

Data Modeling Design Standards

© 2008 Nationwide Mutual Insurance Company. All Rights Reserved.



TABLE OF CONTENTS

DATA MODELING.....	1
DESIGN STANDARDS.....	1
1. EXECUTIVE OVERVIEW	3
2. STANDARDS	4
2.1 DEVELOPMENT GUIDELINES	4
Logical Data Model.....	4
Referential Integrity.....	6
2.2 OBJECT NAMING STANDARDS	7
General Naming Standards	7
Physical Naming Standards.....	8
2.3 DEFINITION GUIDELINES	8
3. APPENDIX	9
APPENDIX A: CLASS WORDS	9
Class Words.....	9
APPENDIX B: ABBREVIATIONS & ACRONYMS.....	12
4. DOCUMENT INFORMATION.....	13
EXCEPTIONS.....	13
DOCUMENT HISTORY.....	13
APPROVALS AND CONTRIBUTORS.....	14
STANDARDS CURRENCY: LOCATION, ACCESSIBILITY, REVISIONS	15

1. Executive Overview

The purpose of this document is to provide a set of enterprise-wide Standards for Data Modeling at Nationwide. The Nationwide enterprise includes all Nationwide entities regardless of geographic location, line of business or channel of distribution.

Standards are comprised of statements that define what *must* be done to contribute to the most effective performance of a discipline.

This document addresses relational and dimensional data modeling in general without regards to specific tools or type of database. This version focuses on Standards with procedures to be addressed at a later time. These Standards were based upon a broad range of experience and knowledge by Nationwide Data Modelers and Architects across the enterprise. Existing Standards and Best Practices were reviewed and incorporated where appropriate.

The Standards within this document are intended to be used on design and development of new databases. Maintenance to existing data models should follow these Standards as much as possible realizing that concessions may need to be made for purchased packages or existing databases where different standards were used when originally created (grandfather clause).

The Standards are defined by a number used for easy reference and tracking purposes. Each Standard has been assigned a unique number in chronological order by which they were developed. Thus, Standards may not appear in numeric order as new Standards can be inserted between older Standards. Each Standard has a short and long description statement and rationale followed by supporting statements as needed.

2. Standards

2.1 Development Guidelines

Logical Data Model

<p>Data Modeling Standard #S1: Adhere to ESDm for Logical Data Modeling.</p> <p>Description: Use the ESDm Logical Data Model instructions when creating logical data models.</p> <p>Rationale: Proper data design ensures that all data elements are understood both individually and in relation to other data elements. This improves the ability to reuse and extend data assets without major redesign and reduces the need for structural changes to the data assets after initial deployment.</p>	<p>Source: P&C AD1.2.1 – Logical Models (OCTO)</p> <p>Status: Data COE Approved</p>
<p>Data Modeling Standard #S2: The Logical Data Model shall be independent of physical implementation considerations.</p> <p>Description: When designing the database, logical data modeling will be independent of the specific database the design will be physically implemented in as well as the tool used to access the data.</p> <p>Rationale: A logical model is a model that describes the logical (non-physical) data and data relationships used without considering physical aspects of either the current or anticipated implementation.</p>	<p>Source: DCDirect-Data-4</p> <p>Status: Data COE Approved</p>
<p>Data Modeling Standard #S3: Adhere to third normal form or greater.</p> <p>Description: The logical data model will be represented in third normal form or greater.</p> <p>Rationale: Third normal form reduces the duplication of data and to minimize create, update, and delete anomalies of data that can negatively affect data integrity and consistency. At the physical level, denormalization may be appropriate for performance issues, but all reasonable efforts must be made to eliminate the performance problem prior to de-normalization.</p>	<p>Source: NF Logical Data Modeling and Object Naming Standards and DCDirect-Data-5</p> <p>Status: Data COE Approved</p>
<p>Data Modeling Standard #S4: Data elements shall be defined and follow ISO Definition standards.</p> <p>Description: Follow ISO/IEC 11179-4, an industry standard adopted by Nationwide. This standard contains explicit rules and guidelines on the formulation of data definitions.</p> <p>Rationale: A precise, well-formed definition is one of the most critical requirements for a shared understanding of an administered item. Well-formed definitions are imperative for the exchange of information. Only if every user has a common and exact understanding of the data item can it be</p>	<p>Source: Data COE</p> <p>Status: Data COE Approved</p>

