

《論文》

インドにおけるオープンデータへの取り組み —data.gov.in の開設と発展を中心に—

Grappling with Open Data in India —Establishment and Development of data.gov.in—

坂田 大輔 Daisuke Sakata

In 2013, G8 leaders agreed on the Open Data Charter. The agreement encourages the idea that publishing open data can offer significant benefits. Many countries have started to publish or improve access to their open data. Accordingly, case studies of emerging countries in the field of open data have become extremely important. This article studies the case of India achieving a high score in the 2014 Open Data Index through the establishment and development of India government's portal site for open data: data.gov.in. The article mainly makes two points. Firstly, data.gov.in was designed to draw user suggestions and evaluations for improving open data. Secondly, the number of open data sets is increasing, but the open data sets from the central government and related organizations have been making up the overwhelming majority. Very few open data sets from state governments and related organizations are available. In addition, very few states publish open data in data.gov.in.

Keywords : Open Data, Open Data Index, Open Government Platform, data.gov.in, India

キーワード : オープンデータ, オープンデータ指数, オープンガバメント・プラットフォーム, data.gov.in, インド

I はじめに

2013年6月17日から18日にかけて、イギリスのロック・アーンで開催されたG8サミットでは、「オープンデータ憲章 (Open Data Charter)」および技術的な別添に対するの合意がなされた。この憲章では、オープンデータを「市民のニーズをより良く満たし、繁栄のためのイノベーションと成功を可能とする、より強くより相互の結びつきを深めた社会構築を促進しうる、非常に大きな潜在能力を持つ未開発の資源」(Cabinet Office et al., 2013, p.2)と位置付けた。

オープンデータをこうした「資源」と位置付けた点は、この憲章の非常に重要かつ興味深い部分である。もちろん、オープンデータをこのように捉えることは希少な見解というわけではない。しかし首脳コミュニケによれば、G8各国には遅くとも2015年末までにオープンデータ憲章及び技術的な別添で示されたオープンデータ環境を実現することが要求されている (Prime Minister's Office, 10 Downing Street et al., 2013)。国際的に影響力の大きいG8諸国がオープンデータを資源とみなし、その開発を国家の目標に据えるという状況は、G8以外の各国をもオープンデータ環境の整備へと向かわせるのに十分な圧力を持っていた。

各国によるオープンデータ環境の整備が進むにつれて、これまでオープンデータ環境が整っているとはみなされてこなかった国々が台頭し始めてきている。世界的にオープンデータの発展を推進してきたオープン・ナレッジ (Open Knowledge) (旧オープン・ナレッジ・ファウンデーション (Open Knowledge Foundation)) が、世界センサス (Global Census) の結果に基づき計算、公開しているオープンデータ指数 (Open Data Index) では、2013年に第1位であったイギリスと第2位であったデンマークはその座を保ったものの、それ以外の国々の順位は大きく変動している。例えば、台湾とチェコ共和国は2013年には60か国中の30位台であったが、台湾は25位順位を上げて97か国中の第11位に、チェコ共和国は19位順位を上げて第12位となった。また、2013年にはオープンデータ指数の作成

が行われなかったコロンビア、スウェーデン、ウルグアイがチェコ共和国と同じ第 12 位となった。

日本におけるオープンデータ研究では、高木（2013）や渡辺（2014）による欧州における最新のオープンデータ事情の紹介とそれを通じた日本のオープンデータ整備の今後に対する考察、林（2014）による世界各国のオープンデータ整備状況の紹介と整理、東（2014）による世界各国での幅広いデータ活用事例に対する考察、などといったように海外での興味深い先進事例の紹介や分析が多く行われている。ただし、こうした研究の関心は主にヨーロッパのオープンデータ先進国およびアメリカ合衆国の事例にあり、それ以外のこれまでオープンデータ先進国とはみなされてこなかった国々における取り組みについては言及されたとしてもわずかな分量にとどまってきた。しかしながら、こうした従来のオープンデータ先進国以外についての事例研究も日本のオープンデータ整備の今後に対して十分示唆に富んだものとなるであろうし、加えて、世界中の誰もが利用者となりうるというオープンデータの性質から鑑みても、より広範な事例研究が今後ますます重要となるであろう。

そこで本稿では、上述した国々と同様に、近年大きく台頭したインドにおけるオープンデータへの取り組みに焦点を当てることで、欧米先進諸国以外の事例研究を行うものである。インドは、2013 年には日本と同じ第 27 位であったが 2014 年には世界第 10 位となった（日本は第 19 位）。この順位はインドがオープンデータの最先進国の一つとみなされているアメリカ合衆国（第 8 位）とほぼ同程度の評価であったことを意味している。

オープンデータの対象は幅広く、様々な団体、そして個人が公開主体となり得る。しかしそれらを全て研究の対象とすることは困難であるため、本稿ではまず、国家のオープンデータの基礎となるインド政府もしくは州政府、およびそれらに関連する組織によって公開されたオープンデータに研究対象を限定している。

本稿の構成は以下の通りである。まずⅡにおいて、G8 によるオープンデータ憲章を概観する。すでに述べたように、G8 の国々の国際的な影響力を鑑みれば、このオープンデータ憲章は各国政府がオープンデータ環境の整備を進めるうえで、今後一層重要になっていくと考えられる。本章では、特に憲章において定められたオープンデータ公開における 5 つの原則をもとに、今後の各国政府がオープンデータ環境の整備を進めるにあたり、少なくとも検討の対象とするであろう基本的な道程を確認する。次にⅢにおいて、インドにおけるオープンデータ環境の整備状況に対してなされた評価について、先に述べたオープンデータ指数や各国の一部地域におけるオープンデータ環境を調査したローカル・オープンデータ・センサス（Local Open Data Census）の結果を通じて概観していく。そしてこれらを踏まえたうえで、現在のインドにおけるオープンデータ整備の中心となっているオープンガバメント・プラットフォーム（Open Government Platform）、“data.gov.in”を研究素材とし、インドにおけるオープンデータ環境整備がどのように進んでいるのかを明らかにしていくこととする¹⁾。

