

第 3 回 情報共有基盤推進委員会 議 事 次 第

日時 平成 27 年 2 月 26 日（木） 15 : 00～17 : 00

場所 経済産業省第 6 共用会議室（本館 17 階東 3）

<議事次第>

1. 開会
2. 共通語彙基盤 WG 報告、及び推進方針に係る審議
3. 文字情報基盤 WG 報告、及び推進方針に係る審議
4. 閉会

<配布資料>

- 資料 0 議事次第
- 資料 0-1 委員名簿
- 資料 0-2 席次表
- 資料 1 政府・自治体での推進状況
- 資料 2 共通語彙基盤整備事業の進捗状況
- 資料 3 構築中のツールについて
- 資料 4 コア語彙、IEP について
- 資料 5 共通語彙基盤整備事業 2015 年度実施項目(案)
- 資料 6 文字情報基盤整備事業の進捗状況と 2015 年度実施項目(案)
- 資料 7 縮退マップについて

参考資料 0 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

参考資料 1 共通語彙基盤パンフレット

参考資料 2 文字情報基盤パンフレット

参考資料 3 IEP の例（住所 IEP）

参考資料 4 コア語彙テクニカルガイドライン

第 3 回 情報共有基盤推進委員会 委員名簿

(2015 年 2 月 26 日)

委員長

須藤 修 東京大学 大学院 情報学環・学際情報学府学環長・学府長

委員

武田 英明 共通語彙基盤ワーキンググループ 委員長

国立情報学研究所 情報学プリンシプル研究系 教授

橋田 浩一 一般社団法人電子情報技術産業協会 (JEITA) ビッグデータ工学専門委員会
委員長

東京大学大学院情報理工学系研究科ソーシャル ICT 研究センター 教授

林 史典 文字情報基盤ワーキンググループ 委員長

聖徳大学/聖徳大学短期大学部文学部文学科 教授 人文学部長/人文学部日
本文化学科長

伏見 諭 一般社団法人情報サービス産業協会 (JISA) 技術強化委員会 標準化部会長

合同会社 ソフデラ代表

(委員 50 音順)

オブザーバ

内閣官房 IT 総合戦略室

内閣官房 社会保障改革担当室

総務省 行政管理局

事務局

和田 恭 経済産業省 情報プロジェクト室 室長

宮里 孝則 経済産業省 補佐情報プロジェクト室 室長補佐

平本 健二 経済産業省 CIO 補佐官

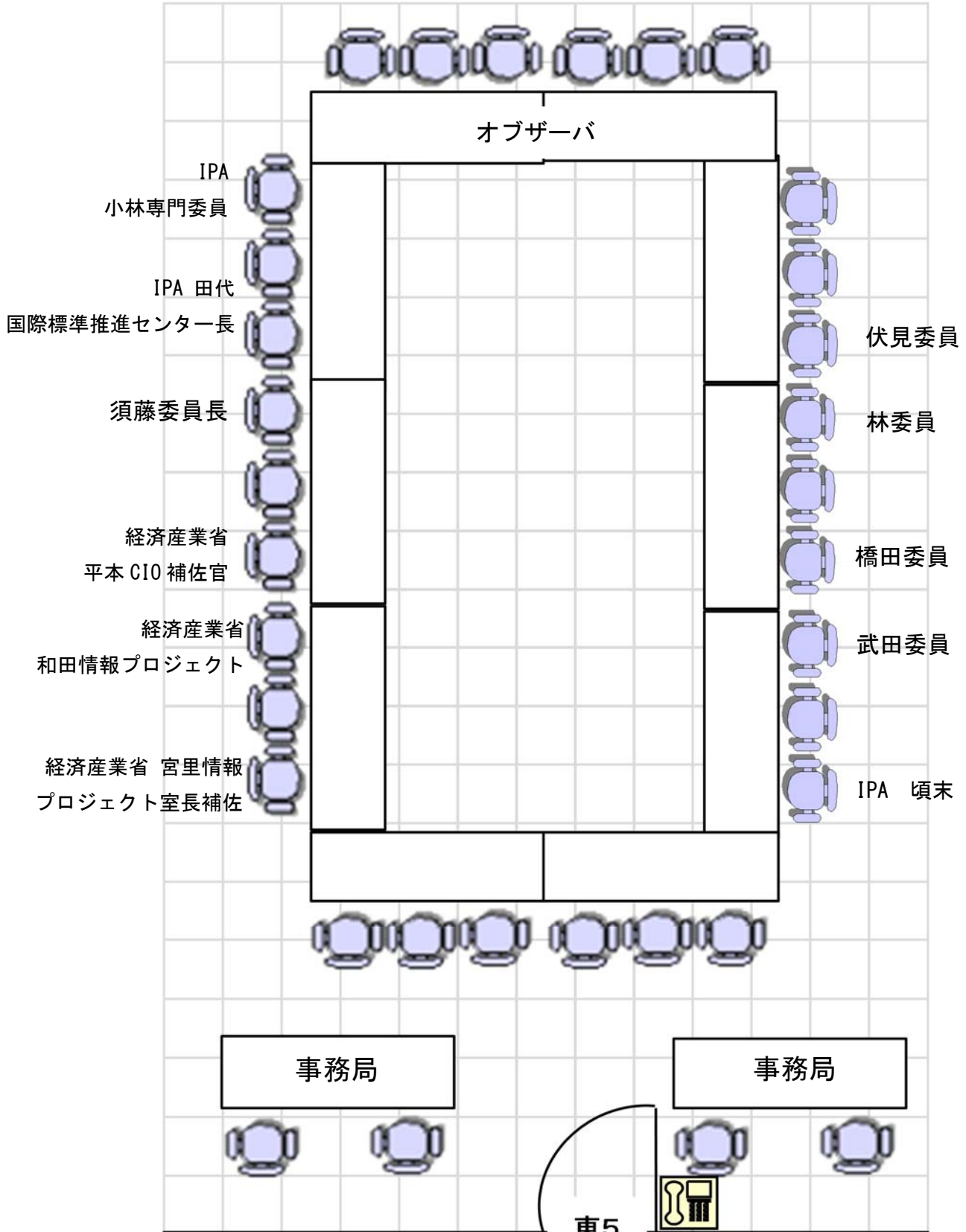
田代 秀一 独立行政法人情報処理推進機構 国際標準推進センター長

第 3 回情報共有基盤推進委員会席次表

日時：2015 年 2 月 26 日（木）15:00

場所：経済産業省第 6 共用会議室（本館 17 階東 3）

内閣官房
IT 総合戦略室



- 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（内閣官房）
 - H26の計画であるコア語彙と住所等のIEPを公表済。
- 経済産業省
 - 法人情報に関する情報を、共通語彙基盤の観点から整理中
- 中小企業庁
 - 官公需情報提供等に関し、共通語彙基盤の観点から支援
- 公共クラウド
 - 観光情報収集で、共通語彙基盤を参照
- 国税庁
 - 法人番号関連情報提供に関し、共通語彙基盤の観点から支援
- 農水省
 - 農業情報標準化を、共通語彙基盤の観点から支援
- 横浜市金沢区
 - 共通語彙基盤を使った、アプリコンテストを実施
 - ・ 防災、子育て、公園等の関連アプリが受賞
- 千葉市
 - 千葉市オープンデータの推進に関する指針
 - ・ 「氏名や住所等の普遍的用語の記述については、国で整備を進めている情報連携用語彙データベース等に準拠することとする。」

