

## Requirements for a new nuclear data structure Part 2: Implementation Plan

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

Dated: May 12, 2014

**Introduction:** This document summarizes the implementation plan developed during the second meeting of the WPEC Subgroup 38 (SG38), which was organized to develop a new evaluated nuclear data structure<sup>1</sup> and then oversee the transition from the current standard (ENDF-6) to the new structure. Part 1 of this document, adopted by WPEC in May 2013, lays out the vision and goals for the new structure. In this second step, SG38 develops a community plan to address these needs and requirements. The plan laid out here represents a consensus on how to execute the project, what work will be done, and to some degree how it will be done and the people involved.

During the development of the vision and goals for the new format it was recognized that the application of modern programming and database practices will have significant benefits for nuclear reaction databases, both for those of us engaged in producing this data and those who utilize the data for applications. It was also appreciated that additional benefits would be realized if the low-level data containers used for storing reaction data were general enough to be shared with other nuclear data products, such as EXFOR, RIPL, and ENSDF, so that codes interfacing with these different databases could share the same set of routines for reading and writing data structures. However, it was also acknowledged that adoption of these new tools and capabilities will be difficult without some supporting infrastructure in place to use the new data structure, specifically open source codes to manipulate, search, plot, and process the data, as well as tools to translate data to other formats in current use and to check the data for quality.

In order to address this broad set of goals, the SG38 project decided to organize the work around seven different products:

1. Low-level data structures
2. Top-level reaction hierarchy
3. Particle properties hierarchy
4. Visualization, manipulation, and processing tools
5. API for reading and writing data in the new structure
6. Testing and quality assurance practices
7. Governance

---

<sup>1</sup> In this document we differentiate between a 'structure' and a 'format'. A format includes both a structure (that is, the overall organization of data) plus the details of how it is represented in a file.

In each case, product team leads were identified from amongst the subgroup's members, who were responsible for guiding the development of an implementation plan for each product. At SG38's May 2013 workshop, team leads proposed a path forward which were then discussed and refined. The purpose of this document is to lay out the plan for what we would like to accomplish before the Subgroup 38 is closed and the work is handed to a long-term Subgroup to modify, adopt and maintain into the future. Ideally, the plan has sufficient detail to not only guide and focus the work, but also to serve as the basis for planning work and acquiring resources at each participant's home organization.

It is important to recognize that the work plan laid out here is geared toward developing a nuclear data structure or language, and does not develop a style guide that completely serves to replace the current standard (ENDF-6). It was decided that there was too much work at this point to tackle the joint desires to develop a language with maximum flexibility, while also providing a style guide that constrains evaluated data products for reasons of completeness, efficiency, and best practices. It was decided to leave that activity to a future long-term subgroup (Task 7 in the work plan below), which will serve to maintain and govern the nuclear data structure and the requirements for acceptable reaction data evaluations into the future. Nor did it seem necessary to define a product team to deliver the desired open source infrastructure. Instead, there seemed general willingness from several institutions (e.g. LLNL, CEA, Kurchatov) willing to both develop and provide their respective data infrastructure as open source. So instead the work plan here seeks to coordinate these activities to make code comparisons easier.

While developing the work plan, it was decided that in order to shepherd the adoption process, each data project should identify a main point of contact whose responsibilities are to be first adopters of the technology and to assist their data project's members in transitioning to this new technology. The LLNL team is committed to providing their FUDGE infrastructure in a timely manner to facilitate this process and provide tools for these first adopters.

## Product 1: Low-level data structures

**Statement of work:** Provide specifications for low-level data containers, APIs to read and write these containers (C, C++, Fortran, Java), and a proto-type substantiation of an API that is tested by relevant stakeholders (ENDSF, EXFOR, GND, RIPL, and those involved in Product 5 below).

**Resources required:** Team leads, M. White & V. Zerkin, will coordinate the development of requirements, specifications, APIs, and proto-type interface. Representatives of the various stakeholders -- including A. Sonzogni, B. Beck, C. Mattoon, R. Capote, and W. Haeck -- will be asked to review the major deliverables of this activity.

