A MODEL-BASED METHODOLOGY FOR MANAGING T&E METADATA

John Hamilton, Ronald Fernandes, Timothy Darr, Michael Graul
Knowledge Based Systems, Inc. (KBSI)
1408 University Drive East
College Station, TX 77840
{JLHamilton, RFernandes, TDarr, MGraul}@kbsi.com

Charles Jones and Annette Weisenseel
812 TSS/ENTI and /ENTT
Edwards AFB, CA
{Charles.Jones, Annette.Weisenseel}@edwards.af.mil

ABSTRACT

In this paper, we present a methodology for managing diverse sources of T&E metadata. Central to this methodology is the development of a T&E Metadata Reference Model, which serves as the standard model for T&E metadata types, their proper names, and their relationships to each other. We describe how this reference model can be mapped to a range's own T&E data and process models to provide a standardized view into each organization's custom metadata sources and procedures. Finally, we present an architecture that uses these models and mappings to support cross-system metadata management tasks and makes these capabilities accessible across the network through a single portal interface.

KEYWORDS

T&E Metadata, Reference Model, Information Retrieval, Ontology, UML

INTRODUCTION

The product complexity and testing requirements of modern test articles have driven the evolution of complex Test & Evaluation (T&E) processes. However, existing T&E processes are lacking in data management support; particularly with respect to data storage and retrieval, and metadata support. The T&E metadata is documented in a diverse set of formats (XML, specialized databases, Microsoft® Office® documents, etc) on a diverse set of systems (SQL Server, Oracle, document management systems, etc). Some of this information is stored in a structured form, such as instrumentation information in the Instrumentation, Loading, Integration, Analysis and Display toolset (ILIAD)'s [1] database. Other information, such as test plans, requirements, and reports, is stored in unstructured formats. Emerging and existing text and XML-based standards such as Telemetry Attributes Transfer Standard (TMATS) [2], TMATS-XML, Data Display Markup Language (DDML) [3], and Instrumentation Hardware Abstraction Language (IHAL) [4] represent a category of "semi-structured" data.

In many cases the same type of information is stored in different formats and on different systems for different test articles. Further, the terminologies used to refer to one type of metadata may vary not only across ranges, but across different groups within the same range. As a result, it is extremely difficult to perform useful metadata management tasks such as cross-referencing and verifying different sources of T&E metadata against each other, quickly retrieving all metadata for a given test or test article, or issuing a single keyword search across all sources of metadata.

The vision of the DoD's Net-Centric Data Strategy aims to make data from multiple, divergent domains visible and accessible. This includes the accessibility of T&E data within the T&E community in a net-centric fashion that is usable for both anticipated and unanticipated users and applications.

As part of a Small Business Innovative Research (SBIR) project with Edwards AFFTC and KBSI, we have developed the Test and Evaluation Metadata Plaza (TEMPL), which consists of a reference T&E metadata model, a prototype architecture, and a methodology for enabling the management of diverse sources of metadata across the network through a single portal interface.

The end-user applications explicitly supported by TEMPL include the following:

1. Metadata browsing, search and retrieval
2. Metadata repository construction
3. Verification, Validation, and Completeness (VV&C) checking.

THE NEED FOR A T&E METADATA REFERENCE MODEL

In its simplest definition, metadata is data about other data. For the purposes of this paper, we view the "T&E data" to be the measurements obtained during a test. Hence, "T&E metadata" is any information that provides additional description or context to the T&E data. This covers a broad spectrum of information, ranging from the initial requirements and motivation for the test, to the test article and instrumentation modifications required to perform the test, to the description of the packet format in which the data is transported.

Having easy access to all of this information through a single interface would enable a test engineer to quickly browse and search for metadata relevant to a given test, and add these individual pieces of metadata into a "metadata repository" for the test. With this metadata repository, ongoing tasks could be validated (e.g. Has all of the necessary information been documented? Does the test article's configuration meet the requirements? etc.), enabling mistakes to be caught earlier in the process. Additionally, post-test forensics could be conducted much more efficiently with all of the relevant information already assembled and on-hand.

