THE GMELIN INFORMATION SYSTEM

THE GMELIN HANDBOOK AND GMELIN FACTUAL DATA BANK

Ekkehard Fluck

Gmelin Institute for Inorganic Chemistry of the Max Planck Society for the Advancement of Science, Varrentrappstr. 40-42, D-6000 Frankfurt 90, Fed. Rep. of Germany

INTRODUCTION. DEVELOPMENT OF CHEMICAL LITERATURE IN EUROPE.

Allow me from the outset to review quickly the development of chemical journals in Europe. A considerable number of the earliest publications in the area of chemistry had their beginnings in Germany. In 1778 the first journal appeared: "Chemisches Journal für die Freunde der Naturlehre, Arzneygelahrtheit, Haushaltungskunst und Manufacturen" ("Chemical Journal for the Friends of Physics and Chemistry, Pharmacology, Household Skills, and Manufacturing"). The editor was Lorenz von Crell, who corresponded with many scientists in other countries, so that his journal contained contributions from France, England, and Sweden, in addition to those from Germany. Scheele, for example, published much of his original work in this journal. Besides this journal from Crell, who is seen today as the founder of chemical journalism, were a series of other publications. The pamphlet series, "Die neuesten Entdeckungen in der Chemie" ("The Newest Developments in Chemistry"), was among them; it appeared every six months. In its twelfth edition, von Crell justified the necessity of a new series, which would be published with greater frequency. His justification is as valid today as it was then. Namely: "After the establishment of my journals, the chemical notices, which I had collected for my readers, could be placed in their hands only once every six months, ... their praiseworthy curiosity for more knowledge was only later satisfied: Many perhaps used their efforts, time, and money in the meantime on such subjects, which they so often found previously treated, often still better: or they could perhaps finish their necessary work more easily, cheaper, and also even better ...".

How justifiable this concern still is today, is shown by estimates of costs to the German economy amounting up to 15 billion DM per year through unnoticed research results, even though they are published.

Shortly after von Crell's journal, the oldest chemical journal that is still published, "Annales de Chimie", appeared in France during the Revolution in 1789. It and others were edited by Lavoisier and Berthollet.

Up to 1841, Henry Carrington Bolton recorded 74 chemical journals in his "Catalogue of Scientific and Technical Periodicals 1665 - 1895", of which 38 were published in Germany, 8 each in France, England, and Italy, 5 in Holland, 3 in Belgium, 1 in Sweden, and 1 in Russia.

It is valid to this day that the use of journals in comparison to that of books is much more widely distributed in chemistry than in the other sciences.

THE GMELIN HANDBOOK OF INORGANIC CHEMISTRY

It is interesting that a long time ago, in spite of the relatively small number of journals, the necessity became recognized that the published knowledge therein has to be critically assessed and condensed, and organized according to substances. In 1817, the first edition of "Gmelins Handbuch der theoretischen Chemie" ("Gmelin's Handbook of Theoretical Chemistry") appeared as three volumes. Beginning with the fourth edition of 1842, an English translation was published by the Cavendish Society. With the beginning of the fifth edition, organic chemistry was published separately as the "Gmelins Handbuch der organischen Chemie" ("Gmelin's Handbook of Organic Chemistry") and later dealt with in an independent work, the "Beilstein Handbuch der Organischen Chemie" (the "Beilstein Handbook of Organic Chemistry").

The Gmelin Handbook is prepared by the Gmelin Institute of Inorganic Chemistry, which is part of the Max Planck Society of West Germany. The present Gmelin Handbook is already in its eighth edition. There are no plans for a ninth edition. Rather the eighth edition is considered as a framework to be continuously updated und expanded by additional volumes.

The format of this framework is provided by the Main Volumes - in general, one Main Volume for each element. To the Main Volume for each element are periodically added one or more Supplement Volumes. Of course, the Main Volumes and the Supplement Volumes are all parts of the single Gmelin Handbook.

In order to report even more promptly the newest developments in inorganic chemistry while maintaining scientific interrelationships, the number of Supplement Volumes to the eighth edition was increased in certain fields in recent years. Within the framework of recent Supplement Volumes, there are, e.g., series dealing with boron compounds, transuranium elements, perfluorohalogenoorgano compounds, radium and the actinide elements, water desalinization, and organometallic compounds.

