

“Halfway up the Stairs”

Antony N. Davies

ISAS, Institute für Spektrochemie und Angewandte Spektroskopie, Postfach 10 13 52, 44013 Dortmund, Germany.

*“Halfway down the stairs
is a stair where I sit.
There isn’t any other stair
Quite like it.
I’m not at the bottom,
I’m not at the top;
So this is the stair
Where I always stop.”*
A.A. Milne, *When We Were Very
Young*, Nov. 6th 1924.

Introduction

As we write the vendors are packing up and the delegates are climbing into their cars for the journey home from the conference on Linking and Interpreting Spectra through Molecular Structures. Over the last three days we had witnessed a very successful conference covering all aspects of spectroscopic and structure data handling.... but for a full write-up you will have to wait for the proceedings.

What we are going to present in this column is the results obtained from the “Warwick Challenge”. All vendors, the instrument manufacturers and software

Exhibitor	Formats	Hardware/Networking
HP	NetCDF	HP-UX Series D HPX Server
Bio-Rad	JCAMP-DX	Pentium 90 NT 4.0, TCP/IP PCMCIA Ethernet
Labware	HPGL/SDF/TGF/ MOL	Pentium and 486, Windows 95
Waters	NetCDF	4004/233 Dec Alpha Open VMS serving 2 Windows for Workgroups clients, “Millennium” Software
Synopsys	SDF/MOL/Smiles/ any graphics	PC Based not networked
Sun/OMF	SGML/MOL/ JCAMP-DX/ NetCDF/PDB/ Swiss Prot	Sun Ultra workstations running Solaris 2.5 on SunOS 5.5.1, Samba to serve Windows 95 and NT, NFS for UNIX
E-Scribe	HPGL	Pentium133, Windows NT 3.51, Windows 3.1 TCP/IP, Ethernet
LabSystems Chemical Concepts	NetCDF JCAMP-DX/MOL	Windows NT 3.51 Silicon Graphics serving ISIS Base interface on PC
Cherwell Scientific MDL Daresbury	Molfiles/SDF	X-Windows link over SuperJANET to CCLRC Daresbury (Sun)

Figure 1. A partial list of exhibitors at the LISMS conference, the data exchange formats they wished to exchange, and an attempt to define the computer systems they were running.

houses who participated in the exhibition were told in their instructions for the conference:

Participants in the exhibition agree to the following.

- Demonstrate their software not the instrument.
- Bring sample data.
- Connect to the local network and demonstrate exchange of information.
- Show what they can add to the objectives of the meeting.
- Exhibit conformance to standards and highlight inadequacies in the standards.
- Allow reasonable access to their systems by the delegates.

So we hoped they would all come armed and ready to show their system inter-operability compliance to open systems approaches and their keen interest in the use of international data exchange standards.

Networking

Sun were exhibiting with a number of systems and were kind enough to donate the central data archive server which was used as a depository for the data to be exchanged. Each company then had to find a method to reach the server and this was achieved only thanks to an enormous effort by Richard Kinder of Sun whose knowledge of everyone else's hardware turned out to be a key feature of the Challenge!

Figure 1 shows the hardware and some information as to the systems running from some of the exhibitors. For those I have missed out or mis-assigned, apologies for not getting the details down in time, things were rather hectic!!

Each vendor was supplied with a twisted pair connection to the Warwick University network as well as an IP address which they would have to use for the period of the conference.

Once the usual questions of clarifying the access privileges were sorted out, most exhibitors were able to link to the depository. One of the interesting aspects was the ease with which most of the UNIX world inter-operated virtually from the word go, and the PC world after some tinkering. FTP was the first service available to move data from the exhibitor systems into the depository. Once this had been shown to work several exhibitors moved on to mount the depository as a NFS drive onto their own systems.

Unfortunately the overlap of exhibitors' field of work was quite

small being divided into chromatography system vendors, spectroscopic data system vendors and LIMS type systems. This meant that the type of data which they could be expected to read or write was also split into these fields.

Results

Figure 2 is a table of exhibitors and their success or lack of it in transferring data to the depository and then a cross-reference of those exhibitors whose systems managed to read the data files from other exhibitors.

Unfortunately for Labware their technical director Mark Gonzalez was taken ill and so they were not able to take part in the exchange, although he had managed to prove the network links to the other PC servers. Best wishes and a quick recovery!

Thanks must go to David Fletcher from the UK Chemical Database Service. He agreed at very short notice to sit in front of an X-terminal window and try to import the data that was being collected into the SpecInfo and REACCS packages in Daresbury.

