

データ連携基盤の取組



IT総合戦略室

デジタル社会に向けての推進方針

世界最先端IT国家創造宣言(IT戦略)[2017-5-30]

- 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[法律:2016-12-14、計画:2017-5-30]

- 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、実行計画[方針:2017-5-30、計画:2018-1-16]

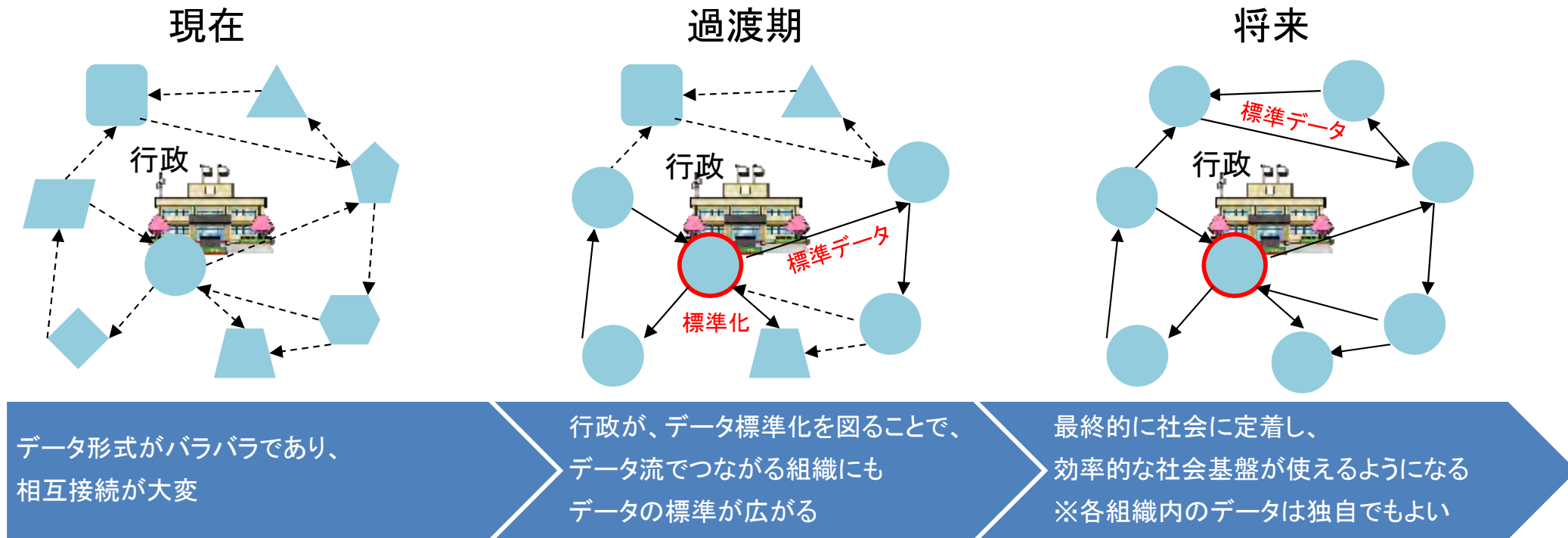
- サービスデザイン
- 官民連携を実現するプラットフォーム
- 文字情報基盤、行政データ連携標準、共通語彙基盤

オープンデータ基本方針[2017-5-30]

