INTERNATIONAL UNION OF PURE
AND APPLIED CHEMISTRY

COMMITTEE ON PRINTED AND
ELECTRONIC PUBLICATIONS

Subcommittee on Electronic Data Standards

JCAMP-DX V.6.00 for CHROMATOGRAPHY and MASS SPECTROMETRY HYPHENATED METHODS

(IUPAC Technical Note 2005)

Prepared for publication by
J. Hau[1], P. Lampen[2], R.J. Lancashire[3], R.S. McDonald[4],
P.S. McIntyre[5], D.N. Rutledge[6], W. Schrader[7], A.N. Davies[8]

[1] Nestlé Research Center, Nestec Ltd, Vers-chez-les-blanc, Case Postale 44, CH-1000 Lausanne 26, Switzerland
[2] ISAS, Inst. für Spektrochemie und Angewandte Spektroskopie, Bunsen-Kirchhoff-Str.11, 44139 Dortmund, Germany
[3] Department of Chemistry, University of the West Indies, Mona, Kingston 7, Jamaica
[4] 9 Woodside Dr., Burnt Hills, New York 12027, United States of America
[5] School of Applied Sciences, University of Glamorgan, Pontypridd, Mid-Glamorgan, CF37 1DL, United Kingdom
[6] Inst. National Agronomique, Lab. de Chimie Analytique, 16, rue Claude Bernard, 75231 Paris Cedex 05, France
[7] Max-Planck-Institut für Kohlenforschung Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany
[8] External Professor, University of Glamorgan, Wales, UK.
c/o Waters Informatics, Europaallee, 27-29 50226 Frechen, Germany.

*Author to whom comments should be sent.

JCAMP-DX V.6.00 for CHROMATOGRAPHY and MASS SPECTROMETRY HYPHENATED METHODS (IUPAC Technical Note 2005)

J. Hau[1], P. Lampen[2], R.J. Lancashire[3], R.S. McDonald[4], P.S. McIntyre[5], D.N. Rutledge[6], W. Schrader[7], A.N. Davies[8]

ABSTRACT

The increasing power of modern spectroscopic data systems has led to increasing use of multidimensional analytical techniques. These lead to multidimensional data sets, which need to be stored and exchanged between computer systems. In the definition of a generic structure for multi-dimensional data sets the framework was laid down for NTUPLE IUPAC/JCAMP-DX file formats. This technical note builds on the NTUPLE structure with specific recommendations for Chromatography data exchange and hyphenated techniques of chromatography / mass spectrometry.

This definition of this format is based on the IUPAC JCAMP-DX protocols, which were developed for the exchange of infrared spectra [1] and extended to chemical structures [2], nuclear magnetic resonance data [3] and mass spectra [4]. This standard of the **J**oint **C**ommittee on **A**tomic and **M**olecular **P**hysical Data is of a flexible design. The International Union of Pure and Applied Chemistry (IUPAC) now have responsibility for the support and development of these standards. General developments in the information technology field have constantly to be addressed such as extensions to cover year 2000 compatible date strings and good laboratory practice [5]. The aim of this paper is to use the new Generic IUPAC/JCAMP-DX version 6.00 protocol to define the special requirements for chromatography and extensions for liquid chromatography/mass spectrometry and gas chromatography/mass spectrometry. Optional extensions to the protocol have also been made to allow for lossless data migration to JCAMP-DX from the old ASTM E1947-98 netCDF ANDI standard for chromatography [9].

KEY WORDS

Data exchange, chromatography, mass spectrometry, LCMS, GCMS, hyphenated techniques, JCAMP-DX, data standards

1 INTRODUCTION

JCAMP-DX is an evolving, open, extensible, machine-independent, self-documenting file format for exchanging and archiving data from computerized laboratory instruments. JCAMP-DX was originated to meet the need for exchanging infrared spectra between similar instruments of different manufacture. The present document is the result of ongoing international collaborative initiatives efforts by users and manufacturers to extend JCAMP-DX to Chromatography and especially hyphenated techniques in mass spectrometry.

A major objective of JCAMP-DX is to enable routine capture of data at the source to make it available for exchange, archiving, and entry into databases. The emphasis is on readability, flexibility and longevity rather than on size or speed for rapid search and retrieval. All data are represented as labelled fields of variable length using printable ASCII characters (ASCII codes 32 through 126). JCAMP-DX is compatible with all electronic data storage media.

Where necessary, such as in an USFDA 21CFRpart11 environment for long-term archiving, a JCAMP-DX file can store all the content of an original raw data file without loss.

The JCAMP-DX recommendations are non-proprietary. These specifications are copyrighted by the International Union of Pure and Applied Chemistry (IUPAC) solely for the purpose of linking them with the name JCAMP-DX. The right to copy these specifications for scientific purposes is hereby granted.

Use of the name JCAMP-DX to describe data files implies that they conform to the form and style described in the relevant protocols and the information content defined for a particular DATATYPE.

Use of the name JCAMP-DX in the description of software capability implies the ability to read and write JCAMP-DX files as defined in the relevant published protocols for a particular DATATYPE.

This version of JCAMP-DX provides examples of the IUPAC/JCAMP-DX NTUPLES file structure for use in the exchange and long-term archiving of, for example, liquid chromatography/mass spectrometry (LC/MS) data sets. It complies with the requirement laid down in the IUPAC/JCAMP-DX v.6.00 protocol for different user communities and instrument vendor groups to define in consensus and adhere to the voluntary IUPAC/JCAMP-DX protocols.

This technical note will focus on:

1) generating an agreed list of DATATYPE-SPECIFIC labels. This list will need to be

agreed by consensus amongst users and instrument vendors, for their own techniques and applications.

2) providing examples of the use of the IUPAC/JCAMP-DX v6.00 file structure to assist in the implementation of this new standard in the field of chromatography and hyphenated techniques in mass spectrometry.

Inasmuch as it is very desirable for instrument data systems to be able to read and write files in a standard format directly, instrument vendors are encouraged to develop JCAMP-DX software for the instruments which they currently support. It is feasible for vendors to do so for a tightly defined CORE of information, which is described in Section 3.

