

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

概念モデル書

平成 26 年 11 月 28 日

実施企業: 株式会社 日立製作所

独立行政法人 情報処理推進機構(IPA)

「本ドキュメントは、クリエイティブ・コモンズ 表示・継承 2.1 日本 ライセンス (CC BY-SA 2.1 JP)によって許諾されています。ライセンスの内容を知りたい方は <http://creativecommons.org/licenses/by-sa/2.1/jp/> でご確認ください。」

目次

1. 共通語彙基盤の事業概要	13
1.1 背景	13
1.2 共通語彙基盤の全体像	14
1.3 「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（以下、「本事業」という。）」の概要	15
1.4 「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（以下、ツールプロジェクト）」の概要	15
1.5 用語の定義	16
2. 共通語彙基盤整備に向けた取り組み	19
2.1 現状の課題整理	19
2.1.1 行政の業務向上とサービス向上、各種産業の発展のためには	19
2.1.2 ツールプロジェクトの成果から抽出した現状の課題	20
2.2 本事業に先行する共通語彙基盤整備についての取組	22
2.3 本事業の実施内容	22
2.3.1 検討会の設置	22
2.3.2 語彙の整備	24
2.3.3 語彙データベースの要求事項の整理	25
2.3.4 共通語彙基盤の運用についての検討	27
2.3.5 パイロットシステムの構築と運用	28
2.3.6 概念モデル書の取りまとめ	28
3. 本事業における語彙の設計方針	29
3.1 先行事例：IMI コア語彙 v1	29
3.1.1 IMI コア語彙 v1 の特徴および整備方針	30
3.1.2 IMI コア語彙 v1 の用語の全体像	32
3.1.3 IMI コア語彙 v1 のデータ構造	35
3.2 IMI コア語彙 v1 の整備において明確となった課題	36
3.2.1 表記の明確化・関連付けについて	37
3.2.2 意味の明確化について	38
3.2.3 用語の構造化について	38
3.2.4 用語の相互運用性について	38
3.2.5 用語のデータ構造について	39
3.2.6 NIEM への依存について	39
3.2.7 コードリストおよびコードの具体的な取扱いについて	39
3.3 共通語彙基盤における語彙の設計方針	40

3.3.1	語彙の整理（コア語彙とドメイン語彙、ドメイン共通語彙による分類）	40
3.3.2	語彙の位置付けについて	41
3.3.3	用語の表記について	41
3.3.4	用語の意味および構造化について	42
3.3.5	用語の相互運用性について	42
3.3.6	用語のデータ構造について	42
3.3.7	用語の物理形式について	43
3.3.8	論理形式と物理形式の相互変換について	43
3.3.9	XML Schema への依存の軽減について	43
3.3.10	NIEM への依存の軽減について	44
3.3.11	コードリストおよびコードの参照と格納について	44
4.	コア語彙の整備	45
4.1	意味的検討：コア語彙の内容の精査・見直し	45
4.1.1	用語の表記の見直し	46
4.1.2	用語の意味の見直し	47
4.1.3	用語の構造化の見直し	48
4.2	構造的検討：コア語彙のデータ構造の拡充	51
4.2.1	データ構造の構成要素の分類・整理	51
4.2.2	語彙データの表現形式の整理	61
4.2.3	語彙データの論理形式	62
4.2.4	語彙データの物理形式	64
4.2.5	語彙データのシリアライズ・デシリアライズ	70
4.3	IMI コア語彙 v2 案の整備結果	84
4.3.1	IMI コア語彙 v2 案の用語の全体像	84
5.	ドメイン語彙サンプルの整備	86
5.1	ドメイン語彙サンプルの整備の目的	86
5.2	ドメイン語彙収集・整備プロセスの仮説	87
5.2.1	ドメインの選定	87
5.2.2	ドメイン語彙の収集・整備プロセスの仮説	91
5.3	ドメイン語彙サンプルの整備	94
5.3.1	ドメイン語彙サンプルの整備の実施内容	94
5.3.2	ドメイン語彙の収集・整備プロセスを見直す必要があった事例とその対応方法	108
5.3.3	本事業で整備したドメイン語彙サンプル	118
6.	語彙データベースの要求事項	130
6.1	要求事項の整理方針	130

6.2	想定される利用シーンについて	131
6.2.1	利用シーンの初期仮説	131
6.2.2	語彙データベース検討会およびドメイン語彙検討会を通じて導き出した利用シーン	132
6.2.3	ツールプロジェクトで設定した利用シーン	134
6.2.4	海外における先進事例で実績のある利用シーン	136
6.2.5	想定利用シーンの整理結果	139
6.3	想定されるツール群について	149
6.3.1	利用シーン検討結果に基づくツール群	149
6.3.2	ツールプロジェクトでの実証から得られたツール群	153
6.3.3	海外における先進事例で実績のあるツール群	156
6.3.4	想定ツール群の整理結果	157
6.4	実装機能案及びAPI案の策定	165
6.4.1	語彙データベースに対する要求機能の洗い出し	165
6.4.2	API案の策定に関する検討結果	168
6.5	データベースとして実装するためのデータ構造の拡充	174
6.5.1	API実装に必要な属性の洗い出し	174
6.5.2	RDFストアでの実装を前提としたデータ構造案の策定	176
6.5.3	データ構造案の策定に関する検討結果	177
6.6	セキュリティ要件の整理	191
6.6.1	パイロットシステムの構成	191
6.6.2	情報システムに求められるセキュリティ特性	192
6.6.3	語彙データベースに対するセキュリティ要件の整理	193
6.7	スケーラビリティを配慮したアーキテクチャの検討	196
6.7.1	情報システムに求められるスケーラビリティの観点	196
6.7.2	スケーラビリティ要件の整理	197
7.	本事業にて整備した共通語彙基盤の運用についての検討	199
7.1	先行事例の運用体制に関する調査	199
7.1.1	NIEMの運用体制	199
7.1.2	ISAの運用体制	206
7.1.3	Schema.orgの運用体制	208
7.1.4	先行事例における運用体制・運用プロセスの比較	210
7.2	共通語彙基盤の運用体制の提案	212
7.3	共通語彙基盤の運用プロセスの提案	214
7.3.1	共通語彙基盤の運用プロセスのポリシー	214
7.3.2	共通語彙基盤の運用プロセス案	215

7.3.3 海外関連機関との将来的な相互運用プロセス確立に向けた提案.....	216
8. 共通語彙基盤の実用化に向けた提言	217
8.1 「①語彙の収集・整備」における課題と提言	218
8.1.1 コア語彙の検討プロセスの可視化.....	218
8.1.2 コードリストの将来的な整備およびメンテナンスの方針.....	219
8.1.3 コミュニティベースの分散型語彙整備.....	220
8.2 「②データテンプレートの設計」における提言	221
8.2.1 利用者にとっての用語の分かりやすさの向上.....	221
8.2.2 用語の意味の補強	221
8.2.3 利用シーンに特化したパッケージの整備	223
8.3 「③データテンプレートに基づく実データの入力」における提言	224
8.3.1 データ入力者のニーズに基づく RDF 形式の継続的な検討	224
8.3.2 用語が取り得る値の文字セットを指定する方法について	224
8.4 「④実データの利用・活用」における提言.....	225
8.4.1 既存のサービスとの親和性の向上	225
8.4.2 多言語対応について.....	225
8.4.3 用語間の関係性の詳細化	225
8.4.4 JSON 形式のサポート.....	226
9. まとめ.....	227

目次

図 1.2.1 共通語彙基盤の全体像	14
図 2.1.1.1 意味の明確化とデータ構造の整備の必要性.....	19
図 2.3.1.1 本事業で設置した検討会	23
図 3.3.1.1 本事業における語彙の整理	40
図 4.2.1.1 語彙の構成要素 (UML クラス図)	52
図 4.2.2.1 語彙データの表現形式	61
図 4.2.3.1 語彙データの論理形式： 語彙テーブル	62
図 4.2.3.2 語彙データの論理形式： 用語テーブル	62
図 4.2.3.3 語彙データの論理形式： 関係性テーブル	63
図 5.2.2.1 ドメイン語彙の収集・整備プロセスの仮説.....	91
図 5.3.2.1 「路線」の概念のイメージ	109
図 5.3.2.2 「列車」の運行パターンの概念のイメージ.....	109
図 5.3.2.3 「列車」の概念のイメージ	110
図 5.3.2.4 「駅」の概念のイメージ	111
図 5.3.2.5 見直し後 ドメイン語彙の収集・整備プロセス（移動・交通ドメイン）	112
図 5.3.2.6 市区町村予算費目標標準語彙と税どこ標準カテゴリ語彙の関係性のイメー ジ.....	113
図 5.3.2.7 見直し後 ドメイン語彙の収集・整備プロセス（財務ドメイン） ...	114
図 5.3.2.8 防災ドメインにおける複数の主体が存在するイメージ.....	115
図 5.3.2.9 見直し後 ドメイン語彙の収集・整備プロセス（防災ドメイン） ...	117
図 5.3.3.1 地理空間・施設ドメインの語彙サンプルのクラス図.....	118
図 5.3.3.2 移動・交通ドメインの語彙サンプルのクラス図	121
図 5.3.3.3 財務ドメインの語彙サンプルのクラス図.....	124
図 5.3.3.4 防災ドメインの語彙サンプルのクラス図.....	127
図 6.2.4.1 NIEM IEPD ライフサイクルの概要図.....	138
図 6.5.3.1 拡充後のデータ構造案 (UML クラス図).....	177
図 6.6.1.1 パイロットシステムの概要図	191
図 7.1.1.1 NIEM 運用組織体制	200
図 7.1.1.2 NIEM の運用プロセス・ルール	202
図 7.1.2.1 ISA 運用組織体制	206
図 7.1.2.2 ISA の運用プロセス・ルール	207
図 7.1.3.1 Schema.org 運用組織体制	208
図 7.1.3.2 Schema.org の運用プロセス・ルール.....	209
図 7.1.4.1 共通語彙基盤の運用体制案	213
図 7.3.1.1 共通語彙基盤の運用プロセスのポリシー案.....	214
図 7.3.2.1 共通語彙基盤の運用プロセス案.....	215

図 7.3.3.1 海外関連機関との将来的な相互運用プロセス確立に向けた取り組みイメージ.....	216
図 7.3.3.1 共通語彙基盤の実用化に向けた提言の対象.....	217
図 8.1.3.1 コミュニティベースの分散型語彙整備体制案.....	220
図 8.2.3.1 IEP を活用した共通語彙基盤の利用イメージ.....	223

表目次

表 2.1.1.1 用語の定義.....	16
表 2.3.3.1 海外聴取先①.....	25
表 2.3.3.2 海外聴取先②.....	25
表 2.3.3.3 海外聴取先③.....	25
表 2.3.3.4 海外聴取先④.....	26
表 2.3.3.5 海外聴取先⑤.....	26
表 2.3.3.6 海外聴取先⑥.....	26
表 2.3.3.7 海外聴取先⑦.....	26
表 2.3.3.8 海外聴取先⑧.....	26
表 2.3.3.9 海外聴取先⑨.....	26
表 3.1.1.1 IMI コア語彙 v1 の整備方針に関する主な特徴.....	30
表 3.1.2.1 IMI コア語彙 v1 に含まれる主要な用語一覧.....	32
表 3.1.3.1 IMI コア語彙 v1 のデータ構造.....	35
表 3.1.3.1 平成 24 年度事業で示された、IMI コア語彙 v1 に関する残課題.....	36
表 4.1.1.1 用語の表記の見直し.....	46
表 4.1.2.1 用語の意味の見直し.....	47
表 4.1.3.1 用語の構造化の見直し.....	48
表 4.2.1.1 語彙の構成要素.....	51
表 4.2.1.2 「語彙」のデータ構造.....	53
表 4.2.1.3 「用語」のデータ構造.....	55
表 4.2.1.4 「クラス用語」のデータ構造.....	56
表 4.2.1.5 「プロパティ用語」のデータ構造.....	57
表 4.2.1.6 「コードリスト」のデータ構造.....	59
表 4.2.1.7 「コード」のデータ構造.....	60
表 4.2.5.1 XML 形式および RDF 形式で参照される語彙（名前空間）.....	70
表 4.2.5.2 「語彙」の論理形式と XML 形式の対応.....	72
表 4.2.5.3 「クラス用語」の論理形式と XML 形式(xsd:complexType 要素)の対応.....	73
表 4.2.5.4 「クラス用語」の論理形式と XML 形式(xsd:element 要素)の対応.....	74
表 4.2.5.5 「プロパティ用語」の論理形式と XML 形式(トップレベルの xsd:element 要素)の対応.....	75
表 4.2.5.6 「プロパティ用語」の論理形式と XML 形式(xsd:complexType 中の xsd:element 要素)の対応.....	76
表 4.2.5.7 「コードリスト」の論理形式と XML 形式の対応.....	77
表 4.2.5.8 「コード」の論理形式と XML 形式の対応.....	78
表 4.2.5.9 「語彙」の論理形式と RDF 形式の対応.....	79
表 4.2.5.10 「クラス用語」の論理形式と RDF 形式の対応.....	80
表 4.2.5.11 「プロパティ用語」の論理形式と RDF 形式の対応.....	81

表 4.3.1.1 IMI コア語彙 v2 案に含まれる主要な用語（構造をもった用語）	84
表 5.2.1.1 ドメイン選定の考え方	87
表 5.2.1.2 国内外の各戦略や取り組みに規定されるドメインの抽出結果一覧	88
表 5.2.1.3 試行的に選択したドメインとその概要	90
表 5.2.2.1 ドメイン語彙収集・整備プロセスの仮説の詳細	92
表 5.3.1.1 「①利用シーンの選定」での実施作業	94
表 5.3.1.2 「② キーワードの抽出」での実施作業と作業結果	95
表 5.3.1.3 「③既存語彙の収集」での実施作業と作業結果	96
表 5.3.1.4 「④ 既存語彙からの用語抽出」での作業概要及び作業観点	99
表 5.3.1.5 「⑤用語の整列」での作業概要及び作業観点	100
表 5.3.1.6 「⑥標準用語の決定」での作業概要および作業観点	101
表 5.3.1.7 「⑦階層化」での作業概要及び作業観点	103
表 5.3.1.8 「⑧用語の構造化」での作業概要及び作業観点	104
表 5.3.1.9 「⑨データタイプの決定」での作業概要及び作業観点	104
表 5.3.1.10 「⑩他の語彙との調整」での作業概要及び作業観点	105
表 5.3.1.11 「⑪フィールド適用による検証」でのテンプレート作成作業	106
表 5.3.1.12 「⑫フィールド適用結果のフィードバック」の作業概要及び作業観点	107
表 5.3.2.1 「路線」の概念の整理結果	108
表 5.3.2.2 「列車」の運行パターンの概念の整理結果	109
表 5.3.2.3 「列車」の概念の整理結果	110
表 5.3.2.4 「駅」の概念の整理結果	111
表 5.3.3.1 地理空間・施設ドメインのドメイン語彙サンプルに含まれる主要な用語（構 造をもった用語）	119
表 5.3.3.2 移動・交通ドメインのドメイン語彙サンプルに含まれる主要な用語（構 造をもった用語）	122
表 5.3.3.3 財務ドメインのドメイン語彙サンプル（自治体一般会計歳出予算用）に 含まれる主要な用語（構造をもった用語）	125
表 5.3.3.4 財務ドメインのドメイン語彙サンプル（「税どこ」）に含まれる主要な用語 （構造をもった用語）	126
表 5.3.3.5 防災ドメインのドメイン語彙サンプルに含まれる主要な用語（構 造をもった用語）	128
表 6.2.1.1 利用シーンの初期仮説	131
表 6.2.2.1 語彙データベース検討会およびドメイン語彙検討会で得た意見から導き出 した利用シーン	132
表 6.2.3.1 ツールプロジェクトで設定した利用シーン	134
表 6.2.4.1 海外における先進事例で実績のある利用シーン	136
表 6.2.5.1 利用シーンの一般化	139

表 6.2.5.2	想定利用シーン (S: Scene)	144
表 6.2.5.3	語彙データベースの想定利用者 (U: User).....	148
表 6.3.1.1	利用シーンの検討結果に基づくツール群	149
表 6.3.2.1	ツールプロジェクトでの実証から得られたツール群	153
表 6.3.3.1	海外における先進事例で実績のあるツール群.....	156
表 6.3.4.1	ツール群の一般化	157
表 6.3.4.2	語彙データベースの想定ツール群 (T: Tool).....	162
表 6.4.1.1	ツール群が求める語彙データベースの要求機能.....	165
表 6.4.2.1	語彙データベースの要求機能と API の関係	168
表 6.4.2.2	語彙データベースの API 案 (A: API).....	170
表 6.4.2.3	API 詳細の例： A5. 個別取得 API	173
表 6.5.1.1	API 実装時に必要となる処理および当該処理に必要な属性.....	174
表 6.5.3.1	各種プレフィックスと名前空間 URI の対応.....	178
表 6.5.3.2	「語彙」クラス	179
表 6.5.3.3	「語彙」クラスの推奨プロパティ	179
表 6.5.3.4	「用語」クラス	181
表 6.5.3.5	「用語」の推奨プロパティ	181
表 6.5.3.6	「クラス用語」クラス.....	183
表 6.5.3.7	「クラス用語」の推奨プロパティ	183
表 6.5.3.8	「プロパティ用語」クラス	184
表 6.5.3.9	「プロパティ用語」の推奨プロパティ	185
表 6.5.3.10	「コードリスト」クラス.....	186
表 6.5.3.11	「コードリスト」の推奨プロパティ	186
表 6.5.3.12	「コード」クラス	188
表 6.5.3.13	「コード」の推奨プロパティ	188
表 6.5.3.14	「利用者意見」クラス	190
表 6.5.3.15	「利用者意見」の推奨プロパティ	190
表 6.6.2.1	機密性、完全性、可用性の説明.....	192
表 6.7.1.1	情報システムに求められるスケーラビリティの観点	196
表 6.7.2.1	語彙データベースに対するセキュリティ要件の整理結果.....	197
表 7.1.1.1	NIEM 運用組織名称とその役割	201
表 7.1.1.2	NIEM における語彙の公開(リリース)の種別.....	203
表 7.1.1.3	NIEM HLVA のステップの概要.....	204
表 7.1.2.1	ISA 運用組織名称とその役割	206
表 7.1.3.1	Schema.org 運用組織名称とその役割.....	208
表 7.1.4.1	運用組織体制の比較.....	210
表 7.1.4.2	運用プロセスの比較.....	211
表 8.2.2.1	【参考】 ISO11179-4 のデータ定義に関する要件 (抜粋)	222

1. 共通語彙基盤の事業概要

1.1 背景

我が国において、電子行政における情報連携及び情報公開の基盤として、データ連携に不可欠な語彙の相互運用性を確保するための基盤を整備し、行政コストの削減及び民間によるデータ活用の促進、さらにはデータ活用による経済活性化を図ることが急務となっている。内閣官房情報通信技術総合戦略室は、(1)行政の透明性・信頼の向上、(2)国民参加・官民協働の推進、(3)経済の活性化・行政の効率化の3つを目的とし、公共データの活用促進のため基本戦略として、2012年7月に「電子行政オープンデータ戦略」を決定した。ここでは、公共データを二次利用可能な形で提供することが重要課題であると指摘し、それを実現するため、電子行政システムにおいて、データ形式・構造等の標準化を推進することを求めている。この戦略の下、2012年度に経済産業省等で実施された調査の結果、電子政府におけるデータ連携やデータ公開を効率的に行うためには、機関・分野横断的に検索可能であり、自動的再利用やマッシュアップを効率的に行えるよう、データ中で用いられる語彙の意味やデータ作成ルール等を整備し、公開することが不可欠であることが明らかとなった。

上記の調査結果等を受け、「世界最先端 IT 国家創造宣言」(2013年6月14日閣議決定)では、公共データの民間開放(オープンデータ)推進及び利便性の高い電子行政サービスの提供を目的として、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(以下、「共通語彙」という。)の基盤構築に取り組むことが明記されている(以下、このような基盤を「共通語彙基盤」という)。

共通語彙基盤は、行政機関等が情報の電子的な公開や交換を行う際、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) 表記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、データの構造、用語の意味定義、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である。共通語彙基盤を整備することにより、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ようになり、また、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ことが期待される。そして、再利用性の高い公共情報が普及することにより、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、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等が期待される。¹

¹ 1.1の本文は、「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の仕様書より抜粋。

<http://www.ipa.go.jp/about/kobo/tender-20130902.html>

1.2 共通語彙基盤の全体像

共通語彙基盤は、共通語彙基盤を支えるデータベース（以下、「語彙データベース」という。）と、それと連携するツール群により構成される。

語彙データベースは、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（以下、「語彙データ」という。）を格納し、外部へ提供する機能を有するデータベースである。

ツール群は、再利用性の高いデータを作成するためのデータ構造の設計、既存データを再利用性の高いデータへの変換等の業務を支援し、語彙データベースと密接に連携することにより、共通語彙の利用における諸課題を解決し、共通語彙基盤の活用を促進するものである。

具体的には、①データ構造設計支援のためのツール、②データ入力支援のためのツール、③データの意味確認のためのツール、④情報連携の際等にデータを変換するためのツール、⑤語彙データを作成するためのツール等が考えられる。

以下に、共通語彙基盤の全体像について示す。

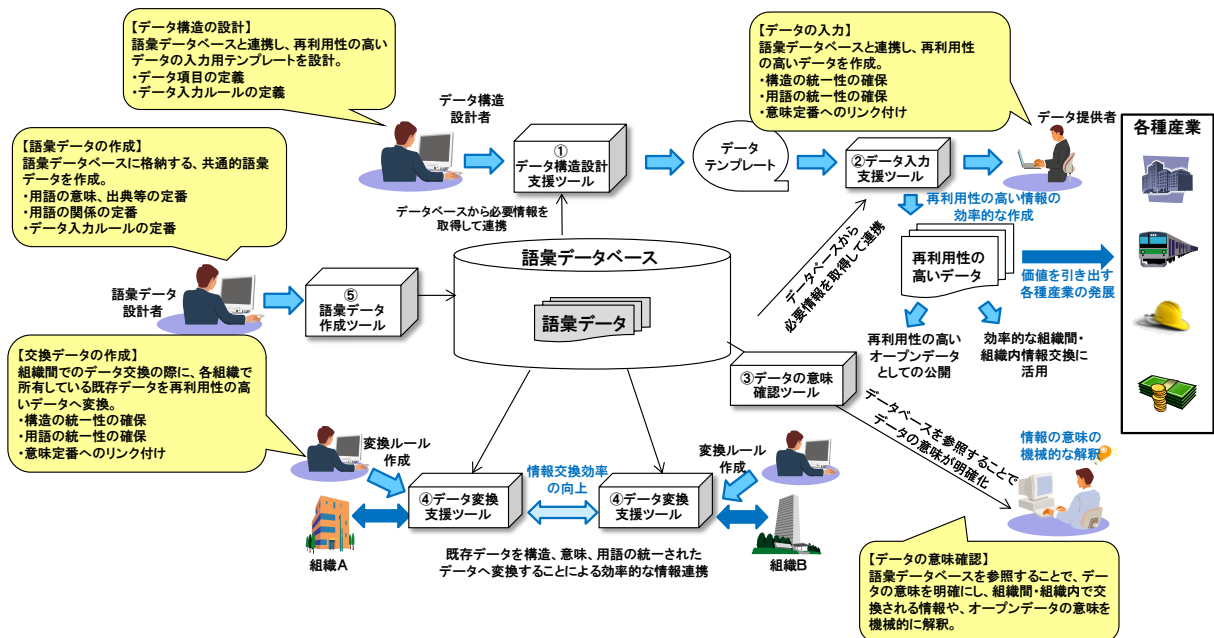


図 1.2.1 共通語彙基盤の全体像

1.3 「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（以下、「本事業」という。）」の概要

官民にわたる多くの組織がオープンデータを活用し、分野を超えた情報交換を行うには、個々の用語について表記・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必要がある。

例えば、「市役所」という用語であれば、まず「市役所」「市庁舎」など複数ある表記の統一を図ると共に、用語の属性として、「建物」、「行政施設」、「防災拠点」などを与える必要がある。

また、位置情報については、住所や緯度経度など複数の定義が既に存在している状況である。

これら多くの情報を、統一フォーマット（構造）でデータ化し、誰もが共通利用できるようにするため、本事業では、共通語彙基盤の導入に向け、専門家や将来の共通語彙基盤の利用者として想定される関係者を交えた客観的検討を通じて、試行的なサンプル語彙の収集・整備や、共通語彙基盤を支える語彙データベースとツール群についての要求事項を洗い出した。

また、実運用を想定したパイロットシステムの開発・運用を通じて、共通語彙基盤導入に向けた課題を洗い出した。

さらに、上記検討結果や課題を踏まえ、諸外国の先進事例の調査を行った上で、実運用に向けた課題や提言を含めて概念モデル書としてまとめた。

1.4 「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（以下、「ツールプロジェクト）」の概要

ツールプロジェクトでは、語彙データベースと連携し、再利用性の高いデータの作成、データ構造の設計、既存データの再利用性の高いデータへの変換等の業務を支えるツールを試作した。

また、試作したツールを本事業で整備したパイロットシステムと連携し、実際にデータの作成を試行して評価した結果やツールの要求事項について、概念モデルとしてまとめた。

1.5 用語の定義

本書で表現する言葉の定義について、以下に示す。

表 2.1.1.1 用語の定義

項番	用語	定義
1	共通基盤システム	IPA が運営するシステムインフラで、仮想化技術等を利用し、サーバハードウェアの統合・運用の共通化等を実現したシステム。
2	用語	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)をもつ場合がある。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、同一の意味をもつ用語が、異なる表記で示される場合がある。尚、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。
3	語彙	用語の集合。
4	語彙データ	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の総称。
5	語彙データベース	「語彙データ」を格納し、外部へ提供する機能を有するデータベース。
6	ドメイン	特定の業務領域。
7	ドメイン語彙	特定の業務領域(ドメイン)で用いられる用語の集合。
8	ドメイン共通語彙	ドメイン語彙の内、他のドメインでも参照される主要な語彙。ドメイン間の情報連携やオープンデータ公開等に利用される。本事業におけるサンプル収集・整備の対象。
9	ドメイン固有語彙	ドメイン語彙の内、各ドメインで専門業務等での情報交換・連携に特化した語彙。ドメイン内の各種標準や既存システムの語彙等を含む。本事業においては参照のみにとどめ、サンプル収集・整備の対象とはしない。
10	コア語彙	複数の業務領域(ドメイン)で共通に用いられる用語の集合。将来的には、語彙は、まずドメイン毎に整理され、後に各ドメインに共通なものとして洗い出されたものをコア語彙としてまとめてゆくことが想定される。コアボキャブラリと同義である。本概念モデル書では、共通語彙基盤において利用されるコア語彙は「IMI コア語彙」と表記する。
11	IMI (Infrastructure for Multilayer Interoperability)	経済産業省の「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)」で提案された、公共情報交換のための標準スキーム。情報に係る意味レベルを含む高い相互運用性を目指す。NIEM(米国)、Joinup(欧州)及び国内の事例を参考にして提案された。共通語彙基盤と同義である。 http://imi.ipa.go.jp/ns/200/
12	NIEM (National Information Exchange Model)	米国において、行政組織(連邦政府・州政府・地方政府など)の間で、意味レベルの互換性を向上させ正確かつ効率的な情報交換を実現するための基盤。語彙の型定義等を XML Schema として提供するとともに、その利用・追加・拡張などを実施するための方法・手順・規則を体系的に整備。 https://www.niem.gov/

項番	用語	定義
13	ISA(Interoperability Solutions for Public Administration)	2010年1月に欧州委員会情報科学総局下に設置された組織/プロジェクト。政府情報システムの相互運用性拡大に係る施策立案を担当。Core Business Vocabulary, Core Location Vocabulary, Core Person Vocabulary, Core Public Service Vocabularyの4つのコア語彙を整備・公開している。 http://ec.europa.eu/isa/
14	Joinup	ISAプロジェクトの一環として推進されている、欧州の政府情報システムにおける情報の、意味レベルの相互運用性や再利用性を実現するためのプロジェクト。語彙の定義等の整備、それを活用するためのツール等の整備を進めている。 https://joinup.ec.europa.eu/
15	XML(Extensible Markup Language)	データの意味や構造を記述するためにW3C(World Wide Web Consortium)で仕様が決められた言語。 http://www.w3.org/TR/xml/
16	XML Schema	XML文書の構造を規定するためにW3Cで策定されたスキーマ言語。 http://www.w3.org/TR/xmlschema-1/
17	RDF(Resource Description Framework)	主語、述語、目的語の3つの要素(トリプル)で情報間の関係を記述する、W3C(World Wide Web Consortium)が勧告する記法。 http://www.w3.org/TR/rdf11-primer/
18	API(Application Program Interface)	サーバ等とクライアント間をHTTP(Hyper Text Transfer Protocol)プロトコルで結ぶインタフェース。
19	LOD(Linked Open Data)	RDF形式を活用し、データベース間でオープンに接続できるようにしたデータベース。
20	オープンソースライセンス	The Open Source Definition (http://www.opensource.org/osd)の規定した定義に合致するものとして同サイトに掲載されたライセンスのいずれかとする。
21	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」事業。
22	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」事業。
23	概念モデル	ツールプロジェクトを通じて得られたドメインごとの語彙データ定義やツールの要求事項などをまとめたもの。
24	概念モデル書	概念モデルを記載した文書。
25	パイロットシステム	DBプロジェクトでパイロットとして構築・運用される語彙データベース。
26	RDFa	RDF形式のデータをHTML文書の中に埋め込む記法。HTMLの構造を利用してアノテーションとして記述する。
27	オープンデータ	オープンライセンスに基づいて公開されたデータ。オープンライセンスとしてはCreative Commonのライセンスがよく利用される。 またデータの形式のオープンかつ再利用可能な形式(csv, rdfなど)が推奨されている。
28	データカタログ	複数ドメインのデータをライセンスやデータ管理者名等を整理したカタログとして公開したもの。日本政府のデータカタログサイト「 http://www.data.go.jp/ 」や経済産業省のオープンデータカタログサイト「 http://datameti.go.jp/ 」などが挙げられる。

項番	用語	定義
29	CMS	コンテンツ・マネージメント・システムの略。ホームページなどの Web コンテンツを構成するテキストや画像などのデジタル・コンテンツを管理し、配信などを行うシステムのこと。
30	schema.org	HTML ページに特定の形式でマークアップ(構造化データ)を埋め込むことで、検索エンジンに詳細情報(リッチスニペット)を表示させ、検索ユーザの利便性を向上させるための仕組み。構造化データの語彙及びフォーマットが定義されている。Google、Microsoft、Yahoo!の検索大手 3 社が共同で開発・サポートしていくことを発表している。 http://www.schema.org/
31	JSON(JavaScript Object Notation)	テキストベースのデータフォーマット。JavaScript のオブジェクト表記構文のサブセットとなっており、XML と比べると簡潔に構造化されたデータ記述であり、人間が理解しやすい。JavaScript だけでなく、様々なソフトウェアやプログラミング言語間におけるデータの受け渡しに使えるように設計されている。2006 年に「RFC 4627(http://www.rfc-editor.org/rfc/rfc4627.txt)」として公開されている。
32	JSON-LD	JSON で Linked Data を表現するための仕様。2014 年 1 月 16 日に W3C の勧告となった。

2. 共通語彙基盤整備に向けた取り組み

2.1 現状の課題整理

2.1.1 行政の業務向上とサービス向上、各種産業の発展のためには

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や、オープンデータとして公開された再利用性の高い情報の活用による各種産業の発展を推進するために、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ことと、行政機関などから公開される情報の意味の解釈が機械的に可能となる必要がある。

(1) データに関する意味の明確化と構造整備の必要性

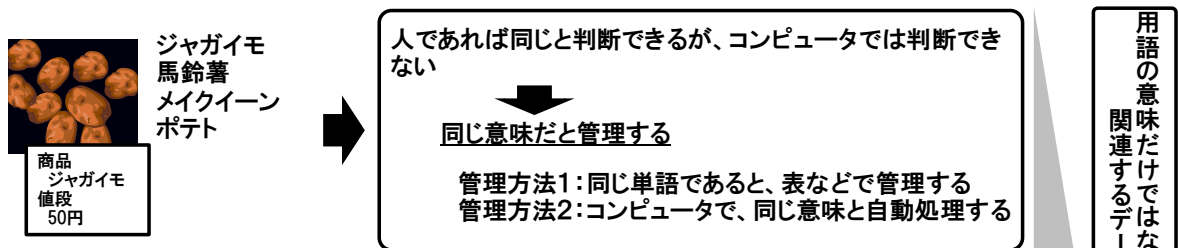
データ公開やデータ交換を行う際に、「じゃがいも」、「馬鈴薯」、「メイクイーン」や「ポテト」のように異なる表記の用語がばらばらに使用されていると、人であれば、同じ意味であると判断できるが、コンピュータでは、同じ意味の用語であると判断出来ず、それぞれ別の用語として処理してしまうことで、不都合が生じることがある。

また、同じ表記であっても、業界や部門によって意味が異なるような用語については、コンピュータでは同じ用語であると処理してしまい、不都合が生じることがある。

例えば、「水」という用語の場合、飲料水や蒸留水、消火用水等の複数の意味が存在しており、データに関する意味が明確になっていない場合、間違ったコンピュータ処理が行われてしまう可能性がある。さらに、飲料水や蒸留水、消火用水等の一つ一つの用語の意味を明確にただけでは、機械処理に適したものであるとは言えない。

そこで、用語の意味だけではなく、その用語で表現するものを、関連するデータの集まりとして表現する用語の構造化が必要となる。

① 違う表現の単語を同じ単語とみなす



② 同じ表現であるが意味が違うモノを明確にする

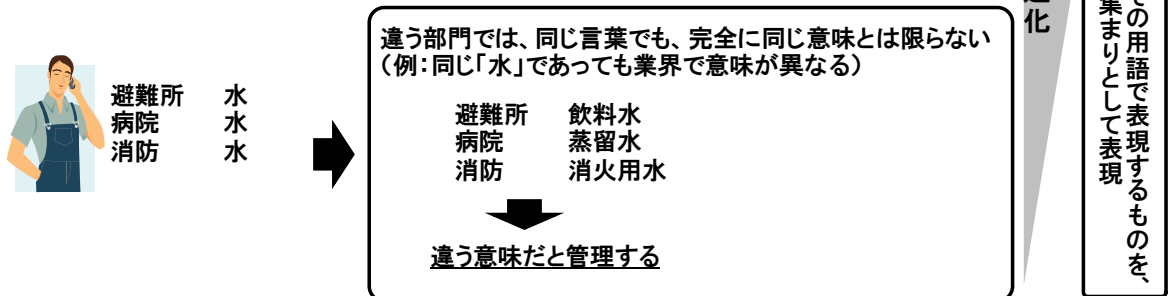


図 2.1.1.1 意味の明確化とデータ構造の整備の必要性

(2) 再利用性の高いデータ形式の整備

システム間連携によるデータ交換や、オープンデータとしてのデータ公開を行う場合のデータ形式は、それぞれの目的や用途によって、フラットな表形式やツリー形式、グラフ形式など、種々の形式が想定される。そのため、本事業では、CSV、XML、RDF 等、多様なデータ形式で利用できるように用語を整備する必要がある。

2.1.2 ツールプロジェクトの成果から抽出した現状の課題

(1) データの表現形式の共通化の必要性

神奈川県横浜市金沢区の「かなざわ育なび.net」で使用されるデータは、区職員が、基となるデータを入力し、委託業者が、入力されたデータを RDF 形式に変換することで生成されている。

区職員が入力するデータのうち、日付データは、ISO8601 準拠の YYYYMMDD の他、YYYY 年 MM 月 DD 日、昭和 YY 年 MM 月 DD 日など、多くの表現形式が混在しているため、データを入力する際に、混乱を招いてしまい、データの表現形式を頻繁に間違ってしまう。また、委託業者によるデータチェックや、区職員によるデータ修正に時間を要し、オープンデータの公開をスムーズに行うことが出来ていない。

そのため、多様なデータの表現形式が存在していても、データ入力時に混乱を招くことなく、オープンデータの公開をスムーズに行うためには、データの表現形式の共通化を図る必要がある。

(2) データの意味や関係性の明確化の必要性

現在、各自治体はオープンデータの提供に積極的に取り組んでいるが、各自治体が独自ルールの下で作成したデータを公開している例が少なくない。

例えば、島根県松江市の松江 SNMS では、観光情報やイベント情報で使用する用語に一意の意味を持たせるよう、オープンデータ化するデータは、「経済産業省 平成 24 年度事業 公共情報交換標準スキームの整備に関する調査研究」¹（以下、「平成 24 年度事業」という。）で整備された共通語彙である IMI コアボキャブラリ v1 を参考にして作成している。しかし、実際には、松江 SNMS のデータ登録者は、必ずしもオープンデータや IMI コアボキャブラリに精通しているわけではないため、独自に設定した項目が多数存在している。

また、「乳幼児等医療費助成制度」に関する情報提供について、川口市の Web サイトでは、「支給の方法」という項目の中に医療費の支払い免除の条件が示されている。一方、千葉市の Web サイトでは、「助成の内容」という項目の中に保護者の負担額が示されている。このように、同じ行政制度に関するデータであっても、自治体ごとに整理方法が異なるため、項目名やデータの構造などが異なっている。

¹ 「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」

<http://datameti.go.jp/data/dataset/report-002-2012/resource/318c2a2d-be84-42e7-be44-50a725b135ea>

上記のように、自治体ごとに独自の項目が多数存在し、項目名やデータの構造が異なっている状態で、データが作成・公開される状況下では、異なる自治体間でのデータ交換の際、それぞれの項目や構造の対応関係をお互いに確認し合う必要が生じる。また、自治体から公開されたオープンデータを、民間事業者や個人などが活用する場合にも、自治体ごとに項目や構造を確認しながらデータを収集しなければならず、活用が困難となる可能性がある。

そのため、データの交換や公開、活用をスムーズに行うためには、自治体ごとに整理された項目の意味や関係性を明確化する必要がある。

(3) 多様なデータ形式の整備の必要性

島根県松江市の松江 SNMS や、神奈川県横浜市金沢区の「かなざわ育なび.net」では、RDF 形式のデータを対象に取り扱っている。一方、他自治体におけるデータ公開やデータ交換を行う際のデータ形式には、CSV、XML、JSON 等、多様なデータ形式が採用されている。

これらのデータ形式は、データ交換先の組織のデータ特性やデータ生成時のファイル形式、技術的潮流などに依存して決定されるため、データ形式を統一することは現実的でない。また、データ形式の異なる自治体間でのデータ交換やデータ連携の際には、データの項目の意味や項目間の関係性が明確化されていたとしても、データ形式の変換作業などが必要となるため、スムーズなデータ交換やデータ連携が出来ない。

そのため、様々なデータ形式が採用されている状況下において、データ交換やデータ連携をスムーズに行うためには、データの項目の意味や項目間の関係性を明確化する以外に、多様なデータ形式に対応する必要がある。

2.2 本事業に先行する共通語彙基盤整備についての取組

「2.1.1 行政の業務向上とサービス向上、各種産業の発展のためには」に示した通り、データに関する意味の明確化や、構造の整備、再利用性の高いデータ形式の整備を推進するため、本事業に先行し、平成 24 年度事業として、海外や国内の既存の取り組みについて調査が行われ、日本の現状を踏まえた語彙群および関連フレームワークや、国内の基本となる体系化された共通語彙基盤(IMI)のコア語彙 v1 が整備された。

2.3 本事業の実施内容

平成 24 年度事業を基に、試行的なサンプル語彙の収集・整備や、共通語彙基盤を支える語彙データベースとツール群についての要求事項の洗い出し、実運用を想定したパイロットシステムの構築・運用を通じて、課題を洗い出し、その課題に対する考察や解決策の検討、実運用に向けた提言をまとめた。

2.3.1 検討会の設置

本事業の推進にあたっては、専門家や将来の共通語彙基盤の利用者として想定される関係者を交えた客観的検討を行うため、各種の有識者で構成されるコア語彙検討会、語彙データベース検討会、ドメイン語彙検討会を設置した。

コア語彙検討会では、共通語彙基盤の語彙データベースが格納・提供する語彙の中で、複数の業務領域(ドメイン)で共通に用いられる用語の集合である「コア語彙」について、用語の見直しや追加に関する検討に加え、コア語彙のデータ構造に関する検討を実施した。

語彙データベース検討会では、語彙データベースの機能や API、また関連するツール群についての検討や、共通語彙基盤の運用についての検討を実施した。

ドメイン語彙検討会では、本事業で試行的に選定した 4 つのドメインである地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインのそれぞれについて、ドメイン語彙の収集・整備プロセスについて検討するとともに、ドメイン毎に語彙のサンプル(ドメイン語彙サンプル)を試行的に整備した。

語彙データベース検討会は、データ構造をドメイン語彙検討会へ提供し、ドメイン語彙検討会で検討した内容をデータ構造に対するニーズの提示やコア語彙の修正依頼として、相互に連携しながら、検討を進める関係とした。

以下に、本事業で設置した検討会の概要について示す。

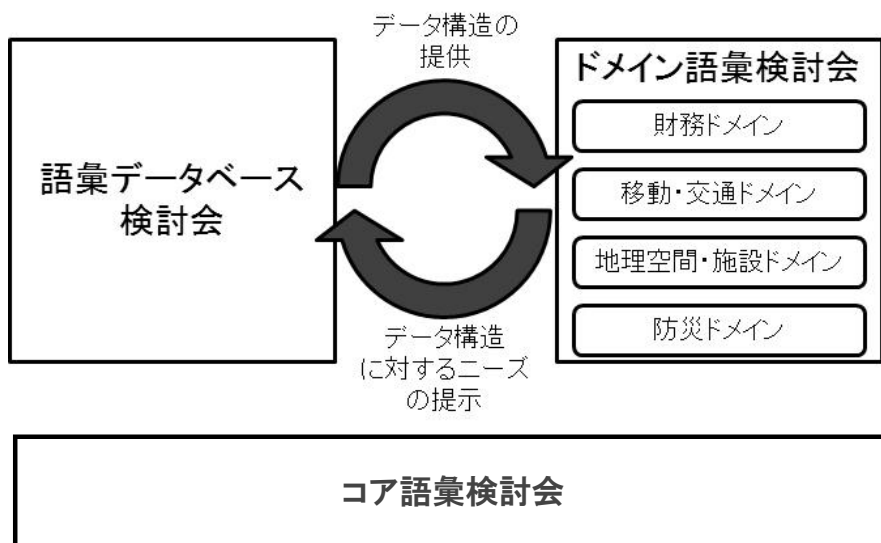


図 2.3.1.1 本事業で設置した検討会

2.3.2 語彙の整備

(1) コア語彙のブラッシュアップ

コア語彙検討会を通じて、平成 24 年度事業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した。

また、既に定義された「用語」を使用して、さらに別の新たな「用語」を定義できるといった、柔軟性のあるデータ構造を検討した。

また、データ構造の検討にあたっては、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（米国 NIEM、欧州 ISA 等）との整合性、既存データ（UN/CEFACT 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、データ入力ルール及びその記述法について留意した。（詳細は、「4. コア語彙の整備」を参照）

(2) ドメイン語彙サンプルの整備

将来の中央官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について、「電子行政オープンデータ推進のためのロードマップ」¹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した結果、地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインを試行的に選定し、対象分野の専門化を交えた十分な検討の場として、ドメイン語彙検討会を通じて、検討したデータ構造に従い、リストアップした用語をドメイン語彙サンプルとして整備した。（詳細は、「5. ドメイン語彙サンプルの整備」を参照）

¹ 電子行政オープンデータ推進のためのロードマップ（平成 25 年 6 月 14 日、高度情報通信ネットワーク社会推進戦略本部決定）、www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryous3.pdf

2.3.3 語彙データベースの要求事項の整理

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上で、語彙データベースについて、語彙データベース検討会を通じて、利用シーンの洗い出しやツール群についての検討、APIについての検討、データ構造の拡充、セキュリティ要件の整理やスケーラビリティ要件の整理を行った。（詳細は、「6. 語彙データベースの要求事項」を参照）

(1) 利用シーンの想定

「1.2 共通語彙基盤の全体像」で示したデータ構造設計に関わる利用シーンや、既存データを再利用性の高いデータへ変換するためのデータ入力の利用シーン、データの意味確認のための利用シーン等を初期仮説として設定した。そして、語彙データベース検討会、ドメイン語彙検討会を通じて、専門家や将来の共通語彙基盤の利用者として想定される関係者を交えた客観的検討を行いつつ、ツールプロジェクトでの実証や、海外聴取で得た意見や知見を通じて想定利用シーンの整理を行った。

以下に海外聴取先について示す。

表 2.3.3.1 海外聴取先①

日時	2014年7月17日（木） 11:30～14:45
面会組織	W3C
目的	・ IMI 語彙データベースのデータ構造や IMI コア語彙等に関する意見聴取 ・ W3C における語彙の標準化動向調査

表 2.3.3.2 海外聴取先②

日時	2014年7月18日（金） 10:00～12:00
面会組織	EC ISA オフィス
目的	・ IMI 語彙データベースのデータ構造や IMI コア語彙等に関する意見聴取 ・ ISA における語彙およびオープンデータ関連活動の調査

表 2.3.3.3 海外聴取先③

日時	2014年7月21日（月） 14:00～16:00
面会組織	Data.gov.uk
目的	・ IMI 語彙データベースのデータ構造や IMI コア語彙等に関する意見聴取 ・ data.gov.uk における語彙およびオープンデータ関連活動の調査

表 2.3.3.4 海外聴取先④

日 時	2014年7月22日(火) 13:00~15:00
面会組織	Schema.org
目 的	・ IMI 語彙データベースのデータ構造等に関する意見聴取 ・ schema.org における語彙関連活動の調査

表 2.3.3.5 海外聴取先⑤

日 時	2014年7月24日(木) 10:00~12:00
面会組織	OMG、NTAC
目 的	・ IMI 語彙データベースのデータ構造や IMI コア語彙等に関する意見聴取 ・ OMG における語彙関連活動の調査

表 2.3.3.6 海外聴取先⑥

日 時	2014年7月24日(木) 13:00~14:00
面会組織	Best of NIEM 受賞プロジェクト「EOC-I」
目 的	・ Best of NIEM 受賞プロジェクト「EOC-I」における NIEM 活用事例の調査

表 2.3.3.7 海外聴取先⑦

日 時	2014年7月24日(木) 16:00~17:00
面会組織	Best of NIEM 受賞プロジェクト「ASAP」
目 的	・ Best of NIEM 受賞プロジェクト「ASAP」における NIEM 活用事例の調査

表 2.3.3.8 海外聴取先⑧

日 時	2014年7月25日(金) 9:00~10:00
面会組織	ISE、DHS、NTAC、NBAC、OMG
目 的	・ 日本政府の相互運用性フレームワーク政策と IMI プロジェクトの状況に関する意見聴取 ・ ISE, NIEM のビジョンや動向に関する調査

表 2.3.3.9 海外聴取先⑨

日 時	2014年7月25日(金) 13:00~15:00
面会組織	NTAC、NBAC、OMG
目 的	・ IMI 語彙データベースのデータ構造や IMI コア語彙等に関する意見聴取

(2) ツール群の想定

上記「(1) 利用シーンの想定」にて整理した語彙データベースの想定利用シーンの検討結果やツールプロジェクトでの実装、海外聴取で得られた利用者の業務を支援するための汎用的なツール群を想定ツール群として整理した。

(3) 実装機能および API の策定

上記「(2) ツール群の想定」にて洗い出したツール群それぞれについて、実装する機能及びそれを実現するために必要な API を特定し、その一案を策定した。

(4) データベースとして実装するためのデータ構造の拡充

上記「(3) 実装機能および API の策定」にて策定した API 案について、個々の API を実装するために語彙データベースが備えるべきベータ構造を特定し、「4.コア語彙の整備」にて整備したデータ構造案に対して、データ構造の拡充を検討した。

データ構造の拡充にあたっては、API 実装に必要な機能と、RDF ストアでの実装を行う上でのデータ構造について検討を実施した。

(5) セキュリティ要件、スケーラビリティ要件の整理

上記「(3) 実装機能および API の策定」および「(4) データベースとして実装するためのデータ構造の拡充」にて策定した API 案およびデータ構造案をシステムとして実現する際のセキュリティ要件、スケーラビリティ要件について、整理した。

2.3.4 共通語彙基盤の運用についての検討

共通語彙基盤の導入に向け、ドメイン語彙の拡充や、コア語彙のバージョンアップ等、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った。検討にあたっては、データベース検討会を通じた議論や、海外聴取を通じて得られた先進事例の知見を活かし、運用体制の他、メンテナンスのプロセスや必要な検討会も含めて、実現可能性が高く持続性の期待できる方法を検討した。(詳細は、「7. 本事業にて整備した共通語彙基盤の運用についての検討」を参照)

2.3.5 パイロットシステムの構築と運用

ツールプロジェクトに連携する語彙データベースとして、ツールプロジェクトの試行に必要な機能を実装し、APIとして提供するパイロットシステムを構築した。

また、ツールプロジェクトにて準備したドメイン語彙を都度追加・更新しながら、パイロットシステムの運用を実施した。

パイロットシステムには、IMI コア語彙 v1 及び、ツールプロジェクトにて準備したドメイン語彙を格納した。

2.3.6 概念モデル書の取りまとめ

共通語彙基盤の導入に向け、諸外国の先進事例の調査、専門家や将来の共通語彙基盤の利用者として想定される関係者を交えた客観的検討を行った。試行的なサンプル語彙の収集・整備の結果や、共通語彙基盤を支える語彙データベースとツール群についての要求事項の洗い出し結果、実運用を想定したパイロットシステムの開発・運用を通じた共通語彙基盤導入に向けた課題を洗い出した結果を整理するとともに、その課題に対する考察や解決策の検討、実運用に向けた提言を概念モデル書としてまとめた。

3. 本事業における語彙の設計方針

本事業では、語彙データベースに格納される語彙データの整備およびそれらのデータ構造に関する要求事項の洗い出しに先立って、平成 24 年度事業において整備された IMI コアボキャブラリ v1¹（以下、「IMI コア語彙 v1」という。）をベースに、当該調査研究および本事業を通じて明確となった課題を整理した上で、共通語彙基盤における語彙の設計方針案を策定した。

3.1 先行事例：IMI コア語彙 v1

平成 24 年度事業においては、米国 NIEM の枠組み（命名規則、運用方法、データ型、交換方法）を踏襲し、欧州 ISA Core Vocabulary を参照しながら、コア語彙の第一版である IMI コア語彙 v1 の整備が実施された。

¹ 共通語彙基盤検討会議（平成 25 年 8 月 21 日開催）<<http://goikiban.ipa.go.jp/node56>> にて、「参考資料 1 IMI コアボキャブラリ」として公開されたもの。

3.1.1 IMI コア語彙 v1 の特徴および整備方針

IMI コア語彙 v1 の主な特徴および整備方針について、以下に抜粋する。

表 3.1.1.1 IMI コア語彙 v1 の整備方針に関する主な特徴

項番	分類	IMI コア語彙 v1 において取られた整備方針
1	コアとドメインの分類	複数のドメインにまたがって利用される基本的な語彙をコア語彙として定義した。具体的には、ISA Core Vocabulary、NIEM を参考に、共通的に広く利用されると見込まれる語彙を対象とした。 それ以外の語彙をドメイン語彙として定義した。(2層構造を採用した)
2	コア語彙・ドメイン語彙の運用	コア語彙・ドメイン語彙ともに、段階的な導入を想定し、定義した。ドメインを拡張・増加する過程において、複数ドメインで利用される語彙は順次コアに追加し、拡充される仕組みを想定した。
3	用語の表記	英語名については、基本的には NIEM に使用されている英語名称を利用し、定義した。
4	用語の整備範囲	「人」「組織」「場所」「住所」「証明」については、NIEM、ISA Core Vocabulary 双方で定義されているデータタイプであり、基本的には NIEM および ISA Core Vocabulary の双方で定義されている語彙を抽出した。 それ以外の用語については、基本的には NIEM を踏襲しつつ、中間標準レイアウトや選定ドメインの語彙を参考に、日本で必要とされる語彙を追加した。
5	用語の日本語名	中間標準レイアウトや APPLIC による標準仕様で対応するものがある語彙については、それらの語彙に含まれる日本語名を利用した。
6	用語の構造	用語の構造については、文書が整備され、活用が進んでいる NIEM を踏襲した。
7	用語「物」「有体物」について	物型については、今回整備する防災ドメインで利用される有体物型 (TangibleItem Type) まで定義した。 語彙については、現状の防災のユースケースでは不要と考えられる項目は排除した。

項番	分類	IMI コア語彙 v1 において取られた整備方針
8	日本独自の用語の追加	<p>NIEM や ISA Core Vocabulary が利用される欧米では「出生地」が一般的であるが、中間標準レイアウトや APPLIC による標準仕様において定義されている「本籍」を追加した。</p> <p>また、日本では欧米とは異なり、カタカナやローマ字といった文字が存在することを受け、対応したカナ姓やローマ字姓などの項目を追加した。</p>
9	NIEM 独自スキーマの参照	<p>NIEM で独自に定義されているスキーマの一部(XML Schema の拡張 [niem-xsd]や、基礎的な構成要素を集めた[structures]など) については、IMI コア語彙 v1 でもそのまま参照・利用した。</p>
10	外部コードリストの参照	<p>国名コードや各種単位系のコードなど、コア語彙の中で利用するコードリストについては、NIEM に含まれるスキーマを直接参照した。具体的には、NIEM2.1 に含まれる ISO4217、ISO639-3、ISO3166、UNECE、NGA (National Geospatial Agency)、TWPDES(Terrorist Watchlist Person Data Exchange Standard)などのコードリストを表現した XML Schema 形式のスキーマファイルに含まれる型を直接参照・利用した。</p>

