

# オープンデータのCOMMONユース —増大するデータの共有と活用の可能性と課題—

金 村 公 一

A study of Open Data as Common

Kouichi KANAMURA

## 概 要

デジタル情報とデジタルメディアの進展は、膨大なデータの蓄積と活用をもたらした。プライベートセクターでは、検索エンジンにおけるインデックスの価値付けが行われ、個人の様々な履歴データの収集・蓄積・活用に対するプライバシーとの軋轢が生じている。パブリックセクターでは、2009年に誕生した米国オバマ政権が、米国史上初の連邦最高情報責任者(CIO)を設置し、「クラウドファースト」と称する情報政策スローガンが掲げて全省庁を対象に政府情報のクラウド化を推し進めている。政府情報がクラウド化され、データの蓄積・運用・管理がアウトソーシングされることによるプライバシー、セキュリティの問題に関する危惧も起こっている。これら2つのセクターだけでなく、市民社会がデータを可視化し、共有することについて、多様な立場の言説を検証し、既存のマスメディアのジャーナリズムと比較して可能性、課題を検討した。

**キーワード：**デジタル情報、オープンデータ、ジャーナリズム、ネットワーク、マスメディア、地域

### 1. はじめに

デジタル情報、デジタルメディアを介したコミュニケーションに代表される人々の情報行動(閲覧、検索、撮影、文書作成、データ収集など)は、ネットワーク上のいづこかのストレージに蓄積される。蓄積の主体は数多あり、その素性、目的、加工方法、利用方法を全て確認することは、至難である。しかしデジタルメディアを介した便利な情報・コミュニケーション環境を活用する多くの人々にとって、圧倒的な利便性を日常のものとし

て手にしたなかで、かかる蓄積者の存在に関する意識は、希薄化しているといえよう。今日、あらゆるデジタルメディアを介して収集・蓄積される情報量は急激に増殖している。本論では、増殖する一方のデジタルデータに対し、その特性を活用してこれまでつかみ切れなかった社会の課題を新しい形で共有し、解決するアプローチに活用できないかという視点で多様な言説を検証し、通底する問題意識を検討する。

## 2. ネットワークトラフィックとデバイスの増殖

CISCOの調査<sup>1</sup>によると、図1のとおり、2011年の世界のIPトラフィック量は30.7エクサバイト<sup>2</sup>/月に達し、2010年の21.8エクサバイト/月から41%増加している。2016年には110.3エクサバイト/月に達し、2011年の実績の3.6倍に増加すると予測している。日本に限ると、2011年2.0エクサバイト/月で2010年1.4エクサバイト/月から43%増加しており、ほぼ世界の増加率と同じ値を示している。2016年には6.4エクサバイト/月で2011年の実績の3.2倍に達すると予測している。ひとびとが送受する情報は今後さらに増加傾向が続くといえる。

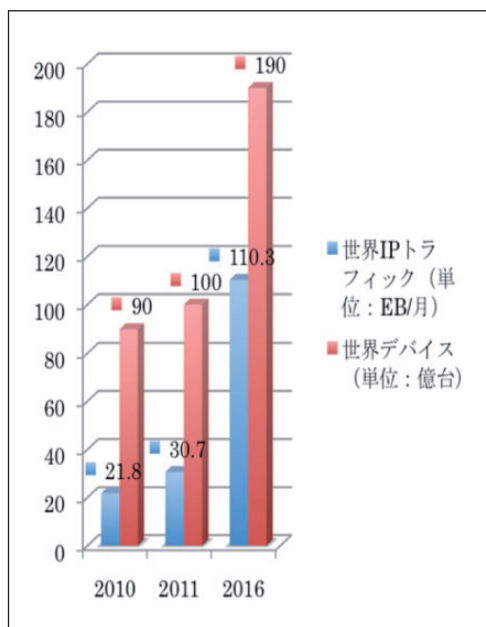


図1 世界のIPトラフィックとネットワークに接続されたデバイスの台数の推移  
(出展: CISCO VNI Forecast のデータより作成)

