OUTLINE OF THE ISS CODE FOR THE CLASSIFICATION OF CHEMICAL SUBSTANCES FOR USE IN COMPILING A NATIONAL INVENTORY

S. CAROLI and N. VIOLANTE

Laboratorio di Tossicologia Applicata, Istituto Superiore di Sanità, Rome, Italy

Summary. - A coding system for categorizing all types of compounds was developed specifically as part of the National Inventory of Chemical Substances to be applied by the Istituto Superiore di Sanità (ISS) in order to comply with the EEC present legal provisions on chemicals in trade. A number of applications were also described to illustrate the coverage and the adaptability of the code.

KEY WORDS: classification code for chemical substances, national inventory of chemical substances.

Riassunto (Struttura del codice ISS per la classificazione delle sostanze chimiche da impiegarsi nella realizzazione di un inventario nazionale). - Un sistema di codificazione per la classificazione di qualunque tipo di composto è stato sviluppato per l’uso preciso da parte dell’Inventario Nazionale delle Sostanze Chimiche gestito dall’Istituto Superiore di Sanità (ISS) nell’ambito dei provvedimenti normativi attualmente vigenti nella CEE per i prodotti chimici in commercio. Vengono inoltre dati alcuni esempi applicativi per illustrare la potenzialità e la flessibilità del codice.

PAROLE CHIAVE: codice di classificazione per sostanze chimiche, inventario nazionale delle sostanze chimiche.

Introduction

The numerous and diversified undertakings of the more industrialized countries in the late 1970s, devised to control and make congruent the production of new substances, led to a number of acts and regulations laying emphasis on the essential role of prevention rather than on a policy of intervention after damage has occurred. The achievements of those years, realized through international and regional organizations like the OECD and EEC, had a strong and permanent impact on governments' attitude for the control of chemicals and their negative consequences on human health and environment.

It soon became evident that there was a definitive need for exhaustive and systematic data banks capable of storing information on both existing compounds and those produced and commercialized for the first time. While several undertakings were successfully launched in the same period (for example IRPTC, an acronym for International Register for Potentially Toxic Chemicals), the EEC set up the so-called EINECS (European Inventory of Existing Chemical Substances) as an indispensable means to put into effect the Directive no. 79/831, better known as the VI Amendment. At a national level this Directive has been enforced through the Presidential Decree DPR No. 927 (24.11.1981); the preceding Act of Sanitary Reform No. 833, Article 9 (23.12.1978), however, already called for the compilation of a national inventory entrusting this to the Istituto Superiore di Sanità. From its very inception it was clear that if the inventory was to be efficient it was mandatory to develop a coding system which would allow the univocal categorization and retrieval of substances. This paper reports in detail the approach adopted for the said classification.

Description of the code

An efficient system of classification should not simply list substances in numerical order but give general information on their principal characteristics.

The classification outlined below endeavours to combine in a single expression, and as concise as possible, overall information on the nature of the substance in a way that the substance cannot be mistaken. The form used is as follows:

\[ \text{XO XX XXXX XXXX} \]
\[ 1 \quad 2 \quad 3 \quad 4 \]

where X indicates a number and O corresponds to a letter in the English alphabet.
The first segment of the code (1) is a number-letter combination where the two symbols correspond respectively to the General Class to which the substance belongs (inorganic compounds, aromatic organic compounds, etc.) and the Functional Sub-class (inorganic acids, phenols, etc.). Therefore, in the binomial the meaning of the letter can be ascertained only on the basis of the number which precedes it. Attachments 1 through 8 list these numbers and the substances to which they correspond.

Part (2) of the code, in the case of classifications beginning with 1 or 2 (that is, specifically for inorganic and organometallic substances), indicates the Atom Number of the main element of the compound, or, when this is not possible, the one, among elements of equal importance, with the largest atomic number. For all other substances, i.e. codes beginning with X = 3, part (2) corresponds to the overall number of carbon atoms present in the organic structure. If this number is in excess of the code's capacity, a conventional sign (for example **) will be used to indicate a complex product with more than 99 carbon atoms.

The next four numbers which form Part (3) of the code indicate the molecular weight of the substance. This indication therefore is entirely independent of the initial binomial assigned as opposed to what took place in Part (1) and (2). Molecular weights exceeding four numbers will be indicated in a similar manner to (2) above, for example *****. An alternative to this, in the case of molecular weights of more than four numbers, could be to group together the 6 places of Parts (2) and (3) instead of the division as it stands. In that way it would be possible to register up to a maximum of six digits. The elimination of information in Part (2) for substances of very high molecular weight is more than compensated for by the fact that for these substances it is more important to know this parameter rather than their number of carbon atoms. This change is indicated by placing Parts (2) and (3) together to form a single expression of six consecutive digits.

