Knowledge Representation of Time Series Data: A Comparison Analysis of Standardized Ontologies

João Moreira[1,*,†], Cornelis Bouter[2], Laura Daniele[2], Mateus Peixoto[3] and Marcos Machado[1]

[1]University of Twente, Drienerlolaan 5, Enschede, 7522 NB, The Netherlands
[2]TNO, Netherlands Organization for Applied Scientific Research, The Netherlands
[3]Pontifical Catholic University of Rio de Janeiro, Rua Marquês são Vicente 225, Rio de Janeiro, 22451-900, Brazil

Abstract

This paper explores the growing relevance of using ontologies to represent time series data for effective knowledge management across various domains and for different purposes. Time series data, fundamental in fields like statistics, econometrics, and healthcare, are typically stored without semantic context. Ontologies address this gap by providing a structured way for integrating time series data with contextual knowledge, improving semantic interoperability and enabling more accurate analysis and predictions. In this paper we revisit two well-established extensions of the Smart Applications REFerence Ontology (SAREF) standard. We describe how each extension addresses different types of requirements on time series representation for e-health and aging well (SAREF4ehaw) and for energy flexibility (SAREF4ener). In this paper we provide a comparison analysis between these approaches through a discussion on their trade-offs. Through case studies on electrocardiogram data exchange and energy forecasting, we compare their advantages and limitations, and we recommend further research to enhance semantic interoperability for time series analysis.

Keywords

Semantic Interoperability, Ontology, Time Series data, SAREF

1. Introduction

The use of ontologies to represent time series data is increasingly relevant due to the growing need for effective knowledge management across various domains [1]. Fields such as healthcare, and energy flexibility demonstrate a critical need to integrate and combine time series data across organizations to enable effective knowledge management and interoperability. Time series, which consist of sequences of data points recorded at regular intervals, are fundamental in such fields where they are used to uncover patterns or forecast future trends [2]. Traditional data formats and specialized time-series databases are commonly used for storing such data, but they lack the ability to provide semantic context. Ontologies offer a structured approach to bridge this gap by enabling the formal representation of time series within a broader knowledge framework [3]. This enhances interoperability, allowing data to be shared, interpreted, and reused across different systems and disciplines. By integrating time series data with ontologies, numerical models can be connected with relevant contextual information and reused across organizational boundaries if needed, ensuring accurate interpretation and explanation, and supporting more reliable predictions [4].

In this paper we discuss some existing initiatives on ontology engineering for time series analysis, highlighting the importance and the rise of research results in this topic in recent years. We give emphasis for extensions of the Smart Applications REFerence (SAREF) ontology[1], which is standardized by the European Telecommunications Standards Institute (ETSI). We discuss how two extensions address different types of requirements for time series representation in the domains of e-health and

EKAW 2024: EKAW 2024 Workshops, Tutorials, Posters and Demos, 24th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2024), November 26-28, 2024, Amsterdam, The Netherlands.

*Corresponding author.
j.luizrebelomoreira@utwente.nl (J. Moreira)
0000-0002-4547-7000 (J. Moreira); 0000-0002-5448-0543 (C. Bouter); 0000-0002-9267-7160 (L. Daniele); 0009-0005-4081-6615 (M. Peixoto); 0000-0003-1056-2368 (M. Machado)

[1]This work is based on SAREF core version 3.1.1 available at https://saref.etsi.org/core/v3.1.1/

aging well (SAREF4ehaw)[2] and energy flexibility (SAREF4ener)[3]. In SAREF4ehaw we emphasized the non-functional requirements on IoT data exchange performance motivated by an use case on electrocardiogram data, where we measured the verbosity of messages sent from field to cloud gateways along with its associated costs. In SAREF4ener we emphasized the functional requirement on demand-side flexibility, using time series data to track and predict variations in energy consumption over time. The main contribution of this paper is the comparison analysis of these approaches through a discussion, highlighting their trade-offs in terms of their advantages and limitations. In addition, we provide a set of recommendations on the adoption of the time series concept in ETSI SAREF, proposing further research to enhance semantic interoperability for time series data analysis.

This paper is structured as follows. Section 2 motivates the research on knowledge representation of time series data, discussing existing work. Section 3 describes how time series is designed in the SAREF4ehaw ontology for electrocardiogram data exchange. Section 4 describes how time series is designed in the SAREF4ener ontology for demand-side flexibility analysis. Section 5 discusses the comparison between both approaches, highlighting the lessons learned, limitations and recommendations. Section 6 concludes this paper.

