IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 29, 2019, accepted June 22, 2019, date of publication July 8, 2019, date of current version July 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927398

Managing Data, Information, and Technology in Cyber Physical Systems: Public Safety as a Service and Its Systems

MONICA DRĂGOICEA[1], (Member, IEEE), MICHEL LÉONARD[2], SORIN N. CIOLOFAN[3], AND GHEORGHE MILITARU[4]

[1]Department of Automatic Control and Systems Engineering, University Politehnica of Bucharest, 060042 Bucharest, Romania
[2]Institute of Information Service Science, University of Geneva, CH-1227 Geneva, Switzerland
[3]Department of Computer Science, University Politehnica of Bucharest, 060042 Bucharest, Romania
[4]Department of Entrepreneurship and Management, University Politehnica of Bucharest, 060042 Bucharest, Romania

Corresponding author: Monica Drăgoicea (monica.dragoicea@upb.ro)

This work was supported in part by the Ministry of National Education under PubArt Grant.

ABSTRACT This paper introduces a unified representation for collaborative development of complex services in public safety to explain the infusion of digital technologies into the design of the community resilience processes with respect to the potential hazardous events. It aims at helping to understand how ICT-based knowledge may help stakeholders situated in various positions inside society to participate collaboratively to service development actions. It presents a roadmap to conceive, design, and operationalize the creation of digital artefacts to compound its supporting systems, including the digital one. This representation includes a *way of describing* the domain of interest to conceive complex services employing human-oriented development, a *way of reasoning* on the resilience processes complexity, using semantic reasoning along with the time series quality assurance (TSQA) solution ontology, and a *way of developing* data processing components as internal (technical) services in enterprise information systems to support the design of novel environmental monitoring digital services. A unified semantic reasoning-based approach to evaluate data quality in cyber–physical systems is described to exemplify the creation of a complex public service ecosystem that promotes collaborative knowledge sharing to formalize the domain expertise through the information intensive services. The TSQA ontology integrates knowledge from other domain-specific ontologies to define and share concepts designating observations acquired from sensors, quality issues, methods for detecting quality issues and correcting data, and tags applied to data objects to assure the data traceability. A semantic component that manages the TSQA ontology and the SWRL-encoded rules are introduced in the data acquisition module of a cyber–physical system application for environmental monitoring to solve a specific problem of data cleaning associated with the water resources management. This method is applicable to any time series of measured data.

INDEX TERMS Collaborative development, complex services, ontologies, public safety, resilience, semantic reasoning, water resource management.

I. INTRODUCTION

A growing number of activities related to hazard assessment and evaluation of vulnerability to hazardous events in public safety has been associated with a laborious data collection effort that is being accompanied by digital systems and technologies such as ICT, Internet, Semantic Web, Big Data, sensors, blockchains, mobile systems, machine reasoning, machine learning, and robotics, or intelligent immersive systems (such as augmented reality and mixed reality). This digital evolution incents the implementation of novel decision-making activities in public safety in times of need through modern Emergency Management Information Systems implemented worldwide, to improve the city resilience processes by automating their activities with various digital technologies [1].

Public authorities are increasingly relying today on digitalization to realize complex and dynamic public service

The associate editor coordinating the review of this manuscript and approving it for publication was Jinsong Wu.

interactions [2]. New safety functionalities may be enabled through the Internet of Services (IoS) by exploiting connectivity, integrating various operational processes to increase co-created value through novel public service offerings facing complex situations in Society. Consequently, new public service operating modes using digitized artefacts supporting digital innovations are required [3], [4].

Public awareness has become an important factor in disaster risk reduction [5], recognizing today the role of *citizens as contributors* to the community shared knowledge [6], as a key participation to the development of services in public safety [7]. The enactment of improved working procedures to support specific global sustainability and resilience targets is driven by the interactions between the physical world of natural phenomena and the digital world created through active participation of various stakeholders [8]. This aspect strengthens the capability of upgrading the city resilience-building processes with information-based intelligence. They appear at different levels of interaction: enterprise instrumentation networks [9], *crowdsourcing* [10], *crowdsensing* [11], or *participatory sensing* [12]. Several approaches are operationalized to support low-cost mitigation measures, aiming to increase community involvement. Some examples include the use of smartphones as mobile sensors [13]–[15], *crowdsourcing* and *crowdsensing* with smartphone-based Earthquake Early Warning Systems [16], smartphone-based participatory sensing for air pollution monitoring [17], or automated collecting and querying macroseismic data based on community reports [18], [19]. This continuous interaction between the user of various types of mobile devices and its environment led to the definition of new types of interaction spaces [20] and user-generated content management frameworks [21]. A concept aiming to describe the possibility to automate most of the informational interactions that improves the user experience is the Smart Space. Its main enabler is the semantic web and its related technologies. The Smart Space software development model fosters information sharing-based interactions blurring the line between physical and information worlds [22].

A worldwide perspective on public safety is enforced by the United Nations Office for Disaster Risk Reduction (UNISDR) aiming at a multi-stakeholder coordination of the regional platforms for disaster risk reduction [23]. The seven global targets of the Sendai Framework for Disaster Risk Reduction 2015-2030 drive towards increasing the "*availability of and access to multi-hazard early warning systems and disaster risk information and assessments to the people by 2030*" [24]. To promote Key Priority 1, Understanding disaster risk, Sendai Framework advises for the "*collection, analysis, management and use of relevant data and practical information and ensure its dissemination, taking into account the needs of different categories of users, as appropriate*", at national and local levels.

A resilience-based approach to address expected and unexpected events, implementing proactive and reactive actions, has been proposed to be an integrative, holistic vision to involve all relevant stakeholders across complexly interconnected systems and identifiable public safety services [5], p.24, [25].

Following these general directions in public safety, connecting people sustainably and resiliently with information in times of need becomes both a fundamental problem and a research opportunity. As such, a growing number of position papers start to acknowledge the role of digital technologies in supporting the concretization of the 17 Sustainable Development Goals (SDGs) defined by the United Nations General Assembly [26] that stays at the heart of the United Nations 2030 Agenda for Sustainable Development. Reference [27] presents a comprehensive, visionary study aiming at identifying correlations among ICT and the SDGs. It stresses the need of a collective, collaborative effort among individuals in different disciplines, various industries, and agencies, in such a way to produce further synergies between the two fields. A novel ICT framework that approaches on three levels - data, sustainability, and governance - the implementation of the SDGs has been introduced in [28]. Reference [29] introduces a discussion related to public safety and its correlation to the four priorities of action described by the Sendai Framework for Disaster Risk Reduction, identifies key data intensive activities from these recommendations, and describes two managerial implications in service ecosystems, with a discussion on a Viable Systems Approach (vSa) perspective [30] on the response to disasters operating rules. Other position papers explain specific corporate and agency activities aiming to support the achievement of the SDGs though various ICT services and related initiatives [31]–[33].

These new complex emerging situations arising today require to be addressed intelligently by integrating information and knowledge that stay at the core processes of a knowledge society. Therefore, research and development efforts may naturally undertake scientific activities aiming to expand the service mindset through active exploration in all domains and critical aspects of our Society, being recognized today that the law of interaction among the large number of Society-level entities is the service [34], [35]. As well, complexity, "*a key characteristic of the world we live in and of the systems that cohabit our world*" [36] must be understood against Society needs. By means of service, being understood as the utilization of specific competences such as knowledge, skills, and technologies of one entity for the benefit of another entity, as defined by Service Dominant Logic [34], knowledge can be considered both as the primary production-resource and the main instrument of value co-creation [37] in service systems, the main value creation entities [38], [39]. This intelligence of service, acting across disciplines, tying economic, social, business, and ICT aspects, addresses identification, analysis, concretization, realization, and implementation of novel ideas approaching the challenges induced by these situations. Therefore, service intelligence fosters the expansion of human intelligence, helping to set up concrete knowledge and skills exchanges inside Society towards the progression facing complex situations. They are generally understood as

exchanges between a large governance entity, its citizens, and its influencers, like in the case of public services, and particularly, exchanges between a provider and a beneficiary, at an intangible level.

This paper introduces an exploration perspective on the development of creative artefacts at various levels of active collaboration to obtain a good design of complex services and their supporting systems, including the digital, in the domain of public safety. Here, a data-centric situational perspective for complex service engineering to support resilience building with respect to hazardous events is built and explored. It aims to operationalize the Public Safety as a Service vision that is defined as an "*inclusive and responsible value co-creation design vision for liveable regions, fostering the expansion of service knowledge to multiple public service contributors, potentially transforming data into information using service intelligence to create sustainable, resilient, and trustworthy service ecosystems*" [29]. This definition advances public safety as a major dimension of action that requires innovated resilience building activities integrating safety critical service processes within the scope of the Sendai Framework for Disaster Risk Reduction [24] and the 17 Sustainable Development Goals defined by the United Nations General Assembly [26].

The remainder of the paper is organized as follows. In Section II, a data-centric situational perspective is formulated to address public safety concerns triggered by hazardous events in water resources management. Specifically, it formulates the problem of Data Quality and Data Cleaning for large volume of data achieved from sensors in Cyber Physical Systems.

A general working methodology is introduced in Section III to exemplify the operationalization of the Public Safety as a Service vision. From a practical point of view, addressed in a specific case study, it is intended to expand service knowledge in understanding water natural variations in terms of quality, using ontologies to embed the expert domain knowledge and semantic reasoning to mine on available data. We introduce a unified description of the main service activities aiming to help identifying service processes, resources, and information, to increase resilience of communities with respect to potential critical hazardous events affecting public safety.

Section IV and Section V describe and operationalize a unified semantic reasoning-based approach for improving data quality that combines both semantic technologies and data mining algorithms. The method is applicable to any time series data (any type of measured data). To evaluate the proposed solution against real measured data, a water quality data situation has been constructed to be analyzed following the proposed working methodology. Exploratory Data Analysis (EDA) is employed here to obtain a general evaluation and understanding of the "big picture" of the data to work on. EDA [40] assumes going through all data and generating summary statistics, plotting distributions using box plots, understanding relations between variables using scatter-plot matrices. This helps to better understand both the data as the primary input for later processing and the phenomena that produced the data [41]. Based on this understanding, the actual problem may be solved, such as a prediction or classification problem, implementing specific algorithms, e.g. Machine Learning algorithms or statistical methods. The results are presented as reports and communicated to the various stakeholders to support other decision-making processes. In this way, complex Information Intensive Services (IIS) emerge [42] and the "data product" is built and deployed back into the Real World.

A summary of the achieved results is presented in Section VI. The method presented in this paper introduces a semantic component in the data acquisition module of a Cyber-Physical System for environmental monitoring to solve specific problems of data quality assurance, here the difficult problem of separation of True Positive data from False Positive data. This is a major concern in data cleaning and the proposed solution addresses this problem embedding semantic technologies. Further possible developments are discussed in Section VII and final conclusions are presented in Section VIII.

II. REVIEW ON DATA QUALITY EVALUATION IN CYBER PHYSICAL SYSTEMS

Various approaches to advance with data the development of information intensive services (IIS) in public safety for the specific case of water resources management may be integrated today, such as water pollution control [24], [43], with information-based intelligence [44], stressing the role of digital technologies for public safety concerns [45], [29] in managing urban water sustainably [46]. To create, evaluate, and understand the general situation of data integration through IIS for water resource management and to apply Exploratory Data Analysis [40] for the assessment of critical situations in public safety, in this section we explain two general use cases on data collection in Cyber-Physical Systems (CPS).

The Cyber-Physical Systems term generally refers the tight integration between computational (cyber) and physical elements and processes [47], where things can be sensed by means of sensor networks and actuated by means of actuator networks [48]. Within the scope of this work, the term Cyber-Physical Systems is better understood with the meaning of "*a new way of cooperation among distributed and intelligent smart networked devices as well as with humans*" [48], where different interaction technologies such as the Internet and the Internet of Things are employed to provide networking and connectivity, support communication, and enable complex information exchanges between various CPS [49], [50], [51].

Sensing in CPS is responsible with the production of an enormous amount of data that must be stored, processed, and analyzed with specific tools and digital systems and technologies, other than the conventional ones [49], [52]. The importance of Big Data, as "*high volume, high velocity, and/or high variety information assets that require new forms*

of processing to enable enhanced decision making, insight discovery and process optimization'' [53], has been acknowledged today not only in the Cyber-Physical Systems perspective [49], but also in its various application domains such as intelligent transportation systems [54], environmental monitoring [55], resource efficiency and sustainability [52], [56], offering support for science, technology, engineering, and mathematics (STEM) education for extending information-based advances in various disciplines [57], or explaining complex situations in disaster risk reduction [58], [59].

Cyber-Physical Systems often embody as system of systems, with an intrinsically heterogeneous composition of distributed, concurrent, large scale complex systems that require specific tools and technologies for interoperability and functionality interlinking, such as domain specific ontologies, interoperability standards, human-machine interfaces [48], [60], hybrid systems modeling and simulation [61], and complex architectural frameworks based on new modeling languages and standards such as the Systems Modeling Language (SysML) [57], [62].

Ontologies provide shared formalization of a domain, by encapsulating knowledge and human experience in a machine understandable way [63], [64]. An ontology represents a formalized modeling of data that defines concepts and their attributes, taxonomies that allow to classify concepts using generalization and specification, relationships among concepts, rules (axioms) that become true if some conditions evaluate to *true*, and instances of concepts (also called ''individuals''). Formal languages that are used to define ontologies are called ''ontology languages''. Two languages often used are RDF (Resource Description Framework) [65] and OWL (Web Ontology Language) [66]. In [67] it is pointed out that despite the fact that both are intended for semantic modeling of data, OWL has a more extended vocabulary, is based on XML (Extensible Markup Language), and has become a web standard (a W3C recommendation).

