

The following is an excerpt (Chapter 2) from the *Surveillance and Data Standards for USDA/APHIS/Veterinary Services, version 1.0 (July 2006)* that was developed by the Centers for Epidemiology and Animal Health in July 2006. This section provides guidelines for epidemiologists and database developers on the type and format of data associated with animal health surveillance programs.

Chapter 2 Data Concepts and Data Classes

2.1 Introduction

Animal health surveillance requires timely, accurate, and accessible data that facilitate analysis and reporting and contribute meaningful information for decision-making. Although surveillance databases are designed to collect specific data that will enable analysts to address the goals and objectives of a specific system, some data classes, or fields, are recorded in nearly all surveillance systems (e.g., dates, location information, species names, disease names, and others). The use of standards when capturing common data classes assures that the resulting databases not only provide data necessary to address system-specific objectives, but also for broader information inquiry and analysis. Within VS, data standards enhance our ability to provide timely and useful animal health information.

This chapter identifies the commonly used data classes in animal health surveillance systems and provides standard specifications for recording these fields. This chapter may also serve as a useful "library" of common data classes to be considered when developing a data collection system for animal health surveillance purposes. Additionally, appendices at the end of this document provide specific codes for several commonly used data classes.

The data categories and classes described here do not include all fields needed to address the specific needs of a particular surveillance system, but provide a starting point for a data dictionary. Similarly, it is not expected that every data class listed will be useful for every surveillance system. Surveillance planners, data analysts, and database designers should determine their specific data needs and use the appropriate standards and specifications for their system. Business rules, entity relationships, and other "best practices" in the design and implementation of database systems are beyond the scope of this chapter, although these factors may significantly impact data quality and accessibility. For surveillance systems within USDA/APHIS/VS, surveillance planners and database designers should request assistance from the VS Application and Information Management (VS AIM) group for more assistance with database design, implementation, and management.

Table: Summary of data concepts listed in this chapter.

Data Concept
2.3.01 Event
2.3.02 Subject
2.3.03 Population
2.3.04 Sample
2.3.05 Observation or Test Results Information
2.3.06 Premises
2.3.07 Person
2.3.08 Geographic Information

Terms. This chapter identifies *data concepts* and *data classes*. A *data concept* is a general grouping of data classes that are common to surveillance systems. *Data classes* are the individual data items (e.g., fields or variables) recorded within a concept. Although some concepts or classes are common to most surveillance systems, the use or importance of many will vary according to objectives and scope of the particular system. Data classes are distinguished into “tiers” based on frequency of use and how common they are in most surveillance programs.

The grouping of data classes into concepts facilitates the presentation of standards information so that data classes related to similar topics can be considered as a group. Data concepts may or may not parallel tables in a relational database. For example, classes in the concept “event” may all occur in a single database table, but classes in the concept “geographic information” will likely occur in many tables within a database.

Term	Description	Examples
Data Concept	A general grouping of classes for a surveillance system. These are generally non-redundant, non-ambiguous, and necessary for any surveillance system	<ul style="list-style-type: none"> ➤ Event ➤ Sample ➤ Results ➤ Premises ➤ Person ➤ Geographic Info
Tier 1 Data classes	The most frequently used data classes in a concept	<ul style="list-style-type: none"> ➤ Disease ➤ Species
Tier 2 Data classes	Data classes used in some but not all surveillance databases	<ul style="list-style-type: none"> ➤ Breed

2.2 General Data Standards and Specifications

String values coded as integers:

- Some variables specified as string variables in this chapter may be coded as integers IF the database includes a related reference table to explain the code meanings (e.g., "gender" may be coded as integers 1, 2, etc. when related to a reference table with a description for each code.)
- Boolean Values: For Boolean values that are coded as integers, no=0 and yes=1.
- Date fields: The standards format for date fields is mm/dd/yyyy
- Metric vs. English Measurement Units: Metric or English units may be used for measurement (e.g., weight), but the selected measurement system must be used consistently throughout the database.

2.3 Data Concepts

2.3.01 Data Concept: Event

Animal health surveillance data are typically collected in association with an event such as market testing, slaughter, on-farm testing, and others. Classes in this concept describe or are associated with animal health events.

