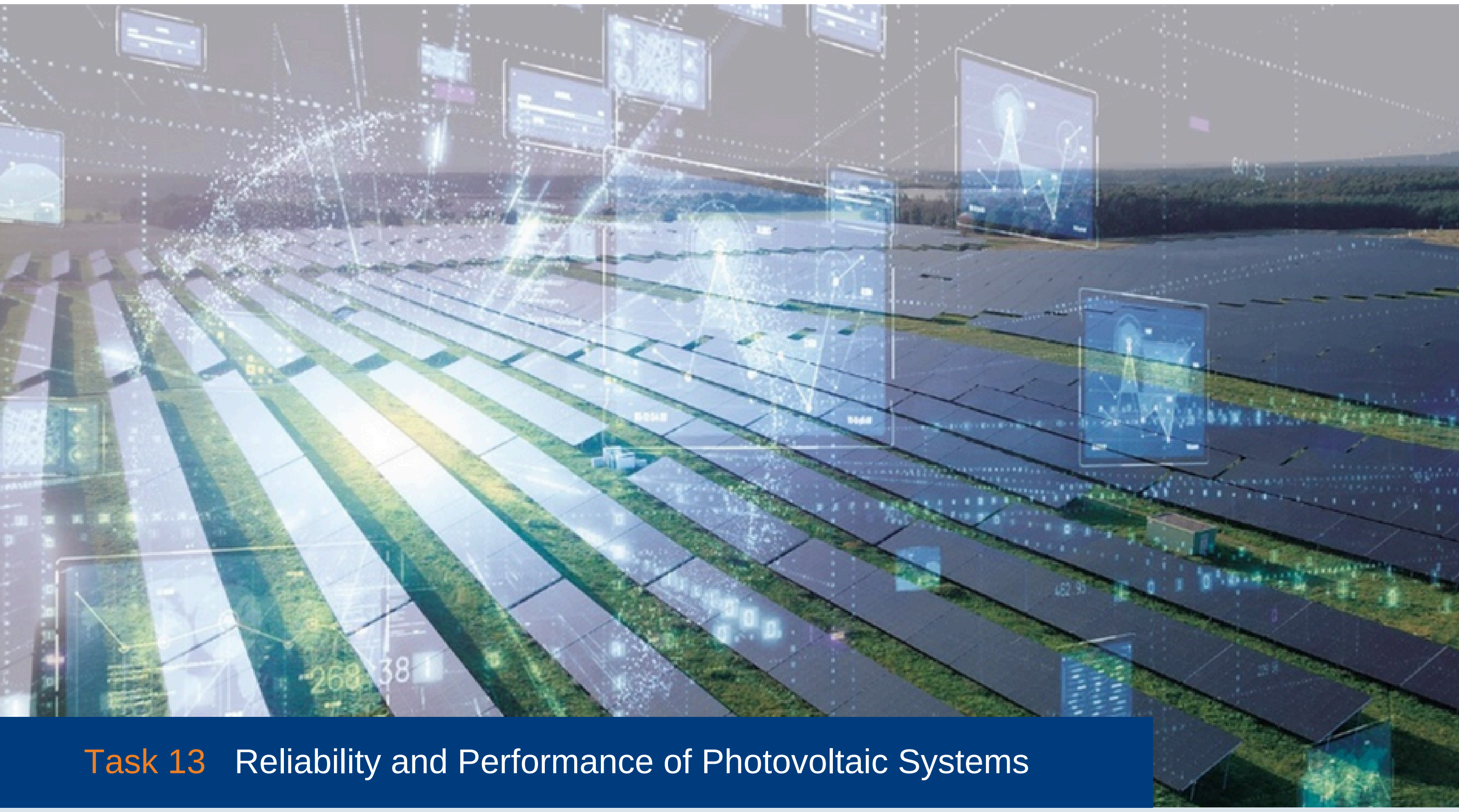
Technology Collaboration Programme
by IEA

International Energy Agency
Photovoltaic Power Systems Programme



Task 13 Reliability and Performance of Photovoltaic Systems

PVPS

Digitalisation and Digital Twins in Photovoltaic Systems

2026

Report IEA-PVPS T13-34:2026

What is IEA PVPS TCP?

The International Energy Agency (IEA), founded in 1974, is an autonomous body within the framework of the Organization for Economic Cooperation and Development (OECD). The Technology Collaboration Programme (TCP) was created with a belief that the future of energy security and sustainability starts with global collaboration. The programme is made up of 6.000 experts across government, academia, and industry dedicated to advancing common research and the application of specific energy technologies.

The IEA Photovoltaic Power Systems Programme (IEA PVPS) is one of the TCP's within the IEA and was established in 1993. The mission of the programme is to "enhance the international collaborative efforts which facilitate the role of photovoltaic solar energy as a cornerstone in the transition to sustainable energy systems." In order to achieve this, the Programme's participants have undertaken a variety of joint research projects in PV power systems applications. The overall programme is headed by an Executive Committee, comprised of one delegate from each country or organisation member, which designates distinct 'Tasks,' that may be research projects or activity areas.

The 28 IEA PVPS participating countries are Australia, Austria, Belgium, Canada, China, Denmark, Finland, France, Germany, India, Israel, Italy, Japan, Korea, Lithuania, Malaysia, Morocco, the Netherlands, Norway, Portugal, South Africa, Spain, Sweden, Switzerland, Thailand, Türkiye, the United Kingdom and the United States of America. The European Commission, Solar Power Europe and the Solar Energy Research Institute of Singapore are also members.

Visit us at: www.iea-pvps.org

What is IEA PVPS Task 13?

Within the framework of IEA PVPS, Task 13 aims to provide support to market actors working to improve the operation, the reliability and the quality of PV components and systems. Operational data from PV systems in different climate zones compiled within the project will help provide the basis for estimates of the current situation regarding PV reliability and performance.

The general setting of Task 13 provides a common platform to summarize and report on technical aspects affecting the quality, performance, reliability and lifetime of PV systems in a wide variety of environments and applications. By working together across national boundaries, we can all take advantage of research and experience from each member country and combine and integrate this knowledge into valuable summaries of best practices and methods for ensuring PV systems perform at their optimum and continue to provide competitive return on investment.

Task 13 has so far managed to create the right framework for the calculations of various parameters that can give an indication of the quality of PV components and systems. The framework is now there and can be used by the industry who has expressed appreciation towards the results included in the high-quality reports.

The IEA PVPS countries participating in Task 13 are Australia, Austria, Belgium, Canada, Chile, China, Denmark, Finland, France, Germany, Israel, Italy, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, Thailand, and the United States of America, and the Solar Energy Research Institute of Singapore.

SUGGESTED CITATION

Louwen, A., Schill, Ch. (2026). Louwen, A., Schill, Ch, Bruckman, L., Jahn, U. (Eds.), *Digitalisation and Digital Twins in Photovoltaic Systems* (Report No. T13-34:2026). IEA PVPS Task 13. https://iea-pvps.org/key-topics/t13-digitalisation-twins-pv-systems-2026/ DOI: 10.69766/RMPH3089

COVER PICTURE

Digitalised power plant, Fraunhofer ISE, Overlay metamorworks Shutterstock

INTERNATIONAL ENERGY AGENCY

PHOTOVOLTAIC POWER SYSTEMS PROGRAMME

Digitalisation and Digital Twins in Photovoltaic Systems

IEA PVPS
Task 13
Reliability and Performance of Photovoltaic Systems

Report IEA-PVPS T13-34:2026
February 2026

ISBN: 978-1-7642902-6-5
DOI: 10.69766/RMPH3089

AUTHORS

Main Authors

Atse Louwen, Eurac Research, Italy
Christian Schill, Fraunhofer ISE, Germany
Lluvia Ochoa, TotalEnergies, France
Nikola Hrelja, TotalEnergies, France
Roger H. French, Case Western Reserve University, USA
Erika I. Barcelos, Case Western Reserve University, USA
Mengjie Li, University of Central Florida, USA
Magnus Herz, TUV Rheinland, Germany
Bernhard Kubicek, AIT, Austria
David Dassler, Fraunhofer CSP, Germany

Contributing Authors

Franz Baumgartner, ZHAW Switzerland

Editors

Christian Schill, Fraunhofer ISE, Germany
Atse Louwen, Eurac Research, Italy
Laura S. Bruckman, Case Western Reserve University, USA
Ulrike Jahn, Fraunhofer CSP, Germany

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This paper received valuable contributions from several IEA-PVPS Task 13 members and other international experts.

The contributors to the report have received funding of their work through several projects and funding bodies, as listed below.

This report is supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under contract no. 03EE1120B.

This contributions of RHF, ML and EIB are supported by under the US Department of Energy's under Award Number(s) DE-NA0004104 and DE-EE0009347.

LIST OF ABBREVIATIONS

AC	Alternating Current
AI	Artificial Intelligence
BFO	Basic Formal Ontology
BIM	Building Information Model
BMWK	Bundesministerium für Wirtschaft und Klimaschutz, Germany
CCO	Common Core Ontologies
CSV	Comma Separated File
DC	Direct Current
DO	Digital Object
DOI	Digital Object Identifier
DT	Digital Twin
EL	Electroluminescence
EoL	End-of-Life
EPC	Engineering, Procurement and Construction
ERP	Enterprise Resource Planning
FAIR	Findable, Accessible, Interoperable and Reusable
FMEA	Failure Mode and Effect Analysis
GW	Gigawatt
IEA	International Energy Agency
IEC	International Electrotechnical Commission
IIoT	Industrial Internet of Things
IoT	Internet of Things
IR	Infrared
IRR	Internal Rate of Return
I-V	Current Voltage
JSON-LD	JavaScript Object Notation for Linked Data
KPI	Key Performance Indicator
LCOE	Levelized Cost of Electricity
LeTID	Light and elevated Temperature Induced Degradation
LLM	Large Language Model
MDS-Onto	Materials Data Science Ontology
ML	Machine Learning
NPV	Net present Value
NRMSE	Normalized Root Mean Squared Error
O&M	Operations and Maintenance
OWL	Web Ontology Language

PID	Potential Induced Degradation
PV	Photovoltaic
PVPS	Photovoltaic Power Systems
R&D	Research and Development
RDF	Resource Description Framework
SCADA	Supervisory Control and Data Acquisition
SKOS	Simple Knowledge Organization System
SWRL	Semantic Web Rule Language
TCP	Technology Collaboration Programme
UAV	Unmanned aerial vehicle
URI	Uniform Resource Identifier
VIS	Visible Light
WEEE	Waste Electrical and Electronic Equipment
WWW	World Wide Web

GLOSSARY

3D model (of a PV system)	A 3D computer model of a photovoltaic (PV) system is a three-dimensional representation that visually depicts the components and layout of a solar power generation system. This model typically includes elements such as PV modules, inverters, mounting structures and other auxiliary equipment.
Basic Formal Ontology (BFO)	Upper-level ontology that is designed for use in supporting information retrieval, analysis and integration in scientific and other domains.
Common Core Ontology (CCO)	A mid-level ontology consisting of 11 component ontologies covering many areas that map terms and concepts to BFO
Critical (Energy) Infrastructure	Assets, facilities, or systems, whether physical or virtual, considered vital, whose incapacity or destruction would severely impact national security, stability, or energy supply
Cyber Threats	Circumstances or events that could negatively impact information systems through unauthorized access, such as ransomware, phishing, malware, destruction, exploitation, disclosure, or manipulation
Cybersecurity	Protection of digital infrastructure (including hardware, network, accounts, data, and software) against unauthorized access and damage
Data Flow	The exchange of (standardized) information between the elements of a Digital Twin (between virtual and physical entity)
Data Model	Model of attributes and their relations within a domain
Data Structure	The method of organizing and storing data in a computer system.
Data-driven models	Statistical (machine learning, regression) approaches to learn system behaviour from a large amount of data
Digital Model	A Digital Model is a digital representation of a physical entity or system that captures its attributes and behaviour through various data structures and models. Unlike a Digital Twin, which continuously updates and interacts with real-time data and allows for simulations and decision-making, a Digital Model may not have these interactive capabilities and can be static or limited in its dynamic representation.
Digital Shadow	A Digital Shadow refers to a digital representation of a physical entity that is updated with real-time data but lacks the interactive capabilities found in a full Digital Twin. Unlike a Digital Twin, which not only reflects the current state of a system but also allows for simulations, data analysis, and decision-making, a Digital Shadow primarily serves as a passive repository of data that

	mirrors the physical entity's state without active feedback or control mechanisms.
Digital Twin (DT)	“A digital twin is a virtual representation of a PV system or systems, at the appropriate level of detail, that can span its lifecycle, is updated from real data, and uses simulation, machine learning or reasoning to help decision making.” [Task13A2.4]
Hybrid model	Combines data driven and physics equations
JSON-LD	JavaScript Object Notation for Linked Data. A Lightweight easy to read and write Linked Data format used to encode FAIR data.
Ontology	Ontology is the branch of metaphysics dealing with the nature of being, and it defines the names of things and their relationships. Therefore, an ontology is a set of concepts, categories and classes in a subject area or domain that shows their properties (as does a taxonomy) and the ontology also defines the relationships between these classes.
Physical device	A Generalisation of the concept “PV” (a PV System, a fleet, a component) (“the thing you can touch”)
Physical Entity	“A box for everything, that can be physically touched and is being described via attributes in a data model” *“Physical entities refer to the study object ontologies and their ancillary resources. Bevilacqua et al. [51] proposed a digital twin reference model, and the physical entities consist of physical industry resources such as products, personnel, equipment, material, process, environment, and facility.”*
Physics-based (or -driven) models	Models to describe the physical sibling based on physical formula (2-diode-model etc). (physics-approximated models)
PMDCo	Lightweight mid-level ontology for materials science and engineering (MSE) that maps terms and concepts to BFO
PV component	The components that a PV system is comprised of
PV Fleet	A group of PV systems
PV System	A single PV asset
Python	Programming Language Python
R	Programming Language R
Resource Description Framework (RDF)	RDF is a World Wide Web Consortium (W3C) standard used for describing and exchanging data represented as triples. The newest version is referred to as RDF-star and is being released in 2025.
Rule-based model	A rule-based model is a system that makes decisions or predictions based on a set of predefined rules or heuristics. These rules dictate how the system should respond to specific inputs or conditions. Rule-based models can be found in various domains.
Semantic Reasoning	Semantic reasoning is the ability of a system to infer new facts from existing data based on inference rules or ontologies. Can also be referred to as Machine Reasoning.

Semantic Web	The Semantic Web is an extension of the current World Wide Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation
Semantics	Semantics is the branch of linguistics and logic concerned with meaning.
Taxonomy	A taxonomy is a scheme of classes to which things can be allocated. It defines names, terms or keys for these organizational classes.
Virtual Entity	Conceptualize the “real thing” by a set of models (physical equations or data driven models)

EXECUTIVE SUMMARY

The report "Digitalisation and Digital Twins in Photovoltaic Systems" provides a comprehensive overview of the transformative role of digitalisation within the photovoltaic (PV) sector, particularly through the integration of digital twins and advanced data models. We underscore the necessity of adopting digital technologies to enhance the operational efficiency, reliability, and overall performance of PV systems throughout their lifecycle.

Digitalisation is increasingly recognized as a crucial driver in the evolution of the PV industry, enabling stakeholders to improve decision-making processes, optimize system designs, and facilitate predictive maintenance. By integrating and leveraging vast amounts of data generated during the lifecycle of PV systems—from manufacturing through to operation and maintenance—digitalisation helps address critical challenges related to efficiency and sustainability in energy production.

Central to this transformation is the concept of the digital twin. We adopt the following definition:

“A digital twin is a virtual representation of a PV system or systems, at the appropriate level of detail, that can span its lifecycle, is updated from real data, and uses simulation, machine learning or reasoning to help decision making.”

Digital twins enable stakeholders to simulate various operational scenarios, analyse performance under different conditions, and predict maintenance needs. This capability is particularly valuable in enhancing the reliability and performance of PV systems by allowing for proactive risk management and informed decision-making.

The report emphasizes the importance of robust data models and structures, which are foundational to the successful implementation of digital twins. The establishment of standardized taxonomies and ontologies is essential for ensuring data interoperability and facilitating effective data sharing among various stakeholders in the PV industry. This is particularly relevant in the context of the Materials Data Science Ontology (MDS-Onto), which aims to unify terminologies and improve data integration across the sector. A critical aspect of interoperability is the so-called “FAIRification” of data (i.e., making data findable, accessible, interoperable, and reusable).

Moreover, the report highlights the significant advancements in digitalisation along the entire PV value chain, from the manufacturing of components to the operational phase. Each stage benefits from digital tools that enhance automation, data analytics, and real-time monitoring, thereby reducing costs and improving performance outcomes. The integration of Internet of Things (IoT) technologies and artificial intelligence (AI) further enhances these capabilities, allowing for continuous improvement in PV system management.

We also highlight the challenges that remain, particularly in achieving full integration of digital processes across the PV value chain. Many current digitalisation efforts are still disconnected, and achieving a cohesive digital strategy requires collaborative efforts from all stakeholders involved. Additionally, the report stresses the critical need for effective cybersecurity measures as reliance on interconnected digital systems increases, emphasizing that robust security protocols must be integrated into every phase of digital twin development and deployment.

In conclusion, the report outlines the potential for digitalisation, particularly using digital twins and robust data models, to support the photovoltaic sector. By optimizing operational efficiency, enhancing predictive maintenance capabilities, and enabling better decision-making processes, digitalisation can not only improve the performance and reliability of PV systems

but also contribute significantly to the broader goals of sustainability and energy transition. The continued advancement and adoption of innovative digital technologies will be crucial in shaping the future landscape of the PV industry.

1 INTRODUCTION

Digitalisation has emerged as a transformative action across various sectors, and the photovoltaic domain is no exception. This process encompasses the integration of digital technologies into all aspects of PV systems, from design and manufacturing to operations and maintenance. At its core, digitalisation enhances efficiency, reliability, and performance, enabling stakeholders to utilize and analyse vast amounts of data and information generated throughout the lifecycle of PV systems. The usage of digital twins (DTs)—a virtual representation of physical assets—serves as a pivotal tool in this context, facilitating real-time monitoring and simulation of PV systems' behaviour under varying conditions, facilitating decision support.

The significance of digitalisation in the photovoltaic domain cannot be overstated. It allows for improved decision-making, as stakeholders can simulate different scenarios, optimize designs, and predict maintenance needs based on data analytics. Digitalisation also fosters greater transparency and collaboration among stakeholders, including manufacturers, operators, and researchers, ultimately driving innovation and reducing costs.

Moreover, digitalisation plays a facilitating role in addressing the challenges of climate change and energy transition. As the world increasingly shifts towards sustainable energy sources, the efficiency and reliability of PV systems become paramount. Digital tools and technologies enable the continuous improvement of PV system performance while ensuring compliance with quality and regulatory standards. By adopting digitalisation, the photovoltaic sector can enhance its competitiveness and increase its contribution to global sustainability goals.

The following sections of this report will delve deeper into specific aspects of digitalisation in PV, including the role of digital twins, data models, and the implications for system reliability and performance. Here, sections 3 and 4 on data models and the definition of DTs are specifically targeted towards software developers and AI experts.

Digital twins integrate data from various sources, including operational data from PV systems, weather conditions, and historical performance metrics. This integration enables stakeholders to assess risks and make informed decisions. For example, by employing machine learning algorithms on monitoring data, operators can predict potential failures before they occur, thereby reducing downtime and maintenance costs.

The role of data models and structures is crucial in this process. Well-defined data models ensure that the information flowing through the digital twin is accurate, consistent, and easily interpretable. Ontologies provide a means in standardising terms and improving data interoperability across the PV sector, enabling the FAIRification of data (i.e., making them findable, accessible, interoperable and reusable). This standardisation is essential for effective data sharing among different stakeholders, including manufacturers, operators, and researchers.

Moreover, the digitalisation of the PV domain facilitates enhanced risk analysis. By leveraging digital twins, stakeholders can quantify risks related to component failures, design flaws, and environmental factors impacting system performance. This proactive approach to risk management contributes to the overall reliability and profitability of PV projects.

In summary, digitalisation, particularly through the implementation of digital twins and robust data models, represents a significant advancement in the photovoltaic domain. It not only optimizes operational performance but also fosters innovation and collaboration among industry stakeholders. Digitalisation in the photovoltaic (PV) sector encompasses various key aspects that contribute to the optimization and efficiency of PV systems throughout their lifecycle. It involves several key elements that enhance the efficiency and effectiveness of PV systems.

First, automation has revolutionized PV manufacturing, using robotics and artificial intelligence to streamline production processes and reduce costs. Next, data analytics plays a crucial role, enabling real-time system performance monitoring and predictive maintenance through the analysis of large datasets.

Digital twins serve as virtual replicas of PV systems, allowing for real-time performance monitoring and the simulation of various operational scenarios, which helps optimize system performance and maintenance planning. Interoperability is also vital, with standardized data models facilitating seamless data sharing among stakeholders.

Risk management is enhanced through digitalisation, as advanced monitoring systems collect and analyse data to quantify risks associated with component failures. Predictive maintenance strategies leverage historical performance data and real-time analytics to forecast potential failures, reducing operational costs and improving system availability.

Enhanced monitoring technologies, such as drones and IoT devices, improve inspection accuracy and speed, while cybersecurity measures are increasingly important to protect sensitive data and maintain operational integrity.

Finally, user engagement is fostered through digital platforms that provide stakeholders with accessible data and analytics, empowering informed decision-making regarding PV systems.

The glossary summarizes all the relevant terms and definitions that are used throughout the report.

2 DIGITALISATION IN THE PV SECTOR

In this chapter, we discuss the role of digitalisation in the PV value chain. First, we show how digitalisation efforts are applied in different stages of the value chain. Next, we discuss a key benefit of digitalisation, which is enhanced risk analysis. Finally, we introduce the digital twin as the core concept in PV digitalisation.

2.1 Digitalisation along the lifecycle of a PV project

As discussed in the introduction, this report focuses on the role of digitalisation in the PV sector and presents digital twins as a core concept in this development. The ideal digital twin combines dataflows from the whole PV system value chain into actionable information for PV stakeholders. However, as we will discuss in this report, disconnected digitalisation initiatives are ongoing in all PV system value chain stages, and the full integration of data along the value chain is an effort that still requires huge efforts from all the involved stakeholders.

In the following subsections, we detail these “disconnected” digitalisation efforts for key phases in the PV value chain, from manufacturing of PV system components, via development, Engineering, procurement and construction (EPC), operation and maintenance (O&M) to end-of-life of the PV system.

The fragmentation of digital tools along the PV project lifecycle leads to information being trapped at each stage, undermining efficiency and impeding strategic decision making. It is characterised by an interoperability of tools from one phase to the other, and impedes information flows between different stakeholders involved and collaboration between them. To fully leverage the value of digitalisation and the knowledge it can generated, there is a need for standardisation, interoperability within and along the PV project lifecycle.

2.1.1 Manufacturing Stage

The first stage of the PV value chain, the manufacturing of PV system components, is arguably the stage where digitalisation has been applied most thoroughly already. The drastic cost reductions in PV components over the last decades, for instance, module costs declining by over 90% since 2000, can be explained by massive advancements in the PV manufacturing industry. The sector’s annual market size grew from a few hundred megawatts in the year 2000 to an estimated number of more than 500 GW in 2024 [1]. This means that every year, billions of PV modules and hundreds of billions of PV cells are being manufactured, and the market for balance-of-system (BOS) components like inverters, mounting systems and tracking systems grew accordingly. Therefore, digitalisation of manufacturing became key to manage such an enormous industrial development. In the manufacturing stage, digitalisation is focused in three key areas: automation, inspection and tracking, and process optimisation.

Underlying the overall development in automation and efficiency improvements in PV manufacturing is the application of the *Industry 4.0* concept [2]. Combining automation tools with concepts as artificial intelligence, Industrial Internet of Things (IIoT), data flows from production equipment sensors as well as inline inspections tools have enabled the nearly fully automated production of PV cells and modules and have resulted in drastic improvements in overall process efficiency and yield, as well as enabling the application of technological improvements in terms of novel cell processes and design, enhanced material efficiency from wafer to module, higher process throughput and substantial cost reductions due to huge economies of scale as typical manufacturing facilities grew by several orders of magnitude.

The PV manufacturing industry is the first example in this report to showcase the applications of a digital twin. While this implementation is completely different than the digital twin which we will discuss in detail in Chapter 4, its fundamental definition is the same: a digital entity that replicates a physical entity, in this case the PV manufacturing line. These virtual replicas of physical manufacturing lines enable precise simulation of process alterations and their impacts on final product specifications without interrupting the physical production, creating large savings in resources and enabling more streamlined optimisation of design and planning.

A related example of using the digital twin concept in PV module manufacturing is presented in the work of Lüer et al [3]. This study proposes the application of a digital twin for PV materials research. By leveraging machine learning approaches, this study mimics chemical and physical materials properties using a digital twin to identify improvements and novel materials that could potentially be used to enhance PV module performance, balancing efficiency, longevity and recyclability. Developing such a material digital twin would enable the identification of novel materials and processes that could allow for improved PV module performance balancing efficiency, longevity and recyclability [3].

The role of digitalisation and the mechanisms it enables (like automation) is not limited to the manufacturing of PV modules, but also applies to manufacturing of BOS components like inverters and mounting systems. While the cost reductions in the PV module industry have been the result of very specific improvements in processes, material quality, tools and automation, the innovations in BOS components are more a result of innovations in other industries like the electronics industry and general advancements in automated manufacturing [4].

2.1.2 Design/Development Stage

In the design and development phase of the PV system lifecycle, digitalisation plays a significant role in all the key activities. Commercially available design software and services, and the high-quality datasets they provide, allow system developers to perform site and yield assessment and system design, ultimately aiming for the development of more reliable and cost-effective PV systems. The application of digitalisation has not only served to enhance efficiency by substantially reducing the required development time but has also improved precision of techno-economic assessment. These improvements combined allow for optimisation of the PV system to achieve the highest possible yields. More recently, software tools used in the early design and development phase are being further developed to allow for more efficiency in next phases of the PV system lifecycle, leading to quicker and more precise engineering designs, easier onboarding of PV system data in monitoring platforms and the creation of digital twins from the early design phase.

The initial step in PV system development is site assessment, where digital tools have significantly enhanced the accuracy and efficiency of evaluating and selecting potential locations. Advanced Geographic Information Systems (GIS) and remote sensing technologies enable detailed analysis of solar irradiance, shading, and terrain [5]. These tools provide high-resolution data which helps identifying optimal sites with minimal shading and maximum solar exposure. Several commercial suppliers of solar irradiance data also provide relevant weather data allowing more detailed modelling of PV plant lifetime yield to be used in financial evaluations. In addition to resource and yield assessment, commercial tools also include a broad variety of GIS datasets for locating suitable sites for greenfield PV development. This includes data layers on local grid connection points, terrain slope/relief, local electricity pricing data and trends thereof affecting financial viability, and many more data layers that help developers minimise project risks.

