



データ集約的開発——『世界開発報告 2021 年版：生活向上のためのデータ活用』を読んで

Data intensive development: A review of *World Development Report 2021, Data for Better Lives*

伊藤 成朗

Seiro Ito

2021 年 10 月

(9,580 字)

\* 画像、図は文末に掲載しています

## はじめに

インターネットが一般家庭に普及して 25 年、グーグルが検索で支配的になって 20 年強、SNS や iPhone が使われ始めてから 15 年あまり。無料 SNS の売っている商品が実は自分たちの情報だった、と多くの利用者が気づき始めたのはここ 5～6 年である。GPU（画像処理のためのプロセッサ）によって深層学習の計算速度が格段に早くなることが分かってからは、データは人工知能（AI）を導く生産要素となった。ゲーマーの熱狂が世界を変える起爆剤になるとは、財界も学界も予想しなかっただろう。

いまや相手の要望を細かく分類・予測して最も受容されやすい商品・情報を時宜を得て提示可能になったのに加え、機械が生産したデータを IoT（モノのインターネット）によって機械同士がやりとりし、保守の予測やラインの組み替えも容易になって生産工程管理も自動化され始めた。コード化できる熟練は機械に置き換えられ始め、AI＋データは生産費用を下げるだけでなく、自動運転や創薬をはじめとする技術進歩も促すようになった。

AI が生産技術やマーケティングに絶え間なく用いられていることから、AI に仕込むデータを効率的に生産し利用すれば、生産性を高めることができる。生産性の上昇は、サービスの質が高まり価格も下がるのに加え、生産者の利潤を増やす。法人税収も増えるし、（そうした企業に）雇用される労働者を筆頭に労働分配率も高まるはず……法人税収と労働分配率

がこの予想と正反対の方向に動いているのはさておき、AI＋データが人々の暮らしを豊かにすることは間違いない。途上国でもそのはずである。

2021 年版世界開発報告（World Bank, 2021）は、途上国の貧しい人たちの生活を良くするためにデータをどのように活用すべきか、包括的に論じている。データを公目的（public intent）と私目的（private intent）に分け、両者の生産や共有を促しつつ、両者の組み合わせにも便益があることを示す。データには、公共財に近い性質のために過小供給になりやすい、発生源となった主体の意思に反して利用される、などの市場の失敗がある。このため、市場を補完するアプローチ、つまり、データに関する個人、企業、政府による社会契約を提唱する。データを生産要素として捉え、データの生産と利用を効率化させるためのデータ統治という視点から、必要な要素を丁寧に議論している。

## 公目的データと私目的データ——貧困層支援への活用例

公目的データとは、行政データ（住民台帳、納税、医療保険、水道・エネルギー提供量、交通情報）、センサスや標本調査（家計調査、企業調査）、市民発生データ（治安や環境）、機械発生データ（気象、衛星画像）、地理データである。多くが調査対象を全国および全住民、もしくは、母集団を全国および全住民としたデータである。よって、包括性に特徴がある。

私目的データとは、企業が営利のために生産したデータである。消費者選択の情報（通話詳細記録、アンテナ基地局接続記録、検索や SNS 利用記録、電子マネー利用記録）、生産過程の情報（生産機器発生情報、輸送機器発生情報）など、特定の生産活動に伴って発生する企業の私有情報であり、対象（母集団）、目的は絞られている。よって、特化に特徴がある。

貧困層支援の策定には、家計調査などの公目的データが伝統的に用いられてきた。近年では、衛星写真、気象情報などの非伝統的な公目的データに加え、通話詳細記録などの私目的データも活用され、支援の精度を高めたり、いままでになかったサービスを提供したりするようになっている。

非伝統的な公目的データを貧困対策で活用した代表例に貧困地図がある。貧困層支援の第一歩はどの地域にどれだけ貧困者がいるかの把握である（各地域の貧困者数を記載した地図を貧困地図という）。ただし、その資料となる家計調査を全国で網羅的に継続して実施するには費用がかかる。このため、全国を継続的に網羅する非伝統的な公目的データを再利用して家計調査データと接合し、全国で継続的および網羅的に貧困層数を予測する。たとえば、衛星画像から得られる夜間輝度データ（luminosity data）で輝度と地域の貧困度合いの関係を機械学習によって推計し、その関係を当てはめれば、夜間輝度データしかない地域や時期でも貧困度合いが予測できる。統計学的な推計なので、予測値の誤差範囲も示すことができる。衛星画像は気象や測量など本来の目的の下に費用が支出されているので、貧困地図作成に用いても追加的な費用は少なく、安価に貧困率予測ができる。

