
**Microbeam analysis — EMSA/MAS
standard file format for spectral-
data exchange**

*Analyse par microfaisceaux — Format de fichier standard EMSA/MAS
pour échange de données spectrométriques*



Reference number
ISO 22029:2012(E)

© ISO 2012



COPYRIGHT PROTECTED DOCUMENT

© ISO 2012

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

	Page
Foreword	iv
Introduction	v
1 Scope	1
2 General considerations	1
3 Format description	2
3.1 General	2
3.2 Required keywords	3
3.3 Spectral data	4
3.4 Optional keywords	5
3.5 Ending a file	8
4 Example of a data file in the EMSA/MAS spectral format	8
Bibliography	10

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 22029 was prepared by Technical Committee ISO/TC 202, *Microbeam analysis*.

This second edition cancels and replaces the first edition (ISO 22029:2003), which has been technically revised.

Introduction

The original EMSA/MAS Standard File Format for Spectral Data Exchange was published in October 1991. Since then, advances in both microbeam analysis techniques and in PC technology have meant that this original standard is not fully able to meet modern requirements. The members of ISO/TC 202 (the International Organization for Standardization Technical Committee for Microbeam Analysis) propose this updated version. Every effort has been made to only make those changes that improve or update the applicability of the standard, while minimizing incompatibility with the original version. The remit of TC 202 does not include surface analysis techniques, which are addressed by TC 201, and so references to these techniques have been removed from the original standard where necessary. The original document also included examples of coding and telecommunications protocols. Since these are now largely outdated, and not integral to the formatting of the data, these have also been removed.

It is noted that one of the originating societies (EMSA) has modified its name since the original document was published. The society is now officially known as the “Microscopy Society of America”, or MSA, the term “Electron” having been dropped to more fully indicate the work and interest of the membership of the society in all forms of microscopy.

Copyright International Organization for Standardization
Provided by IHS under license with ISO
No reproduction or networking permitted without license from IHS

Microbeam analysis — EMSA/MAS standard file format for spectral-data exchange

1 Scope

This International Standard presents a simple format for the exchange of digital spectral data that has been designated as an EMSA/MAS standard. This format is readable by both humans and computers and is suitable for transmission through various electronic networks, the phone system (with modems) or on physical computer storage devices (such as removable media). The format is not tied to any one computer, programming language or computer operating system. The adoption of a standard format would enable different laboratories to freely exchange spectral data, and would help to standardize data analysis software. If equipment manufacturers were to support a common format, the microscopy and microanalysis community would avoid duplicated effort in writing data analysis software.

2 General considerations

The virtues of a single standard data format have been admirably related by various authors^{[1],[2],[3],[4]}. It would often be convenient, after visiting another laboratory to use a different type of microanalytical spectrometer, to be able to return to one's own laboratory to analyse the data, or for a laboratory to be able to send a spectrum to another group at another location for analysis on their computer. A common format would also enable test spectra to be transported between data acquisition systems, in order to compare different data analysis routines, and would give users greater choice of analysis procedure, based on commercial or public-domain software.

Obviously, an ideal solution would be for the manufacturers to represent data in a standard format, but they are unlikely to agree on this without some direction from their customers (the microanalysis community). Therefore it is highly desirable for EMSA and MAS to proceed with the adoption of a standard format. Such a format does not preclude any research group or manufacturer from having their own, possibly proprietary, format. Spectral data can be stored internally in any format, as long as there is an option to convert it to the external standard (and vice versa) for the purposes of exchange. We believe that a standard format should possess the following attributes:

- a) It should be capable of representing the data exactly (without altering the scientific content).
- b) The format should be simple and easy to use.
- c) It must NOT be tied to any particular computer, programming language or operating system. It should work on a large number of computers of all sizes, although we cannot guarantee that it will work on all possible computers.
- d) The format should be both human and machine (computer) readable.
- e) It should be compatible with existing electronic communication networks and with the phone system (using modems). Future networks will likely retain compatibility with these.
- f) The format should support spectra of interest to the microanalysis community (such as XEDS, EELS, AES) and should be flexible enough to accommodate future data sets not yet specified.
- g) Each file should contain enough information to uniquely identify the type and origin of the spectral data and to reconstruct its significance.
- h) Where possible, the format should be compatible with various commercial data plotting or analysis programs (i.e. spreadsheets, or graphical-analysis packages).

