



**GEORGIA DEPARTMENT OF PUBLIC HEALTH
POLICY # CO-12009
DATA STANDARDS POLICY AND PROCEDURES**

Approval:		10/1/13
	James C. Howgate, Chief of Staff	Date

1.0 PURPOSE

To reduce redundancy in data management; to assure data integration is possible; to leverage DPH information assets for purpose of better decision-making.

1.1 AUTHORITY – The Georgia Department of Public Health (DPH) Data Analysis and Reporting Policy and Procedures is published under the authority of DPH and in compliance with the following:

1.1.1 DPH Governance Council

2.0 SCOPE

This policy applies to all data collected and managed in the department that will be used for DPH analysis and reporting.

3.0 POLICY

The policy of the Department of Public Health is that all data collected and managed in the department that will be used for DPH analysis and reporting shall adhere to the DPH Data Quality Principles and with the DPH Data Property Standards.

3.1. The policy of DPH shall be guided by the following principles:

- 3.1.1. A variable shall have one and only one name.
- 3.1.2. A variable shall have one and only one definition.
- 3.1.3. A variable shall be stored in one and only one data type.
- 3.1.4. A variable shall have one and only one field length.
- 3.1.5. A variable shall be stored in one and only one unit of measurement.
- 3.1.6. A variable shall be stored in one and only one level of measurement.
- 3.1.7. A variable shall represent or store only those values specified in its definition.
- 3.1.8. Data objects shall have one and only one source.

Department of Public Health POLICY AND PROCEDURES	Policy No.	CO-12009		
	Effective Date:	10/1/2013	Revision #:	
DATA STANDARDS	Page No.	2 of 9		

- 3.1.9. No duplicate sources of data objects shall exist.
- 3.1.10. No duplicate records shall exist in data objects.
- 3.1.11. All data domains, data objects and variables shall be free of data anomalies.
- 3.1.12. Unknown, missing and inapplicable values shall have respectively unique representations.
- 3.1.13. Unknown, missing and inapplicable values in all data domains shall have consistent representations as defined in 3.1.12.
- 3.1.14. All data dictionaries shall define the following data properties for each variable: Variable Name, Variable Storage Name, Variable Definition, Variable Data Type, Unit of Measurement, Level of Measurement, Unit of Analysis, Level of Analysis, Variable Field Length, Variable Precision, Variable Time Stamp, and Variable Associated Standards. If a data property is not applicable for a variable, "n/a" shall be noted.

4.0. ACCOUNTABILITY

- 4.1 DPH application business owners collecting and managing data.

6.0 DEFINITIONS

- 6.1 Appendix A shows definitions of each Data Quality Principle

7.0 RESPONSIBILITIES

- 7.1 The DPH Model Team shall ensure compliance to this policy (or procedure).
 - 7.1.1 DPH application business owners collecting and managing data.

8.0 PROCEDURES

Consult and adhere to the Standard Data Properties as found in Appendix B.

Applies to:

- New system development
- Existing system modifications
- Active acquisition of existing data sources destined for the DPH data warehouse.

Department of Public Health POLICY AND PROCEDURES DATA STANDARDS	Policy No.	CO-12009		
	Effective Date:	10/1/2013	Revision #:	
	Page No.	3 of 9		

9.0 REVISION HISTORY

REVISION #	REVISION DATE	REVISION COMMENTS
0	10/01/2013	Initial Issue
1		

10.0 RELATED FORMS

- Appendix A: Data Quality Principles – Definitions
- Appendix B: Standard Data Properties

Data Quality Principles – Definitions

These principles serve as a reference for all data quality management.

1. **A variable shall have one and only one name.** Example: the data item “sex” is named “sex” as opposed to “gender.” Or, if a database collects information about “permits,” then that construct shall not also be referred to as a “certification.”
2. **A variable shall have one and only one definition.** Example: The definition of white race is:
A person having origins in any of the original peoples of Europe, the Middle East or North Africa (OMB directive-15, 1997).
3. **A variable shall be stored as one and only one data type.** For example, string data such as ICD codes (cause of death codes with values such as A018, 001.1) should not be stored as a numeric field.
4. **A variable shall have one and only one field length.** Example: collection of street address should allow for 60 characters, not less.
5. **A variable shall be stored in one and only one unit of measurement.** Example: Birthweight is stored in grams, but not pounds and ounces.
6. **A variable shall be stored in one and only one level of measurement.** Example: a *nominal* variable such as “race” shall not be stored as *interval* data.
7. **A variable shall be represented by or stored as only those values specified in its definition.** Example: if 1=yes and 0=no, there should be no other values (e.g., 9, 8, z, abc) found in that field.
8. **Data objects shall have one and only one source.** For example: official Georgia Birth statistics will come from the Office of Health Indicators for Planning, Georgia Department of Public Health.
9. **No duplicate sources of data objects (storage or collection) shall exist.** Example: A central data repository for analytic health information shall contain each data domain (such as vital records, notifiable diseases, immunizations).
10. **No duplicate records shall exist in data objects.** Example: One and only one record for each birth in the birth data domain.
11. **All data domains, data objects and variables shall be free of data anomalies.** All data assets will be examined for invalid values and such values will be processed such that their values are in a known state.
12. **Unknown, missing and inapplicable values shall be represented by one known value.** Example: 99 (unknown), nulls, blanks, or out of range values all set to = -1.
13. **Unknown, missing and invalid values in all data domains shall have consistent representations as defined in 12.**
14. **All data dictionaries shall define the following data properties for each variable.**

