A metamodel for the representation of solar plant implantation contexts for the creation of collective self-consumption loops in rural areas.

1st Yohann Chasseray
Centre Génie Industriel
IMT Mines Albi-Carmaux
Albi, France
https://orcid.org/0000-0001-5267-8508

2nd Anne-Marie Barthe-Delanoë
Centre Génie Industriel
IMT Mines Albi-Carmaux
Albi, France
https://orcid.org/0000-0002-6619-7382

3rd Aurélie Montarnal
Centre Génie Industriel
IMT Mines Albi-Carmaux
Albi, France
https://orcid.org/0000-0002-4549-7597

4th Éva Petitdemange
Centre Génie Industriel
IMT Mines Albi-Carmaux
Albi, France
https://orcid.org/0000-0003-3200-0446

***Abstract*—Self-consumption of photovoltaic electricity, which has a long payback period for small installations, is increasingly favored due to the emergence of regulations encouraging collective self-consumption. However, in rural or semi-rural areas, the viability of self-consumption is conditioned by the ability of actors to cooperate and collectively consume locally produced electricity. Moreover, the fact that the actors involved often have different profiles in terms of consumption and production adds complexity to satisfying this cooperation constraint. Since the conditions for the construction of a self-consumption loop - as well as the actors likely to be part of it - are subject to variation, this paper proposes to build a knowledge structure in the form of a business metamodel describing the contexts of (i) the implantation of photovoltaic power plants and (ii) the construction of collective self-consumption loops. This metamodel can then be used as a support for the construction of one or more knowledge graphs to provide decision support system with domain related knowledge for the construction of a collective self-consumption loop in a given context. An example of instantiation on a concrete use case is illustrated in the paper. The article also opens on perspectives for the use of such a knowledge graph with the ambition to automate the creation of collective self-consumption loops based on the consumption and production volumes of the different actors involved.**

***Index Terms*—Metamodel, Knowledge base, Knowledge modeling, Collective self-consumption**

I. INTRODUCTION

The increasing development of renewable energies and their introduction into the energy mix and the energy distribution network are forcing countries such as France to review the legal modalities of energy consumption [1]. Historically, energy distribution has been organized from and around the consumer demand. However, renewable energy sources have characteristics that make them incompatible with demand-side management. Renewable energy - and solar energy in particular - is an energy whose production level is highly unstable, both in time and in space. Moreover, this production variability depends on many factors, mainly meteorological, the long-term prediction of which is sometimes complicated and often impossible. Furthermore, the energy sources involved are often geographically dispersed. The network is then forced to consider not only the distribution of energy from a production point to different consumption points, but also the distribution of the electricity produced by these same consumption points from their location [2].

A. Collective self-consumption

In France, the financial incentives for self-consumption and the recently adopted legal mechanisms for collective self-consumption make it possible to consume the electricity that has been produced in a decentralized manner on a smaller scale [1], [2]. At the same time these measures lead the actors of the same geographical area to collaborate through collective self-consumption loops. The possible contexts in which a photovoltaic power plant with self-consumption objectives can be set up vary greatly depending on the location (geography, political context, climate), so that a project that is viable in one context will not be successful if it is reproduced elsewhere. This context of collective self-consumption motivates new research topics ranging from economic profitability studies of self-consumption loops [3] to optimizing the implementation of the latter [4].

In this line, this research proposes a knowledge structure to represent all the components involved in the set-up of a collective self-consumption loop. The generic knowledge was first formalized in the form of a metamodel. A part of

---

This work is founded by the French government through a *France Relance* project.

this metamodel was then transformed into an OWL formatted ontology to facilitate instantiation. The former ontology is the result of a collaborative work with domain experts who have acquired generic knowledge about the construction of photovoltaic projects.

B. The Ecotrain project

This research is related to the French ECOTRAIN project. The ECOTRAIN system is designed to be a system of lightweight automated shuttles powered by solar energy. One of the opportunities identified in the project is the possibility of reusing existing railways, that are no longer used by the French national railway company. The global ECOTRAIN project involves several scientific and technical fields, including logistics and simulations for the planning of the whole system in regard of the demand or Internet of Things and materials science for the design of an autonomous shuttle that is compatible with existing railways.

This paper is motivated by a subproject of ECOTRAIN, more precisely called ECOTRAIN-PV. This subproject focuses on the aspects of solar power supply for the train shuttle.

In the ECOTRAIN project, each new system is to be paired with a solar plant and storage facilities. However, the shuttle system has periods of activity that do not necessarily match the production profile of a solar plant. Since it is not economically or ecologically viable to simply sell the unused electricity to the grid and draw the needed electricity from the grid during peak hours, some self-consumption with nearby socio-economic activities is considered. This presupposes that it is possible to quickly define which local actors will be involved and in what proportions in each of the created collective self-consumption loops created along the railway. In order to help decision makers to choose which actors to involve and the conditions of their inclusion in the collective self-consumption loop for a given context, the need for a generic knowledge representation of this context was raised in the project.

II. RELATED WORK

The aim of this section is to summarize the current research work related to self-consumption and the modeling of related knowledge. The need for knowledge structures is justified, in section II-B, by the lack of automated organization of raw data and information when dealing with optimization issues for smart-grids in general and collective self-consumption in particular.

A. Economic studies

As mentioned in the introduction, the concepts related to self-consumption are quite new. In recent years several studies have focused on the economic description and profitability assessment of collective self-consumption loops. [5] and [6] both included self-consumption considerations in their respective studies on the residential battery storage systems and their impact on the electric grid. One of the findings of [5]'s work is that the incorporating residential self-consumption habits into the power grid would be in favor of better load management through the battery use, especially during the winter season. However, these studies only address the issue of residential self-consumption and do not consider collective self-consumption.

[7] studied the economic viability of net-metering and net-billing models for self-consumption in the context of Ecuador and Spain. As a result, they show that self-consumption models are not necessarily profitable for small installations, which mainly concerns residential photovoltaic systems. However, they emphasize the fact that collective self-consumption policies should contribute to the integration of small installations in larger self-consumption and underline the challenges involved, one of which is the determination and optimization of the repartition key of a self-consumption loop.

Some contextualized work dedicated to collective self-consumption can also be found in the literature. [8] and [9] propose economic models to evaluate the profitability of collective self-consumption policies in the Italian and Spanish context, respectively.

B. Self-consumption knowledge representation for optimization

The shift from a centralized electricity production model to distributed networks has focused much research on the optimization of these networks . As a result, many smart-grid optimization models can be found in the scientific literature [10]–[13].

[14] as an example, proposed a multi-objective optimization model for microgrids including several typology of consumers such as electric vehicles, storage systems, industries or residential apartments, and producers (solar PV plants and wind turbines). The multi-objective function used by the authors tends to reduce the costs of electricity production and storage and the costs implied by electricity exchanges with the grid. [15] also explores multi-objective optimization, but in the specific context of a global temperature increase, in order to evaluate the impact of this increase on the resilience of smart grids.

