

A new NMR data standard for the exchange and archiving for multi-dimensional data sets

Antony N. Davies

ISAS, Institut für Spektrochemie und Angewandte Spektroskopie, Bunsen-Kirchhoff-Str.11, Postfach 10 13 52, 44013 Dortmund, Germany.

With multi-dimensional spectroscopic techniques playing an ever increasing role in the life of an NMR spectroscopist it didn't take long for the need for a new NMR data exchange standard for multi-dimensional data sets to arise. This requirement was formally established last year at the Experimental Nuclear Magnetic Resonance Spectroscopy Conference (ENC) in Asilomar, USA. An international task group was set up to develop this data exchange standard for multi-dimensional NMR experiments and is currently working under the auspices of the International Union of Pure and Applied Chemists (IUPAC). This article describes how the process began, how the need for a new data exchange standard was established, and when we can expect to be able to make use of implementations in our software updates.

Introduction

In 1997 a discussion session at the Experimental Nuclear Magnetic Resonance Conference (ENC) in the USA raised the question of exchange and long-term storage of multi-dimensional data (see Figure 1). The ensuing discussion revealed that there was a need but also a severe lack of information on the subject so a discussion meeting was organised for the following ENC held last year at Asilomar entitled "A Standard Format for NMR Spectroscopy".

Following a presentation covering the current state-of-the-art a lengthy discussion was held as to the desirability of a new standard and the best way forward. It was agreed that a new internationally recognised exchange standard was urgently required. A working group was established and many delegates volunteered to help design and build the new standard—including key people representing the major NMR manufacturers, industrial and academic NMR users and representatives of

<u>Data Handling</u>	<u>1-D NMR</u>	<u>n-D NMR</u>
Short Term (original manufacturer format)	YES	YES
Long Term Archive (JCAMP-DX)	YES	NO
Free (JCAMP-DX)	YES	NO

Figure 1. Although solutions are available for short and long term NMR data storage, exchange and long term archiving for conventional NMR data there is currently nothing available for n-D NMR data.

independent software houses. In addition, following the ENC a number of experts active in this field have been contacted who responded positively to the proposal and have agreed to work on the Task Group.

Proposal to IUPAC

Following the discussion meeting a formal proposal was made for IUPAC to organise a Limited Term Task Group to develop, document and validate a new spectroscopic data standard for multi-dimensional NMR data sets. The proposal was presented to the IUPAC Standing Committee on Printed and Electronic Publications¹ in June and the project agreed.

This task group is limited to the lifetime of this project and function as proposed in the "Changes in Organisation and Management of IUPAC Scientific Activities" from the IUPAC Strategy Development and Implementation Committee.²

The Task Group would be co-ordinated through the current CPEP Working Party on Spectroscopic Data Standards (JCAMP-DX),³ with the finished documentation being laid before the CPEP for approval before being published in *Pure and Applied Chemistry*.

Work plan

The key to the success of this project is the production of a generic (manufacturer independent) data dictionary describing a multi-dimensional NMR experiment. It is envisaged that most of the initial work developing the data dictionary will take place via the Internet within the first six months. However, experience has shown that when developing these standards at least one round-table discussion by the key participants is necessary to iron out difficulties of principle and to reach compromises between different points of view. The use of email does not lend itself to arbitrating over difficult semantic differences!

During development all documentation will be available freely over the Internet to registered participants, testers and members of the task group. It will be made clear that these are draft documents only and are not to be cited until officially published.

A progress review will take place at the end of February 1999 before the ENC'99 in Orlando, USA.

An initial report including the draft data dictionary will be made available by, and presented to the ENC'99 in March 1999, and the IUPAC CPEP meeting during the 37th IUPAC General Assembly in Berlin in August 1999.

With the finalisation of the data dictionary test files will be produced to aid development work by the manufacturers and independent software houses.

Development work will be supported throughout 1999 including Round Robin testing of the new files.

It is planned to have the protocol ready for publication in *Pure and Applied Chemistry* and software implementations available in releases by the manufacturers the following year. Finalised implementation documentation should also be made available during this phase.

