

Mecklenburg County

**Data Modeling Standards & Guidelines**

**Draft Version 0.7**

Contents

[Mecklenburg County 0](#_Toc473280215)

[**1** **Overview** 3](#_Toc473280216)

[1.1 Introduction 3](#_Toc473280217)

[1.2 Scope 3](#_Toc473280218)

[1.3 Purpose of Data Model 3](#_Toc473280219)

[1.4 Benefits 3](#_Toc473280220)

[**2** **Data Modeling Concepts** 4](#_Toc473280221)

[2.1 High Level Data Model Definition 4](#_Toc473280222)

[2.1.1 Data Model Types 4](#_Toc473280223)

[2.1.2 Entities 4](#_Toc473280224)

[2.1.3 Relationships 5](#_Toc473280225)

[2.1.4 Attributes 5](#_Toc473280226)

[2.1.5 Class Words 5](#_Toc473280227)

[2.1.6 Data Domain 5](#_Toc473280228)

[**3** **Data Modeling Standards and Guidelines** 6](#_Toc473280229)

[3.1 Common Standards and Guidelines 6](#_Toc473280230)

[3.1.1 Naming Standards 6](#_Toc473280231)

[3.1.2 Definition Standards 6](#_Toc473280232)

[3.1.3 Model Template Usage 6](#_Toc473280233)

[3.1.4 Model Naming 6](#_Toc473280234)

[3.1.5 Model Diagram Naming 6](#_Toc473280235)

[3.2 Subject Area Model Standards and Guidelines 7](#_Toc473280236)

[3.2.1 Development Standards 7](#_Toc473280237)

[3.2.2 Definition Standards 7](#_Toc473280238)

[3.2.3 Naming Standards 7](#_Toc473280239)

[3.3 Conceptual Data Model Standards and Guidelines 7](#_Toc473280240)

[3.3.1 Development Standards 7](#_Toc473280241)

[3.3.2 Definition Standards 7](#_Toc473280242)

[3.3.3 Naming Standards 7](#_Toc473280243)

[3.4 Logical Data Model Standards and Guidelines 7](#_Toc473280244)

[3.4.1 Development Standards 7](#_Toc473280245)

[3.4.2 Definition Standards 8](#_Toc473280246)

[3.4.3 Naming Standards 8](#_Toc473280247)

[3.4.4 Datatype Standards 9](#_Toc473280248)

[3.5 Physical Data Model Standards and Guidelines 9](#_Toc473280249)

[3.5.1 Development Standards 9](#_Toc473280250)

[3.5.2 Definition Standards 9](#_Toc473280251)

[3.5.3 Naming Standards 9](#_Toc473280252)

[3.5.4 Datatype Standards 9](#_Toc473280253)

[**4** **Appendix** 10](#_Toc473280254)

[4.1 Standard Datatype Domains 10](#_Toc473280255)

[4.2 Standard Abbreviations 11](#_Toc473280256)

[4.3 Standard Class words 12](#_Toc473280257)

[4.4 Acronyms 13](#_Toc473280258)

[4.5 Data Model References 13](#_Toc473280259)

# **Overview**

## Introduction

A Data Model is an abstract model that organizes elements of data and standardizes how they relate to one another. The use of standardized data facilitates data sharing, reduces data handling costs and leads to better data accuracy, consistency and timeliness.

## Scope

The scope of this document is to:

* Define the data modeling standards and guidelines including naming conventions which will help emerging and current applications develop and maintain their physical data models in a consistent manner.

## Purpose of Data Model

The data models are developed because of what it represents and should achieve:

* A means of graphically depicting how data can be shared and stored non-redundantly
* A way of defining data requirements in accordance with a set of rules and common methodology
* Means of defining and certifying business data requirements
* Provides the data structure definition for a project.
* The integrator for all data sharing, master data, and data warehouse efforts
* A basis for database design

## Benefits

* Improved data quality
* Accelerated development
* Better documentation
* Reduced cost and maintenance

