# COMMONWEALTH OF PENNSYLVANIA

# HEALTH & HUMAN SERVICES DELIVERY CENTER

# INFORMATION TECHNOLOGY GUIDELINE

|  |  |
| --- | --- |
| Name Of Guideline: **Data Administration** | Number: |
| **Submission Best Practices** | **GDL-DMS001** |
| Domain: | Category: |
| **Data** | Database Modeling |
| Date Issued: | Issued By: |
| **9/19/2003** | **Glenn McDonel, Data Administration, Service Strategy Unit** |
| Date Revised: |
| **6/4/2020** |

**Abstract:**

The Health and Human Services Delivery Center (HHS DC) Technology Service Office (TSO) directs the management of information for the Departments of Health (DOH), Human Services (DHS), Aging (PDA), Drug and Alcohol Programs (DDAP) and Military and Veterans Affairs (DMVA).

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:**

Data Administration has an extensive review process for reviewing the structure of a database. The approved routing process starts with a Logical Data Model (LDM) and Physical Data Model(PDM) review meeting. These meetings consist of representatives from Database Operations, Data Administration Unit, and the Application Developer Team***.***  This group reviews andanalyzes the database structure and design for compatibility and performance enhancements.

Data Administration reviews the model, verifying the accuracyof names, data types, lengths, precisions, descriptions, and business information thoroughly. Any concerns with the model are directed back to the contact person, and through collaborative effort, the modelis finalized, and development may begin. After Data Administrationapproval, the Database Administrators move the request through the appropriate environments.

The purpose of this document is to focus on the requirements of the Data Administration review process. The review process starts with the Packet submission and ends with Data Administration’s approval of the data model.

Please note for Mainframe DMS readers, this document uses the term “table” for a record, and “column” for a data item/field. This is relational database terminology compatible with the Oracle, SQL Server, and Mainframe RDMS database management systems.

The DHS Intranet website is found at [**http://mydhs.dpw.lcl/**](http://mydhs.dpw.lcl/) or [MyDHS Website](http://mydhs/).

On the left scroll down to HHS IT Delivery Center, move mouse to right to Business and Technical Standards then to Data Domain or use the hyperlink: <http://mydhs.dpw.lcl/hhs-it-dc/busandtechstandards/datadomain/index.htm>

**Submission Documentation**

## Open Systems

1. [Request for Database Services Form](http://mydhs/cs/groups/webcontent/documents/form/p_031710.doc) found in Service Now > Catalog Health and Human Services > Database > Database Requests > Database Services

2. CA ERwin data model including the meta data.

## Mainframe Systems

1. [Request for Database Services Form](http://mydhs/cs/groups/webcontent/documents/form/p_031710.doc) found in Service Now > Catalog Health and Human Services > Database > Database Requests > Database Services

2. [Spreadsheet Template](http://hsavsatweb100/datadictionary/Document.aspx).

The spreadsheet template is used for mainframe requests and can be found at:

<https://dhhsintra.dpw.state.pa.us/datadictionary/Document.aspx>

**Request for Database Services Form**

The Database Services Request (DBSR) is used for database structure changes in addition to other database services. Please follow [Instructions for Completing a Database Request](http://mydhs.dpw.lcl/cs/groups/webcontent/documents/document/p_031711.docx)

**Data Models**

ERwin data models are required for open systems database structure changes. The [template for the ERwin data model](http://hsavsatweb100/datadictionary/Document.aspx) contains all necessary User Defined Properties (UDP) required by Data Administration. To obtain an ERwin data model template contact a member of Data Administration. **The complete system data model is required, not just areas being modified.**

**Spreadsheet Template**

This template is only required for all Mainframe submissions.

**Hard Copy Submissions**

All forms are required to be in electronic format. Hard copy submissions are no longer being accepted.

**Naming Convention Standards**

Standards for naming conventions have been developed for table and column names. Please reference the [Data Modeling Standards](http://mydhs/cs/groups/webcontent/documents/document/p_032212.docx) for in-depth details regarding these standards. In general, table names consist of T\_ and a descriptive name. Column names consist of a class word, underscore, and a descriptive name which may consist of any combination of words less than seven characters, underscores, and abbreviations with a total length not to exceed 30 characters. These standards have evolved over time; many existing names would not be allowed under the current standards. For existing tables, a request submitted to add a name that does not meet the current standards will be reviewed by Data Administration on a case-by-case basis. This situation may be permitted for tables that are legacy or import tables where a need exists to match a table with nonstandard names. In most cases, however, if a new table is being created, the names must meet the current standards.

Due to the high volume of certain column names, some common column names have been standardized. These columns and their uses are listed below.