Data Modeling Standards

consumed easily. Reference: ISO/IEC 11179-4 can be found at http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html	
Data Modeling Standard #S5: Each attribute will have one unique objective. Description: Attributes should have one and only one intention. Attributes that serve multiple purposes should be created as separate attributes. Multiple attributes with the same objective should be combined into one data element. Rationale: Attributes that store different data values for multiple purposes require system and resource overhead to understand and use.	Source: Derived from DCDirect-Data-18 Status: Data COE Approved
Data Modeling Standard #S6: Optionality of all attributes shall be identified. Description: The data model should reflect whether or not data is required for each attribute. Rationale: Validation of the data model is easier with optionality noted on the model. The optionality is required for the physical data model and eventual DDL creation. Defining nulls at a later phase in the project causes unplanned discussions and rework in design. It is easier to define nulls at the time of detailed design since it is fresh at that point of time.	Source: NF Logical Data Modeling and Object Naming Standards Status: Data COE Approved
Data Modeling Standard #S7: Identify all natural primary keys. Description: Whether a surrogate key is used or not, the attributes that make a given row unique should be identified. Rationale: In order to properly normalize the data structures, it is important to understand whether each of the data elements is fully dependent on nothing but the entire key (third normal form). Implementation: Features within data modeling tools may handle this display differently. Tools that identify unique indexes (thus the combination of attributes making a unique row) may have display values built into them; i.e., Erwin uses AK#.# to identify the order within a unique index. A simpler tool may not have that ability in which case changing the font to bold or italics could be used. A legend on the model should always be used to explain the notation.	Source: NF Logical Data Modeling and Object Naming Standards Status: Data COE Approved
Data Modeling Standard #S8: Identify all relationships between entities. Description: The data model should document the link between entities describing the business rules governing how the entities interact in the real world. Rationale: Noting relationships on the data model is required for data design sessions where the data model is being validated. A clear understanding of the relationships between entities is required for input into the physical data model to ensure the corresponding linkages (i.e., indexes, foreign keys) are built.	Source: NF Logical Data Modeling and Object Naming Standards Status: Data COE Approved
Data Modeling Standard #S9: Identify the cardinality of all relationships. Description: The number of instances that each entity can be related to a	Source: NF Logical Data

Data Modeling Standards

<p>single instance of another entity should be instantiated in the data model with numeric representation. In the logical data model, cardinality between two entities may be one-to-zero, one-to-one, one-to-many, or many-to-many.</p> <p>Rationale: The cardinality embedded in a data model documents the business rules about relationships between data in a manner that is relevant to the scope. The implication of specifying the incorrect cardinality is significant.</p>	<p>Modeling and Object Naming Standards</p> <p>Status: Data COE Approved</p>
---	---

Referential Integrity

<p>Data Modeling Standard #S10: Relationships and Foreign keys must be documented in the logical data model.</p> <p>Description: The relationships between entities must be noted in the data model along with the corresponding foreign keys created as a result of the relationship.</p> <p>Rationale: Identification of foreign keys in the logical data model allows for verification of relationships of the data. It is necessary to understand this is the logical model before going to the physical model where the relationship is translated into indexes and constraints to enforce data integrity of the relationship.</p>	<p>Source: Data Modeling Working Group</p> <p>Status: Data COE Approved</p>
--	---

<p>Data Modeling Standard #S11: Natural keys must be documented in the Logical Data Model.</p> <p>Description: Attributes familiar to the data consumers that uniquely identify individual occurrences of an entity must be documented in the logical data model.</p> <p>Rationale: The logical data model serves as a communication tool for validation of data requirements and relationships of data. As such, it is important to ensure it represents the attributes that comprise the natural key for each entity so the resulting physical data model can be used for properly building the database indexes and constraints to enforce the uniqueness of the natural key.</p>	<p>Source: Data Modeling Working Group</p> <p>Status: Data COE Approved .</p>
---	---

2.2 Object Naming Standards

General Naming Standards

<p>Data Modeling Standard #S12: Do not use double negatives when naming objects.</p> <p>Description: Names should avoid the use of negatives when possible, but should never use double negatives.</p> <p>Rationale: Double negatives can be easily misunderstood and require additional time to accurately process.</p>	<p>Source: Data Modeling Working Group</p> <p>Status: Data COE Approved</p>
---	---

Entities

<p>Data Modeling Standard #S13: Entities must have meaningful, reasonably unambiguous names that is reflective of its contents.</p> <p>Description: Some ambiguity is unavoidable, but avoid genuinely misleading names and names that have multiple distinct meanings to the business. Avoid entity names made up of numbers exclusively.</p> <p>Rationale: Entity names that are self-explanatory of the contents are more user-friendly and less likely to be misunderstood. Entity names such as '1' or '1099' are confusing to users and reflect incomplete data modeling analysis and design.</p>	<p>Source: Data Modeling Working Group</p> <p>Status: Data COE Approved</p>
--	---

