WPEC Sub-Group Proposal

Title: Validation of Nuclear Data Libraries (VaNDaL[1])

Justification:
The nuclear data community continues the struggle to test our data libraries in a transparent, reproducible manner. Individuals and institutes spend considerable time building validation suites that often have errors that others have already found so that we do not gain the full benefit from our parallel efforts. Standardized quality assurance (QA) processes can provide us benchmark suites for the validation of nuclear data libraries that can meet a basic tenet of science: our colleagues can reproduce them. We can then go further and build upon our shared work enabling even greater rigor and thought to be applied to this most important activity.

Subgroup Monitor
Andrej Trkov (NDS/IAEA)

Subgroup Coordinators
Morgan White (LANL, ENDF/B) and David Bernard (CEA, JEFF)

Subgroup Participants
TBD.

Context

The challenge for any nuclear data evaluation project is to periodically release a revised, fully consistent and complete library – with all the necessary data and covariances – and to ensure that it is robust and reliable for a variety of applications. Within such evaluation efforts, benchmarking activities are the final crucial step in validating the proposed library. The major data evaluations – JEFF, ENDF/B, JENDL, CENDL, ROSFOND/BROND and TENDL – all aim to provide such a library. Thus, they each require a coherent and efficient benchmarking process. In the past, this has been achieved through ad-hoc efforts by many participants typically using many different benchmarking suites. This process is prone to error and misunderstanding and considerable time can be wasted tracking down, for example, typographical or modeling errors that can be avoided by a more systematic approach.

The last two decades have seen the rise of tremendous new resources to address this issue: the international handbooks for criticality safety, reactor physics, fuel burnup, and radiation shielding. These peer reviewed references comprise our best understanding of the benchmark experiments by which we validate the nuclear data that we use for predictions of neutron

[1] ***Vandal*** *– In this case, one who seeks to determine the weaknesses of, that is 'break', our data libraries.*

reactivity, criticality safety, radiation shielding and other aspects of particle transport; and, more recently, for predictions of isotopic transmutation for waste disposal and dosimetry. *But we have failed to take the necessary steps to create, maintain and enhance tools to facilitate widespread use of these benchmarks*. This proposal seeks to (1) provide a methodology to assemble Quality Assured (QA) versions of these inputs for the MCNP and other transport codes; (2) to provide an initial repository of the major collections of such inputs and begin the QA process for them; and, (3) to provide the tools necessary to extract standardized information from such validation tests presented in a harmonized way.

The initial focus of this sub-group will be limited to neutron transport – i.e. validation of the Boltzmann equation – and particularly focused on neutron reactivity, also commonly referred to as k-effective or neutron multiplicity. There are many other aspects of benchmark testing that should be included and these will be discussed more below. Two examples of the advanced test suites used to validate our nuclear data libraries are the criticality test suite of more than 1000 MCNP inputs (developed by Skip Kahler) used heavily by the ENDF/B community and the 2000 tests (developed by Steven van der Marck) used by the JEFF community. Another example of such a test suite is more than 400 tests in the Oak Ridge National Laboratory SCALE code VALID test suite. In an inspiring gesture of good will, these individuals and institutes have agreed to make these inputs available enabling a remarkable starting point for this project. We hope other individuals and institutes will follow their example and contribute their existing inputs to this worthy cause. ORNL has also offered the procedures they have developed to verify the VALID test suite inputs as a starting point for developing the quality assurance process. The rigorous QA of a complete suite of international benchmarks is a task no one person or institute has the resources to tackle. We are stronger if we work together.

The first task is to decide how to lay out a repository of inputs and by what manner to review them for quality assurance. The question of repository layout is not as simple as it seems. There are too many inputs to be reasonably run by hand. Efficient automation is required with standardized scripts and file naming and storing conventions. We must consider that for many benchmarks, there are multiple configurations and, sometimes, more than one description of a configuration – e.g. a simple and detailed model. There are also auxiliary values associated with each configuration that must be stored, e.g. the benchmark k-effective value and, potentially, supplemental information like experimentally measured reaction rate ratios and quantities commonly used for trend analyses, or sensitivities profiles needed for analysis.

