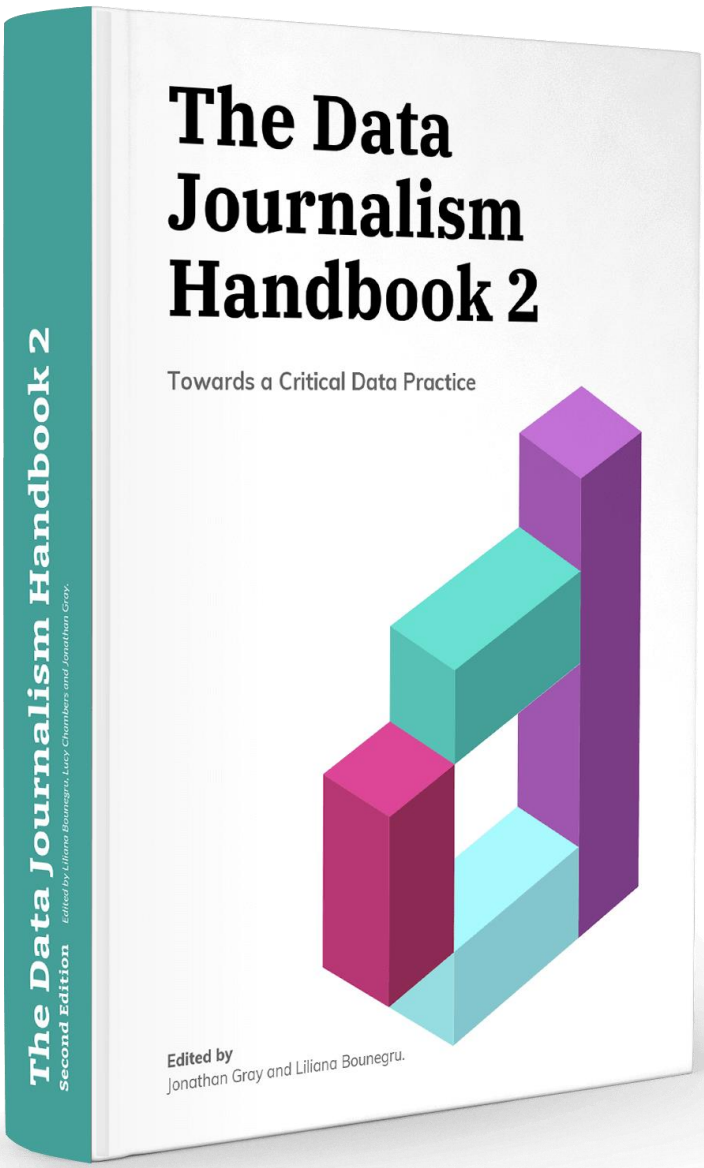
データジャーナリズム・ハンドブック 2



Google News Initiative　DataJournalism.com　European Journalism Centre

目次

01. はじめに
02. コーヒーから植民地主義まで：貧困国が富める国を養う姿を追うデータ調査
03. 国勢調査データを再利用して米国の人種差別を測定する
04. ボゴタの樹木を探訪し、それにまつわる思い出を喚起する
05. 数字の裏にあるもの：占領下の東エルサレムでの住宅解体
06. フィリピンでの交通安全の提唱に向けた事故事例のマッピング
07. トルコにおける労働者の死を追跡する
08. 独自のデータセットを構築する：英国におけるナイフ犯罪の記録
09. 数を語り、困難と共にあること
10. 先住民のデータ主権:データジャーナリズムへの影響
11. 中国における新方式のデータ・プラクティス
12. インド全土の土地紛争を記録するデータベースの作成
13. キューバの公共データを再構築する：情報が欠落、未更新、不足している場合のコラボレーション
14. ラ・ナシオンで読者とともにデータを作る
15. 調査報道のためのアンケート
16. データジャーナリズム：フェミニズムは IT とどう関係する？
17. パナマ文書とパラダイス文書を巡るコラボレーションの構築
18. データとしてのテキスト：収集したテキストからストーリーを見つける
19. ニュースルームでのデータを使ったコーディング
20. データジャーナリズムの手法を説明する：スプレッドシート、スクリプト、プログラミング［統合開発環境の Jupyter］ notebook
21. データジャーナリズムにおけるオープンな働き方
22. 報道に役立つアルゴリズムの構築
23. 機械とジャーナリズムの協働とは？計算論的思考から分散型認知へ
24. データジャーナリズムの態様
25. データ・ビジュアライゼーション:ニュースルームのトレンドと日常的な関わり
26. データをスケッチする
27. データ・ビジュアライゼーションの媒体としての Web

28. ニュース・グラフィックの四つの最新動向
29. ジャーナリズムの産物としての検索可能なデータベース
30. データとインタラクティブ・コミックで水紛争を語る
31. データジャーナリズムは、人とストーリーを重視すべし
32. アルゴリズムの専門性：視座と調査方法
33. ソーシャル・ウェブでストーリーを伝える
34. デジタル・フォレンジック：Google Analytics の ID を再利用する
35. データ調査のためのアプリとアフォーダンス
36. 脚光を浴びるアルゴリズム：シュピーゲル・オンラインでの共同調査
37. Twitter の#ddj ハッシュタグ
38. データジャーナリズムを記録する
39. ガーディアンから Google ニュースラボへ：データジャーナリズムに携わった 10 年
40. データジャーナリズムとシビックテックの連帯
41. データジャーナリズムにおけるオープンソースコーディングのプラクティス
42. データ封建主義：プラットフォームはどのように国境を越えた調査網を形成するか
43. データ駆動型編集って？視聴者測定基準を使う際の注意点
44. データジャーナリズム、デジタル普遍主義と周縁部のイノベーション
45. ジャーナリズムのデータ化：データ駆動型ストーリーテリングと産学連携戦略
46. 疎外された人々による、疎外された人々についての、疎外された人々のためのデータジャーナリズム
47. データジャーナリズムを教える
48. ラテンアメリカの女性やマイノリティとのデータプロジェクトを組織化する
49. データジャーナリズムの系譜学
50. データ駆動の規範：データジャーナリズムは何を評価し、社会のデータ化とともにジャーナリズムはどう進化していくのか？
51. クリックとシェアを乗り越えて:データジャーナリズムのプロジェクトのインパクトを測定する方法と意義
52. データジャーナリズム：誰の利益になるのか？
53. インパクトのあるデータジャーナリズム
54. データジャーナリズムは何のためにあるのか？キャッシュ、クリック、そして試行錯誤
55. データジャーナリズムとデジタル・リベラリズム

はじめに

文：ジョナサン・グレイ、リリアナ・ボーネグル

話題のデータジャーナリズム

データジャーナリズムとは何か？何のためのもの？何ができるのか？どんな可能性と限界があるのか？誰が、そして何がデータジャーナリズムを形作り、データジャーナリズムを理解することができるのか？

本書は、こうした疑問に応える共同実験であり、*The Data Journalism Handbook: How Journalists Can Use Data to Improve the News* [データジャーナリズム・ハンドブック、邦訳は https://datajournalismjp.github.io/handbook/]（O'Reilly Media, 2012）の続編になっています。どちらの本も、データジャーナリズムの発展しつつある分野を説明する、さまざまな意見と視点を収めています。「データジャーナリズム・ハンドブック」は 2011 年にロンドンで開催された Mozilla Festival の（ワークショップ）「book sprint」をきっかけに誕生しました。序文で述べたように、この本は「新たな動きの情熱と熱狂、展望とエネルギーを記録すること」「ストーリーの背後にあるストーリーを提供すること」「さまざまな意見や見解を輝かせること」を目的としています（Gray et al.、2012）。

12 年に出版されたこの本は現在、アラビア語、中国語、チェコ語、フランス語、グルジア語、ギリシャ語、イタリア語、マケドニア語、ポルトガル語、ロシア語、スペイン語、ウクライナ語など、十数カ国語に翻訳されており、多くの主要大学や教育・研修センターで使われているだけではなく、この分野の研究者もよく引用する資料となっています。この本は今でも広く使われていますが（本書はそれに取って代わるというより、補完することを目的としています）、12 年以降多くのことが起きました。

データジャーナリズムは定着しつつあります。11 年のデータジャーナリズムはほんの一握りの人々しか知らず、まさに「発展途上」の分野でした。その後、専門組織、トレーニングコース、求人、専門チーム、アワード、アンソロジー、ジャーナルの記事、レポート、ツール、オンラインコミュニティ、ハッシュタグ、カンファレンス、ネットワーク、ミートアップ、メーリングリストなどを通じ、社会に広まり、制度化されました。また、この用語は内部告発者のエドワード・スノーデンが「データジャーナリズム史上最大のリーク」（Snowden, 2016）と評したパナマ文書のように、データに関連した出来事によっても広く認識されるようになりました。

一方、データジャーナリズムはさらなる論争を呼んでもいます。13 年のスノーデンのリークは、憶測ではなく事実として、州やテクノロジー企業による国境を越えた監視体制をはっきりと知らしめました。これらのリークは、ビッグデータによって人々［の生活］をどのように知ることが出来るようになったのかを示唆しており、身近なデータ作成を伴うデバイス、アプリ、プラットフォームには暗い側面があることを示しています（Gray & Bounegru, 2019）。

米国では、ネイト・シルバーが 14 年に、データジャーナリズムに特化した FiveThirtyEight を立ち上げましたが、特定の定量的手法に偏重し、「オピニオン・ジャーナリズム」を軽視しているとして反発を受けました（Byers, 2014）。シルバーは、12 年の大統領選結果の予測に成功し、［米国の政治風刺番組］「ザ・デイリー・ショー」のジョン・スチュワートから「アルゴリズムの主にして神」と称賛されました。しかし 16 年のドナルド・トランプの当選後は彼の統計手法は批判され、異議を唱えられることになったのです。

これらの選挙は英国でのブレグジット投票や世界中でのポピュリスト的な右翼指導者の台頭とともに、「ポスト真実」の象徴とされています（Davies, 2016）。公的機関や専門家の知識、それらに関連する事実への信頼が広く失われ、オンライン・プラットフォームが公的、政治的な橋渡し役を果たすようになり、利用者はターゲティング、情報操作、誤報に対して脆弱になりました1。

いわゆる「ポスト真実」が失敗を裏付けるものか、それとも行動を促すものなのかは別にして、一つだけはっきりしていることがあります。データはもはやあって当然のものではなく、データジャーナリズムもまた然りということです。データは中立的で分かりやすい世界の表象を提供するだけではなく、政治や文化、金や権力と絡み合っています。調査から統計、気候科学、ソーシャルメディアのプラットフォームに至るまで、データ作成を支える制度やインフラが疑問視されています。

この原稿を書いている時点では、COVID-19 の流行が世界中で続いているため、数字、グラフ、ランキングは広く共有され、テーマ化され、政治化され、［または］非政治化されています。今や感染状況を示す平滑化曲線がどこでも目にすることができることがその象徴です。同時に、こうしたデータの脆弱性や暫定性についても広く報告され、症例の過少報告、非報告、分類に関する懸念や、病院の数値、調査による推定値、自己報告、追跡アプリの取引データなど、さまざまな情報源からのデータの社会的・政治的意味についての認識が高まっています。パンデミックは、数字やデータを単に利用するだけでなく、批判的に報道する意識を高めました。

このように、ジャーナリズムにおけるデータの活用が問われています。どのデータ、誰のデータ、どのような手段で生み出されたデータを使うのか？どのような課題について、どのような目的でデータを取得するのか？どのような種類の課題はデータ量が多く、どのような課題がデータに乏しいのか？誰がデータから恩恵を受けることができるのか？データはどのような公共性を持ち、どのような能力をサポートし、どのような政治を実現し、どのような政治参加を生み出すのだろう？

吟味的データ・プラクティスを目指して

本書では、そのような疑問や懸念についての議論を避けることなく、著名なフェミニスト研究者ダナ・ハラウェイが言うように「困難と共にある」（2016）ことを目指しています 2。私たちはデータジャーナリズムの関連性と重要性を自己主張するのではなく、複数の方法で対応できる問題として扱います。各章は、データジャーナリズ

ムとは何か、誰と、誰のために行うのかについて、より豊かなストーリーを提供することを目的としています。

私たちは本書の編集を通じて、データジャーナリズムのプロジェクトに可能なもの、そしてプロジェクトが成功するための条件として、熟考と謙虚さを推奨してきました。データジャーナリズムの実践を説明するにあたっては、それが発展し、運用される状況を特定するという異なる種類の精度を要します。このような正確さを担保するには、データがどのように分析され、作成され、どのようなジャーナリズムの文脈で使用されているかだけでなく、そのような慣行が埋め込まれている社会的、文化的、政治的、経済的な状況についても視野を広げる必要があります。

この本の副題は「吟味的データ・プラクティスを目指して」となっています。データジャーナリズムの実践に批判的な考察を加えたいという編集者の意気込みと、データジャーナリズムの実践者の批判的なスタンスの両方を反映しています。「吟味的データ・プラクティス」という概念はフィリップ・E・アグリの「吟味的技術実践」に由来するもので、彼はそれを「片足をデザインに、もう一方の片足は再帰的な批評に」と表現しています(Agre, 1997, p. 155)。他の所でも書いたように、本書の関心は、批判的にデータと関わることがデータ・プラクティスをどう修正し、データに基づく政治に対する公共の想像力と介在のためのスペースをどのように作り出すかを理解することにあります(Gray, 2018; Gray et al., 2018)。

(本書には)データジャーナリズムの実践に関するデータジャーナリストたちの寄稿に加え、人類学、科学技術研究、(新しい)メディア研究、インターネット研究、プラットフォーム研究、数量化の社会学、ジャーナリズム研究、先住民研究、フェミニスト研究、デジタル方法論、デジタル社会学などの分野の研究者による、データジャーナリズムの実践に対する批判的考察に役立つ章も含まれています。研究者が批判的に考察し、ジャーナリスト側が役立つヒントや助言を提供するといった、伝統的な分業にとどまらず、研究者は実用性を重視し、ジャーナリストには日々の締め切りに追われるだけではなく、何をすべきかを内省する空間を提供することを目指してきました。これらの異なる視点はこの分野を語り尽くすものではありません。私たちの目的は、データジャーナリズムがどのように実践されているのかについて複数の側面に読者の注意を引きつけることです。言い換えれば、本書は学際的な対話のきっかけとして、またコラボレーションの触媒となることをも意図しているのです。

私たちは、「データジャーナリズム」が統一されたプラクティスの集合を指すとは考えていません。むしろ、実証的に研究し、特定し、実験することができる多様なプラクティスの集合を指す、明確な標識なのです。最近、ある批評で指摘がありましたが、あるレビューが言うように、私たちは「事実だけではなく、[事実を導くための]数値化がどのように行われているのか」をただす必要があり、その影響は「意図と手段に依存する」のです(Berman & Hirschman, 2018)。私たちは、データジャーナリズムを固定するのではなく、むしろその多様な側面に注目し、異なる方法で実践するための可能性を拓きたいのです。

集合的な実験

本書が何のための本ではないのかについて簡単に述べておきましょう。従来の意味での単なる教科書やハンドブックではありません。各章は、確立された知識の集合体ではなく、さらなる調査や実験につながる興味深い方向性を示すことを目的としています。本書は単なるチュートリアルや「ハウツー」の実践的なガイドブックではありません。データ分析やデータ可視化など、データ・プラクティスのさまざまな側面に関する資料やコースはすでに無数にあります。単なる「裏技」的事例集ではなく、プロジェクトがどのように行われたかを示す記事やブログ記事、クリエイターへのインタビューなどが豊富に掲載されています。最近の学術論文ばかりではなく、あまたの書籍や雑誌に散在するデータジャーナリズムに関する新たに出現した文献にも目を配っています 3。

むしろ本書は、データジャーナリズムを説明しようとする実験の集積と、その実践がどのように修正されうるのかを探求する集合的な招待状としてデザインされています。「データジャーナリズム・ハンドブック」と同様、短い本としては比較的多い（70 以上の）寄稿者を擁し、またその寄稿者の助言を編集に役立てているので集合的と言えるでしょう。編集の過程では、電子メールでのやりとりの中で、寄稿者の方々からご推薦をいただいたことが役立っています。18 年にペルージャで開催された国際ジャーナリズムフェスティバルでは、多数の寄稿者によるワークショップが行われ、交流と考察の機会が提供されました。本書の「ベータ版」をオンラインで公開し、印刷本の出版前に一部の章を試読する場を設け、本書が最終的な形をとる前にコメントや出会いを引き出す機会を提供したのです。一種のキュレーションによる「雪だるま式編集」ともいうべきものによって、私たちは、さまざまな人々が、さまざまな場所で、さまざまなトピックで、さまざまな手段を使って、データジャーナリズムをどのように行っているのかを追跡してきました。その過程で、私たちは多くの候補作、ロングリスト、アウトレット、データセットに目を通し、データジャーナリズムの実践に関するさまざまな視点を集めました。より多くの寄稿者に参加してほしかったのですが、印刷可能な本という制約の中で運営する必要があり、性別、地域、テーマのバランスを考慮しなければなりませんでした。

本書は、各章がデータジャーナリズムについて異なる視点と刺激を提供する点で実験的といえます。ツール、データセット、方法、テキスト、書籍や出版物を組み合わせることで、読者にさらなる探求を促しています。公式データセットやソーシャルメディアのデータなどに見られるようなお決まりの見方を継承するのではなく、読者が自分の調査のためのサービスに参加することを推奨しています。これは、今問われている課題や問題の枠組みを修正することを目的とする「批判的分析」と「発明的な［問題解決の］手法」の精神に沿ったものなのです（Lury & Wakeford, 2012; Rogers, 2018）。データジャーナリズムは、物事がどう表現されているのかという観点だけでなく、関係性をどのように整理するかという観点からも見ることができます。そうした観点からは、（データの収集、分析、可視化、叙述）を通じてデータのストーリーを生み出すだけではなく、これらのストーリーが誰に何をもたらすか（視聴者や読者、情報源、方法、機関、ソーシャルメディアのプラットフォームを含む）にも注意を払う必要があります。ノーチェ・マーレスが最近述べたように、私たちは次のように問いかけることができます。「私たちは問題を別の方法で解決することができる空間を創造するために、どんなキュレーション方法、素材、技術、配置を取れば良いのか？」4。本書の各章では、データジャーナリズムがいかにして創意と想像力に富み、協働に適した技術でありうるかを示します。また、データジャーナリストが公式データの出典

に疑問を持ち、独自のデータを作成して編集し、新しい視覚的でインタラクティブなフォーマットを試し、効果を考察し、その方法を説明責任のあるものにし、コードを再現可能にする方法を提供します。データジャーナリズムの将来が不透明であるならば、読者の皆さんには、ジャーナリズムとは何か、これまでのジャーナリズムとは何かを批判的に評価し、その将来を形作る取り組みに加わってほしいのです。すべての作品と同様、本書の実験としての成功、失敗、そして究極の運命は、最終的には読者の皆さんが本書で何をし、どのような反応を引き出すかにかかっているのです。

本書の表紙は 、ルクセンブルク近代美術館にあるサラ・ジーの［日用品で構成したインスタレーション作品］「*Fixed Points Finding a Home*」の写真であり、アーティスト、彼女のギャラリー、美術館の許可を得て転載しています 5。作品をジャーナリズムに関する本の表紙に載せるのは変かもしれませんが、このイメージは、データジャーナリズムを一種のキュレーター的工芸品として捉え、多様な素材、コミュニティ、インフラを組み立て、異なるスケールと時間軸で世界を知り、語り、見る方法を生み出す関係性にぴったりだと考えたのです。（ビジュアライゼーションやインタラクティブのスクリーンショットといった）データジャーナリズムのアウトプットに焦点を当てるのではなく、データを使ったジャーナリズムに関わるさまざまな種類のプロセスやコレクティブを反映させたいと考えたのです。ルクセンブルク・ジャン大公美術館、ヴェネチア・ビエンナーレ、カールスルーエ・アート・アンド・メディア・テクノロジー・センター、［英国の］テートなどでジーの展覧会に偶然出合い、深く魅了された私たちは彼女の作品が、データジャーナリズムの実践に関する異なる（そしてできればあまり知られていない）視点を提供し、情報インフラや「データの集合体」に関する関係性の視座と共鳴するのではないかと考えたのでした 6。彼女のインスタレーションは、拾った素材を再利用するという正確かつ遊び心にあふれた感性を体現しており、私たちがこの本のためにデータジャーナリズムのさまざまな説明を編集する際に強調したかったことと視覚的に一致しています。ブリュノ・ラトゥールは最近、素材を組み立てるジーのアプローチは「構成的不連続性」を肯定するものと考えられるとしています（Latour, 2020）。これは、後述するように「吟味的データ・プラクティス」を奨励し、データジャーナリズムに関わる多様な素材と人々のストーリーを伝えたいという私たちの希望と、本書の各章の異なるスタイル、声、方言、興味を後押しする編集アプローチと一致しています。

本書のあらまし

設定に重点を置いた編集に忠実であるために、私たちは、この本の方向性と章の選択が、気候変動、環境破壊、大気汚染、租税回避、（新）植民地主義、人種差別、性差別、不平等、排外主義、権威主義、アルゴリズムの不公正、プラットフォーム労働を巡り増大する懸念を含む、現代における私たちと友人、同僚、ネットワークに影響を受けていることに留意しています。各章では、データジャーナリズムがどのようにしてこのような問題を理解しやすく、身近なものにするのか、またどのような反応を惹起できるのかについて検討します。

各章の選択には、私たち自身の学術研究、ジャーナリズム、アドボカシーの間の揺れ動きや、執筆スタイル、データ実践の違いも反映されています。私たちは、これらの異なる分野の仲間たちの出会いが生み出す可能性を確信しています。いくつかの章は、分野を超えたコラボレーションが成功したことを紹介しています。相乗効果や共通点を探るだけでなく、（鋭い読者ならお気づきでしょうが）各章で紹介されている視点には相違や緊

張、摩擦があることにも注目すべきでしょう。たとえば、異なる歴史や起源、手法やデータ、新しいテクノロジーに対する異なる見解、従来のやり方や異なるアプローチの実験の望ましい点に関する異なる見解、データジャーナリズムとは何か、何のためのものか、その条件や制約、組織化の方法、それがもたらす可能性に関する異なる視点などが挙げられます。

「はじめに」の後は、データで問題を解決するための「試食セット」が続きます。複数の国にわたる農産物、犯罪、汚職、植民地主義のつながりをたどるもの（Sánchez and Villagrán）、米国における人種隔離のマッピング（Williams）、ボゴタにおける木の記憶の追跡（Magaña）、東エルサレム占領下における住宅取り壊しの数字の背後にある人々と場面を見るもの（Haddad）、フィリピンで交通安全を実現させるための体制づくり（Rey）およびトルコにおける労働者の死亡の追跡（Dağ）といった実践例を紹介しています。この章の各章では、可視化のテクニックからキャンペーンの構築、Instagram 上のデータを使った読者のエンゲージメントまで、幅広い事例を紹介しています。

第 2 部では、本書と関連する研究（Gray et al., 2018; Gray & Bounegru, 2019）が焦点を当てた重要な新興領域である、ジャーナリストのデータ収集方法に焦点を当てます。これには、ナイフ犯罪（Barr）や土地紛争（Shrivastava and Paliwal）といったテーマに関するプロジェクトの制作を探ることや、キューバ（Reyes, Almeida and Guerra）や中国（Ma）など、データの入手が容易ではない国でのデータの入手や作業方法についての説明も含まれます。データを集めることは、読者と関わりを持ち（Coelho）、関心を持つ関係者を問題に引き込むことでもあり、それ自体がプロジェクトの重要な成果にもなりえます。データの収集は、世論調査やアンケート調査といった他の形式の知識生産を、ジャーナリズムの文脈に合致させることを含む可能性を開きます（Boros）。先住民のデータ主権に関する章（Kukutai and Walter）では、公式データをめぐる社会的、文化的、政治的問題や、データによる社会構成に周縁化された視点をどう取り入れるかを探求しています。データジャーナリストは、数字をストーリーの材料として使うだけでなく、数字がどのように作られるのかについてのストーリーも語ることができます。

第 3 部は、データを扱うさまざまな方法に関するものです。これにはアルゴリズム（Stray）、コード（Simon）、機械（Borges-Rey）が含まれる。また、テキストデータ（Maseda）のような情報源から生じる新たな問題や機会についても検討されている。また、データジャーナリズムの仕事を透明化し、説明責任を果たし、再現可能にするプラクティスについても考察します（Leon; Mazotte）。データベースはまた、大規模な調査プロジェクトにおける共同作業の機会を与えてくれるかもしれません（Díaz-Struck、Gallego、Romera）。フェミニストの思想と実践は、データを使ったさまざまな作業方法を触発します（D'Ignazio）。

第 4 部では、データジャーナリズムが取りうる態様を検討する（Cohen）ことから始め、データを体験するためのさまざまな方法を検討します。いくつかの論文では、現代の可視化の実践（Aisch and Rost）や、読者が可視化に対してどのように反応し、意味を理解するために参加するのか（Kennedy et al.）、また、データベース（Rahman and Wehrmeyer）、Webベースのインタラクティブ（Bentley）、テレビとラジオ（de

Jong)、コミック(Luna)、データのスケッチ(Chalabi and Gray)などを通して、データがどのように媒介され読者に提示されているかに注目する章もあります。

第 5 部では、データ、プラットフォーム、アルゴリズムを調査するための新たなアプローチを紹介します。最近のジャーナリズムのプロジェクトでは、デジタルはジャーナリストに新しい技術や機会を提供するだけでなく、新しい調査の対象にもなると考えています。その例として、Bellingcat や BuzzFeed News がバイラルコンテンツ、誤報、デジタル文化について広く共有する活動を行っています 7。この部の各章では、アルゴリズムに関するさまざまな報道のあり方(Diakopoulous)や、この分野で長期的なコラボレーションを行う方法(Elmer)などを検証しています。他の章では、ストーリーテリングのアプローチ(Vo)、デジタル手法研究とデータジャーナリズムの親和性(Webトラッキング・インフラの調査(Rogers)、アプリとその関連プラットフォームについての調査(Weltevrede))など、ソーシャルメディア・データの活用方法について考察しています。

第 6 部は、データジャーナリズムの組織化について取り上げ、この分野において不可欠だと考えられているにも関わらず、明確に認識されていない様々なタイプの仕事について述べています。データジャーナリズムはこの 10 年でどのように変化したか(Rogers)、プラットフォームとギグ・エコノミーが国境を越えた調査ネットワークをどのように形成したか(Cândea)、オープンデータとシビックテックに関する動きとデータジャーナリズムとの連係(Baack)、オープンソースのコーディング・プラクティス(Pitts and Muscato)、視聴者測定の実践(Petre)、データジャーナリズムの記録(Broussard)、Twitter 上のデータジャーナリズム・コミュニティをつなぐ#ddj ハッシュタグ(Au and Smith)などが含まれています。

第 7 部では、データジャーナリズムのトレーニング・プログラムや世界におけるデータジャーナリズムの発展など、共同作業としてのデータジャーナリズムを学ぶことに焦点をあてています。米国の大学におけるデータジャーナリズム教育(Phillips)、社会から疎外されたコミュニティが自分たちのストーリーを伝える力をつける(Constantaras;Vaca)、「デジタル万能主義」と、「周縁部」のイノベーションを過小評価することに対する注意喚起(Chan)、ジャーナリストと研究者のコラボレーションに対するさまざまなアプローチ(Radcliffe and Lewis)などの章が含まれています。

データジャーナリズムは真空のように孤立しているものではありません。最後の第 8 部では、さまざまな社会的、政治的、文化的、経済的環境の中に位置付けようと試みます。米国におけるデータジャーナリズムの系譜に関する章は、それを形成するさまざまな歴史的実践と思想についての考察を促します(Anderson)。他の章では、データジャーナリズム・プロジェクトが賞の選考でどのように評価されているか(Loosen)、データジャーナリズム・プロジェクトのインパクトを測定するさまざまなアプローチ(Bradshaw; Green-Barber)、データジャーナリズムと植民地主義を巡る問題(Young and Callison)、刺激的な実験、相互作用、遊びのフィールドになるという初期の願望にデータジャーナリズムは応えられるか(Usher)、データジャーナリズムとデジタル・リベラリズム(Boyer)などを紹介します。

吟味的データ・プラクティスに向けた 12 の課題

本書の開発を通じてデータジャーナリズムの実践を探求してきた時間を踏まえ、「吟味的データ・プラクティス」実現への 12 の課題を挙げて、本書の序章を結びたいと思います。これらは、異なるアクター間の関係を形成 し、世界についての表現を生み出す能力という観点から、データジャーナリズムを考察するものです。キングス・カレッジ・ロンドンとパブリック・データ・ラボが協働した「研究主導型教育」でテストされたもので、データジャーナリスト志望者や学生グループによるプロジェクトや調査、研究者とジャーナリストのコラボレーション、データが生成されるインフラ、環境、実践を当然視せずデータを使った集団的探求を組織しようとするあらゆる活動を促すものとして意図されています 8。

1. データジャーナリズムのプロジェクトは、データが作られる様々なアクター、プロセス、制度、インフラ、知識の形態を含むデータについて、いかにしてストーリーを語ることができるのか
2. データジャーナリズムのプロジェクトは、どのようにしてスケールの大きな問題（気候変動、不平等、多国籍企業の課税、移民など ）を伝えながら、数字の生成に関わるモデル、仮定、不確実性を認識し、その暫定性を確認すべきか
3. デジタル技術とデジタル文化の相互作用を含め、デジタルデータ、プラットフォーム、アルゴリズム、オンラインデバイスの集合的特性を、データジャーナリズムのプロジェクトはどのように説明できるか
4. データジャーナリズムのプロジェクトは、支配的な制度やインフラ、慣例からデータに「織り込まれた」知の方法を無批判に進めるのではなく、 データを通じて物事を理解し、意味付け、相対化する独自の方法をどうしたら育めるのか
5. データジャーナリズムのプロジェクトは、データ・ビジュアライゼーションと他の視覚的資料の組み合わせなどを通じて、どのようにビジュアル文化や美学を認め、実験することができるのか
6. データジャーナリズムのプロジェクトは、どのようにして、 一般市民の参加と介入の場を用意し、確立されたデータソースを疑い、取り扱う問題がどのようにデータによって説明されるかを再考することができるのか
7. データジャーナリストは、どのようにして データを扱う独自のスタイルを培い、意識的に確認することができるのか。それは、統計、データ科学、ソーシャルメディア分析といった分野を参考にしながらも、それとは異なるものだ
8. データジャーナリズムの分野では、自分たちの作品をアーカイブし、保存するための記録法をどのように発展させればよいのか、また、自分たちが利用している実践や文化との関連にどう位置付けることができるのか
9. データジャーナリズムのプロジェクトは、どのようにしてプラットフォームや植民地の論理を避け、周縁部でのイノベーションを肯定するような方法で、国境を越えた問題を中心に協力できるだろうか
10. データジャーナリズムは、どのようにして疎外されたコミュニティがデータを利用して、彼らのために［内向きの］物語を語るのではなく、［外部へ］自分たちの言葉で声をあげられるように支援することができるのだろうか

11. データジャーナリズムのプロジェクトは、どのようにしてソーシャルメディアの測定基準や他の分野で確立されたインパクト測定の方法論を超えて、世界における価値とインパクトを説明する代替的で独創的な方法を開発できるのか
12. データジャーナリズムは、どのようにして世界に介入し、社会における異なる主体間の関係を形成する上での自らの役割を最小化せず、肯定する客観性のスタイルを発展させることができるのか

謝辞

アムステルダム大学出版局、特にマリーズ・エリオットには、オンライン・ベータ版の出版や、オープンアクセスのデジタル版への支援を含め、この実験的プロジェクトに協力してくれたことに感謝します。数名の寄稿者が、2010 年にアムステルダムで開催されたデータジャーナリズムに関する最初の欧州会議の一つに参加したことを考えると、同出版局[の協力を仰いだこと]は適切な選択だったと思います。オープンアクセスの資金調達は、アムステルダム大学のリチャード・ロジャースが獲得したオランダ科学研究機構（NWO, 324-98014）の助成金で賄われています。本書の構想は、（学際的ネットワークの）Public Data Lab に関連する友人や同僚との議論を通じて生まれました。特に、アンドレアス・ビルバク、エリック・ボラ、ノーチェ・マーレス、リチャード・ロジャース、トマソ・ベンチュリーニ、エスター・ウェルトフレーデ、ミシェル・マウリ、ガブリエーレ・コロンボ、アンヘレス・ブリオネスと本のさまざまな側面を巡る議論から恩恵を受けました。 また、コロンビア大（ブルーノ・ラトゥールとの議論）、ユトレヒト大、カリフォルニア大バークレー校、スタンフォード大、アムステルダム大、マイアミ大、オールボー大コペンハーゲン校、パリ政治学院、ケンブリッジ大学、ロンドン・スクール・オブ・エコノミクス、カーディフ大、ランカスター大、ペルージャの国際ジャーナリズムフェスティバルでのイベントや訪問を通して、本書の方向性が定まりました。

キングス・カレッジ・ロンドンのデータジャーナリズム修士課程に在籍する大学院生たちは、本書の中核をなす「吟味的データ・プラクティス」の概念を検証するのに協力してくれました。また、European Journalism Centre と Google News Lab の支援を取り付けてくれたツバキ・リナのおかげで、待望の再版が実現しました。また、アダム・トーマス、ビアンカ・レメンス、ビバ・クロンプ、レティツィア・ガンビーニ、アルネ・グラウルス、サイモン・ロジャースは、編集の独立性と取り組みの規模を拡大するための息の長い支援を提供してくれて感謝しています。ダニエラ・デマルキとオアナ・ボーネグルの編集協力は、テキスト、脚注、参考文献、電子メール、共有文書、バージョン履歴、スプレッドシートなどの膨大な資料の中から明確な道筋をつけるために、大いに役立ちました。そして何より、（執筆、通信、議論などを通じて）このプロジェクトに関わってくださったデータジャーナリズムの実務者や研究者の皆様に感謝します。時間、エネルギー、資源を割き、またアイデア提供などによりこの試みを支えてくださったこと、それなくしてこのプロジェクトは実現しなかったでしょう。この本は、集団で取り組むべきものであり、これからもそうあり続けます。

謝辞

1. この用語に対する批判的な視点は、Jasanoff, S., & Simmet, H. R. (2017)を参照。No funeral bells: Public reason in a“post-truth” age. Social Studies of Science, 47(5), 751-770.

2. 本書のヴェランの章では、この作業を参照しながら、データジャーナリストが価値と数字の問題にどう対処していくかを探っている。

3. www.zotero.org/groups/data_journalism_research

4. ノーチェ・マーレスがランカスターで開催された EASST 2018 のプレナリー寄稿で投げかけた質問:twitter.com/jwyg/status/1023200997668204544

5. サラ・ジー,*Fixed Points Finding a Home*, 2012 (details). Mixed media. Dimensions variable. Mudam Luxembourg Commission and Collection. Donation 2012—Les Amis des Musées d’Art et d’Histoire Luxembourg. © Artwork:Sarah Sze. アーティストとビクトリア・ミロの好意により提供された。© Photo: Andrés Lejona/Mudam Luxembourg.

6. データインフラに関する関係性の視点については、例えば、Susan Leigh Star の代表的な著作を参照。: Star, S.L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces.*Information Systems Research*, *7*, 111-134; Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, *43*, 377-391.「データ・アッサンブラージュ」に関する最近の研究としては、例えば以下のようなものがある: Kitchin, R. (2014).*The data Big data, open data, data infrastructures and their consequences*. SAGE; Kitchin, R., & Lauriault,T. (2018).吟味的データの研究は: Charting and unpacking data assemblages and their work. In J. Thatcher, A. Shears, & J. Eckert (Eds.), *Thinking big data in geography: New regimes, new research* (pp. 3-20). University of Nebraska Press.

7. www.buzzfeednews.com/topic/fake-news, www.bellingcat.com

8. www.kcl.ac.uk/research/engaged-research-led-teaching and publicdatalab.org

参考文献

Agre, P. E. (1997). Toward a critical technical practice: Lessons learned in trying to reform AI. In G. Bowker, S. L. Star,

B. Turner, & L. Gasser (Eds.), *Social science, techni- cal systems, and cooperative work: Beyond the great divide* (pp. 130-157). Erlbaum.

Berman, E. P., & Hirschman, D. (2018). The sociology of quantif ication: Where are we now? *Contemporary*

Sociology, *47*(3), 257–266. doi.org/10.1177/0094306118767649

Byers, D. (2014, March 19). Knives out for Nate Silver. *Politico*. www.politico.com/blogs/media/2014/03/knives-outfor-nate-silver-185394.html

Davies, W. (2016, August 24). The age of post-truth politics. *The New York Times*. www.nytimes.com/2016/08/24/opinion/campaign-stops/the-age-of-post-truth-politics.html

Gray, J. (2018). Three aspects of data worlds. *Krisis: Journal for Contemporary Philosophy*, *1*.

Gray, J., & Bounegru, L. (2019). What a difference a dataset makes? Data journalism and/as data activism. In J.

Evans, S. Ruane, & H. Southall (Eds.), *Data in society: Challenging statistics in an age of globalisation* (pp. 365-374). The Policy Press. doi.org/10.5281/zenodo.1415450

Gray, J., Chambers, L., & Bounegru, L. (Eds.). (2012). *The data journalism handbook: How journalists can use data to improve the news*. O'Reilly Media.

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, *5*(2), 1–13. doi.org/10.1177/2053951718786316

Haraway, D. J. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press. doi.org/10.1215/9780822373780

Latour, B. (2020) Sarah Sze as the sculptor of Critical Zones. In B. Latour & P. Weibel (Eds.), *Critical zones: The science and politics of landing on earth* (pp. 158–159). The MIT Press.

Lury, C., & Wakeford, N. (Eds.). (2012). *Inventive methods: The happening of the social*. Routledge.

Rogers, R. (2018). Otherwise engaged: Social media from vanity metrics to critical analytics. *International Journal of Communication*, *12*, 450–472. dare.uva.nl/search?identi er=e7a7c11b-b199-4d7c-a9cb-fdf1dd74d493

Snowden, E. (2016, April 3). Biggest leak in the history of data journalism just went live, and it's about corruption.

Twitter. twitter.com/Snowden/status/716683740903247873

コーヒーから植民地主義まで：貧困国が富める国を養う姿を追うデータ調査

文：ラウル・サンチェス、ヒメナ・ビヤグラン

概要

発展途上国の農産業における違法なビジネス手法、持続的な環境破壊 、労働者の奴隷的状況を、データを使って明らかにしました。

キーワード： 越境調査、農業、植民地主義、データジャーナリズム、環境破壊

2016 年初頭、ジャーナリストの小グループが、チョコレートバーやバナナ、カップコーヒーが、農園から自分の机に届くまでの道のりを調査することにしました。私たちの調査は、これらの製品がすべて貧しい国で生産され、そのほとんどが豊かな国で消費されているという報道を受けて進められました。

そのデータをもとに、私たちはいくつかの問いを立てました。これらの農園の労働条件はどのようになっているのか？少数の地主に土地が集中しているのか？これらの製品は、地元でどのような環境被害をもたらしているのか？こで（独立した二つのデジタルメディアである）El Diario と El Faro が協力し、発展途上国の農業のビジネスモデルに潜む暗部を調査しました 1。「Enslaved Land［隷属の地］」プロジェクトは、国境を越えたデータに基づく1年間の調査で、「貧しい国はこうして富める国を養うために利用されている」とストレートな副題が付いています 2。

実際、このプロジェクトでは植民地主義が大きな問題となっています。ジャーナリストとして、体系的な検証抜きに、貧しい先住民の姿を伝えることはしませんでした。私たちは、土地の所有権、汚職、組織犯罪、地域紛争、特定の製品のサプライチェーンが、今でも植民地主義のシステムの一部であることを説明したかったのです。グアテマラ、コロンビア、コートジボワール、ホンジュラスの砂糖、コーヒー、ココア、バナナ、パーム油の五つの作物が欧米で広く消費されていることを明らかにしました。データ駆動型の調査として、データを使ってパターンからストーリーを導くことにしました。作物と国の選択は、過去の国連商品貿易統計データベース（Fig. 1.1）の記録 6800 万件の分析に基づいて行われました。

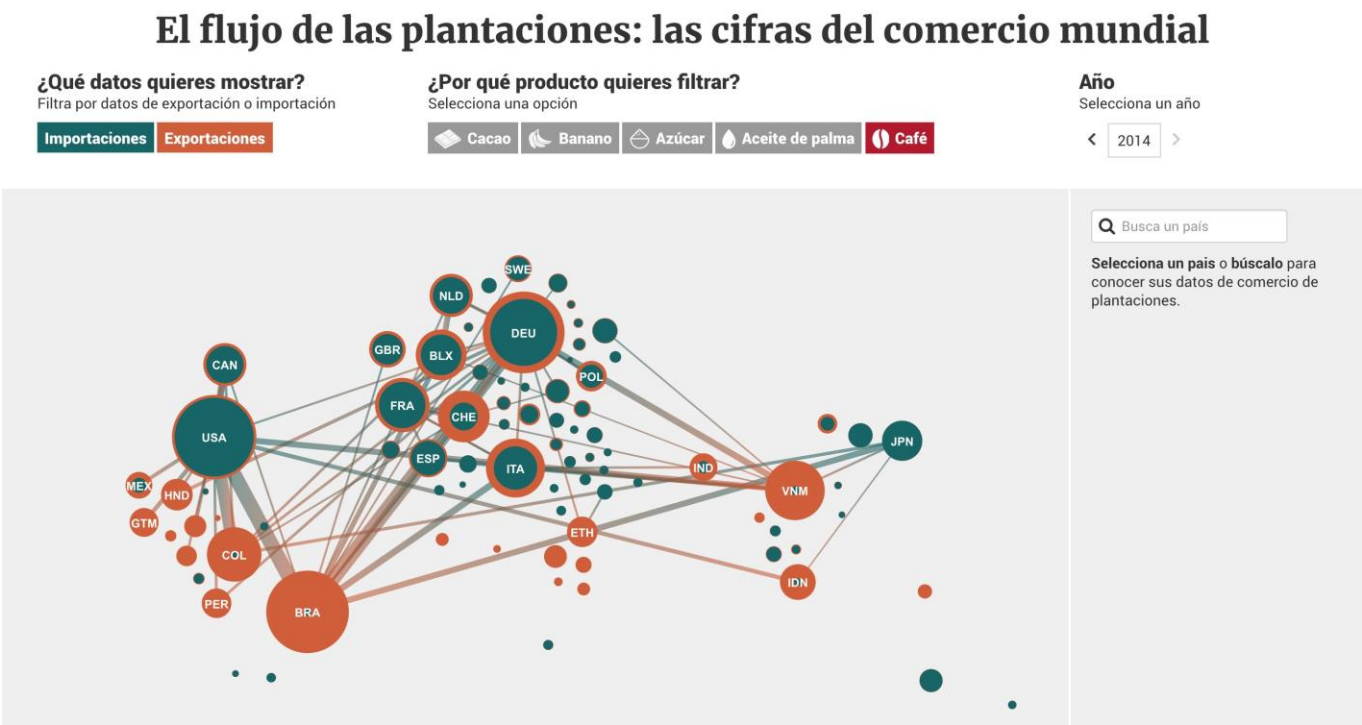


Figure 1.1. 2014 年の世界のコーヒーの輸入と輸出を示すネットワークグラフ 出典：eldiario.es.

この調査では、15 世紀から現在に至るまで、富裕国と貧困国の間のパワーバランスがどのように変化してきたかを示し、これらの作物が搾取的で奴隷のような労働条件、違法な商習慣、長きにわたる環境破壊の下に生産されていることを明らかにしました。こうしてデータに基づき、ストーリーの焦点が絞られました。 ホンジュラスでは、地理情報を使ってストーリーを伝えることに重点を置きました。（私たちは）国土の土地利用アトラスをまとめ、ヤシの植林地の表面と保護区を照合しました。その結果、国内の保護区にはパーム油（を採取するためのアブラヤシ） が違法に 7000 ヘクタール植えられていたことを突き止めました。私たちの記者は、保護区内のパームプランテーションがある地区を調べました。ホンジュラスのジャネット・カワス国立公園でアフリカヤシを栽培する農民「モンチート」の事例など、個々のケースを用いて制度的な虐待を浮き彫りにし、語りかけています。

このプロジェクト（が焦点を当てているの）は土地利用だけではありません。グアテマラでは、国内のすべての製糖工場のデータベースを作成しました。地元の会社の登記簿を調べて、工場の所有者や役員を把握しました。そして、パナマ、バージン諸島、バハマの公開されている企業情報を使い、これらの人物や事業体とオフショア企業を結びつけました。彼らがどのようにしてオフショア構造を作り、管理しているかを調べるために、El Faro はパナマ・ペーパーズのデータベースにアクセスしました。

第三世界の国々の汚職や企業の不正を発見するため、国境を越えた調査が課題となっています。政府の統治が十分及ばない地方で仕事をしなければならず、報道にはたいがいリスクがつきまといました。また、透明性

が著しく欠如している国、オープンでないデータ、場合によっては、どのような情報を持っているのか分からない行政機関などをめぐってもやりくりが必要でした。

ホンジュラスとグアテマラは調査のほんの一部です。10 人以上が協力してこの資料を作成しました。この仕事はすべて、スペインの El Diario とエルサルバドルの El Faro のオフィスを拠点に、コロンビア、グアテマラ、ホンジュラス、コートジボワールのジャーナリストと協働しました。ジャーナリストだけでなく、編集者、写真家、デザイナー、開発者などが、総合的な Web コンテンツを作る開発・制作プロセスに参加してくれました。このプロジェクトは彼らあってこそ成立したのです。

私たちは、それぞれの調査のために、統合されたスクローリーテリングの手法を使用しました。私たちにとって、ユーザーがどのようにストーリーを読み、どのように対話するかは、調査そのものと同じくらい重要です。読者が消費する商品と、その生産に関わる農家や企業、その他の関係者とのつながりを理解してもらいたいと考え、衛星画像、写真、データ・ビジュアライゼーション、そしてナラティブを組み合わせることにしました。

この構造により、個人のエピソードとデータ分析を組み合わせて説得力がある物語ができました。例えば、武力紛争中に準軍事組織やバナナ企業に土地を奪われたコロンビアの農民ジョン・ペレスの話では、彼の農園からコロンビアのバナナ生産の最終目的地までをズーム可能な地図で伝えました。

このプロジェクトは、データジャーナリズムが従来の報道手法をより豊かにし、個人に関するストーリーをより広範な社会的、経済的、政治的現象に結びつけることができることを示しました。グアテマラの Plaza Pública、チリの Ciper からも出版され、グアテマラのラジオ番組 ConCriterio にも取り上げられました。後者は、グアテマラ国税庁から、製糖工場の脱税を摘発するための情報提供を求められました。

脚注

1. www.eldiario.es（スペイン語）, elfaro.net（スペイン語）

2. latierraesclava.eldiario.es（スペイン語）

国勢調査データを再利用して米国の人種差別を測定する

文：アーロン・ウィリアムズ

概要

国勢調査データを使って米国の人種隔離を可視化します。

キーワード：プログラミング、マッピング、人種隔離、国勢調査、データ・ビジュアライゼーション、データジャーナリズム

どうすれば人種隔離を測ることができるのでしょうか。米国では建国以来、人種隔離の試みが続けられてきました。国（の姿勢）が変わり、人種隔離などの差別政策が禁止されると、アフリカ系米国人などを白人から引き離す法律が生まれました。多くの米国人がこれらの法律の影響を受けていますが、住んでいる地域によって（異なる）影響を測る方法があるのではないかと思いました。

ジェフ・チャンのエッセイ集「*We Gon' Be Alright: Notes on Race and Resegregation*[俺たち大丈夫：人種と新たなる隔離について]」に触発されました。このエッセイ集は、人種と場所を結びつけたテーマを追究しています。サンフランシスコ、ロサンゼルス、ニューヨークなどの人口動態の変化について書かれた章が印象的で、チャンが書いたアイデアを数値化するプロジェクトに取り組みたいと思いました。

人種隔離を示す地図の多くは、実態を反映していません。このような地図は大抵、特定の人種や民族のメンバーが地図上の点として表示され、その点がその人の人種によって色分けされています。これらは、人々がどこに住んでいるかという魅力的な人口地図を示していますが、これらの地域がどれだけ多様であるか、あるいは隔離されているかをうかがい知ることはできません。どうやったら分かるのでしょう。隔離と多様性という二つの言葉は人によって定義が大きく異なります。また、多くの人が自分の住んでいる場所は（人種的に）隔離されていると認識しているかもしれませんが、隔離をどのように測定するかによって、その答えは変わってきます。私はエピソードだけに基づいて行動したくありませんでした。 そこで私は、学術的な意味での「隔離」を測る方法を探し、そこから取材のベースを作っていきました。

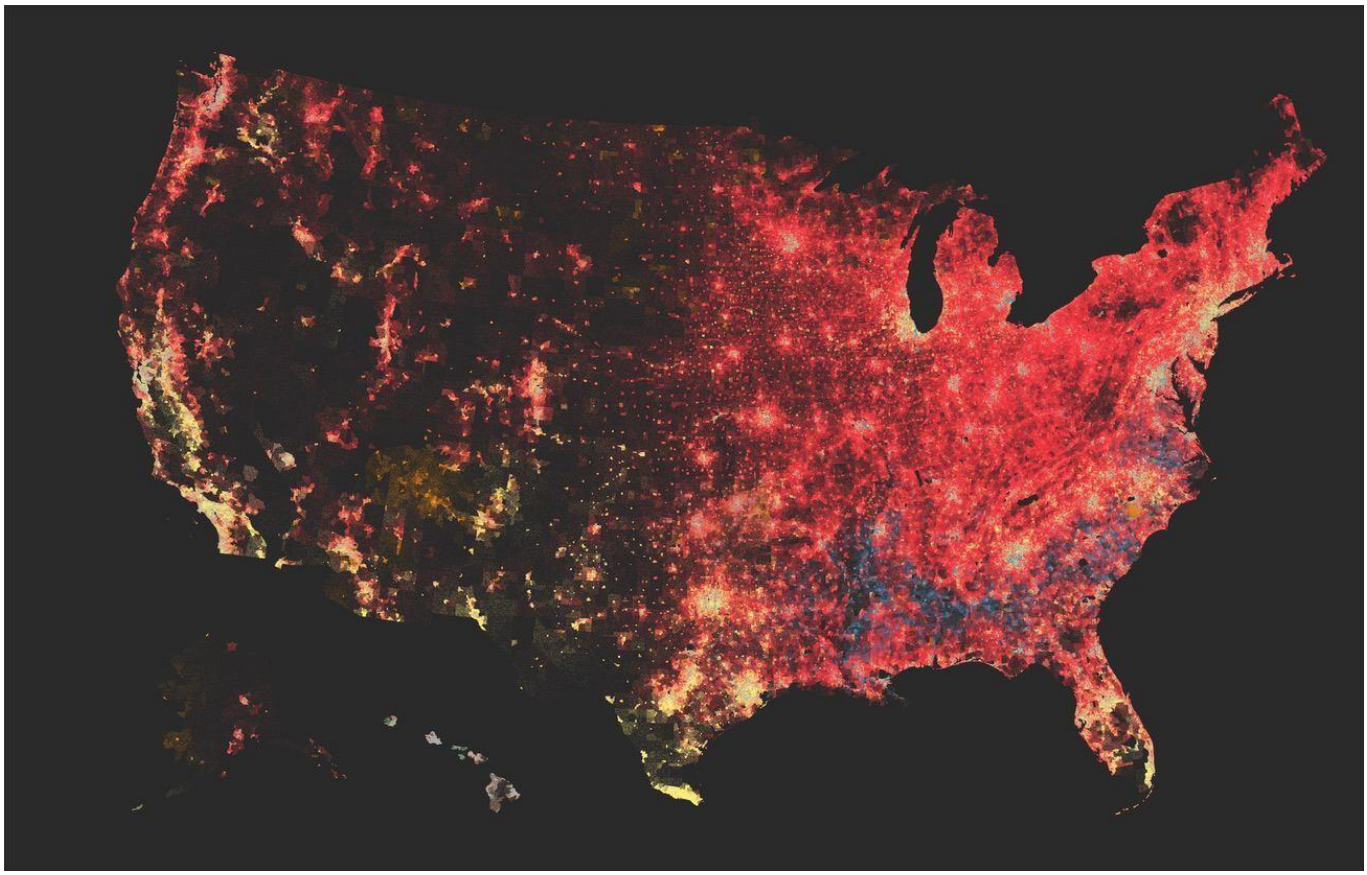


Figure 2.1. 国勢調査推計による米国内の人種のドット密度人口マップ(2018 年) 出典: ワシントン・ポスト https://www.washingtonpost.com/graphics/2018/national/segregation-us-cities/

ワシントン DC にあるアメリカン大学の社会学准教授、マイケル・ベイダーから、複数の人種グループの空間的分布を同時に判定する統計指標「多変量群エントロピー指数(タイル指数)」を教えてもらいました。私たちはこれを使って、アメリカのすべての国勢調査ブロックグループを、そのグループが属する郡の人種[別]人口と比較して点数化しました。完成までにおよそ 1 年を費やし、そのほとんどを、データや様々な隔離度の測定方法の調査に充てました。調査の過程で、偏在を測るにはいくつかの方法があることを知りました。多変量群エントロピー指数は均等性の指標で、ある地理的範囲内の人口の空間的分布を比較するものです。

また、「露出指数」のように、同じ地域で二つのグループが互いに接触する可能性を測る方法もあります。隔離を証明する、あるいは証明しない唯一の尺度はありませんが、これらの尺度は、コミュニティがどのように構成されているかを説明するのに役立ちます。私は、国勢調査の人口統計に関する多くの研究を読み、既存の文献を反映させようとしました。そこで、国勢調査局が実施している人種と人種差別に関する調査に基づいて、このプロジェクトに含まれる六つの人種カテゴリーを選択しました。そして、複数の人種グループを 1 回の分析で比較することができる多変量群エントロピー指数を選びました。そして、各国勢調査ブロックグループの構成

を、その周辺の郡の人種構成と比較することにしました。その後、同僚のアルマン・エマンジョメと、データ分析の原動力となるパイプラインの構築に数ヶ月を費やしました。

これまで、Python、R、SPSS などのツールで行われた人口統計調査を数多く見てきましたが、JavaScript を使ってこの調査ができるかどうか興味がありました。JavaScript と node.js エコシステムは、データを処理して Web に表示するための豊富なツールセットを提供しています。課題としては、分析関数のいくつかを手作業で書かなければならないことでしたが、その代わり、分析のすべてのステップを理解し、同じ関数をウェブ上で使用することができました。Mapbox と d3.js は、地理空間データを扱うための非常に強力で成熟したツールを持っており、分析の各段階で役立ちました。

公開の約 2 カ月前からは、ストーリーのデザインやレイアウトを何度も練り直しました。当初は、画面全体に地図を表示し、その上にテキストをスクロールさせる「スクロールテリング」という手法を採用するつもりでした。この手法は確立されており、[ワシントン]ポストの私のチームでもよく使われていますが、作成した美しい静止画像を全体的に見せることには不向きでした。最終的には、アメリカにおける人種差別や住宅差別の歴史を、三つの都市のケーススタディを交えて紹介する伝統的なストーリーレイアウトを選択し、その下に歴史的な[経緯をたどる]インタラクティブ・マップを掲載しました 1。

この記事は、私がジャーナリストとして発表したプロジェクトの中で最も多く読まれました。読者がストーリーの後にデータを探索することで、物語の中に自分自身を位置づけることができるパーソナライズの要素が加わったのだと思います。データジャーナリズムでは、言葉やアイデアを超えたストーリーを伝えることができます。読者をストーリーの中に直接招き入れ、彼ら自身に語らせることができるのです。

脚注

1.www.washingtonpost.com/graphics/2018/national/segregation-us-cities

ボゴタの樹木を探訪し、それにまつわる思い出を喚起する

文・マリア・イサベル・マガーニャ

概要

コロンビアのボゴタで、思い出を生み出し、透明性を向上させ、市民を物語に参加させるために、樹木に関するデータをどのように使用したか[について紹介します]。

キーワード： データジャーナリズム、シチズンシップ、透明性、オープンガバメント、マルチモーダルストーリーテリング、樹木

ボゴタは、わずか 1.775 平方キロの面積にコロンビアの人口の約 16％が集中しています。さぞごみごみして喧噪に満ちていると思うでしょう。しかし山に囲まれた緑豊かな街でもあり、いろいろな木が植えられています。木は市民生活の中で、意識に上ることはほとんどありません。しかし、私たちのデータチームのうちの1人は違いました。彼女は樹木が好きで、それを無視して街中を歩くことはできませんでした。彼女は、[通りにある]全ての樹種に詳しいのです。混沌とした都会における彼女の自然への愛が、私たちに気づかせてくれました。これまで、街中に植えられている木について話す人はいただろうか?

そして、その単純な質問が、多くの人々を呼び集めるきっかけとなりました。私たちは彼らについて何を知っているのか？誰が手入れを担当しているのか？街の汚染を浄化するのに本当に役に立つのだろうか？街にはもっと木が必要なのだろうか？豊かな地域だけに背の高い木があるのは本当か？街には歴史的な木があるのか？

日常的に目にする緑の「巨人」と市民をつなぐこと、そして市の緑化計画の実態を理解することの二つを目指して調査に着手しました 1。まず、植物園が 2007 年に作成したボゴタの植樹に関する都市センサスを分析しました。これは、毎月更新される唯一の情報セットです。植物園は、私たちが法に基づく情報公開請求を何度も提出したにもかかわらず、完全なデータを提供することを拒みました。彼らはそっけなく、「すでにデータは DataViz ポータルにあるじゃないか」といいました。私たちは「ダウンロードできるのは 1 万エントリだけで、データベースは 120 万エントリで構成されている。公共のデータなんだから、使えるようにして」と求めました。彼らの反応は「じゃあ 5 万エントリダウンロードできるようにアプリを改修しておきます」というものでした。

私たちの解決策は？植物園のデータ収集に協力したことがある他の組織に働きかけたのです。そのうちの一つが、市の土地台帳に関連するあらゆる情報を収集している Ideca[イデカ]です。彼らは直ちに全データセットを提供してくれました。私たちは当然、誰もがアクセスできるように、それを公開することにしました（これを「不透明性に対する私たちの小さな復讐」と呼んでいます）。植物園はこれに気づき、私たちとのそれ以上の論争を

やめたので、私たちは訴訟を取り下げました。また、ボゴタ市長室や国勢調査の公開データを盛り込み、樹木との関連で分析できる情報を相互参照できるようにしました。

最後に、環境の専門家や林業従事者へのインタビューを行い、市の課題を把握することができました。彼らは、緑化の実態だけでなく、街の木々の歴史を分析し、多くの調査を行っていました。 そして、そのほとんどは、当局やジャーナリスト、その他多くの人たちにはほとんど知られていませんでした。最終的に完成したのは、八つのデータプロジェクトで、都市の緑化計画の実態を示し、木の高さ、樹種、街への貢献度などの情報を盛り込みながら一本一本の木をマッピングし、植樹にまつわる多くの神話を覆し、街の歴史的な木の物語を実現したのです。

インタラクティブな操作を可能にするため Leaflet（地図データを扱うための JavaScript ライブラリ）と SoundCloud（音声ファイル共有サービス）を使用し、弊社の優秀なコーダーグループがデザインを担当しました。また、StoryMapJS（地図と連動してテキストや画像をスライド表示できる無償ツール）を利用して、歴史ある街の木々を探索できるようにしました。私たちは、他の類似プロジェクトを調査して、どの作品がこのストーリーにとって重要であるかを見極め、デザイナーと共に素晴らしい UX エクスペリエンスを作成しました。 これは私たちにとって初めてのビッグデータプロジェクトであり、試行錯誤と探究に満ちたものでした。さらに重要なのは、市民を巻き込んで、共同で樹木のカタログを作るのに協力してもらったり、マッピングした樹木に関する経験談を共有してもらったりしたことです。ソーシャルメディアを使って、樹種に関する情報をスプレッドシートに追加するように呼びかけました。

ボゴタの住民は、今日に至るまでカタログの充実に協力してくれています。また、WhatsApp の番号を共有し、人々が木にまつわる話を音声でメモを送ることができるようにしました。初めてキスをした木、登り方を教えてくれた木、泥棒から守ってくれた木、伐採されてしまって行方不明になった木など、100 件近くの音声メッセージが寄せられました。私たちは、この音声を可視化アプリの追加フィルターとして組み込むことにしました。その後、メインの記事と図表は（紙とオンラインの両方で）全国紙に掲載され、自分たちのストーリーを伝え、人々の環境との関係を変えたいと願う自治体や多くの市民に共有されました。今に至るまで、人々はこの地図を使って街の自然を調べたり、街の木々の調査に役立てたりしています。弊社にとって、これはこれまでに開発したプロジェクトの中で最も高度なものの一つでした。しかしそれは、データジャーナリズムが単なる数字以上のものになりうることを示している点で、最も価値のあるものの一つでもあるのです。文化や記憶を創造し、収集し、共有し、（グラフやチャートを超えて）人々が自分の住む場所についての気付きを与え、都市空間における人、植物、物語の関係を増幅し、変化させる役割を果たすこともできるのです。

脚注

1. especiales.datasketch.co/arboles-bogota

数字の裏にあるもの：占領下の東エルサレムでの住宅解体

文：モハメド・ハッダード

下のチャート(Figure 4.1)を見てみると、オレンジと黒のバーが並んでいて、2016 年には大きな伸びが認められます。キャプションを確かめると、このチャートが破壊された建物数とイスラエルの住宅解体政策の影響を受けた人々を示していることが分かります。Flowing Data の著者、ネイサン・ヤウが言うように、「データは現実の生活を抽象化したものである」(2013)。それぞれの数字は家族を、またストーリーを物語っています。Broken Homes は、イスラエルに半世紀占領されてきたパレスチナの東エルサレムで進行する住宅解体を追跡する、これまでで最も包括的なプロジェクトです 1。アル・ジャジーラは国連と密接に協力して、16 年に東エルサレムで行われた家の解体工事をすべて追跡しました。その結果、190 の建造物が破壊され、1200 人以上のパレスチナ人が避難したり、影響を受けたりした記録的な年であることが判明しました。私たちは 15 年末にイスラエル人とパレスチナ人の間で暴力が激化しているのを目の当たりにしてから、このプロジェクトに取り組むことにしました。その目的は、イスラエルの住宅解体政策が緊張の高まりによってどのような影響を受けるのかを理解することと、データの背後にある人々の物語を読者に伝えることの二つでした。このプロジェクトは、ビデオ証言、360 度画像、そしてそれぞれの解体が行われた場所、頻度、影響をハイライトするインタラクティブな地図を通して、パレスチナの家庭への影響を明らかにしています。

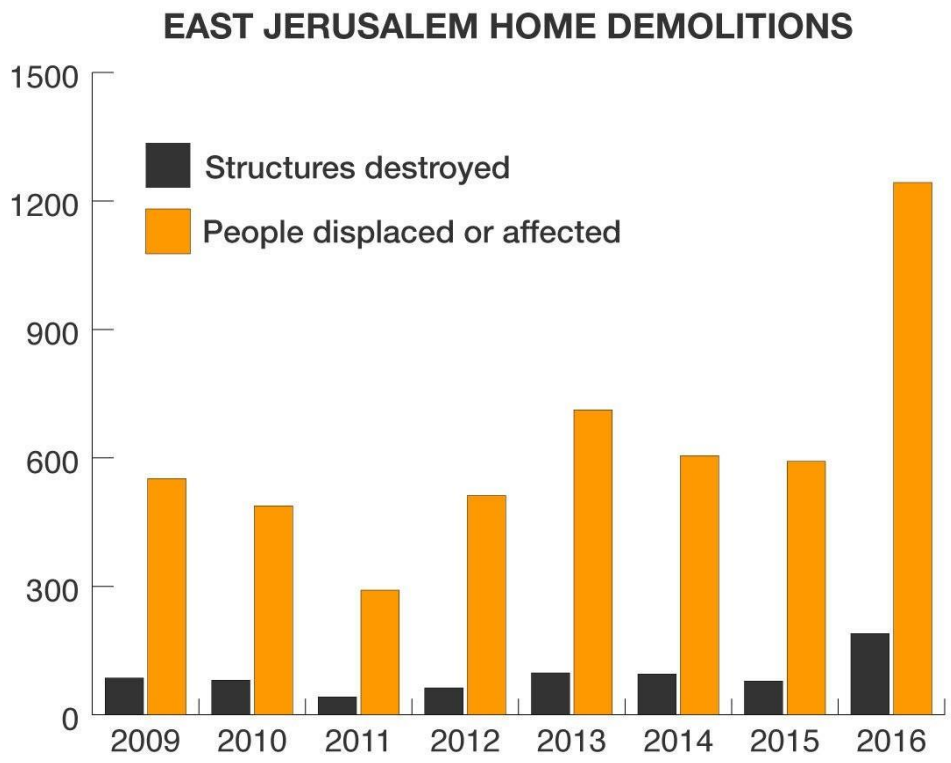


Figure 4.1. 東エルサレム住宅取り壊しを示すグラフ(2009 年～2016 年) 出典:アル・ジャジーラ

ドーハのプロデューサーは、2015 年後半から国連との調整を開始し、プロジェクトの枠組みづくりを進めてきました。国連は日常的に住宅解体に関するデータを収集しており、その一部はオンラインで公開されていますが、GPS 座標を含む他の部分は内部向けにのみ記録されています。私たちは、すべての解体現場を地図上に表示できるようにしたいと考え、国連から毎月のデータセットを入手し始めました。それぞれの解体について、取り壊しの日付、被害を受けた人や建物の数、何が起こったのかの簡単な説明、そしてその場所を示す東エルサレムの地図上のポイントが含まれています。

私たちは、これらの情報をニュース報道や住宅解体に関するその他の現地情報と照合しました。懲罰的なものから行政的な取り壊しまで、幼い子供から高齢者まで誰もが影響を受けるイスラエルの政策のさまざまな側面を示すものとして、私たちは毎月一つの事例を取り上げました。 現地の記者は、1年を通じて東エルサレムを訪れ、被災した多くの家族と話をし、彼らの損失をより深く掘り下げ、取り壊し現場を撮影、記録しました。

Figure 4.2. 2016 年 5 月に取り壊された住宅のパノラマ写真 出典:アル・ジャジーラ

被災されたご家族からの回答は多岐にわたりました。インタビューは解体現場で行われなければならず、被災者にとっては困難な経験となる可能性があるため、ミーティングの設定から資料の収録まで、すべての段階で繊細さと忍耐力が求められました。全体として、被災家族からの反応は上々でした。時間を惜しまず、経験を共有してくれたのです。ある例では、言いたいことのリストを書き留めていた男性がいました。別のケースでは、家

族を参加させるのに数回の説得が必要でした。ある家族が私たちとの面会を断ったので、国連と連絡を取り合い、家の解体について話してくれる別の家族を探したこともありました。

アル・ジャジーラを含む多くの報道機関は、長年にわたって住宅の取り壊しを報道してきました。今回、データ駆動型のアプローチをとった大きな理由の一つは、住宅解体を数え上げることで、ストーリーの規模を明確にしていくことでした。このような文脈と新鮮な視点は、現在進行中のトピックを報道する際には特に重要であり、読者の関心を引きつけることができます。データジャーナリストを目指す人へのアドバイス：データ駆動型のアプローチで記事を書くことは、技術的なことやお金がかかることではありません。ある出来事の発生を時系列順に辿り、数えるだけで、問題の規模がわかることもあります。データ収集の方法が一貫している限り、他の人が報道してこなかった、データを使って伝えられるストーリーはたくさんあります。そしてまた、忍耐力が試されます。このストーリーを伝えるために丸1年間データを集めました。

記者を現場に送る前に、必要なデータを徹底的にストーリーボードに載せることが肝要です。ほとんどの場合、特別な機材も必要ありません。私たちは、すべての 360 度画像を撮影し、特定の GPS 座標を取得するために iPhone を使用していました。17 年 1 月に英語、アラビア語、ボスニア語で発表されたこのプロジェクトは、イスラエルがパレスチナ人申請者の 98%に建築許可を与えず、増え続ける大量の人口に対する圧力を強め続けている中で警鐘を鳴らしています。

脚注

1. interactive.aljazeera.com/aje/2017/jerusalem-2016-home-demolitions/index.html

参考文献

Yau, N. (2013, June 28). Understanding data—Context. *Big Think*. bigthink.com/experts-corner/understanding-datacontext

フィリピンでの交通安全の提唱に向けた事故事例のマッピング

文：アイカ・レイ

概要

フィリピン・カガヤン州の交通事故に関するデータストーリーが、どのようにしてポジティブな政策と社会変革をもたらしたのか。

キーワード：データジャーナリズム、交通安全、コミュニティエンゲージメント、マッピング、フィリピン、政策と社会変革

データによると、フィリピンでは車両事故の死亡者数は、年々着実に増加しています。交通事故による負傷は、今やフィリピンの若者の死因のトップを占めているのです。

そこで私たちは、交通安全に関する関連情報をまとめたマイクロサイトを構築しました。市民に道路の安全性について知ってもらうため、データを収集・分析し、洞察を導き出し、記事を発表し、市民参加の機会をデザインしました 1。また、「Right of Way」と題して、マニラ首都圏の自動車運転手や通勤者の問題を取り上げたビデオシリーズの運営も開始しました。これが[ソーシャルニュースサイト]ラップラーの#SaferRoadsPH キャンペーンの始まりです。

交通事故死傷者数のデータを集めるのは大変な作業でした。交通事故の件数に関する全国的な総合データベースがないため、さまざまな市や省の警察署を含む 10 以上の国や地方自治体（の機関）をそれぞれ訪問し、データを入手しました。これらのリポジトリから取得したデータは標準化されていません。作業の大部分は、分析に使うデータのクリーニングです。大きな課題の一つは、位置情報が不完全であったり、全く記録されていなかったりするデータをどのようにマッピングするかということでした。

オープンソースのデータクリーニングアプリケーション OpenRefine を使って、異なる政府機関から入手した情報を正規化したデータベースを作成することができました。これにより、事故の発生場所、発生日、被害者の数を把握することができました。まだ不完全ではありますが、私たちが収集したデータは、現時点でフィリピンの交通事故に関する最大の一元的なデータだと言えます。それだけでなく、収集したデータに基づいてストーリーや分析、可視化を行った上で、それをネット上だけでなく、現場の活動で関係するコミュニティに直接提示しました。データ分析が市民参加型の活動につながったのです。

特筆すべきは、マニラの北約 600km に位置し、車両事故による死亡者数が最も多いカガヤン州に関する詳細な記事でした。カガヤン州の主要な役場を訪問して交通事故のデータを入手し、被害者や地元警察、

公共サービス関係者にインタビューを行いました。こうした活動の末、ラップラーは州都トゥゲガラオで交通安全意識向上フォーラムを開催し、調査結果を発表しました。このフォーラムでは、同州での交通安全について一般市民を啓蒙するとともに、政策ギャップに対処するために政府関係者に影響を及ぼそうとしました 3。

フォーラムでは、重大事故が最も多く発生する時間帯を示すグラフのほか、Google Fusion Tables を用いて作成した、カガヤン州で事故が最も多く発生する場所を示すヒートマップを発表しました（Figure 5.1）. 関係者は、これらの数字は歩道がないことなどが原因だと考えています。市内の学校を調べても、学校の前には歩道がありませんでした。フォーラム終了後、学校の前にチョークで歩道を描くという社会実験が行われました。学生が横断する際に自動車が止まるかどうかを確認しようとしたのです。ラップラーは、この実験についてのビデオストーリーを Facebook に掲載しました 4。

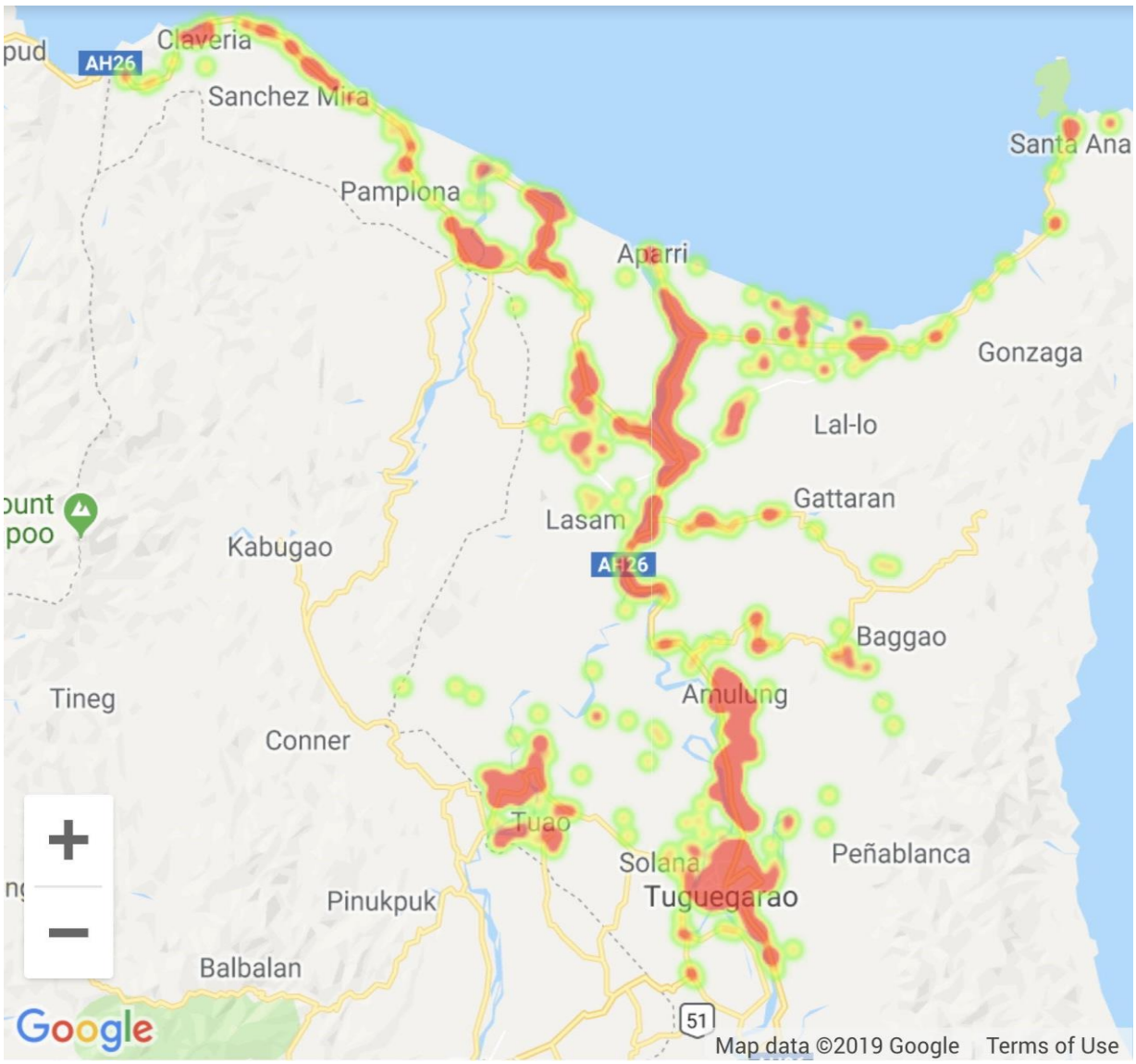


Figure 5.1. カガヤン州の濃い灰色で示された地域は、交通事故発生件数が多い 出典：ラップラー

このビデオは多くの関心を呼び、ラップラーの読者の中には、歩道用のペンキを提供してくれる人もいました。数か月後、自治体とボランティアが力を合わせて、市内の学校[の通学路]にようやく歩道ができました。塗装プロジェクトは 2017 年 9 月 30 日に完了しました。その 2 年後、市政府は交通安全に関する条例を承認しました。このプロジェクトは、データ駆動型報道は、編集者が公開[ボタン]をクリックしただけでは終わらないことを示しています。データジャーナリズムとオンライン・オフラインのコミュニティエンゲージメントを組み合わせることで、政策や社会にポジティブな変化をもたらすことができることを証明したのです。

脚注

1. www.rappler.com/saferroadsph

2. データソースについての詳しい説明はこちら: r3.rappler.com/move-ph/issues/road-safety/171657road-crash-numbers-data-sources

3. r3.rappler.com/move-ph/issues/road-safety/171778-road-crash-incidents-cagayan-valley

4. r3.rappler.com/move-ph/issues/road-safety/172432-cagayan-police-pedestrian-lane-chalk

トルコにおける労働者の死を追跡する

文：プナル・ダグ

概要

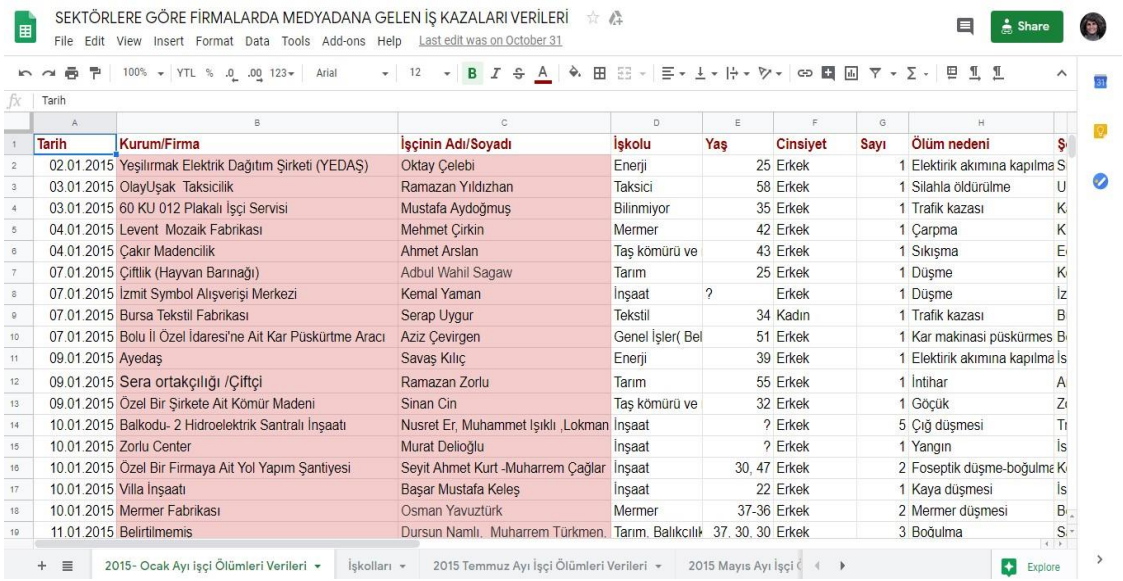
トルコでの労働者の死を記録し、労働条件の改善を提唱しました。

キーワード ：トルコ、ソマ、労働条件、データジャーナリズム、オープンデータ、情報公開（FOI）

2014 年にトルコで発生したソマ炭鉱の［爆発］事故をきっかけに、労働者の状況を記録することが非常に困難であることが判明しました。労働組合による数値とは矛盾があり、過去数十年間の労働者の死亡に関するデータも不足していました。

また、過去数十年間の労働者の死に関するデータも不十分でした。私たちは、このデータに広くアクセスできるようにして、他業界の労働者の死も明らかにしたいと考えました。そこで、プログラマーと編集者、そして私の 3 人で、データジャーナリズムのポータルサイト「Dağ Medya」が主催するプロジェクト「Open Database of Deceased Workers in Turkey［トルコにおける死亡労働者のオープンデータベース］」を立ち上げ、複数のソースからデータを集めて検証し、誰でもアクセスして利用できるようにしました 1。

トルコでは月に少なくとも 130 人の労働者がさまざまな原因で亡くなっています。このプロジェクトの最も重要な目的は、これらの死とその頻度に対する認識を高め、犠牲者と彼らが耐え忍んだ劣悪な労働条件を明らかにすることでした。このプロジェクトは、埋め込み可能な地図、グラフ、さまざまな形式のデータで構成されています 2。2011～14 年にかけ、20 業種以上の労働者の死を取り上げました。プロジェクト終了後も毎月、定期的なモニタリングを通じて、労働者の死を報告しています。このデータベースには、彼らを雇用していた企業の名前が含まれていることは注目に値します（Figure 6.1）。



SEKTÖRLERE GÖRE FİRMALARDA MEDYADANA GELEN İŞ KAZALARI VERİLERİ

Tarih	Kurum/Firma	İşçinin Adı/Soyadı	İşkolu	Yaş	Cinsiyet	Sayı	Ölüm nedeni
02.01.2015	Yeşilırmak Elektrik Dağıtım Şirketi (YEDAŞ)	Oktay Çelebi	Enerji	25	Erkek	1	Elektirik akımına kapılma
03.01.2015	OlayUşak Taksicilik	Ramazan Yıldızhan	Taksici	58	Erkek	1	Silahla öldürülme
03.01.2015	60 KU 012 Plakalı İşçi Servisi	Mustafa Aydoğmuş	Bilinmiyor	35	Erkek	1	Trafik kazası
04.01.2015	Levent Mozaik Fabrikası	Mehmet Çirkin	Mermer	42	Erkek	1	Çarpma
04.01.2015	Çakır Madencilik	Ahmet Arslan	Taş kömürü ve	43	Erkek	1	Sıkışma
07.01.2015	Çiftlik (Hayvan Barınağı)	Adbul Wahil Sagaw	Tarım	25	Erkek	1	Düşme
07.01.2015	İzmit Symbol Alışverişi Merkezi	Kemal Yaman	İnşaat	?	Erkek	1	Düşme
07.01.2015	Bursa Tekstil Fabrikası	Serap Uygur	Tekstil	34	Kadın	1	Trafik kazası
07.01.2015	Bolu İl Özel İdaresi'ne Ait Kar Püskürtme Aracı	Aziz Çevirgen	Genel İşler(Bel	51	Erkek	1	Kar makinasi püskürmes
09.01.2015	Ayedaş	Savaş Kılıç	Enerji	39	Erkek	1	Elektirik akımına kapılma
09.01.2015	Sera ortakçılığı /Çiftçi	Ramazan Zorlu	Tarım	55	Erkek	1	İntihar
09.01.2015	Özel Bir Şirkete Ait Kömür Madeni	Sinan Cin	Taş kömürü ve	32	Erkek	1	Göçük
10.01.2015	Balkodu- 2 Hidroelektrik Santralı İnşaatı	Nusret Er, Muhammet Işıklı ,Lokman	İnşaat	?	Erkek	5	Çığ düşmesi
10.01.2015	Zorlu Center	Murat Delioğlu	İnşaat	?	Erkek	1	Yangın
10.01.2015	Özel Bir Firmaya Ait Yol Yapım Şantiyesi	Seyit Ahmet Kurt -Muharrem Çağlar	İnşaat	30, 47	Erkek	2	Foseptik düşme-boğulma
10.01.2015	Villa İnşaatı	Başar Mustafa Keleş	İnşaat	22	Erkek	1	Kaya düşmesi
10.01.2015	Mermer Fabrikası	Osman Yavuztürk	Mermer	37-36	Erkek	2	Mermer düşmesi
11.01.2015	Belirtilmemis	Dursun Namlı, Muharrem Türkmen,	Tarım, Balıkçılık	37, 30, 30	Erkek	3	Boğulma

Figure 6.1. Google アラートによるメディアモニタリングに基づく、企業名と組み合わせたスプレッドシート
提供：プナル・ダグ

このプロジェクトは 2015 年に始まりました。まず、情報公開請求（FOI）を行い、さまざまなソースからデータを抽出して公開している信頼できる NGO からデータを収集しました。最初に直面した課題は、情報公開請求でオープンデータを入手するのが容易ではないことでした。時には 2 週間、またある時には 4 ヶ月もかかったこともありましたが、それでもデータを入手することができました。次に、もっと想定外の問題が発生しました。それは、この種の仕事に対する相反する見解が表面化したことです。このプロジェクトを発表したとき、データを使用したプロジェクトの一つである İSİG Meclisi（Health and Safety Labour Watch）は、データを使われることに不満を示しました 3。

自分たちが苦労して集めたデータを単に再公表しているだけだというのです。私たちが許可を求め、彼らのデータが公開されているにもかかわらず、彼らはこのプロジェクトに反対し続けました。彼らは、私たちがビジュアライゼーションと表を使って労働者の死を「ポルノ化」していると非難しましたが、私たちのプロジェクトは、彼らが収集したデータを視覚的にアクセスでき、簡単にダウンロードできるフォーマットで提供することで、公共の価値を生み出しています。 ヒューマンストーリーも重要ですが、こうした不正をより体系的に把握するためには、構造化されていない生のデータが重要だと考えました。しかし、この論理では合意を得るのが難しく、オープンデータを共有する価値を論証するのには苦労しました。結局、情報公開請求によって集められた公式データと、私たちがこれらのケースを独自にモニタリングして収集したデータを比較することで、毎月の労働者の死亡数を公表することにしました。

その後、私たちが情報開示請求を行った機関は、より構造的な方法でデータを共有し、それを視覚化するようになりました。これは、私たちの目標の一つである「情報に広くアクセスできるようにすること」が達成された証拠だと思います。このプロジェクトは 15 年のデータジャーナリズム・アワードの最終選考に残りました。

脚注

1.http://community.globaleditorsnetwork.org/content/open-database-deceased-workers-turkey-0

2. platform24.org/veri-gazeteciligi/451/turkiyede-isci-olumleri-veritabani-hazirlandi（トルコ語）

3. www.isigmeclisi.org（トルコ語）

独自のデータセットを構築する：英国におけるナイフ犯罪の記録

文：カエレイン・バー

概要

調査と強力なストーリーテリングに役立つデータセットを構築します。

キーワード：データジャーナリズム、犯罪、説明責任、人種、英国、データベース

2017 年初め、ガーディアンのニュースルームで、ゲイリー・ヤングとダミアン・ゲイルという同僚 2 人が声をかけてきました。彼らはイギリスのナイフ犯罪を調査したいと考えていました。（個別に）被害者の死を詳細に記した記事や、容疑者の追跡調査のフォローアップ、犯人の裁判や有罪判決に関する記事には事欠かないものの、全ての殺人事件を俯瞰的に見ている人はいなかったのです。私はまず、「近年、ナイフで殺された子供や若者は何人いるのだろう」と考えました。簡単な質問に思えましたが、データを見つけようとすると、誰も把握していないことがわかりました。データはどこかにあるのですが、公開されていないのです。私には二つの選択肢がありました。あきらめるか、自分でアクセスし、構築し、検証できるものをもとに、ゼロからデータセットを作るか。私はデータセットを作ることにしました。

なぜ独自のデータベースが必要か

データジャーナリズムは、既存のデータセットだけに基づく必要はありません。独自のデータを作成することには大きな意味があります。日常的に公開されていないデータや、場合によっては収集さえされていないデータには、豊富な情報が含まれています。独自のデータセットを作成すれば、独自の情報セット、つまり専用のソースを作成し、ストーリーを探ることができます。データとそれに続くストーリーは独占的なものになる可能性が高く、他の記者が簡単には見つけられないストーリーを掘り起こす競争力を与えてくれます。また、独自のデータセットは、専門家や政策立案者が気づかなかったトレンドを見出すのにも役立ちます。 データはジャーナリズムにおける情報源です。ジャーナリズムでデータを活用する基本的な姿勢は、構造的思考です。ジャーナリストはデータを最大限に活用するために、プロジェクトの最初の段階で構造的に考える必要があります。自分が伝えたいストーリーは何か、それを伝えるためには何が必要か。記事のためのデータセットをうまく構築するには、ストーリーを構造的に考え、ジャーナリストとしての好奇心を持ってあらゆる情報源に問い合わせることが重要です。

独自のデータセットを構築することは、構造的思考、計画的なストーリーテリング、創造的な方法でデータを見つけることなど、データジャーナリズムの重要なスキルを多く含んでいます。また、プログラミングのスキルがなくてもできるので、比較的参入障壁が低いのも特徴です。スプレッドシートへの入力と表のソートができれば、データジャーナリズムの基本的なスキルを身につけることができます。データジャーナリズムが簡単という

わけではありません。込み入っていて、徹底したデータプロジェクトは、非常に複雑で時間のかかる作業になりますが、いくつかの重要なスキルを身につければ、データをストーリーテリングに活用する強固な基盤を築くことができます。

独自のデータセットを一歩ずつ構築する

必要なものを準備する。分析に使うデータを作成または収集するのに求められる最初のステップは、何が必要で、それが入手できるかどうかを評価することです。どんなプロジェクトでも、最初にストーリーメモを作成する価値があります。このメモには、ストーリーが何を伝えようとしているのか、どこにデータがあるのか、それを見つけるのにどれくらいの時間がかかるのか、どこに落とし穴があるのかをスケッチします。このメモがあれば、作業にどれくらいの時間がかかるか、その努力に見合う成果が得られるかどうか判断できます。また、このメモは、作業の途中で見返すことができます。

書き出しを考える。データが存在しないデータ駆動型ストーリーの最初の段階では、ストーリーどう始めるかを問うべきです。データに何が含まれているべきかを知ることは、データに何を問うのか決める上で不可欠です。これは、データに含まれているものしか質問に答えることができないからです。したがって、ニーズを満たすデータセットを作成するには、何を調べたいのか、それを調べるためにはどのような情報が必要なのかを明確にする必要があります。

データはどこに保存されているか。次のステップは、データがどのような形式で保存されるかを検討することです。一つの方法は、自分の足取りをたどることです。ここにストーリーが潜んでいることをどうやって知ることができるのでしょう。このアイデアはどこから生まれたのか、その背後には潜在的なデータソースがあるのか。調査は、何が存在するのかを明らかにするのにも役立つので、関心のある問題に言及しているあらゆる情報源を調べ、データを収集したり扱ったりしている学者や研究者、統計学者に話を聞いてみましょう。そうすることで、そのデータを使用する際の欠点や落とし穴を特定することができます。また、データを入手するための他のソースや方法についてもアイデアが浮かぶはずです。データセットの構築に着手する前のこうした準備は、融通のきかない政府機関との共同作業が必要になった場合や、データ収集に別のアプローチを取ることになった場合に、非常に大きな力を発揮します。

倫理的な問題。ストーリーの企画や情報収集には倫理的な問題を考慮する必要がありますが、データを扱う場合も同様です。データセットを構築する際には、情報を収集するために使用している情報源と方法が、可能な限り正確で完全なものであるかどうかを検討する必要があります。これは分析の場合も同様で、複数の角度から情報を検証し、現実を反映していないことを言わせるためにデータをいたぶることはしません。ストーリーを発表する際には、データのソース、分析、限界について透明性を保つようにしてください。これらの配慮は、より強いストーリーを構築し、読者との信頼関係を築くのに役立ちます。

データの取得。潜在的な情報源が特定できたら、次はデータの入手です。これには、スプレッドシートへのデータ入力、PDF に記録されている情報を分析可能な構造化データに変換する、人力による文書の入手や情報公開法（FOIA）による文書の入手、文書や Web ページからデータをスクレーピングするプログラミング、アプリケーション・プログラミング・インターフェース（API）によるデータ取得の自動化などの方法があります。

自分が使いやすく。シンプルさを犠牲にしてはいけません。情報を分析可能なデータセットにするために、最も簡単な方法を探してください。可能であれば、作業プロセスを再現できるようにしましょう。そうすれば、自分の作業を確認したり、必要に応じて後からデータセットを追加したりするのに役立ちます。データを取得する際には、ストーリーのアウトラインを参照し、次のように問いかけます。そのデータで、テーマを十分に掘り下げられるか？私が興味を持っているトップラインにつながる情報が含まれているか？

構造。紙の文書の束に含まれる情報とデータセットの違いは、構造です。分析に適したきれいなデータセットを構築するには、構造と反復が不可欠です。最初のステップは、その情報に慣れることです。その資料には何が含まれているのか、それによって何が言えるのか？そのデータでは何が言えないのか？その情報を組み合わせることができる別のデータセットはあるか？このデータセットを構築する際に、他のものと組み合わせることができるような手順を踏むことができるか？

最終的にデータセットがどのようになるべきかを考える。分析したい列や変数を検討します。他の類似したデータセットの方法論や構造からインスピレーションを得ましょう。まず、収集できるあらゆるデータを考慮して網を広げ、ストーリーに必要なものは何か、それを手に入れるにはどのくらいの時間がかかるのかを見極めて、絞り込みます。集めたデータは、必ず似たもの同士を比較するようにします。フォーマットを決め、それに従うことで、最終的に時間を節約できます。また、作成するデータセットのサイズも考慮してください。時間と日付があれば、時系列で情報を分析することができ、地理的情報があれば、データをプロットして空間的な傾向を調べることができます。

自分の作業を記録し、確認しながら進めること。データセットの作成に使ったソースを記録し、常にオリジナルのドキュメントとデータセットのコピーを保管してください。方法論とデータ辞書を構築し、ソース、データの処理方法、各列の内容を記録します。これにより、データを収集して分析を開始する際に、疑問点を指摘したり、潜在的なエラーを洗い出したりすることができます。

何も仮定せず、発見したことをすべてさらなる報告で確認する。専門家や統計学者に相談して、自分のアプローチや結果をチェックしてもらうのをためらってはいけません。データを収集した後は、自分の仕事を確実にする責任がいっそう大きくなります。データ、分析、記述が正しいことを確認するためにあらゆる手段を講じてください。

事例研究: Beyond the Blade

2017 年初め、データ・プロジェクトチームは、ゲイリー・ヤング、ダミアン・ゲイル、ガーディアンのコミュニティ・ジャーナリズムチームとともに、英国のナイフによる殺人事件で犠牲になったすべての子供と 10 代の若者を記録し始めました。この問題の本質を理解し、ナイフ犯罪にまつわる重要なテーマを探るために、チームにはデータが必要でした。国内で刺されて亡くなっている若者は誰なのか？幼い子どもたちなのか、ティーンエイジャーなのか。性別や民族はどうなのか？どんな若者がいつ、どこで殺されているのか？

統計学者や警察官、犯罪学者に聞いてみると、データは存在するが公開されていないことがわかりました。この疑問に対する答えを見つけようとすることが、翌年の私の仕事の大半を占めました。私が必要としていたデータは、内務省が保有する「Homicide Index」と呼ばれるデータセットでした。この数字は、イングランドとウェールズの警察から内務省に報告されたものです。情報を得るためには、内務省に情報公開請求をするか、各警察に請求をするか、二つのルートが考えられました。あらゆる事態に対応するため、私はその両方を行いました。これで 1977 年以降の数字が得られます。今年の死亡者数を把握するためには、発生した死亡者数を［すぐに］数える必要がありました。公表データや集計データがないため、警察の報告書、ニュースの切り抜き、Google アラート、Facebook、Twitter などを利用して、自分たちで情報を追跡することにしました。

名前、年齢、事件の発生日など、絶対に記録しておきたいことはブレインストーミングで決めました。しかし、死亡の状況を巡る他の側面はそれほど明らかではありませんでした。私たちは、ナイフ犯罪を巡る先入観を議論しました。それは、ナイフ犯罪の被害者はほとんどが男性で、黒人男性の被害者が多いということでした。仮説を確認するために、性別と民族性の欄を追加し、英国内の警察に照会してすべての数字を確認しました。私たちが把握していなかった事例が明らかになったこともあり、報道する前に調査結果をクロスチェックすることができました。情報公開請求が何度も却下され、作業は難航しましたが、最終的に内務省はデータを公開しました。このデータには、約 40 年間にナイフで殺されたすべての人の年齢、民族、性別が、警察の管内別に記載されていました。このデータと現在のデータを組み合わせることで、どのような人が殺されているのか、時系列での傾向を調べることができました。このデータによって、17 年にイングランドとウェールズでナイフ犯罪により 39 人の子供とティーンエイジャーが死亡したことが明らかになりました。若者の死亡率は直近 10 年で最悪でした。これらの数字は、何年にもわたる警察の人員削減の中で見えなくなっていた公衆衛生危機への懸念を引き起こしました。また、ナイフ犯罪がどのような人に影響を与えるかに関して、通説を覆す結果となりました。2015 年までの 10 年間、イングランドとウェールズでは、被害者の 3 分の 1 は黒人でした。しかし、首都圏外では、若者間の刺殺事件のほとんどは黒人に関係がなく、同時期のロンドン以外の犠牲者で黒人は 5 分の 1 以下でした。ナイフ犯罪は大きな話題となっていましたが、政治家や政策立案者は［実態を示す］数字を簡単に知ることができない状態でした。そしてそれは、誰がナイフ犯罪の影響を受けているかの基本的な詳細が入手できないとき、効果的な政策を作ることができるのかという疑問につながったのです。このデータは、ナイフ犯罪に関する議論に新たな展開をもたらし、受賞に至ったプロジェクトの礎となりました。このプロジェクトは、独自のデータセットを構築しなければ実現しなかったでしょう。

数を語り、困難と共にあること

文：ヘレン・ヴェラン

概要

数字は一見、単純明快な価値尺度ですが、数字には道徳的、政治的な問題が隠されていることに気を付けましょう。

キーワード： 生態系価値、オーストラリアの環境ガバナンス、定量化の社会学、データジャーナリズム、科学技術研究

オーストラリアの［ビクトリア］州は今世紀初頭に環境政策を打ち出し、土地所有者が環境悪化の回復を目的とした造園工事を行う場合、その人件費を全額補助することにしました。しかし、当時、新自由主義のイデオロギーがこの政策分野をほぼ完全に支配していたため、この政策は生態系サービスの価値というややこしい二枚舌で説明されたのでした。政策文書では、「国家の環境価値を高めるための環境介入を購入する」と表現されたのです。［こうして］2009 年、州政府のある部局は、この政策の成功について、ほとんど理解不能の主張をすることになります。コランガマイト自然資源管理区域の農村部の土地所有者による生態系の価値を高めるための環境介入を補助する取引を行い、オーストラリア全体の GDP に貢献した金額は、494 万豪ドルと算出されました。ここで紹介した数字は、同年、ビクトリア州政府が発表したプレスリリースからの引用です。このリリースは、オーストラリアにある 57 の自然資源管理地域（Natural Resource Management Regions: NRMR）で、州政府による環境保全への投資が成功したことを発表したのでした。

州南西部に東西に広がる玄武岩の草原からなる環境管理地域は、19 世紀半ばにタスマニアからこの国に入植した初期の英国人牧畜民が付けた地名に代わり、アボリジニの言葉で「Corangamite［コランガマイト］」と名付けられました。英国人たちはこの地域を「Australia Felix［オーストラリア・フェリックス］」と呼び、あらゆる木を伐採しました。地主となった不法占拠者たちは、1 世紀も経たないうちに植民地時代の地主階級のような存在になったのです。2008 年、州政府は、コランガマイト国立森林管理局で［オークション型環境支払い政策の］エコテンダー・プログラムを実施し、不法占拠者の子孫から生態系サービスの価値を金額に換算して入札制オークションで購入しました。これらの取引は 2009 年のオーストラリアの GDP に、494 万豪ドル貢献したと算出されました。私がこの価値を知ったのは、取引額を報じるメディアの発表がきっかけでした。この報告書で公表されたニュースを、その価値を含めて取り上げたジャーナリストはいなかったのではないでしょうか。この数字は決してホットなニュースではありませんでした。

プレスリリースの文脈では、［環境への投資を］具体的な数値に紐付けることは安心感を与えてくれます。GNP は重要かつ現実的なものであり、この州政府の介入が国民経済に貢献する価値として機能するなら、政府の介入は明らかに良いことに見えます。価値の特定は、政府の介入がもたらす改善の現実性を主張

し、この政策が良い環境ガバナンスにつながることをほのめかします。もちろん、数字（494 万豪ドル）が指し示す実際の価値が暗黙のうちに何を指標化しているのかということに、関心を持つ人は多くありません。数字が、価値のある何かに対応しているように見えることは、安心感を与えるには十分なのです。

この数字に関する私の記事は、その数字がどんな社会技術によって命を吹き込まれたのかを、くどいくらいに詳細に説明しています。この記事は、無味乾燥なメディアリリースのありふれた数字が隠蔽と紙一重であることを明らかにします。深遠な問題が潜んでいるのです。ありふれたことであっても、[安易に]評価を行うことに影を落とす深遠な問題について説明を始める前に、データジャーナリズム・ハンドブックの読者が質問しそうなことに先回りして答えてみましょう。第一に、ある数字がどのようにして命を吹き込まれたかを語るのではなく、その数字がある文脈で何を意味するのかを可視化する手段を見つけることは、現代のデータジャーナリズムでは珍しいアプローチです。データジャーナリストがそのようなストーリーテリングがうまくいくのかと疑っているのは想像に難くありません。これに対する答えは、それが二者択一ではなく、解釈と解読において、ストーリーと可視化を絡めて作業することが、アイデアを伝える効果的な方法であることを思い出させることです。ジャーナリストは、ナラティブとビジュアライゼーションの合わせ技には二つの基本的な立場があることを常に覚えておく必要があります。 一つは、ビジュアライゼーションをナラティブに組み込んで話を進める場合です。ビジュアライゼーションはナラティブの中で表現したり、説明したりしているように見えるでしょう。ナラティブをビジュアルに埋め込むこともできます。データジャーナリズムではあまり一般的ではありませんが、これから紹介するストーリーにはこの方法が使えます。もちろん、一つの記事の中で[適切な]方法を切り替えるのが最も効果的な戦略です（ Verran & Winthereik, 2016 をご覧ください）。

第二に、意思決定や政策設計において明らかに重要で、データジャーナリストが関心を持つのは、さまざまなアルゴリズムを利用してデータセットから何ができるかということなのに、特定の数字[だけ]をあげつらうのはおかしいと思うかもしれません。この心配に駆られて、数字とデータセットの関係について尋ねることもあるでしょう。このような質問に対する答えは、かなり単純明快で、面白みには欠けます。データセットの中にはたくさんの数値があります。数字は精密な配列で組み立てられていますが、関係は一対多の関係です。より興味深いのは、数字とアルゴリズムの関係についての質問です。私の答えは、アルゴリズムがデータベースに埋め込まれた関係の処理を精巧に記述したプロトコルである一方、数は集合的な存在の関係をどのように処理するかを規定したプロトコルを表しているということです。ナンバリングはアルゴリズムの一形態であり、その逆もまた然りです 1。

数字とアルゴリズムは、種子とそこから発芽する植物のように、いわば鶏と卵のような関係にあります。従来のアナログ的な手段（認知的、言語的、グラフィック的なリソースを混ぜ合わせたもの）、小学生に教えるような数え上げ、デジタル計算によって数え上げた価値を生成することは、興味深いほどに異なる社会技術的特性がありますが、ここで重要なのは同質性の方なのです。 494 万豪ドル[という金額]はアルゴリズム的に生成さ

れ、特定のデータセットに埋め込まれた、ある関係を表現していますが、単なる数値として提示されているのです 2。

さて、ここで私のストーリーに移りましょう。ここでストーリーとして紹介する数字の生成をめぐる詳細な説明を読んだジャーナリストは、政府がメディアを通じてひそかに誘導しようとする[一見]良いニュースが、一筋縄ではいかないことを認識することでしょう。 これは(むしろ)政治的な暴露記事というべきかもしれません。数字がどのように作られているか詳細に見てみると、この官民連携による環境介入プログラムが、国家が非常に裕福な土地所有者の土地の価値を高めるためにお金を払っているものなのは明らかです。私の記事は、ジャーナリストがどうやって、こうした数字を称賛するか[裏側を]暴露するかを見極められるのかといった疑問を呼ぶかもしれません。ストーリーの終わりには、それが的確な質問ではないとお分かりいただけるでしょう。

数を語ること

買い手である政府と売り手である土地所有者の間で、生態系サービスの価値を取引するために、この PPP(官民連携)プログラムは、どうやって生態系サービスの価値を算定するのでしょう。また、この取引の経済的価値は、オーストラリアの経済活動全体、つまりオーストラリアの国内総生産(GDP)に、どのように貢献するのでしょうか。 私は、この二律背反的な問いかけに対し、土地所有者が「生態系サービスの価値」という商品を生み出し、政府主催のオークションでサービスの提供を競うために必要なことを、順を追って説明することにしました。この製品を生み出す作業とは、土をこねくり回し、木の苗を植え、フェンスを修復することです。これは、土地所有者の先祖が、穀物や羊毛や一族の財産を増やそうと、木を伐採し、水を必要とする穀物を植えることで土地に与えたダメージを修復することが含まれています。[すなわち]生態系サービスの価値は、環境プロセスに介入することで生まれるのです。 取引される商品としての価値は、ビクトリア州政府の省庁(当時は DSE、Department of Sustainability and Environment)職員が作り出します。職員が集まって、行政が入札を「実行」する州内の地域をまとめて決定します。その際、環境システムモデリングプラットフォームである EnSym は非常に重要なツールとなります。この計算能力は驚異的で、これまで科学者が知らなかった「外の自然」を知ることができます。正確で綿密な表現が、一夜にして可能になるのです。

このソフトウェアは ecoMarkets チームによって開発され、DSE で開発された科学、標準、指標、情報、そして多くの国際的、国内的な科学モデルを取り入れています。EnSym には、現地調査用の「Site Assessment Tool」、資産の優先順位付けとメトリック構築用の「Landscape Preference Tool」、流域計画に用いる「BioSim」の三つの主要なツールが含まれています (DSE, 2018)。オークションが行われる地域の優先順位付けとマッピングを行い、環境利益を特定して定量化し、保全・緑化作業によって向上する可能性のある生態系の価値を数値で記録しています。これは、「そこにある」土地の生態系の特性を表しているとされます。コンピュータプログラムはそれだけではなく、人間が介入するためのスクリプトを作成することもできます。劇の台本は制作なくして生まれませんが、このスクリプトも然りです。そして、その台本が出来上がると、[人間界の]外側の自然がより身近に感じられます。それは完全に[人間から]隔絶した「外部の自然」では

なく、人間の生活インフラとしての自然であり、「配管」を修理するようにあちこちをつつく[ことができる]インフラになるのです。

人の力を結集したスクリプトができあがったら、政府はプロジェクト地域の土地所有者に興味を持つよう呼びかけます。そして、応募があったすべての物件を訪問します。担当者は EnSym で生成された大まかなスクリプトを持って、実際の場所を訪れることになっています。そこには、手ごわい「翻訳」が待ち受けています。現地の担当者は、スクリプトの舞台となる候補地を評価します。生態系サービスの(価値の)創出を後押しすることが目的なので、ある場所における特定の行動が、生態系サービスの増加をもたらすこと、すなわち特定の生態系サービスの価値を高める可能性を評価する必要があるのです。そうした増加を足し合わせて、州全体での価値の創出を計算するのです。

土地所有者と行政が一緒に計画を練ります。継続的な交渉の中で、対象区画の正式な管理計画が考案されることになります。現場担当者は、契約可能な条件でこの計画を作り、土地所有者は、計画を実行するために行う作業を詳細に記述します。このようにして、特定の「生態系サービスの価値」という形をとった製品は、一定期間内に完了する一連の作業として設計されます。例えば「このパドックのこの角に、この種の苗をこれだけの数植え、このサイズの保全区画を、この材料を使ったフェンスで囲む」といった感じです。土地所有者は、国が指定した工事の費用を計算しますが、その中には間違いなく手厚い労務費も含まれています。そして、この製品を購入するために政府が支払わなければならない価格、つまり「生態系サービス価値」が出来上がるのです。[土地所有者は]指定された工事を請け負い、指定された期日までに生態系サービス価値を提供する対価を明示します。そうして関連書類を封筒に入れ、政府に提出します。

では、その後のオークションはどのように行われるのでしょうか？EnSym は入札を評価する上で再び重要な役割を果たします。EnSym は「外部の自然」を知り、それに介入するスクリプトを書くだけでなく、ワーナー・ブラザーズが映画製作の入札を評価するように、スクリプトをつくって行われた入札を評価できる審判者でもあるのです。入札は、計算された「環境便益指数」と土地所有者が提示した価格によってランク付けされます。私たちは、政府が単価あたりの「環境便益指数」が最も高い製品を購入すると仮定しなければなりません。

入札の評価：すべての入札は、環境成果の変化の推定値、環境成果の変化の価値、これらの変化によって影響を受ける資産の価値(重要性)、ドルコスト(土地所有者が決定する価格)に基づき、客観的に評価されます。(DSE、2018 年)

オークション結果が発表されると、選定された入札者は、管理計画と提出された工事スケジュールに基づき、空間的・時間的構成が定義された最終的な契約に署名します。すべての書類に署名されると、報告書が作成され、支払いが開始されます。DSE は請求書を受領した後、署名した土地所有者に支払いを実施します。支払いは、管理契約書に明記されている行動に対して十分な進捗があった場合に行われます(DSE, 2018)。

これっていいことだよね？

私は、生態系サービスの価値を売買する方法を正確に説明した記事をレイアウトすることにしました。ここで、プレスリリースに話を戻しましょう。ざっと読むと、この政府のプログラムが生み出した自然資本の付加価値が、494 万豪ドルであるかのような印象を読者に与えるかもしれません。一見すると、494 万豪ドルは、このプログラムで達成されたオーストラリアの自然資本価値の増加分に見えます。しかし、それは間違いです。494 万豪ドルは自然資本価値を表しているのではありません。この数字の名前が何を指しているのかは、以下で説明します。このオークションで売買された商品にじっくり向き合ってみましょう。この商品こそ、私が「ともにあるべき」問題なのです。

この精巧でかなりコストのかかる政府のプログラムが達成した「生態系サービスの価値」の増加は、いかにして実現したのでしょうか？価値の増大を実現するための作業を注意深く見てみると、そのプロセスのどこにも、そしていかなる時にも、その価値が定義されていないことがわかります。厳密には、売買されている商品は存在しないのです。さらに悪いことには、そうでなければならない［価値の増大を計算する方法は］ありえないのです。このプログラムは、お金を配るための、非常に手の込んだ手段です。これが分かってしまえば、このプログラムが何なのかという実態も明らかになります。結局のところ、このプログラムは、国の財源から土地所有者の手にお金を移すための合法的手段なのです。これが自由な議会制民主主義における環境ガバナンスのプログラムであり、政党という社会技術が重要であることを認識した上で、ナレーターの私が一時的に政党政治の帽子をかぶってみましょう。コランガマイトは、ビクトリア［州］議会において、中道左派（労働党）の議員を選ぶか、中道右派（自由党）かの間で揺れ動いてきた歴史がある選挙区です。有権者にアピールすることは、左寄りの政党であれ右寄りの政党であれ、どの政党の利益にもなることは明らかです。そのためには、国家から有権者の銀行口座に合法的に資金を移動させる方法を見つけるのが一番です。国が購入し、土地所有者が販売する製品の価値を数値で表せないということは、この際どうでもいいことなのです。

まとめましょう。経済的には、このプログラムは環境サービスの価値を生み出すものとして正当化されています。この観点からすると、これは良いニュースです。納税者のお金を使って環境を改善し、ビクトリア州の過剰な二酸化炭素の発生を改善するために植樹しているのですから。問題は、ビクトリア州の自然資本の価値の増加は、それが売買される製品であるにもかかわらず、数字として明確化し、説明することができないことです。まだ技術的な問題はあると言っても、［自然資本の価値が増えるであったならば］明らかに良いことでしょう。しかし別の経済学的視点からすると、このプログラムは、私有地所有者の財産価値を高めるために植樹の労働に資金を提供しているとも言えます。これは、地主の祖父母が利益を得ていた可能性が高い開墾という誤った労働を助成し、地主が引き続き享受する利益を生み出すという、過去の政策によって引き起こされた損害を増幅する介入手段なのです。この解釈では、州政府が（グリーン・オークションの）エコテンダーで実施した政策は、納税者のお金を合法的に配るための高く付くプログラムです。明らかに、これは悪いことです。

数字とアルゴリズムを軽視しない：困難と共にあること

では、ジャーナリストは何をすればいいのか。私はジャーナリストとしてではなく、研究者としての立場で書いているので、この疑問に漠然としか答えられません。冒頭で[述べた]、プレスリリースに使われる数字は、潜んでいる問題から注意をそらす隠蔽工作と紙一重であるという主張に戻りますが、私の考えでは、評価は常にモラルの問題を引き起こします。私が考える正しい質問は、「データジャーナリストはそのモラルの問題にどのように対応するのか？」というものです。まず、494 万豪ドルの件から片付けましょう。この数字は何ですか？この整然とした金銭的価値はどこから来ているのでしょうか？これは、エコテンダー・プログラムに批判的な学術論文に書かれているものです。

この市場ベースのモデルの下では、生態系サービスの経済的価値は、保全契約を遵守するための単位当たりのコストが、オークションの落札者に与えられる単位当たりの価格よりも低い場合に生み出される。これらの売り手が市場化された商品の生産を見送る可能性はあり、その場合、経済的価値が失われるが、合理的な土地所有者の参加制約により、オークションの実施によって生み出される「経済」価値の純増は確実である（G. Stoneham et al., 2012）

この政策の経済モデルによれば、土地所有者は、インフラとしての自然に介入する政府のスクリプトを作成するため－自然インフラの働きをより効率的に発揮させるために－負担するコストを効率的に計算することが前提となっています。もちろん、土地所有者が利益を得る代わり、計算ミスをして損失を出すこともあり得ますが、土地所有者の労働で生み出される価値の買い手である政府にとっては、どうでもいいのです。 政府が関心を持っているのは、このような経済取引をいかにうまく表現できるかという点です。売買される商品がオークションの中だけに存在している場合には、かなりの問題となります。この問題のある商品の存在形態を解決するのが、GDP の精巧で複雑な技術なのです。政府は自然界に変化をもたらしていることをアピールしようとしています。そして、GDP は、とても便利なことに、これを貨幣的に示すことができます。「環境便益指数」とは、政府が環境サービスという商品を購入する根拠となった価値ですが、それは便宜的な存在です。このような存在形態の難しさにもかかわらず、巧妙な工夫を凝らすことで、一回限りの刹那的な存在である代物を売買する手段が実現し、具体的な経済活動として GDP に組み込むことができるようになるのです。しかし、一部の経済学者はその正確性に重大な懸念を示しています（G. Stoneham et al.）。

[このように]494 万豪ドルという金額は、エコテンダー・プログラムの活動とも、それが改善するはずの自然ともかけ離れたものです。政府が、自然の改善に成功したと発表するとき、改善の程度を示すことができれば[分かりやすくて]いいでしょう。どんな数字でもないよりましに見えます。確かにこれ[494 万豪ドル]はハッピーでポジティブな数字です。政府の会計担当者が間違いなく把握しているはずの不幸でネガティブな数字、つまり政府のプログラムを運営するコストの価値は、ここでは決して明らかにならないでしょう。なぜこんな妙な名前の数字にこだわるのか？少しやり過ぎではないか？比較的簡単に見破ることが出来る小細工の何が悪いのか？

私は、これが意図的な数字の誤用ではないかと心配しています。数字に真摯に向き合わず、ナンバリング、この場合はアルゴリズムが常に引き起こしうる問題を決して認めようとしないのは、プロトコルを台無しにする行為です。政府の Web サイトを訪れて見つけた数字についての記事は、社会財と害を同時に生成する政府のプログラムの実態を明らかにしました。記事で取り上げた（494 万豪ドルというメディアリリース中に現れた）手品のような数字の創出は、モラル的な緊張と困難の場である評価（が必要であること）を指し示しています。ここで、価値とは永遠に論争が続く道徳的な問題なのだろうかと強く問いたいと思います。価値論はあらゆる哲学に古代から根差した深遠なテーマであり、私はこの広大なウサギの巣穴（のような迷宮）には深入りしないでおきます。ここではただ、過去 30 年間によく耳にした、市場の見えざる手が価値に関する道徳的問題を抑制するという主張は、危険な誇張であると指摘しておきます。私の記事で示したように、市場は一時的には価値（の問題を）手なずけるかもしれません。しかし、いずれは必ず価値観の問題に立ち戻ることになるのです。それに注意を払うことこそジャーナリストの使命です。ここでは、データジャーナリストが数字とアルゴリズムをプロトコルとして尊重する方法について、いくつかの提案をします。表面的に問題がないように見えても、何かが潜んでいるように思えたら耳をそばだててください。数字やアルゴリズムに慣れ、数字が覆い隠す（壁の）のぞき穴を広げるよう試みましょう。数字やアルゴリズムに敬意を払い、好奇心を規律ある方法で実践する姿勢を身につけましょう。数字には固有の性質と、[数字同士が]出合った現れる特性があることを知っておいてください。それを把握することが、[数字を扱う]実践の本質なのです。これらを十分にこなせるようになれば、驚くべきことが待ち構えているはずです。数字の中から面白いことが出てきますよ。

脚注

1. 数字とアルゴリズムが同じものであるという考え方は、多くの読者にとって新鮮だろう。数字を「抽象的」なものとして考えることに慣れてしまっている。私の（一風変わった）数字の説明では、数字は実在するごく普通の物質的な記号論的実体に位置づける。集合的な存在の一つの瞬間の中で、異なるプロトコルが関係を動員するという説明については、Watson, H. (1990)を参照。Investigating the social foundations of mathematics:Natural number in culturally diverse forms of life.*Social Studies of Science*, *20*(2), 283–312. doi.org/10.1177/030631290020002004, または Verran, H. (2001). ナンバリングの二つの論理については In *Science and an African logic* (pp. 177–205). University of Chicago Press を参照。
2. アナログまたはデジタル環境において様々に出現する 3 つの数の異なる社会技術的特性についての説明は、Verran, H. (2015)を参照のこと。Enumerated entities in public policy and governance. In E. Davis & P. J. Davis (Eds.),*Mathematics, substance and surmise: Views on the meaning and ontology of mathematics* (pp. 365–379). Springer International Publishing. doi.org/10.1007/978-3-319-21473-3_18

参考文献

DSE. (2018). Innovative market approaches: EcoMarkets. EcoTender and BushT- ender. Department of Sustainability and Environment [now Department of Environment, Land, Water and Planning (DELWP)], Victoria State
Government, Australia. www.environment.vic.gov.au/innovative-market-approaches

Roffe, J. (2015). *Abstract market theory*. Palgrave Macmillan UK.

