

共通語彙基盤 (IMI) 事業 進捗状況

2014 年9月

独立行政法人情報処理推進機構

IMI (Infrastructure for Multilayer Interoperability) 共通語彙基盤整備事業

IPAの事業：

- 行政の情報システムから公開される情報、行政の情報システム間で交換される情報について、
 - 意味・定義の明確化
 - 記述様式の統一化
 - 関係性の明確化
 - を図るために必要な基本情報(語彙データ)を整備し、データベースとして提供
 - あわせて、この基本情報を活用し、「オープンデータ」の作成等を効率的に行うためのツールを整備。リファレンス実装として公開
-
- 2013年度事業開始
 - 2013年度～14年度：
概念設計、パイロットシステムの構築運用、自治体を現場とした実験
 - 2015年度～17年度：
語彙データベース、ツール類(オープンデータ作成支援)を整備、公開

共通語彙基盤とは？

その概要と期待される効果

IMI

意味

•IMI: Infrastructure for Multilayer Interoperability (共通語彙基盤)

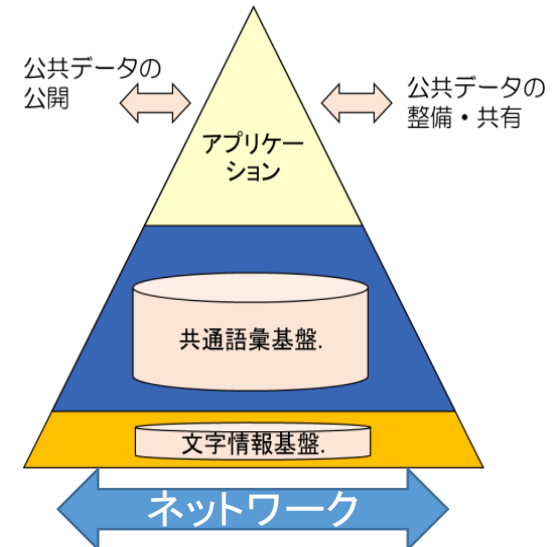
- 情報システムの「各層」(電気信号→データ(文字→単語)→アプリケーション)がきちんと繋がるための「基盤」を提供する。

そのために、情報機器が交換するデータの中で用いられる「用語」について、

表記の共通化

意味の明確化

を目指す。



人は柔軟: 放っておくと、さまざまな表現が使われる。しかし、コンピュータは混乱する。



お名前は

田中です

恵子です

田中恵子です

戸籍では
田中蕙子です

旧姓は
鈴木です

たなかです

I'm Keiko

通称は
鈴木けい子
です

たなかけいこ
です



行政現場：同じことが知りたいのに、多様な様式を作ってしまう

- 各所でシステム開発時等に「車輪の再発明」を繰り返す →業務のムダ
 - 情報の集約に支障 →業務の非効率
- （住所、氏名、組織など、業務特性に依存しないものについては）
政府内で共通的にルールを作ることが望ましい

100-0013 東京都千代田区霞が関1-1-1 経済産業ビル2階
株式会社田中商店
代表取締役 田中一郎

入力したいデータ

各種申請の入力様式

申請Aの記述様式

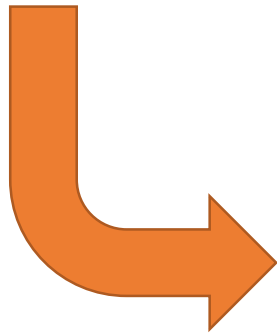
100-0013 千代田区 霞が関一丁目1-1 経済産業ビル2階
株式会社 田中商店 カブシキカイシャ タナカショウテン
田中 一郎 タナカ イチロウ

申請Bの記述様式

100-0013 千代田区 霞が関1丁目1-1 経済産業ビル2階
チヨダクカスミガセキ
株式会社田中商店 タナカショウテン
代表取締役 田中 一郎 タナカ イチロウ

