NBS TECHNICAL NOTE

290

Information Handling in the National Standard Reference Data System

FRANZ L. ALT

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards

National Standard Reference Data System

The National Standard Reference Data System is a government-wide effort to give to the technical community of the United States optimum access to the quantitative data of physical science, critically evaluated and compiled for convenience. This program was established in 1963 by the President's Office of Science and Technology, acting upon the recommendation of the Federal Council for Science and Technology. The National Bureau of Standards has been assigned responsibility for administering the effort. The general objective of the System is to coordinate and integrate existing data evaluation and compilation activities into a systematic, comprehensive program, supplementing and expanding technical coverage when necessary, establishing and maintaining standards for the output of the participating groups, and providing mechanisms for the dissemination of the output as required.

The NSRDS is conducted as a decentralized operation of nation-wide scope with central coordination by NBS. It comprises a complex of data centers and other activities, carried on in government agencies, academic institutions, and nongovernmental laboratories. The independent operational status of existing critical data projects is maintained and encouraged. Data centers that are components of the NSRDS produce compilations of critically evaluated data, critical reviews of the state of quantitative knowledge in specialized areas, and computations of useful functions derived from standard reference data.

For operational purposes, NSRDS compilation activities are organized into seven categories as listed below. The data publications of the NSRDS, which may consist of monographs, loose-leaf sheets, computer tapes, or any other useful product, will be classified as belonging to one or another of these categories. An additional "General" category of NSRDS publications will include reports on detailed classification schemes, lists of compilations considered to be Standard Reference Data, status reports, and similar material. Thus, NSRDS publications will appear in the following eight categories:

Category	*Title*
1	General
2	Nuclear Properties
3	Atomic and Molecular Properties
4	Solid State Properties
5	Thermodynamic and Transport Properties
6	Chemical Kinetics
7	Colloid and Surface Properties
8	Mechanical Properties of Materials

UNITED STATES DEPARTMENT OF COMMERCE • John T. Connor, Secretary
NATIONAL BUREAU OF STANDARDS • A. V. Astin, Director

NBS TECHNICAL NOTE 290

ISSUED July 1, 1966

Information Handling in the National Standard Reference Data System

Franz L. Alt

Office of Standard Reference Data
Institute for Basic Standards
National Bureau of Standards
Washington, D.C., 20234

NBS Technical Notes are designed to supplement the Bureau's regular publications program. They provide a means for making available scientific data that are of transient or limited interest. Technical Notes may be listed or referred to in the open literature.

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402
Price 25 cents

Contents

Page

1. Introduction ---- 1
  - 1.1. Plan of approach ---- 1
    - 1.1.1. Development of NSRDS ---- 1
    - 1.1.2. Recommendations ---- 2
  - 1.2. Organization of NSRDS ---- 3
  - 1.3. Information services ---- 4
    - 1.3.1. General ---- 4
    - 1.3.2. Compilation-production services ---- 4
    - 1.3.3. Inquiry services ---- 4
2. Mechanized information system for NSRDS ---- 4
  - 2.1. General considerations ---- 4
    - 2.1.1. Philosophy of computer selection ---- 4
    - 2.1.2. Arguments in favor of mechanization ---- 5
    - 2.1.3. Obstacles to mechanization ---- 5
    - 2.1.4. Size and location of computer ---- 6
    - 2.1.5. Computer characteristics ---- 7
  - 2.2. Functions of the system ---- 7
    - 2.2.1. Information retrieval ---- 7
    - 2.2.2. File updating ---- 8
    - 2.2.3. Aids to publication ---- 9
    - 2.2.4. Other computer functions ---- 9
  - 2.3. Input to the file ---- 10
    - 2.3.1. Keypunching ---- 10
    - 2.3.2. Print reading ---- 10
    - 2.3.3. Data available from data centers ---- 10
    - 2.3.4. Equipment and format ---- 11

P

2.—Continued
  - 2.4. Information storage ----
    - 2.4.1. Form of data ----
    - 2.4.2. Arrangement of data ----
    - 2.4.3. Compact storage ----
    - 2.4.4. Storage of instructions ----
    - 2.4.5. Storage devices ----
  - 2.5. Access to the computer ----
    - 2.5.1. Man-machine interaction ----
    - 2.5.2. Long-distance access ----
    - 2.5.3. Administrative arrangements ----
3. Proposed interim system ----
  - 3.1. Characteristics of the operation ----
  - 3.2. Inquiry services ----
    - 3.2.1. Getting started ----
    - 3.2.2. Referral ----
    - 3.2.3. Reference ----
    - 3.2.4. Documentation ----
    - 3.2.5. Data service ----
  - 3.3. Publication services ----
    - 3.3.1. Types of services ----
    - 3.3.2. Preparing for mechanization ----
  - 3.4. Preparatory activities ----
    - 3.4.1. Information gathering ----
    - 3.4.2. Bibliography ----
    - 3.4.3. Classification ----
    - 3.4.4. Indexing and abstracting ----
4. Conclusions ----

Information Handling in the National Standard Reference Data System

Franz L. Alt

A preliminary plan is presented for the selection, acquisition, intellectual organization, and storage of the information which will underlie the Information Services Operation of the National Standard Reference Data System, as well as for methods of locating desired information items in storage, retrieving, and displaying or communicating them. Questions of the use of computers for these purposes are discussed, including selection of equipment, arrangement of digital storage, input format, remote access, and the economics of choosing certain functions of the system for mechanization. Also, an interim system, based on conventional and, in the main, manually operated files, is described.

Key Words: Computer-aided inquiry service, data retrieval, file mechanization, information retrieval, standard reference data.

1. Introduction

1.1. Plan of Approach

1.1.1. Development of NSRDS

The present report describes proposed plans for e handling of information in the National andard Reference Data System (NSRDS). It concerned with the information with which the stem deals, with the logical organization of this formation, its acquisition and physical storage; th methods of locating desired information items storage, retrieving and displaying or comunicating them.

It is expected that the information system of SRDS will undergo evolutionary changes for a umber of years. The later stages of this develment can not yet be foreseen with complete arity, because NSRDS itself is developing. etails of the system design will depend on certain rameters (e.g., size and rate of growth of the ta collection) whose eventual values are not yet own, or on changes in technology which are kely to develop during the next few years, and the solution of some of the research problems hich will be undertaken by NSRDS itself.

Despite all this uncertainty about the future velopment of the information system, two facts nerge which have been adopted as basic policy cisions and which underlie all other decisions be made: (1) Ultimately the NSRDS informaon system will involve the use of large electronic gital computers in a crucial way (though not ecessarily to the complete exclusion of manual ethods). (2) At present the introduction of ch computers into the principal operations of SRDS would be premature.

These two statements are plausible in themlves and will become more so in the course of scussing the economics of computer operation section 2 of this report. From them we infer mediately that we should distinguish at least vo periods in the operation of NSRDS: the longnge future, during which the operation will be naracterized by computer use; and the near future, which will have a different regime—as we shall argue, conventional and in the main manually operated files, combined with information stored in human minds.

This leaves several questions still to be answered. First, it may appear plausible that the transition to the computer system should be gradual or in a number of steps, rather than all at once. Second, one may ask whether there should be one or more intermediate periods, characterized by methods which differ both from the ultimate large computer system and the initial manual file system. Such intermediate methods would differ from mere transitional steps on the way to full computer operation, by definition, in that the former call for some investment in hardware or procedures which would be retained more or less without change. A more detailed discussion of these questions given later in the present report will favor a gradual transition from manual to computer operation, but without recourse to costly intermediate techniques that would be discarded before economic use is made of the investment.

A third open question is that of time. A precise prediction of how soon a computer can be used effectively and economically is not possible at present, but it seems likely that the initial manual information system will remain in operation for at least three years, possibly longer. The speed of phasing from a preponderantly manual to a preponderantly machine system will depend upon the activities and development of the Technical Data Centers and other as yet unforseeable factors. At each stage during this gradual transition there should be ample opportunity for us to make decisions more confidently as we observe the system in operation.

Not only will the transition to machine methods be gradual; in some instances it should not take place at all. For example, it is not clear that inquiries can be answered reliably *in toto* by a machine. We may find that machine methods can be used to narrow a search to a few choices, the final selection to be done by humans, or vice versa, humans to perform a preliminary screening and switching of inquiries.

The successive systems are not quite independent of each other. It is easy to see that a choice between alternative procedures in the computer-based system might be influenced by the way in which the same feature has been handled in the preceding (manual or intermediate) system. Similarly, and more importantly, the design of the initial and any intermediate system should preferably avoid anything that might impede transition to the most desirable form of the ultimate computer system. In order to facilitate the exposition, we therefore propose to discuss the long-range information system first, and the short-range one afterwards.

Thus, the present report is organized as follows. The next section summarizes the main results. The remaining portions of section 1 describe the organization and functions of NSRDS to the extent needed for our discussion of the information system. A more complete description has been given elsewhere,[1] and readers familiar with it may bypass these sections. Next, section 2 develops the ultimate computer-based system as far as our present ideas go. This is followed in section 3 by a description of the system envisaged for the next few years, and finally by a few comments on the transition between the two.

We have referred to the long-range system as the "ultimate" one. By this we do not mean that it will be frozen forever; rather, that we have taken into account everything that we can foresee about it at this time. The system will undoubtedly undergo further changes, but they must be disregarded in our present planning.

We envisage NSRDS not as an entity all by itself but as one of a number of information activities constituting the emerging National Scientific and Technical Information System currently under study by COSATI. In particular, we endeavor to keep NSRDS compatible with the information systems of AEC, NASA, DDC, the Clearinghouse for Federal Scientific and Technical Information, and other agencies.

1.1.2. Recommendations

In this section we summarize briefly the principal results of the study, especially as they lead to recommendations for action. It has already been mentioned, and will become more evident in later sections, that these results are still somewhat tentative. This is unavoidable, in view of the uncertainty of many of the premises on which they are based. The best that can be done at this time is to give a full presentation of the pros and cons for each of the major decisions. In order to enable the reader to find this information selectively, if he so desires, the following list of recommend tions is cross-referenced to those sections of th Note in which he may find a discussion of unde lying assumptions, facts, and arguments and where applicable—of alternatives which were co sidered. It will become clear that many of the recommendations, especially those intended for in plementation several years hence, are subject change in the interim, if such change should b come advisable, e.g., through new informati about the availability and rate of generation data, about the operation of data centers, abo performance and cost of computers, and the ope ating experience of the Office of Standard Refe ence Data itself.

Summary of Recommendations

1. A conventional manual data file system f the near future. (1.1, 3)
2. A system based on a digital computer for t more distant future. (1.1, 2)
3. Begin rendering services using System (1) once. (1.1, 3.2.1, 4)
4. Plan on the bulk of System (2) being imp mented in 3 to 5 years. (1.1)
5. Transition from (1) to (2) in steps, but wi out major investment in any temporary interme ate system. (1.1, 4)
6. Aim at man-machine cooperation rather th at complete mechanization of information trieval. (2.1.3, 2.5.1)
7. Share time on a large general-purpose co puter operated by NBS. (2.1.4, 2.5.3)
8. Obtain an external storage unit of about 1 million words capacity, with a transfer rate of least 1 to 2 megabits per second, for exclusive by OSRD. (2.1.5, 2.4.5)
9. Establish in the office of OSRD a console direct on-line access to the computer. (2.1 2.5.1)
10. At a later stage, similar consoles should available throughout the country, for connecti to the computer by long-distance telepho (2.1.5, 2.5.2)
11. Use of the computer to include informat retrieval, file updating, aid to publication (edit and typesetting), and housekeeping. (2.2)
12. OSRD to devise standard formats for k punching of data, e.g., into 80-column punc cards, these formats to be observed as far as p sible by OSRD and all Data Centers. (2.3.4)
13. The burden of keypunching by OSRD to relieved by using machinable data punched other sources, or for other purposes, and data g erated by computers or automatic print read (2.3.3)
14. Data in the master file to be arranged properties or, more generally, by homogene groups of related properties. (2.4.2)
15. Functions of one or more variables to be r resented, where appropriate, by approximat

[1] E. L. Brady and M. B. Wallenstein, *National Standard Reference Data System—Plan of Operation*, NSRDS-NBS 1, U.S. Government Printing Office, 1964.

olynomials (or other series) using maximal ıtervals. (2.4.3)

16. Use of the computer to be in "batch" mode henever possible; remote on-line access to computer from OSRD when necessary for efficient ıan-machine interaction. (2.5.1)

17. Future results of critical evaluation of data ɔ be assembled in a separate file in OSRD. When o standard reference data available, data inuiries to be answered with best data in the literaıre, with suitable disclaimer. (3.2.1, 3.2.5)

18. Initially, a large fraction of all inquiries eceived will be referred to experts; number of eferrals should gradually drop but not to zero. 3.2.2)

19. Technical area managers, Data Centers, IBS scientists and occasionally others to be used s experts for replying to inquiries. (3.2.2)

20. Use of citation indexes, bibliographic couling, and other aids to literature referencing to e explored. (3.2.3)

21. Graphical information to be handled by omputer-controlled curve plotter, or if this is ot possible, by microfiche. Control of the latter ystem by central digital computer to be studied. 3.2.4)

22. Publication of NSRD Series to be coninued; a periodical, especially with bibliographic ıformation, to be considered later. (3.3.1)

23. Computer aids to publication to be coninued, and further development especially of diting codes to be pushed. (3.3.2)

24. Needs of potential users as to frequency, ypes, and form of information to be established hrough user surveys. (3.4.1)

25. Concurrently with the start of information ervices, OSRD to broaden its information by a ibliographic survey of existing data compilaions, and by a questionnaire to prospective users. 3.4.2)

26. Establishing a small thesaurus of subject ndex terms, indexing the present OSRD library ollection, and abstracting of library books to be oursued in this order, and concurrently with inuiry-answering service. Library personnel to be ısed in inquiry answering in order to gain experince. (3.4.3, 3.4.4)

27. Mechanization of a small part of the colection to be attempted soon. (4)

1.2. Organization of NSRDS

Physical properties of materials, which are the ubject matter with which NSRDS deals, have een divided into a number of *technical areas*. To late seven such areas have been defined: (1) nuclear data, (2) atomic and molecular data, (3) solid state data, (4) thermodynamic and transport roperties, (5) chemical kinetics, (6) colloid and urface properties, and (7) mechanical properties. Other areas may be added later, but these seven eem to come close to exhausting our present concern.

