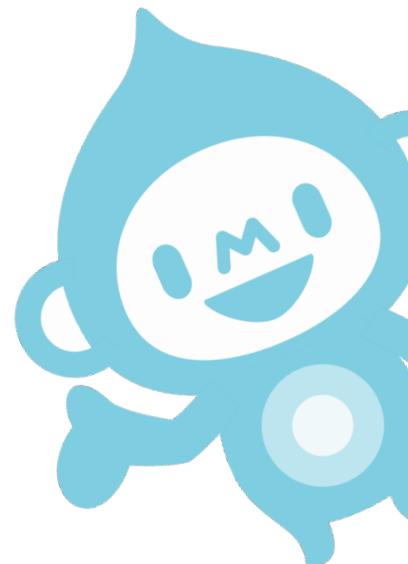


つながるデータに 必要な基盤とは？

2017年6月29日

経済産業省 CIO補佐官
平本 健二



つながるデータの対象は何か

社会の基本情報 & IoTデータ



つながるとは？



Connect

Link

Join



Compatibility
互換性

Interoperability

相互運用性

組織、業務プロセス

- 「データ→プリント→入力→データ」の防止

ルール

- データ:政府統一利用規約

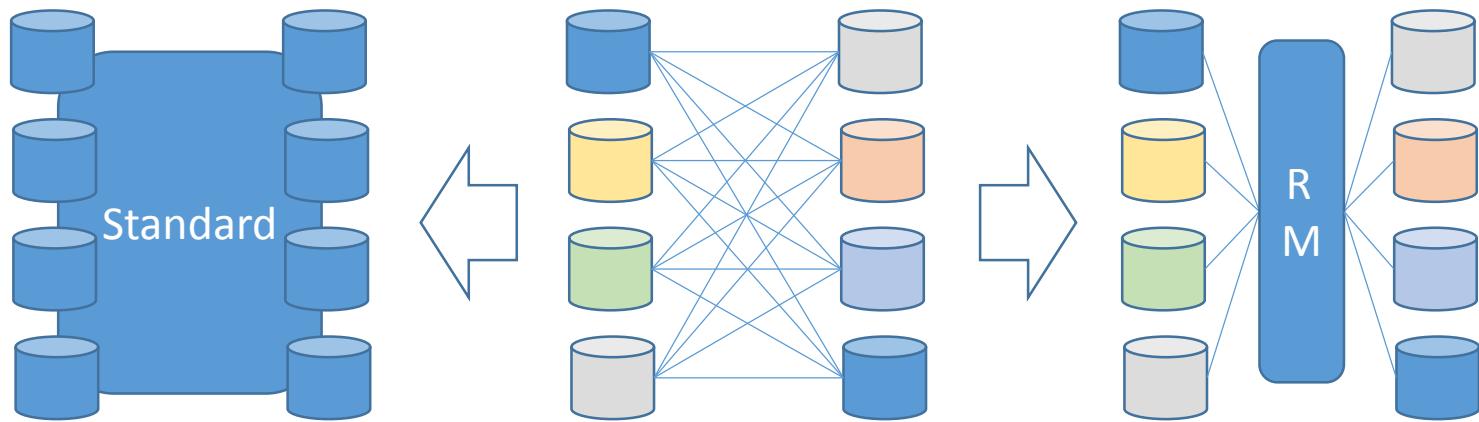
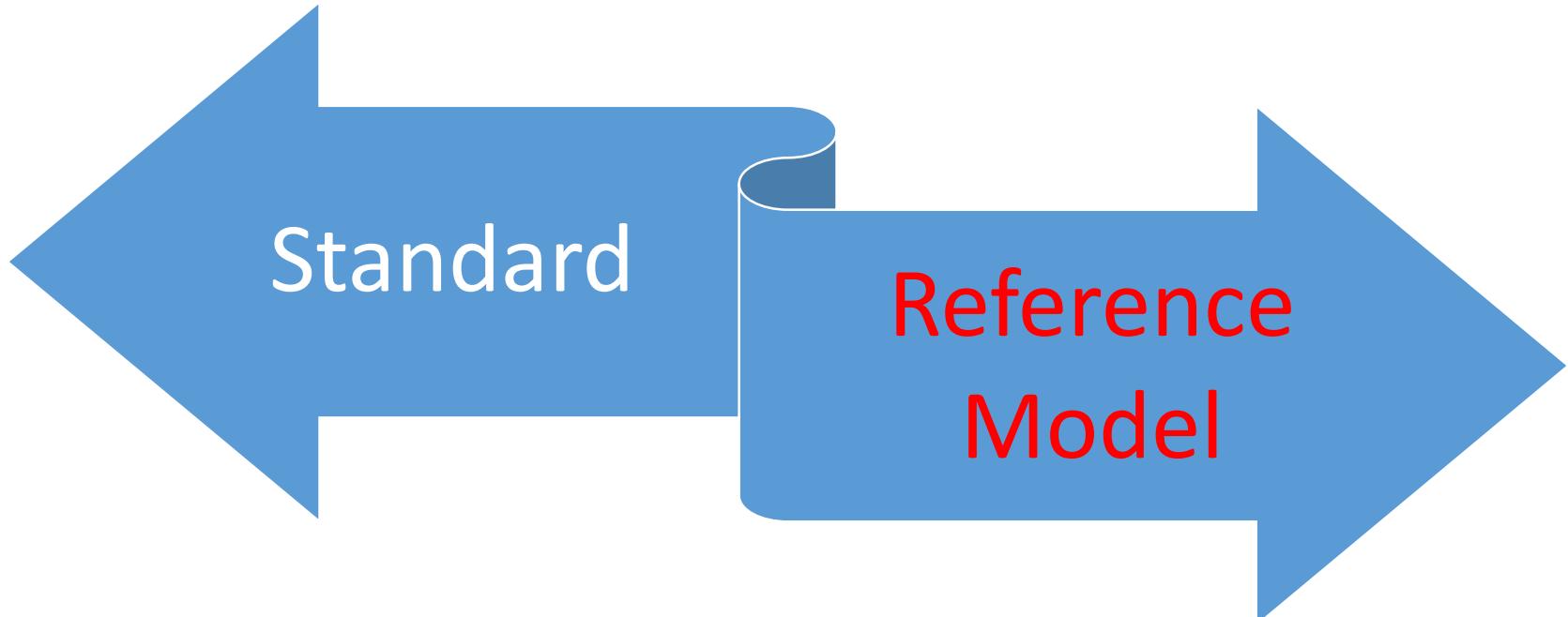
語彙

