Published in: Proc. of ISORC 2014, Reno, Nevada, June 2014. IEEE Press.

A Conceptual Model for the Information Transfer in *Systems-of-Systems*

Hermann Kopetz
Institute für Technische Informatik
Vienna University of Technology
1040 Vienna, Austria
e-mail: H.Kopetz@gmail.com

***Abstract*—A *system-of-systems* (SoS) is a large information processing system formed by the integration of autonomous computer systems, physical systems and humans for the purpose of providing new synergistic services and/or more efficient economic processes. The integration is achieved by the exchange of *information* among the constituent systems. In a *monolithic system*, where a single context for the representation of information prevails, the issue of *context dependence of information representation* is of little concern. With the advent of SoSs the context dependence of the information representation in the diverse constituent systems becomes a central issue that must be dealt with by a proper conceptual model. In this paper we present such a conceptual model for the information transfer in a system-of-systems that takes account of the dissimilar representations of the same *semantic content* in the diverse constituent systems. We introduce the key concept of an *Information Atom*, called *Itom* that comprises data and an explanation of the data in an *atomic unit*. Depending on the context of interpretation, the representation of the data and the explanation may change, while the semantic content, the *information* carried by an Itom must remain invariant. In the second part of the paper we focus on the impact of the progression of real-time on the utility of information.**

Keywords: System-of-Systems; SoS; data; information; conceptual model, Itom, utility of information, time.

I. INTRODUCTION

The domain of *Systems-of-Systems (SoS)* is a relatively new field of computer science that is concerned with the design and operation of large information processing systems that are composed of existing or new *autonomous constituent computer systems (CS), physical machines and humans.* It is assumed that the integration of the CSs will improve current economic processes and provide new synergistic services. The integration of CSs into SoSs is already happening on a wide scale. It is achieved by the exchange of information processed and stored in the diverse CSs with the human users of the SoS and the physical environment.

In [Jam09, p.5] Jamshidi states that *Integration is probably the key viability of any SoS. Integration of SoS implies that each system can communicate and interact (control) with the SoS regardless of their hardware, software characteristics, or nature.*

The *cyber space* of a SoS is formed by the interacting computer systems, while the observed/controlled equipment and the human users/operators form the *environment*.

The constituent computer systems (CSs) of a SoS are designed and managed by different organizations. Each autonomous organization follows its own purposes and adheres to its own idiosyncratic standards and rules that characterize its *architectural style* of system development (e.g., how data is structured and named, what type of protocols are in use, etc.). It is thus quite probable that the architectural styles of two different constituent systems that must interact within a SoS will be incompatible with regards to the representation of data and the use of protocols. A conceptual SoS model for the information transfer among the CSs of a SoS that provides means to resolve these incompatibilities among the CSs is thus of utmost relevance in the domain of SoS design.

A conceptual model is an abstraction, i.e., a *simplification* of reality with the goal to explain precisely questioned phenomena, such as the exchange of information or the composite behavior of the CSs of a SoS at a proper level of abstraction. In order to be able to understand a model, we must take account of the limited rational capabilities of the human mind when designing the model [Kop11, p.40]. Finding a proper level of abstraction for a model that is supposed to explain precisely specified phenomena of a SoS is more of an art than a science. According to [Bat92] conceptual modeling *is by far the most critical phase of database design and further development of database technology is not likely to change this situation*. Integrating the different databases that are part of the CSs of a SoS makes the conceptual modeling activities even more relevant in SoS design.

A conceptual modeling process starts with the identification and formation of stable core concepts that capture the essential characteristics in *the Universe of Discourse* (UoD) [Cas10] during the *Interval of Discourse* (IoD). In a second phase the relationships among these concepts and the technical terms that are employed in the UoD by technicians and by human users must be identified and integrated within the conceptual model.

It is the objective of this paper to present a conceptual model that describes the information transfer among the autonomous constituent computer systems of a SoS, the human operators and the physical machine environment at a proper level of abstraction.

The paper starts with a fresh look at the concepts of *data* and *information* and introduces the notion of an *Information Atom*, abbreviated *Itom,* as a core concept of the paper in Section II. In our model, an Itom is the smallest atomic unit that can carry *information*. After elaborating on the properties of an Itom and the representation of an Itom in cyber-space and at the human-machine interface, Section III establishes a classification schema for Itoms, while Section IV looks at the impact of the progression of time on the utility of the information contained within an Itom. In Section V we elaborate on the communication among diverse constituent systems of a SoS, the physical environment and the human operators. The different types of processing Itoms are shortly discussed in Section VI. The paper terminates with a conclusion in Section VII.

II. Data verus Information

The colloquial use of the term *information* includes structured and unstructured data, text, pictures and different type of *media content* that *inform* about some part of the state of the world. Colloquially, the term is used as a noun with substantially different shades of meaning in different communities. A survey of the different meanings conveyed by the terms *data* and *information* shows that these fundamental concepts are not well defined in the domain of information science [Zin07].

