Development and Operation of a Specialized Technical Information and Data Center (The Cryogenic Data Center)*

VICTOR J. JOHNSON
Institute for Basic Standards, National Bureau of Standards, Boulder, Colorado 80302

Received April 26, 1968

As is typical of many specialized areas of science and technology, cryogenics has had an explosive development in the last decade and a half. The Cryogenic Data Center was established in 1958 to cope with the problem of organizing the world's literature pertinent to the field of cryogenics and furnishing this rapidly developing industry with reliable information and data. The development of this facility over the 10-year period of its existence is discussed. Included is a description of the conversion from a manual to an automated bibliographic retrieval system, the development of cataloging and indexing techniques, the development of comprehensive awareness and literature acquisition procedures with public announcement of new literature, procedures for selecting and compiling data, and finally how the needs of the sponsors and the industry are being served. The discussion emphasizes the problems that are typical in facilities of this type and the need to consider standard or generally accepted solutions. The growing trend for a network system of technical information and data centers requires a high order of compatibility among them.

Cryogenic studies have been going on in physical science laboratories since the two Polish scientists, Wroblewski and Olszewski, first liquefied air in 1883. After Sir James Dewar liquefied hydrogen in 1898 and Kamerlingh Onnes liquefied helium in 1908, cryogenic studies were carried to very low temperatures (within a few degrees of absolute zero) thus greatly increasing the number of physical phenomena of interest to scientists. It wasn't until the early 1950's, however, that a real industrial interest in cryogenics developed, except perhaps for the use of liquid oxygen in smelting steel which started about the end of World War II. The first industrial size liquid hydrogen plant was put into operation by the National Bureau of Standards in 1952 at Boulder, Colorado, in response to needs of the Atomic Energy Commission.[1] At about this same time, liquid helium was also becoming more readily available as a result of the widespread marketing of the Collins helium liquefier, thus making possible a great increase in very low temperature research. Industrial interest soon followed with demands for equipment, instrumentation, techniques, and the cryogens necessary to obtain the low temperatures. To illustrate the rapid growth of industrial cryogenics during the decade of the 1950's, attendance at the Cryogenic Engineering Conferences rose from less than 200 at the first one in 1954, to twice the number at the second one held in 1956, to nearly 1000 for the 1960 conference. The interest was so great by 1956, that the conferences were held annually from then on.

* Presented before the Division of Chemical Literature, 155th Meeting, ACS, San Francisco, Calif., April 3, 1968

The first three conferences were held at Boulder, Colorado, since the Cryogenic Engineering Laboratory (now the Cryogenics Division of the NBS Institute for Basic Standards) had become a focal point for cryogenic research and development and is the nation's principal laboratory devoted to the study of cryogenic applications and to low temperature properties of materials.[2]

The Cryogenic Data Center was established in 1958 to help meet the need for low temperature information. As a Section of the Cryogenics Division of NBS it is not only responsible for maintaining a comprehensive worldwide awareness of cryogenic literature and activities of interest to the Division staff, but also to the cryogenic community and the public who use the Center. Included in this responsibility is the critical evaluation of properties of materials data and furnishing such information to requesters.

EARLY YEARS OF OPERATION

One of the first tasks undertaken was the cataloging of more than 1000 articles of cryogenic literature that had been acquired by the Division's professional staff during the prior years of the Laboratory's development and operation. This material was cataloged and crudely coded for bibliographic retrieval using edge-notched keysort cards (Figure 1). Also a number of literature searches that had been conducted for research projects were reviewed and processed into the infant Data Center system providing about a thousand more accessions.

