

Real-world Application of MDS-Onto: Ontology-driven Knowledge Graphs to Accelerate Discovery in Materials Science

Erika I. Barcelos^{1,3}, Ozan Dernek^{1,3}, Ethan D. Frakes^{3,4}, Rishabh Kundu^{1,3}, Van D. Tran^{1,3}, Holly L. Schreiber^{2,3}, Redad Mehdi^{1,3}, Shawn Liang^{2,3}, Quynh D. Tran^{2,3}, Yinghui Wu^{2,3}, Mengjie Li^{3,4}, Laura S. Bruckman^{1,3} and Roger H. French^{1,2,3,*}

¹Department of Materials Science and Engineering, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH, USA

²Department of Computer and Data Sciences, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH, USA

³Materials Data Science for Stockpile Stewardship Center of Excellence, Case Western Reserve University, 10900 Euclid Avenue, Ohio, USA

⁴College of Engineering and Computer Science, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL, USA

Abstract

Ontology-driven knowledge graphs (KGs) offer a powerful framework for accelerating materials discovery through Symbolic AI. By pairing formal logic engines with structured data rules, researchers can establish a robust semantic backbone that ensures data quality, interoperability, and logical consistency. When integrated with probabilistic, data-driven approaches, it can accelerate scientific insights and discoveries. Unlike purely data-driven knowledge graphs, ontology-driven KGs support advanced reasoning capabilities. This allows researchers to uncover new insights while ensuring full logical consistency with existing domain knowledge. In this work, we introduce MDS-Onto Open, an open-source, community-driven effort to generate linked data and KGs based on BFO-compliant ontologies. We demonstrate the practical utility of MDS-Onto tools through two distinct case studies: degradation science using historical datasets, and data quality assessment of photovoltaics time series data.

Keywords

MDS-Onto, Ontology-driven Knowledge Graphs, Materials Discovery

1. Introduction

In data science and semantic or knowledge engineering fields, an ontology is defined as a formal data blueprint that models a specific domain of knowledge. By establishing a standardized hierarchy of classes, properties, and constraints, ontologies provide important contextual information in a field that leads to machine-readable and understandable data that are "AI-ready", i.e. are highly optimized for AI applications. More recently, with advances in ISO standardization of the Basic Formal Ontology [1] as a Top-Level Ontology and associated W3C recommendations, ontologies have expanded beyond single domains to form broader interoperable systems. These interoperable ontologies have domain or application ontologies that connect with Low-level, Mid-level, and Top-level ontologies, which then connect to schema.org of the W3C. Therefore, these are not the isolated "core" ontologies developed since the introduction of FAIR[2, 3] in 2016, but instead are a new effort to integrate scientific ontologies

SeMatS 2026: The 3rd International Workshop on Semantic Materials Science co-located with the 25th International Semantic Web Conference (ISWC 2026), October 25th, 2026, Bari, Italy

*Corresponding author.

✉ eib14@case.edu (E. I. Barcelos); oxd37@case.edu (O. Dernek); ethan.frakes@ucf.edu (E. D. Frakes); rxk857@case.edu (R. Kundu); vxt101@case.edu (V. D. Tran); hls92@case.edu (H. L. Schreiber); mxm1684@case.edu (R. Mehdi); sgl57@case.edu (S. Liang); qdt@case.edu (Q. D. Tran); yxw1650@case.edu (Y. Wu); mengjie.li@ucf.edu (M. Li); lsh41@case.edu (L. S. Bruckman); rxf131@case.edu (R. H. French)

ORCID 0000-0002-9273-8488 (E. I. Barcelos); 0000-0002-5071-9684 (O. Dernek); 0009-0008-8869-5703 (E. D. Frakes); 0000-0002-0719-3939 (R. Kundu); 0000-0002-0877-7063 (V. D. Tran); 0009-0004-6150-3223 (H. L. Schreiber); 0000-0002-0593-5222 (R. Mehdi); 0000-0003-3991-5155 (Y. Wu); 0000-0003-1271-1072 (L. S. Bruckman); 0000-0002-6162-0532 (R. H. French)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

generally into the broader World Wide Web. Furthermore, when computing systems process ontologically grounded data, they can natively enforce data quality, minimize semantic ambiguity, and leverage logical reasoning engines to infer new facts and uncover discoveries.

In materials science, ontologies establish a standardized knowledge framework and shared vocabulary necessary for systems to communicate and exchange data seamlessly. This guarantees long-term interoperability, robust data management processes, and conformity with the principles of FAIR .

Beyond improving data and metadata management, ontologies in materials science are essential for enabling automated reasoning. By integrating data into an ontologically grounded framework to construct a KG, researchers can leverage this logical foundation to deduce new knowledge from the existing information base. When used as the backbone of KGs, ontologies provide a powerful framework for data quality control and error prevention. Their strict logical rules and structural constraints automatically flag inconsistencies or anomalies that conflict with domain design parameters. This structural integrity in KGs is vital for generating FAIR linked data, ensuring that AI systems operate robustly and efficiently.

Similarly to ontologies, KGs organize information in a graph-based structure where concepts are interconnected through explicitly defined relationships. Although both are represented as graphs, they work at different levels of abstraction: ontologies capture the conceptual schema and structural rules, while knowledge graphs are created and populated with actual data, facts, and factual statements (Resource Description Framework (RDF) [4] triples).

There are two main strategies for constructing KGs from data and results: **data-driven** or **ontology-driven**. Data-driven approaches are based on employing natural language processing, such as Large Language Models(LLM) or textual analysis, to transform raw data into a graph structure. Ontologically-driven KGs, on the other hand, have a stricter architecture where the ontology sets rules and constraints to guide the creation of the graph that obeys and follows the rules defined by the ontologies. Although data-driven KGs are highly adaptable and flexible, they lack formal semantic context. Consequently, they do not support automated reasoning, error detection, consistency checking, or logical inference.

