



Infrastructure
for Multi-layer Interoperability

IMI共通語彙基盤 導入の手引き

第2版

2017年3月

独立行政法人情報処理推進機構(IPA)
技術本部 国際標準推進センター

目次(1/2)

1. 公共データを公開するメリット	
1.1 電子行政推進に関する基本方針 (平成23年8月3日 IT戦略本部決定)のポイント	4
1.2 データ公開の意義	5
2. 共通語彙利用の意義	
2.1 IoT、ビッグデータなどとの関連性	7
2.2 データの横断的利用に対する課題	8
2.3 課題解決手段としてのIMI共通語彙基盤	9
3. 共通語彙基盤を利用するメリット	
3.1 相互運用性の向上とデータ活用の促進 (1/2)	11
3.1 相互運用性の向上とデータ活用の促進 (2/2)	12
3.2 開発コストの低減 (1/2)	13
3.2 開発コストの低減 (2/2)	14

目次(2/2)

4.共通語彙基盤の利用

4.1 共通語彙基盤の構成	16
4.2 共通語彙基盤の利用環境	17
4.3 共通語彙基盤の用語とは (1/2)	18
4.3 共通語彙基盤の用語とは (2/2)	19
4.4 語彙とは (1/2)	20
4.4 語彙とは (2/2)	21
4.5 データの項目と共通語彙の関連付け	22
4.6 IDとコード	23
4.7 データモデル記述	24
4.8 IMI構造化項目名記法	25
4.9 語彙整備の取組み	26
執筆者、著作権・責任	28

1. 公共データを公開するメリット

1.1 電子行政推進に関する基本方針 (平成23年8月3日 IT戦略本部決定)のポイント

- ▶ 行政の電子化には、その目的として行政サービス利便性向上のみならず、「**国民、企業等の権利の実現や利益の保護**」「**行政の一層の透明化や国民参加の促進**」が掲げられています。
 - ▶ そのため、行政の業務遂行に伴うデータが、**入手・加工・分析しやすい形態**で提供されることが重要です。
- ▶ 人口減少トレンド、高齢化の進展、地域社会の縮退といった課題に対し、限定された予算で「**国民一人一人の事情に対応した、きめ細かい行政サービスを提供していく**」(「**電子行政推進に関する基本方針** (平成23年8月3日 IT戦略本部決定)」)ためには、公的サービスの実現に対する**民間企業や地域コミュニティの参画を振興する**必要があります。
 - ▶ 民間企業や地域コミュニティといった多様なステークホルダーにとって、行政が公開するデータが**利活用しやすい形態**で提供されることが重要です。
- ▶ 「**官民含めた我が国の社会・経済全体の活性化が図られ、国際競争力や国際的なプレゼンス等を高める**」必要性があります。
 - ▶ 公共セクター／民間セクターの**生産性向上**と、経済活動の**国際化**に対する対応が急務です。

1.2 データ公開の意義

- ▶ 電子行政の最初の目標は、行政オペレーションを電子化することによる執行の効率化、行政サービスの向上、ニーズの多様化への対応でした。
- ▶ しかし、これまで公が独占的に提供してきた役割が、**民間企業や地域コミュニティとの協働によって実現されるマルチステークホルダーアプローチ**が広がりを見せてています。
- ▶ 民間セクターの経済活動において、行政が整備・提供するデータを活用する機会は多いが、**入手・加工といった工程を効率化することによって、コスト低減が実現できます。**

単に行政が情報(文書)を公開するのではなく、民間セクターが入手・加工・分析・活用しやすい形態でデータが提供されることにより、公的サービス設計・供給の効率化、経済活動の活性化が期待できます。

2. 共通語彙利用の意義

近年のデータ社会における、データの横断的利用に向けた共通語彙の有用性・必要性について。

2.1 IoT、ビッグデータなどとの関連性

- ▶ オープンデータの促進は、**公共データを含むビッグデータの分析・利活用**を通じた様々な効果、例えば、新たなサービスや事業の開発をはじめ、地域住民、コミュニティ、地方公共団体等の課題の発見（見える化）・解決、さらには超高齢社会の到来に備えた我が国全体の課題の発見（見える化）・解決等につながることが期待されているものです。
「新たなオープンデータの展開に向けて（平成27年6月30日 高度情報通信ネットワーク社会推進戦略本部決定）」
- ▶ データ収集や各種コードによる**データの横断的利用**が機械で自動的に可能になることからコスト圧縮ができ、新しいサービスを提供するビジネスが可能となります。（例えば、気象、地質、交通その他の観測・調査データのような専門的データを収集・分析してビジネスに活用するなど）
「二次利用の促進のための府省のデータ公開に関する基本的考え方（ガイドライン）（平成25年6月25日 各府省情報化統括責任者（CIO）連絡会議決定、平成27年12月24日 改定）」

IoT 技術普及によるセンサーデータ流通量の増加など、ビッグデータ社会が到来しつつあるなか、データの横断的利用実現が重要であり、公共データが相互運用性をもつことは不可欠です。

2.2 データの横断的利用に対する課題

- ▶ 「企業等が行政の保有する(中略)震災関連情報を広く周知しようとしても、(中略) 行政機関ごとにフォーマットが異なり、情報の収集や整理に多くの時間が必要とされるケースが発生するなど、様々な課題も指摘された。」
「電子行政オープンデータ戦略（平成24年7月4日 高度情報通信ネットワーク社会推進戦略本部（IT戦略本部）決定）」
 - ▶ 公開されたデータが活用されるということは、不特定多数による編集・加工が前提となります。つまり、多様なデータが混合されて用いられる(マッシュアップ)が前提となります。
 - ▶ データのフォーマットや、項目内容(語彙)等が異なる場合、複数データを混合して用いるのは困難です。

