Molbank and the Convergence of Open Access, Open Data, and Open Source in Chemistry

Richard Apodaca

Blog post published November 30, 2006 in *Depth-First*
https://doi.org/10.59350/h2wh9-q5463

Abstract

Molbank, published by Molecuar Diversity Preservation International, is one of the oldest of a handful of Open Access journals in chemistry. Although its longevity is a remarkable accomplishment in itself, there is much more to Molbank than meets eye. Just below the surface is a feature so revolutionary, yet simple, that chemistry publishers years from now will wonder why *they* didn't implement it sooner.

Keywords

Physical and Theoretical Chemistry (Subfield)

Recommended citation

Apodaca, R. (2006, November 30). Molbank and the Convergence of Open Access, Open Data, and Open Source in Chemistry. *Depth-First*. https://doi.org/10.59350/h2wh9-q5463

Copyright

Copyright © Richard Apodaca 2006. Distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Molbank, published by Molecuar Diversity Preservation International, is one of the oldest of a handful of Open Access journals in chemistry. Although its longevity is a remarkable accomplishment in itself, there is much more to Molbank than meets eye. Just below the surface is a feature so revolutionary, yet simple, that chemistry publishers years from now will wonder why *they* didn't implement it sooner.

A Molbank article consists of a short monograph on a single compound, or possibly two. This may strike some scientists as a strange way to publish results, and it is unusual. On the other hand, this system offers vast potential to capture useful, but "unpublishable" findings that would otherwise be lost. Back when scientists actually read hardcopy journals, such a system would never have been feasible. Today, with hard drive space measured in terabytes, fiber optics cables crisscrossing the planet, Internet connectivity for almost everyone, and servers that can be had for virtually nothing, this system not only looks perfectly feasible, but preferable in many ways to the status quo.

Here's the revolutionary part: each article that Molbank publishes is accompanied by a publicly-available, machine-readable file encoding the structure of the article's subject molecule. That's it. There's nothing tricky or high-tech about it. In fact, the practice is about as low-tech as you could imagine. The file format in which structures are encoded, molfile, dates back at least fifteen years, and nearly every piece of chemistry software - both end-user and developer tools - can handle it. What makes Molbank's practice revolutionary is that not a single chemistry journal, Open Access or subscription-based, currently does this.

Why does the simple inclusion of a publicly-available molfile encoding molecular structures in a paper matter so much? This is where the second two entities of the trinity named in this article's title come into play: Open Source and Open Data. By providing a mechanism for a computer to decipher the chemistry in a paper, Molbank has opened the door to a host of highly-productive integration activities that nobody outside of Chemical Abstract Service has even been able to contemplate, let alone prepare for.

This article is the first in a series aimed at exploring the wide-open space that Molbank has created. Rather than arguing my point with words, I'll actually build working demonstrations of what is now easily within reach. At the same time, I'll document my work on this blog. I'm not sure where all of this will end up, but I do hope to shine some light on a vital, although currently obscure, component of the Open Access debate.