SOP-0175 Data Models Standard Naming Convention - Table and Column Names

Document Control #	SOP-0175
Process Title	Data Models Standard Naming Convention - Table and Column Names
Version #	Version 1.0, 03/13/24
Effective Date	03/13/24
Review Date	March 13, 2024
Owner	Data Platform & Product Engineering
Approved By	VP, Data Platform & Prod. Engineering

Version Control

Version #	Date	Author	Requestor	Change Summary
1.0	03/13/24	L. Sims	L Sims	Initial Creation

1. PURPOSE

This document outlines the standard procedure to follow when naming a table and/or column in the Production or "Live" instance of our data platform (currently known as 'hv-runway-prod'). This standard will be reviewed on a reoccurring basis to validate adherence. Initial creation or modification of a table and/or column should be validated against this standard (see "Specific Procedure" section below for more information).

2. SCOPE

The scope of this standard only pertains to the creation and/or modification of a table or column name in the production instance of Healthverity's Data Platform for all three zones (Red, Orange and Green). However, supplier schemas are exempt from this standard in the Red Zone. This standard will not be associated with the development and test environments of the data platform. Also, these standards do not currently exist for internal use such as for finance, human resources, etc.

3. DEFINITIONS
- **Data Model -** tables and views used to house data in the various data platform layers

- **"Red" (Raw) Zone** - Represents supplier specific data models and PHI type products
  - Original supplier tables, partitioned by batch_id. The Data Models are not compliant with HV Standards.
  - Data Products with PHI - because these products have PHI, they will live in Red and follow these standards
- **"Orange" (Transformation) Zone** - Represents conformed model per data type - One model for all data suppliers and all records. Partitioned by batch_id
- **"Green" (Delivery) Zone** - Represents the final model - Data model is ready for consumption, all privacy rules and add-on rules applied. Clustered by service month, supplier, and HVID

4. POLICY STATEMENT

There is an expectation that tables and columns in the data platform fall under this standard. To validate adherence to this standard, a validation script will be automatically ran against all tables and columns in the data platform, within scope. If a table or column is found out of compliance, it is the data table owner's responsibility to fix the issue. The table owner will be notified via email that corrections are to be made. The table owner will have access to the validation script to use a preliminary test against the table to confirm that the table and column names are in compliance. Please see "Data_Governance_Naming_Standards" script link under Internal Reference section below

A dictionary has been created that is currently holds words, abbreviations and acronyms

found in the data model table and column names: Standard Naming Convention - Data Models However, new additions or changes may be needed. Please see how to proceed with this need below.

4.1 Requests for Newly Added, Words, Acronyms or Abbreviations

Utilizing a new word, new industry standard acronym, or newly used abbreviation in a table or column name will result in error when the validation script is run. Therefore, the new additions will need to be added to the standard naming convention dictionary. To do this, please submit a Data Governance Master Change Request (Data Governance) so that the following updates can be completed:

a. The "Standard Naming Convention - Data Models" is updated (also found in the internal reference section see below)
b. The "Abbreviations" sub-section of this SOP is updated, if applicable. (see below)
c. The script used to validate adherence of this standard is updated to include the new additions (see Internal Reference section below)

4.2 Requests to Modify Currently Used Acronyms or Abbreviations

Changes to currently used acronyms, words or abbreviations in a table or column name without permission will result in system disruptions. Therefore, these requests must be handled via the submission of a Data Governance Master Change Request (Data Governance)

5. ROLES AND RESPONSIBILITIES

The following leadership and organization roles and responsibilities apply to this procedure:

1. VP - Data Platform & Product Engineering - owns compliance for this procedure
2. Assigned Product Data Architect - creates compliance for this procedure
3. Data Governance Lead - maintains and manages compliance for this procedure

All departments, all roles - all department managers - Responsible for adhering to and complying with the requirements of this procedure

6. SPECIFIC PROCEDURE

6.1 General Naming Standards

Both tables and columns are to follow these same standards listed below:

- Singular words are separated by underscore
- Acronyms are to be used only for groups of words of two or more and should follow healthcare industry standards
- No special characters are allowed
- All singular words, abbreviated singular word exceptions, and acronyms are in lowercase
- If words are not found in the dictionary, then they are considered new words that need to be added to the dictionary. These additions must be submitted via the Data Governance Master Change Request (Data Governance) for approval so that the following updates can be completed:
    a. The "Standard Naming Convention - Data Models" is updated (also found in the internal reference section see below)
    b. The "Abbreviations" sub-section of this SOP is updated, if applicable. (see below)
    c. The script used to validate adherence of this standard is updated to include the new additions (see Internal Reference section below)
- All words not listed as an abbreviated exception must be spelled out
    o (See exceptions section below)
- All names should be unique within a context
- Nouns are used in singular form only

6.2 **Initial Validation**

A validation script has been created for general use to confirm that a new created or modified table and/or column name meets these standards. All are to utilize the following script to test and correct what is required based on this standard:

- **Data_Governance_naming_standards script**: https://dbc-eb84c807-40a3.cloud.databricks.com/?o=3671402501693277#notebook/1624199157137572

7. ASSUMPTIONS AND CONSTRAINTS

This procedure assumes financial, labor and technical resources will be available to support the activities outlined in this SOP.

8. EXCEPTIONS

8.1 Acronyms

Acronyms are to be used only for groups of words of two or more and should follow healthcare industry standards.

	Industry Standard Acronyms	Long Name
1	aca	Affordable Care Act
2	ahrq	agency for healthcare research and quality
3	apr	should usually be used with _drg as well (the full field name would be apr_drg) apr = all patient refined
4	cbsa	core based statistical area (about geography)
5	ccu	critical care unit
6	cdm	charge description master
7	cob	Coordination of benefits
8	cpt	Current Procedural Terminology
9	cvx	Vaccine Administered Code
10	drg	Diagnosis Related Group
11	eob	explanation of benefits
12	esrd	End-Stage Renal Disease (ESRD)
13	hcpcs	Healthcare Common Procedure Coding System code
14	hpid	healthcare plan identifier
15	icu	intensive care unit
16	ime	indirect medical education (IME)
17	loc	Location Of Care

18	mcc	multiple chronic conditions
19	msp	Medicare Secondary Payer (MSP)
20	naic	National Association of Insurance Commissioners code
21	ncpdp	National Council for Prescription Drug Programs
22	ndc	national drug code
23	npi	National Provider Identifier
24	poa	present on admit
25	pos	Place Of Service
26	pps	Prospective Payment System (PPS)
27	rxnorm	NIH UMLS RxNorm
28	ssn	Social Security Number
29	tob	Type of Bill
30	uom	Unit of Measure
31	upin	Unique Provider Identification Number
32	yob	Year Of Birth

8.2 Abbreviations

The following list represents the words with approved abbreviated exceptions. Therefore, if the "Long Name" is to be used in a column name, the abbreviation of that word (see Abbreviated Exceptions") must be used in its place:

	Accepted Abbreviations	Long Name
1	addl	additional
2	addr	Address
3	adv	Advantage
4	amt	Amount
5	bin	Bank Identification Number
6	cat	Category
7	desc	Description
8	diag	Diagnosis
9	dispro	disproportionate
10	eff	Effective
11	emr	Electronic Medical Record
12	enc	Encounter
13	geo	geographical
14	hv	Healthverity
15	ind	Indicator
16	ins	Insurance
17	lic	License
18	med	Medical
19	metro	Metropolitan
20	org	Organization
21	orig	Original
22	pct	percentage

23	prim	Primary
24	proc	Procedure
25	prof	Professional
26	prov	Provider
27	psych	Psychiatric
28	qty	Quantity
29	qual	Qualifier
30	sec	Secondary
31	seq	sequence
32	src	Source
33	stat	Statistical
34	ver	Version

9. INTERNAL AND EXTERNAL REFERENCES

9.1 INTERNAL REFERENCES

The data dictionary used to store currently used industry standard acronyms and approved abbreviated words is attached below. Only those listed in the roles and responsibilities section of this document are allowed to make changes to this document.

- **Standard Naming Convention - Data Models**
- **Data_Governance_Naming_Standards** script

9.2 EXTERNAL REFERENCES

N/A

10. DEPARTMENT CONFLUENCE SPACE

Standard Naming Convention (Tables & Columns)

- https://healthverity.atlassian.net/wiki/x/uQAMgg