INTERNATIONAL UNION OF PURE AND

APPLIED CHEMISTRY

COMMITTEE ON PRINTED AND ELECTRONIC PUBLICATIONS

WORKING PARTY ON SPECTROSCOPIC DATA STANDARDS (JCAMP-DX)

AN EXTENSION TO THE JCAMP-DX STANDARD FILE FORMAT, JCAMP-DX V.5.01

(IUPAC Recommendations 1999)

Prepared for publication by:

PETER LAMPEN[1], JÖRG LAMBERT[1], ROBERT J. LANCASHIRE[2],

ROBERT S. McDONALD[3], PETER S. McINTYRE[4], DOUGLAS N. RUTLEDGE[5],

THORSTEN FRÖHLICH[6] AND ANTONY N. DAVIES[1]*

[1]ISAS, Inst. für Spektrochemie und Angewandte Spektroskopie, Bunsen-Kirchhoff-Str. 11, 44139 Dortmund, Germany
[2]Department of Chemistry, University of the West Indies, Mona, Kingston 7, Jamaica
[3]9 Woodside Dr., Burnt Hills, New York 12027, USA
[4]School of Applied Sciences, University of Glamorgan, Pontypridd, Mid-Glamorgan, CF37 1DL, UK
[5]Inst. National Agronomique, Lab. de Chimie Analytique, 16, rue Claude Bernard, 75231 Paris Cedex 05, France
[6]LabControl GmbH, Max-Planck-Str.l7a, 50858 Koln, Germany

*Author to whom comments should be sent.

An extension to the JCAMP-DX standard file format, JCAMP-DX V.5.01 (IUPAC Recommendations 1999)

Abstract: Version 5.00 of the JCAMP-DX specifications were published for NMR and Mass Spectrometry file formats in *Appl. Spectrosc.* 47, 1093-1099 (1993) and *Appl. Spectrosc,* **48,** 1545-1552 (1994). Since publication of these protocols developments in spectroscopy have led to a large number of requests for additions for applications not originally covered. Following careful consideration, it has become apparent that a few minor modifications will significantly increase the range of possible applications.

In addition, new data labels have been introduced to ensure that files are year 2000 compliant and allow for conformity with good laboratory practices (GLP). These modifications are detailed in this publication as well as examples of the official NTUPLE JCAMP-DX definition as applied to NMR data.

1 INTRODUCTION

The transfer of data between computers of different origin, running differing operating systems and application software has been greatly enhanced by the introduction of openly agreed standards. In the field of spectroscopy, the publication and successful implementation of the JCAMP-DX (Joint Committee on Atomic and Molecular Physical Data—Data Exchange) protocols have made it possible to move infrared and Raman spectroscopy [I], chemical structure [2], nuclear magnetic resonance spectroscopy [3], and mass spectrometry [4] data sets.

The JCAMP-DX protocols are based on delivering scientific data using ASCII encoding to ensure maximum compatibility and longevity of the encoded data. Depending on the encoding software, it is possible to produce data exchange files with absolutely no loss of data. Built-in compression algorithms can often yield ASCII exchange files that are smaller than the original binary spectrometer files, again with no loss of information. More recent interest in the long term, safe storage and archiving of original experimental data sets in electronic form has delivered a new dimension to the use of independent international standards.

2 SCOPE

2.1 JCAMP-DX Version 5.01

These specifications define additional content for global notes and NMR datatype specific records. Files being written to include these options should carry the JCAMP-DX version number 5.01. (##JCAMP-DX=5.01 see examples). Additions are made to cover the use of the IUPAC name in JCAMP-DX files, to include a definition for storing dates in a year 2000 compliant fashion, as well as a specification for storing audit trail information. For NMR additional technical data specifications handling pulse sequences, axes labeling and solid state magic angle spinning information are included. For a particular sample, additional data can now be stored on concentration and solvent parameters, including either internal or external referencing.

The new definitions must be used in conjunction with the original protocol and are best studied after gaining a full understanding of the file format from the available literature.

