
データ連携基盤を通して提供される データの品質管理ガイドブック

【概要】

現在、スーパーシティ/スマートシティをはじめとする様々な地域で、データ連携基盤の整備が進められ、データの連携と利用によるサービスの提供が進んできています。一方で、データを活用したサービスの提供にあたっては、データに誤りがあると適切なサービス提供が行われないなど、元となるデータの品質を確保することが重要となります。

このような状況を踏まえ、本ガイドブックでは、データ連携基盤の整備主体、データ提供者、データ利用者が効率的、効果的にデータの品質管理を実施するために必要な役割分担や、データの品質評価の具体的手法に関する参考モデルを提示します。

【キーワード】

スーパーシティ/スマートシティ、データの品質、データの品質管理、データ連携基盤/都市 OS、データ連携基盤の整備主体、データ提供者、データ利用者

2023年9月

内閣府地方創生推進事務局

目次

1. はじめに	1
1.1. 背景と課題	1
1.2. 本ガイドブックが対象としている想定読者	2
1.3. 本ガイドブックの位置づけ	3
1.4. 本ガイドブックの全体像	4
2. データ連携基盤におけるデータの品質管理の考え方	6
2.1. データの品質とは	6
2.2. データの品質管理モデル	6
2.3. 3者のステークホルダーの役割と実施すべき事項	7
2.4. 責任分界点に関する基本的な考え方	11
3. データ連携前に実施すべき事項	12
3.1. データ連携基盤の品質管理のための環境の整備	12
3.2. データ提供者のデータの品質管理プロセスの評価	14
3.3. データ提供者及びデータ利用者からのデータ連携基盤への利用（接続）申請の受付	17
4. データ連携時に実施すべき事項	18
4.1. データの品質情報とデータの提供	18
4.2. データの品質情報の提示とデータの仲介	30
4.3. データの品質情報の確認とデータの利用	31
5. データ利用者による品質改善要求時に実施すべき事項	32
5.1. データ利用者による改善要求／要望の起案	32
5.2. データ連携基盤の整備主体による改善要求／要望の受付	33
5.3. データ提供者による改善要求／要望への対応	33
6. データ生成時の品質向上のための工夫	34
6.1. データ設計	34
6.2. データ収集	35
6.3. 外部データ取得	35
6.4. データ統合	35
A. 付録	37
A.1. 簡易評価ツールに関する操作説明	37
A.2. 仮想ユースケースにおけるデータの品質評価の適用	39
B. 用語説明	45
参考文献	48

1. はじめに

1.1. 背景と課題

センサー、ネットワーキング、ビッグデータ分析、人工知能（AI）などの最先端のデジタル技術により、都市の様々なデータを収集し、分析することが可能になりました。こうした様々なデータを交通管理、エネルギー管理、公共サービス等に利用することは、交通流の最適化や渋滞回避、効率的なエネルギー供給など、住民の生活の質の向上や改善につながります。

我が国でも、現在、スーパーシティ/スマートシティ（以下、スーパーシティ等という）をはじめとする様々な地域で、データ連携基盤の整備や、データ連携基盤を利用した様々なサービスの実装が進められています。

このようなサービスは住民の生活の質の向上に資する一方、その実施にあたって注意すべき点もあります。例えば、データ連携基盤を通して得られたデータを利用し、ドローン配送や自動運転等をサービスとして実装する場合、現在の状況とは異なっている、もしくは間違っている地理データを利用すると、運行中の事故のリスクが増大し、サービスの信頼性が低下するのみならず、かえって住民の安全や安心を脅かす可能性があります。

最先端のデジタル技術を利用して、このような課題を解決し、住民の安全や安心を確保しつつ生活の質の向上に資するサービスを提供するには、データ連携基盤から提供されるデータの正確性の確保など、データ連携基盤の利用における安全性及び信頼性の確保、すなわち“データの品質管理”が極めて重要となります。

なお、国においては、令和3年6月に閣議決定された「デジタル社会の実現に向けた重点計画」中の「包括的データ戦略」において、「ベースレジストリをはじめとする基盤となるデータの整備やオープンデータも含むデータの生成、利用、トラストの確保等を幅広く捉えた包括的なデータマネジメントを推進していく」と位置づけています。また、スーパーシティにおけるデータ連携基盤について、「国家戦略特別区域法及び構造改革特別区域法の一部を改正する法律」（令和5年法律第20号）により国家戦略特別区域法（平成25年法律第107号）の一部が改正され（令和5年9月1日施行）、国は、互換性の確保に加え、データ連携基盤の利用における安全性と信頼性の確保に関する情報提供等についてもデータ連携基盤を整備する者に対して援助を行うこととされたところです。

1.2. 本ガイドブックが対象としている想定読者

本ガイドブックが対象としている想定読者は、主に以下の3者のステークホルダーです。

- **データ連携基盤の整備主体**

データ連携基盤を整備し、データ提供者からデータ利用者（サービスの提供者）へ、データを仲介する主体を示します。

- **データ提供者**

自らの事業や観測活動によりデータを生成し、それらのデータを整理、加工して提供する者を示します。

- **データ利用者**

データ連携基盤を通して取得したデータを加工等せず利用する者及び取得したデータを利用（加工）してサービスを提供する者を示します。

なお、本ガイドブック中の用語については、各章で簡単に定義や説明を行っていますが、詳細な説明はB.用語説明を参照してください。

1.3. 本ガイドブックの位置づけ

スーパーシティ等におけるデータ連携基盤に関しては、これまで、「スマートシティリファレンスアーキテクチャ」（内閣府：令和5年8月）や「2020年度 データ連携基盤技術報告書」（内閣府：令和3年3月）等において、データ連携基盤に求められるブローカー、API等の機能やアーキテクチャが示されています。また、データの品質管理に関しては、デジタル庁の「政府相互運用性フレームワーク」（以下、GIFという）内において、データ提供者（データ生成者）がデータの品質を向上するための実践ガイドとして「データ品質管理ガイドブック」（デジタル庁：令和4年3月）が公表されています。

しかしながら、データ連携基盤を通して提供されるデータの品質管理においては、データ提供者のみならず、データ連携基盤の整備主体やデータ利用者も含めた複数の者がステークホルダーとして関与しており、これらの者が共通認識のもとデータの品質評価や確認を行うことにより、更なる品質の確保と向上を図ることが重要となります。

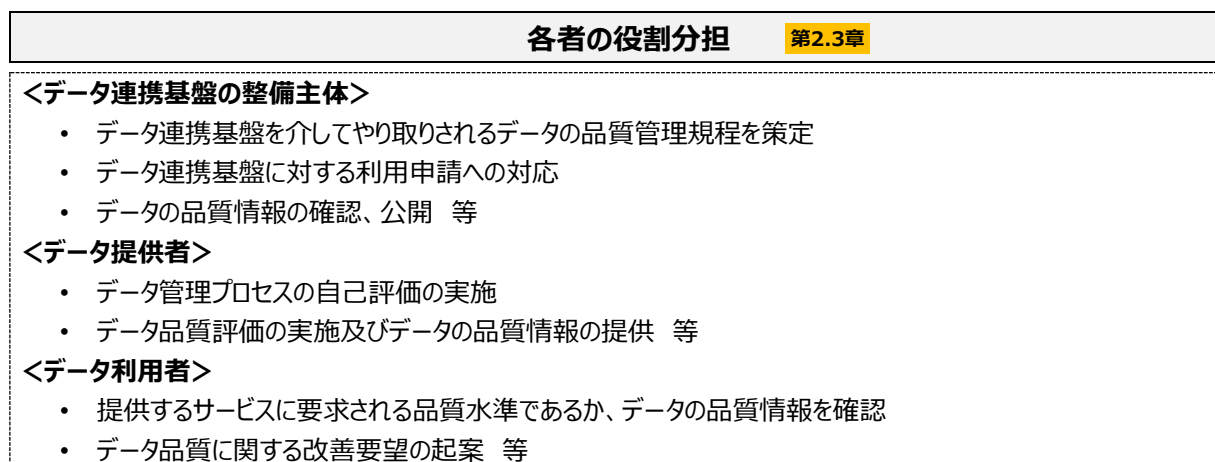
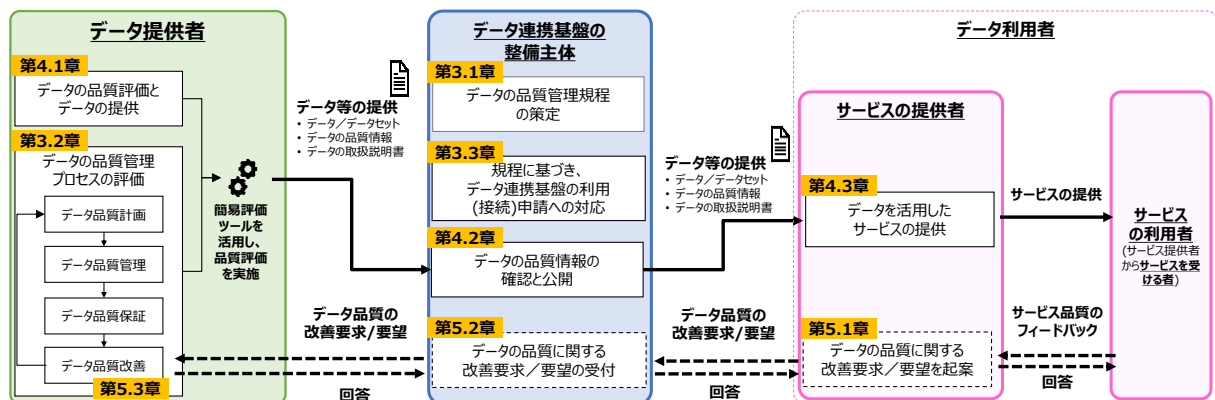
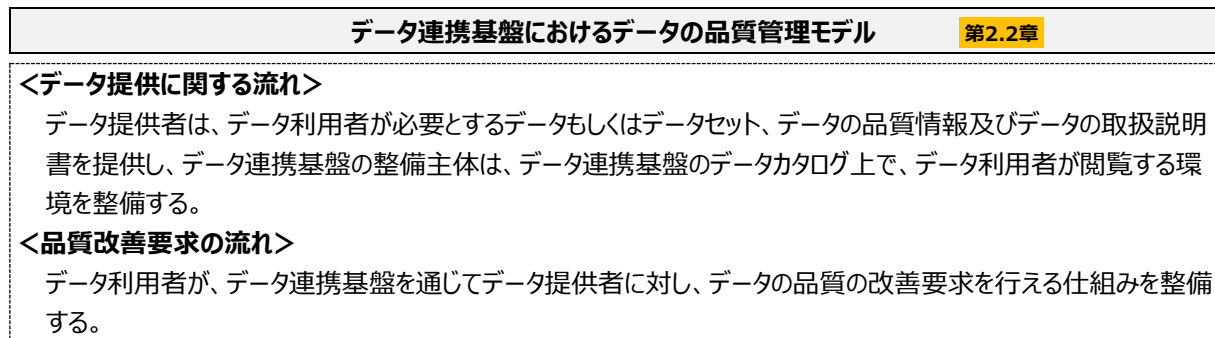
以上を踏まえ、本ガイドブックでは、「データ品質管理ガイドブック」をはじめ、データ連携基盤に関するこれまでの様々な知見を参考にしながら、スーパーシティ等のデータ連携基盤に関与するステークホルダーであるデータ連携基盤の整備主体、データ提供者、データ利用者の3者（以下、3者のステークホルダーという）が、データの品質管理を効率的、効果的に実施するために担うべき役割や、データ提供者以外の者がデータの品質を客観的に確認するための品質評価の方法について参考となるモデルを提示することで、データ連携基盤におけるデータの品質管理の参考となることを目的としています。

なお、本ガイドブックは、第1.2章で示した読者が、データの品質管理の手法を検討する中で、データ連携基盤を通して提供されるデータの品質管理について一般的に推奨される取組を記載したものです。このため、データの品質管理を実施する際は、3者のステークホルダーの関係性、取り扱うデータの内容、提供しようとするサービスの内容等に応じて、個別に検討が必要となります。

また、本ガイドブックでは、データ提供者が実施する評価作業（第3.2章の「データの品質管理プロセスの評価」と4.1.1節(1)の「データの品質評価」）において利用可能な「簡易評価ツール」を提供しています。なお、本ツールは、本ガイドブックで提示するデータの品質管理モデルにおいて、データ提供者が提供するデータについて評価を行い、必要に応じて改善することを目的としたものであり、データ提供者の自己評価に基づくものであることにも十分留意して、評価結果を参照ください。

1.4. 本ガイドブックの全体像

本ガイドブックでは、3者のステークホルダーが、効率的、効果的にデータの品質管理を実施するために必要な役割分担とデータの品質評価の手法等について提示しています。具体的には、第2章では、データの品質管理モデル等のデータの品質管理に関する基本的な考え方について説明し、第3章以降で、データ連携前（第3章）、データ連携時（第4章）、データ利用者による品質改善要求時（第5章）、データ生成時（第6章）の各フェーズにおいて、3者のステークホルダーが実施すべき事項について説明しています。なお、本ガイドブックの全体像と該当する章の番号は以下のとおりです。



簡易評価ツールは、以下の4つのシートから構成されています。

- データの品質管理プロセスを評価する「【A】データの品質管理プロセスの評価」
- 基礎的品質特性を評価する「【B】基礎的品質特性の評価」
- 付加的品質特性を評価する「【C1】付加的品質特性の評価(メタデータの記入)」
- 付加的品質特性の評価結果を表示する「【C2】付加的品質特性の評価(結果出力)」

