EPJ Web of Conferences **294**, 06002 (2024) https://doi.org/10.1051/epjconf/202429406002
WONDER-2023

Integrated, Automated, and Reproducible Nuclear Data Processing at the NEA

Daniela Foligno[1,*], *Andrew* Holcomb[1], and *Michael* Fleming[1]

[1] *OECD Nuclear Energy Agency, Data Bank, 46 Quai Alphonse le Gallo, 92100 Boulogne-Billancourt, France*

Abstract. Ensuring data preservation is a top priority for the Organization for Economic Co-operation and Development (OECD), Nuclear Energy Agency (NEA) Data Bank. Within this context, "preserving data" encompasses activities such as verifying, processing, sharing, improving, and storing the data. The NEA aims to automate these processes to the greatest extent possible, with the goal of providing the JEFF community with reproducible, high-quality data. To achieve this, the Data Bank has chosen to utilize GitLab, a web-based distributed Version Control System that facilitates the collaboration of different users. The author will present the current progress of the NEA pipeline, an ongoing collaborative initiative aimed at standardizing the processing, verification, and validation of nuclear data.

1 Introduction

One of the top priorities for the NEA Data Bank is the long-term storage and maintenance of data in a way that ensures its accessibility, usability, and integrity over time. To achieve this, the NEA has dedicated substantial efforts to implementing strategies and practices that align with the FAIR principles [1]. These principles serve as guidelines to ensure data is *Findable*, *Accessible*, *Interoperable*, and *Reusable*. In order to accomplish this, the IT team at NEA installed a locally hosted GitLab instance. GitLab's comprehensive platform, encompassing software development, version control, collaboration, and CI/CD capabilities, was chosen as a means to facilitate and enhance cooperation among the participants of the Joint Evaluated Fission and Fusion (JEFF) nuclear data project [2].

2 Modernization efforts at the NEA

Prior to these modernization efforts, the exchange of information relied solely on email. Evaluators would submit new files to the JEFF Secretariat, which would then gather and disseminate them via the NEA portal. Users would download these files and independently initiate the processing, verification, and validation steps. Feedback on the files would be provided during the JEFF Nuclear Data Week, occurring every six months. Evaluators would incorporate this feedback to make modifications and enhancements to their evaluations. Unfortunately, this process was characterized by slowness and inefficiency. Moreover, the individualized processing could introduce a potential bias susceptible of compromising the accurate comparison of data quality.

