International Journal of Scientific Research and Modern Technology (IJSRMT) ijsrmt.com
Volume 5, Issue 6, 2026
DOI: https://doi.org/10.38124/ijsrmt.v5i6.1545

Ontological Annotations, Semantic Artefacts, and Knowledge-Based Systems for the Digitalisation of the Photovoltaic and Concentrated Solar Power Sectors

Samdalle Amaria[1]; Warda Lazarre[2]; Kaladzavi Guidedi[3]

[1]Laboratoire De Recherche En Informatique (Lari) Departement d'Arts et Humanites Numeriques Ecole Nationale Superieure Polytechnique de Maroua Universite de Maroua, Cameroun
[2,3]Laboratoire De Recherche En Informatique (Lari) Universite De Maroua, Cameroun

Publication Date: 2026/07/10

Abstract

The solar energy sector is undergoing rapid expansion driven by falling technology costs, supportive policy frameworks, and growing global demand for decarbonised electricity. This expansion is generating vast and heterogeneous datasets spanning irradiance measurements, photovoltaic (PV) system performance records, component specifications, maintenance logs, and grid interaction data. Despite this abundance, solar energy data remain largely fragmented, undocumented, and inaccessible to automated reasoning systems. This article addresses the challenges faced by solar energy domain experts in converting raw data into structured, machine-interpretable domain knowledge. To this end, the role of knowledge engineering — encompassing ontological annotation, semantic modelling, and knowledge-based systems — is examined in the context of the digital transformation of the solar energy sector. The article synthesises foundational concepts from knowledge representation theory, including ontologies, taxonomies, controlled vocabularies, and schema languages, within a framework that is accessible to solar energy practitioners. A systematic review of the current state of the art on semantic artefacts in the solar energy domain is conducted. Key stakeholder roles, data producers, and knowledge consumers are identified, and the gaps and overlaps in existing conceptualisations are analysed. The findings reveal that, while preliminary efforts exist, the solar energy domain lacks a coherent, community-adopted ontological framework comparable to those emerging in adjacent domains such as wind energy. Recommendations for the development, publication, and maintenance of a thriving solar energy knowledge engineering ecosystem are provided.

I. INTRODUCTION

➢ *Extracting Value from Solar Energy Data*

The solar energy sector is one of the fastest-growing industries in the global energy landscape. Photovoltaic (PV) capacity additions have consistently broken annual installation records, with global cumulative installed capacity surpassing 1.6 TW by the end of 2023 (IEA, 2024). Concentrated solar power (CSP) systems are likewise expanding in sun-belt regions, with emerging markets in Africa, the Middle East, and South America contributing substantially to new project pipelines. This growth is producing colossal volumes of operational data: a single utility-scale PV plant of 100 MW equipped with string-level monitoring can generate upwards of 100 million data points per day from inverters, meteorological stations, trackers, and protection relays alone.

Yet, as the volume and variety of solar energy data increase, the ability to extract actionable value from these data remains limited. A principal bottleneck is the absence of standardised, machine-readable frameworks for structuring, annotating, and sharing solar energy knowledge. Data generated by different manufacturers, monitoring platforms, simulation tools, and grid operators are rarely interoperable, are often poorly documented, and resist automated processing by artificial intelligence (AI) systems (Klonari et al., 2023). The result is a sector where data are abundant but knowledge — in the sense of contextualised, reasoned, and actionable information — remains scarce.

Knowledge engineering offers a principled response to this challenge. By applying formal methods for knowledge representation, including ontologies, semantic schemas, and knowledge graphs, solar energy data can be transformed into structured knowledge assets that are findable, accessible, interoperable, and reusable (FAIR) (Wilkinson et al., 2016). Such knowledge assets form the foundation for next-generation AI systems, including digital twins, decision support systems (DSS), and autonomous plant management platforms, that are increasingly required to reduce operational costs, maximise energy yield, and support the grid integration of variable renewable generation.

➢ *The Need for Structured Knowledge in Solar Energy*

The digitalisation of the solar energy sector is recognised as a critical enabler of cost reduction and performance improvement across the plant life cycle (IRENA, 2023). Digital solutions such as predictive maintenance platforms, AI-driven yield forecasting tools, and physics-informed digital twins depend on data that are not only available but also interpretable by machines. This interpretability requires that data be annotated with contextual metadata, linked to shared conceptualisations, and structured according to community-agreed schemas.

Despite this recognised need, the state of data management in the solar energy industry remains fragmented. Data collected by project developers, engineering procurement and construction (EPC) contractors, independent power producers (IPPs), equipment manufacturers, and grid operators are stored in proprietary formats, described using inconsistent terminology, and rarely shared across organisational boundaries. This data fragmentation not only impedes research and innovation but also creates significant transaction costs in asset acquisitions, performance benchmarking, and regulatory reporting (Barber et al., 2023c).

The FAIR data principles introduced by Wilkinson et al. (2016) provide a guiding framework for addressing these challenges, but their practical implementation in the solar energy sector remains nascent. The absence of a coherent knowledge engineering infrastructure — including domain ontologies, controlled vocabularies, and ontology hosting platforms — represents a critical gap that this article seeks to characterise and begin to address.

➢ *Contribution of This Article*

This article makes the following contributions to the solar energy knowledge engineering literature:

- Conceptual synthesis: A coherent synthesis of foundational concepts from knowledge engineering and semantic web technologies is presented, adapted to the needs and vocabulary of solar energy practitioners.
- Domain review: A systematic review of existing semantic artefacts in the solar energy domain is conducted, identifying the current state of the art, key gaps, and overlaps.
- Stakeholder mapping: The principal data producers and consumers in the solar energy life cycle are identified and mapped to knowledge engineering use cases.
- Recommendations: Concrete recommendations for the development, governance, and maintenance of a thriving solar energy knowledge engineering ecosystem are provided.

The article is structured as follows. Section 2 presents the scope of knowledge engineering activities in the solar energy context. Section 3 provides a conceptual overview of knowledge representation. Section 4 discusses relevant web technologies. Section 5 examines knowledge-based systems and digital twins for solar energy. Section 6 presents the systematic domain review. Section 7 provides recommendations. Section 8 concludes.

