

共通語彙基盤概要

コア語彙 2.0 対応版

2014年6月3日12章の前か後ろ（付録）共通語彙基盤の解説

本章では、共通語彙基盤の概要を解説するとともに、その使用方法を例示する。

1.1 共通語彙基盤とは

共通語彙基盤（IMI : Infrastructure for Multi-layer Interoperability）は、分野を超えた情報交換を行うためのフレームワークである。個々の単語について表記・意味・データ構造を統一し、互いに意味が通じるようにすることにより、オープンデータのデータ間の連携はもちろんのこと、行政システムをはじめとした各種システムの連携、検索性の向上等を実現する社会全体の基盤である。<https://www.ipa.go.jp/osc/kyoutsugoikiban>

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。



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



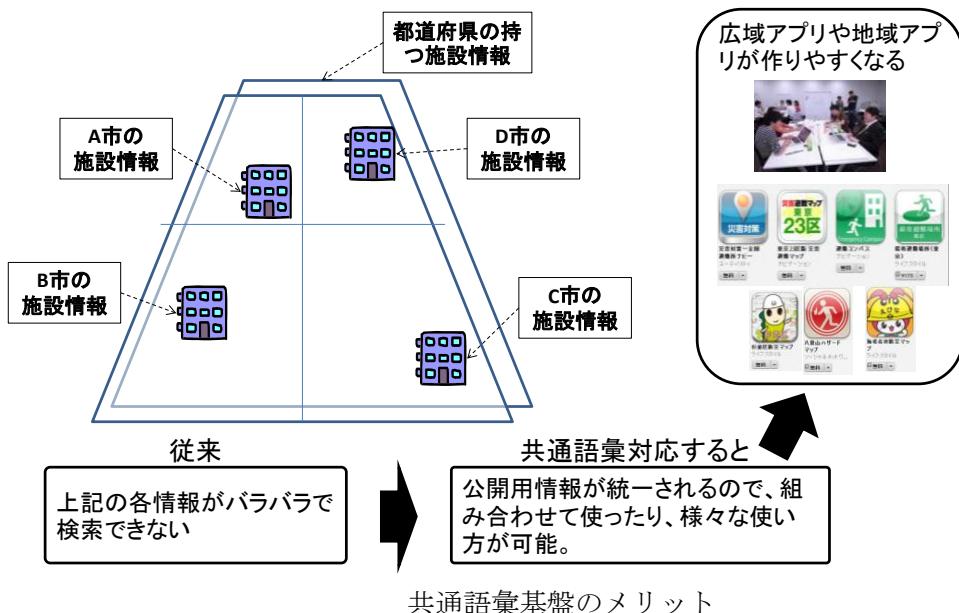
検索

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

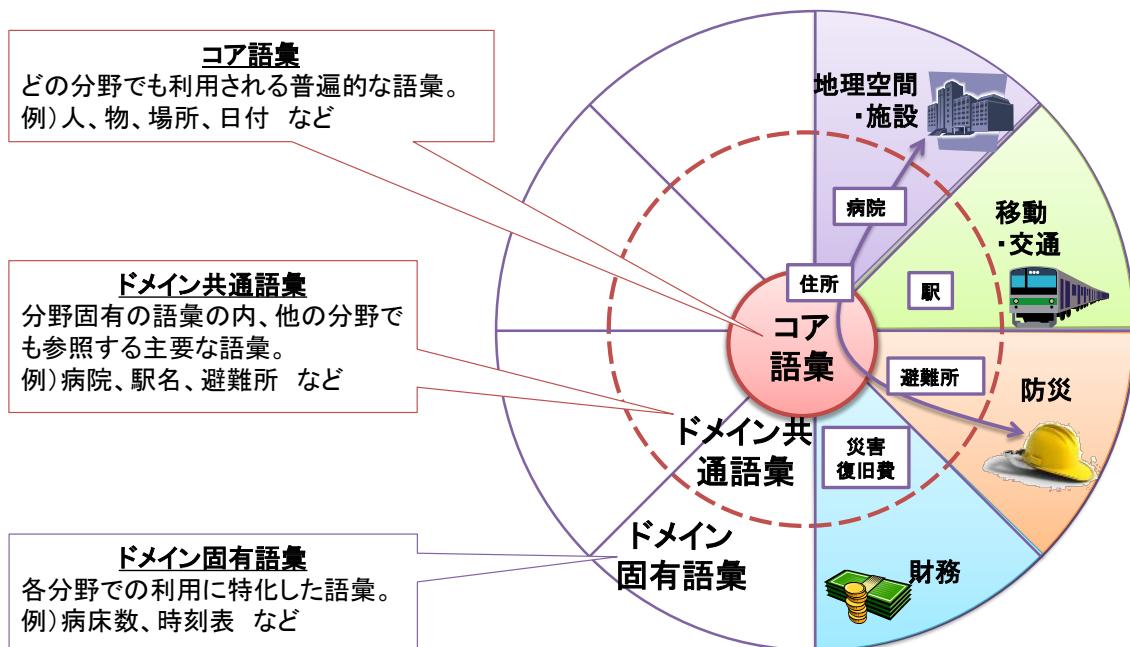
共通語彙基盤が実現する情報共有基盤

このような語彙の基盤を整備し、各語彙が正確に物事を表現できるようになると、同じ単語を違う意味で使うことによる誤解や、違う単語を同じ意味で使うことによる意思疎通の不便さを解消することができる。また、広域での情報連携を促進し、流通性の高いアプリケーションを整備することが可能であることから、データホルダ、開発者、データ活用者等のオープンデータに関する多くの関係者から期待されている。



