【講座】

科学技術におけるデータベースの役割(1)
Role of Databases for Science and Technology (1)

馬場 哲也＊
Tetsuya Baba

１．はじめに

科学の進歩はデータの蓄積と不可分であり，一例として，ヨハネス・ケプラーによる天体運行の３法則の発見がティコ・プラーエによる詳細な天体観測データに基づいてなされたことは有名である．

ゲノム，人工衛星による地球観測データなどの多くのデータが公開される分野では，自らデータを測定するのではなく公開されたデータを解析し新しい知見を見いだす研究が大きな割合を占めるようになってきた．このような手法は"e-Science", "Data intensive science"などと総称され，第 4 期科学技術基本計画に向けた情報通信分野の重点事項に取り上げられている [1].

上記のような研究開発のみならず，センサ技術の進歩やインターネットの進化に伴い大量に生み出され流通するデータ"Big Data" [2]はプライバシー，経済活動，個々の事業展開をはじめとする社会のありかたに，人類史上でも最も急峻かつ広範な変化をもたらしつつある．この変化は情報革命や脱工業化社会，高度情報化社会，知識社会，知価社会への移行とも密接に関係している [3, 4].

ウェブ技術においては Hyper Text Markup Language（HTML）を主体とした文章のウェブから，Resource Description Framework（RDF）によって記述されるデータのウェブ（Semantic Web [5], Linked Open Data, LOD [6, 7]）に進化しつつある．LOD の活用例は増加しており，例えば国立国会図書館の書誌情報 [8]やノーベル賞に関する情報 [9]がある．

英国や米国では政府が率先して，保有する統計情報や調査結果などの公共データを再利用しやすい形式で公開する取り組みを進めている [10, 11].

我が国でも高度情報通信ネットワーク社会推進戦略本部（ＩＴ戦略本部）において，「公共データの活用促進に集中的に取り組むための戦略」が提示されている [12]. その一環として経済産業省では白書や統計情報などを公開するとともに [13]，計量標準や地質情報など知的基盤に関する情報を Open Data として公開する取り組みが開始されている [14].

２．科学技術医学に関するデータ

今日では科学技術医学（Science Technology and Medical, STM）分野のほとんどの学術誌に関してはインターネットで検索できることは周知である．論文検索に良く利用されるサービスとして Web of Science [15], Google Scholar [16]，国立情報学研究所論文情報ナビゲーター (CiNii) [17], J-STAGE [18], 大手学術出版社のウェブサイト，などがあげられる．検索された論文が Open Journal に掲載されていれば制限なく閲覧できるが、商業出版社に掲載されている論文の場合には所属機関が出版社と電子ジャーナルの購読契約をしていることが閲覧の条件となることが一般的である．

一方，大学や公的研究機関において自機関の成果である研究論文をレポジトリー (Repository) に公開することの重要性が認識され，国立情報学研究所（NII）では運営する「学術機関リポジトリポータル Japanese Institutional Repositories Online（JAIRO）」から日本の学術機関リポジトリに蓄積された学術情報（学術雑誌論文，学位論文，研究紀要，研究報告書等）を横断的に検索することができる [19].

主要な論文誌に掲載された論文には国際機関 CrossRef により「デジタルオブジェクト識別子（Digital Object Identifier, DOI）」が与えられており，URL の場合のようにリンク切れが生じる可能性がなく，Web 上での論文へ

＊ 産業技術総合研究所計測標準研究部門,
〒305-8563 茨城県つくば市梅園 1-1-1 中央第３
Metrology Institute of Japan, National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 3, 1-1-1, Umezono, Tsukuba, Ibaraki 305-8563, JAPAN
FAX: 029-861-4236, E-mail: t.baba@aist.go.jp

のアクセスが恒久的に保証される [20]．我が国においても 2012 年にジャパンリンクセンター（Japan Link Center, JalC）が設立され，邦文の文献についても DOI が付与されるようになった [21]．