### **Tasks:**

1. Write requirements document
2. Review requirements document
3. Write specifications document
4. Review specifications
5. Write XML Schema Document (XSD)
6. Review XSD
7. Write API document
8. Review API document
9. Prototype substantiation of an API
10. Review/test API substantiation

**Discussion & Risks:** Existing numerical and string data types in EXFOR, ENDF, and GND were discussed. A broader discussion indicated that there may be a need for specific data types for metadata, bibliography entries/citations and comments in addition to a simple string container. Also, the question was posed whether N-tuples could be used for a rather general numeric container instead of specific ones for each N-tuple. On the other hand, perhaps more specific data types would make interpolation easier. A specific type for a uniform grid in multiple dimensions was also proposed as potentially useful. V. Badikov proposed specific types that came with constraints (e.g. unit normalized or other imposed correlations such as enforcing balance in decay schemes, ie decay in equals decay out, improves data and reduces uncertainty by making this requirement in the analysis). There was general agreement that containers should allow for various clearly defined attributes to be specified including data type (e.g. string, int, float), units (e.g. keV, barns), name (e.g. energy, angle, probability), and also NotANumber should be defined. Arguments were made that the domain region should be defined in header information and checked to make coding easier. Despite all these conflicting needs and requests it was generally desired that the number of low-level data types be kept to minimum level such that this product team's task could come to a close quickly. Keeping the number of data types to a minimum will ensure better readability and that coding is relatively straightforward. In this vein, it was decided that taking on functional primitives at this time contained too much risk and was not well enough defined and so was tabled as a future activity.

## Product 2: Top-level reaction hierarchy

**Statement of work:** Specify how nuclear reaction data should be organized. Many kinds of data need to be handled: cross sections, energy and angular distributions, particle production number and spectra, resonances, atomic processes and covariances. Special consideration also needs to be made for thermal scattering  $S(\alpha,\beta)$ , fission yields, and for documentation. The project team will need to decide whether to make a single hierarchy general enough to handle all types of data, or to have multiple hierarchies for organizing different data.

Our recommended approach is to start at the top of the hierarchy and work down. The first decision would be to determine what order the following elements appear in the structure: projectile, target, reaction channel, type of data (cross section, distribution, etc.), and incident energy.

By first reaching consensus on how to organize these five elements, we can better frame the discussion for how to organize the remaining data that fits inside this hierarchy.

**Resources required:** Team leads, D. Brown & A. Koning, will coordinate reviews of existing GND format specifications for reaction hierarchy. Other reviewers include A. Trkov, J.C. Sublet, S. Kahler, L. Leal, Jouanne, Archer, and V. Badikov. GND representatives B. Beck and C. Mattoon will prepare for reviews and document specifications agreed upon by reviewers. The community will work jointly toward developing the requirements and specifications for the top-level hierarchy

### **Tasks:**

1. Prepare overview of possible ways to organize the following elements: projectile, target, reaction channel, data type, incident energy, resonances and covariances
2. Review these specifications
3. Discuss and decide how the new structure should organize these elements.
4. Prepare an overview of possible ways to organize other elements
5. Review these specifications
6. In a workshop, discuss and converge
7. Write a document giving an overview of the hierarchy and requirements to be met
8. Write specifications document and XML Schema Document (XSD)
9. Review specifications and update

**Discussion & Risks:** An important issue raised was whether or not a default interpretation should be assumed when no data is specified. Some examples included assuming isotropy when no angular distribution provided, assume zero covariance when none provide, but there were many others. However, the spirit of the requirements document was to provide a nuclear reaction data language, and that evaluated data constraints would be provided by the data projects themselves. For example, inclusive data are allowed without requiring a consistent set of exclusive data (though projects can have their own consistency rules). Defaults seem possible, and in some cases might be needed to save space, but need to be clearly laid out and unique. In fact, might be good to define some useful defaults with specific tags, e.g. <isotropic>, <zeroMatrix>, etc. From a language perspective the data should be assumed unspecified if not provided as the case in ENDF-6. For example, it would be difficult to consider defaults for experimental data – only for evaluated data. On the other hand, requiring all data be specified has real benefits from a quality assurance point of view. In the case of units, the requirements state that the physical units must be specified, for example. And in the current era, tools like TENDL can provide data relatively quickly which in many cases may be more realistic than a default. So it would seem likely most data projects will adopt this view.

Another important issue raised was that projectile-target specifications may need to expand in the future to allow for data products not currently provided. These specifications may need to recognize compounds, mixtures, different physical states, and isotopic distributions for composite particles, as well as other baryons, mesons and leptons (to be provided by Task 3 below). One issue raised is that if we will be taking particle and target information from a linked in RIPL, we should require an explicit link to a specific version of RIPL be provided. Or is there a default version based on a date field?

It was noted that we don't currently have any atomic physics expertise on the committee and would find that useful if we plan to extend efforts in that direction. However, our requirements do not require us to specify atomic processes for this project, and so this task could probably be tabled for a future project activity.

In general, it is important to have the right expertise to review the various components of the database structure.

Another issue raised was that "background" cross sections in the resonance region are not physical and are in some sense part of a fit or data form for specifying data in the resonance region.

## Product 3: Particle-properties hierarchy

**Statement of work:** Develop a hierarchy to include masses, level schemes and lifetimes of useful particles, including data that is independent of the reaction used to form the particle. Add isotopic abundance tables and possibly masses for mesons, baryons, and leptons from the Particle Data Group. Provide an XML specification for the database and develop translators to convert current RIPL format into the new format and back as a validation exercise.