More frequently, prospecting of site selection tools is integrated in system design tools for early and engineering design. Modern tools for system design allow users to quickly generate PV system layouts, often with integrated terrain data to be used in 3D designs. These tools enable developers to simulate PV performance in different system configurations to find the optimal system layout that maximises the financial gains. Design tools allow the user to specify the technical characteristics of the modules, mounting systems, module orientation and tilt angles (or in the case of tracking systems, tracker orientation and axis tilt), row spacing or ground coverage to estimate system performance including row to row shading or shading from nearby objects or distant terrain features. A review of PV system design software in 2019 concluded that commercially available PV system design software tools were lacking several important features, such as a lack of local meteorological and terrain data, unavailability of efficient 3D model creation and visualisation, and inability to perform PV system design optimisation automatically [6], among a comprehensive list of issues identified. Modern tools have largely overcome previous challenges by enabling the integration of comprehensive GIS datasets, meteorological and terrain data, as well as incorporating advanced functionalities such as automatic and iterative yield simulation. These developments facilitate more robust analyses and improve the reliability of outcomes.

Commonly integrated in the design tools, sometimes as the key feature, is PV system yield assessment, which is critical to determine the financial feasibility of the PV project under consideration. Digital yield assessment software such as PVsyst, SAM or PlantPredict [7] provide comprehensive simulation features to model PV system yield in detail, including losses due to shading, soiling, module mismatch, cabling losses, etc. These tools often have financial models integrated to evaluate economic KPI’s such as net present value (NPV), internal rate of return (IRR) and levelized cost of electricity (LCOE).

The use of digital tools applied to site identification, site assessment, early design and yield assessment contribute to the potential creation of a digital twin, a virtual representation of the physical system to be built. Especially the design and yield assessment tools could provide digital twins to other PV system lifecycle phases as an extension of their functionality, as these software tools are already commonly used and trusted. An example of this functionality was shown in the TRUST-PV project [8]. Digital twins enable designers to optimize system configurations, assess potential issues, and make data-driven decisions to enhance system reliability and efficiency [9]. In the design/development phase, the digital twin integrates solar resource and meteorological data with the system design to provide accurate yield simulations that include the effects of local weather conditions, self-shading, nearby shading and terrain shading and potential losses due to soiling and system electrical design for a detailed performance model that can be used throughout the system’s lifecycle [8].

2.1.3 Engineering, Procurement, and Construction

The application of digitalisation in PV has substantially improved practices in the Engineering, Procurement and Construction (EPC) phase. Digital tools used here improve efficiency and accuracy, and aid significantly in the management of the increasingly complex PV projects.

A key advantage of digitalisation in EPC is the enhancement of project design and planning. Many tools that aid in early design of PV systems also enable the creation of detailed engineering designs that combine terrain data, PV system layout over the whole plot, required cabling, etc. These computer-aided design tools [8] can provide an overview of all necessary components and materials and estimate e.g. how much groundworks are needed. These detailed designs also allow for the creation of digital twins combining different data streams. As an example, the TRUST-PV project demonstrated the creation of digital twins for new and existing plants, combining high-resolution terrain data from drone surveys with engineering designs from PV design software to create a digital twin that was used for detailed yield simulations [8]. Digitalisation practices such as Building Information Modelling (BIM) can provide a common data repository for managing, sharing and federating information throughout the PV system lifecycle [10]. This has the potential to make EPC and other phases of the lifecycle more efficient, and could reduce or eliminate work repetition, and enable information to be reused. However, many stakeholders still silo their data due to confidentiality or competition concerns.

Digitalisation also streamlines procurement processes, which become increasingly complex due to growing PV system sizes and a large selection of potential suppliers for all components including PV modules, inverters, mounting systems, cabling and interconnectors. As outlined in the SolarPower Europe EPC Best Practice Guidelines [11], the procurement phase is extremely complex, entailing supplier selection and qualification; product selection, qualification, and testing; supply review; and delivery monitoring and pre- and post-shipment inspection. The most important goals in this phase are quality assurance and risk management, and digitalisation and value chain data integration can provide important insights. For more details on this topic, refer to Section 2.2.

The construction phase again represents a complex sequence of events. Here digital construction management platforms aid in planning the overall construction activities, including the necessary component transport and (temporary) placement, civil, electro-mechanical and ancillary works [11]. Tracking all these operations is a key task for these software tools, using field input from e.g. smartphone applications, GPS equipment and other digital tracking solutions such as RFID tags and IoT sensors. Particularly for PV modules, the transportation, handling and installation can have profound effects on long-term performance and reliability. Pre-commissioning or post-installation checks and test are commonly conducted to ensure the installation was performed properly, and any damage incurred to PV modules is properly identified and logged. Aerial surveys using UAVs can aid in both mechanical, visual and performance testing by combining LiDAR data, visual imagery and infrared (IR) or electroluminescence (EL) imagery respectively. The detailed engineering designs integrated in digital platforms serve as the point of reference to compare the design with the as-built measurements. Finally, the EPC contractor is typically responsible for inspection and tracking the data of the system as it goes into operation.

2.1.4 Operational/Maintenance Phase

The operation and maintenance (O&M) phase is arguably the phase in the lifecycle of PV systems where digitalisation plays the most important role. In Chapters 4 and 5 this becomes more apparent as we discuss the definition and workings of the digital twin, and the main roles of digitalisation and the digital twin in PV O&M respectively. Dedicated digital O&M platforms aim to improve most O&M related activities, including (remote) monitoring, fault detection and diagnostics, predictive and corrective maintenance activities, performance and condition monitoring and analytics, automated analysis and acquisition of visual, infrared, and luminescence-based field inspection imagery data. The platforms can also be used for asset management and spare part management, and ticketing and workforce management activities are streamlined by applying dedicated or integrated digital toolsets.

Remote monitoring systems represent a significant advancement in the digitalisation of PV O&M. These supervisory control and data acquisition (SCADA) systems typically collect data from at the module string or inverter level, combined with data from meteorological stations to provide real-time visibility into plant performance [12]. Modern monitoring platforms transmit performance metrics including DC/AC power output, voltage, current, and environmental conditions to cloud-based platforms. Remote and automated monitoring systems can detect underperformance issues within seconds or minutes rather than days or weeks, enabling prompt intervention before significant energy losses occur. Effective remote monitoring can improve annual energy yields through faster fault detection and resolution [13] [14], and has been essential for improving system performance over the past decades [15].

Unmanned aerial vehicles (UAVs) with IR, EL or photoluminescence (PL) imaging cameras have greatly enhanced PV system inspections by quickly identifying hotspots, cell cracks, and other defects. For a 1 MW PV system, IR inspections can be conducted in 1 hour when conforming to relevant International Electrotechnical Commission (IEC) standards of image accuracy [16]. Nighttime, drone-based EL inspections of the same plant would take from 1 to 2 hours [17], which implies an increase in inspection speed by a factor of <5 compared to an inspection using an EL camera on a tripod [17]. The field of automated detection and classification of faults from IR or EL/PL imagery is very active, with numerous research groups presenting progress in acquisition, image segmentation and PV module identification, fault detection and fault classification [18] [19] [20] [21]. Drone based inspections are 50-70% cheaper than manual ones and offer more comprehensive data coverage [15] [22].

Hence, remote monitoring and automated image processing are prerequisites for automated fault detection and diagnosis (FDD), which is a rapidly developing area in the PV sector. FDD typically employs machine learning algorithms to automatically analyse PV performance data, identifying and potentially classifying faults in near real time. These automated systems significantly reduce the need for manual inspection [17] while improving detection rates for issues that might otherwise remain unnoticed until performance degradation becomes severe.

Data analytics is essential to transform raw operational data into actionable insights in monitoring and fault detection tasks. For example, Lindig et al. created a statistical analysis method to differentiate underperformance caused by degradation, soiling, and operational issues [23]. Data analytics coupled to comprehensive reporting capabilities also streamline the ability of asset managers and O&M operators to provide reports to regulatory bodies, asset owners and investors, reducing overhead costs while improving reporting accuracy and consistency [24] [25].

Digitalisation has also enabled a shift to predictive maintenance for PV systems. By analysing performance data, weather patterns, and observed degradation of component performance, predictive analytics can forecast failures and recommend timely maintenance. Predictive maintenance has several advantages that result in cost savings and reliability improvements, such as increasing the availability, energy production and performance of PV systems, reducing time for repairs, reducing spart parts replacement costs and reducing or eliminating the need for some maintenance activities [24]. For further details about predictive maintenance, refer to Section 5.3.4.

2.1.5 End-of-Life

Digitalisation could enable key advances in the end-of-life (EoL) management of PV systems, mainly by providing tools that generate the necessary data about the state-of-health of PV system components to make informed decisions on EoL management. It must be noted that reuse of PV components is still a topic in its infancy, and important challenges need to be addressed in policy, standardisation while financial viability is still questionable. Recent research provides an improved workflows driven by digitalisation that could partly address these challenges.

The first advancement enabled by digitalisation is data-driven decision-making. Digital tools enable the collection and analysis of data on a highly granular scale down to the module level. Within the TRUST-PV project, the application of wireless sensor networks and IoT technologies can facilitate real-time data collection of PV module level performance, degradation and failure modes [26]. The availability of these data is essential for taking the best course of action for each PV component and deciding, for example, whether to relocate/reuse, substitute and/or recycle existing PV modules.

The use of these module level monitoring devices, in combination with potential other on-site inspections enable a workflow that qualifies and triages PV components for reuse. As described by Tsanakas et al [27], this workflow has been developed in the TRUST-PV project and SolarPower Europe's Lifecycle Quality Workstream and aims to improve the cost-effectiveness of module reuse by the triage approach. The workflow begins with a desktop study using system or module monitoring data to assess if a site has modules suitable for reuse collection. Alternatively, all modules are dismantled and sent for recycling. If reuse is deemed generally feasible from the analysis of monitoring data, more detailed non-contact inspections using infrared (IR) imagery and electroluminescence (EL) imaging are carried out to further divide modules towards recycling or reuse. Finally, deeper technical checks determine if selected modules are safe and functional for reuse on the second-life market [28].

Whether fed into the second-life market or towards recycling, digital platforms and tools such as digital product passports and digital material passports create transparency in EoL of PV components that ensures compliance with regulatory requirements such as the Waste Electrical and Electronic Equipment (WEEE) Directive [28]. Additionally, digital records enable the certification and resale of second-life PV modules, which enhances market confidence.

On the system level, digitalisation can support repowering and EoL decisions through the same monitoring systems described in the previous section by monitoring the state of health on non-PV module components and a set of KPIs that enables decision making regarding repowering [REF]. Commercial PV system performance modelling tools also allow asset managers to evaluate system performance and cost benefits for different repowering scenarios. Some of these tools are marketing dedicated functionality for evaluating repowering, such as PVFARM[1].

2.2 Enhancing Risk Analysis in PV Projects through Digitalisation

The PV industry faces diverse risks across the entire lifespan of PV systems, starting from manufacturing all the way to their end-of-life. Quantifying these risks contributes significantly to guarantee the reliability and efficiency of these systems. Digitalisation is transforming how we assess risks in PV projects by making it easier to collect and analyse critical data. Technologies like drones and advanced monitoring systems help capture detailed operational parameters. Big data analytics and machine learning algorithms can screen large amounts of data to detect patterns and predict failures before they happen [29]. By integrating diverse datasets, digital risk assessment models can quantify technical risks much more accurately.

The following table presents different risks faced at each stage of the PV system lifecycle along with their potential consequences and methods to calculate and mitigate these risks [30].

[1] https://www.pvfarm.io/solutions/repower

Table 1: Risk Analysis Across the Lifecycle Stages of Photovoltaic Systems

Lifecycle Stage	Risk	Consequence	Risk Mitigation
Manufacturing	Defective solar modules and inverters due to errors in the manufacturing process	Lower energy output, reduced efficiency and increased failure of PV modules and inverters	Analysing data from quality control tests and historical defect rates to better predict and mitigate the risks of defects.
Design and development	Suboptimal system design (e.g., incorrect tilt angle, inadequate components)	Decreased energy yield, increased operational costs, and long-term efficiency losses of the PV system	Utilizing digital twins for computer-based simulations and performance modelling to optimize tilt angles, system configuration, and component selection during the design phase.
Transportation and installation	Damage to PV modules during transportation or improper installation of modules and BOS affecting performance	Physical damage can cause micro-cracks, leading to decreased module efficiency over time	Utilizing sensors and tracking systems to monitor the condition of components during transit and ensuring installation best practices through rigorous training and certification programs for installers.
O&M phase	Environmental factors (e.g., weathering, soiling, extreme weather) and insufficient maintenance practices resulting in accelerated module degradation or the progression of BOS component failures.	Reduced energy output, increased system downtime, higher repair costs, and diminished overall system efficiency.	Implement advanced monitoring systems to collect real-time environmental and performance data for predicting degradation rates and detecting potential faults. Utilize data-driven analytics in conjunction with maintenance records to establish predictive maintenance strategies, minimizing downtime and preventing critical failures in a timely manner.
End-of-Life	Challenges in recycling and disposing of aged PV modules and BOS	Increased environmental impact due to improper disposal	Developing comprehensive data on material composition to create effective recycling protocols and end-of-life management strategies.

2.2.1 Data flow risks and Importance of High-Quality Data

Data from various sources, such as O&M records, environmental sensors, and performance logs, form the backbone of precise risk analysis. Accurate data inputs support the reliability of predictive models and decision-making tools. This reliability depends on high-quality data flows, but risks such as timestamp misalignment, sensor inaccuracies and data gaps, can undermine their effectiveness [31]. Accurate and representative data is mandatory for key performance indicator (KPI) evaluation, predictive maintenance, and fault detection as discussed in section 5.3. The following recommendations aim to improve efficiency, data integration, and stakeholder collaboration across the PV system lifecycle. Key actions include:

- **Standardization and Interoperability:** Adopting taxonomies, ontologies (s. section 3.1) and standardized data formats and communication protocols is mandatory for smooth data integration and improves managing data across the entire lifecycle of PV systems. International standards such as IEC 61724-1 ("Performance Monitoring of Photovoltaic Systems") for sensor types, placements, and data formats ensure consistency across all monitoring systems while enabling easier integration with external data.

- **Investment in Digital Infrastructure:** Investing in advanced digital infrastructure significantly enhances the effectiveness and efficiency of data collection and analysis capabilities. Cybersecurity should be a priority in digital infrastructure investments.
- **Collaborative Efforts:** Strengthen collaboration among stakeholders—manufacturers, operators, researchers, and policymakers—to promote data sharing and the development of integrated risk management frameworks. [32]

The increasing reliance on interconnected digital technologies can lead to cybersecurity risks. PV systems, using IoT devices, cloud platforms, and digital twins, are vulnerable to unauthorized access, data breaches, and cyber-attacks, which can compromise operational continuity, grid stability, and data integrity. Implementing robust cybersecurity measures is highly recommended to effectively mitigate these risks. A deeper analysis of cybersecurity frameworks, standards, and mitigation strategies is provided in section 6.3 of this report.

2.3 The digital twin as a central concept of digitalized PV

In the previous subchapters, digitalisation along the lifecycle of PV projects was described, ranging from the manufacturing stage to end-of-life, and it was acknowledged at the beginning of the chapter the fact that many already digitalised processes along the lifecycle are fragmented and should be fully integrated in an interoperable way to fully leverage the benefits of digitalisation. This integration process not only requires a common understanding of the entities in the PV domain - an information model or an ontology, see chapter 3 – but also a dedicated information management and stewardship (storing, finding, accessing this information for the whole duration of PV projects, i.e., for several decades in varying levels of detail or levels of information). Once the underlying PV domain knowledge is abstracted and long-term data storage is ensured, the concept of digital twinning can realize its full potential to optimize the management and maintenance of PV systems.

Digital twinning in photovoltaic systems refers to the concept of creating a virtual representation of a PV system that can span its entire lifecycle. Chapter 4 will discuss in detail the definition of digital twins in the scope of photovoltaic projects.

The heart of digitalisation are digital twins, that provide virtual replicas of assets, enabling improved data collection, analytics, simulation, control, and information sharing, and ultimately improved and informed decision making. A digital twin (DT) integrates technologies such as artificial intelligence (AI) and the Internet of Things (IoT) to optimise decision making, thus playing a crucial role in the broader digitalisation of PV or future energy applications in general [33],[34].

The definition of a DT can vary depending on the context, such as the scope of the digital twin, the use of real-time data, or its adaptation to specific sectors or phases, like manufacturing, construction, or operation. However, the three main elements of the original definition remain consistent: the physical entity, which represents the selection and organization of information about the PV system; the virtual entity, which comprises models mirroring specific aspects of the physical entity; and the data flows between the real and virtual entities.

In this report, a variant of IBM's definition [35] is proposed, describing a DT as a virtual representation of a PV system that is continuously updated with real data and is used to support decision making by using simulation, machine learning, reasoning, or a combination thereof.

The DT concept can be applied throughout the lifecycle of PV systems, supporting activities such as planning, design, operation, and maintenance. For example, during the design stage, physics-driven models can be used to simulate the PV system based on historical weather data and component specifications. During the operation and maintenance stage, machine learning models can be developed for performance monitoring and maintenance using real-time weather and electrical data. Ultimately, subchapter 4.3 will discuss the concept of implementing a data driven DT not for a single PV system but for a whole fleet of systems.

3 THE ROLE OF DATA MODELS & DATA STRUCTURES

Data models and data structures represent a fundamental component of data science and analysis in organizations that handles streams of data. It defines the structure and organization of data and metadata to support organizational strategies and decisions [36]. In a data model, different data elements are arranged and organized in a way that facilitates their access and retrieval at any moment as well as the associated metadata information while assuring data integrity, proper data storage, analysis and provenance.

Taxonomy and ontologies are important concepts related to data models. Taxonomy is related to classification and categorization of concepts and provides the hierarchical structure of different data elements. Although taxonomies are important to structure knowledge and support access to information, they are static and provide limited information about relationships between concepts. Ontologies on the other hand, are dynamic and offer a richer knowledge representation where both data elements (variables and their names or “terms”) and their relationships are linked and can be utilized to enable semantic or symbolic reasoning.

In photovoltaics, a well-defined and consistent taxonomy is available as the Orange Button Taxonomy [37], [38]. Ontologies, on the other hand, are still in their initial stages of development in the PV field. In this chapter an extensive literature review on ontologies for PV and progress and advances in the field towards term and vocabulary standardization and knowledge representation are provided. A framework for ontology creation developed by some of the authors is also introduced along with initial ontologies of the PV domain. An open-source tool for creating FAIR (Findable, Accessible, Interoperable and Reusable) [39] [40] Linked data aligned with interoperable ontology terms previously defined by domain experts is also introduced in this report.

3.1 PV Taxonomies & Ontologies: Current Status & Literature Review

Ontologies are formal dictionaries used to represent knowledge in a domain. They were developed aiming to facilitate and enhance semantic information and knowledge sharing across different fields [41]. They were designed to contain information about core concepts (terms) and additional relevant information about terms such as the definition, alternative identifiers (Alt. Labels) and unique identifiers (the Uniform Resource Identifier: URI) [42], [43].

Ontologies have become extremely popular in different research communities due to their capabilities in articulating heterogeneous information and enriching contextual information. They extend the ability of taxonomies in categorizing and classifying concepts by introducing more semantic meaning, contextualization and relationships between core concepts. Ontologies also have a fundamental role in achieving the FAIR principles [39] for data, analyses and models, and for enabling semantic reasoning on data. In FAIR, ontologies are strongly connected to the interoperability principle, facilitating data integration and reuse by using standard and common vocabularies. Proper data integration enables data to be easily discovered, accessed and shared across different teams and research communities. The connection of ontology with the FAIR principles has motivated the rise, acceptance and adoption of ontologies across different research communities. Ontologies for PV expands upon previous experience of IEA PVPS in handling installation PV data from over 20 years where important considerations for rigorous inventories and fit-for-purpose PV installation data are discussed. They enable a more structure and robust way to organize and process data.

3.1.1 Ontologies presence and adoption in the PV Community

In the PV domain, standardization of terms and vocabulary initiatives started in 2016 with the Orange Button Taxonomy [2]. The Orange Button Taxonomy is a stand-alone taxonomy focused on financial contract interoperability for power plant developers, contractors and vendors. It was developed using existing solar standards and solar-finance vocabulary from industry subsectors like project finance, insurance, construction finance, and portfolio management. Orange Button is implemented as an OpenAPI specification [44], that streamlines and standardizes financial and design aspects in the PV value chain for designing, installing and operating power plants [45]. Designed to fully encapsulate data in the sectors, the taxonomy is also comprehensive and facilitates the reuse of its terms in different contexts without any need to define additional metadata.

The taxonomy is organized into hundreds of entry points, each of which provides a list of terms and relationships that are useful for a specific user or purpose. These entry points are grouped into three categories by the Orange Button Taxonomy: Data, Documents, and Processing. Through its broad coverage of terms related to the solar industry as well as its thorough organization of terms into different entry points for their seamless retrieval, the Orange Button Taxonomy is also a great resource for retrieving terms and definitions that can be incorporated into solar-related ontologies. Even though taxonomies are useful for terms matching, it does not capture all the relevant information in a domain as it does not provide relationships between different terms. Ontologies are therefore essential to achieve semantic interoperability as they introduce properties connecting concepts and facilitate knowledge understanding.

There have been solar ontologies proposed in the literature; **Error! Reference source not found.** lists some of the research where PV ontologies were proposed, with their sources and connections to mid and top-level ontologies. Those studies were often developed ontologies following the semantic web rule (SWRL) and employ methodologies for ontology constructing using semantic web knowledge-based systems. However, the terms created are not clearly mapped to existing mid- or top-level ontologies.

Table 2: Research where PV ontologies were proposed.

Ontology	Connection to Mid/Top level Ontologies	Overview	Tool
A Photovoltaic System Model Integrating FAIR Digital Objects and Ontologies, [46]	Acknowledge the importance of mapping that it might be addressed in the long run	This work proposes integration of FAIR Objects and Ontologies. Follow RDA recommendations for Kernel Information (KI) for describing digital objects (DOs)	Protege
Proposing an Ontology Model for Planning Photovoltaic Systems, 2021 [47]	No clear reported terms mapping to other ontologies	Proposed an ontology for planning of PV planning PV systems focus on maximum power point tracking (MPPT) method feature with SWRL	Protege
PV-TONS: A photovoltaic technology ontology system for the design of PV-systems, 2013 [48]	No explicit reported mapping to other ontologies, follow Ontology Development 101	Ontology developed for PV System and main components. It applies SWRL, but it does not explicit if there is a connecting to mid and top-level ontologies	Protege

Table 2 (continued)

Ontology	Connection to Mid/Top level Ontologies	Overview	Tool
Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis [49]	Maps to Basic Formal Ontology (BFO)	The ontology includes different modules covering specific aspects in the energy systems domain including social and economic aspects, models and data and the physical side of energy systems. It is not specific for PV	Protege
Ontologies as a Basis for Constructing Digital Twins in Energy [48]	No clear statement of mapping to mid/top level ontologies	Unclear how the ontologies are built and what are their components	Not stated
Materials Data Science Ontology (MDS-Onto): Unifying Domain Knowledge in Materials and Applied Data Science [50]	Mappings to mid-level ontologies and BFO	MDS-Onto sub-domain and domain ontologies map to MDS-Onto Concept that maps to mid-level Ontologies aligned to BFO	FAIRmaterials

Error! Reference source not found. References: [46], [47], [48], [49], [49], [51].

Additional reports of ontologies related to PV, but not specific to physical PV components can be found in the literature. Liu et al. [52], using 3LConOnt ontologies [53], improved the accuracy of data collection to build predictive models. However, these ontologies were isolated, with no direct mapping and their construction was unclear. Ontologies related to urban planning and household usage and urban planning have also been proposed [54], [55].

3.1.2 PV Ontologies creation and mapping to Mid- and Top-Level Ontologies

Regardless of the specific applications, no clear mapping, or term matching and alignment of PV specific ontologies to mid- or top-level ontologies is available [56]. If one develops and uses an "isolated"," core" ontology, then its broad applicability is challenging, since it probably conflicts with terminologies in other established ontologies being used by other parts of the domain's community - in our case the PV community. This means the fundamental challenge arises comparable to competing standards such as the VHS vs. Beta fights [57] of the past; as adoption of for example, an isolated PV ontology grows, conflicts with other PV ontologies used in other sectors, increases. By connecting all terms and relationships in an ontology up to the world wide web consortia's (W3C) schema.org[2] [58] and registering the ontology on an ontology portal [59], [60], then the initial ontology developers can check term and relationship mappings with other registered ontologies, and find resolutions early for term "clashes" (usually by adopting the other term). This process is referred to as ontological alignment and expands out of the single domain to all concepts even in other domains. By having an ontology that is registered on an ontological portal such as MatPortal [61], then all terms in the domain ontology can be connected through low-, mid- and top-level ontologies directly to schema.org, and

[2] https://schema.org

W3C's schema.org. Schema.org provides the common terminology, as defined by Simple Knowledge Organisation for the Web (skos), for all general terms used on the world-wide web [62], [63], [64]. The Semantic Web, initially proposed in 1993, and in active development to the present, has the goal of having the web contain entities that are both human and machine readable and actionable [65], [66]. Therefore data visible on the web, needs to be FAIR so that it is self-describing, and contains all relevant metadata, so as to form Linked Data, with its metadata expressed according to the Linked Data standards as XML or more commonly today as JavaScript Object Notation for Linked Data (JSON-LD) v1.1 [67], [68]. In addition, the next versions of RDF, SPARQL and JSON-LD, all referred to as RDF-star, SPARQL-star and JSON-LD-star, are in process to be released as a W3C recommendations during 2026 [69] [70]. The release of RDF-star and JSON-LD-star will simplify semantic reasoning by making it easier and more straight forward to make statements about statements (i.e., to make an RDF triple statement about another RDF triple) [71]. Ontology "registration", mapping and alignment is also important to guarantee a smooth interoperability between ontologies, which fundamentally important for the same domain and overlapping fields. An exception is the Open Energy Ontology [49] that was designed for the domain of "energy systems" and maps its terms to the ISO standard ontology BFO [72]. Even though it is not specifically applied to solar, it encompasses different aspects and concepts within the PV domain.

The main tool historically used to create ontologies is Protégé, as observed in **Error! Reference source not found.**. Protégé [67] is currently the most used open-source ontology software that allows users to create, edit and visualize ontologies. Its main capabilities include manually creating and editing ontological terms and relationships, visualizing ontologies, checking the logical consistency of ontologies, and querying ontologies for specific information. While Protégé has extensive functionalities which includes several plugins options, the complexity of the interface is a barrier for those who have little experience with ontology creation. This difficulty, especially for non-experienced users prevents certain researchers from creating and integrating ontologies with their own datasets entirely. In addition, adding terms and properties requires user manual input, which becomes a tedious task for systems with many variables. Therefore, there is a need for a tool that can create ontologies with an interface that is easily understandable and provides ample documentation on how to use it in addition to a PV ontology with clear mapping to top and mid-level Ontologies.