The organization dealing with these subjects consists of the *Office of Standard Reference Data* (*OSRD*), located at the National Bureau of Standards, and a number of *Technical Data Centers*, mostly outside of NBS. Many of these Data Centers are sponsored and operated by other agencies; some antedate the existence of NSRDS. A few are located at NBS, and some operate elsewhere under contract with NBS. Each Data Center has cognizance over a certain domain, usually falling within, but narrower than, one of the seven technical areas. The domain of a technical Data Center may be characterized by a set of physical properties (e.g., infrared spectra) or materials (e.g., metals) or occasionally of other criteria (e.g., low temperature), or by a combination of such criteria. The designation of certain organizations as Technical Data Centers of NSRDS, the delimitation of their scope, and coordination of their activities are among the responsibilities of OSRD. In addition there are data compilation projects directly under the cognizance of OSRD.

It is recognized that the presently existing Data Centers cover only a small part of the entire domain of standard reference data. It is desirable that new centers should be established, or old ones expanded, at a rapid rate. It must be recognized, however, that even in the best of circumstances it will take years before Data Centers will even approach complete coverage of the entire field of standard reference data, and it is unlikely that such completeness will ever be attained. As a result, a larger burden will have to be placed on OSRD, at least initially.

In particular, OSRD will have to engage in a survey of existing data compilations in all fields, which can be used as a basis for information services until a better foundation is furnished by the Data Centers. It will also, in some cases, contract with organizations or individual scientists for producing compilations, critical evaluation and reviews, computation of certain useful functions derived from standard reference data, and even experimental measurements. All these activities can be provided by OSRD as opportunities offer themselves, but a more systematic and exhaustive coverage of the field will depend on the expansion of the Data Center system.

Within OSRD there is an Area Manager for each of the seven technical areas, plus one for information system design and research; in addition, OSRD operates an Information Service at NBS. In a certain sense, the organization and activities of this *Information Services Operation* (*ISO*) are the main subject of the present report, although some relevant questions about Data Center activities will also have to be discussed.

ISO is expected to consist of four units, concerned with (1) compilation-production services, (2) inquiry services, (3) the data file operation, and (4) analysis and user relations.

1.3. Information Services

1.3.1. General

It is the responsibility of the Information Services Operation "to provide the services to the technical community that are determined to be useful and maintain the collection of data that will constitute the data center at the National Bureau of Standards—indexing, filing, storing, and retrieving data as required."

We distinguish two kinds of services: scheduled and nonscheduled. It is too early to make a firm estimate of the relative size of these two kinds of information activities. Indications are that they will be of comparable magnitude.

1.3.2. Compilation-Production Services

Scheduled services include the dissemination of information, either periodical or occasional, on our own initiative. It is expected that some of this will be handled, as it is now in a few cases, by the technical data centers, with or without assistance from OSRD. The central information service operation will not duplicate any of these efforts but will attempt to provide additional publications covering the field of reference data in general, or cutting across the lines of several data centers, or falling between them. A periodical current awareness service is one of the likely activities contemplated for ISO. Preparation of revised editions of data handbooks is an example of activities of the technical data centers.

In general, the publication of monographs, as the primary means of supplying data to users, is perhaps the most important single function of NSRDS. Such monographs, while concentrating on the tabulation of critically evaluated data, will in addition contain such relevant information on the generation and application of the data as is likely to be helpful to the user (cf. sec. 3.3.1).

1.3.3. Inquiry Services

We may visualize four kinds of action taken in response to requests for information: (a) referral; (b) reference; (c) documentation; (d) data information. They are increasingly specific in the order in which they are listed. Referral means that the question is referred to another organization, generally one of the technical data centers though occasionally an organization outside NSRDS. It is expected that the requestor would be informed of the referral. The reply may be sent to the requestor, preferably via OSRD. By reference is meant a listing of relevant literature. Documentation goes one step further and include furnishing of micro-stored or hard copies of the referenced literature. "Data information" impli furnishing not only a listing of the requested dat but also any necessary explanation, caution, et (cf. sec. 3.2).

The choice among these actions must be kept flexible at all times. It would be undesirable to decide that OSRD will furnish only one kind of reply.

2. Mechanized Information System for NSRDS

2.1. General Considerations

2.1.1. Philosophy of Computer Selection

In this section we discuss the advantages and drawbacks of mechanization primarily in the light of their concrete effects, rather than of their imponderable consequences.

The use of digital computers for information retrieval is one of the most widely discussed issues in science administration today. A number of entire new organizations are being set up for this purpose, and large vested interests are at play. In such a situation one is easily tempted into extraneous considerations: having a computer is considered "good advertising," it lends an appearance of progress and importance to an organization, it attracts prospective customers. We propose to resist these temptations and to examine the possible uses of computers in NSRDS strictly on their merits. Accordingly, we will compare the cost of computer versus manual operation and the cost differences among different computer types; we will examine the advantages in speed and ease of distant access to a mechanized file, and the possible loss in quality and convenience of direct access to a manual file kept on the spot. We must keep in mind, however, that an analysis based on these fa tors alone is likely to understate the value of aut mation. Experience in other fields has shown th the introduction of computer methods is often fo lowed by unforeseeable, or at least unforesee rapid progress in other respects. Noteworthy e amples are in the analysis of x-ray crystal graphic data and of bubble chamber observatio where the use of computers was followed by aut mation of data acquisition and has led to a man fold expansion in the amount of scientific info mation obtained by these methods.

At present there is not enough informati available to enable us to make a quantitat study of the cost and performance of computers be used several years from now. We have to r on some general and qualitative observations a experience in other fields, but there will be ti to verify the findings so obtained before comm ting ourselves definitely to one or another cou of action. The arguments to be presented in e next two sections give qualitative support to e contention that a digital computer will ultimat y be economical for NSRDS. In addition we sh ll argue that sharing time on a large compute is preferable to operating a smaller computer r

NSRDS alone. Finally, rather than insisting on complete mechanization of every phase of the process, we shall aim at an optimal division of labor between man and machine.

2.1.2. Arguments in Favor of Mechanization

(a) Size of collection. There is as yet no reliable basis for estimating the magnitude of our information collection. For planning purposes we are using an order of magnitude of 100 million words. (See sec. 2.2.1.) It seems plausible to us that the ultimate size may differ from this estimate by possibly a factor of 10, but probably not by a factor of 100, in either direction. It should be several years before we get to figures of this magnitude. Size alone is rarely a sufficient reason for automation, unless it gets to be truly excessive; if a collection of 100 million words were to be used only in the way in which one uses, say, a telephone directory or library card file, mechanization would not be worthwhile. Taken together with the following arguments, however, the estimated size is an added consideration in favor of mechanization.

(b) Mechanization becomes advantageous when there are frequent occasions for searching through the entire file or large portions of it. An example is the search for materials whose boiling points lie between given limits. Whether or not searching is required depends on the way in which the file is organized; for example, the telephone directory is not properly organized for finding people living on a given street. Indexing is a method of organizing a file so as to allow the answering of certain types of questions without a major file search. It seems extremely unlikely that we should be able to anticipate all our information needs by measures of this kind, and therefore file searches will be needed from time to time. It is impossible at present to estimate their frequency, but the argument lends weight to the demand for mechanization.

(c) Updating a file, correction of errors, addition of new results, etc., are greately facilitated by mechanization. This consideration has prompted us, for example, to use cards and machine methods of type composition in preparing the next edition of "Crystal Data Tables." It is likely that the extra cost incurred will be more than offset by savings in preparing the first subsequent edition.

(d) Methods of bibliographic coupling and of citation referencing are greatly aided by mechanization. Arguments will be advanced in section 2.2.4 below to show that these methods are of particular importance for the Standard Reference Data Program.

(e) The operation both of OSRD and of the data centers will be facilitated by remote access to the file, which in turn presupposes mechanization. More will be said about this below. (See secs. 2.5.1, 2.5.2.)

(f) In some of the data centers, the introduction of machine methods will be aided by the fact that some of the experimental measurements used in the generation of data are set up for automatic recording and digital encoding of results.

2.1.3. Obstacles to Mechanization

The introduction of computers into the process of information retrieval is hindered by the same difficulty which characterizes most other computer applications: our present inability to give a rigorous description of the procedure which the computer is to follow. In human information retrieval, e.g., in a library, not only is the memory of the librarian an important tool, but every user of a library uses clues, lines of reasoning, and other mental processes of which he himself is not aware. Such an imperfectly formulated procedure is well adapted to the human mind but is of no help with computers.

There are two ways to overcome our lack of understanding of the problem. One is a program of research into the formulation of the information retrieval problem. For this, in turn, there are two alternatives: investigate and, if possible, formalize the customary human procedure; or develop new methods which are more suitable for machines. It is, of course, not true that the computer would have to use the same procedure as is used by humans; but it has to use some procedure which is completely formulated, and one way to formulate it is to start with the familiar human procedure: try to make explicit the mental processes which we use without being aware of them, see whether they contain any elements which can be formalized, and if so, translate them into computer programs. If this is possible, it may be preferable to inventing an entirely new and untried approach to the problem.

The other way out is partial mechanization: limit the use of the computer to those fragments of the whole problem for which a rigorous formulation can readily be found, and leave the rest to humans as before. In this case one has to pay special attention to the interaction between man and machine, to the smooth transfer of information from one to the other. This brings up another problem possibly as formidable as the first: often input and/or output are the principal bottlenecks in automatic computation. Indeed, there are classical cases of systems in which the introduction of computers was not profitable until all functions of the system had been mechanized, thus reducing the relative magnitude of input and output. There are other examples in which, on the contrary, the use of computers was made uneconomical by the attempt at complete mechanization, resulting in excessively complicated computer programs; a more modest approach, limited to the most frequently required functions of the system, gave most of the benefits for a small fraction of the cost.

Thus, a reasoned choice between the two possible approaches must be made. The decision does not necessarily have to go all one way; compromises are possible. For the information system which we are considering here, it seems likely that some functions should be reserved to humans for a long time to come, possibly forever. We are thinking especially of the checking, screening, and editing of output, and of the formulation of questions by successive steps. The latter problem may be viewed as that of assigning a sufficient number of pockets for information to assure that no single pocket contains more than a few items. Retrieved information would then be in the form of a pocket of information or a small number of such pockets. Resolution beyond this point by machine methods alone may not be economical.

On the other hand, the problems in man-machine communication which these functions generate, although they are severe, do not seem insurmountable. Indeed, we envision that the human user will interfere repeatedly in the computer process, aided by conveniently designed facilities for inserting instructions and for display of intermediate computed results. Such collaboration or *dialogue* between user and machine is expected to be a characteristic feature of our information system.

2.1.4. Size and Location of Computer

Anyone in need of the services of a computer should first examine whether to acquire a computer exclusively for his own use or to obtain time on somebody else's computer. (We may disregard the third alternative, of acquiring a computer and making some of its time available to outsiders.) For prospective users in the Federal government such an examination is specifically prescribed by the Bureau of the Budget.