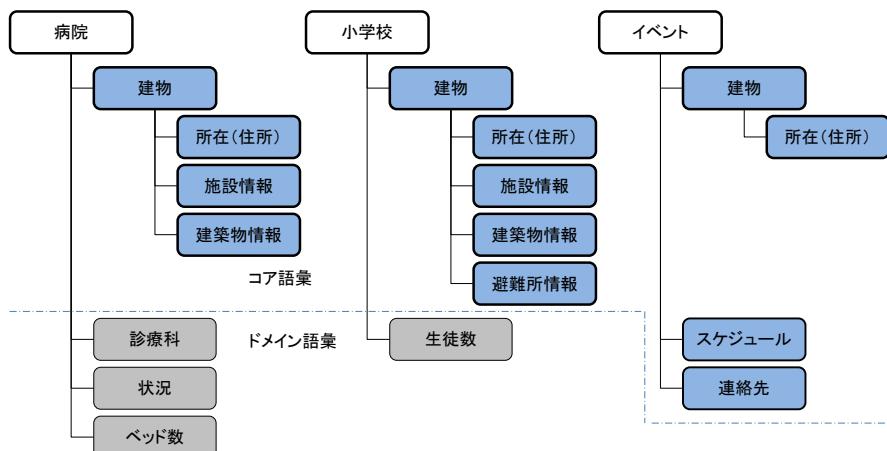
語彙（ボキャブラリ）の整備には、グローバルな視点が重要であり、共通語彙基盤は、米国政府の NIEM (National Information Exchange Model) や欧州の Joinup のボキャブラリとの連携を図るとともに、W3C、UN/CEFACT 等の枠組みを参照しながら、国際整合性を取って整備が進められている。

語彙の整備はこれまでに行われてきたが、これまでの取組みとの大きな差は、単に単語を集めて意味を明確化するだけではなく、語彙のフレームワークを整備し、構造化を図っているところである。まずは、語彙は 3 階層で考えている。その中核となるのが「コア語彙」である。コア語彙は、氏名、住所、組織名、施設表記方法等、あらゆる分野で使用される語彙である。それらのコア語彙をベースに、各種社会活動や業務で使用される語彙を「ドメイン語彙」として整理している。これは、社会共通的に整備するフレームワークの基盤的な語彙と、専門家が整理すべき語彙を分けて整理するためである。「ドメイン語彙」は、さらに、複数分野で共通的に使われる「ドメイン共通語彙」とその分野でのみ使われる「ドメイン固有語彙」の 2 分野で整理している。例えば「駅」は、市民ガイド、観光ガイド等の複数分野で使うが、時刻表で使う語彙は、交通の分野でしか使わないので、これを「ドメイン共通語彙」と「ドメイン固有語彙」と分けている。このように管理することで、分野横断の語彙の整備の効率化と相互運用性の確保を両立させている。