申請Cの記述様式

100-0013 東京都 千代田区 霞が関 1 1 1 経済産業ビル2階
株式会社田中商店
田中 一郎 タナカ イチロウ

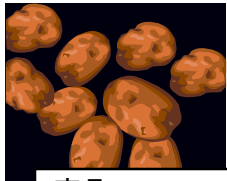


行政の申請においても、
基本的な語彙（住所、氏名等）
の記載方法が異なる

コンピュータは、頭が固い

→ データの側に工夫が必要

① 違う表現の単語を同じ単語とみなせる工夫が必要



ジャガイモ
馬鈴薯
メイクイーン
ポテト

商品
ジャガイモ
値段
50円



人であれば同じと判断できるが、コンピュータでは判断できない



同じ意味だと管理する

管理方法1: 同じ単語であると、表などで管理する

管理方法2: コンピュータで、同じ意味と自動処理する

② 同じ表現であるが意味が違うモノを明確にする必要



至急「水」の
手配を。

避難所 水
病院 水
消防 水



違う部門や状況では、同じ言葉でも、完全に同じ意味とは限らない
(例: 同じ「水」であっても業界で意味が異なる)

避難所 「水」→飲料水
病院 「水」→医療用蒸留水
消防 「水」→消火用水



違う意味だと管理する

複数のサービスやシステムを連携するときには、核となる技術である。

実際の現場では多様な用語が使われている

		自治体A	自治体A-1	自治体B	自治体B-1
避難所	仮の定義				
避難所	避難所の総称				避難施設
指定避難所	自治体があらかじめ指定した避難所	避難所	避難所	避難所	
福祉避難所	生活に解除などが必要な方が入る避難所	二次避難所			
仮設避難所	災害時に避難所に特別に解放されてできる仮の避難所				
広域応援活動拠点	広域で避難者支援を行うための活動拠点。臨時ヘリポートや備蓄倉庫等を保有。			広域応援活動拠点	地域防災拠点
避難場所	避難場所の総称				
緊急避難場所	避難場所へ避難する前に、近隣の避難者が一時的に集合して様子を見る場所又は避難者が避難のために一時的に集団を形成する場所	一時集合場所	一時避難場所		
広域避難場所	大地震時に発生する延焼火災やその他の危険から避難者の生命を保護するために必要な面積を有する大規模公園、緑地等のオープンスペースをいう。		広域避難場所	広域避難地	広域避難場所
帰宅困難者受入施設	帰宅困難者が一時的に滞在する施設	一時滞在施設、一時待機施設			帰宅困難者一時滞在施設
災害時帰宅支援ステーション	帰宅者に対してトイレの提供等の各種支援をする民間施設	災害時帰宅支援ステーション		災害時帰宅支援ステーション	災害時帰宅支援ステーション
一時避難施設	都道府県の施設などで、市町村が一時的に利用する避難施設				一時避難施設、一時避難所、一時受入施設

都道府県と市町村の語彙が違っており、情報をマッシュアップできない



- 現場の状況に基づき、ボトムアップに意味を整理し、コンピュータでも理解できる形式として整備し、基盤として公開する必要。

共通語彙基盤

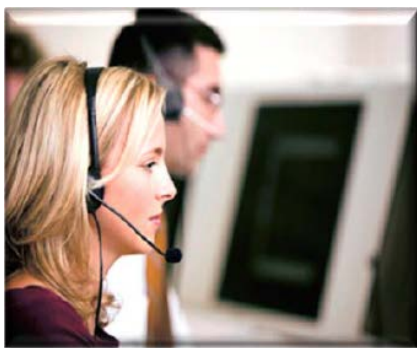
米国の活用事例

- NIEM (National Information Exchange Model)
政府が、基本的な用語について、意味や構造を整備
- NIEMで整備された、住所、氏名、事件の状況、人の特徴などに係る
統一的表記を用い、警備会社と警察との連絡をコンピュータ化。
- Washington DC, Richmondなど次々と採用が進む

警備会社

政府緊急対応機関

従来



電話による
通報



3分／通報
地図で住所を調べor手でカーナビにセットして
現場に急行

現在



NIEMの整備
した共通語彙
を用いたコン
ピュータ通信
(情報連携)



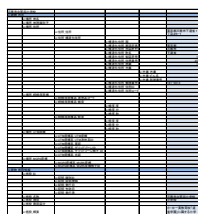
15秒／通報
さらにカーナビに
自動登録して現
場に急行

共通語彙基盤 (IMI: Infrastructure for Multi-layer Interoperability) の整備

- 情報を正しく効率的に交換、活用していくためには、人名、住所、物等、データを体系的、かつ、構造的に定義して行く必要がある。

共通語彙基盤は、用語の参照辞書を整備することで、各種データの同一性の確認を容易にし、その結果として、システム間の連携やオープンデータの活用を容易にできるようにする仕組み。

