ガバメント・オープンデータの品質管理に向けた取り組み

2022-6-9 (Thu.)

10:00 - 12:00

Japan Open Science Summit 2022

D4 研究データの品質表示を考える

下山 紗代子

リンクデータ 代表理事/

インフォ・ラウンジ株式会社 取締役/

デジタル庁 データスペシャリスト

※本セッションは個人の立場で参加しており、所属組織の見解を示すものではございません

Sayoko Shimoyama

- Enterprise
  - Open Data Platform — LinkData.org
  - Medical Institutions Database — me
  - Web & Data Management — INFO LOUNGE
- Public Sector
  - Data Specialist — デジタル庁 Digital Agency
  - local government advisor — MIC
  - Open Data Missionary — OPEN DATA JAPAN
- Civic Tech
  - Code for Japan
  - Code for YOKOHAMA — code for YOKOHAMA
  - Code for AMAMI

Background:

分子生物学→バイオインフォマティクス

データは
共有されないと…

価値は埋もれたまま

2022-06-09

Sayoko Shimoyama

Enterprise

Open Data Platform

LinkData.org

Medical Institutions Database

me

Web & Data Management

INFO LOUNGE

Public Sector

Data Specialist

デジタル庁
Digital Agency

local government advisor

MIC

Open Data Missionary

OPEN DATA JAPAN

Civic Tech

Code for Japan

{ }

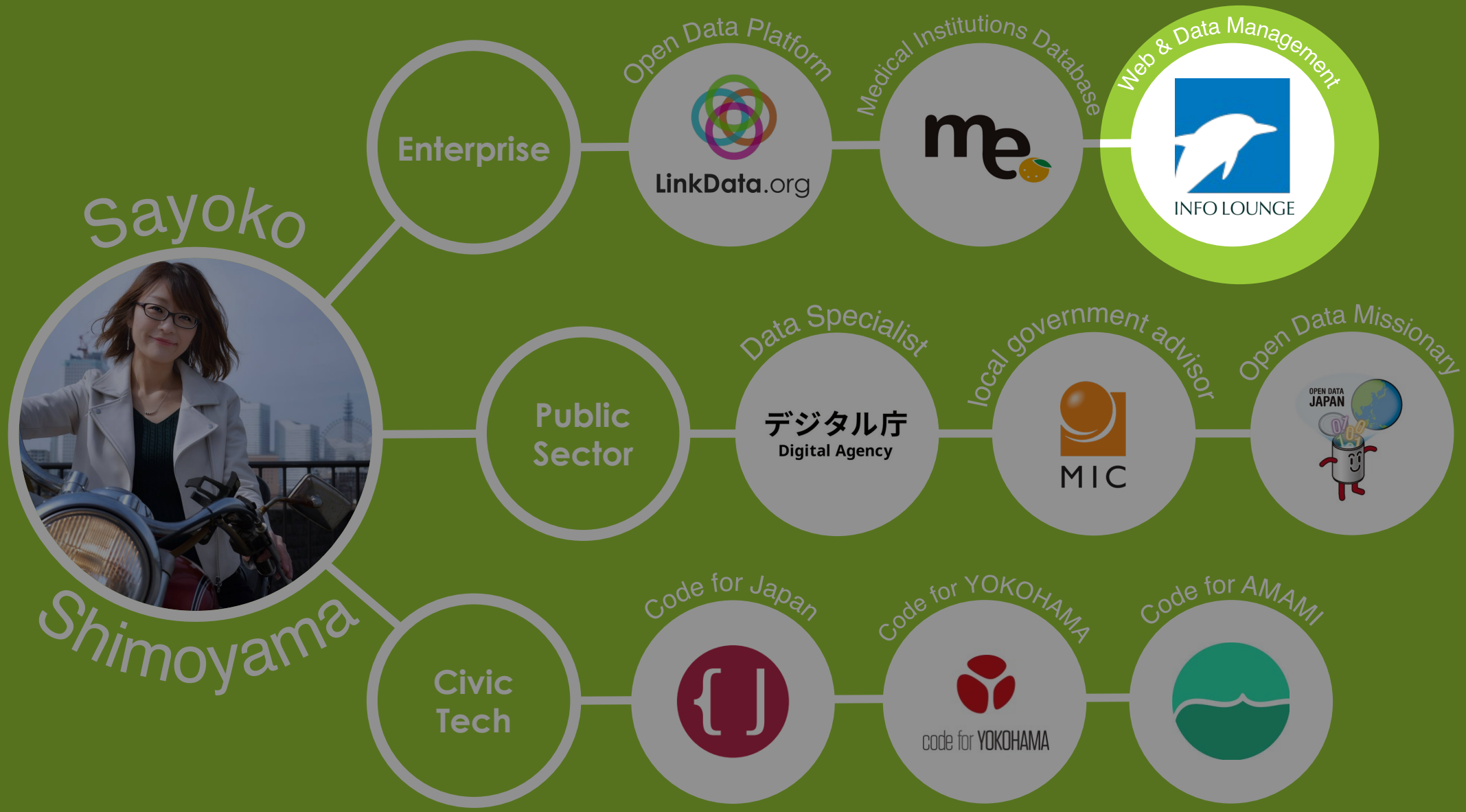
Code for YOKOHAMA

code for YOKOHAMA

Code for AMAMI

LinkData.org

オープンデータの作成と活用を支援するプラットフォーム

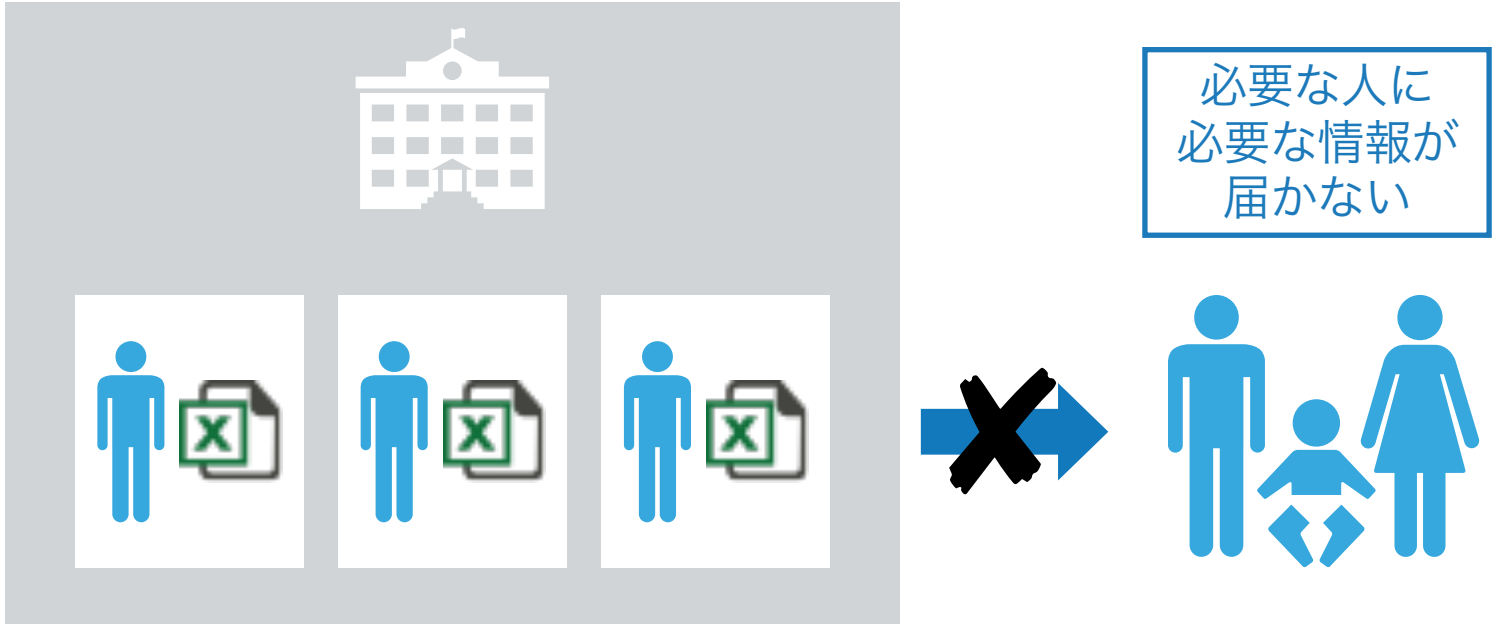

Sayoko
Shimoyama
Enterprise
Public Sector
Civic Tech
Open Data Platform
LinkData.org
Medical Institutions Database
Web & Data Management
INFO LOUNGE
Data Specialist
デジタル庁
Digital Agency
local government advisor
MIC
Open Data Missionary
OPEN DATA JAPAN
Code for Japan
Code for YOKOHAMA
code for YOKOHAMA
Code for AMAMI

