VIPERLAB

FULLY CONNECTED **VI**RTUAL AND **P**HYSICAL
P**ER**OVSKITE PHOTOVOLTAICS **LAB**

D 3.5
VIPERLAB Ontology

DELIVERABLE
REPORT

Version: 1
Date: 30. 05. 2022

VIPERLAB

FULLY CONNECTED VIRTUAL AND PHYSICAL PEROVSKITE PHOTOVOLTAICS LAB
VIPERLAB

DELIVERABLE

D 3.5 VIPERLAB ONTOLOGY

Project References

Project Acronym	VIPERLAB
Project Title	Fully connected **vi**rtual and physical **per**ovskite photovoltaics **lab**
Project Coordinator	Helmholtz-Zentrum Berlin
Project Start and Duration	1st June 2021, 42 months

Deliverable References

Deliverable No	D 3.5
Type	Report
Dissemination level	Confidential, only for members of the consortium (including the Commission Services)
Work Package	WP3
Lead beneficiary	Forschungszentrum Julich gmbh (FZJ)
Due date of deliverable	31 May 2022
Actual submission date	30 May 2022

Document history

Version	Status	Date	Beneficiary	Author
1.0	Final Draft	30-05-2022	FZJ	M. Sytnyk
1.1	Reviewed	30-05-2022	FZJ	J. Hauch
1.2	Reviewed	30-05-2022	HZB	N. Maticiuc

DISCLAIMER

'Fully connected virtual and physical perovskite photovoltaics lab' VIPERLAB is a Collaborative Project funded by the European Commission under Horizon 2020. Contract: 101006715, Start date of Contract: 01/06/2021; Duration: 42 months.

TABLE OF CONTENT

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N°101006715

1. EXECUTIVE SUMMARY

Impacts of lead halide perovskites (PSK), growing interest in PSK photovoltaics (PV), and waning research interest in conventional silicon PV provide numerous technological avenues for advancing PSK solar cells. However, these PSK technologies generate a tremendous amount of information that often is in proprietary data structures which overwhelms interested researchers and thus makes it difficult for them to reuse this data in a sensible way that fits their needs. The emergence of Semantic Web technologies, and with it the subject of ontologies promis to significantly enhance knowledge representation by facilitating the sharing and reusing of data for efficient data-driven PSK research and easy data integration with machine learning (ML) algorithms and artificial intelligence (AI) software. An ontology is a generic data model representing knowledge about a subject as a map of interconnected concepts or definitions and their relationships.

This report investigates how an ontology can represent information and knowledge about PSK PV energy technologies and facilitate system decision-making by recommending the most suitable options for researchers in various situations. As part of the VIPERLAB initiative, a prototype of the VIPERLAB Ontology system for managing knowledge about PSK PV systems is under development to demonstrate specific Semantic Web capabilities.

2. INTRODUCTION

In recent years, the development and use of ontologies have become prominent in the scientific data infrastructure. Ontologies provide the basis for formal specification of the concepts and relations that characterize a particular domain of knowledge, and their formal structure provides knowledge representation for machine processing. By definition, an ontology is a specification of conceptualizations, a formalized structure for designing and using various information resources and written in a particular Web Ontology Language (OWL) terminology. Among the possible applications of ontologies are: standardization of terms, eliminating their ambiguity and misinterpretation, linking documents of related topics with Semantic Web methods, database design, and integration with ML and AI. Awareness of the enormous potential inherent in ontologies for the systematization and dissemination of scientific knowledge has stimulated the active development of subject-oriented ontologies for various domain-specific science topics like genes, proteins, batteries, etc. The next generation of the Web, commonly referred to as the Semantic Web depends on the creation and application of ontologies. The objective of the Semantic Web is for machines to "understand" and analyze information, offer structure to the meaningful content of Web data, and create an environment where software agents surfing from page to page may do complex activities for researchers with ease. Thus, an ontology as a part of the Semantic Web is a foundation for the next level of data interoperability, reuse by machines, and common FAIR data principle.

3. SEMANTIC WEB AND ONTOLOGY FOUNDATION

The Semantic Web enables computers to read word-based data and execute logic in a manner similar to how humans interact with traditional word-based documents. The Semantic Web supplements subject-oriented data with metadata and tags that indicate its meaning and relationship to other data subjects. The essential components of the Semantic Web are the Resource Description Framework (RDF) and the Web Ontology Language (OWL).