- データ定義(意味)、データ構造

技術

- XML、CSV、JSON — USB、DVD — wifi

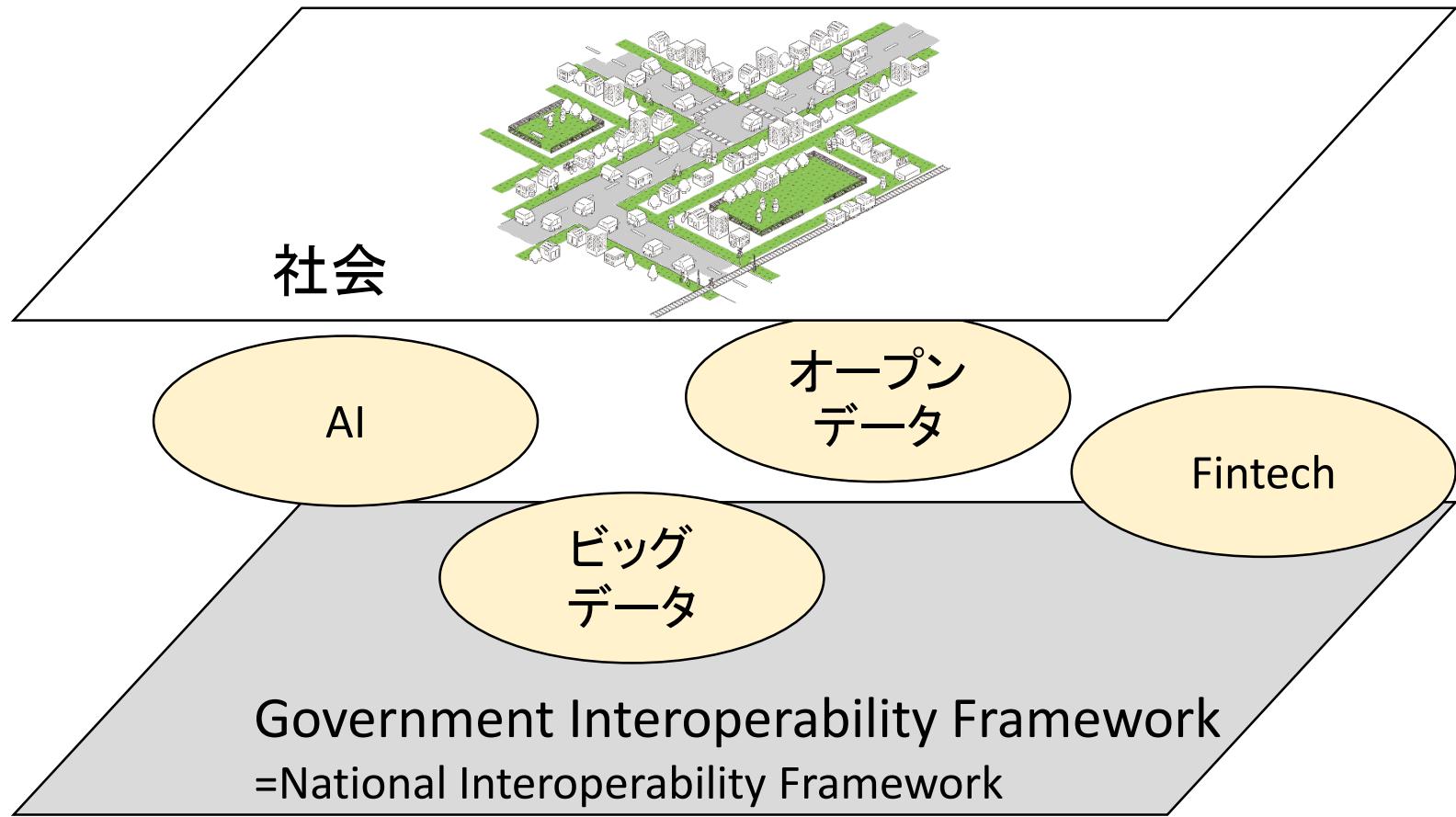
Interoperabilityを確保する仕組み



なぜInteroperabilityにこだわるのか

IMI
Infrastructure
for Multi-layer
Interoperability
2017/6/29 METI/IPA IMI関連イベント公開資料

- インタオペラビリティは社会の基盤。
- 話題の最新技術は全てInteroperabilityの上にのっている。



世界最先端IT国家創造宣言

- データ標準化(語彙、コード、文字等)、API

官民データ活用推進基本法、基本計画

- 相互に連携して、自らの情報システムに係る規格の整備及び互換性の確保、業務の見直しその他必要な措置(基本法15条)
- 情報システムの相互の連携を確保するための基盤の整備その他の必要な措置(基本法15条)
- データ連携のためのプラットフォーム整備(基本計画)

デジタル・ガバメント推進方針、(基本計画)

