



Whitemarsh
Information Systems Corporation

Tag and Post
vs
Data Standardization
Approaches to Achieve Net-Centricity

Whitemarsh Information Systems Corporation
2008 Althea Lane
Bowie, Maryland 20716
Tele: 301-249-1142
Email: Whitemarsh@wiscorp.com
Web: www.wiscorp.com

Table of Contents

1.0	Background	1
2.0	Tag and Post Approach	3
2.1	XML as the Basis for Data Interoperability	6
2.2	Discovery Metadata for Data Assets	7
2.3	The DoD Metadata Registry	7
2.4	Courses of Action Alternatives	7
3.0	Data Standardization Approach	9
3.1	DISA's Failed Attempt at Data Standardization	9
3.2	Smart, Well Engineered Data Standardization Approach	10
3.3	Thoroughly Vetted and Validated	11
4.0	Comparison of Tag and Post vs Data Standardization Approaches	13
4.1	XML as the Basis for Data Interoperability	13
4.2	Discovery Metadata for Data Assets	16
4.3	The DoD Metadata Registry	16
4.4	Courses of Action Alternatives	17
5.0	Tag and Post Approach vs Data Standardization Approach Decision Matrix	18
	Appendix 1 Overheads from the MITRE Presentation (6/30/2004)	22



Acknowledgments

This material is an evolution of documents that were updated during the time frame: September 2003 through December 2004. The primary contributors were Bruce Haberkamp, James Blalock, and Michael Gorman of the Office of the CIO, United States Army. The foundational components of this work has been favorably reviewed by subject matter experts within the U.S. Department of Defense.

1.0 Background

Two approaches are coalescing to achieve Net-Centricity: Tag and Post Approach, and the Data Standardization Approach. The two approaches are characterized as follows:

The Tag and Post Approach requires that the owner of a data asset accomplish its conformance to Net-Centricity through only two activities:

- Tagging its data assets with discovery metadata tags, and
- Creating a single XML schema based information exchange requirement (IER) for that data asset (or, as a variant, the data asset of a community of interest).

Thereafter, the owner of the data asset merely posts the data asset discovery tags to the DoD Metadata Catalog component of the DoD Metadata Registry, and the XML schema to the XML Registry component of the DoD Metadata Registry to then achieve Net-Centricity.

In contrast, the Data Standardization Approach accomplishes data standardization either within a program, if no community of interest exists, or within a community of interest, to the maximum extent practical. If higher levels of communities (Service, Joint, or Federal) exist, owners harmonize their local data standardization results with these higher levels of data standardization. Next, the program or the community of interest:

- Identifies each of its data exchange transactions
- Associates discovery metadata tags, through a completely automated means with these data exchange transactions,
- Creates XML schemas through completely automated means for each of these data exchange transactions

Finally, the owner of the data asset posts, for each data exchange transaction, the data asset discovery-tags to the DoD Metadata Catalog, and the XML schema to the DoD Metadata Registry to then achieve Net-Centricity.

While outside the scope of this paper, it is recommended that the DoD NII fully integrate all the various catalogs, registries, and other products within the DoD metadata registry so that the DoD Metadata repository is then able to fully support the DoD Net Centric Data Goals.



Also outside the scope of this paper is that the Tag and Post alternative serves only XML-based information exchange environments while the Data Standardization Approach serves as the critical foundation for XML exchanges, other types of data exchanges, and supports metadata standardization across all data assets (e.g, training, doctrine, and legacy systems) regardless of whether they participate in data exchanges.

Factors involved in the comparison of two approaches include, at the very least, the DoD Net-Centric Data Goals (See Table 1), and also:

- Cost Long Term
- Cost Short Term
- Portability
- Return on Investment
- Risk Avoidance
- Scalability

These additional six items, described in Table 2, serve as additional criteria to distinguish the Tag and Post Approach from the Data Standardization Approach.

A comparison between the two approaches is then provided in the Decision Matrix contained in Table 3 (Tag and Post Approach vs Data Standardization Approach). The ratings for each of these categories is given as Pass or Fail.

The remainder of this paper contains a description of the Tag and Post Approach (Section 2) followed by a description of the Data Standardization Approach (Section 3). These two approaches are then compared on their four common elements: XML as the basis for data interoperability, Discovery metadata for data assets, the DoD Metadata Registry, and Courses of Action Alternatives (Section 4). The paper concludes with Section 5 that consists of a Decision Matrix comparison between the two approaches supported by an explanation for a “pass” or a “fail.”



2.0 Tag and Post Approach

The Tag and Post Approach is presented in a 30 slide MITRE briefing of June 30, 2004. The title of the briefing is: Net-Centric Overview. There are no additional supporting materials. The key points to the briefing are contained on only six slides. These six slides are at the end of this paper (see Appendix 1). They are referred to by their Power Point slide number: 1, 2, 3, 4, 5 and 6. The other slides have nothing to do with data management. They merely address "envelopes" for data, and the exchange of envelopes, not the content of the envelopes, that is, the data.

