

# DataOpsを活用したオープンデータ利活用に関する報告書

---

Version : 0.1

公開日 : 2024年xx月xx日

著者 : C3Lab



By C3Lab

このドキュメントの内容物は [クリエイティブ・コモンズ 表示 4.0 ライセンス](#) [^1]の下に提供されています。

# 目次

## 1 はじめに

### 1-1 本報告書の位置づけ

### 1-2 背景

## 2 検証イベント実施内容

### 2-1 オープンデータDataOps検証イベント

### 2-2 検証イベント詳細

### 2-3 イベントにて実施したアンケート結果

## 3 考察

## 4 まとめ

## 5 今後の展望

# 1 はじめに

## 本報告書の位置づけ

2022年度版の「[DataOpsを活用したオープンデータ利活用に関する報告書](#)」（以下、2022年度報告書）で提案したDataOpsプロセスのオープンデータ利活用への有効性を確認する為、その最初のアプローチとして、DataOpsプロセスの作業を体験する開発イベントを開催した。

本報告書は開発イベントで得た成果や課題を元に、有効性を確認する為の改善案をとりまとめたものである。

## 背景

2022年度報告書ではオープンデータ利活用の阻害要因例として以下の問題が確認されている。

- 必要なデータが探しづらい
- 必要なデータが公開されていない
- オープンデータの認知度が低い
- 利用者のニーズが解らない

この阻害要因を解消する為の手法としてDataOpsの概念に焦点を当てている。

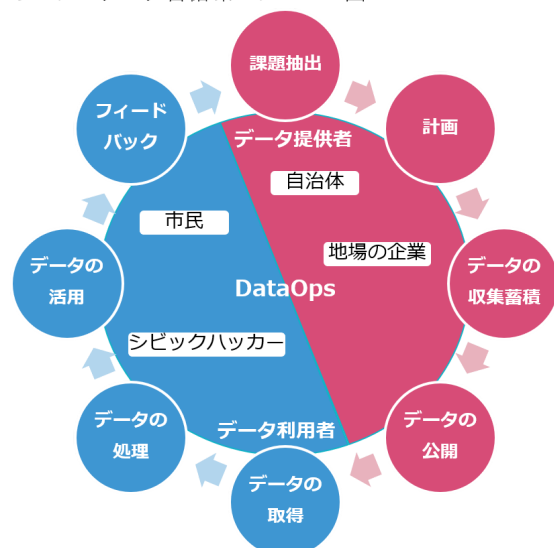
## DataOpsとは

ガートナー社の定義[<sup>2</sup>]によると「組織全体のデータ管理者とデータ利用者の間のコミュニケーションの向上と、データフローの統合と自動化の改善に焦点を当てた共同作業によるデータ管理の手法」と定義している。DataOpsの概念を利用することで、利用者が必要なタイミングで信頼できるデータや分析結果を入手可能となり、データの利用者と提供者が協調することでプロセス、ツール環境や組織文化の継続的な改善が期待できる。

この概念を利用することでこれらの阻害要因を解消することが可能であるという仮説に基づき、DataOpsのプロセスを提案した。

以下の図は、提案したDataOpsのプロセス図とその内容である。

オープンデータ善循環のプロセス図



プロセス	内容
課題抽出	<ul style="list-style-type: none"><li>データで解決可能な課題の発見・言語化・登録</li><li>課題の状況の把握</li></ul>
計画	<ul style="list-style-type: none"><li>オープンデータ活用計画の立案</li></ul>
データの収集蓄積	<ul style="list-style-type: none"><li>新規データの収集・加工</li><li>収集済みのデータの加工</li><li>データの蓄積</li><li>リアルタイムデータの公開準備</li></ul>
データの公開	<ul style="list-style-type: none"><li>データの公開</li><li>リアルタイムデータの収集・加工・公開</li><li>カタログ/検索インデックス登録</li></ul>
データの取得	<ul style="list-style-type: none"><li>オープンデータの検索</li><li>オープンデータの取得・確認</li><li>ソフトウェアからのオープンデータの取得</li><li>オープンデータの再取得・確認</li></ul>
データの処理	<ul style="list-style-type: none"><li>データ前処理</li><li>取得データ保存</li><li>取得データのメタ情報保存(データ取得元など)</li></ul>
データの活用	<ul style="list-style-type: none"><li>課題へ対応</li><li>課題の解決</li></ul>
フィードバック	<ul style="list-style-type: none"><li>オープンデータの利活用状況のフィードバック</li></ul>

※各ステークホルダーやユースケースの詳細については[2022年度報告書](#)を参照

この提案の有効性については検証中であり、有効性の確認の為、イベントを行うこととした。

また、OECD（経済協力開発機構）が取り組む、各国のオープンデータの取り組みを評価する為のデータ指標「OURdata Index」によると、2019年の調査結果では日本の総合順位が36カ国中4位だったものの、2023年の調査結果では40カ国中25位という調査結果が報告されている。

