SATO

Self Assessment Towards Optimization of Building Energy

Deliverable 2.1

Concept of the SRI enabled SATO platform

Deliverable Lead: FC.ID

Deliverable due date: 30/04/2021

Actual submission date: 30/04/2021

Version: v2.0

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 957128.

Document Control Page	
Title	Concept of the SRI enabled SATO platform
Editor	FC.ID
Description	This deliverable identifies the main concepts that will help the SATO platform to support the dynamic self-assessment and self-optimization of the SRI
Contributors	EKAG and SAGOE
Creation date	01/04/2021
Type	Report
Language	English
Audience	☐ public ☒ confidential
Review status	☐ Draft ☐ WP leader accepted ☒ Coordinator accepted
Action requested	☐ to be revised by Partners ☐ for approval by the WP leader ☐ for approval by the Project Coordinator ☐ for acknowledgement by Partners

Table of Contents

List of Figures

Revision history

Version	Author(s)	Changes	Date
v0.1	Vasco Ferreira (FC.ID), André Gil (FC.ID), Frederico Melo (FC.ID)	Initial Version	29/03/2021
v1.0	Vinicius Cogo (FC.ID), André Gil (FC.ID), Vasco Ferreira (FC.ID)	Final Draft	19/04/2021
V1.1	Vinicius Cogo (FC.ID), Pedro Ferreira (FC.ID), Guilherme Graça (FC.ID), Daniel Albuquerque (FC.ID)	Revised Version	30/04/2021
V2.0	Pedro Ferreira (FC.ID)	Revised version after comments from Project Officer	29/05/2022

DISCLAIMER

EXECUTIVE SUMMARY / ABSTRACT / SCOPE

The SATO project aims to create a new energy self-assessment and optimisation platform (i.e., the SATO platform) that integrates and monitors energy consumption of building equipment and appliances. This platform will support a self-assessment framework and optimisation services that will contribute to lower the energy consumption of buildings and increase their energy flexibility, efficiency, and occupants' satisfaction.

The Deliverable D2.1 identifies the main concepts that will help the SATO platform to support the dynamic self-assessment of the Smart Readiness Indicator (SRI) and the subsequent self-optimization towards more energy-efficiency and comfort of building's occupants. This document is a direct output from Task T2.1 (Development of the SRI enabled SATO platform concept) from the WP2 (Development of integrated technical platform for SATO) and will guide the design and development tasks in WP2 of the SATO project.

More specifically, this deliverable presents the definition of the SRI, discusses its benefits, compares some existent initiatives that implement it, and presents open opportunities the SATO project may harness to handle the dynamic nature of the SRI. Additionally, we identified seven concepts that help the SATO platform to support the SRI in more effective and dynamic ways than it is supported by existing works in the literature. These concepts are the systematic construction of the building inventory, automatic calculation of the SRI preliminary scores, security and privacy mechanisms, event-based communication, software-defined infrastructures, data lakes, and interoperability (semantical, syntactic, and technical). Deploying, configuring, and integrating the SATO platform with the pilots from the SATO project will be easier and more efficient if using the concepts we identify and discuss in this deliverable.

1. Introduction

1.1. Motivation

Smart buildings integrate multiple Information and Communication Technology (ICT) solutions to monitor and optimize the energy-efficiency of devices and appliances within a building. They may enable the integration of energy flexibility based on availability, price, and demand into daily operations. The Smart Readiness Indicator (SRI) of a building is a concept that refers to the ability of a building to efficiently adapt to changes in variables related to its internal (e.g., occupants) and external (e.g., weather and energy grids) environments [1].

The introduction of the SRI motivates evaluating and enhancing the use of smart devices that can timely accommodate users' and energy grid's needs while reducing energy consumption. The SATO project intends to directly contribute to the accomplishment of these (and many other) objectives. Figure 1 illustrates some of the expected advantages of enabling a higher degree of energy-management automation and operation optimization.

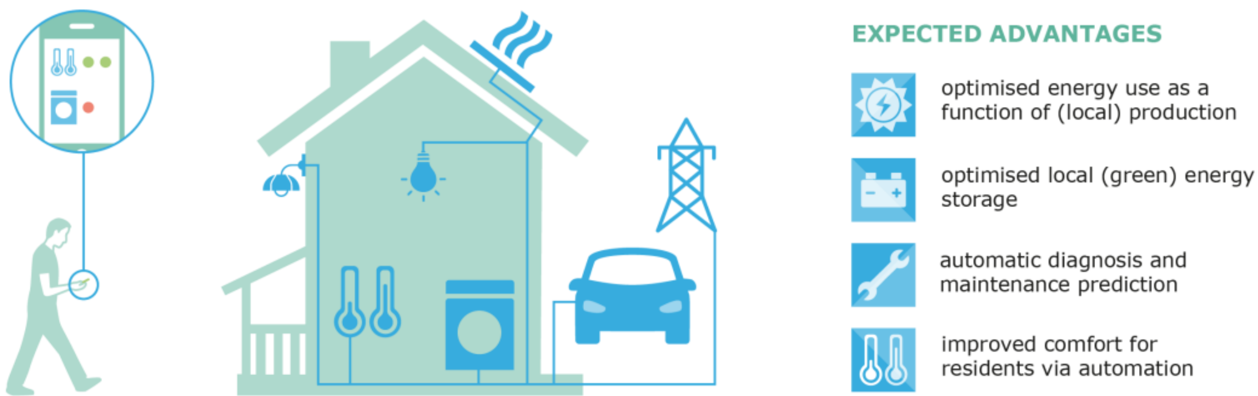

Figure 1: Expected advantages of monitoring and optimizing the energy-efficiency in smart buildings. Image adapted from [2].