3.1.2 IMI コア語彙 v1 の用語の全体像

IMI コア語彙 v1 に含まれる用語の全体像を示すため、その意味あるいは概念の一部として、他の用語との関係をもつ用語について、以下に抜粋する。

表 3.1.2.1 IMI コア語彙 v1 に含まれる主要な用語一覧

項番	分類	項目説明
1	人型	人の情報を表現するためのデータ型。
2	氏名型	氏名を表現するためのデータ型。
3	住所型	住所を表現するためのデータ型。
4	構造化住所型	要素化された住所を表現するためのデータ型。
5	方書型	方書を表現するためのデータ型。
6	連絡先型	連絡先を表現するためのデータ型。
7	電話番号型	電話番号を表現するためのデータ型。
8	組織型	組織を表現するためのデータ型。
9	組織関連型	組織の関連を表現するためのデータ型。
10	関連型	関連を表現するためのデータ型。
11	系列型	組織の関連を表現するためのデータ型。
12	場所型	場所を表現するためのデータ型。
13	証明型	証明を表現するためのデータ型。
14	物型	物を表現するためのデータ型。
15	有体物型	有体物を表現するためのデータ型。
16	輸送機関型	輸送機関・乗り物を表現するためのデータ型。
17	航空機型	航空機を表現するためのデータ型。
18	自動車型	自動車を表現するためのデータ型。
19	船舶型	船舶を表現するためのデータ型。
20	輸送関連型	輸送機関と人、貨物などを関連付けるためのデータ型。
21	建物型	建物のためのデータ型。
22	施設型	特別なサービスをする建物や場所や構造のためのデータ型。

項番	分類	項目説明
23	建物構造型	(未定義)
24	活動型	活動を表現するためのデータ型。
25	測定単位型	測定単位及び方法を表現するためのデータ型。
26	数量型	数量を表現するためのデータ型。
27	容量型	容量を表現するためのデータ型。
28	面積型	面積を表現するためのデータ型。
29	重量型	重量を表現するためのデータ型。
30	長さ型	長さを表現するためのデータ型。
31	物品価値型	物品の価値を表現するためのデータ型。
32	金額型	金額を表現するためのデータ型。
33	実体型	実体を表現するためのデータ型。
34	状況型	状況を表現するためのデータ型。
35	期間型	期間を表現するためのデータ型。
36	日付型	日付を表現するためのデータ型。
37	スケジュール型	施設等のオープン日とオープン時間を表現するためのデータ型。
38	経緯度座標系型	経緯度座標系を表現するためのデータ型。
39	UTM 座標系型	UTM 座標系を表現するためのデータ型。
40	MGRS 座標系型	MGRS 座標系を表現するためのデータ型。
41	緯度型	緯度を表現するためのデータ型。
42	経度型	経度を表現するためのデータ型。
43	緯度値型	緯度を表現するためのデータ型。
44	経度値型	経度を表現するためのデータ型。
45	緯度値単純型	緯度を表現するための単純型。-90 から 90。
46	経度値単純型	経度を表現するための単純型。-180 から 180。
47	分型	分を表現するためのデータ型。

項番	分類	項目説明
48	秒型	秒を表現するためのデータ型。
49	分単純型	分を表現するための単純型。0 から 60。
50	秒単純型	分を表現するための単純型。0 から 60。
51	パーセンテージ型	パーセンテージを表現するためのデータ型。
52	パーセンテージ単純型	パーセンテージを表現するための単純型。0 から 1。
53	固有名型	固有名を表現するためのデータ型。
54	テキスト型	テキストを表現するためのデータ型。
55	カタカナテキスト型	カナ文字列を表すデータ型。データ型についての ガイドライン（案） http://www.soumu.go.jp/denshijiti/pdf/051201_08.pdf の「フリガナ用の片仮名文字列」用のパターンに準拠。

3.1.3 IMI コア語彙 v1 のデータ構造

IMI コア語彙 v1 で採用されている、用語のデータ構造について、以下に概要を示す。
 前述の通り、本データ構造は米国 NIEM の枠組みを踏襲して整備されたものである。

表 3.1.3.1 IMI コア語彙 v1 のデータ構造

項番	項目	概要	利用例
1	項目名 (Type/Sub-properties)	用語の名称を日本語で表記したもの。	姓名
2	項目名 (エントリー名)	用語を一意に特定するための識別子。 (接頭語+項目名)	ic:氏名_姓名
3	英語名	用語の名称を英語で表記したもの。	FullName
4	データタイプ	用語の取りうる型を日本語で表記したもの。	ic:テキスト型
5	データタイプ(英語)	用語の取りうる型を英語で表記したもの。	ic:TextType
6	cardinality	用語が要素の最小出現回数と最大出現回数。	0..1
7	項目説明	用語の意味を日本語で記述したもの。	氏名(姓、名)。
8	項目説明 (英語)	用語の意味を英語で記述したもの。	Full name of a Person
9	Mapping to NIEM	NIEM の用語で対応するものがあれば示す。	nc:PersonFullName
10	Mapping to ISA	ISA Core Vocabulary の用語で対応するものがあれば示す。	fullName

3.2 IMI コア語彙 v1 の整備において明確となった課題

平成 24 年度事業では、IMI コア語彙 v1 について、以下のような残課題が提示された。

表 3.1.3.1 平成 24 年度事業で示された、IMI コア語彙 v1 に関する残課題¹

項番	項目	概要
1	日本語名の定義	語彙自体は、NIEM 及び ISA joinup から抽出し、日本語名は中間標準レイアウトや APPLIC による標準仕様で該当する語彙がある場合にはその名前を利用した。しかしながら、これらの仕様に含まれていない語彙も多い。そのため、日本語名には各分野で利用されている語彙を詳細に調査した上で、精査が必要と考えられる。
2	カタカナやローマ字、漢字の制限（文字コード）	英語圏ではアルファベットが基本であるが、日本ではかな、カタカナ、漢字、ローマ字、英数字等、多くの文字が存在する。また、氏名や行政文書では多様な漢字も使われており、それらの文字に対応したデータの定義を行う必要がある。NIEM では XML Schema を用いるため、文字セットの制限を行うことは可能であるが、拡張を行うには検討が必要である。
3	日本語名、英語名の利用	今回は、日本語名、英語名ともに定義し、XML Schema では日本語名を名前として扱った。XML Schema では二つの名前を併用できないため、日本語名を用い、annotation の形で、英語の説明等を付与している。日本語、英語のそれぞれの Schema を作成することも考えられるが、その場合、別々に管理を行う必要が生じ、互換性も損なう可能性が高い。一つの Schema で英語でも利用が可能となるようにするために、英語の表現方法について検討が必要である。

本事業では、日本における共通語彙基盤の整備に向け、これらの残課題を踏まえた上で、コア語彙の整備に関して解決すべき課題を再度、検討し、以下のような整理を実施した。

¹ 「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」

<http://datameti.go.jp/data/dataset/report-002-2012/resource/318c2a2d-be84-42e7-be44-50a725b135ea>

より引用

3.2.1 表記の明確化・関連付けについて

IMI コア語彙 v1 の用語は、NIEM 及び ISA Core Vocabulary から抽出された、英語表記の用語が中心であった。用語の日本語表記については、総務省中間標準レイアウトや APPLIC による標準仕様で該当する用語がある場合には、その名前が採用されたものの、これらの日本の仕様のない用語については、独自に和訳が行われた。そのため、各分野で利用されている用語については、詳細な調査を行う等、十分な精査が継続的に必要となる。

また、コア語彙の用語は、多くのドメインに共通する普遍的で汎用的な概念を記述するための用語であることから、その呼び名は利用主体によって様々であることが自然である。

例えば、IMI コア語彙 v1 には NIEM の Activity という用語をベースに「活動」という用語が整備されているが¹、オープンデータ提供などの利用シーンにおいては自治体内で企画・開催されるイベントの情報を記述するための「イベント」という用語としてこの用語を利用したいというニーズがあることが判明した。このような事情を鑑みるに、「活動」という表記で表現された概念に、「イベント」という別の表記を付与するなど、関連する用語同士の表記を何らかの形で関連付けるための仕組みが必要となると考える。

さらに、IMI コア語彙 v1 の用語は XML Schema での利用を前提としているため、構造をもった用語の表記には全て「型」という接尾語が付与されている（「人型」「氏名型」など）。「型」という呼び名は、XML Schema での複合型(ComplexType)などに相当する用語であることを明示するためには有用であるが、XML 以外、例えば RDF データモデルでの利用においては、用語の表記として適切でないと考えられる。

¹ IMI コア語彙 v1.00 では「イベント」という用語が「活動」とは別の用語として存在していたが、v1.01 へのリビジョンアップの際に削除されていた。

3.2.2 意味の明確化について

上述の通り、IMI コア語彙 v1 の用語の多くは、NIEM 及び ISA Core Vocabulary から抽出された用語をベースに、それらを日本語化するアプローチで整備されたため、用語の意味や用語の構造化について、さらなる検討の余地が残されている。

用語の意味に関しては、NIEM や ISA Core Vocabulary における英語での用語の説明文を和訳して取り込んだ経緯もあり、一部の用語については用語の説明文が不明確なものや、理解が容易でないものが存在している。

以上より、本事業では、用語の意味の明確化に関して、既存の用語それぞれについての内容を精査、見直しを実施する必要があると考える。

3.2.3 用語の構造化について

用語の構造化に関しては、「人」のように、関連する用語（例えば「氏名」「住所」等）を含む用語（構造をもつ用語）について、その構造が利用者の理解を妨げる可能性があるものや、実際の利用を想定した場合に、より最適化できる余地が残されているものが存在した。

例えば、IMI コア語彙 v1 には、病院や学校、公民館など、施設でもあり建物でもあるような概念を記述するために「施設型」と「建物型」という二種類の用語が用意されているが、実際の利用を想定した場合、両者を区別せずに扱う場面や、両者を区別した上でそれらの関係を明示したい場面などが考えられる。この場合、「施設型」と「建物型」の共通部分を括り出して「地点型」を作るなどといった対策をとることが望ましい。

以上より、本事業では、用語の構造化に関して、既存の用語の構造の精査、見直しを実施する必要があると考える。

3.2.4 用語の相互運用性について

IMI コア語彙 v1 の用語の内、米国 NIEM や欧州 ISA Core Vocabulary に対応する用語が存在するものについては、その対応の存在が明記されている。

このような用語間の関係性（マッピング）は、語彙の国際的な相互運用性を確保するために本質的で重要な情報である。実際、米国 NIEM と欧州 ISA との間では、「人」や「場所」などの基本的な概念について、NIEM の用語と ISA の用語との間の対応関係が整理され始めている。

本事業においては、共通語彙基盤の国際的な相互運用性をさらに拡大するため、各用語について、国際 EDI 標準である UN/CEFACT Core Components Library の用語や、民間の web サイトにおいて急速に利用が進んでいる Schema.org の用語等、国内外で普及している語彙に対するマッピングについても整備することが望ましいと考える。

3.2.5 用語のデータ構造について

IMI コア語彙 v1 の枠組み（命名規則、運用方法、データ型、交換方法）は、文書の整備状況や活用実績の多い米国 NIEM の枠組みをベースに整備された。NIEM は組織間の情報交換を主要な利用目的としているため、文書交換との親和性の高い XML に強く依存したデータ構造が採用されている。例えば、語彙データは XML Schema で記述され、XML Schema 独自の機構である substitutionGroup(代替グループ)の利用が前提として組み込まれている。

一方、共通語彙基盤が想定する利用シーンには、情報交換のみならず、公共データの民間開放(オープンデータ)も含まれる。こうした多様な利用シーンで語彙を効果的に扱うためには、XML に限らず RDF や JSON 等を含む多様なデータ形式をサポートすることが必要となる。

3.2.6 NIEM への依存について

IMI コア語彙 v1 の枠組みは、米国 NIEM の枠組みをベースに整備されており、語彙の一部には NIEM によって規定・公開されている型やコードリストを直接参照しているものが存在する。この結果、IMI コア語彙は NIEM に強く依存する関係となっており、利用にあたっては NIEM や XML の仕様に対する理解が必要になるとともに、RDF データへの利用等、NIEM が想定していない利用シーンでの活用の際にデータ形式の変換などに関して必要以上のコストが発生する可能性がある。

よって、NIEM を参照せずとも、IMI コア語彙単体で利用可能となるような独立性の確保が必要になると考える。

3.2.7 コードリストおよびコードの具体的な取扱いについて

IMI コア語彙 v1 では、NIEM で XML Schema として公開されているコードリストの一部が参照・利用されている。加えて、一部のコード（性別コード、国籍コード、住所コード、数量単位コード、容量単位コード、面積単位コード、年号コード）については、独自にコードリストを定義する方針で、コードリストおよびコードに関する整備が実施されている。

しかしながら、NIEM 以外のコードリストや、XML Schema 以外の形式で整備されたコードリストを参照するための仕組みについては言及されていない。さらに、上記の独自コードリストについては、型が示されているもの具体的なコードリストの実装結果については言及されていない。

そこで、本事業においては、外部で規定されているコードリストやコードを広く参照・利用するための方針と、内部で独自に規定したコードリストやコードを整備し、語彙データベースへ格納するための方針について、それぞれ具体的・詳細な検討を進めることが必要と考える。

3.3 共通語彙基盤における語彙の設計方針

前節に示した課題を解決するための方策を示すとともに、それらを集約・統合することで、共通語彙基盤の語彙の設計方針案を整備した。

3.3.1 語彙の整理（コア語彙とドメイン語彙、ドメイン共通語彙による分類）

3.1.1 IMI コア語彙 v1 の特徴および整備方針に記載している通り、IMI コア語彙 v1 のコアとドメインの分類については、NIEM の語彙構造を踏襲しているため、「人」、「場所」、「物」といった、複数のドメインにまたがって利用される基本的な語彙をコア語彙、それ以外の語彙をドメイン語彙とした 2 層構造にて定義された。

2 層構造の場合、あらゆるドメインで使用される用語も、複数のドメインで使用される用語もコア語彙として扱うため、コア語彙が肥大化する可能性があり、かつ、専門性が高いドメイン固有の用語まで、ドメイン語彙の整備対象となってしまう、ドメイン語彙も肥大化する懸念がある。

その結果、利用者からは、用語を選定することが困難になってしまい、管理者からは、語彙の管理が煩雑となってしまうことが懸念される。

そこで、本事業においては、今後、共通語彙基盤がドメイン語彙も含めて拡大した場合においても、コア語彙及びドメイン語彙の肥大化を防ぎつつ、複数のドメインを跨いだデータの相互運用性が確保することを目的として、ドメイン語彙をドメイン共通語彙とドメイン固有語彙に再定義し、3 層構造（コア語彙、ドメイン共通語彙、ドメイン固有語彙）を採用することとした。

以下に、語彙とドメインの関係を以下に示す。コア語彙は、どのドメインでも広く利用される普遍的な語彙である。ドメイン共通語彙は、他のドメインでも参照する主要な語彙である。ドメイン固有語彙は、各ドメインでの利用に特化した語彙である。

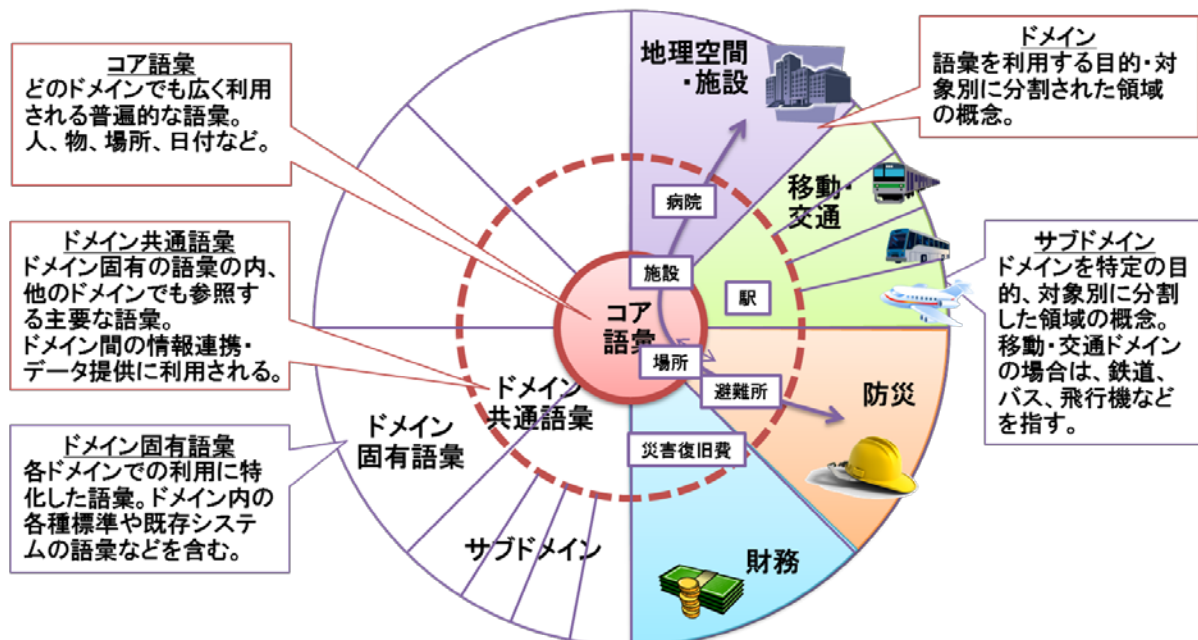


図 3.3.1.1 本事業における語彙の整理

3.3.2 語彙の位置付けについて

本事業における共通語彙基盤の利用シーンには、情報交換のみならず、公共データの民間開放(オープンデータ)も含まれる。こうした多様な利用シーンで語彙を効果的に扱うためには、XMLに限らず RDF や JSON 等を含む多様なデータ形式をサポートすることが必要となるが、IMI コア語彙 v1 の枠組みでは、XML のみのサポートに留まってしまうことから、共通語彙基盤の設計思想として、語彙を「構造をもった概念辞書」として位置づけた設計を行うこととした。

これにより、用語名、概念、概念間の関係、用語と概念間の関係や典拠情報などを十分に記述することが可能となる。

また、本事業において整備対象とするコア語彙は、利用者に対して用語の利用に制限を加えるものではなく、利用者間で情報の意味・構造・表記を共通理解するために用いる「典型(プロトタイプ)」としての位置づけとした。これは、本事業の目的が「情報交換」だけに閉じるものではなく、多種多様なデータ提供者・データ利用者が相互に関係し合う、「データ提供」をもスコープに入れたことに起因するものである。

用語の利用者が多種多様であり、語彙の整備時・提供時に詳細な利用方法まで特定することが困難である以上、厳密に Cardinality (出現回数) などの制限を課すことは利用者にとって必要以上の制約となりかねず、結果として共通語彙基盤やコア語彙の普及・促進を妨げる要因となってしまうことを避ける必要がある。

上記の方針に基づき、本事業で整備するコア語彙については、Cardinality の制約は本質的に必要な箇所を除いては、基本的に制限を課さない方針を採った。また、データの入ルールや文字セットに関する制限などについても、その実現方法を示すにとどめ、コア語彙への適用はいったん保留する方針を採用した。

3.3.3 用語の表記について

本事業では、用語の日本語名については、各分野で利用されている用語を詳細に調査した上で十分な精査を実施する方針を採用した。

また、関連する用語(同義語・類義語など)については、それらの関係を紐付けるためのデータ構造を実現するための検討を実施することとした。

さらに、IMI コア語彙 v1 において構造をもった用語に付与されていた「型」という接尾語については、XML 以外、例えば RDF データモデルへの適用においては適切でないため、用語の表記から排除することとした。

3.3.4 用語の意味および構造化について

本事業では、用語の意味および構造化に関して、共通語彙基盤の導入目的である意味・構造の明確化を実現するため、IMI コア語彙 v1 の用語それぞれについて、用語の意味（項目説明）の見直しと、各用語の構造化の見直しを実施することとした。

また、用語の意味が、法律や条例などの法的根拠に基づく場合や、辞書や標準規格など、広く参照されるような出典情報が存在する場合には、それらへの参照を記述できることが、意味の明確化にとっては望ましい。そこで、本事業では、各用語に対して出典情報を記述できるようなデータ構造を実現するための検討を実施することとした。

3.3.5 用語の相互運用性について

本事業においては、共通語彙基盤の国際的な相互運用性を拡大するため、コア語彙に含まれる各用語について、国際 EDI 標準である UN/CEFACT Core Components Library の用語や、民間の web サイトにおいて急速に利用が進んでいる schema.org の用語等、国内外で普及している語彙に対するマッピングについても整備する方針を採用した。また、それらのマッピング情報、すなわち、コア語彙に含まれる用語と、当該用語に対応する国内外の用語と、それらの用語がどのような関係にあるか（完全に一致する概念を指し示す用語であるのか、一方が他方よりも広義の関係にあるのか、など）を表現するためのデータ構造について、検討を実施することとした。

3.3.6 用語のデータ構造について

本事業では、情報交換に加えて、公共データの民間開放(オープンデータ)も含むような多様な利用シーンで語彙を効果的に扱うことを可能とするため、XML に限らず RDF や JSON 等を含む多様なデータ形式をサポート可能とする必要がある。

そこで、用語のデータ構造の表現方法を、特定のデータ形式に依存しない「論理形式」と、XML、RDF、JSON 等特定のデータ形式で表現した「物理形式」の二つの形式に分割する方針を採用した。語彙データの整備やメンテナンスは、基本的に「論理形式」に基づいて実施することで、語彙データの汎用性の向上を図る。一方で、用語を利用者がデータ交換・データ提供などに利用する際には、個々の利用シーンに応じて最適なデータ形式を選択し、「物理形式」として参照する。

3.3.7 用語の物理形式について

「物理形式」としては、IMI コア語彙 v1 や NIEM と同様、主にエンタープライズ分野で発生する XML ベースの情報交換を利用シーンとして想定し、W3C XML Schema 標準に準拠した「XML 表現」を引き続きサポートすることとした。

また、オープンデータの公開度を評価する手法の一つ「5 Star Open Data」にて、高い公開度(4 つ星)に位置づけられる RDF や、最高の公開度(5 つ星)に位置づけられる LOD (Linked Open Data) の実現の推進に寄与するため、W3C RDF Schema 標準に準拠した「RDF 表現」を新たにサポートすることとした。

加えて、近年 web にて利用が進んでいる JSON (Java Script Object Notation) 形式への対応として、Schema.org でも積極的に採用されている W3C 勧告 JSON-LD 形式を試験的にサポートする方針とした。JSON-LD は RDF データモデルの一表現であるため、「RDF 表現」のバリエーションとして扱うこととした。

さらに、自然言語で記述された人間のためのドキュメント（仕様書や表など）にもコア語彙を活用可能とするため、クラス用語やプロパティ用語を自然言語として記述するための「自然言語表現」をサポートする方針とした。

3.3.8 論理形式と物理形式の相互変換について

「3.3.6 用語のデータ構造について」で示したような、論理形式と物理形式の使い分けを、整合性を保ちながら実現するためには、両者に対応付け、相互変換を可能とするための規則を整備する必要がある。

そこで、本事業では、論理形式から特定の物理形式への変換操作や、その逆に特定の物理形式から論理形式への変換操作についての考え方を整備することで、論理形式と物理形式の相互変換を実現可能とする方針を策定した。

3.3.9 XML Schema への依存の軽減について

用語のシリアライズ・デシリアライズを効率的で実用可能なものとするため、論理形式は特定のデータ形式に依存しないことが望ましい。

そこで、本事業では、IMI コア語彙 v1 や NIEM で採用されているような、XML Schema に特化した substitution Group (代替グループ) の機構の利用を回避する方針を採用した。従来 IMI コア語彙 v1 で substitution Group の機構を用いて記述されていた用語については、代替グループを使わずに、代替可能な要素をフラットに列挙することで表現しなおすこととした。

例えば、IMI コア語彙 v1 では「連絡先」の「連絡方法」が抽象要素（具体的な型をもたない用語）として定義されており、この「連絡方法」を substitutionGroup の機構を用いて、具体的な用語である「名称」「メールアドレス」「住所」「電話番号」「内線番号」などで置換して利用することが想定されていた。これに対して、本事業では、抽象要素である「連絡方法」を廃止し、具体的な用語である「名称」「メールアドレス」などを「連絡先」の下に直接プロパティとして与えることで、substitutionGroup の利用を回避した。

3.3.10 NIEM への依存の軽減について

NIEM への依存を弱め、NIEM に特化した構造の見直しや、NIEM 独自の機構や NIEM の提供する各種コードリストの参照などを最小限にとどめる方針を採用した。これにより、NIEM で想定していないデータ提供（オープンデータ）や RDF 形式への親和性を確保しやすくなるとともに、NIEM 以外の語彙（ISA Core Vocabulary、UN/CEFACT Core Component Library、schema.org など）との相互運用性を図ることも容易になると考えられる。

3.3.11 コードリストおよびコードの参照と格納について

共通語彙基盤の外部で規定されたコードリストおよびコードを広く参照するための方策として、IMI コア語彙の中に、外部コードを参照するための「コード」と、その外部コードが属するコードリストを記述するための「コードリスト」の2つの用語を新たに追加する方針とした。

また、共通語彙基盤の内部で独自に整備されたコードリストやコードを表現するための、適切なデータ構造について、検討・整備することとした。

4. コア語彙の整備

共通語彙基盤の導入目的である「表記、意味、構造の明確化」を実現するためには、複数の業務領域（ドメイン）で共通に用いられる概念の表記、意味、構造を、異なるドメイン間・異なる利用者間で明確に相互理解するとともに、効率的な機械処理を可能とするための共通的な用語が必要となる。本事業では、そのような特性を備えた用語の集合である「コア語彙」について、先行する調査研究の成果をベースに精査・拡充を行うことで整備した。

コア語彙は、共通語彙基盤の語彙データベースが格納・提供する語彙の中で、複数の業務領域(ドメイン)で共通に用いられる用語の集合である。

本概念モデル書では、共通語彙基盤において利用されるコア語彙は「IMI コア語彙」と表記する。

本事業では、前章で示した語彙の設計方針に則り、先行事業の成果である IMI コア語彙 v1 をベースに、用語の精査・見直しおよびデータ構造の拡充を実施することで、語彙データベースによる格納・提供に十分な実用性を有するとともに、各種ドメイン語彙の将来的な整備のベースとしての十分な発展性を有するものとして、IMI コア語彙 v2 案の整備を実施した。

4.1 意味的検討：コア語彙の内容の精査・見直し

前章で示した設計方針に則り、本事業では、IMI コア語彙 v1 の各用語の表記や意味、用語間の関係について精査および見直しを実施した。

4.1.1 用語の表記の見直し

用語の表記に関する見直し内容について、以下に示す。

表 4.1.1.1 用語の表記の見直し

項番	見直しの観点	見直し内容	例	
			見直し前(v1)	見直し後(v2 案)
1	「型」の排除	用語の表記の末尾に付与された「型」を削除した。	「人型」	「人」
2	冗長性の排除	用語の表記の内、用語とその用語が包含する用語で同じ表記が繰り返し利用されているなど、冗長な表記となっている箇所を修正・削除した。	「住所型」の「住所」	「住所」の「表記」
3			「電話番号型」の「電話番号」	「電話番号」の「表記」
4	表現形態の明示	用語の末尾に表現形態を表すために付与された識別子（コード、カナ、ローマ字、など）については、用語名と識別子とを明確に区別できるよう、識別子の部分を括弧で括ることとした。	「人型」の「性別コード」	「人」の「性別(コード)」
5			「氏名型」の「カナ姓名」	「氏名」の「姓名(カナ)」
6			「物型」の「販売業者 ID」	「製品」の「販売業者(ID)」
7	分かりやすさの向上	用語の表記の内、専門的・技術的な傾向が強いものに関しては、利用者にとって理解のしやすい、一般的な用語への置き換えを行った。	「構造化住所型」	「定型住所」
8			「証明型」	「ID」
9			「連絡先型」の「URL」	「連絡先」の「Web サイト」
10	正確性の向上	用語の表記が、その用語の意味を正確に表すように修正した。	「組織」の「状況」	「組織」の「活動状況」
11			「電話番号」の「国際電話番号」	「電話番号」の「国番号」
12			「物型」	「製品」
13	統一性の向上	「名称」や「種別」など、基本的な属性を示す用語については、語彙全体で同じ表記を用いるよう修正した。	「組織型」の「組織名」	「組織」の「名称」
14			「状況型」の「状況名」	「状況」の「名称」

4.1.2 用語の意味の見直し

用語の意味に関する見直し内容について、以下に示す。

表 4.1.2.1 用語の意味の見直し

項番	見直し対象の用語	見直し内容	見直し前(v1)	見直し後(v2 案)
1	「氏名」の「姓名(ローマ字)」	用語の説明が、その用語の意味を正確に表すように修正した。	氏名(姓、名)のローマ字表記	氏名(姓、名)のアルファベット表記
2	「定型住所」の「町名」	用語の説明が、その用語の意味を正確に表すように修正した。	町域	町域、または市区町村の下の行政区分
3	「定型住所」の「方書」	用語の説明が、その用語の意味を正確に表すように修正した。	方書	方書(私書箱も含む)
4	「定型住所」の「部屋番号」	用語の説明が、その用語の意味を正確に表すように修正した。	号棟、階数、号室	階数、号室
5	「組織」の「名称」	用語の説明が、その用語の意味を正確に表すように修正した。	組織名称	組織名称。組合以外は、組織分類を省略して記載
6	「場所」の「地理識別子」	用語の説明が、その用語の意味を正確に表すように修正した。	場所を証明するもの	場所の識別子
7	「製品」の「自動認識ID」	用語の説明が、その用語の意味を正確に表すように修正した。	バーコードID	バーコードや二次元コード等の自動認識IDの画像等
8	「製品」の「価格」	用語の説明が、その用語の意味を正確に表すように修正した。	物の価値	物の価格
9	「施設」の「駐車場」	用語の説明が、その用語の意味を正確に表すように修正した。	駐車可能台数	施設が利用する駐車場
10	「容量」の「単位」	用語の説明が、その用語の意味を正確に表すように修正した。	1 : ml、2 : l	容量の単位
11	「容量」の「面積」	用語の説明が、その用語の意味を正確に表すように修正した。	1 : m ² 、2 : km ² 、3 : a、4 : ha、5 : acre、6 : b、7 : 坪、8 : 歩	面積の単位
12	「イベント」の「名称」	用語の表記変更に伴い、意味の修正も実施した。	活動の名称	イベントの名称
13	「イベント」の「関連イベント」	用語の表記変更に伴い、意味の修正も実施した。	上位のアクティビティ	関連するイベント
14	「座標」の「座標参照系」	用語の表記変更に伴い、意味の修正も実施した。	測地系コード	座標の参照系

4.1.3 用語の構造化の見直し

用語の構造化に関する見直し内容について、以下に示す。

表 4.1.3.1 用語の構造化の見直し

項番	見直しの観点	見直し内容
1	「人」の構造化の見直し	「人」に関わる情報として利用頻度が高いと考えられる「連絡先」「ID」を追加した。また、「国籍(コード)」と同様に、「出生国」についてもコード表現を可能とするための用語を追加した。
2	「国籍」の構造化の見直し	「国籍」には ISO3166 の国コードを表現するために、「ISO3166Alpha2」、「ISO3166Alpha3」、「ISO3166Numeric」の3つの用語が含まれていたが、これらは元々存在した「国籍(コード)」へ集約可能と判断し、削除した。
3	「名称」の追加	「連絡先」「組織」「製品」「地点」「設備」などの名称について、日本においては漢字表記、カタカナ表記、ローマ字表記を選択でき、データ作成者やデータ利用者によって使い分け可能であることが望ましい。そこで、名称の漢字表記、カナ表記、アルファベット表記をまとめて表現し、利用者が使い分け可能とするための「名称」というクラス用語を追加した。
4	「氏名」の構造化の見直し	「氏名」に「種別」を設けることで、「氏名」によって現される氏名がどのような種類の氏名であるか(旧姓、通称名、など)を明記できるようにした。また、これに伴い、v1 にあった「旧姓」関係の用語を廃止した。
5	住所に関する用語の再整理	「構造化住所型」に含まれていた「郵便番号」「住所ID」「住所コード」を、「構造化住所」ではなく「住所」に含めるよう修正した。また、「住所」の構造を簡易化するため、「方書型」を廃止し、「方書型」に含まれていた用語は「住所」に移した。
6	座標に関する用語の見直し	「経緯度座標系型」を見直し、「座標参照系」および「測地高度」を設けた「座標」を整備した。また、上記「座標」で統一的に表すことが可能なため、「UTM 座標系型」「MGRS 座標系型」を削除した。さらに、「緯度型」、「緯度値型」、「経度型」、「経度値型」を「度分秒」へ統合した。
7	「製品」の構造化の見直し	「物」と「有体物」に分かれていた概念を「製品」に統合した。また、個々の製品(個品)を表すための用語として、「製品個品」を追加した。
8	「地点」の追加	「施設」と「建物」を共通的に扱えるよう、両者が共通して継承する上位クラスとして「地点」を追加した。
9	「建物」の構造化の見直し	「建物型」と「建物構造型」に分かれていた概念を「建物」に統合した。
10	「施設」の構造化の見直し	施設が利用する建物を明示するために、「建物」を追加した。例えば、大学という「施設」が、棟という「建物」の群で構成されているような状況を記述するために用いる。

項番	見直しの観点	見直し内容
11	「駐車場」の追加	施設に付随する、ニーズの高い設備情報として、駐車場に関する情報を記述可能なクラス用語「駐車場」を追加した。
12	「設備」の追加	AED など、施設や建物に付随する設備を表現するための用語を追加した。
13	「イベント」の追加	「活動型」の内容を見直した上で、イベント情報を十分に記述可能とするようなクラス用語として「イベント」を新たに整備した。
14	「数量」「容量」「面積」「重量」「長さ」の構造化の見直し	「数量型」、「容量型」、「面積型」、「重量型」、「長さ型」について、それぞれ「測定単位型」を継承して定義されていた構造を単純化した。
15	「時間」の追加	時間の長さ(インターバル)を単位付きで表すために、「時間」を追加した。
16	価格に関する概念の見直し	「物品価値型」を見直し、より利用者にとってイメージのしやすい「価格」、「価格一覧」として再整理した。
17	「実体」の構造化の見直し	「人型」と「組織型」を共通的に扱うための「実態型」について、XML Schema の substitutionGroup を用いて表現されていたものを、より一般的な継承(extension)の機構を用いて表現し直した。
18	「状況」の構造化の見直し	「状況型」に含まれていた「証明」および「証明名」を、「関係者」に集約した。
19	「日時」の構造化の見直し	「日時型」では、各種の日時や日付の表現形式の違いを XML Schema の substitutionGroup を用いて表現されていたが、これを単純化し、階層をもたない「表記」と(定型的な)階層を備えた「表記(定型)」および「表記(定型和暦)」を用いて表現し直した。それに伴い、クラス用語「定型日時」と「定型和暦日時」を追加した。
20	「スケジュール」の再整理	スケジュールに関する記述力を向上させるため、周期的なスケジュールを記述する「定期スケジュール」と、周期的でない(単発的な)スケジュールを記述するための「イベントスケジュール」を追加した。「スケジュール」はこれらを共通的に取り扱うための上位クラスとした。
21	用途の不明確な用語の廃止	「固有名型」、「関連型」、「系列型」については、利用ニーズが不明確であるため、廃止することとした。
22	共通語彙基盤の外部で規定されるコードリスト・コードの表現手段の追加	国内外で定められている、既存の各種コードやコードリストを表現するためのクラス用語として、「コード」および「コードリスト」を追加した。
23	ID・ID 体系の表現手段の追加	国内外で定められている各種の ID や ID 体系を表現するためのクラス用語として、「ID」および「ID 体系」を追加した。それに伴い、「証明型」を廃止した。
24	「アクセス」の追加	施設や建物へのアクセス情報(最寄り駅から施設や建物までの経路など)を表現するためのクラス用語「アクセス」を追加した。

項番	見直しの観点	見直し内容
25	「事物」の追加	コア語彙の全てのクラス用語のベースとなるクラス用語として、「事物」を追加した。
26	「連絡先」の構造化の見直し	施設の連絡先などの記述に用いる「連絡先」用語に関して、将来的な外国人による利用を想定し、「連絡先」に「対応言語」を追加した。
27	組織に関する概念の再整理	組織に関する情報をより詳細に記述可能とするため、新たなクラス用語として「業務組織」、「法人」、「従業員数」、「構成員」を追加した。
28	基本的なデータタイプに関する独自用語の排除	既存サービスとの互換性や、利用者（開発者）にとっての理解・利用のしやすさを考慮し、XML Schema で用意されているデータタイプに置換可能なものはそれらで置き換えた。（「テキスト型」の xsd:string による置換など）
29	NIEM 独自のデータタイプの排除と、XML Schema のデータタイプの積極的な利用	「テキスト型」のような基本的な用語については、既存サービスとの互換性や、利用者にとっての理解や利用の容易性を増すため、W3C XML Schema で定義された標準的なデータタイプである xsd:string 等で置き換えた。 同様に、niem-xsd:anyURI、niem-xsd:string など、NIEM 独自のデータタイプについても、より一般的な xsd:anyURI や xsd:string で置き換えた。
30	コードリストの整理	NIEM が提供しているコードリスト（「iso_4217:CurrencyCodeSimpleType」や「unece:LengthCodeType」など）について、UN/CEFACT が公開しているコードリストで置換した。また、コード定義が存在していなかった独自コードリスト（「codes:容量単位コード型」「codes:面積単位コード型」など）については、UN/CEFACT が公開しているコードリストで相当するものがあればそれで代替し、そうでないもの（「ic:人_性別コード」など）については、新たに整備した「コード」用語で代替することとした。

4.2 構造的検討：コア語彙のデータ構造の拡充

前節における各種検討の結果に基づいて、本事業では、コア語彙のデータ構造案を策定した。データ構造の構成要素、表現形式、データ構造に関するシリアライズとデシリアライズについて、以下に示す。

4.2.1 データ構造の構成要素の分類・整理

コア語彙に関連するデータの内、中核的な構成要素となる「語彙」および「用語」について、以下のように分類・整理を実施した。

表 4.2.1.1 語彙の構成要素

項番	構成要素	概要
1	語彙	用語の集合。
2	用語	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もの。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）をもつ場合がある。用語の特徴に応じて、クラス用語またはプロパティ用語に分類される。
3	クラス用語	用語の内、他の用語(プロパティ用語)を包含するような構造をもった用語。UN/CEFACT CCTS や OASIS UBL における ABIE (Aggregate Business Information Entity) に相当するもの。他のクラス用語を継承することで、継承元のクラス用語が包含するプロパティ用語を包含することが可能。
4	プロパティ用語	用語の内、クラス用語と他のクラス用語またはデータタイプ(文字列や整数などの基本的なデータの種別)とを関連付けるもの。UN/CEFACT CCTS や OASIS UBL における BBIE (Basic Business Information Entity) または ASBIE (Association Business Information Entity) に相当するもの。
5	外部用語	共通語彙基盤の整備・管理する用語と、完全一致、類似、広義、狭義、関連、などの関係にある、外部の（共通語彙基盤以外で整備・管理されている）語彙に含まれる用語。
6	コードリスト	プロパティ用語の取り得る値の表記を限定し、その意味を規定するために、文字列や数字列などの値を列挙したもの。 IMI コア語彙 v2 案に用語として追加した「コードリスト」「コード」は、国内外の既存のコード体系を参照するために導入した用語であったが、本項で示す「コードリスト」「コード」は、共通語彙基盤の内部で独自に整備したコードリストやコードを格納するために、新たに導入したデータ構造を指す。
7	コード	「コードリスト」に含まれる個々の値。

個々の構成要素の関係を UML クラス図により表現した概要図を以下に示す。

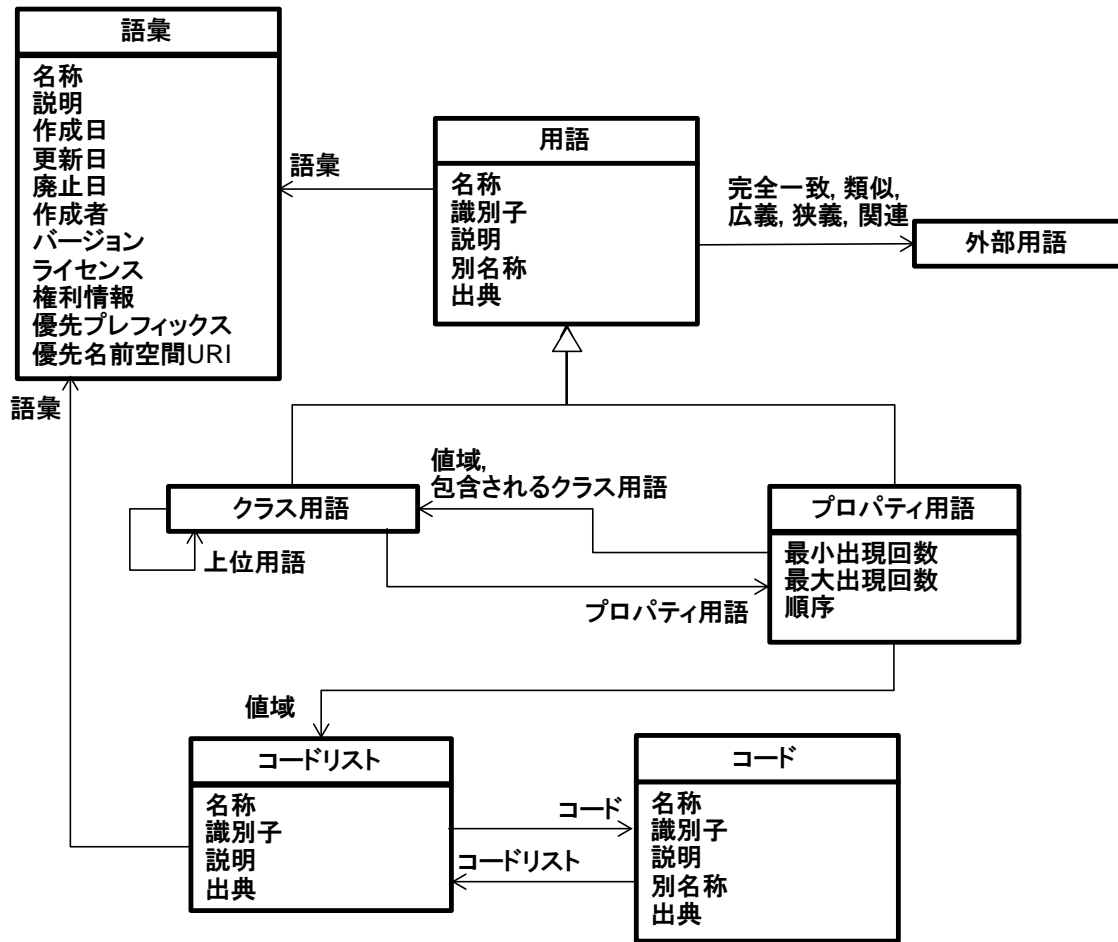


図 4.2.1.1 語彙の構成要素 (UML クラス図)

(1) 語彙の構造

「語彙」は、語彙データベースに格納される用語の集合である。

本事業では、前節までの検討内容に基づいて、「語彙」を特徴付ける属性を集約した以下のようなデータ構造を策定した。

表 4.2.1.2 「語彙」のデータ構造

項番	項目	概要	取り得る値	出現回数 ¹	利用例
1	名称	語彙の名称を自然言語で表記したもの。日本語による名称は必須とする。	文字列	1..n	共通語彙基盤コア語彙
2	説明	語彙の意味を自然言語で説明したもの。日本語による説明は必須とする。	文字列	1..n	コア語彙は、共通語彙基盤の基礎をなすもので、氏名、住所、組織等、あらゆる社会活動で使用される中核的な用語の集合です。多くのシステム間で情報交換のための基礎となる語彙で、データ交換、オープンデータの二次利用等の効率化に役立つものです。
3	作成日	語彙が新規に作成された日時。	日付	0..1	2014-09-09
4	更新日	語彙が更新された日時。	日付	0..n	2014-09-10
5	廃止日	語彙が廃止された日時。	日付	0..1	2014-09-11
6	作成者	語彙を作成した主体。	人または組織	0..n	
7	バージョン	語彙のバージョン。	文字列	0..1	2

¹ 「0..1」：0回又は1回のみ出現、「0..n」：0回以上出現、「1..n」：1回以上出現

項番	項目	概要	取り得る値	出現回数 ¹	利用例
8	ライセンス	語彙の利用に関するライセンス情報。	ライセンス文書へのリンク(URI)	0..n	< http://creativecommons.org/publicdomain/zero/1.0/ >
9	権利情報	語彙の権利に関する情報。一般的に、知的財産権を含む、語彙に関連する様々な所有権に関する文書など。	権利情報へのリンクまたは文字列	0..n	
10	優先プレフィックス	語彙の名前空間のプレフィックスとして優先的に使用されることが望ましいもの。	URI	0..n	ic
11	優先名前空間 URI	語彙の名前空間の URI として優先的に使用されることが望ましいもの。	URI	0..n	< http://imi.ipa.go.jp/ns/core/2# >

(2) 用語の構造

「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ものである。その意味、あるいは概念の一部として、他の用語との関係（構造）をもつ場合がある。用語の特徴に応じて、クラス用語またはプロパティ用語に分類される。

本事業では、前節までの検討内容に基づいて、「用語」を特徴付ける属性を集約した以下のようなデータ構造を策定した。これらは後述する「クラス用語」「プロパティ用語」に共通する属性となることを想定している。

表 4.2.1.3 「用語」のデータ構造

項番	項目	概要	取り得る値	出現回数	利用例
1	名称	用語の名称を自然言語で表記したものの。 日本語による名称は必須とする。	文字列	1..n ¹	人
2	識別子	用語を一意に識別するための識別子。IRI (International Resource Identifier) またはその短縮表現である QName を用いて記述する。	URI	1..1	ic:人
3	語彙	用語の含まれる語彙。	「語彙」へのリンク	1..1	<http://imi.ipa.go.jp/ns/core/2>
4	説明	用語の意味を自然言語で説明したものの。日本語による説明は必須とする。	文字列	1..n	人の情報を表現するためのデータ型
5	別名称	用語の別名。同義語や類義語、表記ゆれを含む。	文字列	0..n	人間
6	出典	用語の定義や用例、法的根拠等が示された文献やリソース。	文字列	0..n	
7	完全一致	用語と完全に一致する関係にある用語。	URI	0..n	
8	類似	用語と類似の関係にある用語。	URI	0..n	
9	広義	用語よりも広義の概念を示す用語。	URI	0..n	
10	狭義	用語よりも広義の概念を示す用語。	URI	0..n	
11	関連	用語と関連する用語。	URI	0..n	

¹ 多言語で記述された名称も格納可能とするため、ここでは出現回数を「1..1」ではなく「1..n」として定義した。

(3) クラス用語の構造

「クラス用語」は、用語の内、他の用語(プロパティ用語)を包含するような構造をもった用語である。UN/CEFACT CCTS や OASIS UBL における ABIE (Aggregate Business Information Entity) に相当するものである。

「クラス用語」は、他のクラス用語を継承することで、継承元のクラス用語が包含するプロパティ用語を包含することが可能である。

本事業では、前節までの検討内容に基づいて、「クラス用語」を特徴付ける属性を集約した以下のようなデータ構造を策定した。「クラス用語」は「用語」の構造を継承し、「名称」や「識別子」をはじめ、「用語」に含まれる属性はすべて「クラス用語」にも含まれるものとする。

表 4.2.1.4 「クラス用語」のデータ構造

項番	項目	概要	取り得る値	出現回数	利用例
1	上位用語	他のクラス用語を継承している場合は、継承元のクラス用語(上位クラス用語)を示す。 RDF スキーマ語彙における rdfs:subClassOf に相当する。	「クラス用語」へのリンク	0..1 ¹	
2	プロパティ用語	クラス用語が包含するプロパティ用語。	「プロパティ用語」へのリンク	0..n	ic:実体

「クラス用語」の「識別子」は、「クラス用語」の「名称」に、「クラス用語」が属する「語彙」の「優先名前空間 URI」を前置した URI か、「クラス用語」が属する「語彙」の「優先プレフィックス」を前置した QName のみを値として取るよう規定する。なお前置の際には、URI の場合はハッシュ記号 “#”、QName の場合はコロン記号 “:” がセパレータとして挿入されるものとする。当該規則は BNF 表記で以下のように記述できる。

```
<「クラス用語」の「識別子」> ::= <「クラス用語」の「語彙」の「優先名前空間 URI」> “#”
                                     <「クラス用語」の「名称」> |
                                     <「クラス用語」の「語彙」の「優先プレフィックス」> “:”
                                     <「クラス用語」の「名称」>
```

上記の規則により、例えば、優先名前空間<http://imi.ipa.go.jp/ns/core/2>および優先プレフィックス「ic」をもつ語彙に属するクラス用語「人」の「識別子」は、<http://imi.ipa.go.jp/ns/core/2#人>または「ic:人」と定められる。

¹ 本概念モデル書では、XML 形式へのシリアライズルールを複雑化することを避けるため、クラス用語の多重継承はサポートしない方針を採用した。(XML 形式で採用している W3C XML Schema が多重継承の機構をサポートしていないため)

(4) プロパティ用語の構造

「プロパティ用語」は、用語の内、クラス用語と他のクラス用語またはデータタイプ(文字列、整数などの基本的なデータの型)とを関連付けるものである。UN/CEFACT CCTS や OASIS UBL における BBIE (Basic Business Information Entity) または ASBIE (Association Business Information Entity) に相当するものである。

本事業では、前節までの検討内容に基づいて、「プロパティ用語」を特徴付ける属性を集約した以下のようなデータ構造を策定した。「プロパティ用語」は「用語」の構造を継承し、「名称」や「識別子」をはじめ、「用語」に含まれる属性はすべて「プロパティ用語」にも含まれるものとする。

表 4.2.1.5 「プロパティ用語」のデータ構造

項番	項目	概要	取り得る値	出現回数	利用例
1	包含されるクラス用語	プロパティ用語が包含されるクラス用語。 RDF スキーマ語彙における rdfs:domain に相当する。	「クラス用語」へのリンク	0..1	ic:人
2	値域	プロパティ用語の値がどのクラス用語またはデータタイプのインスタンスであるかを示したものの。 RDF スキーマ語彙における rdfs:range に相当する。	「クラス用語」または「コードリスト」またはデータタイプへのリンク	1..1	ic:氏名
3	最小出現回数	包含されるクラス用語の中での最小出現回数。省略した場合は、最小出現回数に制限を課さない、すなわち当該プロパティ用語が包含されるクラス用語の中で一度も出現しなくとも良いことを示す。	非負整数	0..1	1
4	最大出現回数	包含されるクラス用語の中での最大出現回数。省略した場合は、最大出現回数に制限を課さない、すなわち当該プロパティ用語が包含されるクラス用語の中で何度でも出現して良いことを示す。	非負整数	0..1	1
5	順序	包含されるクラス用語の中での順序。	整数	1..1	1

NIEM や IMI コア語彙 v1 で採用されていた規則を踏襲し、「プロパティ用語」の「識別子」は、当該「プロパティ用語」が包含される「クラス用語」の識別子と、当該「プロパティ用語」の「名称」とをアンダースコア記号（“_”）で接続した値を用いるものとする。具体的には、「プロパティ用語」の「識別子」は、「プロパティ用語」の「名称」に、「プロパティ用語」の「包含されるクラス用語」の「識別子」を前置した URI または QName のみを値として取るよう規定する。なお前置の際には、アンダースコア記号（“_”）がセパレータとして挿入されるものとする。当該規則は BNF 表記で以下のように記述できる。

$\langle \text{「プロパティ用語」の「識別子」} \rangle ::= \langle \text{「プロパティ用語」の「包含されるクラス用語」の「識別子」} \rangle _ \langle \text{「プロパティ用語」の「名称」} \rangle$

上記の規則により、例えば、識別子として $\langle \text{http://imi.ipa.go.jp/ns/core/2\#人} \rangle$ または「ic:人」をもつクラス用語「人」に包含されるプロパティ用語「氏名」の「識別子」は、 $\langle \text{http://imi.ipa.go.jp/ns/core/2\#人_氏名} \rangle$ または「ic:人_氏名」と定められる。

(5) コードリストの構造

「コードリスト」は、プロパティ用語の取り得る値の表記を限定し、その意味を規定するために、文字列や数字列などの値を列挙したものである。統制語彙ともいう。

IMI コア語彙 v2 案に用語として追加した「コードリスト」は、国内外の既存のコード体系を参照するために導入した用語であったが、本節で示す「コードリスト」は、共通語彙基盤の内部で独自に整備したコードリストを表現するために、新たに導入したデータ構造を指す。

