2008

Community-of-Interest (COI) Model-Based Languages Enabling Composable Net-Centric Services

Andreas Tolk
Old Dominion University, atolk@odu.edu

Saikou Y. Diallo
Old Dominion University, sdiallo@odu.edu

Original Publication Citation
Tolk, A. and S.Y. Diallo (2008) Community-of-Interest (COI) Model-based Languages enabling Composable Net-centric Services. Proceedings of CCRTS 2008: The Command and Control Research and Technology Symposium, June 17-19, Bellevue, WA, USA.

2008 Command and Control
Research and Technology Symposium
"C2 for Complex Endeavors"

Community-of-Interest (COI) Model-based Languages enabling Composable Net-centric Services

Andreas Tolk, Ph.D.
Old Dominion University
241 Kaufman Hall / EMSE
Norfolk, VA 23529, USA
+1.757.683.4500
atolk@odu.edu

Saikou Y. Diallo
Virginia Modeling Analyses & Simulation Center
1030 University Blvd
Suffolk, VA 23435, USA
+1.757.638.6320
sdiallo@odu.edu

Cite as: Tolk, A. and S.Y. Diallo (2008). Community-of-Interest (COI) Model-based Languages enabling Composable Net-centric Services. *Proceedings of CCRTS 2008: the Command and Control Research and Technology Symposium*, June 17–19, Bellevue, WA, USA.

Community-of-Interest (COI) Model-based Languages enabling Composable Net-centric Services

Andreas Tolk, Ph.D.
Old Dominion University
241 Kaufman Hall
Norfolk, VA 23529, USA
+1.757.683.4500
atolk@odu.edu

Saikou Diallo
Virginia Modeling Analyses & Simulation Center
1030 University Blvd
Suffolk, VA 23435, USA
+1.757.638.6320
sdiallo@odu.edu

Abstract

Net-centric services shall be designed to collaborate with other services used within the supported Community of Interest (COI). This requires that such services not only be integratable on the technical level and interoperable on the implementation level, but also that they are composable in the sense that they are semantically and pragmatically consistent and able to exchange information in a consistent and unambiguous way.

In order to support Command-and-Control with Composable Net-centric Services, the human-machine interoperation must be supported as well as the machine-machine interoperation. This paper shows that techniques of computer linguistic can support the human-machine interface by structuring human-oriented representations into machine-oriented regular expressions that implement the unambiguous data exchange between machines. Distinguishing between these two domains is essential, as some requirements are mutually exclusive. In order to get the "best of both worlds," an aligned approach based on a COI model is needed. This COI model starts with the partners and their respective services and business processes, identifies the resulting infrastructure components, and derives the information exchange requirements. Model-based Data Engineering leads to the configuration of data exchange specifications between the services in form of an artificial language comprising regular expressions for the machine-machine communication. Computer linguistic methods are applied to accept and generate human-oriented representations, which potentially extend the information exchange specifications to capture new information not represented in the system requirements.

The paper presents the framework that was partially applied for homeland security applications and in support of the joint rapid scenario generation activities of US Joint Forces Command.

.

1 Introduction

It seems obvious that effective and efficient collaboration must be based on effective and efficient communication between participating organizations and individuals. However, current research often addresses only parts of the overall problem that spans more than one domain. To support efficient communication between organizations and systems, it is essential to understand the overall context of the information exchange and assess existing solutions in the affected domains with respect to their usability in other domains. This paper introduces a framework that allows one to specify a domain, analyze existing solutions, and show how they can play together in support of a more general solution. It is strongly motivated by the NATO Code of Best Practice for Command and Control Assessment (NATO, 2001), in particular by the recommendation not to rely on one model or tool to solve a complex problem, but to apply an orchestrated set of tools to evaluate the different relevant aspects and facets of a problem.

Communication in the Net-centric environment is without a doubt a complex problem. It seems to be logical to divide the overall problem into a set of sub-domains in order to make the problem more understandable and manageable. This results in a set of solutions that when combined solve the original problem. The idea behind the proposed framework is to base the orchestration of solutions on a common model encompassing the set of sub-domains and treat contributing solutions as views on this common model. The common model connects the human components – the members of the Net-centric organizations and users of the supporting systems – and the machine components – the supporting Command and Control Information Technology Infrastructure. Existing solutions can contribute efficiently to their application domain and – at the same time – they are aligned with the constraints of other contributions, as all solutions first contribute to the common model (making sure that the information of interest is part of the model by extending and enhancing it following engineering methods) and the resulting mapping allows to find out immediately which other solutions use the same information (as same pieces of information is mapped to the same elements of the common model) and which constraints need to be fulfilled (as the context of each information element is captured in form of metadata).

Although the framework addresses all aspects of the challenge of coming up with a common understanding of the operations and a supporting language usable for Command and Control by humans and machines, the paper will focus mainly on IT system related issues. The paper will present a general framework resulting from studies conducted in support of Community-of-Interest (COI) Model-based Languages enabling Composable Net-centric Services for NATO's Research & Technology Organization, the US Joint Forces Command, the US Army Test & Evaluation Command, and the US Army Topographic Engineering Center. The paper will discuss the general challenges that have to be addressed by the framework and introduce the framework and sample examples. In addition, the authors will present real world cases in which the framework was applied before summarizing conclusions and recommendations.