However, a collective self-consumption loop could be modeled as a smart-grid where all participants can be considered as prosumers. Following this idea, some studies focus on self-consumption, either in a collective aspect or for a single prosumer. [16] proposed a model for the optimization of photovoltaic installations in commercial buildings. Their goal was to find the sizing and design of photovoltaic panels that would lead to the best compromise between economic benefits and environmental impact. [17] focused on collective self-consumption, considering different actors that have their own objectives when participating in a collective self–consumption loop. The authors then modeled the optimization problem as a game theory problem, in which each actor aims to improve its own situation. [18] focused on a case study where each actor has its own seasonality in terms of consumption and developed a model to find out the implantation choices that suits the best to the contextual data of each actor (consumption profile, environment) while minimizing the implantation cost.

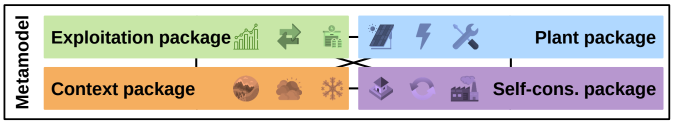

Fig. 1. Modular scheme of the proposed metamodel for solar installation contexts representation

Most of the studies presented here use domain knowledge related to either smart grids, renewable energy or self-consumption and show a strong activity concerning energy-oriented studies to propose models of self-consumption systems [19]. However, none of these studies tend to formalize and structure the mobilized knowledge. Once the mathematical model and the methodology for an optimization task are defined, the remaining work of organizing information to solve real use cases should be systematized in order to easily replicate the same methodology. Each of the study cases used in the previously cited studies presents some specificity (targeted objectives, geographical localization, size of the plant/network considered). Nevertheless, it should be possible to define generic concepts in order to systematize the structuring of the information related to each study case. For this purpose, ontologies and knowledge bases are relevant knowledge modeling tools.

Some structures dedicated to the representation of technical knowledge have been developed in the field of photovoltaic systems but they do not necessarily focus on collective self-consumption. [20] designed the PV-TONS metamodel whose aim is to support the design and sizing of photovoltaic systems. A transcription of the PV-TONS metamodel into an OWL format is also presented in [20], in order to associate reasoning rules with it. [21] proposed the OntoPowSys ontology which is a broader ontology used to describe power systems and help them interact.

[22] worked on an ontology mapping methodology, in order to align energy smart grid agents ontologies, modeling it as an optimization problem. This work highlights the fact that ontologies exist in the generic domain of smart grids.

However, to our knowledge, none of the existing knowledge structures focuses on the collective self-consumption aspects with the aim of representing contextualized self-consumption loop knowledge.

III. A METAMODEL DESCRIBING THE COLLECTIVE SELF-CONSUMPTION LOOPS OF PHOTOVOLTAIC SYSTEMS

The proposed metamodel is divided into four interrelated packages, which are shown schematically in Fig. 1. Each of these packages deals with a dimension of the implementation of photovoltaic power plants in the process of setting up one or more self-consumption loops. If all four packages are designed to be instantiable in a photovoltaic plant implantation project, their classes depict different aspects of an implantation context, meaning that all packages may be instantiated differently at different stages of a project and depending on the nature of the project.

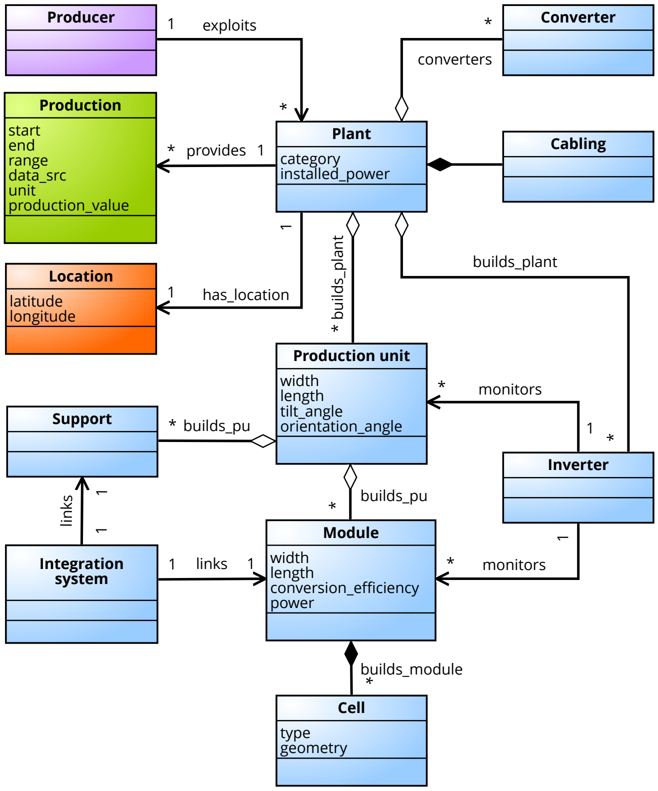

Fig. 2. UML representation of the plant package

A. The Plant package

The *Plant* package makes it possible to describe the composition of a solar power plant, by detailing in a generic way the main components qualifying a photovoltaic installation (modules, cells, etc.) as well as their technical characteristics. The fine-grained components of a solar plant are not necessarily considered in the deployment of a self-consumption loop. However, the installed power and production volumes of the whole plant are highly dependent on these components. Consequently, the installed power, which also determines the production profile of a producer, affects its possibilities to integrate a self-consumption loop. Once these dependencies are established, the plant package can be used either to design a solar plant that must meet a certain installed power (top-down approach) or to evaluate the installed power of an existing plant based on the characteristics of its components (bottom-up approach).

This package, partly based on the previous work of [20] is linked to other packages through the *Plant* class which is the main class of the *Plant* package.

The *Module* class represents a photovoltaic module which is an aggregation of photovoltaic cells (represented by the *Cell* class). The conversion efficiency attribute of a module indicates the amount of photovoltaic power (kW) that can be produced by the module relatively to a standard irradiance. This attribute can be instantiated directly from what is known about a given module or calculated from cells type