Several efforts are currently underway to standardize various portions of T&E metadata, including the Metadata Definition Language (MDL) currently under development as part of the Central Test and Evaluation Investment Program (CTEIP)'s integrated Network Enhanced Telemetry (iNET) project [5], the Test and Training Enabling Architecture (TENA) object model [6], the Range Commander's Council (RCC)'s IRIG 106 standards [2], and KBSI's IHAL [4] and DDML [3] efforts. However, each of these efforts focuses on a single aspect of the T&E process. There is currently no single high-level model that captures all of the metadata generated

in a typical T&E environment. Having such a model would serve as a common terminology for T&E metadata as well as a guide for building a complete metadata description of a test.

T&E METADATA REFERENCE MODEL

The T&E Metadata Reference Model (T&E RefMod) is a high-level model of all of the types of metadata used in the test and evaluation world. The elements of the model represent a common terminology for each piece of information relevant to a test. The purpose of the T&E Ref Mod is to document the information required to completely describe a test, provide common terminology for this information, and provide end-user applications with a common high-level data model.

The T&E Ref Mod is a reference model in that there is only one such model across all T&E centers and test squadrons. Such a model can be developed as an ontology model or as a Unified Modeling Language (UML) class diagram [7]. The reference model must be detailed out as much as is necessary to map to site-specific (or squadron-specific) metadata sets.

The T&E Metadata Reference Model is not a new data model or language or the integration of different T&E metadata models into a giant T&E metadata model. Rather it serves as a reference metadata model against which range-specific or test squadron-specific or test project specific metadata artifacts can be assembled, verified, validated and managed.

The prototype T&E Metadata Reference Model is shown in Figure 1 using a UML class diagram. The scope of this model logically includes a test program and its related entities, test and measurement metadata and their related entities, and test article metadata and its related entities.

The T&E Metadata Reference Model is critical to T&E metadata management because it serves as the starting point for a TEMPL practitioner to map out the AS-IS metadata, data sources and processes of the site or squadron under consideration.