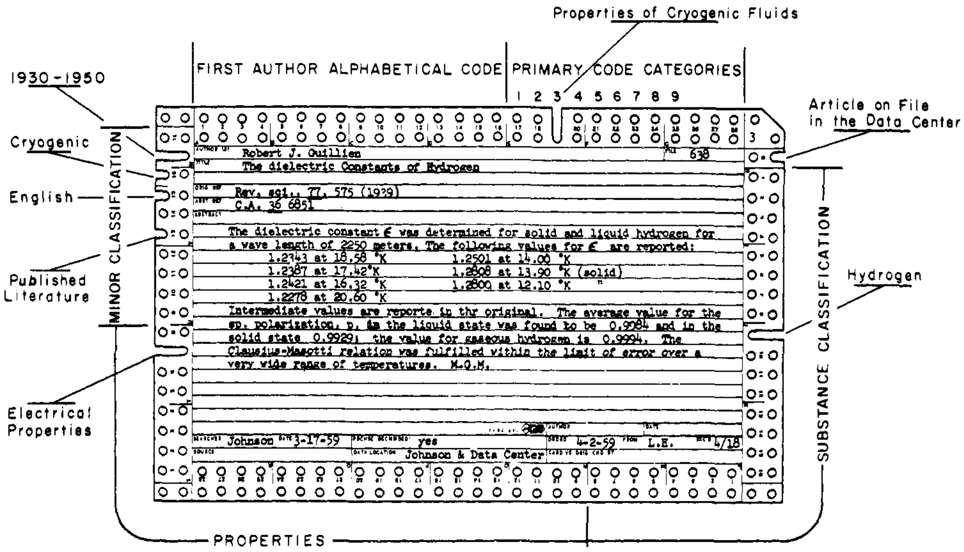

Figure 1. Sample edge-notched catalog card for key-sorting

At this same time, a data compilation project sponsored by the U.S. Air Force's Wright Air Development Center was assigned to the Data Center. Phase I of this project[3] was completed in 1960 and yielded about 2000 references which were entered into the manual storage and retrieval system. The total number of accessions on edge-notched key-sort cards was thus approaching 5000 and had already become too unwieldy to use making it important that we change to a better system. A number of alternate manual systems were studied including the "peek-a-boo" term card system which had undergone considerable development at the National Bureau of Standards in Washington.[4] A decision was soon made, however, to convert directly to a computer based system. This decision was partly based on the fact that the NBS Boulder Laboratories had installed a large central computer and the Director was urging all divisions to use it. A more important influence though was the desire to have all cataloging, coding, and other storage and retrieval information machine-readable so that repeat key-boarding would not be required in future developments or major changes in the bibliographic system.

The first automated program prepared was the *search program* made operational in 1961. Unfortunately, we had to prepare the search program ourselves for no programs seemed to be available at that time that were suitable for a small system such as ours. R. G. Smith, who had worked with a group at the University of Arizona on an information retrieval contract for the U.S. Army Electronic Proving Ground, joined our staff in October of 1960. In his former assignment he had learned to use the information retrieval methods developed by J. W. Perry and Allen Kent at the Western Reserve University.[5] Some modifications in Perry and Kent's search methods were made in that the telegraphic role indicators were omitted and a digital designation of code terms was adopted instead of the alphabetic designation. The coordinate indexing principle was retained though with much of the Boolean search logic, using both "positive," "negative," and "and/or" combinations of search terms in the query. Figure 2 is a greatly simplified outline of our current search program which, although considerably expanded and improved over the original program, is still much the same in principle.

SEARCH PROGRAM

An input query or queries for a search are matched with the search tape. In the preliminary search all possible matches are tagged. If Boolean logic is required by any of the queries, the preliminary results are processed accordingly before going to the numerical sort routine. Accession numbers are then sorted in ascending order by query and for the search. The numerical list of accession numbers is then matched with the catalog tape and the complete bibliography of references, including the coding and other indexing, is printed out. The accession number list for each query is also printed out as an index to the search.