- 公開データの形式等

行政におけるデータ標準化の重要性

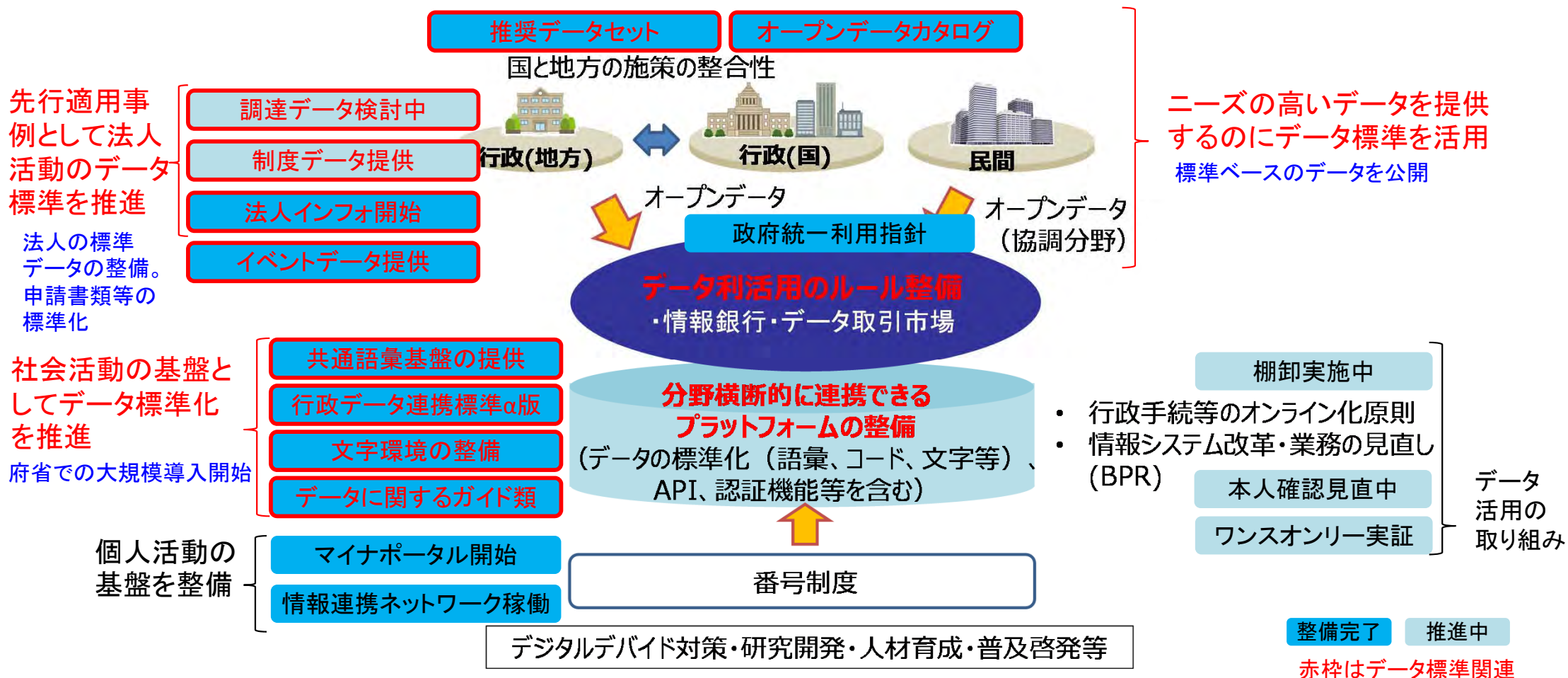
- データ駆動型の社会を作るには、データのインターオペラビリティの確保が重要であり、その核になるのがデータの標準化。
- 申請・届出文書、公開情報の標準化を通じて、データの標準を広げていく。



