Designation: E2145 – 07 (Reapproved 2013)

An American National Standard

Standard Practice for Information Modeling[1]

This standard is issued under the fixed designation E2145; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope*

1.1 Information models are increasingly important in the analysis, design, and sharing a common understanding in information engineering, in business process improvement, in building information systems, in developing informatics standards, and in many other uses.[2]

1.2 The purpose of this practice is to identify best practices for the creation, use, and assessment of various types of information models.

1.3 Included in this practice are recommended organizational policies and procedures, where modeling is best used and recommended modeling methods, best practices and evaluation criteria.

1.4 Excluded from this practice are detailed specifications of modeling techniques that are specified or described in other sources.

2. Referenced Documents

2.1 *ANSI Standards:*[3]

ANSI X3.172-1990 Dictionary for Information Systems

IEEE 1320.1-1998 Standard for Functional Modeling Language—Syntax and Semantics for IDEF0

IEEE 1320.2-1998 Standard for Functional Modeling Language—Syntax and Semantics for IDEF1X (Object 97)

2.2 *ISO Standards:*[3]

ISO 8601-88 Data Elements and Interchange Formats—Representation of Dates and Times

ISO/IEC 1087 Terminology Work—Vocabulary-Part 1: Theory and Application

ISO/IEC 11179 Information Technology—Specification and Standardization of Data Elements, Parts 1-6

ISO/IEC 19501:2005 Information technology—Open Distributed Processing—Unified Modeling Language (UML) Version 1.4.2

ISO/IEC 2382 Information Processing Systems—Vocabulary

2.3 *Other Documents and Standards:*[4]

Unified Modeling Language Specification version 1.5, March 2003

3. Terminology

3.1 The following sections present terms, definitions, and acronyms found in this practice and in various modeling activities. These terms and definitions are referenced or derived from ANSI X3.172 or ISO/IEC 2382 unless otherwise cited.

3.2 *Definitions:*

3.2.1 *activity, n*—a group of logically related tasks performed for a purpose.

3.2.2 *alternate key attribute, n*—in a logical data model, any candidate key of an entity other than the primary key.

3.2.3 *application model, n*—a representation or description of the application software, programs, or components needed to support a business function.

3.2.4 *attribute, n*—a characteristic of an object or entity (see ISO/IEC 11179).

3.2.5 *attribute value, n*—a representation of an instance of an attribute (see ISO/IEC 11179).

3.2.6 *behavior, n*—in the object-oriented methodology, behavior constitutes the observable effects of an operation or event, including any results generated or obtained; and represented by operations, methods, and state machines (see UML 1.5).

3.2.7 *business model, n*—a representation of the strategy, situation, environment, objectives, direction, and similar characteristics of a business enterprise or business area.

3.2.8 *business process model, n*—a representation of business processes, often where these processes are successively decomposed to describe component activities, to identify the events to which the business shall respond, and to identify the results produced.

[1] This practice is under the jurisdiction of ASTM Committee E31 on Healthcare Informatics and is the direct responsibility of Subcommittee E31.25 on Healthcare Data Management, Security, Confidentiality, and Privacy.

Current edition approved March 1, 2013. Published March 2013. Originally approved in 2001. Last previous edition approved in 2007 as E2145 – 07. DOI: 10.1520/E2145-07R13.

[2] An information model in the context of this standard is any representation of process, data, etc., used in any aspect of information technology or information management.

[3] Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036, http://www.ansi.org.

[4] Available from Object Management Group. OMG Headquarters, 250 First Avenue, Needham, MA 02494. http://www.omg.org.

*A Summary of Changes section appears at the end of this standard

Copyright © ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959. United States

Documents
Document Sharing Hub

Full document details: ASTM E2145-07(2013)

3.2.9 *column, n*—a physical data model and relational database structure that is analogous to the attribute of a logical data model.

3.2.10 *concept, n*—a unit of thought constituted through abstraction on the basis of characteristics common to a set of objects (see ISO/IEC 1087).

3.2.11 *conceptual model, n*—an abstract representation of any type of model used to identify the principal components and relationships of the subject of the model, and avoiding unnecessary or confounding detail.

3.2.12 *context, n*—a designation or description of the application environment or discipline in which a name is applied or from which it originates (see ISO/IEC 11179).

3.2.13 *data, n*—a representation of facts, concepts, or instructions in a formalized manner, suitable for communication, interpretation, or processing by humans or automatic means (see ISO/IEC 11179).

