Towards Unbiased Evaluation of Uncertainty Reasoning: The URREF Ontology

Paulo C. G. Costa, Kathryn B. Laskey
C4I Center - George Mason University
Fairfax, VA, 22030
{pcosta, klaskey} @c4i.gmu.edu

Erik Blasch
Air Force Research Lab
Rome, NY, 13441
erik.blasch@rl.af.mil

Anne-Laure Jousselme
Defence R&D Canada-Valcartier
Québec City, QC, G3J 1X5
Anne-Laure.Jousselme@drdc-rddc.gc.ca

Abstract— *Current advances in technology, sensor collection, data storage, and data distribution have afforded more complex, distributed, and operational information fusion systems (IFSs). IFSs notionally consist of low-level (data collection, registration, and association in time and space) and high-level information fusion (user coordination, situational awareness, and mission control), which require a common ontology for effective communication and data processing. In this paper, we describe the ontology reference model developed as part of the uncertainty representation and reasoning evaluation framework (URREF). The URREF ontology is intended to provide guidance for defining the actual concepts and criteria that together comprise the comprehensive uncertainty evaluation framework being developed by the Evaluation of Technologies for Uncertainty Representation Working Group (ETURWG).*

Keywords: *Component, Information Fusion, Performance Evaluation, Uncertainty Reasoning, Knowledge Representation, Ontology, Measures of Effectiveness.*

I. INTRODUCTION

The evaluation of how uncertainty is dealt with within a given IF system is distinct from, although closely related to, the evaluation of the overall performance of the system. Metrics for evaluating the overall performance of IF systems are more encompassing in scope than those focused on the uncertainty handling within the system. The metrics for the overall system include the effects of the uncertainty representation, but there are also effects of other aspects of the fusion system that can affect the performance of the system.

For example, fusion-system-level metrics include time-liness (how quickly the system can come to a conclusion within a specified precision level), accuracy (where can an object be found for a specified localization level) and confidence (what level of a probability match for a defined recall level).Clearly, different choices in uncertainty representation approaches will affect the achievable timeliness, accuracy, and confidence of a system, and therefore must be considered when evaluating both the system's performance as a whole and the specific impact of the uncertainty handling approach. Yet, when evaluating timeliness (or any other system-level metrics), one will likely find some factors not directly related to the handling of uncertainty itself[1], such as object tracking and identification report updates (i.e., Level 1 fusion), situation and threat assessment relative to scenario constraints (i.e., Level 2/3 fusion), overall system architecture (e.g. centralized, distributed, etc.), data management processes and feedback / input control processes (i.e., Level 4 fusion considerations), and user-machine coordination based on operating systems (i.e., Level 5 fusion), and others.

In an ideal situation, evaluating how the management of uncertainty affects the overall performance of a fusion system would be just a matter of isolating the directly related aspects. Unfortunately, real-life information fusion systems are usually too complex to allow for such a clear-cut separation, and most of the aspects considered in system-wide performance are entangled with or influenced by uncertainty representation considerations to some degree. Isolating the impact of uncertainty handling in an IF system is thus a matter of understanding how intertwined the choice of an uncertainty representation and reasoning approach is to the major performance metrics used to evaluate the IF system itself.

The basic premise of the ETURWG is that achieving a level of understanding that supports an unbiased evaluation of the impact of uncertainty in an IF system is a task by itself complex enough to warrant status as an open research area. Understanding what the distinctions are between fusion system-level performance criteria and fusion uncertainty representation performance criteria is the focus of the group. This paper presents one of the initial results of the ETURWG, the URREF ontology, and addresses how it is being currently used to support the development of an unbiased uncertainty reasoning evaluation framework.

The paper is divided as follows. Section II explores the main ideas currently at the state of the art on evaluation of IF systems. Section III introduces some of the aspects of the impact of uncertainty representation and reasoning with some examples from the World Wide Web. Section IV follows with a discussion on the scope of the uncertainty representation and reasoning evaluation framework. Section V presents the main aspects of the URREF ontology, which is the basis to ensure that the framework can be used in different setups without losing consistency of its meaning.

[1] This holds at least at a higher level of abstraction. There might be interactions between the uncertainty representation approach and these system factors, but we begin with a presumption that they are not significant.

II. EVALUATION OF IF SYSTEMS

The objective of this paper is to present the main concepts that support the ETURWG work on evaluating the impact of uncertainty in IF systems. Before addressing evaluation from a broader perspective, it is important do emphasize the different aspects involved in evaluating low-level information fusion (LLF) and high-level information fusion (HLF) systems, since the appropriate set of effectiveness metrics for high-level information fusion (HLF) evaluation is not necessarily standard within the fusion community. We suggest that *HLF effectiveness* has three parts: information gain, quality, and robustness, each of which requires higher fidelity analysis and characterization for uncertainty evaluation. The distinction between high-level and low-level fusion has propagated from the 1980's discussions surrounding the needs and the relevant processes for information fusion. The Joint Director of the Lab's (JDL) model and its subsequent revisions formed the high-low level distinction [1-5].

