

## 〈学内共同研究報告〉

# オープンサイエンス動向の調査と研究

## —その意義と研究データ活用の方策—

藤井 稔也・廣瀬 毅士・高橋 玲・長沼 将一<sup>1</sup>

### Abstract

オープンサイエンスという概念は、オープンアクセス、オープンデータ、シチズンサイエンスなど異なる多数の概念が絡み合った複雑で混沌としたものであるが、端的には科学のオープン化というムーブメントとして総括できるとも云える。研究者は意識するしないに関わらずこのムーブメントに既に巻き込まれており、例えばオープンアクセスによる恩恵に始まり、研究データの預託が義務化される未来もすぐ先であろうとも予想される。本研究ではオープンサイエンスの調査に始まりそれを個々の研究分野と照らし合わせることにより本学に所属する研究者がオープンサイエンスにどのように対峙していくべきかを提示していくことを意義とし、ゆくゆくは必要な研究支援や環境が何かを見極めることにある。

**キーワード (Key Words) :** オープンサイエンス、オープンアクセス、シチズンサイエンス、メタデータ、オープンデータ

### 1. オープンサイエンスとは (藤井)

「オープンサイエンス」は世間一般にはインターネット出現に伴って生まれた新しい概念のように解釈されている。しかし実際にはその萌芽はルネサンス期にみられるものであるが、未だ定見がなく漠然と議論されている概念である。

一方、コンピュータ技術の発展により捨てられるだけであった大規模データが実用的に解析可能となり、その結果データサイエンスという分野が生まれた。オープンサイエンスはそれと混同して扱われることも多いが、少し紐解いてみると、データサイエンスを内包する研究者であれば必ず念頭に置いておかなければならない重要な概念であり、この概念を聞いたことがない研究者であっても、その幾つかは既に指針として持ち実行しているものであることが判ってくる。つまり「オープンサイエンス」は全ての研究者が素養としてもつべきものであるが、その実体は誰もがまだ掴み切れていないものといえるだろう。

#### 1.1. オープンサイエンス共同研究への下地

この共同研究は、代表研究者の藤井稔也と共同研究者の廣瀬毅士による企画で始めたものである。廣瀬は、本学において社会調査士系の科目を受け持ちデータサイエンスを実践している研究者であり本学で最も早くオープンサイエンスの重要性にも気づいた教員である。藤井はメディアデザインを研究の中心に据えインタラクション研究などを専門とする教員であり、共同研究としてデジタルアーカイブなども過去に行っていたことやユーザーインターフェイスのオントロジーの構築にも着手して、オープンサイエンスの趣旨には感じるものがあり共同研究会を立ち上げることになる。その際、多くの分野の研究者に助力を

得ることが重要とも考えたが、我々の持つ漠然としたオープンサイエンスの知識で成果を出すことは危ぶまれた。効率的に短いスパンでの試行錯誤を繰り返すことで行うことができ、スモールスタートといってよい程度の少人数に据え置き、異なった分野の本学在籍の中堅および若手研究者として、経済人類学を専門とする高橋玲、教育工学を専門とする長沼将一の二名を共同研究者として招き入れ研究を開始することとなった。

## 1.2. 研究に関する喫緊の課題との関連性

研究環境が時代によって変化し続けていくのは当然として研究者はそれを積極的に受け入れるのかそれに抗うかを迫られている。そして受け入れる場合でも変化の速度に研究者自身がついていけるのかについては疑問がある。研究に関する喫緊の課題とは大きく二つと云える。それは研究データを含めた成果の公開と事件化することさえあるデータ改竄など研究倫理の問題である。

前者は、科研費等の公的資金を得て推進された研究、収集されたデータは税金を原資とするため、最終的な研究成果としての論文と同様に研究データもまた国民の間で共有すべき知的資産であり、ともに公開が義務づけられるべきものであるという考えでありオープンサイエンスの中のオープンデータという概念と関連付けられるものである。成果としての論文は、学会発表およびその学会が論文をオープンアクセスにしているかに係わるものであり、それなりに整備されているものであるが、研究データの公開は殆ど未整備であり、本研究においても課題として取り上げざるを得ない点であった。

後者は、昨今、生物学系や心理学系で問題化し議論も盛んであることから研究倫理にも絡ませ最も対策がなされていると思われる。本共同研究ではその分野の研究者の参加がないため具体性に欠けるため本共同研究では深入りしていない。

## 1.3. オープンサイエンス概観

科学はギリシャ文化、イスラム文化における古典的知を踏まえ14世紀のルネサンス期に始まったことに異論を唱える人はそう多くはない。ではその科学にオープンと冠を付けたサイエンスが何時頃始まったかについては明確な定義はなく、これに関しての議論は尽くされていない。強いて言えば17世紀に出版され始めた学術雑誌をその萌芽と捉えることができる。また、オープンがつくと何が変わるのか、逆にクローズドなサイエンスという概念を考えた場合、クローズドサイエンスを標榜する研究者は存在しないであろう。つまりはサイエンスにはオープンサイエンスのみしか存在しえないということになる。昔から存在してはいても、概念としては比較的新しいものであり一部ではこれをデジタルサイエンスと言い換えることもあるようにコンピュータ技術との関連性のあるものである。しかしデジタルがオープンサイエンスの主要な要素であるかとも言い切れないが、オープンソース、オープンデータ、オープンアクセスといったデジタル技術抜きには存在しえない概念を要素として構成されつつあることも事実である。また、ArXiv.orgのようなプレプリントサーバの出現やオープンピアレビューといった研究方法やその支援の仕組み、科学の民主化を目的とするシチズンサイエンスやPublicLab.orgのような科学におけるオープンな手法確立、インターネットにより注目された集合知や学術雑誌や国際会議をレポートする原動力にもなっているオンラインコラボレーションの活用、MOOCsに代表されるオープ

ンコースウェアもしくはオープンエデュケーショナルリソースなどをもオープンサイエンスは内包するとされている。

また、シチズンサイエンスに関しても補足しておくべきであろう。ルネサンス以前に戻ると科学者はパトロンの庇護の元に研究を行っていた。そして新しい知識の発見はパトロン名声として還元されており成果は秘匿されることも珍しいことではなかった。時代を経るに従い科学者は横に繋がるようになり学術機関が整備されることにより自身の自立も確立していく。現在においては無数の学術雑誌や学術機関が存在し科学者はその中で切磋琢磨しているが、弊害として購読料の壁が生まれてしまった。あらゆる人が全ての学術情報にアクセスすることが出来ない状況である。本来、科学はあらゆる人が参加できるべきであり、科学の民主化と言い換えることもできるシチズンサイエンスの重要性である。

#### 1.4. オープンサイエンス研究への道筋

共同研究を始めるに当たっての目論みとして我々は研究データを含めた成果の公開とその二次利用の促進を実践的に行うことで道筋が見えるのではと考えていた。しかし、これは勇み足である可能性を考えねばならなくなった。研究データを広く共有・公開することは世界的な潮流であるが、我が国においてはデータ共有・公開を行う機関が少なく立ち遅れている事実から、本学において試行的にそのようなサーバーを立ち上げること、データ共有・公開のための方策やそのプラットフォームを指し示すこと、実行可能な方法を本学の教員に指し示せば事足りるのではないかという仮説に基づくものである。それは特定の研究分野においては可能なことであるかもしれないが、それを全ての分野において成り立つものではないのではという疑念がオープンサイエンスを論文等を辿って調べていく過程や我々の議論において浮上してきたのである。