2. Knowledge representation of time series data

2.1. Time series analysis

A time series is a sequence of data points ordered in time, typically recorded at regular intervals, i.e. within a frequency [5]. It is commonly visualized using run charts and analyzed in fields like statistics, econometrics, and weather forecasting to uncover patterns or predict future values. Time series analysis involves methods to extract meaningful patterns and statistics from time-ordered data, often modeled as a stochastic process. Time series forecasting predicts future values based on past observations. Unlike regression analysis, which tests relationships between different series, time series analysis focuses on relationships within a single series over time. It differs from cross-sectional and spatial analysis due to its natural temporal order, where data points closer in time are more related. Models typically reflect this by expressing current values as influenced by past (not future) observations [6].

In practice, the most common data formats for time series analysis include CSV, spreadsheets (e.g., Excel files), besides JSON and Database Management Systems (DBMS). CSV is widely used due to its simplicity and compatibility with most data analysis tools. JSON is popular for web-based data exchanges and can represent nested time series data. DBMSs, particularly time-series databases like InfluxDB and TimescaleDB, are optimized for handling large-scale time-stamped data, allowing efficient storage and querying. These formats are chosen for their ease of use, flexibility, and integration with various time series analysis software [7].

Demand forecasting is one of the most common real-world applications for time series, forecasting is paramount since it is the main input of sales and operations planning, (S&OP) [8] and therefore forecasting plays a pivotal role in inventory management. Nevertheless, demand forecasting as an activity is typically treated separately from the decision-making processes that derive from it, often affecting inventory management [9]. The critical nature of forecasting calls for an integrated approach as criticized by [2] in a recent systematic literature review, the fully integrated approach suggested as a framework would require a good comprehension of the demand-generating process, therefore associating explanatory events and context to the event log represented by the time series.

Kolassa et al. [10] proposed an ideal forecaster framework, establishing that forecasters must understand the statistical models applied to the time series and the business goals surrounding the demand for the products represented by the time series. This is a challenge since managers typically do not have the statistical knowledge to understand the models, leading to overconfidence bias and poor adjustments [11]. These judgmental adjustments, defined by [12] as altering the output of a computer-generated

[2]This work is based on SAREF4ehaw version 1.1.1 available at https://saref.etsi.org/saref4ehaw/v1.1.1/
[3]This work is based on SAREF4ener version 1.2.1 available at https://saref.etsi.org/saref4ener/v1.2.1/

forecast by an expert with domain knowledge, can have widespread consequences within the decision-making process of the entire organization. Therefore an appropriate standardized ontological structure for time series is the first step to conciliate both aspects, providing managers with the necessary statistical knowledge, and making explicit the context and consequences of adjustments, thus leading to improvements in S&OP.

Representing time series data with ontologies is mostly relevant for effective knowledge management, as it provides a structured way to organize, share, and reuse data across different systems and domains. Ontologies can enable the connection of numerical modeling with the context of the data, ensuring that time-dependent measurements are properly interpreted and integrated. This is particularly important in fields that rely heavily on complex datasets, such as climate science or healthcare, where numerical models are applied to predict trends or behaviors. By using ontologies, time series data becomes more interoperable, enhancing the accuracy of modeling, the reliability of measurement, and the efficiency of knowledge management across various applications [1].

2.2. Ontologies of time series

This section explores some related work on the use of ontologies to represent time series data. We give emphasis to standardized ontologies, but we acknowledge the relevance of other non-standardized ontologies that cover gaps of existing standards.

The first effort to represent time series in the context of SAREF was done in 2017-2020 in the health domain to express electrocardiogram data [3]. The need for time series originated since SAREF could only represent individual measurements (data points) that would result in too large and redundant messages and thus poor performances in real-time exchange of large amounts of measurement data. We will elaborate on this approach in Section 3 of this paper. This work resulted in an extension coined SAREF4health, which was afterwards used as basis by ETSI to create the official extension of SAREF for the E-Health and Aging Well domain, called SAREF4ehaw and published in 2020 as TS 103 410-8 V1.1.1.

Subsequent efforts in the H2020 InterConnect large-scale pilot[4] (2019-2024) aimed at extending the time series proposed in SAREF4ehaw to the energy domain for the purpose of representing forecasts and expressing power curves with corresponding prices in the data exchange between energy users and suppliers. This work was incorporated into the SAREF framework and resulted in the publication by ETSI of a new version of SAREF4ener in 2023 (TS 103 410-1 V1.2.1) which also covers a time series representation. We will revisit this approach in more details in Section 4 of this paper.

Recently, ETSI has published the technical specification on SAREF covering time series for urban digital twins (ETSI TS-103-828) [13]. It leverages on reusing existing ontologies such as the W3C Time and the OneM2M ontologies to cover the concepts of time series and service. The W3C Time Ontology is based on Allen's algebra of binary relations and provides a vocabulary for representing qualitative temporal information and reasoning about time. The temporal predicates for times series proposed by ETSI TS-103-828 overlaps with the ones covered by SAREF4ehaw since they are both based on the same theoretical foundations, i.e., perdurantism and Allen's algebra.

