

IMI 共通語彙基盤に対応したデータ公開の手引き

はじめに	1
■STEP1. 公開の目的を定めデータセットを選ぶ.....	3
1.1 公開データセットの選定とデータセット定義.....	3
■STEP2. データに意味づけを行う	7
2.1 公開対象とするデータ項目の検討	8
2.2 データ項目・記載内容・データ構造の整理.....	10
2.3 データ項目の検証.....	20
2.4 応用語彙の定義.....	22
2.5 コード、ID の整備.....	25
2.6 識別子の定義.....	28
2.7 データ項目定義書の作成	30
■STEP3. 目的に合った記述方法を決めて実装する.....	32
3.1 公開ファイル形式や公開場所などの決定	32
3.2 データレコードの入力・追加・削除	35
3.3 データの変換.....	36
■STEP4. データを公開・運用する	37
4.1 データの公開.....	37
4.2 データ公開の告知.....	38
4.3 公開したデータの運用	40
4.4 データ項目定義書の公開	43
この文書について	46

独立行政法人情報処理推進機構(IPA)
(法人番号 5010005007126)



はじめに

本手引きでは、IMI 共通語彙に対応したデータ公開の流れを、以下の4段階のステップに沿って説明していきます。

- STEP1. 公開の目的を定めデータセットを選ぶ
- STEP2. データに意味づけを行う
- STEP3. 目的に合った記述方法を決めて実装する
- STEP4. データを公開・運用する

それぞれのステップに明確なルールはありませんが、これまでに行われてきた自治体オープンデータ公開の経験等に基づいたものです。各段階について、作業手順、参考情報、作業のポイント等を、具体例を交えてご紹介します。

本手引きについて、広くご意見・ご要望を募集しております。巻末の「この文書について」をご参照ください。

図 1 はデータ公開までの基本的な流れを示したものです。

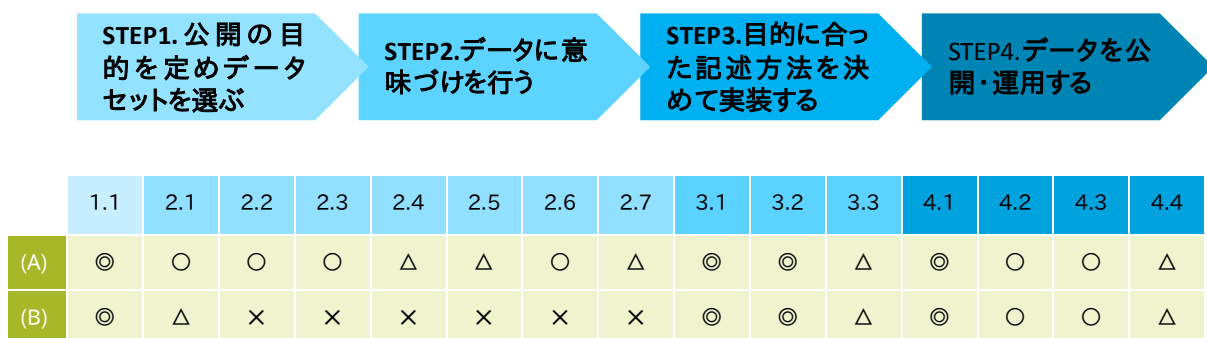


図 1 データ公開までの基本的な流れ

図 1 中にある表の 1 行目に示した数字は、この手引きの章番号と対応しています。

- | | |
|--------------------------|------------------------|
| 1.1 公開データセットの選定とデータセット定義 | 3.1 公開ファイル形式や公開場所などの決定 |
| 2.1 公開対象とするデータ項目の検討 | 3.2 データレコードの入力・追加・削除 |
| 2.2 データ項目・記載内容・データ構造の整理 | 3.3 データの変換 |
| 2.3 データ項目の検証 | 4.1 データの公開 |
| 2.4 応用語彙の定義 | 4.2 データ公開の告知 |
| 2.5 コード、ID の整備 | 4.3 公開したデータの運用 |
| 2.6 識別子の定義 | 4.4 データ項目定義書の公開 |
| 2.7 データ項目定義書の作成 | |

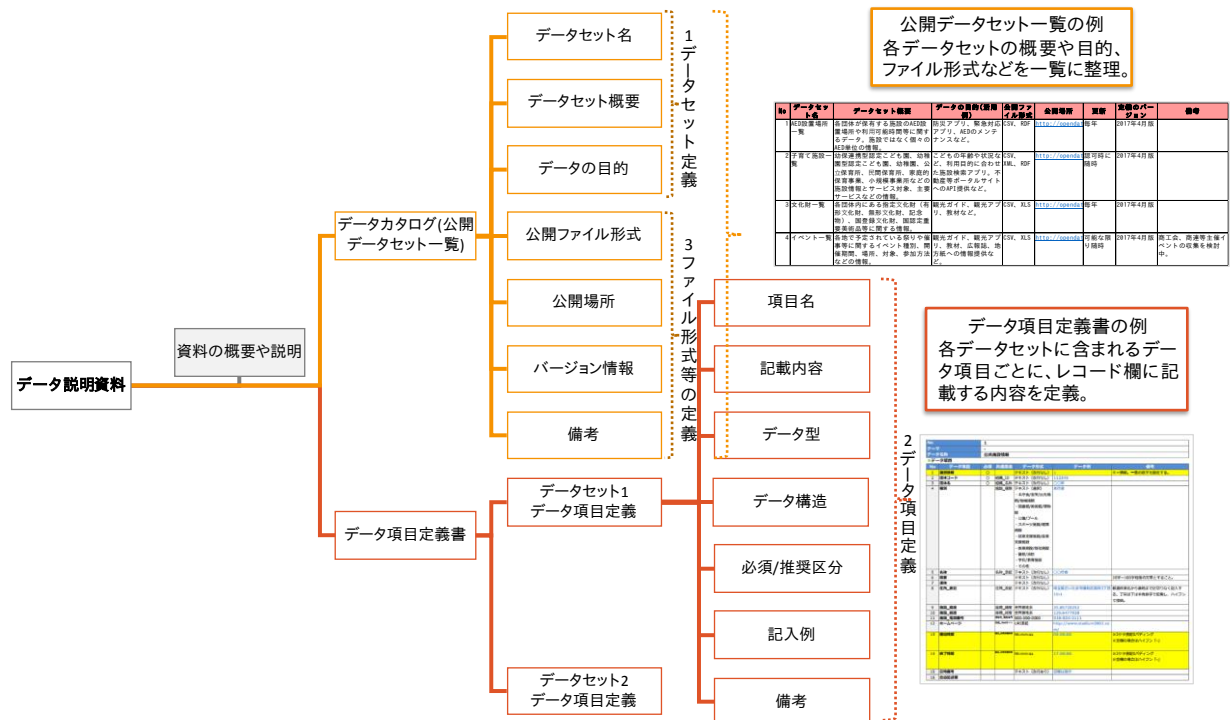
また、図 1 では、データ公開までの手順を次の(A)、(B)のパターンに分けています。

(A) データセットやデータ項目を一から定義する場合の手順

(B) 「推奨データセット」の例などの「既存データセット定義(既存フォーマット)」を利用する場合の手順

また、図 1 中の◎、○、△、×は、以下の意味です。手順(B)を選ぶと、多くの作業を省略することができます。

- ◎ 必ず行うべき作業
- 実施を推奨する作業
- △ 可能であれば実施したい作業
- × 不要な作業



■STEP 1 では公開するデータセットを選び、データセット定義を行う。■STEP 2 でデータ項目定義(赤枠の部分)を定義し、■STEP 3 でデータセットの公開形式を決めてデータセットを作成する。

図 2 データセット定義とデータ項目定義の構成

■STEP1. 公開の目的を定めデータセットを選ぶ

- ▶ このステップではデータ公開の目的を確定し、その目的に沿って、公開するデータを決めます。
- ▶ 「公共施設一覧」、「文化財一覧」など、ある目的や用途のために公開されるデータの集まりを、「データセット」と呼びます。
- ▶ 公開するデータセットの名称、概要、目的などをまとめたものを「データセット定義」と呼びます。

1.1 公開データセットの選定とデータセット定義

● 作業の内容

- ☑ 想定される利用者や利用シーンを明確にし、公開データセットを検討する。
- ☑ 情報開示請求数やホームページ、既存のデータセットへのアクセスログなどを参考にして、あるいはアイデアソンなどを通じて、利用者のニーズを反映することが望ましい。
- ☑ ニーズ把握が難しい場合など、まずはデータセットの相互運用性を優先するという判断もありうる。そうしたケースでは、他団体等がすでに公開しているデータセットと共通のデータセットを選ぶ方法も考えられる。
- ☑ 公開対象とするデータセットの名称、概要、目的、更新頻度などを定義し、データセット定義にまとめる。各データセット定義を一覧にまとめると、データカタログになる。

【開始時に用意する情報】

- ・ 保有する既存データ

【得られる成果物】

- ・ 公開対象とするデータセットの一覧
- ・ データセット定義

【外部参照資料・ツール等】

- ・ 開示請求の数、ウェブサイトのアクセスログなど（利用者ニーズ）
- ・ 政府が公開している「推奨データセット」<https://cio.go.jp/policy-opendata>
- ・ 類似データを扱う他自治体等の団体が作成したデータセット定義
- ・ Dublin Core、Schema.org などその他の団体が作成した語彙

作業例① 公開するデータセットの検討～埼玉県の場合¹

県内市町村が共通フォーマットに沿って作成した 10 種のデータセットをとりまとめて公開している埼玉県では、公開データセットの選定にあたって、公開データを保有する市町村の担当者、データを活用するサービス提供者や支援技術提供者などから構成される官民協働ワーキンググループを開催。データセットへのニーズや活用した場合の価値と可能性、また市町村の対応の可否などを協議しながら検討を行った（図 3、図 4）。

ワーキンググループに先立って実施した自治体アンケート調査や前年度の自治体オープンデータ勉強会で集めた意見、また「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)」²等をふまえ、「オープンデータの意義」「シティプロモーション」「観光」「子育て・健康」「安心・安全」「癒し」5つのテーマを設定した。

共通データ検討用紙	
テーマ	シティプロモーション
データ名称	定住・移住
想定更新頻度	毎年・毎月・毎週・毎日・ <u>随時</u> ・その他()
データ概要	・県内市町村への転入を促進するための情報。 ・定住を促進するための情報。 ・不動産情報サイト。
想定活用先例	・移住・移住サイト ・分譲住宅の場所・間取り、附帯設備、家賃、借年数、名称。(分譲住宅) ・学区。(学区) ・犯罪率。(学区)
備考	公開は困難かもしれないが、学区、学区、スーパー、駅、駅などの情報をよく活用するニーズは高い。
メモ	
・職の情報は提供できない。人事部門。(市職員採用情報) ・学区/LOD での活用が検討されている。	

想定活用先を起点に、どんなデータが有益か検討

既存データ有無などの制約を設けずにアイデア出し

図 3 データセット検討シートの例

¹ 埼玉県の取り組み事例は、IMI サイトでも紹介しています。 <https://imi.go.jp/event/0005.html>

² 政府、独立行政法人、地方公共団体等のオープンデータ推進に関する考え方を示した方針。2013 年 6 月 25 日 各府省情報化統括責任者(CIO)連絡会議決定。 <https://www.kantei.go.jp/jp/singi/it2/cio/dai52/kihon.pdf>

共通データ候補詳細(7/20段階)	
< 優先順位づけの判断基準 >	
基準項目	評価
データ提供の可否の可能性	1 2 3 4 5
活用の可能性	1 2 3 4 5
活用の価値	1 2 3 4 5
データ名称	特産品
テーマ	シティプロモーション
データ概要	・各自治体の銘産品、加
提供団体	
想定頻度	随時
想定更新頻度	随時
想定活用先例	観光ガイドブック 不動産情報サイト カレンダー
備考	可能であれば写真
活用イメージ	
データ項目	品名 販売店 商工部門 分類 問い合わせ先 農業政策部門 製品ジャンル
データ名称	デジタルアーカイブ
テーマ	シティプロモーション
データ概要	・過去の画像データ
提供団体	
想定頻度	随時
想定更新頻度	随時
想定活用先例	教科書 建築、開発現場 観光業への提供
備考	位置 方位 場所 年代 ジャンル
活用イメージ	
データ項目	市史編纂 文化財