The properties are in the following table:

Variable Name(s)	Name of the data item used for storage and if applicable, presentation. Storage names begin with a domain identifier (e.g. Birth), followed by an owner (e.g. Mother), followed by the variable name (e.g. birth.mother.education_level). Presentation name (or label) is used to present data, such as "Mother's education level."
Definition and Variable Associated Standards	A statement containing the reason to collect or use the variable and external standards that apply to the variable.
Valid Values	Acceptable values for the variable being defined (e.g. mother's age range = 10-55 years inclusive.).
Data Type	The characteristic of a variable that determines what kind of data it can hold. For example, data types include Byte, Boolean, Integer, Long Integer, Currency, Decimal, String, Double, and Date.
Field Length	The number of numerical places or characters for a specific field.
Unit of Measurement	(a) refers to the system of measurement: English or metric; (b) the specific unit, within a measurement system, at which measurements for a variable are made such as grams.
Level of Measurement	(1) Nominal: One data element is identified as disparate from another but no direction is implied. Categorical properties or labels; such as the variable race represented by White, Black or African-American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander; or Male/Female. Appropriate descriptive statistics are the mode (most commonly occurring value) and frequency counts; (2) Ordinal: Variables can be ranked to show one value is more or less than another value; however the difference cannot be calculated, such as 'high/medium/low.'" Objects are ordered by some nominal category irrespective of magnitude, and irrespective to the distance between ordered levels; such as variable representing the level of agreement with a statement represented by disagree, somewhat agree, agree, disagree. Appropriate statistics are the mode, frequency, median (middle value), and percentiles. (3) Interval: Variables have an established identical distance between them on a scale; however they do not have a true zero, such as grams, inches, days. Ordering of objects is respective to a nominal category, the distance between objects respective to the nominal category, and without respect to the magnitude of the nominal category such as the number of hours a client waited for service measured in hours represented by 1, 2, 3, 4 etc. Appropriate statistics are the mean (average), median, mode, standard deviation (square root of the variance), range (maximum value – minimum value) and percentiles. (4) Ratio: Have a true zero. Objects are ordered respective to a nominal category, where the distance between objects is known, and each objects measurement is respective to a known zero value such as annual income measured in dollars represented by 20,051, 55,987, 42,042, etc.
Unit of Analysis	The unit of measurement assigned to a variable for analysis.
Level of Analysis	The level of measurement assigned to a variable for analysis.
Derivation	For calculated fields, the variables used and method to derive the calculated variable.
Time Stamp of Standard	The date on which the variable definition was defined or revised.

Standard Data Properties

PERSON (Age, Race, Ethnicity, Sex)

Property	Value
Presentation Name(s)	AGE
Definition	Elapsed time since birth. Age in years at the time of this event. Age is reported as age at last birthday - that is, age in completed years, and calculated by subtracting date of delivery from the reference date, with the reference date being the date of the examination, interview, or other contact with an individual (NCHS).
Valid Values	0 - 44,194 days. 0 – 120 years. (44,194 translates into 120 years); -1=unknown. Four (4) years = (365 x 4) + 1.
Data Type	Long Integer
Field Length	5
Unit of Measurement	Day
Level of Measurement	Interval
Unit of Analysis	Day
Level of Analysis	Interval
Derivation	EVENT_DATE – DOB = AGE IN DAYS. CONVERT TO YEARS FOR PRESENTATION.
Time Stamp of Standard	3/27/2002.

Property	Value
Presentation Name(s)	RACE
Definition	White = a person having origins in any of the original peoples of Europe, the Middle East or North Africa; Black or African-American = A person having origins in any of the black racial groups of Africa; Asian=A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand and Vietnam; American Indian/Alaska Native=A person having origins in any of the original peoples of North and South America (including Central American), and who maintains tribal affiliation or community attachment; Native Hawaiian or Other Pacific Islander=A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. Multiracial = 2 or more of these races (OMB-15, 1997).
Valid Values	1=White, 2=Black or African American, 3=Asian, 4=American Indian/Alaska Native, 5=Native Hawaiian / Pacific Islander, 6=Multiracial; -1=unknown.
Data Type	Integer
Field Length	2
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Time Stamp of Standard	6/12/2000.