In this work, we introduce MDS-Onto, an application-oriented open-source project to create KGs and RDF linked data(JSON-LD) grounded in the MDS-Onto Ontology [5] using open-source packages part of the MDS-Onto Open Ecosystem. While existing platforms exist to support the creation of linked data such as OpenRefine [6] and Protege [7] and open-source packages such as LinkML[8] and Morph-KGC [9] they are largely domain agnostic and require separate mapping files. MDS-Onto Open, in contrast, natively captures experimental and computational metadata, code execution, data transformations, and analysis scripts directly into JSON-LD, offering built-in integration with materials science concepts and physical units.

We introduce the tools part of our framework and showcase its versatility and broad applicability by employing two exemplar cases that illustrate they work together to create linked data, FAIRify metadata, and perform data analysis from raw research data. The first case focuses on a degradation science study of acrylic under accelerated weathering conditions. The second case study illustrates the usage of the tools using domain specific (photovoltaics) time-series data.

2. MDS-Onto Open

MDS-Onto Open is a community-driven effort focused on engaging and training the scientific community to promote FAIRification of data, metadata and results driven by semantic ontologies and RDF Linked Data(JSON-LD). The current MDS-Onto Open Framework includes ontologies, RDF FAIR-linked Data, Knowledge Graphs (KGs), and Semantic Search across Materials Science. MDS-Onto Open has several ongoing partnerships and collaborations, including industry, national laboratories, and universities that combine their efforts to expand MDS-Onto by creating domain or subdomain ontologies, linked data, and KGs in the field of Materials Data Science.

The open-source packages of the MDS-Onto Open ecosystem represent practical solutions for the creation, visualization, and assessment of ontologies and RDF FAIRlinked data and KGs in the Materials

Data Science domains. MDS-Onto Open is a robust end-to-end framework focused on real-world applications of scientific studies that converts raw experimental, computational and analysis data into Linked Data. By directly mapping domain-specific variables to standardized upper ontologies, it delivers fully interoperable metadata without requiring manual semantic alignment. Figure 1 illustrates the main tools of MDS-Onto Open and the workflow to create RDF FAIR Linked data.

2.1. MDS-Onto Tools

TTL-PAWIKAN and FAIRMaterials TTL-PAWIKAN (Parsing and Writing Intelligent Knowledge Assertion Notations)[10] is a VSCode extension available on the Open VSX Registry that provides real-time syntax highlighting and error validation for `.ttl` files. FAIRMaterials [11] is an R package designed to simplify the creation of ontologies for non-experts. Key functionalities include the creation of `.ttl` files and ontologies directly from `.csv` as well as assessing compliance with MDS-Onto design rules.

CEMENTO. CEMENTO is a Python package used for the creation of ontologies based on draw.io diagrams [12]. When designing ontologies, users typically start with a schema of their domain, oriented in the context of the current defined terms and classes of MDS-Onto, to organize ideas, hierarchical concepts and relationships. CEMENTO takes advantage of the existing diagrams to create ontologies with interoperable terms mapped to the MDS-Onto ontology. By default, CEMENTO uses CCO [13], and BFO-compliant ontologies and W3C vocabularies (SKOS, RDFS) as references, which facilitates and streamlines term mapping.

FAIRlinked. FAIRlinked [14, 15, 16] is a Python package designed to create FAIR linked data (JSON-LD [17]) with interoperable terms and JSON-LD files to be findable, accessible, and reusable. FAIRlinked takes a dataframe corresponding to the user data/metadata, identifies mappings and corresponding labels to MDS-Onto Ontology, and produces linked data with parseable and identifiable file names. FAIRlinked requires input data to be structured as a pandas dataframe. Consequently, files in other formats must be imported and appropriately transformed into a dataframe object prior to executing of the functions. The output file includes both the user-defined variables and the corresponding MDS-Onto Ontology term. In practice, FAIRlinked will read the header of the dataframe, map against MDS-Onto classes and replace the dataframe term to the corresponding MDS-Onto class while preserving the original name under `skos:altlabel`. If an exact match is not found, FAIRlinked suggests the closest related term, allowing users to review and approve mappings before export to JSON-LD.

In addition to mapping classes, properties(data and object) are also included in the JSON-LDs. Users can either manually introduce the property and their equivalent domain and range pairs as function arguments or FAIRlinked will scan MDS-Onto and find for the property that matches the domain and range of the user dataframe. If an exact class match is unavailable, it selects the closest related term. Similar to classes, users can evaluate and approve all proposed mappings before JSON-LD conversion. Properties should be reused from OBO ontologies and CCO whenever available. According to MDS-Onto design rules, new properties may only be created if no existing match can be found. Finally, additional axioms and restrictions are derived from the ontology upon loading the JSON-LD.

In many domains, linked data is implemented primarily for data management rather than reasoning. For these use cases, lightweight vocabularies like SKOS are preferred over full OWL ontologies. Conversely, domains that require logical reasoning can incorporate formal axioms into their ontologies and deploy them to a semantic graph database for inference along with the JSON-LDs. FAIRlinked also includes a functionality for deserializing the linked data (JSON-LDs) back into `.csv`, Arrow dataframes, and parquet `.parquet` datafiles. In addition to metadata, FAIRlinked also enables FAIRification of data analysis using a wrapper function that allows the user to track their analysis, recording the metadata of their function executions during the analysis process. This includes, but is not limited to, input and output, package imports, type of data structure, function execution, and file read/write events. In addition to JSON-LD, the wrapper generates a report as a human-readable `.md` version of the metadata.

OntoCheck. Although several tools are available to assess linked and FAIR data [18], ontology-specific evaluation lacks support for broadly adopted standards, and many existing approaches are incomplete or no longer actively maintained. OntoCheck [19] is an open-source Python package developed to provide a unified framework for ontology evaluation by combining domain-independent structural analysis with a query-driven assessment methodology. The framework is designed to support ontology developers in identifying structural deficiencies while also enabling query-specific evaluation for an ontology’s suitability without requiring extensive expertise in formal knowledge representation.

