Subjects under evaluation with the URREF ontology

A.-L. Jousselme*, P. de Villiers†‡, G. Pavlin§, P. Costa¶, K. Laskey¶, V. Dragos‖, E. Blasch**

*NATO-STO CMRE, La Spezia, IT, Anne-Laure.Jousselme@cmre.nato.int
†CSIR, Pretoria, South Africa, jdvilliers1@csir.co.za
‡University of Pretoria, Pretoria, South Africa, pieter.devilliers@up.ac.za
§D-CIS Lab, Thales Research and Technology, Delft, The Netherlands, gregor.pavlin@d-cis.nl
¶George Mason University, Fairfax, VA, USA, [pcosta,klaskey]@gmu.edu
‖ONERA-The French Aerospace Lab, Palaiseau, France, valentina.dragos@onera.fr
**Air Force Research Lab, Rome, NY, USA, erik.blasch.1@us.af.mil

***Abstract*—The question addressed in this paper is "what" is to be evaluated by the Uncertainty Representation and Reasoning Evaluation Framework (URREF) ontology. We thus identify the elements composing uncertainty representation and reasoning approaches, which constitute various subjects being assessed. We distinguish between *primary* evaluation subjects (Uncertainty Representation and Reasoning components of the fusion algorithm), and *secondary* evaluation subjects (source of information, piece of information, fusion method and mathematical model). This paper proposes a list of source quality criteria to be added to the ontology and establishes formal links between the secondary and primary evaluation subjects. A key contribution of the paper is the update of the definitions of sub-criteria of the Expressiveness criterion together with suggestions for complementary concepts to be included in the ontology (type of scale, type of uncertainty expression). Conclusions are drawn to extend the work in using the expressiveness criterion for information fusion analysis.**

Keywords: Assessment; Evaluation; Quality; Uncertainty representation; Reasoning; Criteria.

I. Introduction

Defining criteria to assess information fusion approaches has been the topic of the Evaluation of Techniques for Uncertainty Representation Working Group (ETURWG)[1] since 2011. Through a series of interrelated evaluation criteria and uncertainty-related concepts, the aim of the Uncertainty Representation and Reasoning Evaluation Framework (URREF) ontology would be to support the global and coherent assessment of fusion algorithms as far as uncertainty representation and reasoning is concerned (*e.g.*, [1], [2], [3], [4]).

The most interesting aspect of the development of this ontology addresses specifically the uncertainty representation procedure, where other approaches rather consider the algorithms as complete systems, and evaluate them generally through their outputs only (*e.g.*, [5]). The idea of the URREF is to "open the box" of the fusion algorithm and identify the different elements of uncertainty handling to further assess these components.

In this paper, the elements to be assessed and compared within Uncertainty Representation and Reasoning (URR) approaches will be referred in the following as *evaluation subjects*. Due to the complex and multiple connections between elements it seems difficult to separate the uncertainty representation (*e.g.*, an instantiated probability distribution) and its associated reasoning scheme (*e.g.*, Bayes' rule), from its underlying mathematical framework (*e.g.*, probability theory), from an underlying semantic representation (*e.g.*, possible worlds, Ontology Web Language (OWL) ontology), from the (fusion) algorithm processing the information (*e.g.*, a classifier), from a higher-level (fusion) system possibly including some human interaction. Figure 1 illustrates some system components that could be possibly assessed and which interact to built a full fusion system. As far as the URREF is concerned, the elements of an Uncertainty Representation and Reasoning scheme are the main evaluation subjects. Empirical evaluation techniques require embedding some uncertainty representation and associated reasoning scheme within a fusion algorithm, which serves then as an enabler and becomes often the subject of evaluation itself. Inherently, it is not a trivial problem to isolate the uncertainty representation from its reasoning scheme or from the algorithm which may implement other contributing aspects, albeit minor.

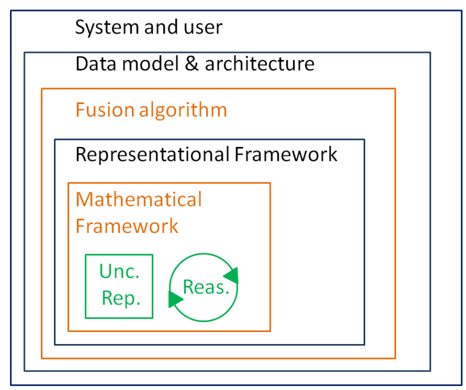

Fig. 1. A not exact but convenient hierarchy of system components as possible evaluation subjects.

For each subject, a series of criteria of interest should be then defined as proposed in the URREF ontology [1]. But it may also happen that the same criterion applies to different subjects (*e.g.*, *accuracy* can be a property of piece of

[1]http://eturwg.c4i.gmu.edu

information and of a source).

