chem-bla-ics

Peer reviewed Chemoinformatics: Why OpenSource Chemoinformatics should be the default

Egon Willighagen

Published July 22, 2008

Citation

Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/7z23e-kmf12

Keywords

Cheminf, Blue-obelisk

Copyright

Copyright © Egon Willighagen 2008. 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.

chem-bla-ics

The battle for scientific publishing is continuing: openaccess, peer reviewing, how much does it cost, who should pay it, is the data in papers copyrighted, etc, etc.

The battle for chemoinformatics, however, has not even started yet. The Blue Obelisk paper (doi:10.1021/ci050400b) has gotten a lot of attention, and citations. But closed source chemoinformatics is doing fine, and have not really openly taken a standpoint against open source chemoinformatics. Actually, CambridgeSoft just received a good investment. I wonder how this investment will be used, and where the ROI will come from. More closed data and closed algorithms? Focus on services? Early access privileges? At least they had something convincing.

There are many degrees of openness, and many business models. I value open source chemoinformatics, or chemblaics, as I call it. There is a striking similarity between publishing and chemoinformatics. Both play an important role in the progress of sciences. A big difference is that (independent) peer review of published results is done in scientific publishing, but not generally to chemoinformatics. Surely, algorithms are published... Ah, no; they are not. They are described. Ask any chemoinformatician why this subtle difference is causing headaches...

Let me just briefly stress the difference between core chemoinformatics, and GUI applications. The first *must* be opensource, to allow **independent** *Peer Review*; the latter is just nice to have as opensource. Bioclipse is the GUI (doi:10.1186/1471-2105-8-59), while the CDK is our peer-reviewed chemoinformatics library (pmid:16796559). I would also like to stress that the CDK is LGPL, allowing the opensource chemoinformatics library to be used in proprietary GUI software. We deliberately choose this license, to allow embedding in proprietary code. The Java Molecular Descriptor Library of iCODONS is an example of this (that is, AFAIK it's not opensource).

So, getting back to that CambridgeSoft investment. I really hope they search the ROI in the added value of the user friendly GUI, and not in the chemoinformatics algorithm implementations, which, IMHO, should be peer-reviewed, thus open source. Meanwhile, I will continue working on the CDK project to provide open source chemoinformatics algorithms implementations, for use in opensource *and* proprietary chemoinformatics GUIs.