共通語彙基盤の構造

また、語彙を構造化して扱うのが共通語彙基盤のもう一つの特徴である。例えば、病院と小学校とイベントに関して情報を記述しようとすると、これまでには、その情報を整理したい人が、「住所」といった共通的な情報についても、それぞれが独自にその記述法を「再発明」し、その結果、表記にばらつきが出るといったことが頻繁に起こっていた。しかし、共通語彙基盤では、情報を構造化して表すようになっており、下図のように表現できる。病院、小学校、イベントを表現するのに、建物や住所等のコア語彙を再利用して組み合わせるとともに、ドメイン固有語彙を付加することで、各分野の情報管理の構造化を推進するとともに、分野横断的な情報交換を容易にしている。以下のように情報を管理することで、災害時に、標高 10 m 以上で鉄筋コンクリートつくりの建物を抽出などの作業も瞬時に行えるようになる。



コア語彙とドメイン語彙の使用イメージ

語彙を効率的に取り組む仕組みとしては、オントロジ等の技術的な解決策も研究されているが、そのような技術で解決を図る場合にも、このような語彙の基盤はその基礎をなすものとなる。

1.2 既存語彙等との共存関係

一方で、社会全体では、既に業界等で分野の語彙を整備しているところがある。しかし、分野毎の語彙では、分野横断の共同作業を行うときに不便なことがある。電機業界と機械業界では、同じ単語を違う意味で使っているかもしれない。現代社会は、様々な分野の融合したサービスが盛んであり、その場合には、業界の語彙を超えた情報交換の枠組みが必要になってくる。行政分野でも、マイナンバーの導入により、様々な分野が接続し、情報交換が行われるので、やはり分野間をつなぐ基盤が重要になる。

共通語彙基盤は、これらの既存の語彙をつなぐ基盤でもある。言葉の定義が違う業界間でも、共通語彙基盤の語彙を中間的に参照して情報交換することで、正確に情報交換できるようになる。例えば国の支援制度の情報を情報交換しようとするとき、「貸付」のことを、ある省では「融資」と定義し、他の省では「貸与」と定義している。これらは、法律などで規定されている等で統一は難しいが、情報交換するときには共通語彙基盤で定義する「貸付」に統一する、あるいは、それへの結びつきを明確にしようと決めることで、情報交換が容易になる。もちろん、オープンデータとして活用したときにも関連情報の収集や整理が容易になる。

統計データの公開でも同様に、語彙の違いによる問題が生じている。各種オープンデータの統計を集めてきたときに、そのデータ項目の持つ意味が異なる場合には、重ね合わせたデータが意味をなさない場合もある。例えば、「一人当たりの水消費量」といったときに、水道水の使用量か、飲料水の消費量か、ミネラルウォーターの消費量かわからない。また、単位も、リットルか、トンか、ガロンかがわからないし、単位表記の語彙も併せて整理しておく必要がある。「リットル」、「l」、「リッタ」、「liter」は人の目では同じと判断できるが、コンピュータでは、判断が難しい。