Of the various uses of the computer in the work of NSRDS, probably the most demanding requirement is the retrieval of specific items of information, or sets of such items, in response to requests. It will be seen that other uses—e.g., in updating the information file, assistance in publication, housekeeping—are less exacting and fit well into a system designed specially for the information retrieval functions.

In a typical problem in this area, a portion of the computer-stored information file is read and each item in the file compared with the question being asked, to see whether the file item is relevant to the question. On a large fast computer only a few microseconds are required for each comparison; for those few items which are found to be relevant, a longer process of evaluation and output takes place. Time is saved if a number of questions are treated simultaneously. One may, for example, collect the questions arriving in the course of a day into one batch and answer them in a single computer run. Each question needs to draw on only a certain portion of the information file, and for each portion of the file there will be, on the average, a small number of questions to be considered.

There are so many details as yet unknown that we can obtain at best an order-of-magnitude estimate of computer time involved. Computers of the incoming generation take about 1 microsecond per (logical) instruction. If each word from the file is compared with an average of 5 questions, and each comparison takes 4 instructions, the computer will spend 20 microseconds per word; with a file of, say, 100 million words the daily computer time would be 2000 seconds. We shall see that the rate of transferring information from store to central processor can just keep up with the speed of computation. If a smaller and slower computer were used, it could be kept busy full time. (On the Bureau's present computer, IBM 7094, the time would be something less than 2 hours per day.) For supporting figures see section 2.4.5.

In the choice between large and small computers, the large ones are normally less expensive; typically, their hourly cost might be greater by a factor of ten, their output by a factor of 100. Small computers can be justified only on grounds other than cost, e.g., convenience of access, or mismatch between internal speed of a large computer and rate of input or output. No serious argument of this kind appears to be valid in our case, as will be seen in section 2.4.5. On the other hand, there is a further argument in favor of a late-vintage large computer: OSRD is expected to lead the way in applying and demonstrating improved ways of retrieving standard reference data; use of front-line equipment offers many opportunities for exercising such leadership.

If OSRD is to use a large fast computer, it will, in the foreseeable future, do so by obtaining computer time from a laboratory operating such a computer, since its own needs would not keep such a computer fully occupied. This leads to the question whether computer programming services should likewise be obtained from another laboratory, or whether OSRD should build up its own programming staff. To this question we do no have a clear-cut answer at present. There is an increasing trend toward professional specializatio in programming, as a result of which a computation laboratory is in a better position to select and train competent programmers, keep them full employed in their special field, and offer them pro fessional advancement. In an organization lik OSRD, a professional programmer is intellectuall isolated, faced with a fluctuating workload of wha to him are "odd jobs." On the other hand, som aspects of computer programming demand inti mate familiarity with its data file organizatio and benefit from the devotion of a staff whose firs loyalty is to OSRD. Another approach is to hav computer programs written by teams consistin of people from both organizations—professiona programmers from the computation laborator teamed with data specialists with detailed pro gramming experience from OSRD; the "interface or communication link between these people migh consist of flow charts and data sheets.

In the absence of a strong argument to the con-ary, it seems natural that OSRD should use the ge general-purpose computer expected to be ilable at NBS. While remote access to com-ters over long distances is increasingly coming o use, it is economically limited to certain spe-l situations—very short problems, or frequently eded large problems which remain unchanged years, so that the program and any tabular data permanently stored at the computer site. In ny ways close proximity to a computer is ad-ntageous, since we envisage the OSRD operation gradually evolving, with frequent changes in nputer programs and additions of large quanti-s of data. Such operations are facilitated by rsonal contact of machine operators, program-rs and users, by hand-carrying of large card cks, and sometimes by the user's ability to influ-ce the policies of the computation laboratory— of which argues in favor of using the general-rpose NBS computer.

The same reasons can be advanced against the oposal that OSRD join with some or all of the hnical data centers of NSRDS for the establish-nt of a common computer laboratory. An addi-nal argument against such a plan is the political ficulty of reaching agreement with the different ta centers, many of whom have their own vested erests in computing laboratories located at their stallations.

There is little likelihood that OSRD would out-ow the sharing arrangement. Present estimates licate that a few years from now OSRD will use mall fraction of the time available on present-y computers. Even if OSRD's requirements ould eventually grow far beyond this estimate, is likely that computers will also have become ich faster by that time.

2.1.5. Computer Characteristics

If, as proposed in the preceding section, OSRD ares in the general-purpose computer of NBS, ovision will have to be made for certain pecu-rities of the OSRD operation.

The amount of data to be stored for OSRD is great that it is impractical to keep them on rds or tape and read them into the computer r each run anew. In this respect OSRD is, and obably will remain, unique among NBS com-ter users. It will be necessary to acquire a parate storage component, which would be pur-ased or rented by OSRD and reserved for its clusive use, and which would be connected to e computer main frame. Suitable storage vices are now commercially available from sev-l manufacturers. A secondary question which n be resolved later is whether this storage unit ould also contain the program instructions—an rangement which is probably economical but ssibly in conflict with the operating system of e computer (cf. secs. 2.4.4, 2.4.5).

It will be mandatory, or at least highly desir-able, to have facilities for remote on-line access to the central computer. A console should be located in the OSRD offices (and there will prob-ably be a demand for a number of similar consoles elsewhere in the Bureau) from which a user can contact the central computer, wait for the end of the current problem (or in the case of a long prob-lem, interrupt it), read into the computer a small amount of instructions and data, have the instruc-tions executed and immediately see a small volume of results. Large-volume output would remain on tape in the computer room and be available to the user there. The program should have access to routines and data permanently stored in the com-puter's internal memory, and should be able to connect the computer, under its own control, to tape stations and special external storage devices such as the one postulated in the preceding para-graph (cf. sec. 2.5.1).

It will also be desirable to have the ability to use similar remote stations in distant cities, using commercial telephone lines. This will enable in-dividual scientists and engineers to obtain needed data and related bibliographic information with a minimum of effort and time loss (cf. sec. 2.5.2).

In order to enable OSRD to be compatible with the Technical Data Centers and, in many cases, introduce recommendations for common practice for their information handling, the NBS computer should be able to accept programs written in the more widely used programming languages, espe-cially those standardized by the American Stand-ards Association. The role of NSRDS will be facilitated if the NBS computer is of a kind com-mercially available throughout the United States.

Finally, as we have already said, the computer should be in the front line of development of large, fast, and powerful computers, in order to enable OSRD to discharge its responsibility of establish-ment of standards of quality, methodology includ-ing machine processing formats and such other functions as are required to ensure the compat-ibility of all units of the NSRDS.

The development of special-purpose computers for information retrieval will have to be watched. It is not yet clear whether current efforts in this direction will be successful nor, if they are, to what extent their novel features will come to be included in future general-purpose computers.

2.2. Functions of the System

2.2.1. Information Retrieval

Under this heading we consider the reaction of the system to (unscheduled) requests for informa-tion. The nature of these requests can be inferred from the experience of existing specialized data centers. It appears that the frequency of such requests ranges from perhaps a few dozen to over a thousand per year, depending on the scope of the center. Since OSRD is broader than any of the centers examined, its work load should at least equal the higher of these figures; that is to say, we should expect to start with several requests per day as soon as the availability of OSRD has become known, and to grow far beyond that number as scientists and engineers learn to use the service. It has been the experience of other centers that a major part of these requests is not for data but for "administrative" information—availability of publications, addresses of organizations, etc.—and technical problems other than data themselves. A large portion of the questions come from the immediate neigborhood of the center—people in other parts of the same organization—which suggests that there is a need for such information in technical laboratories but that the present cumbersome methods of retrieval discourage most potential users. This observation reinforces our argument for remote access to the mechanized data collection (cf. sec. 2.1.5).

Among requests for data we distinguish mainly two kinds. Those of the first kind ask for a specified property or group of properties, of a specified material or group of materials, for specified conditions or ranges of conditions. For instance, one might ask for the optical density of water at 20 °C, at a given wavelength, or one might demand a table of the specific heats, entropies, and enthalpies of the noble gases between 0 and 100 °C. Questions of the second kind ask for materials for which certain properties have specified values, or lie within specified ranges; for instance, alcohols with molecular weights not over 102, whose boiling points at atmospheric pressure lie between 60 and 100 °C. In this second class of questions are also the problems of identifying materials from their spectra, from crystallographic or other properties.

These two types of questions are analogous to the direct and inverse use of a mathematical table—to find values of the tabulated function, or to find those arguments for which the function has given values. In the case of properties of materials, certain other kinds of questions are also possible (e.g., which of the spectral lines of mercury is narrowest?) but they are comparatively rare. Questions of the first kind will probably predominate, if the experience of existing data centers may be taken as a guide.

The ease with which a question of either type can be answered depends crucially on the size of the tables and on the way in which they are organized. The latter will be discussed in a subsequent section of this report. As to size, there is first of all an almost unlimited number of chemical compounds; those on which significant data are available may number 100,000, and this number is growing rapidly. There are perhaps 1000 different properties within the scope of NSRDS. Some of these are represented by single numbers, others are functions of one of more variables; each of the latter is represented by perhaps several hundred numbers (cf. sec. 2.4.3. below). The are few materials for which all this informati exists, especially if the less dependable measu ments are omitted. We estimate vaguely that t average number of reliably measured data each of the 100,000 materials is at present w below 1000, and will reach and pass that numb some years from now. A collection of all the data will then amount to 100 million words.[2]

We can obtain an independent check on tl figure, crude as it is, in the following way. T present rudimentary library of OSRD has abo 600 volumes, of which about 400 are data. At average of 400 pages per volume and 500 words p page, this is 80 million words. Most of the volumes are not entirely filled with data but co tain large sections of text; if tabular they oft include large blank spaces; and there is much ov lap in the contents of different volumes. T number of separate data items in this collecti might be between 10 and 20 million.

2.2.2. File Updating

The maintenance of files is a standard compu problem, common to numerous applications. has been the subject of a substantial techni literature. In many respects the requirements NSRDS are not different from those of other plications, and can be handled by standa methods. File updating—the insertion of new tries into the master file and the replacement any old ones which need correction—is norma done during the same computer runs which made for the purpose of information retriev If any new information for insertion in, or corr tion of, the file is received by the laboratory at time between two such computer runs, it is recor on tape either at once or at any time before next computer run. Immediately before this r all the accumulated new information is arranged in the order in which it is to be ente into the master file. The main computer run tl consists in reading the master file, one entry a time. After reading each entry we first exam the next item on the "new information" list (ta to see whether it is to be inserted before this en or modifies it. If neither, we put the master entry through the information retrieval rout and proceed to the next master entry. If, h ever, the next "new information" item does for insertion or correction, this is carried out, new or corrected item is put through the inf mation retrieval routine, then the following "n information" item is examined in the same w etc.

We expect to use the same procedure in updat the file of standard reference data, except to one of the new erasable mass storage media

[2] Conventionally an average word is thought of as 5 characters. Many contemporary computers use words characters (36 bits). Most data items are only 3 or 4 dec digits (10–13 bits) but because of redundant notation oc space equivalent to about one word.

place of tape. Thus, the only additional effort required for updating is that of actually examining the new information and making the changes in the master file; since normally only a small number of changes occur, this effort is small compared to that of reading the entire master file. The latter operation need not be carried out beyond the extent necessary for information retrieval. At the same time, information retrieval is always based on completely up to date information, since all corrections are made before an old item is used for retrieval.

We have no estimate of the rate at which new information will flow into the system, but it is obvious that if retrieval and updating runs are made daily or even weekly, only a small fraction of the file items will be affected by updating in an average run.

2.2.3. Aids to Publication

"Aids to and from publication" would be an equally appropriate heading; sometimes the existence of machine-readable information files is a help in editing such material for publication; at other times, the creation of such files is facilitated by steps taken primarily for the purpose of publication.

The use of machine-readable material in the publication process can be advantageous in several ways. The most obvious case is that in which the information is numerical and has been produced on a computer, so that the entire costly process of manual typesetting is avoided. For years, tables of such numbers have been produced on typewriters or line printers controlled by punched cards or magnetic tape; these are of limited flexibility, the resulting printed copy suffers from poor readability, and column headings, pagination, etc., present annoying problems. Since the introduction of tape-driven photocomposition devices has made it possible to produce printed output of letterpress quality directly by computer, at costs comparable to those of manual typesetting, there is no reason why computer-produced material should ever have to be hand-set for printing.