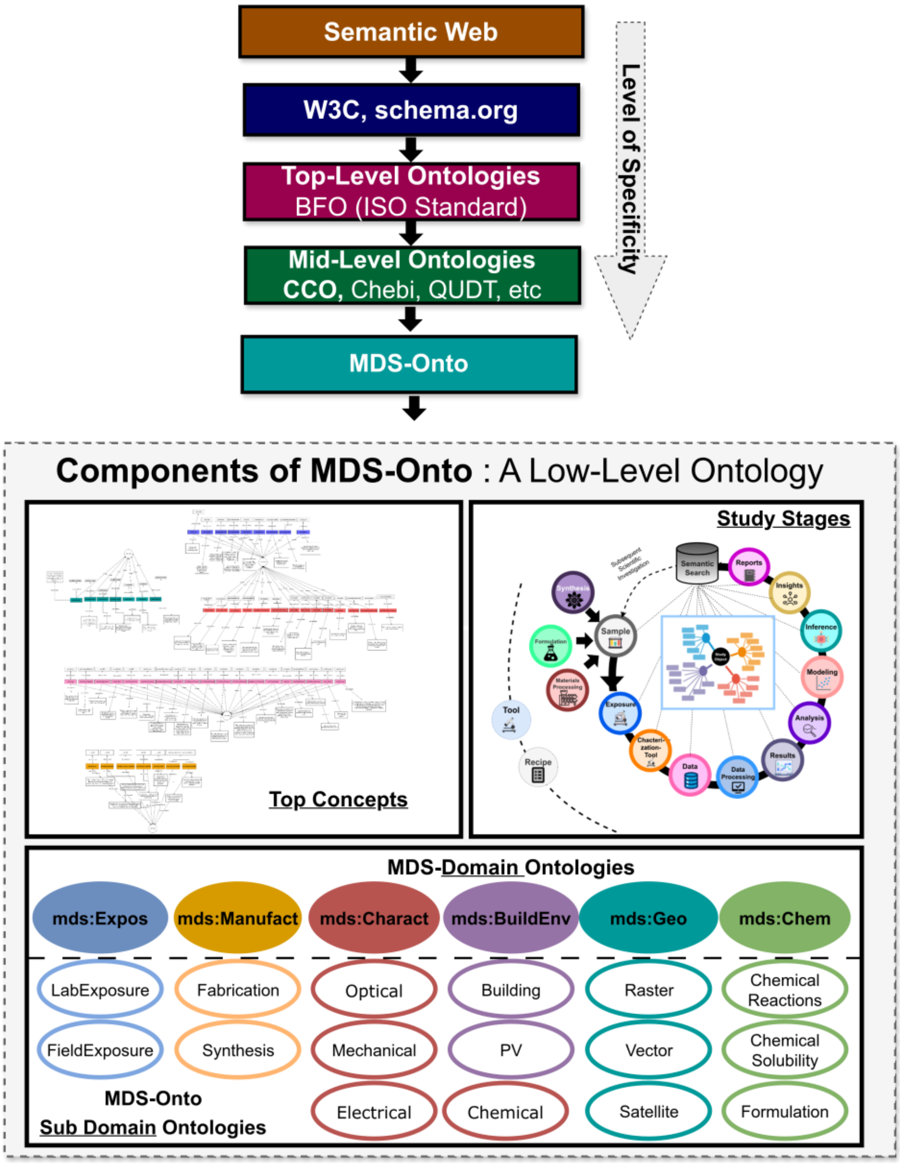


Figure 1: Positioning of MDS-Onto and PV-Onto within the semantic web. PV-Onto as well as MDS-Onto maps to Common Core Ontologies (CCO) that are connected to the Basic Formal Ontology (BFO).

3.2 Towards a Recommendation for Data Modelling: MDS-Onto to overcome the barriers of lack of terminology

To overcome the challenges associated with the lack of consistency in terminologies in the PV fields, we have developed, and herein describe, the Materials Data Science Ontology (MDS-Onto) [50], a modular, low-level and extensive ontology, which also has an associated MDS-Onto Framework which simplifies the process of ontology creation, validation, documentation and visualization, as well as creation of FAIR linked data. The MDS-Onto Framework includes CEMENTO [73], a python package for Ontology creation based on XML diagrams [74] [75], FAIRLinked [76], a python package for FAIR data creation, MDS-Onto *Open website* [77] which hosts documentation, `.owl`, `.json` files and a server for JSON-LD validation and visualization [78] and a WebVOWL server for visualizing ontologies as dynamic graphs [79] [80]. As a component of MDS-Onto, we introduce PV-Onto, the PV domain ontology consisting of multiple subdomain PV ontologies. Figure 1 illustrates MDS-Onto and where they sit in the ontology domain. PV-Onto consists of sub-domain ontologies under mds-BuiltEnv domain.

3.2.1 CEMENTO

CEMENTO [73], [81] is a python package that generates ontologies based on Extensible Markup Language(XML) drawio diagrams [82]. When designing ontologies, users often draw diagrams or schemas to organize concepts and ideas. CEMENTO takes advantage of this natural step in the Ontology design process to streamline the process of creation. CEMENTO can also take ontology files as input and output a drawio diagram. This functionality is particularly useful when users want to enrich an existing ontology part of MDS-Onto; in this case they can either add boxes with new terms or add them directly in the ontology file. CEMENTO uses by default Common Core Ontology [83], a ISO standard mid-level Ontology that is BFO-compliant. This simplifies subclassing and ontology mapping, as users can map their terms using a CCO human-friendly label, which CEMENTO automatically replaces with the corresponding CCO term.

Users can also map their recently created terms to MDS-Onto Concept. The MDS-Onto Concept is a bridge layer between CEMENTO and the domain ontology that was created to facilitate the mapping process. The concept layer contains terms that are intuitive and user friendly. For instance, the class *mds:SiteLocation* is a subclass of *mds:Location* (an existing Concept term). *mds:Location* is a subclass of *cco:GeospatialLocation.* Users have the flexibility to map their terms to MDS-Onto Concept or CCO, depending on their level of familiarity with Ontologies. Since the Concept ontology has been formally mapped to CCO, both methods are acceptable and ensure proper interoperability

Figure 2 illustrate how CEMENTO operates using as an example one term from the PV site ontology for mds;SiteID, a subclass of mds:identifier.

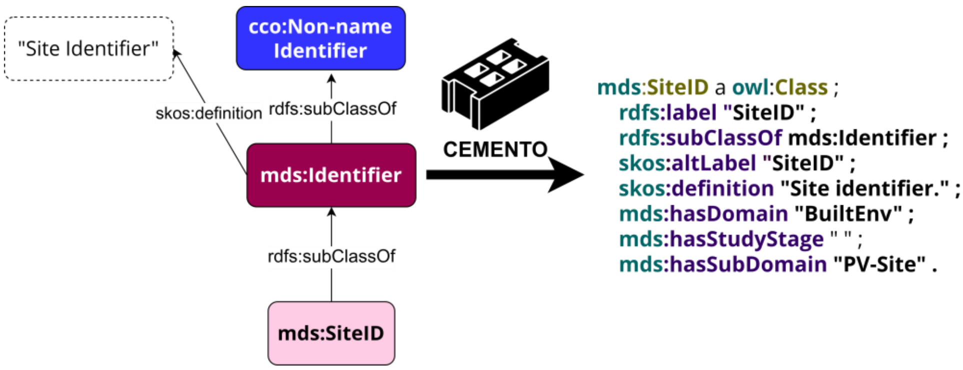


Figure 2 Example of how CEMENTO converts XML drawio diagrams to Ontology files (in turtle format). mds:SiteID has been mapped to mds:Identifier (term part of MDS-Onto Concept) that has been previously mapped to CCO.

3.2.2 MDS-Onto Open Documentation

MDS-Onto Open website [84], as illustrated in Figure 3, is an open-source tool that provides an easy-to-navigate interface that allows users to search for any ontology that has been created through the MDS-Onto Framework and download the MDS-Onto Ontology. In addition to providing MDS-Onto and its components in multiple syntaxes, the website also includes the URI and documentation for terms in every MDS-Onto domain ontology as well as a static visualization of the ontology.

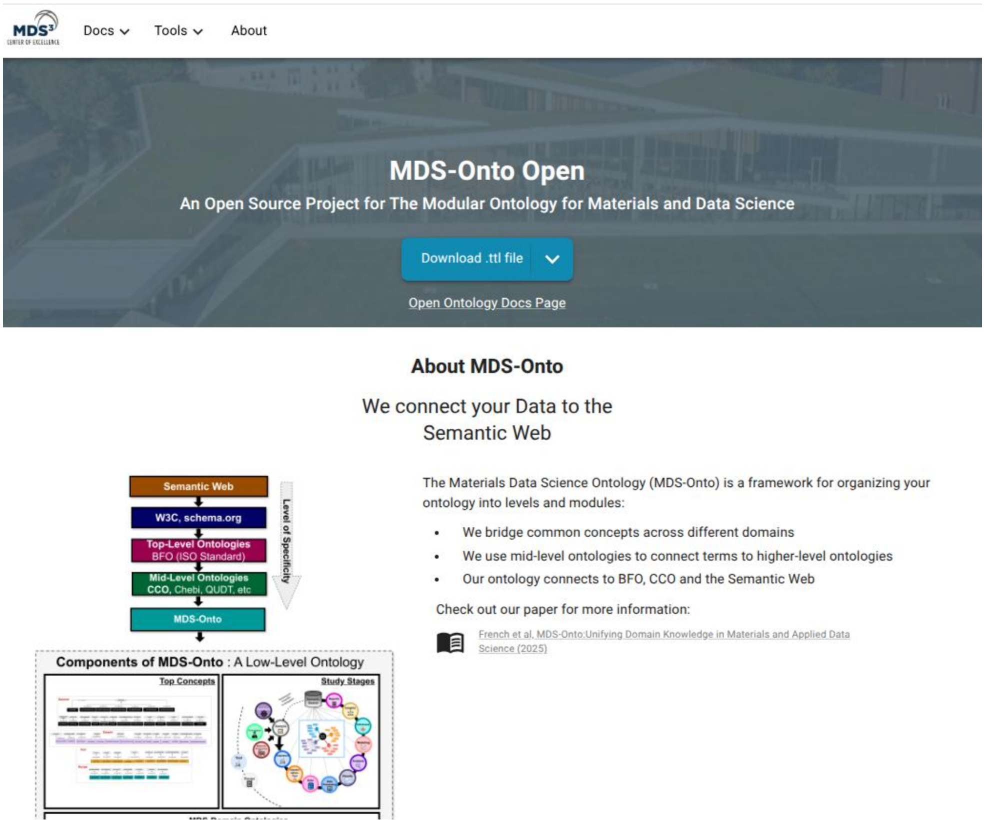


Figure 3: Interface of MDS-Onto Open website.

Furthermore, the website consists of a variety of resources for ontological analysis and visualization, including tools like WebVOWL and JSON-LD Playground, and extensive documentation on how to use the CEMENTO and FAIRLinked Packages.

3.2.3 MDS-Onto Ontology (v.0.3.0.0)

The Materials Data Science Ontology sits directly below other mid-level ontologies in level of specificity. It unifies concepts from various Materials Science domains by integrating domain-specific terminologies under the same overarching low-level ontology.

Each of these domain ontologies can be used for model training, validation, and testing, as well as pre- and post-processing approaches used in different materials science datasets. To create MDS-Onto, we connected specific terms and relationships to a “bridge layer” mid-level Ontology that maps to preexisting generalized concepts, most of them falling into shared concepts with CCO or other mid-level [85] ontologies such as CheBI [86] or QUDT [87]. Users can also map terms directly to CCO as an alternative to MDS-Onto Concept. The ontology also includes a variety of entities to tag different types of data (as primary data, secondary data, or metadata) as well as sample documentation on integrating terms to the MDS-Onto concept ontology or CCO with a variety of different forms of data, making it a prime mid-level ontology choice for materials and data sciences ontologies [83]. Figure 4 introduces the modular and extensive approach of MDS-Onto Framework [50] [88][89][90].

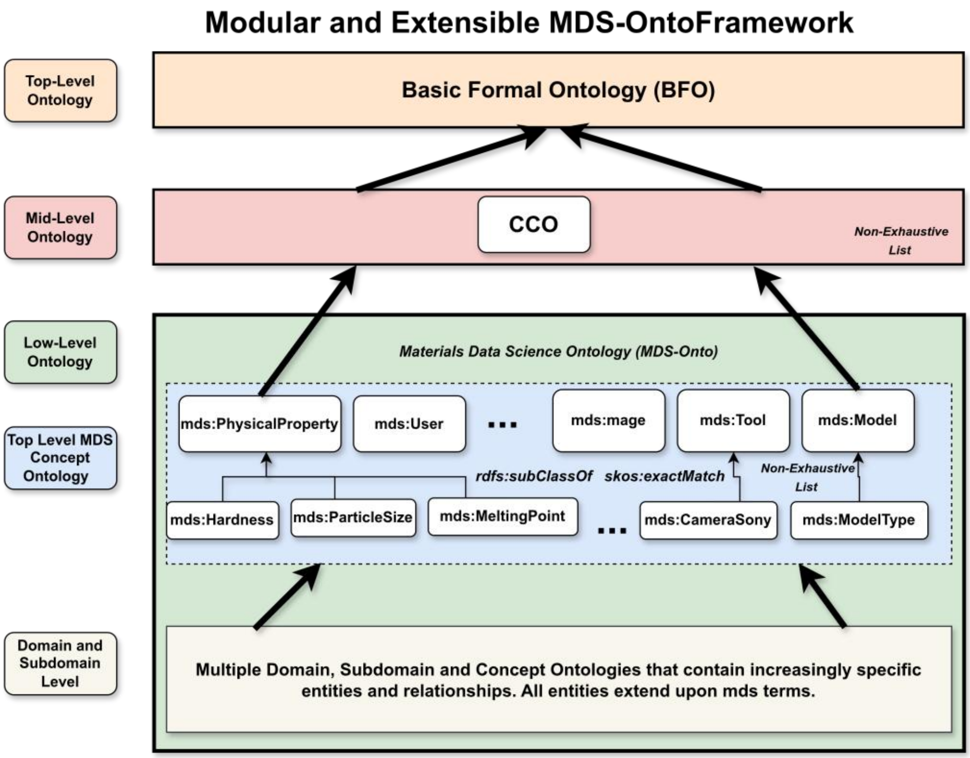


Figure 4: The modular approach of MDS-Onto. Users can map terms to MDS-Onto Concept, a bridge layer connection CCO and domain ontologies.

The newest version of the entire MDS ontology currently version (0.3.1.12) have been published to two ontology repositories: MatPortal [59] and Industry Portal [90]. MatPortal and Industry Portal are publicly accessible ontology catalogues and repositories, having a vast collection of Ontologies for Industry and Materials Science communities. Those tools represent a great resource for Ontology discovery, terms mappings and comparisons.

The website is a clone of the OntoPortal that has been formed from the original BioPortal [59] which consists of over seventy ontologies and engages the biomedical community to review and help improve ontologies via its ontology development features. By publishing our ontology to MatPortal, we are enabling the Materials Science community to review our ontology and invent potential mappings between our terms and terms from other ontologies.

3.2.4 FAIRLinked

Ontologies have a key role in unifying terms and concepts in the PV community and creating ontology-aligned RDF datasets. In addition to enhancing semantic interoperability in the PV domain, they are also fundamental to achieve FAIR data. When the FAIRification process is driven by unified and interoperable concepts for terms, developing pipelines that are able locate, reuse and share the date become a more reliable and streamlined process.

Expanding upon the CEMENTO package for ontology creation, we have developed and published FAIRLinked, a Python package that enables the creation of linked and FAIR data guided by ontology concepts. FAIRLinked operates in synchrony to CEMENTO. CEMENTO is initially used to create ontologies, which are subsequently merged to MDS-Onto. FAIRLinked reads the updated MDS-Onto and the user's dataframe and generates a FAIRified equivalent dataframe, where terms have been replaced by the corresponding MDS-Onto ones, and JSON-LD files populated with data. The JSON-LD files are linked data format that can be easily accessed and integrated into data pipelines. Figure 5 illustrates CEMENTO and FAIRLinked workflow for ontology and FAIR data creation.

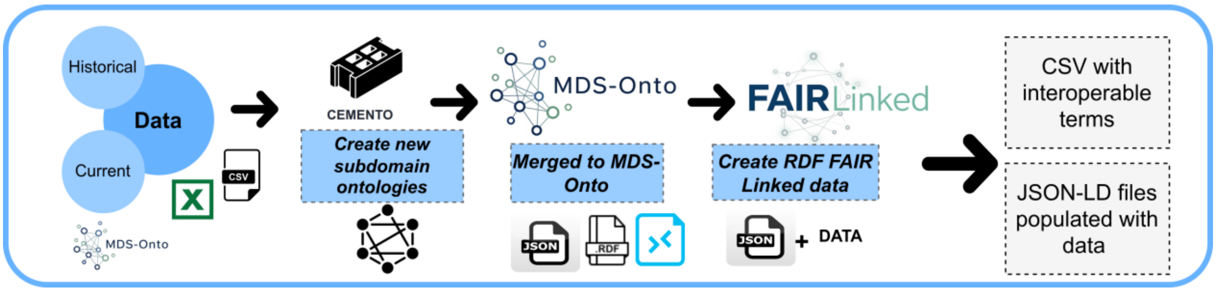


Figure 5: Integrated framework to create FAIR linked data and Ontologies using CEMENTO and FAIRLinked.

3.3 Interoperability of existing and new data

In PV, power plant assets are frequently changing hands. This leads to loss of information including data, metadata and knowledge during the different stages of the ownership transition. In addition, PV instrumentation is often different across manufacturers in a way that it becomes a challenge to translate the raw data into its real representation of the physical world in a way that is consistent and reliable. Initiatives like Orange Button, are important for categorization of PV domains, however, without significant user buy-in, it becomes challenging for a standards-based approach to become adopted and accepted in the PV supply chain. The modelling side of PV also faces challenges given the large diversity of software packages and specific requirements for data input, for example as pvlib-python [43], PVSyst [44], and SAM [45]. Reusing these packages requires additional effort from the user on identifying similarities and differences and implementing additional codes and scripts. As a result, this process leads to loss in time and resources.

From a laboratory perspective, technicians and researchers face many challenges when performing different PV measurements including the assessment, performance and degradation of PV modules. Due to different instrumentation, and experimental and operational conditions used, it's often challenging to reproduce scientific investigations especially when source, conditions, raw metadata, methodologies and procedures are not properly recorded or recorded at all.

All those facts illustrate the importance and growing need of terminology unification at all stages in the PV life cycle and value chain, from data collection, research, experiments, analysis and modelling, for all PV applications. The lack of this leads to slow collaborations and innovation.

Historically, the best solution considered for the lack of terminology and consistency within a scientific domain were taxonomies. However, taxonomies cannot comprehensively and fully describe relationships between concepts and therefore restricts the reasoning within the domain. Ontologies overcome those barriers by adding an additional layer of semantic meaning describing relationships and concepts. Ontologies also provide the foundation for semantic reasoning and open the opportunities of machine reasoning over historical results [91].

To overcome the challenges associated with the lack of consistency and terminologies in concepts in the PV community we extend the MDS-Onto framework to the PV domain, where PV domain ontologies are created, validated and shared with the community of MatPortal, Industry Portal and MDS-Onto *Open website*.

3.4 PV Domain Ontology: Unifying terminologies in PV with MDS-Onto

Ontologies provide a powerful tool for PV power plants as they allow data to be preserved with the correct identifiers and metadata properly recorded at all scales, from PV cell to the PV power plant site. From an asset owner's perspective, encapsulating all of the relevant information into a few key JSON files is extremely valuable, as it provides an easy way for information transfer when PV assets change hands. And using RDFlib and the Apache Arrow packages, JSON-LDs are easily converted to dataframes for analysis [92]. On the modelling aspect, asset owners are frequently concerned with the desirable performance of a PV installation; this requires a detailed digital twin of the system to be developed using FAIR data and metadata. PV-Onto is a sub-domain ontology that sits under the mds-BuiltEnvironment domain ontology. It is formed by six different components that facilitate the creation and organization of those sub domain ontologies: PV Cell, PV Module, PV backsheet, PV Inverter, PV Site, Charge Controller and Battery. A short description of each PV subdomain ontology is provided next [50].

<u>PV System Ontology</u>: A PV Site must have one or more PV Systems. For example, a Site may have multiple similar, but isolated, systems that tie in with different inverters, or a single large grid-connected system and an auxiliary research or quality control system.

<u>PV Inverter Ontology</u>: PV inverters are essential to convert the DC power generated by PV Systems into AC power to be used for grid export. A utility-scale System will have many inverters, as each inverter is only rated for a set amount of input DC power before it reaches its capacity. When an Inverter reaches its maximum DC power input, the remaining power is curtailed and lost. Knowing the DC power limits of the Inverter is crucial for identifying time periods of clipping and where additional hardware may be needed.

<u>PV Module Ontology</u>: The majority and most important information on a PV System depends on the PV Module used. There is a large variety of brands and manufacturers of Modules, and performance depends greatly on different factors that can be recorded in the domain ontology. Module manufacturers tend to have similar, but not standardized, terminology; module testing agencies like the California Energy Commission (CEC) also tend to have such terminology [93].

<u>PV Backsheet Ontology</u>: PV Modules have an PV Backsheet, which is a multilayer polymer laminate component on the backside of the module. In a Module specification sheet, backsheet

materials and compositions are often not explicitly described from the manufacturer. If this information is available, it is valuable to be included, as backsheets can significantly impact how, and the rate of degradation of PV Modules.

PV Cell Ontology: A PV Cell is the smallest unit of a PV Site and is the semiconductor-based p/n junction device that can convert light into electricity. Approximately 95 % of PV Cells manufactured in the world are made using crystalline silicon (c-Si) wafers, and these PV Cells are generally identified by the "so-called" cell technology, or architecture, that dictates the manufacturing processes used to produce the cells [94], [95].

Battery Ontology: A PV System may have some form of Battery storage. Batteries have traditionally been used in PV systems as a form of emergency redundancy; in this operation mode, batteries are maintained full and only discharged during loss of service events. More recently, batteries have become more resilient and less expensive, which makes them more suitable for grid service use, such as for frequency stabilization. Batteries can be filled by PV systems during peak hours and then deployed at sunrise or sunset to battle the "duck curve" problem of solar PV generation.

Charge Controller Ontology: Battery systems require a Battery Charge Controller, which determines at which conditions batteries are charged or discharged. Controllers can operate either from the grid level or be islanded away into the PV System. It is useful for an asset owner to be able to record and store the details of the Charger Controller, as the field has not been standardized yet, and therefore there is a significant variance in how charge controllers might behave.

Figure 6 illustrated a snapshot of the ontologies for PV Site and Module. The images are also available at high resolution at osf [80].

3.5 Conclusions and Takeaways

As the scientific community generates larger amounts and more diverse types of data, there is an increasing need for data standardization, to enable seamless data sharing, integration, accessibility and reuse. The main barrier hindering and slowing the interoperability and longevity of data is the variation in how research groups and industry store, organize, structure and chose to label their data.

Ontologies provide a solution to unify the organization’s data and its labelling by facilitating different terms or aliases for concepts and enable interoperability with other experiment-specific variables. While the ontology creation process has already been streamlined in specific fields of science and applications, there is still significant variation in how ontologies are constructed, across the fields of Materials Science and PV. This variability prevents ontologies from achieving interoperability with one another, hindering, effective and reliable cross-collaboration between research groups, and delaying the use of semantic, or machine, reasoning over historical results.

We have addressed this variability by introducing an ontological framework for the Materials and Data Sciences community, entitled the MDS-Onto Framework, which standardizes fundamental aspects and concepts of the Materials Data Science Ontology development process and enables the creation of FAIR data. The framework’s goal is to standardize the main components of ontology creation, including the positioning of MDS ontologies, including PV-Onto the semantic web, the knowledge representation language and the online locations for where those ontologies are published.

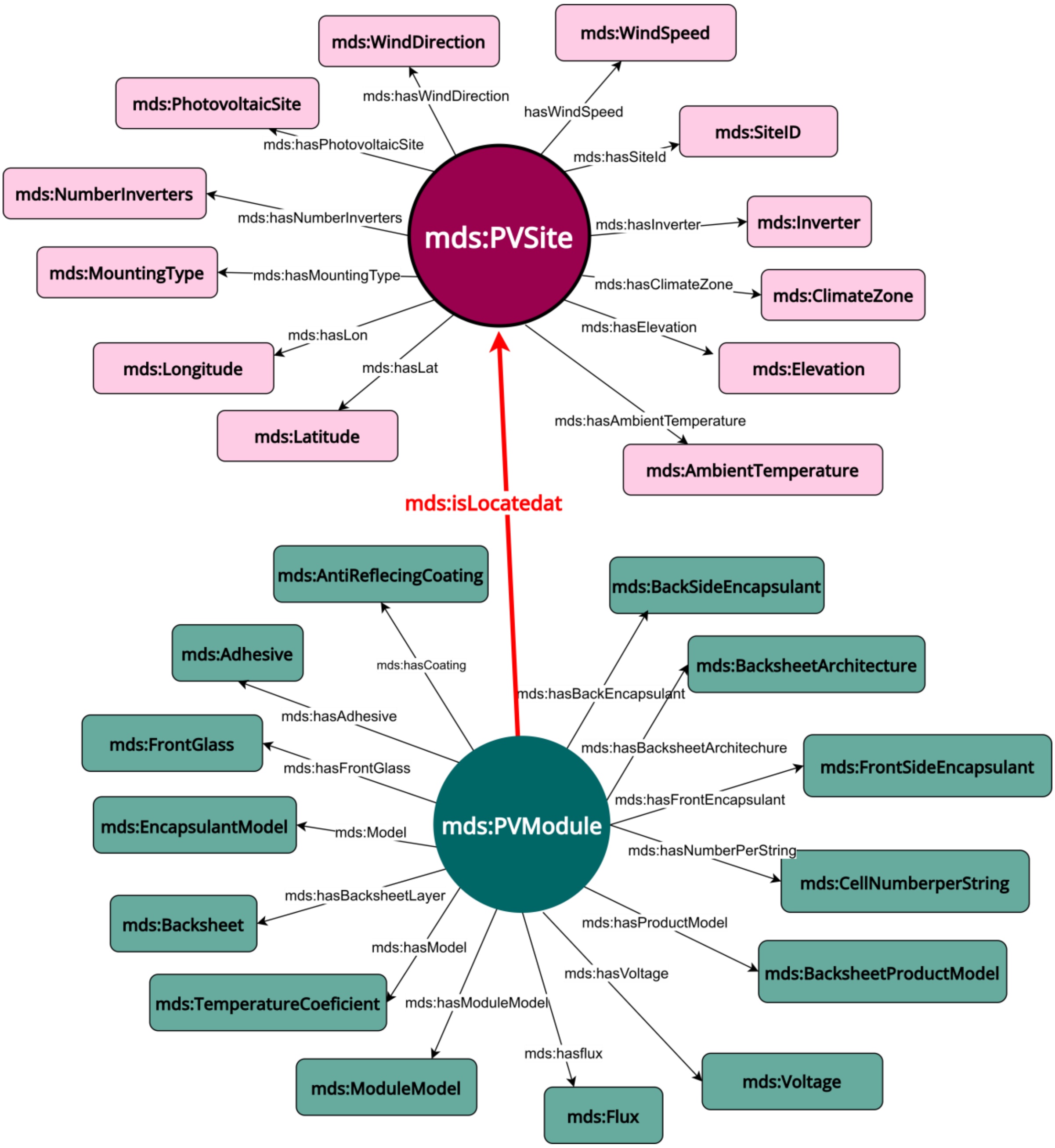


Figure 6: Ontologies for PV module and PV Site illustrated by the respective figures of terms, relationships and classes.