Figure 1 shows the elements of an information fusion model. LLF, including the physical-based parameters, lends itself naturally to quantifiable evaluation techniques. Current directions in measures of effectiveness (MOEs) for information fusion systems (IFSs) [6] need to address approaches beyond estimation (determining parameters from measured data) and fusion rules [7].

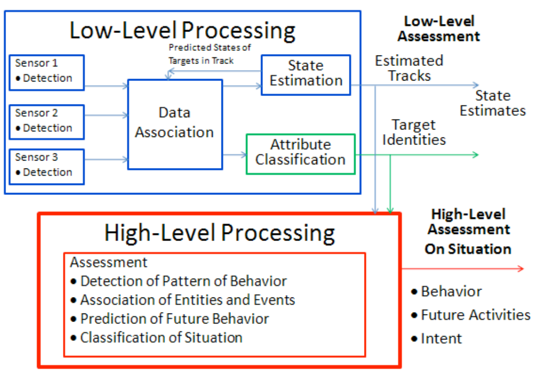

Figure 1 – Elements of Information Fusion

Low-level IFSs typically rely on standard metrics for evaluation such as timeliness, accuracy, and confidence. Given the broader use of IFSs, it is also important to look at high-level fusion processes and determine a set of metrics to test IFSs, such as workload, throughput, and cost. Three types of measures (measures of performance MOP, measures of effectiveness MOE, and measures of merit MOM) are summarized in the literature. In this paper, we seek to describe MOEs that relate to uncertainty for HLF which have been developed from Quality of Service (QOS) and Quality of Information (QOI) or Information Quality (IQ) standards that support the user and the machine, respectively.

Evaluating an information fusion system is not new. There is a large set of literature associated with measures of performance (MOP), MOEs, and measures of force effectiveness (MOFEs) based in estimation. Two excellent summaries include Ch11 from Waltz and Llinas [8] and Ch 20 from Llinas [9]. The compiled information from [8, 9] represent a comprehensive assessment of methods in estimation MOP/MOEs and discussion of HLF issues. However, addressing fusion process MOEs [10] as a system highlights the need for additional discussion on HLF performance metrics in real systems [11]. Systems perform well if they (1) support mission goals [12], (2) enhance operator work tasks [13, 14], and (3) reduce uncertainty [15, 16].

Effectiveness relates to a system's capability to produce an effect. Many benefits of fusion include providing locations of events, extending coverage, and reducing ambiguity and false alarms [17]. The goal for the IFS is to support users in their tasks whether providing refined information, reducing time pressures, or determining completeness, accuracy, and quality in task completion. Effectiveness includes [18]:

- *efficiency*: doing things in the most economical way (good input to output ratio)
- *efficacy*: getting things done, i.e. meeting targets
- *correctness*: doing "right" things, i.e. setting right targets to achieve an overall goal (the *effect*)

Inherent in the definition of effectiveness is a level of performance needed to accomplish a goal. While a large number of ideas and metrics could be postulated; all having their merits and limitations, we focus on three:

- *Information Gain* = valued added from which two pieces of information provide more content than individual pieces of information alone
- *Quality* = Measures of Performance that include accuracy, reduction in uncertainty, confidence, credibility and reliability
- *Robustness* = consistent over testing and application domains

Together, these definitions form the basis of "effectiveness" in (1) presenting high quality data, (2) being derived from more than one source, and (3) being consistent and reliable over the situation. These operating conditions from the object, sensor, and environment form a strategy for looking at effectiveness. Blasch [18] postulated a general HLF MOE as:

$$\text{Effectiveness} = \text{InfoGain} * \text{Quality} * \text{Robustness}$$

which is one formulation that requires updating based on a unified ontology framework. A taxonomy of other metrics could be expanded upon, in order to create an evaluation standard for MOEs. However, different environments would necessitate adaptive metrics tailored for the situation [18, 19, 20, 21].

IFSs include the technology, algorithms, and environment of operation (to include the people). Moving from LLF to the abstract reasoning requires integrating the user into the analysis, such as for command and control. For a system to be operational, it needs to verified (LLF question) and validated (HLF question); described briefly as:

- Verification: "Am I measuring the IFS correctly?"
- Validation: "Am I measuring the correct IFS?"

For instance, in an example of operational maintenance; the normalized timeliness metrics are separated to verify individual machine performance as:

Overall Equipment Effectiveness % =

Available % × Perf. Efficiency % × Quality Rate %

As per a baseline definition of effectiveness, we bring to light the need for information gain in establishing the timeliness, accuracy, and confidence associated with an IFS's ability to help the user reason over data, make decisions, and act on the information. Evaluation of uncertainty management considers the quality of information provided by an IFS. The ETURWG is bringing various contributions from the IF community to bear on this problem.