Stoneham, G., O'Keefe, A., Eigenraam, M., & Bain, D. (2012). Creating physical environmental asset accounts from markets for ecosystem conservation. *Ecological Economics*, *82*, 114–122. doi.org/10.1016/j.ecolecon.2012.06.017

Verran, H., & Winthereik, B. R. (2016). Innovation with words and visuals. A baroque sensibility. In J. Law & E.

Ruppert (Eds.), *Modes of knowing* (pp. 197–223). Mattering Press.

先住民のデータ主権:データジャーナリズムへの影響

文:タフ・ククタイ、マギー・ウォルター

概要

デジタル化がもたらす潜在的な弊害について考察し、科学と行動主義の新たな場としての「先住民のデータ主権(ID-SOV)」が、リスクを調整しながらどのようにして利益を提供できるか考察します。

キーワード:先住民のデータ主権、行動主義、データジャーナリズム、統計的監視、先住民族

モニタリング技術や情報技術、人工知能(AI)などのデジタル技術は、先進国や移行経済圏の先住民の生活の特徴となりつつあります。データ駆動型のテクノロジーはイノベーションを促進し、人間の福利を向上させることができますが、先住民は社会経済的、文化的、政治的に疎外されているというほぼ共通の立場にあるため、これらの利益を公平に享受することはできません。また、政府や企業がリンクして統合されたビッグデータの利用を拡大することは、先住民に大きなリスクをもたらします。例えば、文化的知識や知的財産の流用、土地やその他の天然資源の搾取、差別や偏見の永続化、継続的な社会からの疎外などです。これらのリスクは、先住民の機能不全に関するお決まりの表現を踏襲したジャーナリスティックなストーリーテリングによって増幅されます。

この章では、デジタル化がもたらす潜在的な弊害について考察し、科学と行動主義の新たな場としての「先住民のデータ主権(ID-SOV)」が、リスクを調整しながら利益をもたらす方法を考えます。最後に、ID-SOV の研究とネットワークは、データとデータの専門知識の貴重な情報源であり、先住民や問題に関わるジャーナリズムに対して、より公平で批判的、かつ公正なアプローチを提供することができることを示唆して結びます。

先住民とデータ

世界には 3 億 7000 万人の先住民がいると推定されており、すべての大陸を網羅し、何千もの異なる言語を話しています(United Nations, 2009)。先住民がいる国は、国内のデータ収集を行う際、先住民を特定しないため、実際に世界にどれくらいいるのかを知ることはできません(Mullane-Ronaki, 2017)。このような先住民の「データ砂漠」や先住民の政治的自治、生活水準が世界的に大きく異なるにもかかわらず、先住民はしばしばその国で最も貧しい人口集団のひとつであり、病気や過剰な投獄、広範な不平等といった最も重い負担を背負っているという十分な証拠があります(Anderson et al., 2016; Stephens et al., 2006)。このように周縁化された位置にいるのは偶然ではなく、植民地化され、奪われた人々としての歴史に直接関係しています。しかし、ジャーナリズムの主流は先住民やコミュニティを描写するとき、植民地主義とその仲間である白人至上主義や人種差別の壊滅的な結果を認めることはほとんどなく、批判することもありません。

先住民は、芸術、彫刻、トーテムポール、歌、聖歌、踊り、祈りなどを通じて、情報や知識を記録し、保護するという古い伝統を持ち、現在データとして知られているものに常に積極的に取り組んできました。このような慣習や知識体系を意図的に排除することは、植民地化の過程の一部でもありました。同時に、ヨーロッパの旅行者、探検家、科学者は、先住民とその文化をより客観的に、科学的に、そして信頼できる「知識人」として[の立場で]紹介し、[その方式によって]先住民は理解されるようになりました。やがて、植民地主義を正当化し維持してきた人種的階層は、イデオロギー的構造、制度的取り決め(例えば奴隷制度や人種差別)、国家による分類慣行を通じて自然化し、組み込まれていったのです。

例えば、オーストラリアのアボリジニとトレス海峡諸島民は、1971 年まで国勢調査から除外されており、老齢年金などの基本的な市民権から排除されていました(Chesterman & Galligan, 1997)。

現代では、先住民をどのように数え、分類し、分析し、対処するかを決定する権限は、先住民自身ではなく政府にあり続けています。先住民が自らに関わるデータに対する権力を国家から取り戻すことが ID-SOV の核心です。

ID-SOV の定義

ID-SOV は比較的最近の概念で、このテーマに関する最初の出版物は 2015 年に出たばかりです(Kukutai & Taylor, 2016)。それは、先住民が自分たちに由来し、自分たちの構成員、知識体系、慣習、または領土に関連するデータを所有、管理、アクセス、保有する権利に関するものです(First Nations Information Governance Centre, 2016; Snipp, 2016)1。「先住民族の権利に関する国連宣言(UNDRIP)」に記載されているように、ID-SOV は先住民の民族、国(土地、水、空を含む)、資源に対する自決と統治の固有の権利によって支えられています2。これに込められているのは、先住民の集団的幸福と自決を支え、強化する方法でデータが利用されることを望むという、先住民の NGO、コミュニティ、部族が強調している感情です(First Nations Information Governance Centre, 2016; Hudson et al.)。

実際、ID-SOV は、先住民に関するデータがどのように使用または展開されるかについて、先住民が意思決定者となる必要があることを示しています。そのため、ID-SOV は次のような疑問を喚起します。誰がデータを所有しているのか？誰が、どのような状況で、どのようにデータにアクセスするかを決定する権限を持っているのか？データとそのアプリケーションの想定受益者は誰か？

ID-SOV は、(プライバシー権を含む)個人の権利、リスク、利益と、その個人が属する集団の権利とのバランスをどのようにとるか、というやっかいな問題にも関わります。集団の権利と利益に焦点を当てることは、EU の一般データ保護規則(GDPR)のような政策や規制のアプローチに浸透している、個人データの保護と管理という狭い論点を超越するもので、重要です。個人のプライバシーや所有権といった欧米の法的概念は、個人がより広いグループの一員であり、例えば共通の系譜や遺伝子によって定義されているような先住民族の文

脈では、あまり意味を持ちません。このような状況では、他のグループのメンバーに関する情報をコード化したデータの共有は、個人の同意だけでは成り立たず、集団的な権利や利益も考慮しなければなりません（Hudson et al.、2020）。

ID-SOV と密接に関連しているのが、先住民データガバナンスの概念です。これは広義には、先住民が自らのデータの管理を行うための原則、構造、説明責任のメカニズム、法的手段、政策と定義されます（Te Mana Raraunga, 2018a）。住民データガバナンスは、その本質において ID-SOV を運用する方法です（Carroll et al.、2017）。先住民データガバナンスを通じて、データに関連する先住民の権利と利益を主張できるのです（Walter, 2018）.

統計調査と先住民

データジャーナリズムは、先住民データの 5D を再生産するのではなく、それに挑む立場にあります。データジャーナリストは、先住民やストーリーを表現するためにデータをどのように使用するかを再考し、先住民のデータが政府や産業界によって生産、管理、普及、「利用」される複雑な方法を明らかにする十分な機会があるのです。その際、データジャーナリストは非先住民のデータ生産者や利用者に頼るべきではありません。ID-SOV ネットワークの台頭は、頼れる先住民のデータ専門家が増えていることを意味します。ID-SOV の活動に携わる人々の多くは、自分たちのコミュニティと密接な関係を持ち、データの公正さへの強いこだわりと、「良いデータ」が「良い結果」をもたらす方法を見つけようとする姿勢を持っています。ID-SOV によって提起された問題、特にデータの所有権、管理、被害、集団的利益に関する問題は、先住民のコミュニティを超えて広く適用されます。データジャーナリストは ID-SOV のアプローチと原則に従うことで、先住民の視点や関心を物語の枠組みに組み入れる道を拓くことができ、同時に権力者の責任を追及するレンズを磨くことができるのです。

脚注

1. アオテアロア NZ では、ID-SOV ネットワーク Te Mana Raraunga によるマオリ・データの定義を踏襲し「マオリの人々、言語、文化、資源、環境に関する、デジタルまたはデジタル化可能な情報」とする（Te Mana Raraunga, 2018a）。

2. www.un.org/esa/socdev/unp i/documents/DRIPS_en.pdf

3. www.temanararaunga.maori.nz/tutohinga

参考文献

Anderson, I., et al. (2016). Indigenous and tribal peoples' health (The Lancet–Lowitja Institute Global Collaboration): A population study. *The Lancet*, *388*(10040), 131–157. doi.org/10.1016/S0140-6736(16)00345-7

Berda, Y. (2013). Managing dangerous populations: Colonial legacies of security and surveillance. *Sociological Forum*, *28*(3), 627–630.doi.org/10.1111/socf.12042

Briscoe, G. (2003). *Counting, health and identity: A history of aboriginal health and demography in Western Australia and Queensland 1900–1940*. Aboriginal Studies Press.

Carroll, R. S., Rodriguez-Lonebear, D., & Martinez, A. (2017). *Policy brief (Ver- sion 2): Data governance for native nation rebuilding*. Native Nations Institute. usindigenousdata.arizona.edu

Chesterman, J., & Galligan, B. (1997). *Citizens without rights: Aborigines and Australian citizenship*. Cambridge University Press.

First Nations Information Governance Centre. (2016). Pathways to First Nations' data and information sovereignty.

In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 139–155). Australian National University Press. doi.org/10.22459/CAEPR38.11.2016.08

Henman, P. (2018). Of algorithms, apps and advice: Digital social policy and service delivery. *Journal of Asian Public Policy*, *12*(2), 1–19. doi.org/10.1080/17516234.2018.1495885

Hudson, M., et al. (2020). Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. *Nature Reviews Genetics*, *21*(6), 377–384. doi.org/10.1038/s41576-020-0228-x

Hudson, M., Farrar, D., & McLean, L. (2016). Tribal data sovereignty: Whakatōhea rights and interests. In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 157–178). Australian National University Press. doi.org/10.22459/CAEPR38.11.2016.09

Kukutai, T., & Taylor, J. (Eds.). (2016). *Indigenous data sovereignty: Toward an agenda*. Australian National University Press.

Maiam nayri Wingara Indigenous Data Sovereignty Collective & Australian Indigenous Governance Institute. (2018).

Indigenous data sovereignty summit communique. web.archive.org/web/20190305225218/http://www.aigi.com.%20au/wp-content/uploads/2018/07/CommuniqueIndigenous-Data-Sovereignty-%20Summit.pdf

Mullane-Ronaki, M.-T. T. K. K. (2017). *Indigenising the national census? A global study of the enumeration of indigenous peoples, 1985-2014* [Thesis]. University of Waikato.

Snipp, M. (2016). What does data sovereignty imply: What does it look like? In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda*(pp. 39–56). Australian National University Press.

Stephens, C., Porter, J., Nettleton, C., & Willis, R. (2006). Disappearing, displaced, and undervalued: A call to action for Indigenous health worldwide. *The Lancet*, *367*(9527), 2019–2028. doi.org/10.1016/S0140-6736(06)68892-2

Stoneham, M. (2014, April 1). Bad news: Negative Indigenous health cover- age reinforces stigma. *The Conversation*.theconversation.com/bad-news-negative-indigenous-health-coverage-reinforces-stigma-24851

Te Mana Raraunga. (2017). Statement on social licence. www.temanararaunga.maori.nz/panui/

Te Mana Raraunga. (2018a). Principles of Māori data sovereignty.www.temanararaunga.maori.nz/new-page-2

Te Mana Raraunga. (2018b). Te Mana Raraunga statement on 2018 New Zealand Census of Population and Dwellings: A call for action on Māori census data. www.temanararaunga.maori.nz/panui/

United Nations. (2009). State of the world's indigenous peoples. www.un.org/esa/socdev/unpf ii/documents/SOWIP/en/SOWIP_web.pdf

Van Alsenoy, B., Ballet, J., Kuczerawy, A., & Dumortier, J. (2009). Social networks and web 2.0: Are users also bound by data protection regulations? *Identity in the Information Society*, *2*(1), 65–79. doi.org/10.1007/s12394-009-0017-3

Walter, M. (2016). Data politics and Indigenous representation in Australian statistics. In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 79–98). Australian National University Press.

Walter, M. (2018). The voice of indigenous data: Beyond the markers of disadvantage. *Grif th Review*, *60*, 256.

Walter, M., & Andersen, C. (2013). *Indigenous statistics: A quantitative research methodology*. Left Coast Press.

中国における新方式のデータ・プラクティス

文：馬金馨

概要

中国におけるデータジャーナリズムの状況、キープレイヤー、データ文化、そして実務的なヒントについて、その内部事情を紹介します。

キーワード：中国、データ文化、市民参加、オープンデータ、データジャーナリズム、データ・ビジュアライゼーション

数年前、私は Google News Lab が主催する Google News Summit で中国のデータジャーナリズムを紹介するプレゼンテーションを行いました。美しい冬の日で、シリコンバレーの中心部にある会場は、欧米諸国のシニアメディア専門家ら約 100 人で満席でした。私はまず、中国には良いデータが存在しないと思うかどうか、第二に、中国には本当のジャーナリズムがないと思うかどうか、手を挙げてもらうことから始めました。どちらの質問にも結構な数の手が挙がり、笑いを誘いました。これは、国際的なジャーナリズム会議に出席したり、講演したりする際に、偏見ではないにしても、よく出くわす二つのコメントです。

この 6 年間、私が観察したところでは、中国にはデータがないどころか、膨大な量のデータが日々蓄積されており、その質は向上しています。真のジャーナリズムがないとはいえず、毎日多くのジャーナリストが印象的な記事を書いています。しかしながら、そのすべてが日の目を見るわけではありません。

課題解決型のデータ作成

中国に「データジャーナリズム」という言葉が導入される前から、データ記事は存在していました。現在、中国では「データ駆動型ストーリー」という言葉が普通に使われていますが、データがストーリーを導くのではなく、ストーリーや特定の問題がデータ生成の原動力となっていた時期がありました。それは、大気汚染のような一般市民の心に響く問題です。生態環境部は 2010 年からリアルタイムの大気汚染指数を発表していますが、重要な数値が一つ欠けていました1。PM2.5、つまり人体に不可逆的な害をもたらす直径 2.5 マイクロメートル以下の汚染物質は公表されていなかったのです。大気汚染の深刻さと PM2.5 の公式データの不足を受けて、2011 年 11 月、「我为祖国测空气」という全国的なキャンペーンが開始されました。

このキャンペーンは、環境保護の NPO が始めたもので、市民から資金を集めて検査機器を調達し、興味のあるボランティアにはトレーニングを提供しました 2。その中には、中国で最も広く利用されているソーシャルメディアの一つである新浪微博で 700 万人以上のフォロワーを持つ有名なビジネスリーダー、潘石屹氏も含まれています（Page, 2011）。2012 年 1 月から2年間のキャンペーンの後、PM2.5 のデータは最終的

に政府のデータリリースに盛り込まれました。幸先の良いスタートでしたが、課題も残りました。現地の米大使館が発表したデータとの食い違いがすぐ明らかになり、データの正確性と説明責任に疑問が生じたのです（Spegele, 2012）。

また、機能面でもジャーナリストには使いにくいものでした。100 以上の都市のデータが 1 時間ごとに更新されるにもかかわらず、その情報は Web ページ上で随時更新されるだけで、データセットとしてダウンロードすることはできません。さらに、データは一元化されたものの、過去のデータは一般には公開されていませんでした。1 時間ごとにデータをスクレーピングしてローカルに保存するスクリプトを書くことができなければ、時系列での傾向分析や都市間の比較を行うことができない状態だったのです。しかし、話はこれで終わりではありません。課題解決型のデータ生成は継続しています。データがうまく構造化されていなかったり、ユーザーフレンドリーなフォーマットではなかったり、データジャーナリストが限られた技術力で苦労したりしているときは、市民やテックギークが手助けしてくれることがあります。初期、すなわち 2011 年の例として、大気汚染データをスクレーピングしてクリーンな形式で公開している PM25.in を紹介します。このサイトでは、運営開始以来、10 億回以上の検索クエリを記録しているそうです 3。もう一つの例である非政府組織［所属］の Qing Yue［不詳、上海大の研究者か］は、政府の Web サイトから環境データを収集、クリーニングし、ユーザーが使いやすい形式で一般に公開しています。処理されたデータは、メディアのデータチームだけでなく、政府機関もより良い政策立案のために広く利用しています。

データの生成と、特定の問題に対する意識の高まりは密接に関係しています。15 年、深刻な大気汚染を調査したドキュメンタリーが全国を席巻しました。自主制作映画「穹頂之下」は、全国の有害スモッグの環境危機を暴露し、問題の根源と責任を負う様々な当事者を追求しました（Jing, 2015）。この映画は、スタイルとインパクトの両面で、アル・ゴアの『不都合な真実』と比肩するものです。ストーリーテリングの面では、多くの科学的データ、分析に基づき、長年の傾向を説明するためのチャート、環境とエネルギー産業内の腐敗をソーシャルネットワークで可視化しました。オンラインで公開されるとすぐに、この映画は口コミで拡散され、3 日間で 2 億ヒットを記録しましたが、1 週間後には検閲されて削除されました。

しかし、この映画は、大気汚染データへのアクセスと品質など、この問題に対する国民の意識を高め、国民的な議論に火をつけることに成功したのです。国の指導者たちにこの問題の重要性を認識させることにも成功しました。ドキュメンタリー公開から 2 週間後、全国人民代表大会の記者会見で、この映画に言及した大気汚染に関する質問に答えた李克強首相は、政府が公害抑制を求める国民の要求に応えていないことを認め、汚染規制の緩さなど、ドキュメンタリーで提起された問題のいくつかを認め、有毒スモッグを抑制するために、政府はより重い罰則を課すことを強調しました（Wong & Buckley, 2015）。15 年 8 月末には改正大気汚染防止法が発表され、16 年 1 月に施行されました Lijian et al.、2015）。

大気汚染は、データの入手やアクセスが困難な場合でも、問題に対する市民の関心がデータ生成への貢献につながり、政府の態度や公共部門のデータの入手のしやすさを変えることができることを示す一例に

すぎません。より確立されたシステムにおいては、データはより容易に入手可能で使いやすくなり、ジャーナリストの仕事は、取得したデータを基に記事を書くという、より単純なものになります。中国では、そのプロセスは直線的ではなく、市民、政府、民間組織、メディアの力学は相互作用しています。データは、単にストーリーの出発点としての役割を果たすだけでなく、後の段階で登場し、ジャーナリストと一般市民との間の新しい種類の関係を可能にしているのです。

進化するデータ文化

中国のデータ環境はこの 10 年間で急速に変化していますが、その背景には上記のような力学もあれば、世界的なオープンデータ運動、インターネット企業の急成長、モバイル普及率の驚くべき高さなど他の要因もあります。データ文化もまた、こうした流れの中で進化してきました。政府の法律は、データの入手可能性に関する政策を支えています。多くの人が驚きますが、中国には情報の自由に関する法律があります。

国務院により「政府情報公開条例」が 2007 年に採択され、08 年 5 月 1 日に施行されました。この条例にのっとり、各レベルの政府機関は、データを含む保有する情報を開示するための専用の Web ページを開設しました。しかし、財新が 11 年に発表した調査によると、データや情報を要求する権利を与えられたにもかかわらず、法が施行されてから最初の3年間に、メディアやジャーナリストがデータ開示を要求した事例は確認されていません4。10 年に大手新聞社南方週末が 29 の環境局に法律の順守度合いに関する情報開示を求めたところ、回答率はわずか 44%だったことが明らかになりました、メディア組織内には通常、ジャーナリストの要求を実現するための法的チームのような支援システムはありません。あるジャーナリストは、情報を開示しなかったことで政府を提訴し、結局は職を失うことになったのです。中国のジャーナリストが法的手段を利用する際に直面する困難とリスクは、欧米のジャーナリストよりもはるかに大きいと言えるでしょう。

世界的なオープンデータの取り組みとビッグデータへの関心の高まりを受けて、中国も対応しています。12 年には、上海と北京の両市が、土地利用、交通、教育、公害監視などの分野で、それぞれ数百のデータセットを持つ独自のオープンデータポータルを立ち上げました。その後数年の間に、大都市だけでなく、地方や発展途上の省にも 10 以上のオープンデータポータルが開設されました。開発はどちらかというとボトムアップ型で、地方レベルでデータを公開するための標準的なテンプレートや構造がないため、幅広い比較可能性や有用性に貢献するものではありませんでした。15 年には国務院が「ビッグデータ開発行動計画」を発表し、オープンデータが国家 10 大重点プロジェクトの一つとして正式に認められ、政府データのオープン化に向けたスケジュールが提示されました5。しかし、公的データはジャーナリストの出発点になるとは限らず、公共の利益や関心事と必ずしも一致しているとは限りません。

一方、民間企業、特にアリババやテンセントなどのテクノロジー大手は、長年にわたって膨大なデータを蓄積しています。最新の公式発表によれば、アリババの年間アクティブ消費者数は 2018 年 9 月 30 日までに 6 億 100 万人に達しました（“Alibaba Group Announces,” 2018）。東南アジアの全人口に相当する、

このような強力なユーザー層から得られる e コマースデータからは、取引傾向、人口動態の変化、都市部における移動、消費者の習慣の変化など多くのことを知ることができます。また、中国には Yelp に相当する大衆点評のように、より詳細なデータを利用できるレビューサイトもあります。プライバシーやセキュリティに関する懸念はあるものの、適切に使用されれば、これらのプラットフォームはデータジャーナリストに豊富なリソースを提供してくれます。

ビッグデータ活用の優れた例として、上海メディアグループ傘下で都市生活に関するデータストーリーを専門とするチーム「Rising Lab」を挙げたいと思います6。ラボの設立は、都市化という新たなトレンドに対応したものでした。2016 年の政府報告書によると、中国には 1978 年時点で 193 都市があったのに対し、現在は 600 以上の都市があり、人口の 56％が都市部に住んでいます（“Gov’t Report: China’s Urbanization,” 2016）。ラボは、テック企業とのパートナーシップ契約や技術支援のもと、都市生活者がよく利用するウェブサイトやアプリから、不動産価格、コーヒーショップやバーの数、コワーキングスペースの数、公共交通機関の利用のしやすさなど、都市生活のさまざまな側面を反映したデータを収集しました。独自の手法を駆使して、商業的な魅力、イノベーションのレベル、生活の多様性など、さまざまな側面から都市のランキングを作成しています（Figure 10.1）。ランキングとストーリーは、入手可能な新しいデータに基づいて毎年更新されますが、一貫性を保つために同じ方法論に基づいています。ラボのディレクターである沈从乐氏によると、このコンセプトとストーリーは好評を博しており、都市計画政策や企業のビジネス上の意思決定に影響を与え始めているそうです（Shen, 2018）。

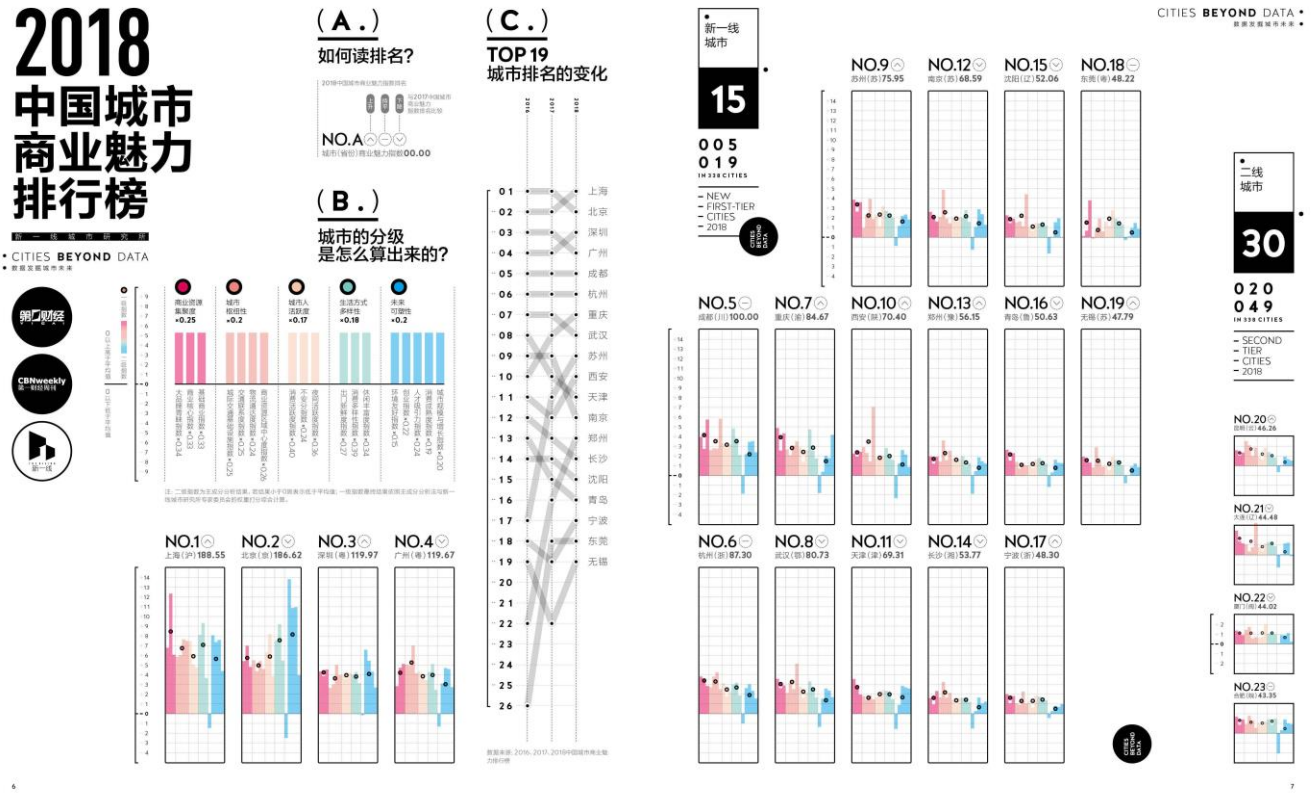


Figure 10.1. Cities Beyond Data 2018：中国における都市のビジネス魅力度ランキング　出典：The Rising Lab.

ラボの成功は、データ提供者、ジャーナリスト、市民の間に新たな動きが生まれていることを示しています。このラボは、環境危機、汚職、司法不正、公衆衛生、マネーロンダリングなど、他の差し迫った問題と並んで、よりソフトなトピックもデータジャーナリズムの舞台になっていることを示しています。加えて、データジャーナリズムの新しいビジネスモデルの可能性や、データベースの製品が政府や企業にどんな価値をもたらすことができるのかを探っているのです。

読者のニュース消費習慣もデータジャーナリズムの発展に影響を与えています。あるものはよりビジュアルに、あるものはよりモバイルなものとなっています。2011 年以降、いくつかの大手ニュースポータルがインフォグラフィックスのストーリーで頂点を極めようと努力するようになり、インフォグラフィックスは人気を博しています。そのほとんどはデータ駆動型です。14 年には、中国で 9 人いる（党）最高幹部の 1 人、周永康元公安部長の失脚が最大のニュースとなりました。財新は取材チームが作成した 6 万字の調査記事をもとに、インタラクティブなソーシャルネットワークのビジュアライゼーションを制作し、関係者 37 人や 105 の企業やプロジェクトを含む周氏を取り巻く複雑なネットワークと、これらの関係性を説明しました（Figure 10.2）。このインタラクティブは 1 週間で 400 万回のヒットを記録し、ソーシャルメディアではさらに 2000 万回のビューを記録したそうです。このプロジェクトの普及は、データ・ストーリーテリングの新手法を一般の人々に紹介し、それまでにはなかった新しいビジュアル・ストーリーへの欲求を生み出しました。

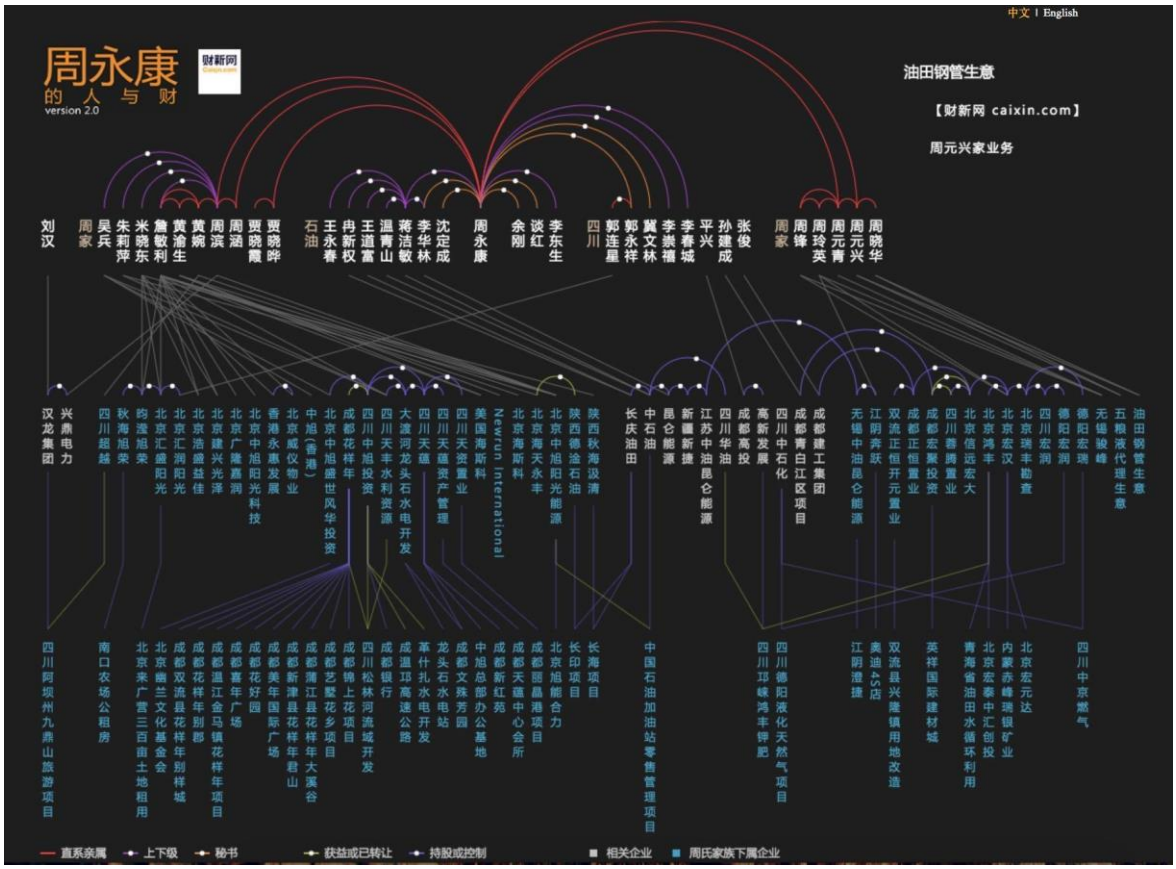


Figure 10.2. 周氏のネットワークを示すインタラクティブな作品　出典：財新

それとほぼ同時期に、メディア業界はモバイル時代を迎えました。ラボもそうであるように、中国の他のオンラインコンテンツでは、今ではデータ記事のほとんどがモバイルで発信されています。中国インターネット情報センター（CNNIC）によると、2016 年には国内のインターネットユーザーの 95％以上がモバイルデバイスを使ってインターネットにアクセスしているそうです（Chung, 2017）。国内で人気のあるメッセージングアプリやソーシャルメディアプラットフォームである WeChat は、18 年 3 月の時点でユーザー10 億人に達しています（Jao, 2018）。モバイルプラットフォームの優位性は、中国におけるデータ記事がモバイルファーストというより、多くの場合モバイルオンリーになっていることを意味します。

つまり、中国におけるデータ文化は、世界の動きから政府の法律まで、国民の要求からメディアの都合まで、新世代のデータプロバイダーから新世代のニュース消費者まで、様々な要因によって進化してきたのです。プレイヤー間の相互依存関係は、制約と機会が共存する、非常に複雑なダイナミクスを生み出している。データジャーナリズムは、中国で独自の道を歩みながら開花してきたと言えるでしょう。

実務上のヒント

ここでは、特に中国関連のストーリーを報じたいと考えている人、どこから手を付けようか悩んでいる人に向けた内容になっています。簡単なことではありません。ほとんどのデータソースは中国語でしか入手できないため、まず言語の壁があります。次に、データの正確性や完全性、不整合などの問題があります。これらの問題に対処する適切なスキルを持っているか、少なくとも学習意欲があると仮定してみましょう。まず、中国におけるデータジャーナリズムの最大の担い手を見極めましょう。あまたの大手メディアがデータチームを立ち上げていますが、彼らのストーリーを追いかけ、記者に話を聞いてヒントを得るのもいいでしょう。ここでは、知っておいた方が良いもののうちいくつかをご紹介します。

財新の Data Visualization Lab；澎湃新聞の Beautiful Data Channel;上海メディアグループの The Rising Lab; DT Finance[8]

第二に、どこでデータを見つけることができるのか（を知りましょう）。包括的なリスト[を挙げれば]別の本になってしまうので、ここではいくつかを紹介するだけにしておきます。中央省庁と地方機関の Web サイトから始めましょう。各省庁のテーマ別分野（例えば、生態環境部）と地方レベルのデータ専用サイトがあれば、その両方をチェックする必要があります。想定もしていなかったデータが出てくるかもしれません-例えば、中国政府が 2014 年以降の数百万件の裁判判決を全文公開したことをご存じですか？法的文書は米国では比較的透明性が高いのですが、中国ではそうではありません。しかし、最高人民法院（SPC）は「China Judggments Online」というデータベースを立ち上げました。オンラインで役立ちそうなデータを見つけたら、ローカルコピーをダウンロードするようにしましょう。ネット上にデータが出てこないこともあります。それはまだよくあることです。時にはそうしたデータは政府の年次報告書の形で公開されており、オンラインで閲覧できます。特定の省庁が所蔵する紙のアーカイブでのみ利用可能なデータもあります。例えば、ある政府機関は、民間企業に関する記録を持っていますが、そのすべてがオンラインで利用できるわけではありません。

行政が公開しているデータが全くない場合は、ユーザーが作成したコンテンツがないか確認しましょう。例えば、公衆衛生のデータは非常に限られていますが、病院登録や高齢者センターなどの専用サイトがあります。それらのデータをスクレーピングしてクリーニングしておけば、そのトピックの概要を把握するための貴重なデータが手に入るはずです。

Hong Kong Companies Registry のような公式データベースから、Webb-site Reports といった独立のデータベースまで、香港のデータベースを活用しましょう。中国本土と香港は政治的にも経済的にも緊密になってきており、透明性の高い環境と法執行が担保されているため、香港ではより多くの情報が入手できるようになりました。これは資金を追跡するのに役立つかもしれません。中国に関するデータは、必ずしも中国にあるとは限りません。中国関連のデータを豊富に持っている国際機関や学術機関があります。例えば、澎湃新聞の最近の記事では、NASA とハーバード大学のデータに基づいたものがあります。最後になりますが、中国

[での報道における]課題や経験には中国特有のものがあるものの、その多くは、社会的、文化的、政治的な違いを超えて、同じような制約がある他の国にも有益な教訓を提供できることでしょう。

脚注

1. www.gov.cn/jrzg/2010-11/25/content_1753524.htm（中国語、新華社通信、2010 年 11 月 25 日に北京でローンチされた全国大気質リアルタイム発表システム）

2. www.bjep.org.cn/pages/Index/40-1699?rid=2782 ..（中国語）

3. chrome.google.com/webstore/detail/pm25b....

4. nance.ifeng.com/leadership/gdsp/20110901/4512444.shtmll（中国語）

5. www.gov.cn/zhengce/content/2015-09/05/content_10137.htm（中国語）

6. zhuanlan.zhihu.com/therisinglab（中国語）

参考文献

Alibaba Group announces September quarter 2018 results. (2018, November 2). *Business Wire*. www.businesswire.com/news/home/20181102005230/en/Alibaba-Group-Announces-September-Quarter-2018Results

Edward Wong and Chris Buckley, 'Chinese Premier Vows Tougher Regulation on Air Pollution', New York Times, March 15, 2015

Chung, M.-C. (2017, February 2). More than 95% of Internet users in China use mobile devices to go online.

EMarketer.www.emarketer.com/Article/ More-than-95-of-Internet-Users-China-Use-Mobile-Devices-GoOnline/1015155

Gov't report: China's urbanization level reached 56.1%. (2016, April 20). *CNTV*. english.www.gov.cn/news/video/2016/04/20/content_281475331447793.htm

Jao, N. (2018, March 5). WeChat now has over 1 billion active monthly users world-wide. *TechNode*. technode.com/2018/03/05/wechat-1-billion-users/

Jing, C. (2015, February 28). Chai Jing's review: Under the dome—investigating China's smog. www.youtube.com/watch?v=T6X2uwlQGQM

Lijian, Z., Xie, T., & Tang, J. (2015, December 30). How China's new air law aims to curb pollution. *China Dialogue*. www.chinadialogue.net/article/show/ single/en/8512-How-China-s-new-air-law-aims-to-curb-pollution

Page, J. (2011, November 8). Microbloggers pressure Beijing to improve air pollution mon-itoring. *The Wall Street Journal*. blogs.wsj.com/chinarea... internet-puts-pressure-on-beijing-to-improve-air-pollution-monitoring/ internetputs-pressure-on-beijing-to-improve-air-pollution-monitoring/

Shen, J. (2018, October). Data journalism in China panel. Uncovering Asia, Investiga-tive Journalism Conference, Seoul. 2018.uncoveringasia.org/schedule

Spegele, B. (2012, January 23). Comparing pollution data: Beijing vs. U.S. embassy on PM2.5. *The Wall Street Journal*. https://blogs.wsj.com/chinarealtime/2012/01/23/comparing-pollution-data-beijing-vs-u-s-embassy-onpm2-5/

Wong, E., & Buckley, C. (2015, March 15). Chinese premier vows tougher regulation on air pollution. *The New York Times*. www.nytimes.com/2015/03/16/world/asia/chinese-premier-li-keqiang-vows-tougher-regulation-on-airpollution.html

インド全土の土地紛争を記録するデータベースの作成

文：クマール・サンバフ・シュリバスタバ、アンクール・パリワル

概要

土地紛争を記録することで、土地とインドの多様な社会との関係について、より深く、よりニュアンスのある、より大きなストーリーを伝えます。

キーワード： 土地紛争、インド、データジャーナリズム、データベース、コラボレーション

インドでは、土地は希少資源です。インドは世界の土地面積の 2.4％しか占めていませんが、世界の人口の 17％以上を支えています。世界でも有数の経済成長を遂げているインドでは、産業とインフラの成長という野心的な課題に取り組むため、広大な土地が必要です。今後 15 年間の開発プロジェクトには、少なくとも 1100 万ヘクタールの土地が必要と見積もられています。しかし、インドの人口の多くは、疎外されたコミュニティ（に属しており）、彼らの生活基盤と生活様式は土地に依存しています。（すなわち）2 億人以上の人々が森林に依存し、1 億 1890 万人が農地に依存しています。

［そうした人々の］競合する要求は紛争を引き起こします。多くの場合、土地は国や企業によって強制的に、または不正に奪われ、反対者は国家機関に冤罪で逮捕され、わずかな補償金でコミュニティは立ち退きを強いられ、家は放火され、人々は殺されています。カースト、階級、性別などの社会的格差もまた、土地紛争を煽っています。気候変動が引き起こす災害は、土地に依存しているコミュニティをさらに追い詰めています。インド全土で土地をめぐる多くの紛争が行われているのには、こうした背景があるのです。インドの開発問題について書き始めたとき、私たちはそのような紛争にたくさん遭遇しました。しかし、辺境のストーリーをニューデリーの編集者に売り込むのは容易ではないことに気がついたのです。主流メディアは、致命的な暴力に発展したものや訴訟になったものを除いて、土地の紛争を報道しませんでした。少数のジャーナリストによる散発的な報道は、ほとんど影響を与えず、紛争の影響を受けた人々の声は聞こえないままでした。彼らの懸念は放置されていたのです。［私たちは］その理由について、記者と編集者が紛争を［それぞれ］孤立した事件として見ていたからではないかと考えました。土地紛争がインドの政治経済に関する最も重要な話の一つであることは知っていました。しかし問題は、それを編集者や読者にどうやって売り込むかということでした。私たちは、ジャーナリストが紛争の個々の事例を巡る報道を拡充し、より広範な傾向を調べることができれば、記事がより多くの人々に届くだけでなく、さまざまな種類の紛争の激しさや、人々、経済、環境への影響を示すことができるのではないかと考えていました。最大の課題は、道路、行政区画、鉱業、野生生物保護地域など、特定の種類の紛争についてどのような傾向が出ているのかをジャーナリストが調べることができるデータベースがないことでした。インドには現在進行中の土地紛争に関するそのようなデータベースはありません。そこで、私たちはデータベースを構築することにしました。

2016 年 11 月には、インドで進行中のすべての土地紛争をマッピングし、文書化することを目的にした、調査ベースのデータジャーナリズム・プロジェクト「Land Conflict Watch」に着手しました。私たちは、土地統治を専門とする学者と相談しながら、文書化の方法論を開発しました。全国の研究者やジャーナリストのネットワークを構築し、この方法論に沿って、各地域の紛争を記録しました。このプロジェクトでは、土地の使用や所有権をめぐって対立する要求や主張があり、コミュニティが当事者の一人となっている状況を土地紛争と定義しました。村レベルから国レベルまで、あらゆる場所でそのような要求や主張が、すでに文書や音、映像で記録されている継続的な紛争も対象です。これらの記録には、ニュース報道、村議会の決議、開発プロジェクトの協議記録、人々が政府当局に提出した苦情、警察の記録、裁判所の文書などがあります。民間人同士の財産紛争や民間人と政府との間の訴訟などは、より広範な国民に直接影響を与えるものでない限り除外しました。研究者とジャーナリストは、自分たちの地域に関する全国および地元のメディアの報道を追跡し、地元の活動家、コミュニティ、弁護士と交流して紛争の事例を見つけ出します。そして、公開されている政府の文書や独立した調査、影響を受けた当事者への聞き取り調査などから情報を収集し、検証します。紛争の場所、紛争の背景にある理由、影響を受けた人々の数、影響を受けた地域、土地の種類（私有地、共有地、森林など）、関係する政府機関や企業の名前、紛争の概要などのデータを記録します。研究者は、すべてのデータを Land Conflict Watch のサイトに組み込まれた報道とレビュー用のソフトウェアにファイルします。データは専任の査読者によって調査・検証されることになっています。データが公開される前に、研究者と査読者との間で双方向の作業を行うために、このソフトウェアは役立っています。ポータル上のダッシュボードでは、進行中の紛争を国レベルでマクロに示すとともに、それぞれの紛争をズームアップし、詳細とデータを裏付ける資料を表示します。

現在、約 35 名のジャーナリストや研究者が貢献しており、18 年 9 月時点で、640 件以上を記録しています。これらの紛争は 730 万人近くに影響を与え、240 万ヘクタール以上の土地にまたがっています。影響を受けたプロジェクトやスキームには、1860 億米ドル相当の投資が行われています。紛争が記録されると、ポータルやソーシャルメディア上に掲載され、全国のジャーナリストや研究者に情報を提供します。その後、プロジェクトチームはジャーナリストと協力して、このデータを使って、土地の権利、土地の紛争、政治、経済、階級、ジェンダー、環境など多角的な観点から詳細な調査記事を作成します。また、国内外のメディアと協力して、これらの記事を公開しています。これらのストーリーの多くは、他の主要メディアに転載されています。ジャーナリストを対象に、データベースを利用して土地統治に関するストーリーを見つけ出し、スケールアップするトレーニングも実施しています。

Land Conflict Watch は現在進行中のプロジェクトです。私たちは、ストーリーをデザインするだけでなく、学者、研究者、学生と協力して公共の議論に取り組んでいます。プロジェクトのデータは、政策シンクタンクの報告書に引用され、土地統治の専門家は、このデータを使って全国紙に論説を寄稿しています。インドや海外の大学の研究生からも、研究にデータを利用させてほしいという依頼が定期的に寄せられています。非営利団体は、紛争の影響を受けたコミュニティの土地の権利のために闘うキャンペーンのために、土地紛争のデータや

文書、事例を利用しています。これらのストーリーは人々に情報を与え、インドにおける土地の権利とガバナンスに関連する問題に関する言論を形成するのに役立っているのです。

キューバの公共データを再構築する：情報が欠落、未更新、不足している場合のコラボレーション

文：ユディビアン・アルメイダ・クルス、サイミ・レイエス・カルモナ

概要

キューバの小さなデータジャーナリズムチームが、データ不足とどう戦ったか紹介します。

キーワード：キューバ、希少データ、データ学習、人工知能、データジャーナリズム、研究者とジャーナリスト共同研究

Postada.club は小さなチームです。ジャーナリストと数学とコンピュータサイエンスの専門家の4人で2014 年、キューバのデータジャーナリズムに挑戦を始めました。当時のキューバでは、データジャーナリズムに特化したメディアはなく、その実践がどのようなものか理解しようとしたのです。現在、私たちのチームは 2 人のジャーナリストと 1 人のデータサイエンティストで、空いた時間を Postdata.club に充てています。サイミ・レイエスは文化系 Web サイトの編集者、ユディビアン・アルメイダはハバナ大学数学・コンピュータサイエンス学部の教授、エルネスト・ゲラは人気の大衆科学雑誌のジャーナリストです。私たちの目的は、メディア組織にとどまらない実験的な空間の実現であり、私たちが住む国について、データを活用して探求し、学びたいと思っています。オープンな公共データを使えるようにするため、私たちの調査とその方法を共有することにしました。そこで、Postdata.club は Github で活動することにしました。伝えたいストーリーの要件に応じて、テキストの拡張子やグラフィック、画像、ビデオ、オーディオなど、使用するリソースを決定します。私たちは、社会的インパクトのあるジャーナリズムに焦点を当てています。私たちは、データを使ってアプローチできるあらゆるテーマに興味を持っていますが、何よりもキューバや国民に関連したテーマに興味を持っています。

データに応じて調査方法は異なります。公開されているデータベースやオープンデータベースにアクセスできる場合もあります。これらのデータベースを使って、私たちはストーリーがあるかどうかを確認するためにデータ分析を行います。（もともと）疑問があって、ストーリーの答えを見つけるために（必要な）データを直接調べることもあります。また、データを調査する過程で興味深いと思われる要素を見つけたり、その答えが関連性があり、そのデータソースや他のデータソースで答えが得られるかもしれない質問が出てきたりすることもあります。これらのデータベースから得られた情報が面白そうであれば、インタビューや他の情報源との比較など、他の情報源で補完します。そして、データから得られた洞察を可視化した上で、調査内容をどのように説明するかを考えます。公開されていても適切に構造化されていない情報をもとに自分たちでデータベースを作成し、それを分析や調査の基礎として利用しなければならないことも少なくありません。例えば、キューバの選挙というテーマに

取り組むためには、さまざまな情報源から得た情報を組み合わせデータベースを構築しなければなりませんでした。

まずキューバ人民権力全国会議のサイトで公開されているデータから始めましたが、完全なものではなかったので、報道やキューバ共産党関連のサイトで公開されている情報でデータベースを完成させました。最近任命指定された閣僚評議会にアプローチするためには、別のデータベースを構築する必要もありました。その場合、人民権力全国会議から得た情報は完全ではなかったので、全体像を把握するために、報道や官報、その他の情報サイトを利用しました。

どちらの場合も、JSON 形式のデータベースを作成し、キューバの選挙や行政府、立法府に関する記事のほとんどに使用しました。ほとんどの場合には、データベースを私たちの Web サイト上で、調査手法の説明とともに共有しています。しかし、キューバでの活動は、公開されていてアクセス可能であるべきいくつかのデータが不足しているために、複雑な状況に陥ることがあります。

利用している情報の多くは政府機関から提供されていますが、キューバでは公開すべき情報をインターネット上で適切に公開していないことがあったりします。いくつかのケースでは、これらの機関を直接訪れ、情報公開請求しています。この手続きはしばしば面倒ですが、重要です。キューバのデータを入手する際の最大の問題の一つは、その古さにあります。探している情報にようやくアクセスできたとしても、その情報は完全ではなかったり、非常に古くなっていたりすることが多いのです。データはウェブサイト上で相談やダウンロードが可能な場合がありますが、例えば、最新データは 5 年前のものだったりします。このような場合は、最新の情報を提供している信頼できる他のウェブサイトを確認するか、印刷物やスキャンした文書、人による情報源に頼ることになります。学生や研究者とのコラボレーションは、情報が不足していたり、古かったり、希少だったりする状況へのアプローチ方法の一つです。17 年以来、私たちはハバナ大学コミュニケーション学部のジャーナリズムを学ぶ学生たちにデータジャーナリズムのコースを教えています。こうした未来のジャーナリストやコミュニケーションの専門家との交流を通じて、私たちは新しい仕事のやり方を学び、情報にアクセスする新しい方法を発見してきました。

これらの授業で行っていることの一つに、学生を巻き込んでのデータベース構築があります。キューバでは、さまざまな分野や活動でのライフワークに基づいて（与えられる）、国家の賞の受賞者名簿の単一の情報源はありませんでした。学生と教師の全員で、27 以上の賞のこれまでの授与歴のデータベースを収集し、構造化しました。この情報から、賞の授与には男女間の格差があることが明らかになりました。女性が受賞したのは 25％に過ぎません。この発見をもとに、さまざまな種類の作品が国民に浸透しているにもかかわらず、ジェンダーの問題が存在することに考察を促す物語を書くことができました。またこれも 2017 年には、既存の公開されているデータベースに依存せず、何が可能で何が不可能なのかについて、あまり早い段階で決めつけてはいけないと思わせる驚くべき体験をしました。

授業の最後に、私たちは学生に小さなチームを作って課題をこなすように求めました。チームは、Postdata.club チームのメンバーの 1 人、ジャーナリズムの学生 2 人、コンピュータサイエンスの学生 1 人で構成されていました。チームの一つは、キューバの自営業の新しい取り組みを取り上げようと提案しました。これらの人々は「cuentapropistas［クエンタプロピスタ］」と呼ばれています。数年前までは稀な雇用形態でしたが、社会に受け入れられるようになり、急速に拡大してきています。そこで、キューバにおける自営業を調査することにしました。この問題は頻繁に取り上げられていましたが、州別の自営業の具体的な内容や、地域活動ごとのライセンス付与数、経時的な傾向などについてはほとんど分かっていませんでした。どのような問題に取り組むべきかを学生と議論した結果、使えるデータを持つ情報源が不足しているという結論に達しました。

この情報が公開されているはずの場所を探しても、何もありません。一部のインタビューや単発の数字を除いて、このテーマに関する情報はあまり見当たりませんでした。データを入手するのは難しいだろうと思われました。しかし、学生が労働・社会保障省にキューバの自営業についての情報を求めました。数日後、学生たちはデータベースを手にすることができました。こうして私たちは多くのキューバ人が興味を持つような情報を持つことになり、市民に共有することができました。実際には、その情報は公開されてはいたのですが、省は最新のインターネットポータルを持っておらず、我々は誤ってデータにアクセスできないと考えていたのでした。

ちょうどそのころ、労働・社会保障省は、非国家雇用のうち 28 の活動に対するライセンス付与の制限を決定したのです。私たちは、これらの新しい措置が国の経済や自営業者の生活にどのような影響を与えるかを分析するために、持っていたデータをすぐに利用することができました。ほとんどの読者は、このデータを入手したこと、しかも比較的簡単に入手できたことに驚いていました。学生たちが省庁に問い合わせをしたことで、このデータにアクセスする道が開かれました。今日に至るまでこの情報が公開されているのは Postdata.club だけで、誰もが閲覧し分析することができます。

キューバでデータジャーナリズムを実践するのは、依然として挑戦であると言えます。特に、データの作成とアクセスを巡る力学、政治的・制度的な文化は、データが容易に入手できる他の国とは異なります。そのため、私たちは重要な問題についての記事を書くため、常にクリエイティブに情報にアクセスする新しい方法を模索する必要があります。Postdata.club は、データが入手しにくい地域でもデータジャーナリズムが可能であることを示すお手本となるよう、常に努力していきたいと考えています。

ラ・ナシオンで読者とともにデータを作る

文：フロール・コエリョ

概要

シビックマラソンと オープンソースプラットフォーム Vozdata を使い、読者、大学、NGO と共同で大規模なデータ駆動型調査を実施しました。

キーワード：シビックマラソン、クラウドソーシング、調査報道、オープンソース、研究者とジャーナリスト共同作業、オープンデータ

［アルゼンチンの日刊紙］ラ・ナシオンは、読者と協力して大規模なデータ駆動型の調査を行ってきました。ここでは、目標設定、調査コミュニティの支援、読者や外部組織・パートナーとの長期的な協力関係の構築など、読者参加型プロジェクトの舞台裏をご紹介します。このようなプロジェクトでは、テクノロジーを利用して大規模な共同作業を行い、ユーザーが調査報道や公式データの公開プロセスに参加できるようにすることで、「不可能」に挑戦しています。

例えば、私たちは、上院の経費に関する 1 万件の PDF を書き写すのに約 5 年、4 万件の電話を聞き取るするのに 2 年、2 万件以上の手書きの選挙用紙をデジタル化するのに数カ月を費やしました1。このようなクラウドソーシングの取り組みには、オンラインのコラボレーションプラットフォーム「Vozdata」を利用しました。Vozdata は、（英紙）ガーディアンの MPs’ expenses やプロパブリカの Free the Files といったクラウドソーシングキャンペーンに触発され、Knight-Mozilla OpenNews と、市民社会団体や活動家の世界的な連合体である（国際 NGO）CIVICUS の支援を受けて開発されました。Vozdata のソフトウェアは、Crowdata としてオープンソース化されています2。

参加者を組織する

これらのプロジェクトでは、主にジャーナリズムを学ぶ学生、市民ボランティア、透明性のある NGO、退職した市民などが協力してくれました。参加者の動機は、プロジェクトによって異なります。公益プロジェクトへの貢献、データチームとの共同作業、ミートアップで他の人たちと知り合いたいなどです。Vozdata には、チームとライブランキングの機能があります。私たちは、「ゲーミフィケーション」によって参加意識を高める方法を模索しています。アルゼンチンの祝祭日にこの方法で市民参加を促したところ、素晴らしい結果が得られました。共同データベース構築への参加は、ほとんどがリモート（オンライン）で行われます。しかしながら、ラ・ナシオンで開催される「オフライン」の市民イベントや、さまざまなイベントで行われるハッカソンへの参加も呼びかけています。また、パートナー大学のジャーナリズム専攻の学生と一緒に、オープンな（自由に再利用できる）データベースを構築することもありました。ハッカソンは通常 1～2 日で終わるイベントですが、私たちのオンラインマラソン

は数ヶ月に及ぶこともあります。プログレスバーには、完成したドキュメントの数と残りのパーセンテージが表示されます。

大きな目標を立てる

「Senate Expenses」と「Electoral Telegrams」プロジェクトの共同研究者の主な役割は、提供された文書から特定の構造化データを収集することでした。これには 1000 人以上のユーザーが参加しました。読者は、詳細なデータを抽出するだけでなく、疑わしいデータや受け入れがたい点を指摘したり、追加情報をコメントとして残したりすることができました。[とはいえ、]後者の機能はほとんど使われませんでした。

クラウドソーシングのプロジェクトでは、締め切りがあると目標を達成できないことがあります。Electoral Telegrams のプロジェクトでは、そうなりました。投票日が近づいてきて、結論を発表する必要がありました。100％達成した県もありましたが、多くの県では 10％から 15％のファイルしか完成しておらず、公開時にそのことを注記しました。

Figure 13.1. 2017年にラ・ナシオンで開催された senate expenses のシビックマラソンのオープニング 出典：ラ・ナシオン

調査コミュニティの支援

ニスマン検事の 4 万に及ぶファイルの調査では、100 人からなる信頼できるネットワークと協力しました。イラン人諜報員の個人的な電話（家族の会話など）を連邦裁判所が盗聴した音声ファイルが多くありました。

有志 6 人のグループは、この調査に没頭しました。WhatsApp グループを作り、誰もが手がかりや好奇心を提案できるようにしました。［すると、］ボランティアの一人が、私たちを 2 ヶ月間にわたり悩ませた謎を解決しました。私たちは、2 人の人間がニックネームと数字を使って暗号で会話している会話にいくつかフラグを立てていました（例：「Mr.Dragon, 2000」）。多くのボランティアがそのような録音を書き起こしていました。私たちは、これらの暗号を分析するために、別のデータベースを作ることを考えました。ある日、ボランティアが、会話の内容が競馬賭博であることを発見したのです！Google で検索してみると、多くの名前が競走馬であることがわかりました。パワーユーザーは必ずいるものです。しかし、プロジェクトの規模にもよりますが、多くのボランティアがそれぞれ数枚のドキュメントを共同作成すると、しばしば「スーパーユーザーの貢献度」を超えることができるのです。

コラボレーションを育む

データ調査に読者を参加させたいと考えているジャーナリストや組織にアドバイスしたいのは、専任のコミュニティマネージャーを任命して、共同作業用のスプレッドシート（Google Sheets など）やメーリングリスト、ソーシャルメディアを利用したコミュニケーションを整理したうえで提供することです。大規模なドキュメントの収集から始めるのが良いでしょう。学習曲線が早く、参加者は大きな［プロジェクト］の一部であることを感じることができます。また、ビデオチュートリアルや文脈に沿った［事例・手法の］紹介を社内やイベントで行い、協力者をサポートすることも重要です。これらの共同プロジェクトに関連する賞を受賞した際には、朝食会を開き、ボランティアの方々と栄誉を分かち合いました。これらは読者との長期的なお付き合いといえるので、イベントでの待ち合わせ、大学訪問、学生プロジェクトのためのインタビューなど、時間とエネルギーを割くようにしました。

大学とのパートナーシップについては、通常、教授が結節点として機能します。毎年、大学には新しいクラスの学生が入学しますが、彼らは通常、私たちと共同プロジェクトを行うことを望んでいます（事前に計画を立てておきましょう！）。透明性の高い NGO は、これらのプロジェクトの利点を示すこともできます。私たちのプラットフォームでは、すべてのタスクを登録することができるので、プロジェクトを簡単に紹介したり、ドナーのためにメディアで紹介したりすることができます。成果物やストーリーを公開する際には、すべてのプラットフォーム（印刷物、オンライン、ソーシャルメディア）や郵送文書で、協働のプロセスや参加組織を認識することをお勧めします。プロジェクトの集合体としての性格を強調することで、責任を負わせたい人たちに、より強いメッセージを送ることができます。

結論

読者とデータを共有するためには、時間とリソースを割いてコミュニティに関わり、リクエストに対処し、アウトプットを分析し、交流を楽しみ、イベントに参加することが不可欠です。

ボランティアが文書を分類するのは、そのプロジェクトが重要だと考えるからです。政府や報道される側にとっては、そのプロジェクトが報道機関の関心事であるだけでなく、市民社会にも影響を与えていることを示します。このようなプロジェクトを通じて、参加者は熱心な支持者となり、コンテンツのオンライン配信者となることができるのです。

脚注

1. blogs.lanacion.com.ar/projects/data/argentina, blogs.lanacion.com.ar/projects/data/prosecutor, blogs.lanacion.com.ar/projects/data/vozdata

2. blogs.lanacion.com.ar/projects/data/vozdata--ii-civic, theguardian.com/news/datablog/2009/jun/18/mpsexpenses-houseofcommons, propublica.org/series/free-the- les, github.com/crowdata/cr...

調査報道のためのアンケート

文：クリナ＝ガブリエラ・ボロシュ

概要

説明責任を果たすジャーナリズムに向けてアンケートを実施する前に、まずこの記事をお読みください。必要なこと、やってはいけないこと、不完全な状況に対処する方法などについて。

キーワード：統計、データジャーナリズム、アンケート、説明責任

その問題は[1回限りの事例が並ぶ]逸話的なものなのか、それとも体系的なものなのか。表形式のデータ-行と列で提供される情報のしゃれた言い方-がないことに気付いたとき、あなたはそれを見極めようとします。どうすればいいのでしょうか？そもそも、データとは何か1？オタク的な定義がたくさん出回っていて、中には人を萎縮させるものもありますが、それらを「情報」というシンプルな概念に置き換えてみましょう。どんな形であれ、情報を集めるには、パターンや異常値を見つける能力が必要です。つまり、ある問題を特定の方法で文書化し、体系的に収集した膨大な量の原材料が必要になるのです（記入式のフォームを考えてみてください）。スプレッドシート、（プログラミングの）開発環境、アプリが、紙とペンのどれを使っても構いません。人の心に閉じ込められた思いや感情、過去の親密な体験などは、時にデータとして表現することができます。このような貴重な情報を得るためには、そのような感情や経験を集めて整理し、自分以外の誰もアクセスできないテーブルやアーカイブ、データベースを作成するような調査を設計することが一つの方法です。

例えば、トムソン・ロイター財団（TRF）は、世界の大都市に住む女性たちが、公共交通機関での性的暴力をどのように受け止めているかを報じるプロジェクトを実施しました 2。これは、この問題に対する認識を高めると同時に、比較対照を行うための調査でもありました（統計学が得意とするところです）。このスポットライトを提供するために、私たちは地獄巡りをしました。アンケートのような社会科学的手法には、ジャーナリストが実務に取り入れる際にも必要とされる厳密な規則があるからです。ここでは、ジャーナリストが知っておくと便利ですが、トレーニングを受けていないことが多い、主なポーリング・ルールを紹介します。

回答者を選ぶことはできない。「代表的」な回答者と見なされるためには、社会的カテゴリー、年齢層、教育レベル、地理的エリアなど、報道すべき対象のすべての人々が回答者として含まれている必要があります。確立された方法によれば、調査対象となる人口のサンプルは代表的である必要があります。

回答者の選択は無作為である必要がある。つまり、誰もが同じ確率で[手品のように無作為に]帽子から名前を取り出せるようにしなければなりません。一般論的な主張をしたいのに、何の基準も方法もなく手近な人に声をかけて世論調査を行った場合は、誤解を招くようなデータを導く危険があります。

アンケートを代表的なものとするには、参加者数が一定の基準に達している必要がある。Raosoft、Survey Monkey、Survey Systems などは便利なオンライン計算ツールを提供しています3。経験則として、信頼度（信頼係数）は 95%、誤差範囲は 5%以下にしてください。回答の選択肢は、回答者が「わからない」または「確信がない」と答えられるものでなければなりません。

記者がこれらの基本的なルールに従えば、調査結果はほとんど批判されることはありません。TRF の公共交通機関の安全性に関する調査でも、私たちの世論調査の方法は、社会科学の保守的なルールに従っています。

今回のテーマは、社会がどう機能しているのか物語るような人間の共通体験を扱ったもので、国連機関からも協力の申し出がありました。光栄なことですが、ジャーナリストとしてはお断りしなければなりませんでした。次の言葉が気に入った人は、統計学のコースを受講してみるといいでしょう。

「厳密な世論調査は時に非現実的だ。だからといって、世論調査をすべきでないわけではない」。

調査には確立された方法がありますが、それだけでは可能なこと、正当なこと、興味深いことを網羅しているわけではありません。あなたの関心事や制約、リソースに応じて、他の方法で調査を行うことができるかもしれません。例えば、[英国の政治サイト]openDemocracy が欧州理事会加盟国 47 カ国の記者にニュースルーム内の商業的圧力についてインタビューを試みたとき、統計的な有意性が得られる可能性はほとんどありませんでした。

「なぜ？」と思われるかもしれません。回答者全員が内部告発者だったからです。内部告発者には保護が必要であり、年齢や性別などの重要な属性情報を開示しないことも必要です。私たちは、言論の自由を行使すると厳しい結果を招く可能性のある国からの投稿を期待していました。個人情報の提供は義務ではなく、提供されたとしても、情報を共有する会社のサーバーに置かれるべきではないと考えました。EU では、地域によってジャーナリストの数が大きく異なり、また不完全であったため、国レベルの代表的なサンプルを作成することは困難でした。会員名簿は非公開なので、すべての報道機関の組合や協会を並べて無作為に回答者を決めることはできませんでした。すべての人が含まれているわけではありませんが、自分たちの限界を正直に話していれば、許容範囲内のベースになるでしょう。さらに、国によっては、透明性を高めるためのプロジェクトが弾圧につながることもあり、どの国で組合の支援を求めれば、監視や処罰の対象にならないのか、専門家のアドバイスを受けました。このような場合、風呂の湯と一緒に赤ちゃんを捨ててしまうことはありません。私たちはそうしませんでした。その代わりに、私たちの報道にとって何が重要なのか、ストーリーを伝えるために世論調査の方法をどのように調整すればよいのかを明らかにしました。私たちは、国内の報道機関における商業的圧力の事例に注目しました。それらの事例にパターンがあるのか、またそのパターンは地域によって異なるのかということに注目しました。

また、報道を通じたイメージロンダリングの疑いで非難される団体はどのような種類なのかといったことにも関心がありました。そこで、ジャーナリストへのインタビュー、メディアの自由に関する報告書、フォーカスグ

ループでの意見などを参考にして、アンケートを作成しました。自由回答欄も設けました。私たちは、すべての綿密に調査されたジャーナリズム組織のチャンネルを通じて調査を進めました。つまり、無作為抽出はしないものの、どの報道機にアンケートを送るかはコントロールしなかったのです。「国境なき記者団」、「ジャーナリスト全国組合」、「欧州ジャーナリスト連盟」などのパートナーがアンケートを周知してくれました。アンケートに寄せられた意見は、専用データベースに登録され、回答に点数を付け、国ごとに回答者をカウントし、逸話的な問題（散発的に報告される問題）と体系的な問題（全般的に報告される問題）を比較できました。

特に便利だったのが、自由回答欄です。回答者は[さまざまな]情報を提供してくました。私たちは経済的な検閲のパターンや、不正行為を行ったとされる人物のタイプに注目しながら、回答者のフィードバックを調査しました。その結果、報道の自由に関する記事を作成することができました4。調査結果の概要は発表しましたが、データの内訳は発表しませんでした。その理由は、無作為抽出ができず、国レベルのサンプルサイズが必ずしも達成できなかったという単純なものです5。しかし、自国のスタッフによれば、報道機関がどの程度自由であるのか、メディアの腐敗はどのようにして起こり、どのように進むのか、そして悲しいことに、記者や真実がどれだけ脆弱であるのかについて、かなり理解することができました6。

ルールを破るためのルールはあるのか？

ほんの一部を紹介します。自分の努力を常に正確に表現すること。政府の経済顧問 3 人に「はい」「いいえ」で済む質問をしたのなら、そう言ってください。10 人のいじめ被害者にインタビューしたのなら、どのようにして彼らを選んだのか、なぜ彼らなのかを説明することです。インタビューを安易にアンケートと表記しないこと。

統計的に有意な研究を行った場合は、その手法を公開する親切さが必要。そうすることで、読者や専門家があなたの記事を信頼するために必要な精査を行うことができます。方法論がなければ、信頼は得られません。

次の大きな「フェイクニュース」の作者にならないように。編集者が、正確なデータ収集ではなく、推論に基づいて相関関係を描くように迫ってきた場合は、因果関係や科学的根拠を示唆しない言葉を使いましょう。私たちの仕事は、事実だけでなく、真実を報道することです。何が真実なのか確信が持てないことを隠すために、事実を利用してはいけません。

あなたのストーリーはどこにありますか？パターンの中、それとも異常値（に着目します）か？この答えに基づいて、どのようなデータを収集するのかを決定します。どこで、どのようにデータを得ることができるかを把握した上で、最も適切な方法を決めます。方法が重要なのではなく、ストーリーが重要なのです。調査を実施する場合は、結果をフィールドテストし、潜在的な問題の指摘から記事を守ります。例えば、自分が住んでいる地域の犯罪率が最も高いという調査結果があったとします。あなたは安全だと感じていたのに、1 年間住んだ別の地区では、ほぼ毎週のように路上暴力を経験したとすれば、まだデータを信用してはいけません。

データが信頼できるかどうかを確認するには、比較対照した場所を訪れ、道行く人、お店、銀行、パブ、学校などで話を聞き、どのようなデータが収集されたのかを調べることです。

ある地域の住民は、他の地域よりも苦情を申し立てる傾向がありますか？どのような種類の犯罪を取り上げようとしているのでしょう？分析で用いた犯罪の種類は重み付けされていますか？それとも窃盗は殺人と同等の扱いになっているのでしょうか？このような「実際の観察」によって、データを評価し、分析結果をどの程度まで信頼できるかを判断することができます。

脚注

1. カイム・ジンズ（2007）のデータ、情報、知識の 130 の定義を参照。Conceptual approaches for de ningdata, information, and knowledge. Journal of the American Society for Information Science and Technology, 58, 479–493.
2. news.trust.org/spotligh...
3. www.raosoft.com/samples..., www.surveymonkey.com/m..., www.surveysystem.com/s...
4. www.opendemocracy.net/...
5. www.opendemocracy.net/...
6. www.opendemocracy.net/...
7. news.trust.org/spotligh...

データジャーナリズム：フェミニズムは IT とどう関係する？

文：キャサリン・ディグネイジオ

概要

データジャーナリズムにフェミニスト的なアプローチをとること。それは、データベースやアルゴリズムに不平等がどのように入り込んでいるかに注意を払い、それらのバイアスを緩和する戦略を練ることです。

キーワード：データジャーナリズム、フェミニズム、ジェンダー、倫理、不平等、データベース

過去 70 年間のテクノロジーの進歩により、人々はかつてないほど多くの情報を保存し、処理できるようになりました。Google、Facebook、Amazon、Microsoft、Apple など、世界で最も成功しているテクノロジー企業は、データを集約することで収益を上げています。ビジネスや政府においても、「データ駆動型」の意思決定がますます重要視されています。しかし、データは平等に分配されていません。データを扱うためのスキルや、データを保存したり処理したりするために必要な技術的リソースも平等ではありません。データを扱う人々は、ふつうの人々を代表しているわけではありません。つまり、男性で白人、グローバルノース出身かつ高学歴な人が多いのです。データエコシステムにこのような基本的な不平等があるだけに、データジャーナリズムにフェミニズム的なアプローチをとることは、情報のパイプラインにおける隠れたバイアスを明らかにします。

フェミニズムとは、簡単に言えば、社会的、政治的、経済的に男女が平等であることを信じ、その信念に基づいて組織的に活動することと定義できます。フェミニストの概念とツールは、ジェンダーを分析の中心（唯一ではない）として社会的権力に疑問を投げかけるのに役立ちます。現代のフェミニズムの特徴の一つは、交差性を主張していることです。これは、権力の不均衡がどのように真実を覆い隠してしまうかを考える際に、性差別だけでなく、人種差別、階級差別、能力差別、その他の構造的な力も考慮しなければならないという考え方です 1。権力に真実を語る」という職業上の慣習に共感するジャーナリストにとって、フェミニスト的アプローチは非常に身近なものに感じられるでしょう。

このエッセイはデータ処理のパイプラインのいくつかの段階（データ収集、データコンテキスト、データコミュニケーション）を見渡し 、フェミニストのレンズを採用する機会と、バイアスの落とし穴を指摘するものです。フェミニスト的アプローチは女性やジェンダーの問題に関連するデータに有効なだけでなく、人間や人間の制度に関するあらゆるプロジェクト（ほぼすべてのプロジェクト）に適していることに着目してください 2。

データ収集

権力がどのように働き、誰が得をするのかを検証することは、常にフェミニストのプロジェクトの中核でした。社会学者のパトリシア・ヒル・コリンズが提唱した「支配のマトリックス」という概念は、権力は複雑であ

り、「純粋な被害者や抑圧者はほとんどいない」ということを理解するのに役立ちます（Hill Collins, 2000, p. 332）。

私たちは、不正を対人関係の領域で考えがちですが（例：性差別的な発言）、変化を起こすためには、理解して暴露しなければならない体系的な力があります（例：データを収集する機関における性差別）。

構造的不平等は、二つの方法でデータ収集に立ち現れます。まず、データ収集の過程で特定の（属性を持つ）人が過大評価されます。過大評価は、一般的に権力者が権力のない者に行う監視に関係しています。例えば、ボストン警察は、2015 年に路上尋問プログラムに関するデータを発表しました。このデータによると、警察は黒人、移民、ラテン系住民の居住区を重点的に巡回し、若い黒人男性を過度に取り締まっています。このような過剰カウントのケースでは、どのグループが権力を持ち、どのグループが監視の対象になりやすいかに注意を払うことが重要です。データジャーナリストの役割は、格差を認識して数値化するとともに、構造的な力（この場合は人種差別）が働いていることを明らかにすることです。

二つ目は（逆に）数が少なかったり、まったく数えなかったりすることです。例えば、メキシコのフェミニサイド（ジェンダーに基づく殺人）に関する最も包括的なデータベースは、なぜ「プリンセサ」というペンネームの女性によって収集されているのでしょう3。［メキシコ北部の］シウダー・フアレスをはじめとして国内の女性の死が増え続けているにもかかわらず、2006 年にフェミニサイドに関する特別委員会が設立されたにもかかわらず、09 年に米州人権裁判所がメキシコ政府に対して下した判決にもかかわらず、メキシコ政府は今のところフェミニサイドを包括的に追跡していません。女性や有色人種に関連する多くの問題では、集計機関が自らに責任のある被害を組織的に無視しているため、過少集計が発生しています。つまり、集計環境が損なわれているということです。集計が不十分な場合、データジャーナリストは、プリンセサが行ったようなことができます。それは、可能な限り自分で数えることです。一般的には、クラウドソーシング、記事の収集、統計的推論を組み合わせて行うことになります。米国では、警察官による殺傷事件や妊産婦の死亡率などが過少集計評価の例として挙げられますが、いずれもジャーナリストによるデータ収集プロジェクトとして取り上げられています。

データのコンテキスト

オープンデータ運動や API の普及は、データジャーナリストにとって好ましいことのように思われますが、「野放しの状態」で得られたデータには、特に人間や社会の現象に関して懸念がつきまといます。フェミニスト哲学者のダナ・ハラウェイは、すべての知識は「状況の中で位置づけられる」、つまり、知識は常に社会的、文化的、歴史的、物質的な文脈の中に位置付けられると言っています。データセットがこれらの文脈の産物であることを解明し、調査することは、権力や特権が真実を覆い隠している可能性を理解するのに役立ちます。

例えば、私のデータ・ビジュアライゼーションのクラスの学生は、大学キャンパスでの性的暴行について最終プロジェクトを行いたいと考えていました 4。米国の大学では、クレリー法に基づき、性的暴行などのキャンパス犯罪を毎年報告することが義務付けられているため、この問題に関する包括的で全国的なデータベース

があると思われました。しかし、このデータは学生たちを悩ませました。例えば、ウィリアムズ大学は、他の都市部の大学と比較して非常に高い数値を示していました。学生たちは、データの背景や収集環境の構造を調べた結果、データの数字が真実とは正反対のストーリーを語っていることに気づきました。性的暴行は汚名を伴う問題であり、被害者はしばしば非難と報復を恐れ、名乗り出ません。そのため、報告数が多い大学は、サバイバーが安心して報告できる環境づくりに多くのリソースを投入しているところでした。逆に、性的暴行の報告数が少ない大学では、サバイバーが沈黙を破ることをよしとしない敵対的な環境がありました。

ここには「落とし穴」と「チャンス」があります。落とし穴とは、ジャーナリストが Web からダウンロードした数字を額面通りに受け取ってしまうことです。これには、権力の差、社会的スティグマ、可視化されることに関する文化的規範などが含まれます。例えば、女性、移民、有色人種などのグループは、一般的に団体数を数えることが難しいのですが、これには極めて大きな理由があります。このように、データを活用したストーリーは、まだまだたくさんあります。データジャーナリストは、常に数字を使って新しい分析を期待するのではなく、数字を使って収集環境を調査し、欠陥のある慣習や権力の不均衡を指摘し、集計機関が本当に重要なものを計上するように方法を変えることができます。

データコミュニケーション

現代の西洋のデータに関する考え方は、合理的で客観的と思われるものが、感情的で主観的と思われるものよりも評価されるという「マスターステレオタイプ[強固な型にはまった見方]」から発展してきました（どの性別が「合理的」でどの性別が「感情的」かを考えてみてください）。

マスターステレオタイプは、感情が判断を鈍らせ、[対象と]距離[を保つこと]が客観性を担保すると言います。しかし、フェミニストの視点は、このマスターステレオタイプのすべてに疑問を投げかけます。感情は判断を曇らせるものではなく、好奇心や関与、もっと知りたいという動機を生み出すものなのです。例えば、パトリシア・ヒル・コリンズ（2000, p. 266）は、理想的な知識の状況を「倫理も感情も理性に従属していない」状況と表現しています。これはデータコミュニケーションにとってどのような意味を持つのでしょうか。

これまでのデータ・ビジュアライゼーションでは、より合理的であるとしてなるべく簡素なチャートやグラフィックが好まれていましたが、研究者やジャーナリストは、ビジュアライゼーションのユニークな特性を創造的なレトリックとして活用することで、より印象的で共有されるようになることを学んでいます。例えば、1984 年にナイジェル・ホームズが作成した「Monstrous Costs」というグラフは、政治運動にかかる費用の高騰を表しています。以前は「ジャンクチャート」の例と揶揄されていましたが、現在ではほとんどの人が直感的に理解していることが実証されています。読者の中には、退屈な棒グラフよりもモンスターの方が好きな人がいるのです（Bateman et al., 2010）。

他のあらゆるコミュニケーション媒体と同様に、データに感情を利用することには倫理的な責任が伴います。研究者は最近、ビジュアライゼーションのタイトルが、解釈に重要な影響を与えることを示しました

(Borkin et al)。典型的なタイトルの付け方は「合理的」な傾向にあります。つまり、「2012～14 年の大学キャンパスにおける性的暴行の報告」のように、データを中立的かつ客観的に描いているのです。しかし、女性やその他の疎外されたグループに関連して、中立的なタイトルがデータによって描かれたグループに害を及ぼすケースが多々あります。

例えば、性的暴行のケースでは、中立的なタイトルは、私たちが持っているデータが真実で完全なものであることを示唆しますが、実際にはそれは全くの誤りであることがわかっています。また、「2012～14 年に警察に殺害された精神疾患の女性」のような中立的なタイトルは、有害な固定観念を永続させることになります。それはまさに、米国で精神疾患の女性が過度に警察からの暴力の犠牲者となっている、能力主義や性差別などの構造的な力を明らかにしていないからです。

Figure 15.1. ナイジェル・ホームズのチャート。Designer's Guide to Creating Charts and Diagrams より (1984)。

データジャーナリズムにフェミニスト的なアプローチをとることは、現行の慣行や制度がエリート男性が頂点に立ち、その他の人々はコリンズの言う支配のマトリックスの様々な交差点に位置づけられる現状支持するというものだということに理解を深めることでもあります。家父長制、白人至上主義、入植植民地主義は、構造的な力であるため、データ駆動型調査やビジュアライゼーションが特に適しています。私たちは、データジャーナリズムの常識に疑問を持ち、不用意に現状を永続させないようにすると同時に、そのツールを使って不正義を暴き、解体する必要があります。この方向にさらに進みたい人は、ローレン・F・クラインとの共著である私の著書「データ・フェミニズム」(2020 年)を参考にしてください。この本では、フェミニズムの概念がデータサイエンスやデータコミュニケーションにどのように適用されるかが詳しく紹介されています。

脚注

1. アイデンティティの他の要素がジェンダーとどのように交差しているかを考慮しないフェミニズムは「ホワイト・フェミニズム」というべきである。交差性は法学者のキンバリー・クレンショーが初めて提唱し、ジェンダーの不平等は人種や階級に基づく不平等と切り離して考えることはできないと主張した黒人フェミニストの知的遺産から出てきたものである。

2. 例えば、ククタイとウォルターの先住民データ主権の章を参照。

3. feminicidiosmx.crowdma...

4. パトリック・トーフィー、ミカエル・ガニオン、ジリアン・ミーハンによる最新話はこちらで公開されている: cleryactfallsshort.ata...

参考文献

Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010). Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems—CHI '10* (pp. 2573–2582). dl.acm.org/doi/10.1145...

Borkin, M. A., Bylinskii, Z., Kim, N. W., Bainbridge, C. M., Yeh, C. S., Borkin, D., P ster, H., & Oliva, A. (2016). Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 519–528. ieeexplore.ieee.org/do...

D' Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

Hill Collins, P. (2000). *Black feminist thought: Knowledge, consciousness and the politics of empowerment*. Routledge. www.hartford-hwp.com/ar...

パナマ文書とパラダイス文書を巡るコラボレーションの構築

文：エミリア・ディアス＝ストラック、セシル・シリス＝ガレゴ 、ピエール・ロメラ

概要

国際調査報道ジャーナリスト連合（ICIJ）は、膨大な量の文書やデータをいかに効率的に掘り起こしているのでしょうか。

キーワード：データリーク、テキスト抽出、ラディカル・シェアリング、国境を越えた調査、データジャーナリズム、国際調査報道ジャーナリスト連合

国際調査報道ジャーナリスト連合（ICIJ）は、1997 年に発足したジャーナリストの国際ネットワークです。ICIJ の大規模な共同作業に参加しているジャーナリストは、多様な経歴とプロフィールを持っています。データやコーディングに強い記者もいれば、最高の情報源と現場取材に長けた能力を持つ記者もいて、それぞれが異なるスキルを持っています。全員がジャーナリズム、コラボレーション、データへの関心で結ばれています。ICIJ を率いるジェラード・ライルは、オーストラリアの消防局の不祥事を 3 年間にわたって調査した末、世界中のタックスヘイブンや人々に関連する企業データが入ったハードディスクをオーストラリアで受け取りました。そのときはそれがジャーナリズムにおけるコラボレーションをどれほど変えるか想像していなかったそうです。2011 年に ICIJ に着任した彼は、260 ギガバイトを超えるオフショア企業に関するデータ、約 250 万ファイルを持っていました。結果的に 46 カ国 86 人以上のジャーナリストの協力を得て、『オフショア・リークス』（2013 年に出版）という形で公開されました1。

オフショア・リークスの後には、大規模なデータセットや数百万のファイルを扱う調査プロジェクトが増え、それらを調査するためのアドホックな開発技術や、それを報道するジャーナリストのネットワークも増えました。例えば最近、私たちはパナマ文書の調査の中心となった法律事務所、モサック・フォンセカから流出した 120 万件の文書を新たに入手し、パートナー企業と共有しました2。これらの文書は、15 年にドイツの南ドイツ新聞から持ち込まれた 1150 万件のパナマ文書のファイルと、その後のパラダイス文書の調査の基礎となった 1,360 万件の文書に追加されました3。一人のジャーナリストがパラダイス文書のファイルをそれぞれ 1 分で読むとしたら、すべてに目を通すには 26 年かかるそうです。もちろん、それは現実的ではありません。

どうすれば近道を進めるのか？どうすればもっと効率的に、時間をかけずに調査できるのか？どうすれば、この巨大な文書の山から新しい手がかりを見つけ出し、コラボレーションモデルをサポートすることができるのか？

ここでは、流出した大量の文書を扱う際に、高度な「ビッグデータ」技術だけでなく、以下のようなアドホックな分析装置を用いて対処する方法をご紹介します。（a）国際共同ネットワーク（b）安全な通信手段とインフ

ラ(c)非構造化文書から構造化データを作成するためのプロセスとパイプライン(d)グラフデータベースと探索的ビジュアライゼーションによる関連性の検索

パートナーとの連係

ICIJ のモデルは、世界的なジャーナリストのネットワークを使って世界の税制を調査することです。私たちは、5 大陸のトップ記者たちを集めて、調査活動を改善し、国と国とのデータの点をつなぎます4。税のストーリーは、ピースのないパズルのようなものです。エストニアの記者がある部分を理解し、ブラジルの記者が別の部分を見つけることがあるかもしれません。それを一緒にすれば全体像が分かります。ICIJ の仕事は、こうした記者をつなぎ、記者がデータから見つけたものをすべて共有できるようにすることです。私たちはこの理念を「ラディカル・シェアリング」と呼んでいます。ICIJ のパートナーたちは、作業中に発見したことを、身近な同僚だけでなく、地球の裏側にいるジャーナリストたちにも伝えています。

協働を促進するために、ICIJ は、オープンソースのソフトウェア・コンポーネントをベースにした「Global I-Hub」と呼ばれるコミュニケーション・プラットフォームを提供しています5。このプラットフォームは、ユーザーから「プライベートな Facebook」と呼ばれ、実際のニュースルームで行われるような直接的な情報共有を可能にしています。記者は、国、スポーツ、芸術、訴訟など、特定のテーマを追うグループに参加します。そのグループ内では、データの中から見つけた政治家や、調べている特定の取引など、さらに具体的なトピックについて投稿することができます。ここでは、ジャーナリストが情報を照合したり、メモや興味深い文書を共有したりして、ほとんどの議論が行われます。ICIJ は記者が I-Hub を使いこなせるようにするため、いくつかのプロジェクトを実施しました。ICIJ の地域コーディネーターは、プラットフォームへの参加を容易にし、技術的な問題に対処する手助けをしています。これは、記者に必要とされるセキュリティ基準を満たすために重要です。

あらゆるものを暗号化する

396 人のジャーナリストとともに調査を行う場合、セキュリティを現実的に考えなければなりません。すべての個人が攻撃の標的となる可能性があり、[セキュリティを]突破されるリスクは高いのです。このリスクを軽減するために、ICIJ は複数の防御策を講じています。ICIJ の調査に参加する際には、電子メールを暗号化するための PGP キー・ペアを設定することが義務づけられています。PGP の原理は簡単です6。あなたは二つの鍵を所有します。一つは公開鍵で、暗号化された電子メールを送る通信相手になりうる人に伝えられます。二つ目の鍵は秘密鍵で、あなたのコンピュータから漏洩してはいけません。秘密鍵の役割はただひとつ、公開鍵で暗号化されたメールを復号化することです。PGP は、メッセージを保管できる金庫と考えてください。その金庫を開けてメッセージを読むための鍵を持っているのはあなただけです。あらゆるセキュリティ対策と同様、PGP にも脆弱性があります。例えば、スパイウェアがあなたのコンピュータ上で動作し、あなたが入力した言葉を記録したり、ディスク上のすべてのファイルを盗聴したりすれば、簡単に危機にさらされてしまいます。セキュリティを何重

にも積み重ねることが重要です。一つの層が破られても、他の層が侵入の影響を緩和してくれることが期待できます。

パートナーの身元を確認するために、ICIJ はすべてのプラットフォームに二要素認証を導入しています。この技術は、Google、Twitter、Facebook などの主要な Web サイトでよく使われています。これは、ログインに必要な二つ目の一時的なコードをユーザーに提供するもので、通常は（携帯電話などの）別の端末で生成され、すぐに消えてしまいます。一部の機密性の高いプラットフォームでは、三要素認証を追加することもあります。それがクライアント認証です。これは基本的には、記者がラップトップに保存し、設定する小さなファイルです。ICIJ のネットワークシステムは、この証明書を持っていないデバイスからのアクセスを拒否します。

ICIJ がセキュリティを改善するために使用しているもう一つの注目すべきメカニズムは、Ciphermail です。このソフトウェアは、私たちのプラットフォームと記者のメールボックスの間で動作し、ICIJ からの電子メールが確実に暗号化されます。

非構造化データへの対処

パラダイス文書は 1360 万件の文書のキャッシュでした。それらの文書を調査する上での主な課題の一つは、リーク元が様々であるということでした。Appleby、Asiaciti Trust、そして 19 カ国の法人登録文書です[7]。文書を詳しく見てみると、その内容や性格が多様であることや、電子メール、PDF、Word 文書など、構造化されたデータを分析するソフトウェアでは直接解析できない「非機械可読」形式の文書が多く存在することにすぐに気がつきます。これらの文書には、ICIJ が調査した二つのオフショア法律事務所の内部活動が反映されています。ICIJ のエンジニアは、記者がこれらの文書を検索できるよう、複雑で強力なフレームワークを構築しました。クラウドコンピューティングの拡張性を利用して、文書は暗号化されたディスクに保存され、「データ抽出パイプライン」に送られました。データ抽出パイプラインとは、文書からテキストを取り出し、検索エンジンが使用できるデータに変換する一連のソフトウェアシステムです。ファイルのほとんどは、PDF、画像、電子メール、請求書などで、簡単には検索できません。Apache Tika（メタデータとテキストの抽出）、Apache Solr（検索エンジンの構築）、Tesseract（画像のテキスト化）などの技術を用いて、チームは Extract というオープンソースのソフトウェアを構築し、これらの文書を検索可能で機械で読めるコンテンツにするという無二のミッションを達成しました。

このツールは、アクセス可能になったデータを最大 30 台のサーバーに分散させる際にとりわけ役立ちました。ICIJ は、「非構造化データ」から抽出された洗練された情報、つまりさまざまな情報源からのさまざまなタイプの文書の集合体をジャーナリストが探索できるよう、ユーザーインターフェースも構築しました。

ここでもまた、Blacklight というオープンソースのツールを利用することにしました。Blacklight は、ジャーナリストが文書を調べたり、高度な検索クエリ（近似文字列のマッチングなど）を使って、リークに潜む手がかりを特定したりできる、使いやすい Web ポータルを提供するものです9。

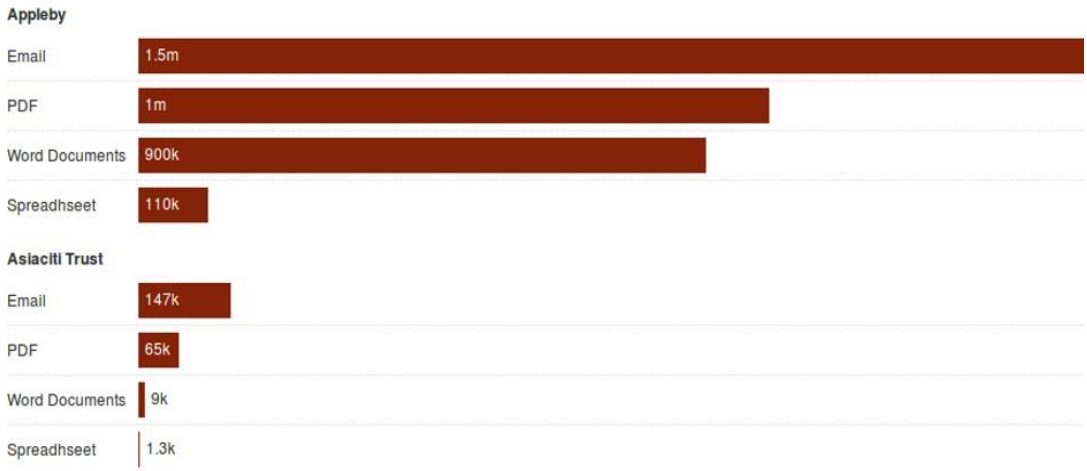


Figure 16.1. リークの内側

グラフを使って隠された宝石を見つけ出す

ICIJ は、2013 年に「オフショア・リークス」データベースの第 1 版を発表しました。このデータベースでは、グラフデータベースを用いて、10 万以上のオフショア企業と役員のつながりを調べることができます。このデータベースは、パナマ文書やパラダイス文書などの後続のリーク情報を含め、本稿執筆時点で 78 万 5000 以上のオフショア企業に拡大しています。ICIJ は、スイス・リークスで初めてグラフデータベースの利用を試みましたが、調査・報道の段階でグラフデータベースが重要な役割を果たすようになったのはパナマ文書からです。1150 万件の複雑な財務・法律記録、2.6 テラバイトのデータを調査するのは容易なことではありませんでした。ICIJ は、Neo4J や Linkurious などのネットワークグラフツールを使用することで、パートナーが人物とオフショア企業間のつながりを迅速に調査できるようにしました。私たちのデータチームとリサーチチームは、ファイルから情報を抽出し、構造化し、Linkurious でデータを検索できるようにしました。

例えば、アイスランドの当時の首相であるシグムンドゥル・グンロイグソンが Wintris という会社の株主であることを発見しました。この発見を可視化して保存し、世界の他の地域で調査に取り組んでいる同僚と共有することができました。その後、ドキュメントプラットフォームの Blacklight に戻って、より高度な検索を行い、Wintris に関連する記録を探ることができました。Blacklight は後にパラダイス文書のナレッジセンターに進化しました。データや文書を調査して得られた重要な知見は、「Global I-Hub」で共有され、現場取材で得られた知見も共有されました。

グラフデータベースとテクノロジーが、ICIJ のラディカル・シェアリングを支えています。「まるで魔法のようだ！」と何人ものパートナーが言っていました。データを調べるのに、コーディングのスキルは必要ありません。ICIJ は、調査やセキュリティのために我々の技術を使用するためのトレーニングを行いました。そして、380 人以上のジャーナリストが、何百万もの文書をマイニングし、グラフデータベースを使用し、バッチ検索を含む高度な検索を行い、調査結果や報道の結果だけでなく、クエリ戦略に関する有益なヒントを共有できるようになったのです。

パナマ文書プロジェクトでは、グラフデータベースや、ナレッジセンターや Global I-Hub などのアドホックな技術を用いて、25 カ国語で活動する約 80 カ国のジャーナリストをグローバルなバーチャルニュースルームでつなぎました。

大量の文書と結びついた構造化されたデータが「オフショア・リークス」のデータベースを通じて視聴者と共有されたことで、新たなジャーナリストが新たな手がかりを求め、「Alma Mater」や「West Africa Leaks」プロジェクトのような新たなコラボレーションに取り組むことができるようになりました。

また、市民や公的機関が独自に調査や研究に利用できるようにもなりました。19 年 4 月現在、パナマ文書の調査の結果、各国の政府は 12 億米ドル以上の罰金や追徴課税を課しています。16 年にパナマ文書が初めて公開されて以来、ICIJ の技術を利用するジャーナリストのグループは増え続け、500 人以上が金融関連の流出文書を調査し、これらの数百万件の記録に関連した公益性の高いストーリーを発表し続けています。

脚注

1. www.icij.org/investiga...

2. www.icij.org/investiga...

3. www.icij.org/investiga...

4. www.icij.org/journalist...

5. https://www.icij.org/blog/2014/07/icij-build-global-i-hub-new-secure-collaboration-tool/を参照のこと。I-Hub のようなジャーナリスティックなプラットフォームに関する別の視点については、本書のカンデアの章を参照。

6. www.gnupg.org/

7. www.icij.org/investigations/paradise-papers/...,
www.icij.org/investigations/%20paradise..., www.icij.org/investigations/paradise-papers/roll-roll-asiacitis-u-s-marketing-tour/

8.https://github.com/ICIJ/extrac...

9. https://github.com/projectblac...,
https://en.wikipedia.org/wiki/...

データとしてのテキスト：収集したテキストからストーリーを見つける

文：バーバラ・マセダ

概要

文書やスピーチのコレクションからデータストーリーを見つけるには。

キーワード：データジャーナリズム、非構造化データ、テキスト分析、テキストマイニング、計算論的ジャーナリズム

ここ数年のデータジャーナリズムの制作状況を見ると、非構造化データ（テキストなど）をベースにしたストーリーは、構造化データをベースにしたストーリーに比べてはるかに少ないことに気づくかもしれません。例えば、2012～16 年にデータジャーナリズム・アワードにノミネートされた 200 件以上の作品を分析したところ、応募作品は主に地理的データと金融データに依存しており、次いでセンサー、社会・人口統計、個人データ、メタデータ、投票など頻繁に使用されるタイプのソースであることがわかりました（Loosen et al, 2020）。つまりほとんどが構造化データです1。しかし、ニュースルームでは、ソーシャルメディアへの投稿、スピーチ、電子メール、長文の公式報告書などを扱う機会が増えており、これらのソースを処理・分析するための計算論的アプローチが重要になっています。

あなたもこの方法で作られたストーリーをご覧になったことがあるかもしれません。例えば、トランプ大統領のツイートを統計的にまとめたものや、アメリカの選挙で大統領候補者が公の場や討論会で取り上げた主なトピックを可視化したものなどがあります。テキストをデータとして扱うことは簡単なことではありません。ドキュメントは、フォーマット、レイアウト、内容が多様で、画一的なソリューションや、ある調査を別のドキュメントセットで再現しようとすると、手間がかかります。データのクリーニング、準備、分析は、文書のコレクションごとに大きく異なります。また、ニュース価値のある主張をしたり、研究者だけでなく広く一般の人々にとっても意味のある調査結果を発表したりする前に、さらに人の目で確認しなければならないステップもあります2。

ここでは、ジャーナリストがストーリーを伝えるためにテキスト分析を利用する五つの方法を、データジャーナリズムのさまざまな模範的プロジェクトを参照しながら検討します。

長さ：どれだけ書いたか、語ったか

文章や単語を数えることは、文書に対する最も単純な定量的アプローチです。昔からある作業で、ほとんどのワープロで簡単に実行できます。学生や記者で、字数制限のある課題を提出したことがある方なら、特別なデータトレーニングを受けなくても理解できるでしょう。単語数の問題点は、意味のあるベースラインに照らし

合わせて結果を解釈することです。このような尺度は、温度や速度のように広く知られているわけではないので、スピーチが 2000 語であるという事実から意味を導き出すのは、それほど簡単ではないかもしれません。実際には、比較のためのベースラインやリファレンスを自分たちで作成するしかない場合が多く、それがさらなる作業を要することもあります。場合によっては、調べようとしているイベントやスピーカーの歴史に文脈を見出すことも可能です。例えば、2016 年に行われたアメリカ大統領の一般教書演説を巡り、[ニュース解説メディア] Vox は全ての歴史的な演説の長さを計算して、「オバマ大統領は史上最も言葉数の多い一般教書演説者の一人である」と判断しました(Chang, 2016)。

複数の人が話すイベントでは、それぞれの人がいつ、どれだけ話したかを、全体の単語数と関連付けて調べることができます。Figure17.1 をご覧ください。

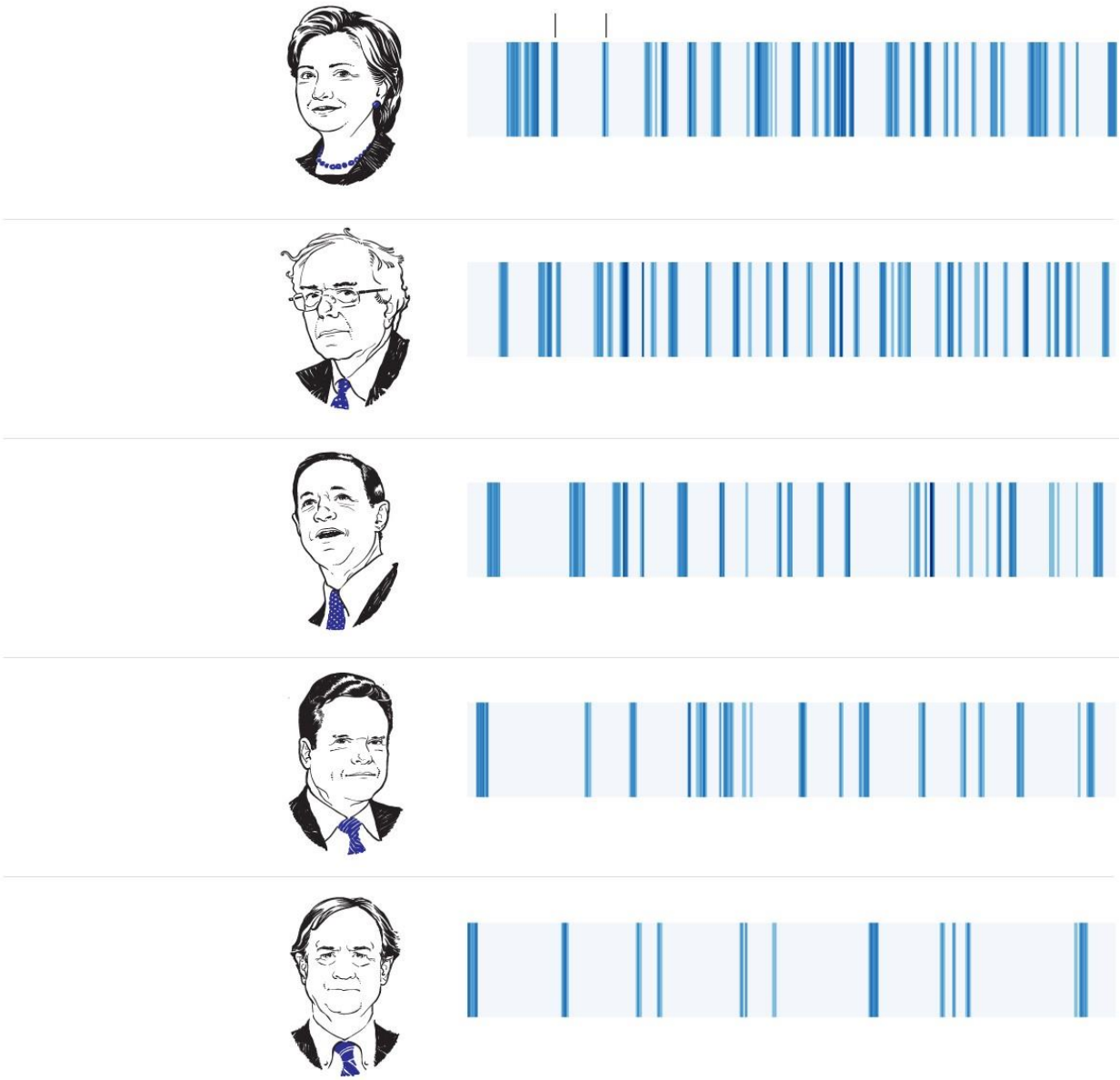


Figure 17.1. 民主党の討論会のビジュアル化（2015 年） 出典：ワシントン・ポスト https://www.washingtonpost.com/graphics/politics/2016-election/debates/oct-13-speakers/

メンション：誰が、いつ、何を、何回言ったか？

ある用語や概念が会話や文章の中で何回使われたかを数えることも、データの統計的な概要を把握するのに有効な単純作業です。そのためには、最も適切な要素を選択してカウントすることが重要です。データから探ろうとする課題に応じ、「ステミング」や「レンマ化」などの正規化操作を使用して、各単語の繰り返しを数えたり、共通の語幹を持つ単語の繰り返しを数えたりすることができます3。また単語出現頻度-逆文書頻度（TF-IDF）と呼ばれる重み付けを用いて、各文書の中で最も関連性の高い用語に焦点を当てるというアプローチもあります4。次のような例があります。

頻出する用語やトピック。ガーディアンは、2016 年のロンドン市長選挙で、2 人の候補者が選挙前の 6 年間に英国議会で様々な争点となる問題（犯罪、汚染、住宅、交通など）を取り上げた回数を分析しました（Barr, 2016）。分析対象となるトピックは、今回のようにあらかじめ決めておき、比較可能なテキストコレクションや類似のテキストコレクションに含まれる多数の関連キーワード（またはトピックに関連するキーワード群）から探索することができます。検索語は類似していても、必ずしも同じではないこともあります。例えば、同じメディアが 2017 年に発生した 3 つの異なるハリケーン（ハービー、イルマ、マリア）をどのように報道したかという FiveThirtyEight の分析を見てみましょう（Mehta, 2017）。もう一つ、トピック検出戦略として単にテキスト内の最も一般的な単語に注目するアプローチもあります。

スピーチの時系列分析。スピーチを時系列で見ることは、これまで言及されなかった、あるいは長い間取り上げられていなかったトピックを指摘する方法でもあります。例えば、ワシントン・ポストが18年の一般教書演説を報道する際に選択したアプローチは、どの大統領がどの言葉を最初に使ったかを強調した記事でした（Fischer-Baum et al.2018）。読者は、何百ページもの演説を読まなくても、トランプがウォルマート（2017 年）や[安全保障の]ただ乗り（2019 年）に言及した史上初の大統領であることをすぐ把握でき、テキストデータの要約と可視化がいかに効果的であるかを示しています。

省略。言及の数が少なかったり、全く言及がなかったりすることも、ニュースになることがあります。言及漏れは時系列で分析できますが、それに加えて、人や組織がある文脈で何かに言及するという期待に基づいて分析することもできます。16 年の米大統領選挙期間中、FiveThirtyEight は、ドナルド・トランプ候補のツイートに世論調査に関するキーワードの言及が比較的少ないことを発見し、世論調査に関するツイートをやめたと報じました（Mehta & Enten, 2016）。このような省略は、同じ発言者を長期的に監視することで発見することができます。今回のケースでは、FiveThirtyEight は数ヶ月前から、トランプが自分を勝者に見せる世論

調査をたくさんツイートしていることを発見していました（Bialik & Enten, 2015）。これは、上述したテキスト統計の文脈化の問題を解決する方法として、テキスト分析に基づいたニュース報道が、後にフォローアップ記事に役立つことを示す良い例でもあります。あるトピックの不在は、人や組織がある文脈でそのトピックについて言及するという期待に基づいて測定することもできるのです。

人、場所、名詞、動詞。自然言語処理（NLP）ツールは、（固有表現抽出、NER と呼ばれる）固有名詞や地名、企業名などの要素の抽出 、（品詞タグ付け（POS）と呼ばれる）名詞や形容詞などの単語の識別を可能にします。先ほどのワシントン・ポストの記事では、企業や宗教用語、動詞などに焦点を当てたフィルタリングを行い、可視化しています。

比較

複数の文書がどれだけ類似しているかを判断することが、さまざまな種類のストーリーの出発点になります。近似文字列マッチング（「ファジー・マッチング」とも呼ばれる）を利用して、盗作を暴いたり、公人の交友関係を明らかにしたり、ある法案がどのように変更されたかを説明したりすることができます。2012 年、プロパブリカは、選挙運動で有権者に送られた電子メールの変化を追跡するために、同じメッセージの連続したバージョンを並べて表示し、削除、挿入、および変更されていないテキストを可視化しました（Larson & Shaw, 2012）。

分類

テキストは、機械学習アルゴリズムを用いて、あらかじめ定義された特定の特徴に応じてカテゴリーに分類することができます。このプロセスは一般に、与えられた特徴に基づいて項目を分類するモデルを学習し、それを新しいデータの分類に使用する構成となっています。例えば、ロサンゼルス・タイムズは 15 年、公文書公開請求で入手した警察の 40 万件以上に上る報告書を分析し、市警が推定 1 万 4000 件の重大な暴行を軽犯罪に誤分類していたことを明らかにしました（Poston et al., 2015）。

記者は、データ量の少ない以前の調査で行ったように、MySQL を使って凶悪犯罪を示すキーワード（刺す、ナイフなど）を検索するのではなく、機械学習の分類器（SVM と MaxEnt）を使って 8 年分のデータを再分類し、1 年分を対象とした最初の調査の半分の時間で検証しました（Poston & Rubin, 2014）。この例は、機械学習アプローチがいかに時間を節約し、調査力を倍増させるかを示しています。

感触

多くのジャーナリストは、問題となっているテーマに対する発言者の態度に応じて、文章や文書をポジティブ、ネガティブ、ニュートラル（他の評価尺度も可能）に分類することの価値を認識しています。応用例としては、トピック、ハッシュタグ、Twitter ユーザーの投稿などを分析して問題に対するセンチメントを評価したり、プレスリリースや Web サイト上のユーザーのコメントに対して同様の計算を行ったりすることが挙げられます。例えば、エコノミストがヒラリー・クリントンとドナルド・トランプの党大会でのスピーチのトーンを比較しています

(「How Clinton's and Trump's Convention Speeches," 2016」)。これらの候補者と過去の候補者が使用した言葉の極性を分析し、トランプは「最近の記憶の中で最もネガティブな演説をした」、クリントンは「過去 40 年間で最も平静なスピーチの一つ」であることを示すことができたのです。

テキスト・マイナー・ジャーナリストを目指して

市販のテキストマイニングソフトウェアを使用すれば、基本的なテキスト分析操作とその結果(単語数、固有表現抽出、文書間のつながりなど)にスムーズに慣れることができます。DocumentCloud や Overview など、ジャーナリスト向けに設計されたプラットフォームには、これらの機能の一部が含まれています5。Google Cloud Natural Language API は、[感情を分析する]センチメント分析、[固有名詞に着目する]エンティティ分析、コンテンツ分類、構文分析など、さまざまなタスクを処理することができます6。テキストマイニングについて詳しく知りたい方には、Python(NLTK、spaCy、gensim、textblob、scikit-learn)や R(tm、tidytext など)のリソースを含む、よりパーソナライズされた分析を可能にする無料のオープンソースツールがあります。すでにこれらの言語に慣れ親しんでいるジャーナリストにとっては、その方が便利かもしれません。

また、正規表現や、テキストの収集(Web スクレーピング、API クエリ、情報公開請求)や処理(光学式文字認識(OCR)、ファイル形式の変換など)に必要なツールやテクニックを使いこなせることも必要です7。そしてもちろん、情報検索、関連モデルやアルゴリズム、テキストデータの可視化など、テキストデータ作業の背景にある理論や原理を把握することも有効です8。

結論

文書に関する新たな洞察を読者に提供できる可能性や、読むのに何ヶ月も何年もかかるような長いテキストを分析する能力を高めることができることは、ジャーナリズムの有用なツールとしてテキスト分析の開発を真剣に検討する十分な理由となります。曖昧さの問題(コンピュータは人間よりも言語の文脈を「理解」するのが難しいかもしれない)から、言語特有の問題(ドイツ語よりも英語の方が解決しやすい場合がある、ある言語では他の言語よりも多く取り組まれてきた)まで、課題はまだまだたくさんあります。ジャーナリストとしての私たちの仕事は、この分野の発展に貢献することができます。多くの報道プロジェクトは、利用可能な注釈付きデータセットの数を増やし、課題を特定し、新しいアプリケーションのアイデアを生み出すとも言えます。このアプローチで制作された最近の記事の数が増えていることからしても、テキストマイニングはデータジャーナリズムの中でも有望でエキサイティングな成長分野なのです。

脚注

1. データジャーナリズム・アワードについては、ルースンの章を参照。

2. データのクリーニングと準備には、次の手順の一つ以上が含まれる場合がある。テキストをユニットまたはトークンに分解する（「トークン化」と呼ばれる処理）、共通のファミリーまたはルートを共有する単語を「グループ化」する（ステミングおよびレンマ化）、ストップワードや句読点などの余分な要素を削除する、テキストの大文字と小文字を変更する、単語に焦点を合わせ、その順序を無視する（「バッグ・オブ・ワーズ」というモデル）、およびテキストをベクトル表現に変換することだ。

3. 語幹解析とレンマ化は、派生した単語を語根に還元する操作で、「reporter」「reporting」「reported」の出現をすべて「report」の下にカウントできるようにするもの。これらの操作は、アルゴリズムが単語の根を決定する方法において異なる。レンマ化とは異なり、語幹分析は、品詞を考慮せずに、単語の接尾辞を除去する。

4. TF-IDF は、コレクション内の単語の重みを理解するためにアルゴリズムで使用される尺度。TF-IDF Weight (w, d) = TermFreq(w, d) - log (N / DocFreq(w)), ここで TermFreq(w, d) は文書 (d) における単語の頻度、N はすべての文書の数、DocFreq(w) は単語 w を含む文書の数である (Feldman and Sanger, 2007).

