

# 公共財としてのビッグ・データの倫理

## ：どの公共？誰の財？

リネット・テイラー

### 出典

Taylor, Linnet<sup>1</sup>. [2016] "The ethics of big data as a public good: which public? Whose good?." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083

### キーワード

- ビッグ・データ(big data)
- 公共財(public goods)
- 通話詳細記録(Call Detail Record: CDR)
- 同意(consent)
- 識別可能性(identifiability)

### アブストラクト

国際的な開発援助団体や人道支援団体は、デジタル・データを公共財として扱うべきだとますます主張している。これは、デジタル・データが、国家の統計データの不足を補ったり、緊急時を含めた、国家による介入に情報を提供する点で有益であることを理由としている。この主張を受けて、データ共有のための倫理的原則の確立に向けたガイドラインやフレームワークを議論する「責任あるデータ(responsible data)」運動が展開されてきた。しかし、この運動は、最も価値の高いデータを保有する人々、特に、中低所得国で収集されたデータに仲介者を介してアクセスできるようにすることに難色を示しているモバイル・ネットワーク事業者から支持を得られていない。本論文では、データ共有のための既存のモデルについて検討しつつ、「公共財としてのデータ(data as a public good)」を擁護する議論が、ビッグ・データを扱う企業の実態にどのように適合しているかについて評価を行う。データが持つ人道支援上の価値を理由としてデータの共有が必須であるとする功利主義的な主張に対

---

<sup>1</sup> リネット・テイラーは、オランダのティルブルフ大学法・技術・社会研究所に属する研究者である。統治や人間・経済開発の研究にデジタル・データを使用する際に生じる影響について研究を行っている。

して、私は、企業が保有するデータは、しばしば相反する権利、義務、主張を含むエコシステムを構成しているという考え方に依拠する。また、公共財としてのデータという考え方が示唆するパワー・ダイナミクスや、インセンティブの違いが、どのような仕方でデータ使用に関する特定の倫理的立場をアクターに採用させることになるのかについても検討を行う。

この記事は、テーマ別特集「データ・サイエンスの倫理的影響(The ethical impact of data science)」の一部である。

## 1. 導入

データ分析が進歩すると同時に、データの使用に伴う倫理やプライバシーの問題が議論されるようになった。特に、携帯電話からのデータが人道支援や国際開発の分野で利用されるようになるにつれて、「公共財としてのデータ(data as a public good)」という考え方が広く使用されるようになった。本論の目的は、携帯電話事業者が開発や人道支援のために携帯電話の通話詳細記録(Call Detail Record: CDR)を共有した2つの事例を用いて、「公共財としてのデータ」という議論がビッグ・データの現実に則しているかどうかを問うものである。

公共財とは、他者の財の使用を阻止することが困難であるという意味での非排他性と、財の使用が他者の使用する分の財の減少を招かないという意味での非競合性を備えた財のことである。ティルブルフ大学の法学者である Purtova は、デジタル・データを(1)人間、(2)プラットフォーム、(3)個人情報、これらのエコシステムから構成される「システム資源」として特徴づけ、知識が公共財であることを示す議論をデジタル・データの場合に適用することは不可能であると指摘している<sup>2</sup>。

本論は、Purtova の議論に依拠しつつ、データを公共財とみなすレトリックへの批判を行う。以下では、データの共有を図る企業の背景には、時として矛盾するような様々な要因が作用していることを示したい。また、データが公共財になりうるという主張には、一部を犠牲にして特定の勢力に権力を与え、データがポジティブな社会的影響を与える機会を減少させるような、データをめぐる策略や特定の力関係を下支えする恐れがあることも示したい。データ共有の際に作用している要因を探るため、本論では研究目的で携帯電話の CDR

---

<sup>2</sup> テイラーがここで依拠する Purtova の議論については、テイラーの議論と少し毛色が異なるので、補足したい。Purtova の見立てでは、個人情報がかつて財産権が保障されるべき資源とみなされず公共財とみなされてきた背景には、個人情報は我々が日々の生活を送る過程で生まれる副産物であり、その産出に労力を必要としないという考え方が存在する。換言すれば、この考え方は、個人情報をある個人についての均質的な情報とみなしている。こうした考え方に対して、個人情報をエコシステムの中の様々なステージで様々な形で具現化する情報と捉え直すことを Purtova は提唱している(Purtova, Nadezhda. "The illusion of personal data as no one's property." *Law, Innovation and Technology* 7.1 (2015): 83-111.)。