Slide 1 identifies three data interoperability alternatives: Common Database Elements, Point to Point Interfaces, and Network Centric.

The first alternative, Common Database Elements, has been know for the past 10 years or more to be an unacceptable alternative. At the center of the Common Database Elements approach is DDDS (DoD Data Dictionary System). DISA tried the DDDS approach, and, while DISA's intentions and program goals were excellent, the DDDS's engineering was flawed from the very start.

The second alternative, Point to Point Interfaces, is also identified as unacceptable. This alternative has also been known to be unacceptable for the past 10 or more years. Most American Corporations have been moving away from Point to Point Interfaces since the late 1980s.

Table 1: DoD Net-Centric Data Goals	
Net-Centric Data Goal	Addressed within Data Management by
Make data visible	Identifying data assets within natural contexts of mission, organization and function. Standardizing taxonomies, ontologies and classification schemes that then enable viewing data semantics including where and how implemented through data assets.
Make data accessible	Including a Discovery Metadata definition in every data asset. Standardize names, definitions and structures first through templates and then incorporate these standardization artifacts, where practical, within data assets or at least map these standardization artifacts to data assets.
Institutionalize data management	Standardizing strategies for data definition. Creating multiple layers to ensure define once use many times. Using ISO standards 11179 (part 3) for enterprise data elements and SQL for data models.



Table 1: DoD Net-Centric Data Goals	
Net-Centric Data Goal	Addressed within Data Management by
Enable data to be understood	Creating standard vocabularies, commonly inherited semantics, commonly used data model templates, automatic names and definitions based on well defined words. Automatically creating abbreviations where necessary.
Enable data to be trusted	Creating consistent semantics, standard reference tables within enterprise data elements, and employing those to govern/map data models across multiple levels of abstraction.
Support data interoperability	Standardizing data structures. Establishing well engineered data transactions, automatic XML wrapping of data, supported by accessible data definitions and contexts.
Be responsive to user needs	Supporting reuse of already defined data assets metadata, central knowledge of all data assets and distributed access to same.

Table 2: Additional Criteria to Distinguish the Tag and Post Approach and the Data Standardization Approach	
Criteria	Description
Cost Long Term	The overall cost effect on the complete family of Army systems and on the ability of new systems to be able to create interoperable data exchanges.
Cost Short Term	The cost to an existing program that may be underway, or to a legacy system that is in maintenance.
Portability	The ability to develop an infrastructure and to enable it to be employed in multiple communities of interest with relatively the same result. In short, reliability and repeatability.
Return on Investment	The ratio of value received from the costs extended.
Risk Avoidance	The ability to inject a new system, COTS, or maintenance of an existing system into the Army's infrastructure without unpredicted consequences.



Table 2: Additional Criteria to Distinguish the Tag and Post Approach and the Data Standardization Approach	
Criteria	Description
Scalability	The ability to roll out policies and procedures to the entire Army with either a linear or lesser increase in cost.

Examples of companies that have forsaken the point-to-point approach are USAA and the Mars Corporation. The Federal Enterprise Architecture materials have moved away from Point to Point in several of their guidance documents. For example, in the document, E-Government Strategy (www.cio.gov/documents/2003egov_strat.pdf), agencies are advised to make sharable data available in portals that are then easy to access.

The Tag and Post Approach is then left with its preferred alternative, Net Centric. This alternative, depicted in the remaining slides (2, 3, 4, 5, and 6), centers on four topics. These are:

- XML as the basis for data interoperability
- Discovery metadata for data assets
- The DoD Metadata Registry
- Courses of Action Alternatives



2.1 XML as the Basis for Data Interoperability

In the Network Centric approach, the centerpiece is Published XML Schemas. This approach is then characterized as follows:

- Supports COTS
- Supports new users and systems
- Only one translator per system
- System developers not constrained
- One “IER” (information exchange requirement) for each COI.

In the center of the Tag and Post Approach is a “cloud” that contains Published XML Schemas. It is well known in the data management industry (both DBMS vendors and sophisticated data management industry users), that an XML schema is just an alternative form for the data structure specifications of a transaction. Thus, within DoD, there are likely to be from hundreds of thousands to possibly millions of XML schemas. Therefore, to say that each system is reduced to just one IER is an unsupportable conclusion. In all of SAP logistics, or PeopleSoft HR, is there one IER for each?

The key issue then is how can either a user and/or a system discover an appropriate XML schema, and then know that the selected XML schema will, in turn, enable access to data that conforms to the DoD Net-Centric Goals? This discovery process is presented on slide 2. On this slide, there is an intersection among all the systems called CoT, that is, Cursor On Target. This is a stylized term for an XML schema that is to represent a single data interface for all the systems, i.e., TCT-F, TACP, TADIL, DCGS, TBMCS, and FMSS. For such intersections to exist semantic agreement or understanding must exist for:

- Units of Measure,
- Formats,
- Reference Systems, and
- Naming Conventions

There is no explanation of the processes, policy, infrastructure, or whatever that will cause the semantic agreement or understanding to happen. Presumably, there will be some processor that will deduce and know how XML schemas are related one to the other. Presumably, some unknown infrastructure and/or process will know that the same names with different meanings are different, and/or that different names with the same meanings are the same. Presumably, there is sufficient precision, scale, and transformation processes necessary to transform data of one type to data of another type without loss of meaning or precision.