- i) The proposed format need not be the most efficient storage mechanism. Its primary goals, stated above, will generally prevent storage efficiency. If anything, this format will err on the side of simplicity and ease of use.

The format originally employed by the Electron Microscopy and Microanalysis Public Domain Library (EMMPDL) at Argonne^[1] has the virtue of simplicity, but is too rigid for general use. A recent revision^[5] corrected some inadequacies, but a more serious re-examination is in order. The format proposed by a previous EMSA task force^{[3],[4]} addresses many of the problems, but is thought by some microscopists to be too complicated for everyday use. The VAMAS format, whose description^[6] runs to 60 pages, is also too complex for our perceived purpose. A format, named JCAMP-DX, used by the infrared-spectroscopy community^[2] is specific and detailed, but is somewhat off target for the spectroscopies of interest to our community.

The format proposed here follows JCAMP-DX in many ways, but is less complicated and has features tailored to X-ray, energy-loss and Auger spectroscopies. We circulated a preliminary version of this proposal to several manufacturers of XEDS systems and have received back comments and suggestions, many of which have been incorporated into this document.

The companion problem of a standard format for digital-image storage is similar to that of spectral data, but is sufficiently different to warrant its own standard. Whereas most spectra are sufficiently compact that they can be stored in a human-readable form, image data are usually so extensive as to require storage of “raw” binary numbers. There exist formats for image storage that are in widespread use. One of these (probably TIFF) should be endorsed by EMSA and MAS, allowing the microanalysis community to take advantage of the large amount of commercial and public-domain software already available.

3 Format description

3.1 General

The general structure of the data file format can be summarized as a simple sequential text file. It begins with a series of header lines which serve to define the characteristics of the spectrum. These header lines are identified by unique keyword fields which occupy the first 15 positions of each line, followed by a data field. These are described in detail below. After the header lines, a keyword indicates the start of data, and the data then follow on successive lines in a manner which is defined explicitly within the header. Finally, after all the data are presented, an end-of-data keyword indicates that the data set is complete. This is diagrammatically illustrated below:

Header lines

*
*

Successive lines beginning with EMSA/MAS-defined keywords, some of which are required and some are optional

*
*

Start-of-data keyword

*
*

Experimental data

*
*

End-of-data keyword

In general, each line of the file either contains a keyword and its associated value or spectral data. All header lines are readily identified as they each begin with “#” in the first character field or column. This symbol demarks the start of a 13 character keyword field, providing descriptive information about the data followed by an associated value. EMSA/MAS-defined keywords (whose definition may be changed only by EMSA/MAS) begin with a single # and occupy the first 13 columns (characters) of each header line. The keyword itself consists of at most 12 characters, which directly follow the #. Shorter

keywords may be employed, and any remaining spaces following the defined keyword may be filled with descriptive text such as unit designation for ease of legibility when the file is printed (see example in Table 1). If a position in the keyword field is not used, it must be filled with a space character. The keywords are not case-sensitive, so that #Xunits is interpreted as being the same as #XUNITS.

The 14th and 15th character positions (columns) in each header line are occupied by keyword/value field separators, which consist of a colon followed by a space. The value associated with each keyword starts in column 16 and may be either textual or numeric as defined by the keyword. Each line of the file may contain no more than 79 characters (for compatibility with the largest number of computers and computer networks, and for general legibility on standard-width CRT screens). Since the keyword and its separator occupies the first 15 positions, this means that all remaining information following the keyword is limited to a maximum of 64 (= 79 - 15) character positions. The end-of-line indicator is a carriage return/linefeed combination (<CR><LF>).

