

# **Specification**

## **Analytical Data Interchange for Mass Spectrometry (ANDI/MS)**

David Stranz, Fisons Instruments  
Scott Campbell, Fisons Instruments  
Ray Christopher, Finnigan MAT  
Jim Watt, PE Nelson  
Don Zakett, Dow Chemical Company

**Version 1.0**

### **Revision History:**

Version 1.0: May 1994

**Copyright 1994  
Analytical Instrument Association**

## Table of Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Overview	1
1.2. The netCDF Data Interchange System	1
1.3. Other Data Interchange Models	2
1.4. The Analytical Instrument Association	2
<b>2. Purpose of this Document</b>	<b>3</b>
<b>3. Mass Spectrometry Information Model</b>	<b>4</b>
3.1. Analytical Information Categories	4
3.2. Analytical Information Classes	5
<b>4. Mass Spectrometry Information Classes</b>	<b>6</b>
<b>4.1. Administrative Information Class</b>	<b>7</b>
4.1.1. data set completeness	7
4.1.2. template revision level	7
4.1.3. administrative comments	7
4.1.4. data set origin	7
4.1.5. data set owner	8
4.1.6. experiment title	8
4.1.7. experiment date/time stamp	8
4.1.8. experiment type	8
4.1.9. experiment cross-references	9
<b>4.1.10. netCDF file date/time stamp</b>	<b>9</b>
4.1.11. netCDF revision level	9
4.1.12. operator name	9
4.1.13. source file reference	9
4.1.14. source file format	9
4.1.15. source file date/time stamp	9
4.1.16. external file references	9
4.1.17. languages	9
4.1.18. number of times processed	9
4.1.19. number of times calibrated	10
4.1.20. calibration history	10
4.1.21. post-experiment program name	10
4.1.22. pre-experiment program name	10
4.1.23. error log	10
<b>4.2. Instrument-ID Information Class</b>	<b>11</b>
4.2.1. instrument component number	11
4.2.2. instrument component name	11
4.2.3. instrument component id	11
4.2.4. instrument component manufacturer	11
4.2.5. instrument component model number	11
4.2.6. instrument component serial number	11
4.2.7. instrument component id comments	11
4.2.8. instrument component software revision	12

4.2.9. instrument component firmware revision _____	12
4.2.10. operating system revision _____	12
4.2.11. application software revision _____	12
<b>4.3. Sample Description Information Class _____</b>	<b>13</b>
4.3.1. sample owner _____	13
4.3.2. sample receipt date/time stamp _____	13
4.3.3. internal sample id _____	13
4.3.4. external sample id _____	13
4.3.5. sampling procedure name _____	13
4.3.6. sample preparation procedure _____	13
4.3.7. sample state _____	14
4.3.8. sample matrix _____	14
4.3.9. sample storage information _____	14
4.3.10. sample disposal information _____	14
4.3.11. sample history _____	14
4.3.12. sample preparation comments _____	14
4.3.13. sample id comments _____	14
4.3.14. manual handling precautions _____	14
<b>4.4. Test Method Information Class _____</b>	<b>15</b>
4.4.1. separation experiment type _____	16
4.4.2. mass spectrometer inlet _____	16
4.4.3. mass spectrometer inlet temperature _____	16
4.4.4. ionization mode _____	17
4.4.5. ionization polarity _____	17
4.4.6. electron energy _____	17
4.4.7. laser wavelength _____	17
4.4.8. reagent gas _____	17
4.4.9. reagent gas pressure _____	17
4.4.10. FAB type _____	17
4.4.11. FAB matrix _____	17
4.4.12. source temperature _____	17
4.4.13. filament current _____	18
4.4.14. emission current _____	18
4.4.15. accelerating potential _____	18
4.4.16. detector type _____	18
4.4.17. detector potential _____	18
4.4.18. detector entrance potential _____	18
4.4.19. resolution type _____	18
4.4.20. resolution method _____	18
4.4.21. scan function _____	19
4.4.22. scan direction _____	19
4.4.23. scan law _____	19
4.4.24. scan time _____	19
4.4.25. mass calibration file name _____	19
4.4.26. external reference file name _____	19
4.4.27. internal reference file name _____	19
4.4.28. instrument parameter comments _____	19
<b>4.5. Raw Data Information Classes _____</b>	<b>20</b>
<b>4.6. Raw Data Global Information Class _____</b>	<b>20</b>
4.6.1. number of scans _____	20
4.6.2. starting scan number _____	21
4.6.3. number of scan groups _____	21
4.6.4. mass axis scale factor _____	21

