Public Health Conceptual Data Model

Premiere Edition

U.S. Department of Health and Human Services

Public Health Service
Centers for Disease Control and Prevention (CDC)
Atlanta, Georgia 30333

July 2000

Table of Contents

INTRODUCTION

Introduction

This document presents the Public Health Conceptual Data Model (PHCDM). The document is comprised of four sections. Section one includes this introduction, model background, a summary of the goals and objectives for the model, and a project scope statement. Section two is a guide to understanding the data model. Section three includes the graphical representation of the data model and the supporting data dictionary. Section four consists of a series of appendices including datatype definitions, a model scenario, frequently asked questions, a glossary, and a bibliography.

The purpose of the Public Health Conceptual Data Model is to document the information needs of public health so that the Centers for Disease Control and Prevention (CDC) and its state and local partners in public health can:

- Establish data standards for public health, including data definitions, component structures (such as for complex datatypes), code values, and data use;
- Collaborate with national health informatics standards setting bodies to define standards for the exchange of information among public health agencies, and healthcare providers;
- Construct computerized information systems that conform to established data and data interchange standards for use in the management of data relevant to public health.

This model is a work in progress, and is being developed with the participation of our public health partners. CDC is developing a process to receive comments from its partners, address issues raised, and provide feedback on the disposition of the comments. In the interim, contact the CDC Health Information and Surveillance Systems Board (HISSB) Executive Secretariat with questions and comments regarding the model: hissb@cdc.gov, or phone 770-488-8301 or 770-488-8302.

Background

The PHCDM is one of many interrelated activities supporting CDC's National Electronic Disease Surveillance System (NEDSS) initiatives. The long-term vision for NEDSS is a collection of complementary computerized information systems that automate the process of gathering health data, facilitate the monitoring of the health of communities, assist in the analysis of trends and detection of emerging public health problems, and provide information for setting public health policy.

The focus of the NEDSS initiative is the development, testing, and implementation of information management technology standards that will support more complete and comprehensive integration of computerized health information systems for use in public health. The NEDSS standards focus on five important areas:

1. Data Architecture (data model, data definitions, and coding rules);
2. User Interface;
3. Information Systems Architecture (based on industry standards);
4. Tools for interpretation, analysis, and dissemination of data;
5. Secure data transfer.

The PHCDM is a major component of the NEDSS data architecture standards. Together with Common Information for Public Health Electronic Reporting (CIPHER) guidelines, the PHCDM provides a foundation for standardization of public health data collection, management, transmission, analysis, and dissemination.

The development of the PHCDM began in May 1999. The first step was the construction of a high-level data model depicting the major subject areas to be included in the PHCDM. The subject area data model was developed by conducting an analysis of selected CDC disease surveillance systems, the Health Level Seven (HL7) Reference Information Model (RIM), and other health-related data models. Brainstorming sessions were held with CDC staff working on the integration of public health surveillance systems to identify additional subject areas that were overlooked in the analysis of existing data models. The subject area model was used to define the project scope, estimate the work effort, and develop the project plan.

In June 1999 an initial "class" diagram was created. A class is something about which data are collected. The class diagram is a depiction of the major classes of data within each subject area. It includes a description of the classes, as well as a description of their inter-relationships. This class diagram was reviewed with groups of CDC epidemiologists in July 1999 and revised based upon their feedback. Attributes (i.e., information about the classes) were added to the class model in August and reviewed within CDC and with state and local agencies in September and early October 1999. Based upon feedback from these review sessions the decision was reached to continue the enhancement of the data model by:

- Developing a public health process model to provide context and clarify scope for the data model;
- Adopting the HL7 Reference Information Model (RIM) representation of health-related activities;

- Validating the data model by using it to develop a prototype database based upon the information needed for a subset of CDC disease management and surveillance systems such as the National Electronic Telecommunications System for Surveillance (NETSS), the Sexually Transmitted Diseases Management Information System (STD*MIS), and the Laboratory and Epidemiological Public Health Information Tracking and Reporting System (LITS+).

From October to December the model was revised to include data constructs from the HL7 RIM and the Missouri Department of Public Health data model. Meetings were held with the NETSS project team to discuss the objectives of PHCDM and its potential application to the NETSS project. An approach to developing the process model was devised in December.

The need for explicit description and publication of the rationale and objectives for the model was acknowledged in January 2000. The PHCDM is critical to meeting the data standardization objectives of the NEDSS initiative. The various model-related activities have multiple objectives but their overall emphasis is on applying the data model to data standardization issues facing CDC and its partners.

The model is expected to undergo continual refinement as it is used. It is a living document that will need to be revised as public health information needs change, as our understanding of those needs are improved, and as available technologies increase the applicability of automation.

Goals and Objectives

The overarching purpose of the Public Health Conceptual Data Model is to document the information needs of public health and facilitate the development of data standards as part of the National Electronic Disease Surveillance System initiative. One might ask, how does a data model facilitate the development of data standards? That question is best answered by examining the goals and objectives for the development of the PHCDM.

PHCDM Goals

1. **Provide a framework for organizing data standards and guidelines**

   The initial CIPHER effort defined standards and guidelines for data representation and code values. It included specifications for representing concepts, such as dates, addresses, and person names as well as standard code lists for coded elements, such as race, ethnicity, and sex. The CIPHER standards can be linked directly to attributes in the data model. These attributes represent characteristics of particular classes of information and will eventually become fields in computerized information systems. The PHCDM provides a context for these standards. By describing this context for the CIPHER guidelines, CDC staff and public health partners working on data

standards are better able to envision the potential impact of their guidelines and the implications of that impact. Also, by examining portions of the PHCDM that do not map directly to CIPHER data standards, persons continuing CDC's work on data standards can readily determine additional areas for which data standards and guidelines are required.

2. Reduce development effort for computerized information systems used for public health

CDC and its state and local partners develop computerized information systems for use in public health. Invariably these development projects expend effort on requirements gathering, data requirement analysis, and database design. The PHCDM will significantly reduce the effort expended by these development projects by providing reusable data analysis and database design and developing a common starting platform that can be used or modified as necessary, resulting in reduced development time and cost.

By using the PHCDM at the beginning of analysis and design each individual development project team can avoid rethinking data analysis and database design issues. New findings or required revisions noted by a given development team can be reflected in the PHCDM as part of its routine maintenance so that they become available to other project teams for reuse.

3. Enhance data sharing through consistency

An additional advantage gained by using the PHCDM on multiple development efforts is the reuse of PHCDM constructs in database design. This reuse increases the consistency in data meaning and representation across independently developed software systems. This increase in data consistency will make it easier to share data, where appropriate, across the suite of information systems used in public health. Use of the PHCDM will minimize the need for complex data mapping and transformation processes prior to sharing or reusing data.

Data consistency will permit data comparisons and linkages to be established across multiple systems. Data consistency is an important system characteristic that will facilitate the analysis of trends and detection of emerging public health problems, and the use of information for setting public health policy.

4. Represent public health data needs to national standards setting bodies

A critical aspect of this project is the ability to collaborate with national health informatics standards setting bodies to define standards for the exchange of information among public health agencies and healthcare providers. Health Level Seven (HL7) is an essential part of the effort to establish standards for information exchange in healthcare. The HL7 message development methodology includes processes for constructing a reference information model, defining datatypes, coordination of vocabulary domains, and

specification of Extensible Markup Language (XML) data interchange standards.

HL7 is the undisputed leader in the establishment of standards for interoperability among computerized information systems in healthcare. A key aspect of the HL7 methodology is the HL7 Reference Information Model (HL7 RIM). The HL7 RIM is the source for the data content of all HL7 version 3 standards. HL7 has defined an extensive process for ensuring that the information needs of its constituencies are reflected in the RIM. It is extremely important that the public health information needs be reflected in the HL7 RIM.

The PHCDM will reuse as much of the RIM as is applicable to the needs of public health and it will be the source of additions or modification to the RIM, to represent unique public health data requirements. By including public health needs in the HL7 RIM, we ensure that those needs are available to the large body of information system vendors and provider organizations participating in HL7, so that they can include them in the design of healthcare information systems that serve as original sources of data relevant to public health. CDC staff will also explore ways to coordinate public health information needs with other accredited standards organizations.

5. Facilitate collaboration between CDC and its state and local partners in public health

Collecting, analyzing, and reporting data related to public health is done at local, state, and national levels. These data are used to monitor the health of the public, identify public health problems and priorities, take immediate public health action to prevent further illness, plan appropriate longer-term interventions, and develop public health policy. Data on diseases and conditions of importance to public health are reported on a regular basis to Local Health Departments, which pass these data onto the states, which in turn report voluntarily to CDC. CDC aggregates and reports these data back to state and local authorities and to the public on a regular basis. The system of surveillance, intervention, and planning requires collaboration between all of the parties involved. Some states and localities rely upon and use CDC-supplied software applications to electronically collect, report, and analyze cases of notifiable conditions. Other states have the resources to develop such applications on their own. Inconsistencies in data definitions, formats, and code sets make the integration of data from the various sources and systems difficult. The PHCDM will serve as a vehicle for collecting and reconciling the information needs for public health at all levels. The data standards defined by the PHCDM and CIPHER can be used by all parties involved in public health. Through collaboration with state and local entities, the PHCDM will lay a foundation for a unified view across the full spectrum of public health activities.

PHCDM Project Objectives

The goals of the PHCDM provide a basis of planning and evaluating the assignment of resources for the PHCDM project. The goals of the PHCDM are long term and will be achieved over the life of the PHCDM. The objectives of the PHCDM project are more near-term and action-oriented. The PHCDM is to be used for several years. The goals are expected to endure while projects and project objectives are established on an annual or semi-annual basis.

The following objectives were achieved in 1999:

- Created a data model for public health based upon analysis of the database structures of current CDC disease surveillance systems, CIPHER data standards, the HL7 RIM, and other health-related information models, and defined a maintenance process for the model;
- Reviewed the PHCDM with epidemiologists from within certain program areas of CDC and from state and local public health agencies and updated the PHCDM based upon their critique;
- Selected a computerized information system development project to make use of the PHCDM and CIPHER data standards within CDC and develop a workplan for applying the standards to the development effort.

Immediate next steps in 2000 are:

- Document and distribute copies of the PHCDM;
- Validate the PHCDM by using it to create a prototype database based upon the information requirements of a selected set of CDC surveillance information systems;
- Participate in the HL7 RIM harmonization process to introduce public health-related information needs and to more closely align the PHCDM with the HL7 RIM;
- Participate in HL7 to use the HL7 RIM in the development of data exchange messages specific to public health;
- Develop a high-level process model for public health to provide context for the PHCDM;
- Continue to coordinate with state and local public health entities to ensure that their information needs are represented in the PHCDM.

Project Scope

The ultimate intention in developing the PHCDM is to represent the information needs of all public health activities and entities. The immediate scope is to document the information needs of public health surveillance including case identification, reporting, investigation, intervention, and follow-up. For the purpose of this project, public health surveillance will be viewed from the perspective of CDC as a whole, not from any particular disease or condition/program. To the extent possible the scope will include the perspectives of State and Local Health Departments and other partners in public health surveillance.

In addition to the data model, the scope of PHCDM project includes development of a high-level process model for public health, harmonizing the model with the HL7 RIM, and development of a prototype database based upon the PHCDM. In addition the PHCDM project team will provide assistance to CDC system development teams and state and local public health entities using the PHCDM in their information management initiatives. These activities are some of many anticipated sources of enhancements to the PHCDM.

Ongoing development of the PHCDM will follow an iterative approach. The model will be continuously maintained and published on an annual basis. Each year projects will be defined that specify the scope, process, and products to be produced. Future project scopes might focus on areas such as public health intervention programs, public health financing, or public health research. The public health high-level process model will be used to help scope future projects.

GUIDE TO UNDERSTANDING THE PUBLIC HEALTH CONCEPTUAL DATA MODEL

Guide to Understanding the Public Health Conceptual Data Model

The purpose of this section is to provide guidance to the reader to aid in understanding the PHCDM. As a conceptual data model the PHCDM attempts to present the information needs of Public Health in a way that lends itself to validation by subject matter experts and has sufficient rigor and formality to be used by experts in information technology in the development of database design specifications.

To meet this objective the PHCDM avoids many of the details generally found in logical and physical data models such as normalized data structures, primary and foreign keys, and specification of field details such as length and decimal positions. It makes extensive use of examples and explanatory text to describe model classes and attributes. Its primary goal is to ensure that the concepts of importance to public health are adequately depicted and documented.

The PHCDM uses a fairly high level of abstraction to document public health concepts. This high level of abstraction extends the applicability of the model and minimizes the need for maintenance. However, it can sometimes make it difficult for subject matter experts to recognize specific details they might expect to find in a public health data model. For example, "Where are items of interest to public health, such as risk behaviors, infectious or environmental agents, drug-resistance, case investigation, or populations-at-risk?", might be questions a subject matter expert might have after first browsing the data model. Rest assured that these concepts are indeed included in the PHCDM. This guide to understanding the PHCDM is intended to assist you, the reader, in finding answers to these questions and others you may have as you review the model.

The first step in understanding the PHCDM is to become familiar with its components, data model terminology, and the standards and conventions used. The PHCDM uses the Unified Modeling Language (UML) modeling conventions. UML is a widely used data modeling standard maintained by the Object Management Group. References to information sources about UML can be found in the bibliography included at the end of this document. The following components of UML are used in this model:

- Subject Areas
- Classes and Relationships
- Attributes and Datatypes

Subject Areas

A subject area is a useful partitioning of a model into a cohesive collection of classes. Subject areas are a way to subset a model into chunks that permit the

model to be more readily digested. There are four subject areas in the PHCDM: Health-related Activities, Locations, Materials, and Parties.

The ***Health-related Activities subject area*** contains information about services, conditions, and actions of interest to public health. A health-related activity might be an observation, an intervention, a referral, or a notification. Typical examples of health-related activities include observation of patient signs and symptoms, clinical diagnoses, surgical operations, laboratory tests and results, as well as public health notifications, case or contact investigations, population-oriented health education campaigns, food item recalls. Cases, case reports (notifications), and outbreaks are classes or health-related activities of particular importance to public health surveillance.

The ***Locations subject area*** contains information about addresses associated with Parties, Health-related Activities, or Materials. Location information may be a postal location, a telecommunication location, or a physical location. Typical examples of locations include street addresses, post office boxes, telephone numbers, e-mail or web-site addresses, geographic coordinates, and spatial references such as three miles east of town on interstate 95.

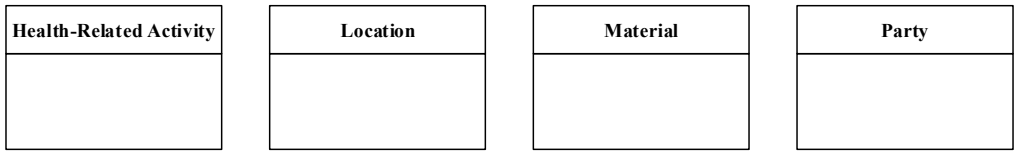
The ***Materials subject area*** contains information about substances, equipment, products such as food and medication, physical entities, and other tangible items of interest to public health that are associated with health-related activities or Parties. Typical examples of materials include food items, pesticides, blood samples, specimens, medications, durable medical equipment, prosthetic devices, and medical supplies. Physical entities of interest to public health may include such items as a lake, pool, ship, or airplane that are potential sources of exposure to health hazards.

The ***Parties subject area*** contains information about the participants of health-related activities. A party may be an individual person or non-person living organism, or a formal or informal organization. Typical examples of parties include patients, physicians, public health nurses, epidemiologists, hospitals, and laboratories, as well as organizations such as the Association of State and Territorial Health Officials, the Council of State and Territorial Epidemiologists. Groups of parties with common characteristics, such as smokers or children under 5 years of age, are also included.

Classes and Relationships

There are 29 classes in the PHCDM. A Class is anything about which information can be collected. Classes can be persons, places, things, concepts, or events. There are four core classes in the PHCDM. The four core classes correspond with the four subject areas. The four core classes are Health-related Activity, Location, Material, and Party. Classes are depicted in the data model diagram by a rectangular box with a line dividing the box into two vertical sections. The name

of the class appears in the top section of the box. The following diagram illustrates the four core classes of the PHCDM.

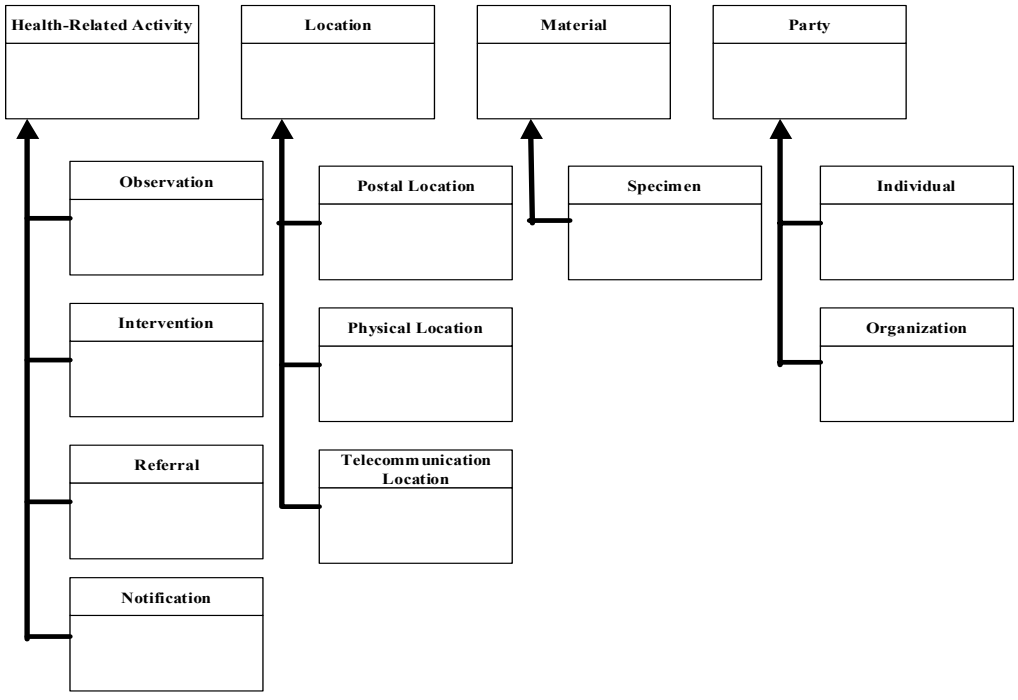


The 29 classes of information in the PHCDM are all interrelated. Direct relationships between classes are depicted in the model diagram by lines connecting the related classes. The UML modeling language defines many ways in which classes can be related. The PHCDM uses three methods of relating classes: Supertype/Subtype Relationship, Relationship Association, and Participation Association.