Unlike conventional ontology assessment tools, OntoCheck integrates 17 domain-agnostic evaluation metrics that assess labeling quality, structural consistency, accessibility compliance, and naming conventions. Compliance with BFO terms is also an available metric in ontocheck. In addition, it incorporates a task-oriented evaluation mode that quantifies both vocabulary coverage and utilization density using SPARQL competency questions generated from natural language. The package is available through both a command-line interface and a Python API, facilitating integration into automated analysis pipelines. Its modular architecture also allows users to incorporate custom evaluation metrics and domain-specific competency questions, while producing structured summaries that support a transparent interpretation of assessment results. By continuing to expand its suite of metrics based on community needs and incorporating community contributions, OntoCheck is intended to contribute to the development of more standardized ontology evaluation practices.

MDS-Onto Ontology. The MDS-Onto Ontology is a low-level ontology designed for the field of Materials Data Science. It was developed to capture various domains and subdomains within the field, organized under seven overarching categories (MDSDom): builtenv, characterization, materials processing, geospatial, exposure, general, and chemistry. To ensure semantic interoperability, all the concepts and properties of MDS-Onto in all domain and subdomain ontologies are aligned with the Common Core Ontologies (CCO)[20], BFO, and other BFO-compliant ontologies[21]. As a low level ontology, MDS-Onto has terms that are more domain specific and closely related to scientists’ actual captured data during a scientific study. In level of specificity, it sits below PMDCo[22] and EMMO [23]. Within each overarching domain, subdomains are established to encapsulate granular terms and concepts. For example, the mdsdom:Characterization domain branches in subdomains such as mdsdom:SEM, mdsdom:AFM, mdsdom:FTIR and mdsdom:UV-Vis. The branching choice is based on domain-specific decisions of relevance for that community. For instance, in Photovoltaics, the subontologies are split into PV-Site, PV-Module and PV-Cell. In characterization they are split according to techniques and tools. MDS-Onto is mainly designed to facilitate practical data FAIRification and automated dataframe/csv serialization to linked data and domain and subdomains are flexible to be changed in the future according to community needs and feedback.

MDSDom is a domain taxonomy in which a hierarchical structure is used to classify and organize material science terms into their domain and subdomain categories. Each domain is defined as a class and the subdomains are defined as a subclass of their domain. Individual terms defined in MDS-Onto are connected to their most applicable subdomain(s) using the annotation property `mds:inDomain`. This relationship allows a term to inherit the broader domain hierarchy without requiring contributors to manually specify every level of the domain of a term.

Study Stages. The study stages are a conceptual framework to structure a scientific study in a more granular way to ensure that metadata is captured comprehensively and properly throughout the study. This modular solution was chosen to enhance the long-term maintainability of the Ontologies. Each “stage” corresponds to a step and a component of the research where metadata management is essential. Developing a single ontology to cover the entire lifecycle of a scientific study, for example, in a study with formulation, exposure, multiple characterization modalities, data analysis, and modeling steps creates an overly complex Ontology that is difficult to govern, maintain, and query.

The number of these stages varies across projects and are highly dependent on the research group’s specific workflow and study design choices. Studies with limited metadata may choose to consolidate

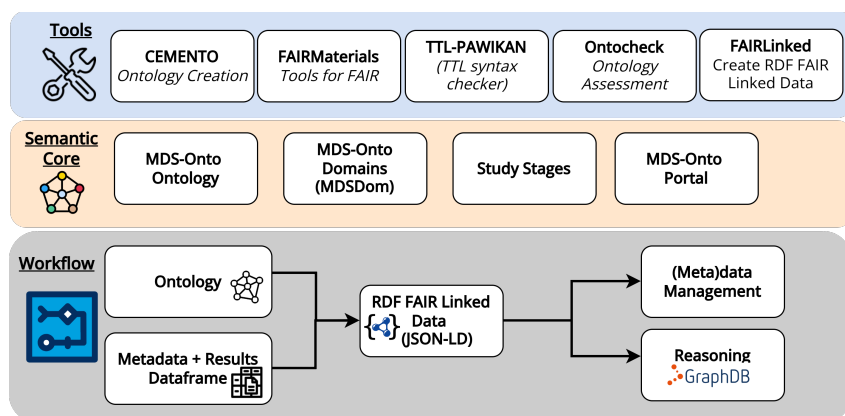


Figure 1: MDS-Onto workflow and tools. In a first step, users define their domain and create ontology terms(if not present on MDS-Onto) for different study stages in their research.

multiple sequential steps into a single phase. For example, if a sample undergoes several distinct materials processing treatments, they can be grouped into a single, comprehensive “Material Processing” stage or isolated into independent stages, such as the “Grinding” and “Polishing” study stage. This depends on the level of granularity preferred by the scientist and their study.

3. FAIRifying data using MDS-Onto Tools

Designing a scientific study typically begins with a research question linked to a hypothesis that a scientist aims to validate. This hypothesis is subsequently tested through a series of experimental or computational methodologies designed to investigate the problem. Once these procedures are executed, the resulting data are analyzed, conclusions are drawn, and the initial hypothesis is validated or not. If the results are not satisfactory, a subsequent interaction is started.

In this traditional workflow, data management is highly centralized entirely around the individual scientist, who often works independently without using standardized practices for recording and labeling data or metadata. Data analysis and interpretation are often localized and manually performed, remaining under the scientist until the polished results are shared and ready to be shared with the broader community. The fundamental limitation of this approach is its heavy reliance on the individual researcher to store data, track metadata, and execute analyses, which severely hinders the longevity and reproducibility of the data.

