A conceptual model for analyzing information quality in System-of-Systems

A Conceptual Model for Analyzing Information Quality in System-of-Systems

Mohamad Gharib, Paolo Lollini, Andrea Bondavalli
University of Florence - DiMaI
Viale Morgagni 65, Florence, Italy
{mohamad.gharib,paolo.lollini,andrea.bondavalli}@unifi.it

***Abstract*—A System-of-Systems (SoS) is an integration of a finite number of Constituent Systems (CSs), which are networked together for achieving a certain higher goal. Therefore, integration is the key viability of any SoS. Although the integration of CSs can be achieved by the exchange of information, no existing work has considered the quality of such information. Without considering Information Quality (IQ), a CS may depend on inaccurate, incomplete, inconsistent, invalid, and/or untrustworthy information, which might lead to its failure, and in turn to catastrophic incidents in the case of critical SoS. The main objective of the paper is proposing a novel conceptual model that provides the required concepts for analyzing for SoS. We illustrate the utility of the model with an example concerning the Intelligent Transportation System (ITS) domain.**

***Index Terms*—System-of-Systems; SoS; Information; Information Quality; Conceptual Modeling**

I. Introduction

A SoS can be defined as an integration of a finite number of independent and operable CSs, which are networked together to achieve a certain higher goal [1]. SoS has a broad application area that covers almost all areas of life (e.g., transportation, energy, healthcare, disaster response, etc.). One main goal of SoS Engineering is ensuring that SoS can function as a single integrated system to support a common mission, that is why integration is the key viability of any SoS [1]. However, the integration of CSs is not an easy task due to the special characteristics that distinguish SoS from other types of systems, such as the autonomy of its components (CSs), emergent behavior, dynamicity, etc. [2], [3].

Several researchers have shed a light on the importance of information in the integration of CSs (e.g., [2]–[6]). One of the most notable work is [5], in which Kopetz argued that the integration among CSs is achieved by the exchange of information among the CSs, and he introduced the concept of Information atom (Itom) that is a tuple consisting of data and an explanation of such data (meta-data). However, the main focus of this work was on how information can be exchanged and understood by CSs, not on the quality of such information. For example, the Itom does not make any assumptions about the truthfulness of data contained within an Itom. Moreover, it is not clear how the temporal property of Itoms can be used to analyze the validity of information for performing a particular task. Without considering IQ, a CS may depend on information that is inaccurate, incomplete, inconsistent, invalid, untrustworthy, etc. Depending on such information may result in undesirable outcome or it may even lead a CS failure and, in turn, to catastrophic incidents since SoS can be extremely vulnerable to sudden collapse due to functionality losses in one of its CSs [4].

The Flash crash (a main stock market crash [7]) is a good example of a SoS failure due to IQ issues. For instance, some CSs (traders) intentionally provide falsified information to manipulate the market and gain extra profit. While other CSs (trading venues) failed to coordinate their activities to halt trading during the crash because they depend on incomplete and inconsistent information. This enables some traders to continue their trades in the venues that did not halt, which escalate the crash due to the unpredictable emergent behavior of such systems [8]. To this end, we advocate that Information Quality (IQ) is a key factor for successful integration of any SoS. More specifically, the integration of CSs not only need to consider how information can be exchanged and understood by CSs, but also the quality of such information.

On the other hand, IQ is a hierarchical multi-dimensional concept that can be characterized by different dimensions, such as accuracy, completeness, timeliness, consistency, etc. [9], [10]. Although there are several models for analyzing IQ (e.g., [7], [10]–[12]), none of them has been designed with the special needs of SoS in mind. For instance, no existing IQ model is able to deal with information that is exchanged through indirect channels (e.g., stigmergic information [13]). Therefore, they are not able to properly analyze the quality of such information. The importance of such information for the performance of SoS has been clearly highlighted in [6], [13].

