

**平成 28 年度**  
**行政におけるデータマネジメント実践に関する調査研究**  
**報告書**

平成 29 年 3 月 31 日

一般社団法人 行政情報システム研究所

## 目次

はじめに 一本調査研究の背景と目的一	3
1. 調査研究の全体像と検討のフレームワーク	4
1.1. 本調査研究の全体像	4
1.2. データマネジメントの検討フレームワークに関する課題	5
① データマネジメント固有のライフサイクルに基づくフレームワークの限界	5
② 情報システムのライフサイクルに基づくフレームワークの限界	7
1.3. 行政データマネジメントのライフサイクルのフレームワーク	8
2. 行政機関におけるデータマネジメントの実践ノウハウ調査	11
2.1. 調査の実施方法	11
① インタビュー対象候補の抽出	12
② インタビュー対象の選定	14
③ インタビュー項目	15
2.2. ノウハウの抽出	16
2.3. ノウハウ探索のためのタイトルおよび問題・課題の設定	18
2.4. 課題の整理	19
2.5. データマネジメントのアプローチの類型	20
① 自主努力	20
② 共通化（共通仕様等準拠）	20
③ 外部化（外部サービス調達等）	21
3. JDMC 研究会における IT 調達時のデータマネジメント課題研究	23
3.1. 検討の実施方法	23
① 研究会の構成	23
② 研究会の開催実績	24
③ 検討の進め方	24
3.2. 抽出された問題の類型	25
3.3. 個別の問題と対応の考え方	27
3.4. まとめ	35
4. 人工知能（AI）の導入・運用におけるデータマネジメントの検討	36
4.1. AI-情報システム間でのデータの用途、必要量、種類の違い	36
4.2. AI の処理内容に応じたデータマネジメントの違い	37
4.3. 有識者へのインタビュー	38
4.4. まとめ	40
5. 今後の取組み（昨年度ハンドブックの見直し）	41
別紙	43

別添資料 ..... 43

## はじめに ―本調査研究の背景と目的―

行政機関における予算や人員の制約が年々厳しくなる中、複雑化・多様化する行政課題に的確に対応し、国民の期待に応えていくためには、行政が保有するデータを積極的に利活用していくことが不可欠との認識が拡がりつつある。

平成 28 年 12 月に成立した官民データ活用推進基本法では、行政機関が行政保有データを主体的かつ計画的に利活用すべきことが初めて立法措置として定められた。また、その実行のための基本計画の策定や制度整備に向けた検討も進みつつある。

行政保有データの利活用を実現していくためには、利活用の源泉となるデータの品質や目的適合性の確保が不可欠である。しかし、行政機関の情報システム部門や業務部門では、そうした取組みの必要性への認識は稀薄であり、「どのデータが最新かわからない」、「コード体系が異なるため、システム同士が繋がらない」といった状況すら散見される。

近年、民間企業では、データの品質や目的適合性などを確保する取組みである「データマネジメント」の研究と実践に向けた活動が拡がりつつあるが、行政機関ではいまだそうした動きは見られない。当研究所ではこうした課題認識に立脚し、昨年度、「行政へのデータマネジメント概念の普及に関する調査研究」を実施し、今後取り組むべき課題と解決の方向性を明らかにした。また、行政分野におけるデータマネジメント普及に向けた第一歩として「行政機関向けデータマネジメント導入ハンドブック」を取りまとめた。

本年度は、昨年度の成果を踏まえ、行政機関の現場でデータマネジメントが展開されるために必要となる工夫・ノウハウや今後取り組むべき課題、またそれらを前提としたデータマネジメントの実践方法を明らかにすべく調査研究を行った。本調査研究の成果が、行政機関におけるデータ活用に向けた取組みの一助となることを期待したい。

なお、本調査研究は、一般社団法人日本データマネジメント・コンソーシアム（Japan Data Management Consortium）（以下 JDMC という）および株式会社リアライズの協力を得つつ当研究所において実施した。また、行政機関および民間企業でデータマネジメントに関わっている多くの方々にインタビューやワーキンググループへの参画を通じて協力いただいた。この場を借りて深く感謝申し上げたい。

一般社団法人 行政情報システム研究所 主席研究員 狩野英司  
研究員 松岡清志

## 1. 調査研究の全体像と検討のフレームワーク

### 1.1. 本調査研究の全体像

本調査研究は、昨年度実施した「平成 27 年度 行政へのデータマネジメント概念の普及に関する調査研究」（以下「昨年度調査」という。）の成果を踏まえ、より具体的なデータマネジメントの実践方法を提示するために実施したものである。

昨年度調査では、行政機関にはデータマネジメントに関する問題が存在すること、データマネジメントを実践することでかかる問題を改善できる可能性があること、そして、民間企業で実践されているデータマネジメントのフレームワークが行政機関にも適用可能であることを明らかにした。また、これらの検討結果を踏まえ、実際に行政機関がデータマネジメントの導入を行うための基本的な考え方を「行政機関向けデータマネジメント導入ハンドブック」（以下「昨年度版ハンドブック」という。）として取りまとめた。

本調査研究では、昨年度整理されたデータマネジメントの基本的な考え方を前提として、以下の調査研究を実施する。

#### 第1章：行政機関におけるデータマネジメントのフレームワークの再整理

データには、情報システムとは別個の固有のライフサイクルが存在する。本調査研究で提示するデータマネジメントの個々の取組みは、行政機関特有の制約条件を踏まえたうえで、データのライフサイクル全体の中で整理し、位置づける必要がある。本調査研究では昨年度調査で策定したフレームワークを見直し、さらに理解しやすい形へと再整理する。

#### 第2章：行政機関におけるデータマネジメントの実践ノウハウ調査

行政機関または行政のデータ利活用等に関わっている団体等が、それぞれの活動を通じてデータマネジメントの観点でどのような工夫や努力を行っているのか、また、その前提としてどのような課題認識を有しているのかを文献調査およびインタビューを通じて把握・整理する。

#### 第3章：JDMC 研究会における IT 調達時のデータマネジメント課題研究

我が国の場合、行政機関のデータマネジメントは実態として、情報システムの構築と運用保守を担当している IT ベンダーが多く部分を担っている。そこで、JDMC の協力を得て、我が国主要 IT ベンダーに対するグループインタビューを実施し、これら企業が認識している現場課題を収集・整理する。

#### 第4章：人工知能（AI）の導入・運用におけるデータマネジメントの検討

最近、様々な行政機関でデータ利活用に向けた取組みに用いる技術の一つとして人工知

能（以下「AI」という。）の導入に向けた検討が進められている。AIは情報システムの一類型ではあるものの、通常の情報システムとは異なる取組みが求められることから、行政機関特有の制約条件を踏まえつつ、AIを導入する際にはどのような点に留意してデータマネジメントに取り組むべきかを検討する。

## 第5章：今後の取組み

本調査研究を通じて得られた知見をもとに、昨年度版ハンドブックをどのように見直し拡充していくべきかを検討する。

以上を図示すると図 1-1 のような流れとなる。

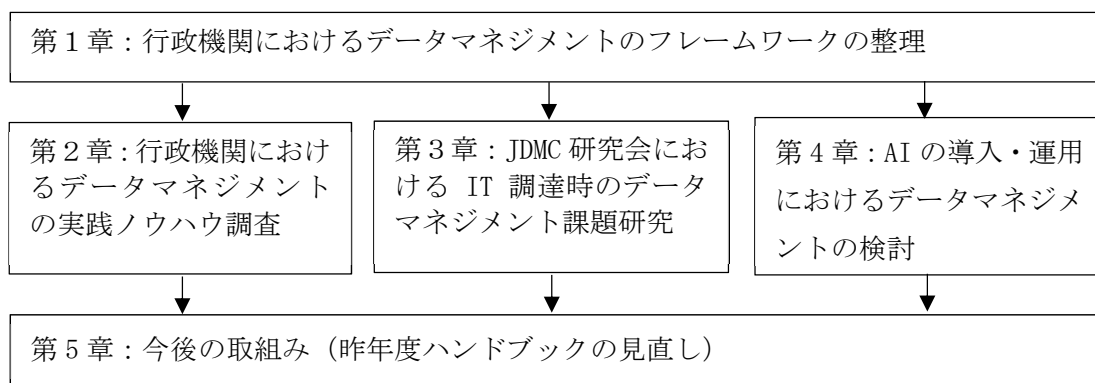


図 1-1 本調査研究の流れ

## 1.2. データマネジメントの検討フレームワークに関する課題

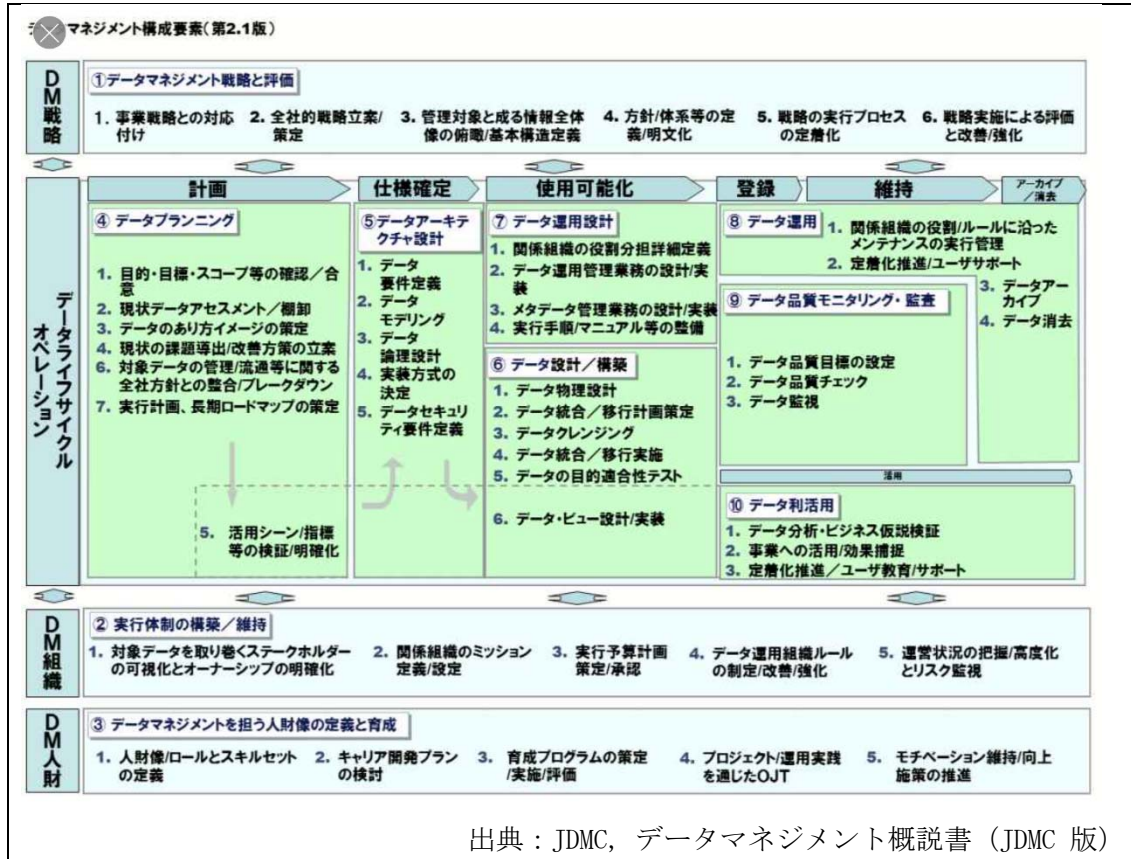
データマネジメントの取組みを体系的に整理するためには、それぞれの活動を、データマネジメントのライフサイクル全体の中でどのように位置づけるべきかを明確にしておく必要がある。行政機関においては、データマネジメントに関わる活動は一般に情報システムの調達・構築・運用プロジェクトの構成要素の一環として位置づけられることが多いことから、昨年度調査では、そのライフサイクルを基礎としてフレームワークを構築した。

