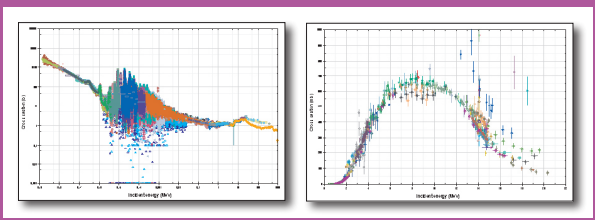
Nuclear Science
2011

OECD

Quality Improvement of the EXFOR Database

International Evaluation Co-operation, Volume 30



AEN NEA

NUCLEAR ENERGY AGENCY

Nuclear Science
NEA/WPEC-30

NEA/NSC/WPEC/DOC(2010)428

International Evaluation Co-operation

Volume 30

Quality Improvement of the EXFOR database

A report by the Working Party
on International Evaluation Co-operation
of the NEA Nuclear Science Committee

Co-ordinator

A. Koning
Nuclear Research & Consultancy Group
The Netherlands

Monitor

A. Mengoni
International Atomic Energy Agency
Austria

NUCLEAR ENERGY AGENCY
Organisation for Economic Co-operation and Development

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 34 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

NUCLEAR ENERGY AGENCY

The OECD Nuclear Energy Agency (NEA) was established on 1st February 1958 under the name of the OEEC European Nuclear Energy Agency. It received its present designation on 20th April 1972, when Japan became its first non-European full member. NEA membership today consists of 29 OECD member countries: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, the Netherlands, Norway, Poland, Portugal, Republic of Korea, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities also takes part in the work of the Agency.

The mission of the NEA is:

- to assist its member countries in maintaining and further developing, through international co-operation, the scientific, technological and legal bases required for a safe, environmentally friendly and economical use of nuclear energy for peaceful purposes, as well as
- to provide authoritative assessments and to forge common understandings on key issues, as input to government decisions on nuclear energy policy and to broader OECD policy analyses in areas such as energy and sustainable development.

Specific areas of competence of the NEA include safety and regulation of nuclear activities, radioactive waste management, radiological protection, nuclear science, economic and technical analyses of the nuclear fuel cycle, nuclear law and liability, and public information.

The NEA Data Bank provides nuclear data and computer program services for participating countries. In these and related tasks, the NEA works in close collaboration with the International Atomic Energy Agency in Vienna, with which it has a Co-operation Agreement, as well as with other international organisations in the nuclear field.

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*.

Foreword

The Working Party on International Nuclear Data Evaluation Co-operation (WPEC) was established under the aegis of the OECD/NEA Nuclear Science Committee (NSC) to promote the exchange of information on nuclear data evaluations, validation and related topics. Its aim is also to provide a framework for co-operative activities between the members of the major nuclear data evaluation projects. This includes the possible exchange of scientists in order to encourage co-operation. Requirements for experimental data resulting from this activity are compiled. The WPEC determines common criteria for evaluated nuclear data files with a view to assessing and improving the quality and completeness of evaluated data.

The parties to the project are: ENDF (United States), JEFF/EFF (NEA Data Bank member countries) and JENDL (Japan). Co-operation with evaluation projects of non-OECD countries, specifically the Russian BROND and Chinese CENDL projects, are organised through the Nuclear Data Section of the International Atomic Energy Agency (IAEA).

The following report has been issued by WPEC Subgroup 30, whose mission was to improve the accessibility and the quality of the EXFOR database to ensure that its invaluable contents can be used in contemporary evaluation work.

The opinions expressed in this report are those of the authors only and do not necessarily represent the position of any member country or international organisation.

Members of Subgroup 30

A. Koning
Nuclear Research & Consultancy Group (NRG)
The Netherlands

E. Dupont
Commissariat à l'énergie atomique (CEA)
France
OECD Nuclear Energy Agency (NEA)
France

V. Zerkin
IAEA Nuclear Data Section
Austria

H. Henriksson
OECD Nuclear Energy Agency (NEA)
France

N. Otuka
IAEA Nuclear Data Section
Austria
Japan Atomic Energy Agency (JAEA)
Japan

N. Soppera
OECD Nuclear Energy Agency (NEA)
France

M. Bossant
OECD Nuclear Energy Agency (NEA)
France

S. Dunaeva
IAEA Nuclear Data Section
Austria

O. Schwerer
IAEA Nuclear Data Section
Austria

V. McLane
Brookhaven National Laboratory (BNL)
USA

D. Brown
Lawrence Livermore National Lab. (LLNL)
USA

R. Capote Noy
IAEA Nuclear Data Section
Austria

J-C. David
Commissariat à l'énergie atomique (CEA)
France

H. Duarte
Commissariat à l'énergie atomique (CEA)
France

R. Forrest
IAEA Nuclear Data Section
Austria
UK Atomic Energy Authority
(UKAEA, now CCFE)
United Kingdom

A. Konobeyev
Karlsruhe Institute of Technology (KIT)
Germany

J. Kopecky
Juko Research
The Netherlands

S. Leray
Commissariat à l'énergie atomique (CEA)
France

S. Mashnik
Los Alamos National Laboratory (LANL)
USA

A. Mengoni
IAEA Nuclear Data Section
Austria

B. Pritychenko
Brookhaven National Laboratory (BNL)
USA

D. Rochman
Nuclear Research & Consultancy Group (NRG)
The Netherlands

A. Trkov
IAEA Nuclear Data Section, Austria
Institut Jozef Stefan (IJS), Slovenia

Table of contents

List of figures

1. Introduction

Among the existing databases with experimental nuclear reaction data, EXFOR [1] is by far the most important and most complete. A document celebrating 50 years of the US CSEWG evaluation committee [2] reveals the effort that has gone into the database over the past half century. Maintained by the various Nuclear Reaction Data Centres (NRDC) [3], the library now contains numerical data of about 147 000 data sets from more than 19 000 experiments performed since 1935. The database mainly contains numerical data and experimental/bibliographic information on experiments for incident neutron, charged particle ($A \leq 12$) and photon-induced reactions on a wide range of isotopes, natural elements and compounds, for incident energies up to about 1 GeV. With a rough estimate of the investment value (salaries, equipment, etc.) of a typical experiment, which will not be undertaken here, it is not too difficult to get an idea of the value that the EXFOR database represents, both in terms of monetary and historical value.

