Fall Simulation Interoperability Workshop
Orlando, FL, September 2006

Merging Protocols, Grammar, Representation, and Ontological Approaches in Support of C-BML

Dr. Andreas Tolk, Saikou Diallo, Chuck Turnitsa
Virginia Modeling Analysis & Simulation Center
(VMASC)
Old Dominion University
Norfolk, VA 23529
atolk@odu.edu

Keywords:

Atomic/Composite/Aggregate Services, C-BML, Composability, Grammar, Ontology, Web Services, XML

ABSTRACT: *The product nomination accepted for the Coalition Battle Management Language (C-BML) Product Development Group identifies four phases to develop a C-BML standard. The initial phase will describe a data model as a subset of the Command & Control Information Exchange Data Model (C2IEDM), an information exchange content and structure specification in the form of an XML schema, and an information exchange mechanism specification using the Web Services Description Language (WSDL). The following two phases will introduce a grammar (syntax, semantics and vocabulary) to define the information exchange content and structure specification. The objective is to formalize the definition of tasks such that they are rigorous, well documented, and parseable. The tasking and reporting grammar sections will make the information exchange more flexible. The ultimate objective is the definition of an ontology-based standard, which will develop a battle management ontology to enable conceptual interoperability.*

This paper summarizes the various approaches, identifies applicable research in neighbored domains, and recommends a holistic approach for all four phases.

1 Introduction

Coalition Battle Management Language (C-BML) is defined as the unambiguous language used to command and control forces and equipment conducting military operations and to provide for situational awareness and a shared, common operational picture. This definition was proposed in [1] and accepted by the Study Group on C-BML [2] for the Product Nomination currently being the basis for the work of the Product Development Group for a C-BML SISO Standard.

In summary, C-BML targets unambiguous information exchange for command and control (*tasking*) and situational awareness (*reporting*) between C2 systems, M&S systems, and robotics.

The standard is going to be developed in phases:

- The first phase will define information exchange between the targeted systems – command and control system, simulation system, and robotics – using web services. The tag set used for information exchange will be based on the Command and Control Information Exchange Data Model (C2IEDM) or respectively the Joint Consultation, Command and Control Information Exchange Data Model (JC3IEDM), which is the C2IEDM successor and a real extension using the same hub and core entities and relations [3].

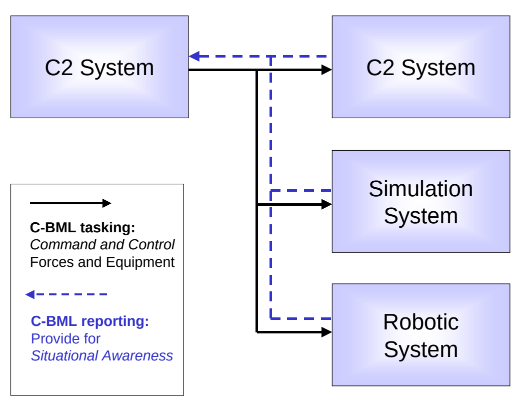

Figure 1: Information Exchange using BML

- The second and third phase will move the work forward from feasibility to extensibility and bi-directional use. Objective is to introduce "generative" grammars: Instead of using enumerations in an XML schema, tasking and reporting grammars will be used to generate valid C-BML expressions. The C2IEDM/JC3IEDM will remain the representation of C-BML, but extensions may be necessary.

- The next and final phase will apply methods of the ontological spectrum. Objective is to generate a C-BML ontology capturing the results in a flexible and yet standardized way enabling real composability for C-BML based services and applications.

The C-BML triangle was introduced to show the different views necessary to describe C-BML. It comprised the three components C-BML Protocol, C-BML-Representation, and C-BML Doctrine. In [4], we introduced our view on where C-BML Grammar and C-BML Ontology need to find their place, namely as connecting elements between the three main components, resulting in the "triangle with five sides" shown in Figure 2.

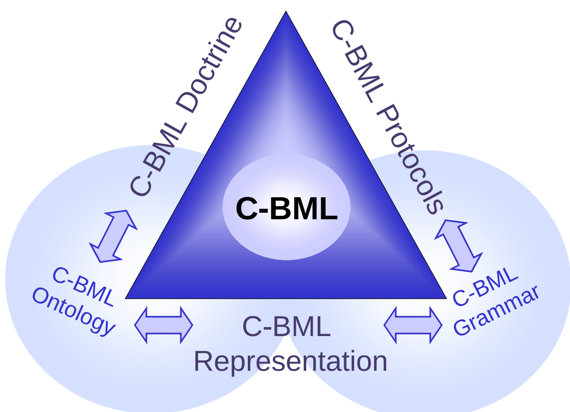

Figure 2: Components of C-BML

The awarded paper [5] envisions a slightly different view in which grammars will play a much greater role than "simply connecting" protocols and representation. This led to some discussions in the C-BML community.

With this paper, we want to show that both recommendations, [4] and [5], are actually different sides of the same medallion and are both supporting the standardization work. However, we will also show that a broader view than legacy system analysis is needed in order to result in a stable and yet extensible standard for C-BML, that allows the application of net-centric ideas as already demonstrated on presented in [6].

2 Components of C-BML

In this section, we will deal with the five components of C-BML: protocols, grammar, representation, ontology, and doctrine. We will give definitions for the terms and explain how they can be used within C-BML.