**Resources required:** Team leads, R. Capote and C. Mattoon, will coordinate the work. Reviews and tests will need to be performed drawing on members of Products 1 & 2.

### **Tasks:**

1. Write short requirements document detailing scope of data to be included in new particle database.
2. Review requirements document
3. Write document that outlines model for a particle data hierarchy
4. Review model outline
5. XML Schema Document (XSD)
6. Review XSD
7. Develop translators to convert current RIPL format into XML and vice-versa
8. Develop translators to convert AME, PDG and IUPAC data for isotopic abundance tables and masses for mesons, baryons, and leptons into XML
9. Write a users manual for a first draft database
10. Review and test draft particle database to serve as a possible starting point for data projects (similar to the role the GND has played)

**Discussion & Risks:** There is an open question regarding the difference between decay libraries and this new database. Should this new particle database include alpha and beta decay to excited states, for example, or is that what distinguishes the two databases? Since decay libraries may change more, perhaps one should allow for two links – one for masses and levels with gamma decay and another for other decay data? There was a fairly clear consensus that this task does not need to include resonance parameters, which are usually part of a nuclear reaction evaluation or optical model parameter sets, though optical model parameters should be included in documentation.

One risk identified was a lack of understanding amongst reaction data evaluators and the community regarding expectations for using RIPL as a link for future evaluations. It was argued that we need a strategy for getting the word out. Specifics were not discussed, however. It was suggested that it would be useful to identify some first adopters to perform evaluations without going through ENDF, with community talks on the experience, including highlighting links to the new particle database.

## Product 4: Visualization, manipulation, and processing tools

**Statement of work:** Develop a list of agreed-upon infrastructure functions and specify their name, purpose, inputs and outputs in order to facilitate infrastructure development and testing. These teams will also begin the process of releasing their infrastructure codes as open source as their individual institutions permit, possibly as partial releases if necessary. The various SG38 contributors will update their infrastructure to work with these new structures.

**Resources required:** Team leads, B. Beck, F. Malvagi, and V. Sinita, will develop an initial list of infrastructure functions and an approach to performing cross comparisons. Other participants, including S. Kahler and M. Dunn, will review and all parties will spearhead efforts to update their infrastructure to work with these new structures and to the extent possible release their infrastructure as open source.

**Tasks:**

1. Document list of data handling, processing, translation, search, and plotting functions for cross comparisons
2. Develop an approach to performing cross comparisons of these functions across different sets of infrastructure
3. Update and test infrastructure software to work with new structures (GRUCON-3, TRIPOLI, NJOY, FUDGE, AMPX)

**Discussion & Risks:** In order to set clear expectations for each members institution, Product 7 below will identify an agreed-upon approach to software development and licensing to be used for the project.

## Product 5: API for reading and writing data in the new structure

**Statement of work:** Document APIs to read and write the new data structure in C, C++, Fortran, and Java. Implement a substantiation of an API that is tested by relevant stakeholders of Tasks 1 & 2 above and the developers of processing and testing codes.

**Resources required:** Team leads, B. Beck & W. Haeck, will develop a conceptual design, iterate with Task 2, develop APIs and then implement. Representatives -- M. White, V. Zerkin, D. Brown, A. Koning, F. Malvagi, M. Coste-Declaux, G. Chiba, S. Kahler, M. Dunn, V. Sinitisa, J.-C. Sublet -- of the various stakeholders will review the major deliverables of this activity.

### **Tasks:**

1. Draft a conceptual design of a nuclear data object
2. Review conceptual design
3. Write API specification document
4. Review API document
5. Implement substantiation of nuclear data object
6. Review/test API of nuclear data object substantiation
7. Implement reading from a file
8. Review/test reading from a file
9. Implement writing to a file
10. Review/test writing to a file

**Discussion & Risks:** It can be very time consuming to iterate coded versions of an object as a way to converge upon a final design and given the activities going on with Products 1&2. So Haeck proposed that he and Beck draft a conceptual design of nuclear data object in a written document first and iterate this design with Product 2. Tasks 3-10 above wouldn't start until Products 1&2 have put forward a rather advanced set of specifications.

One issue that was left unresolved was where does the API end and data processing begin? Should API do some processing, such as interpolation, linearization, convert resonance parameters into point-wise data? The conceptual design document will need to resolve these issues.

A related issue is how low-level or high-level should the interface be? A low-level interface is straightforward: get data as written, but performing simple operations on data for processing with controls makes some sense. There was general agreement that we should lean toward a higher-level interface, but that we want to ensure that the scope be manageable within the project guidelines. The conceptual design document by the team leads should define a delineation between processing and the API, what needs to be stored or represented in a nuclear data object, and how that object might be organized and constructed.