一方、ネットワークに接続された機器の調査結果を見ると、2011年の世界のネットワークに接続されたデバイスの数は100億台で、2010年の90億台から11%増加している。2016年には190億台に達し、2011年の実績の1.9倍に達すると予測している。ひとりあたりに換算して、2011年の1.5台から

2016年には2.5台に達すると予測している。

日本に限ると、図2のとおり、2011年は3億6200万台で、2010年3億100万台から20%増加している。2016年には8億4400万台と2011年の実績の2.3倍に達すると予測している。ひとりあたりに換算すると、2011年の2.8台から6.5台と世界の予測に比べても大きく台数が増えると予測している。

世界的に、様々なデバイスを活用して多様な情報の送受が行われ、その量は増加し続ける。日本でも、同様の傾向にあるが、とりわけデバイスの数が世界に比べて1.9~2.6倍も多く推移する。従って、より多様な手段で多様なサービスを活用することがこれらの調査データから読み取ることができる。

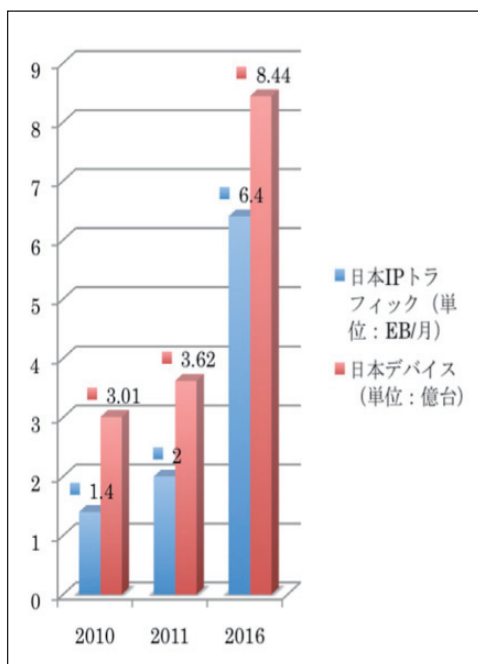


図2 日本のIPトラフィックとネットワークに接続されたデバイスの台数の推移  
(出展: CISCO VNI Forecast のデータより作成)

## 3. 情報・コミュニケーション行動と履歴データ

2012年9月のPew Research Centerの調査報告<sup>3</sup>によると、米国で18歳以上の85%が

インターネットを利用している。若年であるほど、学歴が高いほど、年収が高いほどその利用率は高くなるが、全てを平均した数値が85%である。

ネットワークを使ってひとびとは何を行っているのか。

同報告が調査し、詳細に分けた56の項目を報告している。そのうち、利用者の70%以上の方がインターネットで行っていると回答したのは11項目あった。

表1に示すとおり、上位から順に、91%がサーチエンジンで情報を検索、メールの送受、84%が趣味や興味のある情報の閲覧、地図やドライブルート検索、81%が天気予報のチェック、80%が健康情報を探し、78%が買いたいと思っているサービスや商品を探す。76%がニュースを見る、74%が娯楽のためにネットワークを見る。71%がモノを買うためや、YouTubeやVimeoなど動画共有サイトでビデオを見るためにインターネットにアクセスすると答えている。

これらの情報・コミュニケーション行動の履歴は膨大なデータとして蓄積される。

表1 18歳以上の米国人がインターネットを利用する目的

インターネットの利用目的	%
サーチエンジンで情報を検索	91
メールの送受	91
趣味や興味ある情報の閲覧	84
地図やドライブルートの検索	84
天気予報のチェック	81
健康情報を探し	80
買いたいと思っているサービスや商品を探す	78
ニュースを見る	76
娯楽のためにネットワークを見る	74
商品を買う	71
YouTubeやVimeoなど動画共有サイトでビデオを見る	71

(出典：Pew Research Center Pew Internetより抽出し作成)