- 官民連携を実現するプラットフォーム

オープンデータ基本方針

- 公開データの形式等

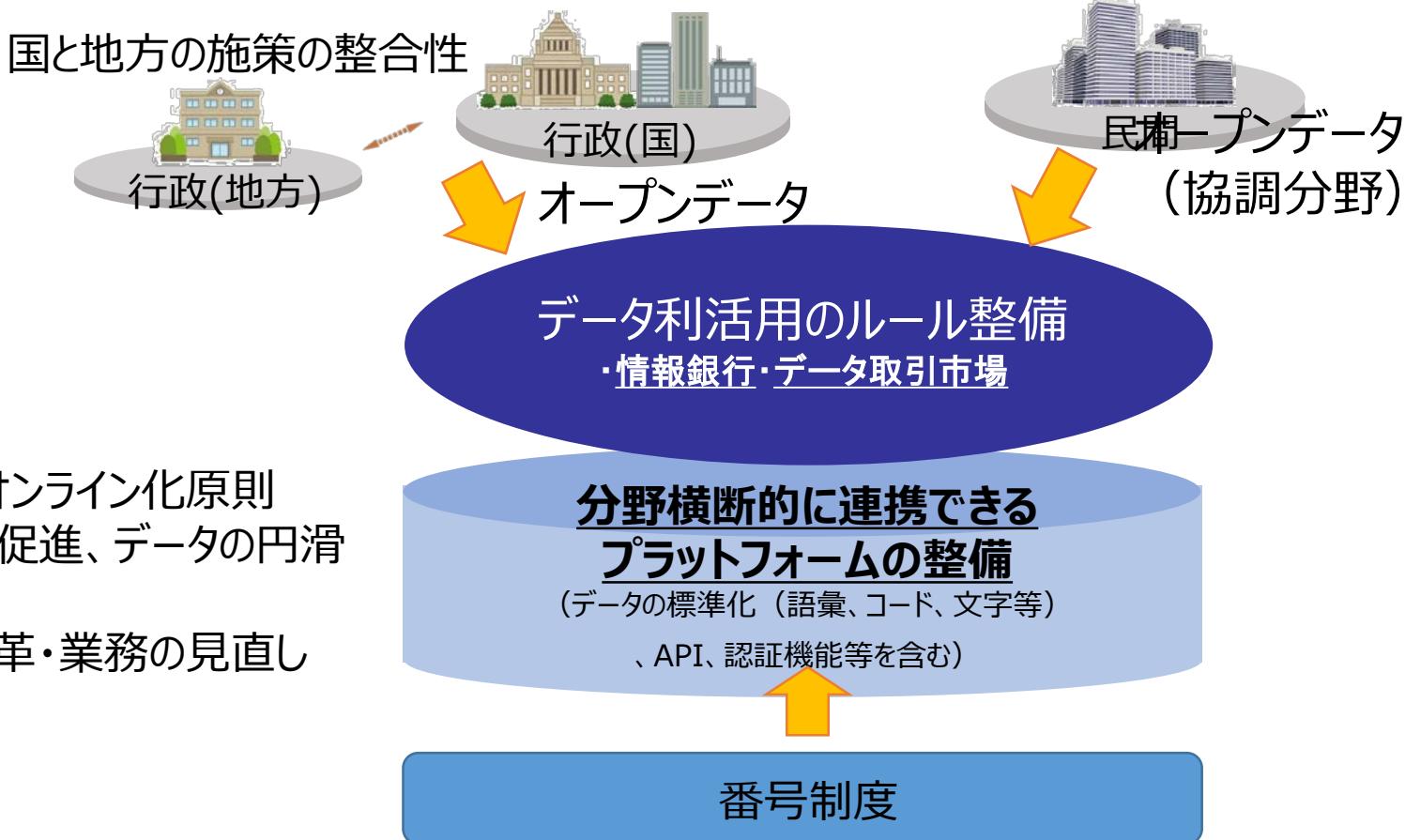
世界最先端IT国家創造宣言・ 官民データ活用推進基本計画

「データ大流通時代」の到来

- あらゆる場面で、ネット上の知識や知恵を共有・活用することにより、我々の生活や産業を一変する可能性。
- 「官民データ活用推進基本計画」の策定・推進により、全ての国民がIT・データの利活用を意識することなく便益を享し、真に豊かさを実感できる社会のモデルを世界に先駆け実現



世界最先端IT国家創造宣言・ 官民データ活用推進基本計画 全体イメージ



- これまでのIT化・業務改革の取組を着実に推進するとともに、官民データ活用推進基本法の成立等を受け、これまで以上に国民・事業者の利便性向上に重点を置き、官民データの流通等に資する新たな取組を推進する必要
- 行政内部の効率化にとどまらない国民・企業等に価値を提供するサービス部分の変革に焦点を当て、行政のあり方そのものをデジタル前提で見直すデジタル・ガバメントの実現を目指す

1. デジタル技術を徹底活用した利用者中心の行政サービス改革

- **サービスデザイン思考**に基づく業務改革（BPR）の推進
- デジタル技術に対応した情報提供のあり方の見直し

2. 官民協働を実現するプラットフォーム

- データ流通を促進する環境の整備
- 官民データ活用のためのインターフェースの整備
- プラットフォームの共用化と民間サービスの活用

3. 価値を生み出すITガバナンス

- サービス改革に対応した推進体制の整備
- ITマネジメントの徹底と投資効果の最大化

考え方

- インターネット前提社会。
- 取り組んでいるかどうかではなく、世界との相対スピードが重要。
- 行政が足手まといになるのではなく、社会活動を加速するプラットフォームに。

デジタルファースト

ワンストップ

ワンスオンリー

オープンデータ基本指針 概要

1. オープンデータの意義

- (1) 国民参加・官民協働の推進を通じた諸課題の解決、経済活性化
- (2) 行政の高度化・効率化
- (3) 透明性・信頼の向上

3. オープンデータに関する基本的ルール

(1) 公開するデータの範囲

各府省庁が保有するデータは、原則オープンデータとして公開。公開することが適当でない公共データは、公開できない理由を原則公開するとともに、限定的な関係者間での共有を図る「限定公開」といった手法も積極的に活用。

(2) 公開データの二次利用に関するルール

原則、政府標準利用規約を適用。

(3) 公開環境

特にニーズが高いと想定されるデータは、一括ダウンロードを可能とする仕組みの導入や、APIを通じた提供を推進。

(4) 公開データの形式等

機械判読に適した構造及びデータ形式で掲載することを原則。法人情報を含むデータは、法人番号を併記。

(5) 公開済みデータの更新

可能な限り迅速に公開するとともに適時適切な更新。

4. オープンデータの公開・活用を促す仕組み

- (1) オープンデータ・バイ・デザインの推進
- (2) 利用者ニーズの反映

2. オープンデータの定義

- ① 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
- ② 機械判読に適したもの
- ③ 無償で利用できるもの

5. 推進体制

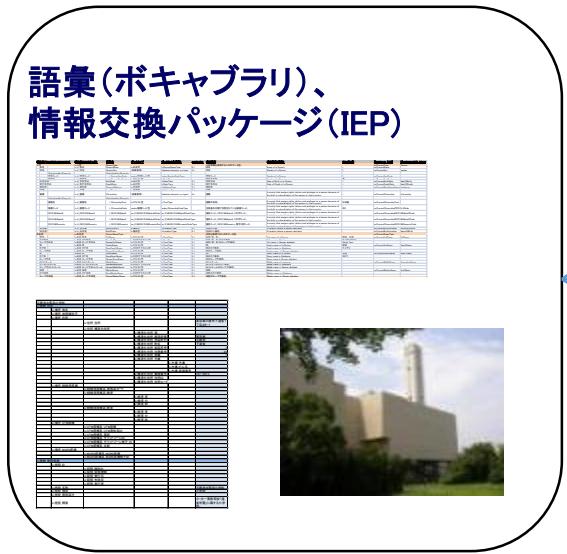
- (1) 相談窓口の設置
- (2) 推進体制

各種戦略や方針の基盤である
データ標準化

コアとなる共通語彙基盤(IMI)

- 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、用語の参考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とする仕組み。



<http://imi.go.jp/>

システム連携



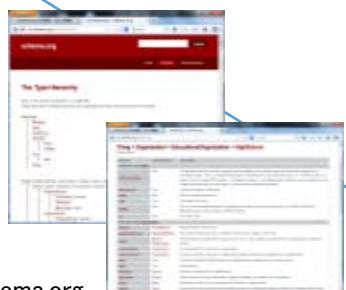
- 情報交換パッケージにより、システム間を連携
- ・高速な情報連携
- ・設計の効率化

