

序文

データの統治^{ガバナンス}は先進国の間で、そして新興の大国において、ますます熱心な議論の主題になっている。しかしながら、多くの複雑な政策問題は未解決のままとなっている。そこで、この『世界開発報告 2021：生活向上のためのデータ活用』は、浮かび上がってきた見解を検討し、政策策定者に問題、機会、およびトレードオフなどを通じて熟慮するための枠組みを提示している。次の一点は明瞭である。すなわち、低所得国の視点はこのような世界的な議論においては今のところほぼ欠如しており、緊急に耳を傾ける必要がある。

データというのは諸刃の剣である。一方では、プログラムや政策を改善し、経済を促進し、市民のエンパワメントを進めることによって、価値を創造する大きな潜在力をもたらす。他方では、データの蓄積は経済力や政治力の集中につながる可能性があり、市民を害する形でデータが不正使用される可能性を高める。データというのは価値を増加させるために繰り返し使うことができる資源ではあるが、次のような問題を孕んでいる。すなわち、データの再利用が増えるほど、不正利用のリスクも高まる。

このような機会と緊張に関して、新型コロナウイルス感染症を超えるほどの劇的な適例を想像するのは困難である。世界中の国々が、ウイルスの拡散をモニターするために、携帯電話の記録のこの目的への利用を迅速に進めてきている。しかし、このことと同時に、このような便宜とプライバシーに関する懸念や乱用のリスクとのバランスをとることに苦闘している。

感染症流行の時期以外においては、コアな経済・社会データを作成して効果的に活用する統計的能力は限定的である。多くの貧困国は、公的財政を正確に追跡すること、対外債務について報告すること、あるいは自国の開発目標をモニターすることができていない。そのようなデータがなければ、政府の責任を維持する能力や、進展状況を追跡する能力は衰えてしまう。

データの幅広い活用を、その誤用を防止しながら促進するためのデータの統治にかかわる取り決めはまだ揺籃期にある。データに関する法規制の枠組みは低所得国では不十分であり、あまりにもしばしば、データ共有のための措置の不足に加えて、その重要な保護措置に関してギャップが存在している。相互運用を可能にし、またより多くのユーザーにデータが行き届くのを可能にする、データに関するシステムとインフラが不完全である。低・中所得国のなかで、コロケーション・データセンターやクラウド・コンピューティング基盤への直接的なアクセスなどの近代的なインフラを有している国は20%以下にとどまっている。初期的なデータ・システムや統治の枠組みを有している諸国においてさえ、必要とされる、行政能力、意思決定の自律性、および財務資源を有する機関が欠如していることは、その有効な実施と執行を妨げている。

このような問題に取り組むために、本報告書はデータに関する新しい社会契約を要求している。すなわち、それは、経済的および社会的な価値を創造するためにデータの利用と再利用を可能にし、データから利益を享受する公平な機会を促進し、提供したデータの誤用によって害されることがないという市民の信頼を醸成するものである。しかし、そのような社会契約を要請することにおいて、低所得国はあまりにも多くの場合に不利な状況にある。というのは、それらの国は、データを取り込み、それを価値のあるものに転換するためのインフラとスキル、グローバルなデータ市場とその統治に公平に参加する規模と担当機関、そしてデータ・システムにおける信頼を生み出すような制度および規制上の枠組みを欠いているからだ。

データについて新たな社会契約を締結することが国内政策のなかでも緊急の優先課題であり、このことは、国家的なデータシステムの強化と国レベルでの全利害関係者の関与を必要とするだろう。データは規模がグローバルであることから、社会契約のなかで最も挑戦的な側面の一部は、——二国間で、地域的に、グローバルに——規制を調和し政策を調整するために、より緊密な国

際協力も必要としている。国際的な関与が必要な最重要な分野に含まれるのは、データ主導型ビジネスにかかわる国際課税権の改革、データによって可能になるサービスの貿易に関する世界貿易機関(WTO)の取り決め、データ・インフラの開発に関わる地域協力、相互運用性を支援するための技術基準の国際的調和、そして法執行および独占禁止規制に関する二国間協調などである。

世界銀行はこのような重要で挑戦的な課題に関して、加盟国を支援する用意がある。この『世界開発報告』で明らかにされていることは、公的および民間部門による投資が最も重要な分野を特定すること、政策改革や技術援助にかかわる豊富なプログラムを策定すること、そして国境をまたぐ協力の招集および促進に役立ち得るグローバルな構想が有効な分野を指摘することなどによって、各加盟国のための支援を形作るだろう。

データの全面的な価値を現実のものにすることは実質的な公約と努力に依存しており、容易ではないだろう。しかし、それを行わないことの代価は、機会を失った、そして不平等が拡大した世界である。



デイビッド・R・マルパス
世界銀行グループ総裁

目次

序文	iii
謝辞	v
略号	ix

概観 1

概観	3
データを通じて開発目標を推進	3
データに関する統治を社会契約に適合させる	10
統合国家データシステムへ向けた動き	16
注	19
参考文献	20

PART I データを通じて開発目標を推進 21

Chapter 1：貧困者のためにデータの価値を活用 23

開発目標に資するデータの潜在力は活用されていない	24
データの略史	25
データの類型学	27
データの経済学と政治経済問題	30
開発のためのデータ：概念枠組み	33
データに関する社会契約の執行に向けたデータ統治の枠組み	38
すべてを統合：統合国家データシステムを確立	41
本報告書の構成	43
注	43
参考文献	44

スポットライト 1.1：コミュニティが自らのデータを収集および分析する能力を獲得するのを支援する 48

スポットライト 1.2：新型コロナウイルス感染症が流行している期間中とその後に、低・中所得国が債務を管理するためには適正なデータが重要 50

Chapter 2：公益を促進する力としてのデータ 53

公的目的データの中心的な役割	54
公的目的データと開発：価値を高める3つの経路	54
公的目的データのカバレッジ、質、そして有用性におけるギャップ	59
データ・ギャップはなぜ持続するのか：公的目的データの政治経済学	65
公的目的データの潜在力を実現	71
注	76
参考文献	79

スポットライト 2.1：女性や少女に対する暴力を削減するためにデータを活用 85

スポットライト 2.2：公的目的データの改善における国際機関の役割 89

Chapter 3：民間部門の資源としてのデータ 91

データ主導型ビジネス・モデルを通じて価値を創造し、開発に関連する挑戦課題を解決	92
企業の生産工程におけるデータの役割	92
データ主導型の企業、およびその価値創造を後押しする技術	93
低・中所得国のプラットフォーム企業に焦点を合わせる	95
経済活動のためのデータ投入財	98
生産プロセスで使われているデータが開発にもたらすプラスのインパクト	100
生産工程におけるデータ利用は、諸産業部門をどのように転換しつつあるか	103
データ主導型企業がもたらす、政策で取り組むべき潜在的なリスクと悪い結果	109
注	113
参考文献	115

スポットライト 3.1：オープンデータのビジネスへの応用は膨大な可能性を秘めている 119

Chapter 4：より一層の価値を引き出すためにデータを創造的に再利用	123
種類や出所が異なるデータの転用と組み合わせが有する力	124
データを転用することや組み合わせることからの新たな洞察	127
開発のために私的目的データを利用するに当たっての制約	136
データの革新に対する投資：データの文化を構築	139
注	146
参考文献	148
スポットライト 4.1：低・中所得国から得られる天気や、水、気候に関する より良いデータを収集し、共有し、そして利用する.....	153
スポットライト 4.2：私的目的の交通データの転用によって道路を安全にする.....	155
PART II ^{ガバナンス} データ統治を社会契約に適合させる	157
Chapter 5：データ・インフラ政策：貧困層や貧困国のために公平なアクセスを確保	159
不公平の原因としてのデータ・インフラ	160
貧困層を接続	162
貧困国を接続	171
結論と勧告	177
注	181
参考文献	183
スポットライト 5.1：新型コロナウイルス感染症の世界的大流行は 妥当なデータ消費に関する期待をどのように再調整し、デジタル格差を露呈したか.....	187
スポットライト 5.2：データの炭素足跡 ^{カーボンフットプリント}	189
Chapter 6：データにかかわる政策、法律、および規則：信頼性のある環境を構築	191
データの安全装置およびイネーブラーに関する信頼の枠組み	192
信頼されるデータ利用のために安全装置を構築	194
データ共有のためにイネーブラーを整備	201
全体論的な法的枠組みを策定するための勧告	209
法および規制の枠組み強化のための成熟度モデル	212
注	212
参考文献	219
スポットライト 6.1：データに関する発展しつつある社会契約：新型コロナウイルス感染症を コントロールすべく接触追跡を円滑化するためにデータ共有とデータ保護のバランスをとる.....	225
スポットライト 6.2：個人データの所有権を巡る論争.....	227
Chapter 7：データ経済の中で価値を創造する：競争、貿易、および税制の役割	229
競争、貿易、および課税を支援するためのデータ規則を考案	230
競争政策	231
貿易政策	240
租税政策	249
結論	254
注	254
参考文献	258
スポットライト 7.1：データ保護と競争政策の間のインターフェイスを理解する.....	262
スポットライト 7.2：データ統治に関連する課題への取り組みにおける 地域的および国際的な協力の役割.....	264
Chapter 8：データ統治の機関：共同での行動を通じて信頼を構築	267
開発のために諸機関はデータ統治をどのように支援できるか？	268
データのライフサイクルを通じたデータ管理	269
データ統治の機能	271
データ統治機能を実例の諸機関にマッピングする	273
データにかかわる仲介と協働	279
データ統治機関を有効にする	282
包容的なマルチステークホルダー統治を通じる持続可能な成果	286
成熟度モデルのレンズを通して諸機関の基盤を評価	288
注	290
参考文献	293

スポットライト 8.1：データに関する新しいグローバル・コンセンサスの必要性：行動の要請…	299
スポットライト 8.2：アマゾン流域で市民科学を奨励…	301

PART III 統合国家データシステムへ向けた動き 303

Chapter 9：統合国家データシステムの創設	305
統合国家データシステムに向けて	306
統合国家データシステムのビジョン	307
ビジョンを実現	311
注	325
参考文献	327
索引	329

ボックス

0.1 統合国家データシステムに向けて：数カ国の事例	18	4.3 海洋保護区における違法漁業を防止	133
1.1 本報告書が「データ」という用語で意味していること	24	4.4 統計的手法と私的目的データを使って 代表性と地理空間的な精度を改善	138
1.2 伝統的な標本調査における革新： 新型コロナウイルスに関するブラジルでの事例	29	5.1 頭脳流出——ICT 専門家	177
1.3 市場経由のデータ取引に関する挑戦課題	32	8.1 ウルグアイにおけるデータ統治に向けた 政府一体のアプローチ	268
1.4 新型コロナウイルス感染症との闘いで私的目的データを 活用	36	8.2 データ統治機関の重要性と複雑性： デジタル身元証明 (ID) システムの例	271
2.1 6 種類の公的目的データ	55	8.3 私的なデータの仲介業者の精査と制約の強化	281
2.2 世界銀行の統計パフォーマンス指標	60	8.4 「デジタル民主制」を通じて マルチステークホルダー型データ統治を スマートシティ・イニシアティブに組み込む	288
2.3 ジェンダーに関するデータと新型コロナウイルス感染症の 大流行	62	9.1 統合国家データシステムと 国家統計システムの関係	306
3.1 データ主導型の意思決定や仲介を支える 技術と方法	95		
4.1 新型コロナウイルスとの闘いに携帯電話を利用	124		
4.2 新型コロナウイルスとの闘いに 私的目的データを活用	128		

図

0.1 データはどのように開発を支えることができるか： 変化の理論	4	1.5 統合国家データシステムではデータは 全利害関係者の間を安全に流れる	41
0.2 データに関する社会契約は価値、信頼、 および公正性に基づく	6	S11.1.1 インド農村部の村において、世帯の意思決定に 関して権限を有する人を突き止める 市民主導型の方法	49
0.3 国内のおよび国際的なレベルにおける データ統治の層	11	S12.1 6 年間で債務の構成は劇的に変化	50
0.4 データの統治に関する法および規制の枠組みは、 所得別国グループの全体を通じて発達途上にある	13	2.1 データがもっている特定の特徴は 公的目的データの価値を最大化できる	55
0.5 1990 年以降、データ主導型サービスにおける グローバル貿易は指数関数的に増加してきており、 今やサービス貿易総額の半分を占めている	15	2.2 水へのアクセスの改善：ケニアでは、 壊れた手押しポンプの修理時間を削減するために センサーによるリアルタイム・データを利用	56
0.6 統合国家データシステムではどのようなことが 生じるか？	18	2.3 地理空間データセットにおける欠陥は 低所得国において特に大きい	61
1.1 国勢調査で把握される人口のシェアは 1850 年の 10 人中 1 人から現在の 10 人中 9 人に増加	26	B2.3.1 190 カ国について性別に分解されたデータ付きで 報告された新型コロナ患者の比率	62
1.2 データのライフサイクル	30	2.4 低所得国、特に脆弱性や紛争から 影響を受けている国は、比較可能な貧困データが 他の国グループよりも少ない	63
1.3 データがそれに沿って開発を促進しうる 3 つの経路	34	2.5 低所得国は他の諸国と比べて、国際的な 最善慣行になっている統計の基準や方法論を 順守している可能性が低い	65
B1.4.1 転用されたデータを利用した新型コロナ感染症の 研究：利用された私的目的データの種類別にみた 発表論文数	36	2.6 正のフィードバック・ループは公的目的データの 成功要因や特徴を、開発にとっての より大きな価値とつなげることができる	66
1.4 国および国際的なレベルにおけるデータ統治の層	39		

