data model in which the metadata and the PV data are separated and a unique alphanumeric row-key string is utilized to map them. The FAIRified metadata and PV data are physically represented by two data tables in HBase. Additionally, the documentations and metadata descriptions are registered in OSF which indexes and renders them human and machine searchable. The infrastructure of the FAIRified PV data are built to support secure access to proprietary data so that a researcher may identify the required data from the descriptions of metadata on OSF and submit a request for access. Further, the machine readable metadata schema of the FAIRified PV data adheres to W3C RDF 1.1 metadata schema documentation.

References

[1] S. Bechhofer, D. De Roure *et al.*, "Research Objects: Towards Exchange and Reuse of Digital Knowledge," *Nature Precedings*, pp. 1–1, 2010-07. [Online]. Available: https://www.nature.com/articles/npre.2010.4626.1

[2] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, pp. 1–9, 2016-03-15. [Online]. Available: https://www.nature.com/articles/sdata201618

[3] M. D. Wilkinson, M. Dumontier, I. Jan Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, Jan Velterop, A. Waagmeester, P. Wittenburg,

K. Wolstencroft, J. Zhao, and B. Mons, "Addendum: The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 6, no. 1, p. 6, 2019-03-19. [Online]. Available: https://www.nature.com/articles/s41597-019-0009-6

[4] R. D. Peng, "Reproducible Research in Computational Science," *Science*, vol. 334, no. 6060, pp. 1226–1227, 2011-12-02. [Online]. Available: https://science.sciencemag.org/content/334/6060/1226

[5] The Statistical Crisis in Science. [Online]. Available: https://bibbase.org/network/publication/gelman-loken-thestatisticalcrisisinscience

[6] A. M. Karimi, Y. Wu, M. Koyuturk, and R. H. French, "Spatiotemporal Graph Neural Network for Performance Prediction of Photovoltaic Power Systems," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 17, pp. 15 323–15 330, 2021-05-18. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/17799

[7] C. Pommier, C. Michotey, G. Cornut, P. Roumet, E. Duchêne, R. Flores, A. Lebreton, M. Alaux, S. Durand, E. Kimmel, T. Letellier, G. Merceron, M. Laine, C. Guerche, M. Loaec, D. Steinbach, M. A. Laporte, E. Arnaud, H. Quesneville, and A. F. Adam-Blondon, "Applying FAIR Principles to Plant Phenotypic Data Management in GnpIS," *Plant Phenomics*, vol. 2019, pp. 1–15, 2019-04-30. [Online]. Available: https://spj.sciencemag.org/journals/plantphenomics/2019/1671403/

[8] R. Devarakonda, G. Prakash, K. Guntupally, and J. Kumar, "Big Federal Data Centers Implementing FAIR Data Principles: ARM Data Center Example," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019-12, pp. 6033–6036.

[9] G. Mayer, W. Müller, K. Schork, J. Uszkoreit, A. Weidemann, U. Wittig, M. Rey, C. Quast, J. Felden, F. Glöckner, M. Lange, D. Arend, S. Beier, A. Junker, U. Scholz, D. Schüler, H. Kestler, D. Wibberg, A. Pühler, and M. Turewicz, "Implementing FAIR data management within the German Network for Bioinformatics Infrastructure (de.NBI) exemplified by selected use cases," *Briefings in bioinformatics*, 2021-02-16.

[10] A. A. Sinaci, F. J. Núñez-Benjumea, M. Gencturk, M.-L. Jauer, T. Deserno, C. Chronaki, G. Cangioli, C. Cavero-Barca, J. M. Rodríguez-Pérez, M. M. Pérez-Pérez, G. B. L. Erturkmen, T. Hernández-Pérez, E. Méndez-Rodríguez, and C. L. Parra-Calderón, "From Raw Data to FAIR Data: The FAIRification Workflow for Health Research," *Methods of Information in Medicine*, vol. 59, pp. e21–e32, 2020-06. [Online]. Available: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0040-1713684

[11] J. Schaaf, D. Kadioglu, J. Goebel, C.-A. Behrendt, M. Roos, D. van Enckevort, &#220, F. Ckert, F. Sadiku, T. O. F. Wagner, and H. Storf, "OSSE Goes FAIR – Implementation of the FAIR Data Principles for an Open-Source Registry for Rare Diseases," *German Medical Data Sciences: A Learning Healthcare System*, pp. 209–213, 2018. [Online]. Available: https://ebooks.iospress.nl/doi/10.3233/978-1-61499-896-9-209

[12] R. Castro, "Data-driven PV modules modelling: Comparison between equivalent electric circuit and artificial intelligence based models," *Sustainable Energy Technologies and Assessments*, vol. 30, pp. 230–238, 2018-12-01. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213138818301929

[13] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, and Q. Xia, "Data-Driven Probabilistic Net Load Forecasting With High Penetration of Behind-the-Meter PV," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3255–3264, 2018-05.

