I-172 – Final Paper for the 13th International Command & Control Research & Technology Symposium – C2 for Complex Endeavors

Title: Lessons Learned in Using and Adapting an Information Exchange Data Model

Topics: C2 Collaborative Technologies for Network-Centric Operations;
C2 Concepts, Theory, and Policy; C2 Architectures

Submitted by: Sam Chamberlain, PhD, Primary POC.
U.S. Army Research Laboratory
In Support of The Joint Staff / J8 / MASO
ATTN: AMSRD-ARL-CI-CT
APG, MD 21005-5067
410-278-8948; Fax: 4988
sam.chamberlain@us.army.mil
chambesc@js.pentagon.mil

and

LTC Ilean Keltz, USA, PhD
The Joint Staff / J-8 / MASO
8000 Joint Staff Pentagon
Washington, DC 20318-8000
703-697-7434
ilean.keltz@js.pentagon.mil

This Page Left Intentionally Blank

I-172 – Lessons Learned in Using and Adapting an Information Exchange Data Model

Dr. Sam Chamberlain

U.S. Army Research Laboratory
In Support of The Joint Staff / J-8 / MASO
ATTN: AMSRD-ARL-CI-CT
Aberdeen Proving Ground, MD 21005-5067
410-278-8948
sam.chamberlain@us.army.mil, or
samuel.chamberlain@js.pentagon.mil

LTC Ilean Keltz, USA, PhD

The Joint Staff / J-8 /
Models and Analysis Support Office (MASO)
8000 Joint Staff Pentagon
Washington, DC 20318-8000
703-697-7285
ilean.keltz@js.pentagon.mil

Abstract

A primary objective of the Global Force Management Data Initiative is the deployment of a suite of information sources called *organization servers* (OS) that provide default organizational and forces structure data for the Department of Defense (DOD). The data in the OSs are produced and maintained by the agencies across the DOD who are responsible for this information. From the net-centric perspective, these seven sources are seven URLs on the Global Information Grid. The GFM Community of Interest (COI) made a decision to create a unified front across the suite of OS so that an information consumer can retrieve the managed organizational and force structure data in the same form regardless of which source was accessed. This decision allows one to present a virtual DOD OS even though multiple sources are used. To accomplish this, a common access format had to be developed and agreed upon. This paper addresses several issues, some contentious, and lessons learned from their resolution which resulted in the creation of the GFM Information Exchange Data Model (IEDM), an augmented subset of an existing IEDM, and its associated transition to an XML Schema Definition (XSD).

1. Introduction

A primary objective of the Global Force Management Data Initiative (GFM DI) is the deployment of a suite[1] of information sources called *organization servers*[2] (OS) that provide access to default organizational and forces structure data for the Department of Defense (DOD). The data accessed via the OSs are produced and maintained by the agencies across the DOD who are responsible for this information; consequently, there are currently seven servers being developed: one by each Service, one by the Joint Staff that includes the combatant command headquarters, and two by the Office of the Secretary of Defense that includes a special OS to handle the needs of the subset of organizations that make up the intelligence community. From the net-centric perspective, these

[1] Suite: a group of software programs sold as a unit and usually designed to work together.

[2] The term "server" is used in its original meaning: a software application program that accepts connections based upon a request / response paradigm. In this usage, it does not mean a physical computer system.

seven sources are seven URLs[3] on the Global Information Grid, and in the GFM DIcommunity they are referred to as the seven GFM DI components.

The semantics of the data accessed via the OSs is defined using the GFM Organizational and Force Structure Construct[4]. The OFSC is independent of any specific representation, to include the IEDM chosen to specify the interfaces for the OS. However, allowing multiple interpretations and definitions of the properties and meta-data of the OFSC complicates the unification and validation process across the GFM DI components, so the GFM Community of Interest (COI)[5] decided to provide a unified front across the OSs to facilitate ease of integration by the substantial and diverse set of users of the OS data.

2. Lessons Learned

2.1 Lesson 1

A somewhat contentious issue was the decision by the GFM COI to create a unified front across the suite of OS so that an information consumer can retrieve the managed organizational and force structure data in the same form regardless of which source was accessed. This allows one to quickly develop a virtual DOD OS even though multiple sources are used. This leads to the first lesson learned:

Lesson 1: The details of the Net Centric Data Strategy (NCDS)[6] can be interpreted and implemented in a variety of ways.

A key decision in implementing the NCSD is the delineation of boundaries among information sources. In this case, the debate was whether each GFM DI component was an independent authoritative data source (ADS), or whether the collective group of components was considered the authoritative source. Several options are illustrated in **Figure 1**. The COI decision was that the suite of OSs would appear as a single community ADS to its external users, and therefore, would have a common interface specification. This is illustrated in **Diagram Z** in **Figure 1**. There were two reasons for this:

First, there is not a single GFM DI application that will access the OSs, but hundreds of applications. If the OSs do not implement a single interface specification, then every application that uses the OS data will have to build an interface to each OS. This is illustrated in **Diagram X** in Figure 1**.** Seven (or more) OSs would mean seven interface specifications, and in the big picture, it is worth expending the effort to provide a unified front across all the OSs to the potential hundreds of applications that will use the GFM DI community org data.