しかしながら、データマネジメントの活動は必ずしもこのサイクルと一致するものではなく、本調査研究の過程でも、情報システムに関わらない職員にとって理解が難しいとの指摘を受けた。そこで、まずデータマネジメントのライフサイクルをどのようなプロセスに区分すべきかを次の手順で再整理することとした。

### ① データマネジメント固有のライフサイクルに基づくフレームワークの限界

データマネジメントのフレームワークは、昨年度調査では、JDMC が作成した「データマ

マネジメント概説書（JDMC 版）」を出発点として検討を行った。同フレームワークでは、データのライフサイクルの中で実施すべき活動を図 1-2 のとおり①～⑩の構成要素（例：「① データマネジメント戦略と評価」）に分類し、さらに戦略策定、オペレーション、組織、人材などより大きなプロセスで括って定義している<sup>1</sup>。



出典：JDMC，データマネジメント概説書（JDMC 版）

図 1-2 JDMC が定義するデータマネジメントのライフサイクル【参考】

ただし、これらのプロセスは、元々JDMC のフレームワークが主に民間企業を念頭に作成されたということもあって、必ずしも行政職員にとって理解しやすい形にはなっていない。そもそもデータマネジメントという概念自体、行政職員はもとより民間企業においてもまだ十分に認知されていないし、概念を理解できたとしても、それを実務の文脈に置き換えて理解することはさらに困難である。従って、行政職員にとってより理解しやすいフレームワークを別途、策定することが必要と考えた。

<sup>1</sup> データマネジメントのフレームワークとしては、このほか DAMA (Data Management Assorciation) International が発行している「データマネジメント知識体系ガイド第一版—DMBOK (Data Management Body of Knowledge)」があるが、昨年度調査の段階で、JDMC のフレームワークの方が我が国での適用においてより親和性や実用性が高いと判断し、検討対象から除外している。

## ② 情報システムのライフサイクルに基づくフレームワークの限界

こうしたデータマネジメントのライフサイクルに対し、情報システムのライフサイクルは、少なくとも情報システムに関連する業務に携わった経験がある者にとっては比較的理解しやすく、現場でも認知が浸透している。そこで、昨年度調査では、その枠組みに基づいて図 1-3 のとおり独自のフレームワークを行政機関向けに設定して検討を行っている。

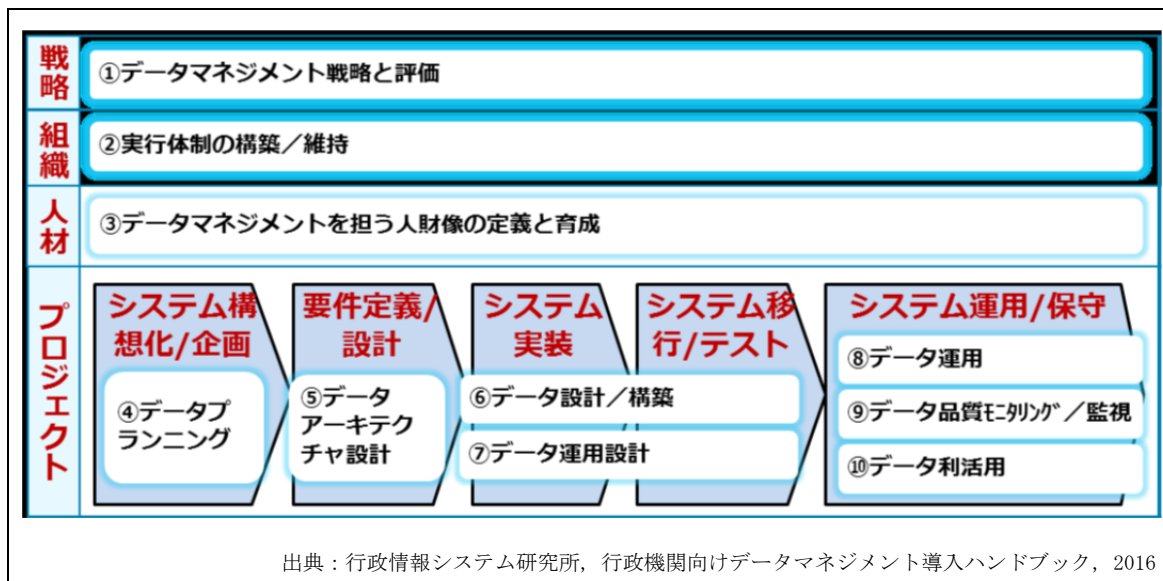


図 1-3 昨年度調査で設定したフレームワーク【参考】

ただし、このフレームワークにも課題がある。すなわち情報システムのライフサイクルを前提としてしまうと、情報システムを前提としないデータのマネジメントを構想することが困難になってしまうのである。また、情報システムありきでプロジェクトが実施されることにより、業務部門におけるデータに対する当事者意識が希薄になってしまうという弊害もある。



### 1.3. 行政データマネジメントのライフサイクルのフレームワーク

このように、(1)データマネジメント固有のフレームワークと(2)情報システムのライフサイクルに基づくフレームワークには、それぞれに長所・短所が存在する(表 1-1)。

表 1-1 JDMC フレームワークと昨年度調査でのフレームワークの比較

	データマネジメント固有のライフサイクルに基づくフレームワーク (JDMC 概説書記載のもの)	情報システムのライフサイクルに基づくフレームワーク (昨年度調査で設定したもの)
プロセスの定義	1. DM 戦略 2. データライフサイクルオペレーション 2-1. 計画 2-2. 仕様確定 2-3. 使用可能化 2-4. 登録 2-5. 維持 2-6. アーカイブ 2-7. 消去、活用 3. DM 組織 4. DM 人材	1. 戦略 2. 組織 3. 人材 4. プロジェクト 4-1. システム構想化/企画 4-2. 要件定義/設計 4-3. システム実装 4-4. システム移行/テスト 4-5. システム運用/保守
長所・短所	× 理解しにくい ○ 情報システムに依存せず、データそのものを軸としてデータマネジメントを構想できる	○ 情報システムに携わったことがあれば比較的理解しやすい × 情報システムに依存しないデータのマネジメントを構想しにくい × 業務部門におけるデータへの当事者意識が希薄化

そこで、本調査研究では、昨年度調査で設定したフレームワークを見直し、(1)データマネジメント固有のフレームワークと(2)情報システムのライフサイクルに基づくフレームワークの両方をつなぐ、共通的なフレームワークを表 1-2 のとおり設定することとした。

また、データマネジメントの活動内容は、大別して、データが使い物になるようにする、いわば「データ品質の確保」の取組みと、行政機関内でのデータの利活用を促進する、いわば「データ利活用の定着化」の2つのテーマに分けられることから、同フレームワークでもこの考え方に沿って個々の取組みを分類・整理することとした。

表 1-2 今回設定したデータマネジメントのフレームワーク

行政データマネジメントフレームワーク				
データマネジメント (DM)のプロセス	1. 戦略・組織	2. 計画・調達	3. 構築	4. 運用 (人材育成含む)
(参考) 対応するシステムライフサイクル	—	システム化構想/企画、要件定義	設計、システム実装、システム移行/テスト	システム運用/保守
DMの主要タスク※	①データマネジメント戦略と評価 ②実行体制の構築/維持	④データプランニング	⑤データ・アーキテクチャー設計 ⑥データ設計/構築 ⑦データ運用設計	⑧データ運用 ⑨データ品質モニタリング/監視 ⑩データ利活用 ③データマネジメントを担う人財像の定義と育成
DMのテーマ				
(a) データ品質の確保				
(b) データ利活用の定着化				

※「行政機関向けデータマネジメントハンドブック」(総論編)p. 15 参照

本フレームワークにしたがえば、後述の第2章で収集する個別のデータマネジメントの活動に関するノウハウは、表1-3のように整理されることとなる。

表 1-3 フレームワークによるノウハウの整理

行政データマネジメントフレームワーク				
データマネジメント (DM)のプロセス	1.戦略・組織	2.計画・調達	3.構築	4.運用 (人材育成含む)
(参考) 対応するシステムライフサイクル	-	システム化構想化/企画、要件定義	設計、システム実装、システム移行/テスト	システム運用/保守
<b>DMの主要タスク</b> (「行政機関向けデータマネジメントハンドブック」p.15参照)	①データマネジメント戦略と評価 ②実行体制の構築/維持	④データプランニング	⑤データアーキテクチャー設計 ⑥データ設計/構築 ⑦データ運用設計	⑧データ運用 ⑨データ品質モニタリング/監視 ⑩データ活用 ③データマネジメントを担う人材像の定義と育成
DMのテーマ	(a) データ品質の確保	1a-1 移行後のデータ品質問題の発生を防止するには 1a-2 データ入カールールの順守を徹底するには 1a-3 システム間連携に伴うデータ形式等の不整合によるトラブルを回避するには	2a-1 データ移行後のデータ格落ちや集計誤り等のトラブルを回避するには 2a-2 システム間でのコード体系不一致を回避するには (連携テスト中に不具合が発生) 2a-3 設計書と実際の機能・データとの乖離に起因する機能設計の手戻りを回避するには 2a-4 データ移行時のフォントや辞書等の権利関係に起因して発生するトラブルを回避するには 2a-5 印影等の画像データの移行に際し、原本性保証に関する解釈の食い違いに起因するトラブルを回避するには 2a-6 コストをかけずに、データの形式違いや表記の揺れを抑制するには	3a-1 人的なマスターデータ入力ミスによる帳票表記間違い等のトラブルを撲滅するには
	(b) データ利活用の定着化	1b-1 データ収集・整備にあたり各部門の協力を得るには 1b-2 業務で取り扱うデータを紙からデジタルに切り替えるには 1b-3 担当者の異動後もデータ活用が組織で継続されるようにするには 1b-4 オープンデータの推進に向けて各部門の協力を得やすくするには 1b-5 部門横断的なデータ利活用の取組についての組織内での正当性を確保するためには 1b-6 各部門の潜在的なデータ活用ニーズを引き出すには 1b-7 オープンデータの利用者である民間企業等の要望やニーズを把握するには 1b-8 公開したオープンデータの民間企業等による利活用を活性化させるには	2b-1 データの発生・入力・利用のタイミングが不明確で、再度手入力する事態を回避するには	3b-1 組織や用途に応じたGISとのデータ連携をバラバラに作らせないためには 3b-2 情報発信サイト等を活性化させるために、各団体からの積極的なデータ提供を促すには

## 2. 行政機関におけるデータマネジメントの実践ノウハウ調査

本章では、行政機関や行政のデータ利活用に関わっている団体等が、それぞれの活動を通じてデータマネジメントの観点でどのような工夫や努力を行っているのか、また、その前提としてどのような課題認識を有しているのかをインタビューを通じて把握・整理する。

### 2.1. 調査の実施方法

現状、データマネジメントという概念は、行政機関はもとより民間企業においても、十分に共通認識となるには至っていない。しかしながら、そもそもデータマネジメントとは、データを利活用するために、データ管理のあり方を企画し、設計し、生成の仕組みを実装し、運用し、利活用していくという、いわば当たり前のことを当たり前に実行する取組みである。したがって、必ずしもデータマネジメントという概念を意識した取組みでなくても、現実に実施されている活動がデータマネジメントそのものである場合がある。

そこで本インタビューでは、データ利活用の成果が公表資料等を通じて公開されており、その内容からデータマネジメントの取組みの存在が推定される事例を収集した上で、前章で整理したフレームワークに従ってある程度網羅的に全体がカバーされるようにインタビュー先を選定した。

その上で、あらかじめインタビューの構成を計画しつつ、話の流れに合わせて話題を拡散あるいは深堀していく、いわゆる半構造化インタビューを実施し、その結果から、データマネジメントに係るノウハウや今後の課題を抽出・整理した。(図 2-1)

