

② データベース化とデータの共有

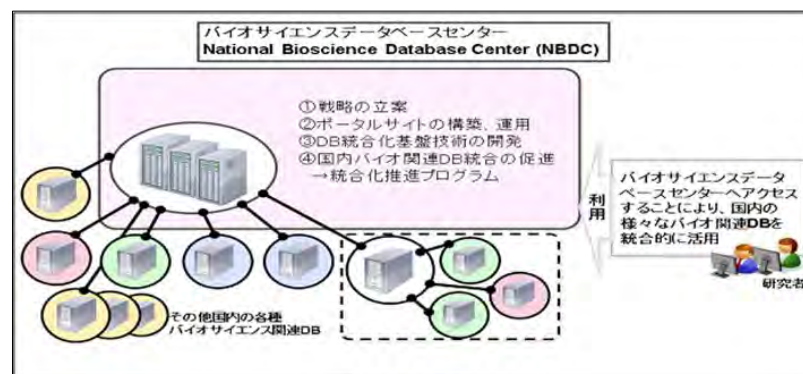
日本学術会議情報学委員会国際サイエンスデータ分科会の報告によると、米国、中国、インドでは、各分野に数十人から数百人のデータベースセンターが存在しており、多くのデータベースを開発している。日本には、このような大規模データベースセンターは存在しないが、研究、教育、産業、行政、医療など各分野で、データベースが生み出されている。しかし、国としてのデータ戦略がないことから、それらは、有機的に繋がることなく、バラバラに作られ、日本全体のデータ活動は把握されていないと指摘している。つまり海外に比べて、日本のデータ利用体制は貧弱で、その原因は、国力の差というよりは、国としての明確なデータ戦略が確立されていないことにあると指摘している。

特に研究分野においては、ライフサイエンス、地球観測、材料科学などの分野でデータベース化が進められているが、データの利活用の最大化が課題となっている。

(a) ライフサイエンス分野

ライフサイエンス分野においては、総合科学技術会議（現総合科学技術・イノベーション会議）ライフサイエンスプロジェクトチーム統合データベースタスクフォース（当時）の示す方針を踏まえ、2011年4月に設立されたJSTのバイオサイエンスデータベースセンター（NBDC）において、ライフサイエンス分野のデータベースの統合が進められている。

NBDCは、ライフサイエンス分野のデータベースを統合し、データの価値を最大化することにより、日本のユーザー、さらには世界のユーザーに貢献できる、日本が誇るべきデータベースセンターとなることを目指し、「戦略の立案」、「ポータルサイトの構築・運用」、「データベース統合化基盤技術の研究開発」、「バイオ関連データベース統合化の推進」の4つの活動を積極的に進められている。

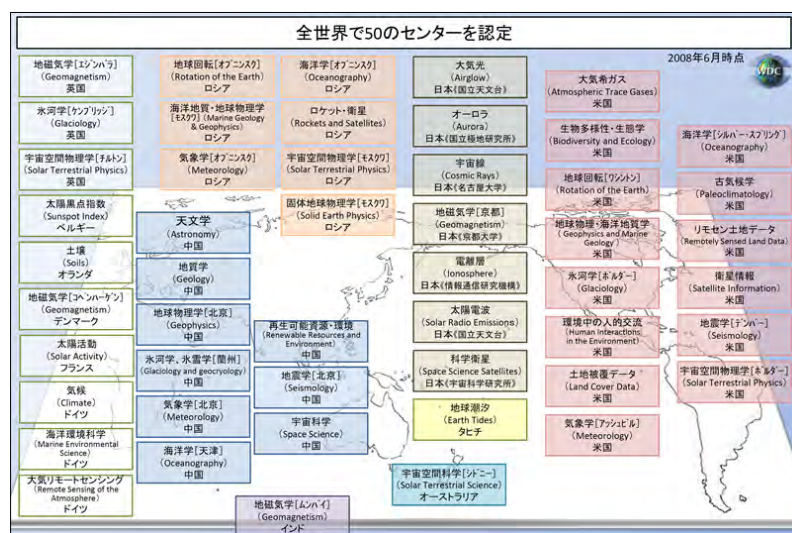


ただし、未だ根強いデータ生産者側のデータへの所有者意識、データ提供に報いるインセンティブ付与の問題、研究者間や分野間の用語の不統一、受入体制整備がライフ分野のデータ量の激増やデータの多様化に比して不十分といった課題が存在する。