To address the MOEs for high and low-level fusion, we need to look at the quality measures being developed over various domains and reasoning methods. Industry standard definitions are Quality of Service (QOS)[2] and Quality of information (QOI)[3]. In Fusion 2005, Johnson and Chang [19] proposed QOI for data fusion in net centric "publish and subscribe" architecture to "update clients in a QOI paradigm rather than a QOS paradigm". They varied the message length in a QOI system versus fixed time metrics in a QOS system. To facilitate end user's needs in a net-centric environment, a QOI was used because of sensor-web enabled ontology development. They applied the QOI/QOS method to a target-tracking example in which they generalized the end user's needs for QOI parameters from which the tracking system conferred the QOS capabilities over state and covariance information. Yu and Sycara [20] addressed the QOI in a distributed decision fusion system by learning the parameters. They applied the technique to determine the QOI information on target reliabilities (or better termed confidences) from a Dempster-Shafer method. Quality of information impacts fusion decisions [18, 21].

Quality of Information is still an emerging topic as information is different for different users and systems. QOI integrity measures whether the data has not been manipulated as it impacts the shared situational awareness [22]. For example, QOI includes: Accuracy, Timeliness, Certainty, and Integrity [23]. There is a need to develop principals and relations between information management and sensor fusion [24].

Closely related to QOI, is *quality of service* (QOS) as it relates to the information flow and availability. QOS has been well vetted in the communications literature as including throughput, delay, error, and jitter.

QOI/QOS requires comparisons of:

- Usability versus Usefulness
- Accuracy versus Precision
- Verification versus Validity

from which we address information gain, quality, and robustness, respectively. As MOPs come from rigorous standard metrics to determine such things as accuracy, there is a need for pragmatic metrics to determine the validity of information aggregation for useful decision making. In the next section we will look at the IF-related literature that addresses issues of information service and type to advance the discussion in HLIF metrics.

III. The Uncertainty Representation Problem

The Information Fusion community envisions effortless interaction between humans and computers, seamless interoperability and information exchange among applications, and rapid and accurate identification and invocation of appropriate services. As work with semantics and services grows more ambitious, there is increasing appreciation of the need for principled approaches to representing and reasoning under uncertainty. Here, the term "uncertainty" is intended to encompass a variety of aspects of imperfect knowledge, including incompleteness, inconclusiveness, vagueness, ambiguity, and others. The term "uncertainty reasoning" is meant to denote the full range of methods designed for representing and reasoning with knowledge when Boolean truth-values are unknown, unknowable, or inapplicable. Commonly applied approaches to uncertainty reasoning include probability theory, fuzzy logic, subjective logic, Dempster-Shafer theory, DSmT, and numerous other methodologies.

To illustrate the challenges of evaluating uncertainty representation and reasoning in information systems, we consider below a few reasoning challenges faced within the World Wide Web domain that could be addressed by reasoning under uncertainty [25].

Uncertainty is an intrinsic feature of many of the required tasks, and a full realization of the World Wide Web as a source of processable data and services demands formalisms capable of representing and reasoning under uncertainty.

- *Automated agents* are used to exchange Web information that in many cases is not perfect. Thus, a standardized format for representing uncertainty would allow agents receiving imperfect information to interpret it in the same way as were intended by the sending agents.
- *Uncertainty-laden data*. Examples include weather forecasts or gambling odds. Canonical methods for representing and integrating such information are necessary for communicating it in a seamless fashion.
- *Non-sensory collected information* is also often incorrect or only partially correct, raising issues related to trust or credibility. Uncertainty representation and reasoning helps to resolve tension amongst information sources having different confidence and trust levels.
- *Dynamic composability* of Web Services will require runtime identification of processing and data resources and resolution of policy objectives. Uncertainty reasoning techniques may be necessary to resolve situations in which existing information is not definitive.
- *Information extracted from large information networks* such as the World Wide Web is typically incomplete. The ability to exploit partial information is very useful for identifying sources of service or information. For example, that an online service deals with greeting cards

2 http://en.wikipedia.org/wiki/Overall_equipment_effectiveness

3 http://en.wikipedia.org/wiki/Information_quality

may be evidence that it also sells stationery. It is clear that search effectiveness could be improved by appropriate use of technologies for handling uncertainty.

These problems are all related with information fusion, involve both LLIF and HLIF, and can be easily extrapolated to represent the more general classes of problems found in the sensor, data, and information fusion domain.

IV. The Uncertainty Evaluation Framework

This section is an initial attempt at determining the distinctions between evaluating the performance of an IF system and evaluating the impact of uncertainty on it. This would then allow us to establish an evaluation framework capable of supporting unbiased assessment of how the choice of uncertainty representation and reasoning impacts the performance of an IF system. The basic idea behind the framework is to analyze an abstract fusion system and define its input data and output products. In a futuristic prototypical information fusion system, the uncertainty representation approach would be "plug-and-playable." That is, one can run it with a Bayesian approach, then switch out the Bayesian approach for a Dempster-Shafer approach, then for a Fuzzy Random Set approach or have a combination of uncertainty reasoning methods. The input data are the same in each case, as are the output products (but not necessarily the values in the output products). Figure 2 below depicts the boundaries of the uncertainty representation and reasoning evaluation framework (URREF).

There are two elements in the picture that are exogenous to the evaluation framework, named in the picture as "World being sensed" and "World being reported." Between these two external elements, the boundary of the evaluation framework encompasses the way uncertainty is handled when data is input to the system, during the processes that occur within it, as well as when the final product is delivered to the IF system's users.