1.2. Objective

The main objective of this Deliverable D2.1 is to introduce several concepts that will help the components from the SATO platform to support the self-assessment of the Smart Readiness Indicator (SRI) of buildings. It directly relates with the specifications and requirements collected in WP1 (Specifications and Requirements for SATO). Moreover, it describes in detail the main concepts that will guide the design and development of the SATO platform in WP2 (Development of integrated technical Platform for SATO). It will serve, together with the Deliverable D1.3 (SATO Platform, SRI and IT Security Requirements), as the basis for the Deliverable D1.4 (Description of the system architecture of the SATO platform), which is the main document that will describe the foreseen architecture of the SATO platform, discuss the existent solutions that can be integrated with the platform, and provide a rationale in the decisions for the implementation of the platform.

1.3. Structure of the Document

The structure of this document is divided into four chapters as the follows:

- **Chapter 1** provides the motivation and introduction of the context for this deliverable.
- **Chapter 2** introduces the Smart Readiness Indicator and describes how the SATO platform will advance the state-of-the-art in self-assessing the SRI score systematically.

- **Chapter 3** presents several concepts that will be integrated into the SATO platform for the appropriate dynamic support of the SRI.
- **Chapter 4** encloses this deliverable with some final remarks and recommendations for the design and development of the SATO platform.

2. Smart Readiness Indicator (SRI)

This chapter discusses several important aspects associated with the Smart Readiness Indicator (SRI). Section 2.1 introduces the concept of the Smart Readiness Indicator while Section 2.2 discusses how the SRI is calculated using the predefined impact criteria and domains. Section 2.3 briefly discusses research work related to the SRI independent from the EU regulation in force. Finally, Section 2.4 presents some open opportunities left by other initiatives, which will motivate the concepts we are adopting in the SATO project to enable the SRI.

2.1. SRI Concept

The Smart readiness indicator (SRI) is a common EU scheme for rating the smart readiness of buildings[1]. The SRI of a building or building unit is and indicator that allows rating and communicating through a certificate, the smart readiness of a building or building unit [2]. As stated in the European Commission (EC) delegated regulation (EU) 2020/2155 [2]:

The smart readiness indicator shall allow for the assessment of the capabilities of a building or building unit to adapt its operation to the needs of the occupant and of the grid and to improve its energy efficiency and overall in-use performance. The smart readiness indicator shall cover features for increased energy savings, benchmarking and flexibility, and enhanced functionalities and capabilities provided by more interconnected and intelligent devices.

The smart readiness indicator shall include the smart readiness rating of a building or building unit and a set of smart readiness scores that reflect the smart readiness of buildings, building units and systems along predefined key functionalities, impact criteria and technical domains.

The main goal with SRI is to raise awareness about energy efficiency and motivate improving the overall performance of buildings through automation and electronic monitoring of technical building systems and components, which ultimately contribute to other initiatives promoting sustainability and decarbonization while addressing climate change and other environmental challenges.

The requirements for establishing the SRI were introduced in 2018 through an amendment to the EPBD – Energy Performance of Buildings Directive [1]. According to this document, the SRI should be a simple indicator that combines multiple factors related to the management and interaction of the building with its occupants and the grid. It should raise awareness, amongst building owners, managers, and occupants, of the value behind building automation and systematic monitoring about the actual savings those enhanced functionalities provide for them [1].

2.2. SRI Assessment

The methodology for calculating the SRI is defined in the EC delegated regulation (EU) 2020/2155 [2]. It is based on the assessment of buildings smart-ready services and their functionality level and is expressed by the ratio of a building SRI compared to the maximum SRI that it could reach.

The methodology for assessing the SRI relies upon stationary analyses of several qualitative aspects that compose a matrix with impact criteria, technical domains, and services impacting the buildings'

[1] Smart Readiness Indicator. https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/smart-readiness-indicator_en

energy management. This SRI framework includes seven main impact criteria grouped in three key functionalities (e.g., energy savings, comfort, flexibility) and nine technical domains (e.g., heating, lighting) as presented in Figure 2.

For each technical domain (i.e., matrix row), a variety of corresponding functionality levels are defined, which represent different smartness stages. Functionality levels range from zero to the highest available stage of smartness functionalities (e.g., 2–5), where the former refers to using only non-smart devices and the latter varies according to the services involved. In the next section, we will describe how the SATO platform enables the systematic assessment of these scores.

The next step in this assessment is to convert the obtained domains and impact criteria into the SRI itself. To accomplish this value assessment, the different functionality levels of each service receive weighting factors according to the climate conditions surrounding the building and its geographic orientation. Then, each criteria score is aggregated by a weighted sum that will produce the building SRI. Figure 3 depicts a hypothetical example of a weight distribution among the several impact criteria to compose the SRI final score of a building in a certain region. The weighting of impact criteria in key functionalities and the weighting factors of technical domains may be defined by EU Member States.