As an example I may mention that there is one Main Volume on boron and its compounds, then the first Supplement Volume which appeared in 1954 and which covers the literature through the end of 1949. Since 1954 the subject has been updated with 27 Supplement and 1 Index Volumes.

The major topics covered in the volumes on boron are divided into chapters and further subdivided into sections and subsections. The data and concepts pertaining to boron and its compounds are organized and summarized within this framework. Thus, related information is compiled in the same place. It is this method of presenting information in context that allows the Gmelin Handbook to catalyze the reflection, the association of ideas, and, perhaps, even the inspiration that make up the innovative thought process. Many a new idea has been prompted or stimulated by browsing through the Gmelin Handbook.

In addition, the researcher has publications available such as Chemical Abstracts and Current Abstracts in Chemistry. These publications and their associated computer-linked information services provide the chemist with current literature references and abstracts. They, however, operate within an abbreviated time frame. Indeed, in order to up-to-date, they strive to make this time frame as short as possible. As a result the material they present is dispersed and highly diluted with respect to the question the individual chemist seeks to answer. It is true that an abstract index or an information service will frequently list several articles per key word but even in such cases comparison of abstracts is difficult if terminology and the manner of presentation are not uniform. It is left to the reader to collect related information, to discard what is not relevant to his needs, and to evaluate and compare what is left.

On the other hand, there are handbooks such as the Gmelin Handbook of Inorganic Chemistry that operate over a longer time frame. For the Gmelin Handbook, coverage begins with the literature closing date of the previous volume and extends up to shortly before publication date of the new volume.

The longer time frame allows Gmelin to compile enough material that has been published on an element or on a particular compound or group of compounds during a period of time, to give a complete picture of that element or compound. The information is critically evaluated, incorporated or discarded where appropriate, and correlated. The results of these efforts are then summarized for each substance under subject entries which follow a specific sequence.

Following the basic arrangement by substance, within each substance section thus defined, the knowledge available regarding the substance is presented in a standard subject matter sequence. The description for an element or a compound usually begins with the historical aspects, the occurrence or deposits for those materials derived directly from natural sources, the minerals, the industrial recovery or production technology, the toxic properties, if any, and the applications or uses. Then follow descriptions of its formation and preparation, the atomic and molecular properties, the physical properties, the electrochemistry, and finally the chemical behavior. In short, the reader is provided with a literature search for each substance covered, a literature search which is valid up to the literature closing date of the volume. It is typical of Gmelin that all properties and characteristics important for the identification of a substance are referenced directly to a literature citation.

Let me at this point describe briefly the history of the Gmelin Handbook. The founder of the Gmelin Handbook, Leopold Gmelin, was born in 1788 in Tübingen. He became Professor of Chemistry at the University of Heidelberg. In 1817 he published the first edition of his "Handbook of Theoretical Chemistry" in three volumes. They were designed to be used mainly by chemistry instructors in preparing their lectures.

The great success of the first edition of the Gmelin Handbook was brought about by three factors. The Handbook was much more comprehensive than any other chemical treatise of that era - it covered the entire field of chemistry. It was organized by subject matter, that is to say, by compounds. Leopold Gmelin put into practice the idea of compiling in one place all the facts pertaining to a particular compound - namely its preparation, its physical properties, and its chemical behavior. And just as is done today, the first edition substantiated each factual statement with a proper reference which unambiguously related all the reported data and concepts to the original literature. Furthermore, the scope and contents of the various chapters were varied according to the amount and type of information available.

Work on the seventh edition began in 1905. It was written by a group of scientists and served as a useful und important reference book when work was started on the current eighth edition of the Handbook in 1924.

