Mnova DB - MyData

Application Note

Mestrelab Research

chemistry software solutions

Turning your spectroscopic data into a searchable resource

Using MyData to efficiently retrieve what you are looking for in a few mouse clicks

Gary Sharman, September 2020

Introduction

You are a coming to the end of your PhD. It's been a long day in the lab, but you made great progress. Your new synthetic route seems to be working and you ping your boss to tell him you have 25mg of the product in a vial. 'That is fantastic' he replies. 'Do me a quick favour though, just so I sleep well' he says. 'Can you send me the NMR along with the one when you made it by the other route last year so I can compare them'.

A feeling of cold horror spills over you. Its 9 pm. You are supposed to be going out tonight. How are you going to find that spectrum amongst all the other data you have? You start wading through your notebooks, and the files you have saved on your laptop. It's here somewhere you say... Finally, you find it! You send it off to him. It's now 11pm and your evening out is ruined.

You vow to sort out all your spectra the next day, so it doesn't happen again. You spend most of the day wading through your notebook and compiling a list in excel of all the compounds with a link to their data files you have saved. Great you think. If he asks me for anything else now, I will be sorted.

Tonight, you are going to the cinema with a friend, it's going to be great. You are just about to leave when a Skype message pops up from your supervisor. 'Great work making that compound by the new route' he says. 'Just one thing, there seems to be an impurity in there and I wonder if it's from the reagent we used. Can you quickly check if it's in any of the other reactions you did?'

You text your friend and cancel the cinema.

If only you had had MyData! MyData is a plugin for Mnova, which indexes all your spectra – NMR, LCMS or other – making them instantly searchable. Spectra can be added to the MyData repository with the click of a button, taking just a second to store all the metadata, processing and any assignments or customizations with it. Once in the repository, you can search by name, date or any parameter. What is more, you can search the spectra

Feliciano Barrera 9B – Bajo 15706
Santiago de Compostela (SPAIN)
www.mestrelab.com
info@mestrelab.com
Phone: +34 881 976 775
Fax: +34 981 941 079

themselves for matching peaks or features, turning an unruly, disc hogging, dark archive of data into an amazing resource.

What is MyData?

MyData is a personal database from Mnova. It is tightly integrated into the Mnova software, such that an open document containing NMR, LCMS or other spectroscopic data like IR or UV can be sent to the DB with one click. When this happens, the data present in the documents is parsed and indexed to make it searchable. Filename, instrument name, date, title, structure and a host of other parameters become searchable fields. But perhaps the killer feature is that the spectral data itself becomes searchable. Multiplets and peaks in the NMR spectrum or masses and retention times in the LCMS all can be queried. It is even possible to directly search spectral raw data via a matching function. Searches can be performed by text, but also graphically from the Mnova interface.

Easy set up – the database with no pain

One of the reasons you use Mnova is probably the ease of set up and use for the core software or the plugins. The very mention of the word 'database' might therefore put you off: doesn't that mean you need to get an administrator to set it up on a server? Won't it force you to do lots of extra work just to 'comply' with what the database expects or needs? The answer to all these questions is no.

MyData runs locally on your laptop or office computer. There is no need for complex remote server database installation, configuration, and testing. You simply install the MyData package and you are ready to start using it. It does not force you to add data in any format or do extra work. In fact, the opposite is true: it takes whatever you have, and with no extra effort on your part, makes it a searchable resource.

As our introduction 'user story' suggests, you also need to consider the time and effort *not* having a database costs you. You spend time organising your data in a way you hope makes it easy to find things. You spend time searching for the data you need. Most likely you spend time doing both. MyData takes care of all of this for you.

So don't see MyData as a database. See it as a personal assistant that takes the drudgery out of organising data and retrieving it.

Adding data with one mouse click

Adding new data into MyData could not be simpler. With a document open in Mnova, you simply click on 'save to database' to create a new record. If you have a structure attached, it will warn you if the same structure already exists in the database, and you will be given chance to create a new record or add to the existing one.