CATALOGING FOR AUTOMATED RETRIEVAL

Some modification of the cataloging was made in the conversion to a computer based operation to facilitate

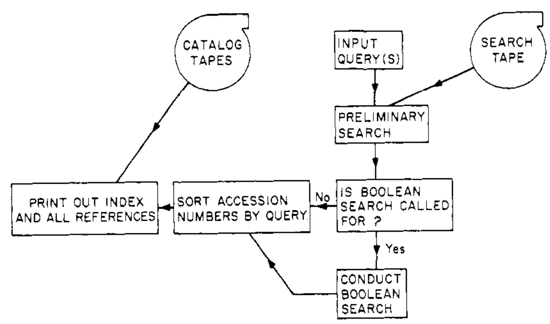

Figure 2. Search program

machine handling and identification of such items as authors, title, source citation, etc. (Figure 3). (Further modification was made to conform to the revised COSATI Standard for Descriptive Cataloging[6] in processing report literature.) At first only the coding terms were key punched and processed onto the search tape (Figure 4). The complete reference citation and coding information used to be stored on punched paper tape which provided machine readability and semi-automated printout for bibliographies. However, numerous local difficulties including the slow speed of the paper tape typewriter made us abandon the use of the paper tapes and go to a fully computerized operation of reference storage and printout. The quality of the printing and readability of computer printed bibliographies was considerably poorer than the paper tape typewriter, but the speed of operation was greatly improved.

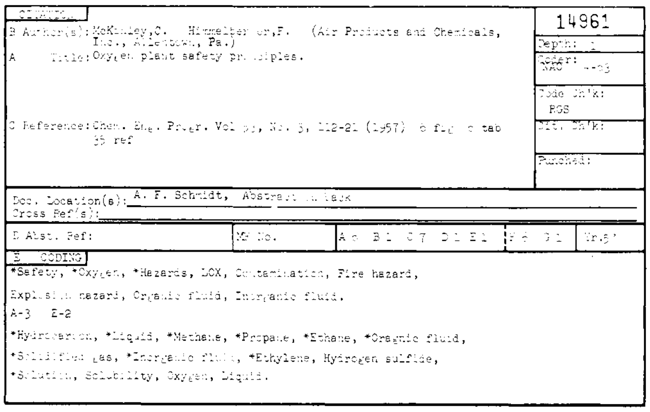

Figure 3. Sample master catalog card used for input to the bibliographic system

CODING VOCABULARIES (DICTIONARY)

Two levels of coding terms were set up to identify the subject matter and our interest in pertinent literature. We designated the upper level as "subject terms" and the others as "descriptors." The two levels are not hierarchical in the usual sense but instead provide a means for designating major and minor subject matter in the articles coded and their pertinence to the cryogenic field of interest. Subject terms are designated with an asterisk by the coder and the numerical designation is a 3-digit number preceded by a minus sign giving it the equivalent of four digits in the search tape. The descriptors are designated with straight 4-digit numbers for search purposes. A maximum of 999 subject terms and 9999 descriptors is all that can ever be used. This choice of limitation was intentional to avoid an excessive proliferation of independent coding terminology.

