EPJ Web of Conferences **292**, 12003 (2024) https://doi.org/10.1051/epjconf/202429212003
NRM2023

Developing a new web service for experimental nuclear reaction database (EXFOR) using RESTful API and JSON

Shin Okumura[1,*], *Georg* Schnabel[1,**], and *Arjan* Koning[1,***]

[1]NAPC-Nuclear Data Section, International Atomic Energy Agency, Vienna International Centre, 1400, Vienna, Austria

Abstract. Efficient data mining from the Experimental Nuclear Reaction Database (EXFOR) has a potential for utilization of modern computational analysis techniques to find trends, shortcomings and hidden patterns in the database, which in turn helps improve our knowledge of nuclear physics. To facilitate the data mining, we have developed two EXFOR parsing computer programs (EXFOR Parser) to convert the data in the database stored in the EXFOR format into the widely adopted JSON format. The converted JSON data are used for further processing to extract individual physical observables and generate tabulated data (x, y, dx, dy) where all units of measurement are standardized. Furthermore, we have developed REST APIs and an open web system for easy access and quick visualizations of these converted datasets.

1 Introduction

Experimental nuclear reaction data are essential for understanding nuclear reaction phenomena, developing nuclear theories and models, and evaluating data for nuclear data libraries. The Experimental Nuclear Reaction Database (EXFOR) [1] comprises both numerical data, i.e. measured cross sections and related quantities, and metadata with details about the experiments, such as the researchers involved and the facility. The compilation of experimental data and metadata into the *EXFOR master files* is undertaken by the International Network of Nuclear Reaction Data Centres (NRDC) under the auspices of the International Atomic Energy Agency (IAEA).

Despite the availability of both metadata and the numerical data in *EXFOR master files*, programmatically searching for and retrieving relevant data is often cumbersome, as the EXFOR format was designed for the transmission of nuclear reaction data in ASCII text among data centers within NRDC decades ago when computing capabilities were limited. Since then, the landscape of information technology has significantly changed and broadly accepted formats emerged for storing and submitting data over the web. Because the EXFOR format is not among those commonly accepted formats outside the nuclear data community, mundane tasks, such as the extraction of data from *EXFOR master files* for plotting or using them in statistical analyses and for Machine Learning (ML) become difficult with programming

*e-mail: S.Okumura@iaea.org
**e-mail: G.Schnabel@iaea.org
***e-mail: A.Koning@iaea.org

languages and tools commonly used for data science tasks, such as Python and R and their respective package ecosystems.

This barrier in dealing with the information given in EXFOR is unsatisfactory as we think that there is a large untapped potential of applying ML algorithms to the database, leveraging its completeness and broad coverage of nuclear reaction observables. The EXFOR includes nuclear reaction cross sections, angular distributions, energy-angle distributions of emitted particles, resonance parameters, fission product yields, multiplicities for particle production, among other more complex reaction observables. Evaluated values of those quantities are found in nuclear data libraries, which are stored in the ENDF-6 format [2]. The richness and variety of nuclear reaction observables makes the entire EXFOR a very valuable resource, yet very complicated.

The IAEA Nuclear Data Section (NDS) is entrusted with the responsibility of maintaining and facilitating user-friendly access to this data. To fulfill this mandate, the NDS has developed several services, such as the EXFOR web retrieval system [3], which allows the retrieval of data from EXFOR in various formats, such as JavaScript Object Notation (JSON) and the tabular format C4 and C5. We anticipate that with the rapid advances of compute infrastructure and the increasing demand to process nuclear data at scale in the context of ML and AI applications, additional flexible retrieval options and open-sourced programs distributed under a permissive license will be helpful to the nuclear data community and facilitate interdisciplinary collaboration among researchers and scientists in the fields of nuclear physics, nuclear data, and nuclear applications.

The use of scientific data in ML and AI algorithms is revolutionizing a wide range of research fields such as design and discovery of medicine [4] and material science [5], to name just two examples. The fields of nuclear physics and nuclear data are also not exceptional in this regard as ML techniques have already started to guide our next experimental, theoretical, and evaluation efforts [6–13]. In 2021, a new Subgroup (SG50) on the topic of "Developing an Automatically Readable, Comprehensive and Curated Experimental Reaction Database (SG50)" [14] was initiated within the Working Party on International Evaluation Cooperation (WPEC) under the Nuclear Energy Agency (NEA). One goal of SG50 was to establish the requirements for an automatically readable, comprehensive and curated experimental nuclear reaction database.