Beyond lab studies, scientific investigations can generate data that varies significantly in scale and complexity, from small in-lab dataset to massive data such as spatiotemporal data involved in in-situ characterization of 3D-printed components, or time-series data in a PV Power Plants. Alternatively, other studies can use open-source datasets, where scientists leverage existing data to answer a research question. Regardless of the data source or the study design, the main challenges remain consistent, as data management practices are poorly implemented, hindering the long-term utility of the data.

To address these limitations, we propose the concept of ontology-driven study protocols grounded in MDS-Onto, where Ontologies govern every stage of the scientific workflow *study stages* to prioritize reproducibility. Within this framework, both data and metadata are systematically captured throughout the research lifecycle and transformed into structured linked data.

Workflow on MDS-Onto. The FAIRification journey starts with the user defining their subdomain, followed by identifying the required study stages and ontologies necessary to comprehensively capture the relevant metadata in that domain. Next, they will use CEMENTO, Protege [24, 25] or their preferred ontology editor to create the ontology file(.ttl). In this process, users should use the MDS-Onto website and pyLODE page [26] and the available documentation to ensure compliance with the MDS-Onto

design rules, which provide additional annotation properties for the subdomain and the study stages of the defined terms.

It is essential that existing terms from MDS-Onto and CCO be reused to avoid duplication. If a term is already present in MDS-Onto, it should be excluded from the new user .ttl file. However, if modifications to an existing core class are required, such as modifying definitions or assigning a new subdomain, users can submit an addition change request. Once the .ttl file is complete, containing only new terms that were systematically mapped to MDS/CCO/BFO-compliant Ontologies, users can execute OntoCheck to evaluate structural quality and MDS-Onto design rules conformance.

Next, users submit new terms and modification requests through a Tally form [27, 28] and join the project's Slack workspace[29]. The submission triggers an automatic alert in the #submit-new-mds-onto terms channel, initiating a review. If adjustments are necessary, feedback is provided directly in the channel; otherwise, the new terms are merged into the core MDS-Onto and the user is notified on Slack. Once the notification is received, users can see their terms on MDS-Onto by directly downloading from the MDS-Onto Open Website. As an alternative to form-based submissions, contributions to the MDS-Onto ontology and its packages can be made directly via GitHub. Through this pipeline, users submit pull requests containing the new or modified terms for review and integration into the MDS-Onto core ontology.

To generate JSON-LD users input their dataframes into FAIRlinked, which aligns dataset variables with formal MDS-Onto classes(preserved under skos:altLabel). Users can choose to export a single JSON-LD per dataframe or one per row, which will depend on their study needs and design. The resulting JSON-LDs can be uploaded to an RDF triplestore graph database, such as GraphDB[30] to populate the KG. In parallel, FAIRlinked analysisTracker can be run on data analysis and modeling scripts, capturing metadata, and generating the JSON-LDs and a report. All the study stage ontologies are linked by domain-defined and globally unique identifier IDs. When they are all loaded into the database, GraphDb will automatically resolve these identifiers, seamlessly linking the records to facilitate comprehensive, cross-study investigations. Subsequently, users can run SPARQL queries to explore the data and derive new knowledge via the selected reasoning engine.

4. Employing MDS-Onto Open on a Polymer Degradation Study

Acrylic(Polymethylmethacrylate, PMMA) is a versatile polymer that is widely used across industries, particularly due to its excellent optical properties. Understanding how acrylic degrades in the field is vital for many of its applications; however, real-world testing is challenging. Depending on the additives used, it can take over 100 years to break down naturally. To overcome this, researchers rely on accelerated weathering tests to study its degradation in simulated real-world conditions.

This study [31] started in 2012, where 6 formulations from 2 acrylic manufacturers containing different types of additives, including UV absorbers, antioxidants, and UV stabilizers, were used to study degradation. The samples were placed in three different accelerated chambers with varying light, irradiance, and humidity rates. Those samples were exposed to a total of 22,500 hours, where every 500 hours they were removed and characterized chemically, optically, and mechanically. The results of the characterizations were used in the modeling of network structural equations (NetSEM), a statistical learning pathway to study relationships between stressors[32, 33]. To assess degradation pathways, models were developed to analyze pairwise relationships between predictors and responses derived from characterization data.

Figure 2 illustrates the workflow of the degradation study with the respective study stages.

Ontologies were designed for all study stages using CEMENTO. Metadata relative to the sample, exposure, formulation and characterizations were recorded under MDSDom:Characterization MDSDom:Formulation and MDSDom:Exposure domains. Five study stages were considered for this study: sample & formulation, exposures, characterizations, analysis, and modeling. For all the study stages, excluding analysis and modeling, one JSON-LD was created for the respective metadata. For analysis and modeling, one JSON-LD was created for each row of the data frame, where each row

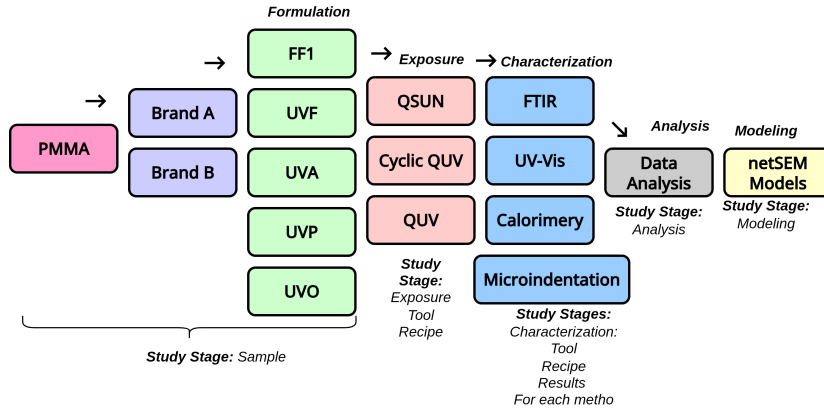


Figure 2: Scientific workflow for PMMA degradation study with their respective study stages. Sample, formulation and exposure metadata along with results and modeling ontologies were created and populated with data to form the PMMA KG.