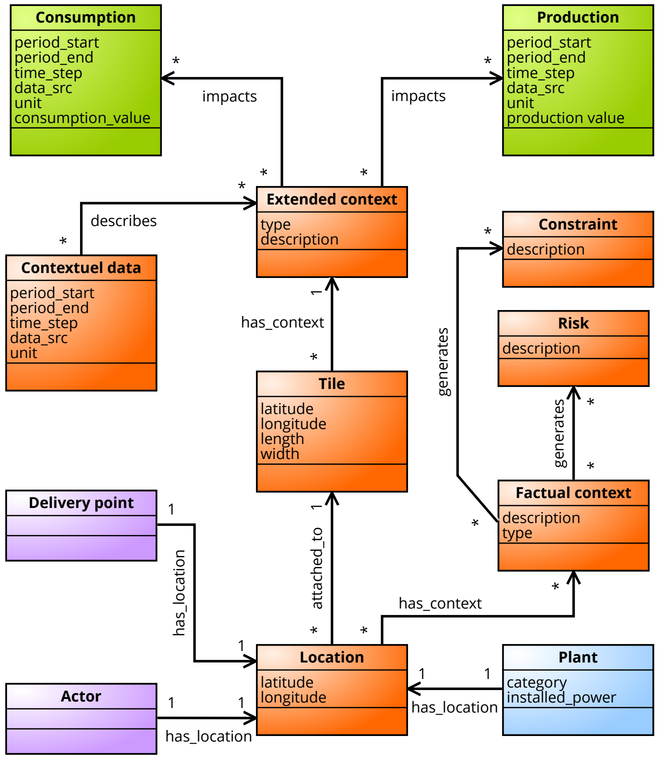

Fig. 3. UML representation of the context package

(amorphous, monocrystalline, polycrystalline) and geometry (monofacial, bifacial). The global installed power of a system can then be derived from the module efficiency. In most cases however, the installed power of a module is known for a given module, which means that the global installed power (*Plant* class attribute) can directly be deduced from the power of each module. Thus, another attribute is added to the *Module* class to directly indicate the power of the module when it is known.

Other classes of the *Plant* package (*Converter*, *Cabling*, *Inverter*, *Support* and *Integration system*) are defined in order to have a complete description of the plant components. This list of components is the result of interviews with experts in the field of solar plants installation. The proposed metamodel remains generic with respect to these components in order to avoid applications-centric restrictions. Furthermore, the definitions of cells, modules and production units is considered as sufficient enough in order to estimate installed power of a photovoltaic plant. Resulting production however do not only depends on the characteristics of the plant but will also require additional information concerning the context. Among the classes presented in the next section, some can be used to estimate production profiles of a given plant.

B. The Context package

The *Context* package defines the contextual data that can influence the production and consumption levels of the actors of a collective self-consumption loop, or even the possibilities of installing a solar power plant on a given field. In this package, information is organized around the *Tile* class, which represents a geographical area defined by its position (latitude, longitude) and its size (length, width). Contextual information is then defined either on each tile or relatively to a more precise location and is represented in the metamodel through two classes, namely *Extended Context* class and *Factual context* class. *Extended Context* class characterizes time-related contextual information for a given tile that may have an impact on either the production of a plant or the consumption of an actor. This impact exists only if the affected plant or actor is located on the tile for which the contextual information is described. With this information any decision support system that relies on the knowledge graph can adjust the production and consumption based on contextual information. Each extended context can change rapidly over time which is why the *Contextual data* class belongs to the *Context* package as well and is associated with the *Extended Context* class. A contextual data instance is a temporal numerical description of a measured characteristic of a tile (irradiance, temperature, cloud cover, etc.). The type of data that can be represented by this class is kept very large considering that the availability of descriptive data is highly dependent on the use case under study. To ensure wide coverage, contextual data class contains metadata about a possible dataset and embeds a link to the represented dataset.

Contextual information is also modeled through the *Factual Context* class. This class describes contextual facts that are not expected to change quickly over time and are more likely to be related to a single location (part of a protected area, aerial zone, public/private location). Unlike extended context, factual context is less likely to affect any consumption or production. However, it gathers situational information that could be useful in the the process of conceiving the implantation of self-consumption related plants. Factual context may indeed generate risks and/or constraints that would challenge the implantation of a plant in a given location. This is the main difference between the extended context which affects a self-consumption loop in its exploitation state, and the factual context that has an impact much earlier, in the design state of solar power plants.

C. The Exploitation package

The *Exploitation* package models the production, storage capacities and the consumption needs of each of the actors (producers or consumers) involved in a given territory and who could participate in a self-consumption loop.

The end goal of having consumption, storage and production information is to support decision making in the process of designing a collective self-consumption loop. With respect to this objective, production and consumption are necessary to approximate the amount of electricity that each collective consumption loop may require. When batteries are considered in the self-consumption loop, their storage capacity can also be used to improve the profitability of some loop configurations making possible configurations that wouldn't have been considered without storage capacity.

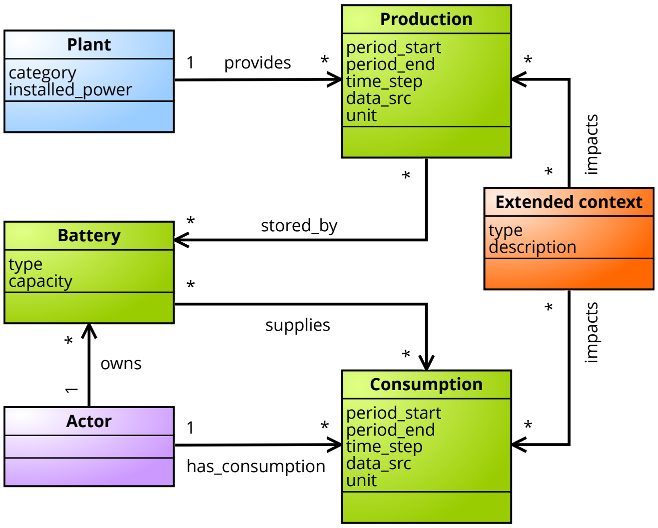

Fig. 4. UML representation of the exploitation package

When considering the exploitation phase of a solar plant in a collective self-consumption loop, some constraints may apply to the possibilities of sharing stored production between actors. To integrate these constraints in the metamodel, associations between battery instances and production/consumption instances have been semantically chosen to indicate whether or not a stored amount of energy can come from a given production source and be used to satisfy a given demand.

D. The Self-consumption package

The *Self-consumption* package describes all the actors and interactions within a collective self-consumption loop (legal entity, producers, consumers, suppliers) as well as the legal constraints associated with such a loop.

Organizations involved in a self-consumption loop, be it a household, industry or other infrastructure, are all represented by the *Actor* class. An actor can be a *Consumer* and/or a *Producer*, which are subclasses of the *Actor* class. A consumer comes with consumption data, modeled by the *Consumption* class, from the *Exploitation* package. A producer owns and exploits a solar plant that provides him with a certain amount of produced electricity and is represented through the *Plant* class, main class of the *Plant* package.

Each actor is located through the *Location* class. In some cases, it is possible to have two actors associated with the same location instance. This mainly represents the situations where a producer owns a plant that is located on a consumer's property, which is the case in the example presented in section IV.

The *LegalEntity* class is used to model the official moral entity that legally represents all members (actors) of a self-consumption loop. An actor can then be a member of the legal entity which means that he participates in self-consumption and exchanges electricity with other members of the loop, or can be excluded from the loop which means that he can't exchange electricity with other members.