本事業では、前節までの検討内容に基づいて、「コードリスト」を特徴付ける属性を集約した以下のようなデータ構造を策定した。

表 4.2.1.6 「コードリスト」のデータ構造

項番	項目	概要	取り得る値	出現回数	利用例
1	名称	コードリストの名称を自然言語で表記したもの。	文字列	1..n ¹	
2	識別子	コードリストを一意に識別するための識別子。IRI (International Resource Identifier) またはその短縮表現である QName を用いて記述する。	URI	1..1	
3	語彙	コードリストの含まれる語彙。	「語彙」へのリンク	1..1	
4	説明	コードリストの意味を自然言語で説明したもの。日本語による説明は必須とする。	文字列	1..n	
5	出典	コードリストの定義や用例、法的根拠等が示された文献やリソース。	文字列	0..n	
6	コード	コードリストに含まれるコード。	「コード」へのリンク	1..n	

¹ 他言語で記述された名称も格納可能とするため、ここでは出現回数を「1..1」ではなく「1..n」として定義した。

(6) コードの構造

「コード」は、「コードリスト」に含まれる個々の値である。統制語ともいう。

IMI コア語彙 v2 案に用語として追加した「コード」は、国内外の既存のコードを参照するために導入した用語であったが、本節で示す「コード」は、共通語彙基盤の内部で独自に整備したコードを格納するために、新たに導入したデータ構造を指す。

本事業では、前節までの検討内容に基づいて、「コード」を特徴付ける属性を集約した以下のようなデータ構造を策定した。

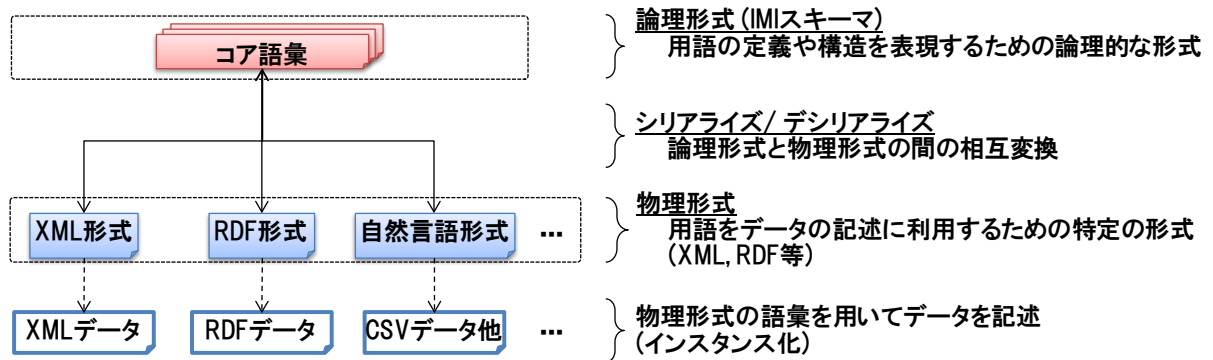
表 4.2.1.7 「コード」のデータ構造

項番	項目	概要	取り得る値	出現回数
1	名称	コードを自然言語で表記したもの。	文字列	1..n ¹
2	識別子	コードを一意に識別するための識別子。IRI (International Resource Identifier) またはその短縮表現である QName を用いて記述する。	URI	1..1
3	コードリスト	コードの含まれるコードリスト。	「コードリスト」へのリンク	1..1
4	説明	コードの意味を自然言語で説明したもの。日本語による説明は必須とする。	文字列	1..n
5	別名称	コードの別名。同義語や類義語、表記ゆれを含む。	文字列	0..n
6	出典	コードの定義や用例、法的根拠等が示された文献やリソース。	文字列	0..n
7	完全一致	コードと完全に一致する関係にある他のコード。	URI	0..n
8	類似	コードと類似の関係にある他のコード。	URI	0..n
9	広義	コードよりも広義の概念を示す他のコード。	URI	0..n
10	狭義	コードよりも広義の概念を示す他のコード。	URI	0..n
11	関連	コードと関連する他のコード。	URI	0..n

¹ 他言語で記述された名称も格納可能とするため、ここでは出現回数を「1..1」ではなく「1..n」として定義した。

4.2.2 語彙データの表現形式の整理

語彙データの表現形式として、特定のデータ形式に依存しない論理形式と、利用シーンに特化した物理形式との二層に分割し、両者の間を「シリアライズ」および「デシリアライズ」という処理により、相互変換可能とする構成を採用した。語彙データの表現形式に関する概要図を以下に示す。



論理形式は、用語の表記、意味、構造、関係性を表現するための、特定のデータ形式によらない論理的な形式である。語彙データの整備やメンテナンスは、基本的に論理形式に基づいて実施することで、語彙データの汎用性の向上を図る。

物理形式は、XML や RDF といった特定のデータ形式によるデータの記述や妥当性検証に適した、物理的なスキーマである。データ交換・データ公開の場面では、その際に採用されるデータ形式に応じて、当該データの記述に最適な物理形式を選択し、利用することを想定している。

シリアライズは、語彙の論理形式から物理形式への変換処理である。また、その逆である、語彙の物理形式から論理形式への変換処理はデシリアライズと呼ぶ。

4.2.3 語彙データの論理形式

語彙データの論理形式に関しては、ドメインエキスパート(業務専門家)による理解・利用に適したテーブル(スプレッドシート)を用いた表現を採用した。前節で示した「語彙」「用語」などの構成要素を表現するため、以下に示す複数テーブルを整備した。

(1) 語彙テーブル

「語彙」のテーブルによる表現として、表 4.2.1.2 に示した「語彙」の各属性を列にもつ表を整備した。

語彙テーブルの例として、IMI コア語彙 v2 案に関する情報を格納したものを以下に示す。

詳細は、別添「コア語彙 v2 案」を参照のこと。

語彙の優先プレフィックス	語彙の優先名前空間URI	語彙の名称	語彙の説明	語彙の作成者の名称	語彙の作成者のURL	語彙のライセンス	語彙のバージョン	語彙の作成日
ic	http://imi.ipa.go.jp/ns/core/210#	共通語彙基盤コア語彙	コア語彙は、共通語彙基盤の基礎をなすもので、氏名、住所、組織等、あらゆる社会活動で使用される中核的な用語の集合です。多くのシステム間で情報交換のための基礎となる語彙で、データ交換、オープンデータの二次利用等の効率化に役立つものです。	経済産業省	http://www.meti.go.jp/	http://creativecommons.org/publicdomain/zero/1.0/	2.10	2014-09-11
				IPA 独立行政法人情報処理推進機構	http://www.ipa.go.jp/			

図 4.2.3.1 語彙データの論理形式：語彙テーブル

(2) 用語テーブル

クラス用語およびプロパティ用語のテーブルによる表現として、表 4.2.1.3、表 4.2.1.4、表 4.2.1.5 に示した「用語」「クラス用語」「プロパティ用語」の各属性を列にもつ表を整備した。クラス用語とそれが含むプロパティ用語の関係を把握しやすくするため、用語テーブルではクラス用語とプロパティ用語の両者を一つの表で管理する。

用語テーブルの例として、IMI コア語彙 v2 案に含まれるクラス用語「人」と、「人」が含むプロパティ用語「氏名」「性別」「性別(コード)」「生年月日」の情報を格納したものを以下に示す。

詳細は、別添「コア語彙 v2 案」を参照のこと。

用語の名称	用語の識別子	プロパティ用語の値域	クラス用語の上位用語	プロパティ用語の出現回数	用語の説明	用語の出典	用語の利用例
人	ic:人		extends ic:実体型		人の情報を表現するためのデータ型		
氏名	ic:人.氏名	ic:氏名型		0..n	氏名		-
性別	ic:人.性別	xsd:string		0..1	性別の名称		男
性別(コード)	ic:人.性別-コード	ic:コード型		0..1	性別コード		1
生年月日	ic:人.生年月日	ic:日時型		0..1	生年月日		-

図 4.2.3.2 語彙データの論理形式：用語テーブル

(3) 関係性テーブル

語彙データベースの格納する用語と、各種の外部用語(NIEM, ISA Core Vocabulary, UN/CEFACT Core Component Library 等)との関係性に関するテーブル表現として、表 4.2.1.3 に示した「用語」の「完全一致」「類似」「広義」「狭義」「関連」を表現するための表を整備した。

また、「用語」の「別名称」についても関係性の一つとして、本テーブルにて管理・表現することとした。

関係性テーブルの例として、IMI コア語彙 v2 案に含まれるクラス用語「人」と、「人」が含むプロパティ用語「氏名」「性別」「性別(コード)」「生年月日」に関する関係性情報を格納したものを以下に示す。

詳細は、別添「コア語彙 v2 案」を参照のこと。

マップ元		マップ先		
識別子	→ 関係 →	用語識別子 または 値	語彙体系	備考
ic:人	関連	Person. Details	UN/CEFACT CCL	
ic:人 氏名	関連	nc:PersonName	NIEM Core 2.1	
ic:人 氏名	別名	名前		
ic:人 性別	関連	nc:PersonSexText	NIEM Core 2.1	
ic:人 性別	別名	性		
ic:人 性別-コード	関連	nc:PersonSexCode	NIEM Core 2.1	
ic:人 性別-コード	関連	Person. Gender. Code	UN/CEFACT CCL	Code Type
ic:人 生年月日	関連	nc:PersonBirthDate	NIEM Core 2.1	
ic:人 生年月日	関連	schema:birthDate	schema.org	
ic:人 生年月日	関連	Person. Birth. Date Time	UN/CEFACT CCL	Date Time Type

図 4.2.3.3 語彙データの論理形式：関係性テーブル

4.2.4 語彙データの物理形式

「3.3 共通語彙基盤における語彙の設計方針」の検討内容に基づいて、本事業では、主に情報交換に適した「XML 形式」、主に情報提供に適した「RDF 形式」(Turtle, RDF/XML, JSON-LD による表現をサポート)、主にテーブルの列名や人間のための仕様説明に適した「自然言語形式」の三種の物理形式を整備した。

(1) XML 形式

シリアライズ後の XML Schema の概要を示すため、以下に IMI コア語彙 v2 案の XML 形式を一部抜粋したものを以下に示す。

XML ヘッダ

```
<?xml version="1.0" encoding="utf-8"?>
```

スキーマ定義：名前空間のプレフィックスと URI の対応等を記載

```
<xsd:schema
  targetNamespace="http://imi.ipa.go.jp/ns/core/2#"
  version="2"
  xmlns:cc="http://creativecommons.org/ns#"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:ic="http://imi.ipa.go.jp/ns/core/2#"
  xmlns:uncefactCode="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCode:9"
  xmlns:uncefactDuration="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeDuration:4"
  xmlns:uncefactISO4217="urn:un:unece:uncefact:codelist:standard:ISO:ISO3AlphaCurrencyCode:2012-08-31"
  xmlns:uncefactLinear="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeLinear:4"
  xmlns:uncefactVolume="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeVolume:4"
  xmlns:uncefactWeight="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeWeight:4"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema">
```


語彙の定義：語彙の名称や説明等を記載

```
<xsd:annotation>
  <xsd:documentation xml:lang="en">IMI Core Vocabulary</xsd:documentation>
  <xsd:documentation xml:lang="ja">コア語彙は、共通語彙基盤の基礎をなすもので、氏名、住所、組織等、あらゆる社会活動で使用される中核的な用語の集合です。多くのシステム間で情報交換のための基礎となる語彙で、データ交換、オープンデータの二次利用等の効率化に役立つものです。</xsd:documentation>
  <xsd:appinfo>
    <dct:title xml:lang="en">IMI Core Vocabulary</dct:title>
    <dct:title xml:lang="ja">共通語彙基盤コア語彙</dct:title>
    <dct:license>http://creativecommons.org/publicdomain/zero/1.0/</dct:license>
    <cc:license>http://creativecommons.org/publicdomain/zero/1.0/</cc:license>
    <dct:issued>2014-09-11</dct:issued>
    <dct:creator>
      <foaf:name xml:lang="en">Ministry of Economy, Trade and Industry</foaf:name>
      <foaf:name xml:lang="ja">経済産業省</foaf:name>
      <foaf:homepage>http://www.meti.go.jp/</foaf:homepage>
    </dct:creator>
    <dct:creator>
      <foaf:name xml:lang="en">Information-technology Promotion Agency, Japan.</foaf:name>
      <foaf:name xml:lang="ja">IPA 独立行政法人情報処理推進機構</foaf:name>
      <foaf:homepage>http://www.ipa.go.jp/</foaf:homepage>
    </dct:creator>
  </xsd:appinfo>
</xsd:annotation>
```

名前空間のインポート：外部ファイルの参照等を記載

```
<xsd:import namespace="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeDuration:4"
schemaLocation="./UNECE_MeasurementUnitCommonCodeDuration_4.xsd"/>
<xsd:import namespace="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeLinear:4"
schemaLocation="./UNECE_MeasurementUnitCommonCodeLinear_4.xsd"/>
<xsd:import namespace="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCode:9"
schemaLocation="./UNECE_MeasurementUnitCommonCode_9.xsd"/>
<xsd:import namespace="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeWeight:4"
schemaLocation="./UNECE_MeasurementUnitCommonCodeWeight_4.xsd"/>
<xsd:import namespace="urn:un:unece:uncefact:codelist:standard:ISO:ISO3AlphaCurrencyCode:2012-08-31"
schemaLocation="./ISO_ISO3AlphaCurrencyCode_2012-08-31.xsd"/>
<xsd:import namespace="urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeVolume:4"
schemaLocation="./UNECE_MeasurementUnitCommonCodeVolume_4.xsd"/>
```

クラス用語の定義（「人」の例）

```
<xsd:complexType name="人型">
  <xsd:annotation>
    <xsd:documentation xml:lang="ja">人の情報を表現するためのデータ型</xsd:documentation>
  </xsd:annotation>
  <xsd:complexContent>
    <xsd:extension base="ic:実体型">
      <xsd:sequence>
        <xsd:element maxOccurs="unbounded" minOccurs="0" ref="ic:人_氏名"/>
        <xsd:element maxOccurs="1" minOccurs="0" ref="ic:人_性別"/>
        <xsd:element maxOccurs="1" minOccurs="0" ref="ic:人_性別-コード"/>
        <xsd:element maxOccurs="1" minOccurs="0" ref="ic:人_生年月日"/>
        <xsd:element maxOccurs="1" minOccurs="0" ref="ic:人_死亡年月日"/>
        <xsd:element maxOccurs="unbounded" minOccurs="0" ref="ic:人_住所"/>
        <xsd:element maxOccurs="1" minOccurs="0" ref="ic:人_本籍"/>
        <xsd:element maxOccurs="unbounded" minOccurs="0" ref="ic:人_連絡先"/>
        <xsd:element maxOccurs="unbounded" minOccurs="0" ref="ic:人_ID"/>
        <xsd:element maxOccurs="unbounded" minOccurs="0" ref="ic:人_国籍"/>
        <xsd:element maxOccurs="unbounded" minOccurs="0" ref="ic:人_国籍-コード"/>
        <xsd:element maxOccurs="1" minOccurs="0" ref="ic:人_出生国"/>
        <xsd:element maxOccurs="1" minOccurs="0" ref="ic:人_出生国-コード"/>
        <xsd:element maxOccurs="1" minOccurs="0" ref="ic:人_出生地"/>
      </xsd:sequence>
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>
```

プロパティ用語の定義（「人の氏名」の例）

```
<xsd:element name="人_氏名" type="ic:氏名型">
  <xsd:annotation>
    <xsd:documentation xml:lang="ja">氏名</xsd:documentation>
  </xsd:annotation>
</xsd:element>
```

(2) RDF 形式

シリアライズ後の RDF 形式の概要を示すため、以下に IMI コア語彙 v2 案の RDF 形式を一部抜粋したものを以下に示す。RDF 形式の表現形式には、Turtle、RDF/XML、JSON-LD などが存在するが、ここでは記述の一例として、可読性の高い Turtle による表記を用いる。

ヘッダ部：名前空間のプレフィックスと URI の対応を記載

```
@prefix cc: <http://creativecommons.org/ns#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ic: <http://imi.ipa.go.jp/ns/core/2#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix uncefactCode: <urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCode:9#> .
@prefix uncefactDuration: <urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeDuration:4#> .
@prefix uncefactISO4217: <urn:un:unece:uncefact:codelist:standard:ISO:ISO3AlphaCurrencyCode:2012-08-31#> .
@prefix uncefactLinear: <urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeLinear:4#> .
@prefix uncefactVolume: <urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeVolume:4#> .
@prefix uncefactWeight: <urn:un:unece:uncefact:codelist:standard:UNECE:MeasurementUnitCommonCodeWeight:4#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

語彙の定義：語彙の名称や説明等を記載

```
<http://imi.ipa.go.jp/ns/core/2> a owl:Ontology ;
  rdfs:label "IMI Core Vocabulary"@en,
    "共通語彙基盤コア語彙"@ja ;
  cc:license <http://creativecommons.org/publicdomain/zero/1.0/> ;
  dct:creator [ foaf:homepage <http://www.ipa.go.jp/> ;
    foaf:name "Information-technology Promotion Agency, Japan."@en,
      "IPA 独立行政法人情報処理推進機構"@ja ],
    [ foaf:homepage <http://www.meti.go.jp/> ;
    foaf:name "Ministry of Economy, Trade and Industry"@en,
      "経済産業省"@ja ] ;
  dct:issued "2014-09-11"^^xsd:date ;
  dct:license <http://creativecommons.org/publicdomain/zero/1.0/> ;
  dct:title "IMI Core Vocabulary"@en,
    "共通語彙基盤コア語彙"@ja ;
  rdfs:comment "IMI Core Vocabulary"@en,
    "コア語彙は、共通語彙基盤の基礎をなすもので、氏名、住所、組織等、あらゆる社会活動で使用される中核的な用語の集合です。多くのシステム間で情報交換のための基礎となる語彙で、データ交換、オープンデータの二次利用等の効率化に役立つものです。"@ja ;
  owl:versionInfo "2" .
```

クラス用語の定義（「人」の例）

```
ic:人 a rdfs:Class ;  
    rdfs:label "人"@ja ;  
    rdfs:comment "人の情報を表現するためのデータ型"@ja ;  
    rdfs:subClassOf ic:実体 .
```

プロパティ用語の定義（「人の氏名」の例）