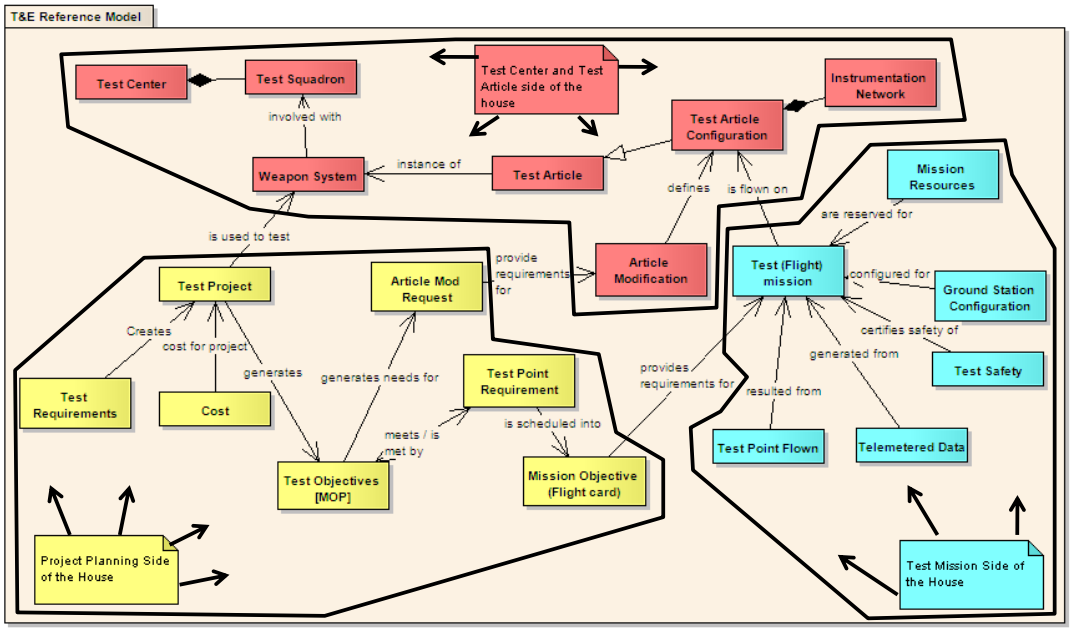

Figure 1: T&E Metadata Reference Model

MODELS AND MAPPINGS: THE TEMPL METHOD

A key aspect of the TEMPL methodology is the use of three types of models: (i) the T&E Ref Mod; (ii) test center data models; and (iii) user process or activity models. As part of the TEMPL method, mappings must be created among these models to support the various TEMPL functions. This mapping process creates associations between the unique naming conventions (*nomenclatures*) at a particular site and the neutral terminology of the T&E Ref Mod. Additionally, mappings must be created between the abstract data model and the data itself.

The T&E Ref Mod, discussed in the previous section, is a neutral abstraction of all of the metadata involved in the T&E world. Its purpose in the TEMPL method is to serve as the common view of the data used by the various end-users and end-user applications.

The test-center data models serve to document all of the data systems and artifacts used to store and maintain metadata at a specific test center. These data models must be mapped to the T&E Ref Mod to enable the TEMPL middleware (discussed in the next section) to access the appropriate metadata. The mappings between data model elements and T&E Ref Mod nodes are many-to-many. That is, each type of metadata may be stored in multiple artifacts or systems, and each artifact or system may contain multiple types of metadata. Once these mappings have been created, the TEMPL middleware can access the various systems and can provide instances of metadata to the various end-user applications without having to re-configure such applications to work with each test center's systems.

The user process or activity models serve to document the functions performed by a specific user or role at the test center. The TEMPL portal and other end-user applications can use these process or activity models to provide a relevant view of the metadata to each user. For the portal to provide this customized view, the user process models must be mapped to both the T&E Ref Mod and the data models. Like the other mappings, these mappings are also many-to-many (one process may involve multiple types of metadata or multiple artifacts, and one artifact or type of metadata may be used in multiple processes). Once these mappings have been created, the end-user applications can organize the metadata in a repository according to the user's process model, and can scope or filter the metadata according to a set of processes or activities that the user performs.

The TEMPL application method also includes a detailed approach for creating the models and the mappings among the various model elements. For data models, the modeler will design a model of data sources, and the data concepts within those sources. For each data concept, two types of attributes will be added to the model element to establish the mappings:

1. 0,1 or more T&E Ref Mod element name(s): These attributes define the T&E Ref Mod element(s) which correspond to the data concept
2. Data linkage information: This attribute contains the necessary information to connect to the data source and retrieve data associated with the data concept. For example, this information could include a URL to a folder containing documents, or a connection string and SQL statement to retrieve data from a database.

The relationships among the various mappings are illustrated in Figure 2 The solid arrows show a run-time relationship—the light blue arrow depicts a support towards development of a

Metadata Repository instance, the green solid arrow provide search filter support for end-user applications, and finally, the solid red arrows show the source input and output for validation support. A logical visualization of a simple example data model mapping involving just one T&E Ref Mod element ("Test Points") is shown in Figure 3.