2 A Framework for COI Model-based Languages

The initial question that needs to be answered is how to use models to support effective and efficient communication within a COI. The approach recommended in this paper is motivated by the research conducted by the authors in support of US Joint Forces Command, the US Army, US Homeland Security, and NATO starts with a common understanding of data exchanged in form of a common reference model (CRM). The CRM is then used to build sentences and ultimately a web-based language to communicate these data in applicable contexts of the COI.

The idea of using COI approaches in support of interoperable IT solutions is not new. In (Renner, 2001), a COI is defined as *the collection of people that are concerned with the exchange of information in some subject area*. The community is made up of the users/operators that actually participate in the information exchange; the system builders that develop computer systems for these users; and the functional proponents that define requirements and acquire systems on behalf of the users. Renner stresses the importance of COI data panels and their task to support Common Data Representations (CDR) to be used within the COI for data exchange.

2.1 Defining the Problem Space

In order to set up the framework, we must first distinguish four categories of communication, as captured in the following figure: humans talking to humans, humans getting information into supporting IT systems, potentially heterogeneous IT systems exchanging information, and IT systems producing reports for humans.

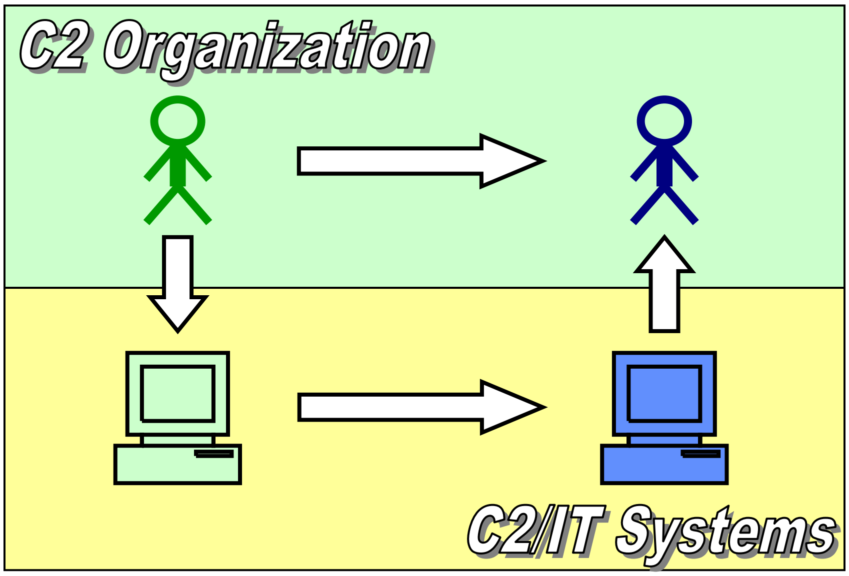

Figure 1: Categories of C2 Communications

- **Human-to-Human Communication:** This category includes organization-to-organization communication as well as human-to-human communication. It has been pointed out in several CCRP publications that organizational changes in support of net-centricity of organizations is often more important than IT system updates (see among others Alberts & Hayes, 2003). Many challenges of C2 are situated in this domain. Difficulties range from overcoming cultural and linguistic challenges to the quest for a common situational awareness. Although a very important aspect – if not the most important one, as the other aspects are merely supportive of this aspect – the focus of this paper is on the other three components. However, as the expertise of the authors of this paper lies in the IT section, we will not focus on these issues in more detail besides recognizing their critical importance.

- **Human-to-Machine Communications:** The Human-Machine-Interface is an important component for every IT application. However, when it comes to common communication concepts as captured in net-centric languages, new aspects needs to be considered. While the spoken language of humans – the natural languages – is ambiguous, often ill-structured, and normally requires a rich context for interpretation, the language of machines is governed by the rules of logic. This category translates the human-oriented view into the machine-oriented view. This task may be the most challenging in the framework. The discipline of computational linguistics deals with related challenges. Until recently, the Human-machine interface was the major element of concern for this component. The programming of user-friendly graphical interfaces with context sensitive pull-down menus to guide the human in constructing expressions valid for the machine was the state of the art. The result was a regular expression understandable and valid for the machine. Messages can be interpreted as special regular expressions. The first prototypes in support of the BML efforts – as described by Sudnikovich et al. (2004) – followed this trend by providing such an interface. More recent research applies computational linguistics to enable a better communication based on information exchange, such as summarized, among others, by Hecking (2004b). Computational linguistics started as the scientific study of language from a computational perspective. As such, it is an interdisciplinary approach combining linguistic theory – including phonology, syntax, semantics, pragmatics – with computer science – including artificial intelligence, theory of computation, programming methods. It applies statistical as well as rule-based models. The traditional focus is the modeling of natural language from the computational perspective. As recent research has shown that human language is much more complex than previously thought, computational linguists often work as members of interdisciplinary teams, including linguists, project specific application experts, and computer scientists. The resulting tools enable to process natural language and transform this information into regular expressions understandable by machines. Hecking (2004a) showed that this process can be facilitated significantly by formalizing the expressions used, which is in particular of interest in the military domain, as warfighters have been trained for generations to use clear and unambiguous expression when giving orders or reports. The work of Schade and Hieb (2007) in support of C-BML contributed significantly to this component as well. The premise is that the use of a structured language to convey orders including the commander's intent as well as reports must be captured in a formal grammar. The grammatical approaches predominantly taught in computer science, the so-called Chomsky grammars, are described as insufficient for C-BML purposes. Instead, the use of lexical functional grammars (LFG) is proposed. From an Internet survey conducted by the authors, the main applications of LFG are translation challenges. The LFG Bibliography maintained by the Universities of Essex, Stanford, and Koblenz[1] gives examples of how LFG is used to parse natural languages into formal structures for further computer manipulations.