In this paper, we report some progress about defining the subjects to be assessed through the different criteria of URREF ontology. One of the aims of this paper is to further clarify the links between subjects and associated assessment criteria to ensure a consistency between the definitions of criteria for different evaluation subjects.

In Section II, an elementary fusion process is presented to single out the primary evaluation subjects in the Uncertainty Representation and Reasoning methodology. Section III defines URREF evaluation subjects and proposes a taxonomy, distinguishing between primary and secondary subjects, and relating them to the associated criteria. In Section IV, the links between the criteria and the evaluation subjects are discussed. We clarify the impact of the secondary evaluation subjects that are the source and the piece of information on the uncertainty representation assessment, with a focus on the *Expressiveness* criterion. Section VI provides conclusions and contributions of this work in the context of the ETUR group, and reflected in the URREF ontology.

II. Elementary fusion process

The **fusion algorithm** is further detailed here through a simple procedure that is generic enough to essentially single-out the main elementary constructs of uncertainty representation and reasoning that are the primary URREF evaluation subjects. The fusion algorithm may be very complex, involving possibly several uncertainty representations, mathematical frameworks, combination or inference rules, etc. We describe here a simple, albeit quite general, fusion process aimed at clarifying the information flow, and which could be considered as an "atomic" process.

Let us consider the following 10 elements of an elementary fusion process, as shown in Figure 2:

① s is the source of information;

② ϕ is a piece of information[2] provided by s. It can be as simple as a measurement (on the scale of real numbers) but could be a fact (*i.e.*, an observation, known to be true), but could also be a natural language statement, an uncertain statement already modeled into a given mathematical formalism (*i.e.*, a probability distribution);

③ h is the uncertainty representation process by which ϕ is transformed into a dedicated mathematical representation with uncertainty. Prior information on source's quality (*e.g.*, reliability), contextual information or comparison with other pieces of information may impact h;

④ $h(\phi)$ is the instantiated mathematical representation as built by h and expresses either the self-assessment of the source or an external assessment by the user based on prior source's quality knowledge;

⑤ ρ is the information combination process which modifies $h(\phi)$ into another $h_{\oplus}(\phi)$ within the same mathematical model. At this point, a series of pieces of information from other sources $\{h(\phi)\}_{i=1,\dots,N}$ are combined;

⑥ $h_{\oplus}(\phi)$ is the aggregated piece of information built from $h(\phi)$ and other related information;

⑦ l is the decision process which transforms $h_{\oplus}(\phi)$ to provide the decision, *i.e.*, the output information ϕ';

⑧ ϕ' is the information output, to be sent to another system. It can be a formal representation (such as a probability distribution), or a single measurement estimated after the decision process (soft versus hard decision). It can thus contain or not contain some uncertainty;

⑨ the reasoning process is $l \circ \rho$;

⑩ the Atomic Decision Procedure (ADP) is $l \circ \rho \circ h$.

As further detailed in [6], the method can distinguish between:

a) information processors (providing pieces of information): Elements ①, ③, ⑤, ⑦, ⑨;
b) the pieces of information provided: Elements ②, ④, ⑥, ⑧;
c) the couples (process; output information): (①,②); (③,④); (⑤,⑥); (⑦,⑧); (⑨,⑧); (⑩,⑧)=(①,②)

From an algorithmic standpoint, we may want to assess each of the 10 items above. However, based on the following observations some simplifications arise:

- Each information processor can be assessed through the information it provides, so it is natural to consider the couples (processor; output information);
- The couple (①,②), (source; input information), is defined as a secondary evaluation subject and its previous characterization should be considered in the assessment of the primary subjects;
- In some cases, the reasoning process ($l \circ \rho$) may be considered as a whole, without separating the combination from the decision.

Thus the important couples are:

- (③,④) - the uncertainty representation process (URP) together with its output;
- (⑨,⑧) - the reasoning process together with its output;
- (⑩,⑧) - the couple (representation, reasoning) together with its output.

which will be evaluated within the URREF.

III. URREF evaluation subjects

Following the previous detailed description of an elementary fusion algorithm, the section defines the different evaluation subjects and identifies the corresponding criteria of the URREF ontology.

A. Introducing the evaluation subjects

Definition 1 (Evaluation subject) *An Evaluation Subject is an item which can be assessed according to the criteria defined in the URREF ontology.*

This distinction helps better specify and communicate the goal of the URREF ontology but also better focus the effort on the primary subjects. Figure 3 lists the evaluation subjects.