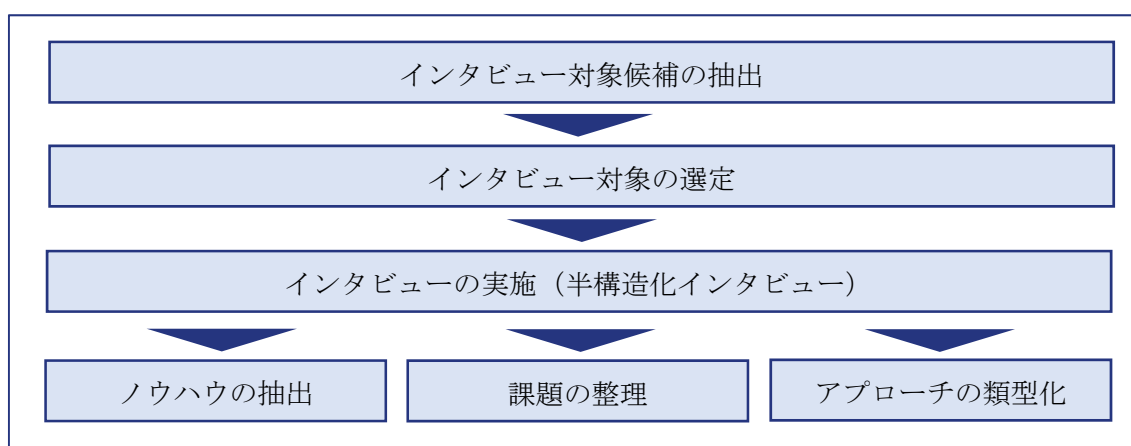


図 2-1 インタビュー調査の流れ

## ① インタビュー対象候補の抽出

以下の手順で公開情報に基づきインタビューの対象候補を抽出した。

### 1) 公開情報の収集

次に示す新聞・雑誌、各団体・企業のプレスリリース、調査報告書等の直近5年間（2012年以降）の公開情報から、データマネジメントと関連があると想定されるデータ利活用事例を、国、自治体および民間企業から商用データベースおよびウェブサイト検索を用いて収集した。

#### a) 新聞・雑誌

(対象) 主要新聞（日経、毎日、読売、朝日）および日経BP（いずれも電子版）

(検索方法) 以下のデータベースを利用

- ・ 「日経テレコン21」：日経4紙記事ほか、業界紙、地方紙を収録。企業、新製品情報ファイルを収録。日経BP出版雑誌、ビジネス誌も同時検索可能。
- ・ 「ELNET」：ビジネス系情報源として、専門新聞（100紙）、専門雑誌（150誌）を収録。日経テレコン21等他の商用DBに収録されていない、流通、通信系等専門性の高い雑誌記事が入手可。

#### b) 団体・企業のプレスリリース

(対象) 中央省庁、自治体、その他行政機関、民間企業のホームページ

(検索方法) 以下の2段階で検索を実施

##### b-1) キーワード検索

google 検索によって「データ利活用」「オープンデータ」「官民連携」「データ連携」をキーワードとして検索。これらでヒットしたものを逐次参照していく過程で、先進的な取り組みを行っている団体や実証事業の名称を抽出した

##### b-2) 目視調査

b-1)で抽出された事例の内容から、関連他事例や取り組みをさらに検索。検索キーワードは団体によって異なる。例えば、青森観光クラウドなら「太宰ミュージアム」、狭山市の取り組みなら「税務システム最適化」等。

#### c) 調査報告書

- ・ 地方公共団体におけるクラウド導入の取組（平成27年度改訂版） 地方公共団体情報システム機構（J-LIS）
- ・ 自治体クラウドグループ一覧（平成28年1月現在） 総務省

## 2) 対象の抽出

上記 1) で収集した情報の中から、データマネジメントに係る特徴的な取組みを行っている好事例、具体的には、各事例における取組み内容の主要部分に、表 1-2 に示す「データマネジメントの主要タスク」のいずれかが含まれるものを抽出した。

以上の結果、インタビュー候補となる以下の事例が抽出された。

表 2-1 抽出された調査対象候補の事例

#	分類	分野	組織・団体名	事例名
1	業務系	観光	観光情報連携プラットフォーム機構	観光クラウド Myルートガイドサービス
2			福井県 鯖江市	データシティ鯖江（総務省実証事業にも参画）
3			裾野市	子育て支援アプリ
4		子育て	横浜市金沢区	かなざわ育ナビ.net
5			川崎市麻生区	「子育て支援アプリ」実証実験
6			福岡市	まちかど安全ガード「けいご君」
7		窓口サービス	埼玉県狭山市	総合窓口対応システムによる、窓口業務の一元化・市民サービスの向上
8		防災	神戸市	神戸市ハザードマップ
9		交通安全	埼玉県	カーナビ情報から急ブレーキ多発箇所の特定
10		インフラ・土地管理	福岡市、佐賀県	情報流通連携基盤の社会資本情報における実証
11			東京都、 国土交通省関東地方整備局	社会資本情報に係る総務省実証実験
12			岡山県真庭市	森林林業クラウドの構築と活用
13		政策立案	富山市	住民基本台帳データに座標情報を付与し政策課題検討
14			浜松市	ソーシャルメディアとアンケートから市民ニーズ分析
15			千葉市	「ちばレポ」による市内の課題把握と共有
16			福岡県 苅田町	庁内 GIS の活用による業務効率化、経費削減
17			大津市	支所統廃合計画策定におけるデータ活用
18	情報基盤系	匿名化	国立がん研究センター	全国がん登録データベース
19			国立長寿医療研究センター	バイオバンク情報管理システム
20		データ統合・集約	中小企業基盤整備機構	小規模事業者統合データベース整備事業
21			静岡市	しずみち info
22		データ提供	姫路市	行政情報分析基盤構築
23			浦安市	統合型 GIS を通じた政策検討とデータの公開
24			倉敷市	高梁川流域圏データポータルサイト・データ分析サロン
25			神戸市	神戸市モバイル空間統計 API
26			国立科学博物館	地球規模生物多様性情報公開
27			国立国会図書館	図書情報オープンデータセット
28		厚生労働省	レセプト情報・特定健診等情報データベース	
29	クレンジング	独立行政法人日本貿易振興機構	顧客情報の一元管理・分析基盤構築	
30	用語統一	経済産業省等	共通語彙基盤	
31	民間・外国	データ利活用	大和ハウス	ココゆれ
32			POLA	美肌県グランプリ
33		農業	富士通	akisai（農業クラウド）
34		オープンデータ		LOD4ALL
35		オープンデータ	EvaCva	
36	オープンデータ	(公財)九州先端科学技術研究所*	OpenGov 他	

※当該団体に対し外国事例についてのインタビューを行ったもの。

## ② インタビュー対象の選定

前節で抽出されたインタビュー候補先について、それぞれの取組内容をインターネット、雑誌、セミナー資料等の公開情報をもとに調査し、行政機関の基本的な業務・サービスの類型（表 2-1 に示す 3 つの分類）、および第 1 章で設定したフレームワークの各プロセスの両方が大きな偏りなくカバーできるよう、表 2-2 のとおりインタビュー先を選定した。

なお、図中の○は、データマネジメントのライフサイクルのうち、該当するプロセスにおいて何らかの取組みが本選定作業を通じて確認されたことを、また、◎は特に特徴的な取組みが確認されたことを示している。

表 2-2 インタビュー対象の選定

区分	インタビュー対象	対象システム・サービス等	データマネジメントのプロセスとの対応			
			1. 戦略・組織	2. 計画・調達	3. 構築	4. 運用 (人材育成含む)
			①データマネジメント戦略と評価 ②実行体制の構築/維持	④データプランニング	⑤データ・アーキテクチャー設計 ⑥データ設計/構築 ⑦データ運用設計	⑧データ運用 ⑨データ品質モニタリング/監視 ⑩データ利活用 ③データマネジメントを担う人財像の定義と育成
国	内閣官房	復興・復旧支援データベース			◎	◎
	内閣府	法人インフォ 公益認定等総合情報システム	○	◎	◎	
自治体	観光情報連携プラットフォーム機構	観光クラウド	○	◎	◎	○
	浦安市	統合型 GIS	◎	○		◎
	狭山市	総合窓口対応システム	○	○	◎	
	埼玉県町村会	自治体クラウド	◎	○		○
	横浜市	オープンプラットフォーム	◎	○		○
	(株)トヨタ IT 開発センター	しず道 info		○	◎	
	富士通(株)	EvaCva	○	○	◎	
広島県	情報戦略総括監					
外国	(公財)九州先端科学技術研究所※	OpenGov 他	◎			◎

※当該団体に対し外国事例についてのインタビューを行ったもの。

### ③ インタビュー項目

インタビュー対象者には、以下の点を軸として各事例における取組内容を踏まえて個別にインタビュー項目を設定した。

- 取組みの概要・インタビュー対象者の役割
- 取組みのデータマネジメント観点での工夫
- 取組みのデータマネジメント観点での課題

(インタビュー項目の一例)

- ①本事例のポリシー、取組内容および貴課の役割について
- 〇〇〇におけるデータの管理や利活用推進のポリシーと具体的な取組内容、その実践に当たっての貴課の役割をご教示ください。
- ②データ品質の確保に向けた取組み
- データ品質を継続的に確保するためのモニタリング等の業務プロセス、体制、規程・ルール等があればご教示ください。
- ③データ項目の追加・変更等への対応
- 利用者からのデータ項目の追加・変更等の要望への対応方法をご教示ください。
- ④データ利活用を推進するリーダーの人材育成および利用者のスキルアップ
- 〇〇〇のデータ利活用を推進するリーダーの人材育成や利用者のスキルアップに向けた取組内容をご教示ください。
- ⑤リソースの確保
- 上記の取組を推進するための人的、予算的リソースをどのように確保したのかをご教示ください。



## 2.2. ノウハウの抽出

インタビューの結果、各事例から確認されたデータマネジメント観点での取組みや工夫を、第1章の表 1-2 で定義されたフレームワークの各プロセスに当てはめて表 2-3 のとおり整理した。なお、各事例の詳細は、「別紙1：インタビューおよび研究会から確認された工夫・ノウハウ」に記載している。

表 2-3 データマネジメント観点での取組み・工夫の抽出

プロセス／観点	機関／事例	取組み・工夫
1. 戦略・組織 / (a) データ品質の確保	内閣府／公益認定等総合情報システム	<ul style="list-style-type: none"> <li>データ入力部門ではなく、データ集計部門（統計担当課）が主体となり、データ形式のチェック等のパトロールを行い、ルール違反者への説得・是正を行った。</li> <li>申請書類毎の審査のボリュームが膨大であり、特に新任者が全体を習得するまで時間がかかることから、間違い発生の予防および教育的配慮から入力チェックとナビゲーションのロックを念入りに作り込んだ。</li> </ul>
1. 戦略・組織～ 2. 計画・調達 / (b) データ利活用の定着化	浦安市／統合型GIS 内閣官房／復興・復旧支援データベース	<ul style="list-style-type: none"> <li>各部門はデータを抽出・作成するところまでを実施し、データ形式の変換やフォーマットへの変換・整形は情報政策課等が行う役割分担とした。</li> </ul>
1. 戦略・組織 / (b) データ利活用の定着化	内閣府／公益認定等総合情報システム	<ul style="list-style-type: none"> <li>文書管理規定に例外として独自規定を設け、システムに入力されたデータを原本とした。</li> <li>電子化および検索機能や差分確認機能などの活用により、変更箇所の特が容易になり、補正作業時の職員の労力が軽減されるなど電子化のメリットを前面にアピールした。</li> </ul>
1. 戦略・組織 / (b) データ利活用の定着化	浦安市／統合型GIS	<ul style="list-style-type: none"> <li>GIS の取組事例を行政機関内部だけでなく対外的に積極的な PR を行うことで、後に引けない状況を作った。</li> <li>取組事例を有名にした結果、他自治体からの問合せや視察等が増加し、担当者がその説明対応を行うことで意識が高まった。</li> <li>対外的な PR 活動に後任者も同行させ、意識の醸成を図った。</li> </ul>
1. 戦略・組織 / (b) データ利活用の定着化	横浜市／オープンプラットフォーム	<ul style="list-style-type: none"> <li>局長・部長級も巻き込んだ庁内横断的な会議体で、オープンデータ指針を1年間掛けて検討・合意することで、全庁的な体制を構築した。</li> </ul>