---

3 URL: Uniform Resource Locator – a web address.

4 See: https://jds.pae.osd.mil/GFM_Secure/GFMCOI_FSC_New.htm (.mil and CAC required)

5 The GFM-COI was established in the summer of 2003 by the Joint Staff, Force Structure, Resources, and Assessment Directorate (J-8) and the Office of the Under-Secretary of Defense for Personnel and Readiness (USD(P&R))

6 See: http://www.defenselink.mil/cio-nii/datastrat/index.shtml

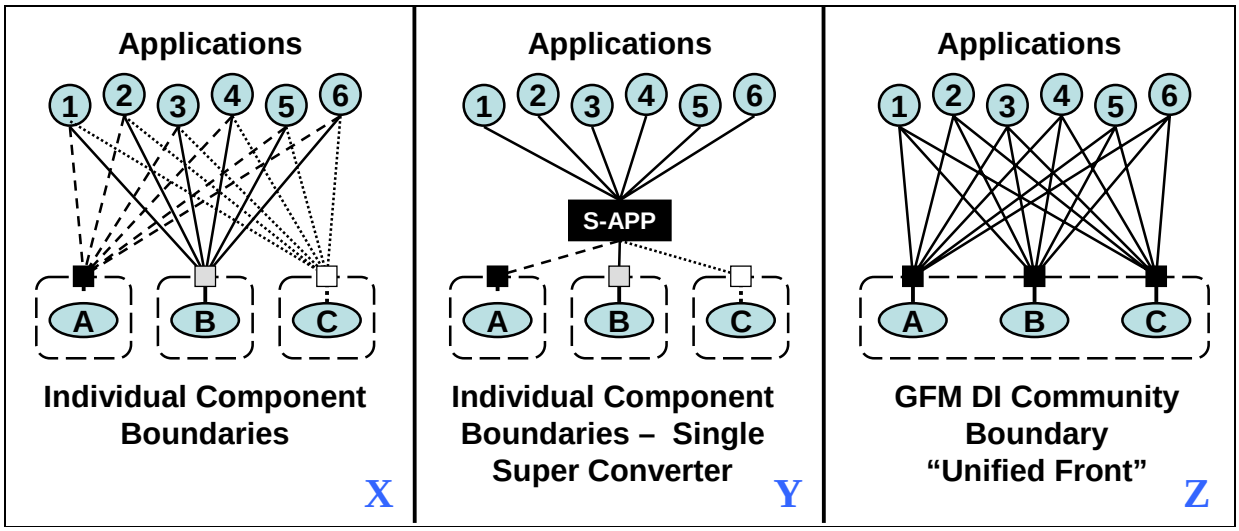

Figure 1: Community Boundary Options for the GFM DI

Second, the Services are indeed different enough in their approaches to the process of developing force structure that *only they* can correctly do the conversions, translations, and mappings to a common semantics and format. It is not realistic to expect to build a "super application"(S-APP) converter that can handle all the diverse aspects of the components to produce a common interface as illustrated in **Diagram Y** in Figure 1. The primary impetus of the GFM DI is to "fix" and enhance the force structure data so that it is conducive to machine manipulation thus allowing trusted results to be routinely calculated. In spite of the optimistic potential of meta-data, without a uniform format (at some level), it is pragmatically impossible to verify that all the OSs represent the same semantics so that the data can be combined consistently. Further, without a uniform syntax accompanying the common semantics, testing and verification of the unified front is much more difficult and complex.

2.2 Lesson 2

Once the decision was made to provide the unified front, a common interface specification had to be developed and agreed upon. There were several alternative ways to accomplish this, but rather than create yet another new interface specification, the COI decided to exploit an existing schema. This leads to the second lesson learned:

Lesson 2: The pros and cons of reuse versus building from scratch are about equal.

In building an interface specification, the GFM COI had two obvious courses of action: exploit an existing specification or create its own. The COI decided to use an existing product from the Multilateral Interoperability Programme (MIP), and in particular, an information exchange data model (IEDM) called the Command and Control IEDM (C2IEDM), that later evolved into the JC3IEDM.[7] The reasons were four fold. First, a subset of the existing product was "good

[7] JC3IEDM: Joint Consultation, Command and Control Information Exchange Data Model, the result of the Multilateral Interoperability Programme (MIP) combining efforts with the NATO Data Administration Group. See: http://www.mip-site.org/

enough" to represent the force structure data for the GFM DI. In the end, only 12 (top level) entities of the 200 plus entities of the JC3IEDM were incorporated. Second, the JC3IEDM evolved out of the intellectual efforts of more than a dozen countries. It already has significant exposure by many coalition partners, and when it becomes ratified as a STANAG (5525), it provides a convenient path to coalition interoperability that is concurrent with the vast efforts already being expended to achieve joint interoperability within the DOD. Third, the theory and tools associated with data modeling are typically well-understood by the data design and implementation community, so it was considered an "easier sell" than more sophisticated approaches. If the COI had decided to develop a new interface, then a more modern and expressive tool, such as OWL, might have been used, but the learning curve and development time would have increased. The focus of the GFM DI was to enhance just a small part of the huge data universe required to conduct the complete business of the DOD and the existing IEDM supported this goal. Finally, the normalized form of the IEDM was perfect for ensuring stability of a key attribute of force structure data, the Organization Unique Identifier, or OUID. When implemented correctly, the IEDM already incorporated the compartmentalization features required to maximize "OUID Retention" as associated attributes evolve. This was a key design tenet of the GFM DI, and although other data representation techniques can be implemented to achieve this property, the IEDM was already proven. The result was the GFMIEDM, an augmented subset of the JC3IEDM, and its associated XSDs.

2.3 Lesson 3

Defining what was meant by (or included within) the term force structure was a challenge in itself. This leads to the third lesson:

Lesson 3: The complexity of the problem increases exponentially with the scope of the data.

The basic GFM DI tenet that "Force structure ties everything together" was fortified during this project. A task that was more challenging than expected was defining the scope of the term "force structure." In the *DOD Dictionary of Military and Associated Terms*[8] the term is redirected to *military capability* where it is defined as: "Numbers, size, and composition of the units that comprise US defense forces; e.g., divisions, ships, air wings." Once again, a debate ensued as to where the boundaries are defined for a "US Defense Force." One perspective was that an organization was only included as part of force structure if the it was assigned to a combatant command. Because all organizations of the DOD were to be included, the word *organizational* was added to make the title of the semantics the OFSC (rather than the original FSC).

A fundamental property of the GFM DI force structure data is that it transcends the traditional unit identification code (UIC) boundary and extends seamlessly down to the billet level.[9] This adds more complexity because two historically separated environments with different perspectives also had to be merged: the maintainers of the upper echelon force structure that extends down to the

---

[8] Joint Publication 1-02, *DOD Dictionary of Military and Associated Terms*; see: http://www.dtic.mil/doctrine/jel/doddict/index.html.

