

Standard Data Item Properties of Georgia Division of Public Health Data

April 2003
Updated 3.26.04

Publication # DPH03-044W



**Georgia Department of Human Resources
Division of Public Health
Office of Health Information and Policy**

Data Property –

A required element stored in a data dictionary that defines a variable's storage and analysis specifications.

**STANDARDS FOR DATA COLLECTION, STORAGE, AND USE BY THE
GEORGIA DIVISION OF PUBLIC HEALTH**

First Released in March, 2000

Prepared by:

**Office of Health Information and Policy
Georgia Division of Public Health**

Architect: Frank H. Millard, M.A

**Developers: Gordon R. Freymann, M.P.H., Michael A. Coletta, M.P.H.,
Rajni Samavedam, M.P.H., Cathie McCabe, M.B.A., Jacqueline R. Bennett-Burkholder, Ph.D., M.S., Robert Attaway
Project Lead: Gordon R. Freymann, M.P.H.**

Many thanks to those that helped develop these standards. These standards for the Georgia Division of Public Health were initially tackled in what was the first Division Data Policy Working Group, led by Gordon Freymann. The people to thank include but are not limited to: Janice Heckler, Cathie McCabe, Robert Attaway, Michael Coletta, Jacqueline Bennett-Burkholder, Frank Millard and Rajni Samavedam – OHIP, Emily Kahn, John Carter, Mohammed Qayad – MCH/EPI, Araine Kraus – Family Health Branch, Tasia Atha – Rome District.

Introduction

Five Dimensions of Public Health: Basic principles of epidemiology are **person, place** and **time**. Therefore the five dimensions of public health are *Race, Age, Sex*, which all describe 'person'; *Place* (that is, where an event occurred and in addition to residence); and *Time* (for example, date of a client visit). These elements are the basic core dimensions for all data collection within the Division and reflect the minimal data needed to create analytical health information. Such information is required for efficient management of Division operations so that maintenance of conditions in which people can be healthy is assured.

Without data standards, the management of Division data assets becomes unnecessarily complex, and the reliability of information production suffers.

Note regarding Unknown values: All missing, unknown, and null values are transformed into one code, -1. This transformation creates an easy way of selecting unknowns when working with row-level data, and addresses issues where some database management systems do not allow for the selection of nulls. This standard also provides: 1) a consistent value across all data items for missing, unknown, invalid, and null; using "9" cannot be the unknown for all data items; 2) a value independent of the field length of a given data item (again, if "9" is used in a field length of 1, often a "99" must be used in a field length of 2, and so on); 3) a value that will not interfere with valid value ranges (again, "9" may represent unknown in one data item, but be a valid value for another).

Explanation of Terms

Property	Definition
Variable Name (s)	Name of the data item used for storage and if applicable, presentation. Storage Names are used in the database. Presentation Names (or “labels”) are used when presenting data.
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 Boolean, Integer, Long Integer, Currency, Decimal, String, and Date.
Field Length	The number of numerical places or characters for a specific field. This is character field length, not data storage field length.
Unit of Measurement	Refers to the specific unit within a measurement system, at which measurements for a variable are made. Examples include birthweights measured in grams, or dates measured in days. "Unitless" refers to cases where a unit per se is not measured.
Level of Measurement	<u>Nominal</u> - categorical properties or labels; such as the variable race; <u>Ordinal</u> - objects are ordered by some nominal category irrespective of magnitude, and irrespective to the distance between ordered levels; such as a variable representing the level of agreement with a statement represented by disagree, somewhat agree, agree, disagree; <u>Interval</u> - 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.; and <u>Ratio</u> - 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. "Unitless" refers to cases where a unit per se is not analyzed.
Level of Analysis	The level of measurement assigned to a variable for analysis
Derivation	For calculated fields, the data items used and method to derive the calculated variable.
Time Stamp of Standard	The date on which the variable definition was defined or revised.

STANDARD PROPERTIES OF AGE DATA ITEMS

Property	Value
Presentation Name	AGE
Storage Name	AGEINDAY (FOR AGE IN DAYS) AGEYEAR (FOR AGE IN YEARS)
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.

STANDARD PROPERTIES OF RACE DATA ITEMS

Property	Value
Presentation Name	RACE
Storage Name	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.

STANDARD PROPERTIES OF ETHNICITY DATA ITEMS

Property	Value
Presentation Name	ETHNICITY
Storage Name	ETHNIC
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	n/a
Time Stamp of Standard	3/16/2000 per U.S. OMB Directive 15, 1997.

STANDARD PROPERTIES OF SEX DATA ITEMS

Property	Value
Presentation Name	SEX
Storage Name	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.