- オープンデータも標準化することにより、申請系だけではなく、行政データを活用したアプリケーションにも波及させることが可能。

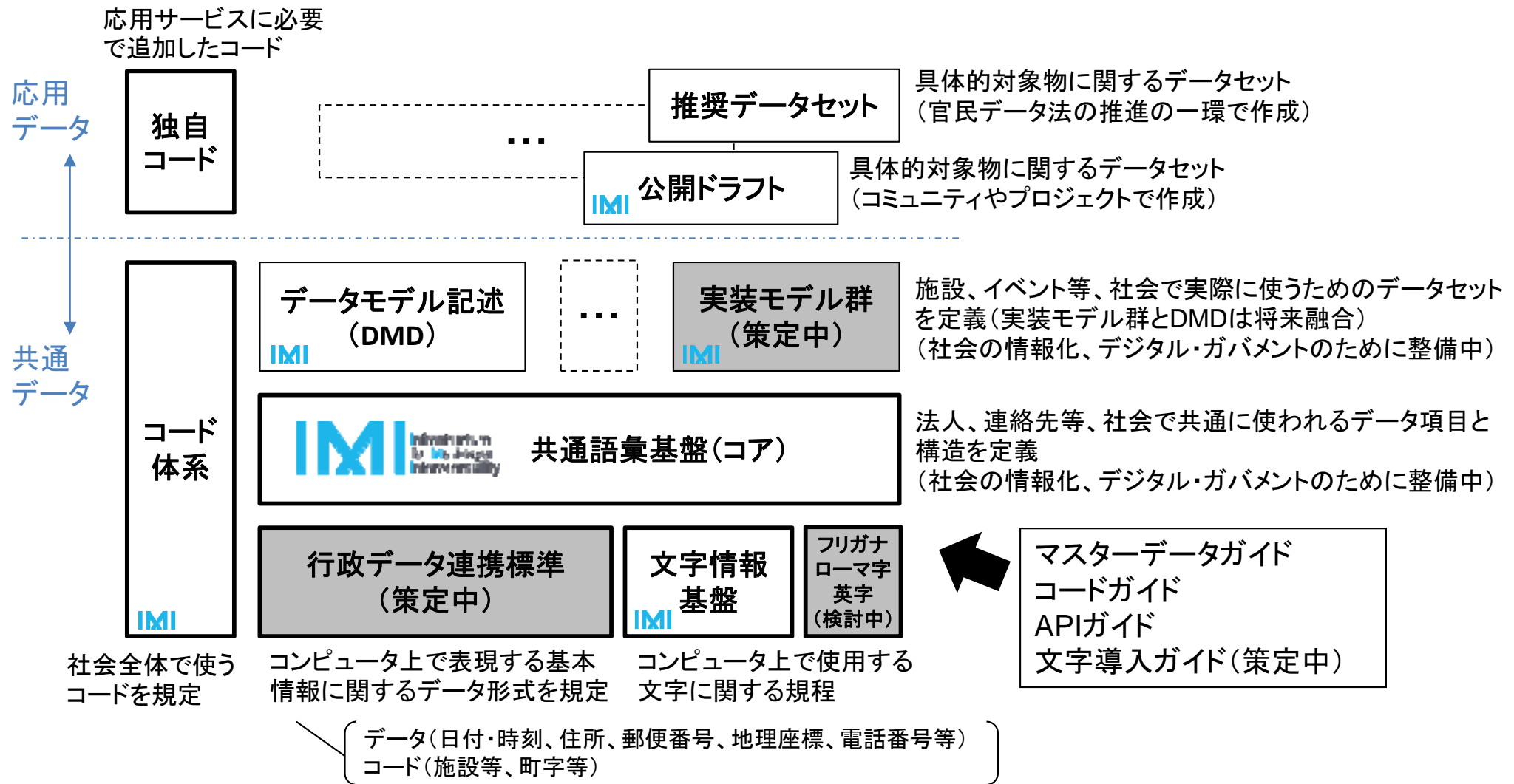
官民データ活用基盤の整備状況

- デジタル100%に向け、分野横断的に横断できるプラットフォームを整備
- デジタル・ガバメントを中核に展開予定。



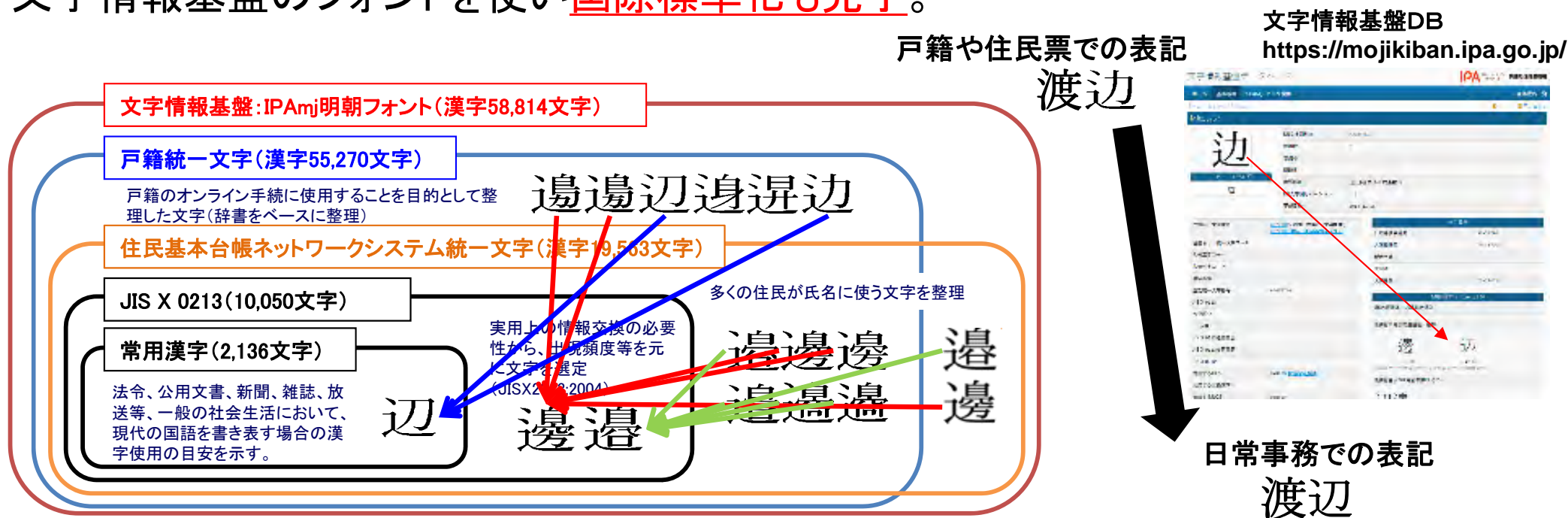
データ体系の全体像

- データ利活用のためには、基本データからの積み上げに基づく体系の中で管理し、相互運用性を確保していく。



文字情報基盤等

- 戸籍システムや住民基本台帳ネットワークの文字を全て網羅し、全ての国民の氏名をコンピュータで扱うことが可能になるよう、無償のフォントを提供。
- 日常生活での利便性を確保するため、これまで外字で表していた戸籍の文字等をJIS第4水準に対応させる縮退マップを提供。(法人番号のシステムでは法人名を縮退マップでJIS第4水準に縮退)
- 行政機関の全ての業務で活用可能であり、外字作成・管理コストの削減、システム間の相互運用性の確保が可能。国や他自治体へのデータ連携等も容易。
- 文字情報基盤のフォントを使い 国際標準化も完了。