In spite of this compilation effort, the retrieval of experimental data from EXFOR runs the danger of becoming one of the main delaying factors in contemporary nuclear data evaluation. In the past, experimental data have been added to EXFOR by various compilers who, although the EXFOR format was unified in 1969, have used different (and often allowed) procedures and entries to store the data. The result is a database that contains the numerical data of almost the entire history of nuclear reaction measurements. The main question now is whether and how the user has access to all these data. For the study of a few detailed reactions, web interfaces are helpful tools to retrieve a few data sets, and one may find all available data by trying various different search options. However, nuclear data evaluation is becoming more automated and quality-assured, so that a genuine step forward in the production of data libraries requires that *all* nuclear data that exist in EXFOR can be retrieved in an unambiguous manner, and without unnecessary delay, and this is lacking at the moment. We all expect flawless, quality-assured and robust usage of evaluated nuclear data libraries in applied (reactor) calculations, so why not expect the same for EXFOR when we use it in the data evaluation process itself?

In addition to the problem of retrievability, EXFOR is known to contain various errors, and admittedly in the past these errors have perhaps not been reported sufficiently by the user community to the NRDC.

There are many reasons why the quality of EXFOR is essential. Some important ones are:

- Nuclear model codes. Evaluated data libraries rely more and more on nuclear model calculations which can provide data for energies, reaction channels and nuclides for which no experimental data exist. An easy comparison with all existing experimental data would be extremely helpful for validating these model codes.
- The evaluation of individual isotopes. To increase efficiency, it is important that an evaluator have easy access to *all* available data, and that these data are correctly represented in the database.
- The global understanding of covariance data as a very important ingredient for advanced reactor studies. It is not only required that the data are good, but also to know how good they are. Once this is properly assessed for key nuclear reactions, on the basis of *all existing* experimental and theoretical information, proposals for new measurements can be more easily justified, as their impact on technological applications will be clearer. An important condition is that the *existing* data in EXFOR be retrievable, complete and reliable.
- Good experimental work should be correctly represented and easily accessible, leading e.g. to an amount of bibliographic references that does the work justice.

These issues alone already require that EXFOR be not only a large and complete database, but also a properly validated database (similar to the Evaluated Nuclear Data Files).

Finally, an unambiguously defined good-quality experimental database fits into a larger and more modern nuclear data framework, in which the entire process of data evaluation and validation is better automated and performed under a more strict quality assurance scheme. Eventually, this could make the evaluation of a data file reproducible, instead of an incremental *ad hoc* process. Note that this does not remove the necessity to study experimental results in detail, to make an appropriate weighting in a nuclear data evaluation. What we are concerned with here is eliminating *unnecessary* delay.

For all of these reasons, SG30 was launched. This subgroup aims to establish EXFOR as an easily accessible and correct database. The ultimate objective could be formulated as an appropriate solution to the following two problems (in this order):

- In past decades, almost all experimental nuclear reaction data has been put into EXFOR. Can we also get it out of EXFOR, in an unambiguous way?
- If we can get the data out of EXFOR, are the data correct and if not, can they be corrected?

An important spin-off, and a necessity for testing, is a computational database that contains the entire EXFOR database in an easy-to-use tabular format. This enables:

- a huge increase in efficiency for the use of experimental data in nuclear data evaluation;
- easy and extensive validation of nuclear model codes;
- more feedback from users to the NRDC to correct data, since more data becomes available.

To reach this goal, SG30 has focused on the following activities:

- attempt to translate the entire EXFOR database into computational format;
- solve the most obvious quantitative errors, using checking codes, plotting packages and comparisons with model codes;
- identify data which are stored incorrectly and attempt to harmonise the format;
- propose steps beyond the scope of SG30.

In this final report, the accomplishments of SG30 will be presented. During the existence of SG30 various Internet sources were made available to its members, and they are open for any further retrieval or provision of information:

- The SG30 mailing list: *sg30@nea.fr*.
- The SG30 website at NEA: *www.nea.fr/html/science/wpec/SG30*. Here, large lists of feedback to the NRDC can be found, as well as SG30 working documents.
- Feedback from WPEC SG30 which can be found at the NDS website: *www-nds.iaea.org/nrdc/error/exfor_err3.html*.
- The IAEA repository of the entire EXFOR and computational format (XC4) databases: *www-nds.iaea.org/x4toc4-master*. These databases were the starting point for the global checking methods outlined here. During the lifespan of SG30, these databases were regularly updated with newer, corrected versions. It is under discussion how, and whether, to proceed with such availability in the future.

The contents of this report are as follows. In Chapter 2, we set the boundary conditions for SG30: which types of EXFOR corrections and extensions have been included, with global approaches, and which not. In Chapter 3, we discuss the transition of EXFOR to the computational database XC4. Next, we outline some statistical tests that can be performed on EXFOR in Chapter 4. In Chapter 5, we present a global comparison of the entire EXFOR database with calculated results from the TALYS nuclear model code. In Chapter 6, we show what can be achieved with visual comparison of EXFOR data and evaluated data libraries. In Chapter 7, we discuss some other tests that have been performed during the existence of SG30, and we then turn to the important role played by NRDC to correct all the problems found in Chapter 8. Finally, we end this report with conclusions and an outlook for future work.

2. Boundary conditions of SG30 activities

The quality assessment, validation and ultimate correction of the EXFOR database can be performed on four levels. In increasing order of effort these are:

- Correction of the most obvious errors:
  - Dimensional errors, e.g. if barns are given where millibarns should be given (or MeV instead of eV, etc.). These errors are often directly visible and readily emerge upon trying to process the entire EXFOR library.
  - Format errors, for either the X4 or computational format. There are errors in the reaction identification in the "mother database", and in the codes used at the Data Centres to handle the numerical data, e.g. X4toC4. This prevents proper use of the reaction data, and worse, reaction data may appear not to exist.
- Correction of more difficult to find, but still obvious dimensional errors, e.g. millibarns instead of barns, through comparison with nuclear model codes or other measurements of the same reaction.
- "Low-level" quality flagging of the data through comparison with other measurements of the same reaction and nuclear model codes.
- "High-level" quality flagging by letting a review team consisting of experimentalists and evaluators judge each experiment (or at least an important subset of the database) in detail, taking into account the measurement method used, quality of data analysis, reputation of the authors and laboratory, etc., and then assigning a quality flag, or use more recent standards data to re-normalise the data values.

In SG30 we considered the first, second and part of the third actions, i.e. identifying, and where possible correcting, errors that have been revealed by global methods.

For completeness, we also list the activities that may appear obvious from the outside, but which SG30 has *not* considered:

- Assign quality "flags" to experiments in EXFOR. Not only would this be a huge task, it also does not do justice to the essence of EXFOR as a compiled, rather than an evaluated, database. Quality flagging is useful but too subjective for SG30.
- Focus on the completeness of EXFOR, i.e. searching for and pointing NRDC to missing data.
- Insist on drastic changes of the EXFOR master format.