Claude Shannon, who developed a scientific theory of *information transmission*, once remarked: *It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field* [Sha93, p. 180]. Colin Cherry in his book on *Human Communication* remarked in [Che80] on page 9: *As a theory it* (Shannon's mathematical theory of information) *lies at the syntactic level of sign theory and is abstracted from the semantic and pragmatic level.*

In the context of a conceptual model of information transfer in an SoS, our focus is on the *semantic and temporal aspects* of information. Our views on the concept *information* have been influenced by the work of Bar-Hillel and Carnap [Bar64] and Floridi [Flo05] on *semantic information*. In [Flo05, p.353]] Floridi has coined the term *infon* that refers to *discrete items of information*. This term *infon* is related to, but has substantially different attributes from our term *Itom* described below.

A. The Concept of an Itom

An *Itom (Information Atom)* is a tuple consisting of *data* and an *explanation* of the data. It is a central concept of our conceptual model.

1) Data in an Itom

A data item is an *artifact,* a *pattern,* created for a specified *purpose*. According to [Hil11, p.1] *an artifact may be defined as an object that has been intentionally made or produced for a certain purpose*. Depending on the generation process, we distinguish between *afferent* data and *efferent* data of a system.

Afferent data is created by an observation of the environment of a system by a data acquisition process. The *properties of an entity* that are observed can exist in space (e.g., an *image of person*) or in time (e.g. an *audio fingerprint* of a song). The *non-occurrence* of an expected event by a specified instant, a *timeout*, is an example for the observation of a *non-physical* phenomenon.

The acquisition of *afferent data* can often be compared to taking a snapshot of an entity to create an *image e.g*., taking a snapshot of a *person*. The *raw data (i.e. the pattern* of the image*)* is determined by the *elementary discriminatory capabilities of the sensor system*, e.g., the *eye* or the *camera* (a given number of pixels of varying intensity). The characteristics of the image depend not only on the properties of the observed entity, but also on the *purpose* that determines the position and view of the camera and the instant of taking the snapshot. The purpose has thus an effect on the characteristics of the image i.e., the *acquired afferent data.*

Efferent data is produced by a system and published at an interface to the environment. An example for efferent data is a control output produced by a system for the control of the environment.

After (*in the temporal sense*) *raw afferent data* has been created by a sensor system it can be stored on a *physical entity* called the *data carrier* in order that it can be interpreted or processed at some later time, producing *refined data*. The data carrier can be biological (e.g., human memory) or technological (e.g., computer memory). The *storage of data* and the *retrieval of stored data* require energy.

Refined data is data that has been created by a purposeful process from the raw data to simplify the explanation of the data in a given context. Refined data eliminates redundancy and abstracts from those properties of the raw data that are not relevant for the given purpose (but maybe relevant for another purpose). The representation of refined data is determined by the *architectural style*, i.e. conventions and standards in the given context.

Encrypted data (called *ciphertext*) is data has been subjected to a transformation (*encryption*) such that an unauthorized user cannot easily interpret the encrypted data.

The process of data creation, i.e., an *observation of an entity* or the *production of a result,* begins at a *start instant* and terminates at a *termination instant* (An *instant* is cut of the physical timeline that can be recorded by a *timestamp—see Section IV*). The *timestamp of the data* is determined by the termination instant of the production process. If afferent data relates to a *dynamic* entity, i.e., an entity that changes its properties as time progresses, then the explanation of an Itom is only complete if the *timestamp of the data* (i.e., the termination instant of the observation*)* is recorded as part of the Itom. If the data is the result of an observation of a *static entity* no such timestamp is needed. In the following the focus of our analysis is on dynamic entities.

2) Explanation of the Data

The *explanation of the data* establishes the links between data and already existing concepts in the mind of a human receiver or the rules for interpreting the data by a machine. It thus reveals the *meaning* of the data. The *explanation of data* forms an integral part of an Itom. The given environment determines the *context* for the explanation. The context provides the structure and meaning of the symbols that can be used in the explanation of an Itom. The *explanation* of an Itom consists of five parts that must give answers to the following five central questions:

- ***Identification***: *What entity is involved*? The entity must be clearly identified in the space-time reference frame, established by the *Universe of Discourse (UoD)* and the *Interval of Discourse (IoD).* With the coming *Internet of Things*, the magnitude of the UoD in cyberspace is vastly expanded.
- ***Purpose***: *Why is the data created*? This answer establishes the link between the *raw data*, the *refined data* and the purpose of the SoS. The answer to this question is addressed to a human user or to the designer of a computer system that creates and processes data. The concept of *purpose* is alien to a machine processing information. Depending on the purpose, we distinguish between *archival data* and *control data*. *Archival data* reports about an observation at an instant and is collected for archival purposes. *Control data* is concerned with the ongoing control of a process in the environment. Control data looses its utility as real-time progresses.
- ***Meaning***: *How has the data to be interpreted by a human or manipulated by a machine*? If the answer to this question is directed towards a human, then the presentation of the answer must use symbols and refer to concepts that are familiar to the human. If a computer acquires afferent data, then the explanation must specify how the data must be manipulated and stored by a computer. If a machine in the environment (e.g., a control valve) accepts *efferent* data from a computer system, it must be explained how the machine has to translate the data into physical action.
- ***Time***: *What are the temporal properties of the data*? Depending on the chosen *Interval of Discourse (IoD)* we can classify data as *static* or *dynamic*. *Archival data* of a *dynamic entity* must include the instant of observation in the Itom. In control applications it is helpful to include a second timestamp, a *utility instant* that delimits the utility of the control data as part of the Itom. If efferent data is output to a machine (e.g. a control valve) it must be made clear exactly when the efferent data has to be presented to the machine.
- ***Ownership***: *Who owns the data?* The question of ownership of *personal data* relates to the legal issues about privacy and is a topic of intense societal debate.

Additionally, an explanation might contain evidence that helps to establish the authenticity and integrity of the data contained in an Itom (e.g., cryptographic keys).

3) Information carried by an Itom

By applying the *explanation* to the *data* of the Itom the *semantic content, i.e. the information item carried by the Itom*, is revealed. *An information item is a timed proposition about some state or behavior in the world, disclosing associations among concepts.* If the proposition informs about a static entity, then the specification of the time is not required. The *information*, i.e., the *semantic content* captures the *deep invariant meaning* of an Itom. *Since data* without an implicit or explicit *explanation* is meaningless, an *Itom* is the smallest *atomic unit* that can carry *information.* This view of *information* is in agreement with the *General Definition of Information (GDI)* presented by Floridi, [Flo05] who defines *information* as *data plus meaning.*

The *representation* of an Itom (the data and the associated explanation) depends on the cultural and/or technological context of the *Itoms environment,* and whether the Itom is destined for a human user or for a machine. If we move an Itom to a new context, the representation of the data and the explanation may change, but the semantic content of the Itom must remain *invariant*. (Consider the example of the translation of the *meaning* of a sentence from one language to another language.) Since an Itom must be represented by *symbols* that are *determined by* and *understandable in* the given context and there exists no absolute context, the representation of an Itom is always relative, while the *semantic content* carried by an Itom, the *information,* has a connotation of *absoluteness*.

To facilitate the exchange of information among heterogeneous computer systems in the Internet, *markup languages,* such as the *Extensible Markup Language XML* [WWW13] that help to explain the meaning of data have been developed. Since in XML the explanation is separated from the data, the explanation can be adopted to *the context of use* of the data. Markup languages provide a mechanism to support an explanation of Itoms.

An example for the *invariant semantic content* of an Itom is an Itom that denotes the *speed of a particular car at a given instant*. Assume that this Itom is transferred from a US context to European context. In an US context, the data will be e.g, *60* and the explanation will say that number has to be interpreted as *miles per hour*. In an European context the data will be *96* and the explanation will say that this number has to be interpreted as *kilometers per hour*. Although on the surface the two representations of the Itom are fundamentally different—with respect to the data as well as with respect to the explanation—the semantic content, i.e., the *information* carried by these two Itoms is the same.

B. Properties of an Itom

Let us now look at some of the properties of an Itom:

1) ***Name***: In human (and machine) communication *names* are assigned to Itoms. The name of the Itom is used to reference the Itom and should also designate the concept that helps to explains the meaning of the Itom to a human. The naming context is determined by the cultural/technological environment of the designer of the Itom and is likely to be dissimilar in different CSs. A *shared ontology* that establishes a uniform naming context for all Itoms in the CSs of an SoS is a desireable but *hardly achievable* goal in a large SoS [Ken78].

2) ***Purpose***: Every artifact is directly or indirectly created for a *purpose by a human author.* The purpose defines why the data is created, what is the aim of this effort, and how this effort is related to the purpose of the overall system. It also defines the *viewpoint* (or perspective) for the data-abstraction from the *raw data* to the *refined data*. The data-abstraction determines what properties of the *raw data* (an observed entity) must be retained and what properties may be disregarded in the data-refinement process for the given purpose [Kop11, p. 41]. Consider the example of a camera that monitors the traffic on a highway for the purpose of toll collection. The image taken by a camera (*raw data*) is analyzed to identify the license plate of a car (*refined data*). The license plate can be considered a symbol [New72,p.25] that indirectly designates the driver's bank account from where the highway toll has to be deducted. Sometimes data is used for another than the original purpose, giving rise to many discussions in the realm of legality and data privacy.