2.7	ほとんどの国は国家統計計画に対して資金手当てを完遂していない	67	5.7	1 ギガバイトのデータにかかる月額料金は低所得国では支払い負担不可能	169
2.8	ある一国の統計法が古いほど、その統計パフォーマンスとデータの開放度は低い	69	5.8	データ消費は市場価格とサービス負担能力に対して非常に敏感	170
2.9	NSO の独立性と報道の自由度が高いことは、統計パフォーマンスがより良いことと正の相関を有する	70	5.9	国内にデータ・インフラが存在することは国のデータ交換を円滑化する	171
2.10	データの需要と供給はデータの作成および利用について好循環 / 悪循環をもたらす	71	5.10	データ・インフラは低・中所得国では相対的に少ない	174
2.11	公的目的データの潜在力を実現するための政策	72	5.11	国は国内のデータ・インフラを段階的に開発	174
S2.1.1	国の所得水準別にみた女性性器切除の広がり (15-49 歳の女性: 2010-19 年)	85	B5.1.1	低・中所得国は ICT 専門家を教育してはいるが保持してはいない	178
3.1	生産プロセスにおけるデータの役割: 開発への経路	94	B5.1.2	ICT 専門家的大幅な賃金格差が特に低・中所得国において頭脳流出を引き起こしている	179
3.2	経済活動におけるデータの役割	96	S5.2.1	データの消費に起因する世界全体の温室効果ガス排出は定常を維持してきているものの、電力消費は増加しつつある	189
3.3	一部の低所得国においては多くのプラットフォーム企業が進出しているが、規模は小さい傾向にある	97	6.1	信頼のための法的枠組みの多元的な性質を描写	193
3.4	上位ウェブサイトに占める企業の割合で示唆されているように、内資系企業と外資系企業の重要性の程度は国ごとに異なる	98	6.2	データ保護の進化	194
3.5	利用者はフェイスブック、グーグル、およびユーチューブを訪問して、他のウェブサイトよりも長い時間を費やしている	98	6.3	サイバーセキュリティのための規制の枠組みの格差は、所得別の国グループ間において明白である	196
3.6	低・中所得国におけるインターネット・トラフィックはアメリカに本拠を置く数社に集中している	99	6.4	個人データ保護法制の整備は所得別の国グループごとに大きく異なる	197
3.7	プラットフォーム企業に由来する市場構造と市場支配力に与えるリスク	110	6.5	e- コマースおよび関連する法制の採択は所得別の国グループ全体にわたって広がっている	202
S3.1.1	アメリカでは公的データの民間企業による利用はアメリカで極めて価値があり、このことはオープンな政府データが有する価値を示唆している	120	6.6	公的目的データへのアクセスや公的データの再利用を可能にする規則の策定は所得別の国グループごとに不均一	203
B4.1.1	ジャカルタにおいてスマートフォンの位置データは利用者が自宅で過ごした時間の変化を明らかにしている	125	6.7	私的目的データ共有のためのイネープレアの採択は、所得別の国グループのすべてにわたり、公的データのそれに後れを取っている	207
B4.2.1	転用されたデータを利用した新型コロナウイルス感染症の研究: 利用された私的データの種類別にみた発表論文数	128	7.1	データ主導型企業の際立った特徴は、競争、貿易、および租税の分野において、政策当局に対して複雑な挑戦課題を提起している	230
B4.3.1	公的目的データと私的目的データは違法漁業を探知するために組み合わせることができる	133	7.2	デジタル経済では、旅客輸送に関連する反トラスト係争は中所得国では高所得国よりも広範にみられる	233
4.1	ネットワーク・カバレッジのギャップは農場規模によって異なり、農業普及指導サービスに悪影響を与えている	137	7.3	世界的な反競争的行為のなかでは支配力の乱用がデジタル経済の多数の部門共通により一般的	233
4.2	教育を受けた国がどこであろうと、AI の専門家はアメリカ市場に引き寄せられている	140	7.4	1990 年以降、データ主導型サービスにおけるグローバル貿易は指数関数的に増加してきており、今やサービス貿易総額の半分を占めている	241
5.1	データ・インフラの供給チェーン	160	7.5	越境データフローの取り扱ひに対する 3 つの相異なるアプローチ	242
5.2	開発途上世界ではデータへのアクセスには圧倒的に無線ネットワークが利用されている	162	7.6	東アジア諸国はデジタル・サービスに現行の VAT ルールを適用していないことによって、膨大な額の税収を失っている	251
5.3	3G の無線ブロードバンド・インターネットの利用可能範囲に関するギャップは縮小してきているが、利用ギャップは執拗に大きいままにとどまっている	163	8.1	データ統治とデータ管理は共にシームレスに機能して社会契約を支援する	270
5.4	世界全体として無線技術のカバレッジはその絶え間ない更新を反映	164	8.2	データ統治の機能	272
5.5	低・中所得国ではインターネットを使っていない人々の約 70% は、デジタル・リテラシー不足を理由に利用を控えている	167	8.3	独立的なデータ統治機関がある低所得国は無く、そのような機能を有する下位中所得国は少数; ほとんどは他の政府機関内に設置されている	275
5.6	国の所得別および地域別にみたモバイル・データ消費における不公平は著しく大きい	169	8.4	国の所得水準が低いほど、データ保護当局 (DPA) がある国は少ない	278

8.5	すべての所得別グループにわたり、各グループに属する国の半数以上が反トラスト当局を有している	278	9.3	公的部門を国家データシステムに統合するためのステップ	313
8.6	低所得国でサイバーセキュリティ機関があるのはわずか約4分の1	279	9.4	市民社会を国家データシステムに統合するためのステップ	315
8.7	有効なデータ統治に対してうまく機能している機関の特徴	283	9.5	学界を国家データシステムに統合するためのステップ	318
S8.2.1	アマゾン流域におけるオープン・サイエンスや市民科学の一連の多様なツールや技術の活用	302	9.6	民間部門を国家データシステムに統合するためのステップ	320
9.1	統合国家データシステムではどのようなことが生じるか?	308	9.7	国際的および地域的な機関を国家データシステムに統合し、国境を越えた協働を行うためのステップ	323
9.2	仮説的な国家データシステムに向けたデータ成熟度モデル	312			

地図

0.1	ガンビアにおける新型コロナ感染症関連の都市封鎖期間中(2020年3-5月)における、週毎の人の流れを追跡することを目的とした携帯電話記録集計値の利用	5	B4.1.1	2020年における、ジャカルタのスマートフォン利用者の自宅位置の地図	124
0.2	ナイジェリアにおける、衛生の改善が必要な地域の高度に精密なデータによるピンポイントでの特定	7	B4.2.1	転用されたデータを利用した新型コロナ感染症の研究：国別にみた公表された論文数	129
0.3	衛星画像と家計調査のデータを組み合わせることによって、タンザニアの貧困地図の解像度が高まった	10	4.1	私的目的データは、正式な銀行口座を保有していない成人数など、各国政府が収集していないユニークで比較可能な情報を提供することができる	135
0.4	データ・インフラは世界の全地域に普及しているわけではない	13	4.2	農業普及指導サービスは、多くの小規模農家にとってアクセス可能な遅くて古いブロードバンド・インターネット向けに調整可能	136
1.1	1854年のロンドンにおけるコレラ感染症に関するジョン・スノウの革新的な地図作製は疾病の追跡に革命をもたらした	27	S4.1.1	基本的な気象データに関するグローバルな報告には大きなギャップが残っている	154
B1.4.1	転用されたデータを利用した新型コロナ感染症の研究：国別にみた公表された論文数	37	S4.2.1	警察の報告書とクラウドソースのデータを組み合わせることで、研究者はナイロビ市における衝突の半数が発生している5%の道路を特定化することができた	156
2.1	貧困削減：クロアチアにおいて貧困地区の地図の作製は、貧困削減のための資金のより良い対象選定を可能にした	57	5.1	グローバルな海底光ファイバー・ネットワークは世界の隅々にまで行き渡っているが、データ・インフラの開発は不均等	172
2.2	公共の安全性を改善：ボゴタ市で市民が収集したデータは市内の自転車路周辺の安全性を高めることにつながった	59	7.1	国境を越えるデータのフローに関する規制モデルの採用	244

表

1.1	元々の意図と収集手段に基づくデータの種類の事例	28	6.2	データの安全装置とイネーブラーに基づき成熟度モデルに従って体系化された勧告	212
2.1	所得別の国グループ別にみたデータ開放性に関する評価	64	7.1	国境をまたぐ個人データの扱いを規制するための政策基準	244
2.2	所得別の国グループ別にみたデータ普及の実施と開放性	64	7.2	近年の貿易協定におけるデジタル貿易に関する主要な規定	247
4.1	私的目的データの開発向け利用を推進するために取り組むべき主要な研究における欠落	146	8.1	データのライフサイクルに沿ったデータ管理の決定	270
5.1	成熟度モデルに基づき配列したデータ・インフラ改善のための勧告	181	8.2	データ統治を遂行する、あるいはデータ統治に情報を提供する諸機関や主体の候補	274
6.1	規制枠組みの特定の要素の整備状況はさまざまであり、パフォーマンスは総じて低い	210	8.3	一般的なデータ仲介業者の構造についての寸評	280



概観

情報がなくてもデータは入手できるが、
データがなければ情報は入手できない

—ダニエル・キーズ・モラン, コンピュータ・プログラマーで
空想科学作家



データは前例のない割合で増加し続けており、至るところでほとんどの人々の日常生活の一部になりつつある。しかし、極度の貧困の中で生活している7億人以上の人々にとって、それはどのような問題をもたらしているだろうか？ データの新しい形態や利用が激増していることは、人々の生活を改善しているだろうか？ それとも貧困者や貧困国は取り残されて、この新たなデータ主導型世界の利益を得る人々とそうでない人々との格差を拡大させるのだろうか？

データの創造的な新しい利用から帰結する革新は、万人にとってこの時代における、人生を変える最大の出来事になりうる。蒸気機関や電気のような多くの汎用技術のように、データ革命がもたらす転換は社会や経済のあらゆる側面に関係する可能性がある。しかし、そういった全面的な変化は自動的に起こるものではない。蒸気機関や電気の生産性の面での価値が実現するまでには、それらが最初に導入されてから数十年間を要した。このような遅れが生じたのは人々がこのような革新の重要性を認識しなかったからではなく——遅かれ早かれ万人が認識することになった——、革新の経済的な潜在力を実現するためのこのような革新に必要な新しい生産システムが一夜では整備できなかったからである。電気それ自体は経済開発に帰結しなかったのと同様に、データだけでは福祉は改善しないだろう。データは社会や経済を改善することができるが、そうなるのは、データが生活を改善するような洞察をもたらす情報を創出するような仕方です系統的に使われる場合のみである。

本報告書では次の2つの根本的な疑問に答えることを目的としている。第1に、どのようにすればデータは開発目的をより改善することができるか？ 第2に、安全で倫理的かつ確実でありながらも、価値を公平に実現する仕方でのデータの創出および利用を支援するためには、どのような種類のデータ統治取り決めが必要か？

本報告書の1つの重要なメッセージは、単により多くのデータを収集することは答えではないということである。データの不足が、特に貧困国において顕著にみられるなかで、本書の目的は、特に貧困国の貧しい人々のために、開発成果を改善するためにデータをより効果的に利用するという方向に注目を移すことにある。

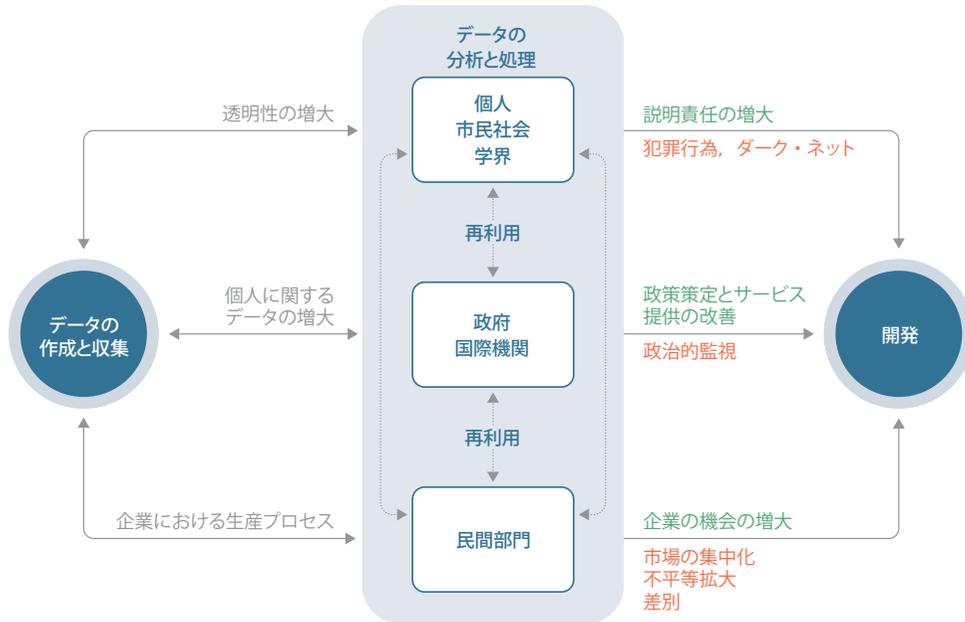
データを通じて開発目標を推進

本報告書のパートIでは、制度的な3つの経路を通じて、データと開発をつなげる概念的枠組みを提示している(図O.1)。中央の経路は、証拠に基づく政策立案やサービス提供の改善を支援するための、政府や国際機関によるデータの利用である。上方の経路は、政府による政策の効果をモニターするための市民社会による、そして公用および商用のサービスをモニターし、利用するための個人による、データの利用である。下方の経路は生産プロセスにおける私企業によるデータの利用——広範な経済成長に加えて自社の成長を後押しするための利用——である。概念的枠組みが示唆している1つの含意は、データだけでは開発問題は解決できないということである。すなわち、データを生計や生活を改善できる有用な情報に転換する中心的な当事者は、(社会、政府、および企業に属する)人々なのである¹。資本や、土地、労働と並んで、データも3つすべての経路に沿って出現してくる開発目標にとっての投入財である。しかし、資本や、土地、労働とは異なり、データは一度使っても価値は減らない。ある1つの意図をもって最初に収集されたデータは、まったく異なる目的のために再利用することができる(第1章)。

データの再利用や新たな目的での利用を増やすための普及、交換、および共有

生産的利用に資するデータの潜在性は本質的に無限

図 0.1 データはどのように開発を支えることができるか：変化の理論



出所：WDR 2021 チーム。

注：プラスのインパクトは緑色、マイナスのインパクトは赤色で表示。

であるため、もしデータが生活の改善につながるならば、データの再利用や新たな目的での利用を可能にすることが極めて重要である。したがって、それが概念的枠組みの中心的な局面になっている。図 0.1 ではこのような関係を示すために双方向矢印を用いている。民間部門と政府・国際機関との間の双方向矢印は、元来は商用目的で収集されたデータの、公的政策に向けた再利用および新たな目的での利用、およびその逆を示している。同様に、個人・市民社会・学界と政府・国際機関との間の双方向矢印は、これら当事者によって交換および再利用されるデータを示している。個人・市民社会・学界と民間部門をつなぐ双方向矢印は、個人・市民社会・学界による民間部門のデータおよびデータ主導型アプリケーションの利用と、個人・市民社会・学界が作成したデータや分析の企業による利用を反映している。しかし実際には、データの保有者はデータの交換に消極的かもしれない。データ保有者は、データの保護や安全性、あるいはデータ収集への投資に対する収益を確保することの必要性などに関して懸念を抱いている公算がある。あるいは、データ保有者は、規模の経済を獲得するためにデータを蓄積することから市場支配力を得る、あるいはデータを退職することによって別の種類の政治的あるいは競争的な優位を得ることを望んでいるかもしれない。

「共有と再利用」という用語は、本報告書では政府の

データ共有のためのオープン・データ・イニシアティブから、私企業を含むデータの市場ベースの取引に至るまでの、再利用を許容するデータのあらゆる種類の取引や交換について使われている簡潔な表現である。理論的には、データを巡って明瞭な経済的財産権を定義することは、データが市場で幅広く取引されることを可能にするだろう。しかし実際の活用においては、データ取引の程度(広告向け市場は除く)は、所有権を巡る競合する主張、データの広範な普及と私的で商業的な利益のためにより多くを蓄積することに対するインセンティブとの間の緊張関係、データの質や正確性を評価することの難しさなどによって限定されてきている。

図 0.1 で例示されている 3 つの経路は、データが生活を改善することができる方法を示しているが、その同じ経路は、人々に害を及ぼす仕方でもデータが使われうる余地も生み出している。政府の経路を介して、データは民族や、宗教、人種、ジェンダー、心身障害、性的嗜好などに沿って監視あるいは差別という政治的な目的のために悪用されることがありうる。個人間に通じている経路には、サイバー犯罪者が機密情報を盗んで操作することによって甚大な被害をもたらす可能性が潜んでいる。「ダーク・ネット」というのは隠匿されたウェブサイトの膨大な並列的なネットワークである。このネットワークは、広範な犯罪行為向けの地

下組織的なデジタル・プラットフォームを提供しており、麻薬や、贋金、盗難品、クレジットカード番号、偽造文書、武器、人の臓器などの違法取引を促進している。同様に、民間部門の経路を通じる、有害な利用の実例に含まれるのは、様々な例の中で特に、説得力のあるメッセージというマイクロ・ターゲティングに基づく攻撃的あるいは操作的な市場技術に従事するための、あるいは市場プレーヤーの間での共謀を促進するアルゴリズムを適用するための、消費者の好みや行動に関する情報の私的利用である²。

公益のためにデータを解放すると同時に不正使用に関して保護措置を講じる：新型コロナ感染症の事例

多くの国が新型コロナ感染症を制御するためにデータを利用してきている。このような利用にはロックダウン期間中における移動パターンをより良く理解する、あるいは感染者接触者追跡を支援するために、人々の位置情報を追跡することが含まれている。個人レベルのデータを伏せるために総計された2020年3-5月の呼詳細レコード(CDR)を使うことによって、ガンビアの政策当局は行政区境をまたぐ人々の移動を示す地図を検討することができた(地図O.1)。このような地図は、当局が都市封鎖が人流の削減に成功している程度を理解するのに役立ち、封鎖の順守と非順守に繋がりのある要因を特定し、対応策を練ることを可能にした。一方、イスラエル政府は、携帯電話から収集

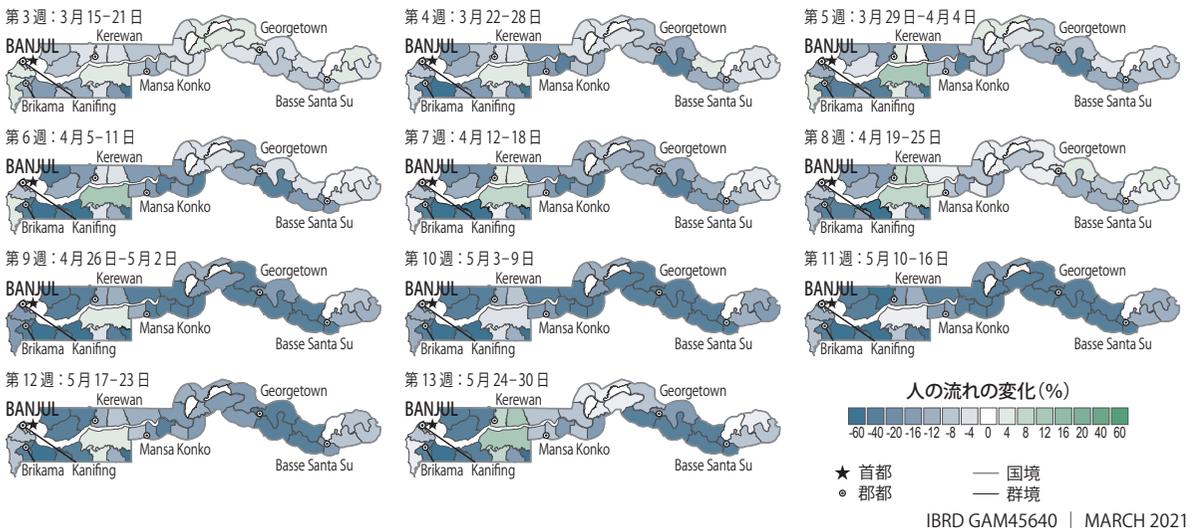
した個人レベルのデータを、人々を追跡し、そして接触の追跡調査を通じて新型コロナの拡散を抑えることに利用できるようにするために、2020年3月に緊急規制を承認した。

CDRは公的政策の策定を支援する、あるいは政府が人々の移動を追跡できるようにするために開発されたわけではないが、データが再利用や新たな目的に利用される事例を示している(図O.1における縦の流れ)。イスラエルでは、このようなデータは感染症流行の以前から収集されていたが、国家安全保障を目的とした場合にのみ利用可能であった³。

感染者の追跡という新たな目的のためにCDRを用いることにおける初期の努力はプラス効果をもたらしたようである。ガンビアでは、この地図はロックダウンが不釣り合いに貧困地域に影響を及ぼしたことを明らかにし、この地域を対象にした救援と回復に向けた取り組みの必要性を示していた。イスラエルでは、携帯電話のデータの分析は次のことを示唆している。すなわち、携帯電話データの利用は、感染症拡散の初期段階においては国内の新型コロナウイルス感染者の3分の1以上(感染者16,200人中の5,500人以上)を特定することにつながり、おそらく、このことが同国で初期における感染者数や死者数を例外的に低くすることに寄与したのであろう。

イスラエルで人口の大部分を追跡するためにCDRデータを新たに活用したことは、政府によるデータ不

地図O.1 ガンビアにおける新型コロナ感染症関連の都市封鎖期間中(2020年3-5月)における、週毎の人の流れを追跡することを目的とした携帯電話記録集計値の利用



出所：Knippenberg and Meyer 2020。データは http://bit.do/WDR2021-Map-O_1。

注：青色の影は人の流出を示す；緑色の影は人の流入を示す。全国的なロックダウンは2020年3月22日に実施。データは呼詳細レコード(CDR)を使って収集。

正使用の可能性を巡る論議や反発を引き起こした。イスラエルでは、多くの議員がプライバシー侵害の懸念を指摘し、最終的には最高裁判所がこのプログラムを停止した。2020年4月下旬、この最高裁は次のように裁定した。政府は携帯電話による追跡を法制化するとともに、「プライバシーの原則と両立する適切な代替的な方法を作り出さなければならない」⁴。

本報告書のテーマの多くがこの事例で示唆されている。民間部門のCDRデータを公的当局とともに共有および再利用することは、新型コロナウイルス感染症の抑制を後押し、そしてそのことによって生命が救済されることで社会的な価値を生み出す。同時に、このようなデータの転送は信頼に関する根本的な懸念を提起した。すなわち市民は、自分たちのCDRデータが次には政府官吏によって公衆衛生以外の意図していない有害な公算のある他の目的向けに転用されうることを懸念した。公平性に関する問題も危うい。イスラエルのような高所得国では、スマートフォンの普及率は93%に達していたが、ガンビアのような低所得国ではわずか75%にとどまっていた。それぞれの場合において、人口の中でスマートフォンを持っていない少数部分はCDRデータを生成することができず、そして、接触者追跡がもたらす公衆衛生上の保護から直接的な利益を必ず得るわけではないだろう。

このような事例はある1つの重要な難問も例示している。すなわち、政策やサービス提供の改善という形で人々が実現する潜在的な利益は、より多くのデータ、特に個人データが共有および再利用されることに

よって急増するかもしれないが、データが誤用されるリスクも高まる。このような潜在的な利益は、データの関係者間での普及ないし交換に依存している。しかし、関係者がそのような交換に進んで関与するためには、交換の安全性を下支えしているシステム、規則、および制度が信頼されていなければならない。

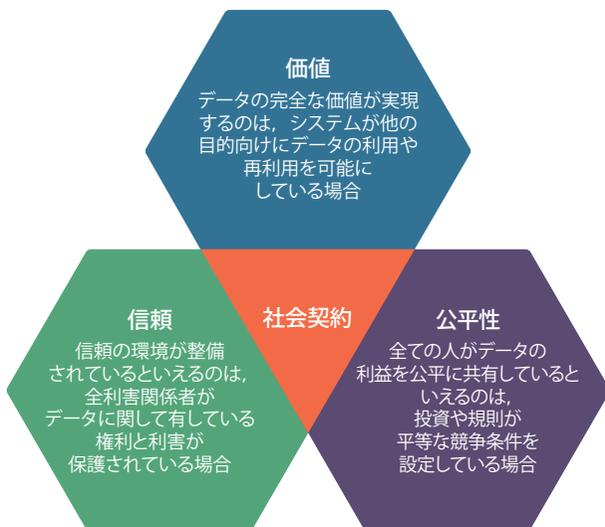
人々は自分に関するデータが保護され、データが生み出すことができる価値を共有することをどのようにしたら信頼することができるだろうか？ そのような懸念が増大しているという事態は、データを巡る新たな社会契約の必要性を示唆している。すなわち、そのような社会契約は、データを作成、再利用、および共有するプロセスの全ての関係者相互間の合意であり、そのような合意は、関係者がデータの交換から害を受けることはなく、そしてデータによって生み出される価値の部分は公平に生じる、という信頼を育むものである(図0.2)。社会がこの種の合意、すなわち社会契約を締結するという考えは、トーマス・ホプスや、ジョン・ロック、ジャン-ジャック・ルソーなどの哲学者の著作にしばしば結び付きながら、数世紀にわたって存在してきている。

法制度、より一般的には統治を、社会契約を締結、促進、そして執行するための道具としてみることができる。社会契約のルールに従うよう関係者を説得するのは容易なことではなく、それはデータ利用から得られる利益が公平な仕方でも共有される——すなわち全ての人が何らかの利益を得る——のを保証することにかかっている。その過程においては、低所得国はあまりにもしばしば不利な状況に置かれている。というのは、これらの国は多くの場合に、データを獲得して価値に転換するためのインフラやスキル、データシステムに対する信頼を生む制度や規制上の枠組み、グローバルなデータ市場やその統治に公平に参加する規模と担当機関を欠いているからだ。

データがわれわれの生活や、社会、より一般的には世界を作り変えるのに伴って、国内的にも国際的にもデータにかかわる社会契約が必要とされている。このことは、とりわけデータの取引や流れの国境をまたぐという性質が理由である。スポットライト8.1では、社会契約に関するこのような考えを国際的な分野にまで拡張し、グローバルな公共財として、また、公平で持続可能な開発を達成するための資源としてデータが保護されるのを確保するための国際的なコンセンサスを求めている。

データが秘める未活用の潜在力、データの作成・利用・再利用にかかわる法律・規制・統治の枠組みの発展、適切な枠組みの策定における各国の状況(歴史や、

図0.2 データに関する社会契約は価値、信頼、および公正性に基づく



文化、統治、政治経済学など)の重要性、ほとんどのデータを安全に作成するための技術的な能力の役割、そしてデータの価値の信頼性とより公平な共有の必要性など、これらすべてが今回の『世界開発報告』の中心的なテーマである。

本報告書のパート I は以下のことの開発への潜在的な影響をより詳しく描写することで始まっている。それは、公的目的のために収集されているデータ——公的目的データ(第 2 章)、ルーティンのビジネス・プロセスの一環として民間部門が収集するデータ——私的目的データ(第 3 章)、そして、異なる型のデータの共同利用から生じる相乗効果(第 4 章)である。公的な目的のデータと私的な目的のデータの間の区別は、データの収集者、あるいはデータ収集に用いられる手法(消費者調査、会計記録、デジタル取引など)とは無関係である。

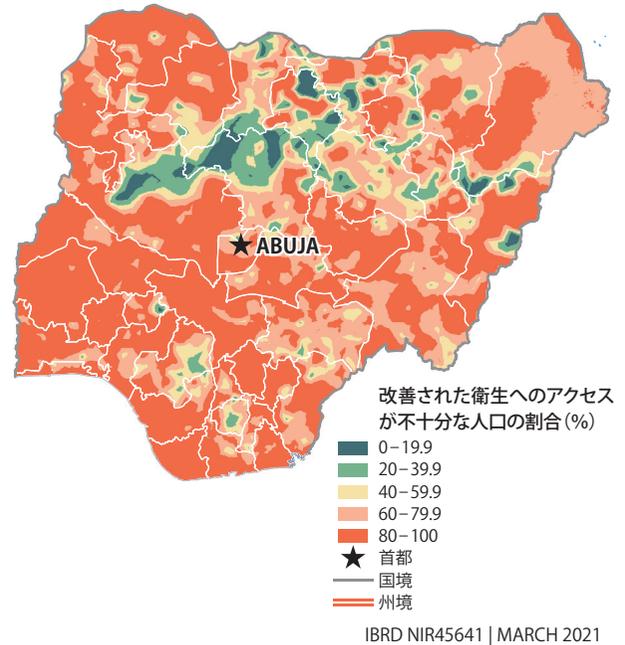
公的利用向けのデータはサービス提供、目標設定、説明責任、およびエンパワメントを改善する

公的な利用が意図されたデータ(公的目的データ)には、公的なプログラムや政策の設計、執行、そして評価をする大きな潜在力がある(第 2 章)。公的利用を目的としたデータは多くの政府機能にとって前提であるため、これらのデータの最も重要な作成者は政府機関である。そして国勢調査や行政上のデータ収集などといった手法を用いて作成されている。市民や、市民社会、非政府組織、学術機関、国際機関などは、調査、クラウドソーシング・プラットフォーム、およびその他の手法を用いて公的な目的のデータ作成に重要な貢献をしている。

この種のデータは次の 3 つの主要な経路を通じて生活の改善につながりうる。すなわち、第 1 に、政策立案やサービス提供を改善することによって、第 2 に、稀少な資源に優先順位を付けて、それを排斥されている人々や地域に届けるために対象を絞り込むことによって、そして第 3 に、より多くの情報や知識を通じてより良い選択をするために個人をエンパワメントすることによって、生活の改善をもたらさう。

ナイジェリアから得られた事例は、サービス提供の改善や絞り込みを行うことについて公的データが有する力を例証している。ナイジェリア政府が委託した『2015 年全国水供給・衛生調査』では、家庭や、給水所、水道網、公共施設(学校や医療機関を含む)などからデータが収集された。このようなデータは、次のような状況を明らかにした。ミレニアム開発目標によって設定されている衛生基準を満たしていない人口が 1 億 3,000 万人(当時の総人口の 3 分の 2 以上)に

地図 0.2 ナイジェリアにおける、衛生の改善が必要な地域の高度に精密なデータによるピンポイントでの特定



出所：World Bank 2017。データは http://bit.do/WDR2021-Map-O_2。

注：改善された衛生へのアクセスが不十分な地理的なホットスポットが、衛生にかかわる国際的な基準を満たしている地域の人口比率という観点でみた厳しさに応じて、最低(■)から最高(■)までの段階階級に表示されている。

達しており、清潔な水へのアクセスが不十分なことが貧困層、および特定の地理的領域で特に問題化している(地図 0.2)⁵。このようなデータに基づいた報告から明らかになったことへの対応において、ムハンマド・ブハリ大統領は当該分野について非常事態宣言を発し、「ナイジェリアの水・衛生(WASH)部門の再生に向けた国家行動計画」を立ち上げた⁶。

データの質が高いほど(時宜性や、正確性、解像度などといった特性の点で)、開発に向けた価値を創造する潜在力は大きくなる。にもかかわらず、さまざまな要因が、各国——特に低所得国——が公益のためにデータからより大きな価値を引き出すことを阻害している。このような障害は、財源や、技術的な能力、データ統治、データがもたらす情報に基づく意思決定の要請などの欠如である。本報告書の一環として発表された、世界銀行の統計パフォーマンス指標は、公的データについて 166 カ国横断的に、時宜性、精度、相互運用のしやすさ、アクセスのしやすさに重点を置いて、入手可能性、質、および使いやすさに見られる欠陥を確認している⁷。

公的な目的のデータの潜在力を完全に引き出すには、政策プロセスにおいてデータに関して明確な優先

順位付けを行う必要がある。そうすることで、政府は高質なデータの作成、および意思決定のためにデータのオープンで透明な利用について優先順位付けを行うことになるだろう。公的な統計の透明性と信頼性は、政府の政策措置に対する信頼を築くのに役立つ。国の債務負担を公開しないとといったような透明性の欠如は、有害な経済的結末をもたらし、政府に対する大衆の信頼を傷付ける可能性がある(スポットライト1.2)。データの潜在力を完全に発揮するためには、データに関連する長期的で安定した財源、統計や技術に関する能力に対する投資、そしてデータの安全な作成と再利用に好意的な法律などが必要であろう。取り組まなければならない他の領域に含まれるのは、データの需要に影響を及ぼしているデータ・リテラシーの水準の低さ、データを利用することに対する政府当局者の動機や関心の欠如、公的な目的のデータの質に対する信頼の低さ、そしてアクセスやデータ利用にかかわるインフラの欠如などである。このような投資やイニシアティブは相互に依存する関係にあり、それ故、1つの分野で成功できないと、データが開発にもたらさる価値全体が危険にさらされる。データを効果的に利用することは、データに対するより一層の需要を生み出し、したがって、より多くのより高質なデータを作成するための投資を正当化するだろう。

私的な利用向けのデータ(私的目的データ)は成長に拍車をかけ開発を促進しうる

商用目的で民間部門が収集および整理したデータも、開発に拍車をかける大きな潜在力を有している(第3章)。企業によるデータの利用と応用における革新は、データ主導型の意味決定の能力を高め、取引コストを削減することによって、莫大な経済価値を生み出しつつある。アメリカの大手企業179社に関する2011年の研究は、データ主導型意思決定を採用している企業は、他の投資や情報技術を利用した際に期待される場合と比較して、生産性を5-6%増加させたことを示している⁸。

データは企業の生産プロセスに様々な仕方で投入されるものの、最近における新しいデータの爆発的増加は、企業経営のデジタル化、個人による携帯電話の利用、デジタル化された取引、そしてソーシャルメディアでのやり取りなどといった経済活動の副産物として生じてきている。このようなデータは、高頻度で収集されて、個人や、企業、経済的成果、現象などに関する詳細な情報を提供することができる。このようなデータは企業自身の経済的効率性を高めるだけでなく、新型コロナ感染症の追跡などのような公的な政策

での必要という新たな目的のために再利用できる潜在力も提供している。例えば、金融サービス提供者は、銀行業務における潜在的な借り手に関するデータの不足(より具体的には、非対称的な情報)という積年の問題を解消するために、代替的な信用スコアリング手法の採用を増やしつつある。このようなテクニックは他の仕方では収集できない、利用者の信用力を評価するために、ユーザーのデジタル・フットプリントを活用している。このアプローチの2つの目立った事例は、フィリピンで営業しているレンド(Lenddo)と、アフリカや、アジア、ラテンアメリカで営業しているシグニフィ(Cignifi)である。

しかし、このようなトレンドには新たなリスクが伴っており、データ主導型経済が社会的厚生を高めることを確実なものにするために、そのようなリスクに取り組まなければならない。過剰なデータの収集、私企業によって保有されているデータの不十分なガバナンス、個人データの不適切な保護、などに関して懸念が高まりつつある。このような懸念の多くは個人データの不正利用を巡って生じている。そのような誤用には、企業が顧客の金融情報を適切に保護していない—利用者資金や身元を盗用にさらしている—、ないしは個人の健康ないし所在に関する機密データを企業が無許可で利用、あるいは保護していない、といったことが含まれる。

企業が自ら所有しているデータで価値を創造するプロセスの多くは、アルゴリズムや機械学習によって牽引されている。このようなモデルにおいて、アルゴリズムは、個人がどのような情報、製品、あるいはサービスに、どのような価格で接しているか、どのような保険商品のパッケージが個人に提供されているか、個人の融資申請が承認されるか否か、個人がどのような仕事に向いているか、医療面で個人がどのような助言を享受するか、などを決定する。

この種の活動のすべてが経済効率性を著しく改善する潜在性を有している。例えば機械学習は、より多くのデータの種類を利用して、一見では無関係に思われるパターンから関連情報を抽出することによって、より多くの個人についてより正確な信用スコアを生成することができるだろう。しかし、もし機械学習に読み込まれるデータに差別的な前提が埋め込まれていれば、機械学習は差別を増幅して、有害な結果を生み出すだけでなく、結果を拡大するだろう⁹。この重要な点は、データ科学で昔から言われている「ゴミを入力すれば出力されるのはゴミ」という格言を想起させる。すなわち、これは、機械学習のようなデータ処理システムは処理のために与えられたデータそのものよりも良いという

わけではない、ということの意味している¹⁰。しかし、より深刻な懸念がある。機械学習の出力結果は典型的には曖昧であり、新データが新システムに入力されるたびに頻繁に変化する。機械学習が不透明なルールを生成するのは、ほぼ設計による。それ故、アルゴリズムの差別的な要素の特定は技術的に非常に挑戦的になりうる。

多くの場合に、データ主導型市場はネットワーク外部性がプラスであることを示しており、このことは、規模に関する収益の逓増と、少数の大企業が支配的になる傾向をもたらす。結果として、地元の起業家活動にとっては有害なことに、小規模または伝統的な企業が排除されることとなり、消費者の厚生にとってはリスクになる公算がある。このような効果は発展途上の市場では一層悪化するかもしれない。このような市場においては、新規参加者はスタートアップのための資本調達がより困難であることを認識する。さらに、このような市場ではデータ科学に長けている人的資本は限定されている。このことに対応するために、政策当局は規模の実現を制約している基本的な要因に取り組むことができる。このような制約の要因には、ジオブロッキング(利用者の地理的場所に基づくインターネット・コンテンツへのアクセスの制限)や、各国相互間におけるデータ政策の調和の欠如などがある。政策当局は、分野ごとの規制や政府の支援制度が全企業にとって公平な競争条件を提供するのを確保すべきである。

データを組み合わせることや、データを新たな目的に再利用することは、開発へのインパクトを深化しうる

異なる種類のデータを組み合わせ、そして新たな目的に再利用することは、データが開発に与える影響を高めることができる(第4章)。開発問題は複雑であり、経済や、文化、環境、人口動態、その他の多くの要因が広く関係している。1つの要素のみが対象となっているデータに基づく政策設計は不完全なものになり、時には思慮不足に陥るだろう。異なる種類のデータを組み合わせることは、データのギャップを埋めて、開発問題に関する新たな視点を提供することができる。

一例として、公共目的の家計調査がある。これは、生活水準や、消費、所得、支出などに関する広範なデータを収集し、ほとんどの諸国で貧困率を推定するためのベースになっている。調査という手段は非常に広範囲にわたり管理に時間がかかるので、標本は比較的小さくなる傾向にある。貧困率の推定は通常は、国全体および一部のより詳細な地理的階層では統計的に妥当であるが、そのような家計調査が、村ないしそれ以下のレベルの地域を対象とした貧困の削減を政策が実行

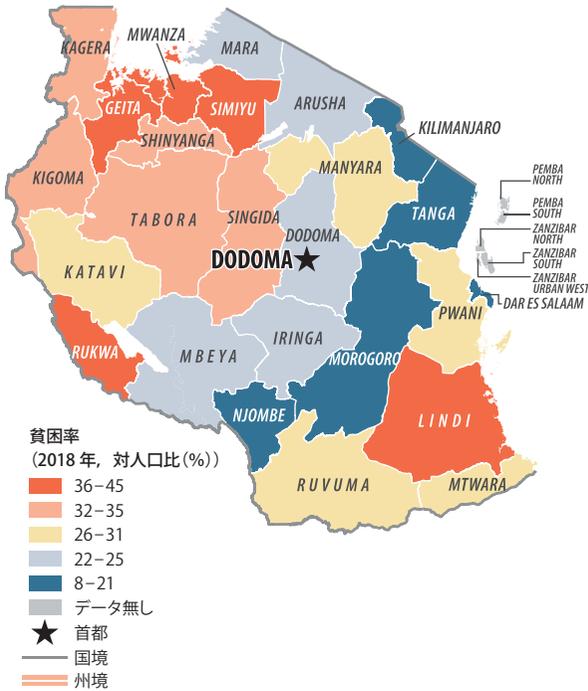
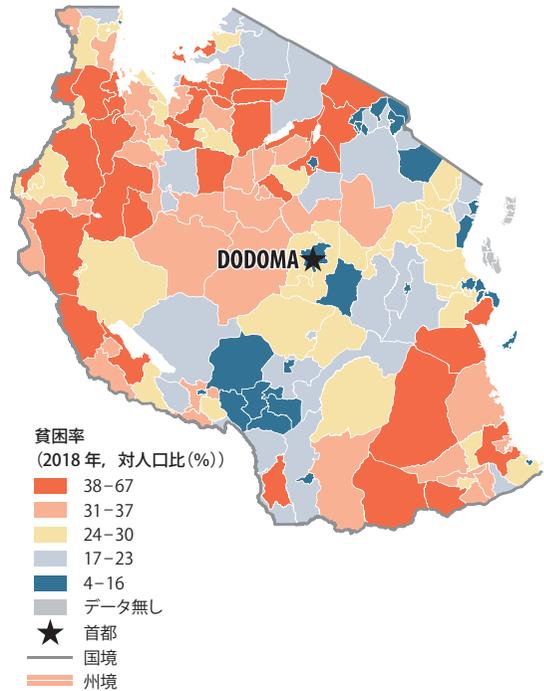
できるような、貧困に関する詳細な状況を提示するように設計されていることはほとんどない。一方、長年にわたって、調査データから貧困のモデルを推定し、このモデルを使って国勢調査データに基づき地図を描くことによって、解像度の高い貧困地図が作製されてきており、国勢調査データにおけるすべての家計に関して貧困状態を推定することが可能となっている。このアプローチの問題点は、国勢調査データが10年に一度しか入手可能でないことである(多くの貧困国では、頻度はさらに低い)。

このアプローチを修正したものが、国勢調査による人口を、CDR データないし様々な種類のリモート・センシング・データ(典型的には人工衛星あるいはドローンによる)で代替している。CDR または人工衛星データのこのような新たな目的での再利用は、解像度と時宜性がより高い貧困地図の提供を可能にしている。例えば、タンザニア政府は、家計調査データを使うだけでは、大陸の20州についてしか貧困水準を図示できなかった。ひとたび、家計調査データと衛星画像データを組み合わせると、タンザニアの169地区の各々について貧困状況を推定することが可能になった(地図0.3)。2つのデータ源を組み合わせることは、本質的に正確性を失うことなく、貧困地図の解像度を8倍に高めた。このような革新的な分析に関する他の実例は、アフガニスタンやルワンダなどといった、世界で最もデータが不足している環境下にある数カ国についても見られつつあり、データの欠陥を埋めるための解決策を提供している¹¹。

データを新たな目的のために再利用する他の方法には、オンライン・メディアとユーザーが作成したコンテンツを使って、水管理と食の安全保障のために水/洪水に関する事件をリアルタイムで地図に示すことや、作物収量をモニターし栄養不良を予測するために、官民が提供する衛星画像データを組み合わせることなどがある¹²。同様に、本報告書で示されている多くの実例は、公衆衛生(疾病の拡散を含む)の監視、危機への対応や資源配分の管理、輸送や交通における道路の安全性の確保、違法な漁業・森林伐採の監視などの分野におけるプログラムや、政策、成果を改善することに対するデータの新たな目的への再利用の潜在力を強調している。

データを作成および利用する新しい手法は、市民社会が政策について政府の責任を問い、腐敗に対する監視を改善することを可能にする¹³。例えば、クラウドソーシングによるデータやウェブ・スクレイピング(ウェブサイトからのデータ抽出)を利用することによって、ソーシャルメディア・ディスカッション

地図 0.3 衛星画像と家計調査のデータを組み合わせることによって、タンザニアの貧困地図の解像度が高まった

a. 家計予算調査を利用した貧困地図
(20 州)b. パネル a のデータと衛星画像を組み合わせた貧困地図
(169 州)

出所：World Bank 2019. データは http://bit.do/WDR2021-Map-O_3.

IBRD TAN45642 | MARCH 2021

ン・ボードが、地方の指導者が腐敗した官吏に対して行動を起こし、腐敗に反対する政策のインパクトに関してリアルタイムの反応を受け取る方法として出現しつつある。2011年にインドで「私は賄賂を払った」というオンラインの構想が Janaagraha Centre for Citizenship and Democracy によって立ち上げられ、世界最大のクラウドソーシングによる腐敗に反対するプラットフォームに発展した。このツールは腐敗行動に関する市民からの報告を収集して、それを問題地域を強調するために地理空間情報と統合する。そうすることによって、このツールは腐敗した行動と戦うために個人や、市民社会、政府をエンパワメントしている。

データ源の新たな目的への再利用および統合に向けた努力を後押しするために、本報告書ではこのような新たなデータ源を公益のためにうまく活用するのに必要とされている、人々や、パートナーシップ、調査に対して、ドナーや、政府、企業が投資するのを可能にする方法を提示している。低所得国が政策イニシアティブと投資を重要視すべき分野は、アナリストや意思決定者のデータ・スキルの構築、データの科学と分析を含んだ高等教育の拡充、高所得国の大学や私

企業とのパートナーシップの促進、政府上級リーダーのデータ・リテラシーの強化、政策立案に際して洗練されたデータや証拠の利用を促進する制度的環境の整備、そして私的な目的のデータを利用して非伝統的な役割を遂行することを目的とした国の統計局の刷新、などである。

データに関する統治を社会契約に適合させる

うまく設計されたデータの統治の枠組みは、国が公的目的データと私的データの両方がもつ経済的および社会的な価値のすべてを捉えることを可能にし、両者間の相乗作用を活用する。このことには、データシステムの誠実性に対する信頼を生成することが含まれる一方で、データの利益が公平に共有されることを確保することも含まれる。そのような枠組みは、一国のデータを巡る社会契約の現実的な表現である。

本報告書のパート II は、データに関する統治を構築している要素を説明している。そのような要素は、データの潜在的な利益を実現することを可能にする一方で、有害な結果を防止できる(図 0.3)。このような要素に含まれるのは、データ・インフラ政策(第 5 章)、データを取りまく政策・法律・規則(第 6 章)、関係

図 0.3 国内および国際的なレベルにおけるデータ統治の層



出所：WDR 2021 チーム。

する経済政策(第7章)、データの統治を行う機関(第8章)である。

データ統治の多くは国内に焦点を合わせているが、データ統治における多くの挑戦課題の効率的で公平な解決は国際的な協力があって初めて可能である。独禁法関連の決定がもたらす国境をまたぐ波及効果を管理し、サイバー犯罪と戦うことを目的として力を合わせるためには、二国間での取り組みが必要である。グローバルなただ乗り問題(データ保護主義やデータによって可能になるサービスにおける脱税など)に取り組んで、データの保護と相互運用性のために法律や技術基準の調和を通じて取引コストを削減するには多国間協力が必須である。同時に、地域的な協力はグローバルなデータ統治交渉において低・中所得国の発言を増幅するのに役立つことができ、データ・インフラの開発において規模の経済を達成するのを支えることができる。

データ・インフラの改善は貧困国における貧しい人たちにとっての公平なアクセスを確保することに役立つ

現代のデータが有するデジタルな特性は、デジタルなインフラを要請している——データの収集、交換、

保存、処理、および配布の前提条件である(第5章)。にもかかわらず、そのようなインフラの利用可能性は国内と各国相互間の両面で、不平等が顕著である。データ・インフラの社会的および経済的な価値はインターネットに接続する市民の数が増加するのに伴って急増することから、サービスの導入を促進する普遍的なサービス政策が長きにわたって存在している。ブロードバンド接続性が個人と国の両方に転換的な機会をもたらすことを認識して、国連のブロードバンド委員会は国際社会が2025年までにブロードバンド・インターネットを利用する人の割合を人口の75%にまで引き上げることを公約した¹⁴。

とは言え、普遍的アクセスに移行する努力は根本的なチャレンジに直面している。第1に、携帯電話に関わる技術の革新は継続的であるため、普及率の達成目標は動的である。2018年には世界人口の92%は信号強度3G(毎秒40メガバイトのデータ転送速度)の範囲内で生活していたが、その割合は4G技術(開発を促進できる洗練されたスマートフォンのアプリケーションが必要とする毎秒400メガバイトのデータ転送速度を提供)になると80%に低下した。最近の

一握りの最先端市場における5G技術(毎秒1,000メガバイトのデータ転送速度)の商業的利用開始は、低所得国をさらに後方に取り残すリスクがある。極めて重要な周波数資源についてより有効な管理手法を導入しつつ、政策当局は下支えになる光ファイバー・ネットワークへの民間部門の投資を支援する環境を整備することによって技術の格上げを迅速化するすることができる。インフラの共有も格上げのコストを大幅に削減しうる。しかしながら、ブロードバンド提供にかかわる競争を可能な限り促進することと、1つ以上のインフラ・ネットワークを支えるには需要があまりに限定的な一部の市場におけるサービス提供者の間での協働を促進することとの間で慎重にバランスをとることが必要である。

第2の挑戦課題は、世界全体で、データ・サービスを利用していない人の40%は、ブロードバンド信号が届く範囲で生活していることである。インターネットにアクセスしていない低・中所得国の人々の3分の2以上は、調査のなかで、インターネットの意味や使い方を知らないと回答しており、このことはデジタル・リテラシーが重要な問題であることを示唆している¹⁵。支払い能力も低・中所得国において利用率が低いことの要因であり、初歩的なスマートフォンでさえ、そのコストは最下位20%層の家計の月当たりの所得の約80%に相当している¹⁶。相対的に高い税金と物品税がこの費用をさらに押し上げている¹⁷。革新や、競争圧力、健全な政府政策などに対する反応としてコストが低下するにつれて、インターネットの利用率は増加するであろう。しかし、インターネット利用者間でさえ、データ・サービスの消費は1人当たり毎月わずか0.2ギガバイトにとどまっている。これは本報告書が基本的な社会的および経済的な機能をオンラインで遂行するのに必要であろうと推定している水準のほんの一部でしかない。

接続性の拡大における第3の挑戦課題は、そのことの地球温暖化に対する潜在的な影響である。接続性向上の気候への影響は一連の複雑なトレードオフを示している。2018年時点で、データ・インフラを支えるのに必要な電気は世界の総消費量の約1%に等しい——環境に影響するかなりの規模といえる。しかし、再生可能なエネルギー支援型のデータ・インフラに対する依存やエネルギー効率の改善によって、データ・インフラに起因する温室効果ガス排出は他の部門との比較では著しく低い。加えて、データ・インフラへのアクセスは、新型コロナウイルス感染症拡散期における旅行の著しい減少やビデオ会議の増加などによって例証されているように、気候変動に対して顕著なプラス効果を

与えることが可能であろう(スポットライト5.2)。

データ主導型経済への完全な参加は、個人としての市民の相互接続だけでなく、国レベルでの十分なデータ・インフラの整備も伴っている。ほとんどの場合、低・中所得国は、自国で生み出されたデータを交換(インターネット相互接続点(IXP)を経由して)、保存(コロケーション・データセンターで)、および処理(クラウド・プラットフォーム上で)するのを可能にする国内施設を欠いている(地図O.4)。その代わりに、多くの諸国は海外の施設に依存し続けているため、大量のデータを自国と外国の間で双方向に転送する必要があり、速度と価格の両面で大幅なペナルティを負担している。

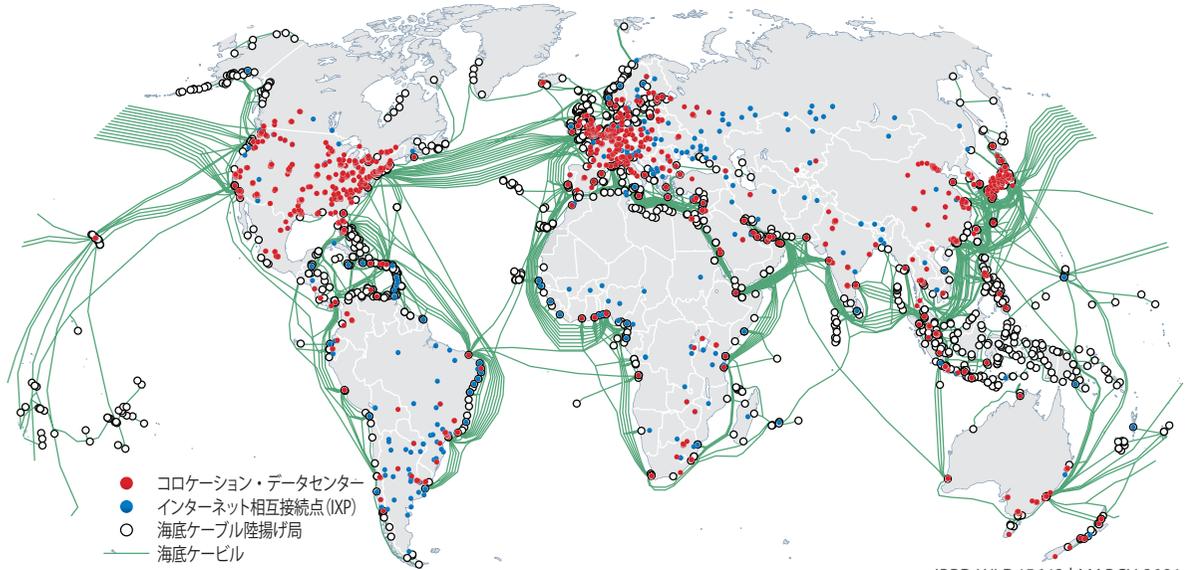
政策立案者はデータ・インフラへのアクセスを漸進的に改善するために多くのことができる。このプロセスは国内IXPの創設を奨励し、コロケーション・データセンターに適した投資環境を育成することによって始まる。このようなセンターでは、人気のあるインターネット・コンテンツは自国で保存することができ、海外のクラウド・インフラへのアクセスは、オンランプの提供を通じて円滑化することができる。各国間に適切な光ファイバーによる接続性が存在し、規制面で十分な調和が図られているような地域のレベルでは、そのような施設は共有することができ、データ・インフラについては、データの炭素足カーボンフットプリントに関する懸念に加えて、極めて高水準の信頼性が必要とされるため、理想的な民間部門の投資環境としては、信頼できるクリーンで低コストの電力、自然冷却、災害リスクがほとんどないといった条件が提供されるべきである。このような条件は、低・中所得国では常に容易に満たされているわけではない。

データ関連の法と規制は信頼の環境を整備することに役立つ

データ取引における信頼は安全装置(safeguard)とイネーブラー(enabler)の両方を包含する強固な法規制の枠組みを通じて支援されうる(第6章)。このような枠組みの確立は所得別の国グループ全てを通じていまだに道半ばにとどまっている(図O.4)。

安全装置はデータの不正使用から生じる害を回避あるいは制限することによって、データ取引に対する信頼を促進する。データシステムの信頼性に対する基本的な前提はサイバーセキュリティである。十分なサイバーセキュリティを達成することは、データの管理者と処理者にデータの安全を確保するための技術的なシステムを採用することを義務付ける法的枠組みの創出を要求する¹⁸。現在までのところ、サイバーセキュリティに対する十分な法的枠組みを採用している低・

地図 0.4 データ・インフラは世界の全地域に普及しているわけではない



IBRD WLD45643 | MARCH 2021

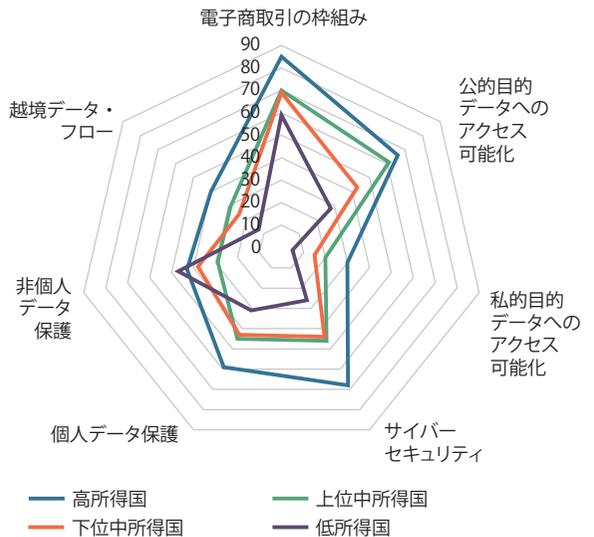
出所：PeeringDB, Interconnection Database, <https://www.peeringdb.com/>；PCH Packet Clearing House, Packet Clearing House Report on Internet Exchange Point Locations (データベース), 2020年12月14日にアクセス, <https://www.pch.net/ixp/summary>；TeleGeography, Submarine Cables (データベース), <https://www.submarinecablemap.com/>. データについては、以下を参照 http://bit.do/WDR2021-Map-O_4.

中所得国はごく少数にとどまっている。ケニアの新しいデータ保護法は包括的なサイバーセキュリティの規定の好例として際立っている。

データ保護のために十分な法的枠組みを整備しておくことも決定的に重要である。そのような枠組みは、個人データ(個人を特定できるデータ)と非個人データ(個人が特定できる情報を含まないデータ)を明確に区別すべきである。中所得国のなかでは、モーリシャスは、個人データに関して比較的適切に開発された保護措置を整備していることで有名である。実際に、モーリシャスはサハラ以南諸国のなかで、ヨーロッパ評議会の「個人データの処理にかかわる個人保護に関する条約」(108号条約)を批准した最初の諸国の1つとして際立っている¹⁹。

個人データの保護は、国際的な人権法に基づいている。この法は、あらゆる種類のデータ取引であろうとも、それが可能となる以前に、データの利害が十分に保護されていることを求めている²⁰。このような保護は、通常はデータの当事者がそのデータの利用について何らかの形で明示的な同意を示すことを義務化することで達成されている。しかし、そのような同意は有意義であろうか？ 証拠は次のことを示している。閲覧したウェブサイトやアプリケーションの各々について同意を求める膨大な開示書類を平均的な人が読みこなすには、1年間当たり76日かかる²¹。この発見が示唆しているのは、データが使用される顧客にとつ

図 0.4 データの統治に関する法および規制の枠組みは、所得別国グループの全体を通じて発達途上にある



出所：本報告書だけのために実施された Global Data Regulation Survey (<https://microdata.worldbank.org/index.php/catalog/3866>)に基づく WDR 2021 チーム。データは http://bit.do/WDR2021-Fig-O_4。

注：本図は各所得別の国グループのなかの調査対象国で導入されている善償行といえる法規制の比率を示している。

て最も都合が良くなるようにデータ・サービス提供者が行動するよう法的義務を強化する必要があるということである。

非個人データは性質的に機密性が低いので、ほとん

どの場合、知的財産権を通じて十分に保護されている。そしてこのことは、データ保護とデータ再利用の間で利害のバランスをある程度図ることを可能にしている。しかし、本報告書は次のことを見出している。調査したほとんどの低所得国では私的な目的のデータに関しては知的財産権が整備されていない。

問題をいっそう複雑にしているのは、個人データと非個人データの区別がますます曖昧になりつつあることである。このような曖昧化が生じているのは精緻なアルゴリズムを使って、さまざまなデータ源を幅広く混合および加工しているからである。そのようなアルゴリズムは、非個人データ（携帯電話から得られるデータ）を個人を特定できるようにする、ないしは少なくとも個別の社会的グループを特定することができるようにするかもしれない。

イネーブラーは、データの完全な社会的および経済的な価値が発揮されるのを確実にするために、利害関係者グループ内や同グループ相互間でのデータへのアクセスと再利用を円滑化する。データ共有を支援する規定の性質とその程度は公的データと私的データとは著しく異なる。世界全体でオープンデータ政策（政府データの積極的な公表を奨励）を通じて、公的データを開示することにに向けた多大な努力が払われてきている。このことには、情報に関する法律へのアクセスも伴っている（開示の強要にかかわる法的に執行可能な権利を市民に付与）。しかし、実際に影響力が発揮されるためには、オープンデータ政策は機密データを区別するための首尾一貫したプロトコルによって支えられていなければならない。また、将来的な再利用を円滑化するために、相互運用可能な技術基準や、機械で判読可能な様式、オープン・ライセンスと組み合わせる必要がある。

私的データの開示については政府の影響力はずっと小さい。そのようなデータの共有は、エジプトにおけるように、市場支配力の集中に対する是正策として機能するかもしれない。同国では配車サービス大手2社の合併申請は、運転手の情報を小規模競争相手と共有することが条件とされた。他の状況では、私的な目的のデータは重要な公的政策の挑戦課題に取り組むことにおいて決定的に重要なものにもなるかもしれない。一例として、新型コロナウイルス感染症の拡散を抑制することを目的とした、接触者追跡のための携帯電話記録の活用がある。しかし、オープンライセンスや、データ・ポータビリティ、各種のデータ・パートナーシップなどの措置を通じて、私的データの交換を促進することの可能性について、これまでのところ、相対的にほとんど関心が払われてきていない。し

かしながら、一部の国——特にフランス——では、公益に資すると考えられる民間部門のデータの共有を義務化する法律が制定されている²²。

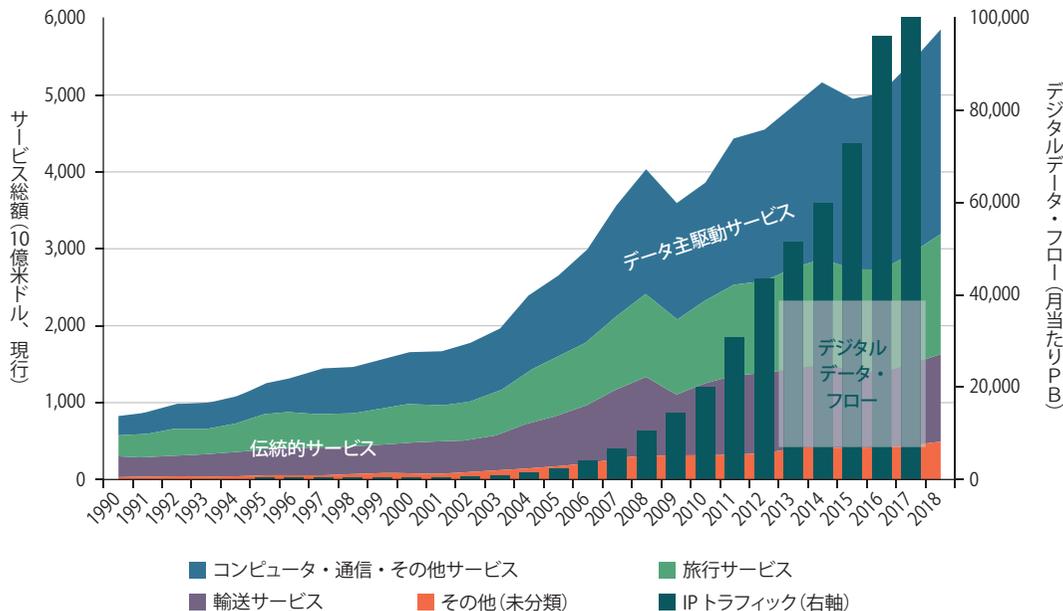
データ規制を経済政策の目的に整合させることは、価値の創造を下支えすることができる