多様なデータが共通のフォーマット、語彙を有することにより、利活用を促進するとともに、防災・災害対応活動や経済活動を効率化、活発化することが期待できます。

2.3 課題解決手段としてのIMI共通語彙基盤

- ▶ 単に行政が情報(文書)を公開するのではなく、民間セクターが入手・加工・分析・活用しやすい形態でデータが提供されることにより、公的サービス設計・供給の効率化、経済活動の活性化が期待できます。
- ▶ IoT技術普及によるセンサーデータ流通量の増加など、ビッグデータ社会が到来しつつあるなか、データの横断的利用実現が重要であり、公共データが相互運用性を持つことは不可欠です。
- ▶ 多様なデータが共通のフォーマット、語彙を有することにより、利活用を促進するとともに、防災・災害対応活動や経済活動を効率化、活発化することが期待できます。

その方法として、

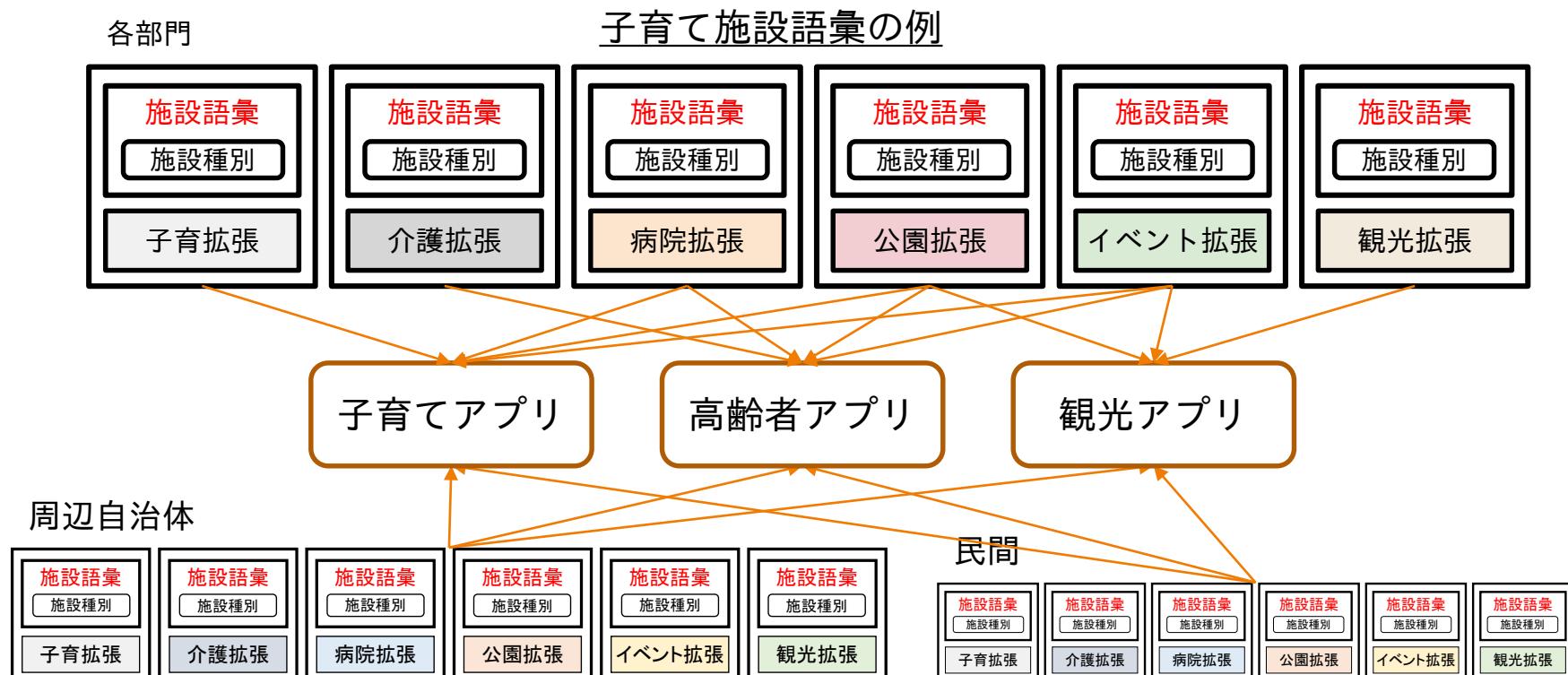
行政によるIMI共通語彙基盤を採用した公開データの整備が望まれています。

3. 共通語彙基盤を利用するメリット

共通語彙基盤による行政内部門間、周辺自治体、民間企業・住民との間でデータの交換・利活用の促進モデルについて。

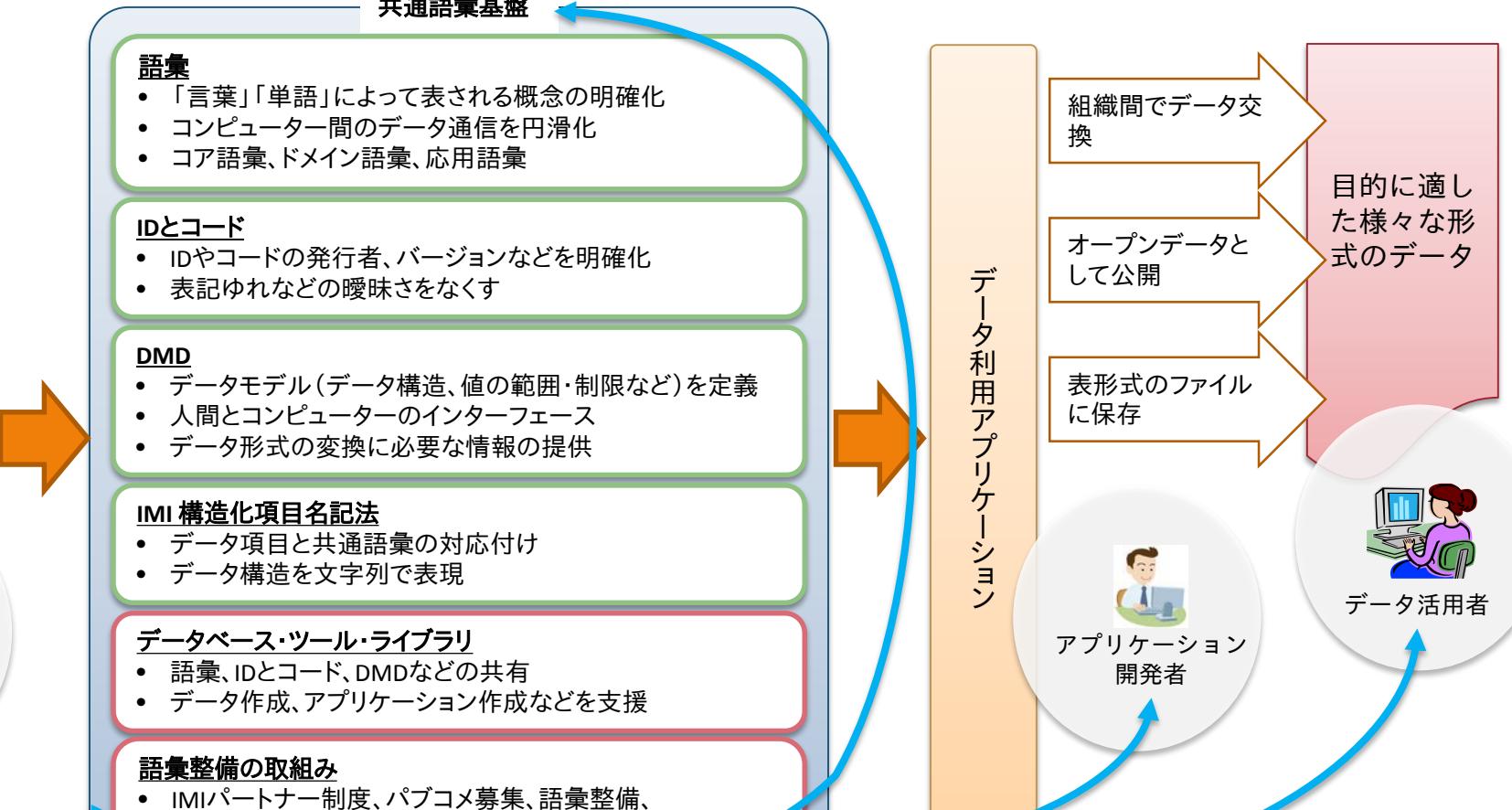
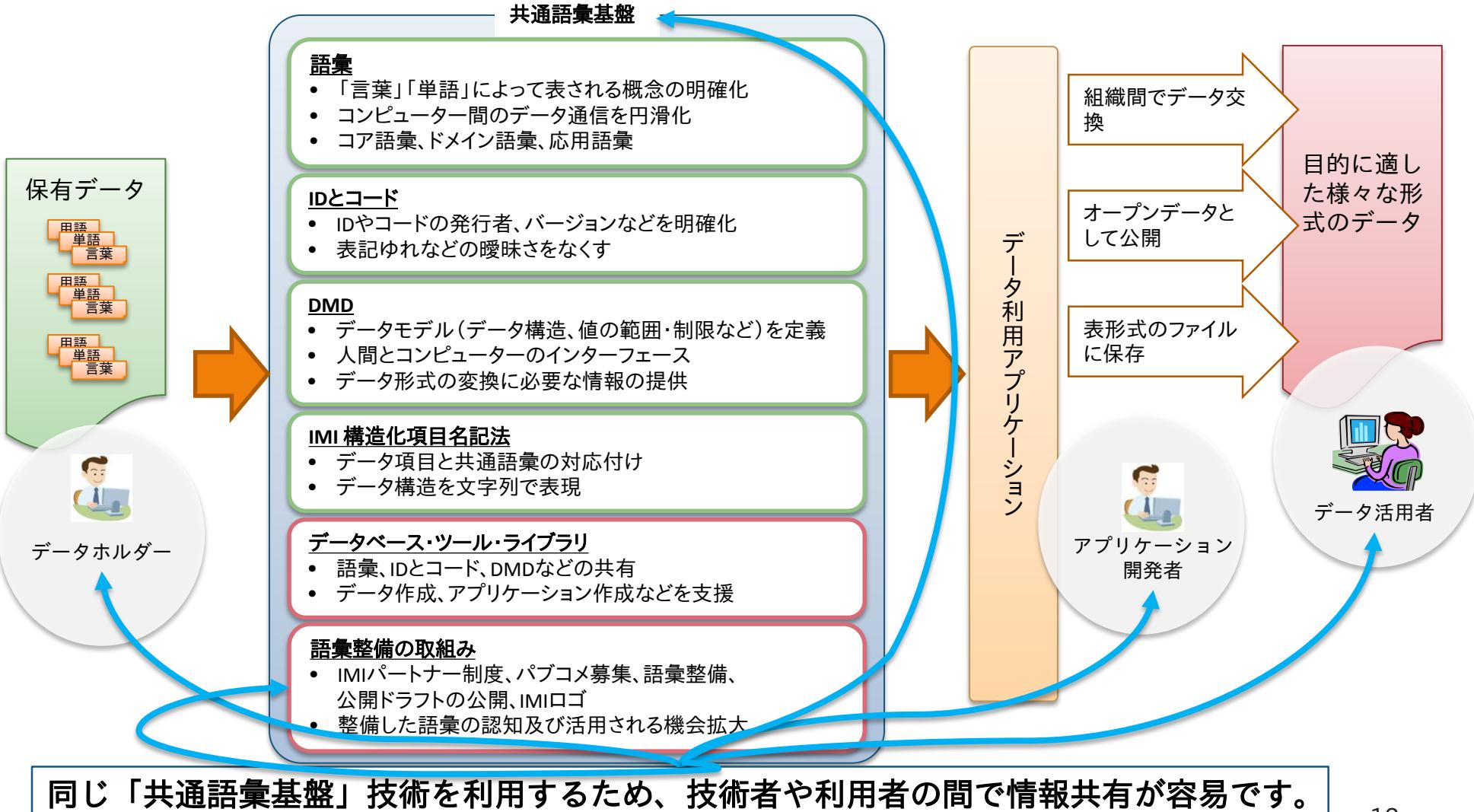
3.1 相互運用性の向上とデータ活用の促進 (1/2)