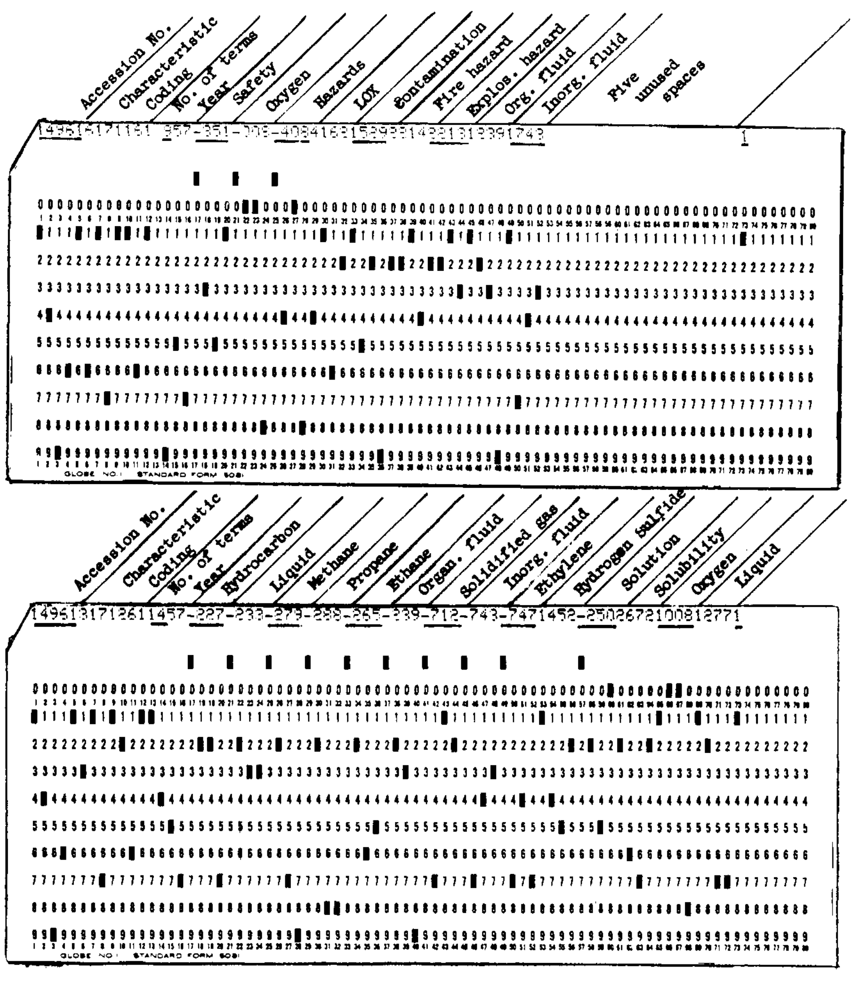

Figure 4. Coding punch-cards for input to the bibliography search tape

Separate vocabularies of terms (both subject terms and descriptors) have been developed for each of the nine categories that we cover in the field of cryogenics. Having separate vocabularies is an aid in maintaining consistent subject coding. Many of the coding terms, however, are used in two or more categories but retain the same numerical designation, no matter which category they are used in. Likewise, many articles being coded are pertinent to two or more categories and are independently coded for each applicable category. Incidentally, this independent coding of articles falling into a number of categories or related subject areas is the equivalent of entering the article into the search system that many separate times, thus providing linkage for each set of coding terms. The maximum number of subject terms presently assigned in any category vocabulary is 162 terms and the maximum number of descriptors is 1122. Many of the category vocabularies have a much fewer number of terms. The total number of subject terms now being used in all categories is 385 and the number of descriptors is 3086.

CHARACTERISTIC CODING

An important adjunct to the specific coding of an article is the "characteristic coding" designating the subject category, language of the article, temperature range, form of data, type of article, availability, and publication status. Figure 5 is the characteristic coding sheet used for such designations. The characteristic coding is not only useful in narrowing a search such as eliminating unpublished literature or articles in a particular foreign language but also is helpful to the user of a search to narrow down the number of references he might wish to review when these have not been so eliminated in the automated search.

DICTIONARY LOOK-UP PROGRAM

By preparing a search program based on 4-digit numbers, it was necessary to assign the appropriate 4-digit number for each subject term and descriptor assigned to an article after it was coded. For example in Figure 4 for the subject terms *Safety, Oxygen,* and *Hazards* the respective numerical terms are: −351, −008, and −408. Similarly 4162, 1529, and 2214 have been assigned for descriptors: *LOX, Contamination,* and *Fire hazard*, respectively. For a few years, this was done manually by a clerk. About five years ago, we asked one of our Computer Laboratory programmers to devise an automated dictionary look-up program to assign the 4-digit numbers. Figure 6 is an outline of this program. By use of a simplified word-matching technique the code terms assigned to an accession are found in the dictionary tape, and are assigned the appropriate 4-digit number for transmittal to the search tape. If a code term is not matched in the dictionary it will be flagged for review. If it is a new term, consideration will be given to adding it, perhaps as a synonym of an existing term with its designated 4-digit number or as an entirely new term. Simple typographical errors are corrected and reentered. Frequent misspellings or term variations such as singular or plural are simply added to the dictionary as synonyms so that they will be recognized in the future.