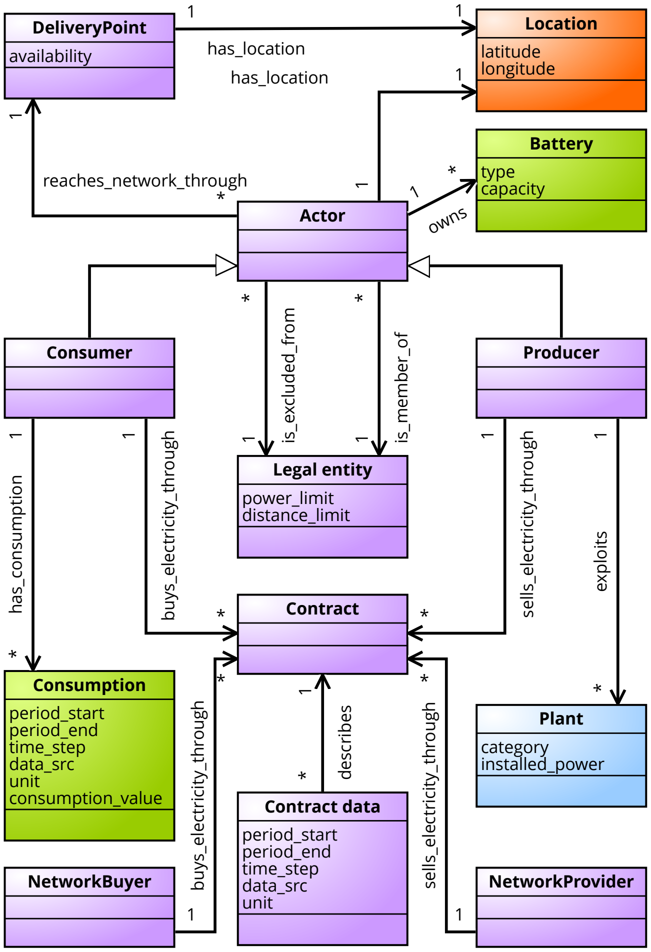

Fig. 5. UML representation of the self-consumption package

The *DeliveryPoint* class abstracts the place where an actor reaches the network. A delivery point, assigned to its location has its importance in the modelling of a self-consumption loop in the sense that any actor who wants to enter a self-consumption loop needs to reach the network at some point. Furthermore, the location of a delivery point can be used as a criteria to exclude an actor that would be too isolated from the others to reach the loop.

Each member of a self-consumption loop is connected to the grid to exchange electricity with other members. This connection also allows to sell or buy electricity from the grid that does not come from the loop. The providers and buyers of the electricity are represented by the two classes *NetworkProvider* and *NetworkBuyer*.

As mentioned above, actors can exchange electricity within the loop or buy/sell electricity from the grid. Although all physical transactions are done through the grid, there are specific contracts between actors of the same loop or between

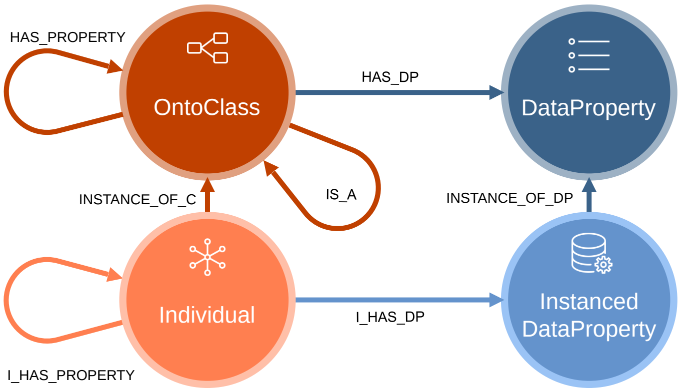

Fig. 6. Schema of the knowledge graph

actors and the grid provider/buyer. The role of the *Contract* class is then to model those contracts that have their proper selling/buying conditions.

IV. BUILDING A KNOWLEDGE GRAPH FROM A CASE STUDY

The presented metamodel has to be used in practice. A first implementation of the metamodel in a knowledge base has been done using a knowledge graph representation. The metamodel is then instantiated on a case study, resulting in an extension of the knowledge graph with instantiated nodes.

A. From UML to the knowledge graph

In the proposed knowledge graph, both the metamodel classes and the instantiated version of the classes are represented. For this purpose, we have classified the nodes of the graph according to the objects of the Web Ontology Language (OWL) [23]. In the graph, the ontological objects are represented by nodes of two categories : *OntoClass*, and *DataProperty*, as shown in figure 6 that represents the knowledge database schema. In order to keep the representation readable, object properties are represented by arcs connecting classes between them, instead of nodes, which is sometimes encountered in common OWL-based representations. The instantiated versions of the ontological objects, are represented by nodes of two other categories : *Individual* and *Instanced-DataProperty*. Individuals are linked to each other according to instanced object property arcs derived from object property arcs. Representing the two levels in the same knowledge graph is necessary because the metamodel level nodes carry semantic information about the model level nodes (meaning of a class, unit system of data properties) that is a mandatory to fully understand the graph.

Given the previous definition of the knowledge graph schema, a transformation from the UML metamodel to a knowledge graph can be performed according to the following rules :

- Each entity (or class) of the UML metamodel is represented as an *OntoClass*-labeled node in the knowledge graph.
- Each association of the UML metamodel is represented as an *OntoProperty*-labeled arc in the knowledge graph.

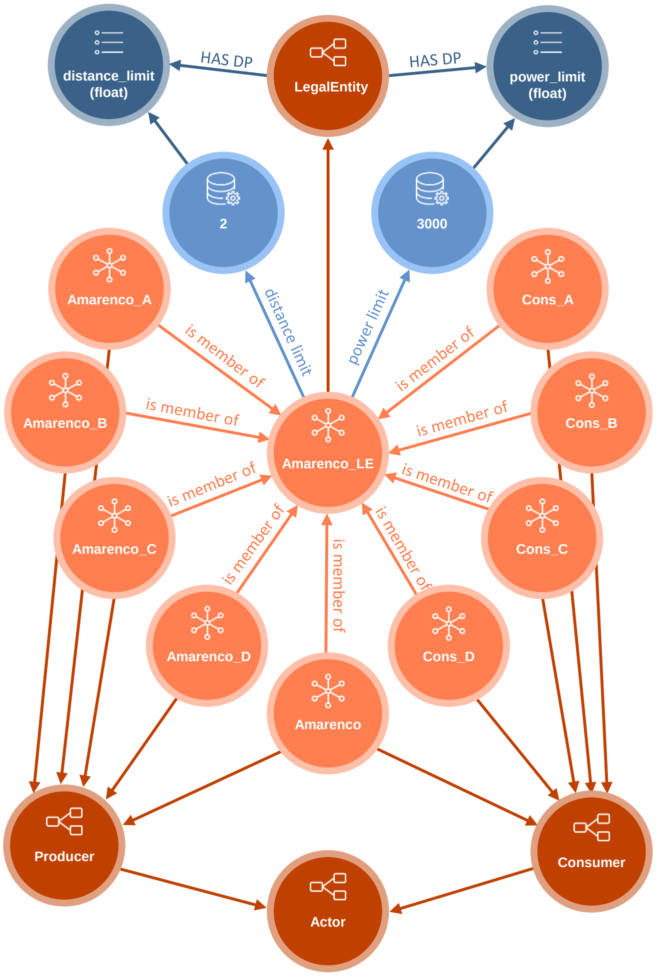

Fig. 7. Extract of the knowledge base representing all the anonymised actors of the self-consumption loop

- Each attribute expressed in the UML metamodel is represented as a *DataProperty*-labeled node in the knowledge graph.

