Glossary of Data Warehouse Terms

It can be hard to keep up with all the acronyms and technical terms associated with the Iowa public health data warehouse. Below are some common terms that may be referenced on the project's Web site, in handouts, at meetings, or in presentations.

- *Data Warehouse*: A data warehouse is, simply put, a central place where data is stored at record or summary level for the purpose of analysis and reporting.

- *Business Intelligence*: Reporting and analysis of data stored in the data warehouse. This term (also known in industry-speak as "BI") speaks broadly of the collection of tools used to provide insights into the business. The challenge is that business "intelligence" doesn't come out-of-the-box; it must first be defined by the business.

- *DW/BI*: Data warehouse and business intelligence system (this is an industry standard term)

- *CHNA & HIP*: Community Health Needs Assessment and Health Improvement Plan. The data warehouse will provide access to health indicators that Local Public Health Agencies can use for their next CHNA & HIP.

- *Community*: A region, state, county, city, or neighborhood.

- *Cubes*: A cube is a data structure that allows the fast analysis of data. Cubes are defined using business rules and processed when new data is loaded into the data warehouse.

- *Data Sets*: A data set is used in the context of this project to describe the category of data. There may be multiple data sources for a single category. For example, births are a data set; however, birth data is obtained from two data sources - the Mainframe and Iowa Vital Records System (IVRS).

- *Dimensional Modeling*: A design technique of facts and dimensions that organizes data in the data warehouse. Facts are the numeric measurements. Dimensions describe the "who, what, when, where, why, and how" of the measurement.

- *Dimension*: This is the "by" aspect of analysis, meaning "we want to look at the data by X." Dimensions provide the list of ways data can be sliced

and diced; for example, common demographic dimensions would be age, gender and race (e.g., "by race", "by age", "by weight").

- *ETL*: ETL is an acronym for the process of extracting ("E"), transforming ("T") and loading ("L") data. This is a crucial part of the data warehouse process that pulls data out of one database and places it in another system. Data are extracted from their source system and transformed using business logic. The transformed data is then loaded into the data warehouse.

- *FACITS*: Family and Community Health Indicator Tracking System. The legacy system developed by Iowa Department of Public Health to disseminate public health data. The data included in FACITS are similar to the data that will be included in data warehouse; however, the functionality will be enhanced.

- *Indicator*: A statistical calculation, generally in the form of a number, rate, or percent, used to measure and monitor public health variables.

- *Iteration*: Iteration is essentially a mini-scope. It defines the data and functionality to be incorporated in the DW/BI in a given timeframe. For this project, it is anticipated that iteration will be 4-6 weeks.

- *Metadata*: All the information that describes the contents, structures and operations of the DW/BI system. Metadata can be split out further into the following: technical, business, and process metadata. An individual piece of metadata can belong to one or many categories.

    - *Technical metadata* defines the objects and processes that make up the data warehouse itself from a technical perspective.

    - *Business metadata* describes the contents of the data warehouse in more user-accessible terms. It tells us what we have, where it comes from, what it means, and what its relationship is to other data in the data warehouse.

    - *Process metadata* describes the results of various operations in the data warehouse.

- *Population-based*: Indicators that pertain to the entire population in a particular area. It does not include indicators based on a subset of the population involved in a select program.