

XML ベースの新フォーマットの提案

Suggestion of New Nuclear Reaction Database Format based on XML

北海道大学知識メディア・ラボラトリー

椿原 康介、松本 琢磨

北海道大学大学院理学研究院

合川 正幸、加藤 幾芳

TSUBAKIHARA Kohsuke, MATSUMOTO Takuma

Meme Media Laboratory, Hokkaido University

AIKAWA Masayuki, KATŌ Kiyoshi

Faculty of Science, Hokkaido University

Abstract

We are now constructing a new nuclear database format based on XML. since it is easy to extend database format and support some applications. Based on “XMLized Nuclear Reaction Data File(NRDF)”, XNRDF, we have almost finished developing how to compile the bibliographic information.

1 はじめに

荷電粒子核反応データファイル (Nuclear Reaction Data File:NRDF) は、日本国内で生産された荷電粒子入射実験を採録するべく、1980年代から整備されてきた。現在はその採録範囲を広げ、光核反応実験の数値データも蓄積されている。ただし、その先進的なアーキテクチャに比べ、現在のNRDFの採録形態は、近年の様々なコンピュータ環境の進展から見るとかなり古く、利用しにくいものになってきている。こうした風潮は国際原子力機関 (International Atomic Engineering Agency; IAEA) と、その協力関係にある核データセンター網が整備している EXFOR(EXTended FORmat) においても共通の問題とされている。前年度までに我々センターは、松本らを中心に、NRDFのデータファイルをXML(eXtensible Markup Language)形式に書き直すことを試みてきた。今年度は更にその作業を進め、採録経験者以外にも理解のしやすく、なおかつ拡張性の高い核データ採録フォーマットの構築を目指すべく、議論を続けてきた。本記事ではその今年度の成果について報告する。

2 新 XML フォーマットの概観

XML について詳しい解説は一昨年度の年次報告 [1] に譲り、本稿では要点だけを示す。XML は 1998 年に World Wide Web Consortium(W3C) により勧告された比較的新しい言語であり、様々な情報をその「意味・種類」と「実際の内容」という二つに分けてテキストにて記述するための比較的平易な方法である。つまり、マークアップ言語を組み立てる際に、従うべき法則と考えることができ、HTML(HTML5) は XML のサブセットの一つとみなせる。その法則に基づき、採録すべき情報にラベル付けを行い、計算機上での処理に適した形を構成することが出来る。その点で、現行の NRDF を XML 形式に単純に置き換えることは、タグを付与する法則を完備すれば、一義的には難しいものではない。但し、例えば将来的に、核データ採録用のエディタを開発する際など、核データ開発に関するソフトウェア開発を考えた時、必ずしも単純な置き換えたものが効果的ではないことが予想される。また、NRDF、EXFOR 共々、各種実験施設や物理量、その他の実験情報について冗長化を避けるためのコード化を行っている。これは計算機の環境が現在ほど潤沢ではない時代の名残りと言えるが、採録者や、その採録された結果を確認する実験者、また核反応データを検索し実際に利用する側からすれば、簡単に理解することが出来るとは言い難いものになりつつある。さらに、理化学研究所のもつ RIBF など、今までの実験とは異なり不安定核の領域に大きく足を踏み入れる実験を行うことが出来る施設が増えてきたため、今後は現状の NRDF、EXFOR 形式での採録が難しい核データが増えてくるおそれもある。それに対して、なるべく拡張性の広いフォーマットを整備し、新しいデータに対応しうる余力を残してあるようなものをベースに採録活動を行っていくべきであると考えられる。

これらの状況を鑑みて、我々は今年度、NRDF、EXFOR を基盤とした新たなデータベース形式開発を目指し、議論および開発活動を行ってきた。特に NRDF は先に述べたように、核データの採録という特定の目的ではあったが、フォーマットとして XML に似た汎用性を本来持っているため、そのコンセプトを受け継ぐような形式を確立する事を目的としている。

3 実際の作業について

新 XML フォーマットの策定を書誌情報に関する部分から始めた。何故なら、書誌情報に関しては NRDF、EXFOR のどちらにおいても、ほぼ同等の採録形式を持っていて、比較的議論する点が少ないと考えられたからである。また、全体的な枠組みを定めなければ、各データ構造の細部を詰める事は出来ない。その上でも、新フォーマットのテストとして、独立的に議論しうる書誌情報から始める、という狙いもあった。

作業としては、策定に関するワーキンググループを設立し、各人が既に採録されたエントリーを XML 化してみるという形から開始した。その際には、椿原と過去に NRDF の XML 化の経験があった松本を中心に作業が行われた。

実際に策定された XML フォーマットを図 1 のように示す。今回の策定に関して

- 各タグ名やその要素について冗長性を持たせ、どの内容についてのタグなのか、またどんな内容に対応しているのかが後の開発者や利用者にわかりやすいものとする
- 先に述べた不安定核などの新しい形式のデータが生じた際に、すぐさま対応できるような拡張性を持ったファイル形式とする
- 複数の子タグを持ちうるタグは複数形で表し、その子タグは単数形で扱う

といった点に留意した。この内、後者については、ある程度 XML 化することによって達成出来ると言えるが、データ構造としてもその点について留意することが重要である。つまりどこを拡張するべきか明確にしておくことが必要になる。その観点から、各種核反応データに関する情報を一つのエントリー中で、複数の場所に散見される形式から、なるべく一括に定義する形式に改めることを議論した。つまり実験施設なら実験施設の項にまとめて、データに関する項でそれを参照するという形式を採用しようということである。この発想は過去の NRDF にもあったもので、その意味で今回の新フォーマットにおいても NRDF の思想というものはある程度であるが引き継いでいると言える。

