SCIENCE AND TECHNOLOGY ORGANIZATION
CENTRE FOR MARITIME RESEARCH AND EXPERIMENTATION

Reprint Series **CMRE-PR-2019-036**

Information and source quality ontology in support to maritime situational awareness

Elena Camossi, Anne-Laure Jousselme

May 2019

Originally published in:

Proc. of the 21st International Conference on Information Fusion (FUSION 2018), Cambridge 2018, pp. 696-703, doi: 10.23919/ICIF.2018.8455754

About CMRE

The Centre for Maritime Research and Experimentation (CMRE) is a world-class NATO scientific research and experimentation facility located in La Spezia, Italy.

The CMRE was established by the North Atlantic Council on 1 July 2012 as part of the NATO Science & Technology Organization. The CMRE and its predecessors have served NATO for over 50 years as the SACLANT Anti-Submarine Warfare Centre, SACLANT Undersea Research Centre, NATO Undersea Research Centre (NURC) and now as part of the Science & Technology Organization.

CMRE conducts state-of-the-art scientific research and experimentation ranging from concept development to prototype demonstration in an operational environment and has produced leaders in ocean science, modelling and simulation, acoustics and other disciplines, as well as producing critical results and understanding that have been built into the operational concepts of NATO and the nations.

CMRE conducts hands-on scientific and engineering research for the direct benefit of its NATO Customers. It operates two research vessels that enable science and technology solutions to be explored and exploited at sea. The largest of these vessels, the NRV Alliance, is a global class vessel that is acoustically extremely quiet.

CMRE is a leading example of enabling nations to work more effectively and efficiently together by prioritizing national needs, focusing on research and technology challenges, both in and out of the maritime environment, through the collective Power of its world-class scientists, engineers, and specialized laboratories in collaboration with the many partners in and out of the scientific domain.

NOTE: The CMRE Reprint series reprints papers and articles published by CMRE authors in the open literature as an effort to widely disseminate CMRE products. Users are encouraged to cite the original article where possible.

Information and Source Quality Ontology in Support to Maritime Situational Awareness

Elena Camossi and Anne-Laure Jousselme
NATO STO Centre for Maritime Research and Experimentation
La Spezia, Italy
{elena.camossi,anne-laure.jousselme}@cmre.nato.int

***Abstract*—To support situation awareness, the benefit of using a variety of sources is undeniable although it brings additional challenges related to possible conflicting information, heterogeneity in data formats, semantics, uncertainty types, and information quality. Information and source quality are intertwined concepts which assessments connect with the evaluation of uncertainty handling in information fusion solutions. While the Uncertainty Representation Reasoning Evaluation Framework (URREF) ontology focuses on assessment criteria, peripheral concepts still play a critical role. In this paper, we propose an Information and Source Quality (ISQ) ontology formalising the relationships between information-related concepts, and discuss information interpretation in support of Maritime Situation Awareness. Specifically, this paper links the concepts of *Information Source*, *Dataset* and *Piece of Information*, and connects them to the corresponding quality concepts. Such concepts link to the upper level concepts of the URREF ontology *Source* (of information) and data *Quality*. The ontology further expands to the uncertainty modelling and the algorithm design. We conclude on future work and identify future avenues, especially the extension to the formalisation of the evaluation process.**

***Index Terms*—Information sources, Information quality, Uncertainty, Maritime Situational Awareness.**

I. Introduction

Maritime Surveillance Systems are designed to support of Maritime Security operators during their daily tasks, ensuring safety and security of maritime navigation and transport, sustainable fisheries and exploitation of ocean resources (e.g., [1]). These systems continuously collect, fuse, elaborate and visually represent information from heterogeneous sensors networks. Operators interact with the system getting understanding of the situation dynamics, requesting further information as needed and taking decisions according to pre-planned sequences of actions.

This information can be semantically augmented through correlation with contextual data, specific databases or registries, and integration with geographical layers, for instance from map services such as the OGC WMS (Web Map Service [1]), GIS data, or environmental data (either *in situ* or models). The information is directly elaborated by the system, contingently to the operational task and upon the guidance of the analyst, whose mission may be triggered by some intelligence reports complemented with other evidence about the situation at hand.

[1]http://www.opengeospatial.org/standards/wms

Last decades technological developments not only increased the number of available sources of information (e.g., new sensors, public databases, social media, open sources, digitalised documents, Internet of Things) but also the variety of features and data formats [2]. The use of multiple sources of information, can indeed improve situation awareness by either complementing or confirming the partial views of independent sources, but nevertheless increases uncertainty. Indeed, different sources may provide conflicting information, for instance because of faulty or intentionally manipulated sensors [2], making uncertainty handling and communication critical issues in information fusion systems.

At the maritime institutional level, standardisation and harmonisation initiatives are ongoing to cope with data and systems interoperability and information exchange, including the EUropean test bed for the maritime Common Information Sharing Environment (EU-CISE) [3] and the Maritime Information Sharing Environment (MISE)[2]. These environments rely on common vocabularies and data models providing specification guidances for efficient information sharing within Maritime Authorities.

Most of standards and data models include a series of concepts related to source and information quality with associated rating scales, acknowledging the importance of communicating this meta-information together with the core information. *Source reliability*, *information credibility*, *confidence*, *source type*, etc, are so many dimensions that influence the global uncertainty of the fusion systems and that should be considered by mathematical uncertainty representation and reasoning schemes.

The URREF (Uncertainty Representation and Reasoning Evaluation Framework) ontology was proposed by Costa *et al.* [4] to identify and manipulate criteria for assessing not the fusion algorithm as a whole but rather its specific components of uncertainty representation and reasoning. The URREF describes the multifaceted aspects of uncertainty in information fusion, and proposes reference criteria for assessing uncertainty handling in support of Decision Support Systems [5], [6]. The formalisation of the URREF ontology may be used and further expanded to characterise the uncertainty representation of the different aspects involved in knowledge

[2]https://www.niem.gov/about-niem/success-stories/maritime-information-sharing-environment

driven systems.