2.1 C-BML Protocols

In order to communicate the reporting data from the systems into C-BML and the tasking data from C-BML to the executing systems, communication protocols are needed. The *C-BML Protocol* standardizes the way the description of the executable tasks and assigned executing military means is transported between the systems using C-BML.

Proposed Definition: C-BML Protocol specifies the interfaces supported by the C-BML standard, the structure of the information, and how to access interfaces and embedded service access points.

The first phase is currently using web services, and consequently the Web Service Definition Language (WSDL) is used to define the protocol data. WSDL comprises data type messages for data type definitions, operations port type bindings for abstract operations, and port services for service bindings. The data type definitions are XML schemas. A port type is a logical grouping of operations. A port is used to expose a set of operations (as specified by the port types) using a given transport mechanism. Service bindings map messages and operations to transport mechanisms needed for the communication when using the services. WSDL uses XML schemas to describe what input parameters are needed, what functions can be called, what output parameters have to be expected, and which protocols have to be used to deliver the input, to invoke the function, and to receive the output.

It should be pointed out that other protocols are also possible. Of particular interest are natural language processing interfaces, such as envisioned in the Canadian contribution to [2], which deals with Natural Language Scripting (NLS). NLS fulfills the criteria of the definition and could become an interface specification for direct human interaction, if the human speaks C-BML.

2.2 C-BML Grammar

A *C-BML Grammar* is used to allow extensible configuration of C-BML Protocols enabling the information exchange based on the operational needs as captured by the C-BML Representation. It also allows the use of different structures or varying protocol means with the same C-BML Representation. A broader view on how grammars can support C-BML is given in [5], in which Schade and Hieb use the early work of Chomsky to motivate their work. We follow their definition.

Proposed Definition: C-BML Grammar is a set of words comprised in a lexicon and a set of rules to generate grammatically valid sentences out of these words.

It should be understood that these "sentences" must be equivalent to the information structure that is specified by the protocol, or – as envisioned by the C-BML study group – that the C-BML tasking and reporting grammars will be used to specify the information structure.

2.3 C-BML Representation

The *C-BML Representation* structures the information. This is currently done using a data model. To be more precise, the Command and Control Information Exchange Data Model (C2IEDM) is used. We strongly disagree with evaluation conducted in [5] that *"the missions listed in the C2IEDM (in the "action-task-activity-code" enumerated values) are merely words with a vague textual description."* This list of words is the lexicon used for the grammars and the constructs exchanged via the interfaces. A data model captures much more, as it identifies entities, attributes of entities, valid values for the attributes – often in the context of other assignments –, and associations. This view is reflected in the working definition.

C-BML Representation is a data model, that groups the allowed values as attribute values to attributes that characterize entities that can be associated with each other. The structure of the data model allows generating the information structure specified by the protocol.

The strong connection between C-BML and supporting data models was documented in [7]: the operational research conducted by the US Army Battle Management Language researchers described in [1] was mapped to the Multi-Source Database (MSDB) utilizing the Joint Common Database (JCDB). As stated in [7]: *"concerning the data, the MSDB is the central piece; it contains all the verbs, nouns, adjectives, adverbs and concepts of BML. Atlantic Consulting Services, Inc and Northrop Grumman developed the initial version of MSDB. The MSDB employs BML data elements and BML concepts. In order to cope with all information exchange requirements, 113 tables were added to the JCDB."* When the same information was mapped to C2IEDM, only four new tables had to be introduced to capture the same information.

2.4 C-BML Ontology

C-BML *Ontology* will formulate an exhaustive and rigorous conceptual schema of the Battle Management domain. As stated by Obrst, who belongs to the leading scientists in the domain of ontologies, it is agreed in practice that an ontology should contain at a minimum not only a hierarchy of concepts organized by the subsumption relation, but other 'semantic relations' that specify how one concept is related to another. The main purpose is the definition of entities and their relationships [8]. Therefore, the role of the C-BML Ontology is to conceptualize the C-BML Doctrine and to relate this to the C-BML Representation.

Proposed Definition: C-BML Ontology is a formal specification of the conceptualization of the Battle Management domain as derived from C-BML Doctrine.

As such, C-BML Ontology comprises all terms defined in the lexicon, the concepts underlying this terms, the relations, etc. One of the main differences between data models and ontologies is that data models represent the symbolic information (structure/syntax of data) while ontologies deal with the contextual information (meaning/semantics of data) and the conceptual and representational level of data elements. We will cope with this issue more in section of this paper. As Schade and Hieb state for data models – and hence C-BML Representations – in [5]: *"C2IEDM is for exchanging facts, but not for communicating meanings and intentions. This, however, is what a language is for."* Again, we agree with this statement; however, we are convinced that only ontological methods can overcome this challenge.

2.5 C-BML Doctrine

As stated from the first reports on Battle Management Language efforts on: every term used within C-BML must be unambiguously defined and must be rooted in military doctrine. In other words, the doctrinal view must comprise a glossary that comprises each term and its unambiguous definition and express the concepts behind the terms and associations. Furthermore, the relations between concepts and terms must be described and executable tasks, their context within missions, and all other relevant information describing the Battle Management domain.