[14] N. Bashir, D. Chen, D. Irwin, and P. Shenoy, "Solar-TK: A Data-Driven Toolkit for Solar PV Performance Modeling and Forecasting," in *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 2019-11, pp. 456–466.

[15] A. Skoczek, T. Cebecauer, and M. Súri, "SolarGIS: Solar Data and Online Applications for PV Planning and Performance Assessment," 2011.

[16] T. Cebecauer and M. Suri, "Site-adaptation of satellite-based DNI and GHI time series: Overview and SolarGIS approach," *AIP Conference Proceedings*, vol. 1734, no. 1, p. 150002, 2016-05-31. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.4949234

[17] Y. Hu, V. Y. Gunapati, P. Zhao, D. Gordon, N. R. Wheeler, M. A. Hossain, T. J. Peshek, L. S. Bruckman, G. Zhang, and R. H. French, "A Nonrelational Data Warehouse for the Analysis of Field and Laboratory Data From Multiple Heterogeneous Photovoltaic Test Sites," *IEEE Journal of Photovoltaics*, vol. 7, no. 1, pp. 230–236, 2017-01.

[18] A. Khalilnejad, A. M. Karimi, S. Kamath, R. Haddadian, R. H. French, and A. R. Abramson, "Automated Pipeline Framework for Processing of Large-Scale Building Energy Time Series Data," *PLOS ONE*, vol. 15, no. 12, p. e0240461, 2020-12-01. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240461

[19] Z. Ma, M. A. M. Capretz, and L. Yan, "Storing Massive Resource Description Framework (RDF) Data: A Survey," *The Knowledge Engineering Review*, vol. 31, no. 4, pp. 391–413, 2016-09.

[20] M. Banane, A. Belangour, and E. H. Labriji, "RDF Data Management Systems Based on NoSQL Databases: A Comparative Study," *International Journal of Computer Trends and Technology*, vol. 58, no. 2, pp. 98–102, 2018-04-25. [Online]. Available: http://www.ijcttjournal.org/archives/ijctt-v58p117

[21] P. A. Harris, R. Taylor *et al.*, "Research Electronic Data Capture (REDCap)–a Metadata-Driven Methodology and Workflow Process for Providing Translational Research Informatics Support," *Journal of Biomedical Informatics*, vol. 42, no. 2, pp. 377–381, 2009-04.

[22] C. Percival and S. Josefsson, "The scrypt Password-Based Key Derivation Function," 2016-05-18. [Online]. Available: https://www.hjp.at/doc/rfc/rfc7914.html

[23] E. D. Foster and A. Deardorff, "Open Science Framework (OSF)," *Journal of the Medical Library Association : JMLA*, vol. 105, no. 2, pp. 203–206, 2017-04. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5370619/

[24] R. H. French, A. Curran *et al.*, *CWRU SDLE Research Center*. Case Western Reserve University, 2019-03.

[25] M. Sporny, D. Longley, G. Kellogg, M. Lanthaler, and N. Lindström, "JSON-LD 1.0," *W3C recommendation*, vol. 16, p. 41, 2014.

[26] D. Tomaszuk and D. Hyland-Wood, "RDF 1.1: Knowledge Representation and Data Integration Language for the Web," *Symmetry*, vol. 12, no. 1, p. 84, 2020-01. [Online]. Available: http://dx.doi.org/10.3390/sym12010084

[27] M. Lanthaler and C. Gütl, "On using JSON-LD to create evolvable RESTful services," in *Proceedings of the Third International Workshop on RESTful Design*, ser. WS-REST '12. Association for Computing Machinery, 2012-04-17, pp. 25–32. [Online]. Available: https://doi.org/10.1145/2307819.2307827

[28] J. Wang, A. Aryani, L. Wyborn, and B. Evans, "Providing Research Graph Data in JSON-LD Using Schema.org," in *Proceedings of the 26th International Conference on World Wide Web Companion*, ser. WWW '17 Companion. International World Wide Web Conferences Steering Committee, 2017-04-03, pp. 1213–1218. [Online]. Available: https://doi.org/10.1145/3041021.3053052

[29] Solar Energy Glossary. Energy.gov. [Online]. Available: https://www.energy.gov/eere/solar/solar-energy-glossary

[30] "Open Data Commons Open Database License (ODbL) v 1.0 Open Data Commons," 2018-07. [Online]. Available: http://opendatacommons.org/licenses/odbl/