論題 Title	第 3 部　データ活用社会を支えるインフラの現状と課題
他言語論題 Title in other language	Part3 Current status and issues of infrastructure for data-driven society
著者／所属 Author(s)	越前　功（ECHIZEN Isao）／国立情報学研究所　教授 ほか
書名 Title of Book	データ活用社会を支えるインフラ：科学技術に関する調査プロジェクト報告書 （Infrastructure for Data-Driven Society）
シリーズ Series	調査資料 2017-6　（Research Materials 2017-6）
編集 Editor	国立国会図書館 調査及び立法考査局
発行 Publisher	国立国会図書館
刊行日 Issue Date	2018-03-30
ページ Pages	79-108
ISBN	978-4-87582-815-0
本文の言語 Language	日本語（Japanese）
キーワード keywords	—
摘要 Abstract	データ活用社会を支えるインフラについて、オープンサイエンスとそれを支える e-研究インフラ、データサイエンティストの育成、個人情報とプライバシー等の動向と課題を概説する。

第3部　データ活用社会を支えるインフラの現状と課題

国立情報学研究所教授　越前　功

第3部では、データ活用社会を支えるインフラの現状と課題に焦点を当てる。Ⅰではデータ活用社会における科学技術面の要となるオープンサイエンスについて、欧米の事例を交えながら概説する。Ⅱではオープンサイエンスを支えるe-研究インフラの動向と課題について、欧米の事例を交えながら概説する。Ⅲではデータ活用社会を支える人材育成に着目し、データサイエンティストの育成事業や人材充足状況について概説する。最後に、Ⅳではデータサイエンスと法制度の側面から、個人情報とプライバシー、著作権及び安全保障について概説する。

Ⅰ　学術研究の在り方の変革触媒としての「オープンサイエンス」

国立情報学研究所准教授　船守　美穂

世界の科学技術政策の潮流として、オープンサイエンスの推進が鮮明となっている。平成25（2013）年のG8科学技術大臣会合（ロンドン）の共同声明(1)後、我が国においても、オープンサイエンスに関する報告書が内閣府(2)や日本学術会議(3)などにより取りまとめられ、第5期科学技術基本計画(4)にもその推進が盛り込まれるなど、布石が着実に打たれている。

一方、「オープンサイエンス」という言葉の持つ意味が極めて曖昧なこと、定義についての共通認識がいまだ形成されていないことなどから、その推進の力点は、推進主体や文脈によって異なる。公的研究助成を得た研究成果の公開義務化や、研究公正の観点からの研究の透明性と再現性の向上という、具体的な課題解決に関わる対応についてはおおむね一致があるものの、オープンサイエンスが本来目的とする学術研究の在り方の変革に関する理念については、十分な理解が形成されていない。

ここでは、オープンサイエンスの諸側面と、それらがどのような背景の下に生まれてきたかを紹介しつつ、欧米でオープンサイエンスがどのような文脈で推進されているか、我が国で最も欠けていると思われるオープンサイエンスの「心」について、指摘する。

1　オープンサイエンスの諸要素

現在、欧米先進国を中心に世界的に推進されているオープンサイエンスに関わる政策や取組は、異なる背景の下に進められていた複数の動きが、オープンサイエンスという用語の下に合流し、より広い概念の取組として位置付けられるようになったもの、と理解することができる。

＊　本稿におけるインターネット情報の最終アクセス日は、平成29（2017）年12月26日である。

(1)　“G8 Science Ministers Statement London UK,” 12 June 2013. gov.uk Website〈https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206801/G8_Science_Meeting_Statement_12_June_2013.pdf〉

(2)　国際的動向を踏まえたオープンサイエンスに関する検討会「我が国におけるオープンサイエンス推進のあり方について―サイエンスの新たな飛躍の時代の幕開け―」2015.3.30. 内閣府ウェブサイト〈http://www8.cao.go.jp/cstp/sonota/openscience/〉

(3)　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「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」2016.7.6.〈http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t230.pdf〉

(4)　「科学技術基本計画」（平成28年1月22日閣議決定）〈http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf〉

以下ではまず、それら異なる背景を持つオープンサイエンスの諸要素を紹介する。

（1）学術論文のオープンアクセス化

学術論文をインターネット上でオープンにアクセス可能（Open Access: OA）とすることに向けた動きは、学術は開かれたものであるべきであるという理念にも基づいているが、より明確な背景要因としては、学術論文が掲載される学術雑誌の購読料が1980年代半ばから現在にかけての過去30年間で4倍になり(5)、ハーバード大学のような裕福な大学も含め、大学が必要な学術雑誌を購入できなくなったことへの反発がある。2000年頃からオープンアクセス運動(6)が学術界を中心に湧き起こり、商用出版社の出版する学術雑誌に対する投稿や査読拒否などのボイコットや、これら雑誌を代替するOAジャーナルの創刊（ゴールドOA）、学術論文の著者最終稿を大学等が運営する機関リポジトリ(7)にて保存・公開する（グリーンOA）といった動きが生まれた(8)。

学術論文をOAで閲覧したいという要求は、社会からも起こった。重病患者が、自身の病状について学術論文に当たろうとして、購読料が高すぎて読めないという問題はかねてからあったが、次のケースがエポックメイキングであった。米国にて、弾力線維性仮性黄色腫（PXE）という希少疾患と診断された子ども二人の親であるシャロン・テリー（Sharon F. Terry）氏が子供の病状について調べるため学術論文に当たろうとしたところ、学術雑誌の購読料が高すぎたため、必要な学術論文が閲覧できなかった。大学等で行われている研究の多くは、公的研究資金、すなわち納税者の税金が財源となっている。研究活動に対して税負担をし、その研究成果を閲覧するために更に負担をしなければならないというのはおかしい。テリー夫妻は、子供の難病の研究を促進するために患者・研究者の支援団体「PXEインターナショナル（PXE International）」を1995年に設立し、世界のPXE患者から疾患に関するデータを集めるバイオバンクを構築する一方、疾患に関わる論文やデータの公開・共有を呼びかけた(9)。

学術雑誌の購読料高騰を背景に、社会からのこうした声が米国の議会において取り上げられた(10)ことが、米国国立衛生研究所（National Institutes of Health: NIH）を始めとする研究助成機関からの公的研究資金によって得られた研究成果のOAを義務化するという規則制定につながった(11)。NIHでは、「Pub Med Central」という、NIHからの研究助成を通じて得られた研究成果を保存・公開するためのリポジトリを用意し、ここでの公開を義務化している(12)。

---

(5)　“Monograph & Serial Costs in ARL Libraries, 1986-2011.” Association of Research Libraries Website〈http://www.arl.org/storage/documents/monograph-serial-costs.pdf〉

(6)　時実象一「オープンアクセス運動の歴史と電子論文リポジトリ」『情報の科学と技術』55巻10号, 2005.10, pp.421-427.

(7)　大学等学術機関がその学術成果をインターネットを通じて無償公開するための、サーバーと情報システムからなる仕組み。

(8)　“Read the Budapest Open Access Institute.” Budapest Open Access Initiative Website〈http://www.budapestopenaccessinitiative.org/read〉

(9)　シャロン・テリー「私の子供は科学的に解明されていない稀少疾患患者だった―私が研究すると決めるまでは―」2016.11. TED Website〈https://www.ted.com/talks/sharon_terry_science_didn_t_understand_my_kids_rare_disease_until_i_decided_to_study_it?language=ja〉

(10)　“Public Access to Federally–Funded Research: Hearing before the Subcommittee on Information Policy, Census, and National Archives of the Committee on Oversight and Government Reform,” House of Representatives, One Hundred Eleventh Congress, Second Session, 2010.7.29, pp.60-62.〈https://www.gpo.gov/fdsys/pkg/CHRG-111hhrg64928/pdf/CHRG-111hhrg64928.pdf〉

(11)　“House Backs Taxpayer-Funded Research Access,” 2007.7.20. Alliance for taxpayer access Website〈http://www.taxpayeraccess.org/news/2007/house-backs-taxpayer-funded-research-access.shtml〉

(12)　Division F Section 217 of Omnibus Appropriations Act, 2009 (Public Law 111-8)

なお我が国は、学術雑誌の購読料高騰が起こった1980年代半ばから、1985年のプラザ合意を経て2011年頃までに円の米ドルに対する価値が約3倍となったことで(13)、海外の購読料の高騰が相殺された。1990年代に一時期外国雑誌の受入数が落ち込んだものの(14)、その後電子ジャーナルの導入による包括的パッケージ契約（いわゆる「ビッグディール」）により購読可能な学術雑誌のタイトル数が飛躍的に拡大したことから(15)、学術雑誌の購読料高騰問題には一般に鈍感であった。近年は購読料高騰により購読を断念する大学が拡大しているものの(16)、これに対応するためにOAが必要であるという認識は学術関係者の間ではいまだに希薄である。日本の代表的な研究助成機関である日本学術振興会および科学技術振興機構は、助成した研究プロジェクトから得られた論文を全て原則公開とし、学術論文のオープンアクセス化を推進している(17)。

（2）研究データの公開・共有

学術論文だけでなく、その研究の基となった研究データも公開･共有しようとする動きがある。

実験等を通じて研究者が取得したデータを別の研究者が異なる視点で分析すれば、新たな成果が生み出される可能性がある。特に､複数分野のデータを組み合わせる学際領域研究や､様々な要因が組み合わされて生起する社会的な課題解決、産業上のイノベーションが促進されると認識されている(18)。研究データは研究目的に応じて実験手法や条件をきめ細かく微調整して取得されたものであり、その解析においては、それら細かい条件等を踏まえる必要があるため、データを単に公開・共有しただけでは、意味のある解析ができないと指摘する研究者も多い。しかし、研究データの公開は、次に挙げる説明責任及び研究公正の観点も背景として、進む方向にある。

公的研究資金を得た研究成果は公開されていくべきであるという、学術論文のOA義務化と同じ理由により、研究助成機関は研究データの共有も求めるようになりつつある。しかし、読まれることを前提として執筆される学術論文と違い、研究データは機密性の高い情報である場合があり、また一度公開されてしまえば、他の研究者が先んじてこれを論文化し、データを取得した研究者のキャリアに大きく影響を与える可能性もあることから、現段階では研究データの全面公開や公開義務化はなされておらず、欧米の先進的な研究助成機関であっても、競争的研究費の申請時又は採択の際に、研究データ管理計画（Data Management Plan: DMP）と呼ばれる計画書を求めるにとどめている。一般にDMPには、研究活動で取得予定の研究データの生成

---

⑬　1985年（プラザ合意前）には1ドル約240円であったが、2011年には1ドル約80円となった。「主要時系列統計データ表　東京市場　ドル・円　スポット」日本銀行時系列統計データ検索サイト〈http://www.stat-search.boj.or.jp/ssi/mtshtml/fm08_m_1.html〉

⑭　日本学術会議情報学研究連絡委員会学術文献情報専門委員会「電子的学術定期出版物の収集体制の確立に関する緊急の提言」2000.6.26, pp.3-4.〈http://www.scj.go.jp/ja/info/kohyo/17pdf/17_44p.pdf〉

⑮　尾城孝一「ビッグディールは大学にとって最適な契約モデルか?」『SPARC Japan NewsLetter』No.5, 2010.5, pp.2-3. 国立情報学研究所ウェブサイト〈http://www.nii.ac.jp/sparc/publications/newsletter/pdfper/5/sj-NewsLetter-5-2.pdf〉

⑯　大学図書館コンソーシアム連合によれば、2016年に海外大手出版社3社の電子ジャーナルパッケージ契約を中止した日本の大学は31校に上る。

⑰　日本学術振興会「独立行政法人日本学術振興会の事業における論文のオープンアクセス化に関する実施方針」2017.3.9.〈https://www.jsps.go.jp/data/Open_access.pdf〉; 科学技術振興機構「オープンサイエンス促進に向けた研究成果の取扱いに関するJSTの基本方針」2017.4.1.〈http://www.jst.go.jp/pr/intro/openscience/policy_openscience.pdf〉

⑱　EU, “Amsterdam Call for Action on Open Science,” 2016, p.2.〈http://openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf〉

方法、保存方法、共有方法、管理体制などを記述する[19]。

一方で、度重なる研究不正を背景に、学術論文などの研究成果の証拠となる研究データや、失敗に終わった実験等も含めた、研究の過程で得られた全ての研究データを保存し、求めに応じて提供できるようにする動きもある。研究データが保存されることにより、研究の再現性や透明性が向上することが期待されている。国内では日本学術会議が「研究データ10年保存ルール」を打ち出し[20]、各大学がこれに基づいて学内規定を設けるに至っている。また国際的な学術雑誌の中には、掲載された学術論文において使用された研究データをサプルメンツ（追補）として要求するものもある[21]。