Another reason for computer-controlled type composition is the facility for rearranging or otherwise revising the material. This is important whenever the same material must be printed in several arrangements. An example is the volume "Crystal Data Tables" now being prepared for publication by photocomposition. The original information is being keypunched in essentially the same format in which it is to appear in print, and this part of the operation offers no great advantage over manual typesetting. But from the keypunched information it will be possible to produce alphabetical indexes for authors, names and formulas of chemical compounds, all by automatic sorting, checking, and editing. In subsequent editions, only the new or revised material will have to be newly keypunched, the computer will insert it in the proper place, change pagination as needed, update the indexes. In addition the computer can perform a large number of checks based on the characteristics of the information: the crystallographic data satisfy certain inequalities, abbreviations for journals occurring in literature reference must be taken from standard lists, the order of items can be checked etc. This saves a large part of the proofreading effort. Against these savings must be reckoned the effort of writing special computer programs.

The publication of "Crystal Data Tables" is an example of a situation in which the desire to use automatic type composition furnishes the incentive for recording the information in machine-readable form, and thus aids in the creation of mechanized files. Other instances of this kind will probably be bibliographies and acquisition lists which are to be published in cumulative, updated form at frequent intervals. There will be other cases in which automatic type composition will become attractive only after the information has been recorded on tape for a different purpose.

2.2.4. Other Computer Functions

Even before a mechanized information file for retrieval and updating has been created, computer methods can be used for a number of housekeeping functions. It has been the experience of other information centers that some of these functions are more easily mechanized than the information storage and retrieval itself.

For example, even while OSRD still uses a manual information retrieval system based on conventional library practices, a computer could conceivably be used to keep track of purchase orders, accessions, shelving, classification, and circulation (loans of books to users). It is possible, however, that mechanization at this stage, while practiced by some other installations, may not be economical for OSRD because of the small size of its library.

It may be desirable to keep statistical information on the requests for information which are acted on by OSRD. Such information will be a promising candidate for automation after a brief initial period of manual handling, during which the staff becomes familiar with the number and types of questions to be expected.

Similarly, machinable records may be helpful in the indexing of the information collection according to properties, materials and certain classes of materials, and some other characteristics (e.g., by checking manually introduced index terms against computer-stored master lists of such terms to insure uniform nomenclature).

It may be desirable to maintain records of sources of information other than OSRD's own collection. These sources are primarily of two kinds: on the one hand, knowledgeable individuals and organizations for referral, and on the other hand, books, papers, and unpublished reports. The former are probably too small in number to warrant use of machine methods. The latter are numerous, and should be the subject of a mechanized file system if it is decided that OSRD will make use of the scientific literature to any extent.

While it is debatable which of these housekeeping operations should be mechanized before the main operation of OSRD is switched over to computer use, all of them are certainly likely to be among the functions of the ultimate computer system.

Apart from housekeeping operations there is an entirely different computer function which promises to be particularly useful in the Standard Reference Data Program, namely the retrieval of information by means of citation indexing and the related method of bibliographic coupling. A citation index is obtained by recording, for each scientific paper or report belonging to a given field of knowledge, all the papers cited in it (customarily these are shown as a "list of references" at the end of the citing paper); and then *sorting* this record in order of the cited papers. Thus we obtain for each paper a list of places where it has been cited. Suppose now, for example, that a scientist wishes to find the latest value for the atomic weight of some element. He knows that this was measured some years ago, and that it may have been revised since. He enters the index with the latest publication on this subject known to him—perhaps 5 or 10 years old— in the hope that the publication of a subsequent revision would reference the previous result. It is obvious that this kind of problem occurs frequently in operations such as the NSRDS data centers. Other applications of a citation index are in bibliographic coupling (finding papers which are related to a given paper in the sense of citing some of the same literature), preparation of bibliographies, current awareness programs, finding reviews of a given paper or corrections to it, etc. These examples may suffice to show that a citation index is not only a useful tool in many scientific undertakings in a general way, but is also particularly applicable in an information system such as NSRDS.

2.3. Input to the File

2.3.1. Keypunching

If, as we have said, the information file will contain about 100 million numbers, then the problem of recording all these numbers initially in machine-readable form is considerable. Let us assume, for example, that the information is to be punched into cards. Experienced organizations like the Bureau of the Census, where keypunching is done on a large scale, estimate the cost at 10 to 20 cents per card. Each card holds 80 decimal digits. Our data may, on the average, have 3 to 4 significant digits each, but since they vary in magnitude one may have to set aside 5 to 6 card columns for each number. Some card columns are needed for identification; on an average, we may get 10 data numbers per card, or a total of about 10 million cards, at a keypunching cost on the order of 1 million dollars. This is for the initial effort; it would be followed by somewhat smaller annual outlays for updating.

This is not prohibitive in comparison with the size and cost of the entire NSRDS operation, but it is large enough to warrant serious study. Fortunately there are alternative ways of original recording for at least part of the collection.

Parenthetically, one might reflect for a moment that 10 million cards fill about 200 file cabinets again a large but not prohibitive number. There should be no reason, however, for storing all these cards simultaneously for any length of time; storage for permanent record would presumably be on tapes or other magnetic media. For instance about 100 to 200 reels of ordinary magnetic tape would suffice to hold the entire collection.

2.3.2. Print Reading

The art of automatic reading of printed cop and recording it in computer-readable form ha been developed to a point where it is possible i most cases, and economical in some. The princi pal difficulties now are not in the machine recogni tion of printed characters but in paper handling turning of pages in books, registration (precis location of copy relative to the reading device) dirt and other imperfections of printed copy, spe cial symbols and unusual type fonts, etc. Printe tables of numbers are relatively simple and fre of most of those difficulties; it is likely that dat which have been assembled into printed volume print style and arrangement remaining complete uniform over many pages, can be handled by auto matic print readers at a fraction of the cost o manual keypunching. To date the main area o practical application of automatic reading ha been to bank checks, but the problem has bee extensively studied, e.g., in connection with ma chine translation of languages, and economica solutions appear to be imminent.

2.3.3. Data Available from Data Centers

Most of the data to be incorporated into th NSRDS come from technical data centers or othe organizations engaged in compiling data, an some of these will provide them in machin readable form for reasons of their own. Fo example, Professor R. Pepinsky's collection o crystallographic data is already on magnetic tap (This, however, is a collection which has not bee subjected to critical evaluation, and cannot be co sidered as standard reference data.) Other exis ing centers use cards or various forms of punche tape, and still others operate manually at prese but will mechanize as they grow. On the othe hand it seems likely that a majority of data cente will always prefer manual operation.

Even then they may have occasion to recor certain sets of data in machine-readable for

As we have indicated, this may be done as an aid to printing, as in the case of the "Crystal Data Tables" (cf. sec. 2.2.3). Or the data may be the consequence or result of numerical computations performed on digital computers, or the result of measurement using instruments which are equipped with automatic recording devices. The latter two uses often occur jointly and reinforce each other; one of the arguments for automatic recording of measurements may be the fact that the results have to be subjected to certain numerical transformations before being used. Spectrometers, x-ray diffractometers, and bubble chambers are examples of instruments which employ automatic recording on a large scale.

2.3.4. Equipment and Format

Machine-readable data coming from so many different sources will appear in a variety of forms. Different media will be used, such as punched cards, punched paper tape of varying widths, magnetic tapes, and possibly others; and the format used with each medium will not be uniform. It will be one of the functions of OSRD to coordinate and standardize these media and formats.

Conversion from one medium to another can be accomplished automatically, at moderate cost, and is at worst a minor nuisance, as long as the formats used correspond to each other in a simple way. On the other hand, conversion from one format to a different one may be easy or hard, depending on whether the source format contains all the information needed, and whether this information appears in approximately the same order as in the target format. Therefore, in order to minimize the difficulty of conversion of data originating in different data centers, OSRD ought to establish one set of standard target formats, and suggest to the data centers that they use, not necessarily these but at least some formats which are easily convertible to the standard ones. For convenience the standard formats would be expressed in terms of one particular medium, for example, the ordinary 80-column punched card, since it is widely used as input to computer systems and for storage of information, and facilities for keypunching are widespread and not expensive. This would leave data centers, and indeed OSRD itself, free to use any other medium, so long as they choose formats which translate easily into the standard ones.

Thus OSRD might establish, in cooperation with the data centers most concerned, standard formats for recording data on punched cards. There would, of course, be a different format for each kind of data; possibly hundreds of formats would have to be agreed on. The established way by which computing laboratories record and exchange detailed information about formats is the "card sheet," a list of instructions for punching each column in a card. Thus this part of the job of OSRD may be described concretely as devising a card sheet for each of its card decks.

The cards may not serve directly as computer input nor as storage medium. In the present circumstances, cards would be transcribed to magnetic tape which would serve both purposes better. This transcription is character-to-character, and is therefore cheap and reversible. The same card sheets which describe the arrangement on cards can be used to document the contents of the tapes. In a few years, a medium other than tape may be preferable; the transcription, again character-to-character, would be no problem, and the same documentation could continue to be used. At present we cannot foresee a medium which would not be compatible with the punched card code (although in the more distant future, there may be too few distinct card codes available). The converse is not true; many tape codes, e.g., cannot be transferred to cards without some added structuring.

It is entirely possible that some of the information may never be physically on cards; it may, e.g., be recorded on paper tape by the originating laboratory, transcribed there to magnetic tape, transmitted to NBS over a radio information link, and recorded again on magnetic tape. It would nevertheless be convenient to think of the arrangement of the information as if it were on cards—as in many cases it will be.

We have so far discussed digital (numerical or alphabetical) information. A few words are in order about graphical information. The digital representation of curves described in section 2.4.3 below is economical, but the conversion (consisting in the computation of a number of coefficients) may be considered too difficult by some of the data centers. Microfilm, mirofiche, video tape, etc., can be used more directly for storage, and facsimile transmission of such information is possible. At present it is not clear how such information would be integrated into the mainstream of digital computer operation envisaged for NSRDS. This subject is further discussed in section 3.2.4 below.

2.4. Information Storage

2.4.1. Form of Data

The simplest kind of information with which the system deals is exemplified by the statement that the atomic weight of hydrogen is 1.00797. This is expressed by three terms:

Atomic weight — Hydrogen — 1.00797

of which one denotes a *property*, another a *material*, and the last a *value*. A more elaborate item is needed to convey the information that the density of water vapor at 500 °K and a pressure of 10 atm is 0.0045967:

Density — Water — 500 — 10 — 0.0045967.

It is not necessary to record the information that the numbers 500 and 10 represent temperature and pressure; this information is implicit in the definition of "density," as are the units in which temperature, pressure and density are given. That is to say, the computer program for retrieving information on density must contain instructions to the effect that following the designation of a material there will be recorded a series of triplets of numbers, namely two parameters representing temperature and pressure, and the corresponding value of density. (Actually, the information will be stored in more compact form, to be discussed in sec. 2.4.3.) There will also be instructions for adding the symbols °K and g/cm^3 after the appropriate numbers in the printed output.

Properties and materials can be represented in the computer by numerical codes. For example, the ACS registry number might be used for materials, while properties might be denoted by arbitrary serial numbers, or by the NSRDS classification number followed by one or two digits which specify the particular property within its class. Again, the computer program has to contain instructions to replace these numbers by English words in the output.

In addition to property, material, parameters, and value, an item may contain comments, comparable to footnotes in a printed table. These may be indications of source, such as a laboratory name or a literature reference, or explanations, cautions, etc. The computer program provides for printing these where appropriate. The literature references could conceivably be used by themselves for an entirely different purpose, namely bibliographic searches, but it is doubtful whether OSRD ought to render services of this kind.

Apart from numerical data, there is frequent need for data in graphical form. One may expect that the demand for graphical data will be somewhat reduced by the easy availability of numerical information, but there will probably be a residue of curves which must be stored and retrieved. It is not yet clear how this is best done. One could use a computer-driven curve plotter, though past experiments with such a system for spectra have not been encouraging. One could keep a manual file of graphs (either on microfiche or as hard copy) and use the computer only to obtain reference numbers to this file. Hardware exists for automatic retrieval of microfiche, but it would be a foreign body incompatible with the main system. It is possible that in the next few years an automatic microfiche retrieval system may be developed which can be connected to, and steered by, the main computer.

2.4.2. Arrangement of Data

Since practically all data in the system are described as properties of materials there are two arrangements which suggest themselves naturally. We could group data by properties, starting with one property and listing the values of this property for all materials, then proceeding to a second property, etc. Or we could use materials as the major subdivision and arrange by properties under each material.

The method which we actually propose to employ is a combination of these two obvious ones. We suggest that the set of all properties be subdivided into homogeneous groups of related properties, and that the entire data file be arranged by this grouping. Within each property group there would be a listing of all pertinent materials, and under each material a listing of parameter values, each followed by the values of the several properties in the group. This may be illustrated by the example shown on the following page (from NBS Monograph 20).