4.6.5. time axis scale factor	21
4.6.6. intensity axis scale factor	21
4.6.7. intensity axis offset	21
4.6.8. mass axis units	21
4.6.9. time axis units	21
4.6.10. intensity axis units	22
4.6.11. total intensity units	22
4.6.12. mass axis data format	22
4.6.13. time axis data format	22
4.6.14. intensity axis data format	22
4.6.15. mass axis label	22
4.6.16. time axis label	22
4.6.17. intensity axis label	22
4.6.18. mass axis global range	23
4.6.19. time axis global range	23
4.6.20. intensity axis global range	23
4.6.21. calibrated mass range	23
4.6.22. actual run time	23
4.6.23. actual delay time	23
4.6.24. uniform sampling flag	23
4.6.25. raw data global comments	23
<b>4.7. Raw Data Per-Scan Information Class</b>	<b>24</b>
4.7.1. scan number	24
4.7.2. actual scan number	24
4.7.3. number of points	24
4.7.4. mass axis values	24
4.7.5. time axis values	25
4.7.6. intensity axis values	25
4.7.7. number of flags	25
4.7.8. flagged peaks	25
4.7.9. flag values	26
4.7.10. total intensity	27
4.7.11. a/d sampling rate	27
4.7.12. a/d coaddition factor	27
4.7.13. scan acquisition time	27
4.7.14. scan duration	27
4.7.15. mass scan range	27
4.7.16. time scan range	27
4.7.17. inter-scan time	27
4.7.18. resolution	27
<b>4.8. Library Data Per-Scan Information Class</b>	<b>28</b>
4.8.1. entry name	28
4.8.2. entry id	28
4.8.3. original entry number	28
4.8.4. source data file reference	28
4.8.5. CAS name	29
4.8.6. other names	29
4.8.7. CAS number	29
4.8.8. chemical formula	29
4.8.9. Wiswesser notation	29
4.8.10. SMILES notation	29
4.8.11. MOLfile reference name	29
4.8.12. other structure notation	29
4.8.13. retention index	29

4.8.14. retention index type _____	29
4.8.15. absolute retention time _____	29
4.8.16. relative retention _____	29
4.8.17. retention reference name _____	29
4.8.18. retention reference CAS number _____	29
4.8.19. melting point _____	30
4.8.20. boiling point _____	30
4.8.21. chemical mass _____	30
4.8.22. nominal mass _____	30
4.8.23. accurate mass _____	30
4.8.24. other entry information _____	30
<b>4.9. Raw Data Per-Scan-Group Information Class _____</b>	<b>31</b>
4.9.1. number of masses in group _____	31
4.9.2. starting scan number _____	31
4.9.3. group masses _____	31
4.9.4. sampling times _____	31
4.9.5. delay times _____	31
<b>5. Acknowledgments _____</b>	<b>32</b>
<b>6. References _____</b>	<b>33</b>

## 1. Introduction

### 1.1. Overview

For several years, members of the ASMS have been working to define a mechanism to permit interchange of mass spectral data between various mass spectrometer data systems. This specification represents the first released results of these efforts. The core working group is composed primarily of the authors of this document, but an extended set of some 100 or more interested individuals representing the breadth of the international mass spectrometric community has been involved as reviewers of previous versions of this document, as implementors of preliminary software versions of the interchange specification, or as attendees at workshops and presentations. Please see the Acknowledgments section for a more complete list of important contributors.

The mass spectrometry working group has closely observed similar data interchange proposals advanced by the ADISS Project and the Analytical Instrument Association (AIA)<sup>1</sup>. The Analytical Data Interchange and Storage Standards (ADISS) Project is an industry-wide, public-domain project that includes scientific instrument and software vendors, end user companies, independent end users, government organizations, and formal standards bodies, working together to produce consistent, comprehensive standards for scientific data interchange and storage. Dr. Rich Lysakowski of the Digital Equipment Corporation is chief architect and head of the ADISS Project. In June 1992, the AIA released its first publicly-available data interchange specification, which provides a model and specification for interchange of chromatography data. The Analytical Data Interchange Specification for Chromatography (ANDI/CHROM) adapts the more general ADISS model.

This mass spectrometry specification, the ANDI/CHROM specification, and the ADISS analytical information model share a common view and terminology for describing analytical chemical experiments and data sets. In all three, there is a clear separation between the *description* of the interchange problem and its *implementation* in computer software. By prescribing only a loose coupling between the analytical information model and the implementation mechanism, the specification can be freed from dependencies on particular hardware or software platforms.

*This is not an implementation document.* The purpose of this document is to lay out the foundations of a general model for mass spectrometric experiments and data sets. It adapts the ADISS and ANDI models to the description of mass spectrometry. A separate implementation guide is available for use by software developers.<sup>2</sup>

### 1.2. The netCDF Data Interchange System

Some brief comments about the implementation are appropriate, however, since the working group's efforts have been more widely known as the "netCDF project". The ANDI chromatography specification and this mass spectrometry specification (as well as other specifications under development by the ADISS Project) are implemented using a public-domain data interchange system known as netCDF (*network Common Data Form*)<sup>3</sup>. The netCDF system was developed and is supported by the Unidata Program Center, a non-profit organization managed by the University Corporation for Atmospheric Research and the National Science Foundation. NetCDF is both a specification for data exchange and a software implementation. The source code has been ported to most of the mainframe, workstation, and personal computer platforms where C language compilers exist. The Unidata Program Center acts as a clearing house for distributing netCDF and applications which use it and as an ongoing programming effort to improve and support the netCDF code. Unidata has a staff of twelve software engineers and scientists engaged in this effort.

The netCDF mechanism was developed to meet the needs of the atmospheric science community. Meteorological data sets are typically large and multivariate. Like mass spectrometer vendors, each developer of an atmospheric science application invented a unique format for storage and retrieval of data sets. As a result, exchange of data sets between applications was very difficult. Through the efforts of a number of individuals in government and the academic community, the netCDF system was developed to provide a common format and exchange mechanism for these large data sets.

The netCDF system consists of several parts: a subroutine library which implements the Application Programming Interface (API), and which is intended to be incorporated into application programs to read or write netCDF files; a network-transparent binary file format which permits the netCDF file to be transferred to any supported computer and read without further conversion; and an external description language (Common Data Language - CDL) for representing the contents of a netCDF binary file in printable form and for generating the binary file. The netCDF system, including all source code, documentation, and test procedures, is distributed as part of the ANDI and ADISS interchange specifications.