「サーチエンジンで情報を検索」することは、日常の情報行動における代表的なものである。

パスキネッリ(2011)は、多様な検索エンジンの中でグーグルの検索精度が他に秀でた理由を「ページランクのアルゴリズムを通じ、集合知を諸価値の所有権に基づく計測器にすることで、価値を生産・蓄積する<sup>4)</sup>」ことによると述べる。インデックス情報を価値付けにリンクという人為的な価値判断の結果を反映することは、ひとびとの情報・コミュニケーション行動の履歴を蓄積し、その膨大な検索の履歴データを蓄積、分析することが必要になる。そのために巨大なデータセンターが必要となり、瞬間にデータセンターには限りなく多くの情報・コミュニケーション行動のデータが蓄積されることになり、クラウドがビジネスとして形成されていった。レビ(2011)は『グーグル』において、その課程を詳細に取材し述べている<sup>5)</sup>。

今日のトラフィックの増大は、単なるデータの増大ではなく、多様な行動履歴を集合知として分析し、価値付けが行われることと共に増大していることを示唆している。しかし、プライベートセクターにおいてはデータとデータの解析や活用はブラックボックス化している。利用者は、便利や無料という便益とプライバシーの狭間に置かれている。

#### 4. クラウドファーストとオープンデータ

2009年1月に発足したオバマ政権は、情報政策遂行にあたり、はじめてCIO(Chief Information Officer)の役職を設置し、初代CIOにVivek KUNDRAを指名した。連邦政府の省庁に対し少なくとも3つのクラウド型サービスを2010年夏までに開始する目標を設定し強力に推進した<sup>6)</sup>。

クラウドファーストの政策の目的、安全性等について検証については本論では触れない。米政府初代CIOのKUNDRAが政権を離れた後、ハーバード大学ジョアン・ショールンスタインセンターのフェローとして発

表した” Innovation through Open Data and the Network Effect”<sup>7)</sup> という論文で主張している点に注目した。

KUNDRA (2012) は、「ヒトゲノムプロジェクト」を例に挙げ、「今日、情報においてデータはきわめて重要である」と述べている。「歴史上のどの時代に比べても、今日、我々が作り出し、蓄積しているデータは多い」とし、Reed の法則<sup>8)</sup>を用いながら、この膨大なデータを有効活用し、社会、政治、経済を変えるためには公開し、ネットワークによるそれもとりわけソーシャルネットワークが活用に参加することが重要であると主張している。

KUNDRA が挙げる 4 つの課題は以下のとおりである。

- ①政府のサービスの拡充
- ②公開性、透明性、参加型へと政府の基本的なあり方を変える
- ③社会の出す信号を読み取り意味のある洞察を生み出す新しいジャーナリズムのモデル
- ④パブリックセクターの持つデータを活用したビッグビジネスの萌芽

①と④は既に米国政府、企業がそれぞれ着手しているが、②と③については大きな伝統的な体質の転換を含めて、新たなプレーヤーとして市民、NGO の参加と協働が必要である。この②と③は、我が国の震災時、原発事故に際しての正確な情報の開示とオープンな議論、そして参加型問題解決にいつこうにつながらない状況に隔靴搔痒の苦痛を味わっていることを考えると、KUNDRA の問題提起を、普遍的な課題として受け入れることができよう。

## 5. データの可視化による可能性と課題

膨大に増殖するデータ。そこにはヒトの発言、行動、写真、映像だけでなく、自然の発するシグナルやその計測結果などありとあら

ゆる種類のデータが存在する。これらのデータは非公開、公開という人為的な操作に基づくものもあれば、今日のソーシャルネットワークの発達により、まさにクラウド上に蓄積され、誰でもいつでも見ることができるといふ自然発生し自動的に蓄積される集合知と称することができるものも多い。

東(2011)は、『一般意志 2.0』において、集合知を「曖昧な経験則ではなく、堅実な実証に支えられた実用的な認識に変わりつつある<sup>9)</sup>」と評価している。「一般意志は政府の意思ではない。個人の意思の総和でもない。単なる理念でもない。一般意志は数学的存在である」と述べ、ルソーの社会契約論の再認識に立ち戻りながら、「総和」よりも「差異の和」に着目している。アーレントは、公共空間を差異の出会いの場としてとらえてユニークに独立したひとりずつの人の「間」にある事柄への関心と言及した。その「間」にある事柄をどのようにして見るのかを考えたとき、今日のデータの増殖とその蓄積、評価、可視化といった様々な試みとの関係を期待せずには居られない。これまで行われてきた代議制や、世論調査、公開の意見交換などの場では限界が見えていた部分にオープンデータの可能性を見いだしたい。