Supertype/Subtype Relationship

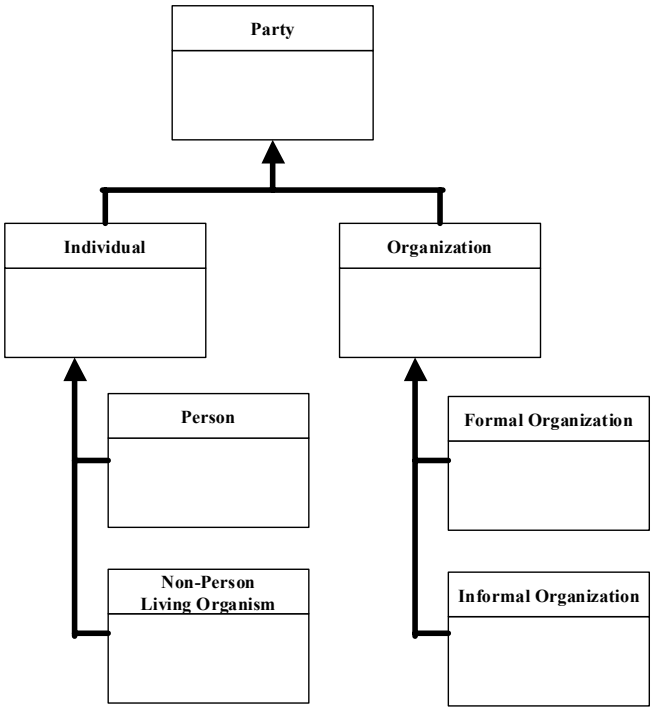
The supertype/subtype relationship is used when generic concepts represented by a class are further represented in one or more specialized classes depicting a subset of the generalized concept. In the supertype/subtype relationship the more generic class, referred to as the supertype, has one or more specialized subtype classes.

Each of the four PHCDM core classes is a generic supertype class with one or more related subtype classes. The supertype/subtype relationship is depicted on the data model diagram by a line drawn between the subtype and the supertype. The line has an arrowhead on one end pointing to the supertype. The following diagram depicts the supertype/subtype relationships to the PHCDM core classes.



This hierarchical structuring of core model concepts makes it easier to understand the model. Once you become familiar with the four core concepts the process of becoming familiar with the specialized concepts in the class hierarchy is very simple. For example, looking at the Health-related Activity hierarchy the classes Observation, Intervention, Referral, and Notification are each special types of Health-related Activity. An Observation *is* a Health-related Activity. All of the data we capture about a Health-related Activity are data that are also collected about an Observation. However, the Observation Class has additional attributes that are captured only when the Health-related Activity is of type Observation. Similarly, an Intervention *is* a Health-related Activity. It too inherits all of the characteristics of Health-related Activity and supplements that information with information unique to an intervention.

A subtype of one class may also be the supertype of one or more of its own subtypes. This is illustrated in the following diagram depicting the full hierarchy of the concepts modeled under Party.

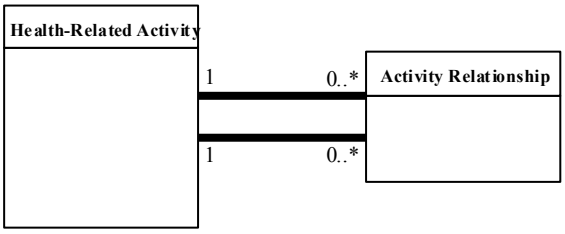


The Individual and Organization classes are subtypes of the supertype class Party. Individual is also the supertype for the Person and Non-Person Living Organism subtypes, and Organization is the supertype for the Formal Organization and Informal Organization subtypes.

Relationship Association

A relationship association is a special type of relationship used in the PHCDM to reflect the relationship an instance of a core class or its subtypes has to another

instance of the same core class or its subtype. These relationships are represented in the PHCDM by four "relationship" classes, each associated with one of the four core classes. The four relationship classes are: Activity Relationship, Location Relationship, Material Relationship, and Party Relationship. The relationship associations are depicted in the model diagrams by a rectangular box representing the relationship class and a pair of association lines connecting the relationship class to the core class that is linked by the relationship. The Activity Relationship is illustrated in the following diagram.

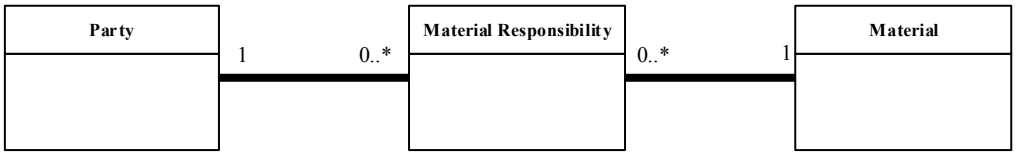


The symbols "1" and "0..*" that appear on the association lines depict the multiplicity of the association between the relationship class and the core class. Multiplicity is an indication of the number of instances of a class that is capable of being involved in any one association. In this case, the multiplicities indicate that each instance of the activity relationship is associated with one and only one health-related activity and that each health-related activity is associated with zero or more activity relationships. Since there are two associations, each with the same multiplicity, an instance of an activity relationship class is always associated with two instances of a health-related activity class. A single health-related activity may be associated with zero or more activity relationships relating it to another health-related activity.

For example, an observation reflecting the presence of a vectorborne disease (with mosquitoes as the vector) can be linked to interventions such as spraying insecticide, using mosquito repellents, and issuing mosquito nets. Both observation and intervention are subtypes of health-related activity.

Participation Association

The participation association is a special relationship used in the PHCDM to depict the relationships that exist between the core classes. Each core class has a many-to-many relationship to all of the other core classes. For example, an instance of the party class may be related to many instances of the material, location, and health-related activity classes. Instances of material, location, and health-related activities classes may also be associated with many instances of the party class. The participation association is depicted using a participation class. There are five participation classes in the PHCDM: Actor Participation, Target Participation, Party Location Participation, Material Responsibility, and Material Location Participation. The following diagram depicts the Material Responsibility participation association:



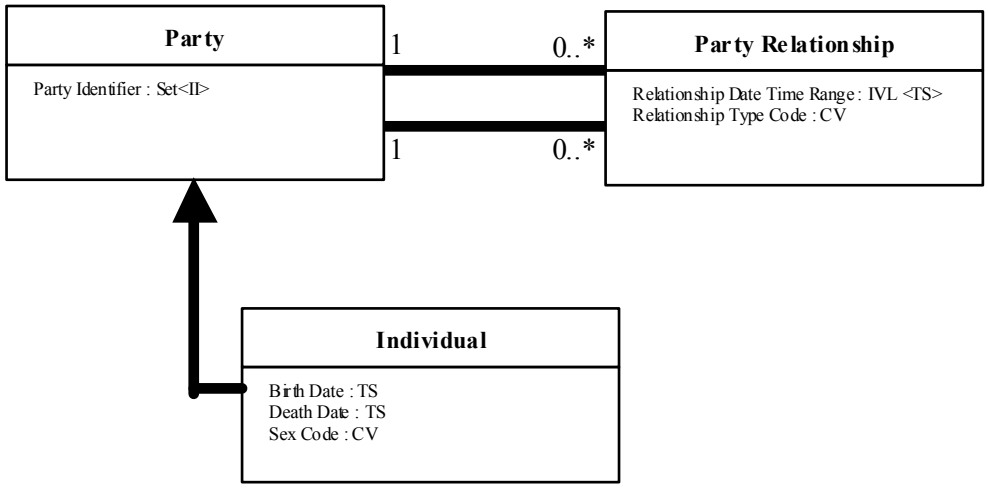
The multiplicities on the associations indicate that a material responsibility is always associated with one party class and one material class. Material and party classes may be associated with zero or more material responsibility classes.

For example, if material is a specimen it may be important to reflect the party that was the source for the specimen and the party that was responsible for obtaining the specimen. This would be captured as two instances of the material responsibility class, one for each party class, each associated with the same specimen material class.

Attributes and Datatypes

Attributes are the specific items of data that can be collected for a class in the PHCDM. Each attribute has a name, a description, and a datatype assignment. An attribute name suggests the meaning of the attribute, while the description defines it, provides examples, and includes relevant discussion. The datatype assigned to an attribute extends the definition of the attribute. A datatype is a specification of the allowed format for the values of an attribute.

Attributes and their datatype assignments are shown in the data model diagram by listing them in the lower section of the rectangle representing the class. The following diagram is an example of three classes and their attributes.



Attributes of a supertype are inherited by its subtypes. In this example, the attribute Party Identifier in the supertype Party class is also an attribute of the subtype class Individual (as well as all the other subtypes of Party).

Attribute names follow the form:

[Class Name] [{Qualifier Name}] Attribute-Type Name

The square brackets […] around class name indicates that the class name may be omitted from the attribute name. The curly brackets {…} around qualifier name combined with the square brackets indicate that there may be zero, one or more qualifier names. Every attribute has an attribute-type name. The attribute-type name provides an indication of the type of data the attribute conveys. The attribute-type names used in the PHCDM are:

- Amount
- Code
- Date
- Description
- Identifier
- Name
- Number
- Quantity
- Text
- Time-Range
- Value

The datatype assigned to an attribute is represented in the data model diagram by the inclusion of the datatype name following the attribute name, separated by a colon (":"). The datatype assignments in the PHCDM appear in one of three forms:

- Attribute Name : Datatype Name
- Attribute Name : SET<Datatype Name>
- Attribute Name : IVL<Datatype Name>

The collection of datatypes used in the PHCDM is drawn from the set of datatypes defined for HL7. A complete list of the datatypes and their descriptions are included in an appendix. The datatype name used in the model diagram is the short name of the datatype.

Most attributes are expected to take on one value at a time. However, if the datatype name is preceded by "SET" and enclosed in brackets, it is an indication that the attribute may repeat. That is, there may be a set of one or more values for the attribute. In the sample model the attribute Party Identifier is a set. That is an indication that there may be multiple identifiers for a single instance of a Party class.

If the datatype name is preceded by "IVL" it is an indication that the attribute represents an interval of values from low to high. In the PHCDM the IVL prefix to a datatype has been limited to intervals of time. In the sample diagram above the attribute Relationship Date Time Range (for the class Party Relationship) is an

interval. The implication is that the relationship date time represents a start and an end date time range.

Key Concepts

To fully appreciate the richness of the PHCDM it is necessary to understand a few key concepts. Although the model contains 4 subject areas, 29 classes and nearly 100 attributes it is still fairly easy to digest. The key concepts included in this section are also described in the model itself. They are described here because of their importance to understanding the data model. The key concepts are:

Health-related Activity Mood Code

The activity mood code is an attribute of the health-related activity class. The attribute is critical to determining the perspective of an instance of the health-related activity class. The activity mood code captures the meaning or context of the activity. Possible values of the activity mood code are:

- **Fact** an activity that has occurred
- **Command** an activity that has been ordered
- **Master** a table entry of possible activities
- **Definition** an algorithm for describing an activity
- **Intent** a goal for an activity

Instances of the health-related activity class with an activity mood code of "master" might be used to represent the list of health conditions under public health surveillance. Other instances of the health-related activity class with an activity mood code of "definition" might be used to represent the case definitions established for the list of health conditions under public health surveillance. And finally, instances of the health-related activity class with an activity mood code of "fact" would be used to capture actual cases or occurrences of health conditions.

The activity mood code makes the health-related activity a very versatile concept and minimizes the number of classes, attribute, and relationships need to represent a wide range of concepts.

Actor / Target Participation

Actor participation is a class that captures the relationship between a party and a health-related activity. Target participation is a class that captures the relationship between a health-related activity and party, material, or location. The actor participation is meant to capture information about parties that carry out the action indicated in health-related activity. The activity type code indicates the nature of

the party's participation. The type code identifies the various types of actors or roles a party may assume. Examples of actor roles might include disease intervention specialist, outbreak investigator, primary care physician, attending physician. This structure permits the party class to assume many roles without having to introduce specialized classes and attributes to the model. Target type code in the target participation class performs a similar function for parties, materials, and locations. The target is not an actor in the health-related activity but is otherwise involved, either as the object of the activity (e.g., the patient) or involved in a more passive mode (e.g., the medication in a medication intervention). Targets may include persons with disease of interest or persons exposed to disease.

Observation

The observation class represents both the act of observing and the results of the observation. It represents objective, subjective, and derived observations. It includes the concepts of test, assessment, vehicle condition, diagnosis, party condition, and health status inquiry. Observations may be made about parties, materials, or locations. Observations may be made about other health-related activities, including other observations. This is a very rich concept. Essentially all facts not included as attributes elsewhere could conceivably be represented by the observation class. Because an observation may take many forms (e.g., numbers, waveforms) depending on the target of the observation and the type of phenomenon being observed, it has a datatype of "any".

Individual Non-Person Living Organism

This subtype of the Party class is intended to be limited to those living organisms that are identified individually. The class is necessary to permit the tracking of health-related activities where the target of the activity is not a person. For example, the target might be an animal such as a pet dog, or a circus elephant. It is unlikely that a life form like a bacterium or a virus would ever be individually identified and so it is not likely for them to be included as non-person living organisms. Even insects, like mosquitoes, are unlikely to be included as non-person living organisms. However, if a particular mosquito is isolated and individually identified it might be valid to include it as a non-person living organism. In most cases, however, non-person living organisms that are not routinely individually identified would be represented either as specimens (a subtype of material) or as members of an informal organization.

Informal Organization

An informal organization is intended to represent any defined population. The members of the population do not have to be enumerated. The population is treated as a unit with respect to its participation as a actor or target in a health-related activity. Informal organizations could include a human family, a herd of

cows, people that smoke, dead crows in New England, or any other defined population.

Implementing the PHCDM

Understanding the PHCDM requires a discussion of how it is to be implemented. This section includes:

- a brief description of the hierarchy of data model types, identifying where a conceptual data model fits within a family of interrelated data models;
- a discussion of the process of deriving multiple physical database models from a common conceptual data model;
- a description of a harmonization process designed to facilitate ongoing enhancements to the PHCDM.

Hierarchy of Data Model Types

A data model is documentation of data from a particular domain, for a specific purpose, using a formal specification. A data model has both a graphical expression and a supporting textual expression or data dictionary. Each expression of a data model uses a pre-defined formalism of symbols, semantics, and rules of construction.

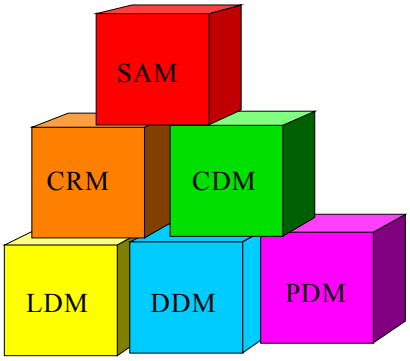
The purpose of a data model is to aid in understanding data in a particular domain. A data model communicates the modeler's understanding of data and allows that understanding to be assessed by others. A data model can be useful in reconciling multiple perspectives of data because it reveals the underlying assumptions, semantics, and constraints expressed in multiple models and requires their harmonization into a single specification. A common use of a data model is to document a database design (existing or planned) so that the design may be evaluated.

There are multiple types of data models. Each type of data model has characteristics that make it more useful than other types for a particular purpose. There are models that are very useful for high-level planning and project definition. These models tend to minimize technical details and focus instead on delineating and defining subject areas and classes of information of interest to executives, high-level decision-makers, and subject matter experts. These models are *not* useful for evaluating or implementing a database design. For a data model to be useful to a database design activity, it needs to include technical details such as database key structures, datatypes, and the physical properties of tables and columns. However, this latter type of model contains too many detail and technical artifacts to make it useful for high-level planning and decision making.

A data model should be constructed with a specific purpose in mind. The model developer should choose from among the various data model types the type that is most suitable for the intended audience and use. The following diagram identifies a hierarchy of six data model types:

Hierarchy of Data Model Types

- **Subject Area**
- **Class Relationship**
- **Conceptual Data**
- **Logical Data**
- **Database Design**
- **Physical Database**



Each type of model is described below.

1. **Subject Area Model (SAM):**

   A subject area model contains only subject areas and their connections, and usually serves as a model for a large domain, such as the entire enterprise or a major functional area. It is used for high-level planning and setting of project scope.

2. **Class Relationship Model (CRM):**

   A class relationship model contains only subject areas, classes, and relationships, and generally depicts a limited domain, such as a single project or enterprise business area. It is used for high-level analysis and estimation of project size.

3. **Conceptual Data Model (CDM):**

   A conceptual data model contains subject areas, classes, attributes, datatypes, and relationships, and generally models a project-specific domain, such as public health, finance, or material management. It results from a relatively detailed level of analysis and is often a primary project deliverable.

4. **Logical Data Model (LDM):**

   A logical data model contains subject areas, normalized classes, atomic attributes, relationships, and candidate/primary keys, and usually serves as a model for an enterprise-specific implementation of a project-specific domain. It signifies the completion of the most detailed level of data analysis and the beginning of database design.

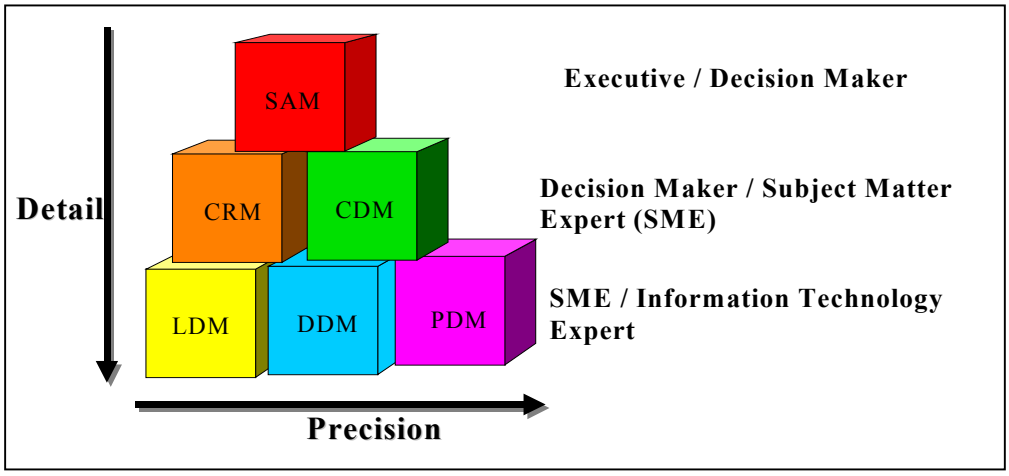
5. **Database Design Model (DDM):**

   A database design model contains table spaces, tables, columns, datatypes, and primary/foreign keys, and generally represents an existing or planned database of a computerized information system. It indicates the completion of database design and the beginning of database construction.