The only characters allowed in the file, except where explicitly stated, are the space (ASCII character 32), carriage return (ASCII character 13) and linefeed (ASCII character 10) characters, plus the printable ASCII character set given below:

```
!"#$%&'()*+,-./0123456789:;<=>? @ (ASCII characters 33 - 64)
ABCDEFGHIJKLMNOPQRSTUVWXYZ (ASCII characters 65 - 90)
abcdefghijklmnopqrstuvwxyz (ASCII characters 97 - 122)
[\]^_`{|}~ (ASCII characters 91 - 96, ASCII 123 - 126)
```

Horizontal TAB characters are NOT permitted in this file format as a substitute for spaces or commas. Examples of keywords, separators and data can be found in Table 1, and are further detailed below. The file should be named using the extension .TXT, .EMSA or .MSA.

3.2 Required keywords

The following keywords are required and must appear at the beginning of the file, in the order specified below. Although some of these may appear arbitrary, it is our judgment that they fulfil a long-term need. After several years of students and outside users passing through a laboratory, the result can be a vast number of data files of unknown origin. Unless some adequate form of labelling is imposed from the start, many worthwhile data files are lost, and useless data sit on a disc taking up valuable space. With the following minimum subset of keywords, it should be possible to reconstruct the significance of most spectra.

Note that there must be exactly one of each required keyword, except for the keyword #TITLE which must appear at least once but may, at the user's discretion, appear more than once to provide an extended-length title.

Required keywords:

#FORMAT	= Character string identifies this format as "EMSA/MAS spectral data file".
#VERSION	= File format version number (TC202v2.0 for this implementation).
#TITLE	= Gives a short description of the spectra. One or more per file. Max. = 64 characters.
#DATE	= The calendar day-month-year in which the spectra were recorded, with the month in letters and the day and year in numbers in the form: DD-MMM-YYYY. For example, 07-JUL-2010.
#TIME	= The time of day at which the spectrum was recorded, in 24-hour format: HH:MM.
#OWNER	= The name of the person who recorded the spectrum.

#NPOINTS	= Total number of data points in X & Y data arrays 1. ≤ NPOINTS. The original EMSA standard set an upper limit of 4096. This was considered to be too restrictive and unnecessary for modern computer systems.
#NCOLUMNS	= Number of columns of data 1. ≤ NCOLUMNS < 5. when DATATYPE = Y 1. ≤ NCOLUMNS < 3. when DATATYPE = XY It is strongly recommended that values > 1 are not used since multicolumn formats cause problems when importing data into spreadsheet packages. They were originally intended to make more efficient use of space on teletype printouts of data files.
#XUNITS	= Units for X-axis data, for example: eV.
#YUNITS	= Units for Y-axis data, for example: counts.
#DATATYPE	= Method in which the data values are stored: Y-axis-only values or X,Y data pairs. The current options are the characters Y and XY.
#XPERCHAN	= The number of X-axis units per channel.
#OFFSET	= A real (but possibly negative) number representing value of channel one in the same units as #XUNITS.
#SPECTRUM	= Indicates that the next line starts the spectral data.
#ENDOFDATA	= Indicates the end of the data file.

3.3 Spectral data

The spectral data must be enclosed between the following keywords:

#SPECTRUM	: Signifies the beginning of spectral data (on the next line).
#ENDOFDATA	: Signifies the end of spectral data.

Between these keywords, the spectrum is listed in one of two ways, as specified by the #DATATYPE value.

In the case of spectra with equally spaced x-values (equal increments per channel), up to #NCOLUMNS y-values may be given per line. For Y-axis data type $1 < \text{\#NCOLUMNS} < 5$. Each y-value is either a real number (containing a decimal point, even if there is no fractional component) or is expressed in exponential format (e.g. 3.142E+3), and is followed immediately by a comma. The corresponding x-values can be obtained from the specified values of #XPERCHAN and #OFFSET or #CHOFFSET. Due to the problems multi-column formats pose for importing data into spreadsheet programs, etc., it is strongly recommended that only a single-column format ($\text{\#NCOLUMNS} = 1$) be used.