（3）研究プロセスのオープン化

学術研究活動はオープンであるべきであるというオープンサイエンスの理念のもと、研究活動のあらゆる側面をオープンにしていく動きもある。

欧州委員会（European Commission: EC）によればオープンサイエンスは、オープンコード（オープンソース）[22]、オープンアノテーション[23]、オープンラボブック[24]、オープンワークフロー[25]、プレプリント[26]、サイエンスブログ[27]などを含む概念であるとしており[28]、これらの用語からは、研究プロセスのあらゆる側面をオープンにしていく思想が見て取れる。しかしながら、コンピュータプログラムを公開し共同開発を進めていくオープンソースや、投稿前の学術論文の原稿を速報性の観点から共有するプレプリントを除くと、これらの研究プロセスのオープン化は、強制力が働かないこともあって、それほどの大きな広がりは見せておらず、関心のある研究者が行っている程度である。

また、インターネット上での人々の協働作業は、オープンサイエンスが実現される上で不可欠な要素と考えられている[29]。ブロードバンドインターネットが普及したことから、インターネットへのアクセスやインターネット上での活動が容易となり、人々の協働作業が促進された。特にアカデミアにおいては、国際共同研究や社会とのコラボレーションが拡大した。インターネット上の協働作業は、関心ある者がそうした作業グループを発見し、グループに参加することを容易にし、また、多様なメンバーによる議論の過程を確認可能とすることから、研究活動のオープン化につながるとされている。

---

⑲ Digital Curation Centre, “DCC CMP Checklist for a Data Management Plan,” v4, 2014.〈http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP-checklist-flyer.pdf〉

⑳ 詳しくは、日本学術会議「科学研究における健全性の向上について」2015.3.6, pp.iii, 8.〈http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-k150306.pdf〉

㉑ 池内有為「研究データ公開に関する学術雑誌のポリシー分析」『三田図書館・情報学会研究大会発表論文集』2013年度, pp.9-12.〈http://www.mslis.jp/am2013yoko/03_ikeuchi.pdf〉

㉒ コンピュータプログラムのソースコードをインターネット上で公開し、自由に利用可能とすること。オープンソースと同じ意味である。

㉓ 複数のインターネットサイトに付加されたコメントを公開・共有できる仕組み。

㉔ 研究ノートへの研究活動の記録をオンラインで公開しながら行う仕組み。

㉕ 研究活動をオンライン上で公開・共有することにより、誰もがその研究活動に確認・参加可能とする仕組み。

㉖ 学術雑誌に正式掲載される前の学術論文を公開し、他の研究者からフィードバックを得る仕組み。

㉗ 科学研究における発見等を公開・共有するインターネット上のブログ。

㉘ European Commission, *Open Innovation, Open Science, Open to the World: a Vision for Europe,* Luxembourg: Publications Office of the European Union, 2016, p.36.〈http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=16236〉

㉙ OECD, “Making Open Science a Reality,” *OECD Science, Technology and Industry Policy Papers,* No.25, Paris: OECD Publishing, 2015, p.7.

（4）査読プロセスのオープン化

前項目の研究プロセスのオープン化の一環として、特に学術論文の査読プロセスをオープンにしていこうという動きがある。

学術論文の査読とは、投稿された論文を学術雑誌に掲載するか否かを判断するための審査プロセスであり、当該学問分野でなされる研究の質の維持と向上の要であるとされている。学術雑誌のエディター（編集者）により指名された複数の査読者が、投稿された論文の採択の是非や改善のポイントについてコメントを記し、エディターはこれらを基に、当該論文の採択を判断する。また、査読者のコメントは匿名で論文著者に送られ、論文著者はこのコメントを基に自身の論文を改善する。査読者は一般には論文著者と同じ分野の研究者であり、査読プロセスは専門家同士で当該分野の研究の品質保証をする仕組みとして、学術界において広く採用されている。

査読が匿名で行われるのは、忌憚（きたん）のない客観的意見を査読者が述べるための配慮ではあるものの、査読に恣意性が生まれる可能性があること、また査読による指摘が論文著者以外の研究者の研究遂行にも参考になる可能性があることから[30]、査読を記名式で行うなどの「オープン査読」が試みられている。また近年、投稿される論文数が膨大となり査読が追いつかなくなってきていることから、却下された論文の著者が他の学術雑誌への再投稿を行った場合に査読内容を次の雑誌に引き継ぐ「カスケード査読」、投稿された論文をひとまずウェブ上に掲載し、後からこの論文に関心を持った研究者がオープンにコメントを付していくという「出版後査読」（Post-publication peer review）など、多様な査読方法が考案・試行されているが、いずれもいまだ十分には定着していない。

なお少し異なる観点ではあるが、査読をアカデミアに閉じずに、社会の意見も取り入れたものとするため、発表後の論文の内容に関するツイッターやフェイスブックなどの SNS におけるインパクトなども評価する「オルトメトリクス」（Altmetrics）も試行されている[31]。

（5）第四の研究パラダイム―データ中心科学―

デジタル時代となり、学問分野によらず、あらゆる分野において膨大なデータが生成・取得されるようになっている中、研究者自身による大容量データに基づく科学の方法論を総称して、「データ中心科学」という概念が提唱されている。科学は、「理論科学」、「実験科学」、「シミュレーション科学」と時代を追って新たなアプローチが加わってきているが、これに「データ中心科学」という新たな次元が加わり、科学は「第四のパラダイム」[32]に移行しつつあるといわれている。

デジタル時代のデータ氾濫としては、インターネット上のライフログ[33]などを中心とする「ビッグデータ」がよく知られているが、これだけでなく、伝統的な学問分野においても収集・利用される研究データが膨大となり、当該分野の学問的知見のみではデータ処理や解析が追いつかなくなっていることが指摘されている。例えば実験科学の場合、ある試料の計測は計測機器における自動化が進み、手動で計測と記録を行っていた時代とは比べものにならないほどの

⑶⓪　Tony Ross-Hellauer, “What is open peer review? A systematic review,” *F1000Research,* 2017.〈https://f1000research.com/articles/6-588/v2〉

⑶①　林和弘「研究論文の影響度を測定する新しい動き―論文単位で即時かつ多面的な測定を可能とする Altmetrics―」『科学技術動向』134 号, 2013, pp.3, 26.〈http://www.nistep.go.jp/wp/wp-content/uploads/NISTEP-STT134J-2.pdf〉

⑶②　Tony Hey et al., eds., *The Fourth Paradigm: Data-intensive Scientific Discovery,* Redmond: Microsoft Research, 2009.

⑶③　インターネット上の人々の情報探索やアクセス、閲覧などの行為の記録。

データが取得される。社会科学などでは大規模アンケート調査が可能となり、人文学でもデジタル・ヒューマニティーズ[34]という新たなアプローチがなされるなど、データ処理と解析のスキルが研究活動を大きく左右するようになっている。一方、あらゆる分野において研究の中心的活動がデータ処理となるにつれ、デジタルデータやデータ解析を媒体とした研究リソースの共有とそれに基づく新たな研究の展開が容易となった。これが研究活動のオープン化、すなわち「オープンサイエンス」へとつながっている。

（6）市民科学と社会的ニーズへの対応

研究活動をより社会に開かれたものにしていこうという動きもある。これには、市民との協力の下で学術的な研究課題について研究活動を行っていく方向と、社会の課題をアカデミアが解決していく方向とがある。

市民との協力の下で行う研究活動は一般に、「市民科学」と呼ばれている[35]。希少種の発見報告や星雲の形状判定など、人手がかかる研究活動について、これらに関心ある市民の協力を得て研究を進めるものである。一般には、インターネット上のサイトが媒体となり、研究者と市民の間の協力活動が行われるが、時には顔を合わせるサイドイベントを設け、目的意識の共有や詳細の説明などを行う場合もある。

他方、社会的課題の解決に向けて研究者が研究活動を行うことで、その解決に向けて、研究者が社会に協力する流れもある。これには、アカデミアの知見を積極的に社会の課題解決に活用したいという社会からの期待が背景にある。近年、研究助成がなされたプロジェクトの研究評価等において社会的インパクトが問われるようになった[36]こともあり、社会的課題解決に前向きな研究者は多いが、一方で、特に研究者キャリアの評価においては、学術的な研究テーマに基づく学術論文による評価の比重が依然として大きく、この状況は社会的課題解決に向けての研究推進の足かせとなっている。

2　欧米におけるオープンサイエンスに向けた考え方

前節で紹介したように、オープンサイエンスは、全般に学術研究を開かれたものにしていこうという動きである一方、それが推進される背景には、「購読料高騰への対応」、「納税者への説明責任」、「学術活動の社会的インパクト向上」、「研究公正への対応」、「査読の恣意性への対処」、「査読負担の低減」、「研究データ過多への対応」など、オープンサイエンスが極めて現実的な諸課題を背景に、推進されていることが見て取れる。

では、オープンサイエンスがこうした現実的な諸課題への対処としてのみ推進され、「学術研究を開かれたものとする」という理念は単にこれらを美しく見せかけるものであるのかというと、そうではなく、幾つかの政策文書や、様々な国際会議における議論から、欧米では政策担当者だけでなく、データ管理者や学術情報流通に関わる者、出版社などが真剣に「オープンサイエンス」に期待している様子が実感される。

⑶④　デジタル媒体による学術資料のアーカイブ構築、文化コンテンツの分析、学術成果の公開や展示の方法などを文系・理系の枠組みを横断して研究する、比較的新しい学問領域。本報告書第1部第Ⅳ章の「5　デジタル・ヒューマニティーズ」を参照。

⑶⑤　マイケル・ニールセン（高橋洋訳）『オープンサイエンス革命』紀伊國屋書店, 2013.（原著名：Michael A. Nielsen, *Reinventing Discovery: The New Era of Networked Science,* 2013.）

⑶⑥　“What is the REF.” Research Excellence Framework Website〈http://www.ref.ac.uk/about/whatref/〉

（1）各国政府・国際機関における共通認識

欧州におけるオープンサイエンス政策の発端となったのは、2014年に実施された「サイエンス2.0」に関するパブリックコメント募集[37]である。そこではサイエンス2.0を、「研究実施及び科学を整理する方法の継続的な進化」（on-going evolution in the modus operandi of doing research and organising science）と表現し、そのような進化がデジタル技術によって可能となり、アカデミアのグローバルな共同研究、そして地球規模の課題解決が促進されるとした。またサイエンス2.0を、研究活動の全般にわたってインパクトを与えるものとして位置付けた。サイエンス2.0の「2.0」には、科学がデジタル時代において、新たな次元に移行するという思いが込められている。他方、このパブリックコメントにおいて、「サイエンス2.0」ではなく「オープンサイエンス」とした方が良いという意見があり、それ以降「オープンサイエンス」という用語が用いられるようになっている。翌年の経済協力開発機構（OECD）の報告書[38]では、オープンサイエンスの定義の範囲を狭めて、「公的研究資金を得た研究成果をデジタル・フォーマットで広くアクセス可能とするための取組」と表現し、オープンサイエンスが推進される背景要因として、「従前からの学術活動のオープン性と、学術界の活動を再形成（reshape）させつつあるICTとの出会い」があり、これが「長期的な研究活動及びイノベーションにつながる」と期待している。

これらの政策文書を結実させた形で、欧州委員会は2016年に、産学官を含む、総合的なオープン化の方針を示す報告書[39]と、学術界に特化したオープンサイエンス政策に関する報告書[40]を提出した。前者ではオープンサイエンスを「協働型研究と知識拡散の、新しい方法に基づいた、科学プロセスへのアプローチ」、後者ではオープンサイエンスを「研究者の研究、協働、コミュニケーション、リソースの共有、研究成果の拡散の方法に関わることである」と表現している。なお、2013年のG8科学技術大臣会合では、オープンサイエンスの理念に沿った「オープンな科学研究データ」として、「公的研究資金を得た研究データは可能な限りオープンであるべき」との共同声明[41]を行い、その後2016年につくば市で行われたG7科学技術大臣会合以降は明示的に、「オープンサイエンス」という用語を用いて、その理念を共同声明[42]として取りまとめている。

これらの政策文書は、強調する点は少しずつ異なるものの、オープンサイエンスが、「ICT技術により可能となった、新しい学術研究の方法で、これまでの伝統的な学術界における価値に変化を与えるもの」であるという点では、おおむね一致している（表1）。

表1　オープンサイエンスに関する共通認識

・学術研究の行い方や科学の整理の新たな方法であること ・ICT技術により可能となったこと ・伝統的な学術界の価値を変えるものであること

（出典）各種資料を基に筆者作成。

---

(37) 詳しくは、“Consultation on ‘Science 2.0’: Science in Transition.” European Commission Website〈https://ec.europa.eu/research/consultations/science-2.0/consultation_en.htm〉

(38) OECD, *op.cit*. (29)

(39) European Commission, *op.cit*. (28)

(40) EU, *op.cit*. (18)

(41) G8 Science Ministers Statement London UK, 12 June 2013, *op.cit*. (1)

(42) 「G7茨城・つくば科学技術大臣会合　つくばコミュニケ（共同声明）」2016.5.17. 内閣府ウェブサイト〈http://www8.cao.go.jp/cstp/kokusaiteki/g7_2016/2016communique.html〉

（2）オープンサイエンスに関する国際会議における認識

オープンサイエンスの方向性を担う国際会議としては、「リサーチ・データ・アライアンス」（Research Data Alliance: RDA）と、「FORCE11」が挙げられる。