よくよく考えてみると研究データの管理、保管、公開といった方法論は、オープンサイエンスが注目されるまでなかったわけではなく、大学や研究機関におけるアーカイブセンターや機関リポジトリなどは存在していた。研究データを二次利用を主眼におくという面でメタデータの整備によりデータ共有を円滑化し研究ワークフローの簡素化が可能となる支援環境の構築などはされており、このことによりデータの保全という主目的だけでなく副次的に研究データの改竄防止が可能であり結果的に研究不正対策をも達成できると目されていた。またデータマネジメントプラン（DMP）は研究者が研究の計画段階において予算計画とも絡め研究データの管理・保管・公開方法を計画し提出を求められるものであるが、どのようにすればそれが可能であるかを研究者に明確に提示できるところまでは至っていないように思える。研究者が個人的努力でデータを公開するケースは増えているとはいえ、これは長期的維持に適した方式ではない。その研究者がリタイアした場合などデータの散逸や消失を未然に防ぐには、法人組織や公的機関での枠組みは大事であろう。本研究での当初の研究目的の半分は試験的な支援プラットフォームの整備であったが、それを行う前にすべき検討や議論が山積していることが明らかになったと言える。

#### 1.5. 研究データの再利用性

データ共有の基準として FAIR の原則というものがある。見つけられる（Findable）、アクセスできる（Accessible）、相互運用可能である（Interoperable）、再利用できる

(Reusable) の頭文字を繋げたものであるが、研究データの再利用性とは、その一つの要素であり、それは論文作成に用いる解析前の研究データを含めた他研究者への公開から始まる。しかし、公開されたとして二次利用が可能かといえばそうではない。調査で得たデータはそのままでは、他人が解析できるものではなくメタデータと呼ばれる付加情報が対にされて意味をもつ。また、人に対する調査研究では、個人を特定可能な情報が含まれることも多く、公開には個人情報保護法に従い秘匿化する必要に迫られるが、それを施したことで無意味となり公開に値しない研究データもある。つまり再利用性は多くの問題に阻まれたままである。

再利用性を高める努力は、副次的に研究不正の防止にも繋がる。つまり、研究データが公開されていることは再現性が担保された環境が整備されているということに他ならず、誰もが研究の正当性を検証可能であることにより、研究者の不正行為を抑制する効果があるということに他ならない。このことにより、公開するしないに拘わらず研究データの一定期間の保存が義務づけるべきという議論もあり、その動向は目が離せない。

また、これに前向きな研究者ほど見落としがちな事実として、研究データの公開は直接に研究成果に結びつかない作業であり、個々の研究者のモチベーションを維持する仕組みが現時点では明瞭ではないということも記しておくべきであろう。

### 1.6. メタデータとは

再利用可能性のキーポイントはメタデータである。メタデータはメタ情報とも呼ばれるもので、いわば「データについてのデータ」であるがその定義では具体性に乏しく実感が湧かないかと思われるので説明すると、データ自体が[山田太郎、45、男性、東京都]であった場合、それぞれのデータ要素が何を表すかは類推可能なものがあるとしても、最後のデータ要素は出身地かもしれないし現住所かもしれないという曖昧性が残る。これは調査を行った研究者が付与することで特定可能である。

メタデータは表計算ソフトウェアであれば最初に説明行を加えることでも可能であるがばらばらな表記や曖昧性を含むなど形式化されていないといえるため、今日の大規模なデータ処理には適さない。公文書館や公共図書館が整備されて歴史のある図書館情報学は最も早くからメタデータ付与がなされ基本語彙セットであるダブリンコア (DC) が作製され、他の研究分野でも利用されている<sup>2</sup>。社会調査分野ではデータ・ドキュメンテーション・イニシアチブ (DDI) がある。この辺りは分野によって温度差がありメタデータ定義を研究者が行う必要性が生じ、結果的に再利用性が下がるということも起こりえる。またメタデータの利用やある程度の専門知識が必要であり、研究者本人の学習コストが無視できない。そのハードルを低減する試みとしてはインタラクティブなオーサリングアプリケーションや自動化を図るツールが考えられる。事実上、殆どの研究者は個票データとして、調査リストはエクセルとして管理していることが多いと考えられ、そのような研究者にどのような動機付けが有効なのかは誰もが明確な答えが出せていない状況である。

### 1.7. 小規模大学や個々の研究者の観点からのオープンサイエンス

オープンサイエンスの中で最も進んでいる要素はオープンアクセスであろう。本学でも

紀要論文集を発行しており東京通信大学機関リポジトリによりオープンアクセスとなっている。これは大学の社会貢献という意味合いも持つが共にオープンサイエンス実践ともいえる。ただ、これはオープンアクセスにしていればいいことではなく、オープンサイエンスの意味を理解して、よりよい方法となるようにしていくべきである。本稿も東京通信大学紀要への報告論文であり、これが掲載される号から CC-BY-NC-ND で公開されることに変更された。CC はクリエイティブコモンズライセンスに従うという意味であり、NC は非商用、ND は改変禁止を意味する。それ以前に比べてライセンスの明確化がされたと云えるが、ここにもまだ議論の余地は残されている。例えば、ND を付けるということは改変を禁じる強い提示であり多国語に翻訳する際も原著者全ての許諾を取る必要が出てくる。これは再利用性を重んじる立場で言えば、本当のオープンアクセスとは云えないというラディカルな主張もある。

電子ジャーナル・データベースの大学機関の購読という意味では、トップレベルの研究機関や大学においては主要なジャーナルの最新号が常に読めるなど研究環境が恵まれているといえる。中小規模大学において同様の環境が揃えられるかといえば潤沢な研究費や図書費を用意することは事実上困難であり、そういう意味では市井の個人研究者は全くの支援なしとなる。これは社会構造変革の必要があるということであるが個々の研究者の手には余ることである。出来ることから始めるとすれば、各機関、各研究者がオープンアクセスやオープンコラボレーションに取り組むことにのみ活路が開けるだろう。

## 1.8. 個々の分野のオープンサイエンス

オープンサイエンスの取り組みはまだ始まったばかりであり、研究データの再利用は十分にされているとは言い難い。では研究データの二次利用の取り組みは無意味かという問をしたときに、その答えとして意味があるという予想は出せるように思う。一つの理由はインターネット技術の浸透であり、二つ目はオープンサイエンスに対する研究者の意識が高まっていることである。あとは、多くの研究分野におけるオープンサイエンスの取り組みや状況を各研究者が把握していく過程が全体を補強するのではと考える。次節以降では、第 2 節では藤井が、オープンソースやデジタルアーカイブといった分野を、第 3 節では廣瀬が社会調査を中心とした社会学での取り組みを、第 4 節では高橋が質的研究の立場から、第 5 節では長沼が教育工学の事例などを踏まえ担当していく。

