Development of a Chemical Use Classification System to Facilitate Reporting under the Toxic Substances Control Act†

WENDY L. BYER* and HERBERT B. LANDAU
Auerbach Associates, Inc., Philadelphia, Pennsylvania 19107

ERNEST W. STALDER
Environmental Protection Agency, Washington, D.C. 20460

Received April 20, 1977

A classification system was developed to enable manufacturers and processors of industrial chemicals to report categories or proposed categories of use of such chemicals to the Environmental Protection Agency in accordance with the Toxic Substances Control Act. To accommodate the two aspects of chemical use (i.e., function and application), a faceted classification scheme was designed. The function facet contains categories denoting the action for which a chemical is specially fitted or used, for example, adhesives or fuels. The application facet contains categories denoting the process or product in which a chemical is used, such as synthetic rubber manufacture. Linking these two facets in a single notation code provides a comprehensive indication of a chemical's use or uses. A variety of existing relevant classification schemes and reference tools were used as input sources for the chemical use classification. The practicability of the classification system was tested using a small sample of manufacturing and processing companies.

INTRODUCTION

The Toxic Substances Control Act (TSCA), which became effective on January 1, 1977, makes the Environmental Protection Agency (EPA) responsible for regulating chemical substances "whose manufacture, processing, distribution in commerce, use, or disposal may present an unreasonable risk of injury to health or the environment".[1] Under the law, manufacturers and processors of chemical substances must submit certain data to EPA prior to the commercial production or processing of a new chemical or of a previously produced chemical intended for a new use. Reported data may include the chemical's molecular structure, its intended use, estimated production volume for each intended use, the byproducts resulting from manufacture, processing, use or disposal of the chemical, estimates of the related occupational exposure levels, and any known environmental and health effects of the chemical. Chemical substances exempted from regulation under the TSCA are pesticides, foods, food additives, drugs, cosmetics, nuclear materials, tobacco, mixtures, and research chemicals. This paper describes the development of a classification system which will allow manufacturers and processors of chemicals to codify chemical use data reported to EPA as required by the TSCA.

In July 1976, the EPA Office of Toxic Substances contracted with Auerbach Associates, Inc. to design a Chemical Use Classification System, consisting of the classification scheme, an alphabetical index to the classification scheme, and an instruction manual describing how to code chemicals according to their use. (Note that chemical uses are to be classified, but not the chemicals themselves.) Reporting of uses of new chemicals to EPA is required by law to begin in December 1977.

CLASSIFICATION SYSTEM SCOPE

Because the specific chemicals or categories of chemicals being reported to EPA via the classification system could not be determined in advance, the system had to be capable of describing all significant current uses of commercial chemicals in the various fields of economic activity. In addition, it had to be expandable to incorporate any significant new uses of chemicals that may arise. It was necessary, however, to limit the notation, or symbolic representation of classification category, to nine decimal digits in order to conform to EPA internal systems design specifications.

We were able to reconcile the somewhat conflicting requirements of broad scope and narrow notation by designing a classification system that would be broadly comprehensive and expandable but restricted in depth, or specificity.

CLASSIFICATION SYSTEM DESIGN

The basic consideration in designing the classification system was to provide EPA with a maximum amount of information at minimal effort on the part of the reporting firm.

For purposes of assessing the environmental and health impact of industrial chemicals, we found that use could be considered to consist of two basic components, i.e., function and application. We therefore decided that a faceted classification would most accurately and succinctly satisfy the requirements for maximum information at minimum effort. In this type of system, each facet represents a different aspect of the chemical being classified, and the system user can combine elements of each facet in a manner which is most appropriate to the particular chemical. The classification system therefore consists of two facets, which together will describe the use of a chemical:

(1) Function
(2) Application

Function denotes the action for which a chemical is specially fitted or used. Thus, terms such as adhesives, flame retardants, explosives, or fuels would constitute functions. Application denotes the process or product in which a chemical is used, such as petroleum refining or synthetic rubber manufacture. Linking these two facets will indicate not only the action for which a chemical is used, but also the process or product to which it is applied.

