Knowledge Representation and Reasoning for Defence

Dr. B. K. Madahar[1], Dr. P. Cripps[1], Dr. D. Barber[1]
[1]Defence Science and Technology Laboratory, CISD
UNITED KINGDOM

1bkmadahar, pcripps, dbarber@dstl.gov.uk

ABSTRACT

This paper describes a foundational shift for Knowledge Representation and Reasoning (KRR) and its potential for Defence and Security applications. Underpinning research for the approach is introduced which shows how techniques from Graph Theory (GT) and Knowledge Engineering (KE) are exploited. Graphs and GT are well-established research areas with successful applications in a number of defence relevant sectors such as IT and telecommunications. KE fuses computational linguistics and semantics with GT to represent knowledge to reason and infer understanding. Furthermore, civil sectors (e.g. finance, marketing, pharma) are embracing new developments in data sciences (e.g. Artificial Intelligence (AI) and Machine Learning (ML)). Mainly to improve decision making, target services and increase the pace at which new solutions can be delivered. There has been some work in Defence in this area e.g. the use of ontology to support semantic interoperability; real-time semantic analysis of multi-modal streams (e.g. video, images, text, audio, social-media) to identify and track multiple entities of interest, including evolving behaviours and relationships. Adding these to the research mix provides an opportunity to expand KRR much further. In particular, how the developments, with machines, can provide the knowledge based systems and analytical support needed by Defence for full spectrum operations at pace. Where, for example, interconnection of observations between the physical, social and cyber domains may not be easily discernible, nor connections of such observations to past knowledge and staff expertise, past and present.

The foundational shift is in fact bi-directional understanding of information and knowledge by humans and machines with fusion of diverse, heterogeneous sources. Enabling machine reasoning (inferencing) methods, which apply rules and formal logic to available data in order to offer higher order deductions. Knowledge Representation (KR) is the expression of knowledge in computer-tractable form in order for it to be exploited (e.g. reasoning). Thus, use of terms symbolic AI and rule-based AI by Knowledge based system. What is missing is the (semi)automated fusion or KRR against observations to support the goals of high level fusion.

In this paper, using analysis of the state-of-the-art, we outline approaches to establish effective KR that can be used by future hybrid systems. We argue that symbolic methods are more adept at dealing with sparse data, support enhanced explainability, incorporate past human knowledge, and can exploit computational methods which excel at pattern recognition and data clustering/classification problems. Furthermore, such approaches/technologies can support future coalition operations (e.g. hybrid warfare), providing the coalition can:

i. develop or adopt domain specific and upper or top-level ontologies and event driven architectures;

ii. assess the inference capabilities, including handling of uncertainty/ambiguity;

1. INTRODUCTION

1.1 Graph Theory

The common saying ‘a picture paints a thousand words’ recognises the power that visual representation has for human perception and understanding. Enabled by human vision it helps the mind build a rich, active representation of the scene, of the identity of entities within it, and the relationships between them. Developing machines with this level of competence remains an enduring challenge requiring collaborative advances in multiple disciplines such as human psychology, neuroscience, perception, cognition, computer science, Machine Learning (ML), statistics, and Artificial Intelligence (AI).

Graphical representations, that is Graphs, provide an enabling component in that direction. Many real world, or abstract, situations, can be depicted by a diagram consisting of a set of points (entities) together with lines joining certain pairs of these points (e.g. a family tree, a road system or a convolutional neural network). Mathematical abstraction of situations of this type gives rise to the concept of a graph and development of underlying science of graph theory [1]. Formally a general graph *G* consists of three things, a set *VG* (called the *vertex (or node)*), a set *EG* (called an *edge (or line)*) and an incidence relation *(or relationship)*, a subset of *VG x EGA*, required to be such that an *edge* incident with either one *vertex* (a loop) or two *vertices* [1]. Because of its broad application, Graph Theory (GT) is well-established area of science, with a strong body of knowledge. Graphs can be used to model types of relations and processes in physical, biological, social and information systems [2]. Often where the term Network(s) is used (in its simplest form a collection of points (*nodes*), joined together in pairs by lines (*edges*)), a graph, in which attributes (e.g. names) are associated with the *vertices* and *edges*, and their *relationship* defined postulated to try to understand the Network and its dynamics (e.g. communications, information, transport, social systems) [3].

Underpinned by GT, the utility of graphs for these application areas has lent it to be extended to address real or abstract problems in other areas of network, information and web sciences. An example is the representation of *knowledge*, as managed by and derived from *data* and *information* processing systems, the subject of this paper.

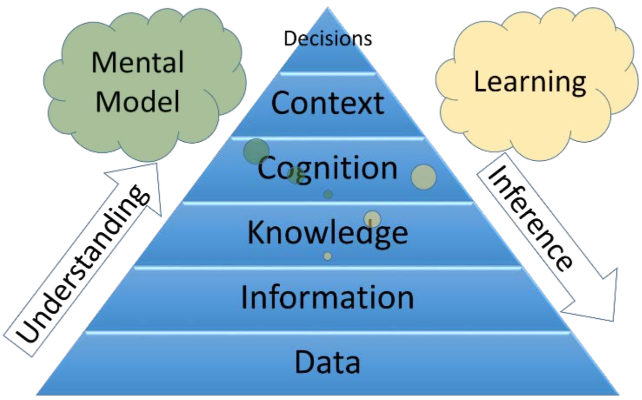
Epistemologically, the definitions of these three terms in italics is beyond the scope of the paper. This is because they are used casually by many researchers, definitions are vague and imprecise - the relationships between them, although non-trivial, of intertwined and interrelated concepts are not sufficiently dealt with [4]. Therefore, we cast their use in commodity processing systems from an Information Science and Technology (S&T) perspective. Outlined in the next section, where we also address other terms in common use such as Knowledge Base (KB), KBased Systems (KBSs) and ontologies. The aim being to establish the basis for Knowledge Graphs (KGs) as they underpin the technical approach to Knowledge Representation and Reasoning (KRR), the foundational shift, presented in the sections thereon. Overall advantages afforded to defence analysts, working as a team with machines, such as more effective and efficient analytical support and better intelligence products, is summarised in the final sections with conclusions as to the actions needed to transition the foundational research towards system implementations.