所属企業の取組事例①

育なび
～39データセットを統合、子育て情報配信をパーソナライズ

- 従来、データは所管課ごとに管理されており、統合的に扱えていなかった
- 住民向けの情報提供も非効率になっていた

Issue



行政では
データもタテ割りで管理されがち

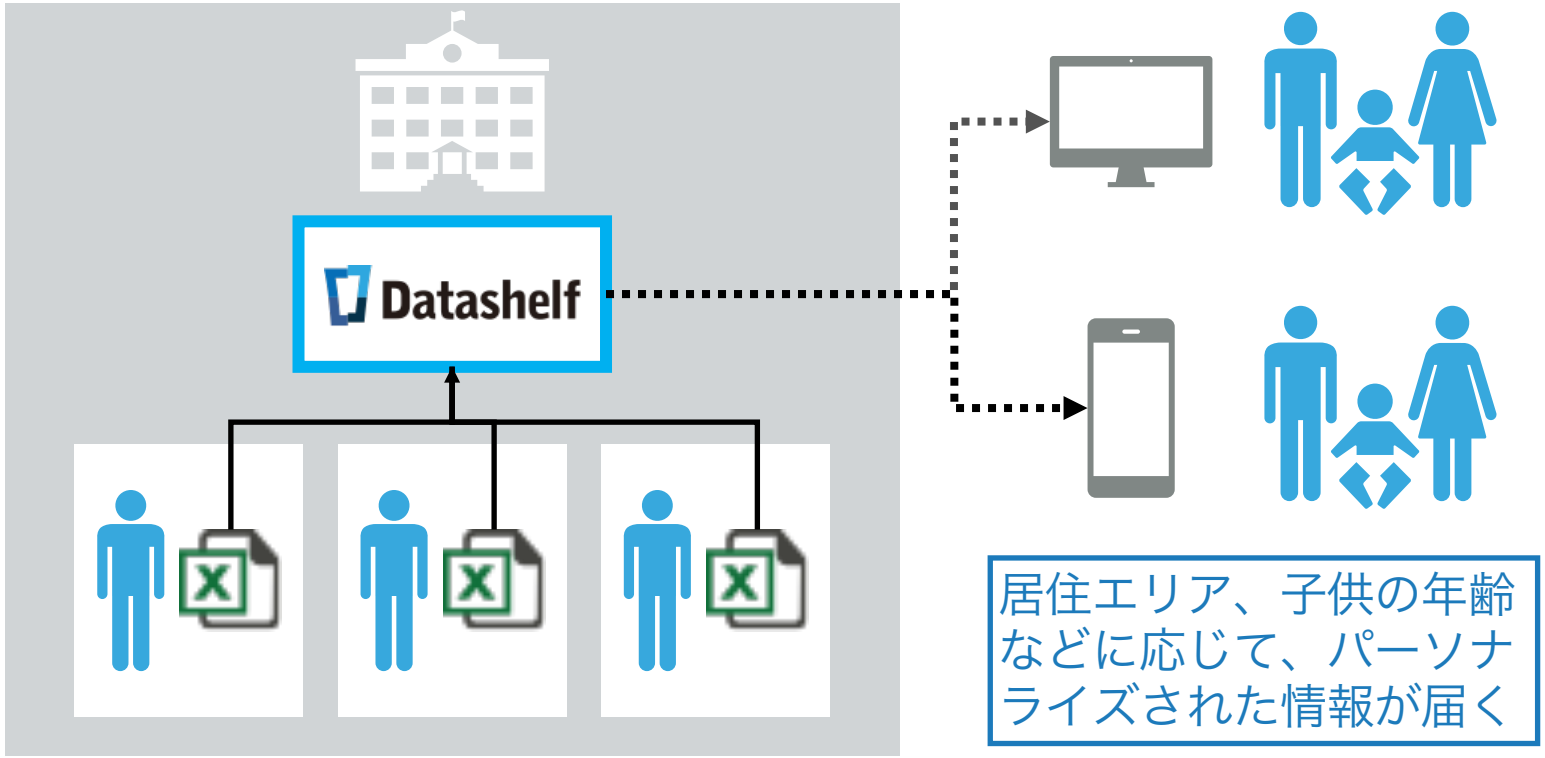
https://datashelf.jp/

所属企業の取組事例①

育なび
～39データセットを統合、子育て情報配信をパーソナライズ

- 子育て関連のデータをデータ管理基盤：Datashelfで統合し、オープンデータ化
- 子育て世帯に向けてパーソナライズした情報配信を可能に

Solution



https://datashelf.jp/

Sayoko Shimoyama

Enterprise

- Open Data Platform — LinkData.org
- Medical Institutions Database — me
- Web & Data Management — INFO LOUNGE

Public Sector

- Data Specialist — デジタル庁 Digital Agency
- local government advisor — MIC
- Open Data Missionary — OPEN DATA JAPAN

Civic Tech

- Code for Japan
- Code for YOKOHAMA — code for YOKOHAMA
- Code for AMAMI

所属企業の取組事例②

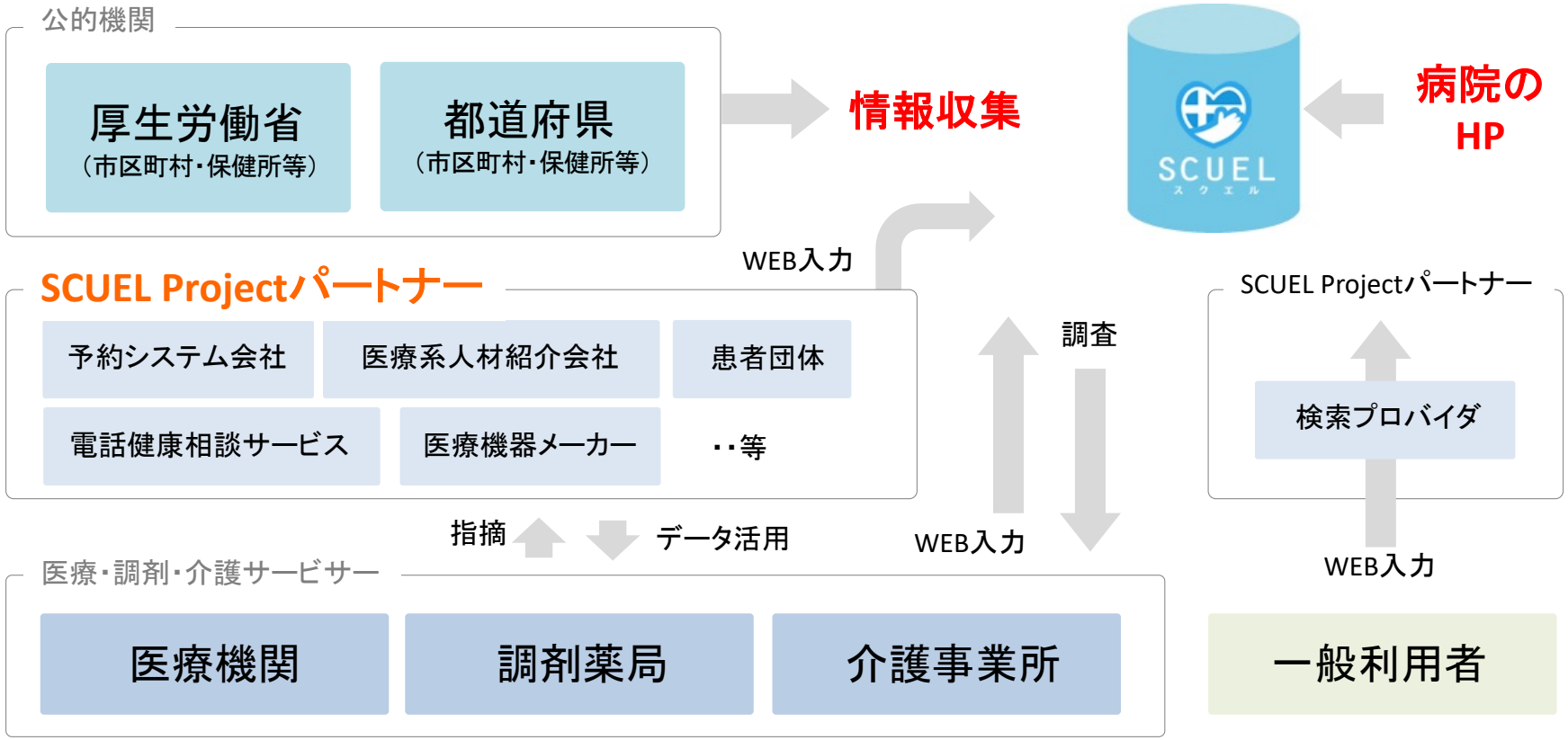
SCUEL データベース
～医薬介護分野のビジネス支援

- 国や自治体が公開している医療機関・薬局・介護施設のデータを統合・構造化してデータベースとして提供
- 新規開業施設や営業日・診療科目等の詳細情報を入手できるため、製薬会社や医療機器メーカーの営業支援、経営分析等で活用されている