3.2.14 *data dictionary, n*—a database used for data that refers to the use and structure of other data; that is, a database for the storage of metadata (see ANSI X3.172-1990).

3.2.15 *data element, n*—a unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes (see ISO/IEC 11179).

3.2.16 *data model, n*—a description of the organization of data in a manner that reflects an information structure (see ISO/IEC 11179).

3.2.17 *data steward, n*—a person or organization delegated the responsibility for managing a specific set of data resources, (see ISO/IEC 11179).

3.2.18 *dependent entity, n*—in a logical data model, an entity that inherits one or more identifying attributes from another entity.

3.2.19 *domain, n*—the set of possible data values of an attribute (see ISO/IEC 2382); also the identification, description or scope of a segment of industry or commerce, or of a program, project, or problem.

3.2.20 *encapsulation, v*—the packaging of an object's data with its corresponding methods with access via a controlled messaging interface (see UML 1.4).

3.2.21 *entity, n*—any concrete or abstract thing of interest, including associations among things (see ISO/IEC 2382).

3.2.22 *external schema, n*—an external schema is the representation of data at the system architecture presentation layer as viewed by the user when interacting or communicating with the information system.[5]

3.2.23 *foreign key attribute, n*—in a logical data model, an attribute or combination of attributes of a child or category entity whose values match those in a primary key of a related or generic entity instance.

3.2.24 *function, n*—an activity, process or transformation identified by a verb that describes that which would be accomplished.

3.2.25 *independent entity, n*—in a logical data model, an entity that does not inherit identifying attributes from another entity.

3.2.26 *inheritance, v*—the transmission of characteristics from a parent process, entity, or object to its children (see UML 1.4).

3.2.27 *internal schema, n*—a schema of the ANSI Three Schema architecture in which views of the information are representations of data structure at the system architecture data layer.

3.2.28 *key attribute, n*—in a logical data model, an attribute used to identify an instance of an entity.

3.2.29 *key inheritance, v*—the transmission of a key attribute from a parent or independent entity to a child or dependent entity.

3.2.30 *lexical, n*—pertaining to words or the vocabulary of a language as distinguished from its grammar and construction (see ISO/IEC 11179).

3.2.31 *location model, n*—a representation or description of the components and interrelationships of a real or virtual location.

3.2.32 *logical model, n*—an expansion of a conceptual model, or a derivation from a physical model, that provides a level of detail sufficient to design or effect a business solution.

3.2.33 *metadata, n*—data that describes other data (see ISO/IEC 11179).

3.2.34 *model, n*—a representation in whatever form of a real or envisioned object, action, or process.

3.2.35 *modeling, v*—the process of creating, revising, documenting and presenting a model.

3.2.36 *model view, n*—a collection of models pertaining to a particular domain and from the perspective of a particular individual's role or viewpoint.

3.2.37 *non-key attribute, n*—in a logical data model, an attribute that is not the primary or part of a composite primary key of an entity.

3.2.38 *object, n*—any part of a conceivable or perceivable world (see ISO 1087).

3.2.39 *object class, n*—a set of objects, ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose properties and behavior follow the same rules (see ISO/IEC 11179).

3.2.40 *object-oriented methodology, n*—any of the formal modeling methods or techniques that employ objects to describe data and the methods or processes acting on those data as a single unit for analysis, design, or development (see UML 1.5, ISO/IEC 11179).

3.2.41 *organizational model, n*—a representation or description of the components and interrelationships of an organization or organizational component.

[5] ANSI Standards Planning and Requirements Committee—a layered model of database architecture comprising a physical schema, a conceptual schema, and user views.

AVA Documents
Document Sharing Hub

Full document details: ASTM E2145-07(2013)

3.2.42 *physical model, n*—a derivation from a logical model, or a physical implementation, providing a highly detailed definition of the solution needed to carry out the activities in business area.

3.2.43 *primary key attribute, n*—in a logical data model, the candidate key selected as the unique identifier of an entity.

3.2.44 *property, n*—a peculiarity common to all members of an object class (see ISO/IEC 11179).

3.2.45 *relational data methodology, n*—any of several modeling and information management methods and techniques that employ structured relationships among entities or tables.

3.2.46 *row, n*—a physical data model and relational database structure that contains a single instance of data in a table.