2. DATA, INFORMATION, AND KNOWLEDGE

2.1 Data and information

A layered view of a processing system offers a convenient view of these terms, as shown in Figure 1, and is in our opinion commonly used. The tapering is not meant to depict the importance of one over the other but simply to recognise the decreasing scale of the entities relative to each other. Nor does the tapering imply uni-directional direction of flow, though common, as higher level entities can lead to changes to lower level entities. The whole should be in fact be considered as a continuum. Data are the base representation of discrete facts, observations, or measurements in a form and format that is amenable to computational processing by systems. What data or their sets, databases, means or conveys that has purpose and relevance to a particular query and in a particular context is information. Conversely, or what can be gleaned from knowledge and codified. Views of data and information, with some similarity to these, were expressed by Stenmark [4], and others listed by him for discussion in his paper (Figure 2). Knowledge however, is a much more elusive and abstract concept and the term is often used interchangeably with information but the two are not the same, though can be similar in some aspects.

Figure 1: Layered system view of information related terminology



2.1 Knowledge

Elaborating on the work of Nonaka and Takeuchi [5], we argue that knowledge relates to the mind, is about beliefs, insights, trust and commitments governed by the mental model[1] held by the beholder. Knowledge creation is a continuous interaction between tacit and explicit knowledge: tacit knowledge is described as knowledge that is 'personal, context-specific, and therefore hard to formalise and communicate'- explicit knowledge is described as 'knowledge that is transmittable in formal, systematic language' (e.g. reports, books, media etc.) [5]. It is the representation of knowledge in the round such that it can be managed, shared and built upon which is the challenge. Moreover, its instantiation from the abstract to a form

Figure 2: Example of definitions for data, information and knowledge

Data	Information	Knowledge
-	Facts organised to describe a situation or condition	Truths and beliefs, perspectives and concepts, judgements and expectations, methodologies and know.how
-	A flow of meaningful messages	Commitments and beliefs created from these messages
Not yetinterpreted symbols	Data with meaning	The ability to asing meaning
Simple observations	Data with relevance and purpose	Valuable information from the human mind
A set of descrete facts	A message meant to change the receiver' s perception	Experiences, values, insights, and contextual information
Text thatdoes not answer questions to a particular problem	Text that answers the questions who, when, what, or where	Text that answers the questions why and how
Facts and messages	Data vested with meaning	Justified, true beliefs

[1] We define as - psychological representation of real, hypothetical, or imaginary situations that is used to anticipate, predict and reason events and underlying explanations such as action/consequence

amenable for processing codified as data and information. Hence data, information and knowledge are intertwined, interrelated and influence each other. Both data and information require knowledge in order to be interpretable, but at the same time, data and information are useful building blocks for constructing new knowledge, such as through learning[2]. A better understanding of the overall relationship between the three, and how this can be codified, must by definition lead to better representation of knowledge and enable its sharing, communication, with others to derive more value[3].

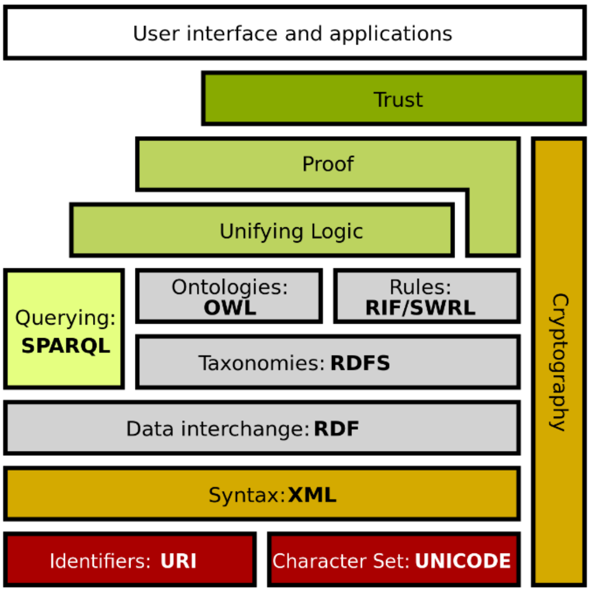
If we consider more than an individual, such representation requires being formal, structured and communicable. Take for example text written by an author. Within that, the author has used their mental model, tried to articulate their thoughts about something based on their knowledge. But the context, concepts and meanings being articulated are understood by the author and can, to an extent, be interpretable by the reader if they share the same language, vocabulary, including definitions, and some of the contextual knowledge needed. If the information communicated between the two becomes too distant from the knowledge required to interpret it, de-conceptualised, then it becomes just data.

2.2 Ontology

A formal and structured way of representing the concepts and relations of a shared conceptualisation [6] has emerged as the commonly accepted definition of ontology in Information Sciences, as opposed to other meanings of ontology used in philosophy or metaphysics[4] (see [7]) for a fuller discussion of this). Conceptualisation of an area of interest, i.e. domain knowledge, where concepts, other entities, and the relationships between them are assumed to exist, needs formal representation if it is to be processed. Domain is referred to as the universe of discourse within which objects, and the describable relationships among them, are reflected in the representational vocabulary with which KBSs represent knowledge [6]. This requires the design, development and creation of ontologies for a domain by extracting relevant instances of information from KB, ontology population, or through automatic ontology learning [8]. Ontological representations with their explicit, machine processable semantics are commonly used as KBs in AI applications (e.g. in the context of KBSs). The application of an ontology as a KB facilitates validation of semantic relationships and derivation of conclusions from known facts for inference through reasoning [9].

With the advent of the internet, the irrepressible growth of the web and related technologies (e.g. communications, networks and services), has led to the continuing and exponential growth in the ubiquitous availability of data and information (across heterogeneous systems, programming languages and network protocols) in most market sectors and aspects of knowledge thereof. But, for this to be of value, consistent and formal representation is needed of the distributed things, the groups of things and relationships between things for any area of interest. The concept of the semantic web emerged as a response to this [11]; a web of data elements, linked to enable search and discovery and online analysis and has continued apace since [10]. Three key enablers, through standardisation efforts of

Figure 3: Semantic web stack (www.w3.org)



[2] We define as - the process of changing knowledge and skills

[3] Benefit through sharing understanding of something and through learning

[4] Nature and essence of things, The Chambers Dictionary, 13th Ed.

the World Wide Web Consortium (W3C) and working groups, have been the Resource Description Framework (RDF[5]), Web Ontology Language (OWL)[5], and query languages for RDF such as SPARQL[5]. These are part of the W3C technology stack, with the first being a standardised graph model, the second being a logic-based specification language for ontologies building on RDF and the third specifies the syntax and semantics of the query language for RDF. This technology stack, and the supporting tools, has been extended much further, with contributions from standardisation bodies (e.g. Open Geospatial Consortium (ogc.org), Internet Engineering Task Force (ietf.org)) to provide the foundations for Knowledge Engineering at web scale **(**Figure 3**)**.