To enhance the accessibility and openness of EXFOR datasets with regards to both the data themselves and also related codes in order to adhere to the requirements established in SG50, (1) open-source EXFOR parsing software to convert the existing EXFOR format into a more computationally accessible format and (2) a web service that implements the FAIR principles (Findable, Accessible, Interoperable, Reusable) [15], are required.

To work towards fulfilling these requirements, we have developed two EXFOR parsing computer programs (EXFOR Parser) in Python to convert data given in the EXFOR format to JSON. These programs co-evolved and mutually benefited from each others development. One parser was conceived as a prototype implementation within SG50 and the other one was conceived as an NDS modernization project. Having said that, both parsers are based on similar design concepts. They both simplify *EXFOR master files* by splitting them into separate files so that only one physical observable type is stored per file. Other simplifications are also performed, such as untangling nested structures, reallocating common experimental condition information, splitting the string representation of reactions into several subfields, and converting all values to common measurement units. Apart from the JSON representation, these parsers also can extract the numerical data in tabular form.

Additionally, we developed REST APIs adhering to FAIR principles so that scientists can programmatically access a large number of datasets for data mining and ML purposes.

Building upon these REST APIs, we have constructed an open web system designed to enable the quick and convenient visualization of the experimental data.

2 Nuclear Data Pipeline

The term *Nuclear Data Pipeline* is often used to describe the flow of data obtained in experiments through several stages until the final data can be used by end-users in nuclear science, engineering, and applications. The stages are (1) the compilation of experimental data and ingestion into EXFOR or similar databases, (2) the evaluation of experimental data and the production of evaluated nuclear data files in ENDF-6 [2], GNDS [16] and related formats, and (3) the processing of evaluated files into application formats understood by transport codes, and (4) the verification and validation of those application files. Figure 1 shows a schematic view of the nuclear data pipeline.

The evaluation stage is a particular complex process because it requires knowledge of both experimental and theoretical nuclear physics to select the best available experimental data and combine it consistently with nuclear models. The resulting evaluation must then be stored in a file adhering to the ENDF-6 format [17]. In order to do this, users (evaluators) must understand both the EXFOR and ENDF-6 formats, which takes some time until sufficient mastery. After downloading *EXFOR master files*, many users probably need to reformat them in a semi-automatic process before they can efficiently work with the data. The manual intervention required is certainly an obstacle, which hinders the flow of data throughout the nuclear data pipeline and also prevents efficient use of the data of the EXFOR for other purposes.

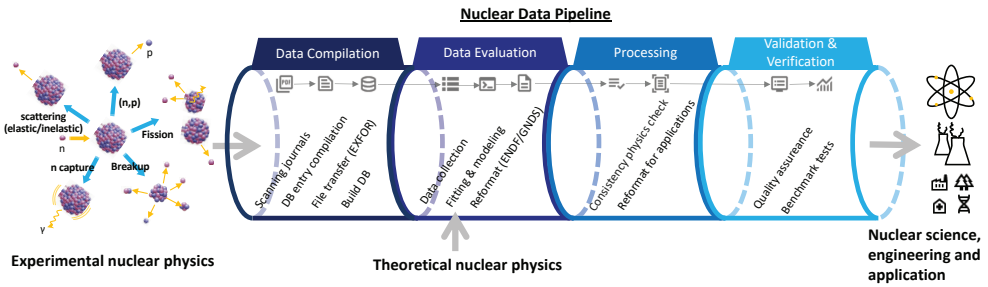

Figure 1. Schematic illustration of the nuclear data pipeline.

