H2020 – LC – SC3 – EE – 2019 – GA 894240

Operating System for Smart Services in Buildings

D3.1 Requirements and Analysis of Service Data

WP3 Common Ontology and Semantics

	Name	Date
Prepared by	Amir Laadhar (AAU), Christian Thomsen (AAU)	18.01.2021
Peer reviewed by	Dominique Gabioud (HES-SO), Yves Stauffer (CSEM)	08.02.2021
Reviewed and approved by	Dominique Gabioud (HES-SO)	28.02.2021

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 894240.

Distribution list

External		Internal	
European Commission	x	Consortium partners	x

Change log

Version	Date	Remark / changes
1.1	18.01.2021	First version by AAU
2.1	10.02.2021	Second version by AAU
3.1	28.02.2021	Final Version by HES-SO

To be cited as

"D3.1 Requirement and Analysis of Service Data" of the HORIZON 2020 project domOS, EC Grant Agreement No 894240.

Disclaimer

Table of contents

List of figures

List of tables

1. Executive Summary

This document is a deliverable of the domOS project, funded by the European Commission under its Horizon 2020 Research and innovation program (H2020), reporting the results of the activities carried out during task T3.1, namely the "Requirements and Analysis of service data". The main objective of WP3 is to provide a common information model, the domOS Common Ontology (dCO) to share a unified understanding of the IoT ecosystem. This will decouple the domOS infrastructure from the software services (e.g., flexibility, energy efficiency, district heating). Therefore, we review in this deliverable the existing ontologies in the scope of domOS project. Existing ontologies will be reused in the development of dCO. We report in this deliverable the input and output data existing in demonstrator sites. This data should be interoperable to interconnect IoT devices and services. We also collected the data functional requirements from the service provider, platform operator and the user. These requirements define the set of functionalities that should be deployed by demonstrator sites using data in the frame of the IoT ecosystem.

2. Introduction

2.1. Purpose of the Deliverable

An objective of the domOS project is to enable the interoperability of data and services for smart buildings. Interoperability can be enabled by using a common information model, an ontology, to share a unified understanding of the information structure among people and software agents (Noy, et al., 2001). Consequently, the main purpose of WP3 is to develop a common ontology dCO to decouple infrastructure from the actual domOS services. Developing a new ontology does not mean reinventing the wheel by building a new ontology from scratch without extending and reusing existing domain ontologies. A core best practice of the semantic web is to reuse existing ontologies when building a new ontology (Gyrard, et al., 2015). The lack of a common data representation model prevents interoperability between smart services and limits the deployment of IoT applications. IoT developers need to map heterogeneous data of different silos, buildings, and infrastructures to a common information model. It is important to unambiguously specify the semantics of data to have a common understanding of this data by different stakeholders and smart services. A common information model needs to be expressive enough to capture the contextual information for buildings, the sensors installed, and the data they generate. Therefore, in this deliverable, we review the existing ontologies related to the scope of domOS project. We also collect and analyse input and output data for each of the 5 demonstration sites of the common (WP4) and specialized (WP5-WP7) services. We present the currently known functional data requirements that define the main functionality that should be achieved using data. We also classify functional data requirements based on the different stockholders (e.g., service provider, platform operator, user).

2.2. Structure of the Document

The rest of the deliverable is structured as follows:

- **Section 3** provides an overview of the main ontologies related to the scope of domOS project.
- **Section 4** presents the set of input and output data for each demonstrator site.

- **Section 5** presents a descriptive analysis of the collected input data, output data.
- **Section 6** includes the collected functional data requirements that will be deployed by the demonstrator sites.
- **Section 7** provides our conclusion and future lines of work.

3. Relevant Existing Ontologies for domOS

In this section, we review the main relevant ontologies related to the scope of the domOS project. These will be further considered and possibly exploited when the domOS Common Ontology (dCO) is developed in T3.2.

3.1. W3C Semantic Sensor Network Ontology (SSN)

The Semantic Sensor Network (SSN) (Janowicz, et al., 2019) ontology is one of the main IoT ontologies for describing sensors and their observations, the involved procedures, the studied features of interest, the samples used to do so, and the observed properties, as well as actuators. SSN is designed in a modular manner by including a lightweight but self-contained core ontology called SOSA (Sensor, Observation, Sample, and Actuator) for its elementary classes and properties. SSN ontology is reused and aligned by multiple other ontologies used in different projects.