3. KNOWLEDGE GRAPHS

The origin of the term KG is attributed to a company Google blog post by Singhal[6] in 2012. It stated a summary of work "on an intelligent model—in geek-speak, a "graph"—that understands real-world entities and their relationships to one another: things, not strings. The Knowledge Graph enables you to search for things, people or places that Google knows about.....and instantly get information that's relevant to your query. This is a critical first step towards building the next generation of search, which taps into the collective intelligence of the web and understands the world a bit more like people do". The overall goal being to enhance Google search through three key ways – a) Find the right thing (e.g. disambiguate language), b) Get the best summary (e.g. concise content summary and key facts about the particular thing), and c) Go deeper and broader (e.g. make unexpected discoveries). Further developments have led to synonymous term to KG, Knowledge Vault, introduced by the company to generate the largest store of knowledge by automatically pulling in information from all over the web, using ML to turn the raw data into usable pieces of knowledge[7].

Ehrlinger and Woess [12] argue that the terms Knowledge Vault, KB and Ontology are used as synonyms, as in the above and as 'buzzwords', and attempt to provide a distinction between them and a definition of KG. They provide a summary of definitions, listed in Figure 4 below, from other researchers, found with references in their paper. From their critique of these, and usage of terminology in research literature, they propose the KG definition as - *"a knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge"*. Arguments in support are that it does not include scale (e.g. as implied by knowledge vault) and aligns with the assumption that a KG is superior and more complex than a KB or ontology. Such a definition is sufficient for the purpose of our paper as the distinctions being cited need to be analysed epistemologically to carry any significance.

Figure 4: Example of definitions of KGs [12]

- *A knowledge graph (i) mainly describes real world entities and their interrelations, organised in a graph, (ii) defines possible classes and relations of entities in a schema, (iii) allows for potentially interrelating arbitrary entities with each other and (iv) covers various topical domains.*
- *Knowledge graphs are large networks of entities, their semantic types, properties, and relationships between entities*
- *Knowledge graphs could be envisaged as a network of all kind of things which are relevant to a specific domain or to an organization. They are not limited to abstract concepts and relations but can also contain instances of things like documents and datasets.*

[5] https://www.w3.org/2001/sw/wiki/OWL.or..RDF.or..SPARQL

[6] https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html [dl Apr, 2020]

[7] https://www.newscientist.com/article/mg22329832-700-googles-fact-checking-bots-build-vast-knowledge-bank/# [Apr, 2020]

- *Knowledge Graph is an RDF graph*
- *Systems exist and use a variety of techniques to extract new knowledge, in the form of facts, from the web. These facts are interrelated, and hence, recently this extracted knowledge has been refered to as a knowledge graph.*

4. DEFENCE AND SECURITY APPLICATIONS

4.1 Background

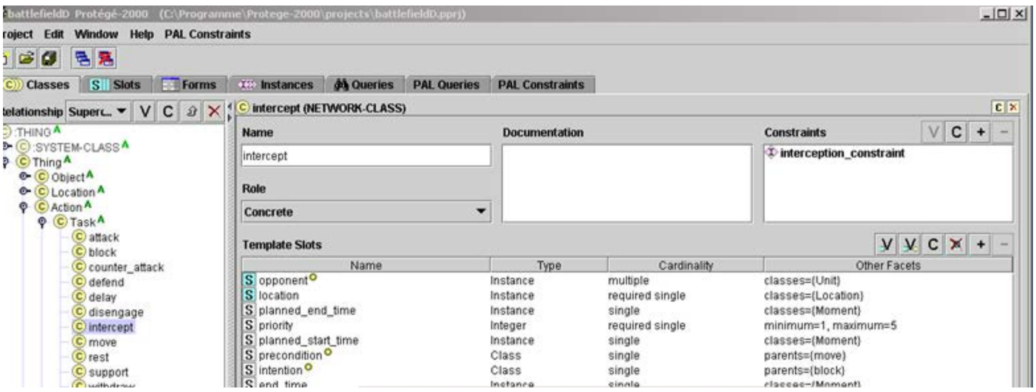
The core research reported in this paper is from the Dstl Underpinning Data Sciences (UDS) research project, under the Information Systems Programme[8]. This is advancing research in the following key areas of which the first four are most pertinent. *a) Development of domain specific ontologies, b) Building effective architectures for event driven processing, c) Deployment and assessment of inferencing capabilities, d) Handling uncertainty and ambiguity in observations, e) Information sharing and observation provenance, f) Approaches for federated deployment and coping with scale, g) Standards and interoperability*. Part of the research has also been through collaborations with academia, industry and under the NATO Information Systems Technology (IST) research collaborative programme of work with coalition partners. Two NATO projects of specific mention are NATO-RTG-144, on content based multi-media analytics [14], and recently completed exploratory research NATO-ET-111, on KRR [15]. Both providing direction to the work reported in this paper.

Regarding KRR the starting point, as eluded to earlier, is determining the domain and context of application supported by defining and designing an ontology that fits purpose. This is a critical step where so far we have essentially commented on static applications. Whereas those dynamic and temporally varying are more typical in general, and representative of defence and security in particular, as addressed in the next sections.

4.2 Ontologies and approach

Ontologies are typically constructed by defining classes of things, with relations, functions, data values and axioms to constrain their interpretation. These classes are based on common characteristics, and subclasses can be seen as specialisms of parent classes in that all individual members of a subclass are also members of any parent classes. This leads to a static representation of the truth of some domain represented in the form of axioms. Example shown in Figure 5, an ontology for the planning and deployment of assets, such as tanks, in the area of operation. Define class *Tank* of which *Battletank* is a sub-class and then specific NATO

Figure 5: Screen shot of simple defence application of tank deployment



---
[8] https://www.gov.uk/guidance/information-systems-programme [May, 2021]

battletanks that are sub-classes, such as *Challenger (UK), Abrams (USA), Leclerc (FRA),* and *Leopard (DEU)*. This ontology not only groups assets by what they physically are, but includes relations to their capabilities and potential actions. Though usable, it is too simple for effective application and fundamentally limited to static domains and static representation of domain knowledge; any shift in this knowledge requires change to the ontology.

The fundamental shift in approach needed, in addition to managing dynamic and temporal changes, a critical requirement for most applications, is the concept of KG as a means to achieve fusion and understanding in a form accessible to both humans and machines.