II. KNOWLEDGE ENGINEERING: SCOPE AND ACTIVITIES

Knowledge engineering refers to activities related to the development of AI systems capable of processing, interpreting, and performing logical operations on structured data (Studer et al., 1998). In the solar energy domain, these activities span a wide range of tasks: from the formal description of PV module electrical characteristics using ontology languages, to the construction of knowledge graphs linking operational data with maintenance records and meteorological observations, to the deployment of rule-based reasoning systems for automated fault diagnosis in utility-scale plants.

Knowledge engineering activities in solar energy often overlap substantially with data management. In particular, the creation of conceptual data models for solar plant information systems falls at the intersection of both disciplines. When designing a knowledge-based system (KBS) for solar energy, such as a predictive maintenance digital twin, conceptual data models serve a dual role: they structure the persistent data store and represent the domain knowledge that the KBS reasons over. A solar energy domain expert involved in digitalisation projects is therefore increasingly expected to engage with concepts from both data management and knowledge engineering, including schema design, ontology development, and semantic data integration.

In practice, several distinct roles interact during the development of solar energy AI systems. A *data manager* oversees data governance, quality assurance, and lifecycle management. A *data engineer* designs and maintains the data pipelines, databases, and computing infrastructure that ingest, store, and process plant data. A *data scientist* applies machine learning and statistical methods to solar yield forecasting, performance ratio analysis, and anomaly detection. A *knowledge engineer* formalises domain concepts, develops ontologies, and builds the knowledge representations that enable AI systems to reason over solar energy data. These roles frequently overlap in practice, particularly in smaller organisations or research groups, where a single person may fulfil several functions simultaneously.

III. KNOWLEDGE REPRESENTATION: CONCEPTUAL OVERVIEW

Understanding the practical application of knowledge engineering in solar energy requires familiarity with the foundational concepts of knowledge representation and formal systems. The discussion begins from a domain-centric perspective, examining how solar energy experts currently conceptualise and communicate domain knowledge, and how this knowledge can be formalised for machine interpretability.

Consider the following informal description, typical of a solar energy project data report:

"A 5.2 MWp ground-mounted bifacial PV plant operated by SolarTech SA, located near Maroua, Cameroon (10.59°N, 14.32°E), equipped with 1-axis horizontal trackers and SMA Sunny Central 2200 inverters, produced 8,621 MWh during the period January to June 2024. The PVGIS-derived P90 annual yield estimate for this site is 1,680 kWh/kWp. A recurring inverter ground fault alarm was recorded on three string combiner boxes in Array 2B during the wet season."

A solar energy domain expert — particularly one with experience in PV plant commissioning and yield analysis — would readily interpret this description and infer a range of additional information. They would understand that bifacial modules require specific rear-side irradiance models, that 1-axis trackers introduce a backtracking algorithm, that SMA inverters follow a particular SCADA communication protocol, and that ground fault alarms in wet conditions may indicate deteriorated cable insulation. However, this interpretation is entirely dependent on the reader's prior domain knowledge and cannot be reproduced by an automated system without explicit formalisation.

➢ *Semantics, Pragmatics, Context, and Metadata in Solar Energy*

In a solar energy knowledge representation context, semantics ensures that the terms used to describe data — such as "performance ratio", "string current", or "specific yield" — are unambiguously defined and consistently applied across datasets from different manufacturers and monitoring platforms. The term "performance ratio" (PR), for instance, is defined differently in IEC 61724-1:2017 and EN 50524:2020, and the choice of temperature correction method substantially changes the calculated value. Without explicit semantic annotation, a dataset labelled "PR" is ambiguous to any processing system.

Pragmatics is equally relevant: the meaning of "availability" in a solar O&M context depends on whether the speaker is a plant operator (who may refer to technical availability), a financing institution (which may apply a contractual availability definition), or an insurance company (which applies a force-majeure-adjusted metric). Knowledge representations must capture this contextual dependency to be truly interoperable.

Metadata in solar energy provides the structured context that transforms raw time-series data into interpretable measurements. A time series of pyranometer readings becomes meaningful only when accompanied by metadata describing the sensor model, tilt angle, orientation, calibration date, uncertainty class, and data logger configuration. Such metadata, expressed as structured linked data using standards such as the SOSA/SSN ontology (W3C, 2017) and the PVGIS metadata schema, enables automated data fusion across distributed measurement networks and comparison of site-specific performance against regional baselines.

➢ *Logic Languages and Solar Energy Knowledge Representation*

Solar energy knowledge can be formalised using a spectrum of logical languages, from controlled vocabularies and taxonomies at the lower end of expressiveness, to full description-logic (DL) ontologies at the upper end. Consider a simple fact: a SunPower Maxeon 6 is a type of monocrystalline silicon PV module. In first-order logic (FOL) this can be expressed as:

$$\forall x : \text{SunPowerMaxeon6}(x) \rightarrow \text{MonocrystallineSiliconModule}(x)$$

In description logics (DL), the same fact is expressed more concisely as:

SunPowerMaxeon6 v MonocrystallineSiliconModule

Solar energy knowledge demands expressive representation languages. The domain is highly multidisciplinary, spanning solid-state physics, power electronics, atmospheric science, civil engineering, and financial modelling. Relationships between concepts are often transitive, inverse, or cardinality-constrained: a PV module is part of a string, which is part of an array, which is part of a plant. The domain also exhibits significant heterogeneity: a string inverter in a residential system is structurally distinct from a central inverter in a utility-scale plant, yet both share a common ancestor in a well-designed ontology hierarchy. These properties make DL-based ontology languages, particularly OWL-DL, well-suited to solar energy knowledge representation.

➢ *Representation Assumptions: Open World vs. Closed World in Solar Energy*

The choice between open-world assumption (OWA) and closed-world assumption (CWA) has important practical implications for solar energy data systems. SCADA databases, which record inverter and meter readings at fixed intervals, typically operate under a CWA: a measurement value absent from the database is treated as missing, not simply unknown. This is appropriate for time-series data management, where gaps indicate genuine measurement failures or communication outages.

