INIS-mf--13438

11th International CODATA Conference
Karlsruhe, FRG, 26-29 September 1988

STANDARDISING EXCHANGE FORMATS

H.D. Lemmel and J.J. Schmidt
International Atomic Energy Agency
Nuclear Data Section
Vienna, Austria

Abstract: An international network of co-operating data centres is described who maintain identical data bases which are simultaneously updated by an agreed data exchange procedure. The agreement covers "data exchange formats" which are compatible to the centres' internal data storage and retrieval systems which remain different, optimized at each centre to the available computer facilities and to the needs of the data users. Essential condition for the data exchange is an agreement on common procedures for the data compilation, including critical data analysis and validation. The systems described ("EXFOR", "ENDF", "CINDA") are used for "nuclear reaction data", but the principles used for data compilation and exchange should be valid also for other data types.

Introduction

In view of the vast amount of numerical data needed in the development of many fields of contemporary science and technology, increased efforts are needed to compile data, to review the data critically, and to make them available in convenient formats to those who need them for projects in pure and applied sciences. The amounts of data make it necessary to provide for an international co-ordination of discipline-oriented specialized data centres who share the workload of data compilation and analysis. Such a worksharing of data centres is possible only if they agree on standardized formats for the exchange of data with the goal that data compiled and validated somewhere in the world become easily available to all potential users in all countries.

In the traditional way of data compilation, a scientist had a private data collection supporting his own research, which he eventually published in a journal or in a handbook. In limited fields such personal efforts may still be possible. But in a longterm view such efforts are in vain, because private data compilations die away when their author changes to other duties or retires. Only if he uses for his compilation an internationally accepted data file format, his efforts will survive and contribute to an international database.

Longterm continuity in the maintenance of scientific data bases can be guaranteed only by data centres with longterm financial and administrative support, and by international co-ordination of data centres, who will jointly not only develop file formats for the exchange of data but who will also agree on high-standard procedures for reliable data compilation, data analysis and validation, as well as services to data centre customers based on a careful analysis of data needs and accuracy requirements.

I will report on an international data exchange system which is in successful operation now since about 20 years, and which has lead to an international database consisting of several general and specialized data files of, together, about six million records or five hundred megabytes, available worldwide to any scientist from any of the co-operating data centres.

I am working in the International Atomic Energy Agency in Vienna, Austria. The mandate of this Agency is to promote the peaceful and safe uses of atomic energy and nuclear sciences. An essential function within this mandate is the international exchange of scientific information, with a special emphasis that may be described by the terms "east-west exchange" and "north-south exchange". In this spirit, a nuclear data programme was started in 1964 which included a Nuclear Data Centre operated by the IAEA Nuclear Data Section. Out of our various activities I wish to report on our data exchange system in the field of "nuclear reaction data". These data are used primarily and originally for the input to computer codes for the design and operation of nuclear power reactors. But the same data are used also for various other applications in science and technology, so that we have a world-wide users' community of several thousands of addresses.

I cannot avoid mentioning various details that are specific to our topic of nuclear data, but I will draw conclusions that should have general validity for various other data types.

The nuclear data exchange system

Let me start with a summary of our exchange system for nuclear reaction data.

The IAEA Nuclear Data Section co-ordinates about a dozen of data centres that compile data, validate data, exchange data, and provide services to customers. Some centres participate with the full data scope, others only for a defined subset. The centres jointly operate several systems

- an exchange system "EXFOR"[1] for the exchange of computer files of experimental data;

- another exchange system "ENDF"[2] for the exchange of evaluated data; (the difference will be explained later);

- an index file "CINDA"[3] which serves as a bibliography and as an index to the EXFOR and ENDF data.

In addition we have several peripheral systems such as a "Dictionary" file[4] which serves as glossary of the agreed keywords and abbreviations used in the computer files; a system called "WRENDA"[5] which contains a list of data to be measured with increased precision in case that the presently achieved accuracy does not satisfy the user needs; a "PROFILE" system containing addresses and scientific interest of data users, data measurers, and data evaluators; newsletters[6] to establish the contacts between data centres and customers; etc. Last but not least we hold scientific meetings on nuclear data topics of special interest.

All activities are guided by the International Nuclear Data Committee with members chosen carefully from the main countries that have nuclear data activities.

By means of the above mentioned systems, the world's nuclear reaction data are speedily compiled, exchanged among the centres, and made available by the centres to users in all countries.

Each participating centre and each data user is free to use these data files for its own purposes, or to derive from the files additional products such as handbooks[7,8] or special data files required for a particular technology.

What are "nuclear reaction data"?

Before talking about our data exchange system I must mention briefly the special features of nuclear reaction data. In short, these data describe "the interaction of radiation with atomic nuclei". But these interactions can be quite complex. The radiations considered are: neutrons, photons, protons, α-particles, other charged particles, heavy ions. Of the known ≃ 2750 elements, isotopes, isomers, about 500 nuclides are of practical interest. For each pair of incident radiation and reacting nuclide the following data must be known:

What reactions occur? There may be three or many more reactions in parallel.
What is the probability that a specific reaction occurs? This probability is called "cross-section", measured in "barns".
What are the properties of the reaction products?
What products? How many?
What secondary radiations are produced? What are their energy spectra and angular distributions?

All these items are highly dependent on the velocity of the incident projectiles, particularly in the case of neutrons. Neutrons that may be produced in reactors or other research facilities, have velocities in the range from some 100 m/s up to almost the velocity of light or, expressed in terms of kinetic energy, from about 10^{-3} to about 10^{8} eV, that is 11 decades! The real problem is, that a change in the kinetic energy by one percent only may change the reaction rate by a factor of 10 or more, as shown in fig. 1 illustrating the "resonance structure" of neutron reaction data.

Consequently, the description of the interaction of neutrons with one nuclide alone requires a data file of many thousand data points, and these data, first of all, must be determined in many expensive experiments. The large volume of the data needed and the high costs of obtaining the data, require international co-ordination.

