

# 自治体におけるオープンデータのデータ品質

## Data quality of open data in municipalities

本田 正美†  
Masami Honda

### 1. 研究の背景

2012 年に電子行政オープンデータ戦略が公表された。これに基づき、日本では公共機関におけるオープンデータの推進が図られてきた。自治体では、同戦略の公表前からオープンデータの取り組みが進められていたところもあり、同戦略の策定や政府の支援も相まって、2021 年 7 月現在で 1000 を超える自治体がオープンデータに取組んでいる[1]。オープンデータは、[2]で次のように定義付けられている。

国、地方公共団体及び事業者が保有する官民データのうち、国民誰もがインターネット等を通じて容易に利用（加工、編集、再配布等）できるよう、次のいずれの項目にも該当する形で公開されたデータをオープンデータと定義する。

- ① 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
- ② 機械判読に適したもの
- ③ 無償で利用できるもの

この定義では、データ公開に関わるルールや形式について言及されているものの、データの品質や量についての言及はない。

全国の自治体においてオープンデータの取り組みが進展するなかで、内閣官房情報通信技術(IT)総合戦略室は、推奨データセットを公表している。これは、政府として公開を推奨するデータと、データの作成にあたり準拠すべきルールやフォーマット等である[3]。

推奨データセットには、基本編と応用編がある。そのうち、基本編は「オープンデータに取り始める地方公共団体向け」とされており、公開を推奨するデータセットとして、「AED 設置箇所一覧」や「介護サービス事業所一覧」などがあげられている。

個別の自治体が独自のルールやフォーマットでオープンデータを公開すると、例えば「AED 設置箇所一覧」として公開されたオープンデータであっても、各自治体間で異なる形式のデータセットが公開されることになり、利用者が自治体を超えて「AED 設置箇所一覧」に関するデータセットを収集して利用する際に、データのフォーマットなどの調整が必要とされることになる。オープンデータとして公開されるデータの品質や量にばらつきがあると、利用者には不都合が生じることもあるといえる。

とりわけ、データの品質はデータの利用にあたって特に重要となる事柄であり、品質の劣ったデータがいくら公開されたところで、それではデータ利用自体に大きな手間が生じることから、オープンデータの取り組みの企図するところのデータ利用も進んでいかなないことになる。

### 2. 既往研究の整理と本研究の目的

自治体において進展するオープンデータの取り組みについては、単にデータを公開するだけでなく、その利活用が指向されていたところであり、それは[2]において「容易に利用（加工、編集、再配布等）できるよう」という一文があることからうかがえる。

オープンデータの公開が広がるなかで、その利用者特性に関わり[4]のような調査がなされている。オープンデータの活用事例に関しては、[5]が研究レビューを行っている。その他、公共機関から公開されたオープンデータの利活用として商用利用の可能性を探った[6]のような研究もある。

オープンデータの公開と利用が広がるなかで、そのデータ公開のあり方も問われるところとなっている。[7]は、統計データの公開のあり方を検討したものである。

[8]のようにオープンデータ間の連携の可能性を検証した研究や[9]のようにオープンデータに対するメタデータとして付与されるタグに着目した研究も見受けられる。オープンデータの公開と利用が進むなかで、あらためてオープンデータのデータそのものにも関心が向かっているのである。

本研究では、オープンデータの取り組みにつき、その数的な広がりがあるなかで、オープンデータとして公開されるデータの品質に焦点を当てる。具体的には、自治体において公開されるオープンデータについて、そのデータ品質がいかなる水準にあるのかを検証する。

なお、データ品質については、内閣官房情報通信技術(IT)総合戦略室が公表した[10]において、政府の策定を目指すデータ戦略のなかで「データ標準と品質の整備」という項目を設けることが示されていることから推察されるように、その重要性が増している。そして、2021 年 6 月には、「データ品質管理ガイドブック」の β 版が公表されるなど、日本政府としてその取り組みを推進しているところである。

### 3. 研究の方法

本研究では、自治体におけるオープンデータのデータ品質につき、都道府県で公開されているオープンデータに着目する。これは、全都道府県がオープンデータに取組んでいることから、その同定が容易であるからである。

