Strategic Health Information Requirements

Executive Summary

In both the private and public sectors, the modern information revolution has created large data stores which, instead of increasing productivity and efficiency, are threatening to drown businesses in a flood of useless information. The greatest challenge to public health today is to sift through its large stores of data for useful information that can be leveraged into valid prevention, promotion and assessment activities.

The data collected by vital records, immunization registries, health programs, WIC, and Medicaid are necessary sources of information to drive health policy, planning, and population based health outcomes. Several fundamental problems exist that prevent the effective use of these data by the health system.

- There is no consistent data quality policy that assures consistency, accuracy, timely delivery, and uniform data packaging;
- The data are collected using systems supplied by many vendors, and are stored in proprietary and incompatible formats;
- Record management is localized at the point of entry or point of service with little or no attention given to making data available for executive management;
- Data stores are fragmented;
- The non-cohesive regional structure of the health system tends to perpetuate and expand all of these problems;
- Timely delivery of health information is difficult due to staffing problems, data access problems, labor intensive data management, and poor use of available data analysis technologies.

A solution to these problems is to use two information technologies that were designed specifically to improve data storage, data access, and decision making. Data warehousing is a data storage and packaging technology that has several benefits to the health system.

- Packages essential business data into a repository that has been standardized according to data quality policies;
- Provides a basis for strategic planning efforts since essential business data are linked and readily accessible;
- Improves business decisions such as, health planning and health promotion, in terms of both quality and quantity by allowing composite views of all essential business data;
- Reduces the reliance on labor intensive data management (data can be accessed without programmers);
- Allows high performance data mining of multi-dimensional data (the standard epidemiological model of who, what, when and where).

Data mining is used on warehoused data to produce business intelligence. Public health is a customer-focused system, whose product is the improvement of the health of the public customer. If a data warehouse is the memory of the health system (its vital events, services and programs), then data mining is the processing of the memory to produce knowledge about the health system.

- Allows for the prediction of outcomes by health programs, prevention and promotion strategies;
- Allows for proactive health monitoring since all essential business data are leveraged into health status knowledge based upon epidemiological methods;
- Creates accurate and reliable needs assessment and planning scenarios for populations and communities at several levels of granularity.
- Presents time as a factor in decision making, resource allocation, intervention actions, and planning.

Data warehousing and data mining are mature technologies that have emerged into the mainstream due to low cost, high performance hardware systems and database technologies. However, the construction of a data warehouse requires planning based upon clear data quality and management policies. Data mining requires a clear understanding of the kinds of business intelligence necessary to drive the health system. Data mining methods should be used to define the requirements for a data warehouse, since a system of memory and access must evolve towards its utility as a decision support system.

Strategic Plan

The purpose of the strategic plan is to outline a scheme for implementing a data warehouse and health knowledge discovery through data mining. The data warehouse is the foundation, or infrastructure, upon which the health knowledge discovery will be positioned. The purpose of the health information system is to leverage health information to manage the Georgia Division of Public Health from a process control model. Public health is a process control system. It monitors occurrences of illness, disease, injury, and medical procedures in order to assess the effectiveness of prevention, service, and service provision and health promotion functions.

Public health is currently experiencing rapid evolutionary change forcing it to focus on its core functions: prevention, health promotion and outcomes assessment. A key feature of this evolution is a departure from a medical/clinical (individual) based model of health delivery, towards population based health management build upon information and knowledge derived from evaluation and assessment of the health of the public. Information is the primary product of public health. Information provides the evidence and credibility needed to promote health education, health programs, assess prevention functions and support the health consumer.

The strategic plan is composed of details of the data warehouse and data mining construction.

Phased Implementation of the Health Information System

The health information system will be built in two overlapping, and largely concurrent, phases. The first phase is the data warehouse and it's supporting data marts. The second phase includes the means by which information and knowledge are acquired from the warehouse, and how this knowledge and information are presented to end-users.

The Data Warehouse

The design and deployment of the data warehouse will follow a generalized method accepted by information technology industries.

