

共通語彙基盤検討状況について

共通語彙基盤の検討について

平成25年8月

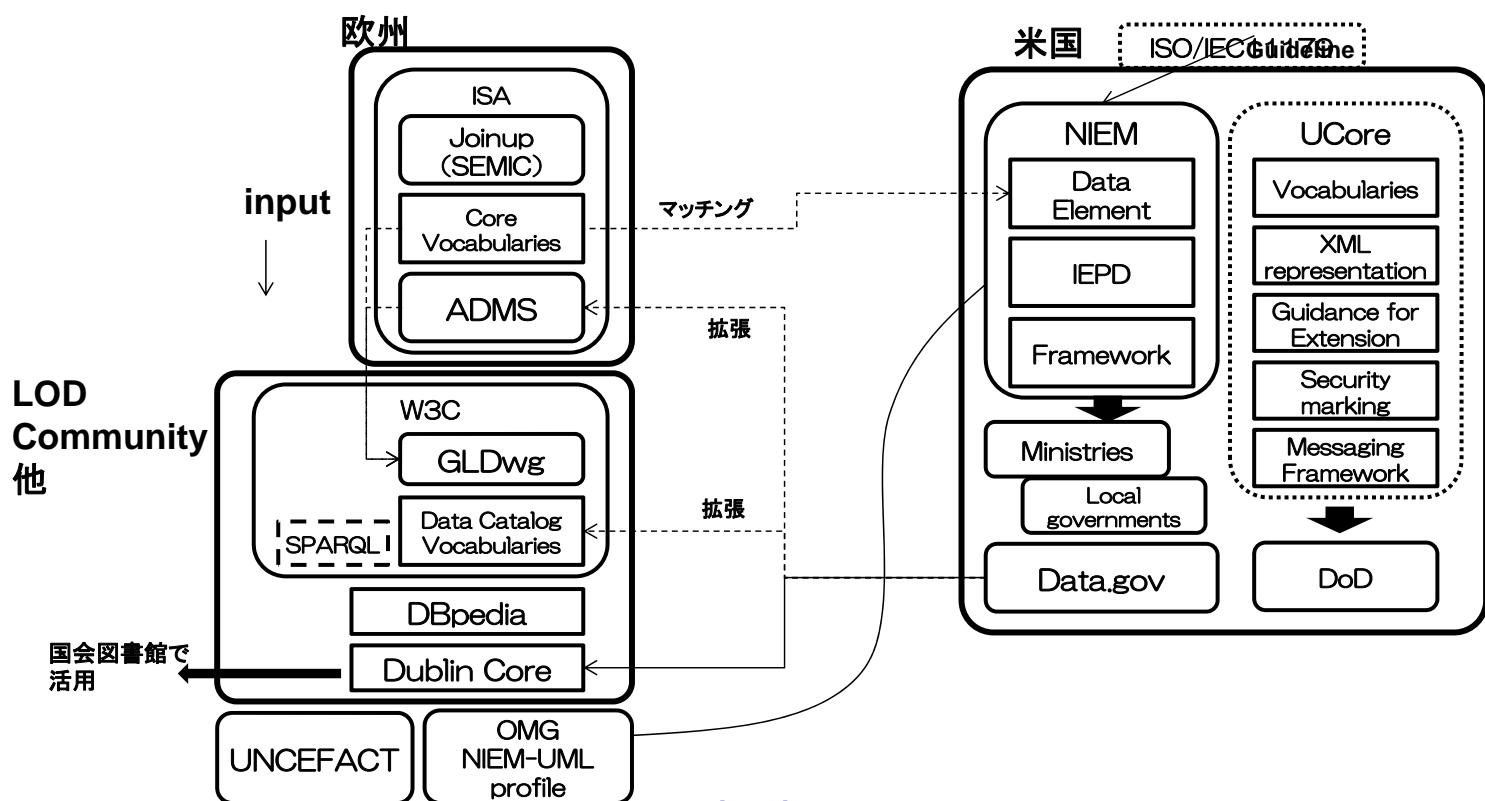
経済産業省

情報プロジェクト室

2

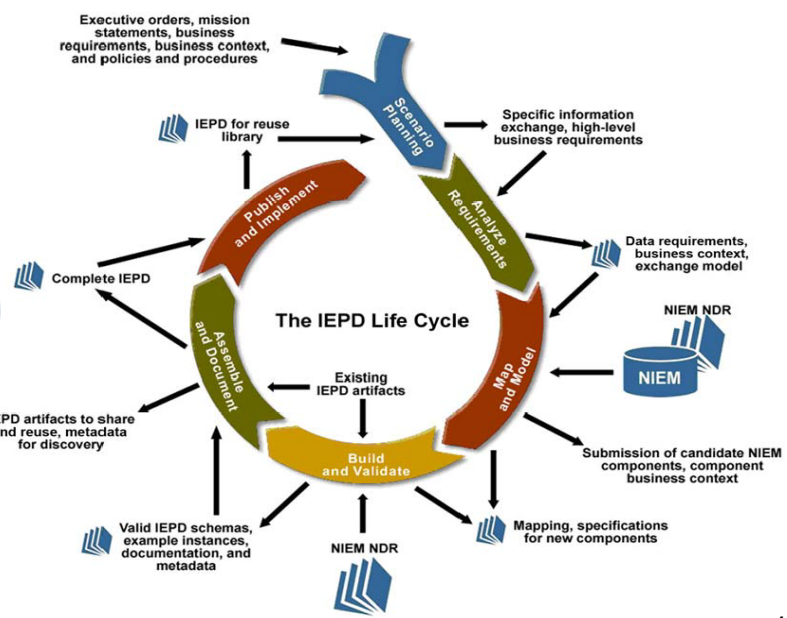
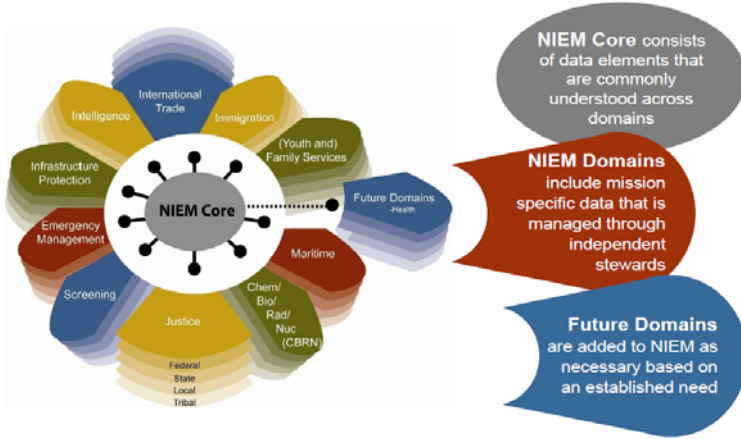
ボキャブラリを巡る世界の取り組み

- 政府機関や標準化団体が積極的に取り組みをおこなっており、各ボキャブラリの連携も始まっている。



■ NIEM (National Information Exchange Model)

- 政府内で情報交換を円滑にするためにボキャブラリの標準化を推進している。全ての分野に共通するコアなボキャブラリと分野別ボキャブラリで構成している。

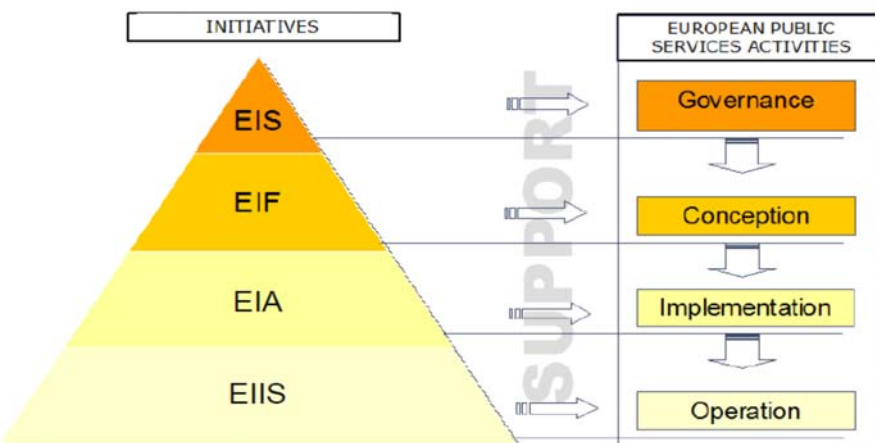


資料:NIEM

ヨーロッパの取り組み

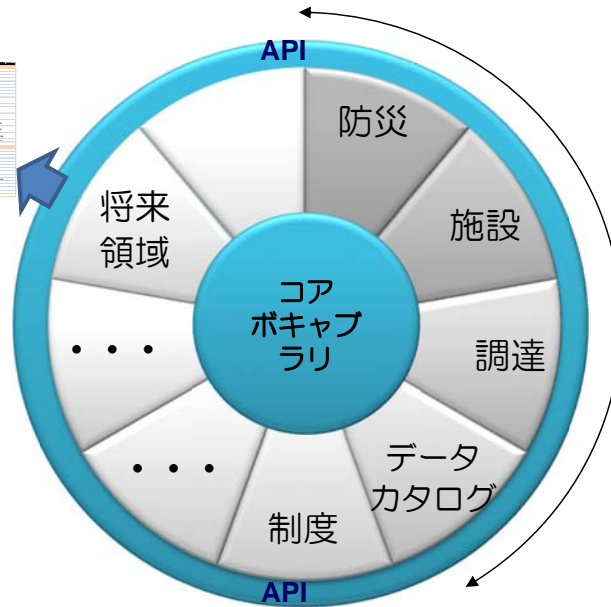
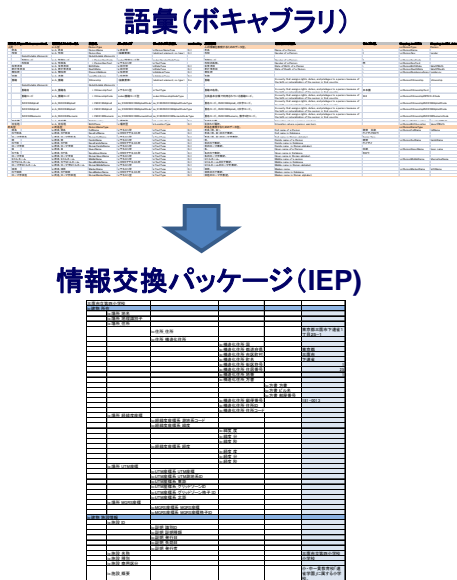
■ EUROPEAN INTEROPERABILITY ARCHITECTURE (EIA)