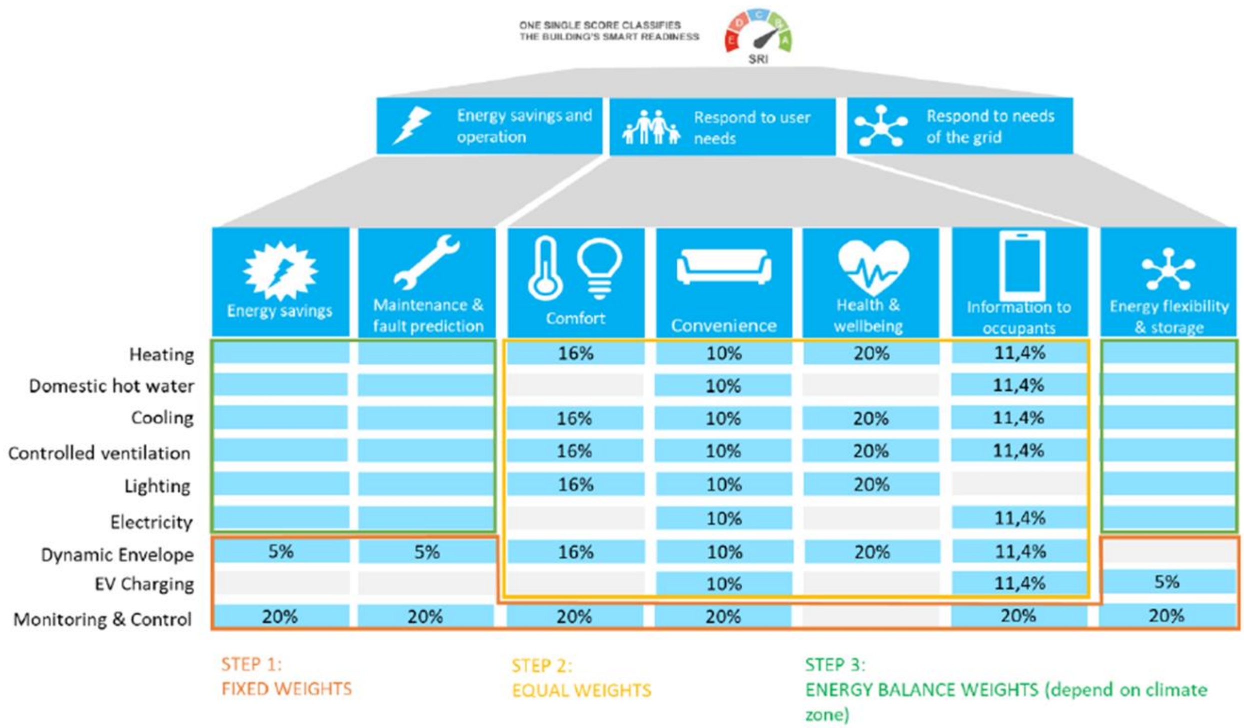

Figure 2: SRI calculation matrix, and its weighting factors. Image adapted from [2].

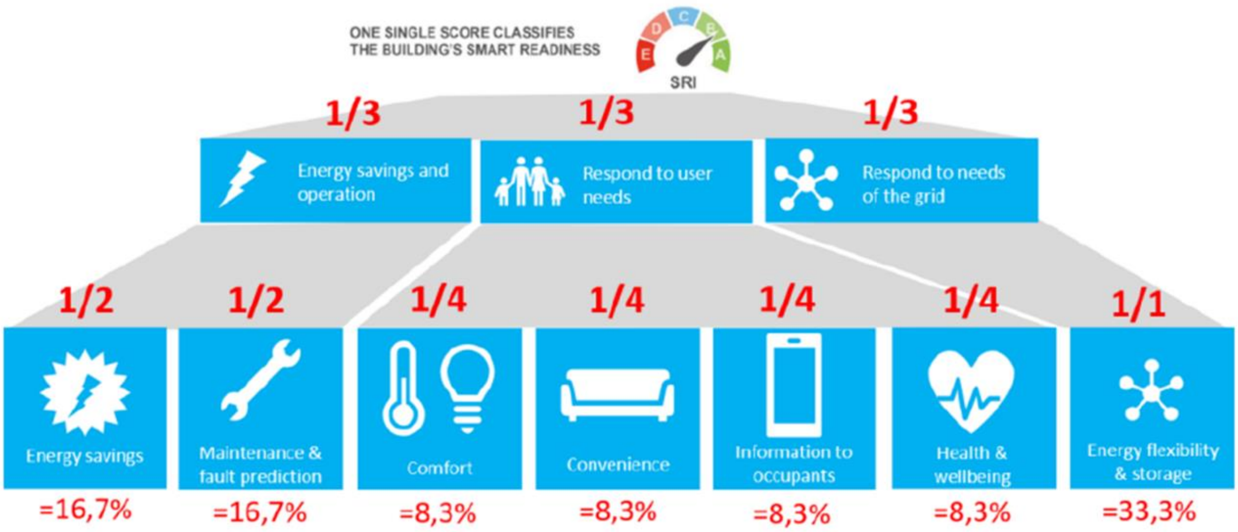

Figure 3: Weight distribution among impact criteria. Image adapted from [2].

2.3. SRI related work

Beyond the EU regulation applicable to the SRI [2] that has entered into force in 2020, research work has also proposed related complementary features or methodologies. One of the first initiatives calculated the smart readiness indicator (SRI) of some buildings as use-case analysis to evaluate how complete the SRI currently is and proposed some improvements for SRI's coverage [3]. Another initiative hired experts to calculate the SRI for other scenarios and provided recommendations for the implementation of SRI to be effective and broad [4]. The same authors of [3] have extended their previous methodology to consider smart districts rather than only individual smart buildings [5]. More recently, work [6] was conducted to allow the integration of SRI concepts into a building automation and control system.

2.4. Open Opportunities

A systematic solution for obtaining the building inventory is an important step for the automatic SRI assessment, but none of the works presented in the previous subsection provides this feature. Additionally, to the best of our knowledge, there is also no scientific study that analyses the move from a stationary assessment of the SRI to a dynamic assessment. Systematically obtaining the building inventory and automatically calculating the preliminary SRI scores are a must for certification experts (resp. building managers) who want to timely (resp. periodically) overview the smartness state of a building.

In this way, the SATO project has the open opportunity to timely address the move to a dynamic SRI assessment, which will contribute to the efficiency and adapting the smart energy management ecosystem. In the next section, we will describe the main concepts and capabilities of the SATO platform that will enable this migration and the resulting added value it provides for the main stakeholders.

3. Enabling the SRI with the SATO platform

In this chapter, we describe how SATO integrates the SRI and the main technical and architectural concepts that will directly contribute to the design and development of the SATO platform towards enabling the SRI self-assessment and optimization.