Other ontology-driven initiatives also cover the representation of time series data. For example, an ontology for time series provenance is introduced in [4]. It presents a domain ontology developed by modular design for time series provenance, which adds semantic knowledge and contributes to the choice of appropriate statistical methods for trend extraction (detrend) within time series analysis. Trend is an interesting time series component that can reveal deterministic features, e.g., statistical measures of mean and variance, used to ensure that correlations become independent over time. In a similar way, a time series core ontology is also introduced in [14], with emphasis on the integration of this core ontology with domain specific ontologies. There is a clear rise on the interest of ontologies for time series data representation. This is shown from the results of the search in Google Scholar for the terms "time series data" and "ontology": 537 results from 2000-2005, 1.770 from 2005-2010, 2.980 from 2010-2015, 4.990 from 2015-2020, and 7.300 from 2020-2024 (present day). This also drives to the need of a systematic literature review on ontologies to represent time series, which is out of the scope here.

[4]https://cordis.europa.eu/project/id/857237

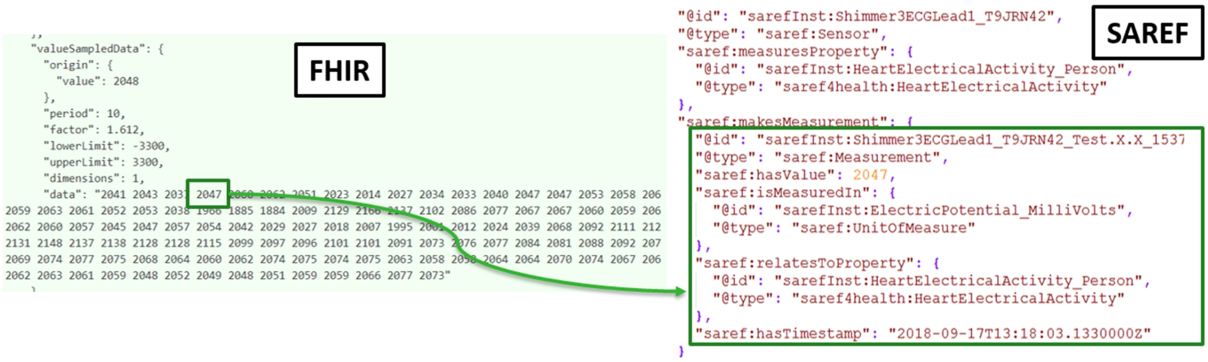

Figure 1: Example of time series data represented with FHIR (left) and with SAREF (right), from [3]

3. Time series data exchange in SAREF for E-Health Ageing Well

In recent years, several ontology-driven e-health solutions supported by IoT technologies have been proposed. In our previous research [3, 15, 16] we exploited knowledge representation of time series driven by a use case on detecting accident risks with truck drivers' vital signs with electrocardiogram (ECG) medical wearable, using the drivers' mobile phones as field gateway. Tracking cardiac-related data is a common necessity in healthcare solutions. We focused on IoT-driven ECG devices that can provide the necessary programming support, e.g., via API or SDK, to allow the connection of the ECG device and real time transmission of high-frequency data to a cloud environment.

3.1. Time series data exchange for IoT applications

Enabling the exchange of lightweight messages among medical devices and cloud infrastructure is crucial for IoT solutions. However, message verbosity can hinder data exchange performance and increase cloud costs. For example, with Microsoft Azure, each IoT hub (message broker) is allocated units within a specific tier, which along with the number of units, establishes the maximum daily message quota. Message size plays a role in calculating this quota, as well as throttling, which is the rate limit imposed by the cloud infrastructure to prevent abuse on message publishing, and is influenced by traffic shaping. Examples of throttles are direct methods, measured in total size (KB) per second per hub unit, identity registry operations (CRUD) and twin updates, which are both measured in number of messages per second per hub unit.

A common solution to minimize the costs associated to throttling is to aggregate measurement data at the field gateway level, periodically (in a lower frequency) transmitting aggregated time series data to the cloud. However, this approach results in the loss of metadata for individual measurements, for example specific timestamps, measurement units, and related properties, which diminishes the ontological richness of the messages. Figure 1 illustrates this difference between the aggregated approach, shown in the left using the Fast Healthcare Interoperability Resources (FHIR) standard, to the individual-based approach, shown in the right using SAREF. The data are serialized in JSON (SAREF particularly in JSON-LD) and reflect an ECG time series, where FHIR represents the aggregation within the *data* field as a sequence of float numbers, while SAREF represents each individual measurement.