- ▶ 分野や地域を超えて広範囲に、データが連携して活用されます。他の組織の情報との相互運用性が確保できます。
- ▶ 行政内部門間、周辺自治体、民間企業・住民との間でデータの交換・利活用が進展することにより、アプリ開発などを通じた効率的な公的サービスの実現が期待できます。



3.1 相互運用性の向上とデータ活用の促進(2/2)

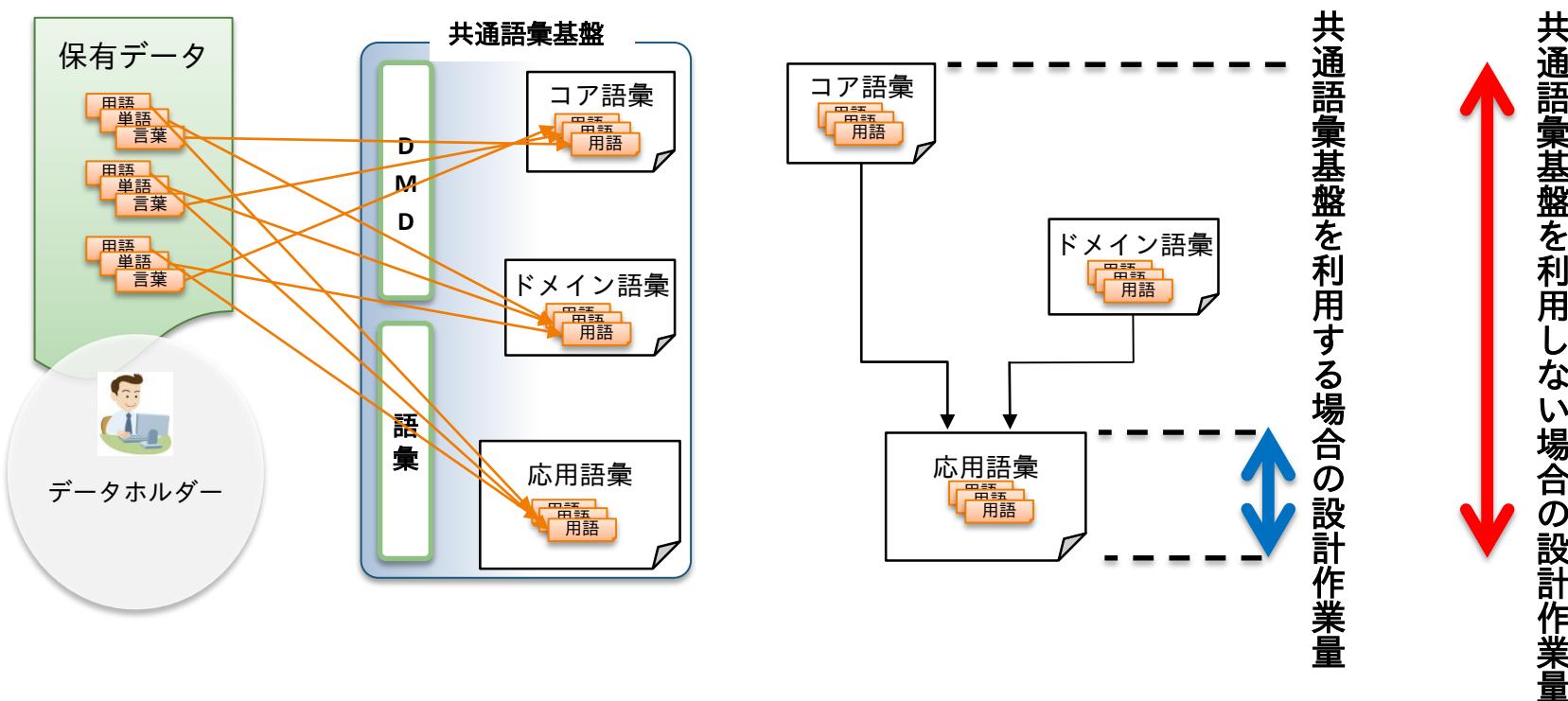
- データホルダー、アプリケーション開発者、データ活用者間でデータに関する情報の共有が容易になり、データ活用が促進されます。



同じ「共通語彙基盤」技術を利用するため、技術者や利用者の間で情報共有が容易です。

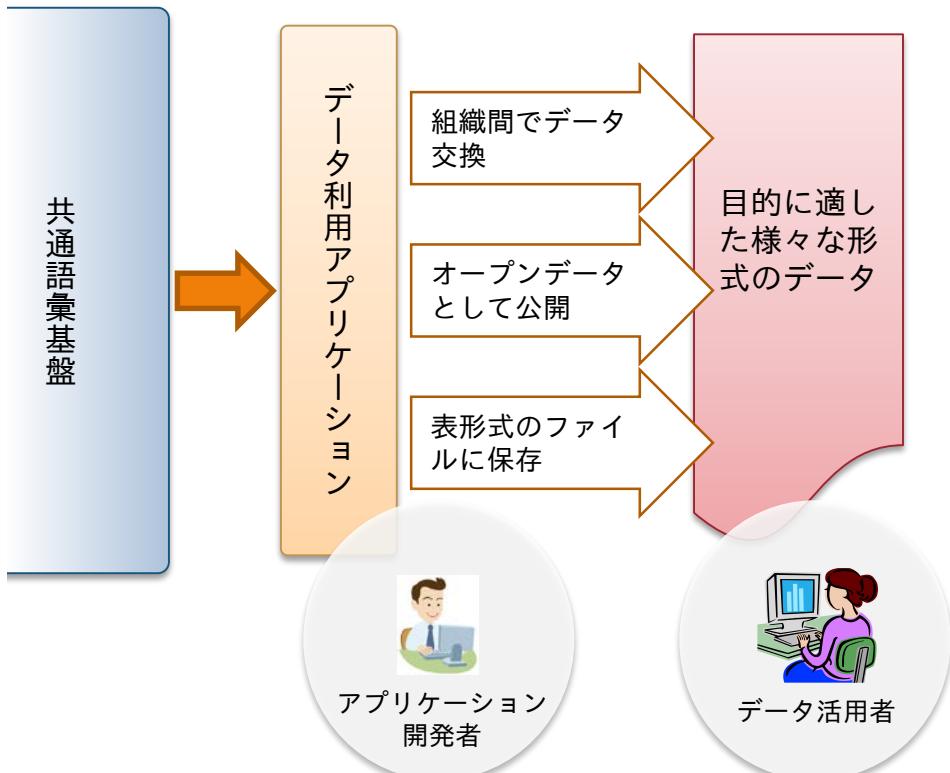
3.2 開発コストの低減 (1/2)

