Whitemarsh
Information Systems Corporation

Data Management Program:
Knowledge Worker Categories and Questions

Table of Contents

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.

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
Business Rules	What is a business rule? Is it merely a data constraint? Or more? Compare and contrast.
	How is a business rule defined with respect to data?
	What are the different types/classes of business rules?
	Where are all the different places within a database that a business rule can be implemented?
	Under which conditions should a business rule be implemented within the database, the view, or the application information system?
	What metadata should support the definition of a business rule?
	Should there be a business rule registry? Should the business rules be managed centrally and then allocated where appropriate? If so, how? How can it be determined that business rules are properly defined within applications and databases? How can it be determined that they are properly executing? How can it be determined that different applications and databases are executing business rules in a non-conflicting and non-redundant manner?
	What is the relationship between the business rule data elements and the various classes of data elements, that is, ISO 11179 data elements, data structure template data elements, database table data elements, view data elements, and application data elements?
	Should business rules be centralized as to definition and distributed where bound? If yes then why? If not then why not?
Capture Context of Database Data	Why is it important to know the contextual business event that was the context for data capture? What purpose would it serve? How would you find out?
	Why is it important to know the contextual business cycle in which the capture occurred? What purpose would it serve? How would you find out?
	Why is it important to know the contextual business calendar in which the capture occurred? What purpose would it serve? How would you find out?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	Why is it important to know the contextual business information system in which the capture occurred? What purpose would it serve? How would you find out?
Communities of Interest (COIs)	What is a Community of Interest? Should they be restricted to organizational units? Functions, Colloborations?
	How should COIs be chartered? Funded? Staffed? Enabled with infrastructure? Time-limited?
	Should there be classes of COIs (e.g., institutional, functional, technical, ad hoc/expedient)? Compare and contrast. Should there be an overall governance structure for COIs?
	What level of authority should COIs have over their members? How should COI business be conducted? Decisions made?
	What product set should a COI produce? Where should these product sets be stored? How should these product sets be interrelated and be related to other COI product sets? How would conflicts be identified and resolved? How would products be developed? Maintained?
	Should COI products sets be mandatory for COI members? How enforced? How would conformance be determined?
	What is a community of interest with respect to data sharing?
	What are the key issues that have to be addressed?
	How would you address them. Provide a general plan.
	What are the key duties and responsibilities?
	How do communities of interest integrate vertically and horizontally?
	How are data issues resolved within and across communities of interest?
	What should the persistent products of a community of interest be?
	What should the COI infrastructure be? What is the general process model?
	How should products be developed? How maintained?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	How can or do you "sunset" a community of interest? What happens to its products?
Data and Process Separation	What does it mean to separate data specification and/or definition from process specification and/or definition?
	How would you determine how much separation is desirable?
	What data products should be separated from process?
	What process products should be separated from data products?
	What is the rationale for the separation?
	When should data products be included in process products?
	When should process products be included in data products?
	What should be the basis for choice? Benefits? Costs?
Data Driven versus Process Driven	What is the difference between process driven and data driven?
	Which is arguably better? Why? How can the better alternative be determined?
	How would you engineer a database design resulting from within a process driven approach? What would be the likely problems or inefficiencies?
	How would you engineer a process design resulting from within a data driven approach? What would be the likely problems or inefficiencies?
	How would you engineer a data architecture and the supporting database designs within a very large and very distributed enterprise-size systems development effort? What metrics would you use to estimate the cost and duration of such an effort? What would be the metrics that would be employed if the approach was data driven? Process driven?
Data Elements	Compare and contrast an ISO 11179 Data Element with a database table data element?
	Compare and contrast an ISO 11179 Data Element with an application Data Element?

