デジタル庁

資料９－３

Digital Agency Dashboard Series

デジタル庁ダッシュボードシリーズ
統計データの活用と課題について

何を公開しているのか？

2022年から開始。他省庁と協力し、約40のダッシュボードを公開予定。

	2022	2023	2024	2025	2026（予定）
デジタル庁	• マイナンバーカードの普及	• マイナンバーカードの利活用 • アナログ規制見直し • 26手続きのオンライン化 • 意識調査	• マイナポータルの利用状況 • GビズIDの利用状況	• 高等学校入学者選抜事務手続のデジタル化	• マイナンバー普及利活用 • 公金受取口座の登録状況
総務省			• 自治体DXの取組 • 自治体フロントヤード改革	• 地方自治体システムの共同調達 • 地方自治体システム標準化	• 地方財政（市区町村/都道府県） • 共同事務処理
厚生労働省			• 電子処方箋の導入状況 • 介護現場の生産性向上	• リフィル処方箋の認知率や利用	• 医療DX • マイナ保険証 • PMH（医療費助成） • 電子カルテ
文部科学省			• 校務DXの取組	• 教員の働き方改革	
国土交通省				• 水道事業等の経営状況	
こども家庭庁				• 保育提供体制	• 子ども家庭相談
内閣府				• 経済・財政・人口と暮らし	• GDP統計・経済循環図 • 国土強靭化

データに基づく政策運営のための「2種類のデータダッシュボード」を推進

日本全体の
社会的指標

Japan Dashboard
2025-

客観データを意識した政策運営

経済・財政・くらし等

- 医療費
- 教育環境、教育費
- 地方財政
- 人口関連
- 等....



