# Requirements for a new nuclear data format

Prepared by WPEC Subgroup #38 (subgroup title: *"A modern nuclear database structure beyond the ENDF format"*)

Dated: September 7, 2012

- Introduction: The purpose of this document is to define the requirements that must be followed in order for a new nuclear data structure to meet all current and anticipated future needs of nuclear data evaluators and users.

  Currently, this document is only a 'skeleton'. It lists the topics that we believe must be addressed by a final requirements document, but still needs input from the nuclear data community in order to become a complete requirements document.

  - History: This section should give an overview of the history and current status of the ENDF format. Topics include a discussion of how ENDF files are being generated and used, what groups are using ENDF-6, its advantages and limitations. This will provide a background for understanding design choices in the replacement format.
  - Purpose: Questions to answer:
    - why is a modernized nuclear data storage technology needed?
    - What should be the scope of the new format: what users must be served, and what range of data products need to be supported (nuclear reactions, nuclear structure, input parameters, etc.)?
    - Can we identify any high-level goals/philosophies for the new format (such as "The new format should use a hierarchy to mirror the underlying physics")?
  - System overview: How will the new format fit into systems that use nuclear data? For example, the data will certainly be used as input by simulation codes, but these codes may use the data in different ways. A flow diagram (such as the example in Figure 1 below) may be a good way to present an organized picture of how nuclear data evaluations are created and used. As seen in the figure, the evaluated reaction database will likely also need to link to other databases, so some form of linking will be required.
- Overall description
  - User characteristics: Many different groups use nuclear data in different ways. In order to meet the needs of all these users, a new format should be designed with their different needs in mind. This section therefore needs to give both a list of the stakeholders who will be using the new format, and an overview of how those stakeholders will be using the data. Specific questions:

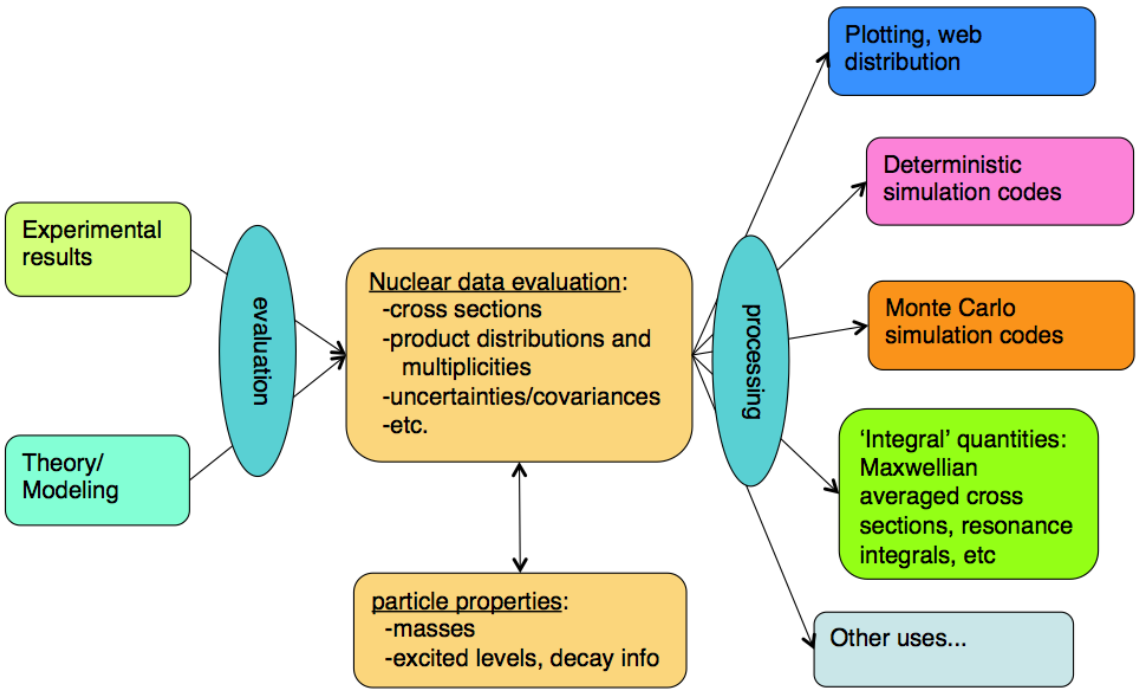


**Figure 1: Partial overview of how nuclear data is assembled and used. A more complete overview is needed.**

- What systems / OS /compilers are being used? Other relevant details?
- How can a new format provide greater flexibility to evaluators, to allow greater fidelity to the underlying nuclear reaction physics?
- How tightly is the nuclear data interface woven into important codes like MCNP, or into the web visualization tools of the NNDC, IAEA, etc.? How difficult will it be to replace the current interface?
- What new features should the new format add in order to serve users better?

