motes equitable evaluations. It is important that those seeking support should consider carefully which agency should be approached, since oftentimes proposals are clearly outside the agency's orbit; yet they must be reviewed. The importance of careful budget calculation and preparation also was stressed. Bad budgeting can influence a review panel into low-rating a proposal that otherwise might have considerable merit. Bad budgets, incidentally, can be too low as well as too high.

Kent Wilson, Head of the Chemistry Section of the National Science Foundation, which supports basic chemical research in universities, stated that a plurality of reviews assures optimum chances for funding a proposal. The program officer in a funding agency and the proposal writer share a common objective—to fund a worthwhile activity. On that basis, Wilson endorsed the careful budgeting suggestion, but added that a "pet peeve" from the funding agencies' side is that of proposals that reflect "bad science." Often, these same proposals reflect that "where's the check?" attitude which proposal reviewer's decry.

OVERLAPPING REQUESTS

Harold Wooster, of the National Library of Medicine's Lister Hill Center for Research in Biomedical Communications, addressed the question of overlapping support requests. He was supported by the other panelists in suggesting that multiple-funding of a worthwhile project is a highly desirable procedure in the allocation of resources for support of research, but suggested complete

disclosure of all sources of support. He declared that grants and contracts were in essence synonymous, at least as administered by the Air Force.

INFORMATION SUPPORT IN BASIC RESEARCH

The final question dealt with the effect on the prospects for obtaining funding for basic research in chemistry if a significant amount of the budget deals with the support of information services, especially those derived from computer-based systems. Panelists pointed out that such support is provided in virtually all proposals. Further discussion disclosed that favorable consideration might be a problem if the amount for information services exceeded 2 or 3% of the total budget. The reasonableness of the higher costs of computer-based services, as contrasted to purchase of books or journal subscriptions, however, was not established, and the panelists concurred that the nature of the work would be a determinant.

The plight of the Federal agency project or program officer was summarized as a "catch 22" situation. If the project officer makes his own decisions without widespread review, he could be accused of highhandedness. If he accepted the concensus of the reviews and in effect became merely a vote teller, he should be accused of cronyism. The audience, although sympathizing with this plight, indicated by its questions that the proposal preparer was very much the underdog in all situations, and therefore deserved more sympathy than he normally is afforded.

A Natural Document Retrieval System for Macromolecular Chemistry

RAIMUND ULBRICH and JUTTA WIERER
Deutsches Kunststoff-Institut,
6100 Darmstadt, Schloßgartenstr. 6 R, Germany

An indexing system for chemistry and technology of macromolecular substances is sketched out, whose characteristics are convenience of use and low cost. The selection mechanism consists of a set of optical coincidence cards. The system is a result of 15 years experience in the German Plastics Institute.

It is a concomitant symptom of the increasing need for documentation facilities in the chemical industry that notation systems for chemical compounds are described more and more frequently in the scientific literature.[1-8] Some of those systems require in practice an expensive employment of personnel and specialized equipment; therefore, they can be supported only by big companies. However, numerous "simple" and inexpensive systems, which would have a quick and convenient manipulating procedure, have been described in the literature. Nearly all of these "simple" systems operate with a sophisticated code which has an attractive universality and an absence of discrepancies, and so they are suitable for documentation by computers. But unfortunately their symbolism becomes a barrier between the creatively working chemist and the information store. There are different reasons

why the laboratory chemist does not learn the "simple" code willingly or does not use the assistance of a documentation expert for his literature searches.

On the other hand, when the chemist himself has personal contact with the information file and does not hesitate to tap it "with his own hand," then psychological and objective benefits are the result. "The chemist must have the sources in his field at his fingertips including . . . the indexing and organizational characteristics . . ." said Dr. Erwin Klingsberg of the American Cyanamide Co.[9] However, the technique of enquiry must be easy for the chemist. He should have to learn as few documentation skills as possible. Therefore, symbolism too must be avoided as far as possible. What matters, is not to provide a big information center with the most sophisticated indexing system, but to place the information store at easy

disposal for the staff of laboratories at any time and to ensure by its organization, that it is used readily and without hesitation; besides this, the information must be obtainable in the minimum of time. Therefore, a great part of proposed indexing systems for chemistry, which use a symbolism, remain only suggestions without having been actually proved over a long period.

For the reasons mentioned above, the use of symbolic notation has purposely been avoided in the documentation system of the "Deutsches Kunststoff-Institut" (German Plastics Institute, DKI). This system operates with clear word-descriptors alone. The latter are chosen in such a way that they should be familiar to each expert in the field in question; it is possible to pass these clear word-descriptors to the information store without the application of any symbolism. The real selection of admissible descriptors is based on a reasoned scheme of definitions. This scheme has been repeatedly modified in the past. So the system of descriptors written up in the following is in one respect the result of a systematic arrangement, and on the other hand, a result of pragmatic accommodation to practical experience. Such a procedure confirms and verifies itself particularly when literature searches are intended to find new concepts.