2 DEFINITIONS

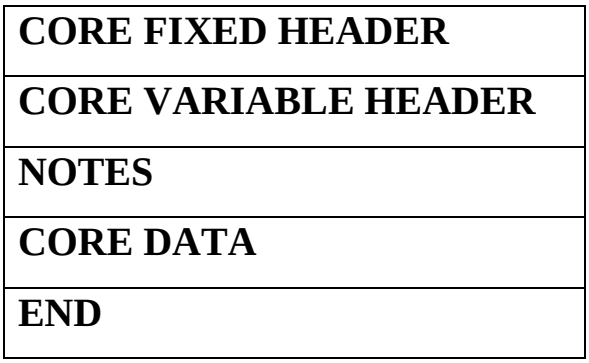

Figure 1. A simple JCAMP-DX file structure consisting of a header and a data section.

The following definitions are important for the understanding of the JCAMP-DX protocol.

JCAMP-DX is a FILE specification. The basic element of JCAMP-DX files is the LABELLED-DATA-RECORD (LDR). LABELLED-DATA-RECORDS are combined into BLOCKS. A simple file will consist of only a single block (figure 1.)

A more complicated JCAMP-DX file may contain more than one BLOCK. This is useful when various associated data sets are to be stored together in one file. Each block in this compound JCAMP-DX file has a full compliment of labels and could stand alone if separated from the original file (figure 2).

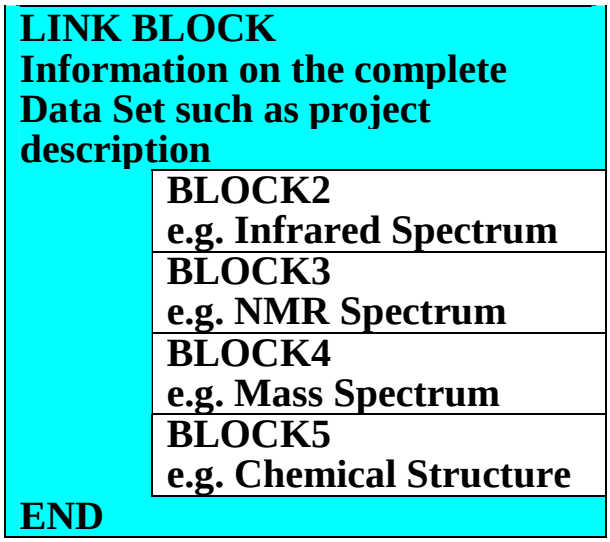

Figure 2. A compound JCAMP-DX file has one block dedicated to the file as a wrapper (the link block) and a series of complete associated blocks containing data

However, for the purpose of storing multidimensional data the NTUPLES format, which consists of multiple DATA PAGEs, is far more convenient method for storing data. As these data pages are very closely associated with each other they do not contain the full header information, as is the case for multiple blocks. They may only vary by representing a different time slice of a hyphenated technique raw data file for example (figure 3).

CORE FIXED HEADER, CORE VARIABLE HEADER, NOTES (All Generic or Datatype Specific)
NTUPLES=
PAGE=1
PAGE=2…...
PAGE=3…...etc
END NTUPLES
END

Figure 3. A simple JCAMP-DX file for multi-dimensional data using NTUPLES and PAGES

Should derived data wish to be stored along with the multidimensional raw data then a multi-block compound JCAMP-DX file using the NTUPLES data storage format for the raw data block is the recommended solution. This combines the more compact PAGES format for the many levels of raw data and the independent BLOCKS for the associated derived data. In an LC/MS files this could mean Total Ion Chromatograms, Single Ion Chromatograms or other associated derived data tables being stored in the blocks following the NTUPLES raw data block. (figure 4.)

All JCAMP-DX files are divided into CORE and NOTES subsets to separate the sections representing tabular data (CORE) and which must be parsed by computer from that which is essentially subsidiary information, mainly for reference by humans. We briefly review the respective definitions but for a complete overview the reader should refer to the original IUPAC Recommendation.

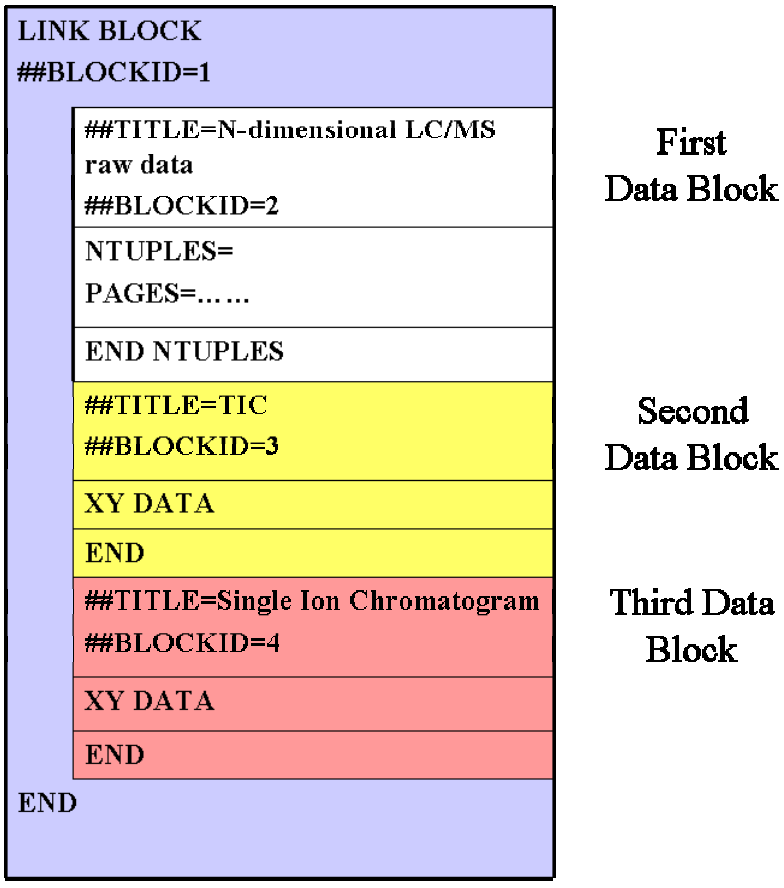

Figure 4. A compound JCAMP-DX file may also use NTUPLES blocks where mutlideimensional data is to be stored.