3) ***Truthfulness***: Our concept of an Itom does not make any assumptions about the truthfulness of the semantic content, the *information*, in the Itom. We thus can attribute factual information as *true information (correspondance theory of thruth [Gla13])*, *misinformation* (accidentially false) or *disinformation* (intentionally false), or just call it *information* if we don't know yet if it is *true* or *false*. It is often the case that only some time after data has been acquired it can be decided whether the information is *true* or *false* (e.g., consider the case of *value error* of a sensor).

4) ***Temporal Aspects****:* The utility of the semantic content of an Itom, the *information*, can be time-dependent. The inclusion of a timestamp that records when the data in an Itom has been created and possibly another timestamp that records when the data in the Itom *looses its utitility* as part of the explanation makes it possible to restrict the *utility* of the *information* to specified real-time intervals in the Interval of Disocourse (IoD).

5) ***Neutrality****:* The semantic content of an Itom does not depend on the state of knowledge of the human receiver of the Itom. The aspect of *newness of information* to the receiver and associated statistical metrics about the *subjective value* or *the utility of information* are not part of our concept of an Itom (see also *Section II D*).

6) ***Physicalism****:* The storage of every Itom requires a *physical data carrier*, i.e., in our model there exists no *stored information* without such an associated physical data carrier. Imprinting data on and retrieving data from the data carrier as well the processing of data require *energy*.

7) ***Relativity****:* Since there exists no *absolute context,* there cannot be an *absolute representation* of information. Since data is created for a *purpose*, the information carried by an Itom is relative to that purpose.

C. Itoms in Cyberspace

In cyberspace all *data* is encoded in *bit strings.* The *explanation* within an Itom consists of two parts, we call them *computer instructions* and *explanation of purpose*.

The *computer instructions* instruct the computer system how the *data bit-string* is partitioned into syntactic chunks and how the syntactic chunks have to be stored, retrieved, and processed by the computer. This part of the *explanation* can thus be considered a *machine program* for a (virtual) computer. Such a machine program is also represented by a bit-string. We call the data bit-string *object data* and the instruction bit-string that *explains* the object data, *meta data*. A computer Itom thus contains digital *object data* and digital *meta data*. The recursion stops when the *meta data* is a sequence of well-defined machine instructions for the *destined* computer. In this case, the *design of the computer* serves as an *explanation for the meaning of the data*.

The second part of the explanation of an Itom, the *explanation of purpose,* is directed to humans who are involved in the design and operation of the computer system. The *explanation of purpose* is part of the documentation of the cyber system and must be expressed in a form that is *understandable* to the human user/designer.

An example for an Itom in cyberspace is the construct of a *variable* in a programming language. The *value* of a variable corresponds to the *data* at the *instant* of accessing the variable and the *name* of the variable points to the explicit or implicit explanation of the *data.* The temporal attributes of this Itom are implicit—on reading a variable the last value written will be retrieved.

D. Itoms for Humans

An Itom destined for a human must be *understandable* to the human receiver. *Understanding* means that the patterns, symbols, names, concepts and relations that are used to represent the data and the explanation of the Itom can be linked with the existing concepts in the *conceptual landscape* in the human mind of the receiver. The conceptual landscape refers to the *personal knowledge base* that is built up and maintained in the mind of an individual over her/his lifetime. This personal knowledge base consists of concepts and links among the concepts [Kop11,p35]. (A concept can be considered a *unit of thought*.)

In many situations, part or all of the explanation of the data in an Itom is taken *implicitly* from the context. This can give rise to serious misunderstandings and has been identified as the cause of fatal accidents, i.e., the crash of an airplane on Jan 20, 2002 [Avi02].

The *understanding of the information* carried by an Itom is improved if the concepts and links used in the explanation of an Itom are tightly integrated in the *conceptual landscape* of the receiver. It can happen that information carried in an Itom is incomprehensible to one receiver, while making sense to another receiver who has the proper conceptual background for understanding the presented information.

The *utility of the information* carried in an Itom is *relative to the human* who receives the information. If an Itom contains information that is new to a receiver, it causes modifications in the *conceptual landscape* of the receiver. The extent of these modifications is a measure of the utility of the information carried by the Itom relative to the receiver. If the receiver already knows the information contained in the Itom, no modification of the conceptual landscape is required—the information contained in the Itom has no utility to this receiver. In general it is *not possible* to quantify the *utility of information* to a human receiver because the state of the conceptual landscape of the human receiver at the instant of receiving an Itom cannot be grasped. Bar Hillel and Carnap [Bar64] have developed a quantitative measure for the *value of information* based on the changes in a *well-defined formal state* of a hypothetical receiver—a proposal that is closely related to our view of the *utility of information*.

If the context of an Itom changes from a machine environment to a human environment, then the representation of the data and the form of the explanation must change to suit the opportunities and constrains of human perception and human cognition in the given context, still keeping the semantic content of the Itom invariant. The information carried by an Itom does not depend on any particular representation.

