

NMRShiftDB RDF #2: Some statistics

Egon Willighagen 

Published September 5, 2009

Citation

Willighagen, E. (2009). NMRShiftDB RDF #2: Some statistics. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/3afny-p1c44>

Keywords

Nmrshiftdb, Sparql

Copyright

Copyright © Egon Willighagen 2009. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

This morning I had some more fun, and since the [statistics](#) view on the [NMRShiftDB](#) server is down, I thought I could recalculate the statistics myself. Because the current RDF version of the data does not include all information yet, I cannot reproduce all of them. On the other hand, I can determine some other interesting statistics.

Spectra per spectrum type

One of the statistics given in the aforementioned page is the number of spectra per nuclei. This can be recalculated with the following SPARQL:

The results for the 1.3.3 release are:

nucleus count

¹³ C	21958
¹ H	3031
¹¹ B	326
¹⁷ O	131
¹⁵ N	79
¹⁹⁵ Pt	68
¹⁹ F	50
³¹ P	38
⁷³ Ge	18
³³ S	8
²⁹ Si	5

I am a bit surprised by the count for the silicon NMR spectra, as I would have thought I alone had entered more than just five.

Molecules with the most spectra

It turns out that the molecules have in the 1.3.3 NMRShiftDB release at most 7 spectra, as I can calculate with:

That is going to change, as the paper I am digitizing now ([doi:10.1021/jo971176v](https://doi.org/10.1021/jo971176v)) has carbon and hydrogen NMR spectra for 7 solvents for each compound :) It should be possible to summarize the number of molecules for each number of spectra per molecule, but did not manage to get this SPARQL to work out well.

BTW, did you know you can find reprint PDFs of a paper (if any; this one happens to have a [PDF copy](#)) with Google using the title in quotes and `filetype:pdf`? Try [this query](#). The top hit was molecule 10016314 ([RDF](#)), which has 4 ¹³C spectra, one ¹⁵N and two proton NMR spectra.

Molecules with the most different nuclei

In the first query, we already saw in the first SPARQL, there are 11 different nuclei in the database, though carbon and hydrogen are by far the most abundant spectra. I like diversity, so one statistic I find interesting, is the molecules which have spectra with the most different nuclei. This is done with the query:

It shows that molecule 10023801 ([RDF](#)) has 5 different NMR types: ^{13}C spectra, one ^{15}N , ^{29}Si spectra, one ^{17}O , and ^1H spectra. Unfortunately, the compound also has chlorines, so it disqualifies as molecule for which NMR spectra are available for all its elements.