STS-IQ – IQ: Analyzing Information Quality Requirements by Examples

In what follows, we show how STS-IQ framework analysis IQ requirements based on seven IQ dimensions:

1- Information accessibility: is the extent to which information is available [1], i.e., accessibility concerns information availability along with the required permission over it to perform a task at hand. In our work, information availability can be analyzed depending on information provision between information producer(s) and its user(s). While we rely on permissions availability (e.g., (P)roduce, (R)ead, (M)odify and (S)end) to analyze whether an actor has the required permissions to perform a task at hand.

For example, in Fig. 1 NYSE needs to read ``trading orders'' that is provided by Pro trading along with a read permission over such information. If Pro trading did not provide ``trading orders'' or it did not provide the read permission over it to NYSE, accessibility issues will arise.

2- Information completeness: means that all parts of information should be available, and information should be complete for performing a task at hand [2]. In our work, information completeness is analyzed depending on two sub dimensions:

(1) ***Value Completeness:*** information is preserved against corruption or lost that might endanger its integrity, i.e., information provision might affect the quality of the transferred information. Thus, we rely on the information provision type (Normal Provision (P)/ Integrity Preserving (IP) provision) to analyze the value completeness f information.
(2) ***Purpose of use completeness:*** information is complete for performing a task at hand, i.e., all the required information for performing a specific task should be available. To analyze information purpose of use completeness, we rely on ``part of" concept to model the relation between a composite item and its sub-items., which helps in analyzing whether information is complete for the purpose of use or not.

For example, in Fig 1 both NASDAQ and NYSE need to read CME CB info that is produced by CME. If CME provide such information through normal provision [P], it will not be considered as value complete. While if CME did not provide such information, to NASDAQ/NYSE, both NSQ CB info/NYSE CB info will be considered as incomplete for the purpose of use.

3- Information timeliness: means to which extent information is valid in term of time (e.g., sufficiently up-to-date) [1]. In our work, to model information timeliness first, we extended ***information*** construct with a ***(V)olatility*** attribute to represent the change rate of information value [3], and we extended ***information provision*** construct with a ***time*** attribute to capture the amount of time an information provision requires (referred to as the transmission time in telecommunication networks [4]). Second, since we have two different relations between goals and information that can be affected by time related aspects (e.g., reads and sends). Thus, we extend them to accommodate attributes that enable to capture information timeliness, as follows:

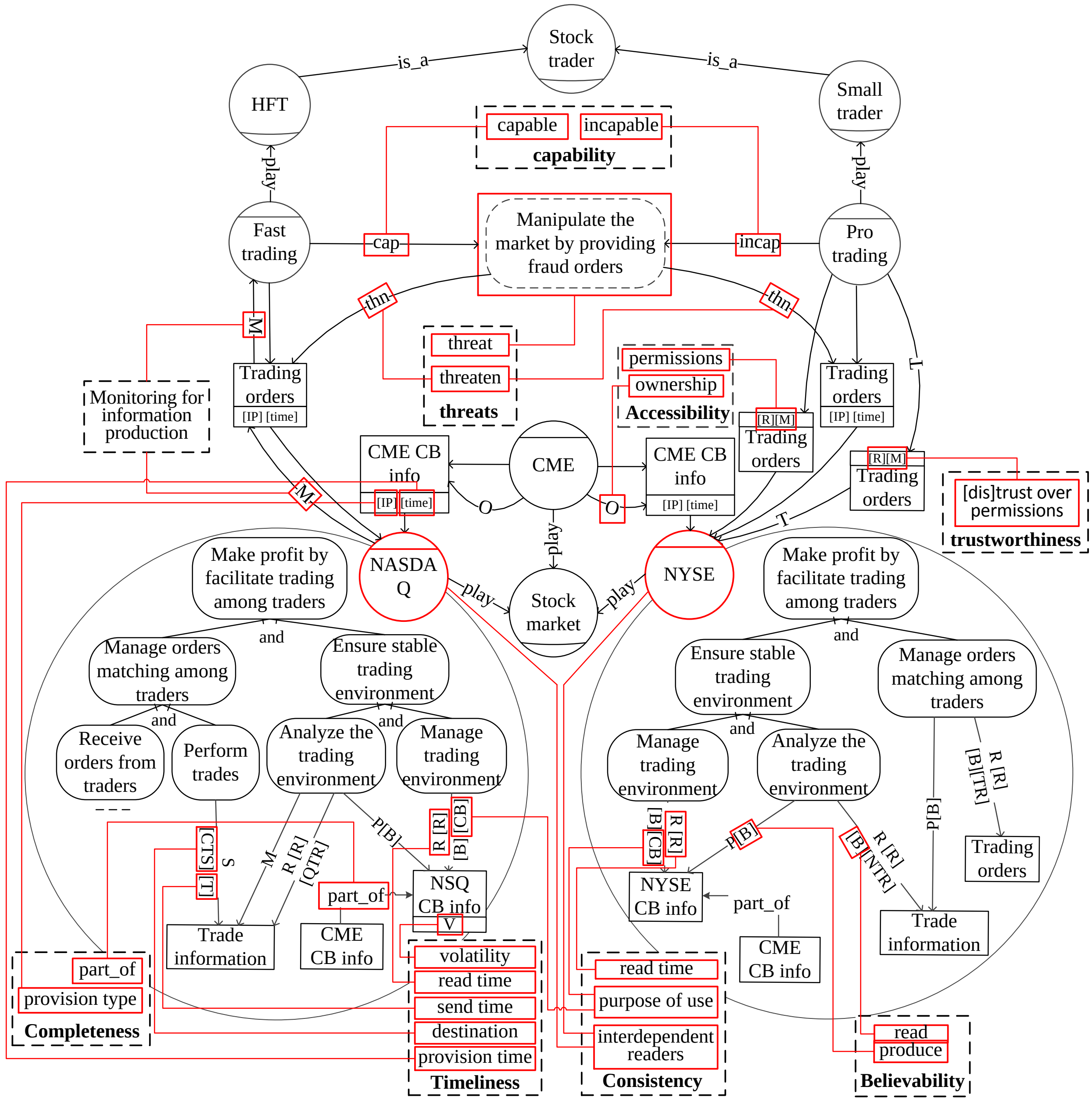

Figure 1 A partial goal model concerning the stock market structure

(1) ***Read timeliness***: in order to capture information validity for read, we propose ***read time***[1] concept that capture the actual ***usage time*** of information, and it enables to determine information ***currency (age)***, which can be used to analyze the timeliness of read information, i.e., information is valid, if its currency is smaller than its ***volatility***, otherwise it is not invalid;

(2) ***Send timeliness:*** in order to capture information validity at its intended destination, we extend ***send*** relation with two attributes: (1) ***(D)estination*** that represents the intended send

[1] Read time can be derived by analyzing the model, i.e., there is no specialized construct or attribute to represent it.

destination of information; and (2) ***(S)end time*** that represents the allowed amount of time for information to reach its final destination.

For example, in Fig 1 the goal ``Manage orders matching among traders'' reads ``trading orders'' that are provided by Pro trading. However, such information is considered as not valid for read, if its read time is higher than their volatility rate. On the other hand, the goal ``Perform trades'' is responsible of sending ``Trade information'' to CTS in time [T], such information is considered invalid for send if its provision time to its destination is bigger than its required send time.

4- Information consistency: means all multiple records of the same information should be the same across time and space [2]. In our work, information consistency related issues arise only when there are multiple records of the same information that are being read by actors for ***interdependent purposes*** (goals), i.e., if such records of the same information are being read by actors for different purposes of usage, information consistency between these actors will not be an issue.