[2]The term "piece of information" used in this paper is used in its more general meaning covering other notions such as evidence, knowledge and/or data

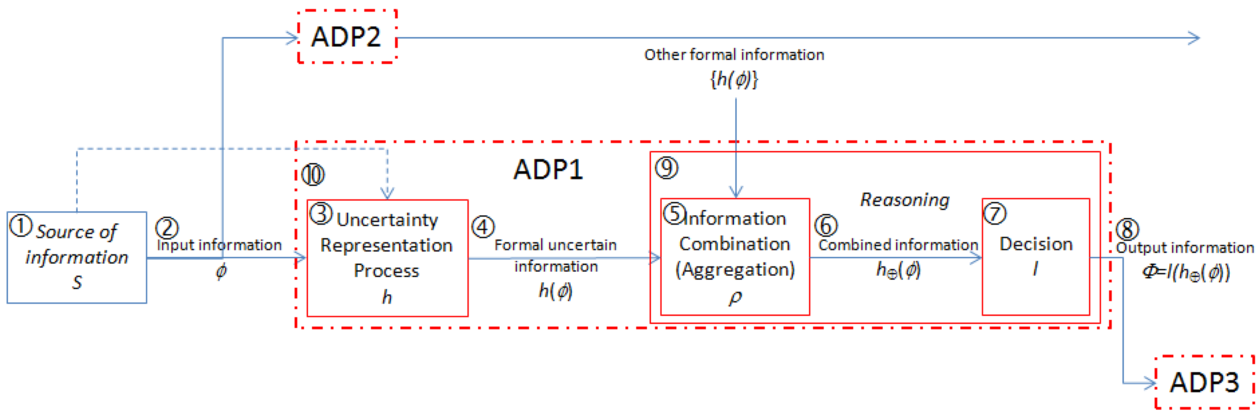

Fig. 2. Basic information flow and evaluation subjects.

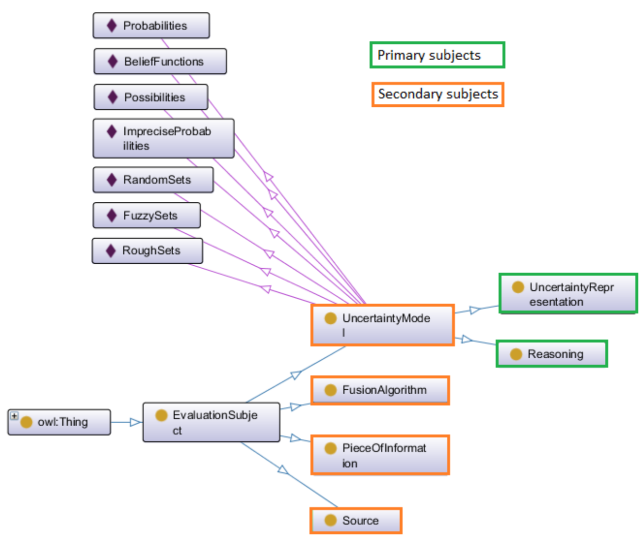

Fig. 3. Proposal of evaluation subjects taxonomy.

Elements with yellow circles are classes. The primary subject classes are drawn with a green box around, while the secondary object classes are drawn with an orange box around. Elements tagged with a purple diamonds are individuals of the class *UncertaintyModel*. This list is non-exclusive and unstructured, as in particular does not reflect the possible relationships between the different models.

We identify the *primary evaluation subjects* of the URREF as being:

- the **uncertainty representation**, which is either instantiated or theoretical: a particular probability distribution or probabilities in general; it may include instantiated uncertainty representations of processes in the real world and how those processes are observed, i.e. uncertainty supports that characterise those processes and observations;
- the associated **reasoning** (or calculus) that comprises the combination, conditioning, updating, inference, decision, transformation rules. The calculus may be assessed while instantiated within an algorithm or theoretically, regardless any application or algorithm focusing on the semantics for instance (*e.g.*, Bayes' rule in general).

Secondary evaluation subjects of the URREF consider other elements which support the assessment of the primary subjects, but which are not the main concern of the URREF ontology:

- the **uncertainty model** (or mathematical framework for uncertainty representation and reasoning), such as probabilities, fuzzy sets, belief functions. It can be assessed either theoretically, based on axioms, properties and original semantics. But also through the output provided by a specific **fusion algorithm** implementing this model;
- the **fusion algorithm**, making use of *instantiated uncertainty representations* as represented by **pieces of information** ϕ, built according to a specific method or the uncertainty representation process h and associated calculus $l \circ \rho$;
- the **source** of information which provides the different pieces of information and which quality may impact the whole fusion process. It is expected that an uncertainty representation will be able to properly capture and handle the information about the source quality;
- the **pieces of information** input, processed and output throughout the algorithm. Input and output pieces of information are only two special cases but others can be considered such as internal steps for instance the aggregated information. The information assessment is unavoidable if one wants to assess some uncertainty representation and reasoning. Hence, as a secondary evaluation subject, the development of an ontology for information quality criteria is not the main purpose of the URREF, and we rather rely on existing works such as [7]. However, when applying these criteria to the specific use of URR assessment some contribution occurred, that will be reflected in the URREF ontology.

