man. This is especially marked during the first six years of this period. The maximum is reached in 1930–32, with over 1,400 such titles. The reasons for this increase are complex. In the beginning it is due to a very marked increase of titles in Russian, but part of this increase, and indeed a considerable part, is due to a rise in national feeling in the smaller countries. Finns felt that they must write in Finnish and Czechoslovakians in Czech. The languages run the gamut from Arabic to Turkish. Before this time, scientists in the smaller countries wrote for the most part in English, German, French, or Italian. These four languages were usually the only ones officially recognized at international scientific congresses. Toward the end of this period an increase in the number of titles in Spanish and Portuguese can be noted, most of which were of South American origin.

The "All Others" curve starts to decline after 1935 and, with the beginning of World War II, declines almost as rapidly as that for German publication, reaching a minimum of slightly over 300 titles in 1945. Of these, nearly one-third are of South American origin.

It would seem that these results have a story to tell. It is true that during the years of the recent war many psychological reports were issued which never reached publication. This is certainly true in the United States, where so much "classified" material was prepared by psychologists within the Services or under contracts with the National Defense Research Committee, the National Research Council, and other agencies. The writer doubts if much of this material will ever be published, partly because of continued classification and partly because the research problems were of a special character which made them useful during the war but frequently of no great scientific or systematic interest after the war is over.

One can only conclude that systematic science cannot flourish during wartime or in a political situation in which the scientists do not have freedom of thought and in which there are continued elements of uncertainty.

A New Classification System for Chemical Compounds[1]

D. E. H. Frear, E. J. Seiferle, and H. L. King[2]
The Pennsylvania State College, State College, Pennsylvania

IN THE COURSE OF SOME RESEARCH INVESTIGATIONS in which the authors have been engaged for the past several years, a simple system for classifying chemical compounds was required. For the purposes of this investigation it was necessary to (a) order the compounds in such a way that an individual compound could be located in a file containing several thousand other compounds with a minimum of effort and possibility for error; (b) arrange the compounds, in so far as possible, so that related compounds would be grouped together in the file; and (c) make possible the collection and statistical study of data on the frequency of occurrence of all chemical groupings appearing in the compounds under study.

After a number of trials with existing classification and indexing systems, including molecular formulas, alphabetical arrangements, and several others, it became apparent that each of these failed, in one way or another, to fulfill the requirements of this investigation. Accordingly, a classification system was devised especially for the work at hand, using an approach which the authors believe to be unique. This system was first applied in substantially its present form in 1943, and since that time it has been used successfully on approximately 8,000 different chemical compounds. These compounds comprise a group on which insecticidal and fungicidal tests had been recorded and cover a wide range of composition. Both organic and inorganic compounds are included.

Briefly, the present classification is based on "code numbers" assigned to each chemical compound. The code number for a particular compound consists of one or more group numbers, depending upon its type and complexity. These group numbers are assigned by referring to a prearranged list in which constituent chemical groupings (not necessarily functional groups) are given numerical designations. *The groups are listed in decreasing order of complexity.* This is probably the most important feature of the present classification system and, in the opinion of the authors, is the greatest single factor contributing to its workability.

The list of major families is given in Table 1, with one or more examples under each.

Parenthetically it should be noted that the presence of the elements O, N, S, or X (halogen) in the constituent group determines into which family the group falls and is thus the criterion of the complexity of the group. The carbon atom may or may not occur in

[1] Authorized for publication on 22 June 1946 as Paper No. 1334 in the Journal Series of the Pennsylvania Agricultural Experiment Station.
[2] The second-named author is grateful to the General Aniline and Film Corporation for approving devotion of part of his time to this project.