The documentation of DKI is at the disposal of a broad and heterogeneous group of external users, and this fact finds its expression in the choice of the descriptor system and the operation method. The tool for enquiry is an inexpensive set of optical coincidence cards which can be used anywhere. It is shown more and more often that optical coincidence cards maintain their importance on account of their specific advantages.[10]

RULES AND DESCRIPTORS FOR ORGANIC COMPOUNDS IN GENERAL

The general subject area of the documentation of DKI covers chemistry, physics, and technology of macromolecular substances, especially the well-known commercial types of thermoplastics and duroplastics, also the materials for fibers and rubbers. An important part of the methodology used in the DKI concerns the notation of chemical compounds as well low molecular as macromolecular ones. Since, in the chemistry of plastics, organic compounds predominate, a kind of fragmentation principle is suitable, which is based on well-known structure concepts of organic chemistry. In this system a given compound—first we treat low molecular compounds—is reduced by certain rules to its basic skeletons and functional groups. For this characterization, the following types of descriptors are used:

1. For carbon-chains and rings (basic skeletons)
2. For the number of chain and ring members
3. For the number of rings
4. For the number of substitutions
5. For functional groups
6. For the total number of carbon atoms

The practical work presumes an accurate definition of these descriptors and of rules for their application. But in this respect, one tries to get along with as few rules as possible. The given rules should be "natural" for the chemist.

Chain skeleton is every linear sequence of carbon-atoms (also a single C-atom) with the bondings between them. Branched structures are reduced to linear main chains and side chains. The main chain is the longest carbon-chain—in the case of chains with equal length, the one that contains the most multiple bonds or substitutions. Chains with multiple bonds have precedence over saturated chains of equal length; chains with triple bonds have precedence over chains of equal length with double bonds. The carbon atoms of such a chain skeleton are the chain members, C-atoms with functional groups (carbonyl-, carboxyl-, nitrilo-groups) are counted as chain members of the basic skeleton. In the description of the chain skeleton, the type of the chain, the number of chain members, the number of substitutions, and the position of multiple bonds are indicated.

A **ring skeleton** consists of the ring forming atoms of a single ring and the bonds between them. For the description of a ring skeleton, one has to indicate the type of ring, the number of ring members, the number of substitutions, and the position of multiple bonds. For the entire compound, the number of rings has to be noted. Two or more ring skeletons with common ring atoms are defined as **ring systems** (fused ring systems, bridge ring systems, spiro compounds). The number of rings in the ring system as well as in the entire compound must be indicated.

Examples of descriptors for skeletons are: *alkane skeleton, alkene skeleton, alkadien skeleton, alkapolyene skeleton, alkyne skeleton, cyclane skeleton, cyclene skeleton, cyclodiene skeleton, cyclopolyene skeleton, heterocyclane skeleton, heterocyclene skeleton,* and so on. Descriptors for skeletons are used only if the concerned chain or ring has at least one substitution; unsubstituted chains or rings have special descriptors (*alkane, cyclane*). For instance the notation of trichlorobenzene uses the descriptors: *benzene skeleton, substitution threefold,* and *chlorine.* The descriptor *benzene,* however, is used only for the characterization of the unsubstituted benzene.
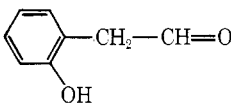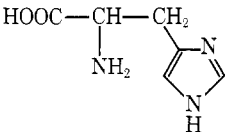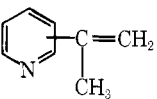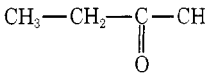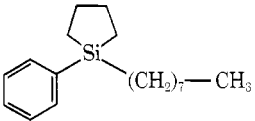
Substitution means replacement of an H-atom at a basic skeleton (by ring, side chain, or hetero atom). The number of substitutions is indicated separately for each basic skeleton. If a compound incorporates a ring system, then the number of substitutions is referred to each entire ring system and not to any ring skeleton belonging to the system. Nevertheless statements about substitutions are not related to the hetero atoms of functional groups or elemento-organic compounds, or to inorganic compounds, but solely to a basic skeleton—e.g., *substitution threefold* may concern the ethane skeleton of trichloroethane, but not the aluminium atom of triethylaluminium.