By saying that the properties in a group are "related" we mean merely that they are frequently used together. We are therefore likely to save time when looking for the answers to a group of related questions.

The property groups are "homogeneous" in the sense that the properties in a group depend on the same parameters, and are meaningful for approximately the same ranges of these parameters. This facilitates the storing and also the retrieval, since the same set of computer instructions can be used for all properties in the group.

Finally, the arrangement by property groups is similar to that usually found in print, and therefore facilitates the original recording of data. On the whole, this arrangement is a natural extension of the one to which data suppliers and users have been accustomed. It hardly needs emphasizing that we retain the option of employing a grouping of properties which differs from the conventional one, whenever this is advantageous for machine retrieval. In many cases we expect that a homogeneous group will contain only a single property so that we will be arranging "by properties."

2.4.3. Compact Storage

It is most important to store information in the least possible space, both because storage capacity in the computer must be paid for and because the time required for every search may increase with the number of words stored. We shall consider two kinds of space savings: omission of identifying information, and condensation of the functional values themselves.

The possibility to omit identifying information depends on details of hardware organization which cannot yet be foreseen. If every storage location were addressable, the identifying information, such as name of property and material and values of parameters, would be replaced by the choice of address. For a simple example, suppose that values of a function of temperature are to be stored at intervals of 10 °K in consecutive addressable storage locations beginning with address 12,288. Computer instructions specifying that the function value for any argument T is stored at address $12{,}288+0.1\ T$ are sufficient for storage and retrieval, and no value of T need be stored. Other identifiers can be handled similarly, or one can store one identifier value preceding the entire group of function values to which it pertains; for instance, record one value of pressure preceding a group of numbers representing density at different temperatures.

ample of Arrangement of Data by Property Groups

[From NBS Monograph 20]

Property Group: Ideal Gas Thermodynamic Functions

C_p°/R	$(H^\circ-E_0^\circ)/RT$	$-(F^\circ-E_0^\circ)/RT$	S°/R

Material: H_2—*normal mixture*

. . . .			
2. 71388	3. 81909	8. 45365	12. 27274
2. 78512	3. 72183	8. 41288	12. 53471
. . . .			

Material: HD

. . . .

[N]ote: In this example, four ideal-gas thermodynamic [func]tions, namely C_p°/R etc., have been selected as one [ho]mogeneous group of related properties" and used as a [maj]or subdivision of the data file. All depend on the same [par]ameter, temperature (°K), and are recorded for the [sam]e range of this parameter. This major subdivision [of t]he file is further subdivided according to materials; [und]er each material there is a listing of values of tempera[tur]e, each followed by the corresponding values of the [four] properties.

[I]n practice it is unlikely that we shall work with [ad]dressable, stored words. In tape and other bulk [sto]rage media, addressing is usually by blocks of [dat]a. These blocks may be of fixed size or they [ma]y be variable but with limits on size imposed [by] the economics of their use. The cost of retriev[ing] a single word is not much less than that for a [wh]ole block of words. A simple procedure is to [sto]re all identifiers common to a block of data at [the] beginning of the block, as long as this results [in] blocks of manageable size. To go into a more [det]ailed design at this time would be premature.

[E]conomy in the functional values leads to con[cep]ts which have long been studied by the makers [of] mathematical tables: interpolation and approxi[ma]tion. For simplicity, consider again the case [of] one function of one variable, say density as a [fun]ction of temperature (at constant pressure):

Density of steam, at 1 atm

[From NBS Circular 500, p. 450]

Temp. °K	*Density g/cm³*	*Differences*	
800	.000 274 64	339	7
810	271 25	332	8
820	267 93	324	9
830	264 69	315	6
840	261 54	309	--
850	258 45	---	--

[In] the example this is tabulated at 10 deg inter[val]s. For temperatures falling between these [val]ues, the density can be obtained by linear inter[pol]ation with an error of not much more than one [uni]t in the last place. With a computer, interpola[tio]n of higher order, say the third, is quite prac[tica]l. For this it would suffice to tabulate values [at] much larger intervals, say every 50 deg. Then [the] density value at an intermediate temperature [is] found by passing a cubic polynomial through [fou]r points, two to the left and two to the right [of] the desired value. To save computation one [doe]s not store the density values at all, but instead stores the coefficients of the interpolating polynomial. This polynomial will reproduce exactly the desired value for $T=50°$, 100°, etc., and will give a sufficiently close approximation at other values of T. Going one step further, we note that there is nothing sacred about these special values of T, and no reason to insist on precise agreement at just these points. So, instead of interpolating between these we use a polynomial which best approximates the density function throughout the entire interval. This, in turn, enables us to make the interval still longer without getting intolerably large deviations. In summary, then, we first choose a class of approximating functions—say, cubic polynomials—and a tolerance limit—say, two units in the last decimal place of the table (2×10^{-8}). We then find the longest interval, beginning at $T=0$, for which a cubic can be found which approaches the given density function within 2×10^{-8}; and we store the end point T_1 of this interval, together with the coefficients of the polynomial of best fit. (The technique which accomplishes this is the method of Chebyshev polynomials.) Then we find similarly a longest interval starting at T_1, etc.

It remains to discuss the choice of the class of approximating functions, which for our example has been cubic polynomials. If we increase the degree of polynomials used, the interval which each will cover increases, so that we need fewer intervals but more coefficients for each, and more computation to evaluate the function. The optimum compromise between these conflicting factors will differ for different functions, but will in any case be for a polynomial of higher order than that used in manual computation. Functions other than polynomials may be considered. Polynomials have the double advantage that they are easy to evaluate and easy to fit (i.e., the coefficients of the optimal polynomial are easily found). Since computers are made to carry out the arithmetic operations of addition, subtraction, multiplication, and division, the only functions which are as easy to evaluate as polynomials are rational functions, and they are hard to fit. They may be used in special cases where singularities or asymptotes are present. Many kinds of orthogonal series, e.g., Fourier series, are just as easy to fit as polynomials but are harder to evaluate. One may put up with this drawback if the nature of the problem seems to call for it. For example, there is some recent work on representing spectra by sums of Gaussians, each of them with three parameters representing the mean frequency, width and intensity of one line. These are not quite orthogonal but almost so.

The stored coefficients, of course, must be derived separately for each table. This effort can largely be mechanized, but even so it is of considerable magnitude.

In some cases it may be possible and desirable to store numerical indicators of the accuracy and/or precision of the tabulated data. Judgments about the reliability of data are among the principal concerns of NSRDS, and should be recorded as far as possible. For the most part such recording will initially not require great sophistication, nor will it add appreciably to the requirements for storage space. For the foreseeable future the large majority of accuracy estimates will be qualitative and will find their expression in the selection of the tabular values from among several competing measured values. Some will be numerical (such error estimates are now being tentatively assigned e.g. to tabular values on heats of formation.) For the tables occupying large portions of memory space, e.g. properties tabulated as functions of temperature and pressure, it will frequently be sufficient to record one single number representing the accuracy of the entire table. The ultimate goal of recording an error estimate alongside each tabulated value is a long way off.

2.4.4. Storage of Instructions

There is an intimate connection between the data to be stored—in the case of the preceding section, the coefficients of approximating functions—and the computer instructions needed to calculate these functions. These instructions are an integral part of the stored information. Inasmuch as they are different for each table (or at least, there will be many different sets of such instructions, although some may apply to more than one table) they might as well be stored with the data. In order to minimize this storage, one will attempt to devise a general retrieval program (e.g., "polynomial approximation"), or perhaps several such programs, applicable to different classes of tables. From such a general program, the specific program needed for a particular table is derived by specifying a few numbers, like degree of polynomial, number of variables, etc.; only these need to be stored with each table.

The point to note is that there is often a tradeoff between data and instructions, and again between special instructions applying to only a small segment of data, and more general ones. Special instructions should be stored with the data, so that no separate lookup is needed; general ones should be in internal computer memory, where they are always accessible. Apart from the limited size of this memory, the major limitation on general purpose instructions is the effort of creating them. They are so important, however, that this programming effort deserves major support.

2.4.5. Storage Devices

As stated before (secs. 2.1.2, 2.2.1) the amount of information to be stored in our system may be vaguely estimated at 100 million words. Internal computer memories store usually 65,000 to 131,000 words. This may increase in the next few years, but not to anything like the volume we require. We shall therefore have to rely on external memory components.

Conventional external memories are magne tapes and drums, and more recently disk files. addition, several new devices have just beco available, and several others are under develc ment and may be expected to be available when need them.

Magnetic tapes have practically unlimited c pacity and are inexpensive. Perhaps 100 ree more or less, depending on length of bloc would hold all our information, at a cost of a fe thousand dollars. However, most computers ha only a small number of tape reading stations, a these have to be shared with other users of t same computer. Tapes have to be mounted a changed manually. The time to find a particul item on tape (random access time) is on the orc of minutes, and the rate of reading successive i formation is too slow for our needs. Therefo tape must be ruled out, except perhaps for an i tial period of transition and for certain auxilia purposes.

Magnetic drums, until recently limited in c pacity, do now have the capacity required for o applications. Disks and other existing or futu storage devices likewise possess the necessa capacity.

Apart from capacity, the transfer rate, i.e., t number of words which can be transferred fro consecutive storage locations into the main fra of the computer per unit time, is critical for o application, since for some of the simpler pro lems it will be necessary to scan a section of su cessive storage entries and perform only a fe simple computer operations on each item, e. comparison with a search request. For this pu pose, in order to avoid delays, the transfer ti must not exceed the time required for, say, t elementary computer operations, about 10 to microseconds with today's computers, correspon ing to a transfer rate of 2 to 4 million bits p second (cf. sec. 2.1.4). For many other purpos the computation to be performed with each item information will be more complex, and therefo the transfer rate less critical.

The random access time is not critical; anythi below, say, one second is certainly acceptable.

Magnetic drums and disk files, which have be in existence for several years and are well teste will easily meet all requirements. They are, ho ever, somewhat expensive. The newer mass sto age media, more reasonable in price, fall somewh short in transfer rates. These are quite new a likely to be improved, and several companies a working on the development of large external sto age devices, so that it is likely that something sui able will be available at a reasonable price by t time it is needed by NSRDS.

Whatever device is used, the file of standard re erence data will be kept permanently in storage it, and will be periodically updated. It will ther fore be necessary (cf. sec. 2.5.3) to have th memory component reserved for the exclusive u of NSRDS.

2.5. Access to the Computer

2.5.1. Man-Machine Interaction

In the preceding sections we have frequently ade passing reference to the ways in which the omputer is interrogated or instructed. A large art of the requests for information received by SRD, as well as of the new data to be incorpo-ated into the files, are collected and run on the omputer, say, once a day, using a general infor-ation retrieval and file updating program. ommunication with the computer will be in the onventional manner, i.e., using a small peripheral omputer or "secretary computer" questions and ew data are manually keypunched into cards (or, a few systems, punched paper tape or a loosely acked magnetic tape), and then loaded into the eripheral computer and there converted to mag-etic tape of a format suitable for the main com-uter, and possibly combined with input to other roblems; then the whole batch is run on the main omputer, resulting in an output tape; and finally e output is printed on the peripheral machine nder the control of the output tape.

A similar regime will govern certain special roblems which have their own special programs ut which can nevertheless be batched with each ther or with other problems. In this class are e housekeeping problems and the preparation of aterial for publication, in which the retrieval information is followed by detailed editing rocedures.

There is, however, another class of special prob-ms which cannot be handled in this way. These e the requests for information which do not fit to the general information retrieval program f. sec. 2.2.1). They are of great importance for e system because through them we learn how to nprove the general-purpose program. In most these questions it will be necessary to "feel one's ay," asking a tentative question, awaiting the swer, modifying the original question, etc. his is anologous to the process of human infor-ation retrieval. Librarians have made serious udies of this process and report that a large frac-on of all requests for information need rephras-g at least once, often several times.

For this reason it is deemed essential to have nvenient facilities for man-machine "conversa-on." We visualize a console in the offices of SRD, with facilities for input by keyboard and nched cards or punched paper tape; typewriter tput; and the ability to connect on-line to the ain computer. It is likely that similar consoles ll be placed in other locations at NBS, where ey will serve a variety of purposes. The main mputer must be operated under a system which ows interrupting long problems in order to mit short requests from the remote stations; the eration is thus characterized as *time sharing* der *remote control*. Naturally there must be ffering to avoid tying up the main computer ile input from, and output to, the remote sta-ns is slowly processed. A small amount of memory is the minimum requirement for such a buffer, but it will probably be more efficient to use one or more small satellite computers which are on-line connected to the main computer as well as to the remote stations. Possibly OSRD, because of the large amount of data which it handles, should have one such satellite computer reserved for its own remote use.