Having decided what to store where, we must next ensure that the correct values are present. There is no one who works on these simulations that has not experienced mistyping one of the hundreds and thousands of numbers that must be entered by hand. It is an effort that requires rigorous cross checks and providing a system that ensures such rigorous cross checks is essential. ***This system will represent a single point of failure and thus must be held to the highest standards.*** Defining the set of requirements for the review process is a key deliverable for the sub-group. A tiered system is envisioned to enable us to capture inputs in a range of states from initial submittal through thoroughly vetted.

A simple but underlying assumption in this process is revision control of the inputs. It is not enough to have a suite of inputs. They will change over time, so versions must be easily traceable and verifiable, and the reasons for the changes must be clearly documented. The software development process faces this same challenge and many tools exist to facilitate this task. A decision must be made on a standard toolset to use. This is also the appropriate time to discuss and implement access control. It is remarkable the consensus that has emerged that these inputs and tools should be publicly available. However, we must have a cadre of gatekeepers who can effectively enforce the procedures by which we modify them.

Simply having the inputs is only the first step. To enable robust, reliable use of these tests we must automate how they are run and standardize the outputs such that they may be easily parsed, stored, reported and compared. It is not always obvious to users of a code the optimal way to run these problems to ensure reliable results; capturing the best practices of how to run these problems and providing tools that automate these practices is essential. We will also need to determine the set of leakage, reaction rate, sensitivity and other tallies that are needed in order to robustly mine the outputs. This standard suite will be driven by the analyses that are proposed. For example, if plots versus above thermal leakage fraction or average lethargy causing fission are desired, these quantities must be tallied. Scripts will likely be needed that modify the inputs in standardized ways to provide the appropriate quantities in outputs.

Knowing what comparisons to make, and how to show the results, is the final crucial step in validating a nuclear data library for use in some application. Capturing and automating the appropriate comparisons is the final step to enhancing the value of this infrastructure. Similar to providing tools to appropriately run these tests, we must work with these communities to provide the tools to mine the results of these simulations and properly compare them with the benchmark values.

By automating the tasks of running the tests and mining the outputs, we reclaim that time to *enable our community to focus on the real task of interpreting their meaning*.

Making these results broadly available is highly desirable. The OECD/NEA DICE and NDaST tools and website present an obvious potential starting point. Standardized outputs that feed the DICE and NDaST databases could ensure that test results are broadly communicated and easily mined to understand differences. This will enable our users to better understand the differences of opinions that are represented between different nuclear data libraries and make choices about which data are best suited for their applications.

It should be noted that this sub-group will not perform the QA of all the inputs. It will produce a framework by which this work can be done. An initial population of benchmarks will be provided and a subset of them will be checked to ensure the QA process works as intended. But many of the inputs will still be in the initial stages of the QA process and require the broader community to help in the process. By working together, we can achieve a system that will enable everyone to perform the robust, reliable benchmarking we all need.

Time-Schedule and Deliverables:

- Year 1:
    - Collect suitable input decks from participants and other stakeholders
    - Define the layout and implement an initial repository
    - Generate a prototype QA requirements specification and tools to help implement this process

- Year 2:
    - Perform QA on a subset of inputs using prototype requirements to determine its suitability and revise as necessary
    - Generate a prototype requirements document for standard outputs
    - Develop tools to run benchmarks and parse these outputs

- Year 3:
    - Finalize QA and outputs requirements
    - Release initial benchmark suite and tools

Potential follow-on projects or stretch goals

This project will set the stage for several obvious follow-ons, for example:

- Expand to include shielding and reactor physics transport benchmarks.
- Expand to include fuel burnup transmutation benchmarks.
- Expand to standardize input decks across multiple transport codes.

As time and resources permit, these items may be considered. Certainly we would expect that follow on subgroups would be considered to expand this effort in these types of ways.