### **1.3. Other Data Interchange Models**

During the course of development of this specification, several other interchange mechanisms were investigated and rejected. The most notable is the JCAMP-DX interchange specification, first developed for interchange of infrared spectroscopic data, and later adapted for nuclear magnetic resonance spectral data. Some mass spectrometer vendors use a modified subset of the infrared JCAMP-DX for export of abbreviated mass spectral data sets and libraries.

All of the alternatives investigated suffer from the same basic shortcoming: they are specifications only, and have no publicly available, portable, supported software implementation. The costs and risks of developing and supporting software for a complex interchange specification are too high for most vendors, and would doom an otherwise successful specification to failure for lack of vendor commitment. The netCDF system overcomes this flaw.

### **1.4. The Analytical Instrument Association**

The Analytical Instrument Association is an industry association representing the interests of analytical instrumentation and software vendors in a variety of arenas. In the past several years, the AIA membership has involved the Association in active development, promotion, and coordination of specifications for the interchange of analytical data. The Association is not a standards body in the sense of the ASTM (American Society for Testing and Materials) or other organizations. It provides a vehicle for industry-wide cooperation in the development of analytical data specifications.

In addition to the mass spectrometry and chromatography specifications, there are on-going efforts to develop similar specifications for infrared, nuclear magnetic resonance, ultraviolet-visible, and X-ray spectroscopies, and for SIMS and other surface science techniques. The AIA itself does not dictate the technical content of the interchange specifications; technical development is done by representatives from member companies and other interested individuals and organizations. Participation in development of these interchange specifications is not limited to the AIA membership.

For more information about the AIA or for additional copies of this or other interchange specifications, please contact:

The Analytical Instrument Association  
225 Reinekers Lane, Suite 625  
Alexandria, VA 22314-2875, USA

## 2. Purpose of this Document

This document presents an Analytical Data Interchange (ANDI) model for the description of mass spectrometric data, and a set of data elements which implement that description. The model and data elements are patterned after those developed by the ADISS Project and the AIA members for describing chromatographic experiments, data, and results.

The model presented here is **Version 1.0** of the **Analytical Data Interchange Specification for Mass Spectrometry (ANDI/MS)**. As the first released version, it 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 two essential types of mass spectral data: single or multiple scan centroided or continuum (profile) data sets, and single or multiple spectrum libraries. This model permits the description and interchange of the most common forms of mass spectral data sets, including full scan, selected ion monitoring (SIM/SIR), chromatography/MS, and library spectra.

The working group members deliberately resisted the temptation to develop an all-encompassing specification. Standards work is difficult, time-consuming, and entails constant compromise to ensure all points of view are considered and either adopted or justifiably deferred. Such is the case with Version 1.0 of this specification. Description of other types of data and instrumentation - MS/MS and FTMS, and complete description of the chromatography component of chromatography/MS in particular - have been deferred for expediency. On-going efforts of the working group (in cooperation with the AIA and ADISS Project) will address these other areas and be presented in later Versions of this specification.

The two major sections which follow describe the Mass Spectrometry Information Model and the Mass Spectrometry Information Classes. Both of these closely follow the ANDI and ADISS Project models. Their purpose is to describe the information content of a mass spectrometric experiment and interchange data set without reference to a software implementation. Because this specification is designed for data *interchange*, parameters or other data elements peculiar to a specific instrument or data system have been excluded. Such items have no relevance outside of their respective instruments.



### 3. Mass Spectrometry Information Model

#### 3.1. Analytical Information Categories

The ANDI Information Model specifies five categories of analytical information. Each of these (except for Category 5) is meant to stand alone; that is, there is no overlap between the information contained in the first four categories. Complete information interchange is achieved with Category 5, which meets the requirements of Good Laboratory Practices. Each of the four categories, taken in order, give an increasingly detailed description of the total experiment, and are intended to facilitate a stepwise implementation of the standard.

The categories and their description are taken directly from the ANDI AIM:

- **"Category 1 - Raw Data Only:** Category 1 is used for transferring raw data arrays that come out of the data system from the instrument. All information necessary to be able to reprocess raw data on a system made by a different manufacturer is given here. All necessary data, relative data scaling end points, and axes labels to allow accurate plotting of the spectrum, chromatogram, or other measurement are interchanged. Some assumptions are made in this category about how to read and scale raw data from one system to another, basically that data will be scaled relatively, not absolutely. Qualitatively correct transfers of chromatographic, spectral images, or other measurements are achieved with this category of information.

"Category 1 also contains minimal administrative information needed to locate the original chemical and data processing methods used with this data set.

- **"Category 2 - Final Results:** All post-calibration 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, calibration results, component identities, sample component amounts, and other derived quantities of interest to the analyst are included in Category 2 data sets. Some quantitation decisions are included as comments in Category 2 data sets to aid the analyst in determining how results were calculated.

"This category is used for data transfer to database management systems, such as a LIMS, research databases, sample tracking systems, data analysis packages, spreadsheets, or visualization packages.