http://mecompany.me/service

参考：SCUELデータベースの開発方法

公的な公開情報を収集・正規化・統合処理を半自動化。
さらに、データ提供先からのフィードバックを元に情報を強化。



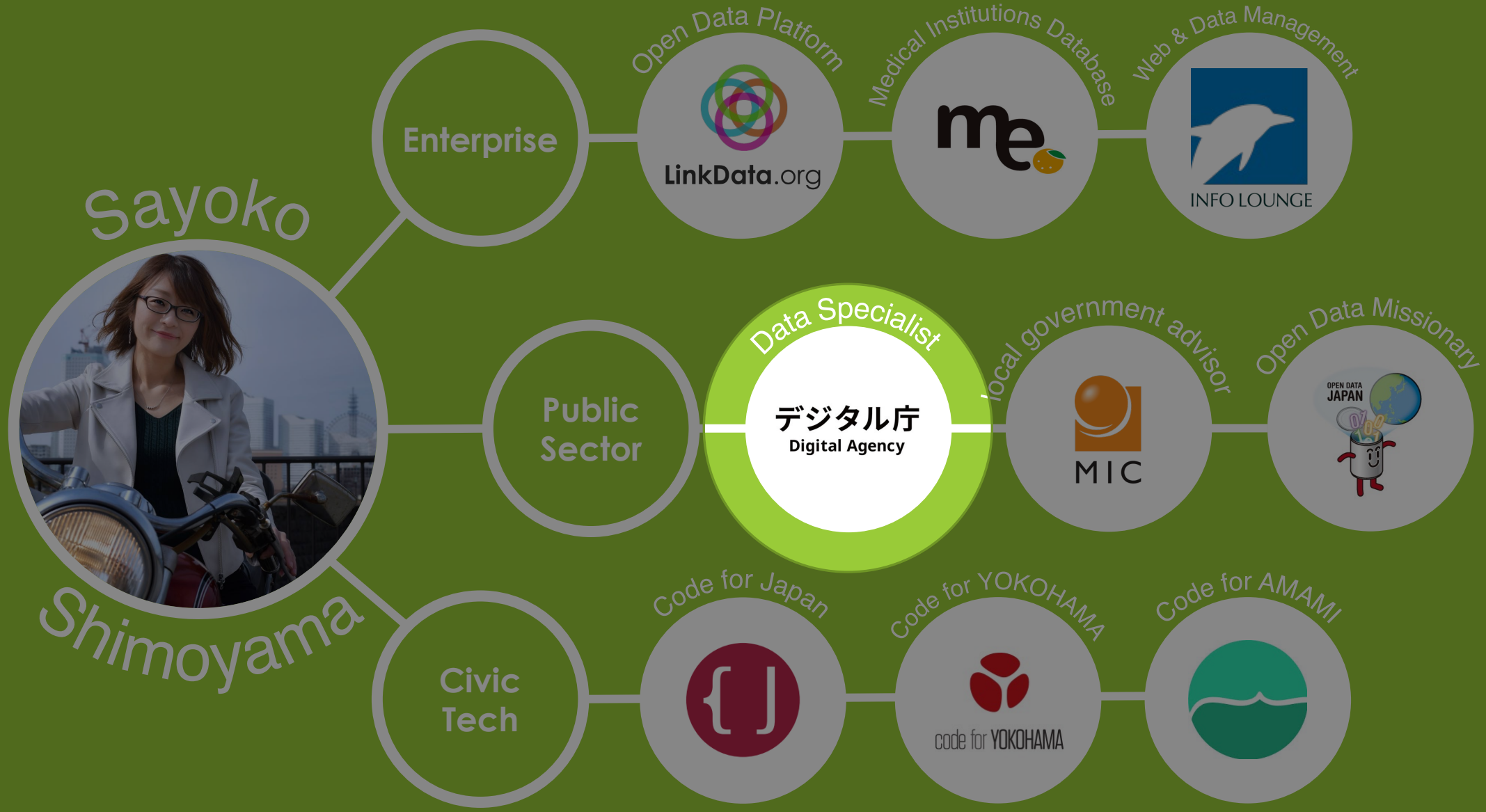
データ利用事業者視点でのガバメント・オープンデータの課題

オープンデータになっていないデータも多く、郵送ベースでの開示請求を頻繁に行わなければならず、非効率

オープンデータになっていても、PDF等機械判読性の低いデータや品質の低いデータが多い

オープンデータポータルが国・都道府県・市区町村で連携しておらず、一箇所で探せない

※一部地域では都道府県-市区町村間で連携されている


Sayoko
Shimoyama
Enterprise
Public Sector
Civic Tech
Open Data Platform
LinkData.org
Medical Institutions Database
Web & Data Management
INFO LOUNGE
Data Specialist
デジタル庁
Digital Agency
local government advisor
MIC
Open Data Missionary
OPEN DATA JAPAN
Code for Japan
Code for YOKOHAMA
code for YOKOHAMA
Code for AMAMI

2020年10月～

データ戦略タスクフォースに構成員として参加

https://www.kantei.go.jp/jp/singi/it2/dgov/data_strategy_tf/dai1/gijisidai.html

参考資料1

データ戦略タスクフォースの開催について

令和２年10月12日
デジタル・ガバメント閣僚会議決定

1　デジタル社会においては、データを十分に活用することが不可欠であり、競争力の源泉である。今般のコロナ危機では、迅速で的確な対応ができず、データ活用基盤（デジタルデータの整備、標準化、データの取扱いルール等）が官民ともに不十分であることが露呈した。デジタル国家にふさわしいデータ戦略を策定するためデータ戦略タスクフォース（以下「タスクフォース」という。）を開催する。

2　タスクフォースの構成は、次のとおりとする。ただし、主査は、必要があると認めるときは、構成員以外の関係行政機関の職員、有識者その他の関係者の出席を求めることができる。

主　　査　内閣総理大臣補佐官（国土強靱化及び復興等の社会資本整備、地方創生、健康・医療に関する成長戦略並びに科学技術イノベーション政策その他特命事項担当）

構 成 員　砂金　信一郎　LINE 株式会社執行役員 AI カンパニーカンパニーCEO
遠藤　信博　一般社団法人日本経済団体連合会審議員会副議長
兼サイバーセキュリティ委員長
日本電気株式会社取締役会長
太田　直樹　株式会社 New Stories 代表取締役
小木曽　稔　一般社団法人新経済連盟政策部長
越塚　登　東京大学大学院教授
後藤　厚宏　情報セキュリティ大学院大学学長
下山　紗代子　一般社団法人リンクデータ代表理事/インフォ・ラウンジ株式会社取締役
庄司　昌彦　武蔵大学教授
手塚　悟　慶應義塾大学教授
村井　純　慶應義塾大学教授
渡部　俊也　東京大学未来ビジョン研究センター教授

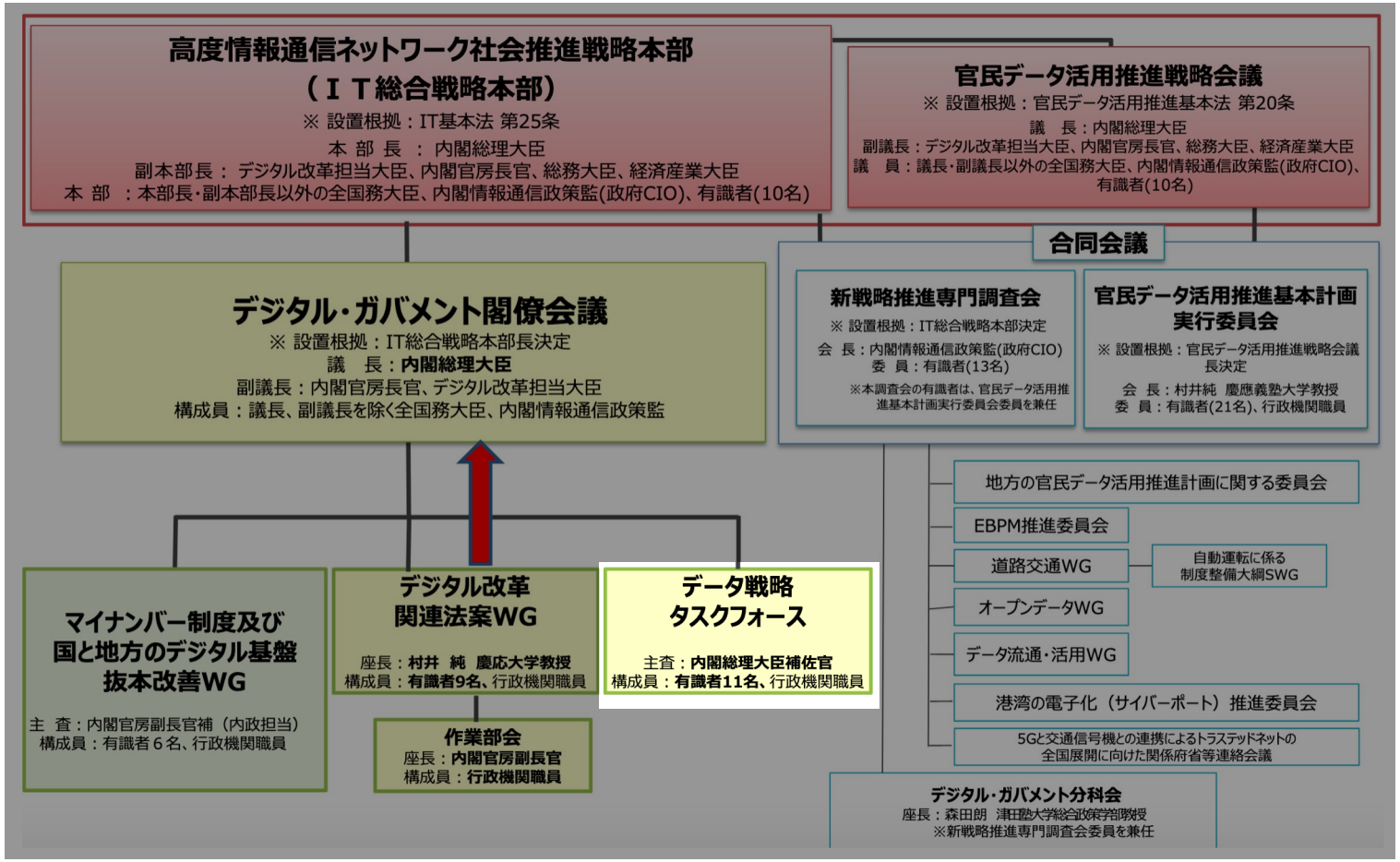
内閣官房情報通信技術（IT）総合戦略室長代理（副政府 CIO）
内閣官房番号制度推進室長（副政府 CIO）
内閣官房イノベーション推進室イノベーション総括官
内閣官房デジタル市場競争本部事務局次長
内閣府大臣官房審議官（科学技術政策・イノベーション担当）
内閣府知的財産戦略推進事務局長
個人情報保護委員会事務局審議官
総務省大臣官房総括審議官（情報通信担当）
文部科学省大臣官房サイバーセキュリティ・政策立案総括審議官
経済産業省商務情報政策局長

データ戦略は デジタル・ガバメントを実現するための重要戦略として策定が進められてきた （2020年度）

出典：2020年10月29日デジタル改革関連法案ワーキンググループ作業部会（第１回）

資料１－１
デジタル社会実現に向けたIT総合戦略本部の推進体制

https://www.kantei.go.jp/jp/singi/it2/dgov/houan_wg/sagyobukai/dai1/gijisidai.html

