

行政データの利活用に関する有識者会議

～論点整理等資料及び中間報告書（案）～

令和4年8月8日

企画調整局 政策課

論点整理等資料

ダッシュボードの庁内共有の基準（案）

○当面の間、下記の基準に従い運用する。

ダッシュボード加工前のデータを基準（以下、元データとする）として、分類した別表中「元データ取扱レベル①から④」に従い、レベル①②については個人を容易に特定できないことを前提に、粒度の別なく職員間で共有可、レベル④については共有不可とする。

レベル③については、個人が特定されないように加工したダッシュボード（統計加工後のデータ）を基準に、明らかに「差別に繋がるもの、市民の利益が害されるリスク」が考えられないものについては、レベル①②と同様に粒度の別なく職員間の共有可、恐れがある場合は不可とする。

レベル③については、不可とした場合であったとしても、区別であれば職員間の共有を可とする。

○上記共有の判断は、データ所管課と企画調整局政策課が協議の上行う。

○上記の基準に従い、職員間の共有を行ったが事後的に問題が発生したと認められる場合、一旦共有を休止し、データ所管課、企画調整局政策課、複数の有識者で協議を行い、再度職員間の共有を可とするか不可とするかについて判断する。

(別表) 統計加工に用いる元データのレベルについて

元データ レベル	元データの種類	元データの状態
元データ 取扱レベル①	オープンデータ、センサーデータなど	統計加工データ
元データ 取扱レベル②	個人が特定された場合の影響度が小さいデータ 住基データ、建築確認申請など	抽象加工データ 個票
元データ 取扱レベル③	個人が特定された場合の影響度が大きいデータ 税データ、生活保護データ、就学援助データなど	抽象加工データ 個票
元データ 取扱レベル④	要配慮個人情報を含むデータ	抽象加工データ 個票

基準（案）の運用イメージ

閲覧権限レベル
 ①：外部公開可
 ②：全職員閲覧可
 ③：個別対応

元データの種類		ダッシュボード				備考
元データレベル	蓄積 → 統計加工 →	共有可能かの判断		全職員閲覧	閲覧権限レベル	
元データ取扱レベル①	オープンデータ センサーデータ など			○	閲覧権限レベル①	
元データ取扱レベル②	住基データ 建築確認申請 など			○	閲覧権限レベル②	共有後疑義が生じた場合 →有識者も入れた協議
元データ取扱レベル③	税データ 生活保護データ 就学援助データ など	個人が特定されないように統計加工	考えられないもの (=センシティブではない)	○		
			考えられる恐れがあるもの (=センシティブである恐れ)	○		
元データ取扱レベル④	要配慮個人情報を含むデータ			×	閲覧権限レベル③	

明らかに「偏見の惹起に繋がるもの、市民の利益が害されるリスク」が考えられないものかどうか

考えられないもの
(=センシティブではない)

考えられる恐れがあるもの
(=センシティブである恐れ)

区別集計結果

区より小さいエリアの集計結果

閲覧権限レベルごとに元データを類型化（例）

※③についてはダッシュボードを基準に分類

元データ レベル	元データ（具体例）	閲覧権限 レベル	統計情報（内容）	
元データ 取扱 レベル①	e-Stat オープンデータ	閲覧権限 レベル①	国勢調査（役職、人口ピラミッド、住居種類）	
	e-Stat オープンデータ		エリア別従業員数	
	e-Stat オープンデータ		鉄道乗降客数ダッシュボード	
	センサーデータ		河川の水位	
	センサーデータ		PM2.5濃度、CO2濃度	
元データ 取扱 レベル②	建築確認申請		住所、主要用途、構造、戸数、敷地面積、建築面積	
	住基データ個票		年齢、住所（町丁目）、続柄、世帯番号、国籍、在留資格	
元データ 取扱 レベル③	国勢調査個票（取り寄せ）	閲覧権限 レベル②	年齢、家族構成、職業、地位、勤務先市町村（政令市は区まで）	ダッシュボード（統計加工後のデータ）を基準で判断 →センシティブではない
	住）固定資産税データ個票		固定資産税納付開始時期、納税額 ※市内の固定資産分（市外在住者含む）	
	法人市民税データ		業種、税別納税額	
	住）市民税データ個票	区別集計 →レベル②	所得、納税額	ダッシュボード（統計加工後のデータ）を基準で判断 →センシティブである恐れ
	住）介護保険データ	区より小さい →レベル③	要介護度、介護保険サービス別利用状況	
	住）生活保護データ		生活保護受給額、受給開始時期	
元データ 取扱 レベル④	住）国保レセプトデータ	閲覧権限 レベル③	医療費、罹患率分布	
	健康診断		健診受診履歴、検診受診結果	

