More Information, Less XML Please!

(Promoting a text-based approach to message design and documentation)

Cheryl L. Connors, MITRE
The MITRE Corporation
Huntsville, AL
cconnors@mitre.org

Introduction

The type of information needed for operational use by the military is, overall, fairly persistent. The need to efficiently and effectively exchange this information with others is also a consistent need. The information's accuracy and exchange mechanisms have improved from the time of Julius Caesar, but a general has always needed to know and be able to inform his army where the enemy is located. This paper examines the United States Department of Defense's (DoD) primary methods for creating the messages used for its modern military operations, explores issues with the implemented approach, and provides recommendations for improvement.

Department of Defense Messaging Standards

The United States military has used a number of different encoding (i.e. representations) for information exchanges (aka "messaging") that leveraged the prevailing technology of the time, from Morse code to ASCII to UTF-8. The military has also developed unique formats, including the binary based formats MIL-STD-6016 (Link 16) and the text-based MIL-STD-6040 (United States Message Text Format (USMTF) to define messages for its operational missions.

A paradigm shift early in the 2000's moved away from more traditional messaging methodologies, emphasized the use of Extended Mark-Up Language[1] (XML) as a message format, and encouraged independent groups of Subject Matter Experts (SMEs) to form Communities of Interest (COIs) to create vocabularies and messages supporting a specific operational area. COIs have been created to support air operations, targeting, ballistic missile defense, etc. While applauding the intent of the COI concept, several issues with the resulting COI products have arisen due to the use of XML as a documentation tool[2], the general lack of XML expertise by the operational SMEs, and the non-inclusion of messaging experts as core members of a COI. This paper proposed a way-ahead that helps resolve these issues by providing guidance to COI participants on how to build a message format, what should be included in it, and how to properly document it. The goal is to provide SMEs with enough understanding so that the data can be used even when its representation changes due to another technology. The next section discusses the issues with using XML as a documentation format. For clarity, this paper strongly

[1] Correct spelling of the words defining the "XML" acronym. Please see http://www.w3.org/standards/xml/core
[2] While XML is touted as self-describing, and an XML instance may be (but isn't necessarily), understanding an XML schema requires training and experience.

advocated the continued use of XML as an information exchange mechanism, but suggests that the documentation of such exchanges should be less formidable.

Message Documentation in XML versus Representation in XML

It's necessary to first make clarify the difference between "documentation" and "representation" as used in this paper because there is a significant distinction between documenting a message standard in XML, and representing and using XML to define information exchanges. "Documentation" in this context refers to a "plain language" compendium of information that describes the overall message format (i.e. standard "template"), the rules for creating valid messages and message instances, a description of each message (e.g. "Air Tasking Order", "Weapon Inventory"), and provides detailed information about each message (e.g. its specific structure and rules) and the data (e.g. data element dictionary) to support each message. "Representation" in this context refers to the encoding (e.g. binary, XML) used to represent each message and each instance. One of the XML's primary benefits is the W3C (XML) Schema language, which can be used to create a template for each message. When used with the right processors, instead of having to write unique code to verify that the content of a message is valid, these schema "message templates" can be used to validate the content of each message instance.

This usage of XML to "document" the COI vocabularies and message formats has led to many COIs to forgo development of a more traditional messaging standard leaves SMEs at the mercy of others to ensure that their requirements are indeed being met. It takes significant time to become XML knowledgeable, and since most operational SMEs do not use XML to perform their tasks, their exposure is limited. The general lack of access to XML specific tools by COI SMEs, such as XMLSpy, makes it difficult for independent evaluation of the validity of the COI XML products. While the graphical views of schema provided by tools like XMLSpy are useful, these images are not intuitive, and abide by tool specific symbols (e.g. in XMLSpy a "dashed line means optional") that may vary from tool to tool. Tests were conducted on several Microsoft Office products including Word and Excel. These tools extract and replicate the XML (e.g.), in their own format. Due to the complexity of most of the schemas, the results were not useful in improving the understanding of the overall message format or its content. Additional issues with the lack of additional documentation is the inability to document rule sets that impact operational behaviors, for example, how the information should be disseminated or reproduced by the receiving system. COIs are currently referring to their message components or elements as "XML Vocabularies", which appears to be a misunderstanding of the relationship between the data and its representation. Historically, while the type of data needed (e.g. "location") remains the same, its representation will change. Like Morse code and ASCII, XML will someday be set aside in favor of a new encoding. This "XML-based" documentation approach will require additional efforts to reverse engineer and extract the data from the representation. An ancillary concern should be noted, while XML is advertised as being self-describing, this is not true for XML representations of military messaging standards. The restrictions on the permissible characters in XML tag names, often causes a loss of information when attempting to create XML tags from military message components, which means the original "English" names cannot be re-created from the XML tags.

