Data Management as an Enabler of Sustainability – Discussion Using the Example of a Digital Data Sheet

Kevin T. Logan*, Marvin M. Meck*, Nils Preuß*, Philipp Wetterich*and Peter F. Pelz*

Technical University Darmstadt, Chair of Fluid Systems, Otto-Berndt-Straße 2, D-64287 Darmstadt, Germany*
E-Mail: kevin.logan@fst.tu-darmstadt.de

Sustainable systems require sustainable data. Imposing the paradigm of sustainability on data corresponds to making it findable, accessible, interoperable and reusable (FAIR). This is vital for planning a fluid system fulfilling functionality as well as satisfactory quality. Therefore, customers are dependent on component suppliers making FAIR product data available. Rather than assuming the context of runtime machine to machine communication and service-oriented interfaces, i.e., solutions based on digital twins, the presented work proposes a *digital data sheet* for components based on citation of information models compliant with W3C standards. This enables suppliers to provide comprehensible and transparent data and customers to implement sustainability already during the design process of fluid systems.

Keywords: sustainability, digitalization, FAIR data, information models
Target audience: Design Process, Pump Manufacturers, Fluid System Planners

1 Introduction

Since the publication of the document "Our Common Future" (also known as the *Brundtland Report*) [1] sustainability has become the paradigm for human development. How this paradigm based on a vague and broadly used term translates into concrete measures is rarely apparent. However, *Sustainable Systems Design* [2] is an approach that shows how to include sustainability as a guiding principle during an engineering design process in a practicable manner. It follows the principle "Maximize Quality subject to Functionality" (cf. Figure 1). The electrical energy consumption of water pumps made up $225\,\mathrm{TWh/a}$ in the year 2015 [3] while a total of $2761\,\mathrm{TWh}$ of electrical energy were generated in the EU in the same year [4]. The energy consumption can be reduced by up to 37% with a system-wide optimization of pump and fluid system [5].

Hence, applying *Sustainable Systems Design* to the process of fluid system planning would be beneficial for contributing to a reduction of energy consumption and in consequence $\mathrm{CO_2}$ emissions. This benefits both economic and ecologic sustainability. Concretely, this implies that the required functionality appears as a constraint in a mathematical optimization, whereas the objective is to maximize the quality of the fluid system. For example, one sub-functionality of the fluid system may be to provide a given pressure difference Δp and volume flow Q. When selecting the corresponding positive displacement pump, the choice is constrained by this requirement but it is motivated by finding the pump with the highest quality. At the same time, quality itself can take the shape of a constraint or objective. As a constraint it can ensure that statutory thresholds are complied with, e.g., rate of efficiency, power consumption, emissions. As an objective, it can determine the direction of development in a continuous improvement process, e.g., minimize material use, maximize efficiency, or minimize carbon footprint. Thus, when planning a fluid system, sustainability in terms of energy and material consumption can be included as a quality that is to be maximized using *Sustainable Systems Design*.

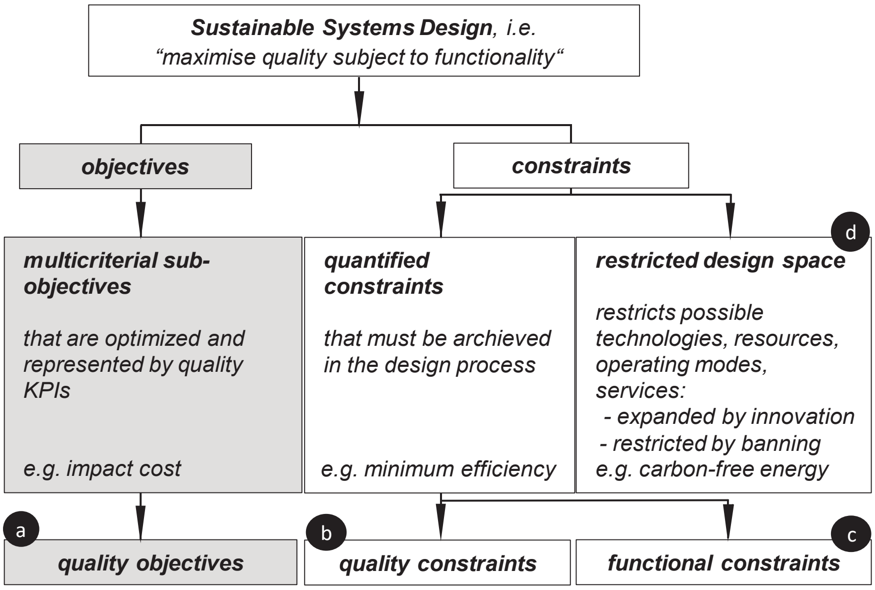

Figure 1: Assignment of quality and functionality to objectives and constraints, whereby these are influenced by the stakeholder's goals and functional specifications. Adapted from [2].

1.1 FAIR Product Data as Enabler for Sustainability

Achieving sustainability is highly dependent on findable, accessible, interoperable and reusable (FAIR) data: During the planning of a sustainable fluid system a positive displacement with a given specification needs to be selected. The playing field is made up of pumps with various specifications from different manufacturers (suppliers) which constitute competing techno-economic systems, cf. Figure 2. The playing field can be restricted by dismissing all pumps that do not fulfil the functional requirements. This is of course only possible if data on the specifications of possibly eligible pumps is findable and accessible. Governing bodies (society) may impose further restrictions to the design space by prohibiting technologies or materials. Furthermore, quality constraints might be legally binding, such as threshold values for efficiency, power consumption or CO_2 equivalent for manufacture. Other technologies may be banned outright (market surveillance). This affects both the designer (customer) and the various pump manufacturers (supplier). The supplier is accountable to society that the offered products comply with those constraints. In practical terms, pump manufacturers need to gather data showing that their products achieve a given rate of efficiency and cause CO_2 emissions only within the scope of emission certificates bought by the company. The customer requires the supplier to provide sufficient and transparent data to ensure compliance with quality constraints applicable to their own product.

