

EXFOR への変換等における NRDF の問題を検討するワーキンググループ報告
A Report of the Working Group to Discuss Problems in Transformation
of NRDF-Data to EXFOR

北海道大学大学院理学研究科
加藤 幾芳
Kiyoshi Katō
Graduate School of Science
Hokkaido University

Abstract

A working group has discussed the problems in the transration of NRDF data to EXFOR. Based on the troubles happened at the transration, some basic problems were investigated. This is an interim report of those problems discussed in the working group, including several proposals for the solutions.

1. ワーキング・グループとその任務

NRDF 活動における重要な課題の1つとして、NRDF データを EXFOR システムに変換して国際的データ交換に参加協力することが挙げられる。すでに、我々の EXFOR データへの寄与は、EXFOR の全データの 5% を越えており¹⁾、決して少なくない。この変換は NTX²⁾と呼ばれるシステムを通して行なわれており、システム自身大変順調に働いてくれている。しかし、現在、NRDF データから EXFOR への変換率は必ずしも高くはない。その理由として NRDF と EXFOR のデータベースに対する基本的な考え方の違いにあることが変換活動をスタートする時点から認識されていた³⁾。

例えば、NRDF では出来るだけ多様なデータを収集できるように、フリー・テキスト⁴⁾と呼ばれる入力形式や 対応するコードが無くても "DATA"⁴⁾ というコードで入力できるようになっている。しかし、そのような形式で入力された NRDF データは EXFOR システムに機械的に変換することを極めて困難なものとする。EXFOR データは始めからデータの比較的特定の利用を目的としていることから、well-definded なデータだけを収集する。well-definded なデータであることは大変望ましいことではあるが、一方、収集するデータは非常に限定されてしまう。

従って、NRDF から EXFOR への変換率を上げるには、

1) RRDF の入力データのうちコードやデータ入力形式について定義や使い方をより明瞭にすると共に新たなコードや入力形式を導入することによって変換できるデータを増やす、

2) NRDF データで EXFOR に対応すべきコード等がないものについては、EXFOR に新たに収集すべき物理量の提案とコードの申請を行なう、ことが考えられる。これらのことを検討するとともに、他に、NRDF から EXFOR への変換率を上げるためにどのような問題があるか検討するためのワーキング・グループを作成した。

メンバーは、私のほかに、NRDF から EXFOR への変換プログラムを作成を作成し、これまで変換作業を行なってきた千葉正喜札幌学院大教授と、NRDF のコード、辞書作成を担当して来た能登 宏北星学園大教授でスタートし、能登氏が外国出張で留守になってからは、NRDF から EXFOR への変換プログラム作成に協力してきた片山敏之北星学園大教授にメンバーとして参加していただいた。

本論は、ワーキング・グループでの議論を整理したものであり、いわゆる中間報告である。最終的な結論を出すにはさらに検討が必要である。

2. NRDF ファイルの検討

1993 年度入力データまでは既に変換が終了しているので、1994 年度入力データについて検討を行なった。NRDF データセットは D 番号で分類され、各 D 番号毎に EXFOR への変換プログラムを実行する。その結果、エラー・メッセージが無く変換終了する場合、また、エラー・メッセージが出てプログラムが終了する場合、さらに、変換するデータが無いまま終了する場合がある。勿論、ここで検討すべき場合は、2、3 つ目の場合である。以下に、検討すべきこととして取り上げた課題を述べる。

(a) Anti-neutron のコード： "ANTIN" を提案する。NRDF の粒子コードを見直し、EXFOR との対応を検討する。

(b) Incident energy の単位について：最近の高エネルギー実験データの入射粒子のエネルギー単位について、

MeV/c, MeV·A, MeV/A

のどれか（あるいは全て）を採用するか。どれでも良しとする場合でも、検索コードをどうするか決めなければならない。また、それぞれの単位の間の変換を用意しておかなければならない。

(c) Beam Intensity (個/sec) について：現在採用している単位は、nA, μ A であるが、荷電が中性な入射粒子のことを考慮すれば、個/sec がより基本的と考えられる。

(d) Table 中の物理量について：Table 中の heading に物理量とその単位が入る。どのような物理量についてのデータかと言う情報が、Table 中の heading だけでなく、NRDF の場合いくつかの場所で現われてくる。例えば、 DATA というセクションで物理量 PHQ が現われるが⁵⁾、そこでの物理量は必ずしも Table 中のデータをユ

ニークに指定しているとは限らない。一方、EXFOR では各 Table の物理量は必ず 1 つ指定しなければならない。そのため、EXFOR への変換に際して NRDF 側の物理量としては Table の heading の物理量を取ることになっている。しかし、NRDF の heading には、採取する現論文の Table, Figure で用いられている物理量が比較的忠実に用いられことが多く、それらの物理量が必ずしも対応するコードを持っているとは限らない場合が多い。その様なとき、Table の heading に "DATA" コードが用いられていることが多く、NRDF から EXFOR への変換を不可能にしている。この問題は、(i) "DATA" で指定される物理量の見直し、(ii) Table の物理量を指定する PHQ の位置付けの検討をすることによって改善を図ることとした。

(e) NRDF で用いられる入力データコード "X" について: NRDF では物理量の値が不明な場合、データコード X を用いる。しかし、このようなデータもまた EXFOR への変換において変換不能の結果をもたらす。そこで、データ・コード X を出来るだけ用いないようにし、用いる場合も "コメント" を付加してその内容を説明するようすることを提案することとした。

(f) Monitor reaction について: NRDF では対応する項目がなく、 $\Psi\Psi$ EXP セクションにおける Data analysis の項目で⁵⁾