The fusion algorithm may be assessed either as a whole (*e.g.*, absolute), assessing only the output, or through the different elements composing it (*e.g.*, relative), that are the instantiated uncertainty representation (process and output information),

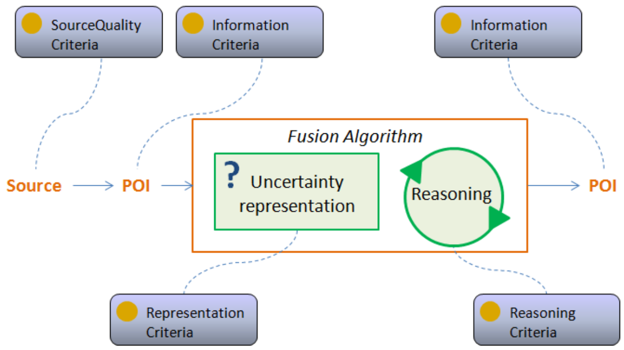

Fig. 4. Evaluation subjects and associated criteria.

and instantiated calculus process (process and output information). Equivalently, the uncertainty model (understood as the mathematical framework for uncertainty representation and reasoning) may be assessed considering the theoretical uncertainty representation (*i.e.*, function of uncertainty such as the probability or belief function) on the one hand or/and the theoretical calculus apparatus (*i.e.*, the set of reasoning tools available to this framework) on the other hand.

For each evaluation subject, there should be a corresponding set of evaluation criteria within the ontology, as illustrated in Figure 4. The quality of the source is assessed by *SourceQualityCriteria*, the pieces of information (POI) provided is assessment by *InformationCriteria*, the uncertainty representation part of the fusion algorithm is assessed through *UncertaintyRepresentationCriteria* and the reasoning part is assessed through *ReasoningCriteria*. In the current version of the URREF ontology there is no such class *SourceQualityCriteria*, while the corresponding class of *InformationCriteria* is named *DataCriteria*.

B. Source quality criteria

Criteria about the source of information are necessary to characterise the information at the input and understand how they impact the representation of uncertainty. The use of these criteria is rather informative than "judgmental". We assume that these initial assessments are known prior to processing the information and the question is if and how the algorithm, and especially the uncertainty representation and reasoning scheme, is able to handle them. As such, the source is a secondary evaluation subject and its impact on the other subjects and their assessment will be detailed in Section IV.

C. Information criteria

Pieces of information (POIs) appear at different places within the process and include in particular, input data or measurement or declaration before any modelisation of uncertainty (*i.e.*, the input information or dataset), the instantiated uncertainty representation (after uncertainty has be modeled), the aggregated information (after the combination or inference process), the output information as presented to the user. Each of these should be characterised according to the same subset of criteria although the expectations may differ. For instance, it is not expected that the input information be precise, nor true. But it would be expected at the output. Also, comparing pieces of information at several steps of the process will provide assessment of relevance (if one has an impact on the other) for instance. Therefore, the same set of evaluation criteria should be used to assess input information, the uncertain information, the combined information, and the output information. Only the values (and the user's expectations) may change, not the criteria themselves. For the input information, the assessment becomes a characterisation since it cannot be changed, while for the pieces of information in the system, the assessment criteria can be turned to optimisation criteria.

D. Uncertainty representation criteria

The Uncertainty Representation criteria are aimed at assessing the primary subject of evaluation within the URREF. Unsurprisingly, expressiveness is the main one. Indeed, before any instantiation of an uncertainty representation, we are interested in knowing the expressive power provided by its mathematical framework in terms of uncertainty representation. This is a *prior* (theoretical) assessment, independent of the application and mainly relies on analyses of (1) the axiomatic constraints of the framework and (2) the current literature about the development of the approaches and tools to support the representation of concepts of interest as identified within the expressiveness list of criteria. The instantiated uncertainty representation should also be assessed along with the subset of criteria. An instantiated uncertainty representation is a piece of information and as such, should be assessed using the criteria of pieces of information described above.

E. Reasoning criteria

This subset of criteria is so far not very detailed within the URREF ontology of criteria since these followings are interrelated:

a) the calculus apparatus of a mathematical framework for uncertainty reasoning, *i.e.*, the set of reasoning tools available to this mathematical framework,
b) a particular instantiation of use of one of these rules, and
c) the algorithm making use of this apparatus.