- ▶ 対象データに必要な項目などを一から考える必要がなくなり、データ項目の設計作業にかかるコストが低減されます。
 - ▶ 対象データ全体の構造を「共通のデータ構造 + 個別のデータ構造」として分けて考えて、共通のデータ構造としてコア語彙やドメイン語彙を採用することで、個別のデータ構造である応用語彙のみを設計するだけで済みます。



3.2 開発コストの低減 (2/2)

- ▶ 公開するためのフォーマットなどを一から考える必要がなくなり、そのための設計作業量が減ります。
 - ▶ 様々な決まった形式の中から、目的に適した形式を選択できます。
 - ▶ 共通語彙基盤を採用することで、項目名、プロパティやタグが決まります。



- ▶ 共通語彙基盤が提供するツールやライブラリを活用できるので、システム開発コストが低減されます。

例えば:

- ▶ 語彙の設計を支援するツールを利用できます。
- ▶ 表計算ソフトで扱うことのできる表形式とオープンデータとして扱いやすい形式やデータ交換に適した形式などの間で相互に変換するツールを利用できます。
- ▶ 使い慣れた表計算ソフトを活用することもできます。
- ▶ 共通語彙基盤が提供するツール以外にも、様々な便利なツールやライブラリが登場すると期待できます。

4.共通語彙基盤の利用

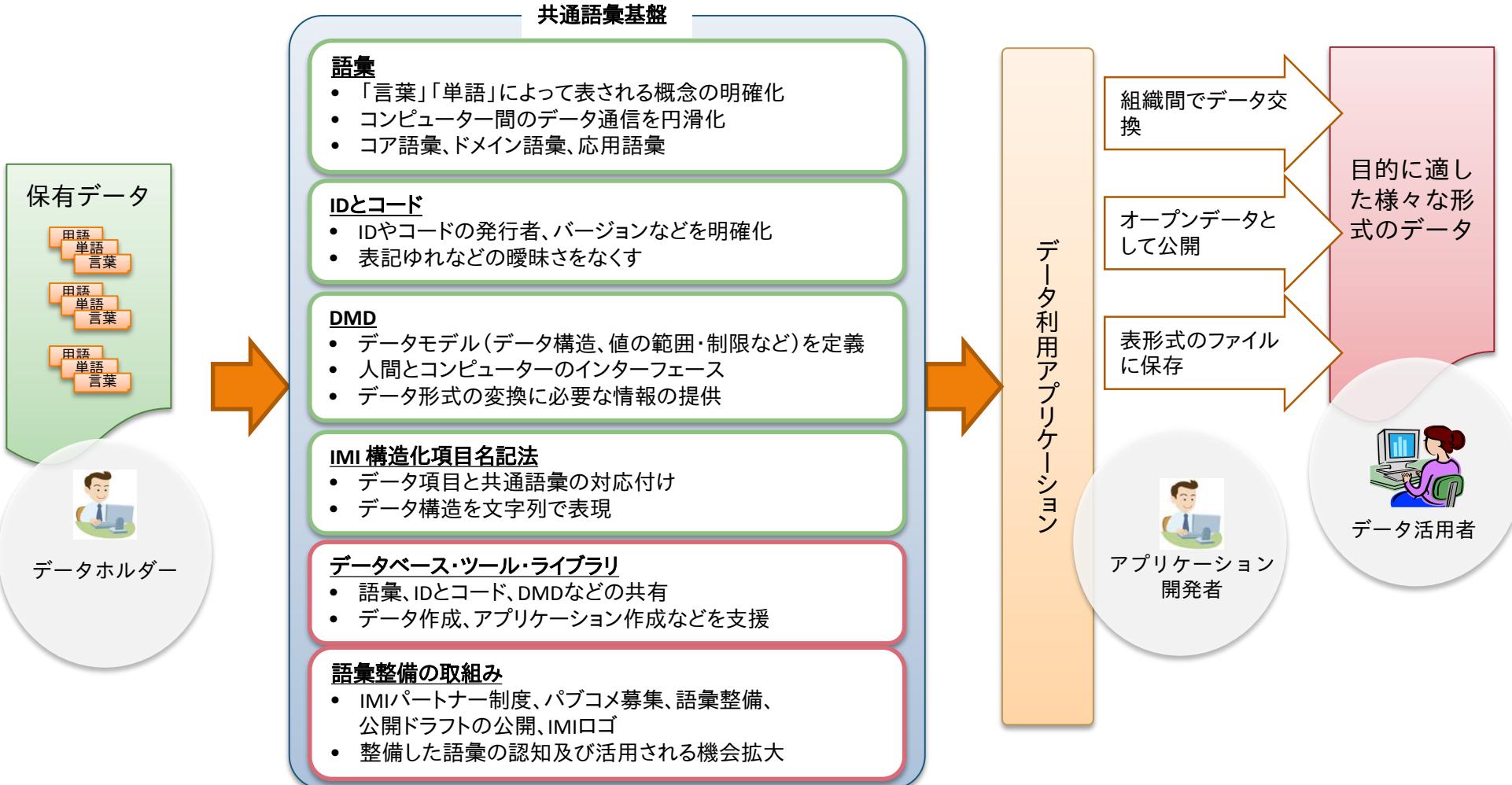
共通語彙基盤を利用するにあたって理解しておくべき語彙の概念、実際の語彙活用において相互運用性を高めるための仕組み、用いる記法などについて。

4.1 共通語彙基盤の構成

共通語彙基盤は次の要素（規定やソフトウェア）によって構成されています。

- ▶ データを表現するための規定など
 - ▶ 共通語彙
 - ▶ IDとコード
 - ▶ データモデル記述（DMD）
 - ▶ IMI 構造化項目名記法
- ▶ 共通語彙を利用するための各種支援
 - ▶ ポータルサイト（imi.go.jp）
 - ▶ 語彙データベース（開発中）
 - ▶ ツール・ライブラリ類（開発中）
 - ▶ 語彙整備の取組み

4.2 共通語彙基盤の利用環境



4.3 共通語彙基盤の用語とは (1/2)

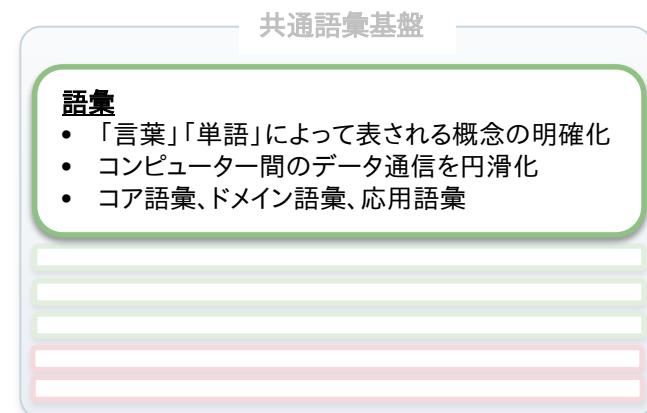
- ▶ 行政に関わる文書は言語（日本語）によって記されています。そこに含まれている用語は専門性の高い用語（例えば法律用語）から日常で用いられる用語まで多岐にわたり、その語数は膨大であるため、類似の用語、重複する用語、分野で意味が異なる同一語といった用語の氾濫による混乱をもたらしています。
- ▶ 人間が文書を読む場合は、これらの用語の混乱は読む側で整理され、多くの場合は問題がありません。しかし、文書をコンピューターで処理する場合は大きな問題となります。例えば、ある事柄を検索しようと文書のデータベースをある用語で検索した時、もしその用語と同じ意味の別の用語があるなら、検索結果は不十分になってしまいます。
- ▶ これは文書だけの問題ではありません。行政に関わるデータは言語で書かれた表題、項目名をつけて生成されています。この表題や項目名においても用語の混乱が影響を及ぼします。
- ▶ IMI共通語彙基盤では、このような混乱を解決するために、次のような考え方に基づいて、データに用いる用語を共通化しています。

4.3 共通語彙基盤の用語とは (2/2)

- ▶ 共通化する対象は事柄を指し示す概念とします。文章や表題名、項目名、項目の値などに出てくる語（言葉、単語）は概念を表現する手段と考えます。例えば、【人】という概念があったとき、文書の中で【人】を表現するときに「人間」という語を使うかもしれませんし、「人」あるいは「Person」という語を使うかもしれません。この場合、IMI共通語彙基盤では、語「人」「人間」「Person」は同じ概念【人】を指しているということで、同じ意味をもつ語と判断します。このように対応させることで、分野における用語の多様性を許容します。ただし、そうはいっても用語が不必要に多様化することは望ましくありません。このため、IMI共通語彙基盤では原則として、一つの概念には一つの語（概念の名称に対応する語）を代表的な表記として用意することで、この語を使うことを推奨します。なお、IMI共通語彙基盤では、特に混同が問題ない場合は、概念を用語と呼びます。

4.4 語彙とは (1/2)

- ▶ 「語彙」は、一般には「語」（言葉、単語）の集合を意味しますが、共通語彙基盤における「語彙」とは、前述のように、単にそのような語を集めたものではなく、ファイルやインターネットなどを介して行われるコンピューター間のデータ通信を円滑かつ確実に行うことの目的として、概念の代表的な表記としての一つの語の意味や構造、そのような語によって表される概念と他の概念の関係などを明確にした概念（＝用語）の集合を指して用います。



4.4 語彙とは (2/2)

▶ 共通語彙の3層構造

- ▶ IMI共通語彙基盤の共通語彙は3層からなります。中心となるのは、**コア語彙**と呼ばれるもので、最も基本的な語彙を対象としたものです。コア語彙は分野に依存せず多くの分野でよく使われるような用語（【人】、【氏名】、【住所】、【組織】など）を収録する語彙です。
- ▶ ドメイン語彙は、分野ごとに用意される語彙です。ドメイン語彙はそれぞれの分野の基礎的な用語であり、また分野を超えてもよく使われる共通性の高い用語を収録する語彙です。ドメイン語彙はコア語彙の拡張として定義されます。すなわち、コア語彙にある用語を継承して、その分野固有の用語を定義します。
- ▶ さらに特定の課題やシステムで用いられる語彙を**応用語彙**といいます。応用語彙は、コア語彙やドメイン語彙の中の用語やこれらの語彙の用語を必要に応じて拡張した用語を用意して、その課題やシステムに必要な用語を取り揃えたものです。