Concerning interlinking of Cyber-Physical Systems applications, semantic technologies and ontologies offer powerful support in integrating domain expert knowledge in autonomous and intelligent systems [63], [68]. A semantic rule engine (SRE) system installed on top of industrial gateways that allows stakeholders to control and monitor industrial devices is proposed in [69]. The SRE is composed of two parts: a Rule Engine (RE) which defines the rules and actions to be executed on actuators and a Semantic Engine (SQenIoT) which allows execution of semantic queries. The novelty of the solution comes from the fact that combining the two components (RE and SQenIoT), the rules do not refer to specific devices ID's but to concepts, thus making the rules valid even in the case of future sensors replacements. A novel publish-subscribe architecture aiming to support interoperability at information level, introduced in [70], extends the Smart-M3 semantic interoperability platform for smart spaces [71]. With a case study related to work-safety regulation compliance, an Open Semantic Framework (OSF) is introduced in [63]. This framework proposes the integration of expert knowledge in domain specific ontologies with domain specific knowledge packs, in such a way to support users understanding between various models in industrial engineering design.

In Cyber Physical Systems applications for environmental monitoring, sensor observations result in many small data objects acquired in real-time, where each object contains several attributes (such as location, humidity, temperature, pressure, wind speed). Storing together all these small data objects results in Big Data repositories where Data Quality becomes an important concern for any category of data, regardless of size and collection method. Wrong data values publicly available to stakeholders have as result, in the best case, loss of credibility in the project, and in the worst case could lead to tragic consequences, especially when the data should be used to support decisions in critical situations (such as floods, water pollution). For this reason global ongoing research efforts try to design the frame for the future Quality Assurance (QA) protocols and standards [72]. Besides natural inherent processes that affect water quality (hydrological, physical, chemical or biological) the most significant impact factor results from human activities (urban sewage, agriculture, industrial and urban waste disposal, dredging, navigation, and harbors) that dispose bacteria, nutrients, pesticides and herbicides, metals, oils and greases, and industrial organic micro-pollutants [46], [73].

The GS1 Data Quality Framework (DQF) [74] specifies that Data Quality must be characterized based on ''*complete, consistent, accurate, time-stamped and industry standards-based*'' data. Completeness means that missing data should be minimized (ideally reduced to zero). Consistency refers to the characteristic of the data to be logically valid across multiple views. Accuracy describes the degree of closeness of observed results to the true values. Time-stamped data are appended the real moment (yyyymm-dd-hh-ss.ms) of generation, to be later ordered on a timescale.

Data Cleaning is the third stage in the Data Science process presented by [75]. Other synonyms used in the literature to denote the same process are ''Error Detection'', ''Data Scrubbing'', and ''Data Cleansing''. Data Cleaning is ''a process used to determine inaccurate, incomplete, or unreasonable data and then improving the quality through correction of detected errors and omissions'' [76]. This stage consists of procedures to detect data anomalies such as outliers, missing values, or duplicates. The first step is the raw (unstructured) data collection (sensor data, genetic data, health data, social media data), followed by the second step, i.e. processing of raw data and their mapping into a file format easily consumed by automated tools, to be eventually exposed through internal (technical) services for further integration.

Physical quantities (such as temperature, pH, pressure, other parameters related to water quality, solar radiation, soil, air) collected through various measuring devices at some specified frequency may be represented as time series that have specific characteristics, such as: stationarity (the mean, standard deviation and autocorrelation does not change over

time; trend (refers to a long-term increase or decrease in data values), seasonal (manifests when the influence of a seasonal factor can be detected, such as month, day of week), and cycle (pattern that refers to fluctuations of data which are not of fixed periods) additive components that may appear in different combinations.

Data Cleaning activities cannot rely entirely on automated procedures because in some cases it is impossible to distinguish between false positive and true positive without access to domain knowledge. Smart solutions require models that can ensure the interoperability and analysis of data. For this reason, semantic technologies may be successfully used to formalize this domain expertise from knowledgeable contributors as a valuable input to the information intensive services development to assess hazardous events. They offer the framework where changes of the to-be-modeled universe are easy to implement. In comparison with relational databases, where a constant and costly redesign is required each time when changes of concepts and relationships between them have to be made, Resource Definition Framework (RDF) requests the data being modeled as a graph [77]. Thus, new concepts and relationships can be added without requiring to change the schema.

Related work in the domain of defining ontologies for quality improvement of sensors data in CPS does not offer yet a unified approach binding together observations from sensors, methods for automatic detection of erroneous data, domain knowledge and correction procedures. Both aspects, ontologies and Data Quality, have been studied in isolation and few attempts were made in the direction of building an ontology-based framework for Data Quality. This research field has been developed in three directions:

- ontologies for Data Retrieval;
- ontologies for Data Integration;
- ontologies for Data Cleaning.

In [78] three specific applications of ontologies in data management for consistency checking, duplicate detection, and metadata management are presented, respectively. In [79] Data Cleaning in multisource information systems such as data warehouses where distributed data sources contribute to an integrated repository is discussed. Based on a study on 61 papers related to the use of ontologies for Data Quality in integrated chronic disease management retrieved from the main databases, in [80] it is concluded that Data Quality does not have a generally accepted conceptual framework and definition and more applications based on ontologies to support automated evaluation of Data Quality are needed. An ontology design pattern is proposed in [81] to assess the quality of spatial data, without including abstractions for detection and correction of data anomalies. An ontology-based Data Quality framework for data streams is described in [82], including three types of metrics for Data Quality: content-based (use of semantic rules defined by the user to measure quality aspects such as consistency), query-based (computes the data quality for query operators such as aggregation operators), and application-based (can use any function which computes an application Data Quality value). The user may define simple constraints on data, but the solution does not offer automatic selection of algorithms for anomaly detection.

The first use case that may be defined following the above discussion concerns the quality of data collected automatically with field sensors used to take measurements of interest (spot sampling) in various environmental monitoring and control activities. This can further involve telemetry (measurement is made in the field and collected data is sent, usually by wireless transfer, to remote monitoring equipment) as presented, for example, in [83]. The quality of data collected through a field sensor depends on several aspects, such as poor calibration, changes suffered by the sensor during transportation to the deployment site, vandalism, accumulation of algae, plants or other microorganisms on the surface of the sensor, extreme natural phenomena (such as extreme cold conditions, high flows), bad circuit boards, or just by ageing. Within a sensor network, it is essential to enforce the same operation standards and procedures in such a way that final data are consistent, and data gathered in different locations can be compared.

The second use case concerns the collection and processing of user-generated content through active participation [6]. *Crowdsensing*, *crowdsourcing*, and *participatory sensing* transform the digitally-enabled users into big data reliable generation sources for environmental monitoring [84]. Inside a Smart Space, software agents running on various computing devices transform these devices into smart objects exposed as participants from the real world [22]. Being that the multitude of information sources creates a shared information pool based on which the software agents interact, specific strategies for service composition in IoT-enabled smart spaces [85] and security concerns of personal mobile data and the intelligent utilization of IoT technologies are needed as well [86]. The problem of subscription notification loss in IoT-enabled smart spaces is addressed in [85]. Five mathematical models for active notification control possible strategies are described, in the case when the client checks on its own about the notification at certain moments rather than passively waiting to receive the notification from the Semantic Information Broker (SIB). A simulation model of a smart space is proposed to estimate the efficiency of the five proposed strategies, including loss assessment metrics. The experiments show that the three adaptive strategies perform well and reduce the loss rate but they require additional resources from the client's mobile devices. In this respect, the software development model in smart spaces uses specific tools, such as multi-source data fusion based on ontologies, knowledge reasoning, and semantic data mining.

Therefore, it becomes mandatory to consider the digital potentialities for advancing public safety services with user-generated data, possibly developing complex services to emphasize hazardous environmental phenomena based on different digital tools [17], [43], [44], [16]. This type of digitization of information favors the creation of shared resource

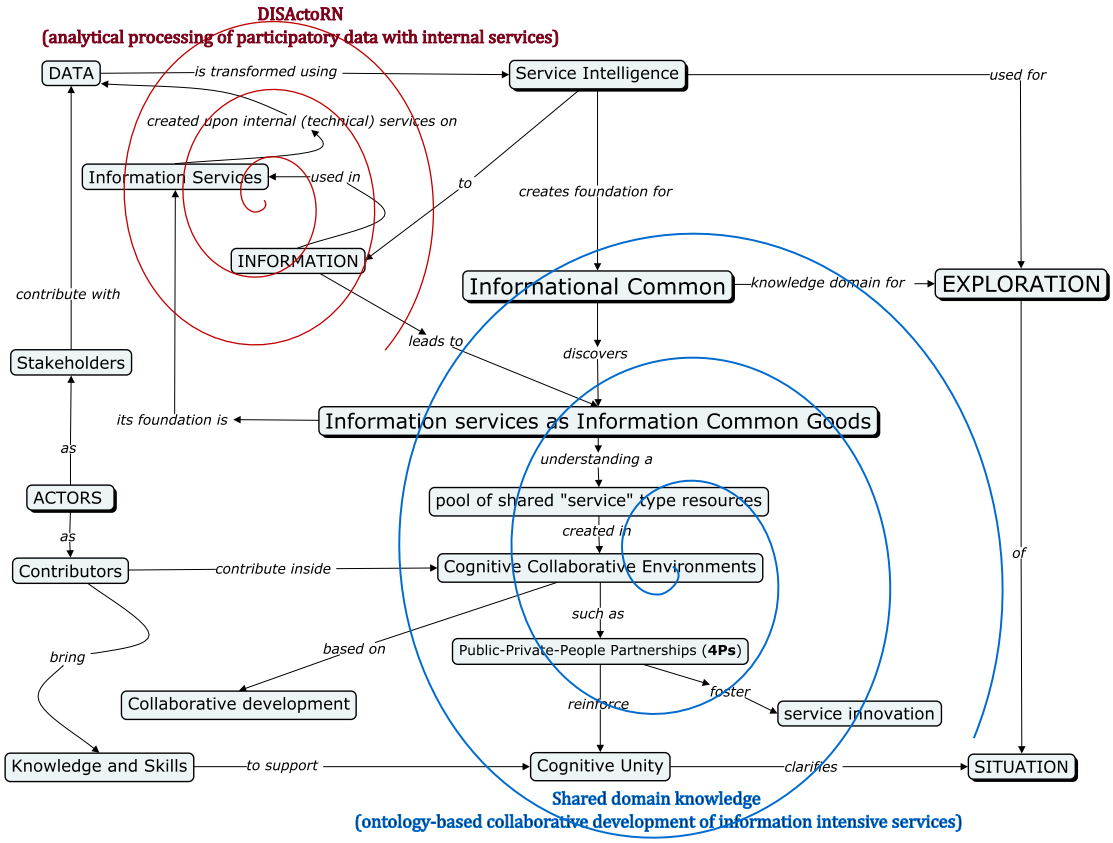

FIGURE 1. Concept map: Complex service ecosystem creation roadmap.

systems to conceptualize data, information, and knowledge as commons [87], [21].

III. WORKING METHODOLOGY

The new situations induced by the inclusion of technological advances in complex services development require both a new way of thinking to transform competencies, to mine expert knowledge, to upgrade digital skills [35]), and a need to redesign organizations based on cognitive social responsibility principles [88].

Henceforth, we formalize our working methodology with a service-related knowledge foundation, aiming to draw a general comprehensive unified representation perspective to develop, evaluate, include, and expose various participants' positions in public safety related activities. Our attempt follows previous other positions related to novel uses of the Digital for the wise management of natural resources, with participatory governance [89], towards citizen-centric data service development [90], [91].

Within the scope of this work we address the development of sustainable services, a concept by which we understand services capable of adapting to their environment and dynamically integrating changing conditions of this environment in such a way that they become sustainably coherent with the surrounding complexity [92]. They require contributory participation of the involved parties, henceforth named **A**ctors. As such, new ways of creating services targeting larger, more complex situations are needed and they require to unify Actors' confederate value co-creation initiatives around the development of services with cognitive unity [35].

Fig. 1 summarizes the roadmap towards the creation of an ecosystem of such complex services, that is judged as a value co-creation network [93] among all the participants in the city resilience-prone processes [29]. It is dedicated to engineering of information intensive services in public safety and their exposure as information common goods. This concept refers to "*rivalrous and non-excludable goods shared by and beneficial for all or most members of a community, or more precisely, the myriad of common goods, which serve the common interest and are free*" [35]. Their creation should be sustainable, environment-oriented, and enabled by the Actors' co-creation initiatives.

A complex service has several systems around, and one of them is the digital one. A complex service is built upon several services, and its systems, including the digital one, are composed upon their systems. In this perspective, the various participants' positions realizing the service activities can be described as organized manifestations of special interests [6]. Here, these manifestations are unified under two broad concepts, *Actors as Contributors* and *Actors as Stakeholders*, from whose interactions the service ecosystem emerges [94]:

- *Actors as Contributors* (people, government, private partners, or individual experts) contribute inside Cognitive Collaborative Environments, such as Public-Private-People Partnerships (4Ps) [25], Public-Private partnerships in disaster management [7], or Public-Private partnerships oriented towards the creation of services [37]. They bring in knowledge and skills to objectify ideas and to concretize their initiatives through services, therefore supporting cognitive unity that clarifies a particular situation;
- *Actors as Stakeholders* is a concept introduced within the scope of this work to address the two specific use cases of data collection and analysis in Cyber Physical Systems presented in Section II. They actively contribute with data upon which specific internal (technical) services act on to develop further the information intensive services in the working domain, by enlarging the analysis scope and providing the basis of future automation of the service integration process [44], [95].