For a more detailed analysis, these three subjects should be clearly distinguished, although the same criterion may be applicable and relevant to all of them. For instance, if we consider the *consistency* criterion:

a) a particular rule of combination could be assessed according to its theoretical ability to provide consistent results,
b) a specific use of the rule necessary relying on other elements such as the universe of discourse selected, the type of uncertainty function to be combine, etc, could be assessed according to the consistency criterion, and
c) an algorithm embedding the rule with the uncertainty function and associated universe of discourse within a higher reasoning scheme (*e.g.*, nearest neighbors approach, back propagation, etc) may also be assessed according to the same criterion of consistency.

IV. Links between criteria of the evaluation subjects

This section further details the crucial notion of *source of information* and its fundamental relationships with the other evaluation subjects. First, the relationship between information quality and source quality is discussed followed by the distinction between the source quality and the fusion algorithm (or fusion system) quality.

A. Source quality versus information quality

A *source of information* is "anything that might inform a person about something or provide knowledge about it"[3]. Thus by definition, source of information provides information and thus should be assessed primarily according to the information it outputs. Consequently, information quality and source quality are by nature strongly connected. The characterisation of information quality in identifying and defining Information Quality Dimensions (IQD) has been widely addressed in the literature (*e.g.*, [7], [8], [9], [10]). In [6], we proposed a list of some source quality criteria which should apply to any type of source, being it a sensor, algorithm or a human. The corresponding taxonomy is displayed in Figure 5. This list of source quality criteria is inspired by the ontology produced by the W3C Semantic Sensor Network Incubator group (the SSN-XG) describing sensors in terms of capabilities, measurement processes, observations and deployments [11], and augmented by quality criteria for human sources associated with the military intelligence (*e.g.*, STANAG 2511 [12]) and the field of justice. As prescribed by [12], the source of information and the information produced by the source should be assessed independently, acknowledging that a "good" source can provide "bad" information at a specific instant in time. The dimensions are named *reliability* for the source quality and *credibility* for the information quality.

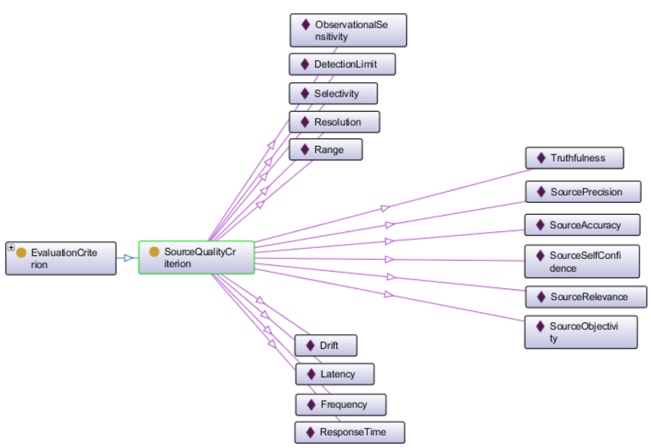

Fig. 5. Source quality criteria.

Source quality (SQ) differs from information quality (IQ) in the sense that SQ is more perennial, the quality of a source being assessed on its *ability* to provide information based on past experiences, experiments, while IQ can be assessed instantaneously. Possibly, any IQ dimension can be translated into a SQ dimension: s is credible if it provides credible information, s is relevant if it provides relevant information, s is objective if it provides objective information, etc.

[3] http://en.wikipedia.org/wiki/Information_source

B. Assessing source quality

The quality of a source (possibly along different dimensions) is a numerical value representing "how good" the source is expected to perform. The assessment is based on past experiments with the source, where its output has been compared to some ground truth information and validated. This is true for sensors characterised for instance by Receiver Operating Characteristics (ROC) curves, for classifiers or detectors characterised by confusion matrices but also for human intelligence sources who are assigned Reliability rates. This phase of characterisation of the source is illustrated in Figure 6(a) where a source (here, a camera associated with a human analyst) is assessed through a series of tests and experiments. For instance, the *precision* is computed and the *accuracy* is validated with some ground truth. The *truthfulness* can be also tested in the case of human sources as a possible tendency to lie.

The quality assessment of the information supplied by a given source over some period of time (and through tests and validation steps) is then *translated* into the source's quality.

In the URREF ontology *self-confidence* is a criterion intended to shed light on the subtle transposition of source's qualities into attributes of provided items. Self-confidence captures some degree of confidence of items as evaluated by the source itself and it is of particular interest for human sources, as most of the time they report on facts by adding appreciations in the form of beliefs, suppositions and plausibility estimations. Estimation of self-confidence assigns to this criterion a numerical value, according to how strong the author stands behind assertions in the statement and can be done on the basis of lexical clues he included when uttering. More generally, markers of author's commitment are in the form of hedges, modal verbs and forms of passive/active language. A hedge is a mitigating word that modifies the commitment to the truth of propositions, *i.e.*, *certainly, possibly*. Their impact can be increased by a booster *(highly likely)* or decreased by a downtoner *(rather certain)*. Modal verbs indicate if something is plausible, possible, or certain *(John could be a terrorist, you might be wrong)*.

