CHEMICAL IDENTIFICATION DATA STANDARD

Standard No.: EX000016.2

January 6, 2006

Approved on January 6, 2006 by the
Exchange Network Leadership Council
for use on the Environmental
Information Exchange Network

Approved on January 6, 2006 by the
Chief Information Officer of the
U. S. Environmental Protection Agency
for use within U.S. EPA

This consensus standard was developed in collaboration by State, Tribal, and U. S. EPA representatives under the guidance of the Exchange Network Leadership Council and its predecessor organization, the Environmental Data Standards Council.

Foreword
The Environmental Data Standards Council (EDSC) identifies, prioritizes, and pursues the creation of data standards for those areas where information exchange standards will provide the most value in achieving environmental results. The Council involves Tribes and Tribal Nations, state and federal agencies in the development of the standards and then provides the draft materials for general review. Business groups, non-governmental organizations, and other interested parties may then provide input and comment for Council consideration and standard finalization. Standards are available at http://www.epa.gov/datastandards.

1.0 INTRODUCTION

The Chemical Identification Data Standard defines the required information for identification of chemical substances. This update to the data standard provides five mandatory and fifteen optional data elements for chemical substance identification. The mandatory data group includes EPA Chemical Tracking Number, Chemical Abstracts Registry Number, Chemical Substance Systematic Name, EPA Chemical Registry Name and EPA Chemical Identifier. The optional data elements provide additional identification information about a chemical substance or chemical grouping.

1.1 Scope

This data standard includes the data groupings needed to consistently and unambiguously identify chemical substances regulated by or of interest to US EPA and other regulatory entities.

1.2 Revision History

Date	Version	Description
February 22, 2001	1-9938:1	Initial Environmental Data Standards Council Adoption of Version 1.
January 6, 2006	EX000016.2	Environmental Data Standards Council Adoption of updated version of the Chemical Identification Standard. Changes included addition of a data element, formatting, and standard renumbering.

References to Other Data Standards

This standard relies on other standards to make it complete and provide the necessary support. As such users should consider the references to other data standards noted below as integral to the Chemical Identification Data Standard. These include:

- ESAR: Analysis and Results [EX000005.1] Data Standard
- EDSC: Attached Binary Object [EX000006.1] Data Standard

1.3 Terms and Definitions

Term	Definition
Chemical Substance	An organic or inorganic material that can be categorized and defined for the purposes of this standard as one of the following: a single fully defined chemical substance; a chemical species; a chemical substance of known composition; a chemical substance of variable composition; a chemical substance of unknown composition; or a generic-confidential business information (CBI) chemical substance.
Substance Registry	The Substance Registry System (SRS) is the official United States System Environmental Protection Agency (US EPA) repository and reference database of names and other identifiers for substances and substance groupings of interest to the Agency and other regulators. The SRS provides tools for search and retrieval of substance identification data, directly through its own Web interface as well as through active linkages from other data systems. The SRS is intended to facilitate substance data integration among US EPA and its external stakeholders and partners. http://ofmpub.epa.gov/sor_internet/registry/substreg/home/overview/home.do

1.4 Implementation

Users are encouraged to use the XML registry housed on the Exchange Network Web site to download schema components for the construction of XML schema flows (http://www.exchangenetwork.net).

1.5 Document Format

The structure of this document is briefly described below:

a. Section 2.0 Chemical Identification Diagram, illustrates the principal data groupings contained within this standard.

b. Section 3.0 Chemical Identification Data Standards Table provides information on the high level, intermediate and elemental Chemical Identification data groupings. Where applicable, for each level of this data standard a definition, XML tag, note(s), example list of values and format are provided. The format column may include the number of characters for the associated data element, where “A” specifies alphanumeric, “N” designates numeric, and “Graphic” designates a diagram or other graphic related binary object.

c. Data Element Numbering: For purposes of clarity and to enhance understanding of data grouping hierarchy and relationships, each data group is numerically classified from the primary to the elemental level.

d. Code and Identifier metadata: Metadata, defined here as data about data or data elements, that includes their descriptions and/or any needed context setting information required to identify the origin, conditions of use, interpretation, or understanding the information being exchanged or transferred. (Adapted from ISO/IEC 2382-17:1999 Information Technology Vocabulary—Part 17: Databases 17.06.05 metadata). Based on the business need, additional metadata may be required to sufficiently describe an identifier or a code. A note regarding this additional metadata is included in the notes column for identifier and code elements. Additional metadata for identifiers may include:

- Code List Identifier, which is a standardized reference to the context or source of the set of codes

Additional metadata for codes may include:

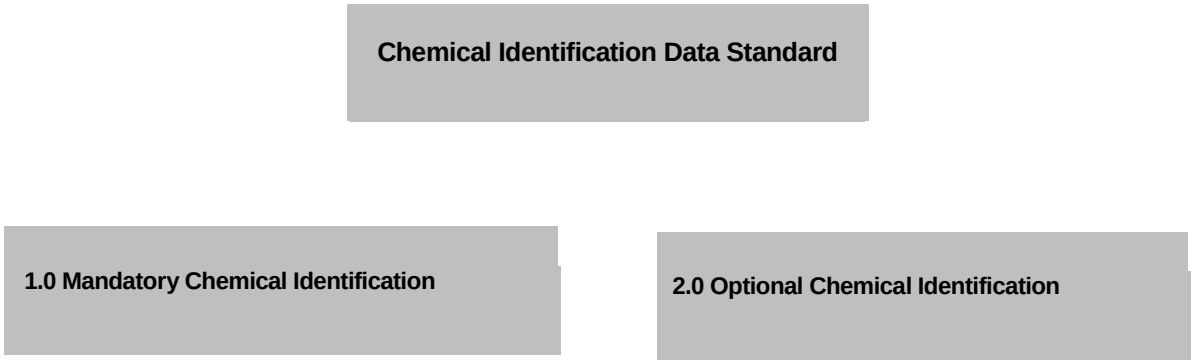
- Code List Identifier, which is a standardized reference to the context or source of the set of codes
- Code List Version Identifier, which identifies the particular version of the set of codes.
- Code List Version Agency Identifier, which identifies the agency responsible for maintaining the set of codes
- Code List Name, which describes the corresponding name for which the code represents

e. Appendix A, Chemical Identification Data Structure Diagram, illustrates the hierarchical classification of the Chemical Identification data standard. This diagram enables business and technical users of this standard to quickly understand its general content and complexity.

f. Appendix B, lists the references for Chemical Identification data standard. **2.0**

CHEMICAL IDENTIFICATION DIAGRAM

The figure below illustrates the major data groups associated with the Chemical Identification Data Standard.



3.0 CHEMICAL IDENTIFICATION DATA STANDARDS TABLE 1.0

Definition: Mandatory information required for identification and designation about the chemical substance.

Relationship: None.

Notes: XML None.

Tag: MandatoryChemicalIdentification

Mandatory Chemical Identification

Name	Definition	Notes	Format	XML Tag

1.1 EPA Chemical Internal Tracking Number	The unique record number assigned to a chemical substance or a chemical grouping for tracking within EPA systems.	This data element is mandatory. It is an electronic key that facilitates data exchange with the Substance Registry System (SRS) and other EPA databases. This identification number must be obtained from the SRS.	A	EPAChemicalInternalTrackingNumber
1.2 Chemical Abstracts Service Registry Number	The unique number assigned by Chemical Abstracts Service (CAS) to a chemical substance.	This is mandatory where that number exists or can be assigned. Example List of Values: • 67-66-3 for Chloroform • 7439-92-1 for Lead • 108-88-3 for Toluene	A	CASRegistryNumber

1.3 EPA Chemical Identifier	The identifier to be created and placed in the SRS for each chemical substance or chemical group in the SRS for which a CAS Registry Number does not exist and cannot be assigned.	This identifier is mandatory when CAS Registry Number does not exist. Example List of Values: • E761429 for Copper Compounds • E17075060 for Polycyclic Aromatic Hydrocarbons, High Molecular Weight *Note:* Based on the business need, additional metadata may be required to sufficiently describe an identifier. This additional metadata is described in the Introduction section 1.6.d.	A	EPAChemicalIdentifier

1.4 Chemical Substance Systematic Name	A standard name assigned to a chemical substance.	*Note:* The name is descriptive of the molecular composition of the substance if the composition is known. If a CAS number exists for a chemical substance, the index name is formulated according to the nomenclature rules set forth for the Chemical Abstracts 9th Collective Indexing Period. This is mandatory where that name exists or can be assigned. Example List of Values: • Methane, trichloro • Lead, tetraethyl • Acetic acid, chloro	A	ChemicalSubstanceSystematicName

1.5 EPA Chemical Registry Name	The name US EPA has selected as the preferred name for a chemical substance.	The US EPA Chemical Registry name cannot be assigned to all chemical substances of interest. The name, however, is mandatory for all chemical groupings and for chemical substances where CAS systematic names do not exist and cannot be assigned. Example List of Values: • Chloroform • Iodine • Anthracene	A	EPAChemicalRegistryName

Optional Chemical Identification

Definition: Additional information about a chemical substance that may be used for identification. Relationship:

None.