Ⅱ オープンデータ憲章

オープンデータの定義として最も普及しているものは、オープン・ナレッジによる定義であろう。このオープン・ナレッジによる定義（Open Definition）は現在バージョン 2.0 が公開されているが、その内容は「誰もが自由にアクセス、利用、修正、共有が可能であるならば、その情報（Knowledge）はオープンだと言える。守る必要があるのは、せいぜいのところ、出所を示すこと、開示性（openness）を保つことだけである。」²⁾と要約されている。

このオープン・ナレッジによる定義と同様に近年重要になってきているのが、上述した、G8 により採択されたオープンデータ憲章である。オープン・ナレッジによる定義が、オープンデータを出来るだけ制限を加えず自由に扱えるデータとして捉えようとしていたのに対して、この憲章では、オープンデータは国家ひいてはそこで生活する人々のために開発

されるべき「資源」と位置付けられている。そして、この資源開発のために G8 各国政府はデータの提供を以下の 5 つの原則にしたがって行うことに合意した (Cabinet Office et al., 2013, p.2)。以下ではこの 5 原則の詳細を見ていくこととする。

- (1) 「原則としてオープンデータであること」 (Open Data by Default)
- (2) 「量と質を保証すること」 (Quality and Quantity)
- (3) 「万人によって利用可能であること」 (Useable by All)
- (4) 「より良いガバナンスのために提供すること」
(Releasing Data for Improved Governance)
- (5) 「イノベーションのために提供すること」 (Releasing Data for Innovation)

(1)の原則では、「オープンデータに対する自由なアクセスとそれに続く再利用が社会経済にとって重要な価値を持つものであるという認識」のもとで、いくつかのデータが正当な理由に基づき公表しえないということは認めつつも、しかしながら、「原則として政府の全てのデータがオープンな形で公開されるであろうという期待感を根付かせる」ことを G8 各国に求めている (Cabinet Office et al., 2013, p.3)。なおここでの政府のデータとは、中央政府だけでなく、広く公的部門が持つすべてのデータを指すとしている (Cabinet Office et al., 2013, p.3)。また、技術的な別添えでは、この原則に関連して、ポータル上でデータを一元的に公開すること、ポータルには全データとメタデータ³⁾をリスト形式にしたレジストリファイルを含めること、API (Application Programming Interface)⁴⁾を備えることが求められている (Cabinet Office et al., 2013, p.6)。

(2)の原則では、「タイムリーで包括的、かつ正確性の高い高品質なオープンデータ」を「データ内の情報が、平易でかつ明確な用語で記述されていることを保証」し、データの利用者がデータの長所短所を把握できるように「こうしたデータが完全に説明されていることを確認」した上で、「出来る限り迅速にデータを公開する」ことを求めている (Cabinet Office et al., 2013, p.3)。

(3)の原則では、利用者の範囲と利用目的の範囲を最大化するために、「可能な限りデータをオープンフォーマット⁵⁾で公開」すること、そして、様々な立場の人々が利用可能なように機械判読可能な形式を含む複数の形式でオープンデータを公開することを求めている (Cabinet Office et al., 2013, p.4)。

(4)の原則では、「オープンデータの公開が、我々の民主主義制度を強固なものとし、市民のニーズを満たすより良い政策立案を推進する」という認識の下で、「誰もがオープンデータの恩恵を享受できるよう、個人間や国家間での技術的専門性と経験の共有」を進めつつ、「我々が保有するデータ収集や基準、公開プロセスの透明化を、関連するプロセスをオンラインでの文章化することにより実施する」ことを求めている (Cabinet Office et al., 2013, p.4)。

(5)の原則では、オープンライセンスや機械判読可能な形式でのデータ公開、API を使ったデータの公開、革新的なデータ利用の奨励 (褒賞や助言) が求められた (Cabinet Office et al., 2013, pp.4-5)。

また、これら原則以外に重要な点として、G8 が「高価値のデータ」とみなすデータの分野と各分野別に具体的なデータセット例を一覧として提示したことが挙げられる (Cabinet Office et al., 2013, pp.9-10)。この一覧は、Ⅲで概観するオープンデータ指数において評価対象となるデータ分野に影響を与えた⁶⁾。

Ⅲ インドのオープンデータ整備環境に対する評価

前節で取り上げた、G8 のオープンデータ憲章の事例のからも明らかなように、すでにオープンデータは国際的な関心事となり、オープンデータ化の推進は多くの国々で基本路線

となっている。そこで本節では、オープンデータ化を推進する上で重要となる、各国のオープンデータ化の程度を評価する試みについて概観したうえで、その評価手法によってインドにおけるオープンデータ整備環境がどのように評価されてきたのか明らかにしていく。対象とするのは、オープン・ナレッジによるインド全体としてのオープンデータ整備状況を評価したオープンデータ指数（Open Data Index）と地域別にオープンデータ整備状況を評価したローカル・オープンデータ・センサスの2つである。

1. オープンデータ指数

オープンデータ指数は、2014年度には世界で97の国や地域から送られてきた情報をもとに作成が行われている（2013年度では60の国や地域についてオープンデータ指数の算出が行われた）。評価対象となったのは以下の10個の分野のデータである。