Categories

A-1: Books, Reviews, Surveys, Bibliographies, Proceedings, etc.
A-2: Properties of Solids
A-3: Properties of Fluids
A-4: Solid State, Theoretical, Phenomena, Basic Physics, etc.
A-5: Cryogenic Techniques, Tricks, Unique Methods, Unusual Procedures, etc.
A-6: Cryogenic Processes, Heat Transfer, Purification, Fluid Flow, Liquefaction, Safety Procedures, etc.
A-7: Laboratory Equipment and Instrumentation
A-8: Cryogenic Equipment
A-9: General Interest Literature, News, Management, Programs, Accidents, Miscellaneous

Language

B-1: English, B-2 French, B-3 German, B-4 Dutch, B-5 Italian, B-6 Japanese, B-7 Russian, B-8 Spanish, B-9 Other

Cryogenic Interest

C-1: Cryogenic Temperature Range (0 to 130° K. where not specifically designated in C-4 through C-7 below)
C-2: Cryogenic Interest but not in Cryogenic Temperature Range (except where designated C-8)
C-3: Not of Direct Cryogenic Interest
C-4: Below 1° K.
C-5: 1 to 10° K.
C-6: 10 to 50° K.
C-7: 50 to 130° K.
C-8: 130 to 300° K.

Form of Data (Omitted where not pertinent)

D-1: Numerical Data Included
D-2: No Data
D-3: Graphical Data Only

Type of Article (Omitted where not pertinent)

E-1: Experimental, Experimental and Theoretical, Original Work
E-2: Review Article, Compilation, Correlation, Discussion
E-3: Theoretical Only, No Specific Data Given

Availability of Document (suggested source)

F-1: Cryogenic Engineering Laboratory
F-2: National Bureau of Standards
F-3: Office of Technical Services (OTS)
F-4: U. S. Government Printing Office
F-5: Armed Forces Technical Information Agency (ASTIA)
F-6: Technical Libraries Generally (Published Literature)
F-7: Technical Libraries—Special (Foreign Literature—Special Periodicals)
F-8: Company Bulletins and Reports (Universities, Research Labs., etc.)
F-9: Other (Patents, Theses, Translations, etc.)

Form of Document

G-1: Published—Open Literature, Journals, etc.
G-2: Books, Proceedings
G-3: Company Periodicals (includes University, Foreign Gov't., State Institutions, etc.)
G-4: Government Periodicals (U. S.)
G-5: Company Reports, Private, Public, Gov't. Contract (includes Foreign Gov't. Repts.)
G-6: Government Reports (U. S.)
G-7: University Theses, Doctoral Dissertations, Master's Theses
G-8: Patents (U. S. and Foreign)
G-9: Other (Unpublished, Informal, Preprints, Letters, Notes, Term Papers, Talks, etc.)