In this paper, we define the links between the URREF concepts of *Source* and *DataCriteria* and propose an ontology of information and source quality formalising the main entities that build up information quality in information systems. The discussion is illustrated in the maritime surveillance context where situational awareness needs to be built.

In Section II, we first briefly review some related work on ontologies for maritime situation awareness and summarise recent work on the Uncertainty Representation and Reasoning Evaluation Framework (URREF) focused on the relationship between sources of information and associated quality. Section III presents the information and source quality ontology, the main contribution of this paper. In Section IV, we refine the relative notion of a source of information and describe how a software tool (implementing a fusion algorithm) can itself be a source of information. Section V finally concludes this paper opening on open issues and future work.

II. Related Work

A. Maritime Surveillance

Maritime Surveillance Systems process a variety of information. The definition of *Information Source*s encompasses anything with the ability of generating relevant information for the specific operational task at hand, disregarding the form and format. The quality of information sources drives the selection of the relevant sources among those available in order to conduct the mission.

Examples of information sources for maritime Intelligence Surveillance Reconnaissance (ISR) are T-AIS (Terrestrial-Automatic Identification System) receivers/base stations, High Frequency (HF)-Radars, Long Range Tracking and Identification (LRIT) Systems, CTD (Conductivity, Temperature, "Depth") sensors, sea state models, underwater hydrophones, tracking software, human operators or analysts.

The information may be structured (e.g., vessel registry), semi-structured (e.g., AIS messages, Meteorological and Oceanographic (METOC) observations), or unstructured (e.g., a warning phone or radio conversation between a coast guard official and a vessel captain possibly recorded in a document). Structured, and in some cases semi-structured, information can be split into smaller information units which may be interpreted separately and which syntax is described in standards and specification documents (e.g., the National Marine Electronics Association (NMEA) format[3] and International Telecommunication Union (ITU)[4] recommendation ITUR M.13714 for AIS). This is the case of some fields of AIS messages (e.g., IMO number, Destination, and Estimated Time of Arrival (ETA), in AIS message number 5), of variables in oceanographic datasets (e.g., surface temperature), of the columns in a database table (vessel name, in the European Fleet Registry database). Each of these units has a value (or sometime multiple values) associated, which semantics is encoded in the specification.

[3]www.nmea.org

[4]www.itu.int

The ontology we propose in Section III aims at capturing the different aspects of source and information quality as exemplified just here for Maritime Situation Awareness.

B. Ontology for Maritime Situational Awareness

Despite being an important and challenging operational area and regardless of the growing interest received by the research community in the last years, an agreed definition of Maritime Situational Awareness (MSA) had long been lacking. However, different research groups have proposed ontological conceptualisation of the underlying aspects of MSA [7], [8]. Kokar *et al.* propose in [7] an ontology-based formalisation for Situation Awareness using OWL-Full and RuleML. The resulting ontology is called Situation Theory Ontology, a high-level ontology that may be applied to formalise Situation Awareness as specified by Endsley [9] in any domain. In [8] the same authors extend the discussion to modelling and querying geographical entities for inferring suspicious activities and discuss examples of formalisation for Maritime Security. Laskey *et al.* [10], [11] propose a probabilistic extension of the Web Ontology Language OWL, namely PR-OWL 2, incorporating a semantic formalisation of uncertainty. Specifically, it includes constructs to formalise Bayesian networks using OWL. The paper [10] addresses compatibility issues of the previous version of the upper ontology PR-OWL with OWL, and proposes a case study for the formalisation of uncertain maritime domain knowledge, discussing the identification of ships of interest exhibiting a suspicious behaviour or having suspicious characteristics. Events the authors formalise include unusual routes, *rendez-vous*, evasive behaviour, or crew members involved in terrorism. A consistent part of the literature on ontology for MSA tackles the semantic formalisation of behavioural patterns for the detection of maritime anomalies and their systematic classification in taxonomies and ontologies [12], [13], [14]. Other work addresses specific features of the maritime information, like uncertainty representation [10], [11], data integration [15], or automatic classification of ship behaviour [16], [17].

C. Uncertainty Representation and Reasoning Evaluation Framework

The Uncertainty Representation and Reasoning Evaluation Framework (URREF) ontology is being developed since 2011 with the aim of providing a sound and rational support for evaluation of uncertainty handling in fusion algorithms [4]. The URREF ontology has evolved further along the years, last contribution being [6], and the specific aspects of uncertainty representation have been discussed in separate work considering different awareness scenarios. For instance, [18] investigates the uncertainty representation in Bayesian networks, while [19] examines the sources of uncertainty in information fusion. In [20], uncertainty assessment criteria in the URREF are provided a detailed categorisation of input information for the comparison of expressiveness of two classical fusion schemes, illustrated on a maritime use case of addressing a typical threat assessment scenario.

Beside the identification and definition of lists of criteria for uncertainty representation and reasoning assessment (classes *RepresentationCriteria* and *ReasoningCriteria*), the classes *DataCriteria* and *DataHandlingCriteria* support the assessment of input and output data and data handling respectively. Peripheral concepts also exist such as *UncertaintyDerivation*, *UncertaintyNature* or *UncertaintyType* which aim at characterising the uncertainty to be captured and represented [20]. Also, the concept of *Source* (of information) relates to information and data to be processed. In [21], several source quality dimensions are identified, some of them being actually closely connected to some *DataCriteria* (e.g., *ObservationalSensitivity*, *SelfConfidence* and *Objectivity*).

