Beyond the ENDF format: Working toward the first specifications

NEA/WPEC Sub Group 38

9-10 May 2016

Dennis P. McNabb

Lawrence Livermore National Laboratory

LLNL-PRES-654634

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

We've made excellent progress toward the development of a modern database structure/representation

- Detailed requirements have been drafted and reviewed
  - Many thanks to David Brown, Bret Beck and Caleb Mattoon for documentation efforts
  - Living document, but done in the sense that these documents capture our thinking and areas to be fleshed out
- Specifications are mature enough to move forward with infrastructure development (will still evolve, but basics are solid)
  - Low level specs in good shape – good enough for EXFOR or other data products
  - Particle properties in good shape –ENSDF community has also been engaged
  - Top level has been fleshed out, less frequent options still need some review
- A lot of supporting infrastructure still needs to be done
  - API, Processing, QA
  - Documentation and governance

My proposal is to close SG38 with a summary report that references all of our documentation that has been prepared, and start two new subgroups.

The two subgroups

Long-term subgroup: International standard for a modern general nuclear database (GND) structure

- A long-term sub-group is needed to serve as a governance body
- Format is reasonably mature, enough to work on infrastructure

Regular subgroup: Code infrastructure to support a modern general nuclear database (GND) structure

- Code infrastructure to work with GND will enable international adoption
- There is a lot of work to do, so I think this deserves a separate focused effort
- After completion, long-term subgroup can oversee efforts

I propose to resurrect Long-Term Subgroup B on Formats and Processing

New Name: *International standard for a modern general nuclear database structure*

- Approve and release the initial version;
  - Deliverable 1: Documentation of the initial version.
- Provide a productive environment to modify and extend the standard;
  - Deliverable 2: Collaborative platform to maintain and discuss the standard
- In close collaboration with the short-term subgroup on infrastructure ensure that important and useful tools for using the new standard are developed and maintained; and
- Release new versions with appropriate documentation as necessary, without overburdening stakeholders.

Infrastructure needs identified by SG38 as being (fundamentally necessary)

- Tools for generating new evaluations using the new structure
- Checking codes to help validate new evaluations and fix problems identified during validation. These include checks for properly formatting, for completeness, and for sensible physics content (e.g. conservation of energy)
- Visualization tools
- An Application Programming Interface (API) for reading and writing data in the new structure
- Initial infrastructure for manipulating and processing nuclear data
- Tools to assist with uncertainty quantification (UQ) studies using the covariance estimates that are being expanded with recent releases of nuclear data libraries

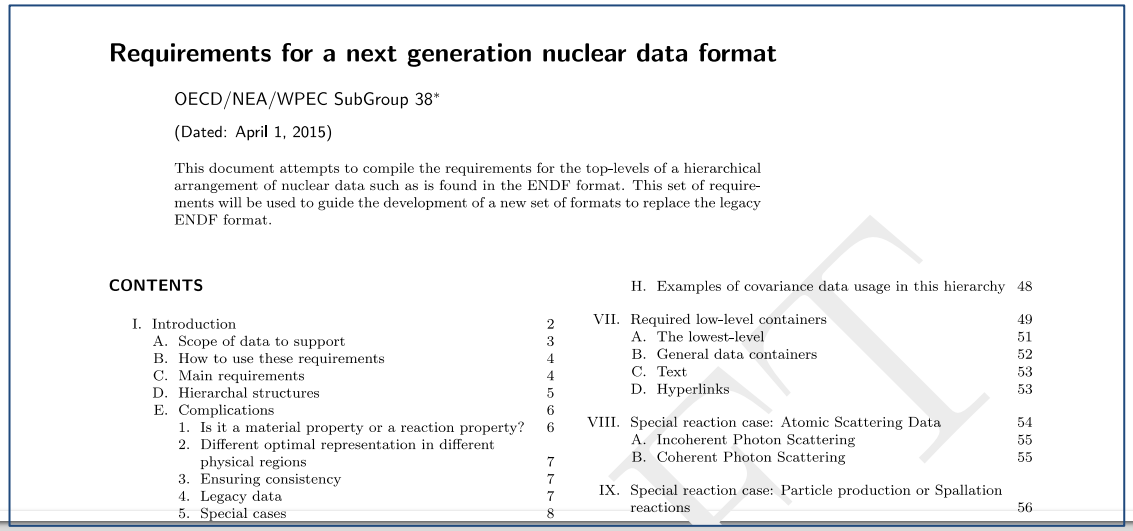
Scope looks too big to me, what are our priorities?

Discussion topics for long-term governance

- Members versus Participants
- General Principles
- Governance Board
  - Chair, Vice-Chair, Secretary (hopefully WPEC secretary)
  - One rep from each member institution
- Getting consensus
- Voting

The main work product this year is 3 draft documents that capture the detailed requirements and draft specifications of the proposed structure

1. Low-level data structures
   - Requirements and specifications more driven by general considerations and computational issues
   - Perhaps applicable to a broader range of scientific data storage applications
2. Top-level reaction hierarchy
   - Requirements and specifications driven by nuclear reaction database considerations
3. Particle properties hierarchy
   - Looking to the future, particle properties not dependent on reaction mechanism should be stored in a separate evaluated database to avoid internal inconsistencies, e.g. masses, decay properties



Requirements for a next generation nuclear data format

OECD/NEA/WPEC SubGroup 38*

(Dated: April 1, 2015)

This document attempts to compile the requirements for the top-levels of a hierarchical arrangement of nuclear data such as is found in the ENDF format. This set of requirements will be used to guide the development of a new set of formats to replace the legacy ENDF format.