3. From EXFOR to a computational database (from X4 to XC4)

The format to store experimental data in EXFOR (EXchange FORmat) is very flexible. Practically all information that should be reported about a measurement can be represented with the EXFOR format. The current EXFOR system is not only a set of EXFOR files (they usually have the extension X4), but also a system of dictionaries describing allowed codes, rules, communication systems based on archived memos, etc. This completeness and flexibility is also EXFOR's weakness: it is rather difficult to retrieve the data in an unambiguous manner, making it difficult to make e.g. a simple plot of the data or automated comparisons with a model code. Also, it is difficult to collect together physically comparable data from different measurements. For example, various units (barns, millibarns, eV, keV, etc.) are allowed and the data can often be found in many different orders of appearance (sorted per energy, per angle, etc.). Essentially, a table with experimental data may be entered exactly as published by an author.

Long before SG30 was established, this particular inconvenience was recognised by Red Cullen (LLNL), who developed a code X4toC4 [4] to translate the EXFOR files into a computational database called C4. This code was later used, maintained and extended by Andrej Trkov at the IAEA. Beyond the code itself, X4toC4 also uses three dictionaries describing the correspondence between EXFOR and C4 data. These dictionaries were filled during the lifetime of the code: they were proposed by users (e.g. working with the EMPIRE code) and added to the official release, also maintained by A. Trkov. Since 2004, X4toC4 has been available under the IAEA-NNDC EXFOR-ENDF Web Retrieval System.

The C4 format is designed to present experimental data in a fixed set of units and column order. By starting from data in the EXFOR format and translating data to the computational format it is possible to combine the advantages of the improved reliability of the data coded in the EXFOR format with the advantages of a fixed unit and column order format for use in subsequent applications. The MF and MT numbers of the ENDF format are used to distinguish between different types of reactions.

A next important step towards further EXFOR user-friendliness was taken by Viktor Zerkin (IAEA), who used X4toC4 to translate the entire EXFOR database into the so-called extended C4 format, or XC4 format, in

which text blocks with basic reaction information are put between different reaction blocks to make the data easier to retrieve. The XC4 format lies at the basis of many tests performed by SG30. The IAEA now regularly provides updated versions of both the entire EXFOR master database and the XC4 database, so that plotting, checking and nuclear model codes of EXFOR users can frequently be re-run with a new version of the database. As can be inferred from the table with statistical results shown in Chapter 5, about half of the EXFOR database can currently be translated into the XC4 database. This mostly concerns cross-sections.

One may process the XC4 database into a directory-structured projectile/nuclide/reaction database, as done at NRG. The next two sections describe EXFOR testing schemes based on such a database.

4. Statistical tests

If all experimental data are available in some logically structured order, it is possible to test different measurements of the same quality against each other. This has been performed at CEA/DSM (Emmeric Dupont) [5]. For any reaction quantity, in a given energy bin, the basic idea is to prepare a set of experimental data that should follow a normal distribution and test whether there exist measurements (outliers) more than a few standard deviations away from the mean value. The main advantage of this approach is that any kind of data (cross-sections, nu-bar, ratios, resonance integrals, spectrum averages, etc.) can be checked with a single test based only on the database contents. It is thus not restricted by limitations of a nuclear model code. The practical implementation of this test requires the definition of three important parameters:

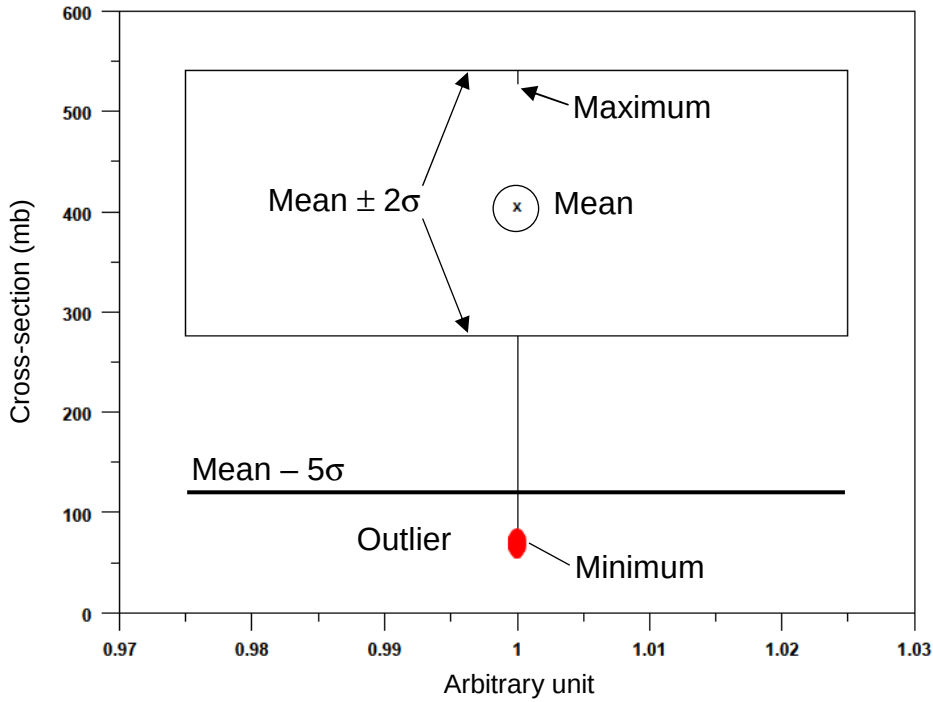
- The bin width. It should be small enough to make sure the data are roughly constant in the bin, but large enough to have acceptable statistics within the bin.
- The minimum number of points per bin. The confidence in the result increases with statistics. However, a large part of the database contains very few measurements of a given reaction in a given energy bin. All these data would be excluded from the test if the minimum number of points is set too high. Of course, no test can be performed when less than three points are available in a bin.
- The outlier criteria k. For a given reaction bin, if $Y \pm \Delta Y$ is one measurement among others with mean value $\overline{Y}$ and standard deviation σ; Then, Y is an outlier if it satisfies $\left|Y - \overline{Y}\right| \geq \Delta Y + k\sigma$.

Note that Y, as a potential outlier, is excluded from the calculation of the mean and standard deviation. Of course, if k is too large we will not detect any outliers, not even errors. However, if k is too small we will detect both compilation mistakes and poor quality data points as outliers. An actual example of outlier identification is given in Figure 1 with $k = 5$.

Several parameterisations have been tested to find the right balance between the number of reaction bins tested and the number of false alarms. The final energy bin widths vary between 10 meV and 5 MeV depending on the energy. The minimum number of points per bin was set to the lowest

Figure 1: Box-plot graph for a given cross-section in a given energy bin

The extreme and mean values are displayed together with the 2σ confidence limits.
In the present case, the minimum value, beyond 5σ, is considered to be an outlier.

