



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Extending Document Models to Incorporate Semantic Information for Complex Standards

Darren Clowes^a, Ray Dawson^a, Steve Proberts^a

^aDepartment of Computer Science, Loughborough University, UK.

ABSTRACT

This paper presents the concept of hybrid semantic-document models to aid information management when using standards for complex technical domains such as military data communication. These standards are traditionally text based documents for human interpretation, but prose sections can often be ambiguous and can lead to discrepancies and subsequent implementation problems. Many organisations will produce semantic representations of the material to ensure common understanding and to exploit computer aided development. In developing these semantic representations, no relationship is maintained to the original prose. Maintaining relationships between the original prose and the semantic model has key benefits, including assessing conformance at a semantic level rather than prose, and enabling original content authors to explicitly define their intentions, thus reducing ambiguity and facilitating computer aided functionality. A framework of relationships is proposed which can integrate with common document modeling techniques and provide the necessary functionality to allow semantic content to be mapped into document views. These relationships are then generalised for applicability to a wider context.

Keywords: *Semantic Modeling; Information Modeling; Aerospace; Defense; Quality*

1 INTRODUCTION

Increasingly the design of complex engineered products and systems (here within referred to as systems) are becoming more reliant on computer-supported models/representations of information, which can be used with computer checking algorithms to ensure consistency and correctness. Computer-supported models enable engineers to represent information graphically, which can aid understanding. These models also give a more rigorous definition of the systems requirements than that of the more ambiguous (open to interpretation) prose representation of the same information. Many system properties can be modeled, such as, behaviour, functionality, verification information, and manufacturing instructions.

For the purpose of this research, a complex engineering domain refers to a domain where the documented materials are large in volume and contain elements that could be modeled semantically. These elements leave no possibility for different interpretations; as such they can be modeled to create the semantic information. Semantic information is the formalised description of the meaning and relationship between elements. This information is often represented in a semantic model. In particular, we use the term semantic model to refer to our models of the case study domain such as the message catalogue and data dictionary. This differs from document model which is a term used to describe a model which represents the structure of the document, i.e. paragraphs, tables etc. This model does not imply any knowledge of the information, more the spatial rendering of the information.

In engineering domains, the documentation is normally highly structured. Engineers are required to use, and have a detailed knowledge of these documents to perform their task. In complex engineering domains, these documents normally consist of large sections of prose. It can, therefore, be hard to understand the precise objective of the section. These sections generally contain elements which could be used semantically mixed with prose which aims to help the readers understanding, but can often be ambiguous.

There are many sectors of the complex engineering domain that could benefit from the use of semantic and static representation of information. The types of domains that will benefit include those that have associated standards documents. The defence industry contains many such documents. Other industries that have similar documents, which this technique may benefit, include the automotive, aerospace, nuclear and legislative sectors.

Most military standards in use are still disseminated as text-based documents. In our experience, this is also the case in most other domains requiring complex detailed standards. These standards can be thousands of pages in volume, which can make locating and composing information in them challenging and laborious. Within the development of military hardware, engineers are typically required to produce text based documents based on a subset of the standards implemented by their product. Product testing will require the engineers to promptly locate information in the standard and validate that their product conforms to it. Interoperability is also crucial in military applications where engineers must verify that their implementation does not impede or conflict with other products. To address these issues, we have found it useful to extract semantic models represented in a structured format which are then amenable to automated querying and processing. In our experience this

approach can greatly enhance both the accuracy and speed of locating and composing information from different parts of a standard. We have also found that in order to make models useful for engineers, it is essential to construct a suitable and familiar user interface for querying and navigating them.

2 BACKGROUND

To evaluate the proposal of hybrid semantic-document models, the military domain of Tactical Data Links (TDL) was chosen as a suitable case study with the support of BAE Systems. Tactical Data Links display many of the characteristics of a complex engineering domain. The domain is specified through standard documents, each of which is many thousands of pages in length. They are highly structured and contain material that can be represented semantically. Regulation of conformance is obtained through production of documents similar in format and style to the original standard.

The TDL domain is a collection of related technologies designed to work within the command, control, communications, computers, and intelligence (C4i) used in the dissemination of information within a battlespace to support joint and combined operations. The TDL domain is often referred to as a family of standards, as several variations of Tactical Data Links have evolved to interface with specific unit types (infantry, aircraft, ships, etc). These variations may differ in waveforms, bandwidths, protocols and capabilities.

The TDL provides one of the backbone technologies underpinning the defence community's goal of Network Enabled Capability (NEC) by providing the information and infrastructure to afford users with both an integrated picture of the battlefield and also provide tasking orders and responses. A number of TDLs are in service with coalition forces, and are implemented on a variety of assets, such as aircraft, ships, land vehicles and command stations, an example of which is illustrated in Figure 1.

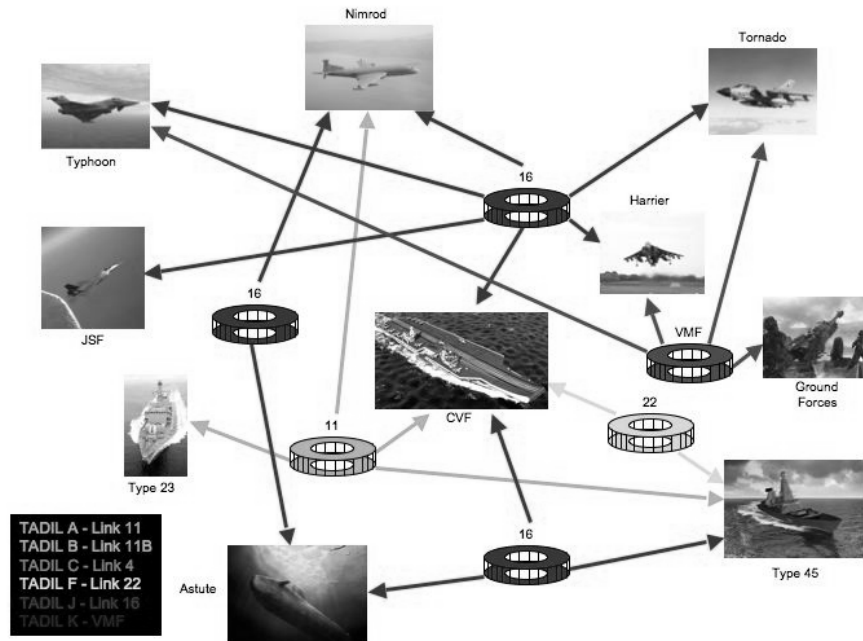


Figure 1. TDL Overview

For the purposes of this project it has been decided to use the Link 16 standard. This is described in the U.S. Department of Defense Military Standard MIL-STD-6016C [1] and NATO Standardization Agreements (STANAGs) 5516 [2] and 4175 [3]. As both STANAGs are required for the complete representation of Link 16, the MIL-STD version will be the basis for this project's interpretation.

The Link 16 TDL is a general purpose TDL, in contrast to some others, e.g. Link 4A or Variable Message Format (VMF). A list of data link characteristics is provided elsewhere [4]. Link 16 has evolved over a number of years, stemming from a requirement identified by the US military in the early 1970s for a TDL offering a broad range of functions that would be applicable for use across multiple forces (e.g. Navy, Marines, Air Force, Army, etc.).

The Link 16 TDL (in MIL-STD and STANAG) is described in the form of narrative combined with many tables and relatively few figures. It is known to feature a number of shortcomings affecting its usability [5]. Of particular relevance to the research undertaken by the System Engineering Innovation Centre at Loughborough are, that the standard:

- is document-based, with no apparent underpinning model
- is largely narrative
- is open to (mis)interpretation
- is not checkable by machine
- contains duplication of material, inviting inconsistency
- has only very limited use of hyperlinking, impeding document navigability
- comprises many interdependent sections and appendices
- is enormous, >8000 pages

3 WHAT IS THE PURPOSE?

The aim of hybrid semantic-document models is to provide a method to facilitate relationships between text-based document views of information and more rigorous computationally accessible representations of its semantic content. Regulatory requirements enforce a document-centric culture, and engineers themselves come from many engineering disciplines and generally consider document based standards or requirements as familiar and comfortable. Although military standards are highly structured and contain many elements that could be modeled syntactically, they are not in a format accessible to computer processing. Computer aided processing of standards offers several beneficial processes that can enhance the efficiency and productivity of engineers using the standards. However, a solely computer modeled standard becomes inaccessible to the engineers.

Therefore, this philosophy aims to achieve a level, whereby the engineers can generate the respective views that they are comfortable with (i.e. traditional document based), as well as having a sound validated semantic model which is computer readable and executable. In reviewing this approach, there are two perspectives to consider when applying the philosophy. Firstly, managing legacy information maintained in a prose/non-model based document and, secondly, the generation of new documents to aid in the consistency and transfer of understanding and knowledge. It is, therefore, believed that such an approach of hybrid semantic-document models will facilitate the following benefits and objectives:

Capture the interpretation of complex prose in a more formal mechanism. This allows the information author to express their intended action explicitly (removing ambiguity) or where the document already exists, an expert can capture their interpretation which can then be shared by document users to gain a shared consistent knowledge.