The intentions and value propositions of a complex service provide a contextual framework in which the intentions and value propositions of the components of this complex service make sense and objectualize. To create a complex service in public safety, there is a need of multidisciplinary teams of creators to create the enabling systems, notably the digital system that will support the service itself. Globally, the intention of the digital system in a complex service is to make the activities of exchange (knowledge and skills) more efficient and to develop new activities.

The two main components of the digital system supporting the operationalization of the Public Safety as a Service vision through the development of information intensive service exposed as information common goods are presented in Fig. 1. One component addresses the creation of a shared domain knowledge integrating ontology-based collaborative development and semantic reasoning (see also Fig. 2). The second component aims to express the public safety service recipient's view on the creation of these sustainable services that would benefit from improved methods of data analytics. Within this component, henceforth named DISActoRN - the *Distributed Information Service Actor Role Network* - a sub-conceptual representation to unify the description of data collection and processing in Cyber Physical Systems is created (see also Fig. 3).

A. ACTORS AS CONTRIBUTORS

To explain how *Actors as Contributors* position in the digital system, we focus further on ontology development to promote collaborative knowledge sharing that formalize domain expertise, and to define sustainable service solutions able to assure the management of data, information and knowledge in Cyber Physical Systems. This is a type of Collaborative Development of sustainable services by means of Actors' creative and motivational application of competences through services [34]. Actors (private, public, individuals) from various knowledge domains assure the cognitive unity in service

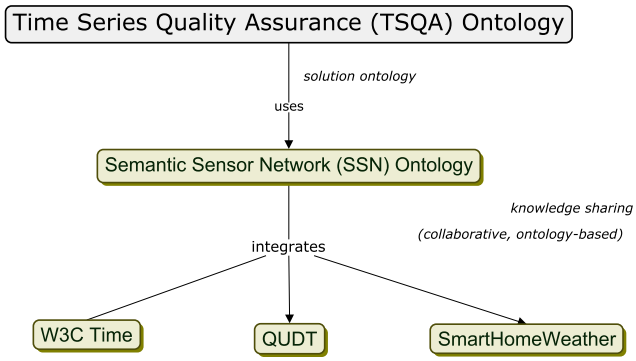

FIGURE 2. Collaborative development for knowledge sharing using the time series quality assurance (TSQA) solution ontology.

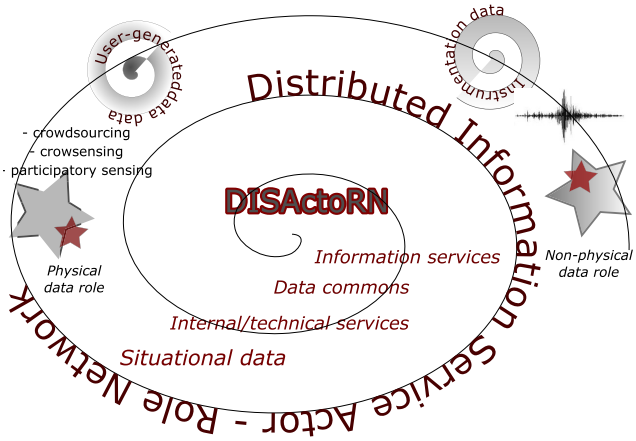

FIGURE 3. The distributed information Service actor - role network (DISActoRN).

creation, through their commitment to service development as a paying effort and having an exclusive right to shape the future domain development [35].

For our specific case, the following *domain expert knowledge* is required to mind rich participatory data and to capture the information richness by means of information intensive services. In the specific case of wise management of water resources and the evaluation of possible hazardous events addressing public safety concerns, this domain knowledge includes: description of sensors, description of observations (such as sensors measurements), description of quantity, units, dimensions and data types, description of time to take measurements, description of weather concepts. Description of sensors should address the device and its capabilities, the operating procedure, and the way the sensor would perform in a specified context. Description of the observations should include measured properties and the actual observed data. Description of time should provide a vocabulary for expressing date-time concepts, durations and relations between instants and intervals.

The conceptual architecture expressing knowledge sharing through Collaborative Development is based on the proposed Time Series Quality Assurance (TSQA) solution ontology that is presented in Fig. 2. It integrates specific domain

knowledge ontologies matching the above-mentioned criteria:

- **SSN** Semantic Sensor Network Ontology [96]. It is a unified ontology for sensors, measurements, and related concepts. It has a hierarchy of classes that offers a higher degree of flexibility in modeling sensors systems and subsystems, processes, deployment procedures, platform sites, measurements values, measurement capabilities and constraints, operating restrictions, etc. At the core of the ontology design is the *Sensor* that is a *Device* which is a *System* that can have as subsystems other systems. Each *System* has an *OperatingRange* that is constrained by some *Condition*. The *Sensor* observes a *Property* that corresponds to some *FeatureOfInterest*. An *Observation* is observed by a *Sensor* and has as result a *SensorOutput* that has value an *ObservationValue*. Each *Sensor* has a *MeasurementCapability* that is constrained by some *Condition*. The *System* has a *Deployment* that is realized on a *Platform*;
- **W3C Time** ontology [97]. It describes temporal concepts such as *Instant* (temporal concept with zero duration) and *DateTimeInterval*, being both specializations of a generic *TemporalEntity* class. An interval has a beginning and an end of type *Instant*, which in turn has a *DateTimeDescription* with a set of properties for representation of year, month, week, day, day of week (an enumeration of strings such as ''Monday'', ''Tuesday'', etc), day of year, hour, minute, second, timezone. Each *TemporalEntity* has a *DurationDescription* to represent decimals for years, months, weeks, days, hours, minutes, and seconds that together describe the duration;
- **QUDT** is the ontology for physical quantities, units of measure, and their dimensions in various measurement systems [98]. It supports the interoperability in a number of ways: units are defined in a non-ambiguous manner avoiding misinterpretation; it distinguishes between variants of the same unit (*day* may refer to solar day, sideral day, etc); it separately defines different units referred commonly by using the same word (e.g ''second'' for time and the same word ''second'' for measuring angles). Core concepts such as *Quantity*, *SystemOfQuantities*, *QuantityKind*,*QuantityValue*, *Unit*, *SystemOfUnits* are defined;
- **Smart Home Weather** is the ontology for weather phenomena and exterior conditions [99]. It offers a vocabulary for modelling weather related data. *WeatherState* has condition a *WeatherCondition* and belongs to a report *WeatherReport* that has a source *WeatherReportSource*. The *WeatherState* has a phenomenon *WeatherPhenomenon*. These five classes are top level concepts. Each of these have sub-concepts defined. For example, sib-concepts for *WeatherState*, *HotWeather*, *CloudyWeather*, *RainyWeather*. The specialization is realized further on several layers.

B. ACTORS AS STAKEHOLDERS

Within the specific scope of this work, we introduce a sub-conceptual representation to unify the description of data collection and processing in Cyber Physical Systems and we formulate two definitions to describe several aspects of data collection through various measurements and instrumentation. This representation, DISActoRN - *Distributed Information Service Actor Role Network* (Fig. 3) - aims to guide the development of the internal (technical) services to assess Data Quality for the specific situation in water resources management described in Section II.

Definition 1: A *Non-physical Data Interaction Role* (NFD-IR) refers the whole set of without-rights devices (such as instrumentation sensors, mobile devices) and with-rights legacy systems (such as automated water pollution data collection and analysis systems) actively participating to the extension of data-centric processing tasks.

Definition 2: A *Physical Data Interaction Role* (FD-IR) refers the whole set of social participants that are actively involved in *crowdsourcing*, *crowdsensing*, and *participatory sensing* activities.

DISActoRN functionality can be described based on two main use cases: `Collect Hazard Related Data` (based on the afore-mentioned interaction roles) and use internal (technical) services to `Transform Hazard Related Data` to ''data products'' to be deployed back into the Real World by means of specific data cleaning information intensive services.

This type of data collection and digitization of information supports the virtualization of various sources of data to provide a unified view on the service activities.

The data resources in the DISActoRN construct are contributed by *Actors as Stakeholders* acting in various roles supporting the resource system provisioning actions. The data pool is mainly produced cumulatively by sensing, within the scope of sharing domain resources to act upon and to efficiently improve resilience processes. Eventually, they will be perceived as community members empowered through pervasive and ubiquitous digital infrastructures, integrating and analyzing data from multiple sources, leading to the improvement of real-time response to fine-tune the delivery of public safety services.

C. EMBEDDING SEMANTIC SERVICES

The components of the proposed architecture aiming for the implementation of the DISActoRN internal (technical) services and its integration with the semantic services through the TSQA solution ontology are presented in Fig. 4.

This architecture integrates components corresponding to detection and correction methods, data acquisition from sensors in various file formats, and semantic processing based on the TSQA and SWRL-encoded rules [100]. These internal services provide the unified access to the data products generated by the data acquisition and analysis software running in the enterprise information system.

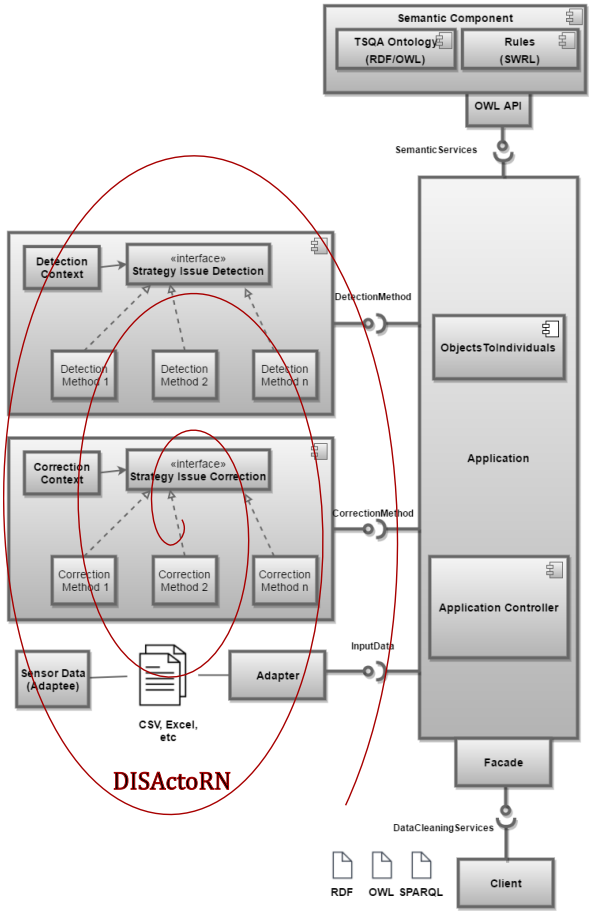

FIGURE 4. DISActoRN - internal (technical) services implementation details: integration with semantic services.

- The *Semantic Component* manages the TSQA ontology and the associated SWRL rules. It will expose the `SemanticServices` interface to fulfill services invoked by the specific application;
- The *Detection Component* is responsible for maintaining and selecting an Issue Detection Method (IDM). A Strategy Design Pattern (SDP) is used in this component to define a family of algorithms, to encapsulate each one, making them interchangeable, and to decouple the client from the algorithm implementation. Detection Method *i* represents a concrete implementation of a generic method (for example, outlier detection using clustering, statistics, distance-based methods);
- The *Correction Component* provides the `Correction Method` interface that is requested by the *Application Component*. It provides a concrete implementation of a Correction Method (for example, Cubic Spline for Interpolation) and it conforms to the same SDP;
- The *Application Component* requires that the corresponding sensor data, acquired in different file formats (such as CSV, tab delimited, Excel, GSF, SIDF), are converted into its specific internal format, and an Adapter Pattern (AP) is used. The Adaptee represents the actual sensor data. The Adapter converts these data and provides the `InputData` interface.

In the *Application Component* (here implemented in J2EE, running on an Apache Tomcat Server), the Application Controller orchestrates the workflow and method invocations. The data read through the `InputData` interface is parsed and for each sensor measurement new individuals connected through object properties are created and inserted into the *Semantic Component*. The reasoner is then asked what detection methods are available for the input data. The end-user either selects one of the detection methods or accepts the default method. Then the data set is sent to be checked against that specific detection method.

When the detection algorithm completes its execution, the input is categorized either as suspect or correct. If suspect, the associated tag of the data is updated accordingly in the *Semantic Component*. According to the rules and the facts that already exists, the reasoner is asked to infer whether the suspect data is TP or FP (described further in Section IV), and to update the tag accordingly. If TP, then the reasoner is asked to list the available correction methods for that problem. The user either selects one correction method or accepts the default one. Then the correction method is executed for the given data and returns a corrected set of data.

Corrected values are set on the corresponding individuals defined in the *Semantic Component*. The *ObjectsToIndividuals* component is part of the *Application Component*. It creates TSQA individuals from Java objects then serialize these in the TSQA ontology (either in RDF or OWL format). The Java objects are created after sensor data is read from file.

The *Facade* module hides the internal complexity of the internal services and provides the `DataCleaning Services` interface used by the end-user to communicate with the *DataCleaning Application*. The end-user can select a specific detection/correction method, retrieve the results, input new facts uploading RDF or OWL files (e.g. describing a storm that was produced), and execute queries (via SPARQL query language) to retrieve information stored in the TSQA ontology.

The engineering process to construct the Time Series Quality Assurance (TSQA) solution ontology and the workflow describing the main activities are presented in Fig. 5. The ontology should be generic enough to be applicable for any type of Data Quality deficiency, being able to distinguish between FP and TP data objects for time series corresponding to sensor measurements.