- EIS、EIF、EIAの全体像は以下の構成になっており、全ての取り組みにセマンティック技術が重要な要素として取り入れられている。
- EIA2.0が公表された2011年11月には、コア・ボキャブラリが完成していなかったため簡単にしか記載されていないが、目指している方向は同じである。



European Interoperability Strategy (EIS)
European Interoperability Framework (EIF)
European Interoperability Infrastructure Services (EIIIS)

- IMI (Infrastructure for Multi-layer Interoperability) とは、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。
- 複数目的、複数組織間でデータを再利用するためには、このような仕組みが必須。



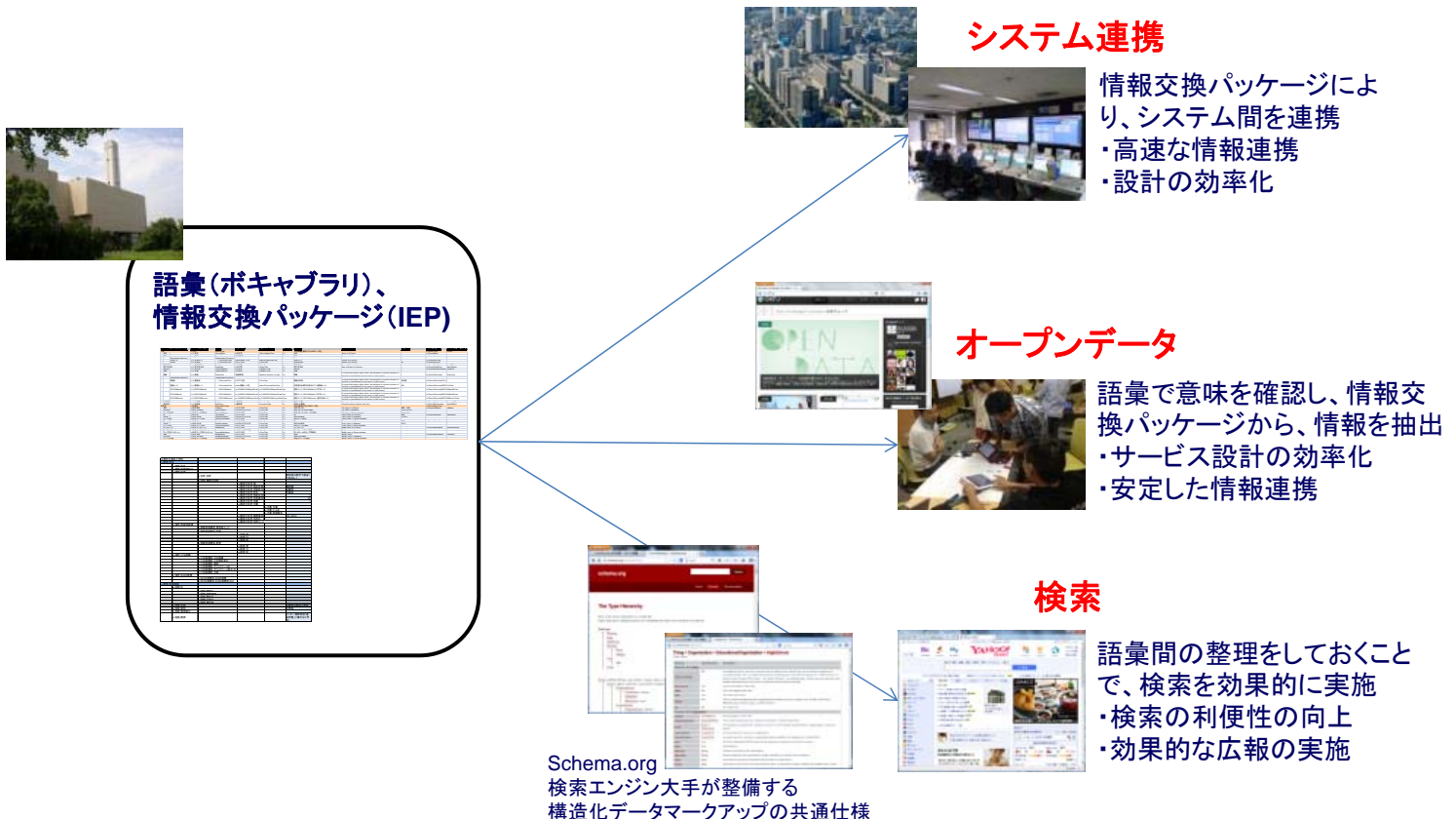
分野横断で情報交換するには、

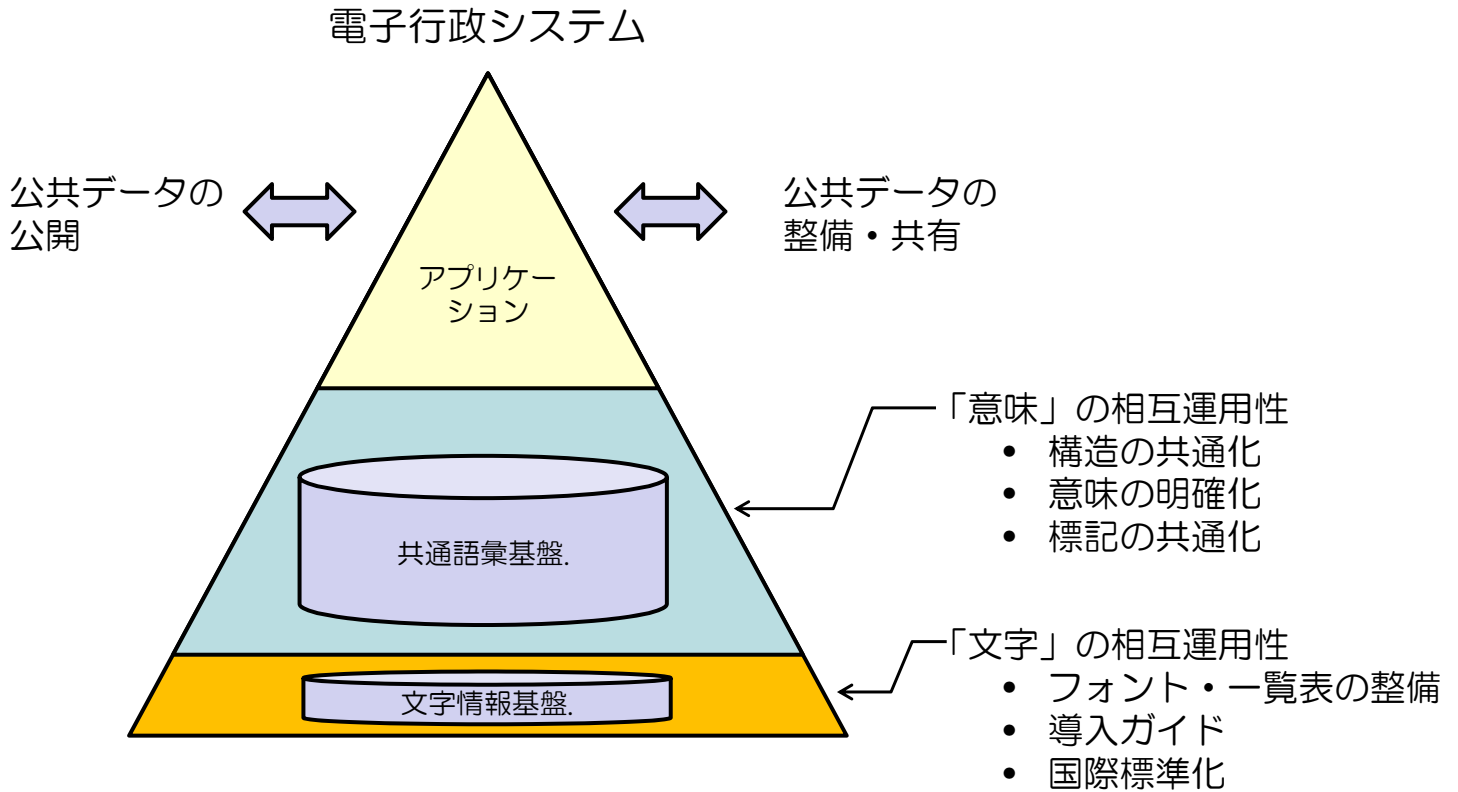
- ・基本的なフォーマットの統一
- ・異なる言葉で表現していても、同じことであれば同じ意味としてとらえること
- ・同じ言葉でも違う意味の言葉の明確化

