28 March 2012

WPEC sub-group proposal

(Dennis McNabb, LLNL)

Title

Beyond the ENDF format: GND, a modern nuclear database structure

Justification for a Subgroup

The ENDF format has been the workhorse of the nuclear data community for nearly 50 years. The international use of this format has allowed and promoted data exchange among major nuclear data centres and their user communities. This broad exchange has greatly enriched our field and thus a change to this underlying pillar in our foundation should be considered only after careful deliberations. However, the international community recognizes significant issues with the ENDF format.

Lawrence Livermore National Laboratory (LLNL) has developed an initial proposal for a Generalized Nuclear Data structure, GND, and the U.S. nuclear data community has been reviewing this structure as a possible successor to ENDF. Although LLNL has taken the first step towards developing and deploying a new, modern nuclear data structure to replace ENDF, it needs critical review and revision by the international nuclear data community in order to ensure that it will adequately address the needs of the different data projects. By forming a long-term subgroup, the WPEC will help ensure this review includes the major stakeholders within the international community and is conducted in a transparent and inclusive manner. Longer term, international support fostered by WPEC and other sponsors will be vital to help encourage the international adoption of a new, modern nuclear data structure and the development or updating of the necessary associated infrastructure.

The current ENDF format dates back to 1964 and the basic data storage records have remained unchanged throughout this time. These records maintain their roots in an archaic punch-card form with its rigid requirements on data position. This leads to files that are cryptic in nature, requiring an extensive manual to decipher. Practitioners spend years learning this format and even experts often make mistakes in its interpretation and use. Further, the rigid nature of these simple records often makes it difficult to extend the format. A versatile and extensible Generalized Nuclear Data (GND) structure is needed that can handle the types of data defined in ENDF but also take advantage of the computing advances made in storage and handling of abstract data types.

The layout of ENDF was designed for serial processing using the FORTRAN-IV language. While this was the standard at its inception, there is a strong desire to access and use these data with a wide range of modern computing languages and paradigms. Newcomers, who have never encountered legacy Fortran code, are typically aghast to

learn that the format uses Fortran specific scientific notation, e.g.“ 2.727472+6” instead of “ 2.72747e+6”. Fortran is the only computing language that handles this gracefully and writing an interpreter for other languages to handle this is difficult and error prone.[1] This onerous task also adds unnecessary time to file I/O. While this convention does allow for one additional significant digit, the ENDF format still does not provide sufficient accuracy for all cases, leading, for example, to problems with accuracy in the neutron resonance region, and to problems with non-positive definite covariance matrices.

More troubling than the issue of numerical precision are several fundamental flaws in the ENDF format design. Among these is the use of 3 digits for the reaction identifier, known as the MT number. Even with a recent update to add missing reactions to previously unused MT values, there are reactions that evaluators – particularly the medium energy community – would like to describe that must currently be handled ad hoc, e.g. by misusing an existing value, or by lumping multiple reactions together until the fundamental data are obscured. A more universal and understandable scheme to identify reactions should be one of the first priorities of any updated structure. Another design flaw is the separation of the reaction data into different segments of the file, e.g. the separation of cross section data into MFs 3, 9 and 10 sections, with particle emission data in MFs 4, 5 and 6, and MFs 31-40 for covariance data. While this allows a data processing code to cleanly access these separate data, it encourages evaluators to address them individually rather than as a whole. In practice, this often leads to evaluations that update these data separately; hence, they are often inconsistent. This might be obvious, e.g. in the case where redundant data across multiple MF sections are noticeably different, or it might be very difficult to diagnose, e.g. the cross section was computed based on one model of the nucleus and the emission data based upon another. Working through a new approach to the structure and procedures for producing and storing these data in a self-consistent and mutually agreeable manner will take considerable time and effort.

Developing a format for nuclear reaction data is only the first step in the overall process. GND must support all the data types defined for ENDF, i.e. cross-section data, resonance parameters, various types of particle emission data, and covariance data. However, before such a format can be adopted, one must also provide translators, quality assurance tools, and processing routines. Translators to and from the current ENDF format are necessary to provide backwards compatibility. Quality assurance tools will provide an important element of quality control parallel with peer review. Processing routines will provide standardized code for converting functional data forms into pointwise data, e.g. Kalbach-Mann, Reich-Moore. To these ends, LLNL has also invested in the development of such tools and has provided prototype code to work with the initial draft data structure. This code is freely available and included in the GND software package. The GND code can be downloaded from https://ndclx4.bnl.gov/gf/project/gnd/.

[1] It is not a simple matter of inserting the ‘e’ because some older evaluations also skip entries in various columns, e.g. “ 2.727 +6”.