2.2 Copyright

This publication was drawn up by the International Union of Pure and Applied Chemistry (IUPAC), Standing Committee on Printed and Electronic Publications (CPEP), Working Party on Spectroscopic Data Standards (JCAMP-DX) following extensive consultation with users and manufacturers. As with all JCAMP-DX protocols they are meant for public use. The specifications are copyrighted by IUPAC for the purpose of linking them to the name JCAMP-DX. The right to copy these specifications for scientific use is hereby granted.

2.3 Use of name

The use of the name JCAMP-DX in the description of data files and software implies the capability of converting between internal data files and JCAMP-DX according to the various published protocols.

3 GENERIC LABELED DATA RECORDS – GLOBAL NOTES

3.1 What are generic labeled data records?

The term generic labeled data records (LDRs) has been introduced to cover those fields which are used in a similar way and have similar content regardless of the type of data stored in the JCAMP-DX data file. These records are to be differentiated from data type specific labeled data records whose content is defined by the particular type of data being stored in the file. In this way new JCAMP-DX protocols will be much easier to develop with the developers concentrating on the task of drawing up a data type specific data dictionary and being able to draw on the generic data dictionary for all of the supporting information.

3.2 IUPAC naming conventions

The previous JCAMP-DX protocols have been able to store substance naming information only in Chemical Abstracts Service nomenclature. In order to save nomenclature in a IUPAC compliant form a new label needs to be defined.

##IUPAC NAME=(TEXT)
The use of IUPAC names has been recommended previously by the Commission on Molecular Structure and Spectroscopy of the Physical Chemistry Division in 1991 [5].

Priority OPTIONAL

3.3 Year 2000 compliance

The date format in JCAMP-DX needs to be addressed, as in its current form it is not year 2000 compliant.

DATE in previous versions of JCAMP-DX represented the year as YY. The new label uses YYYY for the year. Time, fractional seconds and the time zone can be optionally included as follows:

##LONG DATE=(STRING)
Date when the spectrum was measured in the form:
##LONG DATE=YYYY/MM/DD HH:MM:SS.SSSS ±UUUU where

   YYYY is the year

   MM is the month

   DD is the day

   HH:MM are the hours and minutes in local time (optional)

   SS.SSSS are the seconds and fractional seconds (optional)

   UUUU is time zone difference to Universal Coordinated Time (UTC) (optional)

For example:
##LONG DATE=1998/08/12 23:18:02.0000-0500
##LONG DATE= 1988/08/12 Priority OPTIONAL.

3.4 Audit trail

To enable an audit trail, a mechanism to record any changes to the data file is introduced. Separate lines cater for recording the person who made the change and authorization. Parentheses indicate the start and end of each audit trail entry. Each entry must start on a new line.

##AUDIT TRAIL= (AFFN, STRING, TEXT, TEXT, TEXT)

where

AFFN (ASCII Free Format Numeric, see [1]) is the change number

STRING is a longdate

The first TEXT field is the person who made or authorized the change

The second TEXT field is the person's location

The final TEXT field gives details of the change made (and should indicate whether the change is reversible). In the case of authorizations, this field represents the change number that is approved.

The date string and all three text fields should be enclosed in angle brackets.

Example

##AUDIT TRAIL= $$ (NUMBER, WHEN, WHO, WHERE, WHAT)

(1, <1998/09/15 12:00 +0000>, <Joe Bloggs>, <London, UK>, <baseline correction, not reversible>)

(2, <1998/09/16 07:00 -0500>, <Jane Doe>, <New York, USA>, <authorized 1>)

Priority OPTIONAL.

4 NEW DEFINITIONS AND ADVISED USAGE FOR NMR SPECTROSCOPY

4.1 JCAMP-DX for nuclear magnetic resonance spectroscopy

Co-ordinated work on the first JCAMP-DX protocol for NMR began following the sixth INUM conference in Italy in 1990. During this meeting the major NMR manufacturers agreed to implement the protocol when finished. The protocol was published after several years work in 1993 [3] and is available online from: http://www.isas-dortmund.de/projects/jcamp/

This protocol was specifically designed to handle conventional NMR spectra and free induction decays. A IUPAC limited term task group has been established to produce a protocol to handle multidimensional data sets reporting to the CPEP [6].

Within the first NMR protocol [3], 20 NMR specific labels were defined and will not be repeated here. Additionally, rules were laid down as to the use in NMR data sets of some 26 generic labels available from previous JCAMP-DX protocols.