The advocated approach is to use a system engineering approach that as a minimum, results in a deliberate message design based on best practices, a catalog of message formats represented in a table views, a data element dictionary, and a detailed approach defining how an XML representation should be developed from a message in the catalog. A set of XML schema, one for each message should be included as an appendix as a template for use by developers. The main text would provide operational and message formatting rules and explain overall message construction, as well as providing design operational rules for each message.

Figure 1 shows the conceptual relationships between the document, its storage, and its XML representation. The message standard could be stored in a flat file or a database, but user-friendly views, based in plain language should be producible when required. The XML schema used to represent the messages would logically be included as part of the standard. The next section discusses the core set of information that should be considered for inclusion in defining a message format.

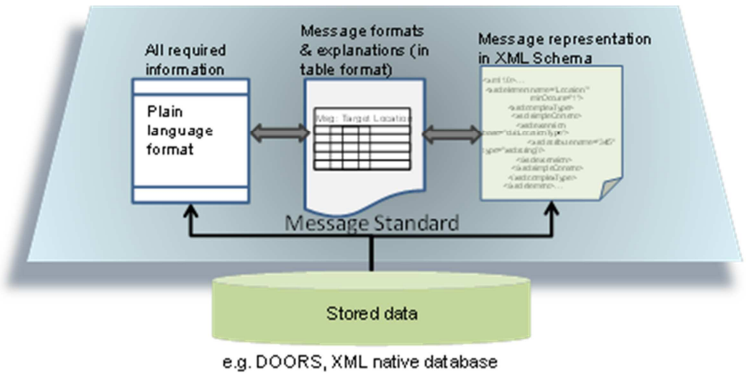

Figure 1. Conceptual Relationships between Message Standard Components

Designing a Military Message

The discussion in the previous sections focused on several issues with the COI approach of designing information exchanges, as well as discussing issues with using XML schema as the documentation of a message catalog. The main complaint is the use of XML to document as well as represent the exchanges, forcing military operators to acquire some understanding of XML. A more traditional approach is recommended instead of relying on XML for documentation and representation. The proposed approach is one that maintains a clear distinction between the data and its representation, and is a format that is more readily understood by non-messaging experts. Though this approach too needs some training, the learning curve is much smaller than needed for learning XML. As indicated earlier in the paper, XML is heartily advocated for use, and this approach supports the use of XML, but is flexible and robust enough to use with any other representation (e.g. a binary encoding). Describing an entire message standard, and the configuration management data it requires, is out of scope for the purpose of this paper.

Messaging Building Blocks

The core building blocks of messaging are the concepts used to describe it. Three DoD military message standards were reviewed and used as a basis for defining the proposed approach that any COI could use to design a new message format. The standards were MIL-STD-6016 (Link 16), MIL-STD-6040 (USMTF) and MIL-STD-6017 (VMF). These standards were selected as they each provide a unique message format that is in use today within and across different DoD organizations, and each standard has been successfully used for decades. The Link 16 and VMF standard use a binary representation and USTMF uses a text-based encoding for information exchanges. Note that MIL-STD-6040 and MIL-STD-6017 VMF both have XML representations of their message exchanges but still maintain an "English" version of the message format, logic and components. This is the basic distinction between the efforts of the more formal, and historical approach which this paper proposes should be adopted by COIs. The review shows that there is no single, definitive approach to creating the "best" message format; all three message standards are in use today. While this paper promotes a particular way-ahead, flexibility should be included in the design goal because there often is no single best answer and the format will always be influenced by context and situational factors. Even when discussing describing data at the lowest level, the answer to almost any question is "it depends". Therefore, the approach used in the following discussion is based on what has been observed to be useful, and it provided to inform potential developers of factors to be considered when they need to create a message standard for their community.

