Developments of the EXFOR database: possible new formats

R. A. Forrest,[1,*] V. Zerkin,[1] and S. Simakov[1]

[1] *International Atomic Energy Agency, PO Box 100, Vienna-A-1400, Austria*

(Dated: February 26, 2013)

The EXFOR database is a collection of experimental nuclear reaction data, maintained by the IAEA on behalf of the International Network of Nuclear Reaction Data Centres (NRDC). The format for the storage of such data was first described in 1969 and while there have been many incremental changes over the years so that the format is now capable of containing a very wide range of measurement results, there is a growing realisation that a major change is required. Consequently the IAEA NDS organised a Consultant's Meeting on 'Further Development of EXFOR' in March 2012. This was an opportunity for a range of international experts to discuss ways of improving EXFOR and while this focused on new formats there was also discussion on ways of storing new data, new output formats and software tools such as editors. This paper will discuss recent and proposed changes to enable new quantities to be stored (such as coincidence measurements and covariances), the range of output formats available (such as C4 and X4+) which make interaction with the data more user friendly and the possible use of XML to modernise the database.

I. INTRODUCTION

EXFOR shares with the other iconic term of nuclear data, ENDF, the property of referring to many things; the data, a format, documentation, an organising committee and also a venerable history. The EXFOR database contains essentially all the experimental reaction data that has ever been measured, at least within a well-defined range of parameters and particularly for neutron-induced reactions, although great progress has also been made with the completeness of charged particle reactions. The format reflects the fact that data are typically extracted from research papers and its hierarchical structure has proved to be well suited to the representation of such data. Compilation is carried out worldwide, the various geographical regions represented by national data centres. The coordination of this work is carried out by the IAEA NDS acting through the International Network of Nuclear Reaction Data Centres (NRDC) [1, 2]. The purpose of this paper is to present some of the technical details of the format and structure of EXFOR, describe the tools used by both compilers and users to interact with it, but mostly to focus on whether we are nearing the time for major changes and point out the problems of a such a change and the benefits that would arise. The answers are not given here; but we hope it will help in the debate which has started about new structures and representations and how EXFOR fits in with similar changes being discussed for the evaluated ENDF files.

* Corresponding author: r.forrest@iaea.org

II. EXFOR AND ITS BRIEF HISTORY

EXFOR stores a large variety of nuclear reaction quantities such as integral, differential and partial cross sections, angular and energy distributions of secondary particles, polarization data, resonance parameters, fission neutron multiplicity and total nu-bar, fission product yields, energy averaged cross sections and reaction rates.

The present policy of compilation makes it compulsory that all published data for neutron, light ($A \leq 12$) charged particle induced reactions, photo-neutron and photo-fission reaction data up to incident energies 1 GeV be included, other data can be compiled on a voluntary basis. Currently EXFOR contains measured results for about 20,000 experiments which provided more than 150,000 datasets. The EXFOR library continues to evolve to meet the needs of diverse user communities. Currently the content grows by 500-700 new Entries every year.

A decision was taken in 1969 [3] to create the EXchange FORmat as the common format to exchange and store neutron-induced reaction data. This was based on the existing formats (SCISRS, NEUDADA and DASTAR) used by Centres. The EXFOR format was originally created to be an 'exchange format' to share data between different nuclear data centres with different information systems and missions. Its design reflects both the type of nuclear reaction data most important in the late 60s, as well as the types of IT technologies available at that time, in particular the method of data storage on punch cards.

In 1976 charged-particle and photon induced reaction data was added to EXFOR and many specialised Centres joined the NRDC Network, it was further agreed that compilation work was to be coordinated by NDS. By 2005 NDS carried out a merging of the EXFOR libraries from all data centres into one central master library.

One lesson that can be learnt from the history of the EXFOR project is that technological changes tend to bring many additional benefits not originally foreseen. The change of the computer platforms used in the data centres (*'Migration'*) from VMS/Fortran/Oracle-DBMS to modern systems at the beginning of 2000 caused a revolution in EXFOR software, with multi-platform systems based on Java and relational databases achieving finally a higher quality of nuclear data services [4]. This included the finding of universal solutions for Web, CD-ROM and application systems, and also making possible the unification of the contents of EXFOR libraries from different data centres in the master library (*'Merging'*), organizing delivery of full contents of master EXFOR library to users with systematic tests and correction of mistakes (which had existed for decades) in the contents. It is extremely likely that a change of format to XML would also have positive side effects unforeseen at this time.