Proposed Definition: C-BML Doctrine is a rigorous documentation of the Battle Management domain in the language of the user, such as Field Manuals, which is sufficient to support a formal specification of its content.

As documented in [1], the US Army BML team started by analyzing more than 70 doctrinal manuals related to tasking and reporting. They began with general manuals as the Field Manual 3-0 on Operations or the Universal Joint Task List as published by the Joint Staff and included the field manuals of branches of the Army, such as Field Artillery, Air Defense Artillery, Engineers, Military Police, and many more manuals down to the platoon level.

In summary of this section on C-BML Components, it is understood that several layers of information in form of terms, concepts, entities and attributes, associations and relations, and the context of their use are interwoven in the five components. These views are mutually supportive and must be approached in a holistic way otherwise structural variances will occur.

3 "Divide and Conquer" or "Partnering for Performance"

Within this section, we want to show that the holistic view is necessary and possible to ensure the success of C-BML standardization and application. We will show that all five components are connected and have their special role in the orchestration. In order to do this, the ontological spectrum, the ISO/IEC 11179, and data models need to be introduced first.

3.1 The Ontological Spectrum

It may be of interest that the ontology domain identifies lexicons and dictionaries as part of the ontological spectrum [8], which can be divided into five categories:

- *Controlled Vocabularies* enumerate all allowed terms and their meanings completely. All terms are well-defined and controlled by a common registration authority. They deal with terms.
- *Thesauri* are controlled vocabularies arranged in a known order and structured so that equivalence, homographic, hierarchical, and associative relationships among terms are displayed clearly and identified by standardized relationship indicators. They deal with terms.
- *Taxonomies* are tree structures of classifications for terms. The root nodes apply to all objects; nodes below these roots are classifications that are more specific. Taxonomies can also be used to introduce the idea of concepts and implementing terms.
- *Ontologies* are formalizations of specifications of conceptualizations. Ontologies describe all the information captured in thesauri and taxonomies plus contain additional relationships and rules, such as theorems and regulations, within the domain concept. They focus on the contextual information (meaning/semantics of data) of data, which are the concepts, but have also references to the structure/syntax of data, which are the terms.
- *Logical Models* are the strongest semantics in the ontological spectrum. First order logic and modal logic are examples.

It should be mentioned that the methods of the ontological spectrum were evaluated with regards to the objectives of the Semantic Web. The two main objectives are the unambiguous definition of information describing content of the Web as well as supporting the search for this content. In particular, the definition ideas directly support the C-BML effort.

Furthermore it should be noted that no standards have been generally accepted for ontologies. No standard format to capture ontologies is established. In addition, no standard to display ontologies for human users is established.

3.2 Syntax and Semantic

We already realized that it is necessary to distinguish between concepts (meaning/semantics of data) and terms (structure/syntax of data).

ISO/IEC 11179 standard on Metadata Registries [11] introduces the following terms to describe a registry:

- *Conceptual Domains* define sets of categories, which are not necessarily finite, where the categories represent the meaning of the permissible values in the associated value domains. They comprise symbolic information on the conceptual level. An example is that we want to communicate the concept of a main battle tank.
- *Data Element Concepts* describe the contextual semantics, i.e., the kinds of objects for which data are collected and the particular characteristic of those objects being measured. They comprise the contextual information on the conceptual level. They define what properties are needed to describe and identify a concept, such as caliber and turret specifications characterizing the main battle tank.

- *Value Domains* comprise the allowed values for an associated data element. Value domains comprise symbolic information on the representation level. These are the allowed values for data elements. At this level, we specify if we measure the caliber in inches or centimeters (unit-of-measure) or list enumeration of allowed values.

- *Data Elements* are the basic containers for data as used in data models. Data elements comprise contextual information on the representation level. Data models comprise these data elements, which are the entities and relations as used by the supporting IT system.

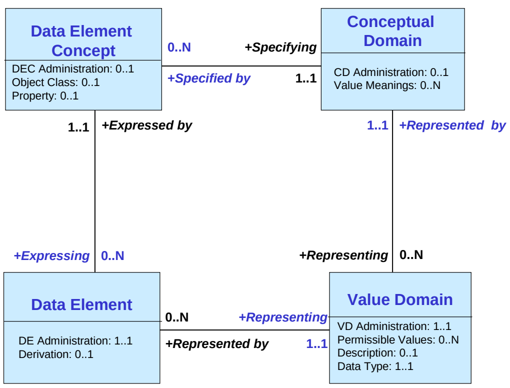

Figure 3: ISO/IEC 11179 Metadata Registry

ISO/IEC Figure 11179 distinguishes four different domains that are often merged and blurred in discussions: What we say (data elements), how we say it (value domain), what we are talking about (data element concept), and in what context we are talking about it (conceptual domain).

These distinctions become very important in the international domain: if nations agree on concepts and representing data, the value domain can be used to support different units-of-measure and different enumerations of terms (such as English, French, and German vocabularies).

These ideas must be followed in the C-BML efforts rigorously. It is necessary to clearly distinguish between the "levels of meaning" we are focusing the standardization efforts on. Table 1 summarizes the ideas. In order to mo In order to model a challenge, we must perceive it first. Our perception is already a subset of the real world, and in C-BML different nations may already have different perceptions (such as on what Armed Forces are eligible to do). This perception must be captured in concepts, which happens by abstracting and conceptualizing. Adding modeling, we add concrete meaning in a context, leading to terms, which are the vocabulary of the communities of interest. When they are implemented, they become entities. When working on C-BML standards the PDG must make sure to make explicit if these standards apply to concepts, terms, or entity.