急速に拡大しつつあるプラットフォーム・ベースのビジネス・モデルにおいて、データは中心的な役割を果たしている。例えば、検索エンジンはユーザーのサイト閲覧に関するデータを収集しており、そのデータをマーケティング会社に販売することができる。マーケティング会社は、そのデータによって広告の対象をより正確に絞り込むことができるようになる。このようなプラットフォーム・ベースのビジネス・モデルが低・中所得国ではより一層重要になりつつある。データに関する法および規制の枠組みの設計は、データ主導型ビジネスの増大に実質的な影響を与えるため、むずかしい政策バランスが必要とされる。例えば、不可欠な市場データ源に対するアクセスを供与することは、プラットフォーム企業間の競争促進にとって極めて重要であるかもしれないが、それはデータ主導型のビジネスへの投資や、そこにおける革新に対するインセンティブにも影響を及ぼすかもしれない。同じく、個人データの保護を企図した規則は、国境をまたぐデータの流れを制限し、急成長しつつある、データによって可能になるサービスの貿易にかかわる一国の競争力に実質的な影響を及ぼすかもしれない（第7章）。

競争・独禁政策。競争政策は、プラットフォーム・ベースのビジネス・モデルによって創造される価値が生産者と消費者によって公平に共有されるのを確保することに、極めて重要な役割を果たす。プラットフォームへの参加者が増加するのに伴ってネットワークの価値が高まる、というデータ収集の外部性における規模の経済の存在は、市場支配力の急速な蓄積につながるかもしれない。そのような市場支配に対処するには次の2つの補完的な戦略が必要になる。

第1に、独禁規制の執行に関して十分な能力を有する諸国では、独禁の事後的な執行が実施されるべきである——ただし、データ主導型ビジネスが提起する挑戦に取り組むには何らかの適応が必要であろう。例えば、市場支配に関する標準的なテスト——市場リーダーによる過剰な値上げ——は、プラットフォーム企業が消費者サービスを日常的に無償で提供している分野では無意味であろう。しかし、エジプトや、インド、メキシコなどの中所得国では、プラットフォーム企業を含む画期的な独禁事例がいくつか発生してはいるものの、本報告書は、同じような事例が決着をみている

図 0.5 1990 年以降、データ主導型サービスにおけるグローバル貿易は指数関数的に増加してきており、今やサービス貿易総額の半分を占めている



出所：世界銀行の WITS (World Integrated Trade Solution) データベース，<http://wits.worldbank.org/WITS/>。データは http://bit.do/WDR2021-Fig-O_5。

注：IP = Internet Protocol; PB = ペタバイト。

低所得国は 1 つもないことを見出している。中所得国と同じグローバルに支配的な企業がこのような市場に進出済みであるにもかかわらず、決着に至っていない。

第 2 に、独禁に向けた取り組みと並行して、不可欠なデータに競合企業や新規参加者がアクセスできるようにする事前的な規制措置も真剣に考慮する価値がある。消費者の個人データの完全な携帯性を義務化することによって、競合するプロバイダー間での変更(プロバイダーの乗り換え)を可能にするよう消費者をエンパワメントすることも、同様に考慮する価値がある。その際には次の点を確認するよう常に注意すべきである。すなわち、データへのアクセスは競争にとっては決定的に重要であり、そしてそれが、データ主導型ビジネスにおける革新に向けたインセンティブに不当に影響を与えることがないように確認すべきである。

貿易政策. プラットフォーム・ベースのビジネスは国際貿易に対して新たなルートを開拓し、データの大量の越境フローをもたらしてもいる(図 0.5)。個人データ保護にかかわる一国の規制枠組みは、そのような貿易に実質的なインパクトを与え、信頼と価値創造の間に、ある程度の緊張関係を生み出す。各国はこれに対処することにおいて多様なアプローチを採用している。一部の国、特にアメリカの連邦管轄区は、民間部門の基準に基づいてオープンなデータ・フローを許

容しており、政府の関与は限定的である。中国や、ナイジェリア、ロシア、ベトナムなどの他の諸国においては、より厳しい規制要件が適用されている。そして特定の個人データについてはそのコピーが国内に保存されることを要請しており(データ・ローカライゼーション)、多くの国際的接続点に対しては国の許可が必要とされている。両者の中間に位置する諸国(EU 加盟国や、アルゼンチン、南アフリカなど、その他諸国が含まれる)は、個人データの国境を越える転送に関しては貿易相手国が十分なデータ保護制度を整備していることを条件としている。本報告書は、次のことを見出している。すなわち、国内の明確に規定された個人データの保護措置と、比較的容易な国境を越える移動の組み合わせが、データで可能になるサービスの国際貿易にとって最も好ましい環境を提供しているようである²³。

租税政策. データ主導型の取引はより多くの経済活動を生み出しているものの、低・中所得国の政府は、関連する税収を動員することによってその価値を公平に共有することに苦勞している。間接税(付加価値税など)については、税収権は最終的な販売がなされた国に明確に割り当てられている。しかし、この税収を獲得する行政能力は典型的には欠如している。東アジア諸国についての推定値は、財政収入の逸失は 2030 年

までには GDP の 1% に達することを示唆している²⁴。直接税(法人税など)に関しては、物理的な進出をしていない市場で営業している企業に対して課税する権利の配分については、合意されている国際ルールが存在しない。そのようなコンセンサスが欠如している状況下では、補償措置としてデジタル・サービス税という特別税の適用に頼る国が増加してきている。

健全な担当機関と統治はデータに関する開発のインパクトを改善する

仮に担当機関がうまく機能しないならば、政策や法・規制が有効に実施あるいは執行される可能性は低く、インフラは潜在力を発揮しないだろう。データの統治のための有効な制度的枠組みは、政策目的の設定、支援的なルールと基準の開発、そのような規制の順守の強要、そして学習と評価を通じた継続的な統治の改善、といったいくつかの重要な機能を果たさなければならない(第 8 章)。例えば、大統領府に近く、政府全体の視野をもって行動する主導機関をウルグアイが創設したことは、同国における 2007 年以降の電子政府改革を成功に至らせることにおいて決定的に重要であった。

データ統治について強固な担当機関を創設しようとしている政府にとって万能のアプローチというものはないが、機関の設計に関する特定の特性は普遍的な重要性をもっている。機関が政府全体にわたって協調的な仕方で機能を有効に果たすためには、機関は、権能を正式に負託され、十分な資源を付与され、有効に機能するのに必要とされる技術的な能力を有しているべきである。このような枠組みのなかで機関の主要な行為者には、データ信託のような新種の機関に加えて、しばしばデータ統治機関、データ保護当局、サイバーセキュリティ機関が含まれる。データ信託は、説明責任指向型のデータ媒介者であり、個人がデータに関する法的権利をプールし、それを明示的な信託義務を有する受託者に委託することを可能にする。データ統治機関を不当な政治的あるいは商業的な影響力から遮蔽するためには、場合によっては、制度的な独立性と機能的な自律性が極めて重要なことであろう。行動と文化の規範や政治経済的な制約は、特に低・中所得国では、改革に向けた努力を妨害することが多く、実施においてギャップをもたらす。管理の仕方の変更や、協力的な指導層、パフォーマンスとインセンティブの文化は、当該機関が実施と協調に対する障壁を克服し、みずからの役割と責任を有効に果たすのに役立つ。

データ統治というエコシステムへの全参加者——より広く社会を含む——の賛同を最大化するためには、データ管理は社会的に包摂的でなければならない、そし

て正当であるとみなされなければならない。正当性は、政府が透明な仕方でデータを管理および利用し、有意義な説明責任の制度に従っている場合に高まる。非政府の行為者やデータ仲介者などといった台頭しつつある仕組みは、政府や最終利用者が誤用や悪用のリスクを防止しながら、データの開発的な価値をより良く利用するためにデータを責任をもって共有および使用するのを支援することによって、エコシステムのなかで重要な役割を果たすことができる。社会全体で、および国際的に利害関係者を協調的かつ透明な形で関与させることは、信頼性と正当性を育成し、データ利用を巡る社会契約を強化するだろう。例えば、東南アジア諸国連合(ASEAN)は地域的な「デジタル・データ統治に関する枠組み」を採択している。これは加盟国のデータ統治取り決めを相互運用性に向けて協調するのに役立つものである。

統合国家データシステムへ向けた動き

うまく機能しているデータ統治の枠組みが保証するのは、個人に属するデータの利用に関する当人の権利を保護する一方で、インフラ、法律、経済政策、および機関がそれぞれの価値と統合的な仕方でデータを利用するのを下支えするために協働することである。この枠組みは、すべての利害関係者によってデータが安全に共有、利用、および再利用される方法についてのルールと関連する順守の仕組みを制定している。

本報告書のパートⅢは総合された国家データシステム(INDS)という切望されるビジョンで締めくくられている。INDSは、高質のデータを作成した上で、そのようなデータを、保護されると同時に、すべての利害関係者によって共有および再利用されるためにアクセスしやすい仕方でオープンにするという約束を実現するものである(第 9 章)。切望される INDS は統治構造と共にシームレスに機能する。仮に統治の枠組みを「交通ルール」の制定と執行として見ることができるならば、INDS はすべての利用者を繋ぎ、データの目的地との安全な往復を保証する「高速道路ネットワーク」とみることができよう。

INDS はデータ統治に向けた、国際的な、政府全体での、多様な利害関係者が関与するアプローチの上に構築されている。それはデータの作成、保護、交換、および政府諸機関全体での企画や意思決定での利用を規定し、市民社会、公共部門、および民間部門からのさまざまな利害関係者をデータのライフサイクルとシステムの統治構造に統合する²⁵。

うまく機能するシステムに必要なとされるのは、高質のデータを生成、加工、および管理する人たち、不正

利用に対してデータを保護する機関に従事する人たち、そしてデータに関する戦略、政策、および規則を起案、監督、そして実施する人たちである。システムでは官民両部門の説明責任を問う人たちや、公共部門における政策改善のために私企業の生産過程から得られるデータを利用する能力を持った人たちも必要とされる。これらすべてのことは、強固なデータ・リテラシーを要請しており、リテラシーを身につけることで、あらゆる階級に属する人たちが INDS から利益を享受する。

健全な INDS のためには、担当機関や当事者はデータを作成、保護、および共有することに対する適切なインセンティブをもち、システムがうまく機能するために必要なインフラや担当機関を実現するのに十分な資金手当てがなされる必要もある。最後に、データ利用の文化はデータの高質な供給を育み、情報に基づく意思決定に対する需要を刺激するのに資する。そのような意志決定は、国のデータシステムの維持に不可欠である。

政府機関、市民社会、学界、および民間部門が国のデータシステムに安全に参加する時には、データの潜在的な利用が拡大し、したがって開発に対する潜在的なインパクトも拡大する。実際に、システムの統合化が進み、参加者が増加するにつれて、潜在的な見返りも大きくなる。仮に 2 人の参加者が安全にデータを交換するなら、データは双方向に流れることができる。仮に 3 人の参加者がデータを交換できるならデータは 6 方向に、4 人なら 12 方向に流れることができる。データは再利用や新たな目的のために再利用できるので、このような繋がりは急速に拡大するだろう。繋がりが高度になると、参加者間での密接な調整や統治の共有が必要となるが、そのような統合は一方で、分権的なデータ構造と両立している。システムは、データが自由かつ安全に流れるのを保証するように設計されている。すなわち、データは、一カ所にとどまてはいない。

ほとんどの国はうまく機能するデータシステムという野心的な目標からは遠く離れてはいるものの、このような目標を設定することは、そのようなシステムの開発における、次のステップに関する指針を国に提供する。INDS のこのようなビジョンに向けて各国がどのように動いていくかは、各国の現在の能力とデータのための社会契約の[媒介変数的な]要素に依存するだろう。INDS をどのように構築するかについての単一の青写真は存在しない。その代わりに、本報告書では進捗を評価するのに役立つ成熟度モデルを提案している。初期段階にある諸国は、統合国家データシステムに関する基本的条件を確立することから最大の利益を

享受できる公算があろう。これに含まれるのは、データ統治の改善を目指す政策や戦略の策定、政府機関や国家統計局によるデータの作成および利用にかかわる技術的能力の強化、教育と訓練を通じたデータ・リテラシーの促進である。基本的条件が整っていれば、政府は国家データシステムの参加者の全体および相互間におけるデータの流れの始動と体系化という作業に着手できる。このことは、データや関連機関の一貫性と相互運用性、およびプライバシー・リスクを緩和するようなデータの安全な交換を可能にするインフラを保証する政策と基準を必要としている。データシステム構築の進捗度が高度なレベルにあるなら、目標は、政府や、国際機関、市民社会、民間部門からなるさまざまな利害関係者の間で共有されているデータ統治や協調を通じて、システムを最適化することである。

このようなシステムの構造は、データの安全な再利用や共有にかかわる各国の規範を反映して国毎にそれぞれ異なるだろう。にもかかわらず、データの再利用や共有から得られる開発利益を実現するためには一定の共通する属性が必要である。うまく機能しているデータシステムは、国内におけるデータの作成、流れ、および使用に関する権限や責任が定義され、そして確立されている。このシステムは本書で議論されているインフラや、政策、法規制、担当機関を土台にして構築されるだろう(図 O.6)。

