

活用事例

法人インフォ機能強化支援

2018年11月13日(火)

独立行政法人 情報処理推進機構
社会基盤センター産業プラットフォーム部
データ活用推進グループ



● 目的

IMI情報共有基盤(以下「IMI」という)を活用して構築されている法人インフォメーションをターゲットとし、IMIを一層使いやすいう、改善する。

これにより、一般的に難解である「語彙」の活用への障壁を下げ、IMIの活用拡大を図り、横断的な官民データ活用を促進するための基盤を整備するとの政府方針に応える。

▶ 法人インフォの機能強化支援

法人インフォを、IMIのユースケースとして確立したものとするため、法人インフォの開発事業者と連携し、機能や運用効率の改善に向けた支援を実施する。

その過程で得られた知見を、汎用性のある語彙の仕様や、語彙構築ノウハウ等として整備し、共通語彙基盤全体の改善、普及広報へ活用できるよう、以下の項目について、実施する。

- (ア) 法人インフォ用「ドメイン語彙」の整備
- (イ) 法人インフォ用DMDの整備
- (ウ) 法人インフォ用変換ツールの整備
- (エ) DMDメンテナンスツールの整備

これまでコア語彙等を開発してきたIPAの経験を生かすとともに、IPAが提供している既存ツールを改良することにより、効率的に整備を進める。



法人インフォ用「ドメイン語彙」の整備

以下の2点を重点に拡張を検討し整備する。

1. 各データ項目に対し、当該項目に係るデータの**変更の履歴**、**データ取得元**(各省庁等)等を記載するための項目を拡張する。

データについて、それがツールによる処理の結果であるのか、あるいはオリジナルのままであるかが区別できる情報を構造化データの一部として書き込む機能を設けるために行う拡張である。

2. 各データ項目に対し、当該項目に記載されたデータが**公開可**であるか、**非公開**であるかといった、**利用制限情報**(利用条件)を記述できるよう、項目を拡張する。

法人ドメイン語彙は、現在は、法人インフォとして公開することを目的として整備されている。しかし、将来的には、法人ドメイン語彙が非公開情報も含め、法人に係る情報を幅広く扱うことが想定されるため、そこから公開情報のみを抽出できる方式をとることが出来るように拡張する。

コア語彙への拡張提案

このような拡張は、法人ドメイン語彙に限らず、他の用途一般においても必要とされることが想定されるため、各語彙の共通した継承元となるコア語彙自体に拡張提案し、全ての語彙がこの拡張の恩恵を受けられるように検討する。



本事業で整備するDMD

IMIにおいて、「DMD」は、特定のデータ、本事業においては法人インフォで提供されるデータの構造、を記述したものである。

DMDには、当該データで使用するデータ項目の一覧である「**データモデル**」と、当該データを構築する際に、元となるデータ(情報源から提供されたオリジナルデータ)のどの項目を、構造化データのどの項目へ割り当てるかの対応関係を記述した「**マッピングファイル**」が含まれる。