Project Startup

SCOPE AND PURPOSE CONFIRMATION

Explicitly defining project boundaries from a business and technical perspective provides positioning (context) and enables a project to move into high gear. Project team members and key stakeholders must have a shared vision of the project and its results. The central issues to be addressed include:

- Does the project add business value? What constitutes success?
- What forms the basis for competition in the marketplace of the future?
- Does the project address the key business drivers? Which are most important?
- What part of the business operations is affected? Which processes, subjects, locations, and organizations are impacted?
- Who are the key stakeholders and what are their expectations?
- Who are the consumers and producers of knowledge? What types of knowledge do they require?
- What technology standards are in place which constrain the project?
- What are the funding, schedule, and resource constraints?
- What is the relationship between the project and other data warehousing solutions or projects?

Knowledge Analysis

KNOWLEDGE REQUIREMENTS DEFINITION

Information is understanding the events of the past. Knowledge, on the other hand, is insight that enables prediction and creation of events in the future. The focus of data warehousing must be on the development of knowledge, and not just the reporting of information.

In this stage, the project team works with business customers to define the knowledge requirements for the data warehousing solution. Dialogue takes place in a natural business language, with many visualizations to help team members and business customers feel confident that their requirements are understood. The central issues to be addressed include:

- What are the business objectives, strategies, and problems that require measurement and analysis? What opportunities and challenges drive the need for knowledge?

- What types of insight, understanding, and learning are needed? Which are historical versus predictive in nature? Who needs this knowledge? Where is this knowledge needed, and how often?
- What are the elements of information that lead to insight, understanding, and learning? Where can these elements be acquired?
- What are the business rules that govern the elements of information?
- What types of analytical methods are used to develop knowledge?
- How are the results of these analytical methods visualized?
- How is information about knowledge requirements represented as metadata?

Note that star schema versus relational schema is not an issue here. These are storage structures which must be considered in the solution architecture. The knowledge models created in this stage are intended for a business audience.

DATA SOURCE AND QUALITY ANALYSIS

As the knowledge requirements are defined, the project team must map the requirements to the discrete elements of internal and external data sources. In many cases, the mapping is to data fields in existing OLTP systems. The team must also address one of the most difficult problems in data warehousing: how to handle "dirty" data. The central issues to be addressed include:

- What are the potential internal and external sources of information? What discrete facts (data) must be leveraged from these sources?
- Which sources are closest to the point of data entry?
- How often are the sources updated? How volatile and current are the sources?
- What business rules constrain each discrete fact of data? Which business rules are inconsistent with the business rules in the knowledge model?
- Are the business rules applied uniformly to the data sources? What percentage of the data is "dirty"?
- Which sources satisfy the knowledge requirements?
- How is information about the selected sources represented as metadata?

KNOWLEDGE USE ANALYSIS

As the knowledge requirements are defined, the project team must also understand the usage metrics (e.g., size, volume, frequency) for the requirements. Scalability is an important issue in data warehousing. The metrics on knowledge usage, which relate to scalability, must be visible and understood early. The central issues to be addressed include:

- What is the estimated size of each element of information?
- How many instances are contained in each information set?
- What analytical operations are performed? Which sets of information are involved? How often are the operations performed?
- How often is each information set refreshed with new data from internal or external sources? What are the sizes of the source data sets?
- How often are instances removed from each information set? What are the criteria for removal?
- How are statistics on information usage represented as metadata?

Solution Design

Solution Architecture Design

The project team—using its understanding of the knowledge requirements, usage, and data sources—designs the data warehousing solution. The solution architecture is a specification containing four of the key components of data warehouse solutions, presented earlier in this article. The central issues to be addressed include:

- What existing standards and infrastructure (that is, hardware, software, network) constrain the solution? How does the project team leverage the existing infrastructure to shorten the development time cycle?
- What data warehousing tools (in current use) must be considered?
- What classes of business customers must the solution support and where are the business customers? What levels of security must be provided?
- What end user tools (for example, query, reporting, analytical, visualization) must be provided?
- How is access to metadata provided to business customers? How is metadata kept consistent with data in the warehouse?
- What is the data architecture of the solution? What storage structures (for example, star schema, relational schema) and indexing methods must be used? How is the data summarized, partitioned, and distributed?
- What is the application architecture of the solution? What programs and logic are needed to extract, transform, cleanse, and load the operational data? How is the application logic partitioned and distributed?

