

e-研究インフラの海外動向 —これからの科学技術の国際競争の分け目か？

船守 美穂

国立情報学研究所 情報社会相関研究系

funamori@nii.ac.jp

The World-wide Trend towards e-Research Infrastructure—Will it be the key factor for S&T international competitiveness?

Miho Funamori

Information and Society Research Division, National Institute of Informatics

概要

研究者の研究活動が ICT と常時連動するようになったのに伴い、欧米を中心に世界では、物理的な研究インフラを補完する「e-研究インフラ」を整備することに力を入れ始めている。たとえば計測機器から出力された実験データが半自動的にこうした e-研究インフラに吸い上げられ、データを保存・共有・公開したり、当該インフラ上で提供される解析ツールにより解析したりすることができるようになる。あらゆる分野の研究者が大量のデータを扱う必要のある「データ集中科学」の時代において、こうした e-研究インフラの整備は科学技術における国際競争の分け目となると認識されている。

他方、日本においてこのような認識はなされていない。このため物理的な研究インフラを補完するデータベースや解析プラットフォームが併設されても、それは物理的研究インフラの全体予算の範囲内で行われ、手薄である。人員が十分に配置されず、担当者が疲弊している場合も多い。しかしこのような状況では日本の科学技術における国際競争力弱くなる一方である。

本小論は、世界における e-研究インフラへの取り組みの動向を紹介することを通じて、日本における e-研究インフラに対する認識を高めることを目的とする。

1 はじめに—研究データ管理への流れ

世界では学術研究活動の過程で生成される「研究データ」を管理しようとするイニシアティブが多層において始動している。

「研究データ」とは極めて広い概念であるが、分野を問わず、研究者が研究活動を行う際に生成したり、利用したりするデータのことを指す。たとえば、実験や観測データ、アンケートやインタビュー等の社会調査のデータ、計算機シミュレーションにより得られたデータなどがある。こうした一次データを解析した結果のデータも、研究データに該当する。また統計データや気象観測等の既存の公開データを自身の研究目的のために加工・整形したデータや、学術研究コミュニティや一般においてデータが利用しやすいよう整備したデータベースなどもこれに含まれる。

これら「研究データを管理する (research data management, RDM)」とは、第一段階においては、

当該研究データを利用・生成している研究者自身が、たとえばバージョン管理やデータの所在、共同研究者とのアクセス権の管理等で混乱せずに、自身の研究データを把握・管理できている状態を指す。しかし次の段階としては、研究の透明性や説明責任の観点から、外部からの要求に対して必要なデータを速やかに公開可能なように管理できている状態を指し、更に進むと、これら研究データが当該研究者以外の主体によって再利用可能な状態で提供される状態を指す。

こうした研究データ管理が追求される背景としては複数ある。一つには、ICT が普及し、学問分野を問わず、研究者が大量のデータを扱うようになり、「データ集中科学」とも言われるような時代に突入ことがある[1]。たとえば計測機器から細かいサンプリングレートで自動計測がなされ、大量のデータが出力されるようになった。人文系など伝統的にデータ解析とは縁の遠かった学問分野においても、デジタル・ヒューマニティーズなどに

見るように、人文系の学問領域に対して、データからのアプローチがなされるようになってきている。また社会の課題解決等学際領域の研究がなされ、複数の分野のデータが統合して利用されるようになった。これまでデータをそれほど扱わない分野でも大量のデータを扱うようになったことから、これらデータを適切に管理するニーズがあらゆる分野で生まれた。同時に、ICTの普及により共同研究が容易となり、データを共有したり、研究プロジェクトへの関わり方によってアクセス権を変えた設定としたりする必要が生まれた。これはいわば研究者や研究グループの、研究遂行上の利便性を高めるための、研究データ管理である。

