Ontology Alignment in Geographical Hard-Soft Information Fusion Systems

Erik P. Blasch  
Defence R&D Canada-Valcartier  
2459 Pie-XI Blvd. North  
Québec City, QC G3J 1X5  
erik.blasch@drdc-rddc.gc.ca

Eric Dorion  
Defence R&D Canada-Valcartier  
2459 Pie-XI Blvd. North  
Québec City, QC G3J 1X5  
eric.dorion@drdc-rddc.gc.ca

Pierre Valin, Eloi Bosse, Jean Roy  
Defence R&D Canada-Valcartier  
2459 Pie-XI Blvd. North  
Québec City, QC G3J 1X5  
eloi.bosse@drdc-rddc.gc.ca

Abstract – Information fusion exists over many forms of hard data (e.g. from physical sensors) and soft data (e.g. from human reports) to interpret observations of real-world objects. As demonstrated from the Geographical Information Systems (GIS) community, there is a growing need for the linking and alignment of both (1) exploited physical imagery products and (2) derived ontological textual labels (semantic markup). Semantic markup can be done on both exploited data (e.g. automated image segmentation), as well as user reports (e.g. weather forecasts). Since the derived information is collected, stored, and displayed into distinct ontological structures by different agencies, ontological alignment is thus required whenever the semantic information is paired with distinct real-world imagery observations. In this paper, we explore issues of fusing hard and soft data as related to ontology alignment. A maritime domain situational awareness example with geographical imagery and textual ontologies is shown to demonstrate the need for ontology alignment to assist users for pragmatic surveillance.

Keywords: Fusion, Situational Assessment, hard-soft fusion, ontology alignment

1 Introduction

The goal of this paper is to put forth issues in semantic and ontological alignment of geographical information fusion system ontologies over hard (e.g. images) and soft data products (e.g. textual labels derived from semantic markup). Level 2 fusion (situational assessment) of the Data Fusion Information Group model [1] incorporates both hard and soft data for user awareness. [2] Ontologies for information fusion include a shared conceptualization and awareness of a specific domain vocabulary. Using semantic vocabulary, properties and attributes can be reasoned over for object relationships and situation event relations. Issues in semantic content for a sensor exploitation system include (a) ontology standardization [3] - multiple semantic labels can be attributed to the same information, (b) ontology alignment [4] - when there is not a shared vocabulary a linking is required, and (c) ontology analysis [5] – same, incomplete, or imprecise labels for different information. By understanding ontology development, effective information fusion could be conducted over the hard and soft data.

With development efforts underway for fusing hard (physics-based) and soft (user-based) information, a good example is Geographical Information Systems (GIS). GIS require ontology standardizations, or a mapping between both hard and soft relations among sensors and their exploited products. A GIS hard-soft exploitation product example is scene labeling, a semantic markup process, to highlight geographical features (e.g. river). Using hard (image) and the soft (labels) also affords display fusion, where information is overlaid such as object detections and road information. Together, GIS hard-soft fusion and display fusion (alignment) assist in situation assessment/awareness (SA).

For future information fusion tasks, such as maritime domain awareness and surveillance, there will be information databases with various products (e.g. GIS, ship traffic, Automatic Information System (AIS) data, and business transactions). The exploited geographical information will be generated and annotated with different ontological labels due to the collection agency, distinct data standards, and storage and retrieval queries. For example, if multiple government organizations (i.e. coast guard, homeland security, commercial shipping) are supplying data to the maritime database (e.g. tourist trips, fishing charters, recreational outings, and general weather reports); then a host of data is available for SA.

Specifically, we are interested in a situation of tracking and identifying a vessel in the waterway, using multiple hard/soft databases to determine the vessels activity and affiliation. Unlike air-traffic control where all aircraft file a flight plan, the coastal waterway traffic does not require everyone to list their activities before venturing into the waterway. Thus, confusion can result from distinguishing legal activity, illegal activity, and normal commercial or recreational activity. However, this complex information product can only be formed through fusion of information elements scattered and disseminated over several hard/soft databases. The fusion process thus relies on appropriate ontological alignments between those hard/soft databases.

1.1 Signs, Ontologies and Information Fusion

Ontologies consist of structured sets of classes, attributes and relations meant to convey meaning. The study of how things gain or convey meaning is the subject of Semiotics.
Since Morris [6], Semiotics is segmented into three components:

- **Syntactics**: The rules that govern how signs relate to one another in formal structures
- **Semantics**: Relation between signs and the concepts they denote (their denotata)
- **Pragmatics**: The systematic study of the use of signs.