In [3], we introduced the SAREF4health extension, a preliminary version of SAREF4ehaw, which was designed to address the aforementioned verbosity problem. Our study emphasizes four key characteristics of semantic models: quality, message size, IoT orientation, and standardization, ultimately combining ontology-driven conceptual modeling with other initiatives. We investigated a number of alternatives, and used a well-founded ontology for Electronic Health Record (EHR) that focused cardiac-related data, which was developed through an analysis of existing health standards and supported by the ontology-driven conceptual modeling practice.

3.2. Extending SAREF for data exchange of ECG time series measurements

As discussed in the previous section, instantiating ECG data with SAREF results in significantly larger message sizes compared to other standardized alternatives . For instance, our experiments of a JSON-LD message comparing SAREF individual measurement to FHIR aggregated approach showed that the SAREF message was fifty times larger. We generated real data with an ECG device at 256Hz, and 1280 measurements every 5 seconds in our experiments. As a result, the message size is 5MB with SAREF, while with FHIR the message size is 100KB [3]. This size discrepancy increases exponentially with the number of measurements. Consequently, we conclude that the core SAREF ontology is unsuitable for exchanging IoT-based ECG time series data due to its excessive message size. We examined a number of semantic models, assessing their strengths and weaknesses considering characteristics of ECG data, including the components involved in ECG recordings. Our study indicated that SAREF is one of the most fitting IoT ontologies, and we concluded that there was a gap in standardized IoT ontologies that adequately balance quality and payload for representing ECG time series data. This realization prompted the development of SAREF4Health as an extension of SAREF to address these needs.

We introduced the term *Time Series Measurements* in SAREF4Health to reflect the concept of "sample sequence" or "sampled data" that existing standards refer to time series data. This element can be used to represent ECG data, i.e., an array of electric potential measurements related to heart activity. We classified *Time Series Measurements* as a type of *Measurement* in SAREF, adhering to definitions and reusing existing structures for properties like *hasTimestamp* and *isMeasuredIn*, and introducing the *hasValues* property to accommodate multiple float values as an ordered sequence, inspired by the notion of *Lists* in JSON-LD . Additionally, we incorporated frequency information by reusing the frequency measurement property from SAREF4envi extension. The frequency of an ECG device can typically be set via an API, reflecting the sampling frequency of each ECG sample sequence collected during a recording session. This helps to differentiate the device's current frequency from those used in previous sample sequences.

Therefore, we define the **Time Series Measurement**[5] term as a sequence of data in a successive equally spaced points in time. The OM ontology (ISO 19156) defines *Time Series Observation* as an "observation whose result is a time-series", while standards like Hl7 aECG and DICOM define the *Series* element as "a sequence of data sharing a common frame of reference". Here this concept is termed as *Time Series Measurements* since this sequence of data refers to time series measured by an IoT device. We validated this approach first by responding to competency questions on a simple ECG data exchange use case[3], and through a working prototype executed in a number of test cases [16]. The results show that the trade-off between ontology quality and lightweight data serialization is a relevant aspect on the design of the time series ontology. From this research outputs, the ETSI SAREF4ehaw task force incorporated the main elements of SAREF4health into the standard[6].

4. Forecasting time series in SAREF for Energy Flexibility

The representation of time series for monitoring, forecasting and optimizing power consumption and production of energy related devices, like Heating, Ventilation, and Air Conditioning systems, Photovoltaic (PV) panels and Electric Vehicles (EV), is a complex topic that requires modeling of sequences of data points linked to contextual data about grid capacity, storage capabilities (e.g., batteries) and weather conditions. Forecasting plays a major role in this context, for example (1) to predict when the sun will be shining in the coming hours or days and, therefore, significant production from PV panels is expected with a consequent low usage of the power grid; (2) to forecast prices in the energy market; and/or (3) to predict behavioral patterns of consumers, such as when most people will be likely to charge their EVs, use hot water, do their laundry, or use air conditioning. From a conceptual point of view, a forecasted data point can be modelled as a "measurement" or an "observation". The OGC

[5]https://saref.etsi.org/saref4ehaw/v1.1.1/#s4ehaw:TimeSeriesMeasurement

[6]It is relevant to highlight that while in SAREF4health we introduced the property *hasValues* with a range of an array of *xsd:float*, in SAREF4ehaw the range was set as *xsd:decimal*

O&M takes the position that the term "measurement" can be reserved for numerical values, but that both "measurement" and "observation"can be applied on forecasted values, as long as there is metadata available to indicate the specific usage [17].

4.1. Time series data for energy forecasting

Europe leads the digital energy transition, aiming for climate neutrality by 2050 through emissions cuts, green tech investments, and environmental protection. Key priorities include affordable energy access, energy independence, local renewable energy production, and data integration across stakeholders to drive this transformation. The H2020 Interconnect project was carried out in this context with the aim to develop and demonstrate energy services for connecting and converging digital homes and buildings with the electricity sector, using SAREF as main pillar. To be able to model these services and support the unambiguous exchange of measurement and prediction data across different platforms, systems and countries, a number of extensions of SAREF have been developed in the project in close collaboration with more than fifty industry and research stakeholders in various workshop over two years[7]. The results were then validated, tested and deployed in seven connected large-scale test-sites in Portugal, Belgium, Germany, the Netherlands, Italy, Greece and France. Given that the majority of the services involved monitoring, forecasting and optimizing energy usage, extensions of SAREF for data points and time series[8] and for forecasts[9] have been created.