Labeled-Data-Records

Labeled-Data-Records (LDR) consist of a flagged *data-label* and an associated *data-set*. An LDR begins with a *data-label-flag* (##) and ends with the next data-label-flag. LDRs are divided into lines of 80 or fewer characters, terminated by <CRLF>. End-of-line is equivalent to a blank, except for certain special cases. Each LDR occupies as many lines as necessary. A data-label is the name of an LDR. It is delimited by a data-label-flag (##) and a *data-label-terminator* (=), for example ##TITLE= is a data-label for the

definition of the working title of the following spectrum. A line contains no more than one data-label. When labels are parsed, alphabetic characters are converted to upper case, and all spaces, dashes, slashes, and underlines are discarded. Thus, XUNITS and X-units are equivalent.
There are two kinds of LDR: Core and Note. Core LDRs are required. Notes are optional. Every definition of an LDR should include if it is a core or a note.

In part 2.2 - 2.5 we discuss the forms used for defined data sets.

2.1 TEXT

The LDR contain descriptive information for humans, not normally intended to be parsed by computers, i.e., title, comments, origin, etc.

2.2 STRING

Some LDR contain predefined text fields intended to be parsed by computers and read by humans. The format of each string field is specified under the LDR in which it is used.

2.3 AFFN

The easiest way to write the data is to use the ASCII free format numeric. This format is important to simplify direct user input. It is a format similar to freeform input in BASIC. A field starts either with +, -, decimal point or digit. E is the only allowed character to give the power of 10 by which the field must be multiplied. It is followed by + or - and two or three digits. The numeric field is terminated either by E, comma or blank.

2.4 ASDF

This is the ASCII squeezed difference form. The LDR contain tabular data using JCAMP-DX data compression scheme (see section 3.4.1).

2.5 Comments

Comments may be specified by a data-label-flag plus a data-label-terminator, with a null data-label (##=). Such comments may continue for more than one line, terminating at the next data-label-flag.
Comments may be entered at any point in a line by prefixing the first word of the comment by $$. Such comments continue only to the end of the current line, and they do not terminate an LDR.

3 THE CORE

This section describes the data labels for the JCAMP-DX core data:
The Core consists of four parts. The first part, called the *Fixed Header Information*, contains generic LDRs which are required for all JCAMP-DX files and which appear at the beginning of each file in a given order. The *Variable Header Information* contains records which are data type specific (in this case LC/MS specific) or which are only used in special types of JCAMP-DX files (e.g. compound files).

Whether a particular piece of information is required or not depends on the application. The third section *Core Data* contains the relevant parameters for the fourth section the *Data Table*. The type of data in the data table determines the parameters which must appear in the core data. Only one data table may appear per JCAMP-DX block (a block being a part of the JCAMP-DX file starting with ##TITLE= and ending with ##END=).

3.1 Core Fixed Header Information

3.1.1 ##TITLE= (TEXT)

Title and/or reason for the measurement suitable for printing on top of a report.
(Required)

3.1.2 ##JCAMP-DX= (STRING)

The version number of this protocol is 6.00. A description of the software used to generate the file follows the version number as a comment.
For example:

##JCAMP-DX= 6.00 $$ JCAMP-DX for Chromatography etc. (Windows XP version 1.0 ISAS-Dortmund, Germany)
(Required)

3.1.3 ##DATA TYPE= (STRING)

Keywords:
CHROMATOGRAPHY
GAS CHROMATOGRAPHY/MASS SPECTROMETRY
LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY
LINK

Distinguishes between different kinds of data sets. The strings GAS CHROMATOGRAPHY/MASS SPECTROMETRY and LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY reports a multidimensional data set arising from a chromatography experiment with mass spectrometry as the major detection method.
(Required)

3.1.4 ##DATA CLASS= (STRING)

Keywords:
XYDATA or
XYPOINTS or
NTUPLES
PEAK TABLE or
ASSIGNMENTS
This label defines the type of tabular data within the data block and is not to be used for link blocks.
(Required)

3.1.5 ##ORIGIN= (TEXT)

Here the name of organization, address, telephone number, name of individual contributor, email, etc., as appropriate must be added. This information is not optional.
(Required)

3.1.6 ##OWNER= (TEXT)

It is possible to set here a copyright linked to the spectrum, that has the form: "COPYRIGHT (C) <year> by <name>". If ##OWNER= contains "PUBLIC DOMAIN", the implication is that the data may be copied without permission on the authority of whoever is named under ##ORIGIN=
(Required)

3.1.7 ##END=

It is important to have a mark at the end of file in the data format to know that transfer has been complete and to distinguish between the blocks of a multi spectrum file.
(Required)

3.2 Core Variable Header Information

The JCAMP-DX standard is easy to understand and expand. Many LDRs are already defined in previous JCAMP-DX protocols and they should be used for Ion Mobility Spectrometry. However, in the case of the equipment parameters the particular requirements of this technique call for some special LDRs. Data-type specific LDRs start with "##." (see 4.2 below). In the following part definitions are given for LDRs which will allow a precise description of the equipment parameters.

3.2.1 ##BLOCKS= (AFFN) and ##BLOCK_ID= (AFFN)

Blocks are used as suggested in the JCAMP-DX 4.24 protocol (3.2) [1] with the extensions defined in the JCAMP-CS protocol (5.13) [2] to provide inter-block referencing.
For ease of use it is recommended to use separate files and to avoid the use of blocks. However, when, for example, XYDATA and a PEAK TABLE must be stored in the same file, these must be written as a compound file [1]. The use of compound data files is, of course, optional but when more than one block per file is stored the ##BLOCKS= LDR has the priority REQUIRED.