※ 住）：住基データと紐づけられたデータ

個人情報由来の行政データ利活用で「判断」が生じる場合の整理

閲覧権限レベル	行為	具体的に何をしているか	アクセス	データ活用の位置付け	行為を行う時の判断
閲覧権限レベル②	外部公開（オープンデータ、ダッシュボードなど）	抽象加工データを、統計加工データに加工し、外部に公開する	市民	個人情報の目的外利用（統計加工）	所管課と政策課が協議の上 <div style="border: 1px solid orange; padding: 5px;"> （問題が生じた場合の対応） 外部に公開後、問題が生じた場合、一旦公開を休止し、上記のメンバーのほか、複数の有識者を入れた協議を行い、改めて判断を行う </div>
	新たにダッシュボードを作り、職員間で共有	抽象加工データを使って、ダッシュボード（統計情報）を作り、ダッシュボード（統計情報）を職員間で共有する	原則全職員	個人情報の目的外利用（統計加工）	所管課と政策課が協議の上 <div style="border: 1px solid orange; padding: 5px;"> （問題が生じた場合の対応） 庁内共有後問題が発生した場合、一旦共有を休止し、上記メンバーのほか、複数の有識者を入れた協議を行い、改めて判断を行う </div>
閲覧権限レベル③	ダッシュボードなどの個別対応	抽象加工データを使って、ダッシュボード（統計情報）を作り、ダッシュボード（統計情報）を、政策立案等の目的で依頼のあった所属に共有・提供する	権限のある職員	個人情報の目的外利用（統計加工）	所管課と政策課が協議の上

中間報告書（案）の内容

1 会議開催の経緯と趣旨

2 論点

- (1) 庁内における行政データの利活用について
- (2) オープンデータについて

3 論点における委員の主な意見

(1) 庁内における行政データの利活用について

- ① (元データ取扱レベル2) 住基データに基づく統計情報 (ダッシュボード) などの共有
- ② (元データ取扱レベル3) 個人情報に基づく統計情報のうち、内容がセンシティブではないと判断された集計結果 (ダッシュボード) の共有
- ③ (元データ取扱レベル3) 個人情報に基づく統計情報のうち、内容がセンシティブの恐れがあると判断された区別集計結果 (ダッシュボード) の共有
- ④ (元データ取扱レベル3) 個人情報に基づく統計情報のうち、内容がセンシティブの恐れがあると判断された区より小さいエリアの集計結果 (ダッシュボード) の共有
- ⑤ (元データ取扱レベル4) 要配慮個人情報に基づく統計情報の共有
- ⑥ 抽象化データ (個人情報) を活用した統計情報 (ダッシュボード) 作成の運用方法

(2) オープンデータについて

(3) その他の議論 (匿名加工情報の有用性について)

論点における委員の主な意見 ①

(1) 庁内における行政データの利活用について

① (元データ取扱レベル2) 住基データに基づく統計情報 (ダッシュボード) などの共有

- 住基情報については、国勢調査の情報が、i)国籍や在留資格が区で1人でも公開されていること、ii)神戸市では、従来から町丁目別に1歳階級別で男女別の人口を庁外に公開しており、町丁目別の男女別合計人口が6以下にならない限り秘匿していないこと、iii)渋谷区では、住基データを全職員が分析できるよう運用していることなどを踏まえると、小学校区別であれば、統計情報を秘匿せず庁内共有しても問題ないと考えられる。
- 町丁目別の統計情報でも庁内ではできる限り共有すべきである。

② (元データ取扱レベル3) 個人情報に基づく統計情報のうち、内容がセンシティブではないと判断された集計結果 (ダッシュボード) の共有

- 税データに基づく統計情報であっても、固定資産税の納税開始時期の集計結果や共働き率、法人市民税であれば企業数が少ない業種は法人が特定できないよう配慮したうえでの集計結果などは、差別につながると思えない情報であり、庁内共有は問題ないと考えられる。
- 国勢調査の個票自体は慎重に取り扱う必要があり、クロス集計の結果が3未満は秘匿処理が必要とされるなど、運用面が厳しいが、統計情報のセンシティブ性は低く、庁内共有は問題ないと考えられる。
- 国勢調査などのデータは、センシティブ性が低く、ポイントは容易照合性の排除であり、エリアが狭まるほど、秘匿処理の必要性が高まり、必要な作業が増える。