We rejected the idea of combining function and application together into a single listing, because it would present sig-

† This work was performed under Environmental Protection Agency Contract No. 68-01-4117. This paper was presented at the 11th Middle Atlantic Regional Meeting of the American Chemical Society, Newark, Del., April 20–22, 1977.

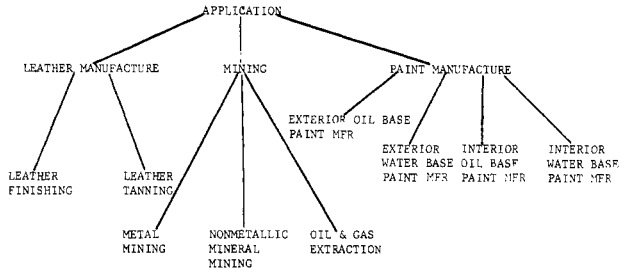

Figure 1. Example of inverted tree structure.

nificant inefficiencies in development and use. Primarily, tremendous redundancy would be built into the system, increasing the amount of work required to construct it and burdening reporting firms (or users) with a multiplicity of terms through which they would be required to search before being able to code a chemical. For example, the function "solvent" would have to be listed under the following application categories:

Dry cleaning (to dissolve foreign matter)
Painting, paper hanging, and decorating (as paint thinner)
Weed control (as herbicide)
Seed oil processing (to extract oil from seeds)

In order to select the appropriate application category or categories for the chemical being classified, the user would have to search an alphabetical index to the classification. To allow the user to do this, the index would have to list all application categories for a function under that function. This requirement also necessitates more work by the developer.

With a faceted system, each function need be listed only once, in the function facet. The user first selects one or more functions from the function facet and then selects one or more applications from the application facet. In this way, the user himself coordinates appropriate terms to describe the use of a chemical. The alphabetical index can then simply list all terms in order without regard to grouping applications under functions.

A secondary, but important advantage of the faceted approach is that it has allowed us to adapt elements of an existing classification system, the "Standard Industrial Classification" (SIC).[2] Since the SIC is widely used both in government and industry, compatibility between it and the Chemical Use Classification system is highly desirable.

Finally, a faceted classification allows the user to complete any or all of the facets independently, thereby providing at least some information, in the event that the user does not know all of the requisite data.

CLASSIFICATION SYSTEM FORMAT AND CONVENTIONS

Classification consists of grouping together similar things and separating dissimilar things. This operation is accomplished by applying increasingly specific criteria to a universe of categories. For example, at the most general level, the Chemical Use Classification System groups various applications together, each of which is subdivided into more specific categories. These categories, in turn, may be further subdivided. As demonstrated in Figure 1, conceptually, the classification system consists of a hierarchical "inverted tree" arrangement of categories from the most general to the most specific. Each of the two facets is characterized by a different level of specificity. Facet I, describing the function of a chemical, extends to two levels of specificity, while Facet II, describing the application of a chemical, extends to four levels of specificity.

Figure 2 (Example of Classification System Format) shows examples of the contents of each of the two proposed facets. The function facet contains approximately 400 major nodes

