

Discoveries and Anti-Discoveries on the Web of Argument and Data

Tim Clark^{1,2,3}, Carole Goble³ and Paolo Ciccarese^{1,2}

1 - Harvard Medical School; 2 - Massachusetts General Hospital; 3 - School of Computer Science, University of Manchester
tim_clark@harvard.edu

The scientific communications ecosystem is a web of both argument and data. Published scientific data are meant to, but too often do not, fully back the assertions in the published argument. There are many examples of this phenomenon, which is nicely exemplified by recent controversy over the so-called STAP (Stimulus-Triggered Acquisition of Pluripotency) methods.

In this dispute, Obakata et al. published an apparent breakthrough in regenerative biology [1] showing a method – STAP – for inducing pluripotency in somatic cells, by the simple method of bathing them in an acidic solution. Other labs trying to reproduce her results failed.

Later, post-publication peer review hosted on a number of web forums indicated problems with Obakata's data, which seemed difficult to explain as the result of mishap. An investigation determined that she had “improperly altered images of DNA fragments” in her publication. However, highly useful discussion on Obakata's papers was spread across multiple websites blogs and forums, is tied to those sites, and has undetermined longevity.

This kind of problem does not seem to be confined to highly scrutinized, breakthrough papers. Begley and Ellis reported in 2012 [2] that only 11% of academic research results they reviewed were reproducible.

We believe a useful approach to discovering and mapping “non-discoveries” would be to annotate articles in review, using the W3C Open Annotation (OA) [3] model. Existing software such as *Domeo*, *Utopia Documents* and *Annotopia Server* [4,5,6], supports this approach. The method can be made more powerful using the Micropublication vocabulary (<http://purl.org/mp>) [7], which creates a formal argumentation network on the web. In our lab such markup routinely takes 15-20 minutes per article, on the Results section of primary research articles.

Automated techniques for detecting citation types or “polarity”, e.g. [8], could further simplify such a process.

With direct citation of data and research resources [9,10] the future pre- and post- publication peer-review process could create a data- and methods- backed biomedical argumentation graph on the web, supporting much better reproducibility and translation to therapeutic development.

Progress on *Annotopia Server* should soon allow diversely developed web clients to create and exchange such annotation freely, in OA format, in a distributed network.

References

1. Obokata H, Wakayama T, Sasai Y, Kojima K, Vacanti MP, et al. (2014) Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature* 505: 641-647.
2. Begley CG, Ellis LM (2012) Drug development: Raise standards for preclinical cancer research. *Nature* 483: 531-533.
3. Sanderson R, et al. (2012) Open Annotation Core Data Model Community Draft, 09 May 2012. W3C.
4. Ciccarese P, Ocana M, Clark T (2012) Open Semantic Annotation of Scientific Publications with DOME. *Journal of Biomedical Semantics* 3: S1.
5. Attwood TK, Kell DB, McDermott P, Marsh J, Pettifer SR, et al. (2010) Utopia documents: linking scholarly literature with research data. *Bioinformatics* 26: i568-i574.
6. Clark T (2014) The Annotopia Open Annotation Services Platform. Annotopia Workshop. European Bioinformatics Institute, Hinxton, UK, May 29, 2014. <http://slidesha.re/1ljFumH>
7. Clark T, Ciccarese P, Goble C (2013) Micropublications: a Semantic Model for Claims, Evidence, Arguments and Annotations in Biomedical Communications. *Journal of Biomedical Semantics*. <http://arxiv.org/abs/1305.3506>
8. Angrosh MA, Cranefield S, Stanger N (2011) Contextual Information Retrieval in Research Articles: Semantic Publishing Tools for the Research Community. *Semantic Web Journal*.
9. CODATA (2013) Out of cite, out of mind: The Current State of Practice, Policy and Technology for Data Citation. *Data Science* 12: 1-75. <http://dx.doi.org/10.2481/dsj.OSOM13-043>
10. Vasilevsky NA, Brush MH, Paddock H, Ponting L, Tripathy SJ, et al. (2013) On the reproducibility of science: unique identification of research resources in the biomedical literature. *PeerJ* 1: e148. <http://dx.doi.org/10.7717/peerj.148>