The current version of the ontology sees *DataCriteria* gathering various aspects of data quality among which *RelevanceToProblem*, *WeightOfEvidence*, *Credibility* and *Quality*. The concept of *Credibility* is itself split into *ObservationalSensitivity*, *SelfConfidence* and *Objectivity*, suggesting that the credibility of a piece of information depends on several quality aspects of the source which provided it. In [22] the following about the relationships between information and source quality on the one hand, and between information and the related criteria of uncertainty handling on the other hand is noticed:

- **Source quality vs information quality**: For a fair rating process, source and information should be assessed independently. This allows to consider that a source of "good" quality may provide occasionally information of "bad" quality. Possibly, any Information Quality (IQ) dimension can be translated into a Source Quality (SQ) dimension. SQ differs from IQ in the sense that SQ is more perennial and IQ is rather instantaneous although source quality can still change over time;
- **Assessing source quality**: By definition, source of information provides information and thus should be assessed primarily according to the information it outputs. In this case, the SQ is assessed based on past experiences, experiments. However, other SQ dimensions can be assessed based on the ability of the source to provide information of good quality, without actually analysing the information provided. This is the case for instance of sources subject to manipulation, or within a context favoring deception;
- **Using SQ**: 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. The uncertainty representation and reasoning process should be able to capture and account for SQ, to correct the pieces of information provided by the source (either discount or reinforce) on the basis of prior knowledge of SQ;
- **SQ vs fusion algorithm quality**: Within a network of heterogeneous agents, the notion of source is relative, and depends on the perspective at hand. Some agent may be a source of information for another agent, who is itself a source for another one. 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 SQ should be assessed according to the source's ability to provide information, and its internal reasoning process should not be characterized in detail.

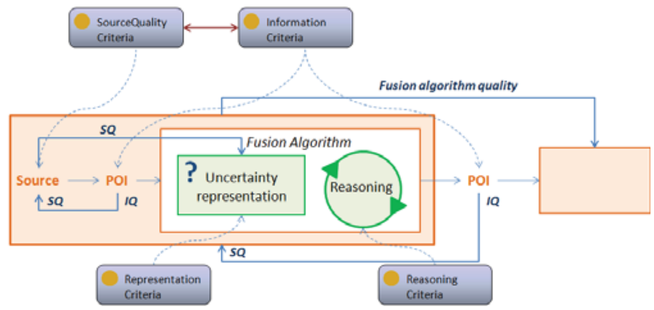

Fig. 1. Embedded assessments of source, information and algorithm [22]

Figure 1 illustrates the different assessment criteria classes involved (*SourceQualityCriteria*, *InformationQualityCriteria* input or output), *UncertaintyRepresentationCriteria* and *ReasoningCriteria*). Moreover, it is displayed how SQ is possibly captured by the uncertainty representation (or reasoning process) and how IQ is possibly transferred to SQ after a learning phase. Finally, the orange boxes illustrate the relative notion of "source" and how the fusion algorithm together with the sources can be wrapped into another source to provide information to be processed by another fusion algorithm.

Both SQ and IQ impact the assessment of uncertainty representation. On the one hand, the choice of uncertainty representation is driven by IQ dimensions as they are related to uncertainty characterisation. The choice depends on the nature of uncertainty and its derivation. It should also capture properly information imperfection such as uncertainty, incompleteness (or imprecision), gradualism or granularity [23]. On the other hand, once the uncertainty model is selected, its instantiation could be expected to account for the SQ and correct the singular statement (discount or reinforce) based on prior knowledge of the quality of this source.

In the remaining of this paper we propose a formalised articulation of IQ, SQ, and software (or algorithm) quality, as a basis for further refining the URREF ontology.

III. Information and Source Quality (ISQ) Ontology

The Information and Source Quality (ISQ) ontology is presented in the following subsections. Figure 2 illustrates an excerpt of the top-level concepts of the ISQ ontology, including different information-related concepts (classes of instances) and the relationships (object properties) among them. The ontology concepts are depicted as rounded labeled rectangles. Associations between concepts are represented by arrows: inheritance relationships (i.e., "is-a") are unlabeled arrows with white/empty head; labeled arrows with black heads are object properties named as the corresponding labels; starred labeled arrows (rel^*) are multivalued associations.

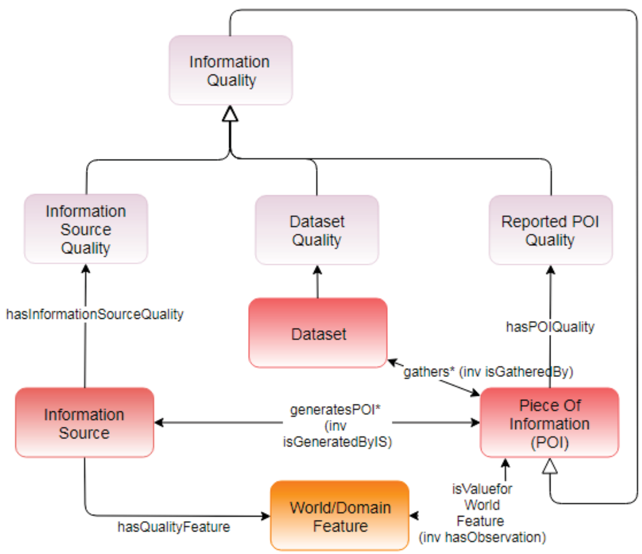

Fig. 2. Excerpt of ISQ ontology: top-level graph illustrating information and information sources and their quality.

A. From Information Source to Datasets

The class *Information source* gathers the instances of all relevant or available sources of information. These may be conveniently organised in subclasses, according to their common characteristics. For example, we can further partition the general class *Information source* into subclasses gathering *Surveillance sensor*s, *Environmental Sensor*s, *Maritime Registry*(ies) *and Gazetteer*s (e.g., port databases, vessel and maritime companies registries, company), *Navigation sources* (e.g., Nautical charts, navigation aids).