プロセス／観点	機関／事例	取組み・工夫
		<ul style="list-style-type: none"> <li>データ利活用やデータマネジメントの推進を IT 部門ではなく、政策部門が担う体制としたことで、利活用の出口を起点とした検討を推進した。</li> </ul>
1. 戦略・組織 / (b) データ利活用の定着化	浦安市／統合型 GIS	<ul style="list-style-type: none"> <li>情報政策課の事務分掌規程に、「GIS の利活用推進」を明記した。</li> </ul>
1. 戦略・組織 / (b) データ利活用の定着化	浦安市／統合型 GIS	<ul style="list-style-type: none"> <li>情報政策課が各部門の GIS 利活用事例や要望を全庁で収集・共有した。</li> <li>各部門からデータ連携に関わる相談があれば、情報政策課がアドバイスを行った。</li> <li>その中で実現に向けて動き出すものがあつた場合は、情報政策課が内部連携部門としての役割を担った。</li> </ul>
1. 戦略・組織～ 2. 計画・調達 / (b) データ利活用の定着化	横浜市／オープンプラットフォーム	<ul style="list-style-type: none"> <li>民間企業等からのオープンデータに係る期待や要望等を把握するため、民間企業等との窓口を一本化した。なお、その役割を担うのは、幅広い視野で課題や政策要件等を俯瞰した上で、どこをポイントとし、どことタイアップすることでイノベーションを起こせるか氏は、データマネジメントを実施している政策部門でないが無理である。</li> </ul>
1. 戦略・組織～ 2. 計画・調達 / (b) データ利活用の定着化	横浜市／オープンプラットフォーム	<ul style="list-style-type: none"> <li>民間企業や学生等を巻き込み、地域振興につながるデータ利活用を目的に、アイデアソンやハッカソン等の取組を継続的に実施した。</li> <li>アイデアソンやハッカソン等のイベントでの優秀作を表彰した。</li> <li>行政側の制度と連携して、優秀作のビジネス化を支援した。</li> </ul>
2. 計画・調達～ 3. 構築 / (a) データ品質の確保	内閣官房／法人インフォ	<ul style="list-style-type: none"> <li>今後のデータ入力にあたっては業界標準データ体系に合わせることをルール化した。</li> </ul>
3. 構築～4. 運用 / (b) データ利活用の定着化	浦安市／統合型 GIS	<ul style="list-style-type: none"> <li>情報政策課が各部門で GIS とどのような連携がなされているかを一覧で整理し、全庁で共有した。</li> <li>GIS と連携しているデータ一覧の変更があれば、各部門から情報政策課に連絡がくるようにした。</li> </ul>
3. 構築～4. 運用 / (b) データ利活用	内閣官房／復興・復旧支援デー	<ul style="list-style-type: none"> <li>支援制度サイトへの各自治体のデータ入力・更新状況を RSS 等で他自治体にも公開し、データ</li> </ul>

プロセス／観点	機関／事例	取組み・工夫
の定着化	タベース	が常に新しくなり、流れ続けていることが見えるようにした。
4. 運用 / (b) データ利活用 の定着化	浦安市／統合型 GIS	・各部門が GIS の利活用を推進する担当者を指名し、その担当者に対して、情報政策課が年に 1～2 回、GIS の利活用に向けた研修を実施した。

### 2.3. ノウハウ探索のためのタイトルおよび問題・課題の設定

上記の取組み・工夫をノウハウとして活用するにあたっては、各プロセスにおける個別具体的な課題から、該当する取組み・工夫を発見できるよう整理することが必要である。そこで、表 2-4 のサンプルに示すような形で、両者をつなぐための「タイトル」や「問題・課題」を整理することが有効である。

表 2-4 ノウハウ探索のためのタイトルおよび問題・課題の設定<サンプル>

プロセス／観点	タイトル	問題・課題	取組み・工夫
1. 戦略・組織 / (a) データ品質の確保	データ入力ルールの順守を徹底するには	・入力フォームを統一しても、入力ルールを無視する入力者がいるため、データ品質が確保されておらず、正確な統計データにするために膨大な手修正が必要となった。	<ul style="list-style-type: none"> <li>・データ入力部門ではなく、データ集計部門（統計担当課）が主体となり、データ形式のチェック等のパトロールを行い、ルール違反者への説得・是正を行った。</li> <li>・入力者自身が入力ルールに違反していることを認識し、是正したくなるよう、ルールを分かりやすくした。</li> </ul>
1. 戦略・組織 / (b) データ利活用の定着化	部門横断的なデータ利活用の取組についての組織内での正当性を確保するためには	・情報政策課が推進するデータ利活用の取組に、各部門の理解・協力を取り付けることが難しかった。	<ul style="list-style-type: none"> <li>・情報政策課の事務分掌規程に、「GIS の利活用推進」を明記した。</li> </ul>
2. 計画・調達～ 3. 構築 / (a) データ品質の確保	コストをかけずに、データの形式違いや表記の揺れを極力抑制するには	・各部門からのデータは、データ形式や表記の揺れ等があったため、データ移行前に手作業でデータを修正する必要があった。	<ul style="list-style-type: none"> <li>・外部団体で策定された業界標準データ体系に、各部門のデータ形式を合わせた。</li> </ul>
3. 構築～4. 運用 (人材育成含む) / (a) データ品質の確保	人的なマスタデータ入力ミスによる帳票表記間違い等のトラブルを撲滅するには	・職員によるマスタデータの入力ミスにより、誤った帳票出力が発生した。	<ul style="list-style-type: none"> <li>・職員一人によるマスタデータの入力及びチェックではなく、複数職員によるマスタデータのダブルチェックなどの運用ルールを設けた。</li> </ul>

3. 構築～4. 運用 (人材育成含む) / (b) データ利活用の定着化	<b>情報発信サイト等を活性化させるために、各団体からの積極的なデータ提供を促すには</b>	<b>・自治体横断の住民向けの支援制度サイトを構築したが、各自治体の Web サイトにも支援制度に係る情報を公開していることもあり、支援制度サイトへのデータ入力・更新頻度が少なかった。</b>	・支援制度サイトへの各自治体のデータ入力・更新状況を RSS 等で他自治体にも公開し、データが常に新しくなり、流れ続けていることを見せた。
---------------------------------------	--	--	---

## 2.4. 課題の整理

データマネジメントを取り巻く課題については、昨年度調査では、データマネジメントに関して問題意識を有する行政職員や CIO 補佐官に対して行ったグループインタビュー等において、以下のような点が指摘されている。

- ・当事者意識が希薄になりやすい
- ・わかりやすい目標設定がしにくい
- ・取組みの必要性を理解するのが難しい
- ・危機意識の乏しさ、保守的な組織文化の影響から取組みへの動機づけが難しい

本年度の調査では、自治体や行政サービスの提供に関わっている民間企業等にもインタビューの対象範囲を広げ、改めて課題の抽出・整理を行った。その結果、以下のような課題が確認された。

### <抽出された課題>

1. 予算や人員が限られる中で、さらに政府からのマイナンバー対応、ネットワーク強化等のセキュリティ対応に追われ、新しい知識を修得するために必要となる十分な時間を確保することが難しい。(広島県)
2. ICT の専門知識を修得するための教育体系・キャリアパスが整備されておらず、職員の ICT スキルを組織的に育成していくことが難しい。(広島県)
3. ほとんどの職員は 2～3 年サイクルで異動となるため、そのことが ICT の知識・ノウハウの継承を阻害するひとつの要因となっている。(広島県)
4. 各原課にとってデータ提供は追加の事務のため、その負荷を低減させる役割分担と仕組みが促進のためには必要 (浦安市、観光情報連携プラットフォーム機構)。
5. 情報システム部門だけでオープンデータ活用はできない。市議会を巻き込むなど庁内横断的な議論が成功の鍵。ハッカソンなどによる民間の活用もある (横浜市)。
6. 利活用の人材育成は、全庁的かつ継続的に行わないと、その場限りになってしまい、モチベーションの維持ができない (浦安市、横浜市)。
7. 利用部門や団体に対しては、効果と合わせて費用負担方法についての十分な説明が重要。(埼玉縣市町村会)
8. 行政機関は紙の文化。データを電子化するだけでも高いハードルがある。(内閣府)

これらの課題はいずれもデータマネジメントに大きな影響を与える要因ではあるものの、以下に示すようなより根深い要因に起因する課題であり、データマネジメントのライフサイクルの中での単独の取組みとして対処するには限界がある。

- a. 予算的・人的資源の不足
- b. ICT人材育成制度の不十分さ
- c. 人事制度の制約
- d. 組織間の役割分担・連携の壁
- e. 組織文化の影響

他方で、これらの課題への取組みも広義のデータマネジメントの課題の一環として捉え得るものである。データマネジメントを実践し、データの利活用を通じた課題解決を推進するためには、これらの課題を避けて通ることはできない。今後は、予算制度や専門人材の育成制度、人事制度、組織体制、職員教育などの組織全体の取組みに対し、データマネジメントの観点から見直しを促していくことも必要になると考えられる。

## 2.5. データマネジメントのアプローチの類型

インタビューを通じて抽出された、各機関におけるデータマネジメントに関する工夫・努力には、大きく以下の3つの方向性があることが確認された。

### ① 自主努力

自らの組織が主導してデータマネジメントに取り組むパターンである。

②共通化や③外部化のパターンの取組みを行う場合であっても、まずは自組織としてのデータマネジメントの取組みへのコミットメントが不可欠であり、そのイニシアチブの下、②共通化や③外部化が行われることとなる。その意味では、①～③いずれのパターンであっても、常にこの自主努力がデータマネジメントの取組みすべての基礎となる。

### ② 共通化（共通仕様等準拠）

標準的な仕様、あるいは共同利用型のサービスに自組織のデータの一部を合致させていく方向での努力である。例えば、

- a. データ項目を地域情報プラットフォーム標準仕様（APPLIC）に準拠して設定する

- b. 語彙を IMI (Infrastructure for Multilayer Interoperability:情報共有基盤)<sup>2</sup>の共通語彙基盤に準拠して定義する
- c. API 仕様を地域の共通的なサービスプラットフォームに準拠して設計する
- d. オープンデータを Linked Open Data の形式で公開する

といった取組みである。住民や企業・団体の視点に立って組織間での縦割りを排除し、ワンストップサービスを実現するためにも、また、組織間での重複投資を回避するためにも、データの共通化の検討は不可欠となりつつある。

### ③ 外部化（外部サービス調達等）

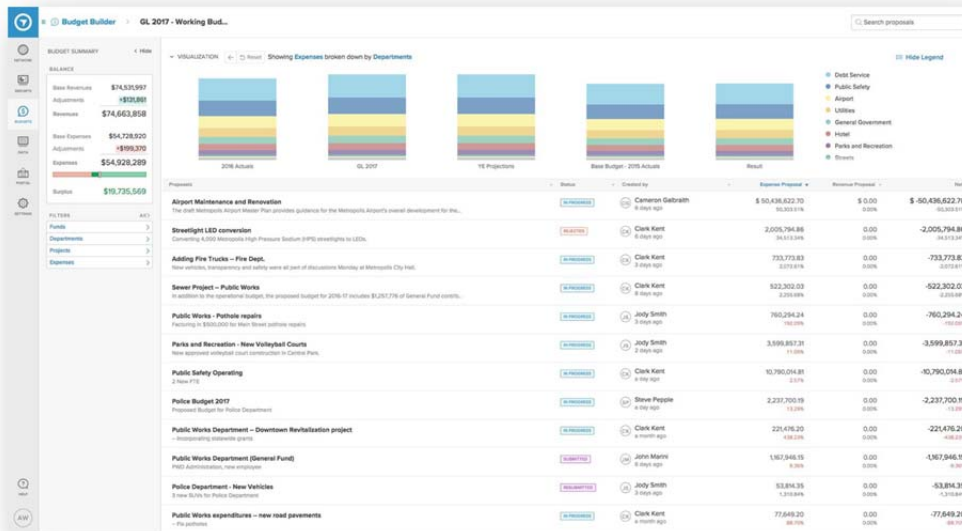
米国等の諸外国でみられる取組みのパターンである。行政機関自らはデータマネジメントを行わず、データの利活用までの一連の作業をまとめて民間事業者へ委託する。例えば、OpenGov の例では、自治体は、予算に関する資料を紙媒体のものも含めて一括して事業者へ渡す。事業者はそれを体系的に整理し、予算書等も含めた集計・分析サービスとして自治体に提供する。これにより、自治体側においては、