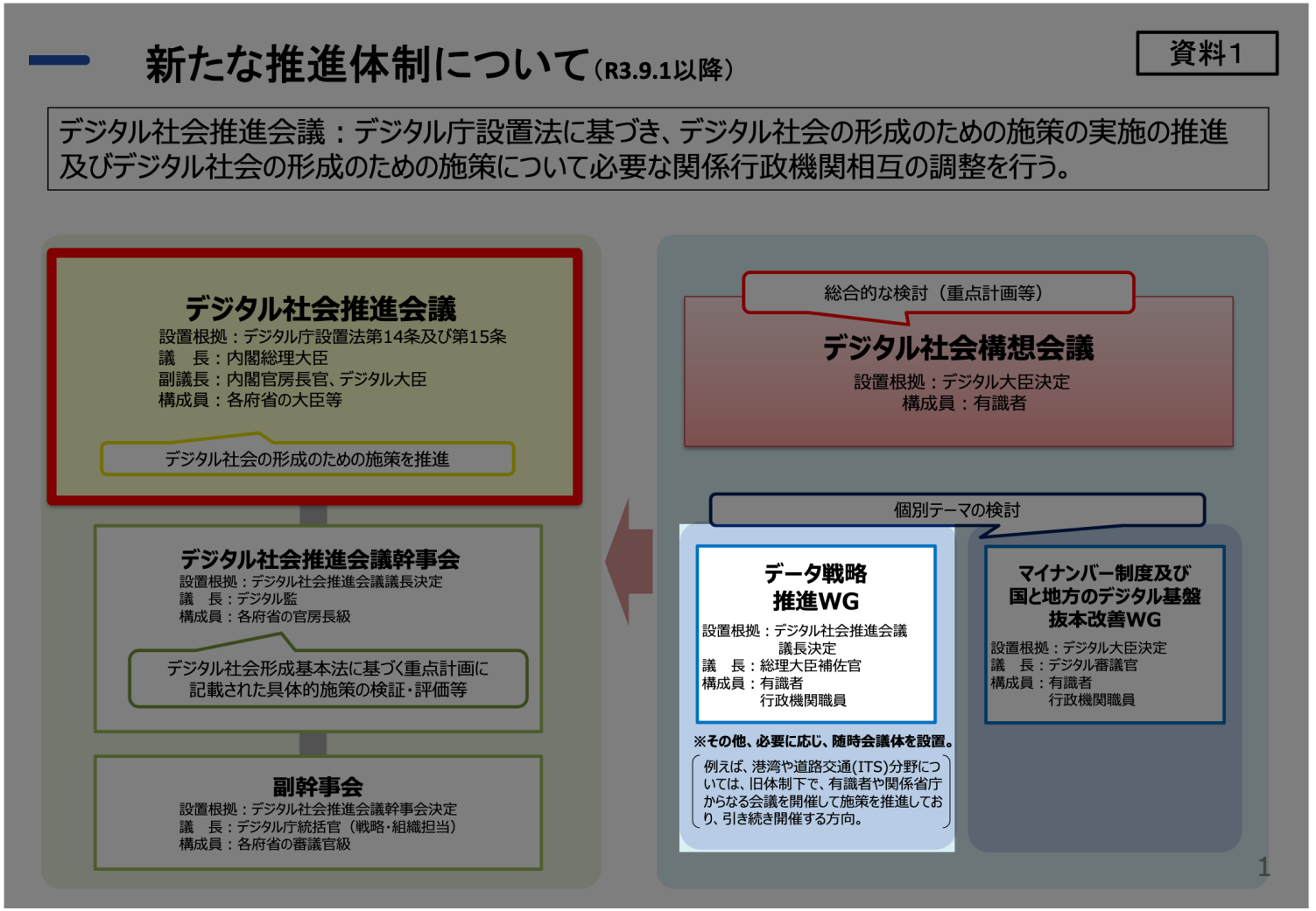


デジタル庁発足以降、データ戦略は推進フェーズへ移行（2021年9月～）

出典：第1回デジタル社会推進会議

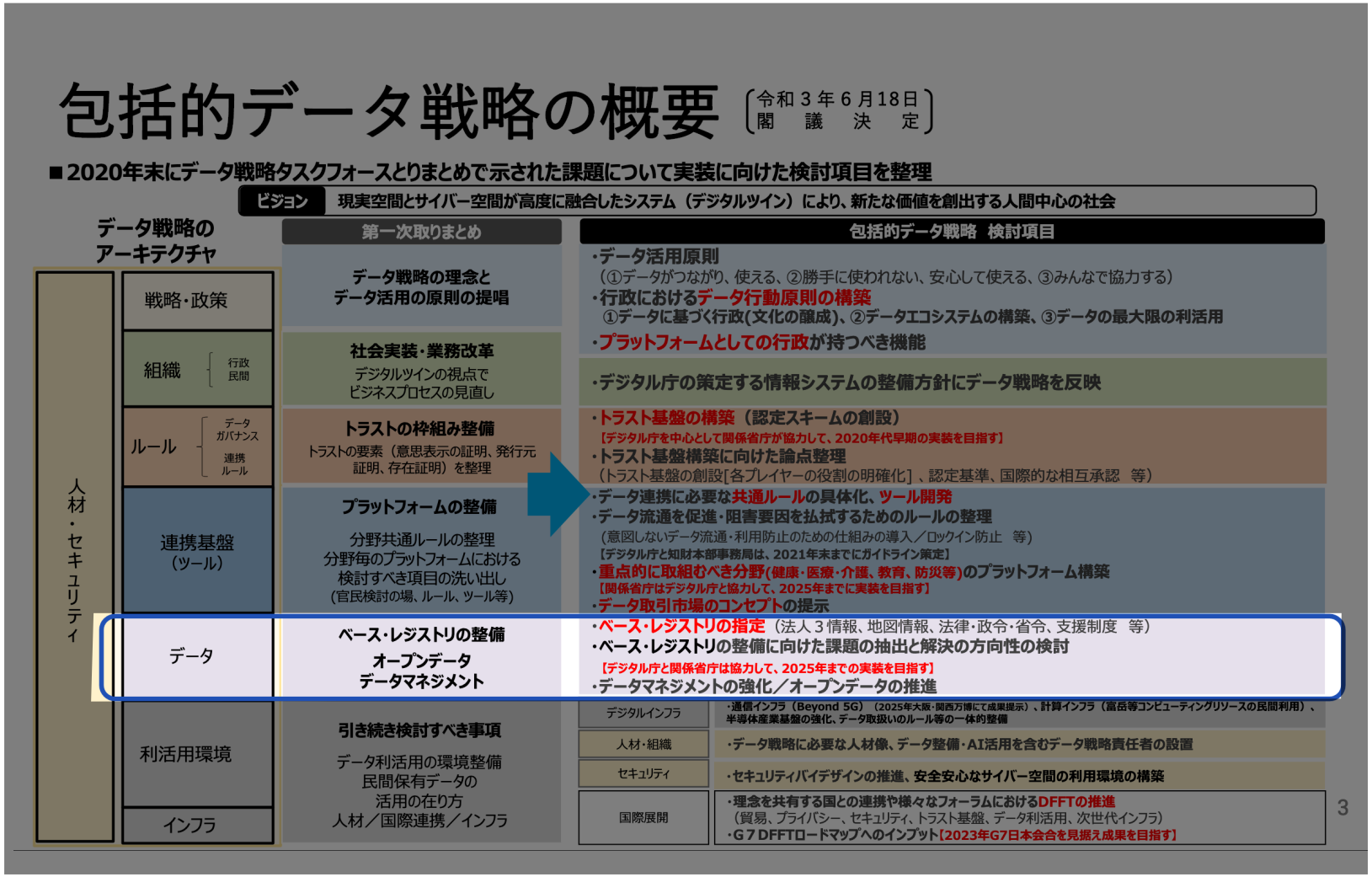
資料1
新たな推進体制について

https://cio.go.jp/sites/default/files/uploads/documents/digital/20210906_meeting_promoting_02.pdf



ベース・レジストリ整備はデータ戦略における大きな柱

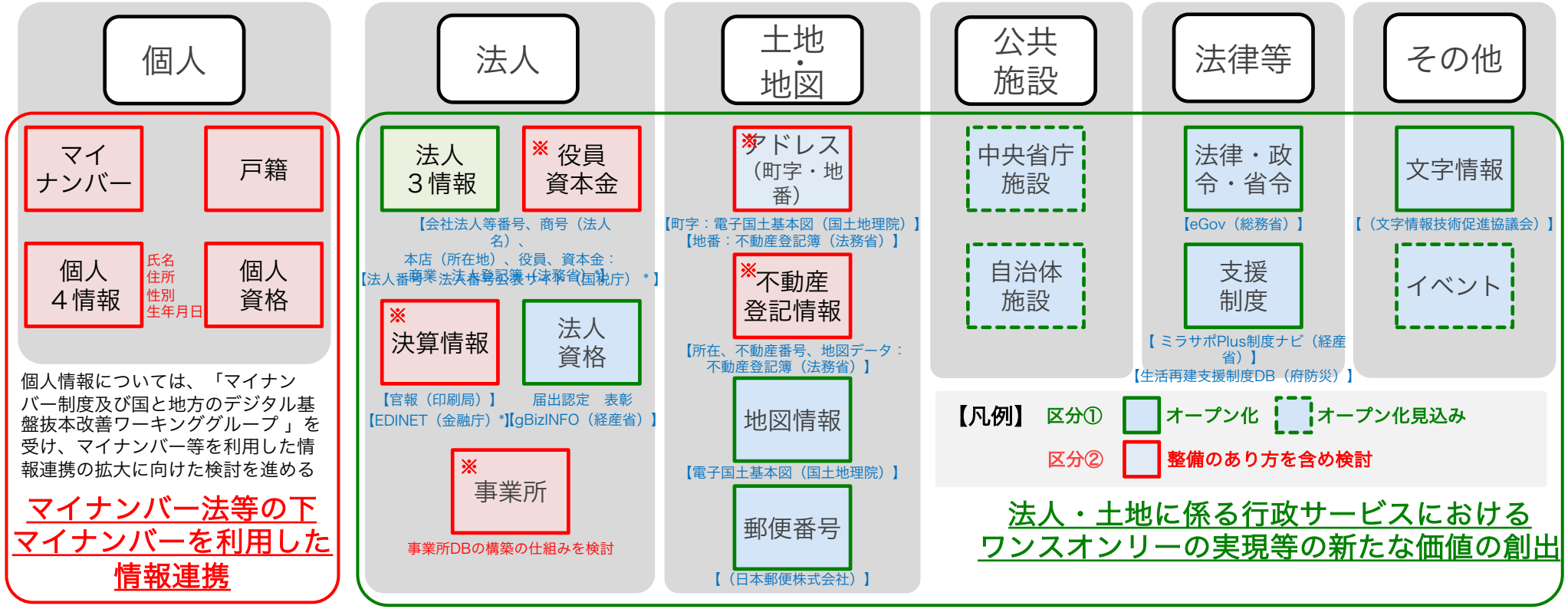
出典：データ戦略推進ワーキンググループ（第1回）
資料2：ベース・レジストリのコンセプト（案）
https://www.digital.go.jp/meeting/posts/icPUoaig



ベース・レジストリの定義

ベース・レジストリとは、公的機関等で登録・公開され、様々な場面で参照される**正確性**や**最新性**が確保された**社会の基幹となるデータベース**

出典：「ベース・レジストリの指定について」, 2021年5月26日,内閣官房情報通信技術（IT）総合戦略室決定, https://cio.go.jp/node/2764