Table 1: Referents, Concepts, Terms, and Entities

Activity	Result
Perception (mental picture of the world)	(real world) *Referent*
Abstraction and Conceptualization (conceptual picture of a subset of the world)	*Concept*
Modeling (make a model of how the concept works)	*Term*
Implementing (make the model executable)	*Entity*

3.3 Data Models

Data Models are used to group terms in a meaningful way for one or more applications. Without giving the usual definitions in detail, data models distinguish between attributes, entities, and associations. Attributes are the data elements that capture terms. Each term must be a valid attribute value in order to be captured in a data model. Entities group of attributes into a meaningful unit. Associations link several entities together; associations can also be attributed.

These artifacts used to build data models – attribute values, attributes, entities, and associations – can be used to implement tools to capture thesauri, taxonomies, or ontological structures. It is therefore not the data model technique but its application that defines on what level a data model is used.

In the context of C-BML, the C2IEDM/JC3IEDM is primarily used as a common reference model capturing the conceptual domains and characterizing data element concepts of C-BML. Although related as well as our own research showed that C2IEDM/JC3IEDM has several weaknesses when used as the basis for an ontology, these are minor and do not disqualify the overall approach.

3.4 Controlled Vocabularies in Protocols, Grammar, Representation, Ontology, and Doctrine

Concerning the authors, the most obvious common factor in all five C-BML components is the lexicon of usable terms. The NATO Code of Best Practice [10] identifies the use of glossaries for Command and Control assessment as a significant success factor, such as the NATO Glossary of Terms [11], that can be downloaded by every interested scientist. National extensions are also published and used.

Controlled vocabularies are of high importance for the C-BML effort, as they are related to all five components:

- C-BML Doctrine is used to extract the terms and meaning from user documentation.
- C-BML Ontology captures the results of C-BML Doctrine in a formal specification, in which each term and its concept is modeled.
- C-BML Ontology also gives meaning to entities, associations, attributes, and attribute-values used in the C-BML representation.
- C-BML Representation uses well-defined numbers and controlled vocabularies as attribute values.
- C-BML Representation must be able to capture all controlled vocabularies used in the lexicon part of the C-BML Grammar.
- C-BML Grammar consists of a lexicon with terminals (terms of the controlled vocabulary) plus non-terminals (starting symbol plus interim placeholders), and a set of rules that can be used to generate valid sentences made up by terms in a grammatically correct sequence.
- C-BML Protocol reads and writes grammatically correct sequences using the elements of the controlled vocabulary.

It should be pointed out that the process of sense-making – mapping syntax and semantics or structure of data and meaning of data to each other – is handled by the C-BML Ontology. Our recommendation is to do this in form of a formal specification based on the controlled vocabulary and identified grammatical constructs under consideration of the representing data model. Furthermore, it should be clearly distinguished if we are talking about concepts (elements of the ontology describing the ideas and meaning), terms (instantiations of concepts in the application domain, such as C-BML vocabularies), and entities (instantiations of terms within the participating applications, such as C2 systems, simulation systems, and robotics).

3.5 Grammar and Representation

In their introduction to [5], Schade and Hieb make the case that data models have an implicit grammar (although, as cited in section 2.3, they later refer to the attribute value enumerations as "*merely words with a vague textual description*"). The authors agree that it is necessary to make the implicit grammar explicit, but this must conducted in an aligned way with the other goals of C-BML and cannot be conducted as an isolated task. To this end, implicit grammars of data models – and guiding ontologies – need to be made explicit in order to enable efficient alignment. The following list of assertions is a recommendation for consideration by the authors:

1) Valid expressions generated by a grammar must be valid selections from the database
2) A valid selection from the database (select on possible views on associated tables) must be a valid expression generated by a grammar
3) The set of all valid attribute values in a data model must be equal to the set of all terminals of a grammar (lexicon of terms)

In other words: if something valid can be extracted from the data model, but it cannot be generated by the grammar, then the grammar is incomplete; and also, if something valid can be generated by the grammar, but it cannot be selected from the database, then the data model is not complete.

One important component of data models often overseen is the guiding business rules. The JC3IEDM has over 150 pages of business rules summarized in Annex G to its current Version 3.0. These rules must be taken into account when populating the data model. They document the interdependencies between attribute values that are not captured in the structure of the model itself.[1] When the concepts of a data model are formally specified, the business rules must become part of this formalization. An ontology is more than an OWL version of the entity-relationship model, it must comprise the logic as well. The C-BML Ontology must therefore represent the entities, attributes, associations, and underlying concepts and it must deal with the business rules.

[1] The four tables mentioned in [7] that were introduced to C2IEDM to be able to capture the C-BML information could also have been established a four new business rules with enumerations.