As represented in Figure 2, each instance of *Information source* may generate multiple *Piece(s) of Information* (cf. the multi-valued object property *generatesPOI** and its inverse *isGeneratedByIS*). For instance, a terrestrial AIS (T-AIS) receiver may generate different types of AIS messages, which may be broken down into different pieces of information according to the different AIS message fields. Each information piece matches a real *World Feature* in the domain of interest. MSA world features concerning vessels include vessel name, vessel dynamics (position and speed at a certain instant in time), but they may also correspond to more abstract features such as behaviour or intent for a single vessel, or to events of interest involving possibly several vessels (e.g., rendez-vous event). This relationship is modelled in the ontology through the object property *isValueForWorldFeature* defined for the concept *Piece Of Information*, and its inverse *hasObservation* which connects any world feature with the piece of information describing it. This relationship, as explained in more detail in Section IV-B, contributes to data and information's *interpretation*.

Piece(s) of Information may be gathered in *Dataset*s (cf. the object property *isGatheredBy*) or conversely, an instance of *Dataset gathers* multiple *Piece(s) of Information*. For the scope of this work, a dataset is a plain collection of pieces of information, fully represented, either as they are generated by the source or in a more convenient format (e.g., post-processed). Other forms of gathering of information pieces in datasets such as sampling, summarisation or aggregation of samples are not discussed here, but the formalisation we propose may be extended to support them as well.

B. Quality of Information and Sources

All information-related concepts introduced above may be quality rated. This is illustrated in Figure 2 with light purple boxes, where concepts *Information Source*, *Piece of Information* and *Dataset* have associated object properties that connect them to *Information Source Quality*, *Piece of Information Quality* and *Dataset Quality*, respectively.

1) Intertwined SQ and IQ: It is important to note that the quality of all instances of information-related concepts may be leveraged to derive the quality of the associated instances (via object properties) in other classes. The quality of a piece of information may be assessed relying on the quality rating, or the *reliability* of the source that generated it [21]: for instance, if we are aware that a camera distorts colours, the quality of the video samples it produces may be lowered accordingly; the quality of an AIS message field may be derived from the reliability of the AIS receiver it comes from. Analogously, the quality of a dataset may be evaluated considering the quality of the pieces of information it is composed of: for instance, the quality of an AIS messages dataset may be derived from the quality of given AIS fields, considering for instance the known global accuracy of kinematic and static features.

When considering multiple sources of information, the accuracy of a piece of information may be estimated relatively to other reported pieces of information by other sources. In this case, some inconsistencies may be revealed and the *credibility* of the information decreased [24].

Conversely, the *reliability* of an information source may be quantified analysing the quality of the information it produces. For instance, after assessing the recordings (i.e., the information samples) of a set of cameras measuring the height of waves, we could conclude that one of them has an erroneous calibration. and graded of “bad” quality. Later, based on this assessment, it could be excluded from the list of sensors to be used for estimating the sea state.

Beside the task or operational scenario, the SQ, and likewise the IQ, may be differently affected by their properties or features, as described below.

2) Quality Features: In the ISQ ontology, the properties that are relevant for quality assessment are defined as sub-properties of a generic, defined at top-level, quality-related property. For instance, in the top-level concept *Information Source*, the upper object property *hasQualityFeature* connecting *Information Sources* with *World Features* may be refined by different sub-properties defined in the information sources subclasses. These should be conveniently defined to leverage the characteristics of different sources.

To exemplify, suppose the class *Information Source* in the ISQ ontology is extended to include a class for *Positional Sensor*s, such as AIS and Radar sources. The quality of these sources is likely dependent on the accuracy of position and speed reported by the sensors for the objects they observe. We may model explicitly this knowledge in the ISQ class *Positional Sensor* by including two properties, *hasPositionQualityFeature* and *hasSpeedQualityFeature*, defined as sub-properties of the object property *hasQualityFeature*. If we want to model also the two quality assessments, we also define other two *Positional Sensor* object properties: *hasPositionAccuracy* and *hasSpeedAccuracy*, connecting *Positional Sensor* to *Information Source Quality*. Supposing that the class *Positional Sensor* is further refined, and that we want to model AIS sensors which accuracy depends on static information, we define a class *AIS* with a property *hasStaticInformationQualityFeature*, sub-property of *hasQualityFeature*. We may also define a property *hasStaticInformationAccuracy* connecting *AIS* to *Information Source Quality* to link the AIS sources with the assessment quality.

As for the quality assessment of information-related concepts, the assessment of implicitly represented concepts depends on the quality evaluation of the related instances. For example, the assessment of a type of source may be inferred from the quality evaluation of all the sources instantiated for the class representing that type, if the class' population is representative of the type itself. The same applies for all common characteristics of a class.

In the example above, once the ISQ ontology is populated with enough source instances, the class of *Positional Sensor*s may be qualitatively assessed relying on the quality evaluation of all the instances of the class, while the assessment for the class of *AIS* would rely on the quality assessment of AIS instances only, which would be based also on the accuracy of static features. Note that, lacking the support for class properties, such *class assessment* is not be explicitly formalised in the ontology. Given the above example, the ontology should be extended to include singleton classes to formalise information source types (e.g., *Information Source Type*). Similarly, the assessment of pieces of information classes may rely on the quality of their instances, and to formally represent them, the ontology should be extended with the concepts *Piece of Information Type* to include instance of the representing type.

Type-related concepts may be useful to exploit the common (global) characteristics, conveniently represented as properties in the corresponding class. For instance, the value of certain fields of AIS messages are manually inserted and therefore are prone to errors and falsifications (e.g., this is the case of the destination and the next port of call fields). Once the instances of this feature types are created, they may have a property to formalise the type quality, and a high uncertainty may be assigned to it, and used to qualify the quality of a AIS message and of the AIS message types that contains those fields. Similarly, the same information may be leveraged to assess the quality of the source type, i.e., AIS, beside the quality of a specific receiver. Indeed, as the quality of a class may be derived from the quality of its instances, the quality of the class may be leveraged to evaluating the quality of its instances.