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

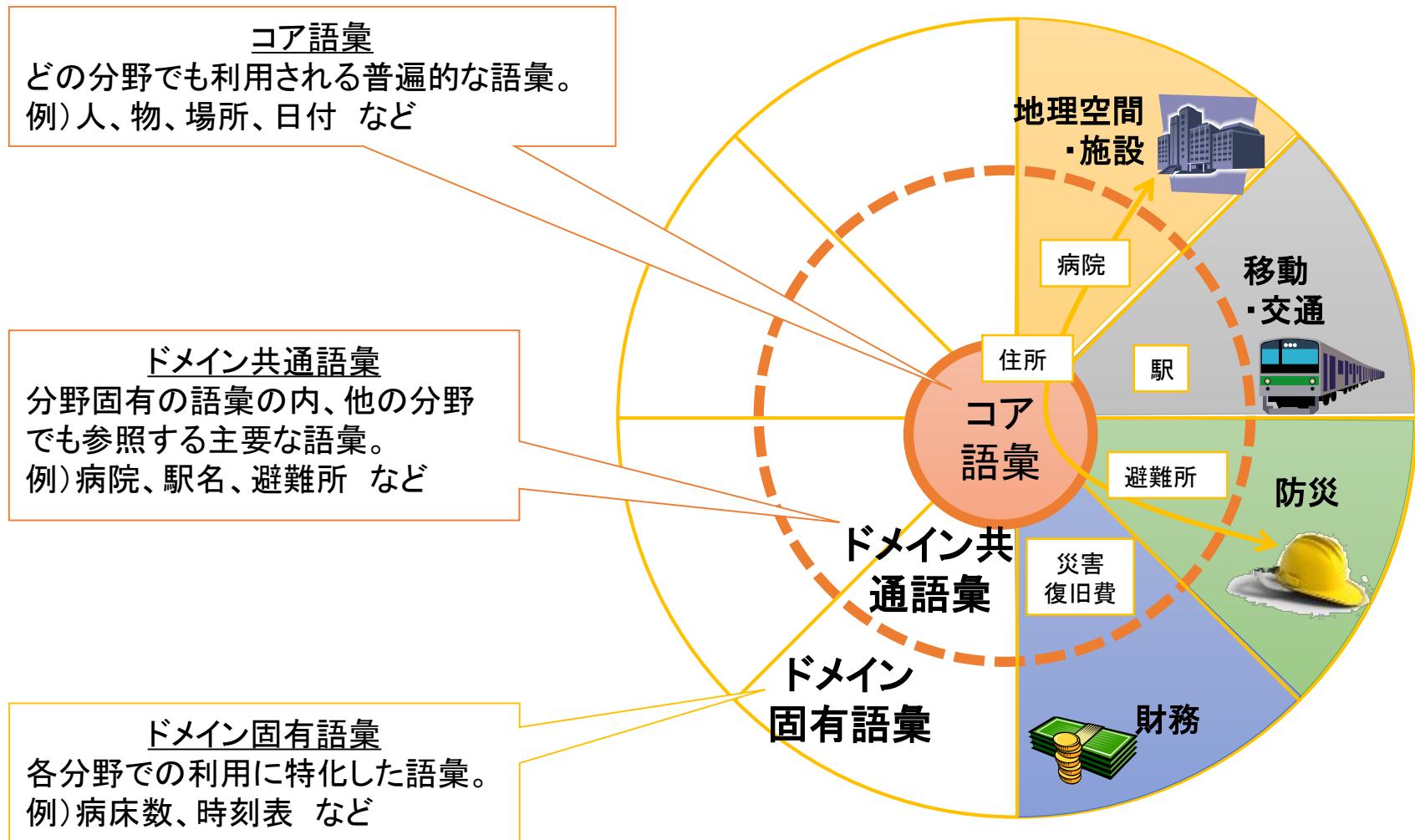


Schema.org
検索エンジン大手が整備する
構造化データマークアップの共通仕様

- 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

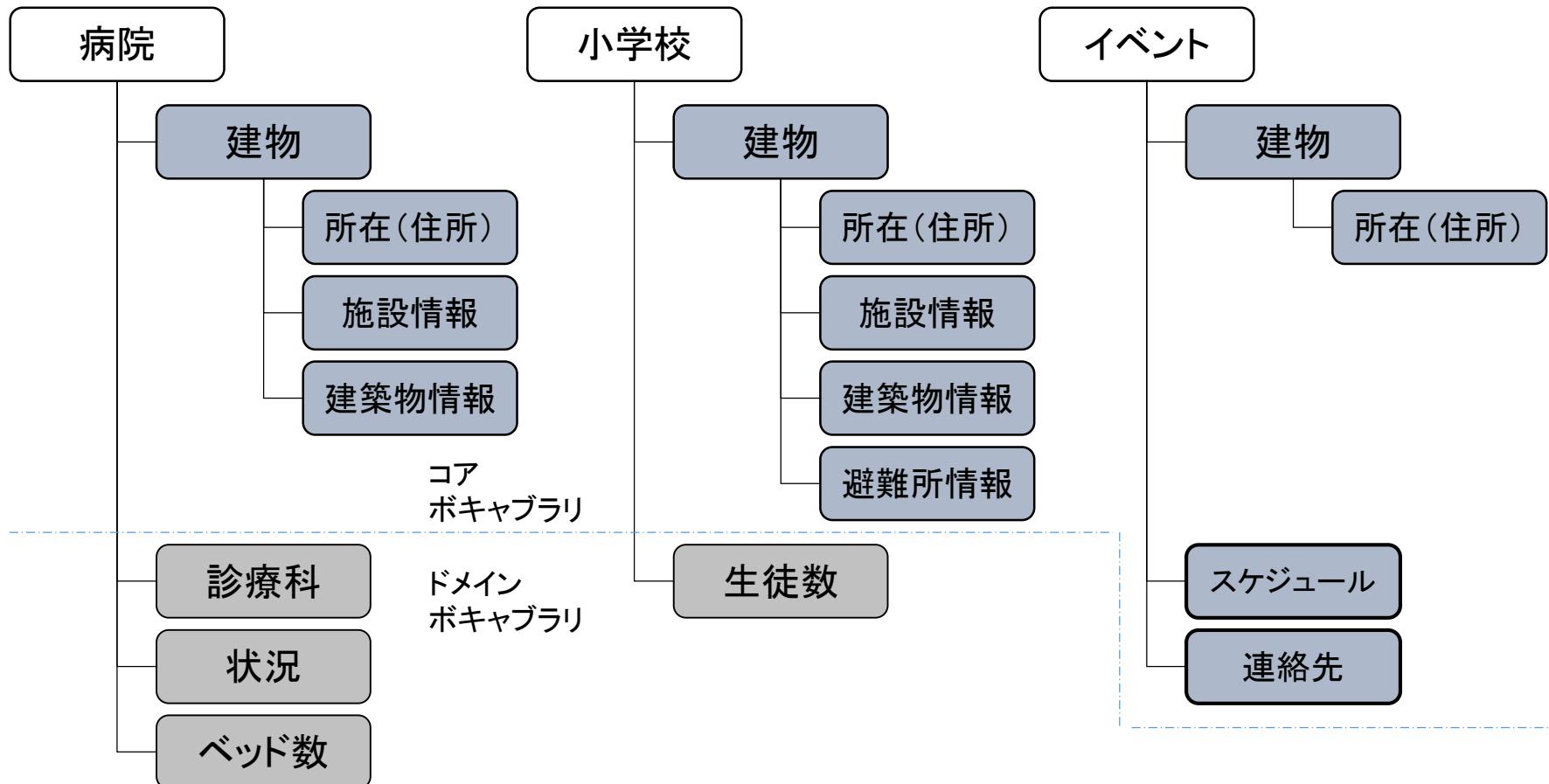
共通語彙基盤の構造

- 共通語彙基盤はコア語彙、共通語彙及びドメイン固有語彙から成り立つ。



情報構造のイメージ

- 施設の情報は、コアのボキャブラリとドメインのボキャブラリの組み合わせで表す。



クラス用語一覧

ID	ID体系	アクセス	アクセス区間	イベント	イベントスケジュール
コード	コードリスト	コード制約	サービス	事物	人
人数	住所	価格	制約	単位コード	参照
名称	土地	地物	場所	定期スケジュール	実体
容量	対象	座標	建物	数量	文書
施設	施設関連	日付	日時	時間	期間
期間スケジュール	期間制約	業務組織	概念	構成員	氏名
法人	活動	測定値	状況	範囲制約	組織
組織関連	製品	製品個品	設備	詳細スケジュール	詳細スケジュール規則
連絡先	重量	金額	長さ	関与	電話番号
面積	駐車場				

● DMD (Data Model Description)

- 法人基本情報 DMD@ja
- 法人活動情報 DMD@ja
- 施設 DMD@ja
- 避難施設 DMD@ja
- 設備 DMD@ja
- 医療機関 DMD@ja
- 氏名 DMD@ja
- イベント DMD@ja
- 住所 DMD@ja
- 組織 DMD@ja
- 地物 DMD@ja

● PD

- PD5474(観光施設に関する語彙の検討)
- PD7706(イベントに関する語彙の検討)
- PD2342(法人情報に関する語彙)
- PD1462(子育て関連施設に関する語彙の検討)

● ドキュメント化待ち

- 調達
- 制度(サービス)

●メリット

➤ユーザ企業、国・自治体

- ✓ 独自データによるロックイン回避
- ✓ 設計費用の低減
- ✓ データ解析の容易化(マッシュアップの容易化)
- ✓ 接続性の向上
- ✓ 段階的導入など拡張性の向上

➤ベンダ企業・開発者

- ✓ データ設計の品質向上
- ✓ データ設計の迅速化
- ✓ 接続性の向上
- ✓ 販路の拡大

●デメリット

➤ユーザ企業、国・自治体

- ✓ 書式変更への現場の理解を得るのが大変な時がある

➤ベンダ企業・開発者

- ✓ 独自仕様によるロックインがしにくくなる

現在どこまでできているのか

1994年12月

「行政情報化基本計画」で
データ標準化の推進を閣議決定

現在どこまでできているのか

準備・検証

- ・動向調査と試行版の作成

2013.2 一次プロジェクト開始
2013.6 IMI1.01

開発・検証

- ・コア語彙の整備

2013.9 二次プロジェクト開始