3.2. Smart Applications REFerence Ontology (SAREF)

The Smart Applications Reference Ontology (SAREF) (Daniele, et al., 2015) is an ontology for IoT devices and solutions published by ETSI in a series of Technical Specifications initially released in 2015 and updated in 2017. Even if its initial objective were to build a reference ontology for appliances relevant for energy efficiency, SAREF is not limited to this scope and can serve as an upper reference model to enable better integration of data from various vertical domains in the IoT. SAREF has been extended to different domains such as energy (SAREF4ENER), environment (SAREF4ENVI), buildings (SAREF4BLDG), smart cities (SAREF4CITY), agriculture (SAREF4AGRI), industry & manufacturing (SAREf4INMA). SAREF4ENER was created in collaboration with Energy@Home (http://www.energy-home.it) and EEBus (http://www.eebus.org/en), the major Italian and German industry associations, to enable the interconnection of their (different) data models. SAREF4ENER model flexibility. However, we cannot assume that this representation is rich enough to model the Flex Offer that will be extended in T3.3 to be able to capture predictions, flexibility, and uncertainty of various dCO measures. SAREF is aligned to many other ontologies such as SSN and SOSA. One of the main criticisms of SAREF is that the ontology is not rich enough to model all the information in the IoT ecosystem. For instance, in SAREF and its extensions, the representation of the various entities related to the actuator is limited (Seydoux, 2018). Therefore, knowledge scientists their IoT ontologies by reusing the multiple state-of-the-art ontologies such as SAREF, its extensions, and SSN.

3.3. oneM2M Base Ontology

The purpose of the oneM2M project (Swetina, et al., 2014) is to develop technical specifications addressing the need for a common M2M IoT service layer that can be readily embedded within various hardware and software. The oneM2M Base Ontology is a minimal ontology (i.e., mandating the least number of conventions) that is required such that other ontologies can be mapped into oneM2M.

External organizations and companies are expected to contribute their own ontologies that can be mapped to the oneM2M Base Ontology. That way oneM2M data can be supplemented with information on the meaning/purpose of the data.

3.4. W3C Web of Things (WoT)

The Web of Things (WoT) (Guinard, et al., 2016) is an extension of the IoT to ease the access to data using the benefits of Web technologies. A Thing Description (TD) describes the metadata and interfaces of things, where a Thing is an abstraction of a physical or virtual entity that provides interactions to and participates in the WoT. The Thing Description ontology is an RDF (Resource Description Framework) (Lassila, et al., 1998) axiomatization of the TD information model. It defines terms associated with the WoT architecture, and focuses purely on interaction with the devices, without considering other elements, such as physical characteristics of the sensor or its deployment. The TD ontology is not rich enough for semantic annotation of the WoT devices. Moreover, the TD ontology is not aligned with the SSN or SAREF ontologies.

3.5. Brick: Metadata Schema for Portable Smart Buildings Applications

Brick (Balaji, et al., 2018) is an ontology representing metadata in buildings, sensors, their subsystems, and relationships among them. Brick focuses on interactions among devices and building spaces as the core of IoT applications. The ontology captures hierarchies, relationships, and properties for describing building metadata and has a clear focus on commercial buildings. Brick’s design is based on more than 17,700 data points supplied by building management systems from six different vendors and have vastly varying subsystems and sensors. To showcase the effectiveness of Brick, the author compared Brick to Haystack (2020), IFC (Bazjanac, et al., 1999) and SARE using three metrics to measure the effectiveness of a schema: (i) the ability to completely map building management system metadata from three existing buildings to the schema, (ii) ability of the schema to capture the relationships required by applications, and (iii) the flexibility of the schema to deal with uncertainty as well as their extensibility to new entities. Figure 1 presents the comparison across Haystack, IFC, SAREF and Brick for metrics (i) and (ii) based on eight representative applications and two buildings. SAREF scored the lowest for both metrics because it models the common entities across different models and systems instead of comprehensively modelling buildings. Brick has the best coverage of both vocabularies and application requirements in this use case. Therefore, we cannot hypothesize that BRICK is comprehensive enough to cover all the possible use cases. Moreover, Brick does not model energy flexibility.