* e-mail: daniela.foligno@oecd-nea.org

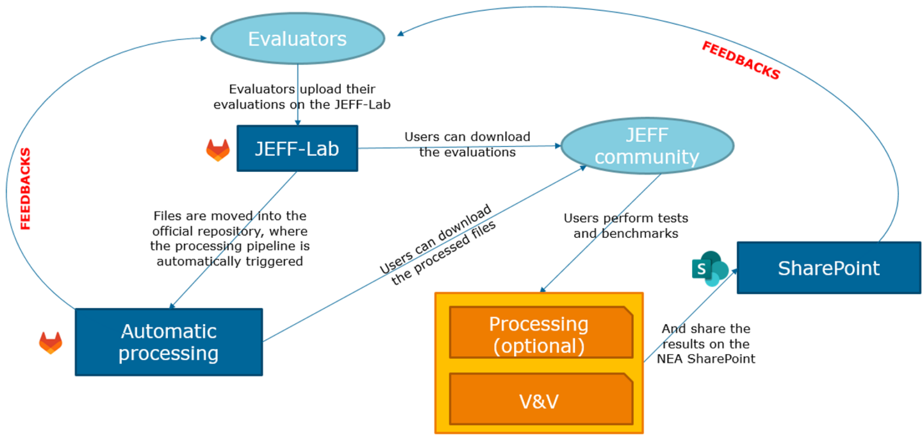


Figure 1: Current JEFF Verification and Validation scheme

To address these issues in the process of producing a new JEFF release, the NEA worked on new modern tools. Figure 1 illustrates a scheme of the current processing, verification, and validation process. The new NEA tools are represented in the blue boxes. The GitLab repository named JEFF-Lab serves as a platform for evaluators to upload and maintain their files. JEFF members gain immediate access to download and test these files as soon as they become available. GitLab enables collaborative work on the files while maintaining a comprehensive record of changes, ensuring complete transparency and traceability. Another GitLab repository defines the *processing pipeline*, which is automatically triggered whenever an evaluation is uploaded or updated. Additionally, SharePoint serves as a platform for users to share their preliminary results and conclusions with evaluators and other users, facilitating archiving purposes. To summarize, the new scheme allows the whole community to have access to the data (new evaluations, working files, preliminary results) at any time, allowing for a faster, more transparent, and more efficient review process. Furthermore, the NEA *processing pipeline* makes the processed data automated, transparent, and reproducible.

3 The current NEA processing pipeline

GitLab is a Web-based collaborative platform built on top of git, a well-known version controlled system. The NEA chose GitLab for the comprehensive set of tools it provides: version control, issue tracking, code review, Continuous Integration / Continuous Deployment (CI/CD), and more. This section describes the current status of the NEA *processing pipeline* and the jobs it contains. Preliminary checks, including CHECKR, FIZCON, and PSYCHE, are succeeded by the essential stages (NJOY basics and PREPRO for FISPACT-II) that involve the reconstruction and Doppler broadening of cross sections. Subsequent steps, such as NJOY ace, OpenMC, and AMPX, are carried out to prepare the files for use in transport codes.

3.1 CHECKR

CHECKR [3] is an ENDF utility code that verifies the conformity of an evaluated data file to the ENDF format. The primary purpose of the CHECKR code is to verify the correctness and internal consistency of the evaluated nuclear data contained in an ENDF file. It checks various aspects of the data, including: format and structure, cross-section consistency, reaction thresholds, and data integrity. By performing these checks, the CHECKR code helps identify potential errors, inconsistencies, or missing information in the ENDF file.

3.2 FIZCON

FIZCON [3] is an ENDF utility code used to check the internal consistency of evaluated nuclear data files. These checks include verifying that data arrays are arranged in ascending energy order, confirming that the widths of resonance parameters add up to the total value, ensuring that Q-values are both reasonable and consistent, verifying the presence of all required sections covering the appropriate energy range, normalizing secondary distributions to 1.0, and confirming energy conservation in decay spectra.

3.3 PSYCHE

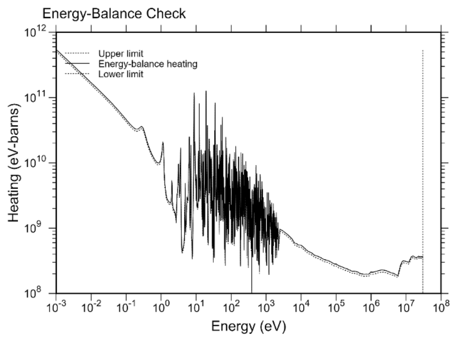
PSYCHE [3] is an ENDF utility code used to verify the physics content of evaluated nuclear data files. It performs various tasks such as checking energy conservation, calculating thermal cross sections and resonance integrals, assessing continuity across resonance region boundaries, and verifying Q-values against mass tables.

3.4 NJOY basics

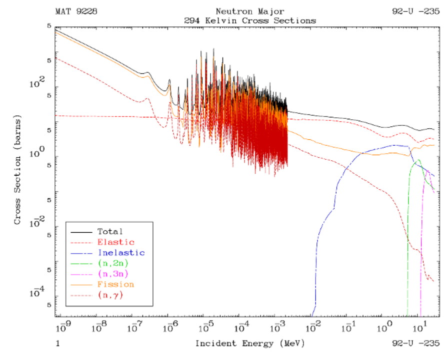
NJOY [4] is the nuclear data processing code developed at Los Alamos National Laboratory. *NJOY basics* provides a thorough procedure for converting, reconstructing, and Doppler broadening cross sections. It also accounts for self-shielding in the unresolved resonance region and incorporates additional reaction data into nuclear data files. This tool is crucial for creating precise and comprehensive nuclear data sets, which are vital for various nuclear physics and engineering applications. In particular, the following NJOY modules are used:

- MODER: Converts files between different formats (ENDF, PENDF, GENDF)
- RECONR: Reconstructs resonance cross sections from resonance parameters. The output is a pointwise-ENDF (PENDF) file containing all the cross sections on a unionized energy grid
- BROADR: Generates Doppler-broadened cross sections by considering the temperature dependence of the nuclear data. The output is in the PENDF format
- UNRESR: Used to generate effective self-shielded cross sections for resonance reactions in the unresolved energy range
- HEATR: Computes point-wise heat production cross sections and radiation damage energy production for specified reactions. It adds this information to an existing PENDF file. These results are essential for assessing the heat and radiation effects in nuclear systems
- GASPR: Adds gas production reactions (MT=203-207) to the PENDF file, relevant for applications involving gas-filled environments or gas-based systems
- VIEWR: Provides plotting capability for NJOY. An example of plot automatically generated by the NEA pipeline is shown in Fig. 2

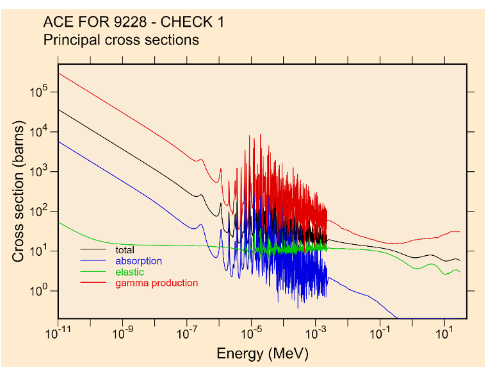
The binary files produced by this job can be used by the Monte Carlo code TRIPOLI-4 [5]. In addition to that, in this step, a python script extracts the thermal neutron constants computed by NJOY and, whenever available, it compares them with the Standards from 2017 [6].



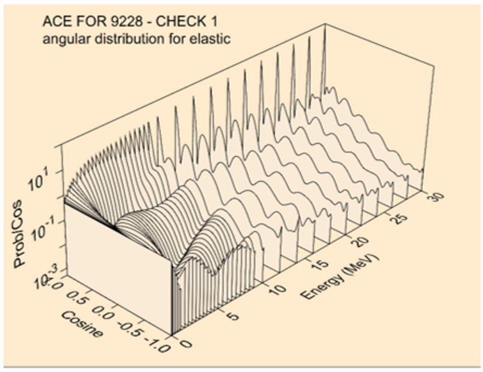
(a) Energy balance from *NJOY basics*



(b) Main cross sections from *PREPRO for FISPACT-II*



(c) Main cross sections from *NJOY ACE*



(d) Angular distribution for elastic from *NJOY ACE*

Figure 2: Examples of plots generated by the NEA pipeline while processing U-235

3.5 NJOY ACE

In the *NJOY ACE* job, two additional NJOY modules are used:

- PURR: It generates probability tables, which can be used in Monte Carlo codes to address unresolved-resonance self-shielding
- ACER: It converts libraries in the ACE format (A Compact ENDF). These ACE libraries are specifically designed for MCNP® [7], but can be used by other continuous-energy Monte Carlo codes, like Serpent [8].

3.6 PREPRO for FISPACT-II

PREPRO [9] consists of 18 pre-processing computer codes created to convert ENDF-formatted neutron and/or photon data into a usable format for various applications. In the NEA *processing pipeline*, the following PREPRO modules are used to generate grouped-wise cross sections for FISPACT-II [10]:

- LINEAR: Converts the tabulated cross sections to linear-linear interpolable form
- RECENT: Reconstructs resonance contributions and background cross sections in linearly interpolable form in the ENDF format
- SIGMA: Doppler broadens neutron-induced cross sections from and to ENDF format
- SIXPACK: Checks all double-differential data (MF=6)
- MERGER: Selectively retrieves data from multiple ENDF files and merges them into a single output
- ACTIVATE: Combines file 3 (cross sections) and file 9 (multipliers) to create file 10 (activation cross sections)
- FIXUP: Performs automatic or optional corrections on evaluated data in ENDF format.
- DICTIN: Creates a reaction index for each material and inserts it in file 1, section 451
- GROUPIE: Computes self-shielded multigroup cross sections and multi-band parameters
- EVALPLOT: Reads evaluated data and plots selected cross sections, parameters, angular distributions, and energy distributions.