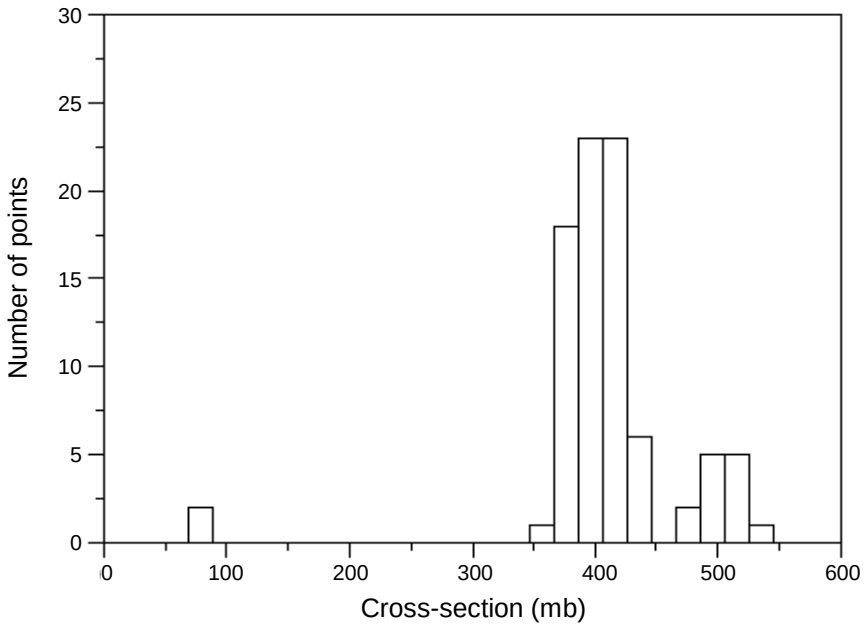


possible value of three. Differential and integral data for neutron, photon, and light-charged particle incident energies from 1 meV up to 150 MeV have been scanned. However, in the case of incident neutrons, the region between 0.2 eV and 100 eV was not tested because of a too-large number of false alarms due to isolated resonances. The statistical data (mean value, standard deviation...) of all reaction bins containing an outlier was automatically plotted for further inspection. Dataplot software [6] was used to display these data in three different formats: standard (*e.g.* cross-section vs. energy), box-plot (Figure 1) and histogram (Figure 2). A quick glance at these hundreds of plots was sufficient to discard most of the false alarms, mainly in the cross-section threshold regions, and to confirm the detection of a statistical outlier.

We note that there are other simple physical considerations to declare a data point suspicious. For example, above 1 MeV, any cross-section should probably not exceed 4 barns [unless it concerns non-exclusive values such as e.g. the energy-integrated (n,xn) spectrum]. With a computational database available, such errors are easily discovered and in fact give rise to the first batch of errors from SG30.

Figure 2: Histogram of a given cross-section in a given energy bin (same data as in Figure 1)

The lowest cross-section value is more than 5σ away from the mean value of the distribution

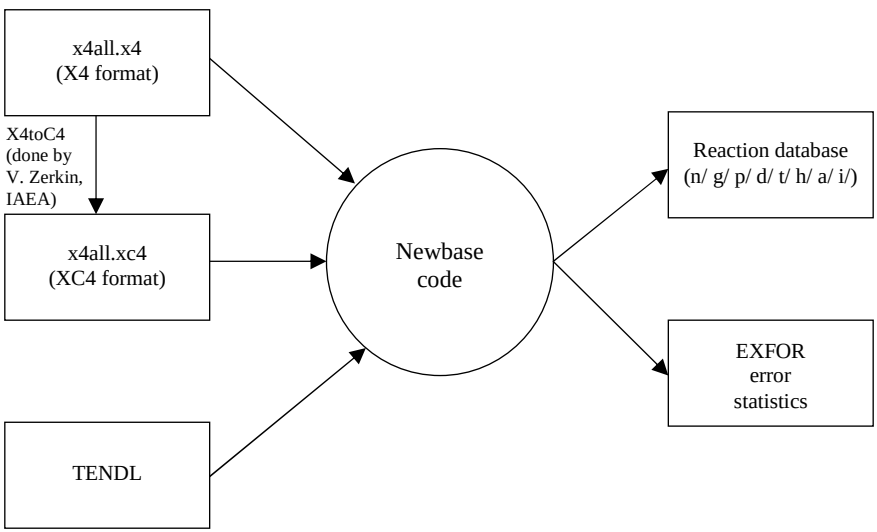


5. Comparison with TALYS nuclear model code

With the current computer power, it is possible to perform complete nuclear model calculations for all projectiles, nuclides and energies within a few days. This has been done with the TALYS code [7] to construct data libraries such as TENDL [8], which is available in both ENDF-6 format and "human-readable" x-y tables. An obvious application of such a complete calculated database is to compare it with EXFOR. This has been done for all nuclear reaction data for which this was possible. Since the XC4 database is not yet well sorted and processable in terms of secondary distributions (angular distributions, spectra), this comparison is currently only possible for cross-sections. With future versions of XC4 and the translation software, it is hoped that more data can be tested.

While this testing is done, a directory-structured projectile/nuclide/reaction database, which has the working name *Newbase*, has been created allowing easy access to the experimental data. The entire translation process is depicted in Figure 3.

Figure 3: Databases produced out of EXFOR



For each experimental energy point, we search for the corresponding energy point, using interpolation, in TENDL and provide a measure for the deviation.

First of all, we like to keep track of how many entries EXFOR contains, and how many can be processed by our system. This information is given in various files.

The main output file is *statistics*, which looks as follows:

```
        Newbase Statistics

Date of XC4 file              :  20100611
Time of XC4 file              :    150234
Time of X4 file               :  20100610

Number of EXFOR        entries:   19290
Number of XC4          entries:   12740
Number of EXFOR     subentries:  146980
Number of XC4       subentries:   75472
Number of obsolete  subentries:     212
Number of valid XC4 subentries:   75260
Number of Newbase   subentries:   73822
Number of Newbase   files     :   73822
Number of TALYS     files     :   27629
Number of XC4      data points:   7107688
Number of Newbase  data points:   7107538
Number of TALYS    data points:   4113822

XC4/EXFOR      conversion rate:   51.34848 %
Newbase/EXFOR  conversion rate:   50.22588 %
Newbase/XC4    conversion rate:   98.08929 %

TALYS/Newbase  comparison rate:   57.87970 %
```

This table shows, among other things, that only half of the EXFOR database has been translated into computational XC4 format, that almost the entire XC4 database has been translated into Newbase, and that we have been able to perform a TALYS comparison for more than 4 million corresponding EXFOR data points.