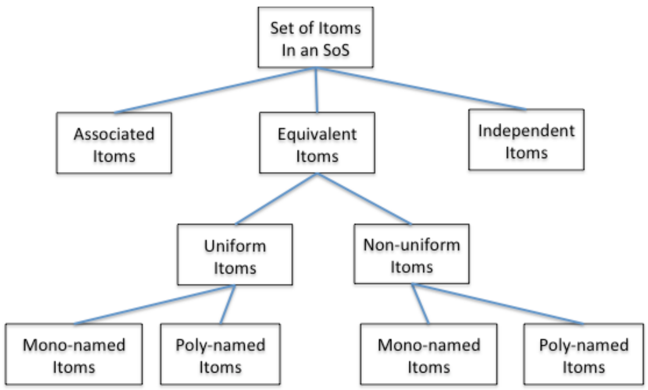

Fig. 1.: Classification of Itoms

III. Classification of Itoms

The first step in the analysis of an SoS concerns the identification of the Itoms that are referenced in the interactions among the CSs of the SoS. In the next step the Itoms should be classified in order to establish their relations. We propose a classification schema as depicted in Fig. 1:

- **Equivalent Itoms:** Itoms that carry the *same semantic content*. Example: Two Itom that refer to the same amount of gold, where in one the weight is expressed in grams and in the other the weight is expressed in ounces. The relation between the units (grams, ounces) is static.
- **Associated Itoms:** Itoms that carry a *closely associated semantic content*. Example: Two Itoms that refer to the value of the same amount of gold, one amount valued in Euros, the other amount valued in Dollars. The relation between the units (Euro, Dollar) is dynamic (i.e., depends on the time). A set of *associated Itoms* is called a *family of Itoms*.
- **Independent Itoms:** Itoms that carry independent semantic content, but can be linked in a particular context. The links among independent Itoms must be identified in the given application scenario.
- **Uniform Itoms**: *Equivalent Itoms* in different CSs that express the data in the same format.
- **Non-uniform Itoms**: *Equivalent Itoms* in different CSs that express the data in different formats.
- **Mono-named Itoms**: *Equivalent Itoms* that are referenced in the different CSs by the same name.
- **Poly-named Itoms:** *Equivalent Itoms* that are referenced in the different CSs by different names.

In an ideal world, all *equivalent Itoms* should be *uniform* and *mono-named*. But in reality, this ideal world is hardly achievable. It is a more realistic approach to translate the representation and the names of equivalent Itoms at the boundaries of the CSs, i.e, in the gateway components of the CSs as outlined in Section VB.

IV. Time

Most entities of interest change their properties as *physical time* progresses. This progression of *physical time* affects an Itom in cyber space in four ways:

- Data about a dynamic entity, i.e., an entity that changes its properties as time progresses, is only a *valid image* of the current state of an observed entity for a limited duration that depends on the dynamics of the observed entity and the purpose of the observation. For example in automotive engine control, the current position of a piston in an automotive engine where the crankshaft rotates at 6000 revolutions per minute is only valid for a few microseconds.
- The relationship among *associated Itoms* is time-dependent.
- Since the acquisition, processing and transportation of data takes physical time, the data carried in a stored Itom is always outdated.
- The local clocks of the constituent systems of an SoS that are used to generate and interpret the timestamps of an Itom cannot be fully synchronized.

In this Section we first draw the attention to the fundamental difference between *physical time* and computational progress (sometimes called *logical time* [Ray00]), before discussing the issues that must be considered when designing a global time base in an SoS. In the final part of this Section we introduce the concept of *time-series* of Itoms.

A. Physical Time versus Logical Time

In our model, *physical time* is based on Newtonian physics. Newtonian physics represents time by a directed line, an arrow (the *time-line*) that extends from the past to the future. A cut of the time-line is called an *instant* and an interval between two instants is called a *duration*. An occurrence that happens at an instant is called an *event*. To enable the measurement of the instant of event-occurrence or the lengths of durations, an international standard of physical time, the *International Atomic Time (Temps Atomique International TAI)* has been established [TAI13]. The *TAI second* is the basis for the definition of the *constants of physics* and forms the *unit of time* in the *laws of physics*. One worldwide distribution system for the TAI is the GPS navigation system [Hof07] that is capable to distribute the global TAI time worldwide with a precision of better than

100 nsec. In Newtonian physics, physical time is an *analog quantity*.

Logical time denotes the progress of computations in computer systems. Logical time expresses the *causality relations* among computational events in a *closed* distributed computer system [Ray00]. There is no concept of *duration* in the domain of logical time.

In the physical part of a cyber-physical system (CPS), it is the progression *of physical time* that matters. A lacking distinction between the concepts *physical time* and *logical time* is leading to a substantial confusion in the domain of real-time computer systems.

Synchronous programming languages [Hal93] make a crystal-clear distinction between the progress of physical time and the progress of logical time. In the synchronous programming model physical time advances in discrete steps, called *ticks*. The computations within a timeless step are considered to happen infinitely fast, i.e., all computations within a step are completed before the next tick occurs.