5. www.documentcloud.org, www.overviewdocs.com

6. cloud.google.com/natural-language

7. regex.bastardsbook.com

8. さらに詳しい情報はダン・ジュラフスキーとジェームズ・H・マーティンの *Speech and Language Processing* やロナン・フェルドマンとジェームズ・サンガーの *The Text Mining Handbook* を参照。また、これらのトピックや関連するトピックに関する無料のオンラインコースも多数ある。

参考文献

Barr, C. (2016, May 3). London mayor: Commons speeches reveal candidates' differing issue focus. *The Guardian*. www.theguardian.com/politics/datablog/2016/may/03/london-mayor-data-indicates-candidates-differing-focus-onissues

Bialik, C., & Enten, H. (2015, December 15). Shocker: Trump tweets the polls that make him look most like a winner. *FiveThirtyEight*. vethirtyeight.com/features/shocker-trump-tweets-the-polls-that-make-him-look-most-like-awinner/

Chang, A. (2016, January 11). President Obama is among the wordiest State of the Union speakers ever. *Vox*. www.vox.com/2016/1/11/10736570/obama-wordy-state-of-the-union

Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.

Fischer-Baum, R., Mellnik, T., & Schaul, K. (2018, January 30). The words Trump used in his State of the Union that had never been used before. *The Washington Post*. www.washingtonpost.com/graphics/politics/presidentiallexicon-state-of-the-union/

How Clinton's and Trump's convention speeches compared to those of their predecessors. (2016, July 29). *The Economist*. www.economist.com/graphic-detail/2016/07/29/how-clintons-and-trumps-convention-speechescompared-to-those-of-their-predecessors

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing*. Pearson.

Larson, J., & Shaw, A. (2012, July 17). Message machine: Reverse engineering the 2012 campaign. *ProPublica*. projects.propublica.org/emails/

Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2020). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*, *21*(9), 1246–1263. https://doi. org/10.1177/1464884917735691

Mehta, D. (2017, September 28). The media really has neglected Puerto Rico. *FiveThirtyEight*. vethirtyeight.com/features/the-media-really-has-neglected-puerto-rico/

Mehta, D., & Enten, Ha. (2016, August 19). Trump isn't tweeting about the polls anymore. *FiveThirtyEight*. vethirtyeight.com/features/trump-isnt-tweeting-about-the-polls-anymore/

Poston, B., & Rubin, J. (2014, August 10). Times Investigation: LAPD misclassi ed nearly 1,200 violent crimes as minor offenses. *Los Angeles Times*. www.latimes.com/local/la-me-crimestats-lapd-20140810-story.html

Poston, B., Rubin, J., & Pesce, A. (2015, October 15). LAPD underreported serious assaults, skewing crime stats for 8 years. *Los Angeles Times*. www.latimes.com/local/cityhall/la-me-crime-stats-20151015-story.html

ニュースルームでのデータを使ったコーディング

文：バシール・サイモン

概要

ニュースルームは、コーダーや技術志向のジャーナリストにとって、ユニークな挑戦の場です。

キーワード： 計算論的ジャーナリズム、プログラミング、データクリーニング、データベース、データ・ビジュアライゼーション

データとコードが一緒になってしまうことは避けられません。データセットが大きすぎて Google Sheets［の処理］が遅くなったとき、Excel の数式が難解になったとき、何百行ものデータを理解するのが不可能になったときなどです。コーディングは、データをよりシンプルに、よりエレガントに、繰り返しを少なく、より再現可能にしてくれます。スプレッドシートを放棄するのではなく、スプレッドシートが数多くのツールの一つになることを意味しています。データジャーナリストは、必要に応じて様々なテクニックを駆使しています。例えば、Python notebooks でデータをスクレーピングし、その結果をスプレッドシートに投げ込んで、コピーして［Open］Refine でクリーニングし、［スプレッドシートに］再びペーストするのです。人によって学ぶプログラミング言語や技術は異なり、ニュースルームによっても制作する言語は異なります。例えば、ロンドンのタイムズでは、データ、ビジュアル、開発のほとんどが R、JavaScript、React で行われていますし、プロパブリカでは、Web アプリの多くに Ruby が使われています。ツールを選択するのは個人であることが多いものの、報道機関の慣習や文化がこれらの選択に大きく影響することもあります。たとえば、BBC はデータ可視化のワークフローを徐々に R に移行しており（BBC Data Journalism team, n.d.）、エコノミスト誌は世界的に有名なビッグマック指数を Excel ベースの計算から R と React/d3.js のダッシュボードに移行しました（González et al.2018）。多くの選択肢があり、唯一の正解はありません。

初心者にとって朗報なのは、多くの基本概念がすべてのプログラミング言語に当てはまるということです。Python でデータポイントをリストに格納する方法（スプレッドシートの行や列に格納するのと同じ）と、さまざまな操作を理解すれば、JavaScript、R、Ruby で同じことをするには、構文を学ぶだけでよいのです。本章では、データジャーナリズムのコーディングは、三つの中核的な領域に分割できると考えます。

データ作業–スクレーピング、クリーニング、統計（スプレッドシートでできる作業）

バックエンド作業–データベース、サーバー、API などの難解な世界

フロントエンド作業–インタラクティブなデータ・ビジュアライゼーションなど、Web ブラウザで行う作業のほとんど

ここでは、データジャーナリストがニュースルームでコードを扱う際に日常的に直面する、(a)学習時間、(b)納期、(c)コードレビューなどの制約によって、これらの異なる作業領域がどのように形成されるかを探ります。

学習時間

データジャーナリズムのコミュニティに通底する素晴らしい特徴の一つに、学習意欲があります。データジャーナリズムの基本を身につけようとする記者も、この分野での就職を目指している学生も、熟練した実務家も、学ぶべきことはたくさんあります。テクノロジーの進化は非常に速く、あるツールが廃れていく一方で、才能豊かで寛大な人々によって別のツールが作られることもあり、常に新しいことが行われ、学ぶことができます。

Twitter の API からデータを取得するためのライブラリのように、あるタスクのためのツールには、次から次へと[短期間での開発を繰り返す]イテレーションやバージョンが登場します。また、以前のツールをベースにして[機能を]拡張することもよくあります(例:D3 のデータ可視化ライブラリの拡張機能やアドオン)。このように、データジャーナリズムにおけるコーディングは、学習のための初期投資に加えて、時間とエネルギーを要する継続的な学習プロセスと言えるのです。プログラミングを学ぶ際に問題となるのは、慣れない概念に取り組むことで、最初のうちはスピードや効率が落ちてしまうことです。プログラミングのブートキャンプを利用すれば数週間で習得できますが、費用がかかります。カンファレンスで行われるワークショップは、初心者から上級者までを対象としており、短期間で安価に参加できます。仕事の一環として、時間をかけて学習する時間を持つことが必要です。そこでは実用的な問題に直面しますし、運が良ければ助けてくれる同僚もいるでしょう。問題を解決するにはコツがあります。何度でも質問を重ね、何が問題の原因なのかを見極める嗅覚を養うのです。

このような時間とリソースへの投資は報われます。コーディングは多くの新しい可能性と収入をもたらします[しかし、]どんな経験を積んでも、タスクにかかる時間を見積もるのが難しいという問題があります。ニュースルームでの仕事は締め切りに追われているので、これは難しいことです。

〆切を守る

ジャーナリズムの仕事では、〆切が大切です。コーディングは、報道のように予測不能なことが起きます。経験の有無にかかわらず、遅れが生じることがあり、というより必ず起こります。初心者にとっての課題の一つは、新しい作業方法を学ぶことによるスピードの低下です。何か新しいことを始めようとするとき、特に学習の初期段階では、使い慣れたツール(表計算ソフトなど)を使ってタスクを完了できるように、十分な時間を確保しておきましょう。学習を始めたばかりで時間がない場合は、使い慣れたツールを使うことにして、実験は時間ができてからにした方がよさそうです。テック企業では、大規模なプロジェクトに取り組む際に、プロジェクトをタ

スクとサブタスクに分解したり、タスクを（所要時間が予想できるようになるまで分割して）リストアップし、重要度に応じて優先順位をつけたりと、さまざまな方法を用いています。データジャーナリストもこの方法を見習いましょう。

英国の警察が解決できた犯罪の割合を調査したサンデー・タイムズのプロジェクトでは、読者の地域の数字を優先して表示しました。これが終わって時間に余裕ができたところで、リストの次の項目を実施しました。読者の地域と他の地域や全国平均を比較できるビジュアライゼーションです。この方法を取ることで、プロジェクトをいつでも公開できるようにしました。このような反復的なワークフローは、集中力と期待値の管理を同時に行うことができます。

コードレビュー

ニュースルームでは、記事の品質を維持するシステムを備えています。記者が記事を提出して［そのまま］印刷されるのではありません。編集者と副編集者が記事を精査するのです。ソフトウェア開発者は、品質を確保し、共同プロジェクトにバグを持ち込まないようにするための独自のシステムを持っています。プログラマーが自分の作品を提出し、他のプログラマーがそれをテストしてレビューする「コードレビュー」や、自動化されたコードテストなどです。

2017 年の Global Data Journalism Survey によると、回答したデータチームの 40％が 3～5 人のメンバーで（構成され）、30％が 1～2 人のメンバーでした（Heravi, 2017）。これだけ少人数だと、内部のコードレビューは困難です。データジャーナリストはこうして、同僚がなく、ピアレビューシステムがなく、あるいはコードをレビューする適切なスキルを持った人がいないために、一人で作業することが多いのです。

すると、内部の品質管理メカニズムは、一部のデータジャーナリズムチームにしかできない贅沢なものになってしまいます（コーディングのための副編集者なんてありえません！）。このような管理を行わない場合、バグが放置されたり、最適なパフォーマンスが得られなかったり、最悪の場合はエラーが見過ごされたりする危険性があります。リソースの制約があるため、多くのジャーナリストにとって、オンラインのコーディングコミュニティなど、自分の仕事に対する意見やコラボレーションを組織の外に求めることが重要なのです1。

脚注

1. データジャーナリズムのコードの透明性とレビューの実践に関する詳細はレオンとマゾッテの章を参照。

参考文献

BBC Data Journalism team. (n.d.). What software do the BBC use [Interview].warwick.ac.uk/fac/cross_fac/cim/news/bbc-r-interview/

González, M., Hensleigh, E., McLean, M., Segger, M., & Selby-Boothroyd, A. (2018, August 6). How we made the new Big Mac Index interactive. *Source*. https://source.opennews.org/articles/how-we-made-new-big-mac-indexinteractive/ Heravi, B. (2017, August 1). State of data journalism globally: First insights into the global data journalism survey. *Medium*. medium.com/ucd-ischool/state-of-data-journalism-globally-cb2f4696ad3d

データジャーナリズムの手法を説明する：スプレッドシート、スクリプト、プログラミング［統合開発環境の Jupyter］notebook

文：サム・レオン

概要

Jupyter Notebooks のようなリテラシーに富んだプログラミング環境が、データジャーナリズムの再現性、エラーの少なさ、協力性を高めるためにどのような役割を果たすかを探ります。

キーワード：Jupyter Notebooks、再現性、プログラミング、Python、文芸的プログラミング環境、データジャーナリズム

データジャーナリズムの台頭により、何をジャーナリズムのソースとみなすのかという考え方は変わりつつあります。ソースは現在、公開されているデータセット、流出した電子メール、スキャンした文書、衛星画像、センサーデータなど、さまざまな形で提供されています。同時に、これらのソースからストーリーを見つけるための新しい方法が登場しています。機械学習、テキスト分析、および本書の別の場所で検討されている他の技術のいくつかは、スクープのためにますます活用されています。しかし、データは、厳密で客観的な真実であるかのようなオーラとは裏腹に、歪められたり、誤って表現されたりすることがあります。データジャーナリストがデータセットの解釈を誤り、誤解を招くような記事を発表する事例には事欠きません。データ収集の時点で問題があり、より広い母集団に対して一般的な推論を行うことができないこともあります。例えば、サンプルの選び方に自己選択バイアスがかかっていることが考えられますが、これはインターネット上で世論調査や調査が行われる現代によく見られる問題です。また、データ処理の段階でエラーが発生することもあります。データ処理やクリーニングには、ジオコーディング、名前のスペルミスの修正、カテゴリーの整合化、または統計的な外れ値などを完全に除外することが含まれます。この種のエラーの良い例としては、政治的信念とポルノ消費との間に相関関係があるとする広く知られた研究における、IP アドレスの不正確なジオコーディングが挙げられます（Harris, 2014）。もちろん、データジャーナリストの仕事の本質である分析も［同様の事例が］あります。相関を因果関係と勘違いしたり、問題のデータセットを要約するために不適切な統計量を選択したりするなど、統計学的な誤りが（この部分に）影響を及ぼす可能性があります。データの収集、処理、分析がナラティブを変えうることを踏まえれば、データジャーナリストはどのようにして、使用したソースが信頼できるものであり、結論を導き出すために行われた作業に問題はないと読者に請け合うことができるのでしょうか？

データジャーナリストが単に第三者のデータや研究結果を報道しているだけなら、多くの大手ニュース・アウトレットが採用している伝統的な編集基準から逸脱する必要はありません。データを収集・分析した機関

への言及で十分です。例えば、英国の平均寿命に関する最近のフィナンシャル・タイムズのチャートには、「出典：EuroStat（EU 統計局）のデータに基づいて Club Vita が計算した」（と注釈が添えられています）。

そうすれば、読者は引用した機関の信頼性を評価できます。責任あるジャーナリストは信頼できると思われる研究のみを報道しますが、第三者機関はその結論に至った方法を説明する責任が大きくなります。学術的には、ピアレビューのプロセスが含まれる可能性が高く、科学的な出版物には、一定のレベルの透明性がある方法論（の裏付け）があるといえます。

[しかし、]ジャーナリスティックな組織がデータ駆動型の調査を行うケースがますます一般的になってきている中では、報道する[調査]結果の信頼性について読者に説明責任を負うことになります。ジャーナリストは、さまざまな方法で自分たちの手法を説明するという課題に対応してきました。一般的なアプローチの一つは、結論に至るために用いた一般的な方法論を説明することです。こうした説明は、できるだけ幅広い読者に理解してもらえるように、技術的ではない平易な言葉で表現されるべきです。

ガーディアンと[国際NGO]グローバル・ウィットネスが、環境保動家の死をめぐる調査の方法を説明したアプローチが良い例でしょう。(Leather, 2017; Leather & Kyte, 2017)。しかし、社会における生活を記述するあらゆる手法がそうであるように、文章で説明責任を果たすには限界があります。最も重大な問題は、分析の作成やデータの準備に使った正確な手順を明記していないことです。

したがって、結論を出すために記者が行った手順を正確に再現することは困難になり、場合によっては不可能になります。言い換えれば、文章による説明は再現性のあるものとは言いがたいのです。上記の例では、データの取得、処理、分析のステップが比較的簡単な場合は、文章による説明でも問題はないかもしれません。しかし、より複雑な技術が採用されている場合には、再現性のあるアプローチを採用する必要が出てきます。

再現性のあるデータジャーナリズム

再現性は、広く現代の科学的方法の根幹であるとみなされています。再現性は、結果を裏付けるプロセスを助け、問題のある発見や疑わしい理論を特定して対処するのに役立ちます。原理的には、ジャーナリスティックな文脈でのデータの誤った使用や誤解を招くような使用を排除するのにも、同じメカニズムが役立ちます。

最近の学術史の中で最もよく知られている方法論的な誤りの一つが、参考になります。ハーバード大学のカーメン・ラインハートとケネス・ロゴフは 2010 年の論文の中で、ある国の公的債務が国内総生産（GDP）の 90%以上に達すると、兵手金的な経済成長率は-0.1%に低下するとしていました（Reinhart & Rogoff, 2010）。この数字はその後、緊縮財政を支持する政治家たちが[主張を裏付ける]弾薬として使いました。

結論から言うと、回帰は Excel のエラーによるものでした。計算式に誤りがあったために、ラインハートとロゴフは、国の列全体の平均ではなく、20 カ国のうち 15 カ国しか集計していませんでした。すべての国を

考慮すると経済成長率は-0.1%から 2.2%の増加になりました。このミスは、大学院生のトーマス・ハーンドンとマイケル・アッシュ教授、ロバート・ポリン教授が、ラインハートとロゴフが作業したオリジナルのスプレッドシートを確認したときに判明しました。このことは、手法がわかりやすく書かれているだけでなく、分析に使用されたデータや技術そのものを入手することの重要性を示しています。しかし、ラインハルト＝ロゴフ[論文]のエラーは、他の何かを示唆しているのではないでしょうか。Excel は、データを扱うプロセスの大部分を設計上隠しています。スプレッドシートの分析を行う数式は、セルをクリックしたときにのみ表示されます。これは、結論に到達するまでの手順を確認することを困難にしています。確かなことはわかりませんが、ラインハートとロゴフの分析が、ステップを明示的に宣言しなければならない言語（プログラミング言語など）で行われていたら、公開前にエラーが発見されていたのではないでしょうか。

Excel ベースのワークフローでは、結論に至るまでのステップを削除することが推奨されます。式ではなく、値が他のシートや列にコピーされることが多く、数値がどのように生成されたかをたどる唯一のルートは「元に戻す」キーしかありません。「元に戻す」の履歴は、もちろんアプリケーションを閉じたときに消去されますので、重要な方法論的情報を保存するのには適していないのです。

文芸的プログラミング環境の台頭:ニュースルームにおける Jupyter notebook の活用

手法の透明性に対する新たなアプローチとして、いわゆる「文芸的プログラミング」環境を利用することが挙げられます。Buzzfeed、ニューヨーク・タイムズ、Correctiv のような組織は、分析のステップを正確に再現するため、機械が実行でき人間が読みやすいドキュメントを提供するこの環境を利用しています1。ドナルド・クヌースが 1980 年代に提唱した文芸的プログラミングはコンピュータコードを書くためのアプローチであり（Knuth, 1992）、コードの間に人間が日常的に用いる言語を挿入して、手順を説明します。今日使用されている二つの主要な文芸的プログラミング環境は、Jupyter Notebooks と R Markdown です2。ともに、HTML でレンダリングして Web 上で公開することができる単一のドキュメントに、平易な英語、ビジュアライゼーション、コードをミックスした、人間が読みやすいドキュメントを作成します。

元のデータは明示的にリンクすることができ、サードパーティのライブラリなど、その他の技術的な依存関係も明確に識別されます。人間が読みやすい説明を重視しているだけでなく、人間の論理を反映するようにコードが並んでいます。したがって、このパラダイムに基づき書かれたドキュメントは、議論のステップや、調査の一連の質問に対する回答のように読むことができます。文芸的プログラミングの実践者は、説明と優れた文体を主眼とするエッセイストです。彼らは作家のように、シソーラスを片手に、変数の名前を慎重に選び、それぞれの変数の意味合いを説明しようとします。書き手は、分かりやすいプログラムにしようと努力します。また、形式的な手法と非形式的な手法の両方を用いて、人間の理解に最も適した順序で概念を導入し、理解しやすいプログラムを目指します（Knuth, 1984）。この形式の典型は、バズフィードの Jupyter Notebook で、カリフォルニア州の山火事の傾向をどのように分析したかを詳細に説明しています3。Notebook には、分析を再現す

るために必要なすべてのコードとソースデータへのリンクが含まれていますが、ドキュメントの主眼は、ソースデータとのナラティブや対話にあります。説明は、問いに沿った見出しの下に設定され、重要なテーマを引き出すために、ビジュアライゼーションや図表を使用しています。プログラミングに対する「文芸的」アプローチのある側面として、（Jupyter Notebook や R Markdown ファイルとして）生成されたドキュメントが、コード自体を読むことができない人にも、結論を出すために取られた手順が健全であることを確認することができることが挙げられます。

このアイデアは、スティーヴン・シェイピンとサイモン・シャッファーの、近世科学における事実を立証する手段としての「仮想目撃者」の説明に似ています。ロバート・ボイルによる［空気ポンプの］実験を例に、シェイピンとシャッファーは「仮想目撃者」が担う役割を設定しました。

仮想目撃の技術は、直接的な目撃や複製を必要とせず、実験のイメージを読者の頭の中に生成するものである。仮想目撃（の手法）によって、原則的には目撃者を無数に増やすことができる。したがって、それは事実を立証する最も強力な手段になるのである。実験を検証し、その結果を事実として認めることは、必然的に心と心の目の実験室での実現を伴う。主張通りの方法で実験が行われたことを信頼し、保証する技術が必要とされた（Shapin & Schaffer, 1985）。

Jupyter Notebook のような文芸的プログラミング環境で作成されたドキュメントは、記事と一緒に公開されている場合、プログラミングに不慣れな読者が、発見を生み出すために行われた手順を可視化できるので、［仮想目撃と］同様の効果があると言えます。プログラミングをしない読者はコード自体を理解したり実行したりすることができないかもしれませんが、ドキュメント内のコメントや説明を読めば、エラーを軽減するために適切な手順が取られたことを理解できます。

例えば、イギリスの児童養護施設について最近 Buzzfeed が行った綿密な調査を見てみましょう4。Jupyter Notebook には、データが正しくフィルタリングされているかどうかをチェックするためのステップが備えられており（Figure 19.1）、ラインハートとロゴフを追い詰めた、単純ながら深刻なミスを防ぐ安全策を提供しています5。コードの正確な内容は、非技術者の読者には理解できないかもしれませんが、テストとエラーに関して適切かつ平易な英語の説明を伴う安全策が存在することは、ジャーナリストが調査結果を生み出すために取った手順が健全だと示すのにいくらか役立つでしょう。

```
In [11]: # Make sure that we've identified all private-sector owners
         assert (
             as_at_data_filtered
             .loc[lambda df: df["Sector"] == "Private"]
             ["Owner"].isnull()
             .sum()
         ) == 0
```

Figure 19.1. バズフィードの Jupyter notebook のセル。生データのフィルタリングが正しく行われたことを確認するコードであることを示す人間が読みやすい説明やコメントがある。 出典：ジェレミー・シンガー＝ヴァイン、バズフィード

再現性だけでなく

文芸的プログラミング環境を用いてデータストーリーを語ることは、再現性を高めるだけではありません。コードを公開することで、組織間のコラボレーションを促進できます。2016 年、[国際 NGO]グローバル・ウィットネスはパプアニューギニアの法人登記簿から企業とその株主を抽出する Web スクレーパーを公開しました6。この調査は、地域コミュニティに破壊的な影響を与える、腐敗しがちな熱帯木材取引の主な受益者を特定することを目的として行われました。グローバル・ウィットネスは、これを再利用する予定はありませんでしたが、コードを人気のコード共有サイト Github で公開しました。間もなく、地域支援団体の ACT NOW!はこのコードをスクレーパーからダウンロードして改良し、iPNG プロジェクトに組み込んで人々が会社の株主や取締役の名前を他の公的情報と照合できるようにしました 7。今やスクレーパーは、サイトのコアデータインフラの一部となっており、年2回、パプアニューギニアの企業登記簿からデータを取得しています。また、文芸的プログラミング環境でコードを書くことは、組織内の同僚が[記事の]公表前に分析を理解して確認する内部プロセスを合理化します。グローバル・ウィットネスでは、Jupyter Notebook が法的検証プロセスを合理化するために使用されています。Notebook には、知見を得るためのステップが論理的な順序で記述されているため、弁護士は[記事に関する]申し立てなどの法的リスクをより正確に評価できるようになります。

調査報道の観点から特に重要となり得るのは、データセットで参照される個人の身元を推定することです。私たちは最近、英国企業の透明性に関する調査の一環として、非常に多くの企業を支配している個人を明らかにしたいと考えました。これは[証拠はありませんが]その人がいわゆる「名義代理人」であることを示しており、企業に支配権を持つ個人（PSC）として登録されている場合などでは違法になることがあります。

最も多くの企業を支配する個人のリストを公開するとき、法務部は私たちがどのようにして「John Barry Smith」が「John B. Smith」と同一人物だと判断したのか示すよう求めました。Jupyter Notebook

で個人の身元を確かめるために使用した機能を手順ごとに表にまとめて提示することで、[法務部は]重複除外の実行方法を明確に把握することができたのです9。グローバル・ウィットネスでは、ファクトチェックにおいても同様のプロセスが採用されています。グローバル・ウィットネスによると、Jupyter Notebook はデータセットを長期にわたって監視する必要がある場合、特に有用です。例えば、グローバル・ウィットネスは 2018 年に、ロンドンの不動産市場における汚職リスクが 2 年間でどのように変化したか検証したいと考えていました10。彼らは、外国企業が所有する不動産の土地登記簿からスナップショットを取得し、2 年前に同じ目的で開発していた Notebook を再利用して公開しました11。こうして、最少のコストで[2年前と]同等の成果が得られました。

グローバル・ウィットネスはライセンス制限があって基礎となるソースデータを公開できなくても、その手法を公開することができたのです。これは、スプレッドシートベースのワークフローでは非常に難しいことです。もちろん、あなたの手法を説明する最も効果的な方法は、使用した生データを公開することです。しかし、ジャーナリストは、著作権やプライバシー、情報源の保護といった理由で再公開できないデータを使用することが多いのです。

文芸的プログラミング環境は、いろいろな利点とともに、ジャーナリストのデータに関わる作業の説明責任と再現性を相当向上させることができますが、いくつかの留意すべき限界があります。その一つは、Jupyter Notebook や R Markdown ドキュメントによるアプローチを再生産する(ただ追従するのではなく、「仮想目撃者」となる)ためには、コードの書き方を知っているか、少なくとも実行する方法を知っている必要があるということです。データジャーナリズムは比較的初期段階にありコードを書けるジャーナリストは少数派です。これはつまり、新聞社の Github は、コミュニティの大半の人がコードを検証できる、学術誌でピアレビューされたコードと同水準の精査を受ける可能性が低いということです。したがって、データジャーナリズムは、より技術的な知識を持つ読者を対象とした学術研究と比較すると、コードに潜むエラーが発生しやすくなります。ジェフ・ハリスが指摘しているように、従来の事実誤認と同じように、メディアがプログラミングの修正を公表するようになるのは、そう遠くないでしょう。Workbench (ジョナサン・ストレイの章でも言及されています)のようなコードを書いたり理解したりする必要なしに、リテラシーの高いプログラミング環境の機能の一部を提供するツールがジャーナリスト向けに開発され始めていることは注目に値します12。ジャーナリズムにおける説明責任のための新しいメカニズムは、既存の「公共」が方法を精査する新しい手段であるだけでなく、新しいタイプの「公共」の形成にも役割を果たしていると考えられます。

アンドリュー・バリーのエッセイ「政治的手段としての透明性」にはこう書かれています。

透明性とは、情報公開だけではなく、公開された情報の価値を認識し、評価し、必要に応じて修正することのできる社会の形成を意味する。透明性の実現は、現地の目撃者に依拠しているが、これらの目撃者は適切に集められ、その存在が検証されることが期待される。このように、政治と石油(に依存した)経済との間には

循環的な関係があり、一方が他方を実現する。透明性とは、情報を公開することだけではなく、情報を得ようとする国民を形成することを目的としたものなのだ（Barry,2010）。

ここまで述べてきた、データジャーナリスティックな作業を説明する方法は、テクノロジーに理解のある市民の新しいグループが、文芸的プログラミング環境のような技術が出現し、使用される以前には不可能だった方法で、記者を精査し、説明責任を負わせることを可能にしたのです。この考えは、グローバル・ウィットネスが取り組む資源採掘業界の説明責任を強化するためのデータリテラシー活動に通じるものです。欧州連合（EU）の画期的な法律では、採掘企業に石油、ガス、鉱業のプロジェクトに対するプロジェクトレベルの支払いを政府に報告することを義務づけています。これは汚職の危険性が高い分野で、どこに収益が蓄積されているのかを詳しく調査する必要があるのです。

グローバル・ウィットネスや「Publish What You Pay」キャンペーンに参加する他の団体は長年、このような［監視の］役割を果たすことができる既存の公共機能が存在しないことを指摘してきました。これらの団体は資源保有国のジャーナリストや市民グループを結集するためのリソースとトレーニング・プログラムを開発し、このデータを使用して企業の責任をより容易に追究するスキルを開発するための支援を行いました。この取り組みの一環として、汚職の可能性のある不審な支払報告書に警告を発するための具体的な方法論が開発され、公表されています13。

文芸的プログラミング環境は、データジャーナリストがその手法をより透明性の高いものにし、説明責任を果たすための手段となっています。データは常に複数の解釈が可能な状態にあるので、記者の仮定を明確にし、その方法を再現可能にする技術は貴重です。これらの技術は、共同作業を後押しし、ますます技術的になっていく学問を様々な人々の精査に開放するものです。現在叫ばれているジャーナリズムの信頼の危機を考えると、再現性のあるアプローチを広く取り入れることは、データチームがその信頼性を維持するための重要な方法の一つとなるかもしれません。

脚注

1. github.com/TheUpshot
2. jupyter.org, rmarkdown.rstudio.com
3. buzzfeednews.github.io/2018-07-wild re-trends
4. buzzfeednews.github.io/2018-07-wild re-trends/
5. github.com/BuzzFeedNews/2018-07-ofsted-inspections/blob/master/notebooks/00-analyze-ofsted-data.ipynb

6. github.com/Global-Witness/papua-new-guinea-ipa

7. pngiportal.org

8. www.globalwitness.org/en/campaigns/corruption-and-money-laundering/anonymous-companyowners/companies-we-keep/

9. github.com/Global-Witness/the-companies-we-keep-public

10. www.globalwitness.org/en/blog/two-years-still-dark-about-86000-anonymously-owned-uk-homes/

11. github.com/Global-Witness/overseas-companies-land-ownershippublic/blob/master/overseas_companies_land_ownership_analysis.ipynb

12. workbenchdata.com

13. www.globalwitness.org/en/campaigns/oil-gas-and-mining/ nding-missing-millions

参考文献

Barry, A. (2010). Transparency as a political device. In M. Akrich, Y. Barthe, F. Muniesa, & P. Mustar (Eds.),

Débordements: Mélanges offerts à Michel Callon (pp. 21–39). Presses des Mines. http://books.openedition.org/pressesmines/721

Harris, J. (2013, September 19). *The Times* regrets the programmer error. *Source*.https://source.opennews.org/ar...

Harris, J. (2014, May 22). Distrust your data. *Source*. https://source.opennews.org/articles/distrust-your-data/ Knuth, D. E. (1984). Literate programming. *The Computer Journal*, *27*(2), pp. 97–111. https://doi.org/10.1093/comjnl...
Knuth, D. E. (1992). *Literate programming*. Center for the Study of Language and Information.

Leather, B. (2017, July 13). Environmental defenders: Who are they and how do we decide if they have died in defence of their environment? *The Guardian*. https://www.theguardian.com/environment/2017/jul/13/environmental-defenders- who-are-they-and-how-do-we-decideif-they-have-died-in-defence-of-their- environment

Leather, B., & Kyte, B. (2017, July 13). Defenders: Methodology. Global Witness. https://www.globalwitness.org/en/campaigns/environmental-activists/defendersmethodology/

Reinhart, C. M., & Rogoff, K. S. (2010). *Growth in a time of debt* (Working Paper No. 15639). National Bureau of Economic Research. https://doi.org/10.3386/w15639

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton University Press.

データジャーナリズムにおけるオープンな働き方

文：ナタリア・マゾッテ

概要

データジャーナリストがよりオープンに活動することで得られる事例やメリット、そして実践するためのいくつかの方法を検討します。

キーワード：データジャーナリズム、オープンソース、フリーソフトウェア、透明性、信頼、プログラミング

Linux、Android、Wikipedia、WordPress、TensorFlow など、多くの著名なソフトウェアや Web プロジェクトは、自由な知識の流れに基づいて共同開発されています1。GNU プロジェクトやフリーソフトウェア財団を設立した著名なハッカーであるストールマン（2002）は、1971 年に MIT で働き始めたとき、ソフトウェアのソースコードを共有することはレシピを交換するのと同じくらい一般的なことだったと述べています。

このようなオープンなアプローチは、長年、ジャーナリズムでは考えられませんでした。ジャーナリストとしてのキャリアの初期に、私はブラジルのオープンソース・コミュニティと仕事をし、ジャーナリズムにとって透明性は唯一の実行可能な道であると考えるようになりました。しかし、ジャーナリストやメディア組織にとって、透明性は優先事項でもコアバリューでもありませんでした。ジャーナリズムは、その近代的な歴史のほとんどにおいて、希少な情報をめぐる競争というパラダイムの中で行われてきたのです。

情報へのアクセスが少数の人の特権であり、重要な発見が目撃者や内部の人にしか得られない場合、説明責任を果たす方法は限られています。文書を引用したり、インタビューの情報源に言及したりするのに、そのような精巧な透明性メカニズムは必要ないかもしれません。場合によっては、秘密を守ることが情報源の安全を確保することにつながり、それが望ましいこともあります。しかし、情報が豊富な場合、「どうやってそこにたどり着いたのか」を共有しないと、読者がストーリーを理解し、納得するための手段を奪うことになりかねません。データやアルゴリズムに依拠した報道を行うジャーナリストなら、オープンソース・コミュニティと同様の倫理観を持つようになるのでしょうか？コミュニティに関連する新しいデジタル手法や価値観を採用するジャーナリストには、どのような利点があるのでしょうか。本章では、データジャーナリストがよりオープンに活動することで得られる事例やメリット、そしてそのための方法を検討します2。

オープン化の事例とメリット

ワシントン・ポストは、数百万種類の鎮痛剤の販売に関するデータベースを掘り下げて、米国におけるオピオイド処方の流行を、前例のない規模で調査し、データセットとその［調査］方法を公開しました3。その結果、30 以上の州の地元記者が、それぞれの地域におけるこの危機の影響について 90 以上の記事を掲載するこ

とができました（Sánchez Díez, 2019）4。［また、］2 人の計算に長けたジャーナリストが Uber の［需要集中期に料金を値上げする］サージ・プライシング・アルゴリズムを分析した結果、同社は白人が多い地域でより良いサービスを提供しているようだと明らかにしました（Stark & Diakopoulos, 2016）。この記事はワシントン・ポストに掲載されましたが、使用されたデータ収集と分析コードは、開発者がコードを保存・管理するためのオンライン・プラットフォーム GitHub で自由に利用できるようになっていました5。このため、データベースのエラーに気付いた読者は、記事の筆者にその旨を報告することができ、筆者はそのバグを修正して記事を訂正することができたのです。私が共同設立したブラジルのデジタルマガジン「Gênero e Número（ジェンダーと数）」は、ブラジルの公共空間から女性が排除されていることを示そうと、80 万以上の通りの名前を分類するプロジェクトを実行しました。Python スクリプトを実行して、通りの名前とブラジル国家統計局の名前のデータベースを相互参照しました（Mazotte&Justen、2017）。同じスクリプトはその後、他の取り組みでも使用され、選挙の候補者や判事のリストなど、ジェンダー情報が含まれていないデータセットを分類しました（Justen, 2019）。

様々なデータセット、ツール、コード、メソッド、プロセスを透明化して利用できるようにすることは、データジャーナリストにとって様々な意味で有益です。まず、仕事の質を向上させることができます。プロセスを文書化することで、ジャーナリストはより整理され、より正確になり、エラーを見逃すことが少なくなります。また、複雑なストーリーの編集やレビューの負担を軽減するとともに、読者が問題を報告できるようになります。

第二に、到達範囲と影響力を広げることができます。蓄積されたストーリーは、さまざまな視点を獲得し、多彩なコミュニティに貢献することができます。プロジェクトは、制作者の当初の範囲や制約にとらわれることなく、独自の道を歩むことができます。

そして三つ目は、ジャーナリストや一般の人々のデータリテラシーを高めることができます。自分の仕事を段階的に説明することで、他の人がそれを追って学ぶことができ、データの生態系、実践、コミュニティを豊かにし、多様化させることができるのです。

いわゆる「ポスト真実」の時代には、メディアに対する国民の信頼を高める可能性もあります6。2018 年のエデルマン・トラスト・バロメーターによると、メディアへの信頼度は最低レベルに達しています。オープンに仕事をすることは、この傾向を減速させ、さらには逆転させることにもつながります。これには、ジャーナリストが自分の偏見や不確実性を正直に伝え、読者との対話を可能にするために、結論に至る方法をよりオープンに話し、より詳細な「ハウツー」を提供することが含まれます7。

しかしながら、データジャーナリズムにおけるオープンで透明性の高い作業の実践と文化は、現在進行形で探求と実験が行われていることに注意すべきでしょう。潜在的なメリットについて理解を深めたとしても、透明性が価値あるものであるか、優先度が低いものであるか、あるいは有害なものであるかを理解する配慮が必要で

す。例えば、パナマ文書の場合のように、調査自体の完全性を守るために、データや技術を秘密にしておくことが重要な場合があります。

オープンな仕事の進め方

そうした障害がない場合（これはケースバイケースで分析する必要があります）、透明性を確保するための一般的なアプローチとして、「ナードボックス」と呼ばれる方法論のセクションがあります。この方法論は、プロセスの複雑さや対象となる読者に応じて、さまざまな形式や長さのものがあります。より多くの人に読んでもらうことを目的としている場合は、記事中のボックスや、方法を簡潔に説明した脚注でも十分でしょう。

どのように報道したかを説明する独立した記事を選択する場合もあります。いずれにしても、専門用語を避け、データの入手方法や使用方法を説明し、重要な注意点を読者が見落とさないようにし、どのようにして結論に至ったのかを明確かつ直接的に説明することが重要です。

FiveThirtyEight、プロパブリカ、ニューヨーク・タイムズ、ロサンゼルス・タイムズといった、データジャーナリズムの分野で有名なメディアの多くは、GitHub などのコード共有プラットフォームにリポジトリを持っています。バズフィードのチームは、あらゆるオープンソースのデータ、分析、ライブラリ、ツール、ガイドなどをまとめたインデックスをも整備しています8。このようにして、データの再現性を高めるとともに（レオンの章で詳しく説明しています）、興味を持った読者が自らデータを調査できるようにしています。科学者たちが何世紀にもわたって行ってきたように、ジャーナリストたちも自分たちの仕事を同僚たちにチェックしてもらい、文書化された手順に従って同じ結論に達することができるかどうかを確かめるよう呼びかけているのです。

多くの報道機関にとって、このようなレベルの文書化やコラボレーションを仕事に取り入れることは簡単ではありません。少ないリソースとチームの縮小に直面する報道機関［の現状］は、自分の調査を記録したいと思うジャーナリストの熱意を削いでしまうかもしれません。ここで、ジャーナリストが直面している制約について考えてみましょう。多くの報道機関は、世界における役割やビジネスモデルの変化に伴い、生き残りをかけて戦っています。これらの課題にもかかわらず、自由でオープンソースなコミュニティのプラクティスを取り入れることは、イノベーションの指標として、また、ますます複雑化し、急速に変化する世界の聴衆との信頼と関係を構築する手段として、際立つ道を示しています。

脚注

1. この章はナタリア・マゾッテが著述し、マルコ・トゥーリオ・ピレスが協力した。

2. データジャーナリズムとオープンソースについては、本書のレオン、バーク、ピッツとムスカートによる章も参照されたい。

3. www.washingtonpost.com/national/2019/07/20/opioid- les/

4. www.washingtonpost.com/national/2019/08/12/post-released-deas-data-pain-pills-heres-what-local-journalistsare-using-it

5. github.com/comp-journalism/2016-03-wapo-uber

6. www.edelman.com/trust/2018-trust-barometer; www.edelman.com/sites/g/les/aatuss191/ les/2018-10/2018_Edelman_Trust_Barometer_Global_Report_FEB.pdf

7. データジャーナリズムにおける不確実性の問題については、アンダーソンの章を参照されたい。

8. github.com/BuzzFeedNews/everything#guides

参考文献

Justen, A. (2019, May 31). Classif icando nomes por gênero usando da- dos públicos | Brasil.IO—Blog. *Brasil.IO*. blog.brasil.io/2019/05/31/classi cando-nomes-por-genero-usando-dados-publicos/index.html

Mazotte, N., & Justen, A. (2017, April 5). Como classi camos mais de 800 mil lo-gradouros brasileiros por gênero. *Gênero e Número. www.generonumero.media/como-classi camos-mais-de-800-mil-logradouros-brasileiros-porgenero/*

Sánchez Díez, M. (2019, November 26). *The Post* released the DEA's data on pain pills. Here's what local journalists are using it for. *The Washington Post*. https://www.washingtonpost.com/national/2019/08/12/post-released-deasdata-pain-pills-heres-what-local-journalists-are-using-it/

Stallman, R. M. (2002). *Free software, free society: Selected essays of Richard M. Stallman* (J. Gay, Ed.). GNU Press.

Stark, J., & Diakopoulos, N. (2016, March 10). Uber seems to offer better service in areas with more White people.

That raises some tough questions. *The Washington Post*. www.washingtonpost.com/news/wonk/wp/2016/03/10/uber-seems-to-offer-better-service-in-areas-with-morewhite-people-that-raises-some-tough-questions/

報道に役立つアルゴリズムの構築

文：ジョナサン・ストレイ

概要

高度なデータ分析アルゴリズムは、調査報道に大いに役立ちますが、その作業のほとんどはデータの取得とクリーニングです。

キーワード：アルゴリズム、機械学習、計算論的ジャーナリズム、データジャーナリズム、調査報道、データクリーニング

計算論的ジャーナリズムの不都合な真実は、ストーリーの「アルゴリズム」は、すべての時間と労力を投じる部分ではないということです。どうか誤解しないでください。洗練されたアルゴリズムは、報道、特に調査報道において非常に有効です。機械学習（コンピュータにパターンを発見させる技術）は、膨大なデータの中から重要な文書を見つけ出すのに使われています。また、自然言語処理（言語を理解するためのトレーニング）により、文書から人名や企業名を抽出することができ、誰がストーリーに関わっているかを理解する近道となります。さらに、記者は不正や偏見を発見するために、さまざまな統計的分析を行っています。

しかし、アルゴリズムを実行するのは簡単です。データを入手し、それをクリーニングし、アルゴリズムのリードをフォローすることこそ大変なのです。これを説明するために、調査報道における機械学習の成功例として、The Atlanta Journal-Constitution の医師による性的虐待に関する驚くべき記事「License to Betray」を取り上げましょう（Teegardin et al,2016）。記者たちは、米国の全州の 10 万件以上の医師の懲戒記録を分析し、患者に性的虐待を加えた医師が診療を続けることを許可されていたケースを 2400 件発見しました。すべての報告書を読むのではなく、機械学習を応用して、性的虐待に関わる可能性の高い報告書を見つけることで、この山を大幅に減らしたのです。その結果、10分の1以下の 6000 件の文書にまで減らすことができ、それを手作業で読み、検証することができました。

記者のジェフ・エルンストハウゼンによると、機械学習がなければ全国的な記事にはならなかったといいます。「もしかしたら、地方版の記事にしていた可能性もあるかもしれない」と後に語っています（Diakopoulos, 2019）。これは、ジャーナリズムにおけるアルゴリズムの比類なき成功例であり、この手法はもっと広く利用できるでしょう。機械学習自体は難しいことではありません。エルンストハウゼンが使用した「ロジスティック回帰」という手法は、どの単語が含まれているかによって文書を分類する標準的な統計的アプローチです。Python では数十行で実装することができ、オンラインには多くのチュートリアルがあります。

ほとんどのストーリーでは、作業の大半がセットアップとその結果を利用することを占めます。データのスクレーピング、クリーニング、フォーマット、ロード、チェック、修正などの作業をたゆまず続けなければなりません。

アルゴリズム分析の結果は、多くの場合、手がかりやヒントに過ぎず、分析ツールではなくコラボレーションツールを必要とする記者チームによる大量の手作業による取材を経て、初めて記事になります。このような作業は、データの仕事の中でも特に地味なものなので、あまり教えられませんし、話題にもなりません。しかし、データ駆動型の記事を書くためには、この準備とフォローアップにほとんどの時間と労力を費やすことになります。

『License to Betray』では、データを入手するだけでも大変な苦労がありました。医師の懲戒に関する全国的なデータベースは存在せず、州レベルのデータベースがいくつかあっただけです。これらのデータベースの多くには、医師が懲戒処分を受けた理由を示す項目がありません。記入欄があっても、性的虐待が確実にコード化されていないことが多いのです。チームはまず、情報公開請求で報告書を入手しようとしました。しかし、州によっては数千ドルを要求してくるところもあり、（累計では）法外に高価なものとなりました。そこでチームは、州の医療委員会の Web サイトから文書をスクレーピングすることにしました（Ernsthausen, 2017）。

これらの文書を OCR（テキスト化）し、共同でタグ付けやレビューを行うためのカスタム Web ベースのアプリケーションに読み込む必要がありました。学習データを作成するために、数百の文書に手作業でタグ付けをしなければなりません。機械学習で残りの 10 万件をランク付けした後、性的虐待について予測された 6000 件の文書と、手作業で選んだキーワードを含む数千件の文書を読むのにさらに数カ月を要しました。さらに、ストーリーを具体化するために、何百もの具体的な事件を調査するなど、その他の取材も必要でした。過去のニュース記事や、関係者への個人的なインタビューなど、他の情報源も利用しました。

機械学習というアルゴリズムの使用は、調査の重要な部分を占めています。しかし、それは（全体に占める）時間と労力のごく一部にすぎません。データサイエンティストを対象とした調査では、仕事のほとんどがデータの「整理」と「クリーニング」であることが一貫して示されており、その割合は 80％にも上ります。ジャーナリズムも同じです。アルゴリズムは、ある種の魔法のようなものだと思われています。一見、複雑で不透明に見えるかもしれませんが、紛れもない力を持っています。

この魔法は、データを準備したり、長いリストをフォローアップしたりするありふれた仕事よりも、話をしていてずっと楽しいものです。技術者は、自分たちの技術を誇示したいのであって、その周辺で行われている同様に重要な仕事には興味がありません。新しく洗練されたツールに対するこのような偏見は、時としてジャーナリズムにも影響を及ぼします。しかし、私たちの第一の責任はジャーナリズムを完成させることであり、それには他のデータパイプラインにも取り組む必要があるのです。

一般的に、私たちはデータの準備に使われるツールを過小評価しています。OpenRefine は、あらゆるクリーニング作業の長年のヒーローです。Dedupe.io は、機械学習を応用してデータベース内の重複した名前をマージします。正規表現のような古典的なテキスト処理方法は、データジャーナリストの教育に含まれる

べきです。私が目下取り組むプロジェクト「Workbench」は、時間がかかるがほとんど表に出ない、報道用のデータを準備する作業、つまり「アルゴリズム」の前に行われるすべての作業に焦点を当てています。このプロジェクトは、記者が大規模なデータプロジェクトに共同で取り組み、機械も含めてお互いの仕事から学ぶことができるように、プロセス全体をより協力的にすることを目的としています。アルゴリズムは報道にとって重要ですが、それを機能させるためには、データ駆動型ジャーナリズムの他のすべての部分について話し合わなければなりません。目立って華やかでハイテクな部分だけでなく、ワークフロー全体の力を引き出す必要があるのです。

参考文献

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Ernsthausen, J. (2017). Doctors and sex abuse. NICAR 2017, Jacksonville. docs.google.com/presentation/d/1keGeDk_wpBPQgUOOhbRarPPFbyCculTObGLeAhOMmEM/edit#slide=id.p

Lohr, S. (2014, August 17). For big-data scientists, “janitor work” is key hurdle to insights. *The New York Times*.

www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html

Teegardin, C., Robbins, D., Ernsthausen, J., & Hart, A. (2016, July 5). License to betray. *The Atlanta Journal-*

Constitution, Doctors & Sex Abuse. doctors.ajc.com/doctors_sex_abuse

機械とジャーナリズムの協働とは？計算論的思考から分散型認知へ

文：エディ・ボルヘス＝レイ

あなたが、それほど遠くない未来のジャーナリストだとしましょう。あなたは記事を書くために、会話型エージェント（愛称：Twiki）に匿名化された 15 以上のデータベースをつなぎ合わせるように依頼します。融合されたデータセットの大きさと複雑さを考えると、可視化ソフトウェアは初歩的すぎて、探している異常を分離することができません。そこで、脳に埋め込まれたインプラントを使ってシステムに接続し、抽象化されたデータセットを簡単に操作します。匿名化されたそれぞれのデータセットは、身元や個人情報を保護するために有効ですが、それらを組み合わせると、著名人の身元や個人情報を推測することができるのです。

名前とそれに付随するデータを公開することが法的課題になる可能性があることを認識したあなたは、この情報を公開することが倫理的または法的課題になるかどうかを判断するためにニューラルネットワークを実行するよう Twiki に依頼しました。このネットワークでは、倫理規定や規制の枠組みに基づいて意思決定を行うバーチャルジャーナリストのシミュレーションを「n＋」回実行します。このシミュレーションがバックグラウンドで実行されている間に、あなたはいくつかの異常値を分離し、いくつかの興味深い傾向を特定することができました。この異常値が単なるエラーではなく、ストーリーに役立つことを確認したいので、Twiki に過去の記録を調べてもらい、異常値が歴史的な出来事と一致しているかどうかを確認します。さらに Twiki には、特定された傾向が今後も続き、憂慮すべき事態を招く可能性を計算するための予測モデルを実行してもらいます。

この短いフィクションは、数年前にタイムズの元データジャーナリストであるニコラ・ヒューズと交わした興味深い会話に基づいています。フィリップ・K・ディックの「マイノリティ・リポート」に出てきそうなシーンですが、実際には、すでに利用可能で広く使われている、あるいは急速に開発されているさまざまなツールやテクニックが登場しています。さらに重要なことは、ジャーナリストがデータを扱う機会が増え、コンピュータが不可欠になっている中で、ニュースルームで生まれつつあるジャーナリズムのワークフローとプロとしての姿勢にも言及していることです。このような最近の変化は、大きな技術革新がニュース制作のワークフローに導入されるたびに、報道そのものが破壊され、結果的に変容してきただけでなく、ジャーナリストの思考プロセスや職業上の理想も必ず修正されてきた歴史を反映しています。ビッグデータの時代から人工知能（AI）と自動化の時代へと移行している今日、［特定の形態や手法によってコンピュータを使用する］コンピューティングやデータサイエンスに由来する原理や作業方法がジャーナリズムに浸透してきています。コロンビア大学の Tow Center for Digital Journalism の創設者であるエミリー・ベルはこう言います。

企業であれ公共団体であれ、あらゆる分野のあらゆる企業や組織は、20 年前に Web 技術を軸に方向性を考えなければならなかったのとまったく同じように、AI を中心に方向性を再考する必要がある。（Bell, personal communication, September 7, 2017）

このような背景から、データや自動化されたプロセスと密接に関わるジャーナリストがコンピュータにまつわる技能を習得し、ジャーナリストとしての能力を高めつつ、他方では仕事のアプローチや理想の根幹を変え始めていることを考察します。ここでは、計算機が普及した環境でのジャーナリズムの認知を構想する枠組みとして、さまざまな理論的概念を検討しています。「拡張された認知」という概念を採用し、ジャーナリズムの認知が、今日では報道に使用される機械に依存している（つまり、分散している）方法について、さらなる議論を促します。この議論を通じて、ジャーナリズムの分散型認知の概念を検証し、さらに明確にするための実証的な研究を含め、ジャーナリスティックな状況における計算の役割を調査する将来の研究を呼び起こしたいと思います。

このような研究は、データ化や自動化がニュース制作のいたるところで行われるようになった場合にジャーナリズムが経験するであろう変化を認識し、それに対処したいと考えているプロのジャーナリストにとって、特に有益なものとなるでしょう。

計算論的思考

計算という概念の歴史的な意味をたどる試みとして、デニングとマーテル（2015）は、「計算とは、数学的な関数を評価するために従う機械的な手順であり、コンピュータとは計算を行う人と考えられていた」と述べています。しかし、1980 年代になると、この概念は、科学を行う新しい方法と結び付けられることが多くなり、その結果、機械から情報プロセスへと重点が移りました（Denning & Martell, 2015）。この変化は、私の議論にとって重要です。というのもそれは、報道と計算の究極の目標を一致させるからです。ジャーナリズムとは、情報プロセスを管理することでもあります。ごく一般的に言えば、ジャーナリストの仕事は、情報の流れを合理化し、それをキュレーションして、読者に受け入れられる形式にパッケージ化することなのです。私は、ニュース報道にコンピュテーションの考え方が浸透しているのは、この二つの職業的実践に共通点があるからだと主張します。コンピューティングもジャーナリズムも、問題を解決するための定型的なものであり、構文を使いこなす必要があります。ウィング（2008） は、「コンピューティングは、『この問題をコンピュータに解決させるにはどうすればよいか』という問いに答えることを目的としている」（p.3719）と述べています。そのためには、比較的高いレベルの計算論的思考が必要となります。

ニュースルームでコンピューディングが当たり前になるにつれ、データストーリーにアプローチするために計算論的思考を取り入れるジャーナリストが増えています。例えばブラッドショーは、計算論的思考は「データジャーナリストの仕事の中心にあり」、「現代のジャーナリズムの多くを構成する問題を解決し、ニュース処理に必要なスピードと正確さでそれを行うことができる」と主張しています（Bradshaw、2017）。

計算論的思考とは、問題を解決するために一連のプログラム的なステップを踏む再帰的なプロセスのことです（Bradshaw, 2017; Wing, 2006, 2008）。ウィングは「計算論的思考の本質は抽象化である」と主張し、問題を解決し、システムを設計し、人間の行動を理解するため、コンピューティングは「時間と空間の物理的な次元を超えて概念を抽象化する」と述べています。そして、「コンピュータにこの問題を解決させるにはどうすればよいか」という質問に答えるために、コンピューティングの専門家は、目の前の問題を解決するためのプログラム計画を設計し、実装に適した適切な抽象化（Wing, 2008, p. 3717）をしなければならないと主張します。

ニュースルームに自動化技術が導入されて以来、コンピュータの専門家と仕事をするジャーナリストは、同じような疑問に直面してきました。「どうすればコンピュータに人間の基準に沿ったニュース記事を調査・執筆させられるのか？」 ユンニルドは、プロのジャーナリズムに計算論的思考が注入されることで、ジャーナリストに「補完的、論理的、アルゴリズム的なスキル、態度、価値観」を身につけさせる「記述的なストーリーテリングから抽象的な推論、自律的な調査、定量的な事実の視覚化までのジャーナリズムの基本的な思考システム」が試されるとしています（Gynnild, 2014）。

むろん「計算論的」抽象化という考え方が、ジャーナリストにとって新しいものだというわけではありません。金融、ビジネス、不動産、教育などの専門ジャーナリストは、市場のパフォーマンス、株式リターン、家計の純資産などの複雑なダイナミクスを理解するために、日常的に抽象化を行っています。興味深いことに、マイルス（2019）が指摘しているように、自動化によってジャーナリストが煩わしい仕事から解放されるという予想に反し、これまでジャーナリストが行っていなかったさまざまな新しい編集活動が導入されています。例えば、AP 通信社のワークフローに画像認識が導入されたことで、ジャーナリストや写真家は、学習データのラベル付け、テスト結果の評価、メタデータの修正、概念の定義の生成など、従来は機械学習に関連していた作業に従事しなければならなくなったそうです（Myles, 2019）。

認知の投影と創造性の拡張

これまで私は、ニュースの自動化がもたらす計算問題に仕事として取り組むジャーナリストは、そのワークフローや編集責任の変容を強いられることを論じてきました。例えば、ウォール・ストリート・ジャーナルは最近、Machine Learning Journalist、Automation Editor、Emerging Processes Editor など、AI と自動化に関連した職種を募集しています。このようなインフラの拡大と、それに伴う編集責任の多様化の結果、ジャーナリストはしばしば、ジャーナリストのように考え、実行しなければならない機械の立場を想像することになります。これは興味深いパラドックスであり、同様に興味深い課題でもあります。

「投影」という考え方は、ニュースの自動化にも通じるものがあると思います。例えば、ジャーナリズムの真髄である「記事の作成」を考えてみましょう。このプロセスを分解すると、一般的には、ジャーナリストは創造力を駆使して、人々を惹きつけたり、情報を提供したりするような出来事の説明をまとめなければなりません。

そこで問題となるのが 人間の記者が書いたかのようなニュースを機械に書かせるには、どうすればい

いのかということです。この 5 年間、ジャーナリストと技術者は、この問題を解決するために、自分自身を投影しながら共同作業を行ってきました。その好例が、記事作成を自動化する自然言語生成（NLG）技術の導入です。

予想に反することでしょうが、このプロセスでは、人間の記者が記事のテンプレートを書きます。そのテンプレートには空欄があり、自動化ソフトウェアがデータベースを使って空欄を埋めるのです。このプロセスは、AP 通信社などの報道機関や、Press Association と Urbs Media が共同で設立した RADAR で成功しており、スポーツ、企業決算、ローカルニュースなどの分野で、ニュース制作作業のスピードと規模の拡大をめざしています。この領域での創造性は、新しい形を取ります。コーダー・ジャーナリストは、ストーリーテリングを、ニュース作成プロセスを解読・再構成する仕組みとして再考せねばなりません。どんなインタビューが説得力を持つか、どんな言葉を選べばより強い見出しになるのかではなく、どの条件文がより効率的で、どんな見出しが効率的に読者の目を引くかを選択することに目標がシフトしています。

ヒューマン・コンピュータ・インタラクション（HCI）やユーザー・エクスペリエンス（UX）デザインの原則に従って、コーダー・ジャーナリストは、ユーザーが自動化された情報体験にどのように関わりたいのか、ユーザーがさまざまな情報の層やニュース記事の制限をナビゲートする潜在的な方法を予測しなければなりません。「創造性の拡張」を提唱したホイラー（2018）は知的創造には、「その思考や思考を実現する物質的な媒体が、脳、身体、世界に空間的に分散している」ケースがあると説明しています。

この創造性の拡張の概念は、データと自動化を扱うジャーナリストの心が、一連の Python ライブラリ、Jupyter Notebooks、データセット、データ分析ツール、オンラインプラットフォームにまたがる一連の自動化［技術］と密接に関連して機能するようになったという考えをうまく説明します。

このダイナミックな動きは、注目すべき課題をもたらします。フル・ファクト社の自動ファクトチェック部門の責任者であるメバン・ババカルは、同社の自動ファクトチェックで直面する課題の一つに、文脈を挙げています。

例えば、英国のテリーザ・メイ前首相は、野党の労働党がマニフェストで約束した以上に多くの資源を国民健康保険サービス（NHS）に割り当てたと主張しました。この主張は正確だと確認されていますが、一般の人々にとって意味のある有用なものとするためには、より広い文脈の中で理解される必要があります。つまり、NHS が効率的に業務を遂行するためには、［なお］配分は十分ではなかったということです（Babakar, personal communication, August 16, 2018）。自動化されたシステムは情報源の間でそのような文脈のつながりを作ることがまだできません。ババカルたちは、「自動化されたファクトチェッカーに文脈のニュアンスを理解させるにはどうすればよいか？」という質問を逃れることはできないのです。

ジャーナリスティックな分散型認知

最後に、ジャーナリスティックな分散型認知の考え方と、それがもたらす疑問についてさらに掘り下げてみたいと思います。アンダーソン、ホイラーとスプレヴァク（2018）は、コンピュータが人間の活動に浸透するにつれ、認知が「脳、非神経系の身体、そして…物、道具、その他の人工物、テキスト、個人、グループ、そして/または社会

的/組織的構造からなる環境に広がる」と論じています。ジャーナリズムの世界では目下、ジャーナリストがネットワーク化されたソフトウェアやハードウェアを使用して、大規模かつ迅速にニュースを作成する能力を高めています。これは、彼らの認知が、使用するさまざまなプラットフォームやツールの範囲に分散するようになることを意味します。これによって、オンラインで人類の知識のほとんどに無制限にアクセスできるようになりました。しかし、このポータブルな知識と分散された認知という考え方は、ジャーナリストが豊富な知識と「自由な」分析力にアクセスすることを誰が管理しているのかという疑問を生じさせます。

ジャーナリストの分散型認知を可能にするのは誰か？この問題は、Google がオンラインデータビジュアライゼーションツール「Google Fusion Tables」を閉鎖したときに経験したように、より深い議論が必要です。閉鎖後、このツールを使って開発された何十ものデータジャーナリズム・プロジェクトは、同社のサポートを受けられなくなったため、利用できなくなりました。ジャーナリストが日常的に計算力学に関わることで、彼らの計算論的思考が常態化し、彼らの認識を日常的なジャーナリズム活動に用いる機械に投影することが容易になるでしょう。ジャーナリズムの認知が分散されると、ジャーナリズムの権威や統制も同じように分散するのでしょうか。

容赦なく[知識の]分散はジャーナリストの日常や職業文化に対するコントロールを可能にする境界をシフトさせ、その結果、彼らの認識論的権威に影響を与えることになります。この章の架空の導入部で行ったように、将来を見据えると、ジャーナリストが重要な倫理的検討や決定を機械に委ねるようになれば、それに伴ってさまざまな関連するリスクを生み出す可能性もあります。データと自動化の時代にジャーナリズムの基礎となる理想を維持するのであれば、彼らが認知に使用するインフラがオープンであり、公共の監視に利用可能であることが重要です。

参考文献

Anderson, M., Wheeler, M., & Sprevak, M. (2018). Distributed cognition and the humanities. In M. Anderson, D.Cairns, & M. Sprevak (Eds.), *Distributed cognition in classical antiquity* (pp. 1–17). Edinburgh University Press.

Bradshaw, P. (2017). Computational thinking and the next wave of data journal- ism. *Online Journalism Blog*. onlinejournalismblog.com/2017/08/03/computational-thinking-data-journalism

Denning, P. J., & Martell, C. H. (2015). *Great principles of computing*. The MIT Press.

Gynnild, A. (2014). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism*, *15*(6), 713–730. doi.org/10.1177/1464884913486393

Myles, S. (2019, February 1). *Photomation or fauxtomation? Automation in the news-room and the impact on editorial labour: A case study.* [Technology]. Computation + Journalism Symposium 2019, University of Miami.

Wheeler, M. (2018). Talking about more than heads: The embodied, embedded and extended creative mind. In B. Gaut & M. Kieran (Eds.), *Creativity and philosophy* (pp. 230–250). Routledge. dspace.stir.ac.uk/handle/1893/26296

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*, 33–35. doi.org/10.1145/1118178.1118215

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of the Royal*

Society A: Mathematical, Physical and Engineering Sciences, *366*(1881), 3717–3725.

doi.org/10.1098/rsta.2008.0118

データジャーナリズムの態様

文：サラ・コーエン

概要

データジャーナリズムがどのように進化してきたか、また、伝統的な調査報道からニュースアプリやビジュアライゼーションまで、さまざまな形態で展開されていることを紹介します。

キーワード：調査報道、ニュースアプリ、データ・ビジュアライゼーション、解説ジャーナリズム、精度ジャーナリズム

データ（dey-tah）：事実や情報のまとまり、個々の事実、統計、情報の項目（“Data,” n.d.）

ジャーナリズム：ニュースの報道、執筆、編集、写真撮影、放送を行う職業、またはビジネスとして報道機関を運営する職業（“Journalism,” n.d.）

このハンドブックを読んでいるということは、データジャーナリズムとして知られる仕事について学ぼうと思っていることでしょう。しかし、データポータルがオープンになり、目を見張るようなビジュアライゼーションがあり、世界中で情報の自由が争われているこの時代に、データジャーナリズムとは一体何を意味するのでしょうか？

この二つの言葉の辞書的な定義はあまり役に立たないものの、まとめると、データジャーナリズムとは、事実や情報で構成されたニュースを作る職業であることを示しています。データジャーナリズムは、電子的に保持された記録や統計を扱うジャーナリズムの行為、つまりジャーナリズムのほぼすべてを意味するようになりました。

そのため、この分野の多くの人々は、自分たちをデータジャーナリストとは思っていません。解説記者、グラフィック/ビジュアル・ジャーナリスト、記者、オーディエンスアナリスト、ニュースアプリケーション開発者など、成長を続けるこの分野をより正確に表す名称を使っているのです。それだけでは不十分なので、ニュースルームで数字を使わなければならないものや、コンピュータープログラミングを必要とするものも加わっています。かつてはガレージバンドだったものが、今ではオーケストラを構成するほどの規模になっています。データジャーナリズムはそれほど新しいものではありません。実際、「データ」をある種の体系的に収集されたものと考えるならば、米国における最古のデータジャーナリズムのいくつかは、18 世紀半ばに遡ります。

「フランク・レスリーのイラスト入り新聞」の発行者フランク・レスリーは探偵を雇ってニューヨーク市内の酪農カートを追跡し、誤表示や汚染された牛乳を記録しました。非営利調査サイト「プロパブリカ」のマネージング・エディター、スコット・クライン（2016）は、新聞が読者に棒グラフの見方を教えていたという、

同じく 18 世紀に遡るデータジャーナリズムの魅力的な歴史を記録しています。クリス・アンダーソンも本編の（別の）章で、1910 年代、1960 年代、2010 年代のデータジャーナリズムの変遷を探っています。このような歴史を踏まえて、データジャーナリズムのさまざまな分野を分類することで、学生や実務者が自分の好みや、成功するのに必要となるスキルを明確にすることができます。ここでは、データジャーナリズムのさまざまな方法を、この分野の発展のおおよその沿革にのっとり紹介します。

検証ジャーナリズム、またはストーリーのためのデータ

ロイター通信のモーリス・タマンは、二つのデータジャーナリズムの伝統を組み合わせる方法として、「検証ジャーナリズム」という言葉を作りました。1960 年代にフィリップ・マイヤーが開発した精密ジャーナリズムは、社会科学的な手法をストーリーに活用しようとするものでした。マイヤーは、デトロイトの暴徒を対象とした調査から、フィラデルフィアの裁判所における人種的偏見に関する調査のデータ収集と分析の指揮まで、さまざまな仕事を手がけました。彼は、一世代にわたって調査の基礎を築いたのです。検証ジャーナリズムは、1990 年代にロードアイランド州プロビデンスでエリオット・ジャスピンが提唱した「コンピュータ支援報道」というジャンルにも含まれます。この分野では、記者は情報や記事のアイデアを調べるために、証拠となる電子書類を探したり、必要に応じて作成したりします。最近では、人工知能や機械学習を利用して、ストーリーの展開を見つけたり、単純化したりすることができるようになりました。医療機器によって被害を受けた患者の性別を政府が隠そうとしていた場合など、単純な疑問に答えるために使用することができます。

あるいは、バズフィードのピーター・アルドースが手掛けたスパイ機の分析のように、複雑なパターンを見極めるために使うことも可能です（Aldhous, 2017; Woodman, 2019）。こうした記者はほとんどニュース収集家というべき存在です。彼らの目的は、データを使って可視化したり、ストーリーを伝えたりすることではなく、データを駆使し潜在的なストーリーを探ることです。彼らの仕事は報道に不可欠なものであり、調査の進展を促します。彼らは通常、ストーリーのプレゼンテーションにはあまり関与しません。この「データジャーナリズム」の世界に新たに参入してきたのは、世界的に拡大している可視化やオープンソースに基づく調査です。 このジャンルは、インテリジェンスや人権にまつわる調査から派生したもので、「データ」の概念を動画やクラウドソーシングによるソーシャルメディア、その他のデジタル成果物にまで拡大しています。コーディングにはさして依存しませんが、独自の調査によって他の人が秘密にしておきたいことを明らかにするという点では、データジャーナリズムの伝統にしっかりと適合しています。最も有名な例の一つ、BBC のドキュメンタリー番組「アフリカ・アイ」の「Anatomy of a Killing」は、カメルーンのどこで、いつ、ある一家が虐殺されたのかを正確に解明し、誰が関与したのかを明らかにしています。カメルーン政府は「フェイクニュース」と否定しましたが、チームは、Google Earth で山の尾根の輪郭を確認したり、Facebook で犯人が着ていた服を記録したりと、さまざまなツールを活用しました。

データ・ビジュアライゼーション

国際的なデータジャーナリズム・アワードの受賞者を見ると、可視化こそがデータジャーナリズムの鍵を握っているのではないかと思えてきます1。統計が貨幣なら、ビジュアライゼーションは[データジャーナリズム]クラブへの入場料です。ビジュアライゼーションは、データジャーナリズムのツールボックスの重要な一部となります。しかし、そのためには、データや統計、報道の世界だけでなく、デザインやアートの世界からツールキットを調達する必要があります。今日、学術界で活躍する最も有名なビジュアル・ジャーナリストの一人である アルベルト・カイロは、雑誌や新聞のインフォグラフィックスの仕事が出発点でした。彼の仕事は、ビジュアライゼーションによってストーリーを伝えることに焦点を当てており、ニュースを収集するだけでなく、ストーリーを伝える役割も担っています。

ニュースアプリ

プロパブリカでは、大規模な調査報道のほとんどがニュースアプリ内で自己完結します。ニュースアプリとは、魅力的で洞察力に富んだインターフェースを通じて、地域や個人のデータにアクセスできるサイトや機能のことです。プロパブリカはニュースアプリで有名になり、キャリアをコーディングから始めたエンジニアは、言葉ではなくコードを使ってストーリーを伝えるジャーナリストへと進化しました。

同社のケン・シュエンケは、開発者としてロサンゼルス・タイムズやニューヨーク・タイムズなどのニュースルームで働いていましたが、プロパブリカのニュースアプリを通じてクラウドソースで収集したストーリーを中心としたサイトの Documenting Hate プロジェクトの一環として、米国のヘイトクライムを取材する全米屈指のジャーナリストとなりました。

データストーリー

「データジャーナリズム」という言葉は、記者や統計学者などの専門家が、データそのものをジャーナリズムの一形態として扱うようになって生まれました。このジャンルを広めたのは、ガーディアンの Datablog の生みの親であるサイモン・ロジャースです。その後、FiveThirtyEight、Vox、そして後にはニューヨーク・タイムズの Upshot がこの分野の旗手となったのです。それぞれの役割は少しずつ異なりますが、統計や分析はそれだけでニュースになるという考えに収斂していきました。中には、アメリカの大統領選のオッズをつける政治予測で有名になった人[ネイト・シルバー]もいます。また、一般の人々の心理を垣間見ることができる、一風変わったデータセットを見つけることでも知られています。

その一例が、Facebook から導き出されたひいきの野球チームを表わす 2014 年の地図です。このジャンルの入り口はデータセットであり、テーマに沿った専門知識を持つことが、これらの実践者が他の分野との差別化を図る方法です。実際、このジャンルを定義したネイト・シルバーらは、ジャーナリズムではなく、統計学や政治学の出身です。ニューヨーク・タイムズの Upshot の編集者であるアマンダ・コックスは、このサイトの役割を、既知の事実未知の事実の間を占めること、つまり、純粋な事実と純粋な意見の境界を行き来しながら、入

手可能なデータの専門的分析による洞察を提供するジャーナリズムだと述べています(Cox, personal communication, n.d.)。

アルゴリズムの検証

データジャーナリズムの新たな分野は，まさにテクノロジーに関するジャーナリズムであり、ノースウェスタン大学のニコラス・ディアコプロスの造語である「アルゴリズム説明責任」の分野です2。記者のジュリア・アングウィンとジェフ・ラーソンは、この専門性を追求するためにプロパブリカを離れ、「The Markup」を設立しました。このサイトは、機械学習や人工知能のアルゴリズムが決定する、懲役刑の判決や消費者の郵便番号に基づく請求価格など、社会にもたらす結果についてテック企業に説明責任を問うとうたっています。この報道を受けて、YouTube はすでにレコメンデーションエンジンを見直し、視聴者をより暴力的な動画に誘う傾向を減らしています。また、Facebook が差別的な住宅広告を出している可能性があるとして責任を追及したり、オンラインストアでユーザーの位置情報に基づく価格差別を確認したりしているのです(Dwoskin, 2019)。

脚注

1. ルースンの章を参照。

2. この分野はディアコプロスとエルマーの章を参照。

参考文献

Aldhous, P. (2017, August 8). We trained a computer to search for hidden spy planes. This is what it found. *BuzzFeed News*. www.buzzfeednews.com/article/peteraldhous/hidden-spy-planes

BBC News. (2018, September 23). *Anatomy of a killing*. BBC Africa Eye. https://www.youtube.com/watch?v=4G9SeoLgX4

Data. (n.d.). In *Dictionary.com*. Retrieved May 20, 2020, from www.dictionary.com/browse/data

Dwoskin, E. (2019, January 25). YouTube is changing its algorithms to stop recommending conspiracies. *The Washington Post*. www.washingtonpost.com/technology/2019/01/25/youtube-is-changing-its-algorithms-stoprecommending-conspiracies/

Journalism. (n.d.). In *Dictionary.com*. Retrieved May 20, 2020, from www.dictionary.com/browse/journalism

Klein, S. (2016, March 16). Infographics in the time of cholera. *ProPublica*. www.propublica.org/nerds/infographicsin-the-time-of-cholera

Woodman, S. (2019, October 22). Using the power of machines to complete impossible reporting tasks. ICIJ.

www.icij.org/blog/2019/10/using-the-power-of-machines-to-complete-impossible-reporting-tasks

データ・ビジュアライゼーション:ニュースルームのトレンドと日常的な関わり

文:ヘレン・ケネディ、ウィリアム・アレン、ローズマリー・ルーシー・ヒル、マーティン・エンゲブレッツェン、アンディ・カーク、ウィブケ・ウェーバー

ここでは、ニュースルームにおけるデータ・ビジュアライゼーション(以下 dataviz)の制作と、読者が日常的に dataviz にどのように関わっているのかについて、二つの独立した研究プロジェクトを通じて考察します。 一つ目は、人々がデータ・ビジュアライゼーションをどのように意味づけているかを研究した「Seeing Data」、二つ目は、社会における記号論的、美学的、論証的な資源としての dataviz を探る「INDVIL」です1。

ここでは、まず INDVIL サブプロジェクトの成果のあらましを紹介します。このプロジェクトでは、ニュースの中の dataviz がさまざまな方法で認識され、多様な目的に利用されていることを明らかにしました。

次に、Seeing Data で得られた主な調査結果を要約します。Seeing Data では、視聴者の dataviz に対する理解が、非常に多様性に富んでいることもわかりました。この多様性は、dataviz の研究者と実践者双方による今後の研究にとって重要なものです。

ニュースルームにおけるデータ・ビジュアライゼーション:トレンドと課題

データ・ビジュアライゼーションはニュースルームの業務にどのように組み込まれているのでしょうか。どんな傾向があり、どんな課題が生じているのでしょう。これらの疑問に答えるべく、2016、17 年に、ヨーロッパ 6 カ国、すなわちノルウェー(NO)、スウェーデン(SE)、デンマーク(DK)、ドイツ(DE)、スイス(CH)、イギリス(UK)の 26 のニュースルームに 60 件のインタビューを実施しました。主要なオンラインニュースメディアのインタビュー対象者には、編集長、データ・ビジュアライゼーション専門チームのリーダー、データジャーナリスト、ビジュアル・ジャーナリスト、グラフィック/データ・ビジュアライゼーションデザイナー、開発者が含まれていました(一部の人は特定の職種に当てはまりませんでしたが、これはこの分野が急速に発展していることを示しています)。ここでは、調査のハイライトをご紹介します。

ジャーナリズムのストーリーテリングを変える

ジャーナリズムの中でデータ・ビジュアライゼーションがますます使われるようになってきていることは、ストーリーテリングの中心的な要素が、主な記号体系としての文章からデータとビジュアライゼーションに移りつつあることを示しています。多くのインタビュー対象者は、データ・ビジュアライ

ゼーションは、たとえそれが単純なグラフィックやダイアグラムであっても、ストーリーを駆動させると述べています。

「統計によると、ストーリーの中にシンプルなデータ・ビジュアライゼーションを挿入すれば、読者のページ滞在時間はやや延びる」（スウェーデン）

dataviz は、「洞察を与える」（英国）、「より平易に説明する」（スウェーデン）、「言葉よりもはっきりと伝える」（英国）、「テキストでは要約でしか表現できない複数の側面を詳細に伝える」（ドイツ）、「アクセスしやすい」（デンマーク）、「悲惨な状況を明らかにする」（スイス）、「人々が世界を理解する手助けをする」（英国）など、コミュニケーションを目的とした幅広い用途に使用されています。

つまり、あるポイントを強調したり、経験的証拠を追加したり、ユーザーがデータセットを探索できるようにしたり、また興味を刺激するための美的魅力として、そして未知のストーリーへの入り口として使われたりしているのです。このような変化に伴い、ニュースルーム内に複数のスキルを持つ専門家グループが出現し、新入社員にはデータや dataviz のスキルが求められるようになりました。

しかし、ニュースルーム内での dataviz 制作組織にはパターンがなく、あるところではデータチーム、あるところではビジュアル・チーム（インタビューした時点で dataviz デザイナーの一人はバーチャルリアリティのプロジェクトにも取り組んでいました）が担当し、また別のチームで行っているところもあります。

このように、新たな視覚的形態に対応する新しい（ビジュアライゼーションの）構造が生まれつつあるのと同じく、ニュースルームのスタッフも、社内および商用の新しいツールを学び、新しいスキルを開発し、効果的なデータストーリーを制作するために、チームや専門分野を超えてコミュニケーションをとる方法を理解する必要があります。

「モバイルファースト」のマントラとその結果

視聴者が小型のモバイルスクリーンでニュースを見る機会が増えていることが広く認識され、ニュースルームでの dataviz 制作においても「モバイルファースト」のマントラが広く採用されるようになってきました。初期には精巧でインタラクティブだったデータ・ビジュアライゼーションは、よりシンプルで直線的な、あるいはインタラクティブ性の低いシンプルな形態へ転換しました。棒グラフや折れ線グラフなどの特定のチャートタイプが優勢になり、スクローリーテリングや、ユーザーがページをスクロールすると展開されるストーリーが、記事に埋め込まれたビジュアライゼーションが適切なタイミングで表示されるようになりました。スクロールすることで、ビジュアライゼーション自体にもズームアウトなどの変化が起こります。

「私たちのストーリーでは、スクロールのテクニックを使っています。クリックする必要はありませんが、下にスクロールすると、ストーリーの中で何かが起こります」（ドイツ）

dataviz を自動で制作したり、専門家でなくても制作を可能としたりするツールの登場により、簡略化されたグラフ形式が普及してきています。しかし、あまり一般的ではない図表（例えば散布図）を、解釈する方法とともに提示することで、読者を教育したいと考えている人もいます。北欧のタブロイド紙は、42 万 7000 枚のピザで埋め尽くすことで、貨物機の大きさを表しました。時間の経過に伴う変化を表示するアニメーションや、ビジュアライゼーションにおける拡大縮小機能の試みに価値を見出す人もいます。

ジャーナリズムが果たす社会的役割

dataviz を情報源にリンクさせ、生データへのアクセスを提供し、方法論を説明することは、透明性を生み出し、選択と解釈の主観性を相殺する倫理的慣行であり、一部の参加者にとっては、データの可視化に不可欠な側面です。しかし、他の参加者にとっては、情報源へのリンクは、視聴者に「すべてのデータ」を与えることを意味し、情報を裏付けてからストーリーを語るというジャーナリズムの規範と相反するものに映ります。

一部の企業では、この対立に対処するために、異なるプラットフォーム（Twitter、Pinterest、Github）上でデータとプロセスの異なる要素を共有する複雑なプロセスを採用しています。データジャーナリストやビジュアライゼーションデザイナーは、どの程度データを共有すべきか、事実の提供者としての役割や、より一般的な社会的な役割について考える必要があります。データジャーナリストのポール・ブラッドショーはこの問題をブログで取り上げました。

人々が私たちの情報を使って語るストーリーに、私たちはどの程度の責任を負うだろう？そして、誰かが必要とするだけの情報を提供する責任はどの程度あるのだろう。（Bradshaw, 2017）

ガーディアンの元編集者アラン・ラスブリッジャーも、ジャーナリズムの社会的役割に疑問を提起し、ジャーナリズムが歴史的に行ってきたこと、つまりデータや公式情報の裏付けを取る人々の取り組みを指摘しています（英国では FixMyStreet や TheyWorkForYou など）。彼は「それがジャーナリズムなのかは分からない。それが重要なのかどうかも分からない」としています（Baack 2018）。

インタビュー対象者には、ラスブリッジャーが論じたような大規模なプロジェクトに取り組んでいる人もいます。例えば、英国の学校に関連するすべての利用可能なデータを照合し、学校選びに役立つように、郵便番号で検索できるようにしました。したがって、幅広いデータと dataviz の観点から何がジャーナリズムとみなされるのかという問いに答えるのは容易ではありません。

さらに、データセットを共有することは、読者がデータセットと対話できることを前提にしていますが、研究によると、オンラインでの対話は、現実同様に神話であり、可視化されたデータセットを意欲的に探索する理想的なイメージは、よりありふれた、ニュースを素早くスクロールする読者とは対照的な姿なのです（例：Burmester et al 2010）。

同様に、北欧データジャーナリズム・アワードに提出されたデータジャーナリズム・プロジェクトの研究では、ほとんどの選択肢はジャーナリストによってすでに決定されているか、事前に定義されているため、インタラクティブ（に見える）要素はしばしばインタラクティブ性の錯覚を提供しているに過ぎないと結論づけています（Appelgren 2017）。このことは、「すべてのデータ」を共有する実践と、ジャーナリズムの社会的役割の変化に疑問を提起しています。

信頼、真実、「野放しの」ビジュアライゼーション

データを可視化するプロセスの他の要素は、信頼と真実の問題を提起し、ジャーナリストがジャーナリズムの社会的役割をどのように考えるかにも関係してきます。こうした問題を指摘するdataviz の側面として挙げられるのは、データの可視化に取り組むジャーナリストが、データとその視覚的表現についてどのように考えているかということです。真実を語る一つの形態として見る人もいれば、選択と解釈のプロセスとして見る人もいます。また、選択によってデータ・ビジュアライゼーションを形成することはストーリーを明らかにする方法であり、まさにジャーナリストがなすべきことだと考える人もいます。これらの考察は、（非）信頼性とプレゼンテーションの関係、そして視点と（非）真実性の関係を浮き彫りにします。読者が十分な事実、データ、専門家にうんざりし、フェイクニュースが迅速かつ広範に流布している「ポスト・トゥルース」の状況下で、参加者は読者がデータ・ビジュアライゼーションにどのような反応を示すのか注視してきました。それには素朴に受け入れること、懐疑的に反論すること、社会的共有を通じて非組織化すること、さらには変更や改ざんさえ含まれています。ある参加者（英国）はジャーナリストには「インターネット文化のソフトな知識」がますます必要になっていると感じています。

それは、オンラインのコンテンツはオフラインのコンテンツよりも［読者からの］問いにさらされる可能性があり、データ・ビジュアライゼーションがテキストよりもオンラインで流通しやすく、数字と画像の組み合わせとして元の文脈から離れて「野放し」になる可能性があることを理解するということです（Espeland and Sauder 2007）。そのためには、説明文をビジュアライゼーションファイルに埋め込み、説明文なしで画像が流通しないようにするなど、これらの危険性に対処する戦略を理解すべきです。これらの問題は、読者のデータやビジュアライゼーションのリテラシーに関する懸念と並んで、ジャーナリストの視聴者に関する考え方に影響を与え、再構築するものなのです。

人々はいかにしてデータ・ビジュアライゼーションに関わっているのか？

このセクションでは、読者の視点で dataviz を見ていきます。読者はどのようにニュースのデータ・ビジュアライゼーションに関わり、意味を見いだすのでしょう。データジャーナリストは忙しくてこの質問に答えることはできません。研究者はそんな言い訳はしませんが、それでもエンドユーザーが目にしたビジュアライゼーションを見てどう思うかに注目することはほとんどありません。

Seeing Data は、人々が日常生活 、特にメディアで遭遇するデータ・ビジュアライゼーションにどのように関わっているかを調査しました。このプロジェクトでは、人々の関心に影響を与える要因と、ビジュアライゼーションを効果的なものにする考え方を探りました。Seeing Data では、[定性調査の]フォーカスグループとインタビューを使ってこれらの疑問を調査し、人々の dataviz との関わりの根底にある態度、感情、信念を知ることができました。調査には 46 名が参加しましたが、その中には、データやビジュアライゼーション、[提示したビジュアライゼーションのテーマである]移民に興味があり、プロジェクトの中心となる問題の一つに「すでに関与している」と思われる参加者と、このような仮定を立てることができない参加者が混在していました。フォーカスグループでは、参加者に八つのビジュアライゼーションを評価してもらいました。これらの資料は、テーマ、チャートの種類、オリジナルのメディアソース、フォーマットが多様で、[役割が]説明的か探索的かのいずれかであるため、議論の末に選ばれました。ビジュアライゼーションの半分は、ジャーナリズム（BBC、ニューヨーク・タイムズ、英国の無料紙「メトロ」、Scientific American）から引用されています。

また、オックスフォード大学の移民観測所、英国の国家統計局（ONS）、経済協力開発機構（OECD）など、仕事の一環としてデータを可視化して共有している組織のものもありました。フォーカスグループの（調査）後、7 人の参加者が 1 カ月間日記をつけ、私たちが選んだのではない「野放し」のビジュアライゼーションとの出合いについてさらなる情報を提供してくれました。

Dataviz のエンゲージメントに影響を与える要因

主題。ビジュアライゼーションは、それが表現する主題から切り離されて存在するわけではありません。主題が参加者の関心を呼ぶものであれば、参加者は積極的に参加します－例えば市民社会の専門家は移民に関心があり、それに関するビジュアライゼーションに興味を持っていました。対照的に、ある参加者（38 歳の英国人農家男性）は、フォーカスグループで見せられたビジュアライゼーションには興味を示さず、時間をかけてみることに自信がないようでした。しかし、（『混乱させようとしている』と感じたという）メディアに対する不信はあったものの、あらゆるビジュアライゼーションを拒んだわけではありません。愛読する[農業雑誌]「ファーマーズ・ガイド」でビジュアライゼーションを見かけると、時間をかけて見てしまうとのことです。

情報源やメディアのロケーション。ビジュアライゼーションの情報源は、ユーザーがそれを信頼するかどうかに関わる重要なものです。多くの参加者は、メディアが混乱を招くのではないかと懸念し、特定のメディアに掲載されたビジュアライゼーションを疑わしく思う人もいました。一方で、オックスフォード大学のロゴが入った移民問題のビジュアライゼーションは、同大学の「ブランド」が品質や権威を連想させると考え、信頼する参加者もいました。しかし、記録を続ける間に、別の姿が見えてきました。参加者は自分が信頼するメディアで見たビジュアライゼーションを信頼する可能性が高いのです。ある参加者（24 歳の英国白人男性農家）は、Daily Mail の購読者ですが、インタビューの中で「The Sun は、間違って掲載されていることがたくさんある」と述べ、このことを実証しました。この二つの新聞はイデオロギー的には似通っており、このコメントは dataviz とユーザーとのエンゲージメントが、メディアのロケーションによって左右されることを示しているのです。

信念と意見。参加者は、自分がいつも読んでいる新聞を信頼し、その新聞に掲載されているビジュアライゼーションを信頼していました。これは、人々が特定のビジュアライゼーションに時間をかけて取り組むかどうかに影響を与える信念や意見の重要性を示しています。参加者の中には、自分の信念や意見を確認できるビジュアライゼーションが好きだという人もいました。しかし、信念が重要なのは、ビジュアライゼーションが信念を裏打ちする場合だけではありません。ある参加者（英国の白人で 34 歳の IT 企業会社員男性）は、figure24.1 の ONS のビジュアライゼーションにある移民データに驚きました。彼は、英国にはアイルランド生まれの人が何人もいることに気づかなかったと言います。このデータは彼の思い込みに疑問を投げかけ、彼はその体験を楽しみました。人によっては、既存の信念に疑問を投げかけるようなビジュアライゼーションのデータが好きだったり、興味を持ったりするものです。それらは[見る人の]視野を刺激し、広げます。信念や意見もこのように重要なのです。

時。ビジュアライゼーションが苦手な人にとって、それに取り組むことは仕事のようなものです。この「仕事」をしたいと思うかどうかは、時間的余裕があるかどうかが重要です。「ビジュアライゼーションを見る時間がない」と答えた参加者の多くは女性で、仕事や家庭の事情で時間が取れないと答えています。あるワーキングマザーは、仕事と家事の両立で疲れてしまい、1 日が終わるとニュースを見たくなくなると話していましたが、その中にはビジュアライゼーションを見ることも含まれていました。このような活動は彼女にとって「仕事」のように感じられ、忙しい一日の終わりにそれを行うには疲れすぎていたのです。ある農業従事者は、労働時間が非常に長いため、フォーカスグループ調査後に 1 ヶ月間のビジュアライゼーションとの関わりを日記に残すことができなかったとメールをくれました。

自信とスキル。読者は、ビジュアライゼーションを理解するスキルを持っていると感じる必要がありますが、多くの参加者はこの点について自信がないと答えています。（参加者の一人である）キャリアアドバイザーは「円と色がたくさんあって、ちょっと大変そうだな、よくわからないなと思いました」と話しました。参加者の多くは、自分にスキルがないことに不安を感じていたり、必要なスキルを持っていないと表明したりしました。そのスキルとはビジュアルリテラシー、言語スキル、［チャートタイプごとの読み方を知っているといった］数学的・統計的スキル、批判的思考スキルなどです。

感情。最後になりましたが、今回の調査では、人々がデータ・ビジュアライゼーションに関わる際に感情が重要な役割を果たすことがわかりました。喜び、怒り、悲しみ、罪悪感、恥ずかしさ、安心感、心配、愛情、共感、興奮、不快感など、データ・ビジュアライゼーションに関連する幅広い感情が浮かび上がりました。参加者は以下に対する感情的な反応を報告してくれました。それは一般的なビジュアライゼーション、表現されたデータ、視覚的なスタイル、データ・ビジュアライゼーションの対象、ビジュアライゼーションのソースまたは［情報の］元々の格納場所、ビジュアライゼーションを理解するための自分のスキルレベルです。例えば、2 人の市民社会の専門家は、figure24.2 に示した英国の移民に関するビジュアライゼーションを見たときの気持ちを、強い言葉で表現しました。彼らはこのデータを見て、英国に来た移民がメディアの移民排斥の見出しに遭遇したとき、どのような気持ちになるのかを思いやり、自分が英国人であることに「罪悪感」や「恥ずかしさ」を感じていると表現しました。また、一部のビジュアライゼーションでは、そのビジュアルスタイルに強い感情を抱いた参加者もいました。ニューヨーク・タイムズによる映画の興行収入のビジュアライゼーションでは、その美しさに惹かれる人と敬遠する人とで意見が分かれました（Bloch et al:）。

イメージとメッセージの連携が取れていて、見ていて楽しくなりました。

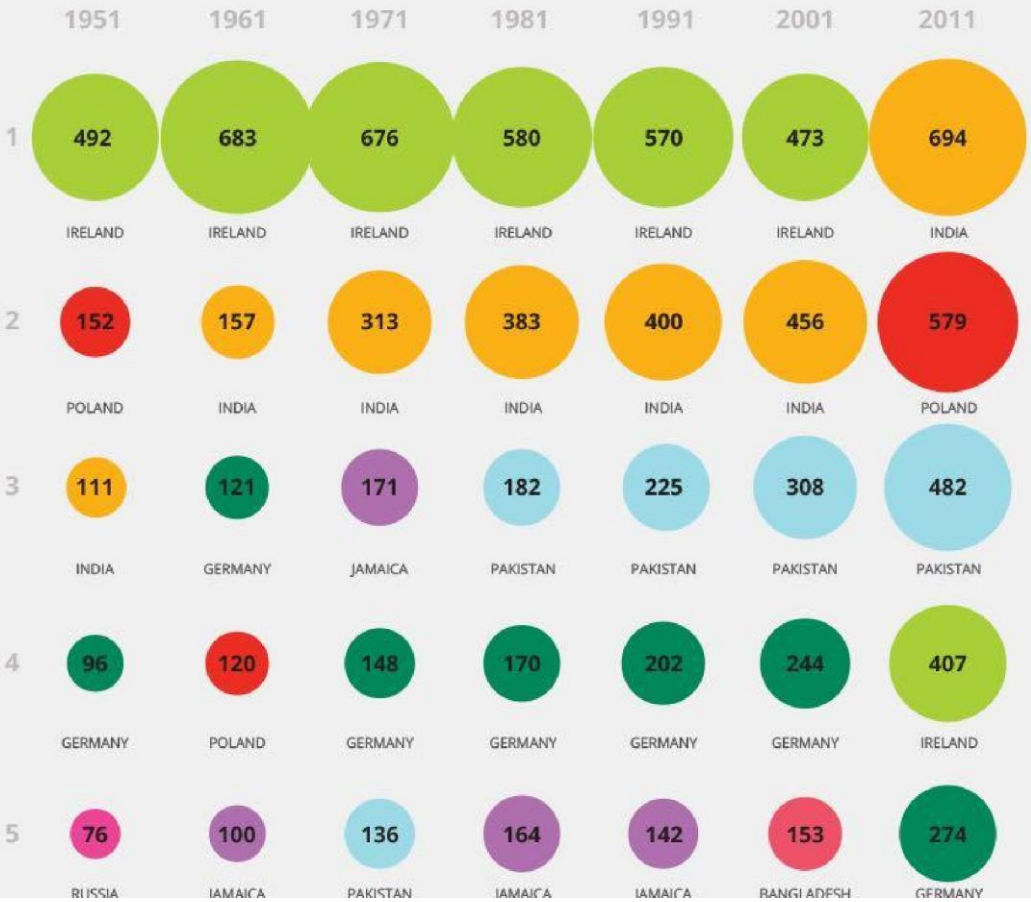
イライラする。そもそも美しくない表現で、はっきりと見づらく、情報もなく、ただごちゃごちゃしているだけだった。

Non-UK born census populations 1951 - 2011

13% (7.5 MILLION) OF RESIDENTS IN ENGLAND AND WALES WERE BORN OUTSIDE THE UK, 2011

TOP TEN NON-UK COUNTRIES OF BIRTH

NUMBERS ARE IN THOUSANDS



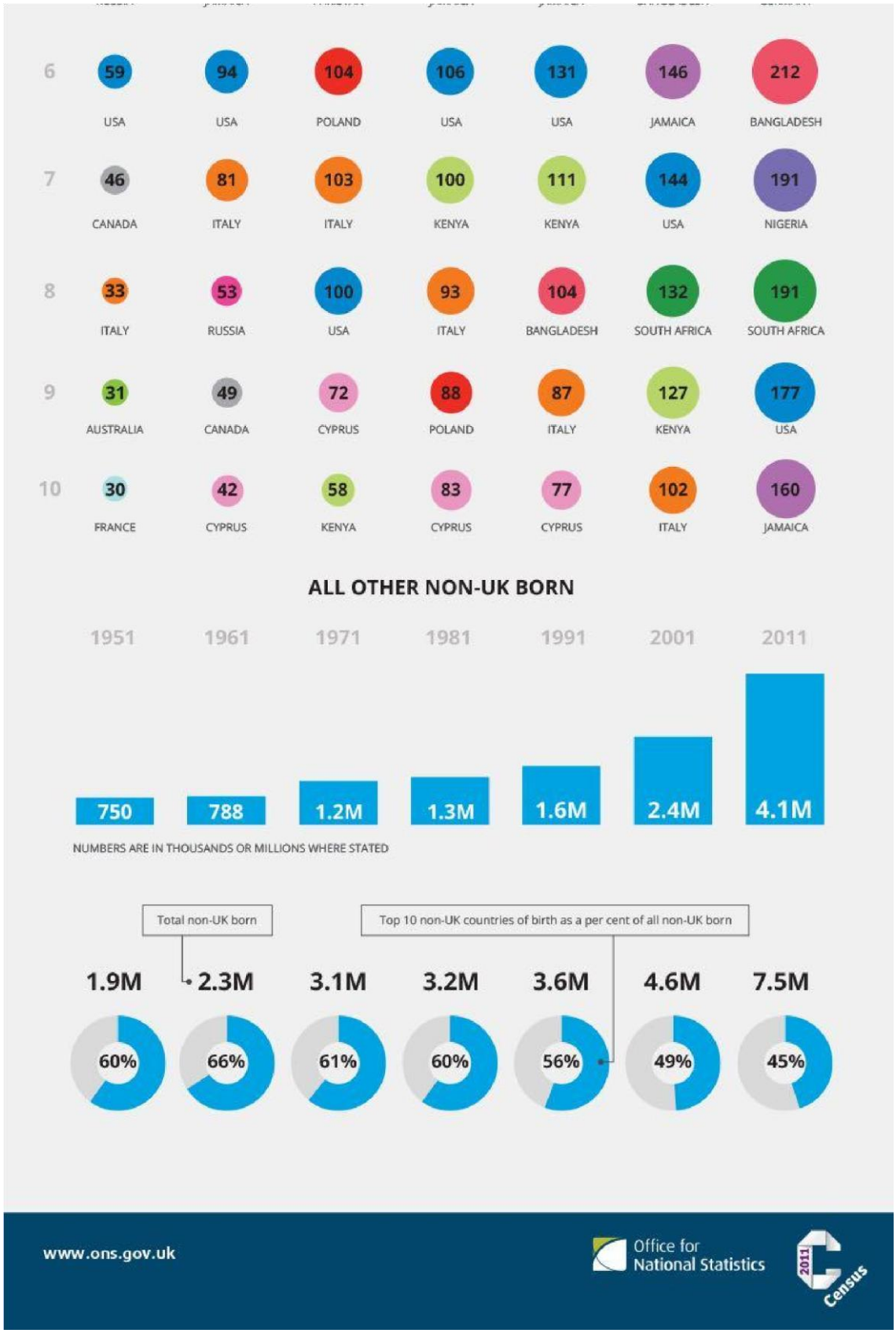


Figure 24.1. 非英国出生者の国勢調査人口 1951～2011 年　出典:国家統計局

Figure 24.2: 国勢調査における移民 オックスフォード大移民観測所
http://www.compas.ox.ac.uk/migrationinthecensus/,
http://migrationobservatory.ox.ac.uk/

効果的なビジュアライゼーションを実現するためのポイント

今回の調査では、効果的なビジュアライゼーションとは何かについて、さまざまなことが分かりました。例えば、専門家ではない人向けのビジュアライゼーションは、説得することを目的としているかもしれません。ビジュアライゼーションのベースとなっているデータを知るために時間を割いてもらうためには、人々を引き付ける必要があります。ビジュアライゼーションは特定の感情を刺激し、人々がより長く、より深く、より遠くを見ようとするきっかけとなるかもしれません。また、興味を持たせる場合もあれば、その逆の場合もあります。効果的なビジュアライゼーションとは次のようなものです。

- 疑問を抱かせる／他の人と議論したいと思わせる
- データの中の他者への共感を生み出す
- ユーザーを惹きつける好奇心の喚起

- 既存の知識の補強やバックアップ
- 驚きをもたらす
- 説得する、あるいは心を変える
- 新しいものを提示する
- dataviz を理解する新たな自信につながる
- 自分の目的に役立つデータを提示する
- 情報に基づいた、あるいは批判的なトピックへの関与を可能にする
- 楽しい経験をする
- 強い感情的な反応を引き起こす

ビジュアライゼーションが効果的であるかどうかは流動的であり、すべての dataviz に当てはまる単一の定義はありません。例えば、ビジュアライゼーションによって楽しませることは、ある文脈では適切ですが、他の文脈では適切ではありません。ビジュアライゼーションには、新しいデータの伝達、読者への情報提供、意思決定への影響、データの探索と分析を可能にすること、驚きと行動への影響など、さまざまな目的があります。

研究で特定したエンゲージメントに影響を与える要因は、異なるビジュアライゼーション、コンテキスト、目的に関連して異なる重みを持つ、効果の次元と見なすべきです。これらの要因の多くは、ビジュアライゼーションの制作ではなく消費に関連しているため、制作者の手に負えません。言い換えれば、ビジュアライゼーションが効果的であるかどうかは、誰が、いつ、どこで、どのようにアクセスするかに大きく依存します。残念ながら、私たちの研究は、普遍的に効果的なビジュアライゼーションを生成するためのチェックリストを示すものではありません。 しかし、アクセス可能で効果的なデータ・ビジュアライゼーションを求めるのであれば、データ可視化に携わるジャーナリストが今回の調査結果に取り組むことが重要です。

脚注

1. seeingdata.org, indvil.org

2. 本章の後半は、オンラインで入手可能な長い論文を要約したものである: Kennedy, H., Hill, R. L., Allen, W., &Kirk, A. (2016). Engaging with (big) data visualizations: Factors that affect engagement and resulting new de nitions of effectiveness. *First Monday*, *21*(11). doi.org/10.5210/fm.v21i11.6389

3. データ・ビジュアライゼーションにおける感情の役割については、以下を参照: Kennedy, H., & Hill, R. L. (2018). Thefeeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology*, *52*(4), 830–848. doi.org/10.1177/0038038516674675

参考文献

Appelgren, E. (2017). An illusion of interactivity: The paternalistic side of data journalism. *Journalism Practice*, *12*(3). https://www.tandfonline.com/do...

Baack, S. (2018). *Knowing what counts: How journalists and civic technologists use and imagine data* [Doctoral dissertation], University of Groningen. www.rug.nl/research/portal/ les/56718534/Complete_thesis.pdf

Bloch, M., Byron, L., Carter, S., & Cox, A. (2008, February 23). Ebb and ow of movies: Box of ce receipts 1986–

2008. *The New York Times*. archive.nytimes.com/www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html?_r=1

Bradshaw, P. (2017, September 14). No, I'm not abandoning the term "storytelling," Alberto—Just the opposite (and here's why). *Online Journalism Blog*. onlinejournalismblog.com/2017/09/14/narrative-storytelling-data-journalismalberto-cairo/

Burmester, M., Mast, M., Tille, R., & Weber, W. (2010). How users perceive and use interactive information graphics:

An exploratory study. In E. Banissi (Ed.), *Proceedings of the 14th International Conference Information Visualisation* (pp. 361–368). doi.org/10.1109/IV.2010.57

Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds.

American Journal of Sociology, *113*(1), 1–40. doi.org/10.1086/517897

Rusbridger, A. (2010, July). *Why journalism matters*. Media Standards Trust Series, British Academy.

データをスケッチする

文：モナ・チャラビ、ジョナサン・グレイ

概要

著名なデータジャーナリストであるモナ・チャラビのインタビューでは、データに親近感を持たせる方法としてのスケッチという手法の開発と評価を探っています。文脈を提供する手段としてのデータ、物事を比較可能にする視覚的実践、彼女の作品におけるユーモアとアナロジーの役割、社会批評としてのデータジャーナリズム、データの不確実性と分析の暫定性を伝えることの重要性についての議論を含んでいます。

キーワード：データスケッチ、データ・ビジュアライゼーション、不確実性、データ公開、データジャーナリズム、ビジュアルプラクティス

ジョナサン・グレイ（JG）: データで使ったスケッチを始めたきっかけは?