6. **Physical Database Model (PDM):**

   A physical database model contains the data definition language (DDL) required to create tables and indexes, as well as data base management system (DBMS)-enforced constraints. It is a machine-processable specification of an existing or planned database of a computerized information system, and corresponds to the final step of database design and construction.

In the diagram of the hierarchical taxonomy of models below, the six model types are arranged from top to bottom by level of detail and target audience, and from left to right by degree of precision and rigor of specification.



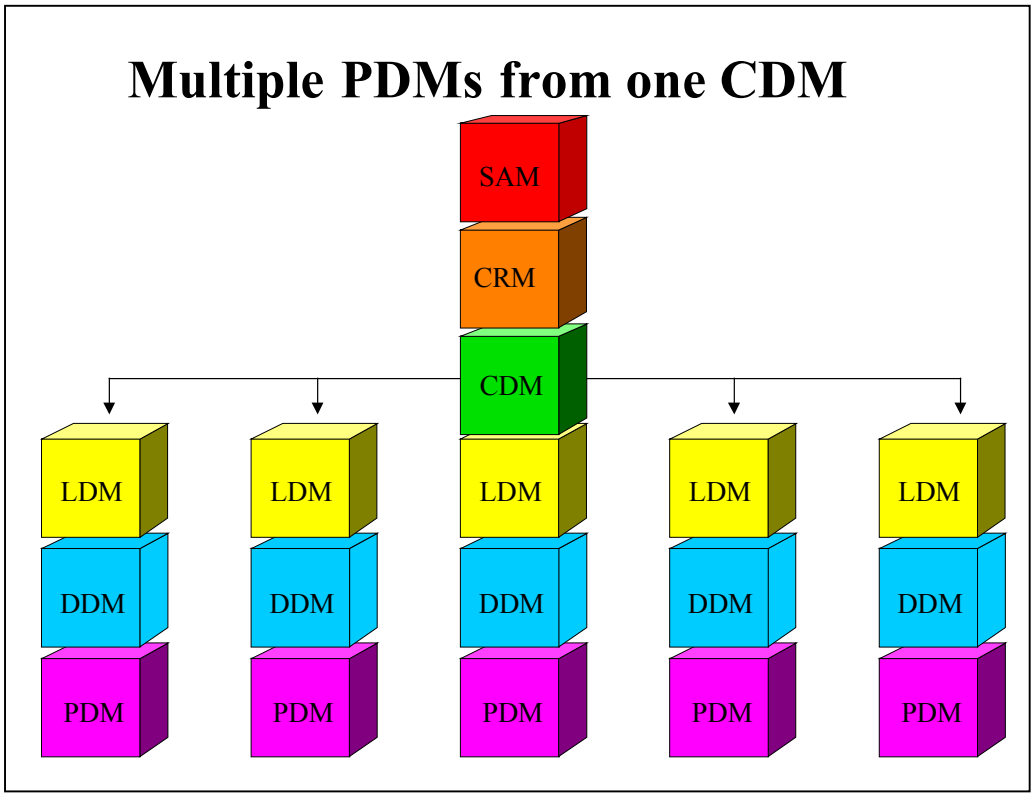
The top three models (subject area, class relationship, and conceptual) are technology-independent and may be applicable to multiple organizations performing the same functions that are supported by the data model. The lower

three models are more technology-specific and may be applicable to multiple organizations that share the same business rules.

The PHCDM is a particular instance of a conceptual data model. A conceptual model was chosen as the style for the PHCDM because of the desire to have a model that is technology-independent and applicable to multiple organizations. A conceptual data model also has sufficient detail to be useful as a definition of information requirements in a specific domain (i.e., public health). The intended audiences for this model are decision-makers and subject matter experts in public health and information technology experts responsible for requirements analysis and design of computerized information systems for use in public health.

Multiple Physical Data Models from one Conceptual Data Model

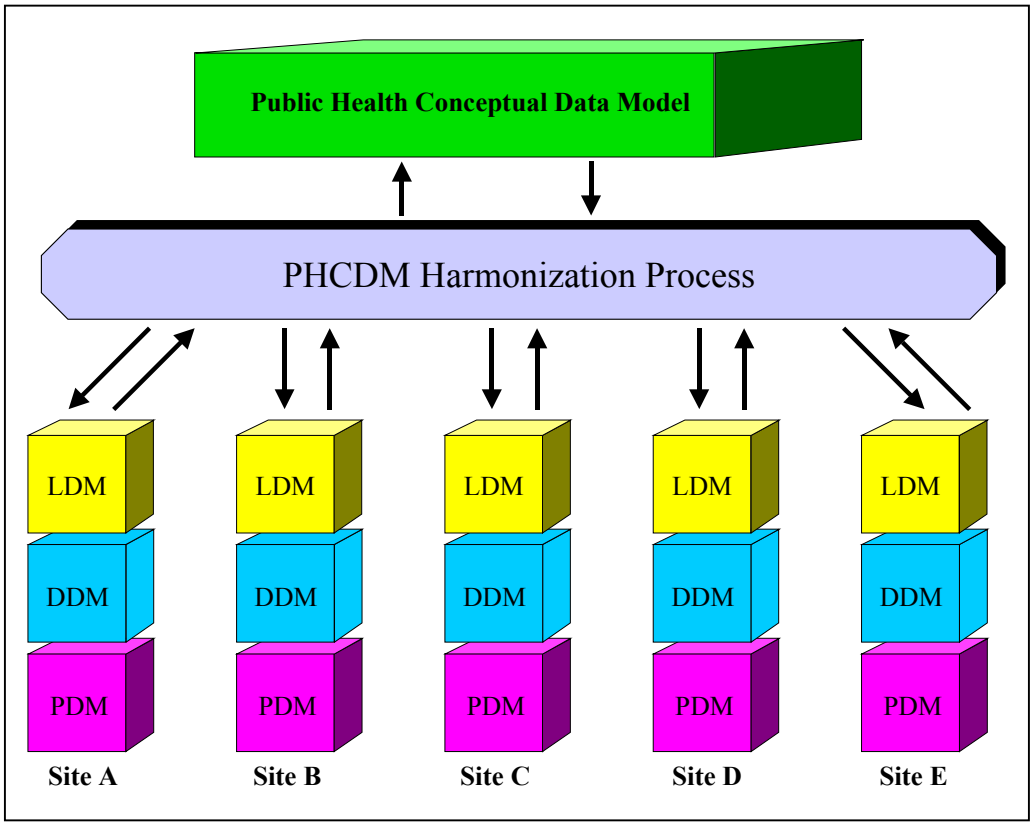
The PHCDM should be considered the base model for multiple physical data models. The process of building a physical database model is expected to use the PHCDM as input to creation of a logical data model (LDM). The LDM may be derived from the entire PHCDM, or simply from a subset of it. Its design is constrained by the business rules and project scope of the entity implementing the model. These constraints will differ from implementation to implementation, resulting in multiple LDMs that are semantically equivalent but that may vary from each other on a technical level (i.e., choice of class identifiers, degree of normalization, and relationship constraints). The database design models (DDM) and the physical database models (PDM) will be derived from the LDMs as depicted in the following diagram:



Each LDM-DDM-PDM triad is a separate implementation of the same CDM. In the case of the PHCDM, these might represent separate or collaborative implementations by Local and State Public Health Departments or by CDC program areas. The physical database models might target different technologies and enforce different business rules, however because they originated from a common CDM, the semantics of the data content would be equivalent on a conceptual level. This will greatly facilitate the sharing of information between these independently developed databases.

PHCDM Harmonization Process

The PHCDM will be enhanced in an ongoing fashion, based upon input from sites that use it in their information system development initiatives. As organizations or programs implement the PHCDM, they will invariably identify omissions and perhaps errors in the model. As errors and omissions are identified, they should be brought to the attention of CDC in the form of proposed changes to the PHCDM. PHCDM change proposals submitted from the multiple sites implementing the model will be considered together and harmonized to ensure that conflicting change requests are reconciled prior to being applied to the model. This harmonization process is illustrated in the following diagram.



The current plan for the PHCDM is to re-issue the model on an annual basis, incorporating as many of the change proposals as it is feasible to handle in one year. The mapping of the PHCDM to CDC-developed information systems like

NETSS, STD*MIS, and LITS+ has already identified possible enhancements for the 2001 version of the PHCDM. Additional input is expected from harmonization of the PHCDM with the HL7 RIM and from use of the model in additional activities related to the National Electronic Disease Surveillance System initiative.

An attempt will be made to minimize the impact that enhancements to the PHCDM have on the sites where the model has already been implemented. Changes from release to release will be highlighted in each publication. An assessment of the impact of the change and suggestions for forward migration will be included in each release.

PUBLIC HEALTH CONCEPTUAL DATA MODEL

Public Health Conceptual Data Model

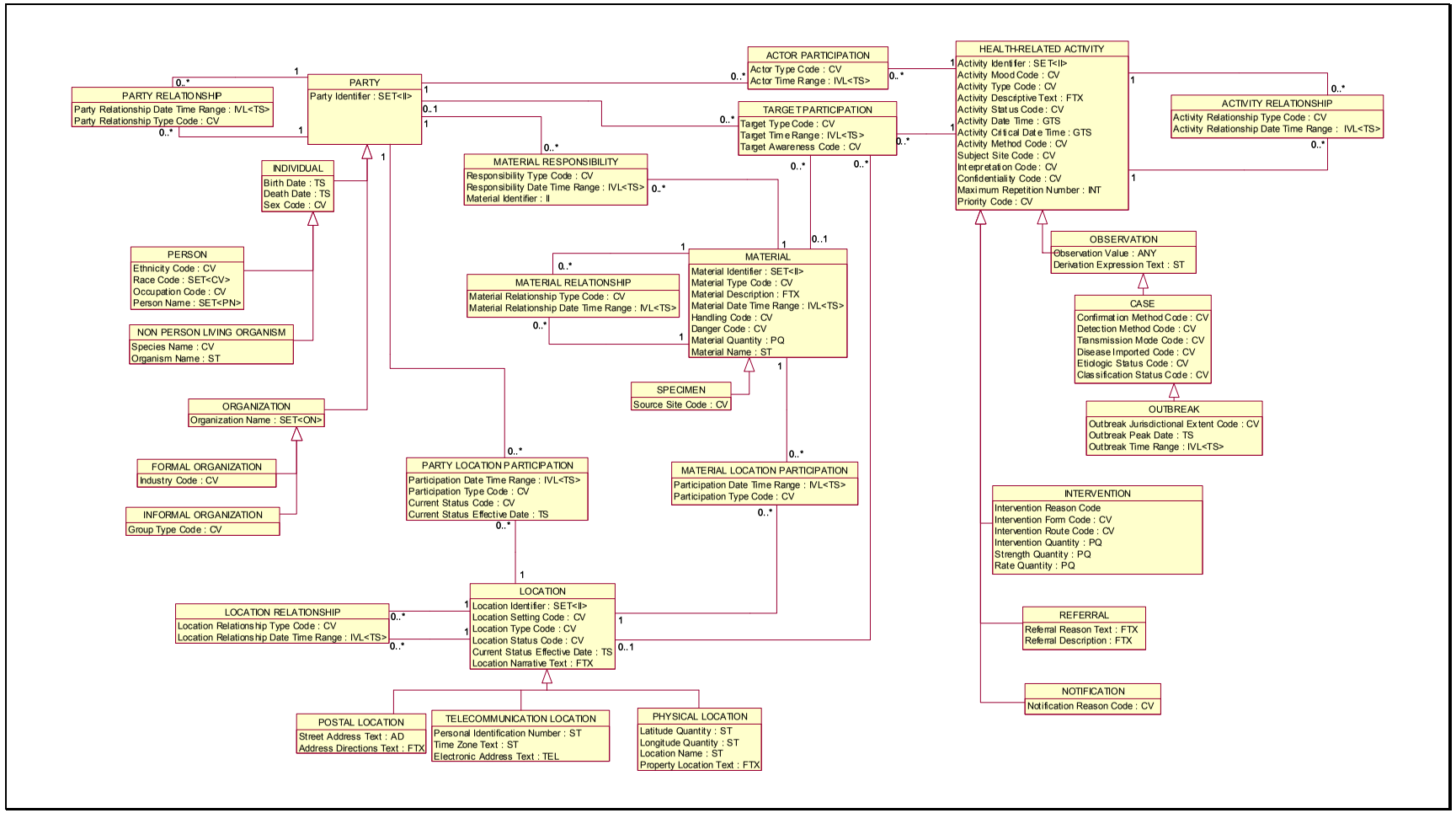


Figure 1. Public Health Conceptual Data Model Diagram

The model consists of 4 subject areas and contains 29 classes. These subject areas are based on the most general categorization of the data relevant to public health concerns. The four subject areas and the classes they contain are listed below.

HEALTH-RELATED ACTIVITIES

- ACTIVITY RELATIONSHIP
- ACTOR PARTICIPATION
- CASE
- HEALTH-RELATED ACTIVITY
- INTERVENTION
- NOTIFICATION
- OBSERVATION
- OUTBREAK
- REFERRAL
- TARGET PARTICIPATION

LOCATIONS

- LOCATION
- LOCATION RELATIONSHIP
- MATERIAL LOCATION PARTICIPATION
- PARTY LOCATION PARTICIPATION
- PHYSICAL LOCATION
- POSTAL LOCATION
- TELECOMMUNICATION LOCATION

MATERIALS

- MATERIAL
- MATERIAL RELATIONSHIP
- MATERIAL RESPONSIBILITY
- SPECIMEN

PARTIES

- FORMAL ORGANIZATION
- INDIVIDUAL
- INFORMAL ORGANIZATION
- NON-PERSON LIVING ORGANISM
- ORGANIZATION
- PARTY
- PARTY RELATIONSHIP
- PERSON

Detailed descriptions of the classes and attributes are contained in the sections for each subject area.

Health-related Activities Subject Area

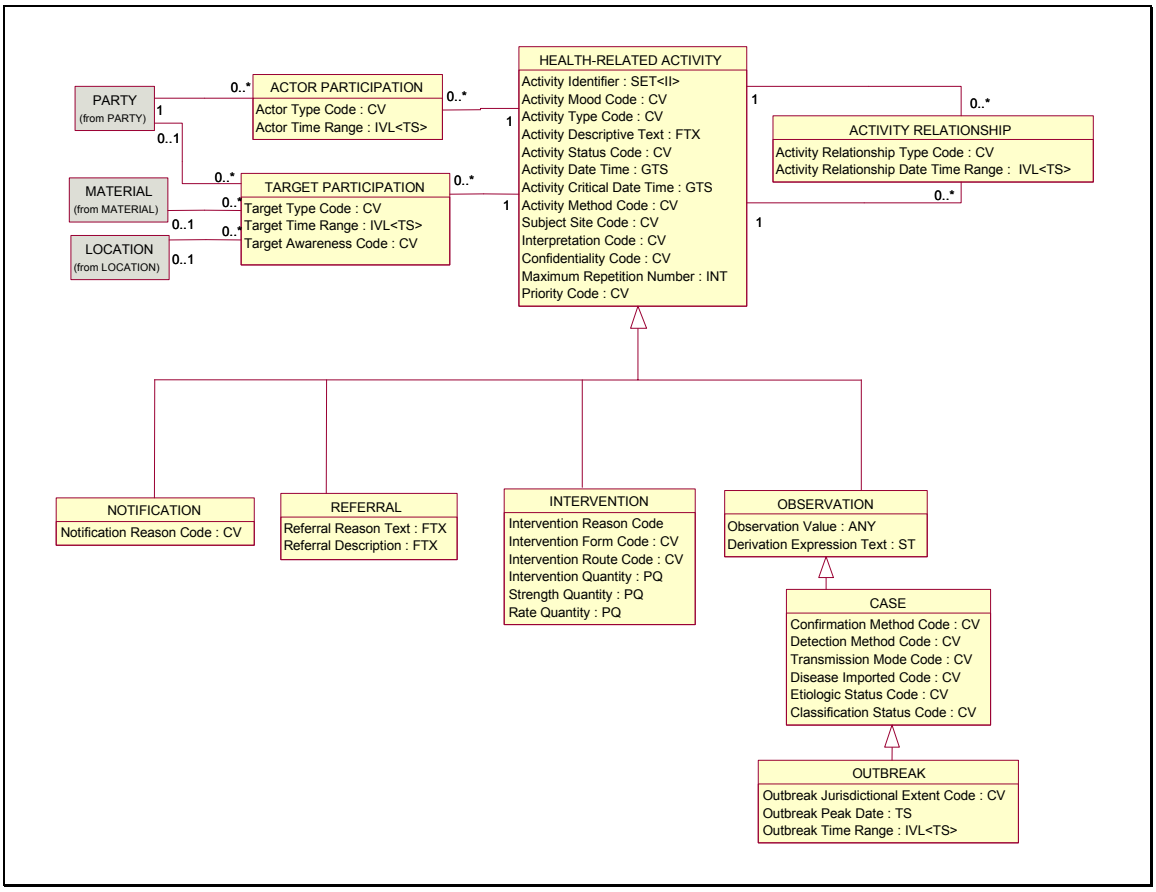


Figure 2. Health-related Activities Subject Area Diagram

The classes and attributes of the Health-related Activities subject area are described below.

Class: **ACTIVITY RELATIONSHIP**

Associated with: **HEALTH-RELATED ACTIVITY**

Description of: **ACTIVITY RELATIONSHIP**

Activity relationship captures the relationship between a pair of health-related activities. Generally, relationships between health-related activities fall into three categories: an activity can be comprised of component activities; one activity can cause another; one activity can be associated with another for any number of reasons.

