Working with Extensional Ontology for Defence Applications

Ian Bailey[1], BEng, PhD, Chris Partridge[2], MA(Oxon)

[1]Model Futures Ltd. – www.modelfutures.com
[2]BORO Centre – http://www.boroprogram.org
Reviewers: Michael Warner (UK MOD) & Fariba Hozhabrafkan (INCOSE/Serco)

The UK Ministry of Defence has been examining the potential for using a 4D extensional Ontology for semantic integration and sharing of enterprise architectures. This paper outlines the findings from the MOD's initial exploratory work, including usage of the BORO Methodology, the IDEAS Group Ontology and associated pilot implementations.

***Index Terms*—Extensional Ontology, IDEAS, Geospatial, MOD**

I. Introduction

MOD's initial interest in 4D ontology came about through its membership of the IDEAS Group[1]. This is a consortium of national defence departments (Australia, Canada, Sweden, UK, US) whose purpose is to develop a common interchange mechanism for military enterprise architecture. The IDEAS Group has a mandate to analyse the national frameworks (DoDAF, MODAF, etc.), establish the common patterns they share, and design a mechanism for sharing architectural data. The IDEAS Group uses the BORO Methodology® to analyse the existing frameworks. This method produces a formal, 4D, extensional, higher-order ontology, which is how MOD began its interest in 4D ontologies.

Aside from the IDEAS efforts, MOD also needed a way to manage common reference information across the various enterprise architecture programmes under way throughout the department. This reference data covers types of equipment, organisations, locations, processes, etc. so is an ideal candidate for investigating the usefulness of ontology. It was decided to re-use the IDEAS ontology for this purpose, with extensions to cover UK-specific requirements.

The MOD has been leading the ontology development effort in IDEAS, but the benefits are beginning to be seen in other nations – for example, version 2.0 the US DoD Architecture Framework is to be underpinned by a conceptual model based on IDEAS. In addition, US contractors have developed pilot IDEAS interfaces to the Telelogic System Architect tool to enable exchange of process models with the other nations (see Figure 1). The UK has also conducted some pilot implementations of IDEAS. They have implemented an IDEAS interface to the Sparx™ EA tool – the US and UK can now freely exchange process models between these tools using RDF files conforming to the IDEAS Ontology.

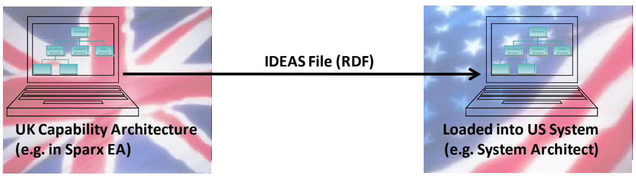

Figure 1 - Graphic Depicting Scenario for Enterprise Architecture Sharing

As well as EA applications, MOD has been keen to investigate broader uses for 4D ontology, and commissioned a demonstration application to integrate geo-political data from multiple sources. It is this work that forms the case-study for this paper. The problem being addressed was one of multiple naming conventions. NATO had ratified its STANAG 1059 for country codes. Most of these matched the ISO3166 coding scheme, but not all. In addition, NATO also had its MOE country codes for logistics purposes, and the US Govt standard is FIPS10-4. MOD wanted to investigate how ontology could be used to integrate these codification schemes and data from other sources (UK Govt Taxonomy and the CIA World Factbook).

II. The IDEAS Ontology

The ontology is stratified in order to maximise re-use of the standard for different purposes:

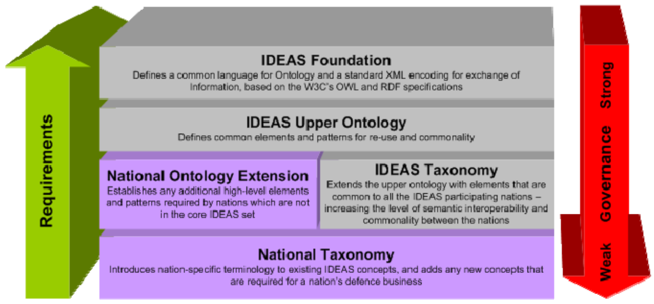

Figure 2 - The Layers of the IDEAS Ontology

The foundation layer defines the basic ontic categories. It is also at this level that the ontology is bound to the W3C RDFS and OWL specifications to enable exchange and publication of the ontology. The ontology is specified and managed in UML. However, the UML has been strictly profiled to match the ontic categories. The foundation elements are summarised in Figure 3. Note that all IDEAS elements and relationships are stereotyped (e.g. <<Type>>), and the application of the stereotype notation implies instance of.

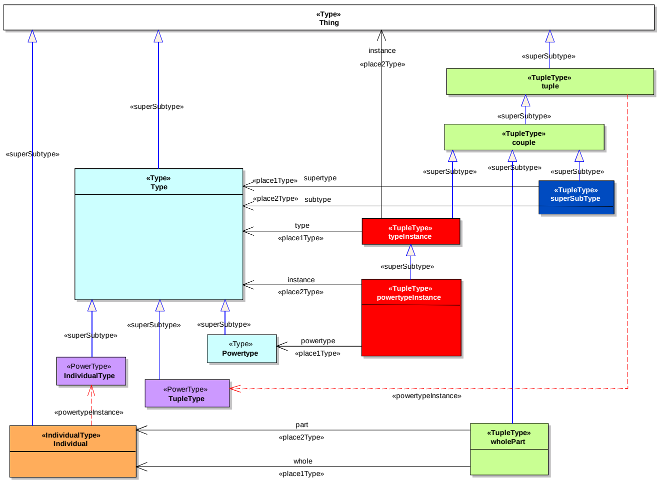

Figure 3 - The Key Foundation Elements in IDEAS

The IDEAS ontology is strictly extensional – individuals are identified by their spatio-temporal extent, and types (classes) are identified by their members. Tuples (relationships) are identified by their places (ends). The foundation defines some key powertypes[2] which are re-used across the ontology – the key ones being IndividualType and TupleType. Instances of IndividualType are Types whose members are Individuals. Instances of TupleType are Types whose members are Tuples. Type-Instance relationships are modelled using red UML dependencies, and Super-Subtype relationships as blue UML generalizations (the colour coding helps with crossing lines).

The IDEAS Upper Ontology level introduces common patterns (overlaps, intentional construction, before-after, set operations, etc.) and subject areas (processes, agents, information elements, etc.). The Upper Ontology also models the key architectural elements and views in national architecture frameworks.

III. The BORO Methodology®

In setting out to develop an ontology for data sharing and integration, it is important to ensure that all parties use the same approach to ensuring the intended interpretation is preserved when mapping legacy information onto the ontology. Without this assurance, there cannot be any integration or sharing at a semantic level[1]. Without formal methods of analysis, even two experts are unlikely to come up with the same ontology for the same domain. This level of unreliability and lock-in of expertise isn't acceptable to IDEAS, so the modelling team cast around for a suitably formal method. The IDEAS Group adopted BORO, not because they wanted to develop an ontology, but because they needed a defensible, solid method that would deliver repeatable results in the hands of any trained analyst.

BORO stands for Business Object Reference Ontology. The purpose of the method is to re-engineer disparate sources of information into a common model. It is particularly good at semantic analysis – establishing whether two concepts are the same, if they overlap, or if they are unrelated. BORO is different to most of the existing data analysis methods. The process itself ignores the names of things. Instead, the analyst is forced to identify individual concepts by their extent. In the case of physical objects, this is their spatial/temporal extent. For types of things, the extent is set of things that are of that type. The BORO methodology can be summarised as a flowchart:

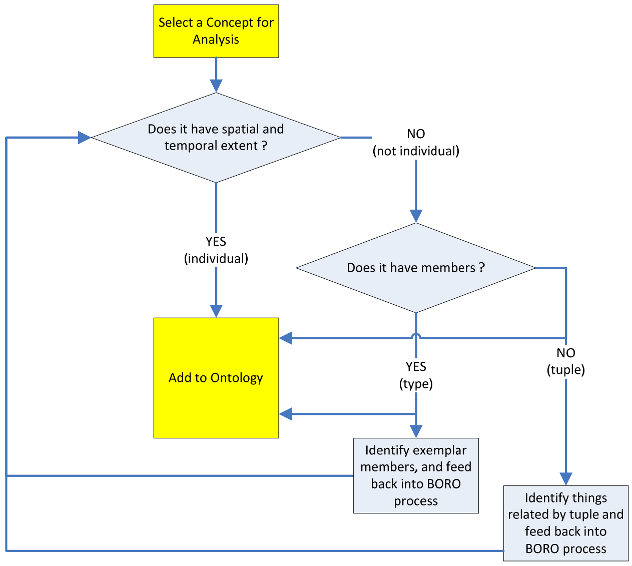

Figure 4 - Flowchart Summarising the BORO Methodology

Because the elements in a BORO-derived ontology are identified by their extent, there can be no question as to whether two things are the same or different (this is the key to semantic interoperability). Once the identity of something has been determined using BORO, names may then be applied to the elements in the ontology.

IV. The IDEAS Naming Pattern

The BORO process deliberately avoids lexical analysis and instead relies on extent as the basis for identification. However, for the ontology to be useful, it must have some human and machine-interpretable names for the elements it contains. The IDEAS ontology is split into two domains – ObjectSpace and NameSpace [3][4]. The ObjectSpace consists of those elements which result from the extensional BORO analysis. The NameSpace consists of names which refer to the elements in the ObjectSpace. Each name has a NameType which allows names to be categorised for a particular community.

[1] By "semantics", we mean the things in the real world that are being referred to by the elements in the ontology. Computer scientists tend to use the term to mean the underlying structure of data rather than the extent of the things the data refers to. This, it may be argued, is one of the root causes of interoperability problems in modern information systems. In terms of true semantics, it is impossible to tell by automated methods whether or not two ontologies are referring to the same extents, hence governance is required as to how the ontologies were produced if they are to interoperate.

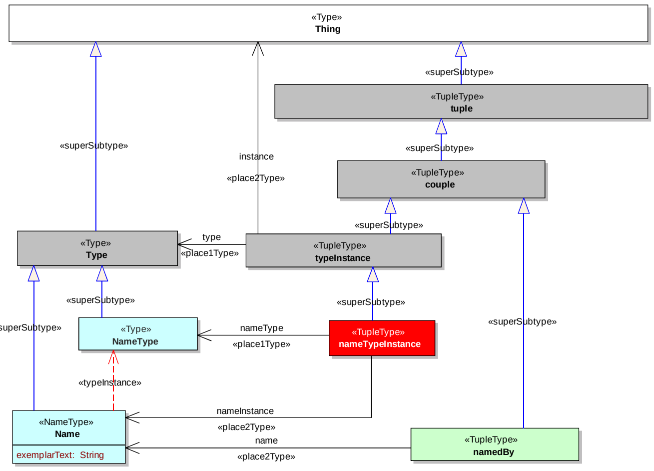

Figure 5 - The IDEAS Naming Pattern

Note that the names are types, not individuals. This is based on Strawson's theory of utterances[5] – where the utterance of a name is an individual event and so has an extent. The name is the set of the utterances that are utterances of the same name. Hence one could say "Fred" twice, referring to the same person. Each utterance of "Fred" is an individual, but the name "Fred" (which is the name of the person Fred) is a type (whose members are all the individual utterances). Other utterances of Fred that do not refer to the same person, do not qualify as instances of this type.