The first external element refers to the events of interest to the IF system that happen in the world and are perceived by the system sources. Note that the implicit definition of sources in this case encompasses anything that can capture information and send it to the system. That is, both hard sources (e.g. imaging, radar, video, etc.) and soft sources (HUMINT reports, software alerts, etc.) are considered external to the evaluation system with respect to their associated sensorial capabilities, while the way they convey their information is within the scope of the system [26, 27, 28].

This is an important distinction between the evaluation of an IF system, which usually encompasses the sensitivity of its sensors, and the evaluation of its handling of uncertainty, which focuses only on how the uncertainty embedded in the sensors' information is captured. The latter comprises what is called in the picture as the *Input step*, which involves assessing the system's ability to represent uncertainty as an intrinsic part of the information being captured. As an example, information regarding trust on the input from a given sensor is an important item to evaluate how the overall system handles uncertainty, although it might not be as critical for its overall performance. A key question for evaluating uncertainty representation is what the uncertainty characteristics of the input data are, and how they affect the use of different uncertainty schemes.

In the ideal system model, having the appropriate data characteristics is critical. If the characteristics do not span the range of uncertainty techniques, then the model may not give meaningful results about the operationally significant differences between the techniques. Correctly identifying the desired input data characteristics will shape the future development of use cases and modeling data sets for those use case.

Once information is in the IF system, it will be processed to generate the system's deliverable that requires uncertainty characterization and reporting in the *Output step*. This process involves fusion techniques and algorithms that are directly affected by the uncertainty handling technique being used, and its impact on the system's inferential process. In this case, the URREF criteria focus on aspects that are specific to the way uncertainty is considered and handled within the fusion process. This is not an evaluation of the system's performance as a whole. We want to understand how the uncertainty

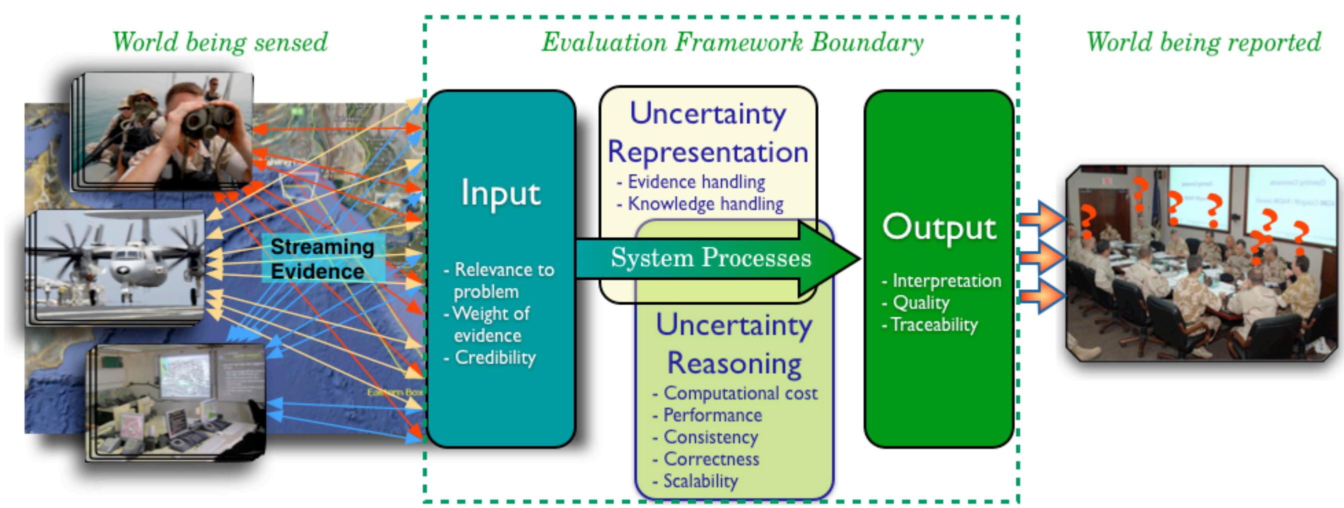

Figure 2 - Boundaries of the Uncertainty Representation and Reasoning Evaluation Framework.

representation affects system performance, and whether different uncertainty representation schemes are more or less robust to variations in the remaining parts of the IF system architecture. But we want to focus specifically on the uncertainty representation aspects, and attempt, as best as possible, to separate those aspects from overall system performance and architecture issues.

After the information is fused and properly treated, then it is conveyed to the system's users. In the figure, these are represented by an image depicting decision-makers who would likely be supported by the IF system in their daily tasks. The URREF output step involves the assessment of how information on uncertainty is presented to the users and, therefore, how it impacts the quality of their decision-making process.

