

共通語彙基盤概要

解説

本文書では、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。

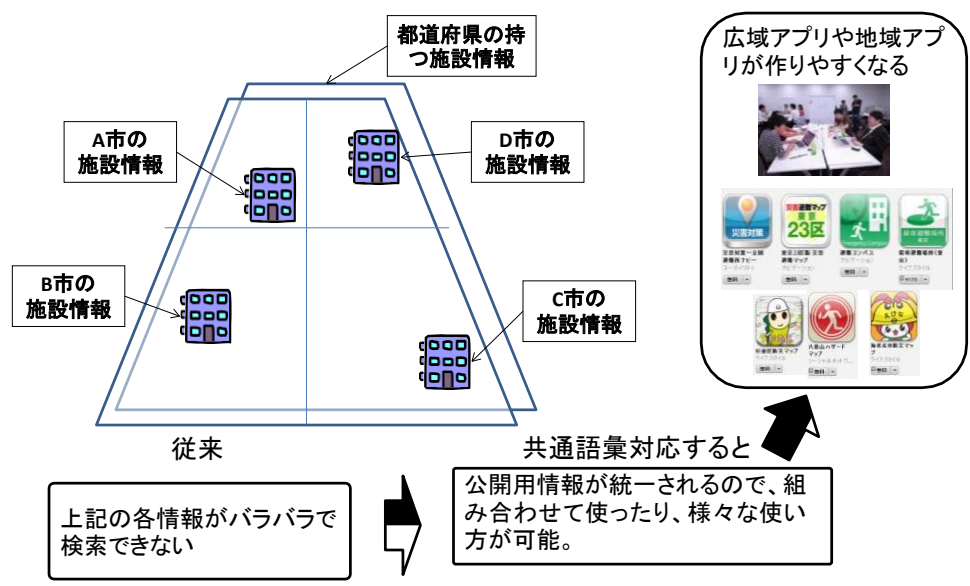
1.1 共通語彙基盤とは

共通語彙基盤 (IMI Vocabulary : Infrastructure for Multi-layer Interoperability Vocabulary) は、情報共有基盤 (IMI) の一部で、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。



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。この基盤は、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開発を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早期の整備が期待されている。

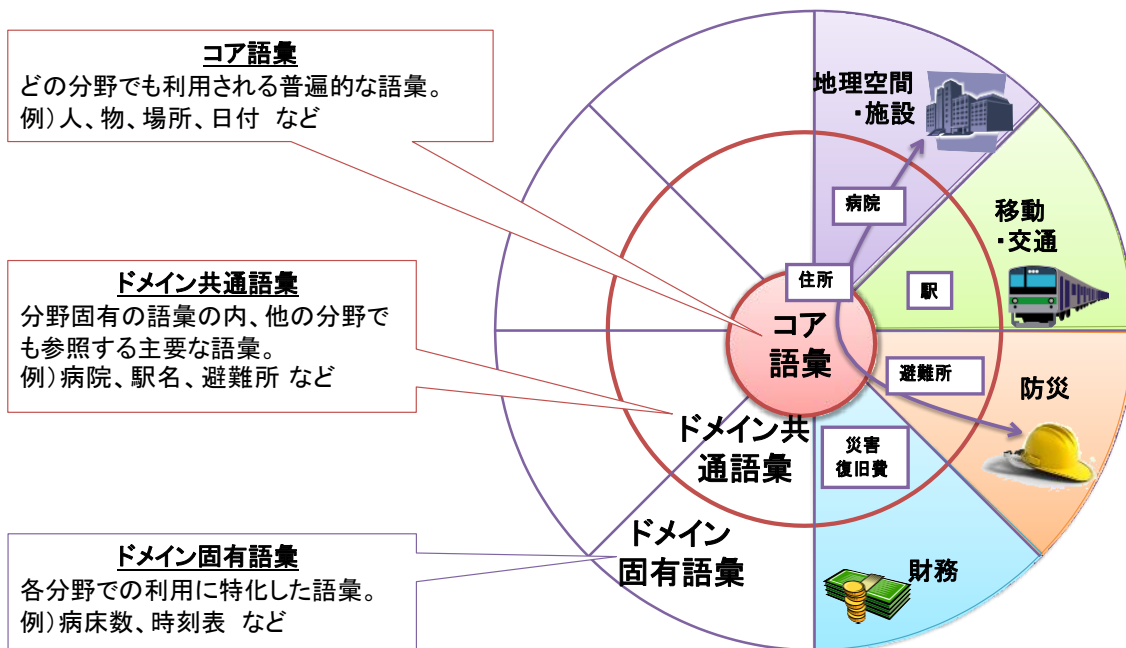


共通語彙基盤のメリット

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。

共通語彙基盤の以前から語彙の整備はこれまでも行われてきたが、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。また、これまでの EDI 等の取り組みは業界毎に
 行われることが多く、社会全体でコンセンサスをとるべき語彙の整備が進んでこなかった。