図 4 スコアリングによるデータセット評価の例

ワーキンググループでの検討の結果、県下の各市町村が公開する 5 テーマ・10 の共通データセット（表 1）が選定された。併せて、データ公開後に想定するデータ更新頻度もデータセットごとに「随時」「毎月」「毎年」などと定義した。

表 1 データセット一覧の例

No	テーマ	データ名称	データ概要	提供団体	想定更新頻度
1	シティプロモーション	公共施設情報	各自治体毎の公共施設の住所、連絡先及び位置情報等のデータ	県・市町村	随時
2		イベントカレンダー	季節のイベントやお祭りなどの行事や地域の時勢に合わせた見どころに関するデータ	県・市町村	随時
3		広報誌URL	各団体が発行する広報誌の掲載URLの情報	県・市町村	随時
4	観光	観光地情報	観光施設、名所、景勝地等に関するデータ	県・市町村	毎年
5		地元グルメ情報	地元B級グルメや地産地消のグルメなどを取り扱う飲食施設に関するデータ	県・市町村	随時
6		文化財一覧	各団体内にある指定文化財（有形文化財、無形文化財、記念物）、国登録文化財、国認定重要美術品等に関するデータ	県・市町村	随時
7	子育て・健康	保育園・幼稚園情報	保育園、幼稚園の基本情報や一時預りに関する情報などのデータ	県・市町村	随時
8	安心・安全	AED設置場所情報	各団体が保有する施設のAED設置場所や利用可能時間等に関するデータ	県・市町村	毎年
9	癒し	景観情報	季節の花や水辺、夕日、花火など景観に関するデータ	県・市町村	随時
10		ご当地キャラ情報	各団体のご当地キャラのプロフィールや出演イベント等に関するデータ	県・市町村	毎月

データセット定義としては、それぞれの要件や活用先を整理し、共通フォーマットのバ

ージョンや更新頻度などを記述した（図 5）。

一般に、公開するデータの名称、概要、目的（活用例）、公開ファイル形式、公開場所、更新頻度、定義のバージョンなどを記述した、一覧を作成することが望ましい。

共通データフォーマット	
No	10
テーマ	産し
データ名称	ゆるキャラ（プロフィール） 各団体のゆるキャラのプロフィールや出演イベント等に関するデータ
データ概要	
提供条件	<ul style="list-style-type: none"> ・ゆるキャラのプロフィール情報。 ・ゆるキャラ一覧。 ・ゆるキャラの各イベントの出演情報。 ・自治体で持っている「ゆるキャラ」に限定する。
想定活用先例	観光ガイドブック、イベント記事サイト、イベントカレンダー、HP、ゆるキャラと写真を撮る（AR など）、ぬいぐるみ商品化
活用イメージ	
備考	・ゆるキャラの画像使用や商品化については、権利関係の検討が必要。
■フォーマット情報	
作成日	
更新日	
バージョン	
ファイル形式	
■提供提供情報	
提供団体	県・市町村
提供課課	商工課・市民活動課
提供更新頻度	毎年
■検討情報	
検討対象	○
検討職位	3
データ提供の可否の可能性	4
活用の可能性	3
活用の価値	2

データセットごとの要件や活用先を整理

図 5 データセット定義の記述例

👉 作業のポイント

- ➡ 利用者のニーズが高く、データを揃えやすいデータセットの公開から着手することが望ましい。
- ➡ 公開後のデータセット運用を含め、データ作成・データ更新の作業量も考慮してデータセットを選ぶ。データ更新の際の過去のデータセットの取り扱いなども定義しておくことが望ましい。
- ➡ 政府や類似団体が作成したデータセット定義を利用することにより、作業の簡略化が期待できる。ただし、公開しようとする既存データセットのデータ項目を、1 対 1 では割り当てられない場合がある。その場合は、■STEP2 の段階で既存データセットの内容を考慮し、データ項目を検討する。
- ➡ 類似するデータセットの例を参考に、データ項目や構成を利用シーンやニーズの観点から検討する。[→作業例①]

■STEP2. データに意味づけを行う

- ▶ このステップでは、■STEP1 で選んだデータセットごとに、公開対象とするデータの項目を選び、データ項目や構造の整理を行います。
- ▶ それぞれのデータ項目に IMI コア語彙の割り当て（意味づけ）をしていきます。
- ▶ 語彙の割り当てに際しては、公開するデータの内容に合わせて各データ項目の項目名（ラベル）・データ型などを段階的に定義していきます。
- ▶ なお、この章で行う作業の多くは、既存のデータセット定義を用いることにより、大幅に削減できます。

■STEP1 で公開対象としたデータセットは、全く新しく作成することもあるかもしれませんが、例えば、表 2 のような表形式の既存データセットが存在することが多いと思われます。こういったデータセットが、本手引きに沿って公開するデータの出発点となります。

■STEP2 では、このような既存データについて、公開するデータ項目の絞り込み、データ項目の意味を明確にし、さらに、その項目名を整理していきます。

表 2 利用可能な既存データセットの例

ページタイトル	施設ジャンル	施設、場所、イベントの名称（読み）	郵便番号	住所	ビル名	フロア数	緯度	経度
●●市民活動支援センター	市役所・公民館	〇〇シミンカ	000-0013	●●市□□2-5-1	●●中央	9 階	nn.609	oo.121
××小学校	非常用井戸	〇〇〇〇シヨ	000-0013	●●市××2-15-1			nn.607	oo.114
●●市 児童相談所	その他市役所	ジドウソウダ	000-0014	●●市△△3丁目2-3			nn.620	oo.065

2.1 公開対象とするデータ項目の検討

● 作業の内容

- ☑ ■STEP1 で選定したデータセットのそれぞれについて、具体的に公開するデータ項目の種類や内容を検討する。データセット定義に記載した用途や利用方法をふまえて、更新の可否などを考慮して公開対象とするデータ項目の取捨選択を行う。
- ☑ データ項目が、公開するデータセットの用途や実際活用される際の利用方法に適合しているかどうかを確認し、追加すべき(足りない)データ項目、削除すべき(該当するデータレコードがない、公開しないなど)データ項目を検討する。
- ☑ 類似したデータセットやデータセット定義が公開されている場合は、データ項目定義の参考にすることが望ましい。

【開始時に用意する情報】

- ・ 1.1 で作成した、公開対象とするデータセット定義（図 5）
- ・ 既存データセットのデータ項目一覧（既存のデータセットを公開する場合）

【得られる成果物】

- ・ 公開データセットごとのデータ項目の一覧

【外部参照資料・ツール等】

- ・ IMI コア語彙 <http://imi.go.jp/ns/core/2>
- ・ 政府が公開している「推奨データセット」 <https://cio.go.jp/policy-opendata>
- ・ 類似データを扱う他自治体等の団体、また Dublin Core、Schema.org などその他の団体が作成したデータセット

1.1 で定義したデータセットごとに、利用目的や概要をふまえ、どのようなデータ項目実際に公開する対象とするかを決め、書き出していきます。その検討例を表 3 に示します。

表 3 データ項目の検討例

データ名称	特産品
テーマ	シティプロモーション
データ概要	・各自治体の銘産品、加工品農産品など
提供団体	
想定原課	
想定更新頻度	随時
想定活用先例	観光ガイドブック 不動産情報サイト カレンダー
備考	可能であれば写真
活用イメージ	

データ項目	品名 販売店 商工部門 分類 問い合わせ先 農業政策部門 製品ジャンル
-------	---

2.2 データ項目・記載内容・データ構造の整理

● 作業の内容

- ☑ 構造が複雑な既存のデータ表を用いる場合は、機械判読性や相互運用性を確保するため、データ項目やレコードの記述方法を整理することが望ましい。
- ☑ 2.1 で選定したデータ項目に対して、項目名（ラベル）、記載内容（記載例）、データ型、必須／任意の種別などを、IMI コア語彙に対応させて整理していく。IMI コア語彙からデータ項目に相当する用語を探し、各データ項目定義に割り当てる。
- ☑ IMI コア語彙で表現できないデータ項目や、独自に設定した方がデータの意味を表現しやすいデータ項目の場合は、IMI コア語彙を割り当てなくてよい。対応欄を空白にしておき、後のステップで応用語彙を検討する（2.4 を参照）。

【開始時に用意する情報】

- ・ 2.1 で作成した公開対象とするデータ項目の一覧（表 3 のデータ項目）

【得られる成果物】

- ・ 公開対象とするデータ項目とデータ構造
- ・ IMI コア語彙との対応有無（データ項目ごとに種別）

【外部参照資料・ツール等】

- ・ IMI コア語彙 <http://imi.go.jp/ns/core/2>

● このステップでの検討・定義の対象

- ・ 項目名（ラベル）
- ・ 記載内容（記載例）
- ・ データ型（文字列、日付、時刻といったデータの形式）
- ・ 必須／推奨の種別
- ・ IMI コア語彙との対応

図 6 に、IMI のコア語彙に沿ってデータの項目名を整理する例を示します。

上表は整理する前のデータ。その一行目に、既存の項目名が並んでいます。データの内容を確認しながら、IMI コア語彙の「施設型」等に対応させ、下表のように項目名を整理していきます。

ページタイトル	施設ジャンル	施設、場所、イベントの名称（読み）	郵便番号	住所	ビル名	フロア数	緯度	経度
●●市民活動支援センター	市役所・市立施設	〇〇〇シミンカ	000-0013	●●市口〇2-5-1	●●中央	9階	nn.605	oo.121
××小学校	非常用井	〇〇〇〇シヨ	000-0013	●●市×2-15-1			nn.607	oo.114
●●市 児童相談所	その他市	ジドウソウダ	000-0014	●●市△△3丁目2-3			nn.624	oo.065



データ項目（ラベル）	ページタイトル	施設ジャンル	施設、場所、イベントの名称（読み）	郵便番号	住所	ビル名	フロア数	緯度	経度
コア語彙との対応	施設>名称>表記	施設>種別							
レコード1	●●市民活動支援センター	市役所・市立施設	〇〇〇シミンカ	000-0013	●●市口〇2-5-1	●●中央	9階	nn.605	oo.121
レコード2	××小学校	非常用井	〇〇〇〇シヨ	000-0013	●●市×2-15-1			nn.607	oo.114
レコード3	●●市 児童相談所	その他市	ジドウソウダ	000-0014	●●市△△3丁目2-3			nn.624	oo.065
レコード4	〇〇小学校	非常用井	〇〇〇〇シヨ	000-0014	●●市口〇区〇〇2-6-23			nn.605	oo.128

図 6 IMI 語彙をもちいたデータ項目定義の例

IMI のコア語彙には、図 7 のような基本的な用語、つまり「クラス用語」が用意されています。このなかから、整理しようとしているデータセットのデータ項目に最も近いもの、ここでは「施設」を選びます。

<http://imi.go.jp/ns/core/2> で公開している IMI コア語彙。図 6 のデータセットの場合、ここで「施設型」をクリックすると「施設型」に用意されたプロパティの一覧（図 8）が表示される。

図 7 IMI コア語彙のクラス用語一覧³

³ クラス用語やプロパティ用語については、「IMI 共通語彙基盤入門」で説明しています。
https://imi.go.jp/doc/IMI_introduction_20171227.pdf