が必要。

共通語彙基盤 (IMI) の利用

- データが構造化されているので、利活用が容易になる。





これまでの成果

■ 昨年度委員会の成果

- コアボキャブラリの整理
 - IMIコアボキャブラリ1.00を整備
- 防災（物資）ボキャブラリの整理
 - IMI防災（物資）ボキャブラリ1.00を整備
- ルール案の整備（名前付けルール（NDR）、情報交換パッケージ（IEP）、メンテナンスルール）

※昨年の検討当初は、社会保険・税も検討候補にしていたが、コアと防災に集中するため実施を見送り。

■ 追加検証

- コアボキャブラリの精査
 - IMIコアボキャブラリ1.01を整備
 - Cardinalityの変更
 - 語彙の小修正（内線の追加、イベントの削除等）
 - 項目説明の追加
- 適用事例の検討
 - 制度、地物（総務省実証実験）、OpendataMETIのカタログ情報、施設

- 検討期間が短かった
 - 今年度は、プロジェクトを早期にスタートし、検討時間を確保
 - 事務局の準備作業も、実践的検証を行うなど、十分に実施
- 海外動向整理などを情報整理と検討を同時並行で行う必要があった
 - 基礎調査が終わっているため、差分の調査を事前に実施
 - NIEM3.0、SEMIC2013、Schema.orgの調査
- コアのボキャブラリ等の基盤がなく、模索しながら推進せざるを得なかった
 - 昨年度に整備したボキャブラリなどがあるため、検証しながら推進が可能
- 机上の検討であり、実証までできなかった
 - 今年度は実証を予定
- 経験者がいなかった
 - 去年のプロジェクトの経験者がいる
 - 成果等を見ることで、キャッチアップが簡単にできる
- 成果の出口が努力目標であった
 - IT戦略に位置づけ、成果の出口を明確化

10

世界最先端IT 国家創造宣言（平成25年6月14日閣議決定）

I. 基本理念

II. 目指すべき社会・姿

III. 目指すべき社会・姿を実現するための取組

1. 革新的な新産業・新サービスの創出と全産業の成長を促進する社会の実現

(1) オープンデータ・ビッグデータの活用の推進

- (2) ITを活用した日本の農業・周辺産業の高度化・知識産業化と国際展開
- (3) 幅広い分野にまたがるオープンイノベーションの推進等
- (4) IT・データを活用した地域（離島を含む。）の活性化
- (5) 次世代放送サービスの実現による映像産業分野の新事業創出、国際競争力の強化

2. 健康で安心して快適に生活できる、世界一安全で災害に強い社会

- (1) 適切な地域医療・介護等の提供、健康増進等を通じた健康長寿社会の実現
- (2) 世界一安全で災害に強い社会の実現
- (3) 家庭や地域における効率的・安定的なエネルギーマネジメントの実現
- (4) 世界で最も安全で環境にやさしく経済的な道路交通社会の実現
- (5) 雇用形態の多様化とワーク・ライフ・バランス（「仕事と生活の調和」）の実現

3. 公共サービスがワンストップで誰でもどこでもいつでも受けられる社会の実現

(1) 利便性の高い電子行政サービスの提供

- (2) 国・地方を通じた行政情報システムの改革
- (3) 政府におけるITガバナンスの強化

IV. 利活用の裾野拡大を推進するための基盤の強化

V. 戦略の推進体制・推進方策

このため、速やかに電子行政オープンデータ戦略に基づくロードマップを策定・公表するほか、2013年度から、公共データの自由な二次利用を認める利用ルールの見直しを行うとともに、機械判読に適した国際標準データ形式での公開の拡大に取り組む。また、各府省が公開する公共データの案内・横断的検索を可能とするデータカタログサイトについて、2013年度中に試行版を立ち上げ、広く国民の意見募集を行うとともに、2014年度から本格運用を実施する。併せて、データの組み合わせや横断的利用を容易とする**共通の語彙(ボキャブラリ)の基盤構築**にも取り組む。

従来政府が担っていたサービスの提供機能を民間にも開放し、官民の協働によって、より利便性の高い公共サービスを創造する。国民がステークホルダーとして積極的に参加できるよう、このためのクラウドを活用したオープンな利用環境を、**データ・フォーマット、用語、コード、文字等の標準化・共通化**、また、アプリケーション・インターフェイス(API)の公開等を行いつつ整備する。特に文字に関しては、今後整備する情報システムにおいては、国際標準に適合した文字情報基盤を活用することを原則とする。

共通語彙基盤 (IMI) モデル検証例 (施設ドメインα)

語彙のイメージ

■ 用語が構造化しており、再利用しやすい構造になっています。

小学校	im:小学校	ElementarySchool	ic:建物型	BuildingType
-----	--------	------------------	--------	--------------

建物型	ic:建物型	BuildingType
所在	ic:建物_所在	BuildingLocation
施設情報	ic:建物_施設情報	BuildingFacilityInformation
建物情報	ic:建物_建物情報	BuildingBuildingInformation
避難所情報	ic:建物_避難所情報	BuildingRefugeInformation
公共設備情報	ic:建物_公共設備情報	BuildingPublicFacilityInformation
連絡先	ic:建物_連絡先情報	BuildingContactInformation

住所型	ic:住所型	AddressType	施設型	ic:施設型	FacilityType
住所表記	ic:住所_住所表記	AddressRepresentation	施設ID	ic:施設 ID	FacilityIdentification
<i>Substitutable Elements:</i>		<i>Substitutable Elements:</i>	名称	ic:施設 名称	FacilityName
住所	ic:住所_住所	+ FullAddress	種別	ic:施設 種別	FacilityCategoryText
構造化住所	ic:住所_構造化住所	+ StructuredAddress	商用区分	ic:施設 商用区分	FacilityCommercialIndicat
構造化住所型	ic:構造化住所型	StructuredAddressType	概要	ic:施設 概要	FacilityDescription
国	ic:構造化住所_国	AdminUnitL1	利用時間	ic:施設 利用時間	FacilityScheduleDay
都道府県	ic:構造化住所_都道府県	AdminUnitL2	料金	ic:施設 料金	FacilityAdmissionFee
市区町村	ic:構造化住所_市区町村	PostName	収容人数	ic:施設 収容人数	FacilityCapacity
町名	ic:構造化住所_町名	AddressArea	駐車場	ic:施設 駐車場	FacilityParking
街区符号	ic:構造化住所_街区符号	Thoroughfare	保有設備	ic:施設 保有設備	FacilityContainsItem
住居番号	ic:構造化住所_住居番号	HouseNumber	アクセス	ic:施設 アクセス	FacilityAccess
地番	ic:構造化住所_地番	LotNumber	リファレンス	ic:施設 リファレンス	FacilityReference
方書	ic:構造化住所_方書	Locator			
郵便番号	ic:構造化住所_郵便番号	PostCode			
住所ID	ic:構造化住所_住所ID	AddressID			
<i>Substitutable Elements:</i>		<i>Substitutable Elements:</i>			
住所コード	ic:構造化住所_住所コード	+ AddressCode			

一見すると複雑な仕組みに見えますが、自動生成される項目や未使用の項目も多いため、これまでのデータ取り纏めより過大な負担をかけることはありません。

語彙のイメージ(三鷹の森ジブリ美術館)

三鷹の森ジブリ美術館				
ic:建物_所在				東京都三鷹市下連雀 1-1-83
ic:住所_住所				
ic:住所_構造化住所				
	ic:構造化住所_国			
	ic:構造化住所_都道府県			東京都
	ic:構造化住所_市区町村			三鷹市
	ic:構造化住所_街区符号			
	ic:構造化住所_住居番号			1
	ic:構造化住所_地番			83
	ic:構造化住所_方書			
		ic:方書_方書		
		ic:方書_ビル名		
		ic:方書_部屋番号		
		ic:構造化住所_郵便番号		181-0013
ic:場所_経緯度座標				
	ic:経緯度座標系_測地系コード			北緯
	ic:経緯度座標系_緯度			35
		ic:緯度_度		41
		ic:緯度_分		46.44
		ic:緯度_秒		
	ic:経緯度座標系_経度			東経
		ic:経度_度		139
		ic:経度_分		34
		ic:経度_秒		13.55
ic:建物_施設情報				
	ic:施設_ID			
	ic:施設_名称			三鷹の森ジブリ美術館
	ic:施設_種別			美術館
	ic:施設_商用区分			商用
	ic:施設_概要			
	ic:施設_利用時間			ジブリに関連した展示を行う三鷹市立のアニメーション美術館
	ic:オープン日			月、水、木、金、土、日
	ic:開始時間			10:00
	ic:終了時間			18:00
	ic:施設_料金			大人・大学生 1,000円 高校・中学生 700円 小学生 400円 幼児(4歳以上) 100円
	ic:施設_収容人数			
	ic:施設_駐車場	ic:数量_単位		0
	ic:施設_保有設備			
	ic:施設_アクセス			三鷹駅南口から徒歩15分。コミュニティバス約9分
	ic:施設_リファレンス			当美術館は、日時指定の予約制です。
ic:建物_建物情報				
	ic:建物構造_敷地面積			
	ic:建物構造_主要用途			
	ic:建物構造_建築面積			
	ic:建物構造_延べ面積			
	ic:建物構造_最高の高さ			
	ic:建物構造_地上階数			2
	ic:建物構造_地下階数			1
	ic:建物構造_構造			RC
	ic:建物構造_竣工日			2001-06-30
ic:建物_連絡先情報				
	ic:連絡先_名称			三鷹の森ジブリ美術館
	ic:連絡先_メールアドレス			museum@ghibli.jp
	ic:連絡先_住所			東京都三鷹市下連雀 1-1-83
	ic:連絡先_電話番号			0570-055777
	ic:連絡先_FAX番号			
	ic:連絡先_URL			http://www.ghibli-museum.jp/