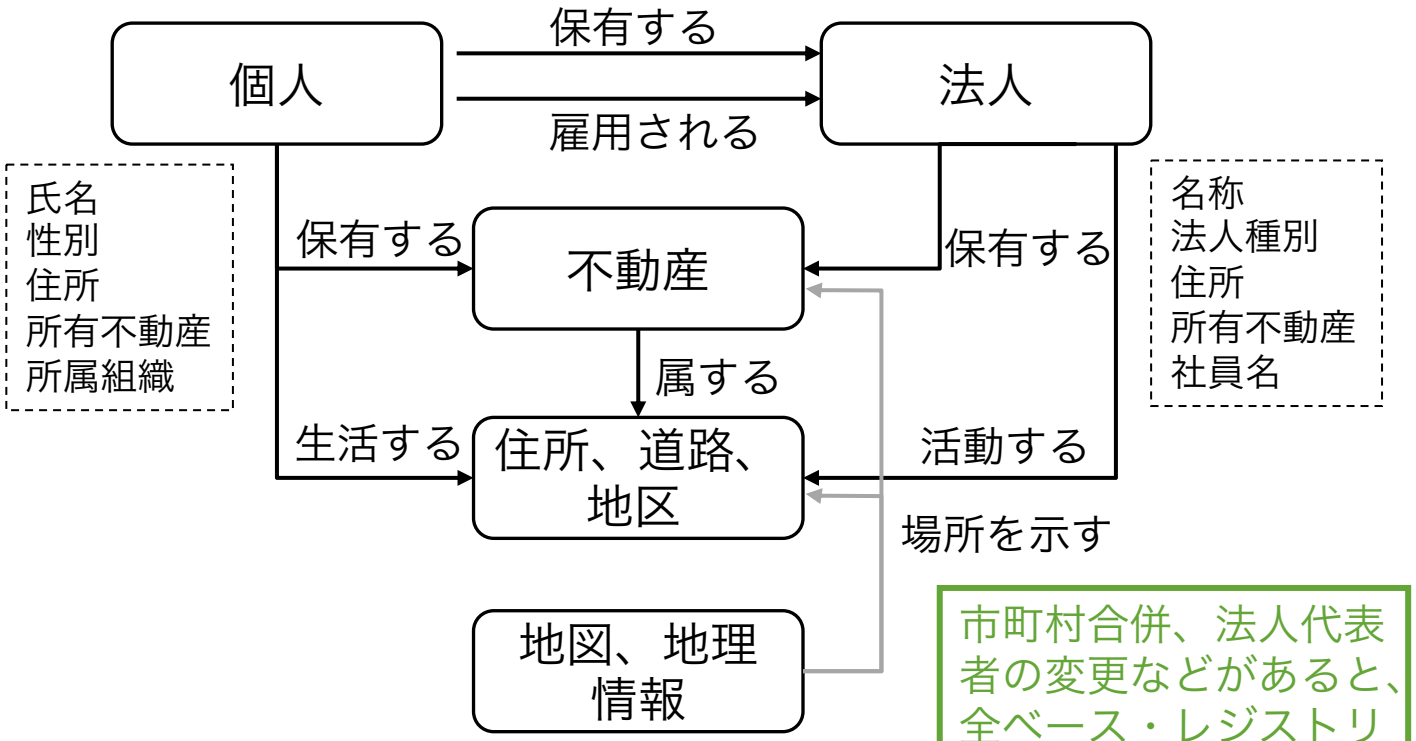


※具体的なデータの活用・整備方法等について要調整　*既にgBizINFOと連携済みでありgBizINFOの活用も検討中

社会の基幹となるデータベース：ベース・レジストリ

ベース・レジストリ同士は連携している必要がある

- ベース・レジストリは複数のベース・レジストリを相互参照することで、効果を飛躍的に増大させることができる。
- そのため、政府横断で方針を決定し、その方針に従い、各データの担当組織が責任を持ち推進する必要がある。



市町村合併、法人代表者の変更などがあると、全ベース・レジストリの情報が連動して更新される必要がある

出典：データ戦略タスクフォース（第１回）, 令和２年１０月２３日, 資料２：ベース・レジストリの概要
https://www.kantei.go.jp/jp/singi/it2/dgov/data_strategy_tf/dai1/gijisidai.html

データ品質管理フレームワークによるベース・レジストリ評価

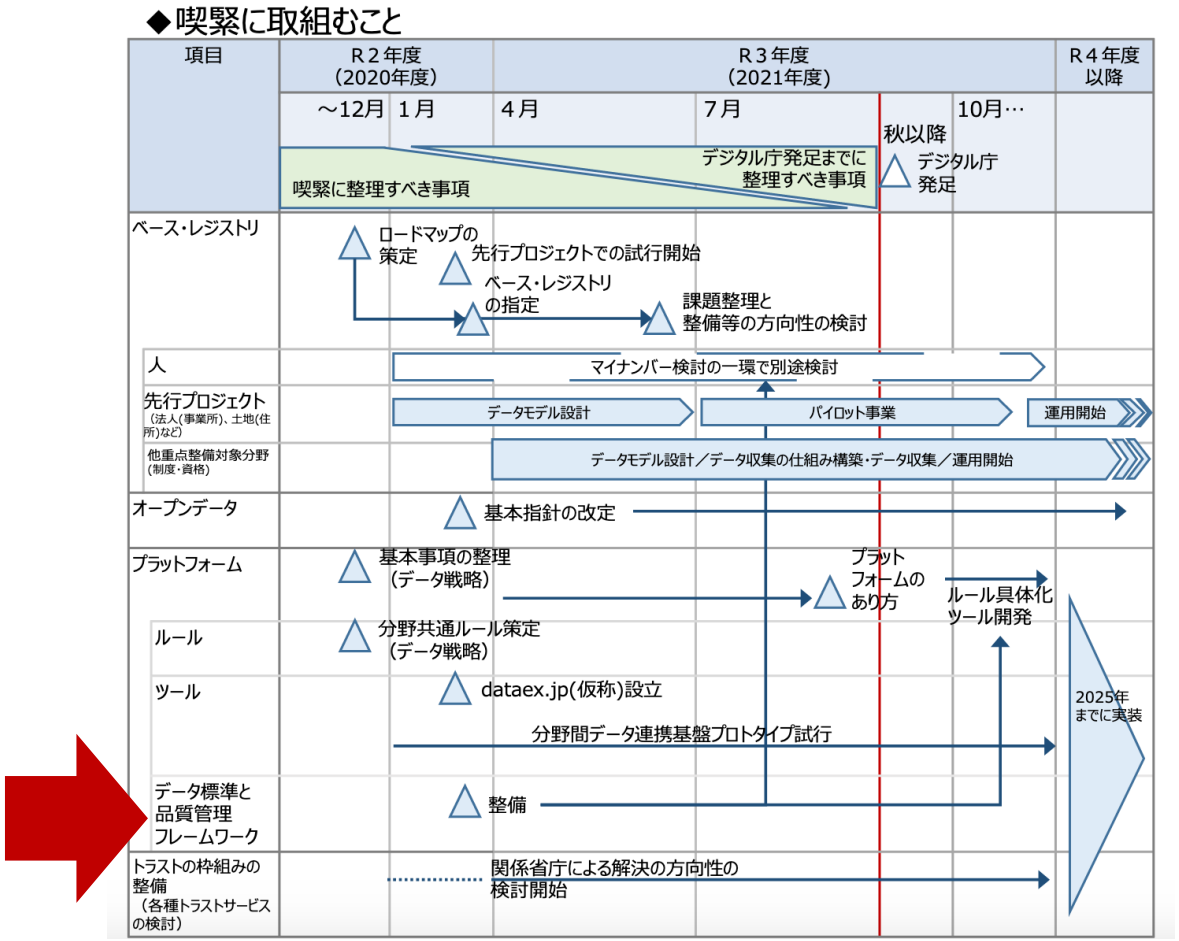
３.４ベース・レジストリ推進のための仕組み

3) データ品質評価と公開

ベース・レジストリ候補になる主要データに対して、別途IT総合戦略室で策定する**データ品質管理フレームワークによる評価を実施**し、その結果を公開する。また、国際的に相互参照可能なように、国際的な情報交換を実施する。2020年度末までにデータ品質評価の試行を実施する。

出典：データ戦略タスクフォース（第４回）別紙 ベース・レジストリ・ロードマップ（案）P.18
https://www.kantei.go.jp/jp/singi/it2/dgov/data_strategy_tf/dai4/gijisidai.html