※JIS X 0213(JIS4水準)は、最新の情報機器において正確な表記が可能

- 実装を進めるため、代替文字、よみがな、英字の扱いを検討開始するとともに、対象も氏名だけでなく、法人名、地名等に拡大。

行政データ連携標準(仮称)

- ほとんど全てのデータ活用や連携に必要なデータ、コードを対象とする。
 - データ(日付・時刻、住所、郵便番号、座標、電話番号等)
 - コード(PoI(公共施設等)、町字等)

基本原則(案)

以下のような内容を行政データ連携標準の基本原則としてはどうか。

- 本標準は、**日本工業標準及び国際標準に準拠**し、行政のデータ連携のために補完するものである。
- 本標準は、**データ連携のための技術標準**であり、**印刷や画面での表現を制約するものではない**。
- 本標準は、**既存、新規を問わず情報システム内のデータに制約を与えるものではない**。
 - データ連携時に必要な標準であることから、情報システムの新規開発、情報システムの更改時に内部データにも本標準を使うことで相互運用性の高いシステムを構築することが可能である。
 - 組み込みシステム等の機器の制約や処理の高速性が重要なものは独自データ形式のほうが良い場合もある。
- 本標準は、**データがグローバルに流通することを前提**として設計される。
- 本標準は、政府内の所定の手続きを経て制定することとする。
 - デジタル・ガバメント技術会議で審査後、CIO連絡会議で審査の上、決定

検討例(日時)

- 日付
 - 現状: 「2017年11月16日」「2017/11/16」「20171116」「2017-11-16」
 - * データ形式が多様な上、実施日、開館日等と休館日等が混在。
 - 案: 「2017-11-16」
 - * ISO8601拡張形式に統一し、実施日、開館日等の利用可能日を記載。
 - * また、「/」はISO8601で期間データのセパレータとして定義されているので準拠。(例:2017-11-16/2017-11-18)
- 時刻
 - 現状: 「10:00-12:00」「10:00~11:00」「10時-12時」
 - * データ形式が多様な上、注記事項が追記される場合が多い。
 - 案: 開始時間「10:00」、終了時間「12:00」、備考「(自由記述)」
 - * ISO8601形式にするとともに、データを分割して処理しやすくする。
 - * 24時間表記とする。
- 曜日
 - 現状: 「月曜日」「月曜」「月」「月曜-水曜」「月曜~水曜」
 - * データ形式が多様。
 - 案: 漢字の場合「月火水」、コードの場合「123」
 - * データを処理しやすくするとともに、コードも活用。