3.6. Discussion

The authors of the Brick ontology show that SAREF, SSN and SOSA are generic and not rich enough to cover all their requirements. Moreover, these ontologies do not provide a comprehensive modelling of energy flexibility, which will be extended in D3.3 and represented by dCO. Since every project has its specificities, knowledge scientists usually reuse existing IoT ontologies such as SSN, SOSA, and SAREF as upper-level ontologies to develop their domain ontologies. Currently, there is no single ontology that can cover all the data points related to every specific project. Knowledge scientists are usually guided by upper-level ontologies in the design of their specific ontologies. They do not only reuse ontologies directly related to IoT, but also other domain ontologies, such as the ontology of unit of measure. Moreover, they align the developed ontology to the existing ones in order to ensure interoperability between ontologies.

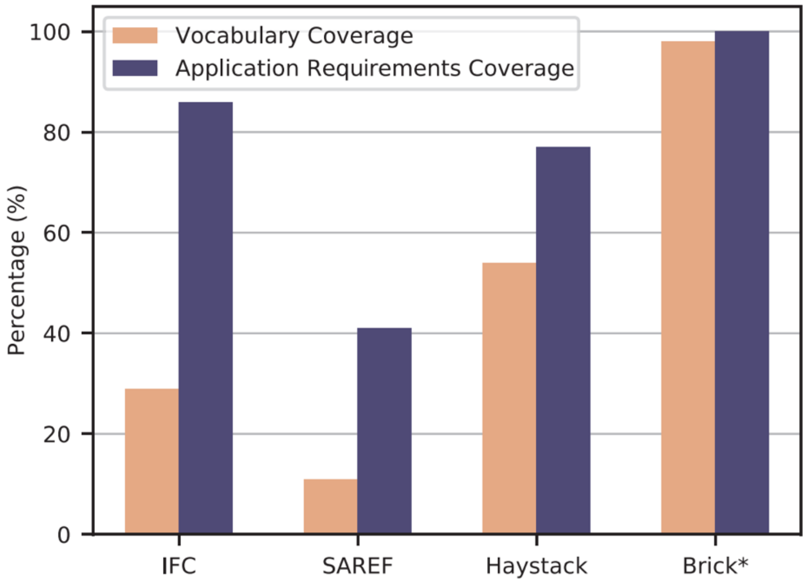

FIGURE 1: COMPARISON OF DIFFERENT SCHEMATA FOR BUILDINGS USING 89 APPLICATIONS (APPS) AND THREE BULDINGS TO EVALUATE IFC, SAREF AND HAYSTACK

4. Services Input and Output Data

4.1. Sources of Input and Output Data

We outline in the following Table 1 the partners that contributed to the definitions of data requirements of their services:

TABLE 1 DOMOS DEMONSTRATION SITES

No	Company	Demonstration site location	Buildings	Work package
WP4	CSEM	Neuchâtel	1 mixed residential / tertiary building	WP4 Common Services
WP5	EDF	Paris	220 households	WP5 Smart Services for Electrical Energy
WP5	OIKEN	Sion	200 single family houses, 2 multi-family houses	WP5 Smart Services for Electrical Energy
WP6	NEOGRID	Aalborg	340 households (26 multi-family buildings, 20 single-family houses)	WP6 Smart Services for District Heating
WP7	SUNTHERM	Skive	6 single-family houses with SUNTHERM heat pumps, 6 single-family houses with legacy heat pumps.	WP7 Smart Heat Generation Control

We set up collaborative repository to allow collaborators to share their schema, input and output data, as defined as follows:

- **Data schema** refers to the skeleton structure that represents the logical view of the entire dataset. It defines how the data is organized and how the relations among them are associated.
- **Input data** refers to the data *required* by each demonstrator site. This data can refer to the data required from external resources (e.g., weather data) or data required from other services (e.g., flexibility).
- **Output data** refers to the data from the demonstrator site that will be *accessible to external entities* (e.g., services). This data can be either already existing in the site or shall exist.

The collected Input data and Output data attributes contain the following columns:

- **Identifier:** the unique identifier of the data attribute.
- **#Data set:** identifier of the dataset associated to the required data attribute (if available)
- **Data attribute name:** the name of the data attribute
- **Description:** the description of the data attribute
- **Consumed by service(s):** the service(s) that will consume the required data attribute.
- **Produced by service(s):** the service(s) that will produce the required data attribute.
- **Sampling frequency:** the sampling frequency of the required data attribute.
- **Delivery frequency:** the delivery frequency of the required data attribute.
- **Data type:** the data type of the data attribute.
- **Attribute value pattern:** the pattern of the attribute value (e.g., dd/mm/yyyy).
- **Unit of measure:** the measurement unit of the data attribute (e.g., energy unit, temperature unit).
- **Allowed values:** the acceptable minimum/maximal ranges of the required data attribute.
- **Example:** an example of the value of the required data attribute.
- **Supporting materials:** links to reports, to datasets, and documentations of the required data (if available).
- **#User story:** the id of the related user story (if available).