3 Code implementation

3.1 EXFOR Format

EXFOR is an ASCII text file and its format is still constrained by the legacy of punch card technology, where each line is limited to 66 characters. This format is characterized by the document-oriented information with a nested structure as shown in Fig. 2. In this structure, each entry, denoted as an `ENTRY`, corresponds to a single research publication. An `ENTRY` can contain multiple `SUBENT` (subentry) items. `SUBENT` 001 typically stores only metadata (bibliography information) defined under each information identifier in `BIB` section using codes such as `TITLE`, `AUTHOR`, `INSTITUTE`. `SUBENT` 001 is followed by other `SUBENT` 002 -

999, each of which contains at least one REACTION field that defines the meaning (observable type) of the numerical data in COMMON and DATA sections.

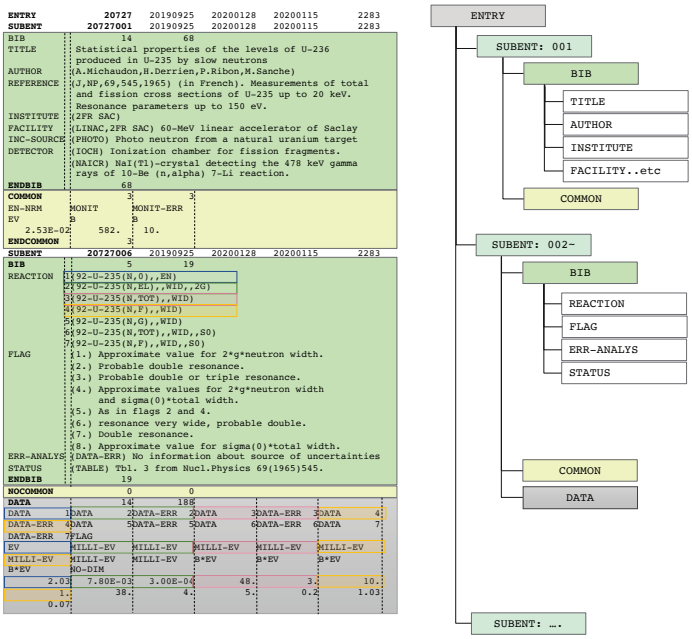

Figure 2. An example of EXFOR entry that contains the COMMON block in SUBENT 001, SUBENT 006 contains multiple REACTION items, and DATA block has 8 columns with word wrapping at the end of 66 characters in each line. Note that some parts of this entry are omitted for simplicity.

In the BIB section, information-identifier keywords must end within the first 10 characters. A POINTER at the 11th character in each line links the specific information in the identifier to the column of data. Characters in 12-66 may contain the coded information as well as accompanying free-text descriptions. If the opening parenthesis appears in the 12th character in any line, coded information is provided until the complementary closing parenthesis appears. The coded information is the keywords to describe the experimental conditions, such as 'REAC' meaning the measurements have been done in reactors. Such keywords have been introduced to reduce the total data size for transmission and those descriptions can be found in the relevant section of EXFOR dictionary, which is not distributed to users and is used only for data compilation. Similarly, many entries have numerical data with special expressions such as '1.000+3' (without exponent 'E') or '.3' (without the digit before the decimal point) that are also allowed to reduce total data size.

The COMMON section contains the common numerical values of experimental conditions such as incident particle energy applicable either to all entries or to specific ones. The DATA table, numerical data and the main part of EXFOR, and COMMON sections have specific limitations. Each column is restricted to 11 characters, and a single line can accommodate a maximum of six columns. Consequently, when the number of columns exceeds six, line breaks are introduced, which worsens readability as demonstrated in Fig. 2.

The EXFOR compilers have the discretion to add supplementary information in a free-text format with various identifiers, decide on the specific structure of ENTRY and how to accommodate numerical data in both COMMON and DATA sections. The data units used in

COMMON and DATA must follow the units reported in the original paper, therefore, different subentries that contain the same measured quantity may use different units, such as milli-barn and micro-barn.

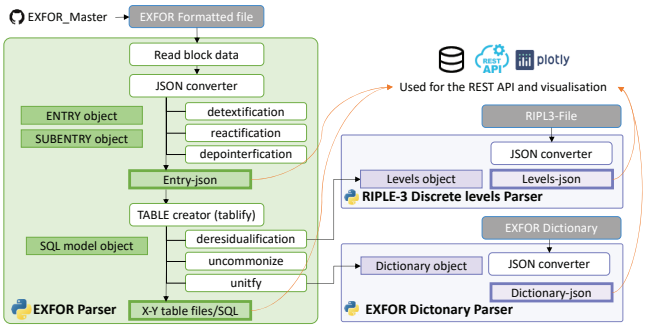

Figure 3. The schematic illustration of the EXFOR Parser, which reads EXFOR entry files, systematically breaks them down into individual physical observables, generates tabulated (X-Y) data, and stores the data in both text and SQL formats.