3.2.47 *structured analysis and design, n*— any formalized methodology for the analysis, design, development or lifecycle management of information systems where processes and data are initially treated individually and subsequently integrated in a system architecture.

3.2.48 *subject area, n*—a subject area is a portion of an entire data model which is created to facilitate understanding of a specific functional area or component task.

3.2.49 *table, n*—a physical data model and relational database structure that is analogous to the entity of a data model.

3.2.50 *technology model, n*—a representation or description of the hardware, system software, and network components needed to support the business area.

3.2.51 *view, n*—a collection SQL queries stored in a relational database under assigned names and represented as a temporary or virtual table that does not permanently store the data it presents.

3.3 *Acronyms:*

3.3.1 *ANSI*—American National Standards Institute

3.3.2 *ASTM*—American Society for Testing and Materials

3.3.3 *CORBA*—Common Object Request Broker Architecture

3.3.4 *CRC*—Class, Responsibility and Collaboration Approach

3.3.5 *DFD*—Data Flow Diagram

3.3.6 *ICOM*—Input, Control, Output and Mechanisms as used in IDEF activity modeling

3.3.7 *IDEF*—Integrated Definition Language

3.3.8 *IDEF0*—The IDEF activity modeling language

3.3.9 *IDEF1X*—The IDEF data modeling language

3.3.10 *IDO*—Integrated Delivery Organization

3.3.11 *IE*—Information Engineering diagramming methodologies

3.3.12 *IEC*—International Electrotechnical Commission

3.3.13 *ISA*—Information Systems Architecture

3.3.14 *ISO*—International Organization for Standardization

3.3.15 *NIST*—National Institute of Standards and Technology

3.3.16 *OMG*—Object Management Group

3.3.17 *RDB*—Relational Database

3.3.18 *RDBMS*—Relational Database Management System

3.3.19 *SADT*—Structured Analysis and Design Technique

3.3.20 *SML*—Standardized Modeling Language

3.3.21 *SQL*—Structured Query Language

3.3.22 *SSADM*—Structured Systems Analysis and Design Methodology

3.3.23 *UML*—Unified Modeling Language

4. Summary of Practice

4.1 This practice describes policies and procedures, best practices for the creation, use, and assessment of information models.

4.2 The foundation for these information modeling best practices is derived from Donnabedian's structure, process and outcome quality constructs (**1**).[6]

4.2.1 *Structure*—Structure constructs identify or describe the organization component where modeling activities occur, the composition of the modeling method or tool, its language and syntax, and the resources allocated to perform modeling processes, for example human resources, technology, materiel, etc.

4.2.2 *Process*—Process constructs identify or describe the appropriate use of modeling, the effective use of modeling methods, and the correct employment of the modeling technique.

4.2.3 *Outcome*—A quality information model that provides a meaningful and usable representation of past, present, or future reality, and can be assessed by five dimensions (**2**) of model quality:

4.2.3.1 *Conceptual Correctness*—The model accurately reflects functional or business concepts where all models are representations of real or envisioned objects or concepts. The model shall accurately describe the intended structures or processes or both.

4.2.3.2 *Conceptual Completeness*—The model contains sufficient objects to describe the full scope of the functional or business domain under consideration. The model describes the full scope or an element of the structures or processes of a concept or both.

4.2.3.3 *Syntactic Correctness*—The objects in the model do not violate syntactic rules of the modeling language. Each modeling method has a predefined language typically consisting of visual structures and symbols where these structures and symbols are assembled and interpreted according to a clearly defined and unambiguous set of rules. These rules shall be widely recognized in the industry and preferably be standardized (**3**). The modeling methodology and processes shall adhere to these rules. The model shall effectively communicate its representations and descriptions to anyone who understands the modeling method and syntax.

4.2.3.4 *Syntactic Completeness*—All essential functional or business concepts are captured at appropriate points in the

[6] The boldface numbers in parentheses refer to the list of references at the end of this standard.

AVA Documents
Document Sharing Hub

Full document details: ASTM E2145-07(2013)

modeling process and represented in the model products. A modeling method may use different objects, symbols or structures at different times in the modeling process. The model shall display the appropriate types and numbers of structures and symbols at the appropriate place in the model and at the appropriate time in the modeling process.