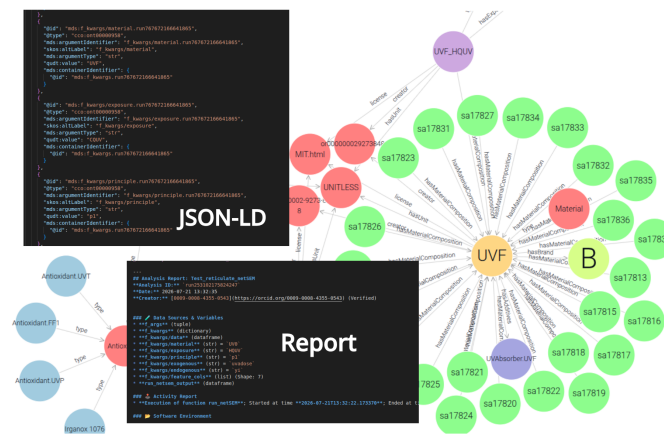


Figure 3: Screenshot of JSON-LD and report generated from a NetSEM model and an overview of PMMA KG on GraphDB.

contains the predictive variables and performance metrics for each model, which was designed for each formulation(material) at specific steps.

Once created, the ontologies were validated using OntoCheck to verify quality and ensure compliance with MDS-Onto design rules before being merged into the core Ontology. Subsequently, FAIRlinked was used to generate all JSON-LD files for the study. These files were then ingested into a GraphDB repository to facilitate SPARQL querying. In addition to the core study datasets, analysis JSON-LDs were created to capture all script parameters and relevant execution metadata for the NetSEM model. Figure 3 displays the GraphDB repository along with the report and the JSON-LD file generated for the modeling workflow. Some of the competency questions investigated with GraphDB in this study were : Which Formulation has the longest life? Which degradation mode is the harshest for the formulations? What grade of PMMA has the worst Residual Monomer degradation?

5. PV time-series data: SunSmart Schools

The next case study extends the MDS-Onto [5] framework from laboratory-scale degradation analysis to fleet-scale photovoltaic (PV) system performance analysis. Time-series data were collected through the SunSmart Schools program [34, 35], where ground-mounted, grid-tied PV and battery storage systems were installed at 30 schools across Florida to provide emergency shelters. Each PV site has three sub-arrays and is monitored at each combiner box (CBOX), producing a 15-minute interval time-series data stream for each node.

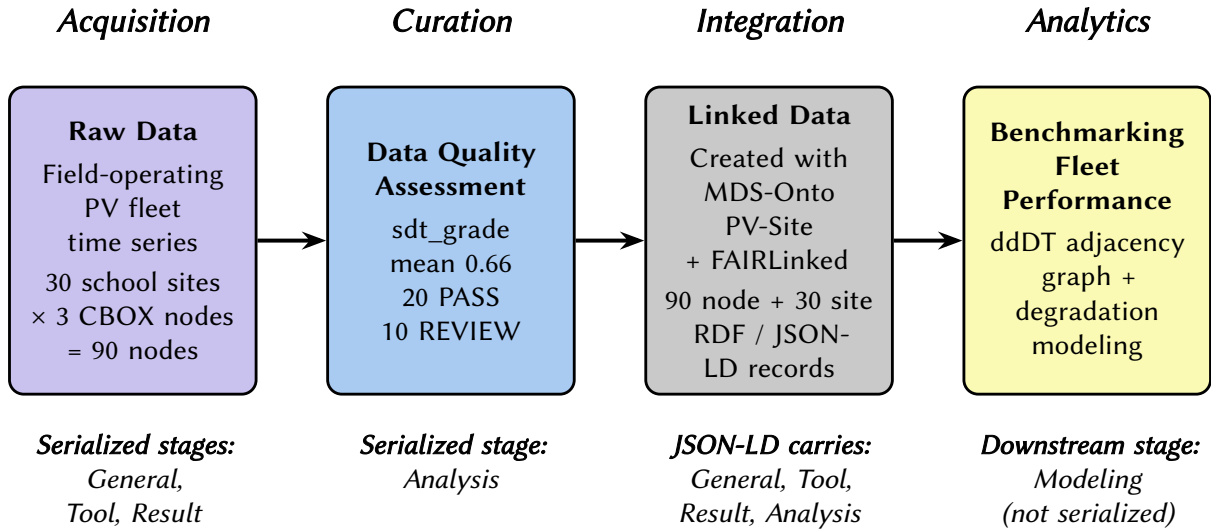


Figure 4: SunSmart Schools PV time-series workflow and its MDS-Onto study stages. Thirty Florida school sites instrumented with three CBOX nodes each yield 90 monitored power time-series streams. FAIRlinked produces 90 per-node JSON-LD records containing descriptive metadata and Solar Data Tools data-quality grades, together with 30 derived per-site quality records.

Figure 4 illustrates the workflow of the SunSmart case study with the respective study stages. The case study comprises 30 school sites with three CBOX nodes each, yielding 90 monitored nodes whose descriptive metadata (including geographic coordinates, site and node identifiers, cluster size, and a per-node data-quality grade) must be captured in a standardized, machine-actionable form before the underlying time series can be reliably analyzed or compared across the fleet. The data quality grade is provided by the Solar Data Tools [36] library, which measures the fraction of a node’s record whose days pass both a density check, requiring sufficient readings above 2% of estimated capacity, and a linearity check, which rejects days bearing the signature of linear-interpolation infill.