■ 施設等の情報は、このように記載することとなるので、機械判読性が非常に高くなります。

料金やアクセスをさらに細分化できます

アクセス型	im:アクセス型	im:Access_Type		
最寄駅	im:最寄駅	NearestStation	im:駅型	im:Station_Type
バス	im:アクセス_バス	AccessBus	im:バス停型	im:BusStop_Type
徒歩時間	im:アクセス_徒歩時間	AccessWalkTime	ic:測定単位型	ic:Measure_Type
総時間	im:アクセス_総時間	AccessTotalTime	ic:測定単位型	ic:Measure_Type
タクシー所要時間	im:アクセス_タクシー所要時間	AccessTaxiTime	ic:測定単位型	ic:Measure_Type
料金型	im:料金型	im:Fee_Type		
大人	im:大人料金	FeeAdult	ic:測定単位型	ic:Measure_Type
子供	im:子供料金	Child	ic:測定単位型	ic:Measure_Type
高齢者	im:高齢者料金	FeeSenior	ic:測定単位型	ic:Measure_Type

語彙のイメージ(バス停)

■ バス停等、住所よりも細かく位置を指定する場合には、住所ではなく緯度経度を使う場合もあります。

➤ 住所と緯度経度はツールでの変換も可能です。

バス停				
im:バス停id				
im:バス停_名称				公園入口
im:バス停_所在				
	ic:場所_経緯度座標			
		ic:経緯度座標系_測地系コード		
		ic:経緯度座標系_緯度		35.701967
			ic:緯度_度	
			ic:緯度_分	
			ic:緯度_秒	
		ic:経緯度座標系_経度		139.576422
			ic:経度_度	
			ic:経度_分	
			ic:経度_秒	
im:バス路線id				
im:バス路線名				
im:バス停_方向				明星学園前
ic:バス停_連絡先情報				
	ic:連絡先_名称			小田急バス
	ic:連絡先_メールアドレス			
	ic:連絡先_住所			
	ic:連絡先_電話番号			0422-22-9009
	ic:連絡先_FAX番号			
	ic:連絡先_URL			http://www.odakyubus-navi.com/

語彙のイメージ(イベント)

- 祭りなどのイベント情報も、同様の方式で管理することが可能です。

秋のフリーマーケット			
ic:活動 活動名			秋のフリーマーケット
ic:活動 識別子			
ic:活動 分類			催事・イベント
ic:活動 所在			
ic:場所 地名			三鷹市リサイクル市民工房
ic:場所 地理識別子			
ic:場所 住所			
ic:住所 住所			東京都三鷹市深大寺二丁目16番13号
ic:住所 構造化住所			
ic:構造化住所 国			
ic:構造化住所 都道府県			東京都
ic:構造化住所 市区町村			三鷹市
ic:構造化住所 町名			深大寺
ic:構造化住所 街区符号			2
ic:構造化住所 住居番号			16
ic:構造化住所 地番			13
ic:構造化住所 方書			
ic:方書 方書			
ic:方書 ビル名			
ic:方書 部屋番号			
ic:構造化住所 郵便番号			
ic:活動 期間			
ic:期間 開始日			2013-10-05
ic:期間 終了日			2013-10-05
ic:活動 時間			
ic:開始時間			10:00
ic:終了時間			14:00
ic:活動 説明			恒例のフリーマーケットです！ご家庭に眠っている不要品などを出品しませんか。
ic:活動 料金			
ic:活動 活動理由			
ic:活動 活動状況			
ic:活動 活動結果			
ic:建物 連絡先情報			
ic:連絡先 名称			三鷹市ごみ対策課フリーマーケット担当
ic:連絡先 メールアドレス			
ic:連絡先 住所			三鷹市野崎一丁目1番1号
ic:連絡先 電話番号			0422-45-1151(内線:2533~2536)
ic:連絡先 FAX番号			0422-47-5196
ic:連絡先 URL			http://www.city.mita.katokyo.jp/e_event/040/040664.html

なぜ、分野別専用語彙では不十分なのか

- 国立天文台は、研究施設であり、文化施設、観光施設であり、災害対応施設です。
- 各部門で、登録フォーマットを作るのではなく、統一した用語形態をとることで、様々な利用が可能になります。

国立天文台			
ic:建物 所在			
ic:場所 住所			東京都三鷹市大沢2-21-1
ic:住所 住所			
ic:住所 構造化住所			
ic:構造化住所 国			
ic:構造化住所 都道府県			東京都
ic:構造化住所 市区町村			三鷹市
ic:構造化住所 町名			大沢
ic:構造化住所 街区符号			2
ic:構造化住所 住居番号			21
ic:構造化住所 地番			1
ic:構造化住所 方書			
ic:方書 方書			
ic:方書 ビル名			
ic:方書 部屋番号			
ic:構造化住所 郵便番号			181-8588
ic:建物 施設情報			
ic:施設 名称			国立天文台
ic:施設 種別			天文台
ic:施設 商用区分			
ic:施設 概要			天文台内の公開展示やガイドツアーがあります。
ic:施設 利用時間			
ic:オープン日			月、火、水、木、金、土、日
ic:開始時間			10:00
ic:終了時間			17:00
ic:施設 料金			
ic:施設 収容人数			
ic:施設 駐車場			
ic:施設 保有設備			54
ic:施設 アクセス			台 武蔵境駅南口からバスで天文台前15分程度。 調布駅北口からバスで天文台前15分程度。
ic:施設 リファレンス			
ic:建物 建物情報			
ic:建物構造 敷地面積			
ic:建物構造 主要用途			
ic:建物構造 建築面積			
ic:建物構造 延べ面積			
ic:建物構造 最高の高さ			
ic:建物構造 地上階数			
ic:建物構造 地下階数			
ic:建物構造 構造			
ic:建物構造 竣工日			
ic:建物 避難所情報			
im:避難所ID			
im:避難所 種別			広域避難場所
im:避難所 期間			
im:避難所 収容人数			
im:避難所 救護設備			
im:避難所 備蓄倉庫			
im:避難所 給水設備			
im:避難所 炊事設備			
im:避難所 発電設備			
im:避難所 風呂			
im:避難所 シェア			
im:避難所 トイレ			
im:避難所 救援資材			
ic:建物 連絡先情報			
ic:連絡先 名称			国立天文台
ic:連絡先 メールアドレス			
ic:連絡先 住所			東京都三鷹市大沢2-21-1
ic:連絡先 電話番号			0422-34-3600
ic:連絡先 FAX番号			
ic:連絡先 URL			http://www.nao.ac.jp/access/mitaka/

防災分野の情報も一括管理

施設IMIが網羅している用語について

- 三鷹市、横浜市、東京都、京都府、京都市、(米)Mapquest、(仏)ミシュラン等を参考に用語を整理しています。
- これらの用語を構造的に表すことができます。
- 英語対応も実施

道路
通り
遊歩道
階段
エレベータ
エスカレータ
スロープ
小径
橋
空港
ヘリポート
タクシー乗り場
レンタカー
レンタル自転車
駐輪場
駐車場
ロープウェイ
リフト
ケーブルカー
船着き場
海上輸送
カーフェリー
旅客船
駅
バス停

広場
噴水
運河
川
湖
池
港
灯台
建物
警察署
交番
消防署
消防署出張所・消防団詰所
消火栓
可搬式消防ポンプ
街頭消火器
AED
郵便局

学校
授乳室・育児室
子育て施設
保育園
幼稚園等
学童保育クラブ
小学校
中学校
高校
専修学校
短期大学
大学
生涯学習施設
文化施設
図書館
児童館
子供遊楽所
病院
保健所等
高齢者支援施設
老人ホーム
介護付き老人ホーム
福祉センター
介護支援センター

都道府県庁舎
市庁舎
区役所
市政窓口
教育センター
消費者活動センター
各種相談所
公営住宅
観光案内所
法務局
税務署
運転免許試験場
職業安定所
ごみ処理施設
資源回収実施場所
市民活動施設・産業支援施設
コミュニティ・センター
上水道施設
下水道施設
投票所
公共施設
大使館等

銀行
ATM
工場
市場
神社
寺
教会
モスク
宗教的建造物
門
城
城壁
展望
展望台
花の名所
記念碑
遺跡
洞窟、鍾乳洞
史跡
温泉
風車
ダム
墓地

スポーツ施設
競技場
テニスコート
ゲートボール場
ゴルフ場
プール
スケート場
スキー場
山小屋
キャンプ場
ヨットハーバー
電波塔
給水塔
塔
宿泊施設
レストラン
レ스토랑
商業施設
買い物エリア
コインロッカー
その他施設
博物館
美術館
ホール
劇場
公園
避難所