4.2.3.5 *Enterprise Awareness*—The model depicts the entire enterprise or can be seamlessly integrated with other models across the entire organization. Enterprise awareness is the ability of a model to describe the entire organization, domain, or a comprehensive segment or component of industry, business, or commerce, either by itself or when combined with other comparable models. Multiple models shall seamlessly integrate across the entire organization with representations and descriptions easily flowing from one model to another. The Zachman Information Systems Architecture (**4**) is one of several frameworks suitable for identifying models that represent the informatics structure and processes of the organization. The Zachman Framework consists of six columns and five rows where the columns describe data, function, network, people, time, and motivation, and answer the fundamental questions *what*, *how*, *where*, *who*, *when*, and *why*, and the rows describe different perspectives as contextual scope, conceptual models, logical models, physical design, and detailed representations (physical implementation). The foot of the framework represents the actual business enterprise.

In the framework, an upper row essentially presents requirements for a lower row and each cell can contain directional, conceptual, logical, or physical models or combinations thereof.[7]

4.3 *Additional Technical Considerations:*

4.3.1 In addition to the above quality dimensions, several qualitative and technical considerations impact the quality of an information model (**3**).

4.3.2 *Nonredundancy*—A fact shall be represented at only one point in the model.

4.3.3 *Business Rules Enforcement*—The model accurately reflects real or envisioned functional or business processes.

4.3.4 *Reusability*—The content of the model is accessible to other than the originally intended users.

4.3.5 *Stability*—The model accepts changing functional or business requirements without significant alteration in the structure of the model.

4.3.6 *Flexibility*—The model is capable of supporting change in domain concepts or business processes.

4.3.7 *Simplicity*—The model is easily understandable and clearly presented.

4.3.8 *Implementable*—The model may be readily and economically used in developing strategy, and in the analysis, design, development, or operation of business processes or information systems.

4.3.9 *Communicability*—The model readily conveys meaning to its users and readers.

4.4 *Assessing the Modeling Process and Model Quality:*

4.4.1 The dimensions of model quality that are derived from structure, process, and outcome quality constructs lead to best practices in the following areas:

4.4.1.1 *Structure*—The organization desiring to employ modeling creates and maintains an environment conducive to the creation, evaluation, communication and use of quality information models.

4.4.1.2 *Process*—Modeling methods are appropriately selected and applied through effective use of modeling language and adherence to syntax to provide a meaningful, comprehensive, and balanced representation of real world functional or business concepts.

4.4.1.3 *Outcome*—The resulting model(s) exhibit a high degree of excellence and wholeness, are identified by author, participants, technique and approach, perspective, view and subject areas, and may include test data for evaluation purposes.

4.4.2 Annex A1 of this practice provides a checklist for assessing the quality of models and modeling activities in terms of structure, process, and outcome.

5. Significance and Use

5.1 Modeling is increasingly used in business, industry, and commerce to develop a common understanding of processes, functions, activities, and supporting data. Typical users of such models are systems developers, operations researchers and business analysts, educators, and executives.

5.1.1 Information models are regarded widely as beneficial by saving cost through realignment of processes, risk reduction and elimination of redundancy. Information models convey ideas and facilitate the analysis and understanding of complex processes and structures. These models form the basis for software engineering practices that build systems and databases, redefine organizational structures, improve business processes, and develop standards.

5.1.2 This practice provides a practical means for developers and users of information models to employ appropriate modeling methods and to objectively determine model quality.

5.2 *Background:*

5.2.1 Models are representations of past, existing, or contemplated reality. Models may assist in the explanation or analysis of complex structures and processes that may exceed human capacity for direct visualization or understanding. Models enable a focus on the key elements of a process or structure while ignoring confounding or irrelevant elements. As such, models make an explicit statement of the meaning of the reality being modeled.

5.2.2 Integrated information engineering models provide a coherent view of the processes and data of an organization or enterprise (**5**). Activity models identify the fundamental tasks performed in a function. Process models accurately describe the detailed collection of these activities within an organization. Data models are derived from and support the functions described in activity models. Object models also characterize the processes and data required to understand business operations. Both structured and object-oriented models may be used to construct information systems that support those business operations. Application models describe in varying levels of

[7] Additional detail about the Zachman Framework is available at http://www.zifa.com/.

AVA Documents
Document Sharing Hub

Full document details: ASTM E2145-07(2013)

detail, the overall architecture and components of the software needed to support the envisioned business functions. Organizational models reflect the current and envisioned future state of the organization, especially as this impacts business processes enabled or supported by information systems. Location models identify and describe the business and geographic position and relationships of structural components of an organizational entity. Technology models describe in varying levels of detail those hardware, system software, and network components needed to operate the information systems supporting a business area.