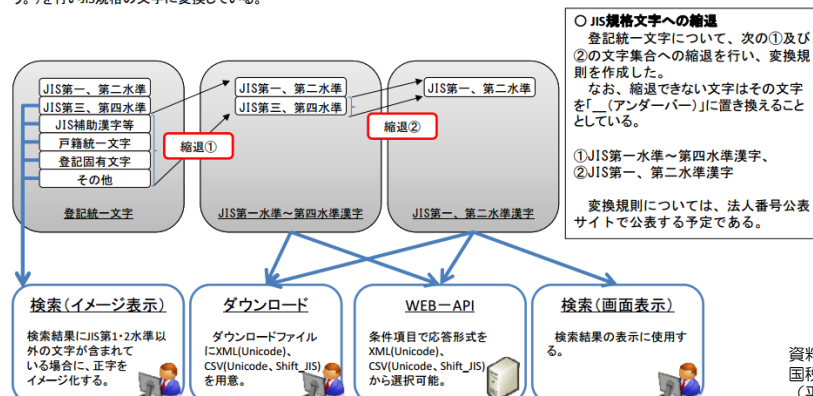
文字情報基盤の政府・自治体での推進状況 1

- 世界最先端IT国家創造宣言、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（内閣官房）
 - 創造宣言およびアクションプランに従い、文字情報基盤の相談が増加
 - 「文字情報基盤導入＝全6万字活用」と誤解されているため、アクションプラン等の表記内容の変更が必要。縮退マップの提供とあわせて実施する必要有り。

■ 国税庁

- 法人番号の公開に伴い、縮退を実施。（IPA縮退マップを活用予定）

設立登記法人の情報提供元である登記情報システム（法務省管轄）は、登記統一文字を利用しているが、これには55,000字ものJIS規格外の文字を保有している。
 そのため、法人番号システムでは、登記統一文字からJIS規格文字への縮退（縮退とは、例えば、「高」⇒「髙」へ置き換えることをい）を行いJIS規格の文字に変換している。



資料：
 国税庁「法人番号の公表機能に係る仕様」（平成26年11月版）

■ マイナンバー

- 通知カード、個人番号カード、マイポータルでの表記、表示、保存方法の検討が必要であり、関係者に情報提供中。

■ 戸籍制度に関する研究会（法務省）

- 第一回研究会（2014/10/29）「戸籍制度に関する検討課題」
 - 「各ベンダーの戸籍情報システムによって使用している文字テーブルを戸籍統一文字に変換するための作業は行っている。ただ、各市町村の文字を完全に戸籍統一文字へ変換できているわけではなく、相当程度の外字が残っている状態である。」
- 第二回研究会（12/3）「番号制度導入について」
 - 現行制度での戸籍電子化の概要説明。正字にも言及。
- 第三回委員会（12/18）「事務処理上の問題点とシステム一元化の是非等」
 - 自治体毎に外字がありネットワーク化が難しいことが指摘される
- 第四回研究会（1/29）で、「システム一元化に伴う制度の見直しの要否」
 - マイポータルによる手続に関して議論。表示する時の文字の問題はあるが、議論なし。

■ 文科省

- 第8回文化審議会国語分科会漢字小委員会（2014/5/23）で「手書き文字の字形」と「印刷文字の字形」に関する検討を開始。
- 第14回文化審議会国語分科会漢字小委員会（2015/1/16）
 - 「「手書き文字の字形」と「印刷文字の字形」に関する指針」の構成例について(素案)の議論
 - 常用漢字表における「字体・書体・字形」等の考え方について(共通理解のための素案)の議論

■ 自治体

- 調達仕様に記載が進む等、自治体の導入意欲は高い
 - ベンダが導入に難色を示しているという意見が未だに寄せられる

共通語彙基盤整備事業の進捗状況

1

IMI

意味

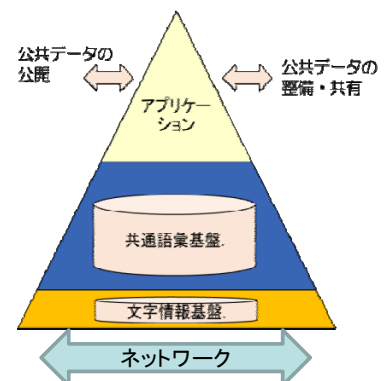
IPA

• IMI: Infrastructure for Multilayer Interoperability (共通語彙基盤)