衛星写真のような機械が発生させる画像情報は、情報量が細やかで更新も頻繁なため、予測される貧困度合いも細やかで時宜を得たものになる利点がある。たとえば、全国の農産物反収を計測すると高い費用がかかるが、衛星画像と機械学習を使えば安価で広域の収量予測値を得られる。

さらに、衛星画像の「正解」として地上現象（ground truth）データを追加で収集し、機械学習のトレーニング・データとして用いることであれば、予測精度を高めることもできる。収量予測で面積あたり生産量を地上現象としてサンプルし機械学習の精度を高めると、収量予測の精度を高められる。特定地域の特定項目だけを調べるので、地上現象のデータ収集費用は広範な項目を網羅的に収集する家計調査よりも圧倒的に低いし、計測機器を用いれば精度を高める余地もある。農村で収量を予測できれば、不作時に作物や地域を特定した貧困対策を事前に検討できるだけでなく、収量をより正確に反映した増産技術を支援できるメリットがある。

私目的データを再利用すれば、企業は貧困層相手の商機を広げることができる。たとえば、モバイル通貨の先駆である *M-Pesa*（エム・ペサ）は、利用者が有料通話時間（エア・タイム）を融通しあって資金移動手段にしていることを知ったケニアの通信会社サファリ・コムが正規に立ち上げた送金・決済サービスである。これに留まらず、サファリ・コムは、*M-Pesa* 利用記録と通話詳細記録を使って貸し倒れリスクを審査し、申し込みから返済までオンラインで完結する無担保デジタル・ローン *M-Shwari* も提供している。世界開発報告は、携帯キャリアによるデジタル・ローンをデータによる金融包摂として高く評価している。ただし、サファリ・コムは小規模ローンの貸倒率が高いことを理由に、少額案件の条件を短期高利商品（オーバードラフト・ローン）に揃えた。リスクの高い借り手が増えたために高金利にした可能性があるため、市場の失敗（逆選択）を注視すべきである。

私目的と公目的のデータを再利用して組み合わせれば、貧困者数予測はきめ細くなる。ほぼ全家計が所有する携帯電話の通信詳細記録（通話時間、通話日時、相手、頻度、SMS テキストデータ、料金）と家計調査データを組み合わせれば、実態により即した経済状態が予測可能である。夜間輝度データは地域単位なのに対し、通話詳細記録は個人単位の情報なので、個人単位で貧困者かどうか予測できる。

私目的データと公目的データを組み合わせて市場の失敗によって欠落していた市場を作る事例も出てきている。たとえば、衛星画像、気象記録、農地の位値座標を機械学習させた予測モデルに入力することで降雨量を推計し、天候指数保険（weather index insurance）を開発販売し、農家の天候リスク軽減を支える企業がある。天候指数保険は、いままで何度も開発されては収益性のある事業として広がらなかった歴史がある。各農地にどれだけ降雨があったかなど、必要な情報を収集する手段がより整ったことで、情報の非対称性を克服する道を歩み始めている。さらに、ブロックチェーンを使って安価で速やかに支払をすることで、降雨確認に加えて事務手続きの人的費用が節約でき、保険価格を下げるができる。

世界開発報告では、ほかにも、津波早期警戒のためにブイや地震計データを整備すること、災害救援のためにアンテナ基地局経由データを活用することなども唱えている。新型コロナ禍でアンテナ基地局接続記録を使って人の動きを感染対策に活かしているが、これも私目的データを用いた公衆衛生対策である。

## データの両刀性

家計調査などの伝統的データは回答者の参加同意を得て収集される。情報は匿名化したうえで公開すること、参加によって回答者に過大な不利益やリスクを与えないこと、などが参加の前提条件になる。一方、非伝統的なデータは、そうした個人の権利保護が不透明な場合がある。通話詳細記録、アンテナ基地局接続記録、検索記録、閲覧記録、SNS の書き込み、電子マネー支払記録、電力メーター記録などは、利用者自身が通信会社などプラットフォーム相手に利用許諾を与えている。しかし、政府を含む第三者に供与する際の条件を利用者が十分理解しているとは言い難い。