**Common Column Names**

* **Dates**
* **DTE\_BEGIN, DTE\_END –** The begin date and end date column names are used when there are begin and end dates for an event (for example, the begin date and end date that a current driver’s license is in effect). Oftentimes, the terms start or effective will be used in place of begin. However, the preferred terms are begin date and end date.
* **DTE\_CHANGE\_LAST –** The last change date column name represents the date an item was updated or changed.
* **DTE\_CREATN –** The creation date column name represents an item’s creation date. This is typically used to track the creation of an entry in the database. It can also be used to track the creation of an object, such as a record. DTE\_CREATN should be used over DTE\_CRTD for consistency. However, DTE\_CRTD will be allowed if previously used.
* **DTE\_DOB, DTE\_DOD –** When using a date of birth or date of death, the standard is to use the abbreviations created for these meanings. The terms BIRTH or DEATH are not spelled out.
* **Identifiers**
* **IDN\_USER –** In cases where the user name column was referring to a user system identifier (such as Commonwealth of Pennsylvania (CWOPA) Identifier) or an application user ID, it was standardized in the Master Provider Index (MPI) system that IDN\_USER was the preferred terminology for this column.
* **IDN\_USER\_CHANGE\_LAST –** Whenever storing the user ID of the person who made the most recent change, the field IDN\_USER\_CHANGE\_LAST should be used for the column name. This name is widely used across all environments.
* **Names**

Use the **NAM** class word for text values

**NAM\_FIRST, NAM\_LAST, NAM\_MI –** The standard for first and last names, as well as middle initials, is to use NAM\_FIRST, NAM\_LAST, and NAM\_MI followed by any additional abbreviations that are needed after it (i.e. NAM\_FIRST\_OFCR, NAM\_FIRST \_WORKER, etc.).

* **Codes**

Application reference tables usually have a code and value pairing (as well as a few other fields at times). When creating a reference table, use the **CDE** class word for the coded value.

* **CDE\_TYPE, CDE\_STATUS –** When using a code that is a type or status code, always put the words type and status next to the class word. This is an effective way to arrange type and status codes among multiple systems.
* **Indicators**
* **IND\_RECORD\_DELTD, IND\_RECORD\_DELETE\_LOGCL –** When using an indicator that is a flag stating the current record has been deleted, use one of these names to represent it. These two fields are widely represented across the systems to track whether a record has been marked deleted in a database.
* **Address Information**

For systems that collect postal and electronic mailaddress information, the columns for collecting it have been standardized:

* **ADR\_LINE\_1 –** first line of the street address.
* **ADR\_LINE\_2 –** second line of the street address.
* **ADR\_LINE\_3 –** third line of the street address.
* **ADR\_CITY –** The city of the address.
* **ADR\_STATE –** The state abbreviation for the address.
* **ADR\_ZIP\_MAIN –** The main zip code 5-digit number of the address.
* **ADR\_ZIP\_EXTN –** The zip code 4-digit extension of the address.
* **ADR\_EMAIL –** The electronic mailaddress column name.
* **CDE\_COUNTY –** County codes or Federal Information Processing Standard (FIPS) code are code values specifying the county.

Please consult with Data Administration or the [Data Dictionary](http://hsavsatweb100/datadictionary/Accepted.aspx) if there are other fields that are needed relating to postal information. Other address information exists within the data dictionary, such as Latitude, Longitude, etc.

* **Numerical Values**

When using numerical values in column names, the usual construct is to place the numerical value, such as CDE\_CAT\_ASST\_OTHER\_1, at the end of the name.

* **Roman Numeral Values**

Sometimes, a roman numeral may be desired. Roman numerals can be easily disguised as an abbreviation and this value should be converted to its numerical equivalent. However, if the database is modeling an official form or other source which includes a Roman numeral, a Roman numeral may be substituted. Please consult with Data Administration if this is the case.

**Length Best Practices**

Certain column names have a standard length to improve compatibility among systems. Varying from this length requires that the reason be documented within the comment section and benchmark tests performed to ensure that the data needs a nonstandard size. The following column names have a standardized length:

**Common Column Lengths**

* **Address Information**
* **ADR\_LINE\_1 –** The lines of an address are a standard length of 26 characters.
* **ADR\_LINE\_2 –** The lines of an address are a standard length of 26 characters.
* **ADR\_LINE\_3 –** The lines of an address are a standard length of 26 characters.
* **ADR\_CITY –** The city column name is a standard length of 23 characters.
* **ADR\_STATE –** The state address has a standard length of 2 characters which is the maximum length of the encoded value.
* **ADR\_ZIP\_MAIN –** The main zip code has a length of 5 digits.
* **ADR\_ZIP\_EXTN –** The zip code extension has a length of 4 digits.
* **ADR\_EMAIL –** The electronic mail address column name is a standard length of 100 characters.
* **CDE\_COUNTY –** County codes are two-digit county code values specifying the county. The maximum length allowed is 2. When the Federal Information Processing Standard (FIPS) code is used, the standard length will be a value of 3.
* **Identifiers**
* **IDN\_USER –** The user ID field has a standard length of 12 which is the number of characters that the Department of Human Servicesallocates towards user names in systems. If the user ID is another type of ID, such as an application user ID, it can vary from this standard.
* **IDN\_USER\_CHANGE\_LAST –** The last change user ID field will match the IDN\_USER length which has a standard of length of 12 in most instances.
* **Names**
* **NAM\_FIRST –** The length of the first name is standardized at 32 characters.
* **NAM\_LAST –** The length of the last name is standardized at 32 characters.
* **NAM\_MI –** The length of the middle name are standardized at 1.
* **Numbers**
* **NBR\_SSN –** The social security number has been standardized at a length of 9 characters or digits.