It is worth noting that modern data storage techniques separate the meta-language that defines the logical schema for the database from its conceptual data model and GND takes full advantage of this philosophy. The advantage of defining a common data model is that it can be implemented in multiple meta-languages or schemas. This means that the same data can be stored both in a human readable format such as XML, and in a high-performance, parallel processable binary format such as HDF5. This allows the user to easily adapt the data to their particular needs, which further broadens the usability of such a new format.

We wish to form a WPEC subgroup in order to encourage international collaboration towards further development of a new nuclear data structure and to work toward a consensus for adopting the new structure as the international standard for transmitting nuclear reaction data between evaluators, data centres and end users. Furthermore, there are parallel efforts to develop new formats/tools for experimental data (EXFOR) and nuclear structure data (ENSDF) and we would like to identify and take advantage of common aspects of these efforts. A unified nuclear data language encompassing experimental and evaluated data for both nuclear structure and reactions would simplify the integration of many processes and greatly enhance the nuclear data user's experience. LLNL's GND lays the groundwork and provides a solid starting point for such efforts.

Subgroup Monitor

Ivo Kodeli (JEFF)

Subgroup Coordinator

Dennis McNabb (ENDF)

Subgroup Participants

USA: Bret Beck (LLNL), David Brown (BNL), Michael Dunn (ORNL), Morgan White (LANL)

Japan: Makoto Ishikawa (JAEA), Go Chiba (JAEA), Osamu Iwamoto (JAEA)

Europe: Arjan Koning (NRG), Cyrille De Saint Jean (CEA), Andrej Trkov (JSI), Jean Christophe Sublet (CCFE)

Int. organizations: Robin Forrest (IAEA), Emmeric Dupont (OECD NEA), Victor Zerkin (IAEA), Nicolas Soppera (OECD NEA)

Definition of the project and proposed activities

This subgroup shall arrange meetings to be held at various international nuclear data centers to review and revise the GND structure and supporting validation and processing infrastructure. The goals of these meetings are to educate the various user communities about the GND structure and supporting infrastructure, and to work with the international stakeholders to revise it to maximize its benefits for the nuclear data community. At the conclusion of this effort, the subgroup will deliver a new nuclear data format/structure proposal to the Cross Section Evaluation Working Group (CSEWG). It is expected that CSEWG, which is charged with maintaining the current nuclear data format ENDF-6, will be the body that formally recommends the adoption of any new format and maintains its definition going forward.

Relevance to Evaluated Data Files

LLNL's GND makes several important improvements to the storage and evaluation of nuclear data. In particular,

- *GND is compatible with modern computer languages and practices.* Nuclear data in GND are stored in a simple hierarchy that mirrors the underlying reaction physics, making it more conducive to object-oriented programming. For example, each reaction contains both a cross section and a list of outgoing products, whereas in ENDF, data for one reaction are mixed with other data throughout the file. The GND hierarchy is easily implemented in any object-oriented language, and has already been coded into LLNL's Fudge package using Python.

- GND simplifies testing nuclear data; checking routines can be rapidly developed using Python or other modern languages. As an example, GND has already been used to improve the quality of the ENDF-VII.1 library: many issues in the library were revealed during the translation from ENDF-6 to GND. These issues include redundancies and inconsistencies (especially in mass and energy levels), and many cases of non-positive-definite covariance matrices. Many of these issues were caught and fixed in time for inclusion in the ENDF-VII.1 release.

- GND is more human-readable than ENDF. This is mainly due to the descriptive names that are used in GND in place of integer flags, but is also due to the simple hierarchical structure of GND files. We anticipate that making GND files more human-readable also makes them much easier to update, and helps streamline the process of finding and repairing bugs in the nuclear data.

We recognize that adopting a new format will add to the workload of an already overstretched nuclear data community. To assist in this transition, LLNL has already released converters for translating ENDF-6 formatted files to and from their GND structure. These converters currently handle nearly all the data types present in ENDF-6 files, and could play an important role in the transition from ENDF-6 to a new nuclear data structure. In particular, being able to freely convert back and forth between ENDF-6 and a new data structure means that users can take advantage of the improved checking and visualization capabilities, while still supplying new evaluations in ENDF-6 format for use in legacy codes.

Time-Schedule and Deliverables:

- Year 1: Educate international stakeholders on proposed GND structure and gather feedback
    - LLNL visits subgroup institutions to present current work
    - Participants capture community feedback in a written form
- Year 2: Revise format specifications
    - Develop revised format specifications
    - Adapt quality assurance and processing tools based on accepted revisions
    - Publish draft format specifications (XML schema and manual)
    - Publish draft validation requirements
- Year 3: Finalize structure and support tool requirements.
    - Iterate specifications and tools as needed.
    - Publish final format specifications (XML schema and manual)
    - Publish final validation requirements
    - Transition format governance to CSEWG

Release final WPEC Report summarizing subgroup project results