These results are supported by current research of other scientists as well. Generally, data modeling is a process starting with the identification of data, the set of components to represent it, and the relationship among those components. Common data modeling approaches include Entity-Relationship (E-R) modeling, Object-Oriented Modeling, and more recently XML-based modeling. However, as Peng et al. point out in [15]:

> *"A data model is the grammar, vocabulary and content that represents all types of* information *stored in one format or another in a* system. *The grammar defines the relationships between elements in the system; the vocabulary defines the terminology used to describe these elements; content defines what is to be included in the system."*

In other words, grammar rules can be applied in data models in the form of primary keys, foreign keys, classifications, associations, and subsumptions, rather than exclusively enumerated as a set of production rules.

Gyssens et al. have shown how a grammar-based approach can be used to build hierarchical data models [16]. Based on the formal language definition of a grammar they offer the following definition:

> *"An information base scheme is a formal grammar G = (V, T, S, P) with V a finite set of attributes, T a finite set of constants, S a set of axioms, $S \subseteq V$, and P a finite set of productions of the form $A \rightarrow s$, where $A \in V$, $s \in (V \bigcup T)$, and each attribute appears at most once in s."*

This type of grammar is referred to as a "context-free" grammar in Chomsky's hierarchy. In order to create a data model based on a grammar, one must first create a grammatical view of the model based on the following algorithm:

- *Create a taxonomy based on a controlled vocabulary*: The controlled vocabulary is the set of terms in the data model. This vocabulary must be organized into a thesaurus then into taxonomy. These are the first three parts of the ontological spectrum.
- *Generate a set of production rules* reflecting the relationships between elements in the model: The most common way is to use an annotated directed graph to represent the system being modeled.
- *Build the grammar tree* representing the grammatical view of the model based on the production rules and the taxonomy: The grammar-based tree must be "a non empty tree the nodes of which are labeled with elements of ($V \bigcup T$) in such a way that each internal node is labeled by an attribute." [16] Figure 4 shows an example of production rules and the resulting grammar tree. The level of detail supported by a grammar tree depends on the level of abstraction of the model.
- *Build a data model representing the grammar tree*: The grammar tree is implementation independent. In addition, it is a highly structured hierarchical model that supports generalizations, typing, and aggregation.

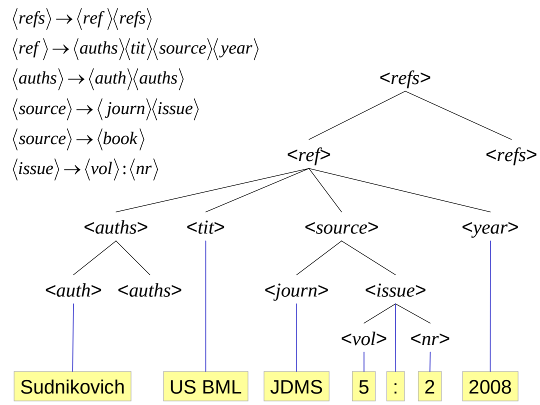

Figure 4: Production Rules and Grammar Tree[2]

Most models, however, have their underlying grammar embedded, and that is the case for C2IEDM/JC3IEDM as well. Nonetheless, it is still possible for such models to exhibit their structure in the form of semantic views. Regardless of the underlying implementation behind a model (object oriented, E-R, XML), it is possible to identify a set of business objects that are of interest to outside systems. These objects can be words (atomic concepts), phrases (composite concepts), or sentences (aggregated concepts). In terms of information exchange, which is the domain of C-BML, identifying these objects greatly improves interoperability between systems; however, business objects are not sufficient if not accompanied by enforceable business rules. These rules can be captured in a grammar based on the following algorithm:

- *Within the reference model identify the set of objects to exchange:* The set of business objects is the set of concepts that two or more interoperating systems have agreed to exchange. By analogy to the definition provided in [16] this is the set of axioms S.

[2] Figure 4 is based on Gyssens et al. [16]

- *For the set of business Objects identify the set of attribute to unambiguously exchange information:* Most interoperating systems distinguish between initialization data and information exchange data. The set of attributes in this case refers to the latter in exclusion of the former. System initialization is another part of C-BML currently investigated by the Military Scenario Definition Language (MSDL) Study Group. By analogy to the definition provided in [16] this is the set of attributes V.
- *Identify the enumeration set pertinent to the information exchange:* Each enumeration set is unique depending on the context, scope, and goal of the information exchange. It is therefore important that only relevant enumerated values are available during the information exchange. For instance, two interoperating army systems should only support the enumerated set of ground attacks. By analogy to the definition provided in [16] this is the set of constants T.
- *Generate the set of production rules:* These rules must reflect the constraints of the information exchange and must conform to the definition given in [16].

The conglomeration of these four steps results in a grammar that is particular to the information exchange needs of participating systems. *With the use of a common reference model such as C2IEDM/JC3IEDM, grammars become reusable for* ***isomorphic*** *sets of business objects.*

3.6 Categories of Rules

When the components of a language such as C-BML are revealed, it quickly becomes apparent that in order to use them to communicate unambiguously, there must be a number of different rules governing how the components come together [12]. These rules can be broadly divided up into two different, but related, categories [13].

The first category applies to the relationships between terms of the language.[3] The first set of rules deals with the questions which types of terms and/or instances of terms are related with each other and when they are valid. This is reflected by the representation of C-BML as well, normally in form of a data model. As this category of rules exists completely within the language itself, we refer to them as internal rules.