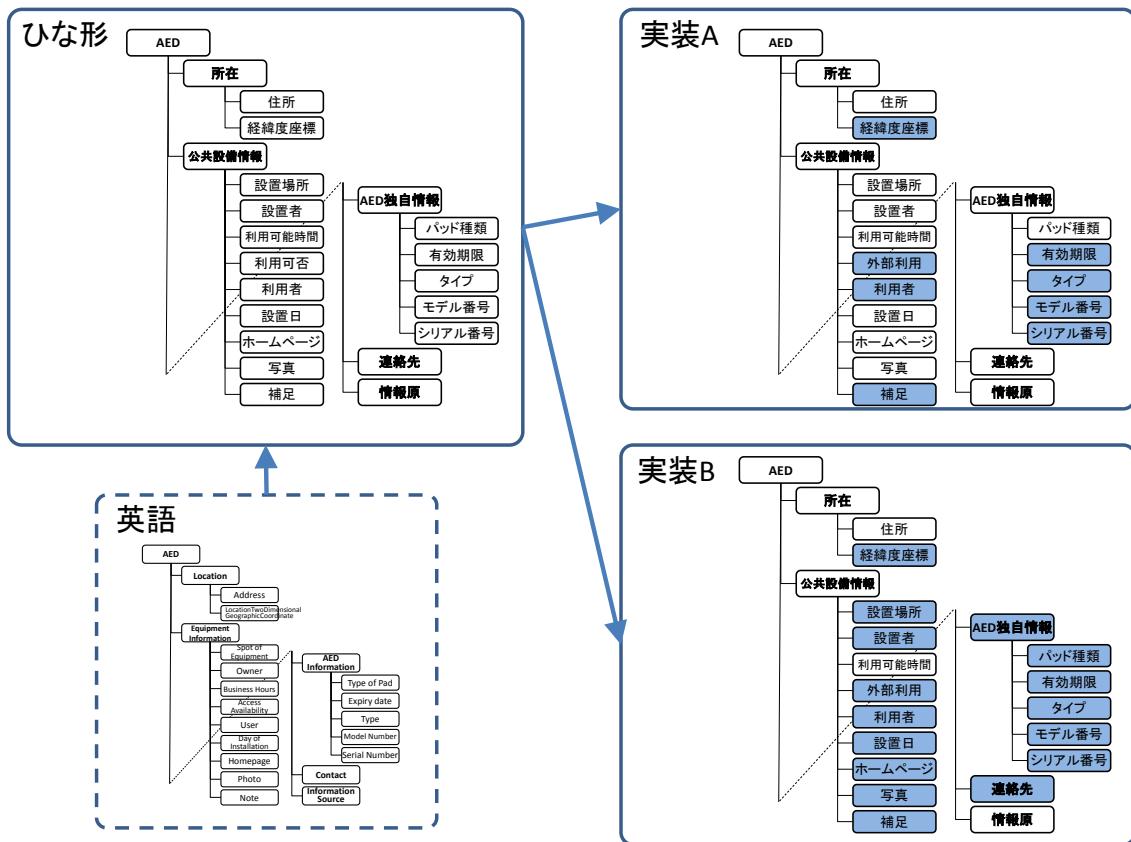
情報流通連携基盤共通 API の中のボキャブラリの整理なども、共通語彙基盤のフレームの中で整理・連携することが可能である。

1.3 共通語彙基盤の導入方法

共通語彙基盤は、現在開発中であるが、フレームワーク化されているので、順次導入することが可能である。2014年2月に行われたインターナショナル・オープン・データ・デーでは、ハッカソンを効率的に進められるよう、開発支援キットを提供するなど、語彙の整備と検証が並行して行われている。(コア語彙2.0は、6月上旬に公開予定)

AED の情報を例にとると、AED の開発支援キットを使えば、既に、AED に関する語彙が洗い出されて体系化されているので、その語彙やその構造について考える手順が省略できる。そのうえで、実装するアプリケーションの目的に応じて使用するデータ項目を選択し、

アプリケーション開発を行えばよい。そうすることで、広域での相互運用性が高く、流通性もあるアプリケーションが開発できる。また、英語名も用意されているので、国際対応も容易に可能となる。



データ項目設定の例

なお、開発支援キット等で使われているひな形の作成は、複数の行政機関やNPO等で実際に使われているデータ項目を精査したうえで整備している。

以下に導入手順を示す。

① 導入の目標を決める

目標を明確にすることで、必要なデータ項目を選択することができる。この部分が明確でないと、データ項目に過不足が生じることとなる。

② 既存語彙や共通語彙基盤の整備状況を確認する

現在、その対象物を管理している場合には、どのようなデータ項目で管理しているかを確認する。また、その分野に関連した、コア語彙を確認するとともに、ドメイン用の支援キットなどが用意されているかどうかを確認する。