B. Sparse Global Timebase

The timestamps contained in an Itom can only be interpreted in the context of an SoS if all constituent systems have access to a *proper global time.* The approximation of the progress of the *analog physical time* of the physical space by the progress of a *digital global time* in cyber space leads to *fundamental limits* in digital time measurement that are discussed in [Kop11, p.61]. These fundamental limits which are caused by the impossibility of perfectly synchronizing digital clocks in a distributed system and the discreteness of the time representation, affect the *truthfulness* and the *consistency* of the timestamps generated in the cyber-system of an SoS. Consistency of the timestamps within the cyber system can be achieved by the introduction of a *sparse global time base*, while the truthfulness problem can be alleviated by an improved precision of the global time. It is however impossible to completely resolve this *truthfulness-consistency conflict*.

In the following we assume that a *sparse global time base of adequate precision* is available in any CS of our SoS and this time base is used for the generation and interpretation of the timestamps of an Itom. Within the SoS, the timestamps are thus *consistent*. We further assume that the consequences of the *imperfect truthfulness* of the timestamps are second order effects that can be disregarded.

C. Time Series of Itoms

In the following we call the data and the timestamps of an Itom the *values of the Itom* and the explanation without the timestamps the *static explanation* of the Itom.

In many computer applications, e.g., in real-time control systems, the same entity is observed cyclically at different instants. This leads to a time-series of Itoms that contain different *values* (data and timestamps), but the same static explanation. We call such a *time-series of Itoms* that can be characterized by a single static explanation a *data type*. The machine specification of the data type does not include the *explanation of purpose*, since *purpose*, meaning *primary reason*, is only relevant for humans but not for machines. In order to reduce the storage space required for the storage of time-series, the parts of an Itom are segregated in the memory of a machine. The static explanation, i.e., the specification of the data type is stored only once and the sequence of values is stored separately. However, the meaning of a data item can only be recovered if the data and the explanation are merged again.

V. Communication

As outlined in Section I, communication among the CSs, communication with the physical environment and human-machine communication are crucial actions in any SoS. The purpose of any communication is the transmission of semantic content, i.e. *information*, from the *sender* to one or more *receivers*.

A. The Concept of a Message

In cyberspace a *message* is a construct that is formed for the purpose of transmitting a bit-string from a sending CS to one or more receiving CSs via a communication channel. A bit-string without an associated explanation contains *data*, but no *information*. In order to recover the information contained in the Itom of the sender, versions of the Itoms that conform to the context of the receivers must be created out of the received data and the explanation of the data.

At our level of abstraction, we are not concerned with the detailed techniques of how the transmission between a sender and a receiver is achieved (e.g., by wireless of wire-bound methods), but take note of the elapsed physical time (the *transport interval*) between the instant of *sending a message* at the sender and the instant of *receiving a correct instance of message* at the receiver. The *transport interval* is important, because it can impact the *temporal utility* of the data contained in the message.

A message is *correctly received*, if the bit-string of the message is not changed during transport. It is assumed that a message contains enough redundant data (e.g., a check field) such that the corruption of a message during transport can be detected and the receiver can discard a corrupted message. A receiver thus receives either correct data or no data at all (*fail-silent fault model* of message transmission).

Since the *cost of transmission* and the *duration of the transport interval* depend on the length of the bit-string in the message, there are convincing economic and technical arguments for reducing the length of messages, particularly in real-time applications. By separating the *dynamic data* from the *static explanation* (e.g., the specification of the *message type*) in an Itom, one can limit the message length to the dynamic data and reconstruct the semantic content of the Itom at the receiver by applying the local explanation to the received data. This requires a static explanation, which must either be part of the design of the receiver or communicated to the receiver in a separate message. If the contexts of the sender and receiver are different (which often will be the case in an SoS), then the explanation of the data at the sender and at the receiver will also be different.

If the endpoint of a communication channel, a *communication port*, is designed to deliver messages of a single data type only, then the association of the data with the static explanation can be established on the basis of the port name. If different messages types can be transported via the same port, then the *message name,* which establishes the link between the data and the explanation must be part of the bit-string in the message.

In some real-time communication protocols, such as TTP [Kop93] that have been designed to minimize the message length of messages carrying real-time data, the *periodic instant of receiving the time-triggered message* is used as a message name on one side and as the timestamp of the data contained in the message on the other side.

B. Communication among Constituent Systems

In this Section we look at the unidirectional transmission of information from a sending CS to a receiving CS in a SoS. We propose that the representational differences among the CSs of a SoS are resolved by dedicated *gateway components* at the boundaries of the SoS (Fig. 2) and that these representational differences do not influence the inner functions of a CS following the design principle that *representational differences are only of concern at system boundaries*.

