Beyond the ENDF Format Workshop
OECD/NEA Data Bank
30 November 2012

On the ENDF Formats and Data Processing

Andrej Trkov
Jožef Stefan Institute, Ljubljana, Slovenia

Historical Overview

Nuclear constants like the cross sections exhibit a strong energy-dependence, which may require several hundred-thousand points to describe accurately. Also, the database must be complete: parameters (e.g. the cross sections) must be defined at all energies, even when no experimentally measured data are available. It is the task of the evaluator to assess the most probable value of a parameter at any energy, resolving issues of discrepant measurement, assigning values (by an educated guess or based on model calculations) where no data are available and providing data in computer-readable format.

Historically, several formats were in existence to store the evaluated nuclear data:

- KEDAK in Germany
- UKNDL in the UK
- ENDL at Livermore National Laboratory, USA
- ENDF in the rest of the USA
- SOKRATOR in the Soviet Union.

The formats were incompatible and hindered free data exchange, which was agreed upon since the Geneva Summit in 1955. Format translation programs were developed locally, but were often limited to selected evaluations. Some attempts were made to develop flexible processing codes to handle "all" formats. This was usually done by translating original evaluations into a common internal format (e.g. FEDGROUP by P. Vertes), but the codes were generally quite unstable.

The ENDF/B-IV evaluated nuclear data library became generally available in the eighties. At the time it was the most comprehensive compilation of general purpose evaluations and soon found broad usage all over the world.

The NJOY-91 data processing system was released, which could handle the data in ENDF format. This fact contributed greatly to the widespread usage of the ENDF/B-IV library.

The ENDF/B-V library was released in the USA in 1979. It was sponsored by the industry and for this reason it was not released to the users outside the USA and Canada. Also, this decision indirectly contributed to the establishment of the joint European JEF library evaluation project in Europe and the JENDL project in Japan. Both of these projects adopted (a subset) of the ENDF format as the basis. The main advantage of the ENDF format was the detailed documentation, which provided rules and procedures to uniquely define the cross sections and other quantities at each point, including the prescription of interpolation rules for tabulated data. The European approach was to build on improvements to the ENDF/B-IV library, which resulted in the JEF-1 library and later on the JEF-2.2 library, which was the

main workhorse for the industry in Europe until not so long ago. Japan followed a more independent route, relying extensively on evaluations from first principles, culminating in the release of the JENDL-2 library. The most recent libraries are ENDF/B-VII.1 in the USA, JEFF-3.1.2 in Europe, JENDL-4.0 in Japan, CENDL-3.1 in China and ROSFOND-2010 in Russia. Other special-purpose libraries also exist and are available from the Nuclear Reaction Data Centres electronically over the Internet. The bulk of the evaluated nuclear data libraries are given in the ENDF format version six, namely the ENDF-6 format.

The ENDF format thus received world-wide recognition in the USA, Europe, Japan, China, as well as in the former Soviet Union. The main advantage of this choice was the ease of data exchange and the availability of a fairly robust NJOY data processing system, which soon enough made other formats and data processing systems obsolete. More so, the AMPX code system from Oak Ridge was poorly maintained for some time. The only real alternative available to general users was the Pre-Pro suite of codes from the IAEA, but their main purpose was data verification, validation and display. They did not include any modules for the preparation of particle energy-group transfer matrices and interfaces to the production libraries for particle transport calculations. In recent years the AMPX system has been brought up-to-date, but NJOY remains the main work-horse for the production of application libraries outside Oak Ridge, while the Pre-Pro codes are focusing on basic data processing for verification and validation purposes.

Data Format Requirements

The ENDF format was designed in the sixties for computer technology, which was based on punched cards for data exchange. Many features of the ENDF format reflect this situation. The ENDF format was evolving with time to accommodate advanced features for better representation of the data and to suit the needs of different users, but the general structure of the format remained quite rigid and was often criticised for the same reason.

Over the years the experience with the ENDF format proved the concept of ordering data structures in a way to minimise duplication and to provide a compact way of data representation. With enormous advances in computer technology some of the features may seem obsolete, but perhaps not entirely as will be discussed below, provided that one is prepared to look at the ENDF format from a broader perspective.