B. Description of the actors considered in the case study

To build a model from the proposed metamodel, an instantiation is driven with 7 potential actors of a collective self-consumption loop initiated by Amarenco, which is a french solar plant company[1]. The aforementioned actors are derived from :

- 4 consumers, representing clients of Amarenco in the context of a collective self-consumption loop.
- 4 producers, representing all delocalized entities of Amarenco that produce electricity on the property of each consumer.
- 1 prosumer, representing Amarenco itself that consumes a part of the electricity produced on its property. The prosumer class is not explicitly defined in the metamodel as any prosumer instance can be defined as an instance of both *Producer* and *Consumer* classes (see Fig. 7).

[1] Associated actors will not be given for confidentiality reasons

For confidentiality reasons, the four consumers are called *Cons_A*, *Cons_B*, *Cons_C* and *Cons_D*. The four producers derived from Amarenco, who exploit plants on the property of the consumers are called *Amarenco_A*, *Amarenco_B*, *Amarenco_C* and *Amarenco_D*.

C. Representing excluded actors and members of a self-consumption loop

In this case study, as represented by the instances shown in Fig. 7, all actors are considered to be members of the self-consumption loop. In some cases, actors (consumers or producers) may be excluded from a self-consumption loop if they do not meet geographical constraints.

Regarding the acceptance of members in a self-consumption loop based on the geographical constraint, the coherence of the knowledge base can then be checked by applying the two following rules :

Property exclusivity :

$$\begin{aligned}&Actor(a) \wedge LegalEntity(le) \wedge isMemberOf(a, le) \\ &\implies \neg\ isExcludedFrom(a, le)\end{aligned} \tag{1}$$

Distance constraint :

$$\begin{aligned}&Actor(a_i) \wedge Actor(a_j) \wedge (a_i \neq a_j) \\ &\wedge LegalEntity(le) \wedge isMemberOf(a_i, le) \\ &\wedge hasLimitDistance(le, d_{lim}) \\ &\wedge Location(l_i) \wedge hasLocation(a_i, l_i) \\ &\wedge Location(l_j) \wedge hasLocation(a_j, l_j) \\ &\wedge distance(d, l_i, l_j) \wedge greaterThan(d, d_{lim}) \\ &\implies isExcludedFrom(a, le)\end{aligned} \tag{2}$$

Similarly, the inclusion of producers in a self-consumption loop should be done by verifying that the legal constraint regarding the installed power of the producers is met. This inclusion constraint states that the accumulated installed power of self-consumption members is restricted to a power limit, preventing actors to include a loop if the addition of their installed power outreaches the power limit of the loop when added to actual members' installed power. Given the following definition, the power limit inclusion constraint can be expressed through the rule (4) given below for each actor (a) of the knowledge base :

- M_{le}, the set of actors that are members of a legal entity (le)
- P_a, the set of plants exploited by an actor (a)
- IP_p, the installed power of a plant (p)
- $PL(le)$, the power limit of a given legal entity (le)

$$\begin{aligned}&\sum_{(m \in M_{le})} \sum_{(p \in P_m)} IP(p) + \sum_{(p \in P_a)} IP(p) \leq PL(le) \\ &\implies isExcludedFrom(a, le)\end{aligned} \tag{3}$$

Unlike the two previous rules, this third rule requires external processing of the knowledge base individuals before its application. Further discussion regarding external rule management is given in the discussion section (section V).

In this case study, all the plants and the location of potential members are already known, which prevents us from designing the plants and considering the factual context for the implantation of these plants. As evoked in the introduction of the case study, Amarenco exploits all the plants that are located on the properties of the consumers. However, in order to keep a reasonable complexity in the representation of the model, the instantiation will focus on two actors, which are Cons-A (consumer) and Amarenco-A (producer). Nevertheless, any information derived on the example actors can be derived with the same process for any other actor of the model.

Fig. 7 shows both the metamodel level (*OntoClass* nodes such as *Actor*, *DataProperty* nodes such as *power_limit*) and the model level (*Individuals* nodes such as *Amarenco_LE*, *InstancedDataProperty* nodes such as *2000* instanced value of the power limit data property). Metamodel level nodes are represented here to illustrate the remark made in section IV-A regarding the semantic meaning carried by those nodes (that the power limit is expressed in kW, for example). The property relations are not explicitly shown here to ensure the readability of the figure, but it should be noted that the edges that exist between two instances (*is_member_of*, for example), also exist at the metamodel level - as property labeled edges - between *LegalEntity OntoClass* and *Actor*, *Producer*, and *Consumer OntoClass* nodes.

D. Representation of consumption information

Given the presented metamodel, each consumer of a self-consumption loop can store its consumption of the past years. Depending on the case, different data source formats can be used. To preserve genericity, the knowledge contained in the knowledge graph only indicates how the data is presented (*start_date*, *end_date*, *time_step*) and where a source file can be found (*data_src*). As shown in Fig. 8, the complete consumption data of Cons-A is easily represented within a single consumption instance. However, it can happen that the same consumer has a consumption that is the addition of different data sources. This is the case when two electricity meters are used to collect consumption data, resulting in two consumption instances linked to two data sources, possibly with different repartitioning of consumption measures. In the case of the Amarenco actor, two meters are considered delivering each a partial information about the total amount of electricity consumed. However, it is still possible to process the metadata and attached files to reconstruct the complete consumption information.

As said in previous sections, the interoperability aspects of the metamodel are ensured by the fact that raw data extracted from different sources are not directly represented in the derived models. However, an external link can be attached to each instance, indicating where raw data can be found.