- ①交通時刻表（Transport Timetables）
- ②政府予算（Government Budget）
- ③政府支出（Government Spending）
- ④選挙結果（Election Results）
- ⑤企業登記（Company Register）
- ⑥全国地図（National Map）
- ⑦国家統計（National Statistics）
- ⑧法律（Legislation）
- ⑨郵便番号（Postcodes / Zipcodes）
- ⑩環境汚染（Pollutant Emissions）

以下では、各データ分野について要求された内容をもう少し詳しく見ていくこととする。①では、全国規模での国営（ないしは国から委託されて運営される）交通サービスについての時刻表が要求されている。②では、分野・部門別などといった、細分化された中央政府予算に関する情報が要求される。③では、実際に行われた政府による支出について、細かな取引レベルでかつ毎月々のデータが求められている。④では、選挙区ないしは地域別での全主要国政選挙の結果が求められている。⑤では、登記された企業名、識別子に加えて所在地や登記された業務内容といったデータが求められる。ただし、バランスシートなどの詳細な財務データを含んでいる必要はない。⑥では、最低でも縮尺が25万分の1の高品質な地図が必要とされる。⑧では、すべての国内法および法令について、情報をオンラインで利用できることが要求される。⑨では、郵便番号に加えて、緯度経度に基づき地理空間の位置情報などが求められる。町名または街路名のためのデータベースでは、他のオープンデータで補完が出来る場合を除き、評価の対象とならない。⑩では、特に健康に対して潜在的に有害と考えられる大気汚染物質の放出についての集計データが要求される（温室効果ガスについての情報は含まなくて良い）。集計は少なくとも全国レベルで、同様に少なくとも毎年行われていることが必要とされる。こうした10分野のデータについて、以下の9項目の評価基準を用いた得点化（単位は%）が行われた。

- ①データが存在しているか（Does the data exist?）
- ②デジタル形式のデータがあるか（Is data in digital form?）
- ③公に公開された利用可能なデータか（Publicly available?）
- ④データは無料で利用可能か（Is the data available for free?）
- ⑤データはオンライン上で利用可能か（Is the data available online?）
- ⑥機械判読可能な形式⁸⁾のデータがあるか（Is the data machine readable?）
- ⑦一括でのデータ⁹⁾利用が可能か（Available in bulk?）
- ⑧オープンなライセンス¹⁰⁾が付与されたデータか（Openly licensed?）

- ⑨データの提供が適切な間隔で行われ、かつ最新のものが提供されているか
(Is the data provided on a timely and up to date basis?)。

これら 9 項目について情報提供者たち (submitters) から得た情報をもとに、オープン・ナレッジが指定した審査員たち (reviewers) もしくはオープン・ナレッジのスタッフによる確認作業、および専門家集団 (panel of expert reviewers) による情報の精査が行われ、得点化がなされる。項目は個々に重みづけがなされており、①～③と⑤は 5 ポイント、⑦と⑨は 10 ポイント、④と⑥は 15 ポイント、⑧は 30 ポイントとして計算がされる¹¹⁾。

表Ⅲ-1 で示されているように、オープン・ナレッジによるインドのオープンデータ指数 (総合) は、2013 年度に 46%であったのに対して、2014 年度では 68%と大幅に上昇した。イギリスやデンマークの値には遠く及ばないものの、68%は世界で 10 番目となる値である (2013 年度は 27 位)。表Ⅲ-2 より、上述の各種データ別にみると、インドは②の政府予算、⑤の企業登記、⑦の国家統計、⑧の法制および⑩の環境汚染の 5 項目で 100%の評価を得ている。反面、得点が低かった項目は③の政府支出、⑨の郵便番号でそれぞれ 10%であった。インドの大幅な指数上昇は②、⑤、⑦、⑧および④ (選挙結果) についてのオープンデータ整備が進んだとみなされたことが大きな要因であった。

2. ローカル・オープンデータ・センサス

ローカル・オープンデータ・センサス (Local Open Data Census) のインド版である、インド都市オープンデータ・センサス (India City Open Data Census) は、アーメダバード (Ahmedabad)、バンガロール (Bengaluru)、チェンナイ (Chennai)、デリー (Delhi)、ハイデラバード (Hyderabad)、コルカタ (Kolkata)、ムンバイ (Mumbai) の 7 都市部を対象に行われており、2014 年に公開された¹²⁾。

ローカル・オープンデータ・センサスの評価方式は、評価対象となるデータ分野にいくつか違いがあるものの、オープンデータ指数に用いられた評価方式とほぼ同様の方式が使われている。評価対象となるデータ分野は以下の計 15 分野である。

- ①リアルタイム交通情報 (Real-Time Transit)
- ②年度予算 (Annual Budget)
- ③財政支出詳細 (Expenditure (detailed)),
- ④選挙結果 (Election results),
- ⑤大気汚染 (Air Quality)
- ⑥交通時刻表 (Transport Timetables)
- ⑦公共施設 (Public Facilities)
- ⑧犯罪統計 (Crime statistics)
- ⑨調達情報 (Procurement contracts)
- ⑩食品の安全性に関する調査 (Food safety inspections)
- ⑪交通事故 (Traffic accidents)
- ⑫建築許可 (Building permits)
- ⑬公共サービスへの要請 (Service Requests)
- ⑭営業許可 (Business Permits)
- ⑮企業リスト (Business Listings)

評価項目は、前述したオープンデータ指数に用いられた 9 項目と同様である。各項目についてイエス、ノー、不詳 (Unsure)、不明 (No Data) のいずれかで評価が示される。得点の重み付けもオープンデータ指数と同様である。