個別の政策の目標

個別の政策の
進捗と現状

政策ダッシュボード
2023-

データを武器にした政策推進

デジタル政策等 個別の政策の進捗

- 医療、介護のDX
- 学校（教員の負担軽減）
- インフラ
- 自治体のDX

など



なぜ取り組んでいるのか？

すべての政策関係者が、手間暇かけず、政策の進捗をモニタリング

従来、関係者がそれぞれデータを探し、異なる時点・切り口でデータを確認。自団体の状況を見るだけもデータ加工に手間暇がかかっていた。

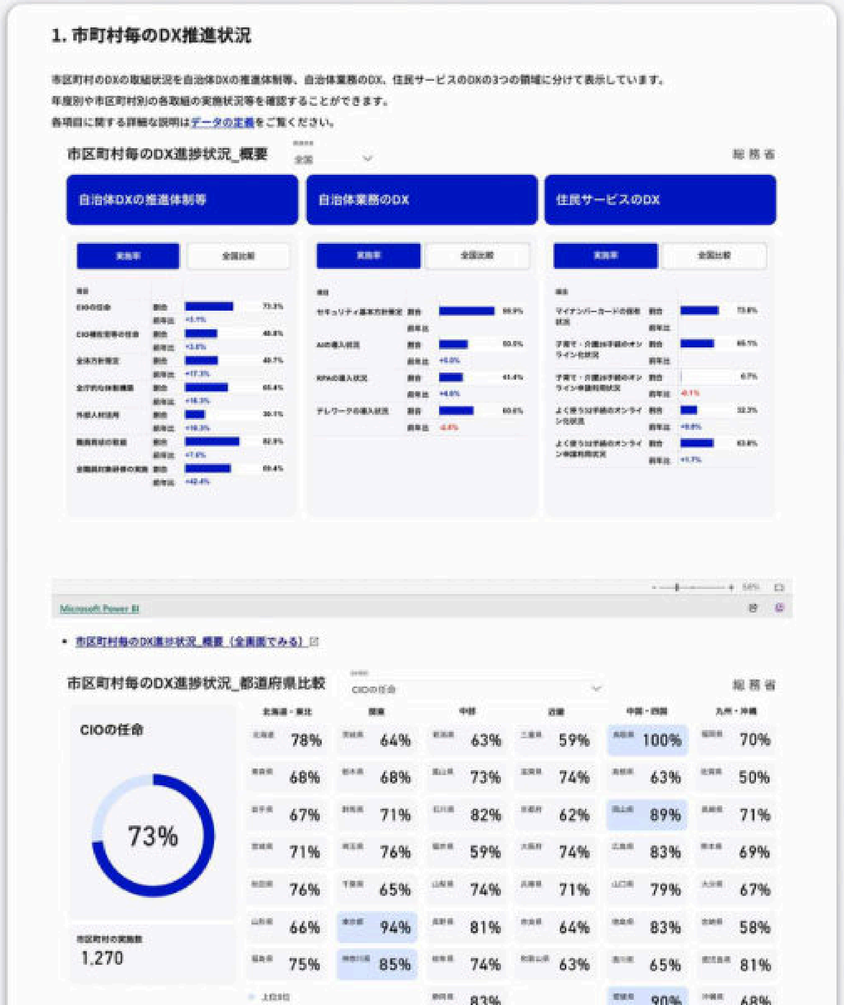
多様な形式のデータソース

（公的統計・業務データ等）



データ加工

e.g.自治体DXに関するダッシュボード



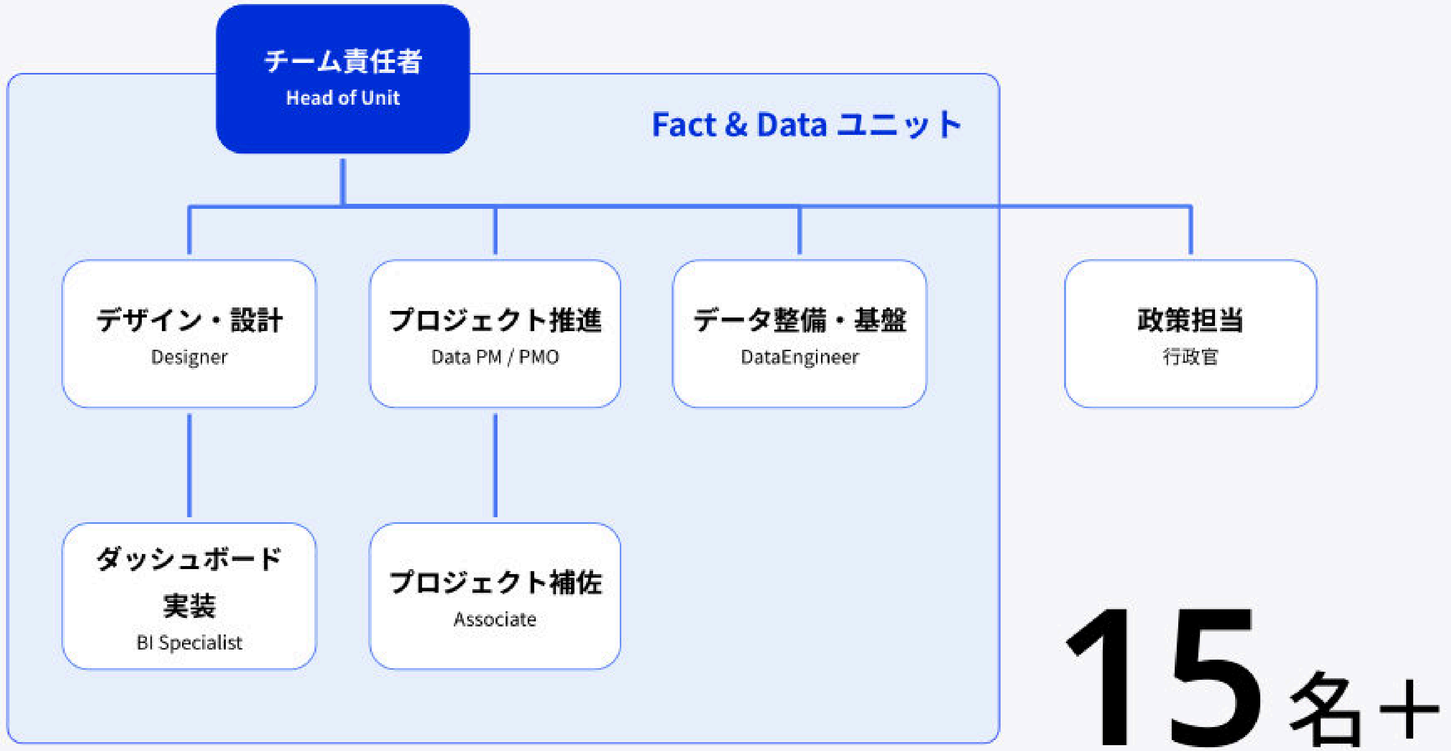
多様な立場の政策関係者

利用

- **中央省庁**
- **地方自治体**
  47の広域自治体・1742の基礎自治体
- **サービス提供主体**
  数十万の医療機関・福祉施設・教育機関
- **受益者**

