

Does ChemSpider really violate Open Data with CC SA?

Egon Willighagen 

Published May 10, 2008

Citation

Willighagen, E. (2008). Does ChemSpider really violate Open Data with CC SA?. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/zebnk-k6z97>

Keywords

Chemspider, Copyright, Nmrshiftdb

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

ChemSpider is afraid they are doing something bad because they release their data as [CC-BY-SA](#). Because, John Wilbanks says in Peter's blog:

*I would add to it that I'd like to see a meaningful discussion of the risks of Share Alike and Attribution on **data integration**. Chemspider's move to CC-BY-SA fits into this discussion nicely - it's a total violation of the open data protocol we laid out at SC, which says "Don't Use CC Licenses on Data" - **but it does conform inside the broader OKD.***

Now, let's take this into pieces.

1. John notes that ChemSpider is in compliance with the [OKD](#). This means, that ChemSpider thinks about Open Data just like the [Open Knowledge Foundation](#) does. I've scanned through the OKD, and it indeed seems to support the BY and SA clauses of the CC. So, Chemspider did not do a bad thing.
2. Data integration is tricky: you have to keep track of license information on an entry-by-entry level. For each fact, you keep to track the source, and associate the source with its original license. For example, the [NMRShiftDB](#) information in ChemSpider should be [GNU FDL](#).
3. OpenX licenses may be viral. This holds for the [GNU GPL](#) as well as for the CC-BY-SA. Nothing new there. It just requires that when you would like to incorporate the ChemSpider data into a larger database, that database has to be CC-BY-SA too, or likely at least CC-SA.

Summarizing, I think ChemSpider did a good thing, and that ChemSpider does **not** violate the OpenData idea, but instead, that the CC-BY-SA and the OKD violates John's requirements for integrating data resources (apparently based on a two year legal study). That has nothing to do with ChemSpider.

Now, people will always have different opinions on Openness. The original BSD clause had a [restrictive 'advertisement' clause](#), not Open enough for at least the [Debian Free Software Guidelines](#) (DFSG), while still open source. The clause was later removed from the BSD license.

Another [Debian](#) example is Firebox, which is named [IceWeasel](#) in Debian, because the 'license' on the Firefox name is not open enough.

Another problem with the definition of Openness, is the viral aspect of some licenses (see earlier). For some, the GPL is not open enough, because it does not give people the freedom to license their software they like themselves, something the BSD and MIT licenses do allow. There is ongoing debate (and that should be ongoing) on how much *freedom* a license must provide to be called Open. The whole OpenAccess discussion is similar (see e.g. [Peter's story on this](#)), where the discussion on the minimal amount of freedom is even worse.

Should we worry about ChemSpider being 'only' CC-BY-SA? Maybe. Data is not software, but I disagree that viral license would be OK for software, but NOT for data. That's just BSD-versus-GPL all over again. I am happy about OpenBabel being GPL, and I am happy about ChemSpider being CC-BY-SA too.

chem-bla-ics

All that said, these discussion are important. And creating good definitions of what freedoms are required, are crucial in deciding whether something is Open. The Blue Obelisk does not have/use such definitions yet, and we should start discussing this, and define a Blue Obelisk ODOSOS Guidelines. Please no funny jokes about how we can boogy then :)

Now, looking forward to hearing what you think about these issues... Looking forward to the other blog items!