This begs a definition for the word *sign* which is a triadic relationship between a *representamen* [7] (the actual labels referred to in this paper), an *interpretant* (purely conceptual) and the *referent* (the actual object referred to by the interpretant and to which the representamen stands for. This trichotomy (Figure 1) was depicted by Ogden and Richards [8] into what is known today as the *Semiotic Triangle*.

![Figure 1. Ogden and Richards’ semiotic triangle.](image)

The triadic relation of a sign allows for the explanation of several meanings attributed to a same pair of symbol–referent. Obviously, information systems manipulate symbolic information and not their associated interpretant nor referent. The action of *semantic labeling* (or creating soft data) aims at aligning heterogeneous data specifications (or ontologies) in that some representamen in hard data are declared *equivalent* in other hard/soft sources. Ontological alignment of hard/soft data sources plays a vital role in the ability to conduct data fusion.

From a semiotic perspective, the act of putting hard and soft data into correspondence (relationship of equivalence) is one that seeks the *representamen* in both sources that denote the same concepts that in turn refer to the same objects. This activity, from now on referred to as *ontological alignment*, is conditioned by the rules that govern how signs related to one another (syntactics), the typical meaning(s) signs convey (semantics) and the meaning(s) signs gain in their usage context (pragmatics). Our goal in designing decision-support systems is not only to correctly “label” information, but also to pragmatically create the ontology that best aids a user in decision analysis [9] and achieve ontological alignment when gathering information from hard and soft data sources. Determining the ontology content, as derived from imagery and user knowledge sources, includes extracting soft (textual) labels from hard (imagery) data.

### 1.2 Hard-Soft fusion

In *Principles of Data Fusion Automation*, [10] Anthony discusses the spatial class indexing structure as mapped to a semantic feature representation (Ch. 08) for database requirements. Semantic labeling for databases is important for compact storage, processing efficiency, and timely and useful information retrieval. Inherent in Anthony’s discussion is a hard-soft relation’s ontology for indexing. Anthony continues to develop database products augmenting hard-sensed data with soft HUMINT data [11] and soft target analysis [12].

Hard-soft fusion (HSF) was brought to light as a research topic during the Fusion08 conference. Hall and Llinas [13] proposed an initial framework for HSF while Pravia et. al. [14], demonstrated an initial data set for HSF. Other discussions focused on contextual data [15], linguistic data [16], and soft-message timing [17]. The information community continues to explore the needs of HSF. Other domains, such as Geospatial Information Systems (GIS) community work to explore ontology development and fusion from various products which serve as useful examples of hard/soft information fusion.

### 1.3 Hard-Soft Ontology Fusion Example

Three types of hard and soft geographical data fusion are

1. fusion of hard-to-soft products (as shown in the top of Figure 2) and
2. fusion of hard data (such as imagery) with soft data extraction from the image set (shown in the bottom of Figure 2), or
3. hard imagery data fusion combined with soft ontological data fusion (shown at middle of Figure 2). Typically, hard data comes from sensing sources, such as imagery. The imagery is exploited to determine class labels (such as forests and urban areas) which are inherently soft data as defined by user definitions of terrain. As shown in Figure 2(a), waterway and ship signature soft labels are created by exploiting imagery data. The exploited products would need to be spatially and temporally aligned to create a fused product for users such as overlapping the two labeled images on top of each other.

Additionally, imagery or video can be registered for soft data extraction such as target tracking and identification as shown by Figure 2(c). However, the soft data extraction utilizes only a single semantic ontology.

Typically, sensors and systems collecting observations are designed by individual companies, forcing an integrated hard-soft product alignment over ontologies and imagery - shown in Figure 2(b). For interoperability reasons, it is best to have multiple forms of image registration and fusion, ontology alignment and fusion, as well as hard/soft calibration for various coordinated references between database information.
The rest of the paper is as follows. Section 2 overviews the ontology applications with emphasis on background research. Section 3 highlights developments in Geographical Information Systems where research is underway in hard/soft fusion. Section 4 demonstrates an example for hard/soft ontological fusion for a maritime example. Section 5 includes discussion and a conclusion.

2 Background

This paper explores issues associated with hard-soft ontology alignment as it pertains to geographical information systems (GIS). Since the 80’s, the Artificial Intelligence (AI) community has been exploring ontological developments building on earlier philosophical concepts.