Notes: None.

2.0

XML Tag: OptionalChemicalIdentification

Name	Definition	Notes	Format	XML Tag
2.1 EPA Chemical Registry Name Source Text	The source of the US EPA chemical registry name.	Example List of Values: • RCRA • CERCLIS • AQS • TRI • CWA311	A	EPAChemicalRegistryNameSourceText
2.2 EPA Chemical Registry Name	The type of source for the US EPA Chemical Registry Name.	Example List of Values: • EPA Regulation Name • US EPA Data System Name	A	EPAChemicalRegistryName

2.3 Molecular Formula Text	The formula that specifies the number of atoms of each element in a molecule of a chemical substance.	SRS will provide a complete formula for all single, fully-defined chemical substances, or a partial formula for mixtures where one or more components can be defined. Example List of Values: • Molecular formula for Chloroform is $CHCl_3$ • SRS matches for $C_2H_4Cl_2$ are Ethane, 1,1-dichloro and Ethane, 1,2-dichloro and Dichloroethane (mixture)	A	MolecularFormulaText

2.4 Chemical Substance Formula Weight Quantity	The sum of the atomic weights of constituent atoms in a chemical substance.	SRS will provide the formula weight where a complete molecular formula for a chemical substance exists. Example List of Values: • Formula weight for Chloroform is 119.38 • Formula weight for Heptachlor is 373.32	A	ChemicalSubstanceFormulaWeightQuantity
2.5 Chemical Substance Type Name	A descriptive name for types of chemical substances.	SRS will store this name for regulatory chemical classes. Example List of Values: • Organic Substances • Radionuclide	A	ChemicalSubstanceTypeName

2.6 Chemical Substance Definition Text	The text that provides clarification to the identity of a chemical substance.	SRS will store this text when needed to completely, uniquely define a chemical substance. Example List of Values: • Definition for Humic Acid: The brown polymeric product from the decomposition of organic matter, particularly dead plants. This combination polymeric product may contain aromatic and heterocyclic structures, carboxylic groups and nitrogen.	A	ChemicalSubstanceDefinitionText

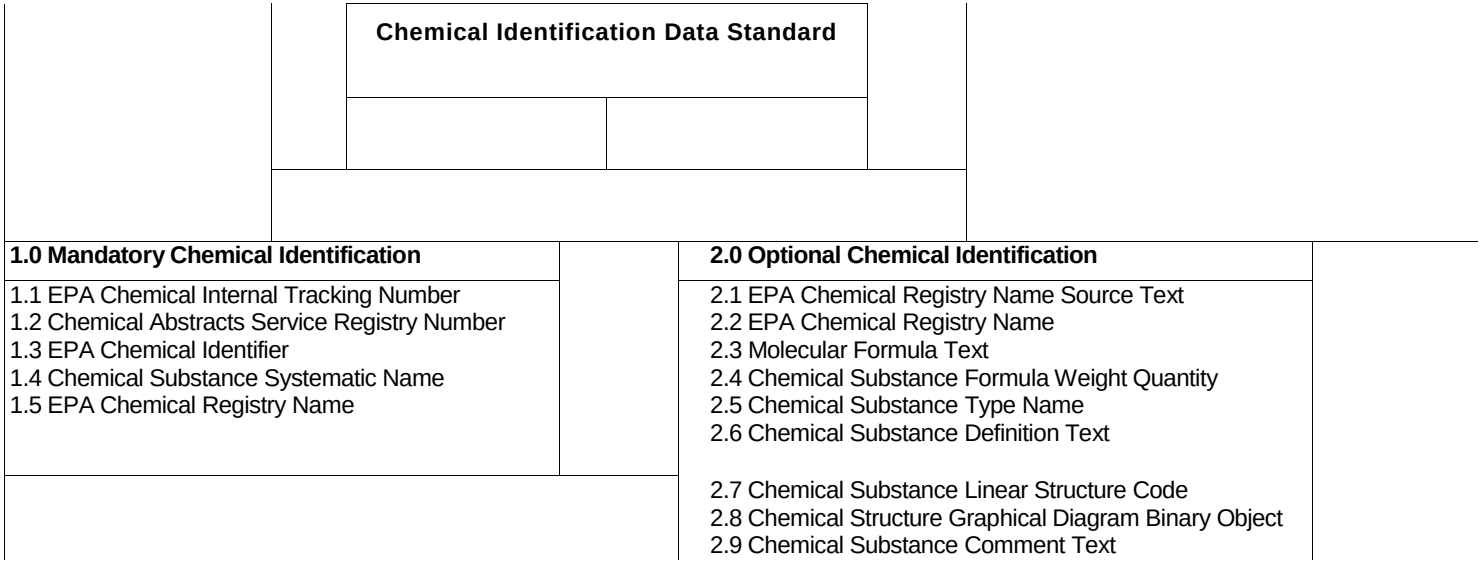
2.7 Chemical Substance Linear Structure Code	The code that represents the connectivity of atoms in a molecule of a chemical substance as a linear formula, such as Simplified Molecular Input Line Entry System (SMILES).	SRS will store this code for all single, fully-defined chemical substances, and a partial code where one or more components can be defined. Example List of Values: • SMILES notation for Triethylamine is CCN(CC)CC • SMILES notation for Ethanol is CCO • SMILES notation for Propionic acid is CCC(=O)O *Note:* Based on the business need, additional metadata may be required to sufficiently describe an identifier. This additional metadata is described in the Introduction section 1.6.d.	A	ChemicalSubstanceLinearStructureCode