Identify and resolve/prevent inconsistencies within the document. Mapping semantic data to its required positions within a document reduces the possibility of errors of inconsistency being introduced through duplicated data.

Identify and resolve/prevent missing data. It is anticipated that by modeling the information, cases where references to non-existent data can be captured, resolved and/or avoided. The use of model validation tools would enhance this ability to detect these instances.

Update information more effectively. An update to the semantic model would be reflected in all positions of the document. This should make updating and maintaining the document easier.

Provide alternative navigation structures. It would be possible to provide alternative navigation menus to the traditional tree like hierarchical document structure. Similarly it would be possible to restrict the amount of information displayed by restricting the information according to semantic constraints rather than document hierarchies.

Enhance document information searching. It should be possible to enhance traditional document searching through the use of the semantics. Search methods could involve querying the semantic models to return related points within the document.

Use model execution facilities. By having the core semantic information of the document modeled, it should be possible to utilise model execution facilities to gain computational benefits from the document. This could include code generation, model querying, model validation, etc.

Perform advanced document comparison. It would be possible to compare the semantics of multiple documents to assess their compatibility and/or similarity beyond the simple text comparison available for a traditional document.

This research, therefore, is carried out to help resolve (1) the knowledge problems of understanding complex documents, interpreting their meaning in particular context and identifying errors of inconsistency and completeness, and (2) the design problem of reforming such complex documents to achieve completeness, eliminate inconsistencies and enable computer execution (REFERENCE TO WIERINGA). In respect of (2) the resulting hybrid semantic-document models can be considered to be a design artifact (REFERENCE?), though this is not their only purpose as the knowledge problems of (1) indicate.

4 WHY IS IT NEEDED?

The primary user base for this research is a wide range of engineers who may not be familiar with modeling and model techniques. However, our audience is highly familiar with their traditional document centric methods of information capture and rendering. They would gain much from utilising formal models but regulation and business practises require much of their utilisation to be through a document. A document alone can be formally specified in terms of its structure, but the information cannot. The document can be ambiguous, creating problems of consistency and completeness when they grow large.

Semantic modeling is a very well-known and board research field, a semantic model alone would provide several of the anticipated benefits such as, model execution, avoiding missing data, etc., but they do not have the traceability back to the originating information source (in our case a document). They also do not facilitate easy rendering of information in document format for regulation and engineers' understanding. Document markup languages provide the easy rendering of information, but do not provide the semantic structure to model the information in a domain specific way. Therefore, in this proposal we try to address the combination of these two areas.

It is possible to maintain separate semantic models and document markup models, however, a second data source is introduced which must be maintained in synchronisation with the original document otherwise the benefits are lost completely. With separate models, it is also difficult to indicate what part of the model relates to elements of prose in the document, this means it would be difficult for engineers with no modeling background to understand.

There appears to be little research in combining a formal model with document markup. Much investigation has been carried out into defining models and/or ontologies to add meaning to complex documents [6, 7], but this does not help with the execution or other model management functions such as code generation or model validation. By using predefined domain markup, more generalised queries can be applied to the document which would only move part way to our overall aim. It is predominantly a technique to assist with information retrieval tasks, for which much has been discussed on how to model the document structure to aid the retrieval process. This too, does not help in achieving our aim.

Similarly much research has been conducted into requirements traceability [8, 9, 10, 11]. Although you could consider the MIL-STD as a requirements document, the research on requirements traceability is focused on tracing between chains of documents. Even if we treated our semantic model as an XML file to facilitate requirements traceability techniques, it still would not help reduce the duplication and would make utilising the models more difficult.

One of closest examples of related work is by Winter and Nittel [12] who discuss the geospatial standard and address the need for more formal modeling. Although they do stress the relationship between prose and the models, there is no direct traceability and the information is duplicated. They make use of functional programming languages which, given the wide range of engineer skills anticipated to use our solution, would be likely to be confusing. During workshops with a cross section of engineers using the MIL-STD, it was found that even UML style modeling was not widely understood.

The closest related area of research is the Health Level 7 Clinical Document Architecture (CDA) framework [13, 14, 15, 16]. The framework has two core parts, the Clinical Document Architecture (CDA) and the Reference Information Model (RIM). This is equivalent to our proposed approach for complex engineering domains, a document model similar to the CDA and a semantic model representing the RIM. The aims of the framework are not too dissimilar to our research:

“The need for a clinical document standard stems from the desire to unlock the considerable clinical content currently stored in free-text clinical notes and to enable comparison of content from documents created on information systems of widely varying characteristics. Given the variability in clinical notes, including structure, underlying information models, degree of semantic encoding, use of standard healthcare terminologies, and platform- and vendor-specific features, it is currently difficult to store and exchange documents with retention of standardized semantics over both time and distance.” [13]

The RIM provides a vocabulary (ontology) and the meta-model to construct semantic information within a document. This inclusion of semantic information in the document is a primary aim for our research. The CDA is a document markup standard that specifies the structure and semantics of clinical documents. Explicit markup tags are defined that facilitate

the encoding of text according to the ontology. The text is not directly retrieved from the underlying RIM, the ontology is used purely to encode the text string. In addition to utilising the ontology, the CDA also allows for the recording of process and interactions through the use of instances of the RIM model. These instances of the RIM require the reuse of the data recorded in the textual elements leading to duplication. There is no attempt made to reuse the information recorded in either part, although CDA does record both the instances of both models within the one document.

There is criticism of the RIM [17, 18], however, the RIM itself provides no benefit to our domain as it is an information model for clinical workflows. The CDA and RIM at present have also deliberately avoided advanced and complex semantics such that a staged approach to the implementation and delivery of a shared structure and semantics can be achieved [13]. This too is something that we consider suitable for complex engineering domains by growing and evolving the semantic model over a period of time allowing for user adoption and providing time to establish the complex semantics.

An astute feature provided with the CDA is the ability to locally define markup [15] which can extend the CDA semantics provided that these extensions do not change the meaning of any of the standard data items, and can be safely ignored by applications without the local knowledge. The CDA approaches the problem where there are multiple sources of information that need to be standardised to enhance interoperability. Our domain works in the opposite direction. The source is singular and well controlled, we however require better methods to disseminate the information to different users with different requirements.

5 CASE STUDY ANALYSIS

Tactical Data Link 16 is one such standard that the authors believe would benefit from the use of hybrid semantic-document models. Within the Military Standards area alone there is a wealth of standards that display similar characteristics to the Link 16 standard, for example VMF, Link 11. In addition NATO and the US DoD produce these standards across a differing set, i.e. Link 16 is defined by US DoD MIL-STD-6016 [1] but NATO define the standard with STANAGs 5516 [2] and 4175 [3]. These standards themselves are huge documents thousands of pages in volume.

Given the wealth of material available and due to the close association between the researchers and the study, the approach was an action research [19] based “single-case” case study [20] method, focusing on the MIL-STD-6016C. Given the volume of the MIL-STD-6016C the authors consider a single-case design sufficient to allow the construct and proof of the hybrid semantic-document models concept. It is also believed that the framework developed by analysing the MIL-STD-6016C can thus be generalised such that it can be applied not just to STANAGs or Military Standards, but to other domains where the capture of complex rigorous information is maintained in a prose based document.

It is intended to produce a framework that will enable the concept of hybrid semantic-document models to be applied to multiple sources. The framework in this context refers to a collection of model components with guidance on where and how they can be used to

construct a hybrid semantic-document model of a whole document. Grounded theory [21] is the discovery of a theory through the systematic analysis of data. In this research we have started with the concept of hybrid semantic-document models and, by using a grounded theory-like approach the aim was to discover the relationships and means of representing these relationships from the MIL-STD-6016C case study to produce this framework.

These methods will be applied and evaluated using an interpretivist philosophy. An interpretivist philosophy is being employed given the multitude of potential valid interpretations of the data and formulation of frameworks. Within this, a constructive approach as per Iivari [22] will be used to define the framework. This is intended to utilise an inductive process by observing the material present in the Tactical Data Link Standards and using these observation to infer a framework which can achieve the objectives in Section 3.

In conducting the case study, a significant form of data capture will be required. To obtain this data we shall use a document analysis method similar to that of Salminen et al [23]. They use multiple sources as part of their document analysis, however it is the authors' opinion that the volume of MIL-STD-6016C and its segregation into discrete sections is sufficient for data gathering. Given the volume, sampling was utilised. The sampling method was performed using cluster sampling. Given prior knowledge of the document's format and structure, random samples were picked to highlight the types of document and information likely to be encountered. The sampling method was performed using cluster sampling. Given prior knowledge of the document's format and structure, random samples were picked to highlight the types of document and information likely to be encountered.