RDAは、研究データの共有を実現する社会的・技術的インフラを構築することを目的とした実務担当者主体の組織で、欧州、米国、豪州の参加国を中心とした助成により2013年に設立された。研究者が自身の研究成果を発表する通常の学会等とは異なり、RDAでは、研究データの共有に関わる実務担当者やこれに関心を有する人々が、表2に示すようなテーマに関連するワーキンググループやインタレストグループを形成し、検討を随時行い、その情報を年に二回開催されるRDAの国際会議で共有、協議する。その成果は、例えば、「データ共有の用語集」、「メタデータ標準」、「データのインターオペラビリティ（相互運用性）」、「関連のデータポリシー」等として取りまとめられ、データを共有していく上での国際的な指針となる。設立後4年経過した段階で、既に6,000名以上のメンバーの登録があり、RDAの会合には通常、400名以上が参加する。参加機関には学術関係、政府機関、IT企業、助成機関、メディア関係、参加者としては研究者、プログラム／プロジェクト・マネジャー、大学図書館員、ITエンジニア、政策担当者、ジャーナリスト、コンサルタントなどがいる。(43)

表2　RDAが対象とするテーマ

・研究の再現性	・データ引用
・データ保全	・データタイプのレジストリ
・分野別リポジトリのベストプラクティス	・メタデータ
	・法的インターオペラビリティ　ほか

（出典）Research Data Alliance, “What does RDA do.” (Slide 4 in “RDA in a Nutshell: December 2017.”) RDA Website〈https://www.rd-alliance.org/sites/default/files/attachment/RDA_in_a_nutshell_December_2017.pptx〉を基に筆者作成。

FORCE11の正式名称は、「研究コミュニケーション及びe-学問の未来」（The Future of Research Communication and e-Scholarship）であり、デジタル時代における研究成果の流通や研究の在り方そのものの未来を構想する国際的な会議体である。2011年1月、カリフォルニア大学サンディエゴ校において、PDFファイルによる学術論文などの機械可読性が低いことを克服したいと考える人々が「PDFを超えて（Beyond PDF）」というテーマの会議で集まり議論を行った。続いて同年8月、情報学の分野において権威のあるダグストゥール・セミナー（Dagstuhl Seminar）(44)において、未来の学術情報流通について議論がなされ、継続的に議論していくための会議体として、FORCE11が結成された。FORCE11の「11」は、これらの会議が2011年に行われたことを記念している。

FORCE11は、研究者や大学図書館員、デジタル・アーキビスト(45)、出版社、研究助成機

---

(43)　Research Data Alliance, “What does RDA do.” (Slide 4 in “RDA in a Nutshell: December 2017.”) RDA Website〈https://www.rd-alliance.org/sites/default/files/attachment/RDA_in_a_nutshell_December_2017.pptx〉

(44)　情報学における世界最高峰のセミナー。ドイツのダグストゥールで毎週のように開催されている。約1週間、合宿形式でトピックに基づいた議論を集中的に行うことで有名。「NII湘南会議」国立情報学研究所ウェブサイト〈http://www.nii.ac.jp/about/international/shonanmtg/〉

(45)　文化資料等のデジタル化についての知識と技能を持ち合わせ、文化活動の基礎としての著作権・プライバシーを理解し、総合的な文化情報の収集・管理・保護・活用・創造を担当できる人。「デジタル・アーキビストの概要」日本デジタル・アーキビスト資格認定機構ウェブサイト〈http://jdaa.jp/about-1.html〉

関、その他の関連機関など多様な主体が集まり、学術情報流通や研究の在り方の未来について議論することに特徴がある。通常の会議は、限られた分野の人々が集まって議論をするため、議論が一定の枠にはまってしまうことが多いが、FORCE11 では、多様な立場の人々が議論することにより、これまでになかった考え方が生まれることが期待されている[46]。実際に、FORCE11 の成果物として、「FAIR データ・プリンシプル」[47] や「データ引用プリンシプル」[48] といった、オープンサイエンスの推進において影響力のある指針が生み出されている。

RDA が実務的課題について検討を行っているのに対して、FORCE11 は未来の学術情報流通と研究の在り方の理念について構想をしており、これら会議の幹事によると、両者はお互いをうまく補完し、オープンサイエンスの推進に寄与しているという。両者に共通するのは、ICT 技術によりもたらされたデータや学術論文などの共有可能性が、学術の在り方を根底から変えるものであるという認識であり、新たな学術の方法や価値体系の創造に向けて、両者とも真剣に検討を行っている。

3　我が国におけるオープンサイエンスの考え方

我が国におけるオープンサイエンスの議論は端緒についたばかりである。現在は、2013 年の G8 の共同声明以降に作成された内閣府、日本学術会議、科学技術・学術審議会などの報告書に記されている施策が着実に実行されつつある[49]。

国内の政府報告書としてオープンサイエンスについて初めて言及した 2015 年の内閣府の報告書は、表 3 のようにオープンサイエンスを定義しており、ここには「公的研究資金を用いた研究成果」に関する説明責任の視点や、「広く容易なアクセス・利用を可能」とする視点、「新しいサイエンスの進め方」であるという視点が含まれている。また、「ICT の様々なツールと結びついて生まれたアプローチ」であることにも言及しており、デジタル時代における新たな物の見方であるという認識が示されている。

表３　内閣府の報告書におけるオープンサイエンスの定義

オープンサイエンスとは、公的研究資金を用いた研究成果（論文、生成された研究データ等）について、科学界はもとより産業界及び社会一般から広く容易なアクセス・利用を可能にし、知の創出に新たな道を開くとともに、効果的に科学技術研究を推進することで、イノベーションの創出につなげることを目指した新しいサイエンスの進め方を意味する。

（出典）国際的動向を踏まえたオープンサイエンスに関する検討会「我が国におけるオープンサイエンス推進のあり方について―サイエンスの新たな飛躍の時代の幕開け―」2015.3.30, p.5.〈http://www8.cao.go.jp/cstp/sonota/openscience/150330_openscience_1.pdf〉

---

⑯　“FORCE2017.” FORCE11 Website〈https://www.force2017.org/〉

⑰　“The FAIR Data Principles.” FORCE11 Website〈https://www.force11.org/group/fairgroup/fairprinciples〉研究データは、FAIR（Findable, Accessible, Interoperable, Reusable）、すなわち発見可能、アクセス可能、相互連携可能、再利用可能なかたちでオープンにされなくてはいけないとしている。

⑱　“Joint Declaration of Data Citation Principles: Final.” *idem*〈https://www.force11.org/datacitationprinciples〉データ引用をすることの目的、機能、データ引用の持つべき特性に関する原則が記されている。

⑲　国際的動向を踏まえたオープンサイエンスに関する検討会「我が国におけるオープンサイエンス推進のあり方について―サイエンスの新たな飛躍の時代の幕開け―」2015.3.30, p.5.〈http://www8.cao.go.jp/cstp/sonota/openscience/150330_openscience_1.pdf〉; 日本学術振興会　前掲注⑰; 科学技術・学術審議会学術分科会学術情報委員会「学術情報のオープン化の推進について（審議まとめ）」2016.2.26〈http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm〉

この内閣府の報告書及び、類似の考え方に基づく後続の報告書を経て、日本学術振興会はオープンアクセス方針[50]、科学技術振興機構はオープンサイエンス方針[51]を取りまとめ、研究助成の過程で方針の履行を助成先に求めるようになった。いずれの方針も、助成を得て生み出された学術論文のオープンアクセス化を原則としている。科学技術振興機構は更に、「研究データ管理計画」の策定とその履行を求め、学術論文のエビデンスデータは公開を推奨、それ以外の研究データについては公開を期待している[52]。また日本学術会議の提言[53]に基づき、国立情報学研究所が国内学術機関の研究者が利用可能で、研究分野を越えた研究データの管理およびオープン化を可能とする「研究データ基盤」も整備しつつある[54]。オープンアクセスリポジトリ推進協会（JPCOAR）では、研究データをリポジトリにて保存・公開する際に必要となる、メタデータの標準の検討も進み[55]、また研究データ管理を行うためのトレーニングツールも、JPCOARにおかれた研究データタスクフォースと国立情報学研究所の協力の下、作成された[56]。これらの基盤やツールが標準的に提供・利用されるようになり、一方で研究助成機関が公的研究資金を得た研究の成果についてのオープンアクセスへの要求を強めれば、オープンサイエンスの理念に基づいた行動をとる研究者も拡大するであろう。

現在行われている「公的研究資金を用いた研究成果」におけるオープンアクセスの推進や、「研究データのオープン化」に向けた環境整備などの目に見える対応を通じて、オープンサイエンスが目的とする「新しいサイエンスの進め方」が自然と生み出される可能性がある。ただし、これらの動きが欧米のように「新しいサイエンスの進め方」を生み出すことを、明確に意識して実施されているかというと疑問が残る。同時に、研究データのメタデータ標準は、我が国でも検討されてはいるものの、表2に挙げたような、研究データの共有化に関連した諸課題に関する専門的な検討は不十分である。我が国では、研究データの管理に専門的に対応している人材又は組織が少なく、それぞれの学問分野の研究者が本来の研究の傍らでデータベース等の管理・運営を行っていることが多いため、データ管理に関わる専門団体の形成や、専門の議論が発展しにくい。結果として、RDAでのワーキンググループにおける我が国のプレゼンスは、相対的に低いものとなっている。FORCE11のような、未来の学術情報流通や研究の在り方を構想する抽象度の高い国際会議には、我が国からの参加は2017年に初めて1人あったのみである。

我が国においても、データ管理やオープンサイエンスに専門に関わる人材と体制を整備し、RDA等の議論に我が国からの視点を導入するとともに、オープンサイエンスの「心」である「新しいサイエンスの進め方」という視点に留意した施策の展開がなされることが望まれる。

⑸⓪　日本学術振興会　同上
⑸①　科学技術振興機構　前掲注⑴⑺
⑸②　小賀坂康志「JSTにおけるオープンサイエンスへの対応（DMP導入施行をはじめとして）」科学技術振興機構, 2017.2.14.〈https://www.nii.ac.jp/sparc/event/2016/pdf/20170214_4.pdf〉
⑸③　オープンサイエンスの取組に関する検討委員会　前掲注⑶
⑸④　「NII研究データ基盤の概要」国立情報学研究所オープンサイエンス基盤研究センターウェブサイト〈https://rcos.nii.ac.jp/service/〉
⑸⑤　オープンアクセスリポジトリ推進協会「JPCORスキーマ説明会資料」2017.10.10.〈https://jpcoar.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=46&item_no=1&page_id=46&block_id=79〉
⑸⑥　「ga088: オープンサイエンス時代の研究データ管理」JMOOC-gaccoウェブサイト〈https://lms.gacco.org/courses/course-v1:gacco+ga088+2017_11/about〉

Ⅱ　21 世紀国際学術競争の要となる「e- 研究インフラ」

国立情報学研究所准教授　船守　美穂

研究活動がサイバー空間で行われるようになってきた。欧米等では、特に欧州を中心に、「e- 研究インフラ」が研究活動の効率化と加速につながると考え、これに重点投資をする動きが鮮明になってきている。本稿は、研究活動の近年の動向とそれに伴う e- 研究インフラ整備に向けての欧米等の動きを紹介し、我が国の現状と課題について概説する。

なお、「e- 研究インフラ」の概念は、欧米では「e- インフラ」や「サイバーインフラストラクチャー」などと呼ばれている。いずれも学術における研究活動を支えるインフラとして構想され、また、歴史的にはネットワークや高速・大規模コンピューティングなどの基盤インフラとして開発・整備され、近年、個々の学問分野固有のニーズに対応する研究インフラとして注目されている。本稿では、学問分野ごとの研究活動を加速、効率化させるためのインフラであることを明確にするため、「研究」の語を付し、「e- 研究インフラ」という用語を用いる。

1　サイバー空間に移行しつつある研究活動

研究者の研究活動がサイバー空間で行われるようになっている。パソコンや電子メール、インターネットにおける情報検索なしで仕事をする研究者はいないであろうし、共同研究者等とのファイルの共有や協働作業はクラウド上のプラットフォームでなされ、データ解析や報告書の作成などもパソコン上で行われる。近年は研究の素材も、デジタルデータとして扱うことがほとんどである。自然科学系の実験や観測データなどは、実験条件などの付帯情報とともに、計測機器から自動的に出力される。それらのデータをグラフ化し、解析するツールもパッケージとして用意されている場合が多い。人文・社会科学系では、アンケート等社会調査や将来予測において数値データが扱われ、文献を中心に行う学問においても、オンライン上のデータベースが利用される。このように、あらゆる分野でデータを扱う必要が出てきたことから、科学が「理論研究」、「実験科学」、「シミュレーション科学」を経て、「データ中心科学」という新たなパラダイムに突入しているとして、2009 年には「第四のパラダイム」の概念が提唱された[57]。

