

官民におけるデータの利活用について

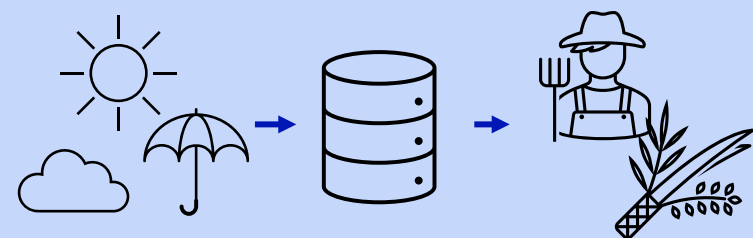
令和 7 年 4 月 1 日
内閣官房デジタル行財政改革会議事務局

行政分野におけるデータの利活用の3つ場面

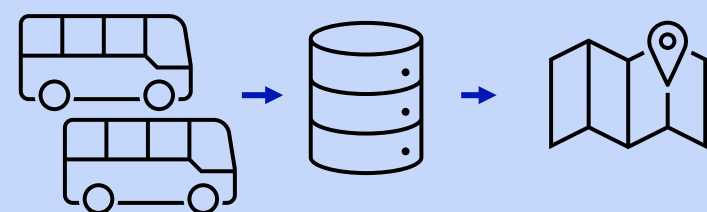
①行政のデータを社会が利活用（オープンデータとして国民や民間企業が活用）、②行政内部データを行政が利活用（政策立案や業務改善など）③行政外部のデータを行政が利用（民間データを取り入れて施策に活用）という3つの場面が想定される。