Section 3.1 describes how SATO will integrate the SRI assessment and how it may provide dynamic updates of the SRI. Section 3.2 describes how the SATO platform will address the challenge of creating the inventory of smart devices and appliances available in the building automatically and systematically. Section 3.3 discusses the concept of having the SRI scores calculated automatically and periodically. Section 3.4 discusses the main concepts that will take place in the SATO platform regarding the security and privacy. Section 3.5 describes the event-based communication approach, which will be crucial for the SATO platform. Section 3.6 presents the idea of software-based infrastructures, which enable separating control and data planes (or layers), as well as programmatically deploying and configuring the platform. Section 3.7 introduces the data lake approach for dealing with data across the whole platform. Finally, Section 3.8 recalls some important aspects associated with the idea of interoperable systems.

3.1. SRI in the SATO platform

The SATO platform and project will support the SRI at three distinct levels:

1. Storing the buildings SRI. The SATO platform will store SRI assessments of buildings or building units executed by qualified SRI experts, for instance through the SRI assessment package[2]. The assessments will be organized chronologically to allow a time-based analysis of SRI features, such as key functionalities, SRI scores, and impact criteria.
2. Applying the SRI. The SATO project will use the SRI assessment package to perform baseline SRI assessments on some of the pilot buildings. This will be done before and after deploying the SATO platform in the pilots. The first baseline enables analysis of the SATO platform contribution to the pilot SRI. The second baseline allows the evaluation of SATO triggered SRI improvements.
3. Updating the SRI. The SATO platform will provide dynamic updates of the SRI by automatically detecting relevant features related to the technical domains and impact criteria. Updates will also be stored chronologically and in relation to a baseline SRI assessment previously stored. Three update triggers are envisaged:
    a. Assessment triggered updates. Resulting from the SATO Self-Assessment Framework, the self-assessments will provide and communicate information with impact on some of the SRI technical domains and impact criteria.
    b. Device triggered updates. These happen whenever a device or system whose data or functionalities affect the SRI calculation is detected as present or absent in the building.
    c. Data triggered updates. Whenever data analysis and inspection allow inferring the presence or absence of technical domain functionalities relevant to the SRI. Two data sources are to be considered:
        i. Metadata present in the communication, semantic models, or knowledge graphs.
        ii. Data values and time-series.

The results storage component of the SATO platform architecture (please consult SATO deliverable 1.4) will be used to store a data structure holding the data employed by qualified SRI experts to calculate SRI assessments. This data structure will reflect the current SRI regulation and will be populated, for instance by importing data from the SRI assessment package files.

[2] https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/smart-readiness-indicator/sri-implementation_en

Then, when triggered, incremental dynamic updates to SRI scores will be stored in relation to reference baseline assessments executed by SRI experts. The triggers will be monitored by an assessment software component that will survey different information sources to trigger the updates, such as BIMs, knowledge graphs resulting from the different ontology components, metadata in IoT level communication, assessment results, or historical data from building systems.

The following subsections provide further details on technical and architectural concepts that will contribute to an efficient SATO platform implementation and to an effective SRI integration.

3.2. Systematic Building Inventory

Enabling the systematic collection of the building inventory is one of the primary steps towards a dynamic SRI assessment. It must include all energy-consuming systems, smart devices appliances, sensors and actors present in a building, including their brands, models, configurations, status, and capabilities (using open standard semantics and data models). This process will assume devices are either SATO-enabled or legacy devices. Devices will self-announce themselves to the platform in the former, whereas gateways in the building will collect and forward data to the SATO platform in the latter. Legacy devices include those that already were in the building before the SATO project and those newly bought that have no self-announcement features. Data can be provided through two main pathways: it can come directly from devices and gateways from buildings or be sent first to a cloud and subsequently forwarded/fetched to the SATO platform. Moreover, this process must monitor the availability status of components (e.g., through heartbeat mechanisms) and must consider efficiency in providing reports by only forwarding state changes instead of complete periodical reports.

All this information will be periodically provided to the SATO platform, which will store up-to-date snapshots of the available devices and will try to determine their physical distribution in the building. Furthermore, the dynamic nature of a building occupancy and the availability of devices must be reflected in useful time in these snapshots. Devices and appliances can (and probably will) fail from time to time. These failures or omissions affect the SRI of a building since they stop providing measures and receiving actuations. Building managers cannot count on these faulty devices during their unavailability period. Obtaining snapshots in useful time provides the opportunity for building managers to timely counteract these failures since only after recovering the faulty devices enables managers to restore the optimal SRI of the building. The most up-to-date snapshot can be presented as an inventory report or through a mobile application to certification experts and building managers when requested or periodically. These reports can inform the certification experts what is the current state of the building, what can be calculated automatically (see the next section), and what need to be manually calculated for obtaining the final SRI scores.

More specifically, the SATO platform will include at least one metadata service for managing the information related to devices present in a building and another service for managing the associate location-based information. Any authorized component of the SATO platform can interact with these services to obtain partial or complete reports or metadata about the devices and appliances available in the whole building (or only portions of it). Additionally, these services will provide information using open standard solutions for semantic (e.g., SAREF [7]) and syntactic (e.g., EEBUS SPINE [8]) interoperability to enhance the data flow and SRI assessment. Finally, the device services will also enable building managers and optimization heuristics to send commands and actuations to the appropriate gateways or devices.

Trustworthy automatic inventory of plug-and-play devices and appliances in buildings will considerably facilitates SRI certification process since experts will need only to validate the reported snapshot of the building to start evaluating the SRI metrics and variables.

3.3. Automatically Calculating the Preliminary SRI Scores

A step that complements the previous one is to calculate, based on the most up-to-date snapshot of the building, the preliminary variables used to obtain the SRI scores. While most existent proposals assume the SRI of a building is stationary, the SATO project considers it an always-changing dynamic metric. For instance, building occupancy and the (un)availability of devices may quickly influence the SRI scores and affect the overall building efficiency and occupants' comfort.