簡易評価ツールの具体的な操作方法については、付録 A.1 を参照ください。

2. データ連携基盤におけるデータの品質管理の考え方

【概要】

第2章では、データの品質に関する基本的な概念を説明した上で、データ連携基盤におけるデータの品質管理モデルを示します。また、データの品質管理モデルに基づいて、3者のステークホルダーの役割と実施すべき事項、責任分界点に関する基本的な考え方について説明します。

2.1. データの品質とは

データの品質とは、データがある目的を満たすために、具備すべき特性を定義したものであり、データの状態（質）を表したものです。一般的にデータの品質が良いということは、データが正確で、かつ信頼できるデータであることを示します。一方、同じデータでも、データを利用する目的が異なれば、最新性或精度など、必要となるデータの特性も異なるため、データの品質評価は目的ごとに変化します。データの品質を評価するには、これらの点を考慮する必要があります。

本ガイドブックでは、スーパーシティ等のデータ連携基盤に関与する3者のステークホルダーが、データ連携基盤から提供されるデータの品質を共通の基準で評価するために、データの品質を測る上で重要な特性（正確性、最新性、精度等）と、各品質特性の評価方法について提示します。詳細は第4章で説明します。

2.2. データの品質管理モデル

本ガイドブックにおけるデータの品質管理モデルは、3者のステークホルダーが、データ連携基盤を通してデータの品質をどのように管理するかを概念的に示したものです（図2-1参照）。

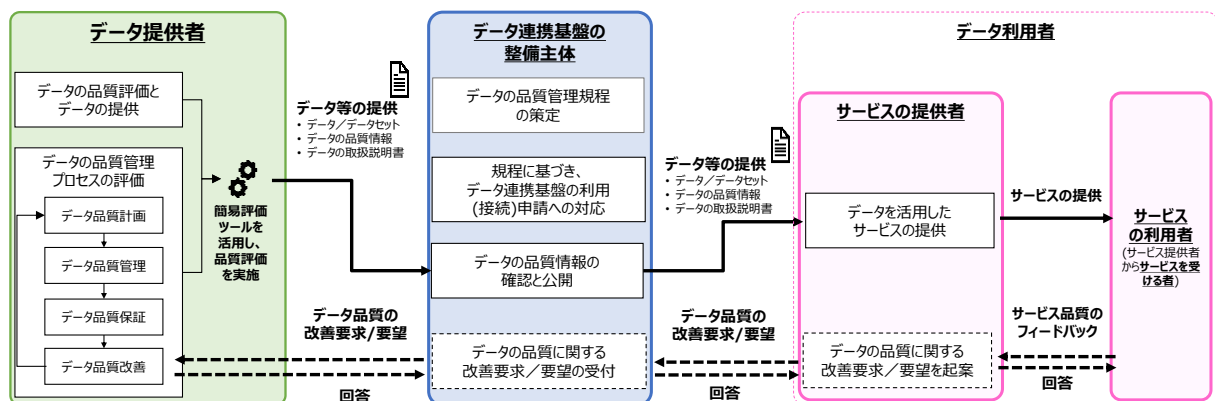


図 2-1 データ連携基盤におけるデータの品質管理モデル

品質管理されたデータを連携するために、3者のステークホルダーが実施すべき事項は、大きく分けて以下の3つとなります。

- データ連携基盤におけるデータの品質管理規程の策定など、品質管理のための環境の整備を起点とした「**データ連携前に実施すべき事項**」
- データ提供者によるデータの品質評価を起点とした「**データ連携時に実施すべき事項**」
- データ利用者（サービスの提供者）からのデータの品質改善要望／要求を起点とした「**データ利用者による品質改善要求時に実施すべき事項**」

3者のステークホルダーの役割と実施すべき事項について、概要は第2.3章で説明し、詳細は第3章～第5章で説明します。

2.3. 3者のステークホルダーの役割と実施すべき事項

2.3.1. 3者のステークホルダーの役割について

(1) データ連携基盤の整備主体

データ連携基盤の整備主体の役割は、データ提供者からデータ利用者（サービスの提供者）へデータを仲介することです。このため、データ連携基盤の整備主体は、主に以下の5つを実施する必要があります。

- **データの品質管理規程の策定**
データの品質管理規程として、データ提供者／データ利用者によるデータ連携基盤への利用（接続）の規程、データ連携基盤上でデータを公開する規程を策定します。
- **データ連携基盤の利用（接続）申請への対応**
データの品質管理規程に基づいて、データ提供者及びデータ利用者（サービスの提供者）によるデータ連携基盤の利用（接続）申請に対応します。
- **データの品質情報の確認と公開**
データ提供者から受領したデータもしくはデータセット¹、データの各品質特性の評価結果（以下、品質評価結果という）、データの取扱説明書²を確認し、品質評価結果及びメタデータ（以下、品質情報という）、データの取扱説明書等をデータ連携基盤のデータカタログ上でデータ利用者（サービスの提供者）にとって分かりやすい形で公開します。

¹ データセットとは、複数のデータから構成されるデータの集合を指します。

² データを扱う際に、そのデータの取り扱い方法や制約、保管、共有、セキュリティ対策などを文書化したもの。データの利用目的や利用方法、制約事項などの項目が含まれます。

- **データの品質に関する改善要求や要望の受付**

データ利用者（サービスの提供者）が起案したデータの品質に関する改善要求／要望を受け付け、内容を確認の上で、データ提供者へ対応を依頼する等の必要な対応を行います。

- **セキュリティ／運用対応**

不正アクセスやデータの改ざん等に備え、利用者のアクセス権限設定やログ等の監視等の必要な対応を行います。

(2) データ提供者

データ提供者の役割は、データを生成し、データ連携基盤へデータを提供することです。このため、データ提供者は、主に以下の2つを実施する必要があります。

- **データの提供**

データの品質を一定の基準に沿って評価し、自らが示したデータの品質に対して責任を負って、データ利用者（サービスの提供者）が必要とするデータもしくはデータセット、データの品質情報及びデータの取扱説明書をデータ連携基盤に提供します。

- **データの品質に関する改善要求／要望の受付と回答**

データ連携基盤を通じて、データ利用者（サービスの提供者）からのデータの品質に関する改善要求／要望を受け付け、これらに対応した結果をデータ連携基盤の整備主体に回答します。

なお、データ提供者には、自らの事業や観測活動などによりデータを生成し取得する、またはそれらのデータを整理、加工、保管、整備する「データ生成者³」としての側面と、他のデータ提供者から取得したデータを整理、加工、保管、整備する「データ流通支援事業者⁴」としての側面の2つの側面があります。

(3) データ利用者

データ利用者の役割は、データ提供者からデータ連携基盤を通して得たデータを利用することです。データ利用者には、「サービスの提供者」としての側面と「サービスの利用者」としての側面の2つの側面があり、以下では「サービスの提供者」と「サービスの利用者」について説明します。

- **サービスの提供者**

サービスの提供者の役割は、データ提供者からデータ連携基盤を通してデータの提供を受け、サービス、製品の開発／運用などに利用するほか、自らの事業に利用することです。サービスの提供者は、データをサービスに利用するために、データの収集及び利用の目的、必要なデータの種類、頻度、精度などを整理し、データに対する期待や必要条件を明確にします。その上で、

³ データ生成者の例としては、国の行政機関、地方公共団体、一般社団法人、民間企業などがあります。

⁴ データ流通支援事業者の例としては、データ共有事業者、パーソナルデータストア、情報銀行、データ処理事業者などがあります（参照：「データ取引市場運営事業者認定基準 D3_解説」、一般社団法人データ社会推進協議会）。

データ連携基盤の整備主体が提示するデータの品質情報をもとに、データの性質及び特徴が、自身の要求を満たすか調べ、データの取得要否を判断します。

また、データ利用者（サービスの提供者）は、提供するサービスについて一義的に責任を負う立場のため、サービスの利用者からの問い合わせ内容に対して回答と説明を行います。また、サービスの利用者からの改善要望等をもとに、データ提供者に対して、提供されるデータの改善要望を起案し、データ連携基盤の整備主体の設ける受付窓口連絡します。

- **サービスの利用者**

サービスの利用者は、サービスの提供者からサービスを受ける者（住民等）を示します。提供されたサービスに対して意見や改善要求／要望がある場合には、サービスの提供者に対して改善要求／要望を起案します。

2.3.2. 3者のステークホルダーの実施すべき事項について

本節では、データの品質管理モデルにおいて3者のステークホルダーが実施すべき事項について、以下の3点を説明します。

(1) データ連携前に実施すべき事項

- **データ連携基盤における品質管理のための環境の整備**

データ連携基盤の整備主体は、データ連携基盤上で取り扱われるデータの品質管理のため、以下の3つの対応を行います。

- ✓ データ連携基盤におけるデータマネジメント推進体制の中で、データの品質管理を実施する環境を整備します。なお、整備にあたっては GIF の「データマネジメント実践ガイドブック（導入編）」を参考にするとともに、提供データの品質確認等を実施する担当者はデータの品質管理に関する知識を有することが望ましいため、必要に応じて、政府等の教育サポート（総務省地域情報化アドバイザー制度⁵等）を受けることが推奨されます。
- ✓ データ連携基盤を利用することができるデータ提供者及びデータ利用者（サービスの提供者）の基準を明確化するとともに、データ連携基盤の利用（接続）申請を受け付ける窓口を構築します。
- ✓ データ利用者（サービスの提供者）からの改善要求／要望を受け付ける窓口を構築します。

- **データ提供者のデータの品質管理プロセスの評価**

データ提供者は、ISO/TS 8000 の「データ品質管理:プロセス参照モデル」の規格である Part 61⁶に基づいて作成した「簡易評価ツール」のシート「【A】データの品質管理プロセスの評価」を用いて、データの品質管理のプロセスを自己評価します。

⁵ 総務省地域情報化アドバイザー制度： https://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/manager.html

⁶ ISO/TS 8000-61： <https://www.iso.org/obp/ui/en/#iso:std:iso:8000:-61:ed-1:v1:en>

- **データ提供者及びデータ利用者の利用（接続）申請**

データ提供者及びデータ利用者（サービスの提供者）は、データ連携基盤の整備主体に対して、データ連携基盤の利用（接続）申請を行います。

(2) データ連携時に実施すべき事項

- **データの品質情報とデータの提供**

データ提供者は、以下の2つの対応を行います。

- ✓ 「簡易評価ツール」のシート「【B】基礎的品質特性の評価」、シート「【C1】付加的品質特性の評価(メタデータの記入)」及びシート「【C2】付加的品質特性の評価(結果出力)」を用いて、提供するデータの品質評価を行うとともに、データの品質情報の表示を行います。簡易評価ツールの具体的な操作方法については、付録 A.1 を参照してください。
- ✓ データ連携基盤に対して、データもしくはデータセット、データの品質情報及びデータの取扱説明書を提供します。

- **データの品質情報の提供によるデータの仲介**

データ連携基盤の整備主体は、データ提供者から受け取ったデータの品質情報等を、データ連携基盤においてデータカタログ情報の一部として、データ利用者（サービスの提供者）にとって見やすく、理解しやすい形で提供します。

- **データの品質情報の確認とデータの利用**

データ利用者（サービスの提供者）は、提供されたデータの内容及びデータ連携基盤の整備主体が提供するデータの品質情報から、自身の要求を満たすデータかを確認します。その上で、自身の要求を満たすと判断した場合は、データを取得し、自身のサービスに利用します。

(3) データ利用者による品質改善要求時に実施すべき事項

- **データの品質に関する改善要求／要望の起案**

データ利用者（サービスの提供者）は、データ連携基盤を通して取得したデータに対して、疑義や要望がある場合には、データの品質に係る内容を含め、改善要求／要望を起案し、データ連携基盤の整備主体の設ける受付窓口に連絡します。

- **データの品質に関する改善要求／要望の受付**

データ連携基盤の整備主体は、データ利用者（サービスの提供者）が起案した改善要求／要望を受け付け、内容を確認の上、データ提供者へ回答する等の対応を行います。

- **データの品質に関する改善要求／要望の対応**

データ提供者は、データ連携基盤を通して受領したデータ品質の改善要求／要望の内容を確認の上、回答を作成し、データ利用者（サービスの提供者）へデータ連携基盤を通して回答する等の対応を行います。

なお、以上の(1)～(3)の3つの実施すべき事項については、第3章から第5章までにおいて、それぞれ詳細に説明します。

2.4. 責任分界点に関する基本的な考え方

第 2.3 章で述べたように、提供するデータの品質に対して責任を負うのはデータ提供者であり、サービスの内容に対して責任を負うのはデータ利用者（サービスの提供者）です。ただし、以下のようなパターンも想定されるため、注意が必要です。

- **データ連携基盤上でサービスを提供する場合**

データ連携基盤の整備主体が、データ連携基盤上でサービスを提供する場合は、データ利用者（サービスの提供者）として取り扱うことが適切です。この場合、データ連携基盤の整備主体が、提供するサービスの内容に対する責任を負います。

- **データ提供者から提供されたデータの変換、加工等を行う場合**

データ連携基盤の整備主体が、データ提供者から提供されたデータの変換、加工等を行った上でデータ利用者（サービスの提供者）にデータを提供する場合は、データ提供者として取り扱うことが適切です。この場合、データ連携基盤の整備主体がデータの品質に対する責任を負いますが、データ連携基盤の整備主体とデータ提供者の責任分界点が不明確となることから、以下の「追跡可能性の確保」、「原本性の確保」といった対策を実施することが有効です。