①行政データを社会が利活用

例：気象データを用いた農業計画策定
（農家によるスケジュールの調整等）

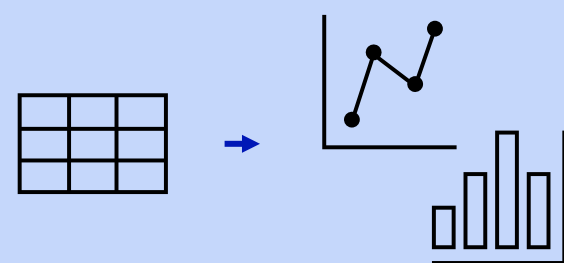


例：市バスデータによる利用状況可視化
（GISによる混雑傾向の地図表示）

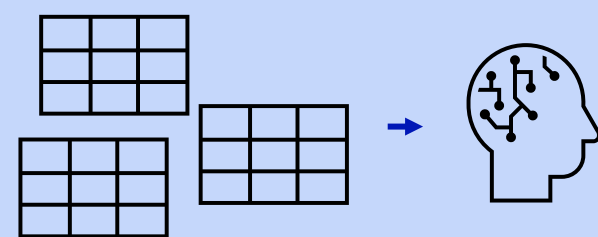


②行政内部データを行政が利活用

例：税務データを用いた分析
（所得分布や税収動向の分析）

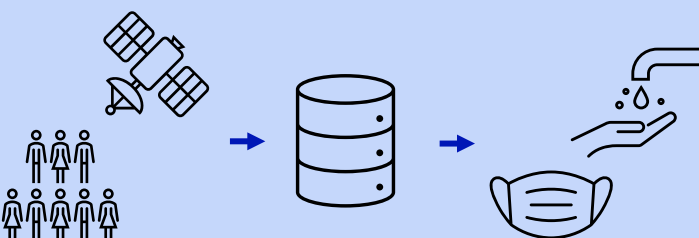


例：住民・就労データ等を用いたAI審査

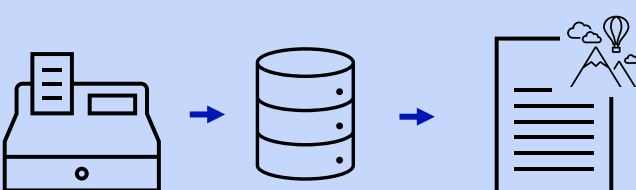


③行政外部のデータを行政が利用

例：人流データを用いた感染症対策



例：購買データを用いた政策企画立案
（観光や地方創生等）

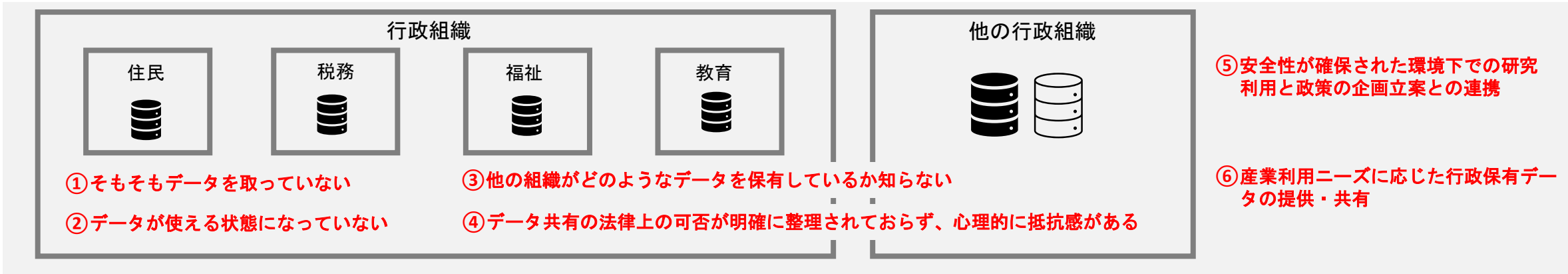


データの取得・整形・活用における課題

データ取得

データ整形

データ分析・利活用



行政データ利活用推進に向けた課題の整理

第7回検討会において事務局から提示した「全体像」をベースに、利用者から聴取した行政データ利活用に関する課題を試みにプロットしたもの。
(※聴取内容の詳細については参考資料参照)

- (1) データが円滑に流通するよう、「形」や「道」を整備する
- (2) データの流通が促されるよう、「勾配」(インセンティブ)をつける
- (3) 安心してデータ利活用するため、「場」をつくる

データの品質確保 (機械可読性)

データが紙ベースで保管されている

デジタルデータも表記ゆれ等が多い

スクレイピングに対応できていない

データ活用のメリットの共有

データホルダー側の利点が分からない

組織内外での活用方法が想像できない

具体のユースケースで利点を説明する必要

法令の運用・分析環境

利活用の具体例の整理・共有は有効

安全性と利便性が両立する分析環境

具体例ベースに法的な論点を検討

データ連携基盤の活用

実業務で使えるか検証が必要

※具体的なデータ形式や連携要件は分野別の検討が必要

官民のリエゾン機能

外部から行政保有データ内容の詳細が不明

複数分野のデータ結合の方法

諸外国の事例の調査研究が必要

(4) データ利活用に向けての司令塔機能の強化 デジタル人材、データ利活用人材の育成

データ利活用推進の体制整備

分野横断的な課題の担当組織の不在

職員のデータリテラシー向上

データに関する基礎知識は全職員に必要

職員のデータ分析能力の向上

実践機会提供・外部人材との協働が有効

AI-OCRによるデジタルデータ化の技術的進展（「機械可読性」は今後も変化する）

大規模言語モデルの革新に伴い、「事前学習なしに」、「低コストで」、これまでのOCRと同水準で画像データをテキストデータに変換可能に。これまで眠っていたデータ活用が容易に実現。

年度	2018年以前		2018年～2023年	2024年	2025年～
技術要素	従来のAI-OCR機能（事前学習が必要）		LLMを使ったOCRの進化（事前学習が不要）		
			ChatGPT等の LLMの誕生初期	画像処理可能な マルチモーダルモデルの登場	画像処理精度の高い モデルの登場
できたこと	・手書き文字認識の精度向上 ・定型帳票のデータ抽出	・AIの学習能力向上 ・多言語対応の強化	※初期モデルは画像認識不可	・画像の前処理技術向上（歪み・汚れ対応、低解像度画像対応） ・非定型帳票への対応	・文脈理解での適切書き起こし（レイアウト・言葉選択） ・専門性の高い表の理解
できなかったこと	・事前学習なしでのOCR ・非定型帳票への柔軟な対応	・事前学習なしでのOCR ・複雑なレイアウトへの対応	※初期モデルは画像認識不可	・文脈を理解した処理 ・スペースや特殊レイアウトが混ざった場合の処理	