Sticking to our example, the bought pump must guarantee that the fluid system meets predefined values for rate of efficiency or CO_2 emissions. To this end, the data further needs to be interoperable and reusable. It is necessary for the designer to be able to feed the product data into their own, possibly very different, software framework in order to calculate the social, ecologic and economic cost the choice of a given pump would imply using specific models. Considering the highly varied models and applications of product data in the design process, the data needs to be made available using formats that are equally versatile, language independent and machine-readable. This also requires adequate metadata to be provided for both humans and machines to understand the origin of the data. This is summarized using the term interoperability of data. The reusability of data is also ensured by supplying it with suitable metadata. Appropriate metadata provides relevant information about context and provenance of the data in a machine-readable form, which allows to evaluate whether it is relevant or applicable for answering a given question. The data gathered by a pump manufacturer needs to be reusable for the designer and in turn, the data generated by them needs to be reusable for their own customers. Data once gathered but used only once, without ensuring reusability, needs to be gathered once again when it is required for addressing a similar problem. This causes dissipation which from an engineering perspective is inefficient and thus detrimental to sustainability. Therefore, FAIR data is sustainable data and sustainable data is required for sustainable fluid systems.

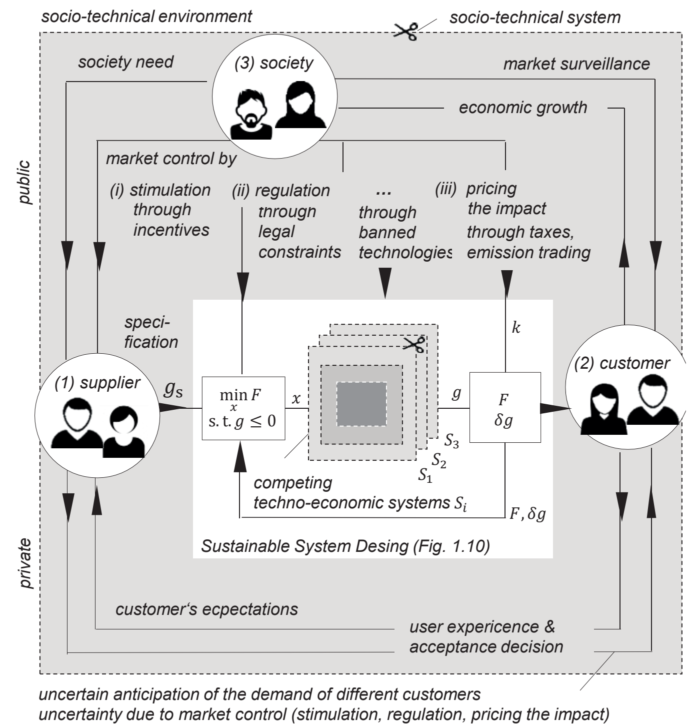

Figure 2: Socio-technical system with the several competing techno-economic systems in its core and the three stakeholder groups supplier, customer, society as human controllers and sensors. Adapted from [2].

1.2 Digital Data Sheets as Vehicle for FAIR Product Data

Currently, a common way for manufacturers of positive displacement pumps to provide data on their products is a data sheet as PDF (portable document format) file for each product line with a selection of technical specifications in tabular form or depicted in figures. However, it is not necessarily the case that these PDF files are freely available for download on the manufacturer's website but rather may only be available on demand with limited specifications given in tabular form as website content. With regard to the requirements derived from the FAIR data principles this is insufficient in several ways which will be detailed in the following.

Findability is impeded by having to go over manufacturers' websites with relevant information spread over several subpages. There is no single point of entry for finding relevant product data on pumps by means of a semantic query for characteristics and specifications relevant to the fluid system under development. When they are available, those PDF files with technical specifications of pumps are themselves often not supplied with metadata, which further complicates findability and narrows the scope of their reusability.

Accessibility is generally measured by Open Access standards. So far, these play a negligible role for manufacturers when it comes to providing product specifications. Procedures for gaining access vary from free download of PDF files to having to ask for an official tender. In either case, PDF files can be considered a front-end barrier withholding access to the actually relevant data, e.g., measurement protocols on which pump curves are based, measurements from which values for maximum pressure difference and volume flow are derived including ranges of uncertainty. So far, these are not generally accessible either through Open Access or licensing agreements. Therefore, it is important that data is provided using an open standard that does not require extra licensing for making product data accessible.

A further drawback of providing product specifications and characteristics in PDF format is that it is not an easily machine-readable format. Attempts have been made to extract data and information from PDF files of data sheets [6]. However, since the data is available before it is brought into PDF format, interoperability could be greatly facilitated by providing data in a machine-readable format directly. A further aspect of interoperability is a standardized controlled vocabulary for describing characteristics and specifications of pumps. Depending on the domain different terminology may be used for describing characteristics of a pump, e.g., pressure difference versus head, and data may be given in different units, e.g., $\mathrm{l/h}$ versus $\mathrm{m^3/h}$. With these peculiarities embedded in PDF files, extracting desired data from them for comparing different products requires one-of-a-kind tools and is burdensome. This also holds true for the case when data is provided in the form of images within a PDF file as is often the case with pump curves.

As already stated, one prerequisite for reusability of data is metadata being available with it, so as to ensure comprehensibility, transparency, and insight into relevance of data. While metadata can be embedded in PDF files the metadata schemes are those of text documents with a focus on catalogues of libraries. This is not the best metadata scheme for the information transported in data sheets and the data it describes. Reusability is further determined by uncertainty of data being quantified as well as the possibility of extending the data provided about a piece of machinery. While the $\mathrm{CO_2}$ foot print of a pump may have been of little interest in the past it is gaining more and more relevance. When new characteristics gain importance, it is necessary to quickly and easily integrate it with the provided data rather than going through a lengthy process of newly publishing PDF data sheets.

To summarize, the imperative "Maximize Quality subject to Functionality" is a practical approach to putting sustainability into practice during an engineering design process. This approach heavily relies on data on technological components. Dissipative work is avoided when this data conforms to the FAIR principles, making these principles the bench mark of sustainability with regard to data itself. Current methods for providing data on fluid technology components fall short of this bench mark.

Using the example of a positive displacement pump, the presented work will show case a method of data management that results in a *digital data sheet* for the pump. The requirements for such a data sheet derived from the issues discussed above are:

(i) a generic, universal information model that can be adapted to the context of specific domains

(ii) usage of a machine-readable format

(iii) an open standard that does not require extra licensing

(iv) adaptability and the possibility of extending the information model to allow for important new characteristics in future development.

The remainder of the presented work is structured as follows. In a section entitled "State of the Art" already available standards for sharing data and information on technological components are presented and discussed with regard to the requirements stated above. Furthermore, the scope of the *digital data sheet* will be delimited against popular concepts such as digital. The central section will illustrate the different approaches undertaken to developing the *digital data sheet* and present the use case of a *digital data sheet* for a positive displacement pump. A following section will discuss to what extent the presented work meets the requirements of the problem statement and illustrate scope for further development. The presented work will be completed by summary and conclusion.