基本的に全てのプロジェクトは内製で完結できる体制

- 他省庁からの相談を元に、要件整理～データ整備～デザイン～ダッシュボード構築といった支援を一元的に提供中。



ダッシュボードを支えるデータ分析基盤: sukuna

Google Cloudの各種プロダクト・サービスを活用。フルクラウドのデータ分析基盤で、ガバメントクラウド上に構築

多様なデータソース

省庁業務で利用、公開される
Excelファイル等

政府情報システムの
APIやデータベース

e-Stat（政府統計の総合窓口）等の
オープンデータ

データ分析基盤「sukuna」

Google Cloud Platform

Integration
- Cloud Functions
- Cloud Storage
- Storage Transfer Service
- Pub/Sub
- Cloud Scheduler

Dataplatform
- Cloud Run
  - dbt
- Cloud Scheduler
- BigQuery
  - Datalake
  - DWH
  - Datamart

BI-Publishment
- Cloud Run
- BigQuery

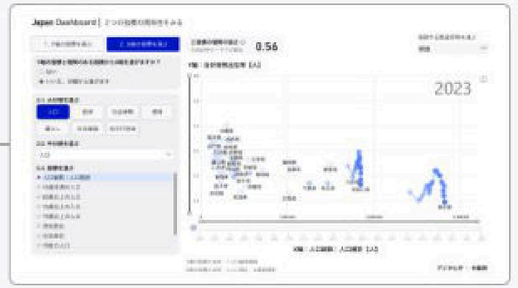
Management
- Security Command Center
- Cloud Logging
- Cloud Armor
- Cloud Build
- Cloud Monitoring
- Cloud Run