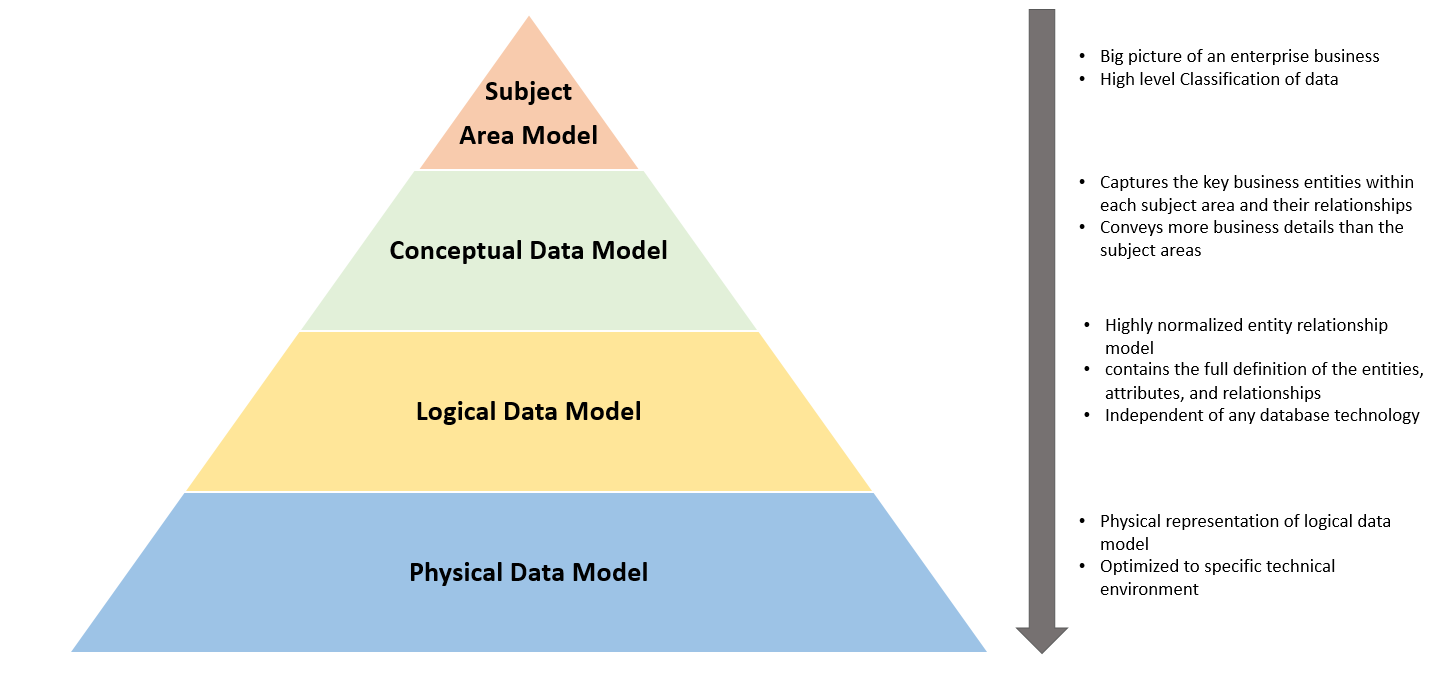
# **Data Modeling Concepts**

## High Level Data Model Definition

This section defines the generic concepts pertaining to Data Modeling which will help the reader of the document to understand the standards and guidelines cited in the consecutive sections.

### Data Model Types

The diagram below shows the different types of data model and how to approach its creation from enterprise perspective.



1. **Subject Area Model**

Subject area is a high-level classification of data representing a generic business concepts pertaining to an organization.

1. **Conceptual Data Model**

An Enterprise Conceptual Model (ECM) is the second level of the Enterprise Data Model (EDM), created from the identification and definition of the major business concepts within each subject area. It identifies the major entities and relationships of importance to business.

1. **Logical Data Model**

A logical data model contains the full definition of the entities, attributes, and relationships needed to satisfy the business and technical requirements for data

1. **Physical Data Model**

A physical data model is the physical representation of the logical data model, with any added physical constructs necessary to build database. The physical data model defines the tables (from entities), columns (from attributes), and keys, foreign keys, and constraints (from relationships) and indexes necessary to find and retrieve data.

### Entities

Entities are the principal data object about which information is to be collected. Entities are

usually recognizable concrete, tangible or abstract concepts, such as person, places, things, or

events important to the organization’s business. Some specific examples of entities are PERSON, ORGANIZATION, or SERVICES.

### Relationships

A Relationship represents an association between two or more entities. An example of a

relationship would be: “Person requesting Service”. Relationships can be either one-to-one,

one-to-many, many-to-many, zero-to-one, or zero-to-many. Relationships between entities

are named, which defines the purpose of the relationship.