3.2 EXFOR master files

The availability of all compiled *EXFOR master files* for download may help users who want to design their own data extraction pipelines. We maintain these *EXFOR master files* on the Git service, available at https://github.com/IAEA-NDS/exfor_master/, making them easily downloadable.

3.3 JSON converter

Two of our EXFOR parsers involve notably similar data processing procedures. The data processing sequence is outlined in Figure 3. The EXFOR parsers first download all *EXFOR master files* to process. For each EXFOR entry, the entire ASCII file is loaded into corresponding 'entry' and 'subentry' objects within a Python program. The process initially involves reading each identifier as a block and storing it within a Python dictionary object, with the identifier serving as the key. Subsequently, the process proceeds to separate coded information from free-text descriptions within the metadata (detextification).

To identify the smallest physical observable, one needs to correctly parse the REACTION and break it down into its constituent subfields (reactification). In the simple case where an EXFOR entry comprises the smallest dataset, data headers, units, and numerical data in the DATA table are stored in a Python dictionary object. In cases where POINTERs are specified within the REACTION, the DATA table will be decomposed based on the POINTERs (depointerfication). Once EXFOR entry is successfully converted into JSON format, the data becomes considerably more manageable and amenable to manipulation.

3.4 Table creator

The difference between the definition of "dataset" as employed in EXFOR and the broader nuclear physics context lies in the granularity of observation. In nuclear physics, a dataset represents the smallest set of a single observable derived from a specific nuclear reaction

involving a target nucleus, an incident particle with a particular energy, and under specific conditions. An EXFOR subentry may contain several unsorted datasets such as multiple incident energies, angles, and/or reaction products. For instance, an EXFOR subentry might contain neutron inelastic scattering cross sections for excitation energies ranging from 0.845 to 10 MeV of ^{56}Fe as a function of neutron energy ranging from 1 to 10 MeV. In this case, the `REACTION` would be '(26-Fe-56(N,INL),PAR,SIG)', and the `DATA` table contains all inelastic scattering cross sections for both excitation energies and incident energies as running parameters. Another example is that many EXFOR entries have 'ELEM', 'MASS', or 'ELEM/MASS' in the product subfield (SF4) in the coded information in `REACTION` such as '(92-U-235(N,F)ELEM/MASS,IND,FY)' for neutron-induced fission of ^{235}U. This can be seen when the measurements have been done for many products such as the fission product yield measurement. In such cases, one cannot know the exact products until you read `DATA` table and cannot query based on the fission products.