CI/CD
- Github
- Github Actions
- Terraform

各種ダッシュボード







ダッシュボード例①：水道事業等の経営状況に関するダッシュボード





作成者

・デジタル庁、総務省、国土交通省

利用データ

・地方公営企業決算状況調査（総務省・e-Stat）
・経営比較分析 類似団体一覧（総務省HP）
・上下水道施設の耐震化状況に関する緊急点検結果（国土交通省HP）

ユースケース

・全国の自治体で水道事業を担当している職員向け。
・近隣の団体や、全国の同規模団体と比較をし、経営改善を検討するための基礎データとして利用する。

ダッシュボード例①：水道事業等の経営状況に関するダッシュボード

画面デモ

ダッシュボード例①：水道事業等の経営状況に関するダッシュボード

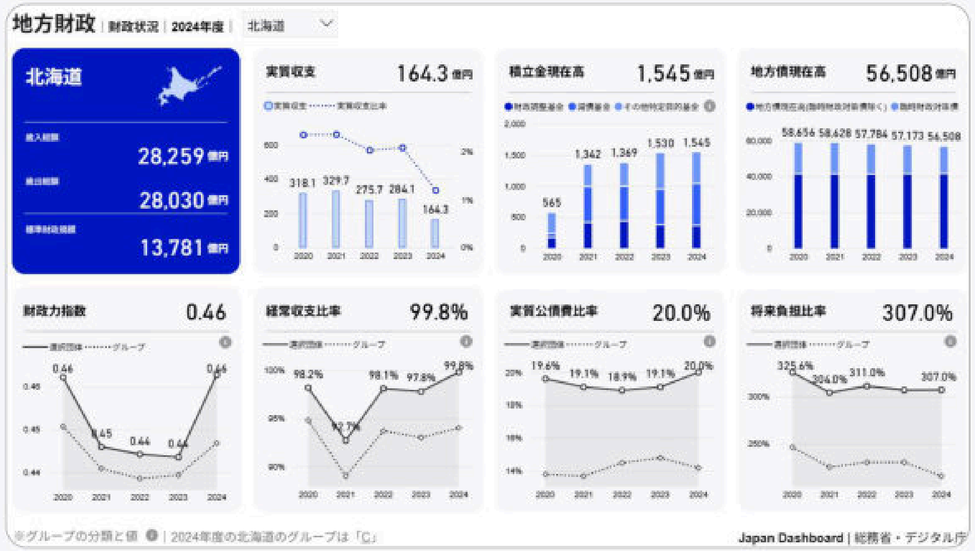
■地方自治体の水道事業の職員の声

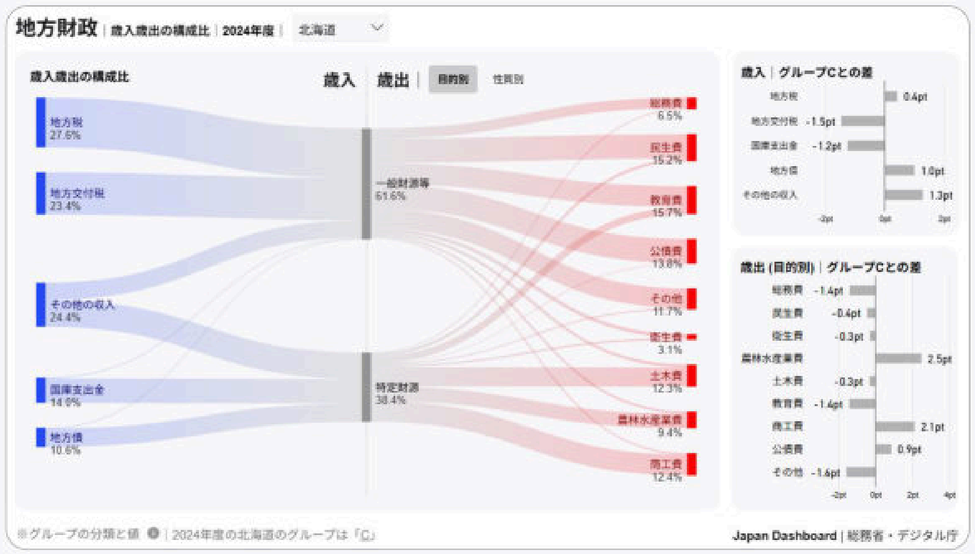
高度成長期に整備した水道管路や浄水場の老朽化に伴い、将来に向けた多額の更新投資が必要となる一方で、人口減少や節水機器の普及により**料金収入は減少傾向**にあります。

事業の安定経営に向けて的確な現状把握が不可欠ですが、従来は予算申請や議会対応のたびにe-Stat等からデータを取得し、**手作業で複雑なクレンジングや集計を行っていたため、多大な時間と労力がかかって**いました。

このダッシュボードは、**画面上のクリック操作だけ**で他都市との比較や経年変化、施設の老朽化状況が直感的に可視化されます。**データ収集・加工にかかる手間が大幅に削減**されたことで、そこに時間を割くのではなく、**本来注力したかった『分析や施策検討』に専念**できるようになり、客観的証拠に基づく議論を役所内ですぐにスタートできています。

ダッシュボード例②：地方財政関するダッシュボード





作成者

・デジタル庁、総務省

利用データ

・地方財政状況調査（総務省・e-Stat）
・財政指標一覧/健全化判断比率/類似団体区分（総務省HP）
・財政状況資料集（各都道府県HP）
・住民基本台帳に基づく人口（総務省・e-Stat）

ユースケース

・都道府県、市区町村で財政部門や企画・政策立案、事業評価を担当している職員向け。

・財政力指数が近いグループや近隣団体と比較をし、歳入・歳出構造の分析や行財政改革・予算編成を検討するための基礎データとして利用する

ダッシュボード例②：地方財政関するダッシュボード

画面デモ

ダッシュボード例②：地方財政関するダッシュボード

都道府県職員の声

- 類似団体・任意団体の比較機能について、率直な感想を申し上げると「すごい」と感じました。

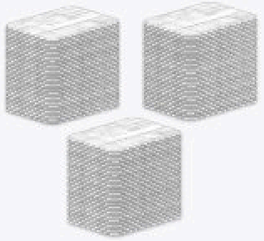
- 本県でも試行的に予算のダッシュボードを作り始めたところでしたが、情報量が多くなると動作が重くなるという課題がありました。一方、デジタル庁のダッシュボードはデータ量が多いにもかかわらず、簡単に比較ができます。**これまでは情報を分析するためにCSVで組み替えるといった操作が必要でしたが、ワンクリックで見比べられる点が非常に優れていると思いました。**

- 活用場面としては、決算の説明などが挙げられます。議会では、基金残高や地方債残高がどうなっているか、必要な財源がきちんと確保できているのか、といった点を問われることが多くあります。**「他県と比較すると多少減っているものの、本県は微増しており、しっかりと予算を確保している」「他県と比較して水準がこの程度であり、財政の健全性を保っている」**といった説明を行う機会が多いのが実情です。こうした説明を**視覚的に行えることが、最も助かる点**だと感じています。

統計データの活用によりEBPMの推進に繋がるが…

データを活用できる基盤を用意する事で国・自治体の政策立案が効率化し精度向上につながっている。
ただ、既存で公開されている統計データの加工には課題が多い状況。この後はデータ整備の課題についてご説明したい。