This global comparison obviously does not replace a "true" evaluation for one particular isotope, which involves careful study of all experimental work, precise nuclear model fitting, etc. However, it has already been shown on many occasions that TALYS provides very reasonable "blind"

estimates for many reaction processes, with the thermal and resonance range and fission as notable exceptions (the TALYS code system can also be used in that capacity, but only adjusted, not global, calculations bring the results somewhere near the experimental data for those processes). Hence, with the exception of certain reactions, TALYS should be able to give a reasonably good prediction of many reaction data, and obviously we will always try to extend such predictions to as many reactions as possible in future versions. At first glance, the problem is simple: if we know that TALYS is usually within e.g. 30% of the experimental data for a certain reaction channel, alarm bells should start ringing if the deviation of a data set for such a channel is suddenly much larger. We note that large deviations may also result from bad TALYS performance, even if the visual agreement on a linear scale is good. For example, for threshold reactions the difference between TALYS and experiment may easily be a factor of 10, close to threshold. In general the rule holds that the smaller the cross-section, the larger the relative error. It is therefore important to judge not only the calculation/experiment (C/E) values, but also the absolute deviation. In several cases, it turns out that there are problems in EXFOR, and many of them are not easily detected with means other than by comparing with a model code, which is why these EXFOR problems have not been discovered in the first place. The problems which are easiest to detect concern C/E values around 0.001 or 1 000, suggesting the well-known error of mistaking barns for millibarns. Unfortunately, the majority of cases is more difficult to judge. The current comparison may also help to solve one of the largest problems of EXFOR: reaction identifiers which are assigned in wrong, inconsistent or even multiple ways, which can be regarded as an "injustice" for otherwise good-quality experimental data. In other words, if TALYS is expected to give a reasonably good prediction for a reaction and we obtain a large deviation, it may be that we are not comparing the TALYS result with the correct quantity, and the EXFOR reaction identifier should perhaps be corrected.

To discover and classify problems, we use three goodness-of-fit estimators. If all three estimators are very large, this is an indication of something being wrong somewhere. They are the F-factor:

$$F = 10\sqrt{\frac{1}{N}\sum_{i}^{N}\left(\log\frac{\sigma_T^i}{\sigma_E^i}\right)^2} \tag{1}$$

the χ^2 estimator:

$$\chi^2 = \frac{1}{N}\sum_{i}^{N}\left(\frac{\sigma_T^i - \sigma_E^i}{\Delta\sigma_E^i}\right)^2 \tag{2}$$

and the absolute deviation:

$$\Delta = \frac{1}{N}\sum_{i}^{N}\left|\sigma_T^i - \sigma_E^i\right| \qquad (3)$$

In these equations, the subscript *T* stands for theory or TALYS and *E* for experimental. In all cases, we average over the number of energy points, *N*, in each data set. Hence, each EXFOR subentry (data set) that contains a cross-section excitation function, or only 1 point, is described by three average numbers: *F*, χ^2 and Δ, while we also keep track of all individual components F_i, χ_i^2 and Δ_i, in an extra column in the Newbase reaction database.

The F-factor is a kind of twisted $C/E = \sigma_T/\sigma_E$ value. In fact, each individual component of the sum inside *F* contributes to C/E if it is larger than 1, and E/C if it is smaller than 1. This is a more appropriate quantity than the average C/E, since averaging C/E values over many points may not be very meaningful if the individual values cross unity at some point. Eq. (1) remedies this. A value of $F = 1.2$ means that for the entire data set we are roughly 20% off on average. We use *F* as the leading indicator in our statistical study, i.e. we sort our results in order of increasing *F* to identify the worst cases. Another standard indicator is of course χ^2, but then the extra complexity is that apart from the central values the uncertainties given in EXFOR need to be reliable as well. This is a separate issue which will be addressed in a later section. Finally, large *F* or χ^2 values may be normal if the underlying quantities have a small value. To identify those cases, the absolute deviation in mb, Δ is helpful. In sum, it is best to look at all three indicators simultaneously.

As stated above, we have deviation factors for no less than 4 million individual data points. These can be grouped and analysed in many different ways. First of all, to get an overall view of the situation, for each reaction type we can distribute the range $F = 1\text{-}1000$ over 100 logarithmically equidistant bins. Hence, the first bin means that TALYS deviates between 0-7% from the experimental data, the second bin between 7-15%, the third bin between 15-23% and so on. All cases with $F > 1\,000$ are put in the last bin. The high peak at the lowest bins means good news for TALYS and EXFOR. The cases with very high *F*-values probably mean trouble for EXFOR (or XC4). The cases in between mean trouble for either TALYS or EXFOR, or both. For error determination in EXFOR, the interesting cases are in the tail of the distribution and it is probably best to start checking and working on the highest values. Note that there is always the possibility of an erroneous XC4 interpretation from our side, leading to a false alarm; hopefully this will improve over time. Figures 4 and 5 show the distribution of the *F*-values for all neutron-induced reactions, and (n,2n), (n,γ) and (p,n) reactions, that we managed to get out of the XC4 database.

Figure 4: Frequency table for the F-values for the cross-sections of all neutron-induced reactions and (n,γ) reactions