[9] For a description of the UIC Boundary, see the 2004 CCRP paper entitled: *A Unifying Strategy for Data Integration for Global Force Management*, http://www.dodccrp.org/events/2004_CCRTS/CD/papers/035.pdf .

UIC level (referred to as the *upper tier*), and the maintainers of the lower echelon force structure that begins at the UIC echelon and extends down to the billet level (referred to as the *lower tier*). Therefore, from either perspective (upper or lower tier), organizational elements (OEs)[10] not part of the traditional organizational structure were added increasing the scope and making the determination of priorities for data properties more difficult. This is illustrated in **Figure 2**.

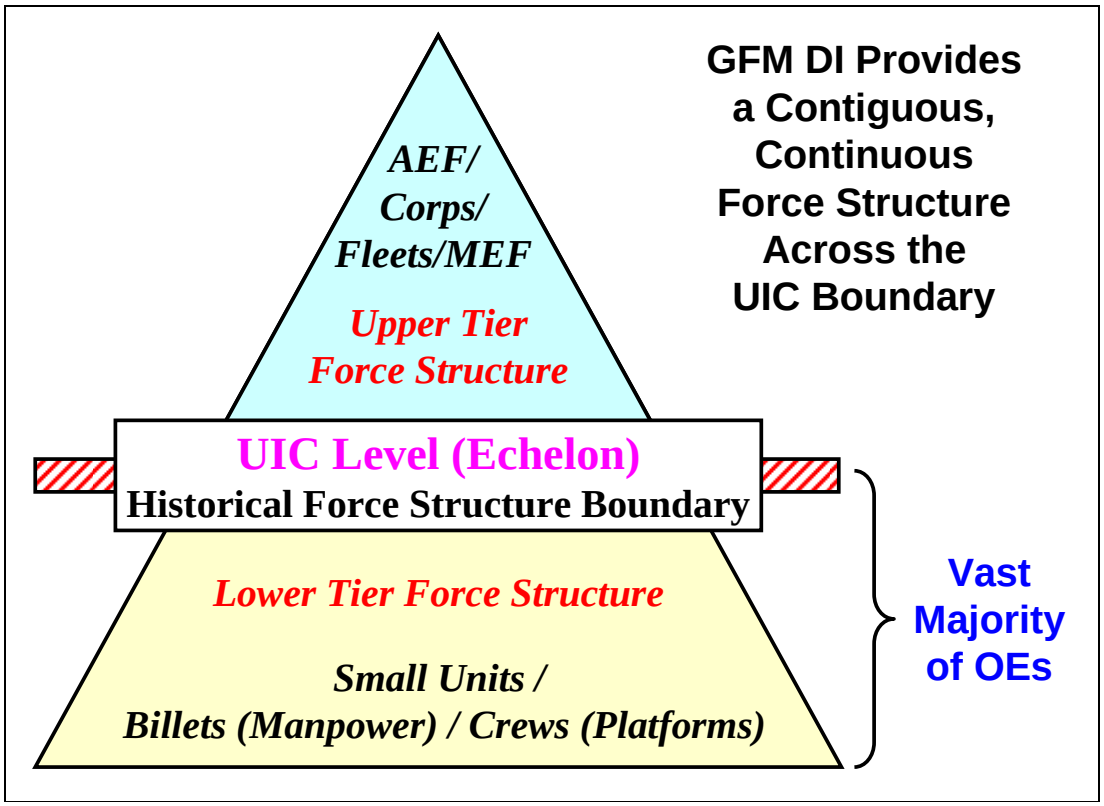

Figure 2: Merging Force Structure Across the UIC Boundary

With this new composite force structure, the COI had to make a decision about how much additional information would be added to the bare bones organization structure that is based upon command and support relationships. This decision was driven in part by the expectations of the OS users who were accustomed to certain content being provided by existing sources of information. The data contained via the OSs will be consumed by a large population of diverse users, all who have expectations driven by their specific requirements and perspectives which vary widely. For example, a person conducting military operations will likely have a different perspective on what is considered useful additional information than someone performing financial tasks. At one extreme, one could provide a sparse "org chart" with nothing more than the name of the organization and its command and support relationships. At the other extreme, one could include extensive information about the properties and features of the organizations, such as that found in the Type Unit Characteristics (TUCHA) data maintained by the Joint Staff J-3 directorate. Therefore, defining and populating the OSs involves both meeting and managing the expectations of a diverse set of users.

---

[10] *OE* – by OFSC definition, an *organizational element* corresponds to a node in an organization tree.

The new composite force structure contains three classes of OFSC OEs: *doctrinal* OEs to support tactics, techniques, procedures, and operations, *billet* OEs to represent jobs, and *crew* OEs to incorporate trackable platforms. The upper tier is composed almost entirely of doctrinal OEs (the few exceptions being ship crews that sit on the tier boundary) and these have had "unit characteristic" data associated with them for a long time. Several well-established sources exist for this information and are maintained by the Services and Joint Staff. However, this is not true for most lower tier doctrinal organizations even though the implementation of doctrinal organizations in the lower tier is no different than those in the upper tier. The characteristics of lower tier doctrinal organizations (e.g., small units) exist, but not in database form. Instead, they are described by Service and Joint operations manuals and documents. Consequently, adding this type of information to the lower tier OEs was a new task for which a process and procedures had to be established thus expanding the scope and adding further complexity to the product.