The analysis of the three military message formats determined that while each used a unique format, there was overall similarity in the type of information being documented. An abstracted view of the core concepts for the contents of a message and its structure are provided in the following graphic, Figure 2.

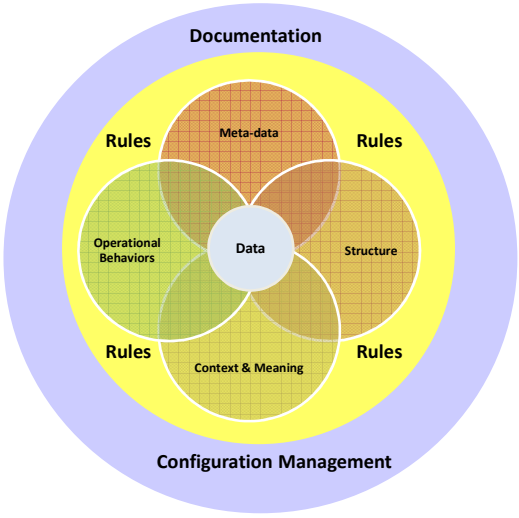

Figure 2. Analysis of Messaging Standard Formatting Content

The outer circle deals with data "outside" of the specific topic area of message formats, but it is important to note that these needs exist; the need to create the overall message standard itself

("documentation"), and to create a formal configuration management process to support changes to the document and the messages it describes. The second circle was created to emphasize the importance of rules for every aspect, from the design view (e.g. creating a prescribed naming convention for creating message components) to the implementation view (e.g., defining "if-then-else" rules for how and when a message component may be used in a specific context). The core is the "data" itself, referring to the information (and rules!) used to effectively exchange meaningful information between two military systems. To have a complete design, structure is needed. To have the data fully defined, metadata is needed. To make the data useful, context and meaning are needed. To help ensure the desired result occurs (i.e. satisfy the purpose or reason that the data was sent), any associated behaviors (either by a machine or a human) need to be described. From these general concepts, it was noted that each standard defined two sets of related information; the set of data points describing the message itself, and a set of data points that help define the structure and the message components. At the message level (e.g. "Weather Report Message"); the metadata generally included the following:

- Message Number
- Message Identifier (usually an acronym)
- Message Name
- Operational Intent and Purpose
- Classification (e.g., "unclassified")
- Metadata used for configuration management

For the message components themselves, the analysis determined that the data points provided in the following list should be considered for inclusion in a formal message design. The sequence of the information is deliberate and supports readability. Also note, that as indicated previously, the set of metadata needed for configuration management (e.g. unique identifiers for each discrete message component) are not included in this list. The message component data points include:

- Classification (e.g., "unclassified")
- Sequence
- Presence Indicator
- Group Name (if any)
- Component Name
- Relationship(s)
- Occurrence
- Repeatability
- Rule(s) (if any)
- Description

The proposed format to document the concepts listed about is a table view, which is easier to humans to use and manipulate. The following paragraphs review the above concepts before providing an example of how they would appear in a table view.

"Classification". While conceptually simple, classification information, which documents the security classification of information (e.g. "unclassified"), it's important to note there are two types of "classification" that need to be documented. A message format needs to have its classification identified, and the classification of the each message component each need to be identified. Also required is a separate classification for an *instance* of a message and each element in that instance needs its own classification. In an XML schema, the classification for a message instance and its component is generally assigned as an XML attribute. Attributes are discussed in more detail later on in this paper.

"Sequence". The first concept to be discussed is "Sequence". A sequence value (e.g. "1.0", "1.1") usually act as an index, which may or may not be needed in a message instance itself, particularly with the use of XML "tagging" where each discrete data received is identified, and so may be processed by "reading" its name. While the sequence information may not be needed to be preserved in the XML schema or forwarded in a message instance, from a human readability format, the sequence number is useful as an organization tool and should be included in a message format.

"Presence Indicator". The Presence Indicator in an artifact of MIL-STD-6017, and indicates where a particular message component should be included in a particular message instance. A message format in this military specification may actually consist of any number of "sub-messages", and each sub-message is defined by which presence indicators are turned off or on. The use of the presence indicator can be useful to save bandwidth and processing time. However, the use of many sub-messages in a single message format makes it much harder for a human to understand, and is included for completeness. Designers may want to exclude the use of the presence indicator in favor of more simplified message designs.