Property	Value
Presentation Name(s)	ETHNICITY
Definition	Ethnicity, currently limited to asking whether the person is "Hispanic or Latino" (A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race) (OMB-15, 1997).
Valid Values	1=Hispanic or Latino, 0=No; -1=unknown.
Data Type	Integer
Field Length	2
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Derivation	Practice in vital records for pre-03 version of certificates: Derived from Origin values 1-5 inclusive.
Time Stamp of Standard	3/16/2000 per U.S. OMB Directive 15, 1997.

Property	Value
Presentation Name(s)	SEX
Definition	Biological sex
Categorical Attributes	Male / Female
Valid Values	1=Male, 2=Female, -1=unknown.
Data Type	Integer
Field Length	2
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Time Stamp of Standard	5/4/2000.

"Place" requires COLLECTION of 4 data items: Street, City, State, and Zip such that geocoding can assign a latitude/longitude.

Specifications for collection of each follow:

Property	Value
Presentation Name(s)	STREET ADDRESS
Definition	Geographic street address of event owner, event, or facility.
Valid Values	Street number followed by street name, followed by street direction (e.g. NW), followed by Apartment identifier, followed by room identifier. Storage of street names will follow standard abbreviations such as RD, BLVD, LN, ST. (See http://www.gis.co.clay.mn.us/usps.htm) -1 = Unknown.
Data Type	String
Field Length	60
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Derivation	Self-report.
Time Stamp of Standard	3/16/2000 2.10.2004 – no need to proclaim "0"=Out of state. 6.10.13 – updated spec to be 60 characters instead of 40.

Property	Value
Presentation Name(s)	CITY NAME
Definition	For use in collection of city names: Geographic city or town or parish or Census Designated Place (CDP) or village where the event occurred or event owner lived.
Valid Values	USPS City Names; -1 = unknown, 0=non-Georgia resident, _UNIN = resident/event in an unincorporated place (not in a city limit).
Data Type	STRING
Field Length	27
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Derivation	A collection item. City is Derived from latitude/longitude via a Geocoding process where applicable, and stored as a city code.
Time Stamp of Standard	3/16/2000. 1/23/07

Property	Value
Presentation Name(s)	ZIPCODE
Definition	Geographic zipcode+4 where the event owner lived or where event occurred, using USPS standard zipcodes.
Valid Values	USPS zipcodes; -1 = unknown. 0=non-Georgia zip.
Data Type	String
Field Length	9
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Derivation	Derived via geocoding where applicable. Originals stored as *_original.
Time Stamp of Standard	3/16/2000

Property	Value
Presentation Name(s)	STATE
Definition	Geographic state of residence or event.
Valid Values	USPS postal abbreviations.
Data Type	String
Field Length	2
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Derivation	Self-report
Time Stamp of Standard	3/16/2000

Place **STORAGE** standards

Property	Value
Presentation Name(s)	LATITUDE
Definition	Latitude of event or residence.
Categorical Attributes	N/A
Valid Values	- - - 0 = non-georgia.
Data Type	Floating point.
Field Length	Float
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless

Level of Analysis	Nominal
Derivation	Derived from geocoding process, using Street, City, State, Zip.
Time Stamp of Standard	3/16/2000. See issue 177.
Access	<input type="checkbox"/> Public Use <input checked="" type="checkbox"/> Datamart Variable
Property	Value
Presentation Name(s)	LONGITUDE
Definition	Longitude of event or residence.
Categorical Attributes	N/A
Valid Values	- - 0 = non-Georgia.
Data Type	Floating point.
Field Length	Float.
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Derivation	Derived from geocoding process, using Street, City, State, Zip.
Time Stamp of Standard	3/16/2000. See issue 177.
Access	<input type="checkbox"/> Public Use <input checked="" type="checkbox"/> Datamart Variable

Property	Value
Presentation Name(s)	COUNTY
Definition	Geographic county of event or residence.
Valid Values	Two digit state FIPS code 00-99 followed by three digit FIPS county code 000-999; range 13001-13321; 0=Non-Georgia county, -1 = unknown.
Data Type	String
Field Length	5
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Derivation	Derived from geocoding where applicable. Note encapsulated rules for event county in specific instances.
Time Stamp of Standard	3/16/2000. 0 for non-ga 10.21.2003.

TIME

Property	Value
Presentation Name(s)	DATE
Definition	Date of an event.
Valid Values	Months (mm)=01-12 / Days (dd)=01-31 / Year (ccyy)=18yy, 19yy, or 20yy; 12/31/9999=unknown.
Data Type	Date
Field Length	10
Unit of Measurement	Day
Level of Measurement	Interval
Unit of Analysis	Day
Level of Analysis	Interval
Time Stamp of Standard	3/27/2002.