2014.6 IMI2.0
2014.9 IMI2.1
2015.2 IMI2.2
2015.12 IMI2.3
2017.3 IMI2.4

普及

- ・法人やイベントなどの事例を軸に展開
- ・官民データ

2017.1 法人インフォメーション
2017.1 埼玉県オープンデータポータル
2017.6 こども霞が関見学デー

動き始めたIMI

● 法人インフォメーション

- 法人番号をキーに、政府内の法人関連情報（基本情報、契約情報、表彰情報、届出情報、処分情報等）を一元的に公開
 - ✓ 法人情報データを、共通語彙基盤で整備
 - ✓ 今後、法人マスターデータや申請書の設計に活用→民間に展開

The screenshot shows the homepage of the 法人インフォ (Business Information) website. At the top, there's a search bar with placeholder text '法人番号または法人名を入力'. Below it, a message states: '当サイトの情報について' (Information about this site) and '法人登録より受けた権利金や表彰、登記等の法人活動情報を掲載しています。ただし、既存の業務事業者であっても、業務執行主体が政府と異なる(都道府県、独立行政法人等)事業者の情報は、現在、掲載していません。ご了承ください。' (We publish information such as rights fees, awards, and registrations received from business registration. However, for existing business operators, if the executing entity is different from the government (prefectures, independent administrative agencies, etc.), their information is not published at present. Please understand.)

● こども霞が関見学デー

- 8月に行われる子供向け体験イベントの情報を集約して提供
 - ✓ イベントデータを、共通語彙基盤で整備
 - ✓ 今後、イベントデータ構造を広く公開(現在もPDで公開)

The screenshot shows the 'こども霞が関見学デー' (Children's Chiyoda Government Building Observation Day) website. It features a grid of event cards for various government agencies:

- [土建省] 4月20日 10:00-12:00
- [霞が関見学デー] 6月25日 10:00-12:00
- [公正取引委員会] 6月25日 10:00-12:00
- [国土交通省] 6月25日 10:00-12:00