V. The URREF Ontology

Within the URREF, a major task is to formally identify the concepts that are pertinent to the evaluation of uncertainty of an IF system, which is a means to ensure that all evaluations follow the same semantic constraints and abide by the same principles of mathematical soundness. The URREF ontology, whose main concepts are depicted in Figure 3 below, is a first step towards this goal and is meant to capture the main aspects to be considered in each step of the evaluation process. The core of the ontology is the Criteria class, which is were the bulk of the development work was focused on. The Uncertainty Classes were either taken or adapted from the Uncertainty Ontology developed by the W3C's URW3-XG [25]. The ontology must also be used as a high-level reference for defining the actual evaluation criteria items that will comprise a comprehensive uncertainty evaluation framework.

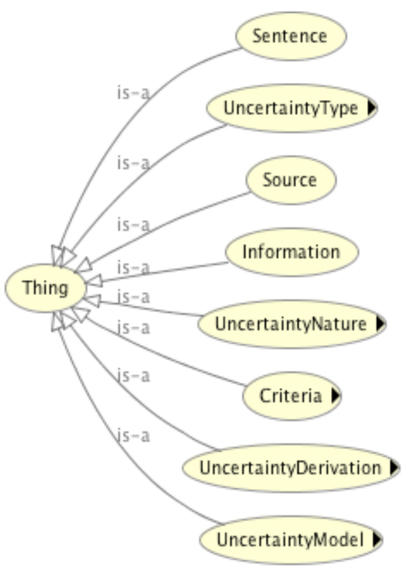

Figure 3 – The URREF ontology main classes.

Information in its most restricted technical sense is an ordered sequence of symbols that can be interpreted as a message. Information can be recorded as signs, or transmitted as signals. Information is any kind of event that affects the state of a dynamic system. Conceptually, information is the message (utterance or expression) being conveyed.

A. Source Class

A source is the origin of the information. A physical sensor is one important example of a source; natural language input from a human is another.

B. Sentence Class

Information is an expression in some logical language that evaluates to a truth-value (formula, axiom, assertion). It is assumed that information will be presented in the form of sentences. So uncertainty will be associated with sentences.

C. Uncertainty Nature Class

This class captures the information about the nature of the uncertainty, i.e., whether the uncertainty is inherent in the phenomenon expressed by the sentence or it is the result of lack of knowledge of the agent. Figure 4 depicts the Uncertainty Nature class and its subclasses.

Figure 4 – URREF Ontology: Uncertainty Nature Class.

1) *Epistemic Subclass:* Uncertainty Nature is considered epistemic when it is caused by lack of complete knowledge. That is, the event itself might be completely deterministic, but there is uncertainty about it due to missing information.
2) *Aleatory Subclass:* Uncertainty Nature is considered aleatory when it comes from the world; that is, uncertainty is an inherent property of the world. In contrast with Epistemic Uncertainty, which is due to the lack of complete knowledge.

D. Uncertainty Derivation Class

Uncertainty derivation refers to the way it can be assessed. That is, how the uncertainty metrics can be derived. Figure 5 depicts the Uncertainty Derivation class and its subclasses.

Figure 5 – URREF Uncertainty Derivation Class.

1) *Objective Subclass:* Uncertainty derivation is considered as objective when it can be assessed in an observer-independent, factual way, e.g., via a repeatable derivation process.
2) *Subjective Subclass:* Uncertainty Derivation is considered subjective when it is assessed via a subjective judgment,

e.g., a subject matter expert's (SME's) estimation, a gambler's guess, etc. Note that even though one might use formal methods for this assessment, it is the assessment itself that defines the Uncertainty Derivation as subjective. For example, a meteorologist may follow a formal procedure to establish a weather forecast, but if the numbers ultimately are derived from his judgment then it is a subjective uncertainty derivation.

E. Uncertainty Type Class

Uncertainty Type is a concept that focuses on underlying characteristics of the information that make it uncertain. Its subclasses are Ambiguity, Incompleteness, Vagueness, Randomness, and Inconsistency, all depicted in Figure 6 below. These subclasses were based on the large body of work on evidential reasoning by David Schum [29].

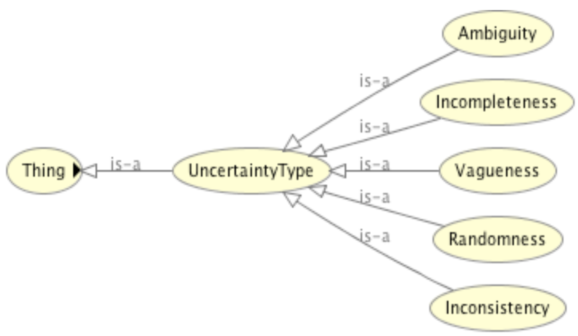

Figure 6 – URREF Ontology: Uncertainty Type Class.

F. Uncertainty Model Class

The Uncertainty Model class contains information on the mathematical theories for the representing and reasoning with the uncertainty types. The specific types of theories include, but are not limited to, the subclasses FuzzySets, BeliefFunctions, RoughtSets, ProbabilisticTheory, and RandomSets.

G. Criteria Class

This is the main class of the URREF ontology, and it is meant to encompass all the different aspects that must be considered when evaluating LLIF and HLIF uncertainty handling in multi-sensor fusion systems. Figure 7 depicts the Criteria class and its subclasses.