研究方法としては、まず都道府県でオープンデータを公開している Web サイトを確認する。都道府県におけるオープンデータの公開場所は[1]などを参照することによって同定可能である。そこで、都道府県においてオープンデータを公開している Web サイトにアクセスした上で、当該団体としてオープンデータの実施にあたって、データ品質を考慮しているのか。当該 Web サイト上に、データ品質に関する何らかの言明があるのか否かを確認した。そして、データ品質に関する言明があれば、その品質の評価方法に基づき、その水準を検証することとした。

本研究が扱うところの情報は、2021 年 7 月 21 日時点のものである。

† 関東学院大学, Kanto Gakuin University

## 4. 結果

以下、都道府県においてオープンデータを公開している Web サイトの名称の一覧である。

表 1 都道府県のオープンデータ公開サイト

団体名	サイト名称	団体名	サイト名称
北海道	北海道オープンデータポータル	滋賀県	滋賀県オープンデータカタログ
青森県	青い森オープンデータカタログ	京都府	京都府オープンデータポータルサイト
岩手県	岩手県オープンデータサイト	大阪府	大阪府オープンデータサイト
宮城県	オープンデータみやぎ	兵庫県	ひょうごオープンデータカタログサイト
秋田県	秋田県オープンデータカタログ	奈良県	奈良県オープンデータカタログサイト
山形県	山形県オープンデータカタログ	和歌山県	和歌山県オープンデータサイト
福島県	オープンデータライブラリ	鳥取県	鳥取県オープンデータポータルサイト
茨城県	茨城県オープンデータカタログ	島根県	島根県オープンデータカタログサイト
栃木県	オープンデータ・ベリ〜とちぎ	岡山県	おかやまオープンデータカタログ
群馬県	群馬県オープンデータサイト	広島県	広島県オープンデータライブラリ
埼玉県	埼玉県オープンデータポータルサイト	山口県	山口県オープンデータサイト
千葉県	千葉県オープンデータサイト	徳島県	徳島県オープンデータポータルサイト
東京都	東京都オープンデータカタログサイト	香川県	香川県オープンデータ
神奈川県	神奈川県オープンデータサイト	愛媛県	愛媛県オープンデータカタログ
新潟県	新潟県オープンデータ	高知県	高知県オープンデータ
富山県	富山県オープンデータポータルサイト	福岡県	福岡県オープンデータサイト
石川県	石川県オープンデータカタログ	佐賀県	佐賀県オープンデータカタログサイト
福井県	福井県オープンデータライブラリ	長崎県	長崎県オープンデータカタログサイト
山梨県	山梨県オープンデータサイト	熊本県	熊本県オープンデータサイト
長野県	長野県オープンデータサイト	大分県	大分県オープンデータカタログサイト
岐阜県	岐阜県オープンデータカタログサイト	宮崎県	宮崎県オープンデータカタログサイト
静岡県	ふじのくにオープンデータカタログ	鹿児島県	鹿児島県オープンデータカタログサイト
愛知県	愛知県オープンデータカタログ	沖縄県	沖縄県オープンデータカタログ
三重県	三重県オープンデータライブラリ		

(出所：47 都道府県の Web サイトを確認の上で筆者作成)

表 1 にあるところの各都道府県でオープンデータを公開している Web サイトにおいて、データ品質に関する言明があるのか否か確認をしたところ、47 団体全てで、その種の言明がなされていないことが分かった。いずれも、その名称にカタログとあることから推察されるように、あくまでもオープンデータを公開することを基本としている。その都道府県におけるオープンデータの取り組みにおいて、データ品質への配慮がなされている現況は確認されなかったことになる。

ただし、長野県と大阪府は、表 1 に記載してあるサイト上で、政府が示した推奨データセットに準拠したデータの公開を行っている旨が明示されていた。推奨データセットへの準拠が直ちにデータ品質への配慮を意味するわけではないが、オープンデータの実施にあたり、日本政府が示す方針に準拠しようとする姿勢がうかがわれ、ここからはデータ品質にも一定の配慮が及ぶ可能性があると考えられる。