語彙(ボキャブラリ)、
情報交換パッケージ(IEP)



システム連携



情報交換パッケージにより、システム間を連携

- ・高速な情報連携
- ・設計の効率化



オープンデータ



語彙で意味を確認し、情報交換パッケージから、情報を抽出

- ・サービス設計の効率化
- ・安定した情報連携



検索



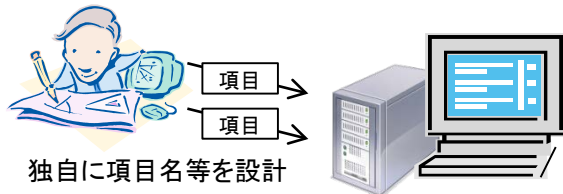
語彙間の整理をしておくことで、検索を効果的に実施

- ・検索の利便性の向上
- ・効果的な広報の実施

Schema.org
検索エンジン大手が整備する
構造化データマークアップの共通仕様

語彙基盤のない世界

「オープンデータ」の提供、申請の電子化等を推進しようとしているが。。。



独自に項目名等を設計

システム設計、情報設計が個別に行われる



- 高い構築・運用コスト
- 意味の不一致による誤解の可能性



同じ項目名でも意味が違うことがある。
 (「世帯」は、同居のことか家計のことか等)

- 様式の不一致

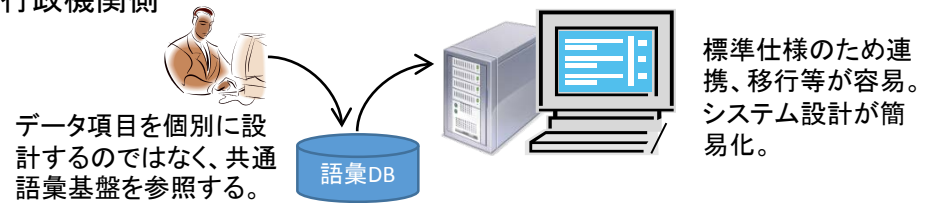
郵便番号 必須 <半角7文字以内、ハイフンは入れないでください>

郵便番号 (8文字) 123-4567 の形式で入力してください。

ECの情報科学総局は、データ様式不統一などによって行政情報へのスムーズなアクセスが行えないことにより、国民一人当たり年間100ユーロ相当の時間を無駄にしていると試算(日本全体に換算すると年間1.7兆円相当の無駄)。

語彙基盤のある世界

行政機関側



データ項目を個別に設計するのではなく、共通語彙基盤を参照する。

標準仕様のため連携、移行等が容易。システム設計が簡易化。



共通化された項目を組みあわせることで情報交換や公開するの項目の調整が容易



- オープンデータの促進、多様な行政情報の発信
- 行政コストの削減
- 電子行政サービスの質的向上
 - 申請の電子化拡大
 - データの様式の統一
 - 情報の意味定義が明確、高い機械可読性



- オープンデータを活用するアプリ、機器、サービス事業の需要拡大、新産業の発達
- オープンデータ作成にかかる新産業(データ構築、データキュレーション、ソフト作成)の需要拡大、新産業の発達

従来（オープンデータ以前）：公開情報を図示するなど、人間にとってわかりやすいよう配慮→しかしコンピュータは理解できない。

将来（オープンデータ）：コンピュータにとってわかりやすい形式で情報を公開→それを活用するスマホアプリなどへの応用が容易

「オープンデータ」形式で公開された情報(左)を活用し、近隣の公園のトイレを自動的に表示するスマホアプリの例(鯖江市)

no	name	localname
1	Nishiyama Park (Central Sq	西山公園(中央広場)
2	Nishiyama Park (Octagonal	西山公園(八角(お祭り広場北
3	Nishiyama Park (Forest Ad	西山公園(冒険の森)
4	Nishiyama Park (Nishiyar	西山公園(西山動物園)
5	Nishiyama Park (Garden ky	西山公園(郷陽庭園(中段))
6	Nishiyama Park (Garden ky	西山公園(郷陽庭園(北の庭))
7	Otani Park	大谷公園
8	Higasi Park	東公園
9	Kaminaka Park	神中公園
10	Hino River green	日野川緑地
11	South Park	南公園
12	Cyureito Park	忠霊塔

man	woman	commonuse	handicappe	babybed
6	5		1	1
12	8		1	
4	3		1	
4	2		1	1
4	2		1	

活用



公開データ(オープンデータ形式)

