



Standard Specification for Analytical Data Interchange Protocol for Mass Spectrometric Data¹

This standard is issued under the fixed designation E 2077; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This specification covers a standardized format for mass spectrometric data representation and a software vehicle to effect the transfer of mass spectrometric data between instrument data systems. This specification provides a protocol designed to benefit users of analytical instruments and increase laboratory productivity and efficiency.

1.2 The protocol in this specification provides a standardized format for the creation of raw data files, library spectrum files or results files. This standard format has the extension “.cdf” (derived from NetCDF). The contents of the file include typical header information like instrument, sample, and acquisition method description, followed by raw, library or processed data. Once data have been written or converted to this protocol, they can be read and processed by software packages that support the protocol.

1.3 This specification does not provide for the storage of data acquired simultaneously to and integrated with the mass spectrometric data, but on other detectors; for example attached to the mass spectrometer’s liquid or gas chromatographic system. Related Specification E 1947 and Guide 1948 describe the storage of 2-dimensional chromatographic data.

1.4 The software transfer vehicle used for the protocol in this specification is NetCDF, which was developed by the Unidata Program and is funded by the Division of Atmospheric Sciences of the National Science Foundation.²

1.5 The protocol in this specification is intended to (1) transfer data between various vendors’ instrument systems, (2) provide Laboratory Information Management Systems (LIMS) communications, (3) link data to document processing applications, (4) link data to spreadsheet applications, and (5) archive analytical data, or a combination thereof. The protocol is a consistent, vendor independent data format that facilitates the analytical data interchange for these activities.

1.6 The protocol consists of:

1.6.1 This specification on mass spectrometric data, which gives the full definitions for each one of the generic mass

spectrometric data elements used in implementation of the protocol. It defines the analytical information categories, which are a convenient way for sorting analytical data elements to make them easier to standardize.

1.6.2 Guide E 2078 on mass spectrometric data, which gives the full details on how to implement the content of the protocol using the public-domain NetCDF data interchange system. It includes a brief introduction to using NetCDF and describes an API (Application Programming Interface) that is intended to be incorporated into application programs to read or write NetCDF files. It is intended for software implementors, not those wanting to understand the definitions of data in a mass spectrometric dataset.

1.6.3 NetCDF Users Guide.

2. Referenced Documents

2.1 ASTM Standards:

E 2078 Guide for Analytical Data Interchange Protocol for Mass Spectrometric Data³

E 1443 Terminology Relating to Building and Accessing Materials and Chemical Databases⁴

E 1947 Specification for Analytical Data Interchange Protocol for Chromatographic Data³

E 1948 Guide for Analytical Data Interchange Protocol for Chromatographic Data³

2.2 Other Standards:

NetCDF User’s Guide⁵

Occupational Safety and Health Administration (OSHA) Standards-29 CFR part 1910⁶

IEEE 488⁷

IEEE -802⁷

EIA - 232⁸

2.3 ISO Standards:⁹

8601:1988 Data elements and interchange formats (First

¹ This specification is under the jurisdiction of ASTM Committee E01 on Analytical Chemistry for Metals, Ores and Related Materials and is the direct responsibility of Subcommittee E01.25 on Laboratory Data Interchange and Information Management.

Current edition approved March 10, 2000. Published July 2000.

² For more information on the NetCDF standard, contact Unidata at www.unidata.ucar.edu.

³ *Annual Book of ASTM Standards*, Vol 03.06.

⁴ *Annual Book of ASTM Standards*, Vol 14.01.

⁵ Available from Russell K. Rew, Unidata Program Center, University Corporation for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.

⁶ Occupational Safety and Health Administration, U.S. Department of Labor.

⁷ Institute of Electrical and Electronics Engineers, Inc., 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331.

⁸ Electronics Industries Alliance, 2500 Wilson Blvd., Arlington, VA 22201.

⁹ Available from ISO, 1 Rue de Varembe, Case Postale 56, CH 1211, Geneva, Switzerland.

edition published 1988-06-15; with Technical Corrigendum 1 published 1991-05-01)

639: 1988 Code for the representation of names of languages

9000 Quality Management Systems

ISO/IEC - 8802

3. Terminology

3.1 *Analytical Information Classes*—The Mass Spectrometry Information Model categorizes mass spectrometric information into a number of information “classes.” There is not a direct mapping of these classes into the implementation categories described further below. The implementation categories describe the information hierarchy; the classes describe the contents within the hierarchy. The model presented here only partially addresses these classes. In particular, the last two (Processed Results and Component Quantitation Results) are not described at all. Only Implementation Category 1 is required for compliance within this specification. Information about the other implementation categories is provided for historical interest. The classes defined here are:

3.1.1 *Administrative*—information for administrative tracking of experiments.

3.1.2 *Instrument-ID*—information about the instrument that generally does not change from experiment to experiment.

3.1.3 *Sample Description*—information describing the sample and its history, handling and processing.

3.1.4 *Test Method*—all information used to generate the raw data and processed results. This includes instrument control, detection, calibration, data processing and quantitation methods.

3.1.5 *Raw Data*—the data as stored in the data file, along with any parameters needed to describe it.

3.1.6 *Processed Results*—processing information and values derived from the raw data.

3.1.7 *Component Quantitation Results*—individual quantitation results for components in a complex mixture.

3.2 *Definitions for Administrative Information Class*—These definitions are for those data elements that are implemented in the protocol. See Table 1.

TABLE 1 Administrative Information Class

NOTE 1—Particular analytical information categories (C1, C2, C3, C4, or C5) are assigned to each data element under the Category column. The meaning of this category assignment is explained in Section 5.

NOTE 2—The Required column indicates whether a data element is required, and if required, for which categories. For example, M1234 indicates that that particular data element is required for any dataset that includes information from Category 1, 2, 3, or 4. M4 indicates that a data element is only required for Category 4 datasets.

NOTE 3—Unless otherwise specified, data elements are generally recorded to be their actual test values, instead of the nominal values that were used at the initiation of a test.

NOTE 4—A table is not to be interpreted as a table of keywords. The software implementation is independent of the data element names used here, and is in fact quite different. Likewise, the datatypes given are not an implementation representation, but a description of the form of the data element name. That is, a data element labeled as floating point may, for example, be implemented as a double precision floating point number; in this document, it is sufficient to note it as floating point without reference to precision.

Data Element Name	Datatype	Category	Required
dataset-completeness	string	C1	M12345
protocol-template-revision	string	C1	M12345
netcdf-revision	string	C1	M12345
languages	string	C1 or C5	...
administrative-comments	string	C1 or C2	...
dataset-origin	string	C1	M4
dataset-owner	string	C1	...
dataset-date-time-stamp	string	C1	M1234
injection-date-time-stamp	string	C1	M1234
experiment-title	string	C1	...
experiment-cross-references	string array[n]	C3 or C4	...
operator-name	string	C1	M4
experiment-type	string	C1 or C4	...
pre-experiment-program-name	string	C2 or C5	...
post-experiment-program-name	string	C2 or C5	...
number-of-times-processed	integer	C5	...
number-of-times-calibrated	integer	C5	...
calibration-history	string array[n]	C5	...
source-file-reference	string	C5	M4
source-file-format	string	C5	...
source-file-date-time-stamp	string	C5	M4
external-file-references	string array[n]	C5	...
error-log	string	C5	...