3) Quality as Relevant Information: Finally, in the ISQ ontology, *Information Quality* assessment is a piece of valuable information to the operator and that we would like to be considered by the information fusion system. This idea is enforced by explicitly modelling the inheritance relationship between the classes *Information Quality* and *Piece of Information* in the ontology, as depicted in Figure 2. As a consequence, sources providing information quality assessment are part of the relevant sources of information to be represented by the ISQ ontology. It is the case of an intelligence analyst providing credibility and reliability rating of a piece of information, or of a classifier which training phase came out with a confusion matrix or predictive recognition rates. This relationship allows then to consider source reliability as a value for a World Feature of interest about which some uncertainty can be expressed.

IV. Information Processing Tools as Sources of Information

As discussed in the previous section and in [21], the notion of "source of information" is relative to the objective for which the pieces of information it produces are used for. A sensor (e.g. a radar) is an example of source of information, but also is a software, consuming information produced by other sources. For instance, tracking software, like other information fusion software elaborate pieces of information produced by positional sensors and output object tracks themselves being further ingested by other higher-level fusion algorithms to produce for instance some threat assessment.

In the ISQ ontology, the class *Software Tool* is defined as a specific category of *Information Source* (cf. Figure 3) and, differently from other information sources, *Software Tool*s use other sources as input (cf. the multivalued relationship *inputs*). Analogously to other sources, a *Software Tool* generates pieces of information (cf. the multivalued relationship *genertatesPOI** inherited from *Information Source*).

To exemplify how software tools relate to information sources, Figure 4 illustrates an instance of a *Software Tool* and its relationships with information concepts. The *Vessel Routes Extraction Tool* implements a *Vessel Routes Characterisation Algorithm* and processes *Positional Information* to generate *Vessel Routes*. This is formalised by an ontology axiom (the dotted line in the figure) on the object property *inputs** originally defined in class *Software Tool*, that restricts the values of this property to *Positional Information* for all the instances of class *Vessel Traffic Characterisation Software*. An example of *Positional Information* is the one provided by the *AIS*. In the figure, the tool processes AIS information and produces the *La Spezia Routes* dataset.

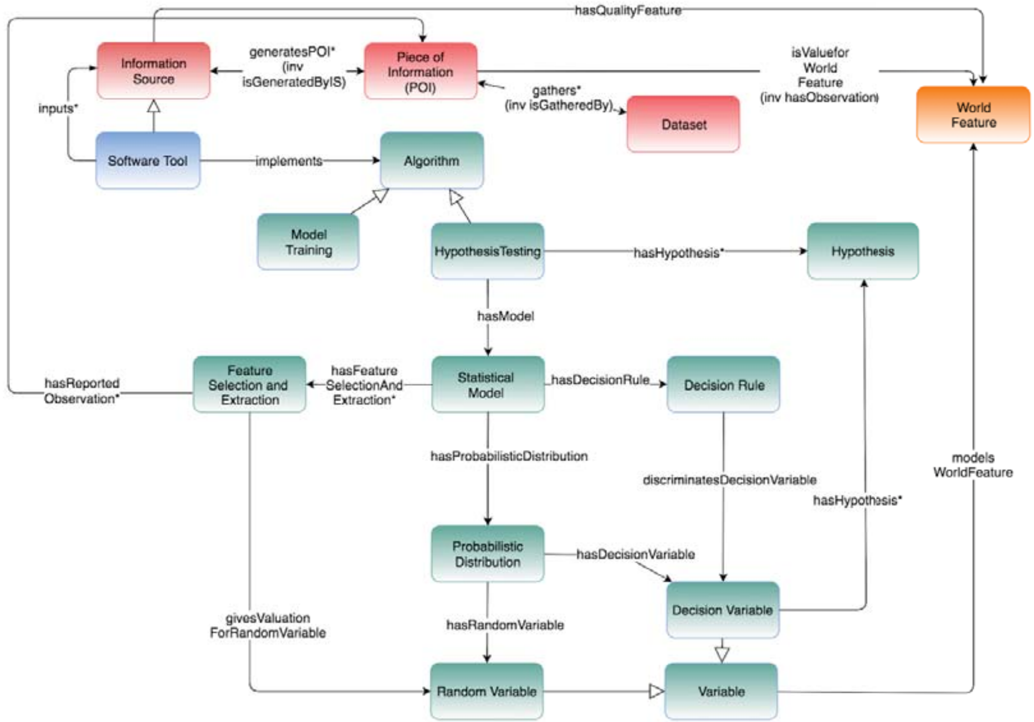

Fig. 3. Excerpt of the ISQ ontology including top-level concepts of statistical information sources for statistical based information processing.