These extensions define the following main concepts: (1) a data point as an atomic piece of information about a certain observable quantity in nature that can contain a numerical value and a corresponding unit of measure; (2) a time series as an ordered sequence of data-points of a quantity that is observed at spaced (not necessarily equally spaced) time intervals; and (3) a clear characterization of the most common forecasting data that reuses data points and time series [10]. When designing forecasts, we incorporated the need to distinguish between point forecasts versus stochastic forecasts, as well as the various ways to express stochastic forecasts. We also consulted standardization bodies such as ETSI and CEN/CENELEC, and the result was a commonly agreed new version of SAREF4ENER that was published in 2023 that allows to model data points, time series and forecasts[11].

4.2. Extending SAREF for demand-side flexibility analysis

The SAREF4ener standard resulted from the InterConnect project is based on two energy flexibility standards, EN 50631 [18] and EN 50491-12-2 [19]. The EN 50631 on "Performance of household and similar electrical appliances" defines the information exchange between smart appliances and energy management systems in homes and buildings. The EN 50491-12-2 on "General requirements for Home and Building Electronic Systems and Building Automation and Control Systems" defines a communication standard for energy flexibility and energy management, which helps to optimize the use of energy of smart devices in homes and buildings. Both [18, 19] require the modelling of time series for the actual energy monitoring of a device and also for forecasting functionalities. The design of this time series incorporates these definitions, which are illustrated in Figure 2. Besides SAREF, it also follows the design choices made in the extension to the SSN ontology [20].

A *DataPoint* is modeled in [22] as a subclass of *saref:Measurement* and, as such, inherits the *saref:hasValue* and *saref:hasTimestamp* properties. Therefore, if the combination of a numerical value and timestamp is sufficient to represent a *DataPoint*, then the SAREF concepts for measurement can be directly reused and the time series is modeled as a container and the measurements are its elements. We highlight these additional properties to represent forecasts as *TimeSeries* of *DataPoints*:

---

[7] https://gitlab.inesctec.pt/groups/interconnect-public/-/wikis/home#interconnect-ontology
[8] https://gitlab.inesctec.pt/interconnect-public/ontology/-/wikis/ic-data
[9] https://gitlab.inesctec.pt/interconnect-public/ontology/-/wikis/ic-fc
[10] An OWL example to model an instance of an energy forecast with corresponding price is available at https://gitlab.inesctec.pt/interconnect-public/ontology/-/blob/master/examples/flex_offer_example.owl
[11] The detailed distinction of different forecasts was not included in the standard for the purpose to stimulate further validation and discussion in the community