FACET I (FUNCTION)

```
206     FERTILIZER CONDITIONERS
207     FERTILIZERS
208     FILLERS (AUGMENTATION)
            (INERT MATERIALS ADDED TO SUBSTANCES TO MODIFY THEIR
             PROPERTIES AND IMPROVE QUALITY)
2080        ADHESIVE FILLERS
2081        PAINT FILLERS
2082        PAPER FILLERS
2083        PLASTICS FILLERS
2084        PRINTING INK FILLERS
2085        RUBBER FILLERS
209     FILLERS (PATCHING)
            (MATERIALS USED TO FILL HOLES IN WOOD, PLASTER OR OTHER
             SURFACES BEFORE APPLYING A COATING SUCH AS PAINT OR
             VARNISH)
210     FILTRATION AIDS
211     FINISHING AGENTS (TEXTILE TECHNOLOGY)
            (THIS IS A GENERAL CATEGORY.  USE SPECIFIC CATEGORIES WHENEVER
             POSSIBLE, I.E., ANTISLIP FINISHING AGENTS, ANTISTATIC AGENTS,
             CREASEPROOFING AGENTS, DELUSTRANTS, FLAME RETARDANTS, GREASE-
             PROOFING AGENTS, MOTHPROOFING AGENTS, ROTPROOFING AGENTS,
             SIZING AGENTS, SOIL RELEASE AGENTS, SOIL RETARDANTS, WATER
             REPELLENTS, WATERPROOFING AGENTS)
212     FIRE EXTINGUISHING AGENTS
213     FIRMING AGENTS (FOOD PROCESSING)
214     FIXATIVES
2140        DYE FIXATIVES
2141        PERFUME FIXATIVES
2142        PHOTOGRAPHIC FIXATIVES
215     FLAME RETARDANTS
216     FLATTING AGENTS
            (ADDITIVES FOR PAINTS OR VARNISHES TO DISPERSE INCIDENT
             LIGHT RAYS TO GIVE THE DRIED SURFACE A MATTE FINISH)
217     FLAVOR ENHANCERS
218     FLAVORS
2180        FOOD SOURS
2181        SEASONINGS (FOOD)
2182        SWEETENERS
219     FLOCCULATING AGENTS
            (SUBSTANCES THAT INDUCE THE AGGREGATION OF SUSPENDED SOLID
             PARTICLES IN SUCH A WAY THAT THEY FORM SMALL CLUMPS OR TUFTS
             RESEMBLING WOOL)
220     FLOUR TREATING AGENTS (FOOD PROCESSING)
```

FACET II (APPLICATION)

```
18      PAINT AND ALLIED PRODUCT MANUFACTURE
180         EXTERIOR OIL BASE PAINT PRODUCT MANUFACTURE
1800            AUTOMOTIVE AND MACHINERY (REFINISH) PAINT AND ENAMEL MANUFACTURE
                    (INCLUDES PRIMERS AND UNDERCOATERS; EXCLUDES LACQUERS)
1801            BARN AND ROOF PAINT MANUFACTURE
1802            EXTERIOR HOUSE PAINT MANUFACTURE
                    (INCLUDES TINTING BASES, PRIMERS AND UNDERCOATERS,
                     INTERIOR-EXTERIOR FLOOR PAINTS)
1803            MARINE PAINT AND ENAMEL MANUFACTURE, EXCEPT INDUSTRIAL
                    MARINE PAINTS
1804            METALLIC PAINT MANUFACTURE
1805            STAIN (EXTERIOR) MANUFACTURE
                    (INCLUDES SHINGLE AND SHAKE STAINS)
1806            TRAFFIC PAINT MANUFACTURE
1807            VARNISH (EXTERIOR) MANUFACTURE
                    (NATURAL AND SYNTHETIC OLEORESINOUS VARNISHES)
181         EXTERIOR WATER BASE PAINT PRODUCT MANUFACTURE
1810            ALL PURPOSE WATER EMULSION PAINT MANUFACTURE AND TINTING
                    BASE MANUFACTURE
1811            MASONRY WATER EMULSION PAINT AND TINTING BASE MANUFACTURE
182         INDUSTRIAL PRODUCT FINISH MANUFACTURE, EXCEPT LACQUERS
1820            APPLIANCE, HEATING EQUIPMENT AND AIR CONDITIONER FINISH
                    MANUFACTURE
1821            INDUSTRIAL MAINTENANCE PAINT MANUFACTURE
                    (INCLUDES INTERIOR AND EXTERIOR PAINTS)
1822            INDUSTRIAL MARINE PAINT MANUFACTURE
1823            INSULATING VARNISH (ELECTRICAL TYPE) MANUFACTURE
1824            MACHINERY AND EQUIPMENT FINISH MANUFACTURE
1825            METAL PAINT MANUFACTURE
                    (INCLUDES PAINTS FOR SHEET METAL AND SIDINGS, CON-
                     TAINERS AND CLOSURES, AND METAL FURNITURE AND FIXTURES)
1826            PAPER AND PAPERBOARD FINISH MANUFACTURE
1827            POWDERED COATING MANUFACTURE
1828            TRANSPORTATION EQUIPMENT PAINT MANUFACTURE
                    (ORIGINAL FINISHING ONLY)
1829            WOOD PAINT MANUFACTURE
                    (INCLUDES PAINTS FOR WOOD COMPOSITION BOARD AND WOOD
                     FURNITURE AND FIXTURES)
183         INTERIOR OIL BASE PAINT PRODUCT MANUFACTURE
1830            INTERIOR OIL BASE WALL AND TRIM PAINT MANUFACTURE
                    (INCLUDES TINTING BASES, PRIMERS AND UNDERCOATERS)
1831            STAIN (INTERIOR) MANUFACTURE
1832            VARNISH (INTERIOR) MANUFACTURE
184         INTERIOR WATER BASE PAINT PRODUCT MANUFACTURE
1840            INTERIOR WATER EMULSION WALL AND TRIM PAINT MANUFACTURE
185         LACQUER MANUFACTURE
                    (INCLUDES ACRYLIC LACQUERS)
1850            AUTOMOTIVE LACQUER MANUFACTURE
1851            FABRICATED METAL LACQUER MANUFACTURE
1852            PAPER AND PAPERBOARD LACQUER MANUFACTURE
1853            WOOD LACQUER MANUFACTURE
186         PUTTY AND ALLIED PRODUCT MANUFACTURE
1860            PAINT AND VARNISH REMOVER MANUFACTURE
1861            PUTTY AND GLAZING COMPOUND MANUFACTURE
1862            WOOD AND TEXTILE PRESERVATIVE MANUFACTURE
                    (NON-PRESSURE TYPE)
187         RELATED PAINT PRODUCT MANUFACTURE
1870            BLEACHED SHELLAC MANUFACTURE
1871            INK VEHICLE MANUFACTURE
1872            PIGMENT DISPERSION MANUFACTURE
188         THINNER MANUFACTURE
                (THINNERS FOR DOPES AND LACQUERS AND OLEORESINOUS
                 THINNERS)
```

