Meta Skepticism

Understanding the limits of metadata and the compromises we make in this uphill battle

This Stuff Costs Real Money

- If not machine generated, it takes serious time by well-trained individuals to create metadata.

- Key driver for industry groups to set standards and share metadata

- Most often, schemes are scaled back or abandoned because of the cost.

Cory Doctorow – Metacrap

- (A riff on and old theme…)

- https://people.well.com/user/doctorow/metacrap.htm

- A basic dope slap about metadata for those who are too enamored with the topic.

- Written in 2001 and just as relevant today

People Lie

- Metadata exists in a competitive world. Suppliers compete to sell their goods, cranks compete to convey their crackpot theories (mea culpa), artists compete for audience. Attention-spans and wallets may not be zero-sum, but they're damned close.

- Meta-utopia is a world of reliable metadata. When poisoning the well confers benefits to the poisoners, the meta-waters get awfully toxic in short order.

People are Lazy

- We are engaged in the incredibly serious business of creating information. Here in the Info-Ivory-Tower, we *pretend to* understand the importance of creating and maintaining excellent metadata for our information.

- But info-civilians *(projection)* are remarkably cavalier about their information.

- Your clueless aunt sends you email with no subject line, half the pages on a Web site are called "Please title this page" and your boss stores all of his files on his desktop with helpful titles like "UNTITLED.DOC."

- Discipline is not our strong suit.

People are Stupid

- Even when there's a positive benefit to creating good metadata, people steadfastly refuse to exercise care and diligence in their metadata creation.
- Take eBay: every seller there has a damned good reason for double-checking their listings for typos and misspellings. eBay uses a host of tools to include misspelled listings in correctly-spelled searches, yet some listings garner fewer bids and lower sale-prices. Try shopping for a kyack, or a kayack,
- We want these people making metadata?

Mission: Impossible -- Know Thyself

- In meta-utopia, everyone engaged in the heady business of describing stuff carefully weighs the stuff in the balance and accurately divines the stuff's properties, noting those results.

- People are lousy observers of their own behaviors. Entire religions are formed with the goal of helping people understand themselves better; therapists rake in billions working for this very end.

- What blind spots or unacknowledged privileges do we bring to the table?

Schemas Aren't Neutral

- In meta-utopia, the lab-coated guardians of epistemology sit down and rationally map out a hierarchy of ideas. This presumes that there is a "correct" way of categorizing ideas, and that reasonable people, given enough time and incentive, can agree on the proper means for building a hierarchy.
- Nothing could be farther from the truth. Any hierarchy of ideas necessarily implies the importance of some axes over others.
- It's wishful thinking to believe that a group of people competing to advance their agendas will be universally pleased with any hierarchy of knowledge

Metrics Influence Results

- Agreeing to a common yardstick for measuring the important stuff in any domain necessarily privileges the items that score high on that metric, regardless of those items' overall suitability.

- Every player in a metadata standards body wants to emphasize their high-scoring axes.

- "No Child Left Behind" emphasized measurable things and diminished the immeasurable.

- The best that we can hope for is a *detente* in which everyone is equally miserable.

There's more than one way to describe something

- "No, I'm not watching cartoons! It's *cultural anthropology*."

- "This isn't smut, it's *art*."

- "It's not history, it's a *crime against humanity* that has been whitewashed by the dominant culture."

- Reasonable people can disagree forever on how to describe something.

- Arguably, your Self is the collection of associations and descriptors you ascribe to ideas. Requiring everyone to use the same vocabulary to describe their material denudes the cognitive landscape, enforces homogeneity in ideas.

- It's hard to account for cultural and demographic differences in perception.

Reliable metadata

- Metadata can be quite useful, if taken with a sufficiently large pinch of salt. Metadata is often a good means of making rough assumptions about the information that floats through the Internet.

- Certain kinds of implicit metadata is useful. Google uses the number of links pointing at a page (and the number of links pointing at each linker), to derive statistics about the number of Web-authors who believe that that page is important. Hence making extremely reliable guesses about how reputable the information on that page is.

- This sort of observational metadata is far more reliable than the stuff that human beings create for the purposes of having their documents found. It cuts through the marketing BS, the self-delusion, and the vocabulary collisions.

The Most Important Piece of Missing Metadata

- Who are you? Why are you here?

- Taken more broadly, this kind of metadata can be thought of as a pedigree: who thinks that this document is valuable? How closely correlated have one person's value judgments been with others' in times gone by?

- Kindred spirits.

- This kind of implicit endorsement of information is a far better candidate for an information-retrieval panacea than all the world's schema combined.

Who's Doing It?

- Let's examine some metadata champions' use of metadata…

- https://www.loc.gov/standards/mets/

- https://ddialliance.org/

- https://www.loc.gov/