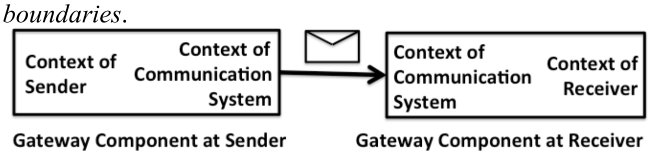

Fig. 2: Contexts in Unidirectional Communication

In the general case of Fig, 2, where third party international standards specify the syntax and semantics of the messages exchanged among the CSs, there are three contexts involved in the communication among the CSs of a SoS:

- The context of the sender, determined by the architectural style of the sender.
- The context of the communication system, determined by third party communication and data standards.
- The context of the receiver, determined by the architectural style of the receiver.

It follows that three versions of an Itom must be considered, each one tailored to the requirements and constraints of the particular context of the Itom.

The transformation of the representation of an Itom from one context to another context is performed by the two gateway components as shown in Fig. 1. The gateway component at the sender transform the *sender Itom* to the standardized data format required by the third party standards. The gateway component at the receiver *constructs a new equivalent Itom* that conforms to the architectural style of the receiver out of the data received on the communication channel and the explanation of the data that is part of the design of the gateway component at the receiver.

In the exceptional case of *uniform mono-named Itoms*, i.e. that the sending CS and the receiving CS conform to the same architectural style, *share the ontology* of the data and no third party standard is involved, gateway components are not needed. However, we take the realistic view that only in few cases it will be possible to establish a *SoS global shared ontology* and force all CSs to commit to these shared ontology.

After the receiver has accepted a syntactically correct message containing *control data*, the receiver must check whether the *temporal utility* of the control data contained in the message has already expired.

The receipt of a message containing information about the environment can be considered an *indirect observation*. In indirect observations issues relating to the *trust in the sender* must be considered

C. Input from the Environment

The explanation of afferent data collected from the physical environment by a sensor system is normally contained in the *specification of the sensor system*. The *raw data*, acquired by a sensor system is often preprocessed in the sensor system to derive at *refined data* that contains the features that are of relevance for the given purpose.

Let us look at an example of temperature measurement. In a wire-bound system the position of the temperature sensor at the end of the connecting wire identifies the temperature data in space. The designer of the sensor system will assign a *meaningful name* to this point in space. The instant of reading the sensor value by the CS is often taken as the timestamp of the data. The transformation of the representation of the *raw measured data* to standardized data is performed in the sensor system such that the user sees the temperature data in a standardized format, e.g., in degrees Kelvin (*refined data*).

In a wireless sensor system the identification of the point in space where the data is collected is more involved, since by moving the sensor in space a different entity is observed.

D. Output to the Environment

A CS of a SoS can manipulate real-world entities by outputting efferent control data via an actuator system to the external physical environment. The actuator system transforms the *information* of the Cyber System to a *physical quantity* (e.g., a force) that produces the intended effect in the physical world. The instant when exactly the intended effect must be achieved can be recorded in a timestamp of an output message. Being aware of the progression of global time, the actuator system will produce the desired effect exactly at the prescribed instant. We call an output message with such a timestamp a *timed output message.* The communication system must deliver a timed output message to the intelligent actuator component *before* the instant denoted by the timestamp in the message.

E. Human/Machine Communication

At the output side of the human-machine interface (HMI) the cyber system must represent the explanation of the efferent data in a form that is understandable to the human user. This can be achieved by assigning meaningful words to the data, thus establishing a *language-based link* between the *conceptual landscape of the human mind* [Bou61] and the *data*, or by building a frame of the environment (e.g., a schematic diagram or a picture) where the dynamic data is displayed at the proper position within the static frame. The static frame serves as the explanation of the data in the Itom.

At the input side of the human-machine interface (HMI) the human actions (e.g., tactile of acoustic) must be performed in a context such that *afferent data* can be interpreted by the machine.

VI. Processing of Itoms

Information processing realizes a function that maps the current state and a set of an *input Itoms* to a new state and a set of *output Itoms*. We distinguish between three types of information processing:

- Transformation of the Representation of Itoms.
- Validation of the Truthfulness of Itoms.
- Creation of Derivative Itoms.

A. Transformation of Equivalent Itoms

Whenever an Itom is copied to a new context, the representation of the data and the explanation within the Itom must change to adapt to the idiosyncrasies of the new context. This transformation of the representation must not influence the semantic content of equivalent Itoms, i.e., the *information* carried within the Itoms must remain unaffected. These transformations of the representation should be carried out in dedicated gateway components that are placed at the boundaries of a CS.

B. Validation of the Truthfulness of Itoms

The input information that is collected at an interface of a CS may be false, due to human error or a sensor malfunction. Since every sensor will eventually fail, it is essential to validate acquired information by comparing it with redundant information that is derived from an independent source, e.g., another sensor, or another human input or a model that examines the consistency of all related information items. In the technical literature the term *agreed data* is used to denote data that has passed the input validation.