Moreover, analyzing information completeness depending on existing IQ models may not fit the needs of SoS, since most CSs may not have complete knowledge about their operational environment, i.e., they may not be able to identify whether the information is complete for performing a specific activity. In addition, due to the decentralized nature of SoS, ensuring information consistency for the overall SoS might not be possible or at least it is not feasible in the case of SoS. In this context, an IQ model that tackles the previously mentioned problems would constitute a great step forward in designing SoS. In this paper, we propose a new conceptual model for analyzing IQ for SoS in terms of its four core IQ dimensions (accuracy, completeness, timeliness, and consistency).

The rest of this paper is organized as follows; Section II describes a motivating example, we use to illustrate our conceptual model, and Section III presents a background concerning information and IQ. We present and discuss the conceptual model for analyzing IQ for SoS in section IV, and in Section V, we illustrate its applicability to a realistic scenario from the ITS domain. Finally, we conclude and discuss future work in Section VI.

II. Motivating Example: Vehicular Communication NETworks

Our motivating example concerns Vehicular Communication NETworks (VANETs) that is an important component of current Intelligent Transportation Systems (ITS). Based on [14], VANETs components can be classified under:

- Drivers/vehicles aim to reach their destinations safely using the shortest and least congested route.
- Road Side Units (RSUs) (infrastructure) collect and disseminate information that assists drivers to reach their destinations. In particular, RSUs provide information concerning road conditions, road congestion, etc.
- Point of Interest (PoI) is an entity that offers different services that might be of interest to passing vehicles (e.g., hotels, gas stations). A PoI depends on RSU to advertise services it offers by means of PoI notification.

While communication links in VANETs can be broadly classified under:

- Vehicle-to-Vehicle (V2V): allows the direct vehicular communication without relying on RSUs and can be mainly employed for safety, security, and dissemination applications.
- Vehicle-to-Infrastructure (V2I): allows a vehicle to communicate with RSUs mainly for information.
- Infrastructure -to-Vehicle (I2V): allows RSUs to communicate with vehicles.

Figure 1 shows a partial diagram of VANETs in terms of its main components along with their communication links. VANETs applications have been proven to be useful in increasing the safety of drivers (e.g., pre-crash sensing/warning, Cooperative Forward Collision Warning (CFCW), hazard location notification, etc.), and increasing the traffic efficiency (e.g., enhanced route guidance and navigation, green light optimal speed advisory, point of interest notification, etc.). However, VANETS component systems depend on information to perform and coordinate their activities among one another, and the efficient performance of such systems heavily depends on the quality of such information.

III. Information and Information Quality

In order to get a better understanding of IQ, we need to clarify how information is related to the knowledge concept, how it can be created/produced by information sources, and how it can be exchanged (information provenance).

Information and Knowledge, the definitions and interrelations of information and knowledge have been discussed intensively within different communities. For instance, Ackoff [15]

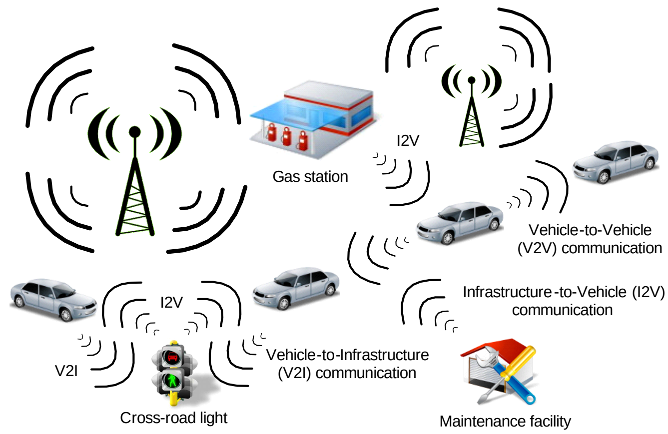

Fig. 1. A partial diagram of VANETs

stated that information is data[1] that has been given meaning, and knowledge is an appropriate collection of information. In [16], information is a commodity, product, or a thing, and knowledge is seen as a belief. In this paper, we adopt the definitions provided in [17], where information is data that has been given a meaning, and knowledge is learned information, which is incorporated in an actors' reasoning resources. On the other hand, information can be transformed into knowledge by learning [15], while knowledge can be interpreted into information [18].