今報告の時点では、参照されるべきタグには付与情報として“no”を与えている。このタグ内で定義される通し番号を用いて、他の部分で参照する形式を現在採用している。例えば図 1 中央では実験の構成について <experimental_setups> とその子タグである <experimental_setup> に置いて定義している。ここでは施設、標的、識別粒子、検出器について定義しているが、要素を番号で参照するに止めて具体的な内容は定義しておらず、各々の情報は別に施設なら施設でまとめて定義されている。

また、実験の解析、施設の概要や標的の情報等を実験のハードウェア、ソフトウェアに関する観点から分類し直し、データの誤差情報についても、その由来をこの二つの分類で分け直すことを議論している。

今年度の議論では、各データと定義された量を結びつける方法に関して完全な合意に至らず、次年度へと課題を残したが、この点に関して合意に至ればラフな形ではあるが一つのフォーマットが出来たといえる。また、今回議論しているフォーマットが、真に XML でのデータベース構成を満たしているかどうか、例えばあるタグで定義された属性値を他のタグで参照するようになっているが、XML でこれを本当に許しているかどうかなど、調査が必要な点が指摘されている。この点についても、次年度以降の課題といえる。

以上の新 XML フォーマットについて、2011 年 9 月の第二回 AASPP ワークショップ(北京にて開催)や 12 月の理研 RIBF ミニワークショップにて講演し、他のセンターの研究者と議論を行い、拡張性のある新 XML コードの必要性について示し、一定の支持が得られた。

4 終わりに

本稿では NRDF の XML 化を受けて、新 XML フォーマットを開発する事を考え、今年度におけるその策定作業について報告した。

具体的には今年度は全体の構造、及び書誌情報の採録方法について議論を行い、大枠ではワーキンググループ参加者間の合意は取れた。今後はより具体的に各核反応データの採録方法やその形式について、更にその開発を進めると共に、具体的な利用を考えていく必要がある。この議論を進める上で、従来の NRDF において意識されてきた事などについて、詳細な理解を進める必要がある。今後その点について研究・議論を進め、新フォーマットの策定に活かしていきたい。

謝辞

本成果の一部は北海道大学大学院理学研究院と独立行政法人理化学研究所の共同研究「RIBF 核反応データの高度利用研究」の助成によるものです。

```

▼<entry no="3">
  ▼<database_info>
    <nrdp_number>D0003</nrdp_number>
    <exfor_number>E0003</exfor_number>
    ▶<history>...</history>
  </database_info>
  ▼<references>
    ▼<reference no="1">
      <reference_type>Main reference</reference_type>
      ▼<journal_info>
        ▶<journal_title>...</journal_title>
        <journal_name>JPJ</journal_name>
        <journal_issue/>
        <journal_volume>27</journal_volume>
        <journal_page>278</journal_page>
        <journal_date>1969</journal_date>
        ▶<purpose>...</purpose>
      </journal_info>
      ▼<affiliations>
        <affiliation no="1">2JPNIPC</affiliation>
        <affiliation no="2">2JPNIT</affiliation>
      </affiliations>
      ▼<authors>
        ▼<author no="1">
          <author_name>T.Fujisawa</author_name>
          <author_affiliation>1</author_affiliation>
        </author>
        ▶<author no="2">...</author>
        ▶<author no="3">...</author>
        ▶<author no="4">...</author>
      </authors>
    </reference>
    <reference no="2">none</reference>
  </references>
  ▼<reactions>
    ▼<reaction no="1">
      <reaction_target>58Ni</reaction_target>
      <reaction_incidentpart>3He</reaction_incidentpart>
      <reaction_emittedpart no="1">INL</reaction_emittedpart>
      <reaction_emittedpart no="2">none</reaction_emittedpart>
      <reaction_residual>58Ni</reaction_residual>
      <reaction_relation>none</reaction_relation>
      <reaction_incidenten>24.15MeV</reaction_incidenten>
    </reaction>
    ▶<reaction no="2">...</reaction>
    ▶<reaction no="3">...</reaction>
  </reactions>
  ▼<experimental_techniques>
    ▼<experimental_setups>
      ▼<experimental_setup no="1">
        <setup_facility>1</setup_facility>
        <setup_sample>1</setup_sample>
        <setup_detectedp>1,2</setup_detectedp>
        <setup_detectors>1,2</setup_detectors>
      </experimental_setup>
    </experimental_setups>
    ▼<hardware>
      ▼<facilities>
        ▶<facility no="1">...</facility>
        <facility no="2">none</facility>
      </facilities>
      ▼<samples>
        ▶<sample no="1">...</sample>
        <sample no="2">X</sample>
      </samples>
      ▶<detectors>...</detectors>
    </hardware>
    ▼<software>
      ▶<experimental_methods>...</experimental_methods>
      ▶<theoretical_methods>...</theoretical_methods>
    </software>
    ▼<errors>
      ▶<hardware_errors>...</hardware_errors>
      ▼<software_errors>
        ▶<systematic_errors>...</systematic_errors>
        <statistical_error>none</statistical_error>
        <total_error>none</total_error>
      </software_errors>
    </errors>
  </experimental_techniques>

```

図 1: 構築中の新 XML フォーマットで採録された NRDF/EXFOR エントリー (D/E0003)

参考文献

- [1] 松本琢磨、千葉正樹、加藤幾芳、「NRDF の XML 形式への変換」, 荷電粒子反応データファイル年次報告 No.23, 2 (2010); 及びその中で紹介された参考文献