2.1 Ontology Research for Situational Awareness

The use of ontologies in information fusion started from a need to explore level 1 fusion (object recognition) to provide level 2 fusion (situational awareness). The integration of information fusion ontologies was motivated by research in situational awareness (SA) from target recognition [20]. Kokar and Matheus combined for multiple papers on developing relations [21], associations [22], and a high-level fusion SA assistant [23, 24]. Recent work is on SA for command and control [25], software architectures [26, 27], and decision support and situation management [28]. These systems are designed for a specific application; however, with multiple systems arises the need to align the respective ontologies to enable systems interoperability.

Interoperability issues are exacerbated by the multiple systems required to communicate with different terminology, software architectures, and users. Examples of such issues are met in geographical information systems [29], the intelligence community [30], agent communication [31], and net-centric warfare [32]. Current developments in other communities include biomedical research [33] and web services [34]. These communities have a plethora of articles describing examples of interoperability topics that could be leveraged by the information fusion community.

2.2 Information Fusion Ontology Research

Ontology representations (i.e. semantic content labels) support user-machine interactions [35, 36]. The synergy between the ontology community using data fusion [37, 38] and the information fusion looking at ontology issues in high-level fusion became popular in 2003 [39, 40]. Applications from the fusion community demonstrated the need for ontology investigations.

Kokar [41] provided details of ontology need in SA scenarios. For a surveillance task, it is important to determine the entities of the situation. When a target is not tracked, and there is no communication, it becomes a potential threat. Little [42], developed a threat ontology based on the target intent, capability, and opportunity.

---

Table 1: Ontologies in AKTiveSA (Smart [18])

<table>
<thead>
<tr>
<th>Hard</th>
<th>Soft</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geography</td>
<td>Humanitarian aid</td>
<td>Geo-political,</td>
</tr>
<tr>
<td>Meteorology</td>
<td>Information sources</td>
<td>Military,</td>
</tr>
<tr>
<td>Datalink</td>
<td>Agent</td>
<td>Organization,</td>
</tr>
<tr>
<td>Weapons</td>
<td>Transportation</td>
<td>Terrorism</td>
</tr>
</tbody>
</table>

While the listing of ontologies {geography, transportation, humanitarian aid, meteorology, information sources, geopolitical, military, datalink, equipment, knowledge system, agent, communication device, weapons, organization, terrorism} could be grouped differently, they are examples of hard/soft fusion. For example, two relevant Ontologies are geography and traffic routes:

1. Geography Ontology: encompasses a wide variety of conceptualizations including terrain features, transport routes, rivers, shorelines, terrain elevation data, etc.
2. Transportation Ontology: overlaps, to some extent, with the geography ontology in the sense that transportation routes, e.g. airways and roads, may also be considered elements of the geographical (geo-spatial) domain. [18]

For a maritime domain awareness scenario, the geography ontology includes the ocean, coastline, forest, and mountains (hard sensed boundaries) while the traffic ontology covers shipping lanes, low and high traffic areas for commercial and recreational systems (soft labels). For exploited geography information, class (semantic labeling) segmentation can be performed to garner soft relations from the hard data. To ensure the robustness of such systems, we need to explore issues in hard/soft ontology alignment and fusion.
Boury-Brisset [43] detailed a goal-driven threat assessment ontology example for weapon assignment. Laskey et al. [44] utilized probabilistic methods for ontologies. Other examples include data sources [45] and scene understanding [46].

2.3 Ontology for Surveillance Research

With the many web services and users available, there is a need to extend the methods of command and control for surveillance. Roy [47] has been developing methods for knowledge-based systems for maritime awareness. Such issues as tracking, situation analysis, and threat assessment, have been included in the efforts to increase techniques for maritime-domain awareness. Examples include anomaly detection [48, 49] and alerting [50].

Other developments include net-centric warfare for enterprise systems that include intelligence and surveillance data as well as cyber information. Examples include knowledge-based mission analysis [51], semi-automated ontology development [52], and semantic-oriented services architectures [53, 54].

With various applications researching ontology developments, the GIS community developments are similar to that of hard-soft ontology information fusion.

3 Geospatial Ontologies

Ontologies have been developed for web-service products. Figure 3 illustrates an example of high-level fusion [55, 56, 57] needs as the user has many databases available. Determining what is needed is as important as how good it is as the user requests decision-quality information at the correct time.

![Figure 3. Information Fusion Decision Quality.](image)

In a typical harbor scenario, imagery is taken and regions are segmented and labeled (semantic markup). The quality of the collection and labeling could vary. For example, the tide comes and goes, so the coastline is not always well defined. Sensor collection resolutions and communication delay can vary, altering the segmentation or producing uncertainty in the labeling.