If the user prefers, data may be stored as (x,y) pairs of data points, at #NCOLUMNS per line. For XY-axis data type $1 \leq \text{\#NCOLUMNS} < 3$. The x- and y-values are expressed as real numbers or in exponential format, and are separated by a comma. Where there is more than one (x,y) pair per line, the (x,y) pairs themselves are separated by a comma followed by at least one blank space. Carriage return and linefeed characters act as a delimiter for the last data value on a line. Additional space or comma delimiters may be added for clarity, but any program reading the data values should treat consecutive delimiters as one. The data value zero must therefore be represented explicitly by at least one "0" text character rather than two consecutive delimiters. Due to the problems multi-column formats pose for importing data into spreadsheet programs, etc., it is strongly recommended that only a single-column format ($\text{\#NCOLUMNS} = 1$) be used.

The reason for avoiding the use of integer numbers is that, in some instances, decimal values are generated, for example if energy-loss spectra have been scaled for normalization or to remove a gain

change. If necessary, the decimal point and the comma can be removed with a text editor (using a global replace) to give integer values, whereas the reverse process may not be straightforward.

Numbers less than unity can be represented either with or without a zero preceding the decimal point (e.g. 0.1 or .1). In the case of negative numbers, there should be no spaces between the minus sign and the numerical value. We recommend that there be no trailing spaces after a number (preceding the comma). The (x,y) option has been included to accommodate segmented spectra, containing gaps where y-data are not specified, and to allow for the possibility that the X-axis scale is non-linear. In addition, it makes the data compatible with most general-purpose graph-plotting software packages. In some future version of the format, this option could be extended by the addition of x and y spatial coordinates, to allow for X-ray maps or energy-selected images, but a more compact representation based on a TIFF standard might be more attractive.

3.4 Optional keywords

All optional keywords except #CHECKSUM and #COMMENT must appear before the keyword #SPECTRUM and after #OFFSET. If used, the checksum value should be the last line in the file and is the integer summation of all characters in the file, excluding the last line containing the checksum itself, and also excluding any trailing spaces after the line terminator. This option is provided to test the integrity of data transmission and/or storage.

With one exception, the keywords listed below require the user to specify an associated value as a real number (with a decimal point and not more than 20 characters in length, including the decimal point) or as an ASCII character string < 64 characters long. The exception is #CHECKSUM, which requires a signed integer value. For ease of classification, we have grouped the optional keywords in order of their function, thus keywords dealing with the spectrum, type of spectroscopy and specimen are presented together in the following listing. This grouping is not required within the file format, but it is strongly recommended. In the keyword list below, the abbreviation [RN] means the keyword is a real number while [nCS] indicates a character string of *n* characters.

Keywords relating mainly to spectrum characteristics:

- #SIGNALTYPE = Type of spectroscopy. Allowed values are [3CS].
 EDS = Energy-dispersive spectroscopy
 WDS = Wavelength-dispersive spectroscopy
 ELS = Energy-loss spectroscopy
 CLS = Cathodoluminescence spectroscopy
 GAM = Gamma-ray spectroscopy
- #XLABEL = X-axis label [64CS].
- #YLABEL = Y-axis label [64CS].
- #CHOFFSET = A real (but possibly negative) number representing the channel number whose value corresponds to zero units on the X-axis scale.
- #COMMENT = Comment line [64CS].
 The comment keyword may be repeated as often as desired within header lines of the file.

Keywords relating mainly to the microscope/instrument:

- #BEAMKV = Accelerating voltage of instrument in kilovolts [RN].
- #EMISSION = Gun emission current in microamps [RN].
- #PROBECUR = Probe current in nanoamps [RN].
- #BEAMDIAM = Diameter of incident probe in nanometres [RN].