事前学習なしでの画像認識は不可

OCRできても、活用しづらい

利活用しやすいデータに変換

			発注書																																																																																																	
アイディニエス 御中			No TH-1001																																																																																																	
			発行日	2024/10/1																																																																																																
下取りあり、発送版しず。			電話商事株式会社 〒100-0001 東京都千代田区千代田1-1-1 アイプリビル南館 TEL：03-1111-2222 担当：電歴太郎																																																																																																	
内容 納品場所	2024年 10月 31日																																																																																																			
支払条件	月末締め翌月払込																																																																																																			
合計																																																																																																				
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 40%;">題目</th> <th style="width: 10%;">数量</th> <th style="width: 10%;">単位</th> <th style="width: 10%;">単価</th> <th style="width: 30%;">金額</th> </tr> </thead> <tbody> <tr> <td>JV - 55T プレート</td> <td>50</td> <td>式</td> <td>1,500</td> <td>57,500</td> </tr> <tr> <td>JV - 55T スト</td> <td>40</td> <td>式</td> <td>1,150</td> <td>46,000</td> </tr> <tr> <td>JV - 66T フラット</td> <td>10</td> <td>式</td> <td>11,500</td> <td>11,500</td> </tr> <tr> <td>JV - 90T スト</td> <td>50</td> <td>式</td> <td>1,600</td> <td>57,500</td> </tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td>式</td><td> </td><td> </td></tr> <tr> <td colspan="3"></td> <td> 小計 </td> <td align="right">0</td> </tr> <tr> <td colspan="3"></td> <td> 消費税 </td> <td align="right">0</td> </tr> <tr> <td colspan="3"></td> <td> 合計 </td> <td align="right">0</td> </tr> </tbody> </table>						題目	数量	単位	単価	金額	JV - 55T プレート	50	式	1,500	57,500	JV - 55T スト	40	式	1,150	46,000	JV - 66T フラット	10	式	11,500	11,500	JV - 90T スト	50	式	1,600	57,500			式					式					式					式					式					式					式					式					式					式					式						小計	0				消費税	0				合計	0
題目	数量	単位	単価	金額																																																																																																
JV - 55T プレート	50	式	1,500	57,500																																																																																																
JV - 55T スト	40	式	1,150	46,000																																																																																																
JV - 66T フラット	10	式	11,500	11,500																																																																																																
JV - 90T スト	50	式	1,600	57,500																																																																																																
		式																																																																																																		
		式																																																																																																		
		式																																																																																																		
		式																																																																																																		
		式																																																																																																		
		式																																																																																																		
		式																																																																																																		
		式																																																																																																		
		式																																																																																																		
		式																																																																																																		
		式																																																																																																		
			小計	0																																																																																																
			消費税	0																																																																																																
			合計	0																																																																																																
備考欄 納品は工場に依頼します。 納入期が厳しく急ぎです。ご検討下さい。																																																																																																				

発 注 書	アイブリー工業 南中	ス 不
No TH-1001		
発注日 2024/10/1		
下記のとおり、発注致します。		
合計		
摘要数量単位単価金額		表 こ
IV - 987 プレート		
50		
1,150 57,500		
IV - 553 プレート		
40		
1,150 46,080		
IV - 661 フラット		
10		
11,500 11,500		
IV - 987 プレート		
50		
1,150 57,500		
合計		
備考		
納品は工場にお願いします。		
納期分りましたらご連絡下さい。		

発注書

文脈を理解して適切 に書き起こし

アイブリー工業 御中

No TH-1001

発注日 2024/10/1

下記のとおり、発注致します。

合計

表形式で書き起こし

摘要	数量	単位	単価	金額
--- --- --- --- ---				
IV-987 プレート	50	式	1,150	57,500
IV-553 プレート	40	式	1,150	46,080
IV-661 フラット	10	式	11,500	11,500
IV-987 プレート	50	式	1,150	57,500
小計 0				
消費税 0				
合計 0				