### Attributes

Attributes describe the entity of which they are associated.

### Class Words

A class word designates the category of data into which a data element fits. Examples of class words are

“Code,” “Name,” and “Quantity”. A complete list of class words is available in Appendix.

### Data Domain

A data domain specifies the type of data that an attribute can have. It can be default domains such as string or Number or user defined domains. For example, for all name type of attributes, we can define the domain as “Name” with the datatype of varchar(50).

# **Data Modeling Standards and Guidelines**

## Common Standards and Guidelines

### Naming Standards

Below are the common practices to be followed for naming all data model objects.

* In general, there are “business names” and “technical names” for data objects.
  + Business names are used in all conceptual and logical level objects
  + Technical names are used only for physical objects and are often generated from corresponding logical object using abbreviation list (when there is one).
* Business names should be fully spelled out English names with no delimiters and are Title Case (e.g.
* “Party Type Code”)
* Technical names can be:
  + Abbreviated names for any database technology that does not permit full names of at least 30 characters.
  + Composite English Names with underscores and all uppercase (e.g. “PARTY\_TYPE\_CD”)
* Object name should always be singular, commonly-used, understood and simple business name.
* Avoid use of organization names, application names, or project names when naming data objects. These are subject to change and the need for data usually outlives organization and applications.

### Definition Standards

Any name used in the data domains requires a corresponding definition. Below are the common practices to be followed for defining all data model objects.

* A definition should be unique and distinguishable from every other data element definition.
* A data definition should be precise, concise, and unambiguous. The definition should be clear enough to allow only one possible interpretation.
* The data definition should be expressed in singular.
* The definition should include the essential meaning or primary characteristics of the concept.
* The definition should only use abbreviations when necessary and the abbreviation must be commonly understood.

### Model Template Usage

Model templates are great way to share reusable standards across organization. Model templates provide the ideal way of maintaining a consistent look and feel across different models, as well as standards for naming, domains, data types, etc.

A model template is created with datatype domains and naming standards. It is available [here](Logical_Data_Model_Template.erwin_tmpl) and should be used for any new logical/physical data model creation.

### Model Naming

* Models are ideally named for what they are intended to represent and serve, such as business function, and not for the project acronym or name.
* Models have a complete definition that defines their scope and intent
* Below is the naming convention to be followed.
  + Enterprise <*Logical/Physical*> Data Model - <*Subject Area Name*>
  + Ex. Enterprise Logical Data Model - Party

### Model Diagram Naming

* Within the data model, an ER diagram needs to be created which contains the entities, relationships and attributes.
* The naming convention to be followed for Model diagram is:

<*Subject Area Name*>\_ER\_Diagram

Ex. Party\_ER\_Diagram

## Subject Area Model Standards and Guidelines

### Development Standards

* Subject area model should be developed working closely with business SMEs and should focus on enterprise business concepts.
* There should be *Only One* subject area model in an enterprise.
* There should not be more than 10 to 12 subject areas in the subject area model.
* Relationships between subject areas must represent significant business interactions and dependencies and should be displayed with verb phrase.

### Definition Standards

* Subject area definition should be as simple and as understandable as possible.
* As subject area definitions are viewed by entire organization so they should be relevant to the organization and not generic.

### Naming Standards

Subject area names should be very clear, concise, and comprehensive; ideally one word

## Conceptual Data Model Standards and Guidelines

### Development Standards

* Concentrating one subject area at a time, the CDM should start at a very high level, showing major entities and primary relationships.
* No attributes should be entered in the conceptual data model.
* Every major entity appearing on a conceptual data model should have at least one relationship to another entity.
* Relationships appearing on an entity diagram must clearly display a verb phrase and the direction in which the action of the verb applies.

### Definition Standards

* Every entity in the CDM should have meaningful definition and adhere to common definition standards and guidelines defined in section 2.3.2

### Naming Standards

* Conceptual entity names should be business oriented; not influenced by systems or applications
* The names should be as simple as possible, yet appropriately descriptive
* Abbreviations and acronyms should not be used

## Logical Data Model Standards and Guidelines

### Development Standards

* Every new logical model should be developed using [model template](Logical_Data_Model_Template.erwin_tmpl) which contains the recommended data standards for naming and defining the objects.
* Logical model should be fully attributed and normalized to Third Normal Form (3NF)
* All entities should be related using identifying or non-identifying relationships
* LDM should identify the primary key attributes in each entity and non-key attribute should be listed below that
* Code attribute in the logical model should be inherited as Foreign Key from their respective code entity

Ex. Gender Code in Person entity should be FK to Gender entity

* Each logical entity and attribute should be provided with meaningful definition.