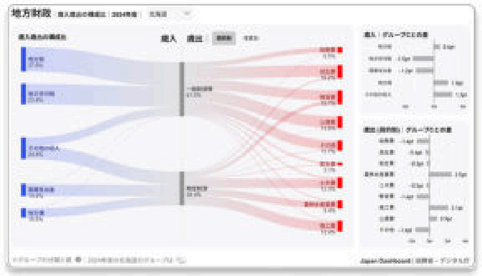
多様な形式の
データソース
（公的統計・業務データ等）



データ加工

デジタル庁で整備したダッシュボード









利用

多様な立場の
政策関係者

- **中央省庁**
- **地方自治体**
  47の広域自治体・1742の基礎自治体
- **サービス提供主体**
  数十万の医療機関・福祉施設・教育機関
- **受益者**

従来、関係者がそれぞれデータを探し、異なる時点・切り口でデータを確認。自団体の状況を見るだけもデータ加工に手間暇がかかっていた。

行政データ利活用を妨げる4つの壁

デジタル庁ダッシュボードシリーズでは、公的統計、行政データなど300種類以上の表データをダッシュボードの要件に合わせてデータ分析基盤に取り込み加工。

データを“使える”状態にする為に、データエンジニアは以下のような４つの壁に苦しめられ続けている。

実態

工程	課題
探す/集める	**1) 探索・解読の壁** カタログや省庁HPの運用が組織や調査によってバラバラで どれが使えるのかわからない
読みとく	**2) 神Excelの壁** 機械が読めない / 事前に解読・解毒が必要 安全に使うための確認コストがすべてのファイルに対して発生
整える・繋げる	**3) 紐づけの壁** そのままでは時系列で繋がらない / 意味解釈ができない / 複数のデータを繋げない 結合するためのリモデリングやマッピングの作成に多大な工数が発生
使い続ける	**4) 継続の壁** 人手が多くかかるので継続的な品質担保が難しい データそのものの改善・修正をしたくともフィードバックの経路が乏しい

（本来）望ましい方向性

問題のレイヤー	アプローチ
データ/ファイル	• 機械可読性の担保/診断 • 公開時に機械可読性を確保する前提での調達・調査
意味識別・接続	• リファレンスデータやメタデータの置き場所と置き方を定める • 1つの概念には一意で不変のIDを採番
運用・方針	• データ作成側へのフィードバック機構 • アジャイル • 行政データの定義と整備の優先度UP

1) 探索・解読の壁

具体例

Web UIにおける目視前提での命名や運用が多く、検索が難しいものや、時系列コレクションの作成が難しいものが散見される。
表番号や統計表名のどちらが信頼できるかは、提供元の運用によるので、利用者側で見極めないといけない。

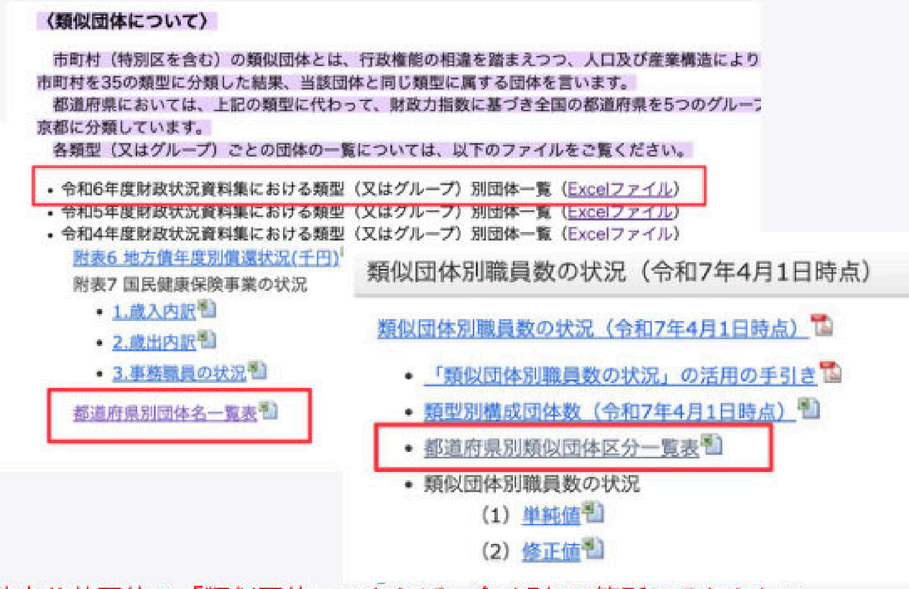
（省庁HP）似たような意味のファイルが散在しており、どれを使うべきか分からない。