C. Creation of Derivative Itoms

The most important type of Itom processing relates to the application specific derivation of *result Itoms* that provide a solution to the stated SoS problem. In a cyber-physical system (CPS), the control data of the result Itom are directly applied to the physical environment by the actuator system.

In many applications, derivative Itom processing leads to a tremendous *data reduction*. For example, in an Automated Emergency Braking System (AEBS) of a vehicle, [Kop13] in every cycle of about 10 msec gigabytes of input data collected by video cameras, radar and laser systems are processed and reduced to a single byte of output data for the timely activation of the brakes of the vehicle.

VII. Conclusion

In this paper we have presented a conceptual model for the information transfer in a System-of-Systems and introduced the abstract concept of an *Information Atom*, called *Itom* that encapsulates *data* and *explanation of the data* in an atomic unit to be able to express the *semantic content*, i.e., the *information* contained in an Itom independently of the concrete realization of the Itom in any CS of an SoS. The inclusion of time in the explanation of an Itom provides the important capability to precisely express information about dynamic properties of changing entities.

A description of an SoS based on Itoms and their relations puts the description on a higher level of abstraction than a description based solely on data. By disregarding representational details of information representation, and focusing on the essential subject matters, the description can be substantially simplified. We conjecture that the *Itoms and their temporal and causal relations* establish the backbone of the architectural framework of any SoS.

VIII. Acknowledgements

This work has been supported, in part, by the European FP7 research project AMADEOS Grant Agreement 610535 on *Systems of Systems*.

References

[Adr12] Adrians, P. *Information*. Stanford Encyclopedia of Philosphy. 2012

[Avi02] Aviation Safety Network. Accident Desription. URL: http://aviation-safety.net/database/record.php?id=19920120-0 accessed on June 10, 2013

[Bar64] Bar-Hillel, Y, Carnap, R. *An Outline of a Theory of Semantic Information.* In Bar Hillel, ed. *Language and Information*. Addison Wesley, 1964, pp. 221-274.

[Bat92] Batini, C. et al. *Conceptual Data Base Design*. Benjamin Cummings, Redwood City. 1992

[Bou61] Boulding, K.E. *The Image.* Ann Arbor Paperbakcs, 1961

[Cas10] Castro, l. et al. A Linguistic Approach to Conceptual Modeling with Semantic Types and Onto UML. Proc. of the 14th IEEE International Enterprise Distributed Object Computing Conference. 2010.

[Che80] Cherry, C. *On Human Communication*. MIT Press. Third edition, 1980.

[Flo05] Floridi, L. *Is Semantic Information Meaningful Data?* Philosophy and Phenomenological Research. Vol. 60. No. 2. Pp. 351-370. March 2005.

[Flo10] Floridi,L., *Information—A Very Short Introduction*. Oxford University Press. 2010

[Gla13] Glanzberg, M. *Truth*. Stanford Encyclopedia of Philosphy. 2013

[Hal93] Halbwachs, N. *Synchronous Programming of Reactive Systems.* Kluwer Academic Press. 1993.

[Hil11] Hilpinen, R., *Artifact*. Stanford Encyclopedia of Philosphy. 2011

[Jam09] Jamshidi, M., Systems of Systems Engineering. John Wiley and Sons. 2009.

[Hof07] Hofmann-Wellenhof, B. et al. *GNSS - Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and more*. Springer Verlag, 2007

[Ken78] Kent, W., *Data and Reality*. North Holland Publishing Company, 1978.

[Kop11] Kopetz, H.. *Real-time Systems–Design Principles for Distributed Embedded Applications*. Springer Verlag, 2011.

[Kop93] Kopetz, H. et al. TTP-A Time-Triggered Protocol for Fault-Tolerant Real-Time Systems. Proc. FTCS 23. IEEE Press. Pp.524-532. 1993.

[Kop13] Kopetz, Poledna, S. *Autonomous Emergence Braking—A System-of-Systems Perspective.* Proc. of the WORCS 2013, IEEE press.

[New72] Newell, A., Simon, H., *Human Problem Solving*. Prentice Hall, 1972.

[Ray00] Raynal, M. *Logical Time in Distributed Systems*. URL: ftp://ftp.irisa.fr/local/caps/DEPOTS/TrUE/ASR/Logical-time.pdf Retrieved on April 8, 2013.

[Sha93] Shannon, C.E. *Collected Papers*. ed. by N. J. A. Sloane and & A. D.Wyner (Los Alamos, Ca: IEEE Computer Society Press).

[WWW13] World-Wide Web Consoritum. Extensible Markup Language. URL http://www.w3.org/XML/ Retrieved on April 18, 2013.

[TAI13] Temp Atomique International. *Internation Atomic Time.* Wikipedia. Accessed on April 4, 2013.

[Zin07] Zins, C., *Conceptual Approaches for Defining Data, Information and Knowledge*. Journal of the Amercian Society for Information Science and Technology, Vo. 58.N. 4. pp.479-493. 2007.