5.1 Sample Selection

Previous extraction and modeling of the semantics at BAE Systems has led to a level of knowledge of the document. As such, a selection of pages from the standard was selected using stratified sampling, key areas of the material can be grouped into strata, and a random sample selected from these strata that represent the semantic model and/or the document structure. The strata were not equivalent in volume, but were intended to cover a cross section of the differing document structures and the semantics that would need to be mapped. For security and sensitivity reasons actual headings and text used in this paper have been substituted with fictional material that maintains the essence of the original.

Three pages from each strata were selected to form the sample. Should the selected page contain a full page graphic, this was ignored and an alternative random page selected. Consecutive pages were also prevented. The random sampling method was non-probabilistic judgement. Judgement was provided by a member of the BAE Systems research team who had knowledge of both the semantic models and the document sections and content. The samples obtained are shown in Table 1.

Table 1. Sample Selection

Strata	Pages
Payload definitions	B-38, B-146, B-824
Message definitions	5.1-15, 5.1-440, 5.2-12
Message rules	F-6, K-14, P-53

Transmit and receive tables	5.4-J7.0-6, 5.4-J12.6-1, 5.5-J14.2-1
Constraint information,	4-23, 4-29, 4-169
General information	1-17, 3-56, 4-130
Message Uses	5.5-J12.0-15, 5.5-J0.3-1, 5.5-J2.5-2
Minimum Specification	W-18, W-27, W-92

On analysis of the samples, it became apparent that, for the message rules, too much meaning and information was lost by reading a single page. The sample selection for message rules was, therefore, changed to a trace. A non-probabilistic random rule was used and all pages necessary to trace an understanding were selected. For example, Rule F.1.2 covers two pages, however, through its references it requires further knowledge of two rules, a transmit table and indirect knowledge of four other messages. These indirect references were ignored as their exclusion would not prevent basic understanding. This resulted in a total of six pages for analysis for this rule.

5.2 Location of Information

In maintaining two sources of information and formatting, the primary location and relationships between sources needs to be considered. The semantic model is the obvious choice as the primary information location. As the document model contains multiple references to the same object, this information can be maintained singularly in the semantic model, resulting in a change to the semantics being reflected at every location in the document. This supports the objective of reduced duplication and hence improves quality. It is also the model which is likely to experience the most computational activity as the document is simply a rendering activity.

However, is it that simple? The two models can be applied to a matrix to aid the analysis. This two-by-two matrix is represented in Figure 2.

		Information Stored in Semantic Model	
		No	Yes
Information Stored in Document Model	No	Case 1	Case 2 No Document View Required Case 4 Document View Required
	Yes	Case 3 No Semantic View Required	Case 5 Semantic View Required Case 6

Figure 2. Information Store Matrix

Cases 1-3 of the matrix can be ignored. These cases have an impact on the location of information as, in case 1, the information does not exist and, in cases 2 and 3, information is not required by the alternative source. In theory, cases 4 and 5 are transferable, in that the information can be located in either and then a link or view of the information could be maintained in the other. In these scenarios, case 5 is considered the better option as mentioned previously; the semantic information will only be stored in one model, hence the better option. The only occasion where case 4 is not a limitation is a singular occurrence, in which case this can be transferred into case 5 by relocating the information into the semantic model. This would improve any future development as well.

As case 5 represents the situation where the data can be directly mapped from the semantics to the document view, or at least derived from the semantic content, case 6 presents potentially the most challenging situation. It represents the case where the same information is represented in two mutually exclusive formats such that one format cannot logically be derived from the other. As such, this case presents the need to maintain a relationship between the two equivalent sets in both sources, this would therefore result in no singular, comprehensive primary location, i.e. both locations would be required for the complete picture.

5.3 TDL Semantics

To perform an analysis of the material, knowledge of the semantic model that supports the document prose must also be established. It is beyond the scope of this paper to present the exact models used, but an overview is provided in this section.

At the lowest level of granularity, there exists a Data Dictionary identifying the set of data types defined for use on the link. These types are identified by a unique key, the Data Field Identifier (DFI) and the Data Use Identifier (DUI) pair, referred to as the DFI/DUI. The set of messages (referred to as J-Messages) that may be transmitted over the link are defined in the form of a Message Catalogue, messages are functionally oriented and contain a number of words, referred to as J-Words, each of which contains a number of fields, the type of which is defined by reference to the relevant item in the Data Dictionary (the DFI/DUI).

Hence, Link 16 messages are tree-structured and must conform to certain well-formed constraints, e.g. all bits in each J-Word must be associated with a DFI/DUI (i.e. all fields must have a defined type). Such constraints have been captured in our models and are described elsewhere [24]. There are a small number of different types of word, and certain elements of the payload are mandated by the word type. This structural set of information covers a large proportion of the standard, covering approximately 4000 pages or some 53% of the document.

In addition, structural modeling has also been undertaken on other sections which detail the network management and Time Division Multiple Access (TDMA) architecture. Once again, there was much information which could be captured and represented by a meta model. Another large section of the standard covers the behavioural aspects of the standard with approximately 2000 pages covering the definition of transmit and receive tables and appendix data defining the behaviour and rules related to message transmission, receipt, storage and display. The appendix rules follow an event (or stimulus) - condition - action type structure. There is a clear process flow within which can be modeled.

5.4 Sample Analysis

Through the analysis of the samples, one key association was observed at an extremely high occurrence rate. It almost exclusively forms entire sections of the samples, specifically the message and data definitions. This association is between an element of text and an instantiated class attribute. In many cases the attribute value of the class is presented directly in the document in the form and format defined in the semantic class. To demonstrate this association consider a semantic model for “*Message*” defined in Figure 3. Within the document, tables simply render all of a *Message*’s parts, in a table i.e. Figure 4.

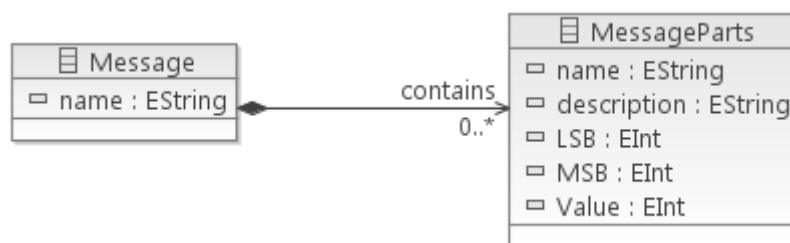


Figure 3. Message Semantic Model

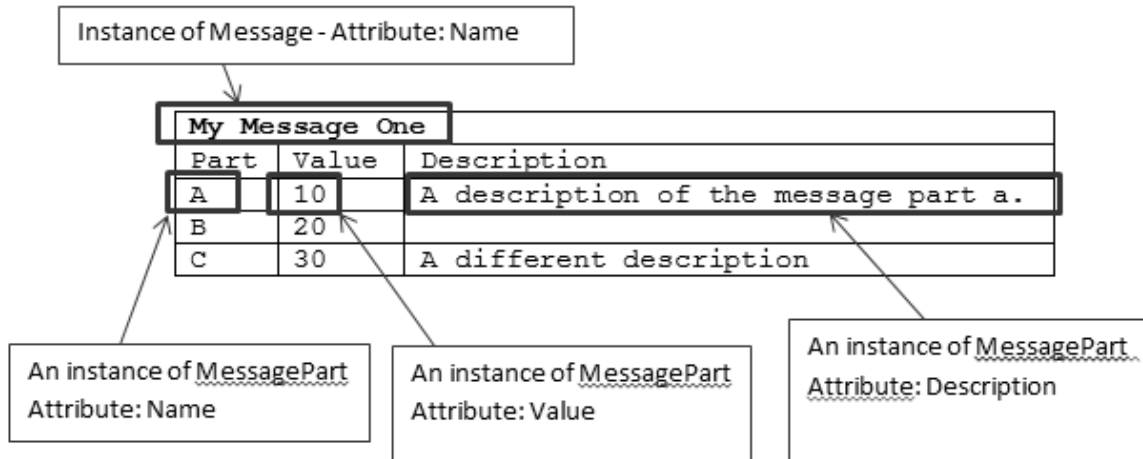


Figure 4. Example Document Content (Message)

To facilitate the aims of hybrid models, a relationship would need to be specified such that this observed association can be modeled. As such, a “*direct*” relationship should be defined where an instantiated class attribute can be directly utilised from within the document model.

This attribute based substitution can also be observed in other forms. It is hypothesised that the document will contain Boolean information, and the storage format and display format of Boolean data could be different. Most computer based systems use the bit as the storage medium using 0 for false and 1 for true. The hypothesis is that there may exist situations where using a direct replacement of Boolean data will result in illiterate sentence/word structures. Therefore, the bit value should be converted to a more meaningful string for display in the prose, such as, True and False or Yes and No. This could be extended to the consideration of numeric display. In some instances the string can be used but it is the actual numeric value which is stored in the model. Similar synonym-based actions can be observed across the samples. For example, the minimum specification defines a set of characters that are used to indicate mandatory and optional settings. The transmit and receive tables also use characters for the source and message use display type. Therefore, boolean/numeric conversion can be considered as part of the larger concept of synonym substitution, whereby the attribute value is substituted by a synonym. A “*synonym*” relationship should, therefore, be defined to allow for this reformatting.