To semantically ground this metadata, PV system-specific classes and properties were developed in MDS-Onto within the `mdsdom:PV-Site` domain. The metadata of each PV site was mapped to an MDS-Onto class, assigned a QUDT unit, and situated within a study stage. Location information was mapped to the classes `mds:Latitude` and `mds:Longitude`, assigned unit:DEG, and annotated with the *Result* study stage. The PV-site and node identifiers were mapped to `mds:PVSite` and `mds:Inverter`, respectively; both were assigned unit:UNITLESS and annotated with the *General* and *Tool* study stages, respectively. Data quality was mapped to `mds:AnalyticalResult`, assigned unit:UNITLESS, and annotated with the *Analysis* study stage. As in the polymer degradation study, the original variable names are preserved under `skos:altLabel` and human-readable definitions under `skos:definition`, so the resulting linked data remains legible to domain scientists while being fully interoperable at the ontology level. FAIRlinked was then used to serialize the metadata into RDF linked data, employing its per-row serialization mode keyed on the node identifier so that each CBOX is emitted as a single self-contained JSON-LD record, carrying provenance through the data curator’s ORCID (`dcterms:creator`) and a dataset-level CC0 license (`dcterms:license`). Because all entities belonging to one node share a common row key, and each record carries its parent site identifier, these records support query-time joins between CBOX nodes and their school sites, enabling fleet-wide queries such as retrieving all nodes of a given site or filtering nodes by data-quality grade before downstream degradation modeling.

Beyond descriptive metadata, the SunSmart case also demonstrates that FAIRlinked can FAIRify the outputs of data analysis itself, corresponding to the Analysis study stage. The raw data was graded on its data quality and recorded in `sdt_grade`. The grade measures usable production data, whereas a high value (> 0.5) means the data passes inspection, and a low value (< 0.5) means the data is flagged for review. In the analysis, the node-level `sdt_grade` values were aggregated into a per-site quality

assessment as well. For each of the 30 school sites, the mean quality grade across its three CBOX streams was computed. We also calculated the fraction of streams with grades exceeding 0.5. In the fleet’s data-driven digital twin (ddDT) workflow [37, 38], only nodes exceeding this threshold may contribute edges to the quality-based adjacency graph. Each site was then assigned a PASS or REVIEW status. Across the fleet, the mean node grade was 0.66; 20 sites qualified on all three nodes, while 10 were flagged for review. Mirroring the PMMA study, where netSEM model outputs are serialized one row per model, this derived table was passed through the same per-row FAIRification workflow, keyed on the site identifier and typed entirely with existing MDS-Onto classes (`mds:NumberOfObservations`, `mds:AnalyticalResult`, `mds:Ratio`, `mds:ProcessingQualityStatus`), yielding one Analysis-stage JSON-LD record per school that resolves, through the shared site identifier, to the node-level metadata records in the graph. The same workflow extends directly to richer analyses of the underlying time series, whose outputs become analysis-stage linked data in exactly the same manner. This case study thus illustrates how the ontology-driven study protocol applied to performance analysis of distributed PV assets, making both time-series raw-data, site metadata and their derived analyses FAIR and ready for fleet-scale performance benchmarking.

6. Conclusion

Ontology-driven KGs provide a standardized and robust data architecture that harmonizes scientific studies including experimental and computational research data across distinct study stages. By enforcing ontologies as the foundation of KGs, researchers can preserve semantic context, ensuring long-term data interoperability and better overall study reproducibility.

In this study, we introduced MDS-Onto Open, an open-source initiative focused on providing practical tools for the materials data science community to generate RDF FAIR linked data and KGs grounded in BFO-compliant ontologies. MDS-Onto offers an accessible, low-barrier workflow for converting dataframes into linked data while seamlessly capturing provenance and metadata from data analysis and modeling pipelines. The tools part of MDS-Onto empower researchers to significantly improve data management practices and unlock reasoning, ultimately accelerating materials discovery.

Acknowledgments

This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University. This material is based on research conducted at the Materials Data Science for Stockpile Stewardship Center of Excellence (MDS3-COE). This material (suspension data) is based upon work supported by the U.S. National Science Foundation Award EEC-2052776, EEC-2052662, and EEC-2435402 in the MDS-Rely IUCRC Center, under the NSF Solicitation: NSF 20-570 Industry-University Cooperative Research Centers Program.

Declaration on Generative AI

During the preparation of this work, the author(s) used Gemini 3.5 and Claude Opus 4.8 for checking grammar and spelling, improving text readability, and creating some of the icons used in the figures.