The continuous assessment of the SRI of a building enables building managers and optimization heuristics to deliver the highest economic and comfort gains possible for a building given its status. Likewise in appliances' energy labels, the SRI may be affected by the surrounding environment and real-world conditions, which usually differ from the near-perfect conditions in which the labels were initially measured. If SRI scores can be continuously assessed, then they can be stored to compose a timeline that also enables certification experts to obtain a historical context of the effective SRI of the building instead of only obtaining an immediate indicator that could otherwise be tampered. This continuous monitoring enables also building managers to store the configuration that has obtained the highest seen SRI score of the building, which can be later used as input for the optimization process as a target value.

Historical data resulting from the self-assessment of SRI can be stored for long periods since it contains aggregate results that usually are smaller than raw data. Raw data (e.g., measurements) will be stored for the longest period of interest needed for the latest SRI assessment. This sliding time window will be maintained while the raw data from previous assessments can be discarded for storage efficiency. Additionally, properly discarding raw data (and keeping only the aggregate results) reduces the pressure for privacy protection (see Section 3.4).

Moreover, continuously monitoring the SRI of buildings can provide bigger contributions to the society since it enables creating knowledge databases on how buildings (and their SRI) from the same region behave and adapt to changes that have common causes. For instance, critical climatic conditions and power fluctuations affect many neighbouring buildings instead of isolated instances. Energy flexibility accounts for one third of SRI scores, which also motivates collaborative scenarios among neighbouring buildings. They may learn from each other on how to optimize their SRI and the practical economic and comfort consequences it entails. These scenarios may extrapolate SATO's results from single buildings to smart districts and beyond.

3.4. Security and Privacy

The SATO platform will collect, manage, and analyse data related with energy consumption of buildings to assess the SRI and will actuate on smart devices to optimize this indicator. Consequently, a considerable amount of data and commands passing through the SATO platform may be considered sensitive for privacy or confidentiality reasons (e.g., personal data, consumption profiles, strategic decisions). Privacy-sensitive data for the SATO project is any data that can be traced back to a single individual or even to a specific group of related individuals (e.g., a family, co-workers). SATO needs to protect the information it manages because the leakage of private data may cause individuals to be discriminated, harmed, or endangered.

A Data Management Plan (DMP) was presented in Deliverable D9.8 (Risk, Innovation and Data Management Plans). This document described what data will be generated, processed, or collected during the project, the standards that will be used, how the research data will be preserved, and what datasets will be published for verification or reuse as open access. One of the main takeovers of this deliverable concerning the data management is that security and privacy protection are primary concerns in all phases of design, development, evaluation, and dissemination of the SATO project.

SATO's applications, components, and interfaces will follow state-of-the-art security methodologies and will use available open-source tools to test the developed software and integrated systems. Examples of these methodologies include (but are not limited to) secure-by-design principles, privacy-aware data flows, strict role-based access control, two-factor user authentication, strong cryptographic methods, and machine-to-machine authentication certificates.

Widely adopted security design principles (e.g., OWASP [9]) promote the assessment of information security aspects prior to any development. Examples of aspects include identifying the main assets and attackers of the system, as well as designing the architecture with focus on the three information security pillars: confidentiality, integrity, and availability. The SATO platform will consider, since conception, the best security-by-design principles such as:

- Minimise the attack surface.
- Establish secure defaults.
- The Principle of Least Privilege (POLP).
- The principle of Defence in depth.
- Fail securely.
- Do not trust external services.
- Separation of duties.
- Avoid security by obscurity.
- Keep security simple.
- Timely fix security issues.

More specifically, a role concept will be deployed to ensure each user and application only has the minimum rights needed to accomplish its tasks and actions. A PKI (Public-Key Infrastructure) will be defined to manage all cryptographic keys and certificates, whereas standard secure communication and protocols will be employed whenever possible to offer integrity, authentication, and encryption. Penetration testing will be conducted to identify vulnerabilities as soon as possible and frequent recovery processes will guarantee their timely patch. Additionally, we foresee the use of cyberthreat intelligence (e.g., Open-source intelligence, OSINT) to increase security-awareness and promote timely detecting vulnerabilities, fixing bugs, and keeping the system protected against security attacks.

Data anonymization will ensure that sensitive data will always be reported in anonymous or aggregated ways, while information sensitivity awareness will enable conducting data flows to components with the appropriate security premises and policies according to the identified risk levels. Sensitive data will be streamed and stored in the SATO platform according to the highest security and protection procedures and in line with the GDPR and other privacy-related legislation.

3.5. Event-based Communication

Event-based communication emerges as a natural technical concept for the SATO platform to support and enable the dynamic SRI assessment and optimization. In this type of communication, events are published every time something important happens in the monitored environment. Supporting this generic concept in the SATO platform opens many opportunities of contributions, such as the ones presented in Sections 3.1 and 3.2.

At the core of event-based communication is the publish/subscribe pattern, where a publisher/producer (e.g., a temperature sensor) publishes a measurement (e.g., the current temperature in Celsius degrees) in an event bus (i.e., a communication channel) and all subscribers/consumers that are registered in this channel will be notified of the produced event [10]. Subscribers/consumers receive the event and decide what to do with it. For instance, a processing component may identify if the measured temperature is within an expected range or is anomalous. If the temperature is anomalous, it may produce another event (e.g., an alarm) that will be sent to the appropriate channel, where the building manager receives for instance the alarm for checking if the sensor is calibrated correctly. This cascade of events is a structure that decouples producers from consumers and facilitates the separation of internal steps from the assessment and optimization workflows.

Various producers (e.g., devices) can publish events in the system at the same time while multiple consumers can read the events associated with the channels they are subscribed to. There are multiple implementations of this communication model, where messaging queues (e.g., Apache Kafka [11], RabbitMQ [12], Eclipse Mosquitto [13]) are the most prominent example.

Since the SATO platform will have to deal continuously with large amounts of data, the event-based communication can also be seen as data streams [14]. Data streaming may provide different performance and reliability guarantees according to the needs of each use case. Resources can be scheduled to prioritize some events while leaving others with the remaining idle resources only. Stream processing enables the continuous transformations to enhance data with metadata, to attest the quality of data, and to structure data for more complex processing tasks in other computing models. Streams can try to deliver data as fast as possible, which can result in a particular case of soft real-time that is considered useful time for building managers.