A more concealed association was observed in the message and data definitions. Within the samples it could be seen that the text could be derived from the semantic class attributes. Extending the example shown for direct data demonstrates this type of occurrence. For this, we have introduced the start and end bits of the message part. The new table, Figure 5, needs to display the start, end and total bits. Using our semantic model, the number of bits can be derived from the start and end bits. Similarly if it had been modeled with the start and length, the end bit could be derived. Therefore, a “*derived*” relationship should be defined where the data displayed in the document is a result of the calculation based on attributes on the instantiated class.

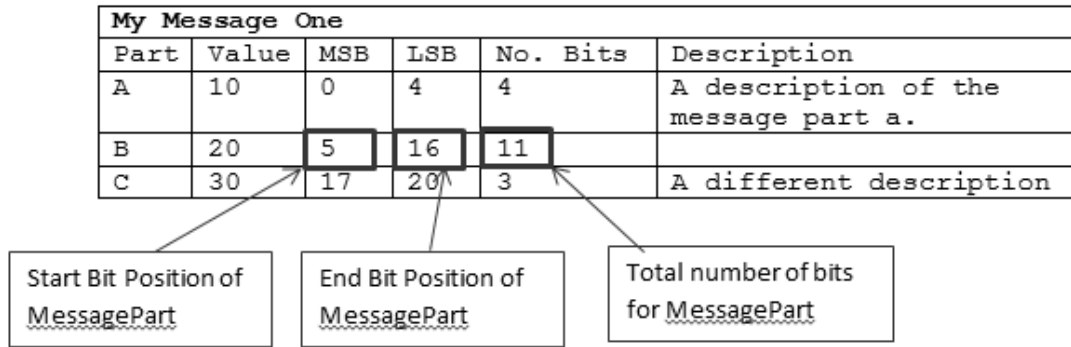


Figure 5. Example Document Usage Semantic Derived Data

The case study document makes use of its highly structured nature to infer further associations. Large sections of narrative are broken into a number of paragraphs. The first paragraph might make reference to a semantic class through the previously discussed “direct” relationship. In doing so there is a clear link established between the paragraph and the semantic data. Any child paragraph can then also infer the relationship through applying the hierarchical structures of a document model. However, the subsequent paragraph, at a sibling level can, in some observed instances, have an implied association with the semantic class referenced in the previous paragraph. The paragraph makes no direct mention or use of data from the semantic class but, through the terminology, spatial proximity and grammar, a human reader can infer the association. Considering an objective of document sub-setting or semantic searching, these sibling paragraphs would be ignored as no relationship would exist. Therefore, some form of “inferred” relationship is required to handle these instances.

During the analysis a common, misleading association was observed multiple times where the document references a table identifier in relation to discussing the semantic class. Because the document model table reference makes use of the semantic attributes from the reference text, it originally was mistaken to be an association to the semantic model. On closer inspection, it is, in fact, a traditional reference within the document model and no hybrid link needs to be established. What it does raise is the association a table has with semantic data. This highlights again the observations of the inferred relationship as the semantic association could possibly be inferred through its child cells or an association would need to be established as per the inferred relationship.

At present the technique for modeling the behaviour within the standard has not yet been implemented at BAE Systems. It is anticipated that further relationships may need to be developed to integrate the behavioural semantics. In particular, it is anticipated that the behaviour may not be directly representable in the document prose. In such cases the observed inferred relationship may be sufficient to handle most cases of behaviour. The main anticipated exception shown in initial behavioural studies implies a need to handle collection associations. Collections are anticipated to handle the situation where a relationship needs to refer to more than one instance of a class. It should be bi-directional in that either a text part is associated with many behavioural objects, or many logically separated text parts define a

semantic behavioural object. As such, there should be a “Collection” relationship defined to handle this anticipated requirement.

Therefore, analysis of the sample section has led to the identification of five possible basic relationship types:

- Direct Attribute Substitution
- Synonym Attribute Substitution
- Derived Attribute Substitution
- Inferred Relationships
- Collection Relationships

6 PROPOSED RELATIONSHIP TYPES

Having identified the required relationship types through the analysis of the sample data, the next stage was to propose the relationship definitions to facilitate this. Throughout this process, consideration was given to applying the definition to the material observed in the sampling. A business decision at BAE Systems, meant that the technology for any model would be the Eclipse Modeling Framework. The Eclipse Modeling Framework (EMF) is a modeling framework developed as part of the Eclipse platform [25]. It also provides a code generation facility for building tools and other applications based on a structured data model. The core EMF framework includes a meta model (Ecore [26]) for describing models and runtime support for the models. Therefore, all examples shown and some evaluated proposals are closely coupled to the Eclipse Modeling Framework and the underlying model format Ecore. The Epsilon [27] meta model management family tools are also the favoured technique for model manipulation at BAE Systems.

This section discusses the definitions and the considerations made during application to the observed material. It should be noted that in the examples shown, a class name in italics implies the class is abstract.

6.1 Direct Relationships

Direct Attribute Substitution appears sufficient to cover most occurrences. A direct substitution would simply require replacing the document text with the value of a semantic class attribute. This relationship can be defined by a relationship to the class and an identifier for the attribute of the class as seen in Figure 6.



Figure 6. Direct Substitution Proposal

However, a question is raised when applied to the identification of the Message Catalogue and Data Dictionary identifiers. These identifiers appear significantly throughout the

semantic information within the samples. For example, DFI/DUIs are identified by the pairing of the DFI id and DUI id to form a seven to eight character identifier, the J-Messages and their sub-parts are identified by the hierarchical combining of their ids. The J-Message or J-Word precisely demonstrates the conceptual problem. Is the identifier for a J-Word, a singular reference, or should the J-Word be a composition of references where the parent classes are also referenced by the selective substring of the J-Word identifier? To highlight this, Figure 7 shows the example of how the J-Word Z9.3X2 could be composed and represented in the document.

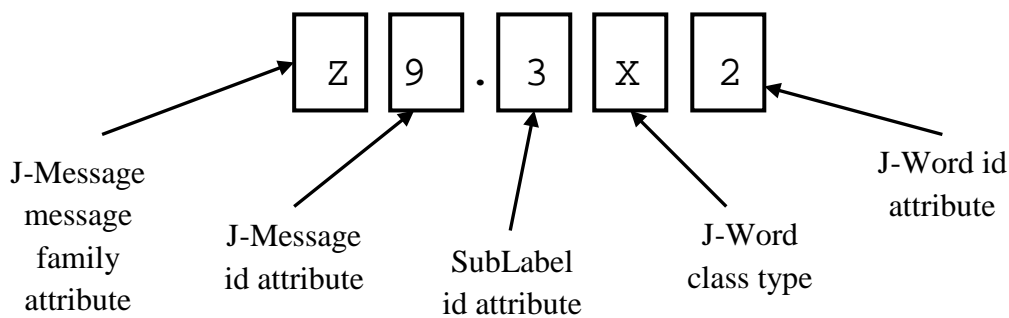


Figure 7. Decomposition of J-Word ID

When referring to the J-Word, the full string reference is used, i.e. “Z9.3X2”. Using simply “X2” would be ambiguous as there are many X2 J-Words of other J-Messages. The critical question is whether instances within the prose are made up of the composition of parts shown in Figure 7, or the full string itself should be the only reference to generate the prose?

By creating all instances within the document as compositions of the hierarchical parts, there is a clear and direct link to each level within the hierarchy. This makes it simple for discovering abstract queries such as finding paragraphs related to high level hierarch objects. However, ignoring our current use of the military standard, considering a scenario whereby the document is being generated through a series of iterations, this composition approach is extremely poor at handling change. For example, consider an iterative step whereby it is decided to move a J-Word X2 or SubLabel Z9.3 to Z9.4 as it was entered incorrectly or due to reorganisation. The use would also need to find all instances within the document of the composite relationships as the reference to the SubLabel would not change and would still be valid. To handle this scenario, it would be better to ensure each hierarchical identifier is a singular reference. By using the object model, relations to the parent hierarchical classes could still be discovered, but the document would not contain any direct knowledge.

To facilitate the singular reference, the relationship needs to have knowledge and ability to traverse the meta model. This traversal is not guaranteed to be fixed depth and the string composition could require differing attributes at each level. Therefore, a programmatic method would be best to generate the necessary text string for the relationship. This method-

based substitution relationship, defined in Figure 8, should exist in addition to the direct substitution relationship.