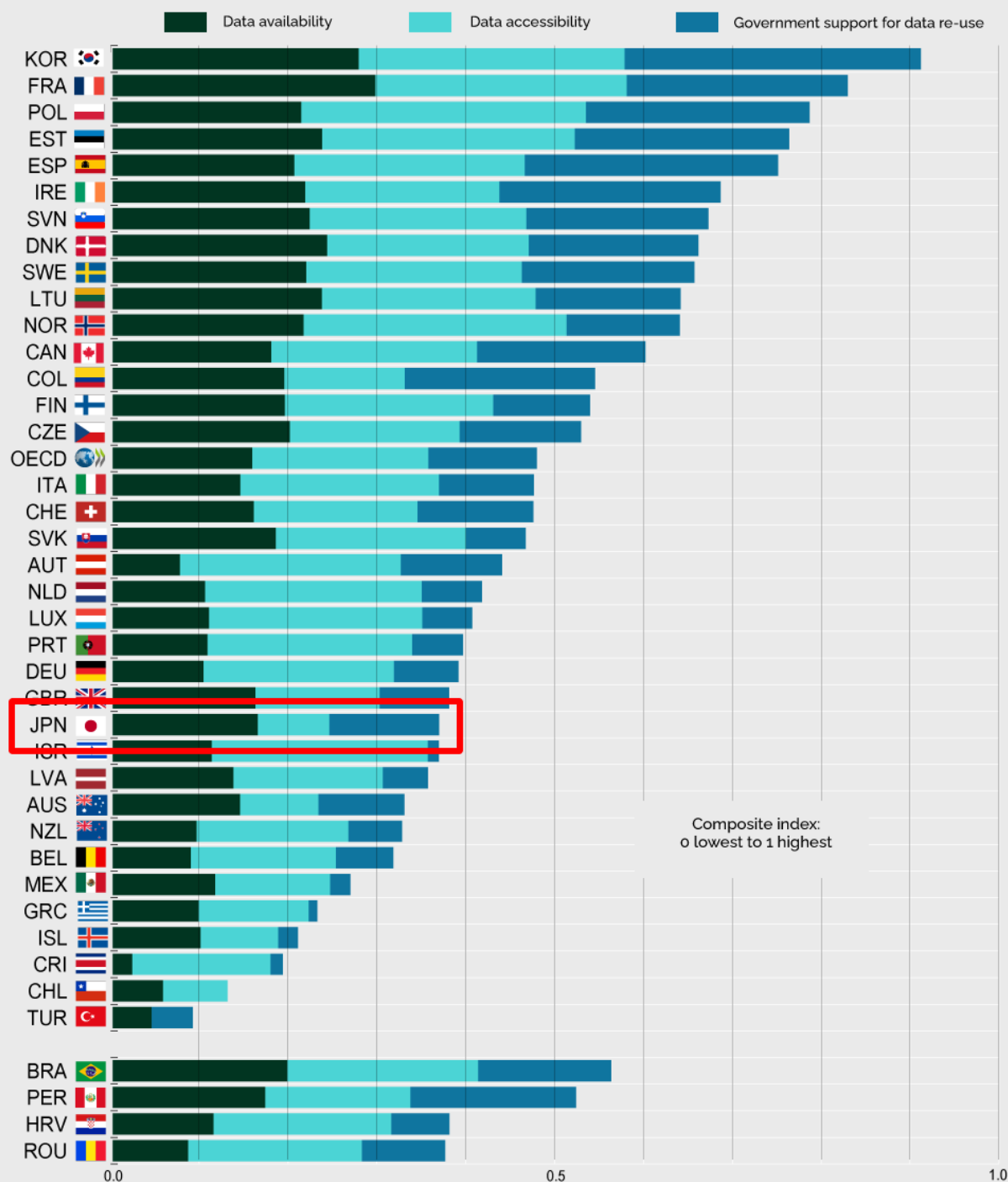
「OURdata Index」は、データの可用性、アクセシビリティ、および政府による利用のサポートに焦点を当てた評価となっている。

以下のグラフ<sup>[^3]</sup>によると、日本においては、特にアクセシビリティの評価が低いことがわかる。



# OECD OURdata Index (2023)

Open, Useful and Re-usable data (OURdata) Index, benchmarks efforts made by governments to design and implement national open government data policies



Source: 2023 OECD Open, Useful and Re-usable data (OURdata) Index



アクセシビリティの内容としては以下の3項目で構成されている。

- データへの自由でオープンなアクセスに関するポリシーの内容(Content of the free and open access to data policy)
  - 機械可読形式とオープン形式に関する要件
  - オープンデータを無償かつオープンライセンスで提供するための要件
  - APIを通じてデータを提供するための要件

- 政府のオープンガバメントデータポータルがある場合、それを通じてデータ提供する為の要件
- データの品質と完全性のためのステークホルダーの関与(Stakeholder engagement for data quality and completeness)
  - 中央/連邦オープンガバメント・データポータルのフィードバックメカニズム
  - ユーザー主導の中央/連邦オープンガバメント・データポータル
- 価値の高いデータセットへのアクセス性(accessibility of high value datasets)
  - 中央ポータルへのアクセス
  - メタデータの質
  - オープンなフォーマット
  - 最新版
  - APIアクセス

これらの評価の平均がアクセシビリティの評価[<sup>4</sup>]となるが、アクセシビリティのみで見ると日本の順位は39位であり、日本の総合順位に影響している主な要因の一つになっていると考えられる。

Table A A.2. Data accessibility

	Content of the free and open access to data policy	Stakeholder engagement for data quality and completeness	Implementation	Data accessibility
OECD	0.68	0.42	0.68	0.59
AUS	0.00	0.19	0.61	0.27
AUT	0.75	0.60	0.89	0.75
BEL	0.50	0.31	0.66	0.49
BRA	0.75	0.54	0.64	0.64
CAN	0.67	0.50	0.92	0.70
CHE	0.67	0.28	0.71	0.55
CHL	0.00	0.00	0.65	0.22
COL	0.00	0.35	0.87	0.41
CRI	1.00	0.00	0.40	0.47
CZE	0.92	0.00	0.81	0.57
DEU	1.00	0.23	0.70	0.64
DNK	1.00	0.13	0.92	0.68
ESP	1.00	0.58	0.75	0.78
EST	0.92	0.83	0.81	0.85
FIN	0.50	0.88	0.74	0.70
FRA	0.79	1.00	0.76	0.85
GBR	0.50	0.00	0.76	0.42
GRC	0.92	0.00	0.20	0.37
HRV	1.00	0.35	0.45	0.60
ICE	0.33	0.00	0.45	0.26
IRE	0.83	0.42	0.72	0.66
ISR	1.00	0.54	0.65	0.73
ITA	1.00	0.44	0.68	0.67
JPN	0.00	0.35	0.37	0.24
KOR	1.00	0.82	0.88	0.90
LTH	0.92	0.42	0.83	0.72
LUX	0.92	0.60	0.65	0.72
LVA	0.58	0.23	0.70	0.51
MEX	0.58	0.00	0.58	0.39
NLD	0.42	0.82	0.96	0.73
NOR	0.92	0.83	0.91	0.89
NZL	0.38	0.63	0.54	0.51
PER	0.83	0.00	0.64	0.49
POL	1.00	1.00	0.89	0.96
PRT	0.92	0.52	0.64	0.69
ROU	0.33	0.83	0.59	0.59
SVK	0.83	0.46	0.63	0.64
SVN	0.75	0.60	0.84	0.73
SWE	1.00	0.55	0.64	0.73
TUR	0.00	0.00	0.00	0.00

Note: Data is not available for Hungary and the United States. On data for Israel, see <https://doi.org/10.1787/888932315602>.