Compact data representation
The capacity of the media for storing the data is incomparably larger than it was when the ENDF format was designed, but the volume of information compiled by the evaluators and the details of data representation also increased enormously. If a new format is to be designed, many of the features of the ENDF format for compact data representation are likely to be retained.

Format rigidity
Formally, the ENDF format is indeed rigid, but in the data centres the evaluated data are not stored in long sequential ASCII files. They are stored in relational databases by blocks; the original ENDF files are assembled together based on the user-request. In general, the way data are stored and handled actually depends on the user, as long as he has the capability to sort and interpret the data correctly. Working the other way round from the point of view of data evaluation, the evaluator can fill the relational database in whichever way he wants and

rely on some expert system to output the data in a consistent way conforming to the format specifications. This will work in most cases, although with time we shall probably encounter data types that will need to be archived and will not fit into the ENDF framework. Until then at least, the ENDF format can serve the purpose of providing a unique and flexible platform for data exchange.

Some of the frequently criticised features of the ENDF format
Sequence numbers: are indeed redundant in a well-tested production library, but when problems are encountered, they provide a useful pointer to locate the data that cause problems. An evaluator should not bother with the sequence numbers. They can be recovered automatically (e.g. by the STANEF code) and should be made optional, except for the officially released libraries for data exchange.

Material numbers: (the so-called MAT numbers) provide an index to an evaluation, which in an assembled library should be unique. Formally, the MAT number assignment according to the ENDF rules is arbitrary, although the ENDF/B Library Project in the USA requires a specific convention. The MAT numbers will retain their significance in sequential ENDF libraries in ASCII format and for the convenience of the legacy data processing codes that can not yet search the materials by their ZA/LIS0 designation (in ENDF terminology).

Reaction numbers: (the so-called MT numbers) are indeed a limitation of the ENDF format and there is hardly a way around it, if reactions are to be represented explicitly for whatever purpose. This deficiency is a matter of principle. Most of the reactions, which can be measured experimentally, have the corresponding MT numbers assigned. The more complex reactions can only be calculated by nuclear model codes. However, for practical purposes (e.g. burnt fuel isotopic assay) only the radionuclide production data are actually needed and they can be accommodate in the current format.

Fixed floating point representation: restricts the numerical resolution of the data to a maximum of nine digits, and more commonly to seven digits when standard number representation is used. There is seldom a real need for more than nine digits, but free format representation might be advantageous for storing the covariances. It would also relieve the evaluators at intermediate steps of the evaluation process to worry about the precision and some users, who read ENDF files by codes written in languages other than Fortran.

Other criticised features are related to the specific data structures, which some users find difficult to understand and interpret. In some cases more elegant ways of blocking the data could be designed, but not always. If a new format is to be produced, designers face essentially the same problems as the ENDF format designers.

Requirements for a new evaluated library format

In the past, several attempts were made to define new formats for storing evaluated nuclear data. The key question (frequently stressed by D.E. Cullen) is: “Does it add value”? Since the files are read by codes, making the numbers look pretty is not good enough. It is necessary to demonstrate distinct advantages of the new format, which are not only cosmetics. Most of the attempts failed on this question, combined with the needed manpower to upgrade and validate the data processing codes.

The so-called GND format has been elaborated at LLNL based on MXL. Its main advantages are the general software support for handling the generic structures of the data in XML, which makes the creation of relational databases easier. An evaluator also has more freedom on how he enters the actual numbers. However, the problem of suitable grouping and blocking of data remains. The GND format is a new development and no doubt, in some cases the designers will find better solution to block the data, but there is no reason to freeze the ENDF format, so useful features could be added to it.

Overall, the GND designers took the optimal approach by creating translation codes from ENDF to GND and vice versa, which offers many advantages:

- LLNL can fill their database with currently the most up-to-date evaluated nuclear data.
- The evaluators at LLNL (and elsewhere) can perform evaluations in XML, using the expert system to convert the data into ENDF format consistently, without having to know all the salient features of the ENDF format.
- The translation codes allow (for a while) the continued use of the legacy data processing codes, into which great many man-years were invested. They will serve as convenient benchmarks for new codes reading directly the GND format, when they are developed and released to the users.
- Increased experience with the GND format might identify cases where certain important features of the data can not be translated accurately into ENDF. This will be the time for the decision to break with the ENDF format and switch to the new one, but by then most of the features of the new format will be thoroughly validated.