筆者が何気なく北海道の賃貸物件を検索したところ、数カ月経っても筆者が使っている天気予報サイトに北海道の物件広告が表示される。データ使用許諾に関する個人の無理解は、こうした軽度のアノイアンス（うざさ）にとどまる問題ではない。日本でも、学生就職支援サイトが利用者の許諾を得ずに収集したデータをもとにはじき出した内定辞退予想率を企業に渡していたり、通信会社が利用者に知らせずに個人データを外国のサーバに保管し、日本の個人情報保護法が適用されない可能性を招いたり、といった事態が発生している。イギリスのケンブリッジ・アナリティカ社は、フェイスブック利用者 27 万人に同意を得てクイズをすることで、利用者とその「友達」の情報を提供させ、300 万人以上の利用者の政治的志向に関するデータを蓄積し、英国の EU 脱退をめぐる国民投票やアメリカ大統領選挙で個人ごとに狙いを定めた政治広告を打てるよう、特定の陣営にそのデータを提供した。その集票効果は疑問視されているが、「友達」であった 273 万人あまりは知らないうちにフェイスブック上の情報を政治活動に利用され、トランプ氏賛美の広告を見る羽目になった。

途上国においては、個人のデジタル・リテラシーが低いだけでなく、法整備も進んでいないために、政府やプラットフォームが個人データを利用することの周知が不徹底になりやすい。法令がないと権利や義務がはっきりしない。本人が知らないうちに機械が発生させたデータに誰が何の権利を持っているのか、その権利をどのように保護するのか、権利を侵害した場合の罰則は何か、誰が違反を摘発するのか、データに関わる権利をどのように個人に周知するのか、さらにはデータを国外に越境させる際の条件は何かなど、個人データの利用を適切に定めた法令とそれを実行するための体制を整備することが急務となっている。

プラットフォームに競争を促す伝統的な規制も、データ利用が非効率を生み出さないために必要である。プラットフォームがデータを蓄積していくと、生産性が高まって独占化

し、消費者に不利益を与える可能性がある。競争制限的な企業買収によって技術革新が停滞したり、買い手独占的な取引慣行が課されて取引先の雇用、ひいては雇用一般が先細る、ということが起きるかもしれない。

データを悪用するのは企業だけでなく個人や団体の場合もある。セキュリティの緩いサーバで管理されている他者のデータを盗んだり、ナイーブな消費者を騙して個人情報を盗み出しダークネット上で売買したりするなど、サイバー犯罪は後を絶たない。漏洩データ被害への補償は不十分で、被害者本人が不利益の大部分を負担する。よって、個人にデータ主導経済への参加を促すには、サイバー犯罪を抑止する必要がある。

全く異なる問題として、政府が個人データを収集して、個人を監視したり主権を制限したりするのに利用する場合もある。中国政府は、SNS の書き込み、ウェブ閲覧記録、アンテナ基地局接続記録、公共交通利用記録などに加え、街中に設置された監視カメラの映像情報などを用いて、政府にとって好ましくない個人を監視している。インド政府は、監視や権利の制限を意図していたか不明だが、国民に生体認証 ID (*Aadhaar*) を発行しながら杜撰な情報端末管理をしていたために、10 億 (!) 件以上の個人情報を漏洩させる事件を起こした。インドはデータに関する個人や団体の権利保護を目的としたデータ保護法制を持たない。中国も 2021 年 11 月 1 日施行の個人情報保護法まで持たなかった。両国で個人情報保護に関する法令の整備が遅れている、という共通項があるのは偶然ではない。中国で個人情報保護法の上層に位置づけられるデータ・セキュリティ法では、「国はデータセキュリティ審査制度を構築し、国の安全に影響し、または影響し得るデータ処理活動に対し国家安全審査を行う」という制度も規定されている ([ジェットロビネス短信、2021 年 6 月 28 日](#))。個人情報保護法で個人情報保護を謳っているとはいえ、「国の安全」が優先されている。

## 社会契約の提唱とグランド・デザインの提示

世界開発報告では、個人、企業、政府によるデータ悪用を防止し、すべての参加者がデータから便益を得つつ、社会全体で効率性を高めるために、社会契約を結ぶことを提唱している。お互いを裏切らないこと（「信用」）、データを共有すること、そうすることで価値を生み出すこと（「価値」）、誰もがその過程に参加できること（「平等」）、が主たる内容である（図参照）。

