



Finalization of the scope, Actions and AOB

Amanda Lewis
WPEC SG50
September 15, 2020

The Naval Nuclear Laboratory is operated for the U.S. Department of Energy and the U.S. Department of the Navy by Fluor Marine Propulsion, LLC.



Discussion – Copyright/IP requirements

- NEA Github can be used for work-in-progress codes
- How should the codes be made available when finished? (Or when versions are complete)




Discussion – On the shoulders of giants

- A significant amount of work has been done already:
 - EXFOR Correction System
 - X4Lite
 - JSON database
 - x4i
 - GNDS format


- How should we build on this work?




Summary – EXFOR and current work

- EXFOR has a hierarchical structure, easy to translate overall format into an object-oriented database
 - Contains valuable “quality” information, including `STATUS` keyword
 - Ambiguities and fluctuations
 - Format of year, tables
 - Free text
 - Old standards/decay data, monitor ambiguities
 - Some reaction string definitions (e.g. FY from heavy-ion reactions)
 - Error analysis sections vary based on paper
- 


Summary – EXFOR and current work

- “Everyone writes their own parsers” – a lot of overlapping work, but also a lot of great options for us to consider
 - EXFOR correction system –
 - Has an archive of monitors to pull from, including old library evaluations
 - Idea of storing the code to apply the corrections, not multiple versions of the same data
 - X4Lite - Translates database into various formats, automatically retrievable
 - JSON database
 - Extracts `ERR-ANALYS` components when in `(ERR-1)` format
 - Powerful search capabilities with database
 - X4i - Treats reaction strings as grammar
- 

Summary – EXFOR and current work

- EXFORTables
 - Format similar to EXFOR, more readable, has C/E values that suggests quality (if you trust the comparison)
 - Can quickly put together distributions, find outliers
 - Uppsala procedure can be used to test new features such as common errors storage
 - Gaussian Process, Measurement Feature ML methods can be used to test new layers for readability
 - Can learn from ENSDF database update process
 - FY evaluation made a “modified” EXFOR format
 - only specific necessary information – good start for format specifications by observable
- 

Summary – Philosophy


- Needs to be machine readable
 - Minimize free text, ambiguities and fluctuations
 - Separate storage and transmission
 - Choose a format that works best for our database, then write codes to translate to the many other formats in use
 - Needs to be curated
 - Quality assessments by experts, documented
 - Real-time feedback, corrected data sets through peer-review process
 - Decide what to store and what to generate on-the-fly
 - Large covariance matrices (SAMMY/AGS method – stat + sys?)
 - Codes to make corrections on data sets? – optional additions, rather than separate files
- 

Summary – Philosophy


- Needs to have complete documentation
 - For all corrections applied
 - Tag and document data as used in evaluations
- Keep consistency with GNDS
 - Resonance parameter and <XYs1d> containers
 - Keep consistent with v2.0 updated documentation format
 - Keep language consistent – “capture” vs “n,gamma”



Summary – User needs

- Curated data to be put right into automatic processes
 - Leaving this up to the user can result in chaos
 - Many automatic processes are very sensitive to outliers
 - Realistic covariance matrices
 - If given by author, use that, otherwise estimate
 - Interpolate onto different energy grids
 - Systematic uncertainties standardized
 - Ability to store correlations between data sets
 - List of corrections that are expected for an observable
 - Experimental conditions for treatment of RR data
- 

Summary – User needs


- “Default” results when users search for a reaction:
 - Tag by library release?
 - Set of data sets used in ENDF/B-V.III evaluation
 - Let users choose library, like they do now with evaluated
 - Collaborate with SG49 for documentation
 - Tag by some measure of quality?
 - Do need positive, negative and neutral flags in this case
 - Some data never makes it to Layer 3?
 - If it can't be corrected or uncertainties correctly applied
 - Need to address issue of pipeline – if nothing can go into Layer 3 until it is looked over by a human, then it might end up being just evaluation-tagged data anyway
- 

Finalization of Scope


- Isotopes/observables:
 - CIELO isotopes
 - Include all differential observables (XS, FY, DDX, PFNS, ...)
 - Include neutron, charged particle, gamma induced reactions
 - Pull out all the skeletons at the start
- Energy regions
 - Thermal?
 - RR
 - Fast



Finalization of Scope

- There is a lot of work to be done, and a lot of people here who would like to help!
 - Organize the subgroup in various sub-subgroups:
 - Metadata/uncertainty keywords
 - Coordination with NRDC
 - Database creation and structure
 - Layer creation codes
 - “Transmission” translation codes
 - Testing output files
 - Correction procedures/documentation
- 

Metadata/Uncertainty Keywords

- Would start as 5 or 6 different groups, one each for different energy regions/observables
 - Each group would take a few months to decide on the metadata they need for their observable
 - Some already have a good base of work
 - RR metadata template (more work still – resolution function, different eqns)
 - unc templates, especially (n,f), PFNS, nu-bar
 - The resulting metadata lists will be compiled into one big list of metadata, given consistent keywords
 - The format will be made consistent with GNDS
 - Alias tables can be constructed to help with translation
 - Continuous improvement based on feedback
- 

Coordination with NRDC

- Define feedback mechanisms to NRDC
- From Metadata sub-subgroup, decide on which keywords to ask NRDC to add to EXFOR
- Improve ability to translate from EXFOR to Layer 1 over time
- Continuous feedback about ease of compilation



Database creation and structure

- Choose a database type
- Figure out where the database should be stored
- Create database by the end of the year?
- Define structure for all three layers
 - In Layers 2 and 3, will need to connect data set to correction documentation
- Work with formats group to define requirements and specifications




Layer creation codes

- Create the code to translate from EXFOR to Layer 1
 - A lot of work already done, can choose best parts of each various code, need to figure out what to do about free text, ambiguities
 - Work can start before format is finished, but is dependent on that, and will provide feedback to formats sub-subgroup
- Create the code to translate from Layer 1 to Layer 2
 - This will include automatic renormalization, flagging of too-small uncertainties, etc
- Create the code to translate from Layer 2 to Layer 3
 - Apply the corrections as determined by the corrections sub-subgroup.



“Transmission” codes

- Create API, and any other interfaces to the database
 - Create codes to translate to current formats (C4, C5, etc)
 - This translation will in some cases require loss of information, but this will entice users to update their codes away from C4, for example, but at their own pace
 - Work closely with formats and testing groups
 - There is already a good base of work to draw from
- 

Testing output files

- Define set of tests to use
 - Uppsala evaluation procedure
 - ML projects (GP, measurement features)
- Perform tests on sample files
- Provide feedback and iterate



Correction Procedures/Documentation

- Decide which corrections are “objective” (Layer 2) and which are “subjective” (Layer 3)
- Define process to document objective corrections, flags
- Define process to make subjective corrections
 - Should include peer-review, similar to ENDF process
- Define minimum documentation for subjective corrections
- Define tagging system
 - Versions used in various library evaluations
 - A current “best” version that is the default result for users?

Finalization of Scope

- There is a lot of work to be done, and a lot of people here who would like to help!
 - Please volunteer for the sub-subgroups! Here or by email after this meeting
 - Metadata/uncertainty keywords
 - Database creation and structure
 - Layer creation codes
 - “Transmission” translation codes
 - Testing output files
 - Correction procedures/documentation
- 