

# Appendix B

## Data Quality Dimensions

---

### Purpose

Dimensions of data quality are fundamental to understanding how to improve data. This appendix summarizes, in chronological order of publication, three foundational definitions of data quality dimensions: those of Richard Wang and Diane Strong, Thomas Redman, and Larry English. These provide context for the choices in the DQAF. In the DQAF, I have not proposed new dimensions of data quality. On the contrary, I draw a subset and have narrowed their scope to define objective measurements that can be taken from within a dataset.

---

### Richard Wang's and Diane Strong's Data Quality Framework, 1996

In the article, “Beyond Accuracy: What Data Quality Means to Data Consumers,” Wang and Strong present results of a survey conducted to understand data quality dimensions from the point of view of people using data. Their starting assumption is that “data consumers have a much broader data quality conceptualization than IS professionals realize” (p. 5). In summarizing previous work on data quality dimensions, they point out the limits of both an intuitive and the theoretical approach to data quality: Both focus on development characteristics rather than use characteristics of data (p. 7).

Wang and Strong define *data quality* as “data that are fit for use by data consumers,” and they define a *data quality dimension* as “a set of data quality attributes that represent a single aspect or construct of data quality” (p. 6). To establish their dimensions, they first collected a set of 118 attributes of data identified by consumers themselves. Next they asked survey respondents to categorize the characteristics and to rate their importance in relation to data use. Wang and Strong performed factor analysis on the results and re-surveyed to understand the association of dimensions with categories of data quality. The result is a conceptual framework of 15 data quality dimensions related to four general categories of data quality: intrinsic, contextual, representational, and accessibility data quality.

- Intrinsic DQ denotes that data have quality in their own right; understood largely as the extent to which data values are in conformance with the actual or true values. Intrinsically good data is accurate, correct, and objective, and comes from a reputable source. Dimensions include: accuracy objectivity, believability, and reputation.
- Contextual DQ points to the requirement that data quality must be considered within the context of the task at hand, understood largely as the extent to which data are applicable (pertinent) to the task of the data user. The focus of contextual DQ is the data consumer's task, not the context of representation itself. For example, contextually appropriate data must be relevant to the consumer, in terms of timeliness and completeness. Dimensions include: value-added, relevancy, timeliness, completeness, and appropriate amount of data
- Representational DQ indicates that the system must present data in such a way that it is easy to understand (represented concisely and consistently) so that the consumer is able to interpret