- Product perspective: This section should define a high level structure for the new format. Questions to be answered:
  - What language or languages should be used to represent the new format? If we decide the format can be represented in multiple languages, should there be a single 'transmission language' defined for web display, sending data between institutions, etc.?
  - What is the most logical hierarchy or structure for organizing nuclear data? See Figure 2 for an example of the hierarchy used in the GND project. Is this sufficiently general?
  - What information must be present for a file to be considered 'complete'?

- reactionSuite (corresponds to a combination of projectile and target)
  - documentation
  - particles (lists all particles involved in the following reactions)
  - resonances (optional: resonance parameters, scattering radii)
  - reaction #1
    - documentation (optional)
    - crossSection
    - outputChannel (including Q-value)
      - product #1
      - product #2
      - ... (each product contains multiplicity and distribution information)
  - reaction #2
  - ...

**Figure 2: Hierarchy used for storing nuclear data in the 'Generalized Nuclear Data' format. Is this general enough to store all types of reaction data and to meet the needs of all users?**

        - How should we handle links, either internally for a single evaluation's file, or externally (to other evaluations, sub-libraries, particle properties database, etc.)?
    - Product structures: In addition to an overall organizational hierarchy, the new format also needs general-purpose basic data containers, similar to TAB1 and TAB2 in ENDF. This section should define what data containers are needed (2-d, 3-d and 4-d interpolation tables, matrices, etc.), along with any special requirements for each data container. Specific questions:
        - Should any of these containers be able to store non-numeric data such as 'NaN', 'undefined', or names of model parameter input variables?
        - Should there be a size limit on these containers?
        - Should axes information (labels and units identifying the contents of the container) be defined once at the beginning of the file, or repeated in each data container?
    - Constraints, assumptions and dependencies: This section of the requirements document will likely be left for later. Once the new format requirements begin to take shape, we will need to explicitly define any assumptions, constraints or dependencies involved in the proposed solution.
- Specific requirements
    - Functional requirements:
        - What types of data must be handled?
        - What types of data should *not* be handled?

    - How should the new format handle 'redundant data'? For example, ENDF permits both MT=51-90 and MT=4 to be present, so codes must check for both. Can the new format be more explicit about cases like this?

We suggest that one important functional requirement is the ability to freely translate back and forth from ENDF to the new format, in order to allow easy backwards-compatibility. Eventually, however, the new format will likely support new data types that have no equivalent in ENDF.

- External interface requirements:
    - Should we define a common API for accessing data in the new format? This could help standardize how many different codes access and use nuclear data, although an API that meets the needs of all users could become very large and hard to maintain.
- Design constraints: Some design constraints will be imposed by the choice of language for representing the data. For example, if the data are represented in XML, element names cannot have spaces, so 'cross section' must become 'crossSection' or 'cross_section'. Also, for maximum portability we may wish to avoid nesting data too deeply, since in HDF5 each 'group' (equivalent to an XML element) takes a minimum of 1 kB (more for older versions of HDF5).
- Software System attributes: The new format (and related tools) must be easily usable on all common platforms. Beyond that basic requirement, are there any special needs that should be considered?
- Performance requirements:
    - Should we impose any limits on the disk space per evaluation?
    - If a common API is defined, should we set performance requirements for the time and physical memory required when accessing the data?

These kinds of requirements run the risk of quickly becoming outdated as computer technology advances.

- Other requirements: Are there other important design factors that haven't already been addressed?
- Exclusions: During the course of preparing this document, the subgroup may decide that some questions should explicitly *not* be answered by the requirements document, either because they are beyond the scope of the current project or because the answers should

be left up to individual users rather than defined as part of the standard. This section is meant for describing these excluded issues.

- Schedule:
  - What timeline should the project follow?
  - Should we define intermediate milestones to help gauge the progress of the project?