モナ・チャラビ （MC）: 私が FiveThirtyEight で働いていたとき、彼らは私のような読者を対象としていないと感じていました。複雑なインタラクション機能を駆使して、別の種類の読者に対応していたのです。私はデータを使ってスケッチを始めました。それは、机に座ってできることでした。始めてみると、データプロジェクトの不確実性を伝えるのに非常に効果的な方法であることがわかりました。設計上の決定はすべて人間が行っていることを人々に思い出させることができるし、誰もができるような方法でデータとコミュニケーションを取ることは非常に民主的だと思います。私は以前、ガーディアンで DIY コラムを書いていましたが、私のプロセスのすべてのステップを人々に紹介していました。ジャーナリストとして、どこでデータを見つけたか、どのように処理したか、何をしたかを読者に伝えるだけでなく、再現できるようにすることで、読者との間の壁を取り除き、新しい種類のアクセシビリティ、参加、読者との関係を生み出すことができるのは、とても楽しいことでした。

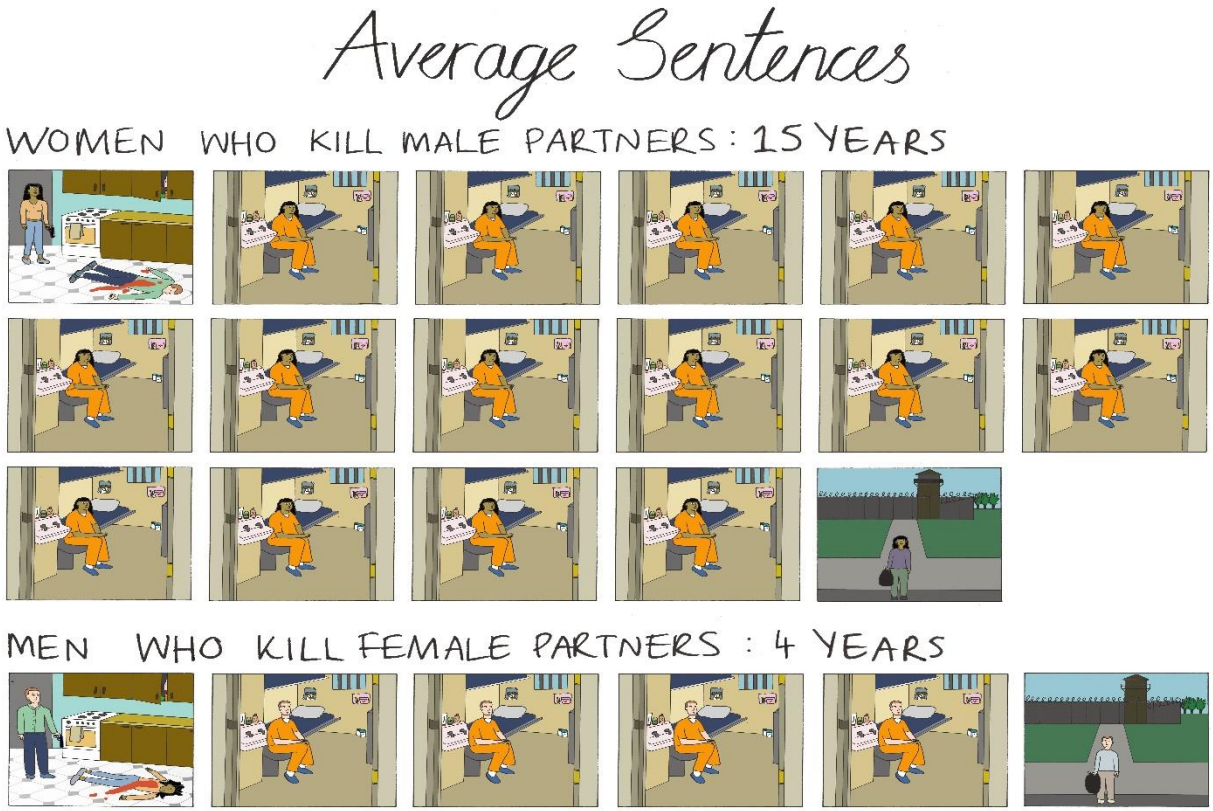


Figure 25.1. モナ・チャラビのイラスト「平均刑期」出典：ガーディアン https://www.theguardian.com/news/datablog/2019/jan/12/intimate-partner-violence-gender-gap-cyntoia-brown

JG：この本では、データジャーナリストが、既存の専門知識（データサイエンスや高度な統計手法など）を反映・強化するだけではなく、他の種類のデータ・プラクティスやデータ文化を促進する方法も探っています。あなたの仕事は、データの扱い方やデータとの新たな関わり方を模索することでもあるのでしょうか？

MC：私には高度な統計学のスキルはありません。私がデータの分析を始めるときは、他の人でも再現できるような比較的簡単な計算をすることが多いですね。そうすることで、データの信頼性を高めることができるからです。他の高度な統計的アプローチでは、ある時点で読者に最後通牒を突きつけることになります：ジャーナリストの分析を信じるか、信じないか。これは、政府の統計や基本的な掛け算を信用するかしないかという命題とは異なります。単純な計算で物事を進めることに

は一定のメリットがあります。これが私の仕事やアプローチの大きな特徴です。データを使って二つのことができます。それは「ズームイン」と「ズームアウト」です。

データジャーナリストとしての私の責任は、ある特定の事件からズームアウトして、データを使って読者にコンテキストを与えることです。例えば、ある事件や攻撃があったとして、どのようにして起こるのか、どこで起こるのか、時間の経過とともにその頻度は増えるのか、他の人よりも狙われやすい人はいるのかといったことを紹介することがあります。読者にとって、より広範なトレンドを理解する、とても有益な情報となります。もしかしたら、ニュースに対してパニックにならずに済んだり、[正常化バイアスに屈せず]きちんと恐れたりするのに役立つかもしれません。

一方、まったく逆にズームインすることもできます。例えば、BLS(米国労働統計局)が失業率のデータを発表していて、他のほとんどの報道機関が既にそれを発表していたとします。私たちデータジャーナリストは、ズームインすることができます。全国の雇用率だけでなく、女性はこう、男性はこう、年齢層はこう、人種や民族別に見るとこうなる、と読者に伝えることができるのです。こうして、読者はデータをより詳しく調べることができます。私の仕事は、この二つのモードを行き来しています。私は FiveThirtyEight のようなメディアに対し、作品が「我々には何ができるか」といった知的な誇張表現になってしまうと批判することがあります。

私はそのようなことはしません。私の目的は、読者、特にギークを自認する白人男性だけでなく、幅広い読者のコミュニティに貢献することです。FiveThirtyEight の読者とジャーナリストはともに、自分をギークと呼びます。しかし、私がジャーナリズムの世界に入ったのはその[人々のみに尽くす]ためではありません。

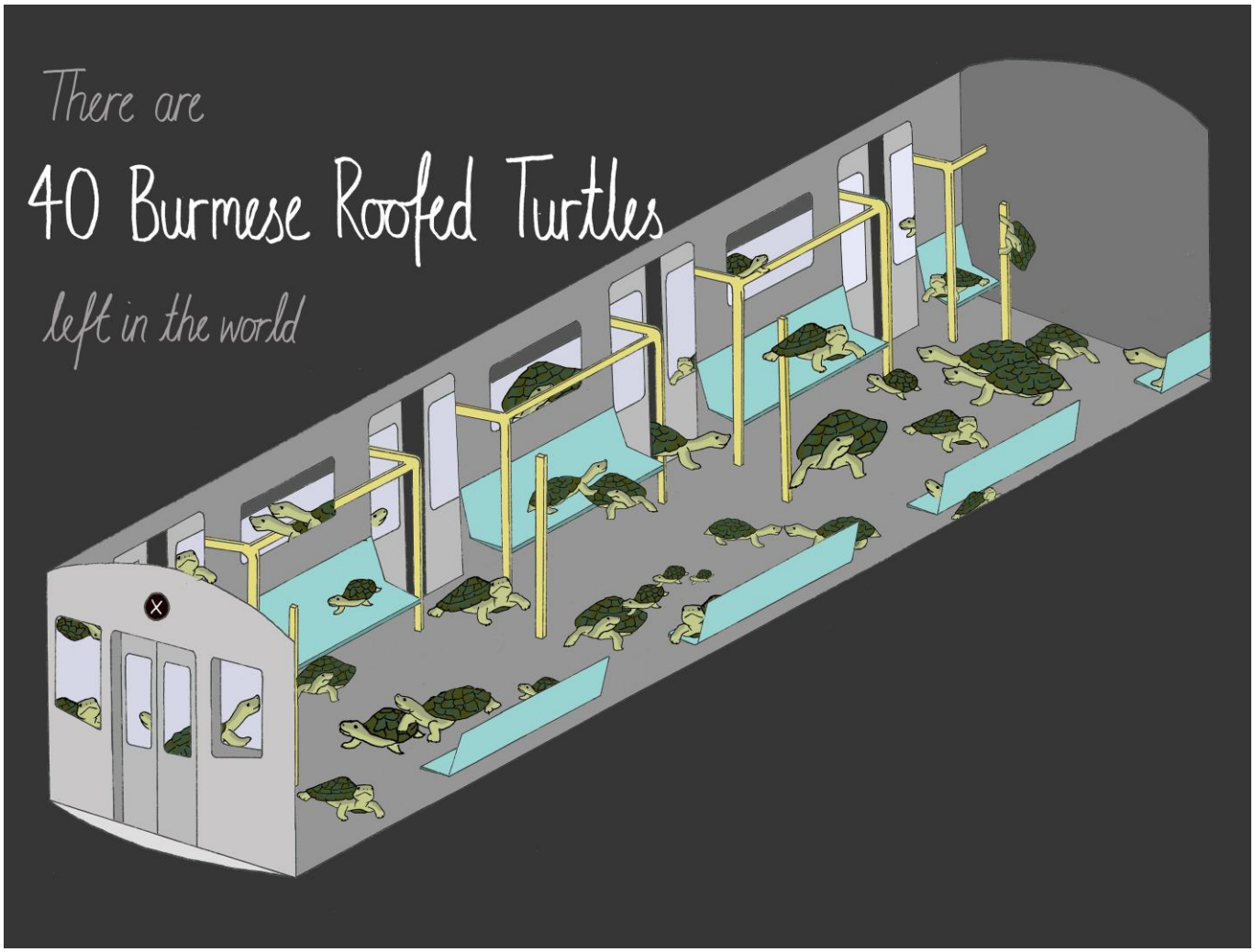


Figure 25.2. モナ・チャラビのイラスト「世界には約40匹のビルマオオセタカガメが残存している」 出典：ガーディアン

https://www.theguardian.com/environment/gallery/2018/sep/17/endangered-species-on-a-train

JG：あなたの最近の作品で、The Guardian に掲載された「Endangered Species on a Train［電車に乗った絶滅危惧種］」について、もう少し詳しく教えてください。どのようにしてこのテーマに興味を持ち、どのようにプロジェクトを立ち上げ、どのようにアプローチしたのでしょうか？

MC：とても風変わりな作品でした。ニュースに触発されたわけではなく、個人的にイラストを描いていて、もう少し野心的なことをしてみたいと思ったのがきっかけです。これらを描き始めた理由の一つは、それらが実に効率的だからです。短納期で、必要ならば数時間で仕上げることができます。私は、もう少し時間をかけて大きなものを作りたいと思いました。まず、人々がすでに親しみを感じている絶滅危惧種のテーマで、既存のビジュアル言語では少し物足りなさを感じていそうなものを選びました。IUCN（国際自然保護連合）のレッドリスト1からデータを取り、絶滅危惧種に関する数字の大半は［推測による］ばらつきがあるので、それぞれの中間値を選びました。一歩下がっ

て、私のイラストをチャートとして見てください。チャートになっているのは、スケール感だけです。私が投稿するイラストにはすべてスケール感があり、それはすべてのチャートに共通しています。

問題点の一つは、国や場所によって尺度が異なることです。たとえば、英国ではミリメートル、米国ではインチです。尺度は人によって意味が異なります。データジャーナリストの多くは、このことを見失っています。誰かにとって「100 万」とは何なのか？誰かにとって「1」とは何を意味するのでしょうか？これらはすべて文脈に依存します。数値が低ければ、次のことを理解しやすくなります。「27」が何を意味するかはご存知でしょう。しかし、それは何を表わそうとしていますか?

データ・ビジュアライゼーションは、物事をより直感的に感じさせることができます。お勧めのもう一つのイラストは、平均的な駐車スペースと平均的な独房を比較したものです（figure25.3）。これは、ジャーナリズムで数字を扱うときの常とう手段です。「ロンドンの銀行員はこれだけ稼いでいる」ではなく、「ロンドンの銀行員はソーシャルワーカーの 7000 倍稼いでいる」とします。類推はどれも役に立ちます。

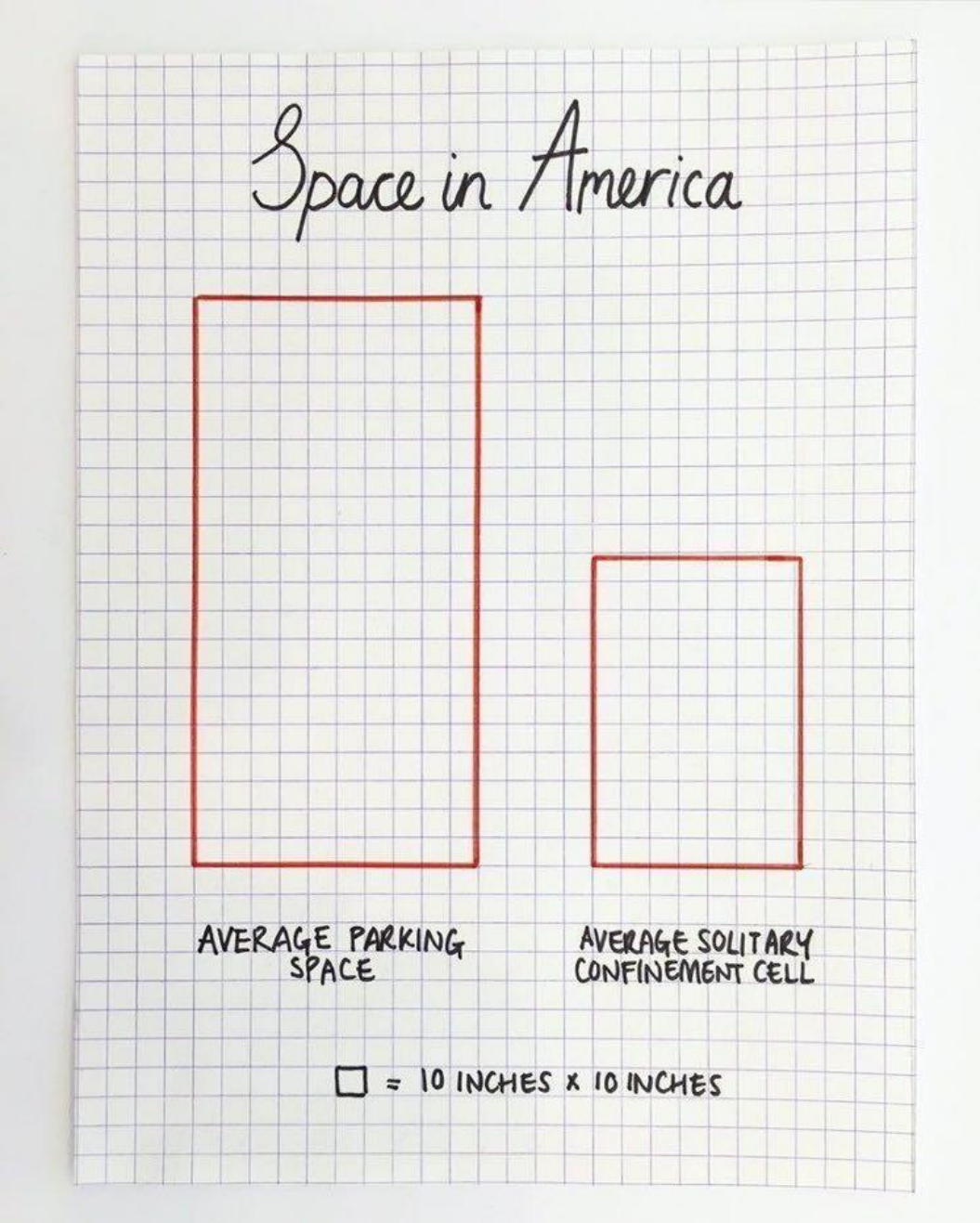


Figure 25.3. イラスト「Space in America」出典：モナ・チャラビ
https://www.instagram.com/p/ BEi-v3tKvBZ/

JG：あなたの活動は、（身近なものと見慣れぬものといった）異なる要素を並置することにも関係しているようですね。ここにはキュレーションの要素もあるのでしょうか？

MC: 私の作品にはユーモアも重要な役割を果たしています。作品が面白いというわけではなく、皮肉を取り入れた作風が多いのです。最高のコメディは、基本的に「こりゃめちゃくちゃだよ」と言っているのです。また、常に何らかの社会的な論評がなされています。データジャーナリズムにそのような要素を少しでも加えることができれば、とても力強いものになります。

JG：「絶滅危惧種」の例に戻りますが、ユーモアで数字に親しみやすくし、視覚的なスペースを使って比較するために、（グラフではなく）客車を描くことにしたのですか？

MC：最初に客車を描いてから、それぞれの動物を 7、8 体ほど描きました。それを Photoshop でレイヤー分けして、色をつけたり、数を数えたりしました。間違いがないように、それぞれの動物は別のレイヤーになっています。私が最初に考えたのは、絶滅危惧種の動物たちを、普遍的で親しみやすいさまざまなものの中に描くことでした。ニューヨークの地下鉄は完璧ではありませんが（ロンドンの地下鉄より大きいのか、それとも小さいのでしょうか）、スケール感を出すには十分です。まず、絶滅危惧種と親しみやすい空間を組み合わせたさまざまな可能性について、スプレッドシートを作成しました。例えば、サメはプールの中に泳がせようと考えていました。しかし、さまざまな空間を頭の中で整理するのは少し難しく、描き分けるととても煩雑だと気づきました。それぞれの場所に描くのではなく、同じ場所に描くことで、より効果的になります。本当は完璧ではありません。すべてのサイをこのスケールに収めるのは、ちょっと疑問です（多くのサイが大人ではなく赤ちゃんでなければならないでしょう！）。しかし、これを見れば、数字について何かを感じることができます。そして、その欠点についても透明性があります。FiveThirtyEight が作成したチャートを見て、特に専門家ではない人は、それがどの程度正確なのか把握できるとは思えません。読者は、「私たちを信じるか、信じないか」という最後通告をされているだけなのです。読者は、絶滅危惧種のイラストを見て、サイを見て、「ちょっと違うけど、わかる」と思うことができます。読者はコンピュータグラフィックスに対してはできない批評ができるのです。

JG：先ほど、あなたは、人々のデータへの関わり方を自分の作品が民主化してくれることを望んでいると言いましたね。もう少し詳しく教えてください。

MC：読者がデータの意味を理解し、自分で判断する能力がなければ、ジャーナリストは政治家と変わらないのではないでしょうか？右翼紙も左翼紙も、「我々を信じるか信じないかだ」と言っているだけです。しかし、私たちは人々が日常生活の中で十分な情報に基づいた判断を下せる力を与えるべきだと考えています。人々に力を与えるということは、単に「これが事実であり、あなたがこれを実行しなければならないのは明らかだ」と言うことではありません。「これが事実であり、私たちはこうしてここにたどり着きました」ということこそ人々を力づけるのです。ジャーナリズムだけではなく、医学の分野でもやるべきことはたくさんあると思います。私は、医療用包装をどのように変えていくかということにもっと取り組んでいきたいと思っています。箱に「こうしなさい」と書かれているのではなく、本当に良い医者になるためには、患者さんに「この薬にはこんなリスクがあります」と言えるようにすべきです。「この薬を飲まないと、こういうリスクがあります。この薬を飲まない場合のリスクはこうです」と患者に言えるようにしておくべきです。そうすれば、患者は自分で判断することが出来るようになるでしょう。私は、優れたデータ・ビジュアライゼーションは不確実性を伝えるべきだと考えています2。不確実性は、あなたの人生において、十分な情報を得た上で意思決定を行

うという、ストーリーの一部なのです。しかし、不確実性を伝えることに時間を割くデータジャーナリストはほとんどいません。ギークではないコミュニティへのアプローチに時間を割くデータジャーナリストもほとんどいません。統計や計算のスキルを持っていないからといって、あなたが賢くないとか、この情報を理解する権利がないということになるのでしょうか？もちろん、そんなことはありません。

しかし、データジャーナリストの中には、このような用語を「毎回説明するのは面倒だ」と安易な気持ちで使っている人がいます。愚かなことです。私のデータジャーナリズムへのアプローチは、賢くなるために特定の語彙や専門知識は必ずしも必要ではないという考えに基づいています。

JG：何が重要かを決めるのに、人々が参加するという要素が影響するのでしょうか？

MC:「Dear Mona」というコラムを始めた理由の一つは、質問をしてもらうためです。インスタグラムでは、皆さんが気になることをどんどん DM で送ってくださるのですが、その中には、私がまったく思いつかないようなこともたくさんあります。

例えば、メンタルヘルスと銃規制の関係を見てみると、心の問題を抱えている人に烙印を押しかねず、本末転倒になるのでやらないという選択もできます。しかし、このことについて知りたいという人からたくさんの DM を受け取ると、複雑だからといって避けて通るのではなく、正面から取り組むべきなのではないかと考えてしまいます。だから私は、読者が何を大切にしているのか教えてくれることを常に期待しているのです。私はこれがジャーナリズムの責任の放棄だとは思いません。これはジャーナリズムの民主的な役割の一部であり、人々はあらゆる面で最終的な製品ができるまで関与するのだと考えます。作る過程は理解する過程であり、「理解できるかできないか」と与えられる物ではないのです。

JG：あなたの作品への反応について、少し教えてください。予想外の反応や注目すべき反応はありましたか？

MC：様々な反響があります。ある人はテーマに注目します。賃金格差をテーマにすると、「いや、黒人女性は労働時間が短いから収入が少ないだけだ」と言う白人男性がたくさんいますが、イラストはフルタイム労働者同士の「似たもの同士」の比較に基づいていて、もし職業上の格差［例えば上級管理職］があれば、それも問題の一部であることを説明しなければなりません。私はいつも、まず批判に注目することにしています。しかし、全体的には批判よりも支持されることの方が多いですね。ときには、私が意見を述べる前から、コメントで批評に応えてくれる人もいます。また、イラストに描かれている人が、「いや、私の個人的な経験で証明されている」と言ってくれることもあります。さらにデータを見たいと言われることもあります。多くの学生が「これをやりたい」と書いてきます（面白いことに、男性よりも女性の方が多く書いてきます）。多くの NGO や慈善団体が、自分たちのデータについて考えるものではなく、何かを感じさせるものを求めて手紙をくれますが、私の

作品がそれを実現することもあります。私の作品には、アメリカの法案に引用されたものもあります。ソーシャルメディア上で、必ずしもデータジャーナリズムそのものに関心のない多くの人々に閲覧、共有されています。これは新しい読者の目に触れるということです。バーニー・サンダースは私の銃乱射事件のイラストをシェアし、マイリー・サイラス、モデルのイマン、公民権活動家のショーン・キングも同じくシェアしました。これらの人々は、面識はなく、私の作品をフォローしているわけでもありませんが、他の人がシェアしているのを見て、自然に目に留めたのだと思います。人々が私の作品に興味を持ってくれるのは素晴らしいことです。有名な誰かがシェアしてくれると、ひとりでに拡散しますからね。

本章で紹介した作品は、Web サイト（monachalabi.com）とインスタグラム（@monachalabi）で見ることができます。

脚注

1. www.iucnredlist.org

2. 編集部より。本書におけるアンダーソンの章、および彼の著書「確実性の使徒」も参照されたい：*Data Journalism and the Politics of Doubt*（Oxford University Press, 2018）.

データ・ビジュアライゼーションの媒体としての Web

文：エリオット・ベントレー

概要

インタラクティブなデータ・ビジュアライゼーション、ゲーム、バーチャルリアリティ（VR）など、Web が可能とするグラフィックスの種類を探ります。

キーワード：インタラクティブ・グラフィックス、データ・ビジュアライゼーション、ウェブ開発、JavaScript、インフォグラフィックス、ニュースゲーム

すべてのメディアが同じように作られているわけではありません。例えば、20 話からなるテレビシリーズは、2 時間の映画とは違ったストーリーを伝えることができます。同じように、地味な Web ページにも、独自のデータ・ビジュアライゼーションの可能性が開けています。Web は元来、単純でハイパーリンクが張られたドキュメント向けに設計されたもので、ほとんどがテキストと静止画像で構成されています。しかし、JavaScript が追加され、新しい機能やツールが徐々に追加されたことで、扱うことのできるパレットが広がりました1。

伝統的なデータ・ビジュアライゼーションの理論や技術（エドワード・タフテやジャック・ベルタンら）は、Web 上のグラフィックにもほぼあてはまりますが、Web 固有の特徴は、新しい形のデータジャーナリズムに大きな可能性をもたらします。これらの作品はしばしば「インタラクティブ」と呼ばれていますが、この言葉は Web のユニークな強みをわかりにくくしています。以下は、Web 上のグラフィックがホスト媒体の利点を生かすいくつかの方法を示したリストです。

膨大でわかりやすいデータセット

「インタラクティブ」の典型例は、巨大なデータセットを読者に提示し、読者が「飛び込んで」好きなだけ探索できるようにすることです。これは、巨大な表の形をとることもあれば、大きな地図の形をとることもあります。この形式は、読者が自分で面白い部分を見つけることを期待しているため、最近では軽視されがちですが、データが十分に魅力的であれば、まだまだ使いようがあります。最も成功しているバージョンは、（記事ではなく）単なるツールであるという事実を受け入れているものでしょう2。例えば、大人気のウォール・ストリート・ジャーナルの米大学ランキングやプロパブリカの公共サービスのニュースアプリなどです。

複雑なチャートを読み解く

現在の一般的なフォーマットは、まず 1 枚のチャートを用意し、それを操作することで、データセットを完全に把握できるようになっています。これは、スクロールテリングと非常に相性がよく、チャートのすべての要素を一度に表示するのに十分なスペースがないモバイルでは、特に有効です3。今や古典となった「A Visual Introduction to Machine Learning」(figure26.1)では、同じデータポイントが複数のチャート形式に切り替わり、読者が機械学習アルゴリズムによる分類を把握できるようになっています 4。もう一つの好例が、Vox の作品「100 years of tax brackets, in one chart」で、他の方法では圧倒されてしまうような[100 年に及ぶ税制優遇措置に関する]データセットを拡大・縮小して提示しています。

最新のライブデータ

グラフに最新の数字が使えるのに、なぜ静的なデータセットで妥協する必要があるのでしょうか。選挙、スポーツ中継、天気予報、金融データなどは、リアルタイムで表示するのに適したライブデータの源です。

さらに素晴らしいのは、これらのライブデータに面白い方法でコンテキストを与えることです。例えば、現在の原油価格からどの国が恩恵を受けているかを示すことができます(figure26.2)。

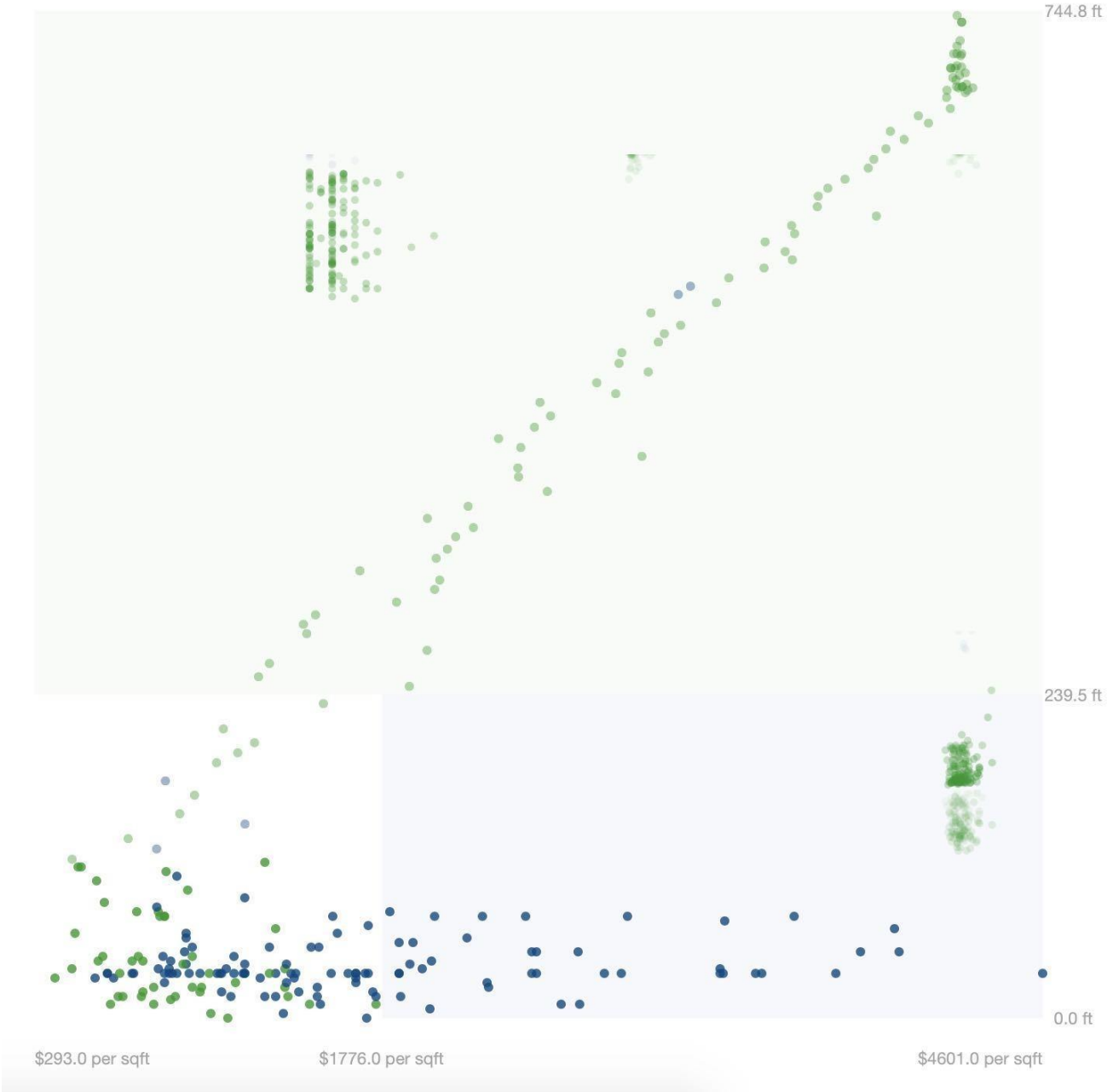


Figure 26.1. A visual introduction to machine learning. 出典: R2D3. http://www.r2d3.us/ visual-intro-to-machine-learning-part-1/

短期間の実験的なポップデータ・ジャーナリズムとして登場した Ampp3d は、ライブカウンターを使って、英国に入国する移民の数やサッカー選手のウェイン・ルーニーの収入など、興味深い方法で数字に命を吹き込みました6。残念ながら、これらはその後オフラインになってしまいましたがね。

読者をデータセットに招き入れる

「巨大なデータセット」というアイデアで使えるもう一つの工夫は、いくつかの個人情報を尋ねることで、読者がデータセットのどこに位置するかを示すことです。ニューヨーク・タイムズの 2013 年の方言クイズマップ（figure26.3）は、12 月 20 日に公開されたにもかかわらず、その年で最も人気のある記事となったことで有名です。英国の BBC は、公共サービスの一環として、「UK fat scale calculator［肥満度計算機］」のような記事を頻繁に掲載しているようです。異なる文化圏の人々がどのように円を描くのかを紹介した Quartz の記事は、読者に円を描くよう求めることから始まり、退屈になりそうだった特集の魅力的な導入部となっています 8。

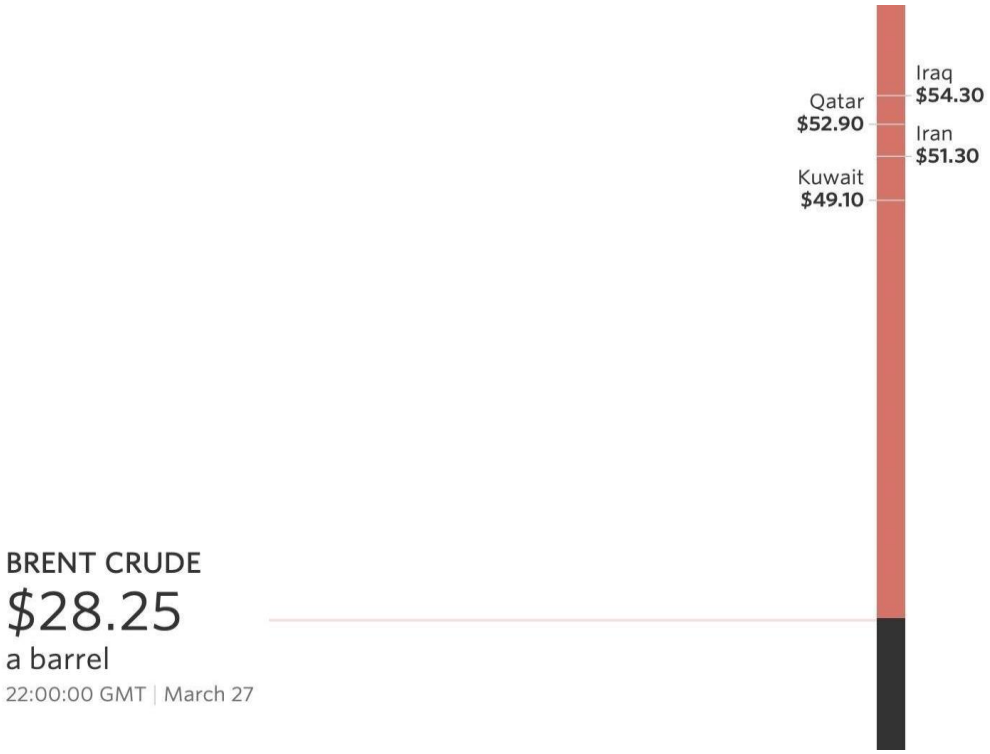


Figure 26.2. 現在の原油価格から利益を得ている国 出典：ウォール・ストリート・ジャーナル http://graphics.wsj.com/oil-producersbreak-even-prices/

オリジナルなデータセットの収集

前述のカテゴリーから一歩進み、読者からデータを受け取って即座に反応するだけでなく、さらなる分析のために新たなデータセットを作成するプロジェクトもあります。オーストラリア放送協会は、政治学者と共同で「Vote Compass」を作成し、読者が自分の政治的立場を理解できるようにし、そのデータに基づいて一連の記事を書きました 9。

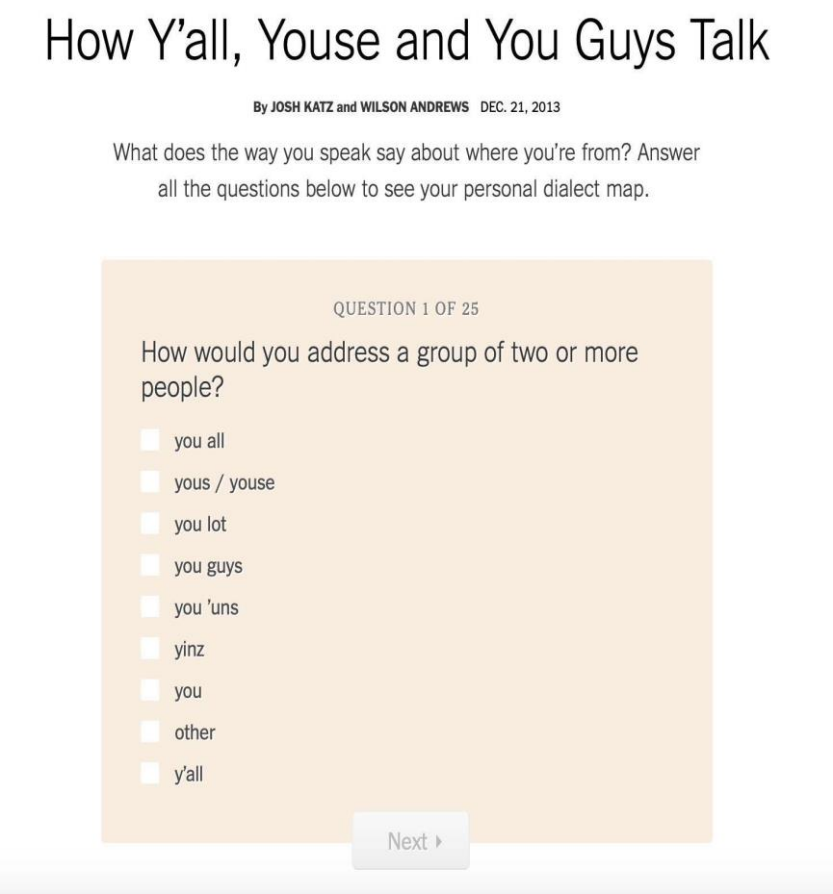


Figure 26.3: ニューヨーク・タイムズの方言クイズマップ(2013年) 出典:ニューヨーク・タイムズ https://www. nytimes.com/interactive/2014/upshot/dialect-quiz-map.html

さらに最近では、ニューヨーク・タイムズが同じアイデアをよりソフトなテーマで使用し、読者に「ゲーム・オブ・スローンズ」のキャラクターを評価してもらい、その結果をライブチャートにプロットしました(figure26.4)。

無限のキャンバス

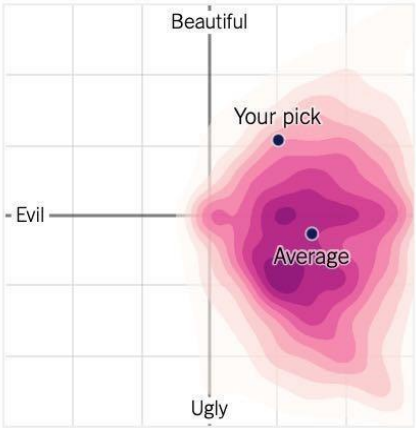
Web の範囲と容量は無限であり、より具体的にいえば、Web ページの幅や高さは自由に変更できます。つまり、作業の場としての 「無限のキャンバス」なのです。私はこの言葉を[「マンガ学」の著作がある]アーティストのスコット・マクラウドから借用しました。彼は「長編コミックをオンラインに移行する際に、ページを分割しなければならない理由はない」と主張しています10。実際、私たちのグラフィックが紙の限界に制約される必要はないのではないでしょう。ワシントン・ポストの「The Depth of the Problem」では、行方不明になった MH370 便を捜索している海域の深さを示すために、高さ 16K ピクセルのグラフィックが使われています(figure26.5)。 この情報は[紙面の]1ページに収めることもできたでしょうが、この非常に背の高いグラフィックの詳細さと感情的

なインパクトには及ばなかったことでしょう。ガーディアンの「How the List Tallies Europe's Migrant Bodycount」では、何万人もの移民の死が、ページをスクロールするたびに一つ一つの点として表示され、力強く表現されています 12。

Your Estimates, and Everyone Else's

Tyrion Lannister

Introduced as "The Imp," Tyrion was a hard-drinking, whoring, black sheep of the royal family; now, he's a respected strategist.



Daenerys Targaryen

A stunning beauty who feels for the oppressed. But she has no problem sending people to their deaths as she conquers kingdoms.

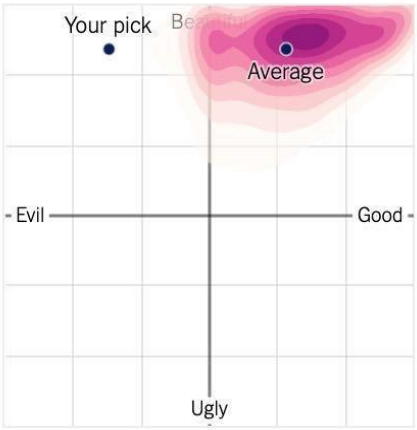


Figure 26.4: ゲーム・オブ・スローンズの登場人物評価表 出典：The Upshot. https://www.nytimes.com/interactive/2017/08/09/upshot/game-of-thrones-chart.html

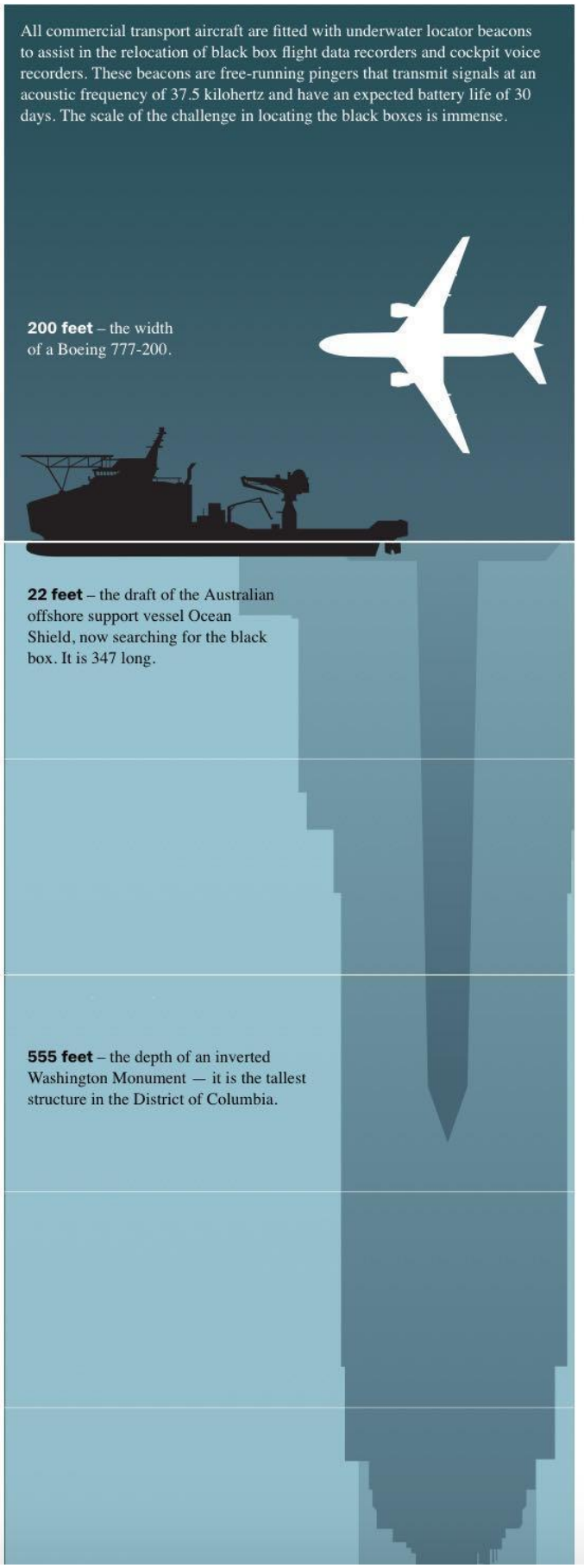


Figure 26.5. 行方不明の MH370 便を捜索している海域の深さを示すグラフィック　出典：ワシントン・ポスト http://apps.washingtonpost.com/g/page/ world/the-depth-of-the-problem/931/

データ駆動型ゲーム

テレビゲームのようにニュースのテーマを探求するインタラクティブな「ニュースゲーム」は、しばらく前に登場し、さまざまな成功を収めています。The Upshot の「You Draw It」シリーズ（figure26.6）は、読者に空白のグラフを埋めるよう求め、その後、答えを明らかにしてそのテーマをより深く掘り下げていくものです。

ゲームの中にはもっと複雑なものもあって、それには BBC の資金調達方法など、現実世界の問題を簡略化して読者に解いてもらい、その難しさを証明するようなものがあります 13。

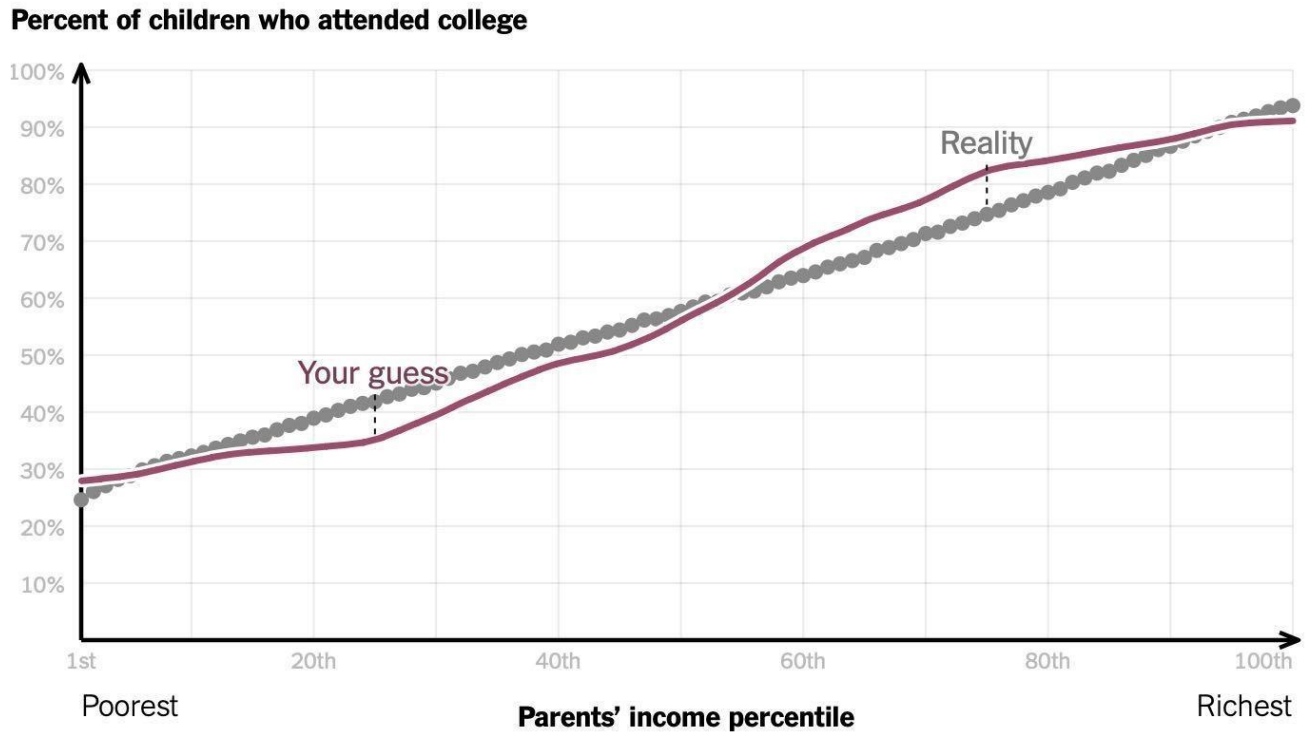


Figure 26.6. The Upshot の「You Draw It」シリーズより　出典：The Upshot. https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how- family-income-affects-childrens-college-chances.html

これらは、読者に表面的な情報を提供するだけのおもちゃとも言えますが、うまくやれば、ありふれたテーマに新鮮な視点を与えることができます。FiveThirtyEight の「How to Win a Trade War」では、読者が取引戦略を選択し、過去にこのページを訪れた人と競い合うことで、無味乾燥な印象を与えがちな経済理論に命を吹き込んでいます 14。

ライブで無作為に行われる実験

関連するフォーマットとして、読者がブラウザ上でライブシミュレーションを実行できるようにする方法があります。単なるアニメーションの説明ではなく、ある程度のランダム性を持たせることで、毎回ユニークな結果が得られ、抽象的な統計的確率を現実のものにする最適な方法です。figure26.7 のガーディアンの作品では、ワクチン接種率が異なる 10 の集団で麻疹の発生をシミュレーションしています。Web グラフィックは、パーセント表示だけでは伝えきれない結果を明確に示しています。

ネイサン・ヤウの「Years You Have Left to Live, Probably」では、単純な折れ線グラフ(「来年まで生きる確率」)が、ランダムに死んでいく「命」を積み重ねることで、より切実なものになっています15。このようなシミュレーションは、架空のデータを使う必要はありません。「The Birthday Paradox」では、過去にこのページを訪れた人のデータを使って、誕生日を共有する確率を検証しています 16。

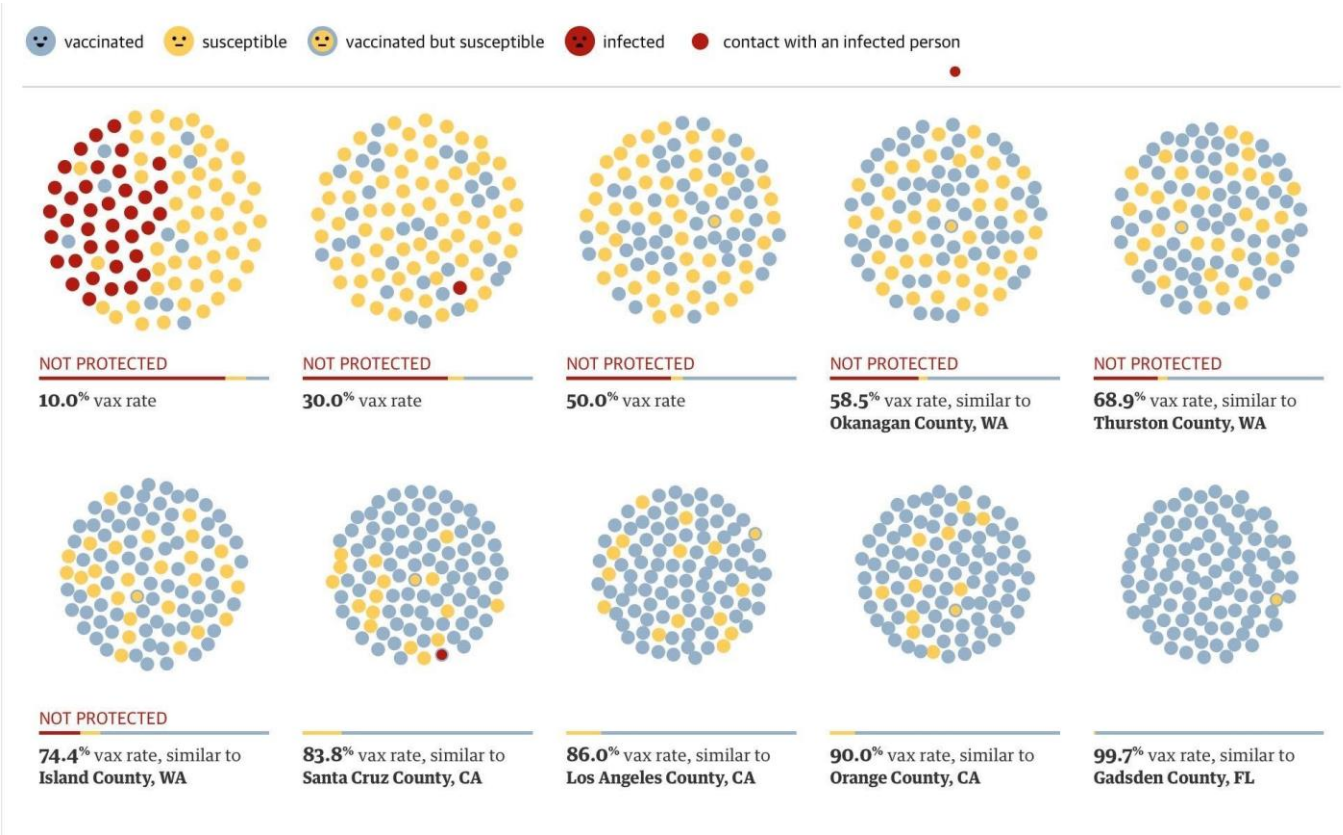


Figure 26.7: ワクチン接種率が異なる 10 の集団で麻疹が発生した場合のシミュレーション　出典:ガーディアン https://www.theguardian.com/society/ng-interactive/2015/ feb/05/-sp-watch-how-measles-outbreak-spreads-when-kids-getvaccinated

3D、VR と AR

3D グラフィックスとバーチャルリアリティは、地形図を除き、データジャーナリズムに活用するのは困難です。2015 年に行われた、二つの注目すべき実験「Is the Nasdaq in Another Bubble?」と「A 3-D View of a Chart That Predicts the Economic Future: The Yield Curve」はともに財務データに関するもので、巧妙な目新しさはあったものの三次元チャートの爆発的な普及には至りませんでした17。おそらくこれが最高到達点でしょう。現実世界のカメラ映像にグラフィックを重ねる拡張現実の可能性は、まだ開拓されていません。

おわりに：新しいフォーマットが生まれるまで

上記の Web グラフィックの一部は、ここ数年で登場したばかりの新しい形式です。中には、複雑なチャートのガイド（主にスクローリーテリングのインタラクティブなパターンを使用）のように、定着したものもありますし、三次元チャートのように一時的な成功で終わるものもありました。しかし、Web 上でどのようなグラフィックが流行するかを決めるのは、好みだけではありません。テクノロジーの進歩や読者の消費傾向によっても、トレンドは変わってきます。たとえば、広く使用されているインタラクティブ・マップを考えてみましょう。

見た目の美しさや理解のしやすさに加え、「Google Maps」や「Leaflet」など、簡単に作成・操作できるツールが普及したことが、普及と親しみやすさを後押ししたことは間違いありません。しかし、確たるデータがあるわけではありませんが、最近ではインタラクティブ・マップの公開数が減っているように感じられます。

この傾向を、インタラクティブな機能（あるいは地図そのもの）が不要だとする認識がジャーナリストの間で高まっているからだと考えるのは簡単ですが、新しいテクノロジーもこの減少に貢献していると思われます。現在、読者の多くは携帯電話で Web にアクセスしており、インタラクティブな地図は小さなタッチスクリーンでは使いにくいのです。さらに、多くの点で優れた新しい技術があります。ai2html は、Adobe Illustrator のファイルからレスポンシブ HTML を生成するスクリプトで、ニューヨーク・タイムズがオープンソースで提供しています。これで作られた地図は、従来の地図製作者の技術を活かしつつ、機械可読性のあるシャープなテキストを実現しています。

このような地図にインタラクティブ性がないことは、多くの点で制限があるとはいえ、恩恵が勝ります。これは、データジャーナリストが Web のユニークな機能をどのように利用すべきかを示す一例です。たくさんの可能性がある中で、それらを慎重に評価し、本当に必要なときを見極めて使うことが大切です。

ニュース・グラフィックの四つの最新動向

文：グレゴール・アイシュ、リサ・シャーロット・ロスト

概要

ここでは、最近のニュース・グラフィックに見られる四つの動きを紹介します。すなわち、「モバイルファースト」の重要性の高まり、インタラクティブ性の重要性の変化、より多くの（自社製）チャートツールの開発、そしてデータ中心のオンライン出版物の増加です。

キーワード：ニュース・グラフィック、モバイル、チャートツール、インタラクティブ性、データ・ビジュアライゼーション、データジャーナリズム

ニュースグラフィックスの分野はまだ歴史が浅く、次のような問いに答えようとしているところです。［世論調査の］データの偏りや不確実性をどのように示すのか？（Cairo & Schlossberg, 2019）。記者とどのように協力すればいいのか？テンポの速いソーシャルメディアで複雑なデータをどのように伝えるか？（Segger, 2018）ここでは、今後数年間影響を及ぼすと思われる四つの重要な動きを取り上げてみます。

「モバイルファースト」が本格的に動き出す

「モバイルファースト」という言葉は広く使われていますが、目まぐるしく変化するニュース・グラフィックの世界では、モバイルエクスペリエンスはしばしば後回しにされてきました。今では、ようやく優先順位が上がってきました。これには二つの意味があります。まず、グラフィックをモバイルで動作させるために、より多くの配慮がなされるようになりました。モバイルユーザーに「このエクスペリエンスはデスクトップでの使用が最適です」とただし書きを表示するのはまずいでしょう。ユーザーの半数以上が離れてしまわないように、図表はレスポンシブでなければなりません。

しかし、モバイルの数ピクセルの枠の中で考えることは、グラフィック担当記者にとってはフラストレーションが溜まるものです。なにせ彼らの多くは、紙の新聞でページ全体を埋め尽くしたり、デスクトップの全画面表示をデザインしたりする「贅沢」に慣れているのですから。一方、小さな画面の限界が、グラフィック担当記者に既成概念にとらわれない創造的な発想をもたらすこともあります。例えば、フィナンシャル・タイムズは、国会の議席数のグラフを 90 度回転させ、新しいタイプのチャートを作成しました1。モバイルファーストのデータ・ビジュアライゼーションがもたらす二つ目の結果は、ニュース開発者や記者が、「モバイル」を単に小さな画面としてではなく、センサーが搭載されたデバイスとして捉えるようになることです。これにより、新しいデータエクスペリエンスが可能になります。

ガーディアンは、2016 年にリオデジャネイロで開催されたマラソン大会と同じ距離のバーチャルオーディオツアーに参加できるアプリを制作しました2。（画面には）「今後 3 週間で、26.2 マイル（42.2km）の全行程を走破しましょう」と書かれています。AR や VR も同様にスマートフォンを利用したものであり、ニュースでも取り上げられるようになりました。

インタラクティブは死んだ、ただし復活もありうる

ここ数年、インタラクティブ機能がシンプルなチャートに使われることは少なくなってきています。ニュースルームが毎年発表する最大のプロジェクトにはなお使われているものの、成功のためにはもはや必須ではありません。フィナンシャル・タイムズや FiveThirtyEight、「National Geographic」などのニュースルームでは、読者にインタラクティブな操作を求めない、ネットで話題のチャートをいくつも発表しています。

インタラクティブ・グラフィックスが減少した理由は、大きく分けて二つあります。第一に、これまで想定されていたよりも、チャートと対話する人の数が減っていることです3。好奇心旺盛でインターネットに精通している人たちはグラフィック担当記者と同じように、ビジュアライゼーションにカーソルを合わせようとします。また、記者は自分の記事がより生き生きとしたものであることを望んでいます。しかし、私たちは、特にモバイルでは、受動的な消費を好む読者を想定して制作しています。そのため、多くのグラフィック担当記者は、そもそも何も隠さないことにしたのです。

二つ目は、グラフィックがニュース速報のサイクルに乗ったことです。グラフィック担当記者のビジュアライゼーション作成のスピードはどんどん上がっており、ニュース速報には、例えば、ある出来事が起きた場所を示すロケーターマップがすぐに掲載されます。しかし、よくできたインタラクティブな作品にはまだ時間がかかります。記事を早く仕上げるために、インタラクティブ性が省かれることもよくあります。

今でもインタラクティブなニュースグラフィックは見られますが、その重要性は変わってきています。ストーリーを補足するのではなく、インタラクティブそのものがストーリーになるのです。読者は居住地や収入、意見などの個人情報を入力し、それが全体の中でどのように位置づけられるかを知ることがでます。ニューヨークタイムズのニュース・グラフィック「You Draw It: How Family Income Predicts Children’s College Chances」や「Is It Better to Rent or Buy? 4」が典型例です。どちらの記事も、読者がデータを入力しなければ意味がありません。インタラクティブが価値を生み出すのです。

ニュースルームでの（自社製）チャートツールの使用が増加

記者は、記事を目立たせることをこれまで以上に強く求められています。チャートを追加することは一つの解決策ですが、グラフィックチームは増え続けるリクエストに対応するのに苦労しています。そこで、数回のクリックで簡単にチャートや地図、表を作成できるチャートツールを採用するニュースルームが増えています。チ

ャート作成ツールには二つの選択肢があります。Datawrapper や Infogram のような外部のチャートツールを使用するか、社内の要件に合わせて自社でチャートツールを構築し、コンテンツ管理システムに統合するかです。

2番目の選択肢は素晴らしいアイデアのように聞こえますが、[その戦略を採用する]多くのニュースルームは予想以上に多くのリソースを要することに気付くでしょう。外部のチャートツールは、専門のチームによって構築され、ツールのメンテナンスやトレーニングが行われます。ニュースルームでは、これらの作業をグラフィックチームやインタラクティブチームが行うことが多く、実際のニュースプロジェクトに割く時間が少なくなってしまいます。内製のチャートツールは、優先順位をつけて初めて成功するものなのです。[スイスの新聞社]Neue Zürcher Zeitung では、チャートツール Q の開発とメンテナンスに専従する3人の開発者がいます。

データ中心の出版物が、イノベーションとビジュアルリテラシーを促進する

数年前までは、データ駆動型のアプローチは個々のストーリーにのみ有効だと考えられていましたが、今では、このアイデアを基にした(そして成功した！)出版物が見られるようになりました。例えば、政治やスポーツに関する「FiveThirtyEight」、ポップカルチャーに関する「The Pudding」、人類の長期的な発展に関する「Our World in Data」などは、情報を伝える手段としてデータを活用しています。同じ分野の他の出版物との最大の違いは、読者層といえます。好奇心旺盛で、データを重視し、グラフを見ることに抵抗のない人たちです。その結果、データ中心の出版物では、線付き散布図のような難解なタイプのチャートを示すことができます。

最近の動向を紹介する章は、すぐに時代遅れになるものです。しかし、今回取り上げた四つの動きは、数年前から議論の中心となっており、今後も継続していくものと思われます。それは、日々の報道活動の中で、唯一の正解のない問いに裏打ちされているからです。「モバイルファーストでプロジェクトを設計するか、デスクトップでしか動作しない複雑なソリューションを採用するか」「このビジュアライゼーションをインタラクティブにして、(読者の 10～20%しかその機能を使わないとされているが)読者の興味を引くことに労力を費やすか」「ビジュアライゼーションをゼロから構築するか、チャートツールを使用するか」「ビジュアライゼーションは広範な読者に向けて作るか、それともデータに詳しい読者を対象とするか」

答えはニュースルーム、グラフィックチーム、プロジェクトによって異なるかもしれません。しかし、次第にモバイルファーストで非インタラクティブであり、チャートツールを使って構築された、データリテラシーの高い読者向けのビジュアライゼーションに収束していくと思われます。

脚注

1. ig.ft.com/italy-poll-tracker/

2. www.theguardian.com/sport/2016/aug/06/rio-running-app-marathon-course-riorun

3. vimeo.com/182590214, medium.com/@dominikus/the-end-of-interactive-visualizations-52c585dcafcb

4. www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-collegechances.html, www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html

参考文献

Cairo, A., & Schlossberg, T. (2019, August 29). Those hurricane maps don't mean what you think they mean. *The New York Times*. www.nytimes.com/interactive/2019/08/29/opinion/hurricane-dorian-forecast-map.html

Segger, M. (2018, June 28). Lessons for showcasing data journalism on social media. *Medium*. medium.com/severecontest/lessons-for-showcasing-data-journalism-on-social-media-17e6ed03a868

ジャーナリズムの産物としての検索可能なデータベース

文・ザラ・ラーマン、ステファン・ウェールマイヤー

概要

データジャーナリズムの調査において、公衆向けの検索可能なデータベースを使用する際の責任あるデータの在り方と透明性を探求します。

キーワード：データベース 、レスポンシブルデータ、クラウドソーシング、エンゲージメント、データジャーナリズム、透明性

検索可能なオンラインデータベースは、ジャーナリズムのフォーマットとしてはいまだに新しい部類に入ります。これは、ニュースルームからデータセットにアクセスできる Web インターフェースです。このフォーマットは目新しいものではないものの、データジャーナリズムのプロジェクトで使われる場面はまだ比較的少ないのです。この記事では、読者の生活に直接影響するトピックをカバーするものから、さらなる調査のために作成されるインターフェースまで、さまざまなタイプのデータベースを検証します。

ここでは、共著者の一人が行った［ドイツの調査報道メディア］Correctiv の「Euros für Ärzte」調査を基に、その概要を説明しています1。また、データ駆動型の調査を行った後に生データを公開することは一般的になっていますが、そのデータに対して検索可能なインターフェースを構築することはまだ普及していないことも注目に値します。ジャーナリズムにおいてデータベースを活用する利点を考えていきますが、データがどのように使用され、アクセスされ、変更され、理解されるのか、プライバシーに関連した倫理的な問題が数多くあることも忘れてはいけません。次に、こうした形でデータを使用した結果生じる、データに対する責任を巡り、情報をオンラインにする影響と、データに内在する力学を検討します。最後に、今後発展していくであろうベスト・プラクティスを紹介します。

ジャーナリスティックなデータベース

データベースは、一調査報道においてさまざまな公共的役割を果たします。パーソナライズの要素が強いデータベースとしては、プロパブリカの「Dollars for Docs」があります。このデータベースは、製薬会社や医療機器会社が医師や教育病院に支払った金額のデータをまとめたものです2。このテーマとアプローチは、Correctiv とシュピーゲル・オンラインによる「Euros für Ärzte」にも踏襲されました（以下に詳しく説明します）。どちらのアプローチも、公開された情報源からデータを集め、読者が自分で検索して、自分の主治医が支払いを受けているかどうかを確認できるように、データへのアクセス性を高めることを目的としています。いずれも取材と継続的な調査が伴うものです。同様に、ベルリナー・モルゲンポストは、親が自分の地域の学校を見

つけるのを支援する「Schul Finder」を構築しました。この場合、データベース・インターフェース自体が製品となっています3。

ニュースルームがデータを集めて準備するタイプのデータベースとは対照的に、もう一つのスタイルは、読者がデータに貢献するものです。「市民によるデータ生成」、あるいは単にクラウドソーシングと呼ばれることもあります。これは、必要なデータが公式な情報源から収集されていない場合に特に有効です。例えば、ガーディアンのクラウドソース・データベース「The Counted」は、2015、16 年にアメリカで警察に殺された人々の情報を収集しました4。このデータベースは、様々なオンライン・レポートと読者からの情報を基に作られています。別のタイプのデータベースでは、既存のデータを利用して、読者が設定した特定の基準に基づいてレポートを作成できるインターフェースを作成します。例えば、「Nauru Files」では、2013～15 年にナウルにある[難民]収容所のスタッフが書いたインシデントレポートの概要を見ることができます5。英国の[非営利団体]調査報道局は、調査によって得られたさまざまな情報源からのデータを「Drone Warfare」というデータベースにまとめています。このデータベースでは、読者が対象となる国や期間を選択して、データを要約して可視化したレポートを作成することができます6。

最後に、データベースは、ジャーナリズムの発展に役立つ、調査支援ツールとしても作成できます。国際調査報道ジャーナリスト連合は、パナマ文書、パラダイス文書、および他の調査のデータを取り込んだ「Offshore Leaks Database」を作成し、管理しています7。同様に、OCCRP（組織犯罪と汚職の報道プロジェクト）は、1900 万件以上の公文書を検索できる OCCRP データを維持・更新しています8。これらのケースでは、ツールの主な利用者は一般読者ではなく、これらのツールを使って見つけた情報を自ら調査するジャーナリストや研究者を想定しています。

ニュース製品としてのデータベースを作る際には、以下のような配慮が必要です。

読者層：直接読者を対象としたものなのか、他のジャーナリストの研究用データベースか

適時性：継続的に更新されるものか、単発の記事か

コンテキスト：調査や物語の一部なのか、データベース自体が製品なのか

インタラクティブ性：データベースを改善するために読者が積極的に意見を述べることを歓迎しているか、あるいは読者は主にデータの受け手であると考えているか

情報源：すでに公開されているデータを利用するか、データベースを通じて新しい情報を公開するか

ケーススタディ：Euros für Ärzte

欧州製薬団体連合会（EFPIA）は、33 の国の団体と 40 の製薬会社が加盟する業界団体です。2013 年、連合会は、16 年 7 月以降、会員企業が事業を展開する国の医療従事者や組織に対する支払いを公表することを決定しました（EFPIA, 2013）。 ドイツの非営利調査報道機関であるコレクティブは、プロパブリカの「Dollars for Docs」プロジェクトに触発され、ドイツの製薬企業の Web サイトからこれらの公表資料を収集し、製薬企業からの支払いを受けた人の検索可能なデータベースを作成して一般公開することを決めました。彼らは調査を「Euro für Ärzte」（「医師へ[支払われた]ユーロ」）と名付けることにしました。ドイツの全国ニュースサイト、シュピーゲル・オンラインと共同で、約 50 の Web サイトから文書とデータを収集し、異なるフォーマットから統一された表データに変換しました。そして、このデータをさらにクリーニングし、複数の企業からの金銭受領者を照合しました。

データクリーニングに要した期間は約 10 日間で、最大で 5 人が参加しました。受領者ごとに個別の URL を持つカスタムデータベースの検索インターフェースをコレクティブが設計し、公開。2017 年にも同様のプロセスでデータベースを更新しました9。コレクティブは、同じ方法論と Web インターフェースを用いて、オーストリアのデータを derstandard.at と ORF の協力で、スイスのデータを Beobachter.ch の協力で公開しました。

報道の目的は、製薬業界がイベントや組織を介して、医療従事者に及ぼす組織的な影響力と、それに伴う利益背反を浮き彫りにすることでした。検索可能なデータベースは、読者がこの話題について医師と議論を始めるきっかけになり、このようなことが起こっているという事実に注意を喚起することを目的としています。 さらに言えば、この取り組みは、自主的な情報開示ルールの不備を明らかにしました。公開条件は法的な義務ではなく、業界のイニシアチブによるものだったため、データベースは不完全なもので、法的に公開を義務付けない限り、これが変わることはないだろうと思われます。前述したように、データベースは不完全で、製薬会社から支払いを受けたことのある多くの人々がデータベースから抜け落ちていました。そのため、ユーザーが自分の担当医師を検索したときに、検索結果が空欄だった場合、その医師が支払いを受けていない場合と、公開を拒否した場合があり、二つの全く異なる結論となってしまいます。

これは、協力的で透明性の高い個人にスポットライトが当たり、より悪質なお金の流れが闇に葬られてしまうと批判されています。そこでコレクティブは、支払いを受けていない医師もデータベースに表示されるよう、オプトイン機能を提供しました。これにより、ストーリーに重要な背景を与えることができますが、検索結果には不確実性が残ります。公開後、コレクティブとシュピーゲル・オンラインの両社は、データベースに掲載された医師から数十件の苦情と法的脅迫を受けました。データは見つけるのが難しいとはいえ、公開された情報源から得たものなので、シュピーゲル・オンラインの法務チームは、ほとんどの苦情を製薬会社に委ねることにして、情報源に変更があった場合にのみデータベースを調整することにしました。

データベース構築の技術的考察

データセットを読者に提供し、利用可能にする方法を検討しているニュースルームには、データセットのサイズや複雑さ、ニュースルームの技術的能力、読者がどのようにデータと対話すべきかなど、考えるべき基準がさまざまあります。報道機関が、データベースが必要だと判断した場合、データベース構築には、オーダーメイドの開発と配備が必要となり、これには少なからぬリソースが必要となります。サードパーティのサービスを利用してデータにアクセスできるようにすれば、通常はよりシンプルで少ないリソースで済みます。

例えば、コレクティブの場合、製薬会社と約 2 万人にも上る受領者との金銭的なつながりを検索し、リストアップするため、カスタムソフトウェアソリューションが必要でした。彼らは、データベース用のソフトウェアをメインの Web サイトとは別のリポジトリで開発して、コンテンツ管理システムに接続できるようにしました。これにより、メインの Web サイトと調査セクションを視覚的、概念的に統合したのです。懸念事項を分離するために、データはコンテンツ・データベースとは別のリレーショナル・データベースに格納されました。公開後に何十もの上流データの修正が入ってきたため、ライブデータベースのエントリを調整するプロセスとインターフェースを持つことが重要でした。

しかし、シンプルな構造を持つ小規模なデータセットであれば、高価なソフトウェア開発プロジェクト抜きでも利用可能にできます。サードパーティ製の表計算ツール（Google Sheets など）には、表を埋め込むことができるものがあります。また、HTML の表を検索、フィルタリング、ソートなどの機能で拡張するフロントエンドの JavaScript ライブラリも数多くあり、数百行程度のデータであれば、読者がアクセスできるようにするのに十分な機能を備えています。

大規模なデータセットにアクセスするための魅力的な仲介者は、API を介する JavaScript ベースの Web アプリケーションです。これは、本格的なアプリケーションを作成することなく、iframe で埋め込み可能な検索インターフェースを実行するのに適しています。API はサードパーティのサービスを介して実行することができ、フロントエンドのスタイルを完全にコントロールすることができます。

データベースがもたらす効果

記事の中に、あるいは記事と一緒にデータベースを利用することで、読者にもニュースルームにも多くの新しいメリットがあります。読者の側から見ると、オンラインのデータベースを提供することで、読者は自分の住んでいる都市や政治家、医師などを検索し、ストーリーを自分の生活に結びつけることができます。これは、より個人的なレベルで記事に関与するチャネルを提供します。また、これらの検索クエリを分析することで、ニュースルームは読者が何に興味を持っているかというデータを得ることができ、今後の仕事のヒントになる可能性があります。

ニュースルーム側では、データベースを長期的な調査のための投資と考えれば、他のデータベースや文書群とエンティティを自動的に相互参照し、リード文を作成することができるようになります。他のニュースルームが

同じようなデータベースを利用できるようになったならば、既存のインフラや方法論を再利用しながら、たやすく共同作業や取材範囲の拡大ができるようになります。また、データベースを利用することで、検索エンジンへの最適化が図られ、ニュースサイトへのトラフィックが増加する可能性があります。データベースが内部のエンティティの個々の URL を提供する場合、検索エンジンはこれらのページをピックアップし、これらの多数のエンティティに関連する頻度の低いキーワード検索の結果で上位にランク付けします。

いわゆる「ロングテール」と呼ばれる Web 検索では、これらのページが検索エンジンで上位に表示され、記事公開元のサイトへのトラフィックが増加することになります。検索エンジンに最適化することは、ジャーナリズムの世界では好ましくない行為とみなされることがありますが、読者が特定の問題を検索しているときにジャーナリズムの情報を提供することは、読者とのエンゲージメントを成功させる一環と考えることができます。公開データベースの目的は、検索キーワードで競争することではありませんが、有機的なトラフィックを促進し、新しい読者を惹きつけることができるという点では、歓迎すべきでしょう。

責任あるデータについての考察

データコミュニティの責任あるメンバーのアプローチを参考にしましょう。[彼らは)データを新たにさまざまな方法で使用する際に直面する、倫理的およびプライバシー関連の課題を考慮したベストプラクティスの開発に取り組んでおり、いくつかの方法で潜在的なリスクを検討することができます10。まず、ニュースルームが個人のデータを含むデータベースの公開を決定するという状況において、権力がどのように分配されているのかという問題があります。通常、その人々は、公開前にそのデータを拒否したり修正したりする能力を持っていません。例えば、このようなデータベースに含まれる政治的暴露を受けた人物(PEP、重要な公的地位にある人物)は、そのような展開を予想し、行動を起こすのに十分なリソースを持っていると思われますが、医療従事者はおそらく調査に巻き込まれることを予想していないでしょう。例えば、Euros für Ärzte データベースに登録されている医師は、自分の名前を検索すると、検索結果の上位にこのデータベースの自分のページが表示されるようになったとフィードバックしています。

読者側の力学にも注目してください。このデータベースは誰にとって最も有用なのか？データベースを利用するのに必要なツールや能力を持っているのか、それとも既に権力を持っている人が自分の利益のためにこの情報を利用するのでしょうか。公開前のユーザーテストの範囲を広げて、対象となる人々にデータベースを適切に説明するための十分な文脈を確保したり、データベースのインターフェースを対象となる人々がより利用しやすくなるような機能を盛り込んだりする必要があるかもしれません。

より多くのデータが社会にとってより良い判断につながるという前提は、近年、多面的に疑問視されるようになってきました。教育学者のクレール・フォンテーヌはこれを敷衍して、米国では「学校のパフォーマンス」について入手できるデータが増えたにもかかわらず(あるいは増えたからこそ)、学校における[人種]隔離が進ん

でいると指摘しています。彼女は「学校選択と横行する（人種）隔離の間の因果関係はまだ確立されていない」と指摘しています。しかし、彼女や他の人々はその関係を理解するために努力しており、より多くの情報がより良い決定につながるという、おそらく過度に単純化された関係に疑問を呈し、「より良い」とは何を意味するのかを問い直しています（Fontaine, 2017）。

二つ目は、データベース自体の問題です。データベースには、何を収集し、何を削除したか、どのように分類したか、どのように分析したかなど、多くの人間の判断が含まれています。客観的なデータは存在しないのですが、データの限界に対するリテラシーや理解度が比較的低いために、読者が結論を誤解する可能性は高いのです。例えば、組織犯罪に関与している政治団体のデータベースにある組織が入っていないのは、その組織が犯罪に荷担していないことを意味しているのではなく、単にその組織の行動について利用可能なデータがなかったことを意味しています。マイケル・フォレビウィスキーとダナ・ボイド（2018）は、このようなデータの欠落を「データボイド」と呼び、データボイドが「社会のバイアスや偏見を受動的に反映している」場合があると指摘しています。データが飽和した（ようにみえる）空間におけるこのようなデータの欠落は、ブルックリンを拠点とするアーティスト兼研究者のミミ・オヌオハ（2016）が「ミッシング・データセット」と呼ぶものにも密接に対応しており、データを収集する際の社会的な選択を浮き彫りにしています。

三つ目は注目の方向性です。データベースを利用することで、人々の関心の焦点が、より広範なシステム上の問題から個人の行動に移り、またその逆も可能です。製薬会社と医療従事者の間の資金の流れは、明らかに公共の関心事ですが、個人レベルでは、医師は自分自身が公共の利益の対象だとは思っていないかもしれません。しかし、ある問題をより広く、構造的に（一過性のものではなく、パターンとして）示すためには、複数の個人からのデータが必要であることは事実です。先に述べた Euros für Ärzte のケーススタディのように、データベースによっては、何が、あるいは誰が、公共の利益になるのかという境界線を変えるものもあります。

個人が自分のデータの公開に同意した場合でも、ジャーナリストはそのデータがいつまで公共の利益になるのか、いつ削除すべきなのかを判断しなければなりません。一般データ保護規則（GDPR）は、ジャーナリストがこの種の個人データをどのように管理すべきか、また、個人が自分のデータの掲載に対する同意を取り消すためにどのような仕組みを利用できるかに影響を与えると思われます。このような課題に対して、私たちは、調査や製品のプロセスと最終的な結果の両方で、人々の権利にどのような影響があるかを検討しています。責任あるデータの取り扱いは、どこかの時点で考慮すべきチェックリストではなく、継続的なアプローチであることを理解することが重要です。データ収集から公開までの調査全体を通して、データに反映された人々の権利を優先するアプローチが、（データ）ジャーナリズムを信頼に値する者にする核心なのです（Rosen, 2018）。

ベストプラクティス

調査結果を公開するためにデータベースの構築を考えているジャーナリストのために、いくつかのベストプラクティスと推奨事項を紹介します。これらは順次改善していくことを想定しており、ご提案をお待ちしております。初めに、公開に先立って、データベースの誤りを修正する方法のプロセスを開発します。データ証明に関する適切なプラクティスは、エラーの原因を見つけるのに役立ちます。第二に、フィードバックチャンネルを設けること。特に、調査中に予期せぬ形で言及された個人から、フィードバック（苦情）が届く可能性があります。苦情を受け付ける優れたユーザー・エクスペリエンスを提供することで、その経験を生かすことができるかもしれません。三つ目は、データベースを最新の状態に保つか、またはメンテナンスが終了したことを明確に示すことです。ジャーナリズムでは、データベースの公開には、記事よりも高いレベルのメンテナンスが求められます。また、データベースにはインタラクティブ性があるため、記事と比較すると最新の状態に対する期待は段違いです。

四つ目は、長期的なメンテナンスのために十分なリソースを割り当てることです。データやデータベースソフトウェアを最新の状態に保つには、かなりのリソースが必要です。例えば、翌年のデータをデータベースに追加するには、新しいデータと古いデータを統合し、ユーザーインターフェースに時間軸を追加する必要があります。五つ目は、読者がどのようにデータベースを利用しているかを観察することです。検索や使用の傾向から、将来の記事や調査のヒントが得られるかもしれません。最後に、透明性を保つことです。データベースが 100％「完全」であることは稀であり、[報じる側の]選択が組み込まれています。これらの選択肢を隠すのではなく、読者が何を見ているのかを明らかにしましょう。

脚注

1. correctiv.org/recherchen/euros-fuer-aerzte/（ドイツ語）

2. projects.propublica.org/docdollars

3. interaktiv.morgenpost.de/schul- nder-berlin/#/

4. www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-police-killings-us-database

5. www.theguardian.com/australia-news/ng-interactive/2016/aug/10/the-nauru- les-the-lives-of-asylum-seekersin-detention-detailed-in-a-unique-database-interactive

6. www.thebureauinvestigates.com/projects/drone-war/

7.offshoreleaks.icij.org/

8. data.occrp.org

9. correctiv.org/thema/aktuelles/euros-fuer-aerzte/

10. .responsibledata.io/what-is-responsible-data/

参考文献

EFPIA. (2013). About the EFPIA Disclosure Code. European Federation of Pharmaceutical Industries and

Associations. efpia.eu/media/25046/efpia_ about_disclosure_code_march-2016.pdf

Fontaine, C. (2017, April 20). Driving school choice. *Medium*. points.data-society.net/driving-school-choice16f014d8d4df

Golebiewski, M., & Boyd, D. (2018, May). Data voids: Where missing data can be easily exploited. *Data & Society*. datasociety.net/wp-content/uploads/2018/05/Data_Society_Data_Voids_Final_3.pdf

Holovaty, A. (2006, September 6). A fundamental way newspaper sites need to change.

Adrian Holovaty. http://www.holovaty.com/writing/fundamental-change/

Onuoha, M. (2016, February 3). On missing data sets. github.com/MimiOnuoha/missing-datasets

Rosen, J. (2018, May 14). Optimizing journalism for trust. *Medium*.

データとインタラクティブ・コミックで水紛争を語る

文：ネリー・ルナ・アマンシオ

概要

ペルーの大手鉱山会社との水紛争を巡る調査報道を、インタラクティブ・コミックで説明してみました。

キーワード：水争い、データジャーナリズム、環境、コミック、インタラクティブ、ペルー

コミック「La guerra por el agua（水をめぐる戦争）」に登場するものはすべて現実です（figure29.1）。主人公であるマウロ・アパサとメルチョーラ・タクレは、彼らが抱く不安や不確実性とともに実在しているのです。

彼らと出会ったのは、2016 年 9 月の暑い日でした。正午で、日陰もなく、風もありませんでした。彼女（メルチョーラ）は手で土の草取りをし、彼（マウロ）は荒地に溝を作っていました。彼らは 70 年以上にわたり、ペルー南部の農業地帯タンボ・バレーにある小さな土地で食料を育ててきましたのですが、そこに鉱山プロジェクトが持ち上がったのでした。この夫婦の歴史は、何千もの農民や先住民のコミュニティと同様に、世界で最も戦略的な資源の一つ、水を採取しようとする強力な産業と農民との間の紛争を物語っています。

ペルーのように、200 以上の環境紛争があり、国家予算がこの分野からの収入に大きく依存している国で、この対立をどのように物語ればいいのでしょう？不安定な農業従事者、多国籍企業の利益、そして徴税を強化しなければならない政府の利益の間の緊張関係をどのように表現すればよいのでしょうか。この問題を理解するのに役立つ物語とは？この緊急の問題に人々を動員するにはどうしたらいいのでしょうか？これらの疑問から生まれたのが、OjoPúblico が開発したペルー初のインタラクティブ・コミック『The War Over Water』なのです。

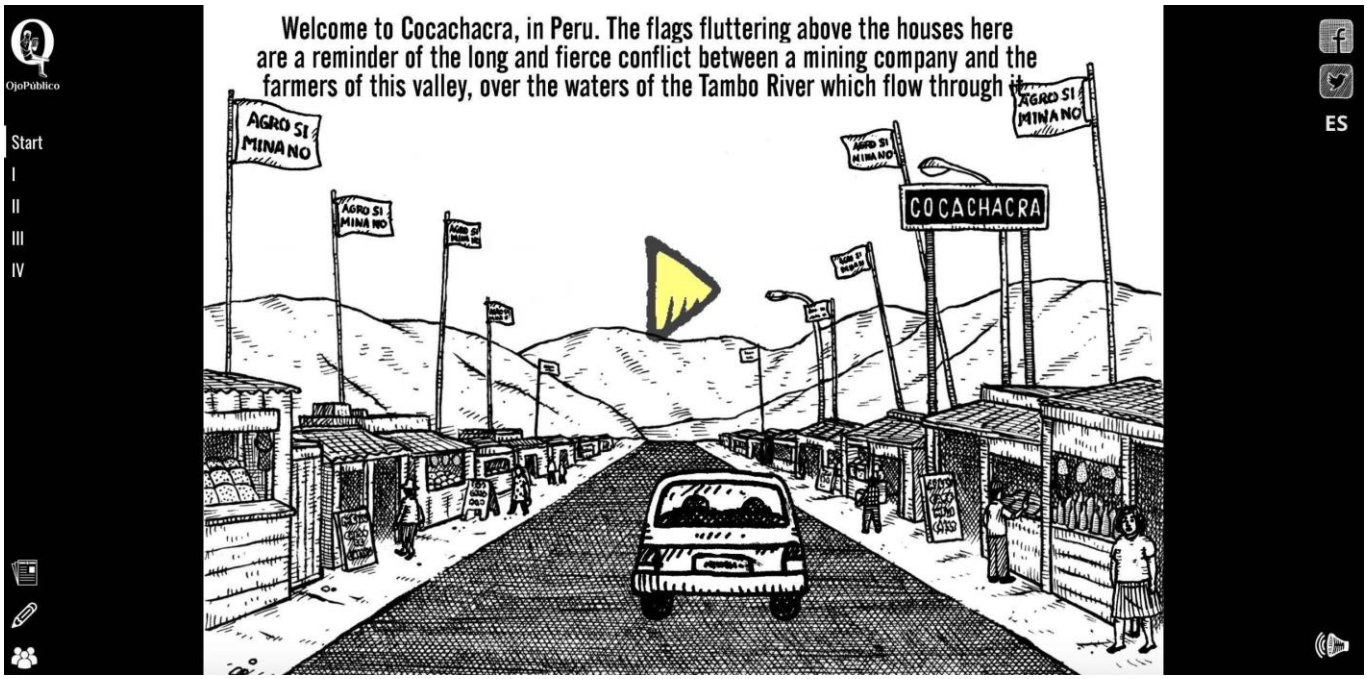


Figure 29.1. インタラクティブ・コミック「The War over Water」のホーム画面 出典：OjoPúblico.

この作品では、データとビジュアライゼーションを統合して、この紛争を巡る物語を描き出しました1。

なぜインタラクティブ・コミックなのか？

プロジェクトは 2016 年 7 月に始まりました。経済的な観点から紛争を語ろうと、最も象徴的な地域への旅行を模して、2 人の農民の視点から読者にアプローチすることにしました。インタラクティブなフォーマットにより、観客は道のりを越えて、紛争の音や対話を体験できます。私たちが選んだのは、メキシコでも世界でも有数の大富豪であるゲルマン・ラレアが所有する世界最大級の鉱山会社、サザン・コッパー社のティア・マリア鉱山プロジェクトの話です。このプロジェクトに反対する地元の人々は、警察による暴力的な弾圧を受け、市民 6 人が犠牲になりました。

このコミックを制作したチームは、ジャーナリストの私、漫画家のヘスス・コジオ、Web 開発者のジェイソン・マルティネスで構成されています。私たち 3 人は、紛争の中心地である[ペルー第二の都市]アレキパのタンボ渓谷に赴き、指導者や農民、当局にインタビューし、その過程を記録しました。メモ、写真、そして後にコミックの最初のスケッチとなる絵を描きました。リマに戻ってからは、最初のプロトタイプとなるものを作成しました。これです。このプロトタイプをもとに、最終的な脚本を書き、インタラクティブな機能を検討し、プロジェクトの開発に着手しました。

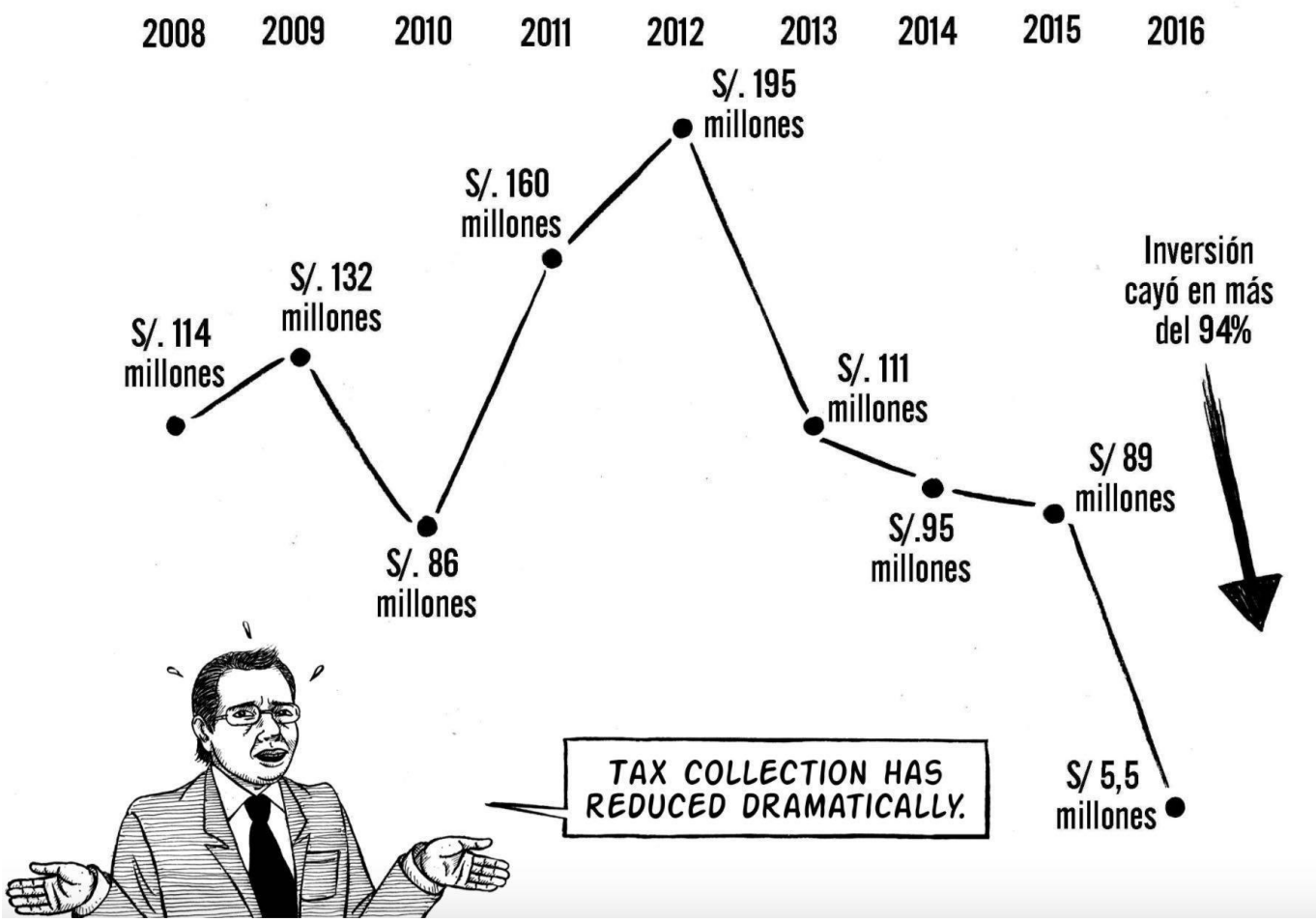


Figure 29.2. 2008 年以降、ペルーでは水をめぐる鉱業紛争の結果、徴税額が減少していることを示すデータ・ビジュアライゼーション 出典：OjoPúblico.

コミックに誠実であること

私たちがコミックというメディアを選んだのは、漫画家のジョー・サッコが言うように、ジャーナリストは「時間均等の名の下に真実を削るべきではない」(2012)と考えるからです。サッコは、プロジェクトの第一章のプレゼンテーションに参加してくれました。彼の作品の一つである「Srebrenica」は、1995 年に 8000 人以上のボスニアのイスラム教徒が死亡した虐殺を巡る Web コミックで、大いにインスピレーションを受けました。「The War Over Water」の制作には 8 ヵ月を要しました。この作品は実際の出来事に基づいており、観客が登場人物の日常生活を体験し、ペルーの経済における最大のジレンマのひとつを表面化できるような物語構造になっています。農業か鉱業か？両方を行うのに十分な水はあるのか？私たちは、マウロとメルチョーラの目と記憶を通して、この対立の物語を語りました。地域の経済的依存関係や鉱山会社の税制上の優遇措置などを示すデータ・ビジュアライゼーションが添えられています。

この作品に登場するシーンやセリフは、現地での取材、当局や地元の人々へのインタビュー、サザン・コッパー社の財務調査などで得られた実在のものです。台詞、人物、インタビュー、設定など、シーンを誠実かつ正確に構成することを目指しました。

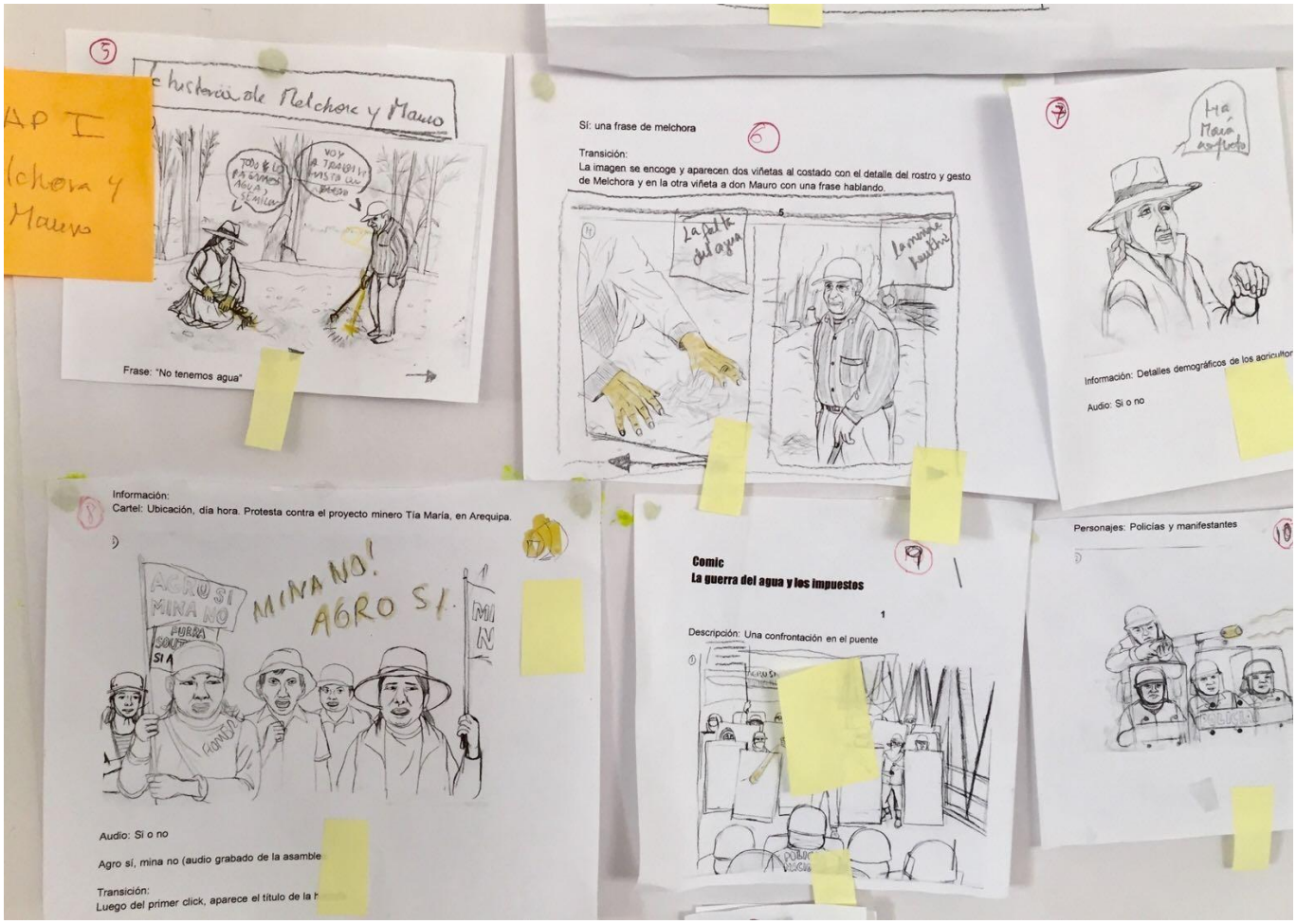


Figure 29.3. OjoPúblico のジャーナリストとイラストレーターは、こうして「The War over Water」というコミックの対話型スクリプトを作成しました。出典：OjoPúblico.

紙媒体から Web へ

漫画家のヘスス・コジオは、インタラクティブ・コミックでの時間の扱い方を再考することに挑みました。「印刷された漫画や静的なデジタルストリップでは、画像のインパクトで読者を立ち止まらせることが目的になりますが、インタラクティブ・コミックでは、より機敏でダイナミックな読書の流れに合わせて構成や画像を変更しなければなりませんでした」。

技術的な観点から見ると、これまでにインタラクティブ・コミックを開発したことがない OjoPúblico のチームにとって、挑戦的なプロジェクトでした。アニメーションやトランジションを作ることができ、シーンやタイム

ラインを標準化することができる[Javascript]ライブラリ、GSAP(GreenSock Animation Platform)を使用しました。それを JavaScript、CSS、HTML5 で補完しました。このコミックには 42 のシーンと 120 以上のカットがあります。ヘスス・コジオは、鉛筆とインクを使って、台本に登場するキャラクター、シーン、風景をそれぞれ描きました。次に、これらのイメージをデジタル化し、レイヤー（背景、環境、相互に作用する必要のある図面の文字と要素）ごとに分離しました。

Web から紙に戻る

The War Over Water は、トランスメディア・エクスペリエンスです。印刷版も発行しています。二つのプラットフォームで、このコミックは異なる読者へのアプローチを目指しています。OjoPúblico チームの最大の関心事は、公共に関する(しばしば複雑な)ストーリーを伝えるためのナラティブやフォーマットを探求することです。私たちは、データ調査報道の受賞歴があります。また、他のプロジェクトでは、暴力というトピックを語るためにコミック形式を使用しています。「Proyecto Memoria(記憶のプロジェクト)」では、1980 年から 2000 年の間にペルーが直面した国内紛争の恐ろしさを伝えました。コミックは、データを使って物語を語るための強力な言語となります。

これが私たちの提案です。調査報道に携わるジャーナリストは、さまざまな読者に物語を伝えるために、あらゆる手法を試すべきです。そして何よりも、ペルーの天然資源管理のような権力の不均衡を告発したいのです。

脚注

1. laguerraporelagua.ojo-publico.com/en

参考文献

Sacco, J. (2012). *Journalism*. Henry Holt and Co.

データジャーナリズムは、人とストーリーを重視すべし

文：ウィニー・デ・ヨン

概要

　ストーリーとストーリーの対象となる人々は、データジャーナリズムを含むジャーナリズムの中心となる太陽であるべきです。

キーワード：ストーリーテリング、データジャーナリズム、ラジオ、テレビ、データ公開、データ・ビジュアライゼーション

　人間と同じように、データジャーナリズムとジャーナリズムには、相違点よりも共通点の方が多いものです1。データ駆動型報道は、別のスキルを必要とする異なるタイプの情報源を基にしていますが、思考プロセスはほぼ同じです。実際、引いて見れば、そのプロセスはほとんど見分けがつかないでしょう。

既知の未知なるもの

　ジャーナリズムの核心は、「既知の未知」を「既知の既知」にするビジネスにあります。既知と未知の概念は、2002 年にアメリカのラムズフェルド国防長官が一般化しました。当時、イラク政府が大量破壊兵器をテロリストグループに提供したという証拠はありませんでした。この件に関する記者会見で、ラムズフェルドは次のように述べています。

何かが起こっていないという報告は、私にとっていつも興味深いものです。なぜなら、私たちは、（自分が知っていることを知っている）既知の既知があると知っているからです。また、既知の未知があることも知っています。つまり、知らないことを分かっているということです。しかし、（知らないことすら分かっていない）未知の未知もあります。我が国や他の自由主義国の歴史を振り返ってみると、［対処が］難しいのは後者のカテゴリーであることが分かります。（米国防総省、2002 年）

　ジャーナリズムのプロセスはすべて、知っていることと知らないことのマトリックスの上で駒を動かすことに尽きます。すべてのジャーナリズムは、質問から、前述のマトリックスで言うと既知の未知のものから始まります（何かわからないことがあるからこそ、質問をするのです）。疑問や予感から公開できる［水準の］ストーリーに移行するためにブートストラップを行う場合、理想的なルートは、すべてのポーンを既知の未知から既知の既知へと「単純に」移動させることです。しかし、ジャーナリストであれば誰もが言うように、現実は異なります。取材をしたり、文書やデータを調べたりしているうちに、自分が知らなかったこと（未知の未知）が出てきて、それにも答えが必要になってくることがあります。運が良ければ、自分が知っていることを知らなかったこと（未知の既知）に

出くわすこともあるでしょう。締め切りに向けて、あなたは三つのカテゴリーの知識を「既知の既知」に変えていくことになります。既知の未知（始めたきっかけとなった疑問）、未知の未知（つまり、質問すべきであったことを知らなかった質問）、未知の既知（自分が知らなかった答え）です。政府とは異なり、ジャーナリストは既知の知識によって行動を起こしたり、公表したりすることしかできません。

堅牢なジャーナリズム

データ駆動型報道と伝統的なブートストラップ［独力で行う報道］がこれほどまでに区別されないのであれば、きっとこの二つは同じ基準を満たすべきでしょう。ジャーナリズムと同様に、データジャーナリズムも常に真実であり、独立性があり、偏りがないものでなければなりません。他の事実と同様に、データは検証される必要があります。既知の既知を作り出す前に、自問してみてください。そのデータは真実か？各数値は実際に何を意味しているのか？データの出典は何か？そのデータはなぜ集められたのか？誰がそのデータセットを作ったのか？データテーブルはどのように作成されたか？データには外れ値があるか？それらは意味があるのか？

そして、忘れられがちで、あらゆるインタビューと同様に、非常に重要なことがあります。ソースに書かれていないことは何か？要件とそのための質問は同じでも、結果として生じる行動はわずかに異なります。

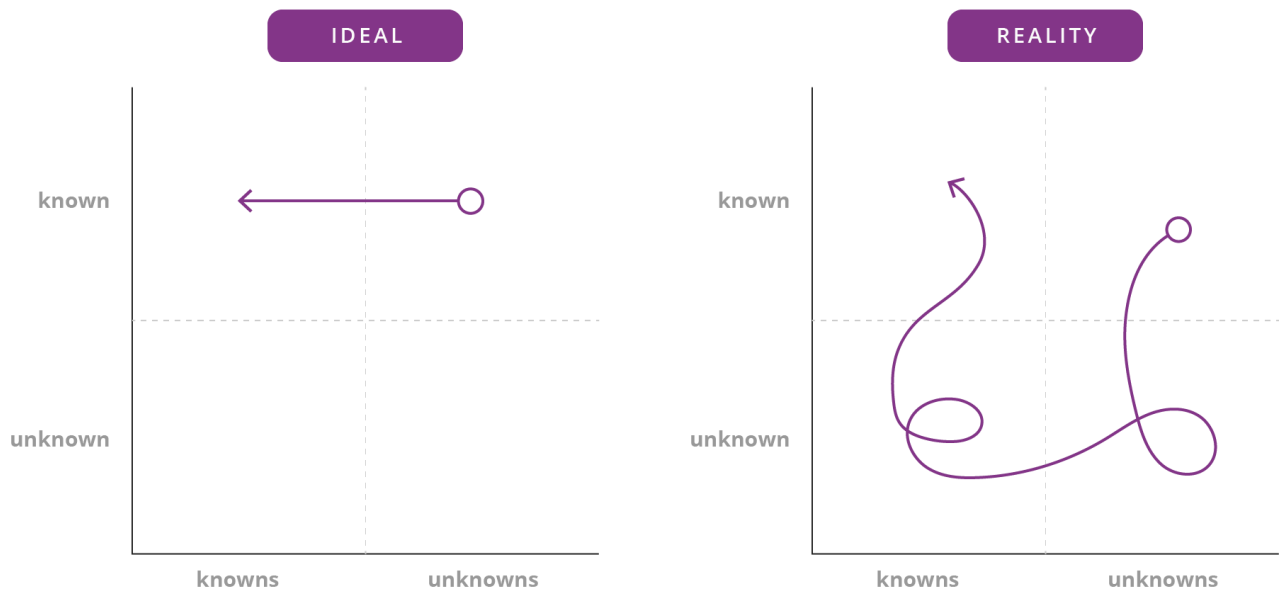


Figure 30.2. ジャーナリズムのための既知と未知のマトリックスを手ほどきする 出典：Lars Boogaard.

ビル・コヴァッチとトム・ローゼンスティールが『ジャーナリズムの原則』(2007)で述べているように、ジャーナリストの仕事はまず、「どの情報が信頼できるかを確認し、人々が効率的に把握できるように情報を整理すること」につきます。データジャーナリスト、特にテレビやラジオで働いている人たちにとって、これは、自分たちが好む数値が必ずしも最終的な作品に反映されないことを意味します。

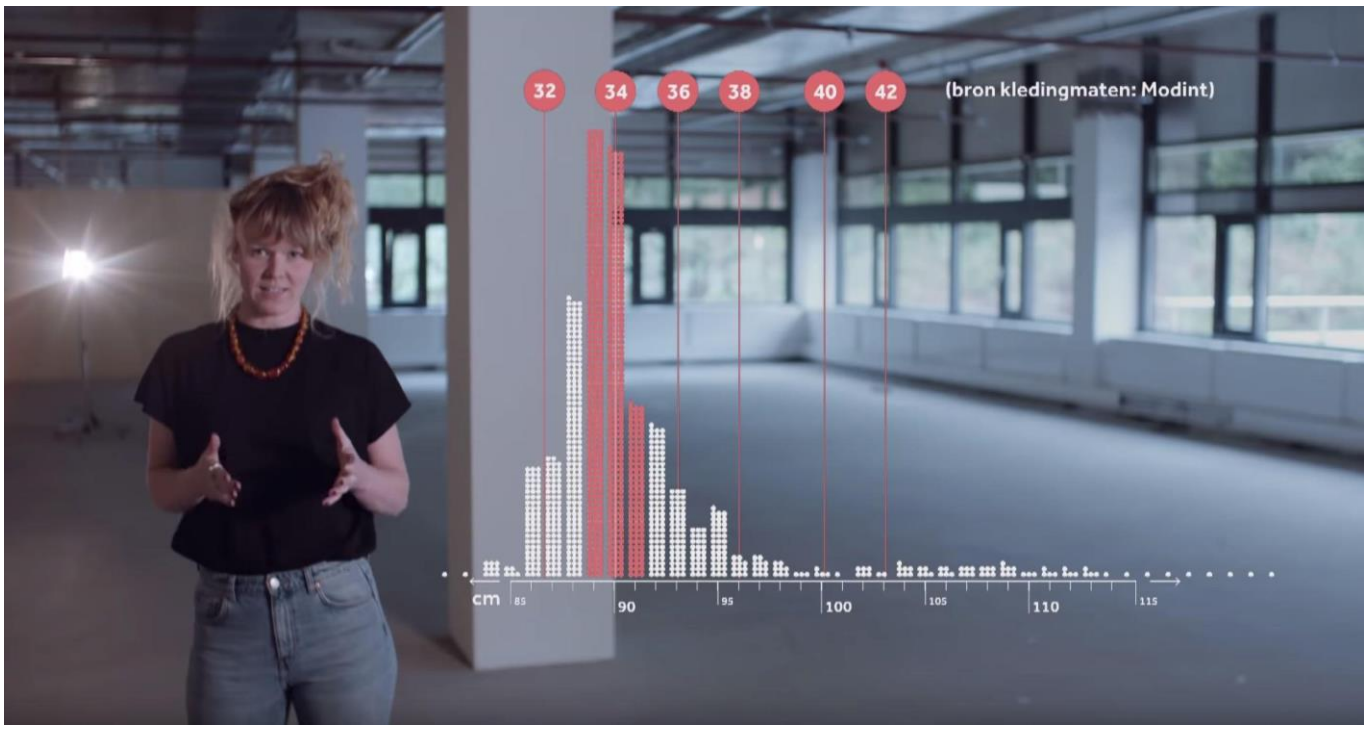


Figure 30.3. ファッションモデルになるにはどれだけ痩せる必要があるか示した NOS のビデオ 出典: NOS.