3

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	What is the likely ratio between an ISO 11179 Data Element and the database table data elements that implement it? What are the factors that affect the ratio?
	What is the likely ratio between an ISO 11179 Data Element and the application data element that implement it? What are the factors that affect the ratio?
	What metadata is necessarily captured to fully define an ISO 11179 Data Element? How is it organized? How is this metadata interrelated?
	How would you map application data elements to database table data elements to ISO 11179 data elements?
	How would you build ISO 11179 data elements? Top-Down? Bottom-up? Why?
	How and where should ISO 11179 data elements and their metadata be stored? How should these be used?
	What is the relationship between XML elements/attributes and application data elements? Database table data elements? View data elements? ISO 11179 data elements? Why?
Data Management	What is a definition and scope of data management? What is the "business case" for data management? Who are the stakeholders? What is the purpose chain (i.e., ultimate objectives of the data owners down to the specific objectives of the data providers)? What is the value chain? (i.e., the chain of goals by which the enterprise purpose is measured, e.g., performance, cost, maintainability, scalability)? What metrics are used to quantify and evaluate these values?
	What major roles exist within data management (i.e. business sponsor, data administrators, database administrators, data stewards (the key players), Subject Matter Experts, Project Management Offices, etc.)? What is the type and extent of participation from business/functional staff? What products do they produce? What is the required metadata support? What is the necessary interaction between functional staff and technical staff? How do the functional and technical products interrelate?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	How can the total cost of data be determined? Total cost of data management? How can the total value of data and data management be determined? How can the ROI on each be determined?
	What is data management? What is data administration? What is database administration? What does each involve? What products do they produce? What responsibilities does each have? How are they similar? Different? Related to each other?
	Assuming that data management encompasses data administration and database administration, what role does data management have with Communities of Interest? Systems Developers?
	Should data management be centralized? Distributed?
	Should data management be responsible to develop/procure/deliver data management courses, workshops, seminars, techniques?
	What is the business case for data management?
	What is the cost for not having data management?
	If there could be multiple relationships between ISO 11179 data elements and application data elements, view data elements, data structure template data elements, and database table data elements, why would that be important and what would be the comparative benefit?
Data Management Assessments	What would constitute a data management assessment of a subordinate organization?
	What would be assessed?
	What would be your metrics?
	How do you assess current data management processes? How do you configure changes? How do you cost and posit benefits for a change? How do you track change management including its cost and benefits?
Data Management Planning	What is data management planning?
	What are the key issues in making data "perform?" What issues are not important?

| Data Management Knowledge Worker Categories and Questions ||
Item	Questions
	What is a general strategy?
Data Management Program Program Planning	What is the function of the program planning layer?
	How is this layer related to the project planning and execution layers?
	What are the items within this layer? Are any missing? What are they if missing. Should any be deleted? Which and why?
	Identify which are the most important?
	Briefly describe what should be in each of the most important components of the program planning layer?
	How would you propose to accomplish the most important components in the program planning layer?
Data Management Program: Project Planning	What is the function of the project planning layer?
	How is this layer related to the program planning and execution layers?
	What are the items within this layer? Are any missing? What are they if missing. Should any be deleted? Which and why?
	Identify which are the most important?
	Briefly describe what should be in each of the most important components of the project planning layer?
	How would you propose to accomplish the most important components in the project planning layer?
	How are artifacts of the various projects able to be interrelated? Integrated? Made non-redundant?
	How can or should projects be planned? Centrally, decentralized? How can one project benefit from another? How can artifacts be shared?
	How should projects be organized into various communities of interest? What would be the staffing overhead? What would be the benefits? How can or should each be quantified?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
Data Management Program Project Execution Layer	What is the function of the project execution layer?
	How is this layer related to the program planning and project plannning layers?
	What are the items within this layer? Are any missing? What are they if missing. Should any be deleted? Which and why?
	Identify which are the most important?
	Briefly describe what should be in each of the most important components of the project execution layer?
	How would you propose to accomplish the most important components in the project execution layer?
	How should project executions be monitored? What should be the metrics? How assessed?
	Should there be common tools and techniques? What should they be? How should they be used?
	How can lessons-learned from various projects be fed back into the project planning and program planning layers?
	How should ADS metadata and ADS data instances be integrated with other data management components and systems?
Data Standards XML	What is XML? What role does it play. When should it be used?
	What is an XML Schema?
	What is the relationship between XML and SQL?
	How should the XML schemas, tags, and supporting metadata be generated? Maintained? Should XML metadata just be another class of metadata that is integrated with all other classes of metadata? If yes, then why. If not, then why not? What would be the benefits, costs, and ROI of an integrated metadata repository that includes XML metadata?