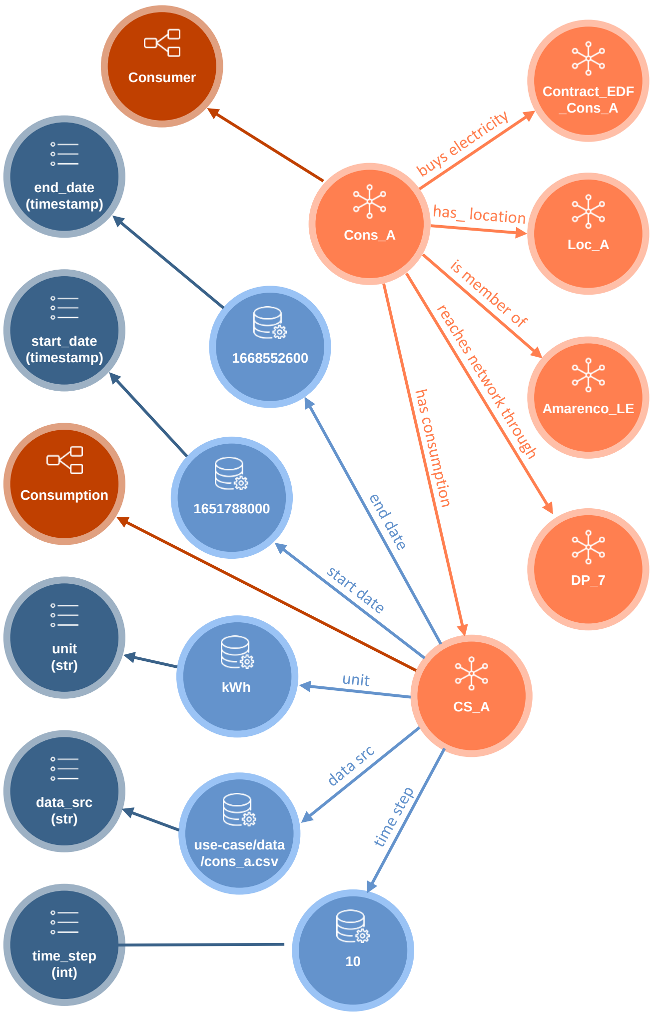

Fig. 8. Extract of the knowledge base representing the consumption of an actor (Cons-A)

Additional metadata are additional information that would help to automatically parse raw data in a decision support system.

Fig. 8 represents the consumption information of actor cons-A in the year 2022. Since the used dataset covers the part of the year going from June 5th to November 15th, with one measure taken every 30 minutes the start date, end date and time step data property of the consumption data are set to 06/05/2022 00:00, 11/15/2022 23:50 and 10, respectively. The unit in which measures are expressed (here kWh) is given as an instance of the *unit* data property.

Similar to Fig. 7, *DataProperty* nodes derived from the metamodel's attributes are shown on Fig. 8 to remind that they bring semantic knowledge that allows correct interpretation of instantiated versions of each class and data property. The information contained in *DataPorperty* nodes depicts how the stored value should be expressed and interpreted (time step is expressed in minutes, start date and end date are given as timestamps, etc.).

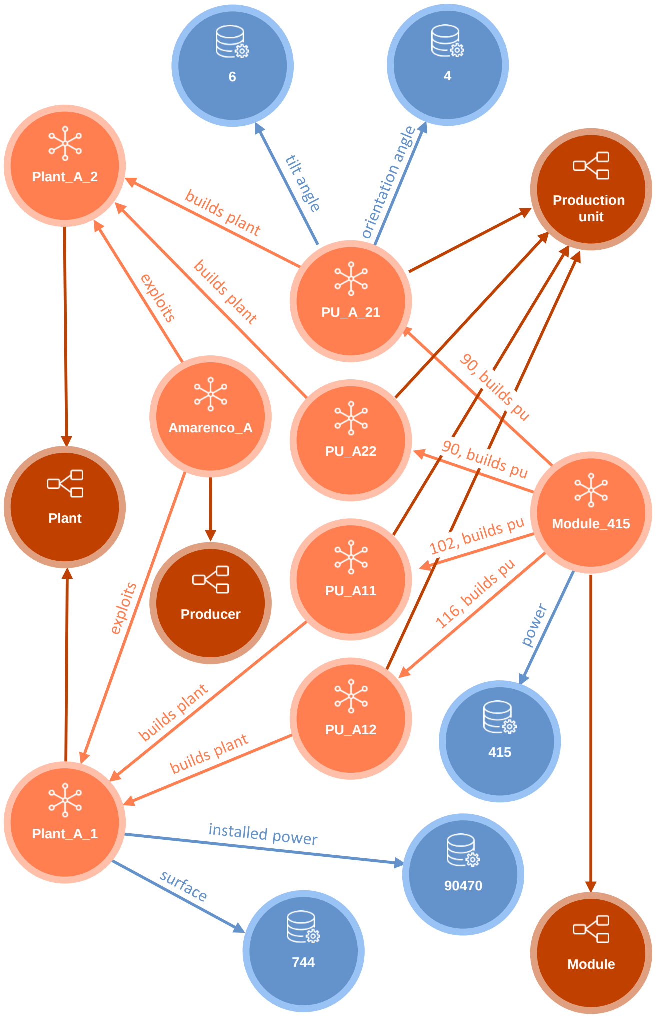

Fig. 9. Extract of the knowledge base representing the plants exploited by Prod-A actor

E. Solar plant instantiation and installed power estimation

In the self-consumption project attached to this case study, *Amarenco_A* will exploit two plants, *Plant_A_1* and *Plant_A_2*. The first one will be installed on an existing building. The other one is depends on the construction of a second building on which it will also be installed. Since the two plants will be built on a double-sided roof with the same orientation, each of them will be modeled with two different production units, with the same tilt angle (6° for the two production units of the first plant, and 10° for the two production units of the second plant) but different orientation angles with a difference of 180° (4° for the production units facing east and 184° for the production units facing west). In Fig. 9, for clarity considerations, instanced data properties are only represented for the instances *PU_A21* and *Plant_A_1*. The same data properties are however also instanced in the knowledge graph for *Plant_A_1*, *PU_A22*, *PU_A11*, *PU_A12* even though they are not represented on the figure.

Fig. 9 shows the instances that are associated with the plants that *Amarenco_A* exploits. Not all the classes are instantiated as only known information is stored in the model. For example, in this study case, the characteristics of the photovoltaic cells used remains unknown, resulting in a model of the photovoltaic plant whose granularity is limited to the module level. Nevertheless, in more advanced stages of the project, this information can be added to the knowledge base along with improved estimates of the installed power.

Each production unit of the plants are composed of several modules of the same type (*Module_415*). Instead of having a represented link for each module instance, the composition property (*builds_pu*) is provided with an additional attribute that indicates the total number of modules that is used to build up each production unit and therefore concerned by the composition property. Still, distinct module instances may be defined if different types of modules, with different power or efficiency, are used. In this case, dedicated composition relations would be derived between a production unit and each module instance that makes up the production unit.

F. Irradiance and production data

A quick estimate of a plant's production can be made from its associated installed power, considering the total amount of energy produced at a given location, during a fixed amount of time. However, this simplistic approach doesn't take into account the seasonality, both during the day and during the year of production. Current smart-grid optimization models use at least hourly production and consumption data. In particular, having accurate hourly production estimates is a necessity to provide sufficient information to decision support systems that would be responsible for designing energy exchanges between producers and consumers.

However, the actual production value depends on a set of various factors (module and cell characteristics, irradiance, temperature, etc.). If they are available in the knowledge graph, dedicated domain-based models can be used to derive hourly production data, and feed a simulation model. The automated generation of production data is discussed from a technical point of view in section V .