一方、説明責任を求める要求も、研究データ管理の流れを後押ししている。説明責任への要求には二つの流れがある。一つは研究不正が多様な分野で勃発するようになり、論文等の研究成果の裏付けする研究データの公開が要求されるようになった。近年、複数の学術雑誌において、論文を執筆する際に利用されたエビデンス・データを、論文とともに公開することが求められるようになってきている。これらは学術雑誌のサプレメントに、あるいは外部のデータ・リポジトリにて公開される。もう一つは、公的研究資金による研究成果は公開し、社会に還元していかなければいけないという流れである。これについては欧米を中心に、研究助成機関が「研究データ管理計画（Data Management Plan, DMP）」を研究費申請書に添えることを義務化しだしている。またこの研究データ管理計画に基づいて実際に研究データが管理されたかをモニタリングすることも想定されている。研究データ管理に関しては、こうした助成機関による義務化が、最も強制力を働かせている。

研究データを管理する、より積極的な理由としては、研究データの再利用を促進しようとするものがある。政策的には、中小企業などが研究データを利用することにより産業上のイノベーションが生まれると謳われているが、より現実的には、学術的な利用による研究進展の可能性であろう。近年、社会課題の解決等につながる学際領域研究に注目が集まり、異分野のデータを組み合わせる機会が多くなっている。また単一の研究領域であっても、複数のデータを組み合わせる解析をする機会が増えている。

2 e-研究インフラ構築に向けての動向

研究データを管理するためには、なんらかの情報インフラが必要である。以下では e-研究インフラ構築に向けての歴史的展開と具体事例を示す。

2.1 e-研究インフラ構築に向けての歴史的展開

(1) 分野別の学術資源データベース整備

学術研究活動を促進するためのデータ管理の重要性はあらゆる分野で認識されており、古くは分野別の研究資源カタログやデータベースの整備が行われてきた。これの特徴は、当該研究コミュニティにおいて、研究者が共通して利用することの多い学術資源をデータベースとして整備しようとするもので、基礎物理定数や地震、ゲノム、古典籍データなど、いわば「標準データ」に関わる学術資源が、データベースとして整備されてきた。

こうしたデータベース整備の努力は主に、それぞれの研究コミュニティにおいて行われ、当該分野の専門家が専門的な見地からデータの収集、加工、提供を行ってきた。データベースシステムは商用で提供されているものなど、汎用的なものが多くの場合用いられてきたが、一部の理工系の分野では、分野ごとのニーズに照らした情報インフラの開発も行われた。この場合も、その分野の研究者が、情報インフラの開発・整備を行った。

(2) 大規模データのための高速・大規模コンピューティングの適用

一方では情報科学分野の立場から、大型計算機に始まり、グリッドコンピューティングやスーパーコンピューティング、HPC（高性能コンピューティング）など、大容量のデータを高速処理をするための技術開発が進み、これらが、大規模データが用いられる地球科学や天文、素粒子物理学などの分野で利用されるようになった。

ここでは計算機技術に関わる情報科学コミュニティが主体となり、情報インフラを開発・整備しているが、それぞれの分野の研究者も協働し、分野ごとのニーズに照らした技術開発を行っている。ここでの問題は、大規模データを扱い、しかもHPCなどの計算機技術を理解し利用することのできる、一部の理工系の分野しか、このような計算機資源を利用するニーズがなかったし、また能力的にも利用できなかったことである。このような事情から、大型計算機を整備・提供する大学等で

は近年、利用者の獲得に苦慮している。

(3) e-サイエンス等に向けた流れ

学術資源のデータベース化や高速・大規模コンピューティングの利用に加えて、ICT 技術の発展に伴いサイエンスが新たな局面を迎えるという考え方が、2000 年ごろから提唱されてきた。「e-サイエンス」が英国でジョン・テイラー科学技術庁長官により 1999 年に提唱された段階では、その概念はまだ(2)に挙げたような、高速・大規模コンピューティングにより支えられたサイエンスを指していたが、その後この概念はより広く捉えられるようになり、ICT 技術により高度化されたサイエンス全般を意味するようになった。より大規模なデータや、それに伴うより高度なデータ解析により、新たな学術的発見を得る可能性、異分野のデータを統合する可能性、研究者間の共同研究の拡大の可能性などが想定されている。

同様の概念は「e-リサーチ」や「Science 2.0」、近年では「オープンサイエンス」といったキーワードで時代を追って推進されている。オープンサイエンスの淵源となった Science 2.0 あたりからは、データ公開や社会における学術データの利用など、サイエンスをよりオープンに進めるといった考え方も含まれるようになってきている。