With few exceptions, billet and crew OEs reside in the lower tier (once again, the exception being ship crews that sit on the tier boundary). The OFSC semantics include decomposition as the default function for defining organization structure. As a result, billet OEs must reside at the bottom of the org tree because they can not be decomposed further into smaller OEs. Therefore, all billets reside in the lower tier and this causes a significant increase in scope because over 70% of the OEs in the org servers are billets (over 3 million active, guard, and reserve military and civilians members[11]). Similarly, most crew organizations (created to represent groups operating platforms) reside in the lower tier. The introduction of the concept of a crew as an OE is new and causes a challenge because current practices focus on the tangible platform hardware that the crew operates rather than on the intangible crew. Therefore, new principles and guidance had to be established to characterize crews. In any case, billet and crew OEs provide the point at which the organizational and force structure domain is integrated with the personnel and materiel domains. Although the complexity incurred by their inclusion presents additional challenges, it is worth the effort because the results can be exploited to combine disparate pieces of routine battlespace information by using the common theme of force structure.

Force structure and manpower have traditionally been closely related and are often referenced together. For billets, the COI had to decide whether to include in the OS specification information about the qualifications to occupy a billet. The starting point to address this question was to evaluate existing Service and Joint manpower documents. It was clear that current manpower documents include not only the name of positions (i.e., job title) and approximately where they are situated in the organization structure, but also a set of requirements to occupy the position. As a common practice, the individuals that build and maintain these documents define both the job titles (the minimal requirement for an OS) and the associated qualifications to hold the job. Further, a small set of common descriptors exists across the Service and Joint documents that define the types of qualifications required to occupy military or civilian positions. Therefore, the COI decided to include these in the OS specification because they are currently expected as part of the force structure information and it would have been more disruptive not to include them.

However, adding this information significantly increased the complexity of the OS specification and was the primary impetus for augmenting the existing IEDM. A primary source of confusion is

---

[11] See "DOD 101" at http://www.defenselink.mil/pubs/dod101/

the ubiquitous practice of using the same qualifications to describe both personnel and manpower. In simple terms, *personnel* refer to people while *manpower* refers to the jobs that they occupy (often nicknamed "faces" and "spaces," respectively). Billets are instances of manpower positions. Confusion arose because, in many cases, the same attribute is used in both domains. For example, the attribute of military pay grade applies both as a qualification of a person and as a requirement to occupy a position, but there are subtle differences. The military pay grade for officers includes cases where a person has prior enlisted experience. This applies to personnel, but not to manpower (i.e., there are no jobs that require enlisted experience). So a restricted version of military pay grade values had to be defined (named just "military grade") for use with manpower. These types of subtle variations added complexity to the expected simple task of added qualifications to positions.

To address the many details associated with manpower data, the GFM COI established a working group with the Human-Resource Management (HRM) COI called the HRM-GFM Collaborative Working Group (CWG). This forum was extremely helpful in getting the right people to answer questions. The HRM-GFM CWG resolved many specification issues early during the development and augmentation of the IEDM. Not all issues were resolved, often because they required coordination across the whole federal government (beyond the DOD). But the realization of the need for interaction with the HRM-GFM COI alone was helpful to the design process.

The same situation occurred with the inclusion of crew OEs. Few concepts in the development of the OS specification caused more debate that the addition of crew OEs. Crews are organizations required to operate platforms. Platforms (by OFSC definition) are equipment that carry its operators. Crews were added to the OS specification for the following reasons: one, they are organizations, two, they are routinely of high interest as they are moving and require tracking, and three, the crew OE concept, with its independence from a particular item of hardware or membership, allows the dynamic nature of these entities to be represented while still offering a degree of stability (as a predefined aggregation point). The details of this example will not be addressed, but it posed a challenge equal to that for billets.

The point of this discussion is that the additional effort incurred by the inclusion of a few seemingly straight-forward enhancements can be subtly deceptive and underestimated as illustrated in **Figure 3**. Every subject is different, but one must be aware of how fundamentally

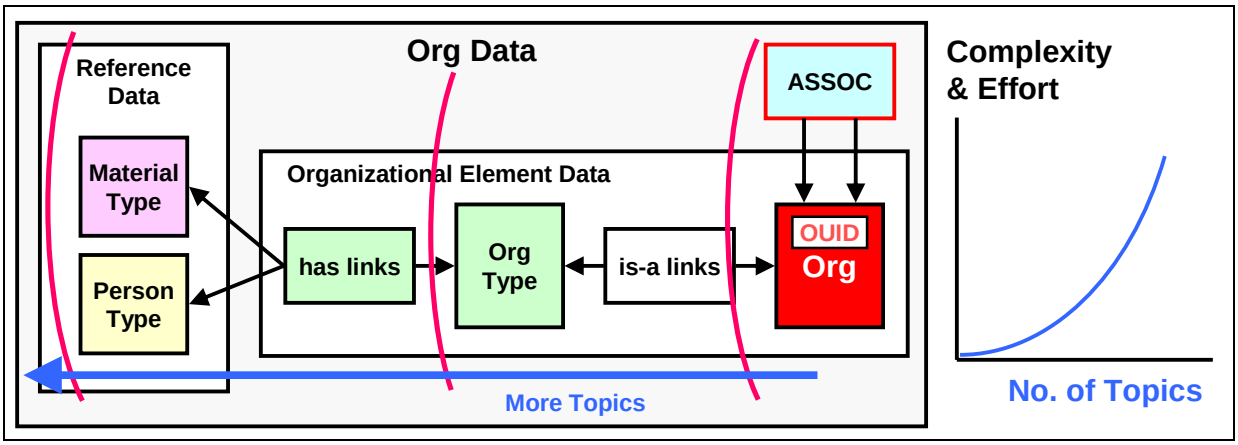

Figure 3: Complexity and Effort Increases Rapidly with Scope

interconnected the topics are to each other and consider this metric when expanding the scope of the domain of information. In many cases, the payoff is worth the effort of expanding the scope, but it may have a significant consequence by increasing the complexity of the task through higher order effects. These effects require more management and coordination due to increased membership of disparate expertise and increase the time required to accomplish milestones.

2.4 Lesson 4

Lesson 4: It is easy to forget what the IE means in IEDM.