Fig. 10 gives an example of how irradiance data can be represented with respect to the metamodel. In this example, the irradiance is modeled as an extended context instance associated with a geographic tile. The irradiance context is associated with a contextual data instance, which depicts - as for the consumption data presented before - both data source and metadata of a irradiance data collections. This way of doing is also replicable for other type of contextual data, such as temperature, as shown in Fig. 10.

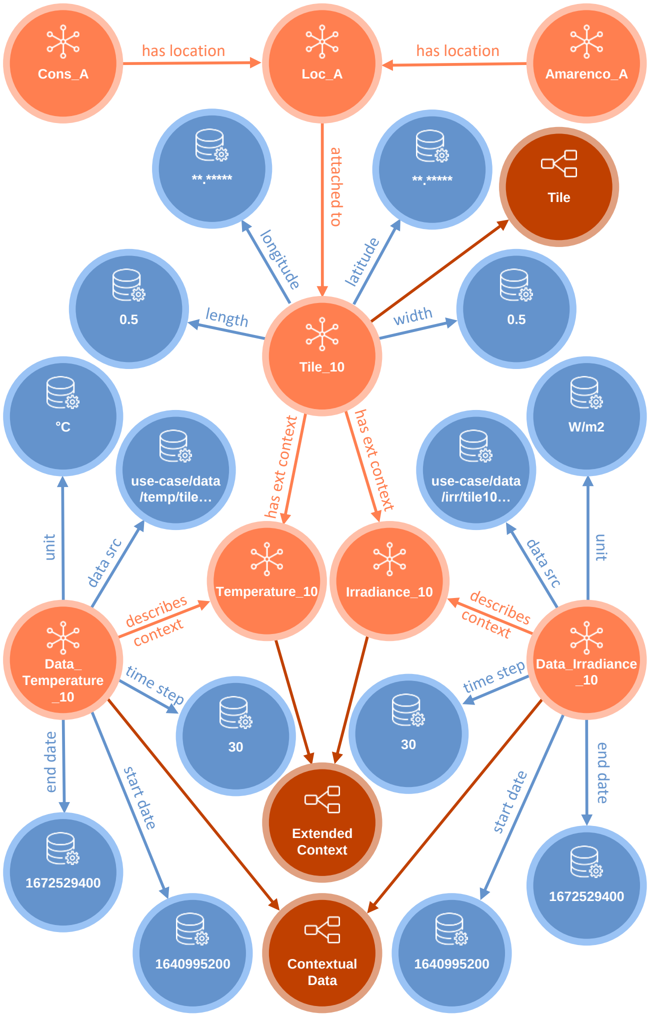

Fig. 10. Extract of the knowledge base representing the contextual information related to Cons_A/_Amarenco_A location's tile (longitude and latitude information have been anonymised)

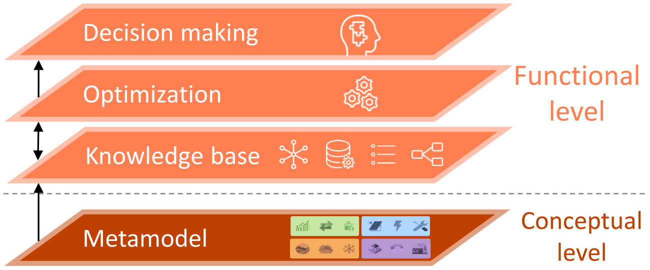

Fig. 11. Representation of the proposed knowledge-based decision support system layers

V. DISCUSSIONS AND PERSPECTIVES

In our use of the metamodel and the derived knowledge base, the installed power of the plant is instantiated directly from the use case. Nevertheless, as mentioned in section III-A, the plant package contains classes that can be used to automatically determine this installed power from fine-grained components. To do so however, the metamodel must be provided with appropriate models and rules that characterize the dependencies that exist between module and cell configurations and overall installed power.

In the early stages of a photovoltaic project, the information available to a decision maker remains incomplete. It is unrealistic to hope for a complete knowledge base that explicitly represents the situation. However, integrating the knowledge base in a more global system is also an opportunity to add tools that will, by interacting with the knowledge graph, either enrich the knowledge of the installation situation through inference, provide new sources of data or guide optimization, as illustrated in figure 11. This decision support system, should also be supported by several modules, presented in this section.

A. Data enrichment modules

To add missing information needed for self-consumption simulations or optimization, external tools can be used to enrich the knowledge base with the information that it already contains. In the case study, for instance, installed power, plant characteristics and irradiance data are detailed in the knowledge base. However, since the plants that will be used by producers have not yet been built, there is no production data available in the knowledge base.

Since these production data will be needed as input data to run simulation or optimization models, further investigation could be conducted to build an additional module dedicated to predicting of production data for each plant based on its characteristics and locally defined extended context elements.

Regarding extended context information, for which data are not always available at the instantiation step, additional modules can also be built to collect this data from external open data sources.

B. Restriction rules manager

Rules 1 and 2 presented in this paper (section IV-C) are expressed using first order logic IF-THEN statements. Nevertheless, the fact that the expression of rule (2) includes a computational step (distance between actors, float comparison) implies the need for an external tool to apply them to the knowledge graph content.

Rule 3 is based on a summation over both members of a self-consumption loop and plants of actors. This kind of rules can't be expressed through simple first order logic statements, and would then be hard to translate into a inference language such as SWRL. Defining an external rule management tool would then allow the expression of more complex rules.

C. Optimizing self-consumption

The final goal of the knowledge base is to be used in a decision support system to guide the elaboration of collective self-consumption loops. Thus one of the tools that would gain efficiency using the knowledge base are optimization tools that need contextual parameters and data, constraints criteria definition, and a set of potential members of a self-consumption loop to check that any optimized loop proposition is compliant with the rules defined on the metamodel level.

The previously listed elements are all available in an instantiated model, as represented by the knowledge base. Using the knowledge base as a source of information for the optimization algorithm makes sense when a generic optimization problem is defined. Even if the final objective of the optimization may vary from one study case to another (maximize self-consumption rates, find an egalitarian repartition of the produced energy among the consumers of an existing loop, ensure economic viability of a loop), the structure of the metamodel still allows to define basic optimization variables and constraints for a wide variety of encountered problems.

It is important to insist here on the egalitarian repartition goal. When a self-consumption loop involves several actors from different sectors (industries, households, commercial buildings, political infrastructures) - which is more likely to happen in rural areas - the diversity of consumption and production profiles is high. Defining the rules for the exchange of self-produced electricity is usually done through a discussion between the actors resulting in a common agreement.

Encoding different rules is a way to support the decision making by evaluating the impact of a given agreement on all actors through the simulation of different scenarios. The activation or deactivation of selected rules, or the prioritized inclusion/exclusion of actors in a self-consumption loop, is also a way to gain insight into how different configurations affect performance indicators for each actor.