2 State of the Art

The demand for summarized digital data regarding technological artifacts is not new. In various fields attempts have been made to establish standards for such digital representations of objects with Building Information Modelling (BIM) and *Industry 4.0* culminating in the phrase *digital twin*. In this section various competing standards will be presented and discussed with regard to their limitations. Furthermore, a distinction will be drawn between the *digital data sheet* proposed in the presented work and common understanding of the so-called *digital twin*.

2.1 Standards for Information Models

Perhaps the most prominent field where standardized data exchange and integrated digital workflows have developed over the past is the buildings sector with building information modelling. Rather than a standard, BIM can be described as an umbrella term to describe a process that involves digital representation of physical assets in the built environment and their integration in various digital workflows. Hence, the domain of BIM provides extensive digital descriptions of the built environment, including buildings and civil infrastructure. While most of them are proprietary, initiatives like buildingSMART [7] work to facilitate open standards for BIM. BIM can be regarded as already very mature. For instance, in 2015 the German Federal Ministry of Transport and Digital Infrastructure published a phased plan to introduce BIM in steps for future infrastructure projects making BIM mandatory as of 2020, cf. [8]. The open, international standard (ISO 16739-1:2018), "Industry Foundation Classes" (IFC) defines physical components of buildings, manufactured products, mechanical/electrical systems, as well as more abstract structural analysis models, energy analysis models, cost breakdowns, work schedules, etc. This does include pumps, as part of the heating, ventilating and air conditioning (HVAC) domain, however the information model is of course heavily focused on the built environment and geometry models [9–11].

Leaving the buildings sector and looking to the process industry, we find similar efforts towards data exchange standardization being made, as for instance the NAMUR Open Architecture [12], NAMUR Asset Life Cycle Data Model [13], or DEXPI [14] initiatives.

Outside of asset information management, standardized information modelling has also become an important topic in process automation with the advent of industry 4.0 and the internet of things (IoT). Here, the Open Platform Communications Unified Architecture (OPC UA) is gaining increasing popularity. OPC UA is also the prioritized interface standard by the German Machinery and Plant Manufacturers Association (VDMA), cf. [15]. For describing the nature and features of products, there are many established standards from traditional Electronic Data Interchange environments. Among those, the arguably most sophisticated one is ECLASS, covering a wide range of products and services (more than 45 000 types) and defines a huge number of properties for their features (more than 19 000 properties). However, the releases are not accessible without a fee-based license [16–18].

2.2 Standards for Representation of Information Models, Digital Twins and Digital Data Sheets

Information models at first are created primarily on a conceptual level, independent of particular implementations or representations. At this stage, they may be specified and communicated using for example class diagrams or tables of entities and properties. While some information models, e.g., IFC, are provided in a variety of modern semantic representation standards, most of the information models introduced in section 2.1 are strongly coupled with specific data models or even file formats that can be used for their representation. For example, the data models and dictionaries specified by the OPC UA or ECLASS standards are available and to be consumed in the form of their respective XML schema [18, 19].

To ensure the interoperability of data, an information model must sufficiently express what kind of information it provides. A suitable representation must be able to express both schema information and domain semantics. Schema information is used to validate data and includes statements about how the data must be structured, or which upper and lower numerical limits apply to a value. Domain semantics include statements that would otherwise only exist implicitly or informally, for example, which physical unit is used to measure a certain value.

With the rise of semantic web technologies, several open *W3C* (World Wide Web Consortium) specifications for formal knowledge representation have emerged as broadly adopted de-facto standards for semantic modeling. They provide a domain agnostic basis for highly interoperable information models, both on a technical level as well as the semantic level. The central specification is the Resource Descriptor Framework (RDF) as a general formalism for conceptual description or modeling of information [20]. Other open *W3C* specifications built an ecosystem around RDF for implementation, description and validation of semantic graphs: The SPARQL Protocol and RDF Query Language (SPARQL) allows querying and transformation of RDF graphs [21]. The Shapes Constraint Language (SHACL) expresses constraints on semantic graphs as well as validation rules, i.e., representing Ontologies and vocabularies [22]. The Web Ontology Language (OWL) remains the prominent method to express rules for semantic reasoning [23].

In the context of current developments of *Industrie 4.0*, the *Open Manufacturing Platform* (OMP) proposes the standardization of *digital twins* and, e.g., the *Asset Administration Shell* based on information models using the open semantic web technologies described above [24]. A *digital twin* is a virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity [25]. *Digital twins* use real-time and historical data to represent the past and present and simulate predicted futures. *Digital twins* are motivated by outcomes, tailored to use cases, powered by integration, built on data, guided by domain knowledge, and implemented in IT/OT systems [26]. The *Plattform Industrie 4.0* and the *Industrial Digital Twin Association* are starting to define so-called sub-model templates to enable interoperability via the *Asset Administration Shell*. These are templates for different data and behavioral aspects a component or an asset can offer [27–29].

The presented work discusses the use-case example of representing relevant properties of a positive displacement pump as well as available data for planning fluid systems. Out of the available information models introduced in section 2.1, the Companion Specifications for Pumps based on OPC UA is arguably the most comprehensive and modern (c.f. section 3.3). However, the available representations are designed primarily with implementations of machine-to-machine (M2M) Communications in mind. For example, an "OPC UA NodeSet" XML model can be used to generate code for client and server applications automatically, but is not intended to represent semantics [19]. Such service-oriented interfaces designed for runtime M2M communications is however not suitable in the context of this discussed use-case. A typical workflow uses "snapshots" of descriptive data that holds true over an intermediate or long time-period rather than runtime data, condition monitoring or predictive models. Middleware and software stacks that provide capabilities to distinguish and combine descriptive data (static information) as well as runtime data and services are being developed. Leveraging the *Asset Administration Shell* introduced above as an abstract information model, an available open source implementation is BaSyx [30], developed as part of the BaSys 4 platform [31]. However, though it is included in the specification documents [29], current software development kits (SDKs) do not support the proposed compatibility with the aforementioned W3C specifications for semantic modeling [32].

The proposed application of *digital data sheets* therefore addresses the following two challenges:

(i) only a subset of the information or behavioral aspects a *digital twin* of a pump would aspire to provide is needed

(ii) interoperability of the data and its semantic context has extremely high priority, which imposes the use of contemporary semantic technologies, formalisms and formats.