3.2.1 *administrative-comments*—comments about the dataset identification of the experiment. This free text field is for anything in this information class that is not covered by the other data elements in this class.

3.2.2 *calibration-history*—an audit trail of file names and data sets which records the calibration history; used for Good

Laboratory Practice (GLP) compliance.

3.2.3 *dataset-completeness*—indicates which analytical information categories are contained in the dataset. The string should exactly list the category values, as appropriate, as one or more of the following “C1+C2+C3+C4+C5,” in a string separated by plus (+) signs. This data element is used to check for completeness of the analytical dataset being transferred.

3.2.4 *dataset-date-time-stamp*—indicates the absolute time of dataset creation relative to Greenwich Mean Time. Expressed as the synthetic datetime given in the form: YYYYMMDDhhmmss ±ffff.

3.2.4.1 *Discussion*—This is a synthesis of ISO 8601, which compensates for local time variations.

3.2.4.2 *Discussion*—The YYYYMMDDhhmmss expresses the local time, and time differential factor (ffff) expresses the hours and minutes between local time and the Coordinated Universal Time (UTC or Greenwich Mean Time, as disseminated by time signals), as defined in ISO 8601. The time differential factor (ffff) is represented by a four-digit number preceded by a plus (+) or a minus (–) sign, indicating the number of hours and minutes that local time differs from the UTC. Local times vary throughout the world from UTC by as much as –1200 h (west of the Greenwich Meridian) and by as much as +1300 h (east of the Greenwich Meridian). When the time differential factor equals zero, this indicates a zero hour, zero minute, and zero second difference from Greenwich Mean Time.

3.2.4.3 *Discussion*—An example of a value for a datetime would be: 1991,08,01,12:30:23-0500 or 19910801123023-0500. In human terms this is 23 s past 12:30 PM on August 1, 1991 in New York City. Note that the –0500 h is 5 full hours time behind Greenwich Mean Time. The ISO standard permits the use of separators as shown, if they are required to facilitate human understanding. However, separators are not required and consequently shall not be used to separate date and time for interchange among data processing systems.

3.2.4.4 *Discussion*—The numerical value for the month of the year is used, because this eliminates problems with the different month abbreviations used in different human languages.

3.2.5 *dataset-origin*—name of the organization, address, telephone number, electronic mail nodes, and names of individual contributors, including operator(s), and any other information as appropriate. This is where the dataset originated.

3.2.6 *dataset-owner*—name of the owner of a proprietary dataset. The person or organization named here is responsible for this field’s accuracy. Copyrighted data should be indicated here.

3.2.7 *error-log*—information that serves as a log for failures of any type, such as instrument control, data acquisition, data processing or others.

3.2.8 *experiment-cross-references*—an array of strings which reference other related experiments.

3.2.9 *experiment-title*—user-readable, meaningful name for the experiment or test that is given by the scientist.

3.2.10 *experiment-type*—name of the type of data stored in this file. Select one of the types in the following list.

3.2.10.1 *Discussion*—The valid types are:

centroided mass spectrum—a data set containing centroided single or multiple scan mass spectra. This includes selected ion monitoring/recording (SIM/SIR) data, represented as mass-intensity pairs. This is the default.

continuum mass spectrum—a data set containing single or multiple scan mass spectra in continuum (non-centroided or profile) form. Scans are represented as mass-intensity pairs, whether incrementally spaced or not.

library mass spectrum—a data set consisting of one or more spectra derived from a spectral library. This is distinguished from an experimental mass spectral data set in that each spectrum in the library set has associated chemical identification and other information.

3.2.10.2 *Discussion*—A required Raw Data Information parameter, the number of scans, is used to define the shape of the data in the file, that is, to differentiate between single and multiple spectrum files. Another parameter, the scan number, is used to determine whether multiple scan files have an order or relatedness between scans.

3.2.10.3 *Discussion*—Some instruments are capable of mixed mode data acquisition, for example, alternating positive/negative EI (Electron Ionisation) or CI (Chemical Ionisation) scans. In order to keep this interchange standard as simple as possible, **each scan mode must be treated as a separate data set** regardless of how the data are actually stored in the source data file. Alternating positive/negative EI data, for example, will generate *two* interchange files (possibly simultaneously, depending on the implementation); one for the positive EI scans and one for the negative EI scans. These files may be made mutually cross-referential using their “external-file-references” fields.

3.2.11 *external-file-references*—an array of strings listing file names referred to from within the raw data file. These could include, for example, tune parameter, method, calibration, reference, sequence, or other files. NetCDF files produced in parallel (such as paired files containing alternating EI/CI scans) should be cross-referenced here.

3.2.12 *injection-date-time-stamp*—indicates the absolute time of sample injection relative to Greenwich Mean Time. Expressed as the synthetic datetime given in the form: YYYYMMDDhhmmss ±ffff. See *dataset-date-time-stamp* for details of the ISO standard definition of a date-time-stamp.

3.2.13 *languages*—optional list of natural (human) languages and programming languages delineated for processing by language tools.

3.2.13.1 *ISO-639-language*—indicates a language symbol and country code from Annex B and D of the ISO-639 Standard.

3.2.13.2 *other-language*—indicates the languages and dialect using a user-readable name; applies only for those languages and dialects not covered by ISO 639 (such as programming language).

3.2.14 *netcdf-revision*—current revision level of the NetCDF data interchange system software being used for data transfer.

3.2.15 *number-of-times-calibrated*—also for GLP compliance, a count of the number of times the data were calibrated before yielding the final results.

3.2.16 *number-of-times-processed*—for GLP compliance, a count of the number of times the data were processed to yield the final results recorded in this file. An audit trail of the file

names of previous processing must be provided.

3.2.17 *operator-name*—name of the person who ran the equipment, which acquired the current dataset.

3.2.18 *post-experiment-program-name*—name(s) of any program(s) used to process raw data after acquisition.

3.2.19 *pre-experiment-program name*—name(s) of any program(s) run prior to the start of acquisition.

3.2.20 *protocol-template-revision*—revision level of the template being used by implementers. This needs to be included to tell users which revision of E 2077 should be referenced for the exact definitions of terms and data elements used in a particular dataset; for example “1.0.”

3.2.21 *source-file-date-time-stamp*—the date and time at which the source file was created. This has the same format as described above for the “experiment-date-time-stamp” field.

3.2.22 *source-file-format*—a string which describes the format of the data file used to produce the interchange file, for example: “HP ChemStation,” “VG Opus I,” “Finnigan IN-COS,” etc.

3.2.23 *source-file-reference*—adequate information to locate the original dataset. This information makes the dataset self-referenced for easier viewing and provides internal documentation for GLP-compliant systems.