2.2 Discovery Metadata for Data Assets

Slide 3 is the next relevant slide to the Tag and Post Approach. It presents material related to the discovery of data assets. The slide does not hint or address how discovery metadata is created, managed, or disambiguated. There is no supporting strategy and/or materials that would ensure that multiple classifiers of the same data asset arrive at the same discovery metadata, or that classifiers of truly different data assets would arrive at truly different discovery metadata. Google searches that return hundreds of thousands or more hits are essentially useless.

2.3 The DoD Metadata Registry

Slide 4 is the next relevant slide in the Tag and Post Approach. It depicts the DoD Metadata Registry. The Tag and Post Approach material on this slide does not identify any of the critical DoD Metadata Registry architecture and integration problems that have been previously identified, described to, and assented by the DoD NII. Without all the metadata integration problems being identified and resolved, the DoD Metadata Registry is no more than a registry space for the hundreds of thousands to possibly millions of entries in the DoD XML Registry and/or the XML catalog.

2.4 Courses of Action Alternatives

The next relevant slide, slide 5, Alternative Course of Action–Phase 1, is essentially the same as was presented in a MITRE March 2004 presentation to the data management staff of the CIO-G6. This slide contained critical data management misunderstandings which were identified and thoroughly discussed. In a follow up meeting several weeks later, upon seeing the Army's Net-Centric approach, the MITRE lead pronounced the Army's approach superior to the one contained in the MITRE slides. The MITRE lead even requested that the data management staff of the CIO-G6 present to MITRE, in detail, the Army's approach so that MITRE could adjust its materials.

The current version of the slide shows that none of the required corrections were made. For example, the first bullet point, separate applications from data, has been accomplished in all sophisticated data management organizations since the late 1970s. When pushed, the MITRE presenter further asserted that all process should be stripped from data. Thus, all dates should only exist as integer numbers, not presented in some format like: mm/dd/yy. Similarly, there should never be any embedded process that would compute total weight, total cost, nautical distance, and the like. All these processes should be reserved for application programs. Thus, if 1,000 application programs all use "dates," then each MUST have a date conversion process custom programmed.

In addition to this class of issues, Slide 5 calls for the posting of data models to the DoD Metadata Registry. To what end? Data standardization? SQL? This conflicts with the assertion



(via alternative 1 of Slide 1), that data standardization is already useless. Further, suppose that the Army, Navy, Air Force, and Marines all post their data models; that is likely to be 500,000 data models. What purpose will they serve? How will they be interrelated with anything else in the DoD Metadata Registry?

The final slide, 6, Alternative Course of Action–Phase 2, suggests strategies that have nothing to do with any of the real Net-Centric Data Goals. Rather, this slide depicts that metadata should be posted to catalogs. Missing from both Slide 5 and Slide 6 is any requirement for or any basis to support the ability for XML schemas to be able to interoperate. Each XML schema will thus become a Tower of Babel. And in the DoD Metadata Registry there will be hundreds of thousands of such towers. How all this will be interrelated, integrated, disambiguated and managed is not addressed in any way in this presentation.

The Tag and Post Approach is thus summarized by two very simple strategies:

- Tag data asset metadata and post it to the Metadata Catalog of the DoD Metadata Registry, and
- Create one XML schema for each system and post that to the XML Registry of the DoD Metadata Registry.



3.0 Data Standardization Approach

The Data Standardization Approach has been set into policy, AR 25-1, and described further in the draft Chapter 5 of DA PAM 25-1-1. The Data Standardization Approach is squarely based on industry best practice, and has been either mandated or called-for by the Federal Enterprise Architecture Framework, the Federal CIO Council, the GAO, and the OMB.

This policy-based approach to achieving data interoperability through data standardization consists of four pillars:

- Enterprise identifiers
- Authoritative data sources
- Information exchange standards specifications
- XML data environment

The ultimate goal and test for the Data Standardization Approach is interoperability. While the demand for interoperability is easy to declare, its achievement is difficult. There are actually no unsolved technical problems. Interoperability consists of two parts: shared value streams, and shared understanding. Both of these are created from within the Communities of Interest and are expressed via the information exchange standards specification. The role of enterprise identifiers (EIDs) within data interoperability is to support technology independent mechanisms to understand both metadata and values (both single value and value sets). The role of authoritative data sources (ADS) is to minimize the versions of the truth. Additionally, it enables the coordinated migration of "truth" from an originating value state through a chain of value states until the data source is either quiesced or deleted. Finally, the role of XML is to take the value streams from an originating system and to transport them to an IEES or vice versa. Embedded within the XML stream are the EID tags that enable users to both understand the authority of the value sets and the supporting metadata.