- ✓ **追跡可能性の確保**

追跡可能性とは、「特定の利用状況において、データへのアクセス及びデータに実施された変更の監査証跡を提供する属性を、当該データがどの程度有しているか」を示します（ISO/IEC 25012⁷）。追跡可能性を確保することで、誰が最後にデータの加工等を実施したのか明確になり、品質に関して責任を負うべき者が明確になります。追跡可能性を確保するためには、一般的に以下のような対応が講じられます。

- データ連携基盤を介してやり取りされるデータのアクセスログやシステムログを取得、保管する。
- データセット内にプロビナンス情報（更新履歴）を追加し、ステークホルダーがデータの加工等を行った際には必ず記載する。

- ✓ **原本性の確保**

原本性とは、そのデータが複製されたものでなく、最初に作成された「原本」であることを証明することを示します。例えば、データ連携基盤の整備主体がデータの原本性を確保することにより、データ利用者（サービスの提供者）からクレーム等が発生した際であっても、自身が加工や改ざん等を行っていないことを証明することができます。なお、原本性を保証するための一般的な方法は、データ連携基盤の整備主体もしくはデータ利用者（サービスの提供者）が持つデータのパリティビット⁸と、データ提供者が持つデータのパリティビットとが一致するかを確認し、データに改ざんがないかを検証することが挙げられます。

⁷ ISO/IEC25012 : <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:25012:ed-1:v1:en>

⁸ パリティビットとは、データの伝送や記録の際に生じる誤りを検知できるように算出、付加される符号の一つで、ビット列中に含まれる「1」の数が偶数か奇数かを表すものです。

3. データ連携前に実施すべき事項

【概要】

第3章では、2.3.2 節（1）で示した「データ連携前に実施すべき事項」について、データ連携基盤の品質管理のための環境の整備（第3.1章）や、データ提供者のデータの品質管理プロセスの評価（第3.2章）、データ提供者及びデータ利用者からのデータ連携基盤への利用（接続）申請の受付（第3.3章）を説明します。

データ連携基盤の環境整備とデータ提供者及びデータ利用者のデータ連携基盤への登録の流れを図3-1に示します。

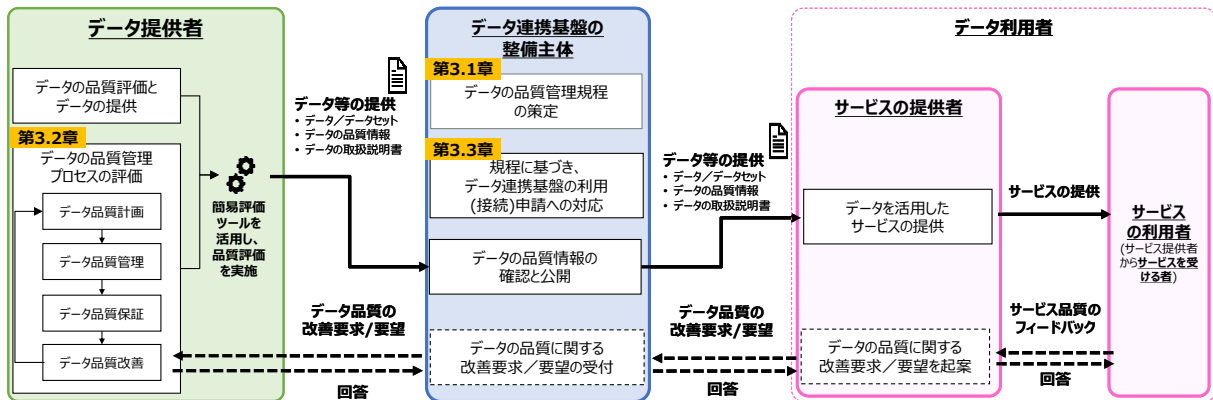


図 3-1 データ連携基盤の環境整備とデータ提供者及びデータ利用者のデータ連携基盤への登録の流れ

3.1. データ連携基盤の品質管理のための環境の整備

3.1.1. データ連携基盤の整備主体の実施作業

データ連携基盤の整備主体は、データの品質管理規程の策定、関連ステークホルダーへの浸透及びその見直し、改善を行います。

(1) データの品質管理規程の策定

データ連携基盤の整備主体は、データの品質管理規程として、データ提供者/データ利用者によるデータ連携基盤への利用（接続）の規程、データ連携基盤上でデータを公開する規程を策定する必要があります。

以下これらの規程について説明します。

- **データ提供者／データ利用者によるデータ連携基盤への利用（接続）の規程**

データを提供しようとする者からデータ連携基盤への利用（接続）申請が行われた場合には、データ連携基盤の整備主体から申請者に対し、付録 A.1.1 で示す「簡易評価ツール」のシート「【A】データの品質管理プロセスの評価」を要求し、申請者の組織内にデータの品質管理プロセスが存在するか確認します。

また、データ連携時において、付録 A.1.2 で示す「簡易評価ツール」のシート「【B】基礎的品質特性の評価」を要求し、データ連携基盤に対して、データもしくはデータセットとともに、データの品質情報、データの取扱説明書を提供することを確認します。

データを利用しようとする者からデータ連携基盤への利用（接続）申請が行われた場合には、データ連携基盤の整備主体から申請者に対し、提供データの利用条件への承諾を要求し、データ提供者の承諾のない限り、提供データの目的外利用や第三者への開示、提供等を行わないことを確認します。

また、提供データの取り扱いに対する安全対策を要求し、提供データに関する不正アクセス又は情報の紛失、破壊、改ざん、漏洩等のリスクに対する合理的な安全対策等が存在することを確認します。

- **データ連携基盤上でデータを公開する規程**

データ連携基盤の整備主体は、データ提供者から、データもしくはデータセット、データの品質情報及びデータの取扱説明書を受け取り、その内容を確認し、受理したデータ及び品質情報に問題がない場合はデータを公開します。

(2) 関連ステークホルダーへの規程の浸透

データの品質管理規程を策定した後は、策定した規程が関連ステークホルダーにおいて遵守されるよう、浸透させる必要があります。具体的には、データの品質管理に関して関連ステークホルダー間の責任範囲を明確化するために、ステークホルダー間において責任範囲、免責事項について合意し、契約や規約でこれを明確に定めることが考えられます。

(3) ガバナンス維持のための取組

データの品質管理を継続的に維持、改善するためには、PDCA サイクルを回し、データの品質管理規程の見直しを継続的に実施することが重要です。また、データの品質管理の知見を保有する人材の育成や、適切な配置など、人的リソースへの投資もガバナンスを維持するためには重要となります。

(4) 改善要求と改善要望の受付体制の構築

データ連携基盤の整備主体がデータ利用者（サービスの提供者）の起案したデータ品質の改善要求／要望の受付のために、データ連携基盤のデータカタログ上に、回答フォームを設置する等、データ品質に関する要望を受け付けることができる仕組みを構築することが重要です。

3.2. データ提供者のデータの品質管理プロセスの評価

3.2.1. データ提供者の実施作業

データの品質を維持し、向上させていくためには、データ提供者がデータの品質管理プロセス及び運用体制を確立することが重要となります。また、データの品質管理において、継続的に PDCA を実施することで、データ品質の改善を図ることが重要となります。

データ提供者が、データ連携基盤にデータを提供するためには、データ提供者のデータの品質管理プロセスが一定水準に達していることを、データ連携基盤の整備主体に対して示す必要があります。このため、データ提供者は、ISO/TS 8000 の「データ品質管理:プロセス参照モデル」の規格である Part 61 に沿って、組織におけるデータの品質管理プロセスを自己評価することが必要です。

本節では、データの品質管理プロセスの評価に関する「基本的な考え方」と「具体的なデータの品質評価の観点」について説明します。

(1) 基本的な考え方

本ガイドブックでは、データの品質管理プロセスの評価方法として、CMMI Institute のデータ管理成熟度モデル⁹を参考に、各項目の評価に成熟度モデル（マチュリティモデル）を採用しています。成熟度モデル（マチュリティモデル）は、データの品質管理プロセスをその成熟度のレベルにより測る評価方法です。具体的には、品質管理に関する計画の策定や保証の実施などの評価項目について、達成度を以下の 5 つの成熟度のレベルで評価します。

低 ----- 高	【レベル 1】： 設計されていない段階 データの品質管理プロセスが設計されていない。
	【レベル 2】： 設計されているが、実施されていない段階 データの品質管理プロセスが設計されているが、実施されていない。
	【レベル 3】： 一部実施されている段階 データの品質管理プロセスが設計され、一部実施されている。
	【レベル 4】： すべて実施されている段階 データの品質管理プロセスが設計され、実施されている。
	【レベル 5】： 改善サイクルを含めてすべて実施されている段階 データの品質管理プロセスが設計され、改善サイクルが回せている。

⁹データ管理成熟度モデル： <https://stage.cmmiinstitute.com/>

以上の考え方に基づき、データ提供者は自身の品質管理プロセスを自己評価することで、品質管理体制に関する現状を把握し、講じるべき対策の検討を行うことができます。また、データ連携基盤の整備主体やデータ提供者は、データの品質管理プロセスの評価結果を確認することで、データ提供者の品質管理プロセスが不十分でないかを確認することができます。

(2) 具体的なデータの品質評価の観点

データの品質管理プロセスの評価は、(1)で説明した成熟度モデルを利用して行います。データの品質管理プロセスは、図 3-2 に示すとおり、データ品質計画、データ品質管理、データ品質保証、データ品質改善、データ関連サポート、リソース提供といった 6 つのプロセスから構成されています。データ品質計画からデータ品質改善までの 4 つは継続的に実施するプロセスであり、データ関連サポートとリソース提供の 2 つは、必要に応じて実施するプロセスです。

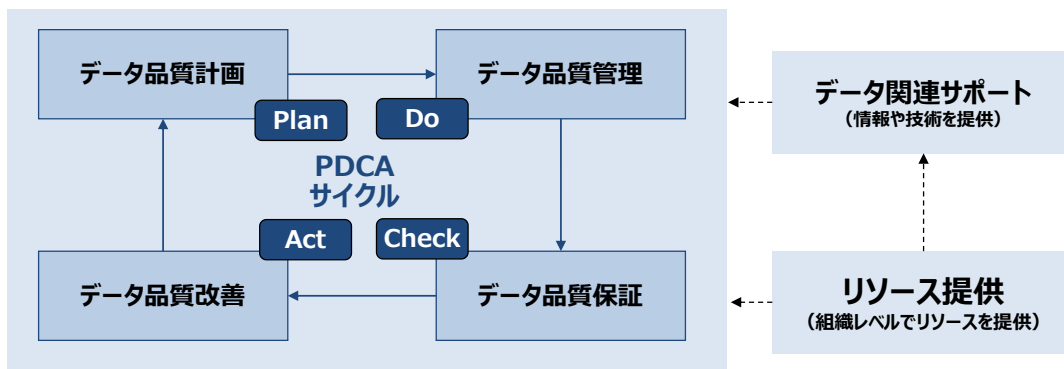


図 3-2 データの品質管理プロセスの構造

ここでは、6 つのデータの品質管理プロセスについて、品質評価の観点からそれぞれ説明します。

① データ品質計画

データの品質を維持、向上させていくために、データの要件とデータの品質の目標を確立し、組織全体で一貫したデータ品質管理を行うための手順をルール化するなど、実現に向けた実装計画が立案されているかを確認し、評価します。

② データ品質管理

①で立案したデータ品質計画に基づいて、データの生成及び更新の作業が実施されていることを管理し、目標とするデータの品質に届かない場合にデータの修正などの対応がされているかを確認し、評価します。

③ データ品質保証

生成されたデータの品質が目標とする品質を満たしているかどうかを測定し、満たせていない場合にはその原因を特定するプロセスが定期的実施されているかを確認し、評価します。

④ データ品質改善

③で特定された品質が満たされていない原因について、根本的な原因の特定と再発防止策を実施するとともに、目標とするデータの品質の達成に向けて自動化ツール等で修正する方法が確立されているかを確認し、評価します。

⑤ データ関連サポート

データの品質管理プロセス全体を支援するためのデータ関連のソフトウェア（ツール）やデータセキュリティの方針などが維持、管理されているかを確認し、評価します。

⑥ リソース提供

組織単位でデータの品質管理プロセスをサポートするための適切な資格を持つ人材の確保や責任者の配置、組織全体でのデータの品質管理に関する知識やスキルの共有や再利用がされているかを確認し、評価します。

上記6つのプロセスの評価は、簡易評価ツールのシート「【A】データの品質管理プロセスの評価」を利用して実施してください¹⁰。具体的な操作方法については、付録 A.1.1 を参照してください。簡易評価ツールを用いた評価を行うことで、図 3-3 に示すとおり、データの品質管理プロセスの最終評価結果として成熟度のレベルが表示されます。

