

The situation

We currently have two databases for nuclear magnetic resonance (NMR) data in the consortium (plus others outside, commercial and academic):

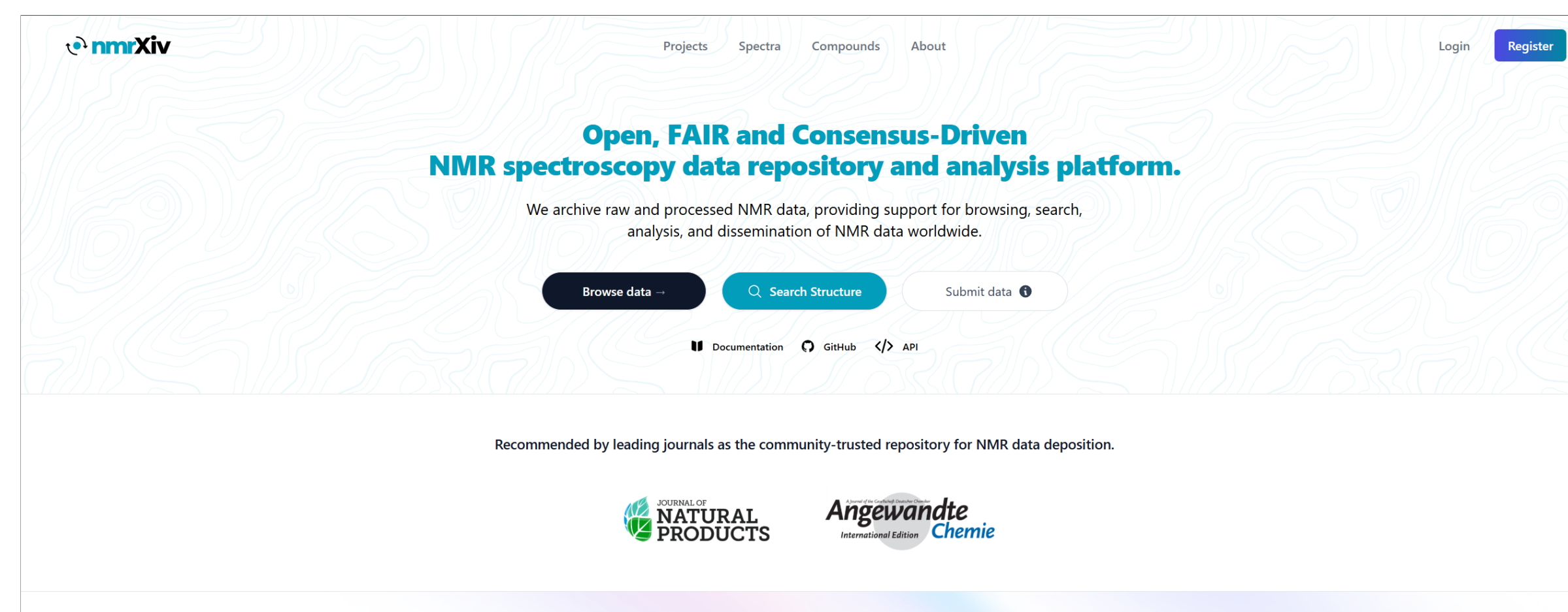


Fig. 1: nmrXiv title page.