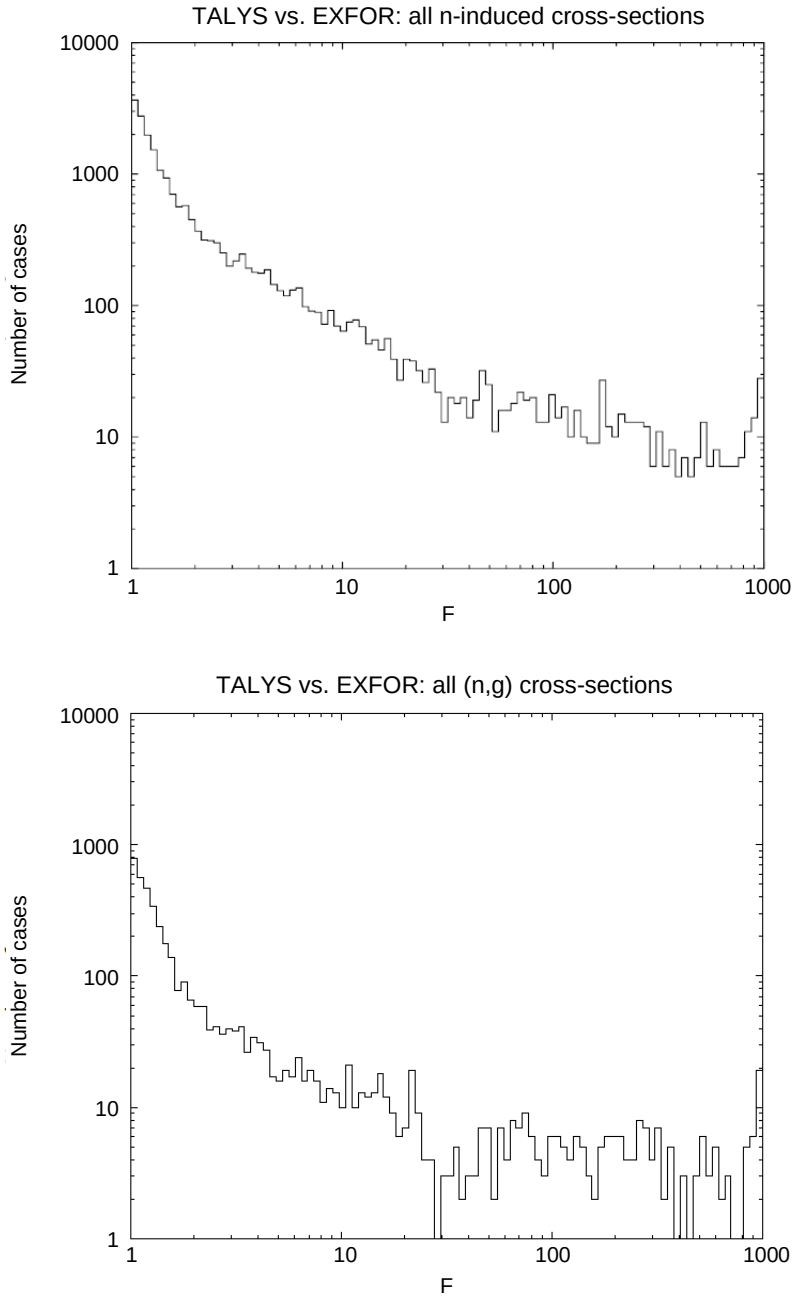
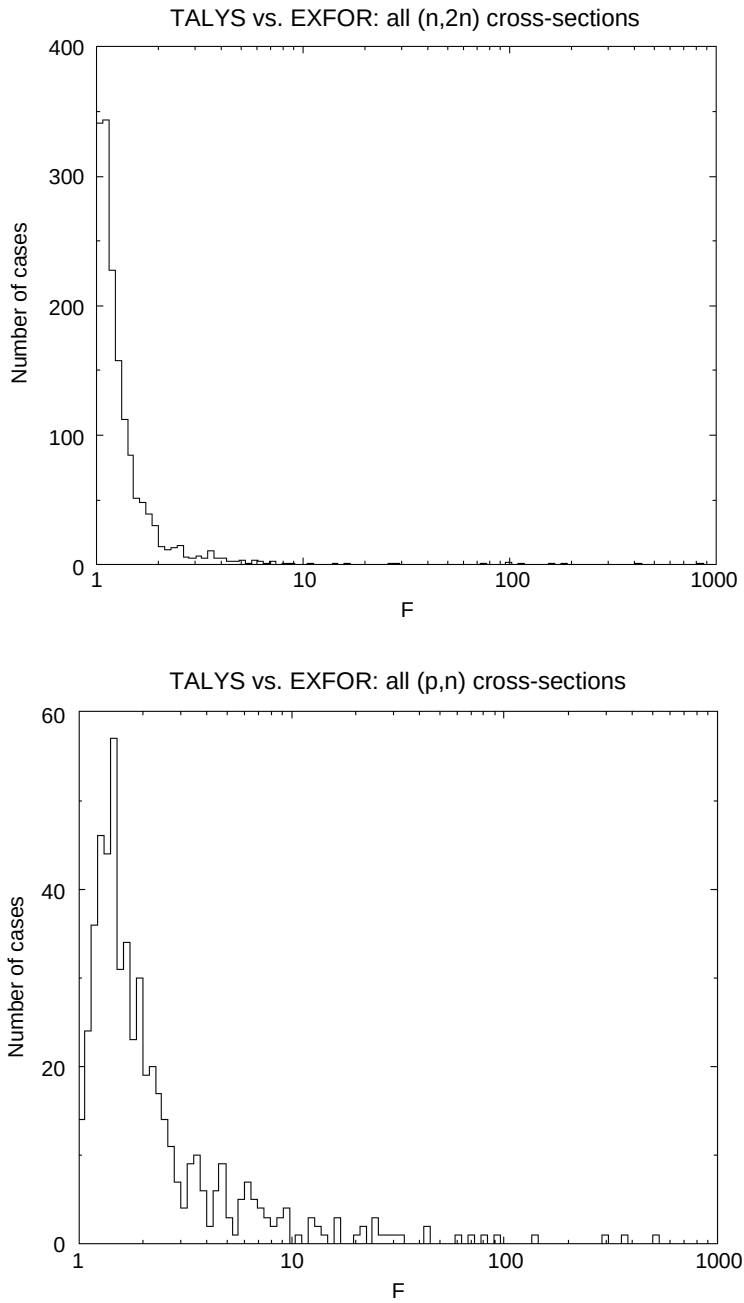


Figure 5: Frequency table for the F-values for the cross-sections of all (n,2n) and (p,n) reactions



A more direct way to trace the actual errors, or at least the suspicious cases, is to sort the *F* values per target isotope, projectile and reaction. This produces a large list of files, whereby each file directly shows the outliers from the average and from TALYS. As an example, here is the file with all ^{89}Y(n,p) reactions, in increasing *F*-order:

EXFOR	N	chi2	F	TALYS	Exp	Delta	E-min	E-max
40223020	1	5.255E-03	1.01	25.4	25.0	0.362	1.48E+01	1.48E+01
31532004	11	0.355	1.13	14.6	13.9	1.34	7.81E+00	1.47E+01
31494005	3	0.670	1.23	24.7	20.7	4.04	1.44E+01	1.47E+01
11462009	18	11.9	1.36	21.9	20.5	3.00	7.00E+00	1.98E+01
11504007	5	6.45	1.50	17.0	11.2	5.83	9.35E+00	1.40E+01
30115008	1	31.2	2.11	25.1	53.0	27.9	1.47E+01	1.47E+01

where all quantities are averaged over the *N* cross-section points in the indicated energy range. Clearly, cases that deviate significantly from the most occurring *F* values deserve closer inspection. Even if TALYS produces a bad prediction of a reaction that has been measured several times, outliers can still be identified since the *F*-values then tend to be clustered around a certain value, not necessarily close to 1.