【A】データの品質管理プロセスの評価			データの品質管理プロセスの最終評価結果 ^{*1}			
データ提供者名:	XXX株式会社		成熟度レベル	説明		
評価者:	XXX		成熟度レベル4	すべて実施されている段階		
データの品質管理プロセスの評価			*1 最終評価結果は、各プロセスの成熟度レベルの平均値から算出しています。			
#	プロセス名	実施作業名	評価項目	評価の記入欄		成熟度レベル
				評価	評価理由	
1		データ品質戦略・方針の策定	組織全体でデータの品質を向上させるためのビジョンや中長期目標、データの品質管理に関する基本的な考え方を示した戦略や方針が策定されているか。 <評価指標> 成熟度レベル5：戦略や方針が策定されており、定期的に戦略等の改善・見直しが行われている状態 成熟度レベル4：戦略や方針が策定しているが、改善・見直しが行っていない状態 成熟度レベル3：不十分ではあるが、一部策定している状態 成熟度レベル2：検討段階 成熟度レベル1：未対応	ポルタワンから選択してください	選択の理由などを記載	3
2	データ品質計画 (Plan)	実施計画の策定	戦略や方針の実現に向けて、具体的なアクションプラン等の実施計画が策定されているか。 <評価指標> 成熟度レベル5：実施計画が策定されており、定期的に計画の改善・見直しが行われている状態 成熟度レベル4：実施計画は策定しているが、改善・見直しが行っていない状態 成熟度レベル3：不十分ではあるが、一部策定している状態 成熟度レベル2：検討段階 成熟度レベル1：未対応			
3		要求/要望の受付体制	データの利用者によるデータの品質に関する改善要求/要望を適切に受け付け、データ品質計画に組み込む等の管理・対応体制が確立されているか。 <評価指標> 成熟度レベル5：受付体制が整備されており、定期的に受付体制の改善・見直しが行われている状態 成熟度レベル4：受付体制を整備しているが、改善・見直しが行っていない状態 成熟度レベル3：不十分ではあるが、一部体制を整備している状態 成熟度レベル2：検討段階 成熟度レベル1：未対応			

図 3-3 データの品質管理プロセスの最終評価（右上赤枠部分）

¹⁰ 簡易評価ツールの評価項目は、デジタル庁発行の「データ品質管理ガイドブック」を参考にし、より評価基準が明確になるように、各プロセスの実施作業単位で評価項目を具体化しています。

3.3. データ提供者及びデータ利用者からのデータ連携基盤への利用（接続）申請の受付

データ連携基盤の整備主体は、データ連携基盤にデータを提供しようとする者及びデータ連携基盤を通してデータを利用しようとする者からの利用（接続）申請を受け付けます。

(1) データ提供者の利用（接続）申請の受付

データ連携基盤の整備主体は、データを提供しようとする者から、データ連携基盤の利用（接続）申請を受け付けます。データを提供しようとする者は、申請書にデータの品質管理プロセスの評価結果を添付して申請します。

データ連携基盤の整備主体は、申請の受付後、申請の内容及びデータを提供しようとする者が実施したデータの品質管理プロセスの評価結果をデータの品質管理規程に基づき確認し、確認の結果、組織内にプロセスが存在するデータ提供者にはデータ連携基盤への接続を許可し、データ提供者として登録します。登録の可否にかかわらず、確認の結果を申請者に通知します。

接続を許可した場合であっても、データの品質管理プロセスが継続的に実施されていることを定期的に確認することが重要です。継続的にプロセスが実施されていない場合は、データ連携基盤との接続を解除する等の措置を講じる必要があります。

(2) データ利用者の利用（接続）申請の受付

データ連携基盤の整備主体は、データを利用しようとする者から、データ連携基盤の利用（接続）申請を受け付けます。申請の受付後、申請の内容をデータの品質管理規程に基づき確認し、基準を満たしていればデータ利用者（サービスの提供者）として登録します。登録の可否にかかわらず、確認の結果を申請者に通知します。

データを提供しようとする者及びデータを利用しようとする者は、データ連携基盤の整備主体から確認結果の通知を受け取り、データ連携基盤におけるデータ提供者又はデータ利用者（サービスの提供者）として登録されたことを確認します。

4. データ連携時に実施すべき事項

【概要】

第4章では、2.3.2節(2)で示した「データ連携時に実施すべき事項」に関して、データの品質情報とデータの提供（第4.1章）、データの品質情報の提示とデータの仲介（第4.2章）、データの品質情報の確認とデータの利用（第4.3章）を説明します。

品質管理されたデータ連携の流れを図4-1に示します。

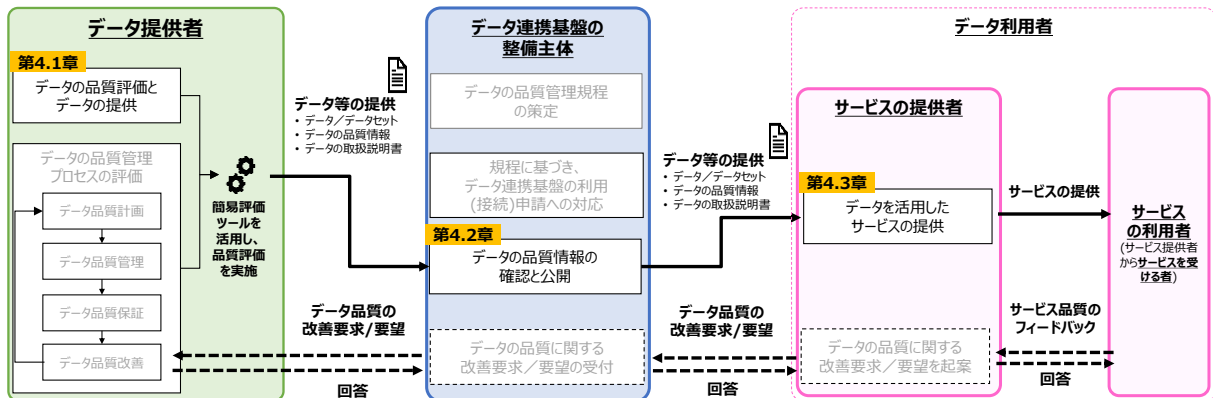


図 4-1 品質管理されたデータ連携の流れ

4.1. データの品質情報とデータの提供

4.1.1. データ提供者の実施作業

データ提供者は、第3章のデータの品質管理規程に基づき、提供しようとするデータについて、その品質に関する特性（以下、品質特性という）ごとに評価を実施し、データの品質情報の作成を行います。その後、データ連携基盤に対して、データもしくはデータセットとともに、データの品質情報及び、データの取扱説明書を提供します。

以下の(1)では、まず、評価対象となるデータの品質特性について説明します。また、(2)及び(3)では、データ提供者による各品質特性の評価の方法について詳細を説明します。

なお、「簡易評価ツール」のシート「【B】基礎的品質特性の評価」、シート「【C1】付加的品質特性の評価(メタデータの記入)」及び「【C2】付加的品質特性の評価(結果出力)」を用いた具体的な操作方法については、付録 A.1.2 を参照してください。

(1) データの品質評価

データの品質については、第 2.1 章に記載のとおり、データが正確で、かつ信頼があるデータあるかを確認する必要があります。正確性など、どのような利用シーンにおいても一定以上の高い品質が要求される基礎的な要素（特性）のみならず、最新性や精度など、利用シーンによって要求される水準が異なる付加的な要素（特性）も含めて考慮する必要があります。

具体的には、これらの要素（特性）を以下の「基礎的品質特性」と「付加的品質特性」と分類した上で、それぞれの評価方法を説明します。

● 基礎的品質特性

データの利用にあたっては、その利用シーンによらず、データが信頼できることが基本となります。信頼できるデータとは、正確で、完全であり、かつ、一貫性を持つデータを示します。本ガイドブックでは、これらの「正確性」、「完全性」、「一貫性」の3つの品質特性を「基礎的品質特性」と定義します。

データの品質を保証するためには、データ提供者は自身が作成（生成）したデータについて、妥当性検査（バリデーションチェック）ツールを利用して、正確性、完全性及び一貫性を保証する必要があります。これら基礎的品質特性の保証されていないデータが提供された場合、データ利用者（サービスの提供者）において本来不要であったデータのクレンジング処理や内容の精査といった作業が必要となり、クレーム等につながる可能性があります。

● 付加的品質特性

データ利用者（サービスの提供者）の立場からすると、データ提供者から提供されたデータが、データ利用者（サービスの提供者）自身の提供しようとするサービスに利用可能かどうかを判断するためには、データの基礎的品質特性のほか、その最新性や精度等のデータの品質情報が重要となります。本ガイドブックでは、これら基礎的品質特性以外の品質特性を「付加的品質特性」と定義します。

データ提供者が、共通の尺度や基準により付加的品質特性に関する情報を提供することで、提供するデータの価値を適切に表現でき、データ利用者（サービスの提供者）は表示された情報をもとに、サービス提供に必要なデータのニーズを満たしているか判断できるようになります。基礎的品質特性のみならず付加的品質特性についても、データ提供者が誤った情報を提供した場合、データ利用者（サービスの提供者）からのクレーム等につながる可能性があります。

データ品質の国際標準である ISO/IEC 25012 においては、15 のデータの品質特性が定義されていますが、これらの品質特性を上記の「基礎的品質特性」と「付加的品質特性」に分類すると、品質特性の正確性、完全性及び一貫性（表 4-1 の #1～3）が「基礎的品質特性」に、これら以外の品質特性（表 4-1 の #4～15）が「付加的品質特性」に整理することができます。

表 4-1 データの品質特性と分類

#	品質特性	品質特性の内容	分類
1	正確性 (Accuracy)	データに誤りがなく、信頼性がある状態か	基礎的品質特性
2	完全性 (Completeness)	必要な項目が網羅されているか、データに誤りや欠落がなく完全な状態であるか	
3	一貫性 (Consistency)	データに矛盾はないか、データ間で整合性が取れている状態か	
4	信憑性 (Credibility)	出所や更新日が明示されているか	付加的品質特性
5	最新性 (Correctness)	更新期間や最終更新日が明示されているか	
6	アクセシビリティ (Accessibility)	環境依存文字等が使用されていないことが明示されているか	
7	標準適合性 (Compliance)	標準に準拠したデータ形式であることが明示されているか	
8	機密性 (Confidentiality)	アクセス制限等の対策が講じられていることが明示されているか	
9	効率性 (Efficiency)	効率化のため、コードの割り当て等を行っていることが明示されているか	
10	精度 (Precision)	データの精度について明示されているか	
11	追跡可能性 (Traceability)	データの変更履歴（変更者／変更日等）が明示されているか	
12	理解性 (Understandability)	データ全体や項目ごとにメタデータが明示されているか	
13	可用性 (Availability)	利用時間帯等について明示されているか	
14	移植性 (Portability)	標準的な形式でデータをエクスポートできることが明示されているか	
15	回復性 (Recoverability)	バックアップ等の障害時の対応方針が明示されているか	

(2) 基礎的品質特性の評価

① 基礎的品質特性の評価方法について

データ提供者は、提供しようとするデータの全体又は一部を対象に、データの基礎的品質特性を、妥当性検査（バリデーションチェック）ツールを用いて確認します。チェックツールを用意できない場合には、必要に応じて Microsoft Excel 等の表計算ソフトのソート機能などを利用することによって、異常値や外れ値を見つけやすくなります。これらの方法でデータ本体に「誤りがないか、漏れがないか、矛盾がないか」等を確認することにより、基礎的品質特性の評価を行います。

なお、バリデーションのチェックを実施する際には、「正確性」、「完全性」、「一貫性」の3つの基礎的品質特性に応じて、次のような観点から確認をします。各品質特性の確認項目の詳細については、「簡易評価ツール」のシート「【B】基礎的品質特性の評価」を参照ください。

- データの書式／値の確認（正確性の確認）
入力されたデータの書式や値に誤りがないかを確認する。
【確認項目の例】
 - ・ 本来はカタカナで記入すべき項目に、平仮名が記入されていないか
 - ・ 本来は日時を入力する項目に、氏名が入っていないか 等
- データの網羅性の確認（完全性の確認）
必要な項目が網羅され、必須項目に空白が含まれていないかを確認する。
【確認項目の例】
 - ・ 購買データで重要となる、購入日、購入金額、購入場所等の項目に空白が発生していないか
 - ・ 機器の不具合等により、センサーで取得している項目に空白が発生していないか 等
- データの整合性の確認（一貫性の確認）
データセット内でデータ間／項目間に矛盾がないかを確認する。
【確認項目の例】
 - ・ 合計値や平均値等の数式に矛盾や誤りはないか
 - ・ 法人名と法人番号が一致しているか 等

例えば、国土交通省の PLATEAU では CityGML に関するチェックツール¹¹を提供しています。各分野で提供データに関連する既存チェックツールがある場合は、既存チェックツールの利用が推奨されます。

¹¹ Project-PLATEAU/CityGML-validation-function :
<https://github.com/Project-PLATEAU/CityGML-validation-function>

また、データの基礎的品質特性の評価は、表 4-2 に示すとおり、評価対象としたデータがすべてのデータ又は一部のデータであるかという観点と、評価・改善の実施有無により評価します。

表 4-2 「3 つの品質特性」に対する評価基準

品質評価の対象データ 品質評価の実施有無	すべてのデータを対象		一部のデータを対象	
	実施	品質改善済	品質改善未対応	品質改善済
A		C	B	C
未実施	品質評価未実施			
	C			

② 基礎的品質特性の評価結果について

データの基礎的品質特性の最終評価は、表 4-3 に示すとおり、3 つの基礎的品質特性の中で最も低い評価結果に対応する形で決定します。

表 4-3 データの品質（基礎的品質特性）の最終評価の基準

最終評価	最終評価の基準
A	3 つの品質特性すべてが A 評価 例) 正確性 : A、完全性 : A、一貫性 : A
B	3 つの品質特性すべてが B 評価以上 例) 正確性 : A、完全性 : A、一貫性 : B
C	いずれかの品質特性に C 評価が含まれている。 例) 正確性 : A、完全性 : B、一貫性 : C

(3) 付加的品質特性の評価

① 付加的品質特性の評価方法について

データ提供者は、表 4-4 で示す付加的品質特性と関連付けされたメタデータの項目（以下、メタデータ項目という）の実体（値）を記入し、「簡易評価ツール」のシート「【C1】付加的品質特性の評価」を活用し、メタデータの記入率を評価します。この際、データ提供者は提供するデータがどのような利用シーンで使われるかを想定し、その利用シーンで重要と想定される品質特性に関するメタデータの記入を充実させることが重要となります。