An information exchange data model is one of several alternatives for specifying the exchange of information over a network. Because the discipline of data modeling emerged from the database design community, it is easy to presume that an IEDM is defining a (physical) database schema rather than an interface specification. Therefore, the limited scope of using a data model to specify information exchange must be frequently reiterated. In practice in the GFM DI, the IEDM served as a preprocessor for the development of an XSD, the actual specification used in the net-centric environment, as illustrated in **Figure 4**.

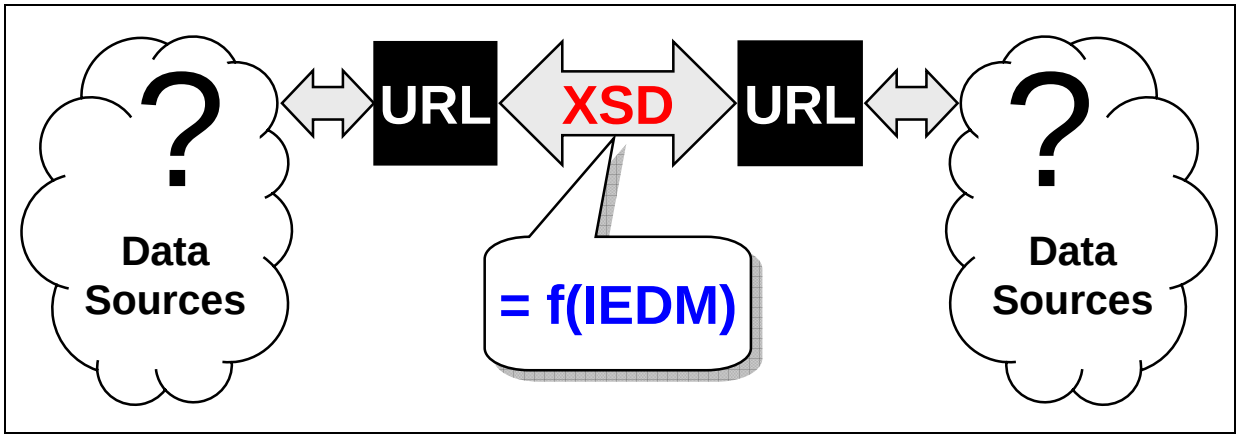

Figure 4: An IEDM Is Not About Physical Storage

Data models have received disparaging reviews due to problems incurred by way they were used and managed by the DOD in the 1990s. General problems have been attributed to data models that are not data model specific, but instead, caused by programmatic decisions independent of their characteristics, most notably, a choice of scope that was much too large. An IEDM is used to specify semantics and syntax, as does an XSD (to widely varying degrees). An issue raised was whether the GFM DI violates the intent of the NCDS by including syntax in the specification for the OS outputs. This argument states that the syntax should be left to the data provider and requiring a specific syntax places an undo burden on them. Instead, each data provider should use the DOD Meta-Data Repository (MDR) to publish its unique syntax (typically, via an XSD) and then the consumer must access the MDR and convert the data provided to the form required.

This is exactly what GFM DI is doing for the suite of OS rather than the individual components. Therefore, this problem is not with data models, but with the decision to provide a unified front across the org servers as described in Lessons 1 and 2. This issue would be present regardless of the technique used to provide the common specification, whether it is based on an IEDM or

another representation scheme to produce a common XSD. Thus, the true culprit was the decision of where to draw the boundaries. The COI decided that the GFM DI XSD would be used by all OS and include a common, rigorous syntax to achieve the results required to achieve OUID stability for the benefit of the force structure data consumers. The complexity (and cost) to the consumers to manipulate, integrate, and interpret several different specifications would greatly outweigh the convenience provided to the producers if allowed to develop their own, unique forms.

In any case, the XSD (or IEDM) does not dictate how the information is maintained and portrayed inside the component systems, but only defines the minimum requirement for how it is made available to the consumers.

2.5 Lesson 5

Lesson 5: Augmentation does not require modification to the original model.

On the first iteration to map the OFSC concepts to the JC3IEDM the approach was simple: directly modify the existing model to meet the GFM DI needs. This involved three basic actions: adding new tables, adding new attributes to new or existing tables, and modifying existing attributes. Soon after the first iteration, a newer version of the existing base model was released that updated some of the attributes modified for the GFM DI. Thus, the shortcomings of this approach were clear and resulted in two realizations triggering a revised strategy.

The first realization was that one should never modify existing attributes. In hindsight this seems obvious, but was initially over looked. In the second iteration, augmentations to attributes were implemented by creating a second attribute with the same name as the original but prefixed with the common string "gfm_". Thus, if additional category codes (an enumerated type) were required for the GFM DI and the original attribute name was "cat_cd," then this original attribute was left unmodified and a second attribute, named "gfm_cat_cd", was inserted that contained the additional category codes, as illustrated in **Figure 5**. Thus, the two attributes could evolve independently in their two development communities.

Legend: Original Codes
New GFM Codes

NOT:
cat_code = { ADMINS, AUGMNT, CMDCTL, NOS, … , COCOM, ISLEDB }

BUT:
cat_code = { ADMINS, AUGMNT, CMDCTL, … }
gfm_cat_code = { COCOM, ISLEDB , NOS }

Figure 5: Augmentation Does Not Require Altering Original Attributes

Because several of the attributes were mandatory, a business rule was developed to delineate which value to use when both versions of the attribute were present. The rule implemented was

simple: check the gfm version first and if it contains the value “NOS” (for “not otherwise specified,” a routine value in the original model), then use the value in the original version. Originally, the attributes were checked in reverse order, but when it was discovered that some original attribute value lists did not contain the value “NOS” (or an appropriate substitute), the order was changed to check the gfm version first.

Iteratively applying this naming convention can be confusing when duplicate tables (with duplicate attributes) were inserted. The original model is based on two fundamental categories of entities: “types” that represent classes (or descriptions) of things, and “items” that represented actual instances of the types. There were several cases which a table existed in one of the categories, but for GFM, was also required in the other category. An example was a table for aliases that existed for items in the original model but not for types. So in the GFM augmentation of the model an alias capability for types was added with many of the same attributes as the original. Using this approach, the new table name was prefixed with “gfm_” as was each of the existing attributes. However, a few of the attributes of the original table already had gfm counter-parts with the “gfm_” prefix, so these ended up with a gfm prefix on the gfm prefix (e.g., “gfm_gfm_”). This naming convention was chosen for consistency rather than elegance.