4.3 Case study and findings

The form of KG adopted is an RDF graph underpinned by an explicit specification (i.e. a formal ontology) grounded in Description Logic, using components from the W3C semantic web stack (Figure 3) in a layered, modular style of semantic architecture. The advantage of this form of KG, compared to other forms, is that it can support the kind of explicit, robust semantics described above. Such a layered modular construct serves our cases well. As our focus is on information fusion and semantic interoperability, sensemaking and understanding, and prediction using widely available Commercial-off-the-shelf (COTS) and Free and Open Source (FOSS) components such as graph databases (also known as triple store) and reasoners wherever possible, supported by developing additional components where necessary.

4.4 Architectural Overview

Our general approach to architecture therefore, is to exploit the W3C semantic web stack (since it provides much of the information framework) and open standards in order to be able to provide interfaces between components and make use of any technology that supports relevant open standards. SPARQL interface, for example, with compliant graph databases thus avoiding any proprietary technology. The use of loosely coupled APIs enabling rapid development of new components to API specifications using the OpenAPI standard avoids brittle architectures. Employ RDF, sort of schema in other information systems, for data interchange as well as to provide content and metadata. All of this structure and content are specified using a layered and modular semantic framework comprising ontologies and other vocabularies.

A core ontology, specified using OWL, to cover all concepts that are stable and common to all use cases and domain extensions to provide coverage of more specialist domains (e.g. maritime situational awareness, pharmaceutical production, and our recent work on information operations). This is analogous to a Tree where the Trunk is the core ontology, and the Branches, the domain extensions. The Leafs, represented using lighter weight semantics in the form of Simple Knowledge Organization System[9] (SKOS), as per [16], provide specifications of taxonomies, thesauri and other classification schema and structured vocabularies.

Using the tree analogy makes it clearer why our layered, modular approach has utility. The trunk, core ontology, comprises all of the essential elements to describe any domain of interest. An upper level generic ontology, high level conceptually, to describe all entities and relationships in a stable way. Simple, lightweight but also rigorous enough to support transformation to other forms of representation as may be required (e.g. Basic Formal Ontology (BFO), Semantic Sensor Network (SSN) and the Business Objects Reference Ontology (BORO)).

The domain extensions then provide further elucidation of the concepts from the core ontology but remaining conformant to its ontological commitments (i.e. extending the core ontology without needing updates to it). Also enabling addition of new domain extensions, modifications to existing ones without any bearing on other domain extensions. This layer is also expressed as a formal ontology, using OWL, to support inference and reasoning, as well as to be able to apply constraints and use closures to restrict the

[9] https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html

open world model as necessary. Reference resources then provide additional detail as per standards agreements and national doctrines. Such material perhaps governed by external organisations, and therefore subject to ownership, change and less stable. SKOS facilitates dynamic approaches to governance enabling users to add new concepts without changing formal ontology.

Attempting to engineer an ontology of such information would be an enormous effort, requiring significant ongoing resource to maintain. By deliberately not choosing to incorporate such resources into an ontology, instead leaving them in the form in which they are maintained, but adding a lightweight semantic layer, it becomes possible to simplify governance processes and avoid any requirement for consensus on each and every concept. Each nation and/or organisation can produce structured vocabularies according to their needs, these structures can be represented using SKOS and, importantly, the SKOS vocabulary provides sufficient semantic relationships to be able to describe how concepts in each structured vocabulary relate to one another. Where additional logics are required, to support particular use cases, these SKOS vocabularies can be extended using OWL constructs.

Finally, common to both, the core ontology and domain extensions, are the related principals of modularity and the use of Ontology Design Patterns (ODP) [17]. Modularity involves using well defined, reusable extant ontologies wherever possible, for example the GeoSPARQL ontology to provide geospatial elements and OWL-Time to provide temporal elements. The establishment of patterns allows for similar subgraphs within the ontology to share common forms of representation, and thus support consistency across the ontology layers.

4.5 Conceptual Overview

Within the semantic framework outlined above, a number of conceptual innovations have been developed to accommodate the kinds of information routinely found in defence and security contexts. The approach to ontology presented here is grounded in the kinds and characteristics of this information to produce a dynamic model capable of handling changing and uncertain information within an explicit semantic framework. Taking in some developments in computable epistemology emerging from research in KRR.

The model draws on a number of theoretical constructs. The first is *phenomenology*, a branch of philosophical discourse concerned with the experiential nature of existence through which the production of knowledge is seen as incomplete and perspectival and is an approach originating in the Digital Humanities [18][19]. Using a phenomenological approach, the production of knowledge is through the description of the world in terms of phenomena, experienced, observed, and documented. The central tenet is the concept that knowledge is the product of some human action and is a subjective perspective on some reality, a perspective that may be partial and uncertain. Crucially, there may be multiple perspectives. The second builds on this, in that perspectives of reality are documented through observational activities. Key to this is the conceptual *SOM/OBP/ABI* approach to GEOINT developed by the NGA [20]. Structured observations are collected through a process of *Structured Observation Management* (SOM), with these observations being used to generate reporting based around entities or objects in a process of *Object Based Production* (OBP) to support analysis focussed on activities in which these entities participate, i.e. *Activity Based Intelligence* (ABI). We have used these concepts to support the phenomenological approach to intelligence gathering and sensemaking.

These two theoretical constructs form the basis of the model that is called the *Phenomenological Observation Model* (POM). The form of this conceptual model is a formal ontology in which the central ontological entity is the *Phenomenon* i.e. happenings in the real world involving entities of interest about which we wish to know. Classes of phenomena are used to describe all things with some spatio-temporal extent, for example an *Event* class represents phenomena with a definable (but not necessarily known) spatio-temporal extent in the real world (e.g. the birth of Caesar, the sinking of the Titanic, a police stakeout) and which bring about some change in state of some thing. A *State* represents some temporally bounded phenomena applying to one or more things (e.g. condition states such as colour, weight, length, names/identifiers; relationship states such as marriage, employment). Associated with *Phenomena* are structured descriptions of them, referred to as *Observations*. Together, it is *Observations* of *Phenomena* that form the major constituents of the core ontology discussed above. This method provides a basic pattern for describing things that happen in the world. Indeed, this basic pattern has been shown to be extensible to all key requirements, as noted below.

The treatment of spacetime associated with *Phenomena* is a good example of how the conceptual model is more closely aligned to how humans understand and make sense of the world. The POM is not dependent on any absolute positioning system for either space or time, unlike most extant information systems dependent on coordinate geometry and/or timestamps to define and describe spatial and temporal units of information. Innovations developed through work of W3C and OGC on semantic models of space and time have been leveraged in the POM by incorporating published standards from those organisations as modules within the POM. The GeoSPARQL ontology [21] is a lightweight ontology concerned with space and place, grounded in human geography, which provides relative positioning in addition to absolute positioning in any coordinate reference system. Similarly, the Owl-Time ontology [22] is analogous in its construction but concerned with time, also providing relative and absolute positioning in supported temporal reference system. Both of these referenced ontologies used within the POM allow for descriptions of places or temporal intervals about which we know nothing, an essential construct for intelligence information where our knowledge is often partial and incomplete. Observations may simply state "…*in London*" or "…*before takeoff*" from which it may be possible to infer some spatial or temporal bounds, but it is essential to be able to define an entity such as "…*the terrorist training camp*" or "…*the hold of the ship*" without knowing anything more about these places at the time of an observation being produced.