IV. DATA CLEANING. EVALUATION CRITERIA FOR DATA QUALITY IN TIME SERIES

This section explains the DISActoRN functionality on the specific case study, focusing on the interactions related to the Non-physical Data Interaction Role (NFD-IR) for the evaluation of data quality in time series for water resources management and the design of the corresponding internal services. We assume that the observed data are available as a time series whose values are available for a specific time

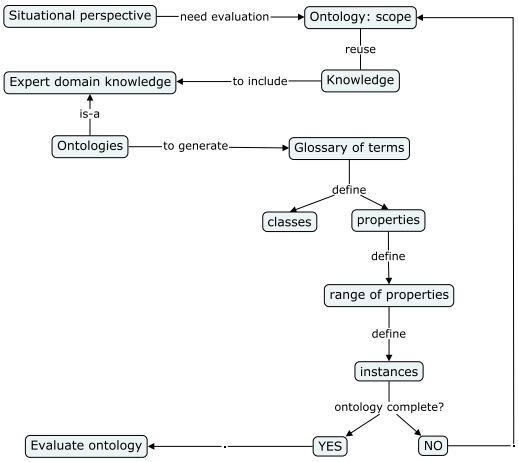

FIGURE 5. The time series quality assurance (TSQA) solution ontology engineering process, adapted from [101].

range. As well, metadata describing the measurement devices and the procedures followed to obtain the data on the field must be available.

Let $\mathbf{O}$ denote an observable domain and a physical quantity $\mathbf{P}$ characterizing a phenomenon that manifests in $\mathbf{O}$. Let $\mathbf{X} = (x_1, x_2, \ldots, x_i, \ldots, x_n)$ be the time series that represents the measurements for $\mathbf{P}$, at time instants $(t_1, t_2, \ldots, t_i, \ldots, t_n)$, where $t_{i+1} > t_i$. Let `dataObj` (data object) be either an individual value x_i or an ordered subset of values $X = (x_i, x_{i+1}, \ldots, x_j)$, where $X \subset \mathbf{X}$.

Different types of wrong data objects may occur in the measured data represented as time series. Common types of anomalies in sensor include spikes, outliers, data not in range, data blocked at constant value, data gaps, null values, mean shift, or wrong timestamps (for example, [102]):

- *Spikes* (peaks, or local maxima). They are points x_i in the time series for which $y_i = f_p(x_i) > \theta$, where f_p is a "peak function" that associates a positive score to the argument, and θ is a user defined threshold. One of the challenges discussed in the literature is proposing a formal definition for the peak function f_p and to evaluate the results [103];
- *Outliers* are data objects that behave far away from the expected behavior. In many domains, such as fraud detection, industrial processes, public safety, healthcare, environmental resources management, the detection of outliers is a critical challenge. It is worth to be noted here that the presence of noise, a random error or variance that occurs in a measured variable, makes difficult to recognize outliers;
- *Data not in range* is a term referring data positioned outside an admissible interval. For example, a bug in the firmware can determine the omission of a decimal point, the value measured by the sensor being correct, while the wrong value is being written in the logger. The admissible interval $[V_{min}, V_{max}]$ can be either user defined (it can be flagged as a soft error and resolved by interpolating N values, or an offset adjustment is applied when a constant bias through time is observed) or defined in the sensor specification;
- *Data blocked at constant value* represent examples of false constants occurring, for example, if the measurement device reported the same value for several consecutive measurements. In this case, for the peak function, f_p, it is advisable to look for values close to 0, finding its minimum $\min_{x_i} f_p(x_i)$. If there is an x_i for which $f_p(x_i) < \mu - 3\sigma$, this indicates with a probability of 99.85% that data is frozen at the constant value x_i in the time windows defined by 2k points. Alternatively, we may choose $\theta =$ resolution of the measurement device, such as to use $\theta = \mu + 3\sigma$ or other statistical indicators;
- *Data gaps* are two or more missing values in the time series. The occurrence of this situation can be checked based on timestamps, and sampling may be done using a specific algorithm (e.g. at constant interval of time). For example, a defect connection between a sensor and the data logger can result in an increased number of dropped data points;
- *Null values* are measurement values reported as 0.0. They can suggest flag a problem of the instrument. In other cases, they can represent real natural phenomena so, again, as for spikes, they cannot be rejected automatically, instead they can be flagged as suspect values. In these cases, further information is needed to decide whether is a sensor problem;
- *Mean shifts* correspond to changes of the mean on some intervals (segments). This problem was studied in literature for signal processing and time series, in general, and for environmental regime shift, in particular [104]. For example, significant changes of the mean can give important indications in the context of water quality monitoring pointing either an increasing/decreasing of pollution or a possible problem with the sensor;
- *Noise* refers to random fluctuations in observed data. In general, smoothing methods are applied (based on Kalman filter, or exponential smoothing known also as "Holt Winters smoothing");
- *Oscillations* are intervals of data that presents high deviation from the mean. As in case of noise, the solution consists in applying smoothing techniques;
- *Wrong timestamps* are data objects with a corrupted timestamp (indicating other instant of time than the measurement time). This type of data deficiency can be corrected using a counter that is incremented each time a new measurement is taken;
- *Surrogates* refer to the situations when chemical/physical parameters of interest, rather than being directly measured, can be deduced from other measured parameters (surrogates) using regressive equations or neural networks [105], [106].

Currently, statistical-based automated spike/peaks detection methods consider that a time series is faulty (for example, containing wrong spikes caused by an interference in the

communication channel) when more than m% spikes are detected in a Δt time interval. "Rule 68-95-99.7" states that 68% of the data with a normal distribution must be located within one standard deviation (σ) of the mean (μ), 95% within two standard deviations, and 99.7% within three standard deviations [107]. Considering $\theta = \mu + 3\sigma$, then only 0.15% of the x_i data will show such a high value of the function $y_i = f_p(x_i)$. This implies that x_i can be considered a candidate to be marked as a "spike".

However, these automated methods cannot discriminate true spikes from the false ones. Real peaks can appear in special circumstances, but the automated methods such as the statistical method discussed above does not consider the context. For example, algal blooms can grow on the water surface in certain seasons, a natural phenomenon that generates peaks in measuring the values of chlorophyll. A decrease of conductivity and an increase of turbidity may appear in case of rain. If a peak for turbidity is detected in the context where no correlated decreasing of conductivity is observed, then it may be concluded that there it is due a sensor problem.

Consequently, it is highly important to have access to *domain expert knowledge* and this knowledge must be correlated with the automated methods. Relying only on automated methods in this case could lead to discarding correct and valuable data misinterpreted as "peaks". Furthermore, it could lead to wrong analysis and consequently to wrong decisions recommended to stakeholders. This expert domain knowledge can be encapsulated in a format understandable by machines, namely ontologies and IF/THEN rules.

Measured data objects can be classified as:

- *True Positive* (TP). An error in measured data that is correctly classified as an error is called a True Positive (TP);
- *True Positive* (FP). A correct data object that is wrongly classified as an error is called a False Positive (FP) or type I error;
- *True Negative* (TN). A correct value that is classified as a correct value is called a True Negative (TN);
- *False Negative* (FN). An error that is wrongly classified as a correct data is called a False Negative (FN) or type II error.

Concerning the general problem of Data Quality, three possibilities can appear if wrong data objects are identified:

- the Data Quality problem is caused by a minor malfunction of the sensor that appears for a small number of times. In this case the solution is to apply a correction procedure that replace the data with interpolated values;
- the Data Quality problem is caused by a serious malfunction of the device that is observed repeatedly (or on a long time-window). The user should be informed that there is a defective sensor which should be removed (and all its collected data rejected);
- the Data Quality problem is a FP. In fact, is a true value but it manifests only in extraordinary circumstances (it is not a typical behavior). The data object is considered FP, but it is reported to the user. The identification of such objects is possible only using semantic technologies to differentiate it from the TP.

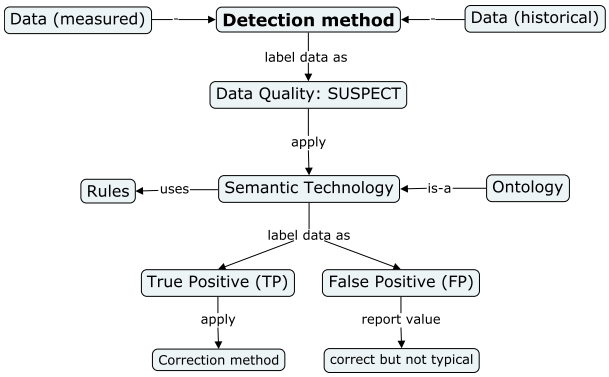

FIGURE 6. Unified semantic reasoning-based approach to improve quality of data acquired by sensors.

The proposed unified semantic reasoning-based approach aiming to dissociate between TP and FP is presented in Fig. 6. It consists of three layers of data processing. It is important to emphasize here that the distinction between TP and FP can be done only using domain expert knowledge, otherwise a FP is automatically assimilated to a TP.

Currently measured or historical data are introduced in the first layer. Then automated methods for detecting wrong data are applied (such as statistical methods for spike detection, distance-based algorithms for outliers, etc.). The term "wrong data" applies to any category of anomalies discussed above (such as data gaps, outliers, data not in range, spikes). The output of this first layer consists of data objects labeled as SUSPECT (i.e. outlier suspect, spike suspect, etc.). At this moment it is still not possible to certainly assert that data are indeed wrong. It is necessary to execute a second check using domain expert knowledge, formalized based on ontologies and rules, that allows taking into consideration the context of the data objects measurement (such as weather condition, location, device). After this processing phase it will be possible to decide if the SUSPECT data is a TP data or a FP data. A corresponding label/tag will be applied to the data object and it will be sent as input to the third layer. Here, the correction method is applied for TP (depending on the type of error detected this procedure can be interpolation, rejection, surrogates, etc.). In case of the FP data, this will be reported to the user as a correct value but not as a typical value (depending on the context, it can mean, for example, a breakdown of an industrial machine or an accidental pollution).

The proposed unified semantic reasoning-based method for improving data quality is presented in **Algorithm 1**. Let **D** be the set of detection methods and $d \in \mathbf{D}$, a particular detection method. The set of correction methods is denoted by **C** and $c \in \mathbf{C}$ is a particular correction method (**C** and **D** are classes of algorithms and c and d are algorithms, namely instances of **C** and **D**). Let `S(Ont,R)` be a semantic technique defined by the ontologies set `Ont` and the rules `R`.

TABLE 1. TSQA ontology: Requirements specification.

Test	Problem verified	Possible tags to be applied	Solution	Threshold	Quality criteria addressed
T1	Spikes	SUSPECT, TP, FP, CORRECTED	Replace with interpolated	Statistic	Accuracy
T2	Blocked at constant value	SUSPECT, TP, FP, CORRECTED	Reject	Statistic / Resolution of the instrument	Accuracy
T3	Outliers	SUSPECT, TP, FP, CORRECTED	If threshold is user defined then interpolate, if it is specification defined then reject	User defined / Specification	Accuracy
T4	Gaps	SUSPECT, TP, FP, CORRECTED	Reject OR Replace with best fit spline segment OR Use surrogates	User defined / Statistic	Completeness
T5	Null	SUSPECT, TP, FP, CORRECTED	If T2 fails then Reject	0.0	Accuracy
T6	Mean shift	SUSPECT, TP, FP, CORRECTED	Reject both segments	User defined	Consistency Accuracy
T7	Oscillation	SUSPECT, TP, FP, CORRECTED	Apply smoothing	User defined	Accuracy
T8	Noise	SUSPECT, TP, FP, CORRECTED	Apply smoothing	autocorellation values = 0	Accuracy
T9	Missing or wrong timestamp	SUSPECT, TP, FP, CORRECTED	Compute timestamp based on Counter and interval of sampling	N/A	Timestamped Consistency Traceability

Algorithm 1 Unified Method To Improve Data Quality

```
1:  function improveQuality(dataObj,a) ▷ input data object and anomaly type a
2:      d ← getDetectionMethod(a);
3:      label1 ←d(dataObj);
4:      if label1 ≠ "SUSPECT" then ▷ data object is correct
5:          return "TN"
6:      else ▷ d method identified a problem
7:          label2 ←S(Ont,R,dataObject) ▷ apply semantic technique
8:          if label2 = "TP" then
9:              c ←getCorrectionMethod(a);
10:             c(dataObj); ▷ apply correction method
11:             return "TP"
12:         else ▷ label2=FP
13:             reportToUser(dataObj,"FP")
14:             return "FP"
15:         end if
16:     end if
17: end function
```

The algorithm receives as input the data object (`dataObj`) and the anomaly type to be checked (for example, "outliers", "gaps", "noise"). It generates as output a label (FP, TP or TN) and applies a correction method for data labeled as TP. Lines 2-5 correspond to the first layer and the method `GetDetectionMethod(a)` returns the detection method applicable to the anomaly type a (for example IF a == "Outliers" THEN the returned result will be the object that encapsulates the "Distance based outliers' detection" algorithm).

If more than one detection methods are available then a second parameter for this method, `pref`, is passed (like in `GetDetectionMethod(a,pref)`). If several options are available, then the `pref` object encapsulates the preferences of the user for one or another algorithm. Line 7 corresponds to the second layer applying inference to decide whether the data object is a TP or a FP. In lines 8-16 (the third layer) appropriate action is taken (either correct the data or just report to the user).

The method `GetCorrectionMethod(a)` returns the algorithm that can correct the anomaly `a`. A preference object can be used as described above when several algorithms are available (such as various smoothing algorithms for noise correcting). Once retrieved, the algorithm c is applied to the `dataObj` input.