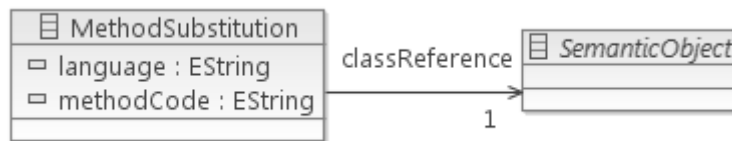


Figure 8. Method Substitution Proposal

Decomposition highlights another potential relationship. Should the referencing be seen as the composition of attributes, then the J-Word type, ('Initial', 'Extension' or 'Continuation') is identified by the inheritance within the semantic class as shown in Figure 9, and not by an attribute of the class. Therefore, a relationship to the class type would be required with the same functionality of synonym such that some text could be used to represent the class type. Although this type of relationship is evident in assessing the composition of the J-Message identifier, there is no other evidence of this elsewhere within the samples. It is, therefore, not deemed necessary to implement this type of relationship. Should it be found that this relationship is required, the functionality could be achieved through the use of the method substitution relationship identified in Figure 8.

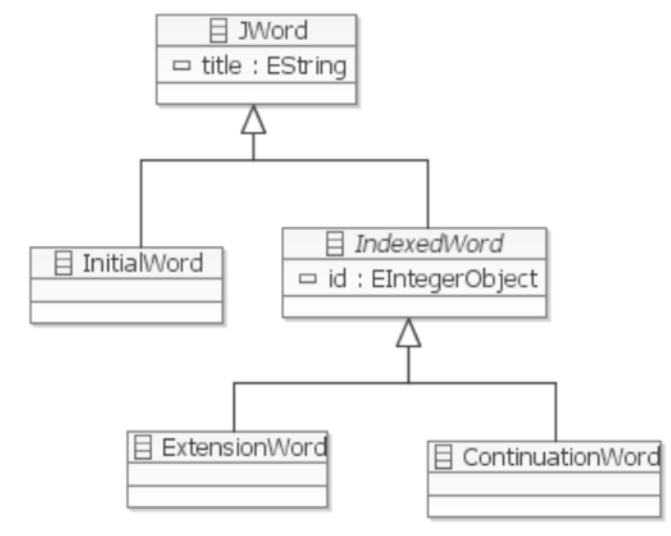


Figure 9. J-Word Inheritance

To provide an example of both direct and method substitution, a common occurrences from within the case study is presented in Figure 10. This represents a paragraph based description for a message which is fictitious for security reasons. This paragraph contains various references to the Message Catalogue. At present, this paragraph is typically captured independently as text in a document model and attributes in the semantic model. The excerpt in Figure 11 shows how the direct and method substitution can be used to markup the

document with a reference to the semantic model. The markup has been simplified for readability.

“Transmit a X7.0 Message Title 0 message, some text describing low level technical data and some more text that requires a second new message in the appropriate X3.x Message Title 1 message or X5.4 Message Title 2 message.”

Figure 10: Example Paragraph

```
<paragraph><text>Transmit a </text><MethodSubstitution language="eol"
method="function buildSubLabelId ..." reference="./messages.model/#_IDX7" /><text>
</text><<DirectSubstitution attribute="title"
reference="./messages.model/#_IDX7"/><text> message, some text describing low
level technical data and some more text that requires a second new message in the
appropriate </text><MethodSubstitution language="eol" method="function
buildMessageId ..." reference="./messages.model/#_IDX3" /><text>.x
</text><DirectSubstitution attribute="title"
reference="./messages.model/#_IDX3"/><text> message or </text><MethodSubstitution
language="eol" method="function buildSublabelId ..." /><text>
</text><DirectSubstitution attribute="title"
reference="./messages.model/#_IDX5"reference="./messages.model/#_IDX5"/><text>
message.</text></paragraph>
```

Figure 11 Example paragraph markup including direct & method substitution

6.2 Derived Relationships

The observed samples demonstrate a relatively simple case of derived data where the result is the output of an operation between two attributes. This type of simple derivation can be achieved through a predefined set of operations and links to the necessary attributes.

Although this type of implementation would facilitate the examples seen in the samples, it is not highly flexible and is fixed to the set of operations that would be defined in the meta model.

Having established the need for a programmatic method substitution to handle the generation of identifiers in the Analysis Section, this method substitution gives the full flexibility offered by the use of a programmatic language whilst proving all the necessary functionality to achieve a derived substitution. It is proposed, therefore, to drop the derived substitution relationship in favour of using the method substitution relationship. In using the Eclipse Modeling Framework with Epsilon it is proposed that this method substitution is implemented using two techniques. Eclipse has the ability to define eOperations for implementing methods on the class instances. Technique one would be to establish the relation to the result of the defined eOperation. Technique two abstracts out of the Eclipse framework and provides the relation as a text string of code and an implementation language. This would allow us to implement Epsilon as the method language to generate the result. The proposed relationships are defined in Figure 12.

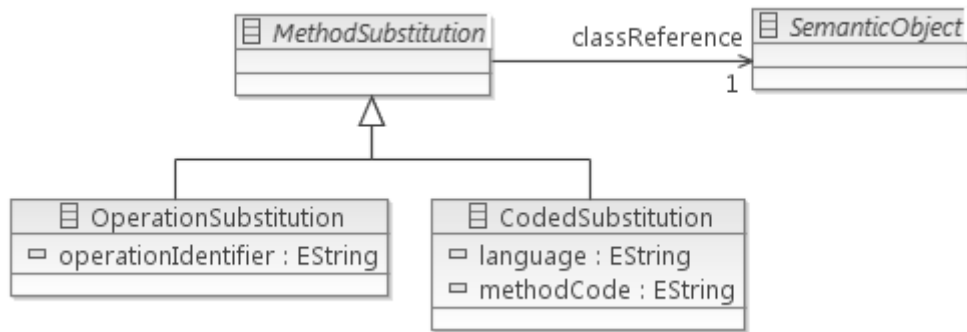


Figure 12. Ecore Modified Method Substitution Proposal

6.3 Synonym Relationships

The Synonym relationship is intended to facilitate the situations where the data held in an instanced semantic class requires formatting into another form. There are three potential methods to achieve this.

- 1) Basic fixed string replacement.
- 2) Mapped replacement.
- 3) Method replacement.

A basic string replacement method would allow the instances of the relation to be assigned a fixed string to be substituted, determined as and when the relation instance is created or modified. Although the basic fixed replacement provides a simple solution for existing documents, similar to that in the discussion of hierarchical identifiers, using a basic fixed replacement is not flexible. Working on the same scenario whereby the document may be created or modified in iterative steps, a fixed replacement suffers from the inability to respond to a change to the semantic attribute value. It would, therefore, be possible to use the method substitution relationship proposed in the derived relationships. This gives complete flexibility but requires a programmatic solution. An alternative would be to use a mapped replacement. A mapped replacement provides a greater flexibility, when defining the relationship; the user must also define a map between values and the required substitution. This method can facilitate the basic fixed method but provides the flexibility to respond to changes to the instantiated semantic class.

The mapped replacement strategy is not immune to faults, consider the scenario where the document author wishes to preserve the language used in the document such that numeric values are represented by their string name i.e. 3 = 'three'. The map necessary to handle this scenario would need to be infinite to support possible numbers. Although this scenario could be more easily supported through the method substitution, it is also an extreme scenario used solely to highlight the issue. As the mapped approach also enables the map to be defined within the model, this approach is favoured. With the method substitution required for other requirements, this too can be used for any complex mappings. Figure 13 shows our proposed substitution relationship.

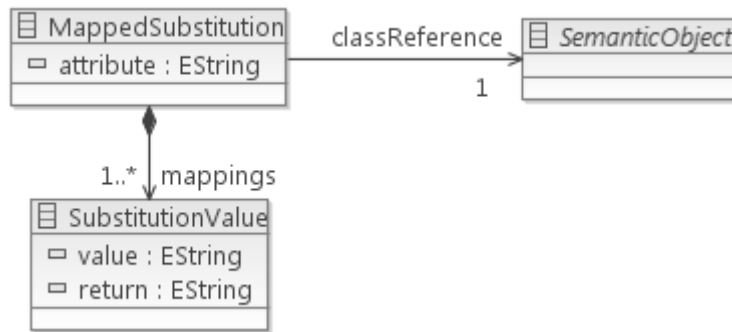


Figure 13. Mapped Substitution Proposal

The mapped substitution is observed mainly in tabular settings. One such example is where an element can be alerted to a user either, textually, graphically or audibly. This is captured in an enumerated type format in the semantic model represented by numeric values. These numeric values should be rendered in the document using a textual representation rather than numeric representation, this is shown in Figure 14.

```

<TableRow><TableCell><MappedSubstitution attribute="title"
reference="/messages.model/#_ID891"><SubstitutionValue Value="0"
return="T"/><SubstitutionValue Value="1" return="G"/><SubstitutionValue Value="0"
return="A"/></MappedSubstitution></TableCell></TableRow>
<TableRow><TableCell><MappedSubstitution attribute="title"
reference="/messages.model/#_ID892"><SubstitutionValue Value="0"
return="T"/><SubstitutionValue Value="1" return="G"/><SubstitutionValue Value="0"
return="A"/></MappedSubstitution></TableCell></TableRow>
<TableRow><TableCell><MappedSubstitution attribute="title"
reference="/messages.model/#_ID893"><SubstitutionValue Value="0"
return="T"/><SubstitutionValue Value="1" return="G"/><SubstitutionValue Value="0"
return="A"/></MappedSubstitution></TableCell></TableRow>
  
```

Figure 14: Example document table containing mapped substitution

6.4 Inferred Relationships & Collections

The inferred relationship type has no display requirement as it will not be used in a document model to render information. It simply implies an association. As such, the inferred relationship is simply a class with a reference to the semantic class.