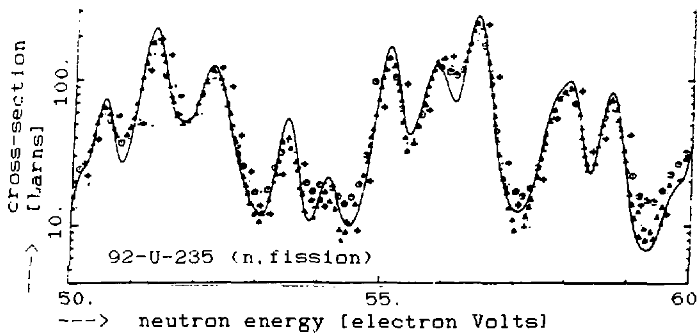

Fig. 1. The neutron-induced fission cross-section for uranium-235 in the small neutron-energy range between 50 and 60 eV (taken from the Japanese report JAERI-M-9823). The solid line shows the "evaluated data" produced by the data evaluator from several sets of "experimental data" indicated by different symbols. The "resonance structure" of the data requires very large data files to describe the reaction for the complete range of neutron energy from milli-eV to Mega-eV.

Coordination of data centres

In 1964, when the nuclear data programme of the IAEA was started, there were three national or regional nuclear data centres: one in the USA at the Brookhaven National Laboratory ("Sigma Center", now "National Nuclear Data Center"), one in the USSR at the Fiziko-Energeticheskij Institut in Obninsk ("Centr po Jadernym Dannym"), and one at the OECD Nuclear Energy Agency in Saclay near Paris ("Nuclear Data Compilation Centre", now "NEA Data Bank"). The IAEA Nuclear Data Section got a double function: firstly to act as a fourth data centre for all countries not served by the above three centres, and secondly to co-ordinate the activities of the centres.

By and by, a data exchange system was established in the way that each centre served a defined geographical area with respect to compilation of data and services to customers. Between the centres, all collected data were exchanged such that the four centres had the same data files. Consequently, data users in all countries had access to the world's nuclear data files, through one of the four centres. This is illustrated in Fig. 2[11].

Whereas the original four centres concentrated on the most important nuclear data, essentially neutron cross-sections and related parameters, other centres and groups joined the network, who used the same data exchange system for other specialized data types. These are the Photonuclear Data Center in Moscow, and several centres and groups devoted to charged-particle induced nuclear reaction data at the Kernforschungszentrum Karlsruhe, at the Kurchatov Institute in Moscow and at two Universities in Japan. More recently, the Chinese Nuclear Data Center in Beijing joined. Other centres, in particular the Nuclear Data Centre of the Japanese Atomic Energy Research Institute, or the nuclear data group of the Lawrence Livermore National Laboratory in USA, primarily contribute evaluated data files. See the Appendix for a list of nuclear reaction data centres.

It is an interesting question whether, for a specific field of data, a centralized effort would be more efficient, or a decentralized effort like that of the four nuclear data centres. For important reasons the decentralized arrangement is preferable. Each data centre must maintain close contacts with the authors of the data and with the users of the data. Obviously, these contacts will be different in, e.g., the USA where highly developed institutes are linked by computer telephone networks, or in the service area of the IAEA Nuclear Data Section which includes the developing countries. Here the data files cannot be transmitted through telephone lines but must be sent on magnetic tapes or diskettes. Also, the transmission of data must be supported by technical advice and training on the know-how of computer processing of the data.

Creation of the exchange format EXFOR

In the beginning, each of the centres used its own data storage and retrieval system. The systems were incompatible, and the data could not automatically be converted from the one system to the other. None of the systems was suitable to serve as an international system. A new system had to be designed.

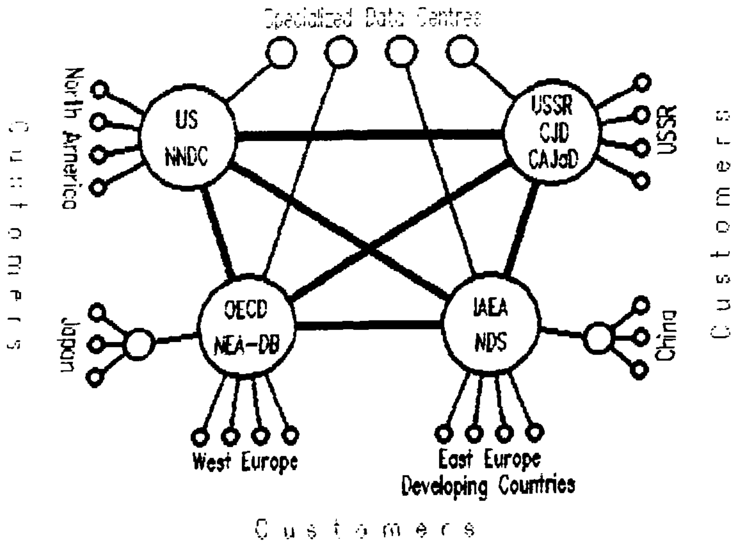

Fig. 2. Simplified scheme of the data exchange between the four co-operating data centres who jointly build up the international database. Their customers include data producers and data users, as well as national sub-centres who may distribute data further. Other specialized data centres feed their data into the exchange system through one of the main centres.

In the negotiations between the data centres it became clear that it was not possible to create a single system to be operated as data storage and retrieval system at each of the centres. The computer facilities available to the centres were too different. For example, the quite large size of the database permitted only one or two of the centres to have the entire database in direct access, whereas the other centres could have only a data index in direct access on a disk file whereas the bulk of the database was stored on a set of magnetic tapes. Also the operating systems of the computers were too unequal to permit that identical systems were used at all centres.

The solution was that the centres did not agree on a joint database system but that they agreed on a joint format for the centre-to-centre exchange of data on magnetic tapes.

The exchange format, called EXFOR, was adopted in 1969[9]. It was agreed that each of the co-operating centres would have its own internal system for storage, retrieval and processing of data, but that each of these internal systems had to be compatible to the data exchange format EXFOR, with easy conversion routines in both directions.

The technical preconditions were solved when the USSR Nuclear Data Center in Obninsk linked an IBM tape unit to their computer of a Russian type. After the exchange of some trial tapes, the centre-to-centre data exchange was agreed and started in 1970. Each data centre compiled the data from its own defined geographical service area, converted its old data files to the new format, and transmitted EXFOR tapes to the other centres in monthly or quarterly intervals. Subsequently, other data centres or groups joined the network so that the world's nuclear reaction data of about 60 000 data sets in 4.5 million records are now available to those who need the data, in EXFOR format or EXFOR-compatible formats.

The task for each of the co-operating data centres can be defined in different terms. The original four centres continue to share their workload by geographical criteria. Other centres use physical criteria; they compile the data of a specific subfield of nuclear reaction data, such as photonuclear data, or data for the production of radioisotopes for medical applications, or nuclear fission-product yield data. Yet all data are sent to the other centres in EXFOR format. Only some of the centres are specialized in services to customers. Others are contributing their compilations and using the database for their own purposes but are not involved in customer services.