```
ic:人_氏名 a rdf:Property ;  
    rdfs:label "氏名"@ja ;  
    rdfs:comment "氏名"@ja ;  
    rdfs:domain ic:人 ;  
    rdfs:range ic:氏名 .
```

(3) 自然言語形式

シリアライズ後の自然言語形式の概要を示すため、以下に IMI コア語彙 v2 案の自然言語形式を一部抜粋したものを以下に示す。

クラス用語の定義（「ic:人」クラスの例）

人

プロパティ用語の定義（「ic:人_氏名」プロパティの例）

人の氏名

プロパティ用語の定義（「ic:人_氏名」プロパティが指す「ic:氏名」クラスの「ic:氏名_姓」プロパティ）

人の氏名の姓

プロパティ用語の定義（「ic:人_住所」プロパティが指す「ic:住所」クラスの「ic:住所_表記」プロパティが指す「ic:定型住所」クラスの「ic:定型住所_都道府県」）

人の住所の表記の都道府県

4.2.5 語彙データのシリアライズ・デシリアライズ

語彙データの論理形式と物理形式の間の相互変換を可能とするため、論理形式と物理形式の対応関係についての整理を実施した。

以下に、語彙データの物理形式である XML 形式、RDF 形式、自然言語形式それぞれについて、論理形式と物理形式の対応を具体的に示す。

なお、XML 形式および RDF 形式では、語彙データの物理形式を記述するために、以下に示す語彙（名前空間）を参照した。

表 4.2.5.1 XML 形式および RDF 形式で参照される語彙（名前空間）

項番	語彙の名称	プレフィックス	名前空間 URI
1	ADMS	adms	http://purl.org/adms
2	Creative Commons	cc	http://creativecommons.org/ns#
3	Dublin Core Terms	dct	http://purl.org/dc/terms/
4	FOAF	foaf	http://xmlns.com/foaf/0.1/
5	OWL	owl	http://www.w3.org/2002/07/owl#
6	SKOS	skos	http://www.w3.org/2004/02/skos/core#
7	VANN	vann	http://purl.org/vocab/vann/
8	XML Schema	xsd	http://www.w3.org/2001/XMLSchema

(1) XML シリアライズ・デシリアライズ（論理形式と XML 形式との対応）

XML 形式へのシリアライズは、論理形式で表現された語彙データを、XML Schema 文書として表現する処理である。XML Schema のデザインパターンとしては、NIEM や IMI(平成 24 年度版)と同様に、全ての複合型(xsd:complexType)と要素(xsd:element)をトップレベルで(xsd:schema 要素の子要素として) 定義し、グローバルに参照・再利用可能とする ” Garden of Eden” パターンを採用した。

「語彙」は、独立した XML Schema 文書として生成される。

「クラス用語」は XML Schema の複合型(xsd:complexType)と要素(xsd:element)として生成され、また、「プロパティ用語」は XML Schema の要素(xsd:element)として生成され、どちらも「語彙」から生成された XML Schema 文書に包含される。

「コードリスト」は XML Schema の単純型(xsd:simpleType)として生成され、「語彙」から生成された XML Schema 文書に包含される。

「コード」は、「コードリスト」から生成された XML Schema の単純型(xsd:simpleType)の中の、列挙(xsd:enumeration)として生成される。

以下、語彙データの論理形式と XML 形式との対応関係について、「語彙」、「クラス用語」、「プロパティ用語」、「コードリスト」、「コード」のそれぞれにおける詳細を示す。

1) 「語彙」に関する論理形式と XML 形式の対応

語彙データの論理形式に含まれる「語彙」は、XML 形式においては、一つの XML Schema 文書に対応する。論理形式の「語彙」の構成要素は、XML 形式においては XML Schema 文書内で、xsd:schema 要素の直下にある xsd:annotation 要素に対応する。

具体的には、「語彙」に関する論理形式と XML 形式の対応は、XPath を用いて以下のように記述できる。

表 4.2.5.2 「語彙」の論理形式と XML 形式の対応

項番	論理形式 「語彙」の 構成要素	物理形式 (XML 形式) XML Schema 内の出現位置 (XPath 表現)
1	名称	/xsd:schema/xsd:annotation/xsd:appinfo/dct:title
2	説明	/xsd:schema/xsd:annotation/xsd:documentation
3	作成日	/xsd:schema/xsd:annotation/xsd:appinfo/dct:issued
4	更新日	/xsd:schema/xsd:annotation/xsd:appinfo/dct:modified
5	廃止日	/xsd:schema/xsd:annotation/xsd:appinfo/dct:valid
6	作成者	/xsd:schema/xsd:annotation/xsd:appinfo/dct:creator
7	整備状況	/xsd:schema/xsd:annotation/xsd:appinfo/adms:status
8	バージョン	/xsd:schema/xsd:annotation/xsd:appinfo/owl:versionInfo
9	ライセンス	/xsd:schema/xsd:annotation/xsd:appinfo/dct:license
10		/xsd:schema/xsd:annotation/xsd:appinfo/cc:license
11	権利情報	/xsd:schema/xsd:annotation/xsd:appinfo/dct:rights
12	優先プレフィックス	/xsd:schema/xsd:annotation/xsd:appinfo/vann:preferredNamespacePrefix
13	優先名前空間 URI	/xsd:schema/xsd:annotation/xsd:appinfo/vann:preferredNamespaceUri

2) 「クラス用語」に関する論理形式と XML 形式の対応

語彙データの論理形式に含まれる「クラス用語」は、XML 形式においては、当該「クラス用語」が属する「語彙」に対応する XML Schema 文書内で、xsd:schema 要素の直下の xsd:complexType 要素と、当該 xsd:complexType 要素を型としてもつ xsd:element 要素に対応する。

具体的には、「クラス用語」に関する論理形式と XML 形式の xsd:complexType 要素との関係は、以下のように XPath を用いて表現される。

表 4.2.5.3 「クラス用語」の論理形式と XML 形式(xsd:complexType 要素)の対応

項番	論理形式 「クラス用語」の 構成要素	物理形式 (XML 形式) XML Schema 内の出現位置 (XPath 表現)
1	名称	/xsd:schema/xsd:complexType/@name ※名称の末尾には「型」を付与する（例：「人型」）
2	説明	/xsd:schema/xsd:complexType/xsd:annotation/xsd:documentation
3	別名称	/xsd:schema/xsd:complexType/xsd:annotation/xsd:appinfo/skos:altLabel
4	出典	/xsd:schema/xsd:complexType/xsd:annotation/xsd:appinfo/dct:source
5	完全一致	/xsd:schema/xsd:complexType/xsd:annotation/xsd:appinfo/skos:exactMatch
6	類似	/xsd:schema/xsd:complexType/xsd:annotation/xsd:appinfo/skos:closeMatch
7	広義	/xsd:schema/xsd:complexType/xsd:annotation/xsd:appinfo/skos:broadMatch
8	狭義	/xsd:schema/xsd:complexType/xsd:annotation/xsd:appinfo/skos:narrowMatch
9	関連	/xsd:schema/xsd:complexType/xsd:annotation/xsd:appinfo/skos:relatedMatch
10	上位用語	/xsd:schema/xsd:complexType/xsd:complexContent/xsd:extension/@base ※名称の末尾には「型」を付与する（例：「ic:実体型」）
11	プロパティ用語	/xsd:schema/xsd:complexType/xsd:complexContent/xsd:extension/xsd:sequence/xsd:element/@ref

また、「クラス用語」に関する論理形式と XML 形式の xsd:element 要素との関係は、以下のよう
に XPath を用いて表現される。

表 4.2.5.4 「クラス用語」の論理形式と XML 形式(xsd:element 要素)の対応

項番	論理形式 「クラス用語」の 構成要素	物理形式 (XML 形式) XML Schema 内の出現位置 (XPath 表現)
1	名称	/xsd:schema/xsd:element/@name
2	識別子	/xsd:schema/xsd:element/@type ※識別子の末尾には「型」を付与する（例：「人型」）
3	説明	/xsd:schema/xsd:element/xsd:annotation/xsd:documentation
4	別名称	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:altLabel
5	出典	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/dct:source
6	完全一致	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:exactMatch
7	類似	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:closeMatch
8	広義	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:broadMatch
9	狭義	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:narrowMatch
10	関連	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:relatedMatch

3) 「プロパティ用語」に関する論理形式と XML 形式の対応

語彙データの論理形式に含まれる「プロパティ用語」は、XML 形式においては、当該「プロパティ用語」が属する「語彙」に対応する XML Schema 文書内で、xsd:schema の直下の xsd:element 要素に対応する。さらに、それが含まれる「クラス用語」（「包含されるクラス用語」属性で特定される「クラス用語」）に対応する xsd:complexType 要素の中で、xsd:element への参照(ref 属性の値)として記述される。

具体的には、「プロパティ用語」に関する論理形式と XML 形式の xsd:element 要素との関係は、以下のように XPath を用いて表現される。

表 4.2.5.5 「プロパティ用語」の論理形式と XML 形式(トップレベルの xsd:element 要素)の対応

項番	論理形式 「プロパティ用語」の構成要素	物理形式 (XML 形式) XML Schema 内の出現位置 (XPath 表現)
1	名称	/xsd:schema/xsd:element/@name ※「包含されるクラス用語」の「名称」と「プロパティ用語」の「名称」をアンダースコア記号 (“_”) で接続した値を用いる (例: 「人_氏名」)
2	説明	/xsd:schema/xsd:element/xsd:annotation/xsd:documentation
3	別名称	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:altLabel
4	出典	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/dct:source
5	完全一致	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:exactMatch
6	類似	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:closeMatch
7	広義	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:broadMatch
8	狭義	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:narrowMatch
9	関連	/xsd:schema/xsd:element/xsd:annotation/xsd:appinfo/skos:relatedMatch
10	値域	/xsd:schema/xsd:element/@type

また、論理形式に含まれる「プロパティ用語」と、それが含まれる「クラス用語」（「包含されるクラス用語」属性で特定される「クラス用語」）に対応する `xsd:complexType` 要素の中の `xsd:element` 要素との関係は、以下のように XPath を用いて表現される。

表 4.2.5.6 「プロパティ用語」の論理形式と XML 形式(`xsd:complexType` 中の `xsd:element` 要素)の対応

項番	論理形式 「プロパティ用語」 の構成要素	物理形式 (XML 形式) XML Schema 内の出現位置 (XPath 表現)
1	最小出現回数	<code>/xsd:schema/xsd:complexType[@name = \${プロパティ用語の包含されるクラス用語の名 称}]/xsd:complexContent/xsd:extension/xsd:sequence/ xsd:element[@ref = \${プロパティ用語の名称}]/@minOccurs</code> ※最小出現回数の値が存在しない場合は、“0” を値として用いる
2	最大出現回数	<code>/xsd:schema/xsd:complexType[@name = \${プロパティ用語の包含されるクラス用語の名 称}]/xsd:complexContent/xsd:extension/xsd:sequence/ xsd:element[@ref = \${プロパティ用語の名称}]/@maxOccurs</code> ※最大出現回数の値が存在しない場合は、“unbounded” を値として用いる
3	順序	(数値としては現れず、 <code>/xsd:schema/xsd:complexType[@name = \${包含されるクラス用語の名 称}]/xsd:complexContent/xsd:extension/xsd:sequence/xsd:element</code> の出現順序として現れる。)

4) 「コードリスト」に関する論理形式と XML 形式の対応

語彙データの論理形式に含まれる「コードリスト」は、XML 形式においては、当該「コードリスト」が属する「語彙」に対応する XML Schema 文書内で、xsd:schema 要素の直下の xsd:simpleType 要素に対応する。

具体的には、「コードリスト」に関する論理形式と XML 形式の対応は、XPath を用いて以下のように記述される。

表 4.2.5.7 「コードリスト」の論理形式と XML 形式の対応

項番	論理形式 「コードリスト」の 構成要素	物理形式 (XML 形式) XML Schema 内の出現位置 (XPath 表現)
1	名称	/xsd:schema/xsd:simpleType/@name ※名称の末尾には「型」を付与する（例：「人型」）
2	説明	/xsd:schema/xsd:simpleType/xsd:annotation/xsd:documentation
3	出典	/xsd:schema/xsd:simpleType/xsd:annotation/xsd:appinfo/dct:source

5) 「コード」に関する論理形式と XML 形式の対応

語彙データの論理形式に含まれる「コード」と、それが含まれる「コードリスト」に対応する xsd:simpleType 要素の中の xsd:enumeration 要素との関係は、以下のように XPath を用いて表現される。

表 4.2.5.8 「コード」の論理形式と XML 形式の対応

項番	論理形式 「コード」の 構成要素	物理形式 (XML 形式) XML Schema 内の出現位置 (XPath 表現)
1	名称	/xsd:schema/xsd:simpleType[@name = \${「コードリスト」の名称}]/xsd:restriction/xsd:enumeration/@value
2	説明	/xsd:schema/xsd:simpleType[@name = \${「コードリスト」の名称}]/xsd:restriction/xsd:enumeration/xsd:annotation/xsd:documentation
3	別名称	/xsd:schema/xsd:simpleType[@name = \${「コードリスト」の名称}]/xsd:restriction/xsd:enumeration/xsd:annotation/xsd:appinfo/skos:altLabel
4	出典	/xsd:schema/xsd:simpleType[@name = \${「コードリスト」の名称}]/xsd:restriction/xsd:enumeration/xsd:annotation/xsd:appinfo/dct:source

(2) RDF シリアライズ・デシリアライズ（論理形式と RDF 形式との対応）

RDF 形式へのシリアライズは、論理形式で表現された語彙データを、W3C RDF Schema を用いた RDF データとして表現する処理である。

以下、語彙データの論理形式と RDF 形式との対応関係について、「語彙」、「クラス用語」、「プロパティ用語」それぞれについての詳細を示す。

1) 「語彙」に関する論理形式と RDF 形式の対応

語彙データの論理形式に含まれる「語彙」は、RDF 形式においては、owl:Ontology クラスのインスタンスとして表現する。

「語彙」について、論理形式と RDF 形式の対応を以下に示す。

表 4.2.5.9 「語彙」の論理形式と RDF 形式の対応

項番	論理形式 「語彙」の構成要素	物理形式 (RDF 形式) owl:Ontology クラスのインスタンス
1	名称	rdfs:label
2		dct:title
3	説明	rdfs:comment
4	作成日	dct:issued
5	更新日	dct:modified
6	廃止日	dct:valid
7	作成者	dct:creator
8	整備状況	adms:status
9	バージョン	owl:versionInfo
10	ライセンス	dct:license
11		cc:license
12	権利情報	dct:rights
13	優先プレフィックス	vann:preferredNamespacePrefix
14	優先名前空間 URI	vann:preferredNamespaceUri

2) 「クラス用語」に関する論理形式と RDF 形式の対応

語彙データの論理形式に含まれる「クラス用語」は、RDF 形式においては、`rdfs:Class` クラスのインスタンスとして表現する。

「クラス用語」について、論理形式と RDF 形式の対応を以下に示す。

表 4.2.5.10 「クラス用語」の論理形式と RDF 形式の対応

項番	論理形式 「クラス用語」の構成要素	物理形式 (RDF 形式) <code>rdfs:Class</code> クラスのインスタンス
1	名称	<code>rdfs:label</code>
2	識別子	インスタンスの識別子(IRI)
3	説明	<code>rdfs:comment</code>
4	別名称	<code>skos:altLabel</code>
5	出典	<code>dct:source</code>
6	完全一致	<code>skos:exactMatch</code>
7	類似	<code>skos:closeMatch</code>
8	広義	<code>skos:broadMatch</code>
9	狭義	<code>skos:narrowMatch</code>
10	関連	<code>skos:relatedMatch</code>
11	上位用語	<code>rdfs:subClassOf</code>

3) 「プロパティ用語」に関する論理形式と RDF 形式の対応

語彙データの論理形式に含まれる「プロパティ用語」は、RDF 形式においては、rdf:Property クラスのインスタンスとして表現する。

「プロパティ用語」について、論理形式と RDF 形式の対応を以下に示す。

表 4.2.5.11 「プロパティ用語」の論理形式と RDF 形式の対応

項番	論理形式 「プロパティ用語」の構成要素	物理形式 (RDF 形式) rdf:Property クラスのインスタンス
1	名称	rdfs:label
2	識別子	インスタンスの識別子(IRI)
3	説明	rdfs:comment
4	別名称	skos:altLabel
5	出典	dct:source
6	完全一致	skos:exactMatch
7	類似	skos:closeMatch
8	広義	skos:broadMatch
9	狭義	skos:narrowMatch
10	関連	skos:relatedMatch
11	包含されるクラス用語	rdfs:domain
12	値域	rdfs:range

(3) 自然言語シリアライズ（論理形式と自然言語形式との対応）

自然言語形式へのシリアライズは、論理形式で表現された語彙データを、自然言語で記述された人間のためのドキュメント（仕様書や表など）に活用可能とするため、プロパティ用語を自然言語として記述するための処理である。

XML形式やRDF形式におけるシリアライズは、語彙に含まれる用語について、用語の表記や意味、構造や関係などの情報をXMLやRDFといったデータ形式で（物理的に）表現する処理であった。一方、自然言語形式におけるシリアライズは、語彙に含まれる個々の用語を、仕様書や表のカラム名などに利用するための自然言語に変換する処理であり、シリアライズの対象は個々のプロパティ用語のみとなる。

自然言語形式へのシリアライズの概要を以下に示す。

ステップ1: シリアライズ対象とするプロパティ用語の特定

シリアライズ対象とするプロパティ用語を選択する。例えば、地点の名称と説明を表現するため、「ic:地点」クラスに含まれるプロパティ用語である、「ic:地点_名称」および「ic:地点_説明」を選択する。

ステップ2: プロパティ用語のシリアライズ

ステップ1で特定したプロパティ用語について、アンダースコア記号(“_”)を日本語の文字である『の』で置換する。例えば、上記の例では、「ic:地点の名称」および「ic:地点の説明」が得られる。

ステップ3: プロパティ用語の指すクラス用語のシリアライズ

ステップ2で得られたプロパティ用語の各々について、以下の処理を実施する。

当該プロパティ用語の値域がクラス用語でない場合（xsd:stringなどのデータタイプである場合）は、処理は不要。上記の例では、プロパティ用語「ic:地点の説明」がこれに該当する。

当該プロパティ用語の値域がクラス用語である場合には、当該クラス用語に含まれるプロパティ用語について、上記のステップ1からステップ3までの手順を再帰的に実施する。上記の例では、プロパティ用語「ic:地点_名称」の値域がクラス用語「ic:名称」であるため、クラス用語「ic:名称」について、ステップ1からステップ3までの手順を実施することとなる。ここでは、ステップ1で、クラス用語「ic:名称」のプロパティ用語の内、「ic:名称_表記」および「ic:名称_表記(カナ)」を選択し、ステップ3で「ic:名称の表記」、「ic:名称の表記(カナ)」を得たとする。「表記」および「表記(カナ)」の値域はどちらもデータタイプ(xsd:string)であってクラス用語ではないため、再帰的な処理はここで終了する。

ステップ 4: 用語の結合

ステップ 3 までで得られた結果について、あるプロパティ用語から再帰的に得られたプロパティ用語については、これらの結果を（文字列として）結合する。上記の例では、「ic:地点の説明」と、「ic:地点の名称」と、そこから再帰的に得られた「ic:名称の表記」および「ic:名称の表記(カナ)」が存在するため、結果として「ic:地点の説明」、「ic:地点の名称の ic:名称の表記」、「ic:地点の名称の ic:名称の表記(カナ)」が得られる。

ステップ 5: 繰り返しの除去

ステップ 4 で得られた結果の内、プロパティ用語の名称とその後に続くクラス用語の名称が同一である場合は、プロパティ用語と後続する『の』の文字を除去する。上記の例では、「ic:地点の名称の ic:名称の表記」および「ic:地点の名称の ic:名称の表記(カナ)」における『名称』がこの規則に該当するため、結果として「ic:地点の ic:名称の表記」および「ic:地点の ic:名称の表記(カナ)」が得られる。

ステップ 6: プレフィックスの除去

シリアライズ結果を利用する範囲で、単一の名前空間しか使われない場合は、先頭以外のプレフィックスを除去することが可能である。上記の例では先頭以外の「ic:」を除去することが可能となり、結果として「ic:地点の説明」、「ic:地点の名称の表記」、「ic:地点の名称の表記(カナ)」が得られる。

上記の 6 ステップの中で例示したシリアライズ結果については、例えば、以下のように表の列名として利用することが可能である。

項番	ic:地点の名称の表記	ic:地点の名称の表記(カナ)	ic:地点の説明
1	富士山頂	フジサンチョウ	富士山の山頂。…
2	…	…	…
…	…	…	…

4.3 IMI コア語彙 v2 案の整備結果

4.3.1 IMI コア語彙 v2 案の用語の全体像

IMI コア語彙 v2 案に含まれる用語の全体像を示すため、構造をもった用語（その意味あるいは概念の一部として、他の用語との関係をもつ用語）について、以下に抜粋する。IMI コア語彙 v2 案の正式な情報の全体については、別紙を参照のこと。

表 4.3.1.1 IMI コア語彙 v2 案に含まれる主要な用語（構造をもった用語）

項番	分類	項目説明
1	人	人の情報を表現するためのデータ型。
2	名称	組織や地点等の名称を表現するためのデータ型。
3	氏名	人の氏名を表現するためのデータ型。
4	住所	住所を表現するためのデータ型。
5	定型住所	要素化された住所を表現するためのデータ型。
6	連絡先	連絡先を表現するためのデータ型。
7	電話番号	電話番号を表現するためのデータ型。
8	組織	組織を表現するためのデータ型。
9	業務組織	業務組織を表現するためのデータ型。
10	法人	法人を表現するためのデータ型。
11	従業員数	従業員数を表現するためのデータ型。
12	構成員	組織等を構成する人を表現するデータ型。
13	組織関連	組織の関連を表現するためのデータ型。
14	場所	場所を表現するためのデータ型。
15	座標	座標を表現するためのデータ型。
16	度分秒	経緯度座標系を表現するためのデータ型。
17	ID	識別子を表現するためのデータ型。
18	ID 体系	ID の種別を表現するためのデータ型。
19	製品	物を表現するためのデータ型。
20	製品個品	製品を表現するためのデータ型。
21	地点	施設や建物など、参照対象となる地点を表すためのデータ型。
22	施設	特定のサービスをする提供する施設を表現するためのデータ型。
23	施設関連	施設の関連を表現するためのデータ型。
24	建物	具体的な建物を表現するためのデータ型。
25	駐車場	駐車場を表現するためのデータ型。

項番	分類	項目説明
26	設備	施設や建物等に備わっている設備を表現するためのデータ型。
27	イベント	イベントを表現するためのデータ型。
28	測定値	測定単位及び方法を表現するためのデータ型。
29	数量	数量を表現するためのデータ型。
30	容量	容量を表現するためのデータ型。
31	面積	面積を表現するためのデータ型。
32	重量	重量を表現するためのデータ型。
33	長さ	長さを表現するためのデータ型。
34	時間	時間の長さを表現するためのデータ型。
35	金額	金額を表現するためのデータ型。
36	価格一覧	価格の一覧を表現するためのデータ型。
37	価格	価格を表現するためのデータ型。
38	実体	実体を表現するためのデータ型。
39	状況	状況を表現するためのデータ型。
40	日時	日付を表現するためのデータ型。
41	定型日時	定型日付を表現するためのデータ型。
42	定型和暦日時	定型和暦日付を表現するためのデータ型。
43	期間	期間を表現するためのデータ型。
44	スケジュール	スケジュールを表現するためのデータ型。
45	イベントスケジュール	施設等のオープン日とオープン時間を表現するためのデータ型。
46	定期スケジュール	施設等のオープン日とオープン時間を表現するためのデータ型。
47	詳細スケジュール	(将来的に整備する予定) 施設等のオープン日とオープン時間を表現するためのデータ型。複数のスケジュールが列挙された場合はそれらの和を取るものとする。
48	詳細スケジュール規則	(将来的に整備する予定) 繰り返し型のスケジュールを定義するための規則。
49	コード	コードを表現するためのデータ型。
50	コードリスト	コードの種別を表現するためのデータ型。
51	アクセス	地点等へのアクセスを表現するためのデータ型。
52	事物	全ての型のベースとなる基本型。

5. ドメイン語彙サンプルの整備

本事業では、共通語彙基盤の導入に向け、語彙データベースに格納する語彙の初期データとして、4章で検討したコア語彙以外にも、3.3.1に示したドメイン共通語彙やドメイン固有語彙を整備し、将来的に格納される語彙を拡大していくことが必要となる。

そこで、本事業では、ドメイン共通語彙やドメイン固有語彙を継続的に整備していくための施策として、国内外の各戦略や取組に規定されるドメインを参考に、試行的に、地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインの4ドメインを選定した。さらに、対象分野の専門家を交えた十分な検討の場としてのドメイン語彙検討会を通じて、ドメイン共通語彙としての用語をリストアップし、4章で検討したデータ構造に従って、ドメイン語彙サンプルとして整備し、その収集・整備プロセスの一案についても示した。

5.1 ドメイン語彙サンプルの整備の目的

先行事例である平成24年度事業では、IMIコア語彙v1の整備の他、試行的に、防災備蓄品のドメイン語彙サンプルの整備を行った。

例えば、施設（地理空間・施設ドメイン）の利用シーンとして、「学校」を表現する場合、コア語彙で整備済である「住所」、「連絡先」などの語彙の他に、「生徒数」、「職員室」、「体育館」などの学校を表現するための固有の用語を組み合わせて表現するため、コア語彙とドメイン語彙との関係を明確化する必要がある。

また、避難所（防災ドメイン）として、「学校」を表現する場合は、コア語彙で整備済みである「住所」「連絡先」などの語彙の他に、「災害種別」、「屋内収容可能人数」などの避難所を表現するための固有の用語を組み合わせて表現する必要があり、一つの用語を表す場合においても、利用シーンやドメインにおいて、表現するための用語や表現方法が異なることから、ドメイン間の語彙の関係を明確化する必要がある。

本事業では、ドメイン共通語彙やドメイン固有語彙を継続的に整備していくため、4ドメインにおいて、コア語彙とドメイン語彙との関係の明確化やドメイン間の語彙の関係を明確化するとともに、ドメイン語彙サンプルを整備することで、その収集・整備プロセスを構築することとした。

5.2 ドメイン語彙収集・整備プロセスの仮説

5.2.1 ドメインの選定

本事業では、ドメイン共通語彙やドメイン固有語彙を継続的に整備していくための施策として、地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインの4ドメインを試行的に選定した。

(1) ドメイン選定の観点

ドメイン選定においては、共通語彙基盤の将来的な活用普及の観点として、現在は具体的なデータが公開されていないものの、政府の戦略やニーズに基づき、今後、公開や活用が見込まれるデータの種類をドメイン選定の観点に加え、ドメインを選定することとした。

以下にドメイン選定の考え方について示す。

表 5.2.1.1 ドメイン選定の考え方

項番	ドメイン選定の観点		戦略や取組等
1	相互運用性	先進諸外国の語彙体系	① 「NIEM」のドメイン ② ISA「JOINUP」のコアボキャブラリ
2		国内の行政事務におけるデータ標準化の取組	③ APPLIC「アプリケーションユニット及び業務ユニット」の対象ユニット
3		国内外のデータカタログサイト	④ 「Open DATA METI サイト」のデータグループ ⑤ 「Data.gov」のコミュニティカテゴリ ⑥ 「Data.gov.uk」、「国立統計局」、「ロンドン市 Data.gov」の公開データ分野
4	将来的な活用普及	IT戦略により将来的に活用が期待される分野	⑦ 「世界最先端 IT 国家創造宣言」の利用シーン
5		オープンデータで公開が期待される分野	⑧ IT 総合戦略本部「電子行政オープンデータ推進のためのロードマップ」の重点分野 ⑨ G8「オープンデータ憲章」の高価値データの種類
6		民間ニーズ	⑩ 経団連「公共データの産業利用に関する調査結果」のニーズの高い公共データ(種類別)

(2) 国内外の各戦略や取り組みに規定されるドメインの抽出

国内外の各戦略や取り組みに規定されるドメインを抽出した一覧を以下に示す¹。国内外の各取り組みにおいて、半数以上の戦略や取組で取り上げられているドメインから、検討対象ドメインを選択する。国内外の各戦略や取り組みに規定されるドメインの中で、該当数が9と一番多い地理空間・施設ドメインをはじめとして、該当数が7の移動・交通ドメイン、財務ドメイン、該当数が6の防災ドメイン、医療・健康ドメイン、該当数が5の気象・環境ドメイン、統計・白書ドメイン、教育ドメイン、人の情報ドメインの9ドメインを選択対象候補とした。

以下に、国内外の各戦略や取り組みに規定されるドメインの抽出結果一覧を示す。

表 5.2.1.2 国内外の各戦略や取り組みに規定されるドメインの抽出結果一覧

ドメイン	該当数	① NIEM	② JOINUP	③ APPLIC	④ Open DATA METI サイト	⑤ Data.gov	⑥ Data.gov.uk、 国立統計局、 ロンドン市 Data.gov	⑦ 世界最 先端 IT 国家創 造宣言	⑧ 電子行政オ ープンデー タロードマ ップ	⑨ オープン データ憲 章技術的 別添	⑩ 公共データの 産業利用に関 する調査結果
地理空間・施設	9		○	○	○	○	○	○	○	○	○
移動・交通	7				○	○	○	○	○	○	○
財務	7			○	○		○	○	○	○	○
防災	6			○	○	○		○	○		○
医療・健康	6	○		○		○	○			○	○
気象・環境	5					○	○	○		○	○
統計・白書	5				○			○	○	○	○
教育	5			○		○	○	○		○	
人の情報	5		○	○			○	○			○
法人	4		○	○						○	○
犯罪と司法	4	○				○	○			○	
エネルギー	4					○	○			○	○
政府の説明責任と民主主義	4			○		○	○			○	
農業	3					○	○			○	
福祉	2	○								○	
研究・科学	3	○				○				○	
経済	2					○	○				
制度	2					○					○
開発・採掘	2					○				○	

¹ 国内外の各戦略や取り組みで規定されているドメインは、多岐に渡っており、そのドメインの分割方法も多岐に渡っている。ドメインの分割方法についての提言は、8章に記載する。

ドメイン	該 当 数	① NIEM	② JOINUP	③ APPLIC	④ Open DATA METI サイト	⑤ Data.gov	⑥ Data.gov.uk、 国立統計局、 ロンドン市 Data.gov	⑦ 世界最 先端 IT 国家創 造宣言	⑧ 電子行政オ ープンデー タロードマ ップ	⑨ オープン データ憲 章技術的 別添	⑩ 公共データの 産業利用に関 する調査結果
金融	2					○	○				
国際	2	○				○					
社会保険・税	1			○							
特許	1										○
生体情報	1	○									
サイバー	1	○									
緊急事態管理	1	○									
移民	1	○									
インフラ防護	1	○									
諜報	1	○									
文化、スポーツ	1						○				
海事	1	○									
労働	1						○				
監視	1	○									

(3) 4ドメインの選定結果

日本のオープンデータの取り組み方針が示されている、「電子行政オープンデータ推進のためのロードマップ（平成 25 年 6 月 14 日 高度情報通信ネットワーク社会推進戦略本部決定）¹」の「2 電子行政オープンデータ推進のための具体的な取組 （2）機械判読に適したデータ形式での公開の拡大 ②具体的な取組」では、統計データその他、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・調達情報から優先的に、費用対効果を踏まえつつ取り組むこととしており、上記ドメインは、日本における重点分野となっている。

よって、「表 5.2.1.2 国内外の各戦略や取り組みに規定されるドメインの抽出結果一覧」より選択対象候補とした9ドメインのうち、⑧電子行政オープンデータロードマップへの記載がされている地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインを検討対象ドメインとして試行的に選択した。

試行的に選択したドメインとその概要を以下に示す。

表 5.2.1.3 試行的に選択したドメインとその概要

項番	ドメイン	概要
1	地理空間・施設	地図データ及びその地図上にマッピングされる施設のデータを扱うドメイン
2	移動・交通	人の移動や公共交通(鉄道・バス)の移動に関するデータを扱うドメイン
3	財務	政府の公開する予算、決算、調達に係るデータを扱うドメイン
4	防災	平時における災害対策及び発災時における緊急情報の機関間の情報連携に必要なデータを扱うドメイン
5	統計・白書	府省の公開する統計や白書のデータを扱うドメイン
6	医療・健康	医療機関間でやり取りされる疾病情報や特定健診データを扱うドメイン
7	気象・環境	天候や大気汚染に関するデータを扱うドメイン
8	教育	学校の情報や成績等に関するデータを扱うドメイン
9	人の情報	住民や消費者のデータを扱うドメイン

¹ 「電子行政オープンデータ推進のためのロードマップ（平成 25 年 6 月 14 日 高度情報通信ネットワーク社会推進戦略本部決定）」 <http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou3.pdf>

5.2.2 ドメイン語彙の収集・整備プロセスの仮説

本事業では、ドメイン共通語彙やドメイン固有語彙を継続的に整備していくための施策として、地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインの4ドメインを試行的に選定した。

ドメイン共通語彙としての用語をリストアップし、ドメイン語彙サンプルとして整備するとともに、その収集・整備プロセスについても検証するため、まず、ISA等の海外の先進事例において公開されている語彙の収集・整備プロセスを参考にしつつ、データ標準化の経験者等を交えて検討を実施し、下記に示すようなドメイン語彙の収集・整備プロセスの仮説を立てた。

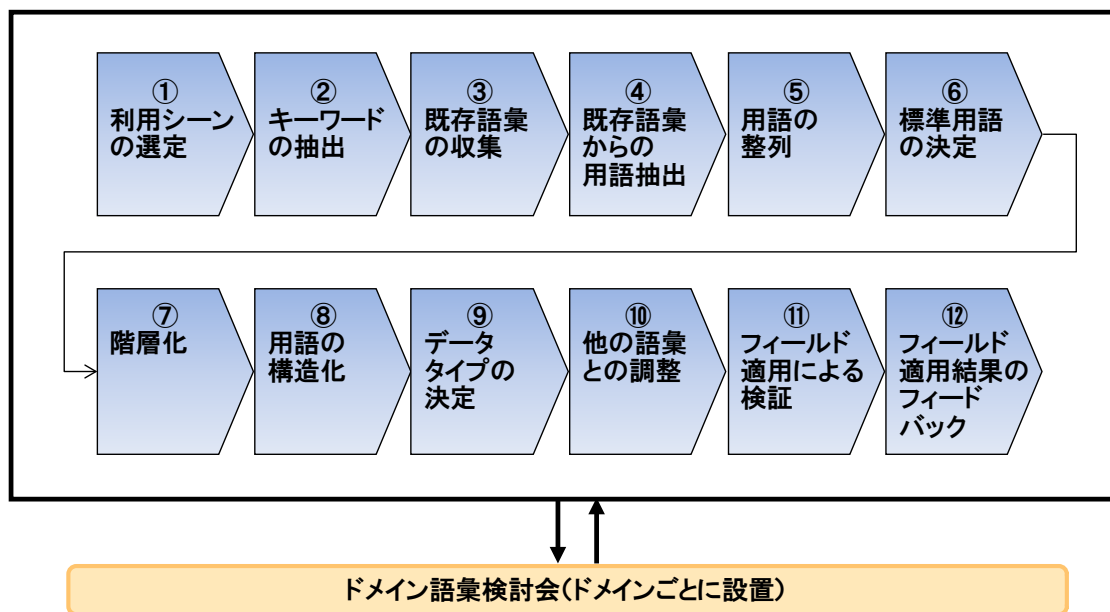


図 5.2.2.1 ドメイン語彙の収集・整備プロセスの仮説

上記、ドメイン語彙収集・整備プロセスの仮説の詳細を以下に示す。

表 5.2.2.1 ドメイン語彙収集・整備プロセスの仮説の詳細

項番	プロセス	プロセスの詳細説明および作業観点	入力	出力
1	①利用シーンの選定	選定したドメインにおいて語彙の利用ニーズの具体化および利用シーンを選定する	—	語彙利用シーン
2	②キーワードの抽出	利用シーンで使用するドメイン語彙のキーワードを選定 例えば、移動・交通ドメインであれば、駅、列車、路線などがキーワードにあたる	語彙利用シーン	語彙利用シーンにおけるキーワード
3	③既存語彙の収集	語彙利用シーンにおけるキーワードに基づき、当該ドメインに関する語彙を収集する。収集にあたっての作業観点としては、可能な限り、以下の3種の情報源から収集する ①コミュニティからの情報源： ドメイン内の標準的な用語をカバーする情報源 ②パブリックからの情報源： ドメイン内の公共系情報源 ③ソーシャル&ビジネスからの情報源： 民間・個人などで構成される情報源	・語彙利用シーン ・語彙利用シーンにおけるキーワード	既存語彙 (構造なし)
4	④既存語彙からの用語抽出	収集した既存語彙から用語の抽出を行う	既存語彙 (構造なし)	用語 (構造なし)
5	⑤用語の整列	抽出した用語の整列を行う	用語 (構造なし)	整列済の用語 (構造なし)
6	⑥標準用語の決定	多数の候補の中からドメイン語彙整備対象となる用語を選定するとともに、同じ概念を示している用語を特定し、リストアップする	整列済の用語 (構造なし)	ドメイン語彙整備対象となる用語群 (構造なし)
7	⑦階層化	抽出したドメイン語彙整備対象となる用語間でレベル合わせを行う	ドメイン語彙整備対象となる用語群 (構造なし)	ドメイン語彙整備対象となる用語群 (レベル合わせ済、 (構造なし))

項番	プロセス	プロセスの詳細説明および 作業観点	入力	出力
8	⑧用語の 構造化	レベル合わせを行ったドメイン 語彙整備対象となる用語群につ いて、用語間の親子関係や参照関 係、汎化-特化関係等の関係性を 整理する	ドメイン語彙整備 対象となる用語群 (レベル合わせ済、 (構造なし))	構造化済の用語
9	⑨データ タイプの 決定	構造化済の各用語のデータタイ プ(文字例、数値、時刻、日時、 URI、真偽値、コード等)を決定 する	構造化済の用語	検証前ドメイン 語彙サンプル
10	⑩他の語彙 との調整	整備した検証前ドメイン語彙サ ンプルに対して、コア語彙および 他ドメインで整備済の語彙を適 用する	・検証前ドメイン 語彙サンプル ・コア語彙 ・他ドメイン語彙	検証前ドメイン 語彙サンプル (他の語彙との 調整済み)
11	⑪フィールド 適用による 検証	検証前ドメイン語彙サンプル(他 の語彙との調整済み)に実データ を適用し、十分に記述可能である かを検証する	検証前ドメイン語 彙サンプル (他の語彙との調 整済み)	検証済ドメイン 語彙サンプル
12	⑫フィールド 適用結果の フィード バック	フィールド適用による検証の結 果、記述できなかった場合には、 ドメイン語彙のサンプルに対し て、用語の追加・修正を実施する	検証済ドメイン 語彙サンプル	検証済ドメイン 語彙サンプル (最終版)

5.3 ドメイン語彙サンプルの整備

「5.2.2 ドメイン語彙の収集・整備プロセスの仮説」で作成したドメイン語彙の収集・整備プロセスの仮説に従い、地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインの4ドメインを対象に、対象分野関係者を交えた十分な検討の場としてのドメイン語彙検討会を通じて、ドメイン共通語彙としての用語をリストアップし、4章で検討したデータ構造に従い、ドメイン語彙のサンプル（以下、ドメイン語彙サンプルという。）を試行的に整備した。

5.3.1 ドメイン語彙サンプルの整備の実施内容

以下、ドメイン語彙サンプルを整備した際の実施内容を示す。

(1) ①利用シーンの選定

選定したドメインにおいて語彙の利用ニーズの具体化および利用シーンを選定した結果について、以下に示す。

表 5.3.1.1 「①利用シーンの選定」での実施作業

項番	作業内容	ドメイン	作業結果/選定結果
1	選定したドメインにおいて語彙の利用ニーズの具体化および利用シーンを選定する	地理空間・施設	利用シーン：病院のマクロ・ミクロ情報を用いた利用者ナビゲーション
2		移動・交通	利用シーン：国内交通網を表す語彙の整備
3		財務	利用シーン：市区町村の予算情報を住民にとって理解しやすい表現で統一的に公開し、税金使途の透明化・自治体間の横比較を可能とする。
4		防災	特定の利用シーンを設定するのではなく、防災ドメインの語彙の収集・整備プロセスの整備と、「避難所」のドメイン語彙サンプルの整備

(2) ② キーワードの抽出

利用シーンで使用するドメイン語彙のキーワードを選定した結果について、以下に示す。

表 5.3.1.2 「② キーワードの抽出」での実施作業と作業結果

項番	作業内容	ドメイン	作業結果/選定結果
1	利用シーンで使用するドメイン語彙のキーワードを選定	地理空間・施設	キーワードの抽出結果： 病院、建物、便所、廊下、階段、 出入口、駐車場、施設、設備、バリアフリー
2		移動・交通	キーワードの抽出結果： 駅、列車、路線、バス、停留所、 運行状況、ダイヤ、
3		財務	キーワードの抽出結果： 一般会計歳出予算
4		防災	キーワードの抽出結果： 避難所

(3) ③既存語彙の収集

「(2) ② キーワードの抽出」にて抽出したキーワードに基づき、当該ドメインに関する既存語彙を収集した結果を以下に示す。

表 5.3.1.3 「③既存語彙の収集」での実施作業と作業結果

項番	作業内容	ドメイン	作業結果/選定結果
1	語彙利用シーンにおけるキーワードに基づき、当該ドメインに関する既存語彙を収集する	地理空間・施設	<p>既存語彙の収集結果：</p> <p>①コミュニティ：ドメイン内における整備済の語彙</p> <ul style="list-style-type: none"> ・ IFC 4 (Industry Foundation Classes Release 4) ・ IMI コアボキャブラリ、IMI 施設ドメインボキャブラリ α ・ IMI 開発支援キット(公共施設、公共設備) ・ NIEM 3.0 ・ Schema.org ・ 国土交通省国土政策局 国土数値情報（公共施設）製品仕様書 第 4.0 版 他 <p>②パブリック：ドメイン内の公共系語彙</p> <ul style="list-style-type: none"> ・ 東京都医療機関案内サービス ・ 確認申請書（建築物）、病院開設等の許可申請書 ・ 上記申請書等が参照している法令（建築基準法・医療法） ・ バリアフリー法チェックシート ・ 施設および固定資産の管理台帳 <p>③ソーシャル&ビジネス：民間・個人などで構成される語彙</p> <ul style="list-style-type: none"> ・ 各種病院の Web ページ ・ 多目的トイレマップ ・ トイレマップ東京都エリア
2		移動・交通	<p>既存語彙の収集結果：</p> <p>①コミュニティ：ドメイン内における整備済の語彙</p> <ul style="list-style-type: none"> ・ 国土交通省自動車交通局 公共交通情報データ標準仕様書 ・ 国土交通省 国土計画局国土数値情報 製品仕様書（鉄道第 2.1 版、バスルート第 1.0 版、バス停留所第 2.0 版） ・ 公益財団法人 国土地理協会 緯度経度付き全国沿線・駅データベース <p>②パブリック：ドメイン内の公共系語彙</p> <ul style="list-style-type: none"> ・ 国土交通省 第 1 1 回大都市交通センサス ・ 国土交通省鉄道局監修 平成 2 4 年度 鉄道要覧 ・ 東京都交通局、大阪市交通局 Web ページ（地下鉄） <p>③ソーシャル&ビジネス：民間・個人などで構成される語彙</p> <ul style="list-style-type: none"> ・ JR 東日本 / 西日本、東京メトロ、広島電鉄の Web ページ ・ 交通新聞社 コンパス時刻表、JR 電車編成表 ・ Wikipedia

項番	作業内容	ドメイン	作業結果/選定結果
3	語彙利用シーンにおけるキーワードに基づき、当該ドメインに関する既存語彙を収集する	財務 (自治体 ¹)	<p>既存語彙の収集結果： 公共機関が提供する、市町村一般会計歳出予算に関する以下の語彙</p> <p>①地方自治法施行規則</p> <p>②総務省「地方財政状況調査 平成22年度市町村別決算状況調」</p> <p>③ ①に基づき作成された、各自治体の一般会計予算資料のうち「款」「項」「目」のレベルで各費目の説明が記載されている語彙</p> <ul style="list-style-type: none"> ・平成25年度 松阪市一般会計予算書 ・平成25年度 津市一般会計・特別会計予算書 (附 予算に関する説明書) ・平成25年度 館山市予算書・館山市予算に関する説明書 ・平成25年度 三重県伊賀市一般会計予算書及び予算に関する説明書 ・平成25年度 小金井市一般会計・特別会計歳入歳出予算事項別明細書 ・平成25年度 新潟市予算説明書(一般会計) ・平成25年度 南魚沼市一般・特別会計予算並びに予算に関する説明書 ・平成24年度 南三陸町一般会計・特別会計当初予算書(案)並びに予算に関する説明書 ・平成25年度 石巻市当初予算書・補正予算説明書
4		財務 (税どこ ²)	<p>既存語彙の収集結果：</p> <p>①各自治体の税金の使われ方を可視化する Web サイト「税金はどこへ行った？」(以下、「税どこ」という。)サイトにおいて、より理解しやすい用語を使用している、松阪市、津市等のサイト</p> <p>②予算分類に使用され、法令や標準等で整備済の語彙</p> <ul style="list-style-type: none"> ・COFOG(政府機能分類)：中央政府の支出項目を分類 ・歳出予算用の語彙について、「⑥標準用語の決定」で選出した、語彙整備対象となる用語群

¹ 財務ドメインのうち、市町村一般会計歳出予算に使用されている語彙を扱うドメイン

²各自治体の税金の使われ方を可視化する Web サイト「税金はどこへ行った？」で使用されている語彙を扱うドメイン

項番	作業内容	ドメイン	作業結果/選定結果
5	語彙利用シーンにおけるキーワードに基づき、当該ドメインに関する既存語彙を収集する	防災	<p>既存語彙の収集結果：</p> <p>①過去に実システムまたは実証実験のシステムで適用実績のある、整備済の語彙</p> <ul style="list-style-type: none"> ・官民協働危機管理クラウドシステムで使われる語彙 ・APPLIC 防災業務アプリケーションユニット標準仕様 ・地震 / 津波 / 洪水 / 土砂災害 被害想定データ製品仕様書 ・公共情報コモンズ(災害情報共有システム(Lアラート)) ・デジタル放送 地域情報共通 XML フォーマット TVCML Version2.1 情報表現ガイドライン 第 1.0 版 ・減災情報共有プラットフォーム 情報スキーマ ・内閣官房 閣副安危第464号 「避難施設データベースの整備について」 <p>②防災業務を規定した計画・マニュアル・実施要領等の語彙</p> <ul style="list-style-type: none"> ・地域防災計画（神奈川県相模原市、千葉県船橋市、愛知県名古屋市、愛媛県松山市、東京都目黒区、大阪府堺市、東京都練馬区、愛知県半田市、兵庫県小野市） ・避難所運営マニュアル（宮城県仙台市、千葉県船橋市、北海道札幌市、愛媛県松山市） <p>③避難所の情報を提供する web ページの語彙</p> <ul style="list-style-type: none"> ・ファーストメディア 防災情報 全国避難所ガイド ・Google Crisis Response 避難所情報 ・Yahoo! Japan 避難所マップ ・ナビット 全国避難所データベース ・PAERA-Hazard 全国避難所施設データベース ・COCOAME / COCOYUKI (福井県の雨・水位マップ / 雪・道路情報)