V. TSQA: A SOLUTION ONTOLOGY FOR IMPROVING DATA QUALITY

This section describes the engineering process to construct the Time Series Quality Assurance (TSQA) solution ontology. The core concepts identified for the TSQA ontology are presented in Fig. 7, and the instantiations of the main concepts used further in the experimental and validation phase are presented in Fig. 8 (classes are depicted with brown circles and instances with violet diamonds).

Following the discussion in Section IV, Table 1 summarizes the main quality concerns (criteria) to be addressed, a number of quality problems to be checked, what solutions can be applied for wrong data, what labels can be applied to the data during the data cleaning workflow, and what type of threshold is required by the detection algorithm (for example, for outliers detection a threshold must be supplied, but for spikes the threshold can be computed internally based on mean and standard deviation). Traceability refers to the propriety of data to be accounted for its provenance at each time step (this goal is achieved in our solution by using labels/tags to mark if data is measured, suspect, corrected, or rejected).

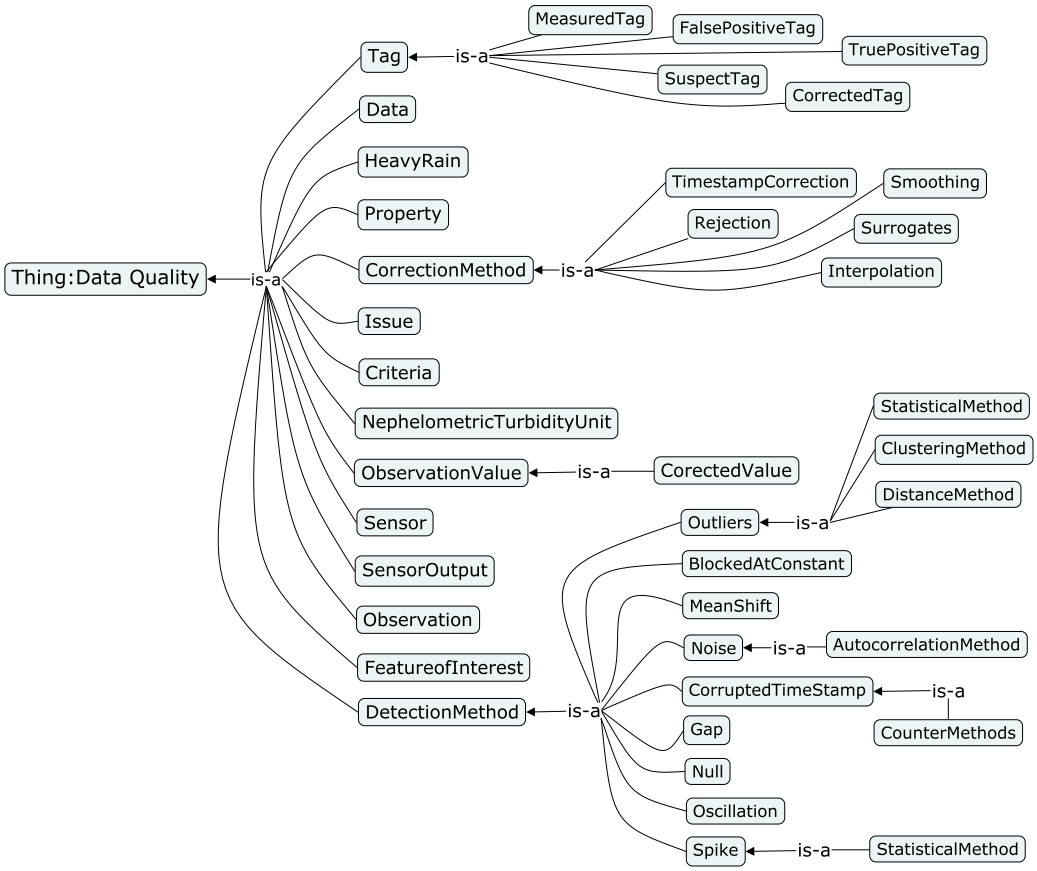

FIGURE 7. TSQA ontology: Concept map.

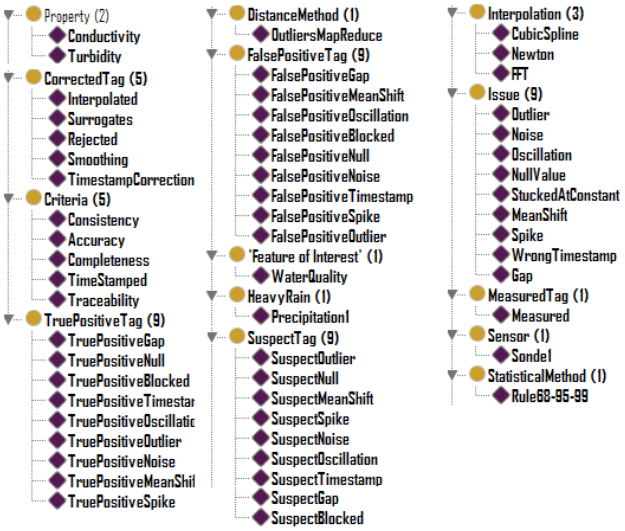

FIGURE 8. TSQA ontology: Individuals by type (Protégé representation).

These requirements are used further to identify the main concepts for the TSQA ontology. We consider the following notations: classes (concepts) that are reused from other ontologies are marked bold underlined and prefixed by the abbreviation of the ontology (like **ssn:Observation**); classes that are defined in our proposed ontology are marked bold (like **Tag**); the properties we define are marked in italics (like *ssn:hasTag*). The instances (individuals) for a given class (concept) are represented with bold italics (like ***Accuracy***).

The **Criteria** concept refers to quality indicators that data should met. The tag refers to the data cleaning process status that an instance of type **ssn:Observation** can take. The **Tag** concept is linked to **ssn:Observation** concept via the *hasObservationTag* property. For the **Tag** concept, the following sub-concepts were defined (**MeasuredTag**, **FalsePositiveTag**, **TruePositiveTag**, **SuspectTag**, **CorrectedTag**) because one observation can go through different stages (measured, problem detection, problem correction), so we need to apply labels to the data object according to these phases.

The **Issue** concept refers to the quality problem that needs to be checked for and has nine instantiations.

Because a given data object may have more than one data issue at the same time (for example, could be suspected to be an outlier and have wrong timestamp in the same time), then Suspect, True Positive, and False Positive tags define 9 individuals, one for each of the defined Issue. Each **Issue** addresses one or more **Criteria** (e.g. issue WrongTimestamp addresses criteria WrongTimestamped and Consistency).

The **Data** concept is the core concept in the TSQA ontology, referring at least one object of type **ssn:Observation**,

TABLE 2. TSQA ontology: Data properties.

Name of property	Domain	Range	Inverse of	Cardinality restrictions
addresses	`Issue`	`Criteria`	*isAddressedBy*	1 `Issue` addresses min 1 `Criteria`
contains	`Data`	`ssn:Observation`	*isIncludedIn*	1 `Data` contains min 1 `Observation`
corrects	`CorrectionMethod`	`Issue`	*isCorrectedBy*	1 `CorrectionMethod` corrects min 1 `Issue`
discovers	`DetectionMethod`	`Issue`	*isDiscoveredBy*	1 `DetectionMethod` discovers min 1 `Issue`
hasCorrectedValue	`ssn:Observation`	`CorrectedValue`	*isCorrectionOf*	1 `ssn:Observation` hasCorrectedValue min 0 `CorrectedValue`
hasObservationTag	`ssn:Observation`	`Tag`	*isObservationTagOf*	1 `ssn:Observation` hasObservationTag min 0 `Tag`
hasSetTag	`Data`	`Tag`	*isSetTagOf*	1 `Data` hasSetTag min 0 `Tag`
hasInput	`Method`	`Data`	*isInputTo*	1 `Method` hasInput min 1 `Data`
hasOutput	`Method`	`Data`	*isOutputOf*	1 `Method` hasOutput min 0 `Data`
corrupts	`Issue`	`Data`	*isCorruptedBy*	1 `Issue` corrupts min 0 `Data`

thus having the capability to represent either an individual measurement or an entire time series. The **Data** is considered as input for the detection algorithms. They are represented by the class **DetectionMethod** whose sub-classes address each of the issues, **Outliers**, **Spike**, **Noise**, etc. Further sub-classes may be defined, for example, for outliers detection: **DistanceMethod**, **ClusteringMethod**, **StatisticalMethod**, etc. Noise class has as sub-class **AutocorrelationMethod**. **Dat**a instances that are considered by a **DetectionMethod** to be wrong (for example, outliers detected by a distance-based detection algorithm) are marked further with an instance of **SuspectTag**.

When SWRL-encoded rules [100] are applied to decide whether the suspect data object is a FP or a TP, the **Data** instance may receive a **FalsePositiveTag** or a **TruePositiveTag**. In the second case, the **Data** will be processed by a **CorrectionMethod**, according to the particular instance of the received tag.

For example, if the tag is ***TruePositiveSpike*** then the correction method will be one instance of **Interpolation** (such as ***CubicSpline***). The **ssn:Observation** has *ssn:observationResult* a **ssn:SensorOutput** that *ssn:hasValue* an **ssn:ObservationValue** which actually stores the value of the sensor measurement.

We need to introduce in our TSQA ontology a new property that will link the **ssn:Observation** with the corrected value (that results from the application of one of the **CorrectionMethod**). For this we created the new class **CorrectedValue** as a sub-concept of **ssn:ObservationValue** and link it to the **ssn:Observation** via the *hasCorrectedValue* property.

The object properties are listed in Table 2. All of them have an inverse property that starts with the word "is…". The cardinality restriction was implemented in Protégé with the Object restriction creator feature that allows to define classes as property restrictions and then to assign the initial class as a sub-class of the property restriction class (e.g. **Issue** subclassOf *addresses* min 1 **Criteria**).

Fig. 9 depicts the integration of TSQA main concepts with SSN concepts as well as the relations between concepts (green rectangles depicts instances connected by doted arrows to their corresponding yellow classes). Choosing the SSN ontology presents also the advantage that it is a unified ontology for both sensors and measurements so it is not necessary to integrate the vocabularies of two distinct ontologies.

VI. SUMMARY OF THE ACHIEVED RESULTS

The proposed unified semantic reasoning-based method for improving data quality is evaluated against a water quality monitoring use case scenario, on a real data set collected from the water monitoring plant in South Branch Tunkhannock Creek, Lackawanna County, Pennsylvania, publicly available for download [108]. A set of water quality parameters, including Turbidity and Conductivity, are measured using an YSI 6920 sensor. Turbidity is a measure of water clarity measured in Nephelometric Turbidity Units (NTU), whose level increases with the level of sediments in the water. Conductivity is a measure of water to pass electrical flow. Significant variations of conductivity are an indicator that a pollutant was released in the water (measured at $25°C$ in milliSiemens/cm).

For this use case scenario, a dataset consisting of 111 sensor observations (15 minutes time stamp) was considered. The data set includes measurements of Turbidity and Conductivity and measured values for Temperature, Dissolved Oxygen, pH, etc. Specifically, we are interested in evaluating only outlier values for Turbidity and Conductivity. These values are represented graphically along with a global trend (computed for the entire time series) in Fig. 10.

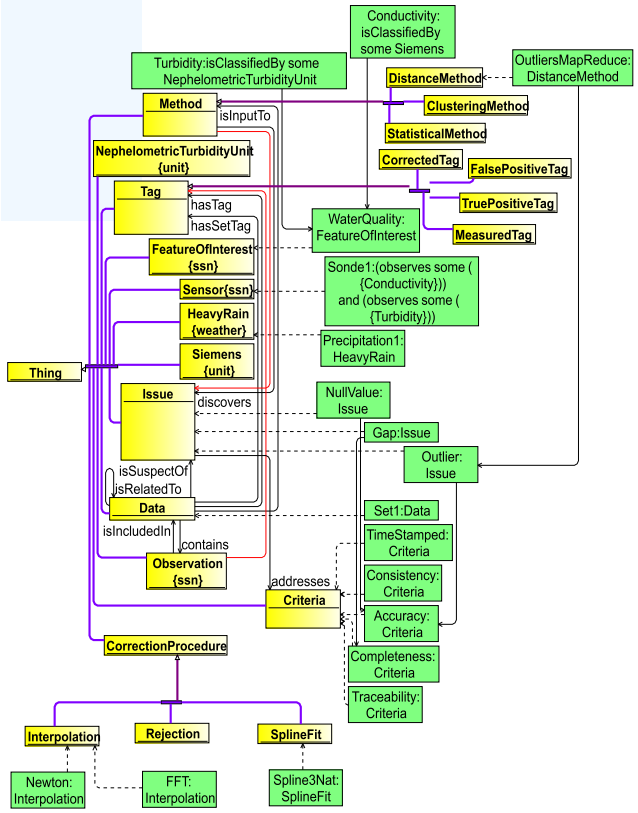

FIGURE 9. Integration between concepts in SSN sensor's ontology and data cleaning ontology.