Figure 2. Example of Classification System Format.

1. AGRICULTURE, FORESTRY AND FISHING
2. CHEMICAL MANUFACTURE
3. CONSTRUCTION
4. CONSUMER AND COMMERCIAL APPLICATIONS
5. DRUGS AND OTHER PHARMACEUTICALS MANUFACTURE
6. FABRICATED METAL PRODUCT MANUFACTURE
7. FABRICATED RUBBER AND PLASTIC PRODUCT MANUFACTURE
8. FABRICATED TEXTILE PRODUCT MANUFACTURE
9. FOOD PROCESSING
10. FURNITURE AND FIXTURE MANUFACTURE
11. INSTRUMENT AND RELATED PRODUCT MANUFACTURE AND PROCESSING
12. LEATHER AND FUR MANUFACTURE
13. LUMBER AND WOOD PRODUCT MANUFACTURE
14. MACHINERY AND EQUIPMENT
15. MINING
16. MISCELLANEOUS PRODUCT MANUFACTURE
17. ORDNANCE AND ACCESSORIES
18. PAINT AND ALLIED PRODUCT MANUFACTURE
19. PAPER AND ALLIED PRODUCT MANUFACTURE
20. PETROLEUM REFINING
21. PLASTICS MATERIALS AND SYNTHETICS (NOT FABRICATED) MANUFACTURE
22. PRIMARY METAL MANUFACTURE
23. PRINTING AND BOOKBINDING SERVICES
24. PRINTING INK MANUFACTURE
25. SANITIZING AGENT AND COSMETICS MANUFACTURE
26. STONE, CLAY, GLASS AND CONCRETE PRODUCT MANUFACTURE
27. TEXTILE MILL PRODUCT MANUFACTURE
28. TRANSPORTATION EQUIPMENT AND FACILITIES

Figure 3. Major nodes of Facet II (application).

and 200 more specific terms. The application facet contains 28 major nodes and approximately 725 more specific terms at three levels of specificity. Figure 3 lists the 28 major nodes of Facet II.

The nine-digit code used in conjunction with the classification scheme consists of a four-digit field representing Facet I and a five-digit field representing Facet II. Thus, digits 1 to 4 always represent function, and digits 5 to 9 always represent application. Note that the notation, as well as the classification, is in hierarchical order.