3.2.23.1 *Discussion*—This data element should include the complete filename, including node name of the computer system. For UNIX this should include the full path name. For VAX/VMS this should include the node-name, device-name, directory-name, and file-name. The version number of the file (if applicable) should also be included. For personal computer networks this needs to be the server name and directory path.

3.2.23.2 *Discussion*—If the source file was a library file, this data element should contain the library name and serial number of the dataset.

3.3 *Definitions for Instrument-ID Information Class*—This class contains the generally experiment-independent information describing the instrument(s) on which the experiment was performed. Because each subcomponent of an instrument may require separate identification, the “instrument-component-...” data element names in Table 2 should be interpreted as occurring once for each identified component. Not all data element names may be relevant for each component.

TABLE 2 Instrument ID Information Class

Data Element Name	Datatype	Category	Required
instrument-component-number	integer	C5	M5
instrument-component-name	string	C5	M5
instrument-component-id	string	C5	M5
instrument-component-manufacturer	string	C4 or C5	M5
instrument-component-model-number	string	C4 or C5	M5
instrument-component-serial-number	string	C5	M5
instrument-component-id-comments	string	C5	M5
instrument-component-software-version	string	C2 or C5	M5
instrument-component-firmware-version	string	C2 or C5	M5
operating-system-revision	string	C5	M5
application-software-revision	string	C5	M5

3.3.1 *application-software-revision*—the name, revision

level, and (optionally, if different from the component manufacturer) manufacturer of each software module (if any) used in acquisition and processing of the data by the data system. This data element name applies only to data system instrument components. Required for GLP compliance.

3.3.2 *instrument-component-firmware-version*—the revision level of the instrument component firmware (if any) when the data were acquired. This data element name applies only to non-data system instrument components. This becomes an Implementation Category 2 field when the revision level affects the data acquisition, processing, or results. An example might be the revision level of a read-only memory (ROM) chip contained on an imbedded controller board.

3.3.3 *instrument-component-id*—the laboratory’s identification code for the instrument component; this might be an internal inventory control number.

3.3.4 *instrument-component-id-comments*—any free-form comments not covered in one of the other fields.

3.3.5 *instrument-component-manufacturer*—the name of the manufacturer of the instrument component. Version 1.0 does not specify an enumerated list; vendor implementations of the specification are expected to standardize on a convention.

3.3.6 *instrument-component-model-number*—the model number or name, or both, used by the manufacturer to identify the instrument component.

3.3.7 *instrument-component-name*—the generic descriptive name of the instrument component. Version 1.0 does not specify an enumerated list of component names, but a future version may. For example: “gas chromatograph,” “data system,” “GC column,” “MS core.”

3.3.8 *instrument-component-number*—provides an index number for the particular instrument component being identified. Note that the total number of instrument components is implicit, and therefore instrument components must be sequentially numbered, beginning with zero.

3.3.9 *instrument-component-serial-number*—the manufacturer’s serial number, if any, for the instrument component.

3.3.10 *instrument-component-software-version*—the revision level of the instrument component software (if any) when the data were acquired. This data element name applies only to non-data system instrument components. This becomes an Implementation Category 2 field when the revision level affects the data acquisition, processing, or results. An example might be a software program for chromatograph run control downloaded from a host data system.

3.3.11 *operating-system-revision*—the name and revision level of the data system’s operating system software (if any) when the data were acquired and processed. This data element name applies only to data system instrument components, of which there might be more than one for hyphenated instruments. Required for GLP compliance.

3.4 *Definition for Sample Description Information Class*—This class contains mostly comment-style information concerning the sample itself, and is intended to be used for minimal GLP compliance. As this standard matures, more explicit chemical method information may be included here. See Table 3.

TABLE 3 Sample-Description Information Class

Date Element Name	Datatype	Category	Required
Sample-owner	string	C5	
sample-receipt-date-time-stamp	string	C5	
internal-sample-id	string	C1	
external-sample-id	string	C5	
sampling-procedure-name	string	C5	
Sample-preparation-procedure	string	C4	
Sample-state	string	C4	
Sample-matrix	string	C4	
Sample-storage-information	string	C5	
Sample-disposal-information	string	C5	
Sample-history	string	C5	
Sample-preparation-comments	string	C5	
Sample-id-comments	string	C5	
manual-handling-precautions	string	C5	

3.4.1 *external-sample-id*—the number or code assigned to the sample by the submitter or submitter’s organization.

3.4.2 *internal-sample-id*—the number or code used to identify the sample within the mass spectrometry laboratory or in a LIMS used by the laboratory.

3.4.3 *manual-handling-precautions*—any safety issues which are of concern when the sample is manually handled.

3.4.3.1 *Discussion*—A future version of this interchange specification, which deals more fully with GLP, will likely be expanded to address other sample management issues.

3.4.4 *sample-disposal-information*—a description of the disposal procedure for the sample (also in accord with the United States Department of Labor Occupational Safety and Health Administration (OSHA) regulations).

3.4.5 *sample-history*—a description of the history of this particular sample, including any special handling, treatments, etc. to distinguish it from others from the same batch.

3.4.6 *sample-id-comments*—any comments not covered elsewhere. This might include laboratory notebook references, etc.

3.4.7 *sample-matrix*—a string describing the natural matrix from which the sample was selected. In a future revision, this field will be made an enumerated set.

3.4.8 *sample-owner*—the name of the sample owner or submitter. This may be different from the data set owner.

3.4.9 *sample-preparation-comments*—any comments concerning preparation not covered in other fields.

3.4.10 *sample-preparation-procedure*—a textual description of the procedure used to prepare the sample for analysis.

3.4.11 *sampling-procedure-name*—the name of the procedure used to select a sample from its natural (bulk) matrix. For example: “supercritical fluid extraction.” This will be made a formal set of choices in a future revision.

3.4.12 *sample-receipt-date-time-stamp*—the date and time the sample was received in the laboratory or submitted for analysis. The ISO 8601 format is used for this field. This date and time is usually earlier than the data set date/time stamp, and may be important when analysis of a sample must occur within a specified period after receipt.

3.4.13 *sample-state*—a string field, specified as one of these choices:

Sample State

- solid
- liquid
- gas
- supercritical fluid
- plasma
- other state

3.4.14 *sample-storage-information*—a description of the storage conditions for the sample, which includes the storage location. This is for OSHA compliance.

3.5 *Definitions for Test Method Information Class*—This class contains the information required to reconstruct the sampling and acquisition of the raw data once the sample has been prepared for analysis. See Table 4.

NOTE 1—None of these data elements are required to be present in the file; where the data element is important to the interpretation of the raw data but is not present, a default value is assumed. The default value for a data element is given in **boldface** type where it is defined.