3.7 OpenMC

OpenMC [11] is an open-source Monte Carlo particle transport code specifically developed for simulating the transport of neutrons, photons, and electrons in intricate systems. As a part of its functionality, OpenMC includes a function that allows for the conversion of processed files from the ACE format to the HDF5 (Hierarchical Data Format) format that is used by OpenMC. This conversion is done in the *OpenMC* job.

3.8 AMPX

AMPX [12] is a collection of computer codes developed at Oak Ridge National Laboratory used to process ENDF-formatted nuclear data files into continuous energy, multi-group, and covariance data libraries for SCALE [13]. AMPX is not included in the NEA *processing pipeline* scheme illustrated in Fig 1 because the version we employ for processing originates from SCALE-6.2, subject to export control. This restriction limits public access to the job. To address this, the NEA has adopted a strategy of processing the nuclear data library using AMPX within a private repository. Subsequently, the resulting binaries intended for use in SCALE are made accessible to the public through the NEA website.

3.9 Collect artifacts

The final job gathers and organizes all the artifacts generated from the previous jobs. The artifacts are sorted and arranged in a folder, which users can then download in a compressed format for convenience. In conclusion, the current NEA *processing pipeline* can produce neutron cross section libraries for the following codes: MCNP®, TRIPOLI-4, Serpent, FISPACT-II, OpenMC, and SCALE.

4 Verification and Validation pipelines

After completing the file processing, the NEA took the initiative of developing new pipelines aimed at automating the verification and validation of the files.

4.1 Verification

The purpose of the *verification pipeline* is to ensure that the processed data can be loaded and utilized by users' codes. Presently, the NEA is able to assess the data's usability in Serpent and SCALE. It is important to emphasize that the verification steps do not provide any assessment of the data's quality, but solely evaluate its format and usability. To confirm the proper processing and usability of ACE files in Serpent, the NEA has developed a criticality calculation input file. A simple configuration is employed, featuring a cylinder composed of 95 at% U-235 and 5 at% of the isotope being tested. Similarly, for the verification of AMPX-generated binaries in SCALE, a comparable approach is adopted. In the case of thermal spectra, the NEA has selected a k_{eff} calculation for an infinite homogeneous box containing U-235, H in H2O, and the isotope being tested. In the case of fast spectra, the calculation involves evaluating the leakage from a sphere of the material being tested, with a Cf-252 source positioned at the center. For further details regarding the NEA Verification pipeline, please refer to reference [14].

4.2 Validation

The objective of the *validation pipeline* is to evaluate the quality of nuclear data by comparing it with experimentally measured quantities of interest. The NEA provides various benchmarking databases such as ICSBEP (International Criticality Safety Benchmark Evaluation Project), SINBAD (Shielding Integral Benchmark Archive and Database), and SFCOMPO (Spent Fuel Isotopic Composition database). Within the WPEC-SG45 framework, over 5000 ICSBEP MCNP® inputs were gathered. Approximately 600 of these inputs were automatically converted to Serpent input using the csg2csg [15] tool and are fully functional. While many others have been partially converted, they still require additional work before they can be used. All the converted inputs, regardless of their functionality, are currently accessible in the NEA GitLab to users who possess a valid Serpent license. For more detailed information regarding the NEA Validation pipeline, please consult reference [14].

5 Conclusions and perspectives

The NEA has made significant efforts to modernize the evaluation, processing, verification, validation, and distribution of nuclear data. The NEA utilizes community-approved methods to process the files, incorporating the latest knowledge available. As a result, the processed files provided by the NEA are of the highest quality. Currently, the NEA system is capable of processing, testing, and deploying neutron cross-section libraries for MCNP®, Serpent, TRIPOLI-4, FISPACT-II, OpenMC, and SCALE. The implementation of the NEA pipeline for processing the evaluations eliminates potential bias and ensures that comparisons solely concentrate on the actual data. The NEA has prioritized transparency and reproducibility at every stage of the process, leading to more reliable and easily accessible data. Thanks to GitLab and SharePoint, the JEFF community can immediately access preliminary data and results, facilitating fast and transparent sharing of information. The objective of the NEA Databank is to expand its system to include a wider range of:

- Nuclear data processing codes
- Neutron transport codes
- Validation suites
- Nuclear data libraries such as ENDF [16], JENDL [17], and more.