7

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	Should XML tags map to an integrated set of view data elements, which in turn are mapped onto database table data elements, which in turn are mapped onto data structure template data elements, which in turn are mapped onto ISO 11179 data elements? If so, why? If not then upon which should the XML tags be mapped and why? What would be the cost, benefit, and ROI of any choice?
	How should XML tags for XML Elements/Attributes be generated? Manually? How, and how maintained? Automated? How and how maintained? Compare and contrast. Detail the benefits of each approach.
	Given that XML metadata will be defined by developers, and COIs how should redundancy and semantic conflicts be resolved?
	What supporting metadata and processes should be in place to ensure certain knowledge of granularity, reference data currentness, and the like?
	What mechanisms in SQL assist with XML schema and XML data stream formation?
	What supporting metadata and processes should be in place to ensure that XML schemas are considered to be the same even if their elements/attributes are named differently?
	What should be the relationship between a Cobol FD and a XML Schema? How might that be affected with a Cobol Redefines?
	Is an XML schema similar to the data structure of a business transaction? SQL View?
	What supporting metadata and processes should be in place to ensure that XML schemas that have the same or similar element/attribute tag names are able to be determined to be different?
Data Standards Enterprise Identifiers	What is an enterprise identifier (EID). What role does it play?
	How should existing data assets be identified and tagged with EIDs?
	How should EID values be engineered into existing data assets?
	How should EID values be generated and managed?
	How should EID values be engineered?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	What metadata should be associated with an EID? How is it collected, entered, updated and managed?
	How do you build a business case for EIDs?
	What problems does an EID solve?
	What problems does an EID cause?
	Should an EID be restricted to a function? A COI? A Service? Multi-Service? Federal Agency? Government wide? What problems are addressed and why? What problems might arise and why?
	Are they functionally dependent or independent? Compare and contrast?
	here have EIDs been built?
Data Standards Information Exchange Standards Specifications (IESS)	What is the relationship between ISO 11179 and an IESS?
	What is the relationship between data structure templates and an IESS?
	How will PMs, COIs, Services, Multi-service functions (Joint), and Federal Agencies participate, identify and resolve conflicts?
	What constitutes an IESS? What metadata should be required to support an IESS? What is built? Where should it be stored? How is its metadata integrated with other metadata?
	Describe the alternatives for its construction, that is, top-down, bottom-up, forward engineering or reverse engineering? What are advantages and disadvantages of each?
	What implementation alternatives are there for an IESS (e.g., shared database, shared tag sets, commonly employed data models)? Compare and contrast. How should a choice be made? What are the factors and metrics?
	What is an IESS? What role does it play?
Data Management Standards ANSI Three Schema Architecture	What is the ANSI Three Schema Architecture?
	How was it developed? By who? When? Who maintains it?
	Which of these are the three schemas? Conceptual? Logical?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	Physical? View/External? Where there more? Why? Why not?
	What is its role? How are the schemas related to each other? Are there any schemas missing? Why?
	Should there be a relationship between ISO 11179 and the ANSI three schema architecture? What would it be? Why?
Data Management Standards ISO/ANSI SQL	What is SQL? Who develops it? Who maintains it?
	What is the relationship between ISO/ANSI SQL and the Relational Model? Has ISO/ANSI SQL implemented the Relational Model? If yes then how? If no, then what's different?
	What are the parts of the SQL standard? What role do the various parts of SQL play within data management?
	Is the SQL language implemented by all the SQL DBMS vendors in the same way? If not, then why not?