Obviously, it requires some control that each of the co-operating centres observes the EXFOR rules strictly. Once a year the IAEA Nuclear Data Section convenes a meeting of representatives of the co-operating centres. New compilation rules are discussed and agreed; the EXFOR Manual is updated; Dictionaries (or glossaries) of agreed abbreviations are maintained and additions to them are agreed; EXFOR check programs are developed in order to guarantee the formal and physical correctness of the data files, so that the data processing computer codes can function smoothly.

Experimental data - data evaluation - evaluated data

In the area of nuclear data, and certainly also in other data areas, one must distinguish between experimental data and evaluated data[10]. Let us analyze the difference, before going into some details of the EXFOR system, which had been designed for experimental data.

A data user requires reliable data over a complete range of parameters. The user wants "recommended evaluated data". In the case of neutron cross-sections, data are required over the full energy range from 10^{-3} eV to 10^{+7} eV or even higher up. The user wants these data in convenient tabulations in handbooks and computer files. Where do these data come from?

Despite significant progress in nuclear theory, all nuclear data must be determined experimentally. Nuclear physics is an area where accurate data measurements require very expensive research facilities (research reactors, nuclear particle accelerators, etc.), sophisticated detector systems (e.g. for the accurate determination of the energy of the neutrons), and isotopic samples of high chemical and isotopic purity. Stimulated by the increasing needs for nuclear data and increasing accuracy requirements, tremendous efforts have been undertaken in about 800 institutes in about 50 countries to measure nuclear data. Yet it lies in the nature of experiments, that a single experiment can cover only a limited range of measurement parameters (e.g. a small range of neutron energies), that the results suffer from experimental uncertainties, and that different experiments may be discrepant. Part of the uncertainties may be of statistical nature which can be estimated; but significant parts of the uncertainties may stem from unidentified systematic errors. Therefore, data must be measured not only by a single method; they should possibly be confirmed by different and independent measurement techniques.

To obtain from the experimental data, which are uncertain and discrepant, reliable data for the user, a process of data evaluation is needed. Data evaluators analyze the experimental uncertainties, try to reconcile discrepant data, fit the experimental data to theoretical models, use theoretical models to estimate data in ranges of parameters that are not accessible by experiments, and finally issue a recommended data set in a form which is convenient to the users. They also issue recommendations to experimenters as to what data should continue to be measured with increased accuracy.

Typically, an evaluated data set, e.g. for the fission cross-section of U-235 over the full energy range of practical interest, may be a data table of several thousand lines, which may have been derived from 100 sets of experimental data. This is only one reaction and one isotope, but there are many different reactions possible, and there are hundreds of isotopes of interest. Obviously, experimental data and evaluated data are so numerous that both must be stored in computerized media.

It is essential to realize that different data storage systems are required for experimental and evaluated data. In both cases it is not sufficient to store just the numerical values. Numbers are meaningless without a minimum of text information explaining how the data were obtained and giving an estimate on the data accuracy and reliability. Contents and format of the information to be stored together with the numerical data, are quite different for experimental and evaluated data.

Experimental details which are essential to be included in experimental data files, are of no interest to the user of evaluated data, who just wants best numbers, disregarding where they come from. But he also must know, how accurate the data are, how old the data

evaluation is, what theoretical model had been used in the evaluation; and this kind of information must be included in evaluated data files, partly as text information and partly in a computer-intelligible way.

In particular in the case of experimental data it is essential to store, with each data set, information with all details on the uncertainty analysis and on the experimental method how the data were obtained. One must also store with the experimental data set all assumptions used in the experiment: standard reference data to which the experimental data were normalized, assumed values of radioactive decay data of the materials used in the experiment, detector efficiencies assumed, and many other details. It is exactly these details which determine the reliability and qualification of the experiment when, at some later time, an evaluator will use the experimental results to produce a best recommended evaluated data set. Often enough it occurs, that in old publications there is insufficient information on the uncertainty analysis; in such case the data evaluator will give this experiment less weight in the evaluation than it might deserve. A measurement is in vain, unless a detailed error analysis is documented together with the measured data. From the well-documented uncertainties of the experimental data, the evaluator will estimate the uncertainty of the evaluated data set.

Data representation

An essential difference between experimental and evaluated data files is the data representation. A data user may wish to have, e.g., angular distributions always in the form of Legendre polynomials. But this is not the form in which experimental data are measured. Measurements are taken at discrete angles and with an angular resolution to be determined. Consequently, experimental data files must be very flexible to include the experimental data exactly as given by the author, including angular resolution and other parameters. Data in any representation and all data parameters must be acceptable in files for experimental data. Only in evaluated data files one may agree to give angular distributions exclusively in Legendre polynomials of a well defined type, and any other representation may be rejected. It is then the function of the data evaluator to produce the required Legendre fit from the available experimental data.

Another feature of experimental data is the fact, that many data are not measured as absolute values but rather as relative values or ratios. There are ratios of one data type against the other, or ratios of data of variable neutron energy versus a reference value of neutron energy. Consequently, a compilation system for experimental data must be flexible to include all kinds of ratio data and relative data. It is then up to the data evalutor to produce from these ratios the absolute data values required in the evaluated data file.

Critical data analysis and data validation

A centre-to-centre data exchange system cannot work without agreed procedures for critical data analysis and data validation. A data user should mistrust the data files unless he can be sure that the data were critically examined by the data compiler. Earlier data compilation systems failed due to lack of critical data analysis. Data were merely copied from the published literature, and this is insufficient. Old nuclear data files without critical analysis contained duplications, when the author published his results more than once, which is a frequent case. They contained contradictions when preliminary data and more final data were both included. They were incomplete, when the compilation was restricted to materials from "archival journals", whereas large data tables can usually be published, if at all, only in "non-archival" laboratory reports. Often, the journal publication will include only a small graph. Old data files were frequently inaccurate because too often they included data that were read from a graph; this is strictly forbidden in a reliable data compilation. In the unfortunate case that the original numbers were lost and only a graph is available, then the data compiled must clearly be labelled as "read from a curve", and the estimated accuracy of the curve-reading procedure must be specified.

The procedure of data compilation and validation as agreed among the co-operating centres typically functions as follows.

In the case of old data where the authors can no longer be reached, all printed material relevant to the experiment must be collected, including progress-reports, laboratory reports, journal articles, conference papers, in order to extract the necessary information. The compilation will include judgments by the compiler such as: "Data in reference 1 seem to be superseded by data in reference 2"; or "Warning: normalization is not up-to-date". Of course, such judgments must be labeled as "comment by the compiler".