V. THE MOD ONTOLOGY DEMONSTRATOR[6]

As mentioned before, the use of IDEAS in MOD was aimed initially at enterprise architecture information. The Information Coherence Authority for Defence (ICAD – a branch of MOD DG-Info) was keen to see where else IDEAS and BORO could be used. The scenario chosen for demonstration was geo-political entities – countries, regions, continents, etc. This scenario was chosen because of the potential of using the BORO process and the IDEAS Naming Pattern to de-conflict the various codification standards used for geo-political data – e.g. ISO3166, FIPS10-4, NATO STANAG 1059, etc. The demonstrator also used the IDEAS Whole-Part and Overlaps patterns to model the structure of geo-political entities and their shared borders (see Figure 6).

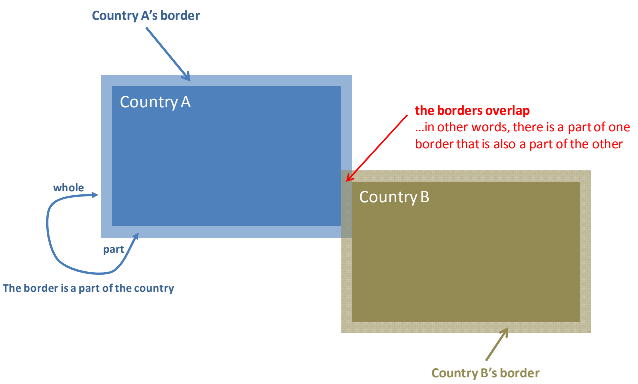

Figure 6 - Simplified Representation of the Whole-Part and Overlaps Patterns as Applied to Countries

The demonstrator uses these patterns to present a geo-political structure, the various names and codes used for geo-political entities and the borders between them:

Figure 7 - Screenshot of the MOD Ontology Demonstrator

Note the use of Google Maps™ in the demonstrator – in this case it is simply an additional feature, but future versions could use the maps as a user interface to the underlying ontology.

The information sources for the ontology were mostly public domain standards. The whole-part structure was provided by an existing UK Govt geo taxonomy. The borders information came from the CIA World Factbook:

Country	b	Borders
Afghanistan	total: 5,529 km	border countries: China 76 km, Iran 936 km, Pakistan 2,430 km, Tajikistan 1,206 km, Turkmenistan 744 km, Uzbe
Akrotiri	total: 47.4 km	border countries: Cyprus 47.4 km
Albania	total: 720 km	border countries: Greece 282 km, Macedonia 151 km, Montenegro 172 km, Serbia 115 km
Algeria	total: 6,343 km	border countries: Libya 982 km, Mali 1,376 km, Mauritania 463 km, Morocco 1,559 km, Niger 956 km, Tunisia 96
Andorra	total: 120.3 km	border countries: France 56.6 km, Spain 63.7 km
Angola	total: 5,198 km	border countries: Democratic Republic of the Congo 2,511 km (of which 225 km is the boundary of discontiguo
Argentina	total: 9,861 km	border countries: Bolivia 832 km, Brazil 1,261 km, Chile 5,308 km, Paraguay 1,880 km, Uruguay 580 km
Armenia	total: 1,254 km	border countries: Azerbaijan-proper 566 km, Azerbaijan-Naxcivan exclave 221 km, Georgia 164 km, Iran 35 km,
Austria	total: 2,562 km	border countries: Czech Republic 362 km, Germany 784 km, Hungary 366 km, Italy 430 km, Liechtenstein 35 km
Azerbaijan	total: 2,013 km	border countries: Armenia (with Azerbaijan-proper) 566 km, Armenia (with Azerbaijan-Naxcivan exclave) 221
Bangladesh	total: 4,246 km	border countries: Burma 193 km, India 4,053 km
Belarus	total: 2,900 km	border countries: Latvia 141 km, Lithuania 502 km, Poland 407 km, Russia 959 km, Ukraine 891 km
Belgium	total: 1,385 km	border countries: France 620 km, Germany 167 km, Luxembourg 148 km, Netherlands 450 km
Belize	total: 516 km	border countries: Guatemala 266 km, Mexico 250 km
Benin	total: 1,989 km	border countries: Burkina Faso 306 km, Niger 266 km, Nigeria 773 km, Togo 644 km
Bhutan	total: 1,075 km	border countries: China 470 km, India 605 km
Bolivia	total: 6,940 km	border countries: Argentina 832 km, Brazil 3,423 km, Chile 860 km, Paraguay 750 km, Peru 1,075 km
Bosnia and He	total: 1,459 km	border countries: Croatia 932 km, Montenegro 225 km, Serbia 302 km
Botswana	total: 4,013 km	border countries: Namibia 1,360 km, South Africa 1,840 km, Zimbabwe 813 km
Brazil	total: 16,885 km	border countries: Argentina 1,261 km, Bolivia 3,423 km, Colombia 1,644 km, French Guiana 730.4 km, Guyana 1
Brunei	total: 381 km	border countries: Malaysia 381 km
Bulgaria	total: 1,808 km	border countries: Greece 494 km, Macedonia 148 km, Romania 608 km, Serbia 318 km, Turkey 240 km
Burkina Faso	total: 3,193 km	border countries: Benin 306 km, Cote d'Ivoire 584 km, Ghana 549 km, Mali 1,000 km, Niger 628 km, Togo 126 km
Burma	total: 5,876 km	border countries: Bangladesh 193 km, China 2,185 km, India 1,463 km, Laos 235 km, Thailand 1,800 km
Burundi	total: 974 km	border countries: Democratic Republic of the Congo 233 km, Rwanda 290 km, Tanzania 451 km
Cambodia	total: 2,572 km	border countries: Laos 541 km, Thailand 803 km, Vietnam 1,228 km
Cameroon	total: 4,591 km	border countries: Central African Republic 797 km, Chad 1,094 km, Republic of the Congo 523 km, Equatorial G