- 情報システムの「各層」(電気信号→データ(文字→単語)→アプリケーション)がきちんと繋がるための「基盤」を提供する。

そのために、情報機器が交換するデータの中で用いられる「用語」について、

意味の共有
を図ることが重要。



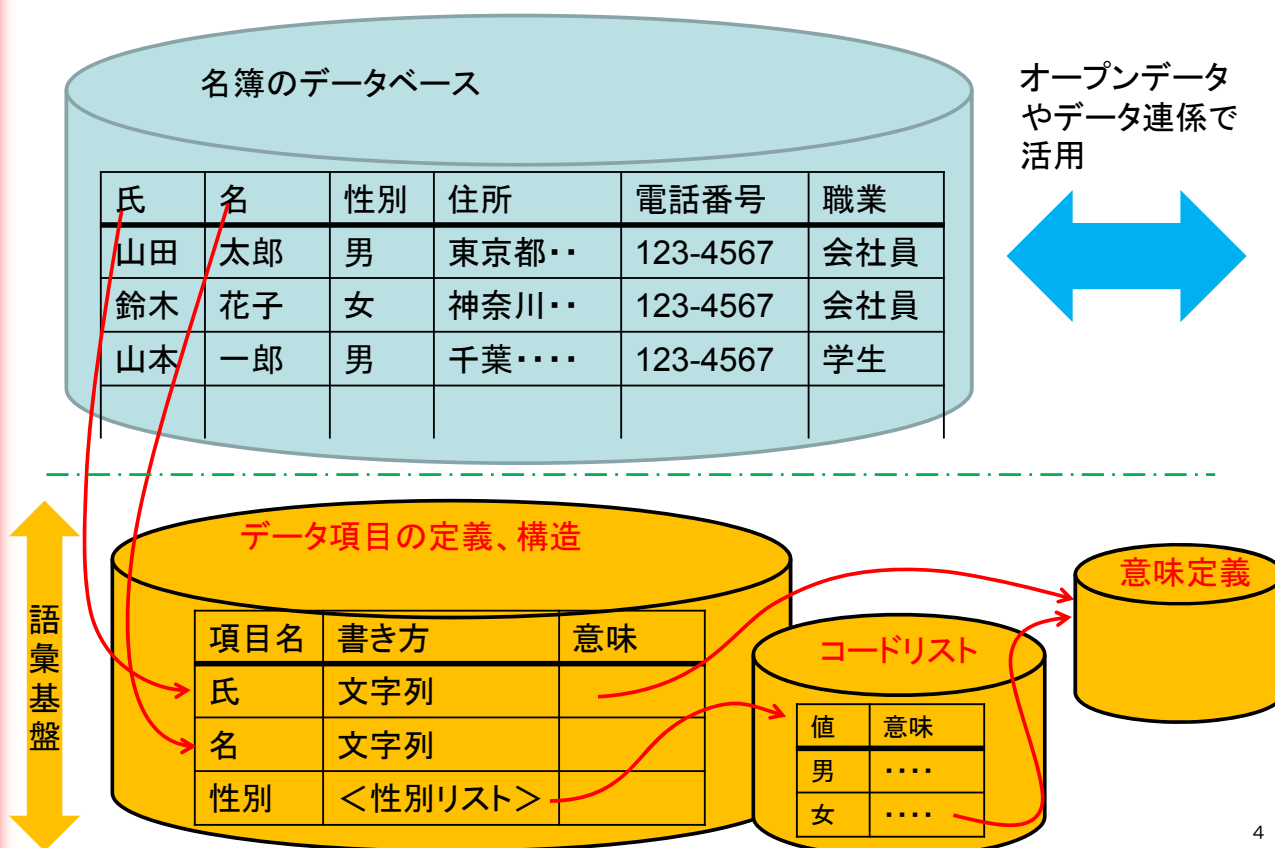
2

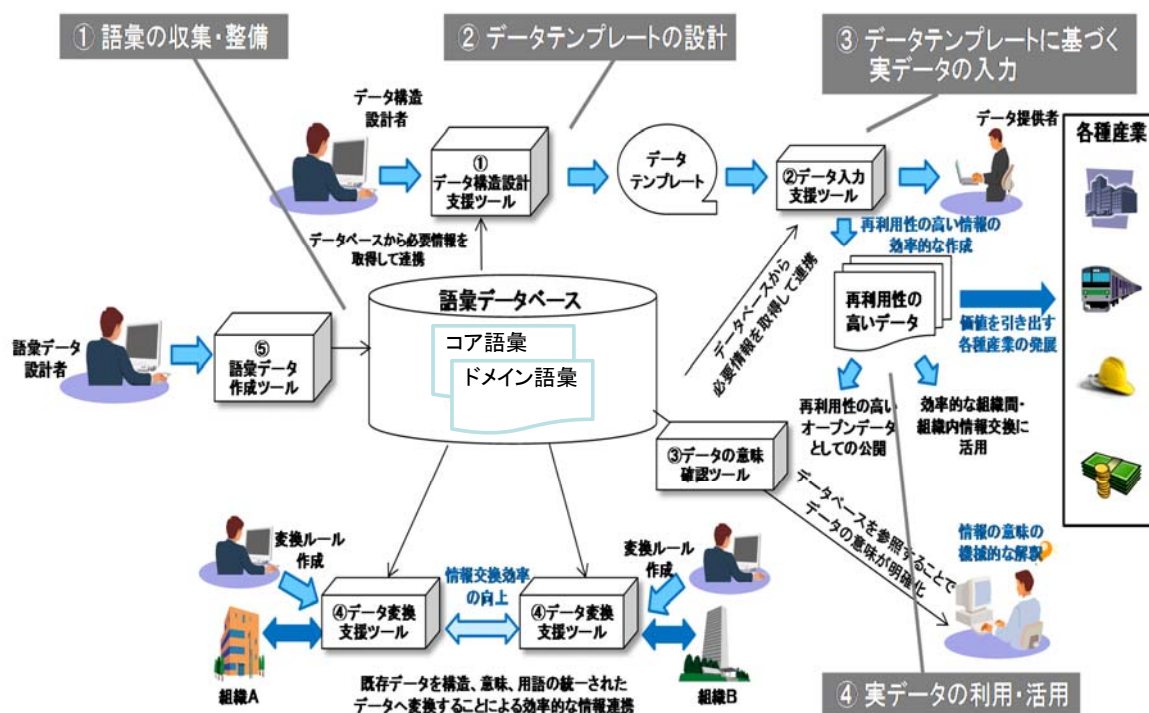
- 交換されるデータの基本的構造を
一致させる／整合させる
- データの中に出てくる「用語」を
統一する／関係付ける

ために

- データの**構造**定義
- データ**項目名**の定義
- データ構築を支援する**ツール**類
等を提供するためのプロジェクト

例) 名簿作成に「語彙基盤」を使う。。



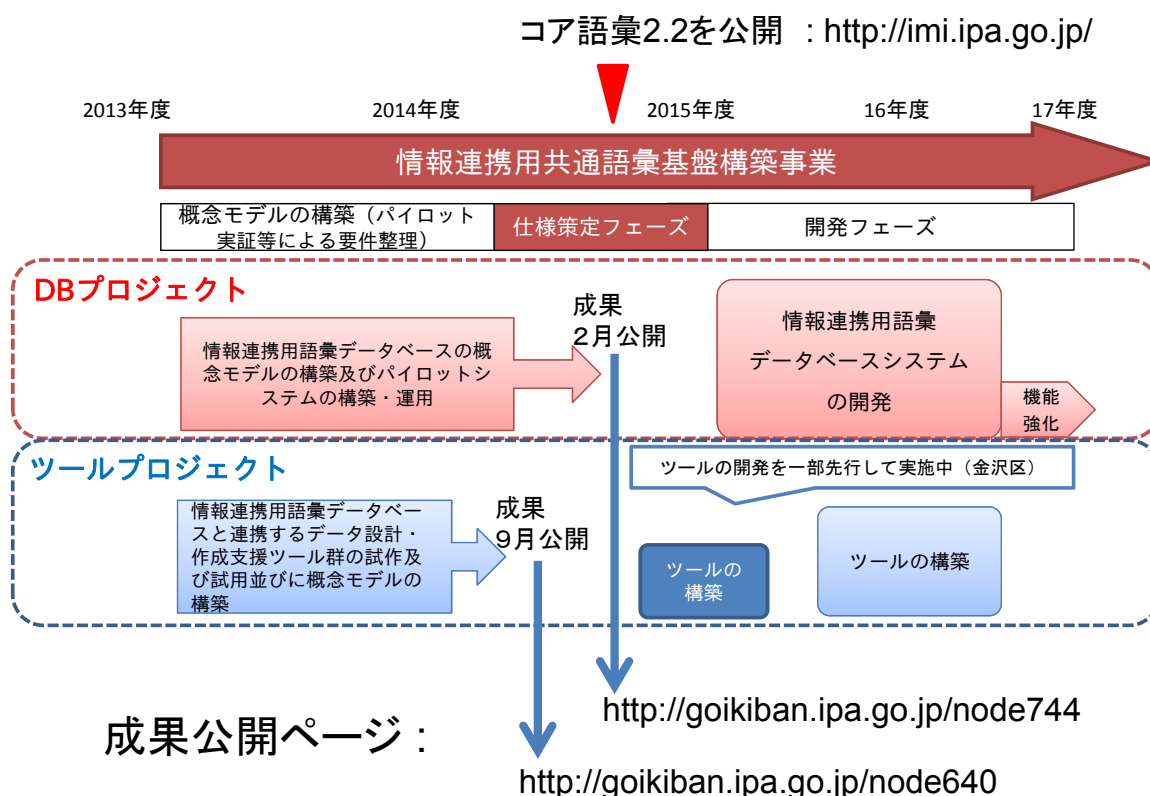


- 行政サービスの向上
 - 透明化
 - 住民参加
 - 防災
- 行政の効率化
 - システムの効率化
 - 業務の効率化
- ビジネス活用
 - データを使う新しいビジネス
 - データ／アプリを作る新しいビジネス

- 「語彙」の整備
- 「語彙」を提供する環境の整備
- 「語彙」を構築するための環境の整備
- 「語彙」を活用するためのツール類の整備



共通語彙基盤事業進捗状況と計画の概要



- 海外先進事例調査
 - 国際的な相互運用も視野に入れ、先行事例である以下の組織に対し、直接の聴取も含めた調査を実施

- ✓ NIEM(米国)
- ✓ ISA(欧州)
- ✓ schema.org

運用組織体制と運用プロセスの比較

		NIEM	ISA	Schema.org
概要		半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とフレームワーク等を整備	webページの内容を検索エンジンに伝えるための語彙
語彙の初版登録プロセス		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
語彙の更新プロセス	編集意思決定機関 実行組織	ESC NIFM PMO	EC	Schema.org PJ (Google, Yahoo, Microsoft, Yandexなど)
語彙設計の規則・方法論	命名 化して 至極委員会	NTAC	WG, RG	W3C SWIG の Web Schemas TF
国際標準化関連	NIM OMG	各委員会 NC&OC	ISA joinup	Schema.org PJ (Google, Yahoo, Microsoft, Yandexなど)
語彙の更新頻度	普通 新しいドメイン語彙	NBAC	なし	なし
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	

ツールの整備

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。

項番	ツール名	概要
1	データモデリングツール	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
2	データ構造設計支援ツール	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
3	データ構造変換ツール	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
4	データ入力支援ツール	「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
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
6	語彙データ閲覧ツール	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
7	妥当性検証ツール	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
8	語彙データ管理ツール	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
9	共通語彙リポジトリ	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
10	語彙データ提案ツール	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