A compound JCAMP-DX file consists of a LINK block (##DATA TYPE=LINK) surrounding the DATA blocks and the total number of DATA blocks (N) must appear in the LINK block header (##BLOCKS=N). A unique positive integer (n) should be assigned to each DATA block inside the compound file (##BLOCK_ID=n). Linking of the blocks can then be achieved using ##CROSS REFERENCE= (see example 1). These records are to be used in compound JCAMP-DX files to provide inter-block referencing.
(Required only for compound files)

EXAMPLE 1.

```
##TITLE= example compound data file      $$ title of the whole compound file
##JCAMP-DX= 6.00                         $$ Name & Version No. of JCAMP-DX software
##DATA TYPE= LINK
##BLOCKS= 2                              $$ number of data blocks
##ORIGIN=                                $$ name of contributor, organization, address, telephone etc.
##OWNER=                                 $$ COPYRIGHT (C)'year' by 'name' or PUBLIC DOMAIN
##TITLE=                                 $$ title of the first data block
##JCAMP-DX= 5.01
##DATA TYPE= MASS SPECTRUM
##DATA CLASS= XYDATA
##BLOCK_ID= 1
##ORIGIN=                                $$ name of contributor, organization, address, telephone etc.
##OWNER=                                 $$ COPYRIGHT (C)'year' by 'name' or PUBLIC DOMAIN
##CROSS REFERENCE= PEAK TABLE:BLOCK_ID=2
. . .
##END=                                   $$ end of first data block
##TITLE=                                 $$ title of the second data block
##JCAMP-DX= 6.00
##DATA TYPE= MASS SPECTRUM
##DATA CLASS= PEAK TABLE
##BLOCK_ID= 2
```

```
##ORIGIN=                          $$ name of contributor, organization, address, telephone etc.
##OWNER=                           $$ COPYRIGHT (C)'year' by 'name' or PUBLIC DOMAIN
##CROSS REFERENCE= MASS SPECTRUM BLOCK_ID=1
. . .
##END=                             $$ end of second data block
##END=                             $$ end of link block and file
```

3.2.2 ##.MASS ANALYSER= (STRING)

The mass analyzer used in the LC/MS experiment.
The following keywords are defined:

QUADRUPOLE
ION TRAP
TIME OF FLIGHT
SECTOR
FOURIER TRANSFORM
TANDEM
TRIPLE QUADRUPOLE
QUADRUPOLE – TIME OF FLIGHT
ION STORAGE
MULTIPLE SECTOR

(Required)

3.2.3 ##.TANDEM SCANNING METHOD= (STRING, {AFFN})

Where a Tandem instrument is in use the scanning method must be recorded followed by the m/z ion under investigation where fixed.
The following keywords are defined:

PRODUCT ION SCAN
PRECURSOR ION SCAN
CONSTANT NEUTRAL LOSS
SINGLE/MULTIPLE REACTION MONITORING

Where ion storage experiments are conducted such as MS^n type experiments with ion traps or Fourier transform instruments AFFN can be an array of m/z values indicating the ions selected for the experiment.
(Required)

3.2.4 ##.INTERFACE= (STRING)

A description of the interface is required**.**
The following keywords are defined:

DIRECT
BATCH
MOVING BELT
MOVING WIRE

GAS CHROMATOGRAPH
THERMOSPRAY
PARTICLE BEAM
ATMOSPHERIC PRESSURE CHEMICAL IONIZATION
ELECTROSPRAY
CONTINUOUS FLOW FAST ATOM BOMBARDMENT
INDUCTIVELY COUPLED PLASMA
(Required)

3.2.5 ##.CHROMATOGRAPHY TYPE= (STRING)

A simplified generalized description of the chromatography experiment.
The following keywords are defined:
AFFINITY
ANION EXCHANGE
CAPILLARY
CAPILLARY ELECTROPHORESIS
CAPILLARY ELECTROCHROMATOGRAPHY
CAPILLARY ZONE ELECTROPHORESIS
CATION EXCHANGE
DISPLACEMENT
ENZYMOPHORESIS
ION EXCLUSION
ISOTACHOPHORESIS
NORMAL PHASE
REVERSED PHASE
SIZE EXCLUSION
SUPERCRITICAL FLUID
(Required)

3.2.6 ##.CHROMATOGRAPHY SOLVENTS= (N,C,U)

A list of the solvents used is stored here with their concentrations where N stands for the name of the solvent, C for concentration or concentration gradient and U for units. Each solvent is enclosed in brackets and each new solvent starts on a new line and may continue onto following lines.
(Required)

3.2.7 ##.ADDITIVES= (N,C,U)

A list of additives used I the mobile phase to improve separation etc. N is for the name, C concentration and U units of concentration.
(Required)

3.2.8 ##. DIMENSIONALITY= (AFFN)

This label provides the data set with a reference value for the dimensionality of the MSn experiment. It is intended to make the dimensionality clear at the start of the file without the need for parsers to read the entire data table to understand the underlying experiment dimensionality. It will also allow software programs that can, for example, only handle simple LC/MS or LC/MS/MS data sets to provide a warning that the file being read has more dimensions than can be displayed. For LC/MS data the label would read ##.DIMENSIONALITY=1, for LC/MS/MS ##.DIMENSIONALITY=2 etc.etc.
(Required)

3.2.9 ##.IONIZATION MODE= (STRING)

The type of ionization is given in the following strings. The polarity is registered with the appropriate sign tagged onto the relevant string.

EI+ or EI-	for electron impact ionization
CI+ or CI-	for chemical ionization
FAB+ or FAB-	for fast atom bombardment
TSP+ or TSP-	for thermospray ionization
ESI+ or ESI-	for electroospray ionization
APICI+ or APICI-	for atmospheric pressure ionization with chemical ionization
LD+ or LD-	for laser desorption

Extra information may be added following the string by use of the $$ comment (e.g. “CI+ $$ <Reaction Gas> or “FAB- $$ <Matrix>/<primary particle>).

3.3 Core Data

3.3.1 ##XUNITS= (STRING) and ##YUNITS= (STRING)

Here the units of the axis can be given. The following keywords are defined:
For ##XUNITS=: SECONDS, MILLISECONDS, MICROSECONDS and NANOSECONDS,
For ##YUNITS=: COUNTS, MICROAMPERES, NANOAMPERES, and PICOAMPERES.
(Required)

3.3.2 ##FIRSTX= (AFFN) and ##LASTX= (AFFN)

First and last actual abscissa values of ##XYDATA=. First tabulated abscissa times ##XFACTOR= should equal ##FIRSTX=.
(Required for ##DATA CLASS=XYDATA)

3.3.3 ##FIRSTY= (AFFN)

Here the actual ordinate value corresponding to ##FIRSTX= is meant. ##FIRSTY= should be equal ##YFACTOR= times the first Y-value in ##XYDATA=.
(Required for ##DATA CLASS=XYDATA)

3.3.4 ##XFACTOR= (AFFN) and ##YFACTOR= (AFFN)