The four digits of Part (4) are a progressive number indicating substances which have identical values for the three preceding parts. The capacity of this section should be sufficient to guarantee all possible combinations of the preceding sections. It would also be possible to extend the capacity of Part (4) to five digits to give an extremely wide scope for registering substances which are identical in the first three parts. When the expression does not fill all the space allotted for that Part, "0" will be inserted to avoid confusion.

From the above it is clear that through this system an unambiguous code for each substance can be established which can be updated from time to time. Within the XO group, i.e. Part (1), the substances are listed in increasing order of complexity through a double entry - atom number or number of carbon atoms/molecular weight (the first can obviously never be greater than the second), thus further distinguishing similar substances. Finally, the last Part of the code assures defining a substance should all the initial parameters be the same.

Furthermore, it is understood that substances having more than one functional group present at the same time will be classified under the function which first appears on the attached Explanation Key.

Organic substances with two or more different structures present simultaneously will be classified under the first category to which they belong in the following order: Aromatic Heterocyclic Compounds - Aromatic Compounds - Alkynic Compounds - Aliphatic Compounds.

The examples given in the following section illustrate what has been outlined above.

**Examples of applications**

1. Substance: Potassium dichromate \((K_2Cr_2O_7)\)
   - General Class: 1
   - Functional Sub-class: F
   - Binomial Value: 1F
   - Atomic Number of main element: \((Cr)\): 24
   - Molecular Weight: 294
   - Progressive Number for substance with identical preceding parameters (simply for use in this example): 7
   - Substance Code: 1F 24 0294 0007

2. Substance: 3-Amino-4-chloro-benzoic acid
   - \[\text{COOH}\]
   - \[\text{Cl}\]
   - \[\text{NH}_2\]
   - General Class: 5
   - Functional Sub-class: C (the three functional groups in order of priority are in fact -COOH, -NH\(_2\), and -Cl, as per Attachment 5, and therefore this substance goes under Sub-group 5)
   - Binomial Value: 5C
   - Number of Carbon Atoms: 7
   - Molecular Weight: 171
   - Progressive Number for substances with identical preceding parameters (simply for use in this example): 18
   - Substance Code: 5C 07 0171 0018

3. Substance: Phenalenone
   - General Class: 5
   - Functional Sub-class: A
   - Binomial Value: 5A
   - Number of Carbon Atoms: 13
   - Molecular Weight: 166
   - Progressive Number for substances with identical preceding parameters (simply for use in this example): 6
   - Substance Code: 5A 13 0166 0006
4. Substance: 1-Methylnaphthalene

General Class: 5
Functional Sub-class: A
Binomial Value: 5A
Number of Carbon Atoms: 13
Molecular Weight: 166
Progressive Number for substances with identical preceding parameters (simply for this example): 7
Substance Code: 5A 13 0166 0007

Examples 3 and 4 illustrate the case of two substances with identical values for the first 3 sub-sections. Consequently, they will be classified under the same group, but can be distinguished by their progressive number in the last section.

From the above it is clear that substances will be classified firstly on the basis of increasing alphabetical-numerical order of the first section. If this first binomial is the same, then the substances will be classified according to the order of the second section and so on for the other sections, thus giving a logical framework which proceeds from inorganic to organic and from simplest to the most complex compounds.

Conclusions

The examples listed above testify to the flexibility and effectiveness of the coding system devised, a system which has actually been in operation since the National Inventory came into existence. Although certainly simpler than the procedure adopted by the Chemical Abstracts Service to generate unambiguous alphanumeric strings for identifying a chemical, the present system not only serves the purpose for which it was intended, but is also applicable in a variety of similar instances. Appendices 1 through 9 list the keys used to set up any specific code together with any additional information requested by the particular class of substances at hand.

Received on November 22, 1990.
Accepted on January 23, 1991.

Appendix 1

Keys for Part 1: Inorganic Substances

1A - Elements
1B - Acids
1C - Anhydrides
1D - Hydrides and hydroperoxides
1E - Oxides and peroxides
1F - Salts
1G - Borides
1H - Carboxylic acids
1I - Hydrides
1J - Nitrides
1K - Organometallic compounds and alloys
1L - Complexes

* Excluding the saline compounds listed in 1G, 1H, 1I and 1J.

