

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 11 日現在

機関番号：14101

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700284

研究課題名(和文) 学術情報流通における研究データの役割に関する基礎的研究

研究課題名(英文) Fundamental Study of the roles and places of research data in scholarly communication

研究代表者

三根 慎二 (MINE, Shinji)

三重大学・人文学部・講師

研究者番号：80468529

交付決定額(研究期間全体)：(直接経費) 1,700,000円、(間接経費) 510,000円

研究成果の概要(和文)：本研究は、学術情報流通における研究データの役割を明らかにすることが最終的な目的である。高エネルギー物理学分野における研究データ提供状況の調査、研究データ関連の政策および研究データの利用・共有に関連する文献レビューを行った。その結果、データが公開されている論文の掲載タイトルは、高エネルギー物理学分野におけるコアジャーナルに集中していること、研究データ共有関連の方策はデータ管理からデータ共有を強調するようになっていること、研究者のデータが共有する際には、1) 個人的、2) 制度的、3) 技術的、4) 文化的・社会的要因が影響していることがわかった。

研究成果の概要(英文)：Purpose: The purpose of this research is to identify the role and place of research data in scholarly communication.

Method&Results: Two main research were conducted; 1) the ratio of making research data open in high energy physics research data database(HepData) and 2) extensive literature review on research data policy and data use/sharing practices by researchers. As a result, open research data in HepData concentrated in core journals of high energy physics, research data-related policies have changed their emphasis from data management to data sharing, factors affected data sharing practices can be summarized as 1) individual, 2) institutional, 3) technological and 4) cultural/social factors.

研究分野：図書館情報学

科研費の分科・細目：図書館情報学・人文社会情報学

キーワード：研究データ 学術情報流通

1. 研究開始当初の背景

eScience は、データ集中型科学、すなわち 1) 以前では不可能なレベルの量・速度・密度・詳細さでの研究データ生成と 2) そのより高度な検索、分析、視覚化を可能にするツールに特徴づけられる、新しい科学研究の形態である。たとえば、ヒトゲノムや気候変動が典型例である。このような eScience は、科学研究の伝統的パラダイムである実験型・理論型・コンピューティング型研究に続く、全く新しい第 4 のパラダイムとも呼ばれている。現在までに eScience を推進する政策的な提言や報告書が出されており、アメリカの Cyber-infrastructure 計画やイギリスの eResearch 計画等が実施されている。

eScience を特徴づける研究データは、研究のアウトプットであると同時に、学術出版物の執筆および今後の研究・学習のインプットでもあり、学術研究の基盤²⁾と言える。よって、研究データがいかに生成、入手、利用・再利用、流通、保存されているのかを理解することは、今後の eScience を遂行・支援するにあたって非常に重要であり基本的知識として求められる。eScience は、グリッドコンピューティングなどの技術的側面から議論されることが多い。しかし、現在までに研究データそのものはもとより、特に学術情報流通における研究データの役割はほとんど関心が当てられておらず、実証的研究がほとんど行われていないため実態がよくわかっていない。

2. 研究の目的

本研究は、学術情報流通における研究データの役割を明らかにすることが最終的な目的である。

3. 研究の方法

学術情報流通における研究データの役割を明らかにするために、以下の調査を行った。

(1) 高エネルギー物理学分野における研究データ提供状況の調査

英国ダーハム大学が中心となって提供している高エネルギー物理学分野の HepData データベースを対象に、Reaction データの提供状況調査を行った。

(2) 研究データの共有に影響を及ぼす政策、研究者の研究活動における研究データの利用および共有に関連する文献レビュー

2000 年から 2013 年に発表された文献を対象とした。科学研究以外の領域でのデータ共有、共有技術、科学研究でも単なる技術紹介を扱った文献をのぞき、残った文献の引用文献に含まれる該当文献を追加確認した結果、

合計で 331 件の文献を収集した。文献の種類は、原著論文、レビュー論文、ニュース・特集記事、エディトリアルとした。

4. 研究成果

(1) 高エネルギー物理学分野における研究データ提供状況の調査

調査の結果、主に以下の結果が明らかとなった。1) データ提供数は 2013 年までに約八千件が登録され、1950 年から 2000 年までは多少の増減があるが増加傾向にあったものの、2000 年以降は半数程度にとどまっていること、2) データが公開されている論文の掲載タイトルは、高エネルギー物理学分野におけるコアジャーナル (Physical Review Letters, Physics Letters B, Physical Review D, Nuclear Physics B, European Physical Journal C など計 10 タイトル) で約 7 割を占め、一部のタイトルに集中していることがわかった。

(2) 研究データの共有に影響を及ぼす政策、研究者の研究活動における研究データの利用および共有に関連する文献レビュー

研究データ共有方針

2000 年代初期以降、英米の政府、研究助成機関、大学、大学図書館を中心に、研究データの管理・共有を推奨・促進する一連の報告書が出され、インフラや関連技術の研究開発がなされてきたが、研究者のデータ共有に大きな影響を与えているものは、その前後に策定された研究助成機関および学術雑誌のデータ共有方針である。以下では、多くの文献で言及されている英米の研究助成機関による研究データ共有方針の動向についてまとめる。

a. 米国

米国では、国立衛生研究所 (NIH) と全米科学財団 (NSF) のデータ方針が大きな影響を与えている。特に、2010 年に公表された NSF の方針は、2011 年 1 月 17 日以降に提出される研究申請に対しデータ管理計画の提出を義務化するものであり、対象範囲の広さから米国の大学および大学図書館に対応を迫るものであった。

b. 英国

英国では、RCUK がデータ管理の牽引役となっている。“Expectations for Societal and Economic Impact”、“Policy and Code of Conduct on the Governance of Good Research Conduct”において、研究者および研究機関はデータ管理および保存の責任があることを指摘している。2011 年には、“Common Principles on Data Policy”を公表している。同方針では、公的助成を受けて

産み出された研究データを公共財と位置づけ、データの発見可能性、データ公表に関する法的、倫理的、商業的制約、データ作成者の先使用権、データ引用、費用対効果の高いデータ管理共有手法に関する原則を示している。これを受けて、英国の 8 研究会議と Wellcome Trust は、それぞれのデータ方針を策定している。

研究者による研究データ共有の実践

a. 研究者のデータ共有の実践

データ共有には 4 つの合理性 (Reproduce/verify, Serve public interest, Ask new questions, Advance research)があるとされており、データ共有を伴う論文の被引用数が増える可能性も指摘されている。実際、研究者間にはデータ共有を行う意志や“良いもの”という認識は見られる。しかし、現時点では、一般的には研究者間でのデータ共有の実践は進んでいるとはいえない。

先行研究における研究データの共有の意志および共有経験の割合をみると、顕著な傾向は見られるというわけではないが、1) 自然科学系の分野が中心であり人文科学系はほとんど見られないこと、2) データ共有の意思は非常に高い一方で、実際の共有経験はそれほど高くない傾向が見られるものの、その割合は分野によるが上昇傾向が見られる。論文が公表された後でさえも、データ共有は容易になされないことも示されている。

データ共有の要因

近年、データ共有の要因を明らかにすることを目的とした調査がいくつかなされており、それらに共通しているのは、大別して、共有の要因を 1) 個人的要因および 2) 制度的要因から説明しようとしている点にあるが、先行研究において、データ共有に影響を与える要因を、1) 個人的要因、2) 制度的要因、3) 技術的要因、4) 文化的・社会的要因の観点から整理する。

a. 個人的要因

個人的要因に関して多くの先行研究で指摘されているのは、データ共有に要する時間である。手持ちのデータを共有する際には、そのままでは不可能な場合が多く、フォーマット、範囲など共有可能な状態にするための付加的な作業が必要になることが多い。そのデータを出すまでに要した時間やコストも同様であり、これらが大きければ大きいほど、データは「自分のもの」という意識が強くなると考えられる。

一方で、自身のデータを共有することによって、科学の進展に貢献、オープンサイエンスの促進、研究の可視性の向上、共著や共同研究の機会につながるといった、研究の促進に関連する要因が言及されることも多い。

b. 制度的要因

制度的要因で多く指摘されている要因は、守秘義務、プライバシー、知財、特許などの法的な問題であり、研究データの共有に制度面から制限がかけられている。一方では、研究助成機関や学術雑誌によるデータ共有方針の制度的圧力も強制力を伴うものとして、データ共有を促進する要因として言及されている。逆に言えば、こうした方針が無ければ、共有が進まないということも考えられる。

c. 技術的要因

技術的要因として挙げられるのは、研究データの共有を可能あるいは促進するための技術の有無である。先に述べたように、研究データはそのままでは共有が不可能な場合が多い。共有を可能にするためには、その背景にある目的や制度に合わせた形態に、データを処理する必要がある。

具体的には、データの共有、交換、引用を可能にする標準の欠如が指摘されることが多い。他にも、データを保管し共有するためのデータベースの長期的維持、容易なデータ登録ツールの欠如があり、それに伴って共有データの量も、共有を左右する要因となる。

d. 文化的・社会的要因

ここでいう文化的・社会的要因としているのは、研究者集団において共有されている文化や規範のことであり、データ共有をすることが、研究者集団内でどのような位置づけになるかが問題とされる。

最も多く指摘される要因は、データ共有による誤解釈・誤用である。データ共有の結果、自分のデータが意図しない形で利用・評価されることの懸念は多くの研究者に共有されている要因と言える。これに類するものとしては、データ共有の結果、他者に出し抜かれる、成果公表の機会を喪失する、データの誤りが見つかる可能性も指摘されている。

次に頻繁に指摘されるのは、データ共有に対して何らかの報償、つまり学術的なメリットや承認が得られるかである。上述したように、データ共有には多くの時間・コストがかかるが、分野内の規範として、学術雑誌での成果公表と同レベルの評価がデータ共有に認められるかが問題となる。

(3) まとめ

以上のように、研究者のデータが共有する際には、1) 個人的、2) 制度的、3) 技術的、4) 文化的・社会的要因が影響していると考えられる。研究者のデータ共有は自然科学系の一部の分野で進展を見せつつも、未だその実態が十分に明らかになっていない。データ共有の要因の組み合わせは、共有を行う研究者個人が置かれている状況ごとに異なっており、促進あるいは阻害する決定的要因もまだよくわかっていないのが現状であると考えられる。研究者による研究データの実態

をより詳細に明らかにし、何が研究者のデータ共有を促進あるいは阻害しているのか、今後も継続して調査研究が必要である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 件)

〔学会発表〕(計 件)

〔図書〕(計 件)

〔産業財産権〕

出願状況(計 件)

名称：

発明者：

権利者：

種類：

番号：

出願年月日：

国内外の別：

取得状況(計 件)

名称：

発明者：

権利者：

種類：

番号：

取得年月日：

国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1)研究代表者

三根 慎二 (MINE, Shinji)

三重大学・人文学部・講師

研究者番号：80468529

(2)研究分担者

()

研究者番号：

(3)連携研究者

()

研究者番号：