- What is the technology architecture of the solution? What hardware, systems software, and networks are used to support the solution? How is the technology distributed to support the data and application architectures?

TESTING STRATEGY AND DESIGN

As the solution architecture is defined, the project team must develop strategies, procedures, and data to test the data warehousing solution. The four key components of the solution should be addressed in the test strategy and design. However, the data bridging component is likely to require the most testing. The central issues to be addressed include:

- What is the potential business impact on the enterprise (for example, decisions made or actions taken) from the use of the data warehouse? What degree of testing rigor must be applied?
- What degree of data correctness must exist? How is data correctness proven?
- How are data bridging programs tested? How is the correctness of extraction, transformation, cleansing, and loading logic proven?
- How are viewing tools tested? How is the correctness of the tool's results and visualizations proven?
- In what sequence are components tested? What procedures and test data are used?
- Who will perform the testing? How are test results reported and evaluated?
- What tests must be completed successfully prior to system acceptance?

PROCEDURE DESIGN

As the design of the solution architecture unfolds, the project team must also focus on the procedures that are required to operate and administer the data warehouse. The design of automated and/or manual procedures must provide availability, security, performance, and storage management. The central issues to be addressed include:

- What are the procedures for backup and recovery?
- What are the procedures for data refresh and archival?
- What are the procedures for monitoring and tuning data warehouse performance?
- What are the procedures for data warehouse security?
- What are the procedures for extending and scaling data storage?

WAREHOUSE ENVIRONMENT INSTALLATION

Before the development of the data warehouse solution begins, the hardware, system software, and network must be installed and/or made available to the project team. Each infrastructure component must be thoroughly tested and certified. Testing and certification of the environment can save the implementation team time in isolating problems during the development of the solution.

Solution Implementation

WAREHOUSE SOLUTION CONSTRUCTION

The project team builds the data warehousing solution according to the specifications of the solution architecture and design procedures. The project team should expect to spend most of its effort on the data bridging component of the solution architecture.

WAREHOUSE LOADING AND TESTING

As the project team assembles the components of the data warehousing solution, an independent testing team must prove the correctness of the solution. The testing strategy and design should guide the testing team as they construct test data, test the data bridging components, load the data warehouse, and test the viewing tool components. Business customers should test the viewing tool components and conduct the final acceptance test before the deployment of the data warehouse solution.

WAREHOUSE SOLUTION DEPLOYMENT

After its acceptance, the data warehouse solution is installed, tested, and made available at the targeted sites. The extraction and load processes bring the data warehouse to a production status. At this stage, end users and the operations staff receive documentation. Training helps end users understand the solution's potential and to become comfortable with the functionality of the viewing tools.

Project Wrap-up

WAREHOUSE SOLUTION REVIEW

After deploying the solution, the project team archives important project documentation and data, and conducts a final meeting to discuss the project process and results. The central issues to be addressed during the discussion include:

- Is the solution complete and correct? How satisfied were business customers with the solution?
- What portion of the original scope was "thrown overboard" and must be incorporated into a future project? What improvements must be made to the current solution?
- How well did the development process work? What improvements must be made to the workflow and techniques?

- How well did the project team work together? What improvements must be made to roles, responsibilities, and allocation of workload?

- How well did the development tools work? What workarounds or utilities must be implemented to increase usability or functionality?

Iterative Application of the Process

It is a well-known fact that implementing a data warehouse is an iterative process. Consequently, the previous process should be performed iteratively, under the control of a process that determines the sequence and scope of each project, as illustrated in Figure 8. Data warehousing scope and iteration management is an ongoing process that employs estimating and release management techniques to implement the data warehouse vision through a series of process iterations. Other ongoing processes, such as configuration management, change control, process and project management, and architecture management, must also interface with each iteration of the data warehouse process.

Georgia Division of Public Health

Data Management Requirements

Purpose

Data management requirements are necessary for the establishment of data quality standards and policies. These derived standards and policies shall drive physical and logical data management within the Georgia Division of Public Health, which shall include: hardware and software selection, data storage, data maintenance, data security, field (variable) and table nomenclature, field properties, data quality control, data collection, and exception handling.

Structure

This document shall be structured to provide data management requirements for basic data quality control methods and definitions, external interfaces of data sources used by the Division, application of basic quality methods to Divisional data sources, storage of Divisional data sources, and maintenance of Divisional data sources.