の共有を行った、2つの通信会社、ノルウェーのテレノール(Telenor)とフランスのオレンジ(Orange)の比較を行う。

## 2. 背景

2000年代に低・中所得国(LMICs)で携帯電話の保有者数が急激に増加して以来、これらの国々のCDRデータはビッグ・データとして解析され、デング熱やコレラなどの病気の蔓延のモデル化や、貧困や人口移動を地図上に可視化するために利用されている。CDRデータを利用した開発・人道支援の研究は、携帯電話事業者と公衆衛生の研究者との間で行われた単発の共同研究に始まったが、この研究を機縁として、より大規模でクラウドソースな形で研究用データの共有を目指す動きが始まった。

LMICsに関する新たな洞察を生み出すCDRの力が実証されたことで、CDRの利用に関心を持つ研究者、資金提供者、仲介者のコミュニティが生まれた。しかし、デジタル・データという新たな資源が持つ力に対する認識が高まっているにもかかわらず、データ共有について広く適用可能なアプローチというものは未だ現れていない。データ共有に対して公共財の観点からアプローチを行うコミュニティは、途上国のデータがより自由に流通するため、途上国の人々を保護するためのルールや規範が必要であるとの認識のもと、倫理的ガイドラインの作成に取り組んできた。

こうした事情を背景として、「責任あるデータ(Responsible data)」をテーマとするグループや議論が生まれている。こうして形成されたネットワークの特徴は2つある。1つ目は、議論の中心になっている機関が8つ存在しており、多極化が生じていることである。2つ目は、データ共有から利益を得る機関ほど、ネットワークをつなぐ中心的な役割を果たしていることである。もっとも、データの利用可能性の向上を求める当初の要求は、特に携帯電話事業者による大規模なデータ共有として実を結ばなかった。この事態を受けて、データ所有者がより安心してデータを共有することを可能にするような、一貫したコミュニティ基準を提示するための戦略として、倫理的なフレームワークの構築を行う議論が中心となった。

### (a)：責任あるデータをめぐる議論におけるCDRの中心性

CDRを用いた研究は、緊急事態や開発の現場で使用されるよりもかなり以前から行われていたが、開発や人道支援のコミュニティが広く関心を寄せるきっかけとなったのは、2010年に発生したハイチ地震でのCDRの使用であった。その後、CDRを使用した応用研究は、科学、人道支援、ビジネス、この3つの分野の関心を伴いつつ、徐々に数を増やし長期的に行われるようになってきている。こうした諸関心は、様々な権利を持った多くの異なるグループがデータ共有に関与することになる、民間のデータ・リポジトリの生態系的な性質によって生じている。3節と4節では、共同研究を行う際の問題点、特にプライバシーと倫理が共同研究を形成する上で果たす役割について検討したい。

### 3. 共同研究者を選択する：エコシステムの影響

2015 年末、ノルウェーの携帯電話会社テレノール社の研究部門では、外部の共同研究者とともに、CDR を使用して LMIC の貧困レベルを検出したり、ユーザーの収入や性別を予測したり、夜間の照明を示す衛星データと組み合わせて人口数を測定したりする研究が行われていた。同社はまた、CDR がデング熱の感染を予測できるという、パキスタンで実施した研究を発表したばかりだった。同じ頃、パリではオランジュ社が、セネガルの CDR データ・セットを用いたプロジェクト「開発のためのデータ(Data for Development)」の後継事業への資金援助を行っていた。またオランジュ社は、途上国のデータを共有するための次のアプローチとして、「開かれたアルゴリズム(open algorithm)」プロジェクト (OPAL) にも取り組んでいた。OPAL は、オランジュ社がデータを公開することなく、研究者がデータに対して質問を投げかけることを可能とするものだ<sup>3</sup>。

これらのプロジェクトは、携帯電話会社が自社のデータ・セットの価値をどのように考えているか、また、携帯電話会社がデータを共有するための新しい選択肢の開発をどのように進めているのかを示す良い具体例となっている。しかし、CDR データの共有には、データの保護や同意取得等について常にリスクが伴う。総合的に判断すれば、データを共有しない方が安全である。しかし、研究のための CDR の共有がますます増加している。では、なぜそのようなことが起こっているのか。