The values of a spectrum may be converted to integer to save space and allow the DIFDUP format (see 3.4.1). It is important to select a convenient scaling to keep the file within reasonable limits but to store all significant digits. In such a case the ##XFACTOR= and ##YFACTOR= LDR contain a floating-point number to be multiplied by the values in ##XYDATA= to arrive at the original data point value.
In most cases ±32767 is sufficient, therefore this is the *recommended ordinate scaling*. If a larger scaling is necessary it is required to give the actual unscaled maximum and minimum of the ordinates in the records ##MAXY= and ##MINY=. This avoids a 2-byte integer overflow in the program reading the data table.
For example if a Y-value with 9 significant figures (e.g. 0.002457194) needs to be converted to an integer value for ASDF coding then:
a) divide by the maximum of the absolute Y-value (say 0.346299765)
b) and then multiply by the largest integer value (MAXINT) necessary to place all significant figures left of the decimal point
c) convert to integers

for this example -

$Y_{integer}$ =INTEGER((Y_{real} / MAX(ABS(Y_{real})))*MAXINT)
=INTEGER(0.002457194/0.346299765)*1.0E+11
=INTEGER(709556935.4487)
=709556935

then ##YFACTOR= (MAX(ABS(Y_{real}))/MAXINT)
(Required for ##DATA CLASS=XYDATA)

3.3.5 ##NPOINTS= (AFFN)

The number of points in the data table is required for all data classes: XYDATA, XYPOINTS, PEAK TABLE, and ASSIGNMENTS.
(Required)

3.4 Core Data Table

Data must be stored in one of the following data formats (3.4.1 – 3.4.4). Only one of these formats is allowed per DATA block and the selected data table is given in the header e.g. ##DATA CLASS=XYDATA.

3.4.1 ##XYDATA= (AFFN or ASDF).

This LDR contains a table of spectral data with abscissa values at equal intervals specified by parameters defined in section 3.3. The label is followed by a *variable list*, (X++(Y..Y)) where .. indicate indefinite repeat of Y-values until the end of line and ++ indicates that X is incremented by (LASTX-FIRSTX) / (NPOINTS-1) between two Y-values. For discrete point the AFFN form is allowed where each Y value is written out in full. This form creates large files, but is easily human readable. The following ASDF forms produce smaller files at the cost of reduced human readability.

SQUEEZED FORM (SQZ)
Data compression is possible using the *squeezed form* (SQZ) in which the delimiter, the leading digit and sign are replaced by a pseudo-digit from Table 1.
For example the Y-values 30, 32 would be represented as C0C2.

DIFFERENCE FORM (DIF)
For a better compression, it is possible to use the *difference form* (DIF) where the delimiter, leading digit and sign of the difference between adjacent values are transformed in a pseudo-digit from Table 1.
For example the Y-values 30, 32 would now be represented as C0K.

To ensure data coding is correct a Y-value check is built-in. Each line starts with the absolute X- and Y-value, which is the same as the last calculated value of the previous line. The last line of a block of DIF data contains only the abscissa and ordinate for a Y-value check of the last ordinate.

DUPLICATE SUPPRESION (DUP)
Another possible variation is to include *duplicate suppression* (DUP) replacing two or more adjacent and identical numbers with pseudo-digits as given in Table 1. This can be used with all ASDF forms.
For example 50 50 50 50 becomes E0V when combining DUP with SQZ.

The best compression (but the least human-readable form) can be achieved when DIF and DUP are combined to provide the form called DIFDUP. In this format the duplicate count is obtained by counting identical differences.
The example above becomes E0%%% in DIF form and E0%U in DIFDUP form.

Table 1:
Pseudo digits used to compress spectra data in SQZ-, DIF- and DUP-format

ASCII digits	0	1	2	3	4	5	6	7	8	9
Positive SQZ	@	A	B	C	D	E	F	G	H	I
Negative SQZ		a	b	c	d	e	f	g	h	i
Positive DIF	%	J	K	L	M	N	O	P	Q	R
Negative DIF		j	k	l	m	n	o	p	q	r
DUP		S	T	U	V	W	X	Y	Z	s

Example for uncompressed data storage

```
##TITLE= Incomplete example file for uncompressed data!
.....
##XUNITS= MILLISECONDS
##YUNITS= NANOAMPERES
##XFACTOR= 1
##YFACTOR= 0.1
##FIRSTX= 4
##LASTX= 56
##NPOINTS= 53
##FIRSTY= 0
##XYDATA= (X++(Y..Y)
4 0 0 0 0 2 4 4 4 7
13 5 4 4 5 5 7 10 11 11
22 6 5 7 6 9 9 7 10 10
31 9 10 11 12 15 16 16 14 17
40 38 38 35 38 42 47 54 59 66
49 75 78 88 96 104 110 121 128
##END=
```

Example for DIFDUP-form

```
##TITLE= Incomplete example file for DIFDUP data form!
.....
##XUNITS= MILLISECONDS
##YUNITS= NANOAMPERES
##XFACTOR= 1
##YFACTOR= 0.1
##FIRSTX= 4
##LASTX= 56
##NPOINTS= 53
##FIRSTY= 0
##XYDATA= (X++(Y..Y)
```

4@VKT%TLkj%J%KLJ%njKjL%kL%jJULJ%kLK1%lLMNPNPRLJ0QTOJ1P
56A28
##END=

3.4.2 ##XYPOINTS= (AFFN)

This LDR contains a table of spectral data with unequal abscissa increments. The label is followed by a variable list, (XY..XY). X and Y are separated by commas, data pairs are separated by semi-colons or blanks. This LDR should not be used for peak tables.

3.4.3 ##PEAK TABLE= (AFFN)

It is recommended to store peak information using ##PEAK ASSIGNMENTS=, however, for backward compatibility this definition is included. This data table contains a table of peaks where the peak data starts on the following line. The label is followed by the variable list (XY) or (XYW) for peak position, intensity and width, where known, on the same line. The function used to calculate the peak width should be defined by a $$ comment in the line below the label. The peak groups are separated by semi-colon or space, components of a group are separated by commas.