名称では判別できない統計表名の例
「〃」や全く同じ名称



目的の統計表に関して、最新年度がいつか、API化有無など
を調査するのは非常に面倒。
上は2018年までしかAPIが無いことを示す例



地方公共団体の「類似団体」であれば、全く別の3箇所にそれらしい
データが存在。
使いやすさが大きく異なり、内容も本当に微妙に異なる。

2) 神Excelの壁

具体例

従来より指摘されるように「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」や「行政データにおける機械可読性に関するルール」に違反するような内容が散見される。

下記は特に初見では発見が難しいパターン。

エッジケースであっても全てのデータに対して常に疑い続けるコスト

が発生する。

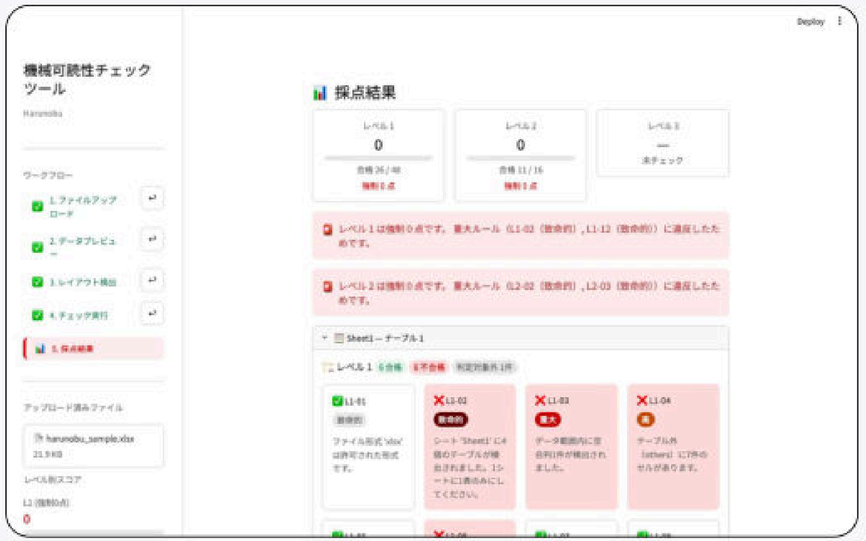
	決算年度	業務コード	団体コード	県名	団体名	団体区分	表番号	表名称	行番号	行名称	54:災害復旧事業支出	055:失業対策事業費支出金	056:委託金	057:普通建設事業	058:災害復旧事業	059:その他	060:財政補助金	061:社会資本整備総合交付金	062:特定防衛施設周辺整備調整交付金	063:電源立地地域対策交付金	064:地方創生関係交付金	065:新型コロナウイルス感染症対応地方創生臨時交付金	066:物価高騰対応分	067:その他	068:子育て世帯等臨時特別支援事業費補助金	069:その他新型コロナウイルス感染症対策関係交付金等	070:その他	071:国有提供施設等所在市町村助成交付金
3080	2022	61	4780369	沖縄県	八重山広	6	4	歳入内訳	1	決算額	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3081	2022	61	4785377	沖縄県	南部広域	6	4	歳入内訳	1	決算額	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3082	2022	61	4782385	沖縄県	北部広域	6	4	歳入内訳	1	決算額	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	152469	0
3083	2022	61	4783393	沖縄県	北谷川行	6	4	歳入内訳	1	決算額	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3084	2022	61	4784407	沖縄県	中部北福	6	4	歳入内訳	1	決算額	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3085	2022	61	4784423	沖縄県	那覇市・	6	4	歳入内訳	1	決算額	0	0	0	0	0	0	0	15478	0	0	0	0	0	0	0	0	0	0
3086	2022	61	4784431	沖縄県	那覇港管	6	4	歳入内訳	1	決算額	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3087	2022	61	4784440	沖縄県	沖縄県介	6	4	歳入内訳	1	決算額	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	312877	0
3088	2022	61	4784458	沖縄県	沖縄県後	7	4	歳入内訳	1	決算額	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3089	2022	61	合計(全国)				4	歳入内訳	1	決算額	251918	14681	91226732	6721652	7785041	83726576	3403144	6.18E+08	21090229	29635641	68617881	1.15E+09	4.67E+08	6.78E+08	9.29E+08	1.83E+09	1.56E+09	37496708
3090	決算年度	業務コード	団体コード	県名	団体名	団体区分	表番号	表名称	行番号	行名称	54:参考・地方	055:参考・分担金及び負担金	056:参考・使用料	057:参考・使用料・授業料	058:参考・使用料・保育所使用料	059:参考・使用料・公営住宅使用料	060:参考・使用料・その他	061:参考・手数料	062:参考・財産収入	063:参考・諸収入	064:参考・その他							
3091	2022	61	11003	北海道	札幌市	1	4	歳入内訳	2	決算額	327678	9587	7730	228	686	3432	3372	78	0	287125	0							
3092	2022	61	12025	北海道	函館市	8	4	歳入内訳	2	決算額	92370	639	0	0	0	0	0	957	449	118426	0							
3093	2022	61	12033	北海道	小樽市	3	4	歳入内訳	2	決算額	14127	0	715	0	0	690	25	726	0	26628	0							
3094	2022	61	12041	北海道	旭川市	8	4	歳入内訳	2	決算額	528737	4758	5093	0	31	5047	15	0	0	139081	0							
3095	2022	61	12050	北海道	室蘭市	4	4	歳入内訳	2	決算額	34410	0	3231	0	0	3231	0	110	0	5180	0							
3096	2022	61	12068	北海道	釧路市	3	4	歳入内訳	2	決算額	105842	471	4996	0	0	4996	0	30	0	38346	0							
3097	2022	61	12076	北海道	帯広市	3	4	歳入内訳	2	決算額	27452	1442	2556	0	960	1596	40	0	0	23280	0							
3098	2022	61	12084	北海道	北見市	3	4	歳入内訳	2	決算額	27002	0	0	0	0	0	0	0	0	0	0							
3099	2022	61	12092	北海道	夕張市	4	4	歳入内訳	2	決算額	6697	0	0	0	0	0	0	0	0	0	0							
3100	2022	61	12106	北海道	岩見沢市	4	4	歳入内訳	2	決算額	7713	1063	0	0	0	0	0	0	0	1	0							
3101	2022	61	12114	北海道	網走市	4	4	歳入内訳	2	決算額	24900	0	7122	0	0	7122	0	0	0	1668	0							
3102	2022	61	12122	北海道	留萌市	4	4	歳入内訳	2	決算額	9601	591	395	0	0	395	0	0	0	2466	0							
3103	2022	61	12131	北海道	苫小牧市	3	4	歳入内訳	2	決算額	29880	0	7435	0	76	7290	69	0	19206	18047	0							
3104	2022	61	12149	北海道	稚内市	4	4	歳入内訳	2	決算額	26385	0	0	0	0	0	0	0	499	0	0							
3105	2022	61	12157	北海道	美唄市	4	4	歳入内訳	2	決算額	10941	0	0	0	0	0	0	0	0	301	0							
3106	2022	61	12165	北海道	芦別市	4	4	歳入内訳	2	決算額	3656	0	0	0	0	0	0	0	0	84	0							