表 4-4 メタデータ項目の実体（値）に記入する具体的な内容について¹²

品質特性	必須区分*	メタデータ項目	メタデータ項目の実体（値）
信憑性	◎	情報の管理主体者	情報の管理主体（組織又は個人）の名称を記入する。
	○	情報の管理主体番号	情報の管理主体（組織）の法人番号があれば記入する。
	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	データセットの提供者	データセットの提供者／作成者（組織又は個人）の名称を記入する。
	○	データセットの提供者番号	データセットの提供者／作成者（組織）の法人番号があれば記入する。
	○	データセットの更新日 又は修正日	データセットが更新 又は 修正された日付を年月日で記入する。 記入形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日）
最新性	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	データセットの作成日	データセットが作成された日付を年月日で記入する。 記入形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日）
	○	データセットの発行日	データセットが発行された日付を年月日で記入する。 記入形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日）

¹² メタデータの項目や必須区分については、GIF 内の「メタデータ導入実践ガイドブック」及び「2020 年度データ連携基盤技術報告書」、「DSA データカタログ作成ガイドライン」に記載されているメタデータの項目をもとに作成しています。また、メタデータの項目とデータの品質情報に関する品質特性の紐づけについては、ISO/IEC 25012 に定義されているデータの品質特性の説明を参考にしています。

	○	データセットの更新日 又は 修正日	データセットが更新 又は 修正された日付を年月日で 記入する。記入形式は以下のとおりとする。 ・YYYY-MM-DD (西暦-月-日)
	○	バージョン	データセットのバージョン情報を記入する。
	○	データセットの対象期間	データセットの対象期間開始日及び 終了日を記入す る。記入形式は以下のとおりとする。ただし、明確に年月 日で示せない場合は、自由記述とする。 ・開始年月日/終了年月日： YYYY-MM-DD/YYYY-MM-DD ※YYYY-MM-DD (西暦-月-日)
	○	データセットの対象期間の 分解能	データセットの対象期間を記入した場合、分解能の情報 がある場合に記入する。 分解能の情報としては、データが時系列データである場 合に、データ項目間の時間差を記入する。あるいは、時 系列データ以外の種類のデータの場合、データ項目の時 間間隔の最小のものを代表値として記入する。 なお、複数の分解能の情報がある場合、サマリとして参 照される値 (例、最小の時間差) を記入する。 【例】3 か月、3 時間、30 分、1 分 30 秒など
	○	データセットの更新頻度	データセットの更新頻度について記入する。 定期的に更新する場合、年/月/週/日あたりの回 数を記入する。 不定期の場合は、不定期と記入する。
アクセシビ リティ	◎	文字	データセット中のデータ記述に使われている文字セット を記入する。 【例】JIS X 0213:2012, 文字情報基盤
	◎	文字コード	データセット中のデータ記述に使われている文字コードを 記入する。
	◎	説明	データセットの特徴を第三者に理解してもらうための説明 を記入する。
	◎	アクセス URL	データセットにアクセス (参照、取得など) できる URL を 記入する。
	◎	公開範囲	データセットの公開範囲を、次のいずれかで記入する。 [公開/制限付き公開/非公開]
	○	公開条件	データカタログの公開範囲が「制限付き公開」の場合に、 制限条件及び制限解除するための条件を記入する。
標準適合性	◎	文字	データセット中のデータ記述に使われている文字セットを 記入する。 【例】JIS X 0213:2012, 文字情報基盤
	◎	文字コード	データセット中のデータ記述に使われている文字コードを

			記入する。
	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	標準モデル	参照とした標準モデルを記入する。
機密性	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	データ種別	産業データ、行政データ、パーソナルデータのうち該当するものを選択し、記入する。
	○	同意区分	パーソナルデータが含まれる場合、区分を記入。同意済、同意不要、同意なし、未確認から該当するものを選択し、記入する。
	◎	公開範囲	データセットの公開範囲を、以下のいずれかで記入する。 ・公開 ・制限付き公開 ・非公開
	○	公開条件	データカタログの公開範囲が「制限付き公開」の場合に、制限条件及び制限解除するための条件を記入する。
	○	秘密保持義務	購入者に同意を求めるか、求めないか、についてテキストで記入する。
	○	データ保護要件	このデータ利用条件が記入の対象としているデータ保護要件として、提供するデータの性質や、当該データの保護のために、データ利用者に求める資格や、管理方法等の要件について記入する。
効率性	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	標準モデル	参照とした標準モデルを記入する。
精度	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	データセットの対象地域	データセットの対象地域を記入する。記述方法は、自治体コード「総務省自治行政局全国地方公共団体コード」や、政府 CIO ポータルサイトの「行政データ連携標準」などを利用する。
	○	データセットの対象地域の分解能	データセットの対象地域を記入した場合、分解能の情報がある場合に記入する。 分解能の情報としては、データが画像やグリッドである場合に、データ項目間の空間距離をメートル単位で記入する。あるいは、画像やグリッド以外の種類の空間データの場合、データ項目の間の距離の最小のものを代表値として記入する。 なお、複数の分解能の情報がある場合、サマリとして参

			照される値（例、最小の距離）を記入する。 【例】30メートル
	○	データセットの対象期間	データセットの対象期間開始日及び 終了日を記入する。記入形式は以下のとおりとする。 ただし、明確に年月日で示せない場合は、自由記述とする。 ・開始年月日/終了年月日： YYYY-MM-DD/YYYY-MM-DD ※YYYY-MM-DD（西暦-月-日）
	○	データセットの対象期間の分解能	データセットの対象期間を記入した場合、分解能の情報がある場合に記入する。 分解能の情報としては、データが時系列データである場合に、データ項目間の時間差を記入する。あるいは、時系列データ以外の種類のデータの場合、データ項目の時間間隔の最小のものを代表値として記入する。 なお、複数の分解能の情報がある場合、サマリとして参照される値（例、最小の時間差）を記入する。 【例】3 か月、3 時間、30 分、1 分 30 秒など
	○	データセットの有効期間	データセットの有効期間開始日及び 終了日を記入する。記入形式は以下のとおりとする。 ただし、明確に年月日で示せない場合は、自由記述とする。 ・開始年月日/終了年月日： YYYY-MM-DD/YYYY-MM-DD ※YYYY-MM-DD（西暦-月-日）
追跡可能性	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	データセットの提供者	データセットの提供者（組織又は個人）の名称を記入する。
	○	データセットの提供者番号	データセットの提供者（組織）の法人番号があれば記入する。
	○	データセットが生成された営み	データセットが生成された営みについて説明を記入する。 不動産登記、防火、防災管理講習、○○補助金申請など。
	○	データセットを生成したシステム	データセットを生成したシステムについて説明を記入する。 ○○市住民記録システム、ボランティア管理システムなど。
	○	データセットの作成日	データセットが作成された日付を年月日で記入する。記入形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日）

	○	データセットの発行日	データセットが発行された日付を年月日で記入する。記入形式は以下のとおりとする。 ・YYYY-MM-DD (西暦-月-日)
	○	データセットの更新日 又は 修正日	データセットが更新 又は 修正された日付を年月日で記入する。 記入形式は以下のとおりとする。 ・YYYY-MM-DD (西暦-月-日)
	○	バージョン	データセットのバージョン情報を記入する。
	○	レコード ID の有無	データセットに含まれる個々のレコードを特定する ID (プライマリーキー) が付与されている場合はあり、されていない場合はなし、未確認の場合は空白を記入する。
理解性	◎	タイトル	データセットのタイトルを記入する。
		サブタイトル	データセットのサブタイトルを記入する。
	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
		関連ドキュメント	データセットに関する情報を持つドキュメントへのリンク (URL) を記入する。
		ランディングページ	データセットに関して、追加及び補足できる情報が公開されている URL を記入する。
	○	標準モデル	参照とした標準モデルを記入する。
可用性	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	データセットの更新頻度	データセットの更新頻度について記入する。 定期的に更新する場合、年/月/週/日あたりの回数を記入する。 不定期の場合は、不定期と記入する。
移植性	◎	文字	データセット中のデータ記述に使われている文字セットを記入する。 【例】JIS X 0213:2012, 文字情報基盤
	◎	文字コード	データセット中のデータ記述に使われている文字コードを記入する。
	◎	使用言語	データセット中のデータ記述に使われている言語を記入する。なお、国際標準化機構 (ISO) が発行する「ISO 639-1:2002」又は「ISO 639-2:1998」に準拠したアルファベット 2 文字又は 3 文字で表記する。 【例】日本語 : ja, 英語 : en
	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	容量	データセットの容量を記入する。
	○	レコード数	提供可能なレコード数を記入する。

	◎	配信方法	API、ダウンロードによるファイル提供、個別に定める方法でのファイル提供などの配信方法を記入する。
	◎	形式	ファイル提供の場合、形式を記入する。文書ファイル、アプリケーションファイル、スキーマ、動画、音声、書籍、記事から選択し、記入する。
回復性	◎	管理 ID	データセットを一意に識別するための ID を設定し記入する。
	◎	説明	データセットの特徴を第三者に理解してもらうための説明を記入する。
	○	レコード ID の有無	データセットに含まれる個々のレコードを特定する ID(プライマリーキー) が付与されている場合はあり、されていない場合はなし、未確認の場合は空白を記入する。

*【必須区分の凡例】 ◎：必須（必ず記入） ○：推奨（できるだけ記入）“(空白)”：任意に記入

② 付加的品質特性の評価結果について

データの付加的品質特性の評価結果は、「必須項目の記入判定」と「付加的品質特性の評価」により決定します。それぞれの表示イメージを図 4-2 と図 4-3 に示します。

図 4-2 は、「表 4-4 メタデータ項目の実体（値）に記入する具体的な内容について」で示す、必須区分が「◎」となっている項目が網羅的に記載されている場合に、「必須項目はすべて記入されています」と表示されます。

また、図 4-3 の「付加的品質特性ごとのメタデータの記入率」は、付加的品質特性に関するメタデータ項目がどの程度記入されているかを示すものです。例えば「最新性」が 100% である場合には、データが常に最新であることを示すものではなく、「最新性」に関するメタデータの項目をすべて記入していることを示すものであることについて、関係者で誤解が生じないよう留意が必要です。

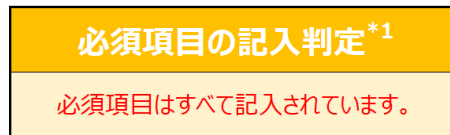


図 4-2 「必須項目の記入判定」の表示イメージ



図 4-3 「付加的品質特性の評価」の表示イメージ

4.1.2. データ連携基盤の整備主体の実施作業

データ連携基盤の整備主体は、データ提供者からデータもしくはデータセット、データの品質情報、データの取扱説明書を受領し、提供されたデータ及びメタデータを確認します。その方法は以下のとおりです。

- 提供されたデータをサンプリングして確認することで、データ全体の特徴や性質を推測し、データの基礎的品質特性を確認する。
- データ提供者がデータをアップロードする際に、データに加え、データの品質情報が表示された状態で提供されているかを確認する。

なお、表 4-4 の必須区分が「◎」となっている「メタデータ項目の実体（値）の記入欄」に必要十分な内容が記入されていない場合や、品質評価が行われていないデータがある場合等の不備があった場合には、データ連携基盤の整備主体は、データ提供者からのデータを受け付けないことも、データカタログ上でデータを公開しない等の措置を講じることも、データの品質を確保するためには必要になります。

4.2. データの品質情報の提示とデータの仲介

4.2.1. データ連携基盤の整備主体の実施作業

データ提供者から受領したデータの品質情報等について、データ連携基盤のデータカタログ上でデータ利用者（サービスの提供者）が分かりやすい形で提示します。その際、各品質特性をカテゴリとして検索できるようにするなど、利用者にとって分かりやすく誤解を招かないよう工夫する必要があります。

また、データの提供方法には、都度データを取得させる場合のほか、定期的・継続的にデータを取得させる場合があり、昨今では、月単位・年単位といった一定の期間、データを利用できるデータの仲介方法（以下、サブスクリプションという）の需要も高まっています。こうした多様なデータの提供／仲介方法に対応するためには、データ連携基盤の整備主体が、必要に応じて、これらのデータの提供／仲介方法に対応した利用規約の策定やデータ連携基盤の機能提供等を行うことも考えられます。

なお、データ連携基盤の整備主体が、サービスを提供する場合とデータ提供者から提供されたデータの変換、加工等を行う場合は、第 2.4 章で示した責任分界点に関する基本的な考え方を考慮する必要があります。

4.3. データの品質情報の確認とデータの利用

4.3.1. データ利用者の実施作業

データ利用者（サービスの提供者）は、提供されたデータ本体及びデータ連携基盤の整備主体が提供するデータの品質情報を確認し、データ利用者（サービスの提供者）自身の要求を満たすか確認した上で、データの取得要否を選択します。データを取得する場合は、データ連携基盤の整備主体の定めるデータの品質管理規程及び提供データの利用条件等に従って、データを取得します。サービスの提供者は、取得したデータを利用してサービスを提供します。

なお、定期的・継続的にデータの取得が必要になる場合は、データの取得方法の一つとしてサブスクリプションを利用することも考えられます。その場合は、データ連携基盤を通じて提供されるデータがサブスクリプションに対応しているどうかを確認してください。

データの利用時には、提供データのメタデータについて、ライセンス及びパーソナルデータに関連する情報（どのデータがパーソナルデータに該当するか、その利用にあたり同意が必要か等）を確認し、利用条件に従ってデータを利用します。