- **Machine-to-Machine Communication:** A lot of research has been done regarding machine-to-machine interoperability. Federated databases, common data replication mechanisms, and common message formats are all examples of IT-system centric communication. The domain of expertise of the authors of this paper falls into this category. It is rooted in the theory of federated databases (Sheth & Larson, 1990). The objective of federated data bases is to merge different data sources, which are – and remain – distributed,

[1] http://www.essex.ac.uk/linguistics/LFG/www-lfg.stanford.edu/bibliography.html

heterogeneous, and autonomous but appear as one coherent federated data schema to the outside application. The definition of the CRM – and capturing the data in the recommended CDR – is closely related to the definition of a federated schema, as already pointed out by Krusche and Tolk (2000). In addition, machines need to follow the principles of logic. Without logic, proofs are not possible. Consequently, construction techniques must be applied to derive the different views from one common model. As shown by Hein (2002), these techniques can be divided mathematically into defined sets and grammars. Defined sets are used for data modeling, XML modeling, and other inductive techniques. Grammars introduce the production rules. It has been shown that for regular expressions the construction techniques can be mapped to each other, as both use finite automata as the underlying logical construct. That means that for every data model, an XML schema and/or a regular grammar can be constructed representing the same finite automata. It doesn't matter if one starts with the data model, the grammar, or the XML schema: if the transformations are consistent, always results in the same finite automata which determine scope, resolution, and structure of the set of valid regular expressions. This leads us to the result that machine-machine communications can and should be governed by regular expressions. The authors of this paper are convinced that in order to reach machine-to-machine interoperability the use of regular expressions is necessary, although it may not be sufficient. The applied solution cannot leave room for interpretation, as machines cannot support such interpretations, if they are not programmed to support this. If they are programmed, the programs will transform the data model into yet another data model, which either represents the same or a different finite automata. In both cases, however, the result is a set of regular expressions, so that the argument holds. In summary, the theory of federated databases is necessary – and potentially sufficient, but there is no proof for this assumption, only experience by application in various domains – to handle the challenges of machine-to-machine communications.

- **Machine-to-Human Communication:** It is important to understand that the translation from the machine-oriented view into the human-oriented view has some challenges as well, but that it is easier to solve than category two. The machine-human-component has the task to map from the logical world of regular expressions to the natural language of humans, which is much easier than the other way round. Many publications deal with the graphical representation of data to support an easier interpretation, which in the military domain leads to improvements in situational awareness. Another aspect of this component is the ability to generate written – or even spoken – reports in human-acceptable form from regular expressions. As mentioned earlier, regular expressions can be transformed into a relational data model, so the example here will utilize data models without limiting the applicability of the underlying ideas for other forms of regular expressions. In relational data models, data are captured in tables that can be related with each other. Each table represents a concept with a common understanding in the COI that is described by its attributes. Attributes can have predefined values. For each attribute, a sentence of the following form can be generated:

```
The <attribute> of the <concept> is <attribute-value>.
```

Examples are *the "Readiness Status" of "2nd Infantry Battalion" is "80%"* or *the "Reporting Data Reliability Code" of "the Current Report" is "Completely Reliable."* What each term exactly means is captured in the definitions of terms in the COI CRM, such as this is the case for the Joint Consultation, Command and Control Information Exchange Data Model (JC3IEDM) within the MIP community.

Similarly, associations between tables are rooted in operational associations between the concepts described by the table. A typical example is the order of battle that builds a hierarchy using the military organizations. The general form is as follows:

```
<subject-name> <association-type> <object-name>
```

Using the order of battle example, a possible sentence is that the *"2nd US/GE Corps" "has operational command of" the "2nd Infantry Battalion."*

Most current database administration systems allow the definition of user-specific reports based on these features. Some system architecture tools use similar ideas to generate system specification documents from the repository, which is filled by defining the artifacts of the supported system modeling and architecture method.

Using speech generators, this technique can be used to generate spoken reports as well. In particular for training purposes, this feature was already supported. The US Air Force uses them to simulate communications between pilots and air traffic controllers. The ability to automatically generate speech is invaluable in situations where it is required to represent a number of external agencies or operators within a simulated environment. Stoner et al. (2003) give additional examples. Kenny et al. (2007) even describe "virtual humans" that report to the training audience using related ideas.

However, as our focus in this paper is a COI-model driven language we will not deal with them in detail. The interested reader is pointed to Tufte's work (1997, 2001) as a good starting point.