研究者に対しても世界共通の ID（ORCID とよばれる）を整備する取り組みが進められている [22, 23]．

また，科学技術データの提供とその解釈との分業が進んできたことに対応して，論文のみならずデータセットに DOI を与え，データセット自体を引用する（DataCite）ことも実施されている [24-26]．

３．基礎物理定数

科学技術の定量的データのなかで最も根源的かつ普遍的な量は，光速度，プランク定数，ボルツマン定数などの基礎物理定数である．これらの量は基本単位の定義に直接関係しており，国家計量標準機関（NMI）や精密な物理測定に取り組んできた研究機関・大学の研究室により測定されてきた．これらの測定結果は科学技術データ委員会（Committee on Data for Science and Technology, CODATA）の基礎物理定数タスクグループ（Task Group on Fundamental Constants）[27]において検討され，合意された値が報告されている [28]．

それらの値は米国国立標準技術研究所（National Institute of Standards and Technology, NIST）のウェブサイトからデータベースとして提供されている [29]．

４．測定の信頼性とトレーサビリティ

Big Data という名称に象徴されるような膨大な情報が容易に取得できる状況においては，それらの情報の信頼性を評価することが根源的な課題となる．

長さ，時間，質量，温度などの定量的なデータはメートル条約 [30]によって定義された単位により表され，測定器を国家標準にさかのぼれる形で校正することにより信頼性が確保される．

それらの量の定義は科学の進歩に伴って明確になり，International Organization for Standardization（ISO）の規格（ISO 80000 シリーズ）により規定されている．日本工業規格（JIS Z 8202 シリーズ）には，ISO 80000 シリーズの旧版である ISO 31 シリーズの内容が反映されている．

これらの量を規格に基づく測定法により SI にトレーサブルな校正がなされた状態で測定することにより"不確かさ"の明らかな普遍的な数値データが求められる．

メートル条約を支える各国の国家計量標準機関（National Metrology Institute, NMI）は上記の定義に基づく計量標準を実現し，その標準が国際的に整合しているかどうかを検証すること，それらの標準を効率的に供給することを使命として設立されている．

現在，長さはセシウム 133 原子の基底状態の超微細構造と共鳴するマイクロ波の振動数と，基礎物理定数のひとつである真空中の光の速さによって定義されているが，2011 年 10 月に開催された第 24 回国際度量衡総会において，次回の国際単位系（SI）の改訂においては以下のように，SI 基本単位を自然界の普遍量（基礎物理定数または原子の特性）により表すことを決議した．

＊質量の単位であるキログラムはプランク定数の値から設定される．
＊電流の単位であるアンペアは電気素量の値から設定される．
＊熱力学温度の単位であるケルビンはボルツマン定数の値から設定される．
＊物質量の単位であるモルはアボガドロ定数の値から設定される．

５．データの誤差と不確かさ

数値データを生み出すための測定には必ず誤差を伴うことは周知であるが，誤差を評価することは測定自体よりはるかに大変なことが多い．誤差は測定結果と"真の値"との違いであるから定義は明確である．ところが"真の値"下記の例外を除き、定義を完全に満たす理想的な測定のもとで得られる値であるので現実には不可知であることが一般的である．例外として前述の光の速さがあげられる．光の速さは定義値なので定義上は"真の値"が既知であり，個々の研究室や事業所などで測定を行った場合には測定によって得られた値と光の速度の定義値との差が測定誤差そのものとなる．

電気素量，プランク定数，ボルツマン定数，アボガドロ数についても上述の SI 改訂以降は定義値となるのでそれらの値の不確かさは 0 となる．

"真の値"が未知である一般の場合には測定結果から"真の値"の推定と，誤差の推定を同時に行うことになり，「不

確かさの評価法」はその推定を実施する具体的手順を示している．

不確かさの評価方法は国際度量衡委員会（CIPM），国際標準化機構（ISO），国際電気標準会議（IEC），国際法定計量機関(OIML)，国際純正応用化学連合(IUPAC)，国際純粋応用物理学連合（IUPAP），国際臨床化学連合（IFCC）により合意された「測定における不確かさの表現のガイド，Guide for expression of uncertainty in measurement（GUM）」に記述されている［31，32］．