## 2. 情報学におけるオープンサイエンス（藤井）

情報学は範囲の曖昧な研究分野であり元々は図書館情報学を源流としコンピュータ科学（CS）と情報技術（IT）を包含すると考えれば間違いはない。厳密には日本学術会議の参照基準により策定される定義がより厳密であると思われるが、ここでは二つの分野に限定して情報学とオープンサイエンスの関わりを紹介する。前者はオープンソースである。これはオープンサイエンスの一つの要素として扱われるが、ある意味オープンサイエンスの源流の一つと言ってもよい重要な概念でありソフトウェア工学の一つの手法である。後者は如何にして情報を残すかという問題についての分野のデジタルアーカイブである。これには上手く行っている事例もないわけではないが、そもそも研究者が、というよりはアマ

チュアが趣味としてなされているというしか云えない現状のものを紹介する。

## 2.1. オープンソースムーブメント

ソフトウェアはコンピュータのハードウェアに付随するものとして認識されていたが、マイクロコンピュータの出現と共にソフトウェアを専業とするマイクロソフトを始めとする私企業の独立製品として提供されるようになった。これらは企業として当然のように使用技術は特許や著作権で守られバイナリコードとして提供されるものである。これらはプロプライエタリと呼ばれるが、その対極としてあるソフトウェアがオープンソースである。バイナリコードが人間の解読が困難なのに対して、料理を作るに当たってのレシピにあたるソフトウェアのソースコードが公開されているという意味と理解できる。公開される意味は著作者以外の誰もがそれを複製し利用できるということであり、企業としては秘匿することがその活動での常識であるが、それを公開することのメリットを広める運動がオープンソースムーブメントである。その歴史については本稿では手に余るので割愛し、ここではオープンソースのソフトウェアとして最も成功したと考えられる **Linux** を取り上げ、それとオープンサイエンスとの関連性について述べる。**Linux** は当時フィンランドの大学生であったライナス・トーバルズが開発した **UNIX** 互換のソフトウェア (OS) であり「ライナスの **UNIX**」という意味で名付けられたものである。アマチュアの営みとしての出自であり当然としてソースコードが開示されていた。当初はマニア向けもしくは玩具と考えられ企業からは無視されていた存在であったが、現在はサーバーサイドの OS として大きなシェアを誇るまでに成長した。この理由としては幾つか挙げる事が可能であるが、ライナス・トーバルズの為人も大きな要因ではあるが、オンラインコラボレーションをその筆頭とすることが出来るだろう。**Linux** は当初、彼の独りプロジェクトであったがインターネットを介して多くのエンジニアと共同で開発する存在へと進化した。これにはメーリングリストやバージョン管理ソフトウェアなど必要とされるツールもまたオープンソースとして用意し協調して遂行するシステムを作り上げたことによるものである。現在のオープンサイエンスを支援するソフトウェアツール群は **Linux** に限らないオープンソースムーブメントから生まれたものであり、オープンデータはオープンソースの延長にある概念であり、オープンアクセスはオープンソースを育んだインターネットの思想そのものであるといっても過言ではないであろう。

## 2.2. デジタルアーカイブ

デジタルアーカイブという恐らく大抵の人は書物をスキャンしてデジタル化することを想像するであろう。それ自体を否定するわけではないが、デジタルアーカイブは情報を残すという意味として、ソフトウェアやデータを残すという意味でそれを稼働させ蓄積するコンピュータやハードウェアを含めた方法論とその実践である。書籍に限れば文字というそもそもデジタル化されたデータをコンピュータで扱える形にデータ化するだけであり、挿絵などは十分な解像度で取り込むことで大半は事足りると言える。しかし、コンピュータのソフトウェアはそういう訳にはいかない。コンピュータはそのどれもが同一ではなく様々なアーキテクチャで設計されたものであり、そのハードウェアは壊れゆくものである。例えば世界初のコンピュータと言われる **ENIAC** のソフトウェアは保存されているかとい

えばその殆どは失われてしまったといえる。ENIAC はパッチケーブルの配線でプログラミングされたものであり、そもそも残すことを前提に組まれたものでない上に稼働するハードウェア自体が保存されていない。同様に、エポックメイキングな大型コンピュータの多くは博物館などで展示されている例はあっても実際に作動する状態で保存されることはほぼない。稀な例として富士通が自社の初期の大型機種である FACOM-128B を動作可能な状態で保存している他、継続的な行いとして米国のコンピュータ歴史博物館では幾つかの歴史的なコンピュータを修繕して動態展示している。失われゆく技術をそのままの状態に維持することはコストの掛かるものであり全ての技術において行うことは不可能であるが、そのことが技術史の研究者にとって足枷となる状況をどのように解決するかはオープンサイエンスの課題の一つといえるだろう。これは古いコンピュータ技術においてだけ起こっているわけではなく比較的新しいコンピュータやそのソフトウェアでも起こることである。

ハードウェアは経年変化で故障するものであり企業が用意する補修部品の保持も 10 年程度を期限としている。また修理のために設計図が公開されていることは稀であり、あったとしても修理には設計並みの知識をもったエンジニアがあたることになるが、高齢化によりこの人的資源も先細る一方である。ハードウェアの保全は諦めエミュレーション技術でもって代えることは可能であるが、ハードウェア依存の部分をどうするかは解決策はない。例をあげれば現在、殆ど使われなくなったフロッピーディスクはドライブ装置が既に生産されることはないが、それに残されたソフトウェアは膨大にあり、それらを研究材料にする場合、ドライブが入手困難であると、その研究に支障をきたすことになる。紙の書籍は読み取り機は人間の眼となるが、このような資産は、眼となるハードウェアが存在しなくなるという危機は想像より容易に起こりえるので別媒体に情報の移設を急ぐか、フロッピードライブの動態保存を計画することになる。この場合、やるべきは両方ということになる。

ネットワークを利用したオンラインゲームという分野がある。これはゲームプレイヤーが持つデバイスとインターネットを介したサーバーに主体がある。これが進化したものにソーシャルゲームがある。これはスマートフォンの台頭と前後したものでサーバーというよりはソーシャルネットワークの形成と存在に主体が移っているため、これを動態保存することは事実上不可能である。この場合、部分としての保全を考える必要がある。

オープンサイエンスの一つの要素としてシチズンサイエンスがある。職業的な研究者だけでなく市井のアマチュア研究者を含めたサイエンスの在り方と考えれば誤解が少ないと思われるが、ハードウェアの動態保存という意味で、この分野のシチズンサイエンス比重は比較的高いといえる。例えば、8 ビットの CPU を利用した初期のパーソナルコンピュータの収集家は技術レベルも比較的高くオークション等での売買も活発であるように一定数が確保されているといえる。しかし、人気のあった機種以外の保全や個人の資産に依存した営みであり収集家の死などにより消滅する可能性も無視することができない。この解決には組織の形成や資金的投入が必須となるであろうが、そこにオープンコラボレーション的手法が解決の糸口になる可能性は高いだろう。

### 3. 社会調査データのオープン化（廣瀬）

#### 3.1. 統計的社会調査データの特徴

本来的にあって科学とは他の研究者が再現可能な手続きに則っているものであり、同じ方法論（分析手続きなど）を、同じデータに適用することによって検証可能なものでなくてはならない。社会科学とりわけ社会学では測定対象が人間の行動や意識に類するものが多いため自然科学のような意味での再現可能性とは異なるものの、少なくとも統計的社会調査においては調査員によって全く異なった結果が得られるということがないように「形式化」することが要諦である。実務目的・学術目的の双方で著しく発展した統計的社会調査の典型的な手法は大規模質問紙調査、つまり大規模な調査対象に対して質問紙（調査票）を用いたデータ収集を行い、統計分析の手法を用いたデータの処理・分析を行うというものである<sup>3</sup>。