## Product 6: Testing and quality assurance practices

**Statement of work:** This team will develop a list of agreed-upon quality assurance (QA) tests and their specifications. The focus will be specifying required unit testing, in order to assist the institutions providing open source data handling infrastructure and the nuclear data centers in meeting their QA requirements. The team will also provide a summary of suggested best practices for regression and data validation testing.

**Resources required:** The team leads, M. Dunn and J.-C. Sublet, will develop the unit testing requirements, and coordinate a QA plan with the data assurance managers (D. Brown, A. Trkov, O. Iwamoto, V. Sinitisa, M. White, V. Zerkin(?)) at each participating data center. Stakeholders for Products 4&6 -- B. Beck, F. Malvagi, V. Sinitisa -- will need to work closely to implement the tests required to meet the unit testing requirements for the QA plan.

### **Tasks:**

1. Draft of goals and requirements for unit testing with initial list of unit tests
2. Review QA requirements document
3. Develop the QA plan with detailed specifications for required unit tests
4. Review QA unit test plan
5. Draft summary of suggested best practices for regression and data validation testing based on industry standards.
6. Review best practices summary
7. Review/test QA tests as implemented by laboratory that first provides QA test implementation
8. Nuclear data centers integrate into their respective QA assurance models

**Discussion & Risks:** A full QA process includes unit testing, regression testing and data validation testing. During the discussion it was decided that for the purposes of providing a nuclear data structure (or language, if preferred), the SG38 project should focus on providing the unit testing necessary to assure that data structures are correctly formed, and let the respective data centers continue to provide the relevant regression and validation testing. However, comparison testing between processing code infrastructures (as discussed also by Team 4 above) is also viewed as an important activity that is expected to emerge as the processing codes agree on infrastructure functions that can be compared. Dunn made a case that the community should begin to think about adopting community standard for QA across data project that meets industry standards such as ISO-9001, NQA-1 Part IV, Subpart 4.2, which are US best industry practices.

Sublet made an initial list of areas for which unit tests need to be specified:

- Resonance parameters
- Physical constants
- Cross sections
- Multiplicities
- Emitted spectra
- Angular distributions
- Derived forms
  - Energy grid, temperatures
  - Thresholds, normalization
  - Dilutions, self-shielding factors
  - Decay data, energy levels and their consistency

Task 5, actual implementation, should probably wait for Product Team 5's activity to be completed.

## Product 7: Governance.

**Statement of work:** Organize and create web interfaces for documentation and links to relevant software and codes. Propose a governance model for the long-term management of the nuclear reaction data structures, formats and infrastructure. This governance model will serve as the proposal for the creation of a long-term WPEC subgroup after SG38 activities are completed.

**Resources required:** The team leads, Dupont & McNabb, will develop the public web interface to the documentation and code infrastructure for the new nuclear reaction data structures/formats. BNL has agreed to and will need to provide some staff to support the web site where the development activities for SG38 will be stored in a Gforge repository. The public interface to the current release with supporting documentation and links to infrastructure will also be hosted by BNL. The team leads will also propose a governance model for the new structure. Team leads and data center leads will need time to review web sites and proposed governance model.

### **Tasks:**

1. Update BNL Gforge repository to support wiki-based coordination amongst Product Teams.
2. Select open source model
3. Web interface for public dissemination of codes and documentation
4. Review web interface and open source model
5. Draft governance model
6. Review governance model
7. Propose long-term WPEC subgroup to maintain nuclear reaction data format/structure

**Discussion & Risks:** Plans for a mirror site were not discussed – this was an oversight.

A first step for a proposed governance model was to decide upon an open source license that the SG38 code products should be developed under. There is some urgency on this point, as each institution typically needs to get approval for shared development projects.

While SG38 is developing a very flexible infrastructure for storing and communicating nuclear reaction and decay data, a concern raised at the WPEC coordination meeting was that evaluated data tables should have additional restrictions and specifications to ensure that the data tables meet the needs of a large-scale, multi-physics simulation codes. The long-term SG38 subgroup may want to consider providing and maintaining an international evaluation style guide (and supporting infrastructure) that would be shared by each nuclear data center.

Some discussion of roles and responsibilities for a long-term subgroup was discussed. A QA manager for ongoing annual assessments and QA improvements, a records custodian to maintain and improve website and documentation, and a data processing coordinator to maintain and improve infrastructure available in the community were all mentioned as important roles. Also a process for adopting new data forms will need to be specified – a useful starting point for this will be the ENDF-6 formats process.

Shmakov presented D-diagrams – a pictorial method for documenting formats/structures. It is used by RFNC-VNIITF to develop and document their codes. There was general agreement that this kind of tool would substantially improve the clarity of our documentation.