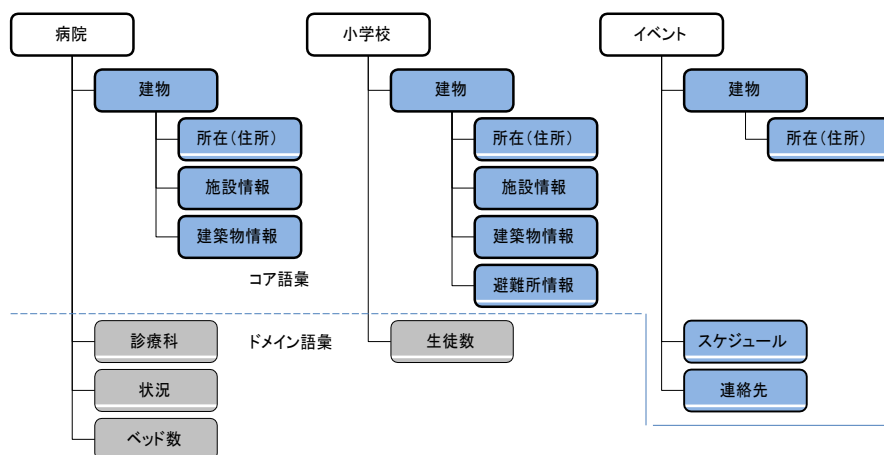
共通語彙基盤では、語彙は 3 階層で考えている。その中核となるのが「コア語彙」である。
 コア語彙は、氏名、住所、組織名、施設表記方法等、あらゆる分野で使用される語彙である。そ
 れらのコア語彙をベースに、各種社会活動や業務で使用される語彙を「ドメイン語彙」として整
 理している。これは、社会共通的に整備するフレームワークの基盤的な語彙と、専門家が整理すべ
 き語彙を分けて整理するためである。「ドメイン語彙」は、さらに、複数分野で共通的に使われ
 る「ドメイン共通語彙」とその分野でのみ使われる「ドメイン固有語彙」の 2 分野で整理して
 いる。例えば「駅」は、市民ガイド、観光ガイド等の複数分野で使うが、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時、これまでは、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う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10m以上で鉄筋コンクリートづ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

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、「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する決めることで、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」 「㍓」 「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 の中で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。