3.1 DISA's Failed Attempt at Data Standardization

From a "rigid point of view," DISA failed in its attempt to design and then impose the "one right data model" onto everybody. To adhere to the DISA (a.k.a. Defense Data Dictionary System (DDDS) and Defense Data Architecture (DDA)) approach you had to adopt the DISA data model without one micron of exception. The DISA approach never has worked and never will work. Not only did DISA want everybody to adopt their data models, they also mandated that every column in every table of every database had to map to a DDDS data element. This approach was known to be wrong by 1994.

DISA, in early 2002, in response to push back, then adopted the 180 degree opposite approach. That is, the "do whatever you want" approach. To then have data interoperability in the face of the "DISA do whatever" approach, DISA, in conjunction with the DoD NII, proposed XML as the silver bullet. Somehow, it was felt that if you post the XML from all the transactions for all



the "DISA do whatever" database schemas, then magic would happen and all these disparate legacy system-based tags would magically get resolved. An alternative to this form of magic is to standardize all the tags across all the legacy systems. If there were only hundreds of systems, this would merely be difficult. However, since there are hundreds of thousands of systems, this effort is both unrealistic and impossible for the very same reasons that caused the DISA DDDS and DDA efforts to fail. In fact, any massive tag standardization effort is just poorly engineered data standardization by another name.

3.2 Smart, Well Engineered Data Standardization Approach

So, what does work? It's rather simple. It's the approach that forms the architecture the Army data standardization effort has espoused for the past five years. The approach consists of four pillars: IESS for shared data semantics, XML for data transport, authoritative data sources, and enterprise identifiers. These four components are mandated by AR 25-1, paragraphs 4-7 through 4-12.

The IESS, information exchange standards specification, is a logical data model that represents the shared data from the various physical data models of legacy systems from members of a community of interest (COI). The COI's end product is not only the IESS, it is also the mapping between the IESS's logical data model and the legacy system physical data models.

To have consistent semantics across all the IESS logical data models, there needs to be two additional data model layers: enterprise data elements, and shared data structure templates (i.e., conceptual data models). The "enterprise data elements" become fact-based, semantic templates for all the columns in the tables of the logical data model. These enable COI logical data models to be interrelated. Where do these enterprise data elements come from? They come mainly from the DDDS by discovering those data elements that are truly unique. For example, we only need ONE supply condition code, not 27. We suspect that there are only 6,000 enterprise data elements in all of DoD. Are these additional models more work? For a single project, yes they are. But, for the enterprise, they are not. These interconnected models increase productivity, increase quality, decrease risk, and decrease cost. These interconnected models make data interoperability practical. Without these models, data interoperability is either prohibitively expensive or not possible.

The shared data structure templates (conceptual data models) facilitate the "manufacturing" of data models from well-engineered collections of commonly employed enterprise data elements (e.g, materiel requisition or disposition, facility location characteristics, and person biographic information). The shared data structure templates should be "mined" from the Defense Data Architecture. When all these data model layers are in place, the problems encountered by DISA are completely avoided. Additionally, these data model layers enable XML tag engineering that causes XML to be useful.



The Army's Net Centric Data Management program, as specified in October 2003, has these components, that is, IESS (enterprise data elements, conceptual, logical, and physical data models), XML, authoritative data sources, and enterprise identifiers. This multi-component approach has been vetted by and practiced within industry for years. It works, period.

Within the IESS component, enterprise data elements are based on the ISO standard 11179, Part 3, and the conceptual, logical, and physical data models are based on ISO/ANSI standard SQL. All these are mandated by OMB.

3.3 Thoroughly Vetted and Validated

Who else in the Federal Government agrees with the Army's Net-Centric Data Management approach? The enterprise data element component is being done by EPA, Census, FAA, and other Federal Agencies. The Navy functional data administrators and DLA also agree. When we presented our approach to various Air Force staff they agreed as well. At the present time, this approach is being adopted into the Node Guidance part of NESI (Net-Centric Enterprise Solutions for Interoperability (an effort of the Army, Navy, and Air Force)). Finally, this approach was presented approximately two years ago to an international data management conference. The over 500 persons in attendance agreed that the approach either mapped onto what they were doing or mapped onto industry best practice.

Now, where does XML fit? XML, to be truly effective, must have a set of tag names that enable XML schemas to be quickly and easily discovered. The Army's Net-Centric Data Management approach has a strategy whereby the XML tags can be automatically generated from the physical data models of the legacy systems. How then are these XML schemas quickly and easily discovered? If the physical data models are associated with logical data models, which are in turn associated with enterprise data elements, then a computer program would be able to determine all other semantically equivalent XML schemas even if their tag names are different. Without this IESS data model infrastructure, searching for semantically equivalent XML schemas is like looking for a snow flake in a blizzard.

The database management system vendor community, through ISO and ANSI standards, is nearing completion of the first version of SQL/XML facilities that will enable automatic composition of XML data streams from SQL data model based data, and automatic shredding of XML data streams into SQL data model based data. It is expected that almost simultaneous with de jure standardization of these facilities will be their release by the major DBMS vendors.