クラス用語「施設」には、図 8 のような「プロパティ」が用意されています⁴。

施設型

概念型 事物型 場所型 地物型 施設型

施設を表現するためのクラス用語

施設型のプロパティ			
識別子	値型	回数	説明
ic:種別	xsd:string	0..n	種別の表記を記述するためのプロパティ用語
ic:種別コード	ic:コード型	0..n	種別コードを記述するためのプロパティ用語
ic:メタデータ	ic:文書型	0..n	データのメタデータを記述するためのプロパティ用語
ic:ID	ic:ID型	0..n	事物に割り振られたID（識別子）を記述するためのプロパティ用語
ic:表記	xsd:string	0..n	文字列による表記を記述するためのプロパティ用語
ic:参照	ic:参照型	0..n	追加情報などを参照するURLを記述するためのプロパティ用語
ic:画像	xsd:anyURI	0..n	画像を参照するURLを記述するためのプロパティ用語
ic:説明	xsd:string	0..n	説明を記述するためのプロパティ用語
ic:記述	ic:記述型	0..n	種別付きの説明を記述するためのプロパティ用語
ic:名称	ic:名称型	0..n	場所を特定する名称（地名など）を記述するためのプロパティ用語
ic:通称	xsd:string	0..n	場所の正式な名称以外に一般的に使用されている名称を記述するためのプロパティ用語
ic:地理識別子	ic:ID型	0..1	場所の地理識別子を記述するためのプロパティ用語
ic:住所	ic:住所型	0..1	場所を表す住所を記述するためのプロパティ用語
ic:地理座標	ic:座標型	0..1	場所の経緯度座標を記述するためのプロパティ用語
ic:要約	xsd:string	0..n	地物の説明の要約を記述するためのプロパティ用語
ic:アクセス	ic:アクセス型	0..n	地物へのアクセス方法を記述するためのプロパティ用語
ic:連絡先	ic:連絡先型	0..n	地物の連絡先を記述するためのプロパティ用語
ic:設備	ic:設備型	0..n	地物に定常的に備わっている設備を記述するためのプロパティ用語
ic:管理者	ic:実体型	0..n	地物を管理している組織や人を記述するためのプロパティ用語
ic:関与	ic:関与型	0..n	地物に関連する組織や人を記述するためのプロパティ用語
ic:利用可能時間	ic:期間スケジュール型	0..n	施設の利用可能日と利用可能時間を記述するためのプロパティ用語
ic:料金	ic:価格型	0..n	施設の利用料金を記述するためのプロパティ用語
ic:収容人数	ic:数量型	0..n	施設の収容人数を記述するためのプロパティ用語
ic:駐車場	ic:駐車場型	0..n	施設にある駐車場を記述するためのプロパティ用語
ic:建物	ic:建物型	0..n	施設のある建物を記述するためのプロパティ用語
ic:関連施設	ic:施設関連型	0..n	関連する施設（上位施設、下位施設など）を記述するためのプロパティ用語
ic:備考	xsd:string	0..1	備考を記述するためのプロパティ用語 ※補足情報などがある場合に記述

施設型の構造

ic:種別 (xsd:string)

ic:メタデータ (ic:文書型)

ic:ID (ic:ID型)

ic:表記 (xsd:string)

ic:参照 (ic:参照型)

ic:画像 (xsd:anyURI)

ic:説明 (xsd:string)

ic:記述 (ic:記述型)

ic:名称

データ型: ic:名称型

ic:種別 (xsd:string)

ic:種別コード (ic:コード型)

ic:メタデータ (ic:文書型)

ic:ID (ic:ID型)

ic:表記 (xsd:string)

ic:参照 (ic:参照型)

ic:画像 (xsd:anyURI)

ic:説明 (xsd:string)

ic:記述 (ic:記述型)

ic:カナ表記 (xsd:string)

ic:ローマ字表記 (xsd:string)

ic:通称 (xsd:string)

ic:地理識別子 (ic:ID型)

ic:住所 (ic:住所型)

ic:地理座標 (ic:座標型)

ic:要約 (xsd:string)

ic:アクセス (ic:アクセス型)

ic:連絡先 (ic:連絡先型)

ic:設備 (ic:設備型)

ic:管理者 (ic:実体型)

ic:関与 (ic:関与型)

ic:利用可能時間 (ic:期間スケジュール型)

ic:料金 (ic:価格型)

ic:収容人数 (ic:数量型)

ic:駐車場 (ic:駐車場型)

ic:建物 (ic:建物型)

ic:関連施設 (ic:施設関連型)

ic:備考 (xsd:string)

施設型の継承関係

ic:概念型

ic:事物型

ic:場所型

ic:地物型

ic:施設型

施設型は以下のクラスに継承されます

クラス	説明
ic:駐車場型	駐車場を表現するためのクラス用語

図 8 IMI コア語彙のクラス用語（施設型）におけるプロパティ一覧

⁴ IMI コア語彙は、<https://imi.go.jp/ns/core/2> をご参照ください。また、語彙の構造等を説明した資料として、「IMI 共通語彙基盤入門」https://imi.go.jp/doc/IMI_introduction_20171227.pdf をご参照ください。

さて、**図 6** の例で、既存の項目名にある「ページタイトル」に、IMI コア語彙に用意された用語を対応づけるにはどうすればよいでしょう？

データを見ると、「●●市民活動支援センター」、「××小学校」などが並んでいますが、これは、施設の名称を文字列で記述したもののようです。

コア語彙の「施設型」には、「名称」というプロパティがあり、これが**図 6** のデータ項目「ページタイトル」中のレコード（「××小学校」といった文字の並び）を入れる先として適しているようです。**図 8** を詳しく見ると、「名称」は「名称型」というクラスを用いることとなっており、「名称型」は、さらに「種別」、「表記」、「参照」などのプロパティから構成されていることがわかります⁵。そのうち「表記」は「値型」が”xsd:string”となっている、すなわち、文字列で自由にそのレコードを表現できるプロパティです⁶。

結果、既存データセットに含まれるデータ項目「ページタイトル」は、コア語彙の「施設型」の中にある「名称」というプロパティの、さらにその中にある「表記」というプロパティ用語に対応付けるのが適切だと判断できます。**図 6** の例では、このデータ項目の構造を理解しやすくするために、「施設＞名称＞表記」と表現します。この表現の方法に関する詳細は、「構造化項目名⁷」の説明を参照してください。

次に、既存データセットにある「施設ジャンル」というデータ項目名について検討してみます。「施設ジャンル」が、各施設の種別を表すものであれば、IMI コア語彙の「施設型」にある「種別」というプロパティへの対応付けが最も適していると考えられます。IMI コア語彙の「施設型」にある「種別」を調べると、このプロパティ用語は「表記」同様に”xsd:string”つまり、文字列を表現できることがわかります。したがって、「施設ジャンル」は「施設＞種別」に対応付けることが適当です。

こうして、既存データセットにあったデータ項目を、コア語彙と結び付けていったのが、**図 6** の下表の青色で示された行です。

では、実際の作業例を見てみましょう。既存データセットは様々な形式をしていること

⁵ もし、この小学校が、正式名称の他に、愛称を持っており、それも書きたいような場合に備え、「名称」というクラスには「種別」（例えば「愛称」）も書けるよう、「種別」というプロパティも用意されています。「コア語彙」は幅広い用途に用いることができるよう、いろいろな用語が用意されていますが、通常使うのはそのごく一部です。

⁶ ”xsd:string” は、「そこは文字列（文字の並び）で記述する」、ということを示す国際標準の言葉です。

⁷ 構造化項目名： <https://imi.go.jp/goi/goi-about.html>

があり、コア語彙と結び付けるために**作業例②**のような前準備が必要な場合があります。

作業例② 複数行で表現されたデータ項目等の整理

図 9 の例では既存データの「区分」のデータ項目が二層構造であり、情報セキユマネ、基本情報、応用情報といった試験群のひとつである「情報セキユマネ」という区分の下に「応募者」「受験者」「合格者」が記述されている。この場合、それぞれのデータ項目を「情報セキユマネ試験の応募者」「情報セキユマネ試験の受験者」「情報セキユマネ試験の合格者」と階層化して整理した方が、機械可読性は高まる。同様に「勤務先」としてまとめられているデータ項目も、「勤務区分の社会人の製造業務」「勤務区分の学生の大学院生」などと、それぞれあらたに升目を設けて明示することが望ましい。

A		B	C	D	E	F	G	H	I	J	K	L
勤務先		区分	情報セキュリティ			基本情報			応用情報			その他
			応募者	受験者	合格者	応募者	受験者	合格者	応募者	受験者	合格者	応募者
1	社会人	ソフトウェア業	3,442	2,756	2,399	11,048	7,480	2,388	14,686	9,167	1,667	7,324
2			81.7	93.1	-3.4	67.7	31.9	-7.1	62.4	18.1	-11.2	
3		情報処理・提供サービス業	4,452	3,681	3,248	7,931	5,268	1,205	7,870	4,890	835	3,444
4			82.7	88.2	-3.2	66.4	22.9	-2.9	61.4	17.3	-1.9	
5		コンピュータ及び周辺機器製造又は販売業	1,313	1,115	935	771	497	133	590	395	100	484
6			84.9	83.9	-22.6	64.5	26.8	-10.2	66.9	25.3	-15.7	
7		農業、林業、漁業、鉱業	20	16	13	17						
8				80.0	81.3	-6.3						
9		建設業	343	284	253	403						
10			82.8	89.1	8.3							
11		製造業	1,146	994	930	1,598						
12			86.7	93.6	-8.2							
13		電気・ガス・熱供給・水道業	153	124	115	163						
14			81.0	82.7	-23.8	65.0	40.6	-9.2	68.1	33.8	-4.2	
15						492	304	89	401	261	97	274
16						-6.3	61.8	29.3	-4.8	65.1	37.2	-9.0
17						305	192	50	157	98	37	93
18						-11.8	63.0	26.0	-8.7	62.4	37.8	0.0
19						603	379	146	514	318	79	390
20						-10.1	62.9	38.5	7.1	61.9	24.8	-2.3
21		サービス業	1,236	1,029	901	928	609	153	567	372	70	265
22			83.3	87.6	-13.6	65.6	25.1	-2.7	65.6	18.8	-20.4	
23		調査業、広告業	56	41	37	57	32	13	10	9	3	9
24				73.2	90.2	-9.5	56.1	40.6	-21.7	50.0	33.3	-52.6
25		医療・福祉業	272	223	198	180	94	34	84	52	11	20
26			82.0	88.8	-11.3	52.2	36.2	-6.7	61.9	21.2	42.9	

勤務区分		勤務/学生の区分	データ種別	情報セキュリティ 応募者	情報セキュリティ 受験者	情報セキュリティ 合格者	基本情報 応募者	基本情報 受験者	基本情報 合格者	応用 応
1	社会人	ソフトウェア業	実数	3542	2895	2696	11048	7480	2388	
2	社会人	ソフトウェア業	割合	31.7	93.1	-3.4	67.7		31.9	
3	社会人	情報処理・提供サービス業	実数	4452	3681	3248	7931	5268	1205	
4	社会人	情報処理・提供サービス業	割合	32.7	88.2	-3.2	66.4		22.9	
5	社会人	コンピュータ及び周辺機器製造又は販売業	実数	1313	1115	935	771	497	133	
6	社会人	コンピュータ及び周辺機器製造又は販売業	割合	84.9	83.9	-22.6	64.5		26.8	
7	社会人	農業、林業、漁業、鉱業	実数	20	16	13	17		1	
8	社会人	農業、林業、漁業、鉱業	割合	80	81.3	-55.3	58.8		10	
9	社会人	建設業	実数	343	284	253	403		237	
10	社会人	建設業	割合	82.8	89.1	8.3	58.8		24.5	
11	社会人	製造業	実数	1146	994	930	1598		1062	
12	社会人	製造業	割合	86.7	93.6	-8.2	65.8		41.5	
13	社会人	電気・ガス・熱供給・水道業	実数	153	124	115	163		29	
14	社会人	電気・ガス・熱供給・水道業	割合							
15	社会人	運輸・通信業	実数							
16	社会人	運輸・通信業	割合							
17	社会人	卸売・小売業、飲食店	実数							
18	社会人	卸売・小売業、飲食店	割合							
19	社会人	金融・保険業、不動産業	実数							
20	社会人	金融・保険業、不動産業	割合							
21	社会人	サービス業	実数							
22	社会人	サービス業	割合							
23	社会人	調査業、広告業	実数	56	41	37	57	32	13	
24	社会人	調査業、広告業	割合		73.2	90.2	-9.5	56.1	40.6	
25	社会人	医療・福祉業	実数	272	223	198	180	94	34	
26	社会人	医療・福祉業	割合		82	88.8	-11.3	52.2	36.2	
27	社会人	教育(学校、研究機関)	実数	575	487	449	195	119	43	
28	社会人	教育(学校、研究機関)	割合		84.7	92.2	-29.1	61	36.1	
29	社会人	官公庁、公益団体	実数	1056	883	837	501	298	94	