Appendix 2

Keys for Part 1: Organometallic Compounds *

2A - Organic compounds containing boron
2B - Organic compounds containing silicon
2C - Organic compounds containing arsenic, antimony, bismuth and phosphorus
2D - Organic compounds of selenium and tellurium
2E - Others

* This term is meant for all those organic compounds in which a carbon atom is directly linked to another atom of any other element other than hydrogen, nitrogen, oxygen, sulphur, fluorine, chlorine, bromine, iodine or astatine. At the same time, however, they can be present in the molecule of the organometallic compound.

Appendix 3

Keys for Part 1: Aliphatic Organic Substances

3A - Non-substituted hydrocarbons
3B - Omium salts (for carbon, nitrogen, phosphorus, oxygen and sulphur)

Appendix 4

Keys for Part 1: Organic Aliphatic Substances

4A - Non-substituted hydrocarbons
4B - Omium salts (for carbon, nitrogen, phosphorus, oxygen and sulphur)
4C - Carboxylic acids (including amino, amic, imidic, hydrazonic, hydroxamic acids, oxy-, hydroxy-, peroxo- and alkoxycarboxylic acids, and acids containing phosphorus)
4D - Sulfonic and sulfuric acids
4E - Salts of 4C and 4D acids
4F - Esters of 4C and 4D acids
4G - Lactones, lactams, lactides and lactams
Appendix 5

Keys for Part 1: Organic Aromatic Substances

5A - Non-substituted hydrocarbons
5B - Onium salts (for carbon, nitrogen, phosphorus, oxygen and sulphur)
5C - Carboxylic acids (including amino, amic, imide, hydrazonic, hydroxamic acids, oxo-, hydroxy-, peroxo- and alkoxycarboxylic acids, and acids containing phosphorus)
5D - Sulphonic and sulenic acids
5E - Salts of SC and SD acids
5F - Esters of 5C and 5D acids
5G - Lactones, lactams, lactates and lactams
5H - Anhydrides
5I - Amides, imides and hydrazides
5J - Amides, formamidines and other compounds containing groups such as \(-\text{N}=\text{C}=\text{N}\), \(-\text{N}=\text{C}=\text{N}-\) and \(-\text{N}=\text{C}(\text{N})=\text{N}-\)
5K - Nitriles and isocyanides
5L - Aldheydes
5M - Ketones and ketene derivatives
5N - Acetals and acylals
5O - Alcohol, glycols, polyglycols and carbohydrates
5P - Thiols
5Q - Amines, imines, hydroxylamines and hydrazine derivatives
5R - Hydrazones and oximes
5T - Schiff's bases
5U - Peroxides and hydroperoxides
5V - Sulfoxides, sulfones and sulfonohalides
5W - Sulfides, polysulfides and hydrosulfides
5X - Halides and halide derivatives
5Y - Nitro- and nitroso-derivatives
5Z - Azo- and azoxy-derivatives

Appendix 6

Keys for Part 1: Organic Aromatic Heterocyclic Substances

6A - Non-substituted hydrocarbons
6B - Onium salts (for carbon, nitrogen, phosphorus, oxygen and sulphur)
6C - Carboxylic acids (including amino, amic, imide, hydrazonic, hydroxamic acids, oxo-, hydroxy-, peroxo- and alkoxycarboxylic acids, and acids containing phosphorus)
6D - Sulphonic and sulenic acids
6E - Salts of 6C and 6D acids
6F - Esters of 6C and 6D acids
6G - Lactones, lactams, lactates and lactams
6H - Anhydrides
6I - Amides, imides and hydrazides
6J - Amides, formamidines and other compounds containing groups such as \(-\text{N}=\text{C}=\text{N}\), \(-\text{N}=\text{C}=\text{N}-\) and \(-\text{N}=\text{C}(\text{N})=\text{N}-\)
6K - Nitriles and isocyanides
6L - Aldheydes
6M - Ketones and ketene derivatives
6N - Acetals and acylals
6O - Alcohol, glycols, polyglycols and carbohydrates
6P - Thiols
6Q - Amines, imines, hydroxylamines and hydrazine derivatives
6R - Hydrazones and oximes
6S - Schiff's bases
6U - Peroxides and hydroperoxides
6V - Sulfoxides, sulfones and sulfonohalides
6W - Sulfides, polysulfides and hydrosulfides
6X - Halides and halide derivatives
6Y - Nitro- and nitroso-derivatives
6Z - Azo- and azoxy-derivatives