Information sources can be broadly classified into three main types [19]: 1- *created internally,* an actor (information source) is able to produce information based on its own knowledge, or elaborate it from information it has/possess (e.g., a driver sets the destination (created information) of her trip in the vehicle driving assistant system); 2- *obtained from objects,* an actor is able to acquire information (called *state variable*) that describes an object or one of its properties (e.g., a driver is able to acquire the situation of a road); and 3- *acquired by communication,* actors can depend on one another for information to be provided (e.g., a driver can depend on a GPS service for positioning information).

Information provenance can be defined as any information that helps in determining the history of an information product, starting from its original sources and the process by which it has been delivered to its destination [20]. Provenance is particularly important for analyzing IQ, since the details of provenance enable for estimating the quality of information.

A meta-model of how information can be created/produced and exchanged among actors is depicted in Figure 2.

Information Quality can be analyzed based on its different dimensions, and several models for IQ analysis have been proposed in the literature (e.g., [7], [10]–[12]), yet none of them is able to deal with the special needs of SoS. Table I lists the main IQ dimensions that have been considered in the literature. Although no general consensus has been reached on which of these dimensions should be considered, four of them

[1] We do not consider data since the main focus of this paper is IQ

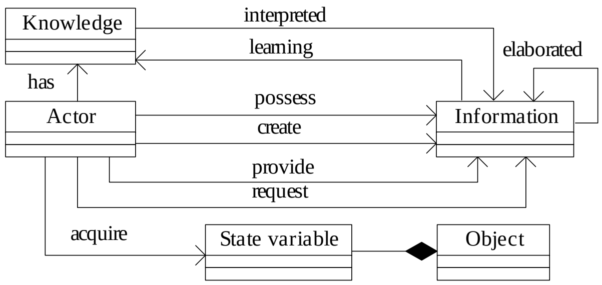

Fig. 2. Information creation and exchange

(e.g., accuracy, completeness, timeliness, and consistency) appear in almost all existing IQ models. Therefore, we propose a model for analyzing IQ in terms of these core four dimensions, namely: accuracy, completeness, timeliness, and consistency, which can be defined as follows:

- **Accuracy** can be defined as the extent to which information is true or error free with respect to some known or measured value [10].
- **Completeness** can be defined as the extent to which information is complete for performing a specific task [10].
- **Timeliness** can be defined as the extent to which information is valid in term of time (sufficiently up-to-date) for performing a specific task [9].
- **Consistency** can be defined as the extent to which all multiple records of the same information are the same across time and space [10].

IV. A CONCEPTUAL MODEL FOR INFORMATION QUALITY IN SYSTEM OF SYSTEMS

In this section, we present a novel conceptual model for analyzing IQ in terms of four IQ dimensions taking into consideration the special characteristics of SoS. As previously mentioned, a main limitation in existing IQ models their inability to deal with stigmergic information that is exchanged through indirect channels among CSs [13]. Therefore, the model shall provide the required concepts/relations for capturing information that is exchanged through direct and indirect channels. Thus, we rely on the ontology proposed by Bunge

TABLE I
IQ DIMENSIONS IN LITERATURE

	Accuracy	Completeness	Timeliness	Consistency	Believability	Accessibility	Trustworthiness
Wang and Strong [21]	X	X	X		X	X	
Bovee et al. [10]	X	X		X			
Pipino et al. [9]	X	X	X	X	X	X	
Liu and Chi [11]	X	X		X		X	X
Gharib and Giorgini [7]	X	X	X	X			
Gharib and Giorgini [12]	X	X	X	X	X	X	X

[22] for modeling the real-world to deal with stigmergic information and analyze their quality. In what follows, first we present the general concepts of the model, followed by concepts and attributes specialized for analyzing IQ in terms of its four core dimensions.