3.1 Resource Description Framework

RDF offers the fundamental relational language layer for data representation on the Semantic Web. RDF is used to create object-descriptive statements with attribute/value pairs. It introduces standardization to descriptions and more intricate semantic links between things in a domain. Formally expressing the relationship between two resources, RDF uses a three-position statement called Triple: a subject, a predicate, and an object. In the instance of PSK PV, a Triple could consist of the simple statement a "perovskite photovoltaic device" (subject) "is a" (predicate) "solar cell"(object). Due to the directionality of the "is a" relationship, it is understood that all perovskite PV devices are solar cells, but yet not all solar cells are perovskite PV devices or other perovskite-type devices. The building blocks of the Semantic Web require RDF triples to be constructed for each subject. The relation between RDF for each subject forms a Linked Data. Thus, in general, Linked Data is structured data that adheres to a set of design guidelines for exchanging linked data on the Web. Linked data consists of both real data about particular entities and semantic metadata specifying the classes of objects, relationship types and characteristics required to explain the data. Ontology is the aggregate term for this grouping of semantic metadata.

3.2 Web Ontology Language

The concept of the Semantic Web also includes strict Web standards to address the need to share large amounts of information that can be interpreted not only by humans but also directly by computers with clear description logic. Hence, to integrate ontologies into the Semantic Web, the World Wide Web Consortium has established a specific standard called the Web Ontology Language (OWL). OWL is a computational logic-based language in which computer programs may leverage information stated in OWL to determine classes, properties, and associated limitations. OWL is based on the RDF notion of Triples, but its vocabulary is more comprehensive and flexible for expressing meanings, classification hierarchies, and basic restrictions. Those factors allow OWL to express more semantically dense machine-interpretable information.

3.3 Ontology Editors

Ontology editors that simplify the construction of ontologies have been created due to the necessity for the automatic creation of ontologies. However, most ontology editors rely directly on ontology languages and methodology, which reflects the ease or complexity of ontology development. KAON, OilEd, Ontolingua Server, OntoSaurus, WebODE, WebOnto, and Protégé are common ontology editors. After an evaluation, Protégé was selected for the VIPERLAB Ontology because it supports ontology libraries and OWL languages and has a straightforward graphical user interface. Additionally, Protégé features robust IT support and a collection of suitable plugins that may be utilized with many add-ons. In addition, Protégé is accessible as both a web application and a desktop program, and it includes an abundance of documentation and tutorials to assist new users in getting started. Sets of functional add-ons and plugins, such as Owlready2, HermiT, and EMMO-Python, are freely accessible. Python library for ontology-oriented programming with Owlready2 can load OWL 2.0 ontologies as Python objects, alter and save them, and reason using the HermiT and Pellet reasoners. Owlready2 is available on Bitbucket as open-source software licensed under the GNU Lesser General Public License. There are extensions of Owlready2, such as EMMO-Python, which integrates reasoning with FaCT++ and focuses on providing tools and features for constructing and validating EMMO-compliant ontologies, but is not limited to them.

3.4 Ontology Methodologies

Different ontology methodologies have been applied in the literature to construct Semantic Web knowledge-based systems. Specifically, these techniques are founded on two basic ideas. The first one outlines the concept, connections, relations, and other distinctions pertinent to representing a domain inside an ontology. The second specifies that object specifications take the form of representational vocabulary (classes, links, etc.) that provides meanings for the language and formal restrictions for its coherent usage. The most frequently used ontology approaches are EMMO, Uschold and King Ontology Development Method, Toronto Virtual Enterprise Method, Methontology, On-To-Knowledge, Horrocks Ontology Development Method, and Ontology Development 101. In addition, EMMO has become one of the most extensively utilized frameworks for the development of material science ontologies. Literature reveals that certain approaches handle ontology production from scratch or by using existing ontologies and that the usefulness of each methodology relies on the ontology's goal. In addition, a number of the approaches were created based on their compatibility with the ontology languages and editors that would be used to generate the requisite ontology. As was already mentioned, the Elementary Multiperspective Material Ontology (EMMO) is an example. This methodological guideline is simply applied using the Protégé editor and the OWL language paradigm. In general, EMMO is a standard framework for depicting applied material sciences. The materials science community has initiated the development of a technique for knowledge capture that adheres to the fundamental principles of materials science. The VIPERLAB Ontology will thus employ the EMMO architecture.