#### (a) ビジネス上のメリット

ビジネス上のメリットと社会的なインパクトという両方の観点から、総体として利益が得られると企業が判断した際に、CDR の共有が行われる。実際、我々がインタビューを行ったテレノール社の研究者は、メディア上のインパクトを重視したと述べている。また、デング熱の感染予測に関しては、他の疾病の疫学モデルを構築することも可能であったが、パキスタン政府との衝突を懸念して、デング熱が研究対象として選ばれた。オランジュ社が、「開発のためのデータ(Data for Development)」や OPAL で CDR 研究を推進する理由も、同様に利他主義とビジネス上の利害が混ざったものであった。プロジェクトの企画者は、データ・サイエンスにとって全く新しいテーマであるアフリカの開発支援に貢献していると研究者たちが思えば、彼ら/彼女らはより革新的で冒険的な仕事をするだろうと考えていたのである。

これらのデータ共有の事例では、利他主義だけではビジネスを始める理由として不十分であり、データを社外で共有することに伴うビジネスリスクを克服するためには、社会的価

---

<sup>3</sup> ここでの OPAL のプロジェクトとは、次頁での質疑応答モデルによる新しいデータ共有の方法を意味している。

値の認識が必要不可欠なのである。

#### (b) データと機会のマッチング

本論文のプロジェクトでインタビューの対象となった、オランジュ社やテレノール社をはじめとする事業者が行った共同研究には、すべて思いがけない偶然(*serendipity*)の要素が含まれていた。何かしらの偶然によってデータの価値を認識しない限り、通常、企業が外部の研究者と CDR を共有する方針を打ち出すことはなかった。こうした偶然は、企業経営者と外部の研究者の社会的なつながりであったり、テレノール社の場合のように、会議での出会いであったりするかもしれない。

共同研究の機会を生み出すもう 1 つの要因は、国家が置かれている状況(*the national context*)、より具体的に言うと、国が保有する補足データの存在（時には欠如）であった。テレノール社の研究者は、「行政が優秀でデータの共有に前向きであったパンジャール州政府によって公衆衛生情報が収集されている」ことから、デング熱が魅力的な共同研究の対象となったと指摘している。一方、オランジュ社の「開発のためのデータ」は、西アフリカのコートジボワールとセネガルという、西アフリカの 2 つの国を中心に展開されている。これらの国では、資源が不足しているために国家統計の収集が困難であり、国家統計の補完や代替がプロジェクトの存在理由の一部でもあった。

#### (c) データ共有のモデルとそれを下支えする戦略

以上の 2 つの考慮要因に加えて、データ共有は、企業がデータを伝達し共有するための適切な技術モデルを構築できるかどうかにかかっている。現在使用されている主な技術モデルとして、(1)リモートアクセス、(2)限定公開、(3)質疑応答モデル<sup>4</sup>、(4)事前に計算処理された指標や合成データの使用<sup>5</sup>、この 4 つが存在している。この 4 つのモデルのうち、オランジュ社の「開発のためのデータ」はデータの限定公開という形をとり、テレノール社とハーバード大学公衆衛生大学院の共同研究の場合は、リモートアクセスモデルが採用されていた。

この 2 つのモデルには共通の特徴がある。それは、いずれの場合も、生データにアクセスできるのは現地の会社だけだという点だ。この事実は、データ共有の観点から見たとき、現地の会社が大きな力を持っていることを示している。多国籍企業の本国以外から CDR を共有する決定を行う際に、かかる特殊な層が決定に関与するという事実は、これまでのデータ共有プロセスの分析では考慮されてこなかった。また、これらのアクターは時間が経って後

---

<sup>4</sup> 「どれほどの人数が都市 A から都市 B まで移動したのか」等の質問を入力し、答えをデータベースから受け取ることで、データにアクセスするモデルのこと。