同じデータ項目でもデータ連携はできない

- データの表記方法が違くとコンピュータは判断できない。



データ項目を定めても表記が揺らぐ



統一ルールが必要



以下もコンピュータでは同じではない

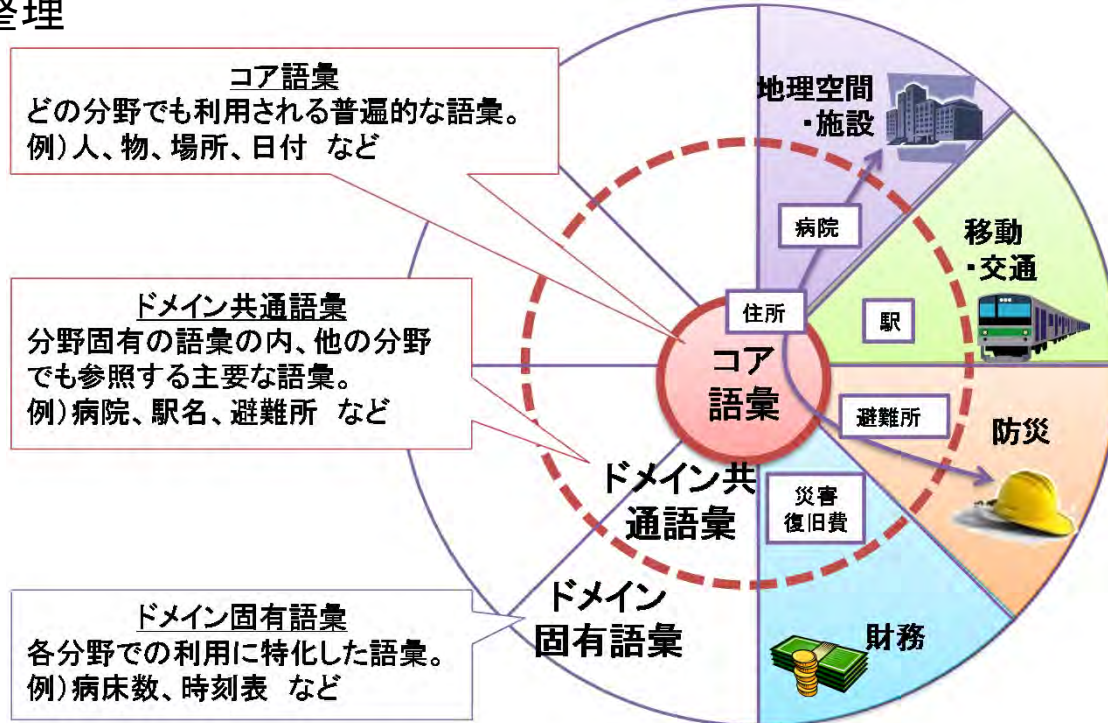
千代田区霞が関2丁目1-2
千代田区霞が関2-1-2
千代田区霞が関2-1-2

変換やAIで解決すればという人もいるが、それ以前にデータをきれいにすればよい。

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

分野横断でのデータ交換を目的としたフレームワーク(世界最先端IT国家創造宣言で推進)

- ・デジタル・ガバメントの基盤として取り組まれるが、将来は社会全体で活用することを前提に設計を実施
- ・米国政府の推進するNIEMと同様に、社会の中核になるコア語彙と分野別の専門分野(ドメイン)語彙を体系的に整理



コア語彙
どの分野でも利用される普遍的な語彙。
例)人、物、場所、日付 など

ドメイン共通語彙
分野固有の語彙の内、他の分野でも参照する主要な語彙。
例)病院、駅名、避難所 など

ドメイン固有語彙
各分野での利用に特化した語彙。
例)病床数、時刻表 など

IMIの特徴

- ・分野横断 (社会基盤のコアな情報を重点推進)
- ・グローバル連携 (EU、米国との情報交換)
- ・IoTへの配慮 (将来的な連携を視野に入れて設計)
- ・オープンデータでの活用 (社会全体のデータ利活用を促進)
- ・検索性向上への配慮 (検索サービス標準の参照)
- ・既存システムへの配慮 (既存データを活かしデータ連携時に活用)

国内の実装状況

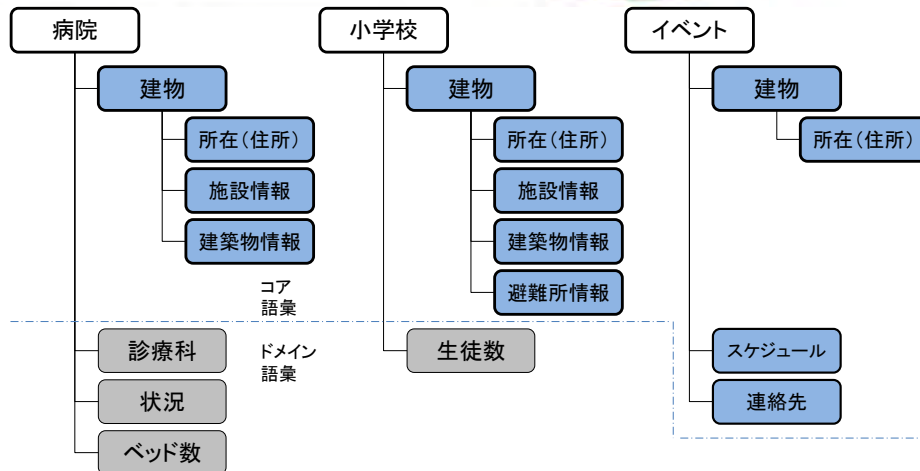
- ・政府機関、自治体、民間で導入が開始
 - ✓ 国が保有する4百万法人の情報を提供する「法人インフォメーション」でIMIを全面採用
 - ✓ 埼玉県では県下市町村を含んだ広域の情報提供に導入
 - ✓ 多くのオープンデータ関連プロジェクトで活用

海外との連携状況

- ・EUのデータ標準プロジェクト総会に毎年参加。(講演等)
- ・米国も2年に一度情報交換。
- ・行政コアデータとオープンデータを同時に進める方式は日本独自の方式で注目されている