ISO 22029:2012(E)

- #MAGCAM = Magnification or camera length [RN].
Magnification in \times or times, camera length in millimetres.
- #CONVANGLE = Convergence semi-angle of incident beam in milliradians [RN].
- #OPERMODE = Operating mode. Allowed values are [5CS].
IMAGE = Imaging mode
DIFFR = Diffraction mode
SCIMG = Scanning imaging mode
SCDIF = Scanning diffraction mode

Keywords relating mainly to the specimen:

- #THICKNESS = Specimen thickness in nanometres [RN].
- #XTILTSTGE = Specimen stage tilt X-axis in degrees relative to the beam normal [RN].
- #YTILTSTGE = Specimen stage tilt Y-axis in degrees relative to the beam normal [RN].
- #XPOSITION = Specimen/beam position along the X-axis [RN].
- #YPOSITION = Specimen/beam position along the Y-axis [RN].
- #ZPOSITION = Specimen/beam position along the Z-axis [RN].

Keywords relating mainly to ELS:

- #DWELLTIME = Dwell time/channel for serial data collection in milliseconds [RN].
- #INTEGTIME = Integration time per spectrum for parallel data collection in milliseconds [RN].
- #COLLANGLE = Collection semi-angle of scattered beam in milliradians [RN].
- #ELSDET = Type of ELS detector. Allowed values are [6CS].
SERIAL = Serial ELS detector
PARALL = Parallel ELS detector

Keywords relating mainly to EDS:

- #ELEVANGLE = Elevation angle of EDS/WDS detector in degrees [RN].
- #AZIMANGLE = Azimuthal angle of EDS/WDS detector in degrees [RN].
- #SOLIDANGLE = Collection solid angle of detector in steradians [RN].
- #LIVETIME = Signal processor active (live) time in seconds [RN].
- #REALTIME = Total clock time used to record the spectrum in seconds [RN].
- #TBEWIND = Thickness of Be window on detector in centimetres [RN].
- #TAUWIND = Thickness of Au window/electrical contact in centimetres [RN].
- #TDEADLYR = Thickness of dead layer in centimetres [RN].
- #TACTLYR = Thickness of active layer in centimetres [RN].
- #TALWIND = Thickness of aluminium window in centimetres [RN].
- #TPYWIND = Thickness of pyrolene window in centimetres [RN].

#TBNWIND	= Thickness of boron-nitride window in centimetres [RN].
#TDIWIND	= Thickness of diamond window in centimetres [RN].
#THCWIND	= Thickness of hydrocarbon window in centimetres [RN].
#EDSDET	= Type of X-ray detector. Allowed values are [5CS]. SIBEW = Si(Li) with Be window SIUTW = Si(Li) with ultra-thin window SIWLS = Si(Li) windowless GEBEW = Ge with Be window GEUTW = Ge with ultra-thin window GEWLS = Ge windowless SDBEW = SDD with Be window SDUTW = SDD with ultra-thin window SDWLS = SDD windowless
#CHECKSUM	= 32-BIT INTEGER NUMBER.

The #CHECKSUM optional keyword and values are used to test for data transmission errors in the file. The checksum is calculated by adding the ASCII values of all ASCII characters in the file, including spaces and line terminators. Trailing blanks (on each line) and the last line containing the checksum itself are excluded from this sum. If the calculated checksum agrees with that stored in the data file, the user can be reasonably sure that data set has been faithfully transmitted.

Optional user-defined keywords may also be included within the header of the data file. These keywords must begin with the double ## sign and follow all EMSA/MAS-defined keywords, except for the start-of-data #SPECTRUM keyword. All field restrictions as outlined above apply to user-defined keywords and their values. These user-defined keywords allow expansion of this file structure to cover items that have not yet been explicitly defined by the EMSA/MAS task force. To aid in consistency of use, the following optional user-defined keywords are specified:

##TCONLYR	= Thickness of contact layer in centimetres [RN].
##TMCONLYR	= Contact layer metal.
##FILENAME	= Name of data file. Should have the extension .TXT, .EMSA or .MSA.