Assigning numerical values to lexical expressions is not an intuitive task, and Rein shows that there are no universal values to be associated in a unique manner to hedges or other uncertainty markers [13]. As Rein argues further, it is rather possible to order those expressions and use this universal ordering as a more robust way to compare combinations of uncertainty expression, and thus highlight different levels of uncertainty in natural language statements. The *self-confidence* of the source can also be assessed and compared with some ground truth. Indeed, after validating the source's statement,

(a) During a characterisation (or learning) phase and through a series of tests, experiments, observations, the information provided by the source is assessed according to different source quality criteria.

(b) During the operation phase, the source quality (previously assessed) along the different dimensions is used to discount or reinforce the information provided by the source at a specific instant in time.

Fig. 6. Information quality versus source quality.

it can be assessed to be over-confident or under-confident in the statements [14].

C. Using the source quality

After this characterisation phase, the source is tagged with many quality measures (not necessarily numerical values) corresponding to the different dimensions. Referring to the STANAG 2511, a completely reliable source would be tagged with an index of "A". During the operation phase, when the source actually provides information to be processed by the fusion system (algorithm), its quality is used to modify the information provided. For instance, if the source is highly reliable and under-confident, the information may be reinforced. While if it is not reliable and over-confident, the information can be discounted and even discarded (see for instance [15]) for some correction schemes with belief functions). For instance, let us consider precision and accuracy as two standard IQ dimensions: The statistical joint assessment of precision and accuracy of the information provided by a source s during laboratory tests can be translated into a perennial notion of *reliability* of the source, which represents its ability to provide predictable results. The reliability factor can then be used to correct, alter, discount the output information but also as predictive parameters.

D. Source quality versus fusion algorithm quality

Steinberg states that "[s]ource characterisation [. . . is] concerned specifically with the information reporting performance, behavior and pertinent relationships of agents" [16]. Consequently, the source's quality should be assessed according to its *ability* to provide information, and its internal reasoning process should not be characterised in detail.

Moreover, it appears that the notion of "source" is relative and depends on the perspective at hand. Indeed, information circulates within a network of heterogeneous agents, mixing physical sensors, artificial agents (algorithms) and humans users. Some agent maybe a source of information for another agent, who is itself a source for another one. Nodes of this network are either named as "sources" (information producer), or "agents" (information processor and actors), etc. In Figure 6, two levels of processing are represented to illustrate the relative notion of the source: The camera is producing an image, itself analysed by a human analyst who will further report to another decision maker. The quality of the camera (source for the human) is assessed based on the quality of the image output, and the quality of the analyst is assessed based on her/his past statements about images analysis, whether they were of good or poor quality. Consequently, a fusion system (algorithm) as a whole should be assessed similarly to sources of information, and should share the criteria based on the output only.

The recursive nature of the process has been illustrated in Figure 2 and briefly discussed in Section II, where the ADP can be seen as a detailed description of the internal process of a *source of information*. The ADP outputs pieces of information Φ to be further input to and processed by another fusion algorithm (another ADP). Figure 7 is a better illustration clarifying the links between the different criteria classes.

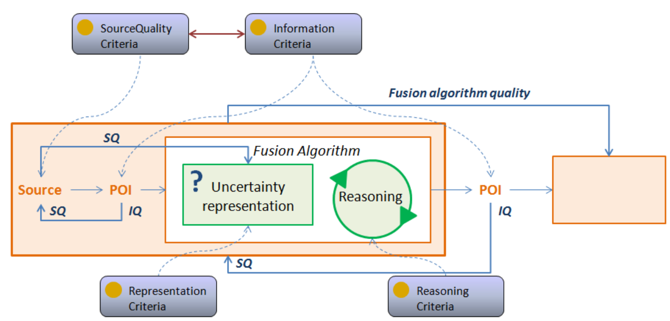

Fig. 7. Recursive process of source, information and algorithm assessment.

The originality of the URREF ontology however is to allow the assessment of the *internal constructs* of a fusion algorithm, and especially the elementary constructs of uncertainty representation and reasoning. In that respect, while the source is assessed only based on the information provided (its output), the fusion system through the URREF ontology will be assessed also internally.

V. Uncertainty representation expressiveness

One important criterion for the assessment of an uncertainty representation is its *Expressiveness*. By that, we would like to assess how many diverse aspects of the information the representation is able to convey. And in particular, its "ability to convey all relevant aspects of a given fusion problem".