Tier	Data class	Suggested Database Name	Description	Specifications (Type, length, precision, units)	Example
1	Event Type	EVENT_TYPE	Type of event (e.g., vaccination event, market testing)	String	Test; vac; insp
	Event Reason	EVENT_REASON	The trigger for the event or reason event is recorded	String or integer	Market Vaccination
	Event Date	EVENT_DATE	Date of the event	See "Date/Time" data concept	See "Date/Time" data concept
	Event Location (requires multiple fields)	EVENT_LOC	Location of the event; One or more fields that describe street address, city, county, state, latitude/longitude, or other location variables. The level of detail recorded for location will depend on management and analysis needs.	See "Geographic Information" data concept	See "Geographic Information" data concept
	Disease/condition	DISEASE	Disease or condition associated with the event; may be recorded as numeric code or string variable	String or integer	See Appendix A "Disease Codes"
	Species	SPECIES	Species associated with the event; may be recorded as numeric code or string variable	String or integer	See Appendix B "Species Codes"
2	Breed	BREED	Animal breed associated with the event; may be recorded as numeric code or string variable	String or integer	See Appendix C "Species-Breed Codes"

Initiating/previous event	---	Provides link to previous event	String or integer
Investigator	---	Person who recorded the event information; often needed for program or data management purposes	See "Person" data concept See "Person" data concept
Person or Complainant	---	Person who requested or initiated the event	String
Results Summary (may require multiple fields)	---	One or more fields that summarize the results from the event, (e.g., number tested, number negative, number positive)	String or double

2.3.02 Data Concept: Subject

Data classes in this concept may describe an individual animal or an entire herd/flock. Surveillance analysis requires information about the characteristics of individual animals (or herds or flocks) and the population under surveillance. An individual animal record, or a record for an individual herd or flock, may include data fields that describe population characteristics; "subject" and "population" fields may occur in the same database table. However, population characteristics are described in a separate data concept because those characteristics should be considered separately while planning data collection and database systems to ensure that the results will provide useful information for analysis.

Tier	Data class	Suggested Database Name	Description	Specifications (type, length, precision, units)	Example
1	Species	SPECIES	Species	String	See Appendix B "Species Codes"
	Individual or group indicator	---	Indicates whether the subject is a group or an individual animal	String or integer	
	Age	AGE	Numeric value of the subject's age, usually in months or years	Single	5
	Age Unit	AGE_UNIT	Unit of measure for age, usually indicates month or year	String or integer	YR, MO 1, 2
	Animal Identification Type/s	ID_TYPE	Type of animal identification methods (e.g. tag type)	String or integer	See Appendix D "Identification Types"
	Animal Identification (may require multiple fields)	---	Variable(s) used to identify the animal or herd in the database (e.g., id numbers for the database record) and/or in the field (e.g., tag numbers)	String	---
2	Breed	BREED	Breed of the subject; may be recorded as a numeric code or string variable	String or integer	See Appendix C "Breed Codes"
	Gender	GENDER	Gender of the subject; may be recorded as a numeric code or string variable	String or integer	See Appendix E "Gender Codes"

Neutered	NEUTERED	Yes/ no / unknown field that indicates if the animal is neutered	String or integer (use 0=no and 1=yes when integers to code yes/no fields)	Yes/No/Unknown 0, 1
Age range (requires multiple fields)	---	When an age range is given for a herd rather than a single age value, the age range values should be reported in separate fields that indicate age and age unit values for the minimum and maximum of the range; use of text strings to describe age ranges should be avoided (e.g., avoid using "15 months to 5 years" to describe age range)	String and integer	Min_age=2 Min_age_unit=mo Max_age=10 Max_age_unit=yr
Subject location (may require multiple fields)	---	Subject's (animal or herd/flock) current residence location One or more fields that describe street address, city, county, state, latitude / longitude, or other location variables. The level of detail recorded for subject location will depend on management and analysis needs.	See "Geographic Information" data concept	See "Geographic Information" data concept
Subject origin (may require multiple fields)	---	Subject's (animal or herd/flock) location of origin One or more fields that describe street address, city, county, state, latitude / longitude, or other location variables. The level of detail recorded for subject origin will depend on management and analysis needs.	See "Geographic Information" data concept	See "Geographic Information" data concept
Herd or Flock size	---	Herd or flock size is the number of animals in the herd or flock. This class is critical when the subject is the herd or flock rather than an individual animal, but herd/flock size may also be recorded when the subject is an individual animal or sample; see "population data" concept.	Integer	1422

Weight	---	Numeric value of animal weight	Single	45.80
Weight Unit	---	Unit of measurement for weight, often pounds, grams, kilograms; unit should be recorded uniformly throughout the surveillance database so that multiple words or abbreviations are not used to indicate the same weight unit (e.g., kilograms, kg, or kilos); also, use metric or English units consistently within a database	Integer	Lb, pounds, gm, kg, kilograms
Use/Function	---	The intended use of the animal or herd from the owner's perspective	String	See Appendix F "Subject Use/ Function Codes"
Status	---	Can be either an individual or herd/flock status	String	Vaccination, Weaning, Infected, Exposed
Number Sick	---	When the subject is a herd or flock, numbers of sick animals should be reported by species on each premises	Integer	22
Number Dead	---	When the subject is a herd or flock., number of dead animals should be reported by species on each premises	Integer	11
Number Clinically Normal	---	When the subject is a herd or flock, number of clinically normal animals should be reported by species on each premises	Integer	33

2.3.03 Data Concept: Population

Population information is often captured in conjunction with event or subject records, and the specific population classes that need recording depend on the objectives and data analysis needs. Whether recorded explicitly during data collection associated with the surveillance activity or through other sources, details on size and demographics are essential for animal health surveillance analysis and should be considered.