レイニーとウエルマン<sup>10)</sup>(2012)が、ソーシャルネットワーク、インターネットとモバイルの3つの革命が相乗して” Networked Individualism”と呼ぶことができる社会の運用システムに言及している点にも注目すると、KUNDRA が挙げた4つの②と③の項目は、①と④の項目よりもずっと重要で至難の課題といえるほど、長く希求されてきた社会の課題とも連動する。

このように考える時、日々吸い上げられるひとびとの履歴によって生成されるデータをプライバシーかパブリックのどちらの視点で捉えるかについてより深い検討が必要とされる。金村(2010)は、「ツイッターやブログの例から抽出できることは、ネットワークコミュニケーションがもつ参加と協働、即時性

と伝播力、ゆるやかな連携の中で茫洋とした事実の集合により明らかな問題が浮上してくるという可能性である<sup>11)</sup>と述べ、クラウド上での個人をベースとした情報・コミュニケーションに何らかの可能性を意識している。また、金村(2011)は、「マスメディアとソーシャルネットワークの連動による創発的なコミュニケーション空間に「私たち」の信頼が宿るヒントを見いだせないだろうか<sup>12)</sup>」との問いを発している。

マスメディアという言葉ジャーナリズムに置き換えると、KUNDRAの③との関連性をみることもできる。

情報工学的な視点では、坂村らが提唱する空間情報科学におけるコモンピクチャー<sup>13)</sup>の活用がある。ネットワークが前提の時代に一つのセクターで全てが完結できず、それぞれが他の機能を利用するためのインフラが必要となり、米国のオープンクラウド化を例に、スタティックにPrivacyを守るという視点だけで新たな可能性は求められないとしている。

## 6. 社会のシグナルを読み取るジャーナリズム

マスメディア、特に報道における重要な役割として「調査報道」が挙げられる。権力の隠蔽を暴き、あるいは長い期間の証拠を積み重ねて事実を解明し、その分析と検証には多大の労力と権力との対峙も要求される。米国では、これまで調査報道を担ってきた報道機関は、デジタルメディアとネットワークの普及に伴い減少の一途をたどる広告収入により、調査報道を維持する力を失いつつあるともいわれる。

この調査報道の構造を図3に示し、オープンデータの共有と可視化という視点で読み直す。目撃証人ニュースの段階は、データの収集や生成となり、これをさらに注意深く収集する段階が追跡調査ニュースにあたる。その際にネットワークを介した参加と協働のもとに、データを解析する手法が検討される段階を説明責任(検証)ジャーナリズムに当ては

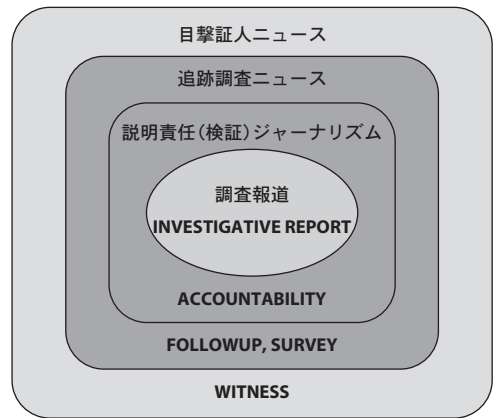


図3 調査報道の階層構造図式  
(出典：『ジャーナリズムの原則<sup>14)</sup>』より作成)

めることができる。このような作業を通して、「差異の和」のようなものが可視化することをコモンピクチャーと呼ぶ。コモンピクチャーを作るために、オープンデータとその活用に必要な情報技術が必要とされる。さらに、その結果を公開し、対話を深めることも必要となる。これを誰が担うのか。既存の報道機関、マスメディアだけで担えるのか。現状の報道機関だけでなく、データの収集、その分析、可視化を行うにあたり専門的な問題に特化した技術や研究者、NPO、市民との協働関係が必要となろう。その前に、伝統的マスメディアのマインドとこの数学的とも発見的ともいえるオープンデータの扱いをどのようにガバナンスするかが課題となろう。そこには、プライバシーとパブリック、マスメディアの関係、線引きに複雑さが増すであろう。

