

Shakespeare wasn't a semantic web guy, Jonathan Rees

Google Sidewiki

What is semantic web? Like google and tagging on steroids. What makes it different from a taxonomy? The ability to associate and build relationships between data. Is it about consequences? Every book has an author is a consequence of being a book. But this isn't what we are supposed to be talking about.

Science Commons: a creative commons project.

Theme: reuse. How do you reuse someone else's work? CC license allows reuse. (How do we define trivial changes in reuse?) Science is about reuse, building on ideas, standing on shoulders of giants.

CC-Zero: statement someone can make that a data set is public domain, like US Law says it is. CC-Zero is not a license, but a waiver. Data is not copyrightable(?). What if some is data, but mixed with creative work? Setting boundaries.

Barren room, people getting up to leave.

Bioinformatics Case Study: open source, open access success story, its small, a community. Journal publishes lists of databases (Nucleic Acids Research, 1,200 databases). A wealth of data. Bioinformatic software library, tools being published everyday. (See wiki for links). Tools refer to databases and databases refer to each other.

Data Reference Risks: you make an investment in someone else's stuff. Stability. Format. Reliability. Institutional failure (TAIR, an NSF resource, NSF is not set up for long-term). Commercial capture. Stability of name. Not a semantic web specific problem, but a URI problem.

Identity Muddle: Paper's Available through many different routes: Different URLs for same content. Comments on one page don't show up on another. Choice of URL for page makes a difference, URL is not always a good key. Also a semantic web issue. Computer Science Solution: add a level of indirection. Use a different ID and an indirection table to collect different URIs to same place, but different resources use different identifiers to key into resources.

Comment: Women who don't change their names after marriage for publication reasons.

Shared Names Project: attempt to get everyone to use the same keys for indirection. (sharedname.org), provide indirection URIs, stability, annotations, errata, mirrors, related resources, protocols. Consumer cooperative, not publisher originated. **Who's responsibility is it to update the reference, publisher consumer? What will prevent a resource from relocating, and a new URI being made for the new reference, but the old reference is now dead.**

Scorecard: preserving content is not the same as preserving name stability.

These organizations understand the problem, because they have 200 years of dealing with it. The web is too young for enough people to be burned by it. **Everyday I post links, many of which will go dead within the year.**

Biodiversity Heritage Library

Web Consortium

DOI and Crossref.

Libraries sit between publishers and consumers to preserve access to data.

Problem: URIs are not tiny. This would be more effective as a tiny url.

Issue of replication. Popular resources get replicated.

Library analogies. Card Catalogue. Books get moved around in space, but catalogue remains the same.

Took the whole session to explain the issue, it's complex/technological and doesn't hold people's interest.

Publishers don't care so much. It's not their job.

There are too many hurdles for this to work. Managing the chaos is insurmountable. Is this a fatalist position? That we simply have to accept it? I always upload a copy of images posted, rather than hotlink because I don't want my content subject to someone else's hosting. Embedded video goes dead all the time, but I don't want to upload gigs of video each year.

Reminds me of Ted Nelson's idea of the web, where there's only one copy of anything stored online and it's manageable.