INTERNATIONAL STANDARD

ISO 22029

Third edition
2022-10

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 22029:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

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

This third edition cancels and replaces the second edition (ISO 22029:2012), which has been technically revised.

The main changes are as follows.

— The file encoding in this version has been changed from ASCII to UTF-8. This allows for the use of non-ASCII characters in specified descriptive fields but still maintains backwards compatibility with older data reading and writing software since the ASCII coding is a sub-set of the UTF-8 coding. Most of the keyword value character limits have also been removed in this version. These were originally imposed to aid readability of files output to teletype printers and are no longer deemed a necessary restriction.

— The definition of the #OFFSET keyword has been modified to specify that the value of the first spectral channel is taken from the channel's mid-point.

— A new required keyword, #TIMEZONE, has been added and the #TIME keyword is now explicitly required to be local time to ensure that the time and date the data has been recorded can be unambiguously determined.

— Eight new optional keywords have been added: #ROTATION, #WORKDIST, #CRC32C, ##MNKRESOL, ##SUPPAREA, ##SUPPTHICK, ##SUPPMAT and ##DETPOS.

— The new optional error checking keyword, #CRC32C, has been added to provide a more robust error checking option than the original #CHECKSUM value.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

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 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 (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.

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

1 Scope

This document 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 enables different laboratories to freely exchange spectral data, and helps 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 Normative references

There are no normative references in this document.

3 Terms and definitions

No terms and definitions are listed in this document.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

4 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 also enables 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. Most importantly, it offers a method for archiving data that does not rely on the continued availability of a given item of equipment or version of software to remain accessible.

An ideal solution is 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. 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 shall 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. 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 the 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 the 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. A preliminary version of this proposal was circulated to several manufacturers of XEDS systems which have received comments and suggestions, many of which have been incorporated into this document.

5 Format description

5.1 General

The general structure of the data file format can be summarized as a simple sequential text file using the UTF-8 encoding. 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 in 5.2 to 5.4. 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 as follows:

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 9). If a position in the keyword field is not used, it shall 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. The previous version of this standard stipulated a line limit of 79 characters for legibility on 'standard width CRT screens' but this is no longer deemed necessary. The end-of-line indicator is a carriage return/linefeed combination (<CR><LF>).

The file encoding shall be UTF-8. However, the only characters allowed in the file, except where explicitly stated otherwise, are the space (UTF-8 character 32), carriage return (UTF-8 character 13) and linefeed (UTF-8 character 10) characters, plus the printable UTF-8 character set given as follows:

!"#$%&'()*+,-./0123456789:; < = > ? @ (UTF-8 characters 33 - 64)

ABCDEFGHIJKLMNOPQRSTUVWXYZ (UTF-8 characters 65 - 90)

abcdefghijklmnopqrstuvwxyz (UTF-8 characters 97 - 122)

[\] ^_`{|} ~ (UTF-8 characters 91 - 96, UTF-8 123 - 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. The file should be named using the extension .TXT, .EMSA or .MSA.

5.2 Required keywords

The following keywords are required and shall appear at the beginning of the file, in the order specified in Table 1. Although some of these may appear arbitrary, it is judged 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 shall be exactly one of each required keyword, except for the keyword #TITLE which shall appear at least once but may, at the user's discretion, appear more than once to provide an extended- length title.

Table 1 — Required keywords

#FORMAT	Character string identifies this format as "EMSA/MAS Spectral Data File"
#VERSION	File Format Version Number (TC202v3.0 for this implementation)
#TITLE	Gives a short description of the spectra. One or more per file.
#DATE	The calendar day-month-year in which the spectra was recorded, with the month in letters and the day and year in numbers in the form: DD-MMM-YYYY. For example, 07-JUL-2021.
#TIME	The local time of day at which the spectrum was recorded, in 24-h format: HH:MM.
#TIMEZONE	Difference of local time relative to UTC in hours. For non-integer differences this shall be expressed in decimal format. For example, India has a local time which is 5:30 ahead of UTC and have a #TIMEZONE value of 5.5
#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 is 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 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 as: 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 the mid-point value of channel one in the same units as #XUNITS.
#SPECTRUM	Indicates the next line starts the spectral data. Note that it is permitted to add descriptive text after the colon on this line (see the example dataset in Table 9)
#ENDOFDATA	Indicates the end of the data file. Note that it is permitted to add descriptive text after the colon on this line (see the example dataset in Table 9)

5.3 Spectral data

The spectral data shall 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 < #NCOLUMNS < 5. Each y-value is either a real number (containing a decimal point as the decimal separator if required) 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.

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 ≤ #NCOLUMNS < 3. The x- and y-values are expressed as real numbers or in exponential format and are separated by a comma. To avoid confusion or ambiguity, a comma shall not be used as either a decimal or thousands separator in the numerical values, only as the separator between values on a line. 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 shall therefore be represented explicitly by at least one "0" text character rather than two consecutive delimiters.

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.

Numerical values between 1 and -1 can be represented either with or without a zero preceding the decimal point (e.g. -0,1 or -.1 are equivalent and allowable). 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 nonlinear. In addition, it makes the data compatible with most general-purpose graph-plotting software packages. In some future version of the format, this option can 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.

5.4 Optional keywords

All optional keywords except #CHECKSUM, #CRC32C and #COMMENT shall appear before the keyword #SPECTRUM and after #OFFSET. If used, the #CHECKSUM or #CRC32C value should be the last line in the file and is the checksum value for the file, excluding the last line containing the checksum itself, and excluding any trailing spaces after the line terminator. This option is provided to test the integrity of data transmission and/or storage. Note either #CHECKSUM or #CRC32C may be used, but not both.

With one exception, the keywords listed in this subclause (see Tables 2 to 8) require the user to specify an associated value as a real number or as a permitted character string (see 5.1). 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, specimen are presented together in the following listing. This grouping is not required within the file format; however, it is strongly recommended. In the keyword list in Tables 3 to 7 the abbreviation [RN] means the keyword is a real number while [nCS] indicates a character string of *n* characters.