TABLE 4 Test Method Information Class

Data Element Name	Datatype	Category	Required
separation-experiment-type	string	C1	
mass-spectrometer-inlet	string	C1	
mass-spectrometer-inlet-temperature	float	C1	
ionization-mode	string	C1	
ionization-polarity	string	C1	
electron-energy	float	C1	
laser-wavelength	float	C1	
reagent-gas	string	C1	
reagent-gas-pressure	float	C1	
FAB-type	string	C1	
FAB-matrix	string	C1	
source-temperature	float	C1	
filament-current	float	C1	
emission-current	float	C1	
accelerating-potential	float	C1	
detector-type	string	C1	
detector-potential	float	C1	
detector-entrance-potential	float	C1	
resolution-type	string	C1	
resolution-method	string	C1	
scan-function	string	C1	
scan-direction	string	C1	
scan-law	string	C1	
scan-time	float	C1	
mass-calibration-file-name	string	C1	
external-reference-file-name	string	C1	
instrument-reference-file-name	string	C1	
instrument-parameter-comments	string	C1	

3.5.1 *accelerating-potential*—this field specifies the accelerating potential in volts.

3.5.2 *detector-entrance-potential*—for detectors in which it is appropriate, this field specifies the (signed) potential at the entrance to the detector relative to system ground, in volts.

3.5.3 *detector-potential*—for detectors in which it is appropriate, this field specifies the (signed) potential across the detector, in volts. Examples include electron multipliers and conversion dynodes.

3.5.4 *detector-type*—this specifies the detection method used, and is chosen from the following set.

Detector Type

electron multiplier

- photomultiplier
- Focal plane array
- faraday cup
- conversion dynode electron multiplier
- conversion dynode photomultiplier
- multi-collector
- other detector

3.5.5 *electron-energy*—this field is relevant for electron impact ionization mode, and contains the electron energy in volts.

3.5.6 *emission-current*—this field gives the filament emission current in microamps. This is also relevant principally for EI and CI ionization.

3.5.7 *external-reference-file-name*—this field specifies the name of an external file which contains the reference spectrum of the material used as an external mass calibrant.

3.5.8 *FAB-matrix*—this field specifies the fast atom bombardment (FAB) matrix used, if any, for the FAB experiment type.

3.5.9 *FAB-type*—this field is relevant for fast atom bombardment, and specifies the atom or neutral used in the bombardment gun.

3.5.10 *filament-current*—this field gives the filament input current in amps. This is primarily relevant for EI and CI ionization modes.

3.5.11 *instrument-parameter-comments*—this is a catch-all field; it might contain instrument tuning parameters, vacuum system pressures, or any other parameter which might be of use in reconstructing the acquisition which is not covered above. As this specification is made more GLP-compliant in later versions, additional formal fields may be defined which contain information on such instrument parameters.

3.5.12 *internal-reference-file-name*—this field specifies the name of an external file which contains the reference spectrum of the material used as an internal calibrant.

3.5.13 *ionization-mode*—this field describes the technique used to ionize the sample. It is also a string, chosen from the following set. **Only one ionization mode is supported per interchange file.**

Ionization Method

electron impact

- chemical ionization
- fast atom bombardment
- field desorption
- field ionization
- electrospray ionization
- thermospray ionization
- atmospheric pressure chemical ionization
- plasma desorption
- laser desorption
- spark ionization
- thermal ionization
- other ionization

3.5.14 *ionization-polarity*—this field describes the polarity of the detected ions and is chosen from the set that follows. **Only one ionization polarity is supported per interchange file.**

Ionization Polarity

positive

negative

3.5.15 *laser-wavelength*—this field is relevant for laser desorption ionization, and contains the laser wavelength in nanometers.

3.5.16 *mass-calibration-file-name*—this field gives the name of the external file which contains the voltage to mass, time to mass, or other mass calibration data.

3.5.17 *mass-spectrometer-inlet*—this field describes the sample introduction interface. It has a string value, from the set:

Mass Spectrometer Inlet

- membrane separator
- capillary direct
- open split
- jet separator
- direct inlet probe**
- septum
- particle beam
- reservoir
- moving belt
- atmospheric pressure chemical ionization
- flow injection analysis
- electrospray inlet
- infusion
- thermospray inlet
- other probe inlet
- other inlet

Electrospray includes ion spray, and is used to describe both the inlet as well as the ionization technique.

3.5.18 *mass-spectrometer-inlet-temperature*—this field specifies the temperature of the spectrometer inlet, if appropriate, in degrees centigrade.

3.5.19 *reagent-gas*—this field is relevant for chemical ionization mode, and specifies the CI reagent gas.

3.5.20 *reagent-gas-pressure*—in CI mode, this specifies the pressure of the CI reagent gas. Units will be agreed upon as part of the implementation.

3.5.21 *resolution-method*—specifies the method for determining spectrometer resolution. For example: “10 % peak valley,” “50 % peak height,” “90 % peak height.”

3.5.22 *resolution-type*—this field specifies the type of instrument resolution: constant over the mass range or proportional to mass. It is chosen from the set that follows. See the description of *resolution*, in the *Raw Data Per-Scan Information* section, (3.8) that follows.

Resolution Type

constant

proportional

3.5.23 *scan-direction*—this field specifies the direction in which the mass range was scanned during acquisition and is chosen from the following set. **It is not necessarily the same direction in which masses are recorded in the interchange file. Masses are always recorded in ascending order in the interchange file.**

Scan Direction

up

down
other direction

3.5.24 *scan-function*—a string specifying an entry from the

following set. Only two scan functions are specifically identified in this version. The mass scan function implies full mass range recording. Selected ion detection is known by various names: selected ion monitoring, selected ion recording, multiple ion detection, etc.

Scan Function

- mass scan**
- selected ion detection
- other function

3.5.25 *scan-law*—this field specifies the mass scan law as a string chosen from the following set:

Scan Law

- linear**
- exponential
- quadratic
- other law

3.5.26 *scan-time*—Specifies the time, in seconds, required to complete one scan of the mass range. This field may not be as precise as the “scan duration” field accompanying each scan.

3.5.27 *separation-experiment-type*—a separation experiment performed as an integral part of the sample introduction is specified here. One from the following set should be chosen:

Separation Experiment Type

- gas-liquid chromatography
- gas-solid chromatography
- normal phase liquid chromatography
- reverse phase liquid chromatography
- ion exchange liquid chromatography
- size exclusion liquid chromatography
- ion pair liquid chromatography
- other liquid chromatography
- supercritical fluid chromatography
- thin layer chromatography
- field flow fractionation
- capillary zone electrophoresis
- other chromatography
- no chromatography**

3.5.28 *source-temperature*—this field gives the temperature of the source in degrees centigrade.

3.6 *Raw Data Information Classes*—These classes contain information generated during the acquisition of the raw data. The parameters are used in the interpretation and further processing of the raw data. The Raw Data Classes have several parts: a global part, which contains information relevant to all the scans in a data set; one or more raw data per-scan parts, each of which contains information relevant to a particular scan; and for library data, one or more library data per-scan parts which occur together with a raw data per-scan part and which contain additional information associated with the library entry. The specification supports both mass and time axis data (either separately or in combination); if both data are supplied, it is assumed that the mass axis has been mass-measured from the time data.

3.7 *Raw Data Global Information Class*—This class contains information relevant to all scans in a data set. See Table 5.