(4) ④ 既存語彙からの用語抽出

「(3) ③既存語彙の収集」にて収集した既存語彙から用語を抽出する際の作業概要及び作業観点を以下に示す。

表 5.3.1.4 「④ 既存語彙からの用語抽出」での作業概要及び作業観点

項番	作業内容	ドメイン	作業概要/作業観点
1	収集した既存語彙から用語の抽出を行う	地理空間・施設	作業観点： ①ドメイン内における整備済の語彙については、用語の表記のみではなく、語彙の構造と共に抽出する。 ②構造をもっていない既存語彙については、見出し語として使用されている用語を中心に抽出する。
2		移動・交通	
3		財務 (自治体)	作業観点： ①各自治体の一般会計予算資料のうち、「款」「項」「目」のレベルで各費目の説明が記載されている既存語彙から用語を抽出する。 ②各地方自治体の税どこサイトで使用されている用語を抽出する。
4		財務 (税どこ)	
5		防災	作業観点： ①ドメイン内における整備済の語彙については、用語の表記のみではなく、語彙の構造と共に抽出する。 ②構造をもっていない既存語彙については、見出し語として使用されている用語を中心に抽出する。

(5) ⑤用語の整列

「(4) ④ 既存語彙からの用語抽出」にて抽出した用語を整列する際の作業概要及び作業観点を以下に示す。

表 5.3.1.5 「⑤用語の整列」での作業概要及び作業観点

項番	作業内容	ドメイン	作業概要/作業観点
1	用語の整列を行う	地理空間・施設	既存語彙から抽出した各用語を比較し、構造化された用語を生成するため、下記の基準で整列を実施する。 ①表記または意味が同一である用語は、同一の行に並べる。 ②包含関係が認められる用語は、行を分けて、隣接する行に並べる。
2		移動・交通	
3		財務 (自治体)	
4		財務 (税どこ)	
5		防災	

(6) ⑥標準用語の決定

標準用語を決定する基準となり得る既存語彙を選定するため、「(5) ⑤用語の整列」で整列した複数の用語から、ドメイン語彙サンプルとなる用語を選定するとともに、同じ概念を示している用語を特定し、リストアップする作業の作業概要及び作業観点を以下に示す。

表 5.3.1.6 「⑥標準用語の決定」での作業概要および作業観点

項番	作業内容	ドメイン	作業概要/作業観点
1	整列した用語の中からドメイン語彙整備対象となる用語を選定するとともに、同じ概念を示している用語を特定し、リストアップする	地理空間・施設	<p>既存語彙の中でも、語彙として構造化されており、かつ最も広範な用語を含むため、IFC 4 および東京都医療機関案内サービスを標準用語の決定基準となり得る既存語彙として、標準用語を選定した。用語の選定の際の基準を優先度が高い順に以下に示す。</p> <p>①整列した用語内で共通して使われる用語 ②行政文書で使われる用語 ③標準用語の決定基準となり得る既存語彙の用語 ④既存語彙の中で使用頻度の高い用語</p>
2		移動・交通	<p>既存語彙の中でも、路線に関する語彙が定義されており概念が具体的に整理されているため、鉄道要覧を標準用語の決定基準となり得る既存語彙として、標準用語を選定した。用語の選定の際の基準を優先度が高い順に以下に示す。</p> <p>①整列した用語内で共通して使われる用語 ②標準用語の決定基準となり得る既存語彙の用語 ③既存語彙の中で使用頻度の高い用語</p>
3		財務 (自治体)	<p>既存語彙の中でも、整備済みで、かつ各自治体の予算編成時に準拠元とされている地方自治法施行規則、および総務省平成22年度市町村別決算状況調を標準用語の決定基準となり得る既存語彙として、標準用語を選定した。用語の選定の際の基準を優先度が高い順に以下に示す。</p> <p>①標準用語の決定基準となり得る既存語彙の用語 ②その他の既存語彙の中で最も使用頻度が高く、かつ2つ以上の自治体で使用された用語 ③用語の重複がない場合は、そのまま標準用語として決定</p>

項番	作業内容	ドメイン	作業概要/作業観点
4	<p>整列した用語の中からドメイン語彙整備対象となる用語を選定するとともに、同じ概念を示している用語を特定し、リストアップする</p>	<p>財務 (税どこ)</p>	<p>税どこサイトにて使用されている用語のうち、各費目の用語（例えば、民生費）が住民にとって、理解しやすい用語（例えば「福祉」）を使用しており、かつ各費目とのマッピング表も提供されているため、松阪市版、津市版税どこサイトで使用されている用語を標準用語の決定基準となり得る既存語彙として、標準用語を選定した。用語の選定の際の基準を優先度が高い順に以下に示す。</p> <p>①標準用語の決定基準となり得る既存語彙の用語 ②その他の既存語彙の中でより住民が理解しやすい用語を使用している用語 ③用語の重複がない場合は、そのまま標準用語として決定</p>
5		<p>防災</p>	<p>既存語彙の中でも、用語数及び利用が多い用語を使用していると想定される PAERA-Hazard 全国避難所施設データベース、Yahoo! Japan 避難所マップ、地震 / 津波 / 洪水 / 土砂災害 被害想定データ製品仕様書、地域防災計画（千葉県船橋市）を標準用語の決定基準となり得る既存語彙として、標準用語を選定した。用語の選定の際の基準を優先度が高い順に以下に示す。</p> <p>①整列した用語内で最も使用頻度が高く、かつ2つ以上の既存語彙に含まれる用語 ②既存語彙の中で使用頻度の高い用語</p>

(7) ⑦階層化

「(6) ⑥標準用語の決定」にてリストアップした、ドメイン語彙整備対象となる用語について、用語の関係性を考慮し、レベル合わせ（階層化）を行った際の作業概要及び作業観点を以下に示す。

表 5.3.1.7 「⑦階層化」での作業概要及び作業観点

項番	作業内容	ドメイン	作業概要/作業観点
1	リストアップしたドメイン語彙整備対象となる用語のレベル合わせを行う	地理空間・施設	用語のレベル合わせ（階層化）の際の基準を、優先度が高い順に以下に示す。
2		移動・交通	①標準用語の決定基準となり得る既存語彙の構造を参考に、標準用語のレベル合わせ（階層化）を実施。 ②同じレベルのもので共通化できる構造・概念があれば可能な限り、汎化を実施する。
3		財務（自治体）	用語のレベル合わせ（階層化）の際の基準を、優先度が高い順に以下に示す。
4		財務（税どこ）	①標準用語の決定基準となり得る既存語彙の構造を参考に、標準用語のレベル合わせ（階層化）を実施。 ②同じレベルのもので共通化できる構造・概念があれば可能な限り、汎化を実施する。 ③同じレベルにおいて各項目の金額の合計金額が、上位のレベルで必ず一致するように、各レベルの項目の下位レベルに「その他」の項目を用意する。
5		防災	用語のレベル合わせ（階層化）の際の基準を、優先度が高い順に以下に示す。 ①標準用語の決定基準となり得る既存語彙の構造を参考に、標準用語のレベル合わせ（階層化）を実施。 ②同じレベルのもので共通化できる構造・概念があれば可能な限り、汎化を実施する。

(8) ⑧用語の構造化

「(7) ⑦階層化」でレベル合わせを行ったドメイン語彙整備対象となる用語群（レベル合わせ済）に対して、用語間の関係性（親子関係や参照関係、汎化－特化関係等）を整理する際の作業概要及び作業観点を以下に示す。

表 5.3.1.8 「⑧用語の構造化」での作業概要及び作業観点

項番	作業内容	ドメイン	作業概要/作業観点
1	レベル合わせを行ったドメイン語彙	地理空間・施設	①既存語彙の中で規定されている用語の親子関係、参照関係は、そのまま活用する。 ②語源や既存語彙の用途から、各用語間の汎化-特化関係(is-a 関係)や包含関係(has-a 関係)を明確化する。 ③状態や種別等、値として特定や制限が可能な用語は、コードリストとして整備
2	整備対象となる用語群について、用語	移動・交通	
3	間の親子関係や参	財務	
4	照関係、汎化－特化関係等の関係性を整理する	防災	

(9) ⑨データタイプの決定

「(8)⑧用語の構造化」で整備した構造化済みの用語に対して、各用語のデータタイプ（文字列、数値、時刻、日時、URI、真偽値、コード等）を決定する際の作業概要及び作業観点を以下に示す。

表 5.3.1.9 「⑨データタイプの決定」での作業概要及び作業観点

項番	作業内容	ドメイン	作業概要/作業観点
1	各用語のデータタイプ(文字列、数値、時刻、日時、URI、真偽値、コード等)を決定する	地理空間・施設	用語が取りうる値の特性に従い、データタイプ(文字列、数値、時刻、日時、URI、真偽値、コード等)を明確化する。
2		移動・交通	
3		財務	
4		防災	

(10) ⑩他の語彙との調整

「(9)⑨データタイプの決定」で整備した検証前ドメイン語彙サンプルに対して、コア語彙および他ドメインで整備済の語彙との重複を避けるため、コア語彙および他ドメインで整備済の語彙を適用する際の作業概要及び作業観点を以下に示す。

表 5.3.1.10 「⑩他の語彙との調整」での作業概要及び作業観点

項番	作業内容	ドメイン	作業概要/作業観点
1	整備した検証前ドメイン語彙サンプルに対して、コア語彙および他ドメインで整備済の語彙を適用する	地理空間・施設 移動・交通 財務 防災	<p>①「(9)⑨データタイプの決定」で整備した検証前ドメイン語彙サンプルの中に、コア語彙もしくは他ドメインで整備済みの語彙と同義の用語または、包含される概念を含む用語が存在する場合、コア語彙もしくは他ドメインで整備済みの語彙を用いて検証前ドメイン語彙サンプルの用語を表現し、検証前ドメイン語彙サンプル（他の語彙との調整済み）として、整備する。</p> <p>②コア語彙もしくは他ドメイン語彙と概念が大きく異なる場合、当該ドメインの検証前ドメイン語彙サンプルとして、そのまま整備する。</p>

(11) ⑪フィールド適用による検証

財務ドメインにおいてはデータ構造が単純であり、実データを検証前のドメイン語彙サンプルで記述可能であることは明白であるため、本作業は残る3ドメインを対象に実施した。

表 5.3.1.11 「⑪フィールド適用による検証」でのテンプレート作成作業

項番	作業内容	ドメイン	作業概要/作業観点
1	検証前ドメイン語彙サンプル（他の語彙との調整済み）に実データを適用し、十分に記述可能であるかを検証する	地理空間・施設	検証前ドメイン語彙サンプル（他の語彙との調整済み）に対して、東京都の某病院の施設情報に関するデータを適用し、十分に記述可能であることを検証した。
2		移動・交通	検証前ドメイン語彙サンプル（他の語彙との調整済み）に対して、JR 東日本の山手線および JR 西日本の東海道本線、JR 京都線、JR 神戸線の路線に関するデータを適用し、路線に関する実データについては、十分に記述可能であることを検証した。
3		財務（自治体）	検証前ドメイン語彙サンプル（他の語彙との調整済み）に対して、平成 25 年度の岐阜県高山市、千葉県八千代市、青森県青森市の一般会計予算に関するデータを適用し、十分に記述可能であることを検証した。
4		財務（税どこ）	税どこ用の検証前ドメイン語彙サンプル（他の語彙との調整済み）に対して、平成 25 年度の岐阜県高山市、千葉県八千代市、青森県青森市の一般会計予算に関するデータを適用し、十分に記述可能であることを検証した。
5		防災	検証前ドメイン語彙サンプル（他の語彙との調整済み）に対して、千葉県某立某中学校の施設情報および避難所情報に関するデータを適用し、十分に記述可能であることを検証した。

(12) ⑫フィールド適用結果のフィードバック

「(11)⑪フィールド適用による検証」の際に、一部記述できない実データが存在した場合、当該用語を記述可能とするために、ドメイン語彙の追加または修正を実施し、ドメイン語彙の記述可能性を向上させることを目的とした作業である。

ただし、本事業にて収集・整備したドメイン語彙サンプルにおいては、当該事象は発生しなかった。

以下に、フィールド検証結果のフィードバックの作業概要及び作業観点を示す。

表 5.3.1.12 「⑫フィールド適用結果のフィードバック」の作業概要及び作業観点

項番	作業内容	ドメイン	作業概要/作業観点
1	検証結果に基づいて ドメイン語彙の 追加・修正を実施	地理空間・施設	フィールド適用の結果、一部記述できない実データが存在した場合、当該用語を記述可能とするために、ドメイン語彙の追加または修正を実施し、ドメイン語彙の記述可能性を向上させる。 ただし、本事業にて収集・整備したドメイン語彙サンプルにおいては、当該事象は発生しなかった。
2		移動・交通	
3		財務（自治体）	
4		財務（税どこ）	
5		防災	

5.3.2 ドメイン語彙の収集・整備プロセスを見直す必要があった事例とその対応方法

ドメイン語彙サンプルの収集・整備を実施する中で、ドメイン語彙の収集・整備プロセスを見直す必要があった事例と、その対応方法について以下にまとめる。

(1) ひとつの用語で複数の意味（概念）をもつ用語への対応

1) 事例

移動・交通ドメインにおける既存語彙の中では、路線や駅等の用語は、その意味が用語の利用者や利用シーンにより様々であり、多義語となっている。そのため、単純に個々の用語を比較するだけでは、ドメイン語彙サンプルの整備が困難であった。

2) 対応方法

既存語彙及び各種文献等に基づいて、各用語の具体的な使われ方を調査し、それぞれの概念の整理を行った上で、用語の収集・整備を実施した。また、ドメイン語彙の収集・整備プロセスに実例を基にした概念整理のプロセスを組み込むこととした。

① 「路線」に関する概念の整理

「路線」に関する概念を整理するため、実際の路線の状況（実例）を基に概念を整理した。また、整理した概念については、鉄道要覧を基に、用語の妥当性について検証を実施した。その結果、「路線」の概念は以下の3概念が存在することが確認できた。

表 5.3.2.1 「路線」の概念の整理結果

項番	概念	概念の説明
1	開業線 (例：赤羽線、 東北線（東京-盛岡）、 山手線（品川-田端）)	国土交通省に事業基本計画として提出済の路線。路線の戸籍に相当する。JR では、運賃計算の基になる距離の算出に用いる他、臨時列車や特急列車等、下記の運行路線として定義されていない特別なルートを走行する列車を表現するのに必要な概念。開業線は東北線東京～盛岡区間と、日暮里～赤羽区間の2区間のように、複数区間をもつ場合がある。(図5.3.2.1 参照)
2	運行路線 (例：京浜東北線、埼京線、 山手線（大崎-大崎）)	路線図や時刻表、駅の案内などで呼称される路線。開業線の組合せにより表現することができる。
3	路線の愛称 (例：宇都宮線、JR 京都線、 東武スカイツリーライン)	運行路線の特定区間において、別名（路線の愛称）を称したもの。複数の運行路線を跨る場合がある。旅客案内上では、運行路線と路線の愛称のどちらも使用する場合がある。

以下に、「路線」の概念のイメージを示す。

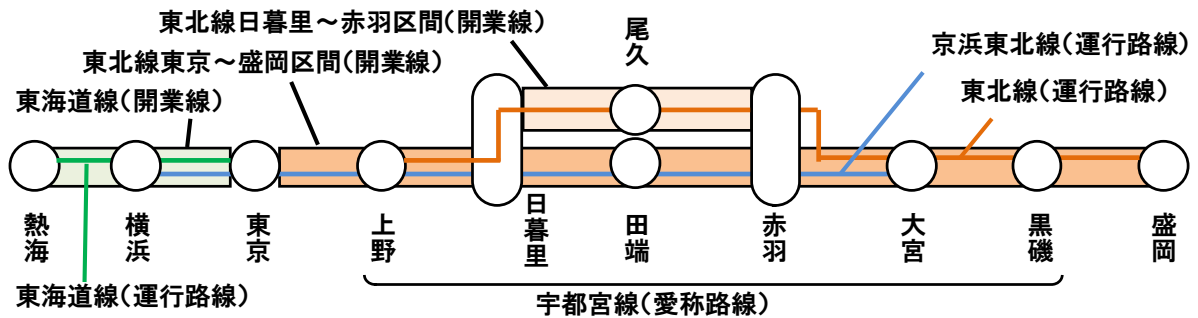


図 5.3.2.1 「路線」の概念のイメージ

②「列車」の運行パターンに関する概念の整理

「列車」に関する概念を整理するため、実際の路線の状況（実例）を基に概念を整理した結果、「列車」の運行パターンについての概念として、以下の2概念が存在することが確認できた。

表 5.3.2.2 「列車」の運行パターンの概念の整理結果

項番	概念	概念の説明
1	運行系統	運行種別（普通、急行など）ごとに定義された、定期的に運行される列車のパターン。複数の路線、他社路線を跨いだ場合や、特急列車や臨時列車の運行系統をあらわすことが出来ない。
2	直通運行での運行系統	複数の路線、他社路線を跨いだ場合や、特急列車や臨時列車の運行系統をあらわすため、運行系統を組み合わせる運行系統。 下記の例では、小田急小田原線の運行系統と江ノ島線の運行系統を組み合わせることで、直通運行での運行系統をあらわすことができる。

以下に、「列車」運用パターンの概念のイメージを示す。

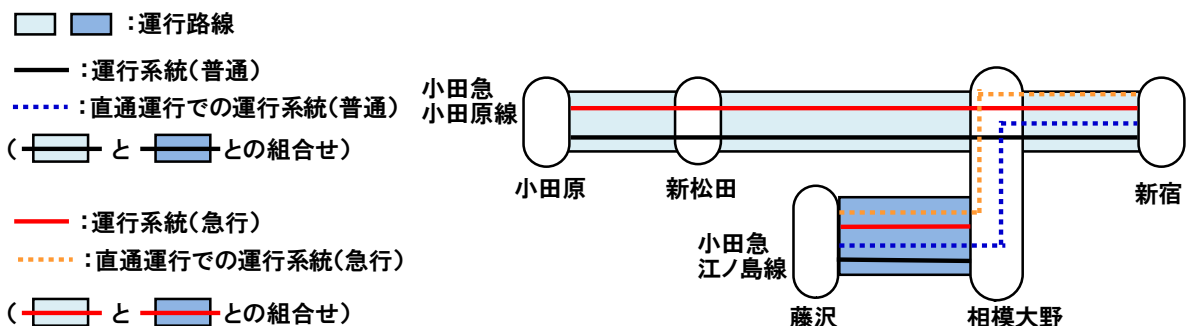


図 5.3.2.2 「列車」の運行パターンの概念のイメージ

③ 「列車」に関する概念の整理

「列車」に関する概念を整理するため、実際の路線の状況（実例）を基に概念を整理した結果、物理的な「列車」についての概念として、以下の2概念が存在することが確認できた。

表 5.3.2.3 「列車」の概念の整理結果

項番	概念	概念の説明
1	列車	路線上を走行する個々の鉄道の便。運行路線ごとに、列車を識別できるように列車番号が割り振られている。ただし、下記の例のように、列車番号は各運行路線に一对一で対応しない場合が存在する。
2	編成単位での列車	各運行路線に一对一で対応しない場合では、列車番号によらず、編成が始発駅から終着駅まで運行する鉄道の便を示した概念として、編成単位での列車の概念が存在する。下記の例では、列車番号 3712 の列車と列車番号 1141S の列車の組合せで、編成単位での列車が表現される。

以下に、「列車」の概念のイメージを示す。

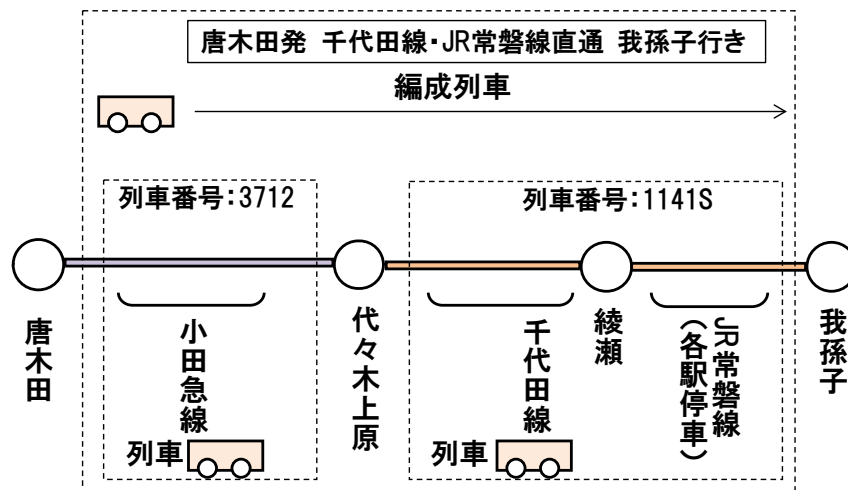


図 5.3.2.3 「列車」の概念のイメージ

④「列車」に関する概念の整理

「駅」に関する概念を整理するため、実際の路線の状況（実例）を基に概念を整理した結果、以下の2概念が存在することが確認できた。

表 5.3.2.4 「駅」の概念の整理結果

項番	概念	概念の説明
1	施設としての駅	鉄道事業者が管理する施設として存在する駅。各路線の列車は、路線ごとに用意されたホームに乗り入れている。
2	運行路線上の駅	各運行路線上に存在する駅。相互直通運転の境界駅のように、複数の路線が一つのホームに乗り入れている場合も、路線ごとに別々の駅として定義されている。

以下に、「駅」の概念のイメージを示す。

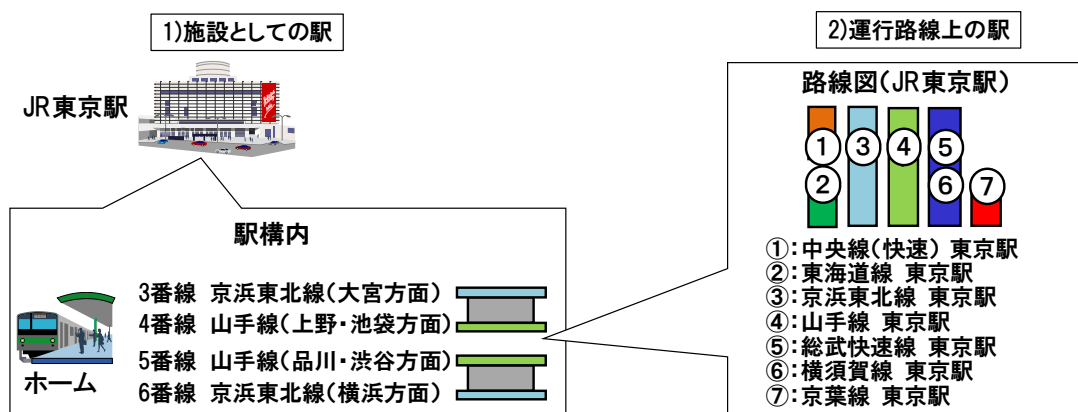


図 5.3.2.4 「駅」の概念のイメージ

3) ドメイン語彙の収集・整備プロセスの見直し

ドメイン語彙の収集・整備プロセスの仮説に対して、既存語彙及び各種文献等に基づいて、各用語の具体的な使われ方を調査し、それぞれの概念の整理を行った上で、用語の収集・整備を実施するプロセスである「実例を基にした概念整理」のプロセスを組み込んだ見直し後のドメイン語彙の収集・整備プロセス（移動・交通ドメイン）を整備した。

以下に、見直し後ドメイン語彙の収集・整備プロセス（移動・交通ドメイン）を示す。

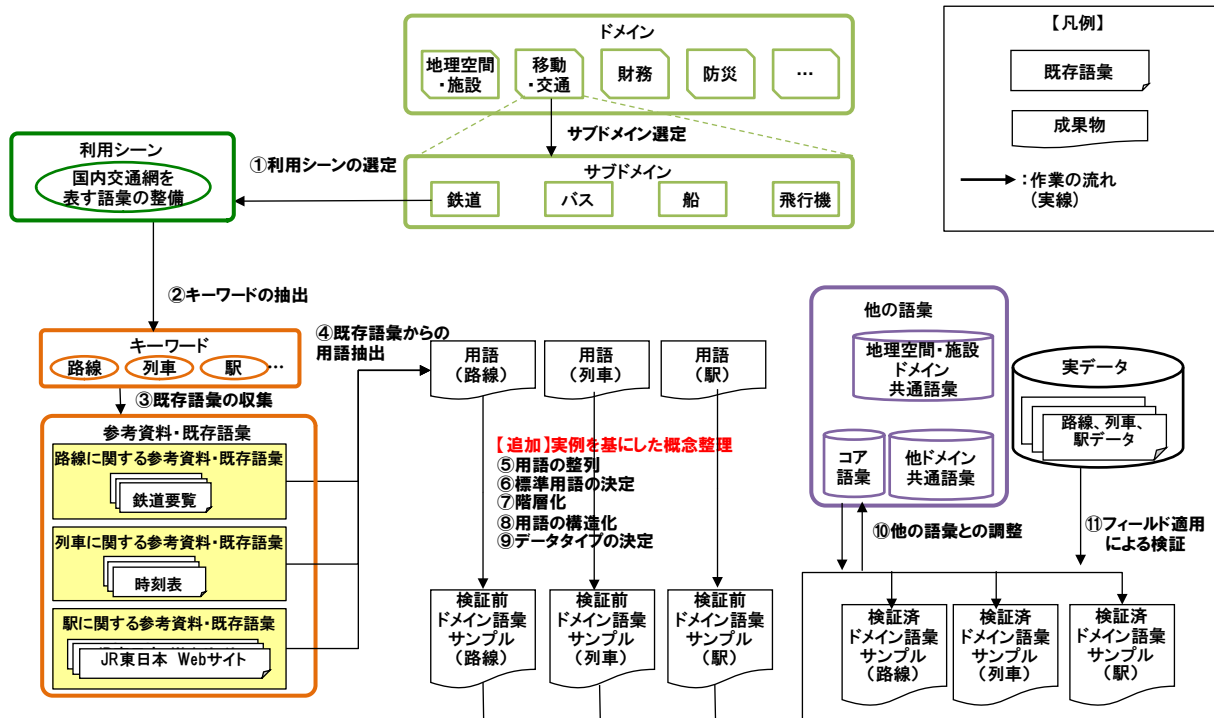


図 5.3.2.5 見直し後 ドメイン語彙の収集・整備プロセス（移動・交通ドメイン）

(2) 理解しやすい用語を使用したオープンデータへの対応

1) 事例

財務ドメインにおいて採用した利用シーン「市区町村の予算情報を住民にとって理解しやすい表現で统一的に公開し、税金使途の透明化・自治体間の横比較を可能とする。」に関連する取組みとして、「税どこ」がよく知られている。「税どこ」は主に有志によるコミュニティ活動を通じて設置されており、2014年4月1日現在で131自治体のサイトが立ち上がっている。

しかしながら、「税どこ」で現在公開されているサイトのうち、一部の自治体では、住民に理解しやすい用語を使用しているが、大半の自治体は、一般会計歳出予算における分類（行政目的の分類）で使用されている用語をそのまま使用しているため、住民にとっては税金の使途が理解しにくい表現となっている可能性がある。

そこで、市区町村の予算情報を住民にとって理解しやすい统一的な表現にすることと同時に、税金使途の透明化・自治体間の横比較を可能とするために、市区町村予算費目標準語彙と税どこ標準カテゴリ語彙の関係性を明確化する必要があった。

以下に、市区町村予算費目標準語彙と税どこ標準カテゴリ語彙の関係性のイメージを示す。

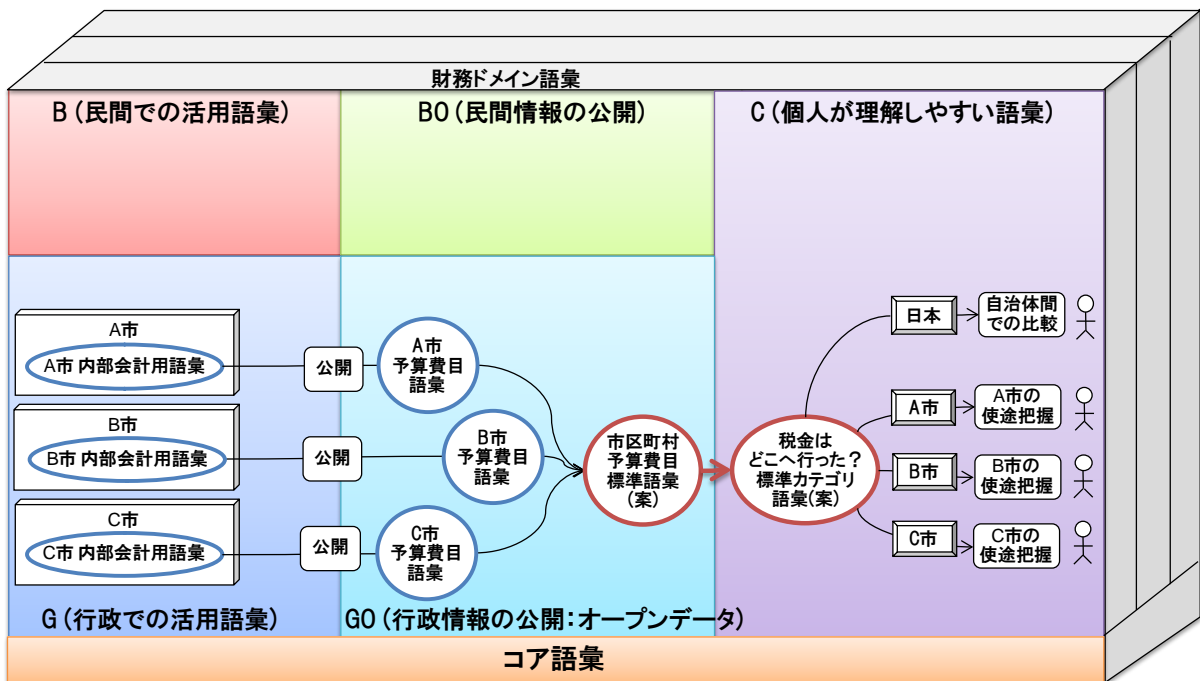


図 5.3.2.6 市区町村予算費目標準語彙と税どこ標準カテゴリ語彙の関係性のイメージ

2) 対応方法

市区町村予算費目標準語彙と税どこ標準カテゴリ語彙の関係性を明確化するため、三重県松坂市や、三重県津市で整備され、公開されてくる市区町村予算費目と税どこカテゴリとのマッピング表を参考に、市区町村予算費目標準語彙と税どこ標準カテゴリ語彙とのマッピング表を整備することとした。

本マッピング表では、「民生費」「公債費」などの市区町村予算費目の用語に対して、「福祉」「借金返済」などの住民にとって理解しやすい統一的な表現をマッピングしている。

また、分類も同様に、福祉関係を表す「民生費」の内訳として「社会福祉費」、「老人福祉費」、「児童福祉費」、「生活保護費」、「災害救助費」などの市区町村予算費目の用語での分類に対して、税どこ標準カテゴリ語彙とのマッピング表を整備した。

さらに、ドメイン語彙の収集・整備プロセスにマッピング表の作成のプロセスを組み込むこととした。

3) ドメイン語彙の収集・整備プロセスの見直し

ドメイン語彙の収集・整備プロセスの仮説に対して、市区町村予算費目標準語彙と税どこ標準カテゴリ語彙の関係性を明確化するため、市区町村予算費目標準語彙と税どこ標準カテゴリ語彙とのマッピング表を整備するプロセスである「マッピング表の整備」のプロセスを組み込んだ見直し後のドメイン語彙の収集・整備プロセス（財務ドメイン）を整備した。

以下に、見直し後のドメイン語彙の収集・整備プロセス（財務ドメイン）を示す。

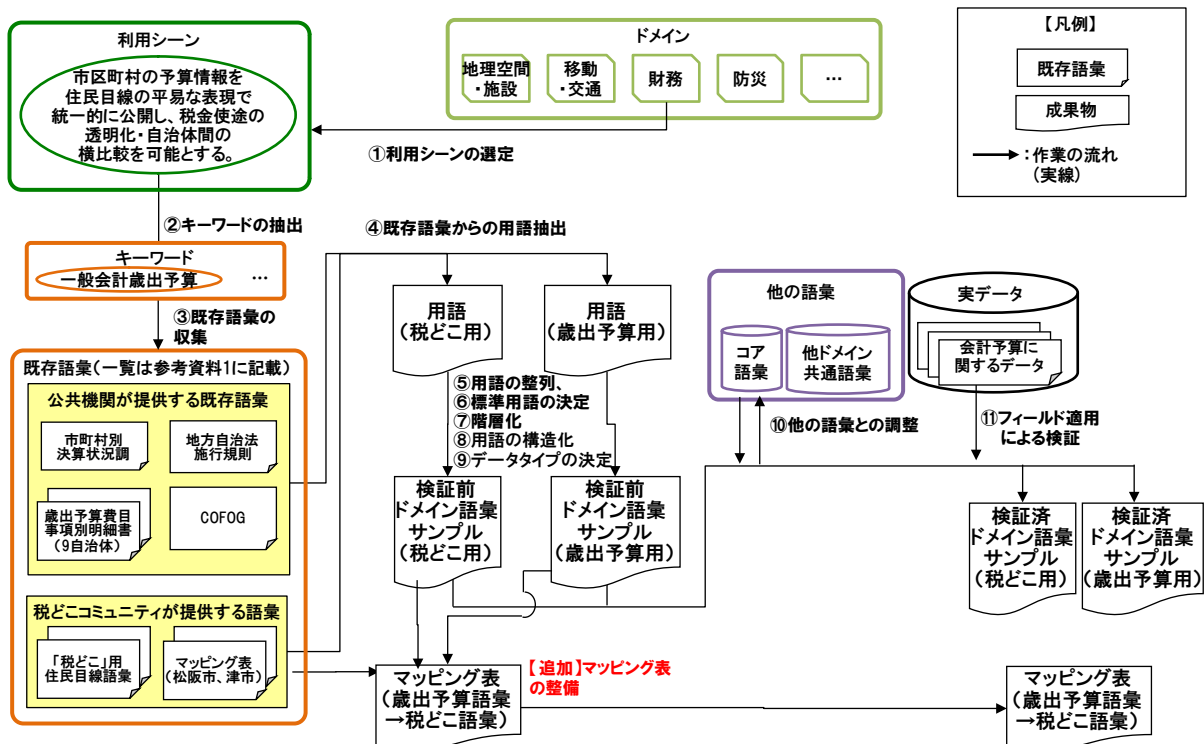


図 5.3.2.7 見直し後 ドメイン語彙の収集・整備プロセス（財務ドメイン）

(3) 視点（どの主体がどの客体に対して使用する用語か）の明確化への対応

1) 事例

防災ドメインにおいては、医療関係者、自衛隊、消防など、非常に高い専門性を備えた主体や客体が、各々の業務に特化し、かつ洗練された語彙を利用している。

例えば、道路管理者が道路を修復し、物理的に「啓開済」になった場合でも、緊急車両／一般車両の区分で通行可能／通行不能に分かれる場合がある。災害対策本部が、道路の復旧状態を道路管理者から受信した場合においても、「道路啓開済」が避難者にとって通行可能／通行不能のどちらであるか確認・判断した上で通知しなければならない。

また、災害対応活動を迅速に行うためには、医療関係者、自衛隊、消防などの様々な主体と客体とが相互に連携を図ることが求められ、その際、業務の視点や、利用する用語やその意味が、情報伝達する主体と客体間で異なる場合においても、適切な情報連携・意思疎通が図れる必要がある。

以下に、防災ドメインにおける複数の主体が存在するイメージを示す。

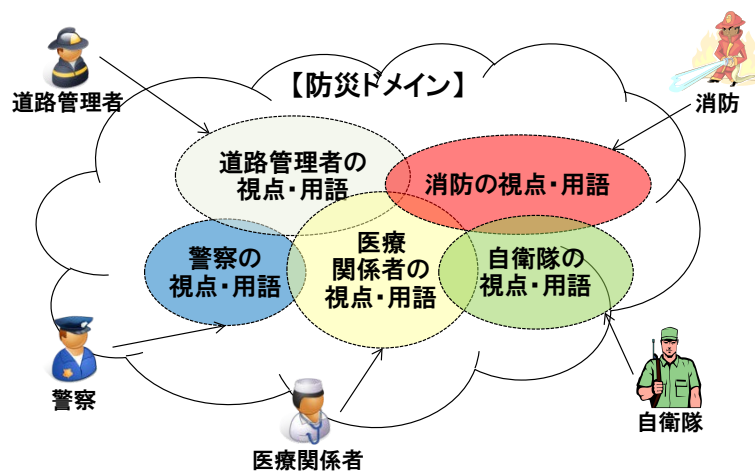


図 5.3.2.8 防災ドメインにおける複数の主体が存在するイメージ

2) 対応方法

共通語彙（各主体と客体間で共通理解が可能な概念や用語）を整備することが望ましいが、防災ドメインにおいては、医療関係者、自衛隊、消防など、非常に高い専門性を備えた主体や客体が、各々の業務に特化し、かつ洗練された語彙を利用しているため、共通語彙を定める事が困難な場合も存在する。

そこで、視点（どの主体がどの客体に対して使用する用語か）を明確化した上で、既存語彙を収集すると共に、各主体-客体間で使用されている標準用語を対応付けた「マッピングルール」を整備することで、各用語間の意味の違いを吸収する仕組みを導入すると共に、ドメイン語彙の収集・整備プロセスに組み込んだ。

また、防災ドメイン等、高い専門性を備えた主体や客体より構成されているドメインにおいては、業務の視点や、利用する用語やその意味を整備する際、情報伝達する主体と客体間で利用する用語やその意味を共有しながら選定することで、より一層適切な情報連携・意思疎通が図れる。

そのため、標準用語の選定時に防災ドメイン有識者・関係者による討論のプロセスをドメイン語彙の収集・整備プロセスに組み込んだ。

3) ドメイン語彙の収集・整備プロセスの見直し

ドメイン語彙の収集・整備プロセスの仮説に対して、視点（どの主体がどの客体に対して使用する用語か）の明確化のプロセス、各主体-客体間で使用されている標準用語を対応付けた「マッピングルール」を整備プロセス、標準用語の選定時に防災ドメイン有識者・関係者による討論のプロセスを組み込んだ見直し後のドメイン語彙の収集・整備プロセス（防災ドメイン）を整備した。

以下に、見直し後のドメイン語彙の収集・整備プロセス（防災ドメイン）を示す。

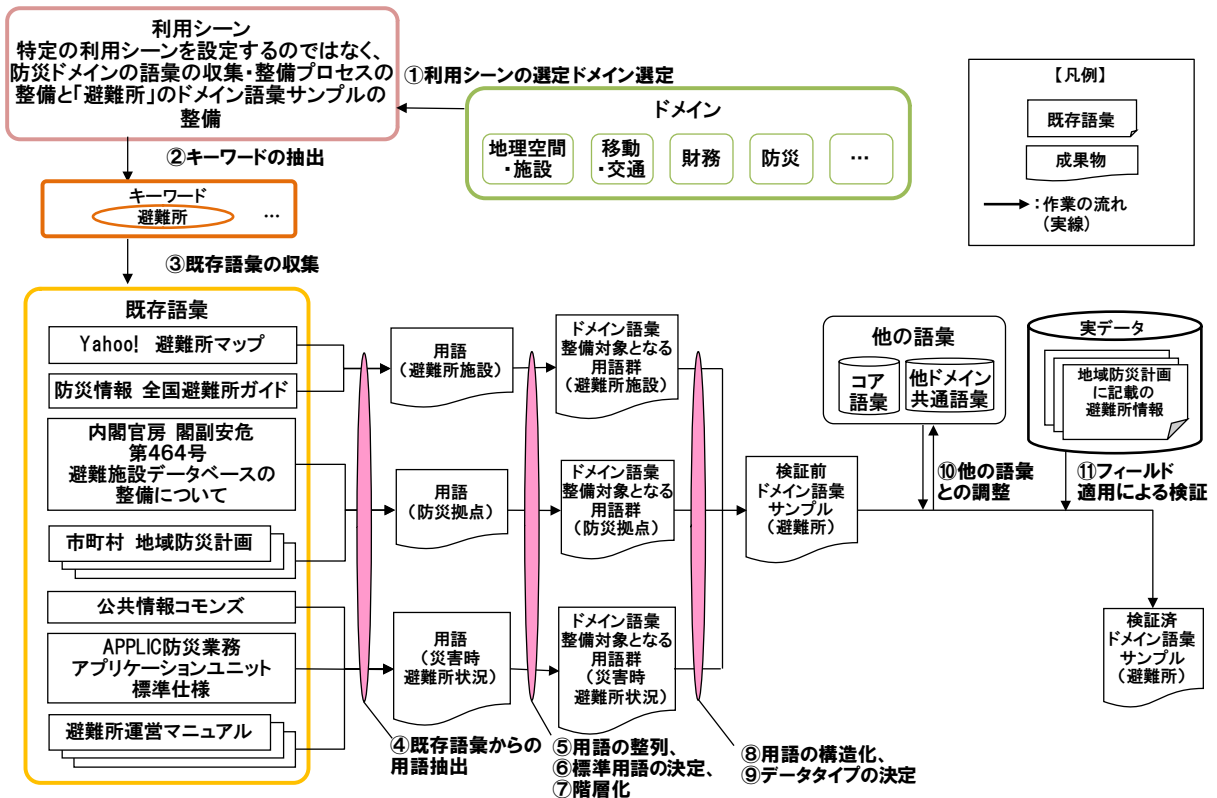


図 5.3.2.9 見直し後 ドメイン語彙の収集・整備プロセス（防災ドメイン）

5.3.3 本事業で整備したドメイン語彙サンプル

本事業では、ドメイン共通語彙やドメイン固有語彙を継続的に整備していくための施策として、地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインの4ドメインを試行的に選定した。

対象分野の専門家を交えた十分な検討の場として、ドメイン語彙検討会を通じて、4章で検討したデータ構造に従い、ドメイン共通語彙としての用語をリストアップし、ドメイン語彙サンプルとして整備するとともに、その収集・整備プロセスについても構築した。

本事業で整備した4ドメインそれぞれにおけるドメイン語彙サンプルを以下に示す。

(1) 地理空間・施設ドメイン

地理空間・施設ドメインでは、見直し後 ドメイン語彙の収集・整備プロセスやドメイン語彙検討会を通じて、建物の構造に関する用語（階、区域など）をコア語彙の建物型を拡張し整備した。また、支払い方法など、サービスの提供時に使用する用語をコア語彙の施設型を拡張し整備した。さらに、病院および医療機関型については、サービスを提供する施設として整備するため、gf:施設型を継承し整備した。

以下に、本事業にて収集・整備したドメイン語彙サンプルについて、クラス図を一部抜粋したものを以下に例示する。

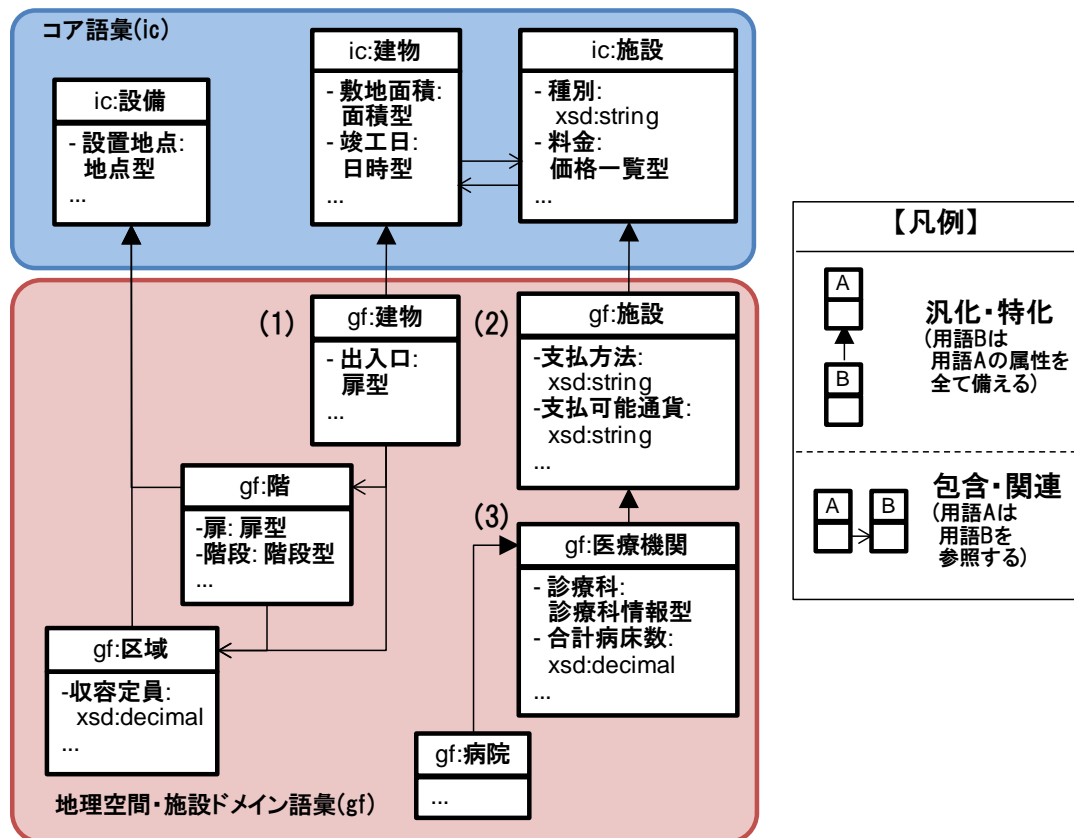


図 5.3.3.1 地理空間・施設ドメインの語彙サンプルのクラス図

以下に、地理空間・施設ドメインのドメイン語彙サンプルに含まれる主要な用語の一覧を示す。

表 5.3.3.1 地理空間・施設ドメインのドメイン語彙サンプルに含まれる主要な用語（構造をもった用語）

項番	分類	項目説明
1	場所	場所を表現するためのデータ型。
2	座標列	線（ライン）や領域（ポリゴン）を構成する座標列を表現するためのデータ型。
3	アクセス	地点等へのアクセスを表現するためのデータ型。
4	道路	道路を表現するためのデータ型。
5	道路区間	道路区間(道路の一部)を表現するためのデータ型。
6	交通規制状態	道路の交通規制状態を表現するためのデータ型。
7	建物	建物を表現するためのデータ型。
8	階	建物内の各階の詳細情報を表現するためのデータ型。
9	区域	部屋(壁のある区域)や空間(壁のない区域)を表現するためのデータ型。
10	扉	扉を表現するためのデータ型。
11	真偽不明値	真偽に加え、不明を表す値を表現するためのデータ型。
12	階段	階段を表現するためのデータ型。
13	移動設備	移動設備(エレベータやエスカレータなど)を表現するためのデータ型。
14	エレベータ	エレベータを表現するためのデータ型。
15	駐車場	駐車場を表現するためのデータ型。
16	便所	便所を表現するためのデータ型。
17	病室	病室を表現するためのデータ型。
18	施設	施設の詳細情報を表現するためのデータ型。
19	医療機関	医療機関の施設としての情報を表現するためのデータ型。
20	病院	病院の施設としての情報を表現するためのデータ型。
21	病床	病床を表現するためのデータ型。
22	医療機関従業員	医療機関の従業員情報を表現するためのデータ型。
23	医療機関内サービス	医療機関内で提供されるサービスを表現するためのデータ型。
24	診療科情報	医療機関に設置された診療科の一覧。
25	宿泊施設	宿泊施設の施設としての情報を表現するためのデータ型。
26	宿泊施設内サービス	宿泊施設内で提供されるサービスを表現するためのデータ型。

(2) 移動・交通ドメイン

移動・交通ドメインでは、見直し後 ドメイン語彙の収集・整備プロセスやドメイン語彙検討会を通じて、駅については、施設としての駅を ic:施設型の拡張として定義し、乗入れ路線や駅構内情報をプロパティ用語として整備した。

区間については路線、経路および地理空間・施設ドメインで整備済の道路に共通に含まれる要素を抽出し、上位概念として整備した。

また、路線や経路は区間型の拡張として追加整備した。

さらに、列車のプロパティ用語である旅客列車編成およびその構成要素である鉄道車両は、人が移動・滞在する空間であるため、地理空間・施設ドメインの区域型の拡張として整備した。

以下に、本事業にて収集・整備したドメイン語彙サンプルについて、クラス図を一部抜粋したものを以下に例示する。

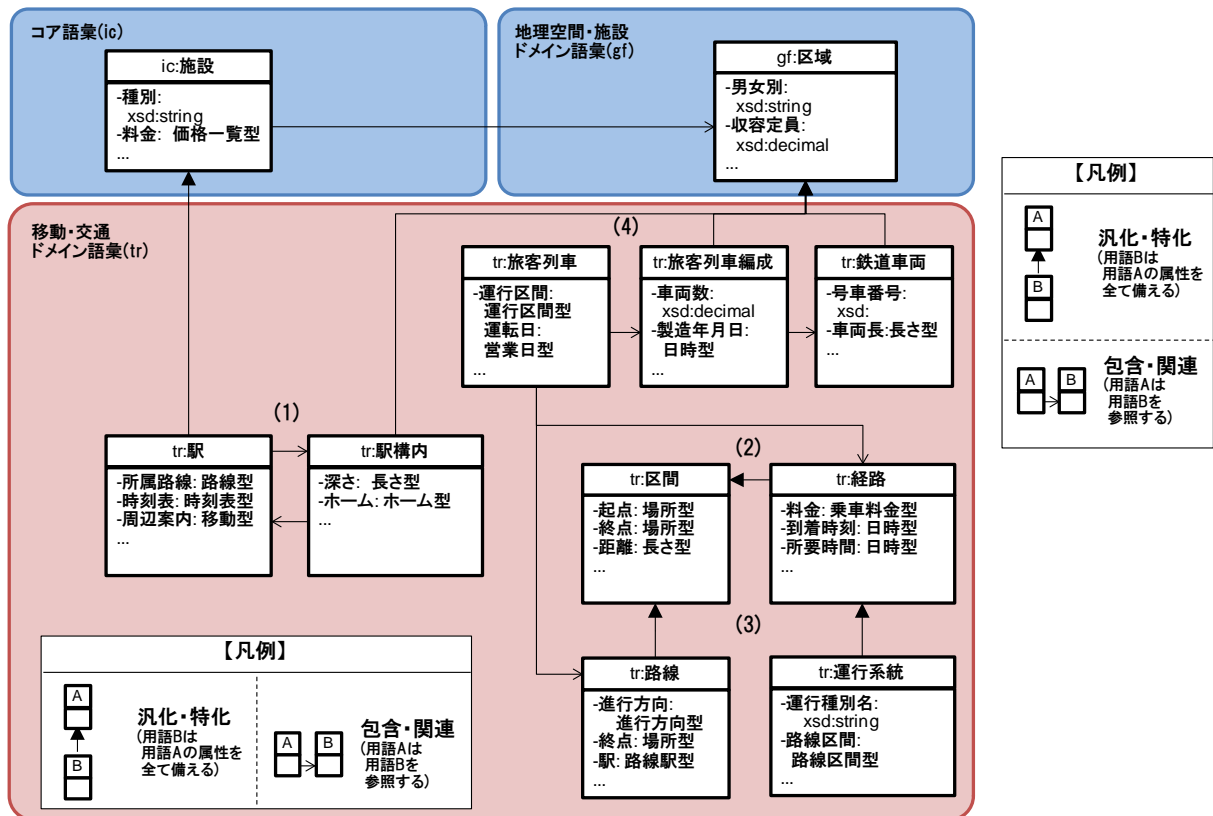


図 5.3.3.2 移動・交通ドメインの語彙サンプルのクラス図

以下に、移動・交通ドメインのドメイン語彙サンプルに含まれる主要な用語の一覧を示す。

表 5.3.3.2 移動・交通ドメインのドメイン語彙サンプルに含まれる主要な用語（構造をもった用語）

項番	分類	項目説明
1	駅	列車への乗車や買い物、待ち合わせ等、施設として利用する駅を表現するためのデータ型。
2	駅構内	駅の構内を表現するためのデータ型。
3	改札口	駅構内に存在する改札口を表現するためのデータ型。
4	ホーム	駅構内に存在するホームを表現するためのデータ型。
5	駅構内施設	駅構内に存在する ATM、売店、駅事務室などの施設を表現するためのデータ型。
6	ATM	駅構内に存在する ATM を表現するためのデータ型。
7	定期券うりば	駅構内に存在する定期券うりばを表現するためのデータ型。
8	駅構内設備	駅構内に設置された設備を表現するためのデータ型。
9	券売機	駅構内に設置された券売機を表現するためのデータ型。
10	移動設備	駅構内に設置された移動設備を表現するためのデータ型。
11	エスカレータ	駅構内に設置されたエスカレータを表現するためのデータ型。
12	エレベータ	駅構内に設置されたエレベータを表現するためのデータ型。
13	階段	駅構内に設置された階段を表現するためのデータ型。
14	バリアフリー情報	駅構内のバリアフリー情報を表現するためのデータ型。
15	便所	駅構内に存在する便所を表現するためのデータ型。
16	授乳室	駅構内に存在する授乳室を表現するためのデータ型。
17	駅バリアフリー設備	駅構内に設置されたバリアフリー設備を表現するためのデータ型。
18	駅出入口	駅構内の出入口を表現するためのデータ型。
19	路線駅	鉄道路線を構成する駅を表現するためのデータ型。
20	停車駅	鉄道路線における列車の停車駅を表現するためのデータ型。
21	乗り換え	鉄道路線における駅の乗換え案内を表現するためのデータ型。
22	移動	目的地までの移動時間および移動距離を表現するためのデータ型。
23	時刻表	鉄道路線における駅の時刻表を表現するためのデータ型。
24	列車発着情報	鉄道路線における駅へ乗り入れる列車の発着情報を表現するためのデータ型。
25	営業日	施設の営業日や列車の運行日を表現するためのデータ型。
26	進行方向	路線や駅を走行する列車の進行方向を表現するためのデータ型。

項番	分類	項目説明
27	旅客列車	旅客用の列車を表現するためのデータ型。
28	区間	区間を表現するためのデータ型。
29	路線	路線を表現するためのデータ型。
30	経路	列車の運行ルートや乗換えルートなどの経路を表現するためのデータ型。
31	乗車料金	運賃や特急料金などを含む、列車の乗車料金を表現するためのデータ型。
32	運行経路	列車の運行経路を表現するためのデータ型。
33	運行路線	列車が貨物や旅客用に運行する路線を表現するためのデータ型。
34	愛称路線	運行路線の一区間に対して愛称を付与した路線を表現するためのデータ型。
35	開業線	路線の戸籍に相当し、国土交通省に事業基本計画として提出を行い、線路を敷設した営業路線を表現するためのデータ型。
36	運行系統	運行路線や列車種別ごとに異なる列車の運行ルート・運行パターンを表現するためのデータ型。
37	直通運行系統	複数の運行路線に跨った直通運転を、運行路線の組合せにより実現する運行系統を表現するためのデータ型。
38	列車運行形態	列車の編成（車両数）や種別（特急・急行など）といった列車の運行形態を表現するためのデータ型。
39	公共交通機関	電車やバス、飛行機、船舶など旅客用に不特定多数の人の輸送を行う公共交通機関を表現するためのデータ型。
40	旅客列車編成	旅客用の列車として走行する列車の編成を表現するためのデータ型。
41	鉄道車両	列車編成を構成する個々の鉄道車両を表現するためのデータ型。
42	貨車	貨物輸送用の鉄道車両を表現するためのデータ型。
43	旅客車	普通車、食堂車、寝台車など旅客輸送用の鉄道車両を表現するためのデータ型。
44	優等列車用旅客車	新幹線や特急列車などの優等列車を構成する旅客車を表現するためのデータ型。
45	優等列車設備	新幹線や特急列車などの優等列車における旅客車の設備を表現するためのデータ型。
46	優等列車座席	新幹線や特急列車などの優等列車における旅客車の座席を表現するためのデータ型。
47	普通列車用旅客車	普通列車における旅客車を表現するためのデータ型。
48	普通列車設備	普通列車における旅客車の設備を表現するためのデータ型。

(3) 財務ドメイン

財務ドメインでは、見直し後 ドメイン語彙の収集・整備プロセスやドメイン語彙検討会を通じて、一般会計歳出予算の各費目の標準語彙、および「税どこ」の各項目の標準語彙をそれぞれ整備した。

また、既存語彙において階層構造の親子関係として定義されている項目を、クラスとそのクラスが備えるプロパティの関係として整備した。

歳出予算の各費目の上位概念として費目型を用意し、費目の名称や金額をプロパティとして整備した。

以下に、本事業にて収集・整備したドメイン語彙サンプルについて、クラス図を一部抜粋したものを以下に例示する。

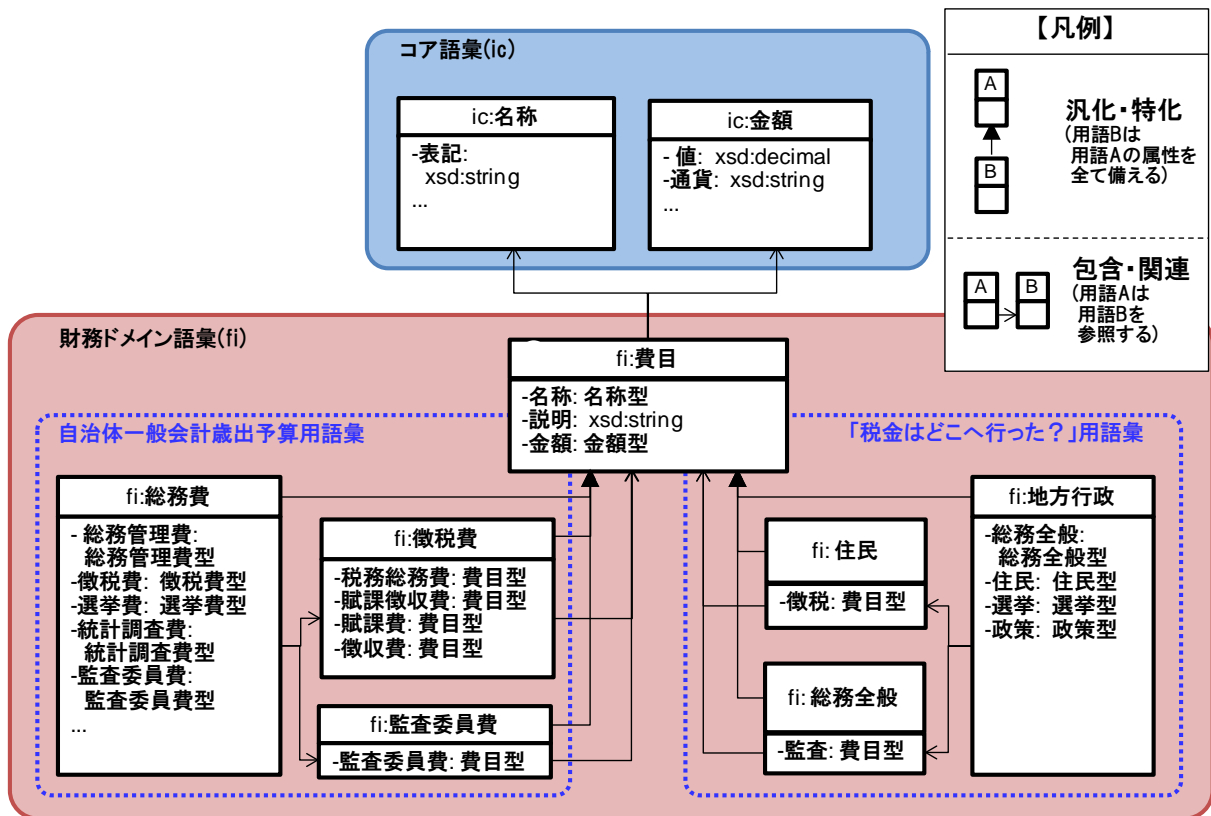


図 5.3.3.3 財務ドメインの語彙サンプルのクラス図

以下に、財務ドメインのドメイン語彙サンプルに含まれる主要な用語の一覧を示す。

表 5.3.3.3 財務ドメインのドメイン語彙サンプル（自治体一般会計歳出予算用）に含まれる主要な用語
（構造をもった用語）

項番	分類	項目説明
1	費目	自治体の一般会計予算の費目を表現するためのデータ型。
2	議会費	主に事務経費、議員報酬など、議会の活動に要する経費を表現するためのデータ型。
3	総務費	市政運営のための一般管理費に位置づけられる費目を表現するためのデータ型。
4	民生費	社会福祉の充実を図るため、児童、高齢者、心身障害者等のための福祉施設の整備及び運営、生活保護の実施等の施策の推進に要する経費を表現するためのデータ型。
5	衛生費	住民の健康を保持増進し、生活環境の改善を図るため、医療、公衆衛生、精神衛生等に係る対策を推進するとともに、し尿・ごみなど一般廃棄物の収集・処理等、住民の日常生活に密着した諸施策の推進に要する経費を表現するためのデータ型。
6	労働費	就業者の福祉向上を図るため、職業能力開発の充実、金融対策、失業対策等の施策の推進に要する経費を表現するためのデータ型。
7	農林水産業費	農林水産業の振興と食糧の安定的供給を図るため、生産基盤の整備、構造改善、消費流通対策、農林水産業に係る技術の開発・普及等の施策の推進に要する経費を表現するためのデータ型。
8	商工費	地域における商工業の振興とその経営の近代化等を図るため、中小企業の指導育成、企業誘致、消費流通対策等さまざまな施策の推進に要する経費を表現するためのデータ型。
9	土木費	土木管理関係人件費、土木共通事務等に要する経費並びに道路橋梁費、河川費、砂防費、港湾費、都市計画費、住宅建設費、災害土木関連費等の費目を表現するためのデータ型。
10	消防費	火災、風水害、地震等の災害から国民の生命、身体及び財産を守り、これらの災害を防除し、被害を軽減するため、消防行政の推進に要する経費を表現するためのデータ型。
11	教育費	教育施策の推進に要する経費を表現するためのデータ型。
12	災害復旧費	災害によって生じた被害の復旧に要する経費を表現するためのデータ型。
13	公債費	元利償還等に要する経費や地方債の発行手数料、割引料等の事務経費など、地方公共団体が発行した地方債の元利償還等に要する経費を表現するためのデータ型。
14	予備費	予算外の支出又は予算超過の支出に充てるための費目を表現するためのデータ型。
15	諸支出金	行政目的を有しない経費のみを計上する計上科目を表現するためのデータ型。

表 5.3.3.4 財務ドメインのドメイン語彙サンプル（「税どこ」）に含まれる主要な用語
（構造をもった用語）

項番	分類	項目説明
1	費目	自治体の一般会計予算の費目を住民にとって理解しやすい形で表現するためのデータ型。
2	地方議会	地方議会に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
3	地方行政	地方行政に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
4	健康福祉	健康福祉に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
5	環境	環境に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
6	経済・観光	経済・観光に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
7	都市開発	都市開発に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
8	消防・防災・安全	消防・防災・安全に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
9	子育て・教育	子育て・教育に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
10	文化・スポーツ	文化・スポーツに関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
11	借金返済	地方自治体の借金返済に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。
12	その他	その他の項目に関して使われた税金の額を住民にとって理解しやすい形で表現するためのデータ型。

(4) 防災ドメイン

防災ドメインでは、見直し後 ドメイン語彙の収集・整備プロセスやドメイン語彙検討会を通じて、コア語彙の施設型をベースとして、防災ドメインで登場する施設の語彙を整備した。また、コア語彙の施設型を拡張し、一時避難場所や避難所など様々な災害時に登場する施設を表現する災害対応施設型を整備した。さらに、避難所は収容可能人数をはじめ独自の項目が必要となるため、災害対応施設型を拡張して避難所型を新たに整備した。

以下に、本事業にて収集・整備したドメイン語彙サンプルについて、クラス図を一部抜粋したものを以下に例示する。

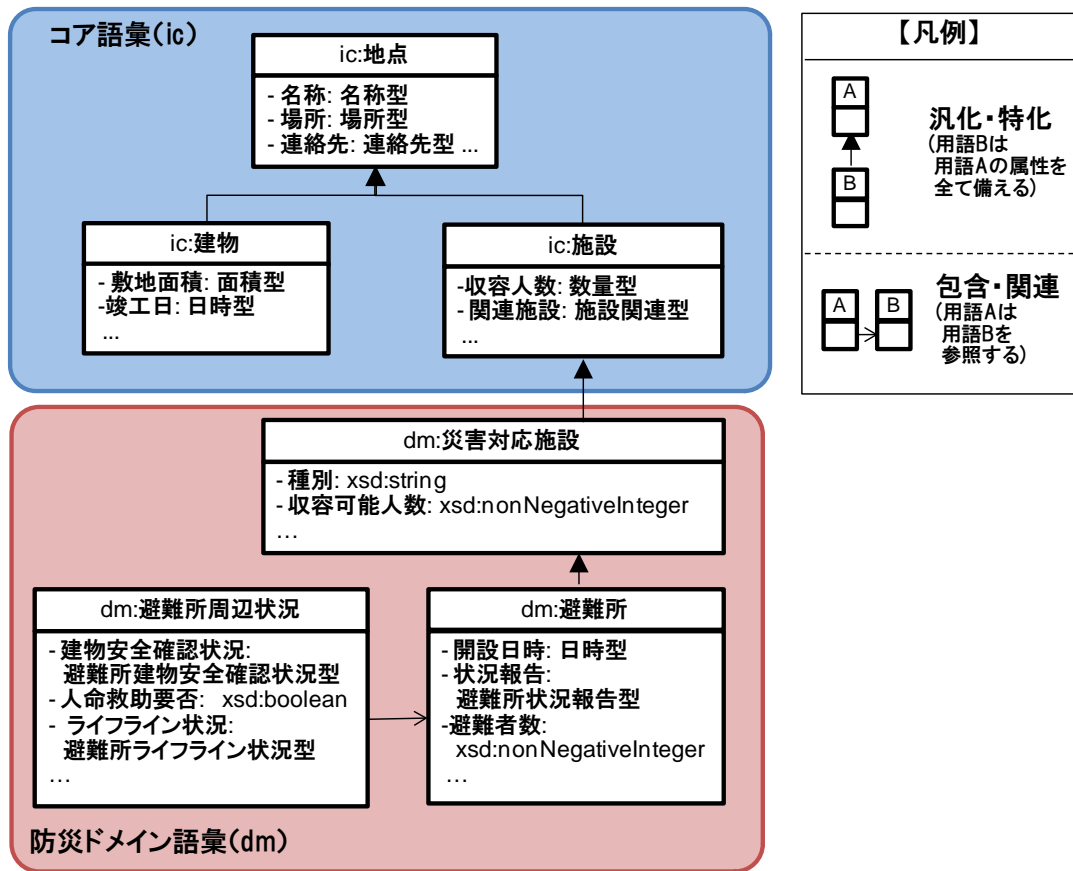


図 5.3.3.4 防災ドメインの語彙サンプルのクラス図

以下に、防災ドメインのドメイン語彙サンプルに含まれる主要な用語の一覧を示す。

表 5.3.3.5 防災ドメインのドメイン語彙サンプルに含まれる主要な用語（構造をもった用語）

項番	分類	項目説明
1	災害対応施設	避難所や広域避難場所、救護所など、災害対応に関わる施設を表現するためのデータ型。
2	災害対応施設種別	災害対応施設の種別を表現するためのデータ型。
3	災害種別	地震、風水害、火山災害など、災害の種別を表現するためのデータ型。
4	避難所	災害発生後、一定期間滞在することを前提とした収容目的の避難所を表現するためのデータ型。
5	避難所開設状況	避難所の開設状況を表現するためのデータ型。
6	避難所混雑状況	避難所の混雑状況を表現するためのデータ型。
7	避難所運営状況	避難所の運営に必要な組織・人員・情報を表現するためのデータ型。
8	避難所状況報告	避難所の災害対策本部に対する状況報告を表現するためのデータ型。
9	避難所公衆衛生状況	避難所の公衆衛生状況（避難者の健康状態や要支援者の内訳など）を表現するためのデータ型。
10	避難所周辺状況	避難所の周辺の状況を表現するためのデータ型。
11	避難所建物安全確認状況	避難所の建物安全確認状況を表現するためのデータ型。
12	避難所火災状況	避難所の火災状況を表現するためのデータ型。
13	避難所ライフライン状況	避難所のライフライン状況を表現するためのデータ型。
14	避難所周辺道路状況	避難所の周辺道路状況を表現するためのデータ型。
15	防災行政無線移動局	避難所に設置している防災行政無線の移動局を表現するためのデータ型。
16	無線移動局種別	防災行政無線の移動局の種別を表現するためのデータ型。
17	給水施設	非常用貯水槽、緊急給水栓、防災用井戸など、給水を行う施設を表現するためのデータ型。
18	給水施設種別	給水施設の種別を表現するためのデータ型。
19	座標	UTM 座標や N コードなど、災害対応で使用する座標系を追加できるように拡張した座標型を表現するためのデータ型。
20	N コード座標	消防防災関係者が主に利用している座標系を表現するためのデータ型。
21	UTM 座標	陸上自衛隊関係者が主に利用している座標系を表現するためのデータ型。

項番	分類	項目説明
22	避難者	住居が被害を受け、住居の場を失った、またはライフラインの被害により、日常生活が著しく困難、または避難勧告が発せられる等により緊急避難の必要がある方を表現するためのデータ型。
23	避難者種別	避難者の種別を表現するためのデータ型。
24	避難状態	避難者の避難状態を表現するためのデータ型。
25	家屋被害状況	避難者の家屋被害状況を表現するためのデータ型。
26	支援対象者	避難所に避難することのできない災害時要援護者、あるいはライフラインが被害を受け、日常生活が著しく困難なため、食料・物資等の支援が必要な方を表現するためのデータ型。
27	災害時要援護者	高齢者、障害者、乳幼児、妊産婦および言葉の通じにくい外国人等、災害時に特に配慮が必要な方を表現するためのデータ型。
28	要援護者種別	災害時要援護者の種別を表現するためのデータ型。
29	アレルギー	災害時要援護者のアレルギーに関する情報を表現するためのデータ型。
30	健康状態	避難者の健康状態を表現するためのデータ型。

6. 語彙データベースの要求事項

語彙データベースは、情報の表記、意味、構造の共通化に必要な語彙データを格納するとともに、住民、中央官庁や地方自治体の職員、民間事業者、個人などの想定利用者に対し、その語彙データを提供することを目的とするデータベースである。

本事業では、語彙データベースがこれらの目的を達成するために必要な、語彙データの格納に資するデータ構造と、語彙データの提供に資するAPIのそれぞれについて、要求事項の整理を実施するとともに、これらの実現案について検討を実施した。

6.1 要求事項の整理方針

本事業では、語彙データベースの要求事項を利用者ニーズに立脚した形で整理するため、まず語彙データベースの利用シーンを想定した上で、APIおよびデータ構造の要求事項を特定するアプローチを採用した。

具体的には、以下の5つのステップに従って語彙データベースのAPIおよびデータ構造に関する要求事項の整理を実施した。なお、各ステップにおいては、ツールプロジェクトにおける実証結果、語彙データベース検討会における検討結果、ドメイン語彙検討会における検討結果、コア語彙検討会における検討結果、海外における意見聴取の結果を適宜フィードバックし、客観的検討を行いつつ、先進事例との相互運用性(マッピング)の確保を配慮した。

ステップ1：利用シーンの想定

語彙データベースの利用シーンの初期仮説、データベース検討会およびドメイン語彙検討会を通じて得られた意見を基に導き出した利用シーン、ツールプロジェクトで設定した利用シーン、海外聴取より入手した海外事例で実績のある利用シーンより、想定利用シーンを整理した。

ステップ2：ツール群の想定

ステップ1にて整理した想定利用シーンより、語彙データベースの利用者が用いることが想定される汎用的なツール群を整理した上で、ツールプロジェクトで設計、開発、検討されたツール群を抽出した結果、海外聴取より入手した海外事例で実績のあるツール群より、想定ツール群を整理した。

ステップ3：実装機能案及びAPI案の策定

ステップ2にて洗い出しを実施した想定ツール群それぞれについて、実装する機能及びそれを実現するために必要なAPIを想定した。

ステップ4：データベースとして実装するためのデータ構造の拡充

ステップ3にて策定した語彙データベースのAPI案について、個々のAPIを実装するために語彙データベースが備えるべきデータ構造を特定し、その一案を策定した。

ステップ5：セキュリティ要件、スケーラビリティ要件の整理

ステップ3およびステップ4にて策定したAPI案およびデータ構造案をシステムとして実現する際のセキュリティ要件、スケーラビリティ要件について、整理した。

6.2 想定される利用シーンについて

語彙データベースの要求事項整理のためのステップ1として、語彙データベースの主な利用シーンおよび利用者についての整理を実施した。

本事業では、「1.2 共通語彙基盤の全体像」で示した利用シーンを初期仮説として設定し、語彙データベース検討会、ドメイン語彙検討会を通じて、専門家や将来の共通語彙基盤の利用者として想定される関係者を交えた客観的検討を行いつつ、ツールプロジェクトでの実証や、海外聴取で得た意見や知見を通じて、想定利用シーンの整理を行った。

6.2.1 利用シーンの初期仮説

利用シーンの整理を始める起点として、まず、「1.2 共通語彙基盤の全体像」で示した5つの利用シーンを初期仮説として設定した。設定した初期仮説について、以下に示す。なお、各利用シーンの概要図については、図2.1.1.1を参照のこと。

表 6.2.1.1 利用シーンの初期仮説

項番	利用シーン	概要
1	語彙データの作成	語彙の設計者や語彙データベースの管理者が、語彙データベースに格納する、共通語彙の語彙データの新規登録を実施する。
2	データ構造の設計	中央官庁や地方自治体の職員など、データを保有・提供可能な主体が、語彙データベースと連携し、共通語彙を用いて、再利用性の高いデータを作成するためのテンプレート（データテンプレート）を設計する。
3	データの入力	中央官庁や地方自治体の職員など、データを保有・提供可能な主体が、共通語彙を用いて設計されたデータテンプレートに基づいて生成された入力フォーム等を利用することで、再利用性の高いデータを効率的に作成する。
4	交換データの作成	中央官庁や地方自治体の職員など、データを保有・提供可能な主体が、各組織で保有している既存データを、共通語彙を用いて設計されたデータテンプレートに基づいて、組織間での相互理解・交換が可能なデータへ変換する。
5	データの意味確認	共通語彙を用いて記述されたデータを利用・活用する主体が、語彙データベースを参照することで、データの意味を明確にし、組織間・組織内で交換される情報や、オープンデータの意味を機械的に解釈する。

6.2.2 語彙データベース検討会およびドメイン語彙検討会を通じて導き出した利用シーン

語彙データベース検討会およびドメイン語彙検討会で、有識者から得た意見の一部より導き出した利用シーンを以下に示す。

表 6.2.2.1 語彙データベース検討会およびドメイン語彙検討会で得た意見から導き出した利用シーン

項番	意見	意見から抽出した利用シーン
1	行政分野と民間分野の間の情報連携時において、直接連携も可能であるが、オープンデータを介した形態をとっているところが特徴である。情報連携を行う際、今までは、それぞれのデータ項目について詳細な検討・決定を行った上で、情報連携を行っていたが、本事業では、一旦、オープンデータを介することで、見えるデータ構造化を狙っており、この点が新たなチャレンジである。	オープンデータを介する情報連携
2	ターゲットとしているエンドユーザが専門家であった場合は、用語のすり合わせと相互変換の議論であるため、相互変換表があれば実現できると考える。ただし、エンドユーザが非専門家であった場合は、語彙そのものも持っている歴史的なしなみがあるが故の分かりづらさ等も紐解いた上で、変換しなくてはならない。	非専門家を対象とした用語変換
3	場所のデータに関しては、住所のデータを、緯度・経度に変換し、マッピングする等の事例が既に存在する。共通語彙基盤を利用することでより、相互変換が簡易化できるのであれば興味深い。	複雑な構造のデータの相互変換
4	位置情報は、ポジション（点）だけでは使い道があまりない。ライン（線）とポリゴン（面、空間）により、相互変換できることが重要である。	複雑な構造のデータの相互変換
5	語彙データベース上に、未承認の用語も登録可能とする試みがあれば、面白いかもしれない。未承認のものでも、それと明示された状態で利用可能となれば、利用実績に応じて、後に委員会などでの承認の候補とするといった利用シーンについても考えられる。こういう仕掛けがあると、国主導のものでなく、オープンなプロセスに立脚した基盤であると印象づけられる可能性がある。	未承認用語の仮登録
6	用語に対する意見やコメントを書く利用シーンを想像した場合、用語の表記、意味、構造のどの部分に対するコメントなのかをどのように指し示すかは課題となり得る。 コメント対象を具体的に特定可能なAPIやツールも必要ではないか。 コメント用の語彙があっても良いかもしれない。	語彙・用語へのフィードバック
7	フェイスブックの「いいね！」のような仕組みを設け、議論を始めた人に通知がいくような利用シーンも考えられる。普及を考えた場合、ランキングを設けるなど、フィードバックの仕組みとその可視化は重要となる。	語彙・用語へのフィードバック

項番	意見	意見から抽出した 利用シーン
8	語彙に関する編集やバージョンアップの利用シーンが考えられる。その際には、各種操作のためのツールやAPIが必要となる。前バージョンとの関係情報が保持できれば、ツールやAPIの自動更新も実現できるかもしれない。	語彙データの編集・更新
9	IMIコア語彙v1.0をベースに、既にAED検索のアプリケーションが作成されている。語彙のアップデートの利用シーンを考えた場合、こうしたアプリやツールの互換性を維持するための方策検討が必要となる。バージョンアップ前後の差分についてもアナウンスが必要だろう。	語彙データの編集・更新
10	公共機関から公共機関、または市民に対するデータ交換・提供だけでなく、市民や事業者から公共機関への申請データの送付なども、共通語彙基盤の利用シーンとして考慮できるのではないか。	市民・事業者から公共機関へのデータ申請
11	事業者から公共機関への申請の利用シーンとして、シンガポールでは、建築確認申請のデータ（機械可読なデータ）をクラウド上に申請できる形になっている。	市民・事業者から公共機関へのデータ申請
12	用語のバージョンが異なった場合に、変換を支援する機能が必要かもしれない。そのためには、新しいバージョンの用語ができた場合に、以前のバージョンとの差異に関する情報を格納しておく必要がある。前のバージョンとの関係情報が残されていればひょっとしたらツールやAPIの自動更新も実現可能かもしれない。	データの マイグレーション

6.2.3 ツールプロジェクトで設定した利用シーン

ツールプロジェクトによる実証結果を参照し、各プロジェクトで設定した利用シーンを整理した結果を以下に示す。

表 6.2.3.1 ツールプロジェクトで設定した利用シーン

項番	ツールプロジェクトのテーマ	利用シーン
1	「施設・サービス」のデータ化	<p>自治体職員による、「かなざわ育なび.net」(子育て世帯に対する情報提供サービス)へのデータ入力業務を支援するため、「データ入力支援ツール」を試作した。</p> <p>本プロジェクトでは、自治体職員が入力したデータを RDF 形式に変換する際、効率的に、日付データ等の表現形式の共通化を図ることを目的として、入力規則等を定義した「①語彙データの作成」、語彙データを用いた「②データ構造の設計」、設計されたデータ構造に基づいた「③データの入力」を利用シーンに設定している。</p>
2	「地理空間」のデータ化	<p>自治体職員による、「道路占用許可申請書類」の電子化業務と調整業務を支援するため、「入力支援機能」と「集計/可視化機能」を試作した。「入力支援機能」は、共通語彙や地理情報 DB 活用することで、プルダウンによる町丁目レベルの指定や、地図画面上での占有対象道路の指定などを可能とし、職員による容易かつ正確なデータ生成を支援するためのツールである。一方、「集計/可視化機能」は、「入力支援機能」によって電子化された道路占用許可申請の情報を可視化し、いつどこで道路の占有が実施されるか/されたかを職員が容易に確認可能とするツールである。</p> <p>本プロジェクトでは、現状は紙ベースでの申請処理が行われている「道路占用許可申請書類」を自治体内部で職員が電子化し、集計・可視化された申請間を調整するための「調整会議」を実施する一連の業務を想定し、利用シーンとして設定している。</p>