- **"Category 3 - Full Data Processing Method:** Exact quantitation decisions and data processing methods are transferred in this category. Quantitatively correct data/information transfer is achieved by this category for all parameters necessary to do peak detection, measurement, and calibration for a sequence of related sample runs. This applies to both samples and standards. Peak processing method parameters and calibration method parameters required to quantitate sample component peaks are included.
- **"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.
- **"Category 5 - Good Laboratory Practice Information:** Any additional information required to satisfy Good Laboratory Practices or ISO-9000 requirements is included at this category. This category generally deals with capturing all product, process, and documentation quality information needed for validation."

The ANDI/CHROM specification currently has detailed description of Categories 1 and 2, and a preliminary description of Categories 3, 4, and 5 as applied to chromatographic data. This mass spectrometry specification considers only Categories 1 and 2 as well, although the complete description is very desirable and is under consideration.

### 3.2. Analytical Information Classes

The ANDI Mass Spectrometry Information Model categorizes mass spectrometric information into a number of classes. There is not a direct mapping of these information Classes into the information Categories described above, which has been the source of some confusion. The information Categories describe the information hierarchy; the Classes describe the contents scattered within the hierarchy. The classes defined by the ANDI/CHROM and adopted for mass spectrometry are:

- **Administrative:** information for administrative tracking of experiments.
- **Instrument-ID:** information about the instrument which generally does not change from experiment to experiment.
- **Sample Description:** information describing the sample, its history, handling, and processing.
- **Test Method:** all information used to generate the raw data and processed results. This includes instrument control, detection, calibration, data processing, and quantitation methods.
- **Raw Data:** the data as stored in the data file, along with any parameters needed to describe it.
- **Processed Results:** processing information and values derived from the raw data.
- **Component Quantitation Results:** individual quantitation results for components in a complex mixture.

The model presented in Version 1.0 only partially addresses these classes. In particular, the last two (Processed and Component Quantitation Results) are not described at all. A more complete model is being developed in cooperation with the AIA and the ADISS Project, and will be part of a future Version of this specification.

#### 4. Mass Spectrometry Information Classes

The tables below provide attribute and data type descriptions for the Administrative, Instrument-ID, Sample Description, Test Method, and Raw Data Information Classes. Following each table is a description of each attribute and its allowed range of values where appropriate.

In the tables which follow, the following abbreviations are used:

- C** Category of analytical information, as described above.
- M** Mandatory field in the given Category level; *e.g.* M1 indicates a particular data element is mandatory in any data set which includes information from Category 1, while M1234 indicates an element which is mandatory in Categories 1, 2, 3, and 4.

Because this document represents a **model** for mass spectrometric information, these tables are **not** to be interpreted as tables of keywords. The software implementation is independent of the phrases used here, and is in fact quite different. Likewise, the data types given are **not** an implementation representation, but a description of the form of the attribute. (That is, a field labeled as floating point may 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).

#### 4.1. Administrative Information Class

The following items are included in the Administrative Information class:

Administrative Information			
Attribute	Data Type	Category	Required
data set completeness	string	C1	M1234
template revision level	string	C1	M1234
administrative comments	string	C1 or C2	
data set origin	string	C1	M4
data set owner	string	C1	
experiment title	string	C1	
experiment date/time stamp	string	C1	M1234
experiment type	string	C1 or C4	
experiment cross-references	string array[n]	C4 or C3	
netCDF file date/time stamp	string	C1	M1234
netCDF revision level	string	C1	M1
operator name	string	C1	M4
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	
languages	string	C1 or C5	
number of times processed	integer	C5	
number of times calibrated	integer	C5	
calibration history	string array[n]	C5	
post-experiment program name	string	C2 or C5	
pre-experiment program name	string	C2 or C5	
error log	string	C5	

##### 4.1.1. data set completeness

A string which describes which Analytical Information Categories are contained in the data set. The string should exactly list the Category values, as appropriate, in a string separated by plus ("+") signs. Examples: "C1+C2", or "C1+C2+C3+C4+C5".

##### 4.1.2. template revision level

The revision level of the netCDF CDL template being used by implementors. This tells users which Version of this document should be referenced for the exact definitions of terms and data elements used in a particular data set. Example: "1.0".

##### 4.1.3. administrative comments

A string into which any comments not covered by another administrative attribute can be placed.

##### 4.1.4. data set 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 data set originated. Example: "Analytical Instrument Association, 225 Reinekers Lane, Suite 625, Alexandria, VA 22314-2875, USA".

#### 4.1.5. data set owner

Name of the owner of a proprietary data set. The person or organization named is responsible for the accuracy of the data. Any copyright statement should be placed here. Example: "Copyright 1994 - Analytical Instrument Association".

#### 4.1.6. experiment title

A optional, user-readable name for the experiment or test given by the analyst.

#### 4.1.7. experiment date/time stamp

Indicates the absolute time of the start of the experiment relative to Greenwich Mean Time, *i.e.* the date and time the raw data acquisition was initiated. This might be the start of the injection for a GC/MS experiment, for example.

This field is a string, and has the format: YYYYMMDDhhmmss±ffff (commas may optionally delimit the date fields, and colons may optionally delimit the time fields). The time differential factor (±ffff) expresses the hours and minutes between local time and the Co-ordinated Universal Time (UTC or Greenwich Mean Time, as disseminated by time signals), as defined in ISO 3307 and other ISO standards<sup>4</sup>. The time differential factor is represented by a four-digit number preceded by a plus (+) or 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 hours (west of the Greenwich Meridian) and by as much as +1300 hours (east of the Greenwich Meridian). When the time differential factor equals zero, this indicates a zero hour, zero minute difference from GMT. Example: 1992,12,01,17:30:23-0500 or 19921201173023-0500. In human terms, this represents 12:30:23 PM, December 1, 1992 Eastern Standard Time (5 hours behind GMT).