悪用を防ぐためにはデータに関する権利・義務と違反行為・罰則を定めた法律を整備し、法執行機関を設置する。データの再利用と共有を促すために、データの形式標準を定める主体を作り、匿名データの仲介活動を推奨する。独寡占規制をプラットフォーマーに適するように運用する。データ作業を担う人材を育成し、利用者のリテラシーを高めるために、人的資本にも投資する。なによりも第一歩としては、誰もがデータ主導経済 (*data-driven economy*) に参加できるように、安価で安心安全な通信インフラと利用環境（ハードウェア、ソフトウ

ェア、情報セキュリティ)を整えることが喫緊の課題である。AI の学習データから一部のグループが除外されれば、その歪みは時間につれて増幅されるので、初期の段階から幅広い参加が望まれる。

世界開発報告は、通信インフラや利用環境の整備を阻む要因とその撤去を議論する。たとえば、国外とのデータ流通量を節約する設備(インターネット相互接続点やコロケーション)を国内に整備し、スマートフォンなどに対する関税率を下げることなどである。法整備の中身は、権利保護を可能にする保護措置(safeguards)とデータ共有を効率化する促進措置(enablers)に区別している。保護措置を通信インフラ整備前に成立させて幅広い個人の参加を募り、非個人データについては知的財産権保護を援用して供給を高めること、政府データのオープン・データ化や企業データのオープン・ライセンス化などの促進措置によってデータ仲介を増やし、データ活用効率の引き上げを推奨する。

世界開発報告は、最終的な到達点として、統合国民データシステム(integrated national data system)と呼ぶデータ生産と管理の体制構築を提唱している。信用、価値、平等を特徴とする社会契約と、その実現を支える通信、人的資本、法律、統治のインフラを総称したシステムである。先進国でも到達している国があるか怪しいが、途上国の目指すべきグランド・デザインを示すことで、政策の方向性と時宜を整理しやすくする狙いがある。

## 世界開発報告の貢献

第1の貢献は、データを生産要素として捉えるという斬新な視点の提供である。データがAIを用いた技術革新の背後にあることは誰もが指摘しながらも、生産要素として捉え直すとさまざまな課題が見えてくる。アナロジーとして生産要素の労働(力)を考えてみよう。労働力の再生産を保証するために労働者の権利を認めた労働基本法や最低賃金規制などがあり、生命の安全や健康を維持するためにさまざまな社会保障制度が構築され、人権を守るために人身売買を禁止する法律がある。これと同様に、データにも生産と利用についての制度が必要になる。世界開発報告は、これを保護措置と促進措置に分けて整理している。

第2の貢献は、既存文献の体系化である。世界開発報告は、社会契約に基づいたデータの生産と利用を促す営為を「データ統治」という言葉で表している。これまでの研究では、データ統治を考えるための部品が個別に分析されていた。途上国でのAI活用例、プラットフォームの行動原理、競争政策、貧困分析における非伝統的データの活用、消費者のデータ権利保護などである。世界開発報告はこれらを総合し、データ統治という言葉の下にそれぞれの部品を位置づけている。これは新たな言葉を作った以上の貢献である。

第3の貢献は、データ統治という整理の枠組みがあらゆる発展段階に適用可能な一般性を持つことである。同じ枠組みで考えることにより、各国がどこに優劣があり、どのような特徴のデータ統治が可能なのか、環境の変化に応じてどのように内容を変えていくべきか、考えやすい。

## 世界開発報告の課題

その一方で、世界開発報告には課題もある。第1に、これまでのすべての世界開発報告と共通の課題として、理想とする到達点を示すものの、どのように到達するかという方法論に乏しいことである。読者として不満を感じるが、すべての世界開発報告に共通することから、意図的に細かな方法論を示していないのであろう。すべての国で適用できる方法論は存在しないので、目標は示すが辿り着き方はそれぞれが考えてほしい、という哲学なのかもしれない。

第2に、低所得国・貧困層におけるデータの有効活用例 (use cases) が乏しいことである。このため、貧困層がデータを提供したときにどれだけの便益が見込めるのか、いまひとつはつきりしない。現時点で貧困層が提供するのは通話詳細記録くらいであり、通話に関わる情報から脱貧困の決定打が出てくるようには思えない。