限られたオタク

データ分析をするときには、当然ながら正確であるべきです。しかし、ストーリーを「効率的に把握できる」ようにするためには、精度に制限を設ける必要があります。例えば、最終的に使用する小数位です。「10 人中 4 人」という数字は、「41.8612%」よりも良いでしょう。私の経験では、適切な精度とは、土曜日の午後、データオタクではない友人に自分の話をするときに使う精度にかなり近いものです。観客がストーリーを理解するために使用した方法やツールについて知る必要がない限り、オタク的な要素は方法論に限るべきでしょう。

読者がデータ駆動型の作品を読んだり、聞いたり、見たりするときに考えるのは、ストーリーであって、ストーリーを浮かび上がらせるデータや分析、テクノロジーではないからです。最高のデータジャーナリズムは、それとはわからない、見えない技術にしてしまうことなのです。その不可視性がストーリーを円滑にし、ジャーナリズムを「効率的に把握できる」ものにしてくれるのであれば、それに越したことはありません。ジャーナリズムは、市民が社会を把握するためのさまざまな地図を作るのですから、すべての人に読まれ、多くの人に読まれる地図を作るべきなのです。

ラジオとテレビ

ラジオやテレビでデータジャーナリズムの記事を発表する際には、「少ない方が豊か(less is more)」です。オランダ最大の報道機関であるNOSのニュースルームでは、記者はストーリーを伝えるのに必要な秒数を考慮しています。ストーリーがどのように作られたのか、なぜ他のデータソースではなくそのデータソースを使うことにしたのかなど、ストーリーやそのストーリーに対する一般の人々の理解に寄与しないなら、あれこれ考えている時間はないということです。ファッションモデルになるためにはどれくらい痩せればいいのかというオンラインビデオでは、方法論を20秒かけて説明しました2。全国ネットのテレビ番組で90秒(の枠)しかないのに、20秒は多いですよね。この場合、「少ない方が豊か」方式では、調査方法を説明する時間が残らないことを意味します。限られた時間とスペースの中では、何よりもストーリーが優先されます。

控えめなビジュアル

もちろん、データ・ビジュアライゼーションにも「少ない方が豊か」の格言は当てはまります。データジャーナリズムはティーンエイジャーのセックスとよく似ています。誰もが話題にしているが、実際にやっている人はほとんどいないのです。ニュースルームがようやくデータをツールキットに加えると、何でもかんでもデータを可視化をして、情報を伝えることになりがちです。確かに、私もビジュアルは好きです。特に革新的でハイエンドなものは好きですが、それはストーリーにプラスになる場合のみです。

ビジュアライゼーションは、様々な方法でジャーナリズムに付加価値を与えることができます。例えば地域レベルでの洞察力を高めて一般の人々の理解を広げることができます。だから、紳士らしく振る舞うことです。ストーリーを伝えるのに役立つ、付加価値のあるデータ・ビジュアライゼーションに限定してください。今日では、ほとんどの人がラジオやテレビを聴くことと他の活動を組み合わせています。これでは、情報摂取量が制限されてしまいます。運転中にニュースを聞くのは二の次、料理中にテレビを見るのも同じです。ですから、視聴者に多くを求めすぎないように注意してください。繰り返しになりますが、私たちの仕事は目に見えないものかもしれません。しかし、私たちはニュースを伝え、ストーリーを伝えるためにここにいるのであって、dataviz の力を見せつけるためではありません。

人について

つまり、あなたのストーリーやジャーナリズム、そして人生において本当に大切なものは、データセットには収まりません。今までも、これからも。何をするにしても、どこで記事を公開するにせよ、データではなく人を語ることです。そして、必要以上にデータやテクノロジー、ニュースのオタク的要素を使いたくなったときは、自分がこの分野では数少ない職人の一人であることを思い出してください。そのこと自体が素晴らしいのです。強調する必要はありません。だからこそ、優れたデータジャーナリズムに見られるような序列にこだわるべきなのです。形式はデータとストーリーを手助けするためにあります。すべてのもの、そして誰もが、私たちの太陽であるストーリーの周りを回る必要があります。ストーリーは王なのです。

脚注

1. アイデアは古い要素の新しい組み合わせなので、このエッセイはウィニーによる 2019 年のニーマンラボでの予測、アムステルダムのスマートニュースデザイン会議での講演、そしてジャーナリストに数学を教えるオランダの Web サイト、alshetongeveer- maarklopt.nl から引用している。

2. www.youtube.com/watch?v=DWRGqmywNY

アルゴリズムの専門性：視座と調査方法

文：ニコラス・ディアコプロス

概要

ジャーナリズムのスキルと技術的なスキルが融合してアルゴリズムにふさわしい精査が行われるようになり、アルゴリズムに関する専門性が形成されつつあります。

キーワード：アルゴリズム、アルゴリズムの説明責任、計算論的ジャーナリズム、調査報道、アルゴリズム研究、情報の自由（FOI）

プロパブリカは 2016 年5月、社会におけるアルゴリズムを調査するシリーズ「Machine Bias」を始めました1。このシリーズで最も目を引いたのは、刑事司法の判断に用いられる再犯リスク評価アルゴリズムの人種的偏りを暴いた調査・分析でした（Angwin et al.2016）。これらのアルゴリズムは、個人再犯のリスクが低いか高いかをスコアリングします。州やその他の自治体は、公判前拘留、保護観察、仮釈放、そして時には判決に至るまで、さまざまな場面でこのスコアを利用しています。

プロパブリカの記者は、フロリダ州ブロワード郡のスコアを情報公開請求し、そのスコアと実際の犯罪歴を照合して、2 年以内に実際に再犯したかどうかを調べました。分析したところ、黒人の被告人は白人よりも高いリスクスコアを割り当てられる傾向があり、2 年後に再逮捕されていないのに、誤って高リスクと判定される可能性が高いことがわかりました（Larson et al.2016）。刑事司法制度におけるスコアリングは、アルゴリズムが社会に展開されている一つの領域にすぎません。「Machine Bias」シリーズではその後、Facebook の広告ターゲティングシステム、地理的差別のある自動車保険料、Amazon.com の不公平な価格設定など、あらゆる分野を取り上げています。アルゴリズムによる意思決定は、官民を問わず、ますます浸透しています。クレジットや保険のリスク評価、雇用システム、福祉管理、教育や教師のランキング、オンラインメディアのキュレーションなどの領域で目にすることができます（Eubanks, 2018; O'Neil, 2016; Pasquale, 2015）。

大規模に運用され、多くの人々に影響を与えるアルゴリズムは、時には論争の対象となる計算、ランキング、分類、関連付け、フィルタリングを行うことがあります。膨大なデータに基づいて稼働するアルゴリズムは、社会で力を行使する新しい有力な方法となっています。プロパブリカの「Machine Bias」シリーズのように、アルゴリズムを通じて権力がどのように行使されるかを調査し、説明責任を果たすため、計算論的ジャーナリズムやデータジャーナリズムの新たな分野が生まれています。

私はこのアルゴリズムを巡る説明責任報道を「ジャーナリズムの伝統的な監視機能を、アルゴリズムを通じて行使される権力に向け再調整したもの」と呼んでいます（Diakopoulos, 2015）2。アルゴリズムは、その客観

性にもかかわらず、ミスを犯したり、バイアスをかけたりすることがあるため、綿密な調査が必要になります。

もちろん、アルゴリズムの説明責任には様々な形態があり、ジャーナリズム以外にも、政治、法律、学術、活動家、芸術など、多様な場で問われる可能性があります（Brain & Mattu, n.d.; Bucher, 2018）3。しかし、ここでは、世論の圧力を動員することで説明責任を果たさせる、独立したジャーナリスティックな試みとしての報道に焦点を当てます。これは、規制や法的基準の策定、市民社会における監査機関の創設、効果的な透明性政策の策定、反射的なアートショーの開催、学術的な批評の発表など、説明責任に貢献しうる他の手段を補完するものです。

ジャーナリズムにおける専門性を決めるには、まずアルゴリズムの何がニュースになるのかを定義することが重要です。技術的には、アルゴリズムとは、特定の問題を解決するため、あるいは定義された結果を達成するために従う一連の手順です。情報プロセスの観点から言えば、アルゴリズムの結果は通常、意思決定です。アルゴリズムのパワーの核心は、コンピュータがこのような意思決定を非常に迅速かつ大規模に行い、多数の人々に影響を与えることができるという点にあります。しかし、アルゴリズムの説明責任は、技術的側面にとどまりません。アルゴリズムは複雑な社会工学的システムにおける設計者、運用者、所有者、管理者などの人々が織り成す技術の複合体として理解すべきなのです（Ananny, 2015; Seaver, 2017）。アルゴリズムの説明責任とは、それらの人々がシステムの中で、またシステムを通じてどのように権力を行使し、システムの決定に最終的に責任を負うのかを理解することです。

多くの場合、アルゴリズムがニュースになるのは、アルゴリズムが何らかの形で「悪い」決定を下したときです。これは、アルゴリズムが想定外のことをしたり、あるいは想定していたことをしなかったりすることを意味します。ジャーナリズムにとっては、間違った判断の社会的意義と結果が重要な要素となります。

個人にとって、あるいは社会にとって、どのような損害が生じる可能性があるのか。間違った決定は、個人に直接影響を与えるかもしれませんし、全体としては構造的なバイアスを強化するかもしれません。また、間違った決定はコストがかかることもあります。ここでは、さまざまな誤った判断がニュースになった例を見てみましょう。

アルゴリズムの視座

ここ数年、ジャーナリズムの世界で展開されてきたアルゴリズムの専門性を観察し、また私自身がアルゴリズムを調査する中で、アルゴリズムの説明責任を巡る議論に通底すると思われる、少なくとも四つの原動力を特定しました。（a）差別と不公平、（b）予測や分類におけるエラーやミス、（c）法律や社会的規範の違反、（d）意図的または不注意によるアルゴリズムの誤用です。以下に、それぞれの例を挙げて説明します。

差別と不公平。差別と不公平を明らかにすることは、アルゴリズムによる説明責任報道に共通するテーマです。冒頭で紹介したプロパブリカの記事は、アルゴリズムがいかに異なるグループの人々の扱いに体系的な格差をもたらすかをはっきり示しました。［再犯］リスク評価のスコアを作成したノースポイント（現在は

Equivant に社名を変更）は、このスコアは人種に関係なく同じ精度であり、それゆえ公平であると主張しました。しかし、彼らは、黒人に影響を与えるミスの数が不均衡であることを考慮に入れていませんでした。差別や不公平の話は、適用される公平性の定義に依存し、異なる政治的推測を反映している可能性があるのです（Lepri et al.、2018）。

私は、アルゴリズムシステムによる不公平さを明らかにするストーリーにも取り組んできました。その中で、Uber のダイナミック・プライシングがワシントン DC の近隣地域ではどう違うのかを調べました（Stark & Diakopoulos, 2016）。待ち時間の違いや、Uber のサージ・プライシング・アルゴリズムに基づいて待ち時間がどのように変化するかが見て取れたため、地域によってサービス品質（すなわち待ち時間）が異なるという仮説を立てました。異なる国勢調査区の待ち時間を系統的にサンプリングすることで、所得、貧困率、近隣の人口密度などの他の要因を調整しても、有色人種が多い国勢調査区は車の待ち時間が長くなる傾向にあることを示しました。Uber のドライバーの行動や潜在的なバイアスなど、他の人的要因もシステムに影響を与えるため、この不公平な結果を Uber の技術的アルゴリズムのみに帰することは困難です。しかし、この結果は、全体として考えると、システムには人口統計に関連した格差があることを示唆しています。

エラーとミス。アルゴリズムは、分類、予測、フィルタリングの決定において、特定のエラーやミスを犯した場合にもニュースになります。Facebook や Google などのプラットフォームでは、アルゴリズムによるフィルタリングによって、ヘイトスピーチ、暴力、ポルノなどの有害なコンテンツへの接触を減らしていることを考えてみましょう。これは、特に子どもにとって安全であるとうたって販売されている製品（Google の YouTube Kids など）では、子どものような特定の弱者を保護するために重要となる場合があります。このアプリのフィルタリングアルゴリズムのエラーは、子どもたちが不適切ないし暴力的なコンテンツに遭遇することがあるため、ニュースになるのです（Maheshwari, 2017）。古典的に、アルゴリズムは 2 種類のミスを犯します。偽陽性と偽陰性です。YouTube Kids の場合、偽陽性とは、実際には子供にとって全く問題ないのに、不適切と誤って分類された動画のことです。偽陰性とは、適切と分類された動画が、本当は子供に見せたくないものであることです。分類の決定は、個人が受けるポジティブないしネガティブな待遇の増減に影響します。アルゴリズムが誤って個人を選んでアイスクリームを無料で提供しても（ポジティブな扱いの増加）、その個人が文句を言うことはありません（ただし、他の人は不公平だと言うかもしれません）。一般的にエラーがニュースになるのは、子供に不適切なビデオを見せてしまうなど、その人にとってネガティブな扱いを増やすことにつながる場合です。また、機会を逃すなど、個人に対するプラスの扱いが減少した場合もニュースになります。 アルゴリズムが誤って除外したため、資格があるのに特別なオファーを受けることができない人がいたとしてみましょう。

最後に、エラーがニュースになるのは、正当でネガティブな注目を減少させる原因となる場合です。例えば、犯罪リスク評価のアルゴリズムが、高リスクの人物を誤って低リスクと判定してしまうことを考えてみましょう（偽陰性）。本人にとっては良いことですが、再び罪を犯す可能性のある人を自由にすることで、公共の安全に大きなリスクをもたらします。

法律や社会的規範の違反。予測アルゴリズムは、時として確立された法的・社会的規範の境界線を試すことがあり、それが別の機会や報道の角度につながることがあります。アルゴリズムによる名誉毀損の可能性について考えてみましょう(Diakopoulos, 2013; Lewis et al.2019)。名誉毀損とは、「ある人物を憎悪、嘲笑、軽蔑の対象とし、仲間内の尊敬を低下させ、敬遠される原因となり、または事業や取引において損害を与える虚偽の事実の陳述」と定義されています4。ここ数年、Google の(検索語句のタイプ中に自動的に検索候補を表示する)オートコンプリートアルゴリズムによって名誉を毀損されたと感じる個人を巡り、数多くの記事や訴訟が発生しています。オートコンプリートは、個人や企業の名前を、犯罪や詐欺、破産、性的行為などあらゆるものと結び付け、その人の評判に影響を与える可能性があります。アルゴリズムは、プライバシーなどの社会的規範を侵害した場合にもニュースになります。

[テックサイトの]Gizmodo は Facebook の「People You May Know」(PYMK)アルゴリズムを大きく取り上げています。PYMK アルゴリズムは、プラットフォーム上で潜在的な「友達」を提案しますが、それが時に不適切であったり、望ましくないものであったりします(Hill, 2017b)。ある記事では、記者が、PYMK によってセックスワーカーが客に身バレしたケースを確認しました(Hill, 2017a)。これは、性風俗につきまとう付随する潜在的な偏見だけでなく、ストーカーになりかねない客への恐怖[という観点]からも問題となります。名誉毀損やプライバシー侵害は、ここで考えられる二つの視点に過ぎません。

ジャーナリストは、アルゴリズムが様々な社会的文脈の中で生み出す、他の様々な法的・社会的規範の侵害に注意を払う必要があります。アルゴリズムは、測定可能なものだけをデータとして取り込む定量化された現実に依存しているため、正確な判断を下すために不可欠な社会的・法的な背景の多くを見逃してしまう可能性があります。特定のアルゴリズムが実際に世界をどのように「見ている」のかを理解することで、判断の裏付けとなる欠落した部分を明らかにし、検証につなげることができます。

人間による誤用。アルゴリズムによる意思決定は、多くの場合、社会工学的なシステムに組み込まれた人間とアルゴリズムの組み合わせによる、より大きな意思決定プロセスに組み込まれています。アルゴリズムの技術的構成要素のいくつかはアクセスできないのですが、その社会工学的性質は、ユーザー、デザイナー、所有者、その他のステークホルダーがシステム全体に対して持つ関係性を調査するための新たな機会をもたらします(Trielli & Diakopoulos, 2017)。社会工学に関係する人々がアルゴリズムを誤用しているなら、これもまたニュース価値があるかもしれません。アルゴリズムの設計者は、システムの合理的な使用状況を予測してガイドラインを定めており、利用者がこれらを無視すると、過失や誤用のストーリーにつながる可能性があります。プロパブリカのリスクアセスメントの記事は、その典型的な例です。ノースポイントは、男性用と女性用の 2 種類の(再犯リスクを予測する)ツールを作成していました。統計モデルは、それが使用される人口を反映したデータで訓練される必要があり、性別は再犯率の予測において重要な要素です。しかしブロワード郡は、男性用に設計・校正されたリスクスコアを、女性用にも使用して誤用していたのです(Larson, 2016)。

アルゴリズムを調査する

アルゴリズムの力を調査するには様々なルートがあり、常に一つのアプローチだけが適切とは限りません。高度な技術を要するリバースエンジニアリングや[人の目でコードを検証する]コードインスペクションの手法、自動化されたデータやクラウドソースによるデータ収集を利用した監査、さらにはアルゴリズムの反応に基づいた試作や批評を行うローテクなアプローチまで、さまざまな手法が登場しています(Diakopoulos, 2017, 2019)5。各記事は、アルゴリズム、そのデータおよびそのコードへのアクセスが可能であるかを含め、視座と文脈に応じて異なるアプローチを必要とする場合があります。

組織的な差別について暴露する記事は、オンラインで収集したデータを使った監査手法をとり、一方で、意図したポリシーが正しく実施されているかどうかを検証するにはコードレビューが必要になるかもしれません(Lecher, 2018)。デザイナー、開発者、データサイエンティストなどの企業内部の人間に話を聞くための伝統的なジャーナリストが行ってきた情報収集や、公文書請求をして影響を受けた個人を探し出すことは、これまでと同様に重要です。

この短い章でこれらの手法のすべてを深く説明することはできませんが、せめてここでは、ジャーナリストが監査技術を使ってアルゴリズムを調査する方法についてもう少し詳しく説明したいと思います。監査技術は、住宅市場などのシステムにおける社会的バイアスの研究に何十年も前から用いられており、最近ではアルゴリズムの研究にも応用されています(Gaddis, 2017; Sandvig et al, 2014)。基本的な考え方は、アルゴリズムへの入力を十分に異なる方法で変化させ、出力をモニターすれば、入力と出力を相関させて、アルゴリズムがどのように機能しているかの理論を構築できるというものです(Diakopoulos, 2015)。ある入力に対してアルゴリズムが[法や社会規範に]反する出力をしたならば、エラーを集計し、体系的に偏っているかどうかを確認するのに役立ちます。アルゴリズムが API やオンラインの Web ページを介してアクセスできる場合、出力データを自動的に収集することができます(Valentino-DeVries et al)。パーソナライズされたアルゴリズムに対応するため、アルゴリズムが独自の「視点」を適用する可能性があるさまざまな人からデータを収集するクラウドソーシングと、監査技術を結びつけることもできます。ドイツの AlgorithmWatch は、この手法を Google 検索結果のパーソナライズの研究に効果的に利用し、ブラウザのプラグインを介してデータを共有した 4000 人以上のユーザーから約 600 万件の検索結果を収集しました(本書のクリスティーナ・エルマーの章で詳しく説明されています)。ユーザーは自分のコンピュータにソフトウェアをダウンロードし、PYMK(知り合いかも機能)の結果を定期的に追跡します。ローカルなのでプライバシーは保たれます。そして記者は、自分の結果が気になる、あるいは意外だと思うユーザーから[報道の]ヒントを募ることができます(Hill & Mattu, 2018)。

アルゴリズムを監査することは、気の弱い人にはできません。情報が不足していると、どこから始めればいいのか、何を求めればいいのか、結果をどう解釈すればいいのか、アルゴリズムの挙動に見られるパターンをどう説明すればいいのかさえ分からないことがあります。また、アルゴリズムに期待されることを知り、それを定

義すること、そしてその期待が文脈によって、また世界的に異なる道徳的、社会的、文化的、法的な基準や規範によってどのように変化するかという課題もあります。

例えば、犯罪リスクを評価するアルゴリズムと、航空座席の値段を変えて請求するアルゴリズムとでは、公平性に対する期待が異なるかもしれません。ニュースになるようなミスや偏りを特定するためには、まず正常や偏りのない状態とはどのようなものかを定義する必要があります。2016 年の米国選挙の際に Google の検索結果のニュースソースを監査したときのように、データ駆動型［報道］の基準に基づいて定義されることもあります（Diakopoulos et al）。アルゴリズムに関する情報への法的アクセスの問題も出てきますが、これはもちろん司法権に大きく左右されます（Bhandari & Goodman, 2017）。米国では、情報自由法（FOI）により、国民の政府文書へのアクセスが規定されていますが、アルゴリズムに関する文書に対する各機関の対応は、決して一様ではありません（Brauneis & Goodman, 2018; Diakopoulos, 2016; Fink, 2017 を参照）。アルゴリズムに関する情報への一般のアクセスがより容易になるように、法改正が必要かもしれません。

情報不足、明確化しにくい期待、不確実な法的アクセスが完全に道をふさいでいるのではないのなら、アルゴリズムが非常に気まぐれなものでもあることを思い出してください。今日のバージョンのアルゴリズムは、昨日のものとはすでに違うかもしれません。例えば、Google は通常、年に 500～600 回、検索アルゴリズムを変更しています。アルゴリズムがどのように変化し、進化しているのかを理解するためには、その変化の大きさに応じて、時間をかけてモニタリングする必要があるでしょう。

今後に向けた提言

アルゴリズムを巡る説明責任報道を始め、最大限に活用するために、私は三つのことをお勧めします。まず、私たちは「Algorithm Tips」というリソースを開発しました。このリソースは、関連する手法や事例、教育リソースをキュレーションし、調査対象となる可能性のあるアルゴリズムのデータベースをホストしています（最初は米国連邦政府のアルゴリズムをカバーしていましたが、その後、より多くの国の司法権をカバーするように拡大しました）7。より多くのことを学び、プロジェクトを立ち上げるのに役立つリソースを探しているのであれば、それが一つの出発点になるかもしれません（Trielli et al）。

第二に、アルゴリズムの意思決定のメカニズムを正確に説明しようとするのではなく、アルゴリズムの結果や影響に焦点を当てることです。アルゴリズムによる差別（＝出力）を特定することは、その差別がどのようにして生じたかを正確に説明するよりも、最初の足がかりとして社会にとって価値があります。ジャーナリストは結果に焦点を当てることで、第一段階の診断を行い、他のステークホルダーが別の場所で説明責任を掘り下げるよう警告を発することができます。

最後に、ここで紹介したアルゴリズムを巡る説明責任の報道の多くはチームで行われていますが、それには理由があります。アルゴリズムを対象とする効果的な報道には、ジャーナリストが必要とする伝統的なスキ

ルである、取材やインタビュー、取材対象に関する知識、公文書の請求と文書の分析、結果を明確かつ説得力のある形で書くことなどが求められます。一方で、データのスクレーピングやクリーニング、監査調査の設計、高度な統計技術の使用など、多くの新しい能力も必要とされます。これらの異なる分野の専門知識は、明確なコミュニケーション、認識、リーダーシップがあれば、チーム内で分担することも、外部の協力者と連携することもできます。このようにして、[アルゴリズムの]手法の専門家は様々な分野の専門家と協力して、より多様な社会部門におけるアルゴリズムの力を理解することができるようになるのです。

脚注

1. www.propublica.org/series/machine-bias

2. アルゴリズムの説明責任という用語はもともと Diakopoulos, N. (2013, August 2). Sex, violence, and autocomplete algorithms. *Slate Magazine*. slate.com/technology/2013/08/words-banned-from-bing-and-googlesautocomplete-algorithms.html technology/2013/08/words-banned-from-bing-and-googles-autocompletealgorithms.html で提唱された。さらに詳しくは Diakopoulos, N. (2013, October 3). Rage against the algorithms. *The Atlantic*. www.theatlantic.com/technology/archive/2013/10/rage-against-the-algorithms/280255/を参照。

3. アクティビストとアーティストの枠組みは以下を参照: Brain, T., & Mattu, S. (n.d.). *Algorithmic disobedience*.samatt.github.io/algorithmic-disobedience/#/. アルゴリズムの力に関する学術的検証は: Bucher,

   T. (2018). *If . . . then: Algorithmic power and politics*. Oxford University Press. 批判的アルゴリズム研究の学術的な幅広いセクションは socialmediacollective.org/reading-lists/critical-algorithmstudies

4. www.dmlp.org/legal-guide/defamation

5. 方法論の選択肢についてのより完全な扱いについては、以下を参照のこと: Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press; see also: Diakopoulos, N. (2017). Enabling accountability of algorithmic media: Transparency as a constructive and critical lens. In T. Cerquitelli, D. Quercia, & F.Pasquale (Eds.), *Transparent data mining for big and small data* (pp. 25–43). Springer International Publishing.doi.org/10.1007/978-3-319-54024-5_2

6. algorithmwatch.org/de/ lterblase-geplatzt-kaum-raum-fuer-personalisierung-bei-google-suchen-zurbundestagswahl-2017/（ドイツ語）

7. algorithmtips.org

参考文献

Ananny, M. (2015). Toward an ethics of algorithms. *Science, Technology & Human Values*, *41*(1), 93–117.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *ProPublica*. www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bhandari, E., & Goodman, R. (2017). Data journalism and the computer fraud and abuse act: Tips for moving forward in an uncertain landscape. Computation+Journalism Symposium. www.aclu.org/other/data-journalism-andcomputer-fraud-and-abuse-act-tips-moving-forward-uncertain-landscape

Brain, T., & Mattu, S. (n.d.). Algorithmic disobedience. samatt.github.io/algorithmic-disobedience Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. *Yale Journal of Law & Technology*, *20*, 103–176.

Bucher, T. (2018). *If . . . then: Algorithmic power and politics*. Oxford University Press.
Diakopoulos, N. (2013, August 6). Algorithmic defamation: The case of the shameless autocomplete. Tow Center for Journalism. towcenter.org/algorithmic-defamation-the-case-of-the-shameless-autocomplete

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation ofcomputational power structures. *Digital Journalism*, *3*(3), 398–415. doi.org/10.1080/21670811.2014.976411

Diakopoulos, N. (2016, May 24). We need to know the algorithms the govern- ment uses to make important decisions about us. *The Conversation*. theconversation.com/we-need-to-know-the-algorithms-the-governmentuses-to-make-important-decisions-about-us-57869

Diakopoulos, N. (2017) Enabling Accountability of Algorithmic Media: Transparency as a Constructive and Critical

Lens. In T. Cerquitelli, D. Quercia, & F. Pasquale (Eds.), *Transparent data mining for Big and Small Data* (pp. 25–44).

Springer.

Diakopoulos, N. (2019). *Automating the News: How Algorithms are Rewriting the Media*. Harvard University Press.

Diakopoulos, N., Trielli, D., Stark, J., & Mussenden, S. (2018). I vote for—How search informs our choice of candidate.

In M. Moore & D. Tambini (Eds.), *Digital Domi- nance: The power of Google, Amazon, Facebook, and Apple* (pp. 320–

341). Oxford University Press. www.academia.edu/37432634/I_Vote_For_How_Search_Informs_Our_Choice_of_Candidate

Eubanks, V. (2018). *Automating inequality: How high-tech tools pro le, police, and punish the poor*. St. Martin' s Press.

Fink, K. (2017). Opening the government' s black boxes: Freedom of information and algorithmic accountability.

Digital Journalism, *17*(1).doi.org/10.1080/1369118X.2017.1330418

Gaddis, S. M. (2017). An introduction to audit studies in the social sciences. In M. Gaddis (Ed.), *Audit studies: Behind the scenes with theory, method, and nuance* (pp. 3–44). Springer International Publishing.

Gillespie, T., & Seaver, N. (2015, November 5). Critical algorithm studies: A reading list. *Social Media Collective*. socialmediacollective.org/reading-lists/

Hill, K. (2017a, October). How Facebook outs sex workers. *Gizmodo*. gizmodo.com/how-facebook-outs-sexworkers-1818861596

Hill, K. (2017b, November). How Facebook f igures out everyone you' ve ever met. *Gizmodo*. gizmodo.com/howfacebook- gures-out-everyone-youve-ever-met-1819822691

Hill, K., & Mattu, S. (2018, January 10). Keep track of who Facebook thinks you know with this nifty tool. *Gizmodo*. gizmodo.com/keep-track-of-who- facebook-thinks-you-know-with-this-ni-1819422352

Larson, J. (2016, October 20). Machine bias with Jeff Larson [Data Stories podcast]. datastori.es/85-machine-biaswith-jeff-larson/

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). How we analyzed the COMPAS recidivism algorithm.

ProPublica.www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Lecher, C. (2018, March 21). What happens when an algorithm cuts your health care. *The Verge*. www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, *31*(4), 611–627. https://doi.org/10.1007/s13347...

Lewis, S. C., Sanders, A. K., & Carmody, C. (2019). Libel by algorithm? Automated journalism and the threat of legal liability. *Journalism and Mass Communication Quarterly*, *96*(1), 60–81. https://doi.org/10.1177/107769...

Maheshwari, S. (2017, November 4). On Youtube Kids, startling videos slip past f ilters. *The New York Times*.

www.nytimes.com/2017/11/04/business/media/youtube-kids-paw-patrol.html

O' Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014, May 22). Audit- ing algorithms: Research methods for detecting discrimination on Internet platforms.

International Communication Association preconference on Data and Discrimination Converting Critical Concerns into Productive Inquiry, Seattle, WA.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2). https://doi.org/10.1177/205395...

Stark, J., & Diakopoulos, N. (2016, March 10). Uber seems to offer better service in areas with more White people.

That raises some tough questions. *The Washington Post*. www.washingtonpost.com/news/wonk/wp/2016/03/10/uber-seems-to-offer-better-service-in-areas-with-morewhite-people-that-raises-some-tough-questions/

Trielli, D., & Diakopoulos, N. (2017, May 30). How to report on algorithms even if you're not a data whiz. *Columbia*

Journalism Review. www.cjr.org/%20tow_center/algorithms-reporting-algorithmtips.php

Trielli, D., Stark, J., & Diakopoulos, N. (2017). *Algorithm tips: A resource for algorithmic accountability in Government*. Computation + Journalism Symposium.

Valentino-DeVries, J., Singer-Vine, J., & Soltani, A. (2012, December 24). Websites vary prices, deals based on users' information. *The Wall Street Journal*. https://www.wsj.com/articles/SB10001424127887323777204578189391813881534

ソーシャル・ウェブでストーリーを伝える

文：ラム・トゥイ・ヴォー

私たちは、歴史上最大のデータ生産者となっています1。オンラインでのクリック、タブレット端末でのスワイプ、スマートフォンでのタップなど、ほぼすべての操作が、仮想リポジトリにデータポイントを生成しています。Facebook は、20 億人以上の人々の生活に関するデータを創出しています。Twitter では、毎月 3 億 3000 万人以上のユーザーの行動が記録されています。マサチューセッツ工科大（MIT）のある研究によると、アメリカの平均的なオフィスワーカーは 1 日に 5GB のデータを生成しているそうです2。それは 2013 年のことですが、今もその勢いは衰えていません。より多くの人々がオンラインで生活し、スマートフォンが世界中のこれまで接続されていなかった地域に浸透していくにつれ、このストーリーの宝庫はますます大きくなっています。

多くの研究者は、ソーシャルメディアのユーザーを、逸話や単一の接触点として、それぞれ独立した対象であるかのように扱う傾向があります。しかし、一握りのユーザーとその個別の投稿を扱うことは、何億人ものユーザーとその相互作用の可能性を無視することになります。研究者やジャーナリストは、大規模なデータ整理の専門知識や分析技術をまだ身につけ始めたばかりで、ソーシャルメディアのユーザーやプラットフォームが生み出す膨大な量のデータには、語られるべき物語がたくさんあります。最近の出来事からも、記者がソーシャル・ウェブをよりよく把握することが重要になっていることがわかります。

2016 年の米大統領選や Brexit に対するロシアの干渉、ヨーロッパの国々やミャンマーにおける Facebook での反イスラムのヘイトスピーチの危険な広がり、世界のリーダーによる Twitter の強引な利用など、これらの動きは、ソーシャルメディアのデータの総体としての有用性と落とし穴について、優れたリテラシーを身につける必要性がますます高まっていることを示しています。

ジャーナリストはソーシャルメディアのデータをどう利用すべきか

報道にソーシャルメディアを役立てる方法はさまざまですが、ソーシャルメディアのプラットフォームから得られるデータを二つのレンズで見てみるとよいでしょう。第一に、ソーシャルメディアは、個人とその行動をよりよく理解する代理人として利用できます。公的発言であれ、私的なやり取りであれ、多くの人々の行動は今日、テクノロジーによって媒介・拡散され、オンライン上に痕跡を残しています。政治家やその他の重要人物を監視する時には特に有効です。彼らの意見は、政策を暗示したり、株価の急落や重要人物の解雇のような結果につながったりするからです。第二に、Web は、ソーシャルプラットフォーム上でストーリーが展開される、それ自体がひとつのエコシステムであると考えることができます。ただし、人間や自動操作によって動かされています。誤報キャンペーン、アルゴリズムによって歪められた情報世界、トロール攻撃などは、すべてソーシャル・ウェブ特有の現象です。

ソーシャルメディアのデータは、抽象的に語るのではなく、特定のストーリーを語るためにどのように利用できるかという文脈で理解すると、より役に立つかもしれません。以下のセクションでは、ソーシャルメディアのデータを利用したいくつかのジャーナリズム作品について説明します。

公人を理解する：説明責任報道に役立つソーシャルメディアのデータ

著名人や公人にとって、ソーシャルメディアは一般の人々に直接働きかける手段となっています。ニュースメディアとのインタビューやプレスリリース、記者会見といった従来の方法ではなく、ステータスの更新やツイート、投稿などを利用するようになっています。

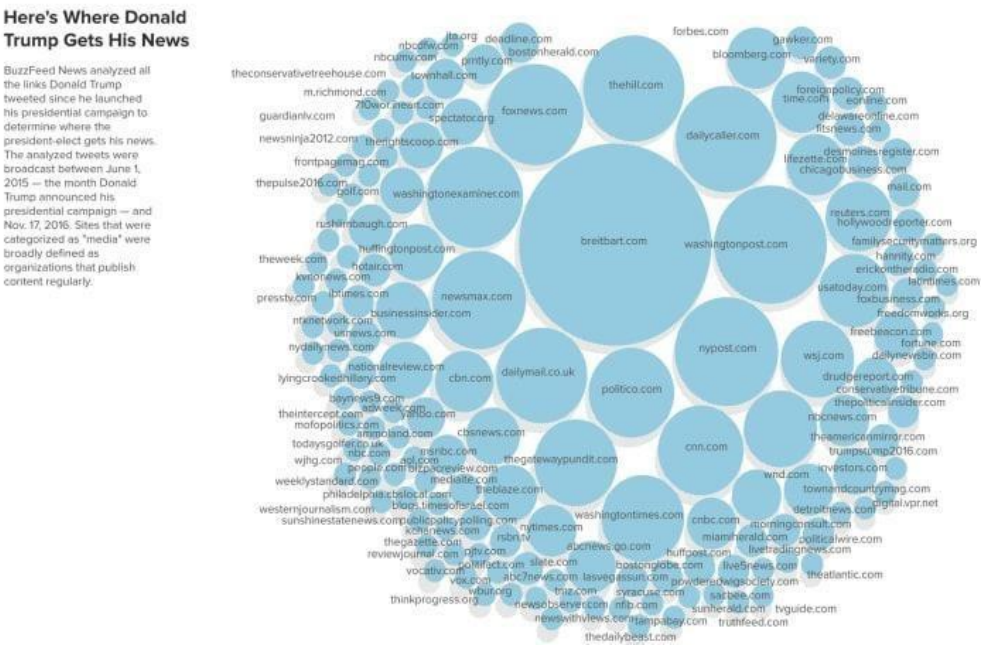


Figure 32.1. トランプが大統領選挙期間中にツイートしたメディアリンクのスナップショット 出典：バズフィード

しかし、政治家による、このような自己投影ともいえる公言は拘束力のある声明となり、有力な政治家なら、未公表の政策の前触れとなるかもしれません。政治家の仕事の一部は公共に関わることなので、政治家のソーシャルメディアアカウントを調査するのは、その人のイデオロギー的な考え方をより深く理解することに役立ちます。同僚のチャーリー・ワーゼルと私は、ドナルド・トランプの 2 万件以上のツイートを集めて分析し、次のような疑問に答えました。「彼はどのような情報を発信しているのか、そしてその情報は、彼が消費する可能性のある情報の代わりになりうるのか？」ソーシャルメディア上のデータポイントは、その性質や、これらのデータセットが不完全であり、個人の解釈に委ねられていることもあって、私たちの実際の姿の完全な引き写しではありません。しかし、それらは補完的に役立つことがあります。トランプ大統領がオンラインでブライトバートと提携していることは、実生活でスティーブ・バノンと強い絆で結ばれていることを早くから示していました。彼が「The

Conservative Tree House」や「News Ninja 2012」などの小規模な保守系ブログをリツイートしていることは、おそらく彼の「主流メディア」への不信感を示唆しているのでしょう。

人間の行動をたどる

ツイートや Facebook のオープンな投稿のような公的・半公的なコミュニケーションは、人々が他者に対して自分をどのように表現しているかを知ることができますが、プライベートメッセージや Google 検索、ジオロケーションデータのように、壁の向こうにあるソーシャルプラットフォームに存在するデータもあります。OKCupid の共同設立者であり、『Dataclysm』という本の著者であるクリスチャン・ラダー（2014）は、この種のデータについて、かなり的確な表現をしています。これらは、私たちが「『誰も見ていない』と思っている」ときの行動が記録された統計なのです。ソーシャルプラットフォームを利用することで、人は自分の行動の縦断的なデータを作り出すことになります。このような個人的なデータの山から、それを作成した人の範囲を超えて多くのことを推測するのは難しいものの、この種のデータはある人物のストーリーを伝えようとするときには非常に強力です。私はこのようなアプローチを「定量化された自撮り」と呼びたいと思っています。これは、モーリーン・オコナーが私の作品を紹介する際に作った言葉です。

［香港の民主派団体「デモシスト」の］ジェフリー・ゴー［敖卓軒］のストーリーを例にしてみましょう。香港出身のニューヨーク大生であるゴーは、2014 年 9 月初めに故郷の香港で民主化デモが始まったとき、行動を起こさなければならないと感じました。ニューヨークやワシントン D.C.に住む香港在住の他の人々に声がけし、最終的には世界中の 86 都市で抗議運動を組織することになりました。彼の物語は、ある問題に対する世界的な怒りから広まる多くの運動の象徴でもあります。ゴーはアルジャジーラ・アメリカの記事のために、個人的な Facebook の履歴の調査を許可してくれました（Vo, 2015）。彼が香港の別の中核的指導者と交わしたメッセージをスクレーピングし、2 人と他の指導者が政治活動について意見を交わした 10 の異なるチャットルームを発見しました。下の図（figure32.2）は、彼らのコミュニケーションの波を記録したものです。まず、「香港警察が平和的なデモ隊に催涙ガスを投げつけた」というニュースが世間を騒がせたとき、通信が急増しました。そして、ベージュ色のチャットルームが出現します。このチャットルームは、最初のニュースの後、中心組織が政治活動を計画するためのチャットルームとなりました。ここでは、ゴーが共同主催者のアグネス・チョウ［周庭］と初めて会ったときのことを振り返ることができました。ゴー本人は最初のやり取りを覚えていませんでしたが、Facebook のアーカイブのおかげで、ゴーがチョウと最初に交わした会話を再現することができました。

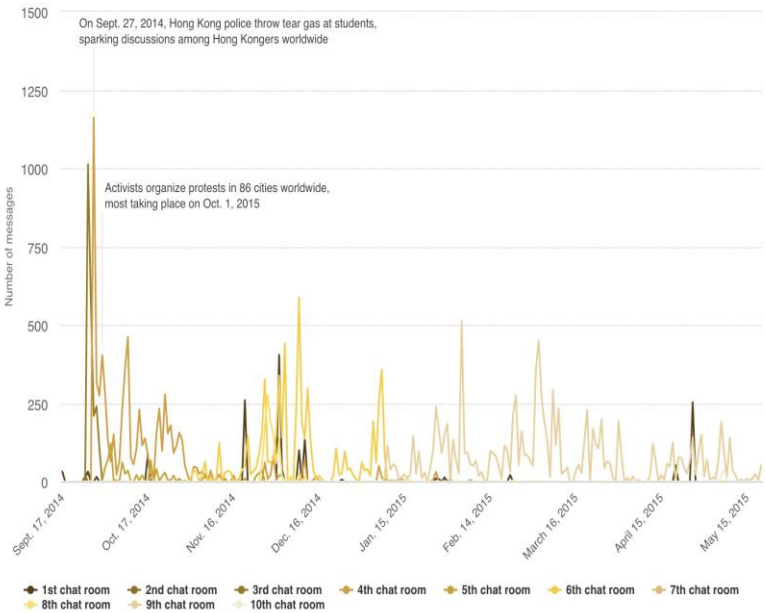


Figure 32.2. United for Democracy: Global Solidarity with Hong Kong の Facebook グループ。Facebook のデータはジェフリー・ゴーの好意により提供されている　出典：バズフィード

ゴーの政治運動家としての道筋は個人的要素が強く、彼の運動に参加したすべての人を代表するものではないことは明らかですが、デジタル時代の政治運動家の歩みを象徴しています。

オンライン生態系に特有の現象

私たちの交流の多くは、もっぱらオンライン・プラットフォーム上に移行しています。社会的行動の多くは、オンラインとオフラインが混ざり合っていますが、オンライン上の人間は強力なツールに助けられているため、オンライン環境は依然として非常に特殊なものとなっています。まず、いじめがあります。いじめは間違いなく人類と同じくらい長い歴史を有しています。しかし今では、いじめっ子たちは何千人もの他のいじめっ子たちに助けられており、彼らは瞬く間に声をかけることができます。いじめっ子は、検索エンジンを使ってその人のデジタル上の人生を追跡、時にはオンライン上のペルソナにまで遡ることができます。そして、彼らには増幅する手段があるのです。1 人のいじめっ子が廊下の向こうで耳をつんざくほどの声で叫ばなくても、何千人ものいじめっ子が一斉に襲ってきます。それが荒らしの本質です。

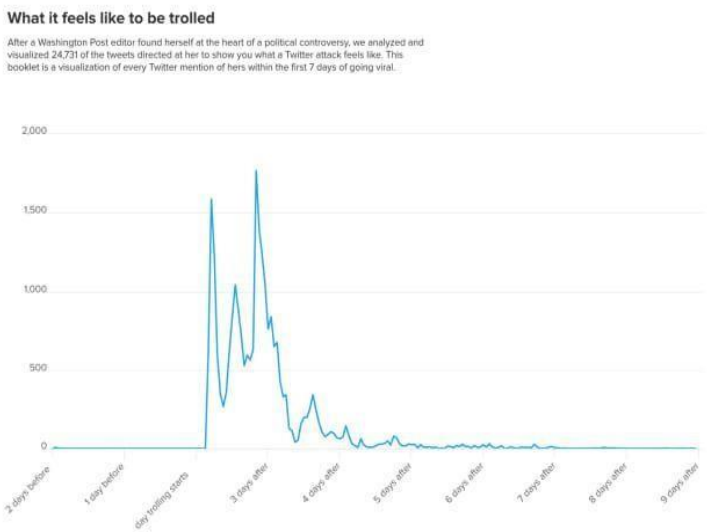


Figure 32.3 – ドリス・チュオンへの攻撃当日からの Twitter メンションチャート　出典：バズフィード https://www.buzzfeednews.com/article/lamvo/heres-what-it-feels-like-to-be-trolled-in-trumps-america

ワシントン・ポストの編集者ドリス・チュオンは、気付くとネット上の政治的な論争のただ中にいました。数日の間に、荒らし（と、彼女を擁護する多くの人々）は彼女に 2 万 4731 件の Twitter メンションを浴びせました。インターネット上で罵詈雑言を浴びせかけられることは、無視できないほど長い間、何らかの精神的負担を強いるものです。荒らしは、他の多くのネット上の攻撃と同様に、今では有名無名にかかわらず、誰もが悩まされる問題となっています。同性愛者のカップルにウェディングケーキを用意することを拒否したケーキ屋のように、Yelp のレビューが炎上することから、旅行に出発する直前に HIV と南アフリカ人に関する不幸なジョークを言った広報担当者ジャスティン・サッコが炎上によって解雇され、公衆の面前で恥をかかされたことまで、私たちの日常生活に影響を与える多くのストーリーが、最近ではオンラインで起きています。

情報戦争

ソーシャルメディアの登場とユビキタス化は、私たちの生活に新たな現象をもたらしました。バイラリティです。

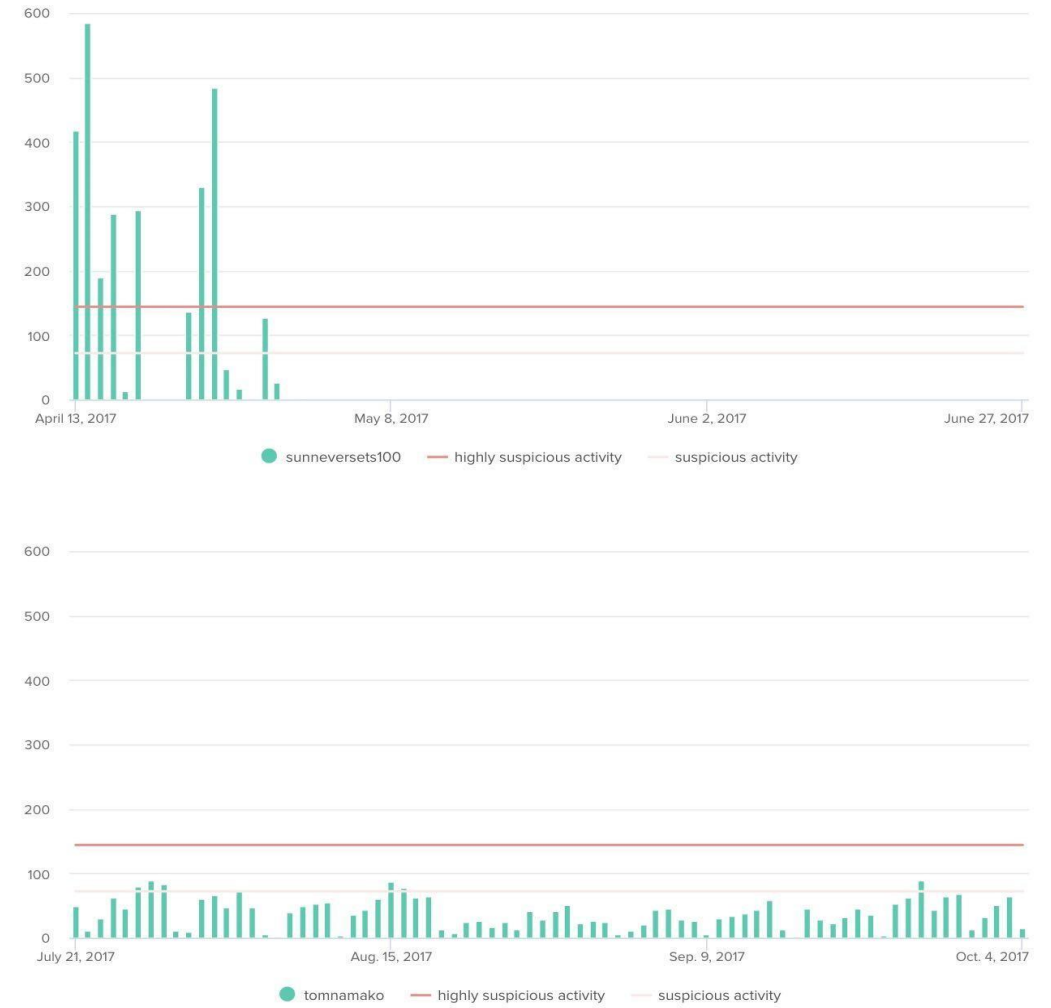


Figure 32.4. バズフィードは、自社の編集者である@tomnamako の Twitter データと、ボット的な活動を見せた複数のアカウントのデータを比較し、ペルソナや行動の違いを浮き彫りにした。上の最初のグラフは、B 編集者の直近の 2955 ツイートが数カ月にわたって均等に分布していることを示している。彼の 1 日のツイート数は、デジタル・フォレンジック研究所が疑わしい活動レベルと指定した 1 日あたり 72 ツイートをほとんど上回っていない。二つ目のグラフは、このボットの直近の 2955 件のツイートを示している。1 日で 584 件を記録するなど、日常的に不審なツイートを連発し、その後、突然停止したようだ

ソーシャルシェアリングによってあらゆる種類のコンテンツが、数百人だけでなく数百万人の人々に見られるようになったので、高価なマーケティングキャンペーンやテレビの放送時間枠を購入しなくてもよくなりました。しかしそれは、多くの人が偽のフォロワーや購入したフォロワー、ボットやサイボーグ（Vo、2017 a)のような（半）自動化されたアカウントを使ってアルゴリズムを操作するようになったことをも意味します。

ボットは元来悪ではありません。気まぐれな俳句やセルフケアのヒントで私たちを楽しませてくれるボットもたくさんいます。しかし、Atlantic Council のフェローで、ボット・アーミーを長年研究してきたベン・ニモがバズフィードの記事で私に語ったように、「 ボットは あらゆる議論をとてつもなく歪める可能性を秘めている・・・6 人のグループを 4 万 6000 人のグループに見せかけることができる」のです。ソーシャルメディアのプラットフォーム自身も、「問題のあるボット」を定義し、それを取り締まる責任を認識しなければならない重要な時期に来ています。一方で、ジャーナリストは、オンライン上で増え続ける非人間的存在とその力を認識すべきです。自動化されたアカウントに関するある記事では、人間のツイートとボットのツイートを比較してみました（Vo, 2017b）。あるアカウントがコード化されたスクリプトによって操作され、人間ではないかどうかを判断する確実な方法はありませんが、ユーザーのさまざまな特徴を見て、その行動が疑わしいものであるかどうかを確認する方法はあります。私たちが調べた特徴の一つは、アカウントの活動量です。そこで、実際の人間とボットの行動を比較してみました。最も忙しい日の最も忙しい時間帯に、私たちが調べたボットは 200 回以上ツイートしていました。一方、人間は 21 回しかつぶやきませんでした。

ソーシャルデータの取得方法

ソーシャル・ウェブからデータを取得する方法は、大きく分けて三つあります。API、パーソナルアーカイブ、スクレーピングです。API データストリームのような公式チャネルが提供するデータの種類は非常に限られています。ソーシャルメディア企業は、消費者行動に関するデータを［膨大に］保管しているのに、API を通じて提供しているのはその一部にすぎません。Facebook の場合、かつては研究者が公開ページやグループのデータを入手することができましたが、ケンブリッジ・アナリティカのスキャンダルを受けて同社がデータの利用を制限したため、この種のデータをマイニングすることはできなくなりました。Twitter の場合、アクセスはユーザーのタイムラインからのツイート数や一定期間内の検索に制限されがちです。さらに、ユーザーがオンライン上の自分の人格や行動について要求できるデータの種類にも制限があります。 Facebook や Twitter などの一部のサービスでは、ユーザーがオンライン上の自分を構成するデータ（投稿、メッセージ、プロフィール写真など）の履歴をダウンロードすることができますが、そのデータアーカイブには、各ソーシャルメディア企業が保有する自分の情報がすべて含まれているとは限りません。例えば、ユーザーがクリックした広告は 3 ヶ月前までしか見ることができないため、ロシアがスポンサーとなっている投稿をクリックしたかどうかを知るのは非常に困難です。

最後に重要なことでが、スクレーピングによってソーシャルメディアのデータをプラットフォームから抽出することは、多くの場合、利用規約に反します。ソーシャルメディアをクレーピングすると、ユーザーはサービスから追い出され、訴訟に発展する可能性もあります（Facebook, Inc. v. Power Ventures, Inc.）。

ソーシャルメディアプラットフォームにとって、スクレーパーを訴えることは金銭的に理にかなっているかもしれません。ソーシャルメディアプラットフォームがユーザーについて収集した情報の多くは、直接［自分たちが使うた

め]ではなく、企業や広告主が広告やマーケティングを通じて利益を得るために使われます。例えば、競合他社がFacebook から情報を収集すれば、同等のプラットフォームを構築できてしまいます。

しかし、訴訟は、経済的な動機によるデータスクレイパーだけでなく、研究目的でソーシャルメディアプラットフォームから情報を収集したいと考えている学者やジャーナリストも、意図せず抑止してしまうかもしれません。ジャーナリストは、これらのストーリーをどのように報道し、伝えるかについて、より創造的になる必要があるかもしれません。ボットを購入してオンラインでの行動をよりよく理解するとか、Facebook の仕組みをよく理解するために Facebook の広告枠を購入することが考えられます(Angwin et al.)。どんな手段であれ、ソーシャルメディア企業が設定した枠の内外で活動することは、刻々と変化するサイバー環境をナビゲートするジャーナリストにとって大きな課題となるでしょう。

ソーシャルメディア・データが苦手とするもの

ソーシャルデータの世界を理解するためには、その注意点についても目配りする必要があります。

ソーシャルメディアを使っている人、使っていない人を理解する。ソーシャルメディア・データの最大の問題点は、Twitter や Facebook で見かける人々が、オフラインのより広範な人々を代表するサンプルではないことです。Facebook や Twitter のアカウントを持っている人は大勢いますが、ジャーナリストは、オンラインで表明された意見が一般の人々を代表するものであると考えることには注意すべきです。2018 年のピュー・リサーチ・センターの調査によると、ソーシャルメディアの利用状況はプラットフォームによって異なります(Smith & Anderson, 2018)。オンラインの米国成人の 3 分の 2 以上が YouTube や Facebook を利用している一方で、Twitter を利用しているのは 4 分の 1 以下です。この種のデータは、ミャンマーの特定の政治家が広めたヘイトスピーチの検証や、陰謀論メディア Infowars の報道を時系列で検証するなど、具体的で具体的なストーリーにより強力な力を発揮します。

すべてのユーザーがありのままの姿をしているわけではない。さらに、すべてのユーザーが必ずしも一人の人間を代表しているわけではありません。自動化されたアカウント(ボット)や、半自動・半人間的にコントロールされたアカウント(サイボーグ)が存在します。また、複数のアカウントを運用しているユーザーもいます。繰り返しになりますが、経済的・政治的な利益のために情報の流れを操作する多数のアクターが存在することを理解することは、ソーシャルメディアのデータを一括して見る際に重要な点です。(ただし、メディアと情報の操作というこのテーマ自体、ジャーナリストがより洗練された方法で伝えようとしているところです)。

声高な人の専制。最後になりますが、すべての人の行動が測定されるわけではないことを認識することが重要です。膨大な数の人々が「沈黙」を選んでいます。より穏健な声が記録されることが少なくなると、記録されてアルゴリズムにフィードバックされるのは過激な反応だけとなり、すでに存在している声の大きい人の優位性が不釣り合いに増幅されてしまうのです。つまり、Facebook や Twitter などのプラットフォームがアルゴリズム

によってソーシャルフィードに表示するコンテンツは、おおむね、声を上げた人の「いいね！」や「リツイート」、「コメント」に基づいています。声を上げなかった人たちは、このプロセスの中で不釣り合いにかき消されてしまいます。そのため、私たちは、測定されるものと同様に、測定されないものにも注意を払う必要があります。測定されたデータポイントや測定されないデータポイントの結果として、情報がどのようにランク付けされ、表面化されるかに目を配りましょう。

脚注

1. この章の以前のバージョンは次のサイトで公開されている: source.opennews.org/articles/what-buzzfeed-newslearned-after-year-mining-data-/

www.niemanlab.org/2016/12/the-primary-source-in-the-age-of-mechanical-multiplication/doi 10.5117/9789462989511_ch32

参考文献

Angwin, J., Varner, M., & Tobin, A. (2017, September 14). Facebook enabled advertisers to reach "Jew haters."

ProPublica.www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters

Facebook, Inc. v. Power Ventures, Inc., No. 13-17102 (United States Ninth Circuit July 12, 2016).

https://caselaw. ndlaw.com/su...caselaw. ndlaw.com/summary/opinion/us-9th-circuit/2016/07/12/276979.html

Rudder, C. (2014). *Dataclysm: Who we are (When we think no one's looking)*. Fourth Estate.

Smith, A., & Anderson, M. (2018, March 1). *Social media use 2018: Demograph- ics and statistics*. Pew Research Center. www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/

www.technologyreview.com/2013/05/07/178542/%20has-big-data-made-anonymity-impossible/Tucker, P. (2013, May 7). Has big data made anonymity impossible? *MIT Technology Review*.

Vo, L. T. (2015, June 3). The umbrella network. *Al Jazeera America*. projects.aljazeera.com/2015/04/loving-longdistance/hong-kong-umbrella-protest.html

Vo, L. T. (2016). The primary source in the age of mechanical multiplication. *Nieman Lab*. www.niemanlab.org/2016/12/the-primary-source-in-the-age-of-mechanical-multiplication/

Vo, L. T. (2017a, October 11). Twitter bots are trying to in uence you. These six charts show you how to spot one.

BuzzFeed News. www.buzzfeednews.com/article/lamvo/twitter-bots-v-human

Vo, L. T. (2017b, October 20). Here's what we learned from staring at social media data for a year. *BuzzFeed*. www.buzzfeed.com/lamvo/heres-what-we-learned-from-staring-at-social-media-data-for

Vo, L. T. (2017c, October 20). What we learned from staring at social media data for a year. *Source*.

source.opennews.org/articles/what-buzzfeed-news-learned-after-year-mining-data-/

デジタル・フォレンジック：Google Analytics の ID を再利用する

文：リチャード・ロジャース

概要

Google Analytics や AdSense の ID を共有している Web サイトを用いてネットワークを発見する手法を説明します。

キーワード：デジタル・メソッド、デジタル・フォレンジック、匿名情報源、ネットワーク・マッピング、Google Analytics、データジャーナリズム

2015 年 7 月、ある調査ジャーナリストが、ウクライナに関する偽情報を提供するロシアの Web サイトの秘密ネットワークを発見しました。この発見は、2016 年の米国大統領選挙での国家主導のキャンペーンを予感させただけでなく1、データジャーナリストや社会研究者がネットワーク発見技術を知る機会になりました（Alexander, 2015）。同じ Google Analytics ID を共有している Web サイトはどれでしょう（figure33.1 参照）？同じ ID を持つ Web サイトは、個人、組織、メディアグループなど、同じ登録者によって運営されています。ジャーナリストのローレンス・アレキサンダーは、2013～14 年に親ロシア派のウクライナ大統領を打倒して親欧米派の大統領を生み出した、ユーロマイダン抗議活動に関する情報を提供しているサイト、「emmaidan.com.ua」のソースが欠如していることをきっかけに、仕事を始めたそうです。その情報源を探し、「匿名性に興味をそそられた」彼は、サイトのソースコードを掘り下げました。

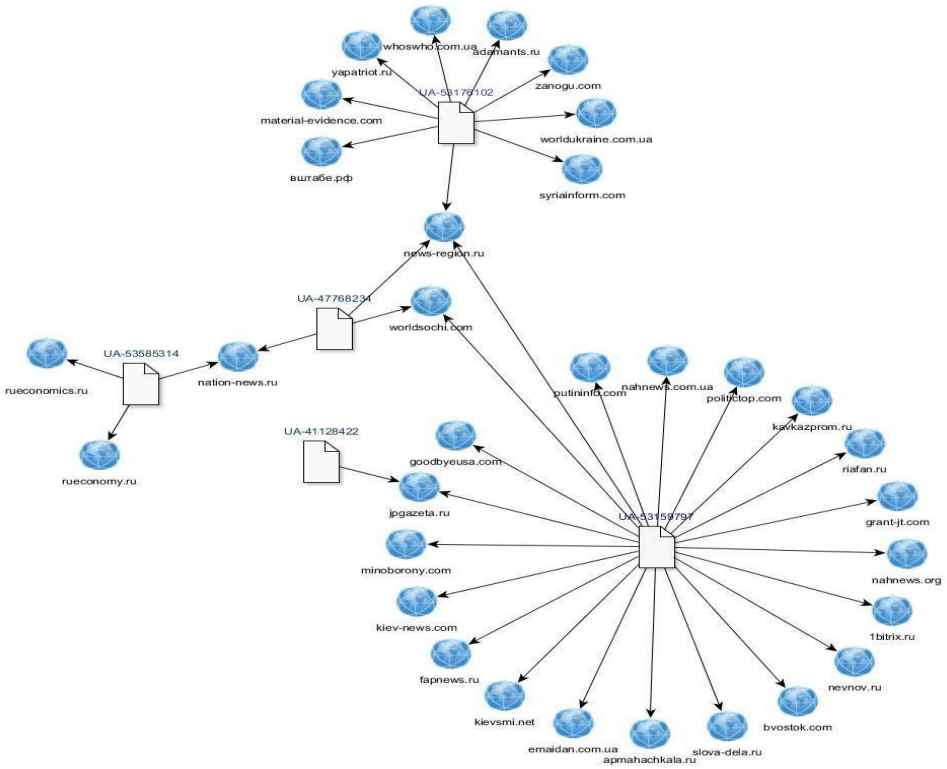


Figure 33.1. Google Analytics の(共有)ID から発見された Web サイトネットワーク 出典: アレキサンダー・ローレンス (2015.7.13). Open-source information reveals pro-Kremlin web campaign. Global Voices. https://globalvoices.org/2015/07/13/open-source-informationreveals-pro-kremlin-web-campaign/.

ソースコードから、Google Analytics の ID を見つけ、同じ ID を使用している他のサイトのリストを提供する逆引きソフトウェアに挿入しました 2。彼は、ある Google Analytics ID が八つの他のサイトにリンクしている(figure33.1 の上段)という(星型の)ネットワークを見つけました。いずれも反ウクライナの立場を取るものでした。それらのサイトの一つは、さらに別の Google Analytics ID を使用しており、これが、同じく同様の政治的影響力を持つ別のクラスタ(figure33.1 の右下)につながっていたのです。これらのドメインの WHOIS 記録を調べると、関連する電子メールアドレスが見つかり、続いて、ロシアのソーシャルネットワーキングサイトである「VKontakte」上の人物のプロフィールと写真が見つかりました。この人物の名前は、ロシア政府が支援する情報工作組織として知られるサンクトペテルブルクのインターネット・リサーチ・エージェンシーから流出した従業員リストに記載されていました(Chen, 2015; Toler, 2015)。アレキサンダーはデータポイント間のリンクを作成し、ロシアのトロール[荒らし]に名前と顔を与えました。また、彼の Pinterest の趣味欄にロシアの宇宙開発の写真があることを明らかにして、トロールに人間味を与えました。このトロールは宇宙飛行士のファンでもあるのです。アレキサンダーらは、いわゆる「オープンソース・インテリジェンス」(OSINT)ツールを「発見」に活用し(Google Analytics や逆引きソフトを再利用する手法もある)、コード、公文書、データベース、リーク

などのリンクを作り、つなぎ合わせて、特定の活動の「背後にいる人物」を説明しています(Bazzell, 2016)。「発見」とは、情報収集と暴露のための調査的、あるいはデジタル・フォレンジックの手法であり、個人、組織、メディアグループを特定してインタビューを試み、自分の仕事を説明する機会を与えるというものです3。ジャーナリストの発見とその相手による二重の説明が、語られるべきストーリーを構成します。その目的は、Web サイトのこんがらがったコードから、秘密裏に行われている政治活動を認めさせることにあります(Latour, 2005)。

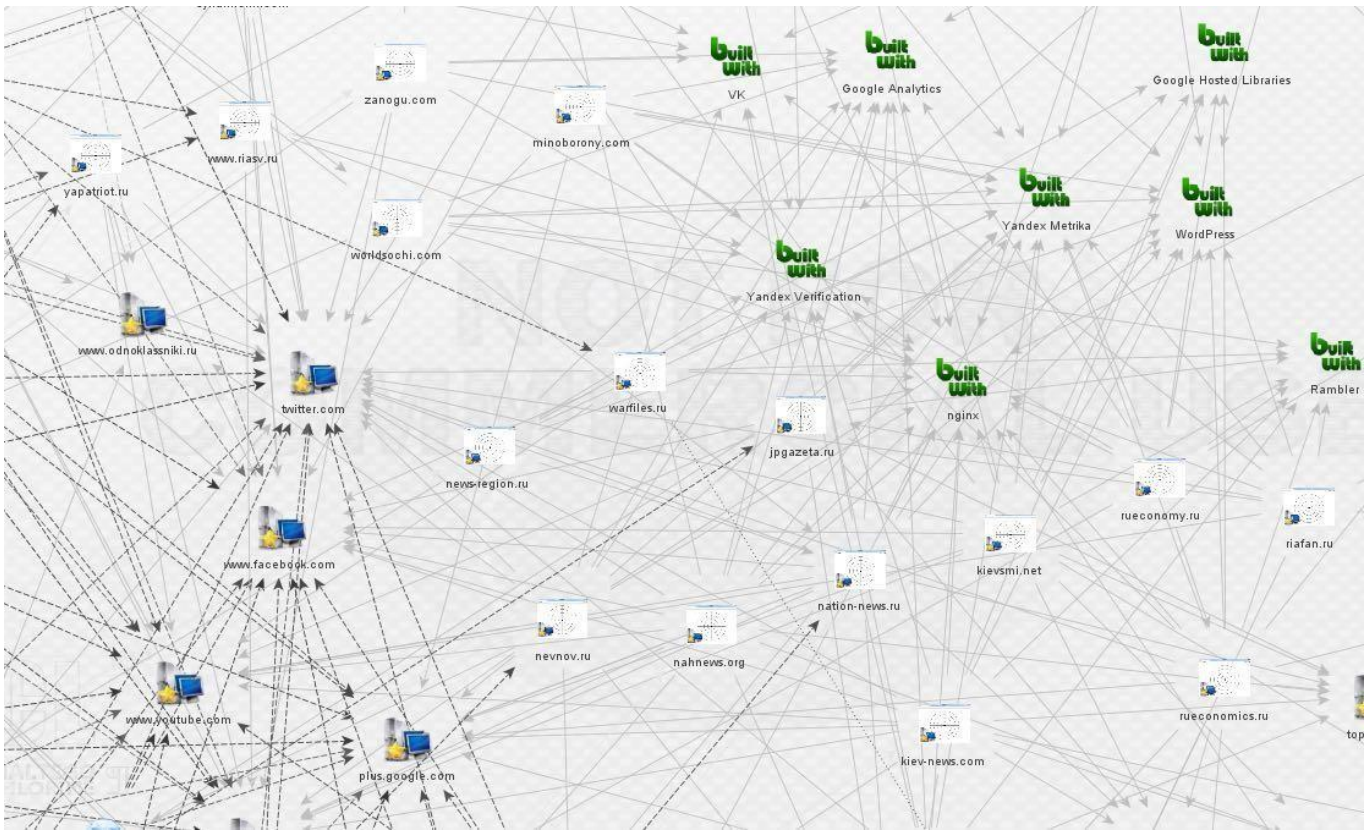


Figure 33.3. ネットワーク図として描かれたウェブサイト上の埋め込み型デジタルオブジェクト　出典:アレキサンダー・ローレンス(2015.7.13). Opensource information reveals pro-Kremlin web campaign. Global Voices. https://globalvoices.org/2015/07/13/open-sourceinformation-reveals-pro-kremlin-web- campaign/.

Google Analytics の ID 検出は、エクスプロイト(個人を特定できるデータへの侵入口)を利用して匿名のオンラインアクターの仮面を剥ぐという、開発者の予期せぬ作業の系譜上にあります。Google Analytics の ID をマイニングしてネットワークの発見やマッピングを行うことは、ソフトウェアを開発者の意図しない形で社会調査に使用する再利用でもあるのです。

このテクニックの創始者である Wired のジャーナリスト、アンディ・バイオは、非常に攻撃的な内容を投稿する匿名のブロガーが、「通常の方法」で正体を隠していた話を紹介しています。このブロガーは、「ドメインレコード

の個人情報を隠し、他のサイトとは異なる IP アドレスを使用し、インストールした WordPress から共有リソースを削除する」といった方法をとっていました（Baio, 2011）。

このブロガーは、Google Analytics の ID を自分が運営する他のサイトと共有しており、それが丸見えだったため、バイオは彼を特定できたのです。この発見と正体を暴く手法についての注意喚起は、メキシコの麻薬カルテルを監視しているといったような正当な理由のある他の匿名ブロガーに対し、その発見によって危険にさらされたり、命を落としたりしないよう安全の手引きを提供することで締めくくられています。また、匿名性が担保されているのかをテストし、脆弱性や悪用の可能性がある場合は、オンラインで覆面調査をしているジャーナリストに知らせることもできます。

最後に、Google Analytics の ID を使ってネットワークを発見するための探査プロトコルを以下のリストにまとめてみました。

■ソースを提供していないサイトのリストを作成する。Google Analytics と AdSense の ID を探す。

■URL リストを dnslytics.com などの逆引きソフトに挿入する。同じ ID を持つサイトを探す。

■ID を共有している Web サイトをテーマ別に分類し、特徴を明らかにする。Gephi によるネットワークの可視化を検討する。

脚注

1. A longer version of this chapter is available in Rogers, R. (2019). *Doing digital methods*. SAGE. The author would like to acknowledge the groundwork by Mischa Szpirt. For more on this approach, see Rogers, R. (2019). *Doing digital methods*. SAGE (Chapter 11), and Bounegru, L., Gray, J., Venturini, T., & Mauri, M. (Comp.) (2017). *A fijield guide to "fake news": A collection of recipes for those who love to cook with digital methods*. Public Data Lab (Chapter 3).

2. The lookup may also yield each website's IP address, Google AdSense ID, WHOIS domain record and other tellinginformation.

3. Digital forensics has its roots in the investigation of corporate fraud through techniques such as "data carving,"which enable the retrieval of deleted fijiles.

参考文献

Alexander, L. (2015, July 13). Open-source information reveals pro-Kremlin web campaign. *Global Voices*.

https://globalvoices.org/2015/07/13/open-source-information-reveals-pro-kremlin-web-campaign/

Baio, A. (2011, November 15). Think you can hide, anonymous blogger? Two words: Google analytics. *Wired*. www.wired.com/2011/11/goog-analytics-anony-bloggers/

Bazzell, M. (2016). *Open source intelligence techniques: Resources for searching and analyzing online information* (5th ed.). CreateSpace Independent Publishing Platform.

Chen, A. (2015, June 2). The agency. *The New York Times Magazine*. www.nytimes.com/2015/06/07/magazine/theagency.html

Latour, B. (2005). From realpolitik to dingpolitik—An introduction to making things public. In B. Latour & P. Weibel (Eds.), *Making things public: Atmospheres of democracy* (pp. 14–41). MIT Press. http://www.bruno-latour.fr/nod...

Toler, A. (2015, March 14). Inside the Kremlin troll army machine: Templates, guidelines, and paid posts. *Global Voices*.globalvoices.org/2015/03/14/russia-kremlin-troll-army-examples/

データ調査のためのアプリとアフォーダンス

文：エスター・ウェルトフレーデ

概要

データ調査のためのアプリとプラットフォームの関係を探ります。

キーワード:アプリ、ソーシャルメディアプラットフォーム、デジタルメソッド、データインフラ、データジャーナリズム、データ調査

最近、Facebook からデータを抽出するツール Netvizz が、Facebook のページ公開コンテンツアクセス機能にアクセスできなくなりました。開発者であるデジタルメソッド研究者のベルンハルト・リーダーが9年間維持してきた Facebook API との不安定な関係に終止符が打たれたのでした1。Netvizz の終焉は、プラットフォームがアプリケーション・プログラミング・インターフェース（API）や開発者ポリシーを通じたデータ収集をさらに制限しようとする、デジタル調査・研究におけるより大きな変化を示しています。ケンブリッジ・アナリティカの手法の実際の有効性は疑問視されているものの（Lomas, 2018; Smout & Busvine, 2018）、このスキャンダルはソーシャルメディアにおけるプライバシーとデータ保護に関する議論を巻き起こし、Facebook は自社プラットフォームからのデータへのアクセスをさらに制限することで対応しました。

2018 年 3 月の最初の発表以来2、Facebook がその大規模なアプリ群の中でデータアクセス制限を段階的に実施する中で、さまざまな目的でプラットフォームに依存するようになったサードパーティ関係者の膨大なネットワークが可視化されました。アプリが動かなくなったり、広告ターゲットが制限されたりしていますが、最も深刻な被害を受けているのはデジタル研究者のようです。データ収集を主な目的とするアプリが許可されなくなったからです。デジタル研究者は、公共の利益に資する研究なのに、その費用を負担させられると主張して、これらの変化に抵抗しました（Bruns, 2018）。

Netvizz 論文の参考文献リスト（Rieder, 2013）には 450 以上の出版物が掲載されていますが、実際にはその量を軽く超えているのです-このツールを利用している多くの学生の研究プロジェクトを考えてみてください。同様に、ベックマンが「API データへのアクセスがなければ存在し得なかった」研究のアドホックな目録を作成したところ3、ジャーナリズム、社会科学、その他のデジタル研究の出版物の印象的なリストができました。データアクセス制限がデジタル研究に与える影響を考慮して、著者らはこれらの動きがあった過去 10 年間を「API ベースの研究」（Venturini & Rogers, 2019）または「API 関連の研究」（Perriam et al, 2020）の時代と名付けました。

これらは、オンライン・プラットフォームが API を通じて提供するデータの抽出に基づくデジタルリサーチのアプローチと定義されています。社会調査のためのデータが用意されている API は、ソーシャルメディア

のデータを使った研究の敷居を下げただけでなくデジタル研究を試みるある世代の学生を増やしたことは言うまでもないことでしょう。

技術的なスキルを必要とせず、Web データ標準の場合は、データは比較的クリーンです。API を使った研究には当初から、API は利便性によって研究のアフォーダンスを操り、研究者の主体性に影響を与えているという批判がありました（Marres, 2017）。ここでは、ベンチュリーニとロジャース（2019）や Digital Methods Initiative による最近の「ポスト API 研究」の呼びかけを取り上げ、ソーシャルメディアのエコシステムにおける発展に対応して生じる機会に焦点を当てています4。ここで採用するデジタルリサーチの意味は、「メディアに従う」という方法論の原則に基づくもので、デジタル環境の発展に対応し、それらの方法と相互作用するものです。

以下では、最近の API 制限について、現在のプラットフォーム環境を形成する鍵となる様々な種類の社会工学的データを創造的かつ独創的に探求することの新たな必要性と可能性を論じることでアプローチします。続いて、デジタル研究者の議論を紹介し、アプリとプラットフォームの関係を研究するための方法論的視点を補足します。ストーリーテリングのためのソーシャルデータに関心のあるデータジャーナリスト（例えば、ラム・トゥイ・ヴォーの章を参照）に、ポスト API の現在において、プラットフォームとそのデータを巡る調査の足がかりを提供します。

デジタル手法は、どれもネイティブなデジタルデータ形式の使用を最適化する一連のデータ収集および分析技術を利用しています。これらは、デジタルメディアが社会生活に導入される過程で出現しました。デジタル手法の研究者は、革新的な方法でこれらのデータ形式を扱えるように、デジタルメディアに触発されたツールを開発しています。すなわちデジタル手法の歴史は、インターネットの主要なデータ形式やデータ構造の歴史をも物語っています。デジタル手法は、メディアの変化に適応し、それらを分析に取り入れています。以下では、データインフラとしてのプラットフォームを研究する視点を提案することで、ポスト API の研究アプローチに貢献します。

大規模なメディア・エコシステムに対するデータアクセス制限の影響は、プラットフォームのインフラとサードパーティアプリとの相互作用に関する高度で微妙な知識が切実に求められることを示しています。サードパーティの企業やアプリがソーシャルメディア環境でどのように動作するかについての知識だけでなく、データインフラのリテラシーを拡充すること（Gray et al.、2018）が必要なのです。

アプリとインフラとしてのプラットフォーム

プラットフォームデータの制限は、ソーシャルメディアをインフラとしてのプラットフォームへと発展させるための重要な要素です。このような動きは、企業とのパートナーシップに焦点を当てたデジタルエコシステムの進化を浮き彫りにしています（Helmond et al.、2019）。プラットフォームが選挙で果たした役割を 1 年にわたり否定的に報道され続け、ザッカーバーグは、「人々をつなぐ」から「社会インフラ」を構築するというプラットフォームの視点の変化をスケッチしたメモを投稿しました（Helmond et al.2019; Hoffmann et al.2018）5。社会インフラという概念は、市場の複数の側面を結びつけて価値を創造するプラットフォームの中核製品としての

社会活動を強調すると同時に、同社がソーシャルネットワークからデータインフラへと[重心を]移し、プラットフォームを自社のウェブサイトや 70 個のアプリの大規模なファミリーにまで拡張していることを示しています6。

このインフラの転換は、プラットフォームがデータインフラをサードパーティのアプリ、プラットフォーム、ウェブサイトに拡張し、内向きの統合を促進するという次のステップを示しています。インフラとしてのプラットフォームが注目されるようになっても(Plantin et al, 2018)、アプリがデータインフラ上やデータインフラ間でどのように動作するのかは、十分に研究されておらず、説明されていないことが多いのが実情です。

しかし、アプリはプラットフォーム環境内の日常的なプラクティスを継続的に変容させ、価値化しているのです。サードパーティ製のアプリに焦点を当てて、アプリの関係性を定義しましょう。サードパーティアプリとは、プラットフォームが所有または運営していない、外部の開発者によってプラットフォーム上に構築されたアプリケーションです。アプリがプラットフォームに接続すると、権限に応じてプラットフォームの機能やデータへのアクセスが許可されます。また、アプリストア、広告主、ユーザーなどの関係者は、アプリを複数の方法で同時に統合し、価値を高めることができます。言い換えれば、アプリには関連付けるという傾向があり、常にさまざまなデータインフラを関連付けようとするのです。

サードパーティ製アプリのこのような位置付けは、インフラとしてのプラットフォームの研究に特に適しています。ソーシャルメディアプラットフォームは、前述のように、ユーザーが生成したデータへのアクセスがますます制限されています。研究者は「ソーシャルデータ」とは何かを新たに考え、別の視点を開拓する必要があり、方法論上の課題となっているのです。ソーシャルメディアのプラットフォームが、API を介して構造化されたユーザー生成データをデジタルリサーチに提供しているのとは逆行して、アプリのデータソースは閉鎖的で独占的な特色をいっそう強めています。

難読化はソフトウェア工学 Matviyenko et al., 2015 年)で広く使用されている技術ですが、コードやデータを判読や利用ができないようにする取り組みは、デジタルリサーチに大きな影響を与えます。プラットフォームやアプリが実証研究をすり抜けようとしてもたらされる課題の増加を、同僚や私は「インフラストラクチャーの抵抗」と呼んでいます(Dieter et al.)。今日のデジタル研究で利用できるデータ形式は、デバイスベースのデータ(GPS など)、ソフトウェアライブラリ(ソフトウェア開発キット、SDK など)、ネットワーク接続(広告ネットワークなど)に至る、異種データ形式が特徴です。アプリはユーザーが生成したデータを収集することができますが、ほとんどの場合、オープンな API を介したアクセスを提供していないため、データリサーチに役立つ既存のデータが存在しません。以下では、三つの異なるボトムアップのデータ調査を紹介します。デジタル研究者やジャーナリストは、これらの調査を進めるために、さまざまな「リサーチアフォーダンス」(Weltevrede, 2016)を活用できます。リサーチアフォーダンスは、研究者の立場に立って、関心に沿った形でソフトウェア内で行動する可能性を追求したものです。このアプローチにより、独創的なデジタル手法の開発が可能になります(Lury & Wakeford, 2012; Rogers, 2013)。そのためには、アプリとプラットフォームの関係の技術的な形式やフォーマッ

トを再考し、その分析機会を探る必要があります。ここでは、同僚と私が行った最近の研究から、課題を指摘し、インフラとしてのプラットフォームに対する理解を深めるために、これらのデータソースを使ってどのような調査が可能か提案します。

フェイクのソーシャルインフラ

まずは、フェイクフォロワーと Facebook のソーシャルインフラとの関係について考察します。ソーシャルメディア環境におけるフェイクフォロワーの存在は、プラットフォームとデジタルリサーチの両面から注目されています。プラットフォーム側から見ると、多面的市場としてのプラットフォームの議論では、フェイクフォロワー市場はたいてい除外されます。フェイクフォロワーのマーケットは「面的」ではなく、もちろん「ファミリー」の一部でもありません。フェイクフォロワーは、プラットフォームは望ましくない誤用として認識する、非公式の関係を構築しています。これらの行為は、プラットフォームが意図したものではないものの、プラットフォームのメカニズムに便乗して機能しています。さらに、これらの行為は、重要な商品である社会的活動の価値を低下させるのです。

同僚と私は、そのデータセットの中で最も頻繁に使用されているアプリに焦点を当てて、Twitter での Brexit 国民投票までの流れを調査しました（Gerlitz & Weltevrede, 2019）（figure34.1 参照）。これらのアプリとその機能を体系的に分析することで、プラットフォームのガバナンス構造の中で自動化されたエンゲージメントとフェイクのエンゲージメントのメカニズムについての洞察が得られます。ヨハン・リンドキストとは、広範囲のプラットフォームでフェイク・エンゲージメントの売買を可能にする 1200 以上の二次流通プラットフォームを調査するプロジェクトを進めています。初期の調査では、フェイクフォロワーが技術的にどのようにプラットフォームと関係しているかを明らかにしました。公式のサードパーティアプリが API を介して接続している場合と、プラットフォームが非公式にソーシャルメディアのプラットフォームに接続している場合があります。こうして、研究は自動化された、あるいはフェイク起源の様々なデータに対応しなければならないということが明らかになりました。自動化された、またはフェイクのアカウントは［単なる］タイプやアクターとして扱うのではなく、メディアのアフォーダンスに関連して登場したものとして扱うべきです。Twitter のツイートが一般的に理解されるように一意的で固有のものではないのと同じく、アカウントも分散的でその時々の状況に応じた存在で、必ずしも人間のユーザーを代表するものではないのです（Gerlitz & Weltevrede, 2019）。

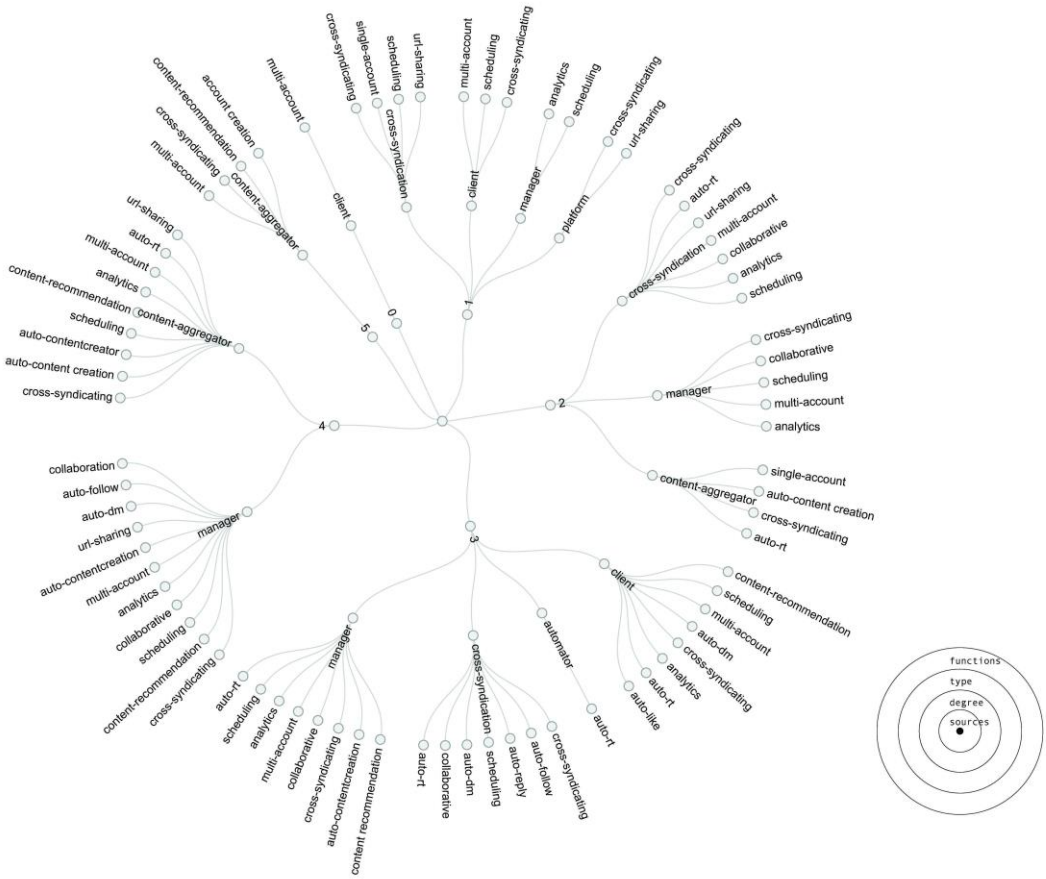


Figure 34.1. この樹形図は、2016 年 6 月 17 日～23 日のブレグジットデータセットにおけるデータソースの階層、自動化の度合い、種類とその機能を可視化したもの　出典：ゲーリッツとウェルトフレーデ(2019). What happens to ANT, and its emphasis on the socio-material grounding of the social, in digital sociology? In A. Blok, I. Farias, & C. Roberts (Eds.), Companion to Actor-Network Theory. Routledge. https://doi.org/10.4324/9781315111667-38

アプリとプラットフォームの関係

次に、アプリのデータインフラとしてのアプリストアについて考えてみます。現在、開発者やユーザーにとって、アプリへの主な入り口はアプリストアであり、ユーザーは個々のアプリを検索したり、アプリのコレクションやジャンルを区分けしたりすることができます。アルゴリズム研究（Rogers, 2013; Sandvig et al, 2014）の手法をもとに、Google Play や App Store における「ランキング文化」（Rieder et al, 2018）の技術的側面に迫ることもできるでしょう。このような取り組みは、アルゴリズム的・経済的な力と社会的影響の両方に関係しています。ランキングのメカニズムについての知識や、文化的コンテンツの流通にとってなぜそれが重要なのかを理解するのにも役立ちます。また、アプリストアは、アプリのコレクションやジャンルを区分けするのにも使われるので、アプリの視点からアプリとプラットフォームの関係を研究することができます。「Regramming the

Platform」(Gerlitz et al., 2019)では、同僚と私が、1 万 8500 本以上のアプリがプラットフォームの機能や特徴とどのように関連しているのかを調べました。この調査で得られた重要な知見の 1 つは、アプリの開発者が公式のプラットフォーム API を回避する創造的なソリューションを見出し、プラットフォームの公式ガバナンスシステムも回避しているということです(figure34.1)。インフラとしてのプラットフォームに対するアプリ中心のアプローチは、ソーシャルメディアのプラットフォームの周辺で開発されるサードパーティ製アプリ、それらのアプリによってサポートされ拡張されるプラクティスと機能、そしてアプリとソーシャルメディア・プラットフォームの間の乱雑で偶発的な関係を巡る洞察を提供します(Gerlitz et al., 2019)。

サードパーティのデータ接続

三つ目の調査では、アプリのソフトウェアと、外部のステークホルダーのデータインフラとの関連性を検討しました。こうした調査によって、ソフトウェアオブジェクトとしてのアプリが、外部のデータインフラをどのように埋め込むのか、また動的なデータフローをどうやってアプリに出入力するのか把握することができます(Weltevrede & Jansen, 2019)。アプリは、私たちには離散的で境界のあるオブジェクトのように見えますが、定義上はデータインフラのオブジェクトであり、データインフラ内で拡張・統合するためのプラットフォームに再配置されるものなのです。入口と出口のデータフローを活性化して探索するために、私たちは一種の「ウォークスルー法」を使用しました(Light et al.)。データの接続に焦点を当てた結果、どのデータがソーシャルメディアプラットフォームやモバイルプラットフォームからアプリに流れ込んでいるのかを可視化しました(figure34.2)。次に、アプリのデータを収益化したり、機能を改善したり、外部へホスティング配信したりするために、アプリが接続している広告ネットワーク、クラウドサービス、アナリティクスなどのサードパーティネットワークをマッピングしました(figure34.3)。アプリに出入りするデータの流れをマッピングすることで、データの循環と再結合の政治経済について、重要な洞察を得ることができます。データ接続がどのように確立され、進められ、どんな型のデータがどのサードパーティに転送されているかということが分かるのです。

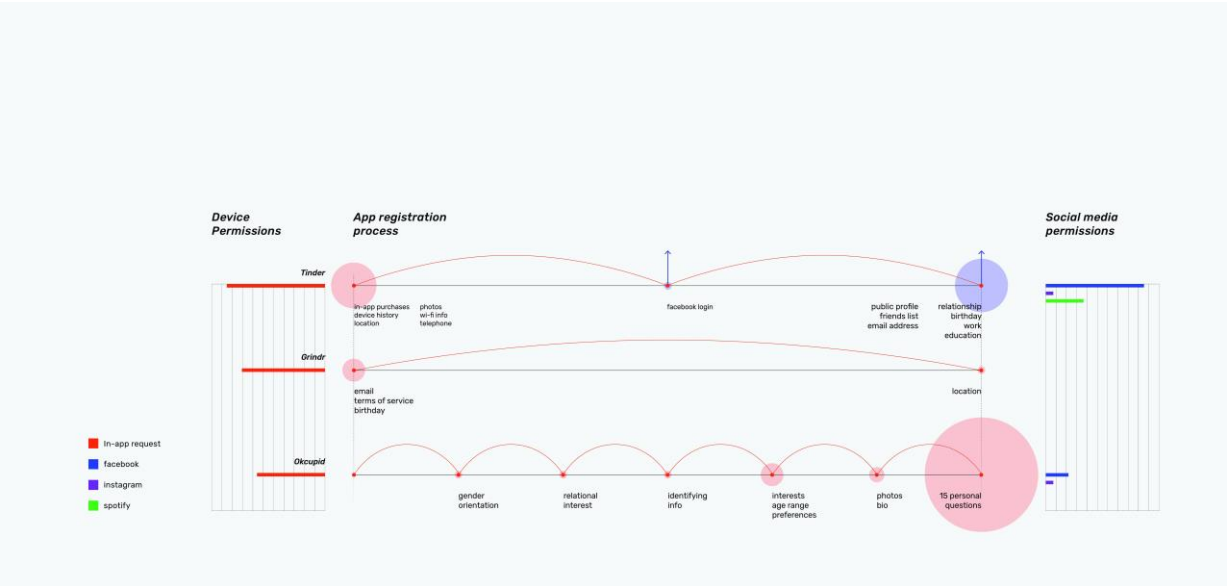


Figure 34.2. 登録プロセス中のデータフローのインターフェイスウォークスルー 出典：Infrastructures of intimate data: Mapping the inbound and outbound data ows of dating apps. Computational Culture, 7. http://computationalculture.net/infrastructures-ofintimate-data- mapping-the-inbound-and-outbound-data- ows-of-dating-apps/

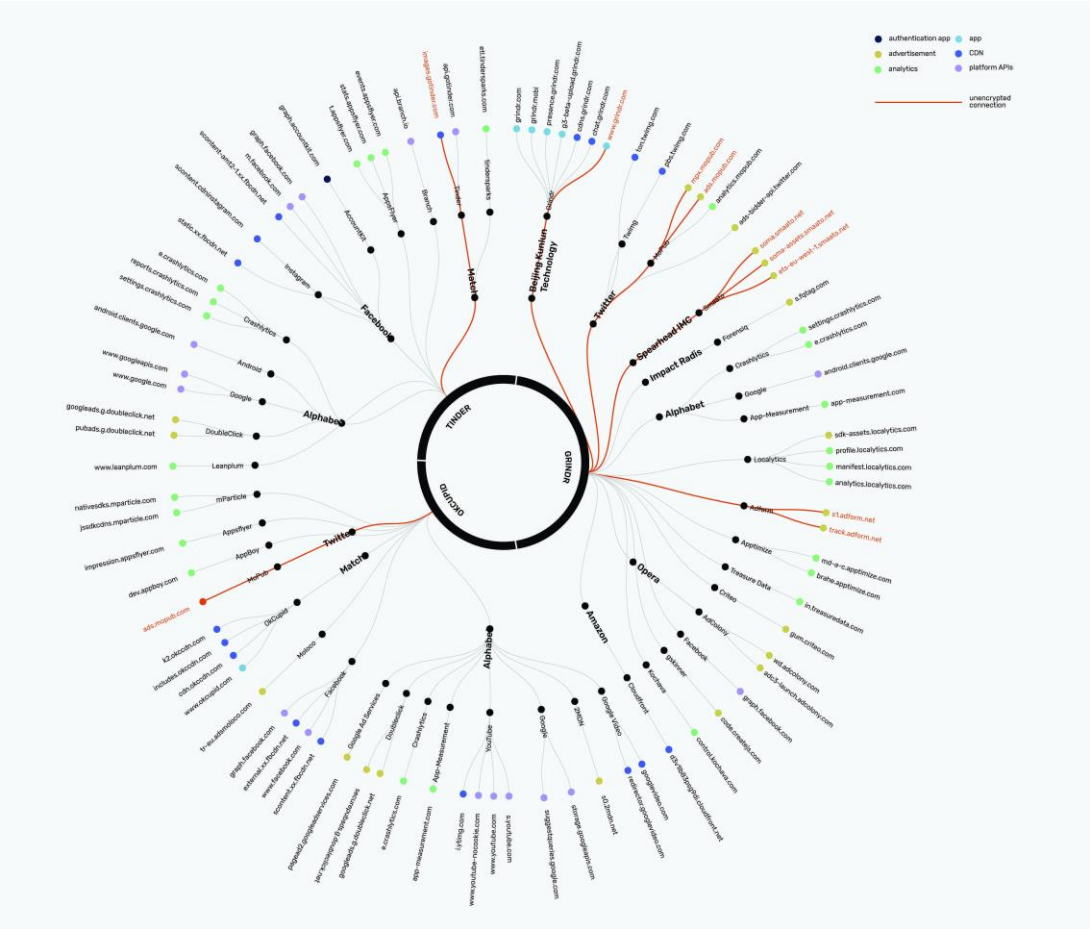


Figure 34.3. 出会い系アプリの Tinder、Grindr、OKCupid とそのサードパーティの間に構築されたネットワーク接続 出典：Infrastructures of intimate data: Mapping the inbound and out- bound data ows of dating apps. Computational Culture, 7.http://computationalculture.net/infra-structures-of-intimate-data-mapping-the-inbound-and-outbound-data-ows-of-datingapps/

結論

プラットフォームやアプリは生活に深く浸透しているため、気づかれないことがよくあります。デジタル研究者、データジャーナリスト、アクティビストは、このような背景があるからこそ、プラットフォームやアプリがどのように機能し、どのような状況で作成・利用されているのかを調査する必要があるのです。データインフラのリテラシーを高めることは、データインフラがさまざまなプラットフォームやネットワークとどのように関連しているのか、それらの間でどのように機能しているのか、また、知られざる利害関係者がどのように関わっているのかを理

解するために重要です。ケンブリッジ・アナリティカのスキャンダルの後、プラットフォームが世論の圧力を受けて、構造化された API を通じてアクセス可能な社会的調査のために用意されたデータを制限することが多くなっています。ここでは、その対応策として、研究者、ジャーナリスト、市民社会グループは、データ調査に適した新しいタイプのデータを創造的かつ独創的に探求すべきであると提案し、アプリを調査するための 3 種類のデータを紹介しました。これをさらに発展させることもできます。ここでは主にアプリを取り上げましたが、スマートシティやモノのインターネット(IoT)など、データが豊富なさまざまな環境を調査する際のヒントになるかもしれません。このように、日常生活を形作るデータ基盤をより深く理解することは、まだまだ現在進行形といってよいのです。

脚注

1. http://thepoliticsofsystems.net/?s=netvizz
2. https://about.fb.com/news/2018/03/cracking-down-on-platform-abuse/
3. https://docs.google.com/document/d/15YKeZFSUc1j03b4lW9YXxGmhYEnFx3TSy68qCrX9BEI/edit
4. https://wiki.digitalmethods.net/Dmi/WinterSchool2020
5. https://www.facebook.com/notes/3707971095882612/
6. https://www.appannie.com/en/

参考文献

Bruns, A. (2018, April 25). Facebook shuts the gate after the horse has bolted, and hurts real research in the process. *Internet Policy Review*. policyreview.info/articles/news/facebook-shuts-gate-after-horse-has-bolted-andhurts-real-research-process/786

Dieter, M., Gerlitz, C., Helmond, A., Tkacz, N., Van der Vlist, F. N., & Weltevrede, E. (2019). Multi-situated app studies: Methods and propositions: *Social Media + Society*. doi.org/10.1177/2056305119846486

Gerlitz, C., Helmond, A., Van der Vlist, F. N., & Weltevrede, E. (2019). Regramming the platform: Infrastructural relations between apps and social media. *Compu- tational Culture*, *7*. computationalculture.net/regramming-theplatform/

Gerlitz, C., & Weltevrede, E. (2019). What happens to ANT, and its emphasis on the socio-material grounding of the social, in digital sociology? In A. Blok, I. Farias,

& C. Roberts (Eds.), *Companion to actor-network theory* (pp. 345- 356). Routledge. doi.org/10.4324/9781315111667-38

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, *5*(2), 1-13. doi.org/10.1177/2053951718786316

Helmond, A., Nieborg, D. B., & Van der Vlist, F. N. (2019). Facebook's evolution: Development of a platform-asinfrastructure. *Internet Histories*, *3*(2), 123-146. doi.org/10.1080/24701475.2019.1593667

Hofffmann, A. L., Proferes, N., & Zimmer, M. (2018). "Making the world more open and connected": Mark Zuckerberg and the discursive construction of Facebook and its users. *New Media & Society*, *20*(1), 199-218. doi.org/10.1177/1461444816660784

Light, B., Burgess, J., & Duguay, S. (2016). The walkthrough method: An ap- proach to the study of apps. *New Media*

& Society, *20*(3), 881-900. doi.org/10.1177%2F1461444816675438

Lomas, N. (2018, April 24). Kogan: "I don't think Facebook has a developer policy that is valid." *TechCrunch*. techcrunch.com/2018/04/24/kogan-i-dont-think-facebook-has-a-developer-policy-that-is-valid/

Lury, C., & Wakeford, N. (2012). *Inventive methods: The happening of the social*. Routledge.

Marres, N. (2017). *Digital sociology: The reinvention of social research*. Polity Press. Matviyenko, S., Ticineto Clough,

P., & Galloway, A. R. (2015). On governance, blackboxing, measure, body, afffect and apps: A conversation with Patricia Ticineto Clough and Alexander R. Galloway. *The Fibreculture Journal*, *25*, 10-29.

twenty ve. breculturejournal.org/fcj-179-on-governance-blackboxing-measure-body-affect-and-apps-aconversation-with-patricia-ticineto-clough-and-alexander-r-galloway/

Nieborg, D. B., & Helmond, A. (2019). The political economy of Facebook's platformization in the mobile ecosystem:

Facebook Messenger as a platform instance.*Media, Culture & Society*, *41*(2), 196–218. doi.org/10.1177/0163443718818384

Perriam, J., Birkbak, A., & Freeman, A. (2020). Digital methods in a post-API environment. *International Journal of Social Research Methodology*, *23*(3), 277–290. doi.org/10.1080/13645579.2019.1682840

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, *20*(1), 293–310. doi.org/10.1177/1461444816661553

Rieder, B. (2013). Studying Facebook via data extraction: The Netvizz application. In *Proceedings of the 5th Annual ACM Web Science Conference* (pp. 346–355). doi.org/10.1177/1461444816661553

Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algo- rithms to "ranking cultures":

Investigating the modulation of visibility in YouTube search results. *Convergence*, *24*(1), 50–68. doi.org/10.1177/1354856517736982

Rogers, R. (2013). *Digital methods*. MIT Press.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014, May 22). Audit-

ing algorithms: Research methods for detecting discrimination on Internet platforms. International Communication Association preconference on Data and Discrimination Converting Critical Concerns into Productive Inquiry, Seattle,WA.

Smout, A., & Busvine, D. (2018, April 24). Researcher in Facebook scandal says: My work was worthless to Cambridge Analytica. *Reuters*. www.reuters.com/article/us-facebook-privacy-cambridge-analyticaidUSKBN1HV17M

Venturini, T., & Rogers, R. (2019). "API-based research" or how can digital sociology and journalism studies learn from the Cambridge Analytica data breach. *Digital Journalism*, *7*(4), 532–540. doi.org/10.1080/216708...

Weltevrede, E. (2016). *Repurposing digital methods: The research affordances of platforms and engines* [Doctoral dissertation], University of Amsterdam.

Weltevrede, E., & Jansen, F. (2019). Infrastructures of intimate data: Mapping the inbound and outbound data ows of dating apps. *Computational Culture*, *7*. computationalculture.net/infrastructures-of-intimate-data-mapping-theinbound-and-outbound-data- ows-of-dating-apps/

脚光を浴びるアルゴリズム：シュピーゲル・オンラインでの共同調査

文：クリスティーナ・エルマー

アルゴリズムの透明性を求める声は、ドイツでは新しいものではありません。2012 年には、デア・シュピーゲルのコラムニストのサシャ・ロボが Google に、たとえ損害を受けるとしても検索アルゴリズムの仕組みを開示するよう求めました（Lobo, 2012）。ドイツでの著名な事例が示したように、Google はオートコンプリート機能などを通じて、私たちの世界の見方を形成することができます。前連邦大統領夫人（ベティーナ・ウルフ）は、自分の名前を検索したときにオートコンプリート機能で問題のある表示が提案されたとして、Google を提訴しました。2 年後、ドイツの法相はこの訴えを再度取り上げ、首相は 16 年にこれを蒸し返しました。メルケル首相は「アルゴリズムはより透明であるべきだ」と要求したのです（Kartell, 2014; Reinbold, 2016）。

ここ数年、シュピーゲルはアルゴリズムの説明責任について議論してきましたが、当初は報道の対象としてのみで、独自の調査や分析プロジェクトはありませんでした。ドイツのメディアが米国のメディアよりも遅れてこの分野の実験を始めた理由は、主に二つあると思われます。第一に、ドイツのジャーナリストには、自由に使える強力な情報公開の権利や手段がないこと。第二に、データジャーナリズムには、米国のような長い伝統がありません。シュピーゲルは 16 年に自前のデータジャーナリズム部門を設けたばかりですが、徐々に、しかし着実にこの分野を拡大しています。もちろん、リソースの少ないニュースルームでも、組織やフリーランサーとの協力などを通じて、この分野で活躍することは可能です。私たちの場合も、これまでに取り組んできたアルゴリズムによる説明責任報道のプロジェクトは、すべてこのような方法で生まれたものです。そこで本章では、コラボレーションとそこから得られた教訓に焦点を当てます。

Google、Facebook、Schufa:三つのプロジェクトの概要

私たち編集部は、[外部との]協力を中心にアルゴリズムの調査をしています。17 年の連邦選挙に向けて、私たちは NGO「AlgorithmWatch」と協力し、Google の検索結果のパーソナライズについての知見を得ました1。ユーザーには、あらかじめ定義された検索を定期的に実行するプラグインを自分のコンピュータにインストールしてもらいました。合計で約 4400 人の参加者が約 600 万件の検索結果を提供してくれたため、少なくとも Google と調査対象地域に関しては、フィルターバブル理論に疑問を投げかける分析データを得ることができたのです。このプロジェクトでは、AlgorithmWatch の共同研究者がシュピーゲルにアプローチしてきました。必要なデータをクラウドソーシングするために、大きなリーチを持つメディアパートナーを探していたのです。取材はインターネットやテクノロジー関連の話題を扱う部門が担当しましたが、データジャーナリズム部門は、取材計画や方法論の評価をサポートしました。さらに、法的に問題のない方法でプロジェクトを遂行するためには、法務部

門のバックアップが不可欠でした。例えば、データ保護の問題は、報道の中で明確にされなければならず、プロジェクトに参加したすべての人が十分に理解できるようにしなければなりません。

並行して、シュピーゲルはプロパブリカとともに、選挙前の数ヶ月間にドイツで AdCollector を展開しました（Angwin & Larson, 2017）。このプロジェクトは、ドイツの政党が Facebook ユーザーをターゲットに広告を出す様子を明らかにすることを目的としたものです。あるプラグインを用いて、ユーザーが自分のストリームで目にする政治広告を収集し、どんな広告がユーザーに表示されないのかを明らかにしました。南ドイツ新聞やターゲスシャウなど通常は競合する他のドイツメディアとも協力しました。この場合、公共の利益に資するため、できるだけ多くの人々に情報を届ける必要がありました。結果を記事として発表することもできましたが、私たちが最も重視したのは透明性です。2 週間後には、約 600 の政治広告が収集され、公開されました。

プロパブリカのジュリア・アングウィンとジェフ・ラーソンはハンブルグで開催されたドイツの調査報道ジャーナリスト協会 Netzwerk Recherche の年次大会でアルゴリズムの説明責任報道に関するセッションを行い、コラボレーションのアイデアを紹介しました。このアイデアは、シュピーゲルの報道部門の複数の部署から集まった技術・方法論の専門家と協力して、一から練り上げられたものです。また、以前からのパートナーである NPO の AlgorithmWatch との交流は、私たちにとって法的背景に光を当て、それを研究に盛り込むために、非常に役立ちました。この会議の後は、定期的に電話会議を行い、さらにアイデアを広げていきました。ドイツの他のメディアのパートナーも、後から参加してくれました。

18 年、デア・シュピーゲルは、ドイツの極めて強力なアルゴリズム、信用情報機関 Schufa を調査する一大プロジェクトに携わりました。このレポートは、民間人の信用力を評価します。このレポートは、誰かが請求書を支払ったり、家賃やローンを負担したりできる確率を示します。そのため、個人の私生活に多大な影響を与え、社会全体にも悪影響を及ぼす可能性があります。例えば、個人に関して利用できるデータの量によっては、スコアが社会的な差別や不平等な扱いを助長することも考えられます。誤ったデータや取り違えは、個人にとって致命的になり得るのです。

アルゴリズムの基礎となるスコアリングは透明ではありません。どのデータがどのような重み付けで考慮されるのかは不明です。また、影響を受ける人は、そのプロセスを知らないことが多いのです。このような状況から，ドイツでは Schufa が議論の対象となっており、OpenSCHUFA のようなプロジェクトはアルゴリズムの説明責任に関する公共の討論に不可欠なのです2。このプロジェクトは、NGO である Open Knowledge Foundation（OKFN）と AlgorithmWatch が中心となって進められました。シュピーゲルは、バイエルン放送とともに提携パートナーを務めました。このプロジェクトのアイデアは、複数の関係者がほぼ同時に思いついたものでした。NGO の AlgorithmWatch と OKFN、そしてバイエルン放送のデータジャーナリズムチームがプロジェクトを成功させた後、シュピーゲルは会合に顔を出しました。この組み合わせは、特別な課題を提示しました。

両メディアチームにとっては、特にクラウドファンディングのプロセスからの独立性を確保するため、NGO とは別に活動することが重要でした。そのため、関係者間での話し合いはあったものの、正式なパートナーシップや共同でのデータ評価はできませんでした。この例は、特にこのような公共性の高いテーマにおいて、ジャーナリストが自らの自律性について考えることがいかに重要かを示しています。OpenSCHUFA を知ってもらうことは、このプロジェクトの中心的な成功要因の一つでした。まず、データを収集するために必要なインフラをクラウドファンディングによって構築しました。結果は、匿名化された形で、1年かけてパートナーが共同で評価しました。その背後にある中心的な疑問は、Schufa アルゴリズムは特定の集団を差別し、社会の不平等を増大させるのかということです。結果を見ると、そのようになっています。

このスコアは、高齢者や女性、転居の頻度が低い人を優遇していることがわかりました。スコア内のアルゴリズムのバージョンが異なると、同じ属性の人でも結果が異なるという、これまで知られていなかったタイプの差別が発見されました。これらの成果は、多くのボランティアや支援者の参加がなければ実現しませんでした。クラウドファンディングのキャンペーンがほぼ成功したことで、ソフトウェアの開発資金を確保することができました3。また、その後のクラウドソーシングでは、約 2800 人から個人信用情報が寄せられました。このサンプルは、もちろん代表的なものではありませんが、今回の調査結果を明らかにするのに十分な多様性を持っています。

インパクトと成功の指標

Facebook と Google の調査は、どちらも結果としては地味だったものの、私たちの仮説を裏付けるものでした。政党は Facebook のターゲティングオプションをほとんど利用していないようですし、よく言われる Google のフィルターバブルはドイツでのクラウドソーシング実験では見つかりませんでした。しかし、私たちにとってこれらのプロジェクトの価値は、社会におけるアルゴリズムの機能性とリスクに関する読者のリテラシーを高めることにありました。記事のリーチ数は、私たちがこのテーマをより広く知らしめることに成功したことを示す指標となりました。Schufa プロジェクトの開始時に掲載した記事は、多くの読者（約 33 万 5000 人）に読まれました4。さらに、このトピックは公共の場で広く議論され、多くのメディアや会議で取り上げられました。ドイツの消費者保護相は、信用スコアリングの透明性向上を求め、「全国民は、自分の信用度を計算する際にどのような特徴が含まれているのか、またそれらがどのように重み付けされているのかを知る権利を持たなければなりません」と訴えました。

日常生活への影響はどうでしょう？最初のステップとして、私たちはこのテーマを一般の人々の意識に根付かせることが重要でした。これまでのところ、政治家が、より広い社会的影響を持つアルゴリズムに対処するまったく斬新な方法は見つかっていません。とはいえ、アルゴリズムの説明責任報道というトピックは、私たちにとって非常に重要です。なぜなら、ヨーロッパでは、社会におけるアルゴリズムの問題を議論し、それにどう対処したいか[といった規範]を形作る余地がまだあるからです。市民が社会におけるアルゴリズムの未来を理解し、形作ることができるように、必要な知識を提供することは、私たちジャーナリストの責務です。可能な限り、アル

ゴリズムとその影響を透明化し、リスクを特定し、責任者と対決することで、権力の監視役としての役割も担っています。そのためには、競合他社や他の分野の人々と、通常では考えられないような協力関係を築かなければなりません。

私たちは、このような協力関係が、最終的にこの分野の法律や透明性の基準に対する圧力を高めることを期待しています。アルゴリズムの説明責任の調査には、より多くの努力と資源が投入されるべきであり、「The Markup」はこの分野で非常にエキサイティングな調査結果を発表しています。アルゴリズムの規制にはまだ行動の余地があることもあり、さらなる調査が求められています。アルゴリズムによる説明責任報道は、ここ数年で発展し始めたばかりです。デジタル化が進む世界の課題に対応するためには、急速に成長しなければなりません。

共同調査の組織化

共同調査を実行するには、新しい、あるいはニュースルームであまり使われていないスキルが必要です。これには、大規模なデータセットの分析やプログラミングだけでなく、プロジェクトの管理も含まれます。後者は見落とされがちなので、ここでは、私たちがこれまでに行ってきた仕事の具体例を挙げながら、詳しく説明します。多様な人々からなる集団での共同作業は、能力やリソースの共有を容易にするだけでなく、役割を明確に定義することができます。シュピーゲルは、これらのコラボにおいては、メディアパートナーとしてプロジェクト自体にはあまり深く関与せず、中立的なコメンテーターとして関わりました。これにより、編集者は独立性を保つことができ、読者の信頼を得ることができました。編集者たちは、プロジェクト内の報道にも[他の記事と同じ]品質基準を適用し続けました。たとえば、報道対象者には、告発についてコメントする機会を常に与えることにしています。このようなメカニズムは、NGO[単独]と比較して活動を遅らせることになるかもしれません。しかし同時に、読者に十分な情報を提供し、長期的には公共の議論を豊かにすることができるのです。

アルゴリズムの説明責任の共同研究では，これらの役割について事前に合意しておくことが重要な成功基準になることがわかっています。また、早い段階で共通のタイムラインを作成し、さまざまなチャンネルでプロジェクトを発表する際の言語ルールを定めておく必要があります。結局のところ、明確な役割分担は、それが一貫して伝えられている場合のみ機能するものだからです。これには、例えば、プロジェクトにおけるさまざまなパートナーの役割に関する明確な用語や、利益相反が発生した場合の免責事項の調整などが含まれます。並行して、プロジェクト管理の方法を慎重に使用し、プロジェクトの目標を明確に設定し、どんなリソースが利用できるのか議論しなければなりません。コーディネーターは全体的なコミュニケーションを助け、それによって参加している編集者に調査に必要なスペースを与えるべきです。誰もが最新の情報を得られるようにするには、情報チャネルをできるだけシンプルにする必要があります。特に、大規模なプロジェクトの立ち上げ時にはそうすべきでしょう。

編集計画上、三つのテーマ分野は難しいものになりました。一般的には妥当性やニュースの価値が問われることはありませんでしたが、幅広い読者を獲得するためには特別なストーリーが必要でした。これらのストーリーは、調べたアルゴリズムの個人的な影響に焦点を当てていました。例えば、Schufa データが正しく割り当てられなかったために、シュピーゲル編集部の同僚がインターネットプロバイダとの契約を結ぶのが困難になったとします。彼の体験を記した記事は、アルゴリズムが個人レベルでどのような影響を及ぼすかを印象的に示しており、その結果、読者の現実生活とつながったのです（Seibt, 2018）。私たちは可能な限り読者の関心に合わせて報道しました。もちろん、関与したデータジャーナリストたちは、調査対象のアルゴリズムの機能にも強い関心を持っていました-この関心は研究目的には極めて有用です。しかし、アルゴリズムがもたらす結果に影響を与える場合のみ、読者がアクセスしやすい方法で報じられるのです。編集部内では、この三つのプロジェクトへの支持は非常に高いものでした。とはいえ、ニュースを中心とした編集チームの日常業務の中で、日々の報道にリソースを割くことは容易ではありませんでした。調査の結果が必ずしも素晴らしいものではなかった場合はなおさらです。

教訓

最後に、これらのプロジェクトから学んだことをまとめてみました。

可能な限りコラボすること。アルゴリズムの説明責任をめぐる優れた調査報道は、他の人と力を合わせ、多様なスキルセットを持つチームを作ることで初めて可能になります。これは、多くのジャーナリストが対処しなければならないリソースの不足と法的規制の両方を考慮すると、重要なことです。しかし、このようなプロジェクトでは、異なる分野の関係者が集まるため、基本的な関連性の基準、要件、能力について事前に話し合っておくことが重要です。

目標を体系的に定義する。このようなプロジェクトでは、アルゴリズムの動作原理に対する認識を高めることが最初の大きな目標になります。もちろん、プロジェクトは可能な限り透明性を高めることも重要です。プロジェクトのパートナーは、アルゴリズムに差別的な効果があるかどうかをチェックする程度でも困難な目標であり、膨大なデータセットとリソースを必要とすることを留意する必要があります。

プロジェクトの実装には注意を払うこと。作業量や関係するジャーナリストの日々のプレッシャーによっては、プロジェクトマネージャーが必要な場合もあります。プロジェクトのスケジュールが報道要件と矛盾する場合があることに注意してください。この点を考慮して、他のパートナーと連絡を取り、可能であれば、そのような場合の代替案を準備します。

調査設計に投資する。有用なデータを生み出す有意義な設計を行うには、専門のパートナーが必要になる場合があります。コンピュータサイエンス、数学、および関連分野の科学者との緊密な提携は、アルゴリズムのより技

術的な側面のいくつかを調査するために特に有用です。さらに、社会・文化研究者と協力し、分類やそこに実装されている規範をより深く理解することも有効でしょう。

ユーザーデータを保護する。ユーザーからのデータ提供は、アルゴリズムを調査する上で有用な場合があります。このようなクラウドソーシングのプロジェクトでは、データ保護を確保し、各国の法規制の要件を考慮するために、法的サポートが不可欠です。会社にデータ保護担当者がいるなら、早い段階からプロジェクトに参加させましょう。

脚注

1. algorithmwatch.org/de/ lterblase-geplatzt-kaum-raum-fuer-personalisierung-bei-google-suchen-zurbundestagswahl-2017// （ドイツ語）
2. www.openschufa.de （ドイツ語）
3. www.startnext.com/open... （ドイツ語）
4. 大半は、ホームページなどの内部チャネルから記事にたどり着いたものです。しかし、記者の体験談を掲載したフィールドレポートでは、約 22 万人に読んでいただくことができました。そのうちの 5 分の 1 はソーシャルメディア経由であり、これは平均を大きく上回っています。つまり、このテーマで新たなターゲット層にリーチできたようです。

参考文献

Angwin, J., & Larson, J. (2017, September 7). Help us monitor political ads online. *Pro-Publica*. www.propublica.org/article/help-us-monitor-political-ads-online

Kartell, S. vor. (2014, September 16). Maas hätte gerne, dass Google geheime Such-formel offenlegt. *Der Spiegel*. www.spiegel.de/wirtschaft/unternehmen/

Lobo, S. (2012, September 11). Was Bettina Wulff mit Mettigeln verbindet. *Der Spiegel*. www.spiegel.de/netzwelt/netzpolitik/google-suchvorschlaege- was-bettina-wulff-mit-mettigeln-verbindet-a855097.html

Reinbold, F. (2016, October 26). Warum Merkel an die Algorithmen will. *Der Spiegel*. www.spiegel.de/netzwelt/netzpolitik/angela-merkel-warum-die-kanzlerin-an-die-algorithmen-von-facebook-will-a1118365.html

Seibt, P. (2018, March 9). Wie ich bei der Schufa zum “deutlich erhöhten Risiko” wurde. *Der Spiegel*.

www.spiegel.de/wirtschaft/service/schufa-wie-ich- zum-deutlich-erhoehten-risiko-wurde-a-1193506.html

Twitter の#ddj ハッシュタグ

文：ユーニス・ウー、マーク・スミス

概要

ソーシャルネットワーク分析・可視化パッケージ NodeXL を使用して、世界のデータジャーナリズム・コミュニティが何についてツイートしているかを調べてみました。

キーワード：#ddj、Twitter、ソーシャルネットワーク分析、データジャーナリズム、ソーシャルメディア、データ分析

データジャーナリズムの分野を一つの用語で追跡するのは簡単ではありません。データジャーナリストは、#datajournalism、#ddj、#dataviz、#infographics、#data など、自分たちの仕事に関連して無数のハッシュタグを使用しています。6 年前、調査報道機関の国際組織であるグローバル調査報道ネットワーク（GIJN）が Twitter でデータジャーナリズムに関する発信を始めたとき、最も人気のあったハッシュタグは#ddj（データ駆動型ジャーナリズム）でした1。データがジャーナリズムを駆動しているのではなく、データは単にジャーナリズムに情報を提供する、あるいは使用するツールに過ぎないという説もあり、「データ駆動型ジャーナリズム」という言葉そのものが議論の的になっています。データは構造化された事実と統計から構成されており、ジャーナリストは記事を作成するためにフィルタリングし、分析し、パターンを発見する必要があります。プロフィール記事を「インタビュー駆動型ジャーナリズム」、公文書に基づいた記事を「ドキュメント駆動型ジャーナリズム」と呼んだりしないように、優れたデータジャーナリズムの記事はデータをその構成要素の一つとしてしか使っていないのです。

#ddj の役割

データジャーナリズムのコミュニティでは、#ddj というハッシュタグが広く使われており、世界中のプロジェクトや活動を共有するための有力な情報源となっています。データジャーナリストはこのハッシュタグを利用して、自分たちの仕事を宣伝し、世界に向けて発信しています。また、データジャーナリズム・コミュニティのメンバーは、ハッシュタグを使用してコンテンツを検索、発見、共有することができ、ソーシャルメディア上での議論の促進にも役立っています。ハッシュタグ#ddj は、選挙予測や確率グラフの誤った解釈から、データ倫理や人工知能の説明責任に至るまで、さまざまな議論を展開しています。

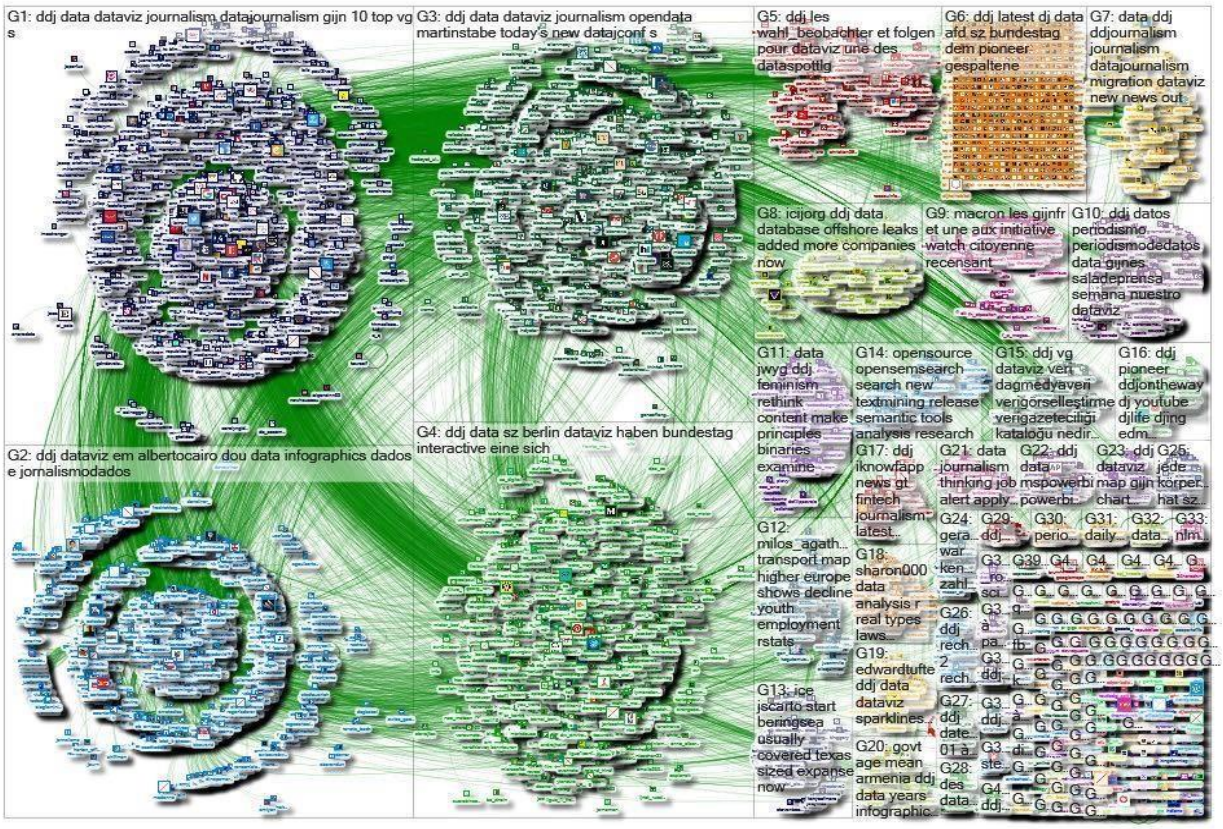


Figure 36.1. 2018年1月1日～8月13日の Twitter 上の#ddj マッピング 出典：NodeXL.