A final report should be made to ENC'2000 in March 2000.

The new data dictionary

The current stage—and probably by far the hardest—is to define the information content required/desired within the NMR data file. To do this the data dictionary must be generic and not manufacturer specific but at the same time contain enough information to fully annotate the experimental data.

By the middle of July a web site was set up to gather suggestions for the data dictionary.⁴

It was pointed out to the task group that they should first make themselves acquainted with the structure of the protocols already agreed and published⁵ to see how the records are built up. Only then would it be advisable to proceed to the data entry web pages where they should register before proceeding to the data dictionary entry pages.

We will gladly take suggestions from anyone with experience in this field so if you wish to participate in this action follow the instructions above. When registering please allow the cookie to be set as this will mean you only have to register once. All further attempts to access these pages will then go straight to the data entry pages.

If you do not have access to the internet then we would be grateful to receive your proposals by returning the form below either by post or fax (+49 231 1392 418).

Geography

Although the Task Group is geographically localised to North America, Europe and Japan this seems to be unavoidable due to the location of the manufacturers and independent software houses. IUPAC is, however, very keen to ensure global access to its activities so the Internet will be used to advertise the presence of the draft documents for comment to obtain a geographical broad review of the proposals before being presented to the CPEP for the ultimate stamp of approval.

Conclusion

Hopefully the activities reported here will quickly return results of benefit to the whole spectroscopic community. The work put into developing these standards is voluntary and a tremendous vote of thanks must go out to those who have devoted so much time and energy into these projects in the last decade. Although such work will probably never win Nobel prizes, in our computer dominated working environment maintaining the availability and accessibility of data might well be the key to many of the more successful industrial and research efforts in the future.

Nowadays we cannot afford to let knowledge die either with the retire-

continued on page 20

n-D NMR Registration			
Name			
Organisation Type	Instrument Manufacturer	<input type="checkbox"/> Independent	<input type="checkbox"/> Academic
Organisation			
Address			
Country			
Tel.:			
Fax:			
Email			
Request for records in the new n-D NMR Protocol			
Remember - the new protocol will be generic and not manufacturer specific. Only fields which comply to this criteria will survive the review process.			
All records require the following elements for the protocol:			
NAME TYPE [OPTIONS] DESCRIPTION			
Please enter the NAME of your requested record. This field should be short but descriptive. e.g. PULSSEQUENCE		NAME and	
Please select one of the following data types. Your field may be one of either	TEXT for free formatted comments etc. <input type="checkbox"/>	AFFN for free formatted numerical values <input type="checkbox"/>	STRING if only options from a pre-defined list are to be allowed <input type="checkbox"/>
If You have chosen STRING. You must now enter the list of possible options. Remember to include definitions for all options in the detailed description field below.			
STRING OPTIONS:			
Finally please provide a detailed description of what purpose this field serves! -how should this data be handled by software?			
should this field be optional or required why?		REQUIRED	<input type="checkbox"/>
		OPTIONAL	<input type="checkbox"/>
WHY?			

continued from page 16

ment of key workers or the phasing-out of a particular computer technology.

Acknowledgements

I would like to thank the Joint Committee on Atomic and Molecular Physical Data and Bruker for financial support of the IUPAC CPEP Working Party on Spectroscopic Data Standards (JCAMP-DX) in 1998. I would also like to thank the Bundesministerium für Bildung Wissenschaft, Forschung

und Technologie, and the Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes Nordrhein-Westfalen, Germany for financial support.

Finally, I have to give a big round of applause to the IUPAC Working Party members who have wrestled for so long with these problems:

Peter Lampen, Robert Lancashire, Bob McDonald, Peter McIntyre and Doug Rutledge, not forgetting Steve Heller at NIST for his unflagging support and encouragement.

References

1. <http://www.iupac.org/standing/cpep.html>.
2. <http://www.iupac.org/news/archives/1998/october/index.html>.
3. http://www.iupac.org/standing/cpep/wp_jcamp_dx.html.
4. <http://www.isas-dortmund.de/projects/jcamp/ndnmr.html>.
5. <http://www.isas-dortmund.de/projects/jcamp/>.