This document shall follow industry standard processes and use industry standard terminology, and shall be based upon biometric data analysis, Unified Modeling Language (UML), Object Oriented Analysis (OOA), and automated rule discovery, which are derived from Texel and Williams (1997), Martin and Odell (1995), Parsaye and Chignell (1993), Warner, Sorensen, and Bouhaddou (1997), Kennedy, Lee, Van Roy, Reed, and Lippman (1998), and Sokal and Rolf (1997).

The outline of this section shall cover, in order,

I. Introduction to Data: Definitions, Properties, and Quality Requirements
II. Data Quality Requirements Modeling with UML and OOA
III. Application of Data Quality Requirements
IV. Rule Based Data Quality Control

1. Introduction to Data: Definitions, Properties, and Quality

Data that are used for Public Health planning and assessment must be standardized so that analytical results are valid and reliable. Data are objects that have names, properties, and methods that assign to them meaning in the processes of Public Health data management, assessment, and planning.

1.1 Definitions used in requirements

BLOB – Binary Large Object, such as a scanned image.

Class – an abstraction of concepts that have the same properties and behavior; such as, vital records.

Concept – the ideas or notions that lead to the collection of observations and measurements, or that represent descriptions of phenomena; such as the notion that births or deaths need to be recorded.

Continuous Variable – a measurement variable that can assume an infinite number of values between any two fixed points, such as birth weight.

Data Anomaly – an exception to a domain specification or variable definitions that indicate errors in the data or failures in the rules contained in domain specifications. If an anomaly is due to a data error, then the data object is corrected. If an anomaly is due to a rule or standard failure, where the domain specification does not accommodate observed data, then the rules or standards should be corrected.

Data Collection – a set or class of data items that are contained within a single file or table; such as, births or deaths.

Data Item – a single collected variable; such as birth weight.

Data Type—The characteristic of a variable that determines what kind of data it can hold. Data types include Binary, having only 2 values; Real, a numeric that has precision as significant digits, i.e. a decimal point can be part of the number; Byte, Boolean, Integer, any positive or negative whole number, including zero. A numeric that represents a single unary value;

Long, Currency, Decimal, Single, Double, Date, a formatted date value; String, Text, ASCII data.

Derived Variables – a variable that is a function of two or more independently measured variables whose relations are expressed in a certain way, such as ratios, percentages, indices, and rates.

Discontinuous Variable – a measurement variable that can assume only fixed numeric values, with no intermediate values, such as counts.

Domain – a selected area of interest that contains a collection of objects that are instances of the domain specification. Domains can be observations, populations, measurements, or variables; vital records is a domain.

Domain Specification – the collection of concepts that apply to a domain; NCHS coding standards and rules are domain specifications.

Independent and Dependent Variables – if F is function or decision space, such that d = F(I), then I is the independent variable and d is the dependent variable, since a value of d is dependent on a value of I.

Individual Observations – observations or measurements taken on the smallest sampling unit.

Level of Analysis – the level of measurement assigned to a variable for analysis.

Level of Measurement – (1) nominal, categorical properties or labels. Used to describe characteristics that have no numerical values (e.g. gender, ethnicity); (2) ordinal, objects are ordered by some nominal category irrespective of magnitude, and irrespective to the distance between ordered levels. Used for characteristics that have an underlying order among them, although the order may be arbitrary. Examples of ordinal scales: frequently, often, sometimes, rarely, never; strongly agree, agree, disagree, and strongly disagree; (3) interval, ordering of objects is respective to a nominal category, the distance between objects respective to the nominal category, and without respect to the magnitude of the nominal category. This type of numerical scale has an arbitrary zero point such as Fahrenheit and Celsius, and thus 0 degrees does not mean no heat at all. Likewise, 40 degrees is not twice as hot as 20 degrees; and (4) ratio, objects are ordered respective to a nominal category, where the distance between objects is known, and each objects measurement is respective to a known zero value.

Measurement Variables – variables whose differing states can be expressed in a numerically ordered fashion.

Nominal Variables – a variable that assigns properties to observations, such as dead or alive, male or female, and black or white.