As detailed above a few minor additions and modifications will significantly increase the range of possible applications of this NMR protocol. This opportunity has also been used to bring the standard into compliance with the IUPAC recommendations on Parameters and Symbols for Use in Nuclear Magnetic Resonance [7]. These changes are detailed below and deal with either additional information on parameters relevant to the measurement of NMR data sets or to the composition of the sample being measured.

4.2 Y-axis labeling

In order to store transformed NMR data it is essential to know whether the data have been transformed to magnitude or power values. To cover this possibility, the labeled data record ##YUNITS= in the core data section of [3] has been modified.

The modified entry to the specification will be:

##YUNITS= (STRING)

Specifies the ordinate units for the data table. The allowed STRING values are:

POWER   for $(R^2 + 1^2)$
MAGNITUDE   for $((R^2 + I^2)^{0.5})$
ARBITRARY UNITS
where R is the real part and I is the imaginary part of the ordinate values.

Priority REQUIRED.

4.3 Use of NMR relevant sample concentration information

Analyte chemical shifts are commonly solvent dependent in NMR spectroscopy. For this reason, the LDR ##SAMPLE DESCRIPTION= in the global notes section may also contain information on the concentration of the solute, if relevant.

This entry to the specification should then be:

##SAMPLE DESCRIPTION= (TEXT)

If the sample is not a pure compound, this field should contain its description, i.e. compositions, origin, appearance, results of interpretation, etc. If the sample is a known compound, the sample information LDRs specify structure and properties, as appropriate (see *Appl. Spectrosc.* 42, 151-162 (1988) section 7.2). For NMR, this LDR may contain information concerning the concentration of the solute.

Priority OPTIONAL.

4.4 Pulse sequences

Unlike most other spectroscopic techniques, the signals observed in NMR are dependent upon the particular pulse sequence used. Therefore, an NMR spectrum, FID or relaxation signal has little or no meaning if the pulse sequence is not given. For this reason, a new LDR has been created to contain this information as an addition to the data type specific notes section of [3].

A list of the most common NMR pulse sequences and guidelines on the information to correctly report them will be available shortly in *Pure and Applied Chemistry* [8].

The new entry to the specification will be:

##.PULSE SEQUENCE= (TEXT)

This LDR will contain all pertinent information concerning pulse sequences, pulse delays, relaxation delays, pulse angles, etc. Whenever possible, this LDR should include the name of the pulse sequence. The text in this field should preferentially be based on the standard reporting format for NMR pulse sequences [8], or may be written in the instrument's native macro language provided enough information is supplied to clearly enable post processing of the data by third parties. The LDR ##SAMPLING PROCEDURE= will no longer be used to contain this information.

Example:

##.PULSE SEQUENCE= $$ Expressed in the instrument macro language

  EXMOD VHMQC

  EXPCM <SV>[HC]VHCDQC:W1(H9),W2(H18),W3(C9),I3(1/2J):V=I2

  PW1 12.5 us

  PD l.000 s

or from the list of standard 2D-NMR pulse sequences [8]:

##.PULSE SEQUENCE= Pulse Acquisition $$ Taken from the IUPAC list

  p1= 12.5 $$ transmitter high power pulse in microseconds
  rd= 1000 $$ relaxation delay in milliseconds Priority

OPTIONAL.

4.5 Solvent effects, pH, etc.

As solvent effects like dipolar or charge transfer interactions, etc. may significantly modify chemical shifts, the solvent should be indicated, if required to correctly interpret the chemical shifts. A new OPTIONAL LDR has therefore been created in the data type specific notes section to contain this information.

The new entry to the specification will be:

##.SOLVENT NAME= (TEXT)
Description of the solvent. In the case of aqueous solutions, pH, ionic strength, etc. should be included, if relevant.

Priority OPTIONAL.

4.6 Chemical shift referencing

Where a solvent reference signal is used instead of TMS then the LDR ##.SOLVENT REFERENCE= is required. However, the name of the reference substance should also be given. It is important to know the source of the referencing as this allows an additional level of quality control. Additionally, in certain types of NMR spectroscopy such as ^{31}P NMR there is still no unified p.p.m. scale based on a single referencing substance. It should be noted that by convention NMR spectra are stored in Hz but displayed in p.p.m.