3.4.4 ##PEAK ASSIGNMENTS= (STRING)

Variable list: (XA), (XYA) or (XYWA)
After this LDR a list of peaks and their assignments for each components are given in the following form:
(X_1[, Y_1][, W_1], <A_1>)
. . . .
(X_i[, Y_i][, W_i], <A_i>)
X and Y indicates the location and height of each peak in units given by ##XUNITS= and ##YUNITS=. W stands for width in ##XUNITS= and A represents a string describing the assignment enclosed in angle brackets.
The parentheses provide a start and end flag of each assignment. Square brackets indicate optional information. It is important for the technical readability to have the same format for the whole peak assignment table and describe it after the ##PEAK ASSIGNMENTS= LDR with (XA), (XYA) or (XYWA). This LDR should be followed by a comment, which gives the method of finding the peak.

For example:
##PEAK ASSIGNMENTS= (XYWA)
(15, 20.0, 1.0,<benzene>)
(30, 40.0, 2.0,<toluene>)
.....
(Optional)

3.4.5 ##PEAK ASSIGNMENTS= (STRING)

A significantly extended PEAK ASSIGNMENTS capability has been defined specifically for chromatography to ensure that no data is lost during migration to JCAMP-DX from netCDF ANDI chromatography data files. Here an extended variable list is allowed containing the following values all but structure coming out of the AIA file.:
(retention time, name, amount, amount units, start time, end time, width, area, %area, height, %height, Kovats index, structure)
In the form (X_i,<$name_i$>,Ci, U_i, $Xstart_i$, $Xfinish_i$, W_i, $Area_i$, $\%Area_i$, $Height_i$, $\%Height_i$ $KOVATS_i$,<A>)
This would look like:
(53.5,<acetone>,10.4,<microgrammes>, 52.1,55.6,3.3,3.2,17,103.4,16.5,47.6,<acetone.mol & block 7>)
……
(Optional)

3.4.6 ##PEAK BASELINE = (STRING)

Another enhancement for the sake of compatibility is the PEAK BASELINE capability. Here the allowed variable list containing the following values
(start time, start value, stop time, stop value, peak start code, peak stop code, manual re-integration) similar to ##PEAK ASSIGNMENTS=

The peak start and stop codes are defined in the ASTM standard E1947 as follows

> *Codes that are used to describe how the baselines have actually been drawn. The peak type may be represented by a two-letter code. The following are examples of peak detection codes:*

- B = baseline peak, that is, the peak begins and/or ends at the baseline.
- P = perpendicular drop, that is, the peak begins and/or ends with a perpendicular drop.
- skimmed peak = skimmed peak, that is, the peak is a shoulder peak that is skimmed.
- VD = vertical drop, that is, the peak begins or ends at a vertical drop to the skim line (such as between two skimmed peaks).
- HP = horizontal projection, that is, the peak baseline starts and/or ends with a horizontal projection.
- EX = exponential skim, that is, the peak starts and/or stops with an exponential skim.
- PT = pretangent skim, that is, the leading edge of the peak is tangent skimmed.
- MN = manual peak, that is, the user forced the baseline at the data level.
- FR = forced peak, that is, the user forced the baseline at a user-supplied level.
- DF = user forced daughter peak, that is, the user forced a baseline on the side of a fused peak (daughter peak).
- LP = lumped peak, that is, peaks values are lumped (added) together until this timed event is turned off.

(Optional)

4 THE NOTES

The notes portion of a JCAMP-DX file or block complements the core. Notes describe an experiment in greater detail than does ##TITLE=, including descriptions of equipment, method of observation, and data processing, as appropriate. Notes may contain information, which is not found in the native file in which data is originally collected by an instrument. Notes are placed before the core data section to permit them to be viewed without listing the whole file. The contents of the notes depend on the user as well as the technique or application. Notes will vary for different samples, sites, data systems, and applications.

4.1 Global Notes

These have been already defined in JCAMP-DX and are common to all spectroscopy types. The file headers, spectral and sample parameters are often the same for different analytical techniques. This allows us to implement many of the standard LDRs from the existing JCAMP-DX protocols. The list given below is only a selection from those allowed. A complete list can be found in the references [1-5].

At least one of the optional LDRs described in section 4.1.4 - 4.1.8 should be included in each JCAMP-DX file. This is important for later archiving as these fields will yield more detailed information on the content of the data stored than a simple ##TITLE= field.

4.1.1 ##LONG DATE= (STRING)

Date of measurement is required by many agencies and recommended in the year 2000 form: YYYY/MM/DD [HH:MM:SS[.SSS] [±XXXX]]. YYYY is the long format of the year, MM the number of the month, DD the number of the day, HH the hour, MM the minutes, SS.SSS the seconds and fractions of a second of the measurement, ±XXXX is the difference to the UTC (e.g. +0100 is one hour difference to UTC).
(Optional)

4.1.2 ##SOURCE REFERENCE= (TEXT)

Here an identification of the original spectrum file in native format or library name and serial number is possible for example.
(Optional)

4.1.3 ##CROSS REFERENCE= (TEXT)

Used to link additional data for the same sample such as other types of spectra or chemical structures for example:
##CROSS REFERENCE=

MASS SPECTRUM: EXTERNAL_FILE= FILENAME.DX
TOTAL ION CHROMATOGRAM BLOCK_ID=16
STRUCTURE: BLOCK_ID=4

.....
(Optional)

4.1.4 ##SAMPLE DESCRIPTION= (TEXT)

If the sample is not a pure compound, this field should contain its description, i.e., composition, origin, appearance, results of interpretation, etc. If the sample is a known compound, the following LDRs specify structure and properties, as appropriate.
(Optional)

4.1.5 ##CAS NAME= (STRING)

Name according to Chemical Abstracts naming conventions as described in the CAS Index Guide is required here. Examples can be found in Chemical Abstracts indices or the Merck Index. Greek letters are spelled out, and standard ASCII capitals are used for small capitals, Sub-/Superscripts are indicated by prefixes / and /\. Example: alpha-D-glucopyranose, 1-(dihydrogen phosphate).
(Optional)

4.1.6 ##NAMES= (STRING)

Here the common, trade or other names are allowed. Multiple names are placed on separate lines.
(Optional)

4.1.7 ##MOLFORM= (STRING)