In the more frequent case of recently measured data, the data centre will see from the progress-report of an institute that a data measurement is being performed. In this stage the centre should contact the author and ask him for early submission of the results. At latest when the author finalizes his manuscript for a publication, he should send his data to the data centre. The data centre will advise the author in the case that the data were not normalized to uptodate values of the standard reference data, or will ask the author for additional information when, e.g., the description of the uncertainty analysis was not detailed enough. Then the data set is compiled and transmitted to the other centres, typically based on a progress report and a private communication. In this stage the data set may be labeled as "preliminary". We regard it as essential to compile and transmit also, such preliminary data because, according to our experience, it is frequently the most recent experimental results that receive high interest in the data users' community. But it is essential that the system provides the mechanism of labeling a data set as "preliminary".

Thereafter the data centre prepares of this data set an "author's proof copy" which is sent to the author for comment and approval. The author will usually suggest some changes to the data entry. Meanwhile the journal publication may have appeared of which the centre adds to the compilation the reference and other pertinent information. With this, and with the changes proposed by the author, the data entry is transmitted a second time to the other centres. In this version the flag "preliminary" will have disappeared, and another label saying "data compilation approved by the author on ... (date)" will appear.

Later on, the author may present his work at a conference and, due to supplementary research, he may decide that his data must be corrected by one percent upwards, because a certain correction had been underestimated. The final corrected data remain unpublished but are sent to the data centre. Thus, the centre will transmit the same data set a third time, including the final data, and including reference to the conference paper.

It may happen that the author had done a similar experiment earlier, and that he determines that the new experiment fully supersedes the previous one. In this case, the compiled data of the previous experiment will be labeled as "superseded"; this way, the data set remains on file for archival purposes but will not show up in data retrievals, so that customers will receive, as far as possible, only valid information.

This is a realistic description of the data analysis and validation process that is required for obtaining a reliable and always uptodate database. It is by this procedure, and in particular by the principle of the "author's proof copy", that the authors themselves share the responsibility for the correct contents of the database, and that the compilation receives a status similar to that of a conventional publication medium. This is particular important at a time, when journals are no longer prepared to include large data tables.

The EXFOR agreement among the centres did not only include the agreement on the exchange formats but included also the agreement to do the compilation and critical data analysis in the way as just outlined. Consequently, the EXFOR database contains many data sets that are not included in any printed material, and also contains data sets giving unpublished revisions to published data. Thus, the EXFOR database has become an archival publication medium in addition to the conventional print media.

Requirements for exchange formats

In addition to exchanging the experimental results in standarized formats, and to validating the data by agreed procedures, the smooth operation of a decentralized data exchange system requires some careful bookkeeping methods.

Accession numbers. Each data set transmitted must be uniquely identifiable. Any updating and retransmission of a data set can be done only by identifying the data set by its accession number. Also in data indexes the data set is identified by its accession-number, e.g. "EXFOR-12345.003". The accession numbers do not carry any physical meaning. They are assigned in sequence of compilation. The first digit of the EXFOR number identifies the centre of origin. Only the centre of origin is permitted to update and retransmit the data set.

History. Each data set must have a history entry indicating when it was compiled, by whom, and from what source. For each update and retransmission a history entry is added, indicating the reason of updating. Accession number and last "history" date together identify a data set and its version uniquely.

Status. Each data set must have status information indicating whether it is preliminary, or final and approved by the authors, or whether it is superseded and, if so, by which more recent data set.

Flexibility. The exchange format must be flexible enough to include any result as given by the author, including unusual representations of data, of uncertainties and of related information. The compilation must look close to the representation of the results as given by the author, so that proof reading by the author is easily possible. Experience has shown that only in this way the number of compilation mistakes can be reduced to a minimun.

Computer processing. In contrast to the required flexibility of the exchange format, format and contents must, at the same time, be sufficiently well defined and identifiable as to make computer processing of the data and conversion of data to other representations and formats possible.

Simplicity and readibility. A centre-internal system may make use of all features of hardware and software available to the centre. To the contrary, a centre-to-centre exchange format must have simple file structure and should be eyeball readable, so that an exchange file received at a centre on magnetic tape, or through telephone transmission, or on a PC diskette, can easily be viewed on a screen. In a decentralized scheme, even a simple exchange format offers unavoidable traps where it is difficult to bring the staff of the co-operating centres to a common understanding of the compilation rules and definition of terms. Therefore, an exchange system should be kept as simple as possible.

The EXFOR system has 80-character records, using only a basic character set. Numbers must be FORTRAN readable. The text information is given in a balanced mixture of coded information and free text, structured by mnemonic keywords and abbreviations that are mostly self-explanatory. One can view the file as it is without much knowledge about the system.

Structure of EXFOR

The structure of EXFOR is illustrated in Fig. 3. An exchange file consists of a series of EXFOR "entries". An EXFOR entry is the compilation unit which contains the result of a given experiment at a given institute. If the result contains several data tables, the entry will consist of several "subentries". Entries and subentries are identified by accession-numbers consisting of entry number and subentry number, e.g. "EXFOR-12345.003". The first digit of the entry number shows the originating data centre. Digits 2 to 5 are a sequential numbering within the centre. "003" is the subentry number within the entry "12345".

A subentry consists of a text section, a section for constant parameters, and the data section containing the data table which is

```
ENTRY        12345       date

SUBENTRY     12345.001   date
TEXT         text information common to the entire ENTRY
AUTHOR       (A.B.AUTHOR, C.D.COAUTHOR)
INSTITUTE    (1USABNL)
REFERENCE    (C,88MITO,AB12,8805) SUMMARY OF WORK, GRAPH
             (J,NIM,123,45,8702) DESCRIPTION OF DETECTOR
             (R,BNL-12345,8712) TABLE, PARTLY SUPERSEDED      record
...          ...
             ...                                                iden-
HISTORY      data compiled when, by whom, from what source
             revised when, bey whom, for what reason             tifi-
STATUS       compilation approved by AUTHOR (date)
             ...                                                  ca-
CONSTANT PARAMETERS
             common to the entire ENTRY                            tion
ENDSUBENTRY

SUBENTRY     12345.002   date
TEXT         text information valid for this SUBENTRY only
             ...
REACTION     (92-U-235(N,F),,SIG) this defines the meaning of
               the numbers given in the DATA TABLE under the
               heading "DATA", whereas other column headings
               are self-explanatory (e.g. "EN" or "TEMP")
METHOD       brief description of experimental method
ERR-ANALYS   summary of error analysis
             ...
CONSTANT PARAMETERS
             valid for this subentry only
DATA TABLE
 ---------------------------------------------------------------
| HEADING | HEADING | DATA    | HEADING |         |         |
| UNIT    | UNIT    | UNIT    | UNIT    |         |         |
|---------|---------|---------|---------|---------|---------|
| ...     | ...     | ...     | ...     |         |         |
|    F O R T R A N            | ...     |         |         |
|        r e a d a b l e                |         |         |
|            n u m b e r s              |         |         |
|                                       |         |         |
|    may be a single line or more       |         |         |
|                than 1000 records      |         |         |
| ...     | ...     | ...     | ...     |         |         |
 ---------------------------------------------------------------
ENDSUBENTRY

SUBENTRY     12345.003   date
...          additional data from the same experiment
...
ENDSUBENTRY

...
ENDENTRY
```