#### 4.1.8. experiment type

A string describing the nature of the data in the netCDF file. Currently, only the following types are valid:

**Centroided Mass Spectrum:** A data set containing *centroided* single or multiple scan mass spectra. This includes selected ion monitoring 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.

A required Raw Data Information parameter, the number of scans, is used to define the shape of the data in the file, *i.e.* 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.

Some instruments are capable of mixed mode data acquisition, for example, alternating positive/negative EI or CI 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, one for the negative EI scans. These files may be made mutually cross-referential using their "external file reference" fields.

In the ANDI/CHROM specification, this field is named "separation experiment type", since their model deals solely with chromatographic data. Because the mass spectrometry specification describes the hyphenated chromatography/MS experiment, there is a separate attribute for "separation experiment type" in the Test Method section.

#### **4.1.9. experiment cross-references**

An array of strings which reference other related experiments.

#### **4.1.10. netCDF file date/time stamp**

The date and time at which the netCDF file was created. This has the same format as described above for the "experiment date/time stamp" field.

#### **4.1.11. netCDF revision level**

The current revision of the netCDF data interchange system software being used for the mass spectrometry data interchange implementation.

#### **4.1.12. operator name**

A string naming the person who ran the experiment which acquired the data set.

#### **4.1.13. source file reference**

A string which contains sufficient detail to completely describe the name of the data file or directory which contains the original raw or library data. This is primarily for GLP compliance for files stored in netCDF format and may not be relevant for data exchange purposes. For example: "hpsid:/chemist/data/ms071001.d".

#### **4.1.14. 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 INCOS", etc.

#### **4.1.15. 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.

#### **4.1.16. 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.

#### **4.1.17. languages**

An optional list of human languages in which parameters, comments, etc. might be expressed. This is primarily for use by language processing tools which automatically parse the contents of a netCDF file into a database, report, or cross-index. Currently only English is defined.

#### **4.1.18. 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.

**4.1.19. number of times calibrated**

Also for GLP compliance, a count of the number of times the data were calibrated before yielding the final results.

**4.1.20. calibration history**

An audit trail of file names and data sets which records the calibration history, used for GLP compliance.

**4.1.21. post-experiment program name**

The name(s) of any program(s) used to process the raw data after acquisition.

**4.1.22. pre-experiment program name**

The name(s) of any program(s) used prior to the start of acquisition.

**4.1.23. error log**

Information describing any error which occurred in the acquisition or processing of the raw data. This is primarily for GLP compliance, where all data must be retained, even erroneous data.

## 4.2. 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 ..." fields below should be interpreted as occurring once for each identified component. Not all fields may be relevant for each component.

Instrument-ID Information			
Attribute	Data Type	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

### 4.2.1. 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.

### 4.2.2. 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. Examples: "gas chromatograph", "data system", "GC column", "MS core".

### 4.2.3. instrument component id

The laboratory's identification code for the instrument component; this might be an internal inventory control number.

### 4.2.4. 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.

### 4.2.5. instrument component model number

The model number and/or name used by the manufacturer to identify the instrument component.

### 4.2.6. instrument component serial number

The manufacturer's serial number, if any, for the instrument component.

### 4.2.7. instrument component id comments

Any free-form comments not covered in one of the other fields.



#### **4.2.8. instrument component software revision**

The revision level of the instrument component software (if any) when the data was acquired. This attribute applies only to non-data system instrument components. This becomes a 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.

#### **4.2.9. instrument component firmware revision**

The revision level of the instrument component firmware (if any) when the data was acquired. This attribute applies only to non-data system instrument components. This becomes a 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 embedded controller board.

#### **4.2.10. operating system revision**

The name and revision level of the data system's operating system software (if any) when the data was acquired and processed. This attribute applies only to data system instrument components, of which there might be more than one for hyphenated instruments. Required for GLP compliance.

#### **4.2.11. 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 attribute applies only to data system instrument components. Required for GLP compliance.

### 4.3. 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.

Sample Description Information			
Attribute	Data Type	Category	Required
sample owner	string	C5	
sample receipt date/time stamp	string	C5	
internal sample id	string	C5	
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	

#### 4.3.1. sample owner

The name of the sample owner or submitter. This may be different from the data set owner.

#### 4.3.2. sample receipt date/time stamp

The date and time the sample was received in the laboratory or submitted for analysis. The ISO 3307 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.

#### 4.3.3. internal sample id

The number or code used to identify the sample within the mass spectrometry laboratory or in a laboratory information management system used by the laboratory.

#### 4.3.4. external sample id

The number or code assigned to the sample by the submitter or submitter's organization.

#### 4.3.5. 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.

#### 4.3.6. sample preparation procedure

A textual description of the procedure used to prepare the sample for analysis.

#### **4.3.7. sample state**

A string field, specified as one of these choices:

Sample State
solid
liquid
gas
supercritical fluid
plasma
other state

#### **4.3.8. 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.

#### **4.3.9. sample storage information**

A description of the storage conditions for the sample, which includes the storage location. This is for OSHA compliance.

#### **4.3.10. sample disposal information**

A description of the disposal procedure for the sample (also in accord with OSHA regulations).