3	「観光」のデータ化	<p>公的団体等の職員による、観光情報のオープンデータ化を支援するため、データ構造を設計支援するツール(「データ構造設計支援ツール」)およびデータ入力を支援するツール(「データ入力支援ツール」)を試作した。</p> <p>本プロジェクトでは、「データ構造設計支援ツール」と「データ入力支援ツール」を利用して、データを保有する各団体が、観光施設情報、イベント情報、観光施設歴史情報、飲食店情報に関するデータを登録し、オープンデータ化を行う過程を利用シーンとして設定している。</p>

項 番	ツールプロジェクト のテーマ	利用シーン
4	「制度」のデータ化	<p>複数自治体間に共通なテンプレートを構築するプロセスのシミュレーションを行うため、テンプレートを作成する際に、他の自治体で作成したテンプレートを活用できることや、複数自治体間でテンプレートを共有できること、自治体の職員など、データ構造や語彙、データ型といった特別な知識を持たない専門外の人であっても、データ構造の設計ができることを目的に、「データ構造設計支援ツール」を試作した。</p> <p>本プロジェクトでは、テンプレートの検討は1自治体の中で情報の構造化を検討することから始め、段階的に参加自治体を増やしながら、徐々に自治体間のデータ構造の違いの検証や、統一を図る過程を利用シーンとして設定している。</p>
5	「イベント」のデータ化	<p>自治体のホームページや共同運営サイト等で配信している子育てイベント情報(HTML データ)に対して、自治体職員及び関連団体職員が、共通語彙や外部語彙(schema.org 語彙)で記述した構造化データを作成し、RDFa 形式のメタデータとして埋め込むことを目的に、「構造化データ付与ツール」、「地理情報正規化 API」、「語彙正規化 API」、「構造化データ抽出ツール」を試作した。</p> <p>本プロジェクトでは、試作ツールを自治体職員が活用し、複数のホームページ等で配信されている子育てのイベント情報から再利用性の高いデータを作成する過程を利用シーンとして設定している。</p>

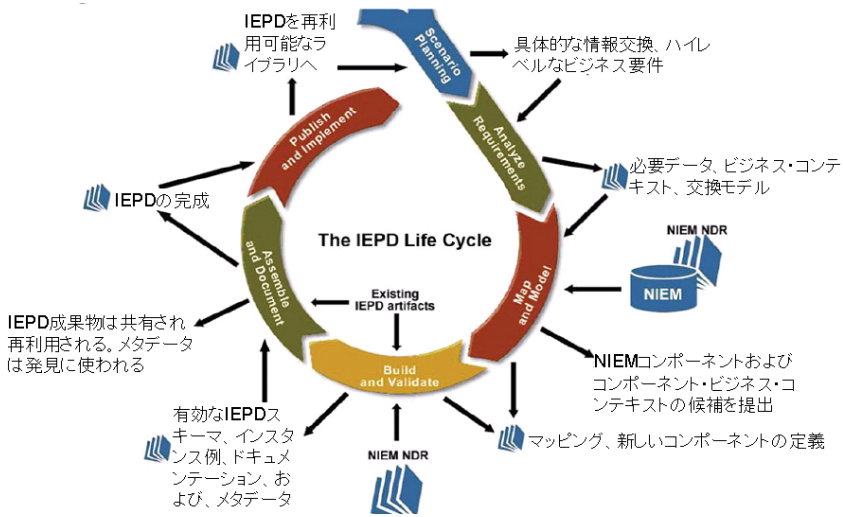
6.2.4 海外における先進事例で実績のある利用シーン

海外聴取や事前調査の結果に基づいて、海外における先進事例で実績のある利用シーンを整理した結果を以下に示す。

表 6.2.4.1 海外における先進事例で実績のある利用シーン

項番	分類	利用シーン
1	web サイトのカスタマイズ検索 (語彙データ作成、データ構造の設計、データの入力、メタデータの埋め込みの利用シーン)	schema.org の語彙を用いて、RDFa などの形式でメタデータが埋め込まれた web サイトについて、schema.org の用語（例えば、求人情報を示す http://schema.org/JobPosting 等）をフィルタの条件として利用することで、求人情報を掲載している web サイトのみに限定したカスタマイズ検索エンジンを簡易に作成することができる。
2	メールへのメタデータ埋め込み (語彙データ作成、データ構造の設計、データの入力、メタデータの埋め込みの利用シーン)	schema.org の語彙を用いてスケジュール等のデータを JSON-LD 形式で記述し、当該データを含むメールを送信する。メールの受信者のデバイス（スマートフォン等）では、当該メールを受信後、その内容を読み込んだ上で、（利用者自身がメールを読む前でも）利用者への通知やスケジュールへの自動登録などを実施することが可能になる。
3	EOC-I (語彙データ作成、データ構造の設計、データの入力、データ交換の利用シーン)	EOC-I(Emergency Operations Centers Interconnectivity)プロジェクトは 2009 年に Best of NIEM を受賞している。 EOC(Emergency Operating Center,災害対策本部)間で事件(Incident)の情報を交換するための IEPD (Information Exchange Package Documentation: データテンプレート相当のパッケージ文書)を開発し、当該 IEPD に従って、実際に情報交換を行うデモシステムを開発した。デモシステムでは、Google Earth や GIS サービスと連携し、EOC 間で交換した情報をインタラクティブな地図上に表示するなどの機能を持たせた。

項番	分類	利用シーン
4	<p>ASAP</p> <p>(語彙データ作成、データ構造の設計、データの入力、データ交換の利用シーン)</p>	<p>ASAP(Automated Secure Alarm Protocol)プロジェクトは、2013年に Best of NIEM を受賞している。</p> <p>ASAP(Automated Secure Alarm Protocol)プロジェクトでは、NIEM 語彙を活用し、警備会社が、公共機関に対して、警報を通知するための標準メッセージやプロトコルを定めた上で、ベンダ非依存の自動通報システムを確立した。これにより、従来の電話(911)による通報を無くし、誤解や聞き直しを排除することで緊急対応を高速化することができ、バージニア州では、従来 1.5~3 分以上かかっていた処理を 15 秒以下まで短縮することが出来た。</p>
5	<p>data.gov.uk</p> <p>(出典情報・法的根拠の付与、リンクの付与の利用シーン)</p>	<p>英国内の法律に関する情報を検索・閲覧可能なサイト (legislation.gov.uk) を利用し、用語の出典・法的根拠を示している。例えば、data.gov.uk のデータの一部には legislation.gov.uk 内の法律へのリンク(URI)が付与されている。</p>
6	<p>欧州パイロット事業</p> <p>(語彙データ作成、データ構造の設計、データの入力、データ交換の利用シーン)</p>	<p>欧州 ISA では、2013 年に、ISA コアボキャブラリを用いて、ベルギー内の住所データの相互接続のパイロット事業を実施した。</p>
7	<p>分散型の語彙整備</p> <p>(語彙の収集・整備、語彙の追加、更新・修正の利用シーン)</p>	<p>Schema.org では、Github、wiki、メーリングリスト(ML)を活用し、分散型の語彙整備プロセスを確立している。</p> <p>本語彙収集方法の利点として、ボトムアップによる語彙の新規追加や修正は誰でも提案可能である点が挙げられる。用語の取り込み判断は、各専門家を交えたワーキンググループ内の検討を経て、判断している。</p>

項番	分類	利用シーン
8	IEPD ライフサイクル (データ構造の設計)	<p>NIEM では、データ構造の設計に関して、データテンプレートに相当する IEPD の作成から公開までの一連の利用シーンが「IEPD ライフサイクル」として明示されている¹。IEPD ライフサイクルは、以下の6つの利用シーンで構成される。</p> <ul style="list-style-type: none"> ・ Scenario Planning (シナリオ計画) ・ Analyze Requirements (要求分析) ・ Map & Model (マッピングとモデル化) ・ Build & Validate (構築と検証) ・ Assemble & Document (組立と文書化) ・ Publish & Implement (発行と実装) <p>以下に NIEM IEPD ライフサイクルの概要図を示す。</p>  <p>図 6.2.4.1 NIEM IEPD ライフサイクルの概要図²</p>

¹ NIEM User Guide Volume 1, <http://reference.niem.gov/niem/guidance/user-guide/vol1/user-guide-vol1.pdf>

² “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）”

<http://datameti.go.jp/data/dataset/report-002-2012> より抜粋

6.2.5 想定利用シーンの整理結果

前節までの語彙データベースの利用シーンの初期仮説、データベース検討会およびドメイン語彙検討会を通じて得られた意見を基に導き出した利用シーン、ツールプロジェクトで設定した利用シーン、海外聴取より入手した海外事例で実績のある利用シーンより、語彙データベースの利用者・利用方法をまとめることで、想定利用シーンを整理した。

(1) 利用シーンの一般化

「6.2.1 利用シーンの初期仮説」で示した初期仮説に加え、「6.2.2 語彙データベース検討会およびドメイン語彙検討会を通じて導き出した利用シーン」、「6.2.3 ツールプロジェクトで設定した利用シーン」、「6.2.4 海外における先進事例で実績のある利用シーン」で得られた利用シーンについて、その内容に基づいて共通性の高いものを集約した上で、以下のように利用シーンを一般化した。

表 6.2.5.1 利用シーンの一般化

項番	利用シーン	6.2.1 初期仮説	6.2.2 検討会	6.2.3 ツールプロジェクト	6.2.4 海外先進事例
1	データ構造設計：シナリオ計画	データ構造の設計	—	「施設・サービス」のデータ化：職員による子育てデータの構造設計 「観光」のデータ化：公的団体職員による観光データの構造設計	NIEM ライフサイクル：シナリオ計画
2	データ構造設計：要求分析	データ構造の設計	—	「施設・サービス」のデータ化：職員による子育てデータの構造設計 「観光」のデータ化：公的団体職員による観光データの構造設計	NIEM ライフサイクル：要求分析

項番	利用シーン	6.2.1 初期仮説	6.2.2 検討会	6.2.3 ツールプロジェ クト	6.2.4 海外先進事例
3	データ構造設 計：マッピングと モデル化	データ構造の 設計	—	「施設・サービ ス」のデータ化： 職員による子育 てデータの構造 設計 「観光」のデー タ化：公的団体職員 による観光デー タの構造設計	NIEM ライフサイ クル：マッピング とモデル化
4	データ構造設 計：構築と検証	データ構造の 設計	—	「施設・サービ ス」のデータ化： 職員による子育 てデータの構造 設計 「観光」のデー タ化：公的団体職員 による観光デー タの構造設計	NIEM ライフサイ クル：構築と検証
5	データ構造設 計：組立と文書化	データ構造の 設計	—	「施設・サービ ス」のデータ化： 職員による子育 てデータの構造 設計 「観光」のデー タ化：公的団体職員 による観光デー タの構造設計	NIEM ライフサイ クル：組立と文書 化
6	データ構造設 計：発行と実装	データ構造の 設計	—	「施設・サービ ス」のデータ化： 職員による子育 てデータの構造 設計 「観光」のデー タ化：公的団体職員 による観光デー タの構造設計	NIEM ライフサイ クル：発行と実装

項番	利用シーン	6.2.1 初期仮説	6.2.2 検討会	6.2.3 ツールプロジェクト	6.2.4 海外先進事例
7	データテンプレートの段階的設計および再利用	－	－	「制度」のデータ化：3自治体による段階的なテンプレート作成・共有・活用	－
8	語彙の見直し	－	語彙・用語へのフィードバック	－	－
9	データ交換	交換データの作成	オープンデータを介する情報連携 複雑な構造のデータの相互変換	－	EOC-I ASAP 欧州パイロット事業
10	データ公開	－	オープンデータを介する情報連携	「施設・サービス」のデータ化：子育てデータの公開 「イベント」のデータ化：自治体サイトへのメタデータ埋め込み	－

項番	利用シーン	6.2.1 初期仮説	6.2.2 検討会	6.2.3 ツールプロジェクト	6.2.4 海外先進事例
11	データ入力	データの入力	—	「施設・サービス」のデータ化：職員による子育てデータの入力 「地理空間」のデータ化：市職員による道路占有許可申請書類の電子化 「観光」のデータ化：公的団体職員による観光データの入力 「イベント」のデータ化：自治体サイトへのメタデータ埋め込み	—
12	データ申請	—	市民・事業者から公共機関へのデータ申請	—	—
13	データのマッシュアップ	—	—	「地理空間」のデータ化：市職員による道路占有許可申請データと地理情報との連携	—
14	メタデータの活用	—	—	—	web サイトのカスタマイズ検索、メールへのメタデータ埋め込み
15	データの記述に使われる用語の意味確認	データの意味確認	非専門家を対象とした用語変換	—	—
16	データのマイグレーション	—	データのマイグレーション	—	—

項番	利用シーン	6.2.1 初期仮説	6.2.2 検討会	6.2.3 ツールプロジェ クト	6.2.4 海外先進事例
17	語彙データの提 案	－	未承認用語の仮 登録	－	分散型の語彙整 備
18	語彙データの提 案承認	－	未承認用語の仮 登録	－	分散型の語彙整 備
19	語彙データの新 規登録	語彙データの 作成	－	全プロジェクト で実施	－
20	語彙データの追 加	－	語彙データの編 集・更新	－	－
21	語彙データの修 正	－	語彙データの編 集・更新	－	－
22	語彙データの廃 止	－	語彙データの編 集・更新	－	－
23	出典情報・法的根 拠の付与	－	－	－	data.gov.uk

(2) 利用シーンの詳細化

「(1) 利用シーンの一般化」で得られた各利用シーンについて、利用シーンの詳細化を実施し、想定利用シーンとしてまとめた。想定利用シーンについて、以下に示す。各利用シーンに現れる利用者については、次節「(3) 利用シーンに現れる語彙データベースの利用者」で後述する。

表 6.2.5.2 想定利用シーン (S: Scene)

項番	利用シーン	概要
S1	データ構造設計：シナリオ計画	データ構造設計者が、データ交換プロジェクトを計画し、プロセスを確立し、データ交換（組織間情報連携）やデータ公開（オープンデータ提供）に関する業務要件を特定する。
S2	データ構造設計：要求分析	データ構造設計者が、特定された業務要件について、業務のコンテキストやデータの要件を理解し、文書化することで、データ交換やデータ公開に関する業務要件を洗練する。
S3	データ構造設計：マッピングとモデル化	データ構造設計者が、業務要件に基づいて、データ項目の定義を行い、交換・提供対象のデータモデルを設計する。 作成したデータモデルに含まれる各データ項目について、共通語彙（語彙データベースに格納される用語）で記述可能なものについては、データモデルと共通語彙の間の対応付け（マッピング）を整理し、用語対応表としてまとめる。 対応付けができなかった用語（語彙データベースに存在しない用語）に関しては、共通語彙に含まれる用語を拡張することで対応するか、独自の用語を設計・追加することで対応し、前ステップで作成した用語対応表に追加する。
S4	データ構造設計：構築と検証	データ構造設計者が、用語対応表に基づいて、作成したデータモデルを実装した、物理スキーマ(XML Schema や RDF Schema 等)を作成する。 さらに、作成した物理スキーマを用いて、交換・公開対象データのサンプルを記述し、スキーマの利用例として整形する。
S5	データ構造設計：組立と文書化	データ構造設計者が、データテンプレートや、データテンプレートに基づいて記述されるデータの内容を説明するドキュメントを作成する。 さらに、前ステップまでで得られた成果物一式を、自己完結・自己文書化された、単一のアーカイブとしてパッケージ化する。
S6	データ構造設計：発行と実装	データ構造設計者が、他のデータ構造設計者やデータ提供者による検索・発見・再利用のために、作成したデータテンプレートを公開する。

項番	利用シーン	概要
S7	データテンプレートの段階的設計および再利用	データ構造設計者やデータ提供者が、他者が設計したデータテンプレートを受け取り、必要な修正・拡張を加えた後、自身のデータ交換・公開に活用する。（データテンプレートの段階的設計） また、データテンプレートが蓄積されたりポジトリの検索を行い、目的に沿ったものがあれば転用・拡張し、データ構造設計のコストを削減する。（データテンプレートの再利用）
S8	語彙の見直し	データ構造設計者が、データテンプレートの設計に利用した共通語彙について、意見や評価、改善のための提案を提供する。
S9	データ交換 （組織間情報連携）	データ提供者が、データ交換の相手先となる組織との間で、データテンプレートについての合意を形成する。 自身のデータベースに格納された実データを、データテンプレートで規定される形式に整形（データ変換）し、共通語彙で記述された再利用性の高いデータ（標準メッセージ）として、他の組織等へ送付する。 標準メッセージの送付にあたっては、データテンプレートを添付するか、データテンプレートへのリンクを付与する。 標準メッセージを受信した組織等は、添付されたまたは参照したデータテンプレートに基づいて、受信したデータの意味を機械的に（または人が読む形で）解釈し、データ構造の変換を行った上で、自身のデータベースへ格納する等の処理を実施し、自身の業務に活用する。
S10	データ公開 （オープンデータ提供）	データ提供者が、自身のデータベースに格納された実データを、データテンプレートで規定される形式（データテンプレート内に含まれる物理スキーマに合致する形式や、web ページに埋め込まれる RDFa などのメタデータ）に変換・整形し、共通語彙で記述された再利用性の高いオープンデータとして、データカタログサイト等を通じて一般に公開する。 オープンデータの公開にあたっては、データテンプレートを同データカタログサイト等で合わせて公開する。 データ利用者は、公開されたオープンデータを他のデータ等と合わせて利用・活用する。
S11	データ入力	データ提供者が、データテンプレートに基づいて生成されたデータ入力フォームに対してデータを入力する。入力されたデータは、データテンプレートに基づいて自動的に物理形式（XML, RDF, JSON など）や、web ページに埋め込まれるメタデータ（RDFa など）へ変換され、オープンデータとして公開される。 データ入力者のデータ入力時には、自然言語で入力した住所または地名を、共通語彙に基づいて自動的に構造化した上で、緯度経度情報を含む形式へ正規化する。また、構造化された住所情報を基に、国土交通省の位置参照情報データ等を参照し、都道府県コード、市区町村コード等を取得・付加する。

項番	利用シーン	概要
S12	データ申請	<p>データ申請者は、申請対象の公共機関が公開しているデータテンプレートで規定される形式（データテンプレート内に含まれる物理スキーマに合致する形式）に変換・整形し、共通語彙で記述された再利用性の高い申請データとして、申請対象の公共機関へ送付する。</p> <p>同様の申請が他の公共機関で利用されている場合、その公共機関が同じデータテンプレートを採用している場合には、申請データをそのまま再利用して申請を実施する。異なるデータテンプレートが採用されている場合には、テンプレートに基づいた申請データの変換を行った上で、申請を実施する。</p>
S13	データのマッシュアップ	<p>データ利用者が、共通語彙を使って表現されたデータ群や、外部語彙(NIEM, ISA Core Vocabulary, schema.org, dublin core 等)を使って表現されたデータ群をマッシュアップして活用する。</p>
S14	メールへ埋め込まれたメタデータの活用	<p>データ利用者が、共通語彙を用いてスケジュール等のデータをJSON-LD形式で記述し、当該データを含むメールを送信する。受信側のデバイスは、当該データを受信後、その内容を読み込んだ上で、デバイス所有者への通知やスケジュールへの自動登録などを実行する。</p> <p>また、同メールに訪問先（「地点」）や関係者（「人」）、関係組織（「組織」）などの情報が記載されている場合は、適宜、民間事業者の公開データや公共機関から提供されるオープンデータを参照することで、自動的にデータの収集・追加・集約などを実施する。</p>
S15	データの記述に使われる用語の意味確認	<p>データ利用者が、加工・利用対象の実データの意味を理解するために、実データ内で利用されている用語の意味・定義を確認する。</p>
S16	データのマイグレーション	<p>データ利用者が、過去に（バージョンアップ前に）共通語彙を用いて表現されたデータテンプレートや、当該データテンプレートに基づいて記述された実データに対して、最新の共通語彙を用いた表現に修正を実施する。</p>
S17	語彙データの提案	<p>語彙設計担当者が、語彙データベースに格納する語彙データの新規整備や修正・更新・廃止のための検討を実施する。</p>
S18	語彙データの提案承認	<p>語彙設計管理者が、語彙設計担当者によって検討された、語彙データベースに格納する語彙データの新規整備や修正・更新・廃止のための案を承認する。</p>
S19	語彙データの新規登録	<p>語彙設計担当者または語彙データベース管理者が、語彙データベースに対して、語彙設計者の検討した用語の新規登録を実施する。</p>
S20	語彙データの追加	<p>語彙設計担当者または語彙データベース管理者が、語彙データベースに対して、語彙設計者の検討した用語の追加登録を実施する。</p>

項番	利用シーン	概要
S21	語彙データの修正	語彙設計担当者または語彙データベース管理者が、語彙データベースに対して、語彙設計者の検討した用語の更新案を反映する。
S22	語彙データの廃止	語彙設計担当者または語彙データベース管理者が、語彙データベースに対して、語彙設計者の検討した用語の廃止案を反映する。
S23	出典情報・法的根拠の付与	語彙設計担当者が、語彙データベースに格納される語彙データについて、出典情報や法的根拠が明確である用語については、当該情報や根拠またはそれらへのリンクを付与する。

(3) 利用シーンに現れる語彙データベースの利用者についての整理

前節「(2)利用シーンの詳細化」で示した想定利用シーンに現れる語彙データベースの利用者について、以下に詳細を示す。

表 6.2.5.3 語彙データベースの想定利用者 (U: User)

項番	利用者名	概要	利用者の例
U1	データ構造設計者	再利用性の高いデータ構造を作成し、語彙データベースに格納された共通語彙の再利用および拡張を実施し、データテンプレートとして、まとめ、提供、公開する主体。 一般に、技術的な専門性を有していることが想定される。	・中央官庁や地方自治体等の行政関係者 ・ドメインの有識者
U2	データ提供者	保有する実データについて、他の組織等との連携（データ交換）や、一般への公開（オープンデータ提供）を行う主体。 一般に、業務の専門家ではあるものの、技術的な専門性を必ずしも有さない場合がある。	・中央官庁や地方自治体の職員 ・インフラ関連の民間事業者
U3	データ利用者	他の組織から送付された実データ（交換対象メッセージ）を受け取り、自組織のデータベースへ格納する主体。	・中央官庁や地方自治体の職員
		公開されたオープンデータを活用し、必要に応じて他組織・他国で提供・公開されているデータとのマッシュアップを実施する主体。	・民間事業者 ・個人
U4	データ申請者	地方自治体等の公共機関に対して、事業者や個人に関する情報の登録や、業務や行為に関する認可を得る等の目的で、データを付与した申請を実施する主体。	・民間事業者 ・個人
U5	語彙データベース管理者	語彙データベースの稼動監視や、ドメインの増減管理、語彙データベースの利用者アカウントの管理など、語彙データベースの運用管理に関わる作業を実施する主体。	・中央官庁や地方自治体の職員
U6	語彙設計担当者	語彙データベースに格納する語彙データの整備や更新のための検討を実施し、語彙データベースに対して用語の新規登録・追加・修正・廃止の要求を出す主体。	・中央官庁や地方自治体の職員 ・ドメインの有識者
U7	語彙設計管理者	語彙設計担当者によって検討された、語彙データベースに格納する語彙データの整備や更新の案について、承認・非承認の判断を実施する主体。	・中央官庁や地方自治体の職員 ・ドメインの有識者

6.3 想定されるツール群について

語彙データベースの要求事項整理のためのステップ2として、前節にて整理した語彙データベースの想定利用シーンの検討結果や、ツールプロジェクトでの実証、海外聴取で得られた利用者の業務を支援するための汎用的なツール群を想定ツール群として整理した。

6.3.1 利用シーン検討結果に基づくツール群

「6.2 想定される利用シーンについて」で示した想定利用シーンそれぞれについて、利用シーンにおける主な業務を整理した上で、利用者の業務効率を高めるとともに、データ交換、データ公開対象のデータの精度や再利用性を高めることが期待される各種の支援ツールを整理した。

表 6.3.1.1 利用シーンの検討結果に基づくツール群

項番	利用シーン	利用シーンにおける主な業務	想定される業務支援ツール
1	S1: データ構造設計 : シナリオ計画	・業務要件の特定	・UML 等を用いた汎用的なモデリングツール
2	S2: データ構造設計 : 要求分析	・業務要件の文書化と洗練	・UML 等を用いた汎用的なモデリングツール
3	S3: データ構造設計 : マッピングとモデル化	・データモデリング ・データモデルと共通語彙の対応付け ・用語の拡張 ・独有用語の追加	・利用者の作成したデータモデルと共通語彙との対応付けを支援するツール ・利用者の要求に応じて、用語の拡張や、利用者独有用語の追加を支援するツール
4	S4: データ構造設計 : 構築と検証	・語彙の物理スキーマ作成 ・語彙の利用例作成	・語彙の物理スキーマ作成を支援するツール
5	S5: データ構造設計 : 組立と文書化	・語彙に関するドキュメント作成 ・データテンプレート作成	・データテンプレートの作成を支援するツール
6	S6: データ構造設計 : 発行と実装	・データテンプレートの公開	・データテンプレートを格納し、利用者による再利用を促進するためのデータベース
7	S7: データテンプレートの段階的設計および再利用	・データテンプレートの段階的設計 ・データテンプレートの再利用	・他者の作成したデータテンプレートに修正を加えることで段階的設計を可能とするツール ・データテンプレートを格納し、利用者による再利用を促進するためのデータベース

項番	利用シーン	利用シーンにおける主な業務	想定される業務支援ツール
8	S8: 語彙の見直し	<ul style="list-style-type: none"> ・ 語彙データに対する意見や評価、改善のための提案 	<ul style="list-style-type: none"> ・ 語彙データの新規登録・追加・更新・廃止内容を提案するためのツール
9	S9: データ交換 (組織間情報連携)	<ul style="list-style-type: none"> ・ 組織内 DB に格納されたデータを共通語彙で記述し、組織間で交換 	<ul style="list-style-type: none"> ・ データテンプレートに基づいて、交換対象データの構造を変換するツール ・ 交換対象データが、それに付随するデータテンプレートに準拠しているか(データテンプレートに含まれる物理スキーマで実データが検証に通過するか)を確認するためのツール
10	S10: データ公開 (オープンデータ提供)	<ul style="list-style-type: none"> ・ 組織内 DB に格納されたデータを共通語彙で記述し、オープンデータとして公開 	<ul style="list-style-type: none"> ・ データテンプレートに基づいて、公開対象データの構造を変換するツール ・ Web サイト等の HTML ファイルに対して、メタデータ(RDFa 形式など)を付与する作業を支援するためのツール
11	S11: データ入力	<ul style="list-style-type: none"> ・ データ入力フォームに対する好感/公開対象データの入力 ・ web サイトへのメタデータ埋め込み 	<ul style="list-style-type: none"> ・ データテンプレートに基づいて、入力支援用のインタフェース(カレンダーによる日付の入力や地図による場所の入力など)を備えたデータ入力フォームを生成し、フォームに入力されたデータをデータテンプレートに合致した形式で出力するツール ・ Web サイト等の HTML ファイルに対して、メタデータ(RDFa 形式など)を付与する作業を支援するためのツール

項番	利用シーン	利用シーンにおける主な業務	想定される業務支援ツール
12	S12: データ申請	・データ入力フォームに対する申請データの入力	・データテンプレートに基づいて、入力支援用のインタフェース(カレンダーによる日付の入力や地図による場所の入力など)を備えたデータ入力フォームを生成し、フォームに入力されたデータをデータテンプレートに合致した形式で出力するツール
13	S13: データのマッシュアップ	・各種語彙で記述されたデータのマッシュアップ	・データの記述に用いられている用語の意味や利用例を確認するツール ・データテンプレートに基づいて、マッシュアップ対象データの構造や用いられている用語を変換し、マッシュアップを容易にするツール ・マッシュアップ対象のデータが、それに付随するデータテンプレートに準拠しているか(データテンプレートに含まれる物理スキーマでデータが検証に通過するか)を確認するためのツール
14	S14: メールへ埋め込まれたメタデータの活用	・メタデータの抽出および変換	・データテンプレートに基づいて、メタデータの構造を変換するツール
15	S15: データの記述に使われる用語の意味確認	・データの記述に使われる用語の意味確認	・データの記述に用いられている用語の意味や利用例を確認するツール
16	S16: データのマイグレーション	・語彙の更新に伴うデータの移行	・更新前の語彙で記述されたデータを、更新後の語彙で記述されたデータへ変換するツール
17	S17: 語彙データの提案	・語彙データの新規登録/追加/更新/廃止に関する提案	・語彙データの新規登録/追加/更新/廃止内容を提案するためのツール

項番	利用シーン	利用シーンにおける主な業務	想定される業務支援ツール
18	S18: 語彙データの提案承認	・ 語彙データの新規登録/追加/更新/廃止に関する提案の承認	・ 語彙データの新規登録/追加/更新/廃止内容に関する提案を閲覧/承認するためのツール
19	S19: 語彙データの新規登録	・ 語彙データの新規登録	・ 語彙データの登録を実施するためのツール
20	S20: 語彙データの追加	・ 語彙データの追加	・ 語彙データの追加を実施するためのツール
21	S21: 語彙データの修正	・ 語彙データの修正	・ 語彙データの修正を実施するためのツール
22	S22: 語彙データの廃止	・ 語彙データの廃止	・ 語彙データの廃止を実施するためのツール
23	S23: 出典情報・法的根拠の付与	・ 語彙データに対する出典情報や法的根拠の付与	・ 語彙データに対して出典情報や法的根拠を付与するためのツール

6.3.2 ツールプロジェクトでの実証から得られたツール群

ツールプロジェクトによる実証結果を参照し、各プロジェクトで設計、開発、検討されたツール群を抽出した結果を以下に示す。(プロジェクト固有の要件に特化したツール、例えば「地理空間」のデータ化における「集計/可視化機能」や、「イベント」のデータ化における「構造化データ抽出ツール」については、語彙データベースの要求事項整理に直接的な関係をもたないため、抽出対象から外すこととした。)

表 6.3.2.1 ツールプロジェクトでの実証から得られたツール群

項番	ツール	ツールの概要
1	データ入力支援ツール (「施設・サービス」のデータ化)	<ul style="list-style-type: none"> ・RDF形式等で記述された意味が明確で再利用性の高いデータを出力するツールであり、広く利用されるオープンデータ作成の効率化、組織内あるいは組織間でのデータ連携を行う際のデータの質的向上といった利用効果が想定されるもの。
2	データ入力支援ツール (「施設・サービス」のデータ化)	<ul style="list-style-type: none"> ・データ構造設計支援ツールで作成されたデータテンプレート(スキーマ)に基づいて入力支援を実施するツール。 ・専門知識のない利用者でも利用可能なツールとして、わかりやすい機能とユーザインタフェースが必要。 ・データテンプレートの情報に基づいて、入力値検証機能、入力値提示・誘導機能、RDFデータ作成支援機能、意味明確化支援機能、表記統一化支援機能、語彙構造化支援機能、データ公開支援機能、語彙情報参照機能等の機能を提供する。
3	データ構造設計支援ツール (「施設・サービス」のデータ化)	<ul style="list-style-type: none"> ・語彙データベースを検索し、そこから必要な語彙を取得し、独自に定義する語彙とあわせてデータテンプレートを作成するツール。 ・データ構造設計者が求める語彙を素早く探し出せるように、語彙の用法の明確な記載と、類似する語彙をすべて抽出して表示できるような検索機能が必要。 ・独自の語彙の定義作成を支援する機能が必要。
4	入力支援機能 (「地理空間」のデータ化)	<ul style="list-style-type: none"> ・地理空間情報関連の共通語彙を個々の業務データに適用するために、データ作成者/提供者向けには郵便番号など日常的に利用される簡易的な地域コードを基点としたサジェスト型の検索機能を提供することにより、人間表示用のラベルである対象住所の構造化などを行いつつ、併せて機械処理用には入力対象に即した地域コード等の地理識別子を付与するような機能。

項番	ツール	ツールの概要
5	データ構造設計支援ツール (「観光」のデータ化)	<ul style="list-style-type: none"> ・オープンデータ化を行うにあたり、そのデータ構造を設計するためのツールである。テンプレートを作成したり、統一化された語彙データを管理したりする機能を有するもの。 ・テンプレートを作成・編集・削除する「テンプレート管理機能」、テンプレート毎の入力項目を管理する「項目管理機能」、データ入力支援ツールの一部設定を管理する「テンプレート設定機能」を有する。
6	データ入力支援ツール (「観光」のデータ化)	<ul style="list-style-type: none"> ・テンプレートを基にオープンデータ化をするデータを登録するためのツール。 ・テンプレートに対して、データを作成・編集・削除する機能。CSVでデータを一括登録する「データ管理機能」を有する。
7	データ構造設計支援ツール (「制度」のデータ化)	<ul style="list-style-type: none"> ・直感的な操作が可能なユーザインターフェースにより、ユーザーにXMLSchemaを意識させること無くパイロットシステムのデータベースへアクセスし、その内容を検索・閲覧することが可能。 ・また、語彙を一覧表示した時の一覧画面や、語彙を検索した際の検索結果画面、テンプレートのプロパティ編集画面上で語彙をクリックすることで、語彙の定義を見ることができ、ストレス無くデータ構造の設計を進めることができる。 ・自治体などで一般的に利用されている複数のブラウザ上での動作をサポートしており、自治体職員が普段業務に利用しているPCで追加のアプリケーションのインストールなしに動作する。 ・さらに、本ツールは複数自治体間のテンプレートの共有化のために、「既存のテンプレートの一覧を表示する機能」と、「テンプレートを閲覧、編集する機能」の、2つをメインの機能としてもつ。 ・「既存のテンプレートの一覧を表示する機能」では、既存のテンプレートの一覧から、閲覧、編集したいテンプレートを選択するか、または「新規作成」を選択することで、テンプレートの閲覧編集画面へ遷移する。 ・「テンプレートを閲覧、編集する機能」では、テンプレートの名称と（テンプレート自体の）説明文、及び使用されている語彙や項目名の閲覧、編集を行うことができる。 ・テンプレートの作成・編集の際には、それを支援するための機能として、「全ての語彙を一覧表示する機能」、「語彙のキーワード検索を行う機能」、「語彙の意味をその場で調べる機能」の3つの機能を有する。

項番	ツール	ツールの概要
8	構造化データ付与ツール (「イベント」のデータ化)	<ul style="list-style-type: none"> ・ユーザがホームページ画面を見ながらマウス操作をすることで、HTMLの複雑な構造を意識することなく RDFa 形式のメタデータを付与した HTML ファイルを作成する機能を提供する。 ・IMI 子育てイベント語彙データ(検証版)の語彙データに基づいてメタデータを付与することができるが、項目名については利用者に分かりやすいように表記を変更している(例:「活動名」→「名称」等)。
9	地理情報正規化 API (「イベント」のデータ化)	<ul style="list-style-type: none"> ・住所または地名から、構造化された住所(緯度経度情報を含む)を取得する。 ・構造化された住所の情報を基に、国土交通省の位置参照情報データを検索し、都道府県コード、市区町村コード等を取得する。 ・取得したデータを加工(正規化)し、JSON 形式で返戻する。
10	語彙正規化 API (「イベント」のデータ化)	<ul style="list-style-type: none"> ・入力された文字列から、対応する型を判断する。 ・型に応じた変換処理を呼び出す。(変換定義は別途設定ファイルから読み出すので、そこに追加は可能) ・代表語彙が設定されている場合は、代表語彙を返戻する。 ・例えば、年号を含む年月日及び日時を、西暦を基にした標準化された形式に変換可能。

6.3.3 海外における先進事例で実績のあるツール群

海外聴取を通じて、共通語彙基盤に関連するツール群に関する情報を収集した結果を以下に示す。

表 6.3.3.1 海外における先進事例で実績のあるツール群

項番	ツール	ツールの概要
1	CAM Editor ¹ (オープンソース)	<ul style="list-style-type: none">・ 情報交換やオープンデータ提供のための、スキーマのモデリングツール。・ NIEM 語彙を取り込み、ドラッグアンドドロップ操作のみで、NIEM 準拠の XML Schema を生成可能。・ ソースコードは OSL (Open Software License) v3.0 の下で公開。
2	MagicDraw® ² (No Magic 社)	<ul style="list-style-type: none">・ ビジネスプロセス、アーキテクチャ、ソフトウェア、システムのモデリングツール。・ Cameo NIEM-UML プラグインを用いることで、OMG によって標準化された NIEM-UML に準拠したモデリングが可能。UML で記述したモデルから、NIEM 準拠の XML Schema を自動的に生成可能。
3	NIEM UML Modeling Tool ³ (オープンソース)	<ul style="list-style-type: none">・ NIEM-UML を活用することで、NIEM の情報交換パッケージ (IEP) の構築を単純化するツール。・ ソースコードは Github 内の ISE のリポジトリで公開され、誰もが開発に参加可能。ライセンスは Eclipse Public License - v 1.0 を採用。

¹ http://camprocessor.sourceforge.net/wiki/index.php/Main_Page

² <http://www.nomagic.com/products/magicdraw.html>

³ <https://github.com/info-sharing-environment/NIEM-Modeling-Tool>

6.3.4 想定ツール群の整理結果

前節までの、利用シーン検討結果に基づくツール群、ツールプロジェクトでの実証から得られたツール群、海外における先進事例で実績のあるツール群より、想定ツール群を整理した。

(1) ツール群の一般化

「6.3.1 利用シーン検討結果に基づくツール群」、「6.3.2 ツールプロジェクトでの実証から得られたツール群」、「6.3.3 海外における先進事例で実績のあるツール群」で得られたツール群について、その内容に基づいて共通性の高いものを集約した上で、以下のようにツール群を一般化した。

表 6.3.4.1 ツール群の一般化

項番	ツール	6.3.1 利用シーン検討結果	6.3.2 ツールプロジェクト	6.3.3 海外先進事例
1	データモデリングツール	UML 等を用いた汎用的なモデリングツール	—	MagicDraw® ¹ NIEM UML Modeling Tool

¹ <http://www.nomagic.com/products/magicdraw.html>

項番	ツール	6.3.1 利用シーン検討結果	6.3.2 ツールプロジェクト	6.3.3 海外先進事例
2	データ構造設計支援ツール	<p>利用者の作成したデータモデルと共通語彙との対応付けを支援するツール</p> <p>利用者の要求に応じて、用語の拡張や、利用者独自用語の追加を支援するツール</p> <p>語彙の物理スキーマ作成を支援するツール</p> <p>データテンプレートの作成を支援するツール</p> <p>作成したデータテンプレートに修正を加えることで段階的設計を可能とするツール</p>	<p>データ構造設計支援ツール (「施設・サービス」のデータ化)</p> <p>データ構造設計支援ツール (「観光」のデータ化)</p> <p>データ構造設計支援ツール (「制度」のデータ化)</p>	<p>CAM Editor</p> <p>MagicDraw®</p> <p>NIEM UML Modeling Tool</p>

項番	ツール	6.3.1 利用シーン検討結果	6.3.2 ツールプロジェクト	6.3.3 海外先進事例
3	データ構造変換ツール	<p>データテンプレートに基づいて、交換対象データの構造を変換するツール</p> <p>データテンプレートに基づいて、公開対象データの構造を変換するツール</p> <p>データテンプレートに基づいて、マッシュアップ対象データの構造や用いられている用語を変換し、マッシュアップを容易にするツール</p> <p>データテンプレートに基づいて、メタデータの構造を変換するツール</p> <p>更新前の語彙で記述されたデータを、更新後の語彙で記述されたデータへ変換するツール</p>	<p>地理情報正規化 API (「イベント」のデータ化)</p> <p>語彙正規化 API (「イベント」のデータ化)</p>	-
4	データ入力支援ツール	<p>データテンプレートに基づいて、入力支援用のインタフェース(カレンダーによる日付の入力や地図による場所の入力など)を備えたデータ入力フォームを生成し、フォームに入力されたデータをデータテンプレートに合致した形式で出力するツール</p>	<p>データ入力支援ツール (「施設・サービス」のデータ化)</p> <p>入力支援機能 (「地理空間」のデータ化)</p> <p>データ入力支援ツール (「観光」のデータ化)</p>	-

項番	ツール	6.3.1 利用シーン検討結果	6.3.2 ツールプロジェクト	6.3.3 海外先進事例
5	構造化データ付与ツール	Web サイト等の HTML ファイルに対して、メタデータ (RDFa 形式など) を付与する作業を支援するためのツール	構造化データ付与ツール (「イベント」のデータ化)	—
6	語彙データ閲覧ツール	データの記述に用いられている用語の意味や利用例を確認するツール	—	—
7	妥当性検証ツール	<p>交換対象データが、それに付随するデータテンプレートに準拠しているか (データテンプレートに含まれる物理スキーマで実データが検証に通過するか) を確認するためのツール</p> <p>マッシュアップ対象のデータが、それに付随するデータテンプレートに準拠しているか (データテンプレートに含まれる物理スキーマでデータが検証に通過するか) を確認するためのツール</p>	—	—

項番	ツール	6.3.1 利用シーン検討結果	6.3.2 ツールプロジェクト	6.3.3 海外先進事例
8	語彙データ管理 ツール	<p>語彙データの登録を実施するためのツール</p> <p>語彙データの追加を実施するためのツール</p> <p>語彙データの修正を実施するためのツール</p> <p>語彙データの廃止を実施するためのツール</p> <p>語彙データに対して出典情報や法的根拠を付与するためのツール</p>	—	—
9	共通語彙 リポジトリ	データテンプレートを格納し、利用者による再利用を促進するためのデータベース	—	—
10	語彙データ提案 ツール	<p>語彙データの新規登録/追加/更新/廃止内容を提案するためのツール</p> <p>語彙データの新規登録/追加/更新/廃止内容に関する提案を閲覧/承認するためのツール</p>	—	—

(2) ツール群の詳細化

「(1)ツール群の一般化」で得られた各ツール群に対して、「表 6.2.5.2 語彙データベースの想定利用者 (U: User)」より、利用者を設定することで、ツール群の詳細化を実施し、想定ツール群としてまとめた。

想定ツール群に関する主な利用者および概要を以下に示す。

表 6.3.4.2 語彙データベースの想定ツール群 (T: Tool)

項番	ツール名	主な利用者	概要
T1	データモデリングツール	・データ構造設計者	UML 等のモデリング言語を用いた、データモデリングを支援するツール。MagicDraw® や NIEM UML Modeling Tool 等に相当する。
T2	データ構造設計支援ツール	・データ構造設計者	交換・公開するデータ(省庁・自治体の保有データ)の仕様に基づいて、当該データの項目に対応する共通語彙を選択し、データの型となる「データテンプレート」を設計するためのツール。 以下の機能を備える。 ・利用者の作成したデータモデルと共通語彙との対応付け ・利用者の要求に応じて、データテンプレートに対して、拡張後の用語や、利用者独自用語を追加 ・語彙の物理スキーマ作成 ・段階的設計（他者の作成したデータテンプレートへ修正を加えることでテンプレート設計を実施）

項番	ツール名	主な利用者	概要
T3	データ構造変換ツール	・データ提供者	<p>ある語彙や形式で表現された実データを、「データテンプレート」に基づいて、別の語彙や形式で表現されたデータへ変換するツール。</p> <p>以下の機能を備える。</p> <ul style="list-style-type: none"> ・データテンプレートに基づいて、交換対象データの構造を変換 ・データテンプレートに基づいて、公開対象データの構造を変換 ・データテンプレートに基づいて、マッシュアップ対象データの構造や用いられている用語を変換し、マッシュアップを容易化 ・データテンプレートに基づいて、メタデータの構造を変換 ・更新前の語彙で記述されたデータを、更新後の語彙で記述されたデータへ変換
T4	データ入力支援ツール	・データ提供者	<p>「データテンプレート」の形式に則ったオープンデータの作成・公開を容易にするためのツール。</p> <p>以下の機能を備える。</p> <ul style="list-style-type: none"> ・データテンプレートに基づいて、入力支援用のインタフェース(カレンダーによる日付の入力や地図による場所の入力など)を備えたデータ入力フォームを生成 ・フォームに入力されたデータをデータテンプレートに合致した形式で出力
T5	構造化データ付与ツール	・データ提供者	<p>Web サイト等の HTML ファイルに対して、メタデータ (RDFa 形式など) を付与する作業を支援するためのツール。</p> <p>ユーザがホームページ画面を見ながらマウス操作をする等により、HTML の複雑な構造を意識することなく操作できることが求められる。</p>
T6	語彙データ閲覧ツール	・データ利用者	<p>語彙データの情報(名称、定義、構造、典拠、同義語・類義語、対応する外部用語、利用例など)を閲覧するためのツール。</p>

項番	ツール名	主な利用者	概要
T7	妥当性検証ツール	・ データ利用者	ある実データが、それに付随するデータテンプレートに準拠しているか(データテンプレートに含まれる物理スキーマで実データが検証に通過するか)を確認するためのツール。
T8	語彙データ管理ツール	・ 語彙データベース管理者 ・ 語彙設計担当者 ・ 語彙設計管理者	語彙データの登録・追加・修正・更新・廃止を実施するとともに、出典情報や法的根拠を付与するためのツール。
T9	共通語彙リポジトリ	・ データ構造設計者 ・ データ提供者 ・ データ利用者	データテンプレートや各種ツール群、その他共通語彙基盤関連の各種資料や情報を格納し、利用者による共有・再利用を可能とするためのデータベース。 利用者の利便性を考慮し、リポジトリ内のコンテンツの検索機能や、コンテンツのメタデータ(テンプレート等の登録・更新日時や登録者の情報など)管理機能等が実装されていることが望ましい。
T10	語彙データ提案ツール	・ 語彙設計担当者 ・ 語彙設計管理者	語彙データの新規登録・追加・更新・廃止内容に関する提案や、当該提案に対する承認・非承認を実施するためのツール。

6.4 実装機能案及びAPI案の策定

ステップ2にて洗い出しを実施したツール群それぞれについて、実装する機能及びそれを実現するために必要なAPIを特定し、その一案を策定した。

6.4.1 語彙データベースに対する要求機能の洗い出し

本事業では語彙データベースのAPI案の策定に関し、「6.3.4 想定ツール群の整理結果」で示した想定ツール群について、当該ツール群が用いられる利用シーン毎に、語彙データベースに求める要求機能の洗い出しを実施した。

ツール群が求める語彙データベースの要求機能について、以下に示す。

表 6.4.1.1 ツール群が求める語彙データベースの要求機能

項番	ツール群	利用シーン	要求機能
1	T1: データモデリングツール	(データモデリングツールは、UML等を用いた汎用的なツールであり、語彙データベースとの直接の連携は想定されない。)	
2	T2: データ構造設計支援ツール	S3: データ構造設計: マッピングとモデル化	キーワードによる用語の検索が可能であること。
3			RDFデータへの適用と親和性の高い、SPARQLを用いた用語の検索が可能であること。
4			ある用語に関連する別の用語を検索可能であること。
5			S4: データ構造設計: 構築と検証
6			サブセットスキーマ生成のため、用語の部分取得が可能であること。
7		T3: データ構造変換ツール	S10: データ公開 (オープンデータ提供)
8	S13: データのマッシュアップ		マッシュアップ対象のデータそれぞれで使われている用語が異なる場合には、それらの用語を可能な限り共通語彙に変換できること。
9	S16: データのマイグレーション		語彙のバージョンアップ(更新)が実施された場合に、更新前の語彙を用いて記述された実データについては、更新後の語彙に自動的に変換できること。

項番	ツール群	利用シーン	要求機能
10	T4: データ入力支援ツール	S11: データ入力 S12: データ申請	自然言語で記述されたデータ（場所名や住所、元号を含む年月日日時など）を、語彙データベースに格納されたコア語彙やドメイン語彙の中で適切な用語（「住所」や「日時」など）を用いて構造化可能であること。
11			データ入力の際に、入力例を表示可能であること。
12	T5: 構造化データ付与ツール	S11: データ入力	自然言語で記述されたデータ（場所名や住所、元号を含む年月日日時など）を、語彙データベースに格納されたコア語彙やドメイン語彙の中で適切な用語（「住所」や「日時」など）を用いて構造化可能であること。
13	T6: 語彙データ閲覧ツール	S13: データのマッシュアップ	マッシュアップ対象のデータで使われている用語の意味や利用例を確認できること。
14		S15: データの記述に使われる用語の意味確認	実データの記述に利用されている用語の意味や利用例を確認できること。
15	T7: 妥当性検証ツール	（妥当性検証ツールは、実データがデータテンプレートに準拠していることを(XMLバリデータやRDFバリデータなどを活用して)オフラインで確認するためのツールであり、語彙データベースとの直接の連携は想定されない。）	
16	T8: 語彙データ管理ツール	S19: 語彙データの新規登録	語彙設計者の検討した用語の新規登録を実施できること。
17		S20: 語彙データの追加	語彙設計者の検討した用語の追加登録を実施できること。
18		S21: 語彙データの修正	語彙設計者の検討した用語の修正を実施できること。
19		S22: 語彙データの廃止	語彙設計者の検討した用語の廃止を実施できること。
20	T9: 共通語彙リポジトリ	（共通語彙リポジトリはデータテンプレートや各種ツール群などの保有庫であるため、語彙データベースとの直接の連携は想定されない）	
21	T10: 語彙データ提案ツール	S8: 語彙の見直し	データテンプレートで使われている用語に対する、意見や評価、改善のための提案を受け付け可能であること。

項番	ツール群	利用シーン	要求機能
22		S17: 語彙データの提案	語彙データベースに格納する語彙データの新規整備や修正・更新・廃止に関する提案を受け付け可能であること。
23		S18: 語彙データの提案承認	語彙データベースに格納する語彙データの新規整備や修正・更新・廃止に関する提案について、権限を保有した者が承認し、新規登録・修正・更新・廃止に向けたワークフローを推進できること。

6.4.2 API案の策定に関する検討結果

「6.4.1 語彙データベースに対する要求機能の洗い出し」から得られた要求機能のそれぞれについて、当該機能を利用者へ提供するためのAPIを整理した。

