TONY DAVIES COLUMN

Who's ahead in the Cloud?—Part three

Antony N. Davies[a,b] and Mohan Cashyap[c]

[a]Strategic Research Group—Measurement and Analytical Science, AkzoNobel, Deventer, The Netherlands
[b]SERC, Sustainable Environment Research Centre, Faculty of Computing, Engineering and Science, University of Glamorgan, UK
[c]MASS Informatics, Harpenden, UK and Tech Mahindra, London, UK

Following on from the two recent articles on how the Cloud may be impacting the availability of scientific software delivery for spectroscopists, in this issue we want to look at what the wider commercial spectroscopy software providers have been doing in this area.

What is the state-of-the-art amongst spectroscopy software vendors?

Initially, Mohan's research into this area was pretty difficult, as with so many areas of science or "trendy" IT-driven concepts, there are as many different understandings of "Cloud "as there are software vendors. The delivery of spectroscopic data processing applications is a niche market without the sales volumes of day-to-day office software products, and therefore does not have the large teams of software developers and product managers behind them driving their product forward at the leading edge of Information Technology. In fact it is probably fair to say that with the increasing rapidity with which operating systems providers change their underlying platforms and drop support for older versions we are lucky that our spectroscopy vendors succeed in keeping their offerings running! However, after the initial stunned silences what has emerged is actually starting to look like a pretty good picture overall. We need to be clear as to what we are actually looking for. So a quick revision of what our NIST Cloud Computing essential characteristics are: **On-demand self-service, broad network access, resource pooling, rapid elasticity and measured service**. Possible underlying Service Models vary through **Software as a Service (SaaS)**, **Platform as a Service (PaaS)** to **Infrastructure as a Service (IaaS)**. With the Deployment Models being variously **Private Cloud**, **Community Cloud**, **Public Cloud** or a mixture known as a **Hybrid Cloud**. For a more detailed discusion on these descriptions in our environments see our earlier column.[1–3]

Now what is interesting is that in discussions with spectroscopy software vendor contacts, we found that the category they actually try to place their product offerings into depends on what their role is within their organisation. The easiest differentiator is of course whether you need to install the software on to your desktop, if so it is not cloud computing... unless of course what you install is some sort of application which actually just works as a front-end for the provision of the functionality from a centralised server in the Cloud and does not have any stand-alone capabilities of its own. If you are processing your data via a web-based application in your browser, then the software might be a Cloud application… but then we need to look at the details of the installation and possibly the contracts your organisation has with the software provider which is thankfully outside the scope of this month's column! So let us look at responses we have received from some familiar spectroscopy software providers and their take on Cloud computing provision.

Starting alphabetically with ACD/Labs. They were an early adopter of the possibilities of providing spectroscopic data processing as a paid service (meeting the "measured service" characteristic) through a web-based front-end as an alternative to their local workstation installed solutions with their I-Lab product.[4] Since then the product has been further developed and has a quite different look with NMR prediction functionality and the expected physical properties also available (Figure 1).

This model is also available for any of the Cloud deployment models should the "public" be unacceptable to a corporate IT/IP policy. Mike Partington (ACD/Labs' Senior Director Worldwide Sales and Strategic Partnerships) was very positive describing their thinking with regard to Cloud services as *"While there is no doubt that organisations are achieving high cost-savings by outsourcing chemistry and analytical work to partners around the globe, this has resulted in a knowledge gap for the sponsor organisation. By no longer directly employing the scientists to do some of the project related work, knowledge and expertise is getting fragmented and lost. Furthermore, many employed data transfer mechanisms are clunky and limited. We've heard all about the inefficient transfers of dead data in the form of PDFs, drawings, spreadsheets and images, and believe there is a better and more efficient way. The solution will require commitment from IT organisations to embrace technologies in the cloud and over the web. It is clear that this is the foundation for the answer to this industry problem."*