## 5. 考察と結論

全ての都道府県でオープンデータに取組済となっているが、その取り組みにあたって、データ品質に配慮されている様子はいかがえなかった。つまり、本研究で企図したところの、オープンデータのデータ品質がいかなる水準にあるのかを検証するという作業は、その前段階の部分で終了したことになる。都道府県においては、データ品質の評価基準を示さずに、現段階では公開出来るデータをオープンデータとして公開している段階にあると結論付けられる。

なお、本研究では、あくまでもオープンデータが公開されている Web サイト上での言明しか確認していないため、例えばオープンデータの実施にあたり実務を担当している

部署の運営上の規則や都道府県に策定が法律で義務付けられている官民データ活用推進計画において、データ品質への言明があり、何らかの基準が組織内部で設けられている可能性は捨てきれない。ただし、そのような基準があったとしても、それを対外的に公表していない、あるいは少なくともオープンデータを公開しているページ上で明記していないという現況には変わりがない。

前述のように、日本政府はデータを公開するだけでなく、その品質も重要な要素となることを認識しており、「データ品質管理ガイドブック」の中には、データ品質評価モデルという項目が設けられている。この取り組み自体、2021 年に β 版が公表されたことから分かるように、データ品質は近年になって注目された事柄ではある。今後、都道府県においてもオープンデータの取り組みを進める中で、政府の方針も参照しながら、データ品質への配慮も行っていくことも予想されるところである。

## 参考文献

- [1] 政府 CIO ポータル「オープンデータ取組自治体一覧」  
[https://cio.go.jp/sites/default/files/uploads/documents/opendata\\_1\\_g\\_list.xlsx](https://cio.go.jp/sites/default/files/uploads/documents/opendata_1_g_list.xlsx) (2021 年 7 月 21 日アクセス)
- [2] オープンデータ基本指針  
[https://cio.go.jp/sites/default/files/uploads/documents/data\\_shishin.pdf](https://cio.go.jp/sites/default/files/uploads/documents/data_shishin.pdf) (2021 年 7 月 21 日アクセス)
- [3] 政府 CIO ポータル「推奨データセットについて」  
[https://cio.go.jp/sites/default/files/uploads/documents/opendata\\_suisyou\\_dataset.pptx](https://cio.go.jp/sites/default/files/uploads/documents/opendata_suisyou_dataset.pptx) (2021 年 7 月 21 日アクセス)
- [4] 小舘亮之, 大室良介, 若原俊彦, 曾根原登 (2021) 「オープンデータのデータ利活用を促進させる利用者特性に関する探索的調査」『電子情報通信学会技術研究報告; 信学技報』, 120(417), pp.35-40
- [5] 井上絵理, 谷口尚子 (2019) 「市民活動をサポートするオープンデータ活用 国内外の実践と研究に関するレビュー」『情報通信学会誌』, 37(4), pp.91-97.
- [6] 本田正美 (2020) 「オープンデータの商用利用の可能性」『経営情報学会全国研究発表大会要旨集 2020 年全国研究発表大会』, pp. 109-112
- [7] 中村英人, 石野洋子 (2019) 「地方公共団体のオープンデータへの取組: 統計データ公開のあり方の検討」『社会情報学』, 8(2), pp.79-94
- [8] 久永忠範, 淵田孝康 (2019) 「述語ベクトルを用いたオープンデータ間の連携可能性の研究」『情報知識学会誌』, 29(2), pp.123-128
- [9] 長谷川誠, 山田泰寛 (2019) 「分散表現を用いた政府オープンデータにおけるタグの階層構造の構築」『研究報告情報システムと社会環境 (IS)』, 2019(7), pp.1-7
- [10] 内閣官房情報通信技術 (IT) 総合戦略室 (2020) 「データ戦略の策定について」  
[https://www.kantei.go.jp/jp/singi/it2/dgov/data\\_strategy\\_tf/dai1/siryou1.pdf](https://www.kantei.go.jp/jp/singi/it2/dgov/data_strategy_tf/dai1/siryou1.pdf) (2021 年 7 月 21 日アクセス)

謝辞

本研究は公益財団法人セコム科学技術振興財団特定領域研究助成「民主制下における地方自治体の情報公開・オープンデータと情報セキュリティとの交錯に関する研究」の研究成果の一部である。