5. データ利用者による品質改善要求時に実施すべき事項

【概要】

第5章では、2.3.2 節（3）で示した「データ利用者による品質改善要求時に実施すべき事項」に関して、データ利用者による改善要求／要望の起案（第5.1章）、データ連携基盤の整備主体による改善要求／要望の受付（第5.2章）、データ提供者による改善要求／要望への対応（第5.3章）を説明します。

データの品質に関する改善要求／要望の流れを図5-1に示します。

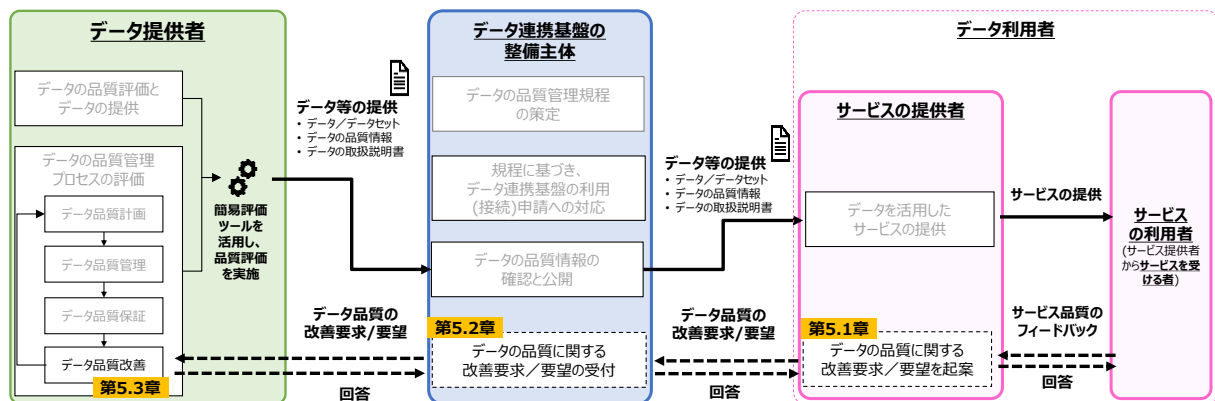


図 5-1 データの品質に関する改善要求／要望の流れ

5.1. データ利用者による改善要求／要望の起案

データ利用者（サービスの提供者）は、住民等のデータ利用者（サービスの利用者）によるサービスへの改善要求／要望を基に、データの品質に関する改善要求／要望を起案し、データ連携基盤の整備主体が整備した受付窓口にて要求／要望の内容を伝えます。具体的な改善要求／要望の内容としては、以下のようなケースが考えられます。

- データ提供者が提示した品質情報を満たしていない不適切なデータに対する改善要求
【改善要求例】
5分単位のデータのはずが、10分単位のデータしかないため、改善してほしい
- データ提供者が提示しているデータの品質よりも、高い品質を有するデータの追加要望
【改善要望例】
5分単位のデータではなく、1分単位のデータを提供してほしい

5.2. データ連携基盤の整備主体による改善要求／要望の受付

データ連携基盤の整備主体は、受付窓口（回答フォームやメール等の通信手段）を通して、データ利用者（サービスの提供者）から改善要求／要望を受け付けます。

なお、改善要望については、既存のデータ提供者では対応できない場合も想定されるため、その際には、データ連携基盤の整備主体が新しいデータ提供者を探索するといった対応も考えられます。

5.3. データ提供者による改善要求／要望への対応

データ提供者は、データ連携基盤の整備主体から共有された改善要望／要求の対応を行います。

特に、データ提供者が提示した品質情報を満たしていない不適切なデータに対する改善要求については、品質情報の訂正を行うといった対応を速やかに講じる必要があります。また、上記、5.1 章で説明した「データ提供者が提示しているデータの品質よりも、高い品質を有するデータの追加要望」については、自身のデータ作成環境を鑑みて対応を検討します。

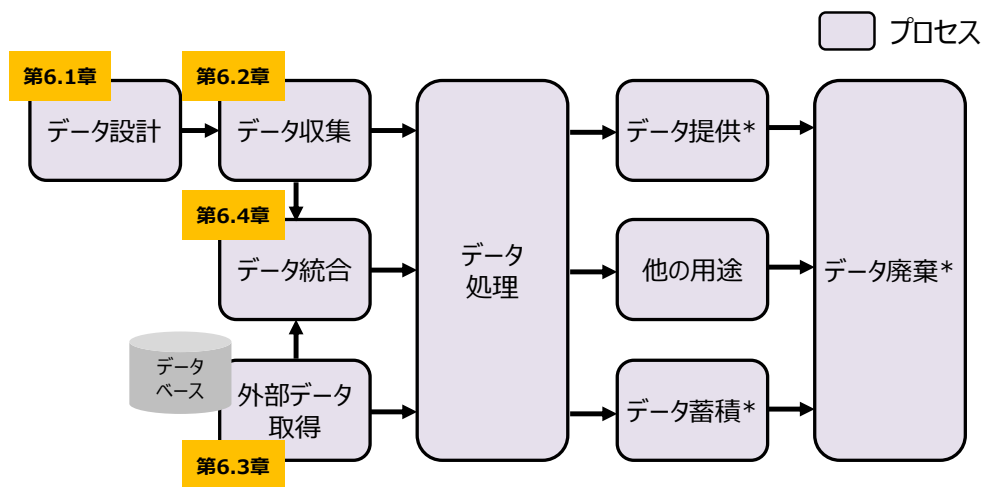
データ提供者は、対応した結果をデータ連携基盤の整備主体に回答します。

6. データ生成時の品質向上のための工夫

【概要】

第5章まででは、生成されたデータの品質管理の手法について説明してきましたが、データ品質を向上させるためには、データ生成時に品質の高いデータを生成することが重要となります。このため、第6章では、ISO/IEC 25024 のデータライフサイクルを参考に、データ生成時の品質向上のための工夫や重要となるポイントを説明します。

データのライフサイクルを図6-1に示します。



*ISO/IEC 25024の記載から、一部記載を変更したもの

図 6-1 データのライフサイクル

6.1. データ設計

選択肢があるデータ項目は、自由記述ではなく、コード入力やプルダウンメニューによる選択式の入力にすることで誤データの混入を防ぐことができます。データ設計時においては、選択肢があるデータ項目に「データ標準」を利用することで、設計品質や相互運用性の高いデータを設計することができます。具体的には、GIF などに内包されている行政データ連携標準、IMI 共通語彙基盤¹³、推奨データセットや schema.org 等のデータ標準が参考になります。

ただし、例えば、「東京都」と入力する代わりに「東京都の地方公共団体コード」を入力させることとした場合、利用者の利便性が低下するため、選択肢により「東京都」を選択するとデータ上では「東京都の地方公共団体コード」が自動的に書き込まれるといった工夫をするなど、利用者の利便性にも配慮す

¹³ IMI 共通語彙基盤 : <https://imi.go.jp/goi/imi-about/>

ることが重要です。この場合の選択肢のことをコントロールボキャブラリ¹⁴と呼びます。コードやコントロールボキャブラリは、独自に設定するのではなく、国際標準や、総務省統計局によるコード一覧など既存の体系を利用することにより、他データとの連携も容易に行えるようになります。また、コントロールボキャブラリの作成時には各データ項目の定義を明確にすることにより他データと連携が容易になります。

6.2. データ収集

データの収集時には、データの重複や誤りなどをあらかじめ排除することが重要となります。この作業を実施することで、次のステップのデータの統合時に行うデータクレンジング作業を効率的に行うことができ、データを提供する前にデータの品質向上を図ることができます。

データの重複や誤りを排除するための工夫として、アナログデータとして収集したデータを手入力によりデジタルデータに変換するといった作業を行うのではなく、なるべく機械的にデジタルデータとしてデータ収集を行うことが考えられます。例えば、アンケートデータ等の場合は、データ入力フォームを利用することで、データ収集時のデータ形式の揺らぎや誤データの混入を防ぐことができます。

また、IoT 機器等により、機械的にデータを取得することもデータの誤り等を排除するための有効な手段となりますが、一時的なデータ欠損やスパイクデータ¹⁵が発生することも考えられるため、その点についてデータ提供時に注意事項として明示することが求められます。なお、データ入力フォームは、自身で作成するだけでなく、外部のサービスを利用することも考えられます。

6.3. 外部データ取得

外部データの取得時に、あらかじめ妥当性検証（バリデーションチェック）ツールなどを用いてデータの品質を確認した上で、データの取得要否を判断する必要があります。妥当性検証を実施しない場合は、データ形式に揺らぎのあるデータや誤データが外部から混入する可能性があります。

6.4. 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、コードの変換などを正確に行う必要があります。データ統合時に品質の劣化を起こさないよう注意が必要です。

データのクレンジングにあたっては、OpenRefine¹⁶（データのクリーニングツール）や IMI コンポーネン

¹⁴ コントロールボキャブラリ：情報検索において、索引語として利用する語を限定し、その意味範囲や使用方法を規定したものです。同義語や語形変化などの表現の違いによる検索漏れを防ぐことができます。

¹⁵ スパイクデータ：センサーが検知するデータの中で、通常のデータと比べて急激に増加する、あるいは異常に高い値を持つデータのことです。

¹⁶ OpenRefine： <https://openrefine.org/>

ト¹⁷（文字列の表記ゆれを正規化するツール）等を利用することも有効です。ただし、ツールを使って一括処理する際には、誤変換を起こさないように細心の注意を払う必要があります。

また、紙のデータを手入力や OCR¹⁸でデジタル化したデータは、誤データが含まれる可能性があることから、既存のデータとの照合、計算による検証、異なる OCR 結果との照合など、必要に応じて確認を行います。

¹⁷ IMI コンポーネントツール： https://info.gbiz.go.jp/tools/imi_tools/index.html

¹⁸ OCR：Optical Character Recognition/Reader の略。光学的文字認識のこと。手書きや印刷された文字を、イメージスキャナやデジタルカメラによって読みとり、コンピュータが利用できるデジタルの文字コードに変換する技術です。

A. 付録

【概要】

付録 A では、第 3.2 章で示した「データ提供者のデータの品質管理プロセスの評価」と、4.1.1 節(1)で示した「データの品質評価」を簡易的に行うための評価ツール（簡易評価ツール）の操作方法について説明します。

簡易評価ツールのシート「【A】データの品質管理プロセスの評価」、シート「【B】基礎的品質特性の評価」、シート「【C1】付加的品質特性の評価(メタデータの記入)」、シート「【C2】付加的品質特性の評価(結果出力)」も併せて参照してください。

また、付録 A.2 では、仮想ユースケースにおいて、簡易評価ツールの具体的な利用イメージや品質評価する際に重要となる観点について説明します。

A.1. 簡易評価ツールに関する操作説明

簡易評価ツールは、データ提供者が「データの品質管理プロセスの評価」と「データの品質評価」を行う際に使用します。以下では、簡易評価ツールの操作方法について説明します。

A.1.1. 「データの品質管理プロセスの評価」に関する簡易評価ツールの操作方法

データ提供者が行う「データの品質管理プロセスの評価」に関する簡易評価ツールの操作方法について説明します。データの品質管理プロセスの評価に関する基本的な考え方は、第 3.2 章を参照してください。

- (1) シート「【A】データの品質管理プロセスの評価」を表示してください。
- (2) シート上部のデータ提供者名と評価者を記入してください。
- (3) 評価項目の内容を参考に、「評価項目(E 列)」それぞれに対応する「評価(F 列)」のプルダウンメニューをすべて選択し、「評価理由(G 列)」は必要に応じて記入してください(赤枠部分)。
- (4) セル「データの品質管理プロセスの最終評価結果(F6、G6)」を確認してください。データの品質管理プロセスの最終評価は、各プロセスの成熟度レベルの平均値が算出されます。なお、「評価(F 列)」の記入がすべて完了していない場合、評価結果は出力されません。

A.1.2. 「データの品質評価」に関する簡易評価ツールの操作方法

データ提供者が行う「データの品質評価」に関する簡易評価ツールの操作方法について説明します。データの品質評価に関する基本的な考え方は、4.1.1 節を参照してください。

(1) 基礎的品質特性の評価

- ① シート「【B】基礎的品質特性の評価」を表示してください。
- ② シート上部のデータ提供者名、評価対象データ名、評価者を記入してください。
- ③ 「評価項目(D 列)」と「確認項目(E 列)」の内容を参考に、それぞれの品質特性に対応する「評価(F 列)」のプルダウンメニューをすべて選択し、「評価理由(G 列)」は必要に応じて記入してください(赤字部分)。
- ④ セル「利用上の注意事項(F9、G9)」は必要に応じて記入してください。
- ⑤ セル「データの品質の最終評価結果(G5)」を確認してください。なお、「評価(F 列)」の記入がすべて完了していない場合、最終評価結果は出力されません。

(2) 付加的品質特性の評価

- ① シート「【C1】付加的品質特性の評価(メタデータの記入)」を表示してください。
- ② シート上部のデータ提供者名、評価対象データ名、評価者を記入してください。
- ③ 「メタデータ項目の説明(F 列)」を参考に、「メタデータ項目の実体(値)の記入欄(H 列)」をすべて記入してください(赤字部分)。
- ④ セル「必須項目の記入判定(H5)」を確認してください。必須項目の記入がすべて完了している場合には、「必須項目はすべて記入されています。」と表示されます。必須項目に未記入がある場合には、「未記入の必須項目があります。」と表示されます。
- ⑤ シート「【C2】付加的品質特性の評価(結果出力)」を表示してください。
- ⑥ シート上部のレーダーチャートと「記入率(F 列)」を確認することで、付加的品質特性に関するメタデータがどの程度記入されているかを確認することができます。なお、それぞれの品質特性に紐づくメタデータの記入率が 50%を下回っているものは赤字になります。