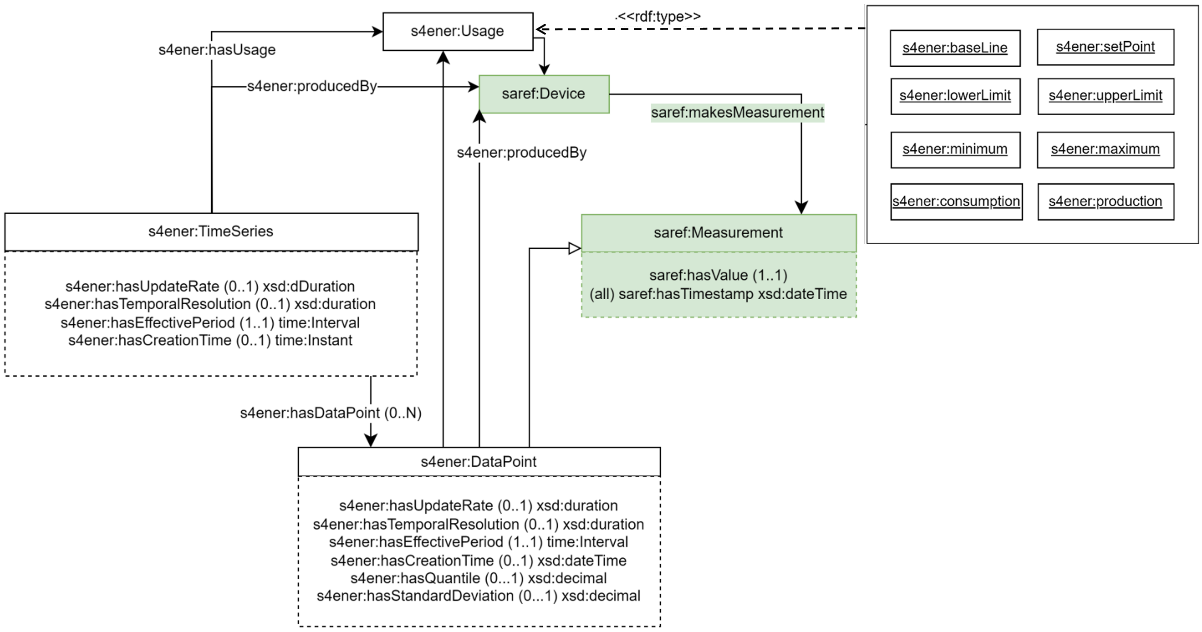

Figure 2: Chowlk [21] diagram of the main SAREF4ENER V1.2.1 time series classes

Creation time: The time instant in which a data point or time series has been created. The creation time differs from the time at which the quantity is in effect, which is expressed by the hasEffectivePeriod property. For example, if a temperature is forecasted on 19-11-2024 at 12:30 (i.e., the creation time of the forecast) for the following day between 14:45 and 15:45 (i.e., the time when the temperature is expected to be in effect), then the creation time is 12:30 of 19-11-2024.

Effective period: This connects to the temporal entity which describes when (time interval) the quantity of this data point was, is, or will be in effect. This is the time interval which is covered by the forecast that in our example begins on 20-11-2024 at 14:45 and ends at 15:45.

Temporal resolution: The distance between two data points measured at different times, which makes sense if the measured data points in a time-series are equidistant in time. For example, the temporal resolution of the forecasted temperature example is 1 hour, i.e., the difference between the end time (15:45) and start time (14:45) of the effective period of each data point in the time-series.

Update rate: The rate at which a data point or time-series or forecast is being updated.

Usage: The purpose (usage) for which the data-point or time-series is used, for example as an upper limit, lower limit or a baseline (i.e., expected value), a maximum versus minimum value, or an energy consumption versus an energy production value.

Produced By: the origin (or provenance) at which the datapoint or timeseries are produced.

5. Discussion

The first point for discussion, on which SAREF4ener and SAREF4ehaw take different approaches, is how to model the ordering of time series elements. The ordering of elements could be left to the specific RDF serialization, as SAREF4ehaw does, by expecting to use the JSON-LD serialization format. RDF implements ordered lists by linking anonymous nodes with *rdf:first* and *rdf:rest*, ending with *rdf:nil* to represent order, requiring a precise (but verbose) way to represent ordered lists [23]. JSON-LD offers a shortcut for ordered lists trough the *@list* annotation. The SAREF4ener extension expects a user to explicitly include a timestamp for each data point associated to the time series. This ensures the validity of the data, but it remains computationally expensive to query for immediately following or preceding elements, in particular on high-frequency time series.

Capturing the timestamp of time series elements is done implicitly with SAREF4ehaw by recomputing the timestamp from the list index, starting time, and measurement frequency. By using explicit timestamps, as SAREF4ener does, data can be shared more reliably across systems where consistency in time-based data alignment is necessary. This explicit approach can be advantageous to mitigate the risk of misinterpreting temporal data sequences — a significant consideration for energy or health-related infrastructures. On the other hand, the implicit timestamp approach taken by SAREF4ehaw prioritizes storage efficiency, which may be beneficial for systems that do not expect variations on the time intervals between data points, and with limited storage and/or bandwidth capacity or where the overhead of recomputing timestamps is manageable.

From a business management perspective, each approach also aligns with different operational priorities. The explicit timestamping method of SAREF4ener may be preferable in applications where real-time data accuracy is crucial, as seen in energy management systems where accurate decision-making can drive cost savings and system resilience. Conversely, the implicit timestamping of SAREF4ehaw may be more suitable for applications focused on long-term data retention at minimal cost, making it a potentially viable option for budget-constrained analytics platforms. Therefore, the choice between these approaches has implications not only for technical implementation but also for aligning with the broader business goals and resource constraints of the organization deploying these models.

From a standardization perspective, it would be beneficial to harmonize the various efforts existing in ETSI on time series not to hinder adoption by stakeholders. To that end, validation and testing of the chosen time series representation in operational environments with real data is essential. This was done on different scales for SAREF4ener and SAREF4ehaw. However, no validation nor testing of the representation suggested in TS-103-828 has been reported and, therefore, it should be addressed. We also suggest to take other alternatives in consideration, and that these are driven by end users' purposes and their associated (functional and non-functional) requirements. We also argue that, in a standard, a trade-off should be reached to keep the time series representation simple and ease of use for industry practitioners, who are typically not ontology experts, while still capturing its essential aspects. As mentioned in Section 2, several research proposed ontologies to represent time series, and, therefore, a systematic literature review should be performed to learn and reason on general characteristics of them.