The collected input and output data can be found in the domOS repository.

4.2. Neuchâtel Demonstration Site

The demonstration site is located in the city of Neuchâtel (Switzerland), within 5-minute walking distance from the CSEM headquarters. The objective of the demonstration in Neuchâtel is to validate the need-based heat generation control service. We depict in Table 2 and Table 3 the collected input and output data, respectively from the Neuchâtel demonstration site.

4.2.1. Input Data

TABLE 2: INPUT DATA FOR NEUCHÂTEL DEMONSTRATION SITE

Data attribute	Description	Consumed by	Delivery frequency
HC1 SP T	Temperature set-point for heating circuit 1	Closed-Loop Energy Efficiency Services	Every 5 minutes
HC2 SP T	Temperature set-point for heating circuit 2	Closed-Loop Energy Efficiency Services	Every 5 minutes
HC3 SP T	Temperature set-point for heating circuit 3	Closed-Loop Energy Efficiency Services	Every 5 minutes
Heat meter power	Measured heat meter power	Closed-Loop Energy Efficiency Services	Every 5 minutes
Heat meter energy	Measured heat meter energy	Closed-Loop Energy Efficiency Services	Every 5 minutes
Heat meter flow	NICE TO HAVE: measured heat meter flow	Closed-Loop Energy Efficiency Services	Every 5 minutes
Heat meter temperatures	NICE TO HAVE: measured forward/return heat meter temperatures	Closed-Loop Energy Efficiency Services	Every 5 minutes
Heat generator power	NICE TO HAVE: thermal (or primary) power of the heat generator	Closed-Loop Energy Efficiency Services	Every 5 minutes
Heat generator status	NICE TO HAVE: status of the heat generator	Closed-Loop Energy Efficiency Services	Every 5 minutes

4.2.2. Output Data

TABLE 3: OUTPUT DATA FOR NEUCHÂTEL DEMONSTRATION SITE

Data attribute	Description	Produced by	Delivery frequency
STV indoor temperature	Indoor temperature measured by STV (STV = Smart Thermostatic Valve)	Closed-Loop Energy Efficiency Services	Every 5 minutes
STV temperature set-point	Temperature set-point for the STV	Closed-Loop Energy Efficiency Services	Every 5 minutes
STV measured opening	Measured STV opening	Closed-Loop Energy Efficiency Services	Every 5 minutes
STV opening set-point	Set point for STV opening	Closed-Loop Energy Efficiency Services	Every 5 minutes

4.3. Paris Demonstration Site

The Ile-de-France region around Paris accounts for 19% of the population of metropolitan France for only 2.2% of its area, making it the most populated and most densely populated region of France. The objective of the Paris demonstration site is to develop services that encourage the members to develop energy-saving behaviours and to shift white appliances operation at times of high grid load / high electricity prices. We depict in Table 4 the collected input data from the Paris demonstration site.