3090行目まで目視しないと折り返しに気付けない例

機械可読性の自動判定ツールの検討

行政データの機械可読性向上の一環として、行政データにおける機械可読性に関するルールをコード化する取組みにおいて、機械可読性ルールをプログラムで自動判定するツール「Harunobu」を開発・公開。

統計ポータル等の公開前のExcel/csvの機械可読性チェックツールや可読性レベルのラベル付けとして使用されることが期待される。



Excel/csvファイルを入力とし、機械可読性ルールへの適合を自動検査・レポートするWeb app/CLI tool “harunobu”

3) 紐づけの壁

具体例

データ部分は整っているが、列名が「列001...」のようになっており、調査様式と照らし合わせなければ内容がわからない。
年度によって対応関係が変化するので、分析対象全年度の対応関係を作成しなければならない。

省庁間で1団体の表記と意味粒度が異なり、機械的なマッチングができず、目検や原課との会話による確認を要することも。



行番号と列番号をもとに項目特定が必要。目検で転記。



調査票のレイアウトと対応する意味が年度間で大きく変更される例
分析可能にするには情報を転記し、年度間共通で使えるマスタを作成しなければ先に進めない

4) 継続の壁

具体例

ある1ダッシュボードの更新にはe-Statから44ファイル、 52のHPから1,659ファイルを毎年取得。

→ データやダッシュボードの**価値を維持するために常に更新・メンテナンスし続ける必要**があるが、
　一つひとつのデータソースの**扱いづらさが運用のコスト高となり続けてしまう**。