(4) オープンサイエンスの推進と EOSC

(3)に紹介したように、e-サイエンスと類似概念は 2000 年ごろから提唱されてきているが、政策アジェンダとした「オープンサイエンス」が明確に用いられるようになったのは、2013 年に G8 により研究データ管理に関わる共同声明がなされてからである。経済開発協力機構 (OECD) が “Making Open Science a Reality” という報告書を 2014 年にとりまとめ、そうした流れを受けて欧州委員会が “Open Innovation, Open Science, Open to the World” という政策提言を 2016 年にした[2][3]。欧州委員会ではオープンサイエンスを学術政策のみとして捉えるのではなく、「デジタル単一市場 (digital single market)」という、産業におけるデジタル化を推し進めるといふ産業政策の一部として提示した。このため 2016 年の報告書では、オープンサイエンスの推進により、産業上のイノベーションが創出される可能性が示されている。

欧州委員会では、オープンサイエンスの推進に向けて精力的な議論を行っており、これを実現す

る重要な手段の一つとして、「欧州オープンサイエンスクラウド (European Open Science Cloud, EOSC)」の形成を図っている[4]。EOSC は、既存の複数の e-研究インフラを連携し、欧州の 170 万人の研究者および 7000 万人の科学技術関連の専門家に対して、研究データを分野横断的に保存、共有、再利用可能な仮想空間を提供することを、構想している。これらは広帯域ネットワークと大容量ストレージ、スーパーコンピューティング能力とともに提供され、研究者はクラウド上の大規模データセットをデータ処理することが効率的にできるようになるとされる。

EOSC 同様、研究インフラの重要性に着目した政策展開は、各国ごと、また学問領域ごとであり、これらを欧州域内で調整し統合を図る組織体として e-IRG (e-Infrastructure Reflection Group) が 2003 年に発足している。e-IRG には、各国の科学技術に関連する政府組織や情報インフラを整備する主体などが集まり、協議を行っている。また欧州における研究インフラの整備計画と予算配分を検討する「研究インフラ欧州戦略フォーラム (European Strategy Forum on Research Infrastructures (ESFRI))」では、e-研究インフラの重要性に鑑み、e-IRG に e-研究インフラに関わる検討を委託し、2016 年の ESFRI ロードマップからは、e-研究インフラの整備計画が、他の分野別の研究インフラの整備計画と並列して、記載されるようになってきている[5]。

2.2 e-研究インフラの具体事例

2.1 節に紹介した e-研究インフラに関わる歴史的展開は主に政策レベルで行われているもので、具体的な事例があまり見えてこない。しかしこれらの元で、具体的な e-研究インフラの整備が行われている。また研究データ管理に関わる政策的・技術的検討事項について国際的に検討を行う「リサーチデータ・アライアンス (RDA)」の会合に行くと、こうした分野別に形成された e-研究インフラの整備・運営に関わる担当者に多数、出会うことができる。

以下にいくつかの具体的事例を示す。

(1) e-IRG で提示された e-研究インフラの具体事例

e-IRG では、2010 年に欧州における e-研究インフラのあり方について、2012 年には特にデータ管理に着目して青写真を描き、2013 年には e-研究インフラの白書を取りまとめている。2010 年および

2012年の報告書（Blue Paper）では、ESFRIが対象とする研究インフラのうち、e-研究インフラの色彩が強いものを一覧している（以下参照）[6][7]。

これらを見ると、これらe-研究インフラが人文社会系から理工系、医薬系など幅広い学問分野にまたがり、また大規模実験施設に付随するデータや、地域的に分散しているデータを統合したもの、人文系など比較的小規模ではあるが多様性のあるデータを整備・提供するものなど、多様なe-研究インフラが整備されていることが見て取れる。