2.8 Chemical Structure Graphical Diagram Binary Object	A graphical representation of a molecule of a chemical substance as a two or three dimensional diagram.	SRS will provide this diagram for all single, fully-defined chemical substances, and a partial representation where one or more components can be defined. *Note:* It may be appropriate to use the Attached Binary Object [DRAFT] Data Standard given the nature of the material being transferred.	Graphic	ChemicalStructureGraphicalDiagramBinaryObject
2.9 Chemical Substance Comment Text	The text that provides additional information about a chemical substance.	Example List of Values: • Chlordane with CAS No. 5774-9 is also identified by CAS No.12789-03-6.	A	ChemicalSubstanceCommentText

2.10 Chemical Substance Synonym Name	The name that is used as an alternative for representing a chemical substance.	Example List of Values: • Methoxyethane, 2-methoxy-2-methylpropane, and tert-butyl methyl ether are Synonyms • Ethanol, Ethyl alcohol and Ethyl hydroxide are Synonyms • Vinyl chloride, Chloroethene and Ethylene monochloride are Synonyms	A	ChemicalSubstanceSynonymName
2.11 Chemical Synonym Source Name	The name of the source of an alternate name for a chemical substance.	Example List of Values: • RCRA • CERCLIS • AQS	A	ChemicalSynonymSourceName
2.12 Chemical Synonym Name	The name that identifies the circumstance in which that name has been used.	Example List of Values: • Iron and Steel Industry • Pulp and Paper Industry • Aniline and Leather Manufacturing	A	ChemicalSynonymName

2.13 Chemical Synonym Name Status Name	The name that documents the correctness of a synonym for a specific chemical.	Example List of Values: • Incomplete • Inaccurate • Ambiguous	A	ChemicalSynonymNameStatusName
2.14 Chemical Substance Classification Name	The name that classifies chemical substances according to structural similarities.	Example List of Values: • Carbamates • Thiophosphates • Freons	A	ChemicalSubstanceClassificationName
2.15 Chemical Preferred Acronym Name	The name the US EPA has selected as the preferred acronym or otherwise abbreviated name in the SRS for a chemical substance, when use of a shortened name is appropriate.	Example List of Values: • MEK for Methyl ethyl ketone • MTBE for Methyl tert-butyl ether • PCB for Polychlorinated biphenyl	A	ChemicalPreferredAcronymName

Appendix A
Chemical Identification Data Structure Diagram



Appendix B
References

1. *Interim Data Standard for Chemical Identification,* Tom Maloney, May 6, 1999.

2. *Reinventing Environmental Information (REI) Interim Chemical Identification Data Standard,* Chemical Data Standard Work Group Report, April 22, 1999.

3. *Summary Report of Standard Data Elements for Chemical Substances,* SDC-0055-057-TC-7015, December 2, 1997.

4. *Rules for Representation of Chemical Data in Envirofacts Pilot Master Chemical Integrator,* SDC-0055-057-LF-3019A, June 10, 1994.

5. *Chemical Abstracts Service Registry Number Data Standard,* IRM Policy, Standards and Guidance/IRM Strategic Planning Documents, EPA Directive No. 2180.1, June 26, 1987.

6. *EPA Chemical Registry Name Selection Procedures,* Drafted by the Chemical Name Selection Subgroup of the Chemical Identification Standard Business Rules Workgroup, Draft 1.1, February 29, 2000 (rev. April 9, 2000).

7. ISO/IEC 2382-17:1999 Information Technology Vocabulary—Part 17: Databases 17.06.05 .