Fig. 3. The structure of an EXFOR entry. The ENTRY consists of SUBENTRIES, and each SUBENTRY consists of sections for text, for constant parameters, and for the actual data table. The text information is identified by keywords, e.g. REACTION, which may be followed by free text (not computer-intelligible) and/or coded information (computer-intelligible), e.g. "92-U-235(N,F),,SIG" (=U-235 neutron-induced fission cross-section).

arranged in columns with headings defining the contents of the columns. The meaning of the data columns is not fixed but is defined for each subentry by column heading keywords. A data table may give data for a single variable, but may as well be multidimensional with several independent variables such as energy of incident radiation, energy and angle of the secondary radiation, plus sample properties such as temperature. A data table may be anything between a single value and a large table of thousands of records.

For example, in the first column the heading keywords EN and EV may indicate that this column contains the incident neutron energy given in units of electron volts. The second column may give the energy resolution ("RESL") in % ("PER-CENT"). The third column headed "DATA" may give the actual results of the experiment. The term "DATA" is further explained in the text section under the keyword REACTION, for example: "(92-U-235(N,F),,SIG)" which would mean that the cross-section ("sigma") for the neutron induced fission reaction "(N,F)" has been determined for the nuclide uranium-235. This is a simple example of the data definition. The string to be coded under the keyword REACTION has been designed such that all occurring nuclear data types can be defined in computer-intelligible way, including ratio data, complex definitions for multi-differential data, etc.

Actually, this coding of the definition of the measured data is the essential key for the entire system. A computer code for data retrieval must search primarily in the string of codes defining the data measured. There are continuing discussions among the data centres to standardize the coding of the data definitions, and to agree on new rules for new data types that had not yet been compiled earlier. A common understanding on the terminology and the coding of the data definitions is the essential condition for the successful operation of a data exchange system.

In addition to the data definition, other essential items to be entered in the text section of an EXFOR entry are:

Bibliographic information: authors, their institute and country, all biographic references relevant to the data set; identified by the keywords AUTHOR, INSTITUTE, REFERENCE.

Physics information under the keywords METHOD, DETECTOR, ANALYSIS, ERROR-ANALYSIS, and several others.

Standard reference data used, entered under the keywords MONITOR, DECAY-DATA and others.

Bookkeeping information as explained earlier, under the keywords STATUS and HISTORY.

For many of these items keywords and abbreviations are used that are collected in a separate Dictionary system agreed and exchanged between the centres. These include abbreviations used for the institutes (e.g. USABNL = Brookhaven National Laboratory, USA), journals (e.g PR/C = Physical Review, Part C), experimental devices (e.g. VDG = Van de Graaff accelerator). All these abbreviations are mnemonic so that they can be remembered easily. But, in addition, there are output formats in which all these codes are converted to normal language (e.g. PR/C to Phys. Rev. C).

Requirements for exchange formats of evaluated data

What I said above about the EXFOR system which was designed for experimental data, is more or less also valid for the exchange of evaluated data, however with one essential criterion in addition.

Evaluated nuclear data files primarily serve as input to computer codes used for power reactor design and operation, for shielding calculations, radiation dosimetry, nuclear activation analysis, or for other applications. Whereas experimental data require flexible data files with various alternate options of entering data in different representations, a flexible data file cannot be accepted as a basis for input to complex computer codes. Where different data representations are possible in principle, computer codes require a high standarization of data representation. Whereas eyeball readability was desirable in EXFOR, computer codes require a compact packing of data whereby the readability may be reduced.

Nevertheless, also in the case of evaluated data, the exchange format uses 80-character records in EBCDIC leaving it up to the centres or data users to convert this, e.g., into more compact binary data formats.

As in the case of EXFOR, also the evaluated data are accompanied by a minimum of text information including a summary of the data evaluation procedure and a history of revisions applied to the first issue of the evaluated data set. As in the case of EXFOR, each evaluated data set is identified by an accession-number plus the date of the last revision.

The ENDF system

For evaluated nuclear cross-section data the international exchange system is called ENDF ("evaluated nuclear data file"). Its origin is different from that of EXFOR.

In the 1960s there were several national evaluated data files in operation, primarily in UK, USA, FRG, Japan and USSR. They were similar in concept but different in details, so that a conversion from one into the other was not possible by a fully automatic procedure. The national efforts were parallel activities, and the results were difficult to intercompare due to the different data representations in the national files. For each nation it was difficult to give up its own data file format, because all the national computer codes for the power reactor design were based on this format.

After years of parallel efforts it became evident that the most powerful computer codes available were based on the ENDF format originating from the USA, and that these computer codes and the ENDF format were taken over by more and more other countries. Finally, the international Nuclear Data Committee recommended in 1984 that ENDF should be considered as the international exchange format for evaluated data.

