OECD-NEA Data Bank
46, Quai Alphonse Le Gallo
92100 Boulogne-Billancourt
France

Memo CP-N/156

Date:	June 23, 2020
To:	Distribution
From:	M. Fleming
Email:	wpec@oecd-nea.org, michael.fleming@oecd-nea.org
Telephone:	+33 (0) 1 73 21 28 22

Subject: New WPEC Subgroup 50

At the 32nd meeting of the NEA Nuclear Science Committee's Working Party on International Nuclear Data Evaluation Co-operation (WPEC) on 14-15 May 2020, a new subgroup approved by the members of the working party. The new subgroup 50 on *Developing an Automatically Readable, Comprehensive and Curated Experimental Reaction Database* will work on the design and implementation of a three-layered database that draws from the EXFOR database, creating a new computational format with an expanded schema and allowing for the addition of 'objective' corrections and non-unique 'subjective' judgements by nuclear data evaluators.

This subgroup will be co-ordinated by:

Amanda LEWIS (US) and
Denise NEUDECKER (US).

The subgroup monitor is:

Arjan KONING (IAEA).

The first meeting has been scheduled for 14-15 September 2020 and will be held remotely via WebEx. The subgroup proposal (LA-UR-20-22620) is provided at the end of this memo. More information on the activities of the subgroup can be found on the WPEC website:

`https://oecd-nea.org/science/wpec/sg50`

Those interested in participating in the activities of this subgroup, including members of the International Network of Nuclear Reaction Data Centres (NRDC), should contact their country's delegate(s) to the WPEC, the parent Nuclear Science Committee or their permanent delegation to the OECD. Please note that membership is limited to NEA member countries, although in certain circumstances, and in compliance with the NEA Participation Plan, experts from non-NEA member countries may be able to participate as invitees. Please contact the NEA Secretariat for more information.

Distribution:

a.koning@iaea.org
abhihere@gmail.com
aloks279@gmail.com
amanda_lewis@berkeley.edu
bknayak@barc.gov.in
daniela.foligno@oecd-nea.org
dbrown@bnl.gov
dneudecker@lanl.gov
draj@barc.gov.in
fukahori.tokio@jaea.go.jp
ganesan555@gmail.com
gezg@ciae.ac.cn
iwamoto.osamu@jaea.go.jp
j.c.sublet@iaea.org

jmwang@ciae.ac.cn
kaltchenko@kinr.kiev.ua
kenya.suyama@oecd-nea.org
kimura.atsushi04@jaea.go.jp
l.vrapcenjak@iaea.org
manuel.bossant@oecd-nea.org
masaaki@nucl.sci.hokudai.ac.jp
michael.fleming@oecd-nea.org
marina-03-08@yandex.ru
mmarina@ippe.ru
nicolas.soppera@oecd-nea.org
n.otsuka@iaea.org
nrdc@jcprg.org
odsurenn@gmail.com
ogritzay@ukr.net
ogrudzevich@ippe.ru
otto.schwerer@aon.at
pikulina@expd.vniief.ru
pritychenko@bnl.gov
s.okumura@iaea.org
samaev@obninsk.ru
sbabykina@yandex.ru
scyang@kaeri.re.kr
selyankina@expd.vniief.ru
sonzogni@bnl.gov
stakacs@atomki.mta.hu
stanislav.hlavac@savba.sk
sv.dunaeva@gmail.com
tada@nucl.sci.hokudai.ac.jp
taova@expd.vniief.ru
tarkanyi@atomki.hu
tatiana.ivanova@oecd-nea.org
vvvarlamov@gmail.com
v.zerkin@iaea.org
vidyathakur@yahoo.co.in
vsemkova@inrne.bas.bg
wpec@oecd-nea.org
yolee@kaeri.re.kr
zholdybayev@inp.kz

Date: 14-April-2020

WPEC sub-group proposal

Amanda M. Lewis (UC Berkeley, USA), Denise Neudecker (LANL, USA)
Arjan Koning (Monitor, IAEA-NDS, Austria)

Title

Subgroup on developing an automatically readable, comprehensive and curated experimental reaction database

Justification for a Subgroup

The importance of the EXFOR database for evaluations of nuclear reactions cannot be overstated. The EXFOR database is the starting point of every evaluation; it either supplies experimental data and uncertainties directly used for the evaluation or it informs nuclear theory parameters for models underlying evaluations. The better the information in EXFOR is–including but not limited to mean values, uncertainties and measurement information–the better our evaluated nuclear data and covariances can become in all international nuclear data libraries including BROND, CENDL, ENDF/B, JEFF, JENDL, and TENDL.