In order to render the links between the Expressiveness criteria (assessing the primary evaluation subject of uncertainty representation), we provide below some refinement and update of some definitions of the current ontology. The links include:

- Types of uncertainty judgment or expression (handled by the *Assessment* criterion),
- Types of scales for the variables (handled by the *Outcomes* criterion), and
- Types of uncertainty (handled by the *Inclusiveness* criterion).

A. Assessment

Definition 2 (Assessment) *Ability to capture the types of uncertainty assessments as provided by the Type of Uncertainty Assessment class, needed for a given problem, and to distinguish them from one another.*

This criterion originates from Walley's list of criteria [17], for what he names "measures of uncertainty" which correspond to "Uncertainty models" (Bayesian probabilities, Coherent Lower previsions, Dempster-Shafer, etc) in the URREF ontology. Uncertainty assessment according to Walley refers to "uncertainty judgment" or "uncertainty expression", that he distinguishes from "reasoning procedures". We provide then the following definition for uncertainty expression:

Definition 3 (Uncertainty expression) *Any statement produced by a source expressing some uncertainty. This encompasses the cases of natural language expressions, simple numerical statements such as a confidence value, complete mathematical expressions such as a probability distribution in the case the automatic processing of source is already framed in an Uncertainty Model.*

An *uncertainty expression* could be for instance:

- Natural language statement (probably, maybe, roughly, likely, unsure that, ...);
- Unique numerical statement (0.5);
- Set of numerical values (*e.g.*, [0.2 0.4 0.5], intervals, $> 0.6, \ldots$);
- Instantiation of an Uncertainty Model (a probability distribution, a belief function, etc).

Some uncertainty expression where previously discussed referring to the source's *self-confidence*. Indeed, the source self-confidence is an uncertainty expression that should be captured by the uncertainty representation.

B. Outcomes

The uncertainty representation should be able to capture different kinds of outcomes for uncertain variables, such as Boolean, categorical, ordinal, discrete numerical, continuous. This is described by the *Outcomes* criterion. To make it more general and flexible, we refer to the types of scale rather than listing them all.

Definition 4 (Outcomes) *Ability to represent required kinds of outcomes for uncertain variables as defined by the Types of Scale.*

There is no real consensus on the types of scale but Stevens' is a good reference and is illustrated in Figure 8. If necessary,

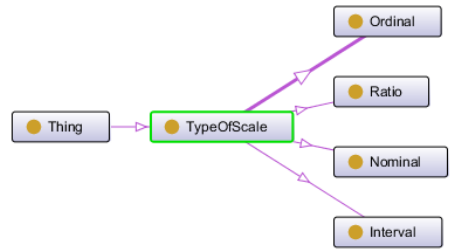

Fig. 8. Stevens' types of scale [18].

we could refer to another list of types of scale and assess the ability of the uncertainty representation to capture the different types of scale.

C. Inclusiveness

There is a need to capture the different types of uncertainty within the URREF.

Definition 5 (Inclusiveness) *Ability to capture different types of uncertainty, among the ones defined by the UncertaintyType item.*

Figure 9 displays the current list of uncertainty types in the URREF ontology. Whatever we mean by "uncertainty type",

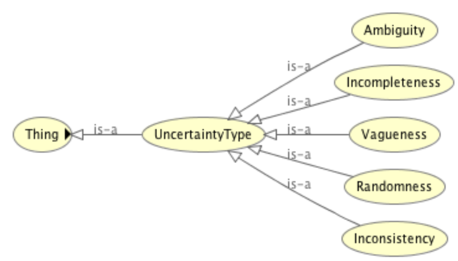

Fig. 9. Some uncertainty types.

this list can be updated, modified, we could refer to other typologies such as Klir and Yuan's [10] or Smets' [8], and assess the inclusiveness criterion.

D. Ability to capture vs ability to distinguish

It seems that two aspects of Expressiveness are interesting to consider: (1) the ability to *capture* the different types of concepts and (2) the ability to *distinguish* between them. For instance, an uncertainty representation process may capture both objective and subjective uncertainty but in the same manner, with no distinction. We could name:

- *Coverage*: the extent to which the uncertainty representation *capture* the different types of concepts";

- *Distinction*: the ability to discriminate between two elements (two types of uncertainty, two types of uncertainty expressions, two types of uncertainty derivation).

E. Toward other criteria

If we define *Expressiveness* as an ability to capture (and distinguish between) characteristics of information, then other expressiveness criteria can possibly be defined in direct relation with the current URREF concepts. In particular, we may want to assess the *ability to capture and make distinction between*:

- different *UncertaintyNature*. For instance, is the uncertainty representation able to capture and distinguish between Aleatory and Epistemic uncertainty?
- different *UncertaintyDerivation*. For instance, is the uncertainty representation able to capture and distinguish between Subjective and Objective uncertainty?
- different *SourceQualityDimensions*. For instance, does the uncertainty representation accounts for source's possible lack of truthfulness? of precision? of observational sensitivity?