The other two components of the Army's Net-Centric Data Management Program for interoperability, authoritative data sources and enterprise identifiers, are indisputably required and are presented in other materials.

In total then, the Army's Net-Centric Data Management Program's data interoperability requires all four pillars: XML for data transport, IESS for shared data semantics (which embraces the



enterprise data elements, conceptual, logical and physical data models), authoritative data sources, and enterprise identifiers. Net-Centric environments that include XML as the mechanism of data transport can be achieved if and only if IESS, ADS, and EIDs are in place.

A critical question is how does the Data Standardization Approach “fix the ERP (Enterprise Resource Planning) problem?” Again, the answer is rather straightforward. Make an IESS for the ERP and “bolt” it onto the ERP system so that it contains, on either a real-time or daily basis, all the data that is to be shared with the Army communities. ERPs are then able to be folded into Net-Centric environments because they will then have IESSs, ADS, EIDs and XML. This solution not only enables ERP systems to interoperate (indirectly, of course), it also enables enterprises to have different ERPs because they are now encapsulated. ERP Interoperability is achieved through the ERP-IESSs.

There is a significant quantity of in-depth materials that present the Data Standardization Approach’s overall engineering and construct. This approach, supported by a detailed framework, has been cross referenced to the DoDAF. It has been demonstrated conclusively to work. The Data Standardization Approach has been presented to meetings, and favorably reviewed by the Army, Navy, Air Force, Marine Corps, and DLA. Additionally, the essential engineering constructs of the Army’s Net-Centric approach has been presented to several international conferences that were well attended by both data management and also information systems development experts. Corrections have been received and materials have been corrected over a five year period.

In summary, the Army's Net-Centric Data Management Program's approach to data standardization and interoperability absolutely replaces the "old" DISA data standardization approach with one that not only squarely fits with industry best practices but also enables XML to fulfill its role as an interoperable data transport mechanism.



4.0 Comparison of Tag and Post vs Data Standardization Approaches

While it is unfair to compare the Data Standardization Approach with the Tag and Post Approach, because the Data Standardization Approach is significantly more expansive, it is instructive to compare the two approaches on the four main points of the Tag and Post Approach, which are:

- XML as the basis for data interoperability
- Discovery metadata for data assets
- The DoD Metadata Registry
- Courses of Action Alternatives

4.1 XML as the Basis for Data Interoperability

XML is not the sole basis for interoperability. In fact, without a smart, well-engineered data standardization infrastructure, accomplished through the Data Standardization Approach, XML cannot possibly succeed. The Tag and Post Approach does not recognize the need to have any strategy for harmonizing tags within a COI, across COIs, or Services. Without pre-existing sets of standardized tags, the only way to harmonize Tag and Post XML schemas is to perform, after the fact, tag-standardization. Because this would be done after the fact, the quantity of XML schemas to be harmonized would be significantly larger than the quantity of tags associated with logical data models. That is because while there is only one logical data model per database, there would likely be many XML Schemas; hence a significantly larger tag-standardization effort. Additionally, if the Data Standardization Approach had the tags associated with conceptual data models or enterprise data elements, the tag standardization effort would be even smaller than the Tag and Post Approach.

XML, to be truly effective, must have a set of tag names that enable XML schemas to be quickly and easily discovered. In contrast with the Tag and Post Approach, the Data Standardization Approach has a strategy whereby the XML tags can be automatically generated from the physical data models of legacy systems. How then are these XML schemas to be quickly and easily discovered to be of interest to an individual? If the physical data models are associated with logical data models, which are in turn associated with enterprise data elements, then when an XML tag set from a XML schema is submitted, a computer program would be able to determine all other XML schemas regardless of their tag names that are semantically equivalent. Solved would be the problem of not knowing that the same names with different meanings are different, and/or that different names with the same meanings are not the same.

The Data Standardization Approach has included lessons learned from the failures associated with the DISA DDDS (DoD Data Dictionary System) approach, and has devised an approach to achieve the data standardization required to meet the requirements of the DoD Net-Centric Data Goals. Standardizing or coming to a common understanding for the meaning, expression, and exchange of critical business facts is an essential component to any data interoperability.



Business facts do not exist in isolation. Rather they exist in well defined collections. The DoD Defense Data Architecture (DDA), like the DDDS, represents hundreds of millions of dollars of investment in the formulation of well defined collections of business facts, rules governing these collections, and the critical relationships that exist between business fact collections. The Data Standardization Approach has incorporated the lessons learned from the DDDS and DDA and has incorporated them—appropriately—into its strategy. To that end, the Data Standardization Approach consists of the creation and management of five levels:

- ISO 11179 data elements (also in this paper, enterprise data elements)
- Conceptual data models
- Logical data models
- Physical data models
- View data models

The ISO 11179 data elements (required by the Federal Enterprise Architecture Framework) act as enterprise level data element templates for use in understanding legacy system data elements. Understanding and mapping is essential because, without common understanding, the creation of meaningful XML schemas and/or discovery metadata is impossible. A key source for the enterprise data elements is the essential and non-redundant set of DDDS data elements.