Notwithstanding, there are many challenges associated with event-based communication that need be tackled by the SATO platform to effectively support SRI. For instance, the diversity of data types and sources, data locality, the scalability of the system to support all pilots from the projects, among others [15]. Interoperability (see Section 3.8) will be the main concept used to address the diversity of data types, meaning, and sources while the scalability of the system is taken into consideration since the concept of the SATO platform and will rely in industry-ready open-source solutions to comply with it. Some of these and other challenges were already identified by the SATO project through the questionaries and interviews from Task T1.1 (Analysis of actors, roles, and interfaces related to A&O service), which culminated into the Deliverable D1.1 (Role of Actors and Design of Stakeholder Framework), and the Task T1.3 (Requirements and System Architecture for the SATO platform and for it to support SRI calculation), which resulted in the Deliverable D1.3 (SATO Platform, SRI and IT Security Requirements). More details on how exactly the SATO platform is subdivided and what specific technologies we are going to use will be provided in the Deliverable D1.4 (Description of the system architecture of the SATO platform).

3.6. Software-Defined Infrastructures

As mentioned in the previous subsection, supporting the SRI in such a dynamic environment requires the SATO platform to be efficient and aware of the priority levels the different data may have. With this in mind, the platform may benefit from several concepts and patterns associated with modern software-defined infrastructures.

The first one is the division of communication into a control plane (e.g., management events) and a data plane (e.g., measurements). This separation directly relates with the previously mentioned priority levels and enables scheduling of resources accordingly. Note that we separate communication into these two levels, but both will require their own processing, storage, and network resources. This concept also relates with the security-by-design principle called separation of concerns [16], which classifies the components of a system into distinctive self-contained sections. Each section is concerned only with its own set of responsibilities, which usually translates into increased efficiency and maintainability for the platform in the long run. It also makes the platform more modular while promoting the "Don't Repeat Yourself" principle. Well-defined interfaces also contribute to this pattern since they hide the complexity of the underlying components.

Another related concept important to the SATO platform is the one of composable infrastructures. This concept promotes the easy deployment and configuration of scalable infrastructures that include a large amount of processing, storage, and network resources. Systematically deploying and configuring the SATO platform accelerates the preparation of the environment for all foreseen project pilots and make these testbed environments similar to real-world scenarios. Additionally, it promotes the scientific reproducibility since it will be easier for other researchers simulate the same experimental environments adopted in the SATO project.

These programmable infrastructures and configurable deployments can be seen as Infrastructure-as-a-Code [17], which is a type of IT process for infrastructure provisioning where systems are automatically built, managed, and provisioned through code rather than manually. It drastically increases the efficiency of resource usage since it enables high-scalability, fast recovery of faulty components, and the timely adapting the system to absorb high-demand periods or save money in low-demand ones [18].

The SATO platform may benefit from the easy deployment and configuration provided by using containerization (e.g., Docker [19]). This approach creates an interesting level of abstraction from the underlying software and hardware stacks and provides sufficient isolation for the components of the SATO platform. Containers facilitate solving software dependency and patch management for timely correcting bugs and vulnerabilities in the diverse components. Moreover, it reduces human errors in configuring the infrastructure and ensures the overall quality of the platform deployment [20]. Finally, embracing such a dynamic configurable environment provides cloud independence for the SATO platform and economic savings for future deployments in case of price fluctuations.

3.7. Data Lakes

Data is at the core of the SATO platform since the devices and gateways will generate large quantities of control events and measurements. A data lake is a modern approach for efficiently handling large data flows in dynamic ecosystems [21]. One of the primary ideas with data lakes is to ingest data as fast (and with the least processing) as possible for an expedite deliver where data is effectively needed. Consequently, data lakes favour schema-on-read rather than structuring data on write. It provides the benefit of making the SATO platform scheme-agnostic and future-prone since data is structured only when it is going to be processed in later steps. Minimizing the processing requirements at initial stages enables the SATO platform to focus in optimizing for low-latency data delivery and protecting the data availability in storing events as soon as possible.

After this initial data ingestion is guaranteed, the SATO platform can focus on providing data enhancements for later processing (including the calculation of the SRI scores). Examples of subsequent steps include (but are not limited to) data cataloguing, privacy protection, data quality control, and enhancing data with metadata. The main idea of these steps (summarized in Figure 4) is to improve the data quality and guarantee that the SATO platform provides trustworthy refined storage for the processing steps from the self-assessment and optimization phases [22]. All these (pre-processing) steps will comply with the previous concepts presented in this deliverable and can be performed using event-based communication.

Figure 4: A workflow example for the Data Lake approach in the SATO platform.

As soon as data is ready to be read by the processing steps, it can be structured according to the multiple frameworks and data models adopted for each goal. Postponing this structuring process once more optimizes the resource usage in the SATO platform and provides performance benefits since it structures data only when it will be used for analyses. It avoids structuring data into multiple models where some of them might be discarded later. Since data is intended to be delivered as soon (and as unprocessed) as possible, the streaming components will not become a bottleneck of the system.

Finally, the data lake approach enables the processing of the structured data, which may include data aggregation, transformation, filtering, among others. These processing tasks range from simple deterministic calculations to complex artificial intelligence. Data analyses generate data insights that will be taken into consideration when optimizing the energy-management of buildings and will later result in alerts for building managers, data visualization in dashboards, or even heuristic-based automatic actuations directly on devices. Interestingly, the data lake storage approach of the SATO platform can be seen as a large repository that will support all dataflow steps, from the data ingestion to the actuation.