LabCognition's[5] Klaus Schürmann and Maren Fiege described the use of
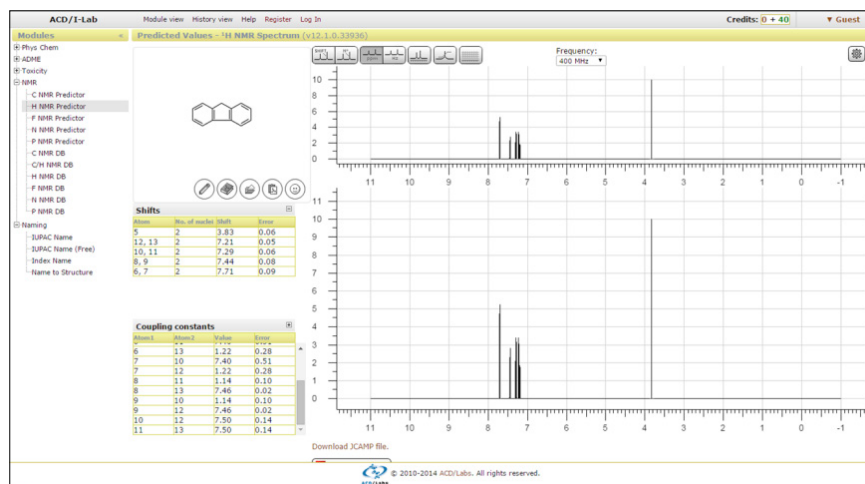
TONY DAVIES COLUMN

Figure 1. ACD/Labs' I-Lab web-served solution.

the Cloud as an obvious way forward for spectroscopic data analysis and solution delivery which has interestingly, and as far as we can tell uniquely, also included instrument control. With their specific experience in understanding and developing multi-vendor instrument control as well as advanced spectroscopic data processing they have recently completed the successful functionality migration and deployment to a specifically Cloud-capable environment which they are calling SciGear. This solution includes instrument control functionality in the optical spectroscopies such as infrared, near infrared, Raman and UV-Vis spectroscopy, as well as the capability to realise spectroscopic and chemometric workflows into simple web-based "apps" (see Figure 2). What was fascinating in talking to them was the capability to deploy this solution in a secure corporate Cloud environment and control instrume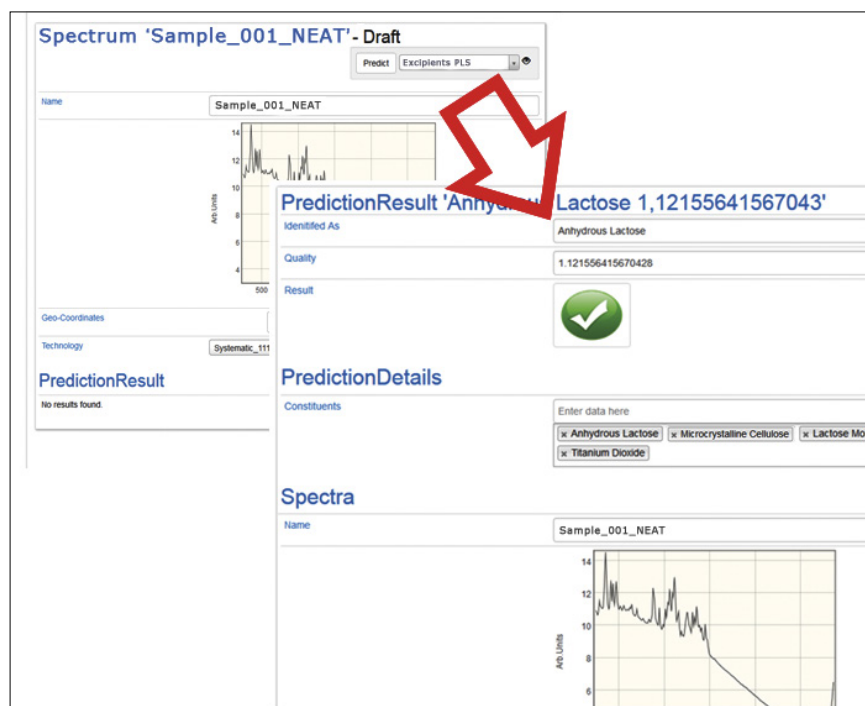nts in various locations around the globe. Obvious really when you have the functionality available; but something which opens up a number of very interesting possibilities for secure instrument deployment strategies where the legal location and storage of the analytical data plays an important role.

MestreLab Research are well known for their NMR and GC/LC/mass spectrometry software solution Mnova, which has always been developed with a focus on integrating advanced data handling capabilities into the working practices of their clients recognising the importance of keeping reporting as simple as possible to help productivity. Carlos Cobas (co-founder and President of Mestrelab Research) explained that their development efforts have recently included the deployment of a new strong chemical structure/reaction oriented electronic laboratory notebook solution "Mbook" capable of Cloud deployment which is currently undergoing end-user acceptance testing (Figure 3) with the capability of running as a native App rather than in web browser. Carlos commented *"Cloud-based solutions pave the way towards the high throughput structure verification of small molecules aimed at calculating the degree of compatibility between a proposed molecular structure and one or more sets of spectral data. The ability by which computing resources can be scaled up and down dynamically to meet a varying workload, for instance, according to the size of the data to be analysed, is a key feature that favours cloud computing technology compared to other traditional models. This concept can be extended to the computationally expensive calculation of NMR parameters such as those involving DFT/GIAO simulations where all the infrastructure can be defined in such a way that the final client is isolated from the complex installation of the QM packages or selection of the basis sets."*

Another pioneering product was developed in a collaboration between Galactic Industries and Nicolet and is now part of the portfolio of Thermo Fisher Scientific called FTIRsearch (Figure 4).[6] We were pleased to hear that it had recently been re-vamped to a Cloud provision. David Joyce (Senior

Figure 2. LabCognition's SciGear solution built for Cloud deployment.
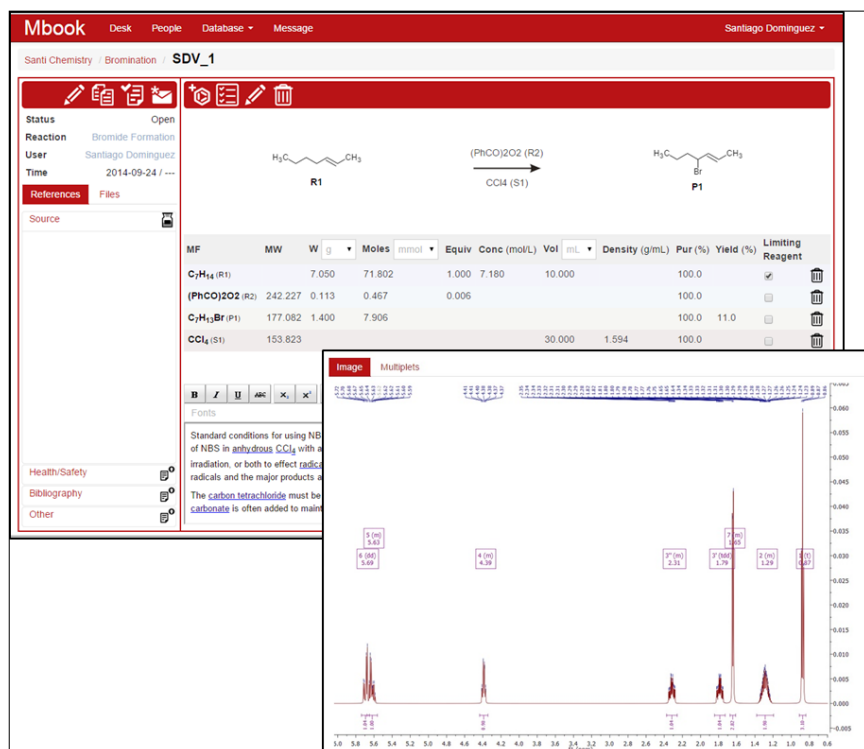
TONY DAVIES COLUMN