- ・ 人件費も含めたトータルコストで見れば行政コストを抑えることができる
- ・ データ分析や可視化などの機能も行政独力の場合よりも高度化できる
- ・ 同サービスを利用する他の自治体との間で共通化・標準化も自然にできてしまう

といったメリットが得られる。現時点で、我が国でこれと同様のサービスとして提供されているのは、オープンデータを LOD 形式に変換するサービスなどかなり限定された範囲にとどまっている。

職員採用や予算が硬直的な縦割構造となっている我が国行政機関では、トータルコストでの優位性で判断するという発想には理解が得られにくいだが、将来的な方向性としては展開の可能性がある取組みと考えられる。

<sup>2</sup> <https://imi.go.jp/>



<https://opengov.com/budgeting-and-planning>

図 2-2 参考：OpenGOV が提供している予算書の作成・分析サービス

これらの3つの方向性の全体像を示したのが図 2-3 である。自主努力を基本としながら、外部標準等は「2. 共通化」として業務・システムの内部で準用する一方、民間企業等でも実施可能な業務は「3. 外部化」してゆく姿を図示したものである。なお、これらはいずれか一つを選択するというものではなく、複合的に取組みが行われるのが一般的である。

以降では、「自主努力」を前提に検討を進めるが、「共通化」や「外部化」であってもデータマネジメントとして実施すべき取組みは異なることはない。異なってくるのは誰がその取組みを実施するかという点のみである。

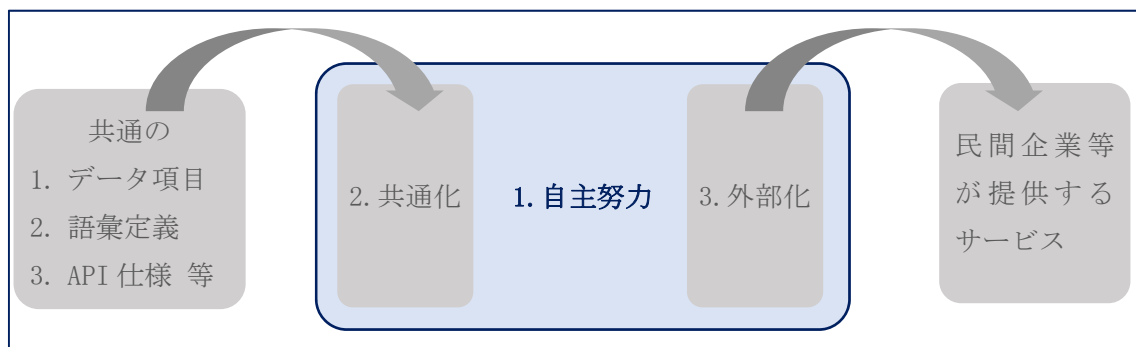


図 2-3 データマネジメントの取組みの3つの方向性

### 3. JDMC 研究会における IT 調達時のデータマネジメント課題研究

我が国の行政機関では、データマネジメントの実践は実態として、情報システムの構築と運用保守を担当している IT ベンダーが多くの部分を担っている。本章では、JDMC の協力により、我が国主要 IT ベンダーの参画を得てグループインタビューを実施し、これらの企業が認識している、行政機関におけるデータマネジメントに関する現場課題の抽出と求められる対応策の考察を行う。

#### 3.1. 検討の実施方法

JDMC の研究会として、本件課題のためのサブワーキンググループ（以下「研究会」という。）を設置し、政府・行政機関における情報システム調達に関する課題に知見の深い有識者の参画を得て、グループインタビューを行った。

##### ① 研究会の構成

研究会には、JDMC の理事会員企業であり、各研究会や各部会でデータマネジメントに関する積極的な活動をしている(株)エヌ・ティ・ティ・データ、日本電気(株)、(株)日立製作所、富士通(株)の 4 社に委員として参加いただいた。

役割	構成員
委員	(株) NTTデータ 第一公共事業本部e-コミュニティ事業部第一ビジネス統括部 部長 松尾 英史 (株) NTTデータ 第一公共事業本部第二公共事業部営業部 部長 渡邊 靖隆 日本電気(株) 第一官公ソリューション事業部行政情報ソリューショングループ 営業部長 新井 隆 日本電気(株) 第一官公ソリューション事業部 シニアエキスパート 青木 秀一良 (株) 日立製作所 全国公共システム第三本部公共システム推進第一部 主管 松田 純一 (株) 日立コンサルティング ディレクター 小池 博 富士通(株) 公共営業本部官庁第二統括営業部第一営業部 部長 押鐘 快之 富士通(株) 公共地域・公共・地域ビジネス推進統括部商談戦略推進室 室長 山本 晃司 <div style="text-align: right;">計 8 名</div>



ファシリテーター	一般社団法人 行政情報システム研究所 主席研究員 狩野英司
事務局	一般社団法人 日本データマネジメント・コンソーシアム事務局 (株) リアライズ 代表取締役社長 大西浩史

## ② 研究会の開催実績

研究会の開催実績および各回の議題は以下のとおりである。

開催回	開催日時	主な議題
第1回	2017年1月20日	<ul style="list-style-type: none"> <li>・サブWGの趣旨について</li> <li>・行政機関向けデータマネジメント導入ハンドブックについて</li> </ul>
第2回	2017年2月24日	<ul style="list-style-type: none"> <li>・行政のデータマネジメントに関する問題点および留意すべき事項について</li> </ul>
第3回	2017年3月13日	<ul style="list-style-type: none"> <li>・成果取りまとめの方向性について</li> <li>・今後のハンドブック等の改善・拡充のあり方 等</li> </ul>

## ③ 検討の進め方

ITベンダーが認識している行政機関におけるデータマネジメントに係る課題の多くは、現行のシステムの機能要件やデータの整備状況等が、発注者側から受託者側に十分に伝達されていないことに起因すると考えられる。その結果、例えば、のちのテスト工程等に至って初めて想定外の要件や制約条件が明らかとなり、プロジェクトの問題が表面化することとなる。したがって、調達前にこれらの情報を把握し、調達仕様書に適切に反映させることが、スケジュール遅延などの問題発生未然防止やシステムの品質確保につながり、ひいては受託者のみならず発注者にもメリットをもたらすことになる。

研究会では、この仮説のもと以下の様式にしたがって情報システム調達時に受託者であるITベンダーが実際に直面した問題事象を洗い出し、その未然の防止に向けて、特に「(7)留意事項・具体的な対策」を中心に発注者側で事前に取り組むべきこと、留意すべきこと、また、調達仕様書に明記すべきこと等について議論を行い、行政機関及びITベンダー双方にとってより効果的な対策や留意事項の具体化を図った。

表 3-1 問題事象の洗い出しで使用した様式

#	回答項目	回答内容
(1)	問題の発生場面	以下のうち該当するものを選択。 ・単独のシステム新規開発・更改の場合[単独] ・複数のシステムを連携させてデータを利活用するシステム構築の場合は[連携]
(2)	構成要素の該当箇所	データマネジメントの構成要素におけるどの領域で発生した問題なのかを選択。
(3)	問題事象	発生した問題事象を記載。
(4)	課題認識の観点	以下のどちらの観点から課題として認識したのかを記載。 (1)データを利活用できるようにするため (2)ベンダー側への過剰なリスク負担を回避するため
(5)	発生した事象	発生した問題事象の具体を記載。
(6)	問題の原因	問題が発生したそもそもの原因を記載。
(7)	留意事項・具体的な対策	どのようなことに留意していればその問題の発生を抑止または回避できたか、また可能な範囲で具体的な対策を記載。 例:「調達仕様書に〇〇〇を記載する」「行政機関側であらかじめ△△を事前調査しておく」等

### 3.2. 抽出された問題の種類

研究会で指摘されたのは大きく以下の a～d の 4 つの事象である。これらは表 3-2 に示す ①～⑭の具体的問題から構成される。

- a. データの品質：システム更改時の移行元データの不備やデータの表記方法の不統一といった、データそのものの品質が確保されていないことにより発生する事象
- b. データの管理：データの責任所在が不明確、データの運用管理ルールの未整備といった、データ管理が十分に行われていないことにより発生する事象
- c. データの運用：人事異動後の引継ぎが不十分、人的な入力ミスといった、主に人的要素により発生する事象
- d. データの権利関係・原本性：外字フォントや辞書等の権利関係、データの原本性の制約等により発生する事象

表 3-2 問題の種類

種類	具体的事象	事象の内容
a.データの品質	① 移行元データの不備	データの桁落ちや欠損といった元データの品質に起因する問題。補間プログラムの作成といった想定外の作業が発生している。
	② データ管理手法の相違	新旧システム間あるいは連携システム間における、データ管理手法(DB管理方法、マスタデータの定義、取り扱う外字等)の相違に起因する問題。
	③ データ表記方法の不統一(表記ルールの未統一)	新旧システム間あるいは連携システム間における、データの入力方法が不均一であることに起因する問題。 データ移行時、データ連携の統合テスト時に発覚して、作業の手戻りが発生している。
	④ 固有のロジックの混在	実態の運用と合わせるために、暫定的に固有のロジックが組み入れられ、またそれら特殊な事情が設計書等に反映されていないことに起因する問題。
	⑤ ドキュメントと実態の差異	旧システムの既存ドキュメントに記載されている内容と、実態のシステムの機能、データ定義との間に差異があることに起因する問題。本番テスト時にデータ自体の不備、データ項目・構造の矛盾が発覚し、設計の見直し等につながっている。
b.データの管理	⑥ データの責任所在の不明確	データの移行時のデータについて、明確な承認者が定められていないことに起因する問題。 移行後にデータの不整合発覚につながっている。
	⑦ 運用ルールの未整備	連携システム間で、マスタデータ等共通データの更新・削除に係る運用ルールが未整備であることに起因する問題。 共通データが更新・削除されることが、連携システムの不具合につながっている。
	⑧ 開発事業者間の情報共有不足	変更内容について、連携先となる開発事業者と情報共有が十分に行えていないことに起因する問題。 データ不整合が発生して、処理の異常終了等につながっている。
	⑨ セキュリティポリシーの相違	セキュリティポリシーの考え方の相違に起因する問題。万一の際の責任の所在が不明確となっている。
c.データの運用	⑩ 移行後の運用想定誤り	移行後における、利用者範囲の拡大に伴うデータ量、ユースケースの変更等の想定が、実態と乖離していることに起因する問題。
	⑪ 移動時の引継ぎ不十分	設計書等に反映されていない暫定的な処理方法について、異動時に十分に引き継がれていないことに起因する問題。
	⑫ 人的な運用ミス	マスタデータの入力誤りといった人的な運用上のミスに起因する問題。
d.データの権利関係、原本性	⑬ 権利関係の制約	外字フォント、辞書等の権利関係の制約に起因する問題。移行の際に想定外の作業が発生している。
	⑭ 原本性の解釈の相違	原本性の解釈の相違に起因する問題。 そのままデータ移行が行えず、紙台帳からデータを取りなおすといった想定外の作業が発生している。

### 3.3. 個別の問題と対応の考え方

以下、表 3-3～3-16 において、前節で示した「①移行元データの不備」～「⑭原本性の解釈の相違」の類型ごとに、研究会での議論を通じて確認された問題事象を「発生した事象」として抽出・整理するとともに、それらが発生した原因を「問題の原因」として整理した。さらに、それぞれの問題事象に対し、行政側として取り組むべき対応を「留意事項・対策の想定」として整理した。