Beyond the storage aspects, a data lake must include other key features like providing diverse APIs and endpoints for enabling secure controlled access to the data. External entities can also access publicly available data (e.g., any third party) or may request access to specific datasets (e.g., trusted parties). This feature is crucial since all data from the platform will be in the data lake and the APIs must cover all foreseen (and unforeseen) use cases. Search and catalogue can include metadata and tagging to provide useful features for organizing and locating data in the data lake. Access control mechanisms must guarantee that only authorized components and actors can access data and this access must be registered for future auditability.

3.8. Interoperability

Deliverable D1.3 (SATO Platform, SRI and IT Security Requirements) introduced the main requirements for interoperability and defined it as one of the key aspects the SATO project needs to focus on since it integrates heterogeneous energy management systems and devices. Only integrating these elements will make possible for the SATO project to provide the promised self-assessment and optimisation.

Based on the European Interoperability Framework (EIF), we have identified in D1.3 three important interoperability layers that directly impact the design and development of the SATO platform: the semantic interoperability, the syntactic interoperability, and the technical interoperability. Deliverable D1.2 (Requirements of the Self-Assessment Framework) advanced these requirements and started proposing some semantic interoperability concepts. In this section, we briefly recall the discussion on the main concepts associated with each of the interoperability levels and the detailed implementations will be described in the Deliverable D1.4 (Description of the system architecture of the SATO platform).

3.8.1. Semantic Interoperability

Semantics is a concept that is usually associated with the human communication and is defined as the meaning in a language. In information technologies, the semantics is essential since computing systems also rely upon languages which need to be precise and unambiguous. On the Internet of Things (IoT) context, this semantic concept can produce an interoperability problem since there are numerous different devices, each one supporting their protocols and having different representations for their characteristics and the features they offer. Even when devices are integrated through a gateway, they can provide the same features but call them differently, which makes their comparison and analysis harder. This interoperability is crucial to the IoT field due to its heterogeneous environment. It is also crucial for the SATO project since it will integrate different IoT platforms and smart energy management systems and eventually would face the mentioned challenges. The Deliverable D1.2 (Requirements of the Self-Assessment Framework) discussed many aspects of this semantic interoperability, identified an open opportunity for proposing a novel ontology related with the context of certifying the energy efficiency of buildings, and provided a KPI (Key Performance Indicators) tool that will serve as the basis for the semantic in designing and developing the SATO platform.

In the SATO platform, a solution will maintain homogeneous metadata about the context of the interconnected heterogeneous devices and platforms, increasing the usefulness and impact of the platform. Such a feature will be accomplished using open standards for the collection of concepts with precise meanings. It will enable other components to access the semantics even when the underlying devices/platforms are heterogeneous. This semantic knowledge in the computer science field is normally used in the form of an ontology, which is a way of describing concepts (sometimes named classes), the relations between them, their properties, features, and attributes.

When analysing the SATO platform requirements from Deliverable D1.3 regarding the semantic interoperability using ontologies, it can be taken into consideration the RESPOND project which is also associated with IoT for energy optimization domain, where an ontology was developed to provide a shared understanding of data. This ontology is available online [23] and may serve as a starting point to help solving the requirement of semantic interoperability.

More specifically, the SATO platform will provide two semantic services for metadata and control events associated with the devices and their location within the building. As stated before, these services will manage and provide this information according to the ontology definition. These components will abstract the heterogeneous environment of the metadata for platform components that need to consult it. Interaction with these services will follow the standard defined by the adopted and proposed ontologies. Additionally, modules that handle the platforms/devices will need to support the mapping of the metadata to the one supported in the SATO platform through plugin connectors.

One of the most known solutions for Semantic Web and Linked Data applications is the open-source framework Apache Jena [24], which provides different solutions for developing semantic components that may directly benefit the SATO project.

3.8.2. Syntactic Interoperability

The integration of the heterogenous environment on the Internet of Things domain will also cause syntactic interoperability problems since the services responses and events from each device or platform might have their own data models. Such heterogeneity hinders the seamless usage of this data for the remaining components. To overcome this limitation, the main solution will be the establishment of a Common Data Model (CDM) that will provide a standard representation of the data. This CDM-compliant data will flow from the integrated platforms and devices to the SATO components, avoiding the numerous mappings needed if no CDM was used, which would also hinder the integration of new devices or platforms.

In the context of the SATO project, defining this CDM is crucial since it will hide the complexity of the underlying platforms and devices different data models for the self-assessment and optimization components of SATO platform. Events and service responses in the platform will be structured only according to the defined CDM.

3.8.3. Technical Interoperability

Technological interoperability is an abstract concept that focuses on the seamless interaction of a set of different components in a system. It is achieved by adopting standard approaches that enable these components to work together and achieve their global purpose, collaborating with each other even if each one has a different specific goal or uses diverse internal mechanisms.

Regarding the SATO platform, an example that requires technological interoperability is the integration of the underlying IoT platforms, smart systems, and devices into a coherent vision of all the entities that compose the building energy management ecosystem. In this case, there will be a solution that communicates with all these underlying devices and adapt to their technologies abstracting them to the rest of the system, making their services available homogeneously. In addition, when considering the SATO platform, one should consider its technology of deployment since it should not be using unique method of deployment. Opting for technologies that provide decoupled deployments, being the private cloud or public cloud, since external platforms may be plugged into the SATO platform and provide data in diverse ways and location that need to be standardized by the SATO connectors (i.e., plug-ins for producing and consuming events from the platform in standardized ways).

Regarding the methodologies used to solve technological interoperability, they can vary given the problem at hand, but there are common possible solutions. For instance, having a single component that contains most of the connectors necessary to abstract the complexity of the different technologies may cause some drawbacks by becoming a single point of failure or a bottleneck. Another methodology would be to use smaller components, each one handling small parts of the technological interoperability. This approach solves the single point of failure and bottleneck, but its feasibility depends on the specific interoperability problems it tries to solve.