TABLE 1

FAMILY	EXAMPLE		
I	(CH)ONSX* groups		
	Chlorosulfonamides	$-SO_2NHCl$	3
II	(CH)ONS groups		
	Thiocarbamates	$-OC(:S)NH_2$	50
	Sulfonamides	$-SO_2NH_2$	56
III	(CH)ONX groups		
	Chloroamides	$-CONHCl$	102
IV	(CH)OSX groups		
	Sulfonyl chlorides	$-SO_2Cl$	151
V	(CH)NSX groups		
VI	(CH)ON groups		
	Amides	$-CONH_2$	185
	Nitro compounds	$-NO_2$	206
	C_4ON ring systems		230
VII	(CH)OS groups		
	Sulfonates	$-SO_3H$	258
VIII	(CH)OX groups		
	Acylbromides	$-COBr$	330
IX	(CH)NS groups		
	Thiocyanates	$-SCN$	401
X	(CH)NX groups		
	Chloroamines	$-NHCl$	477
XI	(CH)SX groups		
	Sulfenyl chlorides	$-SCl$	521
XII	(CH)O groups		
	Carboxylic acids	$-COOH$	541
	Hydroxy compounds	$-OH$	581
	C_4O ring systems		625
XIII	(CH)N groups		
	Azo compounds	$-N:N-$	665
	Amines, primary	$-NH_2$	671
	C_5N ring systems		730
XIV	(CH)S groups		
	Thiols	$-SH$	791
	C_4S ring systems		825
XV	(CH)X groups		
	Organic chlorides	$-Cl$	851
XVI	C(H) groups		
	Fused C_6–C_6 ring systems		924
	C_6 ring systems		951
	Aliphatic C_{10} groups		991
	Aliphatic C_2 groups		1011
XVII	Cationic groups		
	Calcium	Ca	1126
	Potassium	K	1196
	Sodium	Na	1218
XVIII	Anionic groups		
	Bromate	$-BrO_3$	1274
	Phosphate, ortho,	$\equiv PO_4$	1356
	Sulfate	$=SO_4$	1389

* The symbol X is used in this table to indicate any halogen.

each group and, if present, acts solely as a "nucleus" from which depend the other elements; hydrogen may be present coincidentally to complete the valence requirements of one or more of the elements present.

To determine the code number for a given compound, a very simple procedure is followed. The list of constituent groups is read downward until the most complex group present in the compound is encountered. The corresponding number is noted, the perusal of the list is continued until the second group is encountered, and so on until the entire compound is coded.

A few examples and their code numbers follow:

(1) 2-Aminoethanol, $H_2NCH_2CH_2OH$ = 581–671–1011
(2) Metanilic acid, $H_2NC_6H_4SO_3H$ = 258–671–951
(3) 2-Dodecylpyridine, $(C_5H_4N)C_{10}H_{21}$ = 730–991
(4) Acetic acid, sodium salt, CH_3COONa = 541–1011–1218
(5) *N*-Chloro-*o*-nitrobenzenesulfonamide, $O_2NC_6H_4SO_2NHCl$ = 3–206–951

When compounds are arranged according to the numerical order of their code numbers, a regular and orderly progression is apparent. For example, acetic acid, 541–1011, is followed by calcium acetate, 541–1011–1126, potassium acetate, 541–1011–1176, and sodium acetate, 541–1011–1218. In most cases compounds having similar structures are grouped together when arranged according to this system. All simple phenols and monohydric alcohols, for example, are grouped under 581. To locate a compound in a file arranged according to code numbers, it is a relatively simple matter to work out the code number and determine the proper location in the file within a minute or two.

This system is very well adapted for studies involving the correlation between chemical structure and physical, chemical, biological or other properties. By the use of punch cards listing the constituent groups in each compound under study it is possible to segregate from a collection of compounds all those having, for example, an amino group or any other constitutent group whose properties it is desired to investigate.

At the present time the Chemical Codification Subcommittee of the National Research Council Insect Control Committee is revising the system in order better to adapt it to use with machine-sorted punch cards. Full details of the final system will be published as soon as this revision has been completed. In the meantime, comments and suggestions on the coding system and its possible adaptations will be welcomed. Communications may be sent to the authors or to the Chemical Codification Subcommittee, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.