Object – anything to which a concept applies; an instance of a concept; such as, births or deaths.

Population – the totality of individual observations about which inferences are to be made, existing anywhere in a given universe of interest, and definitely specified by space and time.

Ranked Variables – a variable that indicates order in observations without any assumption placed on the magnitude between ranks.

Record – a single instance of an object; one birth or one death.

Reliability – the closeness of repeated measurements to the same value, or a process with inputs of equal value, will always produce outputs of equal value; also known as precision.

Sample Observations – a collection of individual observations, from a population, selected by a specific procedure or program.

Unit of Analysis – the unit of measurement assigned to a variable for analysis.

Unit of Measurement – (a) refers to the system of measurement: English or metric (CGS and MKS) or SI; (b) the specific unit, within a measurement system, at which measurements for a variable are made.

Validity – the closeness of a measured or computed value to its “true” value, or a process that measures exactly one and only one specific phenomena; also known as accuracy.

Variable – the actual property measured by the individual observations, and indicates the degree to which individuals in a sample differ.

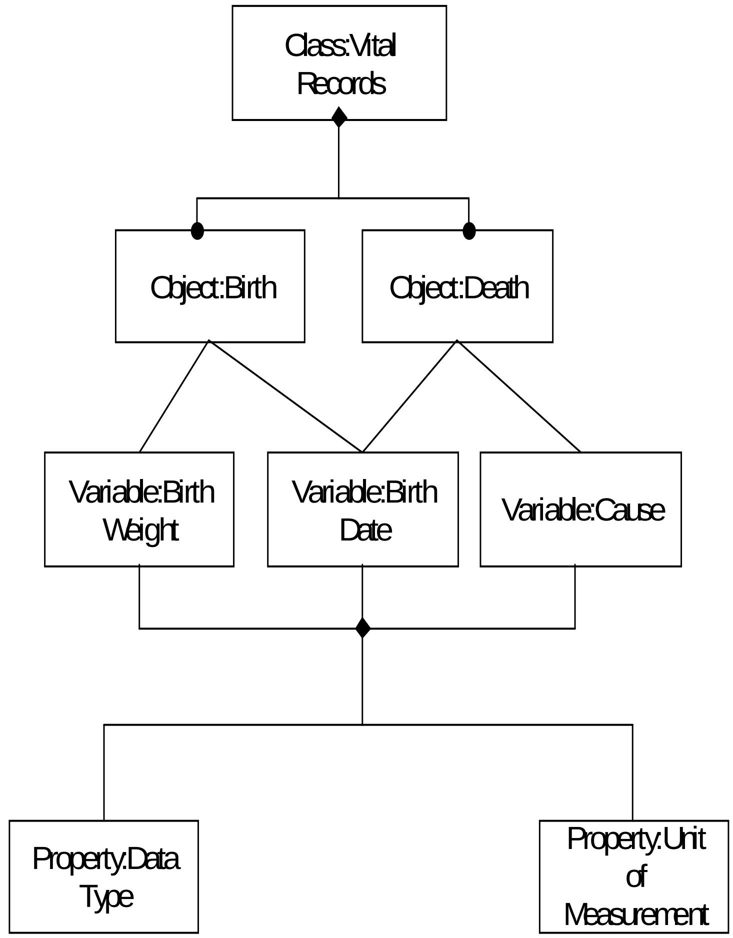
Variate -- a single reading, score or observation of a given variable.

1.2 Data property notation used in requirements

Class.Object.Variable – Defines a variable of a specific class of objects so that a class has objects that have variables; such as, VitalRecords.Birth.BirthWeight .

Variable.Property – Defines a variable property so that a variable takes on that property; such as, BirthWeight.DataType.Double; or, BirthWeight.Unitof Meas.CGS.GM.

Diagrammatically, data property relations are shown on the next page. The notation used and the object diagrams focus on variable and property consistency. For example, both birth and death objects have a date of birth variable, which has a single data type and unit of measurement. The diagram helps to show how requirements for variables shall be determined, since it is important to illustrate variables that are contained in more than one objects must have a unary definition.