Feliciano Barrera 9B – Bajo 15706
Santiago de Compostela (SPAIN)
www.mestrelab.com
info@mestrelab.com
Phone: +34 881 976 775
Fax: +34 981 941 079

Figure 1 – one click is all it takes to make your spectra searchable

If you have large amounts of existing data, Mnova provides efficient tools to make that available within MyData. You can use a standard Mnova script to automate adding them into your MyData store, or use tools such as MyGears to assist the process.

Browsing and Recalling Data

You can browse any database you have created, or the results of any search you have done. You have control over the layout and presentation of the view, choosing from a number of predefined ones or designing your own. In this view (figure 2) we see a list of structures in a table – clicking any one gives a preview of all the NMR data available for that structure. Once you have found a record of interest, you can export a single data item or the whole record back into Mnova for further review.

Figure 2 – A configurable, browsable view of all your spectra

Feliciano Barrera 9B – Bajo 15706
Santiago de Compostela (SPAIN)
www.mestrelab.com
info@mestrelab.com
Phone: +34 881 976 775
Fax: +34 981 941 079

Searches

The real power of MyData is searching. Any field extracted is searchable: a name, a parameter, an instrument. The creation of searchable fields effectively negates the need to organise your data, as MyData makes it possible to retrieve data organised according to any criteria depending on how you do your search. Organising data manually in a tree like folder structure does not achieve this. It makes *one* type of search easy (e.g. by date if you organise by date) but does not help with any other query you might want to do.

In addition to text and numerical fields, chemical structures are searchable for exact matches or for substructures. Most importantly though, the spectra themselves become searchable. Here we will consider these two examples in more detail.

The substructure search is very useful for finding data on compound similar to something you have. For example, perhaps you want to know the NMR signature for a particular functional group, and so want to see all spectra that have that group. To perform the search, you draw the structural group you want to find, and then click on 'molecule search'. Matches are shown in a list, with a score showing the degree of match. You can choose which results you want to view – they are then taken into the browser for you to see them.

Figure 3 – a substructure search in MyData. A substructure is drawn and all records that match that feature are returned. Matching results can then be browsed easily

The spectral search allows you to search by multiplets, peaks or even the raw spectrum. You could use this to find an exact match in the database (did I make the same thing as last time?) or to search for an impurity across a set of spectra, to pick but two examples. Figures 4 and 5 show this process in more detail. Another powerful feature of the search engine is the ability to define your search intent. Perhaps you have a very clean spectrum and you want a close match to all the peaks. Or perhaps your sample is very dirty but you want to pick out a pure

Feliciano Barrera 9B – Bajo 15706
Santiago de Compostela (SPAIN)
www.mestrelab.com
info@mestrelab.com
Phone: +34 881 976 775
Fax: +34 981 941 079

component that is in the database. The search engine lets you change this behaviour to ensure the best results for the kind of sample and MyData store that you have.

Figure 4 – A spectral search matching all the multiplets in a spectrum to find previous examples of this compound

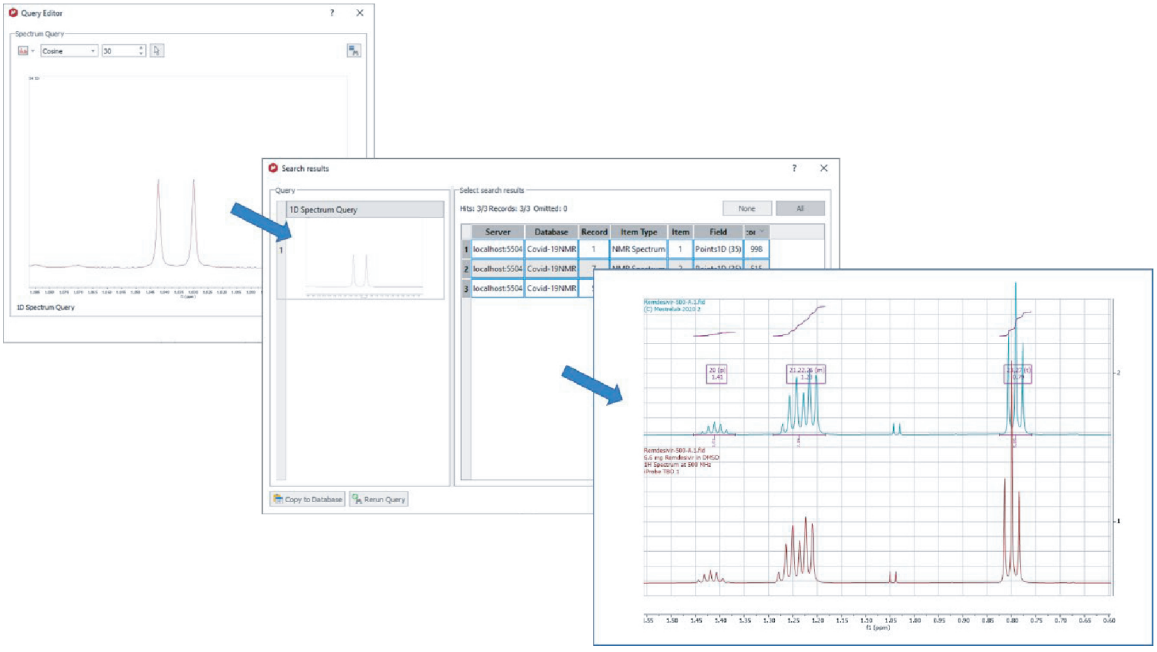

Figure 5 – a direct spectral search for an impurity to see if it has been found in any other samples. Once found the data is stacked in Mnova for easy comparison

Feliciano Barrera 9B – Bajo 15706
Santiago de Compostela (SPAIN)

www.mestrelab.com
info@mestrelab.com

Phone: +34 881 976 775
Fax: +34 981 941 079

Next steps

Perhaps after you have been using MyData for a while, you begin to see value in sharing your database with others. MyData can be expanded in this way, supporting a shared or even corporate level DB. The database 'backend' can be one of many standard ones - MySQL, Postgres or Oracle. Once configured, you use it in an identical way to MyData, with the same searches and views. It's just that now you can share the results with colleagues or across your organisation.

Key Points

- MyData is a personal, searchable and browsable datastore for all of your spectroscopic data
- MyData is ideal for individual researchers such as PhD students, but is equally useful for any scientist in academia or industry that needs to keep track of their own data
- Once installed any data can be made searchable at the click of a button. There are no complex data or metadata requirements
- Its easy to install and runs locally on your computer
- It is readily extendible to a shared database using industry standard DB platforms

Epilogue

You are a coming to the end of your PhD. It's been a long day in the lab, but you made great progress. Your new synthetic route seems to be working and you ping your boss to tell him you have 25mg of the product in a vial. 'That is fantastic' he replies. 'Do me a quick favour though, just so I sleep well' he says. 'Can you send me the NMR along with the one when you made it by the other route last year so I can compare them'.

You do an NMR search in Mnova MyData and send it to him one minute later.

The next day you have a relaxing coffee reading the newspaper.

Tonight, you are going to the cinema with a friend, it's going to be great. You are just about to leave when a Skype message pops up from your supervisor. 'Great work making that compound by the new route' he says. 'Just one thing, there seems to be an impurity in there and I wonder if it's from the reagent we used. Can you quickly check if it's in any of the other reactions you did?'

You do a spectral search in Mnova MyData and send him the results 5 minutes later. You go to the cinema and have a great evening.

Feliciano Barrera 9B – Bajo 15706
Santiago de Compostela (SPAIN)

www.mestrelab.com
info@mestrelab.com

Phone: +34 881 976 775
Fax: +34 981 941 079