4.3.1. Input Data

TABLE 4: INPUT DATA FOR PARIS DEMONSTRATION SITE

Data attribute	Consumed by	Delivery frequency
Time	Smart Services for Electrical Energy	Every 2 seconds
TimeStatus	Smart Services for Electrical Energy	Every 2 seconds
TimeZone	Smart Services for Electrical Energy	Every 2 seconds
ZCLVersion	Smart Services for Electrical Energy	N/A
ApplicationVersion	Smart Services for Electrical Energy	N/A
StackVersion	Smart Services for Electrical Energy	N/A
HWVersion	Smart Services for Electrical Energy	N/A
ManufacturerName	Smart Services for Electrical Energy	N/A
ModelIdentifier	Smart Services for Electrical Energy	N/A
DateCode	Smart Services for Electrical Energy	Every 2 seconds
PowerSource	Smart Services for Electrical Energy	Every 2 seconds
CurrentSummationDelivered	Smart Services for Electrical Energy	Every 2 seconds
Current Tier i Summation Delivered	Smart Services for Electrical Energy	Every 2 seconds
UnitofMeasure	Smart Services for Electrical Energy	N/A
Multiplier	Smart Services for Electrical Energy	N/A
Divisor	Smart Services for Electrical Energy	N/A
CompanyName	Smart Services for Electrical Energy	N/A
Meter Type ID	Smart Services for Electrical Energy	N/A
Data Quality ID	Smart Services for Electrical Energy	N/A
MeasurementType	Smart Services for Electrical Energy	Every 2 seconds
TotalActivePower	Smart Services for Electrical Energy	Every 2 seconds
TotalApparentPower	Smart Services for Electrical Energy	Every 2 seconds
PowerMultiplier	Smart Services for Electrical Energy	Every 2 seconds
PowerDivisor	Smart Services for Electrical Energy	Every 2 seconds
RMSVoltage	Smart Services for Electrical Energy	Every 2 seconds
RMSCurrent	Smart Services for Electrical Energy	Every 2 seconds
ActivePower	Smart Services for Electrical Energy	Every 2 seconds
ApparentPower	Smart Services for Electrical Energy	Every 2 seconds
RMSVoltagePhB	Smart Services for Electrical Energy	Every 2 seconds
RMSCurrentPhB	Smart Services for Electrical Energy	Every 2 seconds
ActivePowerPhB	Smart Services for Electrical Energy	Every 2 seconds
ApparentPowerPhB	Smart Services for Electrical Energy	Every 2 seconds
RMSVoltagePhC	Smart Services for Electrical Energy	Every 2 seconds
RMSCurrentPhC	Smart Services for Electrical Energy	Every 2 seconds
ActivePowerPhC	Smart Services for Electrical Energy	Every 2 seconds

4.3.2. Output Data

Our collaborators from Paris demonstrator sites did not report output data. We may have output data in the future design/development of the services.

4.4. Sion Demonstration Site

Sion, the capital of Canton Valais, is located in the heart of the Swiss Alps, South West of Switzerland. The objective of the Sion demonstration site is to deploy Smart Services related to electricity grid control, flexibility services and open-loop energy efficiency services. We depict in Table 5 the collected input data for the Sion demonstration site.

4.4.1. Input Data

TABLE 5: INPUT DATA FOR SION DEMONSTRATION SITE

Data attribute	Consumed by	Delivery frequency
Active power import +P	Electricity grid control	Every 5 seconds
Active power export -P	Electricity grid control	Every 5 seconds
Reactive power import +Q	Electricity grid control	Every 5 seconds
Reactive power export -Q	Electricity grid control	Every 5 seconds
Current Lx	Electricity grid control	Every 5 seconds
Voltage Lx	Electricity grid control	Every 5 seconds
Active energy import +A (Q1, Q4)	Electricity grid control	Every 5 minutes
Active energy export -A (Q2, Q3)	Electricity grid control	Every 5 minutes
Reactive energy +Ri (QI)	Electricity grid control	Every 5 minutes
Reactive energy +Rc (QII)	Electricity grid control	Every 5 minutes
Reactive energy -Ri (QIII)	Electricity grid control	Every 5 minutes
Reactive energy -Rc (QIV)	Electricity grid control	Every 5 minutes
Active energy import +A (QI+QIV) rate 1	Electricity grid control	Every day
Active energy import +A (QI+QIV) rate 2	Electricity grid control	Every day
Active energy export -A (QII+QIII) rate 1	Electricity grid control	Every day
Active energy export -A (QII+III) rate 2	Electricity grid control	Every day
Indoor temperature	Flexibility Services	Every minute
Domestic hot water boiler temperature	Flexibility Services	Every minute
Power	Flexibility Services	Every 15 seconds
Current Lx	Open-Loop Energy Efficiency Services	Every 15 seconds
Energy	Open-Loop Energy Efficiency Services	Every 15 seconds
Power relay	Open-Loop Energy Efficiency Services	Every 10 seconds
Heated surface	Open-Loop Energy Efficiency Services	N/A
Renovation date	Open-Loop Energy Efficiency Services	N/A
Number of occupants	Open-Loop Energy Efficiency Services	N/A
Heat appliance types	Open-Loop Energy Efficiency Services	N/A
Heat appliance roles	Open-Loop Energy Efficiency Services	N/A

4.4.2. Output Data

Our collaborators from the Sion demonstration site did not report any output data. However, in the future, the demonstration site may define output data related to flexibility or other services. Our collaborators from the Sion demonstration site did not report any output data. However, in the future, the demonstration site may define output data related to flexibility or other services.