Attributes

<p>Data Modeling Standard #S14: The attribute name should be alpha or alphanumeric.</p> <p>Description: Avoid naming an attribute with all numbers. A properly defined attribute should contain alpha characters representing verbiage of the contents. Numbers should only be used when they are an important component of the name.</p> <p>Rationale: An attribute that is all numbers is indicative of incomplete data analysis and definition. For example, an attribute named 1099 to indicate whether a 1099 is needed to be generated for an event should contain additional verbiage to indicate what the column represents.</p>	<p>Source: Data Modeling Working Group</p> <p>Status: Data COE Approved</p>
---	---

Data Modeling Standards

Physical Naming Standards

<p>Data Modeling Standard #S15: The Standard Abbreviation List should be used for words that are abbreviated.</p> <p>Description:</p> <p>Rationale: Use of abbreviations should be consistent within a table. For example, if a column in a table abbreviates a word in one place, every column in the table, including the table name that uses that word should use the same abbreviation.</p> <p>Exception: When the abbreviation is already used in the existing database that is being altered, the abbreviation currently in use shall be used for new tables and columns.</p>	<p>Source: Data Modeling Working Group</p> <p>Status: Data COE Approved</p>
---	---

2.3 Definition Guidelines

Refer to ISO/IEC 11179-4 at <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html> for guidance on Definitions of Entities and Attributes. Tables and columns should inherit the same definition guidelines.

<p>Data Modeling Standard #S16: Entities, attributes, tables, and columns must be defined.</p> <p>Description: Definitions must be comprised of grammatically and structurally correct and complete simple sentence(s) describing the fundamental purpose of the data. Definitions define the person, place, thing, event, or concept/characteristics that are represented.</p> <p>A definition should <i>never</i> be a tautology, i.e. defined by itself. For example: "A unicorn is a beast with one horn" tells us nothing because the word unicorn is a compound word, and its parts are Uni = one and corn =horn. More common examples of tautological definitions are:</p> <ul style="list-style-type: none">• Customer ID is defined as "The identifier of the Customer."• "Metal is something made of metal." <p>Rationale: Definitions facilitate communication to ensure their exact meaning is understood. Good definitions help data consumers make accurate decisions. Data sharing and integration can only occur when the data elements and structures are understood and clearly defined.</p>	<p>Source: P&C Data Warehouse Data Modeling Standards and Guidelines</p> <p>Status: Data COE Approved</p>
--	---

3. Appendix

Appendix A: Class Words

The list below represents class words approved for use in Enterprise models.

Class Words

Name	3 Character Abbreviation	2 Character Abbreviation	Definition
Amount	AMT	AM	Monetary quantity
Audio	ADO	AD	Sound Object
BLOB	BLB	BB	Short for binary large object, a collection of binary data stored as a single entity. BLOBs are used primarily to hold multimedia objects such as images, videos, and sound, though they can also be used to store programs or even fragments of code.
CLOB	CLB	CB	Short for character large object. A collection of formatted character data stored as a single element. Blocks of XML, JSON, NVP or CSV data can be represented this way.
Code	CDE	CD	Alphanumeric character string representing a fact or property of an entity
Count	CNT	CN	Non monetary numeric value arrived at by counting. Counts of transaction and activity occurrences.
Date	DTE	DT	A point in time expressed as fully qualified date or a component of a date, including year, month, quarter, day or any combination of those. Can be stored in any date format: (Gregorian, Julian, etc.)
Datetime	DTM	DM	Full specification of date and time to the maximum precision (at least nanosecond) allowed in the DBMS being used. Do not use for business data - use when system-assigned time sequence is needed.
Description	DSC	DC	A word or phrase containing the meaning of an abbreviation or code. Example: "Premium Personal Interest Free Checking Account" describes product_code PPIFC

Data Modeling Standards

Name	3 Character Abbreviation	2 Character Abbreviation	Definition
Factor	FCT	FR	Number representing a relative measure.
Identifier	(Use 2 Character Abbreviation)	ID	Unique identifier of an entity. Business identifiers uniquely label business objects and may use pneumatic codes, such in a part_id or product_id. Surrogate identifiers usually have no meaning or intelligence. They are unique integers used to create unique table keys.
Image	IMG	IM	Still Image Object
Indicator	IND	IN	Boolean (yes/no) property or condition. An element which has a domain of only 2 states. Y/N
Multi-Media	(Use 2 Character Abbreviation)	MM	Multi-Media Object
Name	NME	NM	Name of person or object
Number	NUM	NB	Non-quantifying alphanumeric string uniquely identifying a person, place, or thing. It does not express measurement. It may express a sequence. Arithmetic operations are not performed on numbers. Use number only if it has common usage, such as Social Security Number (SocialSecurity_Nb) or Sequence Number (Sequence_Nb).
Percent	PCT	PC	Number representing a relative measure or ratio between two objects based on 100
Quantity	QTY	QT	A non-monetary total or value, such as the shares a person owns of a particular stock, distance, length, width.
Rate	RTE	RT	A quantity measured with respect to another measured quantity
Series	SRS	SR	Use for a repeating group of the same data element. For example, Contact Preference Series consists of repeating locality types, such as Home, Office, Roam, Secondary Residence.
Set	SET	ST	Use for a group name of a structured set of data or decomposable taxonomy. For example, SIC (Standard Industry Classification) Set where first two positions are Industry Group and last two are Specific Industry.