Unlike EXFOR, there is not a single international database for

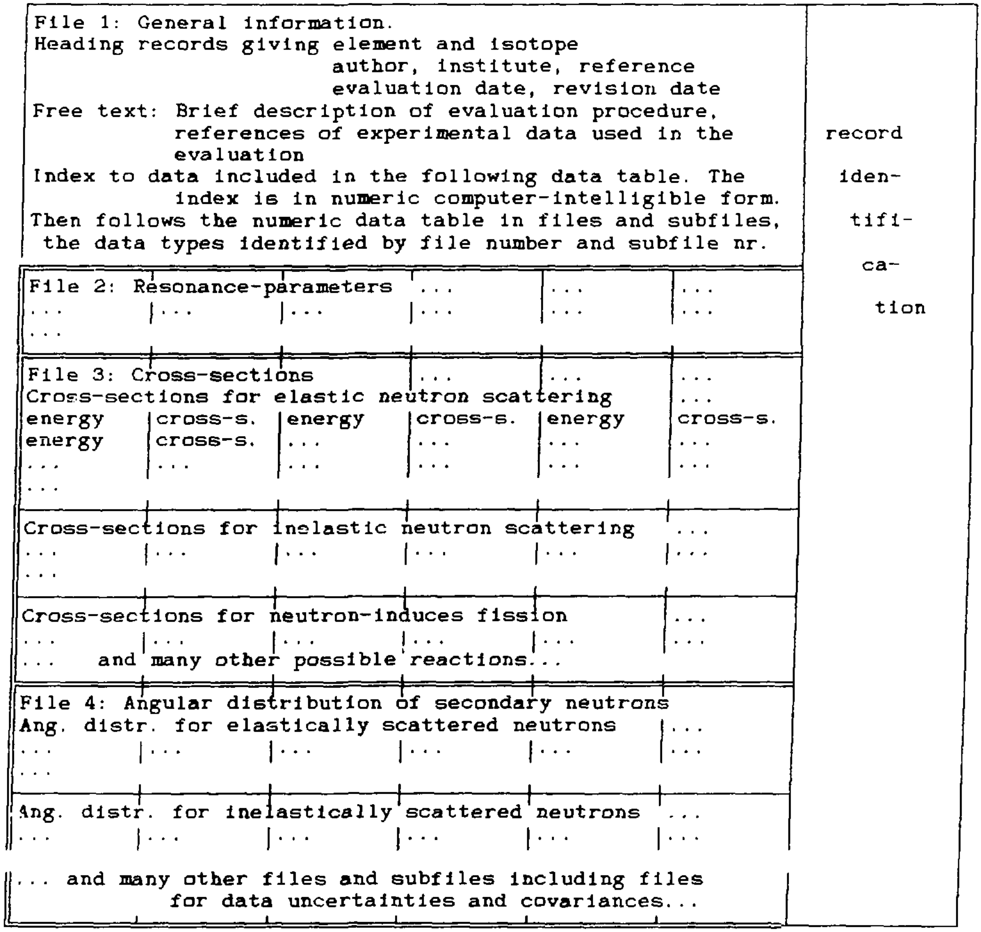

Fig. 4. Simplified structure of an ENDF entry. An entry contains all data for a given material (element-isotope), identified by a MAT-number. Each entry is subdivided into files, identified by MF-numbers, and subfiles identified by MT-numbers. Except for file 1 all information is computer-intelligible. Each possible data type is identified by a pair of MF/MT. In each record, MAT-MF-MT is given in the record identification field, so that a computer program can easily find the information which it requires. Each subfile is preceded by certain control numbers which may indicate, for example, whether linear or logarithmic interpolation must be used for energies or cross-section data.

evaluated data but there are several comprehensive files: ENDF/B[12] and ENDL[13] in USA, JEF[14] in West-European OECD countries, JENDL[15] in Japan, BROND[16] in the USSR; evaluations from other countries are collected by the IAEA in the INDL[17] file. These files are likely to continue as independent efforts; but they all use now the same data file format, so that the data can be exchanged and intercompared and the same computer codes can be used.

The files just mentioned are "general purpose" files. In addition, there are "special purpose" files containing only subsets of data needed for special purposes such as radiation dosimetry [17], nuclear activation[22], or nuclear standard reference data[23]. All use the same format so that the files can be processed by the same data processing computer codes.

A simplified illustration of the structure of ENDF is given in Fig. 4. The data set unit contains all nuclear reaction data for one isotope or element, called "material", identified by an accession number called "MAT-number". The data-types are identified by numerical codes which also determine format and meaning of the numerical data tables. A data set for a given "material" must always include the data for all possible reactions for the complete range of possible parameters, as determined by the requirements of the computer codes.

Together with the data files a program package is available for the processing of ENDF files [18].

It is not necessary here to go into further details. ENDF was designed primarily as input to computer codes; that it is used also as an international exchange format is, in this case, a secondary but very important aspect.

New trends

At present, version 6 of the ENDF format is under development[19]. Based on the analysis of data user needs, the new version will define formats for additional data types that were so far not yet accepted. It will include also agreed formats for the coding of data uncertainties and data correlations. This will enable reactor physicists to analyze how the results of their calculations are affected by nuclear data uncertainties. Such "sensitivity studies" must be accompanied by studies on the accuracy of the data processing computer codes which is another topic covered by the nuclear data centres[20].

In certain areas of nuclear data applications it has been proven that one obtains noticeably different results when correlations of data uncertainties were considered or ignored. This gives rise to new calculational procedures[21] which, in turn, require the inclusion of data covariances in the data files.

CINDA: Bibliography and data index

Before doing a comprehensive data compilation it is necessary to realize what data exist and to collect systematically all information about new data measurements. To this purpose a decentralized computerized system was developed with the name CINDA (computer index to neutron data). CINDA is similar to a bibliography, but it is different. Also this system is maintained in a decentralized scheme with an agreed exchange format. It functions as follows.

CINDA indexers in many countries scan all published literature and also unpublished materials such as preprints, progress reports and private communications. From this they prepare CINDA entries, send them to one of the four neutron data centres who maintain decentralized master files. These are simultaneously updated by means of a centre-to-centre exchange system in an agreed format. In each of the centres this file serves as the basis for the current awareness of new data determinations and hence as a basis for the data compilation. The accession-numbers of the compiled data are also included in CINDA.

The CINDA file, or a suitable retrieval from it, is regularly published in the form of a series of handbooks, which are essential information sources for the data users. From these books the data user can see, what data exist at all in his field of interest and whether these are available as computer files. Thus he is put into the position to ask the data centre well-founded questions and to request exactly those data sets or data files that will meet his needs.

CINDA originated from a private punched-card file. It was then internationalized, first in the OECD countries, then, through the IAEA, world-wide[24]. In 1967 it was one of the very first fully computerized international information exchange systems.

The printed CINDA books may be discontinued in the near future, because more and more CINDA users are getting online access to the CINDA file over telephone lines. The CINDA file presently indexes about 100 000 data sets. The file size of about 25 megabytes is a bit too large for an average size personal computer. But subsets from the CINDA file may well be disseminated on floppy diskettes.

The data scope of CINDA and EXFOR are not quite overlapping. CINDA includes data types that are not included in EXFOR. For practical reasons, a numerical data compilation can be complete only for data of high importance; the compilation of data of lower importance will be less complete due to manpower restrictions in the data centres' staff. A CINDA type index file, however, can be kept complete and up-to-date much easier. Actual, in our case, the international CINDA file existed several years before the agreement on the data exchange format EXFOR could be reached.