These four categories establish the problem space that a framework enabling the definition of COI model-based languages must solve.

2.2 Implications for COI Model-based Languages

In order to define a COI model-based language all four components captured in the problem definition must be taken into account. Even when focusing on a pure system-oriented view (machine-to-machine communication), the other three aspects need to be reflected; otherwise the result is an inconsistent and incomplete representation of the processes within the COI. However, it is also important to understand what challenges have to be solved in which component and how the different aspects of solutions work together, otherwise, suboptimal solutions are developed. The authors offer the following list of challenges and where they should be solved. This list is neither complete nor exclusive. It is not only intended to present already conducted research but also to start additional research:

- ***System-to-System Interoperability***: If the objective is the interoperability of two IT systems supporting a common operation, the assumption is that the IT systems as they are support the sub-tasks of an operation in their application domain. If the IT systems are not sufficient to support the sub-task, this is another challenge to be solved, but interoperability needs to identify which information elements (which are data elements defined by scope, structure, and resolution in a given context) can be mapped to each other. This challenge falls into the machine-to-machine component.
- ***Sufficiency of an IT Solution for an Operation***: The question if an IT system is sufficient to support a task in an operation cannot be answered in the machine-to-machine component. To answer this question requires that the human/ organizational component specifies the operational needs first. These needs, normally derived from doctrine and task lists by applying methods, such as the Military Mission and Means Framework (Sheehan et al., 2003), belong in the human-to-machine component of the framework. The result of such work is to make sure that all information elements needed are represented in the IT systems in the appropriate form.
- ***Ambiguities in Expressions***: The task to reduce ambiguities in expressions is more important than ever if the recipient of such an expression is a machine, as machines don't make interpretations, they need unambiguous structures to read from. As Schade and Hieb (2007) formulate it: "*Without ambiguity means that there is an explicit structure that the Command Intent can be put into and then parsed out of. It also means that only one clear and definite outcome results from the parsing.*" The task of reducing ambiguity therefore falls into the human-to-machine component of the framework. Once the explicit structure is defined and populated, Model-based Data Engineering (MBDE) can be applied to mediate this structure into other representation forms.[2]
- ***Defining a Common Language for System Interoperability***: This task is closely related to the question of the sufficiency of an IT solution for an operation. In order to define a language, the required information elements for possible operations need to be defined. What is needed is derived from the human/ organizational component by operational requirement, but the result is an information exchange model that captures the mandatory information elements in machine adequate form, i.e., using construction techniques representing the necessary finite automata. An example how to apply grammars is given by Schade and Hieb (2007). However, in order to support an interoperable solution, the system-to-system interoperability aspect needs to be solved for the targeted systems in addition.

In summary, a common language must be defined by analyzing the operational needs (human/organizational component) resulting in the definition of information elements

[2] Model-based Data Engineering is based on the Data Engineering as introduced in the NATO Code of Best Practice for C2 Assessment (NATO, 2001). It is characterized by the use of a common reference model to facilitate the data management and alignment process, which is a representation of the COI data model.

(data defined by scope, resolution, and structure in a context) in construction techniques representing a finite automata. Systems that need to support an operation must be sufficient, i.e., their information elements must be mapable to the subset of the common language representing their tasks, which are also represented as information elements. MBDE methods can be used as soon as information elements are defined. To this end, the authors designed a top-down/bottom-up approach that will be presented in the following section.

3 The Top-Down/Bottom-Up Approach to derive a COI Model-driven Solution

In order to derive a common understanding between diverse organizations supported by heterogeneous IT systems, the following multi-layer, model-based approach was proposed, published, and discussed during several conferences (Tolk et al., 2007). It applies the ideas of the framework discussed in section 2.

3.1 Overview of the process

The approach combines top-down analysis with bottom-up systems alignment. The top-down approach concerns itself with deriving accurate conceptual models of the operational goals of the organizations, and then revealing the important data concepts that constitute those models. The bottom-up approach starts with the information models used in the supporting IT systems and results in a CRM describing the information exchange that needs to be supported between the systems regarding the data concepts identified by the top-down approach. The following figure exemplifies these ideas. Figure 2 shows an overview of the steps of this process.

There are three initial top-down steps followed by three bottom-up steps. The result is a COI CRM plus web services implementing a COI model-based web language allowing the systems to exchange information in support of the common conceptual model of the underlying processes.

1. In the first top-down step, the participating organizations agree on the common process to be conducted. This is a high-level business model in which the leaders of the organizations must agree to cooperate and to accomplish common goals.
2. Based on these decisions, a common conceptual model needs to be set up. The recommendation of the research team is to use standardized methods, such as the Unified Modeling Language. Examples how this can be done for business processes are given by Erikson and Penker (2000).
3. The result of step 2 is a model of the business processes to be conducted and the organizations that have to conduct these processes. In step 3, the interfaces between the processes are evaluated in regard to which information needs to be exchanged to orchestrate these processes: Who says what to whom in which contexts based on which results?
4. The first step of the bottom-up process is to document which IT systems are needed in support of the processes. This includes the identification of programs, components, and services plus supported interfaces.