▶ 語彙の実装

- ▶ 共通語彙（コア語彙、ドメイン語彙、応用語彙）の定義は抽象的なデータモデルであり、そのままコンピューターが扱える訳ではありません。それで、実際のデータ公開およびデータ交換に用いる、語彙に基づいたデータ項目とその構造は、具体的な表現であるXMLのスキーマおよびRDFのスキーマとして定義されます。
- ▶ その中で応用語彙のXML、RDFにおけるスキーマ定義では、用語の構造に加え、使用する用語の回数、用語のテキスト値の制限や書式など、実際の語彙データ記述において適用すべき制約事項も定義されます。

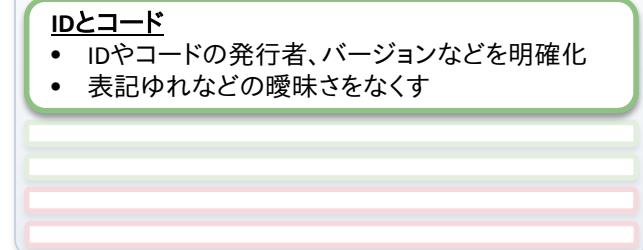
4.5 データの項目と共通語彙の関連付け

- ▶ 実際に語彙のデータを物理的に格納するのはXMLやRDFのスキーマで定義された「データ構造」をもったファイルです。しかし、そのようなXMLやRDFのデータは、利用者（閲覧者、データ入力者など）にとっては直接表示や入力の対象ではなく、利用者は、ブラウザーや入力フォームなどのフロントエンド・アプリケーションを使って自分たちが用いる「データ項目」を扱います。
そこで、語彙に基づいたアプリケーションを構築するには、人間向けの「データ項目」とコンピューター向けの「データ構造」を対応付ける必要があります。
このように、語彙データを実際に利用するには、コンピューターが扱うための、コア語彙やドメイン語彙に基づいた応用語彙を開発するだけではなく、利用者が扱うためのデータ項目との対応を取り必要がありますし、さらに、データ項目で使用可能な文字セット等の情報も与えられなければなりません。さらに、語彙データを利用するアプリケーションを開発するためには、関連する幾つかのデータや仕様も用いられます。

4.6 IDとコード

- ▶ 一般に、データの値として、自然言語による説明文ではなく符号によって対象を表すことがあります。共通語彙基盤では、データを他のデータと識別をする為に符号を用います。共通語彙基盤で用いる符号は「ID」と「コード」の2種類があります。
- ▶ ID
 - ▶ 適用範囲とするグループに属するメンバー各々に一意の符号を割り当て、他と区別する方法です。たとえば、以下2つをID体系の例として挙げる事ができます。
 - 社員番号（会社を適用範囲とし、社員を識別する）
 - 利用者ID（サービスを適用範囲とし、利用者を識別する）
- ▶ コード
 - ▶ 処理対象の持つ情報属性毎にコードを割り当て、他と区別する方法です。たとえば、国コード、文字コード、都道府県コードなどがその例です。区分、種別といったカテゴリーもコードとして表現されることがあります。土地利用種別や日本標準産業分類のコードがそれにあたります。
 - ▶ 参考ドキュメント：<http://imi.go.jp/goi/contents-list.html>

共通語彙基盤

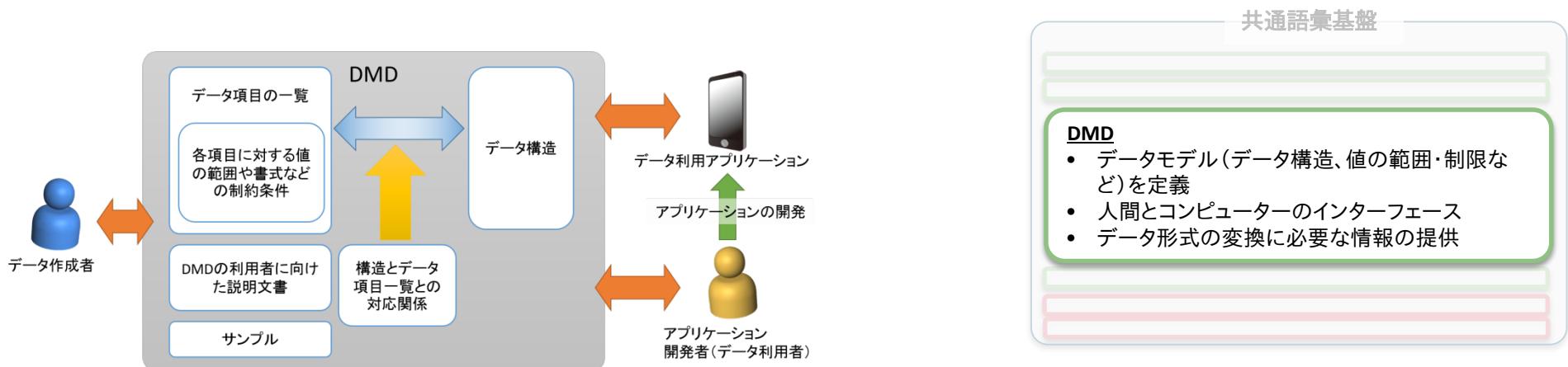


IDとコード

- IDやコードの発行者、バージョンなどを明確化
- 表記ゆれなどの曖昧さをなくす

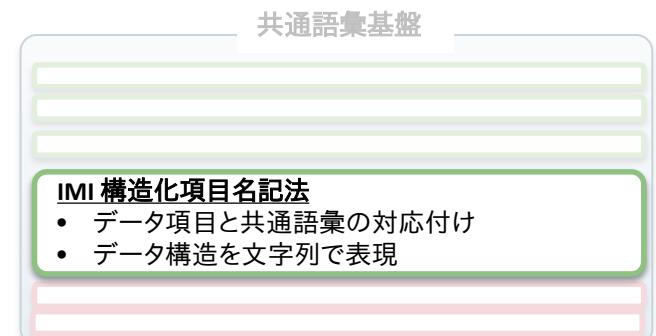
4.7 データモデル記述

- ▶ データモデル記述 (Data Model Description: DMD) は、データモデル、つまりデータの構造や値の範囲などを説明するためのパッケージです。
- ▶ DMDには、データを作成したり、データを利用したりする際に必要な情報の全てが含まれています。
 - ▶ DMDを利用すれば、データを公開する人も、データを利用する人も、公開するデータやアプリケーションで利用できるデータについて、特定のDMDを示すだけで項目ごとに詳細な情報を説明する必要はなくなります。
 - ▶ DMDを共有することで、他の人も同じ構造でデータを公開できるようになります。
 - ▶ 同じDMDに合わせて作られたアプリケーションは、異なる自治体や会社によって作られたデータであっても1つのアプリケーションで利用できるようになります。
- ▶ 参考ドキュメント：<http://imi.go.jp/goi/datamodel-about.html>