The second realization was recognizing that compliancy did not require duplicity. One reason for using the existing data model was for coalition interoperability. Upon reflection of the purpose of an IEDM, the COI realized that one did not have to exactly duplicate the original model provided, but rather be able to automatically reconstruct it. The revised requirement was the automatic conversion between the GFM XSD and the JC3IEDM XSD without any user intervention. This new approach resulted in a major reduction in the size of GFM version of the model. First, any JC3IEDM attribute that was not mandatory and was not identified for GFM DI use was removed from the GFM DI version of the model. Previously, these were left in and annotated as not required for the GFM DI. Although this reduction seems obvious in retrospect, by forgetting the purpose of an IEDM it was easy to leave in attributes to maintain the appearance of compliancy. Second, mandatory attributes that could be derived by the data at hand (i.e., without doing any additional queries) were removed from the gfm version. This ensured that an XML Stylesheet Language Transformation (XSLT, or .xsl) file could be developed to automatically convert a GFM XML file into a JC3 XML file. As a result, the latest GFMIEDM XSD is significantly smaller than its previous versions.

One choice that is being reevaluated was a change of status of a few original attributes from optional (i.e., “Nullable”) to mandatory. This was done in cases where an attribute was important to the GFM DI and its inadvertent omission would be significant. In this case, the original was made mandatory to ensure that it was not forgotten. The presence of a “gfm_” attribute induces a perceived problem that is subtle but important. If an attribute is not mandatory, then in XML it does not have to be sent. In these cases, no special value (like “NOS”) is required in the original attribute to indicate that the attribute does not contain qualifying information. But a mandatory attribute means that some value has to be sent; therefore, some valid values must exist even if it doesn’t make sense. In the GFM DI this is not a problem because in these cases the mandatory

attribute should never be used.[12] However, in the general case, it means that either everyone can pick any value to send, or one value is arbitrarily selected (in the GFM DI, the first value in the list is used). This does not cause a problem, but may cause confusion.

One other discovery was to be wary of over specified data types in the IEDM. An example was for category codes that were specified as a fixed size even when smaller codes were allowed. In the database realm this is often implemented by padding shorter values with spaces; however, when the IEDM is converted to an XSD this can be lost. The result is that if one desires to use the IEDM specification for data insertion, then the arrival of a short value (smaller than the fixed size) has to be handled by padding the value before it can be manipulated by the database. This annoyance can be avoided by simply not using fixed sized data-types in an IEDM unless they really are fixed; instead, use variable sized data-types so that constraints are easily met.

The final result is that the augmented model must be transformable, but does not have to be identical to the original. In the GFM DI, the decision was made to automate this transformation and not require any user action. This significantly reduced the size of the augmented model.

2.6 Lesson 6

Lesson 6: Remove enumerated data values from the IEDM whenever possible.

Traditionally, data models use enumerated types for many attributes. This means that a set of approved, or valid, values are listed for the particular attribute and this was a common feature of the original IEDM.[13] The problem this creates for an IEDM is that any change to a list induces a change to the XSD, which in turn requires re-evaluation and agreement by the COI. To minimize this effect, one strives to keep the XSD (and IEDM) as invariant as possible. To accomplish this, new attributes that would traditionally be defined using a list of valid values are defined by a reference to a table of the values, as illustrated in **Figure 6**. In other words, the values are represented as data outside the XSD just like any other data so that changes to the list do not affect the XSD. This was done for new GFM DI attributes whose values had the potential to be numerous and variant.

A typical example was the list of attribute and associated types that define manpower requirement. Rather than incorporate a list of, for example, all the USAF Officer Specialty Codes, a table was created to hold them so that they could be referenced. Using this approach, new codes can be added and old codes deleted without affecting the XSD because this is just a change to the GFM DI data. The interesting consequence of this approach is that the traditional "data dictionary" is no longer confined to the data model (or XSD), but must also include these special "reference" tables that exist outside the interchange specification. This does not cause any special problems, but must be recognized since the traditional data dictionary is distributed across several sources that collectively describe the properties of the specification.

---

[12] If the original value is to be used, then the "gfm_" sibling will have "NOS" as the value. If the "gfm_" sibling contains a value other than "NOS" then it is to be used and the original attribute, which must be included, is ignored. Its value is irrelevant, but it must be there.

[13] A perfect example is an attribute for aircraft type that includes a list of 4,758 aircraft models world-wide from which to choose. A change to one of these values is a revision to the data model resulting in a new version.