2.5.2. Long-Distance Access

Once the principle of time-shared remote control of the computer has been established, it is only a small step to a system which places similar remote consoles in locations at much greater distance from the computer, say across the country. The hardware techniques for doing this are already in existence; at the time of this writing, stations at the National Bureau of Standards in Washington communicate with computers in Cambridge, Mass., Dartmouth, N.H., and Phoenix, Ariz. Ordinary telephone lines are used for interconnection. This is done not merely for experimental and demonstration purposes but for effective computation, albeit on a small scale. The problem is thus not a technical but an economic one.

That there is a need for such facilities is made plausible by the observation, made by many existing information centers, that a large part of the inquiries which they receive—usually more than one-half—comes from the installation in which they are located. One may well suspect that convenient access to information is an important factor; that there is an equally great need for information in the many outside installations, but this need does not express itself in inquiries because of the slowness and inconvenience of operating over greater distances. Indeed, to satisfy this latent need for information may well turn out to be the greatest accomplishment of NSRDS.

Now there are in principle two ways in which this can be done. One can enable laboratories throughout the country to obtain on-line connection, via long-distance telephone lines, to the central computer and information store at OSRD; or one can duplicate this store in numerous geographically dispersed computing facilities. This saves the cost of a long-distance telephone call for each inquiry, but involves a much greater investment in storage equipment and the considerable difficulty of keeping all these copies of the original store (and of the computer programs which go with it) exactly updated.

There is no point in drawing up a precise balance sheet of costs at this time, since the information file does not yet exist, will take several years to compile, and some cost items may change radically in the meantime. Other organizations are faced with similar problems, in particular the National Library of Medicine, and their experience will be valuable to us. We venture the guess that if a system had to be introduced today, the establishment of a moderate number of "secondary information centers," each a copy of the primary center at OSRD, would be optimal; but that as time goes on, the optimum will shift in the direction of greater centralization. In any case, we envisage the establishment of a nationwide network, either of secondary centers or of telephone access to the primary center, as something to be done only after the primary center itself has been in operation for awhile.

2.5.3. Administrative Arrangements

We have already indicated that it will probably be desirable for OSRD to share the major computing facility of the National Bureau of Standards, but that OSRD will have to procure for its own exclusive use a large external storage component and a remote console. In addition, OSRD will have a heavy share in the use of a satellite computer to act as buffer between the remote console and the main computer; it may even require one entire satellite computer for its own exclusive use. This computer would have to be located in the computing laboratory—transmission at the high pulse rates used by the main computer limits the distance—and it would have to be operated in accordance with the ground rules and operating systems of the laboratory; probably it would be operated by computing laboratory personnel.

The cost of main frame time will depend on the workload. For the example given in section 2.1.4 (something less than 2 hours per day on the IBM 7094) and at today's rates, it would be about $80,000 per year. Future computers will do the same amount of work at far lower cost, but the workload will undoubtedly go up. If the external store is to be acquired by rental, the annual cost at today's rates might be $60,000; only an order-of-magnitude estimate is possible. The decision be- tween renting and purchasing will have to b made just prior to acquisition; usually the advar tages are almost evenly balanced, and the cost di ference is smaller than the uncertainties in th present estimates of cost and workload. The co of a separate satellite computer, if one is require is also hard to foresee, since such smaller compu ers come in wide price ranges; the order of magn tude might be $100,000 of annual rental. The co of the remote console is very small by compariso

It is to be assumed that the computer progran for information retrieval, editing, display, upda ing, etc., will not remain static but will be in a co tinuing state of development. This may be doi by personnel of the Computation Laboratory, of OSRD, or both, but in any case the services several full-time programmers will be require A much larger number of people—possibly b tween 25 and 50—will be needed to prepare inp data and requests for information, accept outp data and send them to their destinations, e These will undoubtedly have to be OSR personnel.

There will be a somewhat larger initial pr gramming effort, extending over the first few yea and costing perhaps several hundred thousand do lars—this cost depending very greatly on how an bitious the initial general-purpose program is, how much is left for later improvement. The in tial keypunching of data, at a few cents per wo is an even bigger investment (cf. sec. 2.3.1).

Nevertheless, all these costs are not large in co parison with the intellectual organization of t information. The latter will represent the pri cipal effort of NSRDS.

3. Proposed Interim System

3.1. Characteristics of the Operation

The real-life conditions under which OSRD will have to operate in the next few years are very different from the ideal situation which has been postulated, explicitly or tacitly, for the mechanized information system described in the foregoing. We assumed the existence of a network of technical data centers, so complete that every physical property of materials for which measured data exist falls into the province of one or the other of those centers. We assumed that each center has collected the existing measured values for all properties for which it is responsible, has critically evaluated them and thus arrived at a collection of standard reference data which it continues to update. Copies of all these standard reference collections form the data file of OSRD, and are used to reply to the numerous requests for information reaching that organization day by day from all parts of the country.

In reality, it will take a long time to establish recognized technical centers in all areas of physical science. For the next few years there will areas not covered by any center, other areas which some work is done by organizations not a hering to NSRDS, along with a small but incre ing number of member centers operating witl NSRDS. Even in the areas for which cent exist, it will take time to set up criteria for t evaluation of data, and more time to perform t actual compilation and evaluation. Therefore, some areas OSRD will have no data at all, in ma other areas it will rely on such compilations as can find in the literature or obtain through ad l correspondence; these will be either unevalua or subjected only to preliminary informal eval tion by OSRD staff. And finally, the demand information will not arise at once in full stren but will build up gradually as OSRD becor known, as scientists and engineers get into habit of using its services, and as they learn h to adapt their working methods and their appro to new problems to the easy availability of t information.

In brief, this period will be characterized by apid change in the quality and quantity of data nd in the volume of demand for services. At the ame time, our own understanding of the situation ill still be deficient, and will become more adeuate only with the passing of time, through our xperience with the operation itself.

3.2. Inquiry Services

3.2.1. Getting Started

The circumstances set forth at the end of the receding section suggest strongly that informaion services to scientists and engineers should bein at once, without delay, not only because the enefits accruing to the technical community from he availability of such services should not be postoned by several years pending the creation of ata files and systems, but also because the creaion of the systems will itself be aided by the exerience which OSRD will acquire in the process f rendering services.

It follows that, in many fields, requests for data ill have to be answered before standard reference ata have been so designated. In such cases, rather han merely indicating to the inquirer that no RD are as yet available, it will be preferable to ive him whatever information can be found in he literature or through personal inquiries. A uitable disclaimer should be appended to such eplies, cautioning the user that the information as not been evaluated by NSRDS.

Within OSRD, the handling of such requests is function of the Information Services Operation ISO). The reaction of ISO to an information equest may take any of the four forms listed in ection 1.3.3 above: referral to an expert, literaure reference, documentation, or data information.

One may wonder whether the willingness to rely in terms of other than SRD would tend to verburden the organization. The charter of SRDS does not commit us to anything beyond iving information on SRD, and one could choose o draw the line there. Actually, however, our orkload for the proposed broader service will be o greater than it would be for the narrower one f standard reference data had already been designated in all fields. Possibly the search for unevalated information is more laborious than the rerieval from an organized file, but on the other and a larger fraction of the inquiries will be anwered by mere referral. Thus the volume of work hich ISO is taking upon itself is no greater than hat to which it will eventually have to get accusomed anyway. Meanwhile, the broader services ill be beneficial to ISO itself as a realistic traing ground, to the customers who need the infornation, and most of all to the entire community by astening the process of acquainting scientists ith NSRDS and getting them into the habit of naking use of it.

As stated in section 2.2.1 above, the experience f other data centers leads us to expect that the number of requests for information will start, once the existence of OSRD has become generally known, at a level of several per day, and will grow from there. A majority of the requests will be "administrative" and can be handled by ISO staff without difficulty. The requests for technical information proper will be screened by ISO staff and processed by one of the procedures outlined in the next few sections.

3.2.2. Referral

In the early years, while NSRDS is still developing, a large portion of technical queries received is likely to be referred to experts. We expect that this practice will gradually decrease but will never cease entirely. There are pitfalls in many seemingly simple technical questions which cannot be avoided by the uninitiated. Until OSRD has accumulated some experience it will be well to have *all* technical answers, even those routinely prepared by OSRD from its own files, checked by a specialist. Later on it will be possible to dispense with this for the more frequently occurring types of questions. Perhaps the hardest problem, in the long run, and one for which we have no ready answer, will be for the ISO staff to recognize when a problem needs a specialist.

The experts to whom questions are referred can be taken from the following groups:

(a) Technical area managers of OSRD.
(b) Data centers which adhere to NSRDS.
(c) Divisions at NBS outside OSRD.
(d) Other scientists.

In general this will be the order of preference in calling on experts, except that sometimes (c) may precede (b). Technical area managers are so few in number that it will often be practical to bypass them. It may be hoped that experts in each category, if they are unable to handle an inquiry, will at least suggest other more suitable experts in the higher categories.

The system used in referral must meet the following conditions, in this order of importance: the inquirer should receive, without undue delay and without further effort or annoyance to him, a reply which is correct and helpful; the replying expert should receive credit and should not be unduly burdened; ISO should be able to add to its storehouse of experience, both technically in regard to the specific question and administratively in regard to statistical distribution of questions in general. Finally, direct back-and-forth contact between inquirer and expert needs to be facilitated for those cases where the formulation of the question itself presents problems.

It is believed that the following stepwise procedure meets all these criteria.

(a) ISO ascertains what information it can furnish from its own files.

(b) If this is inadequate, ISO quickly locates a suitable expert—normally by a series of phone calls—and establishes that he is able and willing to handle the inquiry.

(c) ISO forwards the inquiry to the expert by phone or, if too voluminous, by mail.

(d) Simultaneously ISO informs the inquirer of this action.

(e) The expert replies to ISO, which relays the reply to the inquirer without delay.

(f) ISO follows up if the reply is not forthcoming after a reasonable delay.

(g) ISO maintains statistics on inquiries received and on their disposition.

It is important that ISO consider this procedure not only as a way of satisfying the inquirer but also as an opportunity for its own staff to deepen their understanding of the technical questions asked, so that ISO's own staff will gradually be enabled to handle an increasing portion of inquiries.

In regard to (b), it is likely that ISO will rapidly—probably in a matter of months—become acquainted with a large number of experts both at NBS and in various data centers, with their fields of specialization and with the degree of their willingness to handle inquiries. Initially, the NBS Index of Technical Activities, the NAS–NRC "Directory of Continuing Numerical Data Projects," or inquiry from OSRD technical area managers or NBS technical division chiefs will reveal the names of suitable candidates. Ultimately, an index of such experts, compiled and published by ISO, may in itself become a useful addition to the technical literature.

3.2.3. Reference

An inquiry should be answered by one or more references to the literature if ISO is able to obtain these references without undue effort and if, at the same time, it is impractical to provide copies of the pertinent documents or portions of them. The latter would be the case, e.g., for obscure journals, unpublished reports, documents which are in the main library of NBS but not in the collection of ISO; also for inquiries where the reply involves a large number of pages, too difficult to reproduce; and finally for inquiries answered by phone. In all other cases, namely where an inquiry can be satisfactorily answered by mailing copies of a few pages from a document easily accessible to ISO, it will be preferable to do so rather than merely to give the inquirer a literature reference.

The main limiting factor to the use of both reference and documentation (furnishing of copies) will be the difficulty of locating the references. For this purpose ISO would either have to maintain a large, well indexed and updated file of literature reference, or to undertake an ad hoc search separately for each inquiry. Either alternative is so laborious that there is at present no justification for adopting it, in view of the likelihood that ultimately, some years from now, there will be a mechanism for answering almost all questions from the data file itself, without referring to the literature.

It is therefore proposed that the use of both referencing and documentation in answering to inquiries be limited to those cases where the needed references are obtainable without too much effort.

Potentially, the job of obtaining references, especially those to recent documents, can be greatly facilitated by the use of a citation index. Therefore, the development of citation indexes, and of the related subject of bibliographic coupling, should be closely watched. Presently available citation indexes, namely those of M. M. Kessler at the MIT library and of the Institute of Scientific Information in Philadelphia, are somewhat deficient in coverage and accessibility. Nevertheless, there are reports of encouraging trial uses of these indexes. Similar experiments are contemplated by OSRD. (See sec. 2.2.4.)

3.2.4. Documentation

Under this heading we discuss, as mentioned before, the furnishing of copies of documents in response to data inquiries. For the next few years this will be discouraged by the difficulty of locating references, as discussed in the preceding section. In the long run it will be superseded by the greater ease of obtaining the data themselves from the data file, without going to the source documents. In some cases, however, it will remain desirable to furnish hard copies, in particular where graphical information is involved; phase diagrams, contour lines of electron density obtained from diffraction patterns, shapes of spectral lines. Even though graphical information will become less popular as numerical data becomes more readily accessible—at present some graphs serve merely as a convenient condensed representation of numbers—a hard core of demand for copies of graphs will persist. To this must be added requests for copies of entire tables, sometimes several pages in length.