In order to capture consistency, we extend the ***read*** relation between a goal and information with ***Purpose Of Use (POU)*** attribute that captures the intended purpose of information usage, which enables to identify ***interdependent readers*** that are actors who read the same information for the same ***POU***. To this end, consistency among ***interdependent readers*** can be analyzed based on information ***read times***, i.e., information is consistent among its ***interdependent readers***, if all of them have the same ***read time***, otherwise it is inconsistent.

For example, in Fig 1 both NASDAQ and NYSE read CME CB info for the same purpose of use [POU] (e.g., CB), which make them an interdependent readers for such information. Thus, their read time of such information should be the same otherwise inconsistency issue will arise between them.

5- Information believability can be defined as the extent to which information is accepted or regarded as true [1] [2]. Since only two relations between goals and information (e.g., ***read*** and ***produce***) can be influenced by information believability. Thus, these two relations are extended with an attribute to determine whether these two relations accommodate a ***believability check*** for read/produced information or not. In particular, produced/read information is believable from the perspective of its producer/reader; if the produce/read relations apply a believability check that guarantees the believability of the produced/read information.

For example, in Fig 1 the goal ``Analyze the trading environment'' apply a believability check [B] for ``Trade information'' it reads, which enables the goal to detect whether such information is believable or not, i.e., it allows to reject information with unbelievable values. At the other hand, and in order to avoid producing information with unbelievable values, the goal ``Analyze the trading environment'' apply a believability check [B] for ``NYSE CB info''.

6- Information trustworthiness can be defined as the extent to which information is credible [5]. In our work, information ***trustworthiness*** is analyzed based on the ***trustworthiness of the provenance***. Following [6], we analyze information ***trustworthiness*** depending on the ***trustworthiness of its source*** and the ***trustworthiness of its provision***. Information trustworthiness: can be subject to the trustworthiness of the source and the trustworthiness of the provision:

(1) ***Trustworthiness of the source:*** in [7] [8] we provided constructs to model the trustworthiness of information sources, yet we did not provide any constructs that enable to analyze the trustworthiness of the source. In this paper, we analyze the trustworthiness of information source by identifying and modeling the different ***intentional threats*** [9] (***threats*** for short) that might ***threaten*** the ***trustworthiness*** of information it produces. In other words, information sources (actors) are intentional entities and they might be motivated to produce falsified, fraud, biased information, and such motivations ***threatens*** the ***trustworthiness of the source***, and in turn, it ***threatens*** the trustworthiness of the produced information. Considering the complex nature of such threats, not all information sources (actors) have the required capabilities to achieve them. Thus, we need to identify and model the actors' capabilities toward achieving them as well.
We classify actors' capabilities toward a threat achievement under:

(1) ***Capable:*** when the actor has the competency of achieving the threat;

(2) ***Incapable:*** when the actor does not have the competency of achieving the threat.

In particular, an actor is considered trustworthiness for producing an information item, if such information item does not has related threats that compromise its trustworthiness, or it has a threat but the actor does not have the capability to achieve it. While the actor is untrustworthiness for producing information item, if there is a related threat and the actor have the capability to achieve such threat.

(2) ***Trustworthiness of the provision[2]:*** can be analyzed based on the way information arrives to its final destination [10] [11], taking into consideration the operations that have been applied to it (e.g., modify), and whether the actor who perform such operations was authorized or not (e.g., if it has the required permissions or not), and whether such authorization is trusted or not.

For example, in Fig 1 we have "Manipulate the market by providing fraud orders" as a threat to ``trading orders" trustworthiness, where Pro trading does not have the capability to achieve such threat. While Fast trading has the capability to achieve it, thus, Fast trading is untrustworthiness source for ``trading orders", since it has the capability to achieve the related threat.