III. EXFOR TODAY

A. Structure of EXFOR file

Conceptually, the EXFOR formatted file contains information logically organized as nested and repeated text blocks. This is shown in Figure 1.

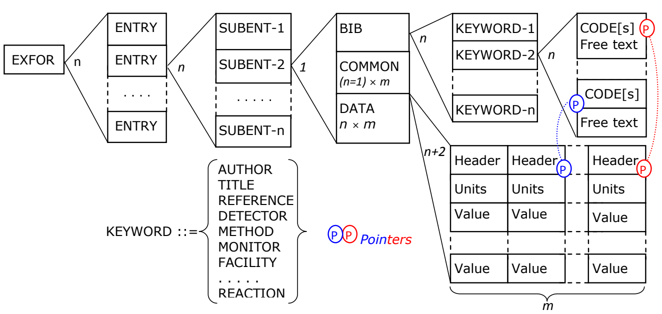

FIG. 1. Structure of an EXFOR file.

This structure was designed to reflect the logic and sequence of information in a publication of experimental data typical in the literature, storing bibliographical information, experimental conditions and the reaction description in a set of keywords in the BIB section, common parameters such as angle or energy level of the product nucleus in the COMMON block and the actual data in a DATA block. This structure is appropriate and does not need significant revision, since it makes the compilation process natural, helping to minimize errors and simplifies cross checking. For example, using pointers to specify different reactions just corresponds to a data table with a common energy grid for different reactions.

In fact, this structure is very similar to the hierarchical logic of XML documents and the data structures in modern programming languages such as C or Java. The extensive system of dictionaries gives an additional level of flexibility allowing extension of codes without any change in software. We have to admire the job done by the EXFOR designers, since the format, compilation system and library have survived almost 50 years.

Low-level structures of EXFOR format do however have a rather old fashion design: parameters are defined by position; data values have fixed length, limit of number of variables in the data line and an 80 column limitation. This makes programming of almost any task using EXFOR from scratch very time consuming. Some of these difficulties can be resolved by software which presents EXFOR information in modern IT data structures.

B. EXFOR and ENDF

Many of the major differences between EXFOR and ENDF formats come from their initial purposes: compilation of experimental data and exchange between data centres (EXFOR) and data evaluation as input to further calculations (ENDF). Essentially the EXFOR format was designed for human-human data flow and ENDF for computer program-computer program interaction.

Reaction-code in EXFOR consists of nine sub-fields SF1(SF2,SF3),SF4,SF5,SF6,SF7,SF8,SF9 that mean: *Target(Projectile,Ejectile)Product,Branch,Quantity,* etc.

The top-level structure of ENDF [5] is based on a TAPE containing many materials (MAT) interacting with a projectile (NSUB). The data are organised in Files (MF) containing Sections (MT) and optionally the product nuclide can be described by a subsection (ZAP). Thus ENDF describes a reaction by the set [NSUB/ZA/MF/MT]. As a simple example:
EXFOR: 13-AL-27(N,A)11-NA-24,,SIG
ENDF: ZA=13027. LISO=0 NSUB=10 MT=107 MF=3

In addition to the native formats used for ENDF and EXFOR it has been found that modern database management systems provide very effective means of storage, search and manipulation of these data. The database design used by both NNDC and NDS was defined in 2000-2004 and has been developed since then.

Using a database as the storage mechanism also enables various extensions of the original libraries, such as: links between databases, information imported from other databases such as NSR, links to the web using DOI, details of PDF files of original publication and various derived data sets.

Today, to understand all details of the EXFOR format requires 400 page manuals and the use of 420 page dictionaries. With its rigid structure this makes editing of EXFOR files difficult especially for a new generation of compilers. For users as well the codes and abbreviations make the EXFOR file cryptic and inconvenient. These two problems can be solved by introducing intermediate level software, namely EXFOR editors for compilers and software providing end-user with interpreted outputs oriented to various user communities. Data in the original EXFOR format are still available as required, but they are used mostly by specialists to investigate problems.