4.5. Aalborg Demonstration Site

Kildeparken is a residential area on the east side of Aalborg that has undergone a low-carbon refurbishment. All the buildings are heated by district heating, where the water is mixed down to a lower temperature, distributed to the apartments. The objective of the Aalborg demonstration site is to deploy smart services for district heating such as an adaptive DH control based on dynamic cost signals. We depict in Table 6 the input data for the demonstration site.

4.5.1. Input Data

TABLE 6: INPUT DATA FOR AALBORG DEMONSTRATION SITE

Data attribute	Produced by	Consumed by	Delivery frequency
Zone_i CO2	IoT sensors	Smart Services for District Heating	Every 5 minutes
Zone_i Temperature	IoT sensors	Smart Services for District Heating	Every 5 minutes
Zone_i Humidity	IoT sensors	Smart Services for District Heating	Every 5 minutes
Building_i energy	External meters	Smart Services for District Heating	Every 5 minutes
Building_i forward temperature	External meters	Smart Services for District Heating	Every 5 minutes
Building_i return temperature	External meters	Smart Services for District Heating	Every 5 minutes
Building_i hot water energy	External meters	Smart Services for District Heating	Every 5 minutes
Building_i mixing-loop forward temperature setpoint	Building Management System	Smart Services for District Heating	Every 5 minutes
Building_i mixing-loop HW forward temperature setpoint	Building Management System	Smart Services for District Heating	Every 5 minutes
Local area mixing-loop forward temperature setpoint	Building Management System	Smart Services for District Heating	Every 5 minutes
Wind direction	External weather API	Smart Services for District Heating	Every 1 hour
Wind speed	External weather API	Smart Services for District Heating	Every 1 hour
Sun radiation	External weather API	Smart Services for District Heating	Every 1 hour
Ambient temperature	External weather API	Smart Services for District Heating	Every 1 hour

4.5.2. Output Data

The output data generated by Aalborg demonstration site is the same as its input data. The output data should be consumed by SUNTHERM for smart heating generation control services. Therefore, we identify a need for semantic interoperability between the two demonstration sites to exchange data in a coherent way.

4.6. Skive Demonstration Site

Skive Municipality is located in the central part of Jutland in Denmark. The municipality has the vision to become CO2 neutral before 2030. The objective of skive demonstration site is to deploy smart heating generation control. We depict in Table 7 the collected input data for Skive demonstration site.

4.6.1. Input Data

TABLE 7: INPUT DATA FOR SKIVE DEMONSTRATION SITE

Data attribute	Produced by	Consumed by	Delivery frequency
Zone_i CO2	IoT sensors	Smart heating generation control	Every 5 minutes
Zone_i Temperature	IoT sensors	Smart heating generation control	Every 5 minutes
Zone_i Humidity	IoT sensors	Smart heating generation control	Every 5 minutes
Building_i energy	External meters	Smart heating generation control	Every 5 minutes
Building_i forward temperature	External meters	Smart heating generation control	Every 5 minutes
Building_i return temperature	External meters	Smart heating generation control	Every 5 minutes
Building_i hot water energy	External meters	Smart heating generation control	Every 5 minutes
Building_i mixing-loop forward temperature setpoint	Building Management System	Smart heating generation control	Every 5 minutes
Building_i mixing-loop HW forward temperature setpoint	Building Management System	Smart heating generation control	Every 5 minutes
Local area mixing-loop forward temperature setpoint	Building Management System	Smart heating generation control	Every 5 minutes
Wind direction	External weather API	Smart heating generation control	Every 1 hour
Wind speed	External weather API	Smart heating generation control	Every 1 hour
Sun radiation	External weather API	Smart heating generation control	Every 1 hour
Ambient temperature	External weather API	Smart heating generation control	Every 1 hour

4.6.2. Output Data

The output data of Skive demonstration site is the same as its input. This data will be used for services related to smart heat generation control. These services are not defined yet.

5. Descriptive Analysis of Input and Output Data

In this section, we analyse the input and output data according to different points of view. We have collected a total of 134 data attributes for input and output data. We have collected 102 input data attributes and 32 output data attributes.

5.1. Comparison Between Static Data and Dynamic Data

We first compare the type of input data and output data, which can be either static or dynamic. Static data is data that does not change after being recorded. It is a fixed data set. Dynamic data may on the other hand change after it is measured and must be continually updated. We identified a total of 114 dynamic data attributes. Whereas the number of static data attributes is 17. The majority of the data in the domOS project is dynamic. The collaborators may need a considerable work to perform data flow verification, storage, and analysis on the dynamic data. Based on the current status, more data attributes will be addressed in T3.2 and documented by dCO in D3.2.