1) *Input Criteria:* This general concept encompasses the criteria that directly affect the way evidence is input to the system. It mostly focuses on the source of input data or evidence, which can be tangible (sensing or physical), testimonial (human), documentary, or known missing [29].

- *Relevance to Problem* assess how a given uncertainty representation is able to capture how a given input is relevant to the problem that was the source of the data request. This is a criterion specific to HLIF fusion systems that work at levels 3 and above of the JDL and the Data Information Fusion Group (DFIG) model [5].

- *Weight or Force of Evidence* assess how a given uncertainty representation is able to capture by the degree to which a given input can affect the processing and output of the fusion system. Ideally, this should be an objective assessment and the representation approach must provide a means to measure the degree of impact of an evidence item with a numerical scale. This criterion is especially useful for determining the value of information in systems that must trade-off their ability to capture more evidence with active sensors with the need to avoid being observed. That is, this criterion is especially important to systems that rely on value of information [24].

- *Credibility,* also known as believability, mainly comprises the aspects that directly affect a sensor (soft or hard) in its ability to capture evidence. Its subclasses are *Veracity*, *Objectivity*, *Observational Sensitivity*, and *SelfConfidence.*

2) *Representation Criteria:* This general concept encompasses the criteria that directly affect the way information is captured by and transmitted through the system. These criteria can also be called interfacing or transport criteria, as they relate to how the representational model transmits information within the system.

- *Evidence Handling* is a subclass of representation criteria that apply particularly to the ability of a given representation of uncertainty to capture specific characteristics of incomplete evidence that are available to or produced by the system. The main focus is on measuring the quality of the evidence by assessing how well this evidence is able to support the development of a conclusion. It has subclasses *Conclusiveness*, *Ambiguousness*, *Completeness*, *Reliability*, and *Dissonance.*

- *Knowledge Handling* encompasses criteria intended to measure the ability of a given uncertainty representation technique to convey knowledge. Its subclasses are *Compatibility* and *Expressiveness* (which is further divided into the subclasses *Assessment*, *Adaptability*, and *Simplicity*)

3) *Reasoning Criteria:* This general concept encompasses the criteria that directly affect the way the system transforms its data into knowledge. These can also be called process or inference criteria, as they deal with how the uncertainty model performs operations with information. It has the following subclasses:

- *Correctness* is a measure of the ability of the inferential process to produce correct results. In cases where there is no ground truth to establish a correct answer (including a simulated ground truth), the representation technique can still be evaluated in terms of how its answers align with what is expected from a gold standard (e.g. SMEs, documentation, etc.).

- *Consistency* is a measure of the ability of the inferential process to produce the same results when given the same data under the same conditions.

- *Scalability* is a measure of how a representational technique performs on a class of problems as the amount of data or the problem size grows very large. Scalability could be broken down into additional sub-criteria.

- *Computational Cost* is a measure of the amount of computational resources required by a given representational technique to produce its results.

- *Performance* include metrics to assess the contributino of the representational model toward meeting the functional requirements of an information fusion system. Other system architecture factors also affect these metrics. This criterion is divided into subclasses *Timeliness* and *Throughput*.

4) *Output Criteria* are usually related to the system's results and its ability to communicate it to its users in a clear fashion. It has the following subclasses:

  - *Quality* is a group of criteria meant to assess the informational quality of the system's output. It includes *Accuracy* and *Precision* as subclasses. It is common to see in the literature the same concepts with different names. For example, accuracy sometimes is used as a synonym of precision; sometimes the terms are used with different meanings. Indeed, accuracy and precision can be inversely related. As one makes the granularity coarser, one can expect that the system will have a better accuracy. Precision can also be used to put bounds on the certainty of the reported result.[4]

  - *Interpretation* refers to the degree to which the uncertainty representation and reasoning can be used to guide assessment, to understand the conclusions of the system and use them as a basis for action, and to support the rules for combining and updating measures (adapted from [30]).

The above concepts are being explored within the ETURWG, which is making use of this ontology to support the development of uncertainty evaluation criteria for a set of information fusion use cases. The interested reader should refer to the group's website[5] for more specific details. Note that the URREF ontology is not supposed to be a definitive reference for evaluation criteria, but simply an established baseline that is coherent and sufficient for its purposes. This approach privileges the pragmatism of having a good solution against having an "ideal" but usually unattainable solution. For instance, a definitive reference would involve having universally accepted definitions and usage for terms such as "Precision." This is clearly infeasible. The approach also takes into consideration that more important than naming a concept is to ensure that it is represented clearly and distinctly within the ontology so to ensure the consistency of the latter.