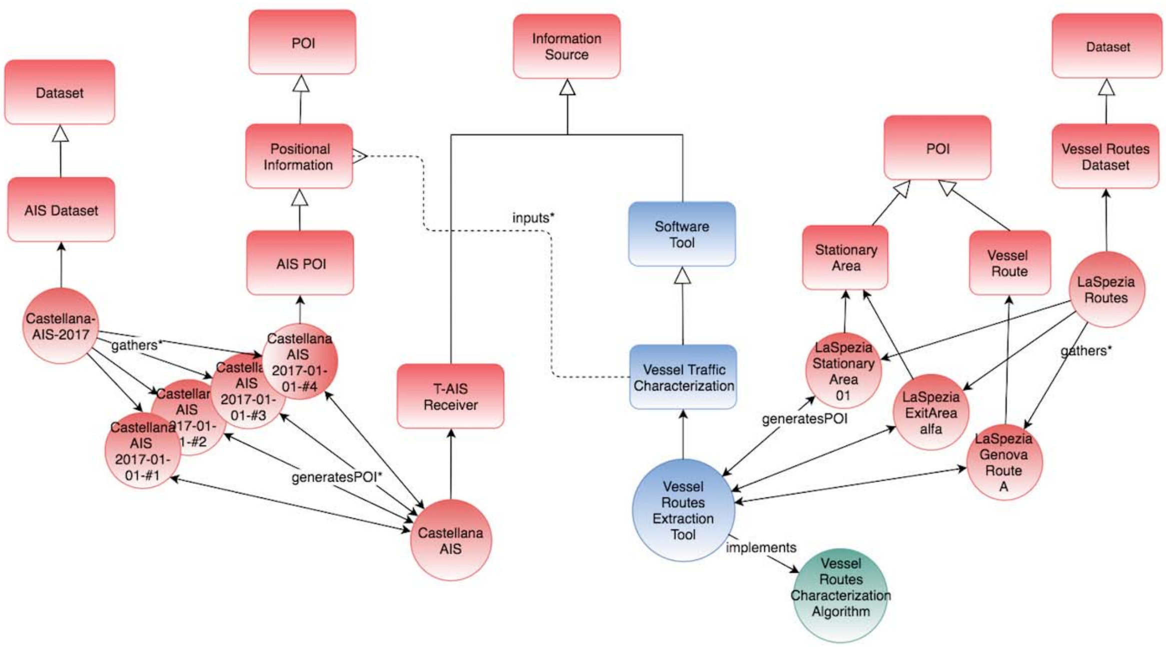

Fig. 4. ISQ ontology excerpt representing two information sources. A Maritime Situational Awareness tool, given positional information, outputs vessels routes and related pieces of information. Circles represent class instances. Dotted lines represent object properties with restrictions formalised by axioms.

A. Algorithms and Hypothesis Testing

Each *Software Tool implements* an *Algorithm*. The excerpt of the ISQ ontology displayed in Figure 3 focuses on a software implementing an algorithm for hypothesis testing relying on a statistical modelling such as for instance the ones described in [25]. The *Algorithm* class may be further expanded to include model training algorithms, quality assessment algorithms, data mining algorithms, or any other type of algorithm producing information about World Features of interest.

As depicted in Figure 3, *Hypothesis Testing* algorithms discriminate among a set of hypotheses, specified as instances of class *Hypothesis* (cf. the multivalued relationship *hasHypothesisToChoose**). The algorithm decides the hypothesis on the basis of a *Statistical Model* (cf. the relationship *hasModel* between *Hypothesis Testing* and *Statistical Model*), which elaborates the *Probabilistic Distribution*s of each hypothesis, based on the knowledge of the observed state given by the data in input and eventually some prior knowledge. The probabilistic distribution corresponds to an uncertainty representation to be further evaluated by the URREF [22]. Observed states are modelled by *Random Variable*s, while hypotheses correspond to possible values of a *Decision Variable*. The decision among the possible hypotheses (corresponding to a decision variable value) is taken according to a *Decision Rule*. *Random Variable*s and *Decision Variable*s are associated to features in the domain, or *World Feature*s. The *Feature Selection and Extraction* step provides valuations for random variables and has a corresponding observation, itself a piece of information. The observations are chosen among the pieces of information produced by the class of sources defined in input for the corresponding software. The excerpt of the ISQ ontology in Figure 5 shows a software tool implementing a vessel destination hypothesis testing algorithm. The software tool *Vessel Destination Prediction Software implements* a *Vessel Destination Testing* relying on the *Vessel Destination Statistical Model* and infers the *Vessel Destination Distribution*. The conditional probability of the *Decision Variable Vessel Destination* given the *Observed Vessel Kinematic State Vector* (positions and speed) is computed and the maximum a posteriori probability (*MAP*) is applied as a decision rule. The modelling proposed in the ISQ is compliant with the URREF, where a statistical model and derived probability distribution implicitly encode an uncertainty representation affecting the reasoning applied to infer new states of variables such as decision variables.

B. Interpreting Information

The interpretation of data and model features is given by ontology relationships that explicitly map the corresponding concepts to domain features: in Figure 3, *Random Variable* and *Decision Variable* connect to *World Feature* through the relationships (*models WorldFeature* defined in class *Variable*. For instance, the *vessel speed* is a world feature that may be represented in a kinematic model either as a decision or a random variable *speed* (in case the model is estimating the vessel speed, or using it for further inference respectively). The speed of the vessel may also be reported or estimated in surveillance data. For instance, values of speed measured by vessel instruments are reported in AIS message 5 in field `speed over ground (SOG)`. Both the model feature *speed* and the `SOG` field of AIS represent vessel speed in mathematical models and data, and their interpretation is given explicitly in the ontology through the relationships described above. These interpretation relationships complement the relation *isValueforWorldFeature* that connects *Piece of Information* to *World Feature*.

V. Discussion and Conclusions

In this paper, we proposed the Information and Sources Quality (ISQ) ontology, a top-level ontology describing the relationships between information quality and information sources quality. We exemplified its use in support to the design of algorithmic solutions for Maritime Situational Awareness. This proposed ontology is compliant with the Uncertainty Representation and Reasoning Evaluation Framework (URREF) ontology and expands the notions of *Source of information* defined therein, and *Quality* listed among the data criteria that relate to information for uncertainty representation. We formalise the construction of datasets from information items provided by information sources, and connects their corresponding quality. We expand the idea that a source is relative to the fusion algorithm and noticeably the implementations of fusion algorithms and quality assessment algorithms are defined in the ISQ ontology as sources of information. In alignment with the URREF approach, they are also source of uncertainty. In the ISQ ontology, we also formalised the algorithmic aspects of hypothesis testing and include the necessary support to model and information interpretation. The ontology is a basis for the development a shared conceptualisation of information and source related concepts together with their quality for an enlightened consideration of uncertainty handling in the design of data models supporting information sharing for Maritime Situational Awareness. This ontology is also an enabler toward an improved harmonisation of vocabulary, understanding and communication between scientists and Subject Matter Experts in this domain.