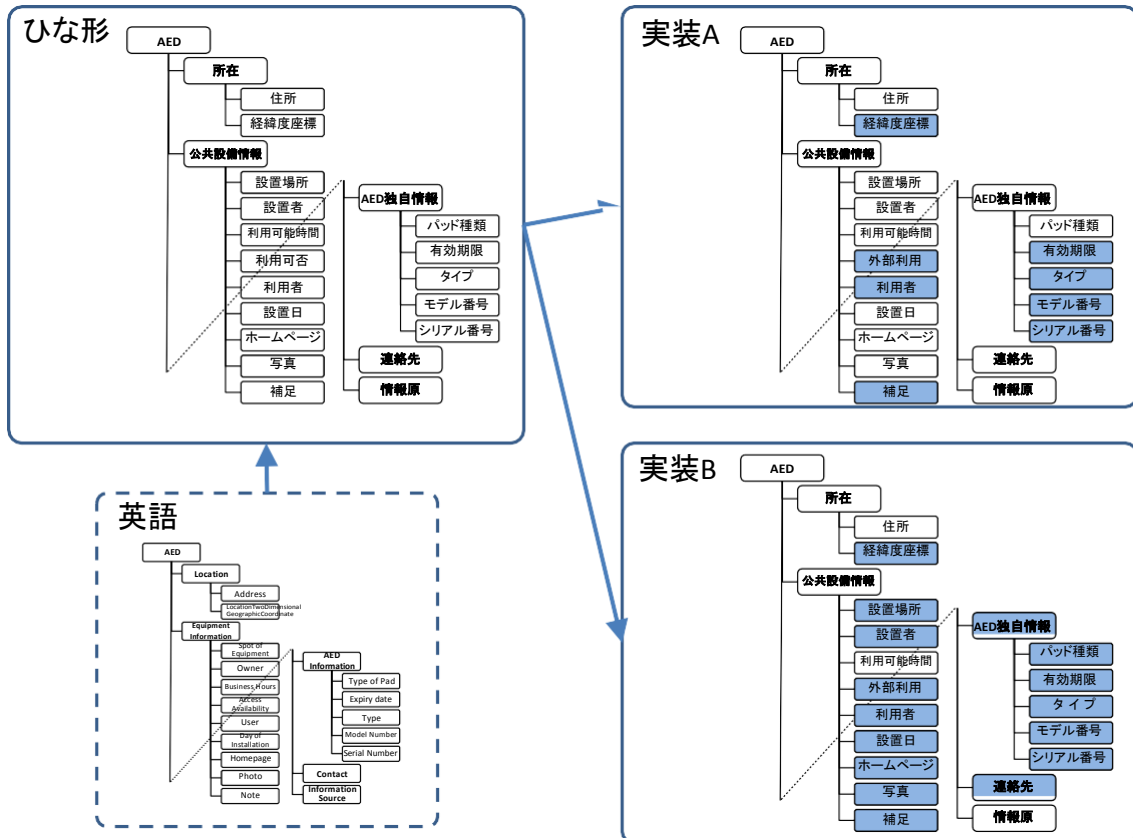
1.3 共通語彙基盤の導入方法

共通語彙基盤は、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月に行われた国際的なオープン・データ・デーでは、ハッカソンを効率的に進められるよう、情報交換パッケージ (IEP : Information Exchange Package) を提供するなど、語彙の整備と検証が並行して行ってきた。その後に語彙全体の見直しが行われ実証を行う等、内容の精査を図っている。また、上述の情報交換パッケージの内容を充実させ、データモデル記述 (DMD : Data Model Description) として仕様の策定が行われた。

AED の情報を例にとると、AED の DMD 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

えて、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、DMD 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、その分野のひな形であるDMDなどの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

DMD 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。DMD がない場合には、データに含める項目をCSV形式などで作成したうえで、DMDエディターを用いて、DMDを作成する。

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）を作成し、DMDに追加する。作成したDMDをDMDエディターに登録することで、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、コンバートして入力する。すべての情報を入力しようとする、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いくらいの気持ちで最初は取り組むことが重要である。データがCSV等の表形式のデータとして作成されている場合は、DMDエディターを利用して表形式のデータを変換して共通語彙基盤形式のデータを作成することができる。

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、支援ツール、共通語彙基盤ライブラリとは

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。

共通語彙基盤ライブラリは、共通語彙基盤の各種の語彙、DMDやデータを容易に扱えるようにするためのライブラリで、データを生成するシステムやデータを利用するアプリケーションを容易に構築できるようにするためのものである。

画面

郵便番号	100-0013	自動入力
住所	東京都千代田区霞が関	
番地名	□ - □ - □	
ビル等	□	

郵便番号	100-0013
市町村コード	131016
都道府県警	東京都
市町村名	千代田区
町名	霞が関

住所入力支援ツールの例

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

イベント情報の項目にデータが設定される

日時・会場	平成25年9月21日(土)
日時	メイン会場: 午前10時から午後4時まで サブ会場: 午後1時から午後4時まで
会場	メイン会場: 千葉市生涯学習センター(千葉市中央区井天3-7-7) サブ会場: 千葉市子ども交流館(千葉市中央区中央4-5-1きぼーる内3~5階)

イベント情報入力支援ツール例

1.5 共通語彙基盤のメンテナンス

導入に当たってデータ項目の変更等が起こる問題を指摘されている。アジャイル型で整備が進められているため、データ項目の変更が全くないとは言えないが、中核部分に変更が起こらないように、十分な検討が行われている。

共通語彙基盤のメンテナンスは、語彙の専門家による委員会において行われている。メンテナンスの方法と指定は以下の方法で行う予定である。

(1) メジャーアップデート

コア語彙の主要項目のアップデート等、影響範囲が大きいアップデートであり、数年に1回の実施を予定している。将来的にドメイン語彙の委員会ができるときには、ドメイン語彙とコア語彙の語彙の調整なども含まれる。

(2) マイナーアップデート

特定分野の追加等、ブロック単位での追加等を行うアップデートである。実証等により特定語彙ブロックなどができた場合に実施される。

(3) マイクロアップデート

ケアレスミスや語彙解説の精緻化等の軽微なアップデートであり、随時行われる。

これらのコア語彙のアップデートと並行して、DMD の追加は随時行われる。

共通語彙基盤の成熟に合わせて上記のように修正が行われていくが、整備は継続的に行われるものであり、共通語彙基盤の導入はその時点のバージョンで進めていくべきである。古いバージョンで実装した場合にもデータが構造化されているので、これまでのデータ移行と異なり、新バージョンへの移行も軽微な負担で行うことができる。