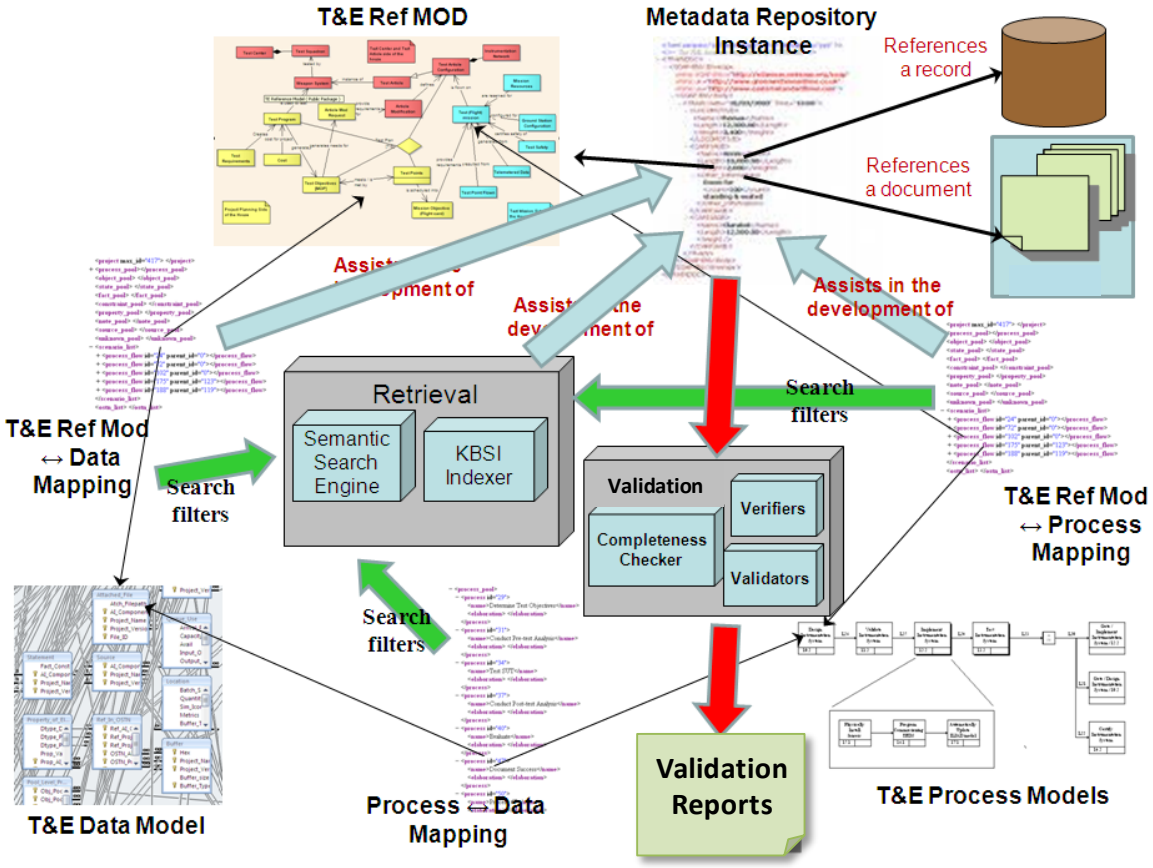

Figure 2: Models and their Relationships in the TEMPL Method

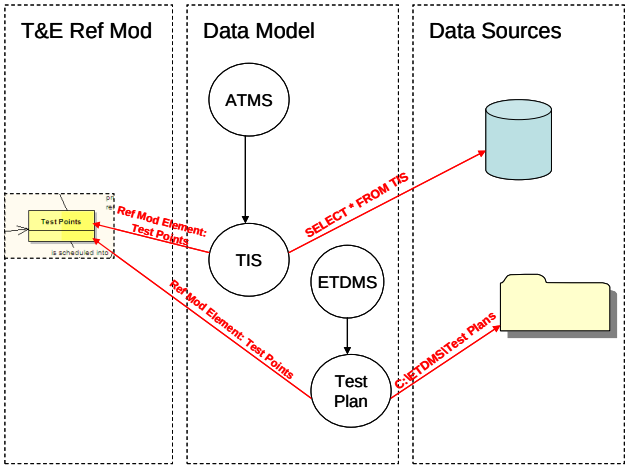

Figure 3: Example of Data Model Mappings

For process models, the modeler will design a standard Integrated DEFinition method (IDEF3) model of the process flow [8]. For each task in the model, two types of attributes can be added to establish the necessary mappings:

1. Zero or more T&E Ref Mod element name(s): These attributes define the T&E Ref Mod element(s) which correspond to the types of metadata being accessed or modified during execution of the task.
2. Zero or more data model element name(s): These attributes define the site-specific data concepts being accessed or modified during execution of the task.