The second category applies to the structuring together of terms by external systems that are making use of the language. In this case, the communication needs of the external system determine which combinations of terms are correct. This set of rules, due to its nature of originating from outside the language, is referred to as external rules. They deal with the mapping of C-BML terms – and phrases and sentences – to application-specific entities – and composites and aggregates of them in form of understandable information groups. The first US Army specific WSDL, grouping C-BML terms in form of an US Army Task Order, is an example for such a grouping.

The close relationship between the two types of rules comes from the fact that they are both governing the combining of terms. The internal rules apply to the correct application of associations to bring terms together and forming associated terms. On the other hand, the external rules must be supported for bringing both terms and associated terms together to create unambiguous statements satisfying the communication needs of the external system.

When considering these two categories of rules in light of the definition given earlier in 2.2, we can see how these rules are connected to the ideas of the C-BML grammar. Different aspects of grammars are supported by a slightly different view of tow classes of rules given here.

- One set of rules support the construction of the terminal symbols of a formal grammar. This set of rules constructs the inner part of the grammar tree shown in Figure 4.
- The second set of rules supports the application of the grammar rules required to construct the equivalence for a non-terminal symbol. This set of rules produces the leaves of the grammar tree.

The first set of these rules can be thought of as being analogous to generate a sentence schema; the second replaces the schema placeholders with words, forming sentences understandable in the described domain.[4]

By examining the definition of a formal grammar [5], we see that it consists of a set of terminal symbols, a set of non-terminal symbols, and a set of rules. In comparing this definition to our definition for rules, we

[3] Remember the separation of concepts, terms, and entities. As long as we refer to C-BML, we talk about terms. When these terms are mapped to application-specific representations, they become entities.

[4] It is easy to make the case that these “sentences” now have to be mapped from “terms” to “entities” of the application.

can see that significant parts of the formal grammar definition already are satisfied.

When we combine the rules described above with a data model, the whole definition of a formal grammar come into view. The addition of the data model provides for all the terms and associations that the internal rules combine to create *words* (terminal symbols), and then the external rules combine those *words* into *sentences* (applying production rules to create non-terminal equivalences). The goal of the formal grammar is satisfied by this combination of rules and a data model.

3.7 Configurable Services

Within the first phase of C-BML development cycle, Web service standards are applied to a common reference model with C2IEDM/JC3IDEM acting as the basis for information exchange between systems. *However, C-BML is not another way to "speak C2IEDM/JC3IEDM," but it should be the foundation for a real language.* In order to fulfill such constraints already in the initial phase, it is important to structure the web services in a way that supports both composition and aggregation of valid expressions independent from C2IEDM specific structures without losing the normative character for the specifications of the meaning derived from the model and application rules as described above.

In order to support composition and aggregation, we recommend that C2IEDM/JC3IEDM must be accessible in a combination of three levels [17]:

- *Atomic Level*: Individual entities are described at the most basic level (tables in a database). In terms of a grammar, atomic services are the vocabulary of the language.
- *Composite Level*: At this level, services are grouped into meaningful semantic information blocks. The reference model is decomposed into views that reflect its underlying concepts (Who, What, Where etc…). Composed services are ordered groupings of semantically related atomic services using existing associations.
- *Aggregate Level*: Aggregate services are ordered groupings of semantically related composite services. These groupings reflect not only grammatically correct sentences but also semantically valid ones ("Who" is doing "What," "Who" is "Where," etc…). In addition, they may aggregate information and add information derived from the reference model, but not explicitly in the model.[5]

An extension of the model results in additional atomic services that can be composed into composite services that can in turn be aggregated into aggregate services. This process does not affect existing services but rather extends and enriches system interoperation. By defining and applying a set of production rules at the atomic level, it is possible to have a web-based grammar dynamically generating composite services that can be aggregated into grammatically correct sentences.[6]

However, ontological descriptions of each service (atomic, composite, and aggregate) are still likely to be necessary in order to ensure semantic correctness. Standards such as the Web Ontology Language (OWL) and its corollary for services (OWL-S) address this issue by providing a way to formalize ontologies of systems and services and allow automated interactions between services. This is beneficial to C-BML in another way because it provides an obvious link between the four phases of standardization described in the first section, but details on this research go beyond the scope of this paper. A good overview of the state of the art can be found in [18].

4 Recommendations for the C-BML Product Development Group

Based on the results of our research, we would like to offer the following recommendations for the C-BML Product Development Group (PDG). The objective is to avoid separate tasks and activities on C-BML Protocol, Grammar, Representation, and Ontology components of C-BML, which would create stove-piped single-technology solution. Instead, we envision a C-BML Doctrine driven standard, which is supported by all these interdependent views.

The first recommendation is to agree on a common set of terms – the ***Introduction of a Controlled Set of Vo-***

---

[5] Typical examples are codes, such as unit-identifier codes. They comprise and codify information on the structure, the place in the unit order of battle, the type of the unit, etc. In the application, this can be very useful, but in a reference model, codes are the introduction of redundant information and they can be reproduced based on the available atomic information.

[6] This approach is currently supported by the Joint Advanced Training Technology Laboratory (JATTL) of Joint Forces Command. Supported by VMASC, our industry partner Gestalt LLC implemented more than 400 web services enabling the access of JC3IEDM based concepts resulting from ISO/IEC 11179 applications.

cabularies – to be used in C-BML. The reason is the following

- When we consider the different levels for capturing ontological meaning [8], each resulting in an increasingly higher level of conceptual specificity, we see that the foundational layer is the controlled vocabulary. Without a controlled vocabulary, none of the higher levels within the ontological spectrum is possible as they all assume access to such a vocabulary.