そうはいつても、貧困層がデータを発生させるサービスを使い、行動実態データを提供しないと貧困層向けのデータ主導型イノベーションは生まれない。たとえば、*M-Pesa* も、貧困層の有料通話時間の使い方が知られたことで生まれている。同様に、電気、ガス、水道のスマート・メータでデータを蓄積すれば、政策にも商売にも役立つかもしれない。そのためには、家庭にインターネットが接続され、電気、ガス、水道なども家庭まで届いていることが前提になる。

通信インフラと利用環境を整備することは、消費者としてだけでなく、労働者・経営者としての貧困層を助ける展望を与える。豊富な無料オンライン教材や安価な学習指導 (エドテック) を用いてデータ補完的な技能を蓄積できれば、労働市場での評価はデータ集約的技術進歩とともに高まるからである。なお、5G が普及してレイテンシー (反応ラグ) が十分に短くなれば、先進国での作業の操作者を途上国で雇用する可能性がある。もしも、これがデータ代替的スキルなのであれば、企業に操作データを取られた後には解雇される一時的な労働需要かもしれない。

経営者としては、サプライチェーンに関わる企業のみが参加する「許可型ブロックチェーン」を取り入れ、さらにはブロックチェーン上で取引完了情報とともに自動執行される「スマート契約」を使うことで、取引費用を下げられる。ブロックチェーンの台帳は改ざんが難しいため、取引記録の証拠となり得る。よって、台帳を示すことで企業・経営者の履歴や評判を客観的に示すことができる。これは、いままでは銀行や取引先が独占していた自らの取引履歴情報の所有権を自分の手に取り戻すことを意味する。そうすることで取引での交渉力が高まるかもしれない。世界開発報告はデータ主導経済を支えるための人的資本投資やブロックチェーンなどに言及しているが、今後は貧困層を対象にした成功例を蓄積して具体的方法が共有されることを期待したい。

第3の課題は、政府を無条件に信頼していることである。中国政府のように、独裁体制の

維持を国民の厚生より優先させている政府が強権をふるうとき、信頼、価値、平等を要素とする社会契約は成り立たない。インド政府のように、意図は国民の厚生引き上げだったかもしれないが、政府が実施能力を欠く場合にも社会契約は崩壊する。世界開発報告が提唱する社会契約に反する政策を実施する政府に対しては、まさにジョン・ロックが唱えた抵抗権をもって、国民は社会契約に対して消極的になるしかないのかもしれない。また、エド・スノウデン氏など一部の亡命者は、民主制のアメリカ合衆国ですら、政府は国民を監視していると非難している。監視の事実がどうであれ、政府を信頼できない国民がいる以上、政府の行動を律する手立てを示すことなしに理想的な社会契約の実現を唱えても、説得力がないと映るであろう。世界開発報告は社会契約破りを罰する方法の乏しさを議論していない。

第4に、データ主導経済を推し進めることで、雇用の伸びが抑えられる可能性である。オターらの研究 (Autor, et. al. 2020) によれば、プラットフォームの登場によって独占化が進み、価格は低下したものの労働分配率の下がる傾向が各国で見られる (伊藤 2021)。労働集約的な生産技術を使って所得を高めていこうとする途上国にとって、データ主導経済が労働者の所得にどのような影響を及ぼすのか、見定めていく必要がある。

## おわりに——世界開発報告でデータを取り上げる意味

理想の到達点を示すことで、各国政府はどこを向いて努力をすれば良いのか、迷いはなくなる。理想先取的な世界開発報告にすることで、各国政府は自らが考え、実行せねばならなくなり、経済厚生を高める素地ができるかもしれない。現時点で貧困層にとって手が届かないデータ主導経済を議論することの意義を問うと、執筆者のひとは、だからこそ先取的な内容にして何が可能かの地図を描き、貧困層の賛同の下に、貧困層にとって役立つデータ統治を作っていく必要がある、と答えていた。

データ主導経済は先進国で勢いを増すばかりである。新興国は身構えて真っ只中に踏み込みつつある。低所得国はまだ時間がかかる。だからこそ、このタイミングでの刊行は低所得国が準備するうえで有益である。低所得国では、体制作りを先導するケニアのような国は少数であり、準備の進まない国が多数である。こうした国に有益な示唆を与えるべく、世界開発報告は貧困のレンズをもってデータ主導経済を見ている。通信インフラと利用環境の整備は、成果が出るまで時間のかかる長期的投資なので、じっくり取り組むべき課題である。今後もデータ統治に関わる研究と成果が求められる。