図 9 データ項目の整理例(その1)

また、図 9 では、データの各セルを上下二段に分け、「実数」と「割合」を表示している。このような場合、図 10 のようにそれぞれに対し「実数」「割合」の行を設け、別のセルに数値を記入したほうが、データの意味が伝わりやすい。データ構造を整理することにより、E 列 12 行が「社会人の製造業勤務者」の「情報セキュリティ受験者」の「実数」の値として特定でき、機械判読性が高まる。

また「実数」と「割合」のようにデータ型が異なるデータは、別のデータセットとして表を分割したほうが使いやすことが多い。

	A	B	C	D	E	F	G	H	I	J	K
	勤務区分	勤務/学生の区分	データ種別	情報ヒューマン応募者	情報ヒューマン受験者	情報ヒューマン合格者	基本情報応募者	基本情報受験者	基本情報合格者	応用情報応募者	応用情報受験者
1	社会人	ソフトウェア業	実数	3542	2895	2696	11048	7480	2388	14686	9157
2	社会人	ソフトウェア業	割合		81.7	93.1	-3.4	67.7	31.9	-7.1	62.4
3	社会人	情報処理・提供サービス業	実数	4452	3681	3248	7931	5268	1205	7870	4830
4	社会人	コンピュータ及び周辺機器製造又は販売業	割合		82.7	88.2	-3.2	66.4	22.9	-2.9	61.4
5	社会人	農業、林業、漁業、鉱業	実数	1313	1115	935	771	497	133	590	395
6	社会人	建設業	割合		84.9	83.9	-22.6	64.5	26.8	-10.2	66.9
7	社会人	製造業	実数	20	16	13	17	10	1	11	8
8	社会人	電気・ガス・熱供給・水道業	割合		80	81.3	-55.3	58.8	10	-21.4	72.7
9	社会人	運輸・通信業	実数	343	284	253	403	237	58	208	131
10	社会人	卸売・小売業、飲食店	割合		82.8	89.1	8.3	58.8	24.5	-7.6	63
11	社会人	金融・保険業、不動産業	実数	1146	994	930	1598	1052	437	1438	940
12	社会人	サービス業	割合		84.1	93.8	-10.1	62.9	38.5	7.1	61.9
13	社会人	調査業、広告業	実数	1236	1029	901	929	609	153	567	372
14	社会人	医療・福祉業	割合		83.3	87.6	-13.6	65.6	25.1	-2.7	65.6
15	社会人	教育(学校、研究機関)	実数	56	41	37	57	32	13	18	9
16	社会人	官公庁、公益団体	割合		73.2	90.2	-9.5	56.1	40.6	-21.7	50
17	社会人	無職、その他無記入	実数	272	223	198	180	94	34	84	52
18	社会人	学生	割合		82	88.8	-11.3	52.2	36.2	-6.7	61.9
19	社会人	短大	実数	575	487	449	195	119	43	145	92
20	社会人	大学院	割合		84.7	92.2	-29.1	61	36.1	-23.7	63.4
21	社会人	大学院	実数	1056	883	837	501	298	94	291	193
22	社会人	大学院	割合		83.6	94.8	-19.5	59.5	31.5	-4.6	66.3
23	社会人	大学院	実数	4498	3576	3111	17447	1913	3548	12576	7595
24	社会人	大学院	割合		79.5	87	-0.8	68.3	26.8	-3.9	60.1

意味構造

- 社会人の製造業勤務者の実数(人)
- 社会人の製造業勤務者の割合(%)

値の意味ごとにデータセットを分割
(C列は不要になるので削除)

A	B	C	D	E	F	G	H
勤務区分	勤務/学生の区分	データ種別	情報ヒューマン応募者	情報ヒューマン受験者	情報ヒューマン合格者	基本情報応募者	基本情報受験者
1	社会人	ソフトウェア業	実数	3542	2895	2696	11048
2	社会人	情報処理・提供サービス業	実数	4452			7480
3	社会人	コンピュータ及び周辺機器製造又は販売業	実数	1313			
4	社会人	農業、林業、漁業、鉱業	割合	20			
5	社会人	建設業	実数	343	284	253	403
6	社会人	製造業	割合	1146	994	930	1598
7	社会人	電気・ガス・熱供給・水道業	割合				
8	社会人	運輸・通信業	割合				
9	社会人	卸売・小売業、飲食店	割合				
10	社会人	金融・保険業、不動産業	割合				
11	社会人	サービス業	割合				
12	社会人	調査業、広告業	割合				
13	社会人	医療・福祉業	割合				
14	社会人	教育(学校、研究機関)	割合				
15	社会人	官公庁、公益団体	割合				
16	社会人	無職、その他無記入	割合				
17	学生	大学	割合				
18	学生	短大	割合				
19	学生	大学院	割合				

実数値だけのデータセット

A	B	C	D	E	F	G	H
勤務区分	勤務/学生の区分	データ種別	情報ヒューマン応募者	情報ヒューマン受験者	情報ヒューマン合格者	基本情報応募者	基本情報受験者
1	社会人	ソフトウェア業	割合			81.7	93.1
2	社会人	情報処理・提供サービス業	割合			82.7	88.2
3	社会人	コンピュータ及び周辺機器製造又は販売業	割合				
4	社会人	農業、林業、漁業、鉱業	割合				
5	社会人	建設業	割合				
6	社会人	製造業	割合			86.7	93.6
7	社会人	電気・ガス・熱供給・水道業	割合			81	92.7
8	社会人	運輸・通信業	割合			85.6	91.1
9	社会人	卸売・小売業、飲食店	割合			80.8	89
10	社会人	金融・保険業、不動産業	割合			84.1	93.8
11	社会人	サービス業	割合			83.3	87.6
12	社会人	調査業、広告業	割合			73.2	90.2
13	社会人	医療・福祉業	割合			82	88.8
14	社会人	教育(学校、研究機関)	割合			84.7	92.2
15	社会人	官公庁、公益団体	割合			83.6	94.8
16	社会人	無職、その他無記入	割合			79.5	87
17	学生	大学	割合			87.6	92.1
18	学生	短大	割合				
19	学生	大学院	割合				

割合値だけのデータセット

図 10 データ項目の整理例 (その2)

作業例③ IMI の「コア語彙」から既存データの項目に相当する用語を探して割り当てる

図 11 は横浜市職員有志勉強会「よこはま YY ラボ」による横浜のおすすめスポット一覧「YOKOHAMA.spot」である。ここでは各スポットの所在地や説明が、IMI コア語彙の「施設型」を利用して記述されている

「施設」のプロパティ

- 施設住所都道府県
- 施設住所都道府県コード
- 施設住所市区町村
- . . .

「施設」のプロパティ

- 施設名称
- 施設説明
- 施設参照
- . . .

	A	B	C	D	E	F	G	H	I	J	K	L
1	#LINK											
2	#lang	ja										
3	#attribution_name	よこはまYラボ										
4	#license	http://creativecommons.org/licenses/by/3.0/deed.ja										
5	#file_name	YOKOHAMA_spot										
6	#download_from	http://linkdata.org/work/rdf15372i										
7	#property	分類										
8	#object_type_xsd	string:ja	ic:都道府県	ic:都道府県コード	ic:市区町村	ic:市区町村コード	ic:区	町丁目	ic:名称	ic:説明	ic:参照	ic:備考
9	#property_context	Assertion	Assertion	Assertion	Assertion	Assertion	Assertion	Assertion	Assertion	Assertion	Assertion	Assertion
10	作成レベル	必須	必須	必須	必須	必須	必須	あれば	必須	必須	あれば	あれば
11	建物・歴史的建造物	神奈川県	14	横浜市	100	中区	中区	中区	中区	中区	中区	中区
12	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
13	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
14	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
15	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
16	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
17	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
18	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
19	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
20	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
21	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
22	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
23	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
24	水辺・広場・公園	神奈川県	14	横浜市	100	西区	西区	西区	西区	西区	西区	西区
25	建物・歴史的建造物	神奈川県	14	横浜市	100	中区	中区	中区	中区	中区	中区	中区
26	建物・歴史的建造物	神奈川県	14	横浜市	100	中区	中区	中区	中区	中区	中区	中区
27	建物・歴史的建造物	神奈川県	14	横浜市	100	中区	中区	中区	中区	中区	中区	中区
28	建物・歴史的建造物	神奈川県	14	横浜市	100	中区	中区	中区	中区	中区	中区	中区
29	建物・歴史的建造物	神奈川県	14	横浜市	100	中区	中区	中区	中区	中区	中区	中区
30	建物・歴史的建造物	神奈川県	14	横浜市	100	中区	中区	中区	中区	中区	中区	中区

IMI コア語彙との対応を明記した公開データセットの例。

http://linkdata.org/work/rdf15372i/YOKOHAMA_spot.html?key=#work_information

図 11 公開データ (YOKOHAMA.spot) における IMI コア語彙との対応例

IMI コア語彙の用語「施設型」を見ると、「名称」、「説明」、「参照」、「備考」、「住所」といったプロパティがあり(図 12 の赤い枠内)、それらはほぼそのまま「YOKOHAMA.spot」データセットのデータ項目定義と対応付けられそうだとということがわかる。

施設型			
概念型 事物理型 場所型 地物型 施設型			
施設を表現するためのクラス用語			
施設型のプロパティ			
識別子	値型	回数	説明
ic:種別	xsd:string	0..n	種別の表記を記述するためのプロパティ用語
ic:種別コード	ic:コード型	0..n	種別コードを記述するためのプロパティ用語
ic:メタデータ	ic:文書型	0..n	データのメタデータを記述するためのプロパティ用語
ic:ID	ic:ID型	0..n	事物に割り振られたID (識別子) を記述するためのプロパティ用語
ic:表記	xsd:string	0..n	文字列による表記を記述するためのプロパティ用語
ic:参照	ic:参照型	0..n	追加情報などを参照するURLを記述するためのプロパティ用語
ic:画像	xsd:anyURI	0..n	画像を参照するURLを記述するためのプロパティ用語
ic:説明	xsd:string	0..n	説明を記述するためのプロパティ用語
ic:記述	ic:記述型	0..n	種別付きの説明文を記述するためのプロパティ用語
ic:名称	ic:名称型	0..n	場所を特定する名称 (地名など) を記述するためのプロパティ用語
ic:通称	xsd:string	0..n	場所の正式な名称以外に一般的に使用されている名称を記述するためのプロパティ用語
ic:地理識別子	ic:ID型	0..1	場所の地理識別子を記述するためのプロパティ用語
ic:住所	ic:住所型	0..1	場所を表す住所を記述するためのプロパティ用語
ic:地理座標	ic:座標型	0..1	場所の経緯度座標を記述するためのプロパティ用語
ic:要約	xsd:string	0..n	地物の説明の要約を記述するためのプロパティ用語
ic:アクセス	ic:アクセス型	0..n	地物へのアクセス方法を記述するためのプロパティ用語
ic:連絡先	ic:連絡先型	0..n	地物の連絡先を記述するためのプロパティ用語
ic:設備	ic:設備型	0..n	地物に定常的に備わっている設備を記述するためのプロパティ用語
ic:管理者	ic:実体型	0..n	地物を管理している組織や人を記述するためのプロパティ用語
ic:関与	ic:関与型	0..n	地物に関連する組織や人を記述するためのプロパティ用語
ic:利用可能時間	ic:期間スケジュール型	0..n	施設の利用可能日と利用可能時間を記述するためのプロパティ用語
ic:料金	ic:価格型	0..n	施設の利用料金を記述するためのプロパティ用語
ic:収容人数	ic:数量型	0..n	施設の収容人数を記述するためのプロパティ用語
ic:駐車場	ic:駐車場型	0..n	施設にある駐車場を記述するためのプロパティ用語
ic:建物	ic:建物型	0..n	施設のある建物を記述するためのプロパティ用語
ic:関連施設	ic:施設関連型	0..n	関連する施設 (上位施設、下位施設など) を記述するためのプロパティ用語