Figure 5. Characteristic coding descriptions for cryogenic literature

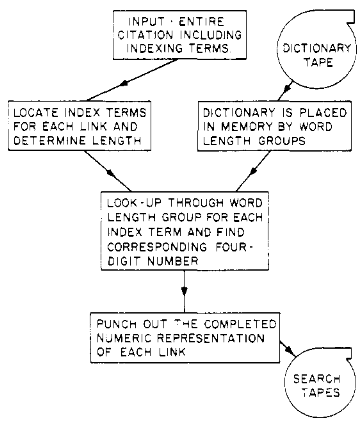

Figure 6. Dictionary look-up program

LITERATURE AWARENESS AND CURRENT AWARENESS SERVICE

It is estimated that the number of articles and published reports in the field of science and technology is approaching 2 million per year. We have found it necessary to scan more than 290,000 titles of journal articles and another 90,000 titles of reports to maintain an awareness of most of the new literature of interest in the field of cryogenics. This is accomplished by reviewing some 200 periodicals cover to cover, scanning through three title announcement bulletins, and 18 abstract bulletins. From these, we note about 15,000 items per year of possible interest. A weekly list of 100 to 200 articles in the area of low temperature physics and chemistry and cryogenic engineering as well as miscellaneous items of interest is prepared and mailed to nearly 1000 subscribers in the United States and abroad. About 8000 articles a year are selected as being sufficiently relevant and important to procure and process into the retrieval system.

INFORMATION ANALYSIS AND DATA EVALUATION

No matter how comprehensive and otherwise efficient a bibliographic service is, it is not what most users want. Their need is for the information contained in the literature and not necessarily the literature itself. As Dr. Alvin Weinberg, Director of the Oak Ridge National Laboratory said in his rather famous report "Science, Government and Information" published by the President's Science Advisory Committee in 1963,[7]

> Retrieval of documents is not the same as retrieval of information: A technical specialist really needs the information contained in the published literature, not the published literature itself. To retrieve information, as contrasted to documents, the technical community has devised the specialized data and information center.
>
> A specialized information center makes it its business to know everything that is being published in a special field—such as nuclear spectroscopy or the thermophysical properties of chemical compounds: it collates and reviews the data, and provides its subscribers with regularly issued compilations, critical reviews, specialized bibliographies, and other such tools.
>
> Specialized information centers, to be fully effective, must be operated in closest possible contact with working scientists and engineers in the field. The activities of the most successful centers are an intrinsic part of science and technology. The centers not only disseminate and retrieve information; they create new information. Making a discriminating selection of data, as was done in preparing the *International Critical Tables*, requires scientific insight of high order, and it itself is an essential scientific activity.

An outgrowth of this study that Dr. Weinberg reported was the establishment of the National Standard Reference Data System (NSRDS) operated at the National Bureau of Standards by Dr. Edward Brady.[8] It seeks to coordinate the activities of information analysis centers and even in some instances provides financial support. These information centers, including our own, work in very specialized fields in close relationship with specialists in these fields. The purpose is to analyze all of the published literature in a subject area, critically evaluate the pertinent data and information, and compile it in a conveniently useful form that can easily be retrieved and disseminated.

Our Cryogenic Data Center is a charter member of NSRDS. Some 60% of the available funds are spent on information analysis and data evaluation. For the past several years, tasks have been pursued for compiling low temperature data on the thermodynamic and transport properties of cryogenic fluids. Some work is done on properties of solids, and long range plans are to increase the effort in properties of solids compilations. The data compilation activities are funded by the National Aeronautics and Space Adminstration, and as a result the most active tasks are those yielding data of importance to the aerospace programs.

CURRENT STATUS AND SERVICES

More than 49,000 articles have now been processed into the storage and retrieval system. Nearly one half of these are in the properties of materials categories, giving us one of the best collections of literature in the world on properties of materials at low temperature. Approximately 100 requests a year are being received for bibliographic searches and many more requests are made for specific data. Figure 7 indicates the approximate distribution among industry, universities, government agencies, foreign countries, and our own staff for services. Some 600 publications and technical reports have evolved from work of the Cryogenics Division staff since establishment of the Laboratory in 1952. These are handled by the Cryogenic Data Center (except for direct public sales