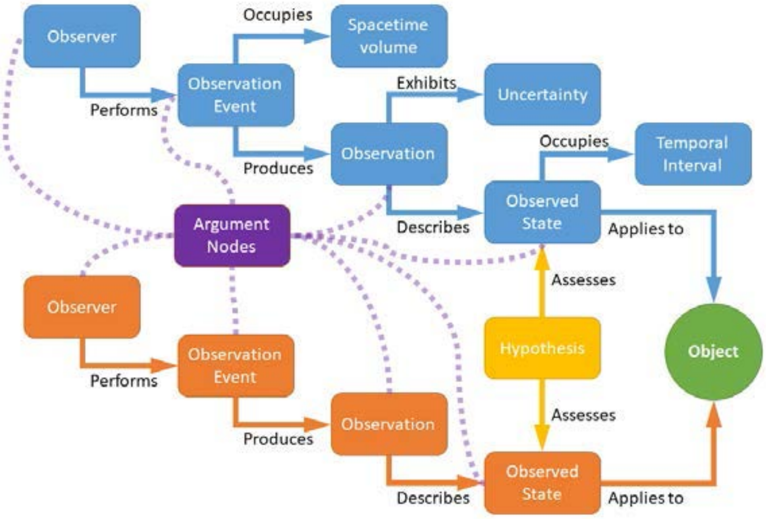
As noted above, these *Observations* are inherently subjective assertions about some perceived reality; they are explicitly produced by some *Agent* (a human or machine) and describe *Phenomena* according to the frame of reference of said *Agent*. By modelling the production of an *Observation* as an *Activity* carried out by some *Agent*, the conceptual model is able to handle multiple observations of *Phenomena*. This uses some ontological sleight of hand in that the ontological truth in the KG based on this ontology is transferred from the axiom itself (e.g. "*the car is red*") to the fact that the axiom was asserted (e.g. "*AgentX observed the car to be red*"); the Observation itself may or may not be true, but that is a different question. In terms of the ontological assertion within a KG based on this ontology, there is no conflict or inconsistency and thus all assertions remain valid according to the ontology, with the potential for further observations to be made regarding observations. This provides a means to be explicit about the nature of observation within the model and epistemological concerns. Taking the example of identifying and classifying a feature from aerial imagery, it is not the case that the feature observed *is* of type X, rather it is the case that an *Agent* (human or machine) has asserted that the observed feature is of type X. The axiomatic truth here is that the classification has been made rather than the classification itself, which may or may not be true. Whereas many ontologies would use a class hierarchy here, with individuals being members of classes, such an approach is problematic. The assertion that an individual belongs to a class is a statement of universality whereas, in most cases, this is not necessarily true, or is only true within bounds and subject to change. Therefore, the POM supports both a) explicit assertion to describe some perceived reality and b) observed phenomena being bounded in spacetime rather than being universally true.

Expanding further on the concept of states and changes of state, the nature of ontological truth becomes more apparent. In many conceptual models, characteristics of entities are described using attributes or properties of said entities. There are two issues with this. Firstly, this does not account for the temporal nature of many characteristics. Secondly, this does not account for the way in which any description of any characteristic is not a statement of fact but rather a description of some observed reality, as described above. The colour of a thing is a good example of this. Colour is rarely an intrinsic property, but can be changed. Equally, assessment of colour is highly subjective and even if measured using a scientific instrument such as a colorimeter, it is only true that the measured value(s) are an assertion of an approximation of the true value. Indeed, all measurements can be seen as a specialism of observation where some approximation of some true state is produced, constrained by the accuracy and precision of the mode of measurement. Length is another good example of this; it is undoubtedly true that a physical object has a length dimension in the real world, but any measurement of this is constrained by the process of measurement and must be presented according to the levels of accuracy and precision afforded by the measurement process. Furthermore, the length of an object is not an intrinsic property of the object but it is a state subject to change, for example due to environmental conditions.

Regarding changes of state, the general principal holds that an event may change a state or states of things. Events are therefore central to the dynamism of the model. Taking the example of *Production*, the *Production* pattern is a specialism of a more generic *Creation* pattern in which some *Event* changes the existential state of some thing; In a creation Event, some thing comes into existence by some means, natural or anthropogenic whilst in a *Production* event, a specialism of the *Event* class (ie subclass) some *Agent* is actively involved in this activity. So we can describe the act of producing an *Observation* (through an *Observation Event*) or *Measurement* (through a *Measurement Event*) using exactly the same design pattern as the manufacture of a car, invention of a mechanism or conceptualisation of a design.

This event driven approach to the ontology underpins the POM. In the core ontology, *Observations* describe interrelated spatio-temporal things (*Phenomena*), which is very much aligned to the principles embodied in ABI; as with ABI, it is events, activities and other happenings that are very much the focus. Furthermore, the application of logical reasoning to the Description Logic based model can be further extended to take advantage of other forms of logical and statistical inference and reasoning. For example, ongoing work to investigate causal inference and forms of Statistical Relational Learning (SRL) applied to the sequences of phenomena, alongside Bayesian methods, such forms of analysis can be seen to align closely with event-driven models. Work has already been undertaken to investigate how POM-like models can be transformed into vector representations to further facilitate ML based inference and reasoning [26]. The concept of *Agent* discussed previously is also important in this regard; as with the Fried Of A Friend (FOAF) project specification of *Agent* [27], any such machine based inference engine or reasoner is treated within the POM as an *Agent*. Therefore capable of making *Observations*, with the result that the outputs of any such algorithm or system become part of the knowledge encapsulated in the KG as further *Observations* based on *Observations*, leveraging the W3C Provenance Ontology [28] .