トップ 10 #ddj の誕生

GIJN の週刊トップ 10 #ddj[を紹介する]シリーズは、2014 年 1 月、私たちの一人が初めて#ddj ネットワークグラフをツイートしたことから始まりました（Smith, 2014）。このグラフは、#ddj というハッシュタグに言及したツイートを、それらのツイートへの返信も含めてマッピングしたもので、Excel をベースにしたソーシャルネットワーク分析・可視化パッケージ「NodeXL」を使って作成されました。これらのネットワークグラフは、返信、@メンション、リツイートなどの行動から生まれる相互関係のパターンを明らかにしています。これらのパターンは、主要な人物、グループ、話題のトピックが浮き彫りにしました。国際的な調査報道機関として、GIJN は常に調査報道とデータジャーナリズムの分野で何が起きているかについての認識を高める方法を探しています。GIJN のエグゼクティブ・ディレクターであるデビッド・カプランは、スミスのネットワークグラフを見て、このマップを使って、データジャーナリズムの人気で興味深い例を紹介する「トップ 10 #ddj」を毎週作成しようと提案しました（彼とスミスは、調査報道を毎週まとめる試みにも取り組みましたが、#ddj ほど役立つ単一のハッシュタグはありませんでした）。

GIJN はネットワークグラフが提案する調査結果に忠実であるものの、重複を排除し、最も興味深い項目を強調するために、人が多少キュレーションしなければなりません。このシリーズが誕生して以来、過去 6 年

間に#ddj のハッシュタグを使ったデータジャーナリズム・コミュニティのディスカッションのスナップショットを 250 以上集めました（GIJN, n.d.）。このシリーズは、#ddj のツイートをすべて追うことができない関係者にとっても、手軽なまとめとして役立っています。スナップショットという言葉は、単なる比喩ではありません。この分析は、フォトジャーナリズムが主要なニュースメディアの第一面に掲載される実際の群衆を描写するのと同じように、データジャーナリズムの Twitter コミュニティの姿を見せてくれます。

#ddj Twitter トラフィックの進化

#ddj を使った Twitter のトラフィックがどのように進化してきたかを理解しようと、私たちは 2014～19 年にかけて収集した#ddj データの大まかな基礎分析を行いました。6 年間それぞれの 2 月と 3 月の 8 週間、計 48 週を少量ずつサンプルとして選びました。さまざまなコンテンツが共有され、最も人気のある項目は、分析と考察、賞、助成金、イベント、コース、仕事、ツール、リソース、調査などでした。共有されるコンテンツの種類は、長年にわたり一貫しています。14 年には、急成長するデータジャーナリズムの分野を論じた記事を目にしました。中には、データジャーナリズムは説明責任と洞察力を高めうると主張する記事（Howard, 2014）や、データ分析はジャーナリストの未来であると予測する記事（Arthur, 2010）などがありました。後年、人工知能、大規模なデータリーク、データの共同調査など、新しいトピックが議論されていることが確認されました。

また、メディア業界がデータジャーナリズムをニュースルームに取り入れるべきかどうかを議論するのではなく、データジャーナリストがデータジャーナリズムのプロセスに関する洞察を提供し（Grossenbacher, 2019）、データベースの最適な活用方法を共有するといった（Gallego, 2018）、掘り下げたハウツー記事が出現したのです。さらに、共有された調査の中には、選挙、移民、公害、気候、サッカーに関する分析が多く含まれていることにも気づきました。GIJN の週刊#ddj ラウンドアップは、最も人気のあるツイートや URL をハイライトするだけでなく、#ddj ディスカッションの中心的な参加者もリストアップしています。データジャーナリズムの専門家であるエドワード・タフテ、アルベルト・カイロ、マーティン・スタブ、ネイト・シルバー、ネイサン・ヤウ、そしてヨーロッパと北米のデータチーム（仏紙ル・テレグラム、スイス紙ターゲス・アンツァイガー、独紙ベルリナー・モルゲンポスト、FiveThirtyEight、Financial Times、The New York Times の The Upshot）といったいつもの面々です。彼らの仕事は、時に教育的で刺激的であり、さらなる議論の引き金となることがあります。データジャーナリズムのコミュニティは、これらのインフルエンサーを活用し、ネットワークを構築することもできます。Connected Action のマッピングで明らかになったように、#ddj には他にも多くのハッシュタグが付くことが多く、コミュニティのメンバーは同様のストーリーを探すことができます。

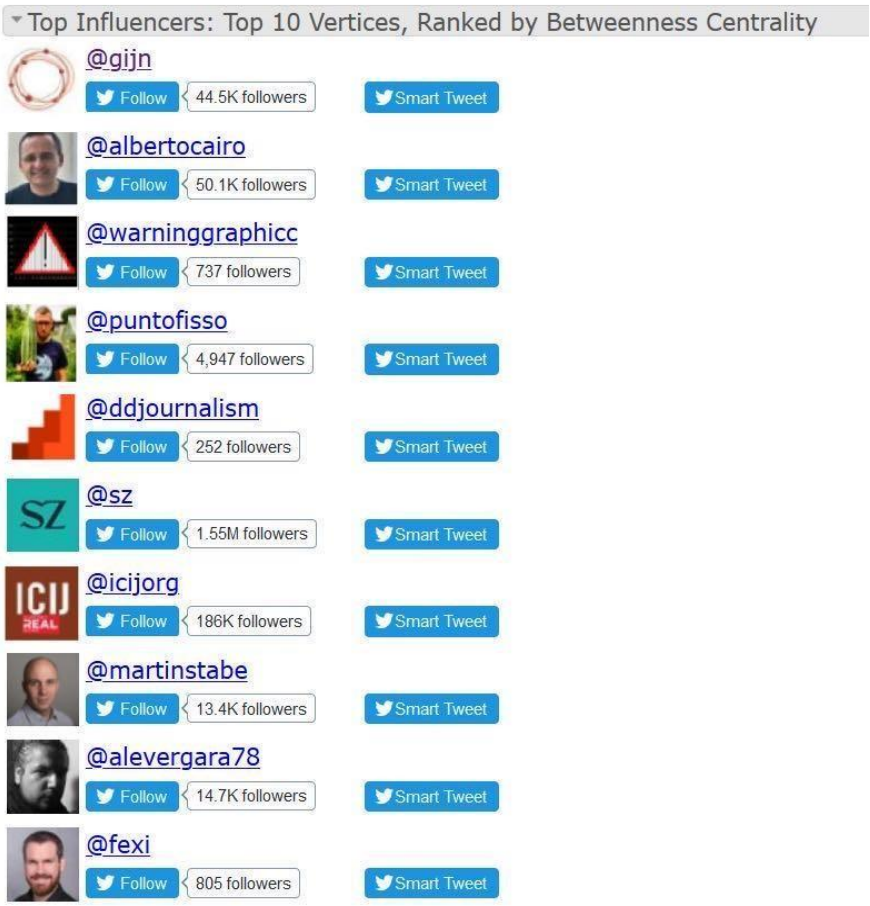


Figure 36.2. トップインフルエンサーの例(2018 年1月1日～8月13日) 出典: NodeXL.

#ddj と並んで表示されるハッシュタグで群を抜いているのは、#dataviz、#visualization、#datajournalism、#opendata、#data、#infographics です。このことは、この分野の人々が、公開データの入手可能性だけでなく、データが読者のために創造的に提示され可視化される方法にも関心を寄せていることを示しているのです。しかし、NodeXL の#ddj マッピングはツイートした人だけを分析するため、決してこの分野全体を代表するものではありません。さらに、一般的に Twitter のフォロワー数が多く、リツイートされる数が多い人は、われわれのまとめでより目立つように取り上げられる傾向があります。また、上位のツイートは通常、ドイツや米国を中心とするヨーロッパとアメリカ大陸からのものが多く、アジアやアフリカからのツイートは少ないことにも気づきました。

これは、他の地域ではデータジャーナリズムのコミュニティが比較的安定していないため、あるいは他の地域のデータジャーナリズムのコミュニティが同じ Twitter ハッシュタグを使って組織化されていないか、Twitter 上でまったく組織化されていないために、Twitter のユーザー・ベースが偏っていることが原因である可能性があります。この 1 年間で、Twitter で広く共有されている著名なデータジャーナリズム団体の活動が、私たちのネットワークグラフに表示されないこともありました。これは、人々がストーリーをツイートする際に#ddj と

いうハッシュタグを使わなかったり、他のハッシュタグを使ったり、全く使わなかったりしたことが原因かもしれません。2017 年 11 月に Twitter がツイート文字数[の上限]を 140 文字から 280 文字に拡大したことも、#datajournalism のような長めのハッシュタグを選択させる一因になったのではないかと推測しています。

Top Hashtags in Tweet in Entire Graph:
[22540] ddj
[6765] dataviz
[1783] datajournalism
[1578] opendata
[1517] vg
[1253] data
[1080] infographics
[589] opensource
[541] datajournalismawards
[534] journalism

Figure 36.3. 上位の関連ハッシュタグの例（2018年1月1日～8月13日）出典：NodeXL.

面白い#ddj を見つける

強力なジャーナリズムや美しいビジュアライゼーションばかりではなく、時にはただ単に面白いだけのものを見つけることもあります。最後に、この 1 年間に私たちが#ddj のハッシュタグを使って発見した、面白い作品をいくつか簡単に紹介します。シャキン・ゴンザレス・ベイラ（2017）は、可愛らしく巧妙なビジュアルエッセイで、さまざまな国の人々がどんな「〇〇を修理する方法」を検索しているのかを表現しました。多くの温暖な国では冷蔵庫、北米や東アジアではトイレ、北欧や東欧の人々は電球の直し方に関する情報を必要としているようです。次に、スミソニアンの「Sally L. Steinberg Collection of Doughnut Ephemera」は、ドーナツの穴の大きさが長年にわたって徐々に縮小しているグラフがあります（Edwards, 2018）。別の作品では、グラフィックデザイナーのナイジェル・ホームズが、『Crazy Competitions』という本の中で、カタツムリのレースから妻運び競争まで、世界中の風変わりで素晴らしい競技を図解して解説しています（Yau, 2018）。

世界中の女性にとって、女性用ジーンズのポケットが実用的でないほど小さいことは周知の事実ですが、The Pudding の作品はそれを証明するための明確なデータと分析を提供しました（Diehm & Thomas, 2018）。最後に、子作りのピークシーズンというのはあるのでしょうか？Visme による国連の出生データの分析は、そう示唆しているようです。彼らは、3 つの異なる変数の間に相関関係を見いだしました。出産が多い月、季節、国の緯度（赤道からの距離）の三つが、国によって性交のリズムに影響を及ぼす可能性があるというのです（Chibana, n.d.）。

脚注

1. gijn.org

参考文献

Arthur, C. (2010, November 22). Analysing data is the future for journalists, says Tim Berners-Lee. *The Guardian*.

www.theguardian.com/media/2010/nov/22/data-analysis-tim-berners-lee

Chibana, N. (n.d.). Do humans have mating seasons? This heat map reveals the surprising link between birthdays and seasons. *Visual Learning Center by Visme*. visme.co/blog/most-common-birthday/

Diehm, J., & Thomas, A. (2018, August). Pockets. *The Pudding*. pudding.cool/2018/08/pockets/

Edwards, P. (2018, June 1). Have donut holes gotten smaller? This compelling vintage chart says yes.

Vox.www.vox.com/2015/9/20/9353957/donut-hole-size-chart

Gallego, C. S. (2018, January 23). How to investigate companies found in the offfshore leaks database. ICIJ. www.icij.org/inside-icij/2018/01/investigate-companies-found-offshore-leaks-database/

GIJN. (n.d.). Top 10 in data journalism archives. Global Investigative Journalism Network. gijn.org/series/top-10-datajournalism-links/

Grossenbacher, T. (2019, March 8). (Big) data journalism with Spark and R. timogrossenbacher.ch/2019/03/bigdata-journalism-with-spark-and-r/

Howard, A. (2014, March 3). Data-driven journalism fuels accountability and insight in the 21st century.

TechRepublic.www.techrepublic.com/article/data-driven-journalism-fuels-accountability-and-insight-in-the-21stcentury/

Smith, M. (2014, January 22). First NodeXL #ddj network graph. Twitter. twitter.com/marc_smith/status/425801408873385984

Xaquín G. V. (2017, September 1). How to fijix a toilet and other things we couldn't do without search. how-to- x-atoilet.com

Yau, N. (2018, May 21). Nigel Holmes new illustrated book on Crazy Competitions. *FlowingData*.

owingdata.com/2018/05/21/nigel-holmes-new-illustrated-book-on-crazy-competitions/

データジャーナリズムを記録する

文：メレディス・ブルサード

概要

データジャーナリズムのプロジェクトを記録する際の課題と、プロジェクトを将来にわたって確実に保存するためにデータチームがすべきことを説明します。

キーワード：データジャーナリズム、アーカイブの実践、アーカイブ、デジタル・アーカイブ、リンク切れ、Web アーカイビング

2012 年に出版された『データジャーナリズム・ハンドブック』初版で、データジャーナリズムのパイオニアであるスティーブ・ドイグは、お気に入りのデータストーリーの一つがトム・ハーグローブによる「殺人ミステリー」プロジェクトだと書いています1。Scripps Howard News Service が発行したこのプロジェクトでは、18 万 5000 件の未解決殺人事件に関する詳細な人口統計データを調べ、どの殺人事件が関連しているかを示唆するアルゴリズムを構築しました。関連性のある殺人は、連続殺人の可能性があります。「このプロジェクトにはすべてがある」とドイグは書いています。「ハードワーク、政府自身よりも優れたデータベース、社会科学技術を使った巧妙な分析、読者が自分で調べることができるオンラインのインタラクティブなデータの提示」。

6 年後のハンドブック第 2 版の時点では、このプロジェクトの URL は途切れていました（projects.scrippsnews.com/magazine/murder-mysteries）。 このプロジェクトが Web 上から消えていたのは、発行元がなくなったからです。何度も合併や再編を繰り返し、最終的にはローカルニュースネットワーク『USA トゥデイ』を発行するガネットと合併したのです。

人は転職し、メディア企業は変遷するものです。しかし、これはデータジャーナリズムのプロジェクトに悲惨な影響をもたらします（この問題については、Boss & Broussard, 2017; Broussard, 2014, 2015a, 2015b; Fisher & Klein, 2016 などを参照してください）。データプロジェクトは、新聞や雑誌の紙面に掲載される「ふつうの」テキストや画像からなる記事よりもはかないのです。通常、リンク切れはアーキビストにとって大きな問題ではありません。LexisNexis や ProQuest などのデータベースプロバイダーを使えば、21 世紀のどの日でもニューヨーク・タイムズの紙面を簡単に見つけることができるからです。

しかし、データストーリーの場合、リンク切れはより深刻です。データジャーナリズムの記事は、従来のアーカイブには保存されておらず、Web から消えつつあるのです。報道機関や図書館が行動を起こさない限り、将来の歴史家は、2017 年のある日にボストン・グローブ紙が発行したすべての記事を読むことはできなくなるでしょう。このことは、研究者にとっても、この分野の集合的な記憶にとっても、重大な意味を持ちます。ジャーナリズムはしばしば、「歴史の第一稿」と呼ばれます。もし、その第一稿が不完全ならば、未来の学者はどうやって現代

を理解するのでしょうか。あるいは、記事が Web から消えてしまった場合、ジャーナリストはどのようにして個人的な仕事のポートフォリオを維持するのでしょう？これは人間の問題であり、単なる計算上の問題ではないのです。

データジャーナリズムがなぜ後世のためにアーカイブされていないのかを理解するには、「ふつうの」ニュースがどのようにアーカイブされているかを知ることから始めるとよいでしょう。すべての報道機関は、コンテンツ管理システム（CMS）と呼ばれるソフトウェアを使用しています。このシステムによって、報道機関は毎日作成する何百ものコンテンツをスケジュール管理し、公開する各コンテンツに一貫した視覚的外観と感触を与えることができるのです。歴史的に、報道機関では紙と Web で異なる CMS を使用してきました。Web の CMS では、各ページに広告を埋め込むことができ、報道機関が収益を上げる方法の一つとなっています。紙の CMS では、印刷デザイナーが異なるバージョンのレイアウトを管理し、そのページを印刷所に送って印刷・製本します。通常、動画は別の CMS になります。ソーシャルメディアの投稿は、SocialFlow や Hootsuite など別のアプリケーションで管理することもありますし、そうでないこともあります。Lexis-Nexis や他の大手プロバイダーへのアーカイブのフィードは、紙媒体の CMS に接続されていることが多いようです。報道機関の誰かが Web の CMS に接続しない限り、デジタルファーストのニュースは図書館やアーカイブが取得するデジタルフィードには含まれないのです。これは、アーカイブが中立的な存在ではなく、人間が将来にとって何を重要か（そして何が重要でないか）判断することに依存していることを思い知らせる事態です。

ほとんどの人はこの時点で、「Internet Archive はどうなの？」と尋ねます。Internet Archive は［まさに］宝であって、このグループはニュースサイトのスナップショットを捉える立派な仕事をしています。その技術は、最先端のデジタルアーカイブソフトウェアの一つです。しかし、彼らのアプローチは、すべてを捕捉しているわけではありません。Internet Archive は、一般公開されている Web ページのみを収集しています。ログインが必要であったり、ペイウォールを導入していたりする報道機関は、Internet Archive に自動的に保存されることはありません。静的コンテンツである Web ページ、つまりプレーンな HTML は、最も簡単に保存できます。このようなページは Internet Archive に簡単に取り込まれます。JavaScript やデータ・ビジュアライゼーションなど、かつて「Web 2.0」と呼ばれた動的なコンテンツは保存が非常に難しく、Internet Archive に保存されることはあまりありません。Internet Archive の FAQ には「動的ページには多くの種類があり、簡単にアーカイブに保存できるものと、完全に崩れてしまうものがあります」と書かれています。「動的なページが標準的な html をレンダリングする場合、アーカイブは機能します。ページが元のホストとの対話を必要とするフォーム、JavaScript、その他の要素を含んでいる場合、アーカイブは元のサイトの機能を再現できません」。動的なデータ・ビジュアライゼーションやニュースアプリは、現在最も先端的なデータジャーナリズムのストーリーですが、既存の Web アーカイブ技術では捕捉することができません。

また、さまざまな制度上の理由から、この種のストーリーは CMS の外部で構築される傾向があります。したがって、たとえデータ・ビジュアライゼーションやニュースアプリのアーカイブが可能だとしても（このアプロ

ーチでは一般的に不可能です)、自動フィードは CMS の内部にないため、それらを捉えることができません。それは複雑な問題で、簡単な答えはありません。私はデータジャーナリスト、図書館員、コフォームター科学者のチームと一緒に、この茨の道を解決するための技術を開発しようとしています。

私たちは、再現可能な科学研究の手法を借りて、人々が将来、コンピュータで今日のニュースを読めるようにすることを目指しています。私たちは、計算科学の実験で使用されるコード、データ、サーバー環境を収集する ReproZip というツールを応用しています。ReproZip は、Webrecorder.io のようなツールと統合することで、ストーリーとソフトウェアの両方(の性格を併せ持つ)ニュースアプリを収集し保存することができそうです。Web やモバイルベースのデータジャーナリズム・プロジェクトは、他の幅広いメディア環境、ライブラリ、ブラウザ機能、Web の実体(これらも継続的に変化する可能性があります)に依存し、それらと関連して存在するため、ReproZip を使用して、複雑なデータ・ジャーナリズム・オブジェクトを Web 上で機能させるためのリモート・ライブラリーとコードを収集し、保存できることが期待されます。私たちの仮説を証明するには、あと 1、2 年かかるでしょう。その間、データジャーナリズムを将来にわたって確実に保存するために、どのデータチームでもできることがいくつかあります。

動画を撮る。この戦略は、テレビゲームの保存から借用したものです。ゲーム機がなくなっても、ビデオでプレイスルーすれば、元の環境でゲームを見ることができます。データジャーナリズムのストーリーにも同じことが言えます。映像は、その映像が示すものを説明するプレーンテキストのメタデータとともに、一カ所に保存します。新しいビデオ・フォーマット(VHS から DVD へ、あるいは DVD からストリーミング・ビデオへ)が登場するたびに、すべてのビデオをこの新しいフォーマットにアップグレードします。

後世に残すために縮刷版を作る。Django-bakery のようなライブラリを使えば、動的なページを静的なページとしてレンダリングできます。これは「ベイクアウト」と呼ばれることもあります。何千ものレコードを持つデータベースでも、それぞれの動的なレコードを静的なページとしてベイクアウトすれば、ほとんどメンテナンスが不要になります。理論的には、これらの静的ページをすべて組織のコンテンツ管理システムにインポートすることができます。ベイクアウトは、ローンチ当初から行う必要はありません。データプロジェクトを動的なサイトとして立ち上げ、数ヵ月後にトラフィックが減少したら、静的なサイトに変更することも可能です。一般的には、できるだけシンプルなバージョンを作成し、同時期に公開された他のすべてのストーリーと同じデジタルロケーションに置いて、作品をアーカイブシステムに適応させるとよいでしょう。

将来について考える。ジャーナリストは、記事を公開して[すぐ]次の仕事に移ることを考えがちです。その代わりに、データストーリーの発表計画と並行し、終了後の計画を立ててみましょう。OpenNews のブログ Source にあるマット・ウェイトのストーリー「Kill All Your Darlings」は、データジャーナリズムの記事のライフサイクルについて考える素晴らしいガイドとなるものです。いずれ、あなたは昇進するか、新しい組織に移ることになります。あなたのデータジャーナリズムが、あなたがいなくなった後も生き残るようにしたいものです。

図書館、知の記録機関、商業アーカイブと協力する。個人のジャーナリストとして、あなたは自分の作品のコピーを絶対に取っておくべきでしょう。しかし、将来ジャーナリズムを探すときに、あなたのクローゼットの中の箱やハードディスク、あるいはあなたの個人サイトを見る人はいないでしょう。彼らは Lexis-Nexis や ProQuest などの巨大な商用リポジトリを探すのです。商業的保存とデジタル・アーカイブの詳細については、キャスリン・ハンセンとノラ・ポールの共著、「Future-Proofing the News:Preserving the First Draft of History（2017）」を参照してください。これは、ニュース・アーカイブの状況と、ニュースの保存に関する技術的、法的、組織的な課題を理解するための標準的なガイドです。

脚注

1. www.murderdata.org/

参考文献

Boss, K., & Broussard, M. (2017). Challenges of archiving and preserving born-digital news applications. *IFLA Journal*, *43*(2), 150–157. doi.org/10.1177/0340035216686355

Broussard, M. (2014, April 23). Future-proofijing news apps. *MediaShift*. mediashift.org/2014/04/future-proo ngnews-apps/

Broussard, M. (2015a). Preserving news apps present huge challenges. *Newspaper Research Journal*, *36*(3), 299– 313. doi.org/10.1177/0739532915600742

Broussard, M. (2015b, November 20). The irony of writing about digital preserva- tion. *The Atlantic*.www.theatlantic.com/technology/archive/2015/11/the-irony-of-writing-about-digital-preservation/416184/

Fisher, T., & Klein, S. (2016). A conceptual model for interactive databases in news. GitHub. github.com/propublica/newsappmodel

ガーディアンから Google ニュースラボへ：データジャーナリズムに携わった 10 年

文：サイモン・ロジャース

概要

データジャーナリズムの過去 10 年を、第一人者の一人であるサイモン・ロジャースの職歴というレンズを通して振り返ります。

キーワード：データジャーナリズム、ガーディアンの Datablog、ウィキリークス、オープンデータ、透明性、表計算ソフト

私がジャーナリストになりたいと思ったのは、小学校 1 年生か 2 年生だった時で、まさかデータを扱うことになるとは思いもよりませんでした。今、毎日データを使って仕事をしていると、自分がいかに幸運だったかがわかります。慎重に練られたキャリアプランの結果ではまったくありません。ただ、適切な時期に適切な場所にいただけなのです。このような経緯は、2009 年、そして 19 年のデータジャーナリズムの状況について多くのことを物語っています。[09 年当時、]シカゴ出身で、ワシントン・ポストに勤務した後、近隣住民を対象としたニュース・ディスカッション・サイト「EveryBlock」を立ち上げた開発者のエイドリアン・ホロバティが、ロンドンのファリングドン・ロードにあるガーディアンの教育センターで、ニュースルーム向けの講演をしに来たのです。

私は当時、オンラインを経て、科学セクションの編集を担当し、紙媒体（当時は重心が置かれていた）のニュース編集者でした。ホロバティが、データを使ってストーリーを伝え、人々が世界を理解する手助けをすることについて話せば話すほど、私の中で何かが動きました。可能だということだけでなく、実際にやってみることができるのです。もしかしたら、私はデータを扱うジャーナリストになれるかもしれない。「データジャーナリスト」です。ニュースエディターとしてグラフィックデスクを担当したことで、マイケル・ロビンソンの才能あふれるチームの中で、私の世界の見方を変えたデザイナーと仕事をする機会ができました。そして、ビジュアル分野のポートフォリオが増えるにつれて、たくさんの数字が蓄積されていることがわかりました。ガーディアンのオープン API を立ち上げたマット・マカリスターは、これを「マザーロード」と表現しています。GDP データ、二酸化炭素排出量、政府支出データ、その他多くのデータをきれいに整理し、すべて Google スプレッドシートとして保存し、次に必要なときにすぐに使えるようにしました。このデータをオープンデータとして公開したらどうなるでしょう？PDF ではなく、誰でもすぐに使える、アクセス可能なデータです。これが、ガーディアンの Datablog で行ったことです。

当初は 200 の異なるデータセットがありました。犯罪率、経済指標、紛争地域の詳細、さらにはファッション・ウィークやドクター・フーの悪役まで。そして、データはあらゆるものに応用できるのだということに気づき

ました。 しかし、それはまだ奇妙な存在でした。データエディターという職業は珍しく、データチームを設置している報道局はほとんどありませんでした。ニュースの打ち合わせで「データ」という言葉を使うだけで、鼻で笑われるような状況でした。これは「まともな」ジャーナリズムではない、というわけです。

しかし、2009 年はオープンデータ革命の始まりの年でした。米政府のデータハブである data.gov は、その年の 5 月にわずか 47 件のデータセットでスタートしました。世界中の国や都市がオープンデータポータルを立ち上げ、活動家たちはより多くのデータへのアクセスを要求していました。それから1年とたたないうちに、読者の協力で何千人もの国会議員の経費をクラウドソース化できるようになったのです。同じ時期に、英国政府は COINS（Combined Online Information System）という究極の支出データセットを発表しました。ガーディアンのチームは読者にその探索を手伝ってもらうために、インタラクティブなエクスプローラーを作りました1。そのデータから記事が生まれると、「どうすればもっと多くのデータを入手できるか」という問いが出てきました。[答えが出るには]そんなに待ちませんでした。

その答えは、スウェーデンを拠点とする新組織「ウィキリークス」が、過激な透明性アジェンダを引っ提げもたらしたのです。ウィキリークスについて今日どう思うかは別にして、この組織が近年のデータジャーナリズムの歴史に与えた影響は相当なものです。ここには、まずアフガニスタン、次いでイラクの戦場から、何千もの詳細な記録が大量に投下されました。それは巨大なスプレッドシートの形で提供され、ガーディアンの調査チームが初めて扱うものとしては巨大すぎるものでした。ベトナム戦争の実態を明らかにしたペンタゴン・ペーパーズ（米国防総省発行の文書）よりも規模が大きく、詳細で、死傷者数、地理的位置、詳細、カテゴリーを含む事件のリストも含まれていました。例えば、イラクでは簡易爆弾の被害が増加し、道路が危険な状態になっていることが分かりました。そして、そのデータと熟練した戦争記者の伝統的な報道技術が組み合わさったとき、世界の戦争の見え方が変わったのです。全世界にインパクトを与えるようなコンテンツを作ることは難しくありませんでした。例えば、スプレッドシートのジオデータはマッピングに適しており、それを支援する新しい無料ツールがあったのです。Google Fusion Tables です。

そこで私たちは、イラクで少なくとも 1 人の死者が出ているすべての事件を大づかみで地図化しました。24 時間以内に、1 時間で作ったコンテンツが世界中で見られるようになり、ユーザーは戦場をよりリアルに感じられるようになりました。また、データが構造化されているため、グラフィックチームは洗練されたリッチなビジュアルを作成し、より詳細なレポートを提供することができました。

ハンドブック初版公開前年の 2011 年末には、「暴動を読む」プロジェクトが、1960 年代にフィル・マイヤーが行ったコンピュータ支援報道の技術を、イングランドで発生した暴力事件に応用しました（Robertson, 2011）。マイヤーは、1960 年代後半のデトロイト暴動に関する報道に社会科学の手法を応用したのです。ガーディアンのポール・ルイスが率いるチームは、その年のイングランド全土で発生した暴力事件に敷衍し、その作業の重要な部分としてデータを取り入れたのです。これらは、一面を飾る、データに基づいたストーリーでした。

しかし、私たちが情報を消費する方法にはもう一つの変化が起きており、それは急速に進みました。2010 年以前は、健康関連の記事以外で「バイラル」という言葉を聞いた記憶がありません。今は違います。データジャーナリズムの台頭は、ソーシャルメディアの普及と軌を一にします。私たちは世界中のユーザーに記事を売り込むのにツイートを活用し、その結果トラフィックが増え、またこの種のデータに基づくストーリーを探すユーザーが増えました。ビジュアルや数値は、数秒で何千人もの目に触れることができるのです。ソーシャルメディアはジャーナリズムを一変させ、データジャーナリズムはニッチから主流へ移行しました。

［具体例の］一つは、消費者との関係が変わったことです。以前は、記者の言葉は神聖なものとされていましたが、今は何百万人の中の一人の声に過ぎません。あるデータセットでミスを犯せば、500 人がすぐに知らせてくれるでしょう。Twitter でデザイナーと地図の配色について長く（そして深く）語り合い、自分の作品に手を加えたことを思い出します。シェアすることで、私の仕事はより良いものになりました。このようなコラボレーションの精神は、今日でもデータジャーナリズムに根差しています。

ハンドブック初版は、ロンドンの Mozilla Festival の参加者によって作られました。データに関するイベントが増えるにつれ、データジャーナリストが一緒に仕事をし、スキルを共有する機会も増えてきました。イラク［戦争の記録］とウィキリークスの発表が大西洋を越えた協力の素晴らしい最初の例であるとすれば、次に紹介するのは何百人もの記者を含む全世界的な報道に成長した実例です。

スノーデンのリークやパナマ文書は、記者が世界中で連携してストーリーを共有し、互いの仕事を発展させたという点で注目されます2。選挙当日の投票問題をリアルタイムで監視するために共同報道の手法を用いた Electionland のような活動をご紹介します。私もリアルタイムで Google のデータを提供し、それらの懸念をリアルタイムで可視化することにしました。今に至るまで、Electionland は 1 日の取材活動としては史上最大規模であり、当日は 1000 人以上のジャーナリストが参加しました。Electionland は、私たちが最初の数年間にやっていたことと直結しています。私はプロジェクトを列挙したいのではありません。ガーディアンだけでなく、世界中のニュースルームに共通するデータジャーナリズム初期の広がりを強調したいのです。ニューヨーク・タイムズ、ロサンゼルス・タイムズ、アルゼンチンのラ・ナシオンなど、世界中のジャーナリストがデータ主導のストーリーを革新的な方法で伝え、新しい働き方を見出しました。これがハンドブック初版の背景だったのです。

アルゼンチンのラ・ナシオンがその良い例です。熱心な記者たちでつくる小チームは、Tableau［当時は新しいツール］を使った可視化を独学で学び、これを情報の自由の記事と組み合わせて、中南米におけるデータジャーナリズムの口火を切りました。データジャーナリズムは、少数の一匹狼の領域から、多くの主要なニュースルームで確立された手法になりました。しかし、当時から一つの傾向は明らかでした。報道に新しい手法が導入されるたびに、データが重要な役割を果たすだけでなく、データジャーナリストが台頭したのです。3 年足らずの間に、クラウドソーシングはニュースルームのツールとして定着し、ジャーナリストはデータを見つけ、データベースを使って膨大な量の文書を管理し、データセットを公開し、複雑なニュース記事にデータ駆動型の分析手法

を適用するようになりました。これは、ジャーナリズムの分野でのみ起きたことではありません。オープンデータ・ポータルの設置にとどまらず、透明性に関する国際的な大きな流れの影響によるものなのです。

［時代の流れが］Free Our Data、Open Knowledge Foundation、シビックテック・グループによるキャンペーンなど、英国政府に対する圧力を高め、ニュース・データセットを一般利用のために開放し、誰もが探索できるように API を提供するよう求めました。また、OpenRefine、Google Fusion Tables、Many Eyes、Datawrapper、Tableau Public など、強力な無料のデータ可視化・クリーニングツールへのアクセス向上も（この潮流を）後押ししました。これらの無料ツールと多くの無料公共データへのアクセスが相まって、より多くのビジュアライゼーションやデータプロジェクトが制作されるようになったのです。テキサス・トリビューンやプロパブリカなどのニュースルームは、このデータを中心に事業を展開するようになりました。この仕組みがお分かりですか？データ、簡単な処理、データの可視化、さらにデータ…という好循環が生まれるのです。データが世に出れば出るほど、そのデータを使った作業が増えれば増えるほど、より多くのデータを公開するよう求める圧力が強くなります。

私が「データジャーナリズムは新しいパンクである」という記事を書いたのは、まさにその点を指摘するためです。私たちは、創造性を自由に発揮できる場所にいるのです（Rogers, 2012）。そしてまた、その仕事はやがて主流となるものでもありました。データは万能ではありません。ジョナサン・グレイ（2012）が書いているように、「データ、データ技術、データ駆動型のあらゆるものに関する現在の盛り上がりは、この機械可読で構造化されたものこそ特別だと思わせるかもしれません」。それは、記者が組み立てなければならない証拠というパズルの 1 ピースに過ぎません。しかし、利用できるデータが増えれば増えるほど、その役割は変化し、さらに重要になります。巨大なデータセットにアクセスし、分析できることが、私の次の転職の大きな動機でした。

2013 年、私はカリフォルニアに移り住み、Twitter 社初のデータエディターになりました。当時、欧米を中心とした出版業界では、データが主流になりつつありました。ニューヨーク・タイムズの Upshot やネイト・シルバーの FiveThirtyEight など、多くのデータジャーナリズムサイトが数週間のうちに立ち上がりました。世界の読者は、複雑なトピックを高度に可視化したものを評価するビジュアルリテラシーを高めていました。世の中の人々がデータ・ビジュアライゼーションに慣れ親しんでいるという根拠は何なのかと聞かれるでしょう。ネット上で大きな反響を呼ぶビジュアライゼーションを制作することは、以前より難しくなっているというのが私の経験上の実感です。以前は誰もが「おおっ」と驚いていたのに、今では肩をすくめる以上の反応を引き出すのは難しくなりました。私がデータジャーナリズムに取り組むために Google ニュースラボに参加した 2015 年頃には、この分野ではかつてないほど大規模で大きなデータセットにアクセスできるようになっていました。毎日、何十億もの検索があり、そのうちのかなりの割合がこれまで見たこともないようなものです。記者はいっそうそのデータを取り込み、ツイートや Facebook の「いいね！」とともに分析するようになっています3。これは、いわば現代生活の排気であり、私たちの今日の生き方についての洞察を提供するものです。

データジャーナリズムもまた、かつてないほど広まりつつあります。2016 年、データジャーナリズム・アワードには過去最高の 471 作品の応募がありました。しかし、2018 年のアワードには 700 件近くが寄せられ、半数以上が小規模なニュースルームからで、世界各地から多くの応募がありました。そして、それらのエントリはますます革新的になっています。バズフィードでのピーター・アルダスの作品（Aldhous, 2017）に見られるように、人工知能、つまり機械学習はデータジャーナリズムのツールとなっているのです。一方、仮想現実や拡張現実のような新しい技術へのアクセスは、データを使って新しい方法でストーリーを語る可能性を開くものです。私はデータジャーナリズムがどのように変化しうるか、そしてそれをサポートするために私たちに何ができるかを想像することが仕事ですが、新しい技術をより多くの記者が仕事に取り入れやすくするにはどうしたらよいかを考えています。例えば、私たちは最近、デザインスタジオの Datavized と協力して、データを音に変換するビジュアルツール TwoTone を構築しました4。Google のデータジャーナリストはどんな仕事をしているのでしょう？大規模で豊富なデータセットを使ってストーリーを伝えるだけでなく、才能あるデザイナーと、ニュースデータの可視化の未来や、ジャーナリズムにおける新しいテクノロジーの役割を想像することができます。

私の役割のひとつは、新しいテクノロジーを適切な使用例や状況にどのようにマッチさせれば、適切かつ有用になるか探求することです。また、ジャーナリストがデータやデジタル技術を使って、どのように新しい方法でストーリーを伝えているかを探ることも含まれます。例えば、最近のプロジェクトのひとつ、エル・ユニバーサルの「Zones of Silence」では、言語処理を使って麻薬カルテルの殺人事件の報道を分析し、公式データと比較して、両者のギャップを報道における沈黙の領域と捉え、ジャーナリズムにおける AI の活用を実証しています。私は、AI の API やデザインリソースにアクセスすることで、彼らの活動を支援しました。

私たち全員にとって、課題は大きいと言わざるをえません。私たちは皆、ますますモバイルで情報を消費するようになり、それに伴う課題がもたらされています。今や半数以上の人がスマートフォンやその他のモバイルデバイスでニュースを読んでおり、（ロイターのニュース消費調査（Newman ら、2017）によると、3 分の 1 の人がトイレでニュースを読んでいるそうです）フルスクリーンの複雑なビジュアライゼーションの時代は終わりを告げました。つまりニュースルームのデザイナーは、小さな画面と注意力の持続時間の短縮に対応してデザインしなければならなくなったということです。

過去を学ぶことに立ちはだかる新しい問題もあります。コードは死に、ライブラリは朽ち、最終的にはジャーナリズムの最も野心的な仕事の多くがただ消えてしまうのです。ガーディアンの MPs' expenses や EveryBlock、その他のプロジェクトはすべて、消えゆく組織の記憶に埋もれてしまいました。データジャーナリズムの消滅という問題は、すでにいくつかの革新的なアプローチの対象になっています（ブルサードの章をご覧ください）。長期的に見れば、これには適切な投資が必要ですが、コミュニティがそれを実現しようとするかどうかは断言できません。そして、私たちはより広く、ますます憂慮すべき問題に直面しています。それは「信頼」です。データ分析には常に解釈や意見の相違がつきまといますが、優れたデータジャーナリズムはそれを克服することができます。ニュースに対する信頼や事実の共有が日々疑われている今、データジャーナリズムは事実

と証拠をわかりやすく明るみに出すことで、私たちに道を示してくれるのです。このように、変化はあっても、この分野では不変のものがあります。データジャーナリズムには長い歴史がありますが5、2009 年当時、データジャーナリズムは、共通の真実に迫る重要な方法で、私たち全員が支持できるものだと考えられていました。今、その必要性はかつてないほど高まっています。

脚注

1. www.theguardian.com/politics/coins-combined-online-information-system

2. パナマ文書にまつわる大規模なコラボレーションについては、本編のディアス＝ストラック、ガレゴ、ロメラの章を参照されたい。

3. これに関するさらなる視点は、「データ、プラットフォーム、アルゴリズムを調査する」の章を参照。

4. twotone.io

5. 例えば、本巻のアンダーソンとコーエンの章を参照。

参考文献

Aldhous, P. (2017, August 8). We trained a computer to search for hidden spy planes. This is what it found. *BuzzFeed News*. www.buzzfeednews.com/article/peteraldhous/hidden-spy-planes

Gray, J. (2012, May 31). What data can and cannot do. *The Guardian*. www.theguardian.com/news/datablog/2012/may/31/data-journalism-focused-critical

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2017). *Digital News Report 2017*.

Reuters Institute for the Study of Journalism. reutersinstitute.politics.ox.ac.uk/sites/default/les/Digital%20News%20Report%202017%20web_0.pdf

Robertson, C. (2011, December 9). Reading the riots: How the 1967 Detroit riots were investigated. *The Guardian*. www.theguardian.com/uk/series/reading-the-riots/2011/dec/09/all

Rogers, S. (2012, May 24). Anyone can do it. Data journalism is the new punk. *The*

Guardian.www.theguardian.com/news/datablog/2012/may/24/data-journalism-punk

データジャーナリズムとシビックテックの連帯

文：ステファン・バーク

概要

データジャーナリズムは他の形態のデータ作品やデータ文化とどのように連帯するのでしょうか。

キーワード：シビックテック、ゲートキーピング、専門家の境界、データジャーナリズム、情報の自由（FOI）、データベース

コンピュータを使った報道は（調査報道）ジャーナリストだけのものとされてきましたが、データジャーナリズムはテクノロジー部門や他の形態のデータワーク、データ文化との連帯という特徴があります。コンピュータ支援報道と比較すると、米国や欧米におけるデータジャーナリズムの出現は、ニュースルーム内外のいくつかの動きと交錯しています。

すなわち、オープンデータへの取り組みや情報漏えいを筆頭に、オンラインでのデータの利用可能性が高まっているのです。また、ニュースルームが開発者を雇用し、編集チーム内に統合することでデータへの対処を改善し、インタラクティブな Web アプリケーションを提供できるようになりました。さらに、技術的スキルを「公益」に役立てる方法としてジャーナリズムに魅力を感じるさまざまな「テック・フォー・グッド」運動が活況です。2000 年代に欧米などでデータジャーナリズムが出現し、人気を博して以来、この動きはニュースルームに技術者が流入することに貢献しています。しかし、結果として生じるデータジャーナリストと他の形態のデータ作業との協働は、地域によって異なっています。

さらに、データジャーナリズムは、持続可能なビジネスモデルを開発するためにメディア組織が長らく苦境にあることに呼応した、新しい起業家的なジャーナリズムの形態と結びついています。こうした新しいタイプのメディア組織、例えば非営利のニュースルームや BuzzFeed のようなベンチャー支援のニュース新興企業は、ジャーナリズムの「復活」または「改善」を熱望するあまり、従来の境界線を疑う傾向があり、テクノロジーとデータはしばしばその傾向を後押しする重要な役割を果たします（Usher、2017；Wagemans et al.、2019 を参照）。

データジャーナリズムと他の形態のデータワークやデータ文化との絡み合いは、新たな依存関係を生み出すだけでなく、セクター間の新たな形態のコラボレーションを可能にする相乗効果も生み出します。ここでは、データジャーナリズムとシビックテックの密接な関係を例に挙げます。なぜなら、両者は多くの場所で同時期に出現し、早い段階から相互に影響し合っていたからです。

シビックテックとは、市民が政府と関わり、その責任を追及することを容易にし、市民の力を高めることを目的としたツールを開発することです。シビックテック・プロジェクトの例としては、議会の監視サイトで、特に議

会の演説にアクセスしやすくする OpenParliament、ユーザーが情報公開請求を提出したり見つけたりするのを手助けするウェブサイト WhatDoTheyKnow、地方自治体への問題の報告を簡単にする FixMyStreet などがあります1。

シビックテクノロジストとデータジャーナリストには、いくつかの重要な共通点があります。まず、両グループの実務家の多くは、オープンソース文化に基づき、共有、オープンソースツールの使用、データ標準化を推進しています。第二に、データジャーナリストとシビックテクノロジストは、公的機関、クラウドソーシング、その他のソースを問わず、データに大きく依存しています。第三に、手段は異なるものの、両者とも市民に力を与え、当局に責任を負わせる公共サービスを提供することを目指しているのです。このように、データのスキル、志、共有への取り組みが重なり合うため、シビックテクノロジストとデータジャーナリストは互いを補完し合う存在として認識しています。また、報道機関やナイト財団などの財団からの支援、Hacks/Hackers などの草の根的な取り組みにより、データジャーナリストとシビックテクノロジストの継続的な交流やコラボレーションが生まれました。

ジャーナリズムの「核」の拡大と強化を巡る緊張関係

ドイツと英国のケーススタディに基づき、データジャーナリストとシビックテックがどのように相互補完しあっているかを検証したところ、両者の関連は二つのプラクティスを中心にできていることが分かりました。それはファシリテートとゲートキーピングです（Baack, 2018）。ファシリテートとは、他者が自ら行動を起こせるようにすることであり、ゲートキーピングは、公に関連する情報の門番となる伝統的なジャーナリズムの役割モデルを指します。

シビックテクノロジストが開発した議会監視サイトは、議会での演説を検索するなどして、利用者が自ら情報を得られるようにすること（ファシリテート）を目的としていますが、専門家が適切と判断した情報を利用者に押し付けること（ゲートキーピング）ではありません。ファシリテーターは個人のエンパワーメントを目的とし、ゲートキーピングは公的な議論を誘導し、影響を与えることを目的としているのです。データジャーナリストとシビックテクノロジストの関連を特徴づけるのは、ファシリテートとゲートキーピングが補完的であり、相互に補強し合うことができるという点です。シビックテックのアプリケーションは一般市民の利便性を高めるだけでなく、データジャーナリストも自らの調査に利用することができます。

一方、ジャーナリストによる調査は、特定の問題に注意を向けさせ、人々が支援サービスを利用するよう促すことができます。さらに、情報権はファシリテートとゲートキーピングの両方に不可欠であり、さらなる相乗効果を生みだします。例えば、データジャーナリストは（職業上の）排他的権利を利用してデータを入手し、シビックテックと共有することができます。一方、ジャーナリストはシビックテックが情報自由権の強化やオープンデータ政策の提唱をすることで利益を得ることができるのです。

データジャーナリズムとシビックテックの関係では、新しい起業家的なジャーナリズムが特に重要な役割を担っています。例えば、プロパブリカは、一般市民と政府との関わりを促進するだけでなく、自らデータを収

集、クリーニング、分析するリソースや専門知識を持たない地方のニュースルームの記者によるジャーナリズム調査を支援するために、検索可能な大規模データベースをいくつか開発しました。また、ドイツの非営利団体 Correctiv は、同様のアプローチで、Open Knowledge Foundation Germany の情報公開サイトをいくつかのアプリケーションに統合しました。この統合により、ユーザーは直接、さらなる情報を要求できるようになり、一度入手した情報は自動的に Correctiv のデータベースに追加されるようになりました2。これらの例は、他のデータ文化圏の慣行や価値観を取り入れることで、従来のジャーナリズムの概念を拡張する組織が増加していることを示していますが、その逆もあります。

データジャーナリストは、ゲートキーパーやストーリーテラーとしての専門的なアイデンティティを受け入れることで、他の分野の慣習や願望の類似性に反応します。こうしたジャーナリストは、必ずしもシビックテックを拒否しているわけではないものの、その反応は、伝統的な調査報道という概念に近い、より専門化したジャーナリズムに即したものといえます。

境界が曖昧になることで生まれる可能性

つまり、データジャーナリズムは、他のデータワークやデータ文化の分野と絡み合うことで、「ジャーナリズム」の理解や実践のあり方をより多様化し、それが伝統的な価値やアイデンティティの拡大や強化につながるのです。ジャーナリスト自身も研究者も、データジャーナリズムをより広範な技術的・文化的・経済的変革の中に組み込まれた現象として捉えるようになります。ここでは、データジャーナリストとシビックテクノロジストの連帯に焦点を当てましたが、こうした特定のケースによらず、データジャーナリストにとって重要な二つの教訓を指摘しておきたいと思います。

曖昧な境界線から利益を得ること。ジャーナリストは、他の分野に対する専門的な境界がないことを問題視しがちですが、シビックテックとデータジャーナリストのシナジーは、曖昧な境界が利点にもなり得ることを示しています。データジャーナリストは、境界を問題視するのではなく、他の分野のデータ作業と相乗効果があるかどうか、そして、どのようにすればその恩恵を最もよく受けられるかを考える必要があります。必ずしもジャーナリストがファシリテーターにならなくてはいけないわけではありません。ジャーナリストも、自分の職業的アイデンティティを犠牲にすることなく、利益を得る方法を見つけようとすることは可能です。

プロのジャーナリズムの多様性を受け入れる。私の研究結果は、「ジャーナリズム」がいかに多様な、より専門化した主体によって提供されるようになっているかを反映しています。私がインタビューしたジャーナリストの何人かは、このような多様化を懸念しています。彼らは、ファシリテーターという慣行手法を採用するメディア組織は、「ハードな」調査報道の概念を弱めるかもしれないといいます。しかし、ジャーナリストは、将来的に単一の明確なジャーナリズムの形態が存在する可能性が低いことを認識すべきです。

とどのつまり、他の形式のデータワークやデータ文化との歴史的・現代的なつながりの両方をより強く認識することは、ジャーナリストが自らの役割を振り返り、新たな依存関係だけでなく、自らの使命を支え、拡大するために利用できる相乗効果の可能性をよりよく認識するために役立つのです。

脚注

1. Openparliament.ca, WhatDoTheyKnow.com, FixMyStreet.com

2. correctiv.org

参考文献

Baack, S. (2018). Practically engaged. The entanglements between data journalism and civic tech. *Digital Journalism*, *6*(6), 673–692. doi.org/10.1080/21670811.2017.1375382

Usher, N. (2017). Venture-backed news start-ups and the fijield of journalism. *Digital Journalism*, *5*(9), 1116–1133. doi.org/10.1080/21670811.2016.1272064

Wagemans, A., Witschge, T., & Harbers, F. (2019). Impact as driving force of journalistic and social change. *Journalism*, *20*(4), 552–567. doi.org/10.1177/1464884918770538

データジャーナリズムにおけるオープンソースコーディングのプラクティス

文：ライアン・ピッツ、リンジー・ムスカート

概要

ジャーナリズムのためのオープンソースコーディングの課題と、成功したプロジェクトに共通する特徴を説明します。

キーワード：オープンソース、プログラミング、コーディング、ジャーナリズム、ツール開発、コードライブラリ

想像してみましょう。数人のジャーナリストが協力して政府のWebサイトから記録をかき集め、データに変換し、そのデータを分析してパターンを探し、読者にストーリーを伝えるビジュアライゼーションを発表します。このようなプロセスは、毎日、世界中のニュースルームで展開されています。多くのニュースルームでは、各ステップが少なくとも部分的にはオープンソースソフトウェアに依存しており、コミュニティでテストされたツールをつなぎ合わせて、以前のどの方法よりも高速なワークフローを実現しています。しかし、今日のデータジャーナリズムのワークフローの一部となっているのは、オープンソースソフトウェアだけではなく、オープンソースの哲学でもあるのです。私たちは、イベントやコミュニティ・チャンネル、ソーシャルメディアを通じて、互いに知識やスキルを共有しています。

私たちは、方法論とデータを公開し、同僚に私たちの仮説を正してもらい、読者に私たちの結果を信頼する根拠を与えます。このようなオープンで協力的なアプローチは、私たちのジャーナリズムをより良いものにします1。フィードバックや外部からの貢献を求めるたびに、私たちの仕事はより弾力的なものになるのです。記事でデータを使用する際の問題点を他の誰かが発見するかもしれませんし、私たちのソフトウェアをより良くする新機能を提供してくれるかもしれません。このようなプラクティスは、自分たちのプロジェクトや組織の枠を超えて、より大きな利益をもたらす可能性もあります。

私たちはふつう、独力で開発したツールや技術だけに頼って、大きなプロジェクトに飛び込むことはしません。その代わりに、メンターから学び、カンファレンスで話を聞き、気に入ったプロジェクトがどのように作られたかを学びながら、他の人たちの仕事の上に構築していくのです。OpenNews では、オープンソースプロジェクトでジャーナリストと協力し、開発者のコラボレーションをサポートし、「The Field Guide to Open Source in the Newsroom」を執筆してきました。ここでは、データジャーナリズムにおけるオープンソースの役割について、共通の課題や成功したプロジェクトの特徴など、私たちが学んだことを振り返っていきます。

共通の課題

オープンな仕事はやりがいがあって楽しく、その過程で多くのことを学ぶことができますが、いつでも簡単というわけではありません。成功には、オープンソースプロジェクトがしばしば直面する課題に目を向けることが必要です。

前例を作る。編集者や法務チームなどに、自分の作品を「無料で提供する」ことが良いアイデアだと説得するのは難しいと感じることがあります。法律、ビジネス、風評、持続可能性などの懸念があるかもしれません。そこで私たちは、ジャーナリストの方々と協力して、オープンソースのツールやプロセスがもたらすメリット（より強固なコード、コミュニティの好意、信頼性の向上など）を文書化しています。

人は移り変わるものであり、テクノロジーも同じ。チームの主要メンバーが別の仕事に就くと、オープンソースのプロジェクトの維持管理に使える時間が、彼らとともに消えてしまうことがよくあります。例えば、数年前、ニューヨーク・タイムズは、巨大なデータセットをブラウザ内で高速にフィルタリングする JavaScript フレームワーク、Pourover をリリースしました。Pourover は広く共有され、コミュニティが形成され始めました。しかし、主要な開発者の一人が転職し、チームは同様の問題を解決する新しいツールを探し始めることになったのです。決して Pourover のコードや計画を非難しているわけではなく、プロジェクトの寿命が想像と違うだけということもあるということです。

成功へのプレッシャー。直感に反するようですが、自分が作ったものに人々が熱狂していることを知ると、作り込まれていない作品が生まれることがあります。突然の爆発的な人気は、作り続け、バグを修正し、コミュニティの貢献に応えなければならないというプレッシャーとなります。エリオット・ベントレーは、本業の問題を解決するために書いた Web アプリ、oTranscribe をリリースした後、これらすべてのことと格闘していました。これは、彼が本業であるインタビューの音声を書き起こしのために書いたWebアプリです。数ヶ月後、彼は何万人ものアクティブユーザーを抱え、プロジェクトの将来について質問攻めに遭いました。

成功するプロジェクトの特徴

ジャーナリズムにおけるオープンソースの素晴らしい例はたくさんあります。あるニュースルームが発表し、他の多くのニュースルームが採用したプロジェクトもあれば、最初からコラボレーションが行われているものまで、さまざまです。これまで見てきた中で最も成功した取り組みには、以下に述べる特徴を一つ以上持っています。

日常で出くわす問題を解決している。多くの場合、誰かがあなたと同じような障害や一連の繰り返し作業に直面しています。マーシャル・プロジェクトは、全国の刑事司法をカバーするため、何百もの Web サイトを監視し、変化や発表がないかを調べています。URL のリストを何度も訪れることは、記者の時間を無駄にしますが、クラウド・サーバーを有効に使えば簡単にできます。Klaxon はこれらの Web サイトを監視し、変更があった場合にはアラートを送信します。非常に高速なので、ニュースルームでは公式発表前に情報を入手することもよくありま

す。このようなトラッキングは、あらゆる種類の報道に有効です。マーシャル・プロジェクトは、記者にとっての問題を解決すると同時に、他の組織にとっての問題も解決しました。Klaxon をオープンソースプロジェクトとして公開することで、開発者は何十ものニュースルームの報道を助け、その見返りとして自分たちのツールをさらに良くするためのコードの提供を受けることができるのです。

七面倒な問題を解決する。NPR のデータ／ビジュアル・チームは、グラフィックスが埋め込まれているレスポンシブ・ページの寸法を変更する方法を必要としていました。読者がモバイルデバイスを使ってニュースコンテンツにアクセスする機会が増えているため、これは重要な機能ですが、大変な作業です。NPR がこの問題を解決するオープンソースのコードライブラリ Pym.js をリリースすると、ジャーナリズムのコミュニティにはあっという間に広まりました。

ドキュメントが充実している。インターネットにコードを放り出すことと、プロジェクトが何のためのもので、どのよう使うのかを具体的に説明することとでは、大きな違いがあります。締め切りに追われると、ドキュメントを書くことの優先順位は低くなりがちですが、プロジェクトはドキュメントなしには成り立ちません。新しいユーザーは、スタート地点を必要としていますし、あなた自身も、後で自分の仕事を見直したときに感謝することでしょう。Wherewolf はちょっとした JavaScript のサービスで、ある住所がどの境界線（例えば学区や郡の境界線）の内側にあるのか教えてくれます。コードはしばらくの間更新されていませんが、ドキュメントが完全で例が豊富なため、ユーザコミュニティは成長を続けています。

貢献者を歓迎している。California Civic Data Coalition は、記者が州の選挙資金データを利用するのに役立つ一連のオープンソースツールを擁しています。このプロジェクトは、二つのニュースルームの数人の開発者のコラボレーションとして始まりましたが、学生、インターン、シビックデータ関係者、関心を持つ市民、さらにはコーディング経験の全くないジャーナリストからの貢献により、大きくなりました。これは偶然の産物ではありません。このイニシアチブでは、構築すべき機能と修正すべきバグをまとめたロードマップを作成し、専門知識のレベルごとにタスクを設定したチケットを作成し、カンファレンスに参加して、誰でも歓迎するスプリントを計画しています。

オープンソースのニュースルームプロジェクトには、成功を測るさまざまな方法があります。コミュニティを形成し、貢献を呼びかけていますか？自分の仕事をより多くの人に見てもらう必要がありますか？それとも、自分にとっての問題を解決するものを作って、他の人が同じ苦労を味わわなくて済むようにするのが良いのでしょうか？成功とはどのようなものかを決めるのはあなた自身です。目標を明確にすることで、自分自身のための正直なロードマップを作成し、他の人に正しい期待を持たせることができます。また、新しい人を迎え入れ、将来的な意思決定を説明するための、親切で事例を重視したドキュメントを作成し、他者を迎え入れる協力的な仕事のやり方を採用します。実際にやってみることで、多くのことを学ぶことができます。

脚注

1. データジャーナリズムにおけるオープンソースの実践と哲学の役割に関する異なる視点については、レオンとマゾッテの章も参照のこと。
2. opennews.org, eldguide.opennews.org/en/latest
3. eldguide.opennews.org/en/latest/Chapter01-Choosing-Open-Source/
4. newsklaxon.org/

データ封建主義：プラットフォームはどのように国境を越えた調査網を形成するか

文：ステファン・カンデア

概要

データ封建主義：プラットフォームは国境を越えた調査報道をどのように形成し、植民地化しようとするのかを探ります。

キーワード：越境調査、ネットワーキングの政治経済、社会技術的アクセス制御、ラディカルな共有、データ封建主義、プラットフォーム

国境を越えた調査報道のプラットフォーム化は、社会のプラットフォーム化（Dijck et al.）という現代の潮流と同じ技術実証主義が支える、台頭著しい現象を指します。国境を越えた調査のためのデータをホストするプラットフォームは、2010 年頃から、流出したデータで調査を行うことで注目され始めました。プラットフォームを利用した大規模なジャーナリズムの共同作業の最も顕著な例は、ピュリッツァー賞を受賞した「パナマ文書」です。パナマ文書の調査に関わった 500 人のジャーナリストによるデータの照会と報告を組織化するため、国際調査報道ジャーナリスト連合（ICIJ）は Global I-Hub というプラットフォームを開発しました（Wilson-Chapman, 2017）1。ライル（2017）はこのプラットフォームを「特別に開発された技術であって、情報を調査し、配信し、オンラインのニュースルームでジャーナリストを結びつけ、ジャーナリストが一つのグローバルなチームとして働くことを保証するのに用いられる」と説明しています。

ICIJ の編集・研究スタッフは「ICIJ の仮想オフィスで、ジャーナリストのための Facebook」と呼んでいます Hare, 2016; Raab, 2016。データと国境を越えた調査は完璧にマッチし、独立したジャーナリズムの共同作業に力を与えるものとされています（Coronel, 2016; Houston, 2016）。ICIJ や OCCRP（Organised Crime and Corruption Reporting Project）などの組織は、世界中の何百人ものジャーナリストの中から厳選したグループに、外部からはアクセスできないプライベートな電子プラットフォーム上で照会可能な専用データセットへの無料（補助金付きとも言える）アクセスを提供しています。また、これらのジャーナリストが作成したストーリーを公開し、宣伝するためのプラットフォームも提供しています。これらの組織にとって、このようなプラットフォームを利用することは、規模の拡大と効率の向上を実現することにつながります。個人のジャーナリストにとっては、リーク情報、会社の記録、情報公開請求の結果、アーカイブ、記者のメモ、過去の記事、デジタル化された検察ファイルや裁判記録など、さまざまなデータに一箇所で安全にアクセスできることは、まさに至高の喜びといえるでしょう。特に、出張が多かったり、データの保存・処理をしたりするリソースがなく、一人で仕事をしたりしている人々にとってはそうでしょう。短期的な利益を認めつつ、こうした調査プラットフォームが、それを利用する

個々のジャーナリストや彼らが属するネットワークの立場や仕事をどのように形成しているかを巡る批判的研究は、まだ発展途上にあります。

国境を越えたジャーナリズムの領域で、このようなプラットフォームを運営する主体が非常に少なく、多数のジャーナリストがそれに依存していることが及ぼす影響は重大です。その一つは、「ビッグテック」の状況下でデータ所有に基づく「超近代的な封建制の形態」と呼ばれているものです（Morozov, 2016）。この概念は、ユーザーのデータとインタラクションの完全なコントロールが、競争に直面しない少数の企業の手に委ねられていると警鐘を鳴らすものです。このモデルには、多くの懸念があります。重要なのは、アクセス制御です。このようなプラットフォームへのアクセスは、多くの正当な理由から、多くのセキュリティ層の背後にあり、すべてのジャーナリストがアクセスできるわけではありません。本質的な問題は、誰が参加し、誰を排除するかといった判断を支配するルールと、そこから生じるかもしれない緊張や対立はどのようなものなのかということです。

このようなプラットフォームへの参加は、通常、基本的な非開示契約またはパートナーシップ契約によって管理されることになりますが、そこには、アクセスを受けるジャーナリストまたはメディアの義務が詳細に記される反面、彼らの権利についてはほとんど言及されません。このようなシステムとその管理スキームは、共有の原則に基づいて設計されたものではなく、むしろユーザーの活動の監視と契約違反の取り締まりを意識した中央集権構造として設計されているのです。

さらに、調査報道でこのモデルを採用することは、他の「シェアリングエコノミー」と同様に、調査報道の領域でプレカリアート[非正規雇用で生計を立てる人]を生み出す危険性があります。このリスクを示唆するのが、こうしたプラットフォームを運営するいくつかの組織の自己紹介です。例えば、OCCRP は、自らを「素晴らしい国境を越えた調査」をしたい「ジャーナリストの AirBnb または Uber」と表現しています（OCCRP, 2017）。実際、多くの場合、ジャーナリストはこれらの組織が所有するデータの調査に無報酬で取り組み、このデータへのアクセス料を記事で支払わなければならず、いつでもプラットフォームから排除されるリスクを抱えています。こうした不利な条件にもかかわらず、ジャーナリストはますますこうしたプラットフォーム上で活動しなければ、ゲームに参加できないようになっています。したがって、今日の大手調査ネットワーク仲介業者のビジネスモデルは、ギグ・エコノミーのデジタル・プラットフォームに似ているとも言えます。

プラットフォームへのアクセスはいつでも取り消すことができ、ガバナンスは議論の余地がなく、ユーザー活動の監視は組み込まれており、「金目は計算に入れないほうがいい」（Buzenberg, 2015）状態なのです。

何百人ものジャーナリストの無報酬の仕事と「過激な共有」のやりとりは、[ジャーナリストに]利益が還元されることなく、寄付者に「売却」されます。リークされたデータの所有権と、そうしたリークを豊かにする情報交換も、ユーザーと共有されることはありません。ユーザー間で交わされた情報によって生み出されたデータは、プラットフォームをより効率的にし、それによってより多くのインタラクション、より多くのユーザー、ひいてはより多

くの寄付者を引き付ける形でのみ提供されるのです。サービスの実質的なコストは、ユーザーにはわからないままです。調査報道界におけるこのような現状を改善するためには、どうしたらよいのでしょうか。

まず重要なのは、調査報道ネットワークにおけるプラットフォームベースのデータ共有には、データやデジタルツールの共有だけでなく、ガバナンスルールや技術設計の議論が必要だと認識することです。これらのネットワークは、公的な行動規範を策定・採用し、あらゆる種類の乱用に対処するためのアカウンタビリティ・メカニズムを備えなければなりません。それがなければ、旧メディアを崩壊に導くだけではなく、個々のジャーナリストの不安定な労働条件を増幅することになります。

第二に、少数の国境を越えた調査ネットワークをそれぞれ数千人にまで拡大することが目標であってはなりません。むしろ、互いに協力し合える多数の独立したネットワークに適用できる優れたモデルを見つけようとすべきでしょう。150 のメディアパートナーからなる単一のネットワークではなく、15 ずつのパートナーからなる 10 のネットワークを持つことが、より望ましいアプローチです。後者は、公正な競争とメディアの多元性を含む、健全なメディアシステムの原則に見合ったものです。このようなアプローチがなければ、国境を越えた調査ネットワークの参加型ポテンシャルは発揮されず、ネットワーク効果に後押しされて、少数のプラットフォームがグローバルな調査データ封建主義システムに統合されてしまうでしょう。

脚注

1. I-Hub プラットフォームに関する別の視点についてはディアス＝ストラック、ガレゴ、ロメラの章を参照。

参考文献

Buzenberg, W. E. (2015, July 6). Anatomy of a global investigation: Collaborative, data-driven, without borders.

Shorenstein Center. shorensteincenter.org/anatomy-of-a-global-investigation-william-buzenberg/

Coronel, S. (2016, June 20). Coronel: A golden age of global muckrak- ing at hand. Global Investigative Journalism

Network. https://gijn.org/2016/06/20/a-golden-age-of-global-muckraking/

Dijck, J. van, Poell, T., & Waal, M. de. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Hare, K. (2016, April 4). How ICIJ got hundreds of journalists to collaborate on the Panama Papers. Poynter.

www.poynter.org/reporting-editing/2016/how-icij-got-hundreds-of-journalists-to-collaborate-on-the-panamapapers/

Houston, B. (2016, April 14). Panama papers showcase power of a global move- ment. Global Investigative

Journalism Network.gijn.org/2016/04/13/panama-papers-showcase-power-of-a-global-movement/

Morozov, E. (2016, April 24). Tech titans are busy privatising our data. *The Guardian.* www.theguardian.com/commentisfree/2016/apr/24/the-new-feudalism-silicon-valley-overlords-advertisingnecessary-evil

OCCRP. (2017). *2016 Annual Report*. OCCRP. www.occrp.org/documents/AnnualReport2017.pdf

Raab, B. (2016, April 8). Behind the Panama Papers: A Q&A with International Consortium of Investigative Journalists director Gerard Ryle. Ford Foundation. www.fordfoundation.org...

Ryle, G. (2017, November 5). Paradise Papers: More documents, more reporters, more revelations. ICIJ. www.icij.org/inside-ic...

Wilson-Chapman, A. (2017, August 29). Panama Papers a “notable security success.” ICIJ. www.icij.org/blog/2017...

データ駆動型編集って？視聴者測定基準を使う際の注意点

文：ケイトリン・ペトレ

概要

ケイトリン・ペトレによる Chartbeat、Gawker Media、ニューヨーク・タイムズのエスノグラフィー研究をもとに、現代のニュース制作における測定基準の役割を探り、編集に測定基準を取り入れるニュースルームに提言します。

キーワード：測定基準、分析、ニュースルーム、ジャーナリズムの実践、エスノグラフィー、編集実務

2013 年 8 月 23 日、風刺ニュースサイト The Onion は、CNN のデジタル編集者メレディス・アートリーが書いたと称する「なぜマイリー・サイラスの VMA パフォーマンスが今朝のトップ・ストーリーだったのか説明します」という論説を掲載しました1。記事によれば、その答えは「とてもシンプル」です。「CNN.com をクリックしてもらい、トラフィックを増やし、その結果、広告収入を増やそうという試みだったのです。全くもって、その日の重要なニュースや、人類にとっての重要な出来事の記録や、世界をよりよくするジャーナリズムの理念に関連するものはありませんでした・・・しかし、そのおかげで、ウェブ上のトラフィックは増加したのです」。

さらに、ここでは、CNN がサイラスの記事をトップページの目立つ位置に掲載した要因として、ページビューや直帰率といった具体的な指標を挙げています。もちろん、アートリーが実際に記事を書いたわけではありませんが、とはいえメディア界の神経を逆なでする内容でした。MTV ビデオ・ミュージック・アワードでのサイラスの悪名高いパフォーマンスに関する記事が CNN.com のトップを占め、後日本物のメレディス・アートリーが明らかにしたように、その日のサイト上のどの記事よりも高いトラフィックをもたらしたからです。この偽の論説は、CNN を非難しているだけでなく、Web 測定法の時代におけるニュース価値判断の残念な現状を解説しているとも言えます。

メディア企業は、常に視聴者の人口統計や行動に関するデータを収集しようとしてきました。しかし、インターネットの追跡能力と、膨大な量のデータを保存し解析する能力によって、視聴者測定基準は近年はるかに洗練されてきています。分析ツールは、前述のページビューや直帰率に加え、訪問者の再訪問率、参照サイト、スクロールの深さ、ページ滞在時間などの変数を追跡します。これらのデータの多くは、報道機関にリアルタイムで配信されます。現代のニュースルームでは、指標ダッシュボードは事実上ユビキタスになっており、いつ、どのように参照すべきかという熱い議論が、指標そのものと同じくらい広まっています。ジャーナリズムの世界で、指標が重要な問題になっているのは驚くべきことではありません。その存在は、商業ニュースメディアにおいて常に存在する数々の緊張関係を露呈させます。中でも、ジャーナリズムの基本的な使命とは何か、そしてその使命が達成されたとき、報道機関はそれをどのようにして知ることができるのかという疑問があります。

さらに、メディア企業は利益の追求と市民的な使命とをどのように調和できるのでしょうか？はたまた、ジャーナリストと読者の区別が意味を持つ限り、ジャーナリストは読者とどのような関係を持つべきでしょうか？視聴者測定基準は報道機関のどこでも見られるようになりましたが、データがどのように作成され、報道室の文化やジャーナリストの日常業務にどのような影響を与えるかについては、これまでほとんど実証的な研究がなされてきませんでした。

私は、コロンビア大学の Tow Center for Digital Journalism の支援を受けて長期的なエスノグラフィー調査プロジェクトを実施し、測定基準の利用が記者の行動をどのように変え、それがジャーナリズムにとって何を意味するかを理解しようとしてきました。調査における主な質問は以下の通りです。まず、指標はどのように作られるのでしょうか。つまり、これらのツールを作り、販売するプログラマー、データサイエンティスト、デザイナー、プロダクトリード、マーケター、セールスマンは、視聴者行動のどの側面を、どのように測定すべきかをどう決定するのでしょう。そうした決定には、その行動を測定される人々（ニュース消費者）とそのツールを使用する人々（ジャーナリスト）の両方について、どのような考え方が組み込まれているのでしょうか。分析企業は、測定基準の価値をどんな風に報道機関に伝えているのでしょうか。

第二に、測定基準はどのように解釈されるのでしょう。測定基準はジャーナリズムにとって良いものなのか悪いものなのか、その対立するスタンスにかかわらず、測定基準の意味は明確でわかりやすいと仮定する傾向があります。しかし、数値は、それを解釈するための概念的な枠組みがなければ、それ自体では何の意味もないのです。誰が、どのように指標を理解するのでしょうか？

第三に、報道業務では測定基準はどのように使われているのでしょう。データはニュースルームがストーリーを割り当て、書き、宣伝するのに役立っているのでしょうか？昇給、昇進、解雇などの人事で、データはどのように利用されているのでしょうか。データは日々の業務と長期的な戦略のどちらでより大きな役割を果たしているでしょう。また、組織が置かれた状況はこれらの質問に対する回答にどの程度影響するのでしょうか。

これらの疑問に答えようと、Chartbeat、Gawker Media、The New York Times という三つの事例研究を通じ、現代のニュースにおけるメトリクスの役割に関するエスノグラフィー調査を行いました。プロダクトマネージャー、データサイエンティスト、レポーター、ブロガー、編集者などへのインタビューや観察から、視聴者測定の根底にある仮定や価値観、測定基準がジャーナリストの日常業務に及ぼす影響、測定基準と組織文化との相互作用を明らかにします。以下では、私が発見した主要な事柄をいくつかまとめていきます。まず、分析ダッシュボードには、見過ごされがちな重要な感情的側面があります。指標や指標がその一部であるより大きな「ビッグデータ」は、一般に合理的なものとして説明されます。つまり、信頼できない直感や判断ではなく、冷静で客観的な情報に基づいて意思決定を行うことができるとされます。この表現は間違っていないものの、不完全です。測定基準のパワーと魅力は、データが興奮、失望、検証、安心感といった特定の感情を引き出す能力によるものです。

Chartbeat は、この感情の価値がダッシュボードの強力な魅力の一部であることを理解しており、ユーザーの感情を喚起するような機能を盛り込みました。たとえば、ダッシュボードは、ジャーナリストの判断に敬意を表し、トラフィックが少ないときの打撃を和らげ、ニュースルームが喜びに沸く機会を提供するように設計されています。

第二に、分析ツールの影響は、それを使用している組織によって異なります。多くの場合、分析ツールがニュースルームの運営を変化させると考えられがちです。しかし、私は、測定基準がニュース制作に影響を与えるかどうか、またどのように影響を与えるかを形成する上で、組織のコンテキストが非常に大きな影響力を持つことを発見しました。例えば、Gawker Media とニューヨーク・タイムズはどちらも Chartbeat のクライアントですが、それぞれの環境で、このツールの活用方法は大きく異なります。Gawker では、指標が非常に目に付くところに表示され、影響力を持っていました。タイムズではそうではなく、編集者がすでに下した決定の裏付けとして使われがちだったのです。アナリティクスがジャーナリズムにどのような影響を与えているかは、特定のニュースルームでアナリティクスがどのように使われているかを調べなければわかりません。

最後に、指標駆動型の文化は、記者のストレスと安心感の源になることがあります。また、この文化は、編集の自由という認識と驚くほど相性が良いのです。Gawker Media の記者は、トラフィックのプレッシャーにストレスを感じていましたが、多くはコメントやソーシャルメディアにおけるオンラインの罵詈雑言にはるかに大きな心理的影響を受けていました。ネット上で敵意や嫌がらせを浴びる中、記者はプロとしての力量を再認識するためにメトリクスを利用することもありました。興味深いことに、記者と編集者は、一般に、会社のトラフィックベースの評価システムが編集の自律性を阻害するものとは認識していませんでした。Gawker Media のようなオンラインのみのメディア企業のジャーナリストは、従来のメディア企業とは編集の自由や制約について異なる考えを持っているようです。

結論として、私は報道機関に次のように提言します。まず、報道機関はアナリティクス関連の問題（すなわち、組織における指標の適切な役割と、データが組織のジャーナリズムの目標と相互作用する方法）について、戦略的思考を優先させるべきです。ほとんどのジャーナリストは日々の業務に忙しく、組織における測定基準の役割や、どの測定基準がジャーナリズムの目標を最もよく補完するかについて、広範囲に、あるいは抽象的に考えることができませんでした。その結果、彼らはその場しのぎの方法で測定基準を参照し、解釈し、利用する傾向がありました。しかし、このようなデータは、その場しのぎで導入するにはあまりにも強力です。ニュースルームは、社内であれ、外部の研究者との提携であれ、アナリティクスの最適な利用方法について、日々の制作のプレッシャーから解放され、熟考する機会を設ける必要があります。

第二に、ニュースルームのマネージャーは、分析サービスを選ぶ際に、ツールだけでなく、どのベンダーの戦略的目標、ビジネスの必須事項、価値観が自社のニュースルームの目標に最も適しているかを検討すべきです。私たちは、数字、ひいてはアナリティクスのダッシュボードを、経験則に基づいた権威ある冷静な世界の

反映と見なす傾向があります。しかし、分析サービスを選択する際には、分析企業も商売しているのだということに注意が必要です。

第三に、測定基準の使用に関する社内方針を策定する際、ニュースルームの管理者は、トラフィックデータが編集内容だけでなく、編集部員にも影響を与える可能性があることを考慮してください。ニュースルームの壁やウェブサイトでランキングが目立つようになると、その影響力は歯止めがきかなくなりがちです。トラフィックベースのランキングは、たとえそれが意図的でない場合でも、他の評価形態をかき消す可能性があります。

最後に、より良い測定基準を開発する努力は必要であり、価値があるものの、ニュースルームと分析会社は測定基準の限界を忘れずに。組織の優先順位や評価システムがますます測定基準に基づいて構築されるようになると、定量的に測定可能なものを価値のあるものと混同してしまいがちです。すべてのものをカウントできるわけではなく、またカウントすべきでもありません。ニュースルーム、分析会社、資金提供者、メディア研究者は、ジャーナリズムの最も魅力的で不可欠な特性、たとえば社会的使命などが、いかに簡単に測定できないものであるか考えてみましょう。データ分析の価値が高まっている現在、定量化できるものと価値があるものを同一視しないように注意しなければなりません。

脚注

1. この作品は、「The Traffic Factories：Metrics at Chartbeat, Gawker Media, and The New York Times」（コロンビア大学ジャーナリズム大学院の Tow Center for Digital Journalism が 2015 年に原論文を発表）から抜粋・翻案した。許可を得て再掲載。

参考文献

Artley, M. (2013, August 26). Let me explain why Miley Cyrus' VMA performance was our top story this morning. *The Onion*. www.theonion.com/let-me-explain-why-miley-cyrus-vma-performance-was-our-1819584893

データジャーナリズム、デジタル普遍主義と周縁部のイノベーション

文：アニタ・セイ・チャン

概要

「デジタル普遍主義の神話」は、公共の物語や想像力をいわゆるイノベーションの「中心」に固定化する手段であるばかりではなく、中心を超えたデジタルな力学への関心をも削いでいます。記者やグローバルデジタル文化の研究者たちは、データや周縁に対する植民地主義的力学の轍を踏んではいけません。

キーワード：デジタル普遍主義、中心、周縁、植民地関係、脱植民地コンピューティング、ローカルイノベーション

「デジタル普遍主義」とは、デジタルをめぐるグローバルな想像力を形成する広範な、しかし誤った枠組みのことで、イノベーションの「中心」が推進する単一の普遍的な物語が、今日世界中で進行中のデジタル開発の形態を正確に表現できると見なしています。つまり、現代の「イノベーション」と技術デザインの中心が、大多数の「残りのデジタル住民」のいる世界のデジタルの未来を決めているというのです。そしてそれは、技術的変容、デジタル生産性とプラクティス、あるいは情報ベースの革新と探求を研究し観察するための最良で最も「正当な」場所は、そうした中心地であるという前提を通じて共鳴しているのです。その最たるものが、シリコンバレーにあるラボやオフィス、研究施設、そして世界中のイノベーションの中心地に散在する、エリート層のデジタル専門知識の集積地なのです。

デジタル文化はこのような中心地から生まれ、[そこでは]最も純粋な形と姿を持ち、他の場所ではそれが再現されるといいます。また、最も正確な、あるいは理想的なデジタルの将来像のビジョンが生まれるのもこの場所であり、ここでは、技術の進歩、すなわちデジタル文化の進歩は、最もダイナミックで生き生きとしていてインスピレーションに満ちていると考えられています。言い換えれば、デジタル文化はそのユニークでグローバルな次元にもかかわらず、その研究を行い、そのダイナミクスを観察するための、より「本物」で生産的な場所が実在すると仮定しているのです。

2000 年代初頭以降、私にはこんな無邪気な質問が投げかけられました。それはペルーとラテンアメリカにおけるデジタル文化の行動主義と政策の若手研究者にとってのその後に大きな影響を与えるものでした。「なぜ、デジタル文化を学ぶためにペルーやラテンアメリカに行くの？デジタル文化を研究するのに、もっと良い場所があるでしょう。シリコンバレーのような場所で活動を観察し、記録した方がいいのでは？」

こうした質問をする人は、ペルーは、かつてインカ文明の中心地であった南米の山岳地帯だとイメージしています。マチュピチュ、アンデス山脈の高地、ケチュア語やアイマラ語を話す人々が多く住む、かつてインカ文明の中心地であった南米の山国を思い浮かべることでしょう。過去の伝統や先住民の文化、あるいは豊かな自

然を覗き見るには理想的な場所かもしれませんが、現代のデジタル文化や高度なテクノロジーの流れ、あるいはそれらに関連する未来志向の開発について教えてくれるようなものはほとんどないと考えられていました。ペルーのような場所には、私たちが忘れてはならない過去の技術の遺物や宝物が散らばっていて、シリコンバレーのような場所では、私たちがまだ見ぬ未来の技術の秘密がきっと明らかになりつつあるのです。

背景に潜んでいるのは、エリートを輩出する中央で、選ばれた技術者たちが想像するデジタルの未来は世界の残りの部分を代弁することができ、現在イノベーションの中心地で展開されているのは、周辺部の未来でなければならないという無邪気な確信です。このように、「デジタル普遍主義の神話」の力は、公共の物語と想像力を、確立された革新の中心地だけに固定させる手段のみならず、そうした中心地を超えたデジタルな力学への関心を削ぐ手段としても現れています。そこでは、さまざまな場所で起きている実際のデジタルの力学をめぐる物語の多様性とグローバルな流通を狭め、多様な形態のデジタル生成物を見えなくし、「イノベーション」資本をデジタル生産性の独占地として人工的に増幅・強化するのです。ここには、グローバルなデジタル文化の記者や研究者が再生産に注意しなければならない、周縁部に対する特殊な植民地的観念があります。

それは、「周縁部」がグローバルな模倣の単なる担い手であり、未来がただ伝播していく領域だということです。実際、周縁部はそれほど受動的でもなければ、非発明的でもありません。中古でリサイクルされ、再組み立てされたコンピュータや部品でいっぱいの活気に満ちたダイナミックな屋外市場やインターネットカフェは、低コストのインターネットアクセスを拡大し、グローバルかつローカルなメディアコンテンツの流通を、農村部と都市部の多様な人々に普及させた、南半球の革新的な技術です。このようなテクノロジーハックや現地での工夫は、周縁部のテクノロジーランドスケープの日常的な一部であり、その活気は、デジタル製品やコンピュータ、インターネットサプライヤーの公式な商業チェーンと比較することによってのみ捉えることができるのです。社会科学者のダニエル・ミラーとドン・スレーター（2001）がトリニダード[・トバゴ]の研究において述べたように、「インターネットは一枚岩のサイバースペースではなく」、さまざまなローカルな現実、慣習、文化政治を取り巻くグローバルな広がりを持つテクノロジーとして存在しているのです。デジタルな実践と接続がどのようなものであるかを想像する方法は、実に多様です。ペルーでは、さまざまな異なるアクターや利害関係者が予期せず、そしてしばしば矛盾する形で接近し、活発なデジタル文化が形成されています。

2003 年、ラテンアメリカにおけるフリーソフトウェア利用に関する国連主催の初の会議―画期的なイベントで、インカの古都クスコで開催されました―の開催に協力したフリーソフトウェア支持団体が、オープンテクノロジーを再定義しようとしました。彼らはそれを、米国における自由でオープンソースなソフトウェア（FLOSS）の提唱のように、個人の自由と選択の問題だけではなく、文化の多様性、国家の透明性、そして南半球の多国籍企業の独占的権力に対する政治的主権として捉え直そうとしたのです。

州が農村部の学校に設置した「デジタル・イノベーション」教室は、わずか数年後には、話題を呼んだ MIT の「One Laptop per Child（OLPC）」イニシアチブの最大の展開拠点ネットワークに変貌を遂げ、すべ

てはデジタルインクルージョンの実現の名のもとに行われることになりました。また、国家のプログラムによって「伝統的な」商品に新たに積極的に適用された知的財産権（IP）タイトルは、成長する情報社会に基づく国の取り組みの一環として、地方の生産者や職人を輸出可能な新しいクラスの「情報労働者」へと作り替えるものでした。

クスコの FLOSS 支持者とハイテク活動家、国が推進する農村部の学校での「イノベーション教室」、新しいグローバルな「情報労働者」としての伝統的な職人などは、デジタル文化の中心地で紡がれる物語の従来の関心や主人公とは異なります。彼らの物語の詳細は、デジタル文化の既存の枠組みや支配的な物語では捉えきれません。新世紀における IT をめぐって、グローバルな想像力は、シリコンバレーのハッカー、ハイテクエンジニアの執着と願望、競争力のあるテクノロジー起業家の戦略的事業などをハリウッド映画の題材にし、Twitter のアカウントをしつこくチェックするくらいしか思いつかなかったのです。

しかし、ペルーでのデジタル文化のダイナミックな取り組みと実験を捉えるには、他にも多くの利害関係者、エージェント、開発者に注意を払うべきです。デジタルを巡る（彼らの）取り組みは、地方と都市、ハイテクと伝統、そして地域への強いコミットメントを持ったグローバルな方向性の空間の間に、新しいつながりと交換を構築しようとしました。今日、データジャーナリストは、世界中のさまざまなローカルな場所でデジタル文化や活動を目撃し、捉え、記憶するデジタルツールや技術的資源には事欠きません。2011 年初頭にアラブの春が起こる以前から、ネットワーク化されたデジタルメディアは、メキシコ（Schulz, 2007）、イラン（Burns & Eltham, 2009; Grossman, 2009）、フィリピン（Uy-Tioco, 2003; Vicente, 2003）やウクライナ（Morozov, 2010）などさまざまな状況でソーシャルメディアを戦略的に活用した運動を捉えようと、新しいグローバル報道力を拡張してきたのです。2011年のアラブの春をきっかけに、スペインの 15 人の（反緊縮運動）インディグナドスから北米のオキュパイ［運動］に至るまで、ハッシュタグによる組織化やソーシャルメディア上での運動が戦略的に活用されました。最近では、米国で始まった#MeToo や#BlackLivesMatter などの運動が、ラテンアメリカの#NiUnaMenos、ナイジェリアの#BringBackOurGirls、オーストラリアの#Sosblakaustralia、カナダの先住民の#IdleNoMore 動員、香港の#UmbrellaRevolution といった世界規模の運動とともに展開しています。

このような運動が押し広げるユーザー生成メディアの潮流は、市民によるデータ活用を拡大し、市場志向のプロファイリングに偏ったソーシャルメディアのプラットフォームにおける「ビッグデータ」の支配的な活用に歯止めをかけます。彼らはデータアーカイブを各地に散らばるドキュメンタリー作家、レポーター、オーガナイザーに貸し出す一方で、既存のメディアやニュース制作の中心地にはない新しい形の物語能力のためにデータを活用します。しかし、プラットフォーム上の社会運動によるオンラインの「データ」アーカイブから、創造的なデータ行動主義の並行形態まで、デジタル・リソースとデータ・リポジトリの増大は、データジャーナリストに新たなリスクも生み出しています。

その最たるものが、ビッグデータとソーシャルメディアのプラットフォームが収集した豊富なデータと情報を追跡して現在の人間活動と社会経験を記録し、その政治的・経済的影響の将来を評価・予測する最善の方法を生み出すかもしれないとする魅惑的な能力によるリスクです。デジタルの中心地で「現在」培われているデジタルの形態は、周辺地域の未来を正確に表現し、またそうなるものというデジタル普遍主義者の時間的前提は、アルゴリズムによるデータ処理の予測能力を自負するデータ産業が新たに裏付けるものになりました。

2016 年の米国大統領選挙から Brexit、Cambridge Analytica のスキャンダル、欧米各地でのオルトライト運動の「サプライズ」な台頭まで、欧米の主流の政治データアナリスト、ソーシャルメディア企業、ニュース専門家が近年の主要なグローバル政治混乱を正確に予測できなかったことは明らかですが、こうした言説は根強く残っています。今日のデータジャーナリストは、データの追跡やアーカイブを、それがどれほど広範で印象的な規模であろうとも、大衆がそれらに依存している多様な社会の現実を記録し、代弁し、評価する唯一あるいは支配的な形態となることを警戒すべきです。ラテンアメリカやポストコロニアルの学者たちから、「脱植民地コンピューティング」の枠組みのもと、情報、テクノロジー、ニューメディア文化を代表するものや人物を拡大するための調査・記録方法を求める声が高まるのと並行して（chan、2018）、デジタル普遍主義の枠組みに批判的なデータジャーナリストたちも、データソースを意識的に多様化し、「ビッグデータ」を経験則上の出来事や社会的現実をマッピングする唯一の、あるいは最も正当な鍵として特権化することは慎むべきです。「知識の脱植民地化」に向けた動きは、南半球の市民や研究者がボトムアップのデータ実践に取り組んでいる多様な方法の重要性を示しています1。これらの実践は、社会変革のためにデータを評価し解釈するコミュニティの実践と人間中心の手段を重視し、抑圧、不平等、社会的被害を拡大するビッグデータの使用に対する抵抗を巡る議論に役立ちます。デジタル普遍主義の台頭に批判的なデータジャーナリストは、クリティカルデータ研究、アルゴリズム研究、ソフトウェアとプラットフォーム研究、ポストコロニアルコンピューティングの研究者たちとの間で、説明責任と責任あるデータ実務を展開しうる協力者と共通する関心事を見出すよう心掛けるべきでしょう。

これには、データ原理主義（Boyd&Crawford, 2012 年）や、アプリケーションのアルゴリズムに関する主流の説明をいまだに取り囲んでいる技術的決定論に対する拒絶の強化が含まれます。また、データ化された世界やデータ産業における人間の根本的な再定義も必要であり、それはビッグデータや「アルゴリズムを対象として読み取る衝動に負けず、技術を説明の主役にのさばらせることにしっかり抵抗する」（クロフォード、2016）ものです。空港のセキュリティ、クレジットスコアリング、病院や患者の追跡、世界の多様なサイトでのソーシャルメディアなど、データポリティクスの新しい経験的、日常的な設定を探求するだけでなく、データを生産的で創造的なやり方で分析する方法論を開発することも必要です（Gillespie、2013、Ziewitz、2016 年）。

最後に、デジタル・イノベーションの中心地を、デジタル発明の独占的な場、あるいはデータの未来を作る場とし続けることは、当然ながら、もう一つの重要な詳細を無視することにもなります。現在の中心地もかつては周縁部にあったということは覚えておいて損はありません。中心がモデルを発明し、それが単に他の場所で採用され、コピーされるようになるという説は、複製機能や力が完全かつ継続的に拡張されることが前提です。より

大きなシステム内部での変化の可能性、つまり、それまでの中心が不安定化し再編成されること、そして同様に、それまでの周辺が再編成されることを説明できないからです。2011 年のアラブの春の「驚き」とその影響は、20 年も前に「発展途上国」とされた国々で非西洋のデジタル市場と経済競争が台頭し、強力だった西洋民主主義が今では不安定化したように、西洋と非西洋の世界の様々な場所に及んでおり、既存の権力の安定と「中心-周縁関係」の永続性に疑問を投げかけています。周縁部からのダイナミックな活動は、単に中心部に遅れをとったり模倣したりするのではなく、かつてマイナーな地位にあったエージェントが、新たな生産性の源泉として出現しうることを示唆しています。(そうした人々が織りなす)その多様な糸は、単一の普遍的な物語が、今日の世界各地に出現する多彩なデジタルの未来と想像力を適切に表すことができるという暗黙の前提を揺るがしています。

脚注

1. 詳しくはククタイとウォルターの章を参照。

参考文献

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662- 679. doi.org/10.1080/1369118X.2012.678878

Burns, A., & Eltham, B. (2009). Twitter free Iran: An evaluation of Twitter's role in public diplomacy and information operations in Iran's 2009 election crisis. In P. Franco & M. Armstrong (Eds.), *Record of the Communications Policy & Research Forum 2009* (pp. 322–334). Network Insight Institute.

Chan, A. (2018). Decolonial computing and networking beyond digital universalism. *Catalyst*, *4*(2). doi.org/10.28968/cftt.v4i2.29844

Crawford, K. (2016). Can an algorithm be agonistic? Ten scenes from life in cal- culated publics. *Science, Technology,*
& Human Values, *41*(1), 77–92. doi.org/10.1177/0162243915589635

Gillespie, T. (2013). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–193). MIT Press.

Grossman, L. (2009, June 17). Iran's protests: Why Twitter is the medium of the movement. *Time*. content.time.com/time/world/article/0,8599,1905125,00.html

Miller, D., & Slater, D. (2001). *The Internet: An ethnographic approach*. Berg Publishers.

Morozov, E. (2010). *The net delusion: The dark side of Internet freedom*. Public Affairs.

Schulz, M. (2007). The role of the Internet in transnational mobilization: A case studyof the Zapatista movement,

1994–2005. In M. Herkenrath (Ed.), *Civil society: Local and regional responses to global challenges* (pp. 129–156). Transaction Publishers.

Uy-Tioco, C. (2003, October 11). The cell phone and EDSA 2: The role of a communication technology in ousting a president. Critical Themes in Media Studies Conference.

Vicente, R. (2003). The cell phone and the crowd: Messianic politics in the contemporary Philippines. *Public Culture*, *24*(47), 3–36. doi.org/10.1080/01154451.2003.9754246

Ziewitz, M. (2016). Governing algorithms: Myth, mess, and methods. *Science,Technology, & Human Values*, *41*(1), 3–

16. doi.org/10.1177/0162243915608948

ジャーナリズムのデータ化：データ駆動型ストーリーテリングと産学連携戦略

文：ダミアン・ラドクリフ、セス・C・ルイス

概要

ジャーナリズムとアカデミアは、それぞれの職業のデータ化にどのように対応しているのでしょうか。また、データ駆動型の仕事において、どのようにすればより効果的な協力ができるのでしょう。

キーワード：ジャーナリズム、アカデミア、コラボレーション、データ化、データワーク、研究者とジャーナリストの協働

私たちは、データに動かされ、情報を得る世界に生きています。データは、政策や政治的決定にますます影響を与え（Höchtl ら、2016；Kreiss、2016）、私たちが住む都市の設計や機能に情報を与え（Glaeser ら、2018）、また、デジタル時代にアクセスし消費するニュース、製品、情報の種類を形成します（Diacopoulos、2019；Lewis、2017；Lewis & Westlund、2015；Thurman ら、2019；Usher、2016；Zamith、2018）。善かれ悪しかれ、データのとてつもない力と可能性は、まだ姿を現し始めたばかりです（Couldry & Mejias, 2019; O'Neil, 2016; Schroeder, 2018）。政府、大学、報道機関は、古くからデータや統計を利用して、世の中のパターンを発見し、説明してきました。しかし、デジタル機器の増加やクリック、「いいね！」、共有、位置情報、連絡先など、それらが生み出す膨大な痕跡データによって生成されるデータの量と、そうしたデータを大規模に利用し分析するコンピューティングパワーの増大は驚異的なものです。

多くの場合、こうした全てのデータの意味を解釈することは、間違いなく最大の課題で、その過程で倫理的な判断を迫られることも少なくありません（Crawford et al.、2014）。これは、政策立案者、企業、研究者、活動家、ジャーナリストなどの議論の的であり、「ビッグデータ・ソリューション」や、今日もてはやされている人工知能が簡単に解決してくれるものではありません（Broussard, 2018; Broussard et al, 2019）。ここでは、データとジャーナリズムに関する実践（ラドクリフ）と研究（ルイス）のそれぞれの観察を踏まえ、ジャーナリズムとアカデミアの世界が、それぞれの職業のデータ化、さらには公共生活の幅広いデータ化にどのように対応しているのかについて概説します。最終的な目的は、歴史的にやや不安定な関係（Carlson & Lewis, 2019; Reese, 1999）を共有してきたこれらの 2 分野が、データ中心の課題に対して、より生産的に連携する方法を提案することです。詩人ジョン・ダンは、「人は島嶼にあらず」と書きました。データ駆動型の世界では、どの職業も島嶼であってはいけません。

ジャーナリズムとデータ駆動型ストーリーテリング：五つの戦略的考察

データを使ってストーリーを伝え、私たちの周りの世界を理解することは、まったく新しいことではありません1。ビクトリア朝のイギリスでは、医師のジョン・スノウがロンドン中心部のコレラ患者をプロットした地図を作成しました。そして、1855 年に大流行したコレラの原因として、ブロードストリートの井戸を特定したのです（fugure44.1 と fugure44.2 参照）。スノーの説得力に満ちた分析は、現代のデータ分析・可視化ツールで作成された疾病マップとあまり変わりません。別の例では、フローレンス・ナイチンゲールの「東洋における軍隊の死因」がクリミア戦争で戦う英国兵の死因として衛生（ないしその欠如）を挙げました。（”Worth a Thousand Words,” 2013）彼女のデザインは今でも驚くほど現代的です。こうした努力と並行して、同じ頃、19 世紀半ばにホレス・グリーリーは『ニューヨーク・トリビューン』に寄稿し、多くの選出議員（若きエイブラハム・リンカーンを含む）が、認められているより多く経費を請求していたことを明らかにしました（クライン、2015 年）。世界は進歩しても（グリーリーは馬で移動する距離を基準としました）、この種の重要な調査報道はなおジャーナリズムの定番です (Larcinese & Sircar, 2017; see also Barrett, 2009; “A Chronology of the Senate Expenses Scandal,” 2016; “Expenses Scandal an Embarrassing Start,” 2017; “MPs’ Expenses: Full List,” 2009; “Q&A: MP Expenses,” 2009; Rayner, 2009; “Senate Expenses Scandal,” n.d.)。

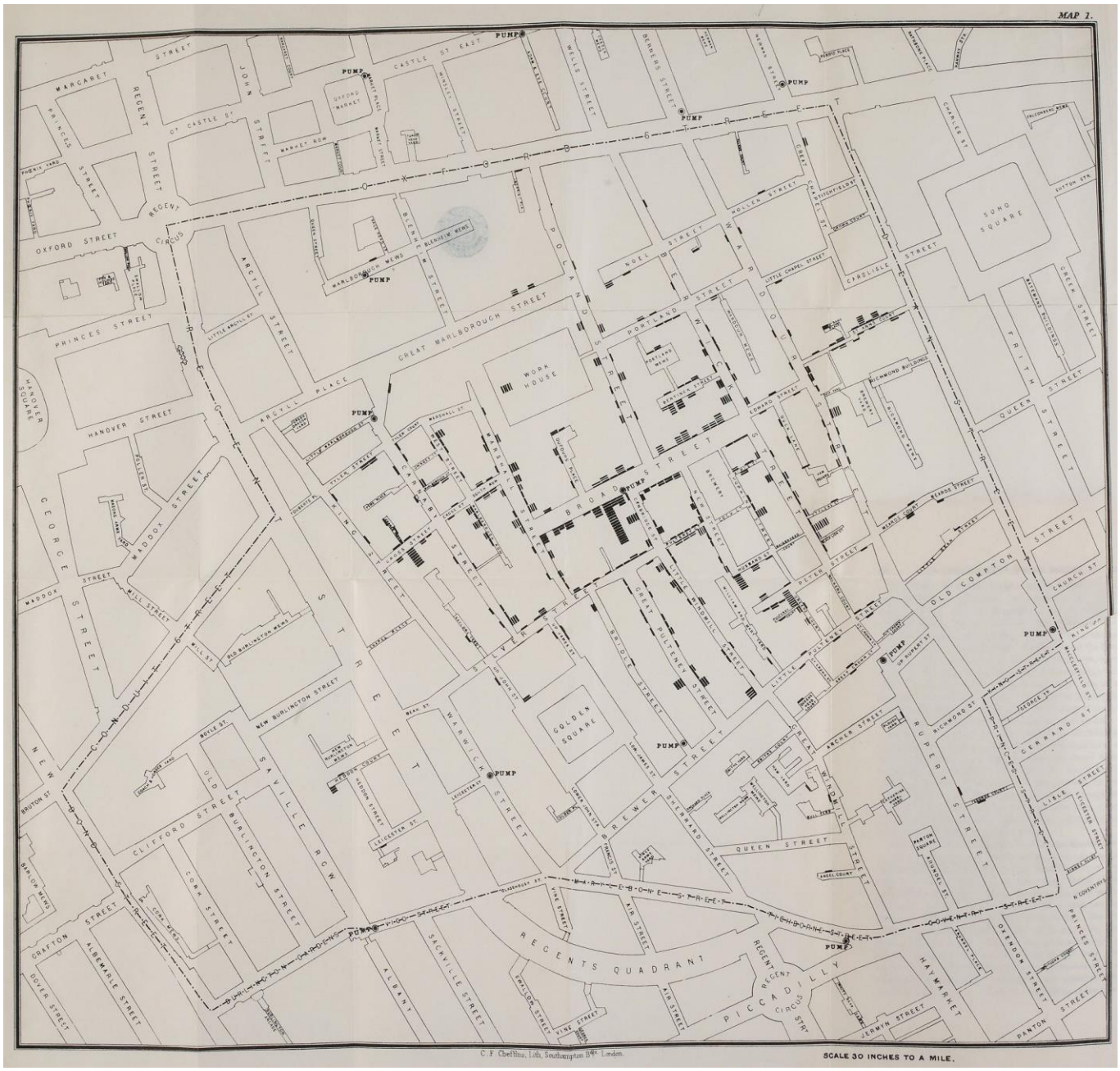
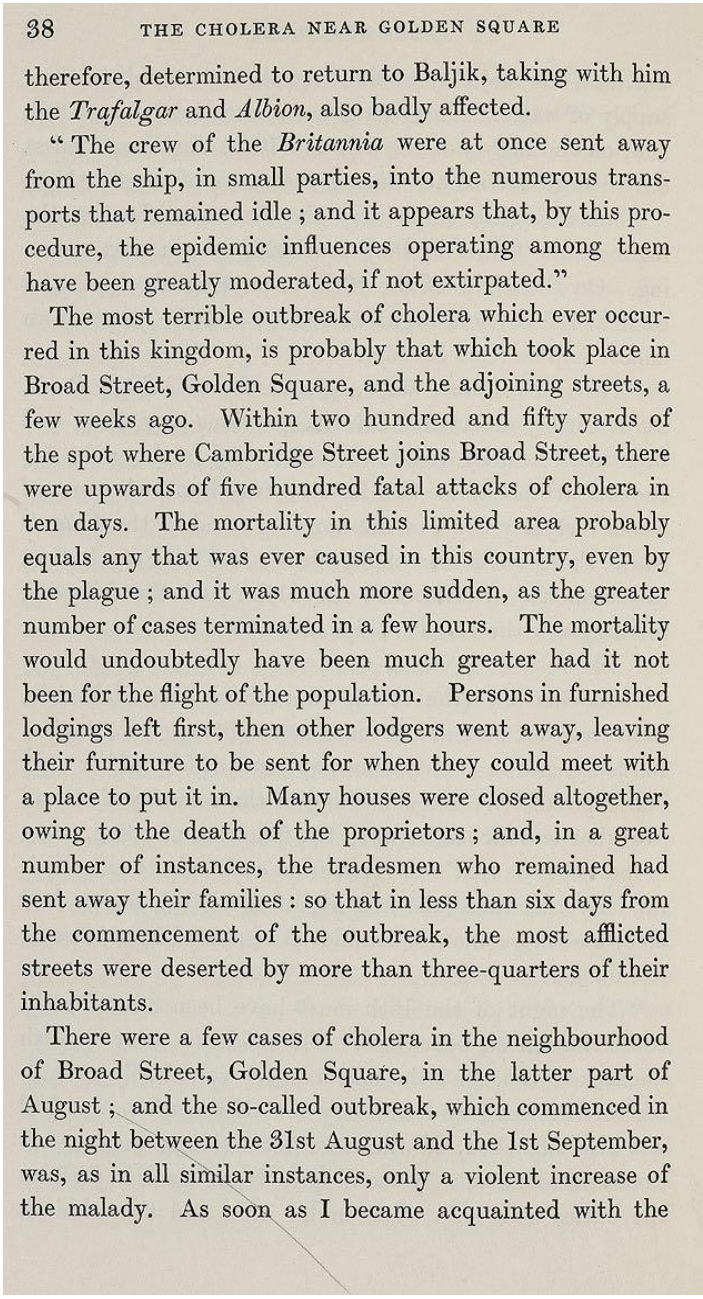


Figure 44.1. 医師のジョン・スノウが作成した、1855 年のロンドン中心部におけるコレラ患者をプロットした地図　出典：大英図書館 https://www.bl.uk/collection-items/ john-snows-map-showing-the-spread-of-cholera-in-soho-london-1855.