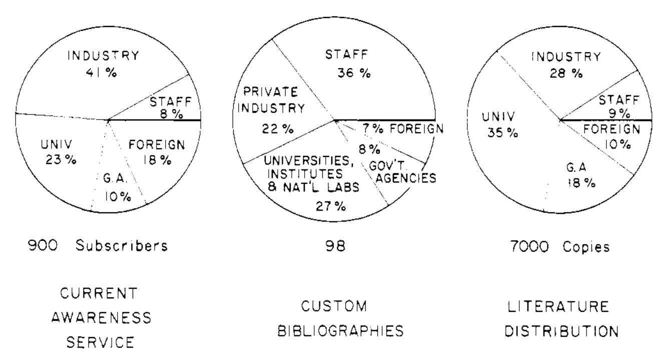

Figure 7. Distribution of customer services

which are now handled by the NBS Clearinghouse for Federal Scientific and Technical Information). Announcements and abstract cards are sent regularly to about 4300 people and institutions on our mailing list. Over 2000 requests for this literature are received annually, and last year about 7000 copies of material were distributed on a complimentary basis. Statistics for the Clearinghouse sales in 1967 are not available, but in prior years when our Data Center handled direct sales, they averaged 20,000 to 25,000 copies.

CONCLUSION

The development and operation of the Cryogenic Data Center has been an extremely interesting experience. Having started 10 years ago, it was among the pioneers of information analysis centers. As a consequence, many operating procedures had to be developed rather than adopted. Since the objective was not simply to develop a system but rather to serve the information needs in the field of cryogenics, only the most basic and necessary procedures were used, choosing the most conventional methods or, where there was no existing convention, to choose the procedure most likely to become conventional. An early decision to process material in machine readable form, I feel, was a wise one and my only regret is that there were essentially no standardized procedures to follow. We are very interested in the growing trend for a network system of technical information and data centers as is actively being sponsored by COSATI, the Committee on Scientific and Technical Information in the Federal Council for Science and Technology. We are looking forward to the establishment of standards such as for catalog format, machine readable characters, thesauri, CODEN, corporate authors, and the like. This can lead to more centralized or source processing and save information and data centers thousands of dollars annually. My advice to anyone engaged in the establishment of an information and data center would be to follow COSATI standards as far as practicable, experiment as little as possible, and plan to join the network of centers as soon as there is such a network. The very large document centers such as NASA, DDC, AEC, and the Clearinghouse have done much in establishing many uniform processing procedures that are a great help. I'm sure that in the development of a national (or international) network of information exchange, these procedures are the ones that are likely to be adopted. I feel that our own Cryogenic Data Center procedures are reasonably in tune with them and that it will be quite easy for us to join the "network."

LITERATURE CITED

(1) "NBS-AEC Cryogenic Engineering Laboratory," *Natl. Bur. Std. (U. S.) Tech. News Bull.* **37**, 152-8 (1953).

(2) Birmingham, B. W., "Cryogenics at the National Bureau of Standards, Boulder Laboratories," *Cryogenics* **5**, 121-8 (1965).

(3) U. S. Air Force, WADD Technical Report 60-56, A Compendium of the Properties of Materials at Low Temperature (Phase I), V. J. Johnson, General Ed. (July 1960), Contract AF 33(616)-58-4.

(4) Wildhack, W. A., and J. Stern, "The Peek-a-boo System," Chapt. 6, "Punched Cards," p. 125, Reinhold, New York, 1959.

(5) Perry, J. W., and Allen Kent, "Tools for Machine Literature Searching," Interscience, New York, 1958.

(6) Standard for Descriptive Cataloging of Government Scientific and Technical Reports; Committee on Scientific and Technical Information, Federal Council for Science and Technology, Washington, D. C., Revision No. 1 (October 1966); AD-641 092 or PB-173314.

(7) "Science, Government, and Information," A Report of The President's Science Advisory Committee, The White House (January 10, 1963).

(8) Brady, E. L., and M. B. Wallenstein, National Standard Reference Data System—Plan of Operation, NSRDS-NBS-1 (December 30, 1964).