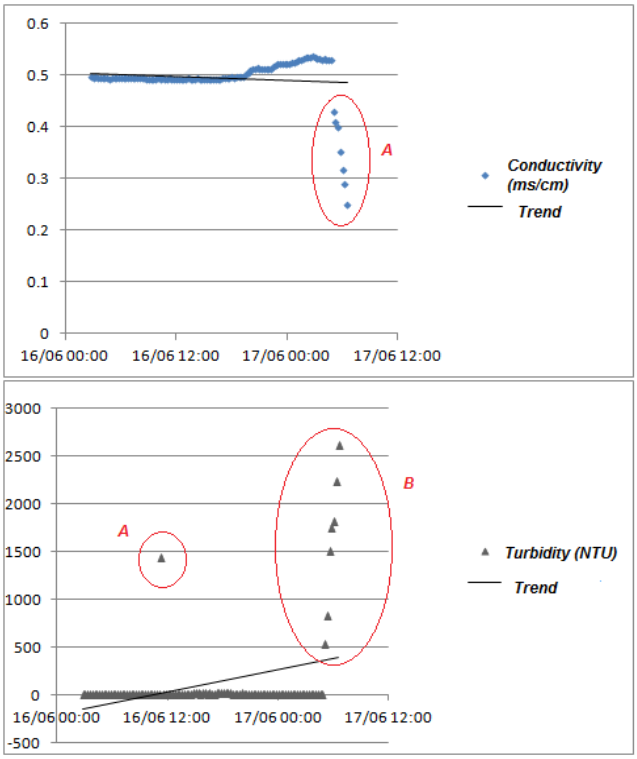

FIGURE 10. Detection of FP and TP outliers based on semantic reasoning using the TSQA ontology.

For our use case, two specific rules (R1 and R2) have been formulated to check whether the measurement context refers exceptionally high values for turbidity can occur, but they are not outliers.

1) **R1**. *If there is an observation for Turbidity that is suspect of being outlier and that observation is taken during a heavy rain and the Turbidity has an ascendant trend then the observation is not an outlier (it is a FP).*
2) **R2**. *If there is an observation for Conductivity that is suspect of being outlier and that observation is taken during a heavy rain and the Conductivity has a descendant trend then the observation is not an outlier (it is a FP).*

Listing 1: SWRL rules for assessment of false positive outliers

```
HeavyPrecipitation(?x),RainyWeatherState(?s),
Interval(?int),hasObservationTime(?s,?int),
Observation(?o),observedsProperty(?o,
Turbidity),
hasObservationTag(?o,SuspectOutlier),
hasInXSDDateTime(?o,?instant),
temporal:contains(?int,?instant),
hasTrend(?o,Upward)-> hasObservationTag(?o,
FalsePositiveOutlier)
HeavyPrecipitation(?x),RainyWeatherState(?s),
Interval(?int),hasObservationTime(?s,?int),
Observation(?o),observedsProperty(?o,
Conductivity),hasObservationTag(?o,
SuspectOutlier),hasInXSDDateTime(?o,
?instant),
temporal:contains(?int,?instant),hasTrend(?o,
Downward)-> hasObservationTag(?o,
FalsePositiveOutlier)
```

Both SWRL-encoded rules refer to outlier suspects for Turbidity and Conductivity time series to check whether a heavy precipitation manifests at the data measurement time, and whether the Turbidity, respectively Conductivity, values have an upward/downward trend. The term trend refers to the long-term increase/decrease in the data values.

For the clarification of the term "heavy rain" in the above rules, we rely on the Smart Home Weather ontology, where the "heavy rain" weather term is precisely defined as being the phenomenon whose precipitation intensity has a value in the range of [20], [50] mm/hour, and the precipitation probability measured in % is expressed as a positive float value.

To express the exact date, time, and duration of the a heavy rain, we need to define an instance ***prec1*** of the class **weather:HeavyPrecipitation**, and an instance ***rainystate*** of the class **weather:RainyWeatherState** (***prec1*** is linked via the property *weather:belongsTo* to the ***rainystate***). The ***interval1*** instance of class **time:DateTimeInterval** is linked with the ***rainystate*** via the property *weather:hasObservationTime*. Two instances (***i1*** and ***i2***) of **time:Instant** are defined to represent the start and end time interval. These are linked with

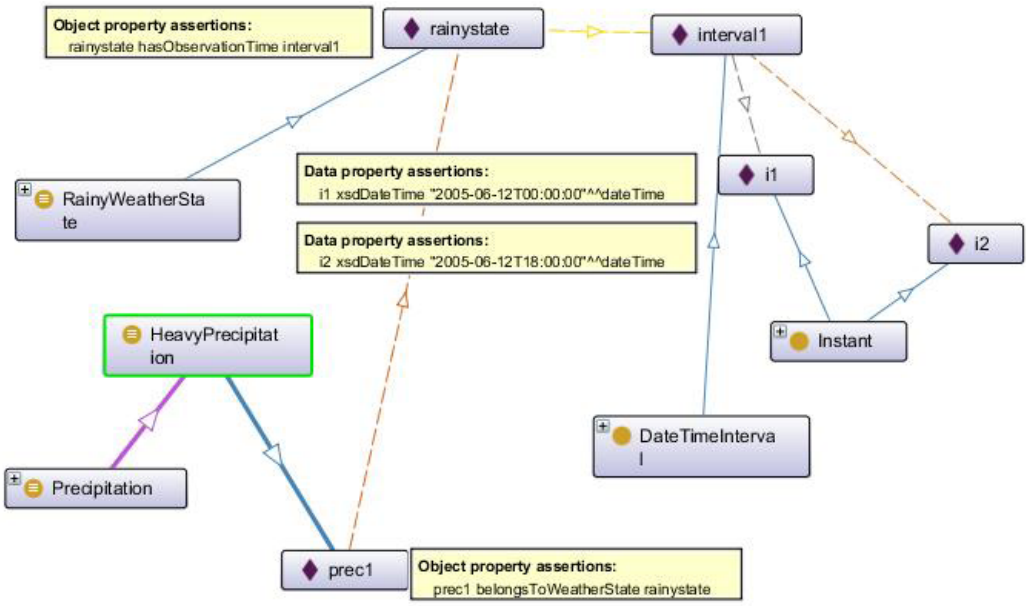

FIGURE 11. Expressing the "heavy rain" weather phenomenon.

the data property *xsdDateTime* to the time label representing a specific moment in time. The relations between individuals for the description of a heavy rain are presented in Fig. 11.

The user-generated RDF-XML document describing the fact that a heavy rain was produced is passed via the `DataCleaningServices` interface to import the file into the existing TSQA ontology. The measured data from the Excel file is filtered by the *Adapter Component*, allowing to pass only the interest values corresponding to Turbidity and Conductivity measurements as a vector of Java objects for the *Application Component*. Later, these are translated into TSQA individuals and linked with corresponding properties, using the OWL API. The internal services execute a query upon TSQA ontology to find what methods are available for issue detection and list them to the user.

As it can be observed in Fig. 10, outliers occur for both Turbidity and Conductivity, mainly in the same time interval. For Turbidity, two different types of outliers may appear: one isolated measurement (noted with A on figure), and a group of continuously increasing outliers (noted with B). Without any extra intelligence, any outlier detection algorithm would treat the same way these two different types of outliers, possibly considering them wrong data, and it will decide to correct them. The data measurements refer to a 28 hours interval between 02:45 AM and 06:30 AM next day. If we know the fact that at 4:50 a thunder heavy rain started and last for 3 hours, we can represent this fact in the *Semantic Component*. This extra knowledge allows by using the proposed internal services to draw the conclusion that the outliers group for Turbidity and Conductivity overlaps with the storm.

The user selects a method for OutlierDetection and start execution. When the outlier algorithm completes, it will detect group A of 7 outliers for Conductivity, and 8 outliers for Turbidity as group A and group B. Therefore, all the 15 values are tagged as ***SuspectOutlier***. For each observation, a new triple in the form ***Obs_MeasuredProperty_i*** *hasObservationTag* ***SuspectOutlier*** is added in the ontology via the `SemanticServices` interface. At this moment, the data cleaning application runs the TrendDetection method. The result shows that the trend for Conductivity is downward, whereas for Turbidity the trend is upward. New triples are then inserted accordingly: ***Obs_Turbidity_i*** *hasTrend* ***Upward*** and ***Obs_Conductivity_i*** *hasTrend* ***Downward***.

If the Pellet reasoner [109] is invoked at this point, the antecedent of both rules evaluates to *true* for 14 values (the outliers in group A for Turbidity and the outliers in group A for Conductivity). For the single outlier in group B, the antecedent evaluates to *false* (*temporal:contains()* evaluates to *false* because that observation is not realized during the heavy rain). Then the SWRL rules are executed, and new property assertions are inferred, ***Obs_MeasuredProperty_i*** *hasObservationTag* ***FalsePositiveOutlier***.

Because negation as failure or modifying ontology facts are not supported by SWRL, we need new rules to assert that suspects that are not given FalsePositive tags should be labeled as TruePositive. In conclusion, the 14 values, marked as false positives are not removed and the only true positive is replaced with an interpolated corrected value.

The unified semantic reasoning-based method evaluated in this section allows to relate sensor observations with meteorological context and with methods for detection and correction of data issues, in three stages:

1) *Execute algorithms* for data issues detection and mark as suspect values that are found to have problems;
2) *Use* contextual *domain knowledge* (such as meteorological situations expressed in rules) and apply the reasoner to label accordingly the data values that proves to be false positives;

3) *Correct* all *data* that was marked in the first stage but not in the second (true positives) using one of the suitable correction methods.

This working methodology has the advantage that it considers the meteorological context of measurements and helps in dissociating between true sensor data problems and exceptional values that can occur in special contexts, but they are not wrong data.

Therefore, it is possible to expand knowledge on the identified informational situation through contributory development using semantic reasoning, based on the necessity to create a unified approach to bind together observations from sensors, methods for automatic detection of erroneous data, domain-specific knowledge, and correction procedures.

VII. DISCUSSION AND FURTHER DEVELOPMENTS

The working methodology employed in this paper defines the main activities aiming to help identifying service processes, resources, and information, to increase resilience of communities with respect to potential critical hazardous events affecting public safety. Taking into consideration the complexity of interactions between various participants in the public safety service ecosystem, we have proposed a unified semantic reasoning-based method that is embedded in a unified representation for collaborative development of complex services.

This way of thinking enforces the importance of data collection, its transformation into information, followed by a knowledge dissemination effort to a larger set of participants. To evaluate, include, and expose various participants' positions in public safety related activities we have employed a service-related knowledge foundation. As such, we highlight the distributed nature of data-centric processing tasks associated to specific activities based on the real-time communication between machine and human, machine and machine, which is made possible in Cyber Physical Systems.

To support this unified representation, we have introduced a conceptual representation detailing the implementation of the internal services acting upon these data, the Distributed Information Service Actor Role Network. Its main functionality is explained along with the case study in the evaluation of hazardous events related to Data Quality in water resource management.

For this case study, we have performed an analysis and we identified five quality criteria and nine types of Data Quality issues that refer to one or more of the quality criteria. This analysis serves as a foundation in designing and implementing the Time Series Quality Assurance (TSQA) solution ontology. We have proposed and implemented an architecture that binds together the TSQA ontology, SWRL-encoded rules, exposed via an OWL API, detection/correction methods, and a controller implemented in Java to exploit the proposed ontology and to infer new domain expert knowledge to be later exposed as intensive information services (IIS).

The TSQA ontology is designed to work on any type of data, and the concepts that were introduced in the ontology are derived directly from the Algorithm 1, introducing the unified method for improving Data Quality. As a specific use case, we have approached a specific situation related to water quality data, thus we re-used other ontologies, such as the Smart Home Weather ontology and Semantic Sensor Network ontology.

Several further development roadmaps may be conceived to enlarge the presentation of this current work. First comes concretizing ideas around the development and evolution of the information common goods in public safety, as a shared resources system, a real platform which can be composed based on several information services. Sustainably managing various types of resource systems as commons [87] and emerging commons in this direction of research is one of most promising applications of Elinor Ostrom's work on institutions [110].

Further developments may be also envisioned for the development of the internal (technical) services in the specific case study of information intensive service development to support communities' resilience facing hazardous events. Ontology design for semantic aware data cleaning is not yet a mature topic and still a field of ongoing research. Semantic technologies are the ideal choice in problems where context matters (context aware computing). In case of the natural resource management, the context plays a key role not only for alerting or forecasting, but also for data acquisition. Ontologies foster knowledge sharing from contributors and establish a common vocabulary such that once defined in a non-equivocal manner, the concepts have the same meaning for all actors in the system (human or machine). Therefore, semantic reasoning and collaborative development of knowledge using ontologies may be a good solution to understand public safety in relation to various occurring hazardous events.

Ontologies are intensively used to infer implicit knowledge from explicit knowledge by applying rules (logical inference, IF/THEN) that often appear in case of environmental applications. Knowledge reuse is a key aspect of ontologies. Once defined, an ontology can be imported, extended and reused. New concepts can be built up starting with the existing one, by dynamically combining existing knowledge, therefore offering explicit reasoning about the problem domain.

As a future direction for improvements, the taxonomy of TSQA ontology could be extended by providing more types of detection/correction methods (for example, only for interpolation we can subdivide into at least four subcategories such as linear interpolation, cosine interpolation, cubic interpolation, Hermite interpolation) and more types of data issues. Then, as another research direction, we may suggest improving the scalability of the architecture for the implementation of the internal (technical services) of the DISActoRN component, both horizontally and vertically, by exposing it in the Cloud. Considering that reasoning over an increased number of facts (millions of individuals corresponding to large time series) is computationally expensive, a Cloud-based solution would offer the elastic allocation of CPU/RAM. However, it is not a trivial topic of research to

establish how reasoners such as Pellet, Hermite, FaCT++ scale in a distributed environment (such as Cloud).

As well, for the future we intend to describe and introduce into practice, at a higher-level conceptualization, a new knowledge domain named Informational Common, dedicated to conceiving complex services (Fig. 1), following previous work on environment-oriented development of services as common goods proposed in [35], [37]. Service intelligence creates the foundation for the development of this knowledge domain, by which we understand a comprehensive vision enabling actionable Exploration of the perceived complex situation in public safety, while it becomes possible to discover new Information Services, exposing them as commons, and crystallizing them as a pool of shared service-type resources created in Cognitive Collaborative Environments.