Data Modeling Standards

Name	3 Character Abbreviation	2 Character Abbreviation	Definition
Text	TXT	TX	Any freeform comment or notes. Data. Text, unlike name, description, and address may not have a pre-determined purpose.
Time	TME	TI	A unique point within a single day expressed as fully qualified time or hour, minute, second, or fraction of, in any combination (HH:MM:SS, HHMMSS, HH, etc). The time class word does not include quantity of time, such as the number of minutes or hours to perform a task. Time quantities are represented with quantity class words and time modifiers.
Timestamp	(Use 2 Character Abbreviation)	TS	Calendar date and time expressed in YYYY-MM-DD HH:MM:SS:MS Exact format is dependent on the DBMS being used. Use for business facing annotation not appropriate for DateTime.
Type	TYP	TP	Type code that is used in the XT (transformation process).
Video	VID	VD	Moving Image Object

Appendix B: Abbreviations & Acronyms

When table and column names need to be shortened, the Standard Abbreviations and Acronyms should be utilized for consistency. Entity and attribute names should be spelled out when possible.

Standard Abbreviations

Acronyms

There aren't hard, fast rules that will work in every instance when developing abbreviations for words. If there is not a standard abbreviation for a word that needs to be abbreviated, the following guidelines can be used to develop an abbreviation:

- Avoid abbreviating words that are 4 characters or less
- Words that are past tense should use the present tense
- Plural words should use the singular word
- Remove vowels as long as it doesn't become unrecognizable
- Remove unnecessary consonants, such as double consonants.

Words that are abbreviated should be uniquely abbreviated within a table, and preferably the entire data model. Do not use the same abbreviation for two different words.

Contact Enterprise Data Governance to have new abbreviations added to the Standard Abbreviation list.

4. Document Information

Exceptions

Standards are defined as rules that must be followed. If unique project requirements dictate deviations from the standard, the deviation should be documented in an Data Decision Document for the project.

If multiple exceptions are granted over time for a given standard, the Enterprise Data Governance should be contacted for possible change to the Standard.

Document History

Version	Date	Author	Reviewer	Purpose	Status
0.02	8/7/2008	Data COE Data Modeling Working Group	Data COE Data Modeling Working Group	Review	Shared
0.03	8/18/2008				
0.04	8/22/2008				
0.05	8/28/2008				
0.06	9/15/2008				
0.07	9/23/2008				
0.08	9/29/2008				
0.09	10/14/2008				
0.10	11/13/2008				
0.20	11/17/2008	Chris Burns	Data COE	Review & Approve	Shared
Enterprise Version 1.00	11/20/2008			Approved	Published
1.1	4/28/2015	Debby Cada	DMUG	Modified Name to not conflict with EDG Standard and updated class word list	Published

Data Modeling Standards

Approvals and Contributors

Approvals

The following individuals have reviewed, contributed and approved the contents of this document:

Name	Role
Linda Macklin	Data Modeling Working Group Team Member
Felix Dashevsky	Data Modeling Working Group Team Member
Padmini Narayan	Data Modeling Working Group Team Member
Deb Cada	Data Modeling Working Group Team Member
Chris Burns	Data Modeling Working Group Team Member
Gauri Dandekar	Data Modeling Working Group Team Member
.....
Dave Jones	Data COE Contributor/Reviewer
Sandi Rodgers	PCIT Reviewer
Sujatha Shiragombi	PCIT Reviewer
Tobi Zappe	Data COE Contributor/Reviewer
Jeff Kanel	Data COE Contributor/Reviewer
Data COE	Data COE Reviewer & Approver

Data Modeling Standards

Standards Currency: location, accessibility, revisions

This document is located on the <Data Center of Excellence website>. Please visit the website at: <http://wiki.nwideweb.net/wiki/display/dcoe/.Best+Practices> for the latest version.

All working documents and all revisions are located in the Data Center of Excellence Data Modeling Standards and Guidelines Sharepoint.

Comments