C. Existing software tools

There are several software tools used for various stages of the EXFOR 'life cycle', and these originate from different data centres. Some of the tools are in common use; others are used only in originating centres. NRDC has no policy concerning EXFOR software, centres are free to donate their software, organize collaborations, provide sponsorship and exchange software as required. This has the consequence that there are no formal specifications, validation or maintenance plan and most importantly no resources for software tools.

In the absence of NRDC guidance, IAEA has taken the lead providing the Web retrieval tools, including visualisation and the checking codes. The other main tools are the EXFOR Editor [6] and data digitizing programs [7].

One of the most important tasks is dissemination of EXFOR data; this is the original purpose of the NRDC activity and the final goal of EXFOR compilation. NDS has well established and long standing co-operation with NNDC, with the EXFOR-CINDA-ENDF system available on both web sites. Dissemination is web based but also includes distribution of stand-alone CDs. Recently developed Web tools [8] are: interactive construction of correlation matrix using partial uncertainties from EXFOR [9]; automatic re-normalization of EXFOR data using old monitors and new standards; access to original PDF article (for authorized users) and a collection of video-guides with 'how-to' instructions.

D. Output formats

The NDS Web retrieval system provides end-users with output in various formats. These consist of: Raw EXFOR (plain text); Interpreted EXFOR (X4+, X4±, T4 and X4XML) which replaces many abbreviations and codes and so is more readable; Standard-Output (plain text and XML) representing EXFOR with interpreted information without Pointers and with data columns in sorted order; and Computational formats (C4, TABLE and XREF, C5, C5M) presenting various computer readable data tables.

It should be noted that XML representations of EXFOR have existed since 2009 and have been available on the web since 2011. Two forms were developed: the first (X4XML) is logically equivalent to raw EXFOR, but is extended by information from dictionaries and uses named attributes for all coded information. The second removes pointers and presents all data tables in self-consistent forms and is suitable for advanced programming tasks. X4XML could be used as a prototype for a future version of EXFOR.

IV. FUTURE OF EXFOR

A. Further Development of EXFOR Meeting

To assess needs and ways for improvement and development of the EXFOR project a Meeting, 'Further Development of EXFOR', was organised by NDS in March 2012 [10] involving nine participants.

A particular problem concerns ways of coding complex reactions into the REACTION string. An increasing number of data are obtained by multi-detector and coincidence measurements. The physical quantities typically reported are fission products correlated with kinetic energies, sequential, break-up and multi-channel reactions at medium and high energies.

The present structure of EXFOR is ill suited for the coding of complex reactions. It uses a rather archaic system of codes and their combinations in the SF5 to SF8 sub-fields of REACTION string. These codings and their combinations attempt to represent an increasing diversity of experiments. Today the most complicated dictionary (#236) contains more than 1100 combinations of subfields 5 to 8 and continues to grow. The codes have short names, are not self-explanatory, the descriptions are limited and it is often difficult to find unique definitions of the reaction type.

New schemes for complex data compilation were discussed based on a more detailed description of the reaction string. The main conclusion was to keep the EXFOR exchange format essentially untouched in the short term. To reduce difficulties the tools should be developed and the default output of data should not be raw EXFOR, but rather the HTML versions X4+ or X4±.

B. XML as the basic exchange format

Today there is no crisis in the EXFOR system forcing the EXFOR community and NDS in particular to urgently migrate to a new format. For compilers the problems are solved by using EXFOR editors. For advanced programming the system offers users two types of XML output files. Many fundamental problems of extension of the EXFOR format and system have been successfully solved by 'clever' software either by introducing new conventions within free text (DOI's, covariance data), or by using the relational database approach importing information from external files (binary PDF files) and databases, but keeping the EXFOR format, the dictionaries and much of the old software unchanged. Such an approach can continue, probably into the next decade, but it will become increasingly difficult and eventually will require essential modernization.

A change of format requires at least: new internationally agreed exchange format and rewriting of almost all programs dealing with EXFOR files. Taking into account efforts/cost of such migration, perturbations due to modernization going in parallel with existing and working system, it seems that straightforward migration to a new exchange format has little benefit since the new system will have comparable functionality with the existing one. Nevertheless, in a long term perspective, such a migration is unavoidable - the main reasons for this are human factors and maintenance cost.