ここで形式化というのは、例えば回答者が「はい」と回答するに際しての心象やその信念の強さは捨棄されるということである。たとえ回答選択肢において「そう思う」「ややそう思う」と段階をつけるのであってもその信念の強度が各回答者を通じて同じとは限らない。しかし、個々の回答者は同じ質問紙を見せられれば異なる調査員が担当しても異なる回答をすることはない。そのように質問項目や選択肢を作成するという意味である。

統計的社会調査データは、多くの場合にデータファイルにおいて回答内容が a, b, c, ... といった記号あるいは 1, 2, 3, ... といった番号によってコード（符号）化されている。すなわち回答が選択肢法によって得られた場合はたいていその選択肢の記号や番号が、自由記述によって得られた回答の場合にはその意味するところを研究者が調査事後にカテゴリズ（アフターコーディング）して記号や番号に符号化してデータ値にインプットされることになる<sup>4</sup>。データファイルにおいてデータが符号化された（あるいは符号化できる）状態であることは、統計的社会調査データが（非統計調査、いわゆる質的調査の語りや観察の記録に比べて）小さいファイルサイズのテキストデータで済むということであり、これが統計調査データの本質的な可搬性の高さ、共有される可能性の高さを示している。

#### 3.2. 統計調査の進歩と調査件数の増加

日本において統計的社会調査が本格的に開始されて活発になったのは戦後であり、1945～1946年には主要新聞社や放送局・通信社の世論調査を担当する部局が開設され、さらに1940～1950年代には現在にも続く調査機関や調査会社が相次いで設立された。この時期の日本では学術の世界においても代表的な大規模統計調査が開始され、1953年には統計数理研究所の「日本人の国民性」調査が始まり、1955年には第1回の「社会階層と社会移動に関する全国調査」（SSM調査）が開始された。

この時期には無作為標本抽出法に基づく標本設計や統計的推測の手法が社会調査に適用されるようになり、その後もデータ分析の技術がコンピュータの計算能力に相まって高度化していったが、コンピュータの発達・進歩はその計算能力だけでなくデータ保存方法・媒体の進歩も促し、分析が簡便化・一般化するという流れもあった。

その時期から現在に至るまで、社会調査が行われる件数自体も増加した。日本で行われている世論調査の現況としては内閣府編『世論調査年鑑』『全国世論調査の現況』において政府機関・地方公共団体・大学・マスコミ・一般企業・各種団体などが調査主体となっ



て実施した世論調査を照会した結果が挙げられているが、それによれば 2019 年度（4 月 1 日～3 月 31 日）に 1906 件の世論調査が実施されたという<sup>5</sup>。1965 年度に実施された世論調査が 480 件であるから、（この間で若干の増減はあるものの）全体として顕著な増加の傾向を示している。

### 3.3. データの私蔵・散逸

本質的なデータの可搬性の高さ、および数多く実施された統計調査の一方で、特に我が国においては調査データが 2 次分析のために広く提供されることは近年までなかった。1 つの理由としては研究者側の問題であり、形式化された質問紙による統計的データの収集とはいえ、質問項目の作成や質問文のワーディング作成には多大な労力と時間、研究の蓄積が必要であるため、その調査データを用いた研究期間が終わった後も調査を企画・実施した研究者およびそのグループによってデータが私蔵（死蔵）され、ついには散逸してしまうものさえあった。

他の理由としては、研究者側がデータを寄託しデータ保持・管理の委託を行う機関およびインフラストラクチャーの欠如である。一部の研究者は個人的に管理する Web ページなどにデータを公開してデータ分析の実習授業のために提供している者もいたが、Web サイトの保守・管理や回答データの匿名化処理などは研究者にとって負担が大きい作業であった。これらの理由から、特に単一の研究者が行うような小規模調査に関しては公開があまり進んでこなかった。

### 3.4. 統計調査に対する重複投資

統計的社会調査を実施するには、相応の金額の研究資金が必要になる。日本でいえば学術研究助成基金助成金／科学研究費補助金（科研費）を資金補助する文部科学省／日本学術振興会、アメリカでいえば NSF（National Science Foundation）、イギリスでいえば ESRC（Economic and Social Research Council）などが有名であるが、他の国でも同様の公的助成の制度がある。これら機関による助成で行われた調査にも関わらず、先述のように調査データが私蔵されてしまうと、学問的意義はあっても先行する調査と類似した調査研究が行われてしまう可能性がある。このことは、上述の助成機関にとっては重複投資である。原資（の一部）が税金である場合にはこれらの助成によって行われた調査のデータは公共の財産ともいえるので、データの死蔵は社会的な損失となる。

また、重複投資が行われて同種の社会調査が繰り返し行われることは同一の個人が同種の調査に繰り返し回答する可能性があるということであり、調査対象となる市民の立場にとってみれば類似の調査が氾濫して「調査公害」ともいわれるような状況になる。このことがひいては市民の側における社会調査への忌避感につながっており、調査を実施する研究者側にとっては回収率の低下という「社会調査の困難」という事態を招きかねない。

### 3.5. データアーカイブの整備、運営

前項までのような問題点を背景に、欧米を皮切りに社会調査データアーカイブ（以下、単に「データアーカイブ」と表記）が整備されるようになった。データアーカイブは、既存の調査データを収集・保管し、調査主体以外の研究者や研究組織による「2 次分析」の

ためにデータを提供する機関である。統計的調査データについていえば、調査データの集計表や分析結果ではなく、個票レベルの回答内容（調査票情報）のデータ（マイクロデータ）が提供される<sup>6</sup>。また、データを採取するにあたって用いた質問紙などの調査データに付随する情報も提供されるのが一般的である。

データアーカイブの嚆矢は1947年に設立された Roper Center といわれるが、これは世論調査データを中心にしている。データアーカイブとしては、その後1962年にミシガン大学で設立された ICPSR (Inter-university Consortium for Political and Social Research)<sup>7</sup>、1962年にはドイツの GESIS (German Social Science Infrastructure Service)<sup>8</sup>が、1967年に設立されたイギリスの UKDA (UK Data Archive) などが代表的な機関である。

日本ではこれらの国にかなり遅れて設立され、札幌学院大学社会情報学部が1990年に設立した SORD (Social and Opinion Research Database Project) が初である。その後、1998年には SSJDA (東京大学社会科学研究所) が設立され、今では日本最大の社会調査データアーカイブとなっている。その後、2004年には大阪大学で SRDQ (Social Research Database on Questionnaires) が設置され、SSM 調査のオンライン集計ができるアーカイブとなっていた<sup>9</sup>。2010年には、立教大学社会情報研究センター (CSI) による RUDA (Rikkyo University Data Archive) が設立されている。他には、労働政策研究・研修機構による「JILPT データアーカイブ」や慶應義塾大学パネルデータ設計・解析センターのように、自機関が行った統計調査のデータを公開する機関もある。