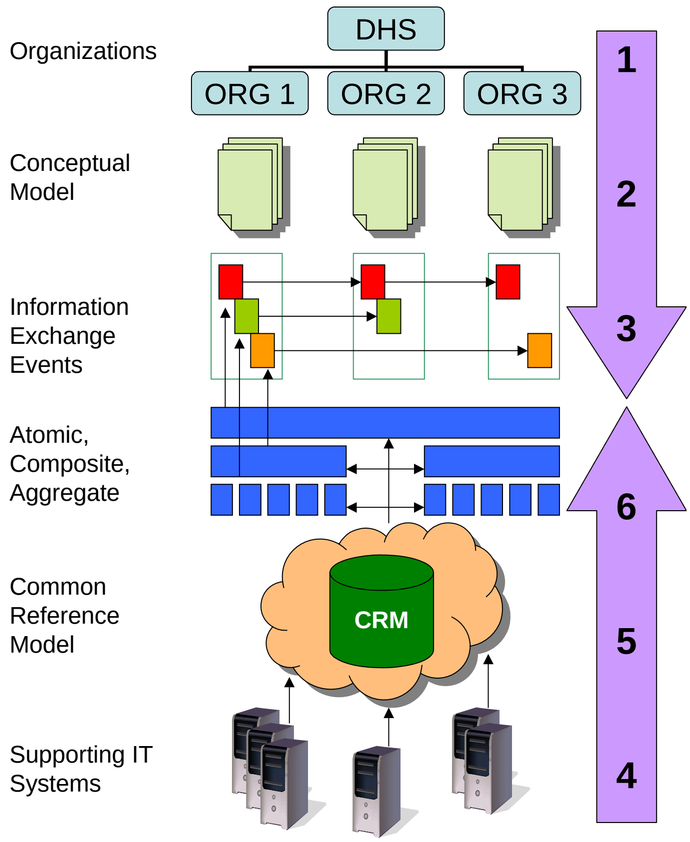

Figure 2: Top-down/Bottom-Up approach to derive COI-Services

5. The next step is to use the interfaces to identify what data in which format (syntax), with which meaning (semantics), in which contexts (pragmatics) must be exchange. The common conceptual model is the guideline directing the definition of information exchange elements provided by the systems. The result is a CRM comprising all information elements provided by the IT systems in support of the conceptual model of the common operation.
6. The last step is the definition of a COI model-based language. The language is based on all possible valid sentences that can be stored into or produced by the CRM. To this end, a logically cascading family of web services is defined starting with atomic web services (exchanging minimal information), composed web services (exchanging more complex information, such as defined by "views" on relational databases), and finally application specific modifications in form of aggregate services (data is transformed, filtered, or aggregated to fulfill the special needs of one or more applications). Tolk et al. (2006) explain this process in more detail.

Each participating program, component, or service knows at the end of this process exactly how to structure the information to be exchanged and which web services to feed with this information.

3.2 A Simple Example for the Top-down/Bottom-up Approach

In order to make the top-down/bottom-up approach easier to understand, this section gives a small example to show the feasibility and applicability of the approach. In the example, two organizations support a common operation. Both organizations are car rental facilities. They give cars to customers, get them back, and check for damage. Now they want to be able to support each other. They agree that they will accept returns of cars from the other organization, check them for damage, and transfer them back to the other organization as soon as possible. Following the top-down part of the model, the following steps are conducted:

- In step one, both car rental companies agree to accept returns of cars from the other company, check them for damage, and transfer them back to the other organization as soon as possible.
- In step two, a common model comprising both organizations – and the customer – is set up and agreed to as a common business model. The sequence diagram in figure 3 shows this common operation.