本事業においては法人インフォ用のDMDとして以下の整備をする。

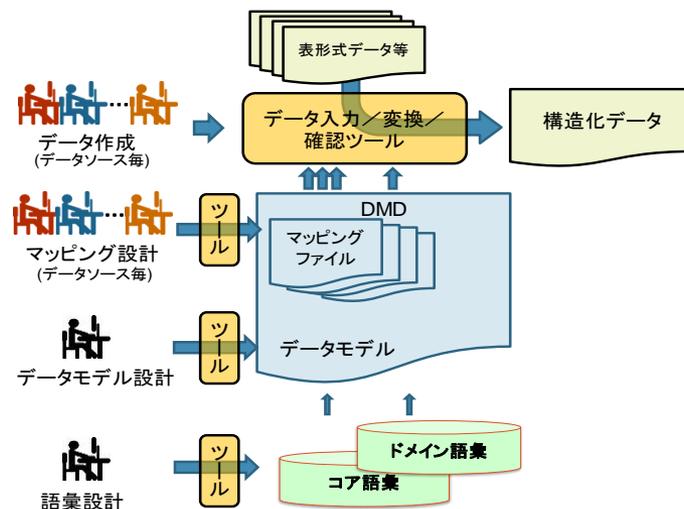
▶ 法人インフォ用のデータモデル

データモデルについては、**法人インフォに共通なもの**を一つ作成する。

データモデルには、**データ項目に関する説明**を含める。

▶ 法人インフォにデータ元となる主要な**情報源**に対応したマッピングファイル

▶ マッピングの情報、正規化規則、文字セット制限、バリデーションルールを含んだものとする。



データ正規化規則の対応

正規化(住所、法人名の分割、全角・半角文字の統一等)の規則を記載する手段は、現在のDMD仕様には含まれていない。そのため、これに対応するための仕様拡張を行う。

正規化規則としては、

- ▶ 一続きに記述された住所の、県、市、町、番地等への分割
- ▶ 県名等、住所の一部が省略された際の補完
- ▶ 番地記述の正規化
「2-2-1」、「2丁目2-1」、「二丁目」等の多様な表記を統一的表記へ変換
- ▶ 法人名から法人種別名の抽出
「株」「株式会社」「(株)」等を法人名から抽出し、正規化して「法人種別」項目へ格納
- ▶ 数字、カナ表記、記号、空白文字における全角-半角統一

(次ページへ続く)



正規化処理モジュールの呼び出しについて、当該モジュールへ渡すパラメータ定義、当該モジュールからの戻り値の格納先定義を記述できるよう、DMDのマッピングファイル仕様を拡張する。

併せて、正規化の変換時にエラーが生じた際の対応につき、

- ▶ エラー情報を、当該データに対応する**データ変更履歴**記述用の項目に書き込む
- ▶ **人の介入**を要求する

等を定義できるよう、**正規化規則記述仕様**を整備する。

正規化については、政府の「行政データ連携標準」を参照して設計する。

なお、**エラー発生時にそのエラー内容を記載**するためのデータ項目は、現在の法人インフォで使われている語彙には存在しないため、語彙に対して必要な拡張を行うものとする。



データバリデーションルールの記述

「法人インフォ用変換ツール」等により、データを変換する際、及び、既存の構造化データを確認する際などに、データを確認し、不正なデータについてどのような反応をするべきかを規定するための記述方式を、マッピングファイルの仕様として追加定義する。

記述方式には、確認すべき条件と、その条件に反したデータについて「何を行うか」の対応のペアを記述する。

条件には

- ▶ 必須項目の欠如
- ▶ データ型の不一致
- ▶ 使用可能文字からの逸脱
- ▶ 指定されたデータ範囲からの逸脱 等がある。

また、対応には

- ▶ 当該データに対応する「メタデータ」へ、エラー内容を格納する
- ▶ 人的介入を要求する 等がある。



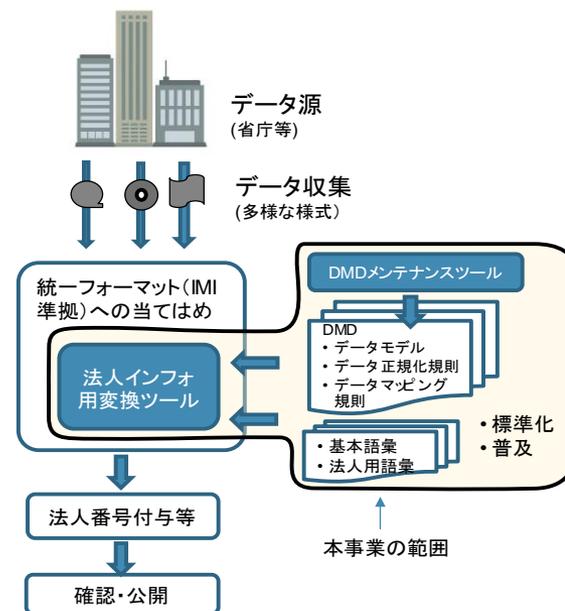
- ▶ 本事業では、「法人インフォ用変換ツール」と、「DMDメンテナンスツール」の2種のツールを整備する。