順位	国名／地域名	得点 (%)	得点 変動	順位 変動	順位	国名／地域名	得点 (%)	得点 変動	順位 変動
1	イギリス	97	+3	±0	48	セルビア	42	-2	▼16
2	デンマーク	83	+4	±0		ブルガリア		-11	▼21
3	フランス	80	+21	△9	51	クロアチア	41	-4	▼21
4	フィンランド	73	+1	△3	53	ベルギー	39	+12	△3
5	オーストラリア	72	+6	△4		コスタリカ		+3	▼8
	ニュージーランド		+6	△4	54	ギリシャ	38	-2	▼13
7	ノルウェー	71	-6	▼3		香港		+11	△2
8	アメリカ合衆国	70	-17	▼3	57	中国	37	-5	▼21
9	ドイツ	69	+8	△2		エルサルバドル		—	—
10	インド	68	+22	△17	59	ブルキナファソ	36	-1	▼15
11	台湾	67	+25	△25		タイ		—	—
	コロンビア		—	—	61	マケドニア旧ユーゴスラビ ア共和国	35	—	—
12	チェコ共和国	66	+21	△18		スロバキア		-4	▼18
	スウェーデン		—	—		バングラデシュ		±0	▼16
	ウルグアイ		—	—		バミューダ		-2	▼18
16	アイスランド	64	+11	△1	63	ネパール	34	+4	▼10
	オランダ		-10	▼11		セネガル		-12	▼5
18	ルーマニア	64	+6	-3		シンガポール		±0	▼16
19	チリ	61	—	—		チュニジア		±0	▼16
	日本		+15	△8	69	グアテマラ	33	—	—
21	マン島	60	+5	▼4	70	リトアニア	32	±0	▼20
22	オーストリア	59	+8	△1	71	フィリピン	31	—	—
	カナダ		±0	▼10	72	米領バージン諸島	30	—	—
24	スイス	58	-1	▼12	73	ナイジェリア	29	-1	▼19
25	イタリア	55	+3	▼5		ルワンダ		—	—
	ブラジル		+6	▼2	74	サウジアラビア	28	±0	▼20
26	スロベニア	54	-19	▼20		カンボジア		—	—
	大韓民国		+10	△5	76	ザンビア	27	—	—
28	メキシコ	53	+6	▼2		コートジボワール	26	—	—
	トルコ		—	—	78	エジプト		-6	▼27
	コソボ		—	—	79	モロッコ	25	—	—
31	マルタ	52	±0	▼11		パナマ		—	—
	スペイン		+6	▼4	82	ガーナ	24	—	—
34	ラトビア	51	—	—		ジンバブエ		—	—
35	グルジア	50	—	—	84	カメルーン	23	—	—
	ハンガリー		+6	±0	85	ケニア	22	+2	▼26
36	アイルランド	48	+8	△5		レバノン		—	—
	南アフリカ		+16	△14		ボスニア・ヘルツェゴビナ		—	—
39	ポルトガル	47	-9	▼23		ボツワナ		—	—
40	イスラエル	46	-2	▼16	87	キプロス	21	+18	▼27
41	パキスタン	45	—	—		レソト		—	—
	パラグアイ		—	—		タンザニア連合共和国		—	—
43	エクアドル	44	+1	▼10	92	ベニン	19	—	—
	モルドバ共和国		-13	▼24	93	オマーン	18	—	—
45	インドネシア	43	+1	▼9	94	シエラレオネ	15	—	—
	ジャマイカ		—	—	95	ハイチ	13	—	—
	ロシア連邦		±0	▼12	96	マリ	12	—	—
48	アルゼンチン	42	—	—	97	ギニア	10	—	—
	ポーランド		+6	▼6					

表Ⅲ-1 2014 年度オープンデータ指数（総合）と順位および 2013 年度からの変動

（資料：Open Knowledge, “Place overview 2013” <http://index.okfn.org/place/2013/> (accessed December 31, 2014) および Open Knowledge, “Place overview” <http://index.okfn.org/place/> (accessed December 31, 2014) より筆者作成。）

注）— は 2013 年度にオープンデータ指数の作成が行われていない国または地域であることを意味する。

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	合計
	交通時刻表	政府予算	政府支出	選挙結果	企業登記	全国地図	国家統計	法律	郵便番号	環境汚染	
2014年度の得点	45%	100%	10%	70%	100%	40%	100%	100%	10%	100%	68%
2013年度の得点	45%	25%	60%	35%	60%	70%	40%	45%	45%	35%	46%

表Ⅲ-2 データ分野別にみたインドのオープンデータ指数 (2014年度・2013年度)

(資料：Open Knowledge, “Places / India” <http://index.okfn.org/place/india/> (accessed December 31, 2014)より筆者作成。)

		①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮
		リアルタイム交通情報	年度予算	財政支出詳細	選挙結果	大気汚染	交通時刻表	公共施設	犯罪統計	調達契約	食品の安全性に関する調査	交通事故	建築許可	公共サービスへの要請	営業許可	企業リスト
コルカタ	①データの存在	○	○	○	○	○	○	○	○	○	×	○	○	×	○	○
	⑤オンライン	×	○	○	○	○	○	○	○	○	×	○	×	×	×	×
	⑥機械判読可	／	×	×	○	×	×	／	×	×	×	×	×	×	×	×
デリー	①データの存在	○	○	○	○	○	○	○	○	○	×	○	○	×	○	○
	⑤オンライン	×	○	○	○	○	○	×	×	○	×	○	×	×	×	×
	⑥機械判読可	×	×	×	×	○	×	○	×	×	×	○	×	×	×	×
バンガロール	①データの存在	○	○	×	○	○	○	○	○	○	×	○	○	×	○	○
	⑤オンライン	○	○	×	○	○	○	○	○	○	×	○	×	×	×	×
	⑥機械判読可	／	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ハイデラバード	①データの存在	×	○	○	×	○	○	○	○	○	×	○	○	×	○	○
	⑤オンライン	×	○	○	×	○	○	○	○	×	×	○	×	×	×	×
	⑥機械判読可	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
ムンバイ	①データの存在	×	○	○	○	○	×	○	○	×	×	×	○	×	○	○
	⑤オンライン	×	○	○	○	○	×	○	○	×	×	×	×	×	×	×
	⑥機械判読可	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
チェンナイ	①データの存在	×	○	○	×	○	○	×	×	○	×	×	○	×	○	○
	⑤オンライン	×	○	×	×	×	×	×	×	○	×	×	×	×	×	×
	⑥機械判読可	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
アーメダバード	①データの存在	×	×	×	○	○	×	○	×	×	×	×	○	×	○	○
	⑤オンライン	×	×	×	○	○	×	○	×	×	×	×	×	×	×	×
	⑥機械判読可	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×