TABLE 5 Raw Data Global Information Class

Data Element Name	Datatype	Category	Required
number-of-scans	integer	C1	M1
starting-scan-number	integer	C1	
number-of-scan-groups	integer	C1	
mass-axis-scale-factor	float	C1	(M1) ^A
time-axis-scale-factor	float	C1	(M1) ^A
intensity-axis-scale-factor	float	C1	(M1) ^A
intensity-axis-offset	float	C1	(M1) ^A
mass-axis-units	string	C1	
time-axis-units	string	C1	
intensity-axis-units	string	C1	
total-intensity-units	string	C1	
mass-axis-data-format	string	C1	(M1) ^A
time-axis-data-format	string	C1	(M1) ^A
intensity-axis-data-format	string	C1	
mass-axis-label	string	C1	
time-axis-label	string	C1	
intensity-axis-label	string	C1	
mass-axis-global-range	float array[2]	C1	(M1) ^A
time-axis-global-range	float array[2]	C1	(M1) ^A
intensity-axis-global-range	float array[2]	C1	
calibrated-mass-range	float array[2]	C1	
actual-run-time-length	float	C1	(M1) ^A
actual-delay-time	float	C1	(M1) ^A
uniform-sampling-flag	boolean	C1	(M1) ^A
raw-data-global-comments	string	C1	

^A These fields are required if mass and time data are present.

3.7.1 *actual-run-time-length*—this field contains the run time, in seconds, between the start of the experiment to the end. For chromatography/MS experiments, for example, this is the time between the injection and the acquisition of the last scan in the data set.

3.7.2 *actual-delay-time*—this field contains the time in seconds between the start of the experiment (for example, the injection) and the start of scan acquisition. Actual delay time plus sampling period should result in the actual run time length.

3.7.3 *calibrated-mass-range*—this field contains the mass range (in low mass, high mass order) over which mass axis calibration is valid.

3.7.4 *intensity-axis-data-format*—this field specifies the format (data type) of the ordinate values as recorded in this file. The same table as for mass axis data format is used. By default, **long** format is assumed.

3.7.4.1 *Discussion*—The ability to choose the data format for abscissa and ordinate permits the construction of an exchange file tailored to the size of the data it contains. For example, nominal mass low-mass data might be most economically stored in 16-bit integer format, while accurate mass high-mass data might require the precision of full 64-bit floating point numbers. These flags guide the exchange file access software to use the proper function to retrieve the raw data.

3.7.5 *intensity-axis-global-range*—this field contains the maximum range of the intensity axis data in low intensity, high intensity order.

3.7.6 *intensity-axis-label*—this field contains the string used to label the intensity axis when plotting file data.

3.7.7 *intensity-axis-offset*—this specifies a constant quantity (in raw data intensity units) which is added to the intensity values as recorded in this file to obtain the actual intensity values as acquired. The intensity offset is added to the intensity

value **after** the scaling factor is applied. The default intensity axis offset is **0.0**.

3.7.8 *intensity-axis-scale-factor*—this specifies a scaling factor to be applied to the intensity axis data. The raw data intensity values as recorded in this file are multiplied by this factor to yield the actual intensity values as acquired. The default intensity axis scaling factor is **1.0**.

3.7.9 *intensity-axis-units*—this field specifies the units for the raw data intensity axis values and is chosen from the following set. The default is “**arbitrary units**” (unitless).

Intensity Axis Units

arbitrary units
counts per second
total counts
volts
current
other units

3.7.10 *mass-axis-data-format*—this field specifies the format (data type) of the mass axis values as recorded in this file. It is a string name from the following table of data types. The 16-bit integer short format is assumed by default.

Name	Data Format
short	16-bit signed integer
long	32-bit signed integer
float	32-bit float
double	64-bit float

3.7.11 *mass-axis-global-range*—this field contains the maximum range of the mass axis data in low mass, high mass order. Although scan range may vary on a scan-by-scan basis, some data systems require advance knowledge of the maximum expected mass range in order to properly assemble mass data. This field is required if mass axis data are present.

3.7.12 *mass-axis-label*—this field contains the string used to label the mass axis when plotting the file data.

3.7.13 *mass-axis-scale-factor*—this specifies a scaling factor to be applied to the mass axis data. The raw data mass values as recorded in this file are multiplied by this factor to yield the actual mass values as acquired. The default mass axis scaling factor is **1.0**.

3.7.14 *mass-axis-units*—this field specifies the units for the raw data mass axis values and is chosen from the following set. The default is “**m/z**” (AMU/charge).

Mass Axis Units

m/z
arbitrary units
other units

3.7.15 *number-of-scan-groups*—this field applies only for experiments in which the scan function is Selected Ion Detection and specifies the number of distinct groups of masses monitored during the course of the experiment. This field is not applicable for other scan function types. A scan group is considered distinct if either the masses, sampling- or delay-times for a mass, or the scan period, during which the masses are monitored, is unique.

3.7.16 *number-of-scans*—this specifies the total number of scans recorded in this file. It is a required parameter.

3.7.17 *raw-data-global-comments*—this string holds any

comments relevant to the raw data not covered by the previous fields.

3.7.18 *starting-scan-number*—in the case where the source data file is only partially converted into interchange format, this specifies the index of the starting scan (relative to the source data file) of the first scan in the interchange file. By default, it is assumed that the first scan in the interchange file corresponds to the first scan in the source data file.

3.7.19 *time-axis-data-format*—this field specifies the format (data type) of the time axis values as recorded in this file. The choices are the same as those for mass-axis-data-format. By default, **short** format is assumed.

3.7.20 *time-axis-global-range*—this field contains the maximum range of the time axis data in start time, stop time order. Although scan range may vary on a scan-by-scan basis, some data systems require advance knowledge of the maximum expected time axis range in order to properly assemble mass data. This field is required if time axis data are present.

3.7.21 *time-axis-label*—this field contains the string used to label the time axis when plotting the file data.

3.7.22 *time-axis-scale-factor*—this specifies a scaling factor to be applied to the time axis data. The raw data time values as recorded in this file are multiplied by this factor to yield the actual time values as acquired. The default time axis scaling factor is **1.0**.

3.7.23 *time-axis-units*—this field specifies the units for the raw data time axis values and is chosen from the following set. The default is “**seconds**.”

Time Axis Units

seconds
arbitrary units
other units

3.7.24 *total-intensity-units*—this field specifies the units for the raw data total intensity values. The default is “**arbitrary units**” (unitless). The same table as for intensity-axis-units applies.

3.7.25 *uniform-sampling-flag*—this field specifies whether the scans in a multiple-scan set are sampled uniformly in time. If the field has a TRUE value, uniform sampling is assumed. A FALSE value specifies non-uniform sampling. In this case, each scan must be accompanied by a scan acquisition time value. The default for this field is TRUE (uniform sampling).

3.8 *Raw Data Per-Scan Information Class*—Data elements in this class may vary on a scan-by-scan basis, or contain information relevant only to a specific scan or library entry. See Table 6.