4. VIPERLAB ONTOLOGY CONCEPT

An ontology segregates a knowledge domain into subdomains and classes with objectives that have properties and relations between them. For example, the renewable energy domain includes the Photovoltaic domain, with PSK solar cells, CdTe solar cells as further subdomains or subclasses. Then, the PSK solar cells subclass can have CVD PSK solar cells, solution-processed solar cells, or lead-free PSK solar cells. Classes can represent groups of individual objectives, other classes, or both**.** These classes can be coupled using "is a" / "same as" / "different from" relations to form a hierarchy (A_lead_free_PSK_solar_cell "is a" PSK_solar_cell "is a" Solar_cell).

Such semantic hierarchy represents a taxonomy, which organizes classes into a logical order utilizing particular word relationships. Taxonomies are the foundation of ontologies, but ontologies go beyond basic "is a"/" same as"/" different from" linkers assertions to specify additional relations, limitations, and axioms that allow for a deeper system description. Ontology structure units are highlighted in Table 1.

Table 1. Ontology Structure

Ontology components	**Explication**
Individuals	Distinct basic entities in an ontology. Examples are a specific object, a specific model version, a specific dataset, etc.
Classes	The collection of individual objects that belong to the class. One can also think about objects as instantiations of a class, for example, a specific person

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°101006715

	is an instantiation of the class Person, the specific model is an instance of the class Model, etc.
Relations/ Taxonomy	Specifications of how classes and objects are related to each other. A significant relation is the "is a" or "is Subclass of" connection, which is used to provide a classification of classes into a hierarchy of subclasses.
Restrictions	A way to define a class by restricting which objects does not belong to the class. Restrictions are often expressed as a relation combined with an existential, universal, or cardinality requirement.
Annotations	Additional content to the entities in the ontology, without being a part of the logical framework itself. Annotations are significant for making the ontology human-understandable.
Axioms	Logical propositions specify relationships between objects and classes.

Complex ontology scientific fields, such as photovoltaics, consist of many connected subdomains, such as semiconductors, materials, and characterization methods. On the other hand, the PV ontology also requires classes that define essential physics and chemical concepts. This kind of complexity necessitates the development of a hierarchical collection of ontologies that derive from the same source of fundamental knowledge and comply with a set of governing rules. Typically, such a hierarchy consists of four levels: the top level, the middle level, the domain level, and the application level(Figure 1). The standard definitions and theoretical foundations shared by all ontologies in the hierarchy are included in top-level ontologies. The top-level ontology is relatively short and consists of only the most basic units and clearly specified logical theories. For instance, SI unit definitions would be included in a top-level ontology. Middle-level ontologies supplement the top-level with ideas common to several domains. For instance, several domains might express the concept of fabrication. In this scenario, the mid-level ontology may specify the relationship between fabrication and properties, generally used physical quantities, units, etc. The domain level defines a specific discipline or field. Domain ontologies offer a common language within the domain while being broad enough to be used by any relevant applications within the domain. A domain ontology may define a domain-specific material class, such as perovskite, but a middle-level ontology may represent the generic idea of fabrication. As with the top- and middle-level ontologies, these ontologies describe classes but rarely objects. The lowest level is occupied by application ontologies, which describe application-specific ideas and objects of minimal general significance. For instance, the application ontology may define the AMANDA equipment used to perform a high throughput perovskite screening for solar cells. There may be subhierarchies of ontologies dependent on one another within each of these broad categories.

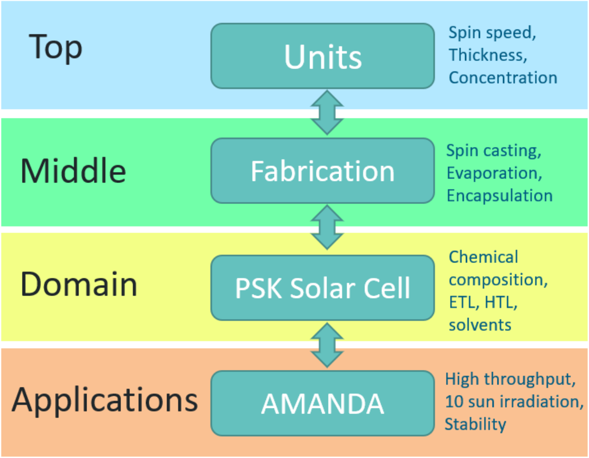

Figure 1. Possible hierarchical collection of ontologies for the VIPERLAB Ontology. There are relatively few generally applicable foundational ideas at the highest level, such as physical units(e.g., spin speed, thickness, concentration). Middle-level ontologies stem from the top-level and include standard views across several domains, such as fabrication(e.g., spin casting, evaporation, encapsulation). Domain ontologies level supplement the intermediate level with domain-specific concepts, such as solar cell materials (e.g., perovskite chemical composition, ETL, HTL, solvents). Finally, the application ontology is the ultimate level that describes a specific use case for the ontology, such as a laboratory with specialized equipment (AMANDA capabilities like high throughput operation, 10 sun irradiation, stability testing).