The steady increase in chemical literature made it quite clear that the eighth edition of the Handbook could not possibly be written by a small group of part-time writers. Alfred Stock, professor of chemistry at Halle and noted for his brilliant research in the area of silanes and boranes as well as for the introduction of Roman numerals to designate oxidation numbers, took the initiative and convinced the German Chemical Society that a permanent team of chemists was necessary to do the work. Work on the eighth edition was commissioned by Professor R. J. Meyer of the German Chemical Society. In 1936 the Gmelin Institute started work in Berlin under the directorship of Professor Erich Pietsch. In 1945 the Institute was evacuated from Berlin to Clausthal-Zellerfeld; in 1957 it was moved to its present location in Frankfurt/Main. From 1969 through 1979 Professor Margot Becke-Goehring was director of the Institute.

Since 1945 the Gmelin Institute has been part of the Max Planck Society for the Advancement of Science. That in a nutshell is the 170-year history of the Gmelin Handbook.

What is the "Gmelin" of today, which elements and compounds are described, and how current is this description? To answer these questions, the Gmelin Institute has made the following material available:

1) the "Alphabetic Reference Chart" and
2) the "Complete Catalog".

The third and last way to find a specific compound in the Handbook is via the so-called "Formula Index". This twenty-volume index contains all defined elements, compounds, ions, and systems which had been described in the Gmelin volumes appearing up to the end of 1979. A supplement to the Formula Index will appear in 1988 and 1989 covering all entries up to the end of 1987. The arrangement is alphabetical according to the element symbol, so that location of a specific element or compound is again simplified.

CHEMICAL LITERATURE AND PROGRESS OF SCIENCE

The invention by Johannes Gutenberg in 1440 of the printing process with single letters, that is, the invention of a technical process had made it possible for the first time, to make information accessible to a large group of people and to exchange scientific ideas over a large geographical area. Scientific work comes about in a steady, self-renewing process of recording, production, and dissemination of information:

1) recording, duplication, and transmission of new scientific results;
2) acceptance of the duplicated information in the scientific community;
3) assimilation of this information in accordance with scientific principles and subsequent production of new results, which are then distributed as new information, starting up the cycle again.

This portrayal makes clear why scientific productivity is closely coupled with the publication and distribution of knowledge by the medium of the printed word. If knowledge is distributed through the printed medium, then the findings and data can be checked and reproduced or falsified anywhere at any time. Information that was distributed by books and journals was of essential importance for scientific development in the centuries since Gutenberg's invention.

After the second World War, the expansion of scientific knowledge was made easier by another factor, namely the standardization of the utilized language; English became the lingua franca of science just as Latin was centuries ago.

The last two decades saw the beginning of a new period that was characterized by the unheard of expansion of scientific results and data and further by the almost simultaneous development of electronic technology in data processing. The background for the mass production of scientific information and data was the intensification of scientific activity around the world. The number of scientists engaged in research and development in the industrialized countries has doubled since 1965. The yield of research and development grew not only in relation to the national product but much more quickly.

With the rising literature flood, the storage problem for libraries became worse and worse. The financial provision of colleges for this purpose, and this applies practically world wide, has not been able to keep pace with the increase of periodicals and monographs. Only a few libraries are in a position today to really ensure good literature support. One of them in Germany is the Technische Informationsbibliothek Hannover (Technical Information Library in Hannover). Why each library can no longer be complete is shown by some figures for the mentioned library. It maintains subscriptions to 3,400 journals which concern chemistry in a narrow sense, and to another 9,000 journals which are relevant to chemistry. Book holdings are alloted some 450,000 volumes and microforms in a narrow chemical sense and some additional 1.2 million volumes and microforms of the customary, chemically relevant literature. This requires more than 4 million DM annually solely for the purchase of journals and books.