In our model, the world is made up of *things*, where a *thing* is a physical entity that has an identifiable existence in the world. A *thing* (e.g., a car) has a *state* (e.g., moving), which has a set of values called *state variables* that represent properties of the *state* (e.g., speed), and the value of each property is contained in a *value* attribute. The real world can be represented in Information System (IS), where *things* can be represented by *information objects*. Each *information object* has a defined set of information items (we call *produced information*). A couple of a *state variable* and a *produced information* item that represent it is called a corresponding couple, and the value of each *produced information* should reflect the value of its corresponding *state variables* at any moment in time.

On the other hand, a System of Systems (SoS) *integrates Constituent Systems (CSs)*, where a *CS is a thing* that is an autonomous sub-system of a SoS, consisting of computer systems and possibly of a controlled objects and/or humans. We differentiate between two types of *CSs*, *intentional CSs* and *unintentional CSs*, where the first have its own intentional objectives that may influence its behavior within a *SoS* (e.g., driver), while the last does not have any intentional objectives (e.g., sensor).

A *CS* is able to *produce information* (*produced information*) by *acquiring* its value from its corresponding *state variable*. For example, a RSU is able to acquire information concerning nearby passing vehicles (e.g., location, speed, direction, etc.). While *Intentional CSs* are able to *create* information (*created information*) based on their knowledge or information they already have. For example, a driver may set the destination of his trip (created information) in the "driving assistance application". The main difference between *produced* and *created information* is that, the former has corresponding *state variable* in the real world, while the last, usually, does not since it is created internally by a CS. *Produced* and *created information* can be generalized into *information* concept that has a *value* attribute. Moreover, a *CS* may *send/receive messages* that *contain information* by *relying* on *message interface* that *transmits messages* depending on a *channel*.

A *CS* may *perform activities* while perusing its objectives, where each *activity* has a *type* attribute that is used to categorize *activities* into coherent related groups. We differentiate between two types of *activities*: (1) *Intentional communicative activities*, which are performed with the intention of changing a *state* of a *thing* to convey a *message*. For example, a driver may turn on a car turn lights to notify other drivers that he/she is planning to turn right/left. In such case, there is a causality between turning flasher on and communicating informing (sending a message) to other drivers that a driver is planning to take a turn. (2) *Unintentional communicative activities*, which are not performed to communicate any kind of information. For example, a driver may increase the speed of its car, yet he may not have increased the speed to convey any message to other entity, i.e., there is no causality between increasing the car speed and sending any message. Note that increasing the car speed is an activity, where information related to it can be captured, interpreted and used by other CSs (e.g., drivers).

In what follows, we present concepts and attributes for analyzing IQ in terms of its four dimensions:

1- Information accuracy, considering the two different types of information, we differentiate between two cases:

Accuracy of produced information, we say that a *produced information* item is accurate at a point of time if its value is equal to the value of its corresponding *state variable*. Inaccuracy is a situation where the value of the *produced information* is not equal to the value of its corresponding *state variable*. Synchronizing the value of *produced information* with the value of its corresponding *state variable* can be used to guarantee the accuracy of the *produced information*, i.e., the value of the *state variable* can serve as an accurate reference for the value of the *produced value*. For example, a GPS system is used to sense the physical location of the car and represent it to the driver, yet in some cases, it might not reflect the accurate position of the car at a point in time.