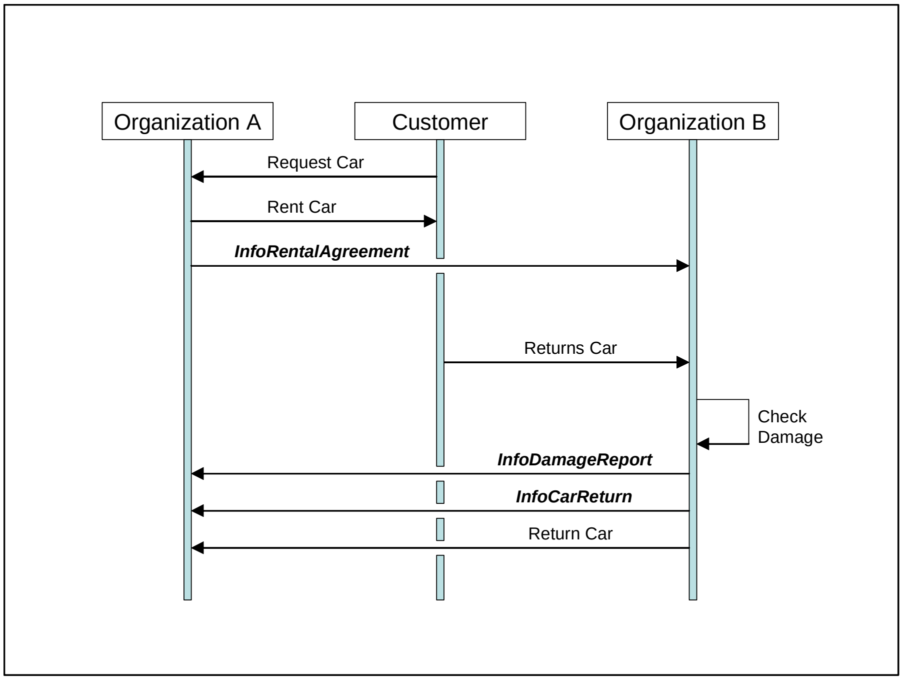

Figure 3: Model of the Common Operation

- Based on this business model, the organizations agree that they need to exchange the following information elements:
  - When the car is rented and the intent to give it back at the other organization is obvious, the rental agreement is transferred to the other system, so that the other organization is aware of the status of the car when it is given to the customer and when and where to expect it back.
  - When the car is returned, the result of the damage check is captured and transmitted to the other agency.

  - When the car is returned, the other organization is notified when and where the car will be returned.

  These are the operational information needs identified in step three. They are written in italic in figure 3.

- Step four, the first step in the bottom-up part, identifies only the two IT systems of the organizations to be relevant.
- In the next step, the data models are evaluated to find out, how the information elements available in the systems can be used to establish a CRM in support of the operational information needs (rental agreement, damage report, and return information). While the first system may, e.g., use the driver's license as the identification for a customer, the second system may use an individual customer ID. This must be known when the return information is transferred. Another possibility is that the first system documents the damage status with each car while the second system has a table to deal with damage descriptions that can be related to cars. As pointed out in several publications describing MBDE (such as Tolk et al., 2006), the CRM should reflect the highest resolution and should be structured to support all participation data structures.[3]
- In the last step, the mapping of the original information elements used in the systems to the information elements in the CRM is done. In practice, this leads to a web-service call using the information elements in support of the three supported activities: when a car is rented, when a car has been checked for damage, and when a car is ready to be returned.

While this example is trivial, the same ideas have been used in support of the US Joint Forces Command to enable rapid scenario generation based on heterogeneous authoritative data sources resulting in several hundred web services utilizing the JC3IEDM as the CRM. An overview of the activities is given by Perme et al. (2007).

4 Application Use Cases

The work on COI model-based languages is nothing completely new. It is possible to utilize and adapt the results of research conducted in several related domains, such as federated databases, common message formats, and other traditional interoperability solutions. New approaches currently evaluated regarding ontological approaches and semantic web-services as well as new applications of computational linguistics support the ideas as well. The next two sections give examples of two application domains. These examples are neither exclusive nor complete. The first example looks at a COI with a potentially usable COI model. The second example looks at a COI that has been build without such a common understanding of domain data.

[3] If system one represents a fact using one attribute A and system two represents the same fact in higher resolution using attributes A1 and A2, the CRM should use A1 and A2 to model this fact.

Furthermore, if system one groups B1 and B2 into one information element while regarding B3 to be an individual element while system two groups B1 and B3 into one information element while B2 is regarded to be an element on its own, the CRM needs to be structured to support both views, which means all three attributes should be modeled as individual elements that can be composed into B1/B2 for system one and B1/B3 for system two.

4.1 The NATO Use Case

Using Renner's definitions for COI and CDR referenced in section 2, NATO is following a comparable approach with the Multilateral Interoperability Programme (MIP) for several years.[4] Related work was presented in CCRTS papers during several symposia. The central idea is the use of a common information exchange data model and data replication mechanisms ensuring that the data exchange is following the underlying business rules and is embedded into the workflow ideas underlying common joint and combined operations of NATO. Similar to the data panels described by Renner, a data administration group ensures the consistency of data regarding its representation within the information exchange data model, but also regarding its meaning.

It should be pointed out that the main objective of MIP is to establish a common information exchange model, not a common data model to be shared and implemented for all systems. The systems must only be able to "speak" as recommended by MIP externally. How the systems store and model the data internally is of national concern. While several nations decided to utilize the information model as their C2 data model as well, this is not a mandatory decision. In particular the new NATO nations are furthermore very interested in commercially supported off-the-shelf solutions executable on affordable standard PC platforms instead of developing often expensive individual solutions.

Finally, new initiatives – such as the NATO Information Systems Technology (IST) panel activity IST-075 "Semantic Interoperability" – are evaluating the use of ontological means to extend the use of the MIP model beyond the data replication between IT systems.

4.2 Homeland Security Use Case

While in military operations the community can utilize a common understanding that has been developed in form of joint and combined doctrine at least to a certain extent (soldiers all over the world share a common understanding of at least the basic activities, such as attack, defend, march, shoot, communicate, etc.), supported by command and control means that are at least understandable to the other partners (although they are not always shared). Other COI groups do not start from such common ground.

When the United States as a response to the 9-11 attacks set up the Department of Homeland Security, its Strategic Plan (USDHS, 2004) identified six guiding principles in support of the philosophy that informs and shapes decision-making and provides normative criteria that govern the actions of policy makers and employees in performing their work in this domain. The guidelines are: protecting civil rights and civil liberties, integrating the actions of the 22 previously disparate agencies, building coalitions and partnerships, developing human capital, and innovation.