The definition of such a domain of interest in conceiving complex services aiming to transform data into information is important for employing human-oriented development when answers are needed for people finding themselves in complex situations inside Society.

VIII. CONCLUSIONS

Extending the data collection effort and its transformation into information is a compulsory step to better understand the critical interactions that govern the co-evolution of the systems as human-centered entities in Society. Taking into account the complexity of the service ecosystems in public safety, in this paper we argue that a clear inter-institutional, inter-disciplinary, and even international context within the United Nations frameworks is mandatory to guarantee the robustness of the exploratory approach to transform data into information using service intelligence to advance the Public Safety as a Service vision. To fully recognize the digital potentialities supported by services, this service intelligence can emerge only through strong concentration and co-design processes with specialists of disasters, public administration, service related research, and digital systems.

To continue the work presented in this article, it is valuable to further explore the transformation of the rich informational situational context, created today by the ubiquitous manifestation of myriads of devices transformed into smart objects that empower the human beings with capabilities never imagined before and hardly envisioned in the near future, into its concrete supporting services through the creation of collaborative co-creative environments fostering service-centric innovations. This will acknowledge that information is the core element in the design, implementation, and management of services, while data, information, and knowledge resources are managed by Actors in various positions in Society through provisioning and appropriation kind of actions.

REFERENCES

[1] V. Grasso, A. Singh, and J. Pathak. (2012). *Early Warning Systems: State-of-Art Analysis and Future Directions*. [Online]. Available: http://www.unep.org

[2] X. Liu, A. Heller, and P. S. Nielsen, "CITIESData: A smart city data management framework," *Knowl. Inf. Syst.*, vol. 53, no. 3, pp. 699–722, Dec. 2017.

[3] M. M. Herterich and M. Mikusz, "Looking for a few good concepts and theories for digitized artifacts and digital innovation in a material world," in *Proc. Int. Conf. Inf. Syst.*, 2016, pp. 1–8. [Online]. Available: http://aisel.aisnet.org/icis2016/DigitalInnovation/Presentations/9/

[4] P. Pasupuleti and B. S. Purra, *Data Lake Development with Big Data*. Birmingham, U.K.: Packt Publishing Ltd, 2015.

[5] (2009). *UNISDR Terminology on Disaster Risk Reduction, United Nations Office for Disaster Risk Reduction*. [Online]. Available: https://www.unisdr.org/we/inform/publications/7817

[6] K. Soma and A. Vatn, "Representing the common goods—Stakeholders vs. Citizens," *Land Use Policy*, vol. 41, pp. 325–333, Nov. 2014.

[7] Z. A. Auzzir, R. P. Haigh, and D. Amaratunga, "Public-private partnerships (PPP) in disaster management in developing countries: A conceptual framework," *Procedia Econ. Finance*, vol. 18, pp. 807–814, Apr. 2014.

[8] Y. Martín, M. R. Mimbrero, and M. Zúñiga-Antón, "Community vulnerability to hazards: Introducing local expert knowledge into the equation," *Natural Hazards*, vol. 89, no. 1, pp. 367–386, Oct. 2017.

[9] H. I. Kobo, A. M. Abu-Mahfouz, and G. P. Hancke, "A survey on software-defined wireless sensor networks: Challenges and design requirements," *IEEE Access*, vol. 5, pp. 1872–1899, 2017.

[10] B. Fitzgerald, "Crowdsourcing software development: Silver bullet or lead balloon," in *Proc. 5th Int. Workshop Artif. Intell. Requirements Eng. (AIRE)*, Aug. 2018, pp. 29–30.

[11] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing Paradigm," *ACM Comput. Surv.*, vol. 48, no. 1, Jul. 2015, Art. no. 7.

[12] J. Wang, F. Wang, Y. Wang, L. Wang, Z. Qiu, D. Zhang, B. Guo, and Q. Lv, "HyTasker: Hybrid task allocation in mobile crowd sensing," *IEEE Trans. Mobile Comput.*, to be published.

[13] J. Reilly, S. Dashti, M. Ervasti, J. D. Bray, S. D. Glaser, and A. M. Bayen, "Mobile phones as seismologic sensors: Automating data extraction for the iShake system," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 2, pp. 242–251, Apr. 2013.

[14] M. Faulkner, M. Olson, R. Chandy, J. Krause, K. M. Chandy, and A. Krause, "The next big one: Detecting earthquakes and other rare events from community-based sensors," in *Proc. 10th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, Apr. 2011, pp. 13–24.

[15] L. Wang and Y. Chen, "An intelligent management of community water and electricity based on wireless sensor network," in *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery*, May 2012, pp. 2581–2585.

[16] F. Finazzi, "The earthquake network project: Toward a crowdsourced smartphone-based earthquake early warning system," *Bull. Seismological Soc. Amer.*, vol. 106, no. 3, pp. 1088–1099, May 2016.

[17] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "Participatory air pollution monitoring using smartphones," *Mobile Sens.*, vol. 1, pp. 1–5, Apr. 2012.

[18] C. Ionescu and M. Drăgoicea, "MACROSEIS: A tool for real-time collecting and querying macroseismic data in Romania," *Romanian J. Phys.*, vol. 55, nos. 7–8, pp. 852–861, Jan. 2010.

[19] J. W. Dewey, D. Wald, L. Dengler, and M. Hopper, "Macroseismic intensity in the Internet age," *Sel. Papers From Vol. Vychislitel'naya Seysmologiya*, vol. 7, pp. 60–65, Jan. 2005. doi: 10.1029/CS007p0060.

[20] S. Balandin and H. Waris, "Key properties in the development of smart spaces," in *Proc. Int. Conf. Universal Access Hum.-Comput. Interact.*, 2009, pp. 3–12.

[21] S. Macbeth and J. V. Pitt, "Self-organising management of user-generated data and knowledge," *Knowl. Eng. Rev.*, vol. 30, no. 3, pp. 237–264, May 2015.

[22] D. G. Korzun, "Designing Smart Space based information systems: The case study of services for IoT-enabled collaborative work and cultural heritage environments," DB&IS, Berlin, Germany, Tech. Rep.183, 2016.

[23] UNISDR annual report. (2019). *United Nations Office for Disaster Risk Reduction*. [Online]. Available: https://www.unisdr.org/we/inform/publications/52253

[24] (Mar. 2015). *Sendai Framework for Disaster Risk Reduction 2015-2030*. [Online]. Available: https://www.unisdr.org/we/coordinate/sendai-framework

[25] P. Marana, L. Labaka, and J. M. Sarriegi, "A framework for public-private-people partnerships in the city resilience-building process," *Saf. Sci.*, vol. 110, pp. 39–50, Oct. 2018.

[26] (2017). *Sustainable Development Knowledge Platform. Goal 11: Make Cities Inclusive, Safe, Resilient and Sustainable*. [Online]. Available: https://sustainabledevelopment.un.org/sdg11
[27] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and communications technologies for sustainable development goals: State-of-the-art, needs and perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2389–2406, 3rd Quart., 2018.
[28] O. Kostoska and L. Kocarev, "A novel ICT framework for sustainable development goals," *Sustainability*, vol. 11, no. 7, p. 1961, 2019.
[29] M. Drăgoicea, N. G. Badr, J. F. E. Cunha, and V. E. Oltean, "From data to service intelligence: Exploring public safety as a service," in *Exploring Service Science*, G. Satzger, L. Patrício, M. Zaki, N. Kühl, and P. Hottum, Eds. New York, NY, USA: Springer, 2018, pp. 344–357.
[30] S. Barile and F. Polese, "Smart service systems and viable service systems: Applying systems theory to service science," *Service Sci.*, vol. 2, nos. 1–2, pp. 21–40, Jun. 2010.
[31] T. Ono, K. Lida, and S. Yamazaki, "Achieving sustainable development goals (SDGs) through ICT services," *FUJITSU Sci. Tech. J.*, vol. 53, no. 6, pp. 17–22, Oct. 2017.
[32] (2018). *Accelerating SDGs Through ICT*. [Online]. Available: https://www.huawei.com/
[33] (2018). *ICT for Transformation and Resilience, Report of the Side Event of the Asia–Pacific Forum for Sustainable Development*. [Online]. Available: https://www.unescap.org/resources/report-side-event-asia-pacific-forum-sustainable-development-ict-transformation-and
[34] S. L. Vargo and R. F. Lusch, "Evolving to a new dominant logic for marketing," *J. Marketing*, vol. 68, no. 1, pp. 1–17, 2004.
[35] M. Léonard and A. Yurchyshyna, "Towards contributive development of services," in *Clean Mobility and Intelligent Transport Systems*, M. Fiorini and J.-C. Lin, Eds. London, U.K.: IET, 2015, pp. 1–21.
[36] H. A. Simon, *The Sciences of the Artificial*. Cambridge, MA, USA: MIT Press, 1996.
[37] A. Yurchyshyna, "Towards contributory development by the means of services as common goods," in *Exploring Services Science*, H. Nóvoa and M. Drăgoicea, Eds. Cham, Switzerland: Springer, 2015, pp. 12–24.
[38] R. F. Lusch, S. L. Vargo, and G. Wessels, "Toward a conceptual foundation for service science: Contributions from service-dominant logic," *IBM Syst. J.*, vol. 47, no. 1, pp. 5–14, Apr. 2008.
[39] P. P. Maglio, S. L. Vargo, N. Caswell, and J. Spohrer, "The service system is the basic abstraction of service science," *Inf. Syst. E-Business Manage.*, vol. 7, no. 4, pp. 395–406, Sep. 2009.
[40] J. Tukey, *Exploratory Data Analysis*. Cambridge, MA, USA: MIT Press, 1977.
[41] (2019). *Visual Analysis for CPS Data*. [Online]. Available: https://ieeeaccess.ieee.org/special-sections/visual-analysis-for-cps-data/
[42] C.-H. Lim and K.-J. Kim, "Information service blueprint: A service blueprinting framework for information-intensive services," *Service Sci.*, vol. 6, no. 4, pp. 296–312, Dec. 2014.
[43] S. N. Ciolofan, G. Militaru, A. Draghia, R. Drobot, and M. Drăgoicea, "Optimization of water reservoir operation to minimize the economic losses caused by pollution," *IEEE Access*, vol. 6, pp. 67562–67580, 2018.
[44] C. G. Chiru, M. I. Mocanu, M. Drăgoicea, and A. D. Ioniţă, "Digital services development using statistics tools to emphasize pollution phenomena," in *Exploring Services Science*, S. Za, M. Drăgoicea, and M. Cavallari, Eds. New York, NY, USA: Springer, 2017, pp. 370–382.
[45] (2005). *Hyogo Framework for Action (HFA) 2005–2015, United Nations Office for Disaster Risk Reduction*. [Online]. Available: https://www.unisdr.org/we/coordinate/hfa
[46] R. C. Brears, *Urban Water Security*. Hoboken, NJ, USA: Wiley, 2017.
[47] W. Li, P. Jagtap, L. Zavala, A. Joshi, and T. Finin, "CARE-CPS: Context-aware trust evaluation for wireless networks in cyber-physical system using policies," in *Proc. IEEE Int. Symp. Policies Distrib. Syst. Netw.*, Jun. 2011, pp. 171–172.
[48] D. P. Möller, "Guide to computing fundamentals in cyber-physical systems," in *Computer Communications and Networks*. Heidelberg, Germany: Springer, 2016.
[49] R. Atat, L. Liu, J. Wu, G. Li, C. Ye, and Y. Yang, "Big data meet cyber-physical systems: A panoramic survey," *IEEE Access*, vol. 6, pp. 73603–73636, 2018.
[50] R. Atat, L. Liu, H. Chen, J. Wu, H. Li, and Y. Yi, "Enabling cyber-physical communication in 5G cellular networks: Challenges, spatial spectrum sensing, and cyber-security," *IET Cyber-Phys. Syst., Theory Appl.*, vol. 2, no. 1, pp. 49–54, 2017.
[51] J. Wu, I. Bisio, C. Gniady, E. Hossain, M. Valla, and H. Li, "Context-aware networking and communications: Part 1 [Guest Editorial]," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 14–15, Jun. 2014.
[52] J. Wu, S. Guo, J. Li, and D. Zeng, "Big data meet green challenges: Greening big data," *IEEE Syst. J.*, vol. 10, no. 3, pp. 873–887, Sep. 2016.
[53] M. A. Beyer and D. Laney, *The Importance of 'Big Data': A Definition*. Stamford, CT, USA: Gartner, 2012, pp. 2014–2018.
[54] A. I. Torre-Bastida, J. Del Ser, I. Laña, M. Ilardia, M. N. Bilbao, and S. Campos-Cordobés, "Big data for transportation and mobility: Recent advances, trends and challenges," *IET Intell. Transp. Syst.*, vol. 12, no. 8, pp. 742–755, Oct. 2018.
[55] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proc. IEEE*, vol. 104, no. 11, pp. 2207–2219, Nov. 2016.
[56] L. D. Xu and L. Duan, "Big data for cyber physical systems in industry 4.0: A survey," *Enterprise Inf. Syst.*, vol. 13, no. 2, pp. 148–169, 2019.
[57] U. J. Tanik and S. Arkun-Kocadere, "Cyber-physical systems and STEM development: NASA digital astronaut project," in *Applied Cyber-Physical Systems*. New York, NY, USA: Springer, 2014, pp. 33–49.
[58] Z. Liu, T. Tsuda, H. Watanabe, S. Ryuo, and N. Iwasawa, "Data driven cyber-physical system for landslide detection," *Mobile Netw. Appl.*, vol. 24, no. 3, pp. 991–1002, 2018.
[59] T. Papadopoulos, A. Gunasekaran, R. Dubey, N. Altay, S. J. Childe, and S. Fosso-Wamba, "The role of big data in explaining disaster resilience in supply chains for sustainability," *J. Cleaner Prod.*, vol. 142, pp. 1108–1118, Jan. 2017.
[60] L. Zhang, "Applying system of systems engineering approach to build complex cyber physical systems," in *Progress in Systems Engineering*. New York, NY, USA: Springer, 2015, pp. 621–628.
[61] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber–physical systems," *Proc. IEEE*, vol. 100, no. 1, pp. 13–28, Jan. 2012.
[62] S. Friedenthal, A. Moore, and R. Steiner, *A practical guide to SysML: The Systems Modeling Language*. San Mateo, CA, USA: Morgan Kaufmann, 2015.
[63] S. Mayer, J. Hodges, D. Yu, M. Kritzler, and F. Michahelles, "An open semantic framework for the industrial Internet of Things," *IEEE Intell. Syst.*, vol. 32, no. 1, pp. 96–101, Jan. 2017.
[64] G. Dodig-Crnkovic and A. Cicchetti, "Computational aspects of model-based reasoning," in *Springer Handbook Model-Based Science*, L. Magnani and T. Bertolotti, Eds. New York, NY, USA: Springer, 2017, pp. 695–718.
[65] F. Manola, E. Miller, and B. McBride, "RDF primer," *W3C Recommendation*, vol. 10, no. 6, pp. 100–107, 2004.
[66] P. Hitzler and M. Krötzsch, B. Parsia, P. F. Patel-Schneider, and S. Rudolph, "OWL 2 Web ontology language primer," *W3C recommendation*, vol. 27, no. 1, p. 123, 2009.
[67] S. Bechhofer, "OWL: Web ontology language," in *Encyclopedia of Database Systems*, M. T. Özsu and L. Liu, Eds. New York, NY, USA: Springer, 2009, pp. 2008–2009.
[68] J. C. Augusto, V. Callaghan, D. Cook, A. Kameas, and I. Satoh, "Intelligent environments: A manifesto," *Human-Centric Comput. Inf. Sci.*, vol. 3, no. 1, p. 12, 2013.
[69] C. El Kaed, I. Khan, A. Van Den Berg, H. Hossayni, and C. Saint-Marcel, "SRE: Semantic rules engine for the industrial Internet-of-Things gateways," *IEEE Trans. Ind. Inform.*, vol. 14, no. 2, pp. 715–724, Feb. 2018.
[70] L. Roffia *et al.*, "A semantic publish-subscribe architecture for the Internet of Things," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1274–1296, Dec. 2016.
[71] D. G. Korzun, A. M. Kashevnik, S. I. Balandin, and A. V. Smirnov, "The smart-M3 platform: Experience of smart space application development for Internet of Things," in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. New York, NY, USA: Springer, 2015, pp. 56–67.
[72] D. V. Chapman, *Water Quality Assessments: A Guide to the Use of Biota, Sediments and Water in Environmental Monitoring*. 2nd ed. Boca Raton, FL, USA: CRC Press, 1996.
[73] R. Ballance and J. Bartram, *Water Quality Monitoring: A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes*. 2nd ed. Boca Raton, FL, USA: CRC Press, 2002.
[74] (2018). *Data Quality Framework, GS1*. [Online]. Available: https://www.gs1.org/services/data-quality/data-quality-framework