Source: OECD (2022), Survey on Open Government Data 5.0

DataOpsプロセスの活用により、アクセシビリティの評価改善となるか否かを確認する。

## 2 検証イベント実施内容

### オープンデータDataOps検証イベント

検証イベントとして以下のイベントを実施した。

- Social Hack Day #56 2023年12月16日開催
- Social Hack Day #57 2024年01月20日開催
- Social Hack Day #58 2024年02月17日開催

### 検証イベント詳細

Social Hack Day #56 (イベント詳細：[Social Hack Day #56 イベントレポート](#))

#### 概要

データ利用者がオープンデータを活用する方法として、データの可視化があげられる。  
本イベントではデータの可視化を行う際に有用なデータ構成について、データ提供者のプロセスである「データの収集蓄積」を通して議論し、このプロセスの有効性を確認する。

#### 成果

成果	関連するDataOpsプロセス							
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用	フィードバック
国の制度関連の取り組みを行っているプロジェクトでDataOpsの概念を組み込んでいきたいという意見が上がった								
自治体がオープンデータを提供するプロセスを体験することで、作業量等の観点から自治体職員が現状の業務に加えて、手作業でデータ加工を行うことは負荷の観点で困難であると感じた			○					
オープンデータ提供者にとって生成系AIによるレビューが有効な手段であることを実感できた			○					
フィードバックをトリガーとしてDataOpsプロセスの課題抽出・計画・データの公開・取得・活用の要素を満たす機能を有しているCKAN拡張機能（ckanext-feedback）をイベント参加者が実際に体験することでDataOpsプロセスにおける善循環サイクルへの理解を促進できた	○	○		○	○		○	○



成果	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
博物館では所蔵品のデータの提供を積極的に実施している為、カルチャーの領域でも活用してみたいという意見が上がった							
ChatGPTを使ってデータ構造についてレビューを実施したところ問題点と改善に関するアイデアを検討してくれた			○				
ChatGPTによる可視化は細かい要求に対しても適切にデータを生成してくれる為、データ利用者にとって活用できるツールであることを実感した							○

#### 課題

課題	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
オープンデータがPDFで公開されている為、データ利用者は扱いづらい			○	○		○	○
自動化して効率的にデータの整形を行いたい			○				
自治体側がファイルに変更を加えた際、その変更がデータ利用時に大きな障害になってしまう場合がある為（想定外の項目が追加されている等）、相互のコミュニケーションが必要だと感じた			○				○

課題	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
<p>以下の理由から初めにフォーマットを用意しておく必要性を実感した</p> <ul style="list-style-type: none"> <li>● 地域ごとにフォーマットが異なる</li> <li>● 半角全角、記号などの統一が行われていない</li> <li>● 自治体がオープンデータに加えた工夫（「市」の行を追加してグルーピングしている）で、かえって扱いにくいデータ形式になってしまっている</li> </ul>			○			○	○
<p>大元のデータが更新されると、リンクが無効になってしまう問題が発生したため、データ更新の際、何かアナウンスがあるべきだと感じた</p>				○	○		
<p>文化資源の領域のオープンデータも利活用したいという意見があったが、オープンデータの認知度や法整備などの課題がある</p>		○					
<p>ChatGPTが出してくれたデータ構造は細かく、確かに正しいが項目が大量になってしまうためデータ作成のコストがかかる</p>			○				
<p>CKANにオープンデータを公開するにあたりバージョンを意識する必要があるが、有識者のサポートが必要になる可能性が高い</p>				○			
<p>ChatGPTによる可視化はデータ利用者にとって活用できるツールではあるが、指示文のちょっとしたニュアンスの違いで結果が変わってしまうこともある為、必ず結果が確かなものかを確認する必要がある</p>							○

Social Hack Day #57 (イベント詳細 : [Social Hack Day #57 イベントレポート](#))

概要

[Social Hack Day #56](#)で議論したデータの収集蓄積プロセス及びフィードバックプロセスの有効性を確認した。

データの収集蓄積プロセスは、データ利用者がデータを有効に活用する為のデータ構造に関して継続議論した。

フィードバックプロセスは、データ提供者のモチベーションをあげる（利活用促進につなげる）ためにはどうすべきかについてデータ利用者・データ提供者それぞれの観点で議論した。

最後にデータの収集蓄積やデータの活用プロセスに使用することを想定して開発したAssistants API UIツールのデモを実施し、意見交換を行った。

成果

成果	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
データをどこまで構造化すべきかはユースケースによって変わる為、ニーズに合わせて変更可能な最小限の構造（名称、住所などの共通部分のみ）であるべきことがわかった			○				
ChatGPTを使用したデータ構造のレビューやデータの可視化の体験を通して、データの構造化や検索用としてChatGPTを活用することの将来性を実感できた			○	○	○		
オープンデータのニーズの収集は東京都職員の方がハッカソンや企業へのヒアリングを通して行っていることがわかった	○						○
オープンデータが循環している流れを可視化することで、さらにサイクルの加速が期待できるという意見があがった							○
オープンデータ作成支援AIの課題（以下課題表中に記載）を把握することができた			○				○