This is in accordance with the current efforts towards standardization of semantic *digital twins* and the modularization or their individual aspects [24].

3 Methods

In the following section, three separate approaches to deriving a suitable information model and choosing an appropriate file format for a *digital data sheet* will be presented and discussed. Of the four requirements mentioned in the problem statement, (i) and (iv) focus on the information model and pose considerable challenges. The first requirement seems initially contradictory in that the information model should be universal and generic on the one hand and domain specific on the other. Considering the second requirement which demands openness to future developments and adaptability the resolution of the contradiction becomes clear in this required flexibility. Three separate approaches for deriving such an information model were tested which will be discussed in the following. First, the current practice of using regular data sheets (unstructured text) is considered as a starting point for conversion into structured data. Second, the current practice of using input or output formats of software tools utilized by designers of fluid systems is considered as part of the status quo regarding practical use of information models. Third, the OPC-UA companion specification for pumps is considered as the favored information model of the available standards introduced in section 2.1.

3.1 Approach Based on Available Data Sheets

Starting from the point of current practice, the initial approach to creating a *digital data sheet* was synthesizing it from the available PDF specification sheets provided by pump manufacturers. Most of the specifications listed in the files are easily represented as key-value pairs, e.g., manufacturer – manufacturer name, catalogue link – hyperlink to the catalogue, maximum operating pressure – value, etc. Attempts have been made to automate and standardize this process and add semantics to the data [6]. However, this representation is not suitable for all data required for characterizing a pump. When considering the behavior of the pump during operation, the design point is still representable as key-value pairs for power consumption, rate of efficiency, pressure difference, and volume flow. Pumps are operated off-design point for considerable amounts of time during their life cycle [5]. Thus, the entire range of possible operating conditions for a pump needs to be representable by the information model. The according data may be provided in the form of figures. One way of transferring this data into a *digital data sheet* is to read value pairs from the figures. This implies discretizing a continuous data set provided in a figure. Since these value pairs do not correspond to the ones the figure is based on, they are subject to uncertainty. A different method is to develop a model to describe characteristic curves as polynomials. This model is then inserted into the *digital data sheet* as key-value pair with the value being a set constituted by the coefficients of the polynomial. As a light-weight, easily adaptable, machine-readable, language independent data-interchange format for the *digital data sheet* JSON (Java Script Object Notation) [33] is chosen. This approach results in manually derived individual *digital data sheets* for each component type.

To summarize, this approach is very straight forward for deriving a *digital data sheet* with the currently available data. However, it leads to a *digital data sheet* that does not use a standardized information model. Neither does the information model use a standard for representation. This results in three considerable insufficiencies:

(i) the vocabulary used in the data sheet may vary considerably,

(ii) semantic modelling is not possible due to a lack of standardized representation of the information model,

(iii) it is burdensome to generate each data sheet individually, i.e., it is not a scalable solution.

3.2 "Smart Data Models" from the Smart Water Domain

A second approach explored in the course of the presented work was to review existing information models that already come with a standardized representation. Specifically, the NGSI-LD (Next Generation Service Interface – Linked Data) graph-based context information model was used. This is based on the representation standard RDF [34]. Various domains have already been serialized using this information model with the JSON-LD format, e.g., smart cities, smart energy, and smart water [35]. The domain smart water includes information models of fluid components, e.g., network, pipe, junction, tank, valve, and pump. The information models of these components in this implementation are based on EPANET, a software tool for simulating water behavior in pipe systems for drinking water distribution [36]. Considering the example of the positive displacement pump, a closer look at the information model for pumps provided here reveals that only three properties, 'id', 'type', 'initialStatus', and two relationships, 'startsAt', 'endsAt', are compulsory, cf. Figure 3 [37]. Properties have a value such as a string or integer in the sense of a key-value pair whereas relationships refer from one entity to another. There are several additional properties and relationships that can further be defined, though the limitations of this information model already become apparent considering these compulsory properties and relationships. 'startsAt' and 'endsAt' are defined as the identifiers of the nodes on the suction and pressure sides of the pump respectively. Hence, it becomes clear that this information model is used primarily in graph-based modelling of water distribution networks and is adapted accordingly.

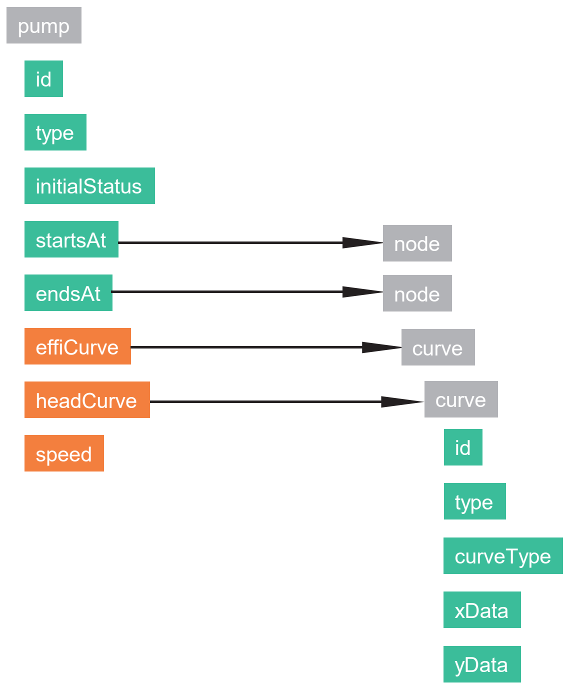

Figure 3: Entities (grey) with required (green) and optional (orange) relationships (with arrows to other entities) and properties (without arrows).

The insufficiency further becomes clear when considering how the operating points and pump characteristics are modelled. For one thing, the model assumes centrifugal pumps thus restricting the scope of the model. Furthermore, the pump characteristics, such as Δp-Q characteristic, are modelled using the entity ‘curves’. In this information model, curves have five required properties, ‘id’, ‘type’, ‘curveType’, ‘xData’, and ‘yData’, cf. Figure 3. For pumps, relevant values of ‘curveType’ are ‘FLOW-HEAD’ for Δp-Q characteristic and ‘FLOW-EFFICIENCY’. Considering Δp-Q characteristic, the information model handles curves differently depending on the number of data points provided in the properties ‘xData’ and ‘yData’. A single-point curve with one value each in ‘xData’ and ‘yData’ for the Δp-Q characteristic assumes the given point to be the design point for operation. A curve is then created around it by assuming two more data points with a maximum Δp and a maximum Q of 133% that of the design point volume flow. A three-point curve uses the data points to identify parameters A, B, C for describing the curve with an equation of the type

$$\Delta p = A - BQ^C \tag{1}$$

A multi-point curve uses linear interpolation between each of the data points to construct the curve. Adaptions of the Δp-Q characteristic depending on pump speed according to the equations

$$\frac{Q_1}{Q_2} = \frac{N_1}{N_2} \tag{2}$$

and

$$\frac{\Delta p_1}{\Delta p_2} = \left(\frac{N_1}{N_2}\right)^2 \tag{3}$$

are possible while efficiency curves may also be provided.