We are sometimes asked why the CINDA system is necessary when more general bibliographic systems exist such as the International Nuclear Information System, INIS. The answer is: The broad scope of INIS and the agreed descriptive terms of the INIS Thesaurus are too coarse for our purposes. The nuclear data users need certain details that cannot be retrieved from INIS.

CINDA is not a bibliographic file in the conventional meaning. CINDA is optimized for a specialized users group, of which the data needs are well known. The information in CINDA is structured such that the main questions asked by this specific users group can be answered without noise. The data classification scheme is not too coarse and not too fine. Of course, such a system which is designed for a specific users group, must be different from a broad scope bibliographic system. The unit of entry is not a given reference but rather a data set. A data set is defined by the target material (element, isotope) by the data type given in the data table (e.g. neutron-induced fission cross-section), the institute, and the year(s) of measurement. For each data set such defined, the following information is included in compact form: all published or unpublished references, name of the main author to be contacted, parameters of the data set (e.g. energy of neutrons used in the experiment), the accession-number(s) of the data set(s) in EXFOR, the size of the data set, plus some free text comment which may contain information on the accuracy of the data, on the measurement method, or any other item that may be essential to the data users, in particular a tag indicating whether the data are experimental or evaluated. When necessary, one will find a warning that data given in that or that reference, are superseded, so that possibly only valid data sets will be used.

Conclusions

For a worldwide information system, a decentralized organization with a network of regional data centres appears to be most suitable to fulfil the user needs in a convenient manner. A decentralized system is suitable to establish close contacts to the data producers and to the data users.

Condition for a decentralized system is an agreed data exchange format. Each of the co-operating centres is free to develop its own data storage and retrieval system, optimized to its computer facilities and user needs, but it must be compatible to the exchange format.

Details of format and contents of exchange formats will depend on the type of data to be exchanged. But some principles should have general validity.

A data exchange system must be designed according to the carefully studied needs of a defined users' group. The exchange system will possibly consist of three files: one for experimental data, one for evaluated data, and one for a data index giving awareness of all data, whether they have been compiled already or not yet.

Together with the formal exchange agreement it is necessary to reach agreement on common procedures for the critical data analysis and validation. It is most essential for the reliability of the data files that the authors receive proof-copies of their data.

A data set has no value if it is not accompanied by certain text information, including

- bibliographic information;
- physics information, in particular uncertainty analysis and standard reference data used;
- bookkeeping information, in particular STATUS (preliminary - final - "approved by author") and HISTORY (compiled when, by whom, and from what source).

All this must be considered when a data exchange is designed.

If a world-wide agreed exchange format is used in combination with agreed procedures for the critical data analysis and data validation, one obtains a decentralized database which has a similar value as the traditional archival journals.

Data centre services to data users in all countries can best be achieved by a network of co-operating national or regional data centres who simultaneously update their common data base by means of an agreed data exchange system.

Captions to the figures

Fig. 1. The neutron-induced fission cross-section for uranium-235 in the small neutron-energy range between 50 and 60 eV (taken from the Japanese report JAERI-M-9823). The solid line shows the "evaluated data" produced by the data evaluator from several sets of "experimental data" indicated by different symbols. The "resonance structure" of the data requires very large data files to describe the reaction for the complete range of neutron energy from milli-eV to Mega-eV.

Fig. 2. Simplified scheme of the data exchange between the four co-operating data centres who jointly build up the international database. Their customers include data producers and data users, as well as national sub-centres who may distribute data further. Other specialized data centres feed their data into the exchange system through one of the main centres.

Fig. 3. The structure of an EXFOR entry. The ENTRY consists of SUBENTRIES, and each SUBENTRY consists of sections for text, for constant parameters, and for the actual data table. The text information is identified by keywords, e.g. REACTION, which may be followed by free text (not computer-intelligible) and/or coded information (computer-intelligible), e.g. "92-U-235(N,F),,SIG" (=U-235 neutron-induced fission cross-section).

Fig. 4. Simplified structure of an ENDF entry. An entry contains all data for a given material (element-isotope), identified by a MAT-number. Each entry is subdivided into files, identified by MF-numbers, and subfiles identified by MT-numbers. Except for file 1 all information is computer-intelligible. Each possible data type is identified by a pair of MF/MT. In each record, MAT-MF-MT is given in the record identification field, so that a computer program can easily find the information which it requires. Each subfile is preceded by certain control numbers which may indicate, for example, whether linear or logarithmic interpolation must be used for energies or cross-section data.

References

[1] H.D. Lemmel, Short guide to EXFOR, report IAEA-NDS-1 Rev. 5, June 1986. V. McLane (ed.), EXFOR Manual, version 87-1, unnumbered report of NNDC, Brookhaven National Laboratory, USA. - H.D. Lemmel (ed.), NDS EXFOR Manual, report IAEA-NDS-3 Rev. of Aug. 1985.

[2] ENDF/B system version 5. Formats Manual, see R. Kinsey, ed., report BNL-NCS-50496 (ENDF-102) 2nd ed. (Oct. 1979. Revision of 1983 see report IAEA-NDS-75 (Sept. 1986).

[3] CINDA 88, the index to the literature and computer files on microscopic neutron data, published on behalf of USA National Nuclear Data Center, USSR Nuclear Data Centre, NEA Data Bank, IAEA Nuclear Data Section (IAEA, Vienna 1988). - CINDA Coding Manual maintained by NEA Data Bank.

[4] O. Schwerer, H.D. Lemmel (ed.), EXFOR Dictionaries, updated quarterly, available on magnetic tape from IAEA Nuclear Data Section. A printed version available as INIS microfiche see report IAEA-NDS-2 (Aug. 1979).

[5] Wang Dahai (ed.), WRENDA 87/88, IAEA report INDC(SEC)-95 (Aug. 1988).

[6] IAEA Nuclear Data Newsletter, issued about annually by the IAEA Nuclear Data Section.

[7] S.F. Mughabghab, R.R. Kinsey, C.L. Dunford (ed.), Neutron Cross-Sections, series of handbooks, started in 1981, Academic Press.

[8] H. Münzel, H. Klewe-Nebenius, J. Lange, G. Pfennig, K. Hemberle, Karlsruhe Charged Particle Reaction Data Compilation, report Physik Daten/Physics Data Nr. 15 by Fachinformationszentrum Karlsruhe, H. Behrens, G. Ebel (ed.) (1979-1982).