## 課題

課題	関連するDataOpsプロセス							
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用	フィードバック
オープンデータの利用者と提供者一人一人のモチベーションをあげる仕組みが必要								○
データを構造化しすぎることによってデータ提供者の作業コストが増加してしまう			○					
データ提供者とデータ利用者の間を仲介する（データを構造化する）ステークホルダーが善循環のサイクルに存在しない			○					
オープンデータを整備する部署と公開する部署が異なる場合、フィードバックコメントを読まないことがある為、データ利用者の声がデータを整備する部署の担当者まで届かない								○
利用者が提供者に対して感謝の意識をもてるような仕組みが必要							○	○
フィードバックには報告、要望などの種類があるが、ネガティブな要素を可能な限り減らす為には「データのこの部分を修正しました」のような報告ベースのフィードバックであるべき								○
提供者はフィードバックとして返ってきたデータを修正する際に、責任やチェック作業にかかるコストに抵抗感を感じる恐れがある			○					○
DataOpsのスタートは利用者のニーズ収集から始まるため、それをサポートする仕組みが必要	○							○
ChatGPTを活用する上で、正しく解釈できるデータ構造とは何かについて検討する必要がある			○				○	
欲しいデータを投票できる仕組みなど、楽にニーズを表明できる環境が必要	○							
オープンデータ作成支援AIに検索させる場合、検索のたびにすべてのデータを参照するのは非効率である為、必要な情報に関してはある程度の構造化が必要			○					

課題	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
オープンデータ作成支援AIを使用する際、プロンプトのテクニックや処理できるデータ件数の限界を把握する必要がある			○				○

概要

[Social Hack Day #57](#)で議論した、データ提供者とデータ利用者の間を仲介する二次データ提供者の作業プロセスにフォーカスを当て、そのプロセスの一部を実際に実施した。

二次データ提供者の作業プロセスは既存のDataOpsプロセスには存在しない為、以下を想定。

- 1. 一次データ提供者(データ提供者：自治体職員など)が公開したデータを取得する（二次データ提供者によるデータ取得プロセス）
- 2. 取得したデータを機械判読可能なデータへ加工する（二次データ提供者によるデータ処理プロセス）
- 3. データ利用者が利用できる形で提供・公開を行う（二次データ提供者によるデータ公開プロセス）
- 4. 一連のプロセスをCIなどで自動化する（二次データ提供者による自動化プロセス）

今回はこのうち、1、2、4の作業を実施した。

実際の作業内容としては以下の通り。

- 厚労省が公開している「緊急避妊薬医療機関」のPDFデータ一括取得（dimによるデータ更新管理含む）
- 取得したPDFデータのCSV変換（フォーマットが異なるPDFについても考慮）
- PDF取得からCSV変換までの GitHub Actions による自動化

成果

成果	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
tabulaとpandasを用いたPDF→CSV変換で、改行コードを置換することで変換することに成功した						○	
dimの機能を活用して、PDFを取得することに成功した					○		
CIによって定期的（1日1回）にPDFを一括ダウンロードし、CSVに自動変換するツールが完成した					○	○	

## 課題

課題	関連するDataOpsプロセス						
	課題抽出	計画	データの収集蓄積	データの公開	データの取得	データの処理	データの活用
オープンデータとして公開されているPDFが定期的に更新され、リンク切れが発生してしまう → dimによるデータ更新管理で対応					○		
リンク切れの判別は、PDFをダウンロードして確認するか、サイトの最終更新で確認するしかない → dimによるデータ更新管理で対応					○		
dimによるPDF取得時に2ファイルほど関係のないファイルが混ざっている					○		
PDFを一括ダウンロードして、csvのような機械可読性の高いものに自動変換したい → イベント内で作成完了					○	○	
Robocorpを使用したPDF→CSV変換ではPDF内のテーブルが認識されずテキストとして出力されてしまう						○	
tabulaとpandasを用いたPDF→CSV変換を行う際、以下の課題を確認 ● 変換後、出力されたCSVがPDFのページ毎にわかれてしまう→ イベント内で対応済 ● PDFのヘッダーが毎回記述されてわかりにくい						○	
各都道府県によってフォーマットの差異がある（1ページ目だけにヘッダーが記載されている、独自の列があるなど）						○	
データが公開されているページにPDFの追加や順番の入れ替わり等が発生するとPDFの更新取得処理に失敗する可能性がある → CIが落ちることで検知可能					○		
今回のPDF→CSV変換処理は、幅広い分野のオープンデータに対しても検証を進めた方がいい						○	

## イベントにて実施したアンケート結果

DataOpsのプロセスがオープンデータ利活用促進につながるかの根拠を得る為、イベント参加者にアンケートを実施した。

- [アンケート結果\(12月開催分\)](#)
- [アンケート結果\(2月開催分\)](#)

## 3 考察

2022年度報告書で提案したDataOpsプロセスのオープンデータ利活用への有効性を確認する為、その最初のアプローチとして検証イベントを開催した。

イベントを開催した結果、背景に記載した以下2点の問題を確認することができた。

- 必要なデータが公開されていない
- 利用者のニーズが解らない

検証イベントの中で行われたディスカッションやアンケートの結果から、新たに以下の課題が見つかった。

1. データ利用者が利活用しづらい形式のデータが存在する
2. データ利用者が利活用しやすい(※1)データへ加工する作業が必要となる
3. データ提供者のモチベーション維持が難しい

※1 利活用しやすいデータとは、機械判読(※2)可能なデータを指す。

公開されているPDFは、データを利用する際に機械判読可能なデータへの変換（PDFからCSVなど）が必要となる可能性が高い。実施したイベントにおいても、PDFからCSVへ変換を実施した。

※2 機械判読とはコンピュータプログラムが自動的にデータを加工、編集等ができること

この課題を改善させる方法としては以下が考えられる。

1. データ利用者が利活用しやすいデータを増やす
2. 二次データ提供者という役割とプロセスを増やす
3. 生成AIを活用する
4. データ利用者からのフィードバックを増やす