データのユーザである研究者の多くは、対象となる現象のデータがないとき、ないものは仕方ない、あるものでなにかするしかない、と諦めるのが普通であった。ないから仕方ないと諦めたり、提唱する内容を理想論と批判したりするのはたやすい。一方で、世界開発報告は、データがないことを真正面に捉え、データを生産し、悪用を防止し、共有を進める体系的営為を論じている。諦観・批判か、理想論の鼓舞か、どちらの態度が途上国社会にとって有益かは明白である。■



※この記事の内容および意見は執筆者個人に属し、日本貿易振興機構あるいはアジア経済研究所の公式意見を示すものではありません。

## 画像の出典

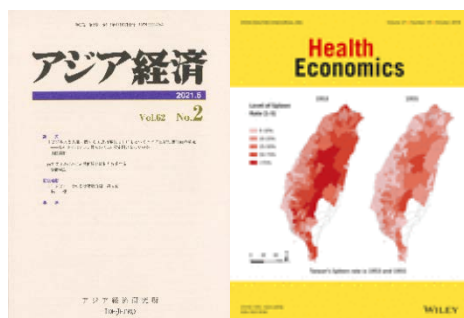
- World Bank. 2021. *World Development Report 2021: Data for Better Lives*. Washington, DC: World Bank. doi:10.1596/978-1-4648-1600-0. License: Creative Commons Attribution [CC BY 3.0 IGO](#). インデックスページの画像については一部を切り取り。

## 参考文献

- 伊藤成朗 2021 「途上国研究の最先端第 52 回 競争は誰を利するのか？ 大企業だけが成長し、労働分配率は下がった」『IDE スクエア』2021 年 10 月。
- Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. “[The fall of the labor share and the rise of superstar firms](#),” *The Quarterly Journal of Economics*, Vol. 135, No. 2: 645-709.
- World Bank. 2021. *World Development Report 2021: Data for Better Lives*. Washington, DC: World Bank.

## 著者プロフィール

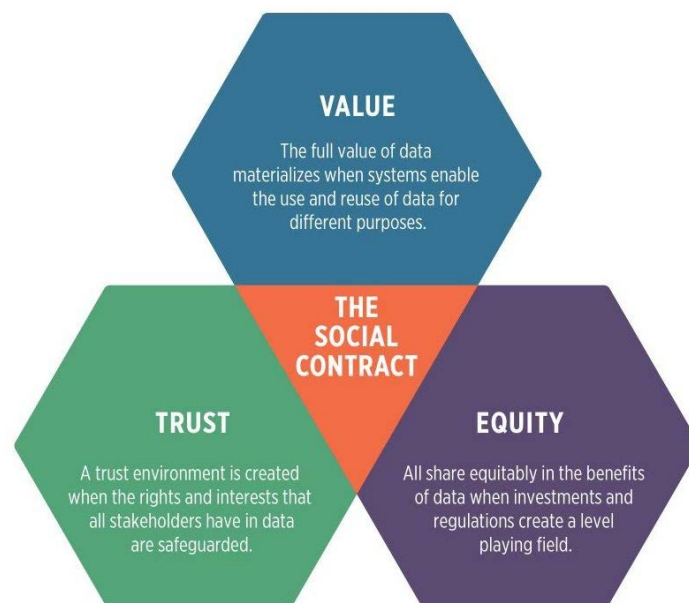
伊藤成朗（いとうせいろう） アジア経済研究所 開発研究センター、ミクロ経済分析グループ長。博士（経済学）。専門は開発経済学、応用ミクロ経済学、応用時系列分析。最近の著作に「南アフリカにおける最低賃金規制と農業生産」（『アジア経済』2021 年 6 月号）、主な著作に”[The effect of sex work regulation on health and well-being of sex workers: Evidence from Senegal](#).”（Aurélia Lépine, Carole Treibich と共著、*Health Economics*, 2018）など。





2021 年版世界開発報告の表紙

図 価値、信用、平等に基づいたデータの社会契約



(出所) World Bank (2021), p. 6. (CC BY 3.0 IGO)