[9] A. Lorenz (ed.), Report of the fifth "Four-Centres Meeting", Moscow, INDC(NDU)-16 (Dec. 1969).

[10] H.D. Lemmel, D.E. Cullen, J.J. Schmidt, Nuclear data files for reactor calculations and other applications. Experimental data - evaluated data. Computer Physics Communications 33 (1984) 161 - 171.

[11] Computer graph by my son Hartmut Lemmel.

[12] R. Kinsey (ed.), ENDF/B summary documentation, report BNL-17541 (ENDF-201), 3rd ed. (July 1979). - B.A. Magurno, P.G. Young (ed.), ENDF/B-V.2 summary documentation, report BNL-NCS-17541, 3rd ed. suppl. 1 (Jan. 1985).

[13] D.E. Cullen (ed.), ENDL-84, LLNL Evaluated Nuclear Data Library 1984 in ENDF-5 format, contents and documentation. Report IAEA-NDS-11 Rev. 4 (1985).

[14] JEF - Joint European File, maintained by the OECD-NEA Data Bank.

[15] T. Nakagawa, Summary of JENDL-2 General Purpose File, report JAERI-M-84-103 (1984).

[16] V.N. Manokhin (ed.), Library of recommended evaluated neutron data BROND, Obninsk (1986), in Russian. English translation report INDC(CCP)-283, IAEA, Vienna (1988). - H.D. Lemmel, P.K. McLaughlin, BROND, Summary documentation of the tape BROND-NDS, report IAEA-NDS-90 Rev. 1 (1988).

[17] D.E. Cullen, P.K. McLaughlin, The International Reactor Dosimetry File IRDF-85, report IAEA-NDS-41 Rev. 1 (1985).

[18] D.E. Cullen, Summary of ENDF pre-processing codes, report IAEA-NDS-39 Rev. 3 (1987). - P.K. McLaughlin, ENDF pre-processing codes, implementing on a personal computer, report IAEA-NDS-69 (1987). - P.K. McLaughlin, ENDF utility codes version 6.4 for ENDF-5 and ENDF-6, report IAEA-NDS-29 Rev. 2 (1988).

[19] P.F. Rose, C.L. Dunford, Data formats and procedures for the Evaluated Nuclear Data File ENDF-6, preliminary issue, report ENDF-102 (May 1988).

[20] D.E. Cullen, The accuracy of data processing, proceedings of the KTG/ENS seminar on "Nuclear Data, Cross-Section Libraries, and their Application in Nuclear Technology", Bonn, 1-2 Oct. 1985, page 62, Kerntechnische Gesellschaft e.V. (1985).

[21] W. Mannhart, A small guide to generating covariances of experimental data, report PTB-FMRB 84 (Physikalisch-Technische Bundesanstalt, Braunschweig, 1981). - W.L. Zijp, Treatment of measurement uncertainties, report ECN-194 (Netherlands Energy Research Foundation, 1987).

[22] V.G. Pronjaev, H.D. Lemmel (ed.), Summary documentation of the ENDF/B-5 Activation File, report IAEA-NDS-38 Rev. 1 (1984).

[23] N. Dayday, H.D. Lemmel (ed.), Contents and summary of the ENDF/B-5 Standards Library, report IAEA-NDS-15 Rev. 2 (1986).

[24] H. Goldstein, M.H. Kalos, A listing of CINDA, report EANDC (US)-69 (1964). - CINDA 65, report EANDC-46, by Columbia University and ENEA Neutron Data Compilation Centre. - CINDA 67, report EANDC-70; this was the first international issue by USAEC Division of Technical Information Extension, USSR Nuclear Data Information Centre, ENEA Neutron Data Compilation Centre, IAEA Nuclear Data Unit.

Appendix: Nuclear Reaction Data Centres

China:

Chinese Nuclear Data Centre, Cai Dunjiu, Institute of Atomic Energy, P.O. Box 275(41), Beijing.

Fed. Rep. of Germany:

Karlsruhe Charged-Particle Nuclear Data Group, H. Münzel (discontinued), Kernforschungszentrum Karlsruhe GmbH, Postfach 3640, D-7500 Karlsruhe.

German Democratic Republic:

Nuclear Data Group, D. Seeliger, Sektion Physik, Technische Universität Dresden, Mommsenstrasse 13, DDR-8027 Dresden.

India:

Nuclear Data Section, S. Ganesan, General Services Building, Indira Gandhi Centre for Atomic Research, Chengalpattu District, Kalpakkam 603 102, Tamilnadu.

Japan:

Nuclear Data Centre, S. Igarasi, Japanese Atomic Energy Research Institute (JAERI), Tokai-mura, Naka-gun, Ibaraki-ken 319 11.

Nuclear Data Group, RIKEN - A. Hashizume, Institute of Physical and Chemical Research, 2-1 Hirosawa, Wako-shi, Saitama 351.

Charged Particle Nuclear Data Study Group, H. Tanaka, Faculty of Science, Hokkaido University, Nishi 8-Chome, Kita 10-Jo, Sopporo-shi 060.

U.S.A.:

US National Nuclear Data Center, S. Pearlstein, Bldg. 197D, National Nuclear Data Center, Brookhaven National Laboratory, Upton, New York 11973.

Nuclear Data Group, R. Howerton, Lawrence Livermore National Laboratory, L-298, P.O. Box 808, Livermore, CA 94550.

Photon and Charged Particle Data Center, M. Berger, Center for Radiation Research, National Bureau of Standards, Gaithersburg, MD 20899.

U.S.S.R.:

USSR Centr po Jadernym Dannym (Nuclear Data Center), V. Manokhin, Fiziko-Energeticheskij Institut, Ploschad Bondarenko, Obninsk 249 020, Kaluga Region.

Centr po Dannym o Stroenii Atommogo Jadra i Jadernykh Reakcikh (Nuclear Structure and Nuclear Reaction Data Center), F.E. Chukreev, Institut Atomnoi Energii I.V. Kurchatova, Ploschad I.V. Kurchatova, Moscow D-182, 123182.

USSR Centr Danykh Fotojad. Eksp., V. Varlamov, Moscow State University, Nauchno-Issl. Inst. Jad. Fiz., Leninskie Gory, Moscow 119899, V-234.

International Organizations:

OECD Nuclear Energy Agency Data Bank, J. Rosen, 38 Boulevard Suchet, F-75016 Paris, France.

IAEA Nuclear Data Section, J.J. Schmidt, International Atomic Energy Agency, P.O. Box 100, A-1400 Vienna, Austria.