【e-研究インフラの色彩の強いESFRI対象プロジェクト】

- **BBMRI:** Biobanking and Biomolecular Resources Research Infrastructure（バイオバンキング）
- **CLARIN:** Common Language Resources and Technology Infrastructure（言語資源）
- **CESSDA - Council of European Social Science Data Archives**（社会科学データアーカイブ）
- **DARIAH - Digital Research Infrastructure for the Arts and Humanities**（デジタルヒューマニティーズ）
- **ECRIN:** European Clinical Research Infrastructures Network（臨床研究）
- **ELIXIR - European Life Sciences Infrastructure for Biological Information**（生物情報）
- **e-VLBI:** Very Long Base Interferometry in Europe（VLBI）
- **ESRF:** European Synchrotron Radiation Facility（シンクロトロン放射光施設）
- **EuroFEL:** Free Electron Lasers of Europe（自由電子レーザー）
- **KM3NET - The Cubic Kilometre Neutrino Telescope**（ニュートリノ）
- **LIFEWATCH - e-Science and technology infrastructure for biodiversity data and observatories**（生物多様性）
- **Preparing for SKA - Square Kilometre Array**（SKA）
- **European XFEL - European X-ray Free Electron Laser**（X線自由電子レーザー）

【特にデータ管理を必要としているESFRI対象プロジェクト】

- **BioMedBridges:** Building data bridges between biological and medical infrastructures in Europe

（生物－医学データ連携）

- **CRISP - Cluster of Research Infrastructures for Synergies in Physics**（物理学）
- **DASISH - Data Service Infrastructure for the Social Sciences and Humanities**（人文社会科学）
- **ENVRI - Common Operations of Environmental Research Infrastructures**（環境研究）

(2) e-研究インフラの個別事例

以下ではいくつかのe-研究インフラを個別に紹介する。通常行われる研究領域に対して、こうしたe-研究インフラが併設整備されることで、研究活動がスムーズで効率的に行われるようになることが特徴である。

a) NFFA—Nanoscience Foundries & Free Analysis

NFFAは、ナノスケールの材料科学分野のためのデータおよび解析のために、欧州で構築されたe-研究インフラである。欧州諸国に点在する各種材料系の計測機器からはき出される計測値が、その計測条件とともにNFFAに転送され、保存される。あるサンプルを計測したいという研究者は、X線回折、NMR、質量分析、電子顕微鏡、微細加工などの計測機器を保有する施設を巡回し、計測をすることとなるが、計測されたデータはNFFA一箇所に吸い上げられる。NFFAは、これら計測値を解析ツールも備えており、研究者はNFFA上でデータ解析を行い、論文作成に挑むことができる。取得されたデータは一定期間を経たら、NFFA上で公開され、他の研究者が利用可能となる。

日本においても材料系の大型計測機器は、複数の大学共同利用機関等にあり、全国の研究者により利用されている。しかし計測されたデータは、個々の研究者が自身が用意した記憶媒体に保存され、研究者自身が用意した商用ソフトウェアや自作のプログラム等で解析される。データ管理も自身ですることとなり、計測を異なる条件で複数回しているうちに混乱することもある。また論文作成後、取得したデータは忘却の彼方にあり、再利用されることも少ないと想像される。NFFAのような仕組みが国内にもあると、研究活動もスムーズになり、成果輩出のスピードアップも図られると想定される。

b) DARIAH—Digital Research Infrastructure for the Arts and Humanities

DARIAHは、欧州におけるデジタル・ヒューマニ

ティーズの取り組みである。ESFRI プロジェクトの一部であるが、各国政府によっても助成され、参加国それぞれがデジタル・ヒューマニティーズに関わる e-研究インフラを提供する。解析ツールやデータ・レポジトリの提供、研修や人材育成のためのワークショップ実施、コンテンツの収集などさまざまな取り組みが行われている。

ドイツが取り組んでいる DARIAH-DE は、ドイツ政府からの強い梃子入れもあり、強力に推進されている。ドイツはインフラ整備に力を入れているが、単なる技術インフラ提供に留めるのではなく、インフラ提供側の技術者と人文系の研究者の共同研究というかたちで、革新的なデジタル・ヒューマニティーズのツール開発を進めている。たとえば史料を時間的・空間的にプロットするツールや、変則的に文字や図がレイアウトされている古典籍について図とテキスト部分とを自動判別するツールなど、複数の研究領域で利用可能なツールを開発する。開発するツールは、人文系および情報科学系の専門家からなる審査委員会にて審査し、決定する。

c) CESSDA—Consortium of European Social Science Data Archives

CESSDA は、欧州各国で開発・整備されている社会科学系データアーカイブのコンソーシアムである。現在、格メンバー国において整備されている社会科学系データアーカイブを横断的に利用可能とするよう、調整・整備を進めている。

