

社説

Information overload

あふれ出る科学データに、いかに取り組むか

Nature Vol. 460(551)/30 July 2009

米国の学術連合組織ナショナルアカデミーが発表した報告書で、ペタ (=1000 兆) バイト時代を迎えた大量科学データ問題への提言がなされた。

2003年、30億塩基対のヒトゲノム完全解読結果が発表されたが、遺伝学者はそのために10年以上の歳月を費やした。しかし、最新の高速シーケンサーを使えば、今では同じ解読が1週間でできる。今日、高速シーケンサーはフル回転状態にあり、全塩基配列が解読された生物種は増え続け、同じ生物種について複数の配列結果も得られている。一方、「スローン・デジタル・スカイ・サーベイ (SDSS)」というプロジェクトに参加した天文学者は、米国ニューメキシコ州に設置された望遠鏡を使って、2000年から2億個以上の天体に関するデータを集め、全天の約25%をカバーした宇宙地図を作成した。ところが2015年にパチョン山 (チリ) 頂上に完成予定の大型シノプティック・サーベイ望遠鏡なら、それだけの大量データを一晩で集められるようになる。

同じようなことが多くの研究分野で見られる。データ不足よりはデータ過多の方がよいに決まっており、研究にとっては喜ばしい状況といえる。しかし懸念材料は、データの扱い方に関する方針ないしは政策や慣行ができあがるまでに、既に研究者が大量のデータを収集してしまうことだ。とりわけ論文誌の編集者は、画像の改ざん、元データの保全、大規模データセットへの継続的アクセスの確保、アルゴリズムとコード共有方式の標準化といった問題と、真剣に取り組まねばならないことがわかった。

こうした懸念から、2006年、科学系の多数の学会と *Nature* を含む研究論文誌数誌は、米国科学アカデミー (NAS) に問題の調査を依頼した。その結果、ナショナルアカデミー (The National Academies) に調査委員会が発足し、ネイチャー・パブリッシング・グループもスポンサーに名を連ねた。マサチューセッツ工科大学 (米国ケンブリッジ) のがん研究者 Phillip Sharp と物理学者 Daniel Kleppner が委員長を務め、7月22日に報告書が発表された (<http://tinyurl.com/datasteward> 参照)。

この報告書では、誠実性の原則、利用の原則、管理の原則という3つの主要原則を中心に据えて、11項目の提言がなされている。データの誠実性の原則では、研究者が作成するデータが真実であり、また正確であることを確保するために、研究者自らが最終的な責任を負う。研究者ひとりひとりとは、それぞれの分野におけるプロの基準に従わねばならず、関係機関は、それが実行できるよう、研究者に研修・訓練を確実に実施すべきである。

データ利用 (データアクセス) の原則では、公開の価値が強調されている。他の研究者による正確さのチェック、解析結果の検証、過去の研究成果に基づく研究の前進など、いずれも研究データの相互利用ができて初めて可能となる。したがって、研究者がデータ公開を差し控える十分な理由がないかぎり、それを公開して、他の研究者の意見を受け付けることが望ましい。できれば早めに、助成研究計画書の作成段階から、研究成果をタイムリーに一般公開するための準備に入るべきである。

データ管理の原則は、長期的な保存の必要性に対応する。科学系学会と科学コミュニティーは、どれが将来の解析のために保有しつづける価値のあるデータなのか、判定のためのガイドラインを定める必要がある。研究機関と資金提供機関は、こうしたニーズに取り組み、支援を行う必要がある。また、研究論文誌は、発表された記録の保存と上記ガイドラインの普及と実施に関して、一翼を担うことができる。さらに、責任あるデータ管理においては、データの専門家が決定的に重要であることを認識しなければならない。これまで、研究者がデータの専門家を支援するケースは少なかったが、これは改めるべきで、データの専門家をもっと尊敬されるべきである。

この報告書は、デジタルデータをめぐる複雑な問題について概観したもので、問題の解決法は提示していない。この点は執筆者も認めている。いま、研究機関、研究コンソー

シム、科学系学会がしなければならないのは、自らが置かれている分野と制約環境において、個別に有効な解決法を見つけ出すことである。資金提供機関はその責任を果たし、個々の助成金を手始めに、データの保全に対する投資

を増やさなければならない。科学活動において必要なのは、データの誠実性を通して一般市民と信頼のぎずなを結ぶことだ。今こそ行動を起こし、そのぎずなを強める時である。
(菊川要 訳) ■