

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

<概要版>

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 ー本調査研究の背景と目的ー	2
1. 調査研究の全体像と検討のフレームワーク	3
1.1. 本調査研究の全体像	3
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	3
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	5
2.1. ノウハウの抽出	5
2.2. 課題の整理	7
2.3. データマネジメントのアプローチの類型	8
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	9
3.1. 検討の実施方法	9
3.2. 抽出された問題の類型	9
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	11
4.1. AIー情報システム間でのデータの用途、必要量、種類の違い	11
5. 今後の取組み (昨年度ハンドブックの見直し)	13

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。

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した。

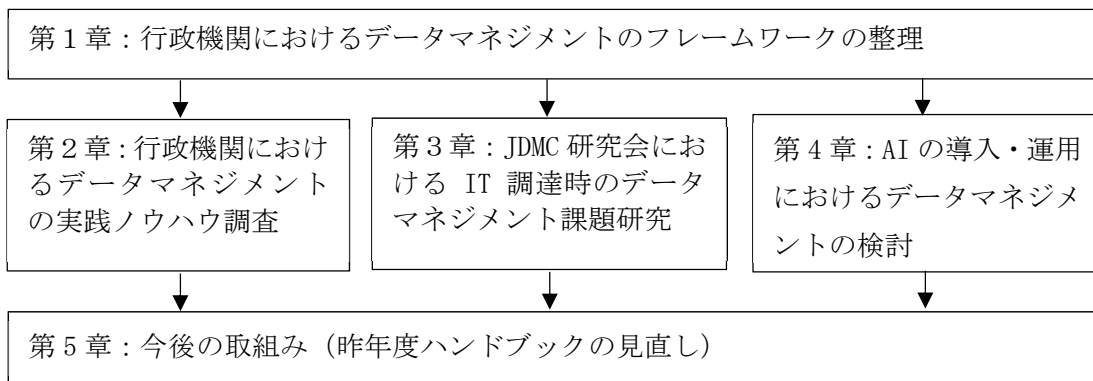


図 1 本調査研究の流れ

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。

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 1-1 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 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

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・整理した。

インタビュー候補先をインターネット、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、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されるようインタビュー先を選定・実施した。

2.1. 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り組みや工夫を、第1章の表1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2-1のとおり整理した。

表2 データマネジメント観点での取り組み・工夫の抽出

プロセス／観点	機関／事例	取り組み・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"> 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。 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ジックを念入りに作り込んだ。
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"> 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。
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"> 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。 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정이容易になり、補正作業時の職員の労力が軽減されるなど

プロセス／観点	機関／事例	取組み・工夫
1. 戦略・組織 / (b) データ利活用の定着化	浦安市／統合型 GIS	<p>電子化のメリットを前面にアピールした。</p> <ul style="list-style-type: none"> 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。 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。 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。
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"> 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。 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。
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"> 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。
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"> 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。 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を行った。 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。
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"> 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を理解し、データマネジメントを実施している政策部門でないと無理である。
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"> 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。 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。 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。

プロセス／観点	機関／事例	取組み・工夫
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	・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。
3. 構築～4. 運用 /(b) データ利活用の定着化	浦安市／統合型 GIS	・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。 ・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。
3. 構築～4. 運用 /(b) データ利活用の定着化	内閣官房／復興・復旧支援データベース	・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。

2.2. 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、ネットワーク強靱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ものだが、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。

2.3. 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。図2-1は、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。



図2 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（㈱エヌ・ティ・ティ・データ、日本電気㈱、㈱日立製作所、富士通㈱の 4 社）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った。

3.1. 検討の実施方法

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"> ・サブ WG の趣旨について ・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
第 2 回	2017 年 2 月 24 日	<ul style="list-style-type: none"> ・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
第 3 回	2017 年 3 月 13 日	<ul style="list-style-type: none"> ・成果取りまとめの方向性について ・今後のハンドブック等の改善・拡充のあり方 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 問題の種類

類型	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。

後述の「行政データマネジメント導入ハンドブック・実践編」では、「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」として整理している。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」として整理している。

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

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の導入にあたってはデータの整備が不可欠となるが、AIを用いない通常の情報システムにおいても、マスタ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ことに変わりはないと考えられる。

他方で、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。

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。

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。

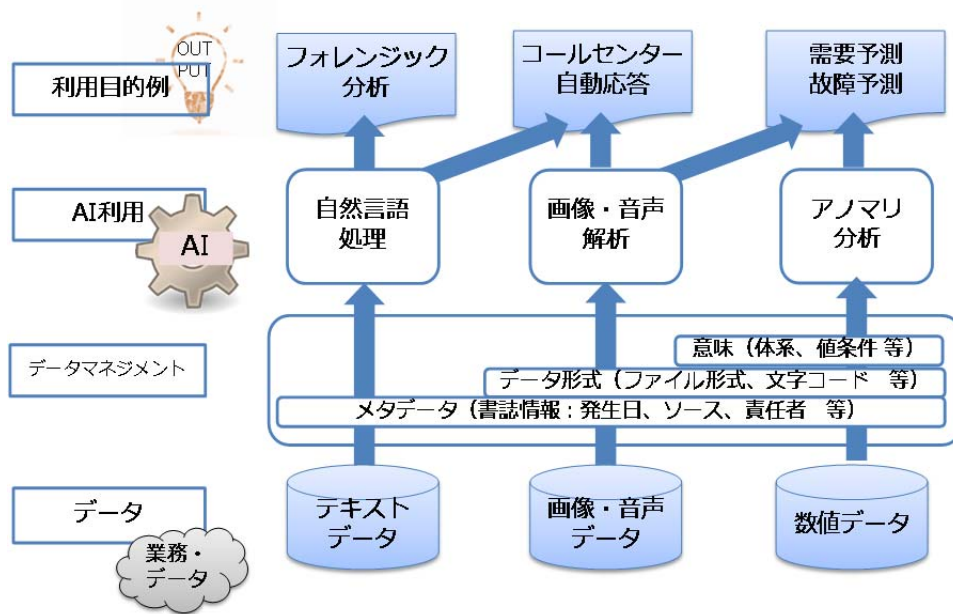


図 3 機械学習型の AI 利用に求められるデータマネジメント (分析種別別)

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（リクルートホールディングス R&D 本部 RIT 推進室長 石山 洸氏（注：肩書きは取材当時のもの）および（株）FRONTEO 執行役員 CTO 行動情報科学研究所長 武田 秀樹氏）へのインタビューを行った（注：インタビューは 2017 年 2 月当時）。

その結果、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）

昨年度策定した「行政機関向けデータマネジメント導入ハンドブック」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、昨年度版の「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、本調査研究第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4 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"> 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。 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。 	<ul style="list-style-type: none"> 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 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。