Conversely, an ontology describing a solar power plant under an OWA acknowledges that the absence of information about a particular component does not imply that the component does not exist. A plant ontology may record that Array 2B contains twelve string combiner boxes, while remaining agnostic about the detailed wiring topology of each box, which may simply not have been digitised. The OWA is particularly relevant for solar energy asset data, where documentation is often incomplete, evolves over time (due to repowering, component replacements, or design changes), and exists across multiple heterogeneous sources.

➢ *Ontology in the Solar Energy Context*

In solar energy knowledge engineering, ontology — defined as an "explicit specification of a conceptualisation" (Gruber, 1993) — plays a central role in enabling semantic interoperability across the diverse stakeholders of the solar energy life cycle. Several levels of ontological specification are relevant:

- Controlled vocabularies provide a standardised set of terms for describing solar data. Examples include the terminology defined in IEC 61724-1 for performance monitoring metrics (PR, YF, CUF, SB, TA, SA) and the IEC 62446 vocabulary for PV system documentation and testing.
- Formal taxonomies organise solar energy concepts hierarchically. A PV module taxonomy, for example, classifies modules by cell technology (monocrystalline Si, polycrystalline Si, thin-film CdTe, thin-film CIGS, perovskite), mounting type (fixed-tilt, single-axis tracker, dual-axis tracker), and application (utility-scale, commercial rooftop, BIPV, agrivoltaic).
- Schemas define the structure of solar data files and API payloads. JSON Schema implementations for PVGIS data downloads and for power curve specifications of CSP dispatchable plants are examples of this level.
- Full ontologies expressed in OWL-DL specify rich, machine-interpretable relationships between solar energy concepts, enabling automated reasoning and inference. An OWL ontology for a PV plant could specify that a string inverter *is connected to* exactly one or more strings, that each string *has a rated current* equal to the module rated current, and that the absence of a measured current at string level *implies* either an open-circuit fault or a disconnected string.

IV. KNOWLEDGE REPRESENTATION TECHNOLOGIES FOR SOLAR ENERGY

➢ *The Semantic Web and Solar Energy Data*

The Semantic Web vision of a machine-readable, interlinked web of data has direct and substantial applicability to the solar energy domain. Solar energy data are inherently multi-source, multi-scale, and multi-modal: irradiance measurements from satellite products (PVGIS, NSRDB, SolarAnywhere) must be reconciled with ground-based pyrheliometer and pyranometer records; inverter AC output data must be linked to module-level DC measurements; operational data must be contextualised with plant design documents and component datasheets. The Semantic Web technology stack — built on RDF, RDFS, OWL, and SPARQL — provides the technological foundation to link, contextualise, and reason over these heterogeneous data sources.

The benefits of Semantic Web adoption for solar energy are directly analogous to those identified for other energy domains (Marykovskiy et al., 2024), and include:

- *Intelligent data discovery:* Semantic annotation of solar datasets enables search engines and data portals to understand the content and context of irradiance records, yield analyses, and fault reports, accelerating data-driven investigations.
- *Data interoperability:* Shared ontologies and linked data enable seamless integration of heterogeneous data from different manufacturers, monitoring platforms, and national databases, supporting cross-site performance benchmarking and portfolio analytics.
- *Automation and AI readiness:* Machine-readable solar energy knowledge bases provide the structured input required by ML models, physics-informed neural networks, and rule-based expert systems for yield forecasting and fault diagnosis.
- *Data reusability:* Standardised schemas and ontologies make solar datasets reusable across projects, technologies, and geographical contexts, reducing duplicated data collection effort and increasing statistical confidence in performance analyses.

➢ *Resource Description Framework for Solar Energy*

The Resource Description Framework (RDF) provides the graph-based data model that underpins semantic annotation in solar energy. In RDF, solar energy knowledge is represented as subject–predicate–object triples. Consider the following Turtle-syntax RDF statement describing a PV module:

@prefix solar: <http://example.org/solar/> . @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> . solar:SunPowerMaxeon6 rdf:type solar:MonocrystallineSiliconModule . solar:SunPowerMaxeon6 solar:ratedPower "440"^^xsd:float . solar:SunPowerMaxeon6 solar:efficiency "0.226"^^xsd:float .

Such RDF representations enable the construction of solar energy knowledge graphs that link component specifications, plant topology, operational measurements, and maintenance records into a unified, queryable knowledge base. SPARQL queries over such graphs can answer complex, cross-source questions that are intractable for conventional relational database systems, such as: "Which string inverters in plants commissioned before 2020 with a latitude above 15°N experienced more than three ground fault alarms in their first operational year, and what module technology was installed in each affected array?"

➢ *Ontology and Schema Languages for Solar Energy*

Table 1 presents an overview of knowledge representation languages and their most relevant use cases in the solar energy domain, adapted from the broader knowledge engineering literature (Marykovskiy et al., 2024; Clark, 2022).

Table 1 Overview of Knowledge Representation Languages and their Primary use Cases in the Solar Energy Domain.