データを活用するスマホアプリ例

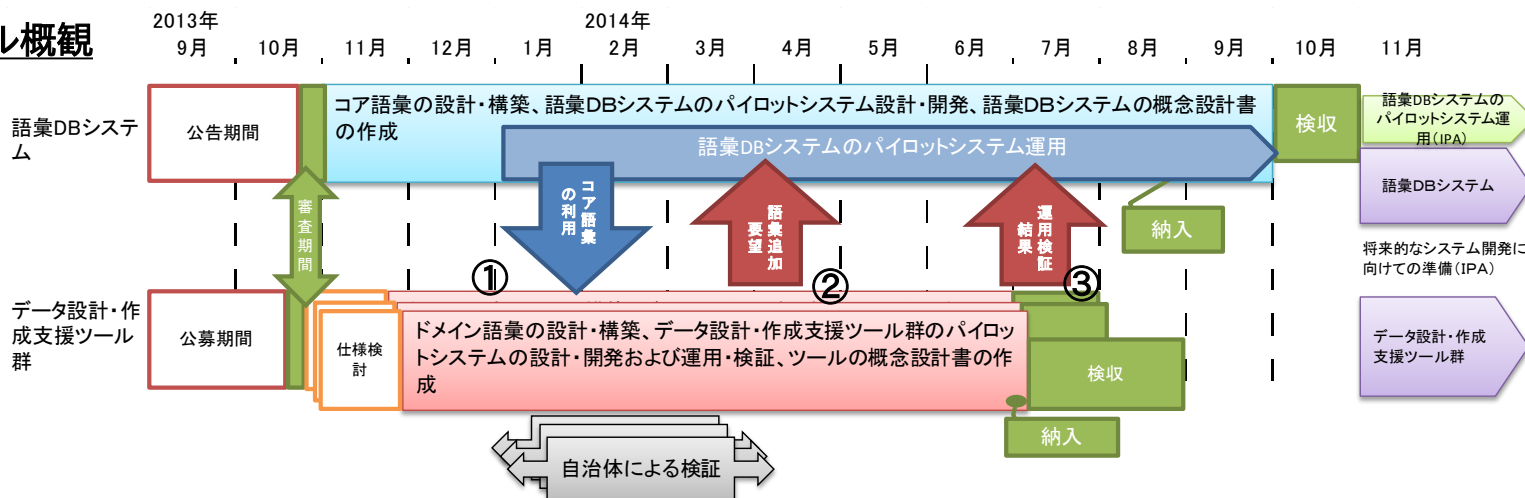
プロジェクト進捗状況

情報連携用語彙データベース事業の全体計画 とスケジュール

全体計画



スケジュール概観



- DBパイロットシステムに用意されるデータ構造、APIおよびコア語彙の仕様がDBプロジェクト担当者から、ツールプロジェクト担当者に渡される。
- 検証用の語彙データは極めて基本的な「コア語彙」以外については、ツールプロジェクト側が用意し、原則DBパイロットシステムへ登録し、運用。
- 評価はツールプロジェクト実施者が独自に行い、概念モデル中にまとめるとともに、それらをDBプロジェクト実施者がDB運用の視点から取りまとめる。

情報連携用語彙データベース事業の実施者

■ DBプロジェクト

語彙データベースのパイロットシステムを構築・運用し、データベースやツールに関する知見や課題を整理

■ ツールプロジェクト

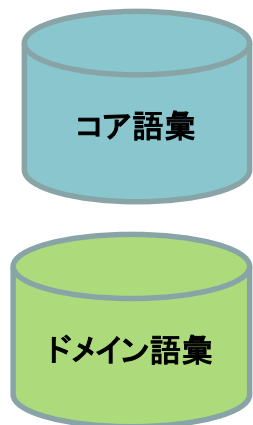
パイロットシステムを活用し、そこから提供される語彙データを活用して、再利用性の高いデータの作成に関わるツールを試作し、自治体の現場で運用する。この試作・運用を通して、語彙データ、データ構造やツールの機能について得られた知見や課題を整理

