



Designation: E1947 – 98 (Reapproved 2022)

Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data¹

This standard is issued under the fixed designation E1947; 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 chromatographic data representation and a software vehicle to effect the transfer of chromatographic data between instrument data systems. This specification provides 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 or results files. This standard format has the extension “.cdf” (derived from NetCDF). The contents of the file include typical header information like instrument, column, detector, and operator description followed by raw or processed data, or both. 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 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

1.4 The protocol in this specification is intended to (1) transfer data between various vendors’ instrument systems, (2) provide 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.

1.5 The protocol consists of:

1.5.1 This specification on chromatographic data, which gives the full definitions for each one of the generic chromatographic 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.

1.5.2 Guide E1948 on chromatographic 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. It is intended for software implementors, not those wanting to understand the definitions of data in a chromatographic dataset.

1.5.3 NetCDF User’s Guide.

1.6 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 ASTM Standards:³

E1948 Guide for Analytical Data Interchange Protocol for Chromatographic Data

2.2 Other Standard:⁴

NetCDF User’s Guide

2.3 ISO Standards:⁵

ISO 2014-1976 (E) Writing of Calendar Dates in All-Numeric Form

ISO 3307-1975 (E) Information Interchange—Representations of Time of the Day

ISO 4031-1978 (E) Information Interchange—Representations of Local Time Differentials

3. Terminology

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

¹ This specification is under the jurisdiction of ASTM E13 on Molecular Spectroscopy and Separation Science and is the direct responsibility of E13.15 on Analytical Data.

Current edition approved Nov. 1, 2022. Published November 2022. Originally approved in 1998. Last previous edition approved in 2014 as E1947 – 98 (2014). DOI: 10.1520/E1947-98R22.

² For more information on the NetCDF standard, contact Unidata at <http://www.unidata.ucar.edu>.

³ For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard’s Document Summary page on the ASTM website.

⁴ Available from Russell K. Rew, Unidata Program Center, University Corporation for Atmospheric Research, P. O. Box 3000, Boulder, CO 80307-3000, <http://www2.ucar.edu>.

⁵ Available from International Organization for Standardization (ISO), ISO Central Secretariat, Chemin de Blandonnet 8, CP 401, 1214 Vernier, Geneva, Switzerland, <https://www.iso.org>.

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.

Data Element Name	Datatype	Category	Required
dataset-completeness	string	C1	M12345
protocol-template-revision	string	C1	M12345
netCDF-revision	string	C1	M12345
languages	string	C5	...
administrative-comments	string	C1 or C2	...
dataset-origin	string	C1	M5
dataset-owner	string	C1	...
dataset-date-time-stamp	string	C1	...
injection-date-time-stamp	string	C1	M12345
experiment-title	string	C1	...
operator-name	string	C1	M5
separation-experiment-type	string	C1	...
company-method-name	string	C1	...
company-method-ID	string	C1	...
pre-experiment-program-name	string	C5	...
post-experiment-program-name	string	C5	...
source-file-reference	string	C5	M5
error-log	string	C5	...

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

3.1.2 *company-method-ID*—internal method ID of the sample analysis method used by the company.

3.1.3 *company-method-name*—internal method name of the sample analysis method used by the company.

3.1.4 *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.

3.1.5 *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.

3.1.5.1 *Discussion*—This is a synthesis of ISO 2014-1976 (E), ISO 3307-1975 (E), and ISO 4031-1978 (E), which compensates for local time variations.

3.1.5.2 *Discussion*—The 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 3307-1975 (E). The time differential factor (ffff) is represented by a four-digit number preceded by a plus (+) or a minus (-) sign, indicating the number of hour and minutes that local time differs from the UTC. Local times vary throughout the world

from UTC by as much -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, and zero second difference from Greenwich Mean Time.

3.1.5.3 *Discussion*—An example of a value for this date element would be: 1991,08,01,12:30:23-0500 or 19910801123023-0500. In human terms this is 12:30 PM on August 1, 1991 in New York City. Note that the -0500 hours is 5 full hours time behind Greenwich Mean Time. The ISO standards permit 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.

3.1.5.4 *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.

3.1.6 *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.

3.1.7 *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.

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

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

3.1.10 *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.

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

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

3.1.11.2 *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).

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

3.1.13 *operator-name*—name of the person who ran the experiment or test that generated the current dataset.

3.1.14 *post-test-program-name*—name of the program or subroutine that is run after the analytical test is finished.