EMMO's ontology at the highest level is basic and succinct, giving simply the essential philosophical basis for values. Table 2 provides a short examination of EMMO's top level units which are typically based on SI unit classification.

Table 2. SI base units

Name	**Symbol**	**Measure**	**Dimension Symbol**
second	s	time	T
meter	m	length	L
kilogram	kg	mass	M
ampere	A	electric current	I
kelvin	K	temperature	Q
mole	mol	Amount of substance	N
candela	cd	Luminous intensity	J

The mid-level of EMMO extends the top-level towards domains from several angles. There are four identified perspectives: reductionistic, physicalistic, perceptual, and holistic. The reductionist viewpoint establishes a direct, nontransitive parthood relationship. Unlike a standard parthood

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°101006715

relation, which is transitive (if x is a part of y and y is a part of z, then x is a part of z), the following equation describes direct parthood between subjects:

$$A_1,\ldots, A_{n-1}, A_n \rightarrow B \qquad (1)$$

Where A_I and B are some objects with relations "is a" / "same as" / "part of." To comprehend and construct the middle level ontology component, EMMO splits it into three conceptual "worlds": the physical world, the material fabrication world, and the modeling world. In the physical world, substances are characterized by their constituent parts, relation, and physical properties (See equation 1). The material fabrication world ontologizes the overall process of object creation. It considers a fabrication to be a process in which an "Observer" detects another "Object" using a specific perception mechanism and provides a fabrication output that is a physical quantity measurement. This may then be applied to the object as a physical attribute. In the world of modeling, models are characterized as consisting of physics equations and material relations, as well as physical variables used to estimate physical values (Figure 2).

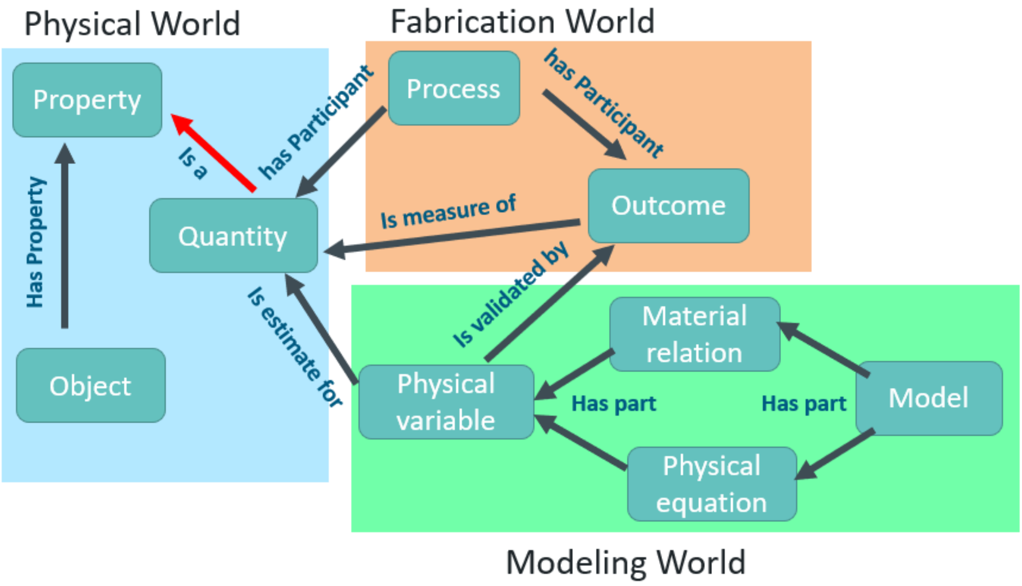

Figure 2. EMMO's simplified conceptual picture of the interdisciplinary connections between the physical, material fabrication, and modeling worlds. Relationships between terms are shown by arrows connecting their blocks from a subject to the object.