"Group Name" and "Component Name". The message standards use abbreviations or acronyms to represent the "Group Name" and "Component Name" and the full name was described "outside" of the message format itself.

"Relationship". Relationship information in the proposed table based format is parent to child only. No peer-to-peer relationships can be identified using a table format. The relationships that can be defined by format are "(parent) message to (child) components, and groups to group components. USMTF defines different types of groups, "Sets" which define a set of elements (i.e. message components) grouped for a specific usage "Target Location", and "Composites", which imply a group of elements required to define a concept (e.g. "Location" composite needs a minimum of "Latitude" and "Longitude"). These concepts are collapsed into the concept of "group".

"Occurrence". notes whether a message component is required to be used or is optional.

"Repeatability". notes how many times a message component may be utilized in one message instance. It is defines with a range of values: "0", "0 to X", and "X to X" (where "X" is equal to 1 or any integer value), or "X" to "Infinity". The Repeatability value should be in sync with the "Occurrence" value. For example, if a message component is "Mandatory", then the Repeatability value should be "1".

"Context or usage". While a name can provide context, like the names "Enemy Location" or "Friendly Unit Location", in some instances, the name may not provide sufficient context, and would need to be supplemented with additional information. Therefore, providing a means of documenting "Context or usage" to every message component is recommended to allow a message designer to add clarification or additional information when needed.

"Rules". Several types of rules are needed. One set of rules governs the creation of a message instance, for example, in USMTF, a message instance rule states that if a specific set (i.e. group) is used than the next set may not be used, and vice-versa. Rules at the message component level often establish "if, then" conditions for the use and non-use of the message components. For example, if "target location" is used, then "Another set of rules govern message exchanges, including triggering events, refresh rates, and timeliness. Then another set of rules govern system to system operation. A place for a rule should be made at every level of design and no design formalism should constrain a message designer from inserting rules where needed, whether it is in a message format or in the main body of the text.

"Description". The description provides a means to explain the purpose and meaning of a message component, which may be different from how it is used within a specific message. For example, a "Location" component has a persistent meaning, which won't change whether it is used within a message group "Friendly Location" or "Enemy Location".

The Role of Attributes

In this context, *attributes* are used to supplement a message component. None of the examined message standards used the concept of attributes as part of a message format. Attributes provide real value when used with message components because they enable multiple uses of a single concept. When not used, the information that could be defined by an attribute must be instead expressed by multiple instances of a message component. A good example is for the component "Distance" where an attribute may be used to define the "Unit of Measure". Without the use of an unit of measure attribute, there's a need to have multiple "Distance in Meters", "Distance in Feet", "Distance in Nautical Miles", message components whose only difference is the type of measurement being used. The use of attributes is advocated to reduce redundancy. No criteria to determine which datum is an attribute vice a message component (or "element") is fixed, as its best usage should be determined on a case by case basis to determine which approach provides the most value for that message component in a particular message. Therefore, the use of attributes may not need to be affixed as part of a message component's core definition, but rather, should only be assigned when assigned in a specific message format.

Example Table View of a Message Format

Based on the analysis, the following example message format is being proposed that includes the common set of message component data points previously discussed and includes the use of attributes. The example is provided in a table format. The contents are notional and are not taken

directly from any message standard. The internal lines were used as a visual aid to assist in seeing grouped components, like “Unit Location”.