Virtually any activity can be decomposed into its parts. In public health, an outbreak of a particular disease can be composed of multiple individual cases of a particular disease. To take a medical example, consider a surgical procedure, e.g., a laparoscopic cholecystectomy. This action consists of many smaller actions that must occur in the right order and relation to each other. In the case of an invasive surgery, preoperative preparation may be required as a precondition, while anesthesia is conducted in parallel to the entire surgical procedure.

Causal associations are used to provide explanations for actions. For example, an episode is defined as a case of a particular disease (event reportable to public health) because of the results of a clinical evaluation combined with laboratory test results. (Note that the definition of the case specifies these criteria.) Another example is the instance of a test that was performed because of the results of two earlier tests.

The notion of "associated with" is more general than "causal" and "comprised of" associations. For example, in public health, a reportable case of disease is commonly associated with multiple observations. These observations record such items as specific behaviors that put the person at risk, the person's visits to locations where they might have been exposed, or the test results that indicate the person has a particular disease.

Associations for: **ACTIVITY RELATIONSHIP**

relates (1,1) :: HEALTH-RELATED ACTIVITY :: is_target_for (0,n)

Attributes of: **ACTIVITY RELATIONSHIP**

Activity Relationship Type Code : CV

The code that reflects the nature of the relationship that exists between two or more associated health-related activities. The possible values include "comprises", "causes", and "is associated with". An example of a "comprises" relationship is a case definition that is *comprised* of laboratory tests, symptoms, and other qualifying criteria. An example of a "causes" relationship is a case notification *causes* a case investigation. An example of an "is associated with" relationship is an outbreak and the *associated* cases.

Activity Relationship Date Time Range : IVL<TS>

The period of time during which the relationship between the two activity instances is effective.

Class: **ACTOR PARTICIPATION**

Associated with: **HEALTH-RELATED ACTIVITY**

PARTY

Description of: **ACTOR PARTICIPATION**

Actor participations include the active roles played by a party in the health-related activity. Examples include an organization that provides physical therapy services, a person who performs a surgical procedure, a public health worker who tracks contacts of an infectious case, a person who conducts a test, a person who conducts an interview.

Additional examples of actor participations are: a) the part played by an epidemiologist or CDC program staff (party) in generating a public health case definition; b) the part played by a provider, State or Local Health Department (party) in the notification of a case.

Associations for: **ACTOR PARTICIPATION**

associates_to (1,1) :: HEALTH-RELATED ACTIVITY :: **associates (0,n)**

associates_to (1,1) :: PARTY :: **associates (0,n)**

Attributes of: **ACTOR PARTICIPATION**

Actor Time Range : IVL<TS>

The time range during which the associated party participated in the health-related activity while taking on the role indicated by the specified actor type code value.

Actor Type Code : CV

Identifies the particular function or a set of functions that a party performs in the health-related activity. Note that the actor type code designates the actual function performed in a particular health-related activity in distinction to other roles or occupation. Examples of actor type codes might include case investigator, interviewer, and disease investigation specialist.

Class: **CASE**

Subtype of: **OBSERVATION**

Supertype of: **OUTBREAK**

Description of: **CASE**

A case is an observation that represents a condition or event that has a specific significance for public health. The case can include a health-related event concerning a single individual or it may refer to multiple health-related events that

are occurrences of the same disease or condition of interest to public health. An outbreak involving multiple individuals is a type of case.

A case definition (a case whose mood code = "definition") includes the description of the clinical, laboratory, and epidemiologic indicators associated with a disease or condition of interest to public health. There are case definitions for conditions that are reportable, as well as for those that are not. There are also case definitions for outbreaks. A case definition is a construct used by public health for the purpose of counting cases, and should not be used as clinical indications for treatment. Examples include AIDS, toxic-shock syndrome, and salmonellosis and their associated indicators that are used to define a case.

Attributes of: **CASE**

Classification Status Code : CV

Code for the classification status of the case. Possible values include confirmed, probable, suspected, not a case, incomplete information. This status code differs from the activity status code inherited from the health-related activity supertype to case. The activity status code captures the lifecycle state of the case (active, inactive, completed).

Confirmation Method Code : CV

Code for the mechanism by which the case was confirmed. This attribute is intended to provide information about how the case classification status was derived. Includes laboratory criteria met, clinical case inclusion criteria (alone) met, epidemiologist- or other public health worker-assigned, epidemiologically linked via investigation, and physician-reported.

Detection Method Code : CV

Code for the method by which the case was identified. Possible values include provider report, patient self-referral, laboratory report, case or outbreak investigation, contact investigation, active surveillance, routine physical, prenatal testing, prenatal testing, prison entry screening, occupational disease surveillance, and medical record review.

Disease Imported Code : CV

Code that indicates whether the disease was likely acquired outside the jurisdiction of observation, and if so, the nature of the interjurisdictional relationship. Possible values include not imported, imported from another country, imported from another state, imported from another jurisdiction, and insufficient information to determine. Note that if the specific jurisdiction is to be captured it is captured as a target participation associated with a jurisdictional party.

Etiologic Status Code : CV

Code for the strength of the causal relationship between the disease-causing agent and the disease. This is particularly relevant for outbreaks where the cause is not yet certain, or emerging/new diseases or conditions where the cause is not clear. For example, in the case of an outbreak of gastroenteritis, blood in the stool may indicate that the agent was most likely a Shiga toxin-producing *E. coli* (strong suspicion), although other infectious or toxic agents may still be included in the differential diagnosis, but to a lesser degree (weak or moderate suspicion). Includes weak suspicion, moderate suspicion, confirmed, and unknown.

Transmission Mode Code : CV

Code for the mechanism by which disease was acquired by the party involved in the case. Includes sexually transmitted, airborne, bloodborne, vectorborne, foodborne, zoonotic, nosocomial, mechanical, dermal, indeterminate.

Class: **HEALTH-RELATED ACTIVITY**

Supertype of: **INTERVENTION**
NOTIFICATION
OBSERVATION
REFERRAL

Associated with: **ACTIVITY RELATIONSHIP**
ACTOR PARTICIPATION
TARGET PARTICIPATION

Description of: **HEALTH-RELATED ACTIVITY**

A health-related activity is an action performed for the purpose of documenting, investigating, or improving the health condition of a party. It may also include documenting the ability to affect the health status of a party. Examples of health-related activities include all of the following:

- interventions such as surgical operations or vaccination;
- administration of a medication;
- referral to another provider;
- diagnostic observations about a patient's condition;
- diagnostic assessment that a condition meets the public health definition of a case;
- a public health notification of a case of a reportable disease or condition;

- public health investigation of all persons exposed to a common source of infection or toxin;
- food or consumer product recalls;
- an intervention targeted at a given population.

An instance of a health-related activity can be captured from several perspectives. Possible perspectives for an instance of a health-related activity are:

- a **fact** about an activity that has occurred, such as the observation of chickenpox in a child;
- a **command**, such as an order to vaccinate a child for chickenpox;
- a **master** table entry of possible activities, such as types of laboratory tests;
- a **definition** algorithmically describing an activity, such as a case definition for chickenpox;
- an **intent** for an outcome of an activity, such as achievement of a 95% immunization rate in children under age 2.

Associations for: **HEALTH-RELATED ACTIVITY**

is_source_for (0,n) :: ACTIVITY RELATIONSHIP :: **relates (1,1)**

is_target_for (0,n) :: ACTIVITY RELATIONSHIP :: **relates (1,1)**

associates (0,n) :: ACTOR PARTICIPATION :: **associates_to (1,1)**

associates (0,n) :: TARGET PARTICIPATION :: **associates_to (1,1)**

Attributes of: **HEALTH-RELATED ACTIVITY**

Activity Critical Date Time : GTS

The "biologically relevant" time for a health-related activity. The concept is best understood with observations, where the time of the observation activity may differ from the time of the observed feature. For instance, in history taking, when the doctor records an episode of Hepatitis A suffered by the patient last year for several weeks. The activity critical date time is the date/time when the patient experienced the episode of Hepatitis A, and not the date and time when the doctor records the history. That is to say, it is the time/dates that the patient actually had hepatitis, and not when the patient tells the doctor, or when the doctor records it. In another example, the provider may order a test, conducted on a blood sample drawn today, for which results will not be available until next week. The activity

critical date time is the date and time of the taking of the specimen, not when the results are available.

Activity Date Time : GTS

The time when the action happened, is ordered or scheduled to happen, or when it can possibly happen. The time specification could be a point in time, a time range during which the activity occurred, or is supposed to occur.

Activity Descriptive Text : FTX

The description of an activity is a piece of free text or multimedia data that describes the activity in all necessary detail. This attribute is a descriptive supplement to an activity type code, not a replacement. There is no restriction on length or content imposed on the description attribute. However, the content of the description is not considered part of the functional information communicated between systems. Descriptions are meant to be shown specifically to interested individuals.

Activity Identifier : SET<II>

This is an instance identifier for a health-related activity. It uniquely identifies a particular instance of a health-related activity class.

Activity Method Code : CV

The activity method code is a parameter of the health-related activity that specifies one of the possible methods used to achieve a given end. The method is specified for a given health-related activity, because there are different methods to achieve results, and knowing the method is important for a more explicit interpretation. For example, when carrying out an assessment of a person's risk-taking behavior, possible methods include: written questionnaire, personal interview, third-party interview (for children), and medical record review. When carrying out interventions for public health education, possible methods include: mass media, billboard, individually targeted automatic messages, and individual counseling.

Activity Mood Code : CV

The activity mood code determines the meaning or context for the activity. The activity (corresponding to a verb in natural language) may be conceived as an event that happened (fact), an ordered service (command), a possible service (master), an algorithm for describing an event (definition), and a goal of health-related activity (intent). Each of these is a different mood.

The activity mood code is critical to the design of this model. Without it, the model described here would be at least three times as big, in order to distinguish between the following:

a) The definition of the health-related activity (e.g., a case or test definition);

b) Health-related activities that are planned;

c) Scheduled health-related activities;

d) Health-related activities that have already occurred or been performed.

Activity Status Code : CV

A code for the state of the action (e.g., intended, ordered, in process, completed). This attribute is not used to describe the classification status of a case; the case classification status code should be used. (See the case attribute: case classification status code.)

Activity Type Code : CV

A code for the kind of activity (e.g., physical examination, person interview, serum potassium, public health notification, product sterilization or pasteurization). The activity type code specifies the service conceptually by using a code from a coding system. The activity type code or "name" is a handle on the concept of the action, not on the individual action instance. Different coding systems cover different kinds of activities, which is why there is not one single coding system to be used for the activity type code.

When observations are recorded for outbreaks, the activity type code captures information to indicate the category of the statistic, e.g., number ill, number exposed, number hospitalized, number treated, number of fatalities, number interviewed, incubation period days/hours, duration of illness (days/hours), number not ill, % female, % male, % less than 18 years of age, ages of affected, and information to indicate the type of statistic, e.g., minimum, maximum, percentage, median, count.

Confidentiality Code : CV

Indicates limitations to disclosure and communication of information about a health-related activity. Includes provider access only, limited to county or state public health department access, disease program access only, or public use/ publicly available.

Interpretation Code : CV

The interpretation code allows for a very rough interpretation of the course or outcome of an activity. These are sometimes called "abnormal flags", however the judgment of normalcy is just one of the common rough interpretations, and is often not relevant. For example, for the observation of a pathologic condition, it doesn't make sense to state the normalcy, since pathologic conditions are not considered "normal." In other words, context is required to make a final

determination, and this code may simply provide a judgment that these data are worth investigating further. For example, this code may be used to indicate that an antibody level is slightly elevated, which may be consistent with disease. However, the interpretation of disease may require additional data, such as a repeated antibody titer, to determine whether the value is rising or falling. This attribute is also used to describe antibiotic susceptibility results as "susceptible", "intermediate", and "resistant".

Maximum Repetition Number : INT

The maximum number of repetitions of a health-related activity. Typical values are 1, some other finite number, and infinity. This is relevant when the health-related activity is a plan or a series of orders.

Priority Code : CV

Code for the priority of the activity. Possible values include routine, emergency, and urgent.

Subject Site Code : CV

Most health care services focus on a particular part of the target on which the health-related activity is performed. Typically, when the target party is a person, this will be a feature related to the anatomic structure of the patient (the "target" of the service). In the case of material entities other categorizations are used. For example, when a sample is ordered from a restaurant to explain a case of food poisoning, sites such as floor, meat grinder, refrigerator, or cutting board could be used.

Class: **INTERVENTION**

Subtype of: **HEALTH-RELATED ACTIVITY**

Description of: **INTERVENTION**

An intervention is the administration of a substance or technique to provide care for or to prevent a condition. This includes vaccinations and preventive therapy as well as medication given directly for therapeutic purposes. An intervention need not be administered solely to individuals, and may include population interventions such as chlorinating or fluoridating the water supply, policies to restrict tobacco sales, pasteurization of milk, and pesticide application in a specific geographic area. Includes therapeutic and preventive treatments, counseling, educational campaigns, needle exchange programs, media campaigns, food recalls.

Attributes of: **INTERVENTION**

Intervention Form Code: CV

The physical form in which the intervention is delivered. For medications, examples include tablet, capsule, suppository, and solution. For environmental interventions, such as chlorination of the water supply, examples might include chlorine in liquid or tablets. For food recalls, examples might include complete meat packages or individual burgers. For media campaigns, examples might include television commercials, radio ads, billboards, or pamphlets.

Intervention Quantity : PQ

The amount of the intervention associated with a single intervention instance. For example, this might refer to the amount of pesticide to be sprayed during a single application or the amount of gas or chemical to be used in a sterilization of a medical device.

In the case of medication, the amount is the dose or amount of the therapeutic or prophylactic agent given at one administration event. This attribute can be used all by itself, or in combination with a strength.

Intervention Reason Code : CV

Code which describes the basis for the intervention. Includes treatment, prophylaxis, post-exposure prophylaxis, high-risk individual or population.

Intervention Route Code : CV

The route by which the intervention is administered to the object of the intervention. For medications, includes oral, intravenous, subcutaneous, subdermal, and intramuscular. Medication route is similar to an anatomic body site through which the therapeutic or prophylactic agent is incorporated or otherwise applied to the body. Other kinds of intervention routes might include: via public health nurse counseling, billboard campaign, newspaper advertisement, helicopter spray (for pesticide treatment), injection of water supply (for fluoridation).

Rate Quantity : PQ

The period of time over which a specified dose is delivered. This attribute only applies to continuously divisible intervention forms such as fluids and gases. In this case, the intervention rate indicates the amount of intervention within a specified period of time. The rate quantity is a duration (physical quantity in time), and it is the denominator of the intervention rate, while intervention quantity is the numerator. For example, pesticide to be used for mosquito abatement may be delivered at a rate of 20 liters per minute from a spray applicator.

Strength Quantity : PQ

The strength of an intervention is the amount of the agent per each unit of administration. This applies to pesticides, chlorination as well as medication. If the intervention form is continuously divisible (e.g., fluid, gas), the strength is a concentration.

When the strength attribute is used, the actual administered amount is the product of intervention quantity and strength quantity.

Class: **NOTIFICATION**

Subtype of: **HEALTH-RELATED ACTIVITY**

Description of: **NOTIFICATION**

A notification is an interaction with a caseworker, person or party to report or document a condition or health-related activity of importance to the health of the public. Includes notification by a provider to a patient that they have a disease, report by a provider or laboratory to public health of a case or positive isolate, report of a gunshot wound to police, reminder of the need for immunization against disease, notification of a possible adverse reaction to a drug.

Attributes of: **NOTIFICATION**

Notification Reason Code: CV

Code for the reason for the notification. Examples might include reportable condition, positive laboratory test, positive screening results, self-motivated, interview, referral, and positive gonorrhea test.

Class: **OBSERVATION**

Subtype of: **HEALTH-RELATED ACTIVITY**

Supertype of: **CASE**

Description of: **OBSERVATION**

Observations are actions performed in order to determine an answer or result value. Observation result values are specific information about the observed object. The type and constraints of result values depend on the kind of action performed.

An observation, according to Webster's, is an "act of recognizing and noting a fact [...] often involving measurement with instruments" and at the same time an observation is also "a record or description so obtained" [i.e., obtained through recognizing and noting]. Thus an observation is both the action and measurement

"procedure" and the resulting information that was obtained. The model understands the result to be entirely dependent on the observation action, and thus models the result as a component (attribute) of the Observation action rather than as an independent entity.

The following concepts are included as observations:

- A **test** is a procedure followed to objectively measure or evaluate the presence or status of a condition. It includes vital signs, physical exams, food tests, animal tests, height, and weight;

- An **assessment of causality** is the relationship between a patient condition and a source that may be causally related to that condition;

- A **vehicle condition** is the circumstances under which the vehicle became a carrier for a disease-causing agent. An example of a vehicle condition includes temperature abuse in storing or preparing food;

- A **diagnosis** is the conclusion drawn from analysis of the signs and symptoms exhibited or described by an individual;

- A **party condition** is the state of health, contamination, or infection of a party;

- A **health status inquiry** is the account of a party's health-related background. This could include an interview conducted anonymously as part of a risk factor survey. It includes description of current symptoms; risk behaviors such as alcohol, tobacco, or other drug use; exposures past and present; medical or surgical history; current or previous medications, vaccinations, or interventions (treatment or prophylactic); reproductive history; occupational history or exposures; sexual habits; eating habits; travel history; educational background; marital status; family history. For example, the patient's, parent's, or guardian's report of drug use, life style, previous medical conditions, and treatments.

In the public health context, case and outbreak information are captured as observations. This includes information such as a count or percentage of cases tracked for public health reporting. It also includes number ill, number exposed, number hospitalized, number treated, number of fatalities, number interviewed, incubation period, duration of illness, number not ill, % female, % male, and % less than 18 years of age.

Attributes of: **OBSERVATION**

Derivation Expression Text : ST

The derivation expression text shows how an observation can be derived from other observations. In this case, the activity relationship links the observations

through the value of the relationship code (activity relationship type code = "derivation").