This expansion is aimed at providing the international nuclear engineering community with the most comprehensive and inclusive publicly accessible data available.

References

[1] F. de Lillo, F. Cecconi, G. Lacorata, A. Vulpiani, Europhysics Letters **84**, 6 (2008)
[2] A. J. M. Plompen, O. Cabellos, C. de Saint-Jean, M. Fleming, A. Algora, M. Angelone, P. Archier, E. Bauge, O. Bersillon, A. Blokhin et al., The European Physical Journal A **56** (2020)
[3] C. Dunfird, ENDF-6 Formats Manual (2018)
[4] R. Macfarlane, D.W. Muir, R.M. Boicourt, A.C. Kahler, J.L. Conlin, https://www.osti.gov/biblio/1338791 (2017)
[5] E. Brun, F. Damian, C. Diop, E. Dumonteil, F. Hugot, C. Jouanne, Y. Lee, F. Malvagi, A. Mazzolo, O. Petit et al., Annals of Nuclear Energy **82**, 151 (2015), joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2013, SNA + MC 2013. Pluri- and Trans-disciplinarity, Towards New Modeling and Numerical Simulation Paradigms
[6] A. Carlson, V. Pronyaev, R. Capote, G. Hale, Z.P. Chen, I. Duran, F.J. Hambsch, S. Kunieda, W. Mannhart, B. Marcinkevicius et al., Nuclear Data Sheets **148**, 143 (2018), special Issue on Nuclear Reaction Data
[7] C.J. Werner, J.S. Bull, C.J. Solomon, F.B. Brown, G.W. McKinney, M.E. Rising, D.A. Dixon, R.L. Martz, H.G. Hughes, L.J. Cox et al., https://www.osti.gov/biblio/1419730 (2018)
[8] J. Leppänen, M. Pusa, T. Viitanen, V. Valtavirta, T. Kaltiaisenaho, Annals of Nuclear Energy **82**, 142 (2015), joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2013, SNA + MC 2013. Pluri- and Trans-disciplinarity, Towards New Modeling and Numerical Simulation Paradigms
[9] D. Cullen, Technical report IAEA-NDS-0238 (2021)
[10] J.C. Sublet, J. Eastwood, J. Morgan, M. Gilbert, M. Fleming, W. Arter, Nuclear Data Sheets **139**, 77 (2017), special Issue on Nuclear Reaction Data
[11] P.K. Romano, N.E. Horelik, B.R. Herman, A.G. Nelson, B. Forget, K. Smith, Annals of Nuclear Energy **82**, 90 (2015), joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2013, SNA + MC 2013. Pluri- and Trans-disciplinarity, Towards New Modeling and Numerical Simulation Paradigms
[12] D. Wiarda, M. Williams, C. Celik, M. Dunn, Technical Report ORNL/TM2016/43 (2016)
[13] W. Wieselquist, R.A. Lefebvre, https://www.osti.gov/biblio/1959594 (2023)
[14] A.H. et al., *Automated, reproducible data processing, verification, and validation at the NEA* (2023), to be published
[15] Davis, Andrew, Lilley, Steven, Shimwell, Jonthan, EPJ Web Conf. **247**, 04012 (2021)
[16] D. Brown, M. Chadwick, R. Capote, A. Kahler, A. Trkov, M. Herman, A. Sonzogni, Y. Danon, A. Carlson, M. Dunn et al., Nuclear Data Sheets **148**, 1 (2018), special Issue on Nuclear Reaction Data
[17] O. Iwamoto, N. Iwamoto, S. Kunieda, F. Minato, S. Nakayama, Y. Abe, K. Tsubakihara, S. Okumura, C. Ishizuka, T. Yoshida et al., Journal of Nuclear Science and Technology **60**, 1 (2023)