### 3.6. メタデータの整備と規格化

前項のデータアーカイブが設立される以前から、共同研究によって実施された大規模統計調査においては研究計画に携わった研究者間でデータが共有されることはあったし、また大学の学部・大学院教育で使用されることもあった。ただしそれらの情報は、①そのデータを使って書かれた論文内で「調査概要」として書かれたセクションの記載内容を読み取り、また②共同研究に携わった研究者間で作成・共有した冊子媒体の「コードブック」を読み取り、③SPSS など特定のソフトウェアに依存したフォーマットで研究者が相互にデータファイルを配布・共有していた場合には、当該ソフトウェアで開くことによってデータファイルの情報を得ていた。これらはすべて、1節にも記した「メタデータ」がばらばらな形式で記録されたものということになる。

ただし、それらには問題もあった。①論文内の「調査概要」で列挙される項目は論文の著者に依存し、調査標題、調査主体／代表者、調査資金、調査時期、調査地域、母集団、標本抽出法、標本規模、観察単位、調査モード、計画標本数、アタック数、有効回収数／回収率など記録された情報は多々あるものの、論文投稿誌の紙幅の限界のため省略される情報も多く、またどこまで深く書かれるか、記載のスタイルもまちまちという問題がある。また、②コードブックでは問番号、データファイルでの入力カラム（テキストデータの場合）、変数名、変数のコード（変数がとりうる値の変域）、各コードの内容、欠損値コード、非該当コードなどが書かれているものの、情報が網羅されていないこともある。また、③特定のソフトウェアのフォーマットで保存されたデータファイルは、調査データそのものの情報や各変数についての情報が得られるものの、ソフトウェアおよびバージョンへの依存が大きく、開発言語の変更などにより再利用できないバージョンの組み合わせも出て

きた。

したがって、統計調査データの再利用可能性のためには、これらの情報の項目および記述内容を第1節で記したようにメタデータ規格として標準化することが望ましい。しかも、広く使われる標準規格ないし国際規格（たとえば、第1節にも記したDDI）の仕様に則っており、データアーカイブの情報システムに対応していることが望ましい。メタデータ体系を国際的に規格化することにより、また技術的には統一規格のもとにメタデータをXMLで記述することで言語横断的な検索が可能になり、論文等の記述言語が違えども国際比較研究・国際共同研究を行うことが容易になる。

国際的なメタデータ規格に準拠することで、第1節にも記したFAIRの原則に則ることができる。すなわちデータが検索可能に（findable）、アクセス可能に（accessible）、相互運用が可能で（interoperable）、再利用が可能（reusable）になる可能性はより高まるだろう。さらにいえば、社会調査を実施してデータが出来上がってからこれらのメタデータ項目をいかに埋めるか、ということを考えるのではなく、社会調査の企画・計画（つまりは、社会調査データ生成のプランニングということになる）そのものを、これらのメタデータに準拠したフォーマットに沿って計画していくデータライフサイクル（data lifecycle）の観点で考えることが必要である。

### 3.7. 組織的・職務横断的な取り組み

これらデータアーカイブ運営・メタデータ整備の取り組みは、現在までのところ我が国ではもっぱら大学の研究者（＝教員）あるいはその所属機関・部局によって担われている。もちろん東京大学の社会科学研究所に附属する社会調査・データアーカイブ研究センターのように大きな組織のバックアップによって研究インフラの構築やデータ整備に取り組んでいる機関もあるし、それに比べれば規模はかなり小さいものの立教大学の社会情報教育研究センターのように大学のメディアセンターと密接な関係のもとに事業を行っている機関がある。しかしこれらの機関でも、データ共有技術の研究を専門とする大学教員が行っているのではなく、社会学や政治学、経済学といったデータを採取する学問分野を専門とする教員がいわば片手間に担っているという脆弱な状況は変わらない。

この分野における国際学会である IASSIST（International Association for Social Science Information Services and Technology）の大会には、大学機関のメディアセンターの情報システム担当職員や図書館職員（ライブラリアン）が数多く参加しており、大学教員が多数派というわけではない<sup>10</sup>。日本の研究機関もまた、これら組織的・職務横断的な取り組みによって専門人材の育成を考える時期にある。

### 3.8. データ・マネジメント・プランの要請

現在のようにいくつかのデータアーカイブが設立された後も、研究期間が終了した後に調査データをアーカイブに寄託するということが残念ながらまだ一般化されていない。それら調査データの保存と社会的共有に供するための準備や努力は、研究者自身の自発的な意思に依存している。アーカイブへのデータ寄託を通じた共有が進まない現実には、研究者にとっては類似の調査を行うことにより研究の効率化を損ない、研究助成機関にとっては重複投資となり、市民にとっては社会の共有財産であるはずの調査データが有

効活用されないということになる。

これについては、研究助成を受ける条件として、助成申請の際に研究期間終了後のデータの取扱い、つまりは調査データの保存と社会的共有といったデータ・マネジメント・プラン (Data Management Plan: DMP) の明示を義務付けることによってデータの共有化が促進されるだろう。既にアメリカでは NSF による助成が、イギリスでは ESRC による助成が、DMP の提示を必須としており、たとえば NSF では研究計画のプロポーザルには DMP に関する別添文書を提出することを要求している<sup>11</sup>。日本では国立研究開発法人科学技術振興機構 (JST) が研究資金を配分し実施する研究プロジェクトについて、DMP の作成と提出を研究者に要請している (国立研究開発法人科学技術振興機構 2017)。

### 3.9. 人文学・社会科学データインフラストラクチャー構築推進事業

なお現在、日本学術振興会が 2018 年度より「人文学・社会科学データインフラストラクチャー構築推進事業」という 5 ヶ年計画の事業を開始しており、人文学・社会科学分野のデータ共有・利活用を促進するインフラストラクチャー構築を推進している。もちろん目的は人文学・社会科学の振興を図るところにある<sup>12</sup>。これは東京大学社会科学研究所や慶應義塾大学経済研究所、一橋大学経済研究所など各拠点研究機関との連携、情報技術に関わる国立情報学研究所への委託業務を含む広範な事業であり、データの共有化、海外発信といった国際化、データ間の連結化など多様な範囲にわたるまさにオールジャパンの取り組みといえよう。ユーザーとしての研究者の立場からいえば、拠点機関が公開するデータを横断的に一括検索できる「データカタログ」を構築する計画が興味深いところである。

## 4. 質的データのアーカイブ化可能性 (高橋)

### 4.1. 質的データとは

社会調査には、量的調査と質的調査という二つの方法がある。前者は、数量的データを扱い、主として統計的研究に用いられる。統計的研究は、事象の少数の側面を全体の中から切り取った調査になる傾向があり、エクステンシブな特徴をもつ。客観的方法で抽出した大量の調査対象者からデータを収集し、計量的に分析をするこの方法は、観察可能性、測定可能性、再現可能性、反証可能性、一般化可能性、などを重視する実証主義の立場と親和性が高い。

後者は、非数量的データを扱い、主として事例研究に用いられる。事例研究は、事象の多数の側面を全体との関連において記述する調査になる傾向があり、インテンシブな特徴をもつ。このデータには、行為の動機や体験の意味付けなどの主観的意味も含まれるため、実証主義の立場との親和性は低い。

ポストモダン以降、質的調査の重要性が叫ばれている。その背景には、事実や言動の「意味」は所与ではなく、社会的文脈の中で構成されるとする「社会構成主義」の視座がある。事実や言動の意味は、主観的評価の相互作用の中で生成されるが、同時に、客観的構造との相関関係の中でもまた、絶えず変転し再構成されていく。量的調査に代表される仮説検証型の方法は、大勢的状況のマクロ的把握に適してはいるが、個別的状況のミクロ的把握とその多元的な意味理解には、観察的なフィールドワークが不可欠である<sup>13</sup>。加え