One first attempted a remedy or at least an allayment by centralizing the institutional libraries. Most colleges today have only one central library. This leads to a reduction in subscriptions of scientific journals, amounting to 60 % in Germany. Likewise, the circulation of scientific monographs has fallen 40 %. This development leads to an increase in costs of single copies, that is, it causes in the end that less information will be distributed for the same price. In addition one must also bear in mind that literature acquisition is linked with the high resulting costs of storage and preparation. This is a reason that more and more of the electronic media must be employed for literature storage. It has already led to scientific information being marketed in considerable quantity by this means. The first data banks in the field of chemistry consisted of bibliographical material and abstracts. The mentioned expansion of information concerns primarily the primary and tertiary literature. It will without a doubt no longer be possible to store in the primary literature all data determined by chemical investigations. They should find direct entrance into the data banks. The means to this procedure was shown, for example, by the Journal of Chemical Research which is published according to the synopsis and microform system. The purpose of the synopsis is to offer the reader an overview of the most important results of the work. All details of the investigation are accessible to the specialist in microform, as either miniprint or microfiche. In the future, the data bank can be in the position to return the detailed publication in full text, or preferably, the data will be processed and stored retrievably in a factual data bank. In this way, the extent of printed primary literature can be reduced to a reasonable quantity without loss of information. At the same time, the researcher is given the opportunity to extract more easily an overview of his speciality and its relation to other areas of his descipline or perhaps even to other scientific disciplines. It is not possible in the long run to store all tertiary data in printed form - this is even now a question of storage and cost. On the other hand it would be uneconomical to let the data pass into oblivion solely on this basis. Their acquisition was associated with a large expenditure in most cases. The answer to this is that the data must be stored in a factual data bank.

The storage of facts in data bases has unassessable advantages: only in this way can the large quantity of data, as exists today and will assuredly grow in increasing quantity in the future, be space savingly archived and handled, and only factual data bases make possible rapid access to the data.

The factual data banks moreover have, in contrast to all other possibilities, the deciding advantage of permitting the combination of data as needed, for example, by the analytical chemist. He normally determines the composition and/or a series of properties of a material in order to establish its nature. Also in applied technology, one is often concerned about materials, which should possess certain properties. Answers to these types of questions cannot usually be given, or only inadequately and certainly not quickly by books, even if a library is reasonably complete. Data, which are archived in the primary literature, do not allow rapid access. Naturally, data comparisons are carried out in the primary literature at best in the narrowest framework. Extensive data comparisons, which for instance are used for structure-property correlations, can be found partly in the secondary literature but they again are selective owing to the abundance of material making the selection more and more fortuitous. Factual data banks simplify statistical analysis to a large degree. They are a prerequisite for chemometrics - yes, they even made this new branch of chemistry possible. Large quantities of data can be handled, assessed, and interpreted meaningfully only by automated data processing. The statistical analysis can on one hand be used for the generation of new data and on the other for data reduction. Another advantage of factual data banks lies in the fact that they permit the automatic checking of numerical data, that is, the implementation of plausibility testing. A single datum or a series of individual data can be extracted from traditional sources. Computer data banks allow, however, the analysis of a large number of related structures and the associated data. Thus one possesses a new and powerful research instrument that leads to results which could not be obtained by other means. In the future, the factual data banks will have their place next to the primary and secondary literature as the tertiary literature has always had up until now. However, they have incomparably larger possibilities than printed matter because they can be more complete, permit more rapid access of the data, and because they allow the data to be combined with each other and thus allow multidimensional formulations of questions. They, together with the primary and secondary literature will constitute a unified information system, in which primary and secondary literature will also be accessible on the computer screen in the course of time. This is already the case with the journals published by the American Chemical Society. Works such as the Gmelin Handbook can also be placed at one's disposal in full text on the computer screen, be it Online or by CD-ROM. Based upon today's technical level, the entire Gmelin Handbook of Inorganic Chemistry of nearly 600 volumes can be stored on about four CD-ROM disks and could in principle exist on every laboratory bench.

Where one finds topics in the Gmelin Handbook, can already be obtained online with the help of the data bank GFI, which consists of the Gmelin Formula Index and the Complete Catalog of the Handbook and all significant information for the location of an article about a specific substance and specific facts contained in the printed Handbooks. GFI is accessible through STN International.

OUTLOOK TO THE GMELIN FACTUAL DATA BANK

Factual data banks will not deliver new ideas. The mind does not think in terms of data but in terms of ideas and this is more an art than a science, and certainly not a technology. If, however, a new idea is born, the analytical thought process sets itself into motion. Today this process requires a great variety of data in the area of natural science. This procedural thought has its important place but it is placed in the corner, to which we must devise a travel plan. We must carefully study the road map, an activity, however, which only appears if we have determined to make the trip and established the destination. Certainly, only if the road map is exact and the instructions thorough will we quickly reach our goal without loss of time and money. Here lies the task of the data banks in general and the factual banks in particular. We need them today, if we want to pursue rational and economical research, achieve research goals, and convert new ideas into practice.