#### **4.3.11. 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.

#### **4.3.12. sample preparation comments**

Any comments concerning preparation not covered in other fields.

#### **4.3.13. sample id comments**

Any comments not covered elsewhere. This might include laboratory notebook references, etc.

#### **4.3.14. manual handling precautions**

Any safety issues which are of concern when the sample is manually handled.

**Note:** A future Version of this interchange specification which deals more fully with Good Laboratory Practices compliance will likely be expanded to address other sample management issues.

#### 4.4. 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.

These fields and their values have been defined to meet the limited goals outlined in the description of purpose at the beginning of this document. None of these fields are required to be present in the file; where the field is important to the interpretation of the raw data but is not present, a default value is assumed. The default value for each field is given in **boldface type** where they are defined.

Test Method Information			
Attribute	Data Type	Category	Required
separation experiment type	string	C1	
mass spectrometer inlet	string	C1	
mass spectrometer inlet temperature	floating point	C1	
ionization mode	string	C1	
ionization polarity	string	C1	
electron energy	floating point	C1	
laser wavelength	floating point	C1	
reagent gas	string	C1	
reagent gas pressure	floating point	C1	
FAB type	string	C1	
FAB matrix	string	C1	
source temperature	floating point	C1	
filament current	floating point	C1	
emission current	floating point	C1	
accelerating potential	floating point	C1	
detector type	string	C1	
detector potential	floating point	C1	
detector entrance potential	floating point	C1	
resolution type	string	C1	
resolution method	string	C1	
scan function	string	C1	
scan direction	string	C1	
scan law	string	C1	
scan time	floating point	C1	
mass calibration file name	string	C1	
external reference file name	string	C1	
internal reference file name	string	C1	
instrument parameter comments	string	C1	

**4.4.1. 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
<b>No Chromatography</b>

**4.4.2. 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
<b>direct inlet probe</b>
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.

**4.4.3. mass spectrometer inlet temperature**

This field specifies the temperature of the spectrometer inlet, if appropriate, in degrees centigrade.

**4.4.4. ionization mode**

This field describes the technique used to ionize the sample. It is also a string, chosen from the set below. **Only one ionization mode is supported per interchange file.**

Ionization Method
<b>electron impact</b>
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

**4.4.5. ionization polarity**

This field describes the polarity of the detected ions. **Only one ionization polarity is supported per interchange file.**

Ionization Polarity
<b>positive</b>
negative

**4.4.6. electron energy**

This field is relevant for electron impact ionization mode, and contains the electron energy in volts.

**4.4.7. laser wavelength**

This field is relevant for laser desorption ionization, and contains the laser wavelength in nanometers.

**4.4.8. reagent gas**

This field is relevant for chemical ionization mode, and specifies the CI reagent.

**4.4.9. 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.

**4.4.10. FAB type**

This field is relevant for fast atom bombardment, and specifies the atom or neutral used in the bombardment gun.

**4.4.11. FAB matrix**

This field specifies the fast atom bombardment matrix used, if any, for the FAB experiment type.

**4.4.12. source temperature**

This field gives the temperature of the source in degrees centigrade.

**4.4.13. filament current**

This field gives the filament input current in amps. This is relevant primarily for EI and CI ionization modes.

**4.4.14. emission current**

This field gives the filament emission current in microamps. This is also relevant principally for EI and CI ionization.

**4.4.15. accelerating potential**

This field specifies the accelerating potential in volts.

**4.4.16. detector type**

This specifies the detection method used, and is chosen from the set below.

Detector Type
<b>electron multiplier</b>
photomultiplier
focal plane array
Faraday cup
conversion dynode electron multiplier
conversion dynode photomultiplier
multi-collector
other detector

**4.4.17. 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.

**4.4.18. 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.

**4.4.19. resolution type**

This field specifies the type of instrument resolution: constant over the mass range or proportional to mass. See the description of **resolution**, in the **Raw Data Per-Scan Information** section, below.

Resolution Type
<b>constant</b>
proportional

**4.4.20. resolution method**

Specifies the method for determining spectrometer resolution. Examples: "10% peak valley", "50% peak height", "90% peak height".

**4.4.21. scan function**

A string specifying an entry from the set below. 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
<b>mass scan</b>
selected ion detection
other function

**4.4.22. scan direction**

This field specifies the direction in which the mass range was scanned during acquisition. **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
<b>up</b>
down
other direction

**4.4.23. scan law**

This field specifies the mass scan law, as the string from the following table:

Scan Law
<b>linear</b>
exponential
quadratic
other law

**4.4.24. 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.

**4.4.25. 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.

**4.4.26. 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.

**4.4.27. 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.

**4.4.28. 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.



#### 4.5. 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. Additionally, for selected ion detection (SID/SIR/SIM, etc.) experiments, a raw data per-scan-group class contains information describing the masses, scan numbers, and sampling- and delay-times for each group of masses monitored during the course of the experiment.

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.

#### 4.6. Raw Data Global Information Class

This class contains information relevant to all scans in a data set.

Raw Data Global Information			
Attribute	Data Type	Category	Required
number of scans	integer	C1	M1
starting scan number	integer	C1	
number of scan groups	integer	C1	
mass axis scale factor	floating point	C1	
time axis scale factor	floating point	C1	
intensity axis scale factor	floating point	C1	
intensity axis offset	floating point	C1	
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	
time axis data format	string	C1	
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	floating point array[2]	C1	(M1)
time axis global range	floating point array[2]	C1	(M1)
intensity axis global range	floating point array[2]	C1	
calibrated mass range	floating point array[2]	C1	
actual run time	floating point	C1	
actual delay time	floating point	C1	
uniform sampling flag	boolean	C1	
raw data global comments	string	C1	

##### 4.6.1. number of scans

This specifies the total number of scans recorded in this file. It is a required parameter.