サイバー空間上の多様な研究活動をサポートするツールが開発され、便利にはなっているものの、それらが必ずしも相互連携していないため、いまだ不便な点が残っている。例えば、計測機器と連結されたパソコンに自動保存される実験データを、データ解析や論文作成に用いるには、外部メモリーを利用してデータを自身のパソコンにコピーしなければならないことが多い。また、データ解析に関する共同研究者とのコミュニケーションが、データ解析等と分断された電子メール等で行われるため、時間が経つとどのような状況に対して行われた議論であったかが曖昧となるといった状況が発生する。そのほかにも様々な場面で、分断された作業をつなぐために人の時間が使われている。しかし、これら一連の研究活動が、分断されず（シームレス）にサポートされるようになれば、研究の効率化につながるだけでなく、多様なデータ連携により、より広く深い研究活動が可能になることが予想される。学術における研究活動は、アイザック・ニュートン（Sir Isaac Newton）が研究活動を「巨人の肩の上にのって洞察を行う営み」と

---

(57)　Hey et al., eds., *op.cit*. (32)

表現したように、先人の蓄積の上に新しい知見を重ねていく営みであり、研究活動の大部分がなされるサイバー空間においてこれがシームレスに実現されれば、研究活動もより着実・堅牢なものとなる。

このように、研究者の研究活動のかなりの部分がサイバー空間において行われるようになったことに鑑み、これら研究活動を加速、効率化する手段として、また、そうした研究活動の相互連携により更にダイナミックに研究を発展させる手段として、欧米では特に、e-研究インフラの開発・整備に期待が向けられている。

2　e-研究インフラ構築に向けた動き

(1) e-研究インフラ構築に向けての歴史的展開

e-研究インフラの構築は、学問分野別の学術データベースの構築と、高速・大規模コンピューティングの技術開発に始まり、徐々にこれらが融合し、総合的なe-研究インフラの構築へと発展してきている。

デジタル化が進むにつれ、まずは学問分野ごとに、共有すると有用なデータのデータベースや研究資源のカタログを整備・構築する動きが生じた。基礎物理定数や材料データ、ゲノムデータ、古典籍のデジタルアーカイブなど、「標準データ」として有意義なデータが収集され、またそれらを管理し利用に供するデータベースシステムが開発された。場合によっては、その分野特有の解析ツールなども併せて開発された。これと並行して、主にビッグサイエンスを意識した高速・大規模コンピューティングの開発も進められた。例えば、高エネルギー加速器を用いる素粒子物理学や天文・地球科学などの分野では大規模なデータを扱う必要があるため、大型計算機に始まり、グリッドコンピューティング[58]、高性能コンピューティング（High Performance Computing: HPC）[59]などの技術開発が進んだ。

こうした分野別データベースと高速・大規模コンピューティングの技術開発はそれぞれ別個に進んでいたが、徐々に融合され、研究活動全体を電子化する方向で進むようになっている。例えば、英国では1999年、当時の科学技術庁リサーチカウンシル局長（Director General of the Research Councils, UK Office of Science and Technology）であったジョン・テイラー（John Taylor）が「e-サイエンス」という政策を打ち出した[60]。これは当初、高速・大規模コンピューティングの開発・整備に軸足があったが、徐々にICT技術で高度化されたサイエンス全般を対象とするようになり、またインターネット上での研究者間の協働作業なども概念として含めるようになっていった。同様の概念として、米国では国立科学財団（National Science Foundation: NSF）の委員会が2003年にまとめた報告書[61]に基づき「サイバーインフラストラクチャー」として、豪州では2006年に取りまとめられた報告書[62]に基づき「e-リサーチ」として、政策展開がなされている。

58　複数のコンピュータに処理を分散させて行う計算処理。
59　単位時間当たりの計算量が非常に多い計算処理。
60　Tony Hey and Anne E. Trefethen, “The UK e-Science Core Programme and the Grid,” *Future Generation Computer Systems*,Vol.18 Issue 8, 2002, pp.1017-1031.〈http://users.ecs.soton.ac.uk/ajgh/FGCSPaper.pdf〉
61　“Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure,” January 2003.〈https://www.nsf.gov/cise/sci/reports/atkins.pdf〉
62　e-Research Coordinating Committee, “An Australian e-Research Strategy and Implementation Framework,” April 2006; Philip Marcus Clark AM et al., “Research Infrastructure Review: Final Report,” September 2015.〈https://docs.education.gov.au/system/files/doc/other/research_infrastructure_review.pdf〉

欧州では、こうしたICT技術により研究活動が新たな次元に移行するという考えの下、「サイエンス2.0」[63]が提唱され、これが2014年の欧州委員会によるパブリックコメントの募集を経て、研究成果や研究活動そのものを開かれたものとするという概念も加わって、「オープンサイエンス」へと変化していった。欧州委員会は、2015年から「欧州オープンサイエンスクラウド」（European Open Science Cloud: EOSC）を、オープンサイエンスの理念を実現するためのインフラとして構想している[64]。なおEOSCは、後述のe-IRGが提唱した「e- インフラ・コモンズ」のビジョンを受け継いでおり、欧州域内の多数のe- 研究インフラを有機的に連携する「システムのシステム（system of systems）」として検討されている[65]。EOSCは、トップダウンで構想されており、既存の研究インフラや研究活動の現状が十分に踏まえられていないとの批判があるものの、オープンサイエンスの実現に向けて、EOSCの整備に大がかりな研究開発投資が進められつつあることは見逃せない。

（2）e- 研究インフラを推進する主体

研究インフラ欧州戦略フォーラム（European Strategic Forum for Research Infrastructures: ESFRI. 2002年設立）は、欧州域内において戦略的に整備・運営していくべき研究インフラ[66]について欧州理事会（European Council）に助言をする組織である[67]。EU加盟各国と関係国が共同投資・運営する研究インフラについて、ほぼ実質的な決定権を持つ極めて影響力のある組織とされ、学術に関わる多くのステークホルダーがESFRIの動きを注視している。ESFRIは2016年に、戦略的に整備・運営すべき研究インフラの選定手順や体制等を明確にし、研究インフラの選定基準の4項目のうちの一つに、「e- ニーズ」を定めた（表4）。「e- ニーズ」では、提案されている研究インフラにどのような「e- インフラ」が想定されているのか、その「e- インフラ」が既存の国別・学問分野別の「e- インフラ」とどのように相互連携し、欧州全体のe- 研究インフラの発展にどのように寄与するかなどが、ESFRIの審査委員会により確認される[68]。ESFRIが、投資すべき研究インフラの選定にe- ニーズの視点を持ち込んだことは注目を集めている[69]。なおESFRIおよび後述のe-IRGは、「e- インフラ」という表現を使っているため、ここではその表現を用いたが、これは本稿全体に関わる「e- 研究インフラ」と同義である。

63　第Ⅰ章の「2　欧米におけるオープンサイエンスに向けた考え方」を参照。

64　European Commission, “Realising the European Open Science Cloud: first report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud,” 2016.〈https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf〉

65　“EOSC Architecture: a System of Systems.” EOSCpilot Website〈https://eoscpilot.eu/content/eosc-architecture-system-systems〉

66　ここでESFRIが対象としている「研究インフラ」は、一般には物理的な研究施設・設備であり、サイバー空間におけるe- インフラではない。ESFRIではこうした物理的研究インフラについて、サイバー空間に形成されるe- インフラとしてのビジョンを全てのESFRI対象研究施設に求めたことが、注目に値する。

67　“About EFSRI.” European Strategy Forum on Research Infrastructures Website〈http://www.esfri.eu/about-esfri〉

68　E-IRG, “Guide to e-Infrastructure Requirements for European Research Infrastructures: An e-ING support document,” March 1, 2017, pp.9-10.〈http://www.hnscicloud.eu/sites/default/files/2017-Supportdocument.pdf〉

69　Keith Jeffery, “The ESFRI Roadmap and its Demands on the E-Infrastructure,” *Data Science Journal,* Volume 9, 2010; Linn Hoff Jensen, “The emerging “e” component of research infrastructure,” 2016.10.2. NordForsk Website〈https://www.nordforsk.org/en/news/the-emerging-201ce201d-component-of-research-infrastructure〉

表4　欧州において戦略的に整備すべき研究インフラのESFRI選定基準

■科学的卓越性
・長期的科学プログラムが明確である。
・科学コミュニティが確立している。
・科学におけるリーダーシップが明確である。
■汎欧州としての関連性（注1）
・当該科学領域における汎欧州のアプローチが明確である。
・対象となるユーザ・コミュニティが汎欧州である。
・各国又は国際的な施設が相互補完する、又はシナジー効果を有する。
■社会・経済的効果
・社会的課題との関係性が明確にされ、経済的効果が予測されている。
■e-ニーズ
・アクセス方針、セキュリティ対策を含む、e-インフラの要件に対するビジョンができている。
・コミュニケーションネットワークや分散コンピューティング、HPC/HTC（注2）との連携がある。

（注1）pan-European Relevance. ESFRIは、欧州の複数の国で共同投資・運営する研究インフラについて検討を行うため、各国のニーズ以上に、当該研究インフラが欧州全体として重要であるかの視点が重視される。
（注2）HPC（High Performance Computing）, HTC（High Throughput Computing）のいずれも、複数のコンピューティング・リソースを用いた大規模な計算処理を目的とするが、前者が短時間の高速処理を追求するのに対して、後者は数か月から数年をかけて、相互に緩やかに関係する演算処理を複数実行することを追求する。
（出典）ESFRI, “Annex II: List of minimal key requirements for scientific case,” *Public Roadmap 2018 GUIDE: Final version dated 9th December 2016*, pp.22-23.〈http://www.esfri.eu/sites/default/files/docs/ESFRI_Roadmap_2018_Public_Guide_f.pdf〉を基に筆者作成。出典には「設計面」（Design）以外に、「準備段階」（Preparation）、「実施段階」（Implementation）、「運用段階」（Operation）について具体的な最低要件を明記しているが、ここでは紙面の制約により割愛した。

「e-インフラ検討グループ」（e-Infrastructure Reflection Group: e-IRG）は、EU加盟国および関係国からなる、e-インフラの統合とサービス連携を促進するための戦略的会議体で、e-インフラに関わる専門組織として機能している。e-IRG加盟国のe-インフラに関わる政策担当者[70]と、各国において学術情報やインターネットなどのサービスを学術界に提供するe-インフラの運営主体[71]が各国を代表して集まり、欧州域内のe-インフラの将来的な発展・維持について検討を行い、指針やロードマップ等として取りまとめている。e-IRGが打ち出す指針等に、各国の政策形成に対する特段の強制力はないが、各国の政策担当者がe-IRGの指針等の策定に関わっているため、各国のe-インフラ整備計画に影響を与える場合が多い。過去にはネットワークや高速・大規模コンピューティングなどの基盤的なe-インフラを検討対象としていたが、近年は、分野ごとに構築されるe-インフラなども含め、包括的なビジョンでの検討がなされるようになっている。

e-IRGは、近年e-研究インフラの重要性が増していることに鑑み、ESFRIや欧州委員会などの機関から、協力要請されるようになっている。ESFRIからは、欧州におけるe-研究インフラの整備の在り方に関する検討の委託を受け、報告書を2010年に発表した[72]。表4の、ESFRIが課すe-ニーズに関する要件も、e-IRGからの助言に基づいている。同時に、e-IRGが

⑺⓪　“e-IRG delegates and observers.” e-IRG Website〈http://e-irg.eu/delegates〉
⑺①　学術界には大容量のデータや高速の計算処理等を行うニーズがあるため、各国とも学術界に特化したインターネットのサービスプロバイダがあり（National research and education network: NREN）、これらがより上位レイヤーの教育・研究活動に直結するサービスも提供している。我が国では学術情報ネットワークSINET5を提供する国立情報学研究所が我が国のNRENであり、我が国の論文検索サイトCiNiiや機関リポジトリのクラウドサービスJAIRO Cloud、科学研究費助成事業データベースKAKENなどが提供されている。現在は日本のe-研究インフラとなる研究データ基盤を構築中（本章第3節参照）。
⑺②　e-IRG, *e-IRG “Blue Paper”* 2010.〈http://e-irg.eu/documents/10920/238805/e-irg_blue_paper_2010〉

2012 年の e-IRG ロードマップ[73] において提唱した、各国に分散した e- リソースを欧州域内で共有して利用するという「e- インフラ・コモンズ」という概念[74] は、その後の欧州委員会における EOSC の提唱へとつながった。

各国レベルでも、e- 研究インフラについて様々な検討がなされている。例えば英国では、英国研究会議協議会（Research Councils UK: RCUK）が「e- インフラ・ロードマップ」[75] を発表し、ここでも様々な e- インフラを融合的に扱うことの重要性が指摘されている。

ドイツでは、「情報インフラ・カウンシル」（Rat für Informations Infrastrukturen: RfII）が 2014 年 11 月に設置された。ここでは科学者ユーザ・コミュニティ、市民、e- インフラ運営主体、連邦政府・州政府からそれぞれ複数名がメンバーとしてカウンシルを形成し、ドイツにおける情報インフラの在り方について検討を行っている[76]。