4 DEFINITION OF DIGITAL TWINS IN PV

4.1 Introduction

The concept of Digital Twins (DT) was first introduced in 2003 to refer to a “Conceptual Ideal for Product Lifecycle Management” [96].

In this definition, three key elements were present: the physical entity, the virtual entity and the data flows between the real entity and the virtual entity, as shown in Figure 7. In this case, the physical entity comprises object ontologies, taxonomies and data models (see Chapter 3 and Section 4.2.1) whilst the virtual entity consist of the model or sets of models that represent specific aspects of the physical entity; e.g. physical or behavioural models (see section 4.2.2 and Chapter 5).

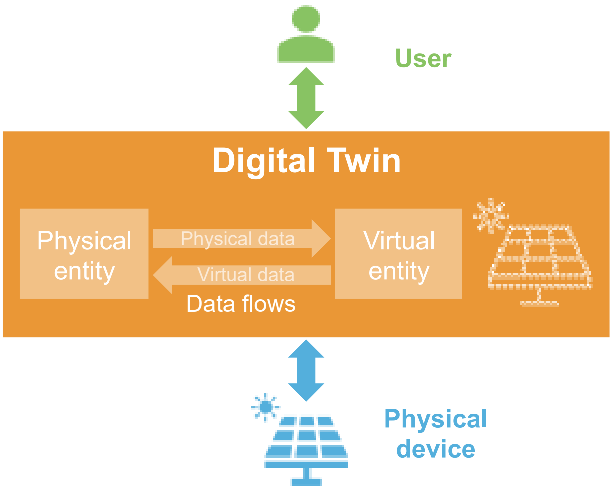


Figure 7: Main elements of a Digital Twin: the physical entity, the virtual entity, and their data flows.

Nowadays, the definition of “Digital Twin” can take multiple variations depending on the context. For example, by explicitly indicating the scope of the DT along the entire asset’s lifecycle [97], [35], a scope “fit for purpose” [98], the use of real-time data [35]; or by adapting the DT concept to the manufacturing [98], construction [99] and PV energy [97] sectors.

Despite the variations in the definitions of the DT, it remains widely accepted that the DT consists of the key three elements of the original definition. Considering these elements and the DT’s purpose to support decision making, the authors propose to adopt a variant of IBM’s [35] definition for this report:

“A digital twin is a virtual representation of a PV system or systems, at the appropriate level of detail, that can span its lifecycle, is updated from real data, and uses simulation, machine learning or reasoning to help decision making.”

In this context, the DT can support the planning, design, operation, and maintenance (O&M) activities in the lifecycle of PV systems. Examples during the design stage include the use of physics-driven models to simulate a PV system before construction based on historical weather data and device specifications data. During the O&M stage, applications include developing machine learning models for performance monitoring and maintenance using real-time weather and electrical data. This DT definition allows the selection of different modelling strategies across the system's lifecycle. For instance, physics-driven models during the design stage, and physics- and data-driven models during the O&M stage.

The DT concept proposed in this report is comprehensive enough; that is, it is not limited to the use of real-time sensor data and is not restricted to using a particular modelling approach. However, it indicates that its main purpose is to provide insights for decision making. This latter characteristic differentiates the DT from simpler models, e.g., process control models, aimed for very specific tasks that may not involve human reasoning.

Figure 8 presents a classification of the DT based on their level of integration; namely, Digital Model, Digital Shadow and Digital Twin [100]. In this classification, a Digital Model does not use any form of automated data exchange, a Digital Shadow only has an automatic data flow from the physical entity to the digital entity, whilst the DT has automatic data flows from and to the physical and virtual entities. In these cases, the DT in Figure 8c represents the most complex integration level. Note that this classification is not used in this report to distinguish the level of integration of the various examples provided, and they are all referred to as Digital Twins.

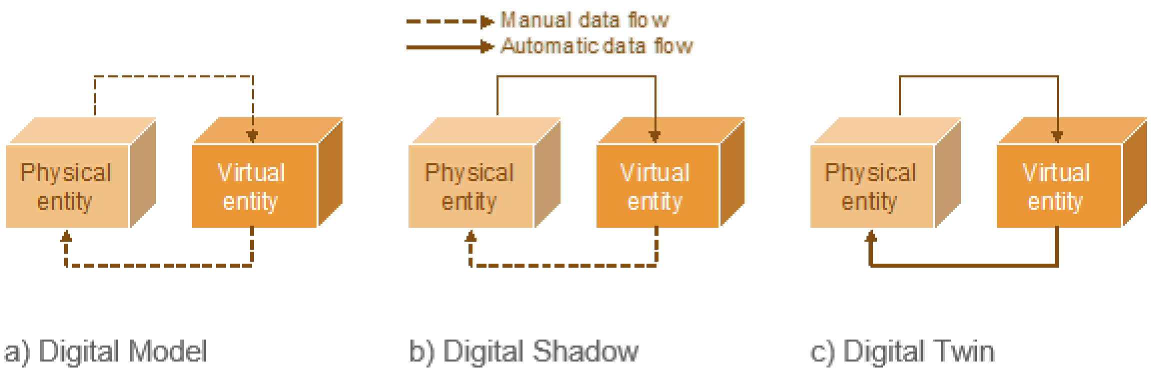


Figure 8: Classification of the DT based on their levels of integration. Figures taken from [100].

4.2 Components of a Physics-based Digital Twin for a PV System

This section defines the three main elements of the DT. Namely, the physical entity, the virtual entity and the data flows between the real entity and the virtual entity.

4.2.1 The physical entity

Physical entities are the foundation to construct the virtual entity. They represent the selection and organization of information related to the physical asset or group of assets in a DT. Physical entities consist of ontologies and taxonomies [101] (see section 3.1). In this respect, ontology can be defined as a "formal specification of a shared conceptual model" [102] with the following characteristics:

- Describes the relationships between the different concepts associated to a specific area or domain.
- Describes in detail such system of concepts.
- Is accepted and used by a specific community.
- Is standardized in a formal language.

The definition of an ontology depends on different factors, such as the purpose and scope of the DT, its target users and the digital resources used. For example, an ontology aimed for the design and construction of a PV plant may focus more on the detailed representation of devices and plant layout, as production data are not yet available. Similarly, an ontology aimed for managers of a PV fleet may focus more on comparing plant-level power data than on comparing those of individual PV devices. Finally, a DT deployed in a low-cost, low-maintenance infrastructure may require a simple ontology involving the minimal use of data.

Several data ontology approaches have been proposed by the PV community. The Orange Button Initiative, while primarily offering a taxonomy for PV, is an active community of public and private actors that proposes a comprehensive data ontology aimed to facilitate the sharing of data between PV energy companies across the lifecycle of PV systems [103]. The TRUST-PV project [97] developed a taxonomy for the PV sector and proposes using a data ontology for the entire lifecycle of a PV system based in the concepts of Building Information Modelling (BIM) described in [104]. For more detailed review of data ontologies, refer to section 3.2.

4.2.2 The virtual entity

The virtual entity refers to the model or sets of models that represent specific aspects of the physical entity. Figure 9 illustrates the four modelling aspects proposed by [105] that can be considered when creating a DT. Namely, geometric, physical, behavioural and rule modelling aspects. It is important to note that a virtual entity can comprise all or less of these aspects depending on the purpose of the DT. For example, having a 3D replica of a PV system may not be needed if other sources of data are available to describe the PV system's behaviour.

These four modelling aspects also comprise different modelling technologies as the type of data and functional requirements differ significantly. Examples of these modelling technologies and applications in the PV sector are also provided in Figure 9 and in chapter 5 .The geometric modelling can also include georeferenced digital plans of the system and its components.

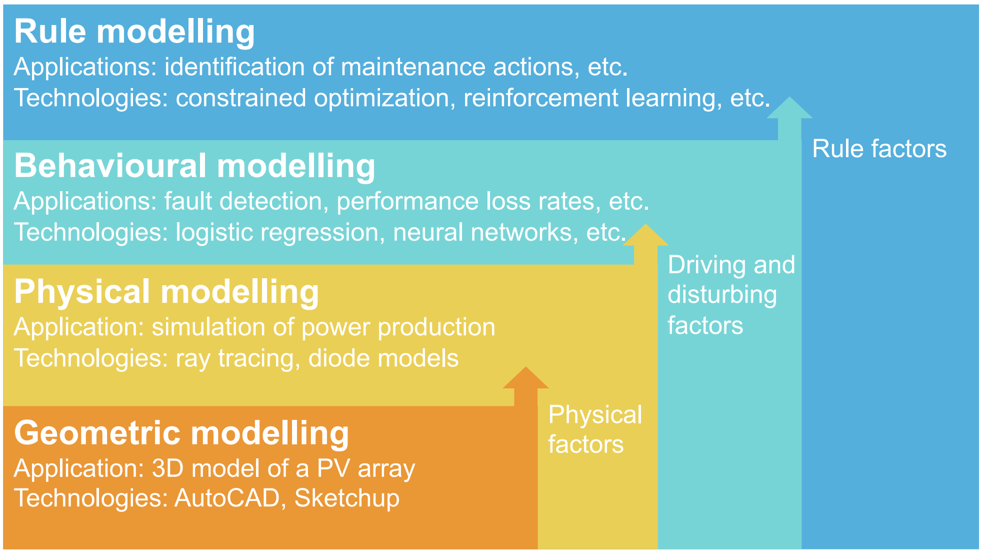


Figure 9: Modelling aspects of a virtual entity as proposed by [11] with examples in the PV sector.

4.2.3 Data flows

Data flows refer to the exchanges of information between the physical and virtual entities and the infrastructure that allows such exchanges. The data generated by these two entities can differ in their variety (e.g., images, timeseries, text), velocity (e.g., real time, offline), volume (e.g., 5-min frequency sensor data, day-ahead forecasts) and veracity (e.g., simplifying assumptions in physical models, sensor noise).

Physical data refers to the data generated by the physical devices and considered by the physical entity. In this case, data collection pipelines are set to gather the physical data from their different sources (e.g. sensors, maintenance logs) and store them according to the object ontologies defined through the physical entity. Data collection can be as simple as manually gathering the essential construction information of the PV system, or as complex as setting up an infrastructure to collect production data from sensors in real-time. In the latter case, data flows can involve a cloud IoT platform using a variety of services such as data bases, software containers, and computing.

On the other hand, virtual data refers to data generated by models that can either use physical or other virtual data. Some examples include the expected production estimated by physics- or data-driven models, LCOE (levelized cost of energy) estimated by economic models, waterfall analyses for performance losses or cleaning schedules calculated by an optimization algorithm.

In Chapter 5, we will discuss some of the key dataflows of the digital twin system and highlight how they enable different applications of the digital twin.

Physical and virtual data interact with the users, services, and devices in a different manner. In the standard for information exchange in the manufacturing sector [106], four types of communication networks are defined to allow the operation and transfer of information. Namely, user, service, access, and proximity networks, as shown in Figure 10. The user network connects the user with the DT; the service network connects the various services within the DT; the access network connects the device communication system, the DT, and the user; and the proximity network connects the device communication system to the physical device. Some examples of these definitions applied to PV plants can be found in Table 3.

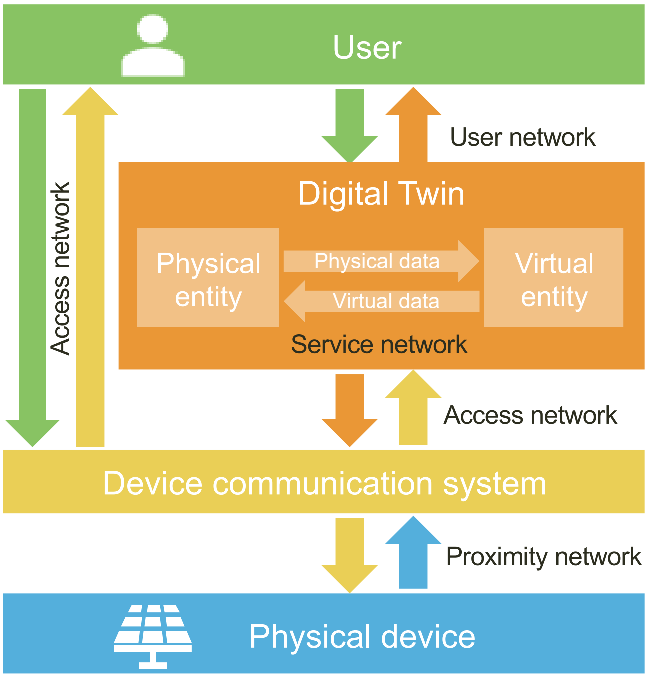


Figure 10: Communication networks that allow the information between the DT, the users (e.g. PV plant operator), and the physical devices (e.g. a PV plant) [106].

Based on the examples provided in Table 3, the reader can imagine a day-to-day scenario where an asset manager logs in to the DT's interface to check the key performance indicators calculated by the DT using data generated by the PV plant and retrieved by the communication system. By interacting with the DT's interface, the asset manager can also inspect the alarms highlighted and a list of actions proposed by the DT. In this scenario, the DT can also execute an action such as rebooting an inverter or raising and O&M ticket after approval by the asset manager. The data to the DT are continuously updated to show the behaviour of the plant and the effect of the actions performed. By executing an action that impacts the PV plant, the DT in this scenario represents an agent that acts on behalf of the asset manager.

Table 3: Examples of information exchange for a DT of a PV plant.

Type	Example
User	User login information is transferred to the DT to access its API. The user provides the information required to configure a DT to an initial state. The user launches a data analysis request. The user approves an operation proposed by the DT; for example, a reboot instruction to the inverter
Service	Transferring selected outputs from one computing block to the next to perform a data analysis. The data cleaning schedule needed for analyses or for model re-training.
Access	The data from the SCADA system is sent to a DT database. The user sends a request through the DT to reboot an inverter. The DT sends this instruction to the SCADA system.
Proximity	The SCADA system sends the reboot instruction to the inverter. Data are transferred from a pyranometer to the SCADA system.

4.2.4 Service systems

Tao et al. [105] introduce the concept of “service systems” to refer to support services for the management and control of the physical entity, and the operation and evolution of the virtual entity. As shown in Figure 11, this includes services related to visualization approaches, software containers, automation pipelines, data processing, user connectivity etc.

Although seldom discussed in the DT literature, the service systems are necessary elements of the DT. They allow the DT to execute its intended computation tasks and interact with the user and the physical device. The development of the service systems requires skills in software development, data architecture, cybersecurity, systems integration, data science and project management. This may contrast with the skills that are more relevant to the development of the physical and virtual entities of the DT, such as domain-specific knowledge, simulation modelling and data analytics.

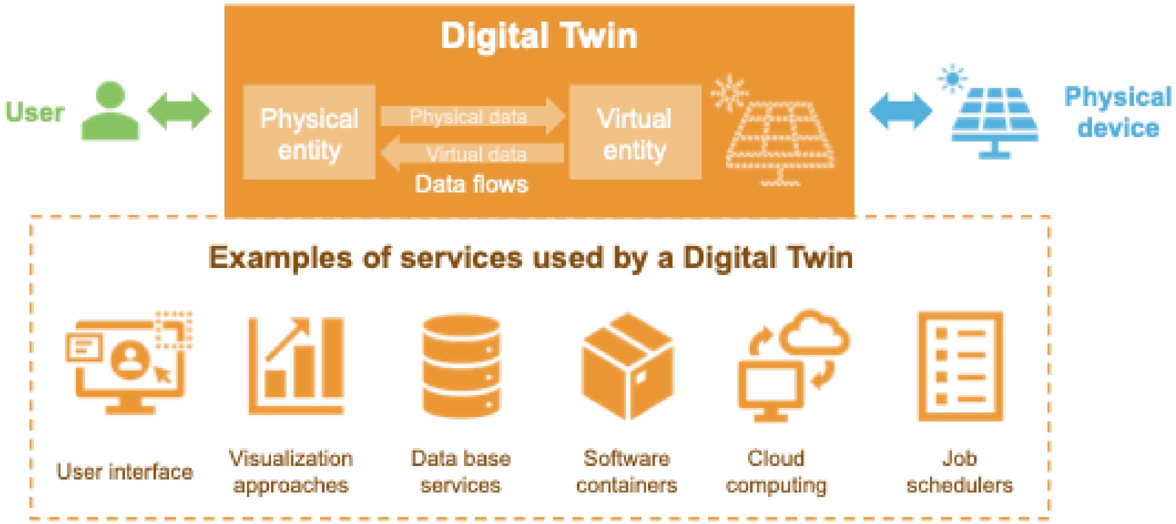


Figure 11: The operation mechanism of the service system that enables the interactions between the physical and the virtual entities through the data flows.

4.3 Data-driven Digital Twins for a Fleet of PV Systems

A data-driven digital twin (ddDT) is an approach focused on modelling the system based on real-world data streams from the system, as opposed to using empirical or physics-based models to reproduce the behaviour of the system. It represents the physical relationships between the system's variables without relying on empirical simulations of *ab initio* models. Data-driven digital twins are extremely helpful when the physics is poorly understood or in situations where one has access to large volumes of real-world data for use in data-driven modelling. ddDT's seek to model the system as it actually is, based on real observations and data streams arising from the device in question. Thus, it's extremely useful to understand the systems behaviour in real-world scenarios and has the advantage of making comparisons among systems easier. For a PV system which is generating large data streams of real-world data, a trained data-driven digital twin can capture a comprehensive and detailed understanding of all of the behaviours of the system, under all conditions it is exposed to, without the blind spots that could arise from physical mechanisms that were not expected and, therefore, not included in a physics-based DT.

4.3.1 Definition of Foundation Model

While many artificial intelligence (AI) models have been used in building DTs and solving real-world problems, creating and deploying each task-specific AI model often requires a considerable amount of time and resources. The new wave of AI application, that's replacing the current task-specific models, is the deployment of foundation models [107]. Foundation models are trained on a broad set of unlabelled data that can be used for different tasks, with minimal fine-tuning. As the name suggests, the foundation model can be used for many applications and can apply information it's learnt about one situation to another.

4.3.2 ddDT's for PV Fleets using Graph Foundation Models

Developing a data-driven digital twin (ddDT) for fleets of distributed power systems is a critical task in the process of energy asset digitization. Notably, data-driven efforts often focus on neural network models trained on singular sites. While traditional neural network model architectures can provide deep site insights, they risk model overfitting and present challenges in model generalization, especially in the face of subpar data quality [108], [109], [110]. Employing a foundation model approach in building data-driven digital twins has significant benefits. Graph based approaches, such as spatiotemporal Graph Neural Networks (st-GNNs) have significant advantages in building a foundation model [111]. Using st-GNNs for data imputation of energy production data, and for power forecasting has great advantages because it leverages the inherent spatial and temporal nature of the distribution of the PV systems and the weather phenomena, they are all experiencing. This methodology inherently considers the spatiotemporal dynamics, and spatial and temporal coherence, influenced by local weather and other environmental factors, with the aim of harnessing fleet-level data to support model predictions. Through this approach, ddDTs seeks to transcend the limitations of single-site models, paving the way for more generalized and accurate predictions across the spectrum of PV system performances [112], [113], [114], [115]. Figure 12 shows an example of st-GNN implementation for a fleet of 3 PV system in which each node represents one system with its strings of PV modules and has a Feature Vector of metadata information about that particular system. The feature vectors are the system's metadata, while each node also contains the time series data streams for electricity production and weather conditions are contained in each st-graph. Figure 13 shows an example of a fleet of 316 PV systems, distributed from Puerto Rico, the continental US and Hawaii. The connectivity of the graph is determined by epsilon, a hyperparameter trained to maximize the predicted power accuracy of the model. For epsilon = 1, each system is isolated, while epsilon = 0 is a fully connected graph, and shown here is epsilon = 0.5. The minimum predicted power forecast is found for epsilon = 0.25. Figure 14 shows how a foundation model is used in building a data-driven digital twin for PV system power prediction. A use case of the data driven digital twin using real world data is presented in section 4.4.2.

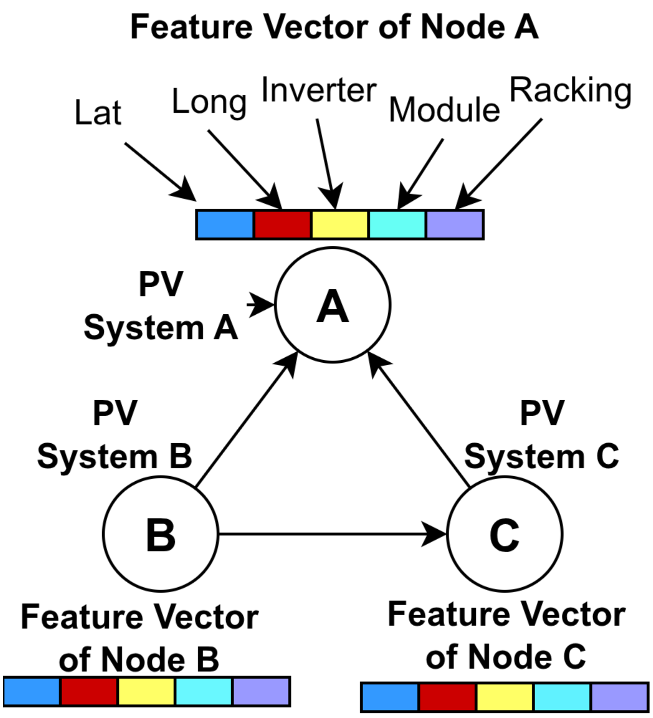


Figure 12 A data-driven Digital Twin model for a PV Fleet of 3 systems consists of a spatiotemporal graph (st-Graph) of the PV systems and inverters, in which each node

represents one system with its strings of PV modules and has a Feature Vector of metadata information about that system. The feature vectors are the system metadata, while each node also contains the time series data streams for electricity production and weather conditions are contained in each st-graph.

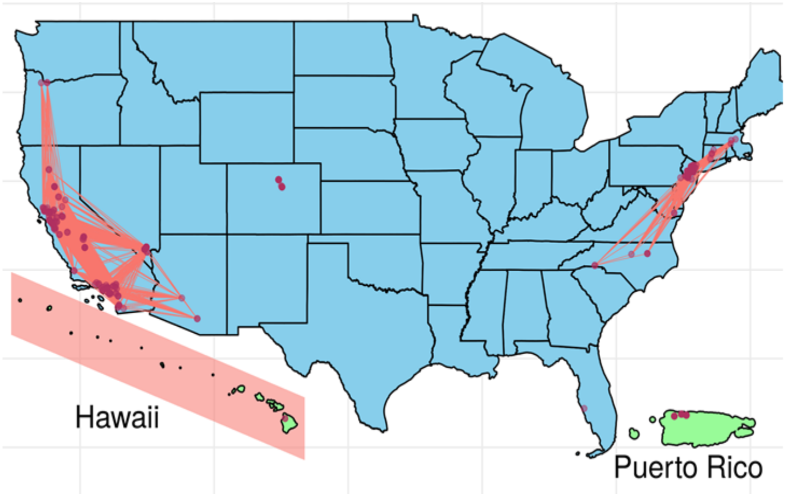


Figure 13: For a ddDT of a fleet of 316 PV systems, distributed from Puerto Rico, the continental US and Hawaii, a spatiotemporal graph foundation model is assembled, and trained [112].

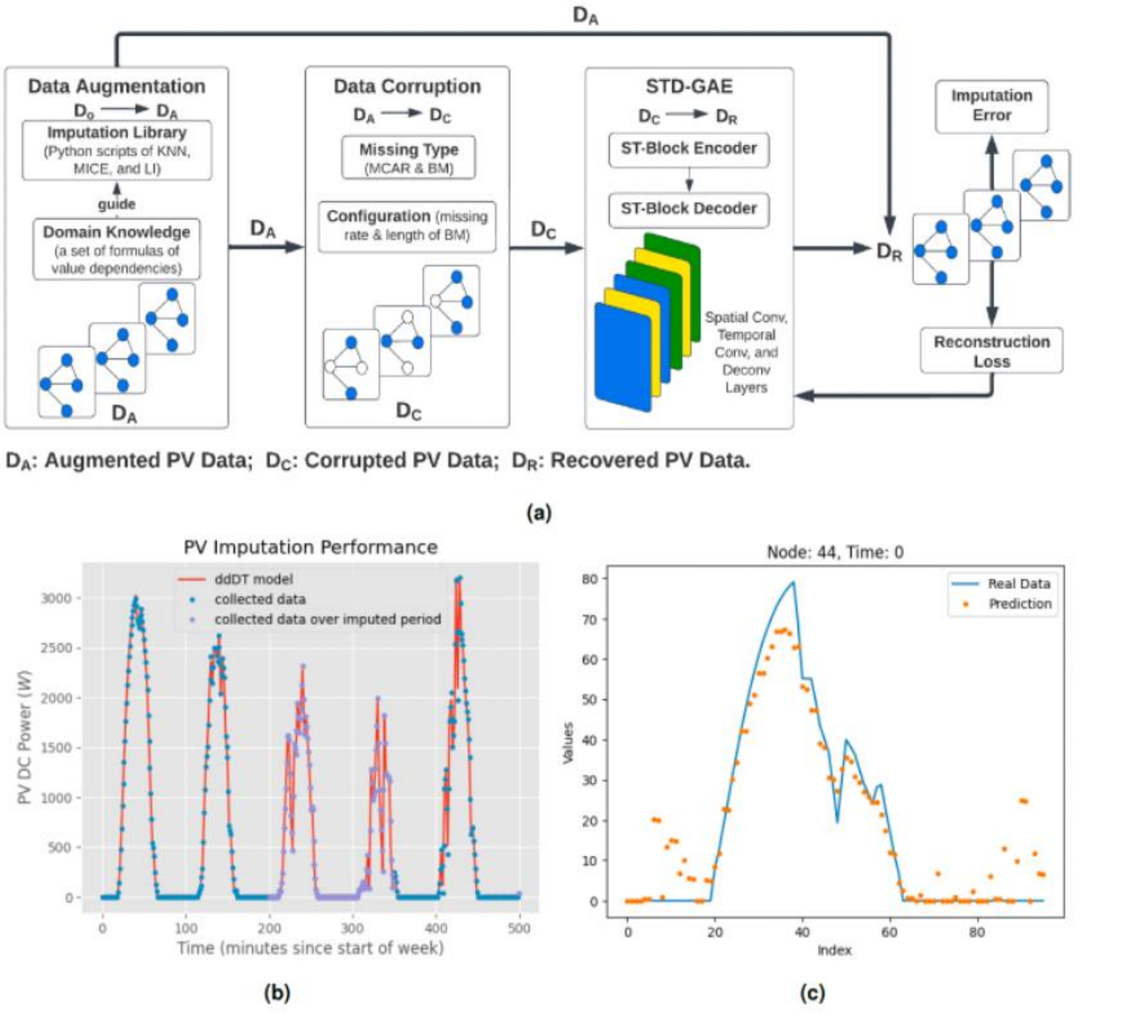


Figure 14: (a) Model structure for data imputation of missing gaps of PV power data. The entire model pipeline (Foundation Model + task specific parts) is referred to as a "STD-GAE", following convention from the literature. (b) An imputed PV power signal. (c) The performance of the ddDT power model on a day of PV power. Note that the values shown on the y-axis are DC power with a normalization factor applied to it. The model closely follows the trend during the daytime. As the day-night cycle is known, the model is not penalized for nonzero values at night. This improves model convergence behaviour and compresses the dataset into a useful representation, as learning that 50% of the data is zeros is unnecessary for a useful model. Taken from Pierce et al. [116].