表 6.4.2.1 語彙データベースの要求機能とAPIの関係

項番	要求機能	要求機能の提供に必要なAPI
1	キーワードによる用語の検索が可能であること。	キーワード検索API：共通語彙データベースに格納されている用語を、キーワードに基づいて検索するためのAPI。
2	RDFデータへの適用と親和性の高い、SPARQLを用いた用語の検索が可能であること。	SPARQL検索API：共通語彙データベースに格納されている用語を、SPARQLで記述した条件に従って検索するためのAPI。
3	ある用語に関連する別の用語を検索可能であること。	関連検索API：用語と完全一致・類似・広義・狭義・関連の関係にある他の用語を検索するためのAPI。
4	オフライン処理を可能とするため、語彙データ全体の一括取得が可能であること。	一括取得API：ある語彙に含まれる用語の全体を、一括して取得するためのAPI。
5	サブセットスキーマ生成のため、用語の部分取得が可能であること。	部分取得API：ある語彙に含まれる用語の一部（サブセット）を取得するためのAPI。 指定された用語だけでなく、当該用語が依存する他の用語も同時に取得可能とする。 個別取得API：ある用語を個別に指定して取得するためのAPI。 指定された用語だけでなく、当該用語が依存する他の用語も同時に取得可能とする。
6	データ公開時に、NIEM、ISA Core Vocabulary、UN/CEFACT、schema.org等の外部用語で記述できるものについては、それらの併用も可能とすること。	外部用語への変換API：共通語彙データベースに格納されている用語から、それと対応する、外部用語(NIEM, ISA Core Vocabulary, UN/CEFACT, schema.org等の用語)への変換を行うためのAPI。
7	マッシュアップ対象のデータそれぞれで使われている用語が異なる場合には、それらの用語を可能な限り共通語彙に変換できること。	外部用語からの変換API：NIEM, ISA Core Vocabulary, UN/CEFACT, schema.org等の外部用語から、それと対応する、共通語彙データベースに格納されている用語への変換を行うためのAPI。
8	語彙のバージョンアップ（更新）が実施された場合に、更新前の語彙を用いて記述された実データについては、更新後の語彙に自動的に変換できること。	用語マイグレーションAPI：更新が行われた用語について、更新前の用語から更新後の用語を解決するためのAPI。

項番	要求機能	要求機能の提供に必要な API
9	自然言語で記述されたデータ（場所名や住所、元号を含む年月日日時など）を、語彙データベースに格納されたコア語彙やドメイン語彙の中で適切な用語（「住所」や「日時」など）を用いて構造化可能であること。	データ構造化 API：住所、スケジュール、組織情報、アクセス情報などの非構造化データを、共通語彙に基づいて構造化して返却する API。共通語彙の構造を完全に理解していない利用者に対して、共通語彙を利用したデータの構造化を支援するもの。
10	データ入力の際に、入力例を表示可能であること。	意味確認 API：用語の意味や利用例を確認するための API。
11	マッシュアップ対象のデータで使われている用語の意味や利用例を確認できること。	意味確認 API：用語の意味や利用例を確認するための API。
12	実データの記述に利用されている用語の意味や利用例を確認できること。	意味確認 API：用語の意味や利用例を確認するための API。
13	語彙設計者の検討した用語の新規登録を実施できること。	用語登録 API：共通語彙データベースに、用語を新規に登録・格納するための API。
14	語彙設計者の検討した用語の追加登録を実施できること。	用語登録 API：共通語彙データベースに、用語を新規に登録・格納するための API。
15	語彙設計者の検討した用語の修正を実施できること。	用語更新 API：共通語彙データベースに格納されている用語を更新するための API。
16	語彙設計者の検討した用語の廃止を実施できること。	用語廃止 API：共通語彙データベースに格納されている用語を廃止するための API。
17	データテンプレートで使われている用語に対する、意見や評価、改善のための提案を受け付け可能であること。	用語評価 API：用語に対する意見や評価を登録するための API。
18	語彙データベースに格納する語彙データの新規整備や修正・更新・廃止に関する提案を受け付け可能であること。	用語提案 API：語彙データベースに格納する語彙データの新規整備や修正・更新・廃止に関する提案の受け付けや、受付済の提案について、権限を保有した者が承認し、新規登録・修正・更新・廃止に向けたワークフローを推進するための API。
19	語彙データベースに格納する語彙データの新規整備や修正・更新・廃止に関する提案について、権限を保有した者が承認し、新規登録・修正・更新・廃止に向けたワークフローを推進できること。	用語登録 API、用語更新 API、用語廃止 API、用語承認 API

上記の結果に基づいて、語彙データベースの API 案を策定した。本 API 案については、語彙データベースの汎用ツール群において、必要となる機能（ツールから API に対する要求事項）を提供するとともに、共通語彙の構造を完全に理解していない利用者でも、共通語彙を利用したデータ交換・データ提供が可能となるような API を備えるよう考慮している。

策定した API 案について、以下に示す。

表 6.4.2.2 語彙データベースの API 案 (A: API)

項番	API 名	種別	要求	応答	利用者制限
A1	キーワード検索 API	共通語彙データベースに格納されている用語を、キーワードに基づいて検索するための API。	・ キーワード	・ キーワードに関する用語の識別子のリスト	なし
A2	SPARQL 検索 API	共通語彙データベースに格納されている用語を、SPARQL で記述した条件に従って検索するための API。	・ SPARQL クエリ	・ SPARQL クエリに対する応答	なし
A3	関連検索 API	用語と完全一致・類似・広義・狭義・関連の関係にある他の用語を検索するための API。	・ 用語 ・ 関係の種類 (完全一致、類似、など)	・ (指定された)用語と、(指定された)関係にある他の用語の識別子	なし
A4	一括取得 API	ある語彙に含まれる用語の全体を、一括して取得するための API。	・ 語彙識別子 ・ 物理形式の種類 (XML 形式、Turtle 形式などの形式名を指定)	・ 語彙に含まれる用語全体の物理形式 (XML 形式、Turtle 形式など)	なし
A5	部分取得 API	ある語彙に含まれる用語の一部 (サブセット) を取得するための API。 指定された用語だけでなく、当該用語が依存する他の用語も同時に取得可能とする。	・ 語彙識別子 ・ 取得対象の用語識別子のリスト ・ 物理形式の種類 (XML 形式、Turtle 形式などの形式名を指定) ・ 依存関係の解消フラグ (指定した用語が依存する用語も同時に取得するか否かを示す真偽値)	・ 語彙に含まれる用語のサブセット (一部)の物理形式 (XML 形式、Turtle 形式など)	なし

項番	API 名	種別	要求	応答	利用者制限
A6	個別取得 API	ある用語を個別に指定して取得するための API。 指定された用語だけでなく、当該用語が依存する他の用語も同時に取得可能とする。	<ul style="list-style-type: none"> ・用語識別子 ・物理形式の種類 (XML 形式、Turtle 形式などの形式名を指定) ・依存関係の解消フラグ (指定した用語が依存する用語も同時に取得するか否かを示す真偽値) 	<ul style="list-style-type: none"> ・用語識別子で特定される用語の物理形式 (XML 形式、Turtle 形式など) 	なし
A7	意味確認 API	用語の意味や利用例を確認するための API。	<ul style="list-style-type: none"> ・用語識別子 	<ul style="list-style-type: none"> ・用語識別子で特定される用語の意味および利用例 	なし
A8	用語評価 API	用語に対する意見や評価を登録するための API。	<ul style="list-style-type: none"> ・用語識別子 ・意見や評価の対象とする用語の属性 (「名称」「説明」など) ・意見や評価 	<ul style="list-style-type: none"> ・成功またはエラーの通知 	要検討 (利用者の認証を設けるか、匿名を許してスパム防止機能の導入程度にとどめる等の選択肢が存在する)
A9	外部用語からの変換 API	NIEM, ISA Core Vocabulary, UN/CEFACT, schema.org 等の外部用語から、それと対応する、共通語彙データベースに格納されている用語への変換を行うための API。	<ul style="list-style-type: none"> ・外部用語の識別子 	<ul style="list-style-type: none"> ・共通語彙に含まれる用語の識別子 	なし
A10	外部用語への変換 API	共通語彙データベースに格納されている用語から、それと対応する、外部用語(NIEM, ISA Core Vocabulary, UN/CEFACT, schema.org 等の用語)への変換を行うための API。	<ul style="list-style-type: none"> ・共通語彙に含まれる用語の識別子 ・変換対象の外部用語を含む語彙の識別子 	<ul style="list-style-type: none"> ・外部用語の識別子 	なし

項番	API 名	種別	要求	応答	利用者制限
A11	データ構造化 API	住所、スケジュール、組織情報、アクセス情報などの非構造化データを、共通語彙に基づいて構造化して返却する API。共通語彙の構造を完全に理解していない利用者に対して、共通語彙を利用したデータの構造化を支援するもの。	・自然言語で記述された非構造化データ（住所、スケジュール、組織情報、アクセス情報など）	・コア語彙の用語（「定型住所」、「スケジュール」、「組織」、「アクセス」）で構造化されたデータ	なし
A12	用語登録 API	共通語彙データベースに、用語を新規に登録・格納するための API。	・登録対象の用語に関するデータ (XML 形式、Turtle 形式などで記述)	・成功またはエラーの通知	語彙設計管理者など、特定の利用者に制限
A13	用語更新 API	共通語彙データベースに格納されている用語を更新するための API。	・更新対象の用語の識別子 ・更新内容 (XML 形式, Turtle 形式などで記述)	・成功またはエラーの通知	語彙設計管理者など、特定の利用者に制限
A14	用語廃止 API	共通語彙データベースに格納されている用語を廃止するための API。	・廃止対象の用語の識別子	・成功またはエラーの通知	語彙設計管理者など、特定の利用者に制限
A15	用語提案 API	語彙データベースに格納する語彙データの新規整備や修正・更新・廃止に関する提案を受け付けるための API。	・提案対象の用語の識別子 ・提案内容 (XML 形式, Turtle 形式などで記述)	・成功またはエラーの通知	語彙設計担当者
A16	用語承認 API	語彙データベースに格納する語彙データの新規整備や修正・更新・廃止に関する提案について、権限を保有した者が承認し、新規登録・修正・更新・廃止に向けたワークフローを推進するための API。	・対象の用語の識別子 ・判断結果（承認 or 非承認）	・成功またはエラーの通知	語彙設計管理者
A17	用語マイグレーション API	更新が行われた用語について、更新前の用語から更新後の用語を解決するための API。	・更新前の用語の識別子	・更新後の用語の識別子	なし

APIの詳細について、個別取得APIを一例として、仕様のサンプルを以下に示す。

表 6.4.2.3 API 詳細の例： A5. 個別取得 API

項番	項目	定義
1	API 名	個別取得 API
2	URL	https://imi.ipa.go.jp/api/100/part/<用語識別子>.<取得形式> <用語識別子>: 取得対象の用語の識別子 <取得形式>: レスポンスの形式。(xml: XML Schema、turtle: RDF の Turtle 表現、rdf: RDF の RDF/XML 表現、json: RDF の JSON-LD 表現)
3	HTTPメソッド	GET
4	利用者制限	なし
5	リクエスト パラメータ	なし
6	レスポンス [ステータス コード]	200: OK - 正常終了。 400: Bad Request - <用語識別子>の指定が無いためエラー。 404: Not Found - <用語識別子>で指定された用語が見つからないためエラー。 500: Internal Server Error - システム上のエラー。
7	レスポンス [ボディ]	<用語識別子>で指定された用語に関する情報を<出力形式>で指定された形式で表記したもの。
8	リクエスト例	GET /api/100/part/人.xml HTTP/1.1 Host: imi.ipa.go.jp
9	レスポンス例	HTTP/1.1 200 OK Content-Length: *** (実際にはコンテンツ長が含まれる) Connection: close Content-Type: application/xml; charset=utf-8 <?xml version="1.0" encoding="utf-8"?> <xsd:schema targetNamespace="http://imi.ipa.go.jp/ns/core/2" ...> <xsd:complexType name="人型"> </xsd:schema>

6.5 データベースとして実装するためのデータ構造の拡充

ステップ3にて策定した語彙データベースのAPI案について、個々のAPIを実装するために語彙データベースが備えるべきデータ構造を特定し、「4. コア語彙の整備」にて整備したデータ構造案に対して、データ構造の拡充を行う必要がある。

データ構造の拡充にあたっては、API実装に必要な属性と、RDFストアでの実装を行う上でのデータ構造についての2点について、検討した。

6.5.1 API実装に必要な属性の洗い出し

「6.4.2 API案の策定に関する検討結果」にて案を提示した語彙データベースのAPIを実装するためには、APIの処理の実行および応答の生成に必要な属性を漏れなく語彙データベースに格納しておく必要がある。語彙のデータ構造を具体化するにあたっては、これらの属性情報を洗い出し、語彙データの構造へ取り入れることが求められる。

そこで本事業では、語彙データベースのAPI案で示した各APIについて、API実装時に必要となる処理および当該処理に必要な属性を精査し、語彙データベースが備えるべき属性について洗い出しを実施した。

表 6.5.1.1 API実装時に必要となる処理および当該処理に必要な属性

項番	API名	API実装時に必要な処理	処理に必要な属性
1	A1: キーワード検索 API	指定されたキーワードを、名称や別名称、説明の中に含むようなクラス用語またはプロパティ用語を検索できること。用語の属するカテゴリなど、検索時にヒントとして活用可能な情報も指定できると望ましい。	<ul style="list-style-type: none"> ・用語の名称 ・用語の別名称 ・用語の説明 ・検索用のキーワードやカテゴリ
2	A2: SPARQL検索 API	指定された SPARQL クエリで記述された条件に従って、語彙データベース内の語彙データを検索できること。	<ul style="list-style-type: none"> ・RDF データとして整備された語彙データ
3	A3: 関連検索 API	指定された用語と完全一致・類似・広義・狭義・関連の関係にある他の用語を検索できること。	<ul style="list-style-type: none"> ・用語間の関係（完全一致、類似、など）
4	A4: 一括取得 API	ある語彙に含まれる用語の全体を、指定された物理形式で記述し、応答できること。	—
5	A5: 部分取得 API	ある語彙に含まれる用語の一部（サブセット）を取得できること。 指定された用語だけでなく、当該用語が依存する他の用語も同時に取得できること。	<ul style="list-style-type: none"> ・用語間の依存関係
6	A6: 個別取得 API	指定された用語を個別に取得できること。 指定された用語だけでなく、当該用語が依存する他の用語も同時に取得できること。	<ul style="list-style-type: none"> ・用語間の依存関係
7	A7: 意味確認 API	用語の意味を確認できること。	<ul style="list-style-type: none"> ・用語の説明 ・用語の利用例

項番	API 名	API 実装時に必要な処理	処理に必要な属性
8	A8: 用語評価 API	用語に対する意見や評価を登録できること。	・用語に対する利用者意見
9	A9: 外部用語からの変換 API	NIEM, ISA Core Vocabulary, UN/CEFACT, schema.org 等の外部用語から、それと対応する、共通語彙データベースに格納されている用語への変換ができること。	・用語と外部用語の関係
10	A10: 外部用語への変換 API	共通語彙データベースに格納されている用語から、それと対応する、外部用語(NIEM, ISA Core Vocabulary, UN/CEFACT, schema.org 等の用語)への変換ができること。	・用語と外部用語の関係
11	A11: データ構造化 API	住所、スケジュール、組織情報、アクセス情報などの非構造化データを、共通語彙に基づいて構造化して返却できること。	・構造化のための変換規則
12	A12: 用語登録 API	共通語彙データベースに、用語を新規に登録・格納できること。	・用語の整備状況 ・用語の登録日時 ・用語の作成者
13	A13: 用語更新 API	共通語彙データベースに格納されている用語を更新できること。	・用語の整備状況 ・用語の更新日時 ・用語の更新者
14	A14: 用語廃止 API	共通語彙データベースに格納されている用語を廃止できること。	・用語の整備状況 ・用語の廃止日時 ・用語の廃止者
15	A15: 用語提案 API	語彙データベースに格納する語彙データの新規整備や修正・更新・廃止に関する提案を受け付け可能であること。	・用語の整備状況
16	A16: 用語承認 API	語彙データベースに格納する語彙データの新規整備や修正・更新・廃止に関する提案について、権限を保有した者が承認できること。	・用語の整備状況
17	A17: 用語マイグレーション API	更新が行われた用語について、更新前の用語から更新後の用語を解決できること。	・用語の更新日時 ・用語の更新者 ・用語の廃止日時 ・用語の廃止者

上記の洗い出し結果より、語彙データベースにおける API 案の実装にあたっては、「4. コア語彙の整備」におけるデータ構造の検討結果には表れなかった、用語の検索に有効なキーワード、カテゴリなどの属性が有用であると判明した。

さらに、個々の用語の整備状況や登録・更新・廃止の日時に関する情報や、用語の用法を利用者が理解しやすくするための利用例、さらには、各用語に対して利用者から寄せられる利用者意見など、語彙データのメンテナンスや利用・促進・発展に寄与する属性も必要となることが判明した。

6.5.2 RDFストアでの実装を前提としたデータ構造案の策定

「4.2 構造的検討：コア語彙のデータ構造の拡充」で示した IMI コア語彙 v2 案のデータ構造は、語彙データの論理形式として、語彙テーブル、用語テーブル、関係性テーブルといった表（テーブル）を用いて整備した。

語彙データベースの実装にあたっては、以下の点から、RDF データモデルに基づくデータ構造を整備した上で、RDF ストアを用いたデータベース実装を推進することが望ましいと考える。

- ・「3.3 共通語彙基盤における語彙の設計方針」で示したように、本事業では、語彙データの物理形式として RDF 形式をサポートしていること。

- ・「6.4 実装機能案及び API 案の策定」で示したように、本事業では、SPARQL を用いた語彙データの検索を可能とする SPARQL API の実現を提案していること。

- ・ツールプロジェクトの大半で、RDF を用いた語彙データの整備が実施されていたこと。

- ・海外の各種先行事例で、RDF を用いた語彙データの整備が実施されていたこと。例えば、欧州 ISA がパイロット事業として開発・整備したメタデータレジストリは、RDF ストアを用いて、ISA Location Core Vocabulary、OASIS UBL、W3C Registered Organization vocabulary といった複数の語彙を集約・検索可能なパイロットシステムを整備していた。

そこで、本事業においては、RDF ストアによる実装に適した、語彙データベースのデータ構造案を示す方針を採用した。

6.5.3 データ構造案の策定に関する検討結果

「4. コア語彙の整備」にて整備した、語彙のデータ構造案をベースに、上記各種 API の実装に求められる属性の追加や、RDF ストアでの実装を考慮に入れた、RDF, RDFS, OWL 等の語彙との関係性の整理などを実施し、データ構造案の拡充を実施した。

拡充後のデータ構造案 (UML クラス図) について、以下に示す。「4. コア語彙の整備」にて整備したデータ構造との差分 (拡充した内容) については赤字で示している。

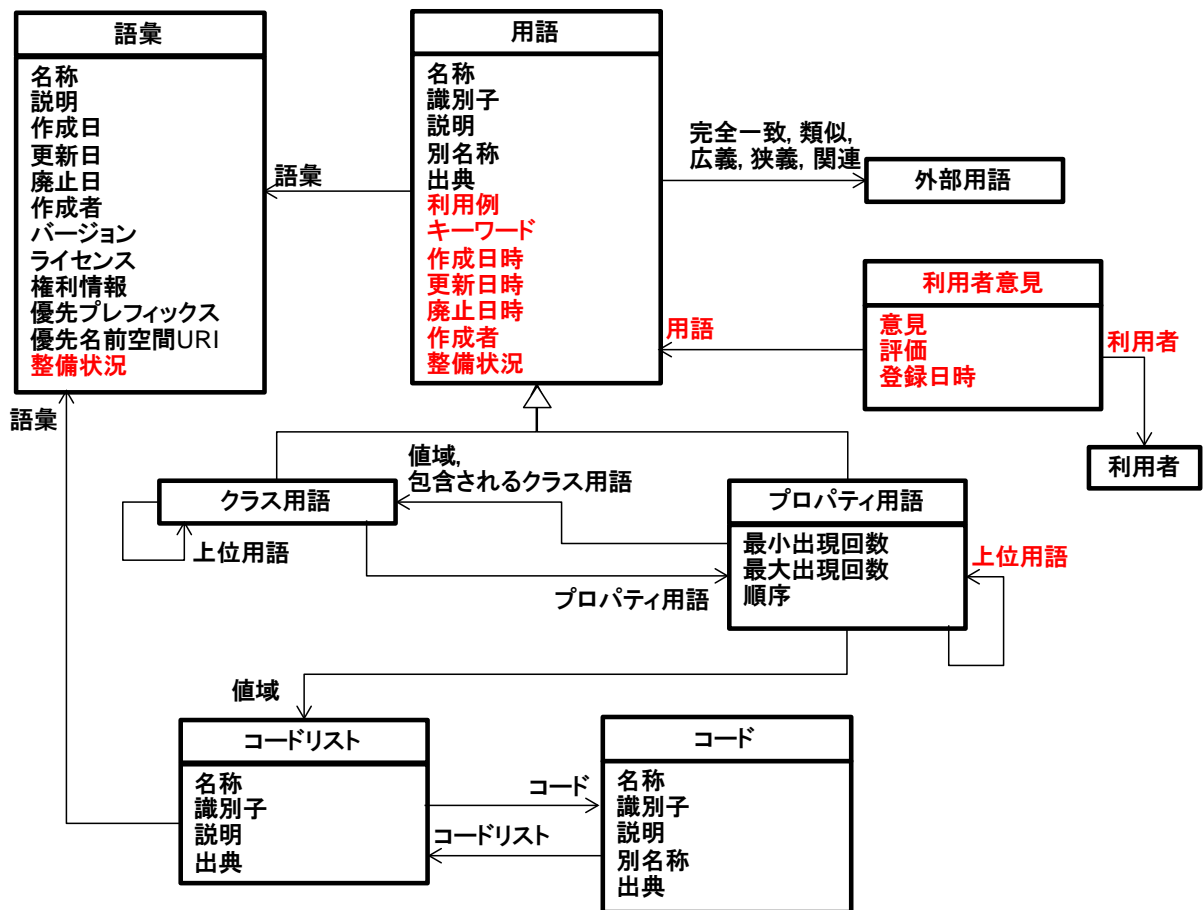


図 6.5.3.1 拡充後のデータ構造案 (UML クラス図)

次ページ以降で用いる各種プレフィックスについて、名前空間 URI との対応を以下に示す。

表 6.5.3.1 各種プレフィックスと名前空間 URI の対応

項番	名称	プレフィックス	名前空間 URI
1	ADMS	adms	http://purl.org/adms
2	Creative Commons	cc	http://creativecommons.org/ns#
3	Dublin Core Terms	dct	http://purl.org/dc/terms/
4	FOAF	foaf	http://xmlns.com/foaf/0.1/
5	OWL	owl	http://www.w3.org/2002/07/owl#
6	SKOS	skos	http://www.w3.org/2004/02/skos/core#
7	VANN	vann	http://purl.org/vocab/vann/
8	XML Schema	xsd	http://www.w3.org/2001/XMLSchema
9	IMI メタ語彙 ¹	imi	http://imi.ipa.go.jp/ns/def/100

¹ 次ページ以降で示す、「語彙」や「用語」といったクラスや、それらに付随するプロパティ（「語彙_名称」、「用語_名称」など）が所属する名前空間。語彙を記述するための語彙ともみなすことができるため、ここでは仮に「メタ語彙」という名称を付与した。

(1) 語彙の構造

「語彙」は、語彙データベースに格納される用語の集合である。

語彙データベースの RDF ストアによる実装を想定し、ここでは、「語彙」を owl:Ontology のサブクラスとして整備した。

表 6.5.3.2 「語彙」クラス

名称	識別子	説明	上位クラス
語彙	imi:語彙	語彙データベースに格納される用語の集合。	owl:Ontology

「語彙」クラスが備えるべき推奨プロパティを、以下に示す。

表 6.5.3.3 「語彙」クラスの推奨プロパティ

項番	名称	識別子	説明	値域	出現回数	同等の RDF 表現
1	名称	imi:語彙_名称	語彙の名称を自然言語で表記したもの。日本語による名称は必須とする。	xsd:string	1..n	rdfs:label, dct:title
2	説明	imi:語彙_説明	語彙の意味を自然言語で説明したもの。日本語による説明は必須とする。	xsd:string	1..n	rdfs:comment, dct:description
3	作成日	imi:語彙_作成日	語彙が新規に作成された日時。	xsd:date	0..1	dct:created
4	更新日	imi:語彙_更新日	語彙が更新された日時。	xsd:date	0..n	dct:modified
5	廃止日	imi:語彙_廃止日	語彙が廃止された日時。	xsd:date	0..1	
6	作成者	imi:語彙_作成者	語彙を作成した主体。	foaf:Agent	0..n	dct:creator
7	整備状況	imi:語彙_整備状況	語彙の整備状況。ADMS で定義される整備状況コード等を用いて表現する。	rdfs:Resource	0..1	adms:status
8	バージョン	imi:語彙_バージョン	語彙のバージョン。	xsd:string	0..1	owl:versionInfo
9	ライセンス	imi:語彙_ライセンス	語彙の利用に関するライセンス情報。	dct:LicenseDocument	0..n	dct:license, cc:license

項番	名称	識別子	説明	値域	出現回数	同等の RDF 表現
10	権利情報	imi:語彙_権利情報	語彙の権利に関する情報。一般的に、知的財産権を含む、語彙に関連する様々な所有権に関する文書など。	dct:Rights Statement	0..n	dct:rights
11	優先プレフィックス	imi:語彙_優先プレフィックス	語彙の名前空間のプレフィックスとして優先的に使用されることが望ましいもの。	rdfs:Resource	0..n	vann:preferredNamespacePrefix
12	優先名前空間 URI	imi:語彙_優先名前空間 URI	語彙の名前空間の URI として優先的に使用されることが望ましいもの。	rdfs:Resource	0..n	vann:preferredNamespaceUri

(2) 用語の構造

「用語」は、ある特定の意味を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ものである。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）をもつ場合がある。用語の特徴に応じて、クラス用語またはプロパティ用語に分類される。

語彙データベースの RDF ストアによる実装を想定し、ここでは、「用語」を owl:Class のサブクラスとして整備した。

表 6.5.3.4 「用語」クラス

名称	識別子	説明	上位クラス
用語	imi:用語	ある特定の意味を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もの。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）をもつ場合がある。クラス用語またはプロパティ用語の二つの種類が存在する。	owl:Class

「用語」クラスが備えるべき推奨プロパティを、以下に示す。これらは後述する「クラス用語」「プロパティ用語」にも共通して推奨されるプロパティとなることを想定している。

表 6.5.3.5 「用語」の推奨プロパティ

項番	名称	識別子	説明	値域	出現回数	同等の RDF 表現
1	名称	imi:用語_名称	用語の名称を自然言語で表記したもの。日本語による名称は必須とする。	xsd:string	1..n	rdfs:label
2	識別子	imi:用語_識別子	用語を一意に識別するための識別子。IRI (International Resource Identifier) またはその短縮表現である QName を用いて記述する。	rdfs:Resource	1..1	
3	語彙	imi:用語_語彙	用語の含まれる語彙。	imi:語彙	1..1	
4	説明	imi:用語_説明	用語の意味を自然言語で説明したもの。日本語による説明は必須とする。	xsd:string	1..n	rdfs:comment
5	別名称	imi:用語_別名称	用語の別名。同義語や類義語、表記ゆれを含む。	xsd:string	0..n	skos:altLabel

項番	名称	識別子	説明	値域	出現回数	同等の RDF 表現
6	出典	imi:用語_出典	用語の定義や用例、法的根拠等が示された文献やリソース。	rdfs:Resource	0..n	dct:source
7	完全一致	imi:用語_完全一致	用語と完全に一致する関係にある用語。	rdfs:Resource	0..n	skos:exactMatch
8	類似	imi:用語_類似	用語と類似の関係にある用語。	rdfs:Resource	0..n	skos:closeMatch
9	広義	imi:用語_広義	用語よりも広義の概念を示す用語。	rdfs:Resource	0..n	skos:broadMatch
10	狭義	imi:用語_狭義	用語よりも広義の概念を示す用語。	rdfs:Resource	0..n	skos:narrowMatch
11	関連	imi:用語_関連	用語と関連する用語。	rdfs:Resource	0..n	skos:relatedMatch
12	利用例	imi:用語_利用例	用語の利用例。	xsd:string	0..n	skos:example
13	キーワード	imi:用語_キーワード	用語の検索に用いるためのキーワードやタグ。	xsd:string	0..n	dcat:keyword
14	作成日時	imi:用語_作成日時	用語が新規に作成された日時。省略された場合は、用語が含まれる語彙の作成日と同等の日時として取り扱う。	xsd:dateTime	0..1	dct:created
15	更新日時	imi:用語_更新日時	用語が更新された日時。	xsd:dateTime	0..n	dct:modified
16	廃止日時	imi:用語_廃止日時	用語が廃止された日時。	xsd:dateTime	0..1	
17	作成者	imi:用語_作成者	用語を作成した主体。	foaf:Agent	0..n	dct:creator
18	整備状況	imi:用語_整備状況	用語の整備状況。ADMSで定義される整備状況コード等を用いて表現する。	rdfs:Resource	0..1	adms:status
19	利用者意見	imi:用語_利用者意見	用語に対して利用者から寄せられた意見や評価。	imi:利用者意見	0..n	

(3) クラス用語の構造

「クラス用語」は、用語の内、他の用語(プロパティ用語)を包含するような構造をもった用語である。UN/CEFACT CCTS や OASIS UBL における ABIE (Aggregate Business Information Entity) に相当するもの。他のクラス用語を継承することで、継承元のクラス用語が包含するプロパティ用語を包含することが可能である。

語彙データベースの RDF ストアによる実装を想定し、ここでは、「クラス用語」を「imi:用語」のサブクラスとして整備した。

表 6.5.3.6 「クラス用語」クラス

名称	識別子	説明	上位クラス
クラス用語	imi:クラス用語	用語の内、他の用語(プロパティ用語)を包含するような構造をもった用語。UN/CEFACT CCTS や OASIS UBL における ABIE (Aggregate Business Information Entity) に相当するもの。他のクラス用語を継承することで、継承元のクラス用語が包含するプロパティ用語を包含することが可能。	imi:用語

「クラス用語」クラスが備えるべき推奨プロパティを、以下に示す。「クラス用語」は「用語」のサブクラスであるため、「名称」や「識別子」をはじめ、「用語」の推奨プロパティはすべて「クラス用語」の推奨プロパティにも含まれるものとする。

表 6.5.3.7 「クラス用語」の推奨プロパティ

項番	名称	識別子	説明	値域	出現回数	同等の RDF 表現
1	上位用語	imi:クラス用語_上位用語	他のクラス用語を継承している場合は、継承元のクラス用語(上位クラス用語)を示す。	imi:クラス用語	0..n	rdfs:subClassOf
2	プロパティ用語	imi:クラス用語_プロパティ用語	クラス用語が包含するプロパティ用語。	imi:プロパティ用語	0..n	

(4) プロパティ用語の構造

「プロパティ用語」は、用語の内、クラス用語と他のクラス用語またはデータタイプ(文字列、整数などの基本的なデータの型)とを関連付けるものである。UN/CEFACT CCTS や OASIS UBL における BBIE (Basic Business Information Entity) または ASBIE (Association Business Information Entity) に相当するものである。

語彙データベースの RDF ストアによる実装を想定し、ここでは、「プロパティ用語」を「imi:用語」のサブクラスとして整備した。

表 6.5.3.8 「プロパティ用語」クラス

名称	識別子	説明	上位クラス
プロパティ用語	imi:プロパティ用語	用語の内、クラス用語と他のクラス用語またはデータタイプ(文字列や整数などの基本的なデータの型)とを関連付けるもの。UN/CEFACT CCTS や OASIS UBL における BBIE (Basic Business Information Entity) または ASBIE (Association Business Information Entity) に相当するもの。 他のプロパティ用語を継承することで、継承元のクラス用語が包含するプロパティ用語を包含することが可能。	imi:用語, owl:ObjectProperty

「プロパティ用語」クラスが備えるべき推奨プロパティを、以下に示す。「プロパティ用語」は「用語」のサブクラスであるため、「名称」や「識別子」をはじめ、「用語」の推奨プロパティはすべて「プロパティ用語」の推奨プロパティにも含まれるものとする。

表 6.5.3.9 「プロパティ用語」の推奨プロパティ

項番	名称	識別子	説明	値域	出現回数	同等の RDF 表現
1	上位用語	imi:プロパティ用語_上位用語	他のクラス用語を継承している場合は、継承元のクラス用語(上位クラス用語)を示す。	imi:プロパティ用語	0..n	rdfs:subPropertyOf
2	包含されるクラス用語	imi:プロパティ用語_包含されるクラス用語	クラス用語が包含するプロパティ用語。	imi:クラス用語	0..1	rdfs:domain
3	値域	imi:プロパティ用語_値域	プロパティ用語の値がどのクラス用語またはデータタイプのインスタンスであるかを示したもの。 RDF スキーマ語彙における rdfs:range に相当する。	imi:クラス用語	1..1	rdfs:range
4	最小出現回数	imi:プロパティ用語_最小出現回数	包含されるクラス用語の中での最小出現回数。省略した場合は、最小出現回数に制限を課さない、すなわち当該プロパティ用語が包含されるクラス用語の中で一度も出現しなくとも良いことを示す。	xsd:nonNegativeInteger	0..1	
5	最大出現回数	imi:プロパティ用語_最大出現回数	包含されるクラス用語の中での最大出現回数。省略した場合は、最大出現回数に制限を課さない、すなわち当該プロパティ用語が包含されるクラス用語の中で何度でも出現して良いことを示す。	xsd:nonNegativeInteger	0..1	
6	順序	imi:プロパティ用語_順序	包含されるクラス用語の中での順序。	xsd:integer	1..1	

(5) コードリストの構造

「コードリスト」は、プロパティ用語の取り得る値の表記を限定し、その意味を規定するために、文字列や数字列などの値を列挙したものである。統制語彙ともいう。

語彙データベースの RDF ストアによる実装を想定し、ここでは、「コードリスト」を skos:ConceptScheme のサブクラスとして整備した。

表 6.5.3.10 「コードリスト」クラス

名称	識別子	説明	上位クラス
コードリスト	imi:コードリスト	プロパティ用語の取り得る値の表記を限定し、その意味を規定するために、文字列や数字列などの値を列挙したものの。	owl:ConceptScheme

「コードリスト」クラスが備えるべき推奨プロパティを以下に示す。

表 6.5.3.11 「コードリスト」の推奨プロパティ

項番	名称	識別子	説明	値域	出現回数	同等の RDF 表現
1	名称	imi:コードリスト_名称	コードリストの名称を自然言語で表記したものの。	xsd:string	1..n	rdfs:label
2	識別子	imi:コードリスト_識別子	コードリストを一意に識別するための識別子。IRI (International Resource Identifier) またはその短縮表現である QName を用いて記述する。	rdfs:Resource	1..1	
3	語彙	imi:コードリスト_語彙	コードリストの含まれる語彙。	imi:語彙	1..1	
4	説明	imi:コードリスト_説明	コードリストの意味を自然言語で説明したものの。日本語による説明は必須とする。	xsd:string	1..n	rdfs:comment
5	出典	imi:コードリスト_出典	コードリストの定義や用例、法的根拠等が示された文献やリソース。	xsd:string	0..n	dct:source
6	コード	imi:コードリスト_コード	コードリストに含まれるコード。	imi:コード	1..n	skos:hasTopConcept

(6) コードの構造

「コード」は、「コードリスト」に含まれる個々の値である。統制語ともいう。

語彙データベースの RDF ストアによる実装を想定し、ここでは、「コード」を `skos:Concept` のサブクラスとして整備した。

表 6.5.3.12 「コード」クラス

名称	識別子	説明	上位クラス
コード	imi:コード	「コードリスト」に含まれる個々の値。	skos:Concept

「コード」クラスが備えるべき推奨プロパティを以下に示す。

表 6.5.3.13 「コード」の推奨プロパティ

項番	名称	識別子	説明	値域	出現回数	同等の RDF 表現
1	名称	imi:コード_名称	コードを自然言語で表記したもの。	xsd:string	1..n	rdfs:label
2	識別子	imi:コード_識別子	コードを一意に識別するための識別子。IRI (International Resource Identifier) またはその短縮表現である QName を用いて記述する。	rdfs:Resource	1..1	
3	コードリスト	imi:コード_コードリスト	コードの含まれるコードリスト。	imi:コードリスト	1..1	skos:inScheme
4	説明	imi:コード_説明	コードの意味を自然言語で説明したもの。日本語による説明は必須とする。	xsd:string	1..n	rdfs:comment
5	別名称	imi:コード_別名称	コードの別名。同義語や類義語、表記ゆれを含む。	xsd:string	0..n	dct:source
6	出典	imi:コード_出典	コードの定義や用例、法的根拠等が示された文献やリソース。	imi:コード	0..n	skos:hasTopConcept
7	完全一致	imi:コード_完全一致	コードと完全に一致する関係にある他のコード。	rdfs:Resource	0..n	skos:exactMatch

項番	名称	識別子	説明	値域	出現回数	同等のRDF表現
8	類似	imi:コード_類似	コードと類似の関係にある他のコード。	rdfs:Resource	0..n	skos:closeMatch
9	広義	imi:コード_広義	コードよりも広義の概念を示す他のコード。	rdfs:Resource	0..n	skos:broadMatch
10	狭義	imi:コード_狭義	コードよりも広義の概念を示す他のコード。	rdfs:Resource	0..n	skos:narrowMatch
11	関連	imi:コード_関連	コードと関連する他のコード。	rdfs:Resource	0..n	skos:relatedMatch

(7) 利用者意見の構造

「利用者意見」は、用語に対する利用者からの意見や評価である。

語彙データベースの RDF ストアによる実装を想定し、ここでは、「利用者意見」を `rdfs:Class` のインスタンスとして整備した。

表 6.5.3.14 「利用者意見」クラス

名称	識別子	説明	クラス
利用者意見	imi:利用者意見	用語に対する利用者からの意見や評価。	rdfs:Class

「利用者意見」クラスが備えるべき推奨プロパティを以下に示す。

表 6.5.3.15 「利用者意見」の推奨プロパティ

項番	名称	識別子	説明	値域	出現回数	同等の RDF 表現
1	利用者	imi:利用者意見_利用者	用語に対する意見の主体。	foaf:Agent	0..1	
2	用語	imi:利用者意見_用語	意見や評価の対象となる用語。	imi:用語	1..1	
3	対象	imi:利用者意見_対象	意見や評価の対象となる用語の属性（imi:用語_名称 や imi:用語_説明 など）。意見の対象を特に限定したい場合に用いる。	rdf:Property	0..n	
4	意見	imi:利用者意見_意見	用語に対する利用者からの意見。	rdfs:Resource	0..1	
5	評価	imi:利用者意見_評価	用語に対する利用者からの評価。1(最低評価)から5(最高評価)までの5段階の数値評価を想定。	xsd:nonNegativeInteger	0..1	rdfs:comment
6	登録日時	imi:利用者意見_登録日時	利用者意見が登録された日時。	xsd:dateTime	0..1	dct:created

6.6 セキュリティ要件の整理

語彙データベースの要求事項整理のためのステップ5として、「6.4.2 API案の策定」に関する検討結果にて策定したAPI案およびデータ構造案をシステムとして実現するために必要となるセキュリティ要件の整理を行った。

6.6.1 パイロットシステムの構成

本事業では、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と共に、本概念モデル書の構築に必要な知見の収集や、実運用上の課題を整理するため、語彙データベースのパイロットシステムは、オープンソースを採用し、構築・運用した。

パイロットシステムは、DMZセグメントに設置するWebサーバと内部Serverセグメントに設置するアプリケーションサーバ兼データベースサーバ（以下、AP/DBサーバ）にて構築されている。WebサーバのOSはRed Hat Enterprise Linux6.4（64bit）、WebサーバのソフトウェアとしてApacheを採用した。

AP/DBサーバのOSはRed Hat Enterprise Linux6.4（64bit）、アプリケーションサーバとしてTomcat、データベースサーバとしてPostgreSQL、開発言語としてJavaを採用している。

以下にパイロットシステムのシステム概要図を示す。

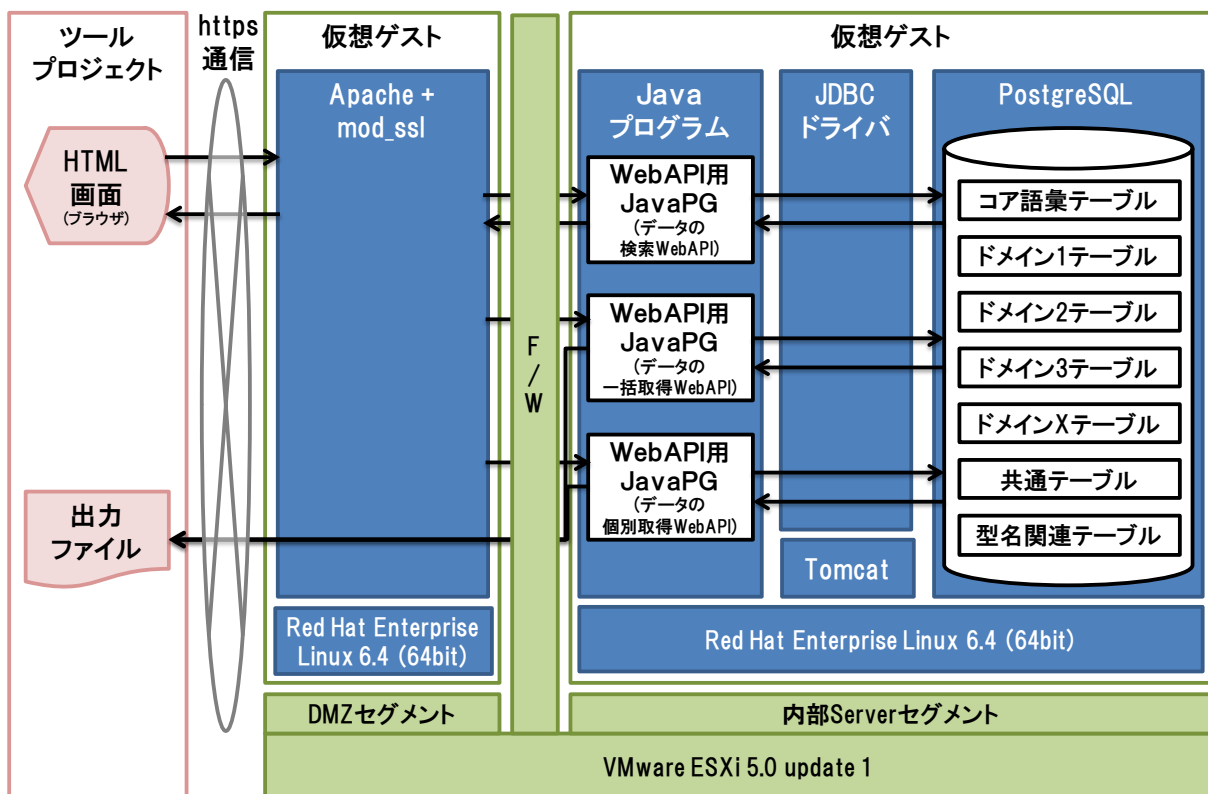


図 6.6.1.1 パイロットシステムの概要図

6.6.2 情報システムに求められるセキュリティ特性

共通語彙基盤は、中央官庁、地方自治体、民間企業や個人に利用される公共性の高いシステムであり、信頼性の高さが求められる。

共通語彙基盤の障害等によって語彙データベースの可用性や信頼性を確保できなくなる等の問題や、管理者を含めた不用意な操作等のセキュリティに関連する問題が生じた場合、語彙データベースの利用者に大きな影響を与える可能性がある。

このことから、本システムの信頼性やセキュリティ確保のための検討が必要となる。

情報システムに求められる3つのセキュリティ特性として、ISO/IEC 27002¹によって「機密性 (Confidentiality Impact)」、「完全性 (Integrity Impact)」、「可用性 (Availability Impact)」からなり、これらを維持することが情報セキュリティであると定義されている。

以下に情報の機密性、完全性、可用性の説明を示す。

表 6.6.2.1 機密性、完全性、可用性の説明

項番	特性	説明
1	機密性 (Confidentiality Impact)	情報へのアクセスを認められた者だけが、その情報にアクセスできる状態を確保すること
2	完全性 (Integrity Impact)	情報が破壊、改ざん又は消去されていない状態を確保すること
3	可用性 (Availability Impact)	情報へのアクセスを認められた者が、必要時に中断することなく、情報及び関連資産にアクセスできる状態を確保すること

¹ <http://www.standardsdirect.org/iso17799.htm>

6.6.3 語彙データベースに対するセキュリティ要件の整理

語彙データベースの信頼性やセキュリティ確保のために必要となる要件として、機密性への脅威、完全性への脅威、可用性への脅威に対して、主要な対策項目を整理し、その対策内容について、整理した。ただし、整理内容は、本事業の期間中に整理した項目であり、引き続き、他要素についても広く検討する必要がある。

以下に語彙データベースに対するセキュリティ要件の整理結果を示す。

項番	主要な対策項目	対策内容	機密性への脅威	完全性への脅威	可用性への脅威
1	管理者ユーザ認証、ユーザ管理	語彙データベースの運用業務の特性に応じて、利用者区分毎に適切な強度の認証レベルを採用する。 暗号化機能を利用して、主体認証情報を暗号化する。 主体認証情報は定期的な更新をチェックし、一定期間以上更新がなされていない場合には変更を促すとともに、ユーザカウントを一時的に利用停止する。	○	—	—
2	アクセス権限管理	語彙データベースを運用する上で、あらかじめ定義された権限を管理する。 暗号化機能を利用して、権限情報を暗号化する。	○	—	—
3	アクセス制御	ユーザ認証とアクセス権限管理により運用要員等が実行可能な運用業務を、あらかじめ許可された権限をもつ業務に制限する。	○	—	—
4	ゾーニング	不正アクセス及びアクセス制御機能のバイパス防止のために、セキュリティレベル毎にファイアウォールで区切られたゾーンを設け、システムを適切に管理できるようにする。	○	—	—
5	ハードニング	サーバ機及びネットワーク機器に対する不正アクセス及びセキュリティ上の脆弱性攻撃を防ぐために、不要サービスの停止及び適切なネットワーク設定を行う。	○	—	—
6	冗長化	システム機器の故障や外部からの不正な攻撃によりシステムが停止した際に、代替機によりシステム利用者様へサービスを継続できるようにする。 また、負荷分散構成とすることで、処理性能を拡張することができる。	—	—	○

項番	主要な対策項目	対策内容	機密性への脅威	完全性への脅威	可用性への脅威
7	侵入検知	侵入検知システムを設置し、外部及び内部からの不正なアクセスやその兆候を検知し、問題発生時には迅速な発見と対応ができるにする。	○	○	—
8	異常検知	情報及び語彙データベースへの不正アクセスや語彙データベースのパイロットシステムの異常を監視し、不正者の存在や異常の原因を検知可能とする。	○	○	○
9	ログ取得 (証跡管理)	語彙データベースへのログインや運用業務の利用時に、そのユーザ ID や操作内容をログファイルに記録する。	○	○	—
10	バックアップ /リストア	データの改ざんや障害によるデータ損失への対策として、事前に取得したバックアップデータからデータの復元を行う。	—	○	—
11	ウィルス対策	対策ソフトウェアを導入し、許可されていないプログラムの導入や起動を検知し、プログラムの除去を行う。	○	○	○
12	セキュリティ ホール対策	修正パッチ適用や、ファームウェアのバージョンアップ、ソフトウェアのバージョンアップを行う。	—	○	○
13	サービス不能攻撃 対策、踏み台対策	以下の対策を徹底することで、サービス不能攻撃に対応する。 □不要なサービスを動作させない □不要なポートを塞ぐ □不要な仮想サーバを動作させない	—	○	○
14	ソフトウェア設定 変更	Web サーバにおいて、初期状態で用意されるサンプルページ及びプログラム等を削除する。	—	—	○
15		Web サーバにおいては暗号化通信及び電子証明書による Web サーバの身元証明を提供するため、HTTPS を利用可能とする。	○	—	—
16	クラウドの利用	パブリッククラウドを利用することで、パブリッククラウドが備えているセキュリティ機能やバックアップ機能を利用することもできる。	○	○	○

項番	主要な対策項目	対策内容	機密性への脅威	完全性への脅威	可用性への脅威
17	開発・保守環境の整備	共通語彙基盤の開発・保守環境を整備する。 開発・保守環境用に、物理的もしくは論理的に本番環境ネットワークと分離されたネットワークを構築する。 開発・保守環境は、新規追加 API やパッチの事前適用、動作確認のための環境である。	—	—	○

6.7 スケーラビリティを配慮したアーキテクチャの検討

本事業を遂行中に、ツールプロジェクトにてパイロットシステムを利用中に、パイロットシステムが停止し、サービスを提供できなかったため、テンプレートの作成が行えなかった事象が発生した。

また、共通語彙基盤は、将来的に中央官庁、地方自治体、民間企業、個人にて活用される基盤であることことから、語彙量の増加や、利用シーンの拡大に伴う利用者の拡大、処理量の拡大が想定される。よって、これらの変化に対応し、継続したサービスを提供できるよう、スケーラビリティを配慮したアーキテクチャの検討が必要となる。

6.7.1 情報システムに求められるスケーラビリティの観点

スケーラビリティは、一般的には、将来想定されるシステム規模の増大に対して対応可能であるように設計することであり、規模の増大に対応するスケーラビリティを考慮する際には、垂直と水平の観点から捉えることができる。

以下に、情報システムに求められるスケーラビリティの観点について示す。

表 6.7.1.1 情報システムに求められるスケーラビリティの観点

項番	観点	方法
1	垂直	サーバーに CPU を追加したり、ハードディスクを増設したりするなど、同一論理単位内でリソースを増強して対応可能なことを垂直スケーラビリティという。
2	水平	クラスタ化、分散システム化により論理単位を追加することを水平スケーラビリティという。

6.7.2 スケーラビリティ要件の整理

語彙データベースのスケーラビリティに配慮するために必要となる観点として、垂直スケーラビリティ、水平スケーラビリティを確保するための主要な対策項目を整理し、その対策内容について、整理した。ただし、整理内容は、本事業の期間中に整理した項目であり、引き続き、他要素についても広く検討する必要がある。

以下に語彙データベースに対するセキュリティ要件の整理結果を示す。

表 6.7.2.1 語彙データベースに対するセキュリティ要件の整理結果

項番	主要な対策内容	対策内容	垂直スケーラビリティでの対応	水平スケーラビリティでの対応
1	冗長化	システム機器の故障によりシステムが停止した際に、代替機によりシステム利用者様へサービスを継続できるようにする。 また、負荷分散構成とすることで、処理性能を拡張することができる。	—	○
2	リソースの増強 (CPU、メモリ、ディスク)	処理量増加等に伴い、リソースを増強(CPU、メモリ、ディスク)する。	○	—
3	機器の増強 (サーバ、ストレージ、ネットワーク)	処理量増加等に伴い、機器(サーバ、ストレージ、ネットワーク)を増強する。	—	○
4	クラウドの利用 (仮想化技術の利用)	仮想化技術の採用によるリソース(CPU、メモリ、ディスク)の追加や変更、機器(サーバ、ストレージ、ネットワーク)の追加や変更が容易に行える。。	○	○
5	オープンな標準の採用	標準化機関等により標準仕様として標準化された技術であるか、または標準化されつつある技術を採用する。また、知的財産権(IPR: Intellectual Property Right)の扱いが標準化の際に明確にされており、不当な課金の可能性無しに誰もが自由にその仕様を実装できる製品を採用する。(OSSも含む)	○	○

項番	主要な対策内容	対策内容	垂直スケーラビリティでの対応	水平スケーラビリティでの対応
6	データベースの分散配置	<p>コアとドメインでの語彙データベースの物理/論理的な分離による、管理や負荷の分散を図る。</p> <p>1つの語彙データベースで全ての語彙データを管理するのではなく、例えばコア語彙用のデータベース、ドメイン語彙毎のデータベースを設けるといように、データベースを分散させる。これにより、ユーザアクセスの分散を図るとともに、新規ドメインの追加等が生じた場合でも、既存のコアやドメインのための容量が圧迫されるのを防ぐ。</p>	—	○

7. 本事業にて整備した共通語彙基盤の運用についての検討