表Ⅲ-3 各都市のデータ分野別にみたオープンデータ整備の状況と可能性

(資料：Open Knowledge, “India City Data Census” <http://in-city.census.okfn.org/> (accessed December 31, 2014)より作成)

注) ○はイエス, ×はノー, /は不詳もしくは不明を表す。

表Ⅲ-3 はこのインド都市オープンデータ・センサスのデータより, ①の「データが存在しているか (表中では「データの存在」)」、⑤の「データはオンライン上で利用可能か (表中では「オンライン」)」、⑥の「機械判読可能な形式のデータがあるか (表中では「機械判読可」)」の 3 つについて抜き出したものである。ここから分かるように, 機械判読が可能なデータとして確認がされたものは, コルカタの選挙結果, デリーの大気汚染, 公共施設 (ただし, オンライン利用不可) および交通事故のみであり, 全体的に見て機械判読が可能なデータはほとんど提供されていないことが分かる。データの存在とオンラインについても見てみると, 特にチェンナイとアーメダバードは, 国内有数の人口規模を持つ都市で

あるにもかかわらず、オンライン利用できる分野は非常に少なく、コルカタやデリーと比較すると、データの存在していない分野もかなり多い。大都市でのこうした状況が事実であるのなら、インドにおいて地域的なオープンデータ公開を推進していくことにはかなりの困難が伴うことが予想される。

以上で、オープンデータ指数とローカル・オープンデータ・センサスよりインドの国家としてのオープンデータ整備状況に関する評価と都市におけるオープンデータの現状について概観してきた。インドでは国家レベルでのオープンデータ整備が進みつつある一方で、地方ではオープンデータ化が可能なデータ自体が不足している可能性が浮かび上がった。オープンデータ・センサスによるオープンデータの評価はまだ始まったばかりであり、その信頼性も十分とは言えないが、今後インドがより積極的なオープンデータ化の推進を目指すのであれば、州政府やより下位の地方自治体におけるデータ収集能力の強化が必要となることは明らかであろう。

IV. data.gov.in の整備状況

1. ウェブサイトの開設経緯

2012年3月17日、科学技術局（Department of Science and Technology）によって作成された「インドにおけるデータの共有とアクセシビリティに関する方針（National Data Sharing and Accessibility Policy：以下 NDSAP と略記する）」が正式に採択されたことが告知された¹³⁾。この NDSAP ではその目標を、「インド政府による各種政策・法令・規範の枠内において、人間が判読できる形式と機械判読が可能な形式の双方で保存された、インド政府の保持する共有可能データおよび情報に対して、全国的なネットワークを通じたアクセスを促進させる。これにより、公的データ・情報に対するより広範なアクセシビリティの提供と公的データ・情報の利用を可能にする」（Government of India, 2012, p.95）と定め、オープンデータ公開における各種用語の定義などが行われた。また、NDSAP には、NDSAP の採択通知後 3 か月以内にすべての省（ministry）や局（department）が最低 5 つの高価値なデータセットを、政府によるオープンデータの公開を一元的に行うため web サイト“data.gov.in”で公開することを求める、といった data.gov.in におけるオープンデータの量的拡大を図るための規定が盛り込まれた（Government of India, 2012, p.97）。この NASDP に基づき data.gov.in の発展が進められている。

2. data.gov.in の構造とオープンデータ公開における工夫

data.gov.in には、インド政府がアメリカ合衆国政府と共同で開発した、データポータルを構築するためのプラットフォーム、オープンガバメント・プラットフォーム（Open Government Platform：以下 OGPL と略記する）が用いられている。OGPL は、データ管理システム（Data Management System）、コンテンツ管理システム（Content Management System）、利用者関係管理（Visitor Relationship Management）、およびコミュニティーユーザー用のモジュールの主要 4 モジュールによって構成されている。現在、OGPL はオープンソースとして利用することが可能になっている¹⁴⁾。

data.gov.in は 2014 年 3 月にリニューアルされ機能の向上が図られた。また後述するように利用可能なデータ数の継続的な向上が進められている。以下では、本サイトにおけるいくつかの特徴について解説していくこととする。

(1) 提案

サイト上では提案（suggestion）という形で、ユーザーのニーズ把握に努めている。この提案はサイト内で閲覧可能となっており、寄せられた提案の内、提供が行われた、もしくはすでに行われている場合は、「関連するデータ」として、リンクが提示されている。

(2) データの評価

登録ユーザーは、データセットごとに、その品質 (Quality)、アクセスのしやすさ (Accessibility)、使い勝手 (Usability) を 5 段階で評価することができる。また、サイト上から、データの管理者と連絡を取ることができる。