For example, to define a derived observation for a change in antibody titer, one will associate the change in titer observation with the acute titer observation and the convalescent titer observation. The derivation expression text would then be "Change in Titer = Convalescent Titer / Acute Titer". If this observation value is abnormal, for example greater than 4, this would be indicated in the Interpretation Code for the Change in Titer observation.

Observation Value : ANY

The result value of an observation activity. This value can be of any datatype. This fact reflects the many different ways in which the value of an observation can be captured. For outbreaks or reporting of aggregate numbers of cases, the number of persons affected would be included as a value here.

It is worth noting that, as a result of the functionality introduced with the activity mood code, reference values or ranges are captured as observation values. The fact that an observation carries a reference value is indicated by the value of the mood code.

Class: **OUTBREAK**

Subtype of: **CASE**

Description of: **OUTBREAK**

An outbreak or cluster is the occurrence in a community or region of cases of a condition of public health importance in excess of those normally expected. The designation of an outbreak implies that a public health assessment of causality or at least of relatedness among cases has taken place. An outbreak is considered to be a special type of case (where a case, in this instance, may include many affected individuals), and may not simply be an aggregate of multiple cases although an outbreak may also be designated as an aggregate of multiple individual cases.

Given that an outbreak is a subtype of observation, the number of parties (which will generally equate to the number of cases) affected by the outbreak is captured as the observation value.

Attributes of: **OUTBREAK**

Outbreak Jurisdictional Extent Code : CV

Code for the qualitative measure of the number of jurisdictions involved. Possible values include single jurisdiction, multi-county, multi-state, and multi-national.

Note that if the specific jurisdictions are to be captured they are captured as target participations associated with a jurisdictional party.

Outbreak Peak Date : TS
Date of onset for the highest number of cases (mode) associated with the outbreak.

Outbreak Time Range : IVL<TS>
The period of time during which the outbreak takes place. The date on which an outbreak starts is the earliest date of onset among the cases assigned to the outbreak, and its ending date is the last date of onset among the cases assigned to the outbreak.

Class: **REFERRAL**

Subtype of: **HEALTH-RELATED ACTIVITY**

Description of: **REFERRAL**
A referral is an introduction of an individual or individuals from one health care organization to another, or from one part of an organization to another for the purpose of diagnosis or treatment. It includes the referral of a case or the referral of multiple exposed persons (or cases) by one State Health Department to another.

Attributes of: **REFERRAL**

Referral Description: FTX
Free form text describing the referral.

Referral Reason Text: FTX
Free form text providing the reason for the referral as well as the action that is expected or requested upon receipt of the referral. Examples might include partner, positive lab test, outside of referring jurisdiction and needs follow-up, possible cancerous lesion for biopsy, and requires surgical intervention.

Class: **TARGET PARTICIPATION**

Associated with: **HEALTH-RELATED ACTIVITY**
LOCATION
MATERIAL
PARTY

Description of: **TARGET PARTICIPATION**

Target participations include the passive parts played by a party in the health-related activity. The target of a health-related activity can be any party or material, including humans, other non-person living organisms, and inanimate material.

For example, within a disease investigation, the person identified as an actual or potential carrier is a target of the activity. If the "patient" is a child, and another person, such as a parent, speaks for them (e.g., answering a questionnaire) that representative is also an activity target.

Associations for: **TARGET PARTICIPATION**

associates_to (1,1) :: HEALTH-RELATED ACTIVITY :: **associates (0,n)**

associates_to (0,1) :: LOCATION :: **associates (0,n)**

associates_to (0,1) :: MATERIAL :: **associates (0,n)**

associates to (0,1) :: PARTY :: **associates (0,n)**

Attributes of: **TARGET PARTICIPATION**

Target Awareness Code : CV

Indicates whether the associated patient or family member is aware of the health-related activity, and especially of the observation made. This is only relevant for persons who are targets of a health-related activity. For example, a patient (or his family members) may not be aware of a malignancy diagnosis, the patient and family may be aware at different times, and some patients may go through a phase of denial.

Target Time Range : IVL<TS>

The time range in which the associated party or material was a target of the specified target type code in the associated activity.

Target Type Code : CV

Identifies the particular role in which the party appears as the target of the health-related activity.

Examples of target type codes include: "State reporting case", "target of case", "location imported from".

Locations Subject Area

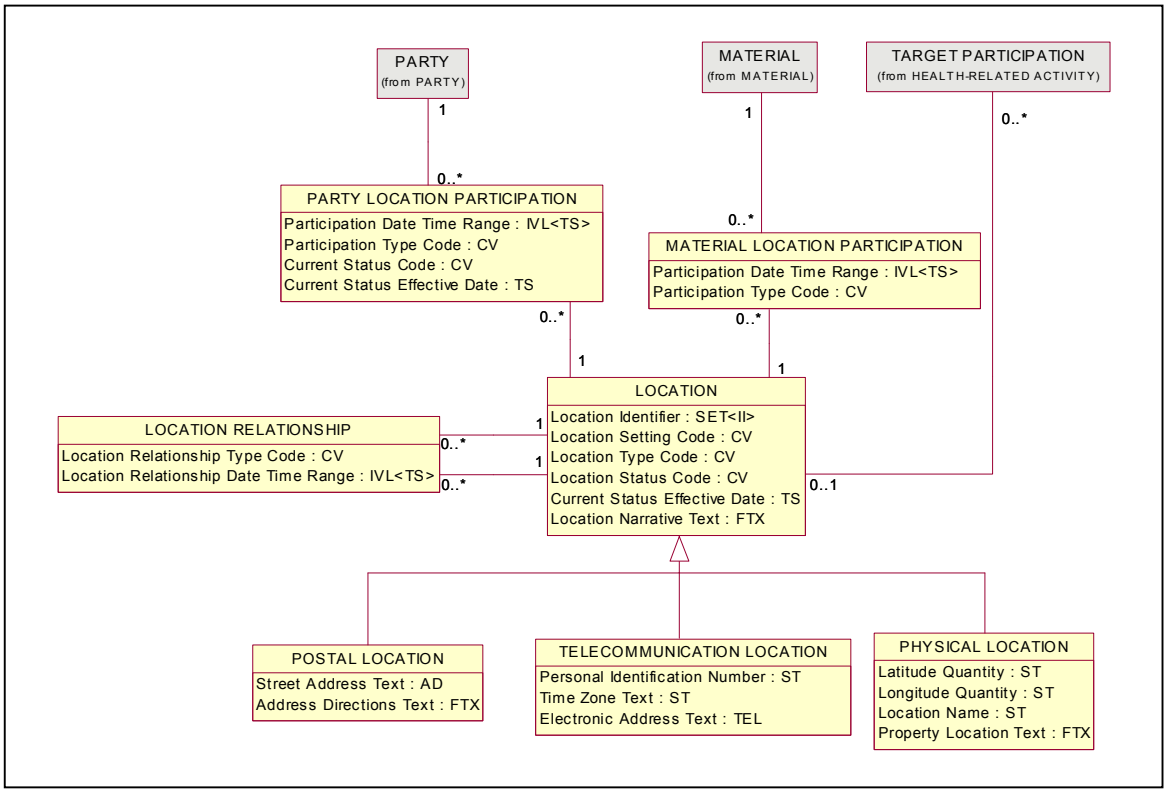


Figure 3. Locations Subject Area Diagram

The classes and attributes for the Locations subject area are described below.

Class: **LOCATION**

Supertype of: **PHYSICAL LOCATION**
POSTAL LOCATION
TELECOMMUNICATION LOCATION

Associated with: **LOCATION RELATIONSHIP**
LOCATION RELATIONSHIP
MATERIAL LOCATION PARTICIPATION
PARTY LOCATION PARTICIPATION
TARGET PARTICIPATION

Description of: **LOCATION**

A location is a site of interest to public health. Examples of locations include buildings, picnic grounds, regional areas, homes, test locations, specimen locations, hospitals, day care centers, prisons, and other potential transmission

locations. It also includes districts - that is to say one location may contain another. The information for a location includes information such as an address that makes it possible to find or to send messages to the location.

Associations for: **LOCATION**

is_source_for (0,n) :: LOCATION RELATIONSHIP :: **relates (1,1)**

is_target_for (0,n) :: LOCATION RELATIONSHIP :: **relates (1,1)**

associates (0,n) :: MATERIAL LOCATION PARTICIPATION :: **associates_to (1,1)**

associates_to (0,n) :: PARTY LOCATION PARTICIPATION :: **associates (1,1)**

associates (0,n) :: TARGET PARTICIPATION :: **associates_to (0,1)**

Attributes of: **LOCATION**

Current Status Date Time Range : IVL<TS>
The time range during which the current location status is or was active.

Location Identifier : SET<II>
An instance identifier that identifies the location. This could include, among other things, identifiers assigned to a property within a registry office or other organization tracking plots of land.

Location Narrative Text : FTX
A free text note that carries additional information related to the location. This could include instructions for finding the location when postal information is inadequate. It could also include information useful to people visiting the location (e.g., "Beware of dog").

Location Setting Code : CV
Code for the location environment. Examples might include public, private, federal, and unknown.

Location Status Code : CV
An indication of the validity of the location. Examples might include verified, unverified, and unable to verify.

Location Type Code : CV

Code that indicates the type of location. Includes residence, office, restaurant, hospital, daycare center, ship, prison, nursing home, or district such as census tract or congressional district.

Class: **LOCATION RELATIONSHIP**

Associated with: **LOCATION**

Description of: **LOCATION RELATIONSHIP**

An association between two locations. This relationship is important in public health reporting and investigations to describe how sites of public health importance are associated, for instance: fourth floor of hospital "has as part" the neonatal ICU. Here, the location relationship, "has as part", describes the association between two locations, a particular ICU and the hospital floor. Another example might be juice maker's apple orchard "is next to" farmer's cow pasture. One can also link telecommunication locations or postal locations to physical locations, for instance, 123 Main Street, Doraville, GA 30256 "is geolocated by" +33 47.966, -84 19.508.

This structure is not needed to link multiple locations, e.g., home address, email address, business address, to a single party. That requirement is supported through linking location information to party with party location participation.

Associations for: **LOCATION RELATIONSHIP**

relates (1,1) :: LOCATION :: **is_source_for (0,n)**

Attributes of: **LOCATION RELATIONSHIP**

Location Relationship Date Time Range : IVL<TS>

The period in time during which the relationship between the two location instances is effective. The time interval can be open at either end. That is, both the start and stop dates for the participation could be indicated, or either start or stop by themselves.

Location Relationship Type Code : CV

Indicates the type of relationship between the two locations. For example, "same as", "adjacent to".

Class: **MATERIAL LOCATION PARTICIPATION**

Associated with: **LOCATION**

MATERIAL

Description of: **MATERIAL LOCATION PARTICIPATION**

Material location participation indicates the location where an item of material is or was to be found.

Associations for: **MATERIAL LOCATION PARTICIPATION**

associates_to (1,1) :: LOCATION :: **associates (0,n)**

associates_to (1,1) :: MATERIAL :: **associates (0,n)**

Attributes of: **MATERIAL LOCATION PARTICIPATION**

Participation Date Time Range : IVL<TS>

Indicates the period in time during which the material item is or was to be found at the location. For example, the date a specimen arrived at the location. The time interval can be open at either end. That is, both the start and stop dates for the participation could be indicated, or either start or stop dates by themselves.

Participation Type Code : CV

Code for the participation role of the material at the location. Examples might include “resides at”, “originated at”, and “destined for”.

Class: **PARTY LOCATION PARTICIPATION**

Associated with: **LOCATION**
PARTY

Description of: **PARTY LOCATION PARTICIPATION**

Party location participation indicates the relationship between a party and a location. The party may be an organization that owns several facilities or locations. The participation role would be that of owner of the facility at this location. Another role for a party would be a person who "works at" a location.

Associations for: **PARTY LOCATION PARTICIPATION**

associates (1,1) :: LOCATION :: **associates_to (0,n)**

associates_to (1,1) :: PARTY :: **associates (0,n)**

Attributes of: **PARTY LOCATION PARTICIPATION**

Current Status Code : CV
Code for the status of the participation between the party and the location.

Current Status Effective Date : TS
The effective date for the current party location role status.

Participation Date Time Range : IVL<TS>
Indicates the period in time during which the party is related to the location. The time interval can be open at either end. That is, both the start and stop dates for the participation could be indicated, or either start or stop by themselves.

Participation Type Code : CV
Code for the participation role of the party at the location. Examples might include owner, occupant, visitor, worker, and client.

Class: **PHYSICAL LOCATION**

Subtype of: **LOCATION**

Description of: **PHYSICAL LOCATION**
Physical location information makes it possible to find the location on a map or by examination of surveyor's documentation or by reference to a land or property registry.

Attributes of: **PHYSICAL LOCATION**

Latitude Quantity : ST
Indicates the latitude of the location as measured in degrees north or south of the equator.

Location Name : ST
The name of the location as it might be referred to on a map or in a registry.

Longitude Quantity : ST
Indicates the longitude of the location as measured in degrees west or east of the prime meridian at Greenwich, England.

Property Location Text : FTX
A description of the property that is sufficiently precise to enable someone to locate the property and to recognize its boundaries. The description can be

formulated as in terms of the property boundaries, or in terms of specific lots or parcels that are located within a legal entity such as a township, county, or other legally defined territorial entity. In some cases the description will be drawn from the legal description of a property as recorded on a deed or other legal paper.

Class: **POSTAL LOCATION**

Subtype of: **LOCATION**

Description of: **POSTAL LOCATION**

Information used to direct mail to a particular location, or to find the location using information to be found on a street map.

Attributes of: **POSTAL LOCATION**

Address Directions Text : FTX

Descriptive information to assist a party in finding a particular location. This information is intended to supplement or replace street address information.

Street Address Text : AD

Text used for an address label. This could include street address information, or postal directions using a box number to send mail to a post office box, a rural free delivery box, or a military post office. It also includes lot or address number when the address refers to an apartment building or housing complex.

Class: **TELECOMMUNICATION LOCATION**

Subtype of: **LOCATION**

Description of: **TELECOMMUNICATION LOCATION**

An electronic address for a party that provides the mechanism to contact the party, to send messages, or to access information relevant to the party. Examples include a telephone number, an email address, a World Wide Web URL. This is distinguished from a postal address.

Attributes of: **TELECOMMUNICATION LOCATION**

Electronic Address Text : TEL

The number or other string that is entered to contact a particular telephone or other electronic location.

Personal Identification Number : ST

An identification number assigned to a person, and used to access a communication device such as a beeper. Often referred to as a PIN.

Time Zone Text : ST

Text indicating the time zone in which a telephone is located.

Materials Subject Area

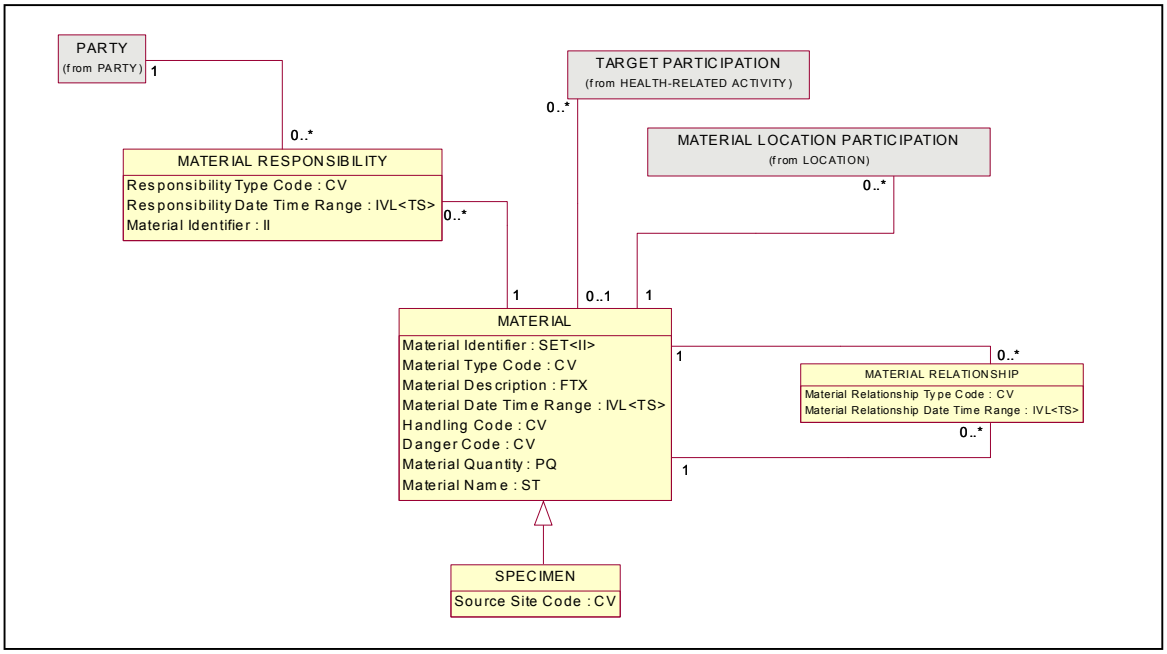


Figure 4. Materials Subject Area Diagram

The classes and attributes for the Materials subject area are described below.

Class: **MATERIAL**

Supertype of: **SPECIMEN**

Associated with: **MATERIAL LOCATION PARTICIPATION**
MATERIAL RELATIONSHIP
MATERIAL RESPONSIBILITY
TARGET PARTICIPATION

Description of: **MATERIAL**

Material is defined according to Webster's: 1) the elements, constituents, or substances of which something is composed or can be made; 2) matter that has qualities which give it individuality and by which it may be categorized.

In public health, interest in materials commonly arises when a material is a vehicle for a disease agent, or is suspected of being such a vehicle. For example, when a case investigation considers the question of whether a bowl of potato salad is contaminated with *Salmonella* organisms, the potato salad might be recorded as an item of material. Note that this assumes that the identity of the potato salad needs to be captured. In some cases it would be sufficient to record an observation that

the contaminated food was potato salad. It is also possible, when collecting information about the bacteria, to capture it as an item of material.

Other materials or entities of interest to public health may include an independent, separate, or self-contained substance or object, such as a lake, a pool, a waterpark, resort, campsite, ship, airplane, or train that might serve as a source or vehicle of exposure to a health hazard. For example, a public health investigation can center around the question of bacterial or other contamination of a site such as a ship or swimming pool. Specimens can be taken from such materials just as specimens can be taken from parties, whether human or otherwise.