VI. CONCLUSION AND FUTURE WORK

In this paper, a packaged metamodel has been proposed for the representation of solar power plant implantation situations in the context of collective self-consumption loop building in rural areas. The proposed metamodel is designed so that to be used in two types of use cases. The first type of use cases concerns the description of the local context in order to support decision making in the selection of a location for new potential solar power plant installations. The second type of use cases - which should be seen more often - concerns the construction of collective self-consumption loops from existing (or planned) solar plants. An instantiation example related to one of these second type use cases has been illustrated. A knowledge graph was then built from the design metamodel, including 11 actors, among which 6 producers, each of which uses at least one plant.

Since the construction of the knowledge base allows the organization of contextualized information for one or more specific use cases, it only constitutes the first milestone of a broader system. In fact, once a knowledge base referencing potential actors of collective self-consumption loops is built, a second objective is to use this knowledge base in a decision support system in order to simulate and optimize the association of actors within a collective self-consumption loop. This second objective will be the main concern of future work in the ECOTRAIN-PV project.

ACKNOWLEDGMENT

We would like to thank Amarenco who is a valuable partner in the ECOTRAIN-PV research project and provided us with fine-grained knowledge and data. Their expertise and dedication have been essential to the writing of this paper.

REFERENCES

[1] C. Sebi and A.-L. Vernay, "Community renewable energy in France: The state of development and the way forward," *Energy Policy*, vol. 147, p. 111874, Dec. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421520305905

[2] B. E. Lebrouhi, E. Schall, B. Lamrani, Y. Chaibi, and T. Kousksou, "Energy Transition in France," *Sustainability*, vol. 14, no. 10, p. 5818, Jan. 2022, number: 10 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2071-1050/14/10/5818

[3] Y. Aknin, C. Bonnefous, M. Bottausci, A. Chouraqui-Quatremain, and N. Seurat, "Method to Determine Auto-consumption Profitability," *Energy Procedia*, vol. 107, pp. 193–198, Feb. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876610216317593

[4] M. S. Simoiu, I. Fagarasan, S. Ploix, and V. Calofir, "Optimising the self-consumption and self-sufficiency: A novel approach for adequately sizing a photovoltaic plant with application to a metropolitan station," *Journal of Cleaner Production*, vol. 327, p. 129399, Dec. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652621035824

[5] H. J. J. Yu, "System contributions of residential battery systems: New perspectives on PV self-consumption," *Energy Economics*, vol. 96, p. 105151, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140988321000566

[6] F. Cucchiella, I. D'Adamo, and M. Gastaldi, "Photovoltaic energy systems with battery storage for residential areas: an economic analysis," *Journal of Cleaner Production*, vol. 131, pp. 460–474, Sep. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652616304462

[7] Ordóñez, E. Sánchez, L. Rozas, R. García, and J. Parra-Domínguez, "Net-metering and net-billing in photovoltaic self-consumption: The cases of Ecuador and Spain," *Sustainable Energy Technologies and Assessments*, vol. 53, p. 102434, Oct. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213138822004866

[8] I. D'Adamo, M. Gastaldi, and P. Morone, "Solar collective self-consumption: Economic analysis of a policy mix," *Ecological Economics*, vol. 199, p. 107480, Sep. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921800922001422

[9] J. T. Villalonga Palou, J. Serrano González, J. M. Riquelme Santos, C. Álvarez Alonso, and J. M. Roldán Fernández, "Sharing approaches in collective self-consumption systems: A techno-economic analysis of the Spanish regulatory framework," *Energy Strategy Reviews*, vol. 45, p. 101055, Jan. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211467X23000056

[10] K. Ullah, G. Hafeez, I. Khan, S. Jan, and N. Javaid, "A multi-objective energy optimization in smart grid with high penetration of renewable energy sources," *Applied Energy*, vol. 299, p. 117104, Oct. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261921005481

[11] P. A. Gbadega and Y. Sun, "Primal–dual interior-point algorithm for electricity cost minimization in a prosumer-based smart grid environment: A convex optimization approach," *Energy Reports*, vol. 8, pp. 681–695, Nov. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484722020807

[12] Y. Lin, X. Tang, and X. Fan, "Optimization of multi-energy grid for smart stadiums based on improved mixed integer linear algorithm," *Energy Reports*, vol. 8, pp. 13 417–13 424, Nov. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484722017942

[13] Y. Mohammadi, H. Shakouri G, and A. Kazemi, "A Multi-Objective Fuzzy Optimization Model for Electricity Generation and Consumption Management in a Micro Smart Grid," *Sustainable Cities and Society*, vol. 86, p. 104119, Nov. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2210670722004322

[14] A. Kumar, M. Rizwan, and U. Nangia, "A hybrid optimization technique for proficient energy management in smart grid environment," *International Journal of Hydrogen Energy*, vol. 47, no. 8, pp. 5564–5576, Jan. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360319921045754

[15] M. Mallaki, S. Najibi, M. Najafi, and N. C. Shirazi, "Smart grid resiliency improvement using a multi-objective optimization approach," *Sustainable Energy, Grids and Networks*, vol. 32, p. 100886, Dec. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352467722001473

[16] A. Allouhi, "Solar PV integration in commercial buildings for self-consumption based on life-cycle economic/environmental multi-objective optimization," *Journal of Cleaner Production*, vol. 270, p. 122375, Oct. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652620324227

[17] M. Stephant, D. Abbes, K. Hassam-Ouari, A. Labrunie, and B. Robyns, "Distributed optimization of energy profiles to improve photovoltaic self-consumption on a local energy community," *Simulation Modelling Practice and Theory*, vol. 108, p. 102242, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569190X2030174X

[18] G. P. Luz and R. A. e. Silva, "Modeling Energy Communities with Collective Photovoltaic Self-Consumption: Synergies between a Small City and a Winery in Portugal," *Energies*, vol. 14, no. 2, pp. 1–26, 2021, publisher: MDPI. [Online]. Available: https://ideas.repec.org//a/gam/jeners/v14y2021i2p323-d477280.html

[19] J. Albouys-Perrois, N. Sabouret, Y. Haradji, M. Schumann, B. Charrier, Q. Reynaud, F. Sempé, and C. Inard, "Multi-agent simulation of collective self-consumption: Impacts of storage systems and large-scale energy exchanges," *Energy and Buildings*, vol. 254, p. 111543, Jan. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778821008276

[20] F. H. Abanda, J. H. M. 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, Apr. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197612002795

[21] A. Devanand, G. Karmakar, N. Krdzavac, R. Rigo-Mariani, Y. S. Foo Eddy, I. A. Karimi, and M. Kraft, "OntoPowSys: A power system ontology for cross domain interactions in an eco industrial park," *Energy and AI*, vol. 1, p. 100008, Aug. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666546820300082

[22] X. Xue and P.-W. Tsai, "Integrating Energy Smart Grid's ontologies through multi-objective particle swarm optimization algorithm with competitive mechanism," *Sustainable Energy Technologies and Assessments*, vol. 53, p. 102442, Oct. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213138822004945

[23] S. Bechhofer, F. Van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, L. A. Stein *et al.*, "Owl web ontology language reference," *W3C recommendation*, vol. 10, no. 2, pp. 1–53, 2004.