用語は実証を通じて精査予定(追加等)
また、カテゴリ等の精査を実施

導入方式

- 既に施設管理用システムを持っている場合
 - IMIの用語との対応表を整理します。外向けAPIでの情報発信などで、IMIを使うことで、相互運用性が高まります。

観光	横浜市	東京都	京都	US	ミシュラン	その他
観光案内所 Tourist Information Center	観光案内所 Tourist Information Center	観光案内所 Tourist Information Center	観光案内所 Tourist Information Center		観光案内所 Tourist Information Center	観光案内所 Tourist Information Center
ATM ATM						ATM ATM
両替所 Currency Exchange	両替所 Foreign Money Exchange	外貨両替所 Money Exchanger				両替所 Currency Exchange
荷物置き場 Baggage Room						荷物置き場 Baggage Room
コインロッカー Locker						コインロッカー Locker
トイレ Rest Room			公衆便所 Public toilet			トイレ Rest Room
AED AED						AED AED
授乳室・育児室 Nursery						授乳室・育児室 Nursery
道路 Road			道路 Road		道路 Road	
通り Street				通り Street	通り Street	
遊歩道 Penedestrian Route	遊歩道 Penedestrian Route				歩行者専用道路 pedestrian walkway	
階段 Stair				階段 Stairway	階段 Stairs	
エレベータ Elevator						
エスカレータ Escalator						
スロープ Slope						
小径 Path					小径 Path	
広場 Place				広場 Place	広場 Place	
噴水 Fountain					噴水 Fountain	
橋 Bridge				橋 Bridge	橋 Bridge	
運河 Canal				運河 Canal	運河 Canal	
川 River			川 River	川 River	川 River	
湖 Lake				湖 Lake	湖 Lake	
港 Port			池 Pond	港 Port	港 Port	
灯台 Lighthouse					灯台 Lighthouse	
建物 Building					建物 Building	
警察署 Police Station	警察署 Police Station	警察署 Police	警察署 Police	警察署 Police Station	警察署 Police	
交番 Police Box	交番 Police Box	交番 Koban	交番 Koban	交番 Police Office		
郵便局 Post Office	郵便局 Post Office	郵便局 Post Office	郵便局 Post Office		郵便局 Post Office	
銀行 Bank	銀行 Bank	銀行 Bank	銀行 Bank	銀行 Bank	銀行 Bank	
学校 School	学校 School	学校 School	学校 School	学校 School	学校 School	
大学 University			大学 University	大学 University	大学 University	
幼稚園 Kindergarten		幼稚園 Kindergarten	幼稚園 Kindergarten			
病院 Hospital	病院 Hospital	病院 Hospital	病院 Hospital	病院 Hospital	病院 Hospital	
消防署 Fire Station		消防署 Fire Department	消防署 Fire Department	消防署 Fire Station	病院 Hospital	

➢ また、他部局のマップとの融合を図っていくことで、活用が広がります。

- 新規整備、または既に持っている情報を根本的に見直す場合
 - 防災等も含めて各施設に情報の入力をお願いします。

● H25年度事業で入力支援ツールは整備予定

導入に当たって

■ IMI対応することのメリット

- 周辺自治体等との個別調整が不要になる
- 多くの分野で再利用が可能になる

■ デメリット、リスク

- 現時点では検証段階なので、将来、情報項目等が変更される恐れがある。

■ 留意事項

- 一見すると大変そうであるが、従来の情報管理と変わらないということを現場の人に理解してもらうことが重要です。
- 共通語彙基盤が普及しなかった場合でも、情報は構造化され、整理されるので再利用が容易になりますので、取り組みはムダになりません。

20

情報交換が必要な例

井の頭公園周辺

概要

- 関連自治体から地図が公表されているが情報がバラバラである。
 - トイレに障害者用があるかどうかの情報が、都の地図にだけある
 - 情報の粒度が異なる。
- 三鷹市は、目的に応じて様々な地図を公表している
- 三鷹市と武蔵野市でデータの内容、公表形式が違うので、地域全体で一体化した案内が難しい。

三鷹市の市民向け地図

三鷹市防災マップ



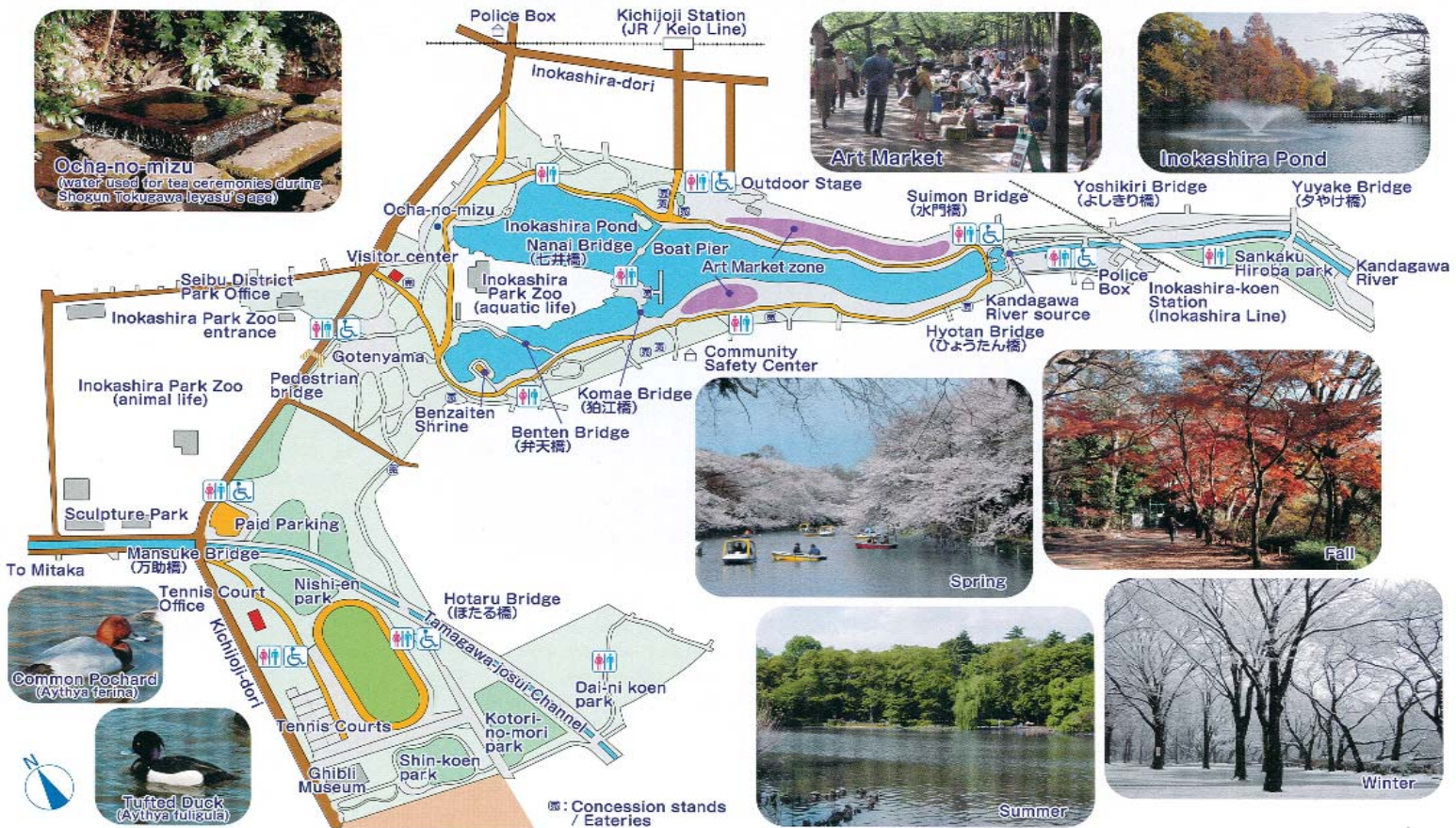
三鷹市の案内図



武蔵野市の案内図



東京都の案内図



情報連携用語彙データベース に係るプロジェクトについて

経済産業省が実施した調査事業

2012年度

1. 「オープンデータに関する調査研究（2012年度）」（2012年3月29日）
 - オープンデータのニーズや経済効果やルール整備の調査
 - 経済産業省及び関係機関が保有するデータを対象とした二次利用可能化のための技術や法的課題の検討
2. 「公共データワーキンググループ報告書（2012年度）」（2013年6月）
 - 経済産業省の保有する公的データ（公共データ）を洗い出し、その一部をオープンデータとして整備。
 - データ公開に係る技術的課題、法的課題を検討
3. 空間位置情報に関連する公共データの活用実証事業（2012年度）
 - 自治体等で保有する公共データと空間位置情報との関係性の整理と有効的な組み合わせについて調査し、サービス利用の促進に必要となる課題などを整理。
4. 「公共情報交換標準スキームの整備に関する調査研究（2012年度）」（2013年3月）
 - 行政の情報システムにおける情報連携の効率拡大のため、データの意味レベルでの相互運用性拡大のための手法につき、海外動向（NIEM(米国)、JOIONUP(欧州)）を中心に調査。
 - 情報交換用語彙セットのプロトタイプを整備
 - 情報交換スキームの有るべき姿として「IMI」を提案