(3) API リクエスト

登録ユーザーが利用できる、API 機能の付与をリクエストするためのボタンがデータセットごとに用意されている。また、このボタンがクリックされた回数はボタン上に表示され、ユーザーがどの程度そのデータセットに対して API 機能の付与を求められているのかを知ることができる。API 機能の付与の可否はこの API リクエストと上述の提案によって判断される¹⁵⁾。1 月 31 日現在で 7 つの API 機能を含むデータカタログが提供されている。

このように、幅広くユーザーからの意見を汲み上げるという試みがなされていることが、data.gov.in の大きな特徴である。

3. オープンデータの利用および公開状況

まず利用状況について見ていくと、図Ⅳ-1 より、登録ユーザーによる累積ダウンロード件数は、2014 年 4 月のニューズレターでは 449,464 件 (登録ユーザー数 39,331 人) と報告されたのに対して、2014 年 12 月のニューズレターでは 827,891 件 (登録ユーザー数 44,934 人) とほぼ倍増していることがわかる。一分間あたりに閲覧されたページ数も同期中に 1.2 ページから 2.07 ページへと上昇した¹⁶⁾。

次に公開状況について見ると、同じく図Ⅳ-1 が示すように、data.gov.in において公開されたデータセット数は 2 年間で着実に増加を続けてきている。2013 年 2 月のニューズレターに掲載された利用可能なデータセット数は 115 件¹⁷⁾であったが、2014 年 12 月のニューズレターに掲載された利用可能なリソース数 (resource : 2014 年 3 月以降、データセット数 + アプリケーション数を表す用語として使用されている) は 12,708 件、カタログ数は 3369 個¹⁸⁾まで拡大している。アプリケーション数は 2015 年 1 月のニューズレター発行時点で 32 個であった¹⁹⁾。フィルター機能を利用して調べたところ、カタログの内、CSV 形式のファイルを含むものは、約 50%であった。しかしながら、より機械判読に優れた XML 形式のファイルを含むカタログは約 15%と少ない。XLS 形式のファイルを含むカタログは約 40%であった。公開主体の数も増えており、2013 年 2 月のニューズレターによれば、当時データセットを公開していたのは 2 つの省と 7 つの局および計画委員会からなる計 10 の組織のみであったが、2015 年 1 月 31 日現在では、中央政府関連の 67 組織、州政府関連の 18 組織、合計 85 の組織からデータセットが公開されるに至っている。

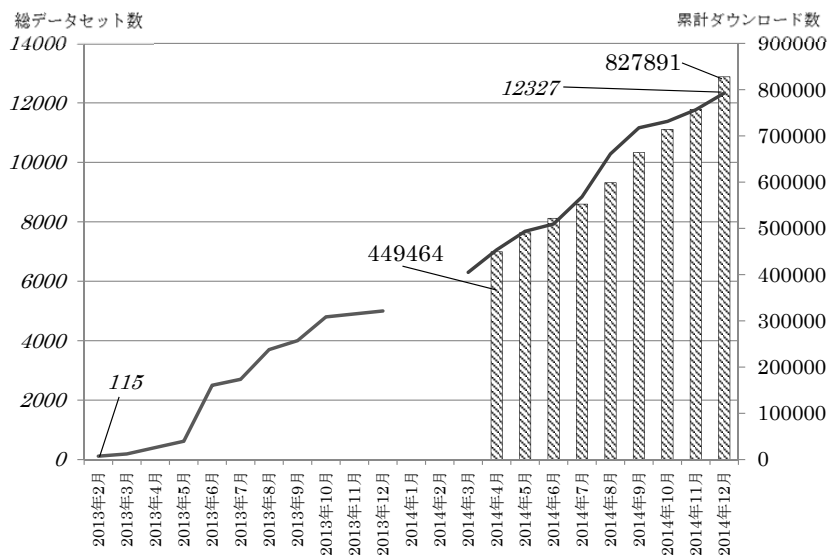
図Ⅳ-2 より分かるように、最も多くのデータセット公開を行っている省は、内務省 (Ministry of Home Affairs) である。データセットの総数が 12,889 件の時点で 3606 件と全体の実に 28%を占めた。これに続いて、農業省 (Ministry of Agriculture) が 16.3% (2100 件)、計画委員会 (Planning Commission) が 12.1% (1560 件)、統計・計画実施省 (Ministry of Statistics and Programme Implementation) が 9.9% (1272 件)、水資源省 (Ministry of Water Resources) が 8.2% (1060 件)、保健家族省 (Ministry of Health and Family Welfare) が 6.7% (868 件) を占めており、これら 4 つの省と計画委員会によって公開されたデータセットだけで、全体の 80%超を占めるに至っている²⁰⁾。図Ⅳ-3 から分かるように、カタログ数で見ると、内務省や農業省の占める割合が大幅に減少する一方で、計画委員会や水資源省の占める割合が大幅に増加し、計画委員会がカタログ数では最大のオープンデータ公開主体になるなどの違いが見られる。しかし、これら 4 つの省と計画委員会が占める割合は 70%と、多くの割合を占めていることに違いはない。

図Ⅳ-2 と図Ⅳ-3 より、data.gov.in での州政府関連によるオープンデータの公開は、データセット数から見ても、カタログ数から見ても、全体に占める割合が非常に小さい。デー

タセット数では 99 件で 0.8%, カタログ数では 99 個で 2.9%を占めるのみである。また、最も多くのデータセット及びカタログを公開しているのがマディヤ・プラデーシュ州 (Madhya Pradesh) であるが、マディヤ・プラデーシュ州が公開しているデータセット数は 60 件 (カタログ数は 60 件) と単独で州政府関連の実に 60%を占めている。さらに、マディヤ・プラデーシュ州の中でも計画・経済・統計局 (Planning, Economics and Statistics Department) は 52 件のデータセット (52 個のカタログ) を公開しており、マディヤ・プラデーシュ州が公開したオープンデータの 9 割近くを占める。