豪州では、連邦政府の主導のもと、e- リサーチ調整委員会（e-Research Coordinating Committee）が「e- リサーチ戦略と実施枠組み」[77] を 2006 年に発表し、これを基に豪州国立データサービス（Australian National Data Service: ANDS）[78] が設立された。ANDS は、豪州で生成された研究データを研究者や研究機関、国に対してより価値あるものとするという使命を有し、「Research Data Australia」[79] という、豪州で生成された研究データを検索・アクセス・再利用できるポータルサイトの運営を行っている。また、豪州政府が発表した「2016 年豪州国家研究インフラ・ロードマップ」は、研究活動がデータ及び e- 研究インフラに依存しているという認識の下、統合された明確かつ信頼性のあるシステムとして「豪州研究データクラウド」（Australian Research Data Cloud）を構築し、データ集約型研究、分野横断型研究、グローバルな共同研究といったニーズに応えるとしている[80]。

米国においても、全米データサービス（National Data Service: NDS）が、研究者にデータ発見、利用、共有・公開を可能とし、データと文献をリンクするサービスの開発を開始している[81]。NDS は、特定の政府機関や研究助成機関により設立されたものではなく、分野にかかわらず、研究データを容易に発見、利用、公開できるプラットフォームが米国においても必要であるという認識を持つ、大学やサイバーインフラストラクチャー[82]、データを多く扱うプロジェクトなどの関係者が集まった[83]、コミュニティ主導の取組であるというところに特徴がある。これらの NDS パートナー機関は、NDS コンソーシアムを形成し、全米データサービスに必要なサービスのデザインや資金集め、サービスの開発・整備などを行う。

（3）e- 研究インフラの具体事例

ここまでで紹介したように、e- 研究インフラは、各分野、各階層で独自に発展した様々なツー

⑺③　e-IRG, “e-IRG Roadmap 2012,” 2012.〈http://e-irg.eu/documents/10920/12353/e-irg_roadmap_2012-final.pdf〉
⑺④　e-IRG, “e-Infrastructure Commons.”〈http://e-irg.eu/documents/10920/290578/e-Infrastructure+Commons+summary.pdf〉
⑺⑤　Research Council UK, “E-Infrastructure Roadmap.”〈http://www.rcuk.ac.uk/documents/documents/roadmapforelc-pdf/〉
⑺⑥　Rat für Informations Infrastrukturen, “2015 Opening Declaration,” June 2015.〈http://www.rfii.de/?wpdmdl=2048〉
⑺⑦　e-Research Coordinating Committee, *op.cit.* ⑹②
⑺⑧　Australian National Data Service Website〈https://www.ands.org.au/〉
⑺⑨　Research Data Australia Website〈http://researchdata.ands.org.au/〉
⑻⓪　“2016 National Research Infrastructure Roadmap.” Australian Government of Department of Education and Training Website〈https://www.education.gov.au/2016-national-research-infrastructure-roadmap〉
⑻①　National Data Service Website〈http://www.nationaldataservice.org/〉
⑻②　米国では「e- 研究インフラ」のことを、「サイバーインフラストラクチャー」と呼んでいる。
⑻③　“Partners.” National Data Service Website〈http://www.nationaldataservice.org/about/partners.html〉

ルやプラットフォームを連携させ、その中でユーザが自由にデータの移動や解析を行うことを想定している。具体的には、複数のe-研究インフラを連携させることが想定されており、そのためにはまず、分野別のニーズに応じたe-研究インフラの構築が進められている。

例えば、欧州の「NFFA」（Nanoscience Foundries & Fine Analysis）[84]というe-研究インフラは、材料科学分野におけるナノ材料合成のコミュニティと精密計測・解析のコミュニティとを融合させるために欧州で構築された。ここでは材料合成の研究者が、自ら合成した材料の物性を精密計測・解析するため、欧州域内に点在するX線回折装置、核磁気共鳴装置（NMR）、質量分析装置、電子顕微鏡、微細加工装置などを利用する。NFFAは、これらの機器から取得されたデータ（実験条件等の情報も含む。）を自動的に取り込む。研究者は、NFFAに用意されたツールでデータの管理や解析を容易に実施できる。このように、NFFAにはe-研究インフラとして、物理的な研究インフラをより効率的・効果的に利用するという視点がある。なお、実験実施から一定期間が経過すると、これらのデータは公開され、他の研究者により再利用される仕組みとなっている。

我が国では、計測機器は大学共同利用施設や大学等で提供され、通常、研究者はその計測値を外部メモリーに保存し、自身のパソコン上で解析する。複数の計測機器からのデータを統合・解析する作業や、投稿した論文の根拠データを公開する作業も研究者が自ら行う必要がある。しかし、NFFAのようなe-研究インフラがあれば、計測データの取得・管理・解析・公開の一連の流れがシームレスに行われ、研究プロセスが効率化される。

欧州・地中海植物保護機関（European and Mediterranean Plant Protection Organization: EPPO）[85]は、植物保護に関する各種基準や標準のガイドラインを各国と調整の上、策定する国際機関である。ガイドラインの基準等を決めるためには、対象地域の植生や害虫、天候等の情報が必要であり、EPPOはこれら情報が集約されるデータベースを運営している。こうした植生等の情報は、対象51か国4,500名の研究者からの報告に基づいており、研究者がインターネット上のプラットフォームにデータを直接入力している。また、このインターネット上のサイトは、植生等について、他地点のデータ等との比較・解析を容易にできる機能も有している。

人文・社会科学系の学問分野においても、このようなe-研究インフラの整備が進みつつある。人文科学系では、デジタル・ヒューマニティーズの流れを背景にEUレベルで「芸術・人文諸科学のための電子研究インフラ構築プロジェクト」（Digital Research Infrastructure for the Arts and Humanities: DARIAH）[86]というプロジェクトが精力的に進められている。EUレベルのプロジェクト下に各国のDARIAHプロジェクトがあり、それぞれ重点領域は異なる。例えば、ドイツはe-研究インフラ提供に力を入れ、デジタル・ヒューマニティーズのための解析ツールを多数開発・提供している[87]。具体的には、①あらゆる人文系コンテンツを時間軸・空間軸に沿って地図上にマッピングできる「Geo-Browser」というツール、②中世の文書のページレイアウトを自動解析できるツール、③文書の転記がなされる過程でどのような変化やミスが生じていったか、どれがオリジナル文書であるかなどを解析するツールが挙げられる。社会科学に

84　Nanoscience Foundries & Fine Analysis Website〈http://www.nffa.eu/〉
85　European and Mediterranean Plant Protection Organization Website〈https://www.eppo.int/〉
86　DARIAH-EU Website〈https://www.dariah.eu/〉
87　“DARIAH-DE in Kürze.” DARIAH-DE Website〈https://de.dariah.eu/dariah-de-in-kurze〉

おいては、社会科学系データアーカイブ[88]が各国で歴史的に形成されているが、これらを統合的に検索できる仕組みを欧州社会科学データアーカイブ協議会（Consortium of European Social Science Data Archives: CESSDA）が用意しようとしている。

表 5 に、e-IRG が ESFRI からの委託を受けて選定した、e- 研究インフラに関する ESFRI プロジェクトを挙げる。表 5 の上段は 2010 年に選定したプロジェクト、下段は研究データ管理（research data management: RDM）に向けての関心が世界的に高まったことを受けて[89]、2012 年に選定した、特にデータ管理を必要とするプロジェクトである。CESSDA 及び DARIAH も、この表に含まれている。

表５　e-IRG が選定した e- 研究インフラの色彩の強い ESFRI プロジェクト

プロジェクト	分野
e- 研究インフラの色彩の強い ESFRI 対象プロジェクト（2010 年選定）	
・BBMRI（Biobanking and Biomolecular Resources Research Infrastructure）	生体試料
・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）	超長基線電波干渉法（天体観測）
・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）	超巨大電波望遠鏡
・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する ESFRI 対象プロジェクト（2012 年選定）	
・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

（出典）e-IRG, *e-IRG "Blue Paper" 2010*, pp.35-54.〈http://e-irg.eu/documents/10920/238805/e-irg_blue_paper_2010〉; e-IRG, "e-IRG "Blue Paper" on Data Management," 30 October 2012, p.4.〈http://e-irg.eu/documents/10920/238805/e-irg-blue_paper_on_data_management_v_final.pdf〉を基に筆者作成。

88　社会科学系データアーカイブ：社会科学の分野（法学、経済学、政治学、社会学、教育学等あらゆる社会科学の分野を含む。）ではデータアーカイブが歴史的に形成されている国が多い。政府統計や社会科学分野の大規模調査等のデータ（個票を含む。）をアーカイブし、利用に供するものである。規模が大きいものとして、米国の ICPSR〈https://www.icpsr.umich.edu/icpsrweb/〉、英国の UK Data Archive〈http://www.data-archive.ac.uk/〉、ドイツの GESIS〈https://www.gesis.org/en/home/〉などがある。我が国では東京大学社会科学研究所附属社会調査・データアーカイブ研究センターが SSJDA を運用している。「データアーカイブをめぐる我が国の動き」東京大学社会科学研究所附属社会調査・データアーカイブ研究センターウェブサイト〈http://csrda.iss.u-tokyo.ac.jp/ssjda/scene/〉

89　2012-2013 年にかけて、英国王立協会（the Royal Society）から "Science as an open enterprise," June 2012 が , 米国大統領府科学技術政策局（OSTP）から公的助成研究の成果物やデータに関するパブリックアクセス拡大のための指令が、G8 科学技術大臣により公的助成を得た研究データを公開していく方針に関する共同声明（*op.cit.*(1)）などが発せられた。

3　e-研究インフラの構築に関わる我が国の現状と課題及び今後の展望

我が国においても、e-研究インフラを整備するという発想が全くないわけではない。例えば日本学術会議では、2008（平成20）年、2014（平成26）年、2016（平成28）年にそれぞれ、「E-サイエンス分科会」、「国際サイエンスデータ分科会」、「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、それぞれ提言が取りまとめられた[90]。これらの提言は海外動向を踏まえたものであり、海外と同様の問題認識を有している。

前述のように、現在、国立情報学研究所では、日本学術会議のオープンサイエンスに関わる報告に基づき、我が国の研究者が生成する研究データを管理・公開できる基盤インフラを開発している[91]。これにより研究分野にかかわらず、研究データを保存したり、各種ファイルを共同研究者と共有したりでき、研究プロジェクト終了時には、研究論文やそれに用いたデータ、プログラム等が研究資源として公開される。他の研究者は公開された研究資源を検索することができ、新しい研究シーズの発見へと繋がることが想定されている。

同時に、本章の「2(3)　e-研究インフラの具体事例」に挙げたような、学問分野ごとのニーズに基づいたe-研究インフラも整備されつつある。例えば、地球規模の課題に対して自然科学のデータや社会経済情報等を融合して解析可能とする「データ統合・解析システム」（Data Integration and Analysis System: DIAS）[92]は、オンライン・プラットフォーム上でデータの統合や解析を可能とするe-研究インフラとしての機能を体現する。

しかし、国内のe-研究インフラの開発・運営担当者がRDA（第1章第2節第2項を参照。）の会議に参加することは少なく、活動が個々のe-研究インフラの開発・整備にとどまっており、国際連携による効果的な研究データ管理に向けた手法開発や基準の策定など国際標準への関与が十分にできていない様子がうかがわれる。日本学術会議で数年ごとに策定される、学術の大型施設および大規模研究に関わる「学術の大型研究計画に関するマスタープラン」でも、ESFRIが言及するような、大型研究インフラの効果を最大化するためのe-研究インフラの必要性については言及されておらず、e-研究インフラに積極的に投資を行う欧州等との温度差が見られる。

欧州は、e-研究インフラに対して域内を横断する仕組みとしての需要があるため、取組が進む素地がある。データの重要性が増し、システムとシステムの間の有機的な連携が求められるデジタル・ネットワーク時代において、複数の国が一つの地域として活動する必要性が、有利に働いているといえる。同様の活動を、欧州のような地域統合体の仕組みを有さない我が国又はアジアに求めるのは難しいのかもしれないが、少なくとも、e-研究インフラの整備が研究活動の加速と効率化につながるという視点と、これへの取組が世界で加速していることの認識を持ち、対応を強化することが求められる。

⑼⓪　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. 〈http://www.scj.go.jp/ja/member/iinkai/kiroku/3-0822.pdf〉, 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「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―」2014. 〈http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-h140930-3.pdf〉, 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「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. 〈http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t230.pdf〉