社会科学の分野では伝統的に、データアーカイブを整備しデータを公開・共有することの重要性が認識されてきた。一部の国では、各国の政府統計のためのデータ収集と管理を社会科学に強い大学等学術機関に委ねてきた、という歴史がある。また英国などでは社会科学系リサーチカウンシルが早い段階から、研究者が行ったアンケート等の社会調査のデータを共有することの重要性を認識し、研究助成を受けた研究者や研究グループに、研究プロジェクトを通じて得た社会調査データを公開することを義務化してきた。このため英国には、UK Data Archive という、これら社会調査データ等を管理・公開するデータアーカイブがエセックス大学内に設置されている。

こうした各国で整備された社会科学系データアーカイブの運営体制やデータの充実度、データポリシーは国によってまちまちではあるが、

CESSDA は ESFRI の一部と認められたこともあり、これらの間を調整し、社会科学系データの利用可能性を最大化しようとしている。

3 考察—日本の現状と展望

(1) 日本における政策展開

研究データ管理および e-研究インフラ整備に関わる政策的展開および、それらに基づいて整備されつつある e-研究インフラの具体事例を紹介した。こうした世界の動向を日本が関知していなかったわけではなく、日本においても学問分野別に学術資源データベースが整備され、また大規模・高速コンピューティングについての世界的な競争にも一、二を争うほどに力が注がれている。しかし「e-サイエンス」や「オープンサイエンス」といった抽象的な政策アジェンダが提唱されるようになったところから、日本は世界の動向に遅れを取りだしているようである。

「e-サイエンス」や「オープンサイエンス」といった標語が日本で認識されていないわけではない。日本学術会議における議論を追うと、2008 年には E-サイエンス分科会、2014 年には国際サイエンスデータ分科会、2016 年にはオープンサイエンスの取組に関する検討委員会が設置され、それぞれにより提言がとりまとめられている[8][9][10]。提言には、これら政策アジェンダに関わる世界の動向が記され、世界と同様の認識が提示されている。しかし多くは報告書止まり、もしくはなんらかの取り組みがなされても小規模で、学術のあり方をデジタル時代に合わせたものにするために根本的に考え直すという流れには、学術的にも政策的にも、なっていない。

物理的な研究インフラに e-研究インフラを併設整備することで、研究活動をスムーズかつ効率的に促進できるようになるという認識も、欠落しているようである。日本学術会議では「学術の大型研究計画に関するマスタープラン」を 2010 年、2014 年、2017 年に発表しているが、そこには e-研究インフラへの言及はない[11][12][13]。欧州の研究インフラ欧州戦略フォーラム (ESFRI) のロードマップにおいて e-研究インフラが特出しされているのとは大違いである。

(2) 日本の研究現場における現状

ひるがえって学問領域ごとに整備されている学術資源データベースやその解析ツールをみると、

老朽化し貧弱な整備状況が浮かび上がる。これらデータベースや解析ツールは多くの場合、物理的研究施設の研究予算の一部を利用して開発・整備されており、一方では全体予算が縮小傾向にあるため、予算削減の影響を最も受けやすい領域となっているのである。

データベース等を開発・整備する人員も貧弱で、疲弊している場合が多い。当初は学術資源コンテンツを収集・整備するなど、当該学問領域に関わる専門性を活かした仕事ができていたのに対して、情報セキュリティなどの問題が多発し、情報インフラの管理に関わる仕事を中心となってきているのである。こうした情報インフラ整備の業務は、当該学問領域コミュニティでは十分には評価されないため、当該分野で研究者としてのトレーニングを受けてきた者にとっては、やりがいをもたない仕事となっており、学術資源データベース等がますます貧弱になるのに拍車をかけている。