There are data features which the current formats do not address yet. Some of them are listed below.

Covariances of the correlated energy/angle distributions of emitted particles would logically fit into ENDF File-36, which is not defined and would imply a three-dimensional matrix. Personally the author of this work is of the opinion that the complexity of such a (tentative) development effort is not justified. The primary component of the uncertainty is the relative uncertainty of the distribution for some energy of the emitted particle. The differences in the relative uncertainties and correlations as a function of angle varying with emission energy are probably smaller that the differences as a function of incident energies, which are currently not accommodated in the format either and would imply the fourth dimension of the matrix. The covariances of the double-differential distributions can be handled by assuming separability in energy and angle (for the purpose of uncertainty calculations only) and storing the covariances in Files-34 and 35. If anything, the methods for defining correlations between the distributions at different incident energies should have priority.

Covariances of the resonance parameters have been analysed for several materials and some interesting features cropped up. Some materials with a very large number of resonances resulted in covariance matrices that exceed 600 MB of information. The question of numerical precision in the handling of such a matrix becomes relevant. A compact format has been proposed to reduce the volume of information in evaluated libraries, but this does not reduce the problem of the order of the matrix. Furthermore, the truncation of the digits for the representation of the matrix may give rise to negative eigenvalues of the matrix, which are unphysical. Methodology needs to be developed to treat the covariances of the important resonances explicitly without destroying the consistency of the evaluation. This is needed, because there is evidence that translating covariances of resonance parameters into the

covariances of the cross sections may in some cases lead to incorrect results when self-shielding (and Doppler-broadening) are present [Ref. Mn-55 paper].

Future needs of the nuclear data processing systems

Nuclear data files are very closely related to the processing codes, which can turn evaluated nuclear data into application libraries. The NJOY Nuclear Data Processing System is the most widely used system for generating application libraries. It was not subjected to severe restriction regarding its distribution and it has many users around the world. The NJOY list-server at the NEA Data Bank is a forum for reporting problems and providing temporary solutions contributed by the users outside the USA, until the official updates are released. This allows traceability of the code used to produce a particular library and avoids many of the ambiguities, which were noted in the past. It also provides suggestions to the official code maintenance team on how the reported problems could be solved.

The official version of NJOY is version 99, with 364 sets of official updates. In parallel, a new version of NJOY is being tested, but it is not officially released. Certain features appear in the new version that are not present in the old one. Sooner or later the decision will have to be taken to switch from the old version to the new one. Hopefully, the new version will not be subjected to more severe restrictions on the distribution, since the present arrangement is beneficial to all sides and will speed up the validation of the new code version. Otherwise, branching of the processing code is likely, which will introduce another source of ambiguity into the calculations, which will make traceability of the discrepancies in the calculated results on international level much more difficult.

The current NJOY-99 code has enhanced capabilities to process covariances, including the covariances of the resonance parameters and cross-material covariances. The validation of the covariances is a very slippery subject, since they are not unique and depend on the methodology used in the evaluation process. In spite of the effort already invested, more is needed to understand the covariances properly.

The idea of "Total Monte Carlo" has been reported in several recent works. The basic idea is to generate perturbed ENDF files by statistically sampling input model parameters on which the evaluation is based, and use each file in a full Monte Carlo transport calculation to calculate some integral parameter. If sufficient samples of the input model parameters are taken, the statistical scattering of the calculated integral parameter should give the average and the standard deviation. It is an extremely powerful technique, but is potentially very dangerous if applied blindly. The problem with this approach is that it works only above the resonance range (but see also next paragraph) and that the nuclear model deficiencies and uncertainties in the measured cross section data are not taken into account (or treated very crudely). In principle the so-called Unified Monte Carlo approach could treat the experimental data rigorously, but there might be an open issue with the convergence, which could mean that this approach is not feasible in practice.

Covariances of the resonances can be taken into account in the Total Monte Carlo method by statistical sampling of the resonance parameters, generating ENDF files from each sampled set and then running the transport calculation (by the Monte Carlo method or otherwise). This has been done at NRG, but a simpler problem was investigated at JSI where the uncertainty of the resonance integral as a function of the self-shielding level was investigated, demonstrating

that in general the resonance covariances can not be substituted by cross section covariances when self-shielding is present. The second problem that was identified in the work was the problem of large uncertainties appearing in some weak resonances which can lead to a large number of sampled negative resonance widths, which are unphysical. The method of correctly treating the sampled distributions was elaborated such that the mean and the average are preserved and that all sampled values are positive [Ref to NIM paper].