VI. Conclusions

This paper is a direct contribution to the development of the URREF ontology. We first presented the evaluation subjects, that are the "what" is to be assessed by the ontology. We distinguished between the *primary* evaluation subjects (Uncertainty Representation and Reasoning components of the fusion algorithm), and the *secondary* evaluation subjects (source of information, information, fusion algorithm and mathematical model), and highlighted them on a simple description of the fusion process and associated information flow.

We proposed a taxonomy of the URREF evaluation subjects distinguishing between primary and secondary subjects, and related them to the associated criteria. The identification of the evaluation subjects revealed the links between for instance, the information quality assessment and the uncertainty representation assessment, between the fusion algorithm and the source assessment or between the information assessment and the source assessment. The paper also highlighted the refinements of the Expressiveness criterion and some of its sub-criteria, by relating them to higher-level concepts of the ontology such as *UncertaintyType*. Other proposed ideas include some missing concepts such as the *TypeOfScale* and the *TypeOfExpression*. The next steps within the ETUR working group will be to validate the ideas put forward in this paper and update the URREF ontology accordingly.

References

[1] P. C. G. Costa, K. Laskey, E. Blasch, and A.-L. Jousselme, "Towards Unbiased Evaluation of Uncertainty Reasoning: The URREF Ontology," in *Proc. of the 15th Int. Conf. on Information Fusion*, Singapore, 2012.

[2] J. P. de Villiers, K. Laskey, A.-L. Jousselme, E. Blasch, A. de Waal, G. Pavlin, and P. Costa, "Uncertainty representation, quantification and evaluation for data and information fusion," in *Proc. of the 18th Int. Conf. on Information Fusion*, Washington, D. C. (USA), July 2015.

[3] J. P. de Villiers, A.-L. Jousselme, A. de Waal, , G. Pavlin, K. Laskey, E. Blasch, and P. Costa, "Uncertainty evaluation of data and information fusion within the context of the decision loop," in *Proc. of the 19th Int. Conf. on Information Fusion*, Heidelberg, Germany, July 2016.

[4] E. Blasch, A. Josang, J. Dezert, P. C. G. Costa, K. Laskey, and A.-L. Jousselme, "URREF self-confidence in information fusion trust," in *Proc. of the International conference on Information Fusion*, Salamanca, Spain, July 2014.

[5] I. Molina, E. Martinez, C. Morillo, J. Velasco, and A. Jara, "Assessment of data fusion algorithms for earth observation change detection processes," *Sensors*, vol. 16, no. 10, 2016.

[6] A.-L. Jousselme, A.-C. Boury-Brisset, B. Debaque, and D. Prévost, "Characterization of hard and soft sources of information: A practical illustration," in *Proceedings of the International Conference of Information Fusion*, Salamanca, Spain, 2014.

[7] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–34, 1996.

[8] P. Smets, "Imperfect information: Imprecision - uncertainty," in *Uncertainty Management in Information Systems. From Needs to Solutions*, A. Motro and P. Smets, Eds. Kluwer Academic Publishers, 1997, pp. 225–254.

[9] M. Smithson, *Ignorance and Uncertainty*. Berlin: Springer-Verlag, 1989.

[10] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice Hall International, 1995.

[11] W3C Incubator Group. (2011, June) Semantic sensor network XG final report. http://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/.

[12] STANAG 2511, *Intelligence Reports (Edition 1)*, North Atlantic Treaty Organization Std., Jan. 2003, NATO UNCLASSIFIED.

[13] K. Rein, "I think it is likely that it might be so ... Exploiting lexical clues for the automatic generation of evidentiality weights for information extracted from english text," PhD thesis, 2016.

[14] V. Dragos and K. Rein, "What's in a message? Exploring dimensions of trust in rereport information," in *Proc. of the 19th interantional conference on Information Fusion*, 2016.

[15] F. Pichon, D. Mercier, E. Lefèvre, and F. Delmotte, "Proposition and learning of some belief function contextual correction mechanisms," *International Journal of Approximate Reasoning*, no. 72, pp. 4–42, 2016.

[16] A. Steinberg, "Agent/source characterization for exploitation of human-generated intelligence," in *Int. conf. on Information Fusion*, Seattle, US, 2009.

[17] P. Walley, "Measures of uncertainty in expert systems," *Artificial Intelligence*, vol. 83, pp. 1–58, 1996.

[18] S. S. Stevens, "On the theory of scale of measurement," *Science*, vol. 103, no. 2684, pp. 677–680, June 1946, new Series.