TABLE 6 Raw Data Per-Scan Information Class

Data Element Name	Datatype	Category	Required
scan-number	integer	C1	M1
actual-scan-number	integer	C1	
number-of-points	integer	C1	M1
mass-axis-values	mass data format array	C1	M1 ^A
time-axis-values	time data format array	C1	M1 ^A
intensity-axis-values	intensity data format array	C1	M1
number-of-flags	integer	C1	
flagged-peaks	integer array	C1	

Data Element Name	Datatype	Category	Required
flag-values	integer array	C1	
total-intensity	float	C1	
a/d-sampling-rate	float	C1	
a/d-co-addition-factor	integer	C1	
scan-acquisition-time	float	C1	
scan-duration	float	C1	
mass-scan-range	float array[2]	C1	
time-scan-range	float array[2]	C1	
inter-scan-time	float	C1	
resolution	float	C1	

^A These fields are required if mass and time data are present.

3.8.1 *actual-scan-number*—this field specifies the actual scan number in the source data file and provides for the case where only part of the source data file is converted into interchange format. If not specified, it will assume the value of *scan-number*.

3.8.2 *a/d-co-addition-factor*—this field specifies the number of A/D samples which are co-added or averaged to produce a single datum point.

3.8.3 *a/d-sampling-rate*—this field specifies the rate (in kilohertz) at which A/D (analog-to-digital) conversions are made.

3.8.4 *flagged-peaks*—this is an array, of dimension *number-of-flags*. The datum point values are the indices (starting at zero) into the mass and time arrays of the peaks which are flagged for that scan. For example, if the first, fifth, and sixth peaks are flagged, then the flagged peaks array will contain three points, with values (1,5,6).

3.8.5 *flag-values*—flag values are characteristic of individual mass or time datum points within a scan. A scan can have multiple peak flags, and any one mass or time datum may have a flag which is a composite of several applicable flags. The flag value datum points in the flag values array correspond one-to-one with the peaks identified in the *flagged-peaks* array. The following flags have been defined, and represent a composite of those used by vendors.

Name	Description
NOT HIGH RESOLUTION	The peak is nominal mass peak (in an otherwise high resolution scan)
MISSED REFERENCE	A reference peak was missed prior to this peak
UNRESOLVED	Peak is an unresolved multiplet
DOUBLY CHARGED	Peak is doubly-charged (that is, has fractional mass)
REFERENCE EXCEPTION	Peak is a reference from the reference file
LOCK MASS	Peak is a reference from the exception file
SATURATED	Peak is a reference mass used to adjust the mass scale during/after acquisition
SATURATED	Peak intensity is saturated (overflows A/D conversion or storage range)
SIGNIFICANT	Peak is a Biller-Biemann significant peak
MERGED	Peak is a composite of two centroided peaks merged during processing
FRAGMENTED	Peak is very wide and generated more than one centroided peak
AREA/HEIGHT	Peak intensity is based on integrated area or height determined through centroiding
MATH MODIFIED	Accurate mass assignment or peak intensity is based on mathematical processing
NEGATIVE INTENSITY	Peak intensity is negative as a result of processing (subtraction or other correction)
EXTENDED ACCURACY	Mass accuracy is derived through mathematical processing
CALCULATED	Peak is artificial (was created through mathematical processing; for example, isotope calculation)

3.8.6 *intensity-axis-values*—this is an array, of dimension

number-of-points, containing the intensity values in *intensity-data-format* data type. It parallels the mass and time axis values arrays (that is, the *n*th entry in the intensity axis array matches the *n*th entry in the mass and time axis arrays). This is also a required field.

3.8.7 *inter-scan-time*—specifies the time delay, in seconds, between the end of one scan and the start of the next for multiple-scan acquisitions.

3.8.8 *mass-axis-values*—this is an array, of dimension *number-of-points*, containing the mass values in *mass-data-format* data type. This is a required field if time data are not present. **Mass axis data must be recorded in low mass to high mass order in the interchange file, regardless of how they were actually acquired.**

3.8.9 *mass-scan-range*—specifies the starting and ending masses of the scan range (in low mass, high mass order). **This is not the same as the minimum and maximum mass datum values in the scan.**

3.8.10 *number-of-flags*—mass or time datum points within a scan may have associated peak flags. This number (generally zero for most normal scans) contains the number of datum points with flags in this scan.

3.8.11 *number-of-points*—this specifies the number of mass-time-intensity triplets, and is a required field.

3.8.12 *resolution*—this field specifies the mass resolution. Resolution can be determined in one of two ways: for instruments with constant proportional mass resolution (such as magnetic sector instruments), resolution is specified in parts per million (mass/D mass); for instruments with constant absolute mass resolution (such as quadrupoles), resolution is specified as mass/charge (m/z). See **resolution type** and **resolution method** (in the “Test Method” section) for the parameters which specify what type of instrumental resolution this value specifies, and how it is determined from a typical peak.

3.8.13 *scan-acquisition-time*—a floating point field which specifies the time (in seconds) from the start of the run (not the start of actual acquisition) at which acquisition of this particular scan was started. It is recognized that a scan requires a finite amount of time to acquire, and that different data systems record the “scan acquisition time” in various ways (start of scan, midpoint of scan, etc.). To force standardization, the interchange specification defines “scan acquisition time” as stated above. For accuracy, implementations which use a different definition should correct their stored time when recording an interchange file.

3.8.14 *scan-duration*—the actual time, in fractional seconds, required to acquire this scan. Data systems which record this value in “clock ticks” must convert to seconds. This avoids an additional field to provide the clock tick period.

3.8.15 *scan-number*—an integer which specifies the index of this scan within the set of scans. For multiple-scan data sets, this is a required field. The first scan in the set has index one (1).

3.8.16 *time-axis-values*—this is an array, of dimension *number-of-points*, containing the time values in *time-data-format* data type. This is an optional field when mass data are present. Time axis data are recorded in increasing time order.

3.8.17 *time-scan-range*—specifies the starting and ending times of the scan range. This is not necessarily the same as the minimum and maximum time datum values in the scan.

3.8.18 *total-intensity*—specifies the total intensity associated with this scan. For a chromatography/MS data set, this series of intensities is used to construct the TIC (total ion current) chromatogram.

3.9 *Library Data Per-Scan Information Class*—Fields in this class occur only for interchange files of the **Library Mass Spectrum** experiment type. Each library spectrum in the file may have values for any or all of these fields. See Table 7.

TABLE 7 Library Data Per-Scan Information Class

Data Element Name	Datatype	Category	Required
entry-name	string	C1	M1
entity-id	string	C1	
original-entry-number	integer	C1	
source-data-file-reference	string	C1	
CAS-name	string	C1	
other-names	string array [n]	C1	
CAS-number	integer	C1	
chemical-formula	string	C1	
Wiswesser-notation	string	C1	
SMILES-notation	string	C1	
MOLfile-reference-name	string	C1	
other-structure-notation	string	C1	
retention-index	float	C1	
retention-index-type	string	C1	
absolute-retention-time	float	C1	
relative-retention	float	C1	
retention-reference-name	string	C1	
retention-reference-CAS-number	integer	C1	
melting-point	float	C1	
boiling-point	float	C1	
chemical-mass	float	C1	
nominal-mass	integer	C1	
accurate-mass	float	C1	
other-information	string	C1	

3.9.1 *absolute-retention-time*—this field contains the absolute retention time (in seconds), measured from the start of the chromatographic experiment in which the library spectrum was acquired.