備考

納品は工場をお願いします。

納期分かりましたらご連絡下さい。

(出典) 表はヒアリングを元に事務局作成、サンプルは  **IVRY** 提供

機械可読性を確保する必要性

画像・図表データについてもシームレスにデジタル化が可能に。ただし、元の図表のレイアウトにより、手作業で対応せざるを得ない部分が残る。

紙データ

デジタル化

表-1 観測機器の更新状況

黄 色 : 目視から機械式へ更新
緑 色 : 機械式から機械式へ更新
着色なし: 更新なし

セル結合部分はうまく処理できない

コード	観測所	観測項目		開始年	機器名称	受感部設置位置		沖合距離	水深	観測機器更新回数
		波高	波向			北緯	東経			
001	沖ノ鳥島	あり	—	1998	マイクロ波式	20°25'00"	136°04'00"			
011	九十九里	あり	あり	1999	超音波式	35°36'31"	140°32'21"	8,000m	22.00m	
101	中の沢	あり	あり	1979	目視	42°26'00"	140°20'00"	400m	8.00m	
111	虎杖浜	あり	あり	1979	目視	42°28'50"	141°15'00"	300m	7.00m	1
				1993	超音波式	42°28'50"	141°15'09"	1,350m	18.00m	2
112	富浜	あり	あり	1979	目視	42°29'20"	142°00'50"			
114	石倉	あり	あり	1979	目視	42°07'00"	140°27'00"	400m	9.00m	
115	稀府	あり	あり	1979	目視	42°27'06"	140°52'16"	300m	8.00m	
116	梟舞	あり	あり	1979	目視	42°13'00"	142°37'10"	300m	7.00m	
121	恵山	あり	あり	1998	海象計 USW-1000	41°46'36"	141°10'40"	1,600m	50.00m	
211	西田沢	あり	—	1979	階段抵抗式	40°51'34"	140°41'38"	280m	4.00m	1
				1985	超音波式	40°51'50"	140°41'54"	400m	11.50m	2
212	平館	あり	—	1979	水圧式ストレンゲージ式	40°59'21"	140°38'26"	290m	10.00m	1
				1987	超音波式	41°09'21"	140°38'26"	290m	10.00m	2
214	広瀬	あり	あり	1995	超音波式	41°00'43"	140°40'02"	1,200m	10.80m	
216	浪江	あり	—	1979	水圧式ストレンゲージ式	37°09'12"	141°04'48"		7.50m	1
				1985	超音波式	37°29'00"	141°03'00"		9.00m	2
311	関屋	あり	あり	1990	超音波式	37°54'35"	138°58'46"	1,600m	16.00m	
321	田中	あり	あり	1979	超音波式	36°57'21"	137°29'41"	517m	12.77m	
323	石田	あり	あり	1998	超音波式	36°51'48"	137°24'42"	470m	17.87m	
331	徳光	あり	あり	1990	超音波式	36°32'30"	136°31'54"	1,500m	15.00m	
411	原	あり	あり	1979	水圧式ストレンゲージ式	35°07'20"	138°46'59"	150m	20.00m	
414	駿河海洋	あり	あり	1985	超音波式	34°47'29"	138°18'44"	250m	6.00m	

表-1 観測機器の更新状況

コード	観測所	波高波向	開始年	機器名称	北緯	東経	沖合距離	水深	観測機器更新回数
001	沖ノ鳥島	あり	1998	マイクロ波式	20°25'00"	136°04'00"			
011	九十九里	あり	1999	超音波式	35°36'31"	140°32'21"	8,000m	22.00m	
101	中の沢	あり	1979	目視	42°26'00"	140°20'00"	400m	8.00m	
111	虎杖浜	あり	1979	目視	42°28'50"	141°15'00"	300m	7.00m	1
112	富浜	あり	1979	目視	42°29'20"	142°00'50"			
114	石倉	あり	1979	目視	42°07'00"	140°27'00"	400m	9.00m	
115	稀府	あり	1979	目視	42°27'06"	140°52'16"	300m	8.00m	
116	梟舞	あり	1979	目視	42°13'00"	142°37'10"	300m	7.00m	
121	恵山	あり	1998	海象計 USW-1000	41°46'36"	141°10'40"	1,600m	50.00m	
211	西田沢	あり	1979	階段抵抗式	40°51'34"	140°41'38"	280m	4.00m	1
212	平館	あり	1979	水圧式ストレンゲージ式	40°59'21"	140°38'26"	290m	10.00m	1
214	広瀬	あり	1995	超音波式	41°00'43"	140°40'02"	1,200m	10.80m	
216	浪江	あり	1979	水圧式ストレンゲージ式	37°09'12"	141°04'48"		7.50m	1
311	関屋	あり	1990	超音波式	37°54'35"	138°58'46"	1,600m	16.00m	
321	田中	あり	1979	超音波式	36°57'21"	137°29'41"	517m	12.77m	
323	石田	あり	1998	超音波式	36°51'48"	137°24'42"	470m	17.87m	
331	徳光	あり	1990	超音波式	36°32'30"	136°31'54"	1,500m	15.00m	
411	原	あり	1979	水圧式ストレンゲージ式	35°07'20"	138°46'59"	150m	20.00m	
414	駿河海洋	あり	1985	超音波式	34°47'29"	138°18'44"	250m	6.00m	

データの機械可読性を高めるための新しいルールのイメージ

遵守すべきルールは既に整備されているが、全てを守ろうとするとハードルが高い。
このため、レベル別に遵守事項を分け、比較的取組やすいところから段階を踏んで質を高めるための工夫は有効ではないか。