**Documentation Best Practices**

**Description/Business Rule**

Business Rule User Defined Properties (UDP) is used to contain the Work Order/Work Initiative number for the packet.

The Description is contained within the ERwin Comment Property for tables and columns. The Comment Sectionis adeclarative expression of the business policies that are to be enforced.

The Comment Section is a very descriptive explanation of the Table/Column. It is to be phrased in proper English, as complete sentences, and all words are to be spelled correctly. All documentation contained within the Comment Sectionshould be prepared so the general population can understand it. If the documentation contains technical terms, these terms must be defined, and any acronyms in the documentation must be spelled out. For fields that have code/value pairs, every code value must be present or a description of the source (a reference table or an external file) must be well documented, an additional spreadsheet may be included with the ERwin model for codes and values.

The Comment Sectionconsists of any constraints and calculations performed on the table or column at the database or application level. For Example:if DHSis part of the name, spell it out followed by the acronym in parentheses, Department of Human Services (DHS).

**English Names**

These are full English names that represent the table and column abbreviated name~~s~~ that they describe. They are entered on the English Name UDP for tables and columns in ERwin.

**Specialized Terms/Acronyms**

Only Data Administration approved acronyms are acceptable in the table and column names. Please refer to the BIS intranet site under Business and Technical Standards/Data Domain***/***Standards***/***[Data Administration Standards](https://dhhsintra.dpw.state.pa.us/datadictionary) for the standardized abbreviations to ensure proper use of acronyms.

The [Abbreviation Search page](http://hsavsatweb100/datadictionary/DataAbbr.aspx) contains a current listing of Authorized Abbreviations, acronyms and class words. If an abbreviation is not found, please use the full word in your model and submit an abbreviation request to Data Administration for approval.

**Default Values**

Please add Default Value to the UDP tab for all Columns. For already existing tables, if you are adding a column and you indicate a Not Null in the Null Option, please provide a default value. You will also need to provide a default value when you are changing a column’s Null option from Null to Not Null. If it is identified in the description, you can leave it there, but you will also need to identify it in the default value line in the UDP tab.



**Full Path**

Full paths including Schema.TableName.ColumnName are only required in descriptions for Class Words IDN, CDE and DTE.  Class Word TXT descriptions do not require a full path and the textual description does not need to be in the description.  If citing a table or column with Class Word TXT, using the table/column English Names will be sufficient.

**Additional Reading**

To find out more information about the standards listed above, please consult the BIS intranet site, select Business and Technical Standards, then [Data Domain](http://mydhs/oa/bis/busandtechstandards/datadomain/index.htm). This section lists all of the standards that are needed to complete a request for a data modelsubmission.

**Refresh Schedule:**

All guidelines and referenced documentation identified in this standard will be subject to review and possible revision annually or upon request by the Health and Human Services Delivery Center Standards Team.

**Guideline Revision Log:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Change Date** | **Version** | **Change Description** | **Author and Organization** |
| 9/19/2003 | 1.0 | Initial creation of the document under the name “H-Net Data Dictionary Naming Guidelines.” | Brian Mains |
| 06/14/2004 | 1.1 | Document renamed to “H-Net Data Dictionary Submission Best Practices.” | Brian Mains |
| 06/24/2004 | 1.2 | Document content revised. | Dale Woolridge |
| 05/17/2006 | 1.3 | Added additional content to document. | Brian Mains |
| 01/12/2007 | 1.4 | Edited content | L. Steele |
| 07/17/2007 | 1.5 | Edited format and new procedures | L. Steele |
| 06/02/2008 | 1.6 | Changed fields w/ user id from 10 to 12 | L. Steele |
| 03/19/2010 | 1.7 | Document reviewed and content revised due to procedural changes. | L. Steele and Patty Gillingham |
| 12/04/2013 | 1.8 | Revised content | W. Hamel, G. McDonel, & K. Romanowski |
| 7/28/2016 | 2 | Renamed Guideline and Revised content | Data Administration Unit |
| 1/30/2020 | 2.1 | Added Default Values and Full Path guidelines. | Glenn McDonel |
| 5/4/2020 | 2.2 | Formatting and content updated. | Glenn McDonel |