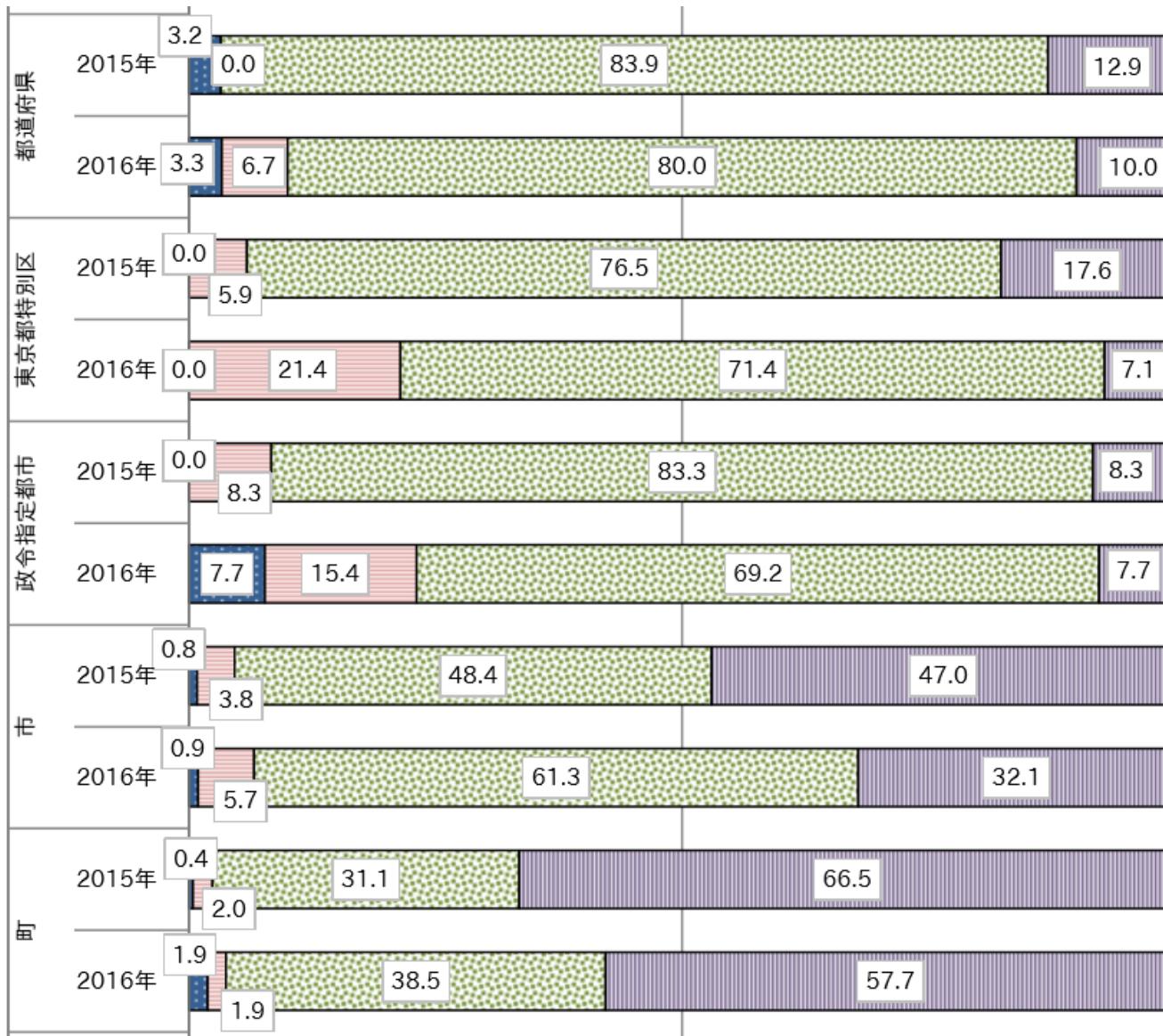
● 埼玉県オープンデータカタログ

- 埼玉県下58自治体のオープンデータのデータ項目や構造を共通化
 - ✓ 10分野を共通語彙基盤で整備
 - ✓ 他自治体でも類似の取り組みが広がる

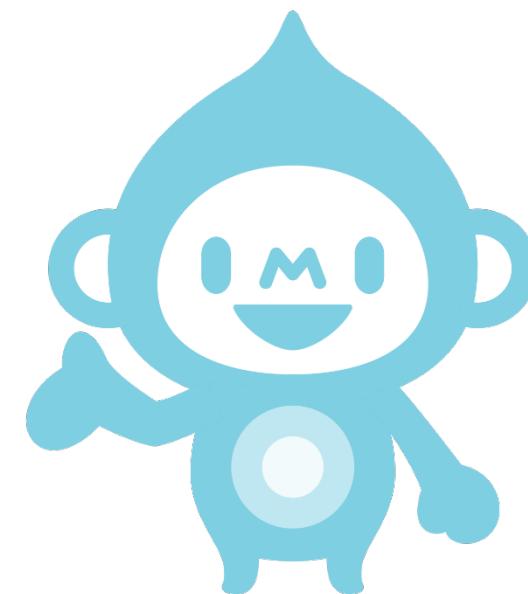
The screenshot shows the 'Open Data Saitama' (埼玉県オープンデータカタログ) website. It displays the following statistics:

- 399 データセット
- 50 組織
- 17 グループ
- 145 タグ

自治体による認知度



- 調達条件や規範として用いている／用いさせている
- 調達の参考にしている／させている
- ▢ 知っている
- ▣ 知らなかった



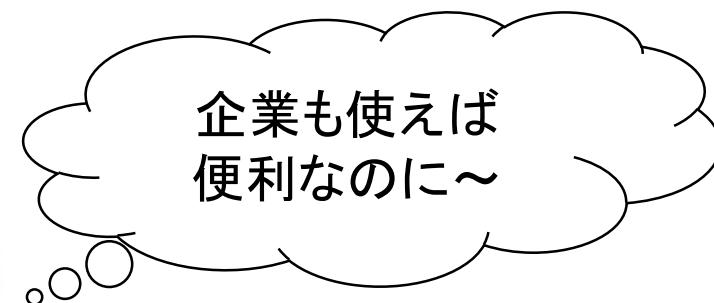
世間での認知度

●wikiにも出ていないし、まだまだ発展途上

共通語彙基盤(キヨウツウゴイキバン)とは - コトバンク

[https://kotobank.jp/word/共通語彙基盤-685831 ▾](https://kotobank.jp/word/共通語彙基盤-685831)

デジタル大辞泉 - 共通語彙基盤の用語解説 - 政府や地方自治体が公開するオープンデータにおいて用いられる語彙集。同じ事柄に対する用語が各機関によって異なる状態では、統計データの集計や情報交換に支障が出るため、共通で使用するための語彙を ...