It is of most importance to the SATO platform to solve its technological interoperability since it will integrate a large variety of components. Each component has its purposes, which together they should accomplish its global goal. To achieve this goal, all these components should intercommunicate even if they use different technologies. FIWARE [25] and OGEMA [26] are examples of technical platforms for interconnecting heterogeneous devices and external platforms. Their recommendations and best practices will be adopted in the SATO project for this goal.

4. Final Remarks

This deliverable identified the main concepts that will enable the SATO platform to support the dynamic self-assessment of the Smart Readiness Indicator (SRI) and the subsequent self-optimization towards more energy-efficiency and comfort of building’s occupants.

We first defined the SRI, discussed its benefits, showed some existent initiatives that implement it, and presented open opportunities the SATO project may focus in addressing to handle the dynamic nature of the SRI. We then identified and discussed seven concepts that help the SATO platform to support the SRI in more effective and dynamic ways than it is supported by existing works in the literature. These concepts are the systematic construction of the building inventory, automatic calculation of the SRI preliminary scores, security and privacy mechanisms, event-based communication, software-defined infrastructures, data lakes, and interoperability (semantical, syntactic, and technical).

The concepts presented in this deliverable consider the use case requirements from previous deliverables and will guide the design and development tasks in WP2 of the SATO project. Deploying, configuring, and integrating the SATO platform with the project pilots from the SATO project will be easier and more efficient if using the previously described concepts.

5. Bibliography

[1] European Parliament, Council of the European Union, «Directive (EU) 2018/844,» 30 May 2018. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2018/844/oj. [Consultato il giorno 2018].

[2] European Commission, Directorate-General for Energy, *COMMISSION DELEGATED REGULATION (EU) 2020/2155,* THE EUROPEAN COMMISSION, 2020.

[3] D. Ö. Thomas Märzinger, «MDPI,» 22 May 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/10/1955/htm.

[4] R. P. G. P. A. G. Ilaria Vigna, «MDPI,» 1 June 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/11/2796/htm.

[5] D. Ö. Thomas Märzinger, «MDPI,» 7 June 2020. [Online]. Available:

https://www.mdpi.com/1996-1073/13/13/3507.

[6] A. Ożadowicz, «A Hybrid Approach in Design of Building Energy Management System with Smart Readiness Indicator and Building as a Service Concept,» *Energies,* vol. 15, n. 4, p. 1432, 2022.

[7] «SAREF,» ETSI, 29 5 2020. [Online]. Available: https://saref.etsi.org/core/v3.1.1/.

[8] schwackenberg, «EEBus SPINE Technical Report,» Cologne,, 2018.

[9] OWASP Foundation, «OWASP/DevGuide,» [Online]. Available: https://github.com/OWASP/DevGuide.

[10] Microsoft, «Docs.microsoft,» Microsoft, 13 1 2021. [Online]. Available: https://docs.microsoft.com/en-us/dotnet/architecture/microservices/multi-container-microservice-net-applications/integration-event-based-microservice-communications.

[11] «Apache Kafka,» [Online]. Available: https://kafka.apache.org/.

[12] «RabbitMQ,» [Online]. Available: https://www.rabbitmq.com.

[13] «Eclipse Mosquitto,» [Online]. Available: https://mosquitto.org/.

[14] M. S.-S. R. P. Dmitry Namiot, «Researchgate,» 22 Oct 2018. [Online]. Available: https://www.researchgate.net/publication/327936728_On_Data_Stream_Processing_in_IoT_Applications_18th_International_Conference_NEW2AN_2018_and_11th_Conference_ruSMART_2018_St_Petersburg_Russia_August_27-29_2018_Proceedings.

[15] J. Yates, «Stream Processing with IoT Data: Challenges, Best Practices, and Techniques,» 4 June 2020. [Online]. Available: https://www.confluent.io/blog/stream-processing-iot-data-best-practices-and-techniques/.

[16] J. Spacey, «Simplicable,» 27 12 2016. [Online]. Available: https://simplicable.com/new/separation-of-concerns.

[17] C. Schults, «What Is Infrastructure as Code? How It Works, Best Practices, Tutorials,» 5 9 2019. [Online]. Available: https://stackify.com/what-is-infrastructure-as-code-how-it-works-best-practices-tutorials/.

[18] Ionos, «Infrastructure as Code (IaC): using code for IT infrastructure management,» 14 10 2019. [Online]. Available: https://www.ionos.com/digitalguide/server/know-how/infrastructure-as-code/.

[19] «Docker,» [Online]. Available: https://www.docker.com/.

[20] A. Patel, «Configuration Management and Continuous Deployment,» 12 4 2018. [Online]. Available: https://medium.com/formcept/configuration-management-and-continuous-deployment-cd0892dce998.

[21] Stitch, «What is a Data Lake? Examples & Solutions,» [Online]. Available: https://www.stitchdata.com/resources/what-is-data-lake/.

[22] J. Harris, «The growing importance of big data quality,» 21 11 2016. [Online]. Available: https://blogs.sas.com/content/datamanagement/2016/11/21/growing-import-big-data-quality/.

[23] I. Esnaola, F. J. Diez, N. Tomasevic e M. Batic, «RESPOND D4.1 Semantic Information Model,» 2019.

[24] The Apache Software Foundation, "Apache Jena," [Online]. Available: https://jena.apache.org/. [Accessed 30 March 2021].

[25] FIWARE Foundation, «FIWARE - The Open Source platform for our smart digital future,» 2021. [Online]. Available: https://www.fiware.org/.

[26] Fraunhofer, «OGEMA - Open Gateway Energy Management Alliance,» [Online]. Available: https://www.ogema.org/.

[27] Y. M. P. V. T. S. B. V. G. O. S. E. P. W. :. K. B. J. A. A. H. M. O. J. G. :. M. U. J. S. Stijn Verbeke, «SUPPORT FOR SETTING UP A SMART READINESS INDICATOR FOR BUILDINGS AND RELATED IMPACT ASSESSMENT,» European Commission, 2018.