Ref

59791

Johnston, T.

Nomenclature Rules for:
~~EDD_ANALYTE_CODE~~ LIST Table Data Fields

Revision 1 (July) 1997.

INTRODUCTION

Maintaining order in the LANL ER databases requires a system for naming and describing data elements that is self-consistent and logical. An added bonus is achieved if the system is also consistent with other naming conventions that aren't necessarily developed at LANL. With these ideals as a target, a system was developed for naming and describing analyte codes and their associated analyte descriptions in FIMAD. These codes and descriptions are found in the EDD_ANALYTE_CODE_LIST table. That table has four data fields called ANALYTE_CODE, ANALYTE_CODE_DESC, SPECIAL DESCRIPTORS and CAS_NO.

It was most expedient to accommodate naming conventions that are already in use such as the US EPA Region IX SAL analyte naming convention, the Institute for Pure and Applied Chemistry (IUPAC) rules which govern chemistry nomenclature, and common names for certain chemicals that are better known under the common name than under the official IUPAC name. However, because the various systems are not unified in their nomenclature, the primary goal was to develop a self-consistent system for LANL ER with as few exceptions as practicable, and to relegate consistency with other systems to a significant, but secondary, level of importance.

An attempt was made to minimize the number of rules and thus limit the complexity of this nomenclature system. Ideally, different people could use these rules to generate analyte descriptions and codes, and the descriptions and codes would be identical for a each analyte. Given the complexity of the FIMAD database, one should not expect this ideal to be attained in all cases, though these rules should ensure a high level of consistency. Before generating new analyte codes and descriptions, consideration should be given to existing code conventions by reviewing existing codes and descriptions.

ANALYTE_CODE_DESC Nomenclature Rules

1. The chemical root name is the main part of the chemical name. Usually, it may be formed by removing the leading prefix numbers or letters from a name. For example the root name of "1,2-dichlorobenzene" is "dichlorobenzene" and the root name of "N-Nitrosodimethylamine" is "Nitrosodimethylamine".

2. Numerical prefixes, and all prefixes except bis, tris, tetrakis, etc., that indicate multiple ligand attachments, are placed in square brackets following the root word

9735

which they modify. For example, "Dimethylbenzidine[3,3']" and "Ethoxyethanol[2-] Acetate" are both correct. "Ethoxyethanol Acetate[2-]" is incorrect.

3. If a chemical name begins with the prefix bis, tris, tetrakis, etc., the prefix is capitalized and retained as the beginning of the name. For example "Bis(chloromethyl)ether" is preferred to "bis(chloromethyl)ether" and to "Ether[bis(chloromethyl)]".

4. The root name and trailing square brackets are not separated by any spaces.

5. Individual words in the root name are capitalized. For example, all words in "Acidity, Methyl Orange" and both words in "Sodium Diethyldithiocarbamate" are capitalized. Words joined with hyphens are not capitalized (e.g., Amino-1-methoxy-4-nitrobenzene[2-]).

6. Except for the letter "N" denoting attachment of a ligand to a nitrogen atom, words and letters in square brackets are not capitalized.

7. Ring positions and other positions of attachment should be denoted using integers rather than "ortho", "meta" or "para". For example, Nitrophenol[4-] is preferred to Nitrophenol[p-].

8. Where attachment of a ligand to symmetrical nitrogens must be denoted, the convention of using "N'" (i.e., N-primed) is preferred. In the case of symmetrical attachment to carbons, integer primes denoting the carbon ring or carbon chain position are used (i.e., 1 and 1', 2 and 2', etc.). For example, Oxybis(1-chloropropane)[2,2'] denotes attachment to an oxygen atom of two different 1-chloropropane chains, each attachment being at the "2" position of the carbon chain.

9. If a CAS number belongs to a mixture of isomers, this should be denoted by following the analyte name with "(Mixed Isomers)" or, if specification of the isomers is desired, the isomeric names should be joined into one analyte code (e.g., Dichloroethene[cis/trans]).

10. The forward slash is used to denote a combination of individual species (e.g., cis/trans indicates cis *plus* trans isomers).

11. In some cases, it might be preferred to use "Total" to describe a mixture of isomers or to differentiate a species representing all forms of a compound or element from individual spaces. For example, Xylene(Total) may be used to represent the three xylene isomers and Cr(Total) may be used to represent total chromium (as opposed to Cr(+3) and Cr(+6)).

12. Root name words such as "Acetate" and "Acid" follow square brackets and are separated from the brackets with one space.

13. Names of metal salts should use the metal's name followed by the oxidation state of the metal in parentheses, and then the accompanying ion name. The oxidation state of the metal should be shown as Arabic (as opposed to Roman) numerals. If the metal appears in nature in only one oxidation state, the oxidation state need not be used. For example, "Aluminum Oxide" may be used instead of "Aluminum(+3) Oxide" because aluminum generally occurs in a single oxidation state, but "Thallium(+1) Oxide" is preferred to "Thallium Oxide" because thallium appears in nature in the "+1" and "+3" oxidation states.

14. Common names, especially for pesticides and herbicides which have otherwise very long chemical names, are allowed if it seems that such usage facilitates efficiency and if it seems that the common name is more commonly known throughout the ER Project than the IUPAC or other chemical name. However, the use of common names should be minimized. This is a very subjective evaluation. Minimizing the use of common names should facilitate adherence to a standard nomenclature.