論点における委員の主な意見 ②

- ③（元データ取扱レベル3）個人情報に基づく統計情報のうち、内容がセンシティブの恐れがあると判断された
区別集計結果（ダッシュボード）の共有
- 要配慮個人情報ではない個人情報に基づく統計情報（税情報など）は、当面個別判断とのことだが、一見センシティブと思えるものでも、区別であれば統計情報として公開されているものも多く、個人が特定されるリスクはほぼないので、庁内共有であれば問題ないと考えられる。
 - 所得情報についても、高額所得者などを配慮して集計すれば、個人が特定されるリスクがほぼなく、生活保護受給世帯数は区別集計であれば各自治体が公開しているほか、所得情報は民間住宅情報サイトで区別の推計平均値が公開されていることなどを踏まえると、区別集計であれば庁内共有は問題ないと考えられる。
 - 政策立案に使うには、区別ではエリアが広すぎるが、データの存在やこういった分析が可能だと示すことには意味がある。
 - 政策立案を行ううえで、詳細なエリアの特性を知ることが重要であり、できるだけ広く庁内共有する方向で考えた方が良い。

論点における委員の主な意見 ③

- ④（元データ取扱レベル3）個人情報に基づく統計情報のうち、内容がセンシティブの恐れがあると判断された区より小さいエリアの集計結果（ダッシュボード）の共有
- たとえば、税データに関しては、地方税法第22条があり、全国的に税情報が慎重に扱われている状況で、統計情報から個人が特定された場合のリスクは大きい。今後、小学校区別のダッシュボードを共有していく場合に、東京大学の研究チームの方法を参考に個票データを匿名加工してから統計情報に加工すれば、個人を特定されるリスクがかなり低くなるのではないか。
 - 元データを匿名加工してしまえば、個人情報保護法の観点からは問題がなくなるが、統計情報に誤差が生じると、抜き取りであれば再現性もなくなり扱いにくくなる。庁内利用でそこまでする必要があるのであるのかと思うが、匿名加工して免責にするという判断はある。
 - 年齢層や所得の幅を広くし、できるだけ個票をそのまま集計した方が良い。
 - 個人が特定できないレベルで統計処理ができるのであれば、地域特性がより精緻に分かる小学校区別のデータはむしろ政策形成に有用であり、積極的に共有すべきである。
 - エリアをメッシュで表現する方法も有効である。境界が変わらないので、毎年度のデータの連続性が担保できる。会津若松市が実施しているように、住民の居住地を緯度・経度で管理できれば可能になる。
 - システム側で、アクセス権やアクセスログの管理などのコントロールを徹底することが重要で、それが確保できれば庁内で共有することについての問題はなくなるのではないか。システムでのコントロールの確保が、行政データの活用について、市民に対する説明にもなる。
 - 今後センシティブな情報を共有するうえで、アクセスログが管理できる仕組みは必須と考える。
 - アクセス権限を管理し、データリテラシーやセキュリティの研修と閲覧権限を連動させて管理できると良い。

論点における委員の主な意見 ④

- ④（元データ取扱レベル3）個人情報に基づく統計情報のうち、内容がセンシティブの恐れがあると判断された区より小さいエリアの集計結果（ダッシュボード）の共有
 - ダッシュボードを共有するうえで、グラフを正しく読める研修が重要である。相関と因果関係の違い、中央値と平均値の違いなどの基礎知識と、実際のデータを見たときに、裏側に何があるのかを解釈する訓練が必要である。
 - 閲覧権限レベル3のセンシティブ性については、独自に定義した方が良いのではないか。
 - 多くの個人情報に基づく統計情報は、レベル3も含めセンシティブ性に基づく共有範囲や共有内容の是非について、当面個別に判断していくしかないと考えられる。

- ⑤（元データ取扱レベル4）要配慮個人情報に基づく統計情報の共有
 - 要配慮個人情報は、犯罪歴や病歴、人種（国籍ではない）など明らかにセンシティブであり、広く共有しないことは当然として、個別対応でも分析する機会は、ほぼないのではないか。
 - 個人情報保護法上の要配慮個人情報と行政が判断するセンシティブな情報は必ずしも一致しない。