これらの改善方法の詳細について、以下に記載する。

### 改善方法詳細

#### データ利用者が利活用しやすいデータを増やす

2024年3月5日時点で公開されているオープンデータ<sup>[^5]</sup>191,534件のうち、43,220件の約23%がPDF形式で公開されている。

自治体によって差異があるが、まだデータ利用者にとって利活用しづらい形式で公開している場所が多く存在する。

また、構造化されていても不要な改行やスペースなどが入っていることで、データ利用者にとって利活用しづらいデータとなっている。

検証イベントで利用したデータに関しても、不要な改行やスペースなどがあり、利用する際のデータ加工作業に時間がかかることがわかった。



対策としては、デジタル庁が公開しているGIF(政府相互運用性フレームワーク)[<sup>6</sup>]や自治体標準オープンデータセット[<sup>7</sup>]の形式に統一して、利用者が扱いやすいデータで公開することが考えられる。

ただし、この形式に変換することは、ITやオープンデータに関する知見が必要であることが想定される為、データ提供者側でこの対応をすることは大きな負担になると考えられる。

その為、この変換プロセスについては別途改善が必要だと考えられる。

## 二次データ提供者という役割とプロセスを増やす

前述の通り、扱いやすいデータを公開するには、デジタル庁が公開しているGIFや自治体標準オープンデータセットに準拠する必要があると考えられる。

しかし、以下の理由から、データ提供者側でこの対応を実施することは大きな負担になることが想定される。

- ・ ITやオープンデータに精通した人の知見が必要
- ・ 作業コストがかかる

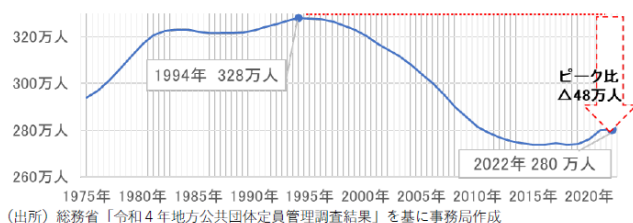
なお、以下内閣官房が公開している資料[<sup>8</sup>]では、DX・情報関係業務の担当職員数は、3名以下の自治体が54.6%(950団体/1,741団体)、1名以下の自治体が16.9%(295団体/1,741団体)と報告されている。過半数が3名以下であるため、自治体職員の負担は大きいと考えられる。