こうした歴史的な事例と、より現代的な事例（データジャーナリズム・アワード受賞作品など）は、ジャーナリストにとって強力なインスピレーションの源として機能します2。これらは、データ駆動型アプローチが、権威に責任を負わせ（ICIJ、n.d）、重要な社会の不正を強調し（Lowenstein,2015）、並外れたものを可視化して紹介するのに使用できる可能性を示します（“2016 Year in Review,” 2016）。20 世紀後半の「コンピュータ支援報道」の出現に反映されるように、データは長い間ジャーナリズムの一部でしたが、データ駆動技術の利用可能性

とアクセス性の近年の発展は、明確にデータ駆動型ジャーナリズムの可能性を広げました（歴史については、Anderson、2018 を参照、データジャーナリズムの概要については、Gray et al, 2012 を参照）。

このような背景から、世界中の報道機関、特にリソースの豊富な報道機関はもちろん、小規模な報道機関も、データを使ったジャーナリズムに貢献できるようになりました。データを使ってストーリーを伝える（Hermida & Young, 2019）だけでなく、データ（デジタル視聴者測定基準）を使ってストーリーの選択に影響を与えたり、仕事の影響を測定して改善したり（Zamith, 2018）しているのです。



38 THE CHOLERA NEAR GOLDEN SQUARE

therefore, determined to return to Baljik, taking with him the *Trafalgar* and *Albion*, also badly affected.

"The crew of the *Britannia* were at once sent away from the ship, in small parties, into the numerous transports that remained idle ; and it appears that, by this procedure, the epidemic influences operating among them have been greatly moderated, if not extirpated."

The most terrible outbreak of cholera which ever occurred in this kingdom, is probably that which took place in Broad Street, Golden Square, and the adjoining streets, a few weeks ago. Within two hundred and fifty yards of the spot where Cambridge Street joins Broad Street, there were upwards of five hundred fatal attacks of cholera in ten days. The mortality in this limited area probably equals any that was ever caused in this country, even by the plague ; and it was much more sudden, as the greater number of cases terminated in a few hours. The mortality would undoubtedly have been much greater had it not been for the flight of the population. Persons in furnished lodgings left first, then other lodgers went away, leaving their furniture to be sent for when they could meet with a place to put it in. Many houses were closed altogether, owing to the death of the proprietors ; and, in a great number of instances, the tradesmen who remained had sent away their families : so that in less than six days from the commencement of the outbreak, the most afflicted streets were deserted by more than three-quarters of their inhabitants.

There were a few cases of cholera in the neighbourhood of Broad Street, Golden Square, in the latter part of August ; and the so-called outbreak, which commenced in the night between the 31st August and the 1st September, was, as in all similar instances, only a violent increase of the malady. As soon as I became acquainted with the

Figure 44.2. ロンドン中心部のコレラ患者を記録した 1855 年の新聞記事　出典：大英図書館.

https://www.bl.uk/learning/images/makeanimpact/publichealth/large12734.html

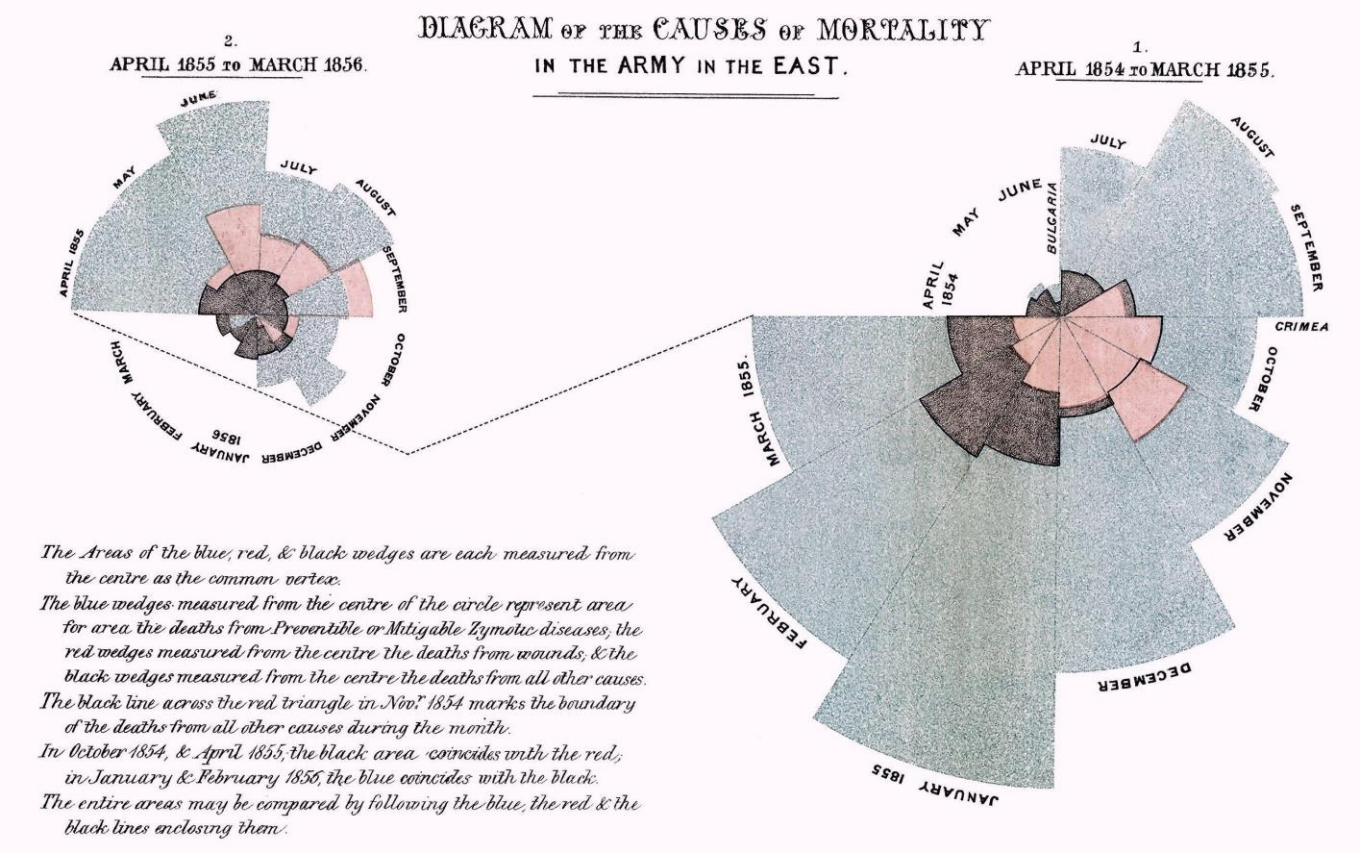


Figure 44.3. フローレンス・ナイチンゲール作「東洋の軍隊における死亡原因の図」出典：Wikimedia. https://upload.wikimedia.org/wikipedia/commons/1/17/nightingale-mortality.jpg

ここでは、データをもっと活用したいと考えるニュースルームやジャーナリストのために、五つの重要なメッセージを紹介します。データだけではストーリーは語れません。やはりジャーナリストが必要です。データにアクセスできても、他の情報源と同じように文脈を踏まえ、データを調査して理解するジャーナリストが必要です。モントクレア州立大学のスティーブ・ジョンソン（2013）が指摘しているように「読者は生のデータには興味がありません。読者はデータの中にあるストーリーが欲しいのです」。初期のオープンデータポータル、EveryBlock が提供したロウアーマンハッタンのデータについて、彼は次のようにコメントしています。

そこには、市が毎月どのような落書きを消したか、地区別に報告されていました。しかし、欠けているのは文脈と写真です。私が落書きに関する記事を書くなら、落書きの前と後の写真を見せたいし、さらに重要なのは、市が落書きアーティストとうまく戦っているのか、つまり誰が勝っているのかを知りたいのです。生データではそれがわからないのです（Johnson, 2013）。

パラダイス文書やパナマ文書など、より最近の「データダンプ」もこの点を強調しています。情報源の相互参照と文脈の特定が必要であり、この作業に多くのジャーナリストが数カ月を費やしました。しかし、このようなジャーナリスト（市民ではなく）による情報源の精査がなければ、データの持つインパクトを十分に発揮すること

はできませんでした。これらの原則は、国内および国際的な重要性を持つストーリーでも、地域レベルでも通用します（Radcliffe, 2013）。

データそのものがストーリーになることはめったにありません。データの背後にある意味を完全に理解するためには、データを解きほぐし、その意味を説明する必要があるのです。一人でやる必要はありません。多くの場合、協働が鍵になります。協働はネットワーク時代のキーワードであり、ジャーナリズムとテクノロジーの感性を融合させる上で重要な要素です。「ハッカー」（ジャーナリスト）と「ハッカー」（コーダー）の融合（Lewis & Usher, 2014, 2016）や、共通の関心事をめぐる報道機関とそのコミュニティの幅広い相互作用が含まれます。こうした「ネットワーク・ジャーナリズム」（Beckett, 2010; Van der Haak et al., 2012）や「リレーショナル・ジャーナリズム」（Boczkowski & Lewis, 2018; Lewis, 2019）の本質は、協働を通じてより多くを達成しうるという信念が根底にあることです。

このアプローチは、大量のデータを含む多くの専門分野やストーリーに適用できます。ガーディアンが 2009 年に行った英国議員の経費分析で示したように、関心を持った市民がジャーナリストと連携してデータセットを分析し、ヒントを提供することができます（「MPs' Expenses: The Guardian,” 2009; Rogers, 2009a, 2009b）。より最近では、スタンフォード大学のジェームズ・T・ハミルトン（2016）らによる研究（Sambrook, 2018）で、高品質でインパクトのある調査報道を行うためには、多くのニュースルームにおいて、組織を超えた協働と分野を超えた展開の両方が重要であることが明らかにされています。多くの新しい組織が抱えるデータ（の膨大な）量は、厳しい経済情勢と相まって、多くのニュースルームにとって、パートナーシップ、専門家、ボランティア、幅広いスキルセットの利用がしばしば必要であることを意味します。そして、経済的な観点からもジャーナリスティックな観点からも、協力的なアプローチはますます不可欠になっています。

データの見せ方は重要です。ジャーナリストは、データを表示し、ストーリーを伝えるためのさまざまなツール、テクニック、プラットフォームを利用できます。そのため、伝えようとするストーリーに最も適したツールを判断することが重要です。データ可視化、グラフ、インフォグラフィックス、チャート、表、インタラクティ ブなど、どれもストーリーを伝え、説得するのに役立つものばかりです。どのツールを使うかによって、大きな違いが生まれます 4。同僚のニコル・ダーメンが指摘するように、データ・ビジュアライゼーションは、そのための一つの方法です。「ビジュアルは読者の注意を引く。[そして]…テキストよりも 60 倍速く処理される」（Frank et al., 2015 から引用）。うまく使えば、テキストだけではできない方法で、ストーリーを生き生きとしたものにすることができるのです。ワシントン・ポストのオンライン特集「The Depth of the Problem」は、行方不明になったマレーシア航空 370 便のブラックボックスが海のどの程度の深さにあるのかを示す好例です（「The Depth of the Problem,” n.d.; figure26.5 を参照）。読者はページをスクロールして、エンパイアステートビルの高さである 1250 フィート、ダイオウイカの（生息する）深さ 2600 フィートを過ぎ、タイタニック号がある 1 万 2500 フィート以下、そしてブラックボックスがあるとされる 1 万 5000 フィートまで到達します。ダーメンは、「機体

がどれだけ深く沈んでいるかを読むだけにとどまりません」と述べています。「あなたは、それほどまでにというほどの深さを、見て、関わって、実体験し、理解することができます」。どのようなアプローチを取るかは、伝えたいストーリー、視聴者のデータリテラシーや嗜好の両方に影響されるかもしれません。いずれにせよ、データ駆動型ストーリーは、オーディエンスが何を見ているのか、どのようにデータを操作すればいいのかわかりやすくなるよう、うまく設計する必要があります（Radcliffe, 2017b, 2017c）。

自分の仕事をより広い文脈の中に位置づけるべきです。データを扱うジャーナリストは、上記の配慮とともに、コンテンツの消費やジャーナリズムに対する姿勢など、より広い範囲の動きも認識する必要があります。

モバイルに配慮を。2012 年、ピュー研究所は、スマートフォンやタブレットを所有している人の半数以上がそれらのデバイスでニュースコンテンツにアクセスしていることを明らかにしました（Mitchell et al., 2012; ”News Consumption on Digital Devices,” 2017）。4 年後、アメリカ人の 10 人に 7 人以上（72%）がモバイル経由でニュースを入手したと報告しています（Mitchell et al., 2016）。モバイルニュースの消費が拡大し続ける中、報道機関も、データ豊富な資料を含むすべてのコンテンツに対して、ポジティブなモバイル体験を提供することが不可欠となっています。

パーソナライズする。パーソナライズとアルゴリズムによるメディア体験の時代には、読者がデータを吟味し、それが自分にとって何を意味するかを理解する機会を作ることが必要です。プロパブリカの Dollars for Docs 調査（Tigas et al.、2019）は、製薬会社や医療機器会社が医師に支払った金額を患者が確認できるようにしたもので、この手法の実践例の一つです。

情報源の保護。ジャーナリストは、データを分析する方法だけでなく、データを保護する方法も知っておく必要があります。自分自身と情報源を保護するためには、機密データや内部告発者を扱うための新しいスキルセットを含む、新しいアプローチが必要になるかもしれません（Keeble-Gagnere, 2015）。（パナマ文書に見られるように）匿名性と結びついた暗号化は、その方法の一つです。

新しい技術を活用する。ブロックチェーンは、データや調査業務を保護・支援しうるツールです（IJNET, 2016）。セーデルトーン大学（スウェーデン）のワリード・アル・サカフ（Bouchart, 2018 に引用されている）が説明しているように「ブロックチェーンはデータを永久に保存し、あらゆる操作や不正を防止します。つまり、政府のデータは、一度公開されたら削除したり変更したりすることはできないのです」。機械学習もこの分野ですでに利用されている技術であり、今後ますます発展していくでしょう（Bradshaw, 2017）。ブロックチェーンとジャーナリズムに関するワリード・アル・サカフの長編記事もご覧ください。

国民の信頼を回復する。ジャーナリズムへの信頼が記録的な低水準にある今、これを改善するために努力することは、すべてのジャーナリストの責務です（Knight Foundation, 2018; Nicolau & Giles, 2017）。データを扱う人にとって、これは、扱っているデータについて透明性を保ち、元の情報源へのリンクを提供し、元のデータファイルがダウンロードできるようにすることを意味します。自分の仕事を見せること-ジェイ・ロ

ーゼン（2017）が「ジャーナリズムにおける信頼のための新しい用語」と呼ぶもの-は、読者にあなたが扱った（そして解釈した）元データを見せることを可能にし、それによってニュースにおける透明性に基づく信頼への扉を開きます。データが仕事に与える影響は、あなたが思っている以上に広く、またいっそう広がるでしょう。

最後に、ジャーナリズムの行為を形成する上でもデータが果たす役割を見過ごすことはできません。ジャーナリズムのデータ化は、データ・ストーリーテリングにとどまらず、ジャーナリズムの職業全般に影響を与えていることを忘れてはなりません（Anderson, 2018; Usher, 2016）。Google Analytics、Chartbeat、Omniture などの分析ツールはニュースルームに普及しており、ジャーナリストは視聴者の閲覧行動について以前より多くの情報を得ることができるようになりました。こうした定量的、定性的な洞察（American Press Institute が開発した Metrics for News などのプログラムを参照）は、大小問わずニュースルームの業務に情報を提供しています。Parse.ly 5が発行したホワイトペーパーや最近の学術研究（Belair-Gagnon & Holton, 2018; Cherubini & Nielsen, 2016; Ponsford, 2016; Radcliffe et al., 2017; Zamith, 2018）で取り上げられているように、データの役割は、記事の配置（文字通りホームページでの配置やソーシャルメディアでの宣伝方法など）やどの記事が取り上げられるかの意思決定の両方で極めて重要な役割を果たしていることは明らかです。シアトルを拠点とする記者・編集者で、シアトル・ポスト・インテリジェンサーの元シニア・エディター、レヴィ・プルキネンは、このデータの多くが、ニュースルームがこれまでとは異なることをする必要があることを示唆していると主張しています。プルキネン（as cited in Radcliffe, 2017a）は「新聞業界は、測定基準が教えてくれることは、内容を変える必要があるということを認めるのにためらいがあるようです」と言い、（他の分野より）パブリック・アフェアーズを巡る報道で変化の機が熟しつつあるとしています。こうも述べています「しかし、私たちが政府の速記者のような役割を果たす必要はないのです」（as cited in Radcliffe, 2017a）。

展望：産学連携のための五つのアイデア

データは、事実上すべてのニュースルームと報道対象の分野を介してジャーナリズム行為を形成し、情報を提供しています。データは、ガーディアンなどの定評ある企業や、FiveThirtyEight、Quartz などの新興企業に見られるように、特定のストーリーを伝えるためのツールであると同時に、編集やリソース駆動型の意思決定のための重要な情報源となり得るのです。しかし、個別のストーリーや戦略を超えて、データはジャーナリズムの大きな変革を予感させます。良くも悪くも、定量化の受け入れは、ビッグデータとジャーナリズムの四つの E と呼ばれるものに大きな影響を与える可能性があります。認識論（ジャーナリズムが知っていること）、専門性（ジャーナリズムがその知識をどのように表現するか）、経済性（ジャーナリズムの市場価値）、倫理性（ジャーナリズムの社会的価値）です（Lewis & Westlund, 2015）。

したがって、データ関連の影響は、私たちがジャーナリズムをどのように教え、実践し、研究するかに至る広範囲に及んでいるのです。私たちは、学界と報道業界が、両者の協力関係を深めることで生まれる可能性を認識し

ていないことがあまりにも多いと考えています（上記のジャーナリズムの協働についての指摘と同様です）。データ化がもたらす可能性を開拓する中で、ジャーナリストと研究者の関係をより緊密にすることは、互いに有益でしょう。以下では、そのための五つの足掛かりを紹介します。

教室とニュースルームのパートナーシップを強化する。ポール・ブラッドショーの研究は、その方法をわかりやすく示しています。英国のバーミンガム・シティ大学で提供されているデータジャーナリズムの修士課程の一環として、ポールと彼の学生は、デイリー・テレグラフ（Bradshaw, 2018）、BBC、ITN、マンチェスター・イブニングニュース、ガーディアン、調査報道センターなど、多くの報道機関と提携しています。この教育ベースのパートナーシップを研究改善につなげるため、これらの報道機関は、民族誌学者による（参加者の）観察にデータジャーナリズムのプロセスを開放できます。そうした研究は、査読済みの学術出版物だけでなく、Tow Center for Digital Journalism や Reuters Institute for the Study of Journalism が作成したような、業界向けの一般報告書にも貢献することが期待されます。

ニュースバリューのあるプロジェクトに参加する。ロンドン西部のキングストン大学でジャーナリズム講師を務めるジャン・グッディは、英国の地方政府における潜在的な利益相反を特定し、講義のプロジェクトを具体的な報道へと発展させる能力を実証しています。彼らの調査は、99 件の情報公開請求、追跡、分析を含むもので、これらの政府が、フラッキング（シェールガス採掘の水圧破砕法を用いる）企業に年金資金を投資しながら、この新興産業が提出する計画案の仲裁役としても機能していることが明らかになりました（Goody & Milmo, 2014）。場合によっては、学生や教授がデータプロジェクトに取り組む時間的余裕が生まれ、絶え間ない日々の締め切りに追われるジャーナリストには難しいデータジャーナリズムを実践できるようになるかもしれません。

これらの関係をリバースエンジニアリングしてみる。ほとんどのニュースルームが直面しているリソースの問題を考慮すると、ジャーナリストは、学生や学者に記事作成の手助けをもっと頻繁にアプローチすることができます。アリゾナ州立大学のスティーブ・ドイグは、マイアミ・ヘラルドで、建築基準法の弱体化と建設［業界の］劣悪な慣行が、前年度のハリケーン・アンドリューの被害を悪化させたことを示したシリーズで7、1993 年に公共サービス部門のピューリッツアー賞を受賞しており8、コンピュータ支援報道の問題についても積極的にコンサルティングを実施しています。また、調査機関 California Watch のために病院の疑わしい請求方法を分析した「Decoding Prime」でジョージ・ポーク賞（2012 年）を受賞しています9。教授陣と学生を巻き込んだアドバイスとコンサルティングのモデルは、他の研究者も見習うべきものです。

研究者や独立した批評に門戸を開く。ジャーナリストは、研究者を記事の情報源として頼りがちですが、研究者の調査に自分たちを開放しようとしたがりません。この問題をさらに深刻にしているのは、上層部の許可なしにアンケートを取ったり、研究者と話をしたりしてはいけないという組織的な指示がますます厳しくなっていることです。しかし、ジャーナリストが仕事をするために良い資料を必要とするように、学者がジャーナリズムについて良い研究をするためには、彼らが現在よりも良いアクセスを持っていることが必要です。これは、研究者が

ジャーナリズムにとってデータ化が何を意味するか(Baack, 2015)、つまり、ジャーナリストがどのように指標を使うか(Belair-Gagnon & Holton, 2018; Christin, 2018; Ferrer-Conill & Tandoc, 2018)、新しい方法でストーリーテリングする方法(Hermida & Young, 2019; Toff, 2019; Usher, 2016)などを、理解しようとしているときには特にそうだといえます。報道機関側のハードルが少しでも下がれば、互恵的な関係を築くのに大いに役立つに違いありません。(そうすれば、)研究者は、データがジャーナリズムにどのように適合するかを理解するより良いアクセスを得ることができます。その一方で報道機関は、自分たちの仕事に対する独立した評価を得ることができ、その結果、自分たちの仕事ぶりを批判的立場からより良く評価することができるのです。

研究がアクセス可能であることを保証する。一方、学術関係者は自分の研究の公開性とアクセス性を確保するために、もっと多くのことができるはずです。これまでにも「ジャーナリズムのデータ化」に関する数多くの学術研究が行われており、特にデータ・ストーリーテリングのためのツールの進化と、ジャーナリズムの倫理やアプローチへの影響に重点が置かれています(概要については、Ausserhofer et al.、2020 年を参照)。これらの研究は、報道機関にとって非常に大きな意義を持ち得ます。しかし、これらの研究は、学術雑誌の有料ページに収められ、専門用語が多用されているために不明瞭で、ジャーナリストにとっては、この研究から得られる教訓を理解することはおろか、アクセスすることも困難な状況に置かれている場合があまりにも多いのです。例えば Nieman Journalism Lab などの業界誌や The Conversation などの一般読者に学術論文を広めるための Web サイト向けに概要や梗概を作ることで、これらの知見が学術誌の枠を超えて反響を呼ぶようにできます。

結論

データジャーナリズムは、有名なデータジャーナリストで現在は Google のデータエディターを務めるサイモン・ロジャースの言葉(2012 年)を借りれば、「偉大な平等主義者」です。その創発的な性質から、誰でもそれを試して熟達することができます。彼は「データジャーナリズムは新しいパンクだ」と言っています。つまり、「多くのメディアグループが、寝室でハッキングしている人と同程度の予備知識と専門知識で始めている」(Rogers, 2012)のです10。もちろん、データジャーナリズムには長い歴史があり、その前身は 1 世紀以上前から存在する科学とストーリーテリングの形態です(Anderson, 2018)。しかし、ジャーナリズムの中にある生まれたての「社会的世界」(Lewis & Zamith, 2017)として、データジャーナリズムはツール、テクニック、ベストプラクティスを報道機関間や世界中で共有する場として、21 世紀の社会の幅広いデータ化の中での転換点に立っているのです。批判的な自己反省の機会があることも補足したいと思います。データジャーナリズムについてこれまでにわれわれが知っていることを検証するため、今後調査すべきことを概説するため、特にジャーナリズムの相互依存を実践と教育、産業と学術として認識する道を追求するためにも。データにあふれた世界をジャーナリズムが理解するためには、データ駆動型アプローチの限界と可能性を吟味することが必要です。それは、何を明らかにするか、何をしないか、どうすれば改善できるかということです。

脚注

1. データジャーナリズムの様々な系譜については、本書のアンダーソンの章を参照されたい。
2. dev.datajournalismawards.org. 本書のルースンの章にある賞についての議論も参照を勧める。
3. www.journaliststoolbox.org/2018/03/11/online_journalism/
4. www.import.io/post/8-f... を参照。異なる手法を用いた八つのストーリーを紹介し、その入れ替えを検討している。
5. www.parse.ly/resources...
6. bcu.ac.uk/media/courses...
7. www.pulitzer.org/winner...
8. www. ickr.com/photos/...
9. cronkite.asu.edu/about...
10. サイモン・ロジャースの章を参照。
11. アンダーソンの章を参照。

参考文献

2016 year in review: Highlights and heartbreaks. (2016, December 29). *Los Angeles Times*. www.latimes.com/local/california/la-me-updates-best-year-review-2016-htmlstory.html

Anderson, C. W. (2018). *Apostles of certainty: Data journalism and the politics of doubt*. Oxford University Press.

Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (2020). The data cation of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journal- ism*, *21*,(7) 950–973. doi.org/10.1177/1464884917700667

Baack, S. (2015). Data cation and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, *2*(2), 2053951715594634. doi.org/10.1177/2053951715594634

Barrett, D. (2009, May 17). MPs' expenses: How they milk the system. *The Telegraph*. www.telegraph.co.uk/news/newstopics/mps-expenses/5294350/Expenses-How-MPs-expenses-became-a-hottopic.html

Beckett, C. (2010). *The value of networked journalism.* POLIS, London School of Economics and Political Science.

Belair-Gagnon, V., & Holton, A. E. (2018). Boundary work, interloper media, and analytics in newsrooms. *Digital Journalism*, *6*(4), 492–508. doi.org/10.1080/21670811.2018.1445001

Boczkowski, P. J., & Lewis, S. C. (2018). The center of the universe no more: From the self-centered stance of the past to the relational mindset of the future. In P. J. Boczkowski & Z. Papacharissi (Eds.), *Trump and the media* (pp. 177– 185). MIT Press. doi.org/10.7551/mitpress/11464.003.0028

Bouchart, M. (2018, February 1). A data journalist's new year's resolutions. *Medium*. medium.com/data-journalismawards/a-data-journalists-new-year-s- resolutions-474ef92f7e8f

Bradshaw, P. (2017, December 14). Data journalism's AI opportunity: The 3 dif- ferent types of machine learning & how they have already been used. *Online Journalism Blog*. onlinejournalismblog.com/2017/12/14/data-journalismsai-opportunity-the-3-different-types-of-machine-learning-how-they-have-already-been-used/

Bradshaw, P. (2018, February 15). Wanted: MA Data Journalism applicants to partner with *The Telegraph*. Medium. paulbradshaw.medium.com/wanted-ma-data-journalism-applicants-to-partner-with-the-telegraph-8abd154260f3

Broussard, M. (2018). *Arti cial unintelligence: How computers misunderstand the world*. MIT Press.

Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C.-H. (2019). Arti cial intelligence and journalism. *Journalism & Mass Communication Quarterly*, *96*(3), 673–695. doi.org/10.1177/1077699019859901

Carlson, M., & Lewis, S. C. (2019). Temporal re exivity in journalism studies: Making sense of change in a more timely fashion. *Journalism*, *20*(5), 642–650. doi.org/10.1177/1464884918760675

Cherubini, F., & Nielsen, R. K. (2016). *Editorial analytics: How news media are developing and using audience data and metrics* (SSRN Scholarly Paper ID 2739328). Social Science Research Network. doi.org/10.2139/ssrn.2739328 Christin, A. (2018). Counting clicks: Quantif ication and variation in web journalism in the United States and France.

American Journal of Sociology, *123*(5), 1382–1415. doi.org/10.1086/696137

A chronology of the Senate expenses scandal. (2016, July 13). *CBC News*.www.cbc.ca/news/politics/senateexpense-scandal-timeline-1.3677457

Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject.

Television & New Media, *20*(4), 336–349. doi.org/10.1177/1527476418796632

Crawford, K., Gray, M. L., & Miltner, K. (2014). Critiquing big data: Politics, ethics, epistemology. *International Journal of Communication*, *8*, 1663–1672. ijoc.org/index.php/ijoc/article/view/2167

The depth of the problem. (n.d.). *The Washington Post*. www.washingtonpost.com/apps/g/page/world/the-depth-ofthe-problem/931/

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Ellis, J. (2015, June 23). Quartz maps a future for its interactive charts with Atlas. Nieman Lab. www.niemanlab.org/2015/06/quartz-maps-a-future-for-its-interactive-charts-with-atlas/

Expenses scandal an embarrassing start to 2017 for Australia's embattled Prime Minister. (2017, January 13). *The Indian Express*. indianexpress.com/article/world/expenses-scandal-an-embarrassing-start-to-2017-for-australiasembattled-pm-4466029/

Ferrer-Conill, R., & Tandoc, E. C. (2018). The audience-oriented editor. *Digital Journalism*, *6*(4), 436–453. doi.org/10.1080/21670811.2018.1440972

Frank, A., Yang, Y., & Radcliffe, D. (2015, December 18). The mainstreaming of data reporting and what it means for journalism schools. *Journalism.co.uk*. https://www.journalism.co.uk/n...

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Enquiry*, *56*(1), 114–137. www.hbs.edu/faculty/Pages/item.aspx? num=51012

Goodey, J., & Milmo, C. (2014, April 27). Exclusive: Local authorities have "con ict of interest" on fracking investments. *The Independent*. www.independent.co.uk/news/uk/politics/exclusive-local-authorities-have-con ictof-interest-on-fracking-investments-9294590.html

Gray, J., Chambers, L., & Bounegru, L. (Eds.). (2012). *The data journalism handbook: How journalists can use data to improve the news*. O'Reilly Media.

Hamilton, J. T. (2016). *Democracy's detectives: The economics of investigative journalism*. Harvard University Press.

Hermida, A., & Young, M. L. (2019). *Data journalism and the regeneration of news*. Routledge.

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 147–169. doi.org/10.1080/10919392.2015.1125187

ICIJ. (n.d.). The Panama Papers: Exposing the rogue offshore nance industry. www.icij.org/investigations/panamapapers/

IJNET. (2016, August 5). How blockchain technology can boost freedom of the press. International Journalists'

Network. ijnet.org/en/story/how-blockchain-technology-can-boost-freedom-press

Johnson, S. (2013, February 8). Sorry EveryBlock, you never learned how to write a headline. *The Hudson Eclectic*. https://hudsoneclectic.com/201...

Keeble-Gagnere, G. (2015, November 5). Encryption for the working journalist: Communicating securely.

Journalism.co.uk. https://www.journalism.co.uk/n...

Klein, S. (2015, March 17). Antebellum data journalism: Or, how big data busted Abe Lincoln. *ProPublica*. https://www.propublica.org/ner...

Knight Foundation. (2018, June 26). 10 reasons why Americans don't trust the media. *Medium*. https://medium.com/trust-media...

Kreiss, D. (2016). *Prototype politics: Technology-intensive campaigning and the data of democracy*. Oxford University Press.

Larcinese, V., & Sircar, I. (2017). Crime and punishment the British way: Account-ability channels following the MPs' expenses scandal. *European Journal of Political Economy*, *47*, 75–99. doi.org/10.1016/j.ejpoleco.2016.12.006

Lewis, S. C. (2017). Digital journalism and big data. In B. Franklin & S. Eldridge (Eds.), *The Routledge companion to digital journalism studies* (pp. 126–135). Routledge. Lewis, S. C. (2019). Lack of trust in the news media, institutional weakness, and relational journalism as a potential way forward. *Journalism*, *20*(1), 44–47. doi.org/10.1177/1464884918808134

Lewis, S. C., & Usher, N. (2014). Code, collaboration, and the future of journalism: A case study of the hacks/hackers global network. *Digital Journalism*, *2*(3), 383–393. doi.org/10.1080/21670811.2014.895504

Lewis, S. C., & Usher, N. (2016). Trading zones, boundary objects, and the pursuit of news innovation: A case study of journalists and programmers. *Convergence*, *22*(5), 543–560. doi.org/10.1177/1354856515623865

Lewis, S. C., & Westlund, O. (2015). Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, *3*(3), 447–466. doi.org/10.1080/21670811.2014.976418

Lewis, S. C., & Zamith, R. (2017). On the worlds of journalism. In P. J. Boczkowski & C. W. Anderson (Eds.), *Remaking the news: Essays on the future of journalism scholarship in the digital age* (pp. 111–128). MIT Press.

Lowenstein, J. K. (2015, February 13). How I used data-driven journalism to reveal racial disparities in U.S. nursing homes. *Storybench*. www.storybench.org/how..https://www.storybench.org/how-i-used-data-driven-journalism-toreveal-racial-disparities-in-u-s-nursing-homes/

Mitchell, A., Gottfried, J., Barthel, M., & Shearer, E. (2016, July 7). *The modern news consumer: News attitudes and practices in the digital era*. Pew Research Center. https://www.journalism.org/201...

Mitchell, A., Rosenstiel, T., & Christian, L. (2012). Mobile devices and news consumption: Some good signs for journalism. In Pew Research Center (Ed.), *The state of the news media: An annual report on American journalism*. www.pewresearch.org/wp-content/uploads/sites/8/2017/05/State-of-the-News-Media-Report-2012-FINAL.pdf

MPs' expenses: Full list of MPs investigated by *The Telegraph*. (2009, May 8). *The Telegraph*. www.telegraph.co.uk/politics/0/mps-expenses-full-list-mps-investigated-telegraph/

MPs' expenses: *The Guardian* launches major crowdsourcing experiment. (2009, June 23). *The Guardian*. www.theguardian.com/gnm-press-of ce/crowdsourcing-mps-expenses

News consumption on digital devices. (2017, August 21). Pew Research Center. www.journalism.org/2012/03/18/mobile-devices-and-news-consumption-some-good-signs-for-journalism/

Nicolau, A., & Giles, C. (2017, January 16). Public trust in media at all time low, research shows. *Financial Times*. www.ft.com/content/fa332f58-d9bf-11e6-944b-e7eb37a6aa8e

O' Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Ponsford, D. (2016, July 31). "Heartbroken" reporter Gareth Davies says *Croydon Advertiser* print edition now "thrown together collection of clickbait." *Press Gazette*. www.pressgazette.co.uk/heartbroken-reporter-garethdavies-says-croydon-advertser-print-edition-now-thrown-together-collection-of-clickbait/

Q&A: MP expenses row explained. (2009, June 18). *BBC News*. news.bbc.co.uk/2/hi/uk_...

Radcliffe, D. (2013). Hyperlocal media and data journalism. In J. Mair & R. L. Keeble(Eds.), *Data journalism: Mapping the future* (pp. 120–132). Abramis Academic Publishing.

Radcliffe, D. (2017a). Local journalism in the Paci c Northwest: Why it matters, how it' s evolving, and who pays for it (SSRN Scholarly Paper ID 3045516). Social Science Research Network. papers.ssrn.com/abstract=3045516

Radcliffe, D. (2017b). Data journalism in the US: Three case studies and ten general principles for journalists. In J.

Mair, R. L. Keeble, & M. Lucero (Eds.), *Data journalism: Past, present and future* (pp. 197–210). Abramis Academic Publishing.

Radcliffe, D. (2017c, November 28). 10 key principles for data-driven storytelling. *Journalism.co.uk*. www.journalism.co.uk/news/10-key-principles-for-data-driven-storytelling/s2/a713879/

Radcliffe, D., Ali, C., & Donald, R. (2017). *Life at small-market newspapers: Results from a survey of small-market newsrooms*. Tow Center for Digital Journalism. doi.org/10.7916/D8XP7BGC

Rayner, G. (2009, May 8). MPs' expenses: Ten ways MPs play the system to cash in on expenses and allowances. *The Telegraph*.www.telegraph.co.uk/news/newstopics/mps-expenses/5293498/MPs-expenses-Ten-ways-MPsplay-the-system-to-cash-in-on-expenses-and-allowances.html

Reese, S. D. (1999). The progressive potential of journalism education: Recasting the academic versus professional debate. *Harvard International Journal of Press/ Politics*, *4*(4), 70–94. doi.org/10.1177/1081180X9900400405...

Rogers, S. (2009a, June 18). How to crowdsource MPs' expenses. *The Guardian*. www.theguardian.com/news/datablog/2009/jun/18/mps-expenses-houseofcommons

Rogers, S. (2009b, June 19). MPs' expenses: What you've told us. So far. *The Guardian*. www.theguardian.com/news/datablog/2009/sep/18/mps-expenses-westminster-data-house-of-commons

Rogers, S. (2012, May 24). Anyone can do it. Data journalism is the new punk. *The Guardian*. www.theguardian.com/news/datablog/2012/may/24/data-journalism-punk

Rosen, J. (2017, December 31). Show your work: The new terms for trust in journal- ism. *PressThink*.
pressthink.org/2017/12/show-work-new-terms-trust-journalism/

Sambrook, R. (Ed.). (2018). *Global teamwork: The rise of collaboration in investigative journalism*. Reuters Institute for the Study of Journalism.

Schroeder, R. (2018). *Social theory after the Internet: Media, technology, and globalization*. UCL Press. Senate expenses scandal. (n.d.). *HuffPost Canada*. www.huf ngtonpost. ca/news/senate-expenses-scandal/
Seward, Z. M. (2016, May 10). Atlas is now an open platform for everyone's charts and data. *Quartz*. qz.com/679853/atlas-is-now-an-open-platform-for-everyones-charts-and-data/

Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, automation, and news. *Digital Journalism*, *7*(8), 980–992. doi.org/10.1080/21670811.2019.1685395 Tigas, M., Jones, R. G., Ornstein, C., & Groeger, L. (2019, October 17). Dollars for docs. *ProPublica*. projects.propublica.org/docdollars/

Toff, B. (2019). The "Nate Silver effect" on political journalism: Gatecrashers, gate-keepers, and changing newsroom practices around coverage of public opinion polls. *Journalism*, *20*(7), 873–889. doi.org/10.1177/1464884917731655

Usher, N. (2016). *Interactive journalism: Hackers, data, and code*. University of Illinois Press.

Van der Haak, B., Parks, M., & Castells, M. (2012). The future of journalism: Networked journalism. *International journal of communication*, *6*, 2923–2938. Worth a thousand words. (2013, October 7). *The Economist*. www.economist.com/christmas-specials/2013/10/07/worth-a-thousand-words

Zamith, R. (2018). Quanti ed audiences in news production: A synthesis and research agenda. *Digital Journalism*,

6(4), 418–435. doi.org/10.1080/21670811.2018.1444999

疎外された人々による、疎外された人々についての、疎外された人々のためのデータジャーナリズム

文： エヴァ・コンスタンタラス

概要

データジャーナリズムは、社会的に無力な立場に追いやられたコミュニティが不公正、不平等、差別と戦うのを後押しします。

キーワード：データジャーナリズム、疎外された人々、不公正、不平等、差別、エンパワーメント

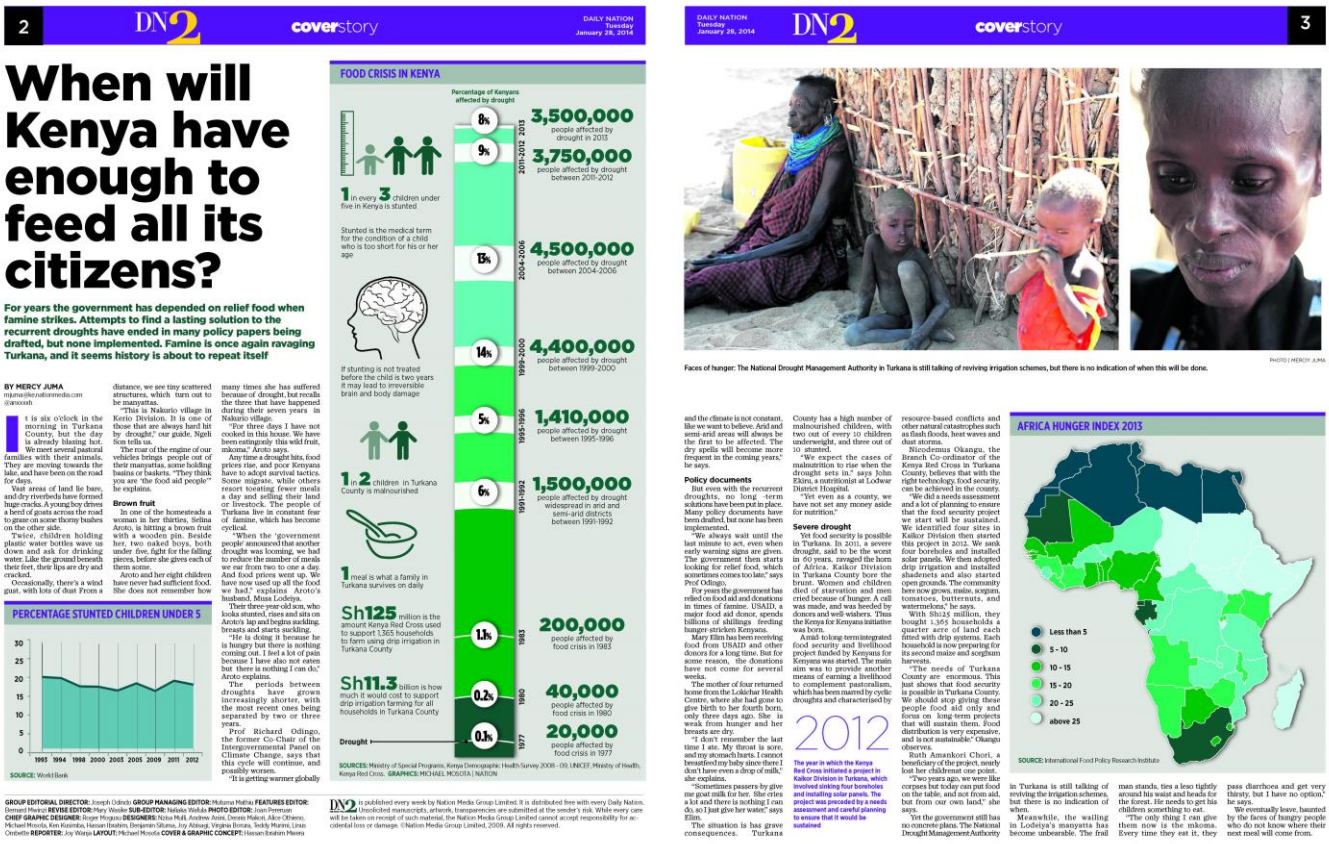
私は、破綻国家とされる国々でデータジャーナリズムを実践しています。そのような国では、汚職が政府を麻痺させ、市民は意気消沈し、市民社会は四面楚歌の状態にあると言われています。事態はひどく悪化しているのです。世界で最も貧しく、教育が行き届かず、治安の悪い地域でデータジャーナリズムを制作することで、私は重要な結論に達しました。不公平、不平等、差別は、ほとんどの国に偏在し、陰湿なまま見過ごされています。私が共に仕事をするジャーナリストたちは、新しいツールを駆使して、事態がどれほどひどいのか、その結果誰が苦しんでいるのか、誰のせいなのか、どうすれば事態が改善されるのかを明らかにしています。このような状況において、ジャーナリストは政策に影響を与え、市民を動かし、プロパガンダと戦う手段としてデータを受け入れています。報道の自由が制限されているにもかかわらず、データジャーナリズムはエンパワーメントの手段と見なされています。この章では、疎外された人々による、疎外された人々についての、社会的に無力な立場に追いやられた人々のためのデータジャーナリズムを探求します。不正、不平等、差別の様々な側面と、それらが疎外された人々の生活に及ぼすより大きな影響に注目することで、我々はそれらを可視化し、測定可能にし、解決できる道を開きます。これらのストーリーは、社会から疎外されたコミュニティに深く根ざしているジャーナリストを巻き込みます。制度的な差別に直面しているグループが懸念する問題に、市民の関与を促します。これらのストーリーは地元のマスメディアを通じて発信され、多くの人々に届けられ、政府が国全体のためによりよい決断を下すよう圧力をかけるのです。以下では、アフガニスタン、パキスタン、ケニア、キルギス、バルカン半島における、社会

から疎外されたコミュニティの関心と問題に取り組む 5 種類のデータジャーナリズムのストーリーを紹介します。

Figure 45.1. パキスタンのデータジャーナリストは、データから得た知見をもとに最初のワイヤーフレームを作成する　出典：Internews.

私たちの国にはすべての人を養えるだけの資源があるのに、なぜ人々は飢えているのか？

ケニアでは、ドナーはまさに間違った食糧計画に資金を提供していました。[ケニアのテレビ局]NTV のマーシー・ジュマが放映した、ケニア北部の孤立した貧困地域トゥルカナについての 12 分間のテレビ番組は、旱魃と飢饉がより頻繁かつ深刻になり、子どもたちの栄養失調が深刻化していることを明らかにしています。お金は緊急の食料援助に回され、長期的な干ばつ緩和には使われていません。議会に提出された政策案によると、1 年間の緊急食糧援助に使われるのと同じお金で、郡全体とその約 100 万人の住民のための食糧維持プログラムを賄うことができます。編集者がデータを編集しようとすると、ジュマは記事を撤回すると迫りました。彼女のストーリーは、主にテレビを通じて、ドナーを引きつけ、市民を激怒させ、政府を困らせるものでしたが、紙媒体の記事にもなり、オンラインでは要約が公開されていました（Juma, 2014）。



2 DN2 coverstory DAILY NATION Tuesday January 28, 2014

When will Kenya have enough to feed all its citizens?

For years the government has depended on relief food when famine strikes. Attempts to find a lasting solution to the recurrent droughts have ended in many policy papers being drafted, but none implemented. Famine is once again ravaging Turkana, and it seems history is about to repeat itself

BY MERCY JUMA
mjuma@ke.nationmedia.com
@amcooh

It is six o'clock in the morning in Turkana County, but the day is already blazing hot.

We meet several pastoral families with their animals. They are moving towards the lake, and have been on the road for days.

Vast areas of land lie bare, and dry riverbeds have formed huge cracks. A young boy drives a herd of goats across the road to graze on some thorny bushes on the other side.

Twice, children holding plastic water bottles wave us down and ask for drinking water. Like the ground beneath their feet, their lips are dry and cracked.

Occasionally, there's a wind gust, with lots of dust From a distance, we see tiny scattered structures, which turn out to be manyattas.

"This is Nakurio village in Kerio Division. It is one of those that are always hard hit by drought," our guide, Nigel Son tells us.

The roar of the engine of our vehicles brings people out of their manyattas, some holding basins or baskets. "They think you are 'the food aid people'" he explains.

Brown fruit

In one of the homesteads a woman in her thirties, Selina Aroto, is hitting a brown fruit with a wooden pin. Beside her, two naked boys, both under five, fight for the falling pieces, before she gives each of them some.

Aroto and her eight children have never had sufficient food. She does not remember how many times she has suffered because of drought, but recalls the three that have happened during their seven years in Nakurio village.

"For three days I have not cooked in this house. We have been eating only this wild fruit, mkoma," Aroto says.

Any time a drought hits, food prices rise, and poor Kenyans have to adopt survival tactics. Some migrate, while others resort to eating fewer meals a day and selling their land or livestock. The people of Turkana live in constant fear of famine, which has become cyclical.

"When the 'government people' announced that another drought was looming, we had to reduce the number of meals we ear from two to one a day. And food prices went up. We have now used up all the food we had," explains Aroto's husband, Musa Lodeiya.

Their three-year-old son, who looks stunted, rises and sits on Aroto's lap and begins suckling, breasts and starts suckling.

"He is doing it because he is hungry but there is nothing coming out. I feel a lot of pain because I have also not eaten but there is nothing I can do," Aroto explains.

The periods between droughts have grown increasingly shorter, with the most recent ones being separated by two or three years.

Prof Richard Odingo, the former Co-Chair of the Intergovernmental Panel on Climate Change, says that this cycle will continue, and possibly worsen.

"It is getting warmer globally and the climate is not constant, like we want to believe. Arid and semi-arid areas will always be the first to be affected. The dry spells will become more frequent in the coming years," he says.

Policy documents

But even with the recurrent droughts, no long -term solutions have been put in place. Many policy documents have been drafted, but none has been implemented.

"We always wait until the last minute to act, even when early warning signs are given. The government then starts looking for relief food, which sometimes comes too late," says Prof Odingo.

For years the government has relied on food aid and donations in times of famine. USAID, a major food aid donor, spends billions of shillings feeding hunger-stricken Kenyans.

Mary Elim has been receiving food from USAID and other donors for a long time. But for some reason, the donations have not come for several weeks.

The mother of four returned home from the Lokichar Health Centre, where she had gone to give birth to her fourth born, only three days ago. She is weak from hunger and her breasts are dry.

"I don't remember the last time I ate. My throat is sore, and my stomach hurts. I cannot breastfeed my baby since there I don't have even a drop of milk," she explains.

"Sometimes passers by give me goat milk for her. She cries a lot and there is nothing I can do, so I just give her water," says Elim.

The situation is has grave consequences. Turkana County has a high number of malnourished children, with two out of every 10 children underweight, and three out of 10 stunted.

"We expect the cases of malnutrition to rise when the drought sets in," says John Ekiru, a nutritionist at Lodwar District Hospital.

"Yet even as a county, we have not set any money aside for nutrition."

Severe drought

Yet food security is possible in Turkana. In 2011, a severe drought, said to be the worst in 60 years, ravaged the horn of Africa. Kaikor Division in Turkana County bore the brunt. Women and children died of starvation and men cried because of hunger. A call was made, and was heeded by donors and well-wishers. Thus the Kenya for Kenyans initiative was born.

A mid- to long-term integrated food security and livelihood project funded by Kenyans for Kenyans was started. The main aim was to provide another means of earning a livelihood to complement pastoralism, which has been marred by cyclic droughts and characterised by resource-based conflicts and other natural catastrophes such as flash floods, heat waves and dust storms.

Nicodemus Okanga, the Branch Co-ordinator of the Kenya Red Cross in Turkana County, believes that with the right technology, food security, can be achieved in the county.

"We did a needs assessment and a lot of planning to ensure that the food security project we start will be sustained. We identified four sites in Kaikor Division then started this project in 2012. We sank four boreholes and installed solar panels. We then adopted drip irrigation and installed shadenets and also started open grounds. The community here now grows, maize, sorgum, tomatoes, butternuts, and watermelons," he says.

With Sh125 million, they bought 1,365 households a quarter acre of land each fitted with drip systems. Each household is now preparing for its second maize and sorghum harvests.

"The needs of Turkana County are enormous. This just shows that food security is possible in Turkana County. We should stop giving these people food aid only and focus on long-term projects that will sustain them. Food distribution is very expensive, and is not sustainable," Okanga observes.

Ruth Amankori Chori, a beneficiary of the project, nearly lost her children at one point.

"Two years ago, we were like corpses but today can put food on the table, and not from aid, but from our own land," she says.

Yet the government still has no concrete plans. The National Drought Management Authority in Turkana is still talking of reviving the irrigation schemes, but there is no indication of when.

Meanwhile, the walling in Lodeiya's manyatta has become unbearable. The frail man stands, ties a leso tightly around his waist and heads for the forest. He needs to get his children something to eat.

"The only thing I can give them now is the mkoma. Every time they eat it, they pass diarrhoea and get very thirsty, but I have no option," he says.

We eventually leave, haunted by the faces of hungry people who do not know where their next meal will come from.

FOOD CRISIS IN KENYA

Percentage of Kenyans affected by drought

1 in every 3 children under five in Kenya is stunted

Stunted is the medical term for the condition of a child who is too short for his or her age

If stunting is not treated before the child is two years it may lead to irreversible brain and body damage

1 in 2 children in Turkana County is malnourished

1 meal is what a family in Turkana survives on daily

Sh125 million is the amount Kenya Red Cross used to support 1,365 households to farm using drip irrigation in Turkana County

Sh11.3 billion is how much it would cost to support drip irrigation farming for all households in Turkana County

Drought

2013	8%	3,500,000 people affected by drought in 2013
2011-2012	9%	3,750,000 people affected by drought between 2011-2012
2004-2006	13%	4,500,000 people affected by drought between 2004-2006
1999-2000	14%	4,400,000 people affected by drought between 1999-2000
1995-1996	5%	1,410,000 people affected by drought between 1995-1996
1991-1992	6%	1,500,000 people affected by drought widespread in arid and semi-arid districts between 1991-1992
1983	1.1%	200,000 people affected by food crisis in 1983
1980	0.2%	40,000 people affected by food crisis in 1980
1977	0.1%	20,000 people affected by food crisis in 1977

SOURCES: Ministry of Special Programs, Kenya Demographic Health Survey 2008 - 09, UNICEF, Ministry of Health, Kenya Red Cross. GRAPHICS: MICHAEL MOSOTA | NATION

PERCENTAGE STUNTED CHILDREN UNDER 5

30 25 20 15 10 5 0
1993 1994 1998 2000 2003 2005 2009 2011 2012

SOURCE: World Bank

GROUP EDITORIAL DIRECTOR: Joseph Odindo GROUP MANAGING EDITOR: Mutuma Mathiu FEATURES EDITOR: Bernard Mwangi REVISE EDITOR: Mary Wanjiku SUB-EDITOR: Nakisa Wafula PHOTO EDITOR: Joan Peersman CHIEF GRAPHIC DESIGNER: Roger Mogusu DESIGNERS: Nzioa Muli, Andrew Amin, Dennis Makori, Alice Otieno, Michael Musula, Ken Kazimba, Hassan Ibrahim, Benjamin Sikuma, Joy Atieno, Virginia Bomas, Teddy Murimi, Linus Ombetto REPORTER: Joy Wanja LAYOUT: Michael Mosofa COVER & GRAPHIC CONCEPT: Hassan Ibrahim Maiera

DN2 is published every week by Nation Media Group Limited. It is distributed free with every Daily Nation. Unsolicited manuscripts, artwork, transparencies are submitted at the sender's risk. While every care will be taken on receipt of such material, the Nation Media Group Limited cannot accept responsibility for accidental loss or damage. ©Nation Media Group Limited, 2009. All rights reserved.

DAILY NATION Tuesday January 28, 2014 DN2 coverstory 3

Faces of hunger: The National Drought Management Authority in Turkana is still talking of reviving irrigation schemes, but there is no indication of when this will be done.

PHOTO | MERCY JUMA

AFRICA HUNGER INDEX 2013

Less than 5
5 - 10
10 - 15
15 - 20
20 - 25
above 25

SOURCE: International Food Policy Research Institute

2012

The year in which the Kenya Red Cross initiated a project in Kaikor Division in Turkana, which involved sinking four boreholes and installing solar panels. The project was preceded by a needs assessment and careful planning to ensure that it would be sustained

Figure 45.2. マーシー・ジュマによるトゥルカナの食糧安全保障に関するテレビ特番を印刷したもの 出典：Internews.

彼女はデータの強みを生かしてドナーを説得したのです。彼女は、政府省庁、公衆衛生調査、ドナー機関、ケニア赤十字社から、気候、農業、健康データを入手しました。USAID ケニアミッションは、USAID の 1 年間の緊急食糧援助が、トゥルカナにおけるケニア赤十字の食糧持続性戦略全体を賄えることを示すデータ・ビジュアライゼーションを目にしました。彼女は、[援助の]遅れが子どもたちに及ぼす健康への影響と、砂漠で食糧を栽培している国々との対照的な状況を示しました。ナイロビの USAID 事務所に招かれて調査結果を発表し、2015 年には USAID の農業・食糧安全保障戦略が人道支援から持続可能な農業にシフトしたのです1。彼女はトゥルカナで飢餓に苦しむ家族の緻密な記録で世論を味方につけました。彼女は、トゥルカナの通訳とビデオグラファーとともに、この作品に登場する家族と 3 日間を過ごしました。

放映中、放送局の電話は鳴りっぱなしで、ケニアの人々は記事に登場した家族にお金を寄付しようとしました。個人、団体を問わず、この記事に対する反響が大きかったため、放送局は数時間のうちにトゥルカナ郡のための救援基金を設立しました。ケニア北部の絶望的な飢餓状況に関するこの記事とその後の記事は、これまでケニア北部の孤立した貧困地域の窮状に関心を示さなかった同国メディアで連日注目を集めました。彼女の主な視聴者は、寄付金をもっと賢く投資できることを示唆するデータではなく、力強く人間的なストーリーに心を

動かされたのです。政府は世論とドナーの圧力に屈しました。干ばつ監視委員会は、ジュマが調査を始めたとき、データへのアクセス料を請求しようとしたのに、状況がこれほど絶望的になっていることを知らなかったと主張し、記事のデータをシェアするよう求めました。ジュマの水不足データに基づき、水資源省はトゥルカナに井戸を増設する予定です。政府は計画・権限委譲省を通じて、トゥルカナでの救援物資配布のために 23 億シリング（2700 万ドル）を拠出し、ジュマはその動向を注視していました。

この記事に対する個人や団体からの大きな反響を受け、同年 5 月、新たな援助先となる食料持続可能性に関する法案が上院に提出されました2。ジュマはその後も、国民の認識とドナーのプログラム、政策の断絶に関するデータ駆動型特集を制作しており、国内の保守的な地域での 10 代の妊娠率に対する避妊具使用の影響に関する調査「#Teen Mums Of Kwale」（「#TeenMumsOfKwale」2016）なども行ってきました。

司法制度が疎外された人々を守っていることをどうやって確かめる？

アフガニスタンでは、Pajhwok Afghan News のデータチームがデータを使って、同国の先進的な取り組みとして称賛されている二つの政策の影響を探りました。それは「女性に対する暴力撤廃法」（2009 年）と「アフガニスタン国家麻薬統制戦略」（2012～16 年）です。しかし、これらの政策によって、二つの予期せぬ犠牲者が出ていることが判明しました。それは、虐待された女性と農村部の労働者です。アフガニスタンには情報公開法がないものの、女性課題省や麻薬対策省など、ドナーから資金提供を受けている多くの機関は、契約上そのデータを公開することが義務づけられています。DV 法が施行されて 5 年、Pajhwok Afghan News は加害者と被虐待者のその後を取材しました。取材班は、女性［課題］省と、登録から最終判決、調停に至るまで事件の記録を担う複数の国連機関から、2 万 1000 件の虐待事件に関するデータを入手しました。その結果、女性にとって世界最悪の国で、広く称賛されている法律が、伝統的な排外主義に凝り固まった地元の調停プロセスを女性に押し付け、通常、女性を加害者のもとに戻していることがわかりました（Munsef & Salehai, 2016）。 2 年後、ヒューマン・ライツ・ウォッチは、PAN の調査結果、すなわち、法律と調停がアフガニスタンの女性の期待を裏切っているとする研究を発表しました（United Nations Assistance Mission in Afghanistan, 2018）。仮に、より多くの女性が加害者の有罪率の高さを誇る裁判制度にアクセスできたとしても、女性が働く風潮のない社会で離婚した女性［の生計］をどうするかという茨の道が残されています。

アフガニスタンの麻薬戦略の実施にも、同様の現実的な課題があります。国連薬物・犯罪事務所は、薬物犯罪で有罪判決を受けた囚人へのアクセスを少しばかり認められ、調査の生データを Pajhwok チームに手渡しました。調査結果の分析からは、この政策によって、貧しく読み書きのできない運転手や農民が投獄され、麻薬王の多くが自由の身になっていることが明らかになりました（Barakzai & Wardak, 2016）。

また、ほとんどの人が、麻薬取引は孤立した地方で家族を支える唯一の方法なので、出所したらすぐに［麻薬栽培］労働に戻るつもりだと報告しています。これらのストーリーは、Pajhwok データチームの三大目的を果たすものでした。西洋の法的観点から開発された政策をチェックすること、経済的疎外がもたらす結果を

性別や場所ごとに明らかにすること、そしてデータ駆動型の公益コンテンツをダリー語、パシュトー語、英語で多様な読者に提供することです。

誰もが質の高い教育を受けられるようにするには？

教育へのアクセスは、しばしば偉大な格差是正装置とみなされます。疎外されたコミュニティは、基本的な公共サービスを提供しない政府の失敗を定量化し、地域のリーダーを改革に向けて後押しできるようになりました。開発者兼ジャーナリストのアブドゥルサラム・アフリディは、一連のストーリーの中で、恵まれない人々の教育へのアクセスに関するビートを構築し、そのポートフォリオでデータジャーナリズム・アワードの最終選考に残りました。

最初の記事では、政府の公式教育統計と全国的な教育調査データを使って、[パキスタンとの国境地帯にある]カイバル峠の辺境部族地域で、増加する一方の子供を必死に私立学校に通わせる親が、間違った投資をしていることを示しました。彼の記事は、公立校、私立校を問わず、卒業生のほとんどが標準テストに落第していることを示しました（Afridi, 2017a）。さらに、カイバル峠とサラム自身の出身地である連邦直轄部族地域の公教育に関するその後の記事では、落第校の背景を探っています（Afridi, 2017b, 2018）。

また、国家職業訓練プログラムの受講生名簿と全国の求人情報をもとに、（受講生が身につける）スキルと市場の需要との間に大きなギャップがあることがわかりました。この調査では、運転手や鉄鋼労働者が必要なのに、IT専門家や美容師を訓練していることが明らかになったのです。このように、卒業生の半数以上が失業したままになっているのは、このプロジェクトの背後にいる人物に責任があります。ドイツ政府の開発基金 GiZ から資金提供を受けたパキスタン政府は、独自の分析で同じ結論に達し、すぐにプログラムを見直して、より必要とされる職業スキルに沿った新しいコースを追加しました（Afridi, 2017c）。

限界集落におけるデータ駆動型の専門報道に固有の利点は、最初のスキャンダルが忘れ去られた後も、ジャーナリストがそのストーリーに留まることができる点です。また、これらのストーリーに共通しているのは、データを使って問題を報告するだけでなく、それに対して何ができるかも伝えている点です。こうしたジャーナリストたちは、問題、影響、原因、解決策を測定するためにデータを集めました。疎外された人々による、疎外された人々のためのアクセス可能なデータジャーナリズムを推進し、彼らの信頼を獲得し、市民生活に参加させることが世界的に求められているのです。

制約条件下でのデータジャーナリズム

データジャーナリズムの長期的な存続に関する学界の意見の分裂はたいがい、その目的が高収益のインタラクティブ商品を生み出すことにあるのか、それとも事実に基づく公益的な報道なのかを巡って生じます。途上国のジャーナリストは、制度化された男女差別、偏見に満ちた司法制度、飢餓に対する意図的な無視などをめぐる根本的な疑問に答えるためにデータを使用し、その情報をできるだけ多くの人に伝えています。彼らは、こ

れらの問題が複雑で、政策が変わる可能性が極めて低いことを承知の上で、このような取り組みを行っているのです。欧米のデータジャーナリストは、より良いリソース、データ、自由なメディアを利用でき、より反応の良い政府を持っていますが、このような激動の時代に、疎外された市民の情報ニーズに対応し、政府の責任を追及する機会を掴んでいないこともあります。

これらの問題のほとんどは、過去には見えなかったものであり、ジャーナリストが監視し続けなければ、再び見えなくなってしまうでしょう。データジャーナリズムは、社会が勘定に入れないと決めた人たちによる、その人たちについての、その人たちのためのものなのです。幸いなことに、市民社会、活動家、学者、政府などは、取り残された人々をよりよく見つめるために協力しています。ジャーナリストは、これらの問題を人々が話題にし、解決するために努力することを保証する重要な役割を担っているのです。私たちがそうした人々を取り上げ、話し続けない限り、過去も現在も、そしてこれからも、すべてが最悪なものとなるでしょう。飢えている人、虐待されている人、投獄されている人、教育を受けていない人、話を聞いてもらえない人など、毎年毎年、取り上げ続けなければならないのです。地球上のいかなる場所も、楽園ではないのですから。

脚注

1. www.usaid.gov/kenya/agriculture-and-food-security

2.kenyalaw.org/kl/ leadmin/pdfdownloads/bills/2014/TheFoodSecurityBill2014.pdf

参考文献

Afridi, A. S. (2017a, February 18). In KP, parents still prefer private over public schools. *News Lens Pakistan*. www.newslens.pk/in-kp-parents-still-prefer-private-over-public-schools/

Afridi, A. S. (2017b, June 16). Half of FATA schools functioning in dire straits. *News Lens Pakistan*. www.newslens.pk/half-fata-schools-functioning-dire-straits/

Afridi, A. S. (2017c, September 16). TVET Reform programmes targeting wrong skills. *News Lens Pakistan*. www.newslens.pk/tvet-reform-programmes-targeting-wrong-skills/

Afridi, A. S. (2018, March 2). Despite huge investment the outlook of Educationin KP remains questionable. *News*

Lens Pakistan.www.newslens.pk/despite-huge-investment-outlook-education-in-kp

Barakzai, N. A., & Wardak, A. (2016, September 28). Most jailed drug offenders are poor, illiterate. *Pajhwok Afghan News*. pajhwok.com/2016/09/28/most-jailed-drug-offenders-are-poor-illiterate/

Juma, M. (2014, January 28). When will Kenya have enough to feed all its citizens? *Daily Nation*. www.nation.co.ke/lifestyle/dn2/

Munsef, A. Q., & Salehai, Z. (2016, May 11). Cases of violence against women: Is mediation the best option? *Pajhwok Afghan News*. pajhwok.com/2016/05/11/cases-violence-against-women-mediation-best-option/

#TeenMumsOfKwale: Primary school girls in Kwale using contraceptives to prevent unwanted pregnancies. (2016, October 2). *NTV Kenya*. www.youtube.com/watch?v=xMx5lRHbw3g&lc=z12vjhuoysrivtgtl22guhgrgrjzsfezd04

United Nations Assistance Mission in Afghanistan. (2018). *Injustice and impunity: Mediation of criminal offences of violence against women*. United Nations Of ce of the High Commissioner for Human Rights. unama.unmissions.org/sites/default/ les/unama_ohchr_evaw_report_2018_injustice_and_impunity_29_may_2018.pdf

データジャーナリズムを教える

文：シェリル・フェリップス

概要

データジャーナリズム教育の第一歩は、吟味思考を教えることです。

テキサス州立大学では、シンディ・ロイヤル教授が Web 開発を教えています1。数千マイル東のフロリダ大学では、ジャーナリズム技術と民主的プロセスのナイト講座で、ミンディ・マクアダムスとノーマン・ルイス准教授が、コーディングから従来のデータジャーナリズムやアプリ開発まで多彩な授業に取り組んでいます。マイアミ大学コミュニケーション学部でビジュアル・ジャーナリズムのナイト・チェアーを務めるアルベルト・カイロは、データ可視化に焦点を当てた包括的プログラムを教えています。北へ行けば、コロンビア大学とニューヨーク市立大学の学生は、NBC やニューヨーク・タイムズの現役データジャーナリストによる授業を受け、データ分析とともに調査報道の基本を学びます。メリーランド大学では、メディア法の授業で、公文書請求を提出するプロセスを定期的に行っています。ネブラスカ（大学）では、マット・ウェイトがレゴを使ってデータを可視化する方法を教えています。スタンフォード大学では、基本的なデータ分析、Python と R によるコーディング、基本的なデータの可視化について、プレゼンテーションよりも理解を深めるために教えています。

データジャーナリズムの教授たちは、その多くが記者として［キャリアを］スタートし、世界中でさまざまな方法で教えています（上記の例は米国のプログラムだけです）。どのプログラムが真のデータジャーナリズムなのか、というのは難しい質問で、すべてそうなのです。では、どのように教えればよいのでしょうか？どのようなジャーナリズムの授業でも、同じように教えることができます。スポーツ・ジャーナリズムからビジネス、科学まで、どんな専門分野でもその分野特有のスキルや知識を習得する必要があります。

しかし、いずれもジャーナリズムの基礎の上に成り立っています。同様にデータジャーナリズムの教育も基礎から始めるべきでしょう。データジャーナリズムの基本的な考え方の多くを理解するのに、スプレッドシートは理想的だと思いますが、それ［のみ］を学ぶという意味ではありません。コンピュータ画面上の小さなマスに情報を入力する実習を行うことで、入力されたデータ固有の乱雑さを理解することは、この上ない収穫なのです。また、Python や R など、特定のコーディング方法を学ぶという意味でもありません。学生がコードを実行し、表計算ソフトでは 4 ステップ以上かかるような結果を得るのを見るのは、何にも代えがたい喜びともいえるでしょう。（しかし、何よりもまず）データジャーナリズムの学習は、情報を吟味する方法と、ジャーナリズムの目的のために情報を収集し、正規化し、分析する方法を理解することから始まります。それは、ストーリーを筋道立てて、結論に到達するための質問をすることから始まります。そして、ジャーナリズムの教育者は、これらの質問がどのような形で行われるかをすでに知っていることでしょう。そのデータは誰が作成したのか？そのデータは何を含んでい

るのか？データの最終更新日はいつか？そのデータは世界のどこを表しているのか？なぜこのデータを使ってストーリーを語る必要があるのか？このデータから、問いたいことの答えをどのように見つけるか？

だから、表計算ソフトや SQL、Python、R を使うカリキュラムを作ればいいのです。私がかつて DOS の[データベース管理システム]Paradox を知っていたことが重要でないのと同じように。重要なのは、データを収集し、分析する手順を知っていることです。分析するにも発表するにも可視化が重要ですが、まずデータを理解するための可視化ができれば、その後の発表はより簡単になります。ここでは、データジャーナリズムの教え方に関して、様々なアプローチと出発点を紹介します。それは、あなたが誰で、どの程度のプログラムを持っていて、どのように共同作業を構築できるかに基づいています。データジャーナリズムを教えるための「スーツケース」アプローチを紹介した後、ワンコースモデル、反転授業モデル、統合モデル、異なる分野間の共同指導について探求しています。

ワンコースでもやり放題：スーツケースに詰め込もう

私はカーキャンプに行くとき、冗談で「台所の流し台も積み込もう」と言ったりします。コツは、何を詰め込んでいいのか、何を詰め込むと生産性が悪くなるのかを知ることです。その「流し台」は、折りたたみ可能な布製の小さなボウルです。もしあなたがたった一つのクラスしか教えておらず、データジャーナリズムの教育者であるなら、スプレッドシートや構造化照会言語（SQL）によるデータ分析、Python によるデータ処理、R による分析、D3 によるデータ可視化デザインなどの全てを四半期または一つの学期に詰め込もうとしないでください。

必要なツールを選びましょう。クラスの少なくとも一部をプロジェクトベースにすることを検討してください。いずれにせよ、ステップを踏んでいくことです。シンプルにすること。選んだツールを使うことで生まれるジャーナリズムに集中しましょう。2014、15 年に、コロンビア大学のチャールズ・ベレットと私は、データジャーナリストとジャーナリズム教育者へのアンケートと広範囲なインタビューを行いました。データジャーナリズムを教える人のほとんどが、スプレッドシートから始めることで、構造化されたデータの概念を学生にわかりやすく紹介できると報告しています。もう一つのステップは、複雑さにも対応してデータジャーナリズムに他の価値ある技術を組み込むことです。ソートやフィルターを超えて「グループ・バイ」クエリに移行したり、異種のデータセットを結合して未発見のパターンを発見したりしてみましょう。しかし、それは無数の新しいツールを追加することでも、最新のツールを選ぶことでもありません。あなたやあなたの教育機関のジャーナリズム・プログラムに適したテクノロジーを使って、学生たちに次のレベルを紹介することができます。もし、すべての学生が[データベース管理ソフトの]MS Access を持っているような大学のプログラムなら、そのようなテクノロジー[を今さら教える]必要はありませんが、使うのならポイント＆クリックのインターフェースの背後にある、各クエリの背後にある SQL を学生が理解できるようにすることです。あるいは、MySQL を使うのもいいでしょう。Jupyter Notebook で Python を使うのもいいでしょう。R と R Studio は SQL のようなクエリのための素晴らしいパッケージを持っています。目標は、生徒たちにジャーナリズムを教えながら、何が必要で、ストーリーを伝えるためにデータを操作する

方法はたくさんあることを理解させることです。 繰り返しになりますが、[とにかく]シンプルにすることです。ツールを優先して学生に輪くぐりをさせるようなことはしないでください。ジャーナリズムをより強力に、より簡単にするためにツールを使用するのです。カーキャンプの例えに戻ると、授業には必要なものだけを詰め込めばいいのです。必要なものが手斧やポケットナイフだけなら、チェーンソーは持っていかないでください。

しかし、クラスを確立したら、そのモデルを超えて考えることも必要です。学部や学校全体でデータジャーナリズムの要素を取り入れる方法を考えてください。他のクラスとモチベーションを共有する方法を見つけてください。基礎的な報道の授業を担当している同僚と協力して、学生にデータの統合を深掘りさせることに興味があるかどうか確認できますか？ジャーナリズムの教授の中には、スキルの指導、吟味思考、理論的考察のバランスをとるために、「反転授業」モデルを実験的に導入している人もいます。学生は自分のペースで個別指導を受け、授業中は講師と問題解決に集中し、データジャーナリズムのさまざまな課題に取り組むための他の手法も学ぶことができます。例えば、フロリダ大学のマクアダムス教授は、Web アプリ設計の授業で反転授業モデルを活用しています。このような授業の利点は、さまざまなスキルレベルのジャーナリストに対応できることです。ジャーナリズムの授業では、コンピュータサイエンスに長けている学生が興味を示すこともあれば、同時にスプレッドシートを使ったことがない学生もいることでしょう。しかし、データジャーナリズムを教えることは、反転授業にとどまりません。データジャーナリズムの概念を教える他の方法を考えることも含まれています。定期的に開催されているアンカンファレンス「SRCCON」で、アリゾナ州立大学でデータジャーナリズムのナイト・チェアーを務め、直近ではニューヨーク・タイムズでピュリッツァー賞を受賞したジャーナリストであるサラ・コーエンは、他のアナログ活動を使って学生の興味を引くことを提唱しています。コーエンとネブラスカ大学の実務担当教授であるウェイトは、世界中の教育者が利用できるモジュールを備えた共通カリキュラムのアイデアを紹介しています。教授がゼロからすべてを構築しなくてもよいシステムを作ることが目的です。

2018 年夏のカンファレンスで、彼らは参加者のグループを率いて、そうしたモジュールを提唱しました。コーエンはグループに「私たちは、そのようなもの（ツール）に宗教を持たないようにしています 」と語り、代わりに 「ジャーナリズムの基本的価値とデータ分析の基本的価値」に焦点を当てるべきだと主張したのです。

現在、GitHub のリポジトリには、データジャーナリズムの教育で使用するためのモジュールが追加・調整されています2。このハンドブックなど、データジャーナリズムを教えるための他のリソースへのリンクも用意されています。モジュールには、世論調査や研究結果の解釈などが含まれます。基本的な計算能力は、ジャーナリズムのコースで重要な要素です。また、オンラインでデータを検索することも、どの授業でもすぐに役立つ方法です。

自分の自由な時間をすべて犠牲にする必要はありません。モジュールやチュートリアルを一度作れば、他の人が何度でも使うことができます。あるいは、すでにある多くの無料のチュートリアルを利用することもできます。Investigative Reporters and Editors（IRE）や National Institute for Computer-Assisted Reporting（NICAR）が毎年開催するカンファレンスでは、ピボットテーブルからスクレーピング、マッ

ピングまで、会員向けにさらに多くのチュートリアルが提供されています。私は四半期ごとに、オンラインでデータを見つける方法のゲスト講師をしています。データジャーナリズムに関心を持つ学生とパイプを作ることができ、また同僚の教員と協力的な雰囲気になれます。可能であれば、同僚が取り組みやすいモジュールを作ってください。例えば、環境ジャーナリストは、スプレッドシートを使って、平均気温の時系列変化に関するモジュールを作成することができます。データジャーナリズムの価値を同僚に示すことができ、これらの実践やアプローチを体系的に統合するカリキュラムのケースを構築するのにも役立つかもしれないのです。

統合モデルの詳細、または国境を越えた教育

完全統合モデルとは、データジャーナリズムの概念を教えるために、複数の人が投資することを意味します。また、ジャーナリズム・プログラムの枠を超える可能性もあります。スタンフォード大学は、「Stanford Open Policing Project」を立ち上げ、警察データの分析をジャーナリストに教えるため[ジャーナリズム研究機関]ポインターと提携しています。工学とジャーナリズムの教授が協力して、[学問の]境界を越えた授業を行い、ジャーナリズムの学生、法学部の学生、コンピュータサイエンスの学生を教育してきました。これは重要なことです。なぜなら、ニュースルームで最高のコラボレーションをするチームには、複数の分野の人々が含まれているからです。

最近では、学術機関はこのような統合モデルを採用しているだけでなく、ニュースルームに届く作品を制作し、同時に学生に教えているのです。スクリップス・ハワード財団は今月、アリゾナ州立大学とメリーランド大学に300 万ドルの助成金を提供すると発表したばかりです。両大学は、調査報道センターを立ち上げます3。これらのセンターは、学生を訓練し、調査研究を行い、報道人とトレーナーの役割を果たします。使命を掲げ、教室の枠を超えた授業は、より説得力があり、より魅力的な学習体験を提供できます。私が参加した中で最も成功したクラスの一つは、2018 年の春に私と(スタンフォード大学)工学部助教授のシャラド・ゴエルが教鞭を執った「法、秩序とアルゴリズム」のクラスです。

クラスのタイトルはゴエルがつけましたが、ひねりを加えました。私の権力監視のためのクラスを、ゴエルのクラスと連携することにしたのです。この二つのクラスで、コンピューター科学やエンジニアリングの学生、法学部の学生、ジャーナリズムの学生を教えました。学生たちは、自分たちのプロジェクトから高度な統計分析や白書、ジャーナリズム[作品]を作り上げました。ゴエルと私は、それぞれ自分の専門分野で講義をしました。私は法律やアルゴリズムの善悪について学びましたし、ゴエル教授は調査やデータジャーナリズムに何が必要か学んだことでしょう。学生たちは、プロジェクトベースの授業ということもあり、自分のチームのプロジェクトの目標を達成するために必要なことを学んでいました。私たちが避けたのは、学生にツールやテクニックを詰め込み、プロジェクトが少しずつしか進捗しないことでした。成功のために必要なことだけを詰め込んだ、カーキャンプのような授業を目指したのです。

脚注

1. この章はコロンビア大学と John S. and James L. Knight 財団の支援により出版された『Teaching Data and Computational Journalism』の共著者、チャールズ・ベレットによる。

2. github.com/datajtext/DataJournalismTextbook

3. 調査報道センター立ち上げのための助成金については Boehm, J. (2018, August 6).Arizona State University, University of Maryland get grants to launch investigative journalism centers, *AZCentral*. amp.azcentral.com/amp/902340002 を参照。

ラテンアメリカの女性やマイノリティとのデータプロジェクトを組織化する

文：エリアナ・アンドレア・バカ・ムニョス

概要

ラテンアメリカの女性やマイノリティとデータプロジェクトを組織することについて論じます。

キーワード：アナログ・データビズ、マイノリティ、女性、ラテンアメリカ、データジャーナリズム教育、データ・ビジュアライゼーション

Chicas Poderosas（パワフルな女たち）はジャーナリスト、デザイナー、開発者による国境を越えたネットワークで、ラテンアメリカ中の女性や社会から疎外されたコミュニティによるデジタルメディアプロジェクトの開発に取り組んでいます。私は Chicas Poderosas のデザイナーとして、文化的遺産、アイデンティティ、社会権力の独占、女性と脆弱な人々の認識を探求するための学際的かつ参加型の研究を含め、文化と多様性のエージェントとしてデザインが果たすことができる役割を追求しています。ここでは、コロンビアと中米にあるいくつかの傘下組織を論じましょう。コロンビアでは社会的・文化的不平等が拡大するなか、少数派の声に耳を傾け、知識を共有し、対等に扱うことが重要になっています。私たちのチームは、データジャーナリズムと関連するデジタルメディアの実践をめぐる一連の共同ワークショップを実施しました。以下のセクションでは、これらのワークショップへの参加を促進するために用いた二つの方法、すなわちアナログデータの収集と可視化について考察します。

これらのアプローチは、アクセシビリティ、デバイス、技術的リテラシーが所与のものではないコミュニティや地域におけるデータジャーナリズムの実践と文化に関連します。

アナログデータの収集

2016 年 5 月、チカス・コロンビアはチョコ県[都]キブドのヴィラ・エスパーニャで、AJODENIU（避難民の青年協会）に所属する女性やティーンエイジャーと活動しました。2002 年以来、このグループは、チョコ、リオ・スシオ、ボジャヤ、ウラバから避難してきた子どもたちの利益と権利を守るために活動してきました。これらの地域はいずれも往来が難しく、インターネットもなく、利用できる支援サービスもほとんどありません。そのため、ワークショップではまずアナログな手法で質的なデータを収集しました。このデータをもとに、強制移住や 10 代の妊娠などの問題について、口語や書き言葉でストーリーを構築していきました1。こうしたアプローチを踏まえ、2018 年 9 月にはホンジュラスの国連開発計画とともに、Observatorios de Municipales de Convivencia y Seguridad Ciudadana（共存と市民の安全に関する自治体観察所）とワークショップを開催しました。

彼らは、男女の暴力死に関するデータを扱っており、性別ごとに集計されたデータを提示しようとしていました。ワークショップを通して、このような難しいテーマについてコミュニティと会話するための感情的な道筋を作ること、そして、限られた資源を使って繊細で重要なデータを共有することの二つを目的としました。最初のステップは簡単で面白い質問でアイスブレーキングを行うことです（figure47.1）。ホンジュラスのワークショップでは、社会規範の違いや言葉の壁から、参加者と暴力について話し合うことが困難でした。そこで私たちは、暴力に関するさまざまな概念や経験を明らかにしようと、さまざまな探索的データ収集活動に重点を置いてワークショップを行ったのです。私たちは、絵や写真を使ってポスターを一緒に作りました。参加者がこれにシールを貼ることで、自分自身のイメージや権利の理解、自分自身の生活の中で様々な種類の DV（身体的、心理的、経済的など）をどのように経験したかといったデータを収集することができました。

アナログデータの可視化

先住民族が抱える問題をより深く理解するため、2017 年はビジア・デル・フエルテ地方のエンベラ族との対話型ワークショップを企画しました。このワークショップは、言葉の壁を超えて、彼らの生活を知る窓を提供しようというものです。歴史的に部族と部外者の交流は男性中心だったため、女性の教育レベルを把握し、エンパワーメントに関する議論を促そうと、女性層へのアクセスを優先しました。

Figure 47.1. “I’m so creative?”: アナログなデータ収集活動 出典：エリアナ・バカ

近代的なテクノロジーがなくても、参加者の生活にもっと有意義にアクセスする手段として、伝統的な文化表現を探りました。織物、ビーズ、工芸品などです（figure47.2）。2018 年 9 月にホンジュラスで、私たちは、被害者や危険にさらされている人々とのレジリエンスプロジェクトを実施し、どのようにデータを「人間らしく」するかという問いを中心にワークショップを行いました。私たちは、ハサミ、紙、ステッカー、ポストイット、風船を使った共感的なデザイン手法で、低コストのアナログデータ可視化ワークショップを行いました。これは、これらのコミュニティをよりよくサポートするために、関連組織と機密情報を共有することを促進する役割を果たし、また、脆弱で識字率の低い人々が自分たちの生活、経験、問題についてのデータを共有するために使えるさまざまな方法を教えるものでもあります。

例えば、私たちは参加者とともに、地域、種類、年齢別の殺人と女性の殺害に関するアナログの可視化資料を作成しました。ベリーズでの別のワークショップでは、犯罪や暴力に関するデータを可視化するためのさまざまな共同アプローチを検討しました。私たちはもともと、ベリーズの暴力観測所からのデータが、さまざまな種類の集団的反応を調整するために使用できることを確認しようとしていました。参加者は高レベルのリテラシーを持っていましたが、技術的なリソースとアクセシビリティははるかに不安定で、基本的なオンライン可視化ツールを使用することは困難だったのです。このことから、オンラインによるデータ可視化の実践について多くの疑問と課題が浮かび上がってきました。オンラインによるデータ可視化は、当然のことと思われがちですが、私たちがいた環境では通用しないものであり、より入手しやすい素材を使ったアナログなデータ可視化の妥当性を改めて示しています。

Figure 47.2. ビーズを使ったアナログ視覚化の例 色は話す言語を表し、ビーズの量はそれぞれの流暢さを表す 出典：エリアナ・バカ

脚注

1. chicaspoderosas.org/2016/11/22/the-paci c-counts-chicas-poderosas-quibdo-colombia

データジャーナリズムの系譜学

文：C.W.アンダーソン

はじめに

なぜデータジャーナリズムの歴史に関心を持つべきなのでしょう。「歴史」は多くの人にとっては学術的で抽象的な話題で、仕事に追われる現役のデータジャーナリストには特に縁遠いものかもしれません。ジャーナリストは、厳しい締め切りの中で、できるだけ多くの読者に複雑な情報を素早く分かりやすく伝えようと仕事をしているのですから、自省に時間を費やすことを嫌うのは当然でしょう。しかし、データジャーナリズムやコンピュータ支援報道の実践や概念を巡り、歴史的思考を忌避することは、質の高いジャーナリズムの生産を妨げかねません。

データジャーナリズムは、今日の世界で最も強力なジャーナリズムの集団的センス・メイキングの形態かもしれません。少なくとも、最もポジティブで実証的なジャーナリズムの形態でしょう。この力（修辞的な力とともに、質の高いジャーナリズムを生み出すデータジャーナリズムの能力）、積極性（ほとんどのデータジャーナリストは、自分たちの分野の将来が上昇傾向にあると確信しています）、実証性（データ記者は、世界に関して証明可能な事実を捉える、方法に基づいた調査能力の信奉者です）により、私が「経験的に自信に満ちた職業」と呼ぶものが生み出されています。この自己肯定感は、データジャーナリズムは常に世界を改善し、向上させているという（進歩主義的）ホイッグ史観のような思い込みも生みかねません。こうした態度は、傲慢さと批判的な自己反省の欠如を招き、ジャーナリズムを、その時間を費やして責任を追及している機関と同じようなものにしてしまう可能性があるのです。

ここでは、歴史にもっと注意を向けることで、データジャーナリズムの日々の活動を実際に改善できることを主張したいと思います。自分たちのプロセスや実践には歴史があることを理解することで、データジャーナリストは、現在の物事がかつてそうでなかったかもしれず、別の方法で行われる可能性があるという事実を受け入れられるのです。特に、データジャーナリストは、実証的な作業で不確実性をクリエイティブに表現する方法について、より深く考えることができるかもしれません。また、単に事実上の証拠を述べるだけでなく、さまざまな政治的感性や説得力のある、読者を引き込むためのテクニックを考えるかもしれません。つまり、科学技術研究者で歴史家のキャサリン・ディグナツィオとローレン・クラインが「フェミニスト・データ・ビジュアライゼーション」と呼ぶ、二項対立を再考し、多元性を受け入れ、権力を検証し、文脈を考慮するような可能性への道を開くことができるのです。

データジャーナリズムは自らを恃みがちなので、これらの変化を達成するためには、ジャーナリズムの他の形態よりもいっそうこの強力な歴史感覚を教え込む必要があります。私は、ミシェル・フーコーによって開拓され、科学の歴史家や科学技術研究者によって受け入れられた概念の発展への系譜学的アプローチほど、自己再帰性を導く能力を備えた歴史の形式はないと思います。フーコーが定義した「系譜学」は、ニーチェの著作を参考にしながら、制度や概念の時間的変遷を研究するユニークなアプローチであり、歴史とは異なります。系譜学

的分析は、過去の慣習や思想の切れ目のない単一の起源を探すものではないし、概念が昨日から今日まで切れ目のない進化的なラインでどのように発展してきたかを理解しようとするものでもないのです。むしろ、現在からみて過去がどうだったかということよりも、不連続性や予期せぬ変化に注目します。ニーチェは、フーコーが引用した『道徳の系譜』の一節で、次のように指摘します。

ある物事、慣習、器官の「発展」は、単一の目標に向かう進歩とは何の関係もなく、ましてや力や資源の最小限の消費で到達する論理的かつ最短の進歩でもない。むしろそれは、その物に起こる多かれ少なかれ深遠で、多かれ少なかれ相互に独立した圧倒のプロセスの連続であり、その都度その圧倒に対して生じる抵抗、防御と反作用の目的で試みられた形態の変化、および成功した対抗措置の結果なのである。形は流動的であるが、「意味」はさらに流動的である。（Foucault, 1980）

「データジャーナリズムの系譜学」は、データジャーナリズムが、その創作者や実践者が予想しなかった方法で、あるいは彼らの願望に反する方法で進化してきたことを明らかにします。歴史が私たちを驚かせ、時には予期せぬ方向へ導いてくれるのを目の当たりにしているのです。このアプローチは、先に述べたように、現代の現役データジャーナリストに特に有効なものです。自分たちは由緒ある伝統の中で働いているのではなく、むしろ、まったく偶発的な方法で、自分たちの手で作り上げているのだということを理解する助けになると思うのです。そしてそれは、批判的な自己反省を促すものであり、現役のデータジャーナリストや記者の（理解できる、しばしばそれに値する）自負心を和らげるのに役立つかもしれないのです。

私は拙著『確かさの使徒たち：データジャーナリズムと疑念の政治学』で、このような系譜の記述を試みています。この後のページでは、この本の主な発見をいくつかまとめ、その教訓が現代に役立つかもしれない方法を論じたいと思います。最後に、ジャーナリズム、特にデータ化されたジャーナリズムは、知らないことをよりうまく示すことができるし、そうすべきです。不確実性を論じるこうした営為は、データジャーナリズムの起源が表象化されたものであると言うよりも、その起源が違法な権力への批判にあるということを尊重するものです。

データジャーナリズムの変遷：1910 年代、1960 年代、2010 年代

より良いジャーナリズムを生み出すために、数値の（記された）紙文書、データの可視化、チャートやグラフといった他の形式の定量化された情報とともに、データを活用できますか。そして、そのジャーナリズムは、国民がよりよい政治的選択をするために、どのように役立つのでしょうか。「確かさの使徒たち：データジャーナリズムと疑念の政治学」に通底するのはこうした疑問で、ニュースの歴史をより長期的な視点から探ろうとしています。本書は、1910 年代、1960 年代、そして現在に立ち戻り、データジャーナリズムの系譜とその物質的・技術的基盤をたどり、報道におけるデータの利用が、国政や計算可能なデータベースの進化、専門科学分野の歴史と不可避的に絡み合っていることを論じています。

社会科学とジャーナリズムの関係を理解しなければ、データのジャーナリズム的利用を理解することは不可能です。また、実証的に検証可能な情報の出版がより公正で繁栄した世界につながるという非常に根強い進歩的信念を最初に理解しなければ、公の真実を語る実証的な形式を解き明かすことも不可能です。この本はこのテクノロジーとプロフェッショナリズムの交差が、より良いジャーナリズムをもたらしましたが、必ずしもより良い政治をもたらしたわけではないと結論づけています。デジタル時代の要求を十分に満たすためには、ジャーナリズムは確かさだけでなく、経験的な疑念を表現することにも慣れていなければなりません。皮肉なことに、この「疑いの受容」によって、ジャーナリズムはより科学に近いものになることができるのです。

社会科学への挑戦

「確かさの使徒たち」の物語は、米国の三つの時代に立脚しており、データジャーナリズムの発展についてそれぞれ異なる視点を提供しています。第一は、いわゆる「進歩の時代」であり、入手可能な最高の統計によって、国家と一般市民がともに世界をより公正で人道的な場所にすることができるという信念を伴ったリベラルな政治の時代でした。第二の画期は、1950 年代と 1960 年代で、改革者の一部が、ジャーナリズムをより経験的で客観的なものにする新しいアイデアと方法の源として、定量的社会科学、特に政治学と社会学に目を向け始めました。彼らを後押ししたのは、ますますアクセスしやすくなったデータベースと強力なコンピューターです。

第三の画期は 2010 年代初頭で、データジャーナリズムの最先端が「計算」あるいは「構造化」ジャーナリズムによって補完された時期です。ビッグデータと「深層機械学習」の現在において、これらのジャーナリストは、ジャーナリズムの客観性は外部参照に依存するのではなく、むしろデータベース自体の構造の中から現れると主張しています。これらの時代において、データ指向のジャーナリズムは、社会科学全般の大きな流れに呼応しながらも、それに部分的に対抗する形で自らを定義しており、この大きな政治・社会学の潮流との関係が、本章で取り上げる事例の選定に役立ちました。

言い換えれば、私はジャーナリズムの歴史において、ジャーナリズムだけでなく、より大きな社会的・政治的構造に光を当てるのに役立つ変曲点を探ったのです。進歩主義時代1、伝統的なニュース報道は、社会学が新たに注目した社会構造や非人格的なコンテキスト情報をはねつけ、有力な人物や重要な出来事に焦点を当てた個人主義的な報道を維持することを好んでいました。しかし、フィリップ・マイヤーやプレシジョン・ジャーナリズムの哲学を中心とする改革者たちが 1960 年代に入って、ジャーナリズムが目指すべき正確さとコンテキストの水準を示すモデルとして、計量社会学と政治学を取り上げ始めました。

21 世紀に入ると、データジャーナリズムの標準的モデルは、社会科学をますます悩ませるようになった再現性と因果性への疑念に取り組み始めました。社会科学のように、「ビッグデータ」と相関的行動主義の非因果的形態が社会活動への洞察を提供できるかどうかを試すようになったのです。『確かさの使徒たち』は、ジャーナリズムの専門性と権威の形態は、決してジャーナリズムの分野そのものから切り離されて構築されるものではなく、完全に内部的なものでもないことを示唆します。データジャーナリズムは、ジャーナリズム内部の事情のみ

でデータジャーナリズムになったわけではないし、このプロセスは、ジャーナリズムの言説や自叙伝的分析で説明できるものでもありません。1960 年代にデータジャーナリズムと呼ばれ始めた専門性のタイプは、データジャーナリストが（より権威と力のある）社会科学の同胞に反応し、交流する方法を調べることによってのみ、関係的に理解できるのです。さらに言えば、このプロセスは、単独であれグループであれ、人間の行動や闘争という観点からだけでは理解できないと思います。

私が『確かさの使徒たち』で提示したモデルによれば、専門知識とは、専門家集団が多種多様な言説的・物質的人工物に対する管轄権を確立しようと争う、ネットワーク化した現象です。データジャーナリズムは、簡単に言えばデータベースの存在なしには不可能だったわけですが、データベースとは何か、どのようにすればジャーナリズムにふさわしい形で展開できるかという特定の専門家の理解によって媒介されたものです。社会科学の権威を理解せずにジャーナリズムの権威を理解することは不可能です（コンピュータ科学、人類学、長編ノンフィクションについても同じことが言えるかもしれません）。ジャーナリズムの専門性と知識は、ジャーナリズムという分野そのものを見るだけでは決して理解できないのです。

政治の永続性

データジャーナリズムは、社会学や政治学といった隣接する専門分野との関連から、系譜学的に理解されなければなりません。そして、これらの分野はすべて、政治に関するより大きな概念を通じて分析され、彼らが発見した「事実」が好むと好まざるとにかかわらず「政治的」であるという事実と、どのように折り合いをつけているかが明らかにされなければならないのです。事実に基づく知識を求めること自体が政治的行為です。データジャーナリズムの歴史を振り返る『確かさの使徒たち』の中で、進歩的な政治目標としか言いようのないものを実現するために、社会科学の中立性に依拠しようとする試みを紹介しました。しかし、この関係が形成される大局的な状況は、時間の経過とともに劇的に変化しています。こうした大きな変化は、ジャーナリズムにおいてわれわれが目撃しているものが、高度化するデジタル機器によって実現されたジャーナリズムの確実性の発展だといった、いかなる熱狂をも抑制するはずです。

進歩主義時代、データジャーナリストの先達は、定量的な事実を収集し積み重ねることを社会的・政治的啓蒙の過程とみなしていたにもかかわらず、その過程には大きな政治的コミットメントがないとしていました。社会調査員たちは、都市の衛生状態、都市空間における貧困の分布、教会への出席や宗教的実践に関する統計、労働条件など、さまざまな事実をきめ細かく収集し、報道という媒体を通じて一般大衆に伝えることによって、社会的有機体は自らの存在条件についてより強固な理解を得られると考えていたのです。社会の理解が深まれば、社会は自力で改善し、政治家に改革を促すことができるというわけです。この場合、世界に関する事実を巡る知識はそれ自体で語られるものであり、それを集め、可視化し、公表するだけで、啓蒙は後からついてくるのです。これは、事実とは何かについての「ナイーブで透明な」概念と呼べるでしょう。事実それ自体には何の解釈も必要なく、その集積が社会のポジティブな変化につながるのです。このようなデータジャーナリズムは、政治性

を明示しないのと裏腹に、実際には政治的だったのです。フィリップ・マイヤーが登場した 1960 年代の時点で、透明な事実と政治との間のこの安易な一致は崩壊していました。

マイヤーとその支持者たちは 1950 年代から 60 年代にかけて、ジャーナリズムには欠陥があると主張しました。ジャーナリズムは客観性を、政治問題のすべての立場の人間が真実だと考えるような記録を単に集め、読者が何が真実かを自ら判断できるようにすることと取り違えていたからです。社会の激動と政治的混乱の時代にあって、ジャーナリズムの客観性はより強固な基盤を必要とし、客観的な社会科学にその足場を見出したのでした。ある問題についてのジャーナリスティックな報道の出発点は、利己的な政治家の言説ではなく、適切な方法による関連データの分析から得られた冷厳な真実であるべきなのです。そのような分析は専門的であっても、政治的ではありません。高度に専門化された真実を語る集団として行動することで、ジャーナリストは政治的混乱を切り抜け、客観的真実の領域に国民を導くことができます。その一方で、この真実が導くはずの方向性については、何の関心も持ちませんでした。以前の世代の、至って単純で進歩的なデータジャーナリストとは異なり、データがもたらす啓蒙的な結果は、当然の結論ではなかったのです。

私はここで、新しい世代のコンピューター・ジャーナリストが、進歩主義時代の先達の政治的・認識論的信念のいくつかを知らず知らずのうちに再吸収していることを指摘します。認識論的には、コンピューター・ジャーナリストの間では、デジタルな事実は何らかの形で「自らを語る」、少なくともこれらの事実が適切に収集、分類、クリーニングされたときにはそうなる、という考えが広まっています。規模が大きく、内部的に一貫性のあるセマンティック・データベースにリンクされている場合、事実は一種の相関的過剰を生み出し、意味や因果関係の問題が計算データの洪水によって洗い流されます。データジャーナリストは、客観性がより大きな職業的解釈プロセスの一部としてではなく、データベースの構造そのものから生まれるものだと理解するようになっていきています。政治的には、最も研究熱心で中立的なデータジャーナリストの多くに、一種の「暗号進歩主義」が戻ってきたと言えるでしょう。より多くのデータが美しく可視化され、強力な報道機関を通じて伝達されれば、西洋民主主義で次第に顕在化している非理性的あるいは病的な政治傾向を断つものとして機能するという政治的願望が根ざしているのです。少なくとも、2016 年、ブレグジットとドナルド・トランプという双子のショックが起こる前は、そのような希望がありました。

確実性と疑心暗鬼

20 世紀に大きな展開をみせた米国におけるデータジャーナリズムは、専門家らしい確かさをますます厳密に主張する一方、すべての事実はその出自を問わず政治に汚染されているという認識も共存している状況下で発展しました。こうしたしばしば矛盾する信念は、もちろんジャーナリズムだけでなく、データを重視するさまざまな分野で顕著に見られます。たとえば、科学コラムニストのエド・ヨンは、2017 年のアトランティックの記事で、「オープンサイエンス」の機運や再現可能性の危機の高まりによって、反科学的な議会が科学研究を貶め、資金を奪うために利用する可能性を論じています。ヨンは FiveThirtyEight の科学記者のクリスティー・アシュワ

ンデンの言葉を借りて「一般の人々が科学について考えることには二つの相反するものがあるように感じる」と述べました。

触れるものすべてを真実に変える魔法の杖なのか、それとも私たちが以前考えていたことが変わってしまったためにでたらめになってしまったのか…真実はその中間にある。科学は不確実性を減らす過程だ。不確実性がプロセスの一部であることを示さなければ、疑い深い人が不確実性を使って物事を台無しにするのを許してしまうことになる(Yong, 2017)。

こうした考えは、科学技術社会論研究者のヘルガ・ノヴォトニー(2016)が『科学と不確実性に立ち向かうわざ』で論じている、「不確実性の克服と確実性の追求の相互作用が、知りたいという欲求を支えている」という作用と一致します。近代科学の本質は、少なくともその理想的な形においては、確実性を達成することではなく、むしろその知識の暫定性を公然と表明することです。科学には何一つ確定しているものはないのです。科学は、しばしばほとんど何も知らないことを認めます。この、最も現代的な逆説によって、科学が主張する知識は、人々の信頼に値するものとなるのです。データジャーナリズムの発展と展開に関するこの系譜学的概観から得られる洞察の一つは、データ指向のジャーナリストは、暫定性と疑いに対するつつましい理解を犠牲にして、正確さと確実性を高めることに執着してきたことです。これまで述べてきたように、20 世紀半ば以降、ジャーナリストたちは、自らの知識の主張をより確かなものにし、文脈に即して説明できるようにする努力を続け、その成果を上げてきました。さまざまな形の証拠、特に定量的な証拠を活用することによって、これを実現してきたのです。

とはいえ、こうしたプロフェッショナリズムの高まり、そして、自分たちは文脈に即した真実の主張をすることができるというジャーナリストの自信の高まりが、必ずしもジャーナリストが期待するような民主的成果をもたらしてきたわけではないことは明らかです。現代アメリカの政治言説は、確実性を強く主張することによって、現代特有の不確実性に対処しようとしてきました。プロのジャーナリズムはこのジレンマを解決することなく、むしろ悪化させてきました。現代世界の複雑さに対処するために、ジャーナリズムは自らの暫定性と不確実性を伝える手段とメカニズムを再考する必要がある、というのが私の結論です。これが正しく行われれば、ジャーナリズムは現代科学のよう[な営み]になるはずです。

脚注

1. 1880～1920 年代の米国では、「進歩主義時代」が続き、一般に、自由主義的な改革を行い、公共政策を産業時代に適合しようとした偉大な時代と見なされている

参考文献

Anderson, C. W. (2013). Towards a sociology of computational and algorithmic journalism. *New Media & Society*, *15*(7), 1005–1021. doi.org/10.1177/1461444812465137

Anderson, C. W. (2018). *Apostles of certainty: Data journalism and the politics of doubt*. Oxford University Press.

D' Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings,1972–1977*. Vintage.

Nowotny, H. (2016). *The cunning of uncertainty*. Polity Press.