### Definition Standards

* Every entity and attribute added in the LDM should have meaningful definition and adhere to common definition standards and guidelines defined in section 2.3.2

### Naming Standards

Entity / Attribute name in logical model should follow below standards.

* Must be unique within the model.
* Must be in Title case and only alphabets are allowed. No special characters.

Ex. Person, Organization

* Multipart names should be separated by spaces

Ex. Party Identification

* Must be singular noun or noun phrase.
* Entity names should follow the below format

|  |  |
| --- | --- |
| **PRIME WORD** | **MODIFIER** |
| *mandatory 1 word* | *optional up to 4 words* |

* Attributes names should follow the below format.

|  |  |  |
| --- | --- | --- |
| **PRIME WORD** | **MODIFIER** | **CLASS WORD** |
| *mandatory 1 word* | *optional up to 4 words* | *mandatory 1 word* |

Where,

PRIME WORD – a word that gives the general grouping or business context. Ex, Person, Party

MODIFIER – word(s) after the Prime Word and before the Class Word to provide the rest of the meaning.

CLASSWORD - word that identifies the general function or purpose of the data, e.g. DATE is used to give a point in time, and NAME is used to identify or classify data.

Ex. Party Identification Issue Date

Prime Modifiers Classword

Word

### Datatype Standards

* The datatypes of the attributes should be defined using the domains in the data model (see Appendix 4.1)
* Domain helps in standardizing the attributes which have common characteristic. For ex. All Name fields can have the datatype of VARCHAR(50) so defining a NAME domain with datatype of VARCHAR(50) will make sure all name fields have consistent datatypes

## Physical Data Model Standards and Guidelines

### Development Standards

* A certain amount of denormalization is usually necessary when implementing the physical data model. Denormalize only if you can demonstrate a performance gain.
* All code tables defined in logical model should be denormalized in physical model into single Code/Value table.
* All many-many relationships should be resolved by introducing bridge table

### Definition Standards

* Generally, all the tables and column in physical model should be inheriting the definitions from the logical model entity and attribute.
* In case the new tables added in physical model during de-normalization, it should have the definitions as per standards. Same is applicable for columns too.

### Naming Standards

Along with the common naming standards stated in section 2.3.1, below are some specific physical model standards that should be taken into consideration.

* Physical model should use the abbreviations (Appendix 4.2) for tables and column naming
* The standard abbreviations are available in the Naming Standard file and applied on the model template.
* Physical model should inherit the names from Logical Model via Naming Standard files. So it is important to use the model template for any new model creation as it has all Naming Standards applied and saved.
* Recommended to limit the name to 35 characters (shorter is better). Use abbreviations if the name is going beyond 35 via naming standard file.
* Table and Column names should be in all UPPERCASE.
* Use underscores for multipart names (ex. PARTY\_IDENTIFICATION)
* Use a letter as the first character
* Each column name should contain all of the elements of the logical attribute from which it was derived, but should be abbreviated to fit within the maximum length.