The above mentioned descriptors for skeletons (chains and rings) are rather general. To save labor, the use of more specific descriptors is prescribed in some cases. The general descriptor *alkane skeleton* is not used if the number of the chain members is n = 1, 2, 3, 4. In these cases, well known expressions such as *methane skeleton, methyl, methylene, ethane skeleton, ethyl, propane skeleton, butane skeleton* are used as descriptors. Besides this, in a few cases, isomerism is taken into account: *propyl-, isopropyl-, n-butyl-, sec-butyl-, tert-butyl-, isobutyl-.* Accordingly there are also descriptors for more specific ring skeletons—e.g., *cyclohexane skeleton, cyclohexene skeleton, benzene skeleton; pyrrole skeleton, pyridine skeleton,* etc. To maintain uniform registration, the obligatory rule orders the use of the most particular of all available descriptors in every case. Using such generally well-known designations for the most commonly occurring skeletons saves a lot of labor in the notation. Thus, enquiries are facilitated especially for people who are not so familiar with the card index.

For more accurate designation of carbon chains, carbon rings and heterocyclic structures, statements about the position of the triple or double bonds serve as further descriptors. In the heterocyclic ring structures, descriptors for the nature and number of hetero atoms are noted.

Besides the number of chain or ring members, the num-

Table I

Compounds	Names	Descriptors
		Phenylene, ethane skeleton, substitution threefold, aldehyde, monophenol, number of rings 1, C-atoms 8
	Histidine	Imidazole skeleton, substitution single, propane skeleton, substitution fivefold, amino acids, number of rings 1, C-atoms 6
	Norbornene	Bridge ring system, number of rings 2, cyclopentane skeleton, cyclopentene skeleton, C-atoms 7
	Isopropenylpyridine	Pyridine skeleton, number of rings 1, isopropenyl, substitution single, C-atoms 8
		Butane skeleton, substitution two-fold, ketone, C-atoms 4
		Heterocyclane skeleton, number of ring members 5, number of hetero atoms 1, silicon, organo silane, substitution twofold, phenyl, alkyl, number of chain members 8, number of rings 2, C-atoms 18

ber of the C-atoms must be indicated for the entire compound.

Functionalities are indicated by corresponding classes of compounds and refer to the entire molecule. C-atoms included in functional groups (e.g., carbonyl-, carboxyl-, nitrilo-groups) are counted as chain members of the skeleton. Descriptors for functionalities are, for example: *alcohols, ethers, aldehydes, ketones, carboxylic acids, amines, thioles, thio ethers, halides* etc.

Reactions are characterized by the initial reactants and the end products; besides this, special descriptors for certain types of reactions are used—for instance, *polymerization, polycondensation, polyaddition, cyclization, hydrolysis, addition, oxidation, reduction, crosslinking,* and others.

A few examples are intended to demonstrate the application of the rules and allude to the sorts of descriptors used in the system (Table I). The procedure sketched above suitable for the indexing of organic compounds is simplified further by distributing particular descriptors to frequently occurring substances. So all important monomers (*styrene* etc.) and additives (*benzoylperoxide* etc.) each have one special descriptor. In this way, one avoids the use of combined descriptors for general conceptions of chemical structure (skeletons, number of C-atoms, etc.) in cases of substances as familiar as methyl methacrylate or chloroform. That means the saving of labor during analysis of the documents as well as a facilitation at enquiry. Of course, several additional descriptors and regulations are prescribed to comprehend the **elemento-organic compounds** and **inorganic substances.**

SPECIAL RULES AND DESCRIPTORS FOR POLYMERS

The method for characterization of low molecular organic compounds described so far is also the basis for the

notation of polymers. For this purpose, it is completed by some special descriptors and a few rules.

A definition of the concept **polymer** is not intended here. It is used as it is common in macromolecular chemistry. The very useful concept of **polymer class** is introduced in the system under purely pragmatic aspects. Therefore, certain terms of polymer chemistry attain the meaning of polymers classes by being named in a list (altogether 53). The scope of validity of these terms is in consensus with the general accepted usage (e.g., *polyamide, polyurethane*), or it is directly discernible from the wording of the descriptor (e.g., *polymers with phosphorus and carbon in the main chain*).

In the scientific literature, there are three important groups of polymer materials which require different ways of notation.

1. Polymers with a simple and well known structure, especially simple addition and condensation polymers. They are characterized most conveniently by their polymer class and their monomers.

Examples of such notations:

poly(vinylidene fluoride)	vinylidene polymers, vinylidene fluoride,
polyformaldehyde	polyacetals, formaldehyde.
poly(ethyl vinyl ether)	poly (vinyl ethers), ether, vinyl, ethyl, C-atoms 4.