The new entry in the data type specific notes section of the specification will be:
##.SHIFT REFERENCE= (STRING, TEXT, AFFN, AFFN)

STRING can be either INTERNAL or EXTERNAL

TEXT is the name of the chemical shift reference compound.

The first AFFN gives the current data point number (starting from the number 1 for the first data point) of the reference compound signal.

The second AFFN is the shift value of the reference compound signal in p.p.m.

e.g. ##.SHIFT REFERENCE= INTERNAL, CDCL3, 10480, 77.0

Priority OPTIONAL.

To construct a p.p.m. scaled display, data in ##.SHIFT REFERENCE= and ##.OBSERVE FREQUENCY= can be used.

The label ##.SOLVENT REFERENCE= was declared as REQUIRED in version 5.00. To maintain backwards compatibility, this will remain but should contain the same value (in p.p.m.) as that given in ##.SHIFT REFERENCE=.

4.7 MAS spinning frequency

In order to cover solid-state NMR spectroscopic techniques, it is necessary to include information on the magic angle spinning frequency. To do this, a new OPTIONAL LDR has been created in the data type specific notes section.

This new entry to the specification will be:

##.MAS FREQUENCY= (AFFN)
Rate of sample spinning at the magic angle in Hz.
Priority OPTIONAL.

5 CONCLUSIONS AND FUTURE WORK

These new definitions extend the original version 5.00 protocol to cover some general concerns and recent developments in NMR techniques. However, one development in spectroscopy is the use of multidimensional data measurements. Although it is possible to code multidimensional data sets using the original protocols, no definition has been published or tested for this purpose. The use of third party software and the move to off-spectrometer data processing makes the need for such a protocol very acute.

As mentioned above a IUPAC limited term task force including all of the major NMR manufacturers and independent software houses has been established [6]. Their goal is to develop and validate a multidimensional NMR data standard within the next two years. Those wishing to help in this activity can register at: http://www.isas-dortmund.de/projects/jcamp/

6 EXAMPLES Example 1: Simple NMR FID

```
##TITLE=            ETHYLBENZENE
##]CAMP-DX=         5.01 $$ ISAS JCAMP-DX for NMR program (Vl.l)
##DATA TYPE=        NMR FID
##DATA CLASS=       NTUPLES
##ORIGIN=           Dr J. Lambert, ISAS Dortmund, Germany
##OWNER=            Copyright (C) 1992 by ISAS Dortmund, Germany
##LONG DATE=        1992/09/14 18:34:48+0100
##SPECTROMETER/DATA SYSTEM=   JEOL GX 400
##.OBSERVE FREQUENCY=         100.4
##.OBSERVE NUCLEUS=           ^13C
##.DELAY=                     (22.5, 22.5)
##.ACQUISITION MODE=          SIMULTANEOUS
##.PULSE SEQUENCE= Pulse Acquisition DEC     $$ This is from the IUPAC list [8]
s 1 = 0                                      $$ power attenuation level for radio frequency pulse in dB
s3= 14                                       $$ decoupler power attenuation level in dB
pl= 5.0                                      $$ transmitter high power pulse duration in microseconds
rd= 3000                                     $$ relaxation delay in milliseconds
##.SOLVENT NAME=     CDC13
##NTUPLES=           NMR FID
##VAR_NAME=   TIME,         FID/REAL,        FID/IMAG,        PAGE NUMBER
##SYMBOL=     X,            R,               I,               N
##VAR_TYPE=   INDEPENDENT,  DEPENDENT,       DEPENDENT,       PAGE
##VAR_FORM=   AFFN,         ASDF,            ASDF,            AFFN
##VAR_DIM=    16384,        16384,           16384,           2
##UNITS=      SECONDS,      ARBITRARY UNITS,ARBITRARY UNITS,
##FIRST=      0.0,          2980,            6216,            1
##LAST=       0.682,        -60242,          -6064,           2
##MIN=        0.0,          -170402,         -165286,         1
##MAX=        0.682,        149235,          161918,          2
##FACTOR=     0.4162852E-04, 0.5200415E+01,  0.5044282E+01,   1
##PAGE= N= 1
##DATA TABLE=(X++(R.R)), XYDATA $$ Real data points
0E73J0801K743n991jl702J2907o922p342r868p9141924J3049K1127R280J0543p471j2800
………………
16383al584
##PAGE=N=2
##DATA TABLE= (X++(I.I)), XYDATA  $$ Imaginary data points
0A232J9544101901994j8571p88K0538J355J6653J718N398o027N305L183j909J1943k4401
………………
16383a202
##END NTUPLES=NMR FID
##END=
```