Although we recognize that C-BML is an international standard, we recommend using the English version on the NATO Dictionary [11] as well as the definition of entities in the JC3IEDM (definitions in annexes B, C, and E) for this purpose. Entities in national applications can be derived from these internationally accepted terms.

The second recommendation is ***Concentration on the Data Model of the C-BML Representation*** to give structure to these terms. The data model is a key for the successful integration of all of the components of C-BML. Following are a few reasons for this:

- One of the benefits that a data model brings to C-BML is that it defines a controlled vocabulary. The categorization of entities provides for the classification of such a vocabulary, and the enumeration of those categories provides for the vocabulary members themselves. The addition of association allows for the building of associated-entities, with are required for the more complex categories and enumerations.
- A data model has an implicit grammar that reflects relations between entities. Explicit grammars must be based on the built in set of rules within the representation layer. By definition, grammars do not exist without controlled vocabularies that data models also provide.
- As mentioned in section 3.4, ontology for C-BML is intended to give meaning to entities, associations, attributes, and attribute-values. Much of this meaning can be captured within the data-model, especially where such meaning is intrinsic to the entities and associations of the data model. The meaning that can be captured at that level is in understanding the attributes and attribute-values that grant definition to the entities, and how the associations tie such entities together.
- The set of atomic, composite, and aggregate services cannot exist without a reference data model that serves as the basis for a common language, C-BML. The configurable services are just a way to manipulate the words of this language into meaningful sentences.
- The migration to higher representations in the ontological spectrum is possible and – as soon as methods, techniques, and the communities are ready for this – can be conducted. The need is already captured in [19].

The third recommendation is the ***Application of ISO/IEC 11179*** to capture the controlled vocabularies (terms) and the represented real world objects (concepts):

- The C2IEDM/JC3IEDM should be used to represent the concepts, as it is the view of the warfighter. It is endorsed by an increasing number of DoD organizations.
- The metadata of ISO/IEC 11179 should be used to map the vocabulary captured from C-BML doctrine and data elements of implementing C-BML enabled systems to this concepts.

The fourth recommendation is the ***Inclusion of Atomic, Composite, and Aggregate Service for the C-BML Protocol.*** This does not contradict recommendations supported so far in the C-BML study group.

- The prototypes described in [2] use services that can be modeled as aggregate services, as they use information captured in the C2IEDM, but they use slightly different structures.
- The first step shall be the definition of atomic services for all tables of the C2IEDM/JC3IEDM. This step must be followed by atomic services for all C-BML specific extensions. The development of composite and aggregate service shall be driven by the information exchange need of the participating C-BML enabled systems, such as the national contributions to the NATO MSG-048 activity described in [20].
- Industry partners implemented more than 400 services for JFCOM proving the feasibility and applicability.
- The general applicability of the approach has been documented and published in [17].

The fifth of our recommendations to the PDG is to make sure that all developments are complementary of each other and the central data model. C-BML will not be possible without the integration of these components. ***Application of a holistic view on all five components of C-BML*** is mandatory and the structure of the product development group tasks and organization must enable and reflect this need.

- The ontological spectrum is a logical continuum based on the principals of data engineering. The XML data elements of the C-BML Protocol, the data elements of the C-BML Representation, the vocabularies of the C-BML Grammar, and the concepts, relations, and terms of the C-BML Ontology are all interconnected and based on the C-BML Doctrine.
- Extensions of one view must be reflected in extensions of the other views. If, e.g., the grammar enables new sentences that reflect a valid view of the doctrine, this view must be captured in the representation and must be expressible by the protocol. The ontological view shall be used as the umbrella holding the elements together.

In this light, ongoing research in individual areas, such as ontology [14] or grammar [5], must be supportive of the integration of *all* C-BML components. Without this support, such work is certainly interesting and valuable, but may not be contributing to the overall success of C-BML.

5 Summary

With this paper, we wanted to show that all five parts of C-BML – protocol, grammar, representation, ontology, and doctrine – are equally important and interconnected. It is not possible to change one part without influencing the others. Therefore, a strong alignment of task forces is mandatory. Activities focusing on one part exclusively may result in solutions that can hardly be integrated with other activities. It is therefore essential to have a strong component that is agreed upon in the center of alignment activities.

Acknowledgments

Significant parts of the work supporting the results presented in this paper were funded by the Joint Advanced Training Technology Laboratory (JATTL) of the US Joint Forces Command. Additional support came from the US Army Test and Evaluation Command (ATEC).

The authors like to thank their colleagues from Naval Postgraduate School and George Mason University for the constructive comments during the writing process. They also want to thank their collaborating industry partners, particularly Gestalt, LLC, for their insights regarding implementation constraints.