Conceptual data models, developed from the DDA, are needed to successfully map to the collections of legacy system data elements. Each conceptual data model data element is mapped to a higher level enterprise data element. The conceptual data models are essential because they enable legacy system data models to be commonly understood. Again, understanding and mapping is essential because without common understanding, the creation of meaningful XML schemas and/or discovery metadata is impossible. The conceptual data models have the potential of being the automated source for all discovery metadata tags, thus reducing or eliminating the Tower of Babel.

Logical data models are either drawn from legacy system databases or represent the shared data understanding across one or more communities of interest. Members of a community of interest, to share data via XML or any other tagging format, must first come to a common understanding of their shared data. Logical data models, either created in a legacy system development environment or through communities of interest, are merely required to map their data element collections to conceptual data models and to map their logical data model data elements to the enterprise data elements. Only when these mappings are accomplished can XML schemas be shown to be different from each other even when the tag names are the same, or to be equivalent even when the tag names are different. Further, there is significant potential that such same/different analyses can be automated.

Physical data models are the data models that are directly employed by the database management systems that operate either the COTS or legacy databases. There can be many different physical data models for the same logical data model. Given the existence of both logical and physical



data models, these differences can be quickly understood and determined to be really different or equivalent.

The final model, the view model, is essentially equivalent to an XML schema model. Views are mapped to physical data models. In fact, views can be automatically generated from physical data models. So too can XML schemas be automatically generated from View models.

The Army's Net-Centric Data Management program, specified as of October 2003, has included the Data Standardization Approach because these data model components, that is, data elements, conceptual, logical, and physical data models, have been vetted by and practiced within industry for about 20 years. It works, period. The enterprise data elements component is based on the ISO standard 11179, Part 3, which is mandated by OMB and is being employed by the Bureau of Census, EPA, and the FAA. The conceptual, logical, and physical data model components are based on ISO/ANSI standard SQL, which is also mandated by OMB.

The Tag and Post Approach does not acknowledge the need for a smart, well-engineered data standardization infrastructure. That causes the Tag and Post Approach to stand apart from the collective wisdom of the data management system vendors, the data management community, and modern industries like Bell South, USAA, and Mars. In this regard, the GAO Report, United States General Accounting Office (GAO) Report to the Chairman, Committee on Governmental Affairs, U.S. Senate, Electronic Government: Challenges to Effective Adoption of the Extensible Markup Language, April 2002, GAO-02-327, contains three quotes that are especially instructive as they directly point to the need for enterprise architectures and data standardization prior to any XML benefits:

- XML's greatest benefits accrue when organizations, such as government agencies, use standard data exchange procedures and agree on standard data definitions and structures. Effectively using XML as a means to share data among disparate systems across the federal government will require agencies to conform to a range of technical and business standards.
- XML's larger promise of facilitating data exchange across broad domains (such as an entire agency, a group of agencies, or a set of external stakeholders and client organizations) will be difficult to realize until critical data elements and structures are identified and standardized across entire agencies and communities of interest.
- This task of identifying and standardizing critical data elements and structures is part of an agency's larger task of developing an enterprise architecture. Well-planned enterprise architectures can also promote the adoption of flexible implementations that can be modified in the future to conform to commercial standards that become established over time. Thus, agency enterprise architectures are key building blocks to effective government wide adoption of XML



4.2 Discovery Metadata for Data Assets

Discovery Metadata for data assets must be automatically generated as there are likely to be hundreds of thousands to millions of data assets. If the discovery metadata tagging process is manual, then it will grind to a halt even before 0.1% of the data assets are tagged. Just suppose there are only 500,000 data assets in all of DoD (hard to believe it's that few as the Navy alone has 100,000 systems), and it takes 15 minutes to manually create discovery metadata. That means that there will be a requirement for 125,000 staff hours, or 60 staff years, to tag the 500,000 data assets. Simply put, unless the data asset discovery metadata tags are automatically generated the effort will be too onerous.

The Tag and Post Approach offers no strategy, guidance, infrastructure engineering or the like to assist in this process. The Data Standardization Approach, in contrast, if implemented, enables the data asset discovery metadata to be automatically generated. Further, the tags employed will be based on the highest level of data asset commonality, that is, enterprise data elements, and/or conceptual data models. Further, the data asset discovery metadata will then be supplemented by an automatic access path into these data models, thus enabling users to quickly and easily understand all the essential data characteristics to then know whether a selected data asset is relevant or not.

4.3 The DoD Metadata Registry

The DoD Metadata Registry, by design and engineering, appears not to be integrated. In short, all the "sins of the past" associated with prior DISA efforts, that is, the DDDS and the DDA, will be repeated. The problems associated with the DoD Metadata Registry have been presented to groups from the Army, Air Force, Navy, Marines, DoD NII, and DLA. There has been 100% concurrence that the DoD Metadata Registry problems exist and, that until fixed, the DoD Metadata Registry will be of no real value.