A.2. 仮想ユースケースにおけるデータの品質評価の適用

効率的、効果的なデータの品質管理を促進するためには、データ提供者が、提供するデータの利用用途を想定して、重要となり得る品質特性に関するメタデータの記述内容を充実させることが重要となります。これにより、より多くのデータ利用者（サービスの提供者）が適切にデータを利用することが可能となり、結果としてデータを活用したサービスの提供が促進されます。

このため、本ガイドブックでは「移動／物流分野」と「健康／医療分野」を取り上げ、それぞれの仮想ユースケースにおいて、重要視される品質特性やデータの品質評価の実践方法を説明します。

A.2.1. 仮想ユースケース① 移動／物流分野

(1) 仮想ユースケース①の概要

本ユースケースは、A社が、「C市データ連携基盤」上のデータを利用し、小型パーソナルモビリティを利用したシェアリング型の移動サービスの提供に取り組む事例となります。小型パーソナルモビリティを運行する際には、以下に示す「走行可否の判断に必要な情報」と「安全な走行ルートの判断に必要な情報」の大きく2つの情報が必要となります。

- **走行可否の判断に必要な情報**

小型パーソナルモビリティの走行可否を判断するデータ

- 天気予報データ : B社

- **安全な走行ルートの判断に必要な情報**

小型パーソナルモビリティの安全な走行ルートを判断するためのデータ

- ゴみの収集場所や回収日時(曜日時間)データ : C市
- 地域イベント情報データ : C市
- 道路占用許可データ : C市
- 道路使用許可データ : D県警察 (C市内管轄)
- 横断歩道信号機情報データ : D県警察 (C市内管轄)
- 通行制限情報データ : E公益財団法人
- 地図情報データ : F社
- 路面情報データ : G社
- 歩道走行精度地図データ : H社
- 車両通行データ : I社

本ユースケースの全体像を図 A-1 に示します。

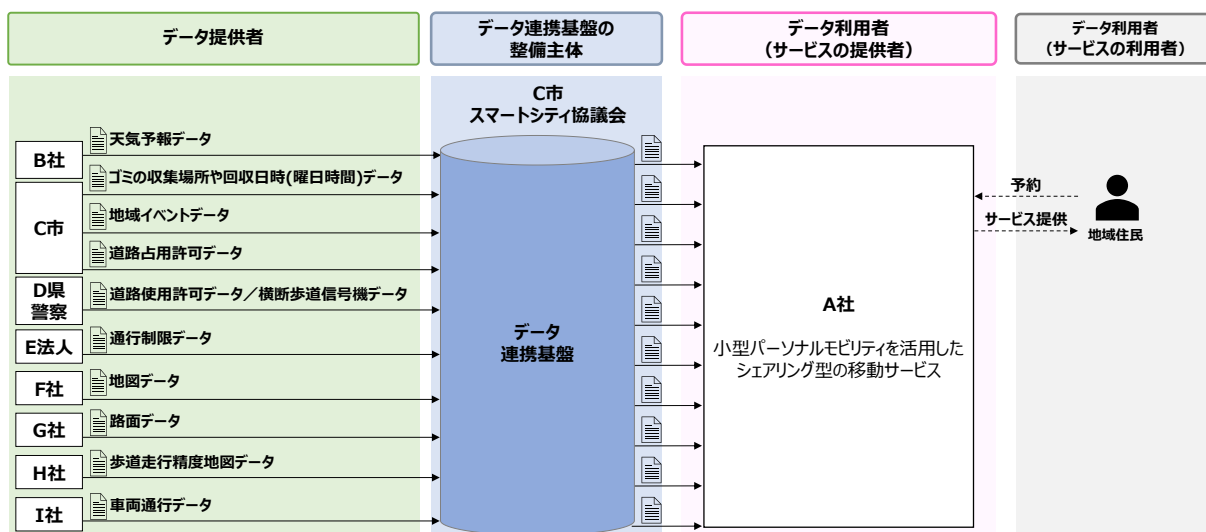


図 A-1 仮想ユースケース①の全体像

(2) 想定される課題

- **走行可否の判断に関する想定課題**

小型パーソナルモビリティは、雨天の走行が困難であるため、天気予報をもとに走行可否の判断を行う必要があります。仮に、データの更新日（最新性）や対象地域（精度）が示されていない場合、走行可否の判断を正確に行えないことが想定されます。

- **走行ルートの判断に関する想定課題**

小型パーソナルモビリティを安全に走行させるためには、歩道等の道路に関する地図情報や障害物に関する情報をもとに、安全な走行ルートを選択する必要があります。仮に、データの更新日（最新性）や地図のメッシュ（精度）が示されていない場合、安全な走行ルートの判断を正確に行うことができず、思わぬ事故が発生してしまう可能性があります。

(3) 仮想ユースケース①におけるデータの品質評価の実践例

本ユースケースにおける、基礎的品質特性及び付加的品質特性の評価方法を説明します。

- **基礎的品質特性の評価の実施**

基礎的品質特性については、ユースケース（利用目的）によって評価方法が変わらないため、A.1.2 に示すシート「【B】基礎的品質特性の評価」の操作方法を参照し、品質評価します。

- **付加的品質特性の評価の実施**

付加的品質特性の評価は、利用目的に応じて、必要となる品質特性が異なるため、まずはデータごとに必要な品質特性の特定をすることが重要となります。その上で、重要となる品質特性に関するメタデータの記述内容を充実させることが必要です。本ユースケースで利用するデータと

重要と考えられる品質特性を、表 A-1 に示します。また、その品質特性を考慮すべき理由を説明します。

表 A-1 利用するデータ及び重要と考えられる品質特性

#	データ	重要と考えられる品質特性
1	天気予報データ	「最新性」、「精度」
2	ゴミの収集場所や回収日時（曜日時間）データ	「最新性」、「精度」、「信憑性」
3	地域イベント情報データ	「最新性」、「精度」、「信憑性」
4	道路占用許可データ	「最新性」、「精度」、「信憑性」
5	道路使用許可データ	「最新性」、「精度」、「信憑性」
6	横断歩道信号機情報データ	「最新性」、「精度」、「信憑性」
7	通行制限情報データ	「最新性」、「精度」、「信憑性」
8	地図情報データ ¹⁹	「最新性」、「精度」
9	路面情報データ	「最新性」、「精度」
10	歩道走行精度地図データ	「最新性」、「精度」
11	車両通行データ	「最新性」、「精度」

● **走行可否の判断に必要な情報に関する重要と考えられる品質特性**

例えば、天気予報データ（表 A-1 の#1）では、「更新期間や最終更新日が明示されているか（最新性）」、「データの精度について明示されているか（精度）」が重要になります。一日前の天気予報では晴天であっても、最新のデータでは雨天に変わっている可能性があります。また、天気予報は様々な範囲（県レベル、地域レベルなど）で情報提供されるため、データにサービス提供するエリアが含まれているというだけでは情報として十分ではなく、できる限りピンポイントの天気予報である方が判断しやすいと考えられます。

このように、走行可否の判断に当たっては、データの「最新性」や「精度」に関するメタデータの記述内容を充実させることが特に重要であると考えられます。

● **安全な走行ルート判断に必要な情報に関する重要と考えられる品質特性**

例えば、ゴミの収集場所や回収日時（曜日時間）データ（表 A-1 の#2）では、「更新期間や最終更新日が明示されているか（最新性）」、「データの精度について明示されているか（精度）」、「出所や更新日が明示されているか（信憑性）」が重要になります。ゴミの収集場所や時間が変わっていた場合、その情報を把握していないと誤った走行ルートを選定する可能性があります。また、ゴミの収集場所が“どの地域、どの道路”というだけでは、その収集場所を通らざるを得ない場合に避けられるのか、避けられないのかの判断が難しくなるため、“道路のどの辺りに障

¹⁹ 仮想ユースケース④では、2次元の地図データを取り扱っていますが、ドローン飛行で用いられる3次元の地図情報についても重要なデータの品質特性とメタデータは基本的には同じです。

害物がある”といった精度が要求されます。さらに、出所や更新日を明示し、信頼に値するデータとして利用してもらう必要があります。

このように、安全な走行ルートの判断に当たっては、データの「最新性」、「精度」、「信憑性」に関するメタデータの記述内容を充実させることが特に重要であると考えられます。

A.2.2. 仮想ユースケース② 健康／医療分野

(1) 仮想ユースケース②の概要

本ユースケースは、K社が、「M市データ連携基盤」上のデータを利用し、救急救命士が救急搬送時に個人を特定する情報を使用して、傷病者の特定及び傷病者の救急関連情報の取得／参照を行うサービスの提供に取り組む事例となります。救急搬送時に搬送先医療機関の医師や救急救命士が傷病者の情報を参照するために「傷病者の救急関連情報」が必要となります。

● 傷病者の救急関連情報

傷病者の救急関連情報を参照するための元となるデータ

- 救急関連データ（PHR²⁰データを含む） : L社
- 母子健康手帳に関するデータ : Q社

本ユースケースの全体像を図 A-2 に示します。

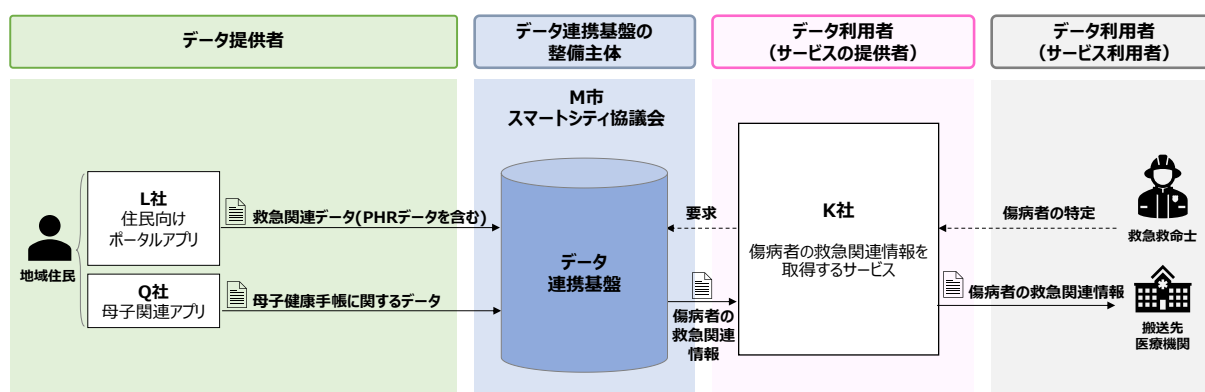


図 A-2 仮想ユースケース②の全体像

²⁰ PHRとは、Personal Health Recordの頭文字をとった略語で、個人の健康、医療、介護に関する情報を示します。

(2) 想定される課題

本サービスで扱う情報には、パーソナルデータが含まれるため、同意区分等のパーソナルデータの取扱いに関する品質情報が必要となります。仮に、同意区分等に関する情報が示されていない場合、傷病者の同意を得られていない可能性があるため、適切なサービス提供を行うことができないと想定されます。

また、古いデータを表示することにより、不適切な医療処置を行う可能性も想定されます。

(3) 仮想ユースケース②におけるデータの品質評価の実践例

本ユースケースにおいても、付録 A.2.1 節 (3) に示した方法と同様に、基礎的品質特性及び付加的品質特性の評価を行います。本ユースケースで利用するデータに関する重要と考えられる品質特性を、表 A-2 に示します。また、その品質特性を考慮すべき理由を説明します。

表 A-2 利用するデータ及び重要と考えられる品質特性

#	データ名	重要と考えられる品質特性
1	救急関連データ (PHR データを含む)	「最新性」、「信憑性」、「機密性」
2	母子健康手帳に関するデータ	「最新性」、「信憑性」、「機密性」

● 傷病者の救急関連情報に関する重要と考えられる品質特性

例えば、救急関連データと母子健康手帳に関するデータ (表 A-1 の#1・2) では、「更新期間や最終更新日が明示されているか (最新性)」、「出所や更新日が明示されているか (信憑性)」、「アクセス制限等の対策が講じられていることが明示されているか (機密性)」が重要になります。データが最新でない場合、傷病者の既往歴や緊急時の連絡先などの情報が不足もしくは誤っている可能性があります。

また、データの取得について傷病者から同意が得られているか不明 (データの出所が不明) である場合、プライバシー保護などの法的要件に抵触する可能性があります。

さらに、不正アクセスにより他人にデータが取得されてしまった場合、データをもとに傷病者の様々な情報が盗まれてしまう可能性があります。

このように、傷病者の救急関連情報を利用する際には、「最新性」、「信憑性」、「機密性」に関するメタデータの記述内容を充実させることが重要だと考えられます。

A.2.3. 3 者のステークホルダーによるデータの品質管理の実施

付録 A.2.1 節及び A.2.2 節においては、「移動／物流分野」と「健康／医療分野」を取り上げ、それぞれの仮想ユースケースにおいて、重要視される品質特性を示しましたが、データの品質管理を効率的、効果的に行うためには、重要と考えられる品質特性を考慮した上で、3 者のステークホルダーが互いに連携し、データの品質管理を実施していくことが重要となります。

具体的には、第 2.3 章で説明した 3 者のステークホルダーの役割に応じて、データ提供者はメタデータの記述内容を充実させ、データ連携基盤の整備主体はその状況を確認後に提示し、データ利用者（サービスの提供者）は重要と考えられる付加的品質特性の内容を確認し、必要に応じ改善要望を行うというデータの品質管理プロセスを実践することで、互いに連携し合ってデータの品質管理を実施していきます。