Tier	Data class	Suggested Database Name	Description	Specifications (Type, length, precision, units)	Example
1	Species	SPECIES	Species of the population	String	See Appendix B "Species Codes"
	Herd/Flock Size	---	The number of animals in the herd or flock; size may provide important "denominator" values for analyzing surveillance data.	Integer	25,632
2	Age	AGE	Numeric value of age, usually in months or years	Single	5
	Age Unit	AGE_UNIT	Unit of measure for age, usually indicates month or year	String or integer	YR, MO 1, 2
	Breed	BREED	Breed of the subject; may be recorded as a numeric code or string variable	String or integer	See Appendix "Breed Codes"
	Gender	GENDER		String or integer	See Appendix "Gender Codes"
	Location (may require multiple fields)		Fields that describe the location of the study population	See "Geographic Information" data concept	See "Geographic Information" data concept

2.3.04 Data Concept: Sample

Data classes in this concept describe samples that are collected. Minimally, the sample identifier and sample type must be recorded.

Tier	Data class	Suggested Database Name	Description	Specifications (type, length, precision, units)	Example
1	Sample identification	---	Identifier used to uniquely identify individual samples; may be recorded as a numeric code or string variable	String or integer or double	A12345
	Sample type	---	Type of tissue sampled	String	See Appendix G "Sample Type Codes"
	Date collected	---	Date the sample was collected	See "Date/Time" data concept	See "Date/Time" data concept
	Collection site (may require multiple fields)	---	Location where the sample was collected; one or more fields that describe street address, city, county, state, latitude/longitude, or other location variables. The level of detail recorded for location will depend on management and analysis needs.	See "Geographic Information" data concept	See "Geographic Information" data concept
2	Animal record identification	---	Key field used to link the animal and sample records in the database; required when animal and sample records are stored in separate tables	String or integer or double	
	Preservation	---	Method used to preserve sample	String	Ice, formalin
	Collector (may require multiple fields)	---	Person who collected the sample	See "Person" data concept	See "Person" data concept

Submitter (may require multiple fields)	---	Person who submitted the sample	See "Person" data concept	See "Person" data concept
Method of shipment	---	Method of getting sample from collection site to testing site (lab)	String	Truck, air
Date shipped	---	Date sample was shipped from collection site to testing site (lab)	See "Date/Time" data concept	See "Date/Time" data concept

2.3.05 Data Concept: Observation or Test Results Information

Data classes in this concept describe results from tests performed on samples or observational data collected during an animal health event.

Tier	Data class	Suggested Database Name	Description	Specifications (type, length, precision, units)	Example
1	Test type	---	Type of test performed or test name	String	PCR, ELISA
	Test observation	---	Quantitative or qualitative values that describe the result of the test. The values recorded for test results, whether quantitative or qualitative, require careful consideration of potential broader applications of the data.	String or integer or single or double; depends on test type	Positive 1.2345 55
	Test date	---	Date test was conducted	See "Date/Time" data concept	See "Date/Time" data concept
	Test location (may require multiple fields)	---	Location where test was performed; may be a laboratory name, laboratory identification number, or multiple fields describing physical location. The level of detail recorded for location will depend on management and analysis needs.	String or integer, or See "Geographic Information" data concept	NVSL, or see "Geographic Information" data concept
2	Date received	---	Date sample was received	See "Date/Time" data concept	See "Date/Time" data concept
	Test date	---	Date the sample was tested	See "Date/Time" data concept	See "Date/Time" data concept
	Clinical Sign	---	Physical observation of symptoms. It is preferable to record individual signs in separate fields for electronic tabulation of data.	String	Recumbent, diarrhea, respiratory distress.

Test interpretation	---	A qualitative judgment regarding the test outcome	String	See Appendix H "Result Interpretation Codes" or positive, suspect, negative
System affected	---	Anatomical system affected by the disease or condition	String	Respiratory, reproductive, GI, CNS
Person (may require multiple fields)	---	Person who interpreted or reported the test information; often needed for program or data management purposes	See "Person" data concept	See "Person" data concept
Remarks	---	Remarks related to the test or observation	String	

2.3.06 Data Concept: Premises

Premises information may be captured in association with event or subject records, or captured separately and linked by common identifiers to event or subject data. Premises must be a physical geographic location (i.e., post office box is not acceptable).

