# COMMONWEALTH OF PENNSYLVANIA

# DEPARTMENT’S OF HUMAN SERVICES, INSURANCE, AND AGING

# INFORMATION TECHNOLOGY STANDARD

|  |  |
| --- | --- |
| Name Of Guideline: | Number: |
| **Data Modeling** | **STD-DMS004** |
| Domain: | Category: |
| **Data** | **Database Modeling** |
| Date Issued: | Issued By Direction Of: |
|  | C:\Users\wtorres\Desktop\CliffSignature (3).jpg |
| Date Revised: |  |
| **07/12/2016** | Clifton Van Scyoc, Dir of Division of Technical Engineering |

**Abstract:**

The purpose of this document is to detail physical and logical data model standards as well as provide data model, table, and column naming conventions and the associated data type.

**General:**

“A **data model** is an abstract model that describes how data is represented and used.” – *Wikipedia.com*

A logical data model represents the relationships and characteristics of data in a system. A physical data model is the physical representation of a logical data design. Whereas a logical data model can be platform neutral, a physical data model is based on the functionality of the specific database management system (DBMS) in which it will be implemented. The physical design should attempt to optimize the functionality and minimize the constraints of the DBMS.

All data models require the approval of the Data Administration (DA) and Database Design (DBD) Units before the database is migrated beyond the development environment. In order to facilitate rapid application development, it is suggested that Data Administration and Database Design participate in the development process. This allows suggestions in the use of existing structures, incorporating database standards in the initial design, and eliminating unnecessary rework after models **and** applications are developed in order to retrofit department standards.

This document contains DHS data modeling and naming standards used within the Bureau of Information Systems and supplements the data modeling standards contained in the Governor’s Office of Administration/Office of Information Technology (OA/OIT) Information Technology [IT Policies](http://www.oa.pa.gov/Policies/Pages/itp.aspx).

[ITP-INF003](http://www.oa.pa.gov/Policies/Documents/std_inf003a.pdf) - Data Modeling Standards

Policy Supplements:

[STD-INF003A](http://www.oa.pa.gov/Policies/Documents/std_inf003a.pdf) - Data Modeling Product Standards

[STD-INF003B](http://www.oa.pa.gov/Policies/Documents/std_inf003b.pdf) - Data Modeling Basics

[BPD-INF003C](http://www.oa.pa.gov/Policies/Documents/bpd_inf003c.pdf) - Data Modeling Best Practices

[BPD-INF003D](http://www.oa.pa.gov/Policies/Documents/bpd_inf003d.pdf) - Core Citizen Data Model and Data Elements

**Standard:**

**Data Model Standards**

A data model must comply with the following standards:

* Models will be constructed using CA ERwin Data Modeler. The Bureau of Information Systems currently has a combination of dedicated desktop and 6 Windows virtual servers with CA ERwin Data Modeling software installed. Check with Database Design to see what server is assigned to your work group.
* Models will be constructed using the standard ERwin template. Using the standard template allows the model to be created with the proper User Defined Properties (UDP) and connects the logical and physical models allowing updates to the model to be applied to both logical and physical models.
* Models contain both the logical and physical data model diagrams.
* Models follow the naming conventions for model, table, and column names for physical data models. This is specified in further detail below.
* All primary and foreign keys must be identified.
* Relationships, including identifying/non-identifying, cardinality, and optionality must be specified for each entity and attribute.
* Table and column descriptions and other metadata as required by the Database Management Section (DMS) must be contained within the model.
* The physical data model and the physical database must be synchronized prior to implementation in the production environment.
* When revising a previously approved model, such as one from an existing system, identify new tables by shading the new table. Columns in existing tables which are undergoing changes such as additional columns, columns for deletion, or modified columns should be in a contrasting color.

**Justification for Data Model Naming Standards and Conventions**

Consistent naming practices are vital. The use of a common methodology when naming database entities and attributes promote the uniform description and definition of the elements while, at the same time, facilitates:

* The reduction of redundant data elements by identifying like elements
* The reduction of element documentation by utilizing existing documents
* The promotion of data integrity by ensuring data singularity and consistency
* The simplification and standardization of system documentation by utilizing existing documents
* Consistency in the use and meaning of abbreviations

The Data Administrator owns and maintains the Standard Abbreviations list. If words over **six** characters are used, consult the Standard Abbreviations list. If an abbreviation does not appear in the list, spell the word out and the data administrator will assign the official abbreviation and include it in the list.

**Data Model Naming Standards**

Standards have been established for naming data models so as to easily identify the data model, data model subsystem and release number. Name data models using the following guidelines:

* The data model name is the application name or acronym.
* Do not include PDM (physical data model) in the data model name.
* Names with more than one word are separated by underscores. The additional words are not abbreviated. Ex. OLTP\_CAS\_SECURITY\_WO6040\_R1V01
* If a Work Order or Work Initiative exists, add the WO or WI number to the end of data model name, but before the Release and Version number. Do not use # after WO or WI. Ex. EDW\_CAPS\_TMSIS\_WO579\_R3V02.erwin
* The data model name must be appended by an underscore, the Release(R) and Version(V) number at the end of the data model name. This must match the release and version number on the Database Services Request Form.
* Additional words in the name representing a subsystem of the data model precede the WO or WI and release number. Ex. OLTP\_HCSIS\_PROVIDER\_QUALIFICATION\_WO1325\_R3V01
* There is no limit for the length of a data model name; however, consider the application or schema name and the lengths permitted within each database management system for database/schema names. Oracle relational database management system (RDMS) limits a database name to eight characters. MS SQL Server limits a database name to 128 characters, but a very long name is cumbersome.
* Platforms and database management systems are not transparent in the data model name; whereas, the type of processing is indicated in the data model name as follows:
	+ OLTP, online transaction processing, data models contain OLTP as a prefix in the name. Ex. OLTP\_eCIS\_\_1095B\_FORM\_ WO1020\_R25V03.erwin is an OLTP data model.
* EDW, enterprise data warehouse, data models including ODS are preceded by EDW\_.