レベル	基準	具体例
レベル 1	表形式＋テキストデータの基準を満たす	- オブジェクトや画像を使わない - 書式でデータの違いを表現しない（黄色い網掛けNG） - セルを結合しない - ファイル形式はExcelかCSV 1セルに1データしか入れない - スペースや改行で体裁を整えない 1シートに1つの表を入れる - 行や列を非表示にしない - 表の外側にメモや備考を記載しない - 機種依存文字を使わない
レベル 2	単体で統計データの基準を満たす	- 数値列には数値しか入れない - 選択枝列と「その他」の詳細記入を分ける - 表側の項目名を省略しない - 回答なし（空欄）の扱いを定める
レベル 3	時系列・横断的に統計データの基準を満たす	- 選択枝回答をコード表記にする - 回答のコード表を別添する（1=取組済、2=検討中、など） - 設問マスター（列・変数）を別添する - 調査概要、数値項目の単位、列・変数の備考等のメタデータを別添する - 列数が多い場合は縦型のレイアウトにする

統計表における機械判読可能な
データ作成に関する表記方法（総務省）

1-1	ファイル形式はExcel かCSV となっているか
1-2	1セル1データとなっているか
1-3	数値データは数値属性とし、文字列を含まないこと
1-4	セルの結合をしていないか
1-5	スペースや改行等で体裁を整えていないか
1-6	項目名等を省略していないか
1-7	数式を使用している場合は、数値データに修正しているか
1-8	オブジェクトを使用していないか
1-9	データの単位を記載しているか
1-10	機種依存文字を使用していないか
1-11	e-Stat の時間軸コードの表記、西暦表記又は和暦に西暦の併記がされているか
1-12	地域コード又は地域名称が表記されているか
1-13	数値データの同一列内に特殊記号（秘匿等）が含まれる場合
2-1	データが分断されていないか
2-2	1シートに複数の表が掲載されていないか

ご議論いただきたい論点(例)

1. データの品質確保に関するルール策定と徹底

- AI-Ready社会において行政データを機械が読み込むことができる状態にしておくことが重要である。
- 機械可読性を確保するためのルールにはどのような内容を盛り込むことが適当か。
- また、このようなルールを徹底するためにはどのような取組が有効か。

2. データ利活用のユースケースの創出と具体例に基づく法的課題の整理

- データ利活用を推進するためには、利活用のメリットをデータ提供者・保有者に理解してもらうことが重要。
- 具体的なユースケースを積み上げることでメリットと安全性を説明していく必要があるが、ユースケースを把握・蓄積し、個々のケースにおける法的な整理のエンドースや現行法令のボトルネックの整理などは、どこがどのように担うべきか。

3. 行政データに関する横断的課題への対応、職員の育成

- 分野毎にデータ利活用に関する課題を議論する場はあるものの、データ一般の政府全体の窓口は存在しない。
- 正確性の確保等に必要となるデータ整備に要するリソースの確保、データ取得のために行う調査の効率化、実務におけるデータ利活用上の課題の相談・調整窓口、具体的なユースケースの事例の蓄積などデータ利活用に関する分野横断的な課題を担当する組織が必要ではないか。
- 公共データは国民共有の財産であり公開が原則との考え方を示した「オープンデータ基本指針」に基づく取組を徹底するため、データ整備や公開等に要するコストや課題に対するルールや技術的ソリューションを提示すべきではないか。また、優先して対応すべき分野にはどのようなものがあるのか。
- データ利活用のユースケースを創出するためには、データ分析に長けた職員の育成・確保が不可欠。その際、単なる研修にとどまらず、実践の機会を提供することに加え、分析環境の整備も重要となる。また、アカデミアや民間シンクタンク等の分析能力の高い職員が行政内部で分析業務に参画することなどが考えられる。こうした取組を実現するためには、どのような工夫が必要か。

(参考)行政データ利活用上の課題(ヒアリング結果のまとめ)

考え方

利用者起点の課題

データ品質

データが使える状態になっているか？



そもそもデータを取っていない、あっても使える状態になっていない、使うためには膨大なコストがかかる状態をいかに避けるかが重要

【機械可読性】

- データが紙ベースで保管されていることもあり、まずはデジタルデータ化が必要
- AIは縦書き・横書きが混在した文書の読み込みが苦手
- 多年度に及ぶデータで、年度間で表記ゆれが生じているケースもある
- データの表記ゆれが随所にありクレンジングにコストがかかるがAIの活用も選択肢
- データ作成上の基本的なルールが徹底されていない
- データの取得・整形にどの程度のコストをかけるかは利用目的別に個別に判断すべき。
- 企業等が当該データを利活用可能か判断するために必要な情報が不足（レコード数、データレイアウト、メタ情報等）
- ベンダが保有しているデータの提供を受ける場合、定義が不明なデータ系列が存在する。

