

Scholarly HTML: Simple, rational modern citations using links

I dashed off a quick post last week about [Scholarly HTML](#) which had a bunch of ideas for a new kind of online journal.

I reckon I was pretty spot on with my timing. Hours after I posted it, I saw the announcement from Elsevier about the winners of their [Article 2.0 competition](#). The [winner](#) has implemented some of the things I was calling for, which is great. I considered having our group enter this competition, but apart from limited time, the biggest issue I had was that the competition was about what you could do with Elsevier's XML, not what you could do about the elephant in the scholarly publishing process which is how the XML gets created. Me, I'd like to see a Journal 2.0 competition that looks at the whole process including the business model and the authoring and review.

The ideas didn't really spell out who does what and what the workflows would look like. I will go through some of the detail now in a few more posts. In this post I'd like to look at one dimension which was not addressed by the winner – how an author should cite their references. The winning entry does [look at this after the fact](#):

Enhancing references with DOIs - Digital Object Identifiers can be added to references in the document, by the document author, or by third parties. This allows references to be made more useful for users. This is a specific example of the reference level assertions described above - DOIs are added by selecting the Add Assertion link next to an item of content, selecting a type of DOI, and entering the DOI.

DOIs are one way to identify articles, but why should this be done by the readers? If they're good identifiers why not let authors cite using DOIs in the first place?

I have one very firm idea about this I'd like to try out. I reckon it should be possible to add citations using links. After all, I think, the main point of a citation is to uniquely identify a reference. I asked Twitter:

Today I'm going to think about citations - what do we need them to do? Identify the reference, right? What else? #research

So far two replies.

[petermurrayrust](#) @ptsefton would be nice if we could find out when a publication actually appears in print

OK – but that's something that would have to happen after we work out a good ID and how to express it is a link. I'll set that aside.

[dorotheasalo](#) @ptsefton: be suckable into bib managers; properly credit authors for impact-factor purposes; allow readers to guess @ article's usefulness

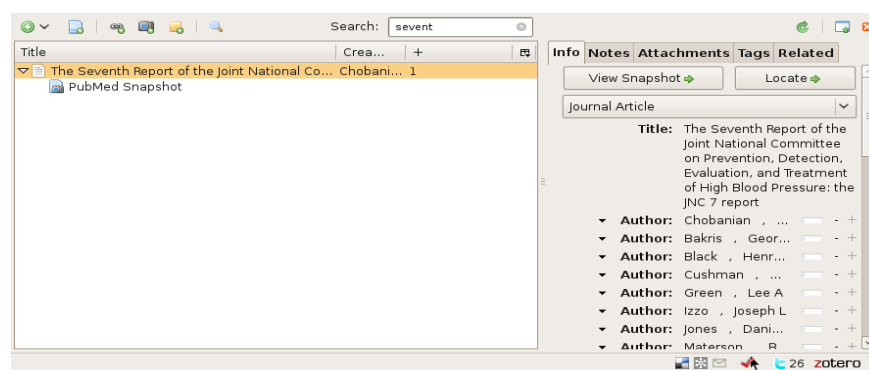
These are all must haves, and they are all about metadata associated with a reference. Given a good ID a machine should be able to do better than a person at managing and presenting this metadata. The way we manage metadata at the moment though, everybody is responsible for maintaining their own metadata in their paper. Some do it efficiently, via shared bibliographies and services like CiteULike, some use tools like Zotero which are edging towards sharing, while apparently some still do it manually. Just this week we have discovered someone at USQ still wanting to teach people how to dot the Ts and cross the Is to make APA references – get the computer to do it, I say.

I can see three main ways to cite by link that I think might work. In order of desirability:

1. If the article is in Scholarly HTML then it will contain suckdownable metadata and be addressable to the paragraph level, so it will be a matter of pointing to the permalink for the paragraph or section heading you want to cite.
2. Reference by a known kind of ID, like PubMed [PMID](#) or a [DOI](#), using one or more sites that journals flag as reliable sources of metadata. For example linking to a pubmed ID like [this](#).
3. Linking to one or more (depends on the discipline) online citation manager service such as CiteULike, Connotea or Zotero. Here's a page from my [Zotero library](#) which I scraped from WordCat.

Note that I don't include linking to any-old repository, journal or website, or Google Scholar as the results would be too unpredictable and the metadata is likely to be crap.

Journal-side systems could scrape the metadata from a page and provide the information Dorothea recommends, like suckdownable metadata, pop up abstracts and maybe author identity. Here's what Zotero makes of it – and Zotero's suckdowners are all available so could be used in other code.



How would this work for editorial and review processes? Upon submission, the computer will fetch down the references and format them in a way that is most convenient for editorial use – this could be whatever the editors and reviewers are used to or some other thing optimized for clarity. They would be able to give feedback on the metadata for a citation as necessary just as they do now.

Another idea would be for the journal itself to keep a reference database, and to ask authors to use it by simply linking to it – thereby making sure that there is a consistent way for the same citation to be referenced across articles and issues. This should vastly improve citation metrics. For example they could put up a Zotero library for the journal with everything cited so far in a canonical format, and require that authors reference that by URL if the thing they wish to cite is in there. If not, then authors can use one of the methods above, but as part of the journal workflow - and this could be automated - the editors will create a Zotero record for the new citation and update the submitted HTML.

And what about page numbers? We could ask authors to append a parameter to their URL but that would be messy and would need tool support. So how about just doing it like this, inline as the anchor text for a link: [p20](#). Remember this is to keep authoring simple it's not how the article will appear on the web later, where the reader may even be able to choose the reference format they prefer. Remember, though, with Scholarly HTML articles everything is on one page with individually addressable paragraphs.

As with the previous post on this I am going into areas where I am far from well informed. Are there already journals with simple, rational, modern citation requirements such as requiring only DOIs?