Accuracy of created information[2]*,* a *created information* usually does not have a corresponding *state variable* in the real world, i.e., its value cannot be synchronized with an accurate reference value. Therefore, its accuracy can be analyzed based on the trustworthiness of its provenance (information source and providence) [20], we analyze the trustworthiness of information source considering the following three aspects:

- The class of the information source (CS) in the SoS, which can be broadly classified under *trusted CSs*, *distrusted CSs*, and *unclassified CSs*.
- The class of the activity that produces such information, which can be classified under *legitimate*, *suspicious*, and *malicious*.
- The class of the content of such information that can be classified under *safe*, *potentially harmful*, and *harmful*.

On the other hand, the trustworthiness of the providence that applies to both types of information can be determine by analyzing whether the transmitted information has been intentionally (e.g., non-authorized, non-trusted and/or malicious entity) or accidentally (e.g., technical failure in the message interface or the channel) modified during its transfer.

2- Information completeness can be analyzed depending on two sub-dimensions [7], [12], namely:

Value completeness, where information is preserved against corruption and/or loss during its transfer. Therefore, we extend the *channel* concept with a new attribute *transmission type* that can be either integrity preserving or normal transmission, where the first preserve the integrity of transmitted information, while the last does not. This enables us to analyze the value completeness of a transmitted information based on the type of the channel used to transmit information, i.e., information is guaranteed to be value complete if it has been transmitted from its source to its final destination through an integrity-preserving channel, otherwise, its value completeness is not guaranteed.

Purpose of use completeness, where information is complete for performing a specific activity, if all required information for performing the activity is available, i.e., an information item should have all its sub-parts for performing a specific activity. We adopt the following three concepts to analyze information purpose of use completeness: *Part_of* concept to model the relation between an information item and its sub-items, e.g., a full integrated map may have several parts related to it (e.g., the location of hotels, police stations, gas stations, etc.), and such relations can be described by the *part_of* concept. *Purpose of use* concept that specifies information purpose of usage, e.g., a driver assistant may need to use an integrated map for the *purpose* of choosing the best route to the driver destination. *Relevant_to* concept to model the relation between the purpose of information usage and any other information that is related to the same purpose of usage. Considering the map for the *purpose* of choosing the best route to the driver destination, the map may need information about any maintenance activities, congestion, etc. since such information is *relevant to* the *purpose* of *use* (e.g., choosing the best route to destination).

3- Information timeliness, considering the two different types of information, we differentiate between two cases:

Timeliness of produced information, to capture timeliness related aspects we extend the *state variable* concept with two attributes, a *time stamp* that specifies when the value of the *state variable* has been created/changed, and a *volatility* (real volatility) that describes the change rate of its value. Consider for example, a position of a car is a *state variable*, the *time stamp* specifies when such location has been taken, and the *volatility* defines until when such value (the location of the car) is valid in term of time. Similarly, we extend the *information* concept with *time stamp*, and we extend the *produced information* concept with a *volatility* (corresponding volatility) attribute that describes the change rate of its value. For example, the value of the previously mentioned information may change in the real world; therefore, the value of its corresponding *produced information* in Information System (IS) needs to be updated accordingly. To this end, the timeliness of *produced information* can be analyzed by comparing its *corresponding volatility* with the *real volatility* of its corresponding *state variable*. If they are identical or close enough, we say produced information is valid in term of time, otherwise, we say it is out-dated (not valid). For example, the GPS system needs to update the value of the *produced information* taking into consideration the volatility of the value of its corresponding *state variable*.

Timeliness of created information, we extend the *created information* concept with a *validity time* attribute that specifies the validity of the information value. For example, a hotel

[2]Applies also to *state variables* created by intentional communicative activities