ちなみに我が国では NBDC のほかに、ゲノム配列と蛋白質立体構造に関して国立遺伝学研究所の DNA Data Bank of Japan や大阪大学蛋白質研究書の Protein Data Bank of Japan が国際的なデータバンクを運営しているが、これらのすべてのセンターの人員体制を合算しても米欧に比べ、極めて小規模な状況に留まっている状況である。

(b) 地球科学・宇宙科学分野

地球科学・宇宙科学分野に関しては、1957年に開始された国際科学会議（ICSU）のWDC（World Data Center）事業で50年以上にわたり、認定されたWDC資料センター（日本は7センター）をはじめとして、データの保存や科学者向けのサービス活動の長い実績を持つ。近年は、科学技術の進展とともに多くのプロジェクトが実施され、新たなデータ取得、データベース整備、データ公開事業が行われ始めたが、必ずしも形式や公開ルール、猶予期間、データ提供者の権利などは統一されてこなかったため、国際的な共通ルールに適用できていない分野もある。また、価値の高いデータを取得・蓄積しておきながら、公開、一般共有ためのデータ整形やメタデータ・カタログ情報作成、利用者向け情報整理等のための人的資源が配分できず、データの利活用が最大化されていないものが散見される。



(c) 材料科学分野（ナノテクノロジー）

材料科学分野に関しては、独立行政法人物質・材料研究機構（NIMS）が、2015年4月から、物質・材料データベース（MatNavi）を運営している。このデータベースは、クリープ（物体に持続応力が作用すると時間の経過とともに歪みが増大する現象）や疲労（物体が力学的応用を継続的に、あるいは繰り返し受けた場合にその物体の機械材料としての強度が低下する現象）のデータシートを基にした構造材料データベース、公表されている学術文献から有用な数値データを採取し、データベース化した高分子、無機材料、金属材料、拡散、超伝導材料データベースなどから構成されている。材料開発、材料の最適な使用、最適な材料選択ばかりでなく、材料の特性予測、材料特性比較、材料の同定（辞書機能）などとして役立てることを目指している。この分野でオープンサイエンスを推進していくためには、国際的連携はもとより、マテリアルズズインフォマティクスが非常に重要であることが認識されるとともに、国際標準化に結びつくようなプラットフォームの構築が望まれている。



(d) 物理学分野

物理学の中では分野により価値観に差があるため、オープンサイエンスへの取り組みの状況は必ずしも一様ではない。

重要な取組の一つに、1991年から物理学分野を中心に国際協力の中で発展してきた重要な研究資料公開リポジトリとして、プレプリント・アーカイブ(arXiv.org)がある。これは物理学、数学、計算科学、定量生物学、定量経済学・統計学の分野をカバーし、査読雑誌の出版に先んじて最先端の研究情報の流通を担っている。全分野合計ですでに100万件を超える論文が登録され、無償でダウンロードできる。現在は国際協力チームの下でコーネル大学図書館が管理・運営しており、我が国では、基礎物理学研究所にミラーサイトが置かれている。リポジトリを活用するに際して、データベース検索システムが大変重要だが、素粒子・原子核・宇宙を中心とする物理分野でこれを担うのがインスパイア(Inspire)である。これは1969年にスタンフォード線形加速器センターが開発したスパイアス(Spires)が発展したもので、世界の主要研究所(我が国では高エネルギー加速器研究機構)が共同で管理運営している。世界の主要出版社やデータベース、プレプリント・アーカイブとも連携して、書誌情報・参考文献・引用などを含む情報が集積されており、利便性が高い。

高エネルギー物理学を中心とする実験での結果とそのデータについては、大きなプロジェクトは国際共同研究であることが通例で、各国際共同実験グループが世界の主要研究機関にデータベースとして保管し、公開も行っている。その集大成が、24か国の200名以上の研究者が参加する素粒子データグループ(Particle Data Group)という国際共同事業によって毎年まとめられて主要学術誌に出版され、年5000件を超える引用を受けている。この事業は1957年に始まったが、我が国の研究者も1970年代以後参加し、現在では活動の中心メンバーに入っている。我が国では、高エネルギー加速器研究機構に素粒子データベースのミラーサイトが置かれている。このように、科学研究の普遍性・