An example of a land-based ontology fusion is provided by Duckham and Worboys [5] where they use soft data labeling from one data set of forest and built-up areas and another data set with woodland and urban areas. They show how different semantic labels from distinct databases can be equated through imagery spatial alignment.

There are many issues to address in hard-soft ontology fusion. Information fusion can address issues of increased spatial coverage, uncertainty reduction, and data aggregation to aid a user in decision making. One example from GIS is to overlay GPS traffic information reports on maps such as tracking targets on roads [58] or group tracking [59].

The inductive nature of extracting information from GIS data to infer the situation has many issues associated with the uncertainty. Duckham and Worboys [5] list many categories to be considered:

- **Semantic alignment:** whether or not the labeling correspond to meaningful categories of alignment.
- **Inaccuracy of information:** the segmentation of the geographical boundaries
- **Imprecision:** the representation of the segmented areas
- **Vagueness:** uncertainty in the boundaries
- **Uncertainty:** the stochastic randomness in the collection boundaries.

**Geo-pragmatics** [60] enhances ontologies by establishing context for users to reason over the data. Using “situated concepts” to help guide ontology development, both method-driven and knowledge (epistemic-driven) determine the dimensions of origin, use, and effects of information for humans and machines.

From the GIS point-of-view, there are many areas in which the general techniques of information fusion can aid a system such as image registration, estimation, and data management; however when labels are created from class structures, there is a need for semantic and syntactic ontology alignment. The ontology alignment is consistent with the information community looking into the developments of hard-soft fusion.

4 Hard-Soft Ontology Example

We use the coastal surveillance example to explain the issues in hard-soft ontological fusion alignment for a geospatial information system. The goal is to provide useful information to a coastal surveillance group and the data products include the imagery and database of vessels. By aligning (or matching) the soft object labels with the hard pixels, we can assist in situation awareness. Detecting an anomaly in routine traffic would aid in the determination of a terrorist activity.

4.1 Ship Automatic Identification System

The Automatic Identification System (AIS) [61] provides valuable hard and soft data about a ship. Hard data available for ship information includes: time, position by GPS, heading, course over ground, speed, and rate of turn. The soft data would include the vessel identification, radio call sign (origin), name, type, size, and destination and estimated arrival. Since the AIS is not available on all

---

1 Valuable comments provided by the ISIF reviewers
ships (or is only required for commercial vessels), it is mainly used for collision avoidance, navigation, and/or search and rescue.

The AIS is an example of cooperative information as the ship intentional supplies its information. However, there are cases for unintentionally cooperative information, and non-cooperative information. [62] The unintentionally cooperative case is if some AIS information is supplied or it is falsified. The non-cooperative case includes ships in which they choose not to supply the AIS information.

Additionally, it is noted that the AIS system can be linked to the GIS in privately operated services. [61] While the International Maritime Organization (IMO) Maritime Safety Committee condemned these products for security reasons, it is still a concern for ontology alignment as different products might differently categorize (especially for the case of soft data) vessel information.

To account for the cases in which the AIS data is not supplied by the vessels of interest, we seek to utilize GIS information to determine the vessel behavior for maritime situational awareness.

4.2 Situation Awareness Example

To determine mission effectiveness, we need to consider information gain, quality, and robustness. Our example comes from a need to protect and address user needs to protect a coast area [63, 64, 65]. Our example comes from the CanCoastWatch (CCW) where Li, et. al. [66] looked at high-level data in a goal-driven net-enabled distributed data fusion system, shown in Figure 4.

![Figure 4. Example of Maritime Domain Awareness.](image)

In the Maritime Domain Awareness system, the goal is to provide search and coverage of possible activities. Parameters used in the search include: search area, time of last visit, detections from radar, sonar and detections from video, and communication. Low level SAR, EO/IR, and radar information provides the object track and identity (ID) information. The situational evidence is gathered from the commanders needs through goals. Observations provided the ability of a user to observe, orient, decide, and act over the track and ID information. Both the IFS and the user is reasoning over the situation and determining a level of confidence in the analysis.

The imagery data collected and developing-scenario databases provide resources from which hard and soft data ontologies can be developed. Linking the scenario ontologies requires ontology alignment, imagery registration, and hard/soft data coordination.

Wehn et. al., [67] looked at the decision function as related to planning and resource management. Goals were to evaluate the system effectiveness for net-enabled operations such as synchronization, scheduling, and search. Using the goals of HLF planning, effectiveness, and information content, we revisit the scenario with the general context from which to establish geographical and traffic soft ontologies from the sensed data.