4 ドメインについての思考実験

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となるデータを扱うドメイン

ドメイン語彙のサンプル

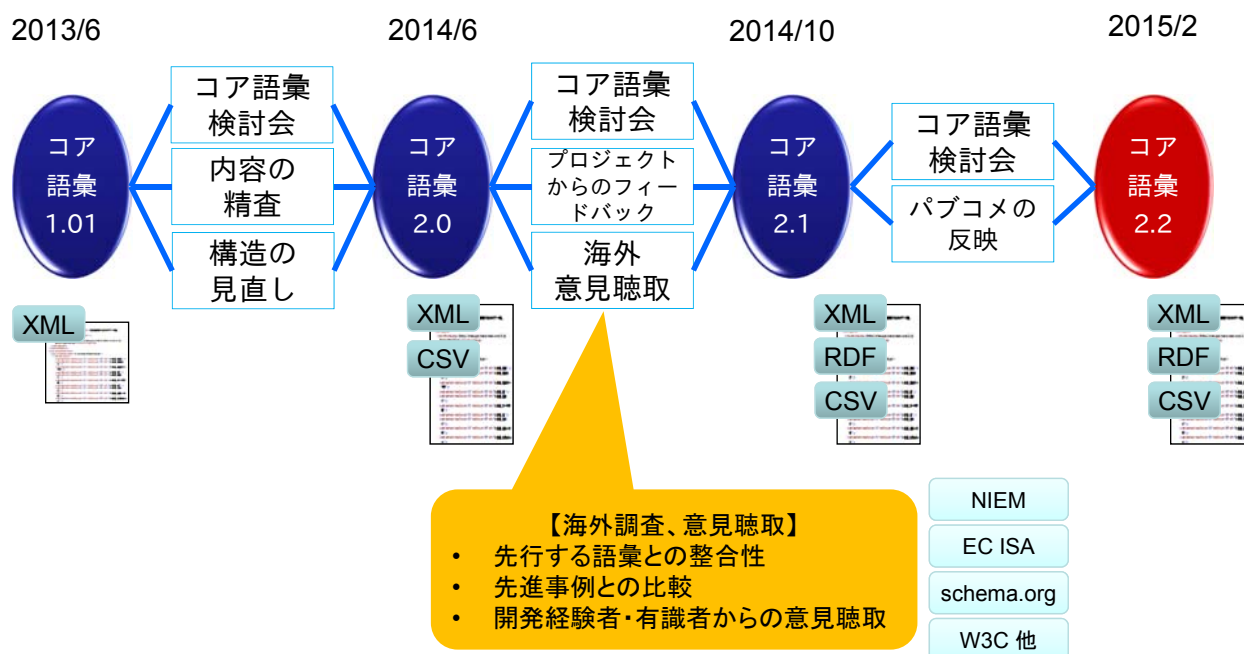
地理空間・施設		移動交通		防災		財務(歳出予算、税ど)	
用語の名称	用語の名称	用語の名称	用語の名称	用語の名称	用語の名称	用語の名称	用語の名称
場所	駅	災害対応施設	災害対応施設	費目	地方議会	名称	地方議会
線	管轄事業者	種別	種別	説明	地方議会	説明	地方議会
領域	開業年月日	対象地区	対象地区	金額	地方行政	金額	地方行政
座標列	一日平均乗車人数	学区	学区	議会費	総務全般	議会費	総務全般
座標	一日平均降車人数	管理者	管理者	議会費	財務・会計	議会費	財務・会計
アクセス	所属路線	災害種別	災害種別	議会費	まちづくり推進	議会費	まちづくり推進
経路	路線別駅情報	収容可能人数	収容可能人数	項議会費	政策	項議会費	政策
道路区間	接続路線	非常用電源有無	非常用電源有無	議会費	住民	議会費	住民
道路	駅構内	指定有無	指定有無	総務費	選挙	総務費	選挙
方向	駅出入口情報	災害対応施設種別	災害対応施設種別	総務管理費	その他	総務管理費	その他
起点	周辺案内	種別	種別	徴税費	総務全般	徴税費	総務全般
終点	駅構内	コード	コード	戸籍住民基本台帳費	電子計算	戸籍住民基本台帳費	電子計算
範囲	駅	補足説明	補足説明	選挙費	公平委員会	選挙費	公平委員会
延長	深さ	災害種別	災害種別	統計調査費	庁舎	統計調査費	庁舎
道路	バリアフリー情報	災害種別	災害種別	監査委員費	支所・出張所	監査委員費	支所・出張所
種別	ホームドア・可動ステップの設置有無	災害種別	災害種別	総務管理費	監査	総務管理費	監査
種別(コード)	AEDの設置有無	コード	コード	一般管理費	その他	一般管理費	その他
ID	駅係員呼出しインターホンの設置有無	補足説明	補足説明	電子計算費	財務・会計	電子計算費	財務・会計
名称	非常停止ボタンの設置有無						
起点	可動式ホーム柵の設置有無						
重要な経過地	ホーム端転落防止柵の設置有無						

報告書・試作ツールを
<http://goikiban.ipa.go.jp/node640>
 から公開

「道路」	<ul style="list-style-type: none"> 浦安市都市整備部市街地開発課液状化対策推進室 	インディゴ株式会社
「観光」	<ul style="list-style-type: none"> 松江市 島根大学 松江商工会議所 	株式会社 ネットワーク応用通信研究所
「イベント」	<ul style="list-style-type: none"> 千葉市総務局情報経営部業務改革推進課 	インフォコム株式会社
「制度」	<ul style="list-style-type: none"> 神戸市こども家庭局 千葉市総務局情報経営部業務改革推進課 川口市企画財政部情報政策課 	株式会社 アスコエパートナーズ
「施設・サービス」	<ul style="list-style-type: none"> 横浜市金沢区 	株式会社三菱総合研究所

11

コア語彙検討の状況



コア語彙の公開ページ: <http://imi.ipa.go.jp/>

12

- 2014.6.6 「姿を現し始めた『共通語彙基盤』」
 - 試作ツールのデモと説明
 - 参加者：約100名
- 2014.10.22 「IPAグローバルシンポジウム」
 - 欧州/日本の語彙にかかる取り組みを紹介
 - 参加者：約140名
- 2015.2.4 「共通語彙基盤（IMI）プロジェクト成果説明会～オープンデータ・情報連携のための基本用語の定義と構造」
 - コア語彙2.2の説明、語彙基盤応用事例の紹介
 - 参加者：約130名

13

- 日・米・欧の関係機関が「Community of Practice on Core Data Models」の設立に合意
 - 現在設立準備中
- 目的
 - Good Practiceの共有
 - 各国の基本語彙間のマッピング
- 初期メンバー
 - 欧州委員会DG Informatics
 - 欧州各国（12カ国）の担当組織
 - 米国NIEM
 - 日本IMI
- 5月にSEMIC国際会議

14

Workshop - Core Data Libraries for public administrations - 12 November 2014, Brussels

Submitted by [Natacha Dewynqaert](#) on August 07, 2014

(⌘ ★★★★★) 5/5 | 1 votes | 62 reads |

WORKSHOP

Core Data Libraries for public administrations

Wed 12 November 2014 - 10 a.m. CET

@SEMICEu
<http://www.semic.eu>

Wed 12 November 2014
10:00 a.m. - 04:00 p.m. (CET)

- [About the workshop](#)
- [Agenda](#)
- [Online registration](#)

REGISTER NOW

About the workshop

On 12 November 2014, the [ISA Programme](#) of the European Commission will host the SEMIC Workshop "Core Data Libraries for public administrations". The workshop invites e-Government experts, standardisation bodies, tool vendors, and ICT service providers to share their views on how Europe can maximally benefit from collections of reusable interoperability assets maintained by Member States, Standardisation Organisations, and non-governmental organisations.