Having data returning methods in collection relationships would lead to ambiguity in which element is generating what, and potential synchronisation issues would be created should items be removed, i.e. is the generation still valid? Collection relationships would, therefore, only need to collate inferred relationships. Figure 15 proposes the inferred relationship and the ability to create collection relationships. It facilitates many-to-many style associations through semantic inference and document inference classes.

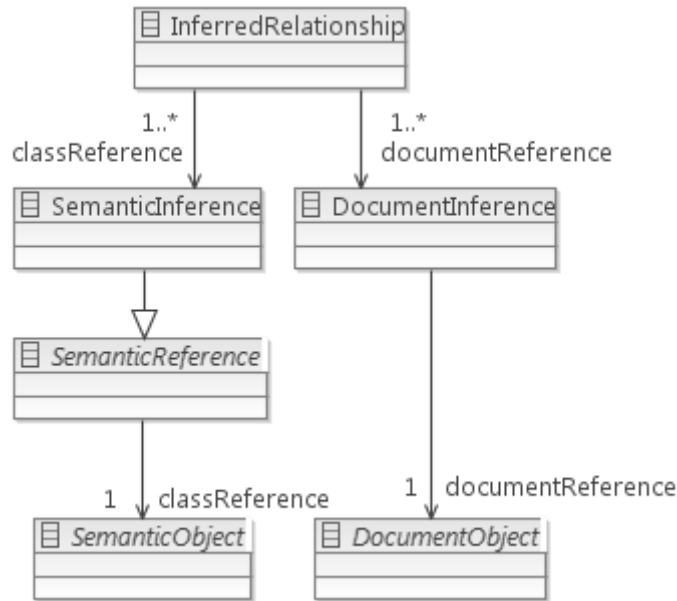


Figure 15. Inferred Relationship Proposal

All of the previously defined relationships are inherently associated with a document model object as they are required for display. The inferred relationship is not and, therefore, needs to create this reference to the document. One method would be to define an object/relationship that resides within the document model, or the semantic model. This would require heavy modification and pre-consideration would be required when defining either model, potentially polluting the actual essence of the model. This would also go against the computing concept of ‘Separation of Concerns’ [28]. By defining a container class such as the Inferred Relationship class in Figure 15, this whole concept can be abstracted out into its own supporting model containing just references to the two other models (like a many-to-many table in relational databases). This ensures that the semantic, document and inferred models stay true to their primary concern, semantics, rendering or mapping, and are not polluted by other concerns.

Within the case study there are paragraphs which describe the user’s actions that occur outside of the system in response to a system message. These paragraphs are related to the message but contain no semantic data more supplementary information. As such inferred relationships could be represented similar to Figure 16.

```

<InferredRelationship><SemanticInference
reference= "./messages.model/#_ID891"/><DocumentInference
reference= "./messages.model/#_PARA1672"/></InferredRelationship>
<InferredRelationship><SemanticInference
reference= "./messages.model/#_ID259"/><DocumentInference
reference= "./messages.model/#_PARA9432"/></InferredRelationship>
  
```

Figure 16: Example Inferred Relationships

6.5 The Semantic & Document Objects

Throughout the definitions discussed above the concept of an abstract Semantic Object is used as a terminal point in all definitions. We propose a simplistic method of achieving this through the definition of the class as an abstract inheritable class. To make a semantic class available to be referenced through our proposed definitions, the semantic model would need to be updated such that the class inherits from this new abstract Semantic Object class. This would allow semantic model owners to also restrict access to elements.

Similarly, the Document Object defined for collection relationships can be achieved in two ways, modifying the base document model such that document model objects also have a reference attribute to the abstract semantic class, or defining an abstract document class that document model elements can inherit from. We favour the construction of an abstract Document Object class that document elements inherit from.

6.6 Final Proposal

Analysing the proposed relationships as a group allows for a level of abstraction which will aid integration with documents. Mapped substitutions can be seen as an extension of direct substitutions as they both require the class attribute. All semantic substitutions have a requirement to relate to a semantic object such that the final proposal defined in Figure 17 defines an abstract class semantic reference which facilitates the abstraction of the semantic association.

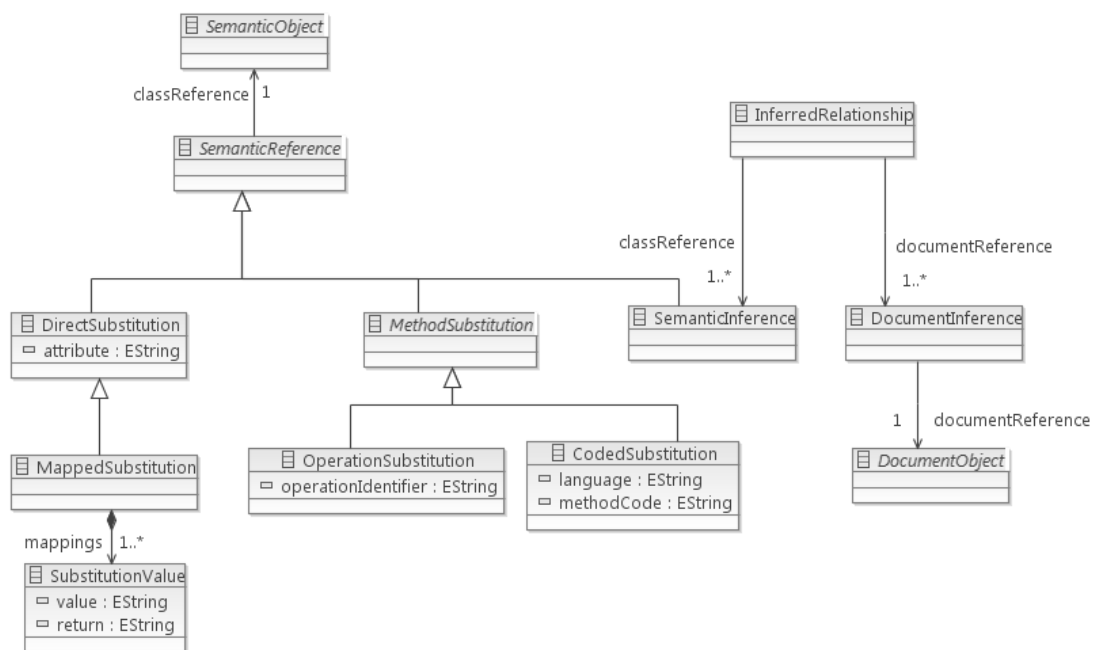


Figure 17. Complete Hybrid Linking Meta-Model

7 INTEGRATION WITH DOCUMENT MODELS

At BAE Systems, for proof of concept and other legacy reasons beyond the scope of discussion in this paper, a custom document model was developed to facilitate the capturing

and rendering of the military standard. The document model itself can be found in a previous paper [29]. With the exception of Inferred Relationships, all other relationships are concerned only at the level of a sub-sentence.

A major design decision on defining these relationships was not to tightly couple their implementation to this BAE Systems proprietary document model. Therefore, to incorporate the relationships, they can be used with any document model that, at least, facilitates the breakdown of sentences into a collection of utterances or text parts. Figure 18 demonstrates the integration of the relationship into the BAE System model at the level of text part. In performing this integration we actually propose another abstract class, Semantic Substitution. This abstract class represents all hybrid relationships that will return a text string, such as substitution relationships. The abstract class makes it easier to factor out a semantic inference as being a viable option as a text part.