既存の調査報道では、衰退が激しい米国のマスメディアを補完する形で、ニューヨークを拠点とする Propublica<sup>15)</sup>に代表されるNPO型の調査報道専門機関が各地で活動を始めている。こうした機関は、Webなどネットワーク上をアウトプットの拠点とし、案件毎に既存の報道機関と連携している。NPO型の調査報道機関に従事するジャーナリストは、伝統的なマスメディアの出身者が多く、従事者を見ていると、既存メディアからNPOに人材がシフトしたともいえる。

従って、オープンデータとその可視化によるコモンピクチャーを前提とした活動には、より地域に密着した問題発掘と問題解決、そのためにデジタルデータやネットワークを活用し、オープンで参加型の構造が必要となる。このような視点で見ると、Knight 財団<sup>16</sup>が MIT メディアラボと行っている協働研究 Center for Civic Media などデジタル技術を民主主義に応用する実験ともいえ、既に多様なアプローチは始まっている。社会のシグナルを読み取るジャーナリズムは、地域コミュニティをエリアとしてより、日常生活に近い課題を地域に根ざす NPO と地域メディアが協働する形で実現することができるかもしれない。

その為には、データをオープンにする仕組み、データを可視化し分析、判読可能な状態にして共有し、これにもとづいて対話と問題解決が進められる仕組みが必要とされる。パブリックとプライバシーが近接するネットワークにつながった個人が、コミュニティのシグナルを読み取り、コモンピクチャーとして共有し、問題解決に役立てる手法を希求する傾向は徐々に高まりを見せている。

## 7. さいごに

オープンデータの可能性は、ノイズからシグナルを発見し、シグナルの意味を解釈することにより、従来見過ごされてきた問題を発見し、その解決を図る可能性である。データを収集、分析、可視化するにあたっては、パブリックの価値を最大化することとプライバ

シーの保護の均衡点を見極めるガバナンスが必要とされるが、その前にデータは誰のものかという根源的な問いを建てるとすれば、オープンデータの視点からは、図4に示す多元モデルの外枠に多様なデータがオープンに提供される姿が求められる。データの活用に関する 5W1H、プライバシーと公益の勘案、データやシステムの危機管理、参加型で共有、公開、協働のメタルールなどガバナンス要件は多岐にわたる。そのガバナンスをどのように実現するのかについて、幅広く議論すべき時が来ている。

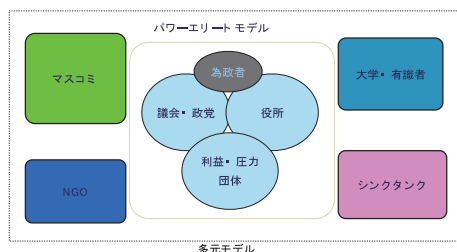


図4 政策に関するプレイヤーの構造

<sup>1</sup> インターネットのルーター技術の先駆的存在である CISCO 社が継続して行っている IP 関連の調査結果を参照した。調査結果は、同社 [http://www.cisco.com/web/solutions/sp/vni/vni\\_forecast\\_highlights/index.html](http://www.cisco.com/web/solutions/sp/vni/vni_forecast_highlights/index.html) のサイトで閲覧可能である。

<sup>2</sup> エクサバイトは、ギガ、テラ、ペタの後に来る単位  
 $(10^3)^6 = 10^{18} = 1,000,000,000,000,000,000$  Byte

<sup>3</sup> Pew Research Center：米国で定評のあるマスメディア、情報コミュニケーションに関する調査機関で利用者の動向をよく調べたものが多い、近年、ネットワーク関連の調査も充実している。本稿で取り上げたデータは、Pew Internet & American Life Project によるものである。  
<http://www.pewinternet.org/Static-Pages/Trend-Data-%28Adults%29/Whos-Online.aspx> のサイトで閲覧可能である。