[75] C. O'Neil and R. Schutt, *Doing Data Science. Straight Talk From the Frontline*. Newton, MA, USA: O'Reilly Media, 2013.
[76] A. Chapman. (2005). *Principles and Methods of Data Cleaning: Primary Species and Species-Occurrence Data*. [Online]. Available: http://www.gbif.org/document/80528a
[77] G. Klyne and J. J. Carroll, "Resource description framework (RDF): Concepts and abstract syntax," W3C, RDF Working Group, Tech. Rep., 2004. [Online]. Available: http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/
[78] S. Brüggemann and F. Grüning, "Using domain knowledge provided by ontologies for improving data quality management," in *Networked Knowledge—Networked Media*. Princeton, NJ, USA: Citeseer, 2008, pp. 251–258.
[79] Z. Kedad and E. Métais, "Ontology-based data cleaning," in *Proc. Int. Conf. Appl. Natural Lang. Inf. Syst.*, 2002, pp. 137–149.
[80] S.-T. Liaw, A. Rahimi, P. Ray, J. Taggart, S. Dennis, S. de Lusignan, B. Jalaludin, A. Yeo, and A. Talaei-Khoei, "Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature," *Int. J. Med. Informat.*, vol. 82, no. 1, pp. 10–24, Apr. 2013.
[81] A. Degbelo, "Short paper: An ontology design pattern for spatial data quality characterization in the semantic sensor Web," in *Proc. 5th Int. Conf. Semantic Sensor Netw.-Vol.*, 2012, pp. 103–108.
[82] S. Geisler, S. Weber, and C. Quix, "Ontology-based data quality framework for data stream applications," in *Proc. ICIQ*, Jul. 2011, pp. 145–159.
[83] H. B. Glasgow, J. M. Burkholder, R. E. Reed, A. J. Lewitus, and J. E. Kleinman, "Real-time remote monitoring of water quality: A review of current applications, and advancements in sensor, telemetry, and computing technologies," *J. Experim. Marine Biol. Ecol.*, vol. 300, nos. 1–2, pp. 409–448, Mar. 2004.
[84] E. D'Hondt, M. Stevens, and A. Jacobs, "Participatory noise mapping works! An evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring," *Pervas. Mobile Comput.*, vol. 9, no. 5, pp. 681–694, 2013.
[85] O. Bogoiavlenskaia, A. Vdovenko, D. G. Korzun, and A. Kashevnik, "Individual client strategies for active control of information-driven service construction in IoT-enabled smart spaces," *Int. J. Distrib. Syst. Technol.*, vol. 10, no. 2, pp. 20–36, Apr. 2019.
[86] D. G. Korzun, I. Nikolaevskiy, and A. Gurtov, "Service intelligence and communication security for ambient assisted living," *Int. J. Embedded Real-Time Commun. Syst.*, vol. 6, no. 1, pp. 76–100, 2015.
[87] B. M. Frischmann, M. J. Madison, and K. J. Strandburg, *Governing Knowledge Commons*. New York, NY, USA: Oxford Univ. Press, 2014.
[88] H. Demirkan, R. J. Kauffman, J. A. Vayghan, H.-G. Fill, D. Karagiannis, and P. P. Maglio, "Service-oriented technology and management: Perspectives on research and practice for the coming decade," *Electron. Commerce Res. Appl.*, vol. 7, no. 4, pp. 356–376, Aug. 2008.
[89] A. Caragliu, C. D. Bo, and P. Nijkamp, "Smart cities in europe," *J. Urban Technol.*, vol. 18, no. 2, pp. 65–82, 2011. doi: 10.1080/10630732.2011.601117.
[90] U. Aguilera, O. Pe na, O. Belmonte, and D. López-de Ipiña, "Citizen-centric data services for smarter cities," *Future Gener. Comput. Syst.*, vol. 76, pp. 234–247, Nov. 2017.
[91] T. Janowski, "Digital government evolution: From transformation to contextualization," *Government Inf. Quart.*, vol. 32, no. 3, pp. 221–236, 2015.
[92] W. Opprecht, A. Yurchyshyna, A. Khadraoui, and M. Lèonard, "Governance of initiatives for e-government services innovation," Tech. Rep., 2010.
[93] S. L. Vargo, M. A. Akaka, and C. M. Vaughan, "Conceptualizing value: A service-ecosystem View," *J. Creating Value*, vol. 3, no. 2, pp. 117–124, Oct. 2017.
[94] S. Fujita, C. Vaughan, and S. Vargo, "Service ecosystem emergence from primitive actors in service dominant logic: An exploratory simulation study," in *Proc. 51st Hawaii Int. Conf. Syst. Sci.*, 2018, pp. 1–3.
[95] A. D. Ioniţă, C.-T. Eftimie, G. Lewis, and M. Liţoiu, "Integration of hazard management services," in *Proc. Int. Conf. Exploring Service Sci.*, T. Borangiu, M. Drăgoicea, and H. Nóvoa, Eds. New York, NY, USA: Springer, 2016, pp. 355–364.
[96] (2005). *Semantic Sensor Network Ontology, W3C Semantic Sensor Network Incubator Group*. [Online]. Available: https://www.w3.org/2005/Incubator/ssn/ssnx/ssn
[97] (2017). *Time Ontology in OWL, W3C*. [Online]. Available: https://www.w3.org/TR/owl-time/
[98] (2015). *Quantity-Unit-Dimension-Type Ontology*. [Online]. Available: http://www.qudt.org/qudt/owl/1.0.0/
[99] (2014). *Smart Home Ontology for Weather Phenomena and Exterior Conditions*. [Online]. Available: https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl
[100] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean, "Swrl: A semantic Web rule language combining owl and ruleml," *W3C Member Submission*, vol. 21, no. 79, pp. 1–31, 2004.
[101] N. F. Noy and D. L. McGuinness. (2010). *Ontology Development 101: A Guide to Creating Your First Ontology*. [Online]. Available: http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html
[102] J. Hamilton, *Time Series Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1994.
[103] A. C. Parnell, "Climate time series analysis: Classical statistical and bootstrap methods," *J. Time Ser. Anal.*, vol. 34, no. 2, p. 281, Apr. 2013.
[104] S. N. Rodionov, "A brief overview of the regime shift detection methods," in *Large-Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges for Management Toward Sustainability*, V. Velikova and N. Chipev, Eds. Varna, Bulgaria: UNESCO-ROSTE/BAS Workshop on Regime Shifts, 2005, pp. 17–24. [Online]. Available: http://www.ecolab.bas.bg/main/Events/unesco-ws/
[105] B. M. Khalil, A. G. Awadallah, H. Karaman, and A. El-Sayed, "Application of artificial neural networks for the prediction of water quality variables in the nile delta," *J. Water Resource Protection*, vol. 4, no. 6, p. 388, 2012.
[106] J. S. Horsburgh, A. S. Jones, D. K. Stevens, D. G. Tarboton, and N. O. Mesner, "A sensor network for high frequency estimation of water quality constituent fluxes using surrogates," *Environ. Model. Softw.*, vol. 25, no. 9, pp. 1031–1044, 2010.
[107] S. Skiena, *The Data Science Design Manual*. Cham, Switzerland: Springer, 2017.
[108] (2000). *Tunkhannock Creek Weekly Data*. [Online]. Available: http://www.waterontheweb.org/data/tunkhannock/realtime/weekly.html
[109] (2018). *Pellet, An Open Source Java based reasoner for OWL*. [Online]. Available: https://github.com/stardog-union/pellet
[110] C. Hess and E. Ostrom, *Understanding Knowledge as a Commons. From Theory to Practice*. Cambridge, MA, USA: MIT Press, 2011.

MONICA DRĂGOICEA (M'18) received the B.S. degree in automatic control from the Faculty of Automatic Control and Computers, University Politehnica of Bucharest, in 1993, the M.S. degree in engineering management from Technische Universität Vienna, Austria, and Oakland University, Rochester, MI, USA, in 1999, and the Ph.D. degree in automatic control from the Faculty of Automatic Control and Computers, University Politehnica of Bucharest, in 2000.

She is currently a Full Professor with the Faculty of Automatic Control and Computers, University Politehnica of Bucharest. She has been involved with theoretical and experimental work in software and systems engineering for the past 20 years. Her research interests include modeling and simulation-based systems engineering, real-time systems, service systems engineering, digital design of services, and computational intelligence. She is also a member of the IEEE Systems, Man, and Cybernetics Society, the International Society of Service Innovation Professionals (ISSIP), and the Robotics Society of Romania (SRR).

MICHEL LÉONARD is currently an Honorary Professor with the Institute of Information Service Science (ISS), University of Geneva. He founded, during his career, since 1977, several educational programs in information systems and service science at all the levels: bachelor's/master's/Ph.D. and continuous education. His current research interests include service science for the society progression in the dimensions of management, IT, engineering, professions, and situational methods. Since 2010, he co-founded a series of annual conferences in service science, called the International Conference on Exploring Service Science (IESS).

SORIN N. CIOLOFAN received the B.S., M.S., and Ph.D. degrees in computer science from the Faculty of Automatic Control and Computers, University Politehnica of Bucharest, in 2003, 2004, and 2017, respectively.

From 2003 to 2012, he held different positions as a Software Engineer, an Research and Development Engineer, a Solution Architect, and a Senior Software Engineer with IBM Ireland, Dublin Software Laboratory, the Foundation of Research and Technology Hellas (F.O.R.T.H), the Institute of Computer Science, Heraklion, Greece, Forte Business Services, Bucharest, Romania, and Qualisoft, Bucharest. He is currently a Lecturer with the Department of Computer Science, University Politehnica of Bucharest. His main research interests include cloud computing, big data, information systems, cyber–physical systems, and modeling and simulation-based systems engineering.

GHEORGHE MILITARU received the B.S. degree in technological physics from the Faculty of Physics, University of Bucharest, in 1984, the B.S. degree in finance from the Bucharest University of Economics Studies, in 2005, and the Ph.D. degree in management and industrial engineering from the University Politehnica of Bucharest, Romania, in 1996.

He was a Certified Business Counselor in Small Business from Washington State University, USA, in 1997. He is currently a Full Professor with the Faculty of Entrepreneurship, Business Engineering and Management, University Politehnica of Bucharest. His research interests include service management, financial management, production management and engineering, information systems for management, digital marketing, and technology entrepreneurship. He is also an active member in Romanian and international professional associations in management and engineering.