Of these six guiding principles, "integrating the actions" and "building coalitions and partnerships" present challenging requirements for migration and interoperation because each of the 22 previously disparate agencies were in existence before Homeland Security established new requirements. Each agency comes with its own policies and processes,

[4] The documentation of MIP, its organizational structure, and the used data model can be downloaded in the current version form the website http://www.mip-site.org

proprietary IT systems supporting these processes, and data models structuring the necessary information within applications.

In order to support the mission of Homeland Security, these stove-piped solutions must be migrated to a common technical standard and integrated to support the new processes, many of which were not foreseen when the existing processes were defined and the supporting IT systems procured. Buying new systems for all participating agencies is out of question, as it is too expensive. Furthermore, employees are already well trained and familiar with their systems.

In support of the endeavor to establish a CRM for Homeland Security applications in order to allow them to federate their systems using a web-based language, the authors developed a top-down/bottom-up approach to identify the business processes engaged, the derived information exchange requirements, the supporting IT systems, the interfaces to be supported, and the IT system specific data representing the information exchange requirements. They recommended Model-based Data Engineering principles as described in by Tolk and Diallo (2006) to generate a CRM. The principles are described in Tolk et al. (2007).

4.3 From Common Data to a Common Language: the C-BML Efforts

These ideas were presented at several conferences in support of the Coalition Battle Management Language (C-BML) efforts. While the work described so far focuses on the definition of data elements, the objective of C-BML is to define a web-based language. C-BML uses standardized data elements, but addresses their use to exchange information in form of valid expressions, that can be interpreted as valid sentences of the language.

The final report of the Study Group (SISO, 2006) recommends a phased approach to define such a language, that should be the an "*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.*" The NATO Code of Best Practice for C2 Assessment defines C2 as the "*organization, process, procedures, and systems necessary to allow timely political and military decision making and to enable military commanders to direct and control military forces. The C2 systems include headquarters, facilities, communications, information systems, and sensors and warning installations.*" It is obvious that a C-BML must support all these components as well, at least the human components and the machine components.

The scope of C-BML was identified as C2 related information, in particular tasks and reports that are exchanged between live forces using their C2 systems, simulated forces, and robotics. Currently, two approaches are evaluated in detail by different groups. The approach favored by the researchers contributing to this article focuses on interoperability between systems: C2 systems, Simulation Systems, and potentially Robotics (although the C-BML group decided not to focus on robotics in the next phases). This approach is rooted in Model-based Data Engineering, such as described by Tolk et al. (2006) and most recently directly applied in Tolk and Diallo (2008). Several recommendations and publications of results of underlying research were awarded.

The alternative approach was very successfully discussed in several conferences as well and also received awards. An overview of the idea to use lexical-functional grammars as the basis for C-BML was recently featured in the CCRP sponsored C2

Journal (Schade and Hieb, 2007). The core of this alternative is to define a C2 doctrine driven grammar to define a C-BML to which participants can describe.

It seems to be at least interesting that both approaches, although they are perceived to be competitive and exclusive, find sponsors and supporters. In the understanding of the authors of this paper the reason is that *both approaches are not really competitive, but mutual supporting*. Both alternatives target to define a web-based language. Both target at the C2 domain. Both alternatives utilize reference models. They are two views on one common model. It is therefore essential that such approaches are aligned: a valid sentence must be part of the data model, and views on the data model must result in valid sentences. A first view on the underlying mathematics has been addressed in (Tolk and Diallo, 2008).

Each C-BML view is supporting the solution of a different problem domain, as hopefully has been shown by the discussions in the context of this paper: The current grammar-driven work focuses on human-machine challenges, such as enabling common planning. The MBDE approach focuses on machine-to-machine interoperability as needed for decision support, operational use of simulation services, etc. They both are needed and need to be applied in an orchestrated manner.

5 Summary

The methods described in this paper allow the definition of a COI model-based language enabling net-centric composable services. The COI model is defined as a model representing information elements, which are data elements defined by their scope, structure, and resolution in a given context, using construction techniques representing regular expressions that can be evaluated by machines. The model needs to be derived from operational information exchange requirements that are based on organizational models of how to conduct the given set of operations. If a candidate model exists, such as the JC3IEDM in the NATO domain, MBDE methods can be applied to extend and enhance the model as needed. If no model exists, the same methods can be used to set up such a model based on the operational information exchange requirements. The proposed framework can be used to orchestrate these different methods, such as different OR methods are orchestrated in C2 analysis by the NATO COBP.

Computational linguistics plays an important part when deriving information elements from operational requirements, as the requirements are described in natural language. In particular recent research can be applied to compute ill-structured language input into regular expressions usable by computers.

In order to support a COI model-based language enabling net-centric composable services, both elements are needed in the definition phase. In agile systems, i.e. systems that use new requirements to define new information elements during operation (or service-based systems that have to integrate new services with new information exchange requirements), the definition phase never ends. In such systems, both elements – Computer Linguistics and MBDE – are needed as operational components in support of their agility. However, it is important to apply the method to solve problems they were designed for to avoid suboptimal or even wrong solutions.

Acknowledgements

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) and the US Army Corps of Engineers Topographic Engineering Center (TEC). We also would like to thank our friends and colleagues for their fruitful discussions and contributions in the course of writing this paper.