Yong, E. (2017, April 5). How the GOP could use science's reform movement against it. *The Atlantic*.

www.theatlantic.com/science/archive/2018/08/scientists-can-collectively-sense-which-psychology-studies-areweak/568630/

データ駆動の規範：データジャーナリズムは何を評価し、社会のデータ化とともにジャーナリズムはどう進化していくのか？

文：ヴィプケ・ルースン

概要

社会のデータ化とジャーナリズムの関係を探り、その進化を研究する手段として、データジャーナリズム・アワードを取り上げます。

キーワード：データジャーナリズム・アワード、データ化、データ化されたジャーナリズム、データ社会、ジャーナリズム研究、共創

はじめに：社会のデータ化に伴うジャーナリズムの対応

データジャーナリズムの初期よりも、この新しい報道スタイルの出現は、社会のデータ化に対するひとつのジャーナリズムの反応であるという理解が浸透しつつあります（Losen, 2018）。データ化とは、私たちの（メディア）環境のデジタル化と、そのような環境で生活することで発生するデジタル痕跡やビッグデータといったデータへのアクセス性が増大していることです（Dijck, 2014）。このプロセスは、私たちの社会生活の多くの側面をコンピュータ化されたデータに変換します。データは、さまざまな目的に合わせて集計され、アルゴリズムによって処理されます。データ活用はさまざまな結果をもたらし、政治の世界では、金融の世界や教育の世界とは異なる形で現れます。しかし、すべての社会領域に共通しているのは、（自己）感覚の形成過程において、これまで以上に多様で大量のデータにますます依存するようになるだろうということです。ジャーナリズムのデータ化をより広い社会のデータ化との関係で位置づけるならば、データジャーナリズムを俯瞰し、ジャーナリズムにおけるデータ化の「唯一の」源として認識し、より多くのデータベース、アルゴリズム化、指標に基づき、あるいは自動化の実践に向かってジャーナリズムが変容していくことをより深く理解できるでしょう（Loosen, 2018）。

特に、ジャーナリズムがカバーすべき対象や話題、別の言い方をすれば、ジャーナリズムが社会の観察者として機能することに当てはまります。ジャーナリズムがカバーすべき分野や社会領域そのものが「データ化」されればされるほど、ジャーナリズム自身がその社会的役割を果たすには、データを意味付け、生産する能力が必要なのです。この関係性は現代のデータジャーナリズムに反映されており、データの利用可能性の増加に依存して、ジャーナリズムの研究やストーリーの特定、伝達に役立つ情報源のレパートリーを拡大しています。

アワード：データジャーナリズムの定義と評価について考察する手段

報道スタイルとしてのデータジャーナリズムの変遷は、そのアウトプットを通じてたどれます。初期のジャーナリズム研究は、その制作者に焦点を当て、主にインタビューに基づいて行われる傾向がありましたが、最近では、コンテンツ分析に基づいてデータジャーナリズムをよりよく理解しようとする研究が増えています（Ausserhofer et al.、2020）。ジャーナリズムアワードは、いくつかの理由から、適切な実証的アクセスポイントだと言えます。第一に、賞の応募作品は、ジャンルやストーリーテリングの側面を分析するための有用な対象であることが証明済みです（Wahl-Jorgensen, 2013）。第二に、データジャーナリズムは拡散した研究対象であるため、内容分析のためにそれぞれの作品を特定することは困難で、むしろ（特定されていることが）前提条件なのです。この戦略は、本質的にジャーナリズムの自己観察を行うための手段であり、この分野自体がデータジャーナリズムとみなし、この報道スタイルの重要な事例だと信じているものを代表するものです。第三に、国際的な賞へのノミネートは、認知度が高く、一種の規範であり、国境を越えた影響力を持つため、この分野全体の発展に影響を与える可能性があります。また、国際的な賞を対象とすることで、地理的・時間的に広い範囲をカバーするサンプルを調査することができるのです。

ただし、（ジャーナリズム）賞の研究には、さまざまなバイアスがかかることも念頭に置くべきです。ここで示す調査は、Global Editors Network1が毎年授与するデータジャーナリズム・アワード（DJA）の 2013 年から 2016 年のノミネート作品 225 点（受賞作品 39 点を含む）の分析に基づいています（Loosen et al.、2020）。つまり、我々のサンプルは二重の選択バイアスにさらされていることになります。ノミネートは、ジャーナリスト自身の推薦によるのですから。第二段階では、多かれ少なかれ毎年変わる専門家の審査員が、どの作品が実際にノミネートされるかを決定します。また、賞やアワードは「文化資本」の一種であり、受賞したプロジェクトがその分野全体に一定のシグナル効果をもたらし、後続のプロジェクトのモデルとなります（English, 2002）。このことはまた、賞が（一定の基準に従って）その分野を代表するだけでなく、それを構成することも意味します。つまり、われわれの場合、（受賞した）コンテンツをデータジャーナリズムと呼ぶことで、賞はさまざまな実践、行為者、慣習、価値を集める役割を担っているのです。データジャーナリズム・アワードは、賞を授与する機能だけでなく、現場を形成する機能も持っているといえます。つまり、受賞に値する作品は、常に応募者と審査員による「共同構築」の結果であり、その期待値は相互に形成されるのです。データジャーナリズムは、まだ比較的新しい報道スタイルであり、この分野のすべての関係者が多かれ少なかれ模索を続けているので、こうした効果は特に影響力が強いようです。

進化しても、革命ではない：（受賞に値する）データジャーナリズムのいくつかの傾向

データ駆動型記事を分析した研究は、一般に、データジャーナリズムの進化は決してニュース手法の革命ではないと実証しています。データ駆動型ジャーナリズムが従来のニュース発見・報道の手法に取って代わり、ジャーナリズムに革命をもたらすという広範な信念に疑問を投げかけているのです。私たち自身の研究は、「日々の」データジャーナリズムのサンプルの他の実証的分析が発見したものとおおむね一致しています（Loosen et al.）。これらはかなり限定的なデータ収集に過ぎないものの、少なくともデータジャーナリズムのアウトプットの

発展や、ある程度の一貫性を追跡可能です。誰が受賞に値するレベルのデータ駆動型ジャーナリズムを生み出しているかという点で、データジャーナリズムの「規範」、つまり同業者の評価に値するものは、新聞社とそのオンライン部門が独占しているという結果がでています。4 年間の分析で、新聞社およびそのオンライン部門は、候補者および受賞者の中で圧倒的に大きな割合を占めています（合計：43.1%、DJA 受賞：37.8%）。

他に目立つのは、プロパブリカや ICIJ などの調査報道を行う組織で、これらの組織は受賞回数が圧倒的に多くなっています。これは、この賞の本質的なバイアスが、確立された知名度の高い関係者に向けられていることを反映しているのかもしれません。一定レベル以上のデータジャーナリズムは、ライター、プログラマー、グラフィックデザイナーからなる分野横断的なチームに投資するリソースと編集のコミットメントを持つ大規模組織向けの事業であるように見えるという他の調査結果にも現れています (Borges-Rey, 2020; Young et al., 2018)。

私たちのサンプルでは、署名のある 192 件のプロジェクトのうち、著者または貢献者として名前を挙げたのは平均して 5 人強で、プロジェクトの約 3 分の 1 は、分析に貢献した、または可視化を設計した外部パートナーとの共同作業により制作されました。このことは、受賞プロジェクトに特に当てはまるようで、我々の分析では、ノミネートされただけのプロジェクトよりも大きなチームによって制作されていることがわかりました（M = 6.31, SD = 4.7 vs M = 4.75, SD = 3.8 ）。今回のコンペティションで評価されたデータジャーナリズムの地域性については、米国が圧倒的であることがわかります。ノミネートの半数近くが米国（47.6%）で、英国（12.9%）とドイツ（6.2%）が僅差でこれに続いています。しかし、ノミネートされる国の数は年々増え、2016 年には 5 大陸から 33 カ国がノミネートされるなど、データジャーナリズムはますますグローバルな現象になっているようです。データジャーナリズムが特定の情報源に依存することは、そのトピックが取り上げられるか否かを左右します。その結果、データジャーナリズムは、データが定期的に作成されていない、あるいはアクセスできない社会領域を軽視することになりかねません。DJA にノミネートされた記事は、政治、社会、経済の問題に常にフォーカスしていることが特徴であり、分析対象のほぼ半数（48.2%）が政治に関するトピックを扱っているのです。

教育、文化、スポーツの割合が少ないのは、他の研究結果とも一致しますが、データジャーナリズム全般を代表するものではなく、（ジャーナリスト）業界の賞に特有の「深刻な」トピックの偏りが原因である可能性があります。しかし、これは異なるドメインやトピックに対するデータソースの利用可能性や利用不可能性を反映しているかもしれません。また、私たちのサンプルの場合、応募者が何を応募に値すると考え、何を審査員に評価されると期待しているかによって、自己選択バイアスがかかっているかもしれません。この極めて重要な点について、より信頼性の高い知見を得るためには、データの入手可能性と入手しやすさ、そして各国のデータ報告でカバーされているトピックを関連づける国際比較研究が必要です。このような研究はまだ文献にはないものの、どの社会領域やトピックがどの分析手法によって、どのデータソースに基づいてカバーされているのかを明らかにすることができるはずです。また、このようなアプローチは、コインの裏側、すなわち（利用可能な）データソースの不足によるデータ駆動型報道における盲点についても貴重な洞察を与えてくれるでしょう。データジャーナリズムに関するコンテンツ関連の研究で繰り返し見られるのは、統計局やその他の政府機関からの「加工前の公共データへの

依存性」を示しているということです（Borges-Rey, 2020; Tabary et al, 2016; Young et al, 2018）。賞レベルのデータ駆動型作品にもあてはまります。公的機関のデータ（データソースのほぼ 70％）、研究機関や NGO などの非商業的組織のデータ、少なくともリクエストに応じて公開されているデータ（ほぼ 45％）に依存していることが判明しました。

このことは、データジャーナリズムがデータソースの利用可能性を高めていることを意味する一方で、この利用可能性に大きく依存していることも示しています。自分で収集したデータ、スクレーピングしたデータ、リークしたデータ、リクエストしたデータのシェアはかなり小さいものです。それにもかかわらず、データジャーナリズムは調査報道と絶えず結びついており、「データジャーナリズムは、ジャーナリズムの勇気と大胆な行動によって獲得された膨大なデータセットがすべてであるという認識をもたらしている」（Knight, 2015; Parasie, 2015; Royal & Blasingame, 2015）とされているのです。パナマ文書などの最近の事例がそのような認識を助長している部分もあるでしょう。しかし、この事例が示すのは、世界的に重要で複雑な問題の中には、異なる報道機関の国境を越えた協力が必要なデータが埋め込まれていることもあるということです。さらに、金融などの国際的なデータの流れを、単にサービスとしてではなく、より深く、調査的なバックグラウンドストーリーとして継続的にモニターするルーチンがさらに開発されれば、このようなケースはさらに増えていくでしょう。そうなれば、例えば金融データの流れを常に監視し、その異常性を探るという、データに基づく新しいタイプの調査データに基づくリアルタイム・ジャーナリズムを喚起することができるのではないでしょうか。

インタラクティブ性はデータジャーナリズムにおける品質基準の一つですが、通常、比較的明確な特徴を持って実装されています。ここでの私たちの［調査］結果は、他の研究や、しばしばデータ関連の対話性における「洗練さの欠如」と表現されるものとも合致しています。ズーム可能なマップとフィルター機能によく当てはまりますが、これはおそらく、使いやすく無料で利用できるソフトウェアの使用が普及し、洗練された可視化およびインタラクティブな機能が少なくなっているためでしょう。しかし、受賞プロジェクトは、少なくとも一つのインタラクティブ機能を含み、複数の異なる（タイプの）ビジュアライゼーションを一緒に提供する傾向があります。インタラクティブなオプションがやや制限される傾向は、高度なインタラクティブ機能（ゲーム化されたインタラクティブ機能、カスタマイズされたデータで作品を調整できるパーソナライズツールなど）に対する視聴者の関心が低いという経験も反映しているかもしれません。ビジュアライゼーションと同じくインタラクティブ機能も、記事のストーリーテリングと説明を支えるには、それぞれのデータ駆動型コンテンツに適したソリューションが必要です。ここ数年の傾向をまとめると、調査対象のカテゴリーのシェアと平均数は長期にわたってほぼ安定しているか、変化があったとしても直線的な増減ではないため、やや複雑なパターンとなりました。むしろ、個々の年における頂点と底は不規則であり、データジャーナリズムのようなまだ新しい分野における試行錯誤的な進化を示唆しています。そのため、この数年間、一貫した動きはほとんど見られません。ビジネス関連記事が急増し、さまざまな種類のビジュアライゼーションの平均数が一貫して大幅に増加し、批判（例：警察の不正な没収方法）や公的介入の要求（例：二酸化炭素排出量）を含む記事の割合が（統計的に有意ではないものの）増えています。この割合は 4 年間で一貫して増加し（2013 年:46.4% 対 2016 年:63.0%）、受賞作品ではかなり高くなっています（62.2%対

50.0%)。 これは、(データ)ジャーナリズムの調査や監視の可能性に対する高い評価の表れであり、おそらく、この新興分野を正統化する方法として解釈することができます。

データジャーナリズムからデータ化されたジャーナリズムへ-データ社会における役割

データジャーナリズムは、社会のデータ化と並行して進化する新しいジャーナリズム分野であり、データの可用性の高まりに、ジャーナリズムが適応していく過程です。しかし、データジャーナリズムは、もはや急成長中の現象ではなく、主流としてしっかりと位置づけられました。それは、データジャーナリズム・アワードを見れば改めて分かることです。2018 年の大会では、「データジャーナリズムにおけるイノベーション」という新しいカテゴリーが導入され、データジャーナリズム自体はもはや革新的な分野とみなされず、現代の実践において斬新なアプローチを模索していることがうかがえるのです3。データジャーナリズムは、社会全体のデータ化の進展、すなわち、感覚や判断、あらゆる社会的行動のデータへの依存度が高まっている社会と共進化していくのではないでしょうか。ジャーナリズム全体とそれを取り巻く環境のデータ化が進み、遠からず「データジャーナリズム」という言葉が不要になることは想像に難くありません。この予言が確かなものであるかどうかは別として、「データジャーナリズム」という言葉は、「データ社会」という言葉と同様に、ジャーナリズムとその周辺における根本的な変革のプロセスを私たちに感じさせてくれるものです。それは、ジャーナリズムがどのように、どのような手段で(データ化された)社会を観察・取材し、そのパフォーマンスを自己監視し、リーチや視聴者参加をコントロールし、(自動的に)コンテンツを制作・配信するかということです。言い換えれば、現代ジャーナリズムの特徴は、よりデータに基づき、アルゴリズム化され、測定基準に基づいた、あるいは自動化された手法へと変化しているのです。しかし、データは「原料」ではなく、社会的な世界への直接的、客観的、あるいは特権的なアクセスを可能にするものではありません(Borgman, 2015)。

このようなデータの捉え方は、社会のデータ化のプロセスが進むにつれて、責任あるデータジャーナリズムにとってより一層重要なものとなっています。データ化の進展とデータ駆動型ジャーナリズムの関連性の高まりは、他の社会領域が(ジャーナリストが)利用できるデータをより多く生産したり公開したりするインセンティブを設定する可能性もあり、「データ PR」すなわち、自らの利益となるために生産・公開されるデータ駆動型パブリックリレーションとの共進化が起きそうなのです。このことは、データの質、出所、重要性をチェックするルーティンが、(データ) ジャーナリズムにとってますます重要になってきていることを意味するとともに、なぜ特定の事実や展開に関するデータが入手できないのかという疑問を呼び起こします。要約すると、七つの「C」(七つの課題と活用されていないデータジャーナリズムの能力) に従って調査結果を整理することができます。これらは、現場で修正された、または代替のプラクティスを提案するためにも役立ちます。

<u>収集</u>。調査的で重要なデータジャーナリズムは、公的にアクセス可能なデータへの依存を克服しなければなりません。データにアクセスし、それらを独自に収集するためには、さらなる努力が必要です。

協働。「日常的な」データ駆動型作品の制作はますます容易になっても、より高みを目指すプロジェクトはリソースと人員を必要とし、グローバルに関連するトピックが増加することでしょう。そのためには、国境や報道機関の枠を超えたデータによる調査や、場合によっては科学やデータアクティビズムといった他分野とのコラボレーションが必要です。

クラウドソーシング。データジャーナリズムの真のインタラクティブな可能性は、高度化するインタラクティブ機能ではなく、ユーザーや市民がデータの収集者、分類者、共同調査者として心から参加するクラウドソーシングのアプローチにあります。

共創。ソフトウェア開発の分野では共創のアプローチは一般的で、長期的なデータ駆動型プロジェクトのモデルとして参考にできます。ユーザーはトピックの発見から開発、そして長期にわたるメンテナンスまで、すべてのプロセスに関与することになります。

コンピテンシー。質の高いデータジャーナリズムには、幅広いスキルセットを持つチームが必要です。ジャーナリストの役割は依然として重要ですが、ジャーナリストはデータ、データ構造、分析手法についてより高度な理解を必要とするようになってきています。報道機関には、他の多くの産業でますます必要とされているデータアナリストを採用するためのリソースが必要になってきます。

組み合わせ。複雑化するデータには、ますます高度な分析が必要となります。データソースを組み合わせたり、様々な視点からデータを見たりする手法は、社会現象をより本質的に描き出し、データジャーナリズムの分析能力を強化するのに役立つでしょう。

複雑性。複雑さには、データそのものだけでなく、様々な社会分野や政治的意思決定におけるデータの重要性が増していることも含まれます。

こうした流れの中で、データジャーナリズムは、データ PR や「フェイクデータ」に直面することになります。これは何を意味するのでしょうか。どのようなデータジャーナリズムが評価され、広く世間の注目を集め、ジャーナリズムに対する一般的な評価に貢献しているかという観点から、（受賞歴のある）データジャーナリズムについて既に分かっていることを考慮すると、私たちはどのようなデータジャーナリズムを望んでいるのでしょう。データ社会における責任あるデータジャーナリズムとは、以下のようなものではないでしょうか。社会的な問題を調査し、データ社会を自らの手段で理解し批判できるようにすること、特定の分野でデータが不足しているのはなぜか、それは良い兆候なのか悪い兆候なのかを問いながら自らの盲点を自覚し、データ操作やデータの乱用を積極的に明らかにしようとすること、最後に、データは決して自明の事実の集まりではなく、しばしば非常に特定の条件や目的に関連して収集された「人工物」だと念頭に置き、説明し、強調することです（Krippendorff、2016 年）。

これは同時に、データジャーナリズムの特殊性、すなわちデータへの依存性が弱点であることも意味しています。この限界は、データの入手可能性、信頼性、品質、操作性に関わるものです。責任あるデータジャーナリズムは、データへの依存について反省的であるべきであり、データジャーナリズムの倫理を議論する中核的な主題であるべきです。

この社会は、私たちが(データを通じて)何を知り、どのように知るのかというジャーナリズムの前提に挑戦する、核心的な認識論的問題に満ちた社会なのです。これらの問いは、より多くの、そしてますます多様化するデータが「ニュースの回路」の様々な地点に組み込まれるにつれて、より緊急性を帯びてきます。ジャーナリスティックな観察や調査の手段として、生産や流通のルーチンの一部として、そして視聴者の消費活動を監視する手段として a)世界を観察し、データからニュースを構築するジャーナリズムの方法、(b)コンテンツ制作の自動化を促進するジャーナリズムのパフォーマンスの核心、(c)データを処理するアルゴリズムとその基本論理によって形作られる環境におけるジャーナリズムの成果の分配と流通、(d)ますますきめ細やかに測定される視聴者のセグメントにとってニュース価値があると理解されるもの、などに影響を与えているのがこれらの方法であり、データ化ジャーナリズムなのです。

こうした状況の進展は、(データ)ジャーナリズムに三つの本質的な責任を与えています。[それは]データ化された社会への発展を批判的に観察すること、それを自らの手段で理解できるようにすること、そして、データのレンズを通して何が再現され、何を見るべきかという限界を可視化することなのです。

脚注

1. www.globaleditorsnetwork.org/about-us, www.datajournalismawards.org

2. panamapapers.icij.org

3. www.datajournalismawards.org/categories/

参考文献

Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (2020). The data cation of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journal- ism*, *21*(7), 950–973. doi.org/10.1177/1464884917700667

Borges-Rey, E. (2020). Towards an epistemology of data journalism in the de- volved nations of the United Kingdom: Changes and continuities in materi- ality,

performativity and re exivity: *Journalism*, *21*(7), 915–932. doi.org/10.1177/1464884917693864

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. The MIT Press.

Dijck, J. van. (2014). Data cation, dataism and dataveillance: Big data between scienti c paradigm and ideology.

Surveillance & Society, *12*(2), 197–208. doi.org/10.24908/ss.v12i2.4776

English, J. F. (2002). Winning the culture game: Prizes, awards, and the rules of art. *New Literary History*, *33*(1), 109–135. www.jstor.org/stable/20057712

Gray, J. (2018, August 8). *New project:* What can citizen-generated data do? Research collaboration around UN Sustainable Development Goals. Jonathan Gray. jonathangray.org/2018/08/08/what-can-citizen-generated-datado/

Knight, M. (2015). Data journalism in the UK: A preliminary analysis of form and content. *Journal of Media Practice*, *16*, 55–72. doi.org/10.1080/14682753

Krippendorff, K. (2016). Data. In K. B. Jensen & R. T. Craig (Eds.), *The international encyclopedia of communication theory and philosophy, vol. 1: A–D* (pp. 484–489). Wiley Blackwell. doi.org/10.1002/9781118766804.wbiect104

Loosen, W. (2018). *Four forms of data ed journalism. Journalism's response to the data cation of society*.

(Communicative Figurations Working Paper No. 18).
www.ko .unibremen.de/
leadmin/user_upload/Arbeitspapiere/CoFi_EWP_No-
18_Loosen.pdf

Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2020). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*, *21*(9), 1246–1263.
doi.org/10.1177/1464884917735691

Parasie, S. (2015). Data-driven revelation? *Digital Journalism*, *3*(3), 364–380. doi.org/10.1080/21670811.2014.976408

Royal, C., & Blasingame, D. (2015). Data journalism: An explication. *#ISOJ*, *5*(1), 24–46. isojjournal.wordpress.com/2015/04/15/data-journalism-an-explication/

Tabary, C., Provost, A.-M., & Trottier, A. (2016). Data journalism's actors, practices and skills: A case study from Quebec. *Journalism: Theory, Practice & Criticism*,*17*(1), 66–84. doi.org/10.1177/1464884915593245

Wahl-Jorgensen, K. (2013). The strategic ritual of emotionality: A case study of Pulitzer Prize-winning articles.

Journalism: Theory, Practice & Criticism, *14*(1),129–145. doi.org/10.1177/1464884912448918

Young, M. L., Hermida, A., & Fulda, J. (2018). What makes for great data journalism? A content analysis of Data

Journalism Awards nalists 2012–2015. *Journalism Practice*, *12*(1), 115–135.

doi.org/10.1080/17512786.2016.1270171

クリックとシェアを乗り越えて:データジャーナリズムのプロジェクトのインパクトを測定する方法と意義

文:リンゼイ・グリーン＝バーバー

概要

データジャーナリズムが個人、ネットワーク、制度にインパクトを与えるユニークな立場にあることを論じ，そのインパクトを測定するための戦略を提案します。

キーワード:インパクト、社会科学、インパクト測定、アナリティクス、データジャーナリズム、オーディエンス・エンゲージメント

ジャーナリズムとインパクト

多くのジャーナリストはジャーナリズムが与えるインパクトという概念を嫌がりますが、現代の職業ジャーナリズムは実際のところ、インパクトの上に成り立っているのです。(それは)公衆が市民として活動し、権力者の責任を追及できるように情報を提供することです。ジャーナリストは、自分たちの仕事がもたらすポジティブな(あるいはネガティブな)影響について考え、話し、戦略を立て、測定することは、ジャーナリズムから提言活動へのレッドラインを踏み越えるのではないかと心配していますが、専門家やコメンテーターも同様に、「フェイクニュース」、誤報、党派報道が個人、社会、民主主義に及ぼす悪影響について多くのコラムの紙幅とピクセルを使って痛烈に訴えています。言い換えれば、ジャーナリストは自分たちの仕事が及ぼす影響について語ることを避けたい一方で、「フェイクニュース」が社会的、政治的、文化的に深刻な影響を及ぼすことは認識しているのです。しかも、19世紀後半から20世紀初頭にかけてジャーナリズムが専門化する以前は、ジャーナリズムは影響力のある活動で、政党に支えられ、政党を支持しその候補者を確実に当選させるという明確な目標を持って制作されていました(Pitt & Green-Barber, 2017)。したがって、歴史的な観点から見ると、ジャーナリズムの専門化と(中立性の)神話は、実は極めて新しいものなのです(Groseclose & Milyo, 2005; Hamilton, 2004)。そして、ジャーナリズムが「中立性」を追求したのは、規範的な決定ではなく、経済モデルの変化と、収益を上げるために可能な限り多くの読者にアピールする必要があったからでした(Hamilton, 2004)。米国や西欧では、報道業界のビジネスモデルとメディアに対する国民の信頼の欠如が同時に、かつ密接に関連していることを考えると、ジャーナリズムがその影響を認識することから目を背けてきたことは、よく言っても責任放棄、悪く言えば失敗だったかもしれません。

しかし、希望の兆しもあります。近年、一部の報道機関が、自分たちが社会に影響を与える存在であることを認め始めているのです。使命感にあふれた慈善財団や個人の支援を受けた非営利のメディアが急増し、

影響力（を把握する）実験のためのペトリ皿が作られたのです。多くの商業メディアも、自分たちの仕事がもたらすポジティブなインパクトを視聴者に伝えることは、信頼とロイヤリティを築くための戦略であり、うまくいけば増益につながると見なしています。例えば、2017 年、ワシントン・ポストは「Democracy Dies in Darkness」をマストヘッドに加え、政治システムにおけるその役割を受け入れました（そして宣伝しました）。また、CNN はウェブサイトに「Impact Your World」セクションを設け、世界の出来事、その報道、「インパクト」のあるストーリー、視聴者がハッシュタグキャンペーンから寄付まで、行動を起こすための経路を開きました。1メディア企業も、自分たちの仕事のポジティブなインパクトを最大化するために新しい戦略を試みるとともに、その効果を把握するために、広告とは異なる調査手法や測定基準を用い始めています。デジタル指標はインパクト測定の有用なプロキシとなることもありますが、ユニークページビューなどの広告指標、あるいはページ滞在時間などのより高度な分析指標は、そのコンテンツが個人に与える影響を考慮するためではなく、コンテンツの到達度を測定するためのものです。介入がもたらす現状の変化、すなわち個人、ネットワーク、制度、公論の4類型からなるメディアインパクトの枠組みはどうでしょう。これらの型は相互に関連しています。ジャーナリズムがしばしば想定するように、報道によって個人の問題に対する知識レベルが向上し、その結果、投票を通じて制度に影響を与えることができます。報道が解雇やリストラなど、組織に直接的な影響を与え、それが個人へと波及する場合もあります。 しかし、ジャーナリズムによってもたらされるインパクトは、多くの場合、時間がかかり、複雑な社会的プロセスを伴うものです。メディア企業も、自分たちの仕事のポジティブなインパクトを最大化する新しい戦略を試みるとともに、その効果を把握するために、広告とは異なる調査手法や測定基準を使い始めています。

デジタル指標はインパクト測定に役立ちますが、ユニークページビューなどの広告指標、あるいはページ滞在時間などのより高度な分析指標は、そのコンテンツが個人に与える影響を考慮することなくコンテンツの到達度を測定するためのものです。ジャーナリズムの種類によってインパクトの種類は異なります。例えば、ジェームズ・T・ハミルトンは、調査報道が、不正や汚職、悪事を暴き、変化を促すことで、機関のコストを削減できることを示しています。そして、ドキュメンタリー映画は、変化を促進するための新たな、または強化されたアドボカシーのネットワークを生み出す上で特に効果的であることが証明されているのです（Green-Barber, 2014）。残りの部分では、データジャーナリズムとインパクトの関係を探り、データジャーナリズムがさまざまなタイプの社会変革にどのように貢献できるかを示します。そして、データジャーナリズムの効果をどのように測定するか、また、ジャーナリストや報道機関がこの情報を使って何を行うことができるか提案します。

データジャーナリズムの意義

ジャーナリストがデータジャーナリズムを採用する理由はさまざまですが、その中でも特に注目すべきものは二つあります。一つ目は、ストーリーテリングの主張を裏付ける信頼できる証拠を提供するため、二つ目は、テキストベースではなく、データとしてオーディエンスに情報を提示するためです。データジャーナリズムは、データには信頼性があり、ひいてはデータを含むジャーナリズム作品にも信頼性があるという価値判断に基づいています。静的な数字、データ、チャート、グラフ、その他のビジュアルとして情報を伝えるために使用されるデータ

記事は、基本的に選択した情報を読者に伝える直線的な形式であるという点で他のジャーナリズム形式（例：テキスト、ビデオ、オーディオ）と類似しています。インタラクティブな機能を通じて読者に提供されるデータ記事は、読者がデータと対話し、それぞれ質問を投げかけ、手元のデータから答えを探すことを想定しているという点で、ストーリーテリングの特殊な形態といえます。したがって、「ストーリー」はジャーナリズムと同様にユーザーにも依存します。このようにデータジャーナリズムのざっくりした類型であっても、四つのタイプのインパクトすべてに関与しているのです。

個人。データジャーナリズムは、変化の単位となり得る個々の視聴者に焦点を当て、視聴者がより知識を深め、ひいてはより情報に基づいた意思決定を行えるように、信頼できる情報を提供する傾向があります。データジャーナリズムは、従来の直線的なストーリーテリングの足場として、コンテンツに対する視聴者の信頼を高める一方、ニュースやデータ・インタラクティブは、データジャーナリズムが個人のレベルに影響を与える最大の可能性を提供します。データ・インタラクティブ、つまり「ニュース・ストーリーを伝える大きなインタラクティブ・データベース」では、利用者は自分で質問を作成し、データを照会して答えを探せます（Klein, 2012）。メディア企業は，データ・インタラクティブによって，視聴者が深く潜ってデータを探索し，関連する情報を見つけ，ストーリーを伝えることができると思いがちです。著者は、ある報道機関によるデータ・インタラクティブの分析で、最も成功したデータアプリ、つまり、トラフィックが多く、深く探求されたものは、他のコンテンツを含む完全な編集パッケージの一部であり、地理的にローカルまたは関連するデータを調べる機能を提供し、高いインタラクティブ性を持ち、美しくデザインされ、読み込みが速かったことを明らかにしました（Green-Barber, 2015b）。プロパブリカの Dollars for Docs はデータジャーナリズムの典型例で、大量のデータ、この場合は製薬会社や医療機器会社の医師への支払いに関するデータにアクセスし、データを構造化して、対話型のデータベースとして視聴者に提示し、個人が自ら調査を行い、行動を起こすよう促すことを目的としています2。このプロジェクトでは、視聴者に医師への支払いを検索するよう求めており、サイドバーに「患者の皆さん、行動する時です。この情報を日常生活でどのように使用したか、または使用する可能性があるかを知りたいと考えています。担当医と話しましたか？そのつもりですか。教えてください」と表示されています3。

ネットワーク。データジャーナリズムは、（公式・非公式にかかわらず）ネットワークがその立場や活動を強化するために利用できる信頼性の高い情報を提供します。たとえば、アドボカシー団体は、特にデータが公開されていない場合に、公開請求や法的手続きで主張を裏付けるためにデータ記事を使用することがよくあります。ジャーナリズムは、公的領域で入手できないデータへのアクセスを要求し、そのデータを分析し、調査結果を公表することで、個人やネットワークにとって他では乗り越えられないようなコストを吸収しているのです（Hamilton, 2016）。

教育機関。データジャーナリズムは、腐敗、不正、不正行為、無能の証拠となるような、組織が懸命に隠そうとするような記事を生み出すことができます。このような情報が明らかになると、政治家にとっては選挙に関連する脅威、上場企業にとっては市場原理による脅威から、組織に制度改革を求める圧力が発生します。例え

ば、国際調査報道ジャーナリスト連合によるパナマ文書の共同調査では、1150 万件以上の記録を分析し、「50 カ国以上の政治家が 21 のタックスヘイブンにあるオフショア企業に関係している」ことが明らかになりました4。この調査により、アイスランドのシグムンドゥル・グンロイグソン首相などの政治家が辞任し、パキスタンのナワズ・シャリフ元首相（2018 年に 10 年間の実刑判決を受けた）などの調査が行われ、その他無数の制度的対応が取られることになったのです。

公の場での議論。 データジャーナリズムは多くの場合、地理的、人口統計的、その他の要因によって小さな部分に分割できるため、さまざまなメディアがいろんなデータを使って多様なストーリーを伝えることができます。データジャーナリズムは、地理的な場所、人口統計的なグループ、その他の社会的な境界を越えて、問題についての公論を促すためにローカライズすることができます。[米国の]調査報道センターは、米国退役軍人省（VA）についての全国的なインタラクティブデータセットを公開しています。一つは退役軍人が退役軍人省（VA）の病院で医療を受けようとする際の平均待ち時間、もう一つは VA システムで退役軍人に処方されているアヘンの数です。どちらのケースでも、地元の報道機関は、問題に関する報道を行うためのベースラインとしてデータセットを使用しました。

では、データジャーナリストはどのように戦略的にインパクトを与えることができるのか

あなたは大変な仕事をこなしてきました。データにアクセスし、数字を計算し、データを構造化して、伝えるべき重要なストーリーを手に入れたのです。さて、ここからどうしましょう？データジャーナリズムのためにインパクトを高める戦略は、以下の五つのステップを踏むとよいでしょう。

目標を設定する。プロジェクトの結果、何が起こりますか？誰が、何が、不正行為に対処する権力とインセンティブを持っているのですか？あなたが明るみに出そうとしている情報には、誰がアクセスできるべきでしょうか？これらの質問を自問自答し、あなたのプロジェクトにとって妥当なインパクトの種類を決定してください。

内容。プロジェクトの目標が決まったら、その作品の重要なターゲットオーディエンスを特定します。対象となる人々は、どのようなニュースや情報を信頼していますか？どのように情報にアクセスするのが最適でしょうか。インタラクティブなコンテンツが必要ですか、それとも直線的なストーリーが効果的でしょうか？

エンゲージメント。あなたやあなたの会社は読者とどのように関わり、読者はあなたの作品にどのように関わるのですか？たとえば、ターゲットとする読者が信頼できる情報源として、自社以外の報道機関を挙げるなら、協力しましょう。また、自社のデータ・インタラクティブが NGO コミュニティにとって重要な情報を持っている場合、その使い方を説明するウェビナーを開催します。

戦略的な調査。目標、内容、および契約プランに応じて、進捗状況を追跡し、何がうまくいっていて何がうまくいっていないかを理解するために、適切な調査方法や指標を選択します。メディアは「測定」という言葉

をよく使いますが、私は「戦略的研究」という言葉を使います。質的、量的な研究方法の両方を考慮する必要があるからです。調査方法や指標を早く知ることができれば、情報の精度も上がります。（次のセクションでは、測定オプションについて詳しく説明します。）

繰り返す。データジャーナリズムのレポート作成、コンテンツ、契約、測定に時間とリソースを投資してきました。何がうまくいきましたか？今度は何を変えますか？どのような質問が未解決ですか？これらの知識をチームと共有し、次のプロジェクトをさらに進めることができます。

仕事のインパクトをどのように「測定」するのか？

先に述べたように、メディアインパクトの研究は、広告指標に支配されてきました。しかし、ページビュー、ページ滞在時間、直帰率などの広告指標は、ある種類のインパクトしか測定できません。これらの指標は、問題に対する意見、新しい情報を学んだかどうか、コンテンツに基づいて行動を起こすかどうかといったことに関係なく、個人へのコンテンツの総接触量を測定することを目的としています。しかし、コンテンツが個人、ネットワーク、組織、公論に与える影響を考える場合、他にこの分野における成功と失敗をよりよく理解するために使用できる革新的な定性的および定量的手法があります。ここでは、データジャーナリズムの影響を理解するための有望な研究手法の一端を探ってみましょう。

分析。メディアの測定基準は、認知度の向上や知識の増加など、望ましい成果の代用品として使用することができます。しかし、メディア企業は、変化をアナリティクスに帰する際に意図的かつ慎重になるべきです。例えば、データジャーナリズムのプロジェクトが制度改革を促進することを目的としている場合、ユニークページビューは成功の適切な指標とは言えません。

実験的研究。実験的研究は、介入の効果を検証するための一定の条件を作り出します。テキサス大学オースティン校の Center for Media Engagement は、ニュースのホームページのレイアウトが視聴者の記憶や感情に与える影響や、解決策を重視した報道が読者の感情に与える影響について、興味深い実験研究を行っています。テクノロジー企業は、さまざまなインタラクティブ要素がユーザーに与える影響を常に検証しています。ジャーナリズム企業も同じように、データ・インタラクティブがユーザーに与える影響をよりよく理解するために、大学との提携や、マーケティング、事業開発、視聴者参加などの分野の研究者と社内で直接連携することができます。

アンケート調査。アンケートは、最先端のリサーチ手法ではありませんが、個人の興味、知識・意見、行動の変化に関する情報を収集する方法として、実績のあるものです。アンケート調査には、再訪問を促すポップアップやニュースレターのクリックスルーを追跡する技術などを活用し、潜在的な回答者のプールを生成することで、調査設計に創造性を発揮できます。

コンテンツ分析。コンテンツ分析は、談話の時間的な変化を判断するために使用される調査手法です。この方法は、あらゆるテキストベースのコーパスに適用できるため、非常に柔軟性があります。例えば、ある組織が国民の言論に影響を与えることを目的にコンテンツを制作した場合、プロジェクト後に全国紙上位 10 紙のコンテンツ分析を行い、その記事の影響力を調べることができます。州議会に影響を与えることを目的とする場合、組織は、公に入手可能な立法議案についてプロジェクト後のコンテンツ分析を行うことができます（Green-Barber, 2015a）。アドボカシーネットワークにデータを提供することが目的であれば、組織のニュースレターを分析します。コンテンツ分析は少なくとも三つの方法で行うことができます。最も基本的なレベルでは、報道機関がプロジェクトの引用を検索し、いつ、どこで引用されたかを記録することができます。例えば、多くの記者は、あるプロジェクトが他のどの媒体で取り上げられたかを調べるために、自分の取材したキーワードと自分の苗字を使って Google News Alerts を作成しています。この方法は方法論的には適切ではありませんが、興味深い情報を提供し、インパクトについて直感的に確認するために使用することができます。また、このプロセスによって、プロジェクトのインパクトについて、さらに深く掘り下げる価値のある疑問が生まれるかもしれません。多くの組織では、Google News Alerts や Meltwater などのニュースクリッピングサービスをこの目的に使用しています。厳密なコンテンツ分析をするなら、プロジェクトにおけるキーワード、データ、フレーズを特定し、有限のテキストコーパスの中で公開前後の普及状況を分析し、変化を記録することができます。コンピュータによるテキスト分析では、さらに一歩進んで、高度な集計・分析技術によって談話の変化を推測します。このようなより厳密な分析は、報道機関が訓練を受けた研究者と提携する必要があるでしょう。

展望：データジャーナリズムのインパクトにジャーナリストが注目すべき理由

報道が重要であり続けるためには、ジャーナリズムが社会に影響を与える事実を積極的に受け入れる必要があります。ジャーナリズムが機能する変化のエコシステムと、このシステム内での特定の役割を理解することで、業界はそのポジティブな影響を最大化し、読者にその価値を示すことができます。データジャーナリストは、定量的データと定性的データの両方の価値と重要性を理解しており、この取り組みに適しているといえます。データジャーナリズム・プロジェクトの目標を明確にし、創造的な視聴者のエンゲージメントと配信戦略を開発し、これらのプロジェクトの成功を測定する洗練された方法を構築することで、記者は内部からこの動きをリードすることができます。

脚注

1. edition.cnn.com/specials/impact-your-world

2. projects.propublica.org/docdollars/

3. propublica.forms.fm/was-the-dollars-for-docs-information-helpful-tell-us-how-you-will-use-it/forms/2249

4. www.icij.org/investigations/panama-papers/

参考文献

Green-Barber, L. (2014). Waves of change: The case of rape in the elds. The Center for Investigative Reporting.

www.documentcloud.org/documents/1278731-waves-of-change-the-case-of-rape-in-the- elds.html

Green-Barber, L. (2015a). Changing the conversation: The VA backlog. The Center for Investigative Reporting.

s3.amazonaws.com/uploads-cironline-org/uploaded/uploads/VA+Backlog+White+Paper+11.10.14.pdf Green-Barber, L. (2015b). What makes a news interactive successful? Preliminary lessons from the Center for

Investigative Reporting. The Center for Investigative Reporting. s3-us-west-2.amazonaws.com/revealnews.org/uploads/CIR+News+Interactives+White+Paper.pdf

Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, *120*(4), 1191–1237. doi.org/10.1162/003355305775097542

Hamilton, J. T. (2004). *All the news that's t to sell: How the market transforms information into news*. Princeton University Press.

Hamilton, J. T. (2016). *Democracy's detectives: The economics of investigative journalism*. Harvard University Press.

Klein, S. (2012). News apps at ProPublica. In J. Gray, L. Chambers, & L. Bounegru (Eds.), *The data journalism handbook: How journalists can use data to improve the news* (pp. 185–186). O'Reilly Media.

Pitt, F., & Green-Barber, L. (2017). *The case for media impact: A case study of ICIJ's radical collaboration strategy*.

Tow Center for Digital Journalism. doi.org/10.7916/D85Q532V

データジャーナリズム：誰の利益になるのか？

文：メアリー・リン・ヤング、キャンディス・キャリソン

概要

データジャーナリズムのプロジェクトが誰の利益になるのかを問い、特に公益を目的とした最近の犯罪関連のデータジャーナリズムに関して、想定される視聴者に疑問を投げかけます。先住民族学の学者たちの研究を紹介し、拒否、誤った表現、植民地主義、データ収集がジャーナリズムの根強い課題であり、より良い倫理的判断が必要であることを示します。

キーワード：植民地主義、先住民族、データジャーナリズム、犯罪コンテンツ、倫理、科学技術研究

米国における初期のデータジャーナリズムの重要な作品の一つに、犯罪統計を重ねたシカゴのオンライン地図である chicagocrime.org があります（Anderson, 2018; Holovaty, 2005, 2008）。創設者のエイドリアン・ホロバティによれば、2005 年に開設された chicagocrime.org は、シカゴ警察の犯罪データとグーグルマップを組み合わせた、マッシュアップ地図の元祖でした。地区ごとのページと RSS フィードを提供し、犯罪データを種類別、場所のタイプ別（歩道やアパートなど）、郵便番号別、通りや住所別、日付別、さらには任意のルートで閲覧することができます（Holovaty, 2008）1。数年後、ロサンゼルス・タイムズはブログ Homicide Report を立ち上げ、警察のデータを利用して郡内の 900 件以上の殺人事件それぞれについての記事を作成しました。いずれのプロジェクトも、米国の主要都市の犯罪データと地理情報を利用しています。そしてどちらも、データ駆動型ジャーナリズムやジャーナリズム全般の目的と影響に関連する、根強い批判や課題に対する洞察を提供します。

ホロバティが chicagocrime.org を立ち上げた動機は、ジャーナリズムがますます技術的なアイデンティティを高め、「クールな技術」に焦点を当てるとともに、地域の「使えるニュース」を生成するという目標と一致していました。Homicide Report の創設者であるロサンゼルス・タイムズのジャーナリスト、ジル・レオヴィの目標は、より批判的なものでした。レオヴィは、特定の殺人事件しか取り上げないという伝統的なジャーナリズムの規範と慣習を脱構築するために、ロサンゼルス郡のすべての殺人事件を説明したいと考えたのです（Leovy, 2008; Young & Hermida, 2015）。2015 年のナショナル・パブリック・ラジオの Fresh Air とのインタビューで、レオヴィは Homicide Report を立ち上げた動機が、ニュースにおける構造的偏向への対応と、犯罪に関する新聞報道が「現実とかけ離れている」ことへの不満にあったと述べています。

新聞の仕事は異常な出来事を取り上げることですが、殺人事件に関しては、常にベルカーブの非常に低い部分を取り上げることになります。もちろん、殺人事件が日常的であることはないのですが、中央の膨らみは普通の殺人事件とされ、決してカバーできません。このような殺人は、特に米国の都市では、同じ形で、同じ

方法で、あまりにも長い間続いてきたため、都市生活の背景のようなものとなってしまいました。それらは当たり前のことであり、新聞社にとって有効な物語やストーリーにすることは非常に難しいのです（Leovy, 2015）。

レオヴィ（2008）は、犯罪ジャーナリストとしての経験と、デジタル・ジャーナリズムという（新聞の一面に比べて）果てしなく階層が少ない空間、そして公開データへのアクセスを組み合わせることで、「ほとんどが若いラテン系で、最も偏りが出ているのは若い黒人男性の」郡内のすべての殺人に関する情報を、できるだけ等価に表現するニュースを構想しました（Leovy, 2015）。反響は強烈で、「（我々）読者は愕然とした。なんてことだ」と、最初の投稿がありました。「その多さは衝撃的だ」と別の人がコメントし、「まるで使い捨ての人間みたい」と 3 人目が書き込みました。これらのデータジャーナリズムの事例は、拡大するサブスペシャリティの新しい表現として、その革新性が評価され、称賛を受けました。ホロバティによると、chicagocrime.org は、ニューヨーク近代美術館の展示品となっています。しかし、これらの代表的なプロジェクトや、公益のためにデータを共有すると称する他のプロジェクトは、誰の利益となり、誰が想像上の観客だったのかという疑問は、ほとんど語られないままになってきました。

科学技術研究（STS）の学者たちは、テクノロジーが新奇で変革的である一方、弱者とある種のデータの間の有害な関係が持続していることを繰り返し証明してきました。人種の社会的構成や白人至上主義の再現にテクノロジーが果たす役割についての広範な批判（Benjamin, 2019; Noble, 2018）がなされているにもかかわらず、データジャーナリズムの実証主義的志向（Coddington, 2019）も（それに）関与していることが暗示されています。さらに、ジャーナリズムの表象、規範、慣行、経済、犯罪ニュースに関する研究は、人種差別、社会的統制、被害、進行中の植民地主義の長い歴史を反映しているのです。ここでは、どのような（データ）構造が支持され、誰のデータが集められやすいのか、あるいは集められないのかを説明し、データジャーナリズムを採用するかどうか、どのように採用するかを決める際に、ジャーナリストが「拒否の倫理」を組み込む必要性があると提唱します。バトラーが主張するように、「脆弱性を配分する方法、ある集団を他の集団よりも恣意的な暴力の対象とする差動的な配分形態が存在する」（Butler, 2004）のです。

我々は、データジャーナリズムが、データジャーナリズム、計算ジャーナリズム、コンピュータ支援報道の三つの形態からなる量的ジャーナリズムとするコディントンの定義（2014, 2019） を引用します。研究やデータ収集の理論的根拠、新技術、社会的問題をまとめる科学的実践や社会的制度に対する根強い批判は、データジャーナリズムの 3 分野すべてに関連しており、脆弱性に関する疑問や、誰の関心が支持されているかについても同様です。STS と先住民研究の研究者タルベアは、アメリカの先住民族のゲノム研究を調査し、「祖先の移動とつながりについて知っていること」という概念に関連するステレオタイプや植民地時代の物語の多くが、潜在的なアイデンティティや健康上の利益に関する記述に加えて、研究の根拠の一部になっていることを発見しました。彼女は、科学の主流における語彙では、遺伝的つながりという概念が人種階層の概念に取って代わったかもしれないが、権力、差異、階層の関係は、私たちのより広範な文化、私たちの制度や構造、科学が行われ、科学の基盤となる文化（TallBear, 2007, p413）にとって不可欠であることを指摘しています。タルベアは、機械

的あるいは実験室ベースの客観性という科学的概念の根底には、制度的特権、コミュニティとの歴史的・継続的関係、研究の根拠や意図された利益を明示する文化的枠組みがあると主張します。「誰のために、なぜ」という視点は常に必要とされています。

そして場合によっては、研究やデータマイニングによる分析は、意味づけのプロセスが固定化された人種、ジェンダー、階級といった概念を前提にしているため、拒否してもよいのです。タルベアが「科学に影響を与え続けている植民地的な仮定と実践」と呼ぶものは、同様にジャーナリズム、ひいてはデータジャーナリズムにも影響を与えます。

先住民はまた、植民地主義のさまざまな形態と歴史のために、人類学と政府の膨大なアーカイブと一貫したメディアの誤った表現とステレオタイプ（Anderson & Robertson, 2011）との戦いにさらされてきました（Tuck & Yang, 2012; Wolfe, 2006）。それだけに、機械による客観性の概念の延長線上にあるデータジャーナリズムの利害関係は深いのです。シンプソン（2014）の人類学批判では、[アメリカ先住民族の]モホークのコミュニティは、こうしたアーカイブと格闘し、入植者植民地の制度や枠組みを中心とした研究に参加する際に、定期的に拒否を表明していることが紹介されています。 シンプソンの枠組みにおける拒否は、多方向的なものです。排除されることへの拒否、自分のアイデンティティ、文化、土地に関する誤った表現を内面化することへの拒否、そして、国家や他者の（この場合、メディアによる）承認が、そうした人々が考える差異についての期待に対する迎合となってしまうことへの拒否なのです。先住民学者によるこのような議論は、歴史と権力の問題を中心に据えているので、データジャーナリズムの意図、理論的根拠、実践に直接的な挑戦を投げかけています。こうした問いは、国家だけでなく、国家の目的と目標、家父長制や入植者植民地主義、白人至上主義といった構造やイデオロギーを支える社会秩序を維持するためのジャーナリズムの役割にも関わります（Callison & Young, 2020）。

先住民コミュニティにとって、データと正確なメディアの表現の両方を見つけることがほとんど常に困難であることと、データが収集、保管、アクセスされる制度の反映であるという事実が事態をさらに複雑にしています。警察統計は犯罪の社会的現実を反映しておらず、「文化、法律、社会構造が組織の目的のために生み出されている」（Ericson 1998, p. 93）という批判は、データ論争だけに焦点を当てるジャーナリストにも関係します。たとえば、ラグナ・プエブロのジャーナリストであるジェニ・モネは、アメリカにおける先住民コミュニティを「アスタリスク国家」と呼んでいますが、これはデータが存在しない人々を意味しています。多くのソーシャルデータチャートでは、アラスカ先住民のデータが存在しないというアスタリスクが表示されているのです。

Facebook のようなデジタルメディアは、先住民のオーディエンスや彼らの関心事に関わり、構造的な不平等やデータのギャップに対処する有意義で正確な表現を生み出すためのジャーナリズムのツールと見なされるかもしれない、有望な代替プラットフォームを提供します（Monet, 2018）。その場合も、参加するかどうか、どのように参加するかという問題は、誰が利益を得るか、データ収集にどのようなプロセスが利用されるか、誰の意味づけプロセスが優位に立つかをめぐって展開されるのです。

ジャーナリズムにとって、広義の意味形成プロセスは、実証主義的な志向の中で、反対、逸脱、対立、あるいは「規範から外れた物や人の行動」(Ericson, 1998, p.84)の問題に結びつけがちです。社会秩序におけるジャーナリズムの役割は、逸脱者として構成された集団に物質的な影響と有害な効果を与えてきましたし、今もそうです(Callison & Young, 2020; Rhodes, 1993; Stabile, 2006)。スタビルは、新聞、テレビの犯罪報道、ラジオ番組、犯罪報道と人種との関係など、アメリカにおける犯罪ニュースの歴史的研究を行い、白人至上主義と営利目的のジャーナリズムというイデオロギー文脈のなかで、逸脱規範が構造的に脆弱な集団に与える影響を明示しました。彼女は、人種とジェンダーに焦点を当て、「逸脱に割り当てられた歴史的意味をめぐる闘争のもっとも重要な場である」(Stabile, 2006, p.5)ことから、メディアが国家と警察などのエージェントによる黒人男性の「犯罪化の過程」を支えていると主張しています。

例えば、カージャックなどの特定の犯罪、犯罪者、被害者に焦点を当てることで、メディアがいかに警察のデータ収集を増幅させ、強化しているでしょうか。彼女は、「アメリカで栄えた貪欲で暴力的な白人社会では、これまでずっと白人テロの虚構が白人テロの物質性に置き換えられてきた」(Stabile, 2006, p.2)ことを見出しています。

ここで、敵や味方を生み出すものとしてジャーナリズムを分析したキャリー (1974)は、国家のジェノサイドや継続的な植民地主義を含む、資本主義や北米の国家に対するジャーナリストの制度的な関係にも関連していると理解される余地があります。事実と知識は社会的に構築されるという概念に対するジャーナリズムのアレルギーと相まって、ジャーナリズムと特にニュースは、社会秩序の言説が共同生成され、複製され、また潜在的に変容するための結合組織となっているのです。

これらの重要な点について、ジャーナリズム、犯罪学、STS、その他の分野の文献は、データジャーナリズムが十分に対処できていない差し迫った一連の懸念に警鐘を鳴らしています。学者たちは、その幅広い効果や結果よりも、類型化(Coddington, 2019)、データジャーナリズムの状況(Heravi, 2017)、データジャーナリズムがより広いジャーナリズムの認識論、文化、実践、アイデンティティに及ぼす影響(Anderson, 2018; Borges-Rey, 2020; Gynnild, 2014; Lewis & Usher, 2014; Young & Hermida, 2015)により時間を割いてきました。英国におけるデータジャーナリズムの成長について横断的に政治経済学的分析を行ったボルヘス=レイ (2016, 2020)を除いて、権力に関連する問題を提起した学者はほとんどいません。しかし、ホロバティのこの発言のように、データジャーナリズムと権力との影響を指摘することができます。chicagocrime.org からは、多くの良いことが生まれました。地域レベルでは、数え切れないほどのシカゴ市民が、この公共サービスに対して感謝の意を表して連絡してきました。市民団体は、このサイトのプリントアウトを警察の巡回会議に持参し、熱心な市民は市会議員にこのサイトのレポートを持って行き、市が明るい街灯の設置を検討すべき危険な交差点を指摘しました(Holovaty, 2008)。この場合、市民団体はデータをもとに、自分なりの意味と行動の根拠を作り上げました。

しかし、権力やメディアの中心地から遠く離れた地方や、すでに不当に監視されているコミュニティ、そしてカナダでもアメリカでも依然として白人が大半を占めるニュースルームにコミュニティが十分に代表されていないケースなど、より大きなスケールでこれがどう機能するかは、より幅広い倫理診断が必要です。

これらの事例やジャーナリズム研究以外の批評的文献から得られた証拠を考えると、「使えるニュース」や技術偏重の実験より、潜在的な被害が優先される余地があり、またそうすべきなのです。ジャーナリストの犯罪ニュースの取り上げ方をデータの観点から見ると、ジャーナリズム（業界）内部だけの意図、成功、スクープを（優先して）考慮する問題点だけではなく、結果についても深い理解が必要となります4。倫理的診断は、拒否の概念、ジャーナリズムによる誤った表現と植民地主義への奉仕の長い歴史、意味づけとデータ収集が起こる不均一なプロセスをより適切に説明する必要があります。データジャーナリズムがなしうる貢献についてジャーナリストが考えるに当たっては、誰のために、なぜ、ということが本質的な問いになってくるでしょう。

脚注

1. これは、2008 年にホロバティが立ち上げ、09 年に MSNBC.com に買収された、誉れ高いコミュニティデータジャーナリズムサイト、EveryBlock の初期版である（Holovaty, A.（2013, 7 February）. RIP EveryBlock. Adrian Holovaty. www.holovaty.com/writing/rip-everyblock
2. その後、アルゴリズム・ジャーナリズムのブログとして再構築された。
3. これらの問題についてはククタイとウォルターの章を参照。
4. ルースンの章を参照。

参考文献

Anderson, C. W.（2018）. *Apostles of certainty: Data journalism and the politics of doubt*. Oxford University Press.

Anderson, M. C., & Robertson, C. L.（2011）. *Seeing red: A history of natives in Canadian newspapers*. University of Manitoba Press.

Benjamin, R.（2019）. *Race after technology: Abolitionist tools for the New Jim Code*. Polity Press.

Borges-Rey, E.（2016）. Unravelling data journalism: A study of data journalism practice in British newsrooms.

Journalism Practice, *10*(7), 833–843. doi.org/10.1080/17512786.2016.1159921

Borges-Rey, E. (2020). Towards an epistemology of data journalism in the devolved nations of the United Kingdom:

Changes and continuities in materiality, performativity and re exivity: *Journalism*, *21*(7), 915–932. doi.org/10.1177/1464884917693864

Butler, J. (2004). *Precarious life: The powers of mourning and violence*. Verso.

Callison, C., & Young, M. L. (2020). *Reckoning: Journalism's limits and possibilities*. Oxford University Press.

Carey, J. W. (1974). The problem of journalism history. *Journalism History*, *1*(1), 3–27. doi.org/10.1080/00947679.1974.12066714

Coddington, M. (2014). Clarifying journalism's quantitative turn. *Digital Journalism*,*3*, 331–348. doi.org/10.1080/21670811.2014.976400

Coddington, M. (2019). De ning and mapping data journalism and computationaljournalism: A review of typologies and themes. In S. Eldridge & B. Franklin (Eds.), *The Routledge handbook of developments in digital journalism studies* (pp. 225–236). Routledge. doi.org/10.4324/9781315270449-18

Dumit, J., & O'Connor, K. (2016). The senses and sciences of fascia: A practice as research investigation. In L. Hunter,

E. Krimmer, & P. Lichtenfels (Eds.), *Sentient performativities of embodiment: Thinking alongside the human* (pp. 35– 54). Rowman & Littlef ield.

Ericson, R. (1998). How journalists visualize fact. *The Annals of the American Academy of Political and Social*

Science, *560*, 83–95. doi.org/10.1177/0002716298560001007

Ericson, R., Baranek, P. M., & Chan, J. B. L. (1991). *Representing order: Crime, law, and justice in the news media*. University of Toronto Press.

Gynnild, A. (2014). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism*, *15*(6), 713–730. doi.org/10.1177/1464884913486393

Hamilton, J. T. (2000). *Channeling violence: The economic market for violent television programming*. Princeton University Press.

Heravi, B. (2017, August 1). State of data journalism globally: First insights into the global data journalism survey. *Medium*. medium.com/ucd-ischool/state-of-data-journalism-globally-cb2f4696ad3d

Holovaty, A. (2005, May 18). Announcing chicagocrime.org. Adrian Holovaty. www.holovaty.com/writing/chicagocrime.org-launch/

Holovaty, A. (2008, January 31). In memory of chicagocrime.org. Adrian Holovaty. www.holovaty.com/writing/chicagocrime.org-tribute/

Leovy, J. (2008, February 4). Unlimited space for untold sorrow. *Los Angeles Times*. www.latimes.com/archives/laxpm-2008-feb-04-me-homicide4-story.html

Leovy, J. (2015, January 25). "Ghettoside" explores why murders are invis- ible in Los Angeles [interview with Dave Davies]. *Fresh Air*. www.npr.org/2015/01/26/381589023/ghettoside-explores-why-murders-are-invisible-in-losangeles?t=1617534739128

Lewis, S. C., & Usher, N. (2014). Code, collaboration, and the future of journalism: A case study of the hacks/hackers global network. *Digital Journalism*, *2*(3),383–393. doi.org/10.1080/21670811.2014.895504

Monet, J. (2020, October 30). Native American voters could decide key Senate races while battling intense voter suppression. *Democracy Now!* www.democracynow.org/2020/10/30/jenni_monet_indigenous_sovereignty_2020

Monet, J. (2018b, March 23). #DeleteFacebook? Not in Indian Country. *Yes! Magazine*. www.yesmagazine.org/socialjustice/2018/03/23/deletefacebook-not-in-indian-country

Nelson, A. (2016). *The social life of DNA: Race, reparations, and reconciliation after the genome*. Beacon Press Noble, S. (2018). *Algorithms of oppression*. NYU Press. Reardon, J. (2009). *Race to the nish: Identity and governance in an age of genomics*. Princeton University Press.

Rhodes, J. (1993). The visibility of race and media history. *Critical Studies in MassCommunication*, *10*(2), 184–190. doi.org/10.1080/15295039309366859

Schudson, M. (2005). Autonomy from what? In R. Benson & E. Neveu (Eds.), *Bourdieu and the journalistic eld* (pp.

214–223). Polity Press.

Simpson, A. (2014). *Mohawk interruptus: Political life across the borders of settler states*. Duke University Press.

Simpson, A. (2016). The state is a man: Theresa Spence, Loretta Saunders and the gender of settler sovereignty. *Theory & Event*, *19*(4). muse.jhu.edu/article/633280

Stabile, C. (2006). *White victims, Black villains: Gender, race and crime news in USculture*. Routledge.

TallBear, K. (2007). Narratives of race and indigeneity in the genographic project. *The Journal of Law, Medicine &*

Ethics, *35*(3), 412–424. doi.org/10.1111/j.1748-720X.2007.00164.x

TallBear, K. (2013). *Native American DNA: Tribal belonging and the false promise of genetic science*. University of Minnesota Press.

Tuck, E., & Yang, K. (2012). Decolonization is not a metaphor. *Decolonization: Indigeneity, Education & Society*, *1*(1), 1–40.

Wolfe, P. (2006). Settler colonialism and the elimination of the native. *Journal of Genocide Research*, *8*(4), 387–409. doi.org/10.1080/14623520601056240

Young, M. L., & Hermida, A. (2015). From Mr. and Mrs. Outlier to central tendencies. *Digital Journalism*, *3*(3), 381–397.

doi.org/10.1080/21670811.2014.976409

インパクトのあるデータジャーナリズム

文：ポール・ブラッドショー

概要

インパクトのあるデータジャーナリズム：インパクトの測定方法と理由、インパクトの変化、そしてインパクトを形成する要因について。

キーワード：インパクト、エンゲージメント、データジャーナリズム、アナリティクス、調査報道、データ品質

ボストン・グローブが児童虐待をめぐる［教会］組織の沈黙を調査した映画『スポットライト』（2015）をまだ見ていないなら、今すぐ見てください。さらにいえば、最後に出てくるタイトルカードまでしっかり見るべきです1。クリーンの下にリストが表示されは、ナイジェリアのアクートからオーストラリアのウーロンゴンまで、この映画以降に虐待が発覚した何十もの場所が詳述されます。しかし、このタイトルカードは、快哉の気持ちを押しとどめるものでもあります。スキャンダルに関わった重要人物の一人が「世界で最も高い地位にあるローマカトリック教会の一つ」に配置換えになったと書かれているのです。

これこそ、データジャーナリズムにおけるインパクトの課題です。問題の認知度を高めることが「インパクト」なのか？そのストーリーは、罰や報酬をもたらす必要があるのだろうか？目に見える政策変更か？インパクトはどの程度重要なのか？そして、誰にとって？この最後の二つの疑問は、最初に取り組む価値があります。従来、インパクトが重要視されてきたのは、主に二つの理由からでした。商業的な理由と文化的な理由です。商業的には、ブランドの認知度や高い視聴者数といったインパクトの尺度は、広告（価格と量の両方を上げる）や購読・コピー販売を通じて出版社の利益率に直接貢献することができます（Rusbridger, 2018）。一方、文化的には、インパクトのあるストーリーは、報道機関や個々のジャーナリストに仲間内で「自慢できる権利」を与えてもきました。両者は、これから見るように、より複雑になってきています。ジャーナリズムにおけるインパクトの測定は、歴史的にみて限られています。利用できるのは、売上や視聴者数の集計、業界賞の限られたプール、そして時折行われる視聴者調査だけでした。

もちろん、現在では、測定基準の急増だけでなく、ビジネスモデルの急増も課題となっています。特に、非営利のニュース提供の拡大により、インパクトがますます重視され、その測定方法に関する議論が高まっています（Schlemmer、2016 年）。さらに、ストーリー単位でインパクトを測定できるようになったことで、視聴者へのインパクトに責任を負うのは、もはや編集者ではなく、ジャーナリストにもなっています。

数字で測るインパクトの大きさ

インパクトの最も簡単な尺度は、おそらくリーチ数でしょう。BBC の「70 億人の人々とあなた:あなたの番号は?」のようなデータ駆動型のインタラクティブコンテンツは、何百万人もの読者を話題の記事に引き込みました。2012 年のある時点で、ネイト・シルバーのデータジャーナリズムは、ニューヨーク・タイムズの閲覧数の5分の1に一人に達していました(Byers、2012)。このような粗雑な指標を嘲笑する人もいるかもしれませんが、これは重要なことです。かつて、ジャーナリストが読者をないがしろにして仲間に感銘を与えようとしていると批判されたことがあったとしても、現代のジャーナリズムは、少なくとも、読者とつながりを持てることを証明することが期待されています。ほとんどの場合、この証明は広告主のために必要なのですが、BBC のような公的資金によるユニバーサル・ニュース・プロバイダーでさえ、資金調達の要件を満たしていることを証明するために必要なのです。

エンゲージメントは、リーチのより高度な関係であり、ここでもデータジャーナリズムはうまく機能しています。例えば、ある新聞社の編集者会議では、ページにデータ・ビジュアライゼーションを追加するだけで、滞在時間が 3 分の 1 増えることが明らかにされました。データ駆動型のインタラクティビティは、退屈なテーマを一変させます。

同社のデヴィッド・ハイガソンは、データチームが貧困の統計に基づいたインタラクティブなウィジェットに 20 万人以上の人々が自分の郵便番号を入力したことを指摘し、「単純な『データが教えてくれる系』のストーリーに想像がつかないほどの」(Higgerson、2015)数の人が参加したと述べている。エンゲージメントは、広告に依存する組織にとって特に重要です(エンゲージメントの高いところでは料金を上げることができる)が、購読、寄付、イベントが重要な組織にとっても重要なのです。これらもエンゲージメントと結びつく傾向があるからです。非営利団体の資金提供や助成金の拡大には、単にリーチするだけでなく、インパクトをモニタリングしたり、実証したりすることが明確に求められている場合が多く、特に、政治的、法的な変化や行動がしばしば言及されます。例えば、国際調査報道ジャーナリスト連合 (ICIJ) は、パナマ文書の調査報道を、ピュリッツァー賞を含む約 20 の賞を受賞するという、より伝統的な評価基準と並んで、「79か国で少なくとも 150 件の照会、監査または調査につながった」と「79か国」の部分を強調しています(Fitzgibbon & Díaz-Struck, 2016; ”ICIJ’s Awards,” n.d.)。

英国では、データジャーナリズムの歴史において、国会議員の経費スキャンダルが特別な位置を占めています。この問題を巡っては、テレグラフが数週間にわたってニュースアジェンダをリードしただけでなく、新しい組織の設立につながりました。IPSA(Independent Parliamentary Standards Authority:独立議会基準局)です。この団体は現在、政治家の経費請求に関するオープンデータを公表しており、政治家の説明責任を果たすことを可能にし、さらなるデータジャーナリズムにつながるものとなっています。

しかし、政策は政治よりもはるかに広い範囲に及ぶことがあります。銀行の融資政策は何百万人もの人々に影響を与えます。1980 年代後半、米国ではビル・デドマンがピュリッツァー賞を受賞した一連の記事「Colour of Money」でその責任を追及したことは有名です。データに基づく調査は、人種差別的な融資慣行

(「レッドライニング」)を明らかにし、調査、新たな融資、訴訟、新しい法律の制定など、政治、金融、法律の変化にもつながりました3。それから 30 年後でも、このアプローチの現代版を見ることができます。プロパブリカの「Machine Bias」シリーズは、アルゴリズムによる説明責任に光を当て、Bureau Local はそのネットワークを利用して、ソーシャルメディア上のアルゴリズムでターゲット化された「ダーク広告」に関する情報をクラウドソーシングしています(McClenaghan, 2017)。

両者とも、Facebook の多くのポリシーの変更に貢献し、プロパブリカの手法は、住宅機会を均等にするために活動する団体が、ソーシャルネットワークに対する訴訟の基礎を確立する中で採用しました(Angwin & Tobin, 2018; "Improving Enforcement and Promoting Diversity," 2017; Jalonick, 2017)。警察の配置に影響を与えることから、非白人地域におけるウーバーの価格設定まで、アルゴリズムの政策が私たちの生活においてますます強力になるにつれ、これらの責任を問うことは、より伝統的な政治的形態の権力の責任を問うことと同様に重要になっています(Chammah, 2016; Stark, 2016)。これらの事例のいくつかで注目すべき点は、その影響が他者とのコラボレーションに依存しており、それによって実証されていることです。例えば、Bureau Local がインパクトについて語るとき、彼らは草の根ネットワークのメンバーによって生み出されたストーリーの数に言及し、他の人々の行動を鼓舞しています。一方、ICIJ はそのネットワークの規模の拡大について、「LuxLeaks (2014) は、26 カ国で 80 人以上の記者を関与させた。Swiss Leaks (2015) 45 カ国で 140 人以上の記者が参加した」(Cabra, 2017)と述べています。パナマ文書の調査では、約 80 カ国、370 人以上の記者に増え、100 の報道機関が 4700 の記事を掲載しました(Blau, 2016)。

さらに、調査の結果、収集・公開されたデータがインパクトの源そのものになることもあります。ICIJ が指摘する Offshore Leaks データベースは、「学者、NGO、税務署によって定期的に利用されている」のです(Cabra, 2017)。公開のプライドから、ファシリテーター、オーガナイザー、データベースマネージャーとしての役割に賞賛が集まるようになったのは注目に値します。その結果、コラボレーションはそれ自体がスキルになりました。多くの非営利団体は、寄稿者やパートナーとの関係を構築・維持するためのコミュニティやプロジェクトマネジメントの役割を担っており、ジャーナリズムのトレーニングもこのような変化を反映するようになってきています。その一部は、初期のデータジャーナリズム文化の影響に遡ることができます。2016 年にカナダでの実践について書いたアルフレッド・ハーミダとメアリー・リン・ヤングは、「組織間のネットワーク型ジャーナリズム関係を優先する進化した分業」と指摘しました。そして、2018 年にロイタージャーナリズム研究所が協働ジャーナリズムの台頭に関する本を出版し、「コラボレーションはそれ自体がストーリーとなり、ジャーナリズムのインパクトをさらに高めることができる」(Sambrook, 2018)と指摘したことで、さらに影響力が認識されるようになりました。

何をどのように数え、正しく理解するかについて変革する

インパクトのあるストーリーを作るのに、高度な技術力は必ずしも必要ではありません。データジャーナリズムのプロジェクトで最も長い歴史を持つものの 1 つ、[英国の]調査報道局の「Drone Warfare」プロジ

ェクトは、5 年以上にわたって米国の無人爆撃機を追跡しています4。その中核となる手法は、一言で言えば「永続性」です5。同局の記者たちは毎週、「フリーテキスト」のレポートを、分析、検索、照会が可能な構造化されたデータセットに変換しています。情報源へのインタビューによって補完されたこのデータは、NGO によって利用され、同局は英国議会の国防委員会に書面による証拠を提出しました6。数えられていない数を数えることは、データジャーナリズムが影響を与える特に重要な方法です—確かに、これはデータジャーナリズムにおける「声なき者に声を与える」に相当するものと言って差し支えありません。15 カ国以上のジャーナリストが参加するプロジェクト「The Migrants' Files」は、データジャーナリストが「ヨーロッパへの移住や滞在を試みて死亡した人々の使えるデータベースがない」と指摘したことから始まりました(The Migrants' Files, n.d.)。その影響は、他の政府機関にも及んでいます。国際移住機関などは現在、独自のデータを収集するようになりました。

政府がカウントしているように見えるときでも、調査する価値がある場合があります。たとえば、BBC イングランド・データ・ユニットと協力して図書館の削減規模を調査していたとき、国会でこの問題に関するデータを求める質問が出されているのを見て、一瞬パニックになったことがありました(「Libraries Lose a Quarter of Staff as Hundreds Close," 2016」)。この回答は、私たちが何ヶ月もかけて行ってきた作業を出し抜くものだったのでしょうか?実はそうではなく、私たちが行ったような深い調査を行っていなかったため、政府自身が削減の真の規模について私たちよりも知らないことが明らかになったのです。また、データの存在だけでなく、その表現にインパクトがある場合もあります。メキシコの新聞『El Universal』によるあるプロジェクト「Ausencias Ignoradas」(無視された不在)は、この 10 年間にメキシコで行方不明になった 4500 人以上の女性に顔写真をつけています(Crosas Batista, 2016)。データはあっても、「人」のレベルまで分解されていなかったのです。Libération の「Meurtres conjugaux, des vies derrière les chiffres」は、家庭内での女性の殺人について同じことをしており、Ceyda Ulukaya の「Kadin Cinayetleri」プロジェクトはトルコの女性殺人をマッピングしています7。

データ不良の場合:データ品質へのインパクト

データジャーナリストとして私が好きなプロジェクトのいくつかは、欠陥や欠落したデータを強調したり、その特定につながったりするものでした。2016 年、BBC イングランド・データ・ユニットは、どれだけのアカデミー校が透明性に関する規則に従っているかを調べました。100 校のアカデミーを無作為に抽出し、公式ルールで義務付けられている、すべてのガバナーの利害関係の登録簿を公開しているかどうかを確認しました。5 校に 1 校はそうしておらず、その結果、規制当局の Ofcom は私たちが特定した学校に対して措置を取ったのでした(「Academy Schools Breach Transparency Rules," 2016」)。しかし、彼らはこれを継続させることに真摯だったでしょうか?後年、この話を検証することは、その影響が単に短期的なものであったのか、それとももっと体系的なものであったのかを確定する上で重要でしょう。

データジャーナリズム・プロジェクトのインパクトは、ストーリーの準備が整い、反応が求められているときにだけ確認できる副産物であることもあります。[英国の調査報道組織]Bureau Local が、イングランドの 18 の議会が金融不安を抑止する引当金を何も準備していないことを発見し、その対応を求めたところ、データが間違っていることが判明しました。「数字をまとめた議会も、(数字を)吟味して公表した住宅・コミュニティ・地方自治省も」含め(Davies, 2018)、誰も間違ったデータであることに気づかなかったということです。彼らの調査は、地方公共団体がより一貫して、よりオープンに、より正確にデータを公表するよう求める運動の高まりに拍車をかけたのでした。

イノベーションを超えたインパクト

データジャーナリズムがより日常的になり、これまで以上に複雑なビジネスモデルに組み込まれるようになるにつれ、その影響力はイノベーションから配信の領域に移行しています。データエディターのデビッド・オットウェルが 2018 年にその違いを書いているように、

イノベーションによって、データジャーナリズムが一面を飾るようになった。配信は、それを毎日一面に掲載し続けることです。イノベーションとは、読者が重要な問題を探求し理解できるような、おしゃれなインタラクティブを構築することです。配信とは、それを実行し、大勢の人に実際に使ってもらうことであり、翌日も、その翌日も、作り続けることなのです。(Ottewell, 2018)

配信はもちろん、仲間を超えたインパクト、印象的なデータビズやインタラクティブ・マップの「すごい」要素を超えた、現実世界でのインパクトのことでもあります。それは、すぐに明らかになることもあれば、じわじわと、目立たずに拡散していく場合もあります。ボストン・グローブのカトリック神父のケースのように、自分たちが変化を起こせなかったと感じることもありますが、変化には時間がかかるものです。報道は、変化の種をまき、何年も何十年も経ってから結果を出すことができるのです。Bureau Local と BBC は、議会や学校のデータの信頼性が将来的に高まるかどうかは分からないが、両者に改善を促すスポットライトが当たっていることは知っています。ジャーナリズムができることは、スポットライトを当て、行動を起こすのは他者の責任だと受け入れることしかできない場合もある一方で、自ら行動を起こし、よりオープンな状態を求めるキャンペーンを行う場合もあります。データジャーナリズムには、より開放するよう強制する能力もあれば、他の人々が行動を起こすことを可能にするツールを作る能力もあるのです。

最終的に、インパクトのあるデータジャーナリズムは、アジェンダを設定することができます。他のジャーナリズムが到達できない聴衆に到達し、他のジャーナリズムが到達できない方法で彼らを巻き込むことができるのです。声なき者に声を与え、他の方法では不明瞭なままであった情報に光を当てます。データの責任を追及し、その権力に真実を語りかけます。このようなインパクトには定量化できるものと、測定が困難なものがあります。イ

ンパクトをモニターする試みは、このことを念頭に置く必要があります。しかし、それは努力しなくてもよいという意味ではありません。

脚注

1. www.imdb.com/title/tt1895587...

2. www.bbc.com/news/world-15391515

3. http://powerreporting.com/colo...

4. www.thebureauinvestigates.com/projects/drone-war

5. www.thebureauinvestigates.com/explainers/our-methodology

6.publications.parliament.uk/pa/cm201314/cmselect/cmdfence/772/772vw08.htm

7. www.liberation.fr/apps/2018/02/meurtres-conjugaux-derriere-les-chiffres/（フランス語）, http://kadincinayetleri.org/（トルコ語）

参考文献

Academy schools breach transparency rules. (2016, November 18). *BBC News*.www.bbc.com/news/uk-england37620007

Angwin, J., & Tobin, A. (2018, March 27). Fair housing groups sue Facebook for allowing discrimination in housing ads. *ProPublica*. www.propublica.org/article/facebook-fair-housing-lawsuit-ad-discrimination

Blau, U. (2016, April 6). How some 370 journalists in 80 countries made the Panama Papers happen. Nieman Reports. niemanreports.org/articles/how-some-370-journalists-in-80-countries-made-the-panama-papershappen/

Byers, D. (2012, November 6). 20% of NYT visitors read 538. *Politico*. www.politico.com/com/blogs/media/2012/11/nate-silver-draws-of-nyt-traf c-148670.html.

Cabra, M. (2017, November 29). How ICIJ went from having no data team to being a tech-driven media organization. ICIJ. www.icij.org/inside-icij/2017/11/icij-went-no-data-team-tech-driven-media-organization/

Chammah, M. (2016, February 3). Policing the future. The Marshall Project. www.themarshallproject.org/2016/02/03/policing-the-future

Crosas Batista, M. (2016, June 22). How one Mexican data team uncovered the story of 4,000 missing women.

Online Journalism Blog. onlinejournalismblog.com/2016/06/22/mexico-data-journalism-ausencias-ignoradas/

Davies, G. (2018, May 2). Inaccurate and unchecked: Problems with local coun- cil spending data. The Bureau of Investigative Journalism. www.thebureauinvestigates.com/blog/2018-05-02/inaccurate-and-unchecked-problemswith-local-council-spending-data

Fitzgibbon, W., & Díaz-Struck, E. (2016, December 1). Panama Papers have had historic global effects—And the impacts keep coming. ICIJ. /www.icij.org/investigations/panama-papers/20161201-global-impact/

Hermida, A., & Young, M. L. (2017). Finding the data unicorn. *Digital Journalism*,*5*(2), 159–176. doi.org/10.1080/21670811.2016.1162663

Higgerson, D. (2015, October 14). How audience metrics dispel the myth that read- ers don't want to get involved with serious stories. David Higgerson. davidhiggerson.wordpress.com/2015/10/14/how-audience-metrics-dispelthe-myth-that-readers-dont-want-to-get-involved-with-serious-stories/

ICIJ's awards. (n.d.). ICIJ.www.icij.org/about/awards/

Improving enforcement and promoting diversity: Updates to ads policies and tools. (2017, February 8). About Facebook. about.fb.com/news/2017/02/improving-enforcement-and-promoting-diversity-updates-to-ads-policiesand-tools/

Jalonick, M. C. (2017, October 27). Facebook vows more transparency over political ads. *The Seattle Times*. www.seattletimes.com/business/facebook-vows-more-transparency-over-political-ads/

Libraries lose a quarter of staff as hundreds close. (2016, March 29). *BBC News*. www.bbc.com/news/uk-england35707956

McClenaghan, M. (2017, May 18). Campaigners target voters with Brexit "dark ads." The Bureau of Investigative

Journalism. www.thebureauinvestigates.com/stories/2017-05-18/campaigners-target-voters-brexit-dark-ads

The Migrants' Files. (n.d.). www.themigrants les.com/

Ottewell, D. (2018, March 28). The evolution of data journalism. *Medium*. towardsdatascience.com/the-evolution-ofdata-journalism-1e4c2802bc3d

Rusbridger, A. (2018, August 31). Alan Rusbridger: Who broke the news? *The Guardian*. www.theguardian.com/news/2018/aug/31/alan-rusbridger-who-broke-the-news

Sambrook, R. (Ed.). (2018). *Global teamwork: The rise of collaboration in investigativejournalism*. Reuters Institute for the Study of Journalism.

Schlemmer, C. (2016). *Speed is not everything: How news agencies use audience metrics*.Reuters Institute for the Study of Journalism. reutersinstitute.politics.ox.ac.uk/our-research/speed-not-everything-how-news-agencies-useaudience-metrics

Stark, J. (2016, May 2). Investigating Uber surge pricing: A data journalism case study. Global Investigative Journalism Network. gijn.org/2016/05/02/investigating-uber-surge-pricing-a-data-journalism-case-study/

データジャーナリズムは何のためにあるのか？キャッシュ、クリック、そして試行錯誤

文：ニッキー・アッシャー

概要

商業的なデータジャーナリズムの金銭的インセンティブと意図せざる結果を取り上げます。

キーワード：ジャーナリズム、データフィケーション、誤情報、政治経済、選挙、実験

2016 年、政治家からジャーナリスト、学生、政府職員まで、ワシントン市民の間では、FiveThirtyEight のインタラクティブな選挙予測マップを毎日更新するのは慣例と化していました。ニューヨーク・タイムズの Upshot 世論調査アグリゲーターを愛用している人もいれば、よりオッズにこだわる人は Real Clear Politics を、よりエキゾチックな趣味を持つ人は The Guardian の米国選挙報道を利用する人もいました。こうした人々にとって、[文明崩壊後の近未来を描く映画]ハンガーゲームの様相を呈した米国大統領選挙でヒラリー・クリントンに有利なオッズがある限り、世界はすべて順調であり、スプレッドが大きければ大きいほど良いということになるのでした。この物語の結末はご存じでしょう。ネイト・シルバーのマップでは、選挙当日でさえヒラリー・クリントンが 71.4％で勝利する可能性が高いとされていました（が、結果は逆でした）。[さて、]2016 年のアメリカ大統領選挙を乗り越えるべき時期が来ているのではないでしょうか。選挙マップへの執着は、定期的に実施される国政選挙がある米国の[人々の]娯楽なのかもしれません-とはいえ、世界中のオーディエンスも注視しています（N. P. Lewis & Waters, 2018）。しかし、リンク切れにならない限り、この時の選挙マップは存在し続け、ジャーナリストとクリントン支持者の両方を悩ませるとともに、共和党が「敵の『時代遅れのメディア』は『フェイクニュース』だ」と[支持者を]引き付け続けるためのネタを提供しています。

政治はともかく、2016 年の米大統領選はデータジャーナリストにとって忘れてはならないものです。たとえ定量化が誰の目にも正しかったとしても、マッピングと可視化の失敗は、ジャーナリストの認識力に関する主張（あるいはもっと単純に、「公認の知識人」であるという主張）を解体するもう一つの道具となったのです。データジャーナリズムを選挙予測と一緒くたにするのは不公平でしょう。特に世界的に見れば、データジャーナリズムはそれ以上のものだということは間違いないものの、データジャーナリズムの極致は、無限に続く地図、クリック可能なチャート、計算の厳密さや統計的な腕前とは関係なく、ユーザーエラーや単純化、[社会的]疎外を踏襲しがちな計算機などに見えることもあります。このハンドブック第 2 版を手にした今、私たちはデータジャーナリズムが成熟し、自省する段階に達したと宣言することができます。そして、それゆえに、「何のためのデータジャーナリズムなのか？」を問うことが重要なのです。データジャーナリズムは今日、ジャーナリズムを再構築し、再活性化するための可能性を示唆しているに過ぎません。このハンドブックの初版は、2011 年に Mozilla Festival

で行われた大人数の共同プロジェクトとして始まりました。この取り組みを私は見ていましたが、すぐに、実際に目に見える成果として実現することはないだろうと考えていました（私は間違っていました）。

この第 2 版は、アムステルダム大学出版局が発行し、米国ではシカゴ大学出版局が寄稿者を募って配布しています。ここでは、データジャーナリズムの自由な性質が、プロ意識、秩序、正当性と代償にいくぶん失われたことを示唆しています。データジャーナリズムは主流となり、ジャーナリズム関連の学校で教えられ、ニュースルームにも組み込まれています（Usher, 2016）。データジャーナリズムは標準化され、過去 5～7 年間ほとんど変化していません。国際的なデータジャーナリズムコンテストを検証すると、形式とトピック（最も多いのは政治）の革新は停滞し、地図とチャートが依然として主流であることがわかります（Loosen et al.、2020）。インタラクティブ性は、情報可視化の関係者が「入門レベルのテクニック」と考えるものに限られており（Young et al.、2018）、さらに、データジャーナリズムは「動的、有向、加重グラフ」の可視化には十分に至っていないのです（Stoiber et al.、2015）。

データジャーナリストはまだ、オリジナルの「ビッグデータ」ではなく、前処理されたデータを扱っているに過ぎません-そしてこのデータは、インターネットサービスプロバイダーが収集するような深さとサイズを有するマルチレベルデータではなく、せいぜい「ビッグデータ的」な政府データです。この批評は、米国中心ではないにせよ、主に西欧中心の視点からなされたものですが、だからといって、私が提唱する本質的な行動を促す呼びかけ[の価値]が損なわれたわけではありません。データジャーナリストは、ジャーナリズムのための革命的な道具箱の上に座っているようなもので、[その中身は]まだ解き放たれていません。しかし、この革命がうまくいかなければ、ユーザー体験とニュース消費者の知識欲はともにさらに毀損し、最悪の場合、ニュースに対する不信感をさらに募らせることになります。

データジャーナリズムが過去 5～10 年間と同じことを続けるだけなら、デジタルとプラットフォームの時代におけるジャーナリズムの大義を推進することにはほとんどならないでしょう。そこで、この本質的な問い（「データジャーナリズムは何のためにあるのか」）を問い直すために、データジャーナリストと、データにそこまで詳しくないものの Web にどっぷり浸かってビデオやオーディオやコードを扱うジャーナリスト、そして彼らをつつき回す学者たちが、データジャーナリズムの原点、現在の根拠、未来を再考する必要があると提案します。

米国におけるデータジャーナリズム:起源神話

起源神話とは、私たちがどのように、そしてなぜ存在するようになったのかについて自ら語る物語であり、バラ色の眼鏡をかけ、事実よりも自慢に満ちていることが多いものです。米国におけるデータジャーナリズムの原点は、次のようなものです。データジャーナリズム以前の原初的な世界では、データジャーナリズムはコンピュータ支援報道として存在し、あるいは米国ではそう呼ばれていて、ジャーナリズムに社会科学の厳密性をもたらす機会を提供していました。データジャーナリズムが Web に導入されるまでの神話では、データジャーナリ

ストは、21 世紀の優れた計算能力を備えパワーアップした調査報道記者となり、データ（または文書）[の力]を解放して、他の方法では伝えられないようなストーリーを伝えるのを助けるとされています。単に記事を調査するだけでなく、データジャーナリストは新しいウェブスキルによってジャーナリズムを救うこともでき、ニュース消費者が喜び、学び、そしてもちろんクリックするニュースに、ある程度の透明性、パーソナライゼーション、そして双方向性をもたらした[とされています]。いわば、かつての Web 上に公開されてきた記事とは、決して同じものではなかったのです。

データジャーナリズムは、ジャーナリズムの質的評価に欠けていた誤りを正し、切望されていた客観的な基盤を提供し、それを、一つの報道機関から計算負荷を取り除く強力なクラウドベースのサーバーを備えた現在のリアルタイム・インタラクティブなデジタル環境以前には想像もつかない規模と能力で行うのです。初期の成功は前進への道筋を示し、一般の読者を調査協力者や市民科学者へと変えてしまうことさえあります。例えば、ガーディアンが議員の経費スキャンダルを報道した事例に加え、ニューヨーク市の公共ラジオ局 WNYC の「Cicada Tracker」プロジェクトでは、地域住民に土壌温度計を作らせて、おそるべき夏の虫[セミ]の到来を記録するのに協力させました。そして、ジャーナリズム、計算、群衆、データ、テクノロジーのこの刺激的な編成は続き、真実を正義へと押し上げることになる[のです]。

現在:「ハッカー・ジャーナリスト」はただの（退屈な）ニュースルーム従業員

現在のデータジャーナリズムは、ビジョンも現実も起源神話から大きく外れてはいません。そこで登場したのが、2 種類のデータジャーナリズムです。ジャーナリズムの崇高なマントルを前面に押し出した「調査型」データジャーナリズムと、最新のバイラル・クリックの興味に合わせて最適化できるデイリーデータ・ジャーナリズムです。これには、ASAP のジャーナリズムの地図作成の取り組みから、世論調査や調査研究を、ジャーナリズムの皮を被ったミームに仕立て上げることまで、幅広い事例が当てはまります。データジャーナリズムは、よく言えば退屈で、過度に専門的になり、悪く言えば、デジタル収入を得るための戦略の一つになってしまったのです。データジャーナリズムは、私たちが知っているジャーナリズムを一変させることができるといっても過言ではありませんが、今のところそうなっていません。2011 年の MozFest では、ボストン・グローブのホームページのモックアップ画像に、誰でも自分の顔を表示できるプラグインのようなものが、このフェスティバルの主役でした。それはそれで楽しかったのですが、ボストン・グローブは、ユーザーが作成したコンテンツを、事前フィルタリングなしに実際のホームページで使用させるつもりはなかったことは確かです。

同様に、最初のデータジャーナリズム・ハンドブックが誕生した頃、データジャーナリストは「ハッカー・ジャーナリスト」であり、テクノロジーからジャーナリズムへ、少なくともオープンソースとハッキングの精神を使って、従来のジャーナリズムのプロセスに逆らうプロジェクトを触発し、実験、不完全、遊び心を提供し、形式や内容では優れていないかもしれないが、それでもジャーナリズムをハックできる何かを生み出すと考えられていました（S.C.Lewis & Usher, 2013）。2011 年には、業界の外側にいる人々がジャーナリズムに進出すると

いう話でしたが、2018 年には、業界関係者がジャーナリズムのプログラミングをプロ化する話になっており、5 年の間に、革新と発明の精神は、明らかに企業的で、明らかにホワイトカラー的で、明らかに楽しくなくなっています(S. C. Lewis & Usher, 2016)。

退屈でもいいし、それなりの役割もあります。データジャーナリズムの専門化は、「データジャーナリストは英雄」という自己認識によって正当化されてきました。データジャーナリストは、異なる価値観(例:コラボレーション、透明性)とスキル(可視化、各種計算スキル)のおかげで、新しい方法で権力に真実をもたらすことができる人たちであると考えられてきました。パナマ文書やパラダイス文書は、このビジョンを最もよく表現しているものの一つでしょう。しかし、調査データジャーナリズムには、単なるデータ解析をはるかに超える時間、労力、専門知識が必要であり、主にインタビュー、現地取材、文書など、データ以外の昔ながらの情報源が含まれています。画期的な調査報道をコンスタントに出し続けろというの[にできないの]は、努力不足ではなく、撞着語法のようなものです。ヨーロッパのデータ・ジャーナリズム・ネットワーク、米国の非営利ニュース研究所、Global Investigative Journalism Network はいずれも、調査活動のための膨大なネットワークを紹介しています。実際、ゲームチェンジャーとなるような調査は簡単には得られません。だからこそ、こうしたハイレベルな成功例は一般に十指に満たず、2010 年のガーディアンによる国会議員の経費告発のようなクラウドソースによる調査の成功は、まだ新しいものに塗り替えられないでいるのです。データジャーナリズムに関していえば、過去[の作品]は序章にすぎません。2012 年にピュリッツァー賞を受賞したニューヨーク・タイムズの革命的な没入型ストーリーテリングプロジェクト「Snow Fall」は、2017 年 12 月に「Deliverance from 27,000 Feet(Everest)」として登場しました。

5 年後、ニューヨーク・タイムズはまた別の長編記事として雪山の災害についての長い記事を取り上げました(著者は同じでジョン・ブランチ)。この 5 年の間に、「Snowfall」または「Snowfalled」は、ストーリーに派手なインタラクティブ機能を加えることを意味する社内外の略語となったのでした。2012 年以降、データジャーナリストが、どんなストーリーでも自動で Snowfall できるテンプレートツールの構築に時間を使うべきか、それとも革新的な単発プロジェクトに取り組むべきか、タイムズに限らず米国や英国のニュースルームで議論が起こりました(アッシャー、2016 年)。

一方、2012 年でもインタラクティブ性は最低限だった「Snow Fall」は、2017 年末でもインタラクティブ性は最小限にとどめたのです。「でも、ちょっと待って」と、古参データジャーナリストは物言いをつけるかもしれません。「Snow Fall はデータジャーナリズムではない。ニュースアプリの開発者が考えたトリックかもしれないが、Snow Fall にはデータがない!」。

これこそ問題なのです。データジャーナリズムの人たちは「Snow Fall」をデータジャーナリズムだと思わない傾向がありますが、なぜそう思うのでしょうか?もし、データジャーナリズムが Web の長所を生かして新しい能力と方法でストーリーを伝えることでないなら、何のためのデータジャーナリズムなのでしょうか?データ

ジャーナリズムは、地図やグラフのためだけのものではありません。また、データを地図やグラフにすれば、データジャーナリズムが没入型のデジタル・ジャーナリズムの取り組みよりも知的優位に立つわけでもありません。地図にできるものはなんでも地図にしてしまう[ことになりがちです]。米国の選挙マッピングはさておき、最新の利用可能なデータを定量化し、クリックできる一貫性のある形で可視化することの倫理的な結果には批判が必要でしょう。

データジャーナリズムは、最も日常的な形式においては、単調なビジュアライゼーションとなります。これは、データジャーナリズムのプロジェクトが日常的かつ定期的な需要に向かっていることを考えると、特に当てはまります。新しい労働統計、都市のサイクリング・データ、リサイクル率、学術研究の結果など、可視化できる(そしてクリックされやすい)データであれば、何も考えずに可視化できてしまいます。最悪の場合、データジャーナリズムは単純化しすぎて、彼らの仕事が明らかにするはずのデータの主題から人間性を奪うことになりかねません。ヨーロッパを流浪する移民とその流れに関する地図は、インタラクティブな矢印や性別のない人物アイコンの形をしています。人文地理学者のポール・アダムスが主張するように、デジタル・ニュースの地図製作は難民危機をクリック可能な実体のない一連の動作に変えてしまい、ジャーナリズムが未知の「難民」を共感させ、数字以上のものにするためにできることとは正反対に作用します(Adams, 2017)。

また、新たな社会問題や学術研究をマッピングする前に、データジャーナリストは、「私たちは何のためにマッピングし、チャート化(あるいはチャーティクル化)しているのか」と問う必要があります。そして Snow Fall と移民マップの間に問題が横たわっています:データジャーナリズムは何のためにあるのでしょうか? 現在は主として専門化と同型化が進んでいます。さらにデータジャーナリズムはニュース消費者が世界を理解するのを助けるだけでなく、報道機関の収益を向上させるという企業のインセンティブが働いています。しかし、データジャーナリズムには他の価値もあります。

未来へ:データジャーナリズムはいかにしてその価値(と楽しさ)を取り戻すか

データジャーナリズムは何のためにあるのでしょうか? データジャーナリズムは、変化と革命、インスピレーションに満ちたハッキングと実験、疲弊しインスピレーションに欠ける業界を駆け抜ける反逆者たちの自己決定的ビジョンというルーツに戻る必要があり、ジャーナリストを知識、物語、流通に対する推定上の権威に立ち向かうよう駆り立てるのです。データジャーナリストはハッカーの着想を認め、かつて約束したようにニュースルームをハッキングすべきです。彼らは、ニュースルーム内での利益とプロ意識に焦点を当てる必要があります。アウトサイダーの地位を取り戻すことで、データジャーナリズムが約束した本質的な提供物に近づくことができるのです。[それこそ、]ジャーナリズムについて異なる考え方をする方法、ジャーナリズムを異なる形で表現する方法、新しいタイプの思想家や行動家をニュースルームに呼び込む方法、さらに言えば、ジャーナリズムを再活性化する方法です。将来的には、データジャーナリズムは「データ」という言葉から解放され、代わりに「ジャーナリズム」という言葉に焦点を当てたものになると想像しています。データジャーナリストは、ニュースルームの他のメン

バーや他のジャーナリストにはないスキルを持っていると思われています。複雑なデータを理解する能力、あるいはそれをコンピュータに代行させる能力、データを意味のある形で可視化する能力、そしてコーディングの能力です。

データジャーナリズムは、地図や図表の作成という反射的な衝動や、360 度動画、拡張現実（VR）、アニメーションなど、データ集約型ではない技術や技能を軽蔑する態度を捨てなければならないのです。BBC の「Secret Life of the Cat」やニューヨーク・タイムズの「Dialect Quizzes」、Journalism 360 immersive news initiative から資金提供を受けた Dataverse の取り組みのように、360 動画や VR とデータを組み合わせたプロジェクトがもっと増えていくでしょう。

選挙マップや、その日のニュースを説明する地図製作も減り、遠くの犠牲者をクリックしやすい線や流れに還元することもなくなるでしょう。米国のトップニュースの新しい傾向である、ライブタイムの投票結果を示すインタラクティブの終焉を見ることができるかもしれません。ジャーナリズムのあるべき姿とその役割について、より多くの独創性と楽しさ、そしてインスピレーションに満ちた破壊が行われるようになるはずです。データジャーナリズムは、説明責任を果たすためのものですが、楽しさと想像力のためのものでもあります。国会議員が辞任するかもしれないとか、トレンド・ラインが明確になるといった理由だけではなく、一般の人々が報道機関やジャーナリストに立ち戻る価値を見出すからです。ジャーナリストはオリエンテーション、娯楽、コミュニティなど、さまざまな人間の情報ニーズを満たす存在なのです。

データに関する優れた知識を主張するために、データを知り、理解できるようにしようとするデータジャーナリストは、自らデータを収集する必要があります。データジャーナリズムは、誰かが収集したデータを新たに可視化するためだけにあるのではありません。[その欠点は]他人のデータを量産することで、データ提供者の思い込みが可視化されるくらいがせいぜいですが、最悪の場合、データジャーナリズムはデータ提供者のプレスリリースに堕落するでしょう。しかし、多くのデータジャーナリストは自分でデータを集めることにあまり興味がなく、自分の役割の枠外だと考えています。

ワシントン・ポストのデータエディター、スティーブン・リッチは、「ポストや他社は、考えるまでもなく政府が収集すべきデータベースを収集・維持する必要はないはずだ」とツイートしました。「これは我々のクソくだらない仕事であってはならない」（Rich, 2018）。しかし同時に、リッチが維持しなければならないことに苛立った銃犯罪統計は、彼が思っている以上に影響力があるものです。政府のデータには、十分な調査と考察を必要とする、何を収集するかについての仮定と決定が埋め込まれています。データは不活性ではなく、どのような事実が重要であるかという推定で満たされているのです。事実性の領域を支配しようとするジャーナリストは、なぜ事実がそうであるかを説明できる必要があります。実際、事実の体系的な生産は、近代ジャーナリズムの大部分において、ジャーナリストが認識上の権威を主張する方法なのです。

データジャーナリズムの目的は、現在よりもはるかに多くのことを行うことにあります。データジャーナリズムは、楽しみ、遊べる実験であり、ストーリーの語り方を変え、ストーリーに関する新たな思考法をもたらすものです。しかし、[それにとどまらず]真実の語り部、事実の提供者としてのジャーナリズムを再確立する上で重要な役割を果たすことも確かです。データを作成し、それを知り、事実に至った観察・データ収集の過程を説明できるようになるのです。データジャーナリズムは、プロのジャーナリストが、他のどの職業、機関、一般人よりも優れた方法で事実を収集することができ、実際に行っていることを示す重要な防衛線となる可能性が高いのです。

参考文献

Adams, P. (2017). Migration maps with the news: Guidelines for ethical visualization of mobile populations. *Journalism Studies*, *19*(1), 1-21.doi.org/10.1080/1461670X.2017.1375387

Lewis, N. P., & Waters, S. (2018). Data journalism and the challenge of shoe-leather epistemologies. *Digital Journalism*, *6*(6), 719-736. doi.org/10.1080/21670811.2017.1377093

Lewis, S. C., & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society*, *35*(5), 602-619. doi.org/10.1177/0163443713485494

Lewis, S. C., & Usher, N. (2016). Trading zones, boundary objects, and the pursuit of news innovation: A case study of journalists and programmers. *Convergence*, *22*(5), 543-560. https://doi.org/10.1177/135485...

Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2020). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013-2016. *Journalism*, *21*(9), 1246-1263. doi.org/10.1177/1464884917735691

Rich, S. (2018, February 15). The @washingtonpost and others should not have to collect and maintain databases that are no-brainers for the government to collect. This should not be our fucking job. Twitter. twitter.com/dataeditor/status/964160884754059264

Stoiber, C., Aigner, W., & Rind, A. (2015). Survey on visualizing dynamic, weighted, and directed graphs in the context of data-driven journalism. In *Proceedings of the International Summer School on Visual Computing* (pp. 49- 58).

Usher, N. (2016). *Interactive journalism: Hackers, data, and code*. University of Illinois Press.

Young, M. L., Hermida, A., & Fulda, J. (2018). What makes for great data journalism? A content analysis of Data

Journalism Awards nalists 2012–2015. *Journalism Practice*, *12*(1), 115–135.

doi.org/10.1080/17512786.2016.1270171

データジャーナリズムとデジタル・リベラリズム

文：ドミニク・ボイヤー

概要

データジャーナリズムの台頭と政治的リベラリズムはどのように交錯しているのでしょうか。

キーワード：リベラリズム、定住型ジャーナリズム、スクリーンワーク、ラテラルメッセージング、自我的個性、データジャーナリズム

この 30 年間で、職業としてのジャーナリズムと、ニュース制作の組織文化が大きく変容してきました。その原因と結果は複雑すぎて、ここでは詳しく説明できません。1990 年代末にはまだかなり強固に見えた印刷物と地上波放送のモデルは、インターネット、検索エンジン、ソーシャルメディアがコミュニケーションと情報システムの主流となり、「新自由主義」という経済哲学によってニュースメディアの金融化と民営化が広く行われたことによって、デジタルファーストのニュースメディアモデルにほぼ完全に取って代わられたと言えば事足ります。本書が主張するように、増殖するデジタルデータの流れとトークンは、今やジャーナリズムの前提となっています。すべてのジャーナリズムは、今やある程度、「データジャーナリズム」なのです。

2013 年に出版した『The Life Informatic』で、私はこのプロセスを「横方向革命」と表現し、ニュースメディアの放射状（たとえば、大部分が単方向、ハブからスポーク）インフラの支配から横方向（たとえば、大部分が多方向、ポイントからポイント）インフラへの生態的変化が起きたことを示唆しています（Boyer, 2013）。レイモンド・ウィリアムズ（1974）がテレビの台頭に関する優れた歴史的研究の中で述べているように、電子メディアは 18 世紀以来、放射状と横方向の両方の潜在能力を示してきました。これらの潜在能力がどこで解き放たれ、制度化されたかは、常に技術そのものを超えた社会的・政治的状況によって導かれてきています。ファクシミリは、社会的ニーズが顕在化する 1 世紀以上前に試作され、正式な「発明」は［社会的ニーズが顕在化するまで］遅れたのです。ウィリアムズ（1974）は、ラジオからテレビに至る放送システムが社会的に必要とされるようになったのは、彼が「モバイルの私有化」と呼ぶ西洋社会が進展し、政府や産業が市民消費者の位置を特定してコミュニケーションをとるのが、全地域に広がる「包括的」放射状メッセージ以外では困難になってからだと論じています。

現代の状況から得られる教訓は、インターネットのような新しい技術やインフラだけがニュースジャーナリズムにおける最近のデータ革命を推進していると考えるべきではないということです。私たちはむしろ、より複雑な社会的勢力の生態系の中で、ニュースメディアがどのように進化してきたか（そして進化し続けているか）に注意を払うべきです。ウィリアムズのアプローチは、私が 2000 年代後半にニュース・ジャーナリストとのフィールドワークで得た、ニュース・ジャーナリストのデジタル情報実践と 1980 年代以降の社会・経済の広範な新自由主義化との間に共生関係があるという直感を捉えるために、『The Life Informatic』で開発した「デジタ

ル・リベラリズム」という概念に生かされました。例えば、私は、報道記者の間でスクリーンワークの重要性が高まっていることに関心を持ちました。スクリーンワークは、データジャーナリズムの台頭のためのインフラ的な前提条件であると考えることができるかもしれません。スクリーンワークは、1970～80 年代、ニュースメディアをはじめとする欧米の企業文化において、パソコンやオフィスベースのデジタル情報システムを活用して新しい生産効率を生み出そうとする組織の取り組みによって、ニュース制作の一側面として初めて登場しました。

報道業界では、コンピュータ化は当初、ワープロの処理速度を向上させ、作文や一部のコピー編集作業を自動化して人件費を削減するための手段として捉えられていました。しかし、制度化の過程で、マーケティングからレイアウト、アーカイブに至るまで、ニュース制作のあらゆる側面にコンピュータが急速に入り込み、これまで人間が直接行っていた作業を自動化し、残った制作業務をより少数の報道関係者に集中させるという新たな機会が生まれたのです。

コンピュータ化以前の仕事を回顧するベテランジャーナリストたちは、「サポートスタッフの人数が増えた」、「机の前で過ごす時間が増えた」、「仕事量が増えた」とよく語ってくれました。確かに、ニュースジャーナリズムは昔から座ったままでする仕事ではありました。たとえば、タイプライターも座っての作業で、携帯電話システム以前の電話も座ったままするものでした。しかし、以前の座りっぱなしのジャーナリズムと現代のジャーナリズムの決定的な違いは、スクリーンワークによって、かつてないほど多くの重要なジャーナリズムの仕事（ワープロ、テキスト編集、アーカイブ調査、速報の監視、競争の監視、オフィス内のコミュニケーションや調整など）が、通常は固定された場所で、単一のインターフェースを通じて行われていることです。スマートフォンの利用と Twitter のようなソーシャルメディアがニュース報道にとって重視されるようになり、モバイルスクリーンワークは少なくともデスクトップのスクリーンワークと同じくらい重要になりましたが、ジャーナリストが「スクリーンに釘付け」という現象［そのもの］はほとんど変化していません。スクリーン・インターフェースがジャーナリズムの実践の中心的な側面になっていることに異議を唱える人は、いまやほとんどいないでしょう。ジャーナリストが行うほとんどすべてのこと、ほとんどすべての情報源、ほとんどすべての専門的なアウトプットには、1 つまたは複数のスクリーンが関与しているのです。このように重要なタスクが同じ場所に配置されることで、利便性が向上する一方で、注意力が散漫になることもあります。多くのジャーナリストは、管理しなければならないデータストリームの数とスピードに圧倒されていると感じています。私のフィールド調査では、ジャーナリズムのスクリーンワーカーは、負担が過剰なために、（例えば、ニュースの価値に関する）判断を他の信頼できるニュースソースに頼らなければならないとしきりに報告しています。

現代のニュースにおける悪評高い「群集心理」にスクリーンワークがいかに寄与しているかは容易に理解できます。データストリームが猛烈なスピードで流れている間、気を取られ、圧倒されたジャーナリストたちはしばしばリアルタイムのガイダンスを互いに依存しあうことは珍しくありません。スクリーンワークの支配が孤立した現象ではないことに鑑みると、新自由主義を並行して研究することは必要です。

古典的リベラリズムは 17～18 世紀、世界中で植民地帝国が形成されていく現実に西欧の知的文化が適応する中で生まれました。中世のキリスト教保守主義やルネサンスの人文主義は、労働、自由、私有財産、生産力を重視する社会哲学に取って代わられていきました。初期のリベラリズムにとって重要な問題は、私有財産の追求と所有をいかに美徳に仕立てるかということでした。なぜなら、それは、人類への神の贈り物の分け前を貧しい人から奪う恐れがあると思われるからです。解決策は、人類の科学と産業は資源の生産的利用を改善する能力を持っており、新しい植民地のフロンティアも豊かなので、私有財産の取得はキリスト教の価値観と背反しないと強調することでした。

この新しい倫理観が形成されたことで意図せずもたらされたのは、理性、行動、自由、美徳の対象としての個人に注目が集まったことです。リベラリズムが資本主義とその現代の生活様式の興隆とともに発展するにつれて、個人は西洋文化においてますます重要な存在となったのでした。当初、個人は「社会」の制限的な力に調和的に対抗しようとしましたが、次第に個人はそれ自体が目的として位置づけられるようになり、すべての社会的・経済的関係が最終的に強固で生産的で自立した個人を育成、実現するのに役立ちました。これらの個人は理想的には社会的決定から解放され、自由に考え行動することができると想像されていました。

私は、人類学者のエリザベス・ポヴィネリに倣い、この個性のモデルを「自己整合」と表現しました。その思想的前提は、個人がかなりの程度「自分を作る」ことができるというもので、今日のリベラリズム政治圏であまねく愛されている命題です。

コンピュータの画面と何の関係があるのだろうかと思われることでしょう。たしかに、アナログとデジタルによる計算の最初の大きな試みは、20 世紀半ばのケインズ主義の社会民主主義諸国が戦争に備えて準備していた 1930 年代に行われました。しかし、現代のスクリーンワークのより直接的な前身であるパーソナル・コンピューティングの発展は、1970～80 年代、新自由主義が政治的にも哲学的にも優勢になったのと同じ時期に起こりました。ケインズ主義の社会民主主義が、1960 年代後半から 1970 年代にかけての複数の地政学的危機（ベトナム戦争、中東紛争、OPEC の形成など）の重みで崩壊しそうになったからです。リベラリズムが長い間、公共の利益を追求する最善の方法は民間の利益に力を与えることであると信じてきたのに対し、新自由主義は公共投資や公共機関を犠牲にしても民間の利益に力を与えるという点で、無慈悲なまでに自己整合的であると言えるかもしれません。政治と政策における新自由主義の転換は、インターネット、衛星放送、携帯電話といった通信・情報インフラへの大規模な新規投資を世界中で推進する一方で、ケインズ的な 20 世紀半ばの規範に追随する公益ニュースジャーナリズムに大きな負の影響を及ぼしました。

これらのインフラを想像し、構築することは、もともとニュースメディアとはほとんど関係がありません。インターネットは、広く知られているように、軍事防衛と研究科学という共通の利害関係によって誕生しました。あまり知られていませんが、同様に重要なのは、裁定取引などの金融実務において、高速の国境を越えた通信が有用であったということです。とはいえ、新しい情報通信インフラは、もちろんニュース制作も含め、社会的コミ

ュニケーションのあらゆる分野に影響を及ぼしました。その正味の効果は、ポイント間の横方向のメッセージング能力の根本的な強化であり、ハブとスポークをつなぐ放送の多元化と再時間化です。国民が一緒に座って夕方のニュースを聞くというモデルは、世界のほとんどの地域、たとえ公共放送の伝統が強く残っているヨーロッパやアジアでさえも、現実的には存在しないのです。

現代のニュース生態系は、コミュニティや信頼できる情報を見つけることをより不安定な冒険にしているにもかかわらず、強固な個人主義を保証するものでもありません。しかし、デジタル・リベラリズムの「泣き所」は、スクリーンワーク（と脚本）の個人化する体験にあります。パーソナルコンピュータ、インターネット、ソーシャルメディアの進化は、個人の行動、コミュニケーション、アイデアの能力を最大化するという新自由主義の社会的重要性によって深く形作られたものなのです。過去 10 年間で、第 5 の手足のように機能する万能ポータブルメディアデバイスを携帯する人々の割合は増えてきています（例えば、米国では 70％以上）。このデバイスは、複数の情報の流れにアクセスし、個人の興味や欲望を反映するようにその流れを管理し、個人の見解や考えを伝え、自己中心的なマイクロコミュニティを構成する無数の方法を可能にします。それは、何世紀にもわたるリベラルな認識論の継承者であると同時に、我々が「デジタル時代」と呼ぶようになった時代に、その認識論の再生産と強化を可能にする重要な装置でもあるのです。バーや電車の中で、見知らぬ人同士がスクリーンに釘付けになっている光景を見かけたことがあるでしょう。もちろん、スマートフォンが社会的疎外感を生み出したわけではありません。スマートフォンが発明したのは、世界のあらゆる場所で見知らぬ人たちに囲まれても、社会的なつながりや説明責任を最小限に抑えつつ、能動的で生産的な個性を体験できるコミュニケーション・インターフェイスです。他の時代には、見知らぬ者同士が非友好な関係だったとしても共存する関係を築く多くの機会があったのでした。

要するに、自己整合的な個性はスクリーン・インターフェースの普及に伴い強化されているのであって、そもそもそうしたインターフェースが存在すること自体が、過去数世紀の間にリベラルな世界観と優先順位を具体化するために開発されたテクノロジーと大いに関係があるのだと私は確信しています。マーシャル・マクルーハンの言葉を借りれば、われわれはスクリーンを操作していると思い込んでいますが、スクリーンがわれわれを操作していることも認識すべきでしょう。このように、携帯可能なスクリーンを使ったメディアと、リベラル派の自己整合的個性との接点は、私が「デジタル・リベラリズム」と呼ぶものであり、このリベラリズムが今後どのように進化していくかは興味深いところです。もし、あの電車の車内にいた見知らぬ人たちが皆、VR ヘッドセットを装着して仮想世界に没入していたとしたらどうでしょう。そのような新しいメディア・インターフェースは、個性と社会性の新しい様式をどのように引き出すのでしょうか。データジャーナリズムは、監視技術やアルゴリズムによる権威主義との親和性を疑われがちですが、実はデジタル・リベラリズムの進化こそがデータジャーナリズムの底流をなす歴史なのです。

参考文献

Boyer, D. (2013). *The life informatic: Newsmaking in the digital era*. Cornell University Press.

Williams, R. (1974). *Television*. Wesleyan University Press.