Standard Properties of all PLACE data items

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

Specifications for each follow:

Collection of STREET DATA ITEMS

Property	Value
Presentation Name	STREET ADDRESS
Storage Name	RESSTAD (FOR RESIDENCE 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.usps.com/ncsc/lookups/usps_abbreviations.html#suffix) Note that storage values can include: -1 = Unknown.
Data Type	String
Field Length	40
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

Collection of CITY NAME DATA ITEMS

Property	Value
Presentation Name	CITY
Storage Name	RESCITY (FOR CITY OF RESIDENCE)
Definition	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. Note that storage values can include: -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	Self-report.
Time Stamp of Standard	3/16/2000

Collection of STATE DATA ITEMS

Property	Value
Presentation Name	STATE
Storage Name	- -
Definition	Geographic state of event or residence.
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

Collection of ZIPCODE DATA ITEMS

Property	Value
Presentation Name	ZIPCODE
Storage Name	RESZIP (FOR ZIPCODE OF RESIDENCE)
Definition	Geographic zipcode+4 where the event owner lived or where event occurred, using USPS standard zipcodes.
Valid Values	USPS zipcodes. Note that storage values can include: -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	Self-report.
Time Stamp of Standard	3/16/2000

STANDARD PROPERTIES OF STATE DATA ITEMS

Property	Value
Presentation Name	STATE
Storage Name	RESSTAT (FOR STATE OF RESIDENCE)
Definition	Geographic state of event or residence.
Valid Values	-2=Remainder of World, 01=Alabama, 02=Alaska, 04=Arizona, 05=Arkansas, 06=California, 08=Colorado, 09=Connecticut, 10=Delaware, 11=District Of Columbia, 12=Florida, 13=Georgia, 15=Hawaii, 16=Idaho, 17=Illinois, 18=Indiana, 19=Iowa, 20=Kansas, 21=Kentucky, 22=Louisiana, 23=Maine, 24=Maryland, 25=Massachusetts, 26=Michigan, 27=Minnesota, 28=Mississippi, 29=Missouri, 30=Montana, 31=Nebraska, 32=Nevada, 33=New Hampshire, 34=New Jersey, 35=New Mexico, 36=New York, 37=North Carolina, 38=North Dakota, 39=Ohio, 40=Oklahoma, 41=Oregon, 42=Pennsylvania, 44=Rhode Island, 45=South Carolina, 46=South Dakota, 47=Tennessee, 48=Texas, 49=Utah, 50=Vermont, 51=Virginia, 53=Washington, 54=West Virginia, 55=Wisconsin, 56=Wyoming, 66=Guam, 72=Puerto Rico, 78=U.S. Virgin Islands, CA=Canada, CU=Cuba, MX=Mexico, -1=unknown, 00=Non-Georgia.
Data Type	String
Field Length	2
Unit of Measurement	Unitless
Level of Measurement	Nominal
Unit of Analysis	Unitless
Level of Analysis	Nominal
Derivation	Derived from geocoding process where applicable.
Time Stamp of Standard	3/16/2000

STANDARD PROPERTIES OF COUNTY DATA ITEMS

Property	Value
Presentation Name	COUNTY
Storage Name	RESCOUN (FOR COUNTY OF RESIDENCE)
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.
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.
Time Stamp of Standard	3/16/2000. 0 for non-ga 10.21.2003.

STANDARD PROPERTIES OF DATE DATA ITEMS

Property	Value
Presentation Name	DATE
Storage Name	EVENDATE (FOR DATE OF THE EVENT FOR WHICH INFORMATION IS BEING COLLECTED)
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.

STANDARD PROPERTIES OF MARITAL STATUS DATA ITEMS

Property	Value
Presentation Name	MARITAL STATUS
Storage Name	MARITAL
Definition	Legally recognized marital status.
Categorical Attributes	Two parent categories are 1=married and 2=unmarried. Married includes 3=married but separated. Unmarried includes 4=widowed, 5=divorced, and 6=never married.
Valid Values	1=married, 2=unmarried, 3=married but separated, 4=widowed (and not remarried), 5=divorced (and not remarried), 6=never married, -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	4/7/2000

STANDARD PROPERTIES OF EDUCATION LEVEL DATA ITEMS

Property	Value
Presentation Name	EDUCATION LEVEL
Storage Name	EDUC
Definition	The last grade of formal education completed by the event owner.
Categorical Attributes	No high school (<9), Some high school but no diploma or general equivalency diploma (GED) (9-11); high school diploma or GED (12); some college, no bachelor's degree (13-15); bachelor's degree or higher (16+).
Valid Values	0=None, 1=1 st Grade Completed, 2=2 nd Grade, 3=3 rd Grade, 4=4 th Grade, 5=5 th Grade, 6=6 th Grade, 7=7 th Grade, 8=8 th Grade, 9=9 th Grade, 10=10 th Grade, 11=11 th Grade, 12=12 th Grade, 13=1 Year College, 14=2 Years College, 15=3 Years College, 16=4 Years College, 17=5 or More Years College, -1 = unknown.
Data Type	Integer
Field Length	2
Unit of Measurement	Ordinal
Level of Measurement	Grade
Unit of Analysis	Ordinal
Level of Analysis	Grade
Time Stamp of Standard	4/7/2000 per NCHS.