### Datatype Standards

Same as Logical Data Model

# **Appendix**

## Standard Datatype Domains

|  |  |  |  |
| --- | --- | --- | --- |
| **Domain Name** | **Domain Parent** | **Logical Data Type** | **Definition** |
| <default> | <root> | CHAR(18) |  |
| String | <default> | VARCHAR(20) | String is free form text with a combination of letters, numbers and characters |
| Number | <default> | INTEGER | Number is a numeric value that can be expressed as a decimal, exponent, positive or negative value, and against which one can perform arithmetic functions. It is a default domain. |
| Datetime | <default> | DATE |  |
| Blob | <default> | LARGE BINARY |  |
| Identifier | Number | INTEGER | Identifier is a whole number assigned to uniquely identify an instance within a designated entity. |
| Name | String | VARCHAR(50) | Name is the wording used to identify an Organization, Group or name of an Individual. |
| Date | Datetime | DATE | Date is a time metric that includes month, day and year only. It does not include time of day. |
| Indicator | String | CHAR(1) | Indicator is a one character code that notes if a condition is met with values of Y = Yes, N = No, blank = Unknown or Not Applicable |
| Status | String | VARCHAR(10) | Status of an attribute like Pending, Active, etc. |
| Description | String | VARCHAR(50) | Description is text that further defines the meaning and use of a code. |
| Code | String | VARCHAR(5) | Code is a value used to represent a condition |
| Id Number | String | VARCHAR(20) | ID Number is an alphanumeric identification number of an entity. |
| Short Desc | String | VARCHAR(20) | Short Description is a generic short text definition for an object |
| Text | String | VARCHAR(100) | Free form text with a combination of letters, numbers and characters |
| Short Name | Name | CHAR(3) | Short Name is used to define Prefix or Suffix Name which is a title or name extension. Examples: Mr., Mrs., Ms., Dr., Jr., Sr., III, M.D., PhD. |
| Count | Number | INTEGER | A non-monetary numeric value expressed in integers. |
| Rate | Number | DECIMAL(9,6) | A quantitative expression that represents the numeric relationship between two measurable units. |
| Quantity | Number | DECIMAL(9,2) | A non-monetary numeric value that does not have to be a whole number. |
| Amount | Number | MONEY | A monetary value. |
| Percent | Number | DECIMAL(9,2) | The ratio of part of whole expressed in hundredth. |

## Standard Abbreviations

|  |  |
| --- | --- |
| **Word** | **Abbreviation** |
| AMOUNT | AMT |
| CODE | CD |
| COUNT | CNT |
| DATE | DT |
| DATETIME | DTTM |
| DESCRIPTION | DESC |
| IDENTIFIER | ID |
| INDICATOR | IND |
| NAME | NM |
| NUMBER | NUM |
| PERCENT | PCT |
| RATE | RT |
| TEXT | TXT |
| TIME | TM |
| TIMESTAMP | TS |
| TYPE | TYP |
| IDENTIFICATION | IDNT |
| METHOD | MTHD |
| PARTY | PARTY |
| CONTACT | CNTCT |
| ADDRESS | ADDR |
| LANGUAGE | LANG |
| PREFERENCE | PRFR |
| ORGANIZATION | ORG |
| EXTENSION | EXTN |
| PRIMARY | PRI |
| EFFECTIVE | EFF |
| DEPARTMENT | DEPT |
| OCCUPATION | OCCPT |
| POSITION | PSTN |
| EDUCATION | EDU |
| DEMOGRAPHIC | DEMO |
| ISSUE | ISSUE |
| EXPIRY | EXP |

## Standard Class words

Below is the list of class words to be used for each attribute or column name in developing the logical and physical data models.

|  |  |  |
| --- | --- | --- |
| **Word/Words** | **Abbreviation** | **Definition** |
| AMOUNT | AMT | A monetary value. |
| CODE | CD | A combination of one or more numbers, letters,  or special characters substituted for a specific  meaning. |
| COUNT | CNT | A non-monetary numeric value expressed in integers. |
| DATE | DT | A specific period of time represented in any  format of month, day, and year. |
| DATETIME | DTTM | A specific period of time represented in any  format of month, day, year, hour, minute and seconds. |
| DESCRIPTION | DESC | An unformatted character string generally in the form of words. |
| IDENTIFIER | ID | A unique sequential value expressed in integer that has no defined meaning. |
| INDICATOR | IND | An attribute having two possible values that are contrary. |
| NAME | NM | A designation of an object or data model  class expressed in a word or phrase. |
| NUMBER | NUM | An identifier that may have alphanumeric characters; an ID number. |
| PERCENT | PCT | The ratio of part of whole expressed in hundredth. |
| TEXT | TXT | An unformatted character string generally in the form of words. |
| TIME | TM | A notation of a specified chronological point within a period of time. |
| TIMESTAMP | TS | A specific point of time, described as a sequential seven-part value, presented in the order of most  significant component to least significant component. |
| TYPE | TYP | A particular kind, class, group, or category. |
| RATE | RT | A quantitative expression that represents the numeric relationship between two measurable units. |

## Acronyms

List of abbreviations used in this document.

|  |  |
| --- | --- |
| **Abbreviation / Acronym** | **Meaning** |
| 3NF | Third Normal Form |
| CDM | Conceptual Data Model |
| ECM | Enterprise Conceptual Model |
| ERD | Entity Relationship Diagram |
| FK | Foreign Key |
| LDM | Logical Data Model |
| PDM | Physical Data Model |
| PK | Primary Key |

## Data Model References

<https://www.niem.gov/technical/Pages/The-Model.aspx>