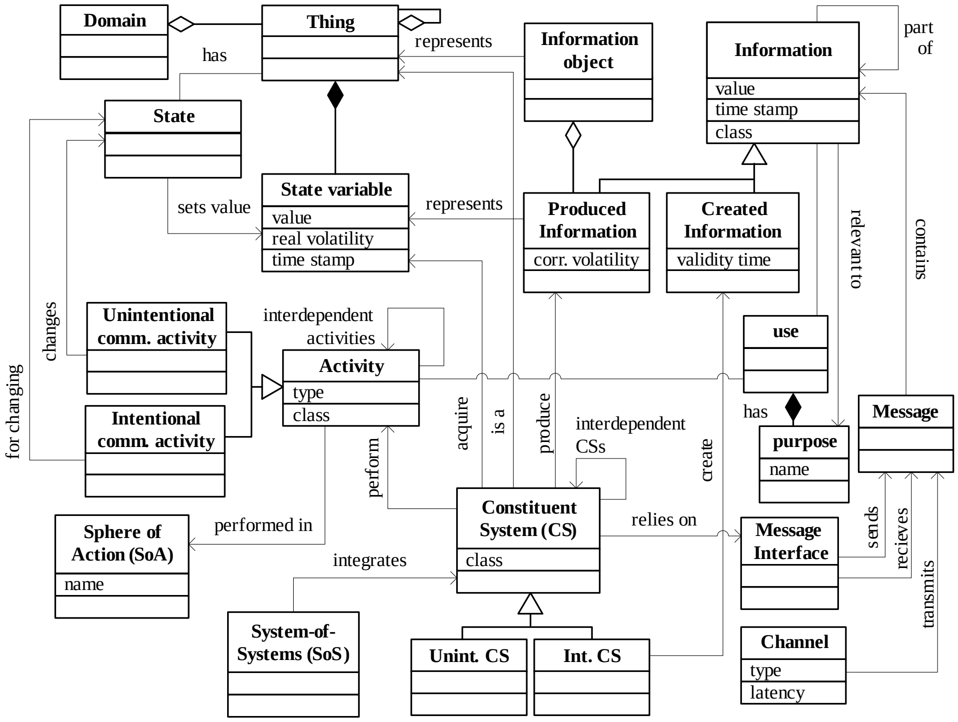

Fig. 3. The meta-model of the propose conceptual model

may advertise its services (created information) as a point of interest notification to passing cars for a defined period of time (validity time). The timeliness of *created information* can be analyzed by comparing its use-time (acquired from its time stamp) with its *validity time*, if its use time is less than its validity time, information is valid, otherwise, it is out-dated.

Information might be transmitted among CSs, which influences its timeliness. Therefore, the *channel* concept is extended with a *latency* attribute to capture the delay of time between information sender and its receiver, which has to be considered while analyzing its timeliness.

4- Information consistency[3] arises when multiple records of the same information are being used by several CSs for *interdependent activities*, where *interdependent activities* are *activities* that belong to the same *activity type*, and performed/executed in the same *Sphere of Action (SoA)*. A SoA can be defined as a partial part of the domain where activities can be performed. CSs who aim to perform *interdependent activities* are called *interdependent CSs*. Inconsistency is a situation where *interdependent users* depends on inconsistent information for performing their *interdependent activities*. This can result from providing/updating information used by *interdependent CSs* with different delay/update time/rate.

Consider for example, a Road Side Unit (RSU) located at highway entrance point that collects and disseminates merge-related information concerning the highway entrance point that can be used by nearby passing vehicles and vehicles that are trying to merge into the highway traffic to avoid collisions. Entering the highway and passing safely through the lanes close to its entrance are *interdependent activities*, and these activities belong to the same *activity type* (pass safely). Moreover, all of these activities are performed within the same *sphere of action* (highway entrance point). Thus, all of these vehicles (drivers) are *interdependent CSs* for merge-related information shared by the RSU, and in turn, such information should be consistent among all of them.

The meta-model of the proposed conceptual model is depicted in Figure 3.

V. Illustrating the Utility of the Conceptual Model

In what follows, we illustrate the utility of the conceptual model by applying it to a realistic scenario concerning VANETs. Consider for example a driver named John aims to reach his destination safely within the shortest possible time.