4.4 Use Cases and Best Practices

4.4.1 SunSmart Case Study

The SunSmart E-Shelter Schools program in Florida, USA is the first in the nation to outfit emergency shelter schools with solar + storage. More than 115, 10 kW PV solar systems are currently installed in schools designated as emergency shelters throughout Florida. FSEC, an energy research centre at the University of Central Florida, coordinated the installations, which began in 2010.

These solar + storage systems have become the centrepiece of community events, with students and teachers acting as ambassadors to educate local citizens about the clean, silent electricity produced for their school. More importantly, in the event of an emergency, these systems use energy stored in batteries to provide power to key aspects of the shelter and provide an emergency shelter for the local population.

Figure 15: More than 115, 10-kW PV solar systems are installed on the emergency shelter schools throughout Florida. More than 50,000 students were introduced to photovoltaics and renewable energy technologies through the SunSmart Schools project. The installation of 10 kW ground-mounted systems with battery backup generated jobs across multiple industries.

Timeseries data spanning three years of PV systems from 29 schools in Florida that participate in the SunSmart E-Shelter Schools program have been gathered [117], [118]. This dataset includes timeseries information on PV system performance, geographic data on their locations, and relevant weather conditions [119], [120], [121], [122].

The initial phase involves building a spatiotemporal graph neural network (st-GNN) to predict power in the system [112]. Timeseries data were processed, and missing values imputed using the pre-trained Spatiotemporal Denoising Graph Autoencoder (STD-GAE) model of Fan et al. [113], which forms the basis of the Foundation Model presented in this work. A distance-based graph is constructed for all sites using a Gaussian kernel to threshold edges. Intuitively, PV systems that are spatially closer together should share more information.

The processed timeseries data and adjacency matrix representation were trained for an imputation task, as shown in Figure 14a. The st-GNN model aims to produce an output with the same dimensions as the input, not just values for specific points. This process, known as "reconstruction," involves the model recreating the input from latent space manifolds or embeddings [123], [124]. However, it is not always informative to calculate the model reconstruction loss for values that had to be filled in with a preliminary imputation method, especially for downstream tasks where fidelity to the original signal is required. Calculating error for these values could bias the model to the methods that were used to fill in the data, especially for datasets with large amounts of missingness, and lose the character of the real dataset. To remedy this, the model can selectively calculate the loss for only values in which real data was measured based on a matrix of flags. This approach allows us to modify our learning targets on the fly and respect domain knowledge when assessing the performance of these deep learning models.

For a block missingness scenario, a raw and imputed PV power signal is shown in Figure 14b. The model performs quite well, especially given the chaotic variability of PV systems' output caused by weather variability. The model can capture both the daily, monthly, and yearly seasonality of power along with matching the high frequency fluctuations due to solar resource variability and local microclimatic effects induced by shading / system failures. Note the generalizability of this approach: although this work chooses to impute PV power, it would be just as simple for system current or voltage as well. As a specific task, it is common in PV to create a "power model", which takes as input measured weather data and then outputs the expected AC or DC power of the PV system. The flexible nature of a st-GNN makes it simple to retrain for these two tasks. During the training step, irradiance, air temperature, and wind speed are used as inputs and targets in the semi-supervised paradigm. The model is then adapted to take the same weather data as input and output the PV power. The benefit of this approach is that the initial training step creates a useful encoding of the weather data, which the power model can then take advantage of. An example of model performance on this task is shown in Figure 14c.

4.4.2 Use Cases

Digital twins are transforming the PV industry by improving real-time monitoring, simulation, and analysis of systems. Acting as a virtual representation of a PV system, they support optimization of performance, predictive maintenance, risk management, and decision-making throughout the PV system lifecycle.

Here we**Error! Reference source not found.** present six key use cases for digital twins, highlighting applications and best practices. These range from performance optimization and risk analysis to electricity market simulation, demonstrating the transformative potential of DTs in enhancing energy production, reducing costs, and improving efficiency in the renewable energy sector.

Table 4: Use cases for DTs in the PV industry

Use case	Description	Best practice
Performance Optimization:	Digital twins can simulate the performance of a PV system under various conditions, helping stakeholders make informed decisions about system design and operation. For instance, a digital twin can help determine the optimal tilt angle for PV modules in a specific location or the best configuration of panels to maximize energy production.	Use high-resolution weather and performance monitoring data integrated into the digital twin model to increase parameter accuracy for simulation scenarios. Implement machine learning algorithms on historical and real-time datasets to recommend deployment scenarios and system-level adjustments. Use sensitivity analysis to test different variables and optimize performance (e.g., tilt angles, row spacing) while meeting financial targets.

Table 4 (continued)

Use case	Description	Best practice
Decision Support System[125], [126]:	Efficient maintenance process automation and standardization are essential for optimizing yield and lifespan. Identifying common issues with a standardized taxonomy can lead to effective solutions. The Cost Priority Number method helps evaluate the economic impact of technical failures in energy systems. Such a methodology offers the opportunity to shift current O&M routines from the predominant reactive approach towards an automated decision support system	Apply failure mode and effects analysis (FMEA) integrated with the Cost Priority Number method to identify critical failure points in PV systems. Standardize O&M (operation and maintenance) decision-making through digital twins by configuring risk thresholds and escalation protocols. Automate alerts and maintenance routines triggered by performance deviations tracked within the digital twin, enabling predictive and proactive O&M workflows.
Risk Analysis[30]:	Risk analysis enables users with statistical and reliability data to develop and run scenarios in which PV performance and costs are affected by components that can fail.	Use Monte Carlo simulations or probabilistic risk models in digital twins to provide insight into failures or yield reductions. Leverage real-time IoT sensor inputs for reliability analysis and degradation modelling of components (e.g., inverters or solar modules). Establish a dynamic insurance risk assessment framework using data derived from digital twin platforms to simulate plant reliability over a 20+ year lifecycle.
Plant Acquisition:	When considering the acquisition of a PV plant, a digital twin can provide valuable insights into the plant's performance. It can simulate the plant's energy output under different weather conditions and over time, helping potential buyers assess the plant's value and return on investment.	Develop a detailed, dynamically updatable model using digital twins to provide a virtual replica of the plant's historical and predicted performance. Include financial KPIs such as ROI (Return on Investment), IRR (Internal Rate of Return), and payback periods derived from twin simulations. Customize acquisition criteria scoring parameters such as component lifetime estimations, irradiation trends, or O&M efforts to quantify the real long-term potential of the system.

Table 4 (continued)

Use case	Description	Best practice
Maintaining Plants:	Digital twins can predict when components of a PV system may require maintenance or are likely to fail. This allows for proactive maintenance, reducing downtime and associated costs. For example, a digital twin might indicate that a particular inverter consistently operates at high temperatures, suggesting that it may require maintenance or replacement.	Implement condition-based monitoring (CBM) within the digital twin framework to track component health. Integrate failure prediction models using techniques like root cause analysis or regression modelling to detect anomalies in thermal profiles, vibration data, and electrical pattern. Schedule maintenance ahead of potential downtime by automating service notifications for spare parts procurement, workforce readiness, and logistics management through digital twin data insights.
Electricity Market:	Digital twins can also simulate a PV system's response to electricity market conditions. For instance, they can model how the system would perform under different electricity prices, helping operators decide when to sell electricity to the grid or when to store it. This can optimize the financial performance of the PV system.	Combine digital twin energy output models with machine learning-based market data analysis tools to simulate forecasted price trends. Optimize energy storage and dispatch decisions by incorporating grid demand predictions, time-of-use rates, and price arbitrage calculations into twin operations. Create actionable electricity trading simulations that enable operators to switch between storage utilization or grid sales based on forecasted financial results.

4.4.3 Best Practices

Digital twinning provides a holistic view of the operation of PV assets, aiming to support decision-making for planning O&M activities or for the selling of the whole asset. One of the significant advantages of digital twinning is the automatization of procedures, such as visualization, reporting, and the calculation of simple metrics, which in turn saves considerable time and reduces manual labour. Additionally, when implemented correctly, a digital twin can serve as a single source of truth, ensuring consistency and accuracy across all data and operations related to photovoltaic assets.

The development of digital twins can be approached incrementally, allowing organizations to stop increasing complexity when it no longer adds value. This approach helps manage costs and ensures the system remains efficient. Digital twins also enable remote monitoring and management of photovoltaic assets, significantly reducing the need for on-site personnel and saving on travel and operational costs. Furthermore, by providing a digital representation of physical assets, digital twins make it easier for a broader range of stakeholders to access and understand important information, thereby enhancing decision-making processes. The use of soft sensors in digital twins can also reduce the need for investing in physical sensors, as these soft sensors use algorithms to estimate measurements, offering a cost-effective alternative.

However, several challenges accompany the implementation of digital twinning. Managing data quality, standardization, control, and security are significant hurdles. Poor data can lead to incorrect decisions, necessitating robust data governance practices. The deployment and maintenance of the necessary IT infrastructure can also be costly and complex, involving regular updates, ensuring connectivity, maintaining privacy, and managing power and storage requirements. Additionally, a lack of standardized modelling approaches and the use of non-explainable models can complicate the implementation of digital twins, with the economic benefits of these models not always being easily verifiable.

The initial and ongoing costs of developing digital twins can be substantial. For instance, at TotalEnergies, developing a minimum viable product (MVP) for a digital twin can take about 12 months and involve around 10 people, including product owners, project managers, developers, IT specialists, data scientists, administrators, and people managers. Further costs are incurred in development, maintenance, and user support post-MVP. The costs associated with sensors, software, and their maintenance can also be high, adding to the overall expense.

Cultural challenges also pose a significant barrier. There is often a shortage of specialists in software development and data science, leading to higher costs for specialized personnel. Additionally, the slow acceptance of new AI technologies can hinder the adoption and effectiveness of digital twins. While digital twins aim to support decision-making, such as buying plants, maintaining plants, and market-driven approaches, the complexity of these decisions can sometimes outweigh the benefits provided by the digital twin.

Finally, implementing digital twins within the industrial sector, particularly for PV power plants, can be challenging. This involves navigating various technological, logistical, and regulatory hurdles. By carefully considering these points, organizations can better assess the potential benefits and challenges of implementing digital twinning in the photovoltaic domain.

Several companies are embracing advanced digital twin technology, improving their abilities through research and development. The widespread use of these innovative technologies has resulted in practical advantages for customers. DTs in the industrial sector offer benefits in surveillance, analysis, prediction, enhancement, and decision-making. They can educate employees through digital models of equipment, settings, and individuals. DTs are commonly utilized in diverse sectors such as construction, healthcare, agriculture, shipping, manufacturing, energy, automotive, and aerospace. [127]

DTs also represent a significant innovation in the PV industry, offering a range of benefits from performance optimization and predictive maintenance to increased cost savings. They provide an effective tool to improve different aspects of PV system management and operation, such as decision-making, plant acquisition, maintenance, or market participation.

5 DIGITAL TWINS IN PV O&M: DATA FLOWS AND APPLICATIONS

5.1 Introduction

In Chapter 4, the concept of digital twins was defined, and related data flows and applications were introduced. For a fully functional DT system, we can identify a large variety of data flows that provide required inputs for the DT system, or present useful output of it. As mentioned before, data flows refer to the exchanges of data between the physical and virtual entities and the infrastructure that allows such exchanges. The combination of DT and data flows is what enables the functionality of the DTs.

In the upcoming sections, we delve deeper into the specific data flows and applications of DTs in PV systems. Section 5.2 focuses on the monitoring data, which are collected in real time from the physical devices that make up the digital twin's PV system or fleet of PV systems. These data are acquired using supervisory control and data acquisition (SCADA) systems and typically consist of timeseries of numerical values, vital for overseeing the PV plant operation. Monitoring data are the basis of the functionality enabled by digital twins in the O&M interface, providing the necessary inputs from the physical to the digital entity.

In Section 5.3, we examine several applications of digital twins enable by these data, and the data flows they provide as outputs, for modelling plant performance, detecting faults and providing predictive maintenance. There is a stream of data originating from the virtual entity, and as mentioned in Section 4.2.3, these are typically data generated by models that aim to emulate the operation of the physical entity at high accuracy, such as data-driven or physics-driven models for the PV plant or PV fleet electrical output parameters. Within the digital twin system, modelled power output can subsequently be used to determine a set of key performance indicators (KPI’s) of electrical or financial performance, and a breakdown of modelled data can be used to analyse power losses in detail and to perform predictive and corrective maintenance. In Section 5.3.2, we detail these data flows.

Through these sections, we aim to provide a comprehensive understanding of how digital twins facilitates and builds upon various data flows to enhance the performance and reliability of PV systems.

5.2 Monitoring PV systems: data from the field

In this subchapter, we describe data streams from the physical entity to the virtual entity of the digital twins. The monitoring data streams are one of the key inputs of the digital twins’ functionality, as they provide (near) real-time data on the performance of the physical PV plant. Here, we describe two distinct flows of data:

- Continuous monitoring data measuring the system’s electrical performance at module, string and/or inverter level, grid feed-in data from transformers and the metered connection point, and measurements of environmental conditions from weather-stations, or arc event sensors.
- Field inspection data collected on regular scheduled intervals or when triggered by the detection of anomalous behaviour of the PV plant from the monitoring data. These data include e.g. visual, IR thermography, or EL (or PL) module inspections from the ground or using UAVs, I-V curve measurements from modules or strings of modules, and structural mechanical tests of mounting systems and trackers.

5.2.1 Continuous Monitoring Data

In a PV system, several physical entities generate real-time measurement data, such as inverters, combiner boxes, weather stations, and electric meters. This data including error messages from devices is typically collected by a central device, which then transmits it to a cloud-based monitoring database. An example of this process is shown in Figure 16.

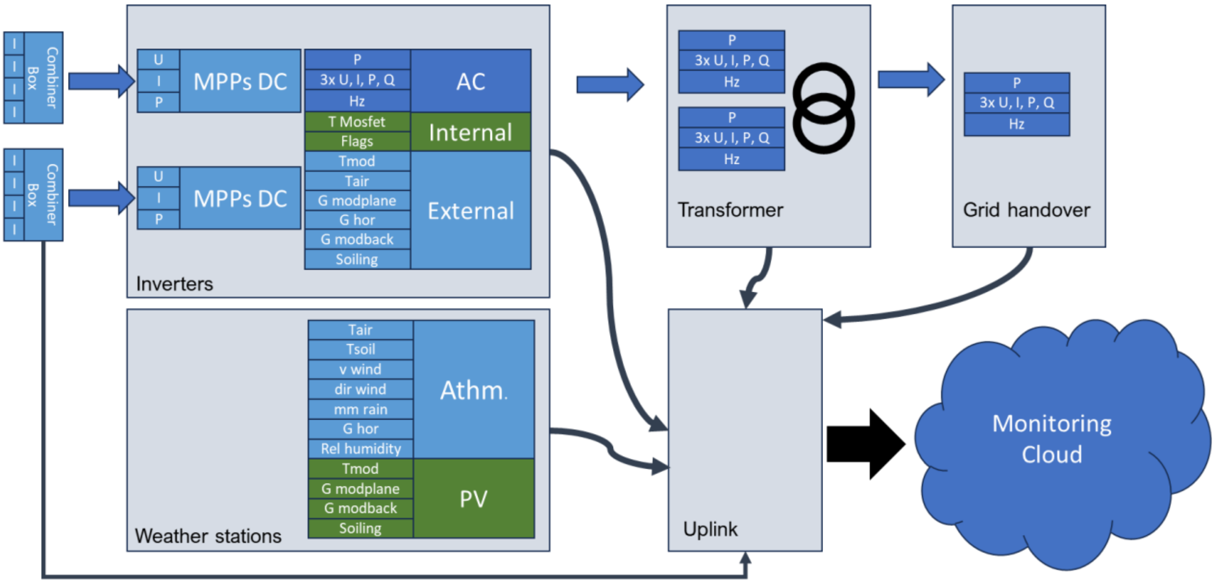


Figure 16: The data (black) and energy flow (blue) in a larger PV system.

As its firmware control procedures depend on them, a typical inverter needs to monitor many different values during its operation:

- DC for each MPP: voltage, current of each parallel string, DC power
- AC total: three-phase currents, active and reactive power, phase angle, produced energy, conversion efficiency
- Grid: three-phase voltages, frequency.
- Internal: self-consumption, MOSFET temperatures, internal time, HVAC of container
- Operational flags: waiting for sun, clipping, reactive power provisioning, overtemperature, faults

Commonly, the DC data has better accuracy than the AC data, where much higher sample rates are required to perform integrations. Often, also the option exists to include additional sensors for air and module temperature, or even irradiation, e.g. using a Modbus interface. For large scale systems, this data is however often collected by dedicated local weather stations that include wind, rain and soiling data. A standard exists for the number and type of sensors within the IEC 61724-1 “Photovoltaic system performance – Part 1: Monitoring”, which for economic reasons is applied seldomly. A common problem of weather stations is that they do not interval average irradiation, resulting in substantial uncertainty of received solar insolation especially for long recording intervals.

Especially smaller inverters might not deliver monitoring data at nighttime, opposed to central inverters, whose grid support functions might be active (i.e., for reactive power provisioning). Not every inverter might be able to perform nighttime reactive power provisioning, as the energy needs to be taken from AC instead of DC thereby.

Apart from the data itself also the timestamp of the data is transmitted, that often originates from the endpoint's individual potentially unsynchronised and hence drifting timekeepers. Some data might be interval averaged (e.g. energy), while other data can be single shot measured (e.g. voltage). This can produce systematic half interval time discrepancies (Figure 17) leading to permanent over and underestimations in morning and evening hours, as well as during rapid changes in irradiation due to cloud movement. If asynchronous data is collected and only transmitted to the centralised database in an aligned fashioned, similar problems arise. The shorter the monitoring interval, the smaller the effect of these artifacts. A maximum of five minutes interval, ideally less, is hence considered good practice.

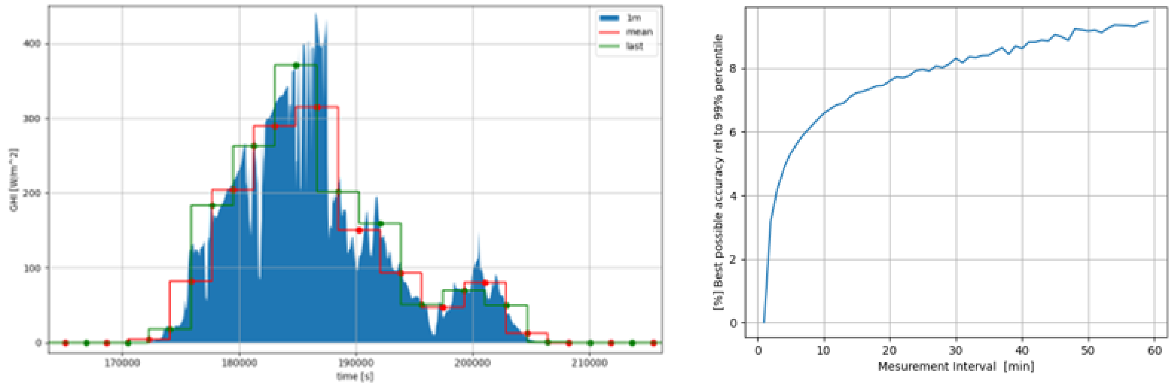


Figure 17: Left: One minute resolved irradiation data in blue. Interval averaging (red) and single shot (green) strategies shows a half interval time shift in the compared data flows. Right: The systematic PV prognosis error depended on the monitoring interval.

In the central monitoring database, a text-based naming scheme exists for all the values that are gathered. Although ideally automatically generated, it can contain spelling irregularities due to human errors. Seldomly, the name of a time series does not fit the content, and data can be erroneous, e.g. exhibiting out of band data or long-term stuck values.

For analysis it is recommended to auto-obtain the monitoring naming scheme using textual analysis of the time series names, and to validate the content by means of amount of zero, strings, not-a-number (NaN) values, quantile values (e.g. 0.5, 0.98), and daily and yearly Fourier analysis. Monotonically increasing values will exhibit integrated quantities, see Figure 18. The Fourier phase amplitude is useful to evaluate both over the day, as well as the year, to find quantities that have day curves (e.g. irradiation, temperature) or seasonal changes. While extreme values should be checked for plausibility, they often indicate sensor faults, data corruption, or unrealistic physical conditions, and may require filtering or flagging before further analysis.

Apart from direct sensors, external data is often integrated, such as surrounding weather station data, satellite image calculated irradiance. To include historic weather forecast might be useful, if a direct trading market participation on a hourly or sub hourly base or e.g. using Virtual Power Plants (VPP) is considered in the future. Derived data, such as performance ratio calculations or inverter efficiencies also are common, but seldomly the algorithms are documented in an appropriate way to trust algorithms in a commercial plant review, and the DC and AC uncertainties are sufficient.

PV system data is often interchanged by 2d text files. Thereby, one can expect a data usage of 0.5-2 Mbyte per Inverter and kWp for 5-minute logging when using 20 kWp Inverters. For 1MW and 30 years, hence 15-60 GBytes are estimated. Using central inverters, the data amount per kWp is largely decreased, although the number of columns per inverter typically rises, e.g. by the logging of the individual DC currents of parallelised PV strings in the combiner boxes. An exemplary large scale 300 MWp using central inverters e.g. creates 6 GBytes per year.

Name1	nrNone	nrNan	nrLimited	nrNums	nrDateTimes	strings	valMin	valMax	Qu1	Qu50	Qu90	Qu99	Year-Amplitude	Year-Phase[days]	Day-Amplitude	Day-Phase[h]	Intraday-Increase	Intraday-Decrease
S 2 (#2, 71)Einstrahlung (W/m2)	0	0	0	455220	0	0	0	1,297.0	0.1	0.1	507.9	924.0	3.5	0	91	12	53%	47%
WR 1 (#1, 1)Udc2 (V)	0	0	0	455220	0	0	0	799.0	0.0	433.0	664.9	703.0	3.3	0	182	12	48%	52%
WR 1 (#1, 1)Pac (W)	0	0	0	455220	0	0	0	18,508.0	- 1,123.8	- 5.0	11,913.4	17,765.5	95.4	0	2,262	12	53%	47%
S 2 (#2, 71)Insolation (Wh/m2)	0	0	0	455220	0	0	0	8,077.0	- 98.0	53.7	4,806.8	7,153.2	39.8	0	75	0	100%	0%
WR 1 (#1, 1)Pdc2 (W)	0	0	0	455220	0	0	0	10,435.0	- 1,272.3	- 7.3	6,104.8	9,076.5	49.6	0	1,162	12	53%	47%
WR 1 (#1, 1)Pdc1 (W)	0	0	0	455220	0	0	0	12,026.0	- 1,653.9	- 8.2	6,248.0	9,398.4	49.5	0	1,189	12	53%	47%
WR 2 (#2, 1)Pdc2 (W)	0	129377	0	325843	0	0	-9	158,620.0	- 14.0	10.5	94,251.1	145,824.2	782.7	183	4,685	0	73%	27%
WR 1 (#1, 1)Etotal_C (WattEver)	0	0	0	455220	0	0	0	67,682,592.0	- 922,739.2	- 8,315.8	35,625,347.6	58,400,698.0	411,603.0	0	329,635	12	100%	0%
WR 1 (#1, 1)Udc1 (V)	0	0	0	455220	0	0	0	817.0	0.0	451.0	684.0	730.0	3.4	0	186	12	49%	51%
S 2 (#2, 71)Modultemperatur (Ã,Â	0	0	0	455220	0	0	-7	255.0	- 2.0	18.0	36.0	55.0	1.1	183	3	12	50%	50%
WR 1 (#1, 1)Uac (V)	0	0	0	455220	0	0	0	243.0	0.0	0.0	237.0	240.0	0.6	0	67	12	50%	50%
WR 1 (#1, 1)Ertrag (Wh)	0	0	0	455220	0	0	0	3,949,464.0	- 16,589.7	60,576.1	2,153,582.6	3,488,307.2	28,614.3	0	34,559	0	100%	0%
dt	0	0	0	0	455220	0											0%	0%
generation	0	0	0	455220	0	0	0	1.0	0.0	0.0	1.0	1.0	0.0	183	0	0	0%	0%
S 2 (#2, 71)Etotal_C (WattEver)	0	325843	0	129377	0	0	0	29,078,055.0	- 166,973.2	2,513,752.4	20,016,169.9	26,685,739.3	230,447.0	0	431,635	0	100%	0%
WR 1 (#1, 1)Ber_PDC_1+2	0	0	0	455220	0	0	0	20,304.0	- 2,175.4	- 8.5	12,354.5	18,508.6	99.0	0	2,352	12	53%	47%
WR 1 (#1, 1)Ber_IDC1	0	203097	0	252123	0	0	0	17.0	- 0.2	2.7	12.3	15.3	0.1	0	1	12	53%	47%
WR 1 (#1, 1)Ber_IDC2	0	203454	0	251766	0	0	0	17.1	- 0.2	2.7	12.2	15.4	0.1	0	1	12	53%	47%

Figure 18: Initial plausibility evaluation of the monitoring data content of a tiny PV system, with cryptic and localized column names. The data type (blue) and the percentiles (yellow) are analysed. The year and day Fourier analysis shows how quantities change by season or daytime, while the data that mostly increase which each timestep is often integrated quantities, e.g. produced energy.

It is not recommended to remove old data from the monitoring system. Especially the initial two years of operation are essential to obtain a reasonable degradation rate, and to quantify effects such as LeTID or PID. Hence, it is also not recommended to first install the modules, and weeks later start logging monitoring data.

It is also necessary that the monitoring platform allows large scale data import and export, as an internet exposed online system cannot be assumed to have 30-year lifetime, e.g. due to changing IT security requirements. This argument becomes even more relevant, if an interactive remote control of the system is possible.

In Operation and Maintenance (O&M) of PV systems, quality management requires to store additional data of PV systems, such as building plans, inverter manuals, or owners financial reports. This is often performed in a file folder structure. Advanced measurement campaigns, such as drone imaging (IR, EL, VIS) are often archived within: As a single drone visit can generate hundreds of GB of images and videos, databases file systems are seldomly applied (see Chap. 5.2.2).

Either using a dedicated ticketing system, or some other database, so called event logs typically exist. They are used to document failures, plant visits, automatic alerts, security fence alarms, maintenance protocols, and similar. While most O&M companies use standard methods for that, international archiving standards that are machine interpretable would be beneficial.

