

# デジタル学術空間の未来に向けて

——縦割りではなく協働的な体制へ

大向一輝

## 1. はじめに

筆者の専門は情報学である。情報学とはその名の通り情報を対象とする学問であるが、あらゆる学問分野が何らかの情報を扱っていることを考えると、情報学特有の課題は存在しないのではないかと疑問に思われる方もおられるかもしれない。実際、多くの学問分野にとって情報技術は単なる研究の手段として捉えられているだろう。例えば人文学分野では、史資料のデジタル化やネットワーク化、そしてコンピューターの高性能化によって手元で扱うことのできる情報量は飛躍的に増大したが、基本的な研究手法は従来と大きく変わるものではない。一方で、情報学、あるいは情報科学や情報工学においても、その成果を利用する側の個別の事情には深く踏み込まず、誰もが等しく恩恵を受けられるように汎用化された技術を高く評価するという文化がある。その結果、今日まで各専門分野と情報学はつかず離れずの相互に独立した関係性が保たれてきた。

筆者のもう一つの活動の場である学術情報流通の分野においても同様の構造が存在する。学術情報流通とは、論文や書籍などの文献情報を通じた研究者同士のコミュニケーション環境を指す。各々の文献は内容こそ違えど、その様式や形態に大きな差はない。そのために学協会や出版社、そして図書館は一連の業務プロセスを整備することで、どのような専門分野の文献であっても同じように取り扱い、出版流通の枠組みの中で情報を循環させることを可能にしてきた。研究者もこの体制の存在を前提として、最終稿の作成までが研究活動であ

ると認識している。こうして研究者と学術情報流通機関の担当者は分業されており、たがいに干渉することは極めて少ない。

このような専門分野と情報学、研究活動と学術情報流通の「ほどよい」距離感は長期にわたって続いてきたが、この関係性が大きく変化する可能性が見えてきた。社会からの期待、そして研究活動のプロセスに直接介入するような情報技術を通じて科学の在り方を変革する動きが徐々に広まりつつある。「Eサイエンス」や「オープンサイエンス」と呼ばれるこの潮流は、近い将来人文学分野にも何らかの影響を及ぼすことになるだろう。本稿ではこれらの潮流の一端について紹介したい。

## 2. デジタル情報の固定化

学術情報流通のデジタル化は1990年代から進められており、電子ジャーナルに代表されるデジタル出版物に完全に占められている専門分野も数多い。電子ジャーナルについては市場の寡占化や価格の高騰などの課題はあるが、データベースや検索エンジンの整備、契約に基づく柔軟なアクセス制御技術の導入によって情報の入手性が飛躍的に高まったことは確かである。また、大規模な文献データベース自体を分析の対象として研究者個人や機関のパフォーマンス評価に用いられるなど、情報の2次的な利用も進められている。

これらの仕組みは、文献が出版というプロセスを経て固定化されていることを前提として機能している。その意味ではデジタル化以前の学術情報流通の体制をそのままの形でコンピューターネットワーク上に移植したものであると言える。一方、ここで用いられている情報技術自体は固定化を必須としない。デジタル情報は容易にコピーでき、常に編集が可能である。現にインターネット上では断片的かつ即興的な言説が飛び交い、時には炎上のような事態を招くが、議論が深まることでより質の高い結論に至ることもある。このような特性を持つ情報技術において、固定化とは人為的に設計されたルールの一つにすぎない。そのルールの存在が科学の発展に有益であるような局面においては残す、あるいは強化し、そうでない局面では捨て去るといった制御を精密に行えることがデジタル化の本質であると言える。