	What has been implemented by SQL vendors in the same way? How would you find out?
	Which SQL DBMSs conform to the ISO/ANSI SQL standard? Which SQL standard year? Which Level? How would you know?
	Should there be Agency/Organization specific SQL language profiles that should be mandated on database application developers? Why? What benefits would accrue? What costs might incur to the agency/organization if a SQL language profile is not followed?
	What types of SQL language profiles should exist? If SQL data structures, then which SQL DDL features should be in the profile? What benefit would accrue? If SQL data manipulation language subsets, then what features should be in the profile? How should these be constructed? Maintained? Enforced? How could and should these be enforced?
	What is the relationship between ISO 11179 standard and SQL? What is the relationship between ISO/ANSI SQL and W3C?
Data Management Standards ISO Standard 11179	What is ISO 11179?
	How was it developed? By who? When? Who maintains?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	What is its role?
	What is the relationship between ISO 11179 and any of the schemas within the ANSI three schema architecture?
	What parts of this standard are normative and non-normative. What's the difference and why are some normative and non-normative?
	What part is critical to be implemented? How? What, if any are the problems? What are the solutions?
	How would this standard be integrated with other CASE and Repository tools? What would be the benefits? Problems?
Data Management Standards W3C	What is the W3C? How do you participate? Is it an open or closed organization?
	What standards does it develop and maintains it?
	Are these ISO/ANSI standards?
	What is the relationship between ISO 11179 standard and any standard of the W3C?
	As to the use of W3C standards within an organization, how would standard vocabularies be handled? How would they be defined, employed, and maintained? How will vocabularies be judged acceptable, stewarded, and tracked? What are the naming standards that should be employed within a XML environment? How are they the same/different from any other data element naming standards? If the same, then how would they be employed? Automatically? Manually? If different, then what and why?
	What is the relationship between a W3C standard and say, for example, ISO/ANSI SQL?
Data Models	What are commonly seen data architecture classes? For example, original data capture, data warehouses, reference data?
	When should each data architecture class be employed?
	What are the normal forms of data? Describe each?
	What are the benefits of adhering to each?
	When would you not adhere to one?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	What is referential integrity? Describe each type?
	What are Referential Actions? Describe each?
	What should the referential actions be for an "associative entity?"
	What is your approach to data modeling, that is, conceptual, logical, and physical? Are these data layers separate and distinct or just transformations?
	How would you manage enterprise wide data models through these conceptual, logical, and physical layers? How would data models be stored and made available for others to employ in the enterprise? What standards and/or rules would be imposed?
	What roles do data elements (11179, data structure template data elements, database table data elements, and application data elements play? How are they interrelated one with the other? How are these all related to the conceptual, logical, and physical data model layers?
Data Quality	What is data quality?
	How would you measure data quality?
	How would you assess data quality?
	What programs and processes would you establish to assess data quality, transform to quality data, and maintain data quality levels?
	How would you measure data quality ROI? What metrics should be employed?
Data Stewardship	What is data stewardship?
	What processes are involved in data stewardship?
	How would you assess whether data stewardship is being properly achieved?
	What programs and processes would you establish to assess data stewardship, transform to a data stewardship environment, and maintain a quality data stewardship environment?
	How would you measure data stewardship ROI? What metrics should be employed?