5.2.2 Field Inspection Data

Aside from the data from the continuous monitoring described in the section above, there is a substantial stream of data generated by preventive and corrective field inspection testing. This involves the direct examination and testing of physical components in the field, at a much lower time resolution compared to the real-time monitoring data described earlier. These inspections aim to detect issues that monitoring systems might miss, such as physical damage, degradation, earth faults, insulation resistance of individual DC strings or potential failures that have not yet resulted in a measurable drop of electrical performance. The primary purpose is to assess the actual condition of components, validate sensor readings, identify hidden defects, and provide ground-truth data to calibrate digital twin models. Table 5 gives an overview of five key field-testing activities generating data from the physical to the virtual DT entity. For typical PV plants, routine field inspection techniques include visual inspection, thermal (IR) imaging, EL testing, I-V curve tracing and inspection of electrical connections [128], [129].

Table 5: Overview of key field inspection tests that are generating data to be analysed in the context of the PV digital twin.

Inspection Type	Typical Frequency	Purpose	Method	Analysis Approach
Visual Inspection (modules + structure)	Annual	Detects physical damage, delamination, discoloration, broken glass, frame corrosion, and structural issues	RGB camera, handheld or UAV	Defect classification, severity assessment
IR Thermography	Semi-annual or annual	Identifies hotspots indicating cell defects, connection issues, or bypass diode failures before they lead to significant power loss or safety hazards	IR cameras handheld or UAV	Hotspot detection, pattern analysis
EL Testing	Every 2-5 years or upon fault detection	Reveals micro-cracks, inactive cells, PID effects, and other cell-level defects invisible to other inspection methods	InGaAs EL camera, tripod or UAV	Cell defect analysis, crack mapping
I-V Curve Tracing	Annual or upon fault detection	Provides detailed performance profile of strings or individual modules	Portable I-V tracers	Parameter extraction, curve comparison
Electrical Connections Inspection	Annual	Verifies integrity of all electrical connections from modules to inverters, earth faults, DC string isolation resistance	Visual check, IR imagery, and selective mechanical testing	Contact resistance evaluation, insulation check

Visual inspection includes inspection by sight by trained technicians, as well as the collection of RGB imagery data using handheld, or UAV mounted cameras. The main aims are to identify visible physical damage in modules (such as glass breakage, delamination) or structural issues with the PV mounting systems. Visual inspection should be conducted at least annually and is often the first step of corrective maintenance when faults are identified from the monitoring data.

IR thermography (IR-T) imaging is a technique that allows for the detection of many faults by identifying abnormal thermal signatures (hotspots) within modules or strings of modules. Hotspots can be caused by a variety of issues, such as cell defects and connection issues, but IR-T cannot usually identify the root cause [128]. Still, as IR-T is a contactless technique that can detect many types of faults and can be conducted using UAVs and offers high inspection throughput at low cost, it is a common field-testing method throughout the lifecycle of the PV plant.

The root cause of faults identified with the techniques described above, cannot always be determined with these techniques, often requiring different or follow-up inspection. Luminescence based techniques such as EL or PL also provide visual inspection of PV modules but can offer more insight into failures' root causes as these techniques can for instance identify micro-cracks, inactive cells, and PID effects. Inspection of the I-V curves of modules or strings using portable I-V tracers can also offer insights into specific failure modes [128], [130].

Field inspections thus help assess the physical condition of a plant, identify early signs of degradation, provide evidence for warranty claims, and collect data that remote monitoring cannot capture. This information is important for maintaining asset health and ensuring long-term plant performance. Digitalisation of field inspection is happening in several ways. First, digital platforms for PV O&M and data analytics help with the processing and (automated) analysis of field inspection data. This includes but is not limited to automated geolocalisation of inspection data, linking inspection data to specific PV system components in the digital twin, image segmentation to detect and identify individual modules, and automated analysis of imagery to detect and classify faults from the image signatures. Secondly, digitalisation is enabling novel inspection methods such as daylight PL based methods, which aim to offer similar fault detection and classification capabilities as EL based inspection, but in a contactless and high-throughput manner [131] [18].

None of these image analysis methods like IR, PL, EL - can yet provide reliable information on the electrical power performance of PV generators. By fusing this image data with precise electrical measurement data, AI or analytical tools could provide a solution in the future.

5.3 Performance Evaluation, Intervention, and Control

As discussed in Chapter 4, a key application of digital twins is during the operational phase of PV systems, to aid in decision support and leverage advanced digital tools for operation and maintenance of PV systems. As illustrated in Figure 19, data flows in this context drive PV system modelling, based on data included in the digital twin, like PV module data and specifications, geometric and geographical data on the physical layout of the system over the entire power plant site including shading objects, other system specifications (e.g., inverters, mounting, cabling) and monitoring data from the physical twin, but also external data like weather data.

Performance assessment, intervention, and control are crucial aspects of any PV system operation and represents the key use cases for the digital twin concept. Digital twins play a vital role in this process by facilitating the continuous flow of data for analysis, decision-making, and optimization. This section explores the key data flows related to Digital Twin modelling, performance assessment, predictive maintenance and intervention.

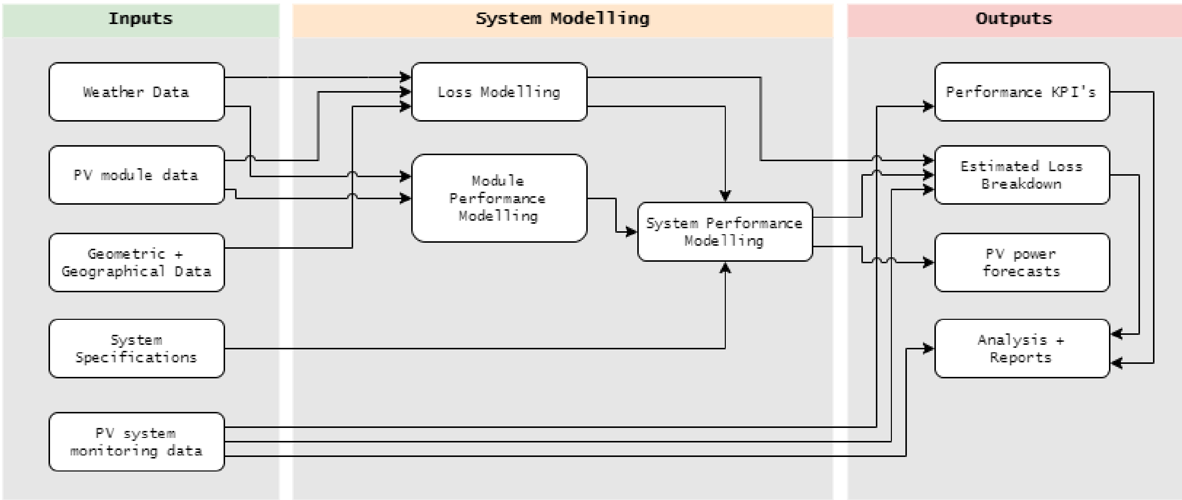


Figure 19: Overview of data and data flows in digital twins.

5.3.1 Modelling PV Power Plant Performance Using Digital Twins

A key functionality and application of digital twins in O&M is the representation of the physical PV plant by a virtual entity, to model the performance of the physical plant. The digital twins allow for a detailed simulation of the performance of the plant and its components, enabling the O&M operator to have 1) an overall estimate of the expected power output and the plant real-time operating conditions, 2) a breakdown of plant losses as a function of different operating conditions and performance KPIs and on the component level. These outputs enable automated condition monitoring and help the O&M operator detect and localize faults, facilitating informed decision-making about potentially required intervention in the plant [132].

In the PV industry, plant performance is typically modelled using either physics-based models starting from the component level, data-driven models leveraging machine learning, or a combination of both. A common method involves generating a parameterized 2D or 3D PV system model in its physical surroundings, (i.e., the digital twin) [133]. The physical properties parametrized in the digital twin are for instance used to model the effective irradiance incident on each module's surface, including the effects of shading from nearby rows of PV modules, nearby buildings or from the surrounding terrain, but also angle-of-incidence effects and rear-side irradiance for bifacial plants on the base of the local and seasonal albedo. Physics-based or empirical models are used to model PV module and cell temperature, and this operating temperature and the effective irradiance are the key inputs for the PV performance model. Alternatively, a data-driven approach levering machine learning and AI algorithms can accurately model the PV plant's behaviour [134]. The latter approach has limitations when applied to other systems. It fails if, for example, two components compensate for each other in the basic data set used for learning.

Digital twin models for PV plants require key data inputs: irradiance, temperature, electrical properties of components, and 3D geospatial context. Irradiance and temperature are sourced from local weather stations and on-site sensors. Electrical properties come from manufacturers' specs and real-time monitoring. The 3D geospatial context is obtained through satellite images and surveys and local measurement for example of nearby shading objects. These inputs are essential for accurate performance simulation and issue identification.

The modelled outputs of the digital twin-based performance models are used in the O&M of PV plants. By comparing real-time performance data with expected performance derived from the digital twin, asset managers can identify deviations and potential faults. This allows for predictive maintenance (see Section 5.3.4), where issues are addressed before causing significant downtime or loss of efficiency. Additionally, modelled data help optimize maintenance schedules, like cleaning [135], reduce operational costs, and improve the overall reliability of the PV plant. Advanced data-driven techniques, such as AI and robotics, are increasingly integrated into O&M practices to enhance the efficiency and effectiveness of maintenance activities.

5.3.2 Plant Performance Assessment by KPIs

To assess whether PV plants are operating according to expectations, it is necessary to compare the data flow from the physical entity, e.g., the monitoring data, to expected values based on predefined performance indicators and to a data flow coming from the virtual entity, detailing the modelled or expected performance parameters. However, this is only feasible within the measurement uncertainties of the real measurement data used in modelling the Digital Twin. Nevertheless, the resulting uncertainties in the output data flow of the Digital Twin should always be considered in high quality decision-making processes.

The “Operation and Maintenance Best Practice Guidelines” by SolarPower Europe [136] give an exhaustive overview of KPIs to be evaluated during O&M of PV plants. In **Error! Reference source not found.**, we present the most relevant KPIs in the context of PV digital twins and discuss the requirements in terms of input data flows for each of these KPIs, as well as applicable standards for calculating these KPIs.

Table 6: Overview of PV plant KPIs calculated from data originating from the physical entity and required input data flows. The list of KPIs is sourced from [136]. The calculation standards shown either provide full norms on how to calculate the KPI itself or (some) of the necessary inputs.

KPI	Short description	Necessary data flows	Calculation standards
Reference Yield	Energy yield attainable assuming operation at STC, in kWh/kWp Can be calculated for the whole system, or for e.g. the PV arrays constituting it	Rated (STC) power of the PV modules Solar irradiance in the plane of the PV modules, measured in-plane or modelled using: • Auxiliary solar irradiance measurements; • 3D geometrical information of the PV system array; • solar irradiance transposition models using measured irradiance, atmospheric parameters, and solar position data.	IEC 61724-1:2021

Table 6 (continued)

KPI	Short description	Necessary data flows	Calculation standards
PV array yield	Measured array DC energy yield in kWh/kWp for a desired evaluation period	Rated (STC) DC power of the array at the desired level of aggregation; Array DC electricity yield at the desired level of aggregation and evaluation period	IEC 61724-1:2021
Final system yield	Measured AC energy yield in kWh/kWp for a desired evaluation period	Rated (STC) DC power of the plant at the desired level of aggregation; Plant AC electricity yield at the desired level of aggregation and evaluation period	IEC 61724-1:2021
Performance Ratio	Yield divided by reference yield, can be calculated for full system or for e.g. PV arrays constituting it	PV array yield or Final system yield; Reference yield for PV array or System	IEC 61724-1:2021
Temperature-corrected Performance Ratio	Performance Ratio, corrected for the influence of temperature on PV module performance	Specific yield; Reference yield; Module temperature - measured or modelled; Module temperature coefficient of power at STC conditions	IEC 61724-1:2021
Technical Availability	Percentage of time a PV plant is operational	Total time during which solar irradiance is above a minimum irradiance threshold; Downtime	IEC TS 63019:2019
Technical Tracker Availability	Percentage of time a tracker is operational	Total time during which solar irradiance is above a minimum threshold; Tracker downtime	IEC TS 63019:2019
Contractual availability	Technical availability, where the downtime is reduced with excluded factors agreed in the O&M contract	Total time during which solar irradiance is above a minimum irradiance threshold; Downtime; Part(s) of Downtime contractually excluded	IEC TS 63019:2019
Contractual tracker availability	Technical tracker availability, where the tracker downtime is reduced with excluded factors agreed in the O&M contract	Total time during which solar irradiance is above a minimum irradiance threshold; Tracker downtime; Part(s) of tracker downtime contractually excluded	IEC TS 63019:2019
Energy based availability	Energy generated as a fraction of the total of energy generated plus energy losses.	Metered energy generation (kWh); Lost energy generation, calculated (kWh)	IEC 61724-1:2021

Table 6 (continued)

KPI	Short description	Necessary data flows	Calculation standards
Expected Performance Ratio	Performance ratio calculated using modelled system power output	Modelled system specific yield Modelled system reference yield determined from: • Solar irradiance in the module plane determined from model input irradiance data; Power plant rated (STC) capacity used in the model.	IEC TS 61724-3:2016
Expected Yield	Reference yield multiplied with the expected performance ratio over a desired evaluation period	Reference Yield over desired period of evaluation; Expected Performance Ratio over desired period of evaluation.	IEC TS 61724-3:2016
Energy Performance Index	Ratio between the specific yield and expected yield.	Specific yield; Expected yield	IEC 61724-1:2001 IEC TS 61724-3:2016

5.3.3 Fault Detection using Machine Learning and AI

As described above, solar PV plants and their operating environment provide a complex stream of monitoring data, with faults and degradation modes introducing different anomalies in the electrical signatures of the PV plants string and inverter level monitoring data. Current data analytics software tools enable the automatic and real-time calculation of KPI's mentioned in Section 5.3.2, in turn, these KPI's can be used for threshold-based fault detection. However, more sophisticated approaches based on the application of machine learning (ML) and artificial intelligence (AI) in fault detection for solar PV plants have seen significant advancements in recent years. ML and AI based algorithms enables fast and accurate fault detection from monitoring and field inspection data [137]. This advancement is seen as necessary to deal with increasingly complex and voluminous data flows from increasingly large PV systems [138], [136].

Modern fault detection techniques leverage a variety of ML and AI methodologies, from regression based techniques used to detect outliers, such as (non)linear, multiple regression or those using regression trees, to classification techniques based on K-nearest neighbours (KNN), logistic regression, artificial neural networks (ANNs) and support vector machines (SVMs), to clustering tools like K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [137]. Additionally, techniques like deep learning, convolutional neural networks (CNNs), and ensemble learning methods such as Random Forest or different forms of gradient boosting are employed to analyse large datasets generated by PV systems, including electrical parameters, thermal images, and environmental data [139] [140].

Approaches for detecting faults from imagery data are commonly based on deep learning and neural networks, such as an example using aerial infrared thermography (aIRT) imagery to detect module faults [20], which employs a coupling of well-known deep learning approaches of real-time object detecting using a single shot method called You Only Look Once (YOLO), and a region-based convolutional neural network (R-CNN). The developed tool can automatically detect individual PV modules from aIRT imagery, detect faults and quantify the associated power losses [20].

The integration of ML and AI with big data and Internet of Things (IoT) technologies has further enhanced fault detection capabilities. By utilizing data from various sensors and monitoring devices, these systems can perform continuous real-time analysis, identifying subtle performance deviations that may indicate potential faults [140]. This approach not only improves fault detection accuracy but also minimizes false alarms, thereby further optimizing maintenance schedules and reducing operational costs.

5.3.4 Predictive Maintenance

The maintenance of photovoltaic plants typically involves technical inspections and repairs to ensure the optimal functioning of the PV plant. In recent years, the maintenance of PV power plants has increasingly become automated. According to maintenance scheduling it can be divided into three groups, namely preventive, corrective and predictive maintenance.

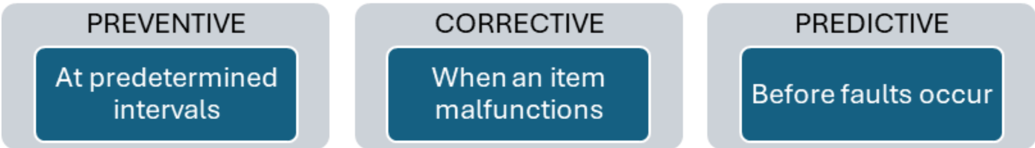


Figure 20: Types of maintenance strategies

Preventive maintenance includes regular verifications that the key components of the PV plant are in working order, it is carried out in regular intervals according to the annual maintenance plan proposed by the individual O&M specifications. It must also ensure that equipment warranties are maintained and reduce the chances of malfunctions. The maintenance plan frequency is defined by the specific PV component manufacturer, as well as the remaining useful operating time before replacement. An example of corrective maintenance would be a thermographic inspection to identify for instance defective PV modules or hot spots within a PV plant.

On the other hand, corrective maintenance corresponds to any activity that requires immediate action and repair to restore PV plant systems, equipment or component to a functioning state. Corrective maintenance includes fault diagnosis, temporary repair and permanent repair. Corrective maintenance involves identifying the root cause of failures, typically related to manufacturer/model/serial number issues, installation errors, or environmental conditions like temperature inside enclosures. Currently both preventive and corrective maintenance are prevalent in PV plant operations.

Predictive maintenance in the PV industry involves using advanced techniques such as machine learning algorithms on monitoring data to anticipate and address potential issues before they occur. This approach ensures the optimal performance and extended lifespan of PV systems. Advanced predictive maintenance uses historical data to create a baseline model, then analyses new data to detect anomalies.

It is defined as a condition-based maintenance carried out following a forecast derived from the analysis and evaluation of the significant parameters of the degradation of the item (according to EN 13306). These parameters can be either based on a (1) monitoring software system or (2) on the analysis of tests and inspections carried out as part of the preventative maintenance and stored in a smart digital representation of the full PV system.

According to the literature review by Bosman et al [141], the current approaches and opportunities for PV predictive maintenance can be divided into four groups based on cost and detection accuracy, namely the manual diagnostic, failure mode and effect analysis, machine learning and forecast and real time sensors and other real time performance data.

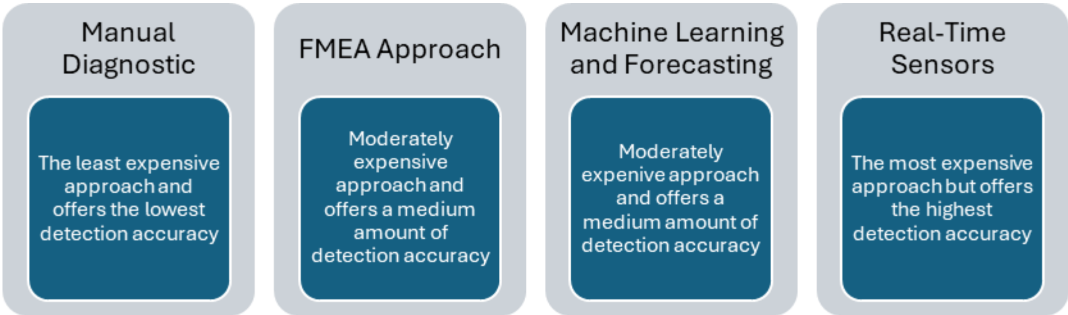


Figure 21: Approaches for predictive maintenance. FMEA: Failure mode and effect analysis Bosman et all [125].

The manual diagnostic covers both quantitative and qualitative approaches. The qualitative approach includes visual inspection of the system and individual components as well as the infrared thermography of PV modules. Quantitative analysis covers IV curve analysis and insulation resistance measurements of PV modules. These methods are effective at identifying issues with the PV modules but do not consider the PV system. One of the challenges associated with using manual diagnostics is the extensive time required to evaluate the entire plant, as well as the potential for human error due to the manual nature of the data collection process and the great uncertainty when drawing conclusions about the electrical performance data of individual PV modules.

Failure Mode and Effect Analysis (FMEA) is a semi-quantitative method used to prevent failures in PV components by identifying their causes and effects [142]. The FMEA has been used to identify components with the highest risk of failure. Analysis generally uses historical data to identify components prone to failure or examines how specific climatic conditions lead to failures. An example of Failure Mode and Effects Analysis (FMEA) involves conducting a failure analysis on an inverter to determine the root cause. The inverters experienced failures under specific climatic conditions, while operating correctly in other environments. Electrical inverter boards were examined in a laboratory setting, where it was determined that the issue was related to the power relay. Further analysis identified moisture ingress as the cause of component failures within the relay. The FMEA study concluded that these inverters should be installed exclusively in dry climatic conditions or alternatively moisture insulation should be improved.

The failure mode, effects, and criticality analysis methodology seek to reduce the impact of potential failures in photovoltaic systems and thus increase the electrical performance. Considering the failure probabilities of PV system elements as well as the amount of the associated loss is essential for predictive maintenance models. Improving these models involves assessing component failure risks.

Machine learning and forecasting as a predictive maintenance approach covers various methods to estimate solar energy based on weather forecast data. The overall idea is to increase the prediction accuracy of PV production and quickly identify any underperformance. Usually, historical data is used to define the baseline PV performance model which is later used to identify anomalous behaviour. The baseline model shows the behaviour across the component's operational conditions. Data analysis involves using new data with the baseline model to identify consistent deviations in the component's behaviour.

Various methods and simulation models have been developed to increase the weather forecast accuracy by identifying any patterns in the weather data which could help predict weather parameters in different time horizons and sizes of the local forecast area. Machine learning methods are utilizing historical data to train weather prediction models to estimate solar irradiance and temperature to give accurate prediction about the solar energy output. Fault detection algorithms can be utilized to compare estimated output to actual output by applying error thresholds based on fault-free systems. These methods allow identification of total and partial productivity loss, enabling pre-emptive maintenance actions.

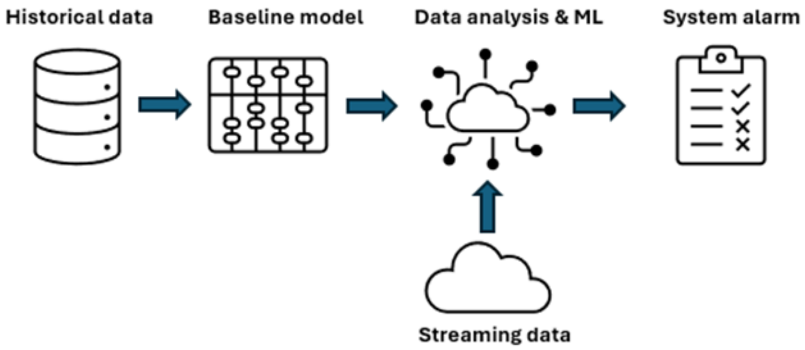


Figure 22: ML based predictive maintenance analytics.

Predictive maintenance using real-time sensors involves onsite devices that provide information about their state, allowing operation and maintenance service providers to evaluate trends or events. Sensor manufacturers should provide a detailed list of status and error codes, along with their meanings and the effects on device function. Status and error codes should be standardized within the same brand's inverters and dataloggers, and eventually across all manufacturers. The Operations & Maintenance (O&M) provider facilitates predictive maintenance through continuous or periodic monitoring, supervision, forecasting, and performance data analysis (including historical performance and anomaly detection) of the solar PV power plant at various levels such as the DC array, transformer, inverter, combiner box, and string level. This can identify subtle trends that would otherwise go unnoticed until the next round of circuit testing or thermal imaging inspection and that indicate upcoming component or system failures or underperformance (e.g., at solar PV modules, inverters, combiner boxes, trackers., etc. level) [143].

To summarize, due to the extensive deployment of PV systems, advanced automation and remote monitoring are required to ensure the quality of system operations. Despite this, there are still challenges and opportunities in this area. These monitoring tools can precisely detect when a PV system is underperforming, but they generally do not provide specific actionable insights that an owner can use to enhance solar performance. Future predictive maintenance systems should be able to differentiate between imminent failures, anomalies that do not result in service disruption, performance degradation, and planned maintenance activities. Additionally, these systems should provide an estimation of the remaining useful life of components and e.g., a recommendation for the replacement of components if, for example, a certain number of PV modules are underperforming. Moreover, they should be able to schedule an optimal plan for corrective maintenance activities.

With the help of predictive maintenance if can be possible to better plan and optimize corrective actions will improve the balance between covered and uncovered corrective maintenance activities in the contracted services, thus producing a greater impact on the O&M budget.

IoT Devices and general Sensors [144]

The Internet of Things (IoT) is a system that integrates various devices, including sensors and actuators, to monitor and control industrial processes. In recent years, IoT has been extensively discussed and implemented for the inspection and monitoring of PV plants and the collection of data for analyses and automation primarily for PV roof systems in urban areas. Employing IoT offers several advantages, such as increased efficiency, improved accuracy, and reduced economic costs, In the sense that once it is installed, then it will decrease the cost of O&M if a constant stream of system information from the sensors flows within the digital twin and supports decision making for operation and maintenance (i.e., fault detection or actuator control). Permanently deployed sensors in the field (beyond the classical sensors like pyranometers or reference cells) comprise for example smart inverters that are capable of IV curve sweeping of whole strings. Another example are cloud cameras for cloud vector motion detection with the aim to enhance power and irradiance forecast that is being used to detect anomalies of the system. Also, especially in soiling-prone region, soiling sensors are permanently installed on site [135]. In the case of integrated systems like FPV or APV, typically more sensors are deployed to monitor the interaction of the systems with its environment or to assess the microclimate within the system. As another example, for an AgriPV system not only PV performance is being recorded but also soil quality, PAR irradiance, ambient Temperature, soil temperature, moisture, dew point etc.

Fault Detection using Peer-PV systems and Machine Learning

The PhD of A. Alcaniz [145] is focused on Machine Learning and analytical PV power prediction in relation to maintenance and in previous publications [13] also on the so-called performance-to-peer approach by adding system characteristics and optimizing with ML techniques and shows promising results. The methodology has been tested in a fleet of more than 12,000 Dutch systems with up to 7 years of data per system. The proposed model achieves an average R^2 of 94.1% and an NRMSE of 0.05, outperforming in terms of R^2 the baseline model by 1.4 points, and the analytical approach by 3.8.

6 OUTLOOK AND DEVELOPMENTS

6.1 Business models for data along the PV value chain

Digitalisation plays a major role in cost reduction through decreased effort and higher efficiency in data curation and stewardship. Especially well-developed information models (Chapter 3 on data models) support the handover of crucial project data in an interoperable way between the different phases of a PV project. Figure 23 shows the circular phases of a project (engineering, procurement, construction, commissioning, operation). Between all these phases, data are being exchanged, analysed, stored and documented, and very often it is still very difficult to access this information, as it is often not properly FAIRified (see Chapter 3).

Figure 23: Circular Phases of PV projects, adapted from Solar Power Europe EPC guide V1.0.

Here, the large amount of data that a PV project and its digital twin generate over its lifetime must be stored and made accessible in a sustainable way. One such way to do that is within a cloud infrastructure of a federation of trust, also known as data spaces.