て、過去に収集された質的データの、適切な「保存」と「再利用」が可能になれば、多様化する現代社会の問題解明が容易になるに違いない。

#### 4.2. アーカイヴ化の現状と取組

質的データの「意味」は、状況依存的であると同時に、文脈に応じて構成される。また、説明がデータに適合しているかどうか、あるいは、データ収集や分析の過程が適切であるかどうかという、「信頼性と妥当性」の検証が、量的調査に比べると困難である。そしてこの二つの特徴こそが、質的データの「保存」や「再利用」の可能性を著しく狭める要因となっている。

「質的データ・アーカイヴ化研究会」は、2012年に、「質的調査データの管理・保存に関するアンケート」[小林 2014][小林 2019]を行った。質的調査活動に携わる研究者から131通の有効回答があり、その約97%で何らかのインタビューが実施されていた[小林 2019: 77-79]。データの記録方法は、9割以上が録音または録画であった[小林 2019: 79]。データの内訳は、写真や動画などの「映像資料」、日記、自分史、手紙、メモ、手記などの「文字資料」、そして、メール記録、ポスター、絵画などの「それ以外」に分類された[小林 2014: 115-116]。

小林は、「インタビューデータのジレンマ」として、①コンテキストから切り離されたデータの価値をいかに確保できるか、②データが生成された調査のコンテキスト記録が必要、③データ保存への社会的コンセンサスが不可欠、という三点を挙げている[小林 2019: 82-83]。「語りそのもの」以外の「現場の生き生きとした状況」、つまり、語り手の、語り口、表情、声色、あるいは、現場や語り手の社会的背景や諸問題、なども、データとして保存する必要があるが、状況の包括的保存は困難である。音声や視覚的データとして形象化されているものが状況の全てではない。

「信頼性と妥当性」の検証が困難であることは、質的データの「再利用」の可能性を狭める一因である。多くの質的研究では、対象選定、実査、記録、データ加工、データ管理、分析などのプロセスが見えにくく、多様な質的研究の方法の間に共通する「言語」がない[木戸 2016: 218-219]。このジレンマの克服を目指す、「コンピュータ利用」や「ミックス法 (Mixed Methods)」などの新たな試みも存在する[林 2015: 68]。前者は、「QDAソフト (Qualitative Data Analysis)」を用いたデータのコード化やカテゴリー化による概念抽出、あるいは、データ集合から何らかの意味ある情報を取り出す「テキストマイニング (text mining)」などである。後者は、量的調査のレビュー結果と質的調査のレビュー結果を組み合わせる方法である。

しかし、仮にこれらの方法でデータの信頼性や妥当性が担保されても、状況が織り成す意味世界を構成することは難しい。「保存」と「再利用」の困難性はそれぞれ独立の問題ではなく、相互に連環している、いわば「質的データの宿命」であり、克服は容易ではない。

#### 4.3. 展望と限界

海外では、データのアーカイヴ化を目指す取組が為されている。例えば、2018年の「IASSIST (International Association for Social Science Information Services and

Technology) 」会議では、データアーカイブ業務の取組や関連動向が発表された[朝岡他 2019: 33-36]<sup>14</sup>。しかし、質的データアーカイブに特化したミーティングでは、20名弱の人が集まったに過ぎず、ほとんどはライブラリアンであったという[朝岡他 2019: 33-36]。武田[武田 2009]は、質的データの保存や二次利用データアーカイブの機関を紹介している。質的データの二次利用による研究の手順は、①コアデータ確定、②オリジナル調査の理解と関連資料群の探索、③リサーチデザイン作成、④二次分析実施、となる。エセックス大学に本部がある ESDS Qualidata (Economic and Social Data Service - the Qualitative Data Archival Resource Centre) は、英国各地の公的機関が所蔵するコアデータのカタログを作成しており、研究者は必要な質的データを利用できるという。

①コアデータ確定については、「保存」と「再利用」の課題がかなり克服されている。例えば、Qualidata が管理するデータソースは、①ESRC (Economic and Social Research Council) から研究助成金を得ている研究、②クラシックスタディーズ (著名な研究者による著名な研究群で、かつては一次研究者が私的に保存していたもの)、③私立財団による助成研究、であるが、このうち①については、「すべての助成対象者は、研究終了後三か月以内に、コンピュータ処理可能データ、非対応データの両方を含めて、質的データセットの複製を Qualidata に提出しなければならない」という「データセット寄託規準」がある[武田 2009: 218-220]。このようなデータ収集システムが確立されれば、「保存」と「再利用」というデータの二次利用に伴う困難の多くは払拭できるだろう。

ただし、②オリジナル調査の理解と関連資料群の探索、及び、③リサーチデザイン作成、に関しては、本質的課題が残る。一次調査者と調査対象者が共振的に構成した、文脈と状況に依存する意味世界を、二次利用者が追体験することは容易ではない。その課題は、武田がコルティの分類を参照して整理した③リサーチデザイン作成に関する六つの方法にも現れている。

1. 歴史的な一次資料として扱う
2. 比較研究の素材として活用
3. 新たな概念で一次データを再解釈
4. 一次調査の調査方法を再検討
5. 一次調査の妥当性を検証
6. 教育目的の利用

[Corti et al.2007][武田 2009: 214-217]

これらのうち、1と3は、対象が「歴史的意義がある一次データ」に限られるため、データの二次利用に関する普遍的な方法とは言い難い。2はやや普遍性があるとはいえ、対象は類似でもテーマが異なれば必要なデータは同じではない。質的調査のテーマは状況依存的であるため、条件と結果が正確に対応する量的データのようにはいかない。4と5は、それ自体が研究テーマであり、データの二次利用とは言えない。

6には可能性がある。オープンサイエンスとは正反対ではあるが、大学内という閉じた世界でのみ利用されるデータとして、学内研究者が収集した質的データを保存し、質的研究の教材として再利用することは可能である。不統一な記録媒体による利用不可能性、個

人情報保護に関わる倫理的問題、相互利用を可能にする物理的な管理体系の構築、など、質的データの二次利用に関わる現代的課題も<sup>15</sup>、学内共通ルールを設定すれば克服可能である。

## 5. オープンサイエンスを支えるオープンエデュケーション（長沼）

### 5.1. オープンエデュケーションの潮流

オープンサイエンスという言葉は Nielsen（2011）が初出とされている。オープンサイエンスにおいては、当初は研究成果への広く容易なアクセスを可能にし、科学技術研究を効果的に推進することに重点が置かれていた。これらはオープンアクセスとオープンデータという概念であり、主として研究者間での知見の交流を促進する取組である。しかしながら、オープンサイエンスの理念としては研究者のみがその利益を享受することは望ましくなく、一般の市民もまた科学の発展に参画していくことが求められている。これがシチズンサイエンスであり、科学の民主化とも言い換えられる概念である。シチズンサイエンスにおいては、市民が科学研究のプロセスであるデータの収集、データの解釈、問題の定義づけ、科学者=研究者との協働まで、関与の仕方に複数の度合いがある。また、科学技術政策の決定にあたって市民が参与することもまたシチズンサイエンスに求められる要素である。このようなシチズンサイエンスの実践にあたっては、学術研究機関に所属している研究者のみならず、市民もまた科学技術について学ぶ必要があることから、教育もまたオープンにする（オープンエデュケーション）必要がある。