Data have properties that define their storage and analysis specifications. The Georgia Division of Public Health shall use the following data properties:
Name – Standard reference name of variable.
Storage name – Standard name used when storing data element in a database.
Presentation of Values –
Standard format used to present the variable's value.
Definition – a) purpose: the reason to collect or use the variable, domains that contain the variable, any external standards that apply to the variable, and its range of values. The definition shall indicate if the variable is derived and the class.object.variable of derivation.
b) Applicable domains: For example, those data domains (such as vital records, cancer) that use the variable *date of birth* or *sex.* c) Applicable standards including version: include any current or proposed standards that have a specification for the variable. The standard may have a version (e.g., HL7 version 2.6) d) Acceptable values: e) Derivation: method used to calculate or "derive" a variable (e.g., infant mortality rate is a derived variable; thus "total infant deaths per 1,000 live births")
Data Type – the kind of data that is represented by the variable. a) Integer: Any positive or negative whole number, including zero. A numeric that represents a single unary value. b) Real: A numeric that has precision as significant digits; i.e. a decimal point can be part of the number. c) String: text, alphanumeric. d) Date e) Currency

Unit of Measurement: a) refers to the system of measurement: (e.g.) English or metric; (b) the specific unit, within a measurement system, at which measurements for a variable are made.
Level of Measurement: (1) nominal, categorical properties or labels. Used to describe characteristics that have no numerical values (e.g. gender, ethnicity, FIPS); (2) ordinal, objects are ordered by some nominal category irrespective of magnitude, and irrespective to the distance between ordered levels. Used for characteristics that have an underlying order among them, although the order may be arbitrary. Example: *frequently, often, sometimes, rarely, never*; (3) interval, ordering of objects is respective to a nominal category, the distance between objects respective to the nominal category, and without respect to the magnitude of the nominal category. Example: *grams, inches, days*; (4) ratio, objects are ordered respective to a nominal category, where the distance between objects is known, and each objects measurement is respective to a known zero value. Example: *Infant mortality rate*.
Unit of Analysis: the specific unit, within a measurement system, at which analysis of variables are made.
Level of Analysis: See level of measurement.
Size or Precision: Size or precision is the amount of storage the variable consumes. Size is applicable to string variables (e.g., 8 characters long), precision is for numeric data both integer and real, so if a numeric variable is integer and can be up to the value 1000, then it must at least 16 bits long, if a number is real and can have values up to 12345.566779 then it is double precision or 64 bits long.
Time Stamp of Definition: The time stamp of the definition is the date the variable definition became a GDPH standard.

1.3. Data quality requirements

1.3.1. A variable shall have one and only one name.

1.3.2. A variable shall have one and only one definition.

1.3.3. A variable shall be stored in one and only one data type.

1.3.4. A variable shall have one and only one field length.

1.3.5. A variable shall be stored in one and only one unit of measurement.

1.3.6. A variable shall be stored in one and only one level of measurement.

1.3.7. A variable shall represent or store only those values specified in its definition.

1.3.8. Data objects shall have one and only one source.

1.3.9. No duplicate sources of data objects shall exist.

1.3.10. No duplicate records shall exist in data objects.

1.3.11. All data domains, data objects and variables shall be free of data anomalies.

1.3.12. Unknown, missing and inapplicable values shall have respectively unique representations.

1.3.13. Unknown, missing and inapplicable values in all data domains shall have consistent representations as defined in 1.3.12.

1.3.14. All domains shall have data quality audits.

Data quality audits shall include:

1.3.14.1. Range checking of integer and real numbers.

1.3.14.2. Value checking of variable contents.

1.3.14.3. Pattern checking of strings and dates.

1.3.14.4. Functional and logical dependency checking.

1.3.14.5. Logical constraint checking within variables, records and objects.

1.3.14.6. Inexact (range) constraint checking.

1.3.14.7. Statistical (outlier) constraints checking.

1.3.14.8. Check for the correct treatment of unknown, missing and inapplicable values.

1.3.14.9. Data quality audits shall be ongoing.

1.3.14.10. Procedures and polices shall be in use to direct and assure data cleaning and schema restructuring meet these data quality requirements.

1.3.14.11. The process of data quality auditing, anomaly detection, data cleaning and schema restructuring shall be automated.

Data Cleaning Strategy

Purpose

Data cleaning is a collection of methods that apply established data quality and standards to data assets. Data cleaning is necessary for the creation of sets of data, that have gone through quality control and quality assurance processes, that can be used to derive business intelligence.