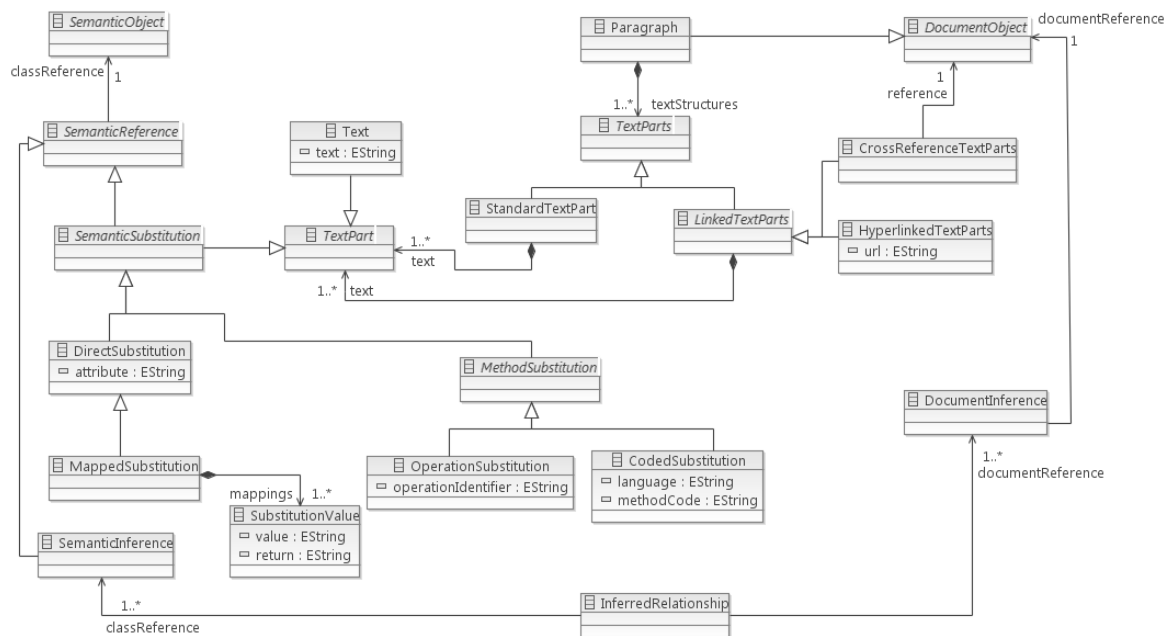


Figure 18. Integrated Hybrid Model

With this proposal, the document model itself contains the semantic relationships (excluding Inferred Relationships). Therefore, the document must contain the instances of the semantic relationships. This raises the philosophical question as to whether the document should contain the hybrid links. One view would be ‘yes’, as all the semantic substitution objects are simply a construct for generating the rendering, which, in essence, is the primary concern of the document model. The other option would be to define a new document model class which is simply a pointer to semantic substitution instances in a separately maintained, hybrid model. The hybrid linking is now separated from the document model, but the document model still needs an element of change to facilitate the reference.

We favour the first view as we consider the purpose of the document model to be the definition of the rendering pattern for the information. As such, the semantic substitution would fall under this concern and be suitably modeled here under the concept of ‘Separation of Concerns’. The Inferred Relationships which do not define any rendering can be separated out into another model/package.

As it stands, another key factor in the definition of these models is the intention for them to be utilisable as a bolt-on package to existing document models. This has been designed-in through inheritance. As a semantic model would need to be custom developed for each domain/document, extending this area is not considered a disadvantage. The main requirement here would be extending classes to inherit from the Semantic Object class such that they can be referenced. The document is a potential problem for this bolt-on concept. To reference Document Objects, the document class would need to be modified such that document classes inherit from the Document Object abstract class. Within the BAE System model this is not a problem as the document model already has a concept “Element”, which all referenceable document objects inherit, so the Document Object class is replaced by a reference to the actual document model class “Element”.

The Document Object is not the only issue, the Semantic Substitution class also needs to inherit from a document class such that it becomes an option when using the document model. The class would need to have no required fields other than a unique identifier and type. In the BAE Systems model, we defined an abstract class “TextPart” from which Text is a child node. We can then make use of this abstract class as the inheritable object for our hybrid meta model. By utilising this abstract class the semantic substitution becomes available as part of building sentences in the document model. Other document models may not have this level of abstraction. If a text part/utterance is defined as a collection of text objects but the text object is left blank, our Semantic Substitution object could inherit from this, maintaining the idea of a bolt-on package. If this is not possible, then another document model would be required or the existing document would need to be ‘tweaked’ to inject the required abstraction to allow for null text parts.

If these changes are simple, as in the BAE Systems model, then changes are only required to the hybrid meta model i.e. replacing Document Object and inheriting Semantic Substitution. In these cases, the meta model does act as a bolt-on feature where no change is required to the document model. A custom rendering agent would still be required such that the data from these hybrid links can be displayed when trying to render the document.

Therefore the complete requirements for a compatible document model are:

1. It should, at least, facilitate the breakdown of sentences into a collection of utterances or text parts.
2. It has an object that all document elements are inherited from, or it can be modified such that the document class can inherit from a new abstract class.
3. It has a class at the text part level that does not have any required attributes beyond a unique id.

7.1 DITA & DocBook

DITA or Darwin Information Typing Architecture and DocBook are both OASIS standards for technical document structure modeling [30, 31]. DITA, in particular, is highly customisable through the use of specialisations [32]. The keyword object of DITA appears to be a suitable point for our proposed Semantic Substitution to inherit. DocBook has a significant number of elements that facilitate the splitting of a paragraph into smaller elements from which we could inherit. With regards to the Document Object, it is not clear if this could be achieved through reference to existing elements. Both appear to support our idea of facilitating the hybrid models as a bolt-on package from the perspective of Semantic Substitution but more detailed analysis would be required to assess the integration of the inferred relationships.

8 EVALUATION

The proposed solution does achieve many of the anticipated benefits discussed in Section 3. Although some of the benefits could be considered achievable through independent models, the novelty of this approach comes from its coupling to the written document. Capturing the interpretation of complex prose is one such example.

The problem of capturing the interpretation can be solved through the sole use of a semantic model. Using just a semantic model introduces a second source of information and duplication but, by using the relationships defined, elements of the duplication and hence synchronisation issues are reduced. The main benefit of this approach is through the inferred relationships, as sections of semantics can be clearly linked to sections of prose. This will enable novice users of the document to trace a complex paragraph to a formal semantic model. It should help reduce ambiguity through the interpretation of prose by different engineers as there is one formal interpretation underlying it.

The reduction of ambiguity aids the objective of resolving/preventing inconsistencies. By having the relationships, any change to an underlying semantic value is immediately reflected in each instance within the document model. This means changes are not missed by having to go page by page to find reference to the item. This one point of change also meets the objective to make updating information easier. Even the inferred relationships help as, before a change is made, it can be inspected to see where it is utilised within the document and any prose changes can then be made.

It, unfortunately, does not directly improve the situation of resolving missing data, beyond the requirement that the data must exist to be related. The process of generating the required semantic model does aid this, however. In reading the prose to generate the model or starting from a model, it is hoped an engineer would be able to identify any missing elements.

Searching is also only slightly improved. The semantic model can be searched and, given the relationships, the result can then be returned with links to elements within the document model. This does provide new search methods, however it can make searching harder. As the data no longer lives directly within the prose, searches from outside of the rendering application are much harder i.e. a system search within files.

As with the searching, the relationships do provide alternative navigation structures through navigating the semantic model and then tracing the relationships back to sections of the document. The biggest potential of this solution is the objective of performing advanced document comparison. As a document is supported by an underlying semantic model, the comparison can be made at semantic level rather than the textual level. It is not a benefit in its own right, however, as this comparison could be achieved by maintaining a separate semantic model, but the tight coupling gives the appearance to the user that the comparison is operating on the document.

The final objective to discuss is the use of model execution facilities. This objective is only achievable through using the semantic model and can be achieved with or without the use of hybrid models.

In summary, although many of the benefits of this approach can be achieved using independent semantic models and documents, this approach provides benefits by combining them into one. The biggest benefit, compared to maintaining independent document and semantic models, is reducing duplication and synchronisation issues. It also potentially hides the semantics from users with little modeling background, enabling them to utilise the benefits of models whilst maintaining their user friendly document view but having advanced options, such as semantic comparison, available.

9 CONCLUSION

We have presented a framework of relationships which allows for the relationship of objects in a semantic model to be represented in a document such that the information can be rendered without the need to maintain the information both in a document form and a semantic modeled form. This framework facilitates the achievement of many of the benefits anticipated in Section ‘What is the Purpose’. Although much of the benefit can be achieved through using independent models, a significant improvement of this technique is the close coupling which reduces duplication and synchronisation.

9.1 Future Work

This is still an active area of research and there is much that is scheduled for investigation. Of primary importance to the BAE Systems project is a demonstration of the exploitation of the ideas. A proof of concept application which can showcase the anticipated benefits would provide business support for extending and improving on the research.

Another major business driver is behavioural integration, analysing how we can incorporate behavioural modeling, currently under development, into this technique and produce a rendering of this behaviour.

Open source document model integration is another area of interest. Although we have briefly discussed the integration into DITA, if it can be demonstrated that this can easily be applied to the mainstream document models, it has greater benefit for the wider community.

9.2 The Success of the Project so Far

The project has developed an evolving model of substantial elements of the MIL-STD-6016C. The research presented in this paper has identified and successfully defined a set of relationships that allow for the data to be maintained in this semantic model whilst being utilised in our document model. These relationships are defined such that they can be utilised as bolt-on enhancements requiring little modification to the existing document and semantic models.

10 ACKNOWLEDGEMENTS

This paper would not have been possible without the help and assistance provided by Julian Johnson of BAE Systems and Chris Holmes before he left BAE Systems for pastures new. Chris and Julian were involved in the very early work of BAE Systems document and subsequent domain modeling. Julian also acted as the project lead for BAE System.