**4.6.2. 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.

**4.6.3. 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.

**4.6.4. 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**.

**4.6.5. 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**.

**4.6.6. 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**.

**4.6.7. 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**.

**4.6.8. mass axis units**

This field specifies the units for the raw data mass axis values. The default is "**m/z**" (AMU/charge).

Mass Axis Units
<b>m/z</b>
arbitrary units
other units

**4.6.9. time axis units**

This field specifies the units for the raw data time axis values. The default is "**seconds**".

Time Axis Units
<b>seconds</b>
arbitrary units
other units

**4.6.10. intensity axis units**

This field specifies the units for the raw data intensity axis values. The default is "**arbitrary**" (unitless).

Intensity Axis Units
<b>arbitrary units</b>
counts per second
total counts
volts
current
other units

**4.6.11. total intensity units**

This field specifies the units for the raw data total intensity values. The default is "**arbitrary**" (unitless). The same table as for **intensity axis units** applies.

**4.6.12. 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:

Name	Data Format
<b>short</b>	<b>16-bit signed integer</b>
long	32-bit signed integer
float	32-bit floating point
double	64-bit floating point

The 16-bit integer **short** format is assumed by default.

**4.6.13. 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.

**4.6.14. 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 data format is used. By default, **long** format is assumed.

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.

**4.6.15. mass axis label**

This field contains the string used to label the mass axis when plotting the file data.

**4.6.16. time axis label**

This field contains the string used to label the time axis when plotting the file data.

**4.6.17. intensity axis label**

This field contains the string used to label the intensity axis when plotting file data.

**4.6.18. 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 data are present.

**4.6.19. 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.

**4.6.20. intensity axis global range**

This field contains the maximum range of the intensity axis data in low intensity, high intensity order.

**4.6.21. calibrated mass range**

This field contains the mass range (in low mass, high mass order) over which mass axis calibration is valid.

**4.6.22. actual run time**

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.

**4.6.23. actual delay time**

This field contains the time in seconds between the start of the experiment (e.g. the injection) and the start of scan acquisition. Actual delay time plus sampling period should result in the actual run time.

**4.6.24. 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).

**4.6.25. raw data global comments**

This string holds any comments relevant to the raw data not covered by the fields above.

#### 4.7. 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.

Raw Data Per-Scan Information			
Attribute	Data Type	Category	Required
scan number	integer	C1	M1
actual scan number	integer	C1	
number of points	integer	C1	M1
mass axis values	<i>mass data format</i> array	C1	M1 (see below)
time axis values	<i>time data format</i> array	C1	M1 (see below)
intensity axis values	<i>intensity data format</i> array	C1	M1
number of flags	integer	C1	
flagged peaks	integer array	C1	
flag values	integer array	C1	
total intensity	floating point	C1	
a/d sampling rate	floating point	C1	
a/d coaddition factor	integer	C1	
scan acquisition time	floating point	C1	
scan duration	floating point	C1	
mass scan range	floating point array[2]	C1	
time scan range	floating point array[2]	C1	
inter-scan time	floating point	C1	
resolution	floating point	C1	

##### 4.7.1. scan number

An integer which specifies the index of this scan within the set of scans in the interchange file. For multiple-scan data sets, this is a required field. The first scan in the set has index zero (0).

##### 4.7.2. actual scan number

This specifies the index of this scan in the source data file, and provides for the case where only part of the source data file is converted into interchange format. By default, it is assumed that the **actual scan number** is the same as the **scan number** value.

##### 4.7.3. number of points

This specifies the number of mass-time-intensity triplets, and is a required field.

##### 4.7.4. 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.**

#### **4.7.5. 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.

The specification does not support centroided mass data in combination with continuum time data. If both mass and time data are present, the arrays must be parallel, that is, the arrays must have an equal number of points, and each point in the mass array is assumed to have been derived by mass measurement from the corresponding point in the time array.

#### **4.7.6. 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 (*i.e.* 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.

#### **4.7.7. 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.

#### **4.7.8. 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 (0, 4, 5).

#### 4.7.9. 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:

NOT HIGH RESOLUTION	The peak is a 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 ( <i>i.e.</i> has fractional mass)
REFERENCE	Peak is a reference from the reference file
EXCEPTION	Peak is a reference from the exception file
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 - <i>e.g.</i> isotope calculation)
LOCK MASS	Peak is a reference mass used to adjust the mass scale during acquisition

**4.7.10. total intensity**

Specifies the total intensity associated with this scan, in floating point format. For a chromatography/MS data set, this series of intensities is used to construct the total ion chromatogram (TIC).

**4.7.11. a/d sampling rate**

This field specifies the rate (in kilohertz) at which A/D (analog-to-digital) conversions are made.

**4.7.12. a/d coaddition factor**

This field specifies the number of A/D samples which are co-added or averaged to produce a single datum point.

**4.7.13. 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.

**4.7.14. 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.

**4.7.15. 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.

**4.7.16. 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.

**4.7.17. 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.

**4.7.18. 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 ( $\text{mass} / \Delta \text{mass}$ ); for instruments with constant absolute mass resolution (such as quadrupoles), resolution is specified as  $\text{mass} / \text{charge}$  ( $m/z$ ). 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.



