

著者からの数値データ提供の取扱いについて

Numerical Experimental Data Sent by Authors

北海道大学知識メディアラボトリー
合川正幸、大塚直彦

Meme Media Laboratory, Hokkaido University
AIKAWA Masayuki, OTUKA Naohiko

Abstract

Numerical experimental data have been attempted to obtain directly from authors to improve quality of NRDF and EXFOR compilation. Many authors kindly sent their numerical data by our request. In addition, some authors gave us other information such as experimental conditions which are also useful to compile. In this paper, we report our attempt to make communications with authors of experimental papers. We also summarize what we have to notice when we request and process the data for our compilation.

1 著者提供データ採録の重要性

NRDF のデータは大きく分けて書誌情報 (BIB)・実験情報 (EXP)・数値情報 (DATA) からなる。このうち数値情報、とりわけ数値データ本体はデータベースの中でも需要の高い部分であり、質の良い数値を利用者に提供することは重要である。論文を入手して書誌情報や実験情報が既知である場合にも数値が図の形でしか掲載されていない場合は多い。例えば理論計算を行なって実験値との比較を行ないたい研究者にとっては論文の図の数値がデータベースから直接入手できればありがたいことである。もちろん数値情報を得たい研究者は直接実験をされた方々に数値送付を依頼することも可能である。しかし実験者がこのような依頼に個別に対応していくのは手間を要することである。

従来より、日本荷電粒子核データ反応グループ (JCPRG) は論文の図の数値データを多数収録してきた。その数値データの多くは出版された論文中の図をスキャンして画像データとし、その画像から数値を読み込むという方法で得られてきた。近江氏らの SyGRD の開発などに見られるように我々はこの数値化で得られる数値の質の向上にも注意を向けてきた。昨年6月のパリでの IAEA 核データセンター長会議で大林氏が SyGRD について発表を行なったところ、翌月にロシアで荷電粒子核反応データを採録する CAJaD からメールがあった。それは実験値が大変に密にプロットされた図 (Physical Review C 62 (2000) 014610-7 Fig.7) の読み取りの依頼であった。これを芦沢氏が SyGRD で読み取り CAJaD に送付したところ驚きの反響を得た (と同時に実験家から得た「正解」を送付してくれた)。このように我々の数値読み取りの技術水準は高いと考えられるが、それでもなお読み取りの精度には限界があることは否めない。例えば測定量に対してその誤差の値が相対的に小さな場合には、誤差値は読み取り分解能の大きさによって離散化されて

出力されることがある。この場合は我々は意味のない桁を落すなどの配慮を必要とする。

以上のような状況を考えるに、JCPRG が著者から数値データの提供を受けて採録結果に取り込み、データが得られた条件などとともに責任を持って配付することは、採録者・利用者・実験者のいずれにも利益があると見える。

荷電粒子核反応の採録を我々と協力して行なっている海外のセンターでの状況はどうであろうか。前述のロシアの CAJaD は相当部の数値を我々と同様にデジタイザーで採録している。一方、アメリカの NNDC の採録データは著者から直接入手した数値データを多く含む。NNDC は更に Physical Review C 誌に掲載された数値データを著者から一括入手する方法について検討している。

NRDF にも著者から入手したデータを採用できた採録がこれまでもないわけではない。2000 年度の採録をみると D1705 (著者から採録者の青木氏が入手)、D1718(API の ftp サイトから入手) の例がある。また、阪大 RCNP の研究計画委員会の構成員であった大西委員は、RCNP で実験を行なったグループによる数値データの JCPRG への提供を依頼した。(以降、RCNP の研究計画提案の募集要項には JCPRG への数値提供協力の件が書かれるようになった)。このように正確な数値データを入手する環境は徐々に整いつつあった。

数値データを入手することは数値化の作業を効率化することにもつながる。昨年度採録した D1750 は数百 MeV 領域の重イオン入射で得られた中性子の二重断面積を様々な角度とエネルギーで系統的に行なったもので核データとして価値の高いものであったが、そのデータ量は膨大なものでありスキャンによる数値化には莫大な時間を要することが予想された。そこで、この論文に関する幾つかの疑問を著者に問い合わせた際に数値データの入手の可否について問い合わせしてみた。その結果、著者は採録論文中の実験数値を下さるばかりか、同グループによる他の数編の論文で発表されたデータの数値も送って下さった。その数値を含む論文は急拠 D1756 と D1759 として採録された。また D1748 はクーロン分解反応を用いてその逆過程の天文学的 S 因子を求めたやはり核データとして価値の高いデータであった。この論文についても採録者の伊藤氏を通じて著者から数値を頂くことができ、また幾つかの物理量の定義について確認をとって正確を期すことができた。

このように著者と交流を持つことは、数値データの入手はもとより採録内容を正確にする上でも非常に意味のあるという認識が高まってきた。そこで今年度は採録品質の向上の一環として、出来るだけ多くの論文について著者に採録上の疑問点と数値入手の可否を問い合わせることで実験者との連絡をとることを試みた。本稿ではその経過について報告する。

2 2002 年度の取り組み、経緯

今回の試みの契機となった D1750 の著者への問い合わせは、そもそも数値データの入手よりも実験条件に関する質問が主たる内容であった。以下はその全文である (一部編集)：

To: ***** <*****@*****.***.*****.***.jp>

Subject: Phys.Rev.C 64(2001)034607

Date: Wed, 23 Jan 2002 22:55:59 +0900

○○○○様

突然のメールで失礼します。私は北海道大学理学部原子核理論研究室の大塚と申します。日本の荷電粒子核反応データを採録してデータベースに登録し、加えて採録結果を国際原子力機関 (IAEA) に送付する活動を行なっております。

今回、〇〇さんが著者に入っておられる Phys.Rev.C 64(2001) 034607 の採録作業を行うにあたり疑問点とお願いがありますのでご連絡させて頂いた次第です。

この論文では標的である C,Al,Cu,Pb、入射核である He,C,Ne の質量数が明示されていません。こちらの採録作業でこれらの核種に関して特定の質量数のものを使っているのか、自然同位体比のものを用いているかの情報が必要です。これについてご教示頂けるでしょうか。

我々のグループでは論文中に数値テーブルがない場合にはデジタイザを用いてグラフから数値を読み取っていますが、データベースの信頼性の点からは著者の方から頂いた数値データを直接登録できれば好ましいです。そこで、可能であれば、Fig.3-Fig.8 と Fig.13-Fig.14 の数値データを頂ければ大変助かります。

以上、お手数をかけますが、よろしくお願ひします。
それでは、失礼します。

このメールに対して論文の著者は数日後に返信を下さった。その返信の中で著者は論文3編の数値が掲載された（主に共同実験者向けであろう）ウェブページの URL を教えて下さった。この中の2編は採録対象論文でありながら、採録予定リストに取り上げられていなかった。それは、掲載雑誌が”Nuclear Science and Engineering”と”Nuclear Science and Technology”という、ともに採録対象外雑誌であったからである。後者の雑誌は研究室で入手できたので早速採録した (D1759)。残り1編の Phys.Rev. の数値は先方の手違いでこちらが依頼した D1750 と別の論文のものであったが、この論文も採録対象であったため、結果的に D1756 として採録させて頂くことにした (D1750 の数値は後日改めて送付頂いた)。

以上のような経過から、2002 年度の採録を開始するにあたっては、”Nuclear Science and Technology”の過去数年分の論文に採録対象論文がないかどうかをまず確認した。その結果として我々は6編の対象論文を発見した。いずれも九大のグループによるもので、それぞれ中性子の二重微分断面積 (DDX) を系統的に調べた論文であり、核データとして価値の高いものであった。そこでこれらを D1760-1765 として採録し、数値データ以外の情報の入力を終了した時点で著者に問い合わせを行なった。その結果、それぞれの実験値を著者から送付して頂くことができた。D1761 はデジタイザでの数値化が不可能な3次元グラフの DDX データを多数含んでいたが、数値が入手できたことでこれらのデータも採録することができた。ただこの論文は幾分古い論文であったために、一部の図に関しては著者も数値データを発見することができなかった。これ以外の論文5編に関しては全ての採録対象のグラフの数値を入手することができた（うち4編については著者が所属する研究室のホームページで公開されていた）。

実験値を著者から得る試みが順調に行なわれたことを受けて、2002 年度は引き続き EXFOR での採録対象となる論文全てに関して著者への数値提供依頼を行なうこととした。基本的には採録と点検が終了した時点で採録上の疑問点の問い合わせとともに数値の提供依頼を行なった。依頼に際してはどの図のどの点が欲しいかを明記し、依頼先にこちらが必要としているものが明瞭に伝わるように心がけた。