<https://imi.go.jp/ns/core/Core241.html#ic:施設型>

図 12 施設のデータを表現するための IMI コア語彙「施設型」

IMI コア語彙「施設型」の「住所」の部分には、値型の欄に「住所型」と記載されているので、これをクリックすると「住所型」のプロパティ一覧が表示される（図 13）。「YOKOHAMA.spot」のデータ項目定義は、この住所型のプロパティから、「都道府県」、「都道府県コード」、「市区町村」、「市区町村コード」を割り当てている。

ただし、「YOKOHAMA.spot」にある「町丁目」という用語は IMI コア語彙にはない。IMI コア語彙では、「町名」、「丁目」、「番地」のように分離した用語となっている。このような場合は、「YOKOHAMA.spot」のように「町丁目」というプロパティ用語を独自に定義する方法（用語彙の作成）と、データレコードを IMI コア語彙に合わせて「町名」と「丁目」に分割する方法がある。語彙はあまり独自に作成せず、既存のものを用いるほうが、データの相互運用性にとっては望ましい。

住所型

概念型 → 事物型 → 住所型

住所を表現するためのクラス用語

住所型 のプロパティ			
識別子	値型	回数	説明
ic:種別	xsd:string	0..n	種別の表記を記述するためのプロパティ用語
ic:種別コード	ic:コード型	0..n	種別コードを記述するためのプロパティ用語
ic:メタデータ	ic:文書型	0..n	データのメタデータを記述するためのプロパティ用語
ic:ID	ic:ID型	0..n	事物に割り振られたID（識別子）を記述するためのプロパティ用語
ic:表記	xsd:string	0..n	文字列による表記を記述するためのプロパティ用語
ic:参照	ic:参照型	0..n	追加情報などを参照するURLを記述するためのプロパティ用語
ic:画像	xsd:anyURI	0..n	画像を参照するURLを記述するためのプロパティ用語
ic:説明	xsd:string	0..n	説明を記述するためのプロパティ用語
ic:記述	ic:記述型	0..n	種別付きの説明文を記述するためのプロパティ用語
ic:郵便番号	xsd:string	0..1	郵便番号を記述するためのプロパティ用語 ※0-9の数字7桁
ic:住所コード	ic:コード型	0..n	住所のコードを記述するためのプロパティ用語 ※地方公共団体情報システム機構の地方公共団体コード住所など
ic:国	xsd:string	0..1	国名を記述するためのプロパティ用語
ic:国コード	ic:コード型	0..1	国のコードを記述するためのプロパティ用語
ic:都道府県	xsd:string	0..1	都道府県名を記述するためのプロパティ用語
ic:都道府県コード	ic:コード型	0..1	都道府県のコードを記述するためのプロパティ用語
ic:市区町村	xsd:string	0..1	市区町村名を記述するためのプロパティ用語
ic:区	xsd:string	0..1	政令指定都市の区の名前を記述するためのプロパティ用語
ic:市区町村コード	ic:コード型	0..1	市区町村のコードを記述するためのプロパティ用語
ic:町名	xsd:string	0..1	町域、または市区町村の下行政区分の表記を記述するためのプロパティ用語
ic:丁目	xsd:string	0..1	丁目の表記を記述するためのプロパティ用語
ic:番地補足	xsd:string	0..1	番地の前、「東」「北」「浜」「甲」「イ」などの文字を記述するためのプロパティ用語
ic:番地	xsd:string	0..1	番、街区の表記を記述するためのプロパティ用語
ic:号	xsd:string	0..1	号、住居番号の表記を記述するためのプロパティ用語
ic:ビル名	xsd:string	0..1	ビル名を記述するためのプロパティ用語

図 13 IMI コア語彙「住所型」

👉 作業のポイント

- ➡ データ項目の語彙（用語）の名称を決める前に、そのデータ項目に含まれる実データ（値）の内容を確認し、実データの意味に合ったデータ項目名を検討することが望ましい。
- ➡ 項目内に階層が含まれている場合、データ表の各列は1つの意味だけをもつようにし、各セルには原則として複数の値が入らないように整理する。[→作業例②]
- ➡ IMI コア語彙として用意されている用語から、利用できる用語をみつけてデータ項目と

の対応を記入する。〔→作業例③〕

- ➡ IMI コア語彙を参考に、既存のデータセットで作成されている一般的な名称の項目（たとえば「連絡先」）を、さらに分割（「連絡先の名称」「連絡先の住所」「連絡先の電話番号」など）できないか検討する。
- ➡ 構造が複雑な項目については、構造の下位にある項目を意識しつつも、データ作成者の心理的負担を考慮し、この段階ではあえて「連絡先」などとしておき、まずは入力を促す工夫も必要である。

2.3 データ項目の検証

● 作業の内容

- ☑ データセット定義の各データ項目定義が IMI 共通語彙等により適切に意味付けされ、データの意味を適切に表しているかどうかを検証する。2.2 で作成した各データ項目を列見出しとして、1 行 1 レコードになるように、いくつか実データを入れてみて、データ項目名やデータ型が合致しているかどうか確認する。〔→作業例④〕
- ☑ IMI コア語彙や既存のデータ項目定義では表現できない項目があれば、そのデータ項目や利用目的を書き出しておく。これらのデータ項目は、2.4 で説明する「応用語彙」としてあらたに定義する。

【開始時に用意する情報】

- ・ 2.2 で作成した公開対象とするデータ項目一覧
- ・ 公開予定の実データ

【得られる成果物】

- ・ 検証済データセット
- ・ 既存語彙で表現できないデータ項目の一覧(→2.4 へ)

● このステップでの検討・定義の対象

- ・ 項目名（ラベル）
- ・ 記載内容（記載例）
- ・ データ型
- ・ 必須／推奨の種別
- ・ IMI コア語彙との対応

以下はデータ項目の検証についての作業例です。このようにして、既存の実データとの整合性を確認し、必要に応じて再調整します。

作業例④ データ項目を検証する

2.2 で作成した各データ項目を列見出し(タイトル行)として、1 行 1 レコードになるように実データを入れる。データ項目名やデータ型が合致しているかどうか確認する。

図 14 の場合、AP 列の「申込み・問合せ先」は、「担当部署」「問合せ先_住所_都道府県」などと個別データ項目に分割した方が機械判読性は高い。ただし、既存データの修正にかかる負担が大きい場合など、「まず公開する」ことを優先するという選択肢も考えられる。

	A	B	D	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	
		主催者	タイトル	開始時間	終了時間	スケジュール備考	就学前	小学生	小学生	小学生	中学生	高校生	高校生以上	対象者備考	募集人数	事前登録	申し込み方法	申込締切	申込み・問合せ先	料金種別	金額
1	CAS001	内閣官庁	遊んで、学んで、君もITマスター！					1	1	1	1	1			1	不要			担当部署:内閣官庁情報通信技術07総合戦略室 郵便番号:100-0013 住所:東京都千代田区豊が岡3-3-1尚友会館2階 直通電話:03-5253-2111 web: https://www.facebook.com/itkodomoday	無料	
2	CAO001	内閣府	アニメ「めくみ」上映と拉致問題管見パネル展示								1	1	1			不要			担当部署:内閣府大臣官庁政策評価広報課 郵便番号:100-8914 住所:東京都千代田区永田町1-6-1中央合同庁舎第8号館 代表電話:03-5253-2111 直通電話: 03-6257-1297(広報課代表) 03-6257-1383(報道室代表) web: http://www.cao.go.jp/	無料	
3	CAO002	内閣府	アニメ「めくみ」上映会と北朝鮮の工作船見学								1	1	1		15名以内(保護者等も含め30名以内)	必要	「往復はがき(7月19日必着)」	7月19日(必着)	担当部署:内閣府大臣官庁政策評価広報課 郵便番号:100-8914 住所:東京都千代田区永田町1-6-1中央合同庁舎第8号館 代表電話:03-5253-2111 直通電話: 03-6257-1297(広報課代表) 03-6257-1383(報道室代表) web: http://www.cao.go.jp/	無料	
4	CAO003	内閣府	「北方領土問題を考えよう！」				1	1	1	1	1	1				不要			担当部署:内閣府大臣官庁政策評価広報課 郵便番号:100-8914 住所:東京都千代田区永田町1-6-1中央合同庁舎第8号館 代表電話:03-5253-2111 直通電話: 03-6257-1297(広報課代表) 03-6257-1383(報道室代表)	無料	

図 14 データ項目の検証

👉 作業のポイント

- ➡ 実データを入れてみた際に、適切な記載方法が判断しづらい場合や、入力すべきデータ項目数がデータ作成者に過度な負荷になりそうな場合などは、2.2 に戻りデータ項目定義を再検討する。

2.4 応用語彙の定義

● 作業の内容

- ☑ 2.2、2.3 で発見された、IMI コア語彙や既存のデータセット定義では表現できない項目を、応用語彙として定義する。
- ☑ 応用語彙を定義する場合も、IMI コア語彙の用語を使用する場合と同様、項目名（ラベル）、記載内容、データ型、データ構造、必須／任意の種別などを定義する。〔→作業例⑤〕

【開始時に用意する情報】

- ・ 2.2 で作成した IMI コア語彙で表現できない項目の一覧

【得られる成果物】

- ・ 応用語彙の定義

● このステップでの検討・定義の対象

- ・ 応用語彙

ここでは、「応用語彙」について触れます。

「コア語彙」は、広い範囲で共通に使われることを目的として設計された語彙ですが、実際のデータセットへの適用にあたっては、コア語彙に定義された用語だけでは表現しきれない場合があります。

そのような場合、コア語彙で定義されているクラス用語に足りない用語を追加し、新しいクラスを作る作業を行います。このように、現場の必要に応じて、現場ごとに新たに作られた語彙を「応用語彙」と呼んでいます⁸。

作業例⑤で、応用語彙が作られた例を紹介します。現場で作られた応用語彙は、作成者が

⁸ 応用語彙の位置づけなどについて、さらに詳しくは「IMI 共通語彙基盤入門」
https://imi.go.jp/doc/IMI_introduction_20171227.pdf をご参照ください。

Web サイト等で公開するほか、データモデル記述（DMD）へ同梱して配布し、利用者やアプリ作成者等へ提供すると、データ構造について共通の理解をしやすくなります（3.3 を参照）。

作業例⑤ 北海道・森町と埼玉県における独自の応用語彙

北海道・森町では、IMI コア語彙とは別に地域のニーズにあわせて「避難施設型」「植物型」「作物型」「年度型」「人口型」といった独自語彙を策定し、公開データに利用している（図 15）。

森町避難場所一覧

使用語彙

- 共通語彙基盤 コア語彙
- 森町独自語彙

項目名	説明
ic:事物型 場所型 名称	避難場所の名称
ic:事物型 場所型 住所	避難場所の住所（番地等含む）
ic:事物型 コードリスト型 URI	避難場所の住所(GeoNames.jp)
ic:事物型 場所型 地理座標 緯度	避難場所の緯度
ic:事物型 場所型 地理座標 経度	避難場所の経度
ic:事物型 場所型 地理座標 測地高度	避難場所の海拔
ic:事物型 電話番号	避難場所の電話番号
ic:事物型 建物型 敷地面積	避難場所の敷地面積
ic:事物型 施設型 収容人数	避難場所の収容人数
moridm:避難施設型 最大収容人数	避難場所の最大収容人数
moridm:避難施設型 避難所種別	一時避難場所又は広域避難場所の種別
moridm:避難施設型 津波避難	津波時の避難場所利用可否