15. Analyte names should provide enough structural information about the molecule that a reasonably competent college sophomore could draw the molecular structure of the compound from the name. Of course, this will not be true in the case of complicated organic molecules represented by a common name.

ANALYTE_CODE Nomenclature Rules

1. Where practical, an attempt is made to use official Chemical Abstracts Service (CAS) numbers for the LANL ER Analyte Code. When available, CAS numbers, as well as the ER Project ANALYTE_CODEs, should be provided in the EDD_ANALYTE_CODE_LIST table. This means that the analyte code and CAS number will be identical for some analytes, and they will differ for others. For many ER analytes, no CAS number exists.

2. Analyte codes for organic and inorganic target analytes are the CAS Nos. for the respective compounds except for CH_4, H_2, CO_2, N_2 and O_2. The analyte codes for these exceptions are CH4, H2, CO2, N2 and O2.

3. Analyte codes for metal target analytes are the *capitalized* chemical symbol (e.g., FE for Fe, MO for Mo, V for V).

4. Analyte codes for isotope target analytes are the *capitalized* chemical symbols followed by a hyphen and the isotopic mass number (e.g., AM-243 and MN-54 for americium-243 and manganese-54, respectively).

5. Analyte codes for isotopic tracers are the analyte codes for the target analytes with a "T" appended to them (e.g., AM-243T, PU-242T).

6. Analyte codes for isotopic carriers are the analyte codes for the target analytes with "Carrier" appended to them (e.g., RA-226Carrier).

7. Generic tracers and carriers may be identified using the appropriate chemical symbol followed by either "-T" or "-Carrier" (e.g., AM-Tracer for Americium-Tracer).

8. Analyte Codes for ions are the chemical symbols for the ions, including electrical charge which is placed in parentheses at the end of the code (e.g., CL(-1), BR(-1), NO3(-1), SO4(-2), NA(+1), CA(+2). Electrical charge always includes the algebraic sign of the charge preceding the magnitude of the charge. For elemental or ionic species for which variants exist, designate each variant with a unique suffix. For example, cyanide amenable to chlorination and reactive cyanide are designated as CN(-1)A and CN(-1)R, respectively. No designation is necessary for indicating "total" species. However, the inclusion of a designator for "total" species (e.g. SIO2TOT) might be useful to guarantee that no confusion will occur among the codes for the various species.

9. For inorganic species which are not elemental or ionic, and for physical analytes/parameters, a logical abbreviation is used that indicates as completely as possible what the intended analyte/parameter is. The length of the abbreviation must be balanced with the intent to be as specific as possible so the analyte code is not confused with another code. In general, analyte codes less than 12 characters are preferred but exceptions do exist.

10. Mixtures of target analytes are generally represented as the analyte codes for each individual target analyte of the mixture separated with a forward slash (e.g., Ba/Cmpds for a mixture of barium and barium compounds). Notable exceptions are HEP/HEPEPOX for heptachlor and heptachlor epoxide (CAS nos. not used here), and radioisotope mixtures which are designated by the Analyte Code of the isotope of lowest mass followed by the mass numbers of the other isotopes of the mixture. The additional mass numbers are presented in increasing numerical order and are separated by forward slashes with no spaces between the mass numbers and slashes (e.g., RA-226/228 and U-233/234).

11. Similar analytes should be represented by similar codes. For example, all silica species are best represented by a code beginning with SIO2, followed by unique designators for individual types of silica (i.e., COLL for colloidal, SUSP for suspended, TOT for total). Phosphates are another example of analytes that should be represented by similar codes. The use of similar codes facilitates the use of alphabetically sorted analyte code lists because it ensures that all similar codes will be grouped together when the list is sorted.

12. Consideration should be given to how the analyte/parameter will be determined and reported. For example, phosphate and nitrogen species are commonly reported in terms of the equivalent amount of elemental phosphorus or nitrogen even though a determination was made of an ionic form such as PO_4^{-3} or NO_3^{-1}. In some cases (e.g., orthophosphate) the species may be determined using different chemical tests and reported either as the equivalent element or the ionic species, so allowance should be made for both. It is then up to the person ordering the analytical test to select the appropriate analyte code.

SPECIAL_DESCRIPTORS Nomenclature Rules

Special descriptors are descriptions that embellish the ANLYTE_CODE_DESC entries. Adherence to the ANALYTE_DESC nomenclature rules is NOT necessarily encouraged. The purpose of these entries is to provide information that may be used to confirm the identity of a particular analyte, analyte code or CAS no., or to provide additional information about analytes that might be useful to have readily available for writing reports, etc. For example, synonyms for the ANALYTE_DESC entries, chemical formulae, and/or descriptions that allow specific identification of operationally defined analytes may be stored in this field. Thus for ANALYTE_DESC entry "Acidity, Methyl Orange....", "As defined in Std. Meth., 17th ed." is located in the SPECIAL_DESCRIPTORS field. This tells the reader that the operational definition for methyl orange acidity may be found in "Standard Methods for the Examination of Water and Wastewater," 17th edition. The use of synonyms can be especially useful in the case of pesticides that have many different commonly used names.

This field is essentially a free-form data field.

CAS_NO Nomenclature Rules

This field is reserved for official Chemical Abstracts Services numbers. Many analytes determined in the ER Project do not have CAS numbers, in which case the field is left blank. CAS numbers consist of a sequence of digits followed by a hyphen, followed by two digits, followed by a hyphen, followed by a single digit. For example, 123-45-6 and 123456-78-9 are of the correct format (although these particular numbers have not necessarily been assigned to a chemical compound).