At present the most economical way to produce copies is the Xerox process. This presupposes that a hard copy is on file in ISO (or less conveniently, in the NBS library). It requires no further investment.

Let us briefly discuss some of the existing methods of partial mechanization which might be invoked to relieve any developing bottlenecks. They can be characterized as micro-optical systems. The first to come to mind is microfilm. Its principal advantage is to reduce the physical size of the library. It also tends to speed up the production of copies. It has, for the application here discussed, the overriding drawback that individual information items cannot easily be corrected, inserted or deleted; the only way to do this is to remake an entire reel of film. We are likely to be faced with numerous cases where a single number in a table, or a single graph out of a set of charts, has to be revised. Since a space shortage for the storing of documents is not likely to be critical for some time, microfilm at present does not appear as a promising prospect.

Microfiche, microcards, and similar systems are easily updated. The saving in space is less pronounced than with microfilm. Handling is often re difficult than with either film or hard copy, there are systems with automatic selection of rds, where handling is no problem. If at some ure date the volume of library holdings and of ly transactions requiring hard copy becomes too ge for manual operation, such a system of crofiche or microcards holds promise.

Before introducing it one should investigate, es- cially in view of the cost of transcribing an en- e document collection to the new system, how it n be coordinated with the digital computer oper- on envisioned for the future. It would be un- tunate to be saddled with a semimanual system copying graphical information which is in- mpatible with the central digital computer sys- n for the handling of numerical information. the other hand, with some development work night be possible to have an integrated system, a microfiche selector which is connected to the gital computer, receives from the computer the ial numbers of items to be copied, automatically rches for these items and copies them, together th identifying information supplied by the com- ter (e.g., serial number of the inquiry).

On the other hand, it may turn out that a curve otter directed by the computer is a more econom- l way of reproducing graphical information. is possibility should be examined before a cro-optical system is proposed for installation. mpare, in this connection, the discussion of in- t, transmission and storage in sections 2.3.4 and .3. above. To date, experiments with digitally produced images of spectra have not been suc- sful, but the door is not closed.

The furnishing of hard copy may be required t only in reply to inquiries but also as a service data centers, in cases where the latter do not ve the facilities for obtaining such copies ectly.

3.2.5. Data Service

It is expected that ISO will be able from the rt to respond to a number of inquiries by di- ctly furnishing the desired data, from its own ta file. As stated above, such information uld be accompanied by a suitable disclaimer ting that the data have not been evaluated for iability and therefore do not constitute stand- l reference data. As technical centers are estab- hed or integrated into NSRDS, as criteria for aluation are set up and the evaluation of data undertaken, ISO must incorporate the results such evaluation into its files.

One way to do this is to set up a file of evalua- e comments which is arranged in the same order the data file itself (cf. sec. 3.4.3 below) and in ich every comment is cross-referenced to the propriate data file item. When an inquiry mes in, the subject is first looked up in this com- nts file. If a comment report is found there, s report, together with the literature items oss-referenced by it and the data given in these ms, furnish the material for the reply. If no mment is in the file, the data are looked up in the data file and used for a reply with a disclaimer as described above.

A more radical approach would be to segregate the entire ISO library into two parts, SRD and other data. This would save one look-up step for SRD but would require frequent rearranging of the collection, possibly affecting different parts of the same publication in different ways. Also this approach would probably oversimplify the problem: quite possibly the result of data evaluation will not be a simple dichotomy into SRD and other data but a more detailed qualitative description of the worth of the data.

In the beginning, as we said above, all technical responses issued by ISO should be checked by technical specialists. Gradually, ISO will learn by experience that certain routinely recurring types of questions do not require such checking. At first the burden of checking will have to be borne by the technical area managers of OSRD and by certain specialists elsewhere in NBS. If this load gets too heavy because the volume of inquiries grows faster than ISO's ability to handle them without assistance, one of two courses are open: either ISO acquires technically competent staff of its own, or the flow of inquiries is restricted by answering only in terms of standard reference data.

3.3. Publication Services

3.3.1. Types of Services

In contrast to the inquiry services discussed in the preceding sections, which are unscheduled and are undertaken on receipt of requests for them, the editorial and publication activities are scheduled by OSRD on its own initiative. We may distinguish periodical and aperiodical publications.

The principal output not only of OSRD but of the entire NSRDS will be monographs, especially data compilations and evaluations. These can take different forms. There is first of all the "National Standard Reference Data Series" of the National Bureau of Standards, published by the Government Printing Office, of which several numbers have already appeared. The series will contain tables of data compiled and evaluated under the auspices of OSRD and related material. It is intended to supplement, rather than supplant, the publication activities of technical data centers and other interested organizations. Thus the NSRD Series will have for primary subjects those compilations produced at NBS, or by organizations which for some reason or other cannot undertake publication, or those for which there is no appropriate technical data center, as well as state of art reports, lists of compilations considered to be standard reference data, reports on classification, indexing, mechanization, and other topics of interest to data compilers, evaluators and users in general.

In addition to producing the NSRD Series, OSRD will publish data (usually taken from monographs) in loose-leaf form, machine-readable media such as tapes or punched cards, or other formats which prove to be widely useful.

OSRD will also endeavor to encourage and assist data centers and other organizations in publishing their results, especially by providing editorial help, advice on mechanization and occasionally financing, especially in situations where such assistance will make the difference between prompt publication and long delay.

As for periodical publication, it may be useful to publish a news or current awareness service, concerned with events in the field of data on properties of materials. New data compilations which have been published, projects undertaken or completed, contracts awarded, new mechanization techniques, etc., would be listed. Undoubtedly there is plenty of material which is of interest—currently OSRD writes unpublished reports on its own activities alone, running to several pages per month—but to collect this material from the many organizations involved would require a considerable effort and should be done only if there is a clear need for it. Some of this information appears in subject-oriented publications; perhaps in time the activity of data compilation and critical evaluation will come to be considered as a field of technical specialization in its own right, and will increase the demand for a periodical publication service of this kind.

Another possible activity, unquestionably useful but requiring an even greater editorial effort, would be a current bibliography on data compilation, evaluation and perhaps generation; i.e., a periodical listing of new published papers and perhaps unpublished reports on these subjects, giving at least the bibliographic description (author, title, place of publication) and perhaps also abstract, critical review, and/or listing of references—the latter for use in connection with a citation index and bibliographic coupling.

OSRD publications, especially those in print or report form, will be distributed by GPO, the Clearinghouse for Federal Technical Information, or other appropriate agencies, rather than by OSRD itself.

3.3.2. Preparing for Mechanization

There are several steps which OSRD has taken or plans to take in the near future in order to assist in the transition to mechanized publication described in section 2.2.3.

The first of these is the acquisition of a linofilm keyboard, which produces the 15-hole punched paper tape needed to drive the linofilm composition machine. One of the advantages of having this device at NBS is that keypunching can now be done under the direct supervision of the scientists responsible for the preparation of a manuscript. This makes it unnecessary to prepare the manuscript to the same degree of perfection as if it were sent to a printer; rather, pencilled corrections and verbal instructions to the keyboard operator are acceptable, and many questionable cases can be settled by discussion as they arise.

Another step in the same direction is the planne procurement of a modified tape typewriter whic will accept a number of special character insert (similar to the commercially available "Typits" and at the same time produce a punched pape tape. The choice of special characters, of differ ent type fonts and other features is far mor limited than with the linofilm keyboard, but th latter is more difficult to handle than the type writer and does not produce an immediatel available typed copy for quick proofreading Therefore, and also because of its lower cost, th tape typewriter is preferred for material of simpl typography.

Both recording methods can be aided by per forming some editing functions on a compute Several computer codes for such purposes hav been written; so far, each of these codes wa tailored to one particular publication. One of th efforts in which OSRD expects to engage in th near future is the production of more generall applicable computer codes for publication editing

An IBM Document Writer has been acquired b one of the data centers located at NBS and being used with great success for material of inter mediate typographical quality.

3.4. Preparatory Activities

3.4.1. Information Gathering

As stated before, the initial period of operatio finds OSRD with inadequate data on both th need for its services and the tools available fo rendering them. One of the first tasks is to ac quire some information on these two problems.

In regard to need for services, an attempt ha been made to survey the field by means of a sho questionnaire sent initially to all members of th American Chemical Society, and later, if this found desirable, to other interested groups. Th questionnaire will attempt to ascertain whic properties of materials are most often sought i the literature, how well the existing literatu satisfies this need, which data compilations a most often consulted, for which properties com pilations need to be prepared and data evaluate It will also try to discover existing or incipie data compilations undertaken by individu scientists and not widely known. This informa tion should assist OSRD in setting priorities an distributing funds among data compilation an evaluation projects; at the same time it shou indicate to ISO what kind of demand for its ser ices is to be expected, and it may point to som existing compilations which should be added ISO's library. Preliminary indications are th the number of obscure compilers to be discovere in this way is substantial.

Another questionnaire, with a small distributi list, will ascertain the characteristics of the know major data centers and similar organization nature, volume, and format of the data they pr duce, store, and distribute; policy in regard answering inquiries; and funding.

Even with all these attempts it is unlikely that ıe full impact of NSRDS on the technical com- unity can be foreseen by an information- athering activity undertaken ahead of time. either the prospective users nor the producers nd distributors of standard reference data are kely to anticipate the changes in project organi- ation and working habits which are potential sults of this "information revolution." When n engineer or scientist can get information on roperties of materials by turning a dial attached his desk—spending less effort than walking to is own bookcase and turning pages in a handbook, nd at the same time receiving answers which are etter evaluated and more up to date—his very hinking will be directed into different channels ways which we cannot foresee at present. Vhen electronic computers were first contem- lated, it was everyone's conviction that one na- onal computing center, or at most half a dozen egional centers, would satisfy all anticipated omputing needs; and no survey of prospective sers could have changed the picture. Similarly, order to insure full utilization of the potential enefits of NSRDS, it will be necessary to obtain continuous feedback—evaluation, complaints, nd new ideas from the users of the system; and it ill be necessary for the management of OSRD to ontinue to look, on its own initiative, for better ıethods and new applications.

3.4.2. Bibliography

A major bibliographic survey of existing com- ilations of quantitative data on physical and emical properties of materials is being under- ıken. It is expected that this survey will cover ll existing compilations of data in the areas out- ned above, whether published in the open litera- ıre or in report form; as far as possible it should lso include unpublished manuscript compilations. n general, the subject of the survey would be sec- ndary publications or manuscript collections, not e primary publications in which newly measured r calculated data are first communicated. A ood example of such a survey is furnished by the Index of Mathematical Tables" by A. Fletcher, . C. P. Miller, and A. Rosenhead. This index is listing of all mathematical tables whose existence e authors were able to ascertain. In the few ears of its existence it has become an indispens- ble reference work. It appears that the proposed urvey of data compilations should be published similar form.

It seems that the survey can best be undertaken y a joint effort of perhaps six to twelve leading cientists, each a recognized authority in one of he major subdivisions of physics and familiar ith all of the important people and projects in hat subdivision. Each of them will be respon- ible for one "chapter" of the entire compilation. here will be one central coordinator in charge of he entire project, whose job it will be to recruit hapter authors, delineate their fields of responsi- ility and set common standards for the chapters. The main responsibility of the chapter authors will be to know all the likely sources of compila- tions in their fields; the job of actually contacting these sources, verifying, and describing the extent of existing tabulations can be left to subordinates. Some of the chapter authors may, however, find that their fields are so large that they have to be subdivided into a number of smaller specialized areas, each with a separate "section author" who would again have to be a recognized authority in his field of specialization and conversant with all data compilation activities in that field. Ex- amples of chapter areas might be: nuclear struc- ture data, infrared spectra, x-ray diffraction pat- terns, other solid state properties, etc. These ex- amples are illustrative only; it should be left to the judgment of the coordinator to delimit the chapters.

The survey will not have to start entirely from scratch. The Office of Critical Tables of the NAS–NRC has collected a list of some of the best known data compilations, and is continuing this work on a small scale. OSRD has on its own as- sembled a modest collection of compilations. These two collections could be used as starting points. There also exist surveys in some special- ized fields: for instance, the Nuclear Data Project at Oak Ridge has made a survey of compilations of nuclear data; it is not yet complete, but the work is being continued.

The publication of such a survey, apart from its importance for the National Standard Reference Data System, would be a most valuable addition to the technical literature and extremely useful to the scientific and technical community.