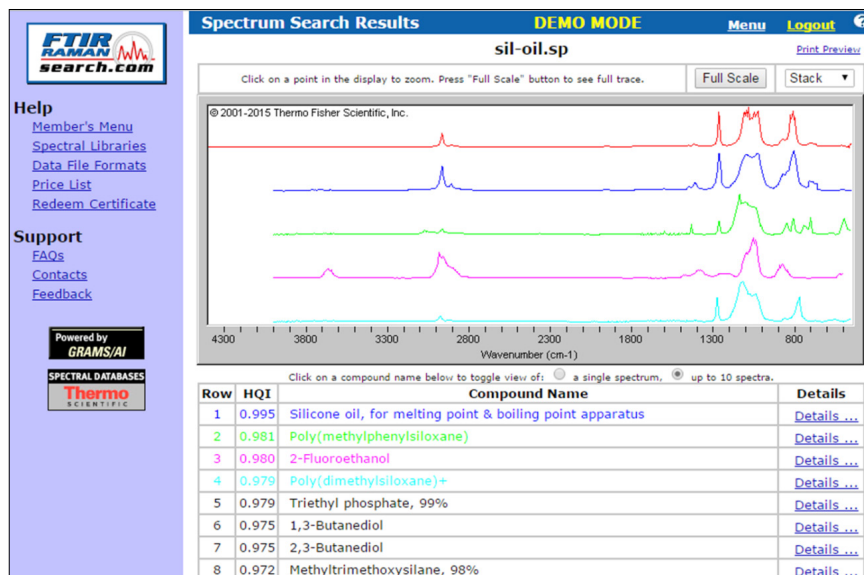**Figure 3.** Mestralab Research Mbook App.

Figure 4. Thermo Scientific's FTIRSearch.

Product Manager, Informatics, Thermo Fisher Scientific) commented: *"Thermo Fisher has been at the forefront of cloud technology for over a decade. FTIRSearch.com, which has been serving spectroscopists for almost 15 years, received its fourth major overhaul in January moving to an entirely virtual infrastructure. More recent initiatives include collaboration in the novel Mass Spectral database MZCloud.org and lifetechnologies.com 'cloud suite' platform. The latter, with its application of cloud-based computational power to huge next generation sequencing datasets, demonstrates a new dimension in instrument data processing beyond the limitations of the desktop PC."*

Jeff Seymour (Modgraph Consultants), a vastly experienced spectroscopist and long-time supporter of this column, drew up a list of some "Cloud capable" systems, but reiterated his fears around the use of poor quality data getting into openly available prediction software.

http://www.hmdb.ca: This is not strictly considered a cloud based resource, but it is certainly a very valuable web server which contains many NMR spectra of metabolites. I think that it even contains some API so that any cloud-based application can use it.

http://www.chemspider.com: It also contains many searchable NMR spectra and an API to build cloud-based solutions.

Universal J Coupling (http://www.perchsolutions.com/juniper.html): Commercial "cloud-based" system for the prediction of NMR scalar couplings. Using Cloud Storage for NMR Data Distribution.[7]

http://www.nmrdb.org: Prediction and databasing.

Other "cloud"-based programs for spectral search and prediction are available—some are paid services such as: Bio-Rad Sadtler's KnowItAll Enterprise server[8] and Wiley's SpecInfo online.[9]

Others offer free spectral search such as: SDBS's http://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_index.cgi and C13 Spectrum similarity search based on 61 million predicted PubChem structures from Wolfgang Robien: http://nmrpredict.orc.univie.ac.at/similar/eval.php.