【分析の質と匿名化・仮名化のレベル】

- 匿名化が必要なレベルは利用目的によって異なる（例：アプリが動くかどうかの確認であれば疑似データでも可能）
- 厳密な分析を行う場合は、限りなく生データを使った方が精度が高まるケースもある
- データの匿名化には技術的な限界もあることから、厳格な要件を課し、セキュアな環境下でユースケースを積み上げていくべき
- トップコーディングすると分析上の価値を損ねるケースもある

連携基盤

データは掛け合わせでより効果を発揮する



分野を超えたデータを繋ぐ連携基盤とは、データ形式・連携要件の標準化という詳細かつ地道な決め事の積み重ねであるとの理解が重要

【データ統合（データリンケージ）】

- データの管理方法やフォーマットが統一されていないことがデータ統合の障害
- 統一コードを一元的に付与・管理する仕組みがないことで、個別案件ごとに手作業で紐づけが必要
- システム間で統一コードが異なることにより、住民情報の更新がシステム間で同期されずコード変更や重複が生じる
- 分析目的によって結合手法は異なる（確率論的な処理でカバー可能な場合もある）
- 法人番号によってデータ連携が容易になったが、個人に関しては識別子がなく共通IDで繋ぐことができる国とは前提条件が異なり中長期的に解決すべき課題。
- 同一IDに複数のデータがくっついているケースもある。
- 組織間（データ間・制度間）の移動を捕捉できていない
- 自治体には、データをまとめて結合する専門部署を持っているところもある
- 基礎自治体のデータを集めたい場合、1,741のホームページを訪問する必要がある
- 分散しているデータを繋ぐために仮名化IDを発行する独立機関の存在によって、匿名化を実現している国も
- API連携のための技術基準など、業務で実際にデータが動くかは実証が必要

人材

データが使える人がいるか？



データ利活用には標準化も含めてコストがかかるため、コストを上回るメリットがあることを使う人が理解することが重要

【リテラシー】

- データ活用の具体的なイメージがないと、手元のデータの価値が分からない。
- 職員にデータに関する知識が不足（自分の部署はデータを保有していないという反応がある）
- データ分析が仕事としてカウントされないので行政職員が時間を割くことに二の足を踏む（専門的な組織がある自治体等では取組が進む）

【人材育成】

- データアナリストなどが政策立案に参加できれば分野は広がるが、研究成果が出るには時間がかかり、政策に活かせる事例が増えない。外部人材が政策立案の実務に携われる仕組みが必要。
- 何が必要なデータなのかは、議論や実践の中から見えてくることもあり、政策立案や行政実務としてデータを取扱う機会が増えることが重要
- 業務レベルでのデータ分析・利活用を行える職員を育成するためには、研修等の知識のインプットだけでは不十分。現実的な課題の解決に取り組み、試行錯誤を重ねながら、実務の中で実践していくことが不可欠。その際、分析に取り組むための環境整備も併せて行うことで、継続的なスキル習得と成果の創出が可能に。

ルール

利活用を阻害するルールはあるか？



ルール上問題がなくても利活用が進まないこともあり、ルールの問題か否かを実態に即して判断していくことが重要

【ルールの遵守を担保する仕組み】

- 法律だけではなく、法令の運用組織、利用するための仕組み、利用者にルールを遵守させるための仕組み、技術的な基盤等、トータルで制度を考える必要
- 官デ法やオープンデータ基本指針など、基本ルールは存在しているが、各主体にどのように徹底するか（その前提としての現状の把握）が課題
- 一般論で議論するのではなく、利用目的や分野毎に具体的なメリットやリスクの程度を丁寧に整理し、具体事例に即してボトルネックを発掘すべき。
- 分析側は従来とは異なるデータを活用することで新たな洞察を得られメリットがあるが、データホルダー側はメリットが見えずリスクのみを背負うという片務性がネックになっている可能性もある
- 機微なデータを使う人と場面を限定すべき

【既存法令に関する知識・理解】

- 法令上当該データを提供して問題ないか現場で判断し切れないケースもある
- 国や自治体のHPに掲載されているデータを二次利用する際に法令上の制約があるのか分からない（スクレイピングの障害になり得る）