さて、このように自由度が極めて高いデジタル空間において、複数の研究者が同一の史資料を対象として議論できるような状況を実現するためには、その史資料にグローバルで一意的な識別子が付与されていなければならない。インターネット上での情報共有を実現してきたウェブ（World Wide Web）において、識別子は URI（Uniform Resource Identifier、通俗的には URL と称する）と呼ばれる標準規格によって定められている。しかしながら、ウェブ自体が持つ技術的な柔軟性に起因して、同一の情報が複数の URI によって呼び出すことができる状態になっている。例えば <http://www.l.u-tokyo.ac.jp> と <http://www.l.u-tokyo.ac.jp/#123> という 2 種類の URI にそれぞれアクセスすると同じようなページが表示される。両者を仔細に比較すれば、完全に同じ情報であることが明らかになるが、これは情報の提供側が同じ結果を送信するように設定しているからにすぎず、あらゆるウェブサイトで同様の結果になるとは限らない。

似て非なる URI が複数存在し、各研究者がそれぞれ異なる URI を引用して議論を行った場合に、それらが同一の史資料を指しているかを第三者が検証するには多大なコストを要する。この問題を回避するために、一つのデジタル情報には代表的な URI を明示的に一つ与えるという運用方針が提案されている。この URI を「永続的（permanent）なリンク」を意味するパーマリンクと呼ぶ。史資料を扱う大規模データベースではすべての情報にパーマリンクを設けることが強く推奨されており、近年では国内外の多くのウェブサイトが対応している。

学術論文においては、被引用数による評価を行うためにも、あらゆる論文に一意的な識別子を与え、この識別子を用いて引用を行うことが求められてきた。そのために、各出版社が自助努力でパーマリンクを維持管理するのではなく、標準的な識別子である DOI（Digital Object Identifier）を定め、DOI と実際の論文の提供サイトとをリンクさせるという戦略が採用された。海外では学術出版社が寡占状態であったことから急速に普及し、今日では目にするのできる英語論文の大半に DOI が付与されている。国内では、主に日本語の文献を対象として DOI を発行するジャパンリンクセンターが 2012 年に立ち上がり、徐々に普及が進んでいる。史資料データベースにおいても DOI は注目されており、国立国会図書館デジタルコレクションや国文学研究資料館が運営する新日本古

典籍総合データベースにおいて画像情報に DOI が与えられている。

DOI による情報の固定化は、文献や冊子といったひとまとまりの情報に限定されるものではない。原理的には文献中の章や節など、より細分化された単位で付与することが可能であり、また個々の図表に DOI が与えられている例もある。また、頻繁に内容が変化する情報のあらゆる版に対して逐一 DOI を与えるような仕組みも整備されつつある。ソフトウェアのように常時改訂が行われるものは、厳密なバージョン管理とそれに伴う DOI 付与が必要な分野である。人文学に必要な史資料や文献においても、このような固定化の技術を駆使することで新たな議論や研究を生み出す土壌となるだろう。

### 3. オープンサイエンスと研究データ

固定化の方向とは反対に、デジタル空間の柔軟性を活かして従来の分業の構造を解体する試みも進められている。その代表例であるオープンサイエンスは、字義としては「科学を開く」ための活動一般を意味するが、中でも研究データの共有および公開についての議論が先行している。ここで、研究データとは、学術論文や書籍などにまとめられた研究成果の証拠となるような情報から、研究の過程で収集あるいは作成される情報を含む幅広い概念である。またその形態も数値情報や画像、音声、映像など多種多様である。研究データの共有と公開によって、研究活動の透明性の向上や新たな共同研究の可能性が開かれることが強く期待されている。

一方、最終成果物だけでなくいわば中間生成物にあたる情報をも共有し、公開するためには数多くの課題がある。情報はネットワーク上にアップロードするだけで自然に利用されるものではない。まず、情報が他者に発見されるためにはその内容を簡易的に理解するためのメタデータを付加するとともに、研究者が日常的に情報収集を行っている場に対してその存在を示す必要がある。文献の場合には、これらの作業はもっぱら出版社や図書館などの機関によって担われてきたが、研究プロセスの中間段階において、研究データを対象とした作業を誰が行うのかは明確でない。また、文献とは異なり研究データの形態や種別は多様であることから、統一的なメタデータの記述ツールを定めることも現