Acknowledgements

This research was supported by NATO Allied Command Transformation (NATO-ACT).

References

[1] F. Berizzi, A. Gabellone, E. Dalle Mese, A. Capria, M. Martorella, and W. Matta, “MSA system concept for Mediterranean sea surveillance,” *Information & Security: An International Journal*, vol. 27, pp. 183–206, 2011.

[2] C. Claramunt, C. Ray, L. Salmon, E. Camossi, M. Jousselme, A-L. Hadzagic, G. Andrienko, N. Andrienko, Y. Theodoridis, and G. Vouros, “Maritime data integration and analysis: recent progress and research challenges [vision paper],” in *Proceedings of the 20th International Conference on Extending Database Technology (EDBT/ICDT 2017)*, March 2017.

[3] EUCISE2020, “EUCISE2020 Data Model Technical Specifications Deliverable 4.3,” tech. rep., 2017.

[4] P. C. G. Costa, K. B. Laskey, E. Blasch, and A.-L. Jousselme, “Towards unbiased evaluation of uncertainty reasoning: The URREF ontology,” in *15th Int. Conf. on Information Fusion (FUSION)*, pp. 2301–2308, 2012.

[5] C. C. Insaurralde and E. Blasch, “Veracity metrics for ontological decision-making support in avionics analytics,” in *IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, Sept 2017.

[6] P. de Villiers, R. W. Focke, G. Pavlin, A.-L. Jousselme, V. Dragos, K. Laskey, P. Costa, and E. Blasch, “Evaluation metrics for the practical application of URREF ontology: An illustration on data criteria,” in *20th Int. Conf. on Information Fusion (FUSION)*, July 2017.

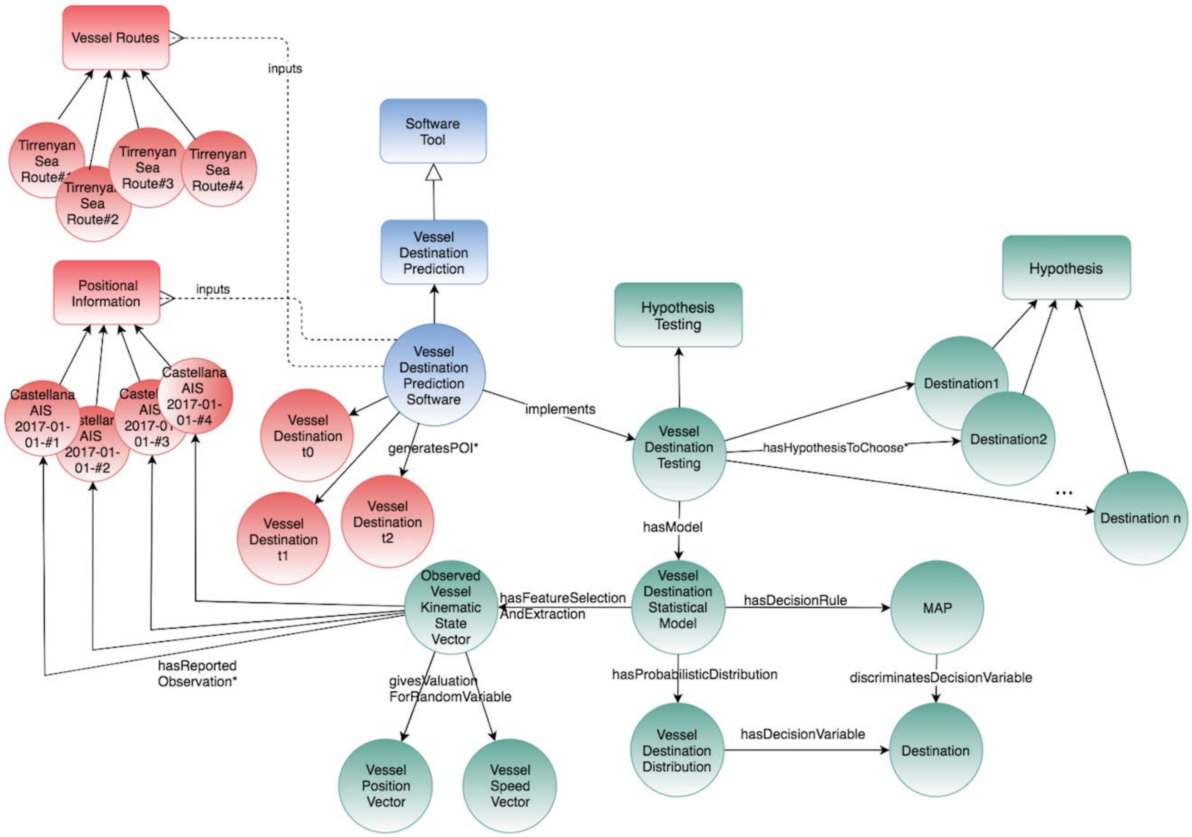

Fig. 5. ISQ ontology excerpt representing a software tool implementing an hypothesis testing algorithm to predict vessel destination. The software, given the observed kinematic state of a vessel and a set of precomputed vessel routes, applies a statistical kinematic model to predict the vessel destination, among a set of given hypotheses.

[7] M. M. Kokar, C. J. Matheus, and K. Baclawski, "Ontology-based situation awareness," *Information Fusion*, vol. 10, pp. 83–98, 2009.