Another possibility of describing the sample is to write down the molecular formula. Elemental symbols are arranged with carbon first, followed by hydrogen, and then remaining element symbols in alphabetic order.
The first letter of each elemental symbol is capitalized. The second letter, if present, is lower case. One-letter symbols must be separated from the next symbol by a blank or digit. Sub-/Superscripts are indicated by the prefixes: / and /\, respectively. Sub- and superscripts are terminated by the next non-digit. Slash may be omitted for subscripts.
For readability, each atomic symbol may be separated from its predecessor by a space. For substances that are represented by dot disconnected formulas (hydrates, etc.), each fragment is represented in the above order, and the dot is represented by *. Isotopic mass is specified by a leading superscript. D and T may be used for deuterium and tritium.
(Optional)

Examples:
C2H4O2 or C2 H4 O2 (acetic acid)
H2 /\17O (water, mass 17 oxygen)

4.1.8 ##CONCENTRATIONS= (STRING)

The list of the known components and their concentrations has the following form, where N stands for the name and C for the concentration of each component in units given with U in the form:
##CONCENTRATIONS= (NCU)
(N_1, C_1, U_1)
. . .
(N_i, C_i, U_i)
The group for each component is enclosed in parentheses. Each group starts a new line and may continue on following lines.
(Optional in JCAMP-DX but in this case strongly recommended)

4.1.9 ##SPECTROMETER/DATA SYSTEM= (TEXT)

This LDR contains manufacturers' name, model of spectrometer, software system, and release number, as appropriate in the form used by the manufacturer.
(Optional)

4.1.10 ##DATA PROCESSING= (TEXT)

Here all mathematical procedures used before storing the data in the JCAMP-DX file are described. This LDR is also important in peak assignments.
(Optional)

4.1.11 ##XLABEL= (TEXT) and ##YLABEL= (TEXT)

These LDRs give the possibility of labeling the axis.
(Optional)

4.1.12 ##MAXX= (AFFN), ##MINX= (AFFN), ##MAXY= (AFFN), and ##MINY= (AFFN

These LDRs give the largest and smallest actual X- and Y-values in a spectrum. ##MAXY= and ##MINY= are required if Y-range exceeds recommended ordinate scaling (3.3.4).
(Optional)

4.2 Data-Type Specific Notes

Data-type-specific-labels are RESERVED labels which are defined by qualified user groups for a particular data-type. A data-type-specific-label is distinguished by a *data-type-specific-label-name* which starts with a period (e.g. ##•MASS ANALYSER=). Choice of period as distinguishing character is by analogy with the convention for data-structure names in Pascal and C. Effectively, the full label name is the concatenation of the data-type name and the label-name, with a period in between, i.e., ##LIQUID CHROMATOGRAPHY/MASS SPECTROMETRY•MASS ANALYSER=.

4.2.1 ##.INLET TEMPERATURE= (AFFN)

Inlet temperature in °C
(Optional)

4.2.2 ##.SOURCE TEMPERATURE= (AFFN)

Source temperature in °C
(Optional)

4.2.3 ##.IONIZATION ENERGY= (AFFN)

Ionization energy where appropriate in eV.
(Optional)

4.2.4 ##.ACCELERATING VOLTAGE= (AFFN)

Accelerating voltage in Volts where appropriate
(Optional)

4.2.5 ##.TOTAL ION CURRENT= (AFFN,STRING)

Total Ion Current as measured with units of measure (COUNTS or VOLTS are defined as allowed string values).
(Optional)

4.2.6 ##.AQUISITION RANGE= (AFFN,AFFN)

The measurement region in amu where appropriate (first, last)
(Optional)

4.2.7 ##.PARENT SCAN REFERENCE= *(STRING)*

Variable list: (AFFN,X) or (AFFN,X,W)

For LC/MSn experiments PARENT SCAN REFERENCE should be used to link parent and daughter scans with the NTUPLES pages. The first AFFN references the parent scan PAGE. The second numeric value refers to the X-axis position of the parent ion. Where an X-axis position and a width are used to define the origin in the parent scan from which the daughter scan arises then the variable W is added.
(Required for MSn where n>=2)

4.2.8 ##.MASS ON COLUMN= (AFFN)

Mass on column from the ASTM standard E 1947 – 98 [9].
A measure of column loading. It is usually reported as the sum of the peak-amount(s).
It needs to be determined against known peaks.
(Optional)

4.2.9 ##.PEAK ASYMMETRY= (AFFN)

Peak-asymmetry from the ASTM standard E 1947 – 98 [9] defined as As= B/A, where A and B are the widths for the front and back parts of a peak, commonly measured at 10 % peak height for USP at 5 %.
In the form (As..As).
(Optional)

4.2.10 ##.PEAK EFFICIENCY= (AFFN)

Peak-efficiency from the ASTM standard E 1947 – 98 [9].
A numerical array in the form (PE..PE)
Also known as the column theoretical plate number for an individual peak.
The peak-efficiency can be expressed as:
[(retention-time / width-at-half-height)^2] x 5.54 or as
[(retention-time / baseline - bandwidth ^2) x 16
where baseline-bandwidth is the peak width along the baseline as determined by the intersection points of the tangents drawn to the peak above its points of inflexion.
(Optional)

5 SUMMARY

The following tables list the LDRs discussed with their basic parameters shown.

THE CORE

Core Fixed Header Information

LDR	Data-form	Keyword	Status	Location
##TITLE=	(TEXT)		Required	3.1.1
##JCAMP-DX=	(STRING)	6.00	Required	3.1.2
##DATA TYPE=	(STRING)	CHROMATOGRAPHY, GASCHROMATOGRAPHY/MASSSPECTROMETRY, LIQUIDCHROMATOGRAPHY/MASS SPECTROMETRY, or LINK	Required	3.1.3
##DATA CLASS=	(STRING)	XYDATA, XYPOINTS, NTUPLES; PEAK TABLE, or ASSIGNMENTS	Required	3.1.4
##ORIGIN=	(TEXT)		Required	3.1.5
##OWNER=	(TEXT)		Required	3.1.6
##END=			Required	3.1.7

Core Variable Header Information

LDR	Data-form	Keyword	Status	Location
##BLOCKS=	(AFFN)		Required for compound files	
##BLOCK_ID=	(AFFN)		Required for compound files	
##.MASS ANALYSER=	(STRING)		Required	
##.TANDEM SCANNING METHOD=	(STRING, {AFFN})		Required	
##.INTERFACE=	(STRING)		Required	
##.CHROMATOGRAPHY TYPE=	(STRING)		Required	
##.CHROMATOGRAPHY SOLVENTS=	(N,C,U)		Required	
##.ADDITIVES=	(N,C,U)		Required	
##.DIMENSIOANLITY=	(AFFN)		Required	
##.IONIZATION MODE=	(STRING)	EI+ or EI- CI+ or CI- FAB+ or FAB- TSP+ or TSP- ESI+ or ESI- APICI+ or APICI- LD+ or LD-	Required	5.1.1