The limitations of this information model can thus be summarized as follows. It is adapted to a specific domain of graph-based modelling of water distribution networks. It further imposes predefined models restricted to centrifugal pumps for the behavior of pumps which are not necessarily suitable for characterizing a pump in every fluid system. The reason for assuming a behavior may be the lack of data. This could be avoided, if manufacturers provide the relevant data and required models and the information model is adaptable to the effect that these models can be integrated. However, the approach of serializing the information model with the JSON-LD standard as a means of addressing interoperability and reusability of data is valuable and should be considered for further use [38].

3.3 Deriving an Information Model from OPC UA Companion Spec for Pumps

Among the information models presented in section 2.1 the OPC UA information model is arguably the most comprehensive while it is also the most modern, only being released earlier this year in April 2021, cf. OPC 40223 [39]. The OPC UA Companion Specification for pumps and vacuum pumps [39] extends on the companion specifications for devices [40] and machinery [41]. At its core, the OPC UA information model for pumps and vacuum pumps comprises seven *FunctionalGroup*s (see OPC 1000-100 for a definition of the *FunctionalGroupType* [40]); they are - in alphabetical order

(i) Configuration

(ii) Documentation,

(iii) Events,

(iv) Identification,

(v) Maintenance,

(vi) Operation,

(vii) Ports,

where only Identification is specified as a mandatory group. *Identification*, contains non-technical data provided to uniquely identify a machine or pump. *Configuration* includes static design, system requirements, and implementation data of the pump, while *Documentation* contains static documentation files of the pump. *Maintenance* records maintenance related data that is mostly (semi-)dynamic but, in some cases, can be static. Lastly, *Events*, *Operation*, and *Ports* are IoT-specific and thus of little interest in a *digital data sheet*. Without going in to the specific details of each group's individual data-structures one can already convince oneself that the OPC UA information model for pumps outside of being useful in IoT applications is also fit for other tasks, for instance planning, asset management, and management of lifecycle records, cf. [42, 43]. This is also reflected in the *Use Cases* section in the standard itself, see chapter 5 in OPC 40223 [39]. Using the information model outside the intended scope, i.e. (I)IoT applications, however, presents itself cumbersome at best. This is owed to the fact that, first, OPC UA heavily relies on code-generation from so-called node-set files in XML-format, and second, most available software stacks (for example open62541 [44], FreeOpcUa [45], S2OPC [46], UAF [47], node-opcua [48]) require one to set up an OPC UA server to make use of OPC UA information models. However, in asset (lifecycle) management for example, interoperability with established database systems is desirable. The node-set approach comes with another disadvantage as the distinction between definition of types and instances becomes blurry.

Hence, a more lightweight and flexible representation of the information model is required to broaden the scope of application. We thus suggest the use of RDF-based representation of a pump and its properties, the *digital data sheet*.

4 Application

In this section, a use case will be presented to illustrate how promising aspects of the approaches described above may be applied to a real-world example for sketching a *digital data sheet* for a positive displacement pump.

4.1 Use-Case: Digital Data Sheet for a Positive Displacement Pump

The information model provided by the OPC-UA companion specification for pumps is chosen as the basis for the use-case application. However, some aspects of a direct use of the available representation in its native XML schema (NodeSet2.xml) for a *digital data sheet* are detrimental to the needs of designers or customers (c.f. section 1) and therefore sustainable systems design. These aspects are: (i) The information model is provided as-is, context specifically needed information needs to be extracted manually, no mechanism is provided to distinguish required properties from optional ones. (ii) & (iii) While the provided formats are machine readable, based on open standards, the ecosystem of available tooling is very narrow. (iv) Extending or adapting the information model as well as the validation of such changes requires tooling specific to this representation schema.

The approach developed in the ongoing AIMS (Applying Interoperable Metadata Standards) project proposes the implementation of metadata application profiles (MAPs) on the basis of W3C standards. This yields a framework for FAIR *digital data sheets* [49–51]: A modeling concept should be used that describes semantic data as MAPs, e.g., a *digital data sheet*, in a modular and hierarchical fashion from elements of established controlled terminologies. This ensures a high degree of specificity can be achieved with maximum applicability and reusability of the metadata schemas. The generated representations can be provided via standard, widely adopted exchange formats like RDF/XML, JSON-LD, Turtle, etc. [52–55]. The presented use-case application uses the turtle-syntax: Manufacturers (suppliers) provide a data graph of facts about their product, illustrated below, using an available RDF based vocabulary or information model. The facts or statements are represented as semantic triplets, shortened to a subject, and a series of indented predicate-object elements. For this use-case application, a small vocabulary of terms was derived from the OPC-UA companion specification information model for pumps.

```
# Declare keywords for namespaces of used vocabularies
@prefix ex: <http://www.example.org/> .
@prefix ua: <http://www.example.org/opcuaRDF#> .
@prefix gr: <http://purl.org/goodrelations/v1#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

# ExamplePump is a positive displacement pump with a geometric displacement
# volume of 0.35 litres and a maximum allowable casing working pressure of 12 bar
# The Terms in <http://www.example.org/opcuaRDF#> are derived from OPC-UA NodeSet
# <http://www.opcfoundation.org/UA/schemas/Pumps/1.0/Opc.Ua.Pumps.NodeSet2.xml>

ex:ExamplePump
    a ua:Pump ;
    rdfs:label "Heavy-duty Pump"@en ;
    ua:Manufacturer “myPumps Inc.” ;
    ua:PumpClass ua:PositiveDisplacementPump ;
    ua:geometricDisplacementVolume ex:QuantitativeValueFloat_GDV ;
    ua:maximuAllowableCasingWorkingPressure ex:QuantitativeValueFloat_MWP .

# UN/EFACT Common Code for “litre” is “LTR”
ex:QuantitativeValueFloat_GDV
    a gr:QuantitativeValueFloat ;
    gr:hasUnitOfMeasurement "LTR"^^xsd:string ;
    gr:hasValueFloat "0.35"^^xsd:float .

# UN/EFACT Common Code for “bar” is “BAR”
ex:QuantitativeValueFloat_MWP
    a gr:QuantitativeValueFloat ;
    gr:hasUnitOfMeasurement "BAR"^^xsd:string ;
    gr:hasValueFloat "12.0"^^xsd:float .
```

Listing 1: Exemplary (meta-) data graph providing information about a pump, using RDF turtle syntax (cf. [20, 55]) and semantic terms derived from the OPC UA companion specification for pumps.