データ品質管理フレームワーク：2020年度末までに整備



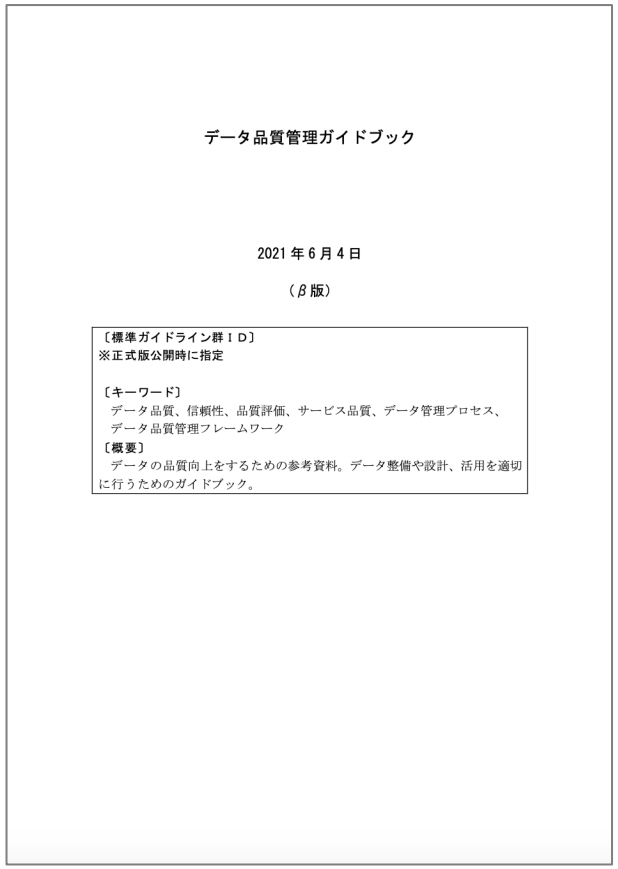
データ戦略タイムラインの「喫緊に取組むこと」に位置付けられた

出典：データ戦略タスクフォース第一次とりまとめ（令和2年12月21日）P.46

https://www.kantei.go.jp/jp/singi/it2/egov/

データ品質管理ガイドブック（β版）2021年6月4日公開

https://cio.go.jp/guides#renkeimodel



データ品質管理ガイドブック

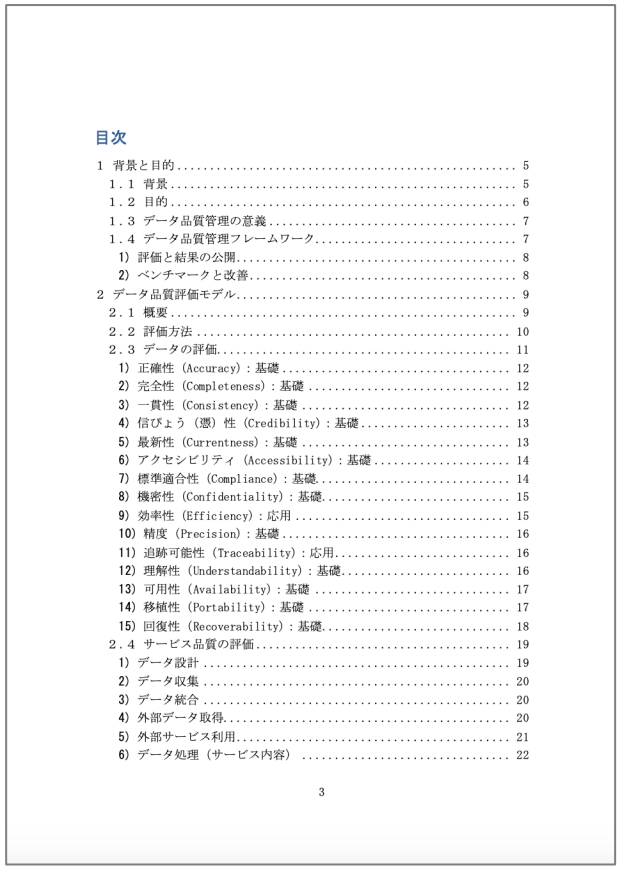
2021 年 6 月 4 日

（β版）

〔標準ガイドライン群ＩＤ〕
※正式版公開時に指定

〔キーワード〕
データ品質、信頼性、品質評価、サービス品質、データ管理プロセス、データ品質管理フレームワーク

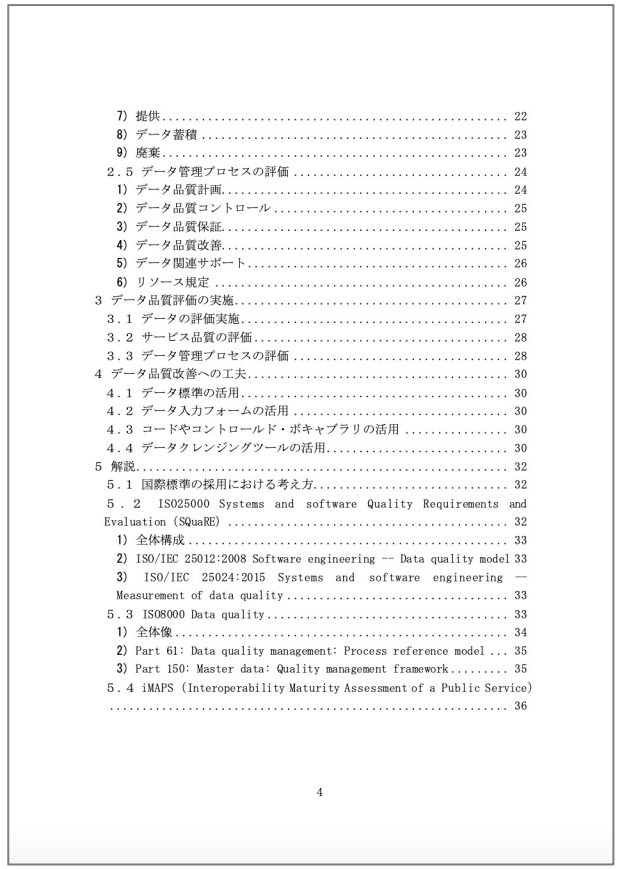
〔概要〕
データの品質向上をするための参考資料。データ整備や設計、活用を適切に行うためのガイドブック。



目次

1 背景と目的 5
1．1 背景 5
1．2 目的 6
1．3 データ品質管理の意義 7
1．4 データ品質管理フレームワーク 7
1）評価と結果の公開 8
2）ベンチマークと改善 8
2 データ品質評価モデル 9
2．1 概要 9
2．2 評価方法 10
2．3 データの評価 11
1）正確性（Accuracy）：基礎 12
2）完全性（Completeness）：基礎 12
3）一貫性（Consistency）：基礎 12
4）信ぴょう（憑）性（Credibility）：基礎 13
5）最新性（Currentness）：基礎 13
6）アクセシビリティ（Accessibility）：基礎 14
7）標準適合性（Compliance）：基礎 14
8）機密性（Confidentiality）：基礎 15
9）効率性（Efficiency）：応用 15
10）精度（Precision）：基礎 16
11）追跡可能性（Traceability）：応用 16
12）理解性（Understandability）：基礎 16
13）可用性（Availability）：基礎 17
14）移植性（Portability）：基礎 17
15）回復性（Recoverability）：基礎 18
2．4 サービス品質の評価 19
1）データ設計 19
2）データ収集 20
3）データ統合 20
4）外部データ取得 20
5）外部サービス利用 21
6）データ処理（サービス内容） 22

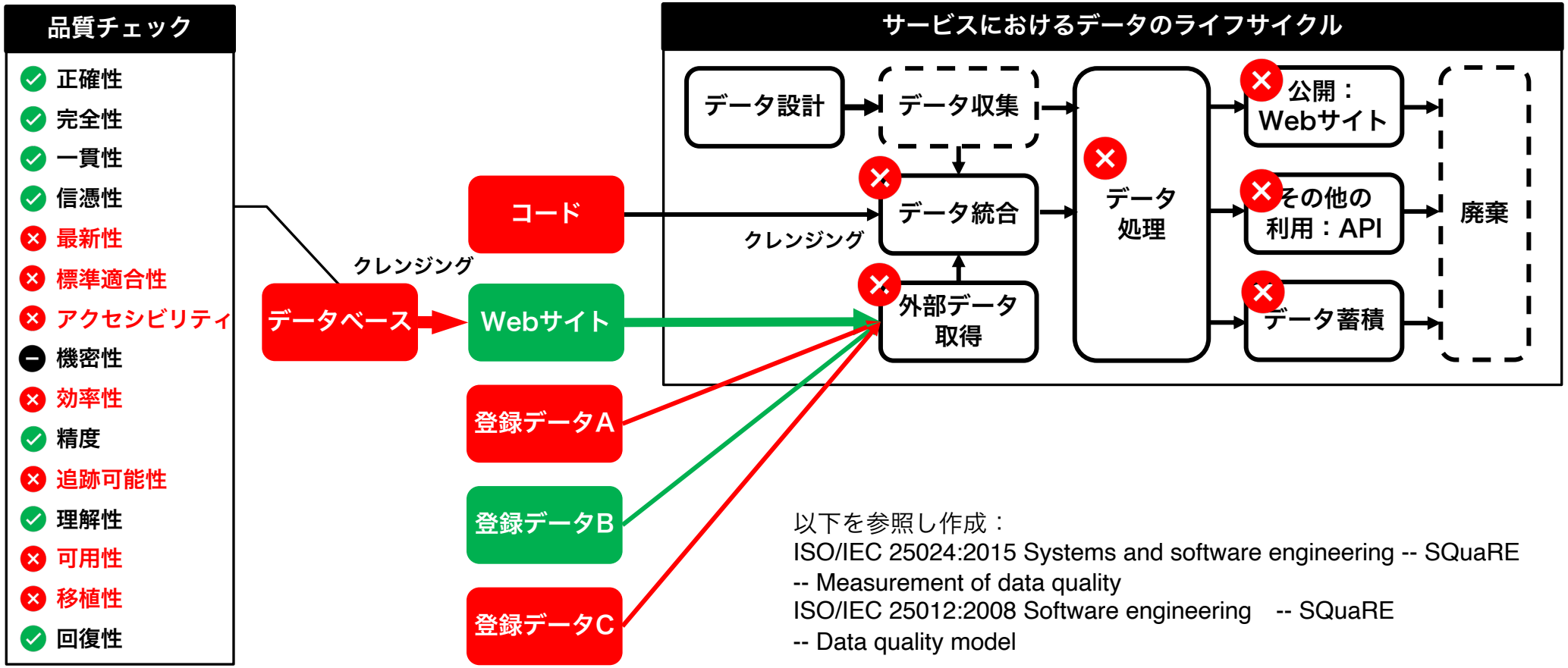
3



7）提供 22
8）データ蓄積 23
9）廃棄 23
2．5 データ管理プロセスの評価 24
1）データ品質計画 24
2）データ品質コントロール 25
3）データ品質保証 25
4）データ品質改善 25
5）データ関連サポート 26
6）リソース規定 26
3 データ品質評価の実施 27
3．1 データの評価実施 27
3．2 サービス品質の評価 28
3．3 データ管理プロセスの評価 28
4 データ品質改善への工夫 30
4．1 データ標準の活用 30
4．2 データ入力フォームの活用 30
4．3 コードやコントロールド・ボキャブラリの活用 30
4．4 データクレンジングツールの活用 30
5 解説 32
5．1 国際標準の採用における考え方 32
5．2 ISO25000 Systems and software Quality Requirements and Evaluation（SQuaRE） 32
1）全体構成 33
2）ISO/IEC 25012:2008 Software engineering -- Data quality model 33
3）ISO/IEC 25024:2015 Systems and software engineering — Measurement of data quality 33
5．3 ISO8000 Data quality 33
1）全体像 34
2）Part 61: Data quality management: Process reference model 35
3）Part 150: Master data: Quality management framework 35
5．4 iMAPS（Interoperability Maturity Assessment of a Public Service） 36

4

低品質のデータがあると
サービス全体に最後まで影響が残る



以下を参照し作成：
ISO/IEC 25024:2015 Systems and software engineering -- SQuaRE -- Measurement of data quality
ISO/IEC 25012:2008 Software engineering -- SQuaRE -- Data quality model

まずは現状把握からはじめよう
→データ品質管理ガイドブック

データ所有者

- 自分の組織で運用しているサービスで使われている/公開されているデータの自己診断
- 改善点を確認し、段階的に改善サイクルを回すことが可能（Maturity Model; 成熟度モデルによる評価）

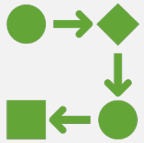
データ利用者

- 国際標準を用いた外形評価が可能
- 客観的な基準を用いて改善の提案ができる

データ品質管理ガイドブックで採用した３つの国際標準

ISO/IEC 25012（**データ品質の評価**）
：データそのものの品質

ISO/IEC 25024（**サービス品質の評価**）
：データを使ったサービス実現プロセスに関する品質



ISO/TS 8000-61
（**データ管理プロセスの評価**）
：データの整備から活用までの管理プロセスに関する品質

ISO/IEC 25012（データ品質の評価）

1. 正確性 (Accuracy)
2. 完全性(Completeness)
3. 一貫性(Consistency)
4. 信憑性(Credibility)
5. 最新性(Currentness)
6. アクセシビリティ(Accessibility)
7. 標準適合性(Compliance)
8. 機密性(Confidentiality)
9. 効率性(Efficiency)
10. 精度(Precision)
11. 追跡可能性(Traceability)
12. 理解性(Understandability)
13. 可用性(Availability)
14. 移植性(Portability)
15. 回復性(Recoverability)