Associations for: **MATERIAL**

associates (0,n) :: MATERIAL LOCATION PARTICIPATION :: **associates_to (1,1)**

relates (0,n) :: MATERIAL RELATIONSHIP :: **is_source_for (1,1)**

associates (0,n) :: MATERIAL RESPONSIBILITY :: **associates_to (1,1)**

associates (0,n) :: TARGET PARTICIPATION :: **associates_to (0,1)**

Attributes of: **MATERIAL**

Danger Code : CV
A code signaling whether there are certain dangers or hazards associated with this material. For example, "Examine under hood", "Wear gloves".

Handling Code : CV
A code to describe how the material needs to be handled to avoid damage. For example: "Do not expose to light", "Keep at certain temperature".

Material Date Time Range : IVL<TS>
An indication of the time interval during which the material is in existence.

Material Description : FTX
A free text description of the material. May contain multimedia, such as a drawing or image depicting the material.

Material Identifier : SET<II>
The identifier assigned to an individual material item.

Ideally each entity will have only one identifier assigned to it. However, since different systems will maintain different material databases, there may be different

instance identifiers assigned by different systems. Note that for serial numbers assigned by specific manufacturers, catalog numbers of specific distributors, or for inventory numbers issued by owners, the attribute Material Identifier in the Material Responsibility class can also be used. This allows clearer expression of the fact that a specific party associated with that material assigns such a code.

Material Name : ST

Name of the material. This is important in special cases such as the name of a lake, an amusement park, or a cruise ship

Material Quantity : PQ

An indication of the amount of material. This could be a count or a quantity. For example, 2 liters of water, 25 vials of blood.

Material Type Code : CV

This code describes the kind of material. No single terminology is expected to provide all concepts that are types of material, since it is simply too broad a domain. Instead of limiting the Material Type Code to a single domain, various coding systems may be used.

For example, specimen types (e.g., whole blood, serum, and urine) can be used in this attribute. For pharmacological substances the U.S. National Drug Code (NDC) may be applicable. For other types of materials of interest to public health, such as lakes, rivers, national parks, trains, planes, or ships, other coding systems will be applicable.

Class: **MATERIAL RELATIONSHIP**

Associated with: **MATERIAL**

Description of: **MATERIAL RELATIONSHIP**

Material relationship captures the relationship between two items of material. Material relates to other material largely in some kind of whole-part or containment relationship. The special functioning of the material relationship depends on the role of material, i.e., whether the material is a discrete thing, a homogenous substance, or a container. Material can be all of those forms.

Associations for: **MATERIAL RELATIONSHIP**

is_source_for (1,1) :: MATERIAL :: **relates (0,n)**

is_target_for (1,1) :: MATERIAL :: **relates (0,n)**

Attributes of: **MATERIAL RELATIONSHIP**

Material Relationship Type Code : CV

Code for the type of material association. Every relationship type implies certain roles for the material on either side of the relationship. For example, there is a relationship between a blood specimen and the species of bacteria cultured from it, between a dish of food and the ingredients used to make it, and a lake and the sample collected from it. Thus examples of material relationship codes include: is cultured from, is an ingredient of, and is a sample from.

Material Relationship Date Time Range : IVL<TS>

The period of time during which the relationship between the two materials is valid.

Class: **MATERIAL RESPONSIBILITY**

Associated with: **MATERIAL**
PARTY

Description of: **MATERIAL RESPONSIBILITY**

Description of the type of relationship between a party and an item of material. Material can have many kinds of relationships with parties. Relationships between material and parties are included here since there are generally one or more parties responsible for managing an item, or for performing particular functions with it.

For example, manufacturing is an activity in which a party or parties acts on material. In some instances we may simply be interested in who made the material. We may also be concerned with how the material item has been processed or treated. For example, if the manufacturing of the material resulted in contaminated or doctored medications, or food was not held at or cooked to proper temperatures, there are significant implications for public health.

An important example of material responsibility is the role of a party as the provider or receiver of a specimen. For example, a lake or a food item or a person may be the source of a specimen, and a public health official may be the person who obtains the specimen (the specimen may be a lake or food sample, a body part, blood sample, sputum, or feces). Owner, distributor, and custodian/holder are additional examples of relationship types between material and party.

Similarly, when a material item is implicated as a vehicle for a disease condition, such as a food item that is contaminated with *Salmonella* organisms, the material responsibility class provides a way to record party responsibility for the food item. This could include recording the party who was responsible for its pasteurization, the party who prepared the food, or the party responsible for storing it. For

medications or intravenous solutions, this might be the party responsible for sterilization or for mixing the solution.

Associations for: **MATERIAL RESPONSIBILITY**

associates_to (1,1) :: MATERIAL :: **associates (0,n)**

associates (1,1) :: PARTY :: **associated to (0,n)**

Attributes of: **MATERIAL RESPONSIBILITY**

Material Identifier : II

An identifier assigned to a material item in the context of its relationship with a responsible party. Different responsible parties may give the same piece of material different identifiers. For example, a manufacturer may assign a manufacturer ID and a distributor may assign a catalog number. All those identifiers can in principle occur under the Material ID attribute, i.e., as a property of the material itself. However, this attribute allows one to make the scope of the ID more clear, i.e., it helps to easily distinguish a specific manufacturer's ID from a distributor's ID much more clearly than can be done using the assigning authority component of the instance identifier datatype.

Responsibility Date Time Range : IVL<TS>

Indicates the period of time during which the responsibility holds.

Responsibility Type Code : CV

Specification of the kind of responsibility that the party takes on with respect to the material. Examples might include owner, responsible for preparation, custodian.

Class: **SPECIMEN**

Subtype of: **MATERIAL**

Description of: **SPECIMEN**

A specimen is a part, fraction, aliquot, component, tissue sample, body fluid, food, or other substance that is collected in a health-related activity to support the assessment, diagnosis, or treatment of a party.

Attributes of: **SPECIMEN**

Source Site Code : CV

The source site code indicates from where, in relationship to the specimen source, the specimen is taken. For persons and non-person living organisms, the valid domain is a list of body sites. This is an attribute of the specimen, since it may be relevant in some cases, e.g., if multiple liver needle biopsies are taken from different lobes and locations of the liver. In the case of material items such as restaurants or lakes, the site code indicates from where the specimen was taken. In the case of a lake, this could be, "near intake", or "at swimming site". In the case of a restaurant, this could indicate a typical site in the restaurant such as within the meat grinder.

Parties Subject Area

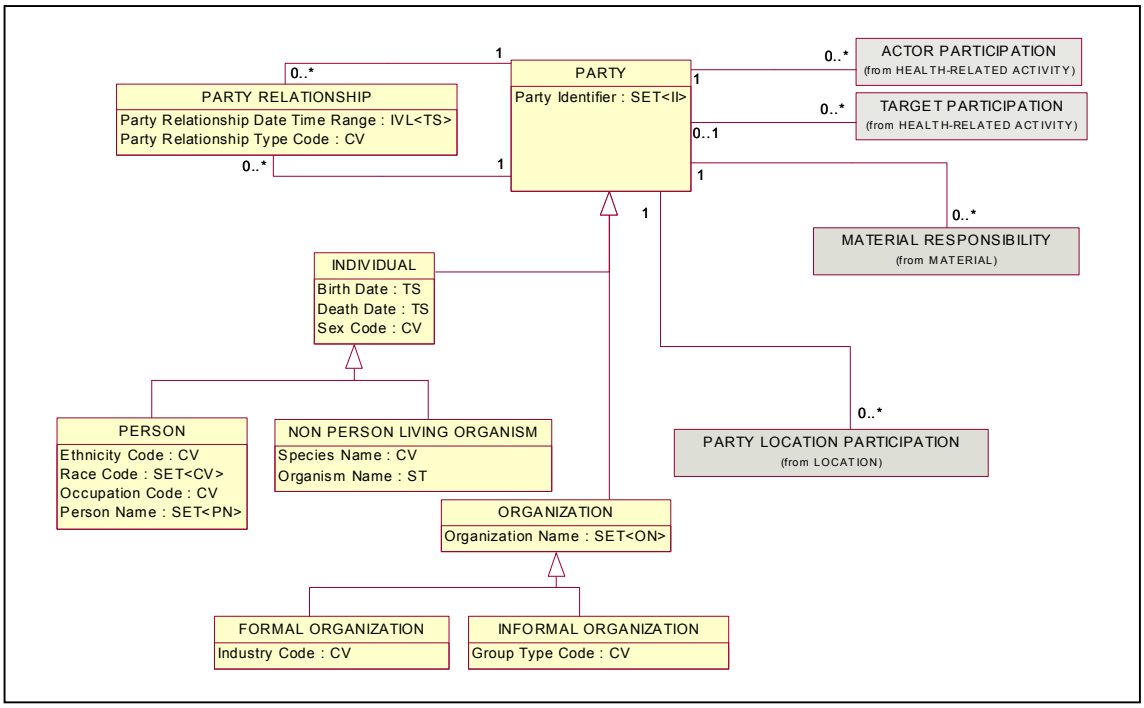


Figure 5. Parties Subject Area Diagram

The classes and attributes for the Parties subject area are described below.

Class: **FORMAL ORGANIZATION**

Subtype of: **ORGANIZATION**

Description of: **FORMAL ORGANIZATION**

A formal organization is an administrative and functional structure with common objectives. Examples in public health might include state-based public health membership organizations such as the Association of Public Health Laboratories (APHL), Association of State and Territorial Health Officials (ASTHO), the Council of State and Territorial Epidemiologists (CSTE), National Association of County and City Health Officials (NACCHO), National Association for Public Health Statistics and Information Systems (NAPHSIS), as well as individual organizations such as California Department of Health Services, Dekalb County Health Department, Blue Cross/Blue Shield Health Plans, Kaiser Permanente Health Maintenance Organization, Quest Diagnostics, Environmental Protection Agency.

Attributes of: **FORMAL ORGANIZATION**

Industry Code : CV
Code for the type of activity or industry in which the organization is engaged.

Class: **INDIVIDUAL**

Subtype of: **PARTY**

Supertype of: **NON-PERSON LIVING ORGANISM**
PERSON

Description of: **INDIVIDUAL**
An individual is a human person or other single organism.

When non-person living organisms are under consideration, their identity should only be recorded as a party when it is reasonable to do so, such as when they need to be recorded in reference to a individual or series of health-related activities. Note that parties can be identified in order to record an association to a material item or to a location. This is not likely to occur for a non-human living organism except in the non-trivial case of specimens. As a general rule, such non-individually identified organisms as microorganisms and viruses will not be recorded as parties. Information about them will be captured as observations.

Attributes of: **INDIVIDUAL**

Birth Date : TS
Date on which the individual was born.

Death Date : TS
Date on which the individual died.

Sex Code : CV
Code for the individual's sex at birth. Includes Male and Female.

Class: **INFORMAL ORGANIZATION**

Subtype of: **ORGANIZATION**

Description of: **INFORMAL ORGANIZATION**
An informal organization is a casual grouping or cluster of individuals with common interests, characteristics or exposures, or relationships. An informal organization can include individuals who do not recognize their relationship to the rest of the group, and in fact, this class is particularly intended to represent populations or groups of interest to public health, e.g., persons who are smokers,

persons of a certain age or race, persons exposed to the same chemical or agent, and persons who are HIV-positive. The concept of informal organizations also includes such clusters as families, neighborhoods, support groups, and groups of migrant workers. The informal organization can group non-human parties as well. Therefore it includes herds of cattle, canine litters, and prides of lions.

Attributes of: **INFORMAL ORGANIZATION**

Group Type Code: CV

Code for the type of informal organization. Examples include groups such as families, Rotary Club members, girl scouts, retired persons, persons with heart disease, alcoholics, persons vaccinated against measles, persons who are chronic typhoid carriers, or patients on a given floor or ward of a hospital.

Class: **NON-PERSON LIVING ORGANISM**

Subtype of: **INDIVIDUAL**

Description of: **NON-PERSON LIVING ORGANISM**

A non-person living organism is an individual living thing other than a human being that is sufficiently important in its own right to model as a party. For example, this includes pets and working or farm animals whose condition is under investigation.

Normally, other living clusters such as bacteria, parasites, viruses, prions, and insects, are modeled as specimens. Information about them is captured as an observation or observations. Such living clusters should only be recorded as parties when it is necessary to capture multiple references to the same individual in the course of a health-related activity.

Attributes of: **NON-PERSON LIVING ORGANISM**

Organism Name : ST

The name assigned to an animal or other organism. For example, the name assigned to a pet or to a working animal such as a racehorse.

Species Name : CV

The name of the species, including both the genus and the species. This value is drawn from a coded domain that contains the names of the known species.

Class: **ORGANIZATION**

Subtype of: **PARTY**

Supertype of: **FORMAL ORGANIZATION**
INFORMAL ORGANIZATION

Description of: **ORGANIZATION**

Organizations provide a way to recognize the grouping and/or collective action of individuals. An organization may be a group of functions operating as a unit. Examples are managed care organizations, hospital systems, State Health Departments, and regulatory agencies. Such an organization is modeled as a formal organization. An organization may also be simply a group of interest that has been assembled or defined in some informal manner. This type of organization is modeled in the PHCDM as an informal organization. Examples of such are social groups or units such as families, boy scouts, day care attendees, and college students.

Attributes of: **ORGANIZATION**

Organization Name : SET<ON>
Name of the organization.

Class: **PARTY**

Supertype of: **INDIVIDUAL**
ORGANIZATION

Associated with: **ACTOR PARTICIPATION**
MATERIAL RESPONSIBILITY
PARTY LOCATION PARTICIPATION
PARTY RELATIONSHIP
TARGET PARTICIPATION

Description of: **PARTY**

A party is an individual or organization that is specifically of interest to public health. This model includes the concept of "party", in order to clearly represent the similar ways that the different kinds of party are related to health-related activities, materials, and locations. These similarities are particularly relevant in the public health context due to the broad range of concerns that come up.

Something is captured as a party when there is a specific interest in its associations with health-related activities. That is to say, information is captured about a particular individual or organization that makes it desirable to record its individual existence. Usually this implies there will be a series of associations with that

individual that need to be linked. This distinction is important because information can also be captured as an observation (i.e., a health-related activity). For example, we expect that pets and specific farm animals such as horses and cows will be captured as parties (non-person living organisms).

Concepts that are not considered parties include purely material entities, such as lakes or parks. These are considered to be a type of material. Bacteria discovered within a specimen will be captured as observations made on that specimen.

The best way to illustrate this point is through the use of examples. Public health interventions are sometimes applied to specific persons. This includes the delivery of treatment to prevent the development of tuberculosis, a vaccination given to a patient exposed to rabies. It also includes the delivery of information, as when a sexual partner of a patient with a sexually transmitted disease is provided with counseling and clinical information about the disease (along with therapy to prevent disease).

Public health interventions are sometimes applied to organizations. Note that this model treats groups of people as informal organizations. Examples include providing vaccinations and information to the members of a boarding school where a case of meningitis was diagnosed, and the delivery of health warnings to the general public when *Shigella* organisms are detected in a commercial food product. Education campaigns related to such topics as AIDS prevention, the dangers of tobacco use, and the importance of calcium in diets are regarded as public health interventions and may be delivered to such "organizations" that include the population of a city, state, or region, or to specific age cohorts or otherwise identifiable groups.

Public health interventions are sometimes applied to non-person living organisms. For example, dogs living as pets within a neighborhood might receive additional rabies inoculations when several dead and infected raccoons were found in the vicinity. Members of a herd of cattle might be treated when disease was encountered in one of them. Note that within this model an informal organization includes relevant groupings of individuals. These individuals could be persons or non-person living organisms. Therefore, a herd of cattle is an informal organization.

Associations for: **PARTY**

associates (0,n) :: ACTOR PARTICIPATION :: **associates to (1,1)**

associated to (0,n) :: MATERIAL RESPONSIBILITY :: **associates (1,1)**

associates (0,n) :: PARTY LOCATION PARTICIPATION :: **associates_to (1,1)**

relates to (0,n) :: PARTY RELATIONSHIP :: **is source for (1,1)**

associates (0,n) :: TARGET PARTICIPATION :: **associates to (0,1)**

Attributes of: **PARTY**

Party Identifier : SET<II>
A party identifier is a value that identifies a party.

Class: **PARTY RELATIONSHIP**

Associated with: **PARTY**
PARTY

Description of: **PARTY RELATIONSHIP**
A party relationship captures the relationship between two parties. Examples of party relationships might include sexual partners, marital relationship, primary caretaker and subject, and employment between parties. Further examples include parent to child, health care provider to patient, health coverage organization to patient. The relationship between a person and their foster parent, adoptive parent, relative, emergency contact, or spouse is captured by this class. This association generally refers to a relationship that exists outside of the particular event of current interest, such as a specific health-related activity.

Associations for: **PARTY RELATIONSHIP**

is source for (1,1) :: PARTY :: **relates to (0,n)**

is target for (1,1) :: PARTY :: **relates to (0,n)**

Attributes of: **PARTY RELATIONSHIP**

Relationship Date Time Range : IVL<TS>
The period of time during which the relationship between the two parties is valid.

Relationship Type Code : CV
Code for the type of party relationship. Examples might include is an employee of, is the sexual partner of, and is the parent/child of.

Class: **PERSON**

Subtype of: **INDIVIDUAL**

Description of: **PERSON**
A person is a human individual.

Attributes of: **PERSON**

Ethnicity Code : CV
Code for the person's ethnic background (e.g., Hispanic, non-Hispanic).

Occupation Code : CV
Code for the occupation in which the person is employed.

Person Name : SET<PN>
A Person Name is a name assigned to a person.

Race Code : SET<CV>
Code for the person's race (e.g., American Indian/Alaskan Native, White, African American, Asian, Hawaiian/Pacific Islander). The attribute repeats in order to record the multiple racial categories to which a person can belong.

APPENDICES

Appendices

Datatypes

This section contains the datatype definitions used within the Public Health Conceptual Data Model. The datatypes are drawn from the HL7 Reference Information Model, and represent a subset of the datatypes defined therein.