⑼①　「NII研究データ基盤の概要」　前掲注⑸④

⑼②　DIASデータ統合・解析システムウェブサイト〈http://www.diasjp.net/〉

Ⅲ　人材育成（データサイエンティスト育成等）

統計数理研究所特任准教授　神谷　直樹

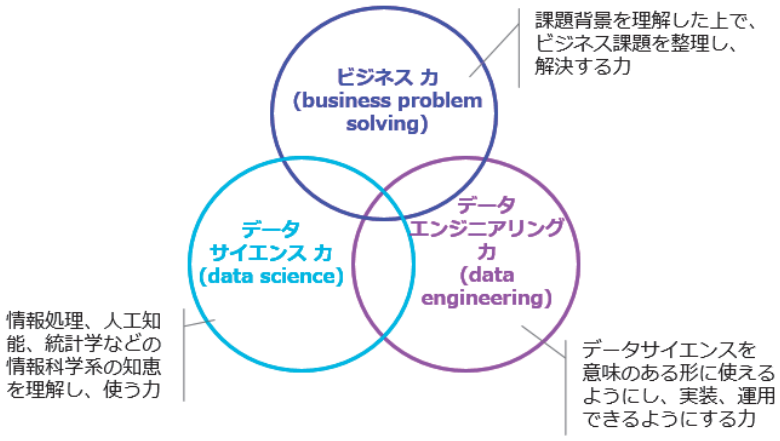
過去 50 年にわたって情報技術が指数関数的に発展してきたことにより、私たちは複雑で巨大なデータ集合の集積、いわゆるビッグデータを手にすることができるようになった。データサイエンティスト（Data Scientist）とは、このようなデータから新しい価値を引き出すプロフェッショナルのことである。データサイエンティストという新しい職種は、ビジネス誌『ハーバード・ビジネス・レビュー』（Harvard Business Review）の 2012 年 10 月号（ビッグデータの特集）で「データサイエンティストほど素敵な仕事はない」(93) という記事が掲載されて以降、注目を集めるようになった。以下ではこのデータサイエンティストに焦点を当て、その育成についてまとめる。

1　データサイエンティストの定義

(1) スキルセットに基づく定義

データサイエンティストには様々な分類が提案されている (94) が、求められている主要スキルの組合せ（スキルセット）(95) は、2010 年以降、現在まで変化していない。データサイエンティストには、ビジネス関連スキル、データサイエンス・スキル（機械学習／ビッグデータ、数学／オペレーションズ・リサーチ (96)、統計学等）、エンジニアリング・スキル（プログラミング等）の三つのスキルが求められる (97)。我が国では、2014 年 12 月、データサイエンティスト協会がスキルセットに基づくデータサイエンティストの定義を提示している（図 1）。

図 1　データサイエンティスト協会によるデータサイエンティストのスキル定義



（出典）データサイエンティスト協会「データサイエンティスト協会、データサイエンティストのミッション、スキルセット、定義、スキルレベルを発表」2014.12.10.〈https://prtimes.jp/a/?c=7312&r=5&f=d7312-20141210-6604.pdf〉

(93) Thomas H. Davenport and Dhanurjay Patil, “Data scientist: The sexiest job of the 21st century,” *Harvard Business Review,* Vol.90 No.10. 2012.10, pp.70-76.

(94) Harlan D. Harris et al., *Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work,* California: O’Reilly Media, 2013, pp.9-18; Nicolaus Henke et al., “The Age of Analytics: Competing in a Data-driven World,” 2016.12. McKinsey & Company Website〈https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world〉

(95) Drew Conway, “The Data Science Venn Diagram,” 2010.9.30. Drew Conway Data Consulting Website〈http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram〉

(96) 限られた資源を有効に利用して目的を最大限達成する意思決定を数学的・科学的に実現する手法を研究する分野。

(97) Dhanurjay Patil and Hilary Mason, *Data Driven: Creating a Data Culture,* California: O’Reilly Media, 2015, pp.2-4.

（2）キャリア類型に基づく定義

欧米と我が国では、データサイエンティストという同じ名称の職種であっても、どのようなワークスタイルで、データからどのような価値を引き出すかについて違いがある。我が国では一つの企業内でデータ分析チームを作って、社内外に対して新たなサービスを提案することが多い[98]。オライリー・メディア（O'Reilly Media）社がプロフェッショナルに行った調査[99]と同等の質問項目を含めた調査を統計数理研究所が我が国の統計検定受験者に行い回答者をクラスター分析したところ、キャリア類型に基づく分類が抽出されている（表6）。

表6　キャリア類型に基づくデータサイエンティストの分類

分類	特徴
メーカーの製品開発・企画部門に勤める中堅のIT系エンジニア	社内では確実にデータの活用が進んでいる。キャリアパスも見えている。
主に中小のサービス系の企業に勤める女性	比較的自由になる勤務形態を望んでいる。
ITサービス業のプロフェッショナル	長年データ分析を実施してきて、仕事に誇りを持っている。
実務経験の少ない若手	データサイエンティストとして活躍する夢を持っている。

（出典）情報・システム研究機構統計数理研究所「文部科学省委託事業　データサイエンティスト育成ネットワークの形成―平成25年度事業報告書―」2014.3. pp.10-20.〈http://www.ism.ac.jp/shikoin/training/dstn/pdf/H25DSTN.pdf〉を基に筆者作成。

（3）T型データサイエンティスト

現実的には、一個人が全てのスキルを備えるという想定に困難があることや、特に我が国では特にデータ分析チームとして業務を行うことが多いことから、T型データサイエンティスト[100]という発想が生まれている。Tの横線は幅広い知識・スキルを表し、縦線はある一つの分野に関する深い知識･スキルを表している。一つの分野に関する深い知識･スキルには、データサイエンス、統計学、ビジネス・コミュニケーション等が考えられている[101]。

2　人材育成事業の種類と対象水準

2015（平成27）年、情報・システム研究機構の「ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会」は、データサイエンティストのスキルレベルと育成対象人数を整理した（図2）。これは、データサイエンティスト育成に対する産業界とアカデミアからの要請を踏まえて､我が国におけるデータサイエンティスト育成の「あるべき姿」をまとめたものである。

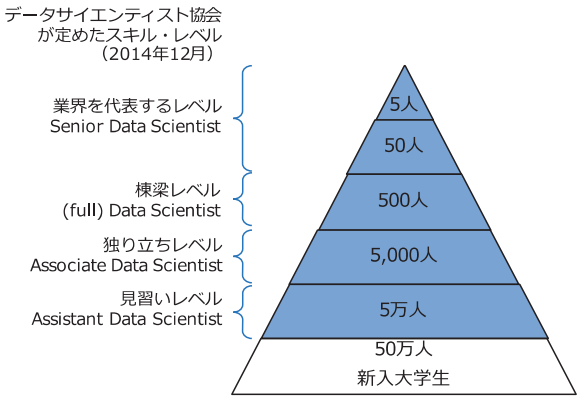
98　情報・システム研究機構統計数理研究所「文部科学省委託事業 データサイエンティスト育成ネットワークの形成―平成25年度事業報告書―」2014.3.31.〈http://www.ism.ac.jp/shikoin/training/dstn/pdf/H25DSTN.pdf〉

99　Harris et al., *op.cit*. 94

100　*ibid*., p.19.

101　統計数理研究所では、Tの横線をデータサイエンス、縦線をデータサイエンスが関わる様々なドメインとするT型人材の育成に2011（平成23）年から取り組んでいる。

図２　データサイエンティスト育成レベルと毎年の育成目標人数



（出典）情報・システム研究機構ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会「ビッグデータの利活用のための専門人材育成について」2015.7.30, p.6.〈http://www.rois.ac.jp/open/pdf/bd_houkokusho.pdf〉

同懇談会による提言書「ビッグデータの利活用のための専門人材育成について」でまとめられたスキルレベルは、次のように要約できる。「見習いレベル」は、全てのデータサイエンティストが持つべき共通スキルレベルである。我が国の理系修士入学者数（約 5 万人）が目標育成人数の目安である。「独り立ちレベル」は、ビジネス、データサイエンス、エンジニアリングのいずれかの分野で専門的な能力を持ち、自らのイニシアチブによる高度なデータ分析・問題解決能力を持ち得るスキルレベルである。これには修士課程修了者（博士課程入学者）が想定され、我が国における資本金 10 億円以上の会社（約 6,000 社）で毎年 1 人程度ずつ採用を検討するとすれば、目標育成人数の目安は 5,000 人程度である。「棟梁（とうりょう）レベル」は、データサイエンティストのチームを率いて組織におけるビッグデータ利活用を先導できるスキルレベルである。複数の応用分野を俯瞰（ふかん）的にマネジメントすることができ、データサイエンスの観点から最適戦略を策定し実行するリーダーシップが求められる。このようなスキルは主に実務を通して育成される。5,000 人の独り立ちレベルを指導統括するリーダーは、組織論の観点から独り立ちレベル 6 ～ 15 人につき 1 人程度必要で、目標育成人数は 500 人程度である。「業界を代表するレベル」（指導的データサイエンティスト）は、アカデミアにおいてはデータサイエンスの最先端を切開くワールドクラスの研究者・開発者として指導的な能力を発揮する人材、産業界においてはデータサイエンスに基づくイノベーションをけん引できる人材である。年間数人～数十人が現実的な目標育成人数である[(102)]。

以下では、このスキルレベルを参照基準として、海外と我が国の人材育成事業の種類と対象水準をまとめる。

---

(102)　情報・システム研究機構ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会「ビッグデータの利活用のための専門人材育成について」2015.7.30, pp.7-8.〈http://www.rois.ac.jp/open/pdf/bd_houkokusho.pdf〉

（1）海外の人材育成事業の種類と対象水準

（i）大学教育プログラム

学位や修了証明書が取得可能な海外の主要各国におけるデータサイエンス教育プログラムについて、ウェブ公開情報を基に主要各国の教育プログラム数をまとめた（表7）。この傾向を反映して、米国における統計学や生物統計学関連の学位取得者は急増している[103]。また、海外の学部・大学院教育における動向に関して次のような知見がある[104]。

①海外の大学には、もともと元々統計学の独立した学部・学科が存在する。

②独立した統計学の学部･学科があると、情報工学、コンピュータ科学、時には数学も加えて、データサイエンス・プログラムが柔軟に生成されやすい。

③米国の学部教育では、統計学専攻のプログラムがデータサイエンスを意識した内容に変更されており、統計学専攻は、近年の理系で最も人気があり卒業後の平均給与も最も高い。

表7　海外のデータサイエンス教育プログラム数（2017年9月時点）

	米国	英国	スペイン	ｱｲﾙﾗﾝﾄﾞ	フランス	オランダ	その他	計
学士	38	5	1	1	1	0	4	50
修士	287	40	8	7	6	7	43	398
博士	18	1	0	0	0	0	2	21
修了証明	94	0	0	1	0	0	1	96
計	437	46	9	9	7	7	50	565

（出典）ウェブ公開情報を基に著者作成。

（ii）民間企業や関連学協会の育成プログラム

海外の民間企業による育成プログラムは、修了後の雇用と結びついていることが多く、人材を求める民間企業の出資により無料で参加できるプログラムもある[105]。

一方、育成プログラムを実施している関連学協会は、以下のとおりである。これらの育成プログラムでは、水準を満たした参加者に修了証明書が発行される。なお、③データサイエンス･セントラルのアプレンティスシップ･プログラム（Data Science Apprenticeship. 見習い実習）は、無料である[106]。

①デジタル・アナリティクス協会（Digital Analytics Association）

②アメリカ統計学会（American Statistical Association）

③データサイエンス・セントラル（Data Science Central: Data Science Apprenticeship）

④統計教育研究所（Institute for Statistics Education）

103　Steve Pierson, “Statistics, Biostatistics Degree Growth Sustained Through 2015,” Amstatnews Website, 2016.10.1.〈http://magazine.amstat.org/blog/2016/10/01/science-policy/〉

104　情報・システム研究機構統計数理研究所「文部科学省委託事業「データサイエンティスト育成ネットワークの形成」平成27年度事業報告書」2016.3. p.25.〈http://www.ism.ac.jp/shikoin/training/dstn/pdf/H27DSTN.pdf〉; 竹村彰通「滋賀大学のデータサイエンス学部創設について」2016.3.7.〈http://www.ism.ac.jp/shikoin/training/dstn/pdf/20160307-Prof Takemura.pdf〉

105　例えば、インサイト・データサイエンスフェローズ・プログラム（Insight Data Science Fellows Program）。“Insight Data Science Fellows Program.” Insight Science Data Website〈http://insightdatascience.com/〉

106　Vincent Granville, “Fast-Track, On-Demand, No-Fee Program to Become a Data Scientist,” 2015.6.12. Data Science Central Website〈https://www.datasciencecentral.com/profiles/blogs/free-on-demand-data-science-program-to-quickly-become-a-data-scie〉

（iii）無料の育成プログラム

その他の育成プログラムで無料のものとしては、大規模公開オンライン講座（Massive Open Online Courses: MOOCs）によるオンライン学習がある（修了証明書が必要な場合は有料となるが受講自体は無料）。2017 年 5 月時点で、「coursera」[107]、「edX」[108]、「UDACITY」[109] といった代表的な MOOCs で修了証明を取得できるコースは 25 コースある（coursera、edX はそれぞれ 9 コース、UDACITY は 7 コース）。これらの目標育成レベルは「独り立ちレベル」以下が想定されている。なお、UDACITY の育成プログラムは就職に結びついた場合に料金が払い戻されることから、職業訓練に近い位置付けと考えられる。

（2）我が国の人材育成事業の種類と対象水準

（i）大学教育プログラム

平成 29 年 4 月に我が国最初のデータサイエンスを中核とする学部として、滋賀大学にデータサイエンス学部が設置された。平成 30 年 4 月には横浜市立大学にもデータサイエンス学部が設置される予定である。このほか、いくつかの大学では学部・大学院とは別にデータサイエンスの教育プログラムを設置したり、履修パターンの一つにデータサイエンスを中心としたカリキュラムを提示したりするなど様々な取組がなされている。しかしながら、多くの学生は、所属する学部・大学院で各分野に特化したデータサイエンスを学んでいる。