体制

- ・創造宣言の中で取り組まれてきたが、官民データ活用推進基本法が整備されたこともあり、重点化。
- ・IT総合戦略室が経済産業省、情報処理推進機構と推進。

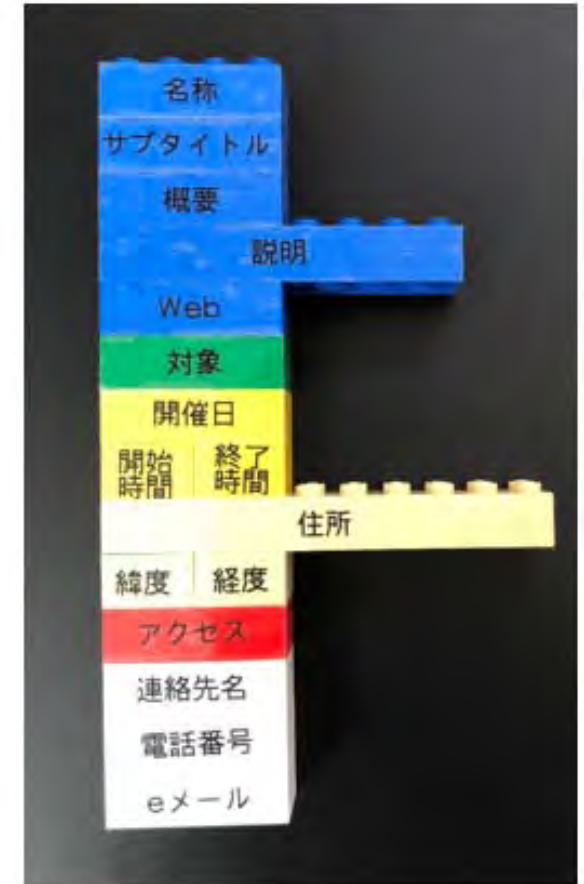
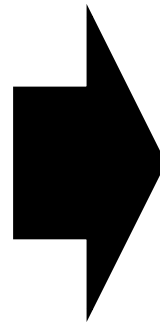


具体的なイメージ

今まで



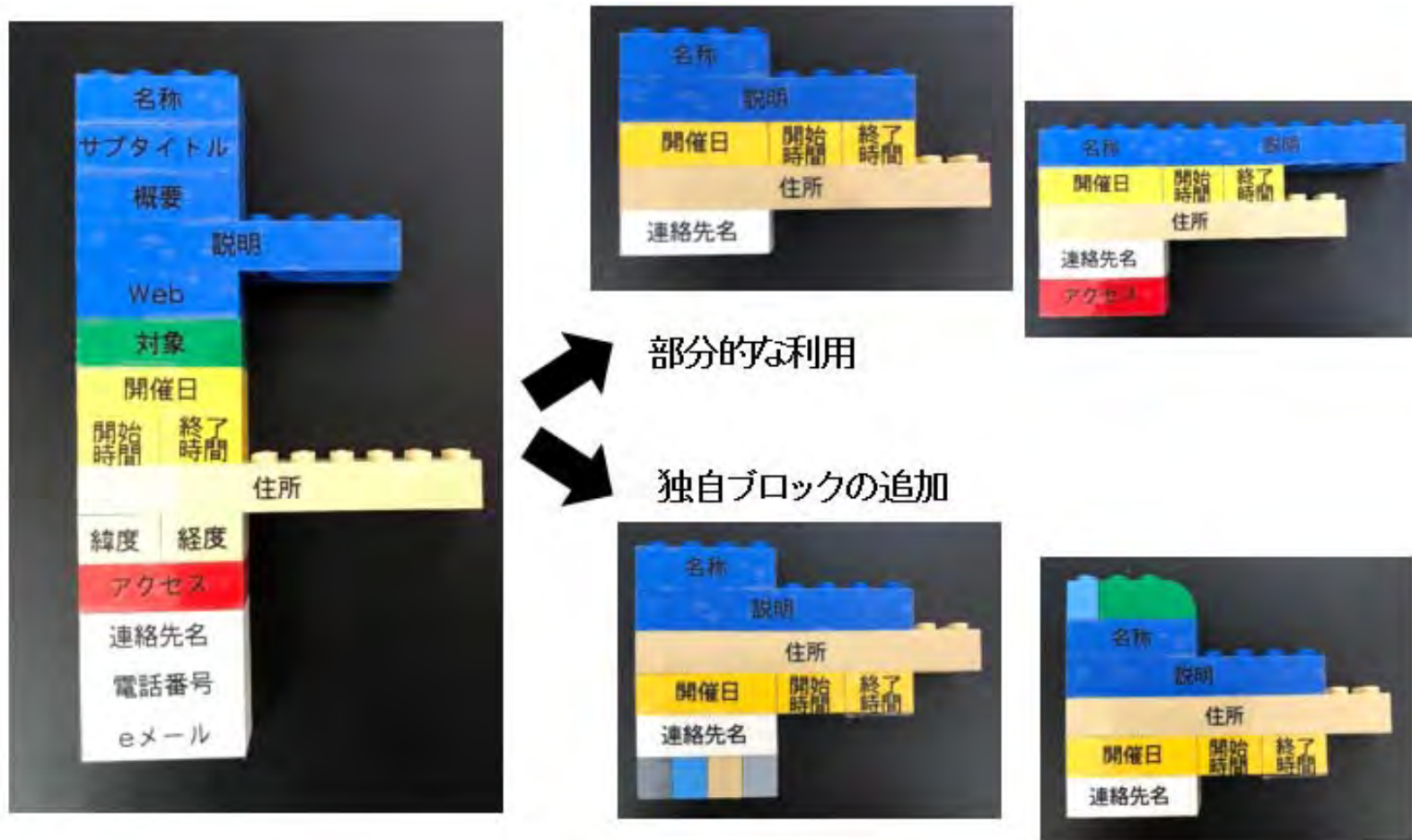
データ連携基盤(共通語彙基盤)