On a European level, the digitalisation of the energy system and data exchange therein are seen as key enablers for a resilient energy system, as set forth in the EU action plan on "Digitalising the energy System" [146]. In principle, this is based on a set of European directives and action plans. Since 2014, the European Commission has implemented various measures to promote the development of a data-driven economy. These measures include the Regulation on the free flow of non-personal data, the Cybersecurity Act, the Open Data Directive, and the General Data Protection Regulation [147]. Ultimately, these initiatives support the "European Green Deal, 2019" and the "European Strategy for Data and AI, 2020"

Possible business models include the integration and connection of a digital twin with an Enterprise Resource Planning system (ERP), or the deployment of AI (e.g., large language models, LLMs) to extract metadata and information from unstructured documentation heaps. During system monitoring and maintenance, machine-learning based Root Cause Analysis and pre-emptive maintenance become increasingly important.

Currently, Business models are enabled by data and integration and other developments in R&D, like data spaces, federations to avoid data silos, a handover of data between phases in an interoperable way.

Digitalisation of O&M processes in PV systems lead to significant cost reductions and enhanced operational efficiencies. The integration of technologies such as virtual reality (VR) and augmented reality (AR) is increasingly being used, enabling remote training and support for technicians. This facilitates complex maintenance tasks without the need for on-site presence, thereby saving time and travel costs.

In the realm of fully unsupervised automated O&M and inspection, advanced solutions like "drone-in-a-box" systems are emerging. These systems autonomously deploy drones for routine inspections and maintenance of PV installations, significantly reducing manual labor and the associated costs, especially for large PV assets. This technology allows for real-time data collection and analysis, ensuring timely interventions when issues are detected.

Also, Vegetation Management Robots, or mowing robots, are actively used for vegetation control around PV installations. While effective on flat terrains, there is still a need for solutions that can operate on slopes. Walking Robots: Future developments may include robotic systems that can navigate uneven terrain to perform visual inspections and maintenance tasks.

The integration of IoT sensors into PV modules allows the deployed of a large number of low-cost sensors. These sensors provide valuable data for predictive maintenance practices. By analysing this data, operators can forecast potential failures and schedule maintenance proactively, thereby minimizing downtime and repair costs.

The commercial integration of these advanced technologies is crucial for the scalability and effectiveness of digitalised O&M in the PV sector. As the industry evolves, we anticipate a future where fully robotic, unsupervised automated maintenance becomes the norm. This will be driven by advancements in AI, machine learning, and robotics, ultimately leading to more efficient and cost-effective operations.

In conclusion, digitalisation is set to transform the O&M landscape of the photovoltaic sector, with significant implications for cost reduction and operational efficiency. The ongoing development and adoption of innovative technologies will play a critical role in shaping the future of PV maintenance and management.

6.2 Outlook for AI and digitalisation in PV

6.2.1 The role of AI and digitalisation in PV

The integration of AI into O&M of PV systems potentially brings large improvements to the solar energy sector. AI-driven digitalisation enhances the O&M of PV plants by breaking down silos and enabling data-driven decision-making [10]. Applications of AI in PV systems, such as predictive maintenance and energy forecasting, are already improving performance optimization and reliability if the underlying data is based on reliable sensors and sources [10]. Where current applications of AI in PV O&M are largely based on “classical” data-driven AI and ML tools for, for instance, failure detection and forecasting, the new wave of Generative AI with language and vision models enables many new possibilities for PV [148]. A strong example is multi-agent drive automated failure detection and intervention in PV O&M, where a combination of AI and robotics for field inspection and intervention can drastically lower O&M costs and labour requirements, decoupling PV growth from labour constraints [148] [149]. Due to its understanding of natural language, generative AI also offers the potential to increase efficiency and accuracy of technical reporting and improve interaction of human operators with data analytics platforms.

By decoupling PV scale and growth from labour constraints, digitalisation and AI are facilitating the massive deployment of photovoltaics within the energy system, where AI's role in PV digitalisation extends to enhancing grid integration and optimizing energy storage solutions. For instance, by combining detailed AI-driven PV system modelling, grid modelling, power and grid forecasting, PV systems, potentially with support of battery storage can provide virtual inertia comparable to the real physical inertia offered by generators in traditional fossil fuel plants [150]. This means that PV systems become connected systems [149], that are “grid-friendly” and offer smooth power output under challenging conditions. As AI continues to evolve, integrating it into PV digitalisation will be crucial for advancing the efficiency, reliability, and scalability of solar energy systems, solidifying the role of PV in the energy transition [150] [151].

6.2.2 Fleet-based modelling, outlook from research to industry

Fleet-based modelling has become increasingly feasible with the advancement of computing techniques. A representative example is presented in Section 4.4.2, showcasing a data-driven digital twin developed for a fleet of real-world PV systems. This digital twin adopts a foundation model approach [152] which overcomes the limitations of single-site models and enables more generalized and accurate predictions across diverse PV system performances. In the research space, large-scale analysis is now possible for fleet-based PV performance assessment tasks such as missing data imputation [113], degradation pattern identification, and performance loss rate (PLR) analysis [114] [115] [153] [154] [155]. Given the inherently spatiotemporal nature of PV system data, graph-based approaches—particularly spatiotemporal Graph Neural Networks (stGNNs)—have shown considerable promise for fleet-wide modelling [114], [113], [154]. Recent work has demonstrated the successful application of st-GNNs to PLR analysis across a fleet of 100,000 systems [115].

6.3 Cyber Security Outlook

As digital twins become increasingly important in the planning, procurement, operation, and administration of PV applications, their complexity, user demands, and reliance on digital property and connectivity introduce new cybersecurity vulnerabilities. The entire energy sector is a critical component of modern infrastructure, requiring robust security policies and practices to protect against evolving cyber threats. Security measures are therefore essential not only to protect sensitive operational and performance data but also to ensure continuous power generation, maintain grid stability, and prevent serious damage.

The importance of cybersecurity also extends significantly to the conceptualization, development, deployment and ongoing application of digital twins in PV. Right from the start, at the conceptual stage, cybersecurity considerations should be integrated into the digital twin’s foundational design, aligning with applicable cybersecurity standards, national regulatory frameworks, and internal company guidelines. This approach prevents inherent vulnerabilities and reduces risks that may become difficult or impossible to correct at later stages.

The development (including also procurement of external components, if necessary) must prioritize products, software and other third-party solutions that possess valid security certificates, recent updates, and trustable sources. Secure deployment requires careful integration into existing digital infrastructure and interfaces, particularly the verification of functional security protection, to avoid the introduction of new vulnerabilities. Throughout the application phase, continuous maintenance, systematic monitoring, and regular cybersecurity audits safeguards digital twins against ongoing threats, ensuring sustained reliability and effectiveness.

Addressing cybersecurity challenges proactively enables PV service providers and operators to establish essential conditions for secure digital twin operations within an increasingly digitized and efficient energy landscape.

The growing number of external cyber-attacks underscores the urgent need for robust cybersecurity measures in energy systems. Operators of critical infrastructures muss comply with minimum IT security standards and implement protective measures to protect their systems against cyber-attacks. Effective cybersecurity strategies require a comprehensive approach involving both organizational and technical actions. Internationally, several initiatives emphasize the significance of protecting critical energy infrastructures. At the European level, the NIS 2 Directive[3] (network and information systems, Directive 2022/2555) establishes standardized cybersecurity requirements for critical infrastructure sectors, including energy. Additionally, the EU Cybersecurity Act[4] (Regulation EU 2019/881) establishes a European cybersecurity certification framework for ICT and EU Directive on the Resilience of Critical Entities[5] (EU 2022/2557) aims to enhance both physical and digital resilience of energy infrastructures. On the global scale, organizations such as the International Electrotechnical Commission (IEC), notably standards such as IEC 62443-5 (industrial communication networks), IEC 62351-9 (energy management systems), IEC 27001 (general guide for SME), IEC 27002 (general security controls), and IEC 27019 (energy utility industry), as well as the National Institute of Standards and Technology (NIST), particularly through its Cybersecurity Framework CSF (particularly version 1.0, on critical infrastructures), develop specialized cybersecurity standards and frameworks tailored specifically to energy sector needs, coordinating security practices across borders.

Cybersecurity involves protecting digital systems, networks, and data – where digital twins are utilized – against unauthorized access or disruptions. For photovoltaic applications, digital twins can manage sensitive operational data, system configuration, service scopes, or performance metrics. The following table describes cybersecurity terms[6] relevant for digital services in energy systems.

Table 7 Cybersecurity terms relevant for digital services in energy systems.

Term	Definition
Vulnerability	Weakness in information systems, system security procedures, internal controls, or implementations exploitable by cyber threats
Incident Response / Handling	Actions and procedures aimed at mitigating security breaches and violations effectively

3 https://eur-lex.europa.eu/eli/dir/2022/2555

4 https://eur-lex.europa.eu/eli/reg/2019/881/oj

5 https://eur-lex.europa.eu/eli/dir/2022/2557/oj

6 https://csrc.nist.gov/glossary

Table 7 (continued)

Term	Definition
Resilience	The capability of information systems to continue functioning under adverse conditions or to quickly recover within an acceptable timeframe
Availability	Ensuring timely and reliable access to and use of information systems and data
Integrity	Guarding against unauthorized modification or destruction of information, including ensuring authenticity and non-repudiation
Authentication	Verifying the identity of a user, process, or device as a prerequisite to allowing access to resources within an information system
Confidentiality	Preserving authorized restrictions on information access and disclosure, including the protecting personal privacy and proprietary information

Effective cybersecurity in respect to digital twins and services in PV applications must be built on practical, actionable principles that involve all relevant stakeholders. According to the Solar Best Practices Guidelines[7] and NREL's report on cybersecurity in PV plant operations[8], cybersecurity begins with clearly defined roles and responsibilities across the value chain. System manufacturers and software developers are responsible for integrating security features into digital twin platforms from outset, such as secure authentication, encrypted data transmission, and role-based access control. Operators and asset owners must implement and maintain these measures by enforcing access restrictions, updating firmware and software regularly, and monitoring PV systems for irregular activity.

Service providers and integrators play a critical role in ensuring secure system architecture during deployment. They must verify component trustworthiness and adhere to recommended integration protocols. Meanwhile, monitoring and maintenance teams are responsible for regular audits, identifying and addressing vulnerabilities, and responding to incidents immediately.

Practical cybersecurity measures include using complex, regularly updated passwords and access controls, implementing multi-factor authentication for access to critical components, isolating networks (e.g., separating SCADA and administrative IT procedures), and establishing comprehensive backup and recovery procedures. Rather than treating cybersecurity as an IT-only issue, it must be approached as a shared responsibility embedded in every phase of digital twinning.

Neglecting cybersecurity in digital twin applications can lead to significant impacts on PV systems and companies:

---

[7]https://solarbestpractices.com/guidelines/detail/data-management-and-high-level-monitoring#chapter354

[8]Walker, Andy, Jal Desai, Danish Saleem, and Thushara Gunda. 2021. Cybersecurity in Photovoltaic Plant Operations. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5D00-78755. https://www.nrel.gov/docs/fy21osti/78755.pdf.

Table 8: Impacts on PV systems when neglecting cybersecurity.

Impacts Area	Consequences
Operational	Unexpected system downtime, reduced availability and energy output, and compromised grid injection; disruption of service, real-time monitoring, technical safety, control and other elemental functions
Economic	Burdens due to incident response, recovery or emergency repair costs, ransom payments, damage to reputation or investor confidence, increase of insurance costs
Data-related	Leakage, manipulation or loss of sensitive data such as key performance indicators, financial figures, operational settings, contractual obligations or personnel details; impair of operational decision-making and compromise competitive advantages
Regulatory and legal	Non-compliance can result in legal action, significant fines, operational restrictions, facing audits, public scrutiny, litigations

7 CONCLUSIONS

This report "Digitalisation and Digital Twins in Photovoltaic Systems" explores the integration of digital technology within the PV sector in five chapters. It emphasizes the importance of digital twins, defined as virtual representations of PV systems that utilize real-time data to enhance decision-making processes across the system's lifecycle. Key topics include the role of data models, the significance of standardized data for interoperability, risk analysis, and digitalisation's impact on the efficiency and reliability of PV operations. The report also highlights the ongoing challenges of terminology consistency and data governance in the PV industry.

Key takeaways from this report are:

- Digitalisation significantly contributes to risk analysis in PV projects, allowing stakeholders to quantify and mitigate risks associated with component failures, design flaws, and environmental factors.
- A digital twin is a virtual representation of a PV system that is continuously updated with real data and used for simulation and decision support throughout the system's lifecycle.
- The emphasis on the digital twin as a core concept signals its potential to revolutionise how PV systems are designed, operated, and maintained, ultimately contributing to the sector's growth and sustainability in the energy transition.
- Robust, standardised data models and ontologies—such as MDS-Onto—are essential for interoperable, FAIR data, enabling effective data sharing and forming the foundation for successful digital twin implementation in the PV industry.
- Two approaches to digital twinning are discussed: physics-based digital twins, which use physical models to simulate behaviour, and data-driven digital twins, which rely on real-world data to model system performance.
- By leveraging data from both the physical and virtual entities, digital twins enable detailed simulations of expected power outputs and help identify deviations from anticipated performance.
- The integration of artificial intelligence (AI) and the Internet of Things (IoT) are key components in optimising operations and maintenance (O&M) processes for PV systems.
- Cyber security is of utmost importance to be considered at all levels of digitalisation.

Chapter 2, “Digitalisation in the PV sector”, investigates the transformative role of digitalisation within the PV sector, emphasizing its profound impact on various stages of the PV project lifecycle. The chapter begins by illustrating how digitalisation enhances each phase, from manufacturing PV components to operation and maintenance, ultimately aiming for a more integrated and efficient system. It highlights the necessity of establishing a cohesive digital strategy that unites disparate digital efforts across the entire value chain, recognizing that many initiatives are currently disconnected and that full integration remains a significant challenge. Digitalisation significantly contributes to risk analysis in PV projects, allowing stakeholders to quantify and mitigate risks associated with component failures, design flaws, and environmental factors. By leveraging data analytics and real-time monitoring, digitalisation facilitates proactive risk management, thereby enhancing reliability and performance outcomes. Central to the discussion is the concept of the digital twin, which serves as a pivotal element in the digitalisation of PV systems. A digital twin is a virtual representation of a PV system that is continuously updated with real data and used for simulation and decision support throughout the system's lifecycle. The chapter argues that the comprehensive implementation of digital twins can significantly optimize performance, maintenance, and overall operational efficiency in PV systems.

Chapter 2 concludes by reinforcing the idea that while digitalisation in the PV sector offers remarkable opportunities for improvement, it also requires collaborative efforts among all stakeholders to ensure effective implementation and integration of digital technologies. The emphasis on the digital twin as a core concept signals its potential to revolutionize how PV systems are designed, operated, and maintained, ultimately contributing to the sector's growth and sustainability in the energy transition.

Chapter 3, “The Role of Data Models and Data Structures”, discusses the critical significance of data models and data structures in the PV sector, particularly in the context of digital twinning. It begins by explaining the concepts of taxonomies and ontologies, highlighting their roles in organizing and structuring data. While taxonomies provide a static classification, ontologies offer a dynamic representation that enhances semantic understanding and interoperability.

The chapter presents the current status of PV taxonomies, notably the Orange Button Taxonomy, which standardizes financial and design aspects within the PV value chain. However, it notes that the development of ontologies in the PV domain is still emerging, with significant work needed to unify terminologies and enhance data sharing. A key focus is the Materials Data Science Ontology (MDS-Onto), which aims to address terminology inconsistencies and improve data integration across the PV sector. The chapter outlines how MDS-Onto provides a framework for ontology creation, thereby facilitating FAIR (Findable, Accessible, Interoperable, Reusable) data practices. The chapter concludes by emphasizing the necessity of adopting standardized and interoperable data models to enhance data sharing and collaboration among stakeholders in the PV industry. It also stresses that well-defined data structures are essential for ensuring the reliability and performance of PV systems.

Chapter 4, “Definition of Digital Twins in PV”, focuses on defining Digital Twins (DT) specifically within the context of PV systems. It begins by outlining the foundational concept of a digital twin, which integrates three key components: the physical entity, the virtual entity, and the data flows between them. The physical entity represents the actual PV system, while the virtual entity is a digital representation that employs various models to simulate the system's behaviour. Data flows refer to the information exchanged between these entities, crucial for real-time monitoring and decision-making. The chapter describes different types of digital twins, including Digital Models, Digital Shadows, and fully integrated Digital Twins, with the latter representing the most complex form of integration. It emphasizes that digital twins can span the entire lifecycle of a PV system, supporting activities such as planning, design, operation, and maintenance. Moreover, two approaches to digital twinning are discussed: physics-based digital twins, which use physical models to simulate behaviour, and data-driven digital twins, which rely on real-world data to model system performance. The latter is particularly valuable when the underlying physics is complex or poorly understood, allowing for a more straightforward representation of system behaviour based on actual operational data. The chapter concludes by highlighting the potential applications of digital twins in the PV industry, including enhanced performance monitoring, predictive maintenance, and improved decision support. Overall, it establishes digital twins as a transformative tool in optimizing the operational efficiency and reliability of PV systems throughout their lifecycle.

Chapter 5, “Digital Twins in PV O&M: Data Flows and Applications”, examines the role of digital twins in the operation and maintenance (O&M) of photovoltaic (PV) systems, emphasizing the significance of data flows in ensuring their functionality. It starts by detailing the continuous monitoring data collected from various physical components of the PV systems, such as inverters and weather stations. This real-time data is crucial for assessing the performance of the PV plant and is typically transmitted to a cloud-based monitoring system. The chapter further explores field inspection data, which are gathered during scheduled maintenance or in response to detected anomalies. These inspections utilize techniques like visual assessments, thermography, electroluminescence testing, and I-V curve tracing to identify issues that standard monitoring might overlook. Such data contribute to a comprehensive understanding of the physical condition of the plant and are essential for maintaining its long-term performance. The chapter also discusses the various applications of digital twins in PV O&M, focusing on their capacity to model plant performance, detect faults, and facilitate predictive maintenance. By leveraging data from both the physical and virtual entities, digital twins enable detailed simulations of expected power outputs and help identify deviations from anticipated performance. This predictive capability allows for timely interventions, enhancing operational efficiency and reliability. To summarize, Chapter 5 illustrates how digital twins, through effective data flows, empower PV systems to optimize performance, streamline maintenance processes, and ultimately contribute to the successful management of solar energy assets over their lifecycle.

Chapter 6, “Outlook and Developments”, discusses the future directions for digitalisation and the role of digital twins in the PV sector, emphasizing the impact of advanced technologies and the importance of robust business models for data management. It highlights the transformative potential of digitalisation in enhancing efficiency and reducing costs across the PV value chain. The chapter also elaborates on the integration of artificial intelligence (AI) and the Internet of Things (IoT) as key components in optimizing O&M processes for PV systems. The role of cybersecurity is highlighted. As the reliance on interconnected digital systems increases, vulnerabilities to cyber threats grow, making strong cybersecurity measures essential. The chapter emphasizes that from the conceptualization of digital twins to their deployment and application, cybersecurity must be integrated into every phase. This includes ensuring that security protocols align with industry standards and continuously monitoring for potential threats. A proactive approach to cybersecurity not only protects sensitive operational data but also ensures the reliability and effectiveness of PV systems. Overall, the chapter underscores the necessity of addressing cybersecurity challenges in the context of digital twins as the "elephant in the room", while also highlighting the advancements in AI and automation that can significantly enhance the performance and management of PV systems in the future.

REFERENCES

[1] SolarPower Europe, “Global Market Outlook for Solar Power: For Solar Power 2024-2028,” vol. 2024.

[2] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, “Industry 4.0,” *Bus Inf Syst Eng*, vol. 6, no. 4, pp. 239–242, 2014, doi: 10.1007/s12599-014-0334-4.

[3] L. Lüer *et al.,* “A digital twin to overcome long-time challenges in photovoltaics,” *Joule*, vol. 8, no. 2, pp. 295–311, 2024, doi: 10.1016/j.joule.2023.12.010.

[4] G. Kavlak, M. M. Klemun, A. S. Kamat, B. L. Smith, R. M. Margolis, and J. E. Trancik, “Nature of innovations affecting photovoltaic system costs,” *PLoS ONE*, vol. 20, no. 8, e0320676, 2025, doi: 10.1371/journal.pone.0320676.

[5] Dr. E. Franklin, “Solar Photovoltaic (PV) Site Assessment,” 2017. [Online]. Available: https://extension.arizona.edu/sites/extension.arizona.edu/files/pubs/az1697-2017.pdf

[6] W. P. U. Wijeratne, R. J. Yang, E. Too, and R. Wakefield, “Design and development of distributed solar PV systems: Do the current tools work?,” *Sustainable Cities and Society*, vol. 45, pp. 553–578, 2019, doi: 10.1016/j.scs.2018.11.035.

[7] M. Theristis *et al.,* “Blind photovoltaic modeling intercomparison: A multidimensional data analysis and lessons learned,” *Progress in Photovoltaics*, vol. 31, no. 11, pp. 1144–1157, 2023, doi: 10.1002/pip.3729.

[8] I. T. Horvath, A. A. Moya, F. Segata, J. Lemmens, T. Hall, and W. Vanheusden, “Automated PV digital twin-based yield simulation framework,” 2022. [Online]. Available: https://trust-pv.eu/reports/automated-pv-digital-twin-based-yield-simulation-framework/

[9] D. Moser, “Digitalisation in the PV sector.: IEA PVPS Task 13 Activity 2.4,” 2021. [Online]. Available: https://iea-pvps.org/wp-content/uploads/2021/10/02_David-Moser_digitalisation.pdf

[10] D. Moser, “Role of Digitalization in Operation and Maintenance of PV Plants: breaking silos,” 2021. [Online]. Available: https://iea-pvps.org/wp-content/uploads/2021/10/02_David-Moser_digitalisation.pdf

[11] SolarPower Europe, *Engineering, Procurement and Construction Best Practice Guidelines (Version 2.0)*. Brussels, Belgium, 2021. [Online]. Available: https://www.solarbestpractices.com/src/Frontend/Files/MediaLibrary/13/epc-best-practice-guidelines-v-2-0-ea4d7d3bc5.pdf

[12] S. Lindig, A. Louwen, D. Moser, and M. Topic, “Outdoor PV System Monitoring—Input Data Quality, Data Imputation and Filtering Approaches,” *Energies*, vol. 13, no. 19, p. 5099, 2020, doi: 10.3390/en13195099.

[13] A. Alcañiz, M. M. Nikam, Y. Snow, O. Isabella, and H. Ziar, “Photovoltaic system monitoring and fault detection using peer systems,” *Progress in Photovoltaics*, vol. 30, no. 9, pp. 1072–1086, 2022, doi: 10.1002/pip.3558.

[14] M. Green, E. Brill, B. Jones, and J. Dore, *Improving efficiency of PV systems using statistical performance monitoring: International Energy Agency Photovoltaic Power Systems Programme : IEA PVPS Task 13, Subtask 2 : report IEA-PVPS T13-07:2017*. Paris: International Energy Agency, 2017.

[15] A. Woyte, M. Richter, D. Moser, S. Mau, N. Reich, and U. Jahn, “Monitoring of Photovoltaic Systems: Good Practices and Systematic Analysis,” 2013, doi: 10.4229/28thEUPVSEC2013-5CO.6.1.

[16] W. Herrmann *et al.,* “Qualification of Photovoltaic (PV) Power Plants using Mobile Test Equipment,” International Energy Agency IEA-PVPS T13-24:2021, 2021.

[17] S. Koch, T. Weber, C. Sobottka, A. Fladung, P. Clemens, and J. Berghold, “Outdoor Electroluminescence Imaging of Crystalline Photovoltaic Modules: Comparative Study between Manual Ground-Level Inspections and Drone-Based Aerial Surveys,” 2016, doi: 10.4229/EUPVSEC20162016-5DO.12.2.

[18] T. Trupke, O. Kunz, and J. W. Weber, “Daylight Photoluminescence Imaging: Quantitative Analysis of String Voltage Mismatch and Balancing Currents,” *Progress in Photovoltaics*, vol. 33, no. 3, pp. 435–444, 2025, doi: 10.1002/pip.3866.

[19] S. Deitsch *et al.,* “Automatic classification of defective photovoltaic module cells in electroluminescence images,” *Solar Energy*, vol. 185, pp. 455–468, 2019, doi: 10.1016/j.solener.2019.02.067.