 Ex. EDW\_HCSIS\_WO1166\_R3V01.

**Table Naming Conventions**

Name database tables using the following guidelines:

* All Enterprise Data Model tables have a prefix of “T”.
* Append a capitalized, abbreviated description (read left to right) of the object to the “T\_”. The table name is singular in nature (T\_AUDIT instead of T\_AUDITS) and uses abbreviations found in the Standard Abbreviations list. Separate all abbreviations by an underscore (“\_”) with a maximum total length of 30 characters. The maximum length of 30 includes underscores. Unofficial abbreviations are not to be used to constrain total length to 30 characters.
* Main frame Relational Data Management System (RDMS) tables should have a prefix of TXXX, where XXX is a three-digit reference number assigned by the database administrator (DBA).
* Main frame Data Management System (DMS) records contain a prefix of RXXX, where XXX is a three-digit reference number assigned by the DBA.

**Column Naming Conventions**

Name database columns using the following guidelines:

* The appropriate class word as listed in Table 1 below.
* A combination of descriptors in the appropriate order which identify the data contained within the column. These descriptors consist of either abbreviations found on the Standard Abbreviations list (<https://dhhsintra.dpw.state.pa.us/datadictionary/>) or words under seven characters long. These components comprise a complete, descriptive name. The column is named with the most generic descriptor first, followed by more specific components. This methodology typically makes the element name read right to left, and allows for easy grouping of like attributes in the data dictionary.
* The column name has a maximum length of 30 characters. Unofficial abbreviations are not to be used to constrain total length to 30 characters. When deriving a column name, please consult the data dictionary for existing names that will suit your need.

An example would be: CDE\_INCRD\_EXP\_TRANSP

The attribute contains the Class Word (CDE) and a description that defines the object. Read as: “This column is a code column that identifies whether something was incurred. It further defines whether an expense was incurred (more specific), and finally determines if a transportation expense was incurred (most specific).”

**Class Words**

Class words classify data in its generic sense, i.e., by its identity or its nature. The function of data is not relevant in this identification process, since the purpose is to describe what it is and not how it is used. See **Table 1** for class words and their definitions. Note that class words are always used in their abbreviated form.

**Table 1**

|  |  |  |  |
| --- | --- | --- | --- |
| **Class Word** | **Classification Category** | **Definition** | **Data Type** |
| ADR | Address | Address data that identifies a specific geographic location, URL or e-mail address. | Character |
| AMT | Amount | A monetary quantity | Number |
| AUD | Audio | Audio or sound related data in the column | As defined by DA |
| CDE | Code | Data that is an encoded representation of a value | Alphanumeric |
| CNT | Count | A non-monetary quantity | Number |
| CTL | Control Data | Control data used to coordinate processing in a system, application, or component  | As defined by DA |
| DOC | Document | Entire forms or documents within the column | As defined by DA |
| DTE | Date | A calendar date | As defined by DA |
| IDN | Identifier | Data that serves as a label or identifier for other data. Data generated by a system or algorithm, as well as manually-generated data, is named using the IDN class word. | As defined by DA |
| IMG | Image | Pictures, video, etc. | As defined by DA |
| IND | Indicator | Binary element that defines a yes/no condition. Any pair of values may be used, but the data must be binary. For examples, the IND class word would be appropriate if the allowed values include only “Y” and “N;” it would also be appropriate if only “0” and “1” are allowed. However, the IND class word would not be acceptable if the column may contain the three values “Yes,” “No,” and “Unknown” because these are not binary". | Alphanumeric |
| NAM | Name | Data that identifies someone or something. | Character |
| NBR | Number | Although IDN is the class word normally used for identifiers, some identifiers, such as social security number, are commonly called “numbers” in common English usage, and in these cases NBR may be used as the class word. | Number |
| PCT | Percentage | Percentage ratio between other data values. The value is stored as a decimal value in the database (i.e., 6% is stored as .06) | Number |
| TME | Time | An instant in which something occurs, or duration of time. | As defined by DA |
| TXT | Text | Data having undefined content and/or unpredictable structure. | As defined by DA |

**Exemptions from this Standard:**

There will be no exemptions to this standard.

**Refresh Schedule:**

All standards and referenced documentation identified in this standard will be subject to review and possible revision annually or upon request by the DHS Information Technology Standards Team.

**Standard Revision Log:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Change Date** | **Version** | **Change Description** | **Author and Organization** |
| 05/10/01 | 1.0 | Initial creation. | Deloitte Consulting |
| 07/25/03 | 1.1 | Changed the definition of class word IDN, NBR and ADR. Changed ADR’s data type. | Dale Woolridge |
| 05/26/04 | 1.1 | Reviewed for content – No change necessary | Dale Woolridge |
| 06/30/04 | 2.0 | Content added to document | Brian Mains |
| 01/15/08 | 2.1 | Reviewed for content and formatted for new standard | DMS |
| 06/24/10 | 3.0 | Reviewed and updated. | L. Steele and P. Gillingham |
| 07/12/016 | 4.0 | Reviewed and updated. | DMS |