Whitemarsh
Information Systems Corporation

Data Management Program: Reference Model

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

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.

This paper briefly explains the data management reference model that is shown in Figure 1. This model consists of the following:

- Three layers (Knowledge online, collaborative interfaces, and net centric applications)
- Three bands (Info-structure migration, data transport, and reference data)
- Four data standards (Authoritative data sources (ADS), Information exchange standards specifications (IESS), Enterprise Identifiers (EID), and XML.
- Three dimensionality arrows (Data to knowledge, data interoperability, and data community)

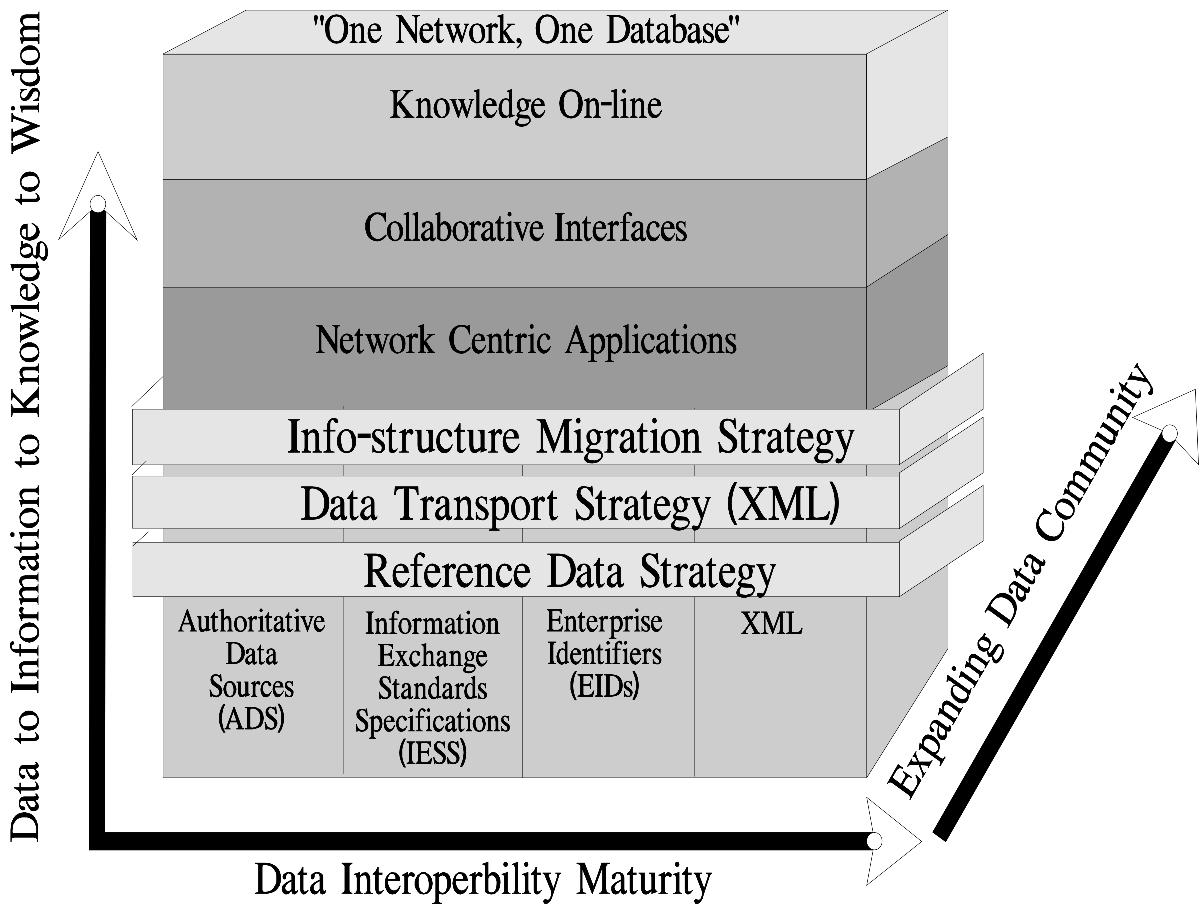

Figure 1. Data management program reference model

The top layer, knowledge online, of this data management program reference model has as its goal the creation of one *virtual* portal of knowledge that represents consistent, non-redundant, and semantically homogeneous data across one network and one database.. This is possible only if there is a high degree of integrated semantics across all the data that is captured, stored, and presented to users.

Integrated semantics can be achieved either through consensus based standardization and implementation through those standards, or through consensus based mapping from one set of data and semantics to another. The latter should be a transition step to the former.

Mapping is greatly eased if there are higher levels of consensus based standards. For example, suppose one legacy system has person gender with the values of M and F, and another legacy system has Sex with the values 1 and 2, if there were a higher level of abstraction with Person Gender with values Male and Female, and further, each legacy system was mapped to these values, then mapping between the two legacy systems could almost be made automatic.

Supporting the top layer is the collaborative interfaces layer that represents all the interfaces with various other systems environments. While it would be greatly eased if these systems were also based on integrated semantics, it is only required that these external semantics be mapped to existing semantics.

The next supporting layer, the environment of all net-centric applications, represents all the application information systems that are employed to capture, store, and make available the information required by the user. Ideally these applications all conform to Net Centric guidelines. That is, to the maximum extent possible they can read and write data via XML according to published XML Schemas and supporting data specifications, and these applications, to the maximum extent possible, form the basis of service oriented architectures that separate data and process from traditional bindings.

Here, it is not enough that all systems merely conform their read/write transactions into XML schemas, wrap their interchange transaction data into XML streams, and then post uniform resource locators (URLs) to where the data asset data transactions reside. If that tactic is taken, then, while connectivity (knowing and obtaining data from a data asset) can be achieved, understandability will be time consuming because the receiving system's environment staff will first have to find the appropriate XML schemas from what could be many millions, understand the XML schema, map the posting and receiving system's semantic understanding, and then construct semantic, value domain, and precision transformations prior to any meaningful data access. As stated at the outset, connectivity is not the measure of success. Understandability with minimum complexity and latency is. Employing XML without the prerequisite steps of smart and intelligent data standardization will only lead to Net Centric failure.