The Six Dimensions of EHDI Data Quality Assessment*

This paper provides a checklist of data quality attributes (dimensions) that state EHDI programs can choose to adopt when looking to assess the quality of the data in the EHDI-IS. It is not a prescriptive list and use of the dimensions will vary depending on the requirements of individual programs.

This is related to data; other attributes such as acceptability and flexibility of the EHDI-IS are not addressed.

The six core data quality dimensions are

- Completeness
- Uniqueness
- Timeliness
- Validity
- Accuracy
- Consistency

*Content adapted from "THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT", DAMA, UK

Title	Timeliness
Definition	The degree to which data represent reality from the required point in time
Reference	The time the real world event being recorded occurred
Measure	Time difference
Scope	Any date item, record, data set or database
Unit of Measure	Time
Related dimension	Accuracy because it inevitably decays with time
Example(s)	1. Time difference between the completion of a patient's newborn hearing screening and information about this screening being sent to and recorded in the EHDI-IS. This time should be in accordance with jurisdictional policy. 2. Time difference between the completion of a patient's audiology diagnostic visit and the information about this visit being sent to and recorded in the EHDI-IS
Comments	1. Need to make the distinction between measuring timeliness of the information system and timeliness of the care and service delivery. For example, whether the EHDI program is meeting the EHDI 1-3-6 performance benchmark should not be used to measure the timeliness of the EHDI-IS 2. Timeliness and its related dimension (Accuracy) are usually affected by the way data are collected. The more steps and intermediate systems involved in data collection, the more delay the EHDI-IS will experience in receiving the needed information.

Title	Completeness
Definition	The proportion of stored data against the potential of "100% complete"
Reference	Business rules which define what "100% complete" represents
Measure	A measure of the absence of blank (null) values or the presence of non-blank values
Scope	0-100% of critical data to be measured in any data item, record, data set or database
Unit of Measure	Percentage
Related dimension	Validity and Accuracy
Example(s)	1. Completeness of the EHDI-IS at database/data set level: number of births documented in the EHDI-IS/number of births documented in the vital records system during a given year x 100% 2. Completeness of the EHDI-IS at record level: percentage of patient records that have all minimum and core data elements (as defined by the EHDI-IS functional standard) populated with non-blank values
Comments:	1. Definition of "100% completeness" depends on 1) the population under measurement 2) optionality of data elements 2. When measuring completeness at the data set level, there is often a need for a reference data set that is considered to be the authoritative source of such data and is 100% complete. For example, vital records data set.

Title	Uniqueness
Definition	Nothing will be recorded more than once based upon how that thing is identified. **It is the inverse of an assessment of the level of duplication**
Reference	Data item measured against itself or its counterpart in another data set or database
Measure	Analysis of the number of things as assessed in the "real world" compared to the number of records of things in the data set. The real world number of things could be either determined from a different and perhaps more reliable data set or a relevant external comparator
Scope	Measured against all records within a single data set
Unit of Measure	Percentage
Related dimension	Consistency
Example(s)	Percentage of duplicate records within a EHDI data set

Title	Validity
Definition	Data are valid if it conforms to the syntax (format, type, range) of its definition
Reference	Database, metadata or documentation rules as to allowable types (string, integer, floating point etc.), the format (length, number of digits etc.) and range (minimum, maximum or contained within a set of allowable values)
Measure	Comparison between the data and the metadata or documentation for the data item
Scope	All data can typically be measured for Validity. Validity applies at the data item level and record level (for combinations of valid values)
Unit of Measure	Percentage of data items deemed Valid or Invalid
Related dimension	Accuracy, Completeness, Consistency, and Uniqueness
Example(s)	1. Validity at data item level: type and severity of hearing loss should be chosen from a given list of allowable values 2. Validity at record level: for any patient, the date/time of hearing screening should be after the date/time of birth.
Comments	Data validation can occur before or after data are collected and entered into the EHDI-IS. The EHDI program should define proper business rules and processes to identify and handle invalid data items or records.

Title	Consistency
Definition	The absence of difference, when comparing two or more representations of a thing against a definition
Reference	Data item measured against itself or its counterpart in another data set or database
Measure	Analysis of pattern and/or value frequency
Scope	Assessment of things across multiple data sets and/or assessment of values or formats across records, data sets and databases
Unit of Measure	Percentage
Related dimension	Accuracy, Validity, and Uniqueness
Example(s)	1. In some state EHDI-IS, three data fields are used to document the results of a hearing screening: left ear, right ear, and overall. The overall field is designed to accommodate cases where ear specific information is not available. When information from both ears are available, this field should be an auto-computed value based on the results of the two ears. (E.g. overall = pass if and only if left=right=pass). Assessment of values across these three fields should be performed to ensure consistency. 2. The EHDI-IS received information about a child's EI enrollment from the EI system, but found out that in the EI system, this child is documented as having bilateral permanent conductive hearing loss. However, in the EHDI-IS, this child's dx evaluation result is still in-process.
Comments	Consistency assessment may not be applicable to all EHDI-IS or all EHDI data items.

Title	Accuracy
Definition	The degree to which data correctly describes the "real world" object or event being described.
Reference	Ideally the "real world" truth is established through primary research. However, as this is often not practical, it is common to use 3rd party reference data from sources which are deemed trustworthy and of the same chronology.
Measure	The degree to which the data mirrors the characteristics of the real world object or objects it represents.
Scope	Any "real world" object or objects that may be characterized or described by data, held as data item, record, data set or database.
Unit of Measure	The percentage of data entries that pass the data accuracy rules.
Related dimension	Validity, Uniqueness, Consistency
Example(s)	A hospital nurse was entering information to the web-based system on May 1, 2014, she recently emigrated from Europe and was not very family with the US system. As a result, she entered all date of birth information in the format of DD/MM/YYYY instead of the required format of MM/DD/YYYY. The data passed the system validation check as the values are all within the legal range. However they are not accurate, the system accepted all children entered as born on Jan. 5, 2014.
Comments	While the other 5 data quality dimensions can be assessed by analyzing the data itself, assessing accuracy of data can only be achieved by either • Assessing the data against the actual thing it represents, e.g. visit the hospital and observe how birth and screening data are collected and entered into the system • Assessing the data against an authoritative reference data set, for example, compare data in the EHDI-IS with the medical records at the audiology clinic.

HOW TO USE DATA QUALITY DIMENSIONS

A typical Data Quality Assessment approach might be:

1. Identify which data items need to be assessed for data quality, typically this will be data items deemed as critical to business operations and associated management reporting. In the context of EHDI, minimum and core data items (as defined in the functional standards) have higher priority for quality assessment compared to extended ones.
2. Assess which data quality dimensions to use and their associated weighting. Among the 6 dimensions, completeness and validity usually are easy to assess, followed by timeliness and uniqueness. Accuracy and consistency are the most difficult to assess.
3. For each data quality dimension, define values or ranges representing good and bad quality data. Please note, that as a data set may support multiple requirements, a number of different data quality assessments may need to be performed
4. Apply the assessment criteria to the data items
5. Review the results and determine if data quality is acceptable or not
6. Where appropriate take corrective actions (e.g. clean the data and improve data handling processes to prevent future recurrences)
7. Repeat the above on a periodic basis to monitor trends in Data Quality

The outputs of different data quality checks may be required in order to determine how well the data support a particular business need. Data quality checks will not provide an effective assessment of fitness for purpose if a particular business need is not adequately reflected in the data quality rules. Similarly, when undertaking repeated data quality assessments, you should check to determine if the business data requirements have changed since the last assessment.

*Content adapted from "THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT", DAMA, UK