オープンエデュケーションとは、社会への参加と包摂性の拡大を意図した、開放性に基づく教育運動である。その具体的な表れとしては大学などの伝統的な教育機関のみならず、様々な団体や個人が教材や学習ツールを無償で提供することで、互いに教えあい学びあうコミュニティを形成し教育機関の枠組みを超えてだれでも自由に学ぶことができる学習環境を構築している。オープンエデュケーションは、もともとは 1960 年代後半の英国を発祥とする初等中等教育における教育改革の 1 つで、教師から生徒への一方通行の教育を改善することを目的とした活動であった。すなわち、学校において教室と教室の境目をなくし、生徒の興味を重視してより学習者が主体となる教育活動を目指していた。この概念が近年のインターネットの普及によって対象と領域を拡大し、教育を受ける機会をあまねく提供しようとする活動全般を指すものとなっている。

### 5.2. オープンエデュケーションの具体例

オープンエデュケーションの活動は教材などのコンテンツの公開とその共有（オープン教育リソース）、そしてそれらのリソースを用いて学びあうコミュニティの運営まで多岐に渡る。ここでは主要な例として、オープンコースウェア（OCW）と大規模公開オンラインコース（MOOC）について概説する。

オープンコースウェア（OCW）は、米国のマサチューセッツ工科大学（MIT）によって 2001 年に創始された、大学で提供する教材をインターネットで無償で公開する試みである。OCW では大学での正規授業の教材であるシラバスや講義ノート、講義の音声や映像に加え、課題やテストなどを取りまとめて教育コースとして提供している。しかし、これらの教材を閲覧したり学んだりしても単位や学位は与えられないこととなっている。

OCW の活動は MIT からコンソーシアムの形で全米、さらには全世界に広がり、日本においても 2013 年時点で日本オープンコースウェア・コンソーシアム (JOCW) に 34 大学が加盟していた。

大規模オープンオンラインコース (Massively Open Online Course(s), MOOC(s)) は、インターネット上で公開される誰でも受講できるオンライン講座を指す。上述した OCW を含むオープン教育リソースでは講師は一般的には存在しない。これに対し MOOC では講師が教授活動 (質疑応答や課題の採点など) を行うことが大きな違いである。MOOC の運営元は、ベンチャー企業 (Udacity、Coursera など) や非営利団体 (TED-Ed など)、大学コンソーシアム (edX、FutureLearn など) など様々である。MOOC の学習コースを受講し到達目標に達したと評価されると、認定証が交付される。この認定証は大学の単位とはならないものが多数であるが、大学が提供しているコースで認定料を支払えば大学の単位として認められるものも存在する。日本においては企業・大学の双方が参画する JMOOC が設立され、配信のプラットフォームである gacco が 2014 年 2 月に開設されている。

### 5.3. オープンエデュケーションと大学の関係

オープンエデュケーションは大学発の取組と大学外からの取組の両者が存在し、一様に議論できるものではないものの、知の拠点としての大学のありように一石を投じる可能性がある。ここではオープンエデュケーションと大学がどのように関わるのか、今後どのように関わっていくべきなのかについて述べる。

まず、豊富な教育リソースが大学外にも蓄積されていることから、大学の教育活動をリッチにするために利用することが挙げられる。一例としては大学の授業としては提供されていない (できない) 初歩の内容 (リメディアル教育) について、自習教材としてオープン教育リソースを用いることである。東京通信大学においても、プログラミングについて学外教材を組み合わせる「プログラミング超入門」が LMS 上に設けられているなど、このような利用方法は今後もありうる。また授業で扱う内容の範囲外の進んだ内容についても、学生の自主学習を支援する目的でオープン教育リソースやオープンアクセスの学術資源、オープンデータを用いることが考えられる。

続いて、広報のために教育リソースを提供することが挙げられる。オープンエデュケーションは広く社会に開放されているため、既存の大学のステークホルダーとは異なる属性の参加者に大学の存在を伝えることができる。高品質な教材や講義を提供することで、大学の教育をアピールすることにも貢献する。提供する教育リソースとしては、公開講座などの小さくパッケージされたものから、講義そのものを提供してさらに有償で単位を発行することまで複数の段階がありうる。

さらに、よりラディカルな発想としては他大学が提供するオープン教育リソースを積極的に利用して単位化していくことが考えられる。教養教育については各大学で個別に運用するのではなく、何らかのハブを設けて一元的に提供するという壮大な計画が構想段階ではあるが提案されている。日本においては放送大学をそのハブとして整備し、各大学の教養教育はすべて単位認定するような形で運用するのが現実的であると思われる。しかしながら、特に私立大学は各大学が独自の理念に基づいて建学していることから教養教育にお



いても他大学と共通化できない部分が多く存在すると考えられるため、実現には多くの議論が必要であろう。

#### 5.4. オープンサイエンスの実現に向けてオープンエデュケーションが貢献できること

ここまで述べたように、オープンエデュケーションはオープンサイエンスとは異なる出自を持ちながら、それぞれの理想の実現に向けてお互いの要素を必要としている。オープンサイエンスをより広めるための方策としては、サイエンスの世界の入り口として初等中等教育に向けたオープン教育リソースを整備することが必要と考えられる。既に文部科学省や学術振興会などが提供している教育リソースは豊富にあるものの、現状では必ずしもすべての児童生徒が恩恵を得られていない。科学的な考え方や態度を涵養するために、学校教育の中でオープン教育リソースを用いることが望ましいと考えられる。また、同様に成人に向けてのサイエンスコミュニケーションがより一層求められている。近年の政策においては科学的知見に基づいた判断が必要であることから、オープンエデュケーションが果たすべき役割が大きくなっていることを指摘して本節の結びとする。

#### 6. まとめおよび今後の展望（藤井）

本稿では、昨今人口に膾炙するようになってきているオープンサイエンスの動向を調査し、本学に所属する教員を始めとする研究者がどのように対処すべきかについて議論等することによりその道筋を探った。その目的や目標は当初のものとは変化している。当初の研究の目標は、AI や IoT、ビッグデータを駆使する新しい科学と共にあるオープンサイエンスとは何かを探ることであると考えていた。オープンサイエンスの中の大きい領域を占めるオープンアクセスは情報学の先端な領域では当然のように扱われ、研究データの預託を行うアーカイブセンターを本学にも設置することなどで、それは事足りるのではないかという目論みであった。

しかし共同研究体制を確立させる端緒において研究分野の異なる研究者同士が集まると、それぞれデータという名称に異なる価値や範囲が在り共通の土台がないことに気づいた。つまりそれぞれの研究データの構造や性質が、単にダブリンコアや DDI といった先人の制定したメタデータ等の拡張で済むものではなく、ましてや拙速にアーカイブセンターのプロトタイプを設置し利用したとしても、限られた利用に留まり本学の組織や機関として有用とはなりえないということである。

1, 2 年といった短い研究期間で実務的な狙いを定めることは無意味である可能性が高く、まずは集まった共同研究者の個々の研究分野でのオープンサイエンス動向を調査し、その上で出来ることを探っていく、地に足のついた目標を定めていくことが重要であろうと考えている。現在、4 名の共同研究者での勉強会、研究会という形態で進めているが、この輪を広げていくことも一つの方法であるかもしれない。