これらの位置付けを右図に示す。

- ▶▶ 「法人インフォ用変換ツール」は、収集したデータを「統一フォーマットへの当てはめる」作業を支援するものである。

- ▶▶ 「DMDメンテナンスツール」は、「法人インフォ用変換ツール」におけるデータ変換に必要な変換規則を記載したDMDおよびマッピングファイルの編集を行うツールである。

- ▶ 情報システムや構造化データに詳しくない担当者でも容易に操作できるように、操作性の向上、ガイダンス機能の追加などを行う。



* IPAでは、データモデルの設計、データの変換などを行うツール(IMIツール)をウェブサービスの形で提供している。
<https://imi.go.jp/tools/v2/>

法人インフォ用変換ツールの整備

表形式データ(CSV形式及びExcel形式)を読み込み、構造化するため、以下の機能を有する変換ツールを構築する。

- ▶ マッピングファイルに示されたマッピング情報に従い、元データから読み取ったデータを、構造化データの対応する項目へ書き込む機能。
- ▶ マッピングファイルのデータ正規化規則に従って、データを分割、変換し、構造化データ中の指定された項目へ割り当てる機能。
 - ▶ 変換に際し、変換のために用いる住所辞書に存在しない文字列がある等により住所を分割できなかった場合、住所に省略があったために補った場合等は、正規化規則中の指定に従い、**人の介入の要求**、当該データ項目に対応するメタデータへの**エラー内容の書き込み等**を行うものとする。
 - ▶ 変換の際、構造化データとして書き込んだデータが、正規化規則に従って**自動変換された結果**であるのか、あるいは変換を経ない**オリジナルのままのデータ**であるのかを区別できる**変換履歴データ**を、当該データ項目に対応するメタデータとして書き込む。
- ▶ データバリデーションルールに従い、エラー処理を行う機能。
 - ▶ **データが指定された範囲等から逸脱する場合には**、データバリデーションルールの指示に従い、**人の介入の要求**、当該データ項目に対応するメタデータへの**エラー内容の書き込み**などを行うものとする。

本ツールは、データの追加・更新の際に日常的に使われるものであることを考慮し、DMD、マッピングファイルの指定、入力データの指定などをわかりやすいユーザインタフェースで指定できるようにするとともに、作業がルーチン化されて実施されることを想定し、定型的作業を指定し、記録することのできるユーザインタフェースを設ける。

あわせて、生成された構造化データを法人インフォへ登録、あるいは法人インフォシステムが情報源から取得するための処理を自動化し、効率化するための**APIを提案**し、本ツールへ実装する。



DMDメンテナンスツールの整備

DMDの構成要素であるデータモデルは、法人インフォのデータ構造を記述したものを一つだけ、マッピングファイルについては、情報源毎にその変換規則を含めて用意する。

データモデルのメンテナンスとマッピングのメンテナンスは、右表のように、その想定利用者や利用頻度が規則異なることが想定される。

ツールの種類	想定利用者	利用頻度
データモデル (データの正規化規則)のメンテナンス	法人インフォ担当者	法人インフォのデータ構造を更新する場合にのみメンテナンスが必要になる。 → 頻度は低い
マッピングファイルのメンテナンス	法人インフォへ格納するデータの提供者(各省庁担当者等)。ただし、法人インフォ担当者等が一括して代行を行うことを想定する。	変換元となる、元データについて仕様変更があった場合。 → 比較的頻度が高い

そこで、これらのツールを明確に区別して提供できること、また、情報システムに不慣れな利用者でも容易に操作できることを考慮し、以下の観点から、現在IPAが提供しているツールの機能追加と改良を行う。

- ▶ マッピングファイルが入っておらず、「**モデル**」のみが格納されたDMDと、**マッピングファイル**とを別個に**管理**し、編集する機能の追加
- ▶ DMDのデータモデルに対し、そのモデルにおける**各データ項目の説明**を入力・編集する機能の追加
- ▶ データモデルに定義された項目と、変換元となるデータに定義された項目名との**対応関係の視認性向上**など、指定を容易に行えるようにする観点での改良
- ▶ マッピングファイルに対し、**対応関係に対する説明やメモ書き**などを、対応毎に記録・表示・編集できる機能の追加



- ▶ 法人インフォをユースケースとして想定利用者を分類
それぞれにとって使いやすいようツールと仕様を改良する