References

- [1] International Organization for Standardization, International Electrotechnical Commission, ISO/IEC 21838-2:2021: Information technology – Top-level ontologies (TLO) – Part 2: Basic Formal Ontology (BFO), <https://www.iso.org/standard/74572.html>, 2021.
- [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. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. '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, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3 (2016). URL: <https://www.nature.com/articles/sdata201618>. doi:10.1038/sdata.2016.18.
- [3] A. Jacobsen, R. de Miranda Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, M. Courtot, M. Crosas, M. Dumontier, C. T. Evelo, C. Goble, G. Guizzardi, K. K. Hansen, A. Hasnain, K. Hettne, J. Heringa, R. W. Hooft, M. Imming, K. G. Jeffery, R. Kaliyaperumal, M. G. Kersloot, C. R. Kirkpatrick, T. Kuhn, I. Labastida, B. Magagna, P. McQuilton, N. Meyers, A. Montesanti, M. van Reisen, P. Rocca-Serra, R. Pergl, S.-A. Sansone, L. O. B. da Silva Santos, J. Schneider, G. Strawn, M. Thompson, A. Waagmeester, T. Weigel, M. D. Wilkinson, E. L. Willighagen, P. Wittenburg, M. Roos, B. Mons, E. Schultes, FAIR Principles: Interpretations and Implementation Considerations, *Data Intelligence* 2 (2020) 10–29. URL: https://doi.org/10.1162/dint_r_00024. doi:10.1162/dint_r_00024.
- [4] F. Gandon, INRIA, G. Schreiber, V. U. Amsterdam, RDF 1.1 XML syntax, 2014. URL: <https://www.w3.org/TR/rdf-syntax-grammar/>.
- [5] B. P. Rajamohan, A. C. H. Bradley, V. D. Tran, J. E. Gordon, H. W. Caldwell, R. Mehdi, G. Ponon, Q. D. Tran, O. Dernek, J. Kaltenbaugh, B. G. Pierce, R. Wieser, W. Yue, K. Lin, J. Kambo, C. Lopez, A. Nihar, D. J. Savage, D. W. Brown, H. Sharma, B. Giera, P. K. Tripathi, Y. Wu, M. Li, K. O. Davis, L. S. Bruckman, E. I. Barcelos, R. H. French, Materials Data Science Ontology (MDS-Onto): Unifying domain knowledge in materials and applied data science, *Scientific Data* 12 (2025) 628. doi:10.1038/s41597-025-04938-5.
- [6] OpenRefine, <https://openrefine.org/>, ????
- [7] Protégé, <https://protege.stanford.edu/>, ????
- [8] LinkML – Bring structure to your data., <https://linkml.io/>, ????
- [9] J. Arenas-Guerrero, D. Chaves-Fraga, J. Toledo, M. S. Pérez, O. Corcho, Morph-KGC: Scalable knowledge graph materialization with mapping partitions, *Semantic Web* 15 (2024) 1–20. doi:10.3233/SW-223135.
- [10] Gabriel Ponon, TTL-PAWIKAN – Open VSX Registry, <https://open-vsx.org/extension/CWRU-SDLE/ttl-pawikan>, ????
- [11] J. E. Gordon, A. H. Bradley, P. Rajamohan, N. Hahn, K. Lin, A. Nihar, H. Cadwell, J. Kambo, J. Jimenez, K. J. Hernandez, H. Htet-Aung, B. Giera, W. Yue, M. R. Mehdi, F. Holt, Q. Tran, G. Ponon, D. Savage, D. Brown, J. Kaltenbaugh, K. Havinal, N. Gray, M. Liggett, B. G. Pierce, R. Wieser, Y. Fan, T. Ciardi, O. J. Akanbi, H. Iawal, W. Oltjen, M. Kalutotage, A. Lino, V. Tran, M. Lu, X. Yu, L. Huang, Z. Li, A. Daundkar, H. Omodolor, D. Bhuvanagiri, M. Rasmussen, S. Nalin-Venkat, T. Wang, R. Chawla, L. Jo, J. M. Yarus, K. O. Davis, M. Li, Y. Wu, P. K. Tripathi, L. S. Bruckman, E. I. Barcelos, R. H. French, FAIRmaterials: Ontology Tools with Data FAIRification in Development, 2024.
- [12] G. Ponon, B. P. Rajamohan, E. Frakes, J. Jimenez, H. Sharma, R. French, *cwru-sdle/cemento*, US Department of Energy (DOE) Software (2026) 65.
- [13] M. Jensen, G. D. Colle, S. Kindya, C. More, A. P. Cox, J. Beverley, The Common Core Ontologies, 2024. doi:10.48550/arXiv.2404.17758. arXiv:2404.17758.
- [14] V. D. Tran, B. Lee, H. Dirks, R. Lamba, B. P. Rajamohan, G. Ponon, Q. D. Tran, O. Dernek, Y. Wu, E. I. Barcelos, R. H. French, L. S. Bruckman, FAIRLinked: A powerful tool for transforming research data into fair-compliant rdf, 2026. URL: <https://pypi.org/project/FAIRLinked/>, case Western Reserve University: SDLE Research Center.
- [15] Van D. Tran, Priyan Rajamohan, Ritika Lamba, Gabriel O. Ponon, Kai Zheng, Erika I Barcelos, Quynh D. Tran, Ozan Dernek, Laura H. Bruckman, Roger H. French, FAIRLinked: Transform research data into FAIR-compliant RDF., 2026.
- [16] V. D. Tran, B. Lee, R. Lamba, H. Dirks, Q. D. Tran, B. P. Rajamohan, O. Dernek, L. S. Bruckman, Y. Wu, E. I. Barcelos, R. H. French, FAIRlinked: Data FAIRification Tools for Materials Data Science, *Journal*