平成 28 年 12 月、文部科学省は「数理及びデータサイエンスに係る教育強化」拠点大学として、6 大学（北海道大学、東京大学、滋賀大学、京都大学、大阪大学及び九州大学）を選定した[110]。この事業は、専門分野の枠を超えた全学的な数理・データサイエンス教育機能を有するセンターを整備し、専門人材の専門性強化と他分野への応用展開の実現及びそれらの相乗効果の創出を目的としている[111]。

（ii）民間企業や関連学協会の育成プログラム

我が国の民間企業による育成プログラムや育成講座は、修了後の雇用と直接結びついていない。平成 26 年 12 月時点で 146 講座[112]、平成 27 年時点で 151 講座[113] が確認されている。

一方、育成プログラムや育成講座を実施している我が国の関連学協会は次のとおりである。これらの中には、例えば、③一般社団法人ウェブ解析士協会の育成講座のように独自資格の認定と連動しているものがある。

①一般社団法人データサイエンティスト協会
②一般財団法人実務教育研究所
③一般社団法人ウェブ解析士協会
④一般社団法人日本マーケティング・リサーチ協会

---

[107] coursera Website〈https://www.coursera.org/〉
[108] edX Website〈https://www.edx.org/〉
[109] UDACITY Website〈https://www.udacity.com/〉
[110] 数理及びデータサイエンス教育の強化に関する懇談会「「数理及びデータサイエンスに係る教育強化」の拠点校の選定について」2016.12.21. 文部科学省ウェブサイト〈http://www.mext.go.jp/b_menu/shingi/chousa/koutou/080/gaiyou/1380792.htm〉
[111] 数理及びデータサイエンス教育の強化に関する懇談会「大学の数理・データサイエンス教育強化方策について」2016.12.〈http://www.mext.go.jp/b_menu/shingi/chousa/koutou/080/gaiyou/__icsFiles/afieldfile/2016/12/21/1380788_01.pdf〉
[112] 情報・システム研究機構統計数理研究所「文部科学省委託事業「データサイエンティスト育成ネットワークの形成」平成 26 年度事業報告書」2015.3. pp.22-50.〈http://www.ism.ac.jp/shikoin/training/dstn/pdf/H26DSTN.pdf〉
[113] 同上, pp.81-123.

（iii）無料の育成プログラム

我が国のMOOCsとしては、株式会社ドコモgaccoが運営する「gacco」がある。データサイエンス関連では6講義が実施されている（日本統計学会と日本計量生物学会の協力の下で作成された3講義と、総務省統計局による3講義）[114]。海外のMOOCsとは異なり講義単位で修了証が発行される。育成目標レベルは独り立ちレベル以下である。

このほか、総務省統計局は「データサイエンス・スクール」という社会人向けのオンライン講座を提供している[115]。この講座の育成目標レベルも独り立ちレベル以下である。

文部科学省も幾つかの育成事業を展開している。大阪大学が中核拠点になっている「成長分野を支える情報技術人材の育成拠点の形成」（enPiT）では、同事業の連携大学に所属する大学生・大学院生が受講できるグループワークを用いた短期集中合宿や分散PBL（Project-Based Learning）[116]を実施している[117]。対象分野は、クラウドコンピューティング[118]、セキュリティ、組み込みシステム[119]、ビジネスアプリケーションの4分野である。また、平成29年からは、データ関連技術（AI、IoT、ビッグデータ、セキュリティ等）を高度に駆使する人材（高度データ関連人材）の発掘・育成・活躍促進を行う企業や大学等における取組を支援する「データ関連人材育成プログラム」が始まった[120]。

3　人材充足状況

人材充足状況の参考資料となるデータがマッキンゼー・グローバル・インスティチュート（McKinsey Global Institute）から2011年に公表されている（図3）。2008年時点で統計学や機械学習に関する高等訓練の経験がある大学卒業生数は、米国24,730人、中国17,410人、インド13,270人などとなっており、我が国は3,400人と11位であった。また、人口100人当たりに換算するとルーマニアとポーランドが23人、英国が13人などとなっており、我が国は2人と25位であった。

欧米では、教育機関や他の人材育成事業が我が国に比べて非常に多く、量的には人材不足に陥っていない[121]。しかしながら、米国では、ビジネス課題やそこで要求されるデータサイエンスの高度化・専門化に伴うデータサイエンティストの質の問題が指摘されている（図4）。米国におけるデータサイエンティストの2024年需要・供給予測では、現在考えられているデータサイエンティスト業務に就く人材は完全に充足し、現状の育成プログラムが対応できない高度に専門化された課題を解決するエリート人材が不足するといわれている。

一方、我が国のデータサイエンティスト充足状況は、従来、次のように指摘されてきた。

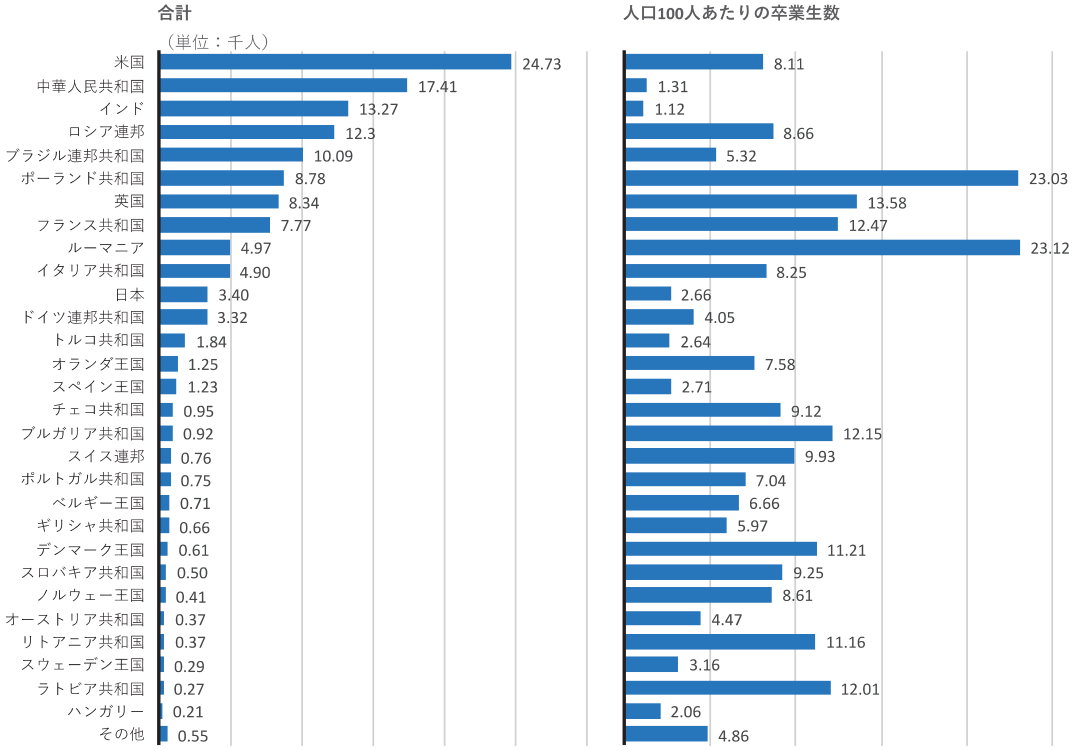
①日本ではビッグデータ関連雇用が36万5,000人分増える見込みであるが、実際に雇用条件

114　「Your Learning, Your Style. 無料で学べる大学講座」gaccoウェブサイト〈http://gacco.org/〉
115　「データサイエンス・スクール　統計力向上サイト」総務省統計局ウェブサイト〈http://www.stat.go.jp/dss/〉
116　学習者グループなどに分かれて行うPBLのこと。
117　「enPiT 成長分野を支える情報技術時内の育成拠点の形成」enPiTウェブサイト〈http://www.enpit.jp/master/〉
118　インターネット上に存在するサーバーを利用してデータを処理すること。
119　特定の機能を実現するために機械等に組み込まれる、代替不可能なハードウェアとソフトウェアによるコンピュータシステム。
120　「データ関連人材育成プログラム」2017.4. 文部科学省ウェブサイト〈http://www.mext.go.jp/a_menu/jinzai/data/index.htm〉
121　Thomas H. Davenport, “The Myth of the Data Scientist Shortage,” 2016.8.11. Wall Street Journal Website〈http://deloitte.wsj.com/cio/2016/08/11/the-myth-of-the-data-scientist-shortage/〉

を満たせる人材は 11 万人程度しかいない[122]。

②将来的には、国内ではデータサイエンティストが約 25 万人不足する[123]。

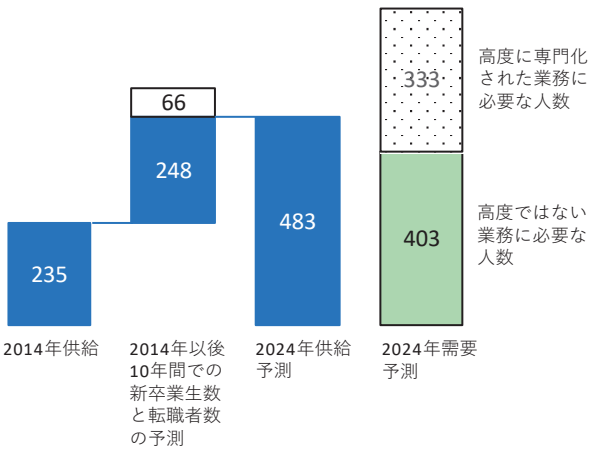
図３　データ分析スキルを有する大学卒業生数（2008 年）



（注）「その他」は、フィンランド、エストニア、クロアチア、スロベニア、アイスランド、キプロス、マケドニア、マルタを含む。

（出典）James Manyika et al., *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, 2011, p.105.〈https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf〉を基に筆者作成。

図４　米国におけるデータサイエンティスト充足予測



（出典）Nicolaus Henke et al., “The Age of Analytics: Competing in a Data-driven World,” 2016.12, p.39.〈https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.ashx〉を基に筆者作成。

---

[122] 國谷武史「201x 年に情報システム部門はどうするべきか?」2012.10.4. ITmedia ウェブサイト〈http://www.itmedia.co.jp/enterprise/articles/1210/04/news117.html〉

[123] 「ビッグデータ、分析に人材の壁」『日本経済新聞』2013.7.17.

経済産業省は2016年6月に、データサイエンティストを含む先端IT人材の動向と将来推計に関する調査結果を公表した。これによれば、我が国では、2016年時点で9.7万人のデータサイエンティストを含む先端IT人材がいるが需要と比べて1.5万人不足しており、2020年には4.8万人不足する。(124)

一方、データサイエンティストに限って人数を推計すると異なる数字が得られる。例えば、上場企業（2017年10月時点で3,566社）がそれぞれ11人のメンバー（リーダー1人を含む。）からなるデータ分析チームを1チームだけ作った場合、最低でも約3.9万人のデータサイエンティストが必要になる。上述した2013年の日本経済新聞記事中には「データサイエンティストIT各社の現状」として、富士通やNECが2015年までの2年間で現状の100名から倍増させることや、SASインスティチュートジャパンは100名から更なる増員を検討していることが紹介されている。各上場企業におけるデータサイエンティストの実際の需要は、ここでの約3.9万人という推計値も上回っているといえる。

４　国内の人材育成に係る課題と職務標準化への動向

各企業がデータサイエンティストを調達するにはコンサルティング会社等の外部サービスへアクセスすること以外に、組織内部での育成･転用、学会･インターンシップやコンペティションの利用、研さんのためのコミュニティ活用等、組織内外における人材育成の取組を含めた様々な方法がある(125)。この調達方法の傾向は、海外においても同様であるため(126)、データ活用社会を支える人材に係る課題は、育成される人材の量と質にあると考えられる。

我が国では、全てのレベルのデータサイエンティストが不足しており、特に重大な問題は棟梁レベルの人材が育っていないこと(127)といわれている。棟梁レベル人材を育成し、育成された棟梁レベルの人材が独り立ちレベル以下の人材の育成に当たることによって、データサイエンティストの効率的な供給が促進され得る。また、棟梁レベルの人材育成によって、そこから業界を代表するレベルの人材が出現することが期待される。

スキルセットによってデータサイエンティストを定義する場合、その職務を標準化することによってデータサイエンティストが提供するサービスの質が保たれ得る。我が国では、2017年4月に、情報処理推進機構がデータサイエンティスト協会とともに、第4次産業革命に対応した新スキル標準「ITSS+（プラス）」を公表した(128)。これは「セキュリティ領域」と「データサイエンス領域」を統合したスキル標準として暫定的に取りまとめられたものである。なお、国際的には、国際標準化機構（International Organization for Standardization: ISO）が、2017年6月に、ISO20252（市場・世論・社会調査－用語及びサービス要求事項）を構成する規格の一つとして、新規格ISO19731（市場調査を目的としたデジタル分析とウェブ解析）を制定・発行した(129)。こ

---

(124)　みずほ情報総研「ITベンチャー等によるイノベーション促進のための人材育成・確保モデル事業　事業報告書　第2部　今後のIT人材需給推計モデル構築等　編」2016.3, pp.216-220.〈http://www.meti.go.jp/policy/it_policy/jinzai/27FY/ITjinzai_fullreport.pdf〉

(125)　丸山宏ほか『データサイエンティスト・ハンドブック』近代科学社, 2015, pp.102-106; 情報・システム研究機構統計数理研究所　前掲注(112)

(126)　Zacharias Voulgaris, *Data Scientist: The Definitive Guide to Becoming a Data Scientist,* New Jersey: Technics Publications, 2014, pp.175-234.