3.4.3. Classification

The most important preparation for the opera- tion of ISO, apart from gathering the information on which it is based, is the organization of this in- formation. The bibliographic survey and ques- tionnaires discussed in the previous section will result in a list of data compilations. Copies of these will be acquired by ISO (many of them are undoubtedly already in their collection) to be used in their inquiry-answering service. It now be- comes mandatory to devise a system of organizing the collection in such a way that any desired item in it can be quickly located. This problem occurs in every library, and conventional solutions are the first to be considered.

Books or documents in a library are located primarily by two devices: systematic classification and subject indexing. These two approaches are discussed in this and the next section.

Classification begins with the selection of a clas- sification system. It has been estimated that a substantial part of the literature in the physical sciences—probably between 20 and 50 percent of all papers published—is concerned with data on properties of materials. This suggests that one should start by looking at existing classification systems for the physical sciences as a whole. Sev- eral such systems are in widespread use: Universal

Decimal (UDC), Library of Congress (LC), a system used by Physics Abstracts, etc. It was decided at an early stage that none of those could be used without change, mostly because they are obsolete. It therefore became desirable to design a new classification system, made specially for the needs of NSRDS, and hope that it could somehow be made compatible with the older and broader systems. The price we pay for having our own system is that we have to do all the classifying, while with a general-purpose system one might have hoped to leave this job in many instances to others. This is a small effort for the present collection of compilations, but a much larger one for the current literature.

Since the proposed classification system is to serve the needs of a large segment of the scientific community, it would be desirable to have it agreed upon by general consensus and in cooperation with the scientific societies; indeed, international uniformity would be most welcome. This, however, would take years to achieve. It therefore becomes expedient to proceed in two directions simultaneously; toward the long-range goal of a broadly based, cooperatively designed classification which can be used on a large part of the literature and is convertible to the older classifications to a reasonable degree; and toward a short-term objective of a classification adequate for the internal operation of ISO during the next few years.

In pursuit of the long-range solution, OSRD sponsored two pilot efforts. One proceeded empirically, in line with modern trends in documentation theory, by collecting statistics on such features as overlap in the vocabulary of pairs of documents, and attempting to derive clusters of documents which ought to have fallen into the same category of the sought-after classification. On completion of this study the results were not judged to be sufficiently promising to warrant continued effort. The second long-range study employed the conventional approach of selecting prominent technical attributes as a basis for classification, but differed from older attempts by its use of the most modern concepts of theoretical physics. Further development along these lines appeared promising but would have required more effort than it was possible at the time to devote to this part of the program. At the same time, some features of this approach proved adaptable to the short-term study being conducted in parallel.

This short-term effort had started while the two experimental long-range studies were in progress, and was completed in the main, except for some details, a few months later. It resulted in a classification of properties developed by OSRD which is currently being used in the operation of the ISO data file. It is strictly limited to physical properties of materials; it avoids using as a basis for any stage of the classification either materials, groups of materials or any other concept. (For instance, "thermal conductivity" appears as a category, but "thermal conductivity of aluminum," or ". . . of metals" or ". . . at low temperatures" ar not categories.) It consists of a few hundre classes, which have proved to be amply adequat for classifying the present library of ISO and wil undoubtedly be adequate for any foreseeable expansion of it. If it were used, e.g., for classifyin all scientific papers dealing with properties of materials, there would probably appear some unmanageably large classes, which would require furthe refinement of the classification.

Once the classification had been designed, th next step was to assign the volumes of the IS(library to classes. This step has been complete for the present holdings but will have to be continued for future acquisitions. It is the success o this operation which constitutes the "proof of th pudding" for the classification system adopted.

The library contains a certain number of documents which are not primarily tables of data These fall outside the classification used. The are kept separately from the main collection an are arranged in groups according to a simple a hoc classification designed for the purpose.

Finally, the books are shelved in accordanc with the classification. This step would not b entirely necessary—conceivably the books could b kept in any order, e.g., by accession number, an a card file be used to locate the volumes bearing desired classification number—but for the manua operation of a small library such as ISO's shelvin by classification number has a number of well recognized advantages.

3.4.4. Indexing and Abstracting

Conventional libraries rely on author and subject indexing, in addition to classification, as principal means of retrieving information.

The preparation of an index card file arrange by authors, separately for personal and corporat authors, is straightforward. For subject indexin there are two methods: one uses derived inde terms, the other uses assigned ones.

Derived index terms are a recent product (modern documentation theory. They are usuall based on statistical analyses of the vocabulary i the document. They have the virtue that they ca be produced by computers, or at least by unskille personnel. These methods are still being devel oped, and have not been shown to be clearly su cessful. There is no more reason—in fact, prob ably less reason—to expect them to be successfu in the field of data on properties of materials tha in other fields.

Assigned index terms for a document are chose by a person who has at least scanned the documen is at least superficially familiar with the subje matter, and makes a decision as to the subje headings under which a user might expect to lo for this document. There are two systems f doing this. In one the indexer (who in this ca is often the author himself) is free to choose ar terms that occur to him. In the other, terms a taken from a master list or "thesaurus" if possibl If the thesaurus contains no suitable term, tl indexer may assign a new term and add it to the thesaurus.

This system, then, requires two steps: first, the building up of a thesaurus, and second, the application of this to a given set of documents. The thesaurus should be generously cross-referenced for synonyms and for inclusion relations ("see" and "see also" references). The difference between the two systems is smaller than might appear: a thesaurus-controlled approach can be handled so liberally that the indexer is in fact free to use any term that occurs to him, and the uncontrolled approach can be augmented by maintaining an alphabetical listing of all index terms used. There remains the difference that in the thesaurus approach the indexer has the responsibility to search, before using any term, for synonymous, more general, and more special terms already occurring in the thesaurus. This task is facilitated if the thesaurus is maintained not only as an alphabetical listing but also in a systematic hierarchical arrangement—thus forming a bridge between indexing and classification.

For reasons which can only be partly detailed here, we believe that a thesaurus-controlled approach to indexing should be taken by ISO. Furthermore, a quite small thesaurus would be adequate for most purposes. In many instances the classification according to physical properties will alone be sufficient to locate desired items of information, without using the index at all. Indexing by materials would be useful but, because of the large number of possible terms, too cumbersome at least in the beginning. It is suggested that a small thesaurus be put together from the names of common classes of materials (e.g., acids, oxides, alcohols, cyclic compounds), common designations of parameter ranges or values (e.g., low-temperature, high-pressure, critical), and a few other terms expected to be useful (e.g., catalysts, refractories, dielectrics). The existing library of ISO should then be tentatively indexed on these terms. This process will result in suggestions for additional index terms, which should be added to the thesaurus and used in a second round of indexing the collection.

A further step in the intellectual organization of the ISO library, after classification and indexing, consists in abstracting. In other environments the value of abstracts lies in the wide circulation which they can be given. In the case of ISO, with its tightly knit organization, this is a minor advantage; it is almost as easy for the staff to work with the documents themselves as with a set of abstracts cards. Perhaps the chief gain accruing from abstracting is that the process will systematically familiarize the staff with the contents of the library.

The information to be put on the abstract card for a document comes under the headings of properties, materials, parameters, and other information. The card should enumerate all properties of materials on which the document contains data. On the other hand, a complete enumeration of all materials referred to in the document is probably impractical; it will be advisable to list only major classes of materials (e.g., gases, metals, hydrocarbons). As for parameters, it will usually be sufficient to list the largest and smallest value of each parameter (temperature, pressure, etc.) for which the document gives data; to indicate other information or parameter values, such as the intervals at which functions are tabulated, would probably be superfluous detail. Finally, under "other information" the abstract card might list applications of the materials, instrumentation of measurement, any theoretical discussion given in the document, evaluation of the quality of data, etc.

In summary: When a request for data is received, the searcher will first ascertain which properties of materials are involved, and will find these properties in the hierarchical classification. This indicates a small group of documents in which the desired information should be looked for. He next consults the evaluations file, which may point to some data compilations of high quality. Along with specialized compilations, the large general data compilations, notably Landolt-Boernstein, must be examined. In doubtful cases, or when the request contains important qualifications which would narrow the search, the subject index can be consulted, which may reduce the number of documents to be searched. If desired, the abstract cards for these documents are looked up, and this may exclude further documents from the search. The remaining documents are then consulted in order to locate the desired information.

4. Conclusions

It emerges from the foregoing pages that during the next few years OSRD should vigorously pursue several activities in parallel. The availability of information services should be announced, questioners should be given the best service possible with the present manpower and information resources of ISO, all personnel of ISO should participate in this service and view it as an opportunity to become familiar with their subject. A large, but gradually decreasing, fraction of inquiries will have to be referred to experts chosen from area managers, data centers, NBS divisions and occasionally others. Where replies are prepared by ISO, all but administrative or routine ones should be checked by area managers or other NBS scientists. Replies should not be limited to the furnishing of standard reference data but would, in the absence of such data, supply literature references or data taken from the literature, accompanied by a suitable disclaimer. If, however, the volume of inquiries grows too fast for the present staff to handle, and if an expansion of the staff at that time is not feasible, then the flow of information would have to be restricted by limiting information services to those cases for which standard reference data are available.

In parallel with the foregoing, ISO should develop a thesaurus of subject index terms for its collection, apply it to the collection, and prepare abstracts of documents. Simultaneously it should broaden its collection in accordance with the results of the bibliographic survey of data compilations now started.

Further exploration is needed of the use of citation indexing and bibliographic coupling in literature searching; use of computers in editing and publishing; remote access to computers.

With these measures OSRD ought to operate satisfactorily for a few years; and meanwhile more information would accumulate on which the transition to mechanized operation could be based.

One other approach, however, appears desirable and should be undertaken simultaneously with all of the above. Rather than wait for several years and then transfer the entire operation to a computer in one move, one should select a segment of the operation which could be computerized before the rest, to serve as a proving ground. For several reasons, the area of thermodynamic properties appears to be an excellent candidate for this role, though atomic spectra or crystal data could be considered alternatively. In these areas there is a body of data already in machine-readable form or now being recorded in this form; and several data centers outside NBS, as well as competent scientists within NBS, have shown interest in such a development.

In the near future we propose to identify a subset of data to serve as a basis for the development of a set of computer codes for retrieval and updating. After such codes are developed—which will take a good deal of time—they should be used for six to twelve months in parallel with manual methods. Thereafter their use could be extended to the entire technical area of which the pilot study was a prototype, and a little later the manual operation for this area could be discontinued. Only then would the time have come to look for the memory component and other special computer features needed in the eventual mechanized operation.

U.S. GOVERNMENT PRINTING OFFICE: 1966 O—219-580

Announcement of New Publications on Standard Reference Data

Superintendent of Documents,
Government Printing Office,
Washington, D.C. 20402

Dear Sir:

Please add my name to the announcement list of new publications to be issued in the series: National Standard Reference Data Series—National Bureau of Standards.

Name ______________________________

Company ______________________________

Address ______________________________

City ____________ State ____________ Zip Code ____________

(Notification Key N337)

THE NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. Its responsibilities include development and maintenance of the national standards of measurement, and the provisions of means for making measurements consistent with those standards; determination of physical constants and properties of materials; development of methods for testing materials, mechanisms, and structures, and making such tests as may be necessary, particularly for government agencies; coöperation in the establishment of standard practices for incorporation in codes and specifications; advisory service to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; assistance to industry, business, and consumers in the development and acceptance of commercial standards and simplified trade practice recommendations; administration of programs in cooperation with United States business groups and standards organizations for the development of international standards of practice; and maintenance of a clearinghouse for the collection and dissemination of scientific, technical, and engineering information. The scope of the Bureau's activities is suggested in the following listing of its three Institutes and their organizational units.

Institute for Basic Standards. Applied Mathematics. Electricity. Metrology. Mechanics. Heat. Atomic Physics. Physical Chemistry. Laboratory Astrophysics.* Radiation Physics. Radio Standards Laboratory:* Radio Standards Physics; Radio Standards Engineering. Office of Standard Reference Data.

Institute for Materials Research. Analytical Chemistry. Polymers. Metallurgy. Inorganic Materials. Reactor Radiations. Cryogenics.* Materials Evaluation Laboratory. Office of Standard Reference Materials.

Institute for Applied Technology. Building Research. Information Technology. Performance Test Development. Electronic Instrumentation. Textile and Apparel Technology Center. Technical Analysis. Office of Weights and Measures. Office of Engineering Standards. Office of Invention and Innovation. Office of Technical Resources. Clearinghouse for Federal Scientific and Technical Information.**

---

*Located at Boulder, Colorado, 80301.
**Located at 5285 Port Royal Road, Springfield, Virginia, 22151.

U.S. DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20230

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE