

NMRスペクトルデータベースを作成した経験から 化学分野の特定目的のデータベースについて

早水 紀久子

株式会社NMRDBTech社
（エヌエムアイールデータテック社）
取締役社長 理学博士

現在、科学技術の研究では既に蓄積されている大量データへのアクセスは重要な課題であり、同時にその信頼性や価値を見極めて役立てることは研究の成否を決めることもある。特に、化学の分野では多様性のある化合物や物質が研究対象になるために、情報へのアクセスはインターネットの利用が日常的に行われる現在において特に重要である。ビッグデータの利用が日常的に語られる現在においても、信頼性の高いデータを見つけたら、僥倖ともいえよう。いくら探してもデータが見つからない場合に、自らデータ作成するか適当に推測することになる。一方で文献検索に始まって、膨大な化学情報の集積と利用は日常的に行われているが、求めるデータが見つからないことが少なくない。

人間力、時間、費用をかけて取得したデータが広く公共の知的財産となり、長い寿命を持つためには、対象物質の普遍性、公共性、信頼性とデータ精度が深くかかわると考えられる。データを作成するのは一般に“分析”と言われる仕事である。日々進歩する分析手法の理解とスキルが要求されるが、縁の下の力持ち的な存在になっていることが多い。分析は地道な研究分野であり、華々しい製品開発の基礎をなしているが、それが前面にでて製品開発の要と言われることは極めて稀である。しかしながら、実に多くの物質に対して似たような分析が繰り返し行われている。化合物（物質）の確かなデータを、組織的に整理してデータベース（DB）化することの要望はありながら、地道な作業であるために、ルーチンワークで創造性がなく誰でもできるレベルの仕事として評価は低い。

1970年代から当時の通産省工業技術院東京工業試験所と日本化学会のNMRデータ小委員会で始めたNMRデータ精度評価とデータ集積の試みのお手伝いの立場からデータ活動に参加した。1979年のつくば移転後、工業技術院の汎用大型計算機にスペクトルデータベース（SDBS）を構築するというプロジェクト（故山本 修博士がシステム構築とデータ集積を主導）が開始された。インターネットで公開するという計画が持ち上がったのは1996年であり、それまでに10年以上にわたる地道なDB活動があった。1996年にインターネットにSDBSの公開システム作成を開始し、2001年3月の定年退職までに、SDBSのインターネット公開とPCによる新規のデータ収録のためのDB構築システムを完成して、工業技術院（現産業技術総合研究所（産総研））から離れた。現在、産総研のSDBSはデータ数を増やしているがシステムの大きな変更はないようである。

2001年3月に定年退職した時に、汎用大型計算機からテキスト-フォーマットで取り出してあったSDBS-NMRのデータをPC-データに変換する仕事が残った。著作権の取得（フォーマット変換に必要!）とベンチャー（NMRDBTech社）を始めた。帰属の付いた化学構造式（大型計算機）をモルファイル（ISIS/BASE）へ変換するというのは単純のようで面倒な仕事であった。PC版SDBS-NMRの著作権は産総研とNMRDBTech社の半々が所有し、下記のKnowItAllへ利用販売ができた。その頃、天然物化合物のNMR-DBの作成の要望があり、PCのDB作成システムは容易に完成したので、深く考えることなく走り出した（ISIS/BASEとMicrosoft Excel、CH-NMR-NPの構築）。

主要な研究論文誌を中心に2000年の文献から2014年3月までの文献に記載してある天然物のNMRデータから約30,000件のデータを集めて一応終止符を打った。収録化合物数は万全ではないが、2000年代に入ってNMR装置の向上に伴ってスペクトルの質は著しく向上している。有料のサービスはユーザ獲得に難があるので、無料でJEOL-Resonanceのホームページから公開サービスしている(2014年12月から<http://www.j-resonance.com/corporate/nmrdb/>)。好評のようである。

1980年代には複数の機関でNMRスペクトルDBを構築していたが、現在、形が残っているものは少ない。大型コンピュータからPCへのデータやシステムの変換が進まない、適正な規模までデータ収録がなされていない、担当者や所属機関の情熱が継続しなかったなど複数の理由があろう。スペクトルDBの草分けは米国SadlerのIRスペクトルで1960年代に図書室の壁一面に冊子が配置されていた。スペクトルデータの重要性にいち早く着目した収録活動であった。現在はBioRad社のKnowItAllとなり、NMR版のKnowItAllもある。この中にSDBS, SpecInfo (Wiley) (^{13}C), RobienのCSEARCHのデータ(^{13}C)も収録され、有料販売されている。SpecInfoはドイツで膨大なデータ集積され、Wileyから有料公開され、 ^1H と ^{13}C のみならず ^{19}F , ^{31}P , ^{29}Si のデータを含む。ウーン大学のRobien教授が ^{13}C シフトの予測システム(CSEARCH)を開発し、 ^{13}C シフトDBを保有する。ヨーロッパを中心に ^{13}C と ^1H シフト予測システムとしてmodgraph社(有料)から提供されている。日本で開発したPolyInfo(無料)にポリマーのNMRが含まれている。本稿を書くために調べて分かったロシアのCheD (Chemical Data)には ^1H , ^{13}C のDBが含まれている(<http://ched.ipac.ac.ru/>)。DBは長年にわたる集積活動が必要である一方、技術の進歩に伴う陳腐化が生じるのは科学DBの宿命であろう。

多様性の学問である化学の分野で、個人、企業あるいは研究機関で化学情報の特定目的のDBの作成中、あるいは作りたいと考える方々は多いと思う。私の経験からDBに収録する個々のデータは正確で価値のあるものに限定すべきである。数が多ければよいというものでもない。捨てるには忍びないので、とりあえず収録するとして作成したDBの信頼性と価値は低い。もし、データが万全でない時には、どこまで確実であるか、残っている疑問点は何かを明記する必要がある。時間がたつと問題点が分からなくなり、データを安心して利用できない。保存する価値のある個々のデータを評価・判断できるのは研究経験者である。DB構築は若者むきの仕事ではない。DB収録の仕事は研究論文にも特許にもならないであろう。だからといって、誰にでもできる活動とはいえない。新しい技術が導入され、データの種類や確度も向上する。その現実に即するようにDBの内容を柔軟に拡充・充実し、システムを改変しないと時代の要求に合わず陳腐化はまぬがれない。研究を推進できる人物がDB構築と集積活動に関与すべきであると、私は考える。



略歴

- 1963年 御茶ノ水女子大学理学部化学科卒業
- 1963年 工業技術院東京工業試験所入所
- 2001年 物質工学工業技術研究所*¹⁾を定年退職
*現在の産業技術総合研究所
- 現在 NMRDBTech社取締役社長、筑波大学数理物質科学研究科研究員

受賞歴

- 2000年 通商産業大臣賞「スペクトルデータベースシステム(SDBS)の構築と利用高度化に関する研究—インターネット公開—」

最近の研究領域

1995年—現在
「リチウム電池用電解質のNMRによる拡散測定の研究」*²⁾に従事、その成果をインターネットで公開 <http://diffusion-nmr.jp/>

*ResearchGateの集計によれば発表論文数はおよそ110件、被論文引用の総数は5,500件を超えた。NMR屋に興味深いサンプルを持ち込んで下さった多くの共同研究者達のおかげである。