[8] M. M. Kokar, B. E. Ulicny, and J. J. Moskal, "Ontological structures for higher levels of distributed fusion," in *Distributed Data Fusion for Network-Centric Operations* (D. Hall, C.-Y. Chong, J. Llinas, and M. I. Liggins, eds.), pp. 327 –346, CRC Press 2012, Nov 2012.

[9] M. R. Endsley and D. J. Garland, *Situation Awareness Analysis and Measurement*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers, 2000.

[10] K. B. Laskey, R. Haberlin, P. Costa, and R. N. Carvalho, "PR-OWL 2 case study: A maritime domain probabilistic ontology," in *CEUR Workshop Proceedings*, vol. 808, pp. 76–83, 2011.

[11] R. N. Carvalho, R. Haberlin, P. C. G. Costa, K. B. Laskey, and K. C. Chang, "Modeling a probabilistic ontology for Maritime Domain Awareness," in *14th Int. Conf. on Information Fusion (FUSION)*, 2011.

[12] A. Vandecasteele, R. Devillers, and A. Napoli, "From movement data to objects behavior using semantic trajectory and semantic events," *Marine Geodesy*, vol. 37, no. 2, pp. 126–144, 2014.

[13] J. Roy and M. Davenport, "Exploitation of maritime domain ontologies for anomaly detection and threat analysis," in *International Waterside Security Conference, WSS 2010*, nov 2010.

[14] B. van den Broek, F. Bolderheij, M. Neef, and P. Hanckmann, "Maritime Anomaly Detection by Fusing Sensor Information and Intelligence," in *Int Workshop on Maritime Anomaly Detection (MAD)*, pp. 33–34, 2011.

[15] S. Brüggemann, K. Bereta, G. Xiao, and M. Koubarakis, "Ontology-based data access for maritime security," in *Lecture Notes in Computer Science*, LN in Artificial Intelligence and LN in Bioinformatics, 2016.

[16] G. de Vries and M. van Someren, "Recognizing vessel movements from historical data," in *Situation Awareness with Systems of Systems* (P. van de Laar, J. Tretmans, and M. Borth, eds.), pp. 105–118, New York, NY: Springer New York, 2013.

[17] W. R. van Hage and D. Ceolin, "The simple event model," in *Situation Awareness with Systems of Systems* (P. van de Laar, J. Tretmans, and M. Borth, eds.), pp. 149–169, New York, NY: Springer, 2013.

[18] J. P. de Villiers, G. Pavlin, P. Costa, K. Laskey, and A.-L. Jousselme, "A URREF interpretation of Bayesian network information fusion," in *17th Int. Conf. on Information Fusion (FUSION)*, July 2014.

[19] 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 *18th Int. Conf. on Information Fusion (FUSION)*, pp. 50–57, July 2015.

[20] A.-L. Jousselme, "Semantic criteria for the assessment of uncertainty handling fusion models," in *19th Int. Conf. on Information Fusion (FUSION)*, pp. 488–495, July 2016.

[21] A.-L. Jousselme, A.-C. Boury-Brisset, B. Debaque, and D. Prévost, "Characterization of hard and soft sources of information: A practical illustration," in *17th Int. Conf. of Information Fusion (FUSION)*, (Salamanca, Spain), 2014.

[22] J. P. de Villiers, G. Pavlin, P. Costa, A.-L. Jousselme, K. Laskey, V. Dragos, and E. Blasch, "Subjects under evaluation with the URREF ontology," in *20th Int. Conf. on Information Fusion (FUSION)*, 2017.

[23] D. Dubois and H. Prade, *Formal representations of uncertainty*, vol. Decision-making - Concepts and Methods, ch. 3, pp. 85–156. ISTE, London, UK & Wiley, Hoboken, N.J. USA, 2009. Invited paper.

[24] M. C. Florea, A.-L. Jousselme, and E. Bossé, "Dynamic estimation of evidence discounting rates based on information credibility," *RAIRO Operations Research*, vol. 44, pp. 285–306, October 2011.

[25] B. Ristic, B. La Scala, M. Morelande, and N. Gordon, "Statistical analysis of motion patterns in AIS Data: Anomaly detection and motion prediction," in *11th Int. Conf. on Information Fusion (FUSION)*, (Cologne, Germany), 2008.

Document Data Sheet

Security Classification		Project No.
Document Serial No. CMRE-PR-2019-036	**Date of Issue** May 2019	**Total Pages** 8 pp.

Author(s)

Elena Camossi, Anne-Laure Jousselme

Title

Information and source quality ontology in support to maritime situational awareness

Abstract

To support situation awareness, the benefit of using a variety of sources is undeniable although it brings additional challenges related to possible conflicting information, heterogeneity in data formats, semantics, uncertainty types, and information quality. Information and source quality are intertwined concepts which assessments connect with the evaluation of uncertainty handling in information fusion solutions. While the Uncertainty Representation Reasoning Evaluation Framework (URREF) ontology focuses on assessment criteria, peripheral concepts still play a critical role. In this paper, we propose an Information and Source Quality (ISQ) ontology formalising the relationships between information-related concepts, and discuss information interpretation in support of Maritime Situation Awareness. Specifically, this paper links the concepts of Information Source, Dataset and Piece of Information, and connects them to the corresponding quality concepts. Such concepts link to the upper level concepts of the URREF ontology Source (of information) and data Quality. The ontology further expands to the uncertainty modelling and the algorithm design. We conclude on future work and identify future avenues, especially the extension to the formalisation of the evaluation process.

Keywords

Information sources, information quality, uncertainty, maritime situational awareness

Issuing Organization

NATO Science and Technology Organization
Centre for Maritime Research and Experimentation

Viale San Bartolomeo 400, 19126 La Spezia, Italy

[From N. America:
STO CMRE
Unit 31318, Box 19, APO AE 09613-1318]

Tel: +39 0187 527 361
Fax:+39 0187 527 700

E-mail: library@cmre.nato.int