原則可能な限りプログラム等による自動化を行っているが、先述の3つの壁はどうしても人手が必要になっている。

そのためデータ加工時の難点をできるかぎり原課にフィードバックし、草の根的な改善を継続している。
しかし、一度公開されたデータ形式の変更には慎重にならざるを得ないのが実態で、中々改善が進みづらい。

AI-readyなデータの土台となるそもそものデータの継続的な改善について、政府としての優先度を上げる必要性がある。

好例

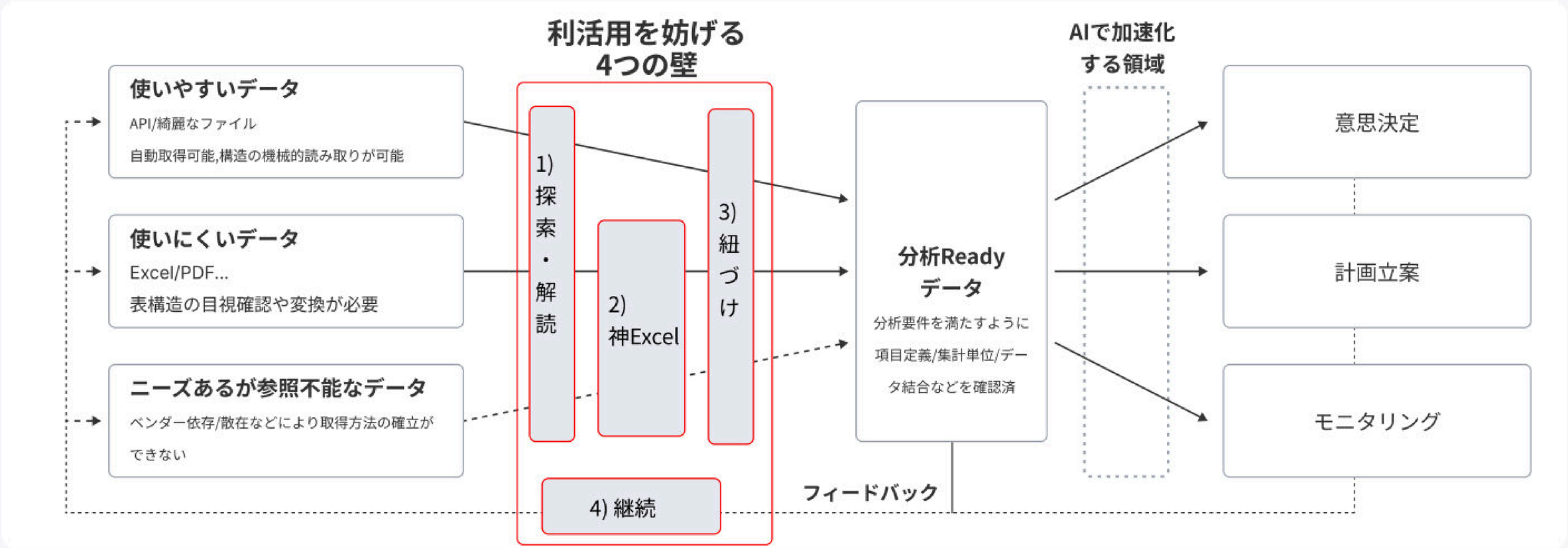
- 省庁HPで公表されている既存の調査結果に自治体コードを追加
- 調査システム更改に併せて、調査項目帳票自体も機械可読性の担保された表データとして公開（検討中）

断念した例

- 47都道府県HPに散在していたExcelの集約と1テーブル化
- すでにe-Statで公開されているExcelの機械可読性の向上（前例優先）

行政データ利活用拡大に向けて供給側で求められる整備内容

AIは整ったデータの上では利活用を大きく加速するが、”使える”データの供給が拡大に向けたボトルネックである。



利活用の現場から見て、優先して整備することが望まれる条件（案）

- （データ/ファイル層）機械的に即時読める状態を確保する
  - e-StatであればAPIが即日公開される
  - Excelであれば機械可読性が担保されている
    - e.g.) harunobu等によって機械可読性のレベルが確認されたもの
- （意味・接続層）調査項目等が機械可読性が担保された状態で公開されている
- （運用・方針）フィードバック機構の構築/データそのものを改善する方針を示す