As expansions of top- and mid-level ontologies, domain ontologies are intended to function as plug-and-play repositories of information about a single topic. The domain levels should be sufficiently specific to resolve the essential knowledge and relationships to represent the subject matter while being well generic to accommodate various applications within the domain. For instance, in order to properly ontologize the design and construction of a perovskite solar cell requires knowledge of solar cells and the physics underlying them, models, characterization tools, data management, fabrication processes, and the quality data for precursor materials material logistics. Even with well-developed top- and middle-level ontologies, it would be challenging to ontologize every PSK solar cell production component from scratch. A well-managed ecosystem of existing coherent and consistent domain ontologies can overcome this difficulty. PSK solar cell domain ontologies are not an isolated activity; instead, PSK ontologies should be developed in close collaboration with relevant domain experts, taking into account the domain community's current standards, checklists, and templates.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°101006715

Characterization of experimental materials is also the target of continuing attempts to standardize data and metadata formats for the PSK solar cell domain ontology.

Figure 3 shows the basic relations for perovskite in solar cells. From the elucidation, it is known that the perovskite must contain a tree or more parts that support the formation of perovskite crystal structure, and it must have a correct bandgap. Furthermore, it must support the charge carriers extraction across the interface (i.e., via matched work functions). It can be assumed that the perovskite must remain physically intact to function. Therefore, the perovskite solar cell must contain some properties of materials that can electronically conduct and participates in an electrical current generation with the adjacent HTL and ETL. Within VIPERLAB ontology, those materials could be termed as active electronic material.

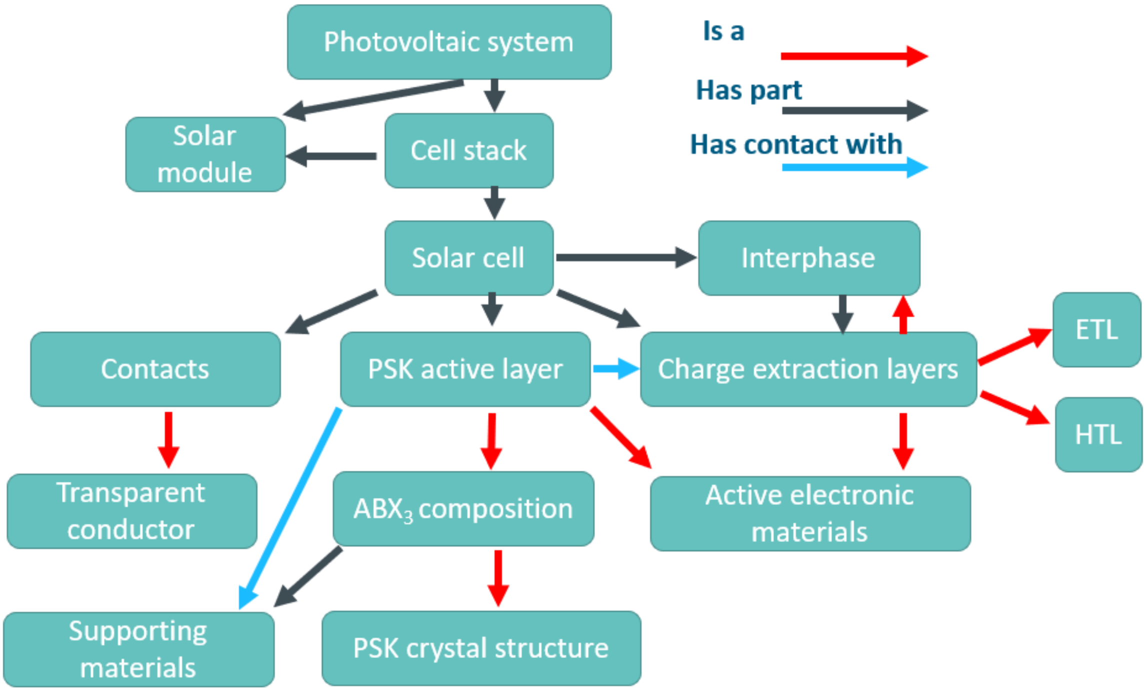

Figure 3. Overview of the fundamental ideas of a PSK solar cell domain level ontology. Arrows point logical links from the subject to the object.

Notably, the VIPERLAB consortium should plan to build each component of the PSK solar cell ontology domain (Figure 3) into an individual domain-level ontology to create a complete VIPERLAB Ontology.

Solar cell application data describes certain observable attributes of a solar cell derived from real or simulated measurements. Using an IV characterization, a researcher may get information about a particular solar cell. For example, maximum power point (MPP) is a feature of a solar cell that may be represented as a physical quantity (i.e., the amount of electric charge created in SI base units A x s). An ontology for the PSK solar cell domain should be able to characterize the measuring procedure, the physical quantities acquired from the measurement, and the relationship between these and object attributes.