Traditionally, classification has referred to a physical or conceptual arrangement of information or physical objects, such as books, where each classified item occupies only one niche in the classification scheme. However, in actual practice, items rarely fit neatly into only one classification category and could properly be classified under multiple categories. In contrast to a traditional classification system, the Chemical Use Classification System permits a chemical substance to be classified in more than one category, owing to the multiple uses of individual chemicals. Thus, this classification system can be used in the coordinate indexing sense; i.e., the number of categories assigned to a chemical is limited only by the capability of the system.

To ensure internal consistency and promote understanding by its users, a classification system must abide by a set of rules, or conventions, governing the various aspects of organization and terminology. In establishing conventions, we have attempted to provide a set of rules which will be compatible with accepted industrial and government organizational and terminological practices. The conventions developed for this classification system cover terminology, display, use of scope notes and qualifiers, and index format. Highlights of these conventions are discussed below.

In choosing terms for the classification, we have emphasized terms common to industrial and government usage, as selected from publications such as the "Standard Industrial Classification", "Chemical Week Buyers' Guide", and the "Materials Handbook". By using terminology that is accepted throughout industry and government, we hope to facilitate user recognition of classification categories. Terms are generally in the plural form and are displayed in upper case only. Each more specific hierarchical level is indented beyond the level above it, and all headings at the same hierarchical level begin in the same column. For multiword terms, direct entry, rather than inverted entry, is used.

To give more specific meaning to individual terms which do not obviously convey the meaning intended without ambiguity, scope notes and qualifiers have been used. Scope notes are used both to define the scope of a category and to distinguish between terms that may appear to have overlapping meanings. Qualifiers are used to distinguish between homographs.

Because the classification system is arranged in hierarchical, rather than alphabetical order, an alphabetical index is required to aid users who are not familiar with the system. Every term that appears in the classification system also appears in the index. Terms are alphabetized, and each term refers to a specific location in the classification by means of the notation for that term. Only terms that are considered valid (i.e., preferred) in the classification are listed with notation.

We have included cross-references in the index to the extent necessary to direct the user to the proper categories of chemical function and application. For example, we have included variations in terminology that are likely to occur in industrial practice, but we do not normally include term inversions, since our conventions state, and users are advised, that the direct entry form is preferred.

CLASSIFICATION SYSTEM DEVELOPMENT

In an attempt to avoid unnecessary duplication of effort, we evaluated existing pertinent classification systems to determine whether any of them could be incorporated into the planned classification system either as is or with modification. We found that while no single classification could serve as the basis for both facets of our classification, the "Standard Industrial Classification" (SIC) could serve, with some modification, as the foundation of the application facet.

Because no single existing classification system contained an exhaustive enumeration of all significant functions for Facet I (function) development, we conducted a literature search to identify a number of sources which were most likely to provide, when used as a group, the comprehensiveness, specificity, and exhaustiveness of function categories required. Each of the 34 sources identified was reviewed to identify appropriate function categories. The titles of identified categories were then standardized and arranged in hierarchical order in accordance with the classification conventions.

For Facet II (application) development, we found that the SIC was the most compatible existing classification system. Because the SIC covers establishments engaged in all statistically significant fields of economic activity, it has sufficiently broad comprehensiveness in the application area. It is also exhaustive and has adequate depth of specificity. Therefore, selected SIC categories were used in the application facet with some modification to accommodate the conventions established for the Chemical Use Classification System.

CLASSIFICATION SYSTEM TESTING

Plans call for testing the classification system to determine its practicability in a real-world situation. To adequately test the classification system, it is necessary to select a representative sample of the classification system's intended users. Thus, we plan to include in our test not only the two categories of users specified in the TSCA (i.e., manufacturers and processors) but also the full range of company sizes. We consider it especially important to include small companies and nonmanufacturers, because they are less likely to have at their disposal chemists who are familiar with the uses of their products, and therefore the test would be more indicative of real-world problems. We have based our selection of test companies on two criteria: (1) the company must be a manufacturer, manufacturer/processor, or processor of chemicals; (2) the company must sell its products for use in

more than one industrial application. The second requirement will ensure more meaningful test results. If a company's entire product line had only one end-use, the coding of its products would be oversimplified, and potential problems may escape notice.