About the Authors

ANDREAS TOLK is Associate Professor in Engineering Management at Old Dominion University (ODU) of Norfolk, Virginia. He is affiliated with the Virginia Modeling Analysis & Simulation Center (VMASC). 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.

How many legs does a dog have if you call its tail a leg? Four!
Calling a tail a leg doesn't make it one.

Abraham Lincoln

6 References

(Alberts & Hayes, 2003) David S. Alberts, Richard .E. Hayes: *Power to the Edge*. Washington: CCRP, 2003

(Erikson & Penker, 2000) Hans-Erik Erikson, Magnus Penker: *Business Modeling with UML – Business Patterns at Work*. OMG Press, John Wiley & Sons, Inc., New York 2000

(Hecking, 2004a) M. Hecking. *Improve Interoperability by Formalizing the Natural Language Parts of Military Messages*. Proceedings of the Information Systems Technology Panel Symposium (IST-042/RSY-014) 'Coalition C4ISR Architectures and Information Exchange Capabilities', The Hague, The Netherlands, 2004

(Hecking, 2004b) M. Hecking. *How to Represent the Content of Free-form Battlefield Reports.* Proceedings of the 2004 Command and Control Research and Technology Symposium, San Diego, California, 2004

(Hein, 2002) J.L. Hein: *Discrete Structures, Logic, and Computability*. 2nd Edition. Jones & Bartlett, 2002

(Kenny et al., 2007) Patrick Kenny, Arno Hartholt, Jonathan Gratch, William Swartout, David Traum, Stacy Marsella, Diane Piepol: *Building Interactive Virtual Humans for Training Environments*. Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), Orlando, 2007

(Krusche & Tolk, 2000) Stefan Krusche, Andreas Tolk. *Information Processing as a Key Factor for Modern Federations of Combat Information Systems*. Proceedings of the NATO IST Symposium "New Information Processing Techniques for Military Systems", RTO-MP-049, Istanbul, October 2000

(NATO, 2001) *NATO Code of Best Practice for Command and Control Assessment*, CCRP Press, 2001

(Perme et al., 2007) Perme, D., S.Y. Diallo, C. Pandolfo, B. Tedesco, H. Tran, A. Tolk: *Joint Event Data Initialization Services (JEDIS) – Implementing a Service Oriented Architecture for Initialization.* Spring Simulation Interoperability Workshop, IEEE CS Press, 2003

(Renner, 2001) Scott A. Renner: *A Community of Interest Approach to Data Interoperability*. Proceedings Federal Database Colloquium '01, San Diego, August 2001

(Schade & Hieb, 2007) Ulrich Schade, Michael R. Hieb: *Improving Planning and Replanning: Using a Formal Grammar to Automate Processing of Command and Control Information for Decision Support*. Journal for Command and Control, Volume 1 Number 2, pp. 69-90, 2007

(Sheehan et al., 2003) Jack H. Sheehan, Paul H. Deitz, Britt E. Bray, Bruce A. Harris, Alexander B. H. Wong: *The Military Missions and Means Framework.* Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), Orlando, 2003

(Sheth & Larson , 1990) A.P. Sheth, J.A. Larson. *Federated database systems for managing distributed, heterogeneous and autonomous databases*. ACM Computing Surveys 22: 183-236, 1990

(SISO, 2006) Study Group: *Final Report on Coalition Battle Management Language (C-BML).* SISO Dokument SISO-REF-016-2006-V1.0, SISO 2006

(Stoner et al., 2003) Donald J. Stoner, Louis Ford, Mark Ricci, Marcia Williams: *Simulating Military Radio Communications Using Chat-Bot Technology*. Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), Orlando, 2003

(Sudnikovich et al., 2004) W.P. Sudnikovich, J.M. Pullen, M.S. Kleiner, S.A. Carey: *Extensible Battle Management Language as a Transformation Enabler.* SIMULATION 80: 669-680, 2004

(Tolk et al., 2006) Andreas Tolk, Saikou Y. Diallo, Charles D. Turnitsa, Leslie S. Winters: *Composable M&S Web Services for Net-centric Applications.* Journal for Defense Modeling & Simulation, Volume 3 Number 1, pp. 27-44, 2006

(Tolk et al., 2007) Andreas Tolk, Charles Turnitsa, Saikou Diallo: *Model-Based Alignment and Orchestration of Heterogeneous Homeland Security Applications Enabling Composition of System of Systems.* Winter Simulation Conference WSC'07, Washington, DC, December 2007

(Tolk and Diallo, 2008) Andreas Tolk, Saikou Y. Diallo: *Strong and Weak Composites for Adaptable Web Languages*. Spring Simulation Interoperability Workshop, IEEE CS Press, 2008

(Tufte, 1997) Edward R. Tufte. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press, 1997

(Tufte, 2001) Tufte, Edward R. *The Visual Display of Quantitative Information*, 2nd Edition, Cheshire, CT: Graphics Press, 2001

(USDHS, 2004) U.S. Department of Homeland Security: *Securing Our Homeland, U.S. Department of Homeland Security Strategic Plan*. Washington, DC, 2004