協業や事業再編時に
独自データは企業にリスク

指摘されている問題

●難しそう

- 簡単に使えるようにDMDやPDなどの活用事例を公開。
- スプレッドシートイメージ(一般的なテンプレート)としての提供。

●ベンダが積極的ではない

- 海外は、パッケージ志向なので、従前から標準化に積極的。
- 国内は、標準化は重要と言うものの、ロックインがあるので消極的。



文字のインター オペラビリティ

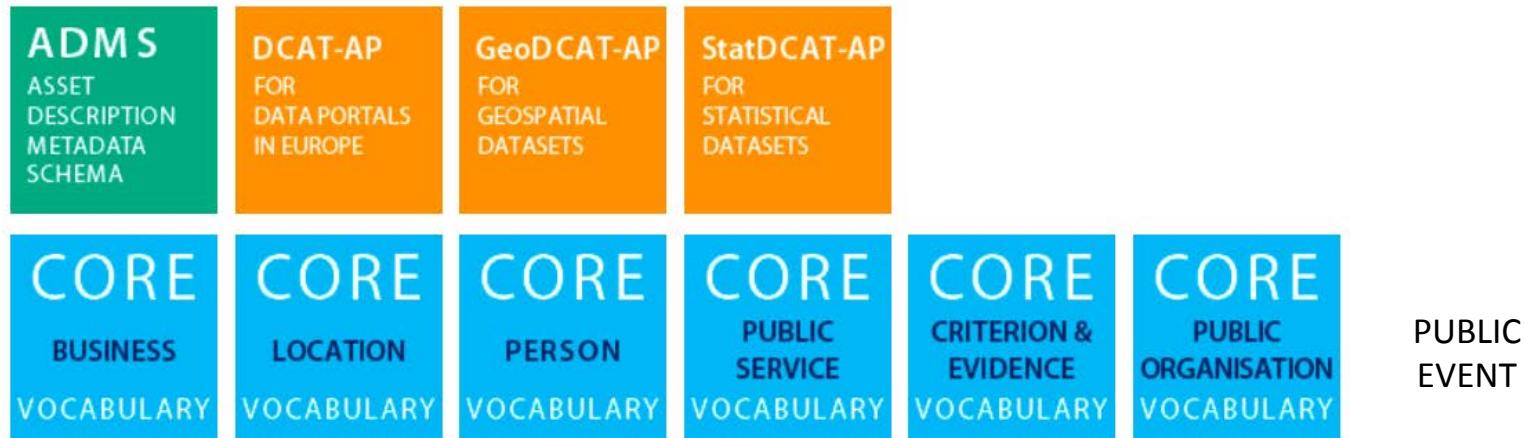
●行政機関で使われる文字は、文字情報基盤で標準化がほぼ完了。

- フリガナやローマ字・英字問題は、番号制度を機に改善中。
- 法人は対応を開始

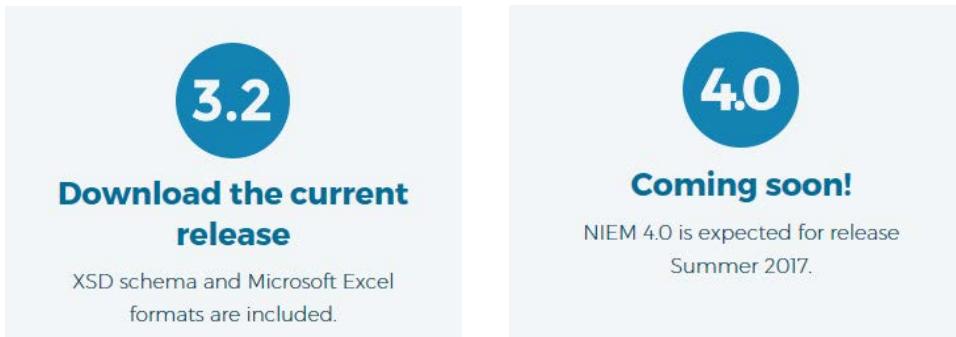
	氏名	法人名
正式名	戸籍統一文字(約6万文字)	登記統一文字(約7万文字)
代替表記名 (PC等での使用を想定)	<ul style="list-style-type: none">•マイナンバー導入時に代替文字を導入•文字の代替ルールは、基本は公的個人認証のルール•JIS第2水準と補助漢字による約1万文字	<ul style="list-style-type: none">•法人番号導入時に文字の縮退(代替)を導入•文字の代替ルールは、IPA縮退マップにより一意•JIS第4水準の約1万文字なので最近の主要な機器で対応可能
フリガナ		2018年度早期に開始予定
ローマ字・英字	総務省自治行政局がマイナンバーカードへのローマ字表記を検討中	2017年4月から国税庁法人番号公表サイトで登録・公表開始 (登記事項にはない(定款内に定義可能))

海外の状況

●ヨーロッパ SEMIC



●米国 NIEM



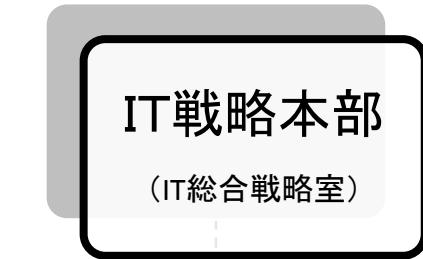
This particular release incorporates approximately 280 new elements and 350 new types, as well as updates to 2,300 existing elements and 230 existing types.

●民間 schema.org



クオリティとサステイナビリティ

推進のための体制



世界最先端IT国家創造宣言
官民データ活用推進基本計画
を推進

協力依頼 ↓ ↑ 報告

情報共有基盤委員会

- 各委員会、WGの設置主体は経済産業省、事務局はIPA
- 各WGの下には、具体的な検討を行なうSWGを設置。

方針の検討
プロセスの検証

共通語彙基盤WG

文字情報基盤WG

利用swg

運用swg

技術swg

語彙、DMDの検証
普及

LOD専門家、XML専門家、
標準化専門家等、
国内トップの陣容

コア語彙の整備方法(トップダウン)

- 行政システムやオープンデータなどのニーズを分解して「氏名」等のターゲットを選定
 - アベイラビリティ以外はほぼ整備
- 既存の様々なデータ表現形態を参照する
- 意味論と抽象化によってモデル化
- スキーマやDMDの整備

- 現実との整合性を保つように表記型等も整備
 - 住所は、市町村名や町名、番地等をすべてバラバラにしたいが、既存の表記が1行で書かれるものがあるため「住所表記」という項目を定義

- 主要な標準やアプリケーションのデータ構造を表で比較
- コア語彙と比較しながらモジュールを組み合わせモデル化
- 導入用のサブセットの整備
- 公開ドラフトとして公開し、意見を募集
- スキーマやDMDの整備
- Referenceモデルなので、利用者は自由に項目を選択し、時には追加することができる
 - IMIの語彙と追加語彙を明確化することでメンテナンス性が向上する

オーサライズに関する考え方

●オーサライズ方式

- 専門家によるWGによる素案→公開→デファクトスタンダード
- コアの用語と実装用モデルを並行して開発
- StandardではなくReference Modelを提供する

➤ 世界各国の進め方もほぼ同じ方式

- ✓ 米国 Core vocabulary +Domain vocabulary +Information Exchange model
- ✓ 歐州 Core vocabulary + Application profile
- ✓ 日本 Core vocabulary + Data Model Description