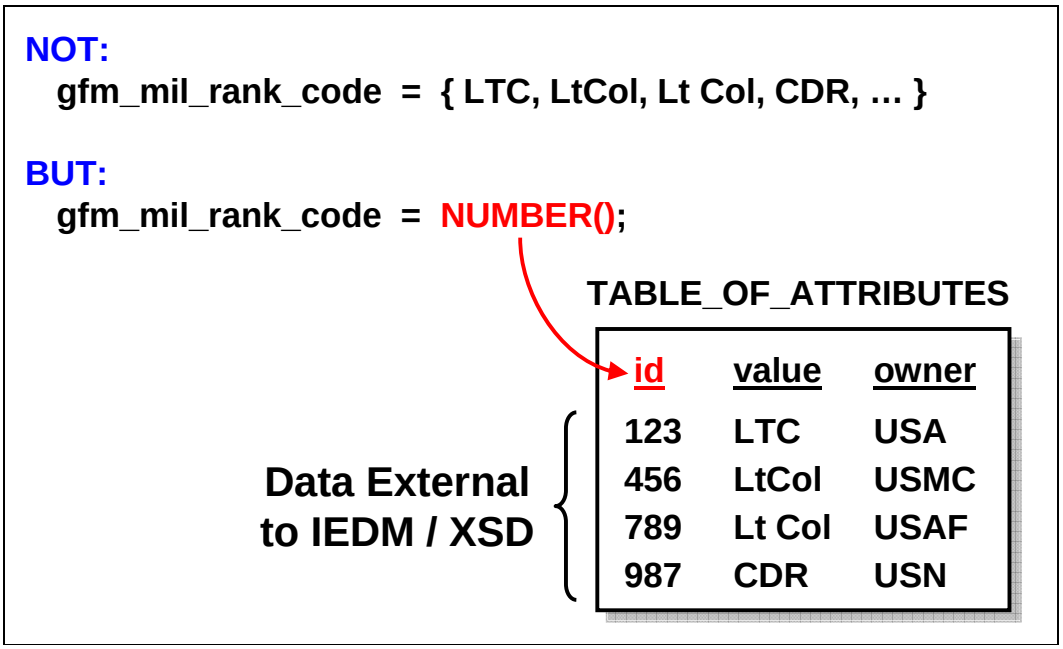

Figure 6: Avoid Enumerated Types Embedded as Part of the Specification

2.7 Lesson 7

Lesson 7: Identifier management remains an issue.

Since its inception, identifier management has been one of the primary elements of the GFM DI. Identifier management issues actually revolved around two separate themes. The first theme, alluded to in Lesson 4, is due to the structure of the XSD that results from using an IEDM as a preprocessor for its development, and the second theme is the characteristics of the identifiers themselves.

For the first theme, the contentious feature is that the resulting XSD reflects the properties of the tool used to generate it; in this case, a normalized, relational data model. Most noticeably, this means that the information is broken into pieces and those pieces require identifiers. Therefore, if one does not use a schema similar to that of the XSD, additional effort is required to map the data to the XSD format. Whether this property is worse for a relational data model versus other representation schemes is debatable, but regardless of the common schema used, similar transformations would have to be invoked. In reality, the ultimate root of this problem is the decision described by Lesson 1 to use a common format to provide a unified front across the systems.

From the identifier perspective, the problem encountered within the GFM DI is that the identifiers are expected to be stable and persistent commensurate with the data. Therefore, the identifiers must be retained for recurring use and this places additional constraints on the legacy systems. The counter argument is that because of the rigor of the semantics and accompanying definitions of the GFM DI, this task can be automated and hidden from the operators. The deeper origin of this matter is that in many cases the data is not maintained in as rigorous a mode as required by the GFM DI, thus the recurring theme that “this in not just an information technology (IT) task” must

be reiterated. The truth is that some of the data (e.g., command and support relationships) do not exist and have to be represented to successfully meet the technical and operational requirements of the GFM DI. The application of pure IT solutions can not address this deficiency. In the meantime, GFM XSD identifiers (called Force Management Identifiers, or FMIDS) must be maintained in concert with existing data sources so that GFM DI data remains consistent.

For the second theme, identifier characteristics, the GFM DI imposes some special requirements. Of particular interest is the possibility that referential links associating disparate pieces of data may be disseminated or made accessible independent of the data to which it references. For example, this can occur during data maintenance that includes only updates to the data, or with task organization data that spans across organizational domains where organizations from two or more Services are associated but do not all reside in the same data source. In these cases, it is plausible that the data to which an identifier refers must be tracked down, consequently, some form of identifier tracking service must be provided.

The requirement to track identifiers was one of the reasons the Enterprise-wide Identifier (EwID) scheme was chosen for the GFM DI.[14] Due to the inclusion of both global and local aspects to the allocation scheme, the features of the global information grid can be exploited to quickly track down the source of an identifier via a series of user defined Uniform Resource Identifiers (URI). To accomplish this, the global portion of the allocation process requires registration with a centralized site. If the tracking service capability is not fully appreciated, then this approach may be perceived as unnecessary, excessive, and burdensome. Therefore, it was important to emphasize the reasons for the selection of this particular identifier strategy. For more details on this identifier technology, see the papers referenced in footnote 14.

As the DOD directive on unique identification is realized more data with unique identifier schemes are emerging.[15] This has caused a welcomed decrease in the scope of the GFM DI and the XSD. For example, as the Real Property community establishes their data sources, data previously resident in the GFM DI regarding home stations (i.e., facilities) is replaced with a reference to a Real Property unique identifier (RPUID). As intended by the NCDS, the RPUID can then be used to access and collect official data about a facility from an official authoritative data sources. Likewise, references via OUIDs will occur from the Real Property community information sources to those of the GFM DI community.

---

[14] Chamberlain, "An Enterprise Identification Strategy for Global Naming Across Arbitrary Data Systems**,"** *Proceedings of the 6th International Command and Control Research and Technology Symposium (IC2RTS)*; US Naval Academy, Annapolis, MD; Presented 19 June 2001. http://www.dodccrp.org/events/6th_ICCRTS/Tracks/Papers/Track2/059_tr2.pdf, and "Implementation of an Enterprise Identifier Seed Server for Joint and Coalition Systems," *Proceedings of the 7th International Command and Control Research and Technology Symposium (IC2RTS)*; Quebec City, Quebec, Canada; Presented 17 September 2002. http://www.dodccrp.org/events/7th_ICCRTS/Tracks/pdf/109.PDF .

[15] DOD Directive 8320.03, *Unique Identification (UID) Standards for a Net-Centric Department of Defense*, 23 March 2005; see: http://www.dtic.mil/whs/directives/corres/pdf/832003p.pdf.

Identifier requirements and implementation details remain a source of deliberation in the GFM DI. Because the mandated identifier scheme is implemented, the characteristics of the identifiers are openly available leaving most of the discussion revolving around the first theme. For those not familiar with the characteristics of a normalized data model, the question of "what needs to be identified" had to be addressed in the context of partitioned, stable data entities. The fact that FMIDS are actually a set of (12) identifiers has to be re-emphasized in light of the attention given to the premier GFM DI identifier, the OUID. FMIDS retention — the objective of keeping the set of FMIDS as stable as the data they tag — continues to be a topic of concern and rules for achieving this objective are being refined through evaluation and the experiences gained through implementation.