5.2. Data Attributes Overlap in Demonstration Sites

We find an overlap between input and output data attributes as the number of identical data attributes that can be the output of a given demonstration site and input of another site. We have identified a total of 56 overlapping data attributes. For instance, some of the data attributes of the Aalborg demonstration site will be set as the input for Suntherm for the smart heating generation controls. Our collaborators from Neogrid and Suntherm stated that they will be collecting a lot more data. We will address this in the next iterations of T3.2, and we will represent these data attributes in dCO, which will be documented in D3.2. We can identify here an interoperability issue that can be resolved using a common information model, an ontology, to ensure semantic interoperability between services. We notice also that all the overlapping data attributes are dynamic. Therefore, our collaborators from the demonstration sites should consider data flow verification to ensure the reception of accurate data. We expect more overlap to be found when the design of the services will end.

5.3. Dynamic Data Attributes Analysis

We count a total of 114 dynamic data attributes. In the following table, we depict the number of data attributes for each delivery frequency for all the five demonstration sites. We can conclude from the table that the majority of dynamic data should be delivered in under 10 minutes. This data flow should be supported by demonstration sites infrastructure. Having such fast delivery frequency should also be coupled with data flow verification (cf. Section 6.2.1).

TABLE 8 THE NUMBER OF DATA ATTRIBUTES FOR EACH DELIVERY FREQUENCY

Number of data attributes	Delivery frequency
23	Every 2 seconds
6	Every 5 seconds
1	Every 10 seconds
3	Every 15 seconds
2	Every 1 minute
59	Every 5 minutes
16	Every 1 hour
4	Every 1 day

6. Functional Data Requirements Specification

We present in this section the functional data requirements. These requirements define the functionalities that should be deployed using data in demonstration sites. Multiple stakeholders can interact with data. Therefore, we collected data requirements using the common user stories defined by the WP2. We have identified data requirements from users, platform operators, and service providers. Each data requirement has an identifier. In the following, we define each of the stakeholders.

- A **service provider (SP)** is a company that offers smart services to users through the Platform Operator.
- A **user (U)** is a person who owns a building or part of a building and consumes smart services.
- A **platform operator (PO)** is an entity that operates a domOS instance. The platform operator must provide some infrastructure for the users and guarantee the security of the system.

We have collected from the user stories a total of 11 functional data requirements for the platform operator, service provider, and the user. Services providers provided 9 data requirements, 7 of them are related to the accessibility to different kind of data (e.g., historical data, buildings metadata). In the following, we describe each of the collected data requirements from every stakeholder.

6.1. Service Provider Data Requirements

6.1.1. Real Time Data Access of Measures (SP4)

Service providers require real-time data access to measures (e.g., indoor and outdoor temperature, consumed energy, humidity). Therefore, service providers could take the right actions at the right time using the dedicated algorithms. For instance, service providers can perform real-time management of heating, ventilation, and air conditioning in buildings.

6.1.2. Exclusive Access to Data Setpoints (SP8)

This requirement addresses the problem of securing access to data setpoints. For example, if a service provider defines a temperature setpoint for a heating system, the system should maintain the temperature and secure it from interference from the other services. Therefore, by default, service providers require an exclusive access to the data setpoints, and no other service should be able to change the setpoints unless having an approval.

6.1.3. Historical Data Access of New Customers (SP11)

Service providers would like to have access to historical data of new customers. This requirement means that if a service provider is entering into a new relationship with a customer, and if the customer has hardware with interesting data, it could be desirable for the service provider to access historical data. Historical data of new customers helps service providers to adapt their service accordingly and measure improvements.

6.1.4. Building Metadata Access (SP13)

Service providers would like to have access to the building’s metadata to predict consumption profiles and design services accordingly. These data allow to design of more reliable prediction algorithms and therefore maximizing the use of available flexibility within the network. Metadata will include information such as the building's surface, volume, location, positioning, and orientation for precise weather data and appliance type/brand for a better understanding of the process.

6.1.5. Demographic Data Access (SP14)

Services providers would like to have access to customers' demographic data to design services accordingly. These data will be used in the framework of the Living Lab Integrative Process to co-design common services. It includes information about age, gender identity, and personal understanding of energy issues and challenges. Personalized interviews could be conducted to collect additional information from customers.