Designers (customers) can then use MAPs based on SHACL, to customize a *digital data sheet* to their needs, based on the knowledge graph of facts about a product, provided by the manufacturer. The example below implements the constraint that for instances of the class *Pump* the property specifying the positive displacement volume is mandatory, leaving other properties optional.

```
# Declare keywords for namespaces of used vocabularies
@prefix ex: <http://www.example.org/> .
@prefix ua: <http://www.example.org/opcuaRDF#> .
@prefix gr: <http://purl.org/goodrelations/v1#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix sh: <http://www.w3.org/ns/shacl#> .

# defines the Shape a dataset needs to provide to fit this digital data sheet
foo:MyPumpShape
    a sh:NodeShape ;
    # this Shape specifies requirements for instances of the Class "Pump"
    sh:targetclass ua:Pump ;
    # "Pumps" in the data sheet must specify the "ua:geometricDisplacementVolume"
    # in the Form of a "gr:QuantitativeValueFloat"
    sh:property [
        sh:path ua:geometricDisplacementVolume ;
        sh:name "Geometric Displacement Volume"@en ;
        sh:minCount 1 ;
        sh:class gr:QuantitativeValueFloat
    ] .
```

Listing 2: Exemplary shapes graph customizing the information provided by the data graph in Listing 1 into a digital data sheet for a pump, using RDF turtle syntax (cf. [20, 55]) and SHACL (cf. [22]).

4.2 Lessons Learned

As the use-case application in section 4.1 shows, the requirements for a *digital data sheet* as stated in section 1 are addressed in the following way: (i) Leveraging W3C standards to represent and validate different views on an information model or vocabulary enables adaptation to different contexts of specific domains. (ii) & (iii) The utilized standards are deployed on the basis of open specifications in machine-readable formats. (iv) Originating from the domains of the world wide web, semantic and linked data, the underlying generic data models are conceptualized with extensibility as a core design paradigm [20]. As illustrated by the use-case application, this enables manufacturers (suppliers) to add provided information to such a knowledge graph as needed. Different designers (customers) can then customize a *digital data sheet* containing the subset of information they need via MAPs. In this way, FAIR product data can be exchanged on a large scale to enable sustainable systems design.

However, state-of-the-art standards for representation of semantic information models are only one part of interoperable *digital data sheets*. The use-case application also highlights the following non-trivial and to some extent non-technical challenges that remain: As stated already in section 2.2, existing information models, categorization standards or vocabularies are strongly coupled with specific, non-semantic data models that are cumbersome to convert to the widely interoperable W3C standards. Several releases of such converted information models are one-time captures of the reused standards, while the standards themselves have undergone significant change in the meantime which limits their usefulness significantly. Research efforts to provide appropriate releases, based on tooling to automate this process compliant with, e.g., the GoodRelations vocabulary for e-commerce [56] and schema.org [57], are severely hindered by the lack of copyright clearance [16, 17]. The availability, and even findability of quality, practically relevant representations of widely adopted information models or vocabularies is therefore extremely limited.

5 Discussion

Providing relevant product data is a key component for the traceability, interpretability and trustworthiness of data, which directly affects its reusability. This in turn is directly connected to the sustainability of products, fluid systems and their operation. Optimizing their sustainability ("Maximize Quality subject to Functionality") as part of the engineering design process is heavily dependent on reliable data and its correct interpretation. A necessary condition for the *digital data sheet* to be put into practice is that pump manufacturers and manufacturers of other fluid components make the relevant component data findable and accessible. The *digital data sheet* primarily addresses the interoperability and reusability aspects of the FAIR principles. As the use-case application shows, findability and availability of existing vocabularies and information models remains a large problem. This is due to a lack of representations following widely adopted, interoperable standards and formats, as well as licensing and copyright objections. This is a major challenge for achieving exchange of context specific, yet interoperable product data on a large scale. An approach to overcoming this challenge suggested in the presented work is for pump manufacturers to make product data available in the form of knowledge graphs. Of course, since data is a valuable resource in the 21st century, accessibility of data invokes the question of monetarization. Since the process of gathering and curating data requires effort, component manufacturers may need to find compensation for providing this service. On the other hand, making such data available free of charge to those engaged in design processes of fluid systems may prove to be an advantage. If the products can be more easily integrated into models of the designed fluid system they may be more likely to be chosen by planners.

In future work, the base of the *digital data sheet* may be expanded on. Provenance of the components described by the data sheets may be included in the data sheets themselves. In this way, the quality of products in terms of sustainability can be assessed easily. From the mining of raw materials via the finished product and its operation to the dismantling and disposal, ecologic impact can be documented and expressed with data included in the *digital data sheets*. For a comprehensive assessment of the sustainability quality, key performance indicators (KPIs) need to be defined and the data required for evaluating performance provided. Transparency is achieved when the provided data follows the FAIR principles resulting in FAIR quality KPIs. A further use for the *digital data sheet* is that it can provide the necessary data to build optimization models for design and operation of hydraulic systems. The target function of such optimization is always to maximize quality or in terms of sustainability minimize effort, i.e., energy and resource consumption. Thus, the data sheet is an important resource for planning and operating sustainable fluid systems.

6 Summary and Conclusion

To conclude, the presented work addresses issues of sustainability in the design process of fluid system planning. It stresses the importance of data as an enabler for making informed decisions to comply with the concept of sustainability. Considering the implications of sustainability for data as a resource, compliance with FAIR data is assumed to be equivalent to sustainable data. A *digital data sheet* is proposed as a means of providing relevant data. This is discussed with reference to commonly available standards for providing data. Three approaches to deriving an appropriate data model for structuring the *digital data sheet* are discussed. In a case study using the example of a positive displacement pump the most promising of the three approaches is illustrated and lessons learned presented. Finally, it is discussed how the *digital data sheet* complies with the principles of FAIR data and how it can contribute to sustainability in the design process.