Overall energy resolution

Calibration of detectors

Monitor reactions

Efficiency of detectors

等の入力フォーマットがあるが、それらが対応するデータ・テーブルと必ずしも結び付いたデータとなっていない。コーディング上の入力形式の改善すべき問題として今後の課題にした。

(g) Error analysis について: NRDF ではデータのエラーについて、 $\Psi\Psi$ EXP セクションの Error analysis の項目で⁵⁾

Total error

Systematic error

Statistical error

の入力フォーマットがあり、また同じセクションで Normalization の入力項目がある。Monitor reactions と同様、EXFOR への変換において、データ・テーブルとの関係が曖昧なため変換がスムーズに行なわれない場合がある。これも、コーディング時にデータ・テーブルと対応させてデータを入力することによって解決されると考えられる。

3. NRDF データ・コーディング

NRDF データの EXFOR への変換の変換率を上げるための議論を進めて行く中で、NRDF の入力形式やコーディング上の問題とも深く関わっていることが明らかになってきた。そこで議論され、検討された問題についてまとめておくこととする。それらは、1) 入力コード・文法エラー・チェック体制、2) 入力データの対象の問題に大別される。

現在、NRDF コーディング入力データのエラー・チェックは3段階で行なわれている。すなわち、1段階で NRDF 文法のチェック、2段階で NRDF コード・チェック、3段階でテーブル・ヘディング・チェックが行なわれている。これらのエラー・チェックはコーディング作業をしながらあるいは作業が終了した時、各段階に対応するチェック・プログラムを実行し、データが正しく入力されているかどうかをチェックし、エラーが見つければそれらを修正するようにしている。このエラー・チェック・プログラムの実行によって、入力データのコード・文法上のエラーをほぼ完全に無くすることができるようになった。さらに、コーディング・入力作業を端末のエディターを用いて、これらのデータ・チェック・プログラムを組み込んだコーディング・プログラムの作成も行なわれ、試作段階を終えている。このように入力エラーのほとんど無い入力データを効率的に作成できるようになった。

しかしながら、EXFOR への変換の問題は別なところにある。前の節で述べたように入力データが文法的にもコード的にも正しい場合でも採取・入力すべきデータが適当な位置に適当な形式で入力されているかどうかによって変換がスムーズに行なわれないことが生ずると言う問題である。これらの問題は上で紹介したチェック・プログラムでは解決できる問題ではない。それらの問題については専門家のプルーフが必要と思われる。本ワーキング・グループとしては適当な NRDF 入力データをサンプリングし、具体的にどのような項目についてプルーフが必要かを調べ、第3者（専門家）によるデータのプルーフを依頼することとした。すなわち、原子核物理学の専門知識を有し、NRDF の文法・コード等について十分な知識を持つ大学院生にコーディング済みデータをチェックすることをアルバイトとして依頼することを今年度入力データから始めた。その成果が EXFOR への変換にどのように現われるか今後の検討課題である。

次に、入力データの対象の問題に移る。この問題は、前節で述べたコード "DATA" や "X" 等に関する問題である。基本的にこれらのデータをできるだけ減らすことである。それが可能かどうか、また、そのためにどのようなことが必要か、ワーキング・グループで議論されたことを報告する。"DATA" や "X" のコードを用いるのは2つのケースがありうる。1つは、論文等から対応する物理量やキーワードが読み取れな

い場合であり、もう2つは、NRDFの辞書に対応するコードや項目がない場合である。後者の問題に対しては、新しいコードの作成・登録作業を充実することによって解決しうる。しかし、問題は前者の場合である。1つの解答として、そのような不確定さ (ambiguity) を有するデータはデータとして価値が低いので収集の対象からはずすというものである。確かに、そのようなデータはデータの検索において検索の対象にならず、データの利用に際してもデータに付随したコメント等を見なければその意味が判らないことやさらに原論文まで逆戻らなければならないということが生ずる可能性がある。

そこで、コーディング作業において出来だけ "DATA" や "X" のコードを使ったデータを減らすことを試みることを提案することとした。ただし、NRDF辞書に無いコードで、重要な物理量・キーワードについては速やかに登録作業を行なうことを同時に提案する。

4. 今後の予定

NRDFシステムは現在大きな曲がり角に来ている。その背景の1つは、情報媒体の急速な発展があり、またもう1つの背景は、NRDFデータの蓄積とその役割の拡大である。本ワーキング・グループは、はじめにも述べたように、当初 EXFOR への変換の問題を検討することが任務であったが、検討を進める中で NRDF 全体の方向と深く結び付いていることがはっきりして来た。その理由はまさに NRDF システムが大きな曲がり角に来ているゆえであり、これまでのデータ収集・入力、データの蓄積方法、データ検索システムについて基本的な考え方に遡って、再検討しなければならない時期に来ているように思われる。

そこで、本ワーキング・グループでの議論と提案を出発点として、さらに NRDF の改善すべき所や新しいシステムの作成などを検討して行くことが必要と思われる。本ワーキング・グループとしては一応本報告をもってその活動を終了し、管理運営委員会で具体的に問題を絞って新たなワーキング・グループ設置して引き続き検討されることをお願いしたい。

参考文献

- 1) 千葉正喜、NRDF Annual Report 92, No.6 (1993), 92.
- 2) 千葉正喜、NRDF Annual Report 87, No.1 (1988), 72.
- 3) 加藤幾芳、原子核研究 30(1995), 63.
- 4) 富樫雅文、田中一、「荷電粒子核反応データファイルユーティリティ開発報告書」(昭和60年3月) 日本荷電粒子核反応データグループ