<sup>5</sup> 生データではなく、電話回数等の生データから得られる指標や合成データを提供するアクセスモデルのこと。

から振り返るとその存在が見えなくなっている場合がしばしばあるため、考慮対象とする価値がある。

## 4. 倫理的問題

### (a)同意

ここで問題となるのが、消費者の同意と現地会社の同意という2つの同意の層である。携帯電話のユーザーは、下請け企業との最初の契約を通じて、非識別化されたデータが研究目的で使用される可能性があることに同意する。次に、下請け企業は、テレノール社やオランジュ社等がデータを使用することに同意しなければならない。こうした仕組みは、消費者側のインフォームド・コンセントが非常に広範なものとなり、データが再利用される際に契約の見直しを行うことができないという潜在的な問題を提起している。テレノール社の場合は、下請け企業からインフォームド・コンセントを得る際に、同意の対象をデータの1つの利用方法にできるだけ具体化し、新たな利用方法が提案された場合には再び確認を行うことで問題に対処している。この方法のデメリットは、下請け企業が同意プロセスにおいて個人の代わりに務めることである。メリットは、使用用途の制限が順守されることである。

### (b)識別可能性

データ共有の過程で生じる2つ目の問題は、識別可能性という概念の明確化である。特に、個人ではなく集団の情報の問題は、プライバシーにとって次の主要な課題になり得るのである。非識別化されたデータであっても、人々が集団やネットワークとして追跡できることを理由として、取り扱いに注意を要する状況を想定することは可能である。

デング熱のプロジェクトは、データやモデル自体を共有するのではなく、モデルから得られた結果を紙の地図の形で公衆衛生当局と共有することができたため、倫理的な観点から見ると、ほぼ理想的なケースであった。しかし、他の多くの疾病の場合、そうした疾病を追跡調査するためには、個別化されたデータが必要となる。たとえば、エボラ出血熱の場合、人々がどこを移動しているかだけでなく、何をしているのか、どのような接触をしているのかを知る必要がある。

## 5. 結論

CDRのようなデータ・セットは、公共のデータとして機能すべきであるという主張は、データは非常に重要なものであり、アクセスを制限することは事実上検閲を意味し、「データを使用しないことは、道徳的には本を燃やすことと同じである」という考えに基づいている。さらに、こうした主張が依拠するのが、倫理的推論についての功利主義的なモデルなのである。

本論では、「公共財としてのデータ」という議論がターゲットとなる企業から支持されな

い理由を理解するためには、当該の企業の観点を理解することが必要だと主張した。企業は、単なるリスク回避のためにデータを公開しないというわけではない。CDR が善と悪、両方の用途で使用されうるという問題に対して、企業は繊細かつ反省的な仕方で対処している。その結果として、データ共有のインフラと実践に対する、新たな倫理的アプローチの進展につながっているのである。加えて、企業によるデータ生産とは、様々なステークホルダーの権利や利害が交錯するエコシステムの間でもあるのだ。以上の事情に基づいて、企業には、低中所得国から得られたデータのコントロールを保持し続ける強いインセンティブが生まれる。というのも、これらのデータには、データが誤用された場合企業が被り得る多大な損失と釣り合うような、ビジネス上の様々な洞察やポテンシャルを含んだ資源として、高い価値があるからだ。

データは公共財であるべきだという主張を検討することが重要なのは、これがデータやデータが生み出すことのできるアウトプットに対する権力の主張だからである。データの共有と利用のための包括的な枠組みを見つけようとする戦いは、道徳的な必要性に基づいてデータに対する権利を主張する戦いでもある。企業のデータ・セットへのアクセスを拡大することは、企業が現在保持している権力の一部を仲介者や研究者に分配することになる。また、データ使用に関する一連の流れに、より多くの意思決定者と使用用途制限の緩和を組み込むことで、データを最初に生成した個人や、誤用の可能性によって最も影響を受けやすい人々から権力を奪うことにもなる。現在構想されている包括的なフレームワークは、法律や規制の手段がまだ存在していない低・中所得国において、そうした手段の実質的な代替案として機能することになるため、こうしたフレームワークに対する企業の抵抗は、データ倫理という新しい分野の発展にプラスの効果をもたらす可能性がある。こうした抵抗は、企業データの新たな利用を監査・承認できる独立した公的機関が存在しないという制度上の欠陥を浮き彫りにしている。独立した監視機関の存在なくして、低・中所得国のデータの自由な流通を期待することは現実的ではないのである。

(福家 佑亮)