The entire checking process produces lists, like the one above, for about 5 400 different reaction channels. Many of these have outliers, like the last experiment in the list above. All lists have been made available to the NRDC to allow further correction of the EXFOR database.

Besides the comparison with TALYS, statistical tests have also been performed on the experimental uncertainties. TALYS is not needed for this: one may simply analyse the uncertainties as given in the XC4 database. Per reaction channel, the experimental uncertainties for all entries are averaged per data set, sorted in increasing uncertainty and printed. Unrealistically small or large experimental uncertainties can then be identified upon closer inspection of the EXFOR file.

We note that the current version of TALYS is supposed to be only reliable up to 250 MeV, although some plans exist to significantly extend this energy range. In any case, as the number of high-energy experimental data sets is not that large, it seems obvious to repeat this exercise with an INC code to discover high-energy errors as well.

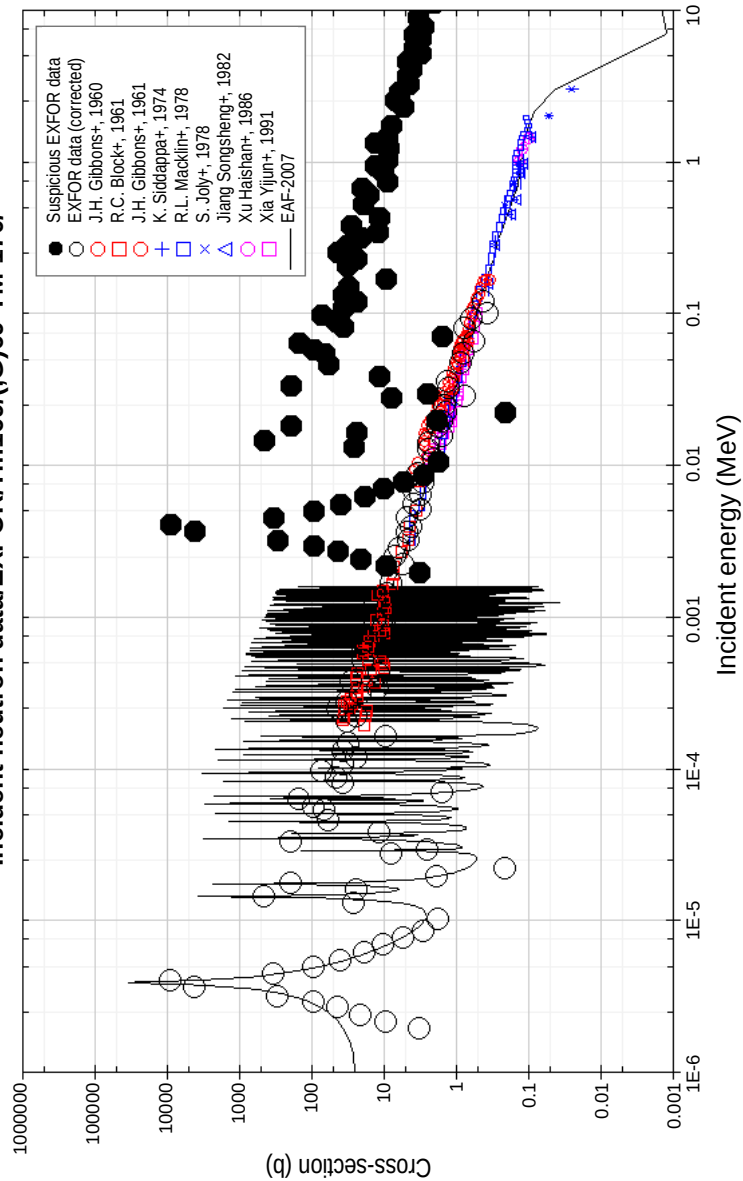
6. Visual inspection of experimental data

In principle, strongly discrepant data that are uncovered by statistical tests or nuclear model code comparisons will also show up as outliers when plotted against other experimental data. Actually, even without statistical goodness-of-fit estimators, it is already possible to reveal erroneous data. This can be done by directly plotting them against other experimental data and nuclear data libraries. This feature has been implemented in the JANIS [9] nuclear data viewer of the NEA Data Bank (Nicolas Soppera and Hans Henriksson) where an extension to JANIS has been created for systematic ENDF-EXFOR comparisons. An example can be found in Figure 6, where various experimental data sets for the same reaction are plotted together with values from nuclear data libraries.

Further, the possibility to read raw EXFOR data into JANIS and to produce C4 tabulated outputs has also been included. As with the statistical tests or nuclear model comparison this, along with the visual comparison, also leads to a list of possible compilation errors which have been sent to the NRDC for further study and/or correction.

Figure 6: Example of visual inspection of experimental data with JANIS

Note the apparent shift in incident energy for the filled circles



7. Other EXFOR tests

The most error-free EXFOR database can eventually be obtained by using as many different, independent checking methods as possible: through automatic chi-2 (or C/E) based comparison with models, statistical methods, visual inspection, etc., using many different translation and database management codes. In addition to the tests outlined in the previous sections, various other initiatives related to SG30 have been started or continued. Some examples are:

- At LLNL (David Brown), a python package x4i has been written to read EXFOR and to put it into python data structures. This route is independent from all the other routes that lead to a computational database, and therefore other types of errors were found and reported by x4i.
- At UKAEA (Robin Forrest) and JUKO research (Jura Kopecky) the SAFEPAQ system has been used to validate the European Activation File (EAF) with cross-section data from EXFOR. This revealed a set of errors/problems that have been reported to the NRDC.
- At KIT (Alexander Konobeyev) large nuclear model code comparisons for incident protons have been performed [10] for codes like ALICE, EMPIRE and TALYS, and statistical goodness-of-fit measures were determined for various subclasses of experimental data (per mass range, energy range, etc.). In total, this comparison comprised thousands of reaction channels. Obviously, such global comparisons only work if the EXFOR data are correct, and through these model comparisons various errors in the EXFOR database were discovered and reported to NRDC.
- Various individual error reports on high-energy data by CEA/DSM (Sylvie Leray, Jean-Christophe David), especially on data for residual products and double-differential spectra, by CEA/DAM (Helder Duarte) and LANL (Stepan Mashnik), were revealed by comparing experimental data with nuclear model codes. Specifically, this was observed when the IAEA benchmark of spallation models [11] was organised. It turned out that not only several EXFOR data sets at high energies are in error, but also various existing data sets have not yet been compiled. For the aforementioned benchmark, 39 experimental data sets turned out not to be compiled; fortunately in the meantime this has been done for 35 of them.

8. Correction of the errors by NRDC

During the lifespan of SG30 in total about ten batches of error reports have been sent to the NRDC. Most of the suspicious entries have been checked, and in about half of the cases they turned out to be wrongly compiled data. A list of entries can be found at *www-nds.iaea.org/nrdc/error/exfor_err3.html*.