This subgroup aims to develop an automatically and unambiguously readable differential experimental data format and generate from accordingly updated EXFOR files a more comprehensive and curated experimental reaction database. This project will specifically tackle the following key issues in the EXFOR database and format that significantly hamper nuclear data evaluations:

- The EXFOR format allows for ambiguities in several instances, incapacitating meaningful automated reading of EXFOR files for larger-scale nuclear-data evaluations or analysing large amounts of data with modern algorithms (e.g., machine learning ones). For instance, the uncertainties are described with uninformative labels such as “ERR-x” with free-text descriptions based on the original work of the author.
- The information in the EXFOR database is either taken directly from the original author’s work or authorized by the authors. When evaluators undertake their uncertainty quantification and analysis of experimental data, they usually pass judgment on these data based on subjective criteria or even time-consuming in-depth investigations. Unfortunately, this information is usually lost. This leads to the well-known fact that nearly every time, when an evaluation is re-done by another evaluator, this person has to “re-invent the wheel” when it comes to the analysis of experimental data. The current EXFOR “correction system” allows for data sets to be renormalized or corrected for updated standards, monitors or

structure values. The system also allows for uncertainty values to be added or changed. This system is not widely known about or used at the present.

If, however, all EXFOR information (including uncertainties and metadata) are stored in a standardized format and language, this information can be more automatically processed and analysed. Also, if a widely-used central repository is available, capturing expert knowledge and in-depth studies teasing out the quality of a particular data set, this information can be retained for future generations. Both efforts combined would save many hours of evaluators performing repetitive analysis work and account for issues in data that might have been otherwise overlooked in a new analysis. This significant speed-up of nuclear data evaluations becomes even more crucial at this time as the number of evaluators is dwindling world-wide. This newly formatted and comprehensive experimental reaction database will also enable more automated testing of nuclear data mean values and uncertainties given available information in EXFOR and allow for the discovery of trends in experimental data with machine learning methods, both distinctly enhancing our ability to judge and improve the quality of nuclear data.

This proposed sub-group ties into SG-49 as it makes past analyses of experimental data for evaluation purposes reproducible. It is also a follow-up to the soon-ending SG-44 that had as one of its aims to assess the quality of nuclear data covariances which becomes a more reachable goal with the outcome of this SG. Lessons learned on developing a code and database infrastructure from SG-43 can be also leveraged.

Subgroup Monitor

Arjan Koning, IAEA-NDS, Austria, (IAEA-NDS)

Subgroup Coordinators

Amanda M. Lewis, UC Berkeley, USA, (ENDF)
Denise Neudecker, Los Alamos National Laboratory, USA, (ENDF)

Suggested Subgroup Participants[1]

CENDL	Z. Ge (CNDC), X. Ruirui (CNDC), J. Wang (CNDC)
ENDF	D. Brown (BNL), Y. Danon (RPI), M. Grosskopf (LANL), M. Herman (LANL), C. Mattoon (LLNL), B. Pritychenko (BNL), V. Sobes (UTK), A. Sonzogni (BNL), P. Talou (LANL), K. Wendt (LLNL), M. White (LANL)
IAEA-NDS	R. Capote, S. Okumura, N. Otsuka, G. Schnabel, J.-C. Sublet, V. Zerkin
JEFF	O. Cabellos (UPM), C. de Saint Jean (CEA), E. Dupont (CEA),

[1] Tentative names only. The data projects will have to identify appropriate participants, which will have to check with their own institutions the time that they can devote to this activity.