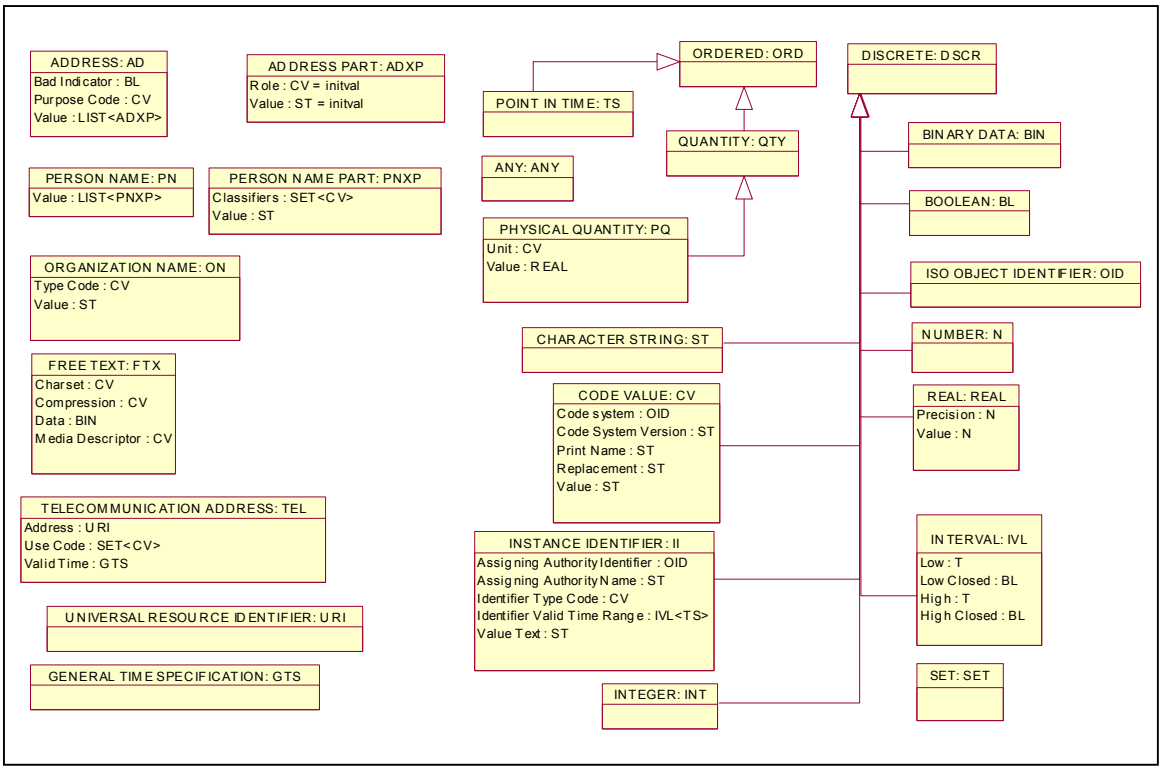


Figure 6. Datatypes Diagram

The classes and attributes that comprise the datatypes are described below.

Datatype: **ADDRESS : AD**

Is a Composite Datatype

Description of: **AD**

This Address datatype is used to communicate postal addresses and residential addresses. The main use of such data is to allow printing mail labels (postal address), or to allow a person to physically visit that address (residential address). The difference between postal and residential address is whether or not there is just a post office box. The residential address is not supposed to contain other

information that might be useful for finding geographic locations or doing epidemiological studies. These addresses are thus not very well suited for describing the locations of mobile visits or the "residency" of homeless people.

Components of: **AD**

Bad Indicator : BL

Indicates that this address is not working

Purpose Code : CV

A purpose code indicates the use for a given address. Examples might include preferred residency (used primarily for visiting), temporary (visit or mailing, but see History), preferred mailing address (used specifically for mailing), and some more specific ones, such as "birth address" (to track addresses of small children). An address without specific purpose code might be a default address useful for any purpose, but an address with a specific purpose code would be preferred for that respective purpose.

Value : LIST<ADXP>

This contains the actual address data as a list of address parts that may or may not have semantic tags.

Datatype: **ADDRESS PART : ADXP**

Is a Composite Datatype

Description of: **ADXP**

This type is not used outside of the Address datatype. Addresses are regarded as a token list. Tokens usually are character strings but may have a tag that signifies the role of the token. Typical parts that exist in about every address are ZIP code, city, country but other roles may be defined regionally, nationally, or on an enterprise level (e.g., in military addresses). Addresses are usually broken up into lines that are indicated by special line break tokens.

Components of: **ADXP**

Role : CV

The role of an address part (if any) indicate whether an address part is the ZIP code, city, country, or post office box.

Value : ST

The value of an address part includes the text for the specific component of the address. It is what is printed on an address label.

Datatype: **ANY : ANY**

Is a Primitive Datatype

Description of: **ANY**

This is a generalized datatype that represents any other datatype within the model. This concept is needed to support observation values, and to let those values take on any datatype.

Datatype: **BINARY DATA : BIN**

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: **BIN**

Binary data is a sequence of uninterpreted raw bytes (8 bit sequences, or octets).

Datatype: **BOOLEAN : BL**

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: **BL**

The boolean type stands for the values of two-valued logic. A boolean value can be either true or false.

Datatype: **CHARACTER STRING : ST**

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: **ST**

A string of characters where every character is represented by a uniquely identifiable entity within the string.

Datatype: **CODE VALUE : CV**

Is a Composite Datatype

Has Super Types: **DSCR**

Description of: **CV**

A code value is exactly one symbol in a coding system. The meaning of the symbol is defined exclusively and completely by the coding system from which the symbol originates.

Components of: **CV**

Code system : OID

An object identifier referring to the code system that defines the code value. The OID supports unambiguous reference to standard coding systems - including HL7 codes, as well as to local codes.

Code System Version : ST

A version descriptor defined specifically for the given coding system.

Print Name : ST

A sensible name for the code as a courtesy to an interpreter of the message. The name should not be considered as carrying the meaning of the code, it should never be sent alone, and it does not modify the meaning of the code.

Replacement : ST

A name for the concept whose meaning is being conveyed. The replacement is used if the concept cannot be captured by a code in the specified coding system. If the value attribute is set, the replacement attribute MUST NOT be set. In no way can a replacement string modify the meaning of the code value.

Value : ST

This is the plain symbol. E.g., "784.0"

Datatype: **DISCRETE : DSCR**

Is a Primitive Datatype

Has Sub Types: **ST**
REAL
OID

N
IVL
INT
II
CV
BL
BIN

Description of: **DSCR**
Abstract generalized type for any discrete type.

Datatype: **FREE TEXT : FTX**

Is a Composite Datatype

Description of: **FTX**
This free text datatype can convey any data whose primary purpose is to be shown to people for interpretation. Free text can be any kind of text, whether written language (formatted or unformatted) or multi-media data.

Components of: **FTX**

Charset : CV
Definition of the character encoding if different from the default encoding.

Compression : CV
Indicates that the raw byte data is compressed, and which compression algorithm is being used.

Data : BIN
Contains the free text data as raw bytes.

Media Descriptor : CV
Allows selection of the appropriate free text data. The default value is "text/plain".

Datatype: **GENERAL TIME SPECIFICATION : GTS**

Is a Primitive Datatype

Description of: **GTS**
This is a primitive datatype that is conceptually an arbitrary set of points in time. It is any combination of 1) a point in time, and 2) an interval of time. This includes

uncertain points and intervals of time. The contents of a GTS instance contains values that are defined in terms of a literal expression syntax that allows statement of any needed frequency or time pattern.

For example, these are some ways in which the GTS datatype is used within the PHCDM:

Dates are represented based on the precision needed and/or supplied. Y1999 represents the year 1999, while Y199909 indicates September, 1999, and Y19990926 indicates September 26, 1999. If only month and day are available, M0926 indicates September 26. The following are some of the period identifiers available: Y for year, M (or MY) for month of the year, D (or DM) for day of the month, H (or HD) for hour of the day.

The tag WY (w for week and y for year) indicates week of the year. Therefore WY23 is the 23rd week of the year.

Date ranges are indicated by a pair of dates separated by a dash, "-". For example, "Y20000110 - Y20000204" indicates a period beginning on January 10, 2000 and ending on February 4, 2000. Open-ended periods can be indicated by a date either preceded by a dash (period ending on the date indicated), or by a date followed by a dash (period beginning on the date indicated).

Durations are indicated by square brackets, "[" and "]". The duration amount and unit are within the brackets. For example "[10 min]" indicates 10 minutes, and "[3 day]" indicates 3 days.

These expressions can be concatenated together to express the union of multiple time concepts. For example, "M09 D26" is an alternate way of indicating September 26. Also, "Y20000224 [8 hour]" indicates a duration of 8 hours on February 24, 2000.

As a general statement, the construction of GTS instances is based on the following ideas:

- Singular time intervals as continuous sets of time points, specified through low and high boundary or width (in case no boundary is known.)
- Periodic time intervals as discontinuous sets of time points, specified through a period duration and a time offset (phase) interval.
- The set-operations intersection and union on such continuous and discontinuous sets of time to form arbitrary sets of time.
- The reduction of any arbitrary set of time into an outer bound interval and a sequence of occurrence intervals, no matter how complex the definition of this arbitrary set is.

- The use of probability distribution datatypes to account for the uncertainty in scheduling and time orders, or, in other words, to allow "fuzzy" constraining of time sets.

Time can be specified in terms of an absolute even flow of time, or events taking place in time can be aligned to calendars.

Datatype: **INSTANCE IDENTIFIER : II**

Is a Composite Datatype

Has Super Types: **DSCR**

Description of: **II**

The datatype is used to uniquely identify an entity that exists within a computer system or other well-controlled identification scheme.

Components of: **II**

Assigning Authority Identifier : OID

The ISO object identifier for the organization or identifier issuing scheme that is responsible for the integrity and validity of the identifier. This field guarantees the uniqueness of the identifier, and permits the origin of the identifier to be determined. If the organization uses OIDs for internal object identifiers, this may be the only field valued.

Assigning Authority Name : ST

The name of the organization or scheme responsible for the identifier.

Identifier Type Code : CV

A code representing the type of identifier. For example, the code might represent the US national provider ID, US national payer ID, medical record number, and social security number.

Identifier Valid Time Range : IVL<TS>

The time range during which the identifier is valid. It may be undefined on either side since in some cases only the start date for ID validity will be known, while, in others, only the end date is available.

Value Text : ST

The character string value of the identifier. For example the character string "123-45-6789" for a medical record number.

Datatype: **INTEGER : INT**

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: **INT**

Integer numbers are precise numbers that are results of counting and enumerating. The set of integers is infinite but countable. Two special integer values are defined for positive and negative infinity.

Datatype: **INTERVAL : IVL**

Is a Generic Datatype

Has Super Types: **DSCR**

Description of: **IVL**

Generic datatype that can express a range or interval of values. An interval is a set of consecutive values of any totally ordered datatype. An interval is thus a continuous subset of its base datatype.

Datatype: **ISO OBJECT IDENTIFIER : OID**

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: **OID**

The ISO Object Identifier is defined by ISO.

Datatype: **NUMBER : N**

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: **N**

The representation of a number. It is used as a generalized type for different numeric representations.

Datatype: **ORDERED : ORD**

Is a Primitive Datatype

Has Sub Types: **TS**
QTY

Description of: **ORD**
Abstract generalized type that at least contains naturally ordered subsets.

Datatype: **ORGANIZATION NAME : ON**

Is a Composite Datatype

Description of: **ON**
A name for an organization, such as "Centers for Disease Control and Prevention".

Components of: **ON**

Type Code : CV
A code identifying the use for an organization name. Possible values include: L - Legal, A - Alias, D - Display, ST - Stock Exchange.

Value : ST
The actual name data as a simple character string.

Datatype: **PERSON NAME : PN**

Is a Composite Datatype

Description of: **PN**
A Person name is one full name of a person. A name such as "Jim Bob Walton, Jr." is one instance of a Person name. The parts of this name "Jim", "Bob", "Walton", and "Jr." are person name parts.

Components of: **PN**

Value : LIST<PNXP>
This contains the actual name data as a list of name parts that may or may not have semantic tags.

Datatype: **PERSON NAME PART : PNXP**

Is a Composite Datatype

Description of: **PNXP**

This type is not used outside of the Person Name datatype. Person Names are regarded as token lists. Tokens usually are character strings but may have a tag that signifies the role of the token. Typical name parts are given names and family names; other part types may be defined culturally.

Components of: **PNXP**

Classifiers : SET<CV>

Classifications of a name part. One name part can fall into multiple categories, such as given name vs. family name and name of public record vs. nickname.

Value : ST

The value of a name part.

Datatype: **PHYSICAL QUANTITY : PQ**

Is a Composite Datatype

Has Super Types: **QTY**

Description of: **PQ**

A physical quantity results from a measurement act. It consists of a value and a unit.

Components of: **PQ**

Unit : CV

The unit of measure. Typically this is a unit, such as kilograms or miles per hour, that is drawn from a table of units of measure. Note that "count" is also included as a unit. This is used, for example, when collecting information about the number of interventions of a particular type.

Value : REAL

The magnitude of the quantity measured in terms of the unit.

Datatype: **POINT IN TIME : TS**

Is a Primitive Datatype

Has Super Types: **ORD**

Description of: **TS**
A point in time is a scalar defining a point on the axis of natural time.

Datatype: **QUANTITY : QTY**

Is a Primitive Datatype

Has Super Types: **ORD**

Has Sub Types: **PQ**

Description of: **QTY**
Abstract generalized type for any quantitative type.

Datatype: **REAL : REAL**

Is a Composite Datatype

Has Super Types: **DSCR**

Description of: **REAL**
A numerical amount. In order to facilitate computer representation, this is, by assumption, a floating-point number.

Components of: **REAL**

Precision : N
The precision of the floating point number in terms of the number of significant decimal digits.

Value : N
The value, expressed as an integer.

Datatype: **SET : SET**

Is a Generic Datatype

Description of: **SET**

SET is an unordered collection of unique items.

Datatype: **TELECOMMUNICATION ADDRESS : TEL**

Is a Composite Datatype

Description of: **TEL**

This is a token assigned as a mechanism for locating a telecommunication device such as a telephone, website, or email address.

Components of: **TEL**

Address : URI

This is an arbitrary address string that uniquely identifies an address in a particular domain.

Use Code : SET<CV>

The purpose of the "use code" is to advise in a system or user's selecting an appropriate telecommunication address to reach a party for a given telecommunication need. The following mandatory value domain is defined: PR - primary residence (home) OR - other residence (other home) WP - work/business/office communication address VR - vacation residence AS - automated answering service EC - emergency contact BP - beeper/pager CL - cellular/wireless phone

Valid Time : GTS

This is a General Time Specification (GTS) that identifies the periods of time during which this telecommunication address can be used. For a telephone number this can indicate the time of day in which the party can be reached on that telephone. For a web address, it may specify a time range in which the web content is promised to be available under the given address

---

Datatype: **UNIVERSAL RESOURCE IDENTIFIER : URI**

Is a Primitive Datatype

Description of: **URI**

The URI is used to refer to addresses of communicating entities used in order to transmit any kind of information. This may be used for messaging addresses, and for non-computer communication.

Model Scenario

This section presents an example public health scenario to show how the classes within the Public Health Conceptual Data Model correspond to data that are encountered over and over in the course of public health work. The scenario is presented in the left-hand column, and it is reinterpreted in the right-hand column in terms of concepts from the PHCDM.

Meningitis Outbreak Scenario

Scenario Text	Scenario Text Reinterpreted in PHCDM Terms
On December 14 the Health Department for County Z in the northern part of State F was notified of a 2 year-old girl who had presented to Hospital H a day earlier with fever, nausea, vomiting, and a petechial rash which was suspected to be meningococcal sepsis.	A Notification, a subtype of Health-related Activity, was recorded on December 14. Two Parties, a Person (the girl) and a Formal Organization (the hospital), are related to the Notification. The girl is related to the Notification as Target Participant. The hospital is related as an Actor Participant. Four Observations (fever, nausea, vomiting, and petechial rash), which are a subtype of Health-related Activity, are linked to the Notification. An additional Observation (suspected meningococcal sepsis) is recorded and linked to the Notification.
On the same day, a blood specimen was drawn. The specimen that was cultured from this girl grew *Neisseria meningitidis*, confirming the suspected diagnosis.	A Specimen, a subtype of Material, was drawn. An Observation (presence of *Neisseria meningitidis*) based upon testing the Specimen leads to a further Observation (confirmation of the suspected diagnosis).

Scenario Text	Scenario Text Reinterpreted in PHCDM Terms
The Local Health Department consulted with the State Health Department regarding recommendations for antimicrobial chemoprophylaxis of close contacts of sporadic cases of meningococcal disease.	The recommended Intervention (antimicrobial chemoprophylaxis), a subtype of Health-related Activity, is based upon another Health-related Activity (consultation) between two Formal Organizations (the Local and State Health Departments). These Health-related Activities are linked by an Activity Relationship.
Based on this consultation, rifampin or ceftriaxone, in recommended dosages and schedules, was administered to members within the girl's household and also to other attendees and staff in the day care center she attended.	An Intervention (rifampin or ceftriaxone) is administered to several Parties that are Informal Organizations (members of the girl's household and other attendees and staff at the day care center she attended). The attendees and staff members as well as the girl herself are related to a Formal Organization (the day care center). Each relationship is a Party Relationship with the day care center.
On December 15, the Health Department was notified of an 18 year-old female who had been admitted to the hospital the day before with fever, headache, and a stiff neck.	Another Notification was recorded on December 15. A Person (18 year-old female) who was the Target of several Observations (the history or presence of fever, headache, and stiff neck) is also the Target of another Health-related Activity (hospital admission). The Observations have Activity Relationships that link them to the admission.
Cultures of cerebrospinal fluid grew *N. meningitidis*.	An Observation (presence of *N. meningitidis*) was recorded after performing a Health-related Activity (culturing) on a Specimen (cerebrospinal fluid).