利用者の視点から外形的に評価することが可能

ISO/IEC 25012（データ品質の評価）

1. 正確性 (Accuracy)

データの基本は正確であること。
データの正しさを以下の点に着目して評価する。

評価項目

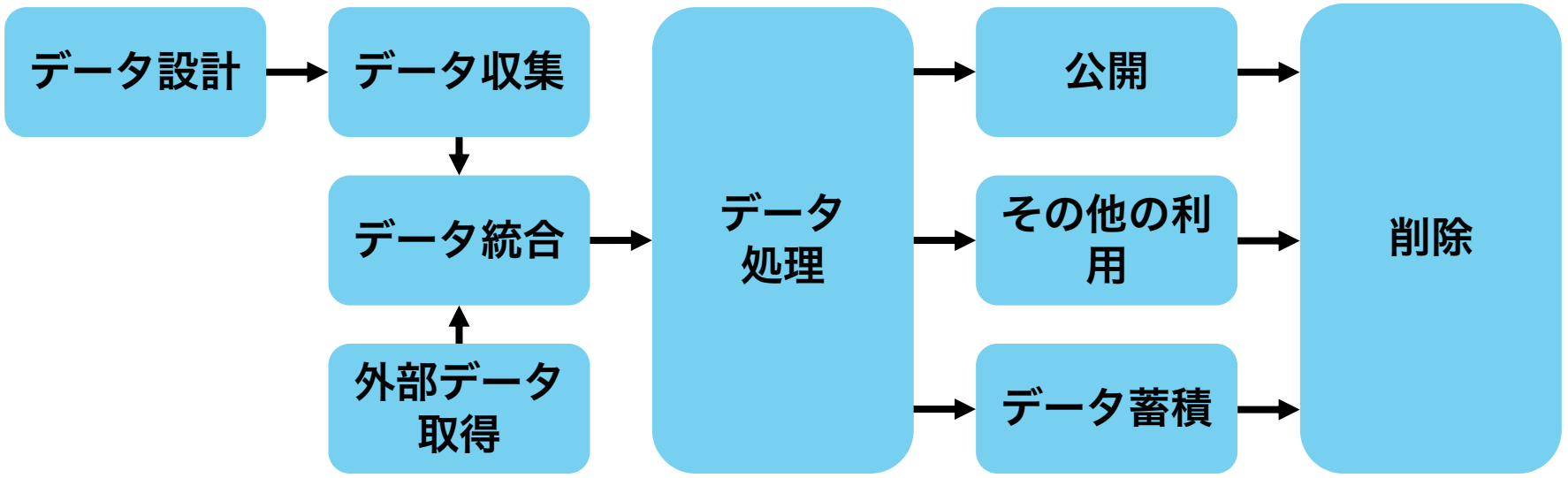
- 書式が正しいか
- 誤字脱字などはないか
- 意味的な誤りがないか
- データに誤りはないか

問題となる例

- 「同上」「〃」などの記述がある
- 日付や数字が入るべき欄に「不明」など数字以外の文字列が記述されている
- 住所が入るべき欄に電話番号が入っている
- フリガナにカタカナとひらがなが混在している

ISO/IEC 25024（サービス品質の評価）

- サービスを提供するプロセスの品質をデータライフサイクルのステップごとに評価
- EC (European Commission) で検討されているインターオペラビリティのための品質評価モデル：iMAPS (Interoperability Maturity Assessment of a Public Service) も参考に評価項目を設定



ISO/IEC 25024:2015 Systems and software engineering -- Measurement of data qualityより翻訳

ISO/IEC 25024（サービス品質の評価）

3)データ統合

■複数データの統合時に、データ項目のマッチング、精度や単位の確認、コードのコンバージョンなどを正確に行う必要がある。データ統合時に品質の劣化を起こさない仕組みが整備できているかを評価する。

評価項目

■重複データの確認プロセスがあるか

■コード変換表など統合に関する情報を公開しているか

■統合元データの出所を明確にしているか

問題となる例

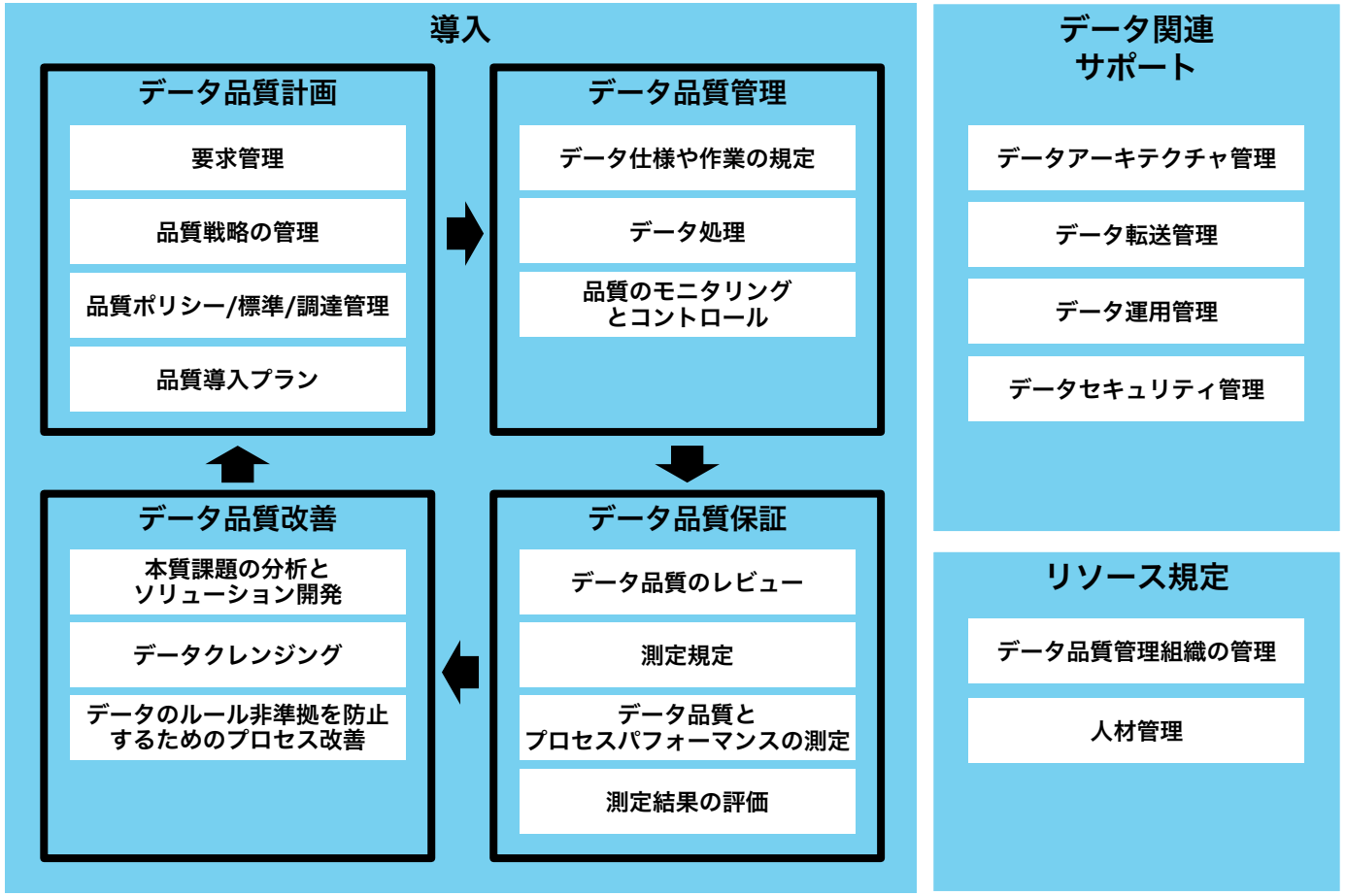
■有効数字の違う数字を単純に加算してしまう

■コード変換表を公開していないので、詳細分析ができない

■同一名の別法人のデータを結合してしまう

ISO/TS 8000-61 (データ管理プロセスの評価)

- データ品質やサービス品質を維持、向上していくためにはデータ管理プロセス及び運用体制が重要
- ISO/TS 8000のデータクオリティのプロセス管理の規格である part61 に沿って評価



Detailed structure of data quality management (ISO8000-61) より翻訳

ISO/TS 8000-61 (データ管理プロセスの評価)

1) データ品質計画

- データの品質を維持・管理していくためには、計画をたてて実行することが重要です。データ品質管理が計画的に行われているかを評価します。

評価項目

- (応用) データ品質に対する要求の管理をしているか
- (応用) データ品質を維持・向上させる戦略の管理ができているか
- データ品質ポリシー／標準／調達管理等のルール化ができているか
- データ品質導入プランを持っているか

問題となる例

- 品質の要求要件や基準が明確になっていないため、何から手を付けて良いのか分からない。

成熟度モデル方式による評価

アドホックレベル

- 独自の方法で実施されている
- 基本機能が実装されていない
- データ作成者による改善がなければ、データとして利用するのが困難

部分対応レベル

- 基本の条件が一部で実装されている
- 利用者側にてある程度の修復が可能であり、限定的な範囲でデータとして利用可能

基本レベル

- 実現すべき機能・レベル
- データとして利用可能な充分な品質を満たしている

サステイナブル・レベル

- 基本機能を継続的に提供し、フィードバックをかけている
- 継続性のあるデータ、コンピュータシステムとして統合が可能

成熟度モデル（Maturity Model）：
単純にできている/できていないで加点/減点する方式ではなく、
現在の自分の組織の立ち位置を確認し、
次のステップを目指せるようにするためのモデル