7 Acknowledgements

This work has been co-funded by the LOEWE initiative (Hesse, Germany) within the scope of the emergenCITY center, as well as the German Federal Ministry for Economic Affairs and Energy (BMWi) within the scope of the HECTOR research project as part of the ENPRO 2.0 Initiative (Funding code: 03EN2006A), as well as the German Research Foundation (DFG) within the scope of AIMS – Applying interoperable metadata standards (Project ID: 432233186), as well as the German government and the governments of the federal states within the NFDI4Ing consortium as part of the NFDI initiative.

Nomenclature

Variable	*Description*	*Unit*
Δp	Surface Pressure	[bar]
Q	Volume Flow	$[\mathrm{m}^3/\mathrm{s}]$
N	Rotational Speed of Pump	$[\mathrm{s}^{-1}]$
A	Maximum Pressure Difference Parameter of the Δp-Q Characteristic	[bar]
B	Base Parameter of the Δp-Q Characteristic	$[\mathrm{bar}\,\frac{\mathrm{s}^C}{\mathrm{m}^{3C}}]$
C	Exponential Parameter of the Δp-Q Characteristic	[-]

References

[1] United Nations World Commission on Environment and Development, *Our Common Future*. Oxford: Oxford University Press, 1987.

[2] P. F. Pelz, P. Groche, M. E. Pfetsch, and M. Schaeffner, *Mastering Uncertainty in Mechanical Engineering*. Cham: Springer International Publishing, 2021.

[3] European Commission, *Ecodesign Pump Review: Study of Commission Regulation (EU) No. 547/2012 (Ecodesign requirements for water pumps).* [Online]. Available: https://www.eceee.org/static/media/uploads/site-2/ecodesign/products/Electric%20pumps%20ENER%20Lot%2011/review_study_on_water_pumps_-_december_2018.pdf

[4] *Electricity production, consumption and market overview.* [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_production,_consumption_and_market_overview (accessed: Dec. 18 2021).

[5] B. Went, “The systems approach to reducing carbon emissions,” *World Pumps*, vol. 2008, no. 500, pp. 52–55, 2008, doi: 10.1016/S0262-1762(08)70153-6.

[6] M. Frank, *From PDF data sheets to shared understanding with serverless SHACL.* [Online]. Available: https://blog.codecentric.de/en/2020/04/pdf-data-sheets-shared-understanding-serverless-shacl/ (accessed: Oct. 17 2021).

[7] buildingSMART International, *What is openBIM?* [Online]. Available: https://www.buildingsmart.org/about/openbim/openbim-definition/ (accessed: Oct. 29 2021).

[8] Federal Ministry of Transport and Digital Infrastructure (BMVI), “Road Map for Digital Design and Construction: Introduction of modern, IT-based processes and technologies for the design, construction and operation of assets in the built environment,” Federal Ministry of Transport and Digital Infrastructure (BMWi), Berlin, 2015.

[9] ISO, *ISO 16739-1:2018.* [Online]. Available: https://www.iso.org/standard/70303.html (accessed: Oct. 31 2021).

[10] buildingSMART Technical, *Industry Foundation Classes (IFC) - buildingSMART Technical.* [Online]. Available: https://technical.buildingsmart.org/standards/ifc (accessed: Oct. 31 2021).

[11] *IFC4.3.x Documentation Cover.* [Online]. Available: http://ifc43-docs.standards.buildingsmart.org/IFC/RELEASE/IFC4x3/HTML/content/cover.htm (accessed: Oct. 31 2021).

[12] Namur, *NAMUR Open Architecture.* [Online]. Available: https://www.namur.net/en/focus-topics/namur-open-architecture.html (accessed: Oct. 31 2021).

[13] Namur, *Asset Life Cycle Data Model.* [Online]. Available: https://www.namur.net/en/focus-topics/asset-life-cycle-data-model.html (accessed: Oct. 31 2021).

[14] *DEXPI – Data Exchange in the Process Industry.* [Online]. Available: https://dexpi.org/ (accessed: Oct. 31 2021).

[15] German Machinery and Plant Manufacturers Association (VDMA), “Industrie 4.0 Communication Guideline: Based on OPC UA,” 2017.