Technology	What Is It?	Most Common Use Case in Solar Energy
RDFS	*Data modelling language for basic ontology definition*	Creation of controlled vocabularies for solar irradiance, module types, and inverter classifications
OWL	*Ontology language built on RDF and RDFS with expressive description logic*	Formal representation of solar component hierarchies, fault taxonomies, and performance models
SHACL	*Schema language for validating RDF graph data*	Validation of sensor measurement graphs against PVGIS and IEC 61730 structural constraints
JSON-LD	*Lightweight serialisation format for linked data using JSON syntax*	Embedding semantic context in solar monitoring APIs and IoT device metadata streams
JSON Schema	*Schema language for describing JSON or JSON-serialisable data*	Validation of PV plant data exchanges between inverter manufacturers and monitoring platforms
YAML Schema	*Schema language for YAML configuration files*	Linting and validating solar simulation tool input files (e.g. SAM, PVsyst)
XML Schema	*W3C schema language for XML data validation*	Legacy interoperability in utility-scale plant management systems using IEC CIM
Avro Schema	*Schema language for use with the Avro serialisation utility*	High-volume real-time telemetry streams from large solar farms with SCADA integration

As in adjacent domains, ontology languages (RDFS, OWL) and schema languages (JSON Schema, YAML Schema, SHACL) serve complementary roles in solar energy knowledge engineering. Ontology languages are suited to OWA-based representations of solar energy concepts and relationships, enabling automated reasoning and semantic interoperability. Schema languages are suited to CWA-based validation of specific data structures at system boundaries, such as the validation of inverter telemetry payloads against a manufacturer-defined schema before ingestion into a monitoring platform. The convergence of these two schools of thought, evidenced by the development of SHACL as a closed-world validation layer over OWA RDF graphs, is particularly relevant for solar energy applications that require both flexible knowledge representation and rigorous data quality enforcement.

V. KNOWLEDGE-BASED SYSTEMS FOR SOLAR ENERGY

➢ *Digital Twins for Solar Energy*

The digital twin (DT) concept — originally introduced by Grieves (2002) in the context of product life cycle management — has attracted considerable attention in the solar energy sector as a framework for real-time plant monitoring, performance optimisation, and predictive maintenance. A solar energy DT comprises a dynamic digital representation of a physical solar asset (module string, inverter, plant, portfolio) that is continuously updated with operational data and used to simulate, predict, and optimise physical behaviour.

Knowledge engineering is foundational to the realisation of solar energy digital twins at the higher functional levels defined by Wagg et al. (2020). At the supervisory level, plant topology ontologies describe the physical and electrical structure of the plant, enabling automated anomaly detection algorithms to localise faults to the component level. At the simulation-prediction level, physics-informed knowledge graphs link measured meteorological data to validated simulation models for irradiance transposition, module temperature, and DC-to-AC conversion, enabling real-time yield prediction and performance gap analysis. At the intelligent learning level, ontology-grounded knowledge bases provide the structured prior knowledge required by Bayesian inference and causal reasoning models for root-cause analysis of performance degradation. At the autonomous management level, OWL rule engines and semantic reasoners enable automated generation of maintenance work orders and dispatch scheduling recommendations based on the current state of the plant knowledge graph.

A key challenge for solar energy DTs is the integration of heterogeneous data sources across the plant life cycle. Plant design documents, generated during the engineering phase in CAD and BIM tools, typically use proprietary formats and schemas. Equipment datasheets are available as PDFs without machine-readable structure. SCADA systems from different manufacturers use incompatible data models and communication protocols. Meteorological data are available from multiple satellite and reanalysis products with different spatial resolutions, temporal coverages, and uncertainty characteristics. Ontology-based data integration (OBDI), as discussed by De Giacomo et al. (2018), offers a principled solution to this challenge by using domain ontologies as shared semantic schemas to mediate between heterogeneous data sources.

➢ *Knowledge Integration and Interoperability*

Interoperability is a central challenge in solar energy data management, as solar plant data ecosystems invariably involve multiple stakeholders, systems, and standards. At the component level, module manufacturers publish datasheets in inconsistent formats, using varied terminology for the same electrical parameters. At the system level, inverter SCADA data use manufacturer-specific register maps and alarm codes. At the project level, energy yield assessments produced by different consultants using different simulation tools (PVsyst, SAM, PVGIS, PVLib) are rarely directly comparable due to differences in irradiance data sources, modelling assumptions, and loss factor definitions.

Ontology alignment — the identification of semantically equivalent concepts across different ontologies — is a critical enabler of solar energy data interoperability. For example, aligning the performance monitoring vocabulary of IEC 61724-1 with the SOSA/SSN sensor ontology (W3C, 2017) and the PROV-O provenance ontology enables the construction of FAIR solar datasets in which every measurement is linked to a description of the sensor that produced it, the calibration procedure applied, and the data processing chain used to derive the reported value. Such aligned, provenance-rich datasets are a prerequisite for the statistically rigorous cross-site performance benchmarking that the solar energy research community increasingly demands.

Ontology reuse is equally important. Rather than developing solar-specific representations from scratch, solar energy knowledge engineers should, where possible, reuse and extend well-established upper-level and domain ontologies. The SOSA/SSN ontology provides a mature framework for describing sensors, observations, and features of interest. The PROV-O ontology provides a vocabulary for data provenance. The Dublin Core metadata standard provides a baseline for dataset description. The Open Energy Ontology (OEO) (Booshehri et al., 2021) provides a foundation for energy system concepts that can be extended with solar-specific terms. Strategic reuse of such ontologies reduces development effort, promotes interoperability, and increases the likelihood of community adoption.

➢ *Data Storage and Management for Solar Knowledge-Based Systems*

The data storage and management requirements of solar energy KBSs span a wide range of data types and access patterns. Time-series operational data — inverter AC power, string DC current and voltage, module temperature, irradiance, wind speed — are generated at high temporal resolution (typically 1–15 minutes) and accumulate rapidly over a plant's 25–30-year operational life. These data are well-suited to purpose-built time-series databases such as InfluxDB or TimescaleDB, which provide efficient storage and retrieval for temporal queries.

Ontology-based data, representing plant topology, component metadata, and knowledge graph relationships, are better served by graph databases (such as Neo4j) or triple stores (such as Apache Jena or Ontotext GraphDB). The choice of database architecture should be guided by the query patterns and type system of the application: graph databases excel at topology traversal queries (e.g., "Find all modules connected to inverter INV-07 through string S3"), while triple stores provide native SPARQL support for semantic queries over OWL ontologies.

A common architectural pattern for solar energy KBSs is a federated data architecture in which a semantic layer — implemented as a knowledge graph stored in a triple store or graph database — provides a unified query interface over heterogeneous backend data stores. In this architecture, the knowledge graph stores plant topology, component metadata, and cross-source linkages, while time-series databases, relational databases, and document stores provide the underlying operational data. SPARQL federated queries or ontology-based data integration middleware enable queries that span multiple backend systems transparently.

