chem-bla-ics

The Molecular Chemometrics Principles #1: access to data

Egon Willighagen

Published August 9, 2010

Citation

Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. https://doi.org/10.59350/srwf0-4gf52

Keywords

Chemometrics, Mcprinciples

Copyright

Copyright © Egon Willighagen 2010. 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.

The meetings in and around Oxford were great! I already wrote that the Predictive Toxicology workshop was brilliant (see Oxford... #1) and Oxford... #2), but I also very, very much enjoyed meeting up with Dan and Nico! During the week, someone (name and address is know at the editorial office) commented on the fact that my blog posts are somewhat difficult to follow; that is, it's often not clear why I am posting what I am posting.

Indeed, I am not particularly one of those bloggers who spends trees after trees, in great detail explaining what is going on. I do make a lot of use of hyperlinking; much more than the average blogger. I actually assume that readers follow links, to read about the perspective of a blog post. But we all know that scientists do not read the cited papers in a paper they are reading, so who am I to assume blog readers would start doing that with blogs :)

Well, since principles seems popular , it might be a good start of my grand scheme that is behind this blog: the Molecular Chemometrics Principles. Hence, this first post about the why. The why is simply to provide a reference frame to what I am blogging about. In the next few posts on these McPrinciples (is that a catchy name, or what?) that will appear over the next two weeks, I will outline the code of chem-bla-ics. And, moreover, from now on, I will tag all my posts with the reaons why I make that post. I am sure that will not be too helpful for the occasional reader, but for anyone who is serious about chem-bla-ics, this will be a genuine gold mine of data for pattern recognition and data mining otherwise.

So, here goes.

Molecular Chemometrics Principles #1: In order to reproduce cheminformatics studies you need access to the input data.

The reason for this is that statistical modeling very much depends on the data on which modeling was done, patterns were recognized, etc. Therefore, without the input data, it is practically impossible to accurately reproduce results. Fortunately, the acceptance of the importance of access to data (e.g. as Open Data) is slowly getting momentum in science.

Further reading: Molecular Chemometrics, 2006 (doi:10.1080/10408340600969601)