The following keywords have been added to allow for non-ASCII text fields to be stored with data:

##TITLE	= Additional TITLE field, which is permitted to be in non-ASCII characters.
##OWNER	= Additional OWNER field, which is permitted to be in non-ASCII characters.
##XLABEL	= Additional XLABEL field, which is permitted to be in non-ASCII characters.
##YLABEL	= Additional YLABEL field, which is permitted to be in non-ASCII characters.
##COMMENT	= Additional COMMENT field, which is permitted to be in non-ASCII characters.
##CHARSET	= An unambiguous identification of the non-ASCII character set used in one or more of the ##TITLE, ##OWNER, ##XLABEL, ##YLABEL, or ##COMMENT fields. Where more than one non-ASCII character set is used, a ##CHARSET statement must immediately follow on the line below each non-ASCII field and identify the non-ASCII character set in the field that immediately precedes it. The list of acceptable character sets is defined by the Internet Assigned Numbers Authority (http://www.iana.org/assignments/character-sets).

3.5 Ending a file

The last line of every file should be either an #ENDOFDATA or a #CHECKSUM line. The #ENDOFDATA keyword must immediately follow the spectral data. If present, the #CHECKSUM keyword must follow the #ENDOFDATA keyword.

4 Example of a data file in the EMSA/MAS spectral format

The following example shows a spectral file with the minimum-required keyword information. Table 1 shows an ELS data set using the (x,y) pair data format.

Table 1 — Example of a single-column (x,y) data set

```
#FORMAT      : EMSA/MAS spectral data file
#VERSION     : TC202v2.0
#TITLE      : NIO EELS OK SHELL
#DATE       : 01-OCT-1991
#TIME       : 12:00
#OWNER      : EMSA/MAS TASK FORCE
#NPOINTS    : 21.
#NCOLUMNS  : 1.
#XUNITS     : Energy loss (eV)
#YUNITS     : Intensity
#DATATYPE   : XY
#XPERCHAN   : 3.1
#OFFSET     : 520.13
#CHOFFSET   : -168
#SIGNALTYPE : ELS
#XLABEL     : Energy
#YLABEL     : Counts
#BEAMKV     : 120.0
#EMISSION   : 5.5
#PROBECUR   : 12.345
#BEAMDIA    : 100.0
#MAGCAM     : 100.
#CONVANGLE  : 1.5
#COLLANGLE  : 3.4
```

```

#OPERMODE      : IMAG
#THICKNESS     : 50.
#DWELLTIME     : 100.
#ELSDET        : SERIAL
#SPECTRUM      : Spectral data start here
520.13,        4066.0
523.22,        3996.0
526.32,        3932.0
529.42,        3923.0
532.51,        5602.0
535.61,        5288.0
538.70,        7234.0
541.80,        7809.0
544.90,        4710.0
547.99,        5015.0
551.09,        4366.0
554.18,        4524.0
557.28,        4832.0
560.38,        5474.0
563.47,        5718.0
565.79,        5034.0
568.89,        4651.0
571.99,        4613.0
574.31,        4637.0
577.40,        4429.0
580.50,        4217.0
#ENDOFDATA     : Spectral data end here

```


Bibliography

- [1] Zaluzec N.J., In the Computer Corner, *EMSA Bulletin*, **17:1** (1987), pp. 93-94, and RWEMMPDL Abstract, *EMSA Bulletin*, **17:2** (Nov 1987), p. 92
- [2] McDonald R.S., Wilks P.A., JCAMP-DX: A Standard for Exchange of Infrared Spectra in Computer Readable Form, *Appl. Spectrosc.*, **42** (1988), pp. 151-162
- [3] Lyman, C., Task Force Committee Report, *EMSA Bulletin*, **19** (1989), pp. 97-100
- [4] Hunt J.A., Fiori C.E., A Proposed EMSA/MAS Data Format for the Transfer of Spectra, *EMSA Bulletin*, **19** (1989), pp. 100-105
- [5] Zaluzec N.J., RWEMMPDL-VERSION 1.1 Abstract, *EMSA Bulletin*, **19** (1989), p. 114
- [6] Dench W.A., Hazel L.B., Seah M.P. and the VAMAS Community, VAMAS Surface Chemical Analysis Standard Data Transfer Format with Skeleton Decoding Programs, *Surf. Interface Anal.*, **13** (1988), pp. 63-122