- of Open Source Software 11 (2026) 9628. URL: <https://joss.theoj.org/papers/10.21105/joss.09628>. doi:10.21105/joss.09628.
- [17] Manu Sporny, Dave Longley, Gregg Kellogg, Markus Lanthaler, Pierre-Antoine Champin, Niklas Lindström, JSON-LD 1.1: A JSON-based Serialization for Linked Data, Technical Report, World Wide Web Consortium (W3C), 2020. URL: <https://www.w3.org/TR/json-ld/>.
 - [18] FAIR-IMPACT Consortium, FAIR assessment tools, <https://fair-impact.eu/fair-assessment-tools>, 2025. Funded by the European Union under Grant Agreement 101057344. Accessed: December 8, 2025.
 - [19] R. Kundu, R. Mehdi, V. D. Tran, E. Frakes, A. Daundkar, M. Sumudumalie, V. S. Mandayam, J. A. Lample, M. Li, L. S. Bruckman, E. I. Barcelos, A. Sehrioglu, R. H. French, Y. Wu, OntoCheck: Query-driven ontology assessments for scientific domain applications, <https://pypi.org/project/OntoCheck/>, 2025. Python software package.
 - [20] M. Jensen, G. D. Colle, S. Kindya, C. More, A. P. Cox, J. Beverley, The Common Core Ontologies, 2024. URL: <http://arxiv.org/abs/2404.17758>. doi:10.48550/arXiv.2404.17758, arXiv:2404.17758 [cs].
 - [21] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, S. Lewis, The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature Biotechnology* 25 (2007) 1251–1255. URL: <https://www.nature.com/articles/nbt1346>. doi:10.1038/nbt1346.
 - [22] M. Schilling, P. von Hartrott, J. Waitelonis, T. Hanke, H. Birkholz, H. B. Nasrabadi, K. Razghandi, K. Zaripova, F. Thonagel, F. Neuhaus, M. Glauer, L. Vogt, H. Sack, L. Mädler, B. Bayerlein, C. Eberl, Semantic Modeling in Materials Science and Engineering With Platform MaterialDigital Core Ontology 3.0 (2026). doi:10.1002/adem.71106.
 - [23] E. G. Bologne, University of, G. G. Consulting, Goldbeck, J. F. SINTEF, A. H. IWM, Fraunhofer, G. S. ACCESS, Elementary Multiperspective Material Ontology (EMMO), <https://emmo-repo.github.io/>, 2025.
 - [24] S. Lohmann, S. Negru, D. Bold, The ProtégéVOWL Plugin: Ontology Visualization for Everyone, in: V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, A. Tordai (Eds.), *The Semantic Web: ESWC 2014 Satellite Events, Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2014, pp. 395–400. doi:10.1007/978-3-319-11955-7_55.
 - [25] N. F. Noy, D. L. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*, Technical Report KSL-01-05, Stanford University, Knowledge Systems, AI, Lab, Palo Alto, California, 2001. URL: https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html, tex.ids= natalyaf.noyOntologyDevelopment101.
 - [26] SDLE Research Center, MDS-Onto Documentation, https://cwrusdle.bitbucket.io/files/MDS_Onto-v0.3.0.3/index-en.html, 2026. Accessed: 2026-07-19.
 - [27] SDLE Research Center, MDS-Onto Submission Form, <https://tally.so/r/KYOLKz>, 2026. Form hosted on Tally. Accessed: 2026-07-19.
 - [28] Tally Forms Team, Tally: The Form Builder, 2026. URL: <https://tally.so/about>.
 - [29] SDLE Research Center, MDS-Onto Slack Community Invitation, https://join.slack.com/t/cwru-mds-onto/shared_invite/zt-3y9dwb2lf~jIbLf0g9zaJ_ngjbQ7T7g, 2026. Slack workspace invitation link. Accessed: 2026-07-19.
 - [30] Graphwise, Graphwise GraphDB, 2026. URL: <https://graphwise.ai/components/graphdb/>.
 - [31] H. H. Aung, D. Li, J. Liu, C. Barretta, Y. Sheng, Y. J. Jo, J. C. Jimenez, E. I. Barcelos, G. Oreski, R. H. French, L. S. Bruckman, Lifetime and degradation study of poly(methyl methacrylate) via a data-driven study protocol approach, *Integrating Materials and Manufacturing Innovation* 12 (2023) 349–370. doi:10.1007/s40192-023-00322-6.
 - [32] L. S. Bruckman, N. R. Wheeler, J. Ma, E. Wang, C. K. Wang, I. Chou, J. Sun, R. H. French, Statistical and Domain Analytics Applied to PV Module Lifetime and Degradation Science, *IEEE Access* 1 (2013) 384–403. URL: <https://ieeexplore.ieee.org/document/6527980>. doi:10.1109/ACCESS.2013.2267611.

- [33] W.-H. Huang, N. R. Wheeler, A. G. Klinke, Y. Xu, W. Du, A. K. Verma, A. Gok, D. A. Gordon, Y. Wang, S. N. Venkat, H. Aung, Y. Jo, X. Yu, K. Ozdemirli, V. D. Tran, J. Gordon, J. C. Jimenez, J. Liu, A. J. Curran, J. S. Fada, X. Ma, J. L. Braid, J. L. W. Carter, L. S. Bruckman, R. H. French, netsem: Network structural equation modeling, 2026. URL: <https://cran.r-project.org/web/packages/netSEM/index.html>.
- [34] W. R. Young, Solar on schools designed for emergency shelters: 39th ieee photovoltaic specialist conference, in: 2013 IEEE 39th Photovoltaic Specialists Conference (PVSC), 2013, pp. 1521–1525. doi:10.1109/PVSC.2013.6744434.
- [35] N. G. Dhere, S. Schleith, Reliability of hybrid photovoltaic DC micro-grid systems for emergency shelters and other applications, in: N. G. Dhere (Ed.), Reliability of Photovoltaic Cells, Modules, Components, and Systems VII, volume 9179, International Society for Optics and Photonics, SPIE, 2014, p. 91790F. URL: <https://doi.org/10.1117/12.2083284>. doi:10.1117/12.2083284.
- [36] S. A. Miskovich, B. E. Meyers, E. Apostolaki-Iosifidou, C. Berschauer, C. Ding, A. Dufour, D. J. F. Rodriguez, J. Goncalves, A. Londono-Hurtado, V.-H. Lian, T. Lin, J. Luo, X. Ming, D. Ragsdale, D. Serbetcioglu, S. Sheng, J. St Louis, T. Takahashi, N. Telang, M. Victoriano, H. Zhang, N. Yadav, Solar data tools: a python library for automated analysis of unlabeled pv data, Journal of Open Source Software 10 (2025) 8478. URL: <https://doi.org/10.21105/joss.08478>. doi:10.21105/joss.08478.
- [37] R. Wieser, Y. Fan, X. Yu, J. Braid, A. Shaton, A. Hoffman, B. Spurgeon, D. Gibbons, L. S. Bruckman, Y. Wu, R. H. French, Large scale, data driven, digital twin models: Outlier detection and imputation, in: 2024 IEEE 52nd Photovoltaic Specialist Conference (PVSC), 2024, pp. 0902–0905. doi:10.1109/PVSC57443.2024.10748985.
- [38] B. G. Pierce, H. H. Aung, T. G. Ciardi, K. J. Hernandez, R. J. Wieser, W. Yue, Y. Fan, A. C. H. Bradley, B. P. Rajamohan, E. I. Barcelos, J. C. Jimenez, B. K. Spears, B. Giera, R. X. Gao, M. Li, K. O. Davis, L. S. Bruckman, Y. Wu, P. K. Tripathi, R. H. French, Generalized graph foundation models as versatile data-driven digital twins for complex technological systems, Scientific Reports (2026). doi:10.1038/s41598-026-61113-5.