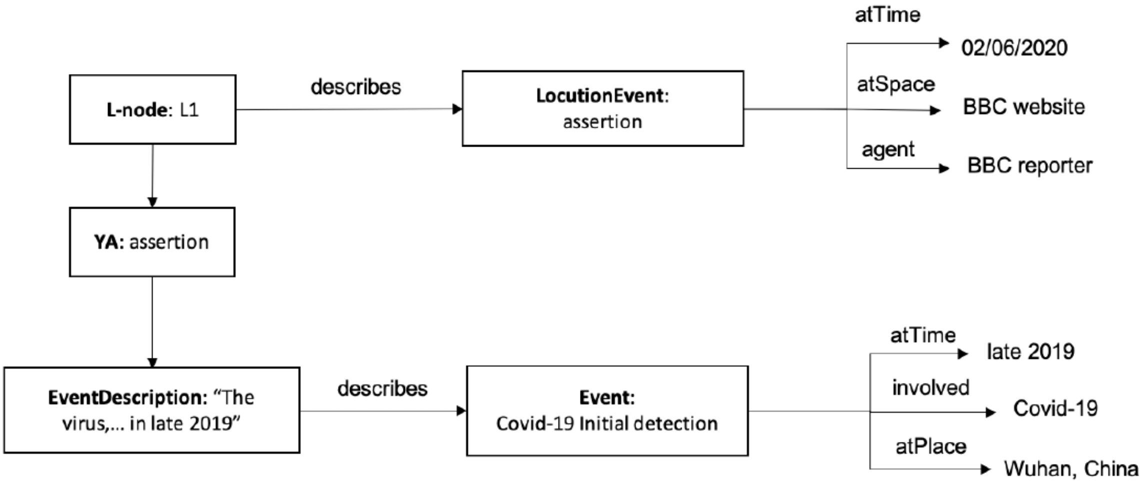
Figure 6: a schematic view of part of the Knowledge Graph based on an early version of the POM. This shows two Observations of States of an object which are then used as argument nodes as part of a hypothesis which can then be evaluated, in this case to determine which of the two perceived realities is more likely



Note also further work undertaken on the way in which *Observations* may be generated as the product of some hypothesis. This approach leverages the Observation pattern to be applied to the products of indirect, as well as direct, observation. To achieve this, the KG underpinned by the POM ontology is aligned to a separate graph based on the Argument Interchange Format ontology (AIF) used to describe individual hypotheses and understandings in terms of arguments based on evidence (Figure 6).

The articulation between the observational ontology and the argumentation ontology allows *Observation* nodes on the former KG to be used as *Evidence* nodes in the latter KG (Figure 7). This articulation makes use of the way in which the AIF ontology already supports disambiguation between events and locution events [28]. The use of Description Logic based ontologies across both models supports reasoning to assist

Figure 7: Part of the articulation between the AIF ontology and an event driven ontology, showing how an event and an event which produces an observation describing said event are represented, derived from a Covid-19 pandemic example.



with inference [29], search and discovery [30], and hypothesis generation, evaluation and explanation [32].

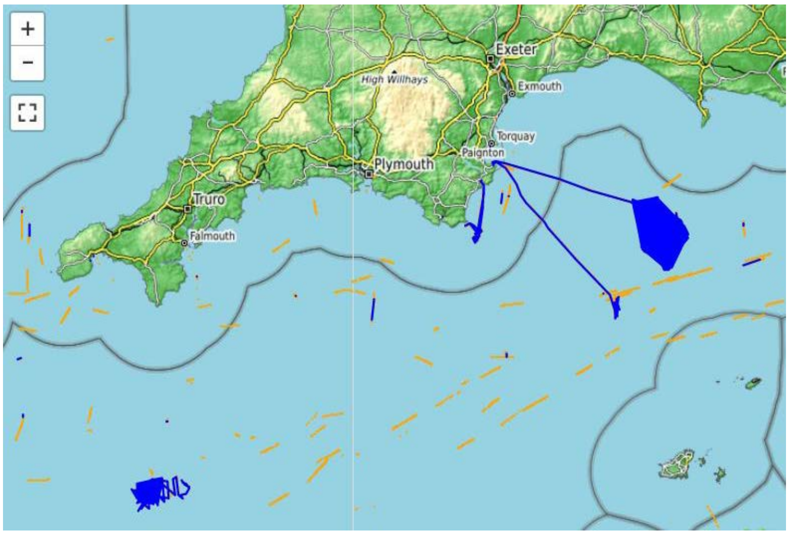
4.6 Use Case: Maritime Situational Awareness

The architecture and conceptual framework described above has, to date, been applied to two domains covering three use cases with a further domain in progress. Selected to be diverse and show the efficacy of the approach and not be restricted to particular kinds of information or problem, but can be generalised. Maritime situational awareness use case is summarise below as an example.

The first stage of work was to develop the core ontology based on prior experimentation. An activity we called OPIS (meaning "description" in Polish) and helped build enough of the core ontology to support the kinds of entities and phenomena needed to describe observations of phenomena [24]. Similarly, the domain ontology and reference components were built to provide sufficient vocabulary to describe the kinds of behaviours exhibited by ships engaged in fishing in the English Channel.

The data used were derived from Automatic Identification System (AIS) data, available commercially. This is by its nature is very simple in form, with observations of position being the primary set of observations generated. From this, a reasoner based on Versatile Event Logic (VEL) [23] was used to infer higher order kinds of movement and these were then used to infer the presence of behaviours of interest, for example trawling activities (Figure 8).

Figure 8: trawling events (shown in blue) inferred from AIS data (shown in orange) using a Knowledge Graph based on the OPIS ontology. These events were identified using logical reasoning given a set of characteristics describing trawling behaviour (ie a pattern) expressed using Versatile Event Logic.



A related piece of work explored how an OPIS-like KG might be populated automatically, using the ontology to support fusion of information from different sources [25]. This piece of work, called EVELLO ("to pluck" in latin), focussed on how to extract semantic descriptions of events to populate a KG; the concept of Semantic Event Extraction can be seen as a subset of the broader topic of Automated Content Extraction and the EVELLO demonstrator made use of Natural Language Processing (NLP) and Image Feature Extraction (IFE) as part of the ensemble processor (Figure 9). This has fused textual information with information extracted from video feeds to produce an observation describing events where ships interacted in the vicinity of a harbour.

Another piece of work investigated how vector based methods can be used to analyse KGs with POM like structures. This work we called VECTOR and, working with an early version of the OPIS ontologies and KGs,it explored how methods for embedding rich semantic graphs into vector space can support vector based methods, for example to undertake subgraph pattern matching [26].