このような継続的にデータの品質を高める仕組みを構築し実施することで、各仮想ユースケースで挙げた「想定される課題」に適切に対処し、住民の安全や安心を確保したサービスの提供を実現することにつながります。

B. 用語説明

【概要】

付録 B では、本ガイドブックの中で使用する用語を説明します。

【データ連携基盤】

データ連携基盤は、自治体や事業者、個人等が有する様々なデータを収集、整理、提供することで、様々なデータやサービスを相互に連携させ、相乗効果を創出する基盤であり、住民課題の解決や住民サービスの向上をめざすスーパーシティ/スマートシティにおける中核となる基盤となります。

なお、本ガイドブックにおいては、「スーパーシティ/スマートシティのデータ連携等に関する検討会 中間とりまとめ」におけるデータ連携基盤と都市 OS の関係性（図 B-1 を参照）の定義に基づき、データ連携基盤と都市 OS を区別することなく記載しています。

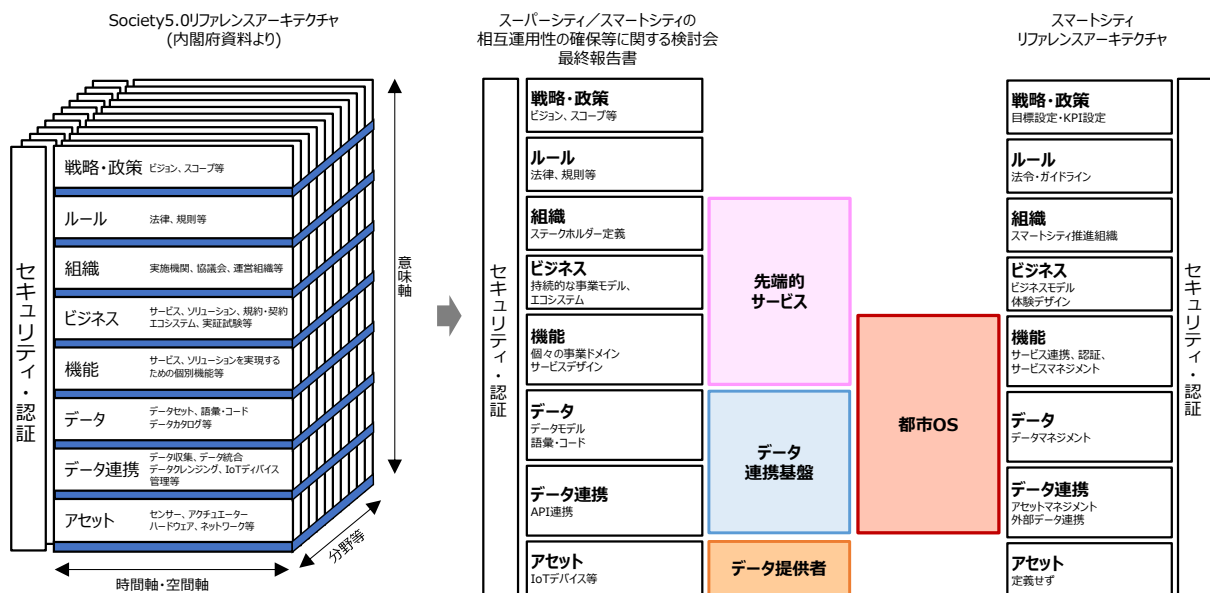


図 B-1 データ連携基盤と都市 OS の関係性

（「スーパーシティ/スマートシティにおける データ連携等に関する検討会 中間とりまとめ」を参考に作成）

【データ連携基盤の整備主体】

データ連携基盤を整備し、データ提供者からデータ利用者（サービスの提供者）へデータを仲介する者を示します。

【データ提供者】

自らの事業や観測活動によりデータを生成し、それらのデータを整理、加工して提供する者を示します。

【データ利用者】

データ提供者からデータ連携基盤を通してデータの提供を受ける者を示します。なお、データ連携基盤を通して取得したデータを加工せず利用する者、取得したデータを利用（加工）してサービスを提供する者（サービスの提供者）、サービスの提供を受ける者（サービスの利用者）に分けられます。

【サービスの提供者】

サービスの提供者とは、データ利用者のうち、データ連携基盤を通して取得したデータを利用（加工）してサービスを提供する者であり、サービス、製品の開発や運用などに利用するほか、自らの事業に利用する者を示します。

【サービスの利用者】

サービスの利用者とは、データ利用者のうち、サービスの提供者からサービスを受ける者（住民等）を示します。

【データの品質】

データの品質とは、データがある目的を満たすために、具備すべき特性を定義したものであり、データの状態（質）を表したものを示します。一般的にデータの品質が良いということは、データが正確で、かつ信頼のあるデータであることを示します。

【データの品質管理】

データ利用者のニーズを満たすためにデータの品質を継続的に高めていくことを示します。データ連携基盤の利用における安全性及び信頼性を確保するために、データの品質管理が極めて重要となります。

【データの品質評価結果】

基礎的品質特性と付加的品質特性に関して、簡易評価ツールを用い、データの品質の評価を行った結果を示します。

【データの品質情報】

データ提供者から受領するデータの品質評価結果及びメタデータをデータの品質情報と言います。

【基礎的品質特性】

どのような利用シーンにおいても一定以上の高い品質が要求される基礎的な品質特性を示します。「正確性」、「完全性」、「一貫性」の3つの品質特性が該当します。

【付加的品質特性】

利用シーンに応じて求められる品質水準が異なる、基礎的品質特性以外の品質特性を示します。「信憑性」、「最新性」、「アクセシビリティ」、「標準適合性」、「機密性」、「効率性」、「精度」、「追跡可能性」、「理解性」、「可用性」、「移植性」、「回復性」の12の品質特性が該当します。

【データセット】

データセットとは、複数のデータから構成されるデータの集合を示します。

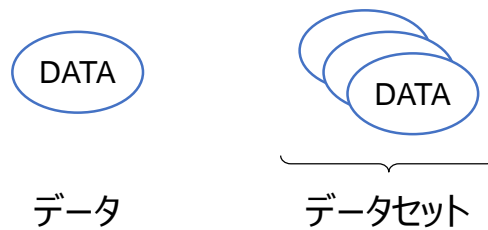


図 B-2 データとデータセットの関係

【メタデータ】

メタデータとは、データ本体を効率的に管理したり検索したりするために、データ本体についてのデータ項目や形式や属性情報等を記述した付帯データを示します。作成日や作成者などの項目が含まれます。

【データの取扱説明書】

データセットを扱う際に、そのデータの取扱方法や制約、保管、共有、セキュリティ対策などを文書化したものを示します。利用の目的や利用方法、制約事項などの項目が含まれます。

【データカタログ】

データカタログとは、データセットやデータの品質情報が一元登録されているカタログを示します。データカタログの利用により、データセットの検索性が向上され、データ提供者、データ利用者のマッチングを促進することが可能となります。

参考文献

本ガイドブックにおける参考文献を以下に示します。

- [1] スマートシティリファレンスアーキテクチャ ホワイトペーパー（内閣府：2023年8月発行）

【概要】

スマートシティに取り組む自治体や企業、関連のサービスを提供する企業等に対し、スマートシティの推進に必要な要素を幅広く提示したものの。

【URL】

<https://www8.cao.go.jp/cstp/stmain/20230810smartcity.html>

- [2] 包括的データ戦略（デジタル庁：2021年6月閣議決定）

【概要】

日本政府が世界トップレベルのデジタル国家を目指す上で必要なデジタル基盤を構築するための戦略として、日本におけるデータアーキテクチャや、プラットフォーム整備の必要性やデータマネジメントの強化の必要性などをまとめたもの。

【URL】

https://www.digital.go.jp/policies/data_strategy

- [3] スーパーシティ／スマートシティの相互運用性の確保等に関する検討会 最終報告書（内閣府：2020年9月発行）

【概要】

データ連携基盤の相互運用性を確保するために具備すべき機能等について検討した「スーパーシティ／スマートシティの相互運用性の確保等に関する検討会」の検討結果をまとめたもの。

【URL】

https://www.chisou.go.jp/tiiki/kokusentoc/supercity/supercity_kentoukai.html

- [4] 「2020年度 データ連携基盤技術報告書」（内閣府：2021年3月発行）

【概要】

2020年9月に公開された「スーパーシティ／スマートシティの相互運用性の確保等に関する検討会の最終報告書」を踏まえ、データ連携基盤として、具備すべき機能、考慮すべき事項に関する調査結果をまとめたもの。

【URL】

https://www.chisou.go.jp/tiiki/kokusentoc/supercity/supercity_kentoukai.html

- [5] スーパーシティ/スマートシティのデータ連携等に関する検討会 中間とりまとめ（内閣府：2021年4月発行）

【概要】

「スーパーシティ/スマートシティの相互運用性の確保等に関する検討会」に引き続き実施した「スーパーシティ/スマートシティにおけるデータ連携等に関する検討会」における検討結果をまとめたもの。

【URL】

https://www.chisou.go.jp/tiiki/kokusentoc/supercity/supercity_kentoukai.html

- [6] デジタル社会推進実践ガイドブック DS-468-1：データ品質管理ガイドブック（デジタル庁：2022年3月発行）

【概要】

データの利用やデータの品質管理が効率的に行われるようにするためのデータ品質管理のフレームワークと評価モデルをまとめたもの。

【URL】

https://github.com/JDA-DM/GIF/blob/main/460_%E5%AE%9F%E8%B7%B5%E3%82%AC%E3%82%A4%E3%83%89%E3%83%96%E3%83%83%E3%82%AF/md/468-1_guidebook_dataquality.md

- [7] ISO/IEC 25012 Data quality model (JIS X 25012_2013 ソフトウェア製品の品質要求及び評価—データ品質モデル) (ISO/IEC JTC1 SC7：2008年発行)

【概要】

全15項目からなるデータの品質特性（正確性、完全性等）を規定してまとめたもの。

【URL】

<https://www.iso.org/obp/ui/en/#iso:std:iso-iec:25012:ed-1:v1:en>

- [8] ISO/IEC 25024 Measurement of data quality (JIS X 25024_2018 ソフトウェア製品の品質要求及び評価—データ品質の測定) (ISO/IEC JTC1 SC7：2015年発行)

【概要】

ISO/IEC 25012 で示すデータの品質特性について、品質測定概念、項目別の品質測定量等を規定してまとめたもの。

【URL】

<https://www.iso.org/obp/ui/en/#iso:std:iso-iec:25024:ed-1:v1:en>

[9] ISO/TS 8000-61 Data quality management : Process reference model (ISO/TC 184 SC4 : 2016 年発行)

【概要】

データの品質評価方法等を定めた ISO8000 のうち、データ品質管理の基本原則、データ品質管理プロセスの構造、実装のための要件等について規定してまとめたもの。

【URL】

<https://www.iso.org/obp/ui/en/#iso:std:iso:8000:-61:ed-1:v1:en>

[10] データ管理成熟度モデル(DMM) (CMMI Institute : 2014 年 6 月発行)

【概要】

データの品質管理プロセスを評価するための手法として、成熟度モデル（マチュリティモデル）をまとめたもの。

【URL】

<https://stage.cmmiinstitute.com/>

[11] デジタル社会推進実践ガイドブック DS-469 : メタデータ導入実践ガイドブック (デジタル庁 : 2022 年 3 月発行)

【概要】

メタデータ導入や活用の考え方を示すとともに、その導入や活用のための方法をまとめたもの。

【URL】

https://github.com/JDA-DM/GIF/blob/main/460_%E5%AE%9F%E8%B7%B5%E3%82%AC%E3%82%A4%E3%83%89%E3%83%96%E3%83%83%E3%82%AF/md/469_guidebook_metadata.md

[12] データカタログ作成ガイドライン V3.1 (一般社団法人データ社会推進協議会 : 2023 年 3 月 公開)

【概要】

ガイドラインの概要や活用方法、データカタログに記述するデータ項目の構造と役割、各データ項目に記述する値の定義や適合性についてまとめたもの。

【URL】

<https://data-society-alliance.org/survey-research/datacatalogguideline/>

[13] データ取引市場運営事業者認定基準 D3_解説（一般社団法人データ社会推進協議会：
2023年6月公開）

【概要】

データ取引市場運営事業者の認定を行う団体として、データ提供者とデータ受領者を仲介する「データ取引市場運営事業者」としての認定要件をまとめたもの。

【URL】

https://data-society-alliance.org/survey-research/cert_datatradingmarket/

[14] デジタル社会推進実践ガイドブック DS-465-1：データマネジメント実践ガイドブック（導入編）
（デジタル庁：2023年2月公開）

【概要】

データ管理を実施するための準備段階（データマネジメント実施計画の進め方等）の進め方について、プロセスごとに実施するためのポイントや実施手順をまとめたもの。

【URL】

https://github.com/JDA-DM/GIF/blob/main/460_%E5%AE%9F%E8%B7%B5%E3%82%AC%E3%82%A4%E3%83%89%E3%83%96%E3%83%83%E3%82%AF/md/465-1_guidebook_datamanagement_intro.md

本ガイドブックは、内閣府地方創生推進事務局が実施した「データ連携基盤から提供されるデータの品質管理の推進に関する調査事業」（受託者：株式会社日立コンサルティング）の一環として、作成されたものです。なお、内閣府及び本調査事業の受託者は、利用者が本ガイドブックの情報や簡易評価ツールによる個々の評価結果を用いて行う一切の行為について、何ら責任を負うものではありません。