To assure utility and acceptability of the URREF ontology, most of its concepts have been drawn from seminal work in related areas such as uncertainty representation [28], evidential reasoning [29], performance evaluation [30, 31]). The ontology has built on the URW3 uncertainty ontology [25]. Also, the structure and viewpoint adopted in the ontology development have been tuned to addressing the uncertainty evaluation problem and its associated perspective (e.g. how information is handled within a fusion system). Finally, it is a goal of the

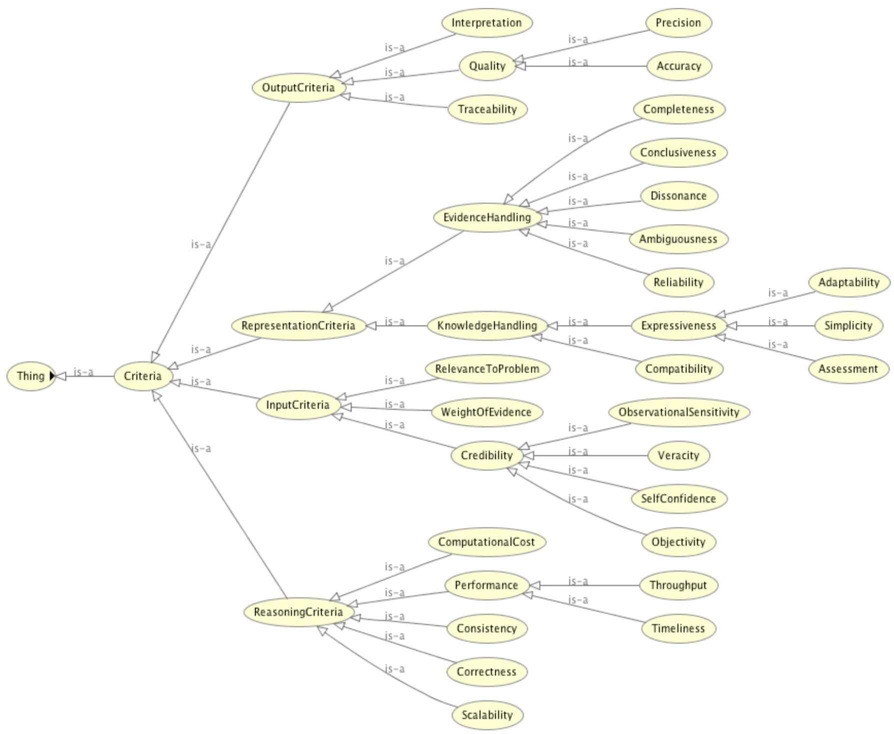

Figure 7 – URREF Ontology: Criteria Class.

[4] http://en.wikipedia.org/wiki/Accuracy_and_precision

[5] http://eturwg.c4i.gmu.edu. Registration is required.

ETURWG to have standardized uncertainty ontology upon which to build for community acceptance.

VI. Discussion

Although not a new research topic, the evaluation of IF systems presents various challenges due to the complexity of fusion systems and the sheer number of variables influencing their performance. [32] In LLF systems, the impact of uncertainty representation is well understood, and generally quantifiable. However, at higher levels of IF the approach chosen for representing uncertainty has an overall impact on system performance that is hard to quantify or even to assess from a qualitative viewpoint. This issue was recognized by the Fusion community when creating the ETURWG, with the main goal of providing an unbiased framework for evaluating the impact of uncertainty in IF systems. From the beginning, it became clear that the various approaches and technical considerations demand a common understanding that is only achievable by a formal specification of the semantics involved. As a result, the group developed the URREF ontology presented in this paper. The ontology is now being employed to support the development of more specific requirements to evaluate a set of use cases and associated data sets designed by the group and accessible through our webpage [http://eturwg.c4i.gmu.edu]. Although it is clear that the URREF ontology is not a definitive reference for these types of activities, its use has proven to be a major asset in developing a common framework. We invite all interested practitioners, developers, and researchers to particapte in the ETURWG.

Acknowledgment

The authors recognize the contributions of the members of the International Society of Information Fusion's ETURWG in formulating this ontology, which was discussed in several of its meetings. Special thanks go to Jürgen Ziegler, Gavin Powell, Amandine Bellenger, Brian Ulicny, Valentina Dragos, Mark Locher, Dafni Stapouli, Kellyn Rein, and Matthew Roberts for their insightful comments during our meetings and at the ETURWG forums.

References

[1] A. N. Steinberg, C. L. Bowman, and F. E. White, "Revisions to the JDL model," *Joint NATO/IRIS Conf.*, 1998.

[2] E. P. Blasch and P. Hanselman, "Information Fusion for Information Superiority," *IEEE Nat. Aerospace and Electronics Conference,* 2000.

[3] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz, & F. White, "Revisiting the JDL Data Fusion Model II", *Int. Conference on Information Fusion*, 2004.

[4] D. L. Hall and S. A. McMullen, *Mathematical Techniques in Multisensor Data Fusion*, Artech, 2004.

[5] E. Blasch, I. Kadar, J. Salerno, M. M. Kokar, S. Das, G. M. Powell, D. D. Corkill, and E. H. Ruspini, "Issues and Challenges in Situation Assessment (Level 2 Fusion)," *J. of Advances in Information Fusion,* Vol. 1, No. 2, pp. 122 - 139, Dec. 2006.

[6] E. Blasch, M. Pribilski, B. Roscoe, *et. al.*, "Fusion Metrics for Dynamic Situation Analysis," *Proc SPIE,* Vol. 5429, Aug 2004.