6.1.6. Data Flow Prediction (SP6)

Service providers should be able to predict the data flow. This allows services to coordinate actions and jointly perform planning ahead in time, thus making them much more capable and powerful. This requirement will focus on extending the FlexOffer concept (originally used for energy only) to be able to capture predictions, flexibility, and uncertainty of various measures (e.g., consumed energy amounts, indoor temperature, no. of occupants).

6.1.7. Common Data Semantics Across all Platforms (SP12)

Semantic data interoperability is the capability of two different agents (human or software) or more to exchange information across all platforms and to have the same interpretation (meaning) of this information. Common data semantic allows service providers to develop services in an interoperable way to generate an added value. A common information model should not only model the set of IoT entities, but also the semantic relationship and constraints between these entities.

6.2. Platform Operator Data Requirements

6.2.1. Data Flow Verification (PO3)

Data flow can contain anomalies. Anomalies can occur as a result of abnormal events or unusual behaviour such as shutdown, failures, and sensor change. Platform operators would like to verify the correctness of the data flow since it is essential to ensure efficient services. Data flow verification could, e.g., focus on the following kinds: data type, range, and consistency. The data flow should contain data with the valid maximum and minimum ranges with the correct data type. Consistency verifications ensure that data is logical. For example, the sampling date of a sensor should not exceed the delivery date.

6.2.2. User Data Storage (PO6)

Platform operators would like to store data flow related to customers. The storage of the data flow will result in historical data that will help the platform operator to provide the adequate infrastructure for the customers.

6.3. User Data Requirements

6.3.1. Control of the Access of User Data (U3)

This requirement concerns mainly the privacy of user data. It has been decided that no service provider will be able to access data unless authorized by the customer. Therefore, there should be a mechanism to allow each user to grant and revoke the access of its data. Prior to the GDPR, the burden was on the user to take privacy protecting measures within a given product or a service; by changing the default settings, opting out, or turning on access controls, for example on location data (EU, 2016).

7. Conclusion

This document defines the data requirements collected from demonstration sites. We have collected input and output data from each demonstration site. The collected data is a primary input that will help the development of dCO ontology. We have also collected functional data requirements. These requirements define the functionalities and needs of the demonstration sites. We have also reviewed the relevant existing ontologies related to the scope of the domOS ontology. As future work, we will collect the ontological requirements using the Ontological Requirements Specification Document (ORSD). The ORSD will be made in collaboration with our partners to define the scope, the users, the use cases, the functional and non-functional requirements of dCO. We will define the dCO development infrastructure and we will use the adequate methodology to build dCO ontology for the deliverable D3.2.

8. References

Balaji B. [et al.] Brick: Metadata schema for portable smart building applications [Journal] // Applied energy. - 2018. - Vol. 226. - pp. 1273-1292.

Bazjanac V. and Crawley D. B. Industry foundation classes and interoperable commercial software in support of design of energy-efficient buildings [Conference] // Building Simulation. - [s.l.] : roceedings of Building Simulation’99. Vol. 2., 1999.

Daniele L., den Hartog F. and Roes J. Created in close interaction with the industry: the smart appliances reference (SAREF) ontology [Conference] // International Workshop Formal Ontologies Meet Industries. - 2015.

EU General Data Protection Regulation (GDPR) [Journal] // Official Journal of the European Union. - 2016.

Guinard D. D. and Trifa V. M. Building the web of things. [Book]. - [s.l.] : Shelter Island: Manning Publications, 2016.

Gyrard A., Serrano M. and Atemezing G. A. Semantic web methodologies, best practices and ontology engineering applied to Internet of Things [Conference] // 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT). - [s.l.] : IEEE, 2015.

Janowicz K. [et al.] SOSA: A lightweight ontology for sensors, observations, samples, and actuators [Journal] // Journal of Web Semantics. - 2019.

Lassila O. and Swick R. R. Resource description framework (RDF) model and syntax specification [Journal]. - 1998.

Noy N. and Mcguiness D. Ontology Development 101: A Guide to Creating Your First Ontology [Conference] // Knowledge Systems Laboratory. - 2001.

Project Haystack [Online] // Project Haystack. - 2020. - https://project-haystack.org/.

Seydoux N. Towards interoperable IOT systems with a constraint-aware semantic web of things [Report]. - [s.l.] : Diss. Toulouse, INSA, 2018.

Swetina J. [et al.] Toward a standardized common M2M service layer platform: Introduction to oneM2M [Journal] // IEEE Wireless Communications. - [s.l.] : IEEE Wireless Communications., 2014. - 3 : Vol. 21. - pp. 20-26.