Data, Information & Knowledge

I find that I often confuse the definition of data, information and knowledge. I have, of course, seen a lot of presentations that blithely assume everyone knows what Acroff meant when he first used the terms, however I have also come across quite a few experienced people whose definitions are clearly quite different from my own. I figured it would be worthwhile to attempt to succinctly define them, so at least I have a definition of what they mean for myself.

Data: the simplest of the levels, this is just a set of values. Data has no significance beyond its existence (in and of itself) and, in isolation, has no meaning. All data can be represented as a sequence of yes/no choices, hence the volume of data you have is measured in the number of bits required to represent it.

Information: a set of data within its context. Information affects the state of a dynamic system, such as our understanding of a situation. It resolves uncertainty, so the measure of how much information is present is related to the change in a probability estimate that it causes. Data that is uninterrupted, untrusted or unsurprising contains no information.

Knowledge: the processing, organisation and structuring of information. It is the synthesis of information from multiple sources that leads to understanding. Admittedly this is an elusive concept which is difficult to define and certainly hard to agree on.

Wisdom: an extrapolative appreciation of the underlying situation. In other words wisdom is the ability to predict future knowledge. This idea is even more slippery than knowledge.

The precise definitions of data and information come from Claude Shannon, who in 1948 provided a clear mathematical basis for the whole topic (his ideas underlie many of today's key technologies). The rather loose descriptions here are my attempt to summarise that work, anyone working with "information" should at least be aware of the original paper (available at **cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf**).

The other two terms are much less clearly defined. There are some computer based systems that I would consider to handle knowledge, for example IBM's "Watson", which competed and won a special version of the TV game show "Jeopardy" and the "Cyc Project" that has been attempting to encode "common sense" since 1984. But, my experience is that the vast majority of claims to handle "knowledge" within computer based systems usually have more to do with sales hype than reality. My own view is that anyone that claims their software handles "wisdom" is clearly just focused on parting you from your money, avoid them.