the data; understood as the extent to which data is presented in an intelligible and clear manner. Dimensions include: interpretability, ease of understanding, representational consistency, and concise representation.

- Accessibility DQ emphasizes the importance of the role of systems; understood as the extent to which data is available to or obtainable by the data consumer. The system must also be secure. Dimensions include: accessibility and access security.

The Wang-Strong hierarchy is a very useful classification. The categories of data quality, in particular, highlight facets of data collection, storage, and use that have a direct impact on data consumers' perceptions of quality. In order for a data consumer to have a perception of intrinsic qualities of data, he or she must understand what the data is intended to represent and how the data effects that representation. If the data is not aligned with a consumer's assumptions about these things, the consumer will perceive it as inaccurate or unbelievable. Representational DQ emphasizes the role played by the specific presentation of data within a system.[1] Data can be perceived as incorrect if the consumer does not understand the conventions through which a system presents it. Contextual DQ points to the idea that the quality of data is defined to a large extent by the intended uses of data. Accessibility DQ points to another aspect of systems design. If a consumer cannot get to the data, then he or she cannot judge its quality by any other criteria.

Thomas Redman's Dimensions of Data Quality, 1996

In *Data Quality for the Information Age*, Tom Redman approaches data quality dimensions from the perspective of data modeling. Within this approach, a data item is defined in an abstract way, as a representable triple: a value, from the domain of an attribute, within an entity. This abstraction is a useful reminder of data's constructedness. Dimensions of quality can thus be associated with the component pieces of data (i.e., with the data model as well as with data values). A further dimension is data representation, which is defined as a set of rules for recording data items (p. 230). Redman identifies 27 distinct dimensions within these three general categories (data model, data values, data representation).

As noted, this approach to quality dimensions is rooted in an understanding of data structure. The first set of dimensions pertains to the conceptual view or data model. While not all data consumers are familiar with data models, models are critical to data use, since they provide a degree of context and meaning to any individual data item. Redman presents 15 characteristics of an ideal view or model of data and boils these down to six dimensions: content, level of detail, composition, consistency, and reaction to change (246–247). These characteristics are interrelated and reflect the choices that sometimes need to be made when developing a model—what to include, what to leave out, and the reasons for doing so.

The content dimensions include the relevance of data, the ability to obtain the values, and the clarity of definition. Relevance of data can be directly connected to its intended and potential uses.

[1] I have used the word "representational" in a sense different from both Strong and Wang's category of "Representational DQ" and Redman's (1999) "Data Representation" category. I have used the word "presentational" to refer to the set of characteristics that Wang and Strong categorize as "representational." "Presentational" refers to how the data itself is presented. I will use the word "representational" to refer to how data functions semioticly to represent aspects of the "real" world.

The ability to obtain data can be seen as a question of completeness of data to populate required attributes or as the appropriateness of data available for particular uses. Redman points out, for example, obstacles to obtaining data, including costs, privacy, and legal considerations, and he recognizes that many organizations resort to the use of surrogates to meet data requirements. His third consideration for content is what I have been referring to as metadata: clarity of definitions for all components of the model (i.e., entities, attributes, and their domains), as well as sources and rules for data.

The dimensions of scope include two potentially conflicting needs: for the model to be comprehensive and for it to contain only essential entities and attributes. Level of detail includes attribute granularity (the number and coverage of attributes used to represent a single concept) and precision of attribute domains (the level of detail in the measurement or classification scheme that defines the domain). Composition includes characteristics of naturalness (the idea that each attribute should have a simple counterpart in the real world and that each attribute should bear on a single fact about the entity), identify-ability (each entity should be distinguishable from every other entity), homogeneity, and minimum necessary redundancy (model normalization). Model consistency refers to both the semantic consistency of the components of the model and the structure consistency of attributes across entity types. The final dimensions of the model include robustness (its ability to accommodate changes without having to change basic structures) and flexibility (the capacity to change to accommodate new demands).

Dimensions related to data values comprise a set of four: accuracy, completeness, currency, and consistency. Redman's definition of accuracy is formulated in mathematical terms: The accuracy of a datum $<e, a, v>$ refers to the nearness of the value v to some value v' in the attribute domain, which is considered as the correct one for entity e and attribute a. If the value v = value v', the datum is said to be correct (255). This equation explains the concept of accuracy, but, as Redman points out, the real challenge lies in knowing the correct value.

The next dimension of data values, completeness, refers to the degree to which values are present in a dataset. Values can be absent for different reasons, some expected, some not. The challenge with completeness comes in knowing whether or not the values are expected to be present.

Currency, the third dimension of data values, refers to time-related changes in data. Redman describes data being *up-to-date* and *current* as special cases of correctness and accuracy. Data can be correct when it is loaded to a database, but incorrect (out-of-date) if there is a change in the status of the entity being represented between the time the data is loaded and when it is used. For some data, the concept of currency (the degree to which data is up to date) is critical. For other data (of low volatility or representing "permanent" characteristics), the concept of currency is not critical.

Consistency is Redman's final dimension related to data values. He notes that the first characteristic of consistency is that two things being compared do not conflict. Consistency problems appear when datasets overlap and represent the same or similar concepts in a different manner, or when their specific content does not correspond.

Redman's classification includes eight dimensions related to data representation. Seven pertain to data formats—appropriateness and interpretability (both related to the ability of a data consumer to understand and use data), portability, format precision, format flexibility, ability to represent null values, and efficient use of storage. The eighth, representational consistency, pertains to physical instances of data being in accord with their formats. Redman concludes his description of dimensions with a recognition that consistency of entities, values, and representation can be understood in terms of constraints. Different types of consistency are subject to different kinds of constraints.