Toward the end I would like to say a few words about the future Gmelin Factual Data Bank. The main features of the data bank are the following. It will be a compound-oriented factual data bank with a fixed data structure containing information on purely inorganic compounds as well as organometallic and coordination compounds. The data are taken from the Gmelin Handbook of Inorganic Chemistry as far as possible, that is, they are evaluated data. The user can recognize this because in addition to the literature reference, the Gmelin reference will be given. Data, taken from literature which have not been treated as yet in the Handbook or which are extracted from the latest current literature, will be referenced to the corresponding literature source. The contents of the data pool are accessed through the data structure using a type of connection table format for covalent compounds and a Gmelin-own format for ionic compounds. The development of a special format for ionic structures was necessary in order to guarantee easy access to all available information on these materials. Besides compounds with integer indices there are those with index ranges or noninteger indices, isomorphous or diadoch compounds, intercalation and inclusion compounds, multicomponent systems of various types, alloys, doped materials, and so on. There will be, however, only one comprehensive data structure for facts on purely inorganic compounds and organometallic and coordination compounds with numerical and alphanumerical data and texts. The data are organized in information units and data fields. There will be 185 information units and 810 data fields. On a final level, this structure is composed of data records for the encoded substances. Data are recorded under headings such as:

1. Identification
2. General Information
3. Formation/Synthesis
4. Structure and Physical Properties
5. Properties of Multicomponent Systems
6. Electrochemistry
7. Chemical Behavior
8. Further Data
9. Bibliography

The structural and physical properties, e.g., are subdivided:

- 4.1 Molecular Properties
- 4.2 Phase Transitions (one component)
- 4.3 Information about Condensed Phases
- 4.4 Mechanical Properties
- 4.5 Thermal Properties
- 4.6 Transport Phenomena

and further subdivided:

4.1.1. Molecular Shape and Symmetry
4.1.1.1 Information about Conformation, Bonding Models
4.1.1.2 General Information On Molecular Symmetrie
4.1.1.3 Point Group
4.1.1.4 Internuclear Distances, Bond Angles, Dihedral Angles

method of determination:
electron diffraction
neutron diffraction
X-ray diffraction
microwave spectroscopy
.....

state of aggregation:
solid
liquid
gaseous
solution
.....

4.1.2. Electronic States
.....

These categories are forming information units in which the individual data fields are combined. One or more references are assigned to each information unit. A data field can belong to different types of classes dependent on its content and its use in subsequent queries.

The primary function of a substance-oriented factual data base is the arrangement of chemical knowledge about substances in such a way that it is possible for a suitable retrieval system to find the information. A few sample applications are:

Search for substances by way of proposed empirical formulas or fragments of formulas.

Search for a specified structural formula or structural fragment.

Search for substances possessing specific properties, for example, melting point, color, dipole moment.

Search for the chemical behavior of substances, for example, reaction of a substance with one or more reactants, as well as reaction with general substance classes, for example, N-oxides, phosphorus heterocycles.

Bibliographical search for authors and publication year.

Search queries can be combined using logical operators such as AND, OR, NOT.

Although the development of this data base is making good progress, the first online service will not be available before 1990. I may point out, however, that the Handbook and the Gmelin Online Data Base will independently continue to meet the information needs in chemistry. They will complement each other. The argument for continuing the Handbook in addition to the Gmelin Factual Data Base was given in an interview of "Chemistry in Britain" with Professor Henry Taube, who in 1983 received the Nobel prize for his work on the mechanisms of electron transfer reactions, especially in metal complexes. With respect to the beginning of his work he said: "So I decided to do something that in retrospect seems a bit unusual: I went to the Gmelin series, which had a whole volume devoted to the chemistry of cobaltammines. I just started to read this stuff! I found it fascinating". And he added: "I believe that chemistry is to a very large extent a descriptive subject".