「ic:」で始まる項目名は IMI コア語彙。応用語彙には森町独自の接頭辞「moridm:」を設定している。
<http://www.town.hokkaido-mori.lg.jp/docs/2015090800018/>

図 15 応用語彙の例（その1）

埼玉県が公開している「ご当地キャラクター」のデータセットは、IMI コア語彙に対応する用語がない「モチーフ」や「性格」といったデータ項目について、独自の応用語彙を作成している（図 16）

また、「身長」「体重」「誕生日」などのデータは、IMI コア語彙の「人型」に含まれるが、2.3 データ項目の検証を行ったところ、身長が「ゆず5個分」、体重が「深谷ねぎ480本分」、誕生日が「古墳時代」などと記述されていることがわかった。IMI コア語彙の「人型」の用語は「身長」「体重」を数量で、「誕生日」を年月日で記載する想定であり、これら実デー

タの意味を適切に表現できないため、「ご当地キャラクター」用に特化した応用語彙をあらたに作成した。

No	データ項目	必須	共通語彙	データ形式	データ例	備考
1	識別情報	○		テキスト (改行なし)	1122345	キー情報。一意の値を設定する。
2	団体コード	○	団体_ID	テキスト (改行なし)	〇〇市	
3	団体名	○	団体_名称	テキスト (改行なし)	コバトン	
4	名称	○	名称_表記	テキスト (改行なし)	シラコバト	
5	モチーフ			テキスト (改行なし)	5/8 31日	20字~100字程度の文章とすること。
6	性別			テキスト (改行なし)	http://www.pref.saitama.lg.jp/a0301/kobaton/images/370370.jpg	
7	身長			テキスト (改行なし)		
8	体重			テキスト (改行なし)		
9	性格			テキスト (改行なし)		
10	趣味			テキスト (改行なし)		
11	誕生日			テキスト (改行なし)		
12	画像	○	画像_画像	テキスト (改行なし)		
13	画像_ライセンス	○		テキスト (選択)	CC BY	画像に対するライセンス表記をクリエイティブコモンズライセンスの表記を

▼データレコードの確認例▼

モチーフ	身長	体重	誕生日	性格	趣味
ゆず	ゆず5個分	ゆず5個分	康平6年(1063年)1月3日	たくさんの人に支えられ	ゆず狩り
ウサギのようでシカの深谷ねぎ5本の	体の大きさを覚えられ(大きくふたり、ちいさくなつたりできる)。	深谷ねぎ480本分	6月28日	やさしい、まけずぎらい	つぶやき
戸田の原っぱに暮らす	不明	不明	不明	不明	戸田ヶ原
男子猿轡	古墳時代(5世紀前半)	古墳時代(5世紀前半)	11月3日(滑川町ででき	男	遠跡め
女子猿轡	古墳時代(5世紀前半)	古墳時代(5世紀前半)	11月3日(滑川町ででき	女	遠跡め
町の魚「ミヤコタナゴ」	ミヤコタナゴと同じ大	ないしょ	11月3日(滑川町ででき	明るく元気いっぱい	子ども

```

PREFIX ic: <http://imi.go.jp/ns/core/rdf#>
PREFIX pd3110:
<http://imi.go.jp/ns/pd/3110/1_0_0#>
PREFIX pd3119:
<http://imi.go.jp/ns/pd/3119/1_0_0#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX schema: <http://schema.org/>

select
distinct ?dantai_code ?dantaimei ?shikibetsujoho
?meisho ?motif ?gaiyo
?shincho ?taiju ?seikaku ?shumi ?tanjobi
?gazo ?gazo_license ?gazo_text
?tantoka ?tantoka_denwabango ?tantoka_email
?jiyukijutsuran ?url

where {

graph<http://data.pref.saitama.lg.jp/graph/10_yuru
chara>
{ ?KEY pd3110:団体コード ?dantai_code;
pd3110:団体名 ?dantaimei;
pd3110:識別情報 ?shikibetsujoho;
ic:名称/ic:表記 ?meisho;
ic:概要 ?gaiyo;
pd3119:モチーフ ?motif;
pd3119:身長 ?shincho;
pd3119:体重 ?taiju;
pd3119:誕生日 ?tanjobi;
pd3119:性格 ?seikaku;
pd3119:趣味 ?shumi;
schema:image/schema:contentUrl ?gazo;
...

```

対応語彙が コア語彙の「人>身長」や「人>体重」、対応語彙がない例
 ない例 「人>誕生日」では表現しづらい例

「pd3110:」「pd3119:」で始まる項目名は、埼玉県オープンデータ共通フォーマットで独自に定義された応用語彙。 https://opendata.pref.saitama.lg.jp/fs/9/6/_/gotoutikyarajyoho.pdf

図 16 応用語彙の例(その2)

作業のポイント

- ➔ 語彙(用語)の名称を決める前に、そのデータ項目に含まれる実データ(値)の内容を確認し、実データの意味に合ったデータ項目名(ラベル)を検討することが望ましい。
- ➔ 実データの性質によっては、IMI コア語彙にすでに同じ用語があっても、応用語彙を設定した方がよい場合がある。

2.5 コード、ID の整備

● 作業の内容

- ☑ データセット内の各レコードを識別する、他のデータセットとの関連性を記述する、複数のデータ間で共有する値をまとめて定義しておくといった目的のために、コードや ID を定義する。
- ☑ 部署コードや職員 ID など、既存の識別子とデータ項目の名称を対応させたリストを整備する。これらのコードや ID を整備した経緯や目的を記録しておく。
- ☑ 「全国地方公共団体コード」や「統計に用いる標準地域コード」のような既存のコードリスト、ID 体系を参照して、可能な限り互換性を保つことが望ましい。利用できるコードや ID がない場合は、必要に応じあらたにコードや ID を整備する。

【開始時に用意する情報】

- ・ 部署コード一覧
- ・ 施設種別コード一覧
- ・ 職員 ID 一覧 等

【得られる成果物】

- ・ コードや ID 体系の定義
- ・ コードや ID の一覧

【外部参照資料・ツール等】

- ・ 全国地方公共団体コード <http://www.soumu.go.jp/denshijiti/code.html>
- ・ 統計に用いる標準地域コード
http://www.soumu.go.jp/toukei_toukatsu/index/seido/9-5.htm

● このステップでの検討・定義の対象

- ・ コードリストの名称
- ・ ID 体系の名称
- ・ コード、ID の発行者
- ・ コード、ID の説明
- ・ コード、ID の発行日時

データの中には、区分、や種別などを文字列による自然言語で表現する代わりに、数字など、統制を取った符号（コード）によって表現することがあります。そのような場合、コードの定義を、別途明確にし、公開する必要があります⁹。

コードが使われたデータ（表 4）と、そのコードの定義の例（表 5）を示します。

表 4 コードの使用例

一連番号	法人番号	処理区分	訂正区分	更新年月日	変更年月日	法人種別	国内所在地	国内所在地
1	1000013050320	1	1	2017/4/17	2015/10/5	101	宮崎県	都城市
2	1000020453820	1	1	2017/4/17	2015/10/5	201	宮崎県	東諸県郡国
3	1010001165096	1	0	2015/11/27	2015/10/5	301	宮崎県	児湯郡都農
4	1010001167927	12	0	2016/5/10	2016/5/2	301	宮崎県	宮崎市

国税庁「法人番号公表サイト」で提供されている法人の基本 3 情報データセットでは、「処理区分」や「訂正区分」、「法人種別」などをコードで提供している。
<http://www.houjin-bangou.nta.go.jp/download/index.html>

⁹ 「コード」について、詳しくは「ID 体系とコードリスト」<https://imi.go.jp/goi/goi-idcode.html> をご参照ください。

表 5 コードの例

リソース定義書（ダウンロードファイル等）

項目名		データ形式等		項目値		提供項目				I
名称	リソース名	形式 (凡例参照)	桁数	コード	項目値 名称	ダウン ロード	Web-API			
							ver1	ver2	ver3	
一連番号	sequenceNumber	ZZZZZZZ9	1～8	-	-	○	○	○	○	月次データ(月末時点の最新情報)や の取得以外の場合(例えば、ある法人 次の更新情報を取得する場合。)には ある。 その際に、法人(法人番号)単位に最 小さい値の情報から順番に処理するこ とを行うことを可能にするために設けた項 目。 なお、ダウンロードでは、1ファイルの ものが分割して提供され、Web-APIではデ ータが分割して提供されるが、いずれの場合も、
法人番号	corporateNumber	9999999999999	13	-	-	○	○	○	○	法人番号の指定を受けた者(以下「法人 番号項目」。
処理区分	process	99	2	-	-	○	○	○	○	法人番号の指定、商号又は所在地に 英語表記又はフリガナの登録に係る 語表記又はフリガナの登録情報を追加
				01	新規					値「01」は、新たに法人番号が公表さ
				11	商号又は 名称の変更					値「11」は、商号又は名称が変更され
				12	国内所在地 の変更					値「12」は、国内における本店又は主 たる事務所がない法人等の場合は、 更又は、市区町村合併等により住所が 変更される。
				13	国外所在地 の変更					値「13」は、国外における本店又は主 たる事務所があることを意味する。
				21	登記記録 の閉鎖等					値「21」は、清算の結了、合併による 消滅が生じたことを表す情報であることを
				22	登記記録 の復活等					値「22」は、法人等が清算の結了、合 併等が生じた後、清算結了等の手続さ れたことにより登記記録が復活等したことを
				71	吸収合併					値「71」は、設立登記法人が、吸収合 併される。
処理区分	process	99	2	72	吸収合併無 効	○	○	○	○	値「72」は、吸収合併後存続した設立 したことを表す情報であることを意味す
				81	商号の登記 の抹消					値「81」は、設立登記法人について商 号の抹消を意味する。
										値「99」は、法人番号の重複指定等0

表 4 のデータとともに公開されているコードの定義資料。「処理区分」や「訂正区分」、「法人種別」に記載された「1」「12」などのコードの意味を説明している。

<http://www.houjin-bangou.nta.go.jp/download/index.html#resource>

👉 作業のポイント

- ➡ コードや ID は、識別値の集合として体系的に整理できる。整理されたコードや ID を組織の内外で共有、参照できるようにしておくと、コードや ID を活用するデータの相互運用性が向上する。

2.6 識別子の定義

● 作業の内容

- ☑ データセット内の各レコードを識別するために、識別子を付与する。識別子がただしく管理付与されたレコードであれば、内外から参照することが容易になる。逆に、識別子がないと、似たレコード同士を識別することが困難になる場合がある。
- ☑ 施設番号や固有の名称など、既存の識別子がある場合は活用する。識別子がない場合は、将来にわたって変わることのない一意の番号や文字列を検討し、識別子のルールとして定義する。

【開始時に用意する情報】

- ・ レコードごとの識別子（ある場合）

【得られる成果物】

- ・ 識別子の定義

【外部参照資料・ツール等】

- ・ 施設番号、制度番号など

● このステップでの検討・定義の対象

- ・ 識別子の名称、説明、記述方法
- ・ 記述方法に沿ったレコードの識別子

データの管理を効率的に行うために、データの各レコードに、ユニークな識別子(id)を付けておくと便利です。そのような識別子を付けた例を次（**図 17**）に示します。適切に識別子を管理することで、名称の変更や、対象物の廃止、統合などの際のデータの管理が容易になります。

2017年の施設一覧

id番号	公共施設の名称(name)
YS000000002	南区役所
YS000000003	南福祉事務所
YS000000004	南保健センター
YS000000005	芝出張所
YS000000006	小宮出張所
YS000000007	霧土出張所
YS000000008	南水区役所
YS000000009	南水福祉事務所
YS000000010	南水保健センター
YS000000011	左水区役所
YS000000012	左水福祉事務所
YS000000013	左水保健センター
YS000000015	清川出張所

2018年の施設一覧

id番号	公共施設の名称(name)
YS000000002	南区役所
YS000000003	南保健福祉センター（健康）
YS000000004	南保健福祉センター（子ども）
YS000000005	芝出張所
YS000000006	小宮出張所
YS000000007	霧土出張所
YS000000008	南水区役所
YS000000009	南水保健福祉センター
YS000000011	左水区役所
YS000000012	左水保健福祉センター
YS000000015	清川出張所
YS0000A0050	堀出張所

それぞれのデータセット内でユニークな識別子を設定する。複数の期間や地域などのデータセットを組み合わせる際にもユニークな識別子が必要である。

施設の例では新設、閉鎖、統合、名称変更といった値の変更が起きるが、いちど使った識別子（閉鎖施設の例）は廃番とし、新たにユニークな識別子（新設の例）を設定する。統合の場合は、ひとつの施設として識別子もひとつに集約する方法も、例のように施設の機能が異なるなどの理由で名称のみ変更する方法もある。利用目的に鑑みて値の意味が変わらないように運用を検討する。

図 17 識別子の例

作業のポイント

- ➔ 識別子はデータセット内でのユニーク性が担保されている必要がある。また、他のデータセットのデータと一緒に扱う場面がある場合には、全体の中でもユニーク性が担保されている必要がある。
- ➔ 規則性のある単純な番号を連続して識別子として使う場合もあれば、名称にユニーク性がある場合（例：明治、大正、昭和など）はその文字列を識別子とする場合もある。また、ウェブのようなグローバル空間でデータを共有するには、URLを採用する場合もある。

2.7 データ項目定義書の作成

● 作業の内容

- ☑ データセットごとに、2.1～2.6 で検討・定義してきた IMI コア語彙/応用語彙の種別、データ項目名、値の範囲、書式、使用するコード・ID 体系、記載例を整理した「データ項目定義書」を作成する。
- ☑ 「データ項目定義書」に従い、データ作成者が入力や整形にもちいるための「データ入力フォーマット」を作成する。

【開始時に用意する情報】

- ・ 2.6 までに作成した各データ項目定義

【得られる成果物】

- ・ データ項目定義書
- ・ データ入力フォーマット

ひとつおりデータの設計が終わったら、それを「データ項目定義書」としてまとめます。データの各項目はどういう意味を持つのか、そこにはどのような様式でデータを入れるのかといった説明をまとめます。図 18 に、データ定義書の例を示します。

2.6 識別子の定義

2.2 IMIコア語彙との対応

2.2 データ型(形式)を定義

2.2 データの階層関係(データ構造)を定義

2.2 データ項目

2.5 コードやID

No	テーマ	データ名称	公共施設情報	データ項目	必須	共通項目	データ形式	データ例	備考
1		施設情報					テキスト(改行なし)		キー情報。一意の数字を設定する。
2		団体コード		団体_ID			テキスト(改行なし)	112345	
3		団体名		団体_名称			テキスト(改行なし)	〇〇市	
4		種別		施設_種別			テキスト(選択)	〇〇市	
5		名称		名称_表記			テキスト(改行なし)	〇〇庁舎	20字~100字程度の文章とすること。
6		住所		住所_表記			テキスト(改行なし)	埼玉県さいたま市浦和区高砂3丁目5-1	都道府県名から所在地まで区切りなく記入する。丁目以下は半角数字で記載し、ハイフンで接続。
7		施設_経度		経度_経度			経度_経度	35.85728252	
8		施設_経度		経度_経度			経度_経度	139.6477938	
9		施設_経度		経度_経度			経度_経度	000-000-0000	
10		ホームページ		URL_表記			URL_表記	http://www.stadium2002.co.jp/	
11		開始時間		開始時間			hh:mm:ss	08:00:00	※2桁表記0(ディンブ) ※空欄の場合はハイフン「-」
12		終了時間		終了時間			hh:mm:ss	17:00:00	※2桁表記0(ディンブ) ※空欄の場合はハイフン「-」
13		日時備考		日時備考			テキスト(改行あり)	日曜は閉庁	
14		自由記述欄		自由記述欄			自由記述欄		

図中の番号(2.2等)は、本手引きの章番号と対応しています。

図 18 データ項目定義書の例

データ項目定義書ができれば、その項目名を見出し行に並べることで、容易にデータ入力用のフォームを作成することができます(図 19)。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	識別情報	団体コード	団体名	種別	名称	概要	通称	住所_表記	施設_緯度	施設_経度	施設_電話(ホームページ)	開始時間	終了時間	日時備考	自由記述欄			
2		112119	本庄市	地域機関	児玉中央小会議室	美術工芸室	本庄市児玉	36.18845	139.1156	0495-72-8851		9:00	21:30	休館日:12月28日から翌年の1月4日				
3		112119	本庄市	地域機関	本庄公民館集会室	小会議室	和	本庄市東	36.24308	139.1977	0495-24-7383		9:00	21:30	休館日:12月28日から翌年の1月4日			
4		112119	本庄市	地域機関	本庄東公民館集会室	会議室	和	本庄市日	36.23306	139.2002	0495-22-3404		9:00	21:30	休館日:12月28日から翌年の1月4日			
5		112119	本庄市	地域機関	本庄西公民館集会室	会議室	和	本庄市小	36.2454	139.1715	0495-21-8786		9:00	21:30	休館日:12月28日から翌年の1月4日			
6		112119	本庄市	地域機関	本庄南公民館集会室	会議室	和	本庄市今	36.22682	139.1643	0495-21-8785		9:00	21:30	休館日:12月28日から翌年の1月4日			
7		112119	本庄市	地域機関	藤田公民館集会室	会議室	和	本庄市牧	36.23609	139.2267	0495-22-2332		9:00	21:30	休館日:12月28日から翌年の1月4日			
8		112119	本庄市	地域機関	仁手公民館集会室	会議室	和	本庄市仁	36.24805	139.2095	0495-22-3538		9:00	21:30	休館日:12月28日から翌年の1月4日			
9		112119	本庄市	地域機関	旭公民館集会室	和室、図書室	本庄市都	36.25993	139.1742	0495-22-5940		9:00	21:30	休館日:12月28日から翌年の1月4日				
10		112119	本庄市	地域機関	北泉公民館集会室	会議室	和	本庄市早	36.21709	139.1882	0495-22-2484		9:00	21:30	休館日:12月28日から翌年の1月4日			
11		112119	本庄市	地域機関	共和公民館集会室	会議室	和	本庄市見	36.20932	139.1514	0495-72-0337		9:00	21:30	休館日:12月28日から翌年の1月4日			

図 19 データ入力用フォームの例

作業のポイント

- ➔ データ入力フォーマットは、データの保有者が入力しやすいように、データ項目ごとに説明や入力例を記載することが望ましい。特に表記ゆれが発生しやすい日付や住所、電話番号等の記載方法は丁寧に説明する(図 18)。

■STEP3. 目的に合った記述方法を決めて実装する

- ▶ このステップでは公開するデータセットの実装を行います。
- ▶ 実装にあたっては公開の目的にあわせてファイル形式を決定し、必要な場合はデータ変換を行います。
- ▶ データセットの用途によっては、DMD（データモデル記述）を作成します。

3.1 公開ファイル形式や公開場所などの決定

● 作業の内容

- ☑ データセットごとに、公開目的に合致するファイル形式（*.csv、*.xlsx、*.xls、*.ttl、*.xml など）を決める。
- ☑ 公開対象とするデータセットを一覧表のかたちに整理する（データカタログ）。公開データセット及び公開フォーマットの流通や再利用を促すため、データセットと併せてデータカタログを公開することが望ましい（図 20）。

【開始時に用意する情報】

- ・ ■STEP1 でデータセットごとに定義したデータの公開目的

【得られる成果物】

- ・ データセットごとの公開ファイル形式
- ・ データカタログ

公開するデータセットの構造が決まったら、データセットを具体的にどのような形式で公開するかを決めます。例えば表形式（.csv、.xlsx 等）、構造化データ(.ttl、.xml)など、目的に応じて形式を決めます。

また、どこから（Web サイトの URL など）公開するかを決めます。

この決定に従い、公開データを整理したデータカタログを作成します。データカタログにどのような情報をまとめるかの例を図 20 に示します。一般的に、データセットを説明した一覧表と、各データの項目名を説明したデータ項目定義書が、カタログに含まれます。

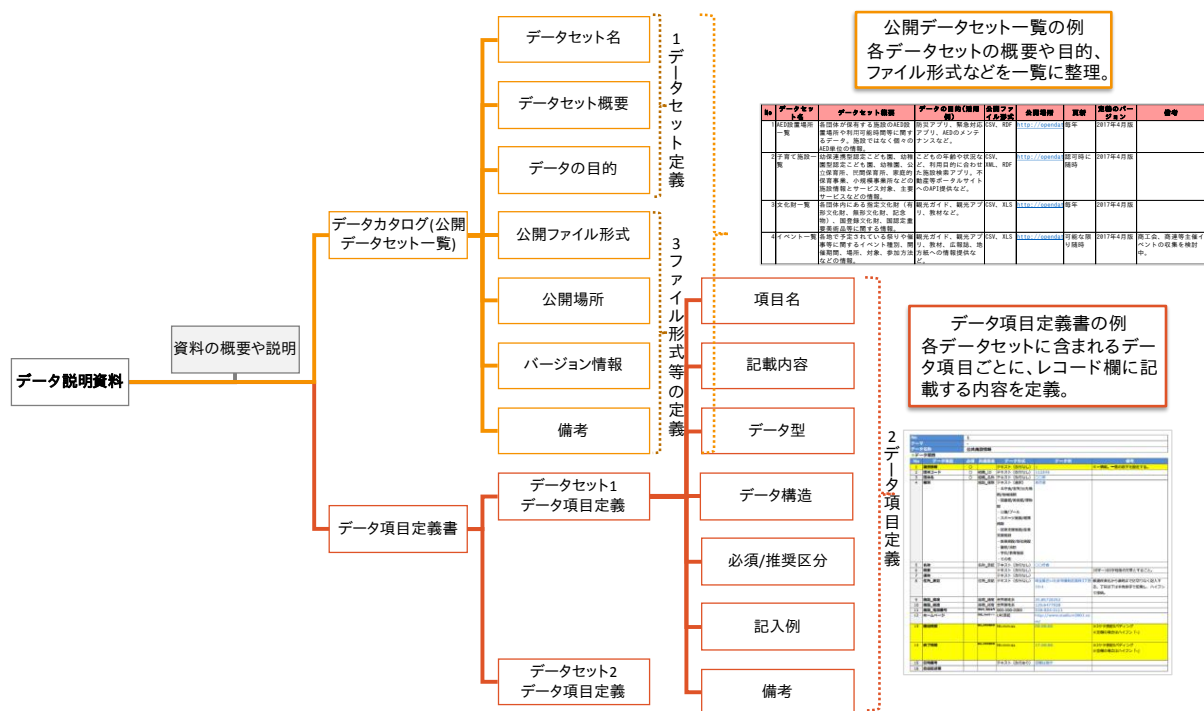


図 20 データカタログとデータ項目定義書

表 6 に、データセット一覧表の例を示します。

表 6 公開データセット一覧の例

No	データセット名	データセット概要	データの目的(活用例)	公開ファイル形式	公開場所	更新	定義のバージョン	備考
1	AED設置場所一覧	各団体が保有する施設のAED設置場所や利用可能時間等に関するデータ。施設ではなく個々のAED単位の情報。	防災アプリ、緊急対応アプリ、AEDのメンテナンスなど。	CSV、RDF	http://opendata.city.kyoto.jp/	毎年	2017年4月版	
2	子育て施設一覧	幼保連携型認定こども園、幼稚園型認定こども園、幼稚園、公立保育所、民間保育所、家庭的保育事業、小規模事業所などの施設情報とサービス対象、主要サービスなどの情報。	こどもの年齢や状況など、利用目的に合わせた施設検索アプリ。不動産等ポータルサイトへのAPI提供など。	CSV、XML、RDF	http://opendata.city.kyoto.jp/	認可時に随時	2017年4月版	
3	文化財一覧	各団体内にある指定文化財(有形文化財、無形文化財、記念物)、国登録文化財、国認定重要美術品等に関する情報。	観光ガイド、観光アプリ、教材など。	CSV、XLS	http://opendata.city.kyoto.jp/	毎年	2017年4月版	
4	イベント一覧	各地で予定されている祭りや催事等に関するイベント種別、開催期間、場所、対象、参加方法などの情報。	観光ガイド、観光アプリ、教材、広報誌、地方紙への情報提供など。	CSV、XLS	http://opendata.city.kyoto.jp/	可能な限り随時	2017年4月版	商工会、商連等主催イベントの収集を検討中。