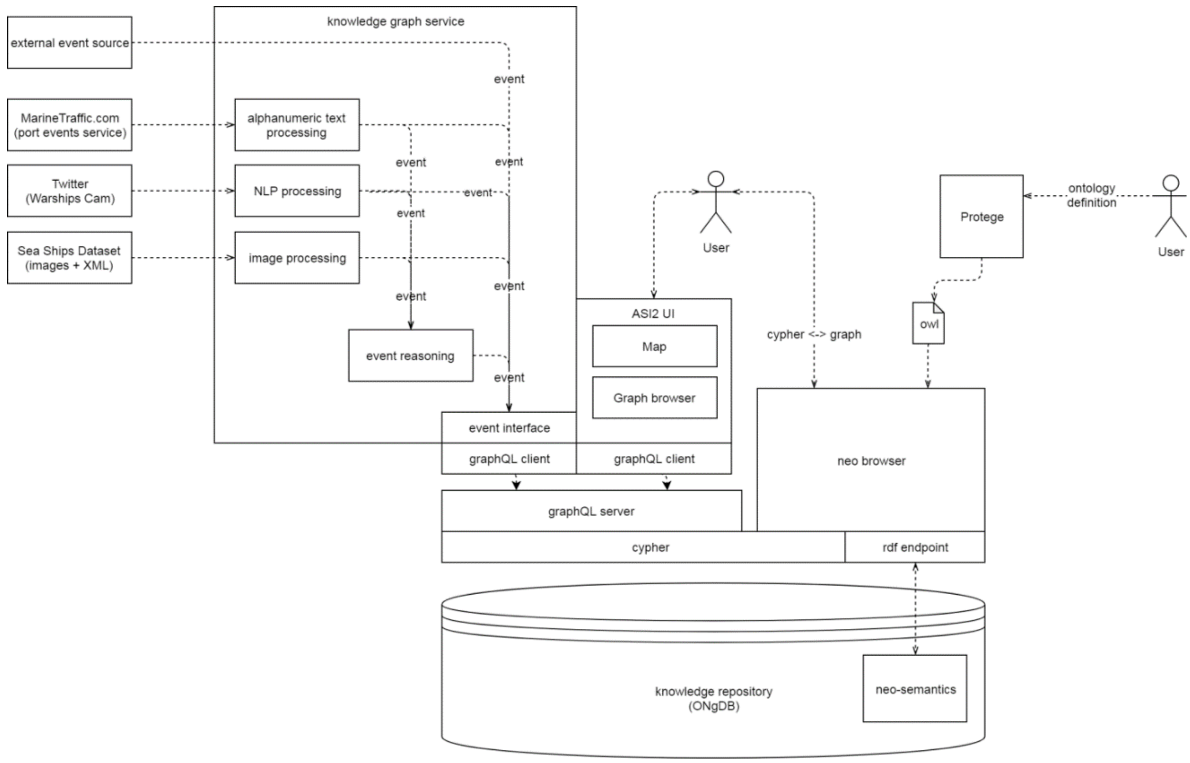
Together, these pieces of work investigated the three main subjects of interest with respect to the use of KGs for defence. Firstly, how to develop forms of ontology which can best support fusion understanding and prediction. Secondly, how to populate KGs based on these forms of ontology. Thirdly, accepting the need to be able to transform knowledge into forms suitable for different forms of analysis, how can forms of knowledge representation support novel approaches to inference and reasoning.

5. CONCLUSIONS AND RECOMMENDATIONS

If we are to achieve effective Human Machine Teams (HMTs) and teaming, and be able to support future generations of AI, it is vital that the forms of KR developed are semantically robust, unambiguous and can support the sharing of information without loss of meaning. The need for semantic clarity is particularly important and while the ultimate aim is arguably for machines to be able to learn context and be able to apply this knowledge, this is currently not possible so there is a need for forms of KR to be able to provide semantic 'glue'.

This paper has described a foundational approach for KRR and its potential application to defence and security operations, such as in hybrid warfare. It builds on underpinning research from GT, KGs, web/data sciences, AI and ML towards the development of a KRR architectural framework. The overall purpose, and the foundational shift, being the bi-directional understanding of information and knowledge by humans and machines and fusion, including as teams. To provide semantic clarity, KGs are developed that are semantically robust, unambiguous and can support the sharing of information without loss of meaning. As a result to overall develop the building blocks for the (semi)automated fusion or KRR against observations, from multiple sources and across organisational boundaries, to support the goals of high-level fusion, at scale and at pace.

Figure 9: the EVELLO architecture used for Semantic Event Extraction, an example of a self-contained, orchestrated suite of fusion and reasoning tools behaving as an Agent capable of populating a POM based KG



A foundational step towards this has been the design of a layered, modular semantic framework and its key components, along with example of exploratory 'use cases' outlined in the paper. The components detailed include core ontologies, to cover all concepts that are stable and common to all use cases, and domain extensions, to provide coverage of more specialist domains with further refinements as needed (i.e. the trunk, branches and leaves respectively of a tree, analogy described earlier). Some other conceptual innovations added include Structured Observation Management, observations used to generate reporting based around entities or objects in a process of Object Based Production to support analysis focussed on activities in which these entities participate, such as Activity Based Intelligence. All these were captured within the Phenomenological Observation Model (POM) we developed. A conceptual model, formal ontology, to address domain dynamics and spatio-temporal phenomena, real world events/happenings involving entities of interest. A model more closely aligned to how humans understand and make sense of the world and hence more amenable to human machine interactions and teaming.

Use case examples from maritime situational awareness, pharmaceutical production, and early work on information operations; show our approach to be on the right track. Some important limitations uncovered so far that need to be addressed include:

1. Socio-technical issues:- development and implementation of the framework and components:

    - Despite hiding computational complexity, the approach requires Suitably Qualified and Experienced Personnel (SQEP) in web/data sciences that is limited in Defence and defence suppliers

2. Information system issues:

    - Information sharing and assurance across domains and organisations (e.g. intellectual property, bespoke developments, different security policies etc.)

    - Federation, scaling, storage and performance of distributed systems within organisations as well as across organisations (coalitions)

3. Science and technology (S&T) issues:

    - Three specific technical areas require much further research and deliberations – Causality, Uncertainty and Argumentation. The first is needed to extract the causal links – the actual DNA of a phenomena, the second to manage quantitative analysis of qualitative/subjective information and assertions, and the third that we can implement logic in support of optimisation of choices in our decision making processes

In summary, we have shown development of a layered, modular semantic framework for defence analysts as an enabler for integration of heterogeneous observations to support Situational Awareness (SA) and Situational Understanding (SU). In particular, SA and SU across an entire operation, enabling more effective HMTs at all levels of command and in multi-national coalition operations (e.g. hybrid warfare). Such a framework can support the coalition to:

i. develop domain specific ontologies and event driven architectures

ii. assess the inference capabilities, including, in future, handling of uncertainty/ambiguity

REFERENCES

[1] N. Biggs, “Algebraic Graph Theory”, Book, Cambridge University Press 1993

[2] Bondy. J. A. and U. S. R. Murty, “Graph Theory with applications”, Book, Elsevier Science Publishing Co., ISBN O~444-19451-7, 1976

[3] M. E. Newman, “Networks An introduction”, Book, Oxford University Press, ISBN 978-0-19-920665-0, 2010

[4] D. Stenmark, "Information vs. knowledge: the role of intranets in knowledge management," Proceedings of the 35th Annual Hawaii International Conference on System Sciences, Big Island, HI, pp. 928-937, 2002

[5] Nonaka. I. and H. Takeuchi, “The knowledge-creating company”, Book, Oxford University Press, 1995.