データ品質評価ツール

- ガイドライン付属の評価ツール
- Excelファイルで各項目の質問に対する実現レベルを答えるとチャートで可視化される
- データ関連サービスの課題を明確にするとともに、他サービスと比較することでベンチマークが可能
- 年に1回実施するなどで、その改善状況を公開するとともに、改善計画などを策定していくことが重要

https://cio.go.jp/guides#renkeimodel

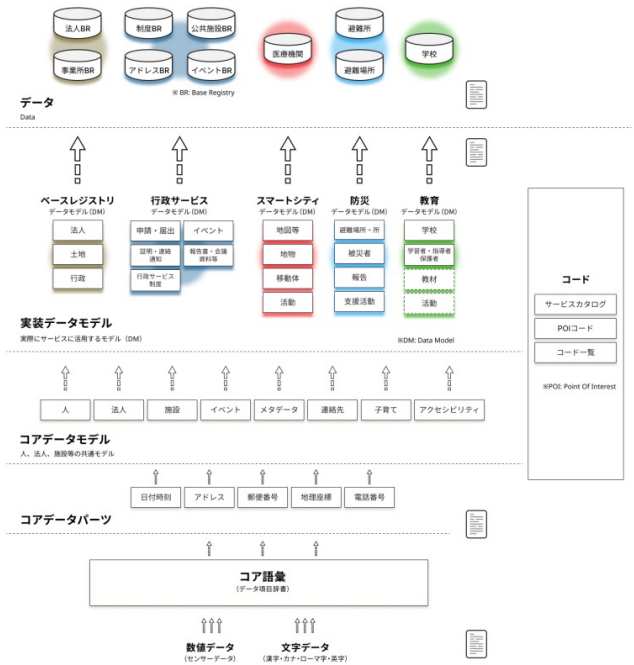
政府相互運用性フレームワーク（GIF）

デジタル庁は、デジタル・ガバメント推進標準ガイドラインのもと、データの利活用、連携がスムースに行える社会を実現するための技術的体系として、「政府相互運用性フレームワーク（Government Interoperability Framework）」（以下、GIFといいます。）を提供します。このフレームワークを利用してデータを整備することで、拡張性が高く、連携が容易なデータを設計することが可能となります。

2022年4月以降に意見受付フォームや研修教材をあわせて公開する予定です。

関連情報

政府相互運用性フレームワーク（GIF）を公開しました



小さなパーツから、組み合わせた実装モデルに展開し、標準化されたデータを整備していきます。

組織情報
政策
会議等
法令
採用
資料
申請・届出
調達情報
お知らせ

注目のトピック

新型コロナワクチン接種証明書アプリ
Visit Japan Webサービス

サイトポリシー
プライバシーポリシー
ウェブアクセシビリティ
コピーライトポリシー　SNS
ご意見・ご要望　サイトマップ

Global Site
プレスルーム

政府相互運用性フレームワーク（GIF）

相互運用性の高いデータの設計をサポートするための技術体系を公開（2022年4月）

出典：政府相互運用性フレームワーク（GIF）（令和4年4月 デジタル庁）
https://www.digital.go.jp/policies/data_strategy_government_interoperability_framework/

環境省 データ マネジメント ポリシー

データマネジメントの実施の一環としてオープンデータを位置付けて推進

【環境省のデータマネジメント実施内容】

【全省的なデータガバナンス（統制）】

全省データアーキテクト及び全省利活用業務推進者は、全省的なデータガバナンスに関する以下の取組を推進 ⇨ データマネジメントの課題解決＋各課室の政策目的実現支援

①環境データ利活用	マチュリティモデル（成熟度モデル）を用いたオープンデータ化の取組評価、環境データショーケース（仮）の公開やアイデア募集イベント、外部団体との連携によるニーズ発掘
②データ連携	省内保有データの全体像を把握し、連携対象システムの選定
③EBPM	データにアクセスし易いよう、省内データ一覧とアクセス方法を整備したデータアクセスガイド（仮）策定
④データ品質向上	各プロジェクトにおける、データ棚卸調査と品質改善計画の推進
⑤データ一元管理	省内保有データの全体像を把握し、一元管理の方針を策定
⑥データ標準化	未整備分野における標準化の推進、政府標準化ルールの確認と遵守
⑦データ利活用基盤構築	環境データショーケース、データカタログサイトの活用、Web-API公開、省内データ一元管理、DHW・BI基盤の構築
⑧省内データマネジメント	プロジェクト計画書、要件定義書等の雛形整備、各プロジェクトのデータベース設計書の収集によるオープンデータ化・データ連携・一元管理・標準化への活用
⑨人材教育・普及啓発	PJMOフォーラム等における施策の説明・事例共有、スキルや用語の説明、勉強会の実施等

【各課室におけるデータマネジメント（管理）】

各課室・PJMOはデータを利活用するプロジェクトの計画・整備・運営・状況把握といった一連の活動において、データマネジメントに関する以下の取組を実施 ⇨デジタル技術を活用した利用者中心のサービス・業務改革を推進するためのデータの整備・管理することで、改革を実現

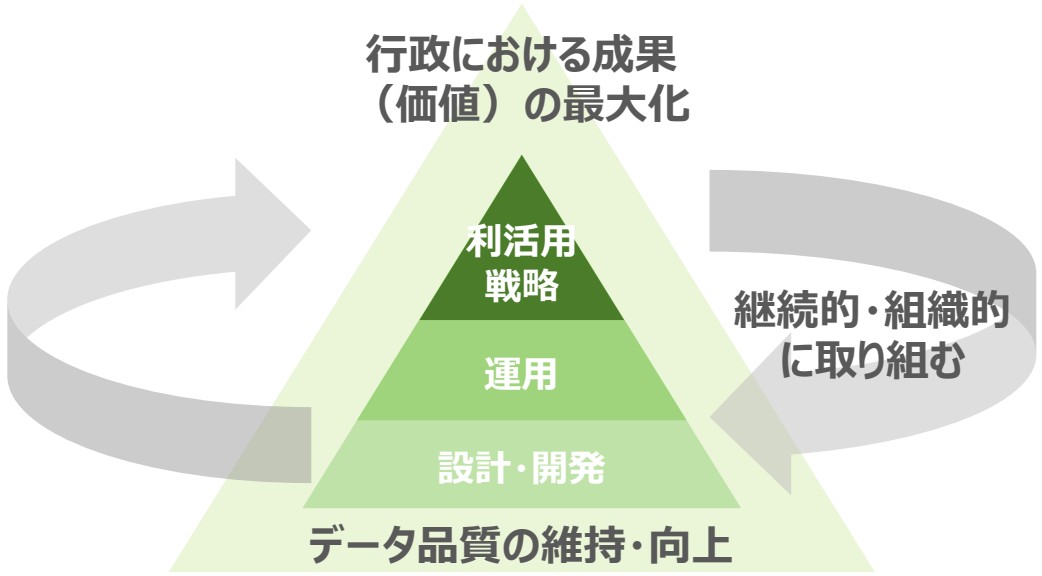
①オープンデータ化	オープンデータ化手順に沿って、方針やロードマップを策定し実施
②データ連携・一元管理・標準化	システム新規構築・更改のタイミングでデータ連携・一元管理・標準化の可能性を検討し、方針を策定
③データ品質改善	定期的なデータ品質の棚卸の実施、品質改善の検討
④状況報告	棚卸を通したPMOへの状況報告

2

出典：https://www.env.go.jp/press/109431.html

データマネジメント：
データを適切に維持・管理するための取り組み

- 業務においてデータを最大限に活かせる状態を作るための取り組み
- その状態を継続的に維持し、よりステップアップすることを目指す
- 目標に向かって活動が進められるように、組織全体の取り組みとして位置付けることが重要

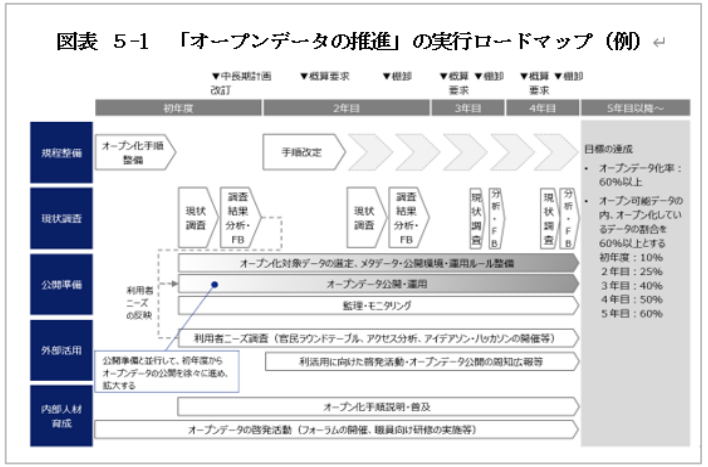


データマネジメント実践ガイドブック

- 具体的な作業を実施するための各種テンプレート（調査票サンプル）や作成例を提供
- 政府相互運用性フレームワーク（GIF）の「実践ガイドブック」のうちの1つ

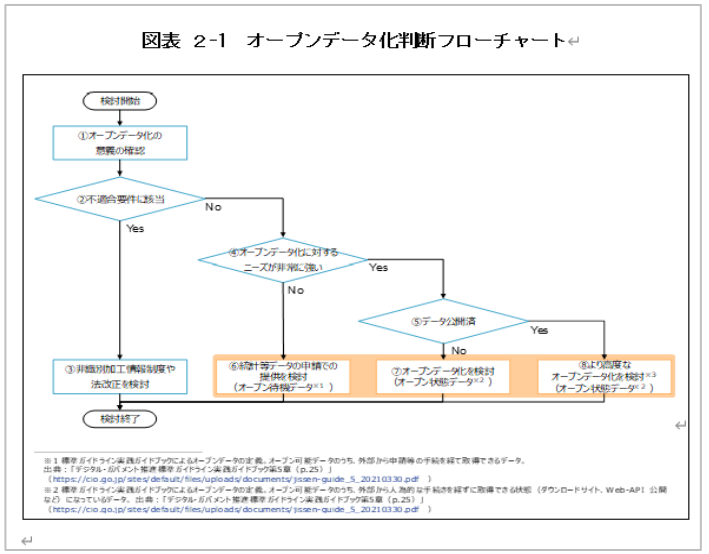
①導入編

データマネジメント実施計画の策定手順をガイド



②運用編

データマネジメントの４つの重点テーマ別に手順をガイド



出典： https://www.digital.go.jp/policies/data_strategy_government_interoperability_framework/

ガバメント・オープンデータにおける品質管理：今後の課題

- データマネジメントの重要性が局所的にしか理解されていない
- データの品質評価をワークフローに組み込む必要がある
- データエンジニアリングに詳しいわけではない、各府省のPMOやPJMOの方々に直接使って頂ける簡易なモデルを用意する必要がある
- 自己評価のための項目と外形評価のための項目を分ける必要がある

Link Data Now!!

LinkData.org

本資料に関するご質問・ご意見や、
データ活用研修・ワークショップ開催のご相談などは
こちらへお寄せください。

一般社団法人リンクデータ
代表理事 下山 紗代子
Email: support@linkdata.org