

Proposal: Establish a Department of Health and Family Services' Information Quality (IQ) Assessment Protocol

This suggested protocol measures aspects of data relative to identified quality characteristics. These characteristics are those that a program analyst (knowledge worker) would find necessary to assess in performance of their job duties.

In the literature on business information quality, there are two basic categories of information quality and related measures. The first, which is the subject of this protocol, relates to *static* data and is the inherent quality of the data that populates a database. The second category, relates to dynamic or *pragmatic* data and includes how intuitive the information is in its presented format and how well it enables knowledge workers to accomplish their objectives.

This assessment protocol relates to static data and inherent information quality characteristics. Static data is a “snapshot” of a database (or other data collection) at a specific point in time. The sample analysis was performed using SAS as the analytic tool.

Static Data Quality Characteristics

- **Definition conformance.** The consistency of the meaning of the actual data values with its data definition.
- **Completeness (of values).** The characteristic of having values for the data fields.
- **Validity or business rule conformance.** A measure of the degree of conformance of data to its domain values and business rules.
- **Accuracy to surrogate source.** A measure of the degree to which data agrees with an original source of data, such as form, document, or unaltered electronic data received from out the control of the organization that is acknowledged to be an authoritative source.
- **Accuracy (to reality).** The degree to which data accurately reflects the real-world object or event being described. Accuracy is the highest degree of inherent information quality possible.
- **Precision.** The characteristic of having the right level of granularity in the data values.
- **Nonduplication.** The degree to which there is a one-to-one correlation between records and the real-world or events being represented.
- **Equivalence of redundant or distributed data.** The degree to which data in one data collection or database is semantically equivalent to data about the same object or event in another data collection or database.
- **Concurrency of redundant or distributed data.** The information float or lag time between when data is knowable (created or changed) in one database is also knowable in a redundant or distributed database.
- **Accessibility.** The characteristic of being able to access data when it is required.

Data quality measures, as in the example is below, are computed by dividing the number of possible values for any one category by the number of actual values. This example is of *static* quality measurements relating to name and address data from the **Technology Systems Interface (TSI) as measured on 4/21/2000**;

Static Data Quality Characteristic	Number of Fields Evaluated	Number Meeting Quality Measure	Quality Statistic (Percentage)
Definition conformance.	Not Assessed	Not Assessed	Not Assessed
Completeness (of values).	Last Name 8,941	8,806	98.49%
	First Name 8,941	8,753	97.89%
	Middle Name 8,941	1,355	15.15%
	Birth Date 8,941	6,379	74.45%
	Complete Name and Birth Date 8,941	1,323	14.80%
	Street Number 6,807	0	0%
	Street Name 6,807	0	0%
	City 6,807	6,807	100%
	State 6,807	6,807	100%
	Zip Code 6,807	6,769	99.44%

Validity or business rule conformance	Not Assessed	Not Assessed	Not Assessed
Accuracy to surrogate source	Address Fields 34,035	20,383	59.88%
Accuracy (to reality).	Not Assessed	Not Assessed	Not Assessed
Precision.	Not Assessed	Not Assessed	Not Assessed
Nonduplication.	Total Client Rows 8,941	Estimated Unduplicated Clients 8,699	97.29%
Equivalence of redundant or distributed data.	Not Assessed	Not Assessed	Not Assessed
Concurrency of redundant or distributed data.	Not Assessed	Not Assessed	Not Assessed
Accessibility	Not Assessed	Not Assessed	Not Assessed
Core Information Quality	Total Assessed 114,909	76,081	66.20%