このような新しい数値収集の展開は採録対象の選定にも影響を及ぼす。典型的な例はヒストグラムの採録である。非常に密なヒストグラムは従来数値化困難という理由から採録されていなかったが、著者に入手依頼することで従来の除外理由がなくなった。そこで従来採録対象から外されていたヒストグラムについても、個別に採録対象かどうかを採録者が判断する必要が生じることとなった。これらについては現在も

個々に判断することになっているが、基本的には測定器のエネルギー分解能（関心のあるレベルの分離精度など）など物理そのものの議論の対象でないものが殆どであり、結果的には採録されたヒストグラムは今までのところ皆無である。

一方、絶対値でなく相対値 (Arbitrary Unit) が与えられた図であっても、測定値が物理の議論の対象となっている場合は採録した (EXFOR では REACTION の SF8 に REL と記載する)。また、カウントで示された物理量についても同様の判断から対象とした場合があった。この場合は採録結果に生データ (Raw Data) である旨を明記した (EXFOR の場合、SF8 に RAW と記載)。

3 実際の著者提供データ採録と注意点

今年度の著者への問い合わせとその後の対応について報告する。2003年1月現在、NRDFとEXFORの採録対象となっている論文中、採録・点検作業が数値データが請求出来る段階に達した論文は

D1766, D1768, D1769, D1770, D1772, D1773, D1775, D1776, D1777, D1779, D1780, D1781, D1782, D1783, D1784, D1785

の16編であった（これらの論文の書誌情報については本年次報告の「資料：2002年入力データ」を参照のこと）。このうち、D1768, D1772, D1782, D1783, D1777の5編は数値を全て論文中に記載していたので、実際に請求の必要が生じたのは11編であり、これらに関して原則第一著者へ問い合わせを行なった。このうちD1784は理論家・実験家の合作論文であり、連絡を取った第一著者が理論家であったため、その後の連絡は紹介頂いた実験家（第二著者）と連絡を取った。

結局11編のうちD1770とD1779を除く著者から返信があった。D1775に関しては後日数値データを頂けるといふ返事であり、その他の論文に関しては早速数値データを頂くことができた（本稿執筆中の2003年3月10日にD1775の数値データが届いた）。同時に著者の方には採録の過程で生じた疑問に答えて頂いた。

数値の受け取り方法として代表的な形式はテキストファイルとエクセルファイルの2通りであり、送付方法としてはメールを介する場合とウェブページを介しての場合があった。エクセルファイルの場合にはテキストデータに出力してから加工を行なった。

何列にも渡る提供データを採録エディタ HENDEL に読み込むためには、必要な列のみを抽出する必要がある。エクセルから出力された場合のように列がタブで区切られている場合などは、表示のされ方がエディタなどに依存し、ファイルを一見しただけでは数値がどの列に属するか分からない場合がある。処理の過程で間違いを作らないためにも、数値の切り分けは perl などで簡単なスクリプトを書いて行なうのが安全である。

しかし数値ファイル加工の過程で人為的なミスをする可能性はやはりある。そこで、HENDEL に読み込ませた数値が著者から送付を受けたものと同一かどうかを確認する必要がある。まず HENDEL のデータ入力欄なり採録結果出力の数値の部分参照し、少なくとも独立変数の上端下端において値が正しく入力されている確認する。次に図の照合を行なう。具体的には HENDEL で表示された図と論文のグラフを照合し、点の位置や個数など目視の範囲で異常のないことを確認する。論文の図に比べて受け取ったデータの点が多い場合、論文に掲載されていない数値を受け取った場合など、その数値を掲載して良いかどうか論文の著者や運営委員と相談するべきである。そのような数値を採録することになった場合は、該当データの部分にコメントをする必要がある。

何らかの理由で数値ファイルに欠損値などの空欄がある場合は、その箇所に X, +-X, UNKNOWN, +-UNKNOWN などを記入するなど適切な処理を行なわなければならない。これを怠ると HENDEL に読み込ませて出力した結果に「列数が十分でない」などのエラーが表示される。欠損値に関して論文に理由等の記述がある場合などは該当データの部分にその内容を記入する。