Because we cannot determine in advance the specific chemicals or categories of chemicals which will be reported via the classification system, our major criterion for selecting test chemicals was that they be commercially marketed and not restricted to research applications. In addition, we have selected chemicals used in representative industrial applications. To the extent possible, we have selected chemicals handled by two or more companies. This overlap among chemicals will enhance the test of the classification system by indicating the correlation between companies in interpreting the use of the classification system. Any changes in the classification system deemed necessary as a result of the test will be incorporated into the system.

CONCLUSION

Development of this classification system addresses a vitally important component of EPA's functions under the TSCA: the collection of chemical use data in a form that permits subsequent evaluation and analysis. The end goal of classification system development is a data collection instrument which will allow manufacturers and processors of industrial chemicals to report codified chemical use data to EPA.

LITERATURE CITED

(1) U.S. Congress Senate, Toxic Substances Control Act, Pub. L. 94-469, 94th Congress, 2nd Session, S. 3149, Oct 11, 1976.
(2) U.S. Office of Management and Budget, "Standard Industrial Classification Manual", U.S. Government Printing Office, Washington, D.C., 1972.

CHEMFILE: An In-House Information System for the Chemical Indexing of *Abstracts on Health Effects of Environmental Pollutants* (HEEP)†

WILLIAM GRAHAM*
BioSciences Information Service, Philadelphia, Pennsylvania 19103

Received June 1, 1977

The inclusion of a Chemical Abstracts Service (CAS) Registry Number index in HEEP has led to the need for a special database designed to link substance names with their appropriate CAS chemical compound Registry Numbers. Begun in 1968 as a tape record with batch mode updates and few record modification capabilities, the information system, which we call CHEMFILE, has evolved to its current form of disk storage with on-line access for file maintenance.

INTRODUCTION

The BIOSIS CHEMFILE is an information system developed to assist in the assignment of Chemical Abstracts Service (CAS) Registry Numbers to substance names occurring in the biological literature. Currently only 23 000, or approximately 10% of articles reviewed by BIOSIS, have substances uniquely identified but the commitment to this technique and the system needed to perform it provide BIOSIS with an operating model that allows us to learn the requirements that must be met to allow this indexing for other BIOSIS products. We now include CAS Registry Numbers as index terms in the printed version of *Abstracts on Health Effects of Environmental Pollutants* (HEEP) and the tape record of abstracts and citations dealing with drug toxicity now sent to the National Library of Medicine for inclusion in the TOXLINE database.

START OF CHEMFILE

In 1972, BIOSIS began publication of HEEP, a subset of abstracts and citations covered by BIOSIS and the National Library of Medicine. The project specifications required that a separate index of Registry Numbers for substances mentioned be included. CAS has cooperated in this work by providing Registry Numbers which could not be located in the printed indexes to *Chemical Abstracts* or other Registry Number sources.

Since the number of different substances encountered in HEEP was relatively small and the frequency of occurrence high, a file which BIOSIS now calls CHEMFILE was developed to save having to locate any given Registry Number more than once. A file similar in nature begun for an earlier project was built upon as new substances were identified. A record was prepared for each substance which contained the following types of data:

BIOSIS accession number
CAS Registry Number
Molecular formula
CAS Type 1 Name
Synonyms such as trade, generic, and systematic names
Wiswesser Line Notation

The serially assigned BIOSIS accession number tied all of the information together. It was used, rather than the CAS Registry Number, since substances and publication items could be processed further while the Registry Number was being located. Further, since not all chemical entities impacting on toxicology studies are registerable, we wanted to retain what information we had—to avoid looking for a Registry Number more than once if it did not exist.

CHEMFILE STRUCTURE

Each record, originally on magnetic tape, contained the accession number, a code for data type (10 for Registry Number, 20 for molecular formula), and the content of the record. The tape record was fully listed periodically to provide a hard copy for reference. The principal tool derived from the file was an alphabetical listing of all substances encountered. A third output was a numerically ordered listing of all Registry Numbers in the file. This was checked manually to assure that no Registry Number would be entered more than once.

The file was updated in large batches by inputting data from keypunch cards. Various validation checks, such as check digit

† Presented at the 11th Middle Atlantic Regional Meeting of the American Chemical Society, Newark, Del., April 20–22, 1977

* Present address: AMP Inc., Harrisburg, Pa.