There are also free prediction programs from NMRShiftDB's http://nmrshiftdb.nmr.uni-koeln.de and Spinus, http://www2.ccc.uni-erlangen.de/services/spinus/.

Finally, there is also the "Robot Referee" from Wolfgang Robien where those publishing articles which include C13 spectral data are encouraged to upload their data and have it automatically checked to see whether it is ready for publication or whether it needs minor or major revision.[10] It is possible that some publishers will insist in future that all data is reviewed by such an automatic program.
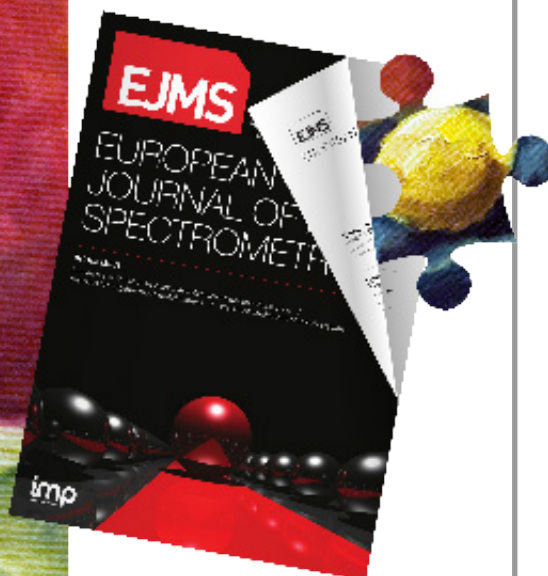
Jeff pointed out some of the problems with the functionality of these ported legacy systems which are now available through web-based front-ends.

"*…where almost all the current programs fall down is that they are too manual to operate. For example, if someone wants to run a spectrum search online, on a spectrum they have just measured, they will want to do so at the click of a button, from within their spectral processing package. They will not want to manually type in all the peaks. There would appear to be a gap in the market for a way to include the online search and prediction capabilities supplied in the programs listed above within the software being used to work with the spectral data.*"

Conclusion

So, as we can see there is some activity in moving software supporting spectroscopists into the Cloud although it initially was something of a fight to identify it. Hopefully, as we train up our next generation of analysts born to a world of apps and tablets more of our service providers will grasp the opportunities now available to make spectroscopic data handling more distributed and user friendly!

References

1. P. Mell and T. Grance, *The NIST Definition of Cloud Computing*. NIST Special Publication 800–145 (2011). http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf
2. A.N. Davies and M. Cashyap, "A head in the Clouds?—Part One", *Spectroscopy Europe* **26(3),** 21–22 (2014). http://bit.ly/tdcloud1
3. A.N. Davies, M. Cashyap, R. Lancashire and R.M. Hanson, "A head in the clouds?—Part Two: exploring distributed, multi-server ^{1}H NMR prediction", *Spectroscopy Europe* **26(5),** 16–17 (2014). http://bit.ly/tdcloud2
4. A.N. Davies, "Online services for your desktop on a pay-as-you-surf principle—ACD/Labs I-Lab", *Spectroscopy Europe* **13(1),** 22–28 (2001). http://bit.ly/1DJtJzw
5. www.labcognition.com
6. A.N. Davies, "More reference spectroscopic data on the Internet! FTIRsearch Nicolet/Galactic Industries new joint venture", *Spectroscopy Europe* **12(5),** 30–36 (2000). http://bit.ly/ftirsearch
7. D. Soulsby, "Using cloud storage for NMR data distribution", *J. Chem. Educ.* **89,** 1007–1011 (2012). doi: http://dx.doi.org/10.1021/ed2005906
8. http://www.bio-rad.com/en-us/product/spectroscopy-software/knowitall-enterprise-server?pcp_loc=catprod
9. http://onlinelibrary.wiley.com/book/10.1002/9780471221692
10. http://nmrpredict.orc.univie.ac.at/c13robot/robot.php