頂いた数値の桁数が異常に多い場合にも注意を要する。著者が論文に図を掲載するにあたって用いた数値は有効数値を意識したものは限らないからである。

また採録結果には、著者から送付された数値を採用している旨を明記しておかなければならない。例えば HENDEL 上では以下のように行なう。

- HISTORY 欄

”Recieved”を選びデータを受け取った日付を記入する。

- STATUS 欄

(1)”Data recieved by center in tabular form”を選ぶ。

(2)”Data (Fig.3 left-top, p034607-4 in reference) received from H.Iwase by e-mail.”あるいは”Data (Fig.6-(a), p534 in reference) are received from author’s ftp site” というような記入をする。

4 まとめ

以上に述べたように、今回こちらから数値データ送付を依頼した殆どの著者は、ご多忙にも関わらず快く数値データの送付に応じ、また同時に採録上生じた疑問にも答えて下さった。今後も引続きデータ送付依頼を継続していくべきであると考ええる。実験家と相談し双方が納得できる採録を行なった上で、データを公開していくことは、採録する側・される側・利用する側のいずれにとっても有益であろう。

近年は大強度陽子加速器の遮蔽計算等の必要性などから、工学系のグループが荷電粒子入射反応実験を行なうケースが増加している。このような実験のデータは NRDF なり EXFOR なりを通じて活用される可能性が大きい。この種のデータは掲載論文が”Nuclear Science and Technology”など我々が採録対象としていなかった雑誌に掲載される場合が多かったこともあり、従来は採録対象から逃していた。今後はこのようなグループとの連絡を密にしていく必要がある。主要な加速器を有する研究所・大学に所属していてそこで行なわれている実験に通じた研究者を助言委員として迎えることも、採録・問い合わせをスムーズに進める上で有益であると考ええる。

採録論文の著者への問い合わせは出版後早い時期に行なうべきである。時間の経過に伴い、著者の関心が移ることが考えられるし、数値ファイルの在処を探すのに余分な時間を使わせてしまうためである。著者からの数値データや意見を反映して質の高い採録を行なう上で、出版後速やかに採録を実施することは大切である。そのようにして速やかに採録された NRDF や EXFOR のファイルを通じて、その論文に関心のある研究者が採録データを発見・使用する頻度が増加する、と考えられる。

我々核データ収集活動をする立場にある者でさえ実験値の入手問い合わせを実現させるのに時間がかかっており、一般の研究者にとっては実験数値の入手の問い合わせはより敷居の高いものであると思われる。我々が責任を持って著者の数値を採録し公開していく活動を続けることで、NRDF がより多くの人に必要とされるものとなることを固く信じている。

5 謝辞

最後にご多忙にも関わらず今回の採録で数値を送付下さりまた論文の内容についてご教示下さった以下の方々にお礼を申し上げます。

D1748:

甲南大学理工学部物理学科原子核研究室 宇都宮弘章氏

D1750, D1756, D1759:

東北大学工学研究科量子エネルギー工学専攻中性子デバイス工学研究室 岩瀬広氏

D1756, D1759:

産業技術総合研究所計測標準研究部門量子放射科放射線標準研究室 黒澤忠弘氏

D1760:

高エネルギー加速器研究機構低温工学センター 中本建志氏

D1761:

日本原子力研究所大強度陽子加速器施設開発センター中性子施設開発グループ 原田正英氏

D1762, D1763, D1764:

九州大学大学院工学研究院エネルギー量子工学部門量子線物理計測研究室 石橋健二氏

D1765:

九州大学大学院工学府エネルギー量子工学専攻量子線物理計測研究室 岩元洋介氏

D1766:

甲南大学理工学部物理学科原子核研究室 山崎かおる氏

D1769:

放射線医学総合研究所重粒子医科学センター加速器物理工学部 岩田佳之氏

D1773:

大阪大学核物理研究センター核理学研究部 若狭智嗣氏

D1775:

日本原子力研究所先端基礎研究センター変形核重元素合成研究グループ 光岡真一氏

D1776:

東京大学大学院理学研究科附属原子核科学研究センター 川畑貴裕氏

D1780:

新潟大学理学部物理学科原子核実験研究室 鈴木健氏

D1781:

東京工業大学理工学研究科基礎物理学専攻中村研究室 佐藤義輝氏

D1784:

東京大学大学院理学系研究科附属原子核科学研究センター 上坂友洋氏

D1785:

東京大学大学院理学系研究科附属原子核科学研究センター 野谷将広氏

この他数編の論文の著者の方に、数値提供の依頼は行ないませんでしたが、採録上の疑問についてご教示頂きました。合わせて感謝致します。