各学問分野の研究者も疲弊している。近年研究データが大量に取得され、その管理・解析が追いつかないのである。一次データだけでなく、中間データや論文作成時のデータがあり、その履歴を記録したプロトコルなどもある。またそうしたデータに関わるファイルだけでなく、学会等の発表資料、論文、学内外での広報資料があり、さらには研究助成を得るための研究提案書やその中間・最終報告、また経費執行に関わる事務書類もあり、ラボが大所帯になるほど、管理が難しくなる。近年は共同研究も多く、これらではクラウド上のドライブでデータが管理されることが多いため、誰が最後まで責任をもって管理すべきデータなのか曖昧なまま、研究プロジェクトが終了し、データや関連資料がそのまま置き去りになることも多い。

データを解析するための解析ツールについても、研究者ごとに解析ツールを用意する必要があり、研究資金の有無により解析可能性が変わってくる。これは解析ツールをまかなえない研究者にとって不利なだけでなく、解析されたデータや論文等の研究成果を横断的に比較し、新たな知見を得ようとする上でも、不便である。

(3) 展望—「e-研究インフラ」の重要性にかかわる認識の必要性

日本は科学技術立国として古くから物理的な研究インフラの重要性は認識しており、各期の科学技術基本計画をみても、重点投資する物理的研

究インフラへの言及には事欠かない。他方、ICTやクラウドなどのバーチャル空間については、現実的な実感をもって検討をすることが出来ていないようである。E-サイエンスやオープンサイエンスなどの標語は掲げられ、それに伴う基盤開発などが行われても、そうした抽象的な次元に留まった施策となっている。

他方、研究の現場では、それぞれの分野の研究者等が、そうしたバーチャルな次元への対応もしており、疲弊している。「e-研究インフラ」の認識がなされ、そうしたところに特別の予算・人員の配備がなされている諸国と競争していく上で、大きなハンディがある。

日本においても、デジタル時代における学術活動を促進する上での「e-研究インフラ」の重要性を認識し、対応を図った方が良いのではないか。物理的な研究インフラに付随する e-研究インフラや、e-研究インフラを媒体とした研究コミュニティにおけるネットワークの緊密化など、複数の形態がありえるように思う。またこうした e-研究インフラの整備にあたり、当該分野の研究者にのみ依存するのではなく、そうした e-研究インフラに関わる研究コミュニティや専門家集団が形成されると良い。そうした分野を直接のターゲットとして予算が継続的配分されることで、そうした専門家集団におけるキャリアパスも形成される。

e-研究インフラ整備が、政策、予算、人員、研究の現場等で充実することによって、科学技術における国際競争に有利になることが期待される。

参考文献

- [1] Toney Hey, “The Fourth Paradigm: Data-intensive Scientific Discovery,” Microsoft Pr, 2009.
- [2] OECD, “Making Open Science a Reality,” 2015.
- [3] European Commission, “Open Innovation, Open Science, Open to the World,” 2016.
- [4] European Commission, “European Open Science Cloud,” 2016.
- [5] European Strategy Forum on Research Infrastructures, “ESFRI Roadmap 2016,” 2016.
- [6] E-Infrastructure Reflection Group, “e-IRG: Blue Paper on e-infrastructure,” 2010.
- [7] E-Infrastructure Reflection Group, “e-IRG: The Blue Paper on Data Management,” 2012.
- [8] 日本学術会議 情報学委員会 E-サイエンス分科会「日本における E-サイエンスの推進に関する諸課題」(2008)
- [9] 日本学術会議 情報学委員会 国際サイエンスデータ分科会「報告：オープンデータに関する権利と義務—本格的なデータジャーナルに

- 向けて一」(2014)
- [10] 日本学術会議 オープンサイエンスの取組に関する検討委員会「提言：オープンイノベーションに資するオープンサイエンスのあり方に関する提言」(2016)
 - [11] 日本学術会議 科学者委員会 学術の大型研究計画検討分科会「提言：学術の大型施設計画・大規模研究計画—企画・推進策の在り方とマスタープラン策定について—」(2010)
 - [12] 日本学術会議 科学者委員会 学術の大型研究計画検討分科会「提言：第 22 期学術の大型研究計画に関するマスタープラン (マスタープラン 2014)」(2014)
 - [13] 日本学術会議 科学者委員会 学術の大型研究計画検討分科会「提言：第 23 期学術の大型研究計画に関するマスタープラン (マスタープラン 2017)」(2017)