- ・モジュール化することで設計を効率化
- ・インタオペラビリティを確保

利用イメージ

- 目的に合わせて部分利用することで、データ連携が可能



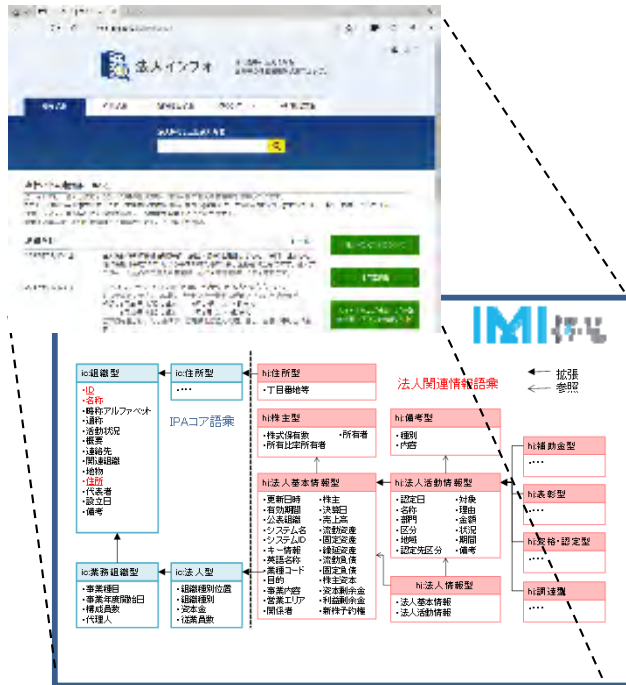
先行活用プロジェクト

■ 法人関連プロジェクト

➤ 法人インフォメーション

- 国内全法人の国との取引や届け出等の情報を公開するサイト
- 共通語彙基盤ベースのデータを活用

法人インフォメーション



■ こども霞が関見学デー

- 夏休みに行く全府省のイベント情報を共通語彙基盤ベースで作成し、検索サイトを整備。
- 従来にない検索性を実現。

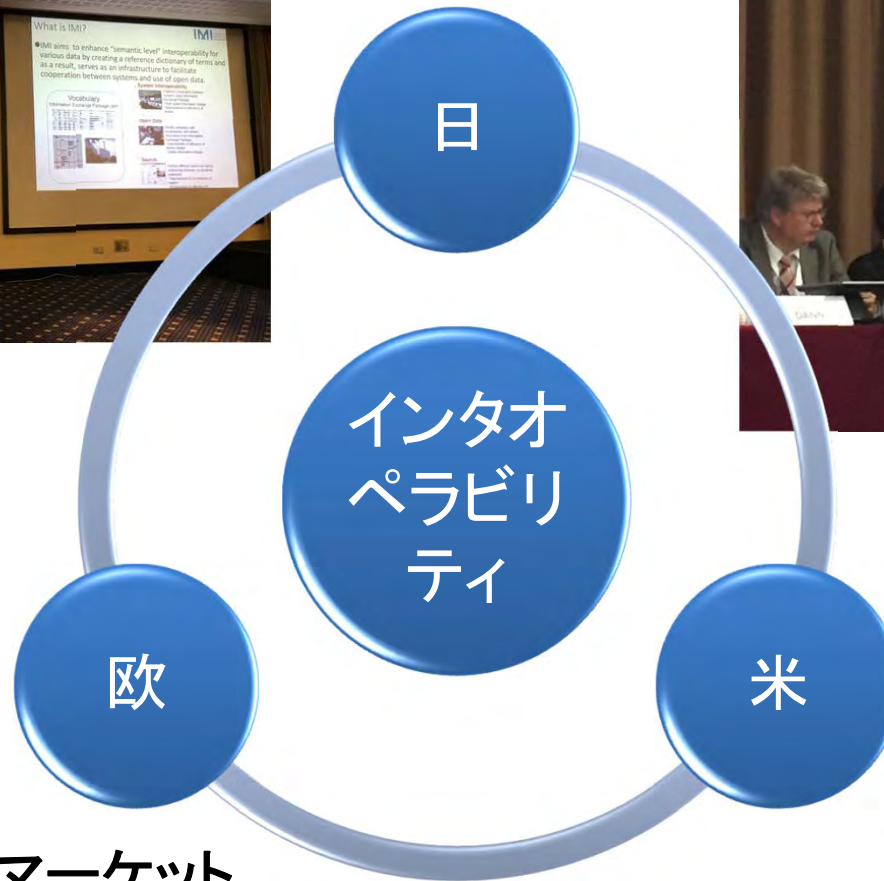


子ども霞が関	IMI2.4	schema.org
都道府県		identifier
主催者	関連組織>役割[主催者]	organizer
共催・協力	関連組織>役割[共催・協力]	contributor
タイトル	名称	name
サブタイトル		alternateName
内容	種別	additionalType
	要約	description
		disambiguatingDescription
サブイベント	関連イベント	subEvent
学ぶ	キーワード	workFeatured
ふれる・感じる	キーワード	workFeatured
体を動かす	キーワード	workFeatured
遊ぶ	キーワード	workFeatured
見る	キーワード	workFeatured
作る・描く	キーワード	workFeatured
収穫・採集する	キーワード	workFeatured
その他	キーワード	workFeatured
集合場所	開催場所>	location
建物	開催場所>	location
住所	開催場所>	location
	座標>緯度	
	座標>経度	
開催種別		
所要時間	数量?	duration
開始日	期間>イベントスケジュール>開始日	startDate
終了日	期間>イベントスケジュール>終了日	endDate
開始時間	期間>イベントスケジュール>開始時間	startDate
終了時間	期間>イベントスケジュール>終了時間	endDate
スケジュール備考	期間>イベントスケジュール>説明	
就学前	利用対象>制約>制約対象[就学前]	typicalAgeRange
小学生低学年	利用対象>制約>制約対象[小学生低学年]	typicalAgeRange