Exhibitor	Formats	Data donated	Read data from:
HP	NetCDF	Yes	Waters, LabSystems
Bio-Rad	JCAMP-DX	Yes	Bio-Rad, Chemical Concepts
Labware	HPGL/SDF/ TGF/MOL	Technical director	taken ill
Waters	NetCDF	Yes	
Synopsys	SDF/any graphics		Chemical Concepts; Cherwell Scientific; MDL
Sun/OMF	SGML/MOL/ JCAMP-DX/ NetCDF/PDB/ Swiss Prot	Yes	
EScribe	HPGL, MOL, Smiles		Labware; Waters
LabSystems	NetCDF	Yes	
Chemical Concepts	JCAMP-DX/ MOL	Yes	
Cherwell Scientific			
MDL	Molfiles/SDF	Yes	
Daresbury (SpecInfo)		N/A	Bio-Rad (5/12)
Daresbury (REACCS)		N/A	MDL

Figure 2. This table shows the exhibitors who donated data and those who managed to read this data out of the depository.

One result which has bearing on the work towards a Chemical MIME standard is the difficulty in recognising the JCAMP-DX spectroscopy type from the file extension. Infrared, mass spectrometry and nuclear magnetic resonance JCAMP files are different in content. This caused problems when a software package, while looking for an infrared file, was in fact reading 2D-NMR data sets. Surprisingly, there was also some problems with carriage-return/line field which we had thought solved.

Conclusions

In parts the table looks at first somewhat bare! However, it should be remembered that the LIMS and lab notebook exhibitors usually have active links from scientific data file names to other programs. These programs would be supplied by other vendors and LIMS packages do not normally attempt to import chromatographic or spectroscopic data.

What was pleasing was to see the netCDF chromatography standard being exchanged without problem between vendors in a real-life situation bearing out the claims made several years ago for this new standard format.

This test cannot be regarded as in any way conclusive. However, the data is planned to be made available on our web site for you to try yourself.

The field of data exchange and open systems are still "halfway up the stairs", we can be confident that we are climb-

ing upwards, but there is still a long way to go before we can put the problem to bed!

Contacting LISMS

Data from the conference can be found at the following web site:

<http://www.isas-dortmund.de/lisms>



The Proceedings will be published by IM Publications in October. Further details are available from IM Publications, 6 Charlton Mill, Charlton, Chichester, West Sussex PO18 0HY, UK, Tel: +44(0)1243-811334, Fax: +44 (0)1243-811711, E-mail: subs@impub.demon.co.uk.

Showing resolve

During the conference it was clear that a political push was needed to be made to make the European Commission see the need for an analytical reference data archive. As discussions are underway as to what priorities should be set for the 5th Framework a proposal was overwhelmingly accepted worded as follows.

“The Need for an Analytical Reference Data Archive: A Resolution”

3 September 1996

The chemical, pharmaceutical and materials industries are a major economic force and job provider in Europe. Keeping research and development abreast of the rest of the world is important to the scientific and economic success of Europe.

Confirming and elucidating chemical structures are major tasks in the discovery and development of new products, in quality control and in environmental analyses. There are currently in excess of 14,000,000 registered chemical compounds, and more than 500,000 new ones are added each year. Although analytical data (from separation science and spectroscopic methods) are used during the synthesis, purification and identification of all of these compounds, few of the data are available, with the chemical structures, in a form useful to the academic or industrial analytical community.

The largest electronic collections of analytical data represent 1% or less of the known chemical structures. It is estimated that as many spectra are recorded in industrial and academic laboratories in a single day as are contained in the largest electronic analytical databases. Nearly all of these spectra are discarded or are unavailable, even to those who acquired them.

Access to large electronically stored collections of spectroscopic and separations data stimulates significant progress in chemical research and in automated methods for structure/spectrum and structure/biological-activity correlation. This has wide implications for human health, new materials, environmental protection, sustainable development and educational progress.

Combinatorial chemistry is a major advance for discovering new materials and new chemical compounds for human health, crop protection or other uses. Rapid methods for confirming chemical identity depend critically on access to large analytical data sets. Time and money are often spent duplicating analyses of known compounds simply because archival data are not available. The efficiencies gained by enabling access to analytical data archives will contribute to maintaining competitive European industrial and academic research and development.

It is unlikely that any single company or institute could take on the effort of building such an electronic repository. It is more appropriate that the

initial funding stimulus for this project come from international public sources. Eventually the repository would become self-funding through fees for access to the data. However, without EU support, the project will not begin or achieve enough momentum to sustain itself.

Because such a collection of analytical data would be an important European scientific resource, members of the European analytical chemistry

community strongly encourage the European Union to include an electronic analytical data repository as a priority area in the forthcoming Fifth Programme.

If readers can support the ideas of this Resolution, they are requested to get it signed by a senior member of their company or institution (Company Director or Department Head) and by any societies that you belong to.