NRDC contributed quite a lot to the extraction of potential EXFOR errors from the EXFOR/TALYS comparison (NEA contribution), as well as to finding actual errors by checking all publications (NDS-Otuka and NNDC-McLane). An unavoidable by-product of automated, global testing schemes is that not all suspicious data are actually errors. This means that NRDC had to invest a lot of effort in filtering the data to remove all false alarms due to format translation issues, weakness of the statistical tests, biases in TALYS, etc. So far, only very suspicious data have been reported to NRDC, i.e. the majority of cases concerned *F*-factors of five or more. Hence, half of the aforementioned suspicious data were not errors, but turned out to be (probably) poor quality data. At some point, SG30 follow-up activities will have to deal with *F*-factors lower than two, which probably means there will be an even more significant amount of poor quality data to deal with, relative to compilation errors. In this range of *F*-factors, we gradually enter the transition from EXFOR correction to data evaluation and that is where the current correcting activities should stop, although no one should be prohibited from making his or her personal quality flagging database. Anyway, it is clear that SG30 was only the "end of the beginning" of EXFOR correction and quality assessment, and that the future quality of EXFOR depends strongly on the resources that NRDC can invest in correction activities.

9. Summary and outlook

WPEC Subgroup 30 has contributed to the long-term objective to establish EXFOR as an easily accessible and correct database, with as most important deliverables an EXFOR master database from which many data and format errors have been removed, and a computational database that contains the entire EXFOR in tabular ("x-y-dy") format.

Several activities that have been launched during the lifespan of SG30 deserve to be continued, and they probably will be continued since most of the software is in place:

- Translation of the regularly updated EXFOR database into computational XC4 format to enable easy further processing and use. (The distribution should be agreed upon with NRDC; regular translation should be requested from the IAEA-NDS.)
- Automatic test of the XC4 database by translation into a directory structured database categorised by projectile/nuclide/reaction and automatic comparison with TALYS/TENDL.
- Efficient testing with JANIS.
- Construction of an "EXFOR correction system" operating outside the EXFOR library (under construction at the IAEA-NDS).
- An increased flow of emails from EXFOR users to the NRDC with error messages.

Thanks to, or at the least during, the lifespan of SG30, EXFOR has significantly improved in quality; however, some important unfinished actions remain to be addressed:

- Ensuring that the same quantities all have the same reaction identifier (requires case-by-case checks): a key step to efficient data evaluation and model code testing.
- Secondary distributions like differential and double-differential spectra, angular distributions, gamma-ray production cross-sections, have not yet been tested with TALYS/TENDL, due to sorting problems in XC4. (We note however that these data are included in

XC4 and have been tested for many years by users under the IAEA Web Retrieval System, Sigma/NNDC, EMPIRE and EndVer packages. This means it is feasible.)

- A wealth of statistical info on EXFOR problems is available, but this has not yet been translated into a plan of attack for large-scale EXFOR correction. This will be taken up by the NRDC. Arguably the best way to proceed is to perform this correction reaction class by reaction class. For example, many activation cross-sections may need to be revisited. The checking procedures outlined in this report have revealed many cases where isomeric and total cross-sections were interchanged, and the only way to correct that is to revisit the EXFOR entries one by one. Other classes of data should be visited in the same way.

SG30 has also given rise to other initiatives to improve EXFOR. For example, an issue under study is the possibility to consistently distinguish between systematic and statistical uncertainties (Otuka, IAEA). The present computational C4 format should be extended to accommodate partial uncertainty information consistent with that available in EXFOR. In addition, a new format should be developed, capable of accommodating experimental covariance information.

Another ambitious plan developed at the IAEA is to extend the XC4 format to enable reproducible corrections to the experimental data, such as re-normalising the data with constant or energy-dependent factors, or to more recent standards data, quality flagging, etc. In this way, a "derived" experimental database can be made out of EXFOR, which is trusted by at least the people who produced it.

The most important objective of SG30 was to help provide an EXFOR database with significantly less data errors and format errors than was previously the case. The translation into the XC4 database and the testing methods described in this report have accelerated the discovery of such errors. This benefits both case-by-case (web) retrieval of data and automated use of the entire database. New, corrected, versions of EXFOR can henceforth always be retested with the software described in this report.

References

[1] Experimental Nuclear Reaction Database EXFOR.

[2] Holden, N., "CSISRS History", BNL document, *www.nndc.bnl.gov/exfor3/exfor00.htm*.

[3] Otuka, N., *et al.*, "The Role of the Nuclear Reaction Data Centres in Experimental Nuclear Data Knowledge Sharing", *Proc. of the International Conference on Nuclear Data for Science and Technology*, 26-30 April 2010, Jeju, Korea, forthcoming; and *www-nds.iaea.org/nrdc*.

[4] Cullen, D.E., A. Trkov, *Program X4TOC4*, IAEA-NDS-80, IAEA Nuclear Data Section (2001).

[5] Dupont, E., A.J. Koning, N. Otuka, "Exploratory Data Analysis of the EXFOR Database", *Proc. of the International Conference on Nuclear Data for Science and Technology*, 26-30 April 2010, Jeju, Korea, forthcoming.

[6] Heckert, A., J. Filliben, "DATAPLOT Reference Manual", *NIST Handbook 148*, National Institute of Standards and Technology Handbook Series (2003).

[7] Koning, A.J., S. Hilaire, M.C. Duijvestijn, "TALYS-1.0", *Proc. of the International Conference on Nuclear Data for Science and Technology*, O. Bersillon, *et al.* (Eds.), Nice, France, 22-27 April 2007, EDP Sciences, p. 211.

[8] *TENDL-2009: Consistent TALYS-based Evaluated Nuclear Data Library Including Covariance Data*, A.J. Koning and D. Rochman, Nuclear Research and Consultancy Group and NEA/WPEC, Nov. 2009, Ref. JEF-DOC-1310.

[9] Soppera, N., *et al.*, "Recent Upgrades to the Nuclear Data Tool JANIS", *Proc. of the International Conference on Nuclear Data for Science and Technology*, 26-30 April 2010, Jeju, Korea, forthcoming; and *www.nea.fr/janis*.

[10] Konobeyev, A.Yu., *et al.*, "What Can We Expect from the Use of Nuclear Models Implemented in MCNPX at Projectile Energies Below 150 MeV? Detailed Comparison with Experimental Data", *Proc. of the International Conference on Nuclear Data for Science and Technology*, 26-30 April 2010, Jeju, Korea, forthcoming.

[11] *www-nds.iaea.org/spallations*.