[6] Gruber, T. R., “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”, Int. J. Hum. Comput. Stud., 43, 907–928, 1995

[7] Guarino, N., Oberle, D. and Staab, S. “What is an ontology?”, 2nd ed. in Staab, S. and Studer, R. (eds), Handbook on Ontologies, Berlin, Springer, pp. 1–17. DOI: 10.1007/978-3-540-92673-3, 2009

[8] Asim, M.-N., et al. “A survey of ontology learning techniques and applications”. Database, Vol. 2018: article ID bay101; doi:10.1093/database/bay101, 2018

[9] Feilmayr. C. and W. Wöß. “An Analysis of Ontologies and their Success Factors for Application to Business”. Data & Knowledge Engineering, 101:1-23, 2016

[10] Maedche. A. and S. Staab. “Ontology learning for the semantic web”. IEEE Intell. Syst., 16, 72–79, 2001

[11] Berners-Lee. T. 1998. “Semantic Web Road Map”. W3C. https://www.w3.org/DesignIssues/Semantic.html

[12] Ehrlinger. L. and and W. Wöß, “Towards a Definition of Knowledge Graphs”, SEMANTICS 2016: Posters and Demos Track, Leipzig, Germany, September 13-14, 2016

[13] Shaoxiong Ji et al, “A Survey on Knowledge Graphs: Representation, Acquisition and Applications”, JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015

[14] NATO-RTG-144, “Content Based Multi-Media Analytics (CBMA)”, NATO, STO Technical Report TR-IST-144, ISBN 978-92-837-2301-1, October, 2020 (download from https://www.sto.nato.int/publications/STO%20Technical%20Reports/STO-TR-IST-144/$$TR-IST-144-ALL.pdf)

[15] NATO IST-ET-111, “Knowledge representation and reasoning – a review of the state of the art and future opportunities”, Pre-release, NATO, STO Technical Report, STO-TR-IST-ET-111, ISBN 978-92-837-2342-4, March, 2021

[16] Bechhofer. S & Miles. A. “Using OWL and SKOS”. W3C. https://www.w3.org/2006/07/SWD/SKOS/skos-and-owl/master.html. 2008.

[17] Janowicz, K., and Gangemi, A.. “Ontology Engineering with Ontology Design Patterns: Foundations and Applications.” Germany, IOS Press. 2016.

[18] Le Boeuf, P., Doerr, M., Ore, C. E., Stead, S., Aalberg, T., Balzer, D., Bekiari, C., Boudouri, L., Crofts, N., Eide, Ø., Gill, T., Goerz, G., Hagedorn-saupe, M., Hiebel, G., Inkari, J., Iorizzo, D., Kotipelto, J., Krause, S., Lampe, K. H., Lindenthal, J., Nyman, M., Riva, P., Rold, L., Smiraglia, R., Stein, R., Stiff, M. and Žumer, M. "Definition of the CIDOC Conceptual Reference Model." ICS Forth. 2012.

[19] Cripps, P. "Places, People, Events and Stuff; building blocks for archaeological information systems" in Earl, G., Sly, T., Chrysanthi, A., Murrieta-Flores, P., Papadopoulos, C., Romanowska, I., and Wheatley, D. (eds), Archaeology in the Digital Era Volume II: e-Papers from the 40th Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, 26-29 March 2012, Amsterdam, Amsterdam University Press, pp. 487–497 [Online]. DOI: oai:ARNO:500. 2014.

[20] NGA. "Geospatial Intelligence (GEOINT) Basic Doctrine." v1.0. 2018.

[21] Battle, R. & Kolas, D. 'GeoSPARQL: Enabling a Geospatial Semantic Web', Semantic Web Journal, vol. 0, no. 0, pp. 1–17. 2012.

[22] Little. C. & Cox. S. "Time Ontology in OWL." W3C. 2006.

[23] Bennett. B & Galton. A. "A Versatile Representation for Time and Events". In Fifth Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense 2001), pp43-52. 2001.

[24] Marshall, S., Wood. M., Bennett. B., Cohn. A. "Opis Foundational and Extensible Descriptive Models." Frazer Nash Consultancy. DSTLX-1000140662. 2020.

[25] King. P & Strickson. B. "Evello". BAE Systems AI Labs ref 4184-1.1. 2020.

[26] Millar, D., Braines, D., D'Arcy, L., Barclay, I., Summers-Stay, D., & Cripps, P. "Embedding dynamic knowledge graphs based on observational ontologies in semantic vector spaces." Proceedings Volume 11746, Artificial Intellig. and Machine Learning for Multi-Domain Operations Applications III; 117461O (2021) https://doi.org/10.1117/12.2585888. 2021.

[27] Brickley. D & Miller. L. "FOAF Vocabulary Specification v0.99" http://xmlns.com/foaf/spec/. 2014

[28] Duthie, R., Zografistou, D., Reed, C., Snaith, M., Visser, J. "Project Dedwi U2 Deliverable D1: Knowledge Repository". Dstl. ITT/ R1000124743. 2020.

[29] Duthie, R., Lawrence, J., Reed, C., Visser, J., Zografistou, D. "Project Dedwi U2 Deliverable D2: Inference over Knowledge Graphs". Dstl. ITT/ R1000124743. 2021.

[30] Duthie, R., Reed, C., Visser, J., Zografistou, D. "Project Dedwi U2 Deliverable D3a: Search". Dstl. ITT/ R1000124743. 2020.

[31] Duthie, R., Lawrence, J., Reed, C., Visser, J., Zografistou, D. "Project Dedwi U2 Deliverable D3b: AO-enhanced Knowledge Graph Analytics" Dstl. ITT/ R1000124743. 2020.

[32] Duthie, R., Lawrence, J., Reed, C., Visser, J., Zografistou, D. "Project Dedwi U2 Deliverable D4: Explanation Generation Service". Dstl. ITT/ R1000124743. 2021.