## 4.8. 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.

Library Data Per-Scan Information			
Attribute	Data Type	Category	Required
entry name	string	C1	M1
entry 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	floating point	C1	
retention index type	string	C1	
absolute retention time	floating point	C1	
relative retention	floating point	C1	
retention reference name	string	C1	
retention reference CAS number	integer	C1	
melting point	floating point	C1	
boiling point	floating point	C1	
chemical mass	floating point	C1	
nominal mass	integer	C1	
accurate mass	floating point	C1	
other information	string	C1	

### 4.8.1. 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.

### 4.8.2. entry id

This field specifies a non-name attribute 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.

### 4.8.3. 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.

### 4.8.4. 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.

**4.8.5. CAS name**

This string gives the name of the entry recognized by the Chemical Abstracts Service.

**4.8.6. other names**

This is an array of strings, and specifies additional names by which this entry is known.

**4.8.7. CAS number**

This is the Chemical Abstracts Service registry number for the library entry, if any.

**4.8.8. chemical formula**

This string gives the chemical formula for the entry, if any.

**4.8.9. Wiswesser notation**

This field specifies the Wiswesser notation for the entry.

**4.8.10. SMILES notation**

This string specifies the SMILES notation for the entry.

**4.8.11. MOLfile reference name**

This string specifies the name of an external file containing chemical structure information for the entry in Molecular Design Limited MOLfile format<sup>5</sup>. 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.

**4.8.12. 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.

**4.8.13. 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.

**4.8.14. retention index type**

This field contains the method by which retention index was determined, for example: "Kovats".

**4.8.15. 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.

**4.8.16. 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.

**4.8.17. retention reference name**

This field specifies the name of the reference material used in measurement of the relative retention of the library spectrum.

**4.8.18. 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.

**4.8.19. melting point**

This field contains the melting point, in degrees Centigrade.

**4.8.20. boiling point**

This field specifies the boiling point, in degrees Centigrade.

**4.8.21. chemical mass**

This field specifies the chemical mass, computed using the average atomic masses for each element in the formula.

**4.8.22. 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.

**4.8.23. 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.

**4.8.24. other entry information**

Some spectral libraries allow association of user-supplied information with entries. This field contains this descriptive information.

#### 4.9. 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 type. 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.

Raw Data Per-Scan-Group Information			
Attribute	Data Type	Category	Required
number of masses in group	integer	C1	M1
starting scan number	integer	C1	M1
group masses	floating point array[n]	C1	M1
sampling times	floating point array[n]	C1	
delay times	floating point array[n-1]	C1	

##### 4.9.1. number of masses in group

This specifies the count of masses monitored during this scan group. This field is mandatory.

##### 4.9.2. 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.

##### 4.9.3. 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.

##### 4.9.4. 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.

##### 4.9.5. 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.

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).

## 5. Acknowledgments

It is impossible to acknowledge all of the 100 or more individuals who have contributed to the development and review of this specification. The authors would like to thank particularly their employers, who have universally recognized the importance of this data interchange specification and who have provided the time and resources required to develop and implement it.

In addition to the authors, a number of individuals have been greatly involved in this effort:

Jeremy Batt, VG Biotech  
Walter Blackstock, Glaxo  
Ron Bonner, PE Sciex  
David Bowen, Pfizer  
Pete D'Arcy, VG Masslab  
Ed Darland, Hewlett-Packard  
Tony Davies, ISAS Dortmund  
Mike Duff, Analytical Instrument Association  
Craig James, Daylight Chemical Information Systems  
James Hegedus, Fisons Instruments  
Sue Kennerley, Kratos Analytical  
Rich Lysakowski, Digital Equipment Corporation  
Bob McDonald, JCAMP-DX  
Brian Mehlman, JEOL USA  
Brian Musselman, JEOL USA  
Jon Onstot, Midwest Research Institute  
Charles Reilly, UC Riverside  
Keith Rollins, ICI  
Jim Shofstahl, Finnigan MAT  
Dave Smith, PE-Sciex  
Jun Tamura, JEOL USA  
Kent Wenger, Extrel FTMS

The authors apologize for any inadvertent omissions.

## 6. References

- 
- <sup>1</sup> Lysakowski, R., *Analytical Data Interchange and Storage Standards; AIA Generic Chromatography Data Standard, Draft Version 1.0*, personal communication to D. Stranz (April 1991)
- <sup>2</sup> Stranz, D.D., *Implementation Guide: Analytical Data Interchange for Mass Spectrometry (ANDI/MS), Version 1.0*, Analytical Instrument Association (May 1994)
- <sup>3</sup> Rew, R.K., *netCDF User's Guide: An Interface for Data Access*, Unidata Program Center (March 1991)
- <sup>4</sup> ISO Standard 3307-1975 (E) "Information Interchange - Representations of Time of the Day"; ISO Standard 2014-1976 (E) "Writing of Calendar Dates in All-Numeric Form"; and ISO Standard 4031-1978 (E) "Information Interchange - Representations of Local Time Differentials"
- <sup>5</sup> Dalby, A., Nourse, J.G., Hounshell, W.D., Gushurst, A.K.I., Grier, D.L., Leland, B.A., and Laufer, J., *Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited*, Journal of Chemical Information and Computer Sciences **32**, 244-255 (1992)