2015年5月 EC主催の国際で発表予定

Login or Sign up

Share and reuse interoperability solutions for public administrations

European Commission > ISA > JoinUp > Communities > Semic > Semantics > SEMIC 2015 - Semantic Interoperability Conference

My Page
Communities
Interoperability solutions
News
Events
e-Library
People

Find communities
My Communities
Recommended
Editor's Choice

3,33/5 | 135 votes

⌘ ★★★★★

SEMIC - Semantic Interoperability Community

Overview Description Presentations Pilots Publications Training Members list Highlights Forums **Communications**

SEMIC 2015 - Semantic Interoperability Conference

Submitted by [Stephan Kotopoulou](#) on November 21, 2014

(⌘ ★★★★★) 5/5 | 2 votes | 1711 reads |

5 May 2015

Riga - Latvia

Attend Event

Bringing Enterprise Information Management and Service Portfolio Management to public administration
 Emerging trends in public administration, such as "digital by default" and the "once only" principle, require public organisations to rethink the way they manage information, especially in the context of public service delivery. "Digital by default" is the approach of putting the online channel at the forefront of public service delivery. The "once only" principle involves eliminating the unnecessary administrative burden involved when users are required to supply the same information more than once to government.
 SEMIC 2015 will explore how Enterprise Information Management and Service Portfolio Management can help public administration improve its modus operandi, reduce costs and improve public services through digital transformation. While these practices are common in the private sector, we observe that **adoption in public administrations remains quite low**.
Enterprise Information Management introduces principles and practices for managing an organisation's information in a correlated, contextual and enterprise-wide manner. In this context, certain types of data become very important:

- **Master data:** data used by various applications and systems across the enterprise. Governments store their master data in what has been called "base registries". Access to base registries should enable public administrations to have access to a single source of "Single Version of Truth".
- **Core Vocabularies (aka Core Data Models):** the data models used to represent master data.
- **Reference Data:** highly reusable value lists like codeslists and taxonomies.

Service Portfolio Management aligns the delivery of digital services with the business objectives of the organisation. It introduces a service perspective on organisations. Service Portfolio Management involves a mix of activities including service documentation, service governance and supporting IT tools. Instead of creating a flat list of services a "service portfolio" captures relationships, information flows and dependencies forming a complex enterprise service ecosystem. Service Portfolio Management principles and practices can help public administrations to manage efficiently and effectively their digital public services and better understand the underlying information flows.
 In this year's conference experts from the private and the public sector will share first-hand experiences, tools and best practices on the above presented topics.

Related Content

- 16 January 2014 | European Union
SEMIC 2014 - Semantic Interoperability Conference
- 10 April 2013 | Ireland
SEMIC 2013 - Semantic Interoperability Conference 2013
- 18 June 2012 | Belgium
SEMIC 2012 - Semantic Interoperability Conference 2012
- 09 March 2012 | Europe
CEBAR Workshop 2012.03.07
- 19 December 2011 | Europe
EU - CEBAR Workshop: Share experiences with setting up and maintaining semantic asset repositories

Newsletter

You will receive news and links of the most interesting developments ...

[Subscribe to newsletter](#) >

Meet our Keynote Speakers

Wada Takashi - [Commerce and Information Policy Bureau, Ministry of Economy, Trade and Industry \(METI\), Japan](#)

Mr. Takashi Wada is the Director of Information Technology Project Office, Ministry of Economy, Trade and Industry (METI), GOJ. He serves for e-government policy planning and administration of major e-government projects including open data demonstration projects and the Infrastructure for Multilayer Interoperability (IMI) project among e-government systems and private sector. The Japan's latest IT strategy in 2013 refers to the development of information exchange infrastructure among government organizations where fragmented authorities at central government and 1700+ local municipalities hinder streaming flow of information. The IMI project is directly responding to such policy need. Mr. Wada joined Ministry of International Trade and Industry (currently, METI) in 1993. After undertaking several IT related positions, he was assigned to the current position in 2013. Since then he has been taking leadership in opendata and government interoperability discussions within the government.

Shulichi Tashiro - [IPA \(Information-technology Promotion Agency\), Japan](#)

Dr. Shulichi Tashiro is a General Manager of the Open Standards Promotion Center of the Information-technology Promotion Agency (IPA), Japan and driving IMI project to enhance semantic interoperability of eGovernment systems. IMI published their core vocabulary on February 2015 after experimental use among some forward-thinking local governments and open-data community. Dr. Shulichi Tashiro began his work in Electrotechnical Laboratory of Ministry of International Trade and Industry as a researcher of computer science and networking in 1987. From 2002 he served in METI to start up projects to promote OSS and Open Standards. He moved to IPA in 2006. He is also a lecturer of University of Tokyo and Chuo University.

METI and IPA are developing the IMI core vocabulary which provides a Concept Dictionary (description of meaning, relationship and hierarchical structure) and sample Information Exchange Package Descriptions for fundamental terms which can be accessed at <http://imi.ipa.go.jp/>.

The purpose of the IMI core vocabulary is to ensure government information is well-understood by public in case the information is published as open-data, and well-understood by each e-government systems in case of the information is used for internal information exchange.

Currently, IMI is at the end of trial stage. 5 pilot projects utilized the IMI vocabulary to publish open-data in local governments. And there are many plans to use the IMI core vocabulary. There are several community-based activities such as "hackathon" that tried to use IMI vocabulary for developing open-data and software.

Registration

If you wish to attend, please pre-register online as soon as possible **and no later than April 21** via the following link: <http://semic2015.eu/authentication.jsp>.

Please note that your registration is subject to approval by the organizers. You will be notified as soon as possible if your registration has been accepted or not.

Spread the word: #SEMIC - @SEMICEu

Related initiatives:

DCAT

ADMS

reference

CORE

CORE

CORE

CORE

構築中のツールについて

2015年2月

独立行政法人 情報処理推進機構

1

開発中ツールの概要



✓ 表計算ソフトやCSV形式等で保存されているデータに対し、IMI語彙による情報の付加や構造の整理を行いコンピュータ処理に適したデータを作成にする作業を支援します。

元データ

ID	名称	よみがな	郵便番号	住所
N7001	金沢さくら保育園	かなざわさくらほいくえん	2360021	神奈川県横浜市金沢区泥亀一丁目21-1
N7002	金沢谷保育園	かなざわやほいくえん	2360042	神奈川県横浜市金沢区金沢谷東一丁目1-3
N7009	金沢保童園	かなざわほどうえん	2360091	神奈川県横浜市金沢区金沢五丁目20-1

- IMIが定義する構造に従って、データの構造を整理する。
- IMIが定義する項目の情報をデータに付加する。

作成ツール

既存のデータを 入力

利用者は項目に対応するIMIの用語を選択

- ID → ic:施設
- 名称 → ic:ID
- よみがな → ic:名称
- 郵便番号 → ic:連絡先の電話番号
- 住所 → ic:住所の郵便番号
- 住所 → ic:住所(分割)

ツールによる助言
「住所」には「ic:住所」を対応付けることができます。

ツールによるデータ構造の整理
住所の分割
電話番号の定型化など

IMI語彙が付加されたデータ

```

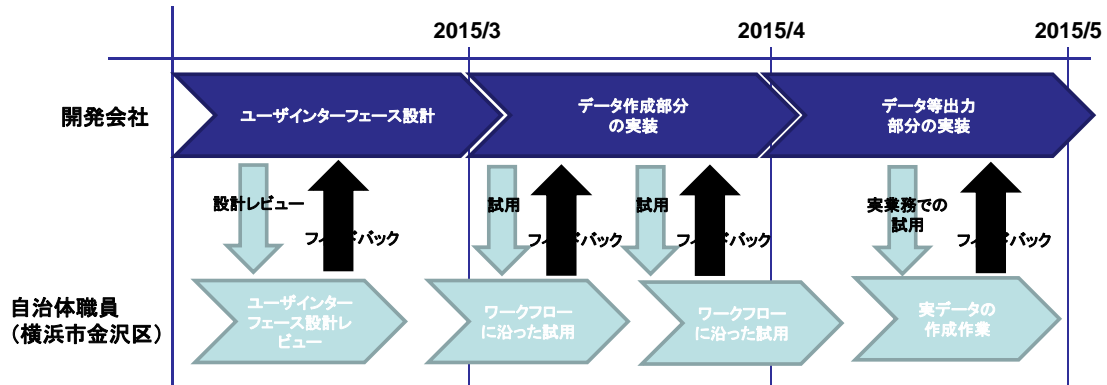
prefix xx: <http://xx#>
prefix ic: <http://im1.ipa.go.jp/ns/core/rdf#>

xx:N7001 a ic:施設 ;
ic:ID [
  ic:ID体系 <http://xx#> ;
  ic:識別値 "N7001"
];
ic:名称 [
  ic:名称型 ;
  ic:漢字 "金沢さくら保育園";
  ic:カナ表記 "カナザワサクラホイクエン"
];
ic:住所 [
  ic:住所型 ;
  ic:郵便番号 "2360021";
  ic:都道府県 "神奈川県";
  ic:市区町村 "横浜市";
  ic:区 "金沢区";
  ic:町名 "泥亀";
  ic:丁目 "1";
  ic:番地 "21";
  ic:号 "1"
];

xx:N7002 a ic:施設 ;
ic:ID [
  ic:ID体系 <http://xx#> ;
  ic:識別値 "N7002"
  