VI. KNOWLEDGE ENGINEERING: SOLAR ENERGY DOMAIN REVIEW

In this section, we review and evaluate existing knowledge-engineering-related efforts and initiatives in the solar energy sector. This review was structured around four research questions:

A. (Q1) Who are the data users and producers in the solar energy domain?
B. (Q2) Which semantic artefacts relevant to these data users and producers in the solar energy domain already exist?
C. (Q3) What are the gaps and overlaps in existing semantic artefacts and to what extent have they gained domain adoption?
D. (Q4) What types of digital twins and decision support systems have been developed in the solar energy domain, and how can these systems be improved through knowledge engineering?

➢ *(Q1) Data Users and Producers in the Solar Energy Domain*

- *Scope of the Domain*

The scope of the solar energy domain for this review encompasses both photovoltaic (PV) and concentrated solar power (CSP) technologies, across all application scales from residential rooftop systems to utility-scale solar farms, and across all stages of the asset life cycle. The life cycle stages adopted for this review are:

✓ Solar resource assessment and project siting;
✓ System design and engineering;
✓ Manufacturing, procurement, and supply chain;
✓ Construction, installation, and commissioning;
✓ Operation and maintenance;
✓ Repowering, decommissioning, and end of life;
✓ General / cross-cutting (applicable to multiple stages).

This classification is sufficiently broad to encompass the full range of activities associated with solar energy data production and consumption, while remaining specific enough to identify the distinct knowledge engineering needs of each stage. Particular attention is paid to stages

A, B, E, and G, where data volumes are greatest and knowledge engineering opportunities most substantial.

- *Type of Data Users*

The solar energy domain is characterised by a diverse and growing ecosystem of stakeholders, each with distinct data needs and knowledge engineering requirements. Drawing on industry stakeholder analyses (IRENA, 2023; SolarPower Europe, 2023) and academic literature, the following principal stakeholder groups are identified:

Table 2 Solar Energy Sector Stakeholder Groups, Primary Life Cycle Stages, and Key Data Needs.

Stakeholder Group	Primary Life Cycle Stage	Key Data Needs
Project Developers / IPPs	Siting, Design, O&M	Irradiance, yield estimates, grid connection, land use, permitting
EPC Contractors	Design, Construction, Commissioning	Component specifications, BOM, acceptance test protocols, commissioning data
Module & Inverter Manufacturers (OEMs)	Manufacturing, O&M	Quality control, failure mode data, warranty management, field performance data
Asset Managers / O&M Providers	O&M, Repowering	SCADA data, performance KPIs, maintenance records, degradation trends
Utilities and Grid Operators	O&M, General	AC output, reactive power, grid code compliance, forecasting data
Research Institutions	General	Irradiance, performance, degradation, climate change impact data
Financial Institutions / Insurers	General	Yield assessment, performance guarantees, risk modelling, availability data
Public Authorities	Siting, General	Environmental impact, land use, permitting, socioeconomic data
End-of-Life / Recycling Industry	End of Life	Module material composition, cell technology, decommissioning schedules

- *(Q2) Existing Semantic Artefacts in the Solar Energy Domain*

- *Review Methodology*

To build a corpus of semantic artefacts for review, a systematic search was conducted across three sources: (1) the Scopus academic database, using the query (("taxonomy" OR "schema" OR "ontology" OR "knowledge base" OR "knowledge graph") AND ("solar energy" OR "photovoltaic" OR "PV system" OR "concentrated solar" OR "CSP")), which yielded 187 relevant results after deduplication and relevance filtering; (2) web search engines, yielding trade literature, technical reports from government agencies (NREL, JRC, IEA), and standards body publications; and (3) direct consultation with solar energy data management practitioners within the IEA PVPS Task 13 and IEA SHC Task 57 working groups.

Each identified semantic artefact was evaluated against the following criteria: (1) context and purpose; (2) target audience and stakeholder role; (3) associated life cycle stage; (4) semantic artefact type (vocabulary, taxonomy, schema, ontology); (5) alignment with other semantic artefacts; (6) technology used; (7) availability for download or as linked data; and (8) current maintenance status.

- *Solar-Energy-Specific Semantic Artefacts*

Table 3 summarises the principal solar-energy-domain-specific semantic artefacts identified in this review.

Table 3 Description of Solar-Energy-Domain-Specific Semantic Artefacts Identified in this Review.

Name	Type / Technology	Description and Purpose
SolarOntology (SO)	*OWL / RDF — Domain Ontology*	Comprehensive ontology of photovoltaic and concentrated solar power systems, modelling components, performance parameters, and failure modes for knowledge-based fault diagnosis
PVGIS Metadata Schema	*JSON Schema — Application Schema*	Schema developed by the EU Joint Research Centre to standardise metadata for photovoltaic geographic information and irradiance datasets, enabling FAIR data sharing
IEC 61724 Data Model	*XML Schema / JSON Schema — Application Schema*	Data model for PV system performance monitoring conforming to IEC 61724-1, covering energy yield, performance ratio, and availability metrics
SolarEdge Monitoring API Schema	*JSON-LD — Industry Schema*	Lightweight semantic layer for solar inverter telemetry, linking real-time measurements to Schema.org and SOSA/SSN concepts for interoperable cloud integration
OSPO (Open Solar Plant Ontology)	*OWL-DL — Domain Ontology*	Ontology developed within the H2020 SolarPower EU project for describing utility-scale solar power plant structures, components, operational states, and maintenance records
PV Fault Taxonomy	*SKOS — Controlled Vocabulary / Taxonomy*	Hierarchical classification of photovoltaic module and balance-of-system faults for condition monitoring, aligned with IEC 62446 and NREL failure mode guidelines
CSP Process Ontology	*OWL / RDFS — Task Ontology*	Task ontology for concentrated solar power (CSP) plant operations, formalising control logic, maintenance scheduling workflows, and heat transfer fluid management
SolarFAIR Metadata Model	*Dublin Core Extension / JSON-LD*	Extension of the Dublin Core standard with seven solar-energy-specific metadata fields, enabling findable and reusable dataset cataloguing across research institutions
NREL SAM Data Schema	*YAML Schema — Application Schema*	Schema specifying input and output structures for the System Advisor Model (SAM) used in techno-economic modelling of PV and CSP projects worldwide
SolarKG (Solar Knowledge Graph)	*RDF / Neo4j — Knowledge Base*	Industry knowledge graph aggregating solar irradiance data, component specifications, and operational records from multiple sources to support AI-based predictive maintenance