Scope and Methods

The scope of data quality policies and standards is divisional; the division should establish and enforce data quality policies and standards. Data cleaning methods, therefore, have divisional scope, since the application of data quality methods should occur throughout the division. Data policies and standards should be applied to historical and current data.

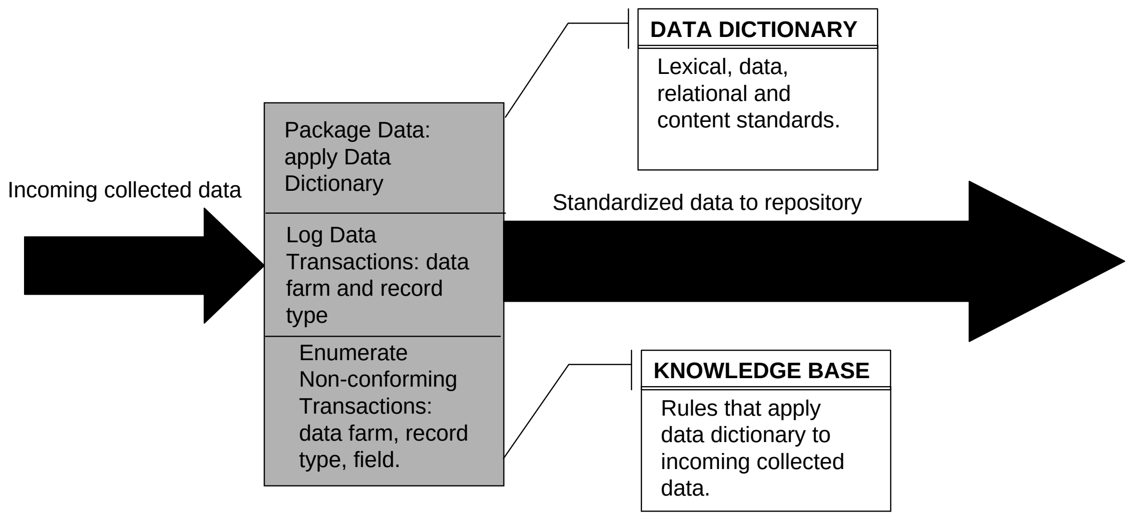
Data cleaning methods are derived from data quality policies and standards. Data policies and standards should address:

- Standardized field and variable names;
- Standardized data types, acceptable ranges, and registration of missing and unknown data;
- Standardized coding for service providers, facilities, service types, geographical properties (e.g., county);
- Standardized demographic properties for race income, education, age, sex, and marital status.

Data policies and standards should be part of the input process into an architected data repository, as shown in the following illustration on the next page. A generic device, a fusion engine, monitors input data from heterogeneous sources, applies data quality standards, and logs exceptions by source, field, and data type. Fusion engine is the industry standard term for a device that integrates data from several sources into a single architected environment; it is independent of technology and policies. The fusion engine allows data cleaning methods to be automated, and monitors the quality of input and output data.