6 References

[1] Sudnikovich, W.P., J.M. Pullen, M.S. Kleiner, and S.A. Carey (2004). *Extensible Battle Management Language as a Transformation Enabler.* SIMULATION, Vol. 80, pp. 669-680

[2] Simulation Interoperability Standards Organization (2004) SISO-REF-016-2006: *Coalition-Battle Management Language (C-BML)* Study Group Final Report, IEEE CS Press

[3] Tolk, A., S.Y. Diallo, and C.D. Turnitsa (2006). *The Command and Control Information Exchange Data Model (C2IEDM) as a Modeling & Simulation (M&S) Standard.* International Training Exhibition and Conference (ITEC) 2006, London, UK

[4] Turnitsa, C.D., and A. Tolk (2006). Battle Management Language: A triangle with Five Sides. Spring Simulation Interoperability Workshop, IEEE CS Press

[5] Schade, U., and M.R. Hieb (2006). *Formalizing Battle Management Language: A Grammar for Specifying Orders.* Spring Simulation Interoperability Workshop, IEEE CS Press

[6] Tolk, A., S.Y. Diallo, K. Dupigny, B. Sun, and C.D. Turnitsa (2005). *A Layered Web Services Architecture to Adapt Legacy Systems to the Command & Control Information Exchange Data Model (C2IEDM).* European Simulation Interoperability Workshop, ACM CS Press

[7] Turnitsa, C.D., S. Kovvuri, A. Tolk, L. DeMasi, V. Dobbs, and W. Sudnikovich (2004). *Lessons Learned from C2IEDM Mappings within XBML.* Fall Simulation Interoperability Workshop, IEEE CS Press

[8] Obrst, L. (2006). *The Ontological Spectrum & Semantic Models.* Open, International, Virtual Community of Practice on Ontologies, http://ontolog.cim3.net

[9] ISO/IEC 11179 (2004). *Information technology – Metadata registries (MDR),* Part I-VI

[10] NATO Research & Technology Organization (2002). *NATO Code of Best Practice for Command and Control Assessment, Revised Edition.* CCRP Press

[11] NATO Allied Administrative Publications No. 6 (AAP-6). *NATO Glossary of Terms and Definitions / Glossaire OTAN de termes et definitions.* http://www.nato.int/docu/stanag/aap006/aap6.htm

[12] Tolk, A., C.D. Turnitsa, and S.Y. Diallo (2006). *Ontological Implications of the Levels of Conceptual Interoperability Model*. World Multi-Conference on Systematics, Cybernetics and Informatics (WMSCI) 2006, Orlando, FL

[13] Turnitsa, C.D. (2006). *A Component Based Formal Ontology Model, A Method for Evaluating Such a Model, And the Results of an Application of that Method*. Master's Thesis, Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA.

[14] Blais, C.L., C.D. Turnitsa, and P. Gustavsson (2006). *A Strategy for Ontology Research for the Coalition Battle Management Language (C-BML) Product Development Group*. Fall Simulation Interoperability Workshop, IEEE CS Press

[15] Peng J., and H. Kincho (2004). *A Brief Review of Data Models for NEESgrid*, Technical Report NEESgrid-2004-01

[16] Gyssens M., J. Paredaens, and D. Van Gucht (1989). *A Grammar Based Approach Towards Unifying Hierarchical Data Models.* ACM 0-89791-317-5/89/CQO5/0263

[17] Tolk, A., S.D. Diallo, C.D. Turnitsa, and L.S. Winters (2006). *Composable M&S Web Services for Net-centric Applications.* Accepted for the Journal of Defense Modeling and Simulation (JDMS), forthcoming

[18] Alesso, H.P. and C.F. Smith (2005). *Developing Semantic Web Services*. A.K. Peters, Ltd.

[19] Tolk, A (2006) *Multi-resolution Challenges for Command and Control M&S Services*. Spring Simulation Interoperability Workshop, IEEE CS Press

[20] Pullen, J.M., W. Sudnikovich, A. Ritchie, P. De Champs, L. Khimeche, and M.R. Hieb (2006). *NATO Exploratory Team 016 Integration Lessons Learned*. European Simulation Interoperability Workshop, ACM CS Press

Authors' Biographies

ANDREAS TOLK is Associate Professor in the Faculty for Modeling, Simulation, and Visualization at the Engineering Management Department of the College of Engineering and Technology at Old Dominion University (ODU) of Norfolk, Virginia. He has over 16 years of international experience in the field of Applied Military Operations Research and Modeling and Simulation of and for Command and Control Systems. He is affiliated with the Virginia Modeling Analysis & Simulation Center (VMASC). His domain of expertise is the integration of M&S functionality into real world applications based on open standards. He received a Ph.D. and an M.S. in Computer Science from the University of the Federal Armed Forces in Munich, Germany.

SAIKOU Y. DIALLO is a Ph.D. candidate at the Virginia Modeling Analysis and Simulation Center (VMASC) of the Old Dominion University (ODU). He received his B.S. in Computer Engineering (2003) and his M.S. in Modeling & Simulation (2006) from ODU. His Ph.D. research under Tolk focuses on the domain of Model Based Data Engineering and web services for M&S applications.

CHARLES D. TURNITSA is a Ph.D. candidate at the Virginia Modeling Analysis and Simulation Center (VMASC) of the Old Dominion University (ODU). He received his B.S. in Computer Science (1991) from Christopher Newport University (Newport News, Virginia), and his M.S. in Modeling & Simulation (2006) from ODU. His Ph.D. research under Tolk focuses on the domain of dynamic and fractal ontology models for M&S interoperability.