### 公共サービスの持続可能性への懸念 ①地方公共団体の職員数

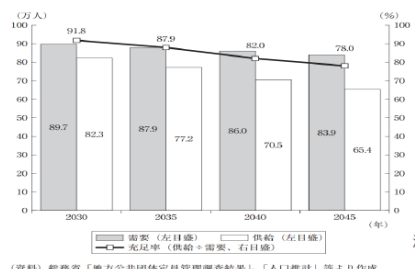
○ 地方公共団体の職員数は、業務の効率化等によりピーク時から減少

○ デジタル化による更なる業務効率化も、3名以下の自治体が55%(1名以下の自治体は17%)であり、課題

#### 《地方公共団体（都道府県・市町村）職員総数の推移》

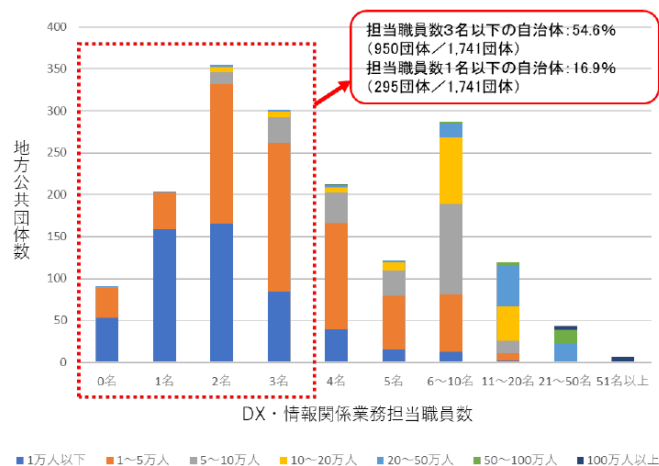


#### 《地方公務員不足の将来推計（市町村・普通会計）》



(出所) 株式会社日本総合研究所「地方公務員は足りているかー地方自治体の人手不足の現状把握と課題ー(2021年)」

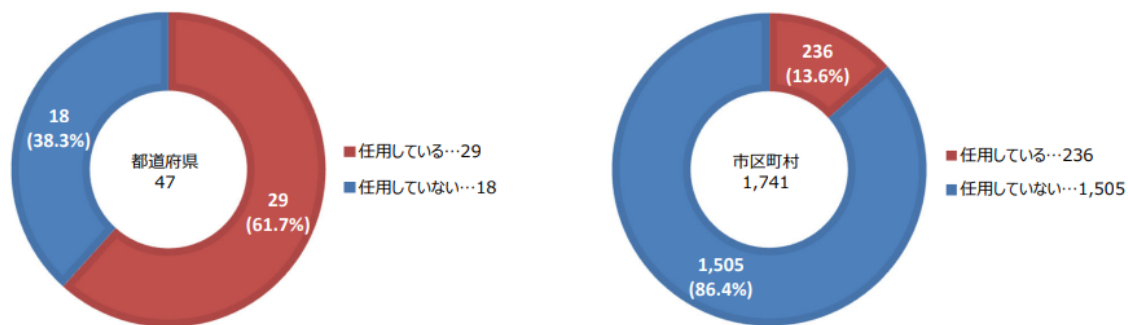
#### 《（人口規模別）DX・情報関係業務担当職員数》



以下総務省のレポート[<sup>9</sup>]からも、自治体職員以外の人材に頼らざるを得ない状況と言える。

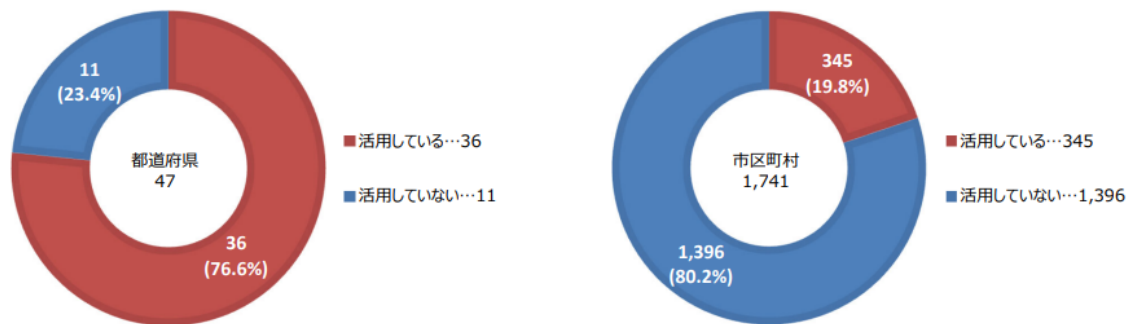
7 DXを推進するための外部デジタル人材の任用

都道府県では29団体（61.7%）、市区町村では236団体（13.6%）が外部デジタル人材を任用している。



8 DXを推進するための外部デジタル人材の活用

都道府県では36団体（76.6%）、市区町村では345団体（19.8%）が外部デジタル人材を活用している。



対策としては、データ提供者とデータ利用者の間を仲介する（機械判読可能なデータに加工する）ステークホルダを設け、二次データ提供者としてデータの取得、データの処理をするプロセスをDataOpsプロセスに追加することが考えられる。

この対策を実施することで、データ提供者に負担をかけずに、扱いやすいデータをデータ利用者が使えるようになると考えられる。

また、データ提供者の負担を軽減することで、データ公開に対する抵抗感をやわらげることにつながり、データ公開及びデータ活用の促進につながると考えられる。

ただし、この改善策を実施する前提として、二次データ提供者を担当する人物が必須となる為、誰がその役割を担うかが課題となる。

二次データ提供者について

二次データ提供者について、その位置づけを以下に記載する。

二次データ提供者の対象プロセス

二次データ提供者は以下プロセスの実施を想定している。

- 一次データ提供者(データ提供者：自治体職員など)が公開したデータを取得する（DataOpsプロセスにおけるデータの取得）
- 取得したデータを機械判読可能なデータへ加工する（DataOpsプロセスにおけるデータの処理）  
加工するフォーマット例：GIF や 自治体標準オープンデータセット

- ・ データ利用者が利用できる形で提供・公開を行う（DataOpsプロセスにおけるデータの公開）
- ・ 一連のプロセスをCIなどで自動化する

検証イベントでは、厚労省が公開している「緊急避妊薬医療機関」のPDFを一括ダウンロードし、CSVデータへ変換するプロセスを自動化することに成功した。

しかし、以下のような課題がまだ残っている為、今後もさらに検討を進める必要があると考えられる。

- ・ 変換したCSVに不要な改行や余計なヘッダー情報が含まれている
- ・ 他のオープンデータに関しても「[Social Hack Day #58](#)」の課題に記載したような問題が発生する可能性がある為、幅広いオープンデータに対して検証を進める必要がある

#### 二次データ提供者となりうるステークホルダ

GIFや自治体標準オープンデータセットに準拠した機械判読可能なデータへの加工、及びプロセスの自動化は、ITやオープンデータに関する知見が必要であると考えられる。

その為、二次データ提供者となりうるステークホルダとしては、どちらにも精通している可能性のあるシビックハッカーが候補として考えられる。

また、将来的にはエコシステムとしての新たなステークホルダが必要になる可能性も考えられる。

今回検証イベントで実施した際は、イベント参加者自身がこの役割を担当した。

#### 二次データ提供者による効果

二次データ提供者が、機械判読可能なデータへ加工する作業コストを担うことで、データ提供者に負担をかけずに扱いやすいデータをデータ利用者が使えるようになると考えられる。

また、データ提供者は作業コストの低減に加え、扱いやすいデータを意識することなくデータを公開することが可能になると考えられる為、データ公開に対する抵抗感をやわらげることにつながると考えられる。

その結果、データ活用及びデータ公開の促進につながると考えられる。

### 生成AIを活用する

検証イベントを通じて、生成AIのChatGPTが、データ加工やデータ活用に利用できる可能性があることがわかった。

作成したデータのレビューなどをChatGPTで実施することで、機械判読可能なデータにする為の提案を作業コストを抑えて実施することが可能になると考えられる。

また、データ可視化などのデータ活用の場においても、少ない労力で位置情報をマッピングしたHTMLデータを生成することが可能になると考えられる。

ただし、以下の課題に関しては引き続き検討する必要がある。

- ・ AIが適切な応答や結果を生成するためのプロンプトのテクニックが必要
- ・ ハルミネーション※3を起こしてしまうことがある為、レスポンス内容が確かなものであるかの確認が必要
- ・ 細かい指摘や修正を命令するためには結局ITの知識が必要となる

※3 ハルミネーションとは人工知能（AI）が事実に基づかない情報を生成する現象

## データ利用者からのフィードバックを増やす

データを有効に活用していく為には利用者のニーズに応じたメンテナンスを継続的に実施する必要がある。ニーズの把握や、データ提供者のモチベーションを維持する目的としても「フィードバック」プロセスは重要な役割を持つと考えられる。

フィードバックの仕組みを実現する方法の一つとして、OSSとして公開されているCKAN拡張機能（ckanext-feedback）<sup>[10]</sup>を活用する方法が挙げられる。