Given the importance of time series and their wide-spread usage in the IoT industry, we finally encourage ETSI, who has been at the forefront of semantic modeling of time series, to actively engage in public consultations with stakeholders and other interested standardization bodies for the validation of (intermediate) results on an ontology for time series, in the same iterative, open and inclusive fashion that characterized the development of the first version of SAREF [24]. The support of the European Commission in facilitating this process for SAREF has already been proven as greatly beneficial and could benefit once again the development of a shared consensus on the representation of time series.

6. Conclusion

In this paper, we have highlighted the increasing relevance of ontologies for representing time series data, particularly in fields where accurate data interpretation is crucial as healthcare and energy management. By comparing two standardized ontology-based approaches — SAREF4ehaw and SAREF4ener — we discussed their potential for improving semantic interoperability in time series analysis. Our research shows that, although both ontologies offer valuable contributions, they also have limitations, particularly regarding message verbosity and handling complex time series datasets.

The main lessons learned from this research underscore the importance of balancing ontology quality with practical concerns, which are typically described as non-functional requirements, like data transmission efficiency, especially in systems that use IoT technologies. While SAREF4ehaw offers a way for representing time series for real-time IoT data exchange with cloud environments, it is clear that the lack of expressiveness of each data point in the time series poses a challenge in knowledge-intensive applications. The SAREF4ener extension takes the opposite approach of including all data elements in the model, but this may bloat the triple store or harm computational efficiency.

Future work should focus on addressing these challenges by further reusing and refining ontologies. This includes exploring new ways to reduce message size without sacrificing the semantic richness necessary for complex applications. We discuss these and other recommendations, such as improving standardization through stakeholder consultations from different domains, which should be taken into account by standardization bodies to create consensus on time series representation. Additionally, conducting a systematic literature review on ontologies for time series representation would provide a deeper understanding of emerging trends and gaps in the field.

Finally, many machine learning models have been developed based on time series data, requiring proper explanations for improved trust, which is one of the main topics within explainable AI (XAI). Integrating ontology-driven XAI techniques into this research could also enhance the interpretability of machine learning models, making it easier to explain temporal relationships and model decisions to end-users. By prioritizing explainability, researchers can ensure that these advanced ontologies support transparent decision-making, especially in critical domains like energy and healthcare.

Acknowledgments

This work is part of the HEDGE-IoT project (https://hedgeiot.eu) which has received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101136216. It also includes results of the InterConnect project (https://interconnectproject.eu/) which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 857237.

References

[1] M. A. Farahani, M. McCormick, R. Gianinny, F. Hudacheck, R. Harik, Z. Liu, T. Wuest, Time-series pattern recognition in smart manufacturing systems: A literature review and ontology, Journal of Manufacturing Systems 69 (2023) 208–241. URL: https://www.sciencedirect.com/science/article/pii/S0278612523000997. doi:https://doi.org/10.1016/j.jmsy.2023.05.025.

[2] T. E. Goltsos, A. A. Syntetos, C. H. Glock, G. Ioannou, Inventory – forecasting: Mind the gap, European Journal of Operational Research 299 (2022) 397–419. URL: https://linkinghub.elsevier.com/retrieve/pii/S0377221721006500. doi:10.1016/j.ejor.2021.07.040.

[3] J. Moreira, L. Pires, M. van Sinderen, L. Daniele, M. Girod-Genet, Saref4health: Towards iot standard-based ontology-driven cardiac e-health systems, Applied ontology 15 (2020) 385–410. doi:10.3233/AO-200232.

[4] L. de Souza., M. S. Marcon Gomes Vaz., M. Sfair Sunye., Domain ontology for time series provenance, in: Proceedings of the 16th International Conference on Enterprise Information Systems - Volume 1: ICEIS, INSTICC, SciTePress, 2014, pp. 217–224. doi:10.5220/0004886502170224.

[5] T. chung Fu, A review on time series data mining, Engineering Applications of Artificial Intelligence 24 (2011) 164–181. URL: https://www.sciencedirect.com/science/article/pii/S0952197610001727. doi:https://doi.org/10.1016/j.engappai.2010.09.007.

[6] T. Warren Liao, Clustering of time series data—a survey, Pattern Recognition 38 (2005) 1857–1874. doi:https://doi.org/10.1016/j.patcog.2005.01.025.

[7] E. Duarte, D. Gomes, D. Campos, R. L. Aguiar, Scalable architecture, storage and visualization approaches for time series analysis systems, in: S. Hammoudi, C. Quix, J. Bernardino (Eds.), Data Management Technologies and Applications, Springer International Publishing, Cham, 2020, pp. 59–82.