⇒ 本「語彙基盤プロジェクトは」上記調査研究で提案されたスキーム「IMI」をベースに、情報共有の基盤の実現を目指すものである。

2013年度（実施中）

1. 「オープンデータの利用に係る民間、地方公共団体の連携基盤の開発・活用実証事業」（2013年度）
 - 地理情報を対象とした情報連携のためのツール（基盤地図情報等のWeb標準（JIS X 7197、JIS X 4197等）準拠に変換するツール等）を整備し、実証実験を実施。
 - ツールは11月ごろプロトタイプを公開、3月末に正式版をオープンソースにより公開予定。

語彙基盤とは

情報の電子的な公開や交換の際、伝えたい内容が正しく伝わるようにするために、

1. 情報の構造に一貫性を持たせ、
2. そこに用いられる用語の意味を明確にし、
3. 用語の統一化または用語間の関連性の明確化を図る

事を目的とし、用語の意味定義や構想、データ作成時のルール等の基本情報を共有し、活用するための基盤。



構造、意味、表記の共通化

語彙基盤の効果

1 次的効果：

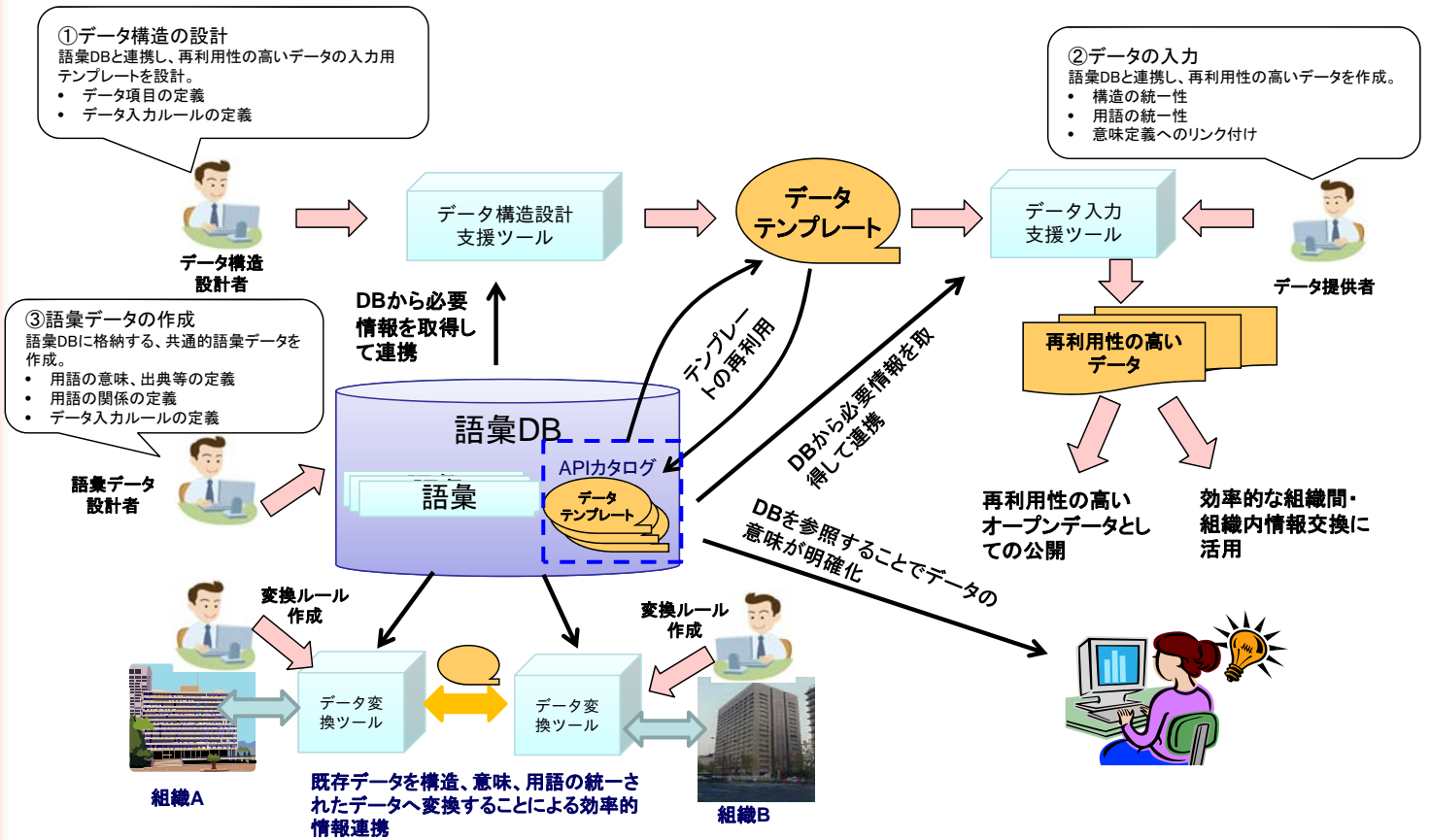
1. 行政機関等が、意味が明確で再利用性の高い情報を効率的に作成できる。
2. 行政機関などから公開された情報の意味の解釈が機械的に可能となる。

(再利用性の高い情報が普及することによる) 2 次的効果：

1. 情報交換効率の向上による行政の業務効率とサービス向上
2. オープンデータとして公開された再利用性の高い情報を活用し、価値を引き出す各種産業の発展

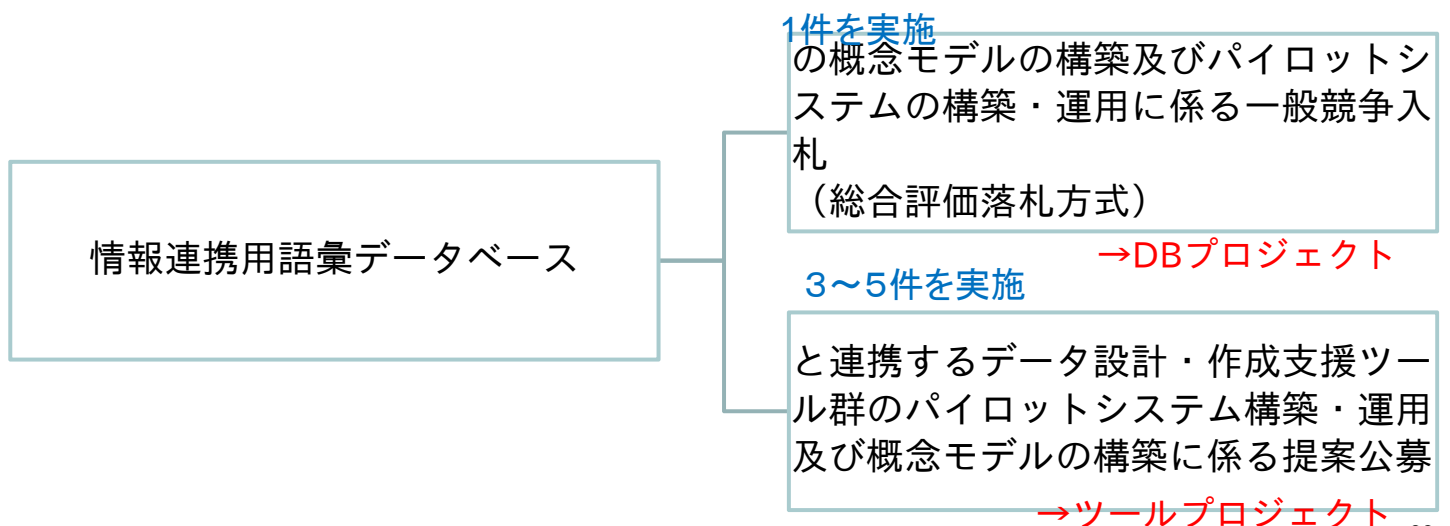
「用語」、「語彙」、「語彙データ」

- 「用語」は、ある特定の意味あるいは概念を示すものであり、文字を組み合わせた表記を伴う。また、用語は、その意味あるいは概念の一部として、他の用語の関係（構造）を持つ場合がある。たとえば、「時刻」という用語を「時」、「分」、「秒」という下位の用語から構成すると定義することができる。また、「融資」と「貸し付け」のように、同一の意味を持つ用語が、異なる表記で示される場合がある。
- 用語の集合が「語彙」である。
- 用語の構造や意味、表記の制限、用語と用語との関係等、用語に係る様々な情報を総称してここでは「語彙データ」と呼ぶこととする。なお、本事業で対象とする用語は、原則、行政の現場で業務上必要とされるものとする。