CONTENTS

The level of detail and iteration has been substantial

(~180 pages)

We've broken the work into seven products

1. Low-level data structures
   - Requirements and specifications more driven by general considerations and computational issues
   - Perhaps applicable to a broader range of scientific data storage applications
2. Top-level reaction hierarchy
   - Requirements and specifications driven by nuclear reaction database considerations
3. Particle properties hierarchy
   - Looking to the future, particle properties not dependent on reaction mechanism should be stored in a separate evaluated database to avoid internal inconsistencies, e.g. masses, decay properties
4. Collaborating on visualization, manipulation, and processing tools
   - Agree on a path forward for comparing and cross-validating tools working with the new format
5. API for reading and writing data in the new structure
6. Testing and quality assurance practices
7. Governance

We have captured our plans in a report

A possible path forward

- I think it is reasonable to expect that we have a useable draft format specification by next year
  - Version 0.1
- However I expect QA will take some time and the format specification will continue to evolve
- At next year's meeting, we are planning to propose a long-term Subgroup in order to establish a governance model for the format going into the future
- For the first few (5?) years, we suggest that each Data Project release two versions of their databases
  - Production version in ENDF102
  - Opportunity version in ENDFX (or whatever name we decide to adopt)

This approach allows the new format to mature (QA)
and the user communities time to become familiar with new tools

Some initial thoughts on governance

- Every member in the long-term Subgroup contributes and has a voice (vote) in changes/extensions to specifications and supporting infrastructure
- The long-term Subgroup will need a Governance Board to organize the Subgroup's activities
  - Each project should appoint a member to the Governance Board
  - A Chairperson, Secretary, and Dissemination Guru should also be appointed to the Governance Board
- The Chairperson should report the Subgroup's activities to WPEC yearly for review and feedback

One year from now, our goal is to have developed a new draft structure and toolset that can be adopted and maintained by the community

- These two days we are in the heart of the project: developing the requirements and specifications for the new structure
  - Basic data containers
  - Top-level hierarchy
- By November meeting I'd like to see convergence on an acceptable draft set of requirements and specifications
  - I propose that the November meeting be held the week after CSEWG at BNL (or LLNL, if BNL unable)
- I updated our Implementation Plan after last workshop in Japan
  - I will ask Emmeric to print copies
  - Please take a moment to review, as I plan to submit it to WPEC as part of our progress report and as an official document of the project

A principle going forward: Try to propose modifications or extensions of the current drafts, instead of general criticisms

- More efficient to the extent that we can move toward improving the documents before us
- Any long-term governance will require a similar approach
  - Might as well get used to it

The governance must perform some specific actions

While ensuring that the standard meets the needs of major international nuclear data communities, this governance body must

- Approve the initial version;
- Provide a productive environment to modify and extend the standard;
- Ensure that important and useful tools for using the new standard are developed and maintained; and
- Release new versions with appropriate documentation as necessary, without overburdening stakeholders.

General principles guiding the governance of the new format standard

- All members of the nuclear data community in good standing and interested in contributing will be allowed an equal voice in the definition of the standard;
- Consensus will attempted for all decisions taken by the formats community when feasible; and
- Members will facilitate the transfer of their technical expertise and general-purpose infrastructure in support of a successful format standard.

Basic governance structure

- Governance Board:
  - Maintain membership lists, organize/lead meetings, organize necessary votes, and provide for the dissemination of new versions of the standard and associated tools.
  - Each participating institution should appoint a member to the Governance Board.
  - The Governance Board will appoint a Chairperson, a Secretary, and a Dissemination Guru. The persons fulfilling these leadership roles will also be on the Governance Board.
- Executive Board:
  - Review and accept (1) new versions of the standard and (2) leadership appointments of the Governance Board.
  - The Executive Board can also overrule any decision by the Governance Board to declare a member of the community no longer in good standing and is expected monitor and provide guidance to the Governance Board.
  - The Executive Board will be the WPEC Executive Board. Any changes to this document outlining the operation of the long-term subgroup will be undertaken under the existing rules for such changes maintained by WPEC.

Governance Board leadership roles

- Chairperson
  - Organizes and leads meetings to discuss and develop the format and supporting infrastructure.
  - There will be at least one meeting of the collaboration per year.
- Secretary
  - Maintains membership lists (members in good-standing and members of the Governance Board)
  - Executes any formal votes, if required.
- Dissemination Guru
  - Maintains a website to disseminate the documentation and infrastructure for the formats standards.
- Institutional members
  - Nominate new members in good standing, and in doing so provide some evidence for the nominee's expertise and willingness to contribute. The Secretary will request from the Governance Board a vote on the acceptance of each new member. Concurrence from a simple majority shall suffice.
  - Institutions that do not participate in community discussions and meetings for a period of three years or more will be asked to step down as members.

Collaboration member roles

- Improve the format and data quality, supporting infrastructure,
- Generally encourage and support the broad use of the new standard by the nuclear data community
- Collaboration members are expected to release their contributions with a relevant open source license that enables the community to make full use of their contributions.
  - For computer codes, a public domain or open source attribution (e.g. BSD) copyright is recommended
  - For documentation or specifications, a public domain or agreed upon attribution-share alike license is recommended

Votes on format specifications and other proposals (e.g. new release)

- Structure to encourage productive, consensus-driven decisions.
- All votes are fully public.
- Members have 3 options during a vote:
  1. Agree with the proposal as it stands
  2. Abstain from voting due to a lack informed knowledge
  3. Submit an argument against the proposal and provide a sensible counter-proposal