●理由

- 変化が速い
- 品質はプロセスと体制で担保
- 導入時のカスタマイズ範囲が大きい

サステイナビリティ確保の仕組み

- IT戦略の中で推進

- 中長期に推進が進む

- IMIドメインの取得

- URI変更などが起こると利用者に混乱が起こるため、組織再編などがあっても継続的に使用できる`imi.go.jp`のドメインを取得

- 各種標準モデルの参照

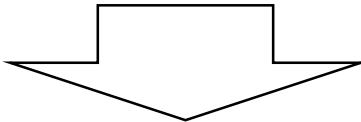
- 他のモデルに乗り換えるにしてもマッピングなどで移行が容易

関連した環境整備

マスターデータ

コード

API



一覧やガイドが必要

The screenshot shows the 'Government CIO Portal' homepage. At the top, there's a banner for the 'World's First IT National Creativity Declaration' and the 'Citizen Data Utilization Promotion Basic Plan'. Below the banner, a section titled 'Code Catalog' is visible, with a sub-section titled 'List of Codes'. A detailed description of the code catalog is provided, mentioning it's a collection of codes released by the government and encouraging users to use them. It also notes that the portal will be updated regularly.

世界最先端 IT 国家創造宣言
官民データ活用推進基本計画

コード一覧

政府機関がインターネットに公開しているコードを、以下に整理しました。
マスター版ではなく、公開しているコード一覧ということをご留意の上お使いください。
政府機関内のコード一覧については、「創造宣言」に詳しい、今後検討をしていく予定です。
政府機関が保有するコード一覧
※ブラウザでファイルを直接開けない場合には、名前を付けて保存したうえで、聞いてください。

特集

アンケート等に便利な基本コード

各種アンケートをした結果を他の調査データと重ねて分析するときがあります。そのため、標準的なコード等を参考して基本的な事項を設定したアンケートにすることによって、収集したデータの活用がしやすくなります。



今後に向けて

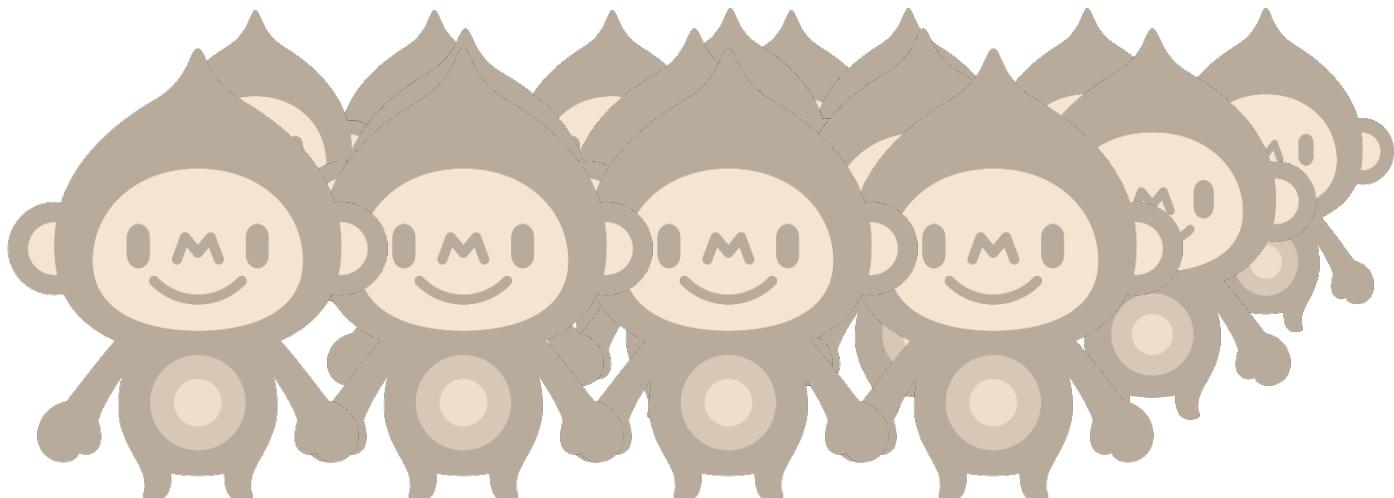
目指すべきビジョンは何か

「官民データ活用推進基本計画」

全ての国民がIT・データの利活用を意識することなく便益を享し、
真に豊かさを実感できる社会

- インタオペラビリティの向上を通じた**社会活動のスピードアップ**
 - 調整時間の削減、正確性の向上（エラー率の減少）
- 情報のValueを最大化
 - = 人、モノ、サービスのValueを最大化

をデータ基盤チームが地味に支える



ロードマップ(イメージ)

FY2017

FY2018

FY2019

FY2020

民間

プロモーション

自治体

都道府県官民データ活用推進基本計画



法人
調達
制度
イベント

マスターデータ一覧
コード一覧
API一覧
データカタログサイト

施設(観光、公園)
制度



イベント

官民データ活用推進基本計画

デジタル・ガバメント実行計画

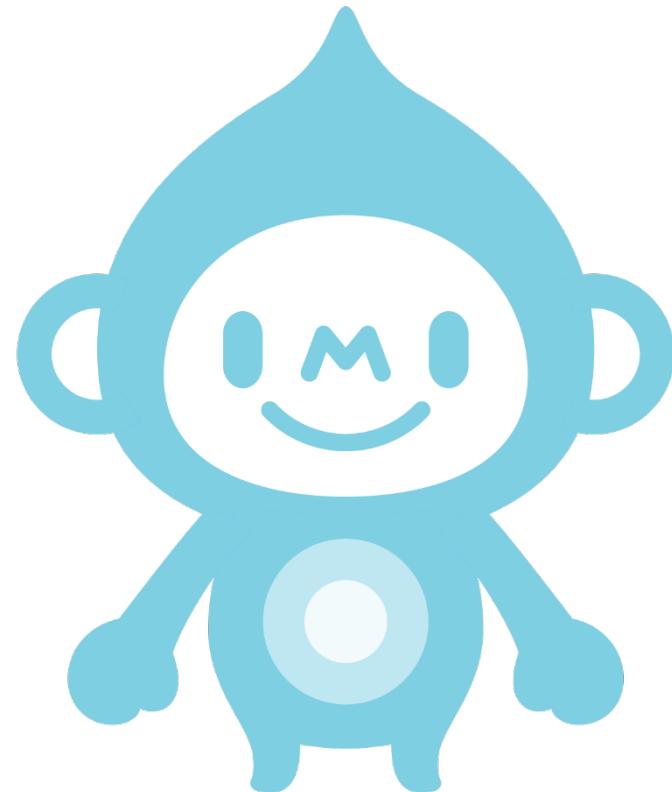
ひな形群の拡大

IMI準拠のマスターデータ、APIの拡大

国際連携の枠組み強化

国

国際



Join us!