この共同研究を進めていくに従い既にこの大学においてはオープンサイエンスの取り組みと云えることが行われていると認識させられた。まず東京通信大学紀要は冊子体と共に東京通信大学機関リポジトリにおいて 2021 年 9 月現在第 3 号まで公開されている。この機関リポジトリでの JAIRO クラウド採用の経緯には、その時点ではオープンサイエンスという概念を特に注目していたわけではなかったが代表研究者の藤井も関わっていた。2021

年2月28日に開催した情報マネジメント・フォーラム2021(廣瀬・藤井ほか2022)も当にオープンサイエンスの取り組みであり、企画運営のメンバーには前述の藤井、共同研究者の廣瀬がおり、オープンデータを中心とした取り組みをその方面の第一人者に講演願ったものとなっている。また、毎年秋開催の東京通信大学公開講座は、所属教員が講師となり本学の教育や研究を広く一般に知ってもらうというシチズンサイエンスを体現する場と考えることができる。この取り組みには共同研究者の高橋が企画運営に携わっている。

これらのことから程度の差はあれ、ほとんどの大学教員はオープンサイエンス活動に参与しているともいえる。それら活動をオープンサイエンスへの知識を踏まえ意識レベルに引き上げることが重要であり、この共同研究のメンバーを核として引き続き研究を進めていくことを考えている。

## 文献

- 朝岡誠、高橋かおり, 2019, 「海外データアーカイブの動向(2)IASSIST 年次大会の報告から」『社会と統計:立教大学社会情報教育研究センター研究紀要』5: 33-41。
- 廣瀬毅士・藤井稔也・宮坂敬造・浜日出夫・安部雅仁, 2022, 「情報マネジメント・フォーラム2021 実施報告:情報×マネジメントのフロンティア」『東京通信大学紀要』4(1): 301-316, 東京通信大学。
- 木戸功, 2016, 「NFRJ と質的研究 -質的データの収集と分析および公開に向けて」『家族社会学研究』28(2): 218-223。
- 小林多寿子, 2014, 「質的調査データの公共性とアーカイヴ化の問題」『フォーラム現代社会学』13(0): 114-124。
- 小林多寿子, 2019, 「オーラルヒストリーとアーカイヴ化の可能性:質的データ・アーカイヴ化研究会調査より」『日本オーラル・ヒストリー研究』15(0): 77-88。
- 国立研究開発法人科学技術振興機構, 2017, 「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」[https://www.jst.go.jp/pr/intro/openscience/policy\\_openscience.pdf](https://www.jst.go.jp/pr/intro/openscience/policy_openscience.pdf) (2021年10月31日アクセス)
- 佐藤健二, 2018, 『文化資源学講義』東京大学出版会。
- 佐藤博樹, 2012, 「実証研究におけるデータアーカイブの役割と課題:SSJ データアーカイブの活動実績を踏まえて」『フォーラム現代社会学』11(0): 103-112。
- 武田尚子, 2009, 『質的調査データの2次分析:イギリスの格差拡大プロセスの分析視角』ハーベスト社。
- 林真帆, 2015, 「質的データを用いたソーシャルワーク研究に関する一考察:事例研究法に焦点をあてて」『別府大学紀要』56: 65-74。
- 原純輔, 2001, 「文化資源学の創設をめざして」『社会情報』11(1): 9-26。
- リチャーズ, L., 2009, 『質的データの取り扱い』大谷順子、大杉卓三訳、北大路書房。
- Corti, L. and Thompson, P. 2007, "Secondary Analysis of Archived Data," in Gubrium, J. F. and Silverman, D. (eds.) *Qualitative Research Practice*, London: Sage.
- Nielsen, M., 2011, *Reinventing Discovery: The New Era of Networked Science*, Princeton University Press.

## 注

- 1 執筆順。執筆分担箇所は、藤井が第1節・第2節・第6節を、廣瀬が第3節を、高橋が第4節を、長沼が第5節を担当した。
- 2 立教大学社会情報教育センターの社会調査データの寄託先となりうるデータ共有機関である RUDA が構築されており、拡張したダブリンコアを利用している。なお、共同研究者の廣瀬はかつてこの事業に携わった。
- 3 これを量的調査と表現する場合もあるが、量的・質的とはそもそもデータの型についての分類であるため、第3節では「統計的」という語を用いた。
- 4 もちろんデータファイルにおいてコードに変換されていること自体が本質的なのではなく [コードに変換可能である] ことが重要であり、google フォームや Microsoft forms で出力されるデータファイルはコードでなく回答選択肢の文言そのものがインプットされている。一般的に作成・配布されている社会調査データファイルでは、容量圧縮のためコード値がデータファイルにインプットされる。
- 5 照会対象となっている世論調査の範囲は、①個人を対象とする調査であること、②調査対象者（母集団）の範囲が明確に定義されていること、③意識に関する調査であること、④対象者数（標本数）が 500 人以上であること、⑤調査事項の数（質問数）が 10 問以上であること、⑥調査票（質問紙）を用いた調査であること、である。
- 6 データアーカイブではおもに質問紙調査の個票データが収録されてきたが、近年、質的調査法（参与観察法、自由面接法）で収集された「語り」「観察記録」「資料」などのデータも収集、提供され始めている（第4節を参照）。
- 7 設立当初は social の語がなく「ICPR」であった。
- 8 設立当初は前身の Zentralarchiv für empirische Sozialforschung（ZA）であった。
- 9 SRDQ は、令和 2 年 3 月末をもって運用を終了したとアナウンスされている。  
[https://www.iser.osaka-u.ac.jp/survey\\_data/application.html](https://www.iser.osaka-u.ac.jp/survey_data/application.html)（2021 年 10 月 31 日アクセス）
- 10 共同研究者の廣瀬は 2011 年から 2016 年まで IASSIST の会員として年次大会に参加しており、この事実を確認している。
- 11 NSF に関しては <https://www.nsf.gov/bfa/dias/policy/dmp.jsp> を参照（2021 年 10 月 31 日アクセス）。
- 12 この事業については、廣瀬・藤井ほか（2022）を参照。
- 13 社会的弱者の立場に置かれている対象では、当事者も気づいていないような行動の意味や意識、人間関係などを捉えることが可能になる質的調査の重要性を指摘できる。
- 14 主な議題は、目的に合致したデータを得るための相談窓口となるべき「データライブラリアンの教育」、情報をいかにして研究データにするのかというデータ活用を教える「学生教育」、取扱注意のデータをいかに扱うかという「プライバシーの問題」であった。
- 15 オーラルデータは、デジタル技術の発展にともなう記録媒体の変転と多様化、近年の調査倫理や著作権と関わる問題、調査データの検証可能性や二次利用を含む公開の問題、データそのものの遺産的価値とその管理保存や公共性の問題など、さまざまな現代的課題に直面している[小林 2019: 77]。

藤井 稔也（ふじい としや）

廣瀬 毅士（ひろせ つよし）

高橋 玲（たかはし りょう）

長沼 将一（ながぬま しょういち）

東京通信大学 情報マネジメント学部 准教授

東京通信大学 情報マネジメント学部 准教授

東京通信大学 情報マネジメント学部 准教授

東京通信大学 情報マネジメント学部 助教