It is necessary to allow mappings between the process model and both the T&E Ref Mod and the data model(s). The usage of one type of mapping over the other will depend mostly on the level of detail of the process model. For instance, for a higher-level, more-general process model, it would be better to define the process model in terms of the abstract metadata types defined in the T&E Ref Mod. However, for a more detailed process model that describes a particular user's process flow, it would be better to define the process model in terms of the actual data sources the user accesses.

Once these site-specific models are developed and mapped to the T&E Ref Mod, the TEMPL Middleware can use this information to provide end-user applications access to the metadata in a site-neutral way, while still allowing filtering and searching based on site-specific model elements.

TEMPL ARCHITECTURE

The high-level logical architecture of TEMPL, supporting the TEMPL methodology and end-user applications, is shown in Figure 4.

The architecture can be divided into four layers – user interfaces; information management; middleware; and data and metadata storage, or *persistence*.

The user interfaces layer includes KBSI's ModelMosaic® enterprise modeling tool (which includes ontology and process modeling, as well as supporting mapping relationships between them) and the T&E metadata portal and search tools. The portal represents the single interface through which end-users can issue searches, build repositories, and perform verification, validation, and completeness checking against all available metadata.

The information management layer consists of the "business logic", including the Verification, Validation, and Completeness checking logic; indexing, search and retrieval engine; and automation logic in ModelMosaic®. Included in the information management layer are various verification, validation and completeness checkers that test various T&E metadata rules. These VV&C checkers support various use cases that perform model and metadata verification, validation, and completeness tests and auto-generate models and artifacts. These components are intended to reduce the cognitive workload of the various stakeholders and users of a test program. The verification, validation, and completeness checking capability is discussed in more detail in [9].