[20] J. A. Tsanakas, M. Stepec, P. Marechal, and D.-L. Ha, "From Pixels to Insights: A Software Prototype for AI-Driven Complete Diagnostics of PV Plants," *41st European Photovoltaic Solar Energy Conference and Exhibition*, 2024, doi: 10.4229/EUPVSEC2024/4BO.7.1.
[21] C. Buerhop-Lutz, T. Pickel, T. Winkler, I. M. Peters, and J. Hauch, "Analyzing the Power Prediction by Deep Learning Algorithm Using EL-Images," in *37th European Photovoltaic Solar Energy Conference and Exhibition. Proceedings of the international conference*, 2020, 4C1.4.
[22] M. Aghaei, A. Gandelli, F. Grimaccia, S. Leva, and R. E. Zich, "IR real-time analyses for PV system monitoring by digital image processing techniques," in *2015 International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, Krakow, Poland, 2015, pp. 1–6.
[23] S. Lindig, I. Kaaya, K.-A. Weiss, D. Moser, and M. Topic, "Review of Statistical and Analytical Degradation Models for Photovoltaic Modules and Systems as Well as Related Improvements," *IEEE J. Photovoltaics*, vol. 8, no. 6, pp. 1773–1786, 2018, doi: 10.1109/JPHOTOV.2018.2870532.
[24] SolarPower Europe, *Operation & Maintenance Best Practice Guidelines / Version 6.0*.
[25] SolarPower Europe, *Asset Management Best Practice Guidelines (Version 2.0)*. Brussels, Belgium.
[26] G. O. Hernandez and P. V. Chiantore, *D4.5 Progressive Revamping and Repowering of Utility Scale PV Plants*.
[27] I. A. Tsanakas *et al.,* "Toward Reuse-Ready PV: A Perspective on Recent Advances, Practices, and Future Challenges,"
[28] SolarPower Europe, *End-of-life Management: Best Practice Guidelines*.
[29] A. Livera, M. Theristis, G. Makrides, and G. E. Georghiou, "Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems," *Renewable Energy*, vol. 133, pp. 126–143, 2019, doi: 10.1016/j.renene.2018.09.101.
[30] M. Herz, G. Friesen, U. Jahn, M. Koentges, S. Lindig, and D. Moser, "Identify, analyse and mitigate—Quantification of technical risks in PV power systems," *Progress in Photovoltaics*, vol. 31, no. 12, pp. 1285–1298, 2023, doi: 10.1002/pip.3633.
[31] S. Lindig *et al.,* "Review of Technical Photovoltaic Key Performance Indicators and the Importance of Data Quality Routines," *Solar RRL*, vol. 8, no. 24, 2024, doi: 10.1002/solr.202400634.
[32] *Supernova Project.* [Online]. Available: https://supernova-pv.eu/
[33] O. Olayiwola, U. Cali, M. Elsden, and P. Yadav, "Enhanced Solar Photovoltaic System Management and Integration: The Digital Twin Concept," *Solar*, vol. 5, no. 1, p. 7, 2025, doi: 10.3390/solar5010007.
[34] Z. Song *et al.,* "Digital Twins for the Future Power System: An Overview and a Future Perspective,"
[35] *What is a digital twin?* [Online]. Available: https://www.ibm.com/topics/what-is-a-digital-twin (accessed: May 4 2025).
[36] M. La Rocca, *Advanced Algorithms and Data Structures,* 1st ed. Erscheinungsort nicht ermittelbar, Boston, MA: Manning Publications; Safari, 2021. [Online]. Available: https://learning.oreilly.com/library/view/-/9781617295485/?ar
[37] *"Orange Button PV Taxonomy." Open-Orange-Button.* [Online]. Available: https://github.com/Open-Orange-Button/Orange-Button-Taxonomy. [Accessed: Feb. 25, 2022]
[38] NREL (National Renewable Energy Laboratory (NREL), Golden, CO (United States)), *"Open data exchange standard for the Distributed Energy Resources industry | Orange Button FactSheet: RevB 2019-09,"* [Online]. Available: https://orangebutton.io/. [Accessed: Feb. 25, 2022]
[39] M. D. Wilkinson *et al.,* "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, p. 160018, 2016, doi: 10.1038/sdata.2016.18.
[40] A. Jacobsen *et al.,* "FAIR Principles: Interpretations and Implementation Considerations," *Data Intellegence*, vol. 2, 1-2, pp. 10–29, 2020, doi: 10.1162/dint_r_00024.
[41] B. Smith, "Ontology," in *The furniture of the world*, G. Hurtado and O. Nudler, Eds.: BRILL, 2012, pp. 47–68.
[42] T. Berners-Lee, R. Fielding, and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax," 2005.
[43] *"RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax (RFC2396)."* [Online]. Available: http://www.faqs.org/rfcs/rfc2396.html. [Accessed: Jul. 28, 2024]

[44] *“OAI/OpenAPI-Specification.” OpenAPI Initiative.* [Online]. Available: https://github.com/OAI/OpenAPI-Specification. [Accessed: Aug. 22, 2024]
[45] *“OpenAPI Specification v3.1.0.”* [Online]. Available: https://spec.openapis.org/oas/v3.1.0. [Accessed: Jul. 27, 2024]
[46] J. Schweikert, K.-U. Stucky, W. Sus, and V. Hagenmeyer, “A Photovoltaic System Model Integrating FAIR Digital Objects and Ontologies,” in *2022 IEEE 10th International Conference on Smart Energy Grid Engineering (SEGE)*, Oshawa, ON, Canada, 2022, pp. 48–57.
[47] F. Khosrojerdi, S. Gagnon, and R. Valverde, “Proposing an Ontology Model for Planning Photovoltaic Systems,” *MAKE*, vol. 3, no. 3, pp. 582–600, 2021, doi: 10.3390/make3030030.
[48] F. H. Abanda, J. Tah, and D. Duce, “PV-TONS: A photovoltaic technology ontology system for the design of PV-systems,” *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1399–1412, 2013, doi: 10.1016/j.engappai.2012.10.010.
[49] M. Booshehri *et al.,* “Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis,” *Energy and AI*, vol. 5, p. 100074, 2021, doi: 10.1016/j.egyai.2021.100074.
[50] B. P. Rajamohan *et al.,* “Materials Data Science Ontology(MDS-Onto): Unifying Domain Knowledge in Materials and Applied Data Science,” *Scientific Data*, vol. 12, no. 1, p. 628, 2025, doi: 10.1038/s41597-025-04938-5.
[51] L. Massel and A. Massel, “Ontologies as a Basis for Constructing Digital Twins in Energy,” in *2021 International Symposium on Knowledge, Ontology, and Theory (KNOTH)*, Akademgorodok, Novosibirsk, Russian Federation, 2021, pp. 1–5.
[52] H. Liu, Q. Gao, and P. Ma, “Photovoltaic generation power prediction research based on high quality context ontology and gated recurrent neural network,” *Sustainable Energy Technologies and Assessments*, vol. 45, p. 101191, 2021, doi: 10.1016/j.seta.2021.101191.
[53] O. Cabrera, X. Franch, and J. Marco, “3LConOnt: a three-level ontology for context modelling in context-aware computing,” *Softw Syst Model*, vol. 18, no. 2, pp. 1345–1378, 2019, doi: 10.1007/s10270-017-0611-z.
[54] D. Saba, F. Z. Laallam, H. E. Degha, B. Berbaoui, and R. Maouedj, “Design and Development of an Intelligent Ontology-Based Solution for Energy Management in the Home,” in *Studies in Computational Intelligence, Machine Learning Paradigms: Theory and Application*, A. E. Hassanien, Ed., Cham: Springer International Publishing, 2019, pp. 135–167.
[55] J. Wu, F. Orlandi, T. AlSkaif, D. O'Sullivan, and S. Dev, “Ontology Modeling for Decentralized Household Energy Systems,” in *2021 International Conference on Smart Energy Systems and Technologies (SEST)*, Vaasa, Finland, 2021, pp. 1–6.
[56] B. Bayerlein *et al.,* “A Perspective on Digital Knowledge Representation in Materials Science and Engineering,” *Adv Eng Mater*, vol. 24, no. 6, 2022, doi: 10.1002/adem.202101176.
[57] Wikipedia, *“Videotape format war,”* [Online]. Available: https://en.wikipedia.org/w/index.php?title=Videotape_format_war&oldid=1186933490. [Accessed: Jul. 27, 2024]
[58] K. Payne and C. Verhey, “Schema.org for research data managers: a primer,” *IJBDM*, vol. 2, no. 2, p. 95, 2022, doi: 10.1504/IJBDM.2022.128449.
[59] ““Welcome to the NCBO BioPortal | NCBO BioPortal.”,” [Online]. Available: https://bioportal.bioontology.org/. [Acces-sed: Jul. 27, 2024]
[60] C. Jonquet *et al.,* “Ontology Repositories and Semantic Artefact Catalogues with the OntoPortal Technology,” in *Lecture Notes in Computer Science, The Semantic Web – ISWC 2023*, T. R. Payne et al., Eds., Cham: Springer Nature Switzerland, 2023, pp. 38–58.
[61] *“Materials Data Science Ontology - Classes | Materials Open Laboratory MatPortal.”* [Online]. Available: : https://matportal.org/ontologies/MDS/?p=classes&conceptid=root#visualization. [Accessed: Jun. 17, 2024]
[62] A. Iliadis, A. Acker, W. Stevens, and S. B. Kavakli, “One schema to rule them all: How Schema.org models the world of search,” *Asso for Info Science & Tech*, vol. 76, no. 2, pp. 460–523, 2025, doi: 10.1002/asi.24744.
[63] A. Miles and J. R. Pérez-Agüera, “SKOS: Simple Knowledge Organisation for the Web,” *Cataloging & Classification Quarterly*, vol. 43, 3-4, pp. 69–83, 2007, doi: 10.1300/J104v43n03_04.

[64] S. Miranda, F. Orciuoli, and D. G. Sampson, "A SKOS-based framework for Subject Ontologies to improve learning experiences," *Computers in Human Behavior*, vol. 61, pp. 609–621, 2016, doi: 10.1016/j.chb.2016.03.066.
[65] T. Berners-Lee and J. Hendler, "Publishing on the semantic web," *Nature*, vol. 410, no. 6832, pp. 1023–1024, 2001, doi: 10.1038/35074206.
[66] P. Hitzler, "A review of the semantic web field," *Commun. ACM*, vol. 64, no. 2, pp. 76–83, 2021, doi: 10.1145/3397512.
[67] G. Kellogg, P-A. Champin, and D. Longley, *"JSON-LD 1.1: A JSON-based Serialization for Linked Data"* [Online]. Available: https://www.w3.org/TR/json-ld11/. [Ac-cessed: Nov. 13, 2022]
[68] T. Heath and C. Bizer, *Linked Data*. Cham: Springer International Publishing, 2011.
[69] A. Gschwend and O. Lassila, *RDF & SPARQL Working Group Charter*: W3C. [Online]. Available: https://www.w3.org/2025/04/rdf-star-wg-charter.html
[70] *JSON-LD-star*: W3C. [Online]. Available: https://json-ld.github.io/json-ld-star/publications/2021-02-18.html
[71] Ontotext, *What Is RDF-star?*: Ontotext-Graphwise. [Online]. Available: https://www.ontotext.com/knowledge-hub/fundamentals/what-is-rdf-star/
[72] *"Basic Formal Ontology (BFO) | Home".* [Online]. Available: https://basic-formal-ontology.org/bfo-2020.html. [Accessed: Jul. 28, 2024]
[73] G. O. Ponon, H. Sharma, and R. H. French, *CEMENTO: A package to view and write ontologies directly from draw.io diagram files.* Accessed: Dec. 9 2025.
[74] A. H. Bradley et al., *"FAIRmaterials: Python."* [Online]. Available: https://pypi.org/project/fairmaterials/
[75] J. E. Gordon et al., *"FAIRmaterials: R."* [Online]. Available: https://cran.case.edu/web/packages/FAIRmaterials/index.html. (Accessed: Jun. 23, 2024)
[76] B. P. Rajamohan et al., *"FAIRLinked: Python."* [Online]. Available: https://pypi.org/project/FAIRLinked/
[77] *FAIR Materials FindTheDocs.* [Online]. Available: https://cwrusdle.bitbucket.io/. [Accessed: Mar. 13, 2025]
[78] *JSON-LD Playground.* [Online]. Available: https://json-ld.org/playground/. [Accessed: Jul. 27, 2024]
[79] V. Wiens, S. Lohmann, and S. Auer, *"WebVOWL Editor: Device-Independent Visual Ontology Modeling"*
[80] *OSF Storage.* [Online]. Available: https://osf.io/jrdm9/files/osfstorage
[81] *CEMENTO — CEMENTO 0.13.0 documentation.* Accessed: Dec. 10 2025 "'.
[82] JGraph, *diagrams.net, draw.io.* [Online]. Available: https://www.diagrams.net/
[83] M. Jensen, G. de Colle, S. Kindya, C. More, A. P. Cox, and J. Beverley, "The Common Core Ontologies," 2024.
[84] *MDS-Onto FindTheDocs.* [Online]. Available: https://cwrusdle.bitbucket.io/
[85] B. Bayerlein *et al.,* "PMD Core Ontology: Achieving semantic interoperability in materials science," *Materials & Design*, vol. 237, p. 112603, 2024, doi: 10.1016/j.matdes.2023.112603.
[86] J. Hastings *et al.,* "ChEBI in 2016: Improved services and an expanding collection of metabolites," *Nucleic acids research*, vol. 44, D1, D1214-9, 2016, doi: 10.1093/nar/gkv1031.
[87] FAIRsharing Team, "FAIRsharing record for: Quantities, Units, Dimensions and Types," 2024.
[88] J. N. Otte, J. Beverley, and A. Ruttenberg, "BFO: Basic Formal Ontology," *AO*, vol. 17, no. 1, pp. 17–43, 2022, doi: 10.3233/AO-220262.
[89] *ISO/IEC 21838-2 Ed. 1: 2021 | Information technology - Top-level ontologies (TLO) - Part 2: Basic Formal Ontology (BFO),* ISO/IEC, Geneva, Switzerland, Nov. 2021. [Online]. Available: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/45/74572.html. [Accessed: Nov. 12, 2022]
[90] M. Stocker *et al.,* "Persistent Identification of Instruments," *Data Science Journal*, vol. 19, 2020, doi: 10.5334/dsj-2020-018.
[91] S. Colucci, F. M. Donini, and E. Di Sciascio, "A review of reasoning characteristics of RDF -based Semantic Web systems," *WIREs Data Min & Knowl*, vol. 14, no. 4, 2024, doi: 10.1002/widm.1537.
[92] Apache Software Foundation, *Apache Arrow: The universal columnar format and multi-language toolbox for fast data interchange and in-memory analytics.* [Online]. Available: https://arrow.apache.org/

[93] A. P. Dobos, “An Improved Coefficient Calculator for the California Energy Commission 6 Parameter Photovoltaic Module Model,” *Journal of Solar Energy Engineering*, vol. 134, no. 2, 2012, doi: 10.1115/1.4005759.
[94] C. Ballif, F.-J. Haug, M. Boccard, P. J. Verlinden, and G. Hahn, “Status and perspectives of crystalline silicon photovoltaics in research and industry,” *Nat Rev Mater*, vol. 7, no. 8, pp. 597–616, 2022, doi: 10.1038/s41578-022-00423-2.
[95] B. Vicari Stefani *et al.,* “Historical market projections and the future of silicon solar cells,” *Joule*, vol. 7, no. 12, pp. 2684–2699, 2023, doi: 10.1016/j.joule.2023.11.006.
[96] M. Grieves and J. Vickers, “Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems,” in *Transdisciplinary Perspectives on Complex Systems*, F.-J. Kahlen, S. Flumerfelt, and A. Alves, Eds., Cham: Springer International Publishing, 2017, pp. 85–113.
[97] TrustPV Project, *Building Information Model (BIM) requirements and design for the operational phase.* [Online]. Available: https://trust-pv.eu/reports/building-information-model-bim-requirements-and-design-for-the-operational-phase/ (accessed: May 4 2025).
[98] *Automation systems and integration — Digital twin framework for manufacturing. Part 1: Overview and general principles*, ISO23247-1, 2021, ISO. [Online]. Available: https://www.iso.org/standard/75066.html
[99] *BIM Dictionary*. [Online]. Available: https://bimdictionary.com/terms/search (accessed: May 4 2025).
[100] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “Digital Twin in manufacturing: A categorical literature review and classification,” *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, 2018, doi: 10.1016/j.ifacol.2018.08.474.
[101] X. Liu *et al.,* “A systematic review of digital twin about physical entities, virtual models, twin data, and applications,” *Advanced Engineering Informatics*, vol. 55, p. 101876, 2023, doi: 10.1016/j.aei.2023.101876.
[102] L. Massel, N. Shchukin, and A. Cybikov, “Digital twin development of a solar power plant,” *E3S Web Conf.*, vol. 289, p. 3002, 2021, doi: 10.1051/e3sconf/202128903002.
[103] *Orange Button - Empowering the Future of Clean Energy.* [Online]. Available: https://myorangebutton.com/ (accessed: May 4 2025).
[104] *Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) - Information management using building information modelling - Part 1: concepts and principles.*, ISO19650-1, 2018, ISO. [Online]. Available: https://www.iso.org/standard/68078.html
[105] F. Tao and M. Zhang, “Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing,” *IEEE Access*, vol. 5, pp. 20418–20427, 2017, doi: 10.1109/ACCESS.2017.2756069.
[106] *Automation systems and integration — Digital twin framework for manufacturing. Part 4: Information exchange.*, ISO23247-4, 2021, ISO. [Online]. Available: https://www.iso.org/standard/78745.html
[107] R. Bommasani *et al.,* “On the Opportunities and Risks of Foundation Models,” 2021.
[108] T. Yalçin, P. Paradell Solà, P. Stefanidou-Voziki, J. L. Domínguez-García, and T. Demirdelen, “Exploiting Digitalization of Solar PV Plants Using Machine Learning: Digital Twin Concept for Operation,” *Energies*, vol. 16, no. 13, p. 5044, 2023, doi: 10.3390/en16135044.
[109] J. Liu, X. Lu, Y. Zhou, J. Cui, S. Wang, and Z. Zhao, “Design of Photovoltaic Power Station Intelligent Operation and Maintenance System Based on Digital Twin,” in *2021 6th International Conference on Robotics and Automation Engineering (ICRAE)*, Guangzhou, China, 2021, pp. 206–211.
[110] G. Zhang and X. Wang, “Digital Twin Modeling for Photovoltaic Panels Based on Hybrid Neural Network,” in *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)*, Beijing, China, 2021, pp. 90–93.
[111] B. G. Pierce et al., *Graph Foundation Models: Code and demos for contructing Data-Driven Digital Twins of Photovoltaic & Advanced Manufacturing systems.* [Online]. Available: https://github.com/cwru-sdle/graph-foundationmodels. [Accessed: Sep. 29, 2024]
[112] A. M. Karimi, Y. Wu, M. Koyuturk, and R. H. French, “Spatiotemporal Graph Neural Network for Performance Prediction of Photovoltaic Power Systems,” *AAAI*, vol. 35, no. 17, pp. 15323–15330, 2021, doi: 10.1609/aaai.v35i17.17799.
[113] Y. Fan *et al.,* “Spatio-Temporal Denoising Graph Autoencoders with Data Augmentation for Photovoltaic Data Imputation,” *Proc. ACM Manag. Data*, vol. 1, no. 1, pp. 1–19, 2023, doi: 10.1145/3588730.

[114] Y. Fan, R. Wieser, L. Bruckman, R. French, and Y. Wu, "Parallel-friendly Spatio-Temporal Graph Learning for Photovoltaic Degradation Analysis at Scale," 2024.

[115] Y. Fan, R. Wieser, X. Yu, Y. Wu, L. S. Bruckman, and R. H. French, "Using spatio-temporal graph neural networks to estimate fleet-wide photovoltaic performance degradation patterns," *PLoS ONE*, vol. 19, no. 2, e0297445, 2024, doi: 10.1371/journal.pone.0297445.

[116] M. Li *et al.,* "Data Driven Digital Twin," *CWRU SDLE Research Center*, doi: 10.17605/OSF.IO/DVRC7.

[117] W. R. Young, "Solar on schools designed for emergency shelters: 39th IEEE photovoltaic specialist conference," in *2013 IEEE 39th Photovoltaic Specialists Conference (PVSC)*, Tampa, FL, USA, 2013, pp. 1521–1525.

[118] N. G. Dhere and S. Schleith, "Reliability of hybrid photovoltaic DC micro-grid systems for emergency shelters and other applications," in *Reliability of Photovoltaic Cells, Modules, Components, and Systems VII*, San Diego, California, United States, 2014, 91790F.

[119] W. C. Oltjen *et al.,* "FAIRification, Quality Assessment, and Missingness Pattern Discovery for Spatiotemporal Photovoltaic Data," in *2022 IEEE 49th Photovoltaics Specialists Conference (PVSC)*, Philadelphia, PA, USA, 2022, pp. 796–801.

[120] K. Lynn, J. Szaro, W. Wilson, and M. Healey, "A Review of PV System Performance and Life-Cycle Costs for the SunSmart Schools Program," in *Solar Energy*, Denver, Colorado, USA, 2006, pp. 153–156.

[121] K. O. Davis, S. R. Kurtz, D. C. Jordan, J. H. Wohlgemuth, and N. Sorloaica-Hickman, "Multi-pronged analysis of degradation rates of photovoltaic modules and arrays deployed in Florida," *Progress in Photovoltaics*, vol. 21, no. 4, pp. 702–712, 2013, doi: 10.1002/pip.2154.

[122] E. J. Schneller and K. O. Davis, "Outdoor Field Testing," in *Durability and Reliability of Polymers and Other Materials in Photovoltaic Modules*: Elsevier, 2019, pp. 279–295.

[123] Y. Lee, "A Geometric Perspective on Autoencoders," 2023.

[124] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013, doi: 10.1109/TPAMI.2013.50.

[125] S. Lindig *et al.,* "Towards the development of an optimized Decision Support System for the PV industry: A comprehensive statistical and economical assessment of over 35,000 O&M tickets," *Progress in Photovoltaics*, vol. 31, no. 12, pp. 1215–1234, 2023, doi: 10.1002/pip.3637.

[126] S. Gallmetzer, S. Lindig, M. Herz, and D. Moser, "Automated Fixing Cost Estimation of Photovoltaic System Failures for the Creation of a Decision Support System," *Solar RRL*, vol. 7, no. 23, 2023, doi: 10.1002/solr.202300562.

[127] F. B. Ismail, H. Al-Faiz, H. Hasini, A. Al-Bazi, and H. A. Kazem, "A comprehensive review of the dynamic applications of the digital twin technology across diverse energy sectors," *Energy Strategy Reviews*, vol. 52, p. 101334, 2024, doi: 10.1016/j.esr.2024.101334.

[128] L. Koester, S. Lindig, A. Louwen, A. Astigarraga, G. Manzolini, and D. Moser, "Review of photovoltaic module degradation, field inspection techniques and techno-economic assessment," *Renewable & Sustainable Energy Reviews*, vol. 165, p. 112616, 2022, doi: 10.1016/j.rser.2022.112616.

[129] W. Herrmann et al., *Qualification of Photovoltaic (PV) Power Plants using Mobile Test Equipment.* [Online]. Available: https://iea-pvps.org/wp-content/uploads/2021/04/IEA-PVPS-T13-24_2021_Qualification-of-PV-Power-Plants_report.pdf

[130] T. Trupke, O. Kunz, J. W. Weber, H. Gottlieb, and A. Slade, "Large Scale Daylight Photoluminescence Imaging of Photovoltaic Systems," in *2024 IEEE 52nd Photovoltaic Specialist Conference (PVSC)*, Seattle, WA, USA, 2024, p. 467.

[131] L. Koester, A. Louwen, S. Lindig, G. Manzolini, and D. Moser, "Large-Scale Daylight Photoluminescence: Automated Photovoltaic Module Operating Point Detection and Performance Loss Assessment by Quantitative Signal Analysis," *Solar RRL*, vol. 8, no. 1, 2024, doi: 10.1002/solr.202300676.

[132] D. D. Angelova, D. C. Fernández, M. C. Godoy, J. A. Á. Moreno, and J. F. G. González, "A Review on Digital Twins and Its Application in the Modeling of Photovoltaic Installations," *Energies*, vol. 17, no. 5, p. 1227, 2024, doi: 10.3390/en17051227.

[133] I. T. Horvath, A. A. Moya, F. Segata, J. Lemmens, T. Hall, W. Vanheusden, *Automated PV digital twin-based yield simulation framework.* [Online]. Available: https://trust-pv.eu/reports/automated-pv-digital-twin-based-yield-simulation-framework/

[134] G. M. Tina, C. Ventura, S. Ferlito, and S. de Vito, "A State-of-Art-Review on Machine-Learning Based Methods for PV," *Applied Sciences*, vol. 11, no. 16, p. 7550, 2021, doi: 10.3390/app11167550.

[135] C. Schill *et al.,* "Soiling Losses – Impact on the Performance of Photovoltaic Power Plants," International Energy Agency Photovoltaic Power Systems Programme (IEA PVPS) IEA-PVPS T13-21:2022, 2022. [Online]. Available: https://iea-pvps.org/research-tasks/performance-operation-and-reliability-of-photovoltaic-systems/

[136] SolarPower Europe, *Operation & Maintenance Best Practice Guidelines/Version 6.0.* [Online]. Available: https://www.solarpowereurope.org/insights/thematic-reports/operation-and-maintenance-best-practice-guidelines-version-6-0

[137] S. Rapaport.,M. Green, C. Ulbrich, P. Graniero, A. Louwen, U. Jahn, *The Use of Advanced Algorithms in PV Failure Monitoring*.

[138] R. Schlatmann, D. Moser, D. Muñoz Cervantes, D., I. Gordon, T. Garabetian, H. Dittmar, *Strategic Research and Innovation Agenda on Photovoltaics. European Technology and Innovation Platform*, 2024.

[139] G. M. El-Banby, N. M. Moawad, B. A. Abouzalm, W. F. Abouzaid, and E. A. Ramadan, "Photovoltaic system fault detection techniques: a review," *Neural Comput & Applic*, vol. 35, no. 35, pp. 24829–24842, 2023, doi: 10.1007/s00521-023-09041-7.

[140] M. Aghaei, *Fault detection for PV systems using machine learning techniques.* [Online]. Available: https://www.pearlpv-cost.eu/wp-content/uploads/2020/05/Aghei-Fault-Detection-for-Photovoltaic-Systems-using-Machine-Learning-Techniques_07072021.pdf

[141] L. B. Bosman, W. D. Leon-Salas, W. Hutzel, and E. A. Soto, "PV System Predictive Maintenance: Challenges, Current Approaches, and Opportunities," *Energies*, vol. 13, no. 6, p. 1398, 2020, doi: 10.3390/en13061398.

[142] M. Herz, G. Friesen, U. Jahn, M. Köntges, S. Lindig, D. Moser, *Quantification of Technical Risks in PV Power Systems.* [Online]. Available: ISBN 978-3-907281-11-6

[143] DNV, "Predictive maintenance of solar pv plants: the time is now,"

[144] P. S. Lakshmi, S. Sivagamasundari, and M. S. Rayudu, "IoT based solar panel fault and maintenance detection using decision tree with light gradient boosting," *Measurement: Sensors*, vol. 27, p. 100726, 2023, doi: 10.1016/j.measen.2023.100726.

[145] A. Alcañiz Moya, "From Waves to Shadows: PV systems yield modeling within H2020 Trust-PV project," Delft University of Technology, 2025. [Online]. Available: https://research.tudelft.nl/en/publications/from-waves-to-shadows-pv-systems-yield-modeling-within-h2020-trus

[146] "Digitalising the energy system - EU action plan: COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS,"

[147] European Commission, *Press release: Shaping Europe's digital future: Commission presents strategies for data and Artificial Intelligence.* [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_273

[148] D. Moser, "Let's Flex Data and AI to Maximise Power Plant Efficiency and System Integration," in 2025.

[149] P-J. Alet *et al.,* "Mapping the Relevance of Digitalisation for Photovoltaics," in *Proceedings of the EU PVSEC 2023*, 2023, 020366-001 - 020366-025.

[150] M. Topic, R. Drozdowski, W. Sinke, G. Arrowsmith, and A. Spoden, "Strategic Research and Innovation Agenda on Photovoltaics," European Technology and Innovation Platform for Photovoltaics (ETIP PV), 2023. [Online]. Available: https://www.etip-pv.eu/publications/strategic-research-and-innovation-agenda-on-photovoltaics

[151] U. Jahn, D. Moser, D. Muñoz, and P. Sánchez-Friera, "Driving the Quest for Reliable and Bankable PV in Europe - Status and Targets in 2030," in *EUPVSEC 2024*, 2024.

[152] B. G. Pierce and others, *Graph Foundation Models: Code and demos for constructing Data-Driven Digital Twins of Photovoltaic & Advanced Manufacturing systems.* [Online]. Available: https://github.com/cwru-sdle/graphfoundationmodels

[153] R. Wieser *et al.,* “PVplr-python: Python Package Implementation of PVplr for Performance Loss Rate Analysis,” in *2024 IEEE 52nd Photovoltaic Specialist Conference (PVSC)*, Seattle, WA, USA, 2024, pp. 1325–1327.

[154] Y. Fan, X. Yu, R. Wieser, Y. Wu, and R. H. French, *PVplr-stGNN: Python.* [Online]. Available: https://pypi.org/project/PVplr-stGNN/. [Accessed: Jan. 30, 2023]

[155] A. Curran et al., *PVplr: R.* [Online]. Available: https://CRAN.R-project.org/package=PVplr. [Accessed: Mar. 14, 2023]

Technology Collaboration Programme

by iea