Figure 8 - CIA World Factbook Data on Country Borders

VI. Conclusions & Future Plans

The demonstrator proved that the extensional approach, coupled with a naming pattern based on Quine's work on reference provides an ontology that is well suited to de-conflicting multiple identification schemes. The only issue is that because time and funding were so limited, the demonstrator itself does not show all the usual benefits of an ontological approach – flexibility, precision, openness. The implementation was restricted to demonstrating the whole-part, overlap and naming patterns, but the ontology itself is capable of much more. In particular, there is much interest in implementing the 4D aspects of the ontology to show events in a geo-spatial context.

The IDEAS Model continues to be developed, and MOD ICAD is still investigating uses for the MOD Ontology – e.g. de-confliction of cost centre codes, etc. There has been some interest in the use of maps as a user-interface to the ontology, and avenues of funding are being investigated. There has also been some suggestion that the ontology should be exposed as a web service to act as a de-confliction facility for geopolitical identification schemes.

It can be argued (and indeed has, by the conference programme committee) that country codes are not a convincing application for an ontology. One could easily conceive of a traditional software tool for this purpose, but it would not be extensible in the way that an ontology implementation is.

The main aim of this demonstrator was to show how the IDEAS naming pattern facilitated de-confliction of reference data. The value (and broader application) of this functionality in defence should not be underestimated. Although machine reasoning is clearly an exciting and useful application of ontology, there is much more potential for business improvement and cost saving in de-confliction of information. The use of an extensional ontology, which makes a clear distinction between names and objects (reference and sense), has great potential for improving the way defence manages its information.

References

[1] The IDEAS Group Website – http://www.ideasgroup.org
[2] Definition of powertype / powerset http://en.wikipedia.org/wiki/Powerset
[3] Goodman's preface to Quine's "Roots of Reference" (ISBN 0812691016)
[4] Business Objects: Re-Engineering for Re-Use, Partridge (ISBN 0955060303)
[5] http://semantics.uchicago.edu/kennedy/classes/f07/pragmatics/strawson64.pdf
[6] http://www.modaf.com/file_download/37/OntoDemo.zip