公開を行っている州は、2015 年 1 月 31 日現在でマディヤ・プラデーシュ州をはじめ、タミル・ナードゥ州 (Tamil Nadu), メーガーラヤ州 (Meghalaya), チャッティスガル州 (Chhattisgarh), そしてカルナータカ州 (Karnataka) とインド全 29 州の内の僅か 5 州にとどまっている。2013 年 11 月のニューズレターによれば、ニューズレターの発行時点で 3 つの州が data.gov.in でのオープンデータ公開を行っている²¹⁾。そして、1 か月後の 2013 年 12 月のニューズレターによれば発行時点で 4 つの州が data.gov.in でのオープンデータ公開を行っていた²²⁾。したがって、data.gov.in で 4 つの州がオープンデータ公開を始めた後の約 1 年間に data.gov.in で新たに公開を開始した州は 1 州しかなかったことになる。

このように、data.gov.in でオープンデータを公開する州が少なく、加えてなかなか公開を開始する州が増加しない、さらに公開を行っている州でも一部を除き公開データセットおよびカタログの数が非常に小さい、といった問題が現在の data.gov.in にあることがわかる。Ⅲでも述べたように、地方にはオープンデータ環境の整備に際して非常に多くの問題が存在していると想定される。今後は、州レベルにおいてオープンデータ公開が進まない原因を積極的に明らかにし、州レベルからのオープンデータ公開を促進していくことが、data.gov.in における最も重要な課題となるであろう。



図IV-1 総データセット数と累計ダウンロード数の推移

(資料: Government of India, “News Letters” <http://www.data.gov.in/newsletters> (accessed January 31, 2015)より筆者作成。)

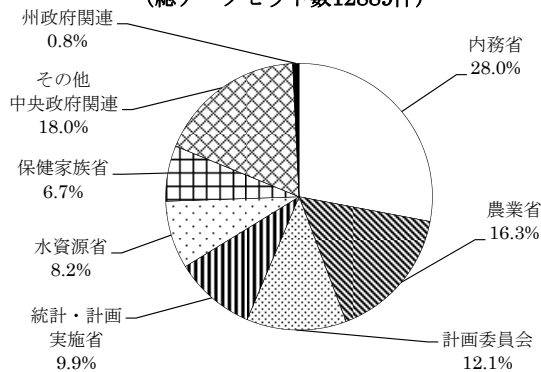
注 1) 年月はデータセット数と累計ダウンロード数を掲載したニューズレターが発行された年月である。

注 2) 2014 年 1 月と 2 月はニューズレターが発行されなかった。

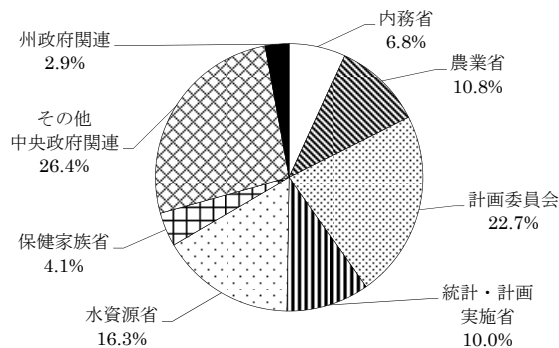
注 3) 2013 年 6 月から 2014 年 3 月までの値はおおよその値である。

注 4) 2014 年 3 月以降のデータセット数の値はデータセット数とアプリケーション数を合わせたリソース数である。

図IV-2 データセットの構成
(総データセット数12889件)



図IV-3 データカタログの構成
(総カタログ数3421個)



(資料：図IV-2と図IV-3ともに、Government of India, “Central Publication Metrics” <http://www.data.gov.in/metrics/central> (accessed January 31, 2015)および Government of India, “State Publication Metrics” <http://www.data.gov.in/metrics/state> (accessed January 31, 2015)より筆者作成)

V. おわりに

まとめと今後の課題を示すことで、本稿を締めくくりたい。

本稿では、2014年のオープンデータ指数で高い得点を得たインドのオープンデータ整備状況を、政府のオープンデータポータルサイト data.gov.in を通じて確認していった。その中で、機械判読により適したXML形式のファイルを含むカタログがまださほど多くなく、API機能の提供もごく少数であるものの、数量面でみると中央政府レベルでのオープンデータの順調な拡大が見られた。加えて、利用環境の改善のため積極的にユーザーからの情報を収集するといった試みがなされていることが確認された。一方で、州レベルでのオープンデータの公開状況は芳しくなく、今後大きな問題になると予想される。

今後の課題として、まず初めに、オープンデータの信頼性や正確性の解明が挙げられる。オープンデータ指数など、現在のオープンデータの評価基準はこうした点を対象としておらず、本稿でも研究対象とすることは出来なかった。しかし、信頼性や正確性の検討なしにオープンデータが本当の意味で普及することはかなわないだろう。特に、州レベルやそれ以下の地方自治体におけるデータの信頼性や正確性に対する研究は必要不可欠だろう。

次に、はじめにでも述べたようにオープンデータの対象は幅広く、様々な団体、そして個人が公開主体となり得る。しかし本稿では公的なオープンデータにその対象を限定した。今後はその他の公開主体も研究の対象に加えていきたい。

最後に、オープンデータについての研究を行う際には、公開側だけでなく利用側も重要な研究の対象となりうる。しかしながら、オープンデータの利活用を研究の対象に含めることは本稿ではかなわなかった。これも今後の課題としたい。

