

Why Semantic Markup Sucks

Don't get me wrong, semantic markup—the process of putting tags on data, text, or images to indicate their meaning or provide some other fact or description about them (metadata)—has immense advantages. One of the biggest is that it enables us to access, analyze, recombine and reuse information independent of the applications and formats used to create them.

But for many of the folks doing the marking up it's like paying into a bank account you never see, and with the tedium and inefficiency of queuing at a teller window. Having to make all those deposits distracts you from your real job of making money—or, to step out of our analogy, the job of creating a great dataset or document.

Applying semantic markup is a distraction from real work because people are frequently unsure of what the correct tag is within the context of their task and organization. In life sciences industries, for example, most tools for annotating research data were designed with no sympathy for the fact that users live in an uncertain world. A project team knows how to annotate the data in a way that makes sense to them, but isn't sure if another team doing tangential research half a globe away is using different terminology, or whose approach will ultimately dominate. Or maybe there's a codified, shared enterprise vocabulary/ontology, but plucking the right tag from it involves searching or browsing through a mass of irrelevant possibilities.

As a result, you get highly paid scientists spending too much time thinking about or trying to locate the right tag—the part of their task that is essentially clerical. It's a constant rub of friction that slows down the workflow. In my experience at pharmaceutical companies, it can also be a psychological irritant that wears down motivation and morale.

One of our primary goals is to remove the uncertainties and friction of semantic markup while keeping all of its power. We want research teams to be able to focus on high-value thinking: what's interesting in the data and how to structure the dataset for the meaningful comparisons and other types of analyses.

We've achieved this goal by combining template-based webforms at the front end with a semantic network at the backend. Webforms put the right tags for the project at user fingertips. The semantic network invisibly manages the relationships between project-specific and enterprise vocabulary/ontology. It continually updates these links as projects evolve and enterprise knowledge expands.

In my next post I'll talk about how to let project teams create knowledge on-the-fly by inventing tags as needed—without blowing up the structure of the enterprise knowledge store.