<sup>4</sup> マッテオ・パスキネリ、「グーグル<ペー  
 ジランク>のアルゴリズム」(永原豊訳)、  
 現代思想 (Vol39-1), p119, 青土社, 2010

<sup>5</sup> スティーブン・レビ、『グーグル』(仲達志、  
 池村千秋訳)、NTT 出版, 2011

- 6 Jeff.GOULD氏が情報通信学会研究会で発表した際に得た情報に基づく
- 7 Vivek KUNDRA, "Innovation through Open Data and the Network Effect", Joan Shorenstein Center on the PRESS・POLITICS, Harvard University, 2012, p4  
尚、2代目の米国連邦政府CIOは、CISCO Robert Bigmanが2012年4月から
- 8 インターネットのような相互接続による開放型のネットワークがもつ可能性を、2のノード数乗(2<sup>n</sup>)とするもので、閉鎖的なネットワークのもつネットワーク数の二乗(メトカーフの法則)や、スター型の中央集権的ネットワークのノード数倍(サーノフの法則)に比べて格段に高いと考えられている。Reedとは、MIT(米国マサチューセッツ工科大学)のコンピュータネットワーク研究者でインターネットのIP接続方式(TCP/IP)の開発にかかわったDavid Patrick Reedの名前である。
- 9 東浩紀、『一般意志 2.0』, 講談社, 2011, p.32, p.47  
古典・現代の多様な言説を用いて展開されるすばらしい論述で示唆に富む箇所が多数ある
- 10 Lee RAINIE & Barry WELLMAN, *NETWORKED*, The MIT Press, 2012
- 11 金村公一「持続できる「調査報道」へ」『総合ジャーナリズム研究』No.212, 社団法人東京社, 2010年, pp.26-27で、米国の報道機関の広告依存度が高いために近年広告収入が減少するなかで調査報道が衰退している事実と、これに対する補完的な動きとしてNPO型調査報道機関の台頭、インターネットを活用した若者や市民が運営する調査報道的活動などを
- 詳しく紹介している。
- 12 金村公一「マスメディアとネットの信頼性」『総合ジャーナリズム研究』No.218, 社団法人東京社, 2011年, pp.4-5  
大震災と原発事故の報道を検証し、当時展開されたソーシャルネットワークの動き、震災後に調査されたメディア信頼度などについて示し、マスメディアとネットワークの信頼性について提言を行っている。
- 13 坂村健(東京大学情報学環), 日本学会議 情報学委員会ユビキタス状況認識社会基盤分科会シンポジウムの発表内容に基づく
- 14 ビルコヴァッチ, トムローゼンスティル, 『ジャーナリズムの原則』(加藤岳史, 斎藤邦泰訳), 日本経済評論社, 2011で調査報道のプロセスに関する記述を図式化した。
- 15 NPO型調査報道専門機関の草分け的存在 ピュリツァー賞を連続して受賞するなど高い評価を得ている。ニューヨークタイムなど既存の報道機関との連携も行っている。
- 16 Knight Foundationは多様な市民活動や地域メディア試行を支援している。MITメディアラボにも資金を提供し、プロジェクトを協働で進めている。  
<http://www.knightfoundation.org>

#### 参考文献

マッテオ・パスキネリ, 「グーグル<ページランク>のアルゴリズム」(永原豊訳), 現代思想 (Vol39-1), p119, 青土社, 2010

スティーブン・レビ, 『グーグル』(仲達志,

池村千秋訳), NTT 出版, 2011

Vivek KUNDRA, “Innovation through Open Data and the Network Effect”, Joan Shorenstein Center on the PRESS・POLITICS, Harvard University, 2012

東浩紀, 『一般意志 2.0』, 講談社, 2011

金村公一「持続できる「調査報道」へ」『総合ジャーナリズム研究』No.212, 社団法人東京社, 2010年

金村公一「マスメディアとネットの信頼性」『総合ジャーナリズム研究』No.218, 社団法人東京社, 2011年

ビルコヴァッチ, トムローゼンスティル, 『ジャーナリズムの原則』(加藤岳史, 斎藤邦泰訳), 日本経済評論社, 2011