多くの諸国にとって、さまざまな参加者の間を高質のデータが流れ、そして安全に利用されているシステムというのはかけ離れたビジョンにとどまっている。高水準の貧困、脆弱さ、貧弱な統治に苦しんでいる低所得国は、データ統治に向けた政府全体による、多様な利害関係者を包摂するアプローチを制定するどころか、最も基本的なデータを作成することにおいてさえ苦闘しているかもしれない。にもかかわらず、このビジョンから目を離さないことがすべての国にとって、データで最も苦戦している諸国にとってさえ、重要である。というのは、自国のデータシステム開発に関する決定を行う際の指針としてこのビジョンは役立つからだ(ボックス O.1)。

データにかかわる新たな社会契約——価値は、公平に分配されるデータから生成されるという信頼の上に構築される——の共通理解を巡って連合し、統合国家データシステム——ユーザーの顔ぶれを広げることによって、創造的で革新的な適用を通じて生活を改善するためにデータの価値を解放する——の構築に向けて適切な青写真を発見することは、極めて野心的な目標である。このような目標を達成することは、データが作成、管理、保護、共有、そして利用される方法につ

図 0.6 総合国家データシステムではどのようなことが生じるか？



出所：WDR 2021 チーム。

ボックス 0.1 統合国家データシステムに向けて：数カ国の事例

正しい方向への重要なステップ。多くの諸国が統合国家データシステムにおいて描かれている側面を具体化している重要なイニシアティブを採用している。南アフリカの計画監視評価局は、政府プログラムの実績を監視している市民が作成したデータを取り入れたシステムを開発している。チリでは、市民社会の参加は「公共管理における連合と市民参加に関する 2011 年法」によって義務化されており、国家統計局は市民社会評議会を設置している。国家データシステムに多様な利害関係者を含めることは持続可能性を促進し、全参加者が国家データシステムを利用し、そして利益を得る機会を持つのを

確実にすることに役立つ。ネパールで 2019 年に創設された「ネパール・データ・リテラシー・プログラム」は、国民の間でデータ利用の文化を高めることを目的として技術スキル構築とそれへの取り組みの両方を支援するために、100 時間のモジュラー形式でカスタマイズ可能な教授法で構成されている。このプログラムは今ではカトマンズ経営大学(KUSOM)と提携して、データ・リテラシーのツールキットを大学の課程に組み込み、データ主導型のコースを開発している。このコースは、他の機関や数千人の学生に対して無料で提供されることになっている。

ボックス 0.1 統合国家データシステムに向けて：数カ国の事例(続き)

完全に実現されたビジョン。エストニアでは、政府機関や参加企業による利用から市民の個人データを安全に管理するために、政府が国家データ・システムを立ち上げた。X-Road は、オープン・ソースのデータ・エクステンジ・レイヤー・ソリューションである。このソリューションは、リンクされた公共および民間のデータベースが自動的に情報を共有するのを可能にし、参加者間のデータ交換における秘匿性、誠実製、および相互運用性を保証している。それは技術面でのソリューション(技術的構造と一連のプロトコルを許容)と統治面でのソリューション(データに関する要請を重ねて行うのを止めることを公的機関に義務化した国法の規定——「一回限り」原則)を組み合わせている。このシステムの下では、市民は政府機関や参加企業に対して自分に関する情報を一度だけ提供すればよい。そうすれば、それは他の参加者に自動的に転送される。X-Roadの暗号プロトコルも透明性を高めている。というのは、

システムへの記入は記録され、当該個人は自分のデータを誰が何のために共有しているかについて詳しい洞察を得られるからだ。

X-Roadの取り決めは、信頼、公平性、および価値を提供することによって、データに関するエストニアの社会契約の上に構築されており、同時に、それを強化してもいる。その透明性は信頼を生んでいる。国全体をカバーし、万人が利用可能であることで、公平性を促進している。容易さと包括性は価値を提供している。うまく機能するために、このデジタルのデータシステムはある程度の「アナログ」的な要素に依存している。政府と民間部門の間で、およびインフラの構成部分の間で、協調が促進されている。変化の管理がシステム全体——国の法律(および社会契約)における基盤からその設計・理解・維持に至るまで——に組み込まれている。信頼と共有(データ共有)の文化が奨励されている。

いて著しい変更を要請するだろう。そのような変更は容易ではなく、そして内容のある決意と努力に依存するであろうが、変更失敗した場合の代価は、より大

きな不公平と多くのチャンスの損失に直面する世界である。

注

1. 第1章ではデータの定義がレビューされており、データという用語の本報告書における使われ方が説明されている。
2. Amnesty International (2019); Zuboff (2019).
3. Scheer and Cohen (2020).
4. Scheer and Cohen (2020).
5. World Bank (2017).
6. FMWR (2018).
7. World Bank, Statistical Performance Indicators (database), <http://documents.worldbank.org/curated/en/815721616086786412/Measuring-the-Statistical-Performance-of-Countries-An-Overview-of-Updates-to-the-World-Bank-Statistical-Capacity-Index>.
8. Brynjolfsson, Hitt, and Kim (2011).
9. この問題と機械学習に関する他の多くの懸念に関する広範な議論については O'Neil (2017) を参照。
10. 「ゴミを入れたらゴミが出てくる」という格言に言及していた初期の統計学文献に関しては Parzen (1964) を参照。
11. Brynjolfsson, Hitt, and Kim (2011).
12. Burke and Lobell (2017); Osgood-Zimmerman et al. (2018).
13. 公共支出の説明責任を改善するために、データをどのように使うことができるかに関する初期の例証は、Reinikka and Svensson (2001) によるウガンダにおける教育予算の使用に関する研究のなかで示されている。
14. 次を参照: Broadband Commission for Sustainable Development, International Telecommunication Union, “Target 3: Connectivity” (accessed October 31, 2020), <https://broadbandcommission.org/Pages/targets/Target-3.aspx>.
15. Chen (近刊). アフリカ, アジア, およびラテンアメリカの22の低・中所得国において, Research ICT Africa によって集められた Access Survey 2017-18 のデータの分析に基づく。
16. GSMA (2019).
17. 「移動体通信ネットワーク向けの電話——「携帯電話」——や他の無線ネットワーク」(HS コード 851712)にかかわる世界貿易機関の従価関税について, 世界銀行が分析したもの。
18. ITU et al. (2018).
19. COE (2018).
20. 個人データの保護措置は国際法に基づく人権の枠組みに根差している。このような保護措置は「法の支配」の確立に起源があり, 啓蒙思想のなかで個人の権利が表明され, 第2次世界大戦後の国際法において成文化された。それは1970年代と80年代にアナログ・データという文脈でいっそう洗練化された。これは, 「公正な情報取り扱い」, ヨーロッパ評議会による1981年の「個人データの自動処理に関する個人の保護に関する条約」(108号条約), およびOECDによって発表された最新の指針による。OECD指針と108号条約は, 1995年にwwwが開始され, 進化を続けているなかで, デジタル・データという文脈で更新された。
21. Madrigal (2012).
22. OECD (2019).
23. Ferracane and van der Marel (近刊).

24. Al-Rikabi and Loeprick (近刊).
25. 統合国家データシステムは、すべてのデータが国家のデータベースに統合される、ということを示唆するものではない。そうではなく、多様な参加者が、データが安全に流れ、そして利用されるシステムに統合されるということである。これは参加者の集まりが共同で公的統計を収集、保護、処理、および配布するという意味で、国家の統計システムに類似している。しかし、国の統計システムとは違って、統合国家データシステムの範囲は公的統計をはるかに上回っており、そして参加者とその役割の統治に関して意図的なアプローチを必要とする。

参考文献

- Aiken, Emily L., Guadalupe Bedoya, Aidan Coville, and Joshua E. Blumenstock. 2020. "Targeting Development Aid with Machine Learning and Mobile Phone Data: Evidence from an Anti-Poverty Intervention in Afghanistan." In COMPASS '20: Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies, 310–11. New York: Association for Computing Machinery.
- Al-Rikabi, Jaffar, and Jan Loeprick. Forthcoming. "Simulating Potential Tax Revenues from Data-Driven Platform Businesses in East Asia." WDR 2021 background paper, World Bank, Washington, DC.
- Amnesty International. 2019. "Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights." Report POL 30/1404/2019, Amnesty International, London. <https://www.amnesty.org/en/documents/document/?indexNumber=pol30%2f1404%2f2019&language=en>.
- Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim. 2011. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" SSRN Scholarly Paper ID 1819486, Social Science Research Network, Rochester, NY. doi.org/10.2139/ssrn.1819486.
- Burke, Marshall, and David Lobell. 2017. "Satellite-Based Assessment of Yield Variation and Its Determinants in Smallholder African Systems." *PNAS, Proceedings of the National Academy of Sciences* 114 (9): 2189–94. doi.org/10.1073/pnas.1616919114.
- Chen, Rong. 2021. "A Demand-Side View of Mobile Internet Adoption in the Global South." Policy Research Working Paper 9590, World Bank, Washington, DC. <http://documents.worldbank.org/curated/en/492871616350929155/A-Demand-Side-View-of-Mobile-Internet-Adoption-in-the-Global-South>.
- COE (Council of Europe). 2018. "Convention 108+: Convention for the Protection of Individuals with Regard to the Processing of Personal Data." COE, Strasbourg. <https://rm.coe.int/convention-108-convention-for-the-protection-of-individuals-with-regard-to-the-processing-of-personal-data/16808b36f1>.
- Ferracane, Martina Francesca, and Erik Leendert van der Marel. 2021. "Regulating Personal Data: Data Models and Digital Services Trade." Policy Research Working Paper 9596, World Bank, Washington, DC. <http://documents.worldbank.org/curated/en/890741616533448170/Regulating-Personal-Data-Data-Models-and-Digital-Services-Trade>.
- FMWR (Federal Ministry of Water Resources, Nigeria). 2018. "National Action Plan for Revitalization of the WASH Sector." FMWR, Abuja.
- GSMA (GSM Association). 2019. "Connected Society: State of Mobile Internet Connectivity 2019." GSMA, London. <https://www.gsma.com/mobilefordevelopment/resources/the-state-of-mobile-internet-connectivity-report-2019/>.
- ITU (International Telecommunication Union), World Bank, ComSec (Commonwealth Secretariat), CTO (Commonwealth Telecommunications Organisation), and NATO (Cooperative Cyber Defence Centre of Excellence, North Atlantic Treaty Organization). 2018. "Guide to Developing a National Cybersecurity Strategy: Strategic Engagement in Cybersecurity." ITU, Geneva. https://www.itu.int/pub/D-STR-CYB_GUIDE.01-2018.
- Knippenberg, Erwin, and Moritz Meyer. 2020. "The Hidden Potential of Mobile Phone Data: Insights on COVID-19 in The Gambia." *Data Blog* (blog). September 10, 2020. <https://blogs.worldbank.org/opendata/hidden-potential-mobile-phone-data-insights-covid-19-gambia>.
- Madrigal, Alexis C. 2012. "Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days." *Atlantic*, March 1, 2012. <https://www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/>.
- OECD (Organisation for Economic Co-operation and Development). 2019. *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies*. Paris: OECD. https://www.oecd-ilibrary.org/science-and-technology/enhancing-access-to-and-sharing-of-data_276aaca8-en.
- O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books.
- Osgood-Zimmerman, Aaron, Anoushka I. Millar, Rebecca W. Stubbs, Chloe Shields, Brandon V. Pickering, Lucas Earl, Nicholas Graetz, et al. 2018. "Mapping Child Growth Failure in Africa between 2000 and 2015." *Nature* 555 (7694): 41–47. doi.org/10.1038/nature25760.
- Parzen, Emanuel. 1964. "Review of Smoothing, Forecasting and Prediction of Discrete Time Series, by Robert Goodell Brown." *Journal of the American Statistical Association* 59 (307): 973–74. doi.org/10.2307/2283122.
- Reinikka, Ritva, and Jakob Svensson. 2001. "Explaining Leakage of Public Funds." Policy Research Working Paper 2709, World Bank, Washington, DC.
- Scheer, Steven, and Tova Cohen. 2020. "Israel Extends Coronavirus Cell Phone Surveillance by Three Weeks." *Emerging Markets* (blog), May 5, 2020. <https://www.reuters.com/article/us-health-coronavirus-israel-surveillance/israel-extends-coronavirus-cell-phone-surveillance-by-three-weeks-idUSKBN22H11>.
- World Bank. 2017. *A Wake Up Call: Nigeria Water Supply, Sanitation, and Hygiene Poverty Diagnostic*. WASH Poverty Diagnostic Series. Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/27703>.
- World Bank. 2019. "Tanzania, Mainland Poverty Assessment 2019: Executive Summary." World Bank, Washington, DC. <https://www.worldbank.org/en/country/tanzania/publication/tanzanias-path-to-poverty-reduction-and-pro-poor-growth>.
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs. <https://www.hbs.edu/faculty/Pages/item.aspx?num=56791>.