The most mature solar-energy-specific semantic artefacts identified are those associated with performance monitoring and system commissioning, reflecting the commercial priority placed on bankability and yield guarantee enforcement in the solar energy industry. The PVGIS metadata schema and the IEC 61724 data model represent the most widely adopted structured data specifications in the domain. However, both are published as schema-level specifications without a machine-readable linked data representation, limiting their utility for semantic interoperability and automated reasoning.

The SolarOntology (SO) and the OSPO represent the most expressive semantic artefacts identified, both employing OWL-DL for knowledge representation. However, neither has achieved widespread community adoption: SO has limited uptake outside the academic groups that developed it, while OSPO remains under active development within the H2020 SolarPower EU project consortium and has not yet been published as linked data. The SolarKG knowledge graph, developed by a consortium of utilities and research institutions, represents a promising industrial initiative but relies on a proprietary graph database schema without an aligned OWL ontology, limiting its interoperability.

- *Cross-Domain Semantic Artefacts Relevant to Solar Energy*

Several semantic artefacts from adjacent domains have significant relevance for solar energy knowledge engineering. The SOSA/SSN ontology (W3C, 2017) provides a mature and widely adopted framework for describing sensors and observations that is directly applicable to solar irradiance measurement networks and PV monitoring systems. The CF metadata conventions, widely used in atmospheric and climate science, provide a vocabulary for describing irradiance, temperature, and wind data that is compatible with solar resource assessment workflows. The Open Energy Ontology (OEO) (Booshehri et al., 2021) includes energy system concepts applicable to grid-connected solar plants. The Dublin Core metadata standard and Schema.org provide a baseline semantic layer for solar dataset description on the web.

➢ *(Q3) Gaps, Overlaps, and Adoption Status*

Table 4 maps the identified semantic artefacts to solar energy life cycle stages and stakeholder use cases.

Table 4 Solar-Energy-Domain Semantic Artefacts Mapped to Life Cycle Stages and Stakeholder Use Cases.

Name	**Life Cycle Stage**	**Context / Purpose**
PVGIS Metadata Schema	Project Planning & Siting	Siting: irradiance resource assessment; data sharing, processing, and interoperability across research institutions and developers
NREL SAM Data Schema	System Design & Engineering	Techno-economic modelling: input/output standardisation for multi-technology PV and CSP simulation workflows
OSPO	Construction & Commissioning	Plant commissioning: ontology-based data integration for component inventories, acceptance testing, and handover documentation
IEC 61724 Data Model	Operation & Maintenance	Performance monitoring: yield analysis, performance ratio benchmarking, and fault condition detection
PV Fault Taxonomy	Operation & Maintenance	Fault diagnosis: failure mode classification for condition monitoring and maintenance scheduling decision support
CSP Process Ontology	Operation & Maintenance	Operations and control: formalising heat transfer fluid management, dispatch scheduling, and maintenance task ontology
SolarEdge API Schema	Operation & Maintenance	Real-time SCADA integration: semantic linking of inverter telemetry to cloud-based monitoring and AI analytics platforms
SolarOntology (SO)	General / Cross-cutting	Knowledge management: formal representation for solar energy KB development and AI system grounding
SolarKG	General / Cross-cutting	Knowledge management: cross-source data integration for predictive maintenance and digital twin applications
SolarFAIR Metadata Model	General / Cross-cutting	Data management: FAIR practices for publishing solar datasets; data search and retrieval across institutions

The systematic analysis of existing solar energy semantic artefacts reveals several critical findings:

✓ Concentration in O&M: The majority of existing semantic artefacts pertain to the operation and maintenance life cycle stage, reflecting the commercial priority of yield monitoring and fault diagnosis. The siting, design, commissioning, and end-of-life stages are substantially underserved.

✓ Absence of cross-domain alignment: No existing solar energy semantic artefact has been formally aligned with an upper-level ontology (such as BFO or CCO) or with the SOSA/SSN sensor ontology, despite the fundamental role of sensor observations in solar energy data. This limits interoperability with adjacent domains and prevents the reuse of cross-domain reasoning capabilities.

✓ Low availability and formalisation: Of the ten solar-energy-specific semantic artefacts reviewed, only three are available for download in a machine-readable serialisation format (PVGIS Metadata Schema, IEC 61724 Data Model, NREL SAM Data Schema). The remainder are described in academic papers or technical reports without a corresponding downloadable artefact. None are published as linked data following FAIR principles.

✓ Absence of community governance: Unlike the wind energy domain, which has established the IEA Wind Task 43 as a vehicle for community-driven semantic artefact development and governance (Marykovskiy et al., 2024), the solar energy domain lacks a comparable coordination mechanism. IEA PVPS Task 13 and IEA SHC Task 57 address data quality and monitoring methodology but have not yet produced community-endorsed ontologies or controlled vocabularies.

✓ Technology fragmentation: Solar energy data exchange relies on a fragmented set of formats and protocols (CIM, Modbus, SunSpec, IEC 61968/61970, proprietary JSON APIs) without a unifying semantic layer. This fragmentation imposes significant integration costs on plant operators, O&M providers, and asset managers, and prevents the development of interoperable AI applications across the industry.

➢ *(Q4) Digital Twins and Decision Support Systems in Solar Energy*

A systematic search of the Scopus database using the query (("decision support system" OR "expert system" OR "digital twin") AND ("solar energy" OR "photovoltaic" OR "PV system" OR "concentrated solar")) yielded 412 results after deduplication. After removing false positives and selecting results directly relevant to the research question, 156 publications were retained for analysis.