3.9.2 *accurate-mass*—this field specifies the exact mass of the entry, based on the carbon = 12 scale, and using the accurate mass of the most abundant isotope of each element.

3.9.3 *boiling point*—this field specifies the boiling point, in degrees Centigrade.

3.9.4 *CAS-name*—this string gives the name of the entry recognized by the Chemical Abstracts Service.

3.9.5 *CAS-number*—this is the Chemical Abstracts Service registry number for the library entry, if any.

3.9.6 *chemical-formula*—this string gives the chemical formula for the entry, if any.

3.9.7 *chemical-mass*—this field specifies the chemical mass, computed using the average atomic masses for each element in the formula.

3.9.8 *entry-id*—this field specifies a non-name data element name of the library entry, such as a user-, corporate-, or library-defined registry code for the library entry or the sample which was used to generate the library entry. An example is the NIST accession number.

3.9.9 *entry-name*—this field specifies the name of the entry, as found in the library. It may not be the same as the CAS name. This string is a required field.

3.9.10 *melting-point*—this field contains the melting point, in degrees Centigrade.

3.9.11 *MOLfile-reference-name*—this string specifies the name of an external file containing chemical structure information for the entry in Molecular Design Limited MOL file format. The specification does not require that data systems on the receiving end of such a file be able to interpret the data contained in it; this field simply allows explicit reference to such an associated file.

3.9.12 *nominal-mass*—this field specifies the integer nominal mass of the entry, using the integer mass of the most abundant isotope of each element in the formula.

3.9.13 *original-entry-number*—this field specifies the index number of the entry as contained in the original (source) library. This number may not have relevance outside the scope of the library, but serves only as a reference back to the source of the entry.

3.9.14 *other-information*—some spectral libraries allow association of user-supplied information with entries. This field contains this descriptive information.

3.9.15 *other-names*—this is an array of strings, and specifies additional names by which this entry is known.

3.9.16 *other-structure-notation*—this string specifies structural information in an ASCII format other than SMILES or Wiswesser. For the present, this provides a mechanism for providers of spectral libraries who use an alternative means of associating structures with spectra to distribute those structures in a NetCDF format. The library provider must specify the format of this field so that the structures can be extracted.

3.9.17 *relative-retention*—this field contains the retention (unitless) of the library spectrum relative to the spectrum of a reference material. The reference material is identified by the **retention reference name** and **retention reference CAS number fields**.

3.9.18 *retention-index*—this field contains the retention index for the entry. The standard by which this index was determined is contained in the **retention index type** field.

3.9.19 *retention-index-type*—this field contains the method by which retention index was determined, for example: “Kovats.”

3.9.20 *retention-reference-CAS-number*—This field specifies the Chemical Abstracts Service registry number for the reference compound used in measurement of the relative retention of the library spectrum.

3.9.21 *retention-reference-name*—this field specifies the name of the reference material used in measurement of the relative retention of the library spectrum.

3.9.22 *SMILES-notation*—this string specifies the SMILES notation for the entry.

3.9.23 *source-data-file-reference*—this field provides a reference to the source data file used to create this library entry (not the library from which the interchange file was extracted—see the field “source file reference” in the Administrative Information class). An example is the original data file name and scan number(s) from which the spectral data were extracted.

3.9.24 *Wiswesser-notation*—this field specifies the Wiswesser notation for the entry.

3.10 *Raw Data Per Scan-Group Information Class*—Fields in this class occur only for interchange files of the **selected ion detection** scan function type. This class is not used for experiments of other scan function types. It is not mandatory that interchange files recording selected ion detection function type data contain this information, but inclusion is recommended to assist in accurately reconstructing the experimental conditions. See Table 8.

TABLE 8 Raw Data Per-Scan-Group Information Class

Data Element Name	Datatype	Category	Required
number-of-masses-in-group	integer	C1	M1
starting-scan-number	integer	C1	M1
group-masses	float array [n]	C1	M1
sampling-times	float array [n]	C1	
delay-times	float array [n-1]	C1	

3.10.1 *delay-times*—this field is an array, containing the delay time (in seconds) between the end of monitoring the corresponding mass in the *group-masses* array and the start of monitoring for the next mass. Note that there is no delay time for the last mass in the group.

3.10.2 *group-masses*—this field is an array, containing the masses (in M/Z units) monitored in this group. Masses are in floating point format, and are recorded in the order in which they are monitored. This array is a mandatory field.

3.10.3 *number-of-masses-in-group*—this specifies the count of masses monitored during this scan group. This field is mandatory.

3.10.4 *sampling-times*—this field is an array, containing the sampling time or monitoring period (in seconds) for the corresponding mass in the **group-masses** array.

3.10.5 *starting-scan-number*—this specifies the scan number (relative to the interchange file, not the source data file) at which this scan group starts. The scan group remains the current group until the starting scan number for the next scan group is encountered. This field is mandatory.

3.10.6 *Discussion*—The sum of all *sampling-times* and *delay-times* for the currently active scan group and the *Inter-scan-time* (from the *Raw Data Per-Scan Information* class) for the scan within the group should equal the *scan-duration* (also from the *Raw Data Per-Scan Information* class.)

4. Objectives and Features of the Analytical Data Interchange Protocol

4.1 Functional Objectives:

4.1.1 *Data Types*—The analytical data interchange protocol for a mass spectrometric data specification is not meant to be an all-encompassing model of every mass spectrometric technique and data type. Rather, the initial goal is to provide a framework for describing the following essential types of mass spectral data:

4.1.1.1 Single or multiple scan, centroided or continuum (profile) full scan data sets.

4.1.1.2 Single or multiple group selected ion monitoring (SIM/SIR) data sets.

4.1.1.3 Single or multiple entry, mass spectrum libraries.

4.2 Technical Objectives:

4.2.1 *Standards Development and Systems Selection*—The technical goals have been to develop a protocol for analytical

data representation and interchange that meets the following criteria:

4.2.1.1 Easy to use by software developers and end users,

4.2.1.2 Readable by humans using some facile mechanism,

4.2.1.3 Open, extensible, and maintainable,

4.2.1.4 Applies to multidimensional,

4.2.1.5 Independent of any particular communication link, such as EIA-232, IEEE-488, IEEE-802 series, ISO/IEC 8802 series, etc.,

4.2.1.6 Independent of a particular operating system like DOS, OS/2, UNIX, VMS, MVS, Windows, etc.,

4.2.1.7 Independent of any particular vendor, and acceptable and usable by all,

4.2.1.8 Coexists with, and does not negate, other standards,

4.2.1.9 Designed for the long-term and implemented for use in the short-term, and

4.2.1.10 Works well for mass spectrometry and does not preclude extensions to other analytical technique families.

4.2.2 *Data Integrity Across Heterogeneous Systems*—The current implementation specifies a mechanism with particular directionality for data transfer integrity. The protocol has unidirectional data integrity for data transfers between heterogeneous systems. This is because source systems and target systems are made by different manufacturers, or if the systems are from the same manufacturers, they may use different hardware or algorithms. An example would be data transfer from Vendor A's data system running on a DOS-based personal computer to Vendor Z's LIMS running on a Unix-based minicomputer; another would be transfer between mass spectrometric data systems made by different manufacturers.