CKAN拡張機能はデータセットに対する評価や「要望、質問、宣伝、感謝」のフィードバックコメントを記載できるようになっている。

データ提供者にとってモチベーションが向上するフィードバックを実施することが、データ提供におけるモチベーション維持につながると考える。

モチベーションを向上させるフィードバックとしては、ポジティブなフィードバックやネガティブな要素を排除したフィードバックが考えられる。

例えば、ポジティブなフィードバックとしては、評価や具体的な活用事例の報告、「素晴らしいデータをありがとう」などの感謝のメッセージなどが考えられる。そして、ネガティブな要素を排除したフィードバックとしては、オープンデータに修正が必要なケースの場合、「データの〇〇が間違っています」のような Issue形式ではなく、「データの〇〇部分を修正しました」のような Pull Request 形式でのフィードバックなどが考えられる。

データ提供者のモチベーションを維持する為には、このようなフィードバックを促す仕組みの検討が今後必要であると考えられる。

## OECD Our Data Index の Data Accessibility の改善

前述の「[二次データ提供者による効果](#)」の通り、二次データ提供者を設けることで、機械可読可能な形式でのデータ公開が可能になると考えられる。

「データへの自由でオープンなアクセスに関するポリシーの内容」の観点で、この二次データ提供者プロセスの追加が Data Accessibility の改善につながると考えられる。

この改善策は Data Accessibility の一部分であり、以下の観点を考慮する必要がある。

- データへの自由でオープンなアクセスに関するポリシーの内容(Content of the free and open access to data policy)
- データの品質と完全性のためのステークホルダーの関与(Stakeholder engagement for data quality and completeness)
- 価値の高いデータセットへのアクセス性(accessibility of high value datasets)

## 4 まとめ

検証イベントでのDataOpsプロセスのオープンデータ利活用への有効性を確認する過程で、新たに以下の課題が見つかった。

1. データ利用者が利活用しづらい形式のデータが存在する
2. データ利用者が利活用しやすいデータへ加工する作業が必要となる
3. データ提供者のモチベーション維持が難しい

これらの課題を改善させる為の方法として、以下を提案した。

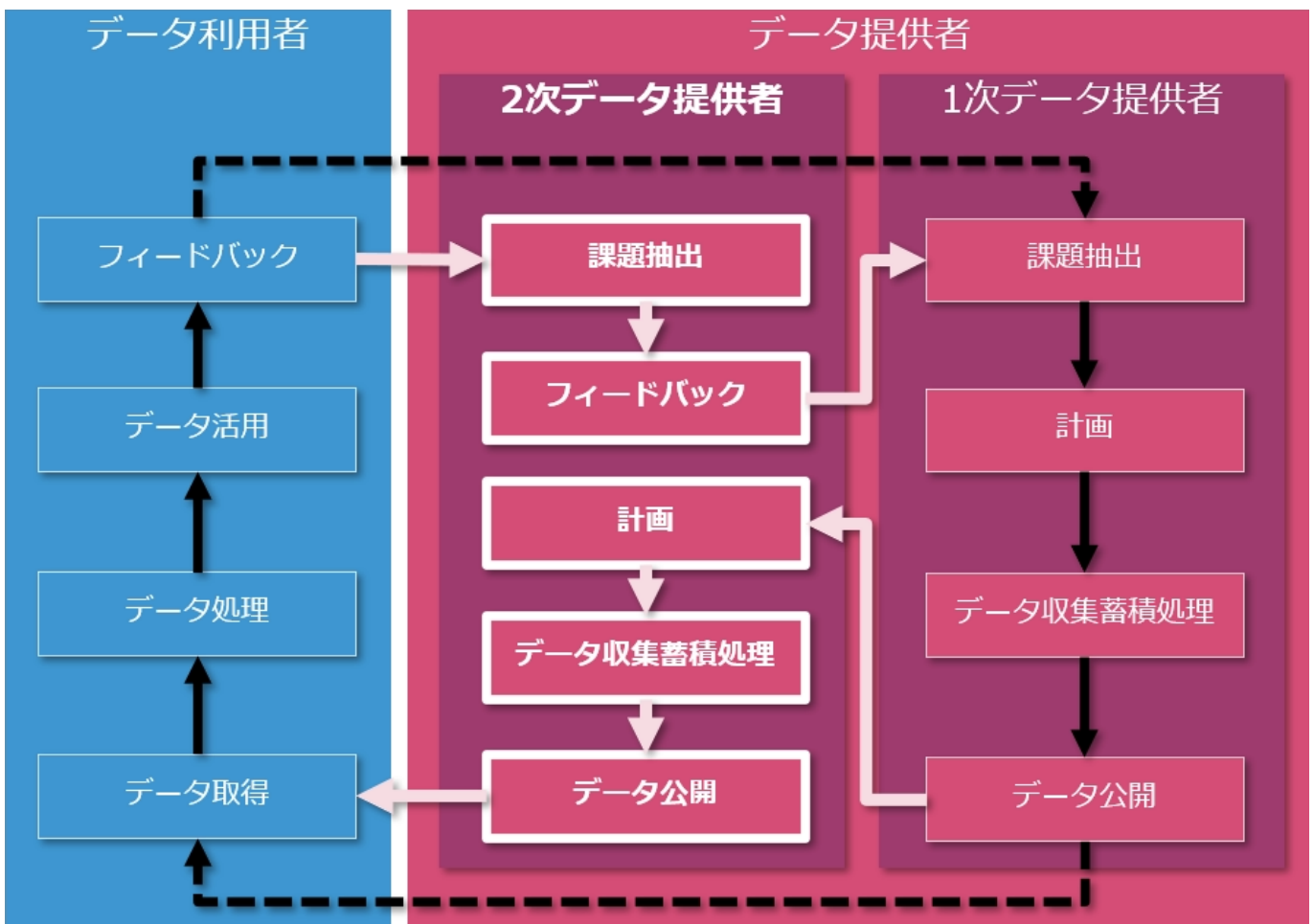
1. データ利用者が利活用しやすいデータを増やす
2. 二次データ提供者という役割とプロセスを増やす
3. 生成AIを活用する
4. データ利用者からのフィードバックを増やす