11 BIBLIOGRAPHY

- [1] U.S. Department of Defense, Department of defense interface standard – tactical data link (tdl) 16 – message standard, Tech. Rep. MIL-STD-6016C, US Department of Defense (2004).
- [2] NATO, Tactical data exchange – link 16, Tech. Rep. STANAG No. 5516, NATO Standardization Agency (2006).
- [3] NATO, Technical characteristics of the multifunctional information distribution systems (mids), Tech. rep., NATO (2001).
- [4] C. Holmes, J. Johnson, Tactical data link – domain dictionary, Tech. Rep. SEIC RP 0393, BAE Systems (2005).
- [5] B. Zeigler, Simulation-based testing of emerging defense information systems, Presentation [Available at: <http://www.acims.arizona.edu/PUBLICATIONS/Presentations/AuburnTalk.ppt>] (April 2006).
- [6] J. Guo, A. Takada, K. Tanaka, J. Sato, M. Suzuki, T. Suzuki, Y. Nakashima, K. Araki, H. Yoshihara, The development of mml (medical markup language) version 3.0 as a medical document exchange format for hl7 messages, *Journal of Medical Systems* 28 (2004) 523–533.
- [7] M. Palmirani, G. Contissa, R. Rubino, Fill the gap in the legal knowledge modelling, in: G. Governatori, J. Hall, A. Paschke (Eds.), *Rule Interchange and Applications*, Vol. 5858 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2009, pp. 305–314.
- [8] G. Antoniol, G. Canfora, G. Casazza, A. De Lucia, E. Merlo, Recovering traceability links between code and documentation, *Software Engineering*, *IEEE Transactions on* 28 (10) (2002) 970 – 983.

- [9] A. Marcus, J. Maletic, Recovering documentation-to-source-code traceability links using latent semantic indexing, in: *Software Engineering, 2003. Proceedings. 25th International Conference on*, 2003, pp. 125 – 135.
- [10] B. Ramesh, M. Jarke, Toward reference models for requirements traceability, *Software Engineering, IEEE Transactions on* 27 (1) (2001) 58 –93.
- [11] G. Spanoudakis, A. Zisman, E. Pérez-Miñana, P. Krause, Rule-based generation of requirements traceability relations, *Journal of Systems and Software* 72 (2) (2004) 105 – 127.
- [12] S. Winter, S. Nittel, Formal information modelling for standardisation in the spatial domain, *International Journal of Geographical Information Science* 17 (8) (2003) 721–741.
- [13] R. H. Dolin, L. Alschuler, C. Beebe, P. V. Biron, S. L. Boyer, D. Essin, E. Kimber, T. Lincoln, J. E. Mattison, The hl7 clinical document architecture, *Journal of the American Medical Informatics Association* 8 (6) (2001) 552–569.
- [14] R. H. Dolin, L. Alschuler, F. Behlen, P. V. Biron, S. Boyer, D. Essin, L. Harding, T. Lincoln, J. E. Mattison, W. Rishel, et al., HI7 document patient record architecture: an xml document architecture based on a shared information model., in: *Proceedings of the AMIA Symposium*, American Medical Informatics Association, 1999, p. 52.
- [15] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, A. S. Shvo, HI7 clinical document architecture, release 2, *Journal of the American Medical Informatics Association* 13 (1) (2006) 30–39.
- [16] R. H. Dolin, L. Alschuler, Approaching semantic interoperability in health level seven, *Journal of the American Medical Informatics Association* 18 (1) (2011) 99–103.
- [17] A. Hasman, et al., HI7 rim: an incoherent standard, in: *Ubiquity: Technologies for Better Health in Aging Societies: Proceedings of MIE2006*, Vol. 124, Ios PressInc, 2006, p. 133.
- [18] C. N. M. Gunther SCHADOW, The hl7 reference information model under scrutiny, in: *Ubiquity: technologies for better health in aging societies: proceedings of MIE2006*, Vol. 124, Ios PressInc, 2006, p. 151.
- [19] B. R. L., W.-H. A.T., <http://www.ingentaconnect.com/content/routledg/rjit/1996/00000011/00000003/art00005>A critical perspective on action research as a method for information systems research, *Journal of Information Technology?* 11 (3) (1996) 235–246. <http://www.ingentaconnect.com/content/routledg/rjit/1996/00000011/00000003/art00005>
- [20] I. Benbasat, D. K. Goldstein, M. Mead, <http://dx.doi.org/10.2307/248684>The Case Research Strategy in Studies of Information Systems, *MIS Quarterly* 11 (3) (1987) 369–386. <http://dx.doi.org/10.2307/248684> doi:10.2307/248684. <http://dx.doi.org/10.2307/248684>

- [21] R. Goede, C. de Villiers., The applicability of grounded theory as research methodology in studies on the use of methodologies in is practices., in: SAICSIT '03, 2003, pp. 208–217.
- [22] J. Iivari, A paradigmatic analysis of comtemporany schools of is development, *European Journal of Information Systems* 1 (4) (1991) 249–272.
- [23] A. Salminen, K. Kauppinen, M. Lehtovaara, [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199707\)48:7<644::AID-ASI12>3.0.CO;2-V](http://dx.doi.org/10.1002/(SICI)1097-4571(199707)48:7<644::AID-ASI12>3.0.CO;2-V)Towards a methodology for document analysis, *Journal of the American Society for Information Science* 48 (7) (1997) 644–655. [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199707\)48:7<644::AID-ASI12>3.0.CO;2-V](http://dx.doi.org/10.1002/(SICI)1097-4571(199707)48:7<644::AID-ASI12>3.0.CO;2-V) doi:10.1002/(SICI)1097-4571(199707)48:7<644::AID-ASI12>3.0.CO;2-V. [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199707\)48:7<644::AID-ASI12>3.0.CO;2-V](http://dx.doi.org/10.1002/(SICI)1097-4571(199707)48:7<644::AID-ASI12>3.0.CO;2-V)
- [24] C. Holmes, J. Johnson, A. Riaz, Tactical data links – a description of the supporting models, Tech. Rep. SEIC-RP-0531, BAE Systems (2007).
- [25] D. Steinberg, F. Budinsky, M. Paternostro, E. Merks, EMF: Eclipse Modelling Framework, 2nd Edition, Eclipse Series, Addison-Wesley Professional, 2008.
- [26] IBM and others, <http://download.eclipse.org/modeling/emf/emf/javadoc/2.6.0/org/eclipse/emf/ecore/package-summary.html>org.eclipse.emf.ecore (emf javadoc), this is an electronic document. Date of publication: Jun 23, 2010. Date retrieved: Nov 11, 2010 (2010). <http://download.eclipse.org/modeling/emf/emf/javadoc/2.6.0/org/eclipse/emf/ecore/package-summary.html>
- [27] D. S. Kolovos, R. F. Paige, F. A. Polack, The epsilon object language, in: Proc. European Conference in Model Driven Architecture (EC-MDA) 2006, Vol. 4066 of LNCS, Bilbao, Spain, 2006, pp. 128–142.
- [28] W. L. Hürsch, C. V. Lopes, Separation of concerns, Tech. rep., College of Computer Science, Northeastern University (1995).
- [29] D. Clowes, R. Dawson, S. Proberts, J. Johnson, C. Holmes, Pilot studies in using the semantic knowledge of information in large technical documents to aid user navigation, in: D. Harorimana, D. Watkins (Eds.), *The 9th European Conference on Knowledge Management: Eckm 2008*, Academic publishing, 2008, pp. 129–138.
- [30] OASIS, <http://www.docbook.org/specs/docbook-5.0-spec-os.html>The docbook schema version 5.0 (November 2009). <http://www.docbook.org/specs/docbook-5.0-spec-os.html>
- [31] OASIS, <http://docs.oasis-open.org/dita/v1.2/spec/DITA1.2-spec.html>Darwin information typing architecture (dita) version 1.2 (December 2010). <http://docs.oasis-open.org/dita/v1.2/spec/DITA1.2-spec.html>

[32] E. Henum, <http://www.ibm.com/developerworks/xml/library/x-dita5/index.html> Specializing domains in dita, Tech. rep., IBM (September 2005). <http://www.ibm.com/developerworks/xml/library/x-dita5/index.html>



Darren Clowes is an Engineering Doctorate student at Loughborough University. He received his masters degree in Computer Science at Loughborough University and recently has been working at the Systems Engineering Innovation Centre in collaboration with BAE Systems on standards modeling and exploitation.



Ray Dawson is Professor of Knowledge Management at Loughborough University, UK. He obtained his bachelor's and masters degrees from Nottingham University before entering Plessey Telecommunications in 1977. At the company he developed an interest in software engineering processes and the working methods for information systems development as practiced in industry. Since 1987 he has continued this interest in industrial working methods at Loughborough University, working with companies to improve their software processes and information systems. Ray Dawson is now Deputy Head of the Computer Science Department and Associate Dean for Teaching of the School of Science at Loughborough University.



Steve Probets is a lecturer who has recently moved from the Department of Information Science to the Department of Computer Science at Loughborough University. He is programme tutor for the Department's 'Information Management and Computing' programme. His research interests include electronic publishing and document engineering and he has supervised a number of PhD candidates looking into mechanisms to improve the usability of both electronic documents and websites. Other research interests include the evolution of scholarly publishing.