[8] A. M. Tavares Thomé, L. F. Scavarda, N. S. Fernandez, A. J. Scavarda, Sales and operations planning: A research synthesis, International Journal of Production Economics 138 (2012) 1–13. URL: https://linkinghub.elsevier.com/retrieve/pii/S0925527311004907. doi:10.1016/j.ijpe.2011.11.027.

[9] A. Syntetos, M. Babai, J. Davies, D. Stephenson, Forecasting and stock control: A study in a

wholesaling context, International Journal of Production Economics 127 (2010) 103–111. URL: https://linkinghub.elsevier.com/retrieve/pii/S0925527310001805. doi:10.1016/j.ijpe.2010.05.001.
[10] S. Kolassa, B. Rostami-Tabar, E. Siemsen, Demand Forecasting for Executives and Professionals, 1 ed., Chapman and Hall/CRC, Boca Raton, 2023. URL: https://www.taylorfrancis.com/books/9781003399599. doi:10.1201/9781003399599.
[11] P. Goodwin, B. Moritz, E. Siemsen, Forecast Decisions, in: K. Donohue, E. Katok, S. Leider (Eds.), The Handbook of Behavioral Operations, 1 ed., Wiley, 2018, pp. 433–458. URL: https://onlinelibrary.wiley.com/doi/10.1002/9781119138341.ch12. doi:10.1002/9781119138341.ch12.
[12] R. Fildes, P. Goodwin, Against Your Better Judgment? How Organizations Can Improve Their Use of Management Judgment in Forecasting, Interfaces 37 (2007) 570–576. URL: https://pubsonline.informs.org/doi/10.1287/inte.1070.0309. doi:10.1287/inte.1070.0309.
[13] SAREF: Ontology Support for Urban Digital Twins and usage guidelines - ETSI TS 103 828 V1.1.1 (2024-03), European Telecommunications Standards Institute (ETSI), 2024. SmartM2M.
[14] B. Božić, W. Winiwarter, Ontology mapping and reasoning in semantic time series processing, in: Proceedings of International Conference on Information Integration and Web-Based Applications & Services, IIWAS '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 443–452. URL: https://doi.org/10.1145/2539150.2539156. doi:10.1145/2539150.2539156.
[15] J. Moreira, L. Daniele, L. Ferreira Pires, M. van Sinderen, K. Wasielewska, P. Szmeja, W. Pawlowski, M. Ganzha, M. Paprzycki, Towards iot platforms' integration: Semantic translations between w3c ssn and etsi saref, 2017. SEMANTiCS conference 2017 ; Conference date: 11-09-2017 Through 14-09-2017.
[16] J. Moreira, L. Ferreira Pires, M. van Sinderen, R. Wieringa, P. Singh, P. D. Costa, M. Llop, Improving the semantic interoperability of iot early warning systems: The port of valencia use case, in: K. Popplewell, K.-D. Thoben, T. Knothe, R. Poler (Eds.), Enterprise Interoperability VIII, Springer International Publishing, Cham, 2019, pp. 17–29.
[17] S. Cox, Observations, Measurements, and Samples, Technical Report, Open Geospatial Consortium (OGC), 2023.
[18] TC 59X, Household appliances network and grid connectivity - Part 1: General requirements, generic data modelling and neutral messages, Technical Report, CEN-CENELEC, 2023.
[19] TC 205, General requirements for Home and Building Electronic Systems (HBES) and Building Automation and Control Systems (BACS) - Part 12-2: Smart grid - Application specification - Interface and framework for customer - Interface between the Home / Building CEM and Resource manager(s) - Data model and messaging, Technical Report, CEN-CENELEC, 2022.
[20] S. Cox, Extension to the Semantic Sensor Network Ontology, Technical Report, OGC/W3C Working draft, 2020.
[21] D. Garijo, M. Poveda-Villalón, Best practices for implementing fair vocabularies and ontologies on the web, 2020. URL: https://arxiv.org/abs/2003.13084. arXiv:2003.13084.
[22] TC SmartM2M, Extension to SAREF Part 1: Energy Domain, Technical Report, ETSI, 2024.
[23] E. Daga, A. Meroño Peñuela, E. Motta, Modelling and querying lists in rdf: A pragmatic study, in: M. Saleem, A. Hogan, R. Usbeck, A.-C. Ngonga Ngomo, R. Verborgh (Eds.), QuWeDa 2019 3rd Workshop on Querying and Benchmarking the Web of Data, CEUR-WS, 2019, pp. 21–36.
[24] L. Daniele., F. den Hartog., J. Roes., How to keep a reference ontology relevant to the industry: A case study from the smart home, in: Post-proceedings of the 12th International Experiences and Directions Workshop on OWL, OWLED 2015, co-located with ISWC 2015, Lecture Notes in Computer Science, vol 9557. Springer, Cham, 2016, pp. 117–123. doi:https://doi.org/10.1007/978-3-319-33245-1_12.