不確かさ評価においては個々の要因によって起こるばらつきを求め，それを合成することで全体のばらつきを算出する．個々の要因によるばらつきを評価するために，GUM では下記の A タイプと B タイプの 2 種類の手順を提示している．

A タイプの評価：実測によりデータを得てばらつきを求める．

B タイプの評価：実測以外の方法でばらつきを推定する．

A タイプの評価法の解析手法は標本の情報から母集団の情報を求める統計手法とのアナロジーで理解することができる．

６．本講座について

科学技術におけるデータをめぐる状況が最近数年で大きく変化し，その重要性に対する認識が飛躍的に高まった．本稿ではそれらの状況を概観するとともに，基礎物理定数や単位の定義に関する近年のメートル条約の取り組み，ならびに数値データの信頼性評価に必要な不確かさ評価について紹介した．

これらの動向と熱物性分野のデータとデータベースとのつながりについては次回以降に記述する予定である．

参考文献

[1] http://www.mext.go.jp/a_menu/kagaku/kihon/main5_a4.htm

[2] 城田真琴：ビッグデータの衝撃 -巨大なデータが戦略を決める-，東洋経済新報社（2012）

[3] ピーター・ドラッカー（著），上田 惇生（訳）：ネクスト・ソサエティ ― 歴史が見たことのない未来がはじまる，ダイヤモンド社（2002）

[4] 常深康裕：スーパーテクノロジー -世界を変えたネットワークとシステムの興亡-，行人社（2001）

[5] http://www.kanzaki.com/docs/sw/

[6] トム・ヒース，クリスチャン・バイツァー（著），武田英明（監訳）：Web をグローバルなデータ空間にする仕組み Linked Data，近代科学社（2013）

[7] http://linkeddata.jp/

[8] http://iss.ndl.go.jp/

[9] http://www.nobelprize.org/nobel_organizations/nobelmedia/nobelprize_org/developer/manual-linkeddata/terms.html

[10] http://data.gov.uk/

[11] http://www.data.gov/

[12] http://www.kantei.go.jp/jp/singi/it2/denshigyousei.html

[13] http://datameti.go.jp/

[14] 「知的基盤整備特別委員会」中間報告書

[15] http://science.thomsonreuters.jp/products/wok/

[16] http://scholar.google.co.jp/

[17] http://ci.nii.ac.jp/

[18] http://www.jstage.jst.go.jp

[19] http://jairo.nii.ac.jp/

[20] http://www.crossref.org/

[21] https://japanlinkcenter.org/top/

[22] http://orcid.org/

[23] http://current.ndl.go.jp/node/22097

[24] http://www.datacite.org/

[25] http://current.ndl.go.jp/node/22904

[26] http://librarylearningspace.com/npg-to-launch-scientific-data-to-help-scientists-publish-and-reuse-research-data/

[27] http://www.codata.org/

[28] Peter J. Mohr, Barry N. Taylor, David B. Newell, "CODATA Recommended Values of the Fundamental Physical Constants: 2010 a", J. Phys. Chem. Ref. Data 41, 043109 (2012), http://dx.doi.org/10.1063/1.4724320

[29] http://www.nist.gov/pml/data/physicalconst.cfm

[30] http://www.bipm.org/en/convention/

[31] ISO/IEC Guide 98-3：2008 [GUM:JCGM 100], Guide to the expression of uncertainty in measurement

[32] TS Z 0033：2012，測定における不確かさの表現のガイド，（ISO/IEC Guide 98-3：2008）