Scenario Text	Scenario Text Reinterpreted in PHCDM Terms
Over the next 2 weeks, five more cases occurred with signs of meningitis.	Several Cases, subtypes of Observations, were recorded.
In three of those cases, CSF cultures were positive for *N. meningitidis*; in the other two, latex agglutination tests were positive though cultures were negative. These cases occurred in persons 4 to 18 yearsold.	In each Case, Health-related Activities (culturing and testing) on Specimens (cerebrospinal fluid) were carried out. Observations related to the tests were recorded.
The occurrence of multiple cases prompted discussions among the Local Health Department, the State Health Department, and CDC personnel.	Formal Organizations (Local and State Health Department, CDC) were involved in consultation and discussion.
Available *N. meningitidis* isolates from the cases with positive cultures were forwarded to the State F Public Health Laboratory where they were shown to be serogroup C.	Specimens (*N. meningitidis* isolates) were forwarded to a Formal Organization (State F Public Health Laboratory). At the laboratory, Health-related Activities (tests) were performed and generated Observations (the isolates are serogroup C).
Based on the conclusion that a cluster of meningococcal diseases due to serogroup C *N. meningitidis* was occurring in Towns A and B, a decision was made to vaccinate.	Public health authorities confirmed the existence of an Outbreak (cluster of meningococcal cases due to serogroup C *N. meningitidis*) in a particular Location (Towns A and B) and decided on an Intervention (vaccination).

Scenario Text	Scenario Text Reinterpreted in PHCDM Terms
Between December 29 and January 1, a vaccination campaign was initiated targeting residents of Towns A and B (total population 33,000) between the ages of 2 and 22. Approximately 13,500 persons were vaccinated with polyvalent meningococcal polysaccharide vaccine.	A large scale Intervention (vaccination campaign) was initiated. The Target for the Intervention was an Informal Organization (residents of Towns A and B between the ages of 2 and 22). Within this context, many individual Interventions (vaccination with polyvalent meningococcal polysaccharide vaccine) were performed.

Frequently Asked Questions

1. **What is data modeling?**

   Data modeling is the process of analyzing and representing the things (i.e., classes) that an organization must understand. A data model represents the business facts (i.e., attributes) that the organization must know about the classes, along with their associated relationships and business rules. Data modeling requires that business personnel, facilitated by a data architect, wrestle with and achieve consensus on the definition of what the specific classes and their associated attributes are.

   The purpose of data modeling is to develop an accurate model, or graphical representation, of the client's information needs and business processes. A well-developed data model is the architectural blueprint that enables stable and flexible database and application development**.** The data model acts as a framework for the development of new or enhanced applications.

2. **What is a conceptual data model?**

   A conceptual data model is a high-level or abstract data model. It is based on an analysis of client requirements and it describes the interests of an organization or business. Conceptual data modeling is one of the most powerful and effective analytical techniques ever developed for understanding and organizing the information required to support any enterprise. This form of model focuses on the big picture, and the really important strategic objectives that will assure the health and prosperity of the enterprise. Data subjects (i.e., classes) are shared across functional, process, and organizational boundaries in the business. As a result, this model (and the applications and databases which will be built from it) is the linchpin for removing waste and unnecessary time and cost in the conduct of the enterprise processes by increasing shared use and avoiding redundancy.

3. **What is a logical data model?**

   A logical data model forms, defines, and standardizes data elements so that they can be shared for all business purposes, instead of being tailored to fit only one business unit or agency's use or point of view. Logical models ensure that the organization of the data is also optimal for all uses. A logical data model really a blueprint of the data across an entire organization, irrespective of platform, operating system, file structure, or database technology. It defines your entire data landscape.

   Logical data modeling is concerned with structuring the data to conform with established database management principles. The logical data model represents the actual data requirements that exist on forms and reports of an

information system. The logical data model for an enterprise is based on its conceptual data model, but it will include an additional level of detail, and more rigorous expression of data structures.

4. What is a physical data model?

The physical data model captures the physical constraints imposed by the business (performance of certain transactions, data volume limitations, physical distribution of the data at different locations, and security protection of selected data), and creates a best-balanced physical design of the computer database which meets those (sometimes conflicting) constraints without sacrificing the natural subject organization of the data. Each physical model is based on the logical data model, but takes account of performance requirements and the special characteristics of the chosen implementation software.

5. How do the three models fit together?

Once user requirements are gathered, a conceptual model is designed. This model provides an overview of the subject areas (classes and their relationships). The next step is to design a logical data model. The logical model focuses on the data required by the system, independent of how the data are stored, the database software, and how the data are processed. When the logical model is complete, a detailed physical model(s) can be developed. The physical model is the logical model implemented according to the constraints of the chosen database management system (DBMS) software.

6. Does a conceptual data model tell me how to set up a computerized system?

No, the conceptual data model expresses the data requirements of the enterprise independently of enterprise functional requirements (such as for volume of transactions, response time, and data security), and independently of the characteristic features of the chosen application and the database software. However, when the enterprise follows a disciplined application development strategy, the logical and then the physical data models will be developed using the conceptual data model as the basis and starting point.

7. If the PHCDM is not a physical data model, then how do I use the PHCDM in developing my system?

The conceptual model provides a framework for the development of new or enhanced applications. It is a blueprint that database and software developers use to design and implement new or enhanced software information systems. First, the logical data model is developed through a process of refinement from the conceptual data model, and later, the physical data model is constructed on the basis of the logical data model.

8. How does the PHCDM relate to data standards?

The PHCDM provides a framework from which we can identify and organize public health concepts and data standards. Below are some examples of how the PHCDM can be used to further data standardization efforts:

- Categorize areas to identify where standard code sets are needed, e.g., codes for demographic characteristics, test names, or risk factors. These code sets may be borrowed from existing national standards or, when not available, may be developed through collaboration between CDC and public health partners.

- Mappings from existing logical data models (CDC, state, local and other relevant systems) to the PHCDM can help determine whether the PHCDM truly reflects and represents the broad spectrum of public health. The mapping process may result in the identification of common concepts between data models (as well as the identification of concepts requiring standardization that are not represented in the PHCDM.) Common concepts may include general demographic, case management, and behavioral risk factor information (e.g., smoking, IV drug use, and unprotected sex).

- The PHCDM can serve as a springboard for discussion with standards development organizations (SDOs). Presenting the PHCDM to SDOs is one way of ensuring public health needs are considered in future revisions of national standards. If the SDO has a data model (e.g., HL7 Reference Information Model), comparisons of relevant areas of the PHCDM to the SDO model may provide an opportunity for further harmonization of data concepts.

A CDC standards body will be involved in such data standards activities. The PHCDM will help facilitate the iterative process of identifying subject areas and data concepts requiring standardization and disseminating data standards as they are adopted. Collaboration with CDC staff, public health partners, and standards development organizations is necessary to resolve differences in the way concepts are defined and represented from each perspective. Once these differences are resolved, and a common concept(s) can be agreed upon, use of the newly standardized concept(s) in subsequent iterations of the model(s) and existing/future system design will allow for the exchange of this information in a more uniform, consistent way.

9. How does this model facilitate data exchange?

The conceptual model, when it is shared across application and organizational boundaries, expresses the data that are common, and that will be shared. In effect, it provides a standard for data definitions and representation. Applications and databases built on basis of the conceptual model need to

store the data in a form consistent with that model or support a mapping to that format, in order to support data exchange.

10. Who should be the users of this model?

Parties within public health who need to use or manage public health data should utilize the model. This includes subject matter experts, end-users that analyze data and create reports, and application and database developers.

11. What is the role of the data model in the National Electronic Disease Surveillance System (NEDSS)?

The National Electronic Disease Surveillance System (NEDSS) is a set of related activities designed to electronically integrate and link together the myriad information systems currently used for public health surveillance. The goal of NEDSS is to facilitate more efficient, accurate, and timely collection, interpretation, and use of data by communities, Local and State Health Departments, and CDC to improve the health of the public. Other NEDSS activities include an information systems architecture based on industry standards, including a standard user interface for CDC surveillance information systems; shareable tools for interpretation, analysis, and dissemination of data; a network for secure internet-based data transfer that is also based on industry standards. Because data are vital to doing our jobs in public health, the PHCDM is a critical element of NEDSS; it documents our information needs in public health, providing a framework for organizing data standards and guidelines and facilitating data comparability and exchange with other systems. For example, we are using the PHCDM as a tool to communicate public health data needs to national health informatics standards setting bodies, to enable development of standards for the exchange of information among public health and healthcare providers.

12. What is the scope of this model? Does it include all public health activities?

At this point the primary focus of this model is to represent data needs for the key "Essential Public Health Functions" (as defined by the United States Public Health Functions Steering Committee in 1994, available at http://www.health.gov/phfunctions/project.htm) of "Monitor health status to identify community health problems" (including public health surveillance) and "Diagnose and investigate health problems and health hazards in the community". We are developing a process model to help explain the scope of the model and to relate the concepts included here directly with specific public health activities.

13. What is a process model, and how does a data model differ from a process model?

A process model is a model of the activities, functions, and processes of an organization. Processes in a process model are often defined in terms of their inputs and outputs. Because of this, a data model that provides definition for the data consumed and produced by a process often accompanies a process model. A data model does not reflect any action or flow of information. Only a static view of data is presented in a data model.

14. Does this represent ALL of public health, i.e., local, state, and federal issues?

It is hoped that this model will be broadly representative of public health issues at all levels. The data model used in Missouri was used as input to the development of this model, and we have held a meeting and requested feedback from our partners in State and Local Health Departments. However, the purpose of publishing the model is to ensure its use and critique by our partners in community organizations, and local, state, and federal agencies. We are requesting your feedback, to ensure the continued growth and improvement of this model, and to ensure that it truly represents all of public health's data needs. For example, due to a relative lack of focus at the CDC level on direct provision of healthcare, contact follow-up, or issues with patient insurance coverage, this model may incompletely represent data needs associated with financial management of patients, or follow-up of contacts of a case of communicable disease. Please provide specific suggestions for possible inclusion in the model.

15. Does this model include subject areas and classes related to both infectious and non-infectious disease?

This model is generally intended to represent data needs in public health. It should be noted that concepts included in the model, such as individuals, locations, risk factors, and interventions, are generic and important to many kinds of public health activities. We have tried to be inclusive in our thinking as we developed this model, although participants in model development have included mostly colleagues with backgrounds in infectious diseases. However, there has been participation from our colleagues in other areas, such as injury prevention and environmental health, and we hope through publication of the model to obtain specific suggestions on how to facilitate its representation of data needs for the broad spectrum of public health issues.

16. How should State or Local Health Departments use this model?

State or Local Health Departments that already have data models should compare their models to the PHCDM in order to determine what changes may be needed to ensure that the PHCDM truly represents public health at all

levels. If a State or Local Health Department does not yet have a data model underlying its information system design and development, it should consider evaluating and adopting the PHCDM.

17. How will the PHCDM be maintained?

A team at CDC consisting of subject matter experts, data modelers, and technical specialists will maintain the PHCDM. As the model is applied to public health initiatives such as system development, database design, interface specifications, and other data standardization initiatives areas where the model is weak will be reviewed and attempts will be made to enhance the model to make it useful.

CDC is developing a process to receive comments from its partners concerning the data model. These comments will be used to identify areas of the model that need refinement or extension. For example, in the early versions of the PHCDM, there was a class entitled, "patient coverage", which referred to the health insurance that would (if it existed) pay for the patient's clinical care. However, it was determined that there were not enough systems at CDC that required this information; thus, we had no data to use to flesh out this class (with its attributes, for example). If our public health partners feel that the PHCDM should include this class, we would welcome their suggestions on its attributes and associations with other classes.

The model will be continuously maintained, but published no more frequently than once per year.

Glossary

American National Standards Institute (ANSI): A voluntary standards organization that serves as the coordinator for national standards in the United States and the U.S. member body to the International Organization for Standards. ANSI accredits standards committees and provides an open forum for interested parties to identify, plan, and agree on standards; it does not itself develop standards. Standards are developed by Standards Development Organizations (SDOs).

URL: www.ansi.org

Association: An association is a structural relationship that specifies that instances of one thing are connected to instances of another.

Attribute: The specific items of data that can be collected for a class.

Centers for Disease Control and Prevention (CDC): An agency of the Department of Health and Human Services that promotes health and quality of life by preventing and controlling disease, injury, and disability.

URL: www.cdc.gov

Common Information for Public Health Electronic Reporting (CIPHER): A set of standards and guidelines for data representation and code values which includes specifications for representing concepts as well as standard code lists for coded elements. The CIPHER standards can be linked directly to attributes in the PHCDM.

URL: http://www.cdc.gov/od/hissb/docs/cipher.htm

Class: A description of a set of objects that share the same attributes, relationships, and semantics.

Data Model: A framework for the development of a new or enhanced application. The purpose of data modeling is to develop an accurate model, or graphical representation, of the client's information needs and business processes.

Datatype: A specification of the allowed format for the values of an attribute. Examples include string, number, code, and text.

Electronic Data Interchange (EDI): A standard format for exchanging business data. An EDI message contains a string of data elements, each of which represents a singular fact, such as a price, product model number, and so forth, separated by delimiters (a character that identifies the beginning and end of a character string). The entire string is called a data segment. EDI is one form of e-commerce, which also includes e-mail and fax.

HIV/AIDS Reporting System (HARS): A computerized management information system developed to assist State and Local Health Departments in managing and analyzing HIV/AIDS surveillance data.

URL: http://www.cdc.gov/hiv/software/hars.htm

Health Information and Surveillance Systems Board (HISSB): A board established by the authority of the Director of CDC/ATSDR. The mission of HISSB is the formulation and enactment of policy concerning the planning, development, maintenance, and use of public health information and surveillance systems.

URL: http://www.cdc.gov/od/hissb/

Health Level 7 (HL7): A standards development organization formed in 1987 to produce a standard for hospital information systems. HL7 received ANSI accreditation as an Accredited Standards Developing Organization in 1994. The HL7 standard is an American National Standard for electronic data exchange in health care that enables disparate computer applications to exchange key sets of clinical and administrative information. HL7 is primarily concerned with movement within institutions of orders; clinical observations and data, including test results; admission, transfer and discharge records; and charge and billing information (coordinating here with ASC X12). HL7 is the selected standard for the interfacing of clinical data for most health care institutions.

URL: www.hl7.org

HL7 Reference Information Model (HL7 RIM): A conceptual model that defines all the information from which the data content of HL7 messages is drawn.

URL: www.hl7.org

International Organization for Standardization (ISO): A worldwide federation of national standards bodies from some 100 countries, one from each country. Among the standards it fosters is Open Systems Interconnection (OSI), a universal reference model for communication protocols. Many countries have national standards organizations, such as the U.S. American National Standards Institute (ANSI), that participate in and contribute to ISO standards development.

URL: www.iso.org

Laboratory and Epidemiological Public Health Information Tracking and Reporting System (LITS+): A computerized information management system that provides seamless integration of laboratory and epidemiologic

data. The system was developed by CDC and is in use in many public health and reference laboratories.

URL: http://www.cdc.gov/ncidod/dbmd/litsplus/newsletters/fall99.pdf

National Electronic Disease Surveillance System (NEDSS): A set of related activities designed to electronically integrate and link together the myriad information systems currently used for public health surveillance. The goal of NEDSS is to facilitate more efficient, accurate, and timely collection, interpretation, and use of data by State and Local Health Departments and CDC to improve the health of the public. When complete, NEDSS will include data standards; tools for interpretation, analysis, and dissemination of data; a network for secure internet-based data transfer that uses industry standards; and policy-level agreements on data access, sharing, burden reduction, and protection of confidentiality.

URL: http://www.cdc.gov/od/hissb/

National Electronic Telecommunications System for Surveillance (NETSS): A computerized public health surveillance information system that uses a standard ASCII record format to provide CDC with weekly data regarding cases of nationally notifiable diseases.

URL: http://www.cdc.gov/epo/dphsi/netss.htm

Public Health Conceptual Data Model (PHCDM): A high level conceptual model, developed as part of the CDC NEDSS initiative, which provides the foundation for standardization of public health data collection, management, transmission, analysis and dissemination.

Sexually Transmitted Diseases Management Information System (STD*MIS): A management information system designed to assist state and local STD control programs in managing and operating their programs. The system allows programs to maintain surveillance data on both laboratory and case reports, allows tracking and monitoring of disease intervention activities such as patient interview and follow up, and allows capture of data regarding patient clinic visits and outcomes.

URL: http://www.cdc.gov/nchstp/dstd/STD-MIS.htm

Subject Area: A way of organizing classes into groups within a model, where classes grouped together into higher-level units. Within the UML, a subject area is referred to as a package.

Subtype: A specialization of another class, which inherits the attributes of its parent class.

Supertype: A generalized class that is related to subtypes that inherit its attributes.

Tuberculosis Information Management System (TIMS): A software program that automates the administration of tuberculosis prevention, surveillance, and control programs. TIMS facilitates the management of TB cases and the tracking and reporting of TB control program activities. Data collection and transcription on all TB clients can be performed daily at the service delivery level at one or more facilities.

URL: http://www.cdc.gov/nchstp/tb/tims/tims.htm

Unified Modeling Language (UML): A graphical language for visualizing, specifying, constructing, and documenting the artifacts of a software-intensive system.

URL:http://www.omg.org/uml/

World Wide Web Consortium (W3C): An industry consortium that seeks to promote standards for the evolution of the Web and interoperability between WWW products by producing specifications and reference software.

URL:http://www.w3.org/

X12: A standards development organization that develops uniform standards for inter-industry electronic interchange of business transactions-- electronic data interchange (EDI). X12N, a subcommittee of X12, develops standards for healthcare insurance and claims processing.

URL: http://www.disa.org/

eXtensible Markup Language (XML): A specification developed by the World Wide Web Consortium. XML is designed especially for Web documents. It allows designers to create their own customized tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations. XML provides a file format for representing data, a schema for describing data structure, and a mechanism for extending and annotating HTML with semantic information.

URL: http://www.w3.org/

Bibliography

Booch G, Rumbaugh J, Jacobson I. *The Unified Modeling Language User Guide*. Reading, MA: Addison-Wesley, 1999.

Institute of Medicine. *The Future of Public Health.* Washington, DC: National Academy Press, 1988.

Fowler M.. *UML Distilled*. Reading, MA: Addison-Wesley, 1997.

Health Level Seven, Modeling & Methodology Technical Committee. *HL7 Message Development Framework*. 1999.

Health Level Seven, Modeling & Methodology Technical Committee. *HL7 Reference Information Model.* Version 0.96. 2000.

Centers for Disease Control and Prevention. *Integrating Public Health Information and Surveillance Systems: A Report and Recommendations from the CDC/ATSDR Steering Committee on Public Health Information and Surveillance System Development*. 1995.
URL: http://www.cdc.gov/od/hissb/docs/katz.htm