4.2.2.1 If the receiving system has algorithms that assume a different analog-to-digital (ADC) converter word length from the sending system, and it calculates results based on its own, different data precision and accuracy, then the accuracy and precision of the original data may not be maintained. For example, if the sending system has an algorithm that assumes a 24-bit internal representation, and the receiving system has an algorithm that assumes a 20-bit internal representation, one may lose data accuracy and precision. If calculations are done by the receiving system, and the data are then sent back to the source system for their calculations, data integrity may not be maintained. Thus, there is an inherent directionality to data transfer given by different algorithms and different hardware systems.

4.2.2.2 The protocol for mass spectrometry data can be used for data round-trips relative to the source system, for example, from the source system to an archive and then back to the source system again. Such round-trip data transfers will maintain data integrity as long as there was no calculation or alteration of the data during transfer that would alter its accuracy or precision.

4.2.2.3 Thus, the protocol is bi-directional for homogeneous source-system round trips and unidirectional for heterogeneous source-to-target transfers.

4.2.2.4 The first implementation allows transfer of mass spectrometry raw data. Plotting or listing of raw data on other vendor's data systems (for comparison purposes) are possible in the first implementation.

4.2.3 *Algorithmic Issues*—Algorithmic issues are not addressed at all by this specification. Users cannot expect to get the same exact processed results from systems that use completely different algorithms.

4.2.4 *Absolute Scaling of Raw Data*—Absolute scaling of raw data across different manufacturer's systems is not possible at this time, due to the lack of general-purpose algorithms that can convert and scale data of different internal representations, from different data acquisition systems, and different computer hardware systems.

4.2.5 *Requirements*—The protocol in this specification does not yet specify all elements needed to meet documentation quality data requirements (Good Laboratory Practices or ISO 9000 – the specific standards will be defined when taking up this requirement).

4.3 *Technical Features of the Protocol:*

4.3.1 *Separation of Concept from Implementation*—There is a clean separation of the protocol into contents (the data definitions within a data model) and container (the data interchange system). This is important because it effectively decouples concept from implementation. Computer technology is changing much more rapidly than analytical data definitions, which are stabilizing for the maturing analytical instrument industry. Producing an accurate analytical information model and having well-defined definitions for data elements within that model actually have higher long-term significance than any particular data interchange system technology.

4.3.2 *General Technical Features*—Two general technical features stand out:

4.3.2.1 *Analytical Information Categories*—A convenience for simplifying the work of developing analytical data specifications. These five categories were chosen based on three practical considerations: (1) which data is of interest to transfer most routinely, (2) which can be standardized most easily in the short-term, and (3) which can be standardized in the long term. The analytical information categories are explained later in Section 5.

4.3.2.2 *The Data Interchange System*—The container used to communicate data between applications, in a way that is independent of both computer platforms and end-user applications. The system has software routines that are used to read, write, and manipulate data in analytical datasets. It has a data access interface, called an Application Programming Interface (API).

4.3.2.3 The data interchange system that most closely fits the scientific and software engineering requirements for a public-domain data interchange software system is the NetCDF (network Common Data Form) system. The Unidata Corporation, which supports the National Center for Atmospheric Research, is the source of NetCDF. NetCDF is copyrighted by the Unidata Corporation. The Protocol used the NetCDF system for its implementation. Engineering tests that prove the applicability of NetCDF for analytical data applications have been completed. An overview of NetCDF is given in Guide E 2078.

5. Analytical Information Categories

5.1 Data and information usage varies widely in complexity and completeness. Information is therefore sorted into logical

categories, called the Analytical Information Categories. These categories serve two very useful purposes.

5.1.1 First, the categories sort analytical information into convenient sets to allow more rapid standardization. This has made it easier for implementors to produce working demonstrations, without the burden and complexity of the hundreds of data elements contained in a full dataset for any given analytical technique.

5.1.2 Second, the categories accommodate different organizations' usage of information more easily. Some organizations may only want to transfer raw data among data systems. Others may want to transfer information to a LIMS or other database systems. Still others may want to build databases of chemical methods, instrument methods, or data processing methods. The first version of the protocol is for a single sample injection, not for sequences of samples.

5.1.3 The information contained in this specification represents the greatest common subset of information end-user and vendor requirements available at this time.

5.2 *Category 1: Raw Data Only*—Category 1 is used for transferring raw data. It includes raw data, units, and relative data scaling information. This will allow accurate replotting of the spectrum, chromatogram and/or reprocessing. Category 1 also contains administrative information needed to locate the original chemical and data processing methods used with this dataset.

5.3 *Category 2: Final Results*—All post-quantitation calculated results are included. This information category includes the amounts and identities (if determinable) of each component in a sample. Final sample peak processing results, component identities, sample component amounts, and other derived quantities of interest to the analyst are included in Category 2 datasets. Quantitation decisions are included here as comments to aid the analyst in determining how the results were calculated.

5.3.1 Category 2 datasets can be used to transfer data to database management systems, such as a LIMS, research database, or sample tracking systems. It can also be used to transfer data to data analysis packages, spreadsheets, visualization packages, or other software packages.

5.4 *Category 3: Full Data Processing Method*—Quantitation decisions and data processing methods are transferred in this category. Quantitatively correct data/information transfer is achieved by the category for all parameters necessary to do peak detection, measurement, and response factor calculation, and calibration for a sequence of related sample runs. This applies to both samples and reference standards. Sample quantitation results are not included here; those are in Category 2. Peak processing method parameters, response factor calculation and other calibration method parameters required to quantitate sample component peaks are included in Category 3.

5.5 *Category 4: Full Chemical Method*—All chemical method information needed to repeat the experiment under exactly the same chemical conditions is included in this category.

5.6 *Category 5: Good Laboratory Practice Information*—

Any additional information required to satisfy Good Laboratory Practices or ISO-9000 requirements are included in this category. This category generally deals with capturing product, process, and documentation quality information needed for validation.

data interchange protocol; detection method; information class; ISO; mass spectrometer; mass spectrometric; NetCDF; peak-processing-results; raw data; sample description

6. Keywords

6.1 analytical information categories; chromatographic;

The American Society for Testing and Materials takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.

This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.

This standard is copyrighted by ASTM, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or service@astm.org (e-mail); or through the ASTM website (www.astm.org).