

# Breaking the Paper Barrier and Publishing At the Speed of Thought

Katie Chapman and Chris Chapman

Pentandra Research Solutions, Logan, Utah, USA  
{katie, chris}@pentandra.com  
<http://pentandra.com>

**Abstract.** We are starting to see Semantic Web technologies being used in our research papers and data. The next revolution will be publishing knowledge at the speed of thought.

**Keywords:** Semantic Publication • Linked Data • Skywriting

In September 2007, Wired Magazine included an article by Thomas Goetz entitled “Freeing the Dark Data of Failed Scientific Experiments” [1], in which Goetz discussed the incredible value of *dark data*—data that is filed away, never to see the light of day. We believe that researchers are already well aware of the understandings gained through research projects that led in unexpected directions, revealed unwanted results, or simply left the researcher feeling painted into a dead-ended corner. Such failed research efforts are powerful opportunities to learn and often lead to important insights. Yet most of the time these insights are not shared. According to Goetz, the reasons dark data remains dark include storage costs, technology barriers, financial constraints, or career incentives that encourage dissociation from negative and inconclusive research. Goetz also shared efforts being made at the time to change the situation, such as the Journal of Negative Results in Biomedicine, open notebook science, and others. After acknowledging these contributions, Goetz wrote, “These are great first steps. But freeing dark data should be the norm, not the exception. Once the storage and format problems are solved, scientists will need easy ways to search and retrieve each other’s data.” Think about how far we’ve come since 2007. Data storage solutions abound and great strides have been made to lower the barrier for the search and retrieval of data by the Linked Data community, introducing outreach standards such as JSON-LD and RDFa. Yet, with all of this progress, dark data has still been left in the dark.

Research—all research—has value. That value is independent of any grant awarded, tenure attained, or publisher’s acceptance. The value lies in learning. Knowledge, truth, and the pursuit of understanding something previously not understood are the goals and purposes of research. Nowhere in any definition of the word *research* have we read that research only has value if it produces cherry-picked outcomes that make administrators, editors, grantors, and publishers happy. And yet, due to the undeniable bias towards “positive” results—or

what have been deemed as such—much of the rest of all research is sentenced to a filing-cabinet death-by-obscurity. It does not have to be this way in today’s scientific technology revolution.

Incentives are the key here. Because future job prospects, tenure, and public acclaim all tend to revolve around journal publication, why would a researcher want to take the time to go back, write out, format, narrate, and then attempt to publish negative results that are likely to end up in unpretentious journals or on personal blogs that often are not recognized as of much worth by the research community, employers, or funding agencies? That’s a lot of effort and time for something that will essentially provide no career or financial advantages for the researcher.

That leaves two things we can do to bring dark data to light: one, change the incentives involved in research to better reflect the value of negative or inconclusive research results, and two, remove the barrier of time and effort that keeps a researcher from sharing this dark data in the first place. Ideally, these two changes would occur simultaneously, but funding agencies and institutional policies tend to be incredibly slow to change. Social expectations are often equally or even more sluggish to evolve. So we have a choice: do we sit around and wait until publishers and institution administrators change incentives to publish dark data? Or, do we find ways to make it worth sharing this dark data now?

If the time and effort barrier were removed from sharing nonconfirmatory research, we believe that many researchers would gladly share those results even if there were no external incentives. Open notebook science gives us a good starting place, but we need to overcome the problem of data deluge [2] inherent in many of the Web 2.0 approaches and start preserving knowledge, not just notes. Semantic Web technologies applied to that sort of “record as you research” mentality would produce human- and machine-understandable knowledge that could be sorted by machines and handed to researchers on a silver platter. Imagine if the data, analyses and conclusions, as well as the thoughts and intents of the researcher—or in short, every aspect of the research process—were preserved in context by intuitive, domain-specific research systems capable of generating the underlying semantic representations automatically. The development of open vocabularies expansive enough to preserve the entire research process are a prerequisite to building such systems. Taking it a step further, if that big-data amount of semantically understandable research was captured as a researcher went along—at the speed of thought [3]—the negative or inconclusive or downright befuddling results would be accessible without any extra effort.

The processes of researching and publishing would become distinct and separate from presenting the research results to a certain audience in, for example, a paper. If the researcher wanted to write and submit a paper based on those nonconfirmatory results, there would be nothing to prevent that. In fact, having the entire research process preserved would make writing up such journal articles easier anyway. Other presentation formats could be created, such as video or interactive media, without worrying that some important parameter or intricate concept might be left out. But the biggest game changer is that all of that dark data would come to light, and that the paper might not always be necessary.

## References

1. Goetz, T.: Freeing the Dark Data of Failed Scientific Experiments. Wired Magazine (2007), [http://archive.wired.com/science/discoveries/magazine/15-10/st\\_essay](http://archive.wired.com/science/discoveries/magazine/15-10/st_essay)
2. Open Notebook Science (Drawbacks), Wikipedia, [https://en.wikipedia.org/w/index.php?title=Open\\_notebook\\_science&oldid=654594797#Drawbacks](https://en.wikipedia.org/w/index.php?title=Open_notebook_science&oldid=654594797#Drawbacks)
3. Harnad, S.: The Future of Scholarly Skywriting. In: Scammell, A. (ed.) "i in the Sky: Visions of the information future" Aslib (1999), <http://users.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad99.aslib.html>