

COMMONWEALTH OF PENNSYLVANIA HEALTH & HUMAN SERVICES DELIVERY CENTER

INFORMATION TECHNOLOGY STANDARD

Name of Standard: Data Modeling	Number: STD-DMS004
Domain: Data	Category: Database Modeling
Date Issued: 05/10/2001	Issued by Direction of:
Date Revised: 04/07/2020	Jon Arnold, Chief Technology Officer Health & Human Services Delivery Center

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 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 Health and Human Services (HHS) Technology Services Office (TSO) Database Service Delivery and Service Operations 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 to retrofit department standards.

This document contains the Health and Human Services Information Technology Delivery Center (HHS IT DC) data modeling and naming standards that supplements the data modeling standards contained in the Governor's Office of Administration/Office of Information Technology (OA/OIT) Information Technology IT Policies.

[ITP-INF003](#) - Data Modeling Standards

Policy Supplements:

[STD-INF003A](#) - Data Modeling Product Standards

[STD-INF003B](#) - Data Modeling Basics

[BPD-INF003C](#) - Data Modeling Best Practices

[BPD-INF003D](#) - Core Citizen Data Model and Data Elements

Standard:

At this time there are two different data model standards for the Department of Human Services (DHS) and the Department of Health (DOH).

Department of Health (DOH)

Database Standard Data Elements & Abbreviations

Standard Data Elements

Standard data elements are created to achieve consistency of names, datatypes, lengths, etc. throughout agency databases. This will benefit the sharing of data between agencies and applications. To set a standard, the database elements and the data dictionary are researched to determine the most used or most accurate element names and attributes.

- Standard elements take precedence over using any abbreviation.
- Prefixes can be added to the standard element upon consulting with a DBA.
 - Example: FirstName could become ClientFirstName or ContactFirstName

Abbreviations

The DOH training department maintains [an acronyms list](#) on the DOH SharePoint site. In lieu of a database naming standards abbreviation list, the agency list can be referenced. Some acronyms, however, are not unique; and a DBA must be consulted for reasonable database abbreviations.

- Abbreviations should be avoided whenever possible.
- Abbreviations should not be used for words with six or less characters. If words over six characters are used, and an abbreviation is required, consult with a DBA.
- To set a standard database abbreviations list, abbreviations are researched and an attempt shall be made to keep the list to a minimum. Page Break

Reserved Words

Microsoft SQL Server uses reserved keywords for defining, manipulating, and accessing databases. Reserved keywords are part of the grammar of the Transact-SQL (T-SQL) language that is used by SQL Server to parse and understand T-SQL statements and batches. Although it is syntactically possible to use SQL Server reserved keywords as identifiers and object names in T-SQL scripts, you can do this only by using delimited identifiers. For more information, go to <https://msdn.microsoft.com/en-us/library/ms189822.aspx>

Naming Conventions

Best Practices – Database Objects

A database object is a database component like a table, trigger, view, key, constraint, default, user defined datatype or stored procedure in a database. The following naming conventions apply for all database objects:

- The most important rule for naming is that names must be significant and meaningful. A name should give an idea of the usage of code objects or contents for collection objects.

- Naming should be consistent.
- Names must begin with a letter followed by letters, numbers or the underscore.
- For readability, names shall utilize both upper and lower case. Utilizing the camelback style can enhance readability and eliminate a need for the underscore character between words.
Example: PurchaseOrderLineItem instead of purchaseorderlineitem or Purchase_Order_Line_Item
- Prefixes/suffixes denoting object type are not necessary except where noted in this document. Object types should be identifiable by usage/context. Redundant prefixes/suffixes reduce available string length. · Names should not be reserved words, keywords (Appendix A) or include spaces or special characters. Use of the underscore character is an exception when elements of a name must clearly be separated.
- Abbreviations may be used as part of a name. The list of standard abbreviations, however, should be checked prior to usage. Example: Use NUM instead of NO. See later section about abbreviations.
- Try to keep names of database objects as short as possible while still observing the previous best practices. Consider front- or back-end user interfaces with stricter limits:
 - SAS EG (Statistical Analysis System Enterprise Guide) has a limit of 32 characters.
 - Oracle has a limit of 30 characters.

Department of Human Services (DHS)

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 TSO Service Strategy 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 TSO Service Strategy 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/DataAbbr.aspx>) 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	Alphanumeric

		"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".	
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 HHS Information Technology Delivery Center Domain Leads.

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
04/07/2020	4.1	Reviewed and updated.	Glenn McDonel PGillingham