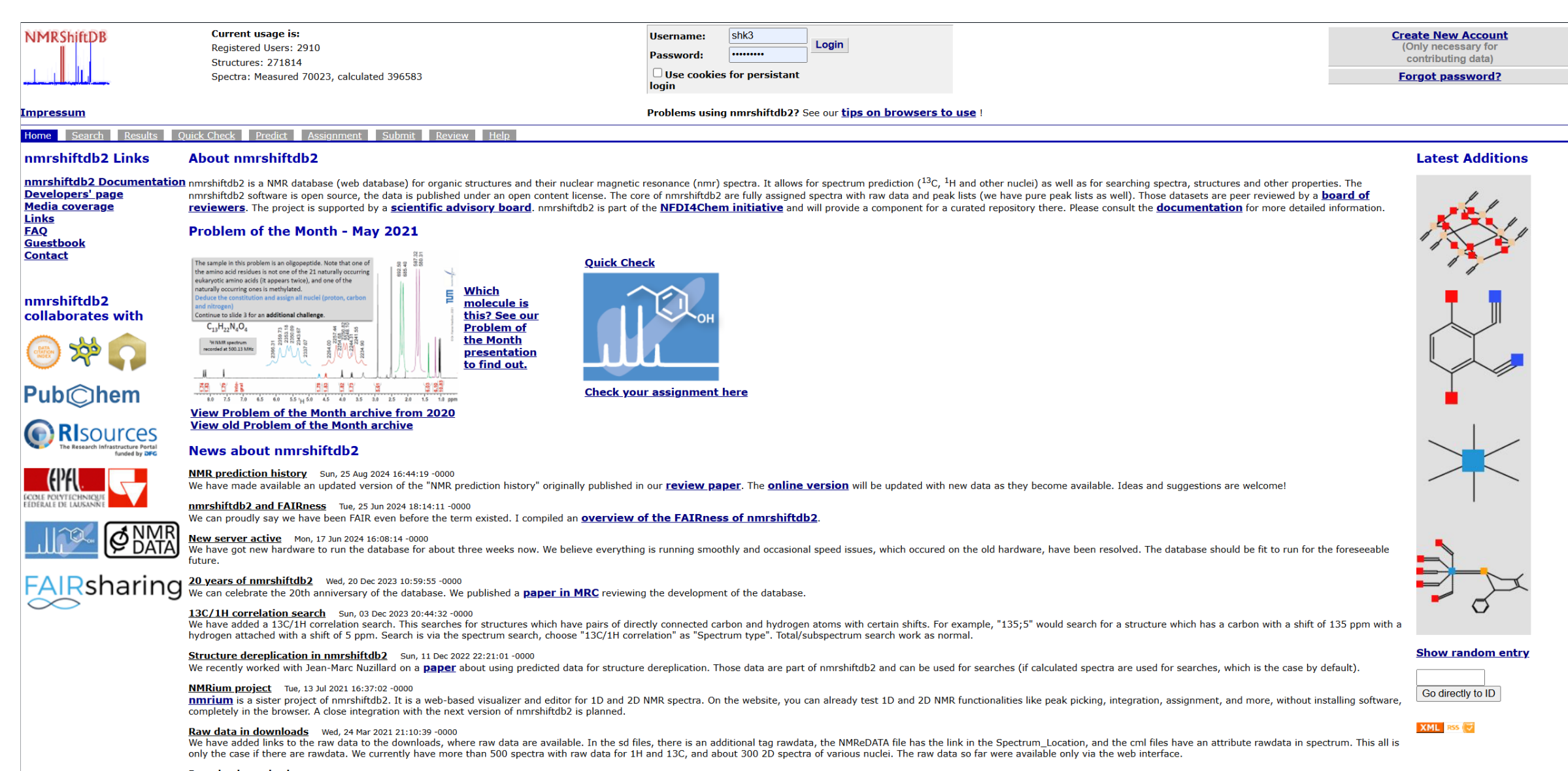


Fig. 2: nmrshiftdb2 title page.

At first, this looks like a redundancy.

Differences

At a closer look, there are significant differences between the two:

nmrshiftdb2

- Focuses on assigned shifts (but can store raw data as well)
- Holds shifts, structures, and some metadata
- Uses a custom database scheme holding molecules and shift data
- Can directly query this for assigned shifts doing predictions or structure elucidation tasks
- Interface is pure html generated by Java on the server
- Offers NMR prediction via the browser or web service, using the current database content

nmrXiv

- Requires raw data for all submissions
- Holds raw data, shifts, structures, and extensive metadata
- Uses the MDL cartridge and the NMRium format for storage
- Has a browser-based interface using Laravel and Inertia.js
- Integration with NMRium
- Feeds into work like the MARGARITAS

Continue both?

Given that merging project is difficult and shutting down not desirable, a continued existence of both databases could be the best solution. For this, the following procedures could be used:

- Both databases allow submissions
- Peak data are replicated to nmrshiftdb2 and full data to nmrXiv (as has happened)
- Separate user interfaces are maintained, either with different functionality or layout

- API functions can be different
- API interface should be common

Case study: Organometallic compounds

Organometallic compounds cannot be handled easily in a standard graph-based organic chemistry framework. In [1], minimal extensions (mainly the use of a zero bond) were suggested to make them compatible with e. g. the mol file format. Since the database backend of nmrshiftdb2 is a custom product and the chemistry is handled by the open-source product CDK [3], it was possible to make nmrshiftdb2 use zero bonds [2].