```

2

- まずは、横浜市金沢区が運営する、「かなざわ育なび.net」により提供されている、子育て、防災等に関する、施設、制度、イベント等すべてのデータをIMI語彙を利用した情報が付加された二次利用がし易いオープンデータとして提供することを目標とする。
- 二次利用し易いオープンデータの作成作業を自治体職員自身ができるようにすることを目指し、利用者の利便性を最優先した設計とする。
- 自治体職員の利便性を高めるため、設計段階から、実装段階に至り、自治体職員によるレビュー・試用を繰り返し行う。



- 2015年5月初旬に開発を終了し、2015年6月中に運用を開始する予定
- IPAのサーバーにてツールを運用するとともに、変換ロジックなどを再利用できるよう、ソースコードをオープンソースとして提供する予定。

コア語彙、IEPについて

2015年2月3日 「正式版」を公開
<http://imi.ipa.go.jp/>

2015年2月
独立行政法人 情報処理推進機構

1

コア語彙とは



【コア語彙】

どのような分野・サービスでも広く利用される普遍的な語彙。「人」「氏名」「住所」「連絡先」「組織」「場所」「日時」「日付」「建物」「施設」「座標」など、特定のものを表現する語彙から、「数量」「面積」「重量」「長さ」「時間」「金額」のような計量のための基礎概念を表現する語彙などが含まれる。

【ドメイン共通語彙】

「病院」「避難所」など、複数分野に渡って使われる用語

【ドメイン語彙】

「診療科」「病床数」など、特定分野での利用に特化した用語



2

- コア語彙 (HTML) <http://imi.ipa.go.jp/ns/core/2/>
データ作成者が、コア語彙として定義されている用語を確認するためのWebページ。
- スキーマ
データの検証やデータの意味を推測するためにコンピュータシステムが参照するコア語彙の定義ファイル (通常のWebブラウザでは表示できません。)
 - コア語彙 XML スキーマ 名前空間: <http://imi.ipa.go.jp/ns/core/2/>
 - コア語彙 RDF スキーマ 名前空間: <http://imi.ipa.go.jp/ns/core/rdf#>
- IEP (情報交換パッケージ) <http://imi.ipa.go.jp/doc/IEP/v1/>
コア語彙を使用して情報交換を行うための雛形です。IEPを共有することにより、同種のデータの定型化を促進することができます。
コア語彙2.2のリリースに併せ、サンプルとして参照できるよう、下記9つの実用レベルのIEPを提供しています。
○住所 ○地物 ○施設 ○避難施設 ○イベント ○医療機関 ○氏名 ○設備 ○組織
- 共通語彙基盤およびコア語彙に関するドキュメント
 - 共通語彙基盤概要 http://imi.ipa.go.jp/doc/IMI_Overview_v2.pdf
自治体職員等、技術者でない方を対象とし、共通語彙基盤の意義や基本的な考え方が理解できるように解説したドキュメントです。
 - コア語彙テクニカルガイドライン http://imi.ipa.go.jp/doc/IMI_Core2_TechGuide_v2.pdf
コア語彙を用いたデータ設計やシステム設計、また、同形式の独自語彙等の定義を行う技術者向けに技術的な詳細を解説したドキュメントです。

- 様々な分野のデータの中で共通に使用される用語で構成、用語には以下の2種類の用語があります。

クラス用語

- 「人」や「組織」などの大きな概念を表す用語。プロパティ用語のリストをもつ。
- 48個の用語が定義されている。

クラス用語一覧

分類	クラス用語
物や事象を表す用語	<ul style="list-style-type: none"> ● 人 ● 施設 ● 駐車場 ● 建物 ● 組織 <ul style="list-style-type: none"> ● 法人 ● 業務組織 ● イベント ● 製品 ● 製品個品 <ul style="list-style-type: none"> ● 地物 ● 設備
物や事象の説明を行う用語	<ul style="list-style-type: none"> ● 氏名 ● 住所 ● 場所 ● 座標 ● ID ● ID体系 ● コード ● コードリスト ● 名称 <ul style="list-style-type: none"> ● 連絡先 ● アクセス ● アクセス区間 ● スケジュール ● イベントスケジュール ● 定期スケジュール ● 詳細スケジュール ● 詳細スケジュール規則 ● 構成員 <ul style="list-style-type: none"> ● 施設関連 ● 事物 ● 実体 ● 状況 ● 組織関連
値に関する用語	<ul style="list-style-type: none"> ● 重量 ● 価格 ● 期間 ● 時間 ● 人数 <ul style="list-style-type: none"> ● 数量 ● 長さ ● 面積 ● 容量 ● 電話番号 <ul style="list-style-type: none"> ● 日時 ● 日付 ● 金額 ● 測定値

プロパティ用語

- クラス用語を構成する項目となり、クラス用語と、他のクラス用語や値との関係を表す用語。
 - 206個の用語が定義されている。
- 例 (カッコ内はプロパティ用語を項目としてもつ主なクラス用語)
生年月日(人)、都道府県(住所)、関連組織(組織)、製造者(製品)、送付先(連絡先) など、

共通語彙基盤整備事業 2015年度実施項目(案)

1

2015年度の実施項目(案)

IPA

- 情報連携用語彙データベースの構築開始
 - ドメイン語彙の作成、検討の支援
 - コア語彙の配布、検索
 - ドメイン語彙のポータル
 - ツールの配布、検索
- 語彙活用のためのツール構築
 - データ設計支援
 - データ入力支援
 - データ変換支援
- コア語彙の充実
 - 用語定義や説明文書の充実
 - 国際連携
- 活用支援、普及・啓発
 - 事例の集積と導入ガイドの作成（アクションプランに提案）
 - 省庁・自治体等への個別対応
 - セミナー（地方へ紹介へも力を入れる）