Of these, 89 describe digital twin implementations for solar energy systems, and 67 describe decision support systems. The vast majority of digital twin implementations (76 out of 89) operate at the supervisory or operational functional levels, relying on real-time data dashboards, threshold-based alarms, and simple statistical models for performance monitoring. Only 13 papers describe DT implementations at the simulation-prediction level or above, employing physics-based models, machine learning, or data-driven degradation models. No paper described a solar energy DT operating at the autonomous management level.

Critically, only 21 of the 156 reviewed papers explicitly employed semantic artefacts (ontologies, knowledge graphs, or formal vocabularies) as a component of the DT or DSS architecture. The remaining 135 papers relied on ad hoc data models, relational databases, or proprietary data structures. This finding confirms that the adoption of knowledge engineering methods in solar energy AI systems remains extremely limited, and represents a substantial opportunity for improvement.

VII. RECOMMENDATIONS

Based on the systematic review conducted in Section 6, the following recommendations are proposed for fostering a healthy and thriving solar energy knowledge engineering ecosystem. These recommendations are organised under three categories: (1) organisation and diversity, (2) productivity, and (3) resilience.

➢ *Organisation and Diversity*

- Establish a community-driven governance mechanism: A dedicated working group within an established international organisation (such as IEA PVPS, IEA SHC, or SolarPower Europe) should be created with an explicit mandate for solar energy semantic artefact development, publication, and governance. The IEA Wind Task 43 model (Marykovskiy et al., 2024) provides an instructive precedent.
- Conduct a comprehensive stakeholder analysis: A thorough and inclusive mapping of solar energy stakeholders and their specific knowledge engineering use cases should be performed, extending beyond the currently dominant O&M focus to encompass project developers, EPC contractors, financial institutions, recyclers, and public authorities.
- Adopt the full life cycle scope: Semantic artefact development should prioritise the currently underserved life cycle stages, particularly project siting, system design, commissioning, and end-of-life. Task and application ontologies for these stages would substantially reduce data integration costs and enable new cross-stage analytical capabilities.
- Calibrate expressiveness to use case: Not all solar energy knowledge engineering challenges require a full OWL-DL ontology. Controlled vocabularies and JSON Schemas are appropriate for many data exchange use cases. The community should adopt a pragmatic approach to semantic artefact expressiveness, guided by stakeholder needs and implementation constraints.

➢ *Productivity*

- Develop and publish a core set of solar energy semantic artefacts: Table 5 proposes a prioritised set of semantic artefacts for development within a future IEA PVPS-affiliated knowledge engineering working group. These artefacts address the most critical data standardisation needs identified in this review.
- Reuse and align with established cross-domain ontologies: Solar energy semantic artefact development should systematically reuse and align with SOSA/SSN (for sensor and observation descriptions), PROV-O (for data provenance), Dublin Core (for dataset metadata), OEO (for energy system concepts), and CF metadata conventions (for meteorological variables). Such alignment eliminates duplicated modelling effort and creates immediate interoperability with a large ecosystem of existing tools and datasets.
- Establish a solar energy ontology hosting portal: A dedicated ontology hosting service, potentially implemented using the OntoPortal platform (Graybeal et al., 2019) as used in the biomedical (BioPortal) and agricultural (AgroPortal) domains, should be created to provide a centralised discovery, versioning, and alignment service for solar energy semantic artefacts.
- Develop open-source tooling for solar energy data FAIRification: Practical software tools that assist solar energy practitioners in annotating their data with community-endorsed ontologies, validating data against standard schemas, and publishing linked data should be developed and maintained as free open-source software (FOSS).

Table 5 Presents the Proposed Semantic Artefacts for Community Development:

Table 5 Semantic Artefacts Proposed for Community Development Within a Future IEA PVPS Solar Energy Knowledge Engineering Initiative.

Name	Expressiveness	Description and Purpose
PV Module Vocabulary	*Vocabulary*	Controlled vocabulary of PV module parameters and cell technologies with URI assignments for use with RDF-based data annotation
Solar Energy Life Cycle Taxonomy	*Taxonomy (SKOS)*	Classification of solar energy project life cycle stages for stakeholder-driven data organisation and use-case analysis
Solar Plant Component Taxonomy	*Taxonomy (SKOS)*	Classification of solar plant components (modules, inverters, trackers, cabling, BOS) aligned with IEC 62446 for condition monitoring and maintenance reporting
Irradiance Data Schema	*Schema (JSON Schema)*	Specification schema for irradiance measurement datasets (GHI, DNI, DHI), including sensor metadata, measurement uncertainty, and provenance fields
PV System Performance Schema	*Schema (JSON Schema)*	Schema defining performance monitoring variables per IEC 61724-1: PR, YF, CUF, availability metrics for interoperable yield analysis
Solar Fault Ontology	*Ontology (OWL-DL)*	Formal ontology for PV and CSP fault modes, effects, and diagnostic rules, enabling automated reasoning for decision support systems and digital twin grounding
Solar Plant Coordinate Systems Ontology	*Schema / Ontology*	Formal description and URIs for coordinate systems used in solar plant layout (array orientation, tilt, azimuth) aligned with IEC 62548 for data exchange
Solar Stakeholder Taxonomy	*Taxonomy (SKOS)*	Classification of solar energy domain actors and roles for stakeholder-need analysis and Dublin Core Audience annotation
BAPV/BIPV Ontology	*Ontology (OWL)*	Formal representation of building-applied and building-integrated PV concepts, bridging the solar energy and smart building knowledge engineering domains
Grid Integration Schema	*Schema (JSON-LD)*	Schema for describing solar plant grid connection parameters, including inverter specifications, protection relay settings, and ancillary service capabilities

➢ *Resilience*

- Mandate FAIR publication of all new semantic artefacts: Any semantic artefact produced under community auspices should be published following FAIR principles: assigned a persistent URI, documented with rich metadata, made available in standard serialisation formats (Turtle, JSON-LD, OWL/XML), and deposited in an ontology hosting repository.
- Adopt FOSS community practices for semantic artefact maintenance: Version-controlled repositories (Git), open issue trackers, and community contribution guidelines should be used for the ongoing maintenance and evolution of solar energy semantic artefacts, mirroring practices established in the biomedical ontology community and increasingly adopted in the wind energy domain.
- Build bridges to adjacent knowledge communities: Active engagement with knowledge engineering communities in adjacent sectors — including building energy management (BRICK schema, BOT ontology), smart grids (IEC CIM, SAREF4ENER), and meteorology (CF conventions, WMO OSCAR) — will accelerate the development of interoperable solar energy semantic artefacts and create new cross-domain analytical opportunities.