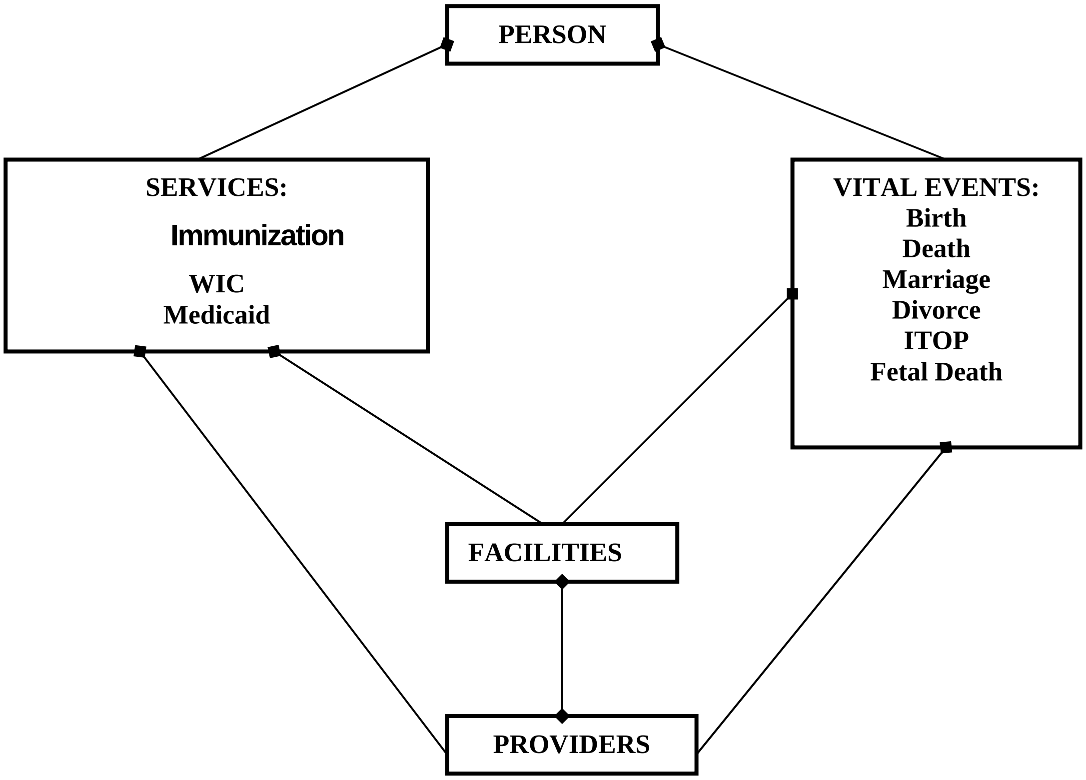
The fusion engine is a rule-based device. It has a data dictionary that contains data content, relation, and type definitions derived from data standards. The fusion engine uses a knowledge base that has methods derived from data policies, which apply the data dictionary standards.

Fusion Engine



Data Deconstruction as a Means to Focus Data Cleaning

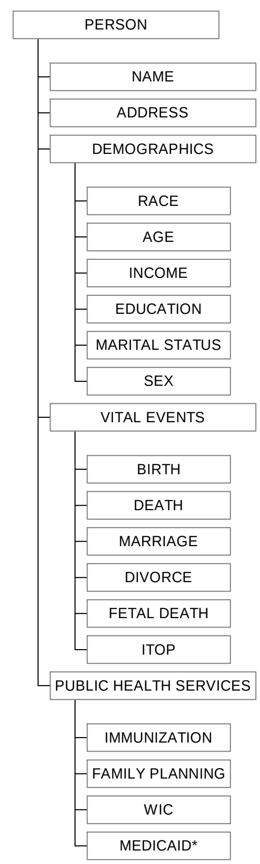
Deconstructing data files into components is useful for exposing data redundancy and data inconsistency. A simple deconstruction of public health data is shown on the next page. Persons are known to the public health system by their participation in programs or services which are monitored by public health, or through the vital records system, which is authorized to record vital events. This kind of deconstruction allows for the extraction of relevant subject matter from heterogeneous sources into a fusion engine, without having to modify or move the source data.



Since all relevant subject matter (e.g., vital records and service delivery) have persons, who in turn have demographic properties (e.g., race and sex), then demographic properties need to be consistent throughout all subject matter. This means, for example, a person's race will have one and only one race identification code, and that race code will be the race code for that person across all subject matter contained in the data warehouse. Each set of subject matter will require a data dictionary that defines the following:

- Demographic codes – race, sex, education, income, age;
- Geographic codes – county, health district, zip code;
- Facility codes – type, service level, allowed services;
- Provider codes – type, license, title, facility, allowed services.
- Service codes – type, category, ICD, CPT.

To the right is a simple data model of this deconstruction.



Appendix A Template for Variable Definition

Variable Name:
Storage Name:
Presentation of Values:
Definition:

a) Purpose:
b) Applicable domains:
c) Applicable standards including version:
d) Acceptable values:
e) Derivation:

Data Type:

a) Integer:
b) Real:
c) String:
d) Date
e) Currency

Unit of Measurement:

Level of Measurement

Unit of Analysis:

Level of Analysis:

Size or Precision:

Time Stamp of definition:

APPENDIX B

Class Interaction Diagrams

1) Data Quality Standards