7- Information accuracy means that information should be true or error free with respect to some known or measured value [2]. Information accuracy is the most important and studied IQ dimension, yet analyzing accuracy is not possible without defining clear criteria. In our work, we analyze information accuracy based on two sub dimensions: ***trustworthiness*** and ***believability***. Dai et al. [6] stated that information accuracy is highly influenced by information ***trustworthiness***. While Wang and Strong [3] argued that accuracy can be analyzed based on several dimensions including ***believability***. Thus, we analyze information accuracy based on these two sub dimensions. In our work, accuracy concerns arise only when information is produced and read[3]. Thus, we analyze accuracy for produce and read as follows:

(1) ***Accuracy of the produced information***: can be analyzed based on its ***believability***, which enables to avoid producing unintended information, and the ***trustworthiness of the production process*** that arise when the producer is not the owner of information[4], since the producer might not produce the owner intended information. Considering an investor who wants to produce orders by

[2] No special constructs are added to capture the trustworthiness of the provision

[3] Information accuracy concerns during the transfer is covered by the trustworthiness of information provenance.

[4] We trust the owner for producing its intended information.

itself, if the produce relation does not apply a believability check, there will be no guarantee that it will not mistakenly produce the wrong information (unbelievable information), i.e., information accuracy is not guaranteed. On the other hand, to analyze the accuracy of the investor's order that is produced by a trader, we need to check whether the trader is has the produce permissions and there is a trust relation for such permissions between the investor and the trader, which guarantee that the trader will produce the orders intended by the investor. With the absene of such trust relation, we cannot guarantee the accuracy of the produced orders from the investor perspective.

(2) ***Accuracy of read information***: can be analyzed based on information ***believability*** and ***trustworthiness***. Considering a stock market that reads a trading order, if the read relation does not apply a believability check, there will be no guarantee that the order has a believable value (price), and in turn information accuracy will not be guaranteed. As previously mentioned applying a believability check will enable for avoiding unbelievable orders such as stub quotes, which were considered as a main reason of the flash crash [12]. On the other hand, to analyze the accuracy of the received orders, we need to analyze the trustworthiness of their provenance (e.g., the trustworthiness their producing along with the trustworthiness of their transfer), which can be done as described earlier.

Bibliography

[1] Leo L Pipino, Yang W Lee, and Richard Y Wang, "Data quality assessment," *Communications of the ACM*, vol. 45, no. 4, pp. 211-218, 2002.

[2] Matthew Bovee, Rajendra P Srivastava, and Brenda Mak, "A conceptual framework and belief-function approach to assessing overall information quality," *International journal of intelligent systems*, vol. 18, no. 1, pp. 51-74, 2003.

[3] R.Y. Wang and D.M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of management information systems*, pp. 5-33, 1996.

[4] A Behrouz Forouzan, *Data Communications \& Networking*.: Tata McGraw-Hill Education, 2006.

[5] Liping Liu and Lauren Chi, "Evolutional Data Quality: A Theory-Specific View.," in *IQ*, 2002, pp. 292-304.

[6] Chenyun Dai, Dan Lin, Elisa Bertino, and Murat Kantarcioglu, "An approach to evaluate data trustworthiness based on data provenance," in *Secure Data Management*.: Springer, 2008, pp. 82-98.

[7] Mohamad Gharib and Paolo Giorgini, "A Framework for Information Quality Requirements Engineering," in *Joint Proceedings of*

[8] Mohamad Gharib and Paolo Giorgini, "Dealing with Information Quality Requirements," in *Enterprise, Business-Process and Information Systems Modeling*.: Springer, 2015, pp. 379-394.

[9] Axel Van Lamsweerde, "Elaborating security requirements by construction of intentional anti-models," in *Proceedings of the 26th International Conference on Software Engineering*, 2004, pp. 148-157.

[10] Peter Buneman, Sanjeev Khanna, and Tan Wang-Chiew, "Why and where: A characterization of data provenance," in *Database Theory—ICDT 2001*.: Springer, 2001, pp. 316-330.

[11] Yogesh L Simmhan, Beth Plale, and Dennis Gannon, "A survey of data provenance in e-science," *ACM Sigmod Record*, vol. 34, no. 3, pp. 31-36, 2005.

[12] Andrei Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun, "The