EMMO representation of measurement is a semiosis procedure that yields a quantitative comparison of an object's quality to a reference standard. In semiosis, an interpreter creates a sign to represent some object. EMMO includes terminology for describing the essential components of a measurement as a process involving a measuring system and producing a measurement result. Consequently, VIPERLAB Ontology development might center on creating classes to ontologize solar cell measuring equipment and the associated physical quantities and attributes. Application data and solar cell measurements that generate solar cell-related data can come from a variety of instruments, including standard laboratory instruments such as multimeters, degradation setups, and MPP trackers, as well as equipment in NREL laboratories and large-scale open-air field testing. Ontologizing solar cell application requires cooperation with other domain ontologies since it is a very diversified field on its whole.

The VIPERLAB Ontology might first concentrate on the ontologization of measurement equipment required to enable research within the VIPERLAB project and the development of application ontologies for the current VIPERLAB infrastructures. For instance, VIPERLAB members' solar cell measuring instrument classes can specify the functions and attributes of the instruments in a measurement process, from which specific models of the apparatus can be specified for the application ontology. As a result, these devices interact with some samples and generate a measurement result inside a quantifiable measurement framework which could be included in application ontology. Figure 4 depicts the possible VIPERLAB structure of the measurement process for application ontology.

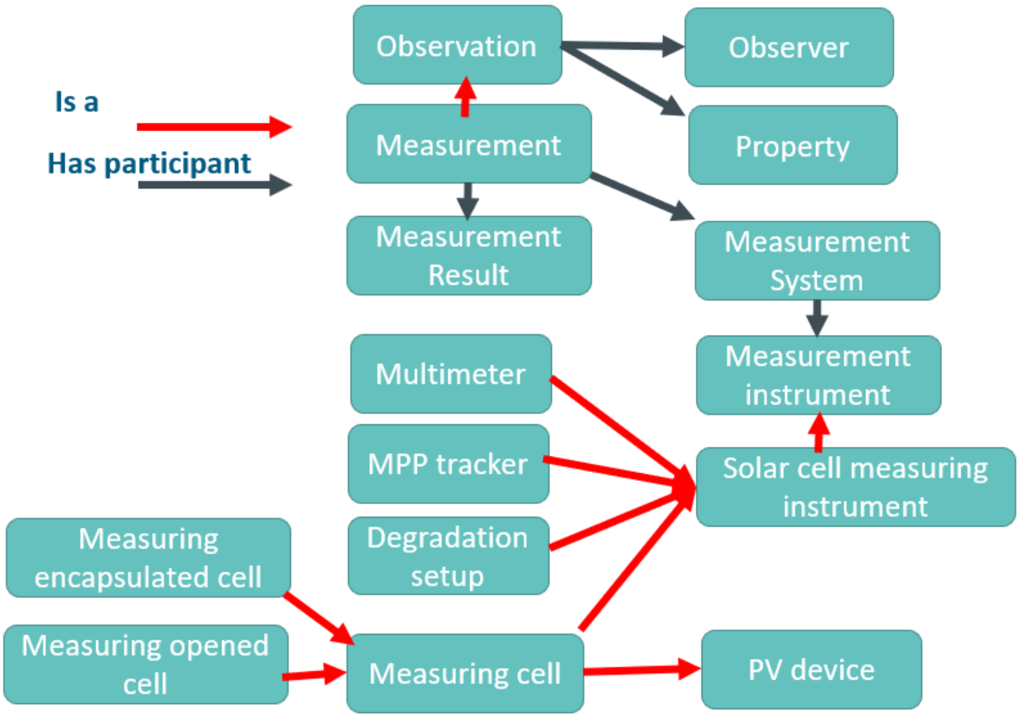

Figure 4. Overview of the proposed VIPERLAB relations for application ontology with arrows pointing From the subject to the object.

.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N°101006715

5. CONCLUSIONS

This study presents the fundamentals for representing the PSK solar cells domain of knowledge for future development of the VIPERLAB Ontology, which will serve as a platform for uniform data organization, distribution, and knowledge machine interoperability. In addition, the VIPERLAB Ontology will pave the path for the integration of VIPERLAB research data into the relevant data space. At this level, the report provides the VIPERLAB Ontology design's fundamental needs. Although this document mainly focuses on PSK solar cell ontology subsets, the EMMO framework will be employed in additional PSK solar cell knowledge domain areas. This will be included in future VIPERLAB deliverables.