実施者と実施内容

プロジェクト	テーマ	プロジェクトの特徴	実施者	試用公的団体
DBプロジェクト			株式会社日立製作所	
ツールプロジェクト	「施設・サービス」のデータ化	Excel入力によるRDF形式でのデータ構造作成	株式会社三菱総合研究所	金沢区
	「地理空間」のデータ化	地理空間の情報化と地理情報との連携についての考察	インディゴ株式会社	浦安市都市整備部市街地開発課液状化対策推進室
	「観光」のデータ化	非熟練者でも共通語彙を使用してオープンデータの作成が可能に	株式会社ネットワーク応用通信研究所	松江市 島根大学 松江商工会議所
	「制度」のデータ化	制度情報を複数の自治体で共有化	株式会社アスコエパートナーズ	神戸市こども家庭局こども企画育成部 千葉市総務局情報経営部業務改革推進課 川口市企画財政部情報政策課
	「イベント」のデータ化	既存語彙とのマッピングについての考察	インフォコム株式会社	千葉市総務局情報経営部業務改革推進課

ツールプロジェクト実施の基本

住所、氏名、施設等、
基本的事項について、
記入すべきデータ項目、
データのタイプ
等を記述したデータ



自治体関係者
などプロジェクト
参加者



取捨選択、
新規作成、
並べ替え

データ構造設計
支援ツール



データ交換用
パッケージ(IEP
/テンプレート)

自治体関係者
などプロジェクト
参加者



具体的データ
の入れ込み

データ入力
支援ツール

語彙基盤の形式
で統一された
データ

制度、メニュー、イベント等、対
象分野に固有の事項について、
記入すべきデータ項目、
データのタイプ
等を記述したデータ

- 各プロジェクトが作成した独自の語彙データ
 - 観光情報（博物館、史跡、店舗など）の表現に係る語彙
 - 子育て助成制度の表現に係る語彙
 - イベント、施設（場所、主催者、参加者、内容、用途など）に係る語彙
 - 道路の形状、管理情報などに係る語彙等
- 各プロジェクトが試作したツール類（オープンソースライセンスで公開）
 - データテンプレートを作成するツール
 - 入力ルールに従い、テンプレートにデータを入れるツール
 - 既存のウェブサイトのデータに、用語の意味、場所に対するGPS座標情報などをメタデータとして埋め込むツール
 - 自然表記された住所、日付などを分析し、テンプレートの該当箇所に自動的に埋め込むツール等

- 10月22日 (水) 東京ミッドタウンホール
 - 欧州委員会情報科学総局ISAのプロジェクト責任者 Vassilios Peristeras氏の講演
 - 欧州の相互運用性拡大施策(欧州版共通語彙基盤) について
 - 平本CIO補佐官から、日本の取り組みについて講演
 - IPAから、プロジェクト成果報告+デモ

ホール B

13:00~13:45

欧州委員会が推進する情報連携のための取り組みについて

国際標準推進

● 講演者：Vassilios Peristeras氏 (欧州委員会(EC) プログラムマネージャ)



Vassilios Peristeras氏

13:45~14:00

共通語彙基盤 (IMI) 構築への日本の取り組みについて

国際標準推進

● 講演者：平本 健二氏 (内閣官房 政府CIO補佐官)



平本 健二氏

14:00~15:00

IPAの推進する共通語彙基盤 (IMI) プロジェクトの紹介とデモンストレーション

国際標準推進

● 講演者：田代 秀一 (IPA技術本部 国際標準推進センター長)、プロジェクト担当者



田代 秀一

その他のイベント

- OECDグローバルフォーラム
 - 10月2日、3日 @ホテルオークラ東京
- IPA主催イベント
 - 2月ころ
 - 防災用情報発信への共通語彙基盤の活用結果発表
 - 防災用語彙、ツール等の公開
- インターナショナル・オープンデータ・デー (IODD) 2014
 - 2月ころ

デモ