この改善方法により、DataOpsプロセスは以下のような善循環のサイクルをまわすことが可能になると考えられる。

結果、必要なデータが公開されるようになり、利用者のニーズも把握することが可能になると考えられる。

1. データ提供者は扱いやすいデータを意識することなくデータを公開する
2. 二次データ提供者は GIF or 自治体標準オープンデータセット に準拠したデータを作成する
3. データが扱いやすい為、データ利用が増える
4. データ利用が増えることでフィードバックも活性化する
5. データ利用者はデータ提供者、二次データ提供者に対してポジティブ、及びネガティブ要素を含まないフィードバックを実施する
6. データ提供者、二次データ提供者はモチベーションを維持したままフィードバックを取込、ニーズのあるデータを提供し続ける

二次データ提供者を追加したDataOpsプロセス案としては、以下のようなイメージである。



この善循環のサイクルを実現する為には、二次データ提供者を担当する人物が必要となる点が今後の課題になると考えられる。

OECD Our Data Index の Data Accessibility についても、二次データ提供者を設けることで、機械可読可能な形式でのデータ公開が可能になると考えられる為、「データへの自由でオープンなアクセスに関するポリシー

ーの内容」の観点で Data Accessibility の評価向上につながると考えられる。

## 5 今後の展望

OECD Our Data Index における日本の総合順位を上げる為、要因の一つである Data Accessibility の評価向上を目指す。

その為に、まずは本報告書で述べた、二次データ提供者プロセスにおける残課題を解決し、善循環サイクルの実現を目指す。

ただし、二次データ提供者による Data Accessibility の改善は、以下の Data Accessibility 評価対象項目の内、「データへの自由でオープンなアクセスに関するポリシーの内容」の一部分である。

Data Accessibility の評価向上の為に、二次データ提供者プロセスを追加したDataOpsプロセスをベースとして、評価対象項目の要件考慮を前提に、課題の洗い出しと改善を今後も進めていきたい。

最終的には、このDataOpsプロセスの概念を普及させることで、日本のオープンデータの改善及び OECD Our Data Index での日本の総合順位回復に貢献できれば幸いである。

### ※Data Accessibility の評価対象項目

- データへの自由でオープンなアクセスに関するポリシーの内容(Content of the free and open access to data policy)
  - 機械可読形式とオープン形式に関する要件
  - オープンデータを無償かつオープンライセンスで提供するための要件
  - APIを通じてデータを提供するための要件
  - 政府のオープンガバメントデータポータルがある場合、それを通じてデータ提供する為の要件
- データの品質と完全性のためのステークホルダーの関与(Stakeholder engagement for data quality and completeness)
  - 中央/連邦オープンガバメント・データポータルのフィードバックメカニズム
  - ユーザー主導の中央/連邦オープンガバメント・データポータル
- 価値の高いデータセットへのアクセス性(accessibility of high value datasets)
  - 中央ポータルへのアクセス
  - メタデータの質
  - オープンなフォーマット
  - 最新版
  - APIアクセス

## 参考文献

- [^1]: クリエイティブ・コモンズ・ライセンス, "クリエイティブ・コモンズ・ライセンス," [オンライン]. Available: <https://creativecommons.org/licenses/by/4.0/>. [アクセス日: 16 09 2022].
- [^2]: Gartner, Inc. and/or its affiliates., "Press Releases," 11 9 2018. [オンライン]. Available: <https://www.gartner.com/en/newsroom/press-releases/2018-09-11-gartner-hype-cycle-for-data-management-positions-three-technologies-in-the-innovation-trigger-phase-in-2018>. [アクセス日: 29 8 2022].
- [^3]: SIP NIIコンソーシアム, "データカタログ横断検索システム" [オンライン]. Available: <https://search.ckan.jp/>. [アクセス日: 05 03 2024].
- [^4]: Digital Agency, Government of Japan, "政府相互運用性フレームワーク ( GIF )" [オンライン]. Available: [https://www.digital.go.jp/policies/data\\_strategy\\_government\\_interoperability\\_framework](https://www.digital.go.jp/policies/data_strategy_government_interoperability_framework). [アクセス日: 20 02 2024].
- [^5]: Digital Agency, Government of Japan, "自治体標準オープンデータセット ( 正式版 )" [オンライン].

Available: [https://www.digital.go.jp/resources/open\\_data/municipal-standard-data-set-test](https://www.digital.go.jp/resources/open_data/municipal-standard-data-set-test). [アクセス日: 20 02 2024].

[^6]: Digital Agency, Government of Japan, "政府相互運用性フレームワーク ( GIF )" [オンライン]. Available: [https://www.digital.go.jp/policies/data\\_strategy\\_government\\_interoperability\\_framework](https://www.digital.go.jp/policies/data_strategy_government_interoperability_framework). [アクセス日: 20 02 2024].

[^7]: Digital Agency, Government of Japan, "自治体標準オープンデータセット ( 正式版 )" [オンライン]. Available: [https://www.digital.go.jp/resources/open\\_data/municipal-standard-data-set-test](https://www.digital.go.jp/resources/open_data/municipal-standard-data-set-test). [アクセス日: 20 02 2024].

[^8]: 内閣官房デジタル行財政改革会議事務局, "「国・地方を通じたDXの推進」に係る討議資料," 2 2024. [オンライン]. Available: [https://www.cas.go.jp/jp/seisaku/digital\\_gyozaiikaikaku/taiwa6/240227taiwa1.pdf](https://www.cas.go.jp/jp/seisaku/digital_gyozaiikaikaku/taiwa6/240227taiwa1.pdf). [アクセス日: 29 03 2024].

[^9]: 総務省自治行政局地域力創造グループ地域情報化企画室, "自治体DX・情報化推進概要," 4 2023. [オンライン]. Available: [https://www.soumu.go.jp/main\\_content/000878726.pdf](https://www.soumu.go.jp/main_content/000878726.pdf). [アクセス日: 29 03 2024].

[^10]: C3Lab, "ckanext-feedback" [オンライン]. Available: <https://github.com/c-3lab/ckanext-feedback>. [アクセス日: 26 02 2024].