The Table Creator (tablify) divides such diverse EXFOR subentries into datasets that typically align more with the requirements of nuclear physicists, such as the neutron inelastic scattering cross sections of ^{56}Fe accumulated in the first excited level or the fission product yield of ^{99}Mo produced from thermal neutron induced fission of ^{235}U, in the above cases. To determine the correspondence between the excitation energy of ^{56}Fe and its level, one needs to refer to the RIPL-3 [18] discrete-level data. Also, to make all fission products contained in the DATA table searchable, all products need to be indexed properly in the database (deresidualification). Such small separated dataset is used in a further process to generate tabulated (x, y) data. Sometimes, common information such as statistical errors in percentage form is defined in the `COMMON` data table. These errors should be redistributed (uncommonized) into the tabulated (x, y, dx, dy) table. Furthermore, to standardize the units (unitfy), one should consult the EXFOR dictionary to identify the standard unit (typically indicated as $\times$ 1.0 magnifier). We have developed similar JSON converters for parsing RIPL-3 and the EXFOR dictionary. An example of tabulated text data is shown in Fig. 4.

```
# entry-subent-pointer  : 23134-002-0
# EXFOR reaction        : ['26-FE-56', ['N,INL'], '26-FE-56,PAR,SIG']
# incident energy       : 8.4700e-01 MeV - 9.5620e+00 MeV
# target                : Fe-56
# product               : Fe-56
# level energy          : 8.4700e-01 MeV
# MF-MT number          : 3 - 51
# first author          : R.Beyer
# institute             : (2GERZFK): Helmholtz-Zentrum Dresden-Rossendorf, Dresden
# reference             : (J,NP/A,927,41,2014)
# year                  : 2014
# facility              : (LINAC): Linear accelerator
# git                   : https://github.com/IAEA-NDS/exfor_master/blob/main/exforall/231/23134.x4
#
#       E_in(MeV)       dE_in(MeV)            XS(B)           dXS(B)
      8.47000E-01      0.00000E+00      8.50000E-02      9.00000E-03
      8.89000E-01      0.00000E+00      2.51000E-01      2.30000E-02
      9.34000E-01      0.00000E+00      3.41000E-01      3.20000E-02
      9.84000E-01      0.00000E+00      5.19000E-01      4.10000E-02
      1.03800E+00      0.00000E+00      5.47000E-01      4.60000E-02
      1.09500E+00      0.00000E+00      6.07000E-01      4.50000E-02
      1.15800E+00      0.00000E+00      6.17000E-01      5.00000E-02
      1.22700E+00      0.00000E+00      5.67000E-01      4.50000E-02
      1.30100E+00      0.00000E+00      5.07000E-01      4.10000E-02
      1.38300E+00      0.00000E+00      7.10000E-01      5.30000E-02
```

Figure 4. The example of the tabulated table generated from EXFOR `ENTRY` 23134 `SUBENT` 002 for neutron inelastic cross section of 0.847 MeV level in ^{56}Fe.

3.5 REST APIs and web interfaces

Finally, an application programming interface (API) has been developed to facilitate responsive interactions between client applications of data utilization, such as data plotting tools,

web user interfaces, statistical analysis programs, and the database. For instance, the converted JSON-EXFOR entries are utilized by the REST API and are accessible through the responsive EXFOR entry preview web user interface (UI) at https://nds.iaea.org/dataexplorer/exfor/.

Another UI located at https://nds.iaea.org/dataexplorer/reactions/ provides interactive plots for scientists to visualize experimental datasets and pre-processed evaluated nuclear data libraries for comparison. Since the web interface is meant for basic data exploration, its data analysis functionality is limited. For more detailed analysis, the users can download the datasets in text format from the interface and analyze them using software packages such as Python, R, and others. For easier analysis across a larger number of datasets, the REST APIs (https://nds.iaea.org/dataexplorer/api/) data access rather than data download is recommended. These REST APIs are available not only for EXFOR datasets but also for EXFOR dictionary and RIPL-3 discrete levels, all returned in JSON format.

Given that *EXFOR master files* are irregularly added and updated, users have the flexibility to create snapshots of the datasets available at any given moment and even establish their own databases if needed, in accordance with requirements such as those of WPEC SG50. These functions will be accessible from the IAEA website by the end of 2023.

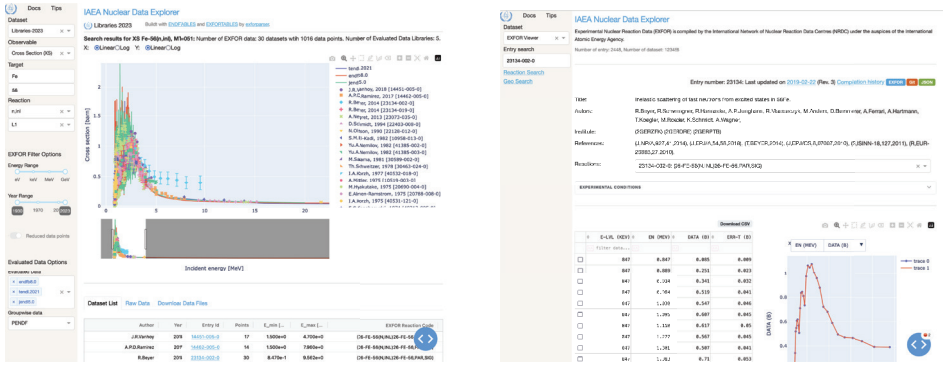

Figure 5. The screenshots of plot interface (left) and EXFOR entry viewer interface (right) in the new dataexplorer web system. An example reaction dataset, inelastic scattering cross sections for excitation energy at 0.847 MeV of ^{56}Fe, and an example EXFOR entry, `ENTRY` 23134 `SUBENT` 002, are presented.