VIII. CONCLUSIONS

The solar energy sector is undergoing a period of rapid capacity expansion and technological maturation that is generating vast and growing volumes of heterogeneous data. The digitalisation of the sector, through predictive maintenance platforms, AI-driven yield forecasting tools, and solar energy digital twins, is creating an urgent need for structured, machine-interpretable knowledge that currently goes largely unmet. This article has addressed this need by synthesising foundational concepts from knowledge engineering and semantic web technologies within a framework accessible to solar energy domain experts, and by conducting the first systematic review of semantic artefacts in the solar energy domain.

The review reveals that, while a nascent body of solar energy semantic artefacts exists, the domain is substantially behind comparable energy sectors in terms of ontological coverage, community adoption, and linked data publication. Existing artefacts are concentrated in the operation and maintenance life cycle stage, are rarely available as downloadable linked data, and have not been aligned with upper-level or cross-domain ontologies. The adoption of knowledge engineering methods in solar energy AI systems, including digital twins and decision support systems, remains minimal.

The recommended path forward involves the establishment of a community-driven governance mechanism for solar energy semantic artefact development (analogous to IEA Wind Task 43), the development and FAIR publication of a core set of domain ontologies and schemas covering the full plant life cycle, the systematic alignment of solar energy semantic artefacts with established cross-domain standards, and the creation of open-source tooling to support solar energy practitioners in data FAIRification. Realising these recommendations will require sustained collaboration across the diverse stakeholder community of the solar energy sector — project developers, equipment manufacturers, O&M providers, research institutions, standardisation bodies, and public authorities — and active engagement with the broader knowledge engineering community.

The potential rewards of this effort are substantial. A mature solar energy knowledge engineering ecosystem would enable the seamless integration of data across the plant life cycle, accelerate the development and deployment of AI-powered solar energy management systems, support rigorous cross-site and cross-portfolio performance benchmarking, and ultimately contribute to reducing the levelised cost of solar energy and maximising the value of the global solar asset base.

REFERENCES

[1]. Barber, S., Sempreviva, A. M., Sheng, S., Farren, D., and Zappalá, D.: A use-case-driven approach for demonstrating the added value of digitalisation in wind energy, J. Phys. Conf. Ser., 2507, 012002, https://doi.org/10.1088/1742-6596/2507/1/012002, 2023c.

[2]. Booshehri, M., Emele, L., Flügel, S., et al.: Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis, Energy and AI, 5, 100074, https://doi.org/10.1016/j.egyai.2021.100074, 2021.

[3]. Clark, T.: How to communicate and collaborate on data: easy-to-use tools and techniques for eliminating overwhelm, confusion and ambiguity, Zenodo, https://doi.org/10.5281/zenodo.7928816, 2022.

[4]. De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., and Rosati, R.: Using Ontologies for Semantic Data Integration, Springer International Publishing, Cham, 187–202, https://doi.org/10.1007/978-3-319-61893-7_11, 2018.

[5]. Graybeal, J., Jonquet, C., Fiore, N., and Musen, M. A.: Adoption of BioPortal's Ontology Registry Software: The Emerging OntoPortal Community, 13th RDA Plenary Meeting, Philadelphia, April 2019.

[6]. Grieves, M.: SME Management Forum Completing the Cycle: Using PLM Information in the Sales and Service Functions, SME Management Forum, Troy, MI, 2002.

[7]. Gruber, T. R.: A translation approach to portable ontology specifications, Knowl. Acquis., 5, 199–220, https://doi.org/10.1006/knac.1993.1008, 1993.

[8]. IEA: Renewables 2024 — Analysis and Forecast to 2030, International Energy Agency, Paris, 2024.

[9]. IRENA: World Energy Transitions Outlook 2023: 1.5°C Pathway, International Renewable Energy Agency, Abu Dhabi, 2023.

[10]. Klonari, V., Papachristos, G., and Fraile, D.: Solar energy digitalisation: current status and future directions, Solar Energy Reports, 8, 100107, 2023.

[11]. Marykovskiy, Y., Clark, T., Day, J., Wiens, M., Henderson, C., Quick, J., Abdallah, I., Sempreviva, A. M., Calbimonte, J.-P., Chatzi, E., and Barber, S.: Knowledge engineering for wind energy, Wind Energ. Sci., 9, 883–917, https://doi.org/10.5194/wes-9-883-2024, 2024.

[12]. SolarPower Europe: Global Market Outlook for Solar Power 2023–2027, SolarPower Europe, Brussels, 2023.

[13]. Studer, R., Benjamins, V. R., and Fensel, D.: Knowledge engineering: Principles and methods, Data Knowl. Eng., 25, 161–197, 1998.

[14]. W3C: Semantic Sensor Network Ontology — W3C Recommendation, World Wide Web Consortium, https://www.w3.org/TR/vocab-ssn/, 2017.

[15]. Wagg, D. J., Worden, K., Barthorpe, R. J., and Gardner, P.: Digital Twins: State-of-the-Art and Future Directions for Modeling and Simulation in Engineering Dynamics Applications, ASME J. Risk Uncertainty Part B, 6, 030901, https://doi.org/10.1115/1.4046739, 2020.

[16]. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al.: The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3, 160018, https://doi.org/10.1038/sdata.2016.18, 2016.