実的ではない。このような混沌とした状況の中で、研究者自身がこういった作業に時間を費やすことは、研究の進展を遅らせる原因にもなり得る。

このように、オープンサイエンスの恩恵を十全に享受するためには、研究方法や体制を見直し、最適化する必要がある。その際に、いままも発展を続ける情報技術や、新たな職能を持つ支援者と一体となって進めることが肝要である。各国において、オープンサイエンスにおける実務を支えるための情報システムが開発されている。例えば、複数の組織にまたがる共同研究者間でデータを安全かつ効率的に保管するためのオンラインストレージや、情報の保存日時を記録して改ざんが行われていないことを証明するためのタイムスタンプ発行システム、文献の公開と同時に証拠となる研究データのメタデータ付加と公開を行うための機関リポジトリシステムなどがあげられる。また、公開された研究データを発見、入手するための検索エンジンも整備されつつある。今後は、外部公開されているデータを取り込み、処理や解析までを自動的に実行するような研究支援システムなど、さまざまな展開が考えられる。

情報システムの開発だけでなく、これらを自在に扱うことのできる人材の育成も重要である。研究データの公開に際しては、データの内容を理解し、適切な形態での公開を選択すること、さらには利活用を促進するための対外的活動までが求められることから、専門知識と情報技術に関する知識、そしてコミュニケーション能力を兼ね備えたデータキュレーターと呼ばれる職種が注目されている。大学や研究機関においては、リサーチアドミニストレータや図書館員などの研究支援職がこういった能力を身に着けるための教材やカリキュラムの開発が求められる。

#### 4. 人文学におけるデジタル学術空間

最後に、人文学ならではのデジタル学術空間の可能性について議論してみたい。筆者は情報学の観点から人文学分野の各研究を眺めている立場だが、多くの研究に共通する問題意識として、史資料群に内在する差異と同一性の検討があげられるのではないかと考えている。特に、この課題は史資料のデジタル化を行う際に重要な論点となっている。例えば異体字が多数存在するような史資

料群においては、それらを明確に区別して記述し、データベースに格納することが何よりも重要である。一方、検索時に個々の異体字を間違いなく入力しなければヒットしないようなデータベースは、利用者の専門知識の多寡によって使い勝手が大きく変わる。ことに初学者にとっての利用が困難なデータベースのみが存在するような分野は後継者育成が危ぶまれる。記録時には差異をできるだけ明確にし、利用時には少々の差異は同一のものとして捉えたいという相異なる要求のどちらを満たすべきかは常に議論の種となってきた。

しかしながら、現在はコンピューターの性能と記憶容量が大幅に向上していることから、記録のためのデータベースと利用のためのデータベースを分離するとともに、記録された異体字を検索時にどのように扱うべきかを明示化した辞書を整備することで、両者の課題を同時に満たすことは実現可能である。また辞書を複数用意することで多様な利用者の要求に応えることもできる。このような辞書、言い換えれば知識の整備こそが今後の課題であるが、人文学としてこれらの知識を通じたデジタル情報の利活用の体制を構築することができれば、他分野に対しても大きな影響を与えられるのではないかと期待する。

ここまで、デジタル学術空間の展望について述べてきた。いずれの方向性においても、研究者だけではなく、関連する機関や支援者の存在が必要不可欠であり、今後はそれらの支援者をどのように評価し遇するかという課題が生じるだろう。すでに、大規模な実験装置が必要な研究分野では、データの取得者がその情報の内容について執筆するデータ論文やその雑誌であるデータジャーナルが発刊されており、また研究論文においてデータの DOI やデータ論文の引用を義務化する動きもある。人文学は他分野と比較してひとりまたは少人数で研究が行われているという特徴があるが、今後はプロセスごとの縦割りではない、協働的な体制によって実施されることが増えるものと予想される。