[7] P. Djiknavorian, D. Grenier and P. Valin, "Analysis of information fusion combining rules under the DSm theory using ESM inputs," *Int. Conf. on Information Fusion*, 2007.

[8] E. Waltz and J. Llinas, "System Modeling and Performance Evaluation," Ch 11 in *Multisensor Data Fusion Systems*, Artech House 1990.

[9] J. Llinas, "Assessing the Performance of Multisensor Fusion Processes," Ch 20 in *Handbook of Multisensor Data Fusion*, (Eds.) D. Hall and J. Llinas, CRC Press, 2001.

[10] A. Theil, L. J. H.M. Kester, E. Bosse, "On measures of performance to assess sensor fusion effectiveness," *Int. Conf, on Info. Fusion*, 2000.

[11] J. Van Laere, "Challenges for IF Performance Evaluation in Practice," *Int. Conference on Information Fusion*, 2009.

[12] E. Blasch, "User refinement in Information Fusion", Chapter 19 in *Handbook of Multisensor Data Fusion 2nd Ed,* (Eds.) D. Hall, and J. Llinas, CRC Press, 2008.

[13] K. J. Vicente, *Cognitive Work Analysis*, *Toward Safe, Productive and Healthy Computer-based Work*, Lawrence Erlbaum, 1999.

[14] E. Blasch, R. Breton, P. Valin, and E. Bosse, "User Information Fusion Decision Making Analysis with the C-OODA Model," *Int. Conf. on Info Fusion*, 2011.

[15] A-L Jousselme, P Maupin, and E. Bosse, "Formalization of Uncertainty in Situation Analysis," *DSTO conference*, 2003.

[16] A. Auger and J. Roy, "Expression of Uncertainty in Linguistic Data," *Int. Conference on Information Fusion*, 2008.

[17] D. A. Lambert. "Situations for situation awareness," In. *Int. Conference on Information Fusion*, 2001.

[18] E. Blasch, P. Valin, E. Bossé, "Measures of Effectiveness for High-Level Fusion," *Int. Conference on Information Fusion*, 2010.

[19] M. E. Johnson and K. C. Chang, "Quality of Information for Data Fusion in Net Centric Publish and Subscribe Architectures," *Int. Conference on Information Fusion*, 2005.

[20] B. Yu and K. Sycara, "Learning the Quality of Sensor Data in Distributed Data Fusion," *Int. Conference on Information Fusion*, 2006.

[21] G. Rogova & E. Bosse, "Information Quality effects on Information Fusion," *DRDC Valcartier, Tech Rept. 2005-270*, May 2008.

[22] W. Perry, D. Signori, and J. Boon, "Exploring the Information Superiority: A Methodology for Measuring the Quality of Information and its Impact on Shared Awareness," *RAND Corporation*, 2004.

[23] M. A. Hossain, P. K. Atrey, and A. El Saddik, "Modeling Quality of Information in Multi-Sensor Surveillance Systems," *IEEE Int. Conf. on Data Engineering Workshop*, 2007.

[24] C. Bisdikian and L. M. Kaplan, M. B. Srivastava, D. J. Thornley, D. Verma, and R. I. Young, "Building principles for a quality of information specification for sensor fusion," *Int. Conf. on Information Fusion*, 2009.

[25] K.B. Laskey, K.J. Laskey, P.C.G. Costa, M. Kokar, T. Martin, and T. Lukasiewicz (eds.) "Uncertainty Reasoning for the World Wide Web: W3C Incubator Group Report." World Wide Web Consortium. Available at http://www.w3.org/2005/Incubator/urw3/XGR-urw3/.

[26] P.C.G. Costa, KC Chang, K.B. Laskey, T. Levitt, and W. Sun, "High-Level Fusion: Issues in Developing a Formal Theory", *Int. Conf. on Information Fusion*, 2010.

[27] P.C.G. Costa, R.N. Carvalho, K.B. Laskey, and C.Y.Park, "Evaluating Uncertainty Representation and Reasoning in HLF systems", *Int. Conference on Information Fusion*, 2011.

[28] A-L Jousselme, P. Maupin, E. Bossé, "Quantitative Approaches" Ch. 8 in *Concepts, Models and Tools for Information Fusion*, Artech House, 2006.

[29] D. Schum, "The Evidencial Foundations of Probabilistic Reasoning," Northwestern University Press, 1994.

[30] P. Walley, "Measures of Uncertainty In Expert Systems," Artificial Intelligence, 83(1), May 1996, pp. 1-58.

[31] K. C. Chang, Z. Tian, S. Mori, and C-Y. Chong, "MAP Track Fusion Performance Evaluation," *Int. Conf. on Information Fusion*, 2002.

[32] P. Hanselman, C. Lawrence, E. Fortunano, B. Tenney, and E. Blasch, "Dynamic Tactical Targeting," *Proc. of SPIE,* Vol. 5441, 2004.

[33] E. P. Blasch, E. Bosse, and D. A. Lambert, *High-Level Information Fusion Management and Systems Design*, Artech House, April 2012.