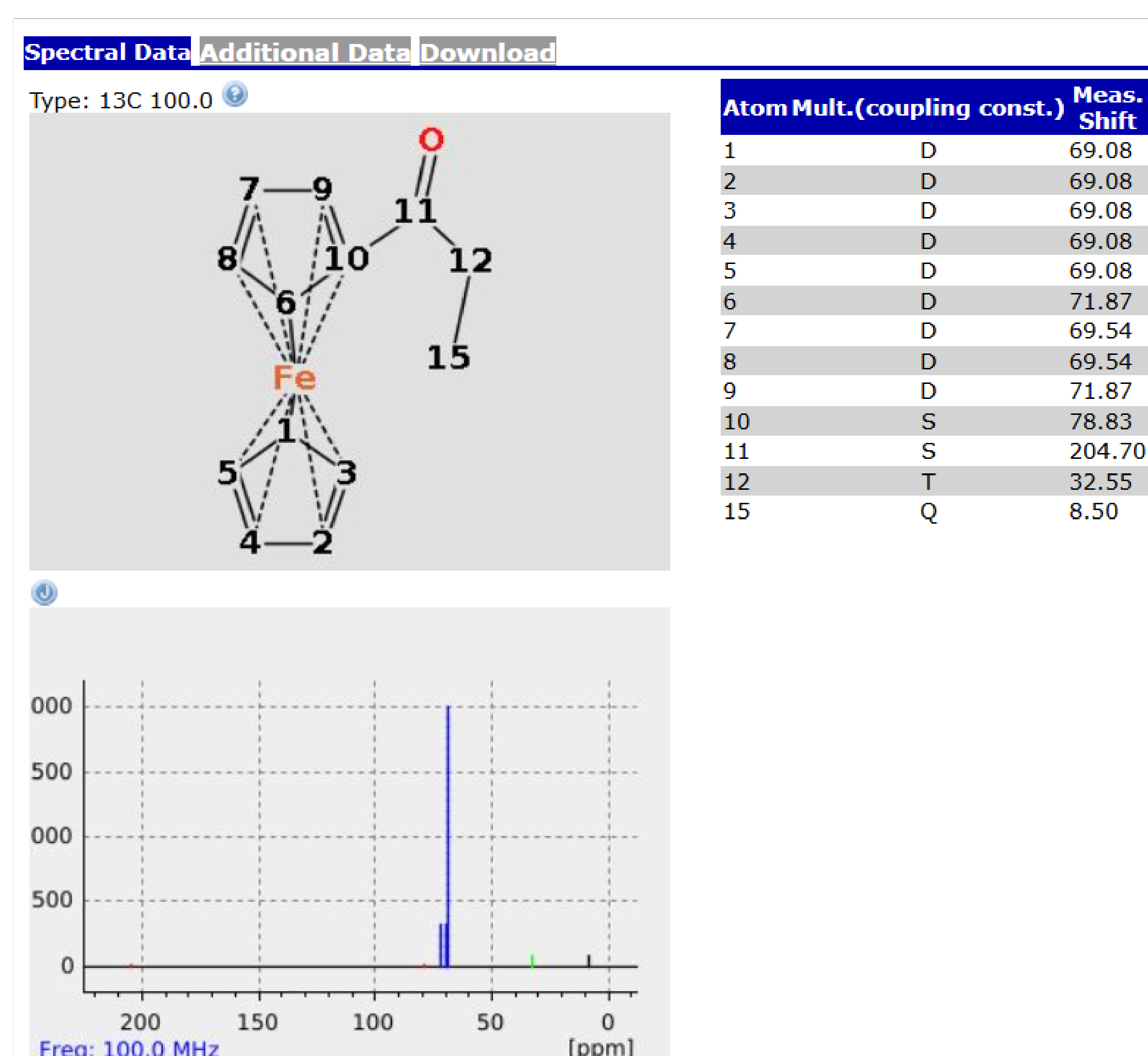


Fig. 3: A ferrocene compound in nmrshiftdb2.

The database can now not only store organometallic compounds, it can also store the assignments as for other molecules. Based on this, functions like prediction or substructure search can be used for organometallic compounds.

The use of zero bonds makes sure that organometallic bonds can be distinguished from covalent bonds. This is important for the prediction. Right now, only HOSE codes are used, but we will extend this to machine learning methods.

Conclusions

There are enough differences between nmrXiv and nmrshiftdb2 to justify their separate existence. Apart from offering features including NMR prediction, nmrshiftdb2 allows adopting for unusual use cases. The organometallic compounds are an example for this. An update of the nmrshiftdb2 user interface is desirable.

References

- [1] Alex M. Clark. "Accurate Specification of Molecular Structures: The Case for Zero-Order Bonds and Explicit Hydrogen Counting". In: *Journal of Chemical Information and Modeling* 51.12 (2011). PMID: 22111697, pp. 3149–3157. DOI: 10.1021/ci200488k.
- [2] Stefan Kuhn et al. "An NMR Database for Organic and Organometallic Compounds". In: *Molecules* 30.16 (2025). ISSN: 1420-3049. DOI: 10.3390/molecules30163442.
- [3] Egon L. Willighagen et al. "The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching". In: *Journal of Cheminformatics* 9.1 (June 2017), p. 33. DOI: 10.1186/s13321-017-0220-4.