In the last example, the monomer **ethyl vinyl ether** is indicated by the combination of 4 descriptors (*ether, vinyl,*

ethyl, C-atoms 4). In many cases, it is necessary instead of using the monomers to indicate the hydrolysis products of polymers to get a uniform notation. This is true especially for polycondensates.

Example of a notation:

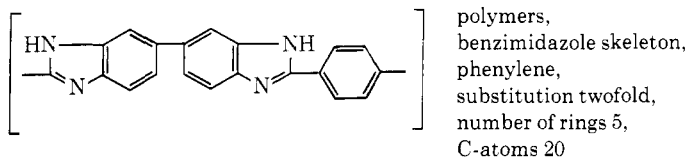poly(hexamethylene adipamide)   polyamide, hexamethylene diamine, adipic acid.

In cases where none of the provided descriptors for polymer classes is suitable, the general descriptor *polymer* is used.

Copolymers—i.e., all addition polymers of two or more comonomers—get the descriptor *copolymer* instead of *polymer* and, if possible, additionally, the descriptor of the polymer class. The monomers are indicated too.
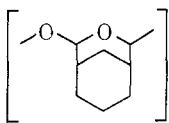
2. Polymers with a well defined structure as said in group 1. In contrast to polymers of group 1, the starting compounds are not recognizable from the polymer structure. Examples are materials like polybenzimidazoles or pyrrone polymers, which have been synthetized in the search for thermally stable polymers. In such cases, the repeating unit is characterized by descriptors in the same manner as a low molecular compound. But, it must be noted that fragments of chain skeletons, which are part of the polymer chain, are always main chains without regard to their length. Furthermore, the descriptor of the polymer class or the descriptor *polymers* is added.

Examples of notations:

Poly(phenylenedibenzimidazole)

polymers, benzimidazole skeleton, phenylene, substitution twofold, number of rings 5, C-atoms 20

Polymer of cis-1,3-diformyl-cyclohexane

polyacetals, bridge ring system, number of rings 2, cyclohexane skeleton, oxacyclane skeleton, number of hetero atoms 1, number of ring members 6, substitution twofold, C-atoms 8

3. Polymer materials with a complex constitution, which is not completely known—e.g., the **crosslinked resins**. Therefore, the structure is unsuitable as a basis of descriptors. These plastics are characterized by a descriptor for a polymer class (e.g. *phenole resins, alkyd resins, urea resins*) and additionally by notes about the starting materials. For instance the reaction products of phenol, *p*-n-butyl phenol and formaldehyde are characterized by the following descriptors: *phenol resins, phenol, formaldehyde, monophenol, phenylene, n-butyl, number of*

rings 1, C-atoms 10. In this case the *p*-n-butyl phenol is represented according to the above mentioned rules by the following descriptors:

p-n-butylphenol   monophenol, phenylene, n-butyl, number of rings 1, C-atoms 10

Sometimes a polymer can be considered as belonging to **several** classes. A product from phenol, urea, and formaldehyde belongs to two classes of polymers: *phenol resins* and *urea resins*.

Finally there are descriptors for special characterization: *block copolymers, graft copolymers,* and *oligomers.* All important and often occurring polymers, for instance, poly(vinyl chloride), or polystyrene, each get one special descriptor as in the case of low molecular compounds.

The compounds characterized by the preceding rules generally have a certain function which can be described by means of additional descriptors—e.g., *monomer, crosslinking reagent, catalyst, solvent,* and others. Further descriptors of the system are used for the notation of processes and machines for plastics processing, semimanufactured products (pastes, foils, tubes, fibres, etc.), technological properties (mechanical, thermal, electrical) and important fields of application. In addition, there are descriptors for the notation of general physical facts, and finally some descriptors for more general concepts.

The system consists of about 4000 concepts on the field of chemistry, physics, and technology of polymers, but only 1700 optical coincidence cards are used. The reduced number of cards is possible owing to the fact that many concepts are represented by a prescribed combination of two or more descriptors.

DOKUMENTATION EXPERIENCE OF DKI

During a year, about 10,000 to 12,000 papers from 200 scientific periodicals are analyzed (also preprints of important meetings). The information file, which was established in 1955, contains more than 170,000 documents today. The information file has been used, even by non-documentation specialists, for handling hundreds of search orders and retrievals.