#### ①移行元データの不備に起因する事象

データの桁落ちや欠損といった元データの品質に起因する事象。補完プログラムの作成といった想定外の作業が発生する場合がある。

表 3-3 移行元データの不備に起因する事象

No.	発生した事象	問題の原因	留意事項・対策の想定
①-1	桁落ちした氏名等が連携され、連携先で不一致が発生し、当該月の処理締切りに間に合わず、翌月処理に回るものが発生した。	マスタデータ及びトランザクションデータの移行で、トランザクションデータのみ桁落ちした氏名が一部登録されていた。	現行システムのデータ品質の把握、データクリーニングが望ましい。
①-2	データ移行時に移行プログラムの追加が発生した。 本稼働後にデータの修正、再印刷が発生した。	データ移行時に一部の項目に欠損が判明、補完するためのプログラムの作成が発生した。加えて、補完方法の仕様に考慮漏れがあり、集計時に誤った金額を算出、除外対象も発生した。	現行システムのデータ品質の把握、データクリーニングが望ましい。

## ②データ管理手法の相違に起因する事象

新旧システム間あるいは連携システム間における、データ管理手法（DB 管理方法、マスタデータの定義、取り扱う外字等）の相違に起因する事象。

表 3-4 データ管理手法の相違に起因する事象

No.	発生した事象	問題の原因	留意事項・対策の想定
②-1	旧システムから新システムへのデータ移行時に、プライマリーキー重複が発生	旧・新システムのデータ管理手法の違いを事前に調査する工程が不足。	システム開発の調達、企画段階で発注者が調査することが望ましい。または、開発契約の範囲にデータ調査&設計を含めることも望ましい。
②-2	税情報の連携（税→国保など）にあたり、全く同じ項目名について解釈が異なっていた（一レベル下の項目をあらかじめ合算するのかもしれないかといった内容）ため、連携したあとに国保料を計算した結果に誤りが生じた。	仕様書におけるデータ項目説明が十分ではなかった。	データ項目説明は、項目そのものの意味的な説明だけでなく、その項目にセットするデータの生成過程（いつ、誰（何）が、何をを用いて、どのように）を明示する必要がある。特に、日付、金額については注意が必要。
②-3	参照する他システムのマスタ項目不備。総合テスト段階で、複数のシステムの連動テストにより発覚するケースが多い。 （例） 開発対象システムでは必須の項目が、連携されてくるマスタで任意または省略可となっている。想定外の「NULL」が設定されている等。	要件定義、データ設計前の既存データの調査工程が不足。	システム開発の調達、企画段階で発注者が調査することが望ましい。または、開発契約の範囲にデータ調査&設計を含めることが望ましい。
②-4	システム間で使用する外字コードが不一致。総合テスト段階で、複数のシステムの連動テストにより発覚するケースが多い。 （例） 住所や氏名で外字が使用されており、受信したシステム側で文字化けが発生する。	要件定義、データ設計前の既存データの調査工程が不足。	システム開発の調達、企画段階で発注者が調査することが望ましい。または、開発契約の範囲にデータ調査&設計を含めることが望ましい。
②-5	県、市、町字等を分けて管理しているシステムと分けずにベタ書きしているシステムの連携時に齟齬が発生した。また、「方書き」の扱いがまちまちで齟齬が発生した。	事前にデータ形式の摺合せをしていなかった。	調達仕様書に住所の管理方法について明記することが望ましい。
②-6	氏と名が分かれている/分かれていないケース、分かれていないが間に全角/半角スペースが入るケースなどがあり、齟齬が発生した。	事前にデータ入力時の制約事項を確認しなかった。	調達時に、データ入力の運用手順を整理しておくことが望ましい。

### ③データ表記方法の不統一に起因する事象

新旧システム間あるいは連携システム間におけるデータの入力方法が不均一であることに起因する事象。データ移行時、データ連携の統合テスト時に発覚して、作業の手戻りが発生している。

表 3-5 データ表記方法の不統一に起因する事象

No.	発生した事象	問題の原因	留意事項・対策の想定
③-1	各地方拠点に存在するサーバーを統合するケース等で発生。 各拠点では、歴史的になじみのある呼称や形式を使用しているが、統合システム側では JIS 規格を使用しているケースなどで、同一住所が別物と判断される。	旧・新システムのデータ表記方法の違いを事前に調査する工程が不足。	システム開発の調達、企画段階で発注者側が調査する必要あり。または、開発契約のスコープにデータ調査&設計を含める。
②-2	Null、空白、あらかじめ決められた初期値、などのルールが不明確。生年月日不詳(年しかわからない等で一部不詳の場合もあり)の場合に「99999999」を指定するようなケースもある。本稼働後にデータの修正、再印刷が発生した。	要件定義における調査が不十分であった。	調達仕様書のデータ項目の説明に加えていただきたい。
④-3	1未満の数の扱い方法に関する認識齟齬。総合テスト段階で、複数のシステムの連動テストにより発覚するケースが多い。 (例) 0.8の場合の整数部の表現方法を省略するかどうか。 「.8」または「0.8」の2通りの設定が可能。	要件定義、データ設計前の既存データの調査工程が不足。	システム開発の調達、企画段階で発注者側が調査する必要あり。または、開発契約のスコープにデータ調査&設計を含める。
④-4	コード値の属性に関する認識齟齬。総合テスト段階で、複数のシステムの連動テストにより発覚するケースが多い。 (例) コード値に含まれるハイフンの形式が多数存在する。「半角-」または「全角-」または「長音-」 (例) コード値に含まれるカナが「半角」または「全角」の両パターンが存在する。	要件定義、データ設計前の既存データの調査工程が不足。	システム開発の調達、企画段階で発注者側が調査する必要あり。または、開発契約のスコープにデータ調査&設計を含める。
④-5	氏名、通称名、併記氏名の扱いがまちまち。重複入力が必要かどうかで齟齬が生じることがあった。	事前にデータ入力時の制約事項を確認しなかった。	調達時に、データ入力の運用手順の整理をしておいてほしい。
④-6	意味が同じなのに名称が異なり連携に手間がかかることがあった。例えば、「納付組合」と「納組」のような略称を用いるケース、「給付額」「給付費」のように表記が異なるケース、や「受付日」「申請日」のように異なる名称を用いるケースがあった。	仕様書の項目説明の内容が不十分であった。	調達仕様書の項目説明は丁寧かつ具体的にしたい。

#### ④固有のロジックの混在に起因する事象

実態の運用と合わせるために、暫定的に固有のロジックが組み入れられ、またそれら特殊な事情が設計書等に反映されていないことに起因する事象。

表 3-6 固有のロジックの混在に起因する事象

No.	発生した事象	問題の原因	留意事項・対策の想定
④-1	過去の法制度改正時に、管理が必要となったデータについて、暫定的に備考欄、メモ欄のような予備的なデータ項目を利用する対応をしていた。そのことが設計書等に明記されていない状態でその後仕様変更などの対応を行い、日付を遡った処理のときに不具合が発生した。	法制度改正時（期間、補助金等の制約あり）の暫定的な仕様について、ドキュメントに明示されていなかった。	暫定対応のまま運用する場合は、それを正式な対応としてドキュメント化するよう IT 事業者に指示することが望ましい。
④-2	計算結果の端数の処理に顧客特有のロジック（従来との整合性をとるため）があり、仕様書に現れない特殊事情が考慮できなかった。特に、税額計算や給与計算の訴求処理等で生じる。	調達前に、計算ロジックに関わるリスクの検討が不十分であった。	調達前に、特殊事情の説明が必要。

#### ⑤ドキュメントと実態の差異に起因する事象

旧システムの既存ドキュメントに記載されている内容と、実態のシステムの機能、データ定義との間に差異があることに起因する事象。本番テスト時にデータ自体の不備、データ項目・構造の矛盾が発覚し、設計の見直し等につながっている。

表 3-7 ドキュメントと実態の差異に起因する事象

No.	発生した事象	問題の原因	留意事項・対策の想定
③-1	既存システムの設計ドキュメントに基づいて新システムの設計を実施したが、識別番号自体は一意であればよいという認識から、新システム切替を契機に新しい体系で番号を採番する方針で設計を進めたところ、既存システムでは番号自体に意味を持たせていた場合があること（何桁目が9であれば〇〇等）が後から発覚し、大きな手戻りとなった。	一つの項目に複数の意味（一意な識別番号以外の意味）を持たせていた。また、現行システムの仕様書において、項目の詳細な意味が明示されていなかった。	本質的には、データ設計においてきちんと正規化をすることが対策になる。 新システムの調達前に、発注者としてデータの意味、使われ方を調査してドキュメントに明記することが重要。
⑤-2	現行システムのドキュメントに基づいて設計を行ったが、実際の機能・データとドキュメント内容に差異があり、本番テストで機能不足、データ自体の不備、データの項目、構造の矛盾が発覚し、設計の見直しが発生した。	現行システムのドキュメント等が適正に更新されておらず、実現機能の要件を明確にする場合、事前に実機による機能要件の確認が必要。	発注者側（コンサルを含む）で機能要件について確認し、設計書等のドキュメントを不備なく提示する必要がある。
⑤-3	設計書に記載の内容と実際のシステムの動きが異なり、データの不備が発生した。ソースを見ないと事実が分からず、想定外の工数がかかった。特に、微改修を重ねているシステムでは、設計書の改訂が追いついていないケースがあった。	改修を重ねる際にドキュメントが反映されない、予算の関係で省略する、などの原因がある。	調達前に、設計書の品質を事前に知らせることが望ましい。

⑥データの責任所在の不明確に起因する事象

データの移行後のデータについて、明確な承認者が定められていないことに起因する問題。

移行後にデータの不整合発覚につながっている。

表 3-8 データの責任所在の不明確に起因する事象

No.	発生した事象	問題の原因	留意事項・想定される対策
⑥-1	新システム移行後のデータの正しさを、誰も承認せずに（責任を持たずに）移行を完了したため、稼働後の運用の中でデータの不整合が多数発覚した。	移行後のデータの正しさを、どのように判断するかの基準が不明確であったため、最終確認が曖昧なまま、稼働してしまった（データの管理、責任）	移行後のデータの確認基準は、IT事業者だけでは策定できないため、発注者、利用する各部局の利用者が一体となって、実運用に即して判断基準を策定する必要がある。 また、運用データは誰のものか（誰が責任を持つのか）を明確にする。

⑦運用ルールの未整備に起因する事象

連携システム間でマスターデータ等共通データの更新・削除に係る運用ルールが未整備であることに起因する問題。共通データを更新・削除することで、連携システムの不具合につながっている。

表 3-9 運用ルールの未整備に起因する事象

No.	発生した事象	問題の原因	留意事項・想定される対策
⑦-1	あるシステムからマスターデータのメンテナンスを実施したところ、別のシステムで不具合が発生する。 （例）論理削除、物理削除の概念の相違。 データの有効期間に関する概念の相違。適用開始・終了の考え方の相違等。	開発元が異なるシステム間での共通データの運用ルールの欠如。	システム開発の調達、企画段階で発注者側が調査する必要がある。



⑧開発事業者間の情報共有不足に起因する事象

変更内容について、連携先となる開発事業者と情報共有が十分に行えていないことに起因する問題。データ不整合が発生して、処理の異常終了等につながっている。

表 3-10 開発事業者間の情報共有不足に起因する事象

No.	発生した事象	問題の原因	留意事項・想定される対策
⑧-1	データの編集、出力処理で、処理が異常終了した。	連携中のデータ内容の桁数が変更されたが、連携先に伝達されておらず、連携先が未対応で処理が異常終了	連携に係るデータ変更時の連絡ルールの整備が必要。

⑨セキュリティポリシーの相違に起因する事象

セキュリティポリシーの考え方の相違に起因する問題。万一の際の責任の所在が不明確となっている。

表 3-11 セキュリティポリシーの相違に起因する事象

No.	発生した事象	問題の原因	留意事項・想定される対策
⑨-1	発注者と受託者で、データ取扱いのセキュリティレベルの認識に相違がある。 (例) 本番データを取り扱い時の事前の手順承認の要否、作業時の発注者の立会の要否などの考え方に相違がある。	情報セキュリティポリシーに対する基本的な姿勢の相違が原因。発注者側のセキュリティ意識の低さに起因することが多い。	保守作業などの契約前の、セキュリティポリシーの確認、契約事項への盛り込みが必要。