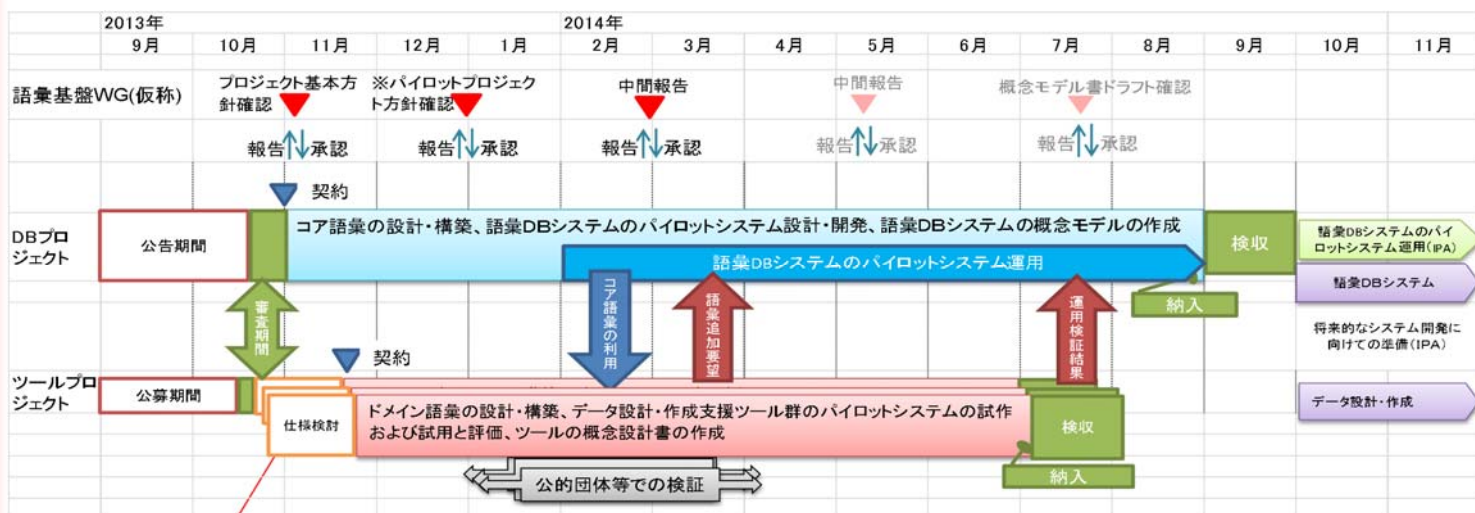


今回のプロジェクト構成

- 語彙基盤を支えるデータベースとツール群について、その要求要件を洗い出した「概念モデルの構築」を行う
- 「概念モデルの構築」にあたっては、パイロットシステムを運用し、実際のデータ作成を試行的に実施し、そこで得られた知見を盛り込む
- プロジェクトは以下のように「DBプロジェクト」と「ツールプロジェクト」の2種類から構成される。



「概念モデルの構築」の全体スケジュール(2013年10月頃～2014年9月頃)



DBパイロットシステムに用意されるデータ構造、APIおよびコア語彙の仕様がDBプロジェクト担当者から、ツールプロジェクト担当者に渡される。

検証用の語彙データは極めて基本的な「コア語彙」以外については、ツールプロジェクト側が用意し、原則DBパイロットシステムへ登録し、運用。

評価はツールプロジェクト実施者が独自に行い、概念モデル中にまとめるとともに、それらをDBプロジェクト実施者がDB運用の視点から取りまとめる。

公共情報交換標準スキームの整備に関する調査研究(2012年度) 参考資料11に収められた語彙とする

「概念モデルの構築」を含む、計画全体の概観



DBプロジェクト

情報連携用語彙データベースの概念モデルの構築及びパイロットシステムの構築・運用に係る一般競争入札（総合評価落札方式）

36

DBの目的

- 情報の「構造、意味、表記の共通化」に必要なとなる基本的データ（語彙データ）を格納し、供給する。
 - 語彙データの、想定される供給先：
 - データ構造設計支援のためのツール
 - データ入力支援のためのツール
 - データの意味確認のためのツール
 - 情報連携の際等にデータを変換するためのツール（「ツール」には一般的Webブラウザも含む）
- 語彙データのメンテナンスを支援する機能を提供する。
- 利用者が設計した「データテンプレート」の共有を支援・促進する。
- その他、データの「構造、意味、用語の共通化に資する資源(ツール類を含む)を格納し、供給する。

1. DBの要求要件整理

- データ構造などに係る外部専門家を交え、客観的、かつ、国際的動向との整合性を配慮した検討を行う（10名程度による検討会を5回程度実施）
- 以下のドメインを設定し、各ドメイン及び共通ドメインにおけるニーズ評価及び利用シーンの洗い出しを行い、各シーンに基づいた要求要件を洗い出す

(ドメイン)

- 「電子行政オープンデータ推進のためのロードマップ」に掲げられた「オープンデータの重点分野（白書、防災・減災情報、地理空間情報、人の移動に関する情報、予算・決算・調達情報）を中心に対象とする分野を検討する。

(シーン例)

- 語彙データの配信に係る利用シーンとそれに対応した要求要件の洗い出し
- 語彙データの作成に係る利用シーンとそれに対応した要求要件の洗い出し
- データテンプレートの共有促進に係る利用シーンとそれに対応した要求要件の洗い出し
- 省庁等が所有する既存データを、標準的データへ変換することによる情報連携
- SNS的機能を活用した、語彙やテンプレートへの評価収集、意見交換機能の検討など

38

- データ構造の検討

- 「公共情報交換標準スキームの整備に関する調査研究(2012年度)」報告書第3章に示された「公共情報交換標準スキーム（IMI）」で示されたデータ構造をベースとし、さらに、
 - 「用語」の意味の記述法
 - 「用語」の構造の記述法
 - 「用語」の関係の記述法について、汎用性の高い構造を検討する。
- 既に定義された「用語」を使ってさらに別の新たな「用語」を定義できるといった、柔軟性、汎用性のある情報構造を検討する
- 国際標準（デファクトを含む）との整合性の検討（XML、RDF等）
- 諸外国先進事例（NIEM, JOINUP等）との整合性
- 既存データ（UN/CEFACT, Dublin Core等）との相互運用性
- 国等が統計等で用いている各種コードとの連携
- 観光・防災ドメイン等で重要となる地理情報の記述方法及び、地域コード等との連携
- データ入カールール及びその記述法の検討
- 基本データ型の検討（文字コード／文字セットを含む）

– APIの検討

- 語彙DBを活用するためのツール群と語彙DBが連携するためのAPIを検討。

– その他

- 実運用時におけるセキュリティ要件の洗い出し
- コアとドメインへの分割など、将来の効率的、持続的運用を考慮した論理構成、物理構成及び管理機能等の要件の洗い出し

– DB運用体制の検討

将来実現する語彙DBについて、語彙のメンテナンスを含めた運用全体の有り方に係る検討を行う。

- 語彙データの整備とメンテナンス、その他DBの運用全般を検討
- 語彙データの作成に係るルールの検討
- NIEMでとられているような、コアとドメインに分けるといった、検討構造につき、持続性の期待できる方法を検討。

40

2. ツール群に係る要件の洗い出し

- 語彙データを活用し、情報連携、情報公開の効率化に資するツールの要件を洗い出す。
 - データ構造設計支援のためのツール
 - データ入力支援のためのツール
 - データの意味確認のためのツール
 - 情報連携の際等にデータを変換するためのツール

3. 語彙データの整備

- 将来の中央省庁及び地方自治体等の情報公開、情報連携の基本となる用語を（3つ以上のドメイン及びそれらに共通するコア語彙につき、構造を持つ用語を合計100用語以上）リストアップ。
- 以下の対象を調査あるいは参照し、対象データを設定する
 - 公共情報交換標準スキームの整備に関する調査研究報告書(2012年度)
 - NIEM(米国)、JOINUP(欧州)、UN/CEFACT Core Component Library (国連)、地域情報プラットフォーム(APPLIC)、中間標準レイアウト仕様(総務省平成23年度調査)
 - 優先度の高い用語を、根拠を明確にして選択
 - 対象分野関係者を交えた十分な検討を実施(10名程度による検討会を5回程度実施)
- 検討したデータ構造に従い、上記でリストアップした用語を語彙データとして整備

4. DBパイロットシステムの構築と運用

パイロットプロジェクトで構築・運用するDBは、本プロジェクトで最終的に構築を目指す機能のうち、「ツールプロジェクト」を試行するために必要最小の機能を実現するものであり、最終的に実現を目指すDBとは必ずしも同一仕様である必要はない

- 語彙基盤の概念モデルの構築に必要な知見を収集するため、IPA共通基盤システム上にデータベースのパイロットシステムを構築し、運用する。
- 「公共情報交換標準スキームの整備に関する調査研究(2012年度)」報告書第3章に示された「情報交換基盤案」に基づくデータ構造に基づき、同報告書参考資料1に収められた「IMI コアボキャブラリ」の情報を格納し、WebAPIを通じて提供する。
- パイロットシステムで用いるAPI及び語彙データの詳細仕様を、契約後1か月以内に公表する（ツールプロジェクトが参照するため）。
- 「ツールプロジェクト」参加者が希望する語彙データのホスティングと配信を行う。ただし、データ構造は上記IMIコアボキャブラリと同等とする。
- その他、実施者が提案するDBテスト項目（**提案書へ提案**）について運用確認を行う
- 限定した利用者に対するアクセスを許可する
- （利用者限定の条件で）IPAが定めるセキュリティテストに合格すること

5. 概念モデルの取りまとめ

- データ構造等重要事項については、少なくとも米国NIEM、欧州ISAのJOINUP、英国data.gov.uk関係者に案を示し、意見聴取を行たうえで概念モデルをとりまとめる。
- 取りまとめた概念モデルに基づく語彙基盤の実現により期待される効果につき、定量的、定性的面から根拠を明確にして取りまとめる。