4.8 IMI構造化項目名記法

- ▶ 構造化項目名記法は、階層構造をもつデータの位置を文字列によって表現するための仕様です。構造化項目名記法を利用することで、階層構造をもつデータを表形式のデータとして表現することができるようになるため、階層構造をもつデータの表計算ソフトウェアによる編集や、表計算ソフトウェアなどで作成されたデータから階層化されたデータへの効率的な変換ができるようになります。
- ▶ 構造化項目名記法では、 クラス用語>プロパティ用語>プロパティ用語>・・・ のように「>」または「>」で区切って一つのクラス用語と 任意の数のプロパティ用語を並べたものになっています。非常に簡単な例を次に示します。ここでクラス用語はデータの対象となるもので、プロパティ用語はそれに関する事項です。クラス用語とプロパティ用語については、<http://imi.go.jp/ns/core/2> を参照してください。
- ▶ 例えば、施設の名称は、構造化項目名記法では「施設>名称>表記」のように記述します。
 - ▶ 参考ドキュメント : <http://imi.go.jp/goi/j-serialize.html>



4.9 語彙整備の取組み

▶ IMIパートナー制度

▶ 語彙の整備等を行う皆様と、相互運用性拡大などの目的を共有するIMIパートナーとしてお互いに協力して成果の展開を図り、IMI共通語彙基盤の一層の普及・充実と相互運用性の拡大を図るとともに、広く社会全般の一層の繁栄と発展につなげていくことを目指します。

▶ 概要

- IMIパートナーのみなさまと協議の場を設け、語彙の整備を支援いたします。
- IMI情報共有基盤のウェブサイト「imi.go.jp」で、公開ドラフトを公開したり、パブリックコメントを募集することができます。また、ご希望によりIMIパートナーの取組みを紹介させていただきます。
- 別途「共通語彙基盤の推進に関する協定書」を取り交わしていただきます。

▶ メリット

- 品質と効率 :IMIの知見と成果をふまえた実効性のある語彙整備
IMI DMD Editorツール利用による省力化
- チャネル拡大:整備した語彙が認知・活用される機会の拡大
- 信頼性 :IMI共通語彙基盤への対応を示すシンボルマーク
- ニーズの確認:パブコメ活用による利用者ニーズをふまえた語彙整備

共通語彙基盤

語彙整備の取組み

- IMIパートナー制度、パブコメ募集、語彙整備、公開ドラフトの公開、IMIロゴ
- 整備した語彙の認知及び活用される機会拡大

4.9 語彙整備の取組み

▶ パブリックコメント募集

▶ 概要

- IMI共通語彙基盤では、公開ドラフト・仕様・記法などについて、IMIポータルサイトに公開し、みなさまからのご意見を募集しております。

▶ 公開ドラフトの公開

▶ 概要

- 公開ドラフト(Public Draft)は、IMIポータルサイトより公開し、皆様に活用いただきつつ、改良を続けてゆくことを想定した語彙、あるいは将来の語彙の基となる検討文書です。
- IMI共通語彙基盤を活用するデータの作成者、利用者にとってより有益な内容にしていくため、公開するとともに広くみなさまのご意見を募集しています。

▶ IMIロゴマーク

▶ 概要

- IMI共通語彙基盤では、共通語彙基盤を利用するみなさまがIMIの提供情報を参照していることを示しやすくするため、ロゴマーク(IMIロゴ)を提供しています。

共通語彙基盤

語彙整備の取組み

- IMIパートナー制度、パブコメ募集、語彙整備、公開ドラフトの公開、IMIロゴ
- 整備した語彙の認知及び活用される機会拡大

▶ 執筆者

IPA技術本部国際標準推進センター
コア語彙検討サブワーキンググループ

武田英明（主査） (以下、50音順)	大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 情報学プリンシップ研究系 教授
奥井康弘	株式会社日本ユニテック 総合研究所
加藤文彦	大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 情報学プリンシップ研究系 特任研究員
神崎正英	Xenon Limited Partners 代表
小林巖生	
頃末和義	
菅又久直	国連CEFACT日本委員会 サプライチェーン情報基盤研究会 業務執行理事
深見嘉明	立教大学 大学院ビジネスデザイン研究科 特任准教授
松澤有三	インディゴ株式会社 シームレス空間基盤研究開発センター 主席研究員
山口 琢	
山本 暖	株式会社日立製作所 研究開発グループ 研究員

▶ 著作権・責任

本書は、CC0 1.0 全世界 (<https://creativecommons.org/publicdomain/zero/1.0/legalcode.ja>) により公開します。
本書の内容を適用した結果生じたこと、また、適用できなかった結果について、執筆者、IPAともその責を負いません。
あらかじめご了承ください。

改訂履歴

IMI共通語彙基盤 導入の手引き（第1版）

2017年1月6日 IMIOverviewBook_v1.pdf 発行

2017年3月31日 IMIOverviewBook_v2.pdf 発行 共通語彙基盤の構成要素変更及び説明ページ追加

発行 独立行政法人情報処理推進機構（IPA）
〒113-6591 東京都文京区本駒込2-28-8
文京グリーンコート センターオフィス16階
技術本部 国際標準推進センター
<https://imi.go.jp/>