2

文字情報基盤整備事業の進捗状況と 2015年度実施項目（案）

2015年2月

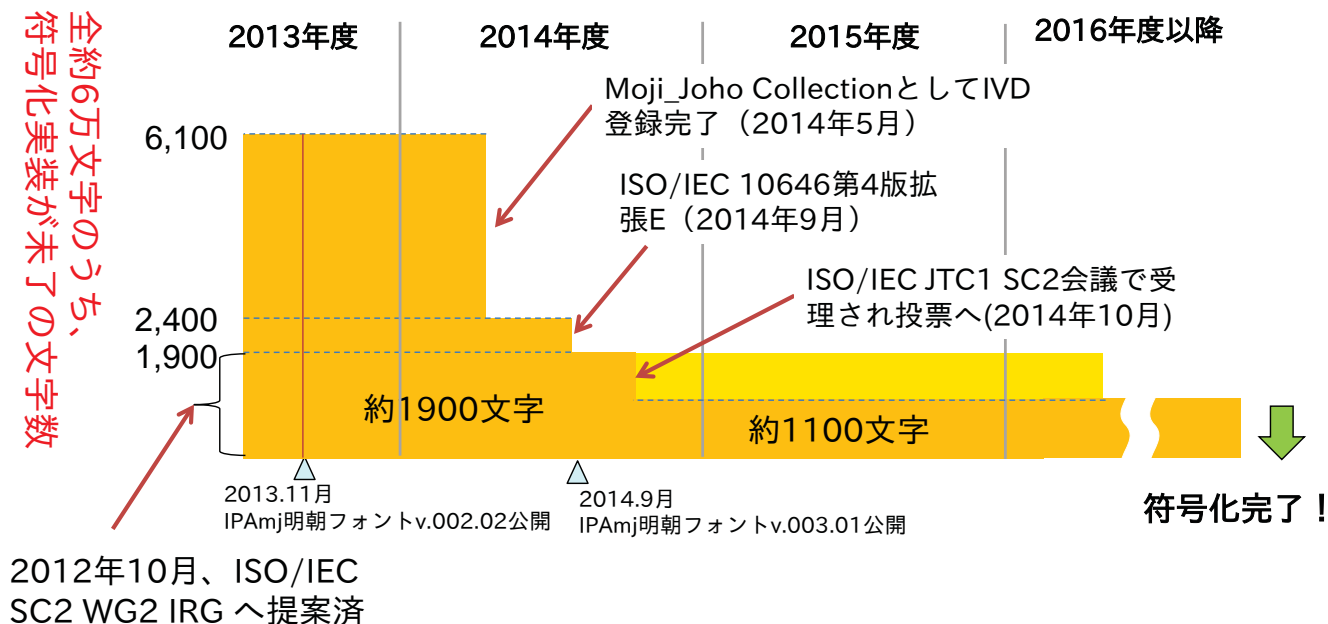
独立行政法人 情報処理推進機構

1

前回委員会（2014年2月）以降の動き IPA®

日付	事象	概要
3月13日	業界団体の設立	文字情報基盤の普及へ向けた業界活動を推進するため、「IVS技術促進協議会」を改組し、「文字情報技術促進協議会」へ
3月20日	文字情報基盤導入ガイド類の公開	IPAから公開 <ul style="list-style-type: none"> 文字情報基盤導入パンフレット 文字情報基盤導入ガイド Ver.1.0 文字情報基盤導入テクニカルスタディ Ver.1.0
3月24日	「電子自治体の取組みを加速するための10の指針」	総務省自治行政局から自治体へ向けた指針 <ul style="list-style-type: none"> 国の取り組みとして文字情報基盤を紹介
4月25日	「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」	各府省情報化統括責任者（CIO）連絡会議決定 <ul style="list-style-type: none"> 導入ガイドの参照 縮退マップの整備
5月16日	Moji_Joho collectionのIVD登録	文字情報基盤の文字図形を収録した” Moji_Joho Collection” が、Unicode コンソーシアムから公開
6月18日	セミナーの実施	幕張メッセで開催されたInterop会場にて 「行政の電子化への活用が進む『文字情報基盤』」を開催
6月30日	文字情報基盤データベース構築開始	文字に係る情報をRDF形式で格納し、公開するデータベース。2015年夏に仮運用開始予定。本格運用開始は2016年4月。
9月1日	ISO/IEC 10646:2014の発行	文字情報基盤の文字のうち、約500文字が規格に追加
9月19日	MJ文字情報一覧表, IPAmj明朝のバージョンアップ	IVD登録、ISO/IEC10646の発行を反映 標準に従って使える文字が約4,000文字増加
10月3日	ISO/IEC JTC1 SC2委員会開催	<ul style="list-style-type: none"> 文字情報基盤から提案中の約1900文字のうち、約800文字が投票へ。 文字情報基盤のグリフの集合の定義を10646規格書に加えることにつき、提案が受理された → 投票処理中
1月7日	変体仮名規格案の準備状況を情報規格調査会SC2専門委員会へ報告	<ul style="list-style-type: none"> 対象文字の選定状況、規格化方針案を報告し、議論

2



参考：国際会議の日程

- SC 2 会議
 - 20th – 2015-10-19/23, Matsue, Japan, collocated with WG 2
 - 21st – 2016-10(tentative), Russia (tentative), USA(backup)
 - 22nd – 2017-10(tentative), China (pending confirmation)
- WG 2 会議
 - 64th – 2015-10-19/23, Matsue, Japan
 - 65th – 2016-10(tentative), Russia (tentative), USA(backup)
 - 66th – 2017-10(tentative), China (pending confirmation)
- IRG 会議
 - 44th - 2015-06-15/19, Seoul, Republic of Korea
 - 45th - 2015-11-16/20, HKSAR

SC2: 符号化文字集合の専門委員会
 WG2: ISO/IEC 10646 (国際符号化文字集合) 維持・作成の作業部会
 IRG: Ideographic Rapporteur Group (漢字標準化作業チーム)

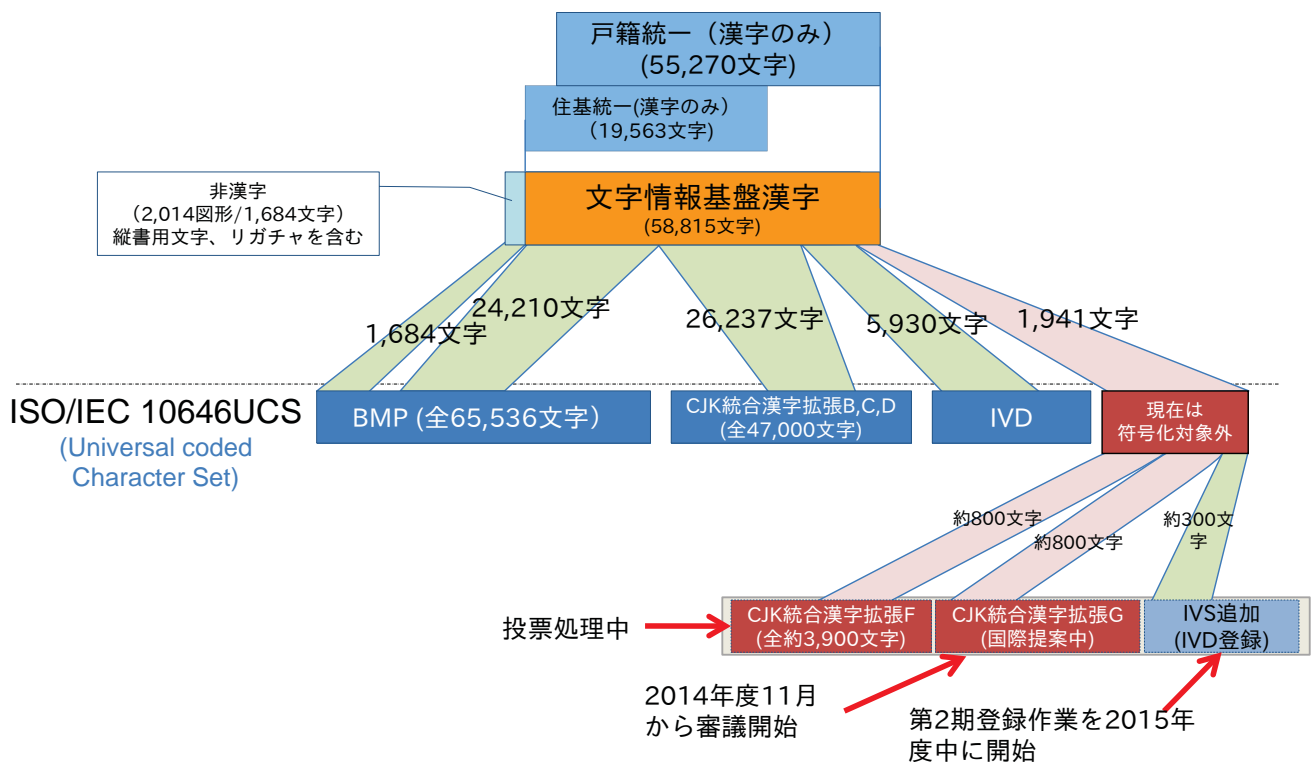
- 2014年9月19日に公開
- MJ文字情報一覧表Ver.004.01
 - 国際規格の更新に伴う掲載情報の追加
 - 行政実務で用いられる字辞典検字番号情報の追加
- IPAmj明朝フォントVer.003.01
 - 2014年9月発行のISO/IEC 10646 第4版に対応
 - 国際標準規格の最新版に対応し、前バージョンと比較して約4,200文字が利用可