12

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
Data Structure Granularity	How should you document the units for a value?
	How should you document the time and/or calendar precision that might be in force?
	How should you know the capture time-zone?
	How should you know the precision of the data type?
	Why would you need to know which DBMS was employed for a numeric data type?
	What about BLOBs?
	What about CLOBs?
Data Structure Templates	Are there data structures (e.g., collections of data elements that serve a specific purpose, e.g., person full name, or address) that commonly appear across database designs? What would be typical examples?
	What benefit would there be if there was a library of data structure templates that could be employed in the building of database designs?
	How much would that save in time and cost? How would it contribute to data structure and semantic standardization? How should they be designed? Maintained?
	How should data structure templates be made available? How would you enforce their use?
	Should relationships exist between ISO 11179 data elements and data structure template data elements? If yes, then how should these relationships maintained? If not, why not? What is the cost of such relationships? Benefits? ROI?
	Should a data structure templates be able to be instantiated more than once in a database table? More than once in a database? If yes, why? If no, then why not?
Data Values	How could you discover that values have come from atomic data elements? What purpose would they serve? How would you find out?
	How could you discover that values have come from derived data elements? What purpose would they serve? How would you find out?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	How could you discover that values have come from compound data elements? What purpose would they serve? How would you find out?
	How could you discover the composition of compound data value? What purpose would they serve? How would you find out?
	How would you know if a compound data element value contains a derived data element as a member? What purpose would they serve? How would you find out?
	How do you know the process for data element value derivation?
	How do you know if a derived value process is unique? Shared? Obsolete?
Database Objects	What should object classes and object be defined as and employed in a persistent database environment? Should they embrace multi-table structures? Table-based processes? States? State change systems?
	How have object classes been specified within ISO/ANSI SQL? How are they the same? How are they different?
	What is the difference between an ISO/ANSI SQL object class and a programming language object class? Under which conditions should each be used?
	How should data life cycles be defined and managed? How would this relate to database object states? State changes?
	How would you approach the enforcement of a database object state? What would be specified? Implemented? Maintained?
	How can you define database object states within ISO/ANSI SQL?
Metadata Repositories	What is a metadata repository? What should be its contents?
	What classes of metadata should exist? Should there be metadata repositories for each class?
	Should a metadata repository be active? Passive?
	Should development of metadata products for the repository be on the critical path of projects?
	Who should "own" repositories?

| Data Management Knowledge Worker Categories and Questions ||
Item	Questions
	Should metadata repository data models be open?
	Should metadata repository process models be open?
	How should metadata be exchanged? Shared? Updated? Versioned?
	Should there be development, test, and production metadata partitions?
	What kind of metadata import and export facilities should exist?
	Should there be a single metadata repository or a federation of repositories? Compare and contrast. Should all the metadata in the repository be integrated? If yes, how? If not, then why not?
	If an integrated metadata repository, how would individual metadata uses be selectable? How multiple uses be selectable? How should reuse be measured? How should a ROI be calculated? What would be the metrics? How will the metadata repository's content and context be measured?
Reference Data	In a data stream (e.g., XML, EDI, or formatted), how would you be able to discover that a value is a data value or a code?
	If a code, how would the meaning be found?
	If a meaning, how would the code be found?
	How should the effective dates of codes be engineered? Implemented? Employed?
	How should the assignment dates of codes be engineered? Implemented? Employed?
	If there are different data streams that need to be compared, how would you know that the codes represented by data elements are current?
	How would the mappings of code values be engineered, audited, and versioned?
	How should code values maintained?
	How should code values be established?

Data Management Knowledge Worker Categories and Questions	
Item	**Questions**
	What metadata about reference data should be captured. How should it be engineered? Stored? Evolved? Maintained? Be made visible?
	What ANSI/ISO standards are employed to identify, capture, and maintain types of reference data? How do you know?
	Who should maintain codes and values? Within a PM? Within a COI? Across COI's? Across Services? Across Federal Agencies?
Security	What is data security? What should it encompass?
	How should data security be defined? Implemented? Maintained?
	How is data security related to process security?
	How should SQL security (access control) facilities be integrated with other computing environment facilities? What components of SQL access control are acceptable? Need improvement?
Transaction Management	What is transaction management?
	How is transaction management implemented in database environments?
	What are Savepoints? How are they implemented?