(127)　情報・システム研究機構ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会　前掲注(102), pp.8-9.

(128)　「ITSS+（プラス）」2017.4.7. 情報処理推進機構ウェブサイト〈https://www.ipa.go.jp/jinzai/itss/itssplus.html〉

(129)　International Organization for Standardization, “ISO 19731:2017,” 2017.6.〈https://www.iso.org/standard/66187.html〉

の規格では、ウェブ上のデータ集合の集積を読み解くためのデータ収集・分析・報告に関するルールが定められている。

Ⅳ　データサイエンスと法制度

国立情報学研究所副所長　佐藤　一郎

従来、科学技術におけるデータの活用の限界は、データ処理するコンピュータの性能によることが多かった。例えば、データ量が多すぎると処理しきれなくなる、又は未対応なデータ形式は扱えないなどが挙げられる。しかし、近年はデータの活用の限界はむしろ法制度に起因することが多くなってきている。本章ではデータサイエンスを推進する上で、想定される法制度的な制約とその対策について概説する。

1　個人情報とプライバシーに関する法制度とデータサイエンス

IT ビジネスの主戦場は、従来のスパコンやデータベースから、ソーシャルネットワークサービス（SNS）に代表されるネットサービスに移行している。SNS では書き込み情報やユーザ間の関係など、人々の日々のプライバシーに関する情報をデータとして蓄積・利用しており、SNS の研究は人々のプライバシーをかいま見ることになる。また、社会学やコンピュータサイエンスなどが対象とするデータは、ネットサービスに関わるデータにも広がっている。この結果、科学技術と社会との境界は曖昧になり、人文科学においても大規模な社会実験やアンケートが広く行われており、研究においても個人情報やプライバシーを直接的又は間接的に収集していることは少なくない。

(1) 個人情報とプライバシーの関係と各国の法制度

さて、個人情報やプライバシーに関わる法律を検討する場合、その保護対象が問題となる。プライバシー情報とそれ以外の境界は曖昧であり、何をもってプライバシーとするかは個々人によって違う。法制度は、定義のない対象を保護することはできない。そこで、1980 年に OECD が個人情報保護の基本となるガイドライン「OECD 8 原則」[130] を策定した。同原則では、プライバシー情報とみなされる個人情報、つまり特定の個人を識別する情報を保護する方針が打ち出され、我が国を含む多くの OECD 諸国は、法律やガイドラインが国ごとに異なると国際ビジネス上の問題が生じることから、そのガイドラインを参考として国内法制度を定めた。しかし、各国で保護される個人情報が同じとは限らず、特に英米と欧州及び日本とではかい離が見られる。例えば英国では住所は公知情報とされることが多く、米国は 18 歳未満の個人に関する情報は欧州や日本以上に厳格に扱われる。また、法制度の運用も異なり、個別事例にきめ細かく対応する国と大枠の法規制で対応する国もある。その背景には、英米はいわゆるコモンロー（Common Law）、欧州や我が国は大陸法（Civil Law）を採用していることがある。前者は判例に基づく事例ごとの対応が可能であるが、後者は事前に規制範囲を設定するため、企業

---

[130] “OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.” OECD Website〈http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm〉で示された基本原則をいう。

等データを活用する側の活動を制約しがちになる。経済界には、その柔軟さから英米に類似した個人情報保護制度を望む声もあるが、その場合、法規制よりも訴訟による解決が多くなり、企業などにとって訴訟コストが膨大になることが予想される。

（2）匿名加工情報と要配慮個人情報

我が国では、「個人情報の保護に関する法律」（平成15年法律第57号。以下「個人情報保護法」という。）が平成27年に改正され、同意なしで第三者提供し得るデータ類型（匿名加工情報）の取扱いなどデータ利活用のための仕組みが導入された。また、同改正により、厳格な取扱いが必要な個人情報の類型として「要配慮個人情報」が規定された。これは、人種や宗教、犯罪歴、病歴などの差別につながる情報を取得する際には、本人の事前同意を必要とするものである。第三者提供する際の条件が厳格化されることから、要配慮個人情報に該当する医療情報を病院や大学が匿名加工情報として共有する場合に限って、その第三者提供を可能にする「医療分野の研究開発に資するための匿名加工医療情報に関する法律」（平成29年法律第28号）が制定された。

（3）個人識別符号

個人情報保護法の平成27年改正では、個人情報範囲を明確化するため、個人情報の定義に「個人識別符号」という概念が加えられた。これは二つの類型から構成され、一つは特定の個人の身体の一部の特徴に関するデータ、もう一つは個人に付番される番号や記号で、旅券の番号、運転免許証の番号、住民票コード、個人番号などのデータが対象となる。データサイエンスにおいて注意が必要を要するのは前者の類型（遺伝情報、顔認識情報、声紋、指紋など）であり、特に機械学習[131]や深層学習[132]において問題となり得る。機械学習や深層学習の「学習済みモデル」[133]は一般に統計情報となるので、個人情報保護法による規制はかからない。しかし、顔認識向けの機械学習や深層学習による学習済みモデルは、個人識別符号に該当し、個人情報保護法上の個人情報となり、同法に基づく保護が求められる可能性がある。

（4）学術研究を目的とする場合の適用除外

個人情報保護法の第76条第1項第3号は、大学その他の学術研究を目的とする機関若しくは団体又はそれらに属する者、学術研究の用に供する目的である場合、個人情報の取扱規定を適用しないこと（適用除外）を規定している。科学データを扱う場合、仮にその科学データに個人情報が含まれていても、学術研究を目的とする機関等が学術研究で利用する場合は適用除外となり得るが、営利企業は除外されない。また、大学などが企業からの受託研究を行う場合、学術研究の用に供する目的とはいえなくなる可能性があり、その場合は除外されないこともあり得る。こうした適用除外の判断を支援するため、所管省庁などがガイドライン等を整備することもあるが、文部科学省は、医学研究などを除き、必ずしもガイドラインを十分に整備して

131　機械学習とは、データに潜むパターンを学習して、その結果を新たなデータに当てはめることで、パターンに従がって将来を予測する手法。

132　深層学習はディープラーニング（Deep Learning）と呼ばれることが多い。多層のニューラルネットワーク（脳機能に見られるいくつかの特性を計算機上のシミュレーションによって表現することを目指した数学モデル）による機械学習手法であり、近年、画像認識等を中心に広く利用されるAIの手法である。

133　機械学習などの処理における学習用データを読み込ませて、その学習用途を実現するために必要なパラメータ（係数）が規定されたデータ群。例えば機械学習による人間の顔に識別では、それぞれの個人の顔画像から、画像上の濃淡や形状を数値化したものが学習済みデータとなる。

いなかった[134]。例えば、小中学校の生徒に対する緊急連絡リストを生徒保護者に配布するときの個人情報保護法上の可否に関わる指針が十分に示されていなかったこともあり、必要な情報が共有できないなどの問題が生じた結果、地方公共団体が指針を出すこととなった。

（5）個人情報提供者の同意と個人情報の管理

個人情報を利用したり第三者提供したりする場合、個人本人の同意が前提となる。そのため、企業や研究機関は個人から個人情報を受け取るときに、個人情報の利用目的や管理体制を具体的に説明すること、つまりインフォームドコンセント（Informed Consent: IC）が重要となる。ただし、医療分野を除き、必ずしもICの体系化や基準の整備は十分でなく研究者に任せられている状況であり、不十分な説明で個人情報を取得しているケースは少なくない。

ICとともに重要となるのは個人情報の管理である。不十分な管理は漏えいなどにつながりやすく、個人情報保護の実効性を担保することが重要である。個人が企業や研究機関に自らの個人情報を提供し、その利用を許すのはその企業や研究機関であれば個人情報を適切に扱うと信頼しているからである。例え少数であっても、一部の研究機関が信頼を失うような行為、例えば個人情報の漏洩や目的外利用をすれば、学術研究全体が信頼を失い、個人情報を集めることは困難となるであろう。

2　著作権とデータサイエンス

（1）科学データと著作物

「著作権法」（昭和45年法律第48号）において、著作物は「思想又は感情を創作的に表現したものであつて、文芸、学術、美術又は音楽の範囲に属するもの」と定義されている（同法第2条第1項第1号）。このため、科学データのうち、観測や実験などの測定データは、思想又は感情とは無関係であり、著作物の対象とならない。また、研究者が実験や事象を記載したテキスト（例えば「2017年10月30日の東京は快晴だった」）に関しても、単なる事実を表記したものは創作的な表現とはいえないことから著作物となり得ない。

（2）情報解析のための著作物複製

一方、研究において著作物を扱う場合も多い。例えば、小説における特定単語の出現頻度を分析する場合、著作物である小説を扱うことになる。また、一般にデータ処理では、データをその処理するコンピュータにコピーする必要があり、複製権の侵害が懸念されていたが、平成21年の著作権法改正により、情報解析の目的で、必要と認められる限度において記録媒体への記録又は翻案する行為は、著作権侵害ではないとされた（同法第47条の7）。ここで「情報解析」とは、データなどから言語、音、映像その他の要素に係る情報を抽出し、比較、分類その他の統計的解析を行うことであり、その目的もデータ分析だけではなく、本人認証、自動翻訳、社会動向調査、情報検索も含まれることになる。

---

[134]　文部科学省研究振興局ライフサイエンス課生命倫理・安全対策室／厚生労働省大臣官房厚生科学課、医政局研究開発振興課「人を対象とする医学系研究に関する倫理指針ガイダンス」平成27年2月9日（平成29年5月29日一部改訂）〈http://www.mhlw.go.jp/file/06-Seisakujouhou-10600000-Daijinkanboukouseikagakuka/0000166072.pdf〉

（3）情報処理・情報提供を円滑かつ効率的に行うための著作物複製

平成21年の改正前の著作権法では、一時的・瞬間的にデータをコンピュータに蓄積することは権利侵害に当たる複製として扱われてきたが、改正後は、コンピュータ内部の通常の技術的過程で生じる情報の記録（メモリーやハードディスクへの著作物の記録）に関しては、著作権侵害に当たらないとされた（同法第47条の8）。こうしたコンピュータ内部における情報の記録が著作権者に不利益を与えるとは考えられないからである。ただし、同条は、記録できるのは「当該情報処理を円滑かつ効率的に行うために必要と認められる限度」としている。

また、平成24年の著作権法改正で、情報通信技術を利用した情報提供を円滑かつ効率的に行うための準備に必要なコンピュータ上の情報処理のための利用が認められた（同法第47条の9）ことも科学データの活用において重要である。これは、例えば動画像配信において、画像データを送信に適したデータ形式に変換・圧縮することなどを想定したものである。科学データの共有・公開では、統一的なデータ形式に変換することが必要となるが、その変換行為自体は著作権法違反ではないことになる。ただし、情報技術の進歩により、低解像度画像データから高解像度画像データに変換する技術や、白黒映像をカラー映像に変換する技術など、著作物の情報量を増やす手法も登場しており、情報技術の進歩に合わせた対応が求められるであろう。

（4）検索サービスのための著作物複製

平成21年の著作権法改正では、ウェブサイト等の検索サービスの提供過程における著作物の記録及び翻案が認められた（同法第47条の6）。こうしたコンピュータ内部における情報の記録が著作権者に不利益を与えるとは考えられないからである。ただし、その対象は、サーバーに記録され、送信可能化された著作物に限定される。なお、「翻案」の意味は、収集データの加工とそのサーバーへの記録、公衆からの求めに応じた検索結果のURL表示や要約情報の提供の範囲内であろう[135]。

3　安全保障とデータサイエンス

通信技術の進歩により、大量のデータであっても短時間に世界各地とのやり取りができるようになった。しかし、データによっては大量破壊兵器、その他通常兵器の開発等に転用されるおそれがある。「外国為替及び外国貿易法」（昭和24年法律第228号）により、大量破壊兵器やその他の通常兵器の開発等に転用され得る特定の技術を、安全保障上の危惧がない国以外に輸出するときは規制を受ける。このとき当該技術に関わるデータを海外に転送する場合も同様に規制を受けることになる。国際的な共同研究において科学データを共有する場合、そのデータの応用可能性や共有する国に応じてその是非を判断する必要がある。ただし、自然科学の分野における現象に関する原理の究明を主目的としたデータであって、実験又はシミュレーションにより得られるものであり、特定の製品の設計又は製造を目的としない場合は同法の規制の対象外となる。このため、宇宙の生成に関するシミュレーションデータ、気象測定データなどは対象外となる。

---

(135)　半田正夫・松田政行編『著作権法コンメンタール2　第2版』勁草書房, 2015.