Sequence	Classification	Group Name	Element	Child Element	Attribute	Occurrence	Repeatability	Description
1.0	U	Message Purpose				Mandatory	1	Message purpose
1.0.1	U				Classification	Mandatory	1	Classification of the Location element
2.0	U	Unit Location						The location of a friendly UNIT
2.0.1	U				Classification	Mandatory	1	Classification of the Unit Location
2.1	U		Unit					Describes a military unit (e.g. squadron, battalion)
2.1.1	U			Unit ID		Optional	0	Formal identification of a military unit as provided by appropriate authority
2.1.2	U			Unit Name		Mandatory	1 to 5	The formal designation of a unit
2.2	U		Location			Mandatory	1 to any number of repetitions	The location of a point on the earth's surface
2.2.1	U			Latitude		Mandatory	1	Angular distance from the equator to a point...
2.2.1.0	U				Unit of Measure	Mandatory	1	Identifies the unit of measurement used (e.g. "Degrees", "UTM 10-METER")
2.2.2	U			Longitude		Mandatory	1	
2.2.2.0	U				Unit of Measure	Mandatory	1	Identifies the unit of measurement used (e.g. "Degrees", "UTM 10-METER")
3.0		Target						The location of a UNIT
3.0.1	U				Classification	Mandatory	1	Classification of the Target element
3.1	U		Location			Mandatory	1	The location of a point on the earth's surface
3.1.1	U			Latitude		Mandatory	1	Angular distance from the equator to a point...
3.1.1.0	U				Unit of Measure	Mandatory	1	Identifies the unit of measurement used (e.g. "Degrees", "UTM 10-METER")
3.1.2	U			Longitude		Mandatory	1	Angular distance from the prime or zero meridian...
3.1.2.0	U				Unit of Measure	Mandatory	1	Identifies the unit of measurement used (e.g. "Degrees", "UTM 10-METER")
3.2	U		Target Type			Optional	0 to 10	The type of target.
4.0	U	Operation				Mandatory	1	Name of the operation "e.g. Eagle Eye")
4.0.1	U				Classification	Mandatory	1	Classification of the Location element

Figure 3. Table View of an Example Message Format

Defining a full message standard is, as stated, out of scope of this paper, as the goal was to define an overall message format that may or may not be supported by a formal message standard. However, for clarity, it is noted that the appearance of a message component in a message format is supported by more detailed component descriptions which are in turn these are supported by an even more detailed set of characteristics. The arrow indicated a link to itself notes that those message components may be defined by a set of other discrete components. “Location” for example, is composed of sub-components that include “Latitude” and “Longitude”. The message component “Latitude” would have its own distinct set of metadata, and its own a set of characteristics, including the allowable range of values (e.g. “0-90”), the number of characters, format rules, etc. The sum of these detailed component descriptions and characteristics logically comprise a data element dictionary.

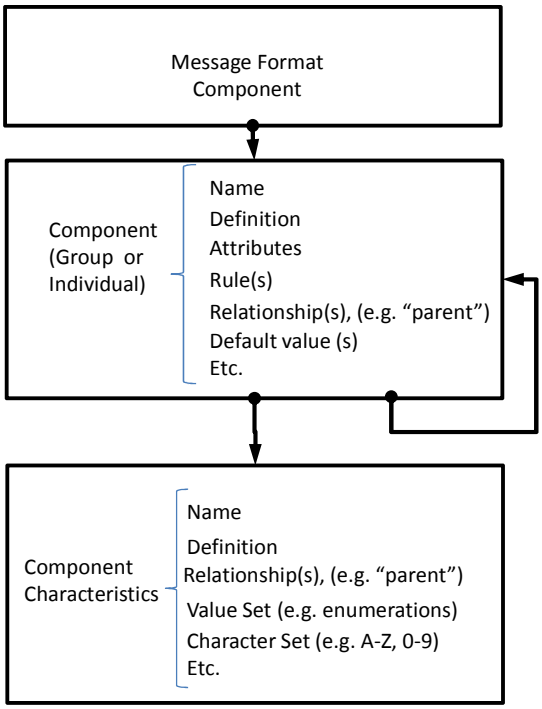

Figure 4. Example of Message Component Documentation

The arrow indicated a link to itself notes that those message components may be defined by a set of other discrete components. "Location" for example, is composed of sub-components that include "Latitude" and "Longitude". The message component "Latitude" would have its own distinct set of metadata, and its own a set of characteristics, including the allowable range of values (e.g. "0-90"), the number of characters, format rules, etc. The sum of these detailed component descriptions and characteristics logically comprise a data element dictionary.

Summary

In summary, an examination of DoD COI schema related products has revealed major opportunities for improvement in their design and documentation. Several factors, including the general lack of understanding of appropriate message format design, limited access to appropriate tools, and lack of XML expertise among the COI's constituents, as well as the convergence on XML as an information exchange format, has resulted in the use of XML as a driver in the design and documentation of COI vocabularies and message formats. While this paper strongly advocated the continued use of XML for information exchanges and the use of XML schema, the overall recommendation is for a text-based approach to message format documentation, and suggested a set of data points and metadata, derived from three core military messaging standards for use in future COI vocabularies and information exchanges development. The result of this effort is an increased assurance by the SME's and developers that the resultant product meets their goals and expectations as the material would be more accessible by the non-XML expert.