The Data Standardization Approach recognized the lack of integration within the DoD Metadata Registry because it's comprehensive, completely integrated metadata model was compared with the DoD Metadata Registry meta model. That comparison additionally showed that the current DoD Metadata Registry is missing critical classes of metadata necessary for it to completely store, interrelate, and manage the data architecture products of the DoDAF.

The Tag and Post Approach offers no recognition of the DoD Metadata Registry's critical problems. Thus, XML schemas, created under the Tag and Post Approach, will be just loaded en masse into the DoD Metadata Registry without any of the critical supporting infrastructure. This will not enable the same XML names with different meanings to be seen as different, and/or different XML names with the same essential meaning to be seen as the same.



4.4 Courses of Action Alternatives

The Tag and Post Approach courses of action (Phase 1 or Phase 2) map onto either strategies that have already been in place for a large number of years by modern IT organizations and thus should already be accomplished by virtually every Army IT program, or strategies that are set squarely on foundations that are not well engineered and cannot hope to succeed.

In contrast, the Data Standardization Approach is set out within the Army's Net-Centric Data Management program in AR 25-1. It is further described in draft DA PAM 25-1-1. The Data Standardization Approach has already been favorably reviewed by the DoD NII, and groups within the Army, Air Force, Navy, and Marines. The Data Standardization Approach has also been favorably reviewed by key data management industry groups as fitting squarely within best practices and/or already being accomplished by successful corporations. The Data Standardization Approach, contained in AR 25-1, and supported extensively by methodology, software, courses, workshops, books, and white papers enables:

- The establishment of the data management infrastructure within the Army that is a smart, well engineered form of data standardization.
- The establishment of communities of interest for the development of Net-Centric conforming products, and the infrastructure to harmonize the work products across communities of interest
- The delivery to communities of interest of already proven policies, practices, and infrastructure so that they can concentrate on their real mission.
- The ability, at the level of the CIO-G6, to identify areas of commonality across communities of interest and to harmonize them quickly.
- The ability to have, at an appropriate level, productive working relationships with other DoD Services, Joint organizations, and other Federal Agencies.

These components have already been identified, described, engineered, and proven.



5.0 Tag and Post Approach vs Data Standardization Approach Decision Matrix

Table 3: Tag and Post (TPA) Approach vs Data Standardization Approach (DSA)			
Evaluative Area	Approach Comparative Evaluation		Difference
	TPA	DSA	
Make Data Visible	Fail	Pass	The Tag and Post Approach does not propose or support the infrastructure to make data visible within mission, organizations, and/or functions. The Tag and Post Approach does not support the completely interrelated metadata infrastructure to truly understand a data asset.
Make Data Accessible	Fail	Pass	The Tag and Post Approach does not enable the automated generation of data asset metadata tags necessary to make the “right” data pop to the surface for subsequent use. Not having an automated generation of these critical tags will inevitably cause different sets of tags for essentially the same data, which will greatly increase access confusion.
Institutionalize Data Management	Fail	Pass	The Tag and Post Approach does not recognize the need for or support the creation of a modern data management environment that is seen by Services, the Federal Government (OMB, CIO, and GAO), and industry as being essential for success.



Table 3: Tag and Post (TPA) Approach vs Data Standardization Approach (DSA)			
Evaluative Area	Approach Comparative Evaluation		Difference
	TPA	DSA	
Enable data to be understood	Fail	Pass	The Tag and Post Approach, because it fails to recognize the need for a smart, well-engineered data standardization infrastructure, cannot possibly enable data to be disambiguated or broadly understood. The Tag and Post Approach does not contain the ability to discover common use, disparate meanings that need to be made common, and the like. These are all essential to an organization that demands information superiority.
Enable data to be trusted	Fail	Pass	The Tag and Post Approach does not enable the infrastructure to be installed that enables consistent semantics, value domains, definitions, and contexts. Without these to govern and manage the vast inventories of data assets there will be an increase in authoritative data sources for the same data and no way to sort out this critical issue.
Support data Interoperability	Fail	Pass	The Tag and Post Approach does not recognize the need for or the ability to enable the same named data to be seen as different, and the differently named data to be seen as the same. This problem extends to data types, precision, scale, value domains, and value transformation processes that must be centrally defined and managed.



Table 3: Tag and Post (TPA) Approach vs Data Standardization Approach (DSA)			
Evaluative Area	Approach Comparative Evaluation		Difference
	TPA	DSA	
Be responsive to user needs	Fail	Pass	The Tag and Post Approach will not enable any of the critical re-use methodologies and techniques to be put into place such that a new data asset can be brought on-line, modified, and discarded well within necessary decision time-frames. In short, the Tag and Post Approach does not enable automation of IT assets critical to the creation and use of data assets.
Cost: Long Term	Fail	Pass	The Tag and Post Approach will greatly increase the overall cost and lack of interoperability within data asset intensive programs. That is because there will be no way the data asset metadata across the entire infrastructure can be recognized as the same. Redundancy greatly exceeding the duplication that occurred during the DDS and the DDA efforts of DISA will occur.
Cost: Short Term	Fail	Pass	The Tag and Post Approach, because of its lack of infrastructure engineering will cause significantly wasted resources within programs of even a small size or a short term. For example, because of the lack of an infrastructure, the C2IEDM IESS, already being adopted by a number of organizations, is starting to be changed in incompatible ways.