注

- 1) オープンデータがあらゆる組織または個人によって公開されたオープンデータを指すのに対して、オープンガバメント・データは、基本的にはガバナンスの改善や公共サービスの向上などを目指して、政府もしくは地方自治体といった行政機関から公開されるデータを指している。本稿では、行政機関によるオープンデータの整備状況に焦点を当てているため、考察の対象は、厳密にはオープンガバメント・データとなる。しかしながら現時点では、その性質がオープンガバメント・データであっても、オープンデータと呼称されることが多い。よって、これらを区別して用いることは今の段階ではかえって無用の混乱を生む恐れがあるため、オープンガバメント・データと呼称の方が適切な場合であっても、オープンデータという呼称を用いて考察を進める。
- 2) Open Knowledge, “Open Definition” <http://opendefinition.org/od/> (accessed December 31, 2014)
- 3) メタデータとは、データについての説明をするためのフィールドや要素を意味している。(Cabinet Office et al., 2013, p.7)
- 4) オープンデータにおける API は、プログラムを通じて機械判読可能なデータを取得できるようにする仕組みを意味している。
- 5) 技術的な別添えでは、オープンフォーマットの事例として CSV 形式を挙げている(Cabinet Office et al., 2013, p.7)。
- 6) Open Knowledge, “Methodology” <http://index.okfn.org/methodology/> (accessed December 31, 2014)
- 7) ①~⑩の内容は Open Knowledge, “Methodology” <http://index.okfn.org/methodology/> (accessed December 31, 2014)にもとづく。
- 8) 解説では、機械判読が可能なデータの例として XLS 形式, CSV 形式, JSON 形式, XML 形式を挙げている (Open Knowledge, “Methodology” <http://index.okfn.org/methodology/> (accessed December 31, 2014))。
- 9) ここでの、「一括でのデータ利用が可能」とは、対象となるデータについて、データの一部が利用不可などになっておらず、一括した全体利用が可能であること意味する (Open Knowledge, “Methodology” <http://index.okfn.org/methodology/> (accessed December 31, 2014))
- 10) ここでの「オープンなライセンス」とは、オープン・ナレッジによるオープンデータの定義に従っていることを意味する。解説には、データの利用、再利用、再配布が自由であることがオープンなライセンスの要素として挙げられている (Open Knowledge, “Methodology” <http://index.okfn.org/methodology/> (accessed December 31, 2014))
- 11) Open Knowledge, “Methodology” <http://index.okfn.org/methodology/> (accessed December 31, 2014)
- 12) Open Knowledge, “India City Data Census” <http://in-city.census.okfn.org/> (accessed December 31, 2014)
- 13) Chattapadhyay, Sumandro., “Timeline of Open Government Data in India.” <http://ajantriks.github.io/oddc/resources/timeline.html> (accessed January 25, 2015)
- 14) Government of India, “About Open Government Data (OGD) Platform India” <http://www.data.gov.in/about-us> (accessed January 31, 2015)
- 15) Government of India, “How to use Datasets APIs” <http://www.data.gov.in/help/how->

- use-datasets-apis (accessed January 31, 2015)
- 16) Government of India, “Communication News Letter April 2014” <http://www.data.gov.in/news-letters/communication-news-letter-april-2014> (accessed December 31, 2014)および Government of India, “Communication News Letter December 2014” <http://www.data.gov.in/news-letters/open-government-data-platform-india-datagovin-newsletter-dec-2014> (accessed December 31, 2014)。また、サイトでは、マップやグラフの公開も行われており、2015年1月31日現在で、411種類が公開されている。
- 17) Government of India, “Communication News Letter February 2013” <http://www.data.gov.in/news-letters/communication-news-letter-february-2013> (accessed December 31, 2014)
- 18) Government of India, “Communication News Letter December 2014” <http://www.data.gov.in/news-letters/open-government-data-platform-india-datagovin-newsletter-dec-2014> (accessed December 31, 2014)
- 19) Government of India, “Communication News Letter January 2015” <http://www.data.gov.in/news-letters/open-government-data-platform-india-datagovin-newsletter-jan-2015> (accessed January 31, 2015)
- 20) オープンデータを提供しているその他の中央政府関連組織については、Government of India, “data.gov.in / Metrics / Central” <http://www.data.gov.in/metrics/central> を参照されたい。
- 21) Government of India, “Communication News Letter November 2013” <http://www.data.gov.in/news-letters/communication-news-letter-november-2013> (accessed December 31, 2014)
- 22) Government of India, “Communication News Letter December 2013” <http://www.data.gov.in/news-letters/communication-news-letter-december-2013> (accessed December 31, 2014)

参考文献

- [1] 高木聡一郎, 2013, 「欧州におけるオープンデータ政策の最新動向」『情報管理』Vol.55, No.10, pp.746-753.
- [2] 林 雅之, 2014, 『オープンデータ超入門』インプレス R&D.
- [3] 東 富彦, 2014, 『データ×アイデアで勝負する人々』日経 BP 社.
- [4] 渡辺智暁, 2014, 「欧州から考える政府のオープンデータ国際戦略」『智場』国際大学 GLOCOM, 119, pp.64-76.
- [5] Cabinet Office, 2013 *G8 Open Data Charter and Technical Annex*, GOV.UK, <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex> (accessed January 25, 2015)
- [6] Government of India, 2012, *The Gazette of India Weekly 2012 March 17 – March 23*, <http://data.gov.in/sites/default/files/NDSAP.pdf> (accessed January 25, 2014)
- [7] Prime Minister's Office, 10 Downing Street, The Rt Hon David Cameron MP, Cabinet Office, Department for Business, Innovation & Skills, Department for International Development, Foreign & Commonwealth Office, HM Treasury and UK Trade & Investment, 2013 *2013 Lough Erne G8 Leaders' Communique*, GOV.UK, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207771/Lough_Erne_2013_G8_Leaders_Communique.pdf (accessed January 25, 2014)