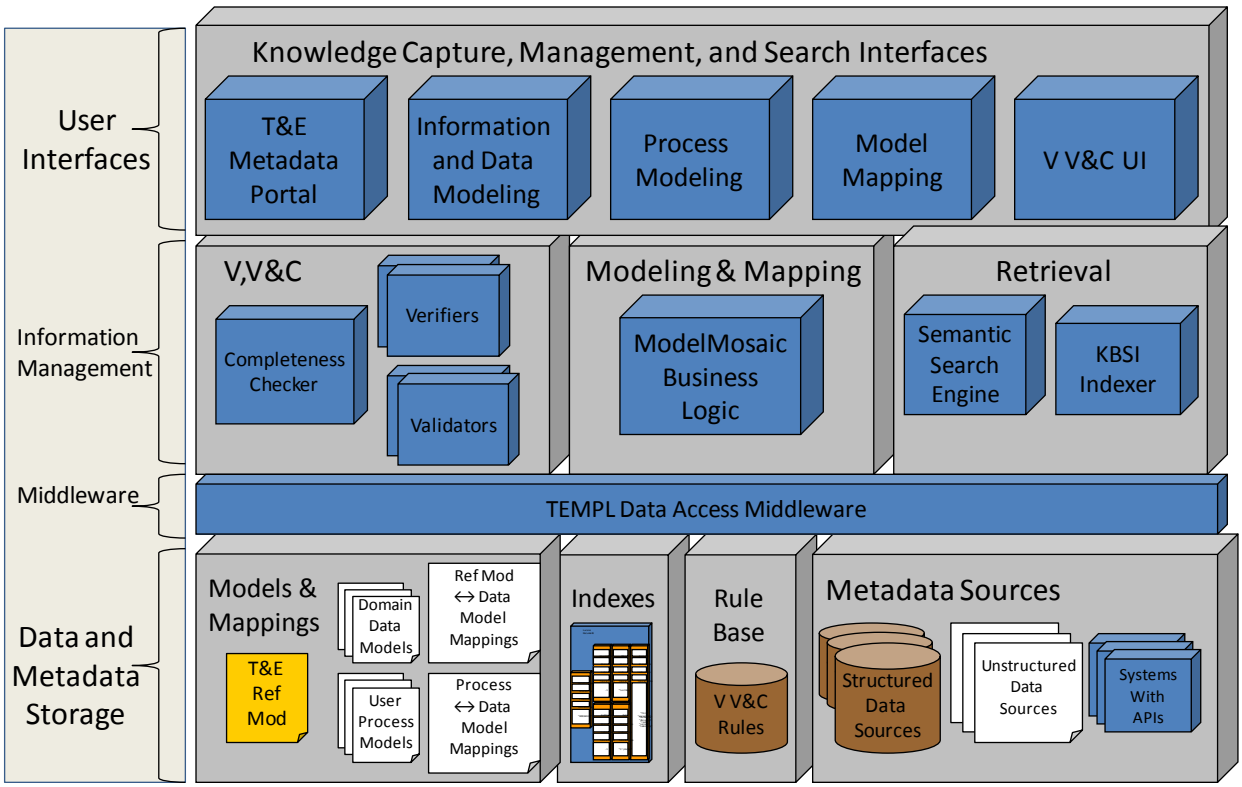

Figure 4: High-Level TEMPL Logical Architecture

The data access middleware layer uses the various models, mappings, database connections, and application program interfaces (APIs) to provide metadata retrieval and navigation services to the information management layer. This engine provides support for mixed-mode (structured and unstructured) data access. The middleware is "configured" by the models and mappings discussed in later sections. Additionally, by providing a common view into the diverse sources of metadata in the persistence layer, this model-based middleware can be used to support end-user applications not explicitly defined in the TEMPL architecture.

Finally, the user persistence layer consists of the various metadata, ontology and process models, and various mappings, index files for search, rule bases, and sources of data and metadata.

CONCLUSION

The complexity and diversity of metadata storage and management practices across as well as within ranges, coupled with the absence of a common terminology for T&E metadata has made it difficult to perform simple data management tasks such as retrieval, storage, and test validation across all of the information relevant to an individual test or set of tests. Through a SBIR project with Edwards Air Force Flight Test Center and KBSI, we have developed the Test and Evaluation Metadata Plaza (TEMPL). The TEMPL methodology defines the way in which the T&E Metadata Reference Model (T&E Ref Mod), when mapped to a site's specific data and process models, can support an architecture for implementing T&E metadata management that

enables repository building, retrieval, and VV&C checking of all information relevant to the T&E process through a single interface.

ACKNOWLEDGEMENTS

The ideas and results described in this paper have been developed as part of a Phase I Small Business Innovative Research (SBIR) project funded by Edwards Air Force Base AFFTC.

REFERENCES

[1] Data General, Instrumentation, Loading, Integration, Analysis, and Decommutation (ILIAD) Toolset tutorial.

[2] Telemetry Group, Range Commanders Council, IRIG Standard 106-01, Telemetry Standards, Secretariat, Range Commanders Council, U.S. Army White Sands Missile Range, NM, (2001).

[3] Michael Graul, Ronald Fernandes, John L. Hamilton, and Charles H. Jones, "An Enhanced Data Display Markup Language," 42nd International Telemetering Conference, San Diego, CA, 2006.

[4] John Hamilton, Ronald Fernandes, Paul Koola, and Charles H. Jones, "An Instrumentation Hardware Abstraction Language," 42nd International Telemetering Conference, San Diego, CA. 2006.

[5] Jeremy Price and Michael Moore, "XML Data Modeling for Network-Based Telemetry Systems," 44th International Telemetering Conference, San Diego, CA. 2008.

[6] Gene Hudgins, "The Test and Training Enabling Architecture, TENA, Enabling Technology For The Joint Mission Environment Test Capability (JMETC) and Other Emerging Range Systems", 44th International Telemetering Conference, San Diego, CA. 2008.

[7] OMG Unified Modeling Language™ (OMG UML) Superstructure, Version 2.2, Feb 2009, http://www.omg.org/docs/formal/09-02-02.pdf

[8] Richard J. Mayer et al. (1993) Information Integration for Concurrent Engineering (IICE): IDEF3 Process Description Capture Method Report. Logistics Research Division, Wright-Patterson AFB, OH 45433

[9] Darr, Fernandes, et. al., "Semantic Web Technologies for T&E Metadata Verification and Validation", 45th International Telemetering Conference, Las Vegas, NV. 2009.