### ⑩移行後の運用の想定誤りに起因する事象

移行後における、利用者範囲の拡大に伴うデータ量、ユースケースの変更等の想定が、実態と乖離していることに起因する問題。

表 3-12 移行後の運用の想定誤りに起因する事象

No.	発生した事象	問題の原因	留意事項・想定される対策
⑩-1	国などに報告する統計資料をシステムで出力することとしたが、正しい数値を算出するために必要となるデータ項目について厳密な運用ができず(窓口ですぐに確認できず、関係機関に確認のうえ後日入力するしかない項目があり、統計処理の締め日には間に合わない等)、結局、紙ベースの申請書類を見ながら、後日数値を補正しながら運用していた。	システムを利用するにあたっての業務運用が整理できていなかった。	システムの設計開始前に、利用者としての事務運用フローを明確にして、データの発生・入力・利用のタイミングを明らかにすることが望ましい。
⑩-2	現行システムからの移行データに不備があり、新システムの機能仕様に合わないデータが多数内在してしまい、本番稼働後に、データ補正作業が発生した。	現行システムと新システムとの移行データの調査、仕様検討は実施したが、現行システムでの使用頻度の低かった機能を新システムで強制的に使い始めたところ、本番稼働後に利用者から不備が指摘された。	システム移行の際には、現行システム側も交えて(プロジェクトに参画)移行仕様の確認・調整を行い、移行仕様を確定する必要がある。
⑩-3	設計時点において、稼働後のデータ増加量を過少見積りしていたため、実際には想定以上のデータ量となってしまう、リソース(CPU、ディスク)不足によりシステム動作が不安定となった。	現行システムでは、各部署の利用者はあまり利用していない状態であったが、その状況を認識せず現行システムの実績を基に、新システムのデータ量を見積もっていた。実際には新システムでシステム利用が当初想定以上に増加したため、見積り値を超えてしまった。	各部署の現状及び今後の利用状況を十分に調査したのちに、新システム稼働後の利用状況を想定すべきであった。発注側で今後の利用状況は事前に調査し、調達時の要件として提示することが望ましい。

### ⑪異動時の引継ぎ不十分に起因する事象

設計書等に反映されていない暫定的な処理方法について、異動時に十分に引き継がれていないことに起因する問題。

表 3-13 異動時の引継ぎ不十分に起因する事象

No.	発生した事象	問題の原因	留意事項・想定される対策
⑪-1	⑩-1の問題について、職員の異動を契機に、この運用自体がきちんと継承されず、長期にわたり誤った数値のまま国に報告される結果となった。	データの取り扱いに関する運用が継承されなかった。	事務運用フローは、データの取り扱いについても明文化して、職員の異動時等も確実に引き継ぐ必要がある。

### ⑫人的な運用ミスに起因する事象

マスタデータの入力誤りといった人的な運用上のミスに起因する問題。

**表 3-14 人的な運用ミスに起因する事象**

No.	発生した事象	問題の原因	留意事項・想定される対策
⑫-1	外部の機関名が出力される一覧系の帳票で、誤った機関名を含む一覧が送付された。	職員による機関名マスタ入力ミス	発注側でダブルチェックなどの運用対応が必要。

**⑬権利関係の制約に起因する事象**

外字フォント、辞書等の権利関係の制約に起因する問題。移行の際に想定外の作業が発生している。

**表 3-15 権利関係の制約に起因する事象**

No.	発生した事象	問題の原因	留意事項・想定される対策
⑬-1	既存システムからデータを移行するにあたり、既存システムの外字フォントや辞書については、利用者に著作権がない（IT 事業者が著作権を留保）していることが発覚し、新たにフォントや辞書を作り直す必要があった。	調達前に、データを構成する要素に関する権利関係が確認できていなかった。	調達前に、現行システムから継承すべき資産の洗い出しと、その権利関係を整理しておくことが望ましい。
⑬-2	利用者登録機能において、パスワードチェック時に辞書にある用語を不可にするチェック機能が必要であったが、辞書に著作権があるため、現行システムの辞書がそのまま使えず、辞書を作り直す必要があった。	調達前に、辞書に関する権利関係を確認しなかった。	調達前に、現行システムから継承する資産の洗い出しと権利関係の整理をしておくことが望ましい。
⑬-3	他システム（国のシステム）のフォントを自庁内システムでも表示・印字できるようにしたいと考え、フォントの移行（もしくは購入）をしようとしたが、権利関係から不可能との判断になり、多くのコストをかけて、紙に印字されたフォントを見ながら、独自のフォントを作成した。	調達前に、データを構成する要素に関する権利関係が確認できていなかった。	調達前に、現行システムから継承すべき資産の洗い出しと、その権利関係を整理しておくことが望ましい。

**⑭原本性の解釈の相違に起因する事象**

原本性の解釈の相違に起因する問題。そのままデータ移行が行えず、紙台帳からデータを取りなおすといった想定外の作業が発生している。

**表 3-16 原本性の解釈の相違に起因する事象**

No.	発生した事象	問題の原因	留意事項・想定される対策
⑭-1	印鑑登録における印影の移行にあたり、既存システムにおける取り込み解像度と、新システムにおける管理・出力解像度が異なることから、一定のアルゴリズムで画像変換を行うこととしたが、「原本性が損なわれる」という理由から変換は不可とされ、紙台帳から新システムで読み取りなおすこととなった。	調達前にデータの原本性を保証する定義（レベル）が確立されていなかった。	調達前に、特に画像、音声などアナログ⇄デジタルの変換が必須となるデータについては、その原本性を保証する定義をしておくことが望ましい。

### 3.4. まとめ

以上の検討を経て、行政機関のデータマネジメントに係る課題として、昨年度調査では明確には確認されていなかったが、実際の業務・システムの構築・運用に影響を及ぼしていると思われる問題が多数確認された。特に、「人的な運用ミスに起因する事象」、「利権関係の制約に起因する事象」、「原本性の解釈の相違に起因する事象」、「データ管理手法の相違に起因する事象」は、新たに確認された観点である。

これらの結果は、確認された問題事象にどのように未然または事後速やかに対処すべきかを示唆するノウハウとして活用可能である。そこで、特にそのうち表 3-17 に示す点については、昨年度版ハンドブックにおいて作成した「第 5 章：調達仕様書の作成時に考慮しておくべき観点」のチェックリストとして追加しておくことが望ましい。

表 3-17 調達仕様書作成時のチェックリストに新たに追加すべき確認視点

データマネジメントに関する記載の確認視点（上段）／確認の不足によって想定されるリスク（下段）	
<p>■ マスタデータ管理</p> <p>職員一人によるマスタデータの入力及びチェックではなく、複数職員によるマスタデータのダブルチェックなどの運用ルールを設けること。 その要件を整理・明確化する実施タスクを受託者に対して明示すること。</p>	<p>【取組み不足による想定リスク】</p> <p>マスタデータの入力ミスにより、後工程で想定外のデータ発生による不具合を招く恐れ</p>
<p>■ データの権利関係の把握</p> <p>既存システムから外字フォントや辞書を引き継ぐ際、その知的財産権の保有者や権利関係を把握にすること。 引き継ぐ資産の利権関係を整理・明確にし、受託者に対して明示すること。</p>	<p>【取組み不足による想定リスク】</p> <p>既存システムの外字フォントや辞書が引き継ぐことができないことにより、新規に作成しなければならず、後工程で想定外の作業を招く恐れ</p>
<p>■ データ（印影等の画像）の原本性保証</p> <p>既存システムから印影等の画像データを引き継ぐ際、画像データの原本性の考え方とその対応方針を受託者に対して明示すること。</p>	<p>【取組み不足による想定リスク】</p> <p>原本性保証に関する解釈の食い違いにより、紙データから画像データを新規に作成しなければならず、後工程で想定外の作業を招く恐れ</p>
<p>■ 連携テスト</p> <p>連携システム間でコード体系やデータ管理手法、データ入力方法等がないか、事前に確認すること。 その要件の整理・明確化する実施タスクを受託者に対して明示すること。</p>	<p>【取組み不足による想定リスク】</p> <p>システム間で解釈やコード値の認識が異なることにより、システム間の連携テストで不具合を招く恐れ</p>

## 4. 人工知能（AI）の導入・運用におけるデータマネジメントの検討

ここ数年、政府では、行政機関におけるデータの利活用に向けて、EBPM(Evidence Based Policy Making)の推進をはじめとする様々な施策が立案・実施されている。こうしたデータ利活用に向けた取組みの中で、期待が高まっているテーマの一つが人工知能（以下「AI」）の導入である。

本章では、行政機関が自らの業務・サービスにAIを導入する場合に求められるデータマネジメントに、通常の情報システム構築の場合との間での違いがあるか、また、どのような点に留意してデータマネジメントを実践すべきかを検討する。

### 4.1. AI-情報システム間でのデータの用途、必要量、種類の違い

AIには大きく、予め人間が設定した判断基準に則って処理を行うルールベースのAIと、大量のデータから機械が習得した判断基準に則って処理を行う機械学習型のAIとがある。どちらの場合も機能させるためにはデータが不可欠であるし、使用するデータの品質に問題があれば、処理結果も問題を含んだものとなる。

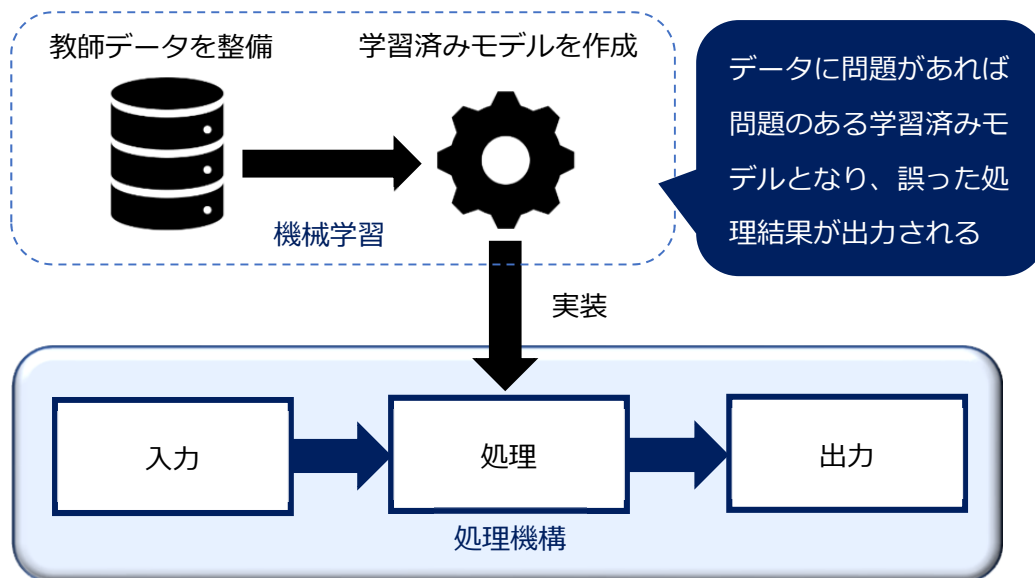


図 4-1 AI と情報システムの関係（教師あり機械学習の例）

他方で、データの用途や必要量、種類は、AI と通常の情報システムの間、また機械学習型のAI とルールベースのAI とでは異なってくる。さらに、機械学習型AI であってもアルゴリズムによって大きく異なってくる。データマネジメントの観点でこれらの違いを概観すれば以下のとおりとなる。

表 4-1 情報システムと AI の間のデータの用途、必要量、種類の違い

	AI		通常の情報システム
	機械学習型	ルールベース	
用途	・データから業務処理の判断基準自体を生成	・業務処理の判断基準は人間が設定。データは主に出力情報として利用	・業務処理の判断基準は人間が設定。分類等の判断基準としてデータを利用
量	・基本的には量が多い程判断の精度が向上する	・量が多い程、回答パターンのバラエティが増える	・サービスや用途に応じて一定量が必要
種類	・あらゆるデータ (例：テキスト、動画、画像、音声、数値)	・数値データ ・人間が読めるデータ (例：テキスト、音声)	・数値データ ・人間が読めるデータ (例：テキスト、音声)