In order to reach its distention within the shortest possible time, John needs to depend on a *map* that is made available to drivers by a trusted maps provider (e.g., Google maps). Moreover, to facilitate choosing the shortest possible time, the *map* should contain information concerning 1- *route condition* information (e.g., maintenance, repair, etc.), which can be acquired from several trusted local authorized; and 2- *traffic congestion* information that can be obtained either from RSU (a trusted source) or other vehicles using the same route. Usually, vehicles may not be trusted for such information, because they may disseminate false information (e.g., bogus information attack) in order to affect the decisions of other drivers. More specifically, both of *route condition* and *traffic*

[3]In this paper, consistency refers only to value consistency

congestion are *relevant_to* the *purpose* of its *activity* "reach destination within the shortest possible time". Therefore, both of them are considered as sub-parts (*part_of*) of the *map*, i.e., if they were not made available to John, the *map* will be considered incomplete for the *purpose of use*.

On the other hand, John needs to avoid collisions in order to arrive at his destination safely, to simplify the scenario we consider John is able to avoid collisions in general and he needs assistant while passing intersections. John needs to slow down its speed and stop if he is not likely to pass while a green light is on. If John seems to violate a red light, RSU starts notifying all vehicles trying to safely pass (*interdependent activities*) through this intersection (*sphere of action*), and this information should be consistent among all of these vehicles (*interdependent CSs*) in order to avoid a possible collision.

Finally, John needs to be notified when there are nearby gas stations; therefore, he is interested in receiving notifications about nearby gas stations through PoI notifications. PoI notifications are produced by PoI and sent to RSUs to be broadcast to passing vehicles. Although PoI notification received from RSUs are trusted (considered accurate), but RSUs should not trust all PoI for such notifications, since PoI may provide falsified (inaccurate) information for their own benefits, e.g., increase traffic flow to their facilities. Thus, PoI are not classified as *trusted CSs* for such information, the activities they are performing to produce such information may not be *legitimate*, and the information content may not be *safe*. Moreover, RSU needs to verify the timeliness of such notification before providing it to passing vehicles. Other PoI notifications might be received by John, but they are discarded if they are not related to the purposes of activities he performs.

VI. Conclusions and Future Work

In this paper, we advocated that IQ is very important for the performance of any SoS, and presented a new conceptual model that provides concepts/attributes specialized for analyzing IQ for SoS in terms of the four core IQ dimensions. Moreover, we illustrated the applicability of our model by applying it to a realistic example from the ITS domain.

For the future work, we are currently working on extending the conceptual model for designing SoSs introduced in AMADEOS (Architecture for Multi-criticality Agile Dependable Evolutionary Open System-of-Systems) project[4] to accommodate the new concepts/attributes concerning IQ, and then we are planning to derive the corresponding SysML profiles [23], [24] from the extended conceptual model . In addition, we intend to extend the IQ dimensions we considered in this paper, and better investigate their inter-dependencies. We aim to provide more refined concepts to perform more expressive analysis for IQ related aspects rather than the binary one. Finally, we aim to better validate our model by applying it to other case studies from different domains.

[4]http://amadeos-project.eu/

Acknowledgment

This work has been partially supported by the REGIONE TOSCANA POR FESR 2014-2020 SISTER "SIgnaling & Sensing Technologies in Railway application", and by the PIRSES-GA-2013-612569 DEVASSES project.

References

[1] M. Jamshidi, "System of systems engineering-new challenges for the 21st century," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 23, no. 5, pp. 4–19, 2008.

[2] Bondavalli, A., Ceccarelli, A., Lollini, P., Montecchi, L., Mori, M., "System-of-Systems to Support Mobile Safety Critical Applications: Open Challenges and Viable Solutions," *(2016) IEEE Systems Journal*.

[3] M. W. Maier, "Architecting principles for systems-of-systems," in *INCOSE International Symposium*, vol. 6, no. 1. Wiley Online Library, 1996, pp. 565–573.

