

What Are Facts (Data) and  
What Is Information?

March 20, 1974

Number 12

An old friend and CC<sup>®</sup> subscriber recently wrote me to protest a grammatical error in one of our advertisements.<sup>1</sup> It included the question: "Is this data the most recent available on the subject?" Answering his letter required an exercise I have little time to enjoy these days: word-study. This essay is part of the result.

"This data is" is, strictly speaking I suppose, grammatically incorrect. *Data* is the plural of *datum*. Most people, including myself, who use the expression occasionally, know that fact, but go along with *Webster's Unabridged* (or *Webster's* goes along with us) in acknowledging "data pl but often sing in constr."<sup>2</sup> *The Oxford English Dictionary* does not concede any such usage; it does not, as a matter of fact, even hint that it may exist.<sup>3</sup> Fowler confirms that "data is often so treated in U.S.; in Britain this is still considered a solecism, though it may occasionally appear."<sup>4</sup> The solecistic use Fowler then quotes happens to be from some British *scientific* report.

The quarrel between *this data* and *these data* suggested to me, however, one of deeper significance, and of more interest to me as an erstwhile linguist and sometime information scientist. I have found that preference for use of *data* as a singular is usually strongest when the writer or speaker really means *information*. Which brings us to interesting philosophical and psychological

problems about the difference between data and information.

At a conference not long ago I was asked to distinguish between the two. To my surprise some people find it difficult to make the distinction. I have found the most effective way of explaining it, as I did at the conference, in exploring the history of the word *information*.

There's no trouble with the history of the word *datum* or *data*; it meant and still means *something given*, that is, *fact* or *facts*. How does history of the word *information* help us distinguish between *information* and *facts*?

*Information* is basically the act or process of informing, that is, of giving something *form* or identifiable and comprehensible *shape*. Very basically, the thing which gives form or shape to anything is *light*. (At this point, the etymology begins to include connotations of mythology and religion). The word *form*, according to *Webster's*,<sup>5</sup> comes to us from a Latin modification of the Greek *morphe*. And both of those words are related thus to a root which gave rise also to the Greek *marmarein* (to shine, to sparkle), and gave rise also to our words *morn* and *morning*. Thus, there may be more than we generally appreciate in one of the names of the greatest of the fallen angels. It is interesting to remember that the sin of Adam and Eve in Eden

was to have eaten of the fruit of the tree of knowledge, to which they were tempted by Lucifer, the Light-Bringer, also known as Son of the Morning.

Thus the word *inform* meant basically to endow with some quality, originally shape, that makes something identifiable and comprehensible in the mind. The meaning was enlarged to include other qualities or principles that act to provide, or merely themselves act as, distinguishing characteristics. Eventually the word *inform* came to be used with reference to that strange and most shadowed of all indescribable things, the human mind, and the verb *inform* took on a meaning of "to mold, form, train the mind or character by imparting learning or instruction." At about the same time the word describing this act of shaping up the mind, that is, *information* came to be used of the thing that (we hope) accomplishes it: data, facts, knowledge of things and events. (*Shape Up!* is an example of unconsciously erudite and allusive slang.)

With this history of the word in mind, one can easily accomplish what some find so difficult—that is, distinguishing between data or facts and

what we call information. The former has no 'shape' that is relevant to a particular viewpoint. It must be given relevance, arrangement, coherence, usefulness within a definite framework of meaning, intent, or interest. Then data or facts become information, they do inform the mind, or going back to the basic concepts, *cast light upon* a subject.

It's frequently asked (usually with sarcasm by those who distrust and haven't learned to use information technology) whether there is or can be any such thing as *information retrieval*. It is, they contend, only extraction of data or selection of facts. They are, of course, wrong. In information retrieval a great deal of 'informing' goes into the composition and structure of the data base. But, as in the case of our ASCA<sup>®</sup> service, even more goes into the construction of the user's profile. It is the profile that sets the framework within which the selected bibliographical data become information. In this very important sense, it is the user, and the light he brings with him, that determines whether any retrieval system spits out mere facts or supplies information.

1. Whalley, W.B. Personal Communication, 18 January 1974.
2. *Webster's Third New International Dictionary of the English Language Unabridged*. Springfield, Mass.: Merriam, 1968. p. 577.
3. *The Oxford English Dictionary* . . . 13 vols. Oxford: Clarendon Press, 1933, repr. 1961. Vol. 3, p. 42-43.
4. Fowler, H.W. *A Dictionary of Modern English Usage*. 2nd ed. New York: Oxford Univ. Press, 1965, p. 119.
5. *Webster's* p. 892.