[16] M. Hepp and A. Radinger, *eClassOWL - The Web Ontology for Products and Services.* [Online]. Available: http://www.heppnetz.de/projects/eclassowl/ (accessed: Oct. 31 2021).
[17] A. Stolz, B. Rodriguez-Castro, A. Radinger, and M. Hepp, *PCS2OWL: A Generic Approach for Deriving Web Ontologies from Product Classification Systems.* [Online]. Available: http://www.heppnetz.de/files/pcs2owl-eswc2014.pdf (accessed: Oct. 31 2021).
[18] CEN - European Committee for Standardization, *Guidelines for the design, implementation and operation of a product property server (ePPS).* [Online]. Available: https://ftp.cencenelec.eu/CWA/CEN/CWA_16100ePPS.pdf (accessed: Oct. 31 2021).
[19] OPCFoundation, *OPC 10000-6 Unified Architecture Part 6 Mappings Overview.* [Online]. Available: https://reference.opcfoundation.org/v104/Core/docs/Part6/F.1/ (accessed: Oct. 31 2021).
[20] World Wide Web Consortium, *RDF 1.1 Concepts and Abstract Syntax.* [Online]. Available: https://www.w3.org/TR/rdf11-concepts/ (accessed: Oct. 31 2021).
[21] World Wide Web Consortium, *SPARQL Query Language for RDF.* [Online]. Available: https://www.w3.org/TR/rdf-sparql-query/ (accessed: Oct. 31 2021).
[22] World Wide Web Consortium, *Shapes Constraint Language (SHACL).* [Online]. Available: https://www.w3.org/TR/shacl/ (accessed: Oct. 31 2021).
[23] World Wide Web Consortium, *OWL 2 Web Ontology Language Primer (Second Edition).* [Online]. Available: https://www.w3.org/TR/2012/REC-owl2-primer-20121211/ (accessed: Oct. 31 2021).
[24] Open Manufacturing Platform, *Open Manufacturing Platform.* [Online]. Available: https://open-manufacturing.org/ (accessed: Oct. 31 2021).
[25] Digital Twin Consortium, *Digital Twin Consortium Glossary – Digital Twin.* [Online]. Available: https://www.digitaltwinconsortium.org/glossary/glossary.html#digital-twin (accessed: Oct. 31 2021).
[26] S. Olcott and C. Mullen, *Digital Twin Consortium Defines Digital Twin.* [Online]. Available: https://blog.digitaltwinconsortium.org/2020/12/digital-twin-consortium-defines-digital-twin.html (accessed: Oct. 31 2021).
[27] Federal Ministry of Economic Affairs and Energy (BMWi), *Submodel Templates of the Asset Administration Shell - Generic Frame for Technical Data for Industrial Equipment in Manufacturing (Version 1.1).* [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/Submodel_templates-Asset_Administration_Shell-Technical_Data.html (accessed: Oct. 31 2021).
[28] Federal Ministry of Economic Affairs and Energy (BMWi), *Submodel Templates of the Asset Administration Shell - ZVEI Digital Nameplate for industrial equipment (Version 1.0).* [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Submodel_Templates-Asset_Administration_Shell-digital_nameplate.html (accessed: Oct. 31 2021).
[29] Federal Ministry of Economic Affairs and Energy (BMWi), *Asset Administration Shell Specifications.* [Online]. Available: https://www.plattform-i40.de/IP/Redaktion/EN/Standardartikel/specification-administrationshell.html (accessed: Oct. 31 2021).
[30] *BaSyx / Documentation / AssetAdministrationShell - Eclipsepedia.* [Online]. Available: https://wiki.eclipse.org/BaSyx_/_Documentation_/_AssetAdministrationShell (accessed: Dec. 21 2021).
[31] *BaSys 4.0 | Basissystem Industrie 4.0.* [Online]. Available: https://www.basys40.de/ (accessed: Dec. 21 2021).
[32] *BaSyx / Download - Eclipsepedia.* [Online]. Available: https://wiki.eclipse.org/BaSyx_/_Download#BaSyx_SDK (accessed: Dec. 21 2021).
[33] *JSON.* [Online]. Available: https://www.json.org/json-en.html (accessed: Oct. 29 2021).
[34] Context Information Management (CIM) ETSI Industry Specification Group (ISG), *GS CIM 006 - V1.1.1 - Context Information Management (CIM); Information Model (MOD0).* [Online]. Available: https://www.etsi.org/deliver/etsi_gs/CIM/001_099/006/01.01.01_60/gs_CIM006v010101p.pdf (accessed: Oct. 23 2021).
[35] FIWARE, *Smart Data Models - FIWARE.* [Online]. Available: https://www.fiware.org/developers/smart-data-models/ (accessed: Oct. 23 2021).
[36] L. A. Rossman, H. Woo, M. Tryby, F. Shang, R. Janke, and T. Haxton, “EPANET 2.2 User Manual,” United States Environmental Protection Agency, Washington, DC, EPA/600/R-20/133, 2020.
[37] *Swagger UI.* [Online]. Available: https://swagger.lab.fiware.org/?url=https://smart-data-models.github.io/dataModel.WaterDistributionManagementEPANET/Pump/swagger.yaml (accessed: Oct. 23 2021).

[38] *Understanding @context - Step-by-Step for NGSI-LD.* [Online]. Available: https://ngsi-ld-tutorials.readthedocs.io/en/latest/understanding-%40context.html (accessed: Oct. 24 2021).
[39] *OPC UA for Pumps and Vacuum Pumps*, OPC 40223, OPC Foundation, 2021.
[40] *OPC Unified Architecture: Part 100: Devices*, OPC 10000-100, OPC Foundation, 2021.
[41] *OPC UA for Machinery: Part 1: Basic Building Blocks*, OPC 40001-1, OPC Foundation, 2021.
[42] *DIN 77005-1:2018-09, Lebenslaufakte für technische Anlagen - Teil 1: Strukturelle und inhaltliche Festlegungen*, Berlin.
[43] M. Schenk, “Nachhaltigkeit bei der Nutzung digitaler Daten in der Produktion,” 2019.
[44] *Introduction — open62541 1.2.0-5-gcc3a3d396-dirty documentation.* [Online]. Available: https://open62541.org/doc/1.2/ (accessed: Dec. 21 2021).
[45] *Python OPC-UA Documentation — Python OPC-UA 1.0 documentation.* [Online]. Available: https://python-opcua.readthedocs.io/en/latest/ (accessed: Dec. 21 2021).
[46] *S2OPC OPCUA Toolkit: S2OPC Toolkit API.* [Online]. Available: https://systerel.gitlab.io/S2OPC/ (accessed: Dec. 21 2021).
[47] *uaf/uaf: OPC UA made easy - main repository of the Unified Architecture Framework.* [Online]. Available: https://github.com/uaf/uaf#readme (accessed: Dec. 21 2021).
[48] *NodeOPCUA reference documentation - public API.* [Online]. Available: https://node-opcua.github.io/api_doc/2.32.0/index.html (accessed: Dec. 21 2021).
[49] Program for Cooperative Cataloging (PCC) Task Group on Metadata Application Profiles, *PCC Task Group on Metadata Application Profiles - Program for Cooperative Cataloging (Library of Congress).* [Online]. Available: https://www.loc.gov/aba/pcc/taskgroup/Metadata-Application-Profiles.html (accessed: Oct. 31 2021).
[50] *AIMS - Applying Interoperable Metadata Standards.* [Online]. Available: https://www.ulb.tu-darmstadt.de/forschen_publizieren/forschen/aims.de.jsp (accessed: Oct. 31 2021).
[51] *AIMS-Projekt.* [Online]. Available: https://www.aims-projekt.de/ (accessed: Oct. 31 2021).
[52] World Wide Web Consortium, *RDF 1.1 XML Syntax.* [Online]. Available: https://www.w3.org/TR/rdf-syntax-grammar/ (accessed: Oct. 31 2021).
[53] World Wide Web Consortium, *JSON-LD 1.1.* [Online]. Available: https://www.w3.org/TR/json-ld11/ (accessed: Oct. 31 2021).
[54] *JSON-LD - JSON for Linking Data.* [Online]. Available: https://json-ld.org/ (accessed: Oct. 31 2021).
[55] World Wide Web Consortium, *RDF 1.1 Turtle.* [Online]. Available: https://www.w3.org/TR/turtle/ (accessed: Oct. 31 2021).
[56] *GoodRelations is the new, official e-commerce core for schema.org.* [Online]. Available: http://wiki.goodrelations-vocabulary.org/Cookbook/Schema.org (accessed: Oct. 31 2021).
[57] Schema.org, *Schema.org.* [Online]. Available: https://schema.org/ (accessed: Oct. 31 2021).