In a cooperative scenario, detected objects can be confirmed; however for the non-cooperative case, the objects cannot be communicated with after detected. Effectiveness includes information gain [IG] (confidence of track and ID information), information quality [IQ] (data fusion coverage area), and robustness (whether or not the information is consistent, or needs multiple verifications). Robustness can be repeated observations or confirmed communications. Likewise, we are interested in the fusion of soft (textual ID labels) fusion with that of the imagery data.

4.3 Ontological Fusion Example

We assume that we have two systems that are collecting information on the maritime scenario. Both systems collect and label their products differently. We need to align (or match) the ontology representations of the physical objects to the situation for coordination with the image displays of the coastal areas.

First, we take a picture and determine the coastal areas. From the geographical information, we can determine what the “land areas” are and what the “waterways” for the vessel traffic are. As shown in Figure 5, by masking out the land, we have determined an initial ontology labeling through reasoning over the imagery. While images are passively collected in the scenario, further efforts would include active sensing to modify the imagery data for the mission requirements.

![Figure 5. Coastal Areas.](image)
Next, we mark up the images with various ontological labeling products. For example, if we have two data sets, one giving the commercial and recreational information and another giving the traffic information of shipping and tourist vessels’ pathways, we need to determine how to fuse the information. Figure 6 shows the case in which the two data ontology mappings need to be fused or aligned. The semantic labeling requires some user-defined mappings as to the coordination. One notional example is that the commercial traffic can be both shipping and tourists while recreational vessels are typically associated with tourist activities and shipping is a commercial event.

![Figure 6. Ontological Alignment Process](image)

There is an issue for anomaly detection in that activities that include shipping of illegal substances might be mapped as tourist activities, however the behavior and traffic of the vessels would not fall into the aligned normal traffic patterns.

From the exploited database information and fusion of hard/soft data, we can gather information on each product. The normal traffic pattern information can be derived from the instance information. Since the waterways demark a clear boundary, it was easy to determine the location of the shipping vessels. However, as shown in Figure 7, we have to make choices on the traffic patterns for the shipping and tourist activities. In this example, we assume that the shipping information comes from the sea to the large ports and that the tourist information is located in the coastal areas. While the exact mapping might be different, the exemplar case provides an ontology mapping for both the geographical and traffic ontologies.

![Figure 7. Fusion with DataSet B Instance Data.](image)

Likewise, we can gather the information from other products such as commercial and recreational vessels. Many times, the commercial vessels would be monitored so as to provide safe traffic to the ports. The requirement for many trans-ocean vessels would require large waterways clearly marked with traffic patterns, lighthouses, and other prescribed coast-guard standards. The recreational vehicles would utilize this information and plan their routes accordingly. For example of the data set A ontology, shown in Figure 8, we extend the coastline for the recreational vehicles (as they might be on short trips) and label the areas with the sea-lanes as the commercial information.

![Figure 8. Fusion with DataSet A Instance Data.](image)

From the geographical to ontology mappings in data set A and data set B, we see that the information is not a one-to-one mapping. Required is ontological alignment of the data. The alignment is both semantic and spatial. The overlay, shown in Figure 9, affords a map or territory of the data sets. However, the overlay of spatial information is not exactly a mapping for the semantic content. We can determine the alignment between the two with a spatial relationship as grounded with the GIS data. Aligning one ontology to another can be coordinated with the soft and hard information.

![Figure 9. Geographical Fusion Modeling](image)
Issues associated with the ontological alignment require efforts to combine traffic data that is gathered both from the vessels themselves (in a cooperative approach) and those detected from surveillance imagery (non-cooperative approach). The traffic data requires methods of tracking, information fusion, and situational awareness. Ontology alignment can be used to determine the vessel activity and identification by way of associated behaviors and class labels from historical data. For the case of anomalies, there would be required some threat ontologies, mappings of the uncertainties associated with the collected ontologies, and many other issues such as culture [68] we will explore in future research.

5 Conclusions

We explored the issues in hard-soft ontology development for geographical information fusion systems. For ontological alignment, many open questions still remain such as (1) how to deal with mismatch and mapping uncertainty in ontologies, (2) coordination between different standards for interoperability, and (3) how to fuse ontologies from both hard and soft ontologies. We presented a simple case for maritime domain awareness with emphasis on using geographical information system (GIS) products to account for cases where the Automatic Information System (AIS) data is not provided by the coastal traffic. The GIS community passively utilizes collected imagery for ontology extraction; however there is a need to balance and coordinate the soft ontology fusion with the hard data collection. Together, the integration of the hard-soft ontology development would aid in command and control, synergistic collection of surveillance information, and user interactions with the system to provide reasoning and understanding over the situation.

6 References