#### 4.2. AI の処理内容に応じたデータマネジメントの違い

AI 利用の場合に、実際にどのようなデータマネジメントが必要とされるかは、AI 側でどのような処理を行うかによって異なってくる。

仮に AI によるデータ処理を、大量の文章の中から不正の予兆や証拠を検出するフォレンジック調査などに用いられる「自然言語処理」、コールセンターでの自動応答などに用いられる「画像・音声解析」、設備・インフラ保守における異常検知などに用いられる「アノマリー（異常値）分析」に分けると、それぞれのデータマネジメントの重点は以下のように変わってくる。

- ・ 自然言語処理：メタデータ（書誌情報、発生日、ソース、責任者等）が重要
- ・ 画像・音声解析：メタデータに加えデータの形式（ファイル形式、文字コード）も重要
- ・ アノマリー分析：メタデータ、データ形式に加えデータの意味（体系や値条件）も重要

これを利用するデータの種類、典型的な利用目的も含めて関係性を整理すると図 4-2 のとおりとなる。

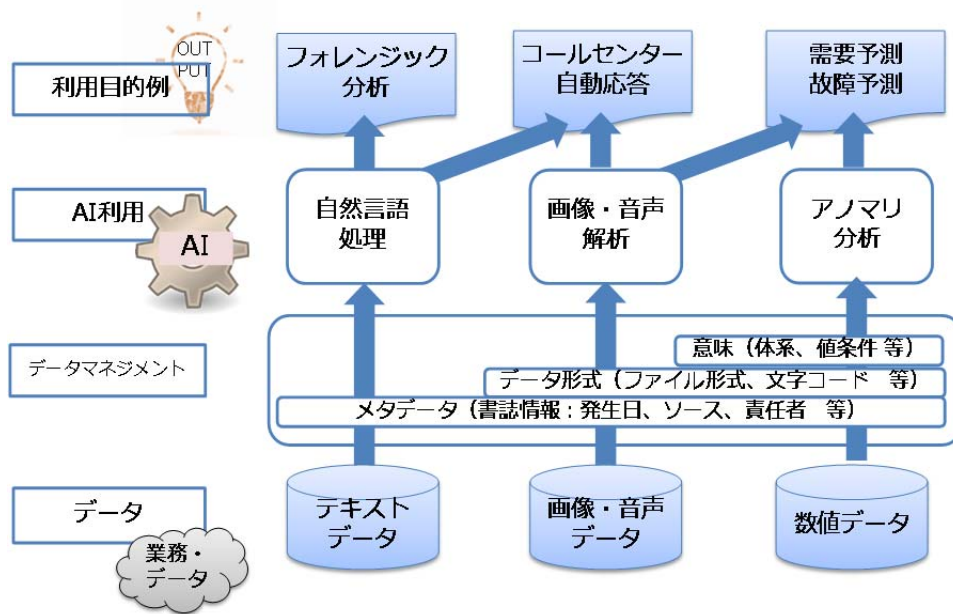


図 4-2 機械学習型の AI 利用に求められるデータマネジメント (分析種類別)

AI、特に機械学習型の AI では、データそのものが判断基準を形成することになるため、通常の情報システム以上にデータの重要性は高い。また、学習用のデータを整備し、チューニングしていくなど独特のプロセスも必要とされる。データそのものの取扱いにも差異が出てくる。AI では利用する手法によって、データマネジメントで考慮すべき重点は異なってくる。また、そこで利用するデータも、用途、量、種類は異なってくる。

しかしながら、データに対して「データ品質の確保」と「データ利活用の定着化」が求められる点は通常の情報システムと変わらない。そしてこれらの取組みにおいて AI 特有の要素はあまりないと考えられる。AI を用いない通常の情報システムにおいても、マスターデータをはじめ様々なデータを必要とする。正確なデータが投入されていなければ正しい出力結果が得られない点も AI の場合と同様である。すなわち、通常の情報システムであっても、AI であってもデータマネジメントが必要とされる場所に本質的な違いはないと考えられる。

したがって、AI 利用におけるデータマネジメントには、前述のような AI 特有の特徴を意識したうえで、通常の情報システム構築・運用で想定されるのと同様のデータマネジメントの取組みを行えばよいはずである。

#### 4.3. 有識者へのインタビュー

以上の仮説を検証するため、AI の実事業への導入に携わってきた有識者へのインタビュー

一を行った。インタビュー対象者とその選定理由は以下のとおりである。

インタビュー対象者	選定理由
① (株) リクルートホールディングス R&D 本部 RIT 推進室長 石山 洗氏 (注: 肩書きは取材当時のもの)	自社事業の強化に AI を活用するとともに、その成果を体系的に世の中に発信している。
② (株) FRONTEO 執行役員 CTO 行動情報科学研究所長 武田 秀樹氏	フォレンジック調査などのサービスを法執行機関に提供しており、行政での AI 活用に関する知見を有する。

各氏へのインタビュー結果のうち AI のデータマネジメントに関する回答内容の概要は以下のとおりである (注: インタビューは 2017 年 2 月当時)。

#### ① (株) リクルートホールディングス

- ・ 昨年度版ハンドブックで示されているコンセプトやフレームワークについて、大きな認識のずれはない。
- ・ データマネジメントのあり方は、システム開発を内製しているかどうかや、実行時の開発/運用プロセスに依存して決まってくる。
- ・ データマネジメントについては、長期的な効果の創出や事業の継続の観点が重要である。そのためには組織マネジメントやプロセス管理が必要。行政機関においても、省庁横断の CTO、または各省の CTO の配置を検討すべき。
- ・ AI では、システム間でデータの授受を行うための API が重要。例えば、チャットボットでは、1 回のクエリに対して複数のサービスにリクエストを投げるため、複数のサービスが API の口を持ち、データが繋がる必要がある。データが繋がっていないと複数の窓口が必要になるといった非効率が生じる。
- ・ 図 4-2 は同じ認識である。コールセンターの自動応答において、画像・音声データとテキストを組み合わせて利用する場合、結合 (名寄せ等) が大変。テキストからの回答の抽出や教師データの作り方への考慮が必要

#### ② (株) FRONTEO

- ・ データ活用を前提としたデータマネジメントを考えることが重要。
- ・ 様々なデータフォーマットに対応し、データ整形ができていなくても解析できる仕組みを用意している。
- ・ データを自ら整備し、整形するよりも、外部のデータ解析サービスを活用した方が早い。
- ・ 解析対象が変わった場合に、プロパティの追加ができるなど、分析やデータ活用の



観点の変化に柔軟に対応できるシステム設計とすることが重要。

- ・ 教師データ作成においては、経験豊富な人の判断を活用することが重要。信頼できる教師データを作ることが品質確保につながる。同質の判断を下すには少数の教師データの方が効率的な場合もある。
- ・ AI 活用の目的に合わせてデータを取り出せるようにしておくことが重要。

#### 4.4. まとめ

データマネジメントの基本的な考え方はAI 利用の場合であっても変わるところはないことが有識者へのインタビュー結果からも明らかになった。他方で、データの取り扱いについては、AI 特有の工夫が必要である。

機械学習の場合、どのようなデータを投入するかによって判断基準が形成されるので、教師データの設定が重要となる。精緻にメタデータを付与したり、欠損値の穴埋め処理をしたりしなければならない場合もあれば、必ずしも大量のデータすべての形式変換や整形をしなくてもよい場合もある。こうした判断は一概には言えないのでケースバイケースで判断するほかない。

また、データが判断基準を形成するという AI 特有の条件から、システム構築において特に設計の柔軟性を確保すること、また、質の高い教師データを作りこむことが重要である。

## 5. 今後の取組み（昨年度ハンドブックの見直し）

本調査研究では、第1章において、行政のデータマネジメントに関する各活動を位置づけるためのデータのライフサイクルのプロセスを定義した。その上で、第2章では、行政機関等に対してインタビューを実施し、データマネジメントに関する活動の好事例を収集・整理した。さらに第3章では、政府の情報システムのデータマネジメントに深く関わっているITベンダーの参画を得てグループインタビューを実施し、行政サイドからは必ずしも把握が容易でない課題を明らかにした。第4章では、AIの導入を行う場合も、他の場合と同様のデータマネジメントを実施することが重要であることが示された。

本章では、これらの調査研究の結果から得られた知見を広く行政部門に展開するために必要となる昨年度版ハンドブックの拡充の方向性を提言する。

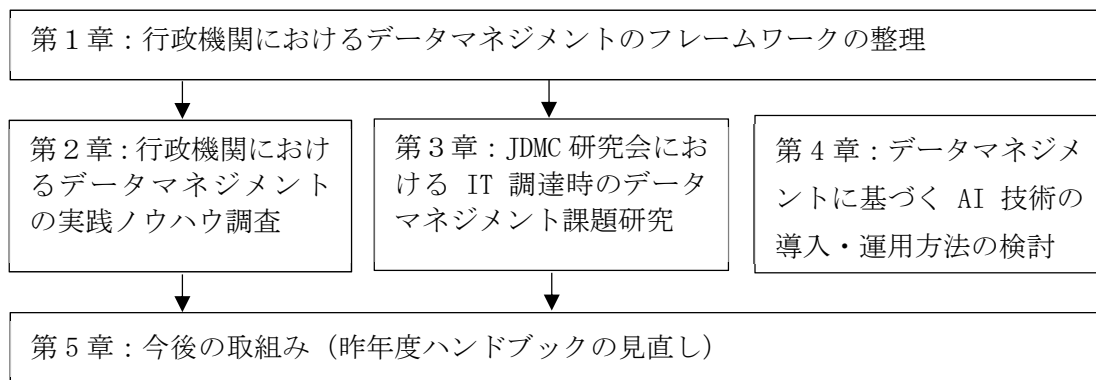


図 1-1 本調査研究の流れ（再掲）

昨年度版ハンドブックでは、行政におけるデータマネジメントの導入に向けた基本的考え方をとりまとめるとともに、実践に向けた第一歩として、「調達仕様書の作成時に考慮しておくべき観点」を示した。

本調査研究では、これらの成果を踏まえ、第1章で定義したデータマネジメントのライフサイクルプロセスに則って、第2章および第3章で示した調査研究の結果として得られたデータマネジメント実践のノウハウを整理し、「行政データマネジメント導入ハンドブック・実践編」として取りまとめる。

なお、昨年度版ハンドブックについては「調達仕様書の作成時に考慮しておくべき観点」を切り離し、第3章まとめの内容を反映したうえで「実践編」に移管し、「行政データマネジメント導入ハンドブック・総論編」として位置づける。すなわち、導入ハンドブックはこれら「総論編」および「実践編」の両文書からなる2部構成とする。

表 5-1 導入ハンドブック（総論編）と（実践編）の位置づけの違い

	総論編	実践編
目的	データマネジメントの全体像を理解し、取組みを動機づけるための素材を提供する。	データマネジメントを実践するためのノウハウを提供する。
内容	(1) データに係る典型的な問題事例 (2) データマネジメントにより実現できること (3) データマネジメントへの取り組みの主要タスク	(1) データマネジメント実践ノウハウ集 (2) 調達仕様書作成時チェックリスト
利用場面 (例)	<ul style="list-style-type: none"> <li>組織内でデータマネジメントの意義や必要性を説明する際の資料作成の素材として引用する。</li> <li>データ利活用の企画を行うにあたり、データマネジメントの全体像を把握する。</li> </ul>	<ul style="list-style-type: none"> <li>データに関する問題が発生した際に参考になりそうな解決策を調べる</li> <li>データ利活用に向けた取組みの各フェーズで、データマネジメント観点での検討の抜け漏れがないかをチェックする。</li> </ul>

## 別紙

別紙1：インタビューおよび研究会から確認された工夫・ノウハウ

別紙2：実践ノウハウ調査インタビュー結果概要

## 別添資料

①行政機関向けデータマネジメント導入ハンドブック（総論編）

②行政機関向けデータマネジメント導入ハンドブック（実践編）