③ データ項目を整備する

ひな形がある場合には、その項目から公開データに使うデータ項目を選択する。ひな形がない場合には、住所等、コア語彙の中から利用できるデータ項目を選択する。必

要に応じて、独自項目を追加する。

④ データ項目の検証をする

いくつかの例を入力して、情報項目に不足しているものがある場合には、追加を検討する。独自項目として追加するものと、共通語彙基盤事務局に改善を呼びかける場合がある。

⑤ データの定義書を作成する

データの定義書（アプリケーション・プロファイル）を作成する。共通語彙基盤事務局に事例登録することで、他組織での活用に寄与することが可能である。

⑥ データを登録する

データの入力をしていく。既存の各部門が情報を持っている場合には、コンパートして入力する。すべての情報を入力しようとすると、入力者の負担になることがある。空欄があってもかまわなくらいの気持ちで最初は取り組むことが重要である。

⑦ 情報を公開する

API、CSV 等の情報が再利用しやすい形で情報を公開する。

⑧ 利用を呼び掛け、フィードバックを呼びかける

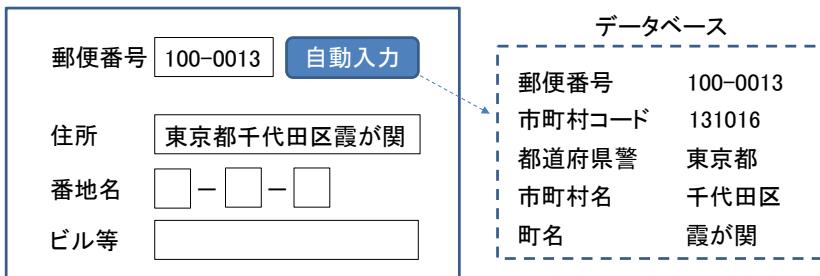
データ活用を推進し、改善のためのフィードバックを呼びかける。

1.4 情報連携用語彙データベースおよび支援ツールとは

共通語彙基盤を効率的に導入するための情報連携用語彙データベースの開発と、関連ツールの開発が進められ、試行運用されている。情報連携用語彙データベースは、使いたい語彙が既に定義されているか、類似の語彙がないかを検索する等、語彙を管理するためのデータベースである。語彙の増加、ひな形の増加、情報交換用パッケージの追加などには必要な基盤である。利用者による語彙の追加要望や既存の語彙等の追加は、事務局で審査のうえで実施し、語彙の品質を確保していく。

支援ツールは、データのテンプレートの設計や、データの入力や表示を支援するツールである。現在でも、郵便番号を入力すると、住所の町まで自動で入力してくれる各種サービスはあるが、システム的には、市町村と町を分離して持ちたい場合もある。その場合には、入力者に負担をかけずに、自動的にデータを生成するツールが必要となる。データを精緻に持ちたいという要望と簡易に入力したいという、相容れない要望を実現するためにはこうした入出力の支援ツールが必要であり、順次整備を行っている。下記の千葉市の例は、ウェブで公開するデータの中に、それをオープンデータとして活用しやすくするために、そのデータの意味等を説明する情報をメタデータとして埋め込む作業を支援するツールの例である。

画面



住所入力支援ツールの例

千葉市

イベント情報の項目のデータとして設定された箇所は反転表示

平成25年度「青少年の日フェスタ」を開催しました！

毎年9月の第3土曜日は、「青少年の日」です。このフェスティバルは、青少年と家庭、学校、地域、行政がつながりを持つことで、青少年のコミュニケーション力を高め、居場所づくりを推進することを目的として、「青少年の日」に実施しています。

毎年9月の第3土曜日は、「青少年の日」です。このフェスティバルは、青少年と家庭、学校、地域、行政がつながりを持つことで、青少年のコミュニケーション力を高め、居場所づくりを推進することを目的として、「青少年の日」に実施しています。

イベント情報の項目にデータが設定される

「青少年の日フェスタ」の様子

イベント情報入力支援ツール例