[4] M. Efatmaneshnik and M. Ryan, "Failure propagation in sos: Why sos should be loosely coupled," in *System of Systems Engineering (SOSE), 2014 9th International Conference on*. IEEE, 2014, pp. 49–54.

[5] H. Kopetz, "A conceptual model for the information transfer in systems-of-systems," in *International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing*. IEEE, 2014, pp. 17–24.

[6] B. Frömel, "Interface design in cyber-physical systems-of-systems," in *SoSE Conference, 2016 11th*. IEEE, 2016, pp. 1–8.

[7] M. Gharib and P. Giorgini, "Modeling and reasoning about information quality requirements,"in *International Conference on Requirements Engineering: Foundation for Software Quality*. Springer, 2015, pp. 49–64.

[8] A. Bondavalli, S. Bouchenak, and H. Kopetz, *Cyber-Physical Systems of Systems*. Springer International Publishing, 2016.

[9] L. L. Pipino, Y. W. Lee, and R. Y. Wang, "Data quality assessment," *Communications of the ACM*, vol. 45, no. 4, pp. 211–218, 2002.

[10] M. Bovee, R. P. Srivastava, and B. Mak, "A conceptual framework and belief-function approach to assessing overall information quality," *Int. Journal of Intelligent Systems*, vol. 18, no. 1, pp. 51–74, 2003.

[11] L. Liu and L. Chi, "Evolutional data quality: A theory-specific view." in *IQ*, 2002, pp. 292–304.

[12] M. Gharib and P. Giorgini, "Dealing with information quality requirements," in *International Conference on Enterprise, Business-Process and Information Systems Modeling*. Springer, 2015, pp. 379–394.

[13] H. Kopetz, B. Frömel, and O. Höftberger, "Direct versus stigmergic information flow in systems-of-systems," in *System of Systems Engineering Conference (SoSE), 2015 10th*. IEEE, 2015, pp. 36–41.

[14] J. Miller, "Vehicle-to-vehicle-to-infrastructure (v2v2i) intelligent transportation system architecture," in *Intelligent Vehicles Symposium*. IEEE, 2008, pp. 715–720.

[15] R. Ackoff, "From data to wisdom," *Journal of applied systems analysis*, vol. 16, no. 1, pp. 3–9, 1989.

[16] E. Quigley and A. Debons, "Interrogative theory of information and knowledge," in *Proceedings of the SIGCPR conference on Computer personnel research*. ACM, 1999, pp. 4–10.

[17] A. Aamodt and M. Nygard, "Different roles and mutual dependencies of data, information, and knowledge – An AI perspective on their integration," *Data & Knowledge Engineering*, vol. 16, no. 3, pp. 191–222, 1995.

[18] D. Stenmark, "The relationship between information and knowledge," in *Proceedings of IRIS*, vol. 24, 2001, pp. 11–14.

[19] M. K. Buckland, "Information as thing," *Journal of the American Society for information science*, vol. 42, no. 5, pp. 351–360, 1991.

[20] C. Dai, D. Lin, E. Bertino, and M. Kantarcioglu, "An approach to evaluate data trustworthiness based on data provenance," in *Workshop on Secure Data Management*. Springer, 2008, pp. 82–98.

[21] R. Wang and D. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of MIS*, pp. 5–33, 1996.

[22] M. Bunge, *Treatise on Basic Philosophy: Ontology I: The Furniture of the World*, ser. Treatise on Basic Philosophy. D. Reidel, 1977.

[23] P. Lollini, M. Mori, A. Babu, and S. Bouchenak, "Amadeos sysml profile for sos conceptual modeling," in *Cyber-Physical Systems of Systems*. Springer, 2016, pp. 97–127.

[24] Mori, M., Ceccarelli, A., Lollini, P., Frömel, B., Brancati, F., Bondavalli, A., Andrea, "Systems-of-systems modeling using a comprehensive viewpoint-based SysML profile," (2017), *Journal of Software: Evolution and Process*, Article in Press.