Appendix 7

Keys for Part 1: Polymeric Substances

7A - Addition polymers
7B - Condensation polymers
7C - Other types of polymers

Appendix 8

Keys for Part 1: Miscellaneous (substances which cannot be classified using the rules given under Appendices 1 through 7)

8A - Inorganic compounds
8B - Organometallic compounds
8C - Aliphatic organic compounds
8D - Alocyclic organic aromatic substances
8E - Non-heterocyclic organic aromatic substances
8F - Heterocyclic organic aromatic substances
8G - Other substances of a mixed nature in which the main component is inorganic
8H - Other substances of a mixed nature in which the main component is organic

Appendix 9

Substances of natural origin

This class covers all products of natural origin which cannot be reduced to well-defined individual chemical structures, i.e. with precise formulas and molecular weights. In fact, when the chemical nature of a substance can be ascertained, the compound is placed in a given class on the basis of the chemical family to which it belongs, regardless as to whether it is more or less of natural origin.

On the other hand, when there are highly complex compounds in natural products, each of which can be chemically defined, or substances having widely variable structures and molecular weights (both cases could arise at the same time), these are classified in an entirely different way.

Given the particular nature of this section, the meaning of the four subparts of the code is considerably different from the eight preceding classes. In this case, the first part of the code is made up of two symbols: the number 9 shows that the substance belongs to this section while the letter gives the chemical sub-group on the basis of its characteristics as follows.

9A - Aminosides
9A - Peptides and polypeptides
9C - Proteins
9D - Enzymes
9E - Hormones
9G - Steroids
9H - Lipids
9I - Nucleic acids
### Code System

- **9I** - Oligo- and polysaccharides
- **9K** - Alkaloids
- **9L** - Vitamins
- **9M** - Antibiotics
- **9N** - Pigments
- **9Q** - Membranes and cell-free extracts
- **9P** - Biological fluids and secretions
- **9Q** - Plant fluids and extracts
- **9R** - Fossil products
- **9S** - Toxins
- **9T** - Mineral products

The second part specifies whether or not the substance has undergone transformation of any kind except, however, simple mechanical separation, elimination of water of hydration, etc., and indicates its origin in the following manner:

- **NU** - Not transformed, of human origin
- **NA** - Not transformed, of animal origin
- **NV** - Not transformed, of vegetal origin
- **NM** - Not transformed, of mineral origin
- **NN** - Not transformed, of not specified origin
- **TU** - Transformed, of human origin
- **TA** - Transformed, of animal origin
- **TV** - Transformed, of vegetal origin
- **TM** - Transformed, of mineral origin
- **TN** - Transformed, of not specified origin

The third part, made up of four places, uses only the first with a letter referring to the principal use of the substance as per the list below:

- **A** - Foods
- **B** - Cosmetics
- **C** - Fuels
- **D** - Colours, Paints, Solvents
- **E** - Detergents
- **F** - Fertilizers

- **G** - Pesticides
- **H** - Therapeutics
- **I** - Industrials
- **Z** - Others or unspecified

When a substance has more than one use and it is not possible to establish which is the main one, each use will be indicated with its appropriate letter in alphabetical order from left to right up to a total of three letters. The last place is always used for a number in progression for substances which have the same first three parts.

For further clarification it is pointed out that in practice it is not possible to have all the combinations foreseen in theory in as much as some of them would be absurd (for example, 9P, biological fluids, is incompatible with NM, not transformed substance of mineral origin).

Some examples will illustrate the above:

1. Chlorophyll: on the market it is an alcoholic or oily watery solution of an intense green colour and contains hydrolysate products as salts with Na and K, of the methyl and phytol esters of both forms of natural chlorophyll (a and b). The substance is therefore classified under section 9N. Furthermore, the substance has undergone hydrolytic and salting treatment, thus allowing letter T in the second part and, being of vegetal origin, the letter V. The principal use of this product is as a colorant, and therefore the third part becomes letter D. For the sake of example a progressive number has been given and the complete code becomes 9N TV D000 0000.

2. Hemoglobin: as it is a protein, this substance becomes 9C in the first part of the code. As its preparation requires only physical means of separation (in this particular case, centrifugation) it is considered "unchanged". Further, in general it can be assumed to be of animal origin, unless specified as being human. The second part becomes NU. The third part is H000, i.e. for therapeutic use. The complete code with a progressive number is 9C NU H000 0005.

3. Algic acid: this polysaccharide is extracted from seaweed and mainly used as a colloid for food products and in the manufacture of paper. The code can be derived, along the same lines as above, as 92 NV AI00 0002.