論点における委員の主な意見 ⑤

⑥ 抽象化データ（個人情報）を活用した統計情報（ダッシュボード）作成の運用方法

- 神戸市個人情報保護条例では、統計情報作成のための個人情報の利用は、目的外利用の類型答申ですでに認められており、また、改正された個人情報保護法でも統計利用と学術研究での利用は、例外として認められている。また、庁内の運用面では、個別にデータ利用課がデータ所管課にデータ利用申請を提出し、承諾を得たうえで利用することになっており、運用上の問題はない。
- 政策形成のためにそのデータを使って良いのかをデータ所管課で判断するとなると、閲覧権限の範囲や個人情報を使う妥当性の判断が難しく、利用承認に慎重になることが懸念される。一方で倫理委員会のような第三者を入れて行う仕組みは様々なコストがかかる。
- ダッシュボードを作る際、共有の可否も含めデータ所管課の許可を取って企画調整局政策課で作成する方法がうまくいっているのであれば、それで良いと思う。
- 一連の統計情報の作成・共有に関する議論は、理解しておくべき細かな内容が多く、現在推進している職員が異動しても、後任が判断できるようにしておく必要がある。

(2) オープンデータについて

- 個人情報から作成した統計情報は、オープンデータとしてできる限り市民に公開していくべき。また、オープンデータそのものをデータとして扱えない市民に対してもダッシュボードとして公開することで、市民がより簡単にデータヘアプルーチできるようになる。
- 国の基幹統計は、調査ごとに秘匿化する基準を決めているのでそれに合わせて決定するのが良い。
- オープンデータとして外に出していくときの判断のプロセスを決めておいた方が良い。
- 思いもよらぬ直接的被害を受ける人たちをどのように保護するかは考えておく必要がある。窓口を作るとか、第三者委員会を開かなくていいように常設で第三者機関を設けておくことが必要と考える。
- 架空のリスクばかりで新しい価値を創出する可能性を狭めてしまうのはよくない。基本的にはオープンにする方針で、懸念される事故が起きたときにどういった対処をするのかを考えた方が良い。何かあったときに平常にすぐ戻ることを考えることが重要である。
- オープンデータにしたとき、通報窓口や審査会のようなクレームを受け付ける窓口を作り、第三者機関を何らかの形で入れた方が良い。
- 各課が判断するとなると出したときのリスクを各課が負うことになり、どうしても出さない方向に振れてしまう。仕組みを作っておくことが重要である。
- 医療分野では、病院ごとにどの手術を行ったかを国が大胆に公開しており、どの地域でその病気が多いかがある程度推察できる。地域性が分かるデータも統計情報であれば、出していけるのではないか。
- ダッシュボードを公開する際は、神戸市のメッセージ性が加わることから客観的に確認する仕組みがある方が良い。

(3) その他の議論（匿名加工情報の有用性について）

- 東京大学政策評価研究教育センター（CREPE）が行った「EBPM推進のための自治体税務データ活用プロジェクト」の匿名加工情報（※）は、総務省にも見解を伺いながら匿名化の手順を作成し、実際に自治体から税情報に基づく所得情報を集め分析につなげる実績ができています。これまで、税情報を外部利用する事例はほぼなかったことから、個票に近い形で外部利用した点では、大きな実績だと考えています。
- 学術研究が目的であれば、個人情報利用は認められる自治体は多いが、税情報については、実際には出してもらえないので、こうした匿名加工が必要になっていると思われる。
- 税以外のデータについても、様々なユースケースがあり、こうした匿名加工情報であったとしても、個票ベースで取り扱うことができるのであれば、研究に非常に役立つ行政データがあると思う。カナダの小児科病院で、児童の医療情報を匿名化したうえで、研究者や民間に提供し、研究・創薬などに使用したケースがある。こうした匿名加工データは、学術研究などに非常に役立つと確信している。

※ 具体的な手順

- ① 氏名、個人番号（いわゆるマイナンバー）、住所等を事前に削除
- ② 特異な世帯（多子世帯等）の世帯番号の秘匿
- ③ 宛名番号、世帯番号のハッシュ化
- ④ 宛名番号をキーにして、複数年のデータを結合
- ⑤ 所得、賦課額等のトップコーディング（上位1%）
- ⑥ 生年月日の月単位への丸め（※前日に変換）
- ⑦ 生年月、性別、郵便番号について、「3 - 匿名性」を判定し、満たさないものは、満たすまで秘匿化
- ⑧ 各世帯番号（のハッシュ値）について、50%の確率でのランダムサンプリング