Example 2: Simple NMR spectrum

```
##TITLE=              ETHYLBENZENE
##]CAMP-DX=           5.01 $$ ISAS JCAMP-DX for NMR program (Vl.l)
##DATA TYPE=          NMR SPECTRUM
##DATA CLASS=         XYDATA
##ORIGIN=             Dr J. Lambert, ISAS Dortmund, Germany
##OWNER=              Copyright (C) 1992 by ISAS Dortmund, Germany
##LONG DATE=          1992/09/14 18:34:48 + 0100
##SPECTROMETER/DATA SYSTEM=    JEOL GX 400
##.OBSERVE FREQUENCY=          100.4
##.OBSERVE NUCLEUS=            ^13C
##.PULSE SEQUENCE=                      $$ This is in the instrument macro language
EXMOD  SGBCM
EXPCM  'B'SGBCMA:Single.pulse.Complete.decoupling.IRRPW.IRATN
IRATN   14
PW1     5.0 us
PD      3.000 s
##.SOLVENT REFERENCE=  77.0
##.SOLVENT NAME=       CDCL3
##.SHIFT REFERENCE=    INTERNAL, CDCL3, 10480, 77.0
##DATA PROCESSING=     Ordinates are re-scaled between -32767 and +32767
##RESOLUTION=          1.470
##XUNITS=              HZ
##YUNITS=              ARBITRARY UNITS
##XFACTOR=    1.467283
##YFACTOR=    29670.15003
##FIRSTX=     24038.50
##LASTX=      0.0E+00
##MAXY=       972201806
##MINY=       -27593240
##NPOINTS=    16384
##FIRSTY=     2254931
##XYDATA=     (X++(Y..Y))
16383G6k53o5J88M13kl2k70J95K00kl3O1q3k25J80k9j20MQ6jl6n7J70J57NL8q9NlM0k95
16356a63J98K3J59m73m8J80J10Mlk28k55J83J45r2K59R5k82p8j9L50jl2QM7j82K30117
………….
………….
30b44k49L53J02Q7QK8j24kl6J86k02J79J87m3K21k34k20L17j32J25N8117J61Nj86J26R4
4G0J71J43kl5jl8
0E1
##END=
```

REFERENCES

1 R. S. McDonald, P. A. Wilks, Jr. *Appl. Spectrosc.* **42,** 151-162 (1988).

2 J. Gasteiger, B. M. P. Hendriks, P. Hoever, C. Jochum, H. Somberg. *Appl. Spectrosc.* **45,** 4-11 (1991).

3 A. N. Davies, P. Lampen. *Appl. Spectrosc.* 47, 1093-1099 (1993).

4 P. Lampen, H. Hillig, A. N. Davies, M. Linscheid. *Appl. Spectrosc.* **48,** 1545-1552 (1994).

5 JCAMP-DX. A standard format for the exchange of infrared spectra in computer readable format (IUPAC Recommendations 1991). *Pure Appl. Chem.* **63,** 1781-1792 (1991).

6 A. N. Davies. *Spectrosc. Eur.* 11(1), 14-20 (1999) (see also: http://www. spectroscopy europe.com/td_col.html).

7 R. K. Harris, J. Kowalewski, S. Cabral de Menezes. *Pure Appl. Chem.* **69,** 2489-2495 (1997).

8 A. N. Davies, J. Lambert, R. J. Lancashire, P. Lampen, W. Conover, M. Frey, M. Grzonka, D. Iverson, D. Meinhart. Guidelines for the representation of pulse sequences for solution-state nuclear magnetic resonance spectrometry (IUPAC Recommendations 1999). Draft for publication in *Pure Appl. Chem.* (1999).