The changes described above do not include modernization of two fundamental parts of the EXFOR system: the principles of coding of reactions integrated with the logic of dictionaries and with checking and conversion codes. The system is very complex, all details of its construction are not fully documented and some have been lost with the retirement of the original designers and its revision needs much more effort then a simple change of form from text to XML.

In summary the basic advantage is a healthier system with the potential for future development. The disadvantage will be XML files with much optional text ('repeating garbage'), dependence on software and limited advantages for end-users. Basic problems are time, resources and a clearly formulated motivation. Initial tasks must be to conclude the necessity of migration, organize an international project, carry out a deep study of the relevant IT-technologies, prepare a realistic plan, and finally to achieve its practical implementation.

Since 2006 there have been efforts in LLNL (Livermore, USA) to design an XML format for evaluated data, to develop software converting ENDF and ENDL formats to the XML format termed Generalized Nuclear Data (GND) [11], and to adopt processing codes for the GND format. Eventually the GND structure based on the XML language could be further developed to also apply to EXFOR and nuclear structure and decay data from ENSDF. This is an extremely ambitious task that will require collaboration from different nuclear data networks, institutions, scientific groups and individuals working in the different fields. In order to determine the feasibility of such a vision a new WPEC Subgroup 38 was formed and started working in 2012. NRDC delegated NDS staff to participate in this activity in order to exchange experience in the development of new formats, taking into account EXFOR needs and to help design the common parts of the format which can be used in software dealing with data from the three types of data libraries.

V. CONCLUSIONS

The EXFOR format works well, and in no sense is it 'broken', but it has reached the stage where great ingenuity is required to add in representations of new data types. As argued above it is really a 'hidden' format that users and even compilers do not need to be fully conversant with since compilers can interact with an editor, shielding them from the format and allowing concentration on the essential physics, while for users a wide range of 'friendly' outputs are provided, meaning that deciphering the codes and abbreviations of EXFOR is not necessary. Even a discussion of a format change means that we are forced to consider in detail how tools such as editors, data digitisers, web retrieval, visualisation and checking are specified, maintained and resourced. The conclusion of a recent Technical Meeting on the future of EXFOR was that an immediate change is not necessary, but it encouraged the use of the existing XML output formats and urged comparison with the GND XML structure to ensure commonalities of approach.

It has been stressed that NDS has prepared two output representations of EXFOR data in XML, the first that keeps the structure but just translates into XML, while the second addresses some of the issues that makes the output confusing to users and difficult for computers to read by removing pointers and common blocks and ensuring that data tables are self-contained. As a participant in WPEC SG38, NDS is committed to help improve a common XML structure and ensure that if NRDC wishes to change the EXFOR format that a suitable option is available.

The data community and the funding bodies need to realise that while changes to a new structure such as XML can bring many benefits and ensure that the data field can attract a suitably trained new generation of scientists, it will not be cheap. In particular many legacy codes will need to be refreshed and many areas that we rely on, but do not invest in, will need to be recognised and funded. This also comes at a time when security issues, rigorous specification and change control are being insisted on by organisations and since change will be forced on scientific software, we should ensure that we get some benefit at the same time by changing our structures and methods moving towards a new phase of nuclear data.

[1] N.Otsuka, S.Dunaeva, INDC(NDS)-0401, IAEA, (2010).
[2] N.Otuka, *et al.*, this conference.
[3] A.Lorenz, INDC(NDU)-16/N, IAEA, (1969).
[4] V.V.Zerkin, *et al.*, ND2004, pp. 586-589, (2004).
[5] M.Herman, A.Trkov, ENDF-102, BNL, (2009).
[6] G. Pikulina, S. Taova, http://www-nds.iaea.org /nrdc/nrdc_2012/present/pikulina.pdf, (2012).
[7] N.Otsuka, V.Semkova, INDC(NDS)-0629, IAEA, (2013).
[8] IAEA Nuclear Data Services, EXFOR Retrieval System, http://www-nds.iaea.org/exfor/, (2000-2012).
[9] V.Zerkin, EPJ Web of Conferences **27**, 00009 (2012).
[10] D.Brown *et al.*, INDC(NDS)-0619, IAEA, (2012).
[11] C.M.Matoon *et al.*, NUCLEAR DATA SHEETS **133**, 3145 (2012).