Core Data

LDR	Data-form	Keyword	Status	Location
##XUNITS=	(STRING)	SECONDS, MILLISECONDS, MICROSECONDS; NANOSECONDS	Required	3.3.1
##YUNITS=	(STRING)	MICROAMPERES, NANOAMPERES; PICOAMPERES	Required	3.3.1
##FIRSTX=	(AFFN)		Required for ##DATA CLASS= XYDATA	3.3.2
##LASTX=	(AFFN)		Required for ##DATA CLASS= XYDATA	3.3.2
##FIRSTY=	(AFFN)		Required for ##DATA CLASS= XYDATA	3.3.3
##XFACTOR=	(AFFN)		Required for ##DATA CLASS= XYDATA	3.3.4
##YFACTOR=	(AFFN)		Required for ##DATA CLASS= XYDATA	3.3.4
##NPOINTS=	(AFFN)		Required	3.3.5

Core Data Table

LDR	Data-form	Keyword	Status	Location
##XYDATA=	(AFFN or ASDF)	(X++(Y..Y))	Optional	3.4.1
##XYPOINTS=	(AFFN)	(XY..XY)	Optional	3.4.2
##PEAK TABLE=	(AFFN)	(XY) or (XYW)	Optional	3.4.3
##PEAK ASSIGNMENTS=	(STRING)	(XA), (XYA) or (XYWA)	Optional	3.4.4
##PEAK ASSIGNMENTS=	(STRING)	Complex extended	Optional	3.4.5
##PEAK BASELINE=	(STRING)	Complex extended	Optional	3.4.6

THE NOTES

Global Notes

LDR	Data-form	Keyword	Status	Location
##LONG DATE=	(STRING)		Optional	4.1.1
##SOURCE REFERENCE=	(TEXT)		Optional	4.1.2
##CROSS REFERENCE=	(TEXT)		Optional	4.1.3
##SAMPLE DESCRIPTION=	(TEXT)		Optional	4.1.4
##CAS NAME=	(STRING)	Names defined by CAS Index Guide	Optional	4.1.5
##NAMES=	(STRING)		Optional	4.1.6
##MOLFORM=	(STRING)	see Def.	Optional	4.1.7
##CONCENTRATIONS=	(STRING)	(NCU)	Optional but strongly recommended	4.1.8
##SPECTROMETER/ DATA SYSTEM=	(TEXT)		Optional	4.1.9
##DATA PROCESSING=	(TEXT)		Optional	4.1.10
##XLABEL=	(TEXT)		Optional	4.1.11

##YLABEL=	(TEXT)		Optional	4.1.11
##MAXX=	(AFFN)		Optional	4.1.12
##MINX=	(AFFN)		Optional	4.1.12
##MAXY=	(AFFN)		Required if Y-range exceeds recommended ordinate scaling	4.1.12
##MINY=	(AFFN)		Required if Y-range exceeds recommended ordinate scaling	4.1.12

Data-type Specific Notes

LDR	**Data-form**	**Keyword**	**Status**	**Location**
##.INLET TEMPERATURE=	(AFFN)		Optional	4.2.1
##.IONIZATION ENERGY=	(AFFN)	RECT or CYL	Optional	4.2.2
##.SOURCE TEMPERATURE=	(AFFN)	RECT or CYL	Optional	4.2.3
##.ACCELERATING VOLTAGE=	(AFFN)		Optional	4.2.4
##.TOTAL ION CURRENT=	(AFFN,STRING)		Optional	4.2.5
##.AQUISITION RANGE=	(AFFN,AFFN)		Optional	4.2.6
##.PARENT SCAN REFERENCE= (STRING)	(AFFN, X) or (AFFN, X, W)		(Required for MSn where n>=2)	4.2.7
##.MASS ON COLUMN=	(AFFN)		Optional	4.2.8
##.PEAK ASYMMETRY=	(AFFN)		Optional	4.2.9
##.PEAK EFFICIENCY=	(AFFN)		Optional	4.2.10

6 REFERENCES

1. R.S. McDonald and P.A. Wilks Jr., JCAMP-DX: A Standard Form for Exchange of Infrared Spectra in Computer Readable Form. *Appl. Spectrosc*. **42,** 151-162 (1988), and 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).

2. J. Gasteiger, B.M.P. Hendriks, P. Hoever, C. Jochum and H. Somberg., JCAMP-CS: A Standard Exchange Format for Chemical Structure Information in Computer-Readable Form. *Appl. Spectrosc.*, **45,** 4-11 (1991).

3. A.N. Davies and P. Lampen, JCAMP-DX for NMR. *Appl. Spectrosc.,* **47,** 1093-1099 (1993).

4. P. Lampen, H. Hillig, A.N. Davies and M. Linscheid, JCAMP-DX for Mass Spectrometry. *Appl. Spectrosc.,* **48,** 1545-1552 (1994).

5. P. Lampen, J. Lambert, R.J. Lancashire, R.S. McDonald, P.S. McIntyre, D.N. Rutledge, T. Fröhlich and A.N. Davies, An Extension to the JCAMP-DX Standard File Format, JCAMP-DX V.5.01 (IUPAC Recommendations 1999). *Pure Appl. Chem.*, **71,** 1549-1556 (1999).

6. J.I. Baumbach, P. Lampen, A.N. Davies, IUPAC / JCAMP-DX: An International Standard for the Exchange of Ion Mobility Spectrometry Data. *Int. J. Ion Mobility Spectrom.* **1,** 64-67 (1998).

7. G.A. Eiceman, Z. Karpas, Ion Mobility Spectrometry, CRC Press, Inc. Boca Raton, FL, 1993.pp 57-86

8. J.I. Baumbach and G.A. Eiceman, Ion Mobility Spectrometry: Arriving On Site and Moving Beyond a Low Profile. *Appl. Spectrosc.,* **53,** 338A-355A (1999).

9. ASTM Standard E1947-98, Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data