Models may be textual, graphic, or mixed graphic and text forms, including, tables and structured lists, flowcharts, process flow diagrams, state diagrams, data flow diagrams, entity-relationship diagrams, and related techniques develop an understanding of business processes and the transformation of data through these processes (**6**). Modeling products may also relate two or more types of models, such as for an Application-Data Matrix, or Technology-Location Plan.

5.3 *Framework for Relating Models to Systems:*

5.3.1 Since proposed by Zachman in 1987 (**4**), the Information System Architecture (ISA) has become one of several means to view and to blueprint information systems within an entire organization, to design business processes and change enterprise strategies, and to describe and interrelate the components of an information system to the entire organization. The cells within the framework represent primitive architectural constructs where the sets of columns collectively represent a consistent integrated model for each perspective (row). These rows describe different types of artifacts, each having a different purpose:

5.3.1.1 Row 1—artifacts define the scope as boundaries of the enterprise.

5.3.1.2 Row 2—artifacts define the conceptual design as envisioned by the enterprise "owners" or stakeholders.

5.3.1.3 Row 3—artifacts present a logical design, how the enterprise concepts will be realized independent of technology.

5.3.1.4 Row 4—artifacts describe the physical design, how the enterprise implementation is generally constrained by technology.

5.3.1.5 Row 5—artifacts specify the physical implementation as a specific application of technology.

5.4 *Principles and Approaches to Modeling:*

5.4.1 Modeling activities have been prevalent throughout the history of the engineering disciplines. This rich history indicates four basic principles (**7**):

5.4.1.1 The choice of what models to create has a profound influence on how a problem is approached and how a solution is shaped.

5.4.1.2 Every model may be expressed at different levels of precision.

5.4.1.3 The best models are connected to past, current, or envisioned future reality.

5.4.1.4 No single model is sufficient; every nontrivial system is best approached through a small set of nearly independent models.

5.4.2 The three approaches to information modeling are characterized as top-down, bottom-up, and inside-out (**8**). Of these, the top-down and bottom-up approaches are most often used.

5.4.2.1 The "top-down" approach follows a progression of activities that first analyzes and understands functionality of the domain-of-interest and only then progresses to model preparation. The top-down paradigm typically consists of preparing a high level or conceptual model followed successively by logical and physical models in increasing levels of detail. The strength of this approach is it provides a broadly-based and comprehensive set of rationally derived models that are generally appropriate across the spectrum of functional or business activities. This paradigm saves time, conserves resources, and reduces risk over the long-term course of system development.

The "bottom-up" approach reflects the experience of subject matter experts in one or more specialized subdomains. These panels of experts prepare small models that are subsequently integrated. The strength of this approach is that it addresses detail that can be easily converted to a specific application, and may be tuned or optimized for a small and contained segment of the business environment. Often, it is used when the entire domain is not well understood or in the absence of a business model foundation. The bottom-up paradigm solves small problems quickly. It produces narrowly focused model components requiring extensive effort for subsequent integration into systems with a broader scope or enterprise solutions.

5.5 *Use of Modeling Techniques for Content*—The complexity and extensiveness of business processes, the information used in these processes, and the information systems supporting these processes inhibit or prevent a comprehensive understanding to the depth of detail frequently required. Formal methods are needed to facilitate the development and sharing of a common understanding about information, the processing of that information, and the systems supporting those processes.

5.6 *Use of Models for System Engineering:*

5.6.1 System engineering is considered in all leading system lifecycle methodologies as a sequential, cyclic, or iterative sequence of four essential activities from a visionary or inception activity to deployment or implementation (**9**) as illustrated in Fig. 1. Performance and outcome of these activities can be substantially improved through modeling. For example, system development models may consist of elementary structures like an organizational chart to detailed structural representation of a software package. Models typically are employed in the planning and analysis phase of software development, in system and software design, and in the coding process.

5.6.1.1 Booch, et al (**7**) , state that the single fundamental reason for modeling a system is to better understand the system being developed, achieving four aims:

(1) To visualize an existing or desired system;

(2) To specify the structure and behavior of a system;

(3) To provide a guiding template for system construction; and

AVA Documents
Document Sharing Hub

Full document details: ASTM E2145-07(2013)