図 21 に、公開するデータセットのファイル形式の記載例を示します。

【和光市】すべての公共施設（市役所・出張所・出先機関・その他）

和光市のすべての公共施設（市役所・出張所・出先機関・その他）

CSV

XLS

TXT

【朝霞市】AED設置場所情報

市内公共施設等のAED設置場所

CSV

XLSX

【深谷市】AED設置場所情報

深谷市が保有する施設等のAED設置場所や利用可能時間等に関する情報です。（次のURLより地図を見ながらダウンロードも可能です。<https://fukaya.geocloud.jp/webgis/?p=0&bt=0&mp=12>）

CSV

KML

SHP

csv、xls、xlsx、txt、kml、shp など、用途に応じて公開データのファイル形式を複数用意する場合もある。

図 21 埼玉県オープンデータにおける公開データのファイル形式の例

👉 作業のポイント

- ➡ 公開データのファイル形式は用途に応じて適切なものを用意する（表 6、図 21）。
- ➡ データセットの大分類や目的、ファイル形式、公開日、更新頻度などを記述して公開しておく、ユーザーがデータセットを選びやすい（表 6）。

3.2 データレコードの入力・追加・削除

● 作業の内容

- ☑ 2.7 で作成した「データ項目定義書」に従い、データ入力フォーマットに公開するデータのレコード（値）を入力する。もとにする実データがある場合は、データ入力フォーマットに合わせて整形する。公開対象としないデータは削除する。
- ☑ あらたに識別子を定義した場合は、2.6 で定義した記述方法に沿い、各データレコードに識別子を付与する。

【開始時に用意する情報】

- ・ データ項目定義書
- ・ データ入力フォーマット
- ・ 入力対象データ（既存データセットなど）

【得られる成果物】

- ・ 公開データセット一式(Excel、csv 等)

データ項目定義書に従い、公開するデータを作成します。さらに RDF 形式などを作成する手順については、3.3 をご参照ください。

👉 作業のポイント

- ➡ 全角文字と半角文字の違いや、空白文字や区切り記号の有無等の整理（データクレンジング）は、データセットとして全てのレコードを入力したあとに作業する方が、効率的な場合が多い。
- ➡ 公開の初期段階では、各レコードの完全性よりも、データレコードの数の充実を優先することが望ましい。

3.3 データの変換

● 作業の内容

- ☑ アプリでの利用など、データ活用先が決まっている場合は、3.2 で作成した Excel データ (*.csv、*.xlsx、*.xls)などを、XML や RDF などの機械判読可能な形式に変換する。
- ☑ 必要に応じてデータモデル記述 (Data Model Description) ¹⁰を作成しておく、アプリ開発者やコンピューターがデータセットの構造を理解しやすくなる。

【開始時に用意する情報】

- ・ 3.2 で作成した公開データセット
- ・ データ項目定義書

【得られる成果物】

- ・ 公開データセット一式(XML、RDF など)

● このステップでの検討・定義対象

- ・ データモデル記述(DMD)

IMI 共通語彙基盤では、データの構造や、各項目へのデータの入力方法などを機械可読な形式にまとめた DMD (データモデル記述) の様式を規定しています。

作業のポイント

- ➔ Excel 等の表形式データで問題はなく、まずはデータセットの公開を優先する。必要に応じて、より機械判読性の高い形式での公開を検討する。
- ➔ ■STEP2 でデータ項目を適切に整理し、3.2 でデータ項目定義に合致したデータを作成してあれば、データ変換の作業は外部委託が容易になることが多い。

¹⁰ DMD については「IMI 共通語彙基盤入門」[「https://imi.go.jp/doc/IMI_introduction_20171227.pdf](https://imi.go.jp/doc/IMI_introduction_20171227.pdf) をご参照ください。

■STEP4. データを公開・運用する

- ▶ このステップでは、共通語彙に対応した項目の定義と整理が完了したデータセット一式の公開を行います。
- ▶ 公開後はユーザーの意見・要望等を積極的に受け付け、より良いものに改善していきます。

4.1 データの公開

● 作業の内容

- ☑ オープンデータサイト等に、公開データセット一式を公開する。

【開始時に用意する情報】

- ・ 公開データセット一式

【得られる成果物】

- ・ オープンデータサイト等でのデータセット公開

データセット公開用のサイトへ、■STEP3 までで構築したデータを配置し、公開します。

👉 作業のポイント

- ➡ データセットを RDF や XML で公開する場合、独自に作成した応用語彙については、データ公開サーバー等に名前空間（namespace）を定義したファイルを格納する。¹¹

¹¹ IMI における応用語彙の作成方法に関しては、支援ツールやドキュメントなどを公開する予定です。

4.2 データ公開の告知

● 作業の内容

- ☑ データを公開したことを公式ホームページ等で告知する。

【開始時に用意する情報】

- ・ オープンデータサイト等

【得られる成果物】

- ・ オープンデータサイト、公式ホームページなどでの告知

データセット公開の告知例を図 22 に示します。



公式ページでも告知を行った例。 <https://www.pref.saitama.lg.jp/a0001/news/page/170119-01.html>

図 22 データ公開告知の例

埼玉県ではデータセット公開の際、共通フォーマット策定の経緯やデータ項目定義書、検討プロセスにおける参加メンバーなども含めて告知を行いました（図 23）。



埼玉県は、共通フォーマット作成の経緯やデータの活用事例なども含めて情報公開を行った。
<https://opendata.pref.saitama.lg.jp/events/news20170119.html>

図 23 データ公開の告知記事例

4.3 公開したデータの運用

● 作業の内容

- ☑ 1.1 で定義したデータセットの更新頻度に従い、データセットを更新または追加していく。
- ☑ レコードの追加、削除、変更等が生じた場合は、可能な限り該当データセットのメンテナンスを行う。

【開始時に用意する情報】

- ・ 更新レコード
- ・ ユーザーからのコメント等

【得られる成果物】

- ・ 更新されたデータセットや更新履歴

データを公開した後は、必要に応じ、更新等のメンテナンスを行います。

図 24 の例（京都市）では、最新のデータセットとともに、その更新履歴も併せて公開しています。

2016年9月30日 v1.0.0

No	データ項目	必須	利用用途	データ形式	データ例	備考
1	施設コード	○	組織_ID	テキスト (改行なし)	112345	
2	施設名	○	施設_名称	テキスト (改行なし)	〇〇〇	
3	施設_名称	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇 (保育園)	住所の名称や建物等の名称を表記。 20字〜100字程度の文章とすること。
4	施設_住所	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇	20字〜100字程度の文章とすること。
5	施設_電話番号	○	施設_電話番号	テキスト (改行なし)	99-9999-9999	
6	施設_メールアドレス	○	施設_メールアドレス	テキスト (改行なし)	abc@city.ooo.jp	
7	施設_URL	○	施設_URL	URL表記	http://www.ooo.jp/event/abc.html	
8	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	郵便番号から施設まで記す必要なく記入する。〒は半角で記し、ハイフンで接続。
9	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	郵便番号から施設まで記す必要なく記入する。〒は半角で記し、ハイフンで接続。
10	施設_電話番号	○	施設_電話番号	テキスト (改行なし)	33-85778252	
11	施設_電話番号	○	施設_電話番号	テキスト (改行なし)	133-6477738	
12	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
13	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
14	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
15	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
16	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
17	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
18	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
19	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
20	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
21	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
22	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	

2017年6月29日 v1.1.0

No	データ項目	必須	利用用途	データ形式	データ例	備考
1	施設コード	○	組織_ID	テキスト (改行なし)	112345	
2	施設名	○	施設_名称	テキスト (改行なし)	〇〇〇	
3	施設_名称	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇 (保育園)	住所の名称や建物等の名称を表記。 20字〜100字程度の文章とすること。
4	施設_住所	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇	20字〜100字程度の文章とすること。
5	施設_電話番号	○	施設_電話番号	テキスト (改行なし)	99-9999-9999	
6	施設_メールアドレス	○	施設_メールアドレス	テキスト (改行なし)	abc@city.ooo.jp	
7	施設_URL	○	施設_URL	URL表記	http://www.ooo.jp/event/abc.html	
8	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	郵便番号から施設まで記す必要なく記入する。〒は半角で記し、ハイフンで接続。
9	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	郵便番号から施設まで記す必要なく記入する。〒は半角で記し、ハイフンで接続。
10	施設_電話番号	○	施設_電話番号	テキスト (改行なし)	33-85778252	
11	施設_電話番号	○	施設_電話番号	テキスト (改行なし)	133-6477738	
12	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
13	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
14	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
15	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
16	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
17	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
18	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
19	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
20	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
21	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	
22	施設_住所_表記	○	住所_表記	テキスト (改行なし)	〇〇〇〇〇〇〇〇〇〇〇〇	

➡

👉 作業のポイント

- imi.go.jp

4.4 データ項目定義書の公開

● 作業の内容

- ☑ ユーザーがデータセットを選んだり活用したりする際にデータの意味が伝わりやすくなるよう、データ項目定義書を公開する。
- ☑ ユーザーや他団体等のコメント、利用ログ等をふまえ、相互運用性向上の可能性を検討する。

【開始時に用意する情報】

- ・ データ項目定義書
- ・ データモデル記述（DMD）

充実したデータ項目定義書を整備し、公開することで、公開したデータをより活用しやすくなります。京都市の例（**作業例⑥**）、厚木市の例（**作業例⑦**）を紹介します。

作業例⑥ 京都市オープンデータポータルサイト

京都市オープンデータポータルサイトは、各データセットに「データセットの情報」を提供。作成者、更新日、公開日、更新頻度、データセットの識別子と存在場所、ライセンス、作成基準日などがわかるようになっている（**図 26**）。

作業例⑦ 厚木市オープンデータポータルサイト

厚木市オープンデータポータルサイトでは、利用者からのオープンデータに関する要望を募っている（図 28）。



厚木市オープンデータポータルの「ご要望」ページ。随時要望を収集している例。
<http://www.city.atsugi.kanagawa.jp/shisei/15001/opendata/portalsite/d033348.html>

図 28 要望・意見の募集例

👉 作業のポイント

- ➔ 各データセットについての情報を提供すると、ユーザーが安心してデータを活用しやすい。〔→作業例⑥〕
- ➔ ユーザーのコメント等をふまえ、データ項目やデータ構造の見直しを行うことが望ましい。〔→作業例⑦〕

この文書について

「IMI 共通語彙基盤」に対応したデータ公開の手順を解説しています。

表題	IMI 共通語彙基盤に対応したデータ公開の手引き
公開日	2018 年 3 月 30 日
改定日	2018 年 5 月 31 日
作成者	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
発行者	独立行政法人情報処理推進機構(IPA) (法人番号 5010005007126)

この文書のご利用にあたって

▶ 著作権

この文書は、IPA が著作権を持ち、CC0 1.0 全世界 (<https://creativecommons.org/publicdomain/zero/1.0/legalcode.ja>) で公開します。

▶ 免責事項

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、IPA 及び IMI 検討部会は、一切の責任を負いませんのでご了承ください。

ご意見を募集しています

広くみなさまのご意見を募集しています。以下ご意見投稿のページに進み、ご記入ください。

<https://imi.go.jp/783/>