Table 3: Tag and Post (TPA) Approach vs Data Standardization Approach (DSA)			
Evaluative Area	Approach Comparative Evaluation		Difference
	TPA	DSA	
Portability	Fail	Pass	The Tag and Post Approach will not encourage portability. That is because there will be no common understanding of data asset metadata such that a data asset can be constructed from already standardized enterprise data elements and collections of enterprise data elements.
Return of Investment	Fail	Pass	The Tag and Post Approach will not enable a positive ROI because there will be no mechanism to have commonly defined data asset metadata. Thus, reuse will be close to impossible.
Risk Avoidance	Fail	Pass	The Tag and Post Approach will greatly increase risk. That is because without a data asset metadata infrastructure within COIs, across COIs, Services, Joint, and Federal Agencies, the effort required to do due diligence will exponentially increase.
Scalability	Fail	Pass	The Tag and Post Approach will certainly not scale. With every new data asset within a COI, across COIs, Services, Joint, and Federal Agency, every other data asset metadata will have to be carefully examined to determine similarity or difference. The entire process will slow further with every new data asset that needs to be included.



Appendix 1

Overheads from the MITRE Presentation (6/30/2004)

Metadata for Discovery Services

- Cross COI Discovery Services require contextual metadata containing rich semantics
 - Enable detailed, sound, meaningful distinctions to be made among the classes, properties, relationships, business rules, and environments
 - May be expressed in a logic-based language
- Semantically-rich, contextual metadata is not
 - More detailed definition
 - How data is related in a data model

Static, Manual Discovery

•Structural, Descriptive Metadata

•White, Yellow, Green pages

Data Element

Data Attribute

Keyword List

Data Schema

XML Schema

XML DTD

Taxonomy

MITRE

Dynamic, Run-time, Cross-COI Discovery

•Semantically-Rich Metadata

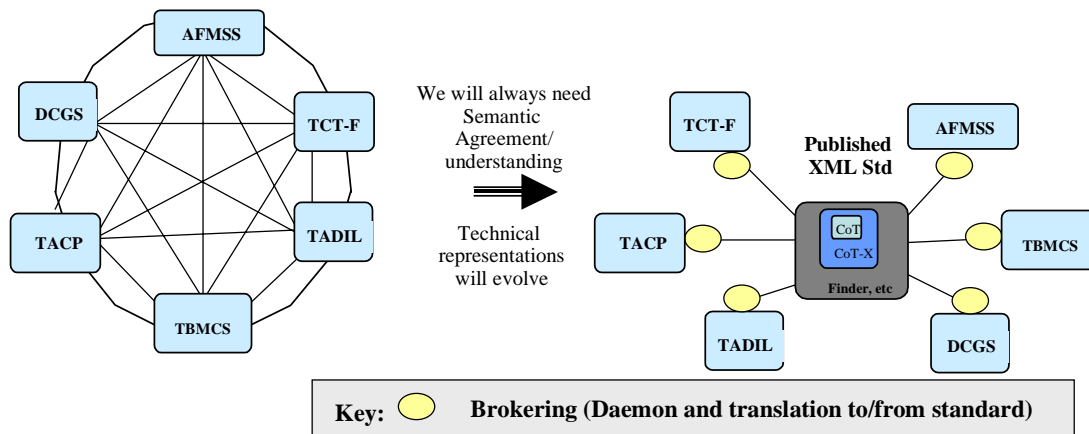
•Brown pages

- Description of the object in the context of its usage
 - A tank has one or more radios
 - A tank can transverse rugged terrain
 - A tank can not transverse vertical drops
 - A tank supports ground forces



To Achieve Data Interoperability (example)

Point to Point Interfaces become Network Centric



Each Point-to-Point interface with their

- Own semantics
- Units of Measure (e.g. decimal degrees **vs** degrees, minutes, seconds)
- Formats (e.g. latitude and longitude **vs** coord, N/S **vs** +/-hemispheric indicators)
- Reference Systems (e.g. MSL **vs** HAE)
- Naming Conventions (e.g. coord_deriv_acc **vs** cerr)

Each in semantic understanding, and mapping to the same

- Units of Measure
- Formats
- Reference Systems
- Naming Conventions





Alternative Course of Action --- Phase 2

- COIs develop extensive discovery metadata
 - Rich semantic content
 - May be logic-based and may use ontologies
 - Contextual (key attributes and concepts)
- Discovery metadata posted to COI Metadata Catalog
 - COIs establish databases for Metadata Catalogs
 - Implement Brown Pages within UDDI data structure
- Full Mediation-Based interfaces using cross-COI discovery services
 - Cross-COI ad hoc queries
 - Dynamic, run-time data discovery

**Tools and standards
maturing**