Validation of the covariances could be performed by the following exercise:

- Choose a system where the resonance range is not important and ignore the covariances of the resonance parameters.
- Perform total Monte Carlo calculation by sampling model parameters in the usual way. In addition, generate the covariance matrices in full detail from the generated ENDF files.
- Restore the full covariance matrix from the information in the ENDF file, statistically sample all cross sections from this global covariance file, generate perturbed ENDF files and perform Monte Carlo calculations with each of them.
- Compare the averages and the standard deviations obtained from the classical Total Monte Carlo and the variant with ENDF files generated from the global covariance matrix. The integral parameters could include the multiplication factor (in case of fast reactor criticality problems), as well as reaction rate ratios and spectra.

The outcome of this exercise would provide an indication how much we can trust the mean values and the uncertainties from the covariance calculations. One has to bear in mind that the results of the exercise are applicable only to problems of similar nature. They shed light on the importance of the correlations, which are lost in the process of assembling the ENDF files, and on the linearity assumption associated with the perturbation method, assuming that the correlations arising from the sampling of nuclear model parameters are more or less realistic (note: the comparison of the mean values and standard deviations is meaningful because they are both based on the same set of perturbed cross sections, even though the correlations in the cross sections themselves might be unrealistic).

Considering the importance of the covariances to estimate the uncertainties and the research effort invested in this field lately suggests that further work on the validation of covariance processing is justified. A common tool for generating perturbed ENDF files from the covariance information might be advantageous and would open the door widely to the extension of the Monte Carlo technique for uncertainty estimation using deterministic codes. One potential problem is related to the assembly of a global matrix, which should include all cross-reaction and cross material correlations, considering that the energy grids for the covariances could be different in each case. The second potential problem is the actual processing of such covariance matrix and avoiding numerical roundoff errors.

Summary of Conclusions

The ENDF formats have served the community of nuclear data users from different fields of applications quite well for decades. Enormous effort has been devoted to the development and validation of the processing codes. Although there is no urgent need for a rapid transition to something completely new, there are signs that the current ENDF format is being pushed close to its limits. The time is right to look for a modern replacement, with due consideration for the following:

- Development of data processing capabilities, starting from the data in the new format.
- Backward compatibility through robust translation codes between the new and the old format until the majority of processing tools have been adequately validated.
- Standardisation of the format features on the international level to maintain the possibility of easy data comparison and exchange.

The NJOY Data Processing System is the most versatile and widely used code system for generating application libraries. The AMPX system is mainly used for generating libraries for codes from Oak Ridge. The Pre-Pro codes are found to be very robust, but their main purpose is data verification, validation and display. These codes do a good job for the present scope of applications, but current trend rely heavily on Monte Carlo simulations and sensitivity/uncertainty calculations. Further developments in the data processing tools should reflect these trends, focusing on the following:

- Further verification and validation of covariance processing methods.
- Development of a common tool for generating a global covariance matrix of nuclear data, including all available cross-reaction and cross-material correlations.
- Consider if we can move away from histogram covariance representation into a piecewise linear domain.
- Having a "global" covariance matrix (that can include the covariance matrix of the resonance parameters), pursue the development of a common tool for statistical sampling of the cross sections and other parameters and generation of "perturbed" ENDF files. This would open the door to a variant of the Total Monte Carlo method that is not based on the sampling of nuclear model parameters. It will require generalisation from the assumption of normal probability distributions (particularly for inherently positive parameters with large assigned uncertainties), in addition to the numerical problems of handling the enormous amount of information.

ADDENDUM

Some thoughts on a different representation of cross sections and covariances in evaluated nuclear data files are given below.

- The TAB1 equivalent record would contain three values for each point, namely the energy, the cross section and the relative uncertainty.
- The uncertainties would be represented by the correlation matrix relating to the pointwise values (and not group averages, as it is now).

The potential advantages of this representation include:

- The range of values in the correlation matrix is much smaller, since the diagonal is always one. It might be easier to identify and eliminate insignificant correlations. The matrix might exhibit better numerical stability in the calculations.
- The uncertainties would not have discontinuities.