4 Concluding remarks

The IAEA Nuclear Data Section is expected to continue providing free and unrestricted access to its basic nuclear data including nuclear reaction datasets from EXFOR as well as other data files, and related software following FAIR Principles. We have developed two open-source EXFOR parsing programs, which are based on similar conceptual considerations, to convert the EXFOR into the JSON format. We also developed REST APIs to provide easier access to the physics-oriented datasets and enable the programmatic retrieval of a large number of datasets for data mining and ML purposes. Upon these REST APIs, open web systems are designed to quickly and conveniently visualize datasets of the EXFOR entries.

References

[1] N. Otuka, E. Dupont, V. Semkova, B. Pritychenko, A. Blokhin, M. Aikawa, S. Babykina, M. Bossant, G. Chen, S. Dunaeva et al., Nuclear Data Sheets **120**, 272 (2014)

[2] A. Trkov, M. Herman, D. Brown, Tech. Rep. BNL-203218-2018-INRE, Brookhaven National Laboratory, Upton, NY 11973-5000 (2018)
[3] V. Zerkin, B. Pritychenko, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **888**, 31 (2018)
[4] D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, R. Tekade, Drug Discov Today **26**, 80 (2021)
[5] Y. Katsura, M. Kumagai, T. Kodani, M. Kaneshige, S.G. Yuki Ando, Y. Imai, H. Ouchi, K. Tobita, K. Kimura, K. Tsuda, Science and Technology of Advanced Materials **20**, 511 (2019)
[6] S. Gazula, J. Clark, H. Bohr, Nuclear Physics A **540**, 1 (1992)
[7] N.J. Costiris, E. Mavrommatis, K.A. Gernoth, J.W. Clark, Phys. Rev. C **80**, 044332 (2009)
[8] R. Utama, J. Piekarewicz, H.B. Prosper, Phys. Rev. C **93**, 014311 (2016)
[9] Z.A. Wang, J. Pei, Y. Liu, Y. Qiang, Phys. Rev. Lett. **123**, 122501 (2019)
[10] Z.M. Niu, H.Z. Liang, B.H. Sun, W.H. Long, Y.F. Niu, Phys. Rev. C **99**, 064307 (2019)
[11] D. Neudecker, M. Grosskopf, M. Herman, W. Haeck, P. Grechanuk, S. Vander Wiel, M. Rising, A. Kahler, N. Sly, P. Talou, Nuclear Data Sheets **167**, 36 (2020)
[12] H. Iwamoto, O. Iwamoto, S. Kunieda, Journal of Nuclear Science and Technology **59**, 334 (2022)
[13] A. Boehnlein, M. Diefenthaler, N. Sato, M. Schram, V. Ziegler, C. Fanelli, M. Hjorth-Jensen, T. Horn, M.P. Kuchera, D. Lee et al., Rev. Mod. Phys. **94**, 031003 (2022)
[14] A. Lewis, D. Neudecker, A. Koning, D. Barry, J. Brown, G. Schnabel, EPJ Web of Conf. **284**, 18003 (2023)
[15] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne et al., Scientific Data **3**, 160018 (2016)
[16] D. Brown, J. Brown, B. Beck, J.L. Conlin, M. Fleming, G. Gert, W. Haeck, A. Holcomb, C. Matoon, M. White et al., Tech. Rep. NEA No. 7647, OECD-NEA, Paris, France (2023)
[17] A. Trkov, M. Herman, D.A. Brown, Tech. Rep. CSEWG Document ENDF-102, BNL-203218-2018-INRE, Brookhaven National Laboratory (2018)
[18] R. Capote, M. Herman, P. Obložinský, P. Young, S. Goriely, T. Belgya, A. Ignatyuk, A. Koning, S. Hilaire, V. Plujko et al., Nuclear Data Sheets **110**, 3107 (2009)