小学生中学年	利用対象>制約>制約対象[小学生中学年]	typicalAgeRange
小学校高学年	利用対象>制約>制約対象[小学生高学年]	typicalAgeRange
中学生	利用対象>制約>制約対象[中学生]	typicalAgeRange
高校生以上	利用対象>制約>制約対象[高校生以上]	typicalAgeRange
対象者備考	利用対象>制約>制約対象[説明]	attendee
募集人数	人数>	maximumAttendeeCapacity
事前登録	?? 独自拡張項目	
申し込み方法	イベント?	ResisterAction>description
申込締切	イベント?	ResisterAction>endTime
申込み・問合せ先	連絡先	ResisterAction>name
		ResisterAction>url
料金種別		isAccessibleForFree
金額	料金>価格>金額	priceSpecification>price
料金備考	料金>価格>備考?	priceSpecification>description

➤ ワンスオンリープロジェクト

- 上記サイトのデータを利用して、過去の提出データを再利用できる申請システムを実証中。
- 共通語彙基盤を使った申請書類を導入

国際的なインターオペラビリティの実現



- デジタル・シングル・マーケット戦略の基盤として重視
- 実装のためのEIRAを提供

- 政府全体で強力に推進
- ベンダも積極的に協力

IoTデータへの展開

- 行政データを中核に進めてきたが、行政でのIoT活用や行政データとIoTデータを組み合わせた活用も増えてきたことを受け、IoTデータの相互運用性についても検討を開始。
 - IoTデータ構造 (SSN: Semantic Sensor Network Ontology等)
 - IoTを活用するためのツール群
 - データ品質
 - 成熟度 等

IoTデータの特徴

- ・精度が必要
- ・誤差が必要
- ・単位情報が必要
- ・測定方法情報が必要
- ・機器情報が必要
- ・リアルタイムデータとバッチデータがある
- ・測定データだけのデータと属性情報を持つデータがある
- 等

Society5.0との連携

■ Society5.0の各分野をつなぐ基盤として官民データの連携基盤を活用。

Society5.0の基盤



官民データの連携基盤

データ連携プラットフォーム