2.8 Lesson 8

Lesson 8: Accessibility does not obligate permission to access.

The NCDS calls for making data visible, accessible, and understandable, and for promoting trust. By far, the most difficult task appears to be promoting trust which in turn impedes visibility and accessibility. Technical accessibility is the first step, but the ultimate challenge is gaining permission to access "the accessible." The GFM DI OSs provide no information about real people, equipment, location, missions, task organization, or any other operational information. Yet, as innocuous as this seems, there is serious objections by some components about sharing force structure data. There are many reasons, but the root factor is mistrust of how the information may be used against the component. In spite of the plethora of public praise and optimism about sharing data, many in DOD are hesitant about sharing even the most routine data. Therefore, the policy component of the GFM DI became as important as any other.

To address these types of issues, a separate Policy, Integration, and Process Working Group (PIPWG) was established with membership at the O-6 level. This is the first echelon to address issues raised by the component OS developers, and many times solutions or alternatives can be developed and agreed upon at this level. Issues not successfully resolved by the PIPWG are forwarded to the next level, the GFM DI General/Flag Officer Steering Committee (GOSC). This body provides further scrutiny and perspective to the issue resolution. As expected, it is highly desirable to achieve agreement at this level because escalation moves to the upper echelons of the DOD. One of the few issues which were not resolved at the GFM DI GOSC was data access. So it is evident that accessibility and visibility are two very challenging topics of the NCDS. As is sometimes the case, the difficulties posed by the technical aspects of the task can be surpassed by the procedural ones.

3. Summary

This paper discusses eight lessons learned from using an information exchange data model (IEDM) as the development tool for an interface specification represented via an XSD. The first lesson discussed what may be perceived as a data representation matter is actually the result of a policy decision:

Lesson 1: The details of the Net Centric Data Strategy can be interpreted and implemented in a variety of ways

In the GFM DI, this was caused by a decision to set interface boundaries that encompassed the GFM DI suite of organization servers that resulted in the use of a common interface. This reaction can occur regardless of the specification representation chosen.

Next are three lessons that deal with general implementation topics such as scoping the task and selecting an approach:

Lesson 2: The pros and cons of reuse versus building from scratch are about equal
Lesson 3: The complexity of the problem increases exponentially with the scope of the data
Lesson 4: It is easy to forget what the IE means in IEDM

Throughout these three lessons are references back to Lesson 1 and complications that arise more from the selection of interface boundaries than from the actual implementation approach.

The next three lessons deal with technical issues associated with the choice of implementation:

Lesson 5: Augmentation does not require modification to the original model.
Lesson 6: Remove enumerated data values out of the IEDM whenever possible.
Lesson 7: Identifier management remains an issue.

The main point with these topics is that technical decisions do have an impact and one can lubricate the implementation process by selecting wisely. Unfortunately, these lessons are often learned through experience.

The last lesson reverts to policy.

Lesson 8: Accessibility does not obligate permission to access.

In spite of the best intentions, one has to be willing to share data even when it is technically visible, accessible, and understandable. As expected, leadership will remain a key element of net-centricity as trust must be instilled among the data providing components.

In conclusion, some difficulties were caused because of data representation conflicts while others were caused by differing interpretation of the Net-Centric Data Strategy thus requiring face-to-face meetings for resolution. To date, face-to-face meetings have been copious, regular, and an important aspect to the success of the GFM DI. There is simply no shortcut to replace human interaction to achieve understanding. Every attempt was made to reduce conflicting approaches to technical issues, but this was not always possible. As a result, compromise and experience were often a key factor in the resolution of issues. But to quote the comedian Steven Wright: "Experience is something you don't get until just after you need it."

4. **References**

Chamberlain & Sprung; "A Unifying Strategy for Data Integration for Global Force Management; "*Proceedings of the 2004 Command and Control Research and Technology Symposium*; Loews Coronado Bay Resort, San Diego, CA; 15-17 June 2004. http://www.dodccrp.org/events/2004/CCRTS_San_Diego/CD/track02.htm

Chamberlain, "Implementation of an Enterprise Identifier Seed Server for Joint and Coalition Systems," *Proceedings of the 7th International Command and Control Research and Technology Symposium*; Quebec City, Quebec, Canada; 16-20 September 2002. http://www.dodccrp.org/events/7th_ICCRTS/Tracks/pdf/109.PDF .

Chamberlain, "An Enterprise Identification Strategy for Global Naming Across Arbitrary Data Systems**,"** *Proceedings of the 6th International Command and Control Research and Technology Symposium*; US Naval Academy, Annapolis, MD; 19-21 June 2001. http://www.dodccrp.org/events/6th_ICCRTS/Tracks/Papers/Track2/059_tr2.pdf

DOD Directive 8320.03, *Unique Identification (UID) Standards for a Net-Centric Department of Defense*, 23 March 2007; USD(AT&L)/USD(P&R), see: http://www.dtic.mil/whs/directives/corres/pdf/832003p.pdf.

DOD Instruction 8260.03, *Organizational and Force Structure Construct (OFSC) for Global Force Management (GFM)*, 23 August 2006, Under Secretary of Defense (Personnel and Readiness). See: http://www.dtic.mil/whs/directives/corres/html/826003.htm.

DOD 8320.02-G: *Guidance for Implementing Net-Centric Data Sharing*; 12 April 2006; ASD(NII) / DoD Chief Information Officer; see: www.dtic.mil/whs/directives/corres/pdf/832002g.pdf .

JC3IEDM: *Joint Consultation, Command and Control Information Exchange Data Model*, Multilateral Interoperability Programme (MIP); See: http://www.mip-site.org/

Joint Publication 1-02, *Department of Defense Dictionary of Military and Associated Terms* 12 April 2001, as amended through 31 August 2005; http://www.dtic.mil/doctrine/jel/doddict/