	M. Gilbert (UKAEA), L. Leal (IRSN), D. Leichtle (KIT), G. Noguère (CEA), A. Plompen (JRC-Geel), D. Rochman (PSI), P. Schillebeeckx (JRC-Geel), H. Sjöstrand (UU), O. Vilkhivskaya (UKAEA)
JENDL	S. Chiba (Tokyo Tech), O. Iwamoto (JAEA), A. Kimura (JAEA), S. Nakamura (JAEA), T. Sanami (KEK), N. Shigyo (Kyushu Univ.)
NEA	M. Fleming, N. Soppera

Definition of the project and proposed activities

Definition: The goal of this sub-group is to provide a prototype of three layers of an experimental reaction database with the specific aim to (a) make information in EXFOR more easily readable with automated methods and (b) store comprehensive and curated information on specific data sets in order to speed up the experimental analysis process of evaluations by years. This database will build on "Layer 0", the EXFOR database in its current form, which is maintained by the International Network of Nuclear Reaction Data Centres (NRDC).

Proposed activity: For the first layer of this experimental reaction database, a standardized and unambiguous format for all meta-data, data and uncertainties in EXFOR and even beyond those already in there will be developed along with codes that convert existing EXFOR files (layer 0) into the new format. These new files will be tested with automated reading and machine learning algorithms. The second layer of this database expands its scope compared to the EXFOR philosophy as information in the data files will be corrected according to objective criteria (for instance, use of new standards or decay data, flagging missing uncertainty sources using the templates of expected uncertainties developed by CSEWG, outlier estimation using statistical methods). The automatic monitor and standards updating in the EXFOR correction system will be used as a starting point for this layer. The third layer will be a curated experimental database using the 2nd layer database as basis. In this database, we will store expert judgment on experimental data and add missing uncertainties based on either expert knowledge or the aforementioned templates. The expert corrections in the EXFOR correction system can serve as a starting point for this layer. For all these different databases, we will develop codes to produce the new files from current EXFOR files and create formats to store the information.

Relevance to Evaluated Data Files

If this project is executed, the resulting experimental reaction database files will enable a significant speed-up of future evaluations because the files can be automatically read and interpreted. Also, these files will contain expert knowledge on past experimental data sets obtained throughout decades of research. Hence, the evaluator can build up on this information rather than having to re-do a time-intensive analysis. In addition to that, we strive to provide comprehensive data files with regards to uncertainties. Having close to complete uncertainty information available will result in more realistic evaluated data and uncertainties.

Last but not least, having these databases available enables building the capabilities of the future. Currently, the data in EXFOR cannot be used on a large scale for testing whether mean values or evaluated uncertainties are realistic because mean values are not corrected for known effects (e.g., new standards) and uncertainties are not comprehensive. Having these databases available, we could perform this testing. Having these files at our disposal also eliminates the barriers to employing large-scale machine learning algorithms to tease out trends in data that currently elude evaluators due to their sheer number.

Time-Schedule and Deliverables:

- *2021–2022:* As the first step, a decision must be reached regarding which part of the EXFOR database will be used as test ground for the new formats and layers of the experimental reaction database. In the same year, we will deliver a document on the agreed-upon naming convention for attributes and changes to the EXFOR format that allow an unambiguous automated reading of EXFOR information, and the type of database to use. We will start on a first prototype of the code that converts existing EXFOR entries (layer 0) into the new format (layer 1).
- *2022–2023:* The prototype code will be applied to convert a limited number of EXFOR entries into the new format. We will test in the same year if ML and other automated algorithms can work with this new format and refine codes and formats as needed. The finalized conversion code is the 2^{nd}-year deliverable. In the same year, we will start the discussion on what the 2^{nd} (corrected) layer of the experimental reaction database should be corrected for and will start developing codes to generate this 2^{nd} layer using templates of expected measurement uncertainties developed by CSEWG and the formats developed by this sub-group.
- *2023–2024:* In the third year we will deliver a limited number of 2^{nd}-layer experimental reaction database entries for testing and review. This information aids in finalizing the format for the 2^{nd} layer and finalizing the conversion codes. The rest of the last year will be spent on extending formats and codes to encompass the development of the 3^{rd} (curated) layer of the experimental reaction database and giving a few example entries.

LA-UR-20-22620