Tier	Data class	Suggested Database Name	Description	Specifications (type, length, precision, units)	Example
1	Premises id	PREM_ID	The unique identifier for the premises Premises identification numbers should be assigned using the National Animal Identification System	String Premises identification numbers should be assigned using the National Animal Identification System (alpha-numeric 7, random, character 7 is a check value)	002FVPL
	Location (requires multiple fields)	---	Location of the premises; this will include multiple fields describing physical address, city, county, state, latitude/longitude, and other location variables. The level of detail recorded for premises location will depend on management analysis needs	See "Geographic Information" data concept	See "Geographic Information" data concept
2	Premises type	---	A description of the type of premises, such as farm, slaughter plant, market, etc.		See Appendix I "Premises Type Codes"
	Species (may require multiple fields)	---	A description or list of species present on the premises; may be needed for program management purposes	String	See Appendix B "Species Codes"

Contact/s (may require multiple fields)		Premises owner(s) or contact person(s) for the premises; often needed for program or data management purposes; may include person name, phone numbers, mailing address	See "Person" data concept	See "Person" data concept
Status	---	Status of the premises (e.g., quarantine, certified, exposed)	String	See Appendix J "Premises Status Codes"
Active status	---	Current operational condition/status	String or integer or Boolean (yes/no)	Active Inactive

2.3.07 Data Concept: Person

Classes in this concept describe the attributes of any person directly or indirectly related to the surveillance data or data flow such as data collectors, animal owners, laboratory personnel, and others. Classes in the person concept may occur in many tables related to the collection of animal health surveillance data. See the event, subject, observation or test results, and premises categories for instances where person is indicated as a data class.

Tier	Data class	Suggested Database Name	Description	Specifications (type, length, precision, units)	Example
1	Last Name	---	Person's last name; should be in a separate field from the first name	String, upper, no spaces	Smith
	First Name	---	Person's first name; should be in a separate field from the last name	String, upper	John or John Jacob
2	Agency Name (may require multiple fields)	---	Agency name or professional affiliation; names or acronyms/codes must be uniform throughout a database so that one agency is not indicated by multiple names or acronyms	String	APHIS
	Role / Concept/ Affiliation	---	The person's role or relationship to the regulatory program	String	
	Phone number(s)(may require multiple fields)	---	Area code plus 7-digit number	Number, xxx-xxx-xxxx	
	Fax Number	---	Area code plus 7-digit number	Number, xxx-xxx-xxxx	

Address (requires multiple fields)	---	Person's mailing address	See "Geographic Information" data concept	See "Geographic Information" data concept
Location (requires multiple fields)	---	Person's physical location; may be needed if physical location is different from mailing address	See "Geographic Information" data concept	See "Geographic Information" data concept
E-mail address	---	Person's e-mail address	String; no spaces	person@agency.gov
Person type	---	The person's role as it relates to the surveillance data or data flow	String	Owner, veterinarian
Person Status	---	The person's activity status	String	Retired, active, inactive
Prefix	---	Prefix to a person's name	String	Dr., Mrs.
Suffix	---	Suffix to a person's name	String	Jr., Sr.

2.3.08 Data Concept: Geographic Information

Classes in this concept describe the attributes of any location information recorded in relation to surveillance data such as event or premises locations, or address information associated with people (see the “Person” data concept). Classes in the geographic information data concept may occur in many tables related to the collection of animal health surveillance data. See the event, subject, observation or test results, and premises categories for instances where location is indicated as a data class.

Tier	Data class	Suggested Database Name	Description	Specifications (type, length, precision, units)	Example
1	Address	ADDRESS	Physical address of the location	Use postal service standards	17454 Roller Coaster Rd
	City	CITY	City name	String	
	State	STATE	Two-character state abbreviation	Use postal service standards	AK, CO, NY
2	County		County name	String	Erie
	Zip Code		U.S. Postal Service 5-digit mail delivery area.	String	80132
	Zip Code plus 4		A geographic segment within the U.S. Postal Service 5-digit mail delivery area.	String	8312
	State / County FIPS code		State/County Federal Information Processing Standard Code	String	08041
	City FIPS code		Metropolitan Areas Federal Information Processing Standard Code	String	CO322
	Front Gate Latitude		Defined as the point of transition from public to private access when entering premises.	Double; decimal degrees	39.097886

Front Gate Longitude	Defined as the point of transition from public to private access when entering premises.	Double; decimal degrees	-104.845889
Coordinate type (front gate)	Method used to determine front gate coordinate (see appendix for values)	String	GPS, GEO See Appendix K "Coordinate Type Codes for Geographic Classes"
Validation code (front gate)	Method used to validate front gate coordinate (see appendix for values)	String	NVG, CBG See Appendix L "Validation Codes for Geographic classes"
Estimated positional error (front gate)	Estimated positional error during GPS data collection.	Double; meters	42