ツールプロジェクト

情報連携用語彙データベースと連携するデータ設計・作成支援ツール群の試作及び試用並びに概念モデルの構築に係る提案公募事業

44

ツール群の目的

- 語彙データベースと連携し、共通化された構造を持ち、意味が明確化され、標記に統一化の図られたデータの作成を支援する。
 - データ構造の設計を支援するツール
 - 設計されたデータ構造へのデータ入力を支援するツール
 - 既存のデータを、語彙基盤が規定する標準的構造、用語に変換することにより、情報連携の効率化を図るためのツール
- 等がある。

- データ入力のテンプレートとなるデータ（データテンプレート）を作成するためのツール。
 - データ構造を定義する
 - 語彙DBに格納されている「語彙データ」、「データテンプレート」を参照し、適切なものを選択して構造定義に活用する。
 - 語彙DBに既存でない語彙データについては、その場で作成することを支援する。
 - データ入力ルールを定義する
 - データの型、値の制限などを定義する。

46

- データテンプレートに基づき、指定された場所に指定された形式のデータを入力するようガイドするツール。併せて以下の機能を検討する。
 - データの意味を明確化するためのリンク等作成支援
 - 語彙データベースが提供する語彙データへのリンクの他、地理座標情報の付与、コード情報、法令情報等の公的データベース等へのリンク作成による意味明確化等も含めた検討を行う。
 - 標記の統一を図るための用語選択支援

- 省庁、自治体などが既に有しているデータを、標準的構造、用語に変換することにより、情報連携の効率化を図るツール。
 - データ変換ルールの設定
 - 既存データの構造やそこで用いられる用語について、語彙データベースで提供される標準的な構造、意味等に変換するための変換規則を設定するためのツール
 - データ変換ツール
 - 上記で設定されたルールに基づき、既存データと、標準形式のデータとの間でのデータ変換を行うためのツール

1. 対象者

- 本パイロットプロジェクトの実施主体は、地方自治体等、公的な業務を行う団体（以後公的団体と呼ぶ）のシステム構築に関わる企業等とする。
- パイロットシステムの構築、運用に際しては当該公的団体の職員等が構築への助言、および運用結果への評価に関与するものとする。

2. 対象業務（以下のいずれかまたは両方とする）

- 再利用性の高いオープンデータの作成と公開
- 業務連携のための再利用性の高いデータの作成とその交換

3. 業務内容

① ツールの要件整理

- 利用シーンの設定と、そのシーンに基づいたツールの要件の洗い出し
(例)
 - データ構造設計支援
 - データ入力ルール設計支援
 - データ意味定義とのリンク作成支援

② 語彙データの構築

- パイロットプロジェクトの実施に必要な語彙データを構築する。
- 語彙データ構築に当たっては「DBプロジェクト」が策定した要件に沿うものとする。
- 語彙データは、「DBプロジェクト」が運用するデータベースパイロットシステムに格納し、そこから参照することを原則とする。

③ 運用実験

- 上記①で要件整理したツールを実際に作成して運用し、データテンプレートの作成、データの入力に活用する。
- 地理データに係るツールについては、「平成25年度電子経済産業省構築事業（オープンデータの利用に係る民間、地方公共団体の連携基盤の開発・活用実証事業）」で実施されるプロジェクトで作成されたものが公開される予定なので、その活用、データとしては、総務省統計局の国勢調査データ（IDとポリゴン）、国土交通省位置参照コード（IDと緯度経度座標）等の活用の検討が推奨される。
- 実験で作成したデータは納品対象としない。データの作成は実験に必要最小限の範囲とし、それを超えたものは実施者の負担によるものとする。
- 作成した「語彙データ」は、納品の対象とする。ただし、実施者がもともと所有していたデータあるいは外部参照したデータについては、その限りでない。その場合データは納品対象とはならないが、当該データを本プロジェクトで活用するために変換するツールなどを本プロジェクトで作成した場合は、そのツールは納品対象とし、IPAからオープンなライセンスで公開する。
- ツールの運用を通し、ツールおよび、DB、語彙データ等の有効性、問題点を明らかにする。

④ 概念モデルの取りまとめ

- ツールを運用し、実際にデータテンプレート及びデータを作成し、作成効率、作成されたデータの質などに付き、評価を行う。
- ツールの運用実験結果を受け、ツール群及びそれが必要とするデータに係る概念モデルを作成する。

1	世界最先端IT国家創造宣言	http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf
2	世界最先端 IT 国家創造宣言 工程表	http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou4.pdf
3	電子行政オープンデータ戦略	http://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryou2.pdf
4	電子行政オープンデータ推進のためのロードマップ	http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou3.pdf
5	オープンデータに関する調査研究	http://datameti.go.jp/data/dataset/report-001-2012
6	公共情報交換標準スキームの整備に関する調査研究	http://datameti.go.jp/data/dataset/report-002-2012
7	IT融合フォーラム公共データワーキンググループ報告書	http://datameti.go.jp/data/dataset/report-003-2012
8	空間位置情報に関連する公共データの活用実証事業	http://datameti.go.jp/data/dataset/report-004-2012
9	平成22年度版「地域情報プラットフォーム標準仕様」	http://www.applic.or.jp/2012/tech/APPLIC-0002-2011/index.html
10	自治体業務アプリケーションユニット標準仕様V2.4	http://www.applic.or.jp/2012/tech/gyomu.html
11	総務省統計局の国勢調査データ	http://e-stat.go.jp/SG2/eStatGIS/page/download.html
12	国土交通省位置参照コード	http://nlftp.mlit.go.jp/isj/data.html
13	総務省 中間標準レイアウト仕様	http://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/lg-cloud/02kiban07_03000024.html
14	IPA 文字情報基盤整備事業	http://mojikiban.ipa.go.jp/
15	米国オープンデータの動向調査報告書	http://www.ipa.go.jp/about/research/20130830.html
16	NIEM	https://www.niem.gov/Pages/default.aspx
17	ISA Joinup	http://joinup.ec.europa.eu/
18	Schema Subset Generation Tool(SSGT)	http://tools.niem.gov/niemtools/ssgt/index.iepd
19	Data Model Browser	http://tools.niem.gov/niemtools/viewer/DataModelViewer.iepd
20	Work With IEPDs	http://tools.niem.gov/niemtools/iepd/index.iepd
21	Migration Assistance	http://tools.niem.gov/niemtools/migration/index.iepd
22	Conformance Testing Assistant(ConTesA)	http://tools.niem.gov/contesaNIEM/
23	Code List Generator	http://tools.niem.gov/niemtools/codelist/index.iepd
24	IEPD Clearinghouse	https://it.ojp.gov/framesets/iepd-clearinghouse-noClose.htm
25	Data.gov.uk	http://www.data.gov.uk/

検討すべき論点

今年度のプロジェクトを通してたたき台を検討し、情報共有基盤推進委員会で検討

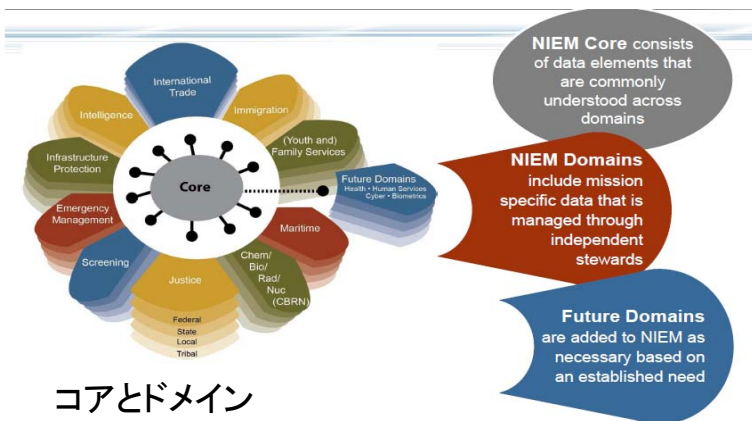
関係省庁の担当者を交えた検討

技術的専門家を交えた検討

将来の運用・検討体制
語彙の収集・整理
データ構造

運用・検討体制の検討

- IMIの将来の安定的な運用・検討体制について、有るべき姿の検討。
(参考)
 - 米国NIEMでは、ドメイン毎に責任者を決めた分散的運用・検討体制を構築している。
 - さらに、国全体としての意思決定機構も構築されている。

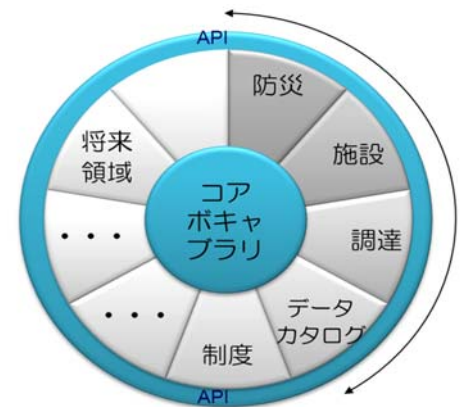


NIEMの意思決定機構



語彙の収集・整理

- 対象とする分野、処理順等の検討
 - 優先度の評価
 - 需要の評価
- 語彙データとして収録する語の選択
 - 分野ごと語彙の優先度を検討
- 語彙データの整理法の検討
 - 典拠情報等の在り方
 - 類義語・同義語の整理
 - 「コア」と「ドメイン」の分け方



56

データ構造の検討

- 柔軟性の高さ
- 利便性の高さ
 - 情報公開（オープンデータ）の視点
 - 組織内/組織間情報共有の視点
- 日本固有の事情への対応
 - 同義語・類義語等の扱い
 - 伝統的スタイルとの整合性
- 先進事例との整合性