The methods developed and proved over 15 years of use at DKI show a way that enables documentation of polymer chemistry and technology by simple means which do not discourage the chemist. It proves that even specialized questions can be answered by the information file, if one uses a suitable system of descriptors. An important part of the documentation method in question is the chemical structure fragmentation system for registration of organic compounds. Although this system has been developed for the chemistry of high polymers, its basis is so general, that it can be transferred with some additions to other domains of organic chemistry. Systems of the above described type attain more and more importance today, especially since the chemical industry seems to show the first indications of a certain weariness of the documentation complexities.[11-13]

LITERATURE CITED

(1)   Kokoropoulos, P., and Kankkanatt, S. V., *J. Chem. Doc.* 8, 179-87 (1968).

(2)   Livingston, H. K., *Ibid.,* 9, 131-4 (1969).

(3)   Lobeck, M. A., *Angew. Chem.* 82, 598-605 (1970).

(4) Maynard, J. T., J. Chem. Doc. 10, 285-9 (1970).
(5) Meyer, E., Angew. Chem. 82, 605-11 (1970).
(6) Rössler, S., and Kolb, A., J. Chem. Doc. 10, 128-34 (1970).
(7) Schier, O., Nübling, W., et al., Angew. Chem. 82, 622-8 (1970).
(8) Skolnik, H., and Hays, J. T., J. Chem. Doc. 10, 243-7 (1970).

(9) O'Dette, R. E., Klingsberg, E., Bloemeke, M. J., Garfield, E., Wigington, R. L., J. Chem. Doc. 9, 183-90 (1969).
(10) Shtejn, V. S., Zh. Vses. Khim. Obshchest. D. I. M. 16(1), 68-73 (1971).
(11) "Dokumentation eingestellt," Nachr. Dok. 22, 206 (1971).
(12) "Zweckentfremdet," Nachr. Dok. 22, 219 (1971).
(13) Sabiniewicz, S., Przem. Chem. 49, 461-5 (1970).

Multipackaging of Data at Source

CHARLES L. BERNIER
State University of New York at Buffalo, Buffalo, N.Y. 14214

Prompt handbooks and computer stores of data can be provided through cooperative multipackaging of data by authors and publication through existing organizations. Data scattered throughout primary and secondary literatures are costly to find, extract, evaluate, organize, report, and share. Techniques, media, and systems for making organized data available exist, but have not been used extensively enough. Reasons for failure to use these are explored and recommendations made for extraction by authors at time of writing, centralized organization, and publication through existing organizations. Datum signatures are described. Centralized extraction is shown to be unfeasible.

Authors embed data in documents. Extraction later is inconvenient, mainly because of scattering. Comprehensive, concise compilations of current, processed data (hereafter called simply "data") are not available despite awareness of need and despite what seem to be practical techniques, media, and systems for dissemination of data (Cf. all references). This paper explores reasons for failure to use these techniques, media, and systems for the benefit of all specialists (such as engineers, professional people, and scientists), and presents possible solutions to the problems involved in this failure.

For at least a century, the scientific community has been well aware that useful data are scattered throughout its primary and secondary literatures. Retrieval of data requires undue amounts of time of scientists to: search indexes, find documents (including computer tapes), copy data, evaluate it, organize it, report it, and sometimes share it by publication.[1,2]

Specialists waste valuable hours searching the wildernesses of their primary and secondary literatures for data. In view of the existence of handbooks and of computer stores of data, they do not understand why prompt, comprehensive handbooks and computer stores of data have not been provided to them.[3,4] Their time, including overhead, may average $200 a day. Handbooks of current data have not kept pace with outpourings of technical presses.[2] Specialists are well aware of the convenience of handbooks, despite need to use indexes to them, as contrasted with the inconvenience and cost of searching the primary literature for data, for example. Access to handy handbooks is a matter of seconds; access to handy computer stores is a matter of minutes; access to data of unknown location in the primary literature is a matter of hours, days, weeks, or even months. The cost of providing data conveniently and promptly is small compared with the cost of developing, calculating, recording, checking, and publishing the data in the first place. The cost is negligible compared with the combined value of such data to our civilization. Such value can be estimated from Gross National Products, longevity, morbidity, mortality, national incomes, etc. Making data more convenient seems to lie in the fields of economics and of psychology. We experience what can be termed, "fractured economics"—dollars (e.g., 21-24 billions per year) made readily available for research, while cents (e.g., hundreds of millions of dollars) are provided for convenient communication of the results. In the United States, we seem heavily to favor getting information from experience (e.g., experiment) over getting it from the record (e.g., the library).

PRIMARY PUBLICATIONS

Primary publications are not convenient sources of data because of:

Word Dilution. The primary paper or report is written to enable repetition of the work described without consulting the author. It is ideal for this purpose. Papers and reports are unsuitable for skimming—the rapid, convenient location of data.[5] Data are diluted by too many words. Words are seductive. Place a word in front of a scientist, and he will read it. He will not know whether