本事業では、語彙のメンテナンスを含めた共通語彙基盤の運用に関わる検討を実施した。検討にあたっては、先行する事業である平成 24 年度事業の成果をベースに、諸外国の先進事例を踏まえた上で、将来の共通語彙基盤における運用組織体制および運用プロセスを示した。加えて、運用プロセスを整備するにあたっては、国際的な相互運用性を確保するため、海外政府機関・民間機関・各種標準化機関との将来的な相互運用プロセスの確立に向けた取り組みについての検討を実施した。

7.1 先行事例の運用体制に関する調査

共通語彙基盤の運用体制を検討するにあたり、米国における先行事例である NIEM、欧州における先行事例である ISA、民間で普及促進が進んでいる schema.org について、運用組織体制および運用プロセスに関する調査を実施した。

7.1.1 NIEM の運用体制

米国 NIEM の運用組織体制および運用プロセスについて、以下に概要を示す。

(1) NIEM の運用組織体制

米国 NIEM の運用は、意思決定機関である ESC（執行運営協議会）を頂点として、その下に実行機関である NIEM PMO（プログラムマネジメントオフィス）を備えた体制で実施されている。NIEM PMO の下には、NIEM の技術アーキテクチャを取りまとめる NTAC（技術委員会）と、NIEM の業務アーキテクチャを取りまとめる NBAC（業務委員会）が組織されている。

NBAC の下には、ドメイン毎に、各ドメイン語彙を取りまとめるための体制が組織されている。これらの組織体制はドメイン毎に異なるものとなるが、一般的には、ドメイン内を取りまとめる統治委員会と、ドメイン内の複数の利害関係機関(COI: Community of Interest)とで構成される。COI は、政府関係者と民間関係者とで構成されるグループである。例えば CYFS（児童・青少年・家族福祉）ドメインの場合、Juvenile Justice（少年法）、Child Welfare(児童福祉)、Child Support Enforcement（小児扶養義務）、Courts(裁判所)等の COI が存在する。

以下に、ESC、NIEM PMO、NTAC、NBAC、およびドメイン内の組織体制で構成される NIEM の運用組織体制の概要図を示す。ドメイン内の組織体制については、一例として CYFS（児童・青少年・家族福祉）ドメインの組織体制が記載されている。

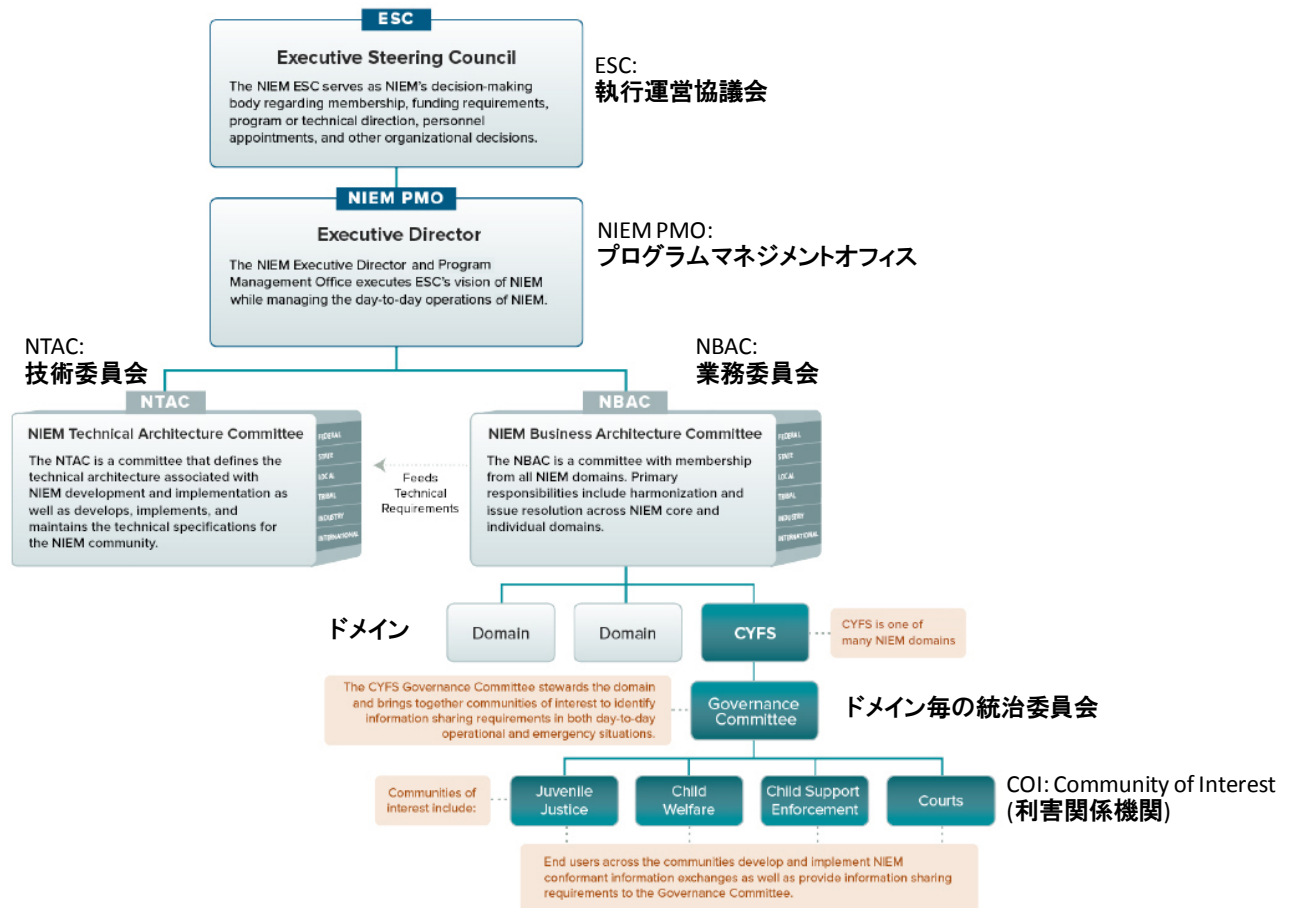


図 7.1.1.1 NIEM 運用組織体制¹

¹ “NIEM Domain Governance” https://www.niem.gov/meet-us/Pages/Domain_Governance.aspx に対して日本語の説明を加筆したもの。

NIEM の運用組織体制を構成する各組織の名称とその役割について、以下に概要を示す。

表 7.1.1.1 NIEM 運用組織名称とその役割

項番	組織名称	役割
1	ESC: Executive Steering Council (執行運営協議会)	NIEM の意思決定機関。メンバーシップ、予算要求、NIEM の実行計画と技術的な方針、PMO および各委員会の人事を決定する。PMO および各委員会のすべての決定事項は ESC の同意を必要とする。DHS(国土安全保障省)、DoJ(司法省)、HHS(保健福祉省)の CIO を中心に、Global (Global Justice Information Sharing Initiative)、NASCIO (全米州政府 CIO 会議)、OMB(行政管理予算局)の代表者や、PM-ISE (Program Manager of the Information Sharing Environment) から構成される。
2	NIEM PMO: NIEM Program Management Office	ESC の定めた NIEM のビジョンを実行する組織。共通語彙基盤に関わる日々の業務運営を行う。DHS の Donna Roy を中心に構成される。
3	NTAC: NIEM Technical Architecture Committee (技術委員会)	NIEM の開発及び実装において、技術的・構造的詳細に取り組む委員会。
4	NBAC: NIEM Business Architecture Committee	NIEM コア語彙の開発や、harmonization(調整)、バージョンアップ、及び、実装のガイドと業務アーキテクチャの観点からの運用プロセスのガイドを行う委員会。
5	Governance Committee	ドメイン毎に設置される委員会。ドメイン語彙の取りまとめや、ドメインに関係するコミュニティをまとめ、ドメイン内の情報交換に関する要件を特定する。
6	COIs: Communities of Interest	組織間連携や、日々の運用、サービスやプログラムの提供を通じて、NIEM に関わりのある、実務家および技術者で構成されるグループ。政府関係者と民間事業者の両者を含む。CYFS (児童・青少年・家族福祉)ドメインの場合、Juvenile Justice (少年法)、Child Welfare(児童福祉)、Child Support Enforcement (小児扶養義務)、Courts(裁判所)等の COI が存在する。

(2) NIEM の運用プロセス

米国 NIEM の運用プロセスの特長は、語彙のバージョン管理や、プロセスを厳格に定義している点である。

具体的には、語彙のバージョン管理アーキテクチャ仕様として NIEM HLVA (High Level Version Architecture)¹ が策定されている。NIEM HLVA では、語彙のリリースから次のメジャーバージョンのリリースまでの一連のプロセスを、12 のステップで構成している。

さらに、NIEM HLVA を実施する際に準拠すべき語彙の命名・設計に関する規則として、NIEM NDR (Naming and Design Rules) が策定されている。

NIEM HLVA および NIEM NDR を含む、NIEM の運用プロセスの概要図を以下に示す。

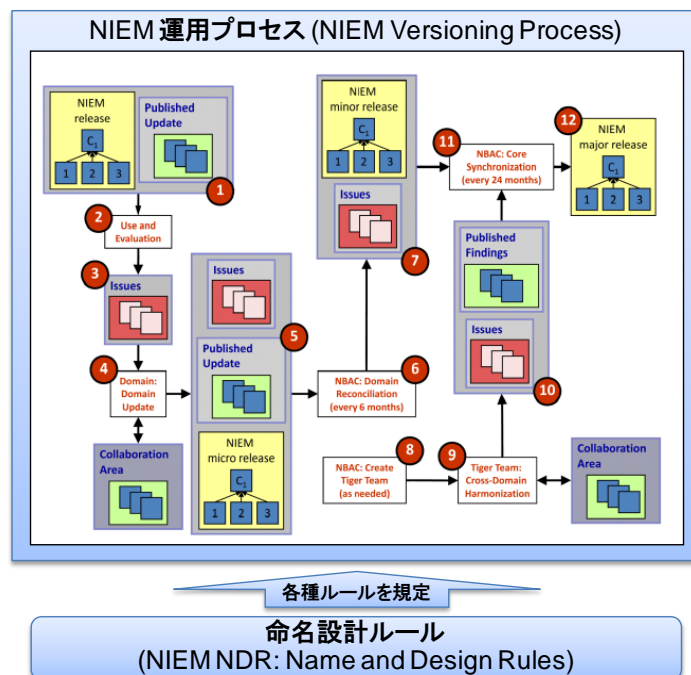


図 7.1.1.2 NIEM の運用プロセス・ルール

¹ <http://reference.niem.gov/niem/specification/high-level-version-architecture/1.0/>

NIEM HLVA では、NIEM における語彙の公開について、その規模に対応して以下の 4 種類のリリース種別を定めている。

表 7.1.1.2 NIEM における語彙の公開(リリース)の種別

項番	リリース種別	担当	頻度	概要
1	ドメイン・アップデート	ドメイン	随時	各ドメインは、ドメイン内で語彙のバージョンアップを実施した場合、更新内容を自由なタイミングで公開できる(NBAC によるレビューは不要)。公開されたドメイン・アップデートは主にドメイン内で利用されるが、ドメイン外での利用も可能。
2	マイクロ・リリース (例：NIEM2.1.3)	ドメイン	随時	各ドメインは、自ドメインのドメイン・アップデートについて、準拠性(conformance)と一貫性(coherency)を検証し、リリース領域(Release Area)に公開することもできる。ただし 2014 年 9 月時点で最新の 3.0 に至る NIEM の歴史で、マイクロ・リリースは一度も実施されていない。
3	マイナー・リリース (例：NIEM2.2)	NBAC	およそ 6 ヶ月毎	NBAC は、約 6 ヶ月毎に、蓄積されたアップデートを見直し、主にドメイン内整合性を配慮して、マイナー・リリースを Release Area に公開する。
4	メジャー・リリース (例：NIEM3.0)	NBAC (+専門家チーム)	およそ 18-24 ヶ月毎	IBAC は必要に応じて専門化チーム(Tiger team)と協力しながら、約 24 ヶ月毎に、ドメイン間の調整およびコアとの整合性を配慮した上で、メジャー・リリースを Release Area に公開する。

NIEM HLVA では、マイクロ・リリース、マイナー・リリース、メジャー・リリースの総称として「NIEM リリース」という表現が用いられている。

NIEM HLVA で定められた、語彙のリリースから次のメジャーバージョンのリリースまでの 12 のステップについて、以下に概要を示す。

表 7.1.1.3 NIEM HLVA のステップの概要

項番	ステップ	担当	概要
1	語彙の公開 (NIEM リリースまたはドメイン・アップデートの公開)	IBAC	NIEM リリース(マイクロ・リリース、マイナー・リリース、メジャー・リリース)や、ドメイン・アップデートが、IEPD での利用やその他の目的のため、IBAC から公開され、誰もが利用可能となる。
2	語彙の利用・評価	利用者	NIEM リリースや公開された語彙の利用や解析を通じて、各ドメインにおける利用者は、ドメイン語彙やコア語彙について、問題や新しい要件を発見する。ここで、利用者とは、IEPD の開発者に加えて、IEPD を利用した情報交換の実装者や利用者も含むものである。
3	課題の発見	利用者	前ステップで発見された、ドメイン語彙やコア語彙に関する問題や要件は、課題として課題管理領域(issue tracking area)に登録される。
4	ドメイン・アップデート (随時)	ドメイン	NIEM ドメインでは、課題管理領域に登録された課題に基づいて、段階的な改良、拡張、及び次期 NIEM リリースへの変更提案などが実施される。このプロセスは「ドメイン・アップデート」と呼ばれる。複数のドメインが参加する場合は、コラボレーション領域(collaboration area)に共同作業のための区画を作成する。
5	ドメイン・アップデートの公開	ドメイン	ドメイン・アップデートは、次の 3 種類の方法で公開される。 (1) 課題が更新され、新しい課題が課題管理領域に登録される。 (2) 公開領域(publication area)のドメイン語彙を更新する。 (3) 必要に応じて、更新後のドメイン語彙は、準拠性(conformance)と一貫性(coherency)の検証を経て、NIEM マイクロ・リリースとしてリリース領域(release area)に公開される。ただし、前述した通り、2014 年 9 月時点で最新の 3.0 に至る NIEM の歴史で、マイクロ・リリースは一度も実施されていない。
6	ドメイン調停	IBAC	ほぼ半年に一度の周期で、IBAC はドメイン・アップデートを調停(reconcile)する。このプロセスには次の 3 つの目的がある。 (1)新規あるいは修正されたドメイン語彙によって発生した矛盾を解決すること。 (2)対象ドメインにおいて、一貫性のある(coherent)リファレンス・スキーマー式を生成すること。 (3)スキーマ・セットに対して、インクリメンタルな調和作業(harmonization)を行うこと。 このプロセスにおいて、IBAC は、課題、公開されたドメイン・アップデート、及び、NIEM リリースを検証し、ドメインへの変更を単一の整合性のあるドメイン・スキーマ・セットへと調停する。IBAC は、すべてのドメインが、一貫性をもって相互参照していることを担保する。

項番	ステップ	担当	概要
7	NIEM マイナー・リリースの公開	IBAC	ドメイン調停の結果は、ほぼ 6 ヶ月に一度の周期で、NIEM マイナー・リリースとして公開する。 NIEM マイナー・リリースに伴うバージョン ID は、4.2 のように、2 個の数字から構成され、2 個目がゼロでないような ID とする。 また、ドメイン調停によって解決されなかった課題については、課題管理領域で、更新、作成、打ち切りなどの適切な処理がとられる。
8	クロス・ドメイン調整のための専門家チーム (Tiger Team) 編成	IBAC または ITAC	クロス・ドメイン調整が必要となった場合、IBAC または ITAC は、他のプロセスとは独立に、Tiger Team を編成し、特定の問題の解決を図る。この Tiger Team は、SME(内容領域専門家、Subject Matter Experts)から構成される小さなワーキンググループであり、特定の問題の解決のために編成され、短期間だけ存在し、任務終了後には解散する。
9	クロス・ドメイン調整	Tiger Team	クロス・ドメイン調整(cross-domain harmonization)とは、IBAC または ITAC によって編成された Tiger Team によるドメインとコアとの間の主要な調停作業の課題の解決や、複数ドメイン内に存在している重複コンポーネントの発見、NIEM として共通の定義と利用が可能になるように、コアへの引き上げ作業である。
10	クロス・ドメイン調整結果の公開	Tiger Team	Tiger Team の作業結果は、NIEM コア名前空間への追加や拡張を行うスキーマが含まれる公開領域にて公開される。 この段階では、各スキーマは、NIEM コア名前空間とは異なる独自に定義した名前空間で公開される。
11	コア同期	IBAC	ほぼ 18 ヶ月から 24 ヶ月ごとに、コア同期(Core synchronization)が行われる。コア同期とは、クロス・ドメイン調整及び、ドメイン調停の結果をまとめる作業である。 コア同期では、課題管理領域や各ドメイン及び Tiger Team により公開されたアップデートを参考にして、コア語彙が更新される。以前の NIEM コアに依存するドメインにおいても、更新されたコア語彙との調整がなされる。
12	NIEM メジャー・リリースの公開 (1 に戻る)	IBAC	コア同期の結果、コア語彙が更新されたものを公開することを、NIEM のメジャー・リリースと呼ぶ。 なお、NIEM のメジャー・リリースでは、バージョン ID は、「4.0」のように、後ろが「0」であるような 2 個の数字から構成される。

7.1.2 ISA の運用体制

欧州 ISA の運用組織体制および運用プロセスについて、以下に概要を示す。

(1) ISA の運用組織体制

欧州における相互運用性確保のためのプロジェクト ISA(Interoperability Solutions for European Public Administrations)によって策定された、ボキャブラリ整備のための組織体制¹を示す。

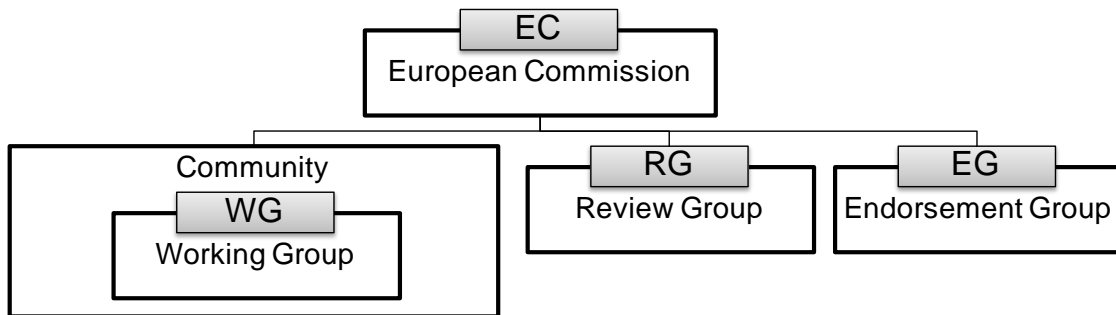


図 7.1.2.1 ISA 運用組織体制

表 7.1.2.1 ISA 運用組織名称とその役割

項番	組織名称	役割
1	EC: European Commission	欧州委員会。プロセス全体および成果物の所有者。戦略的な方向性を提示するとともに、内外のリソースを用いて Working Group と Review Group の取り纏め・サポートを実施する。
2	Community	整備対象のボキャブラリに関するステークホルダー(利害関係者、専門家など)の集合。
3	Working Group	Community の一部の人々で構成され、ミーティングへの参加やドラフト作成などの実務への貢献を求められる。
4	Review Group	Working Group の成果物をレビューする。ISA のメンバである欧州各国の代表により構成される。
5	Endorsement Group	ISA Coordination Group または ISA Trusted Information Exchange (TIE) Cluster Group の代表から構成される。ボキャブラリの最終決定権限を有する。

¹ “PROCESS AND METHODOLOGY FOR CORE VOCABULARIES”

https://joinup.ec.europa.eu/catalogue/asset_release/process-and-methodology-developing-core-vocabularies

(2) ISA の運用プロセス

欧州 ISA の運用プロセスの特長は、コア語彙を新規に収集・整備を対象として、プロセスが定義されている点である。なお、ドメイン語彙の収集・整備については、対象外となっている。

欧州における相互運用性確保のためのプロジェクト ISA(Interoperability Solutions for European Public Administrations)によって策定された、コアボキャブラリ整備のための運用プロセス・ルール¹を示す（図中の日本語は追記したものである）。

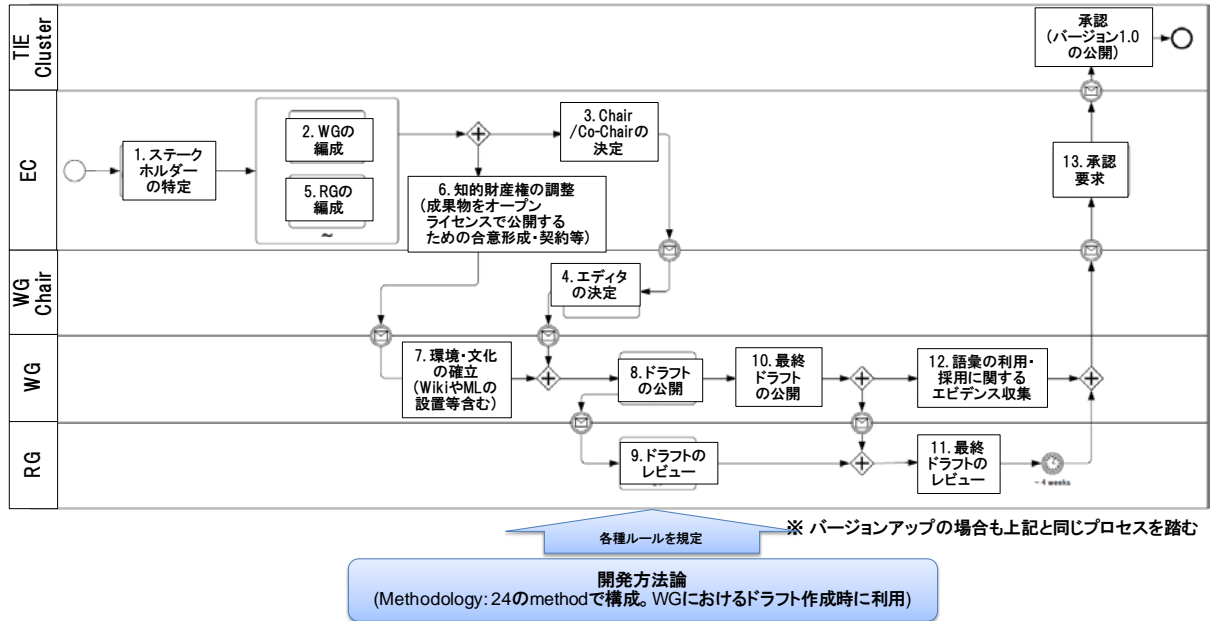


図 7.1.2.2 ISA の運用プロセス・ルール

¹ “PROCESS AND METHODOLOGY FOR CORE VOCABULARIES”

https://joinup.ec.europa.eu/catalogue/asset_release/process-and-methodology-developing-core-vocabularies

7.1.3 Schema.org の運用体制

(1) Schema.org の運用組織体制

Schema.org (Web ページの内容への意味付け作業を単純化する共通のマークアップ語彙)に関する組織体制¹を以下に示す。

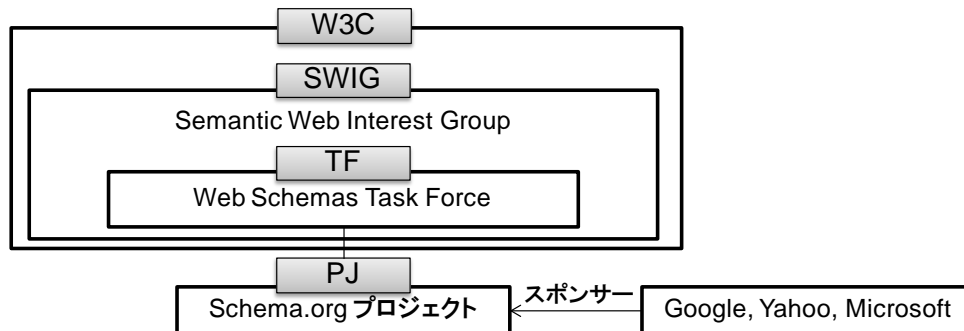


図 7.1.3.1 Schema.org 運用組織体制

表 7.1.3.1 Schema.org 運用組織名称とその役割

項番	組織名称	役割
1	SWIG: W3C Semantic Web Interest Group	セマンティックウェブ技術(RDF, OWL, SPARQL 他)の開発者と利用者をサポートするためのフォーラムとして作られたグループ。チェアは Dan Brickley (Google)。
2	Web Schemas Task Force	SWIG に設置されたタスクフォース。語彙(Schema.org 以外も対象)やマッピング、構文中立な語彙デザイン・ツールにフォーカス。チェアは R.V.Guha(Google)。Schema.org にとって、他のコミュニティと繋がるための公開フィードバックポイントとして位置づけられる。メーリングリスト(public-vocabs@w3.org)やプロポーザルリスト(http://www.w3.org/wiki/WebSchemas/SchemaDotOrgProposals)、issue tracker(http://www.w3.org/2011/webschema/track/)等を備えている。
3	Schema.org プ ロジェクト	Web サイトマークアップ語彙のプロジェクト。Google, Yahoo, Microsoft をスポンサーとする。

¹ “Web Schemas TF and Schema.org“: <http://www.w3.org/2001/sw/interest/schema.org-collab.html>
 “Semantic Web Interest Group - Web Schemas Task Force”:
<http://www.w3.org/2001/sw/interest/webschema.html>

(2) Schema.org の運用プロセス

Schema.org の運用プロセスの特長は、ボトムアップによる語彙の収集・整備プロセスとなっている点である。

Schema.org (Web ページの内容への意味付け作業を単純化する共通のマークアップ語彙)に関する運用プロセス・ルール¹を以下に示す。

Schema.org では、語彙の提案内容についての提案においては、Wiki や GitHub を活用し、提案者との情報共有を行っている。

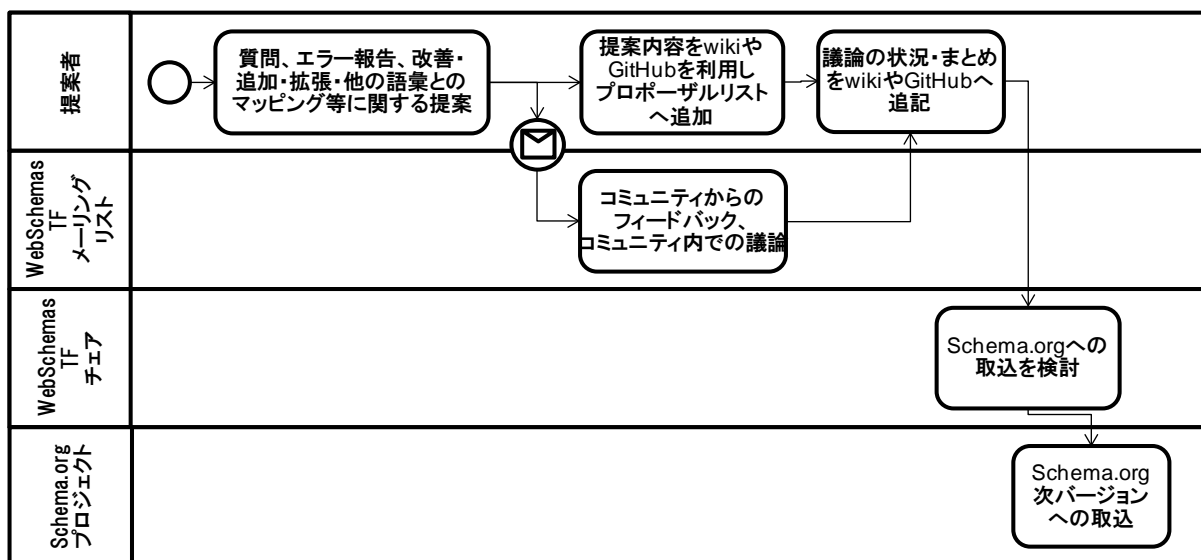


図 7.1.3.2 Schema.org の運用プロセス・ルール

¹ “SchemaDotOrgProcess - W3C Wiki” <http://www.w3.org/wiki/SchemaDotOrgProcess>

7.1.4 先行事例における運用体制・運用プロセスの比較

7.1.1 から 7.1.4 の調査結果に基づいて、各種先行事例の運用組織体制および運用プロセスについて比較した。

(1) 運用組織体制に関する比較

各種先行事例の運用組織体制に関する比較結果について、以下に示す。

表 7.1.4.1 運用組織体制の比較

	NIEM	ISA	Schema.org
概要	米国行政機関間での情報交換に用いる語彙とフレームワーク	欧州内行政機関間の相互運用性向上のための語彙やプロセス等を整備	web ページの内容を検索エンジンに伝えるための語彙
語彙の種類	コア語彙とドメイン語彙(2層構造)	コア語彙のみ	コア語彙のみ
意思決定機関	ESC	EC	Schema.org PJ (Google, Yahoo, Microsoft, Yandex を含む)
実行組織	NIEM PMO		
技術委員会	NTAC	WG, RG	W3C SWIG の Web Schemas TF
業務委員会	NBAC		
普及委員会	NC&OC	ISA joinup	Schema.org PJ (Google, Yahoo, Microsoft, Yandex を含む)
語彙最終決定	NBAC	EG	
ドメイン語彙	統治委員会と COI で運用	なし	なし

(2) 運用プロセスに関する比較

各種先行事例の運用プロセスに関する比較結果について、以下に示す。

表 7.1.4.2 運用プロセスの比較

	NIEM	ISA	Schema.org
概要	米国行政機関間での情報交換に用いる語彙とフレームワーク	欧州内行政機関間の相互運用性向上のための語彙やプロセス等を整備	web ページの内容を検索エンジンに伝えるための語彙
語彙の初期整備プロセス	-	コア語彙の開発プロセスを文書化して公開	Wiki やメーリングリストを利用したコミュニティベースの議論・合意形成
語彙の更新プロセス	語彙の更新プロセスを文書化して公開	-	Wiki やメーリングリストを利用したコミュニティベースの議論・合意形成
語彙設計の規則・方法論	命名規則や設計ルールを文書化して公開	命名規則や設計の方法論を文書化して公開	ガイドラインのみ
国際標準化関連	NIEM 用の UML プロファイルを OMG 仕様として公開	Location (場所) 語彙を W3C namespace 文書として公開	W3C SWIG と関係はあるが、W3C 名義での語彙公開は無し
語彙の変更頻度	普遍性の高い語彙を取り扱っているため変更頻度は低い	普遍性の高い語彙を取り扱っているため変更頻度は低い	修正要求や追加要求が多いため、変更頻度は高い
類型	トップダウン型 (委員会ベース)	トップダウン型 (委員会ベース)	ボトムアップ型 (コミュニティベース)
相互運用性	厳密に確保	厳密に確保	寛容

7.2 共通語彙基盤の運用体制の提案

「表 7.1.5.1 運用組織体制の比較」の「語彙の種類」に示すとおり、NIEM では、語彙のコア語彙、それ以外の語彙をドメイン語彙とした 2 層構造を採用しており、この語彙の構造に適切な運用体制及び運用プロセスが定義されている。また、ISA は、コア語彙のみの構造を採用しており、この語彙の構造に適切な運用体制及び運用プロセスが定義されている。

一方、本事業では、今後、共通語彙基盤がドメイン語彙も含めて拡大した場合においても、コア語彙及びドメイン語彙の肥大化を防ぎつつ、複数のドメインを跨いだデータの相互運用性が確保することを目的として、ドメイン語彙をドメイン共通語彙とドメイン固有語彙に再定義する 3 層構造（コア語彙、ドメイン共通語彙、ドメイン固有語彙）を採用することとした。（3 層構造の詳細は 3.3.1 節に記載の通り。）

そのため、共通語彙基盤の実運用化に向けては、NIEM や ISA との語彙の構造の違いである 3 層構造に適した運用体制を整備することで、語彙の収集・整備対象との整合性を確保することが必要である。

また、NIEM では各ドメインの運用を取りまとめる省庁が明確に割り当てられていることに対して、本事業で試行的に選定したドメインにおいては、1 ドメインに対して複数の省庁・組織が密接に関係し、取りまとめが困難となるケースが発生することも考えられる。例えば、2020 年の東京オリンピックや、地方自治体の観光客誘致のための施策のひとつとして、オープンデータの利活用が想定され、オープンデータ整備に向けたドメイン語彙は、防災ドメインや観光ドメインのような世の中のニーズの高い分野に対して、スピードをあげて整備を推進することが必要となってくる。その際、例えば、防災ドメインにおける避難所に関する語彙の整備においては、施設の情報を所有している組織や、設備を整備している組織、防災担当組織等、複数の組織が関係しているため、語彙の収集・整備に時間を要してしまう。

そのため、各ドメイン内での小規模修正が発生した場合においても、柔軟に対応できる体制の整備が必要となる。そこで、コア語彙に関係する運用については、NIEM や ISA などの海外先進事例を参考に、各委員会や WG などを含む具体的な概念モデルとしての運用体制を整備し、ドメイン語彙に関係する運用については、各ドメインをサブドメインに細分化し、サブドメイン WG を設置するとともに、複数のサブドメイン WG を取りまとめるドメイン語彙 WG を設置する運用体制を提案する。

以下に、共通語彙基盤の運用体制案を示す。

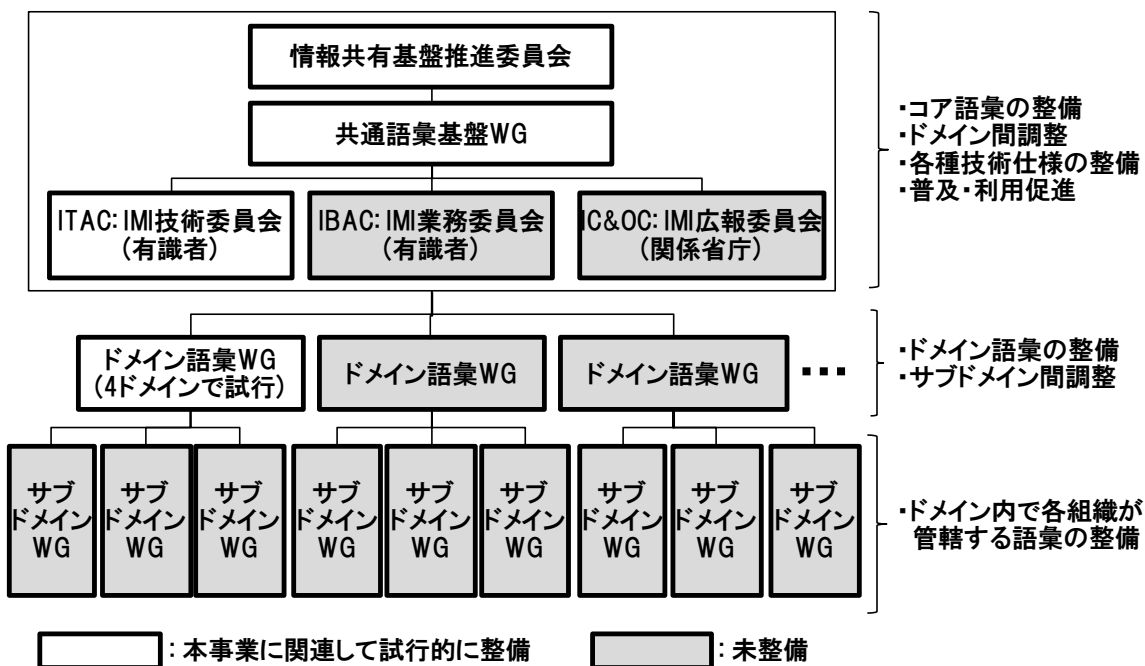


図 7.1.4.1 共通語彙基盤の運用体制案

情報共有基盤推進委員会¹は、オープンデータ等の官民での円滑な情報交換と利活用推進に必要なとなる、情報連携用語彙および文字情報基盤の導入に向けた検討を行うための委員会であり、共通語彙基盤に関する意思決定機関として現時点で最もふさわしいものと想定した。

共通語彙基盤WG²は、情報共有基盤推進委員会の下部に位置するワーキンググループであり、語彙データベースの実現、APIの整備、ツール群の整備が実施項目として挙げられるものである。

ITAC(IMI Technical Architecture Committee: IMI技術委員会)は、共通語彙基盤のコア語彙およびドメイン語彙の開発及び実装において、技術的・構造的詳細に取り組む委員会である。

IBAC(IMI Business Architecture Committee: IMI業務委員会)は、IMIコア語彙の開発や、コア語彙とドメイン語彙の調整、バージョンアップ、及び、実装のガイドや業務アーキテクチャの観点からの運用プロセスのガイドを行う委員会である。

IC&OC(IMI Communication and Outreach Committee: IMI広報委員会)は、共通語彙基盤に関する情報が、重要な意思決定者、各政府機関高官、COI(利害関係機関)、その外の参加者に対して、一貫して、かつ、効果的であるように取り組む組織である。

ITAC、IBAC、IC&OCは、共通語彙基盤WGの下部に属し、共通語彙基盤WGでの検討方針や議論結果に従って、それぞれの役割を担う。

¹ “第1回情報共有基盤推進委員会を開催します”、

<http://www.meti.go.jp/press/2013/09/20130924001/20130924001.html>

² “第1回 共通語彙基盤ワーキンググループ”、<http://goikiban.ipa.go.jp/node90>

ドメイン語彙 WG はドメイン語彙の整備およびサブドメイン間の調整を行う。本事業では、試行的に選定した 4 つのドメインである地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインについてドメイン語彙検討会という形で当該ワーキンググループを試験的に整備した。また、サブドメイン WG は、ドメイン内で各組織が管轄する語彙の整備を行うワーキンググループである。

7.3 共通語彙基盤の運用プロセスの提案

共通語彙基盤の実運用化に向けて、「7.2 共通語彙基盤の運用体制の提案」に示した共通語彙基盤の運用体制案に即した運用プロセスを提案する。

7.3.1 共通語彙基盤の運用プロセスのポリシー

今後のコア語彙の拡充、ドメイン語彙の整備を考慮すると、コア語彙や各ドメイン語彙についての語彙データの運用(提案・承認・登録・更新・失効等)を円滑に行えるよう、運用プロセスの整備にあたっては、運用ポリシーの設定が必要と考える。

運用ポリシーは、「表 7.1.5.2 運用プロセスの比較」の「語彙の変更頻度」、「類型」を基に、コア語彙やドメイン語彙のように修正頻度が低く、普遍性の高い語彙を運用する場合は、NIEM や ISA の運用プロセスに習い、委員会ベースのトップダウン型整備ポリシーを適用し、修正要求や追加要求が多いと想定されるサブドメインについては、ドメイン内で各組織が管轄する語彙を整備することから、Schema.org の運用プロセスに習い、コミュニティベースのボトムアップ型整備ポリシーを適用することを提案する。

以下に、共通語彙基盤の運用プロセスのポリシー案を示す。

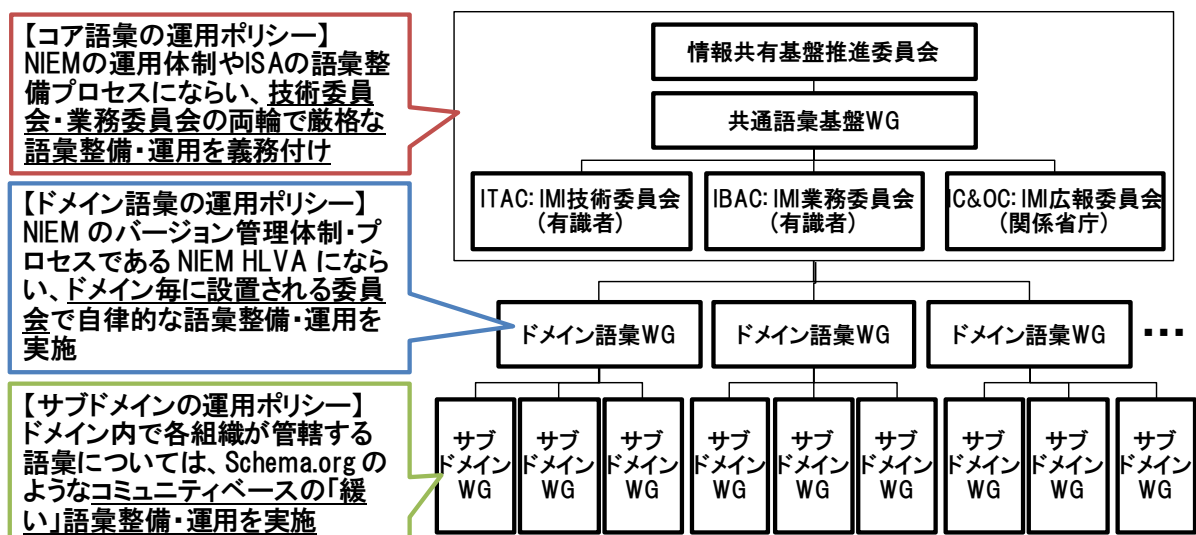


図 7.3.1.1 共通語彙基盤の運用プロセスのポリシー案

7.3.2 共通語彙基盤の運用プロセス案

「図 7.4.1.1 共通語彙基盤の運用プロセスのポリシー案」を基に、ESC、IMI PMO、IBAC、ITAC の運用プロセスのうち、コア語彙の整備プロセスについては、「図 7.1.3.2 ISA の運用プロセス・ルール」の運用プロセスを踏襲することとし、その後のプロセスについては、「表 7.1.1.3 NIEM HLVA のステップの概要」を踏襲した。

また、ドメイン語彙検討 WG での運用プロセス、サブドメインの運用プロセス群は、「図 5.2.2.1 ドメイン語彙の収集・整備プロセスの仮説」を踏襲した。

さらに、サブドメインについては、ドメイン内で各組織が管轄する語彙を整備することから、Schema.org の運用プロセスに習い、コミュニティベースのボトムアップ型整備ポリシーを適用するため、「図 7.1.4.2 Schema.org の運用プロセス・ルール」および、語彙データベース検討会での委員からのボトムアップ型整備ポリシーについての提言の際に、案として意見があった Wiki・ML・リポジトリ等を用いた語彙登録・修正希望の受付をプロセスに追加した。

共通語彙基盤の運用プロセス案を以下に示す。

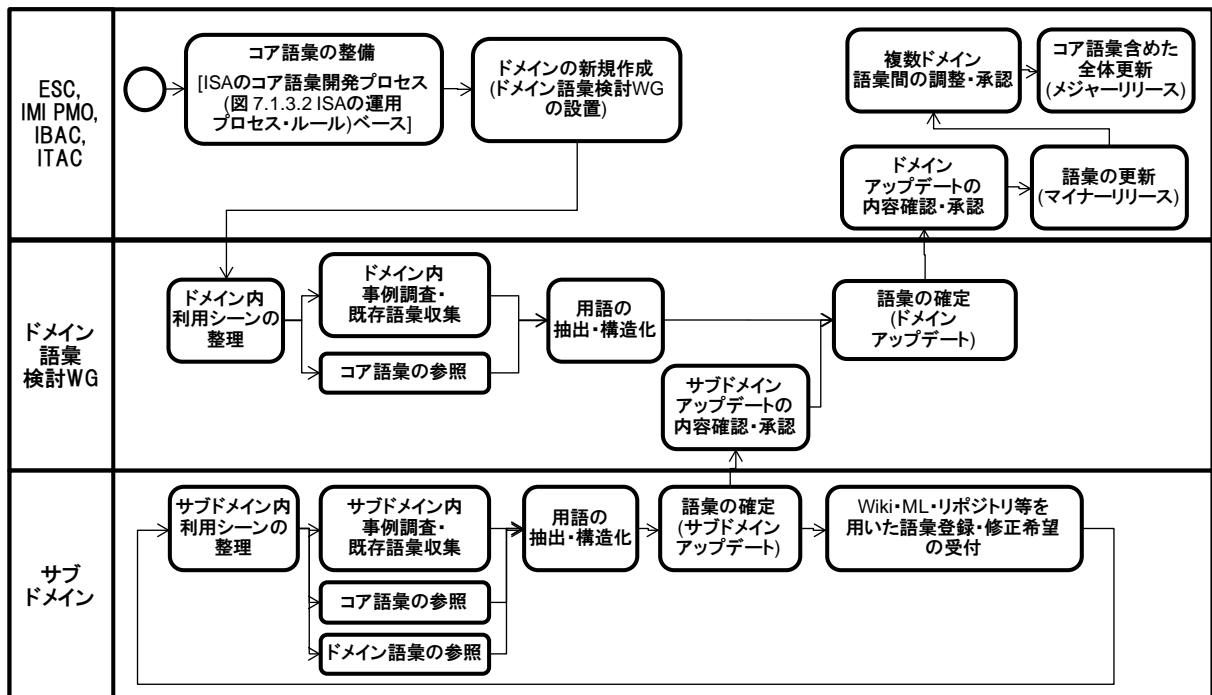


図 7.3.2.1 共通語彙基盤の運用プロセス案

7.3.3 海外関連機関との将来的な相互運用プロセス確立に向けた提案

運用プロセスを整備するにあたっては、国際的な相互運用性を確保するため、海外政府機関・民間機関・各種標準化機関との将来的な相互運用プロセスの確立に向けた取り組みについての検討が必要である。

米国 NIEM と欧州 ISA の間では、定期的なミーティングが開催され、お互いの語彙のマッピングに関する検討などを実施する協力体制が築かれ始めている。

さらに、標準化団体 OMG の Government Information Sharing and Services Domain Task Force (GovDTF) では、米国 NIEM や欧州 ISA などの情報共有基盤に関する概念モデルを整備しようとする Federated International Information Sharing Model が、将来的な標準化プロセスの候補として検討されている。

日本の共通語彙基盤がこれらの取組みに参画することで、国際的な相互運用性の向上や、国家間の情報連携、国をまたいだオープンデータの活用など、多岐に渡るメリットを得られるものと考えられる。

2014 年 11 月に開催予定のワークショップ(EC ISA 主催)にて、EC ISA、NIEM、UN/CEFACT、UBL の担当者による、各語彙間のマッピングに関する議論が予定されており、日本も参加する予定である。

日本としても、今後、W3C や OMG 等、各種標準化団体の活動について、特に、共通語彙基盤と関係の深い活動¹について、動向を継続的に調査するとともに、参画の必要性の検討やその時期、参画に必要なリソース等を具体化・明確化していくことを提案する。

また、国際的な相互運用プロセスが決定した場合、「図 7.4.2.1 共通語彙基盤の運用プロセス案」に追加する必要がある。

以下に、海外関連機関との将来的な相互運用プロセス確立に向けた取り組みイメージを示す。

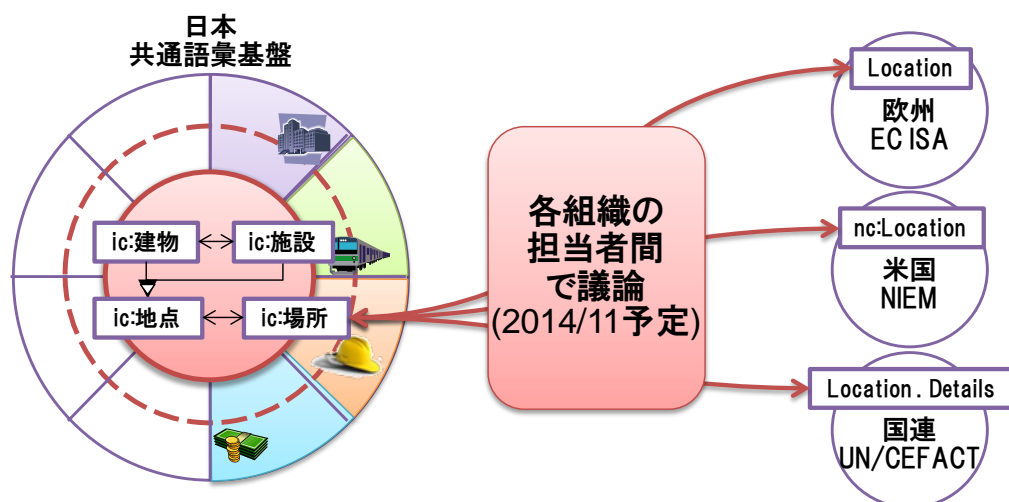


図 7.3.3.1 海外関連機関との将来的な相互運用プロセス確立に向けた取り組みイメージ

¹ W3C: Data on the Web Best Practices WG, CSV on the Web WG, Data Shapes WG

OMG: SIMF Architecture, UML Operational Threat & Risk Model

8. 共通語彙基盤の実用化に向けた提言

共通語彙基盤の利用を促進するためには、より一層の語彙の収集・整備、語彙を活用したデータテンプレートの設計、データテンプレートに基づく実データの入力、その利用・活用といった一連の工程が滞りなく実施できることが重要である。

本章では、共通語彙基盤の実用化に向けて、本事業およびツールプロジェクトによる実証結果を通じて明らかになった課題や、その解決に向けた検討方針案についての提言について、①語彙の収集・整備、②データテンプレートの設計、③データテンプレートに基づく実データの入力、④実データの利用・活用の4つのテーマに沿って示す。

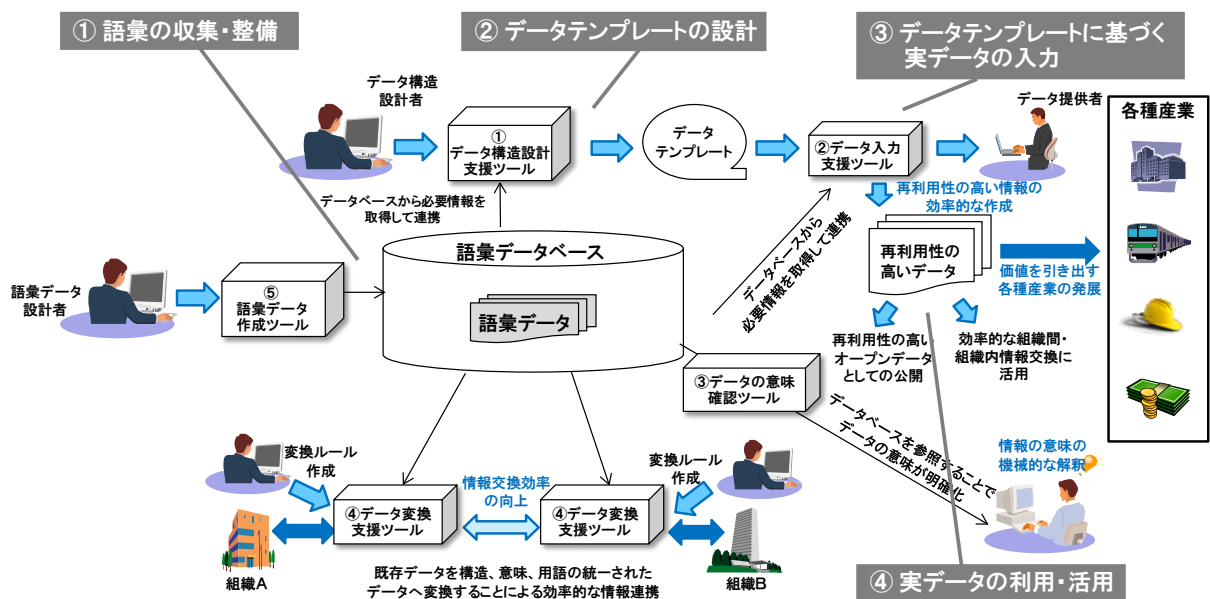


図 7.3.3.1 共通語彙基盤の実用化に向けた提言の対象

8.1 「①語彙の収集・整備」における課題と提言

共通語彙基盤を利用して必要なデータを整備するまでの一連の工程について、本事業及びツールプロジェクトを通して得られた知見より、「①語彙の収集・整備」における提言を示す。

8.1.1 コア語彙の検討プロセスの可視化

ツールプロジェクトによる実証では、試行的運用に向けてドメイン共通語彙を定義する際、本来コア語彙として定義されるべき用語（例：「道路型」）が、コア語彙に存在しないケースが発生していた。多くのドメインで参照される概念であるにもかかわらず、コア語彙に含まれていない場合には、ドメインからコア語彙へ用語の追加要求を挙げることが想定される。こうした結果を踏まえ、コア語彙の検討・整備に関しては、レビューや増補改訂、パブリックコメントの機会を適切に設けることが望ましいとの提言が得られた。

そこで、将来的な共通語彙基盤の推進に向けては、ドメイン共通語彙の内、他のドメインでも広く参照・利用されるような用語をコア語彙へ統合していく作業などを含め、コア語彙の検討プロセスを詳細化することが望ましいと考える。さらには、有識者による専門的な検討の場に加えて、中央官庁や自治体を中心とする各種のステークホルダーや、オープンデータの利用者となる民間企業・一般市民を含む、広範なレビューや増補改訂、パブリックコメントの機会や、それを効率的に実施するための仕組みを設けることが望ましいと考える。

さらに、防災ドメインなど、情報源の収集や用語の選定において、特に高度なドメイン知識を必要とする語彙や、様々な分野の専門家が関わる語彙の整備については、関係者間でコンセンサスを得ながら継続的に語彙を拡張する方策・仕組みを確立することの重要性が明確となった。

これを受け、専門用語の有識者や現場担当者など、関係者による議論・意見を集約するためのドメイン語彙WGを設置することが望ましいと考えられる。

8.1.2 コードリストの将来的な整備およびメンテナンスの方針

平成 26 年 4 月 25 日付け IT 戦略本部 各府省情報化統括責任者（CIO）連絡会議にて決定の「電子行政分野におけるオープンな利用環境整備に向けたアクションプラン」においては、オープンな利用環境の実現のために必要な事項の一つとして「コードの公開の促進、コードの対応関係の整理等の活用促進を図る」としてコード運用の必要性について言及されている。将来的な語彙データベースでは、本事業で整備したコードリストおよびコードの格納のためのデータ構造を活用し、世の中に散在する各種の既存のコードリストについても格納でき、かつ利用者に対して検索・変換・意味確認などの機能(API)を提供できることが望ましい。

既存のコードリストの格納にあたっては、コードリストのメンテナンス方法と、コードリストの利用制限について留意する必要がある。前者については、コードリストの整備主体（標準化機関など）で行われるコードリストの更新を、如何にして語彙データベース側で補足・把握し、どのような周期で当該更新を語彙データベースへ反映させるかについての検討が必要と考える。後者については、オープンでない（利用に制限が発生する）コードリストについて、オープン化を促すための取組みや、一部の利用制限を保持したまま、語彙データベースに格納するための実装方針についての検討が必要と考える。

8.1.3 コミュニティベースの分散型語彙整備

日本では、複数の組織が密接に関係し、複数の組織間での調整が難航する可能性がある。そこで、本事業の将来の運用体制案として「コミュニティベースの分散型語彙整備」を提案する。

専門家や、民間団体・個人等、多様なステークホルダーが関わるサブドメインにおいては、コミュニティや個人による分散型の語彙整備プロセスにより、自由な語彙の整備が可能となる。

分散型の語彙整備プロセスの確立に向けては、Schema.org に倣い、Github のような分散開発プラットフォームを活用することや、そのために必要な語彙データベースの API を精査・明確化していくことが必要である。

以下にコミュニティベースの分散型語彙整備体制案を示す。

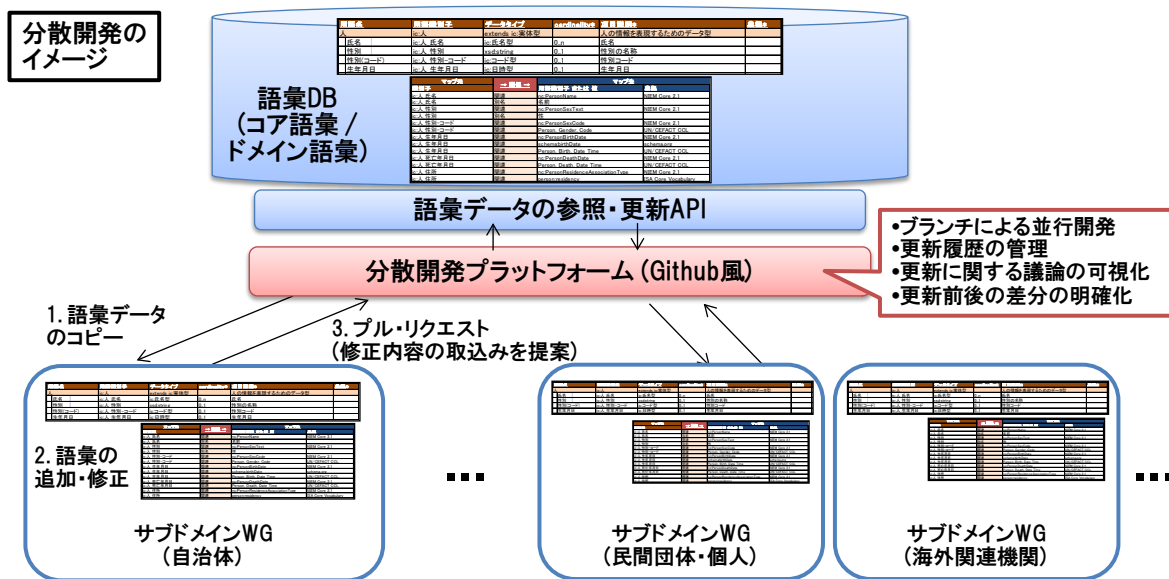


図 8.1.3.1 コミュニティベースの分散型語彙整備体制案

8.2 「②データテンプレートの設計」における提言

共通語彙基盤を利用して必要なデータを整備するまでの一連の工程について、本事業及びツールプロジェクトを通して得られた知見より、「②データテンプレートの設計」における提言を示す。

8.2.1 利用者にとっての用語の分かりやすさの向上

IMI コア語彙を利用したツールプロジェクトによる実証の結果、いくつかの用語の名称（「活動」よりも「イベント」、「市区町村」よりも「市町村」の方が一般利用者には馴染みやすい）や、住所表記の分解方式（町名、丁目、番地、号）に対する混乱（住所表示をどのように町名、丁目、番地、号に分解すればよいのか一般にあまり知られておらず、データ入力の試行において混乱が見られた。また字（あざ）の記入有無についても揺れが生じた）など、一般にあまり馴染みのない表現・概念を含む用語の存在が明確となった。

IMI コア語彙 v1 から IMI コア語彙 v2 案への更新にあたっては、「活動」を「イベント」に見直すなど、いくつかの用語については名称の見直しを実施した。今後、データテンプレート設計者にとっての分かりやすさを向上し、用語利用の容易性をさらに高めるためには、ハッカソンの開催やアプリケーション開発の支援、さらに実証実験など、利用者および開発者に、IMI コア語彙を利用してもらう機会を増やすなど、実践的かつ継続的な検証作業を組み込んだ、コア語彙の整備プロセスの確立が望ましいと考える。

8.2.2 用語の意味の補強

IMI コア語彙 v1 から IMI コア語彙 v2 案へ更新するにあたり、コア語彙に含まれる用語の意味（用語の定義や説明のための文章）の見直しを実施したものの、必要十分な記載とするには、なお修正の余地が残されている。

法的根拠を備えるような用語の場合（例：児童手当法に基づく「児童手当」）は、法律や条例などの文書を出典として、用語の意味を確定させることができるが、厳密な定義よりも分かりやすい表現が好ましい場合や、「人」のような一般的な用語の意味を正確に記述する場合には、市販の国語辞書や概念辞書を調査することも一案と考える。

辞書の出版社の著作権を損なうことなく、可能な限り既存辞書に基づく意味の補強が可能となるような実現案を検討することが望ましい。また、ISO11179-4 で規定されるデータ定義要件¹など、意味を記載する際に注意すべき・守るべき観点についても、将来的な整備が望ましいと考える。例えば、ISO11179-4 で示される以下のような要件・勧告を提示することが一案として想定される。

¹ ISO/IEC 11179-4: “Formulation of data definitions”, <http://metadata-standards.org/11179/>

表 8.2.2.1 【参考】 ISO11179-4 のデータ定義に関する要件（抜粋）

分類	内容	内容（英語）
要件 (requirement)	単数形を用いること	be stated in the singular
	概念が何であることを示すこと。(何でないかを示すだけでは不十分)	state what the concept is, not only what it is not
	記述的なフレーズや文章で示すこと	be stated as a descriptive phrase or sentence(s) (in most languages)
	略称は共通理解のあるもののみを用いること	contain only commonly understood abbreviations
	他のデータや概念を含まないように表現すること	be expressed without embedding definitions of other data or underlying concepts
勧告 (recommendation)	概念の本質的な意味を示すこと	state the essential meaning of the concept
	正確で、かつ冗長でないこと	be precise and unambiguous
	簡潔であること	be concise
	それ単体で自立したものであること	be able to stand alone
	根拠、機能的な使い方、ドメイン情報や手続き的な情報を含まないように表現すること	be expressed without embedding rationale, functional usage, domain information, or procedural information
	循環論法を避けること	avoid circular reasoning
	関連する定義と同じ用語や論理的構造を使うこと	use the same terminology and consistent logical structure for related definitions
	メタデータの種類(type)にふさわしいものであること	be appropriate for the type of metadata item being defined

8.2.3 利用シーンに特化したパッケージの整備

共通語彙基盤が普及すると、語彙データの拡張に伴い、利用者にとって各利用シーンで用いるべき用語を選択することが次第に困難となることが予想される。現に、ツールプロジェクトの実証結果では、コア語彙の利用にあたって選択の容易化が要望の一つとして挙げられているものもある。(例えば「イベント、スケジュール」という分類による検索で、「イベント」や「期間」といった用語が取得できる、等)

そこで、語彙データの利用者が語彙を容易に選択できるよう、代表的な利用シーンに特化した語彙のセットとして、目的に応じて語彙を組み合わせたパッケージ (IEP: Information Exchange Package) を整備・公開することが望ましいと考える。

平成26年1月に、経済産業省より、住所 IEP のワーキングドラフト¹が公開されている。順次、氏名や組織など代表的な概念から、IEP の整備および公開を予定している。今後、代表的な利用シーンを考慮した IEP が継続的に公開されることで、共通語彙基盤の利用がより促進されると考える。

以下に、IEP を活用した共通語彙基盤の利用イメージを示す。

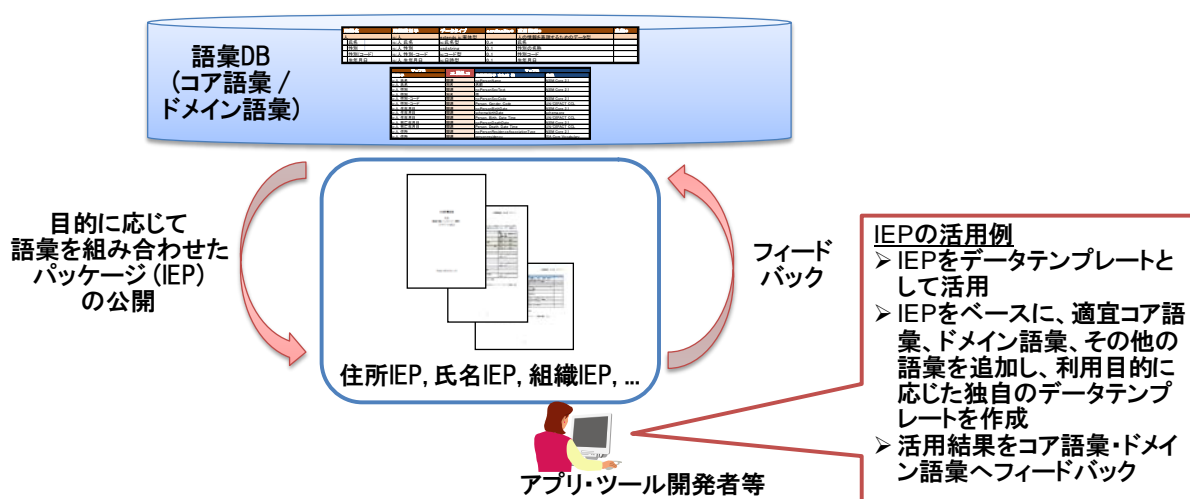


図 8.2.3.1 IEP を活用した共通語彙基盤の利用イメージ

¹ <http://datameti.go.jp/goi-draft>

8.3 「③データテンプレートに基づく実データの入力」における提言

共通語彙基盤を利用して必要なデータを整備するまでの一連の工程について、本事業及びツールプロジェクトを通して得られた知見より、「③データテンプレートに基づく実データの入力」における提言を示す。

8.3.1 データ入力者のニーズに基づく RDF 形式の継続的な検討

コア語彙検討会やツールプロジェクトの実証結果を通じて、RDF 形式に関する課題として、本事業で整備した RDF 形式の語彙データを表現するために利用している RDF Schema だけでは、用語の Cardinality（用語の出現頻度）を規定できないことが判明した。

厳密な Cardinality を求める利用シーンにおいては、RDF スキーマだけでなく OWL の owl:minCardinality や owl:maxCardinality をを用いることで対応可能である。

将来的に、利用者ニーズによっては、RDF Schema に加えて OWL を取り入れることで、RDF 形式の語彙データを拡張することも選択肢となり得る。

8.3.2 用語が取り得る値の文字セットを指定する方法について

日本においては、かな、カタカナ、漢字、ローマ字、英数字等といった、ラテンアルファベット以外の文字も多分に利用される。また、氏名や行政文書では多様な漢字も使用されているため、共通語彙基盤の実用化に向けては、それらの文字に対応した文字種別の取扱いを可能とする必要がある。

そこで、共通語彙基盤の将来的な拡張に向けて、文字列を値として取る用語については、人名、地名、ローマ字などの文字セットを明確に指定できる仕組みを設けることが望ましい。IMI コア語彙 v2 案では、文字列を扱う用語の型（データタイプ）として、既存のアプリケーションでの利用を考慮し、一般に広く使用されている xsd:string を採用した。上記のような多様な文字セットの指定を取り扱う必要がある場面においては、文字セットの制約を付与した新たな用語（例えば「テキスト」クラス用語）の導入も考慮に入れることが望ましい。

そうした新たな用語を実装する場合、XML Schema では「制限」(xsd:restriction)を用いることで文字セットの制限やデータ入力ルールの実現が可能であるが、RDF 形式でも同様の制限を課すための施策や課題の再整理などの検討も必要となる。

8.4 「④実データの利用・活用」における提言

共通語彙基盤を利用して必要なデータを整備するまでの一連の工程について、本事業及びツールプロジェクトを通して得られた知見より、「④実データの利用・活用」における提言を示す。

8.4.1 既存のサービスとの親和性の向上

座標を表現するための用語について、IMI コア語彙は「緯度」と「経度」を異なる用語として分割しているが、Google Maps 等の既存サービスでは、緯度と経度は分割せずカンマで接続するデータ形式が利用されている。

既存サービスで利用されている標準的なデータ形式や、既存サービスの API 仕様との親和性を高めるため、共通語彙の構造化についても実サービスでの検証を通じた継続的なブラッシュアップが必要であると考ええる。

8.4.2 多言語対応について

本事業では、IMI コア語彙に含まれる用語の表記、識別子、意味の説明などについては、主要な利用者が日本人であることを考慮して、日本語による記載を採用した。ツールプロジェクトの中では、既存語彙との連携性を高めるため、schema.org とのマッピングを行い、それにより定義した語彙の構造化データを既存の検索サービスで読み込むことが可能になり、IMI コア語彙の活用が期待できる利用シーンが示されている。

将来的に、利用者の国際化や、海外組織との連携が進むことを想定すると、表記の多言語化、識別子の多言語化、意味の多言語化が必要となるとともに、語彙データベースのデータ構造・API・ツール群の仕様や用法についても、多言語を用いた説明が必要になると考える。

8.4.3 用語間の関係性の詳細化

本事業では、語彙データベースのデータ構造として、用語間の関係性を SKOS 語彙で記述できるような案を策定した。さらに、IMI コア語彙と各種既存語彙(NIEM、ISA、UN/CEFACT、schema.org 等)の間に対応があるものについて、「関連」(skos:relatedMatch)という関係を用いてその対応を表現した。

今後は、共通語彙基盤と各種既存語彙との相互運用性をさらに向上させるため、これらの関係を、完全一致(skos:exactMatch)、類似(skos:closeMatch)、広義(skos:broadMatch)、狭義(skos:narrowMatch)といった、より具体的な関係に代替できないか、精査していくことが望ましいと考える。

さらに、共通語彙基盤と既存語彙との関係の新規整備だけでなく、互いの語彙のバージョンアップに伴う関係性の維持(更新)等、関係性のメンテナンスに関する組織体制およびプロセスを明確化していくことも、将来的な共通語彙基盤の推進に向けて望ましいと考える。

8.4.4 JSON形式のサポート

本事業では、語彙データの物理形式として、主に情報交換用の「XML形式」、データ公開用の「RDF形式」、人間による直接操作の「自然言語形式」をサポートした。

近年、JSON (Java Script Object Notation) 形式の派生系として、Schema.org でも積極的に採用されている W3C 勧告 JSON-LD 形式の普及が進んでいる。JSON-LD は、RDF/XML や Turtle と同様、RDF データモデルの一表現として扱うことができるため、「RDF形式」のために整備したシリアライズの方法を適用することで、JSON-LD 形式の物理スキーマを生成することが可能である。。JSON-LD 形式は JSON 由来であるため、Web ブラウザ上での取り扱いが XML や Turtle に比べて容易である。本事業にて利用シーンとして示した「データのマッシュアップ」や「メールへ埋め込まれたメタデータの活用」において特に有用な物理形式であるため、将来的にサポートすることが望ましい。

今後は、共通語彙基盤の利用・普及に向けて、JSON-LD 形式の語彙データを実際の JSON-LD データの記述に適用・検証し、JSON-LD ならではの使い方を踏まえた上で、データテンプレートの設計方針や、語彙データベースのデータ構造や API、ツール群の要求事項へフィードバックすることが望ましい。

9. まとめ

本事業では、共通語彙基盤の導入に向け、専門家や将来の共通語彙基盤の利用者として想定される関係者を交えた客観的検討、また諸外国の先進事例の調査を行った上で、以下の項目を実施することで、共通語彙基盤を支える語彙データ、語彙データベース、ツール群についての要求事項を整理した。

本事業の推進にあたっては、専門家や将来の共通語彙基盤の利用者として想定される関係者を交えた客観的検討を行うため、各種の有識者で構成されるコア語彙検討会、語彙データベース検討会、ドメイン語彙検討会を設置した。

コア語彙については、コア語彙検討会を通じて、IMI コア語彙 v1 の整備において明確となった課題について整理を行った後に、その課題を解決すべく、語彙の設計方針を整備した。また、語彙の設計方針に従い、「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した。さらに、既に定義された「用語」を使用して、さらに別の新たな「用語」を定義できるといった、柔軟性のあるデータ構造を検討し、IMI コア語彙 v2 を整備した。

データ構造については、語彙データベース検討会や、ツールプロジェクトからのフィードバック、海外聴取を基にした国際的動向を踏まえたデータ構造の一案を作成するとともに、利用シーンの洗い出しやツール群についての検討、API についての検討、セキュリティ要件の整理やスケーラビリティ要件の整理を行った。また、共通語彙基盤の導入に向け、ドメイン語彙の拡充や、コア語彙のバージョンアップ等、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い、運用体制の他、メンテナンスのプロセスや必要な検討会も含めて、実現可能性が高く持続性の期待できる方法を検討した。

ドメイン語彙については、地理空間・施設ドメイン、移動・交通ドメイン、財務ドメイン、防災ドメインの4ドメインを試行的に選定し、ドメイン語彙検討会を通じて、リストアップした用語をドメイン語彙サンプルとして整備するとともに、ツールプロジェクトと連携する語彙データベースとして、ツールプロジェクトの試行に必要なAPIを実装した上で提供するパイロットシステムを構築し、ツールプロジェクトにて準備したドメイン語彙を都度追加・更新しながら、パイロットシステムの運用を実施した。

以上の実施結果を踏まえて、各専門家や海外の先行事例での知見を活かした各種要求事項や課題に対する考察、解決策の検討、実運用に向けた提言を概念モデル書としてまとめた。

本概念モデル書が、今後の共通語彙基盤の実現・普及・発展に寄与することを期待する。