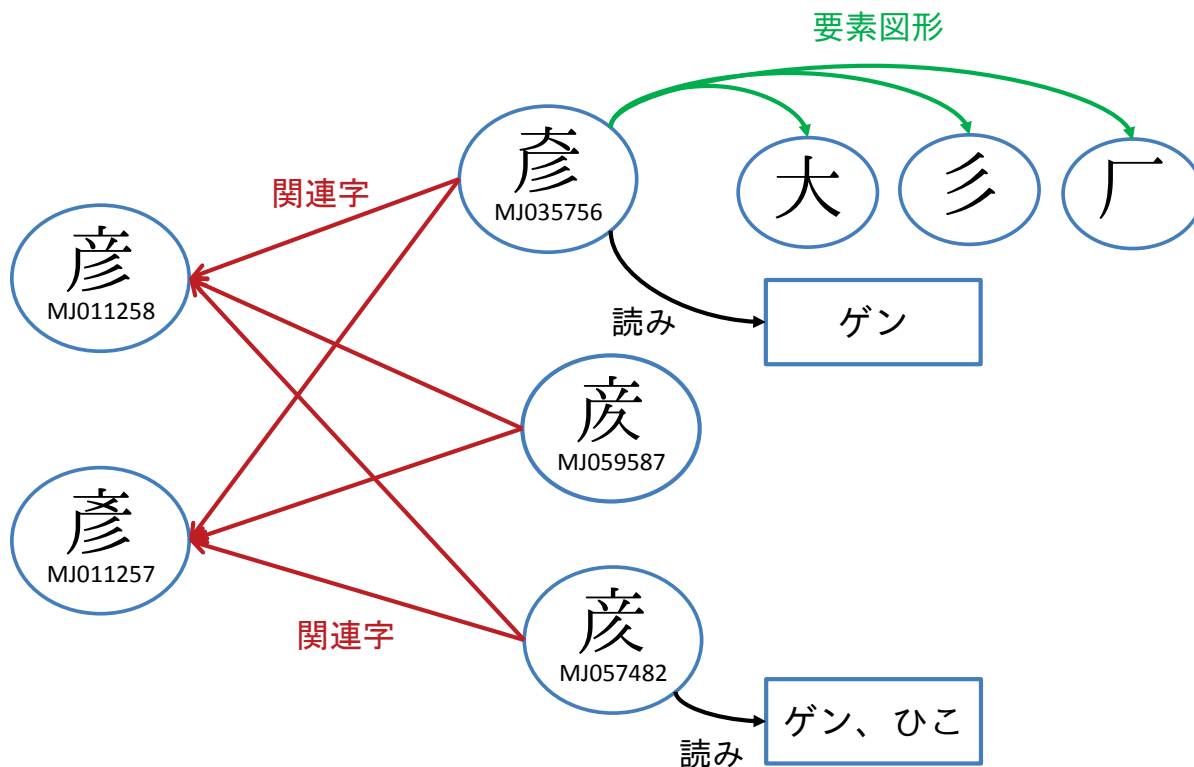
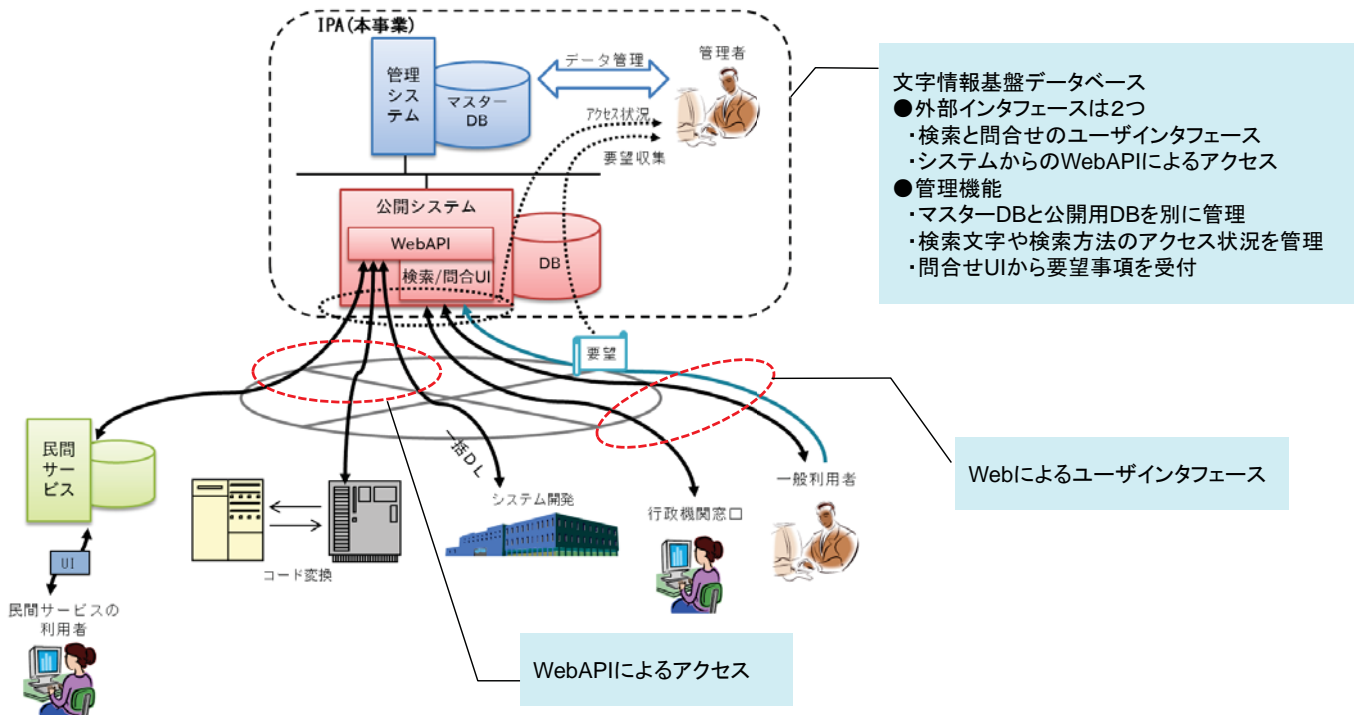
- 国際規格更新に係る項目追加
 - IPAmj明朝Ver.003.01で実装した「Moji_JohoコレクションIVS」
 - IPAmj明朝Ver.003.01で実装した互換漢字に対するSVS(Standardized Variation Sequence)
- ISO/IEC 10646第4版に伴う「対応するUCS」の更新
 - 新たに付与したもの・・・505文字
 - 符号位置を変更したもの・・・2文字
- 以下の字辞典検字番号の追加
 - 講談社新大字典
 - 角川大之源

- UCS、IVSの符号化実装の追加
 - ISO/IEC 10646 第4版に対応し、拡張Eに該当する496文字をUCS符号化実装
 - IVD Moji_JohoコレクションのIVSを10,710文字実装
 - 互換漢字に対するSVSを89文字に対し実装
- 前のバージョンと比較して、利用可能な文字が4,152文字増加

現在のUCS符号化状況

IPAmj明朝フォントVer.003.01の符号化状況





誤字俗字・正字一覧表

「正字等」漢字番号	「正字等」MJ図形名 (Subject)	アザ心差	「誤字等」漢字番号	種別と表示記号 (Predicate)	「誤字等」MJ図形名 (Object)	アザ心差
1554-1-1	MJ011752		1554-1-1-1	民事2842誤字		
1554-1-1	MJ011752	1				
1554-1-1	MJ011752	1				
1554-1-2	MJ010612		1554-1-2-1	民事2842誤字		
1554-1-2	MJ010612		1554-1-2-2	民事2842誤字		
1554-1-3	MJ010114		1554-1-3-1	民事2842誤字		
1554-1-4	MJ009693		1554-1-4-1	民事2842誤字		
1554-1-4	MJ009693	1				
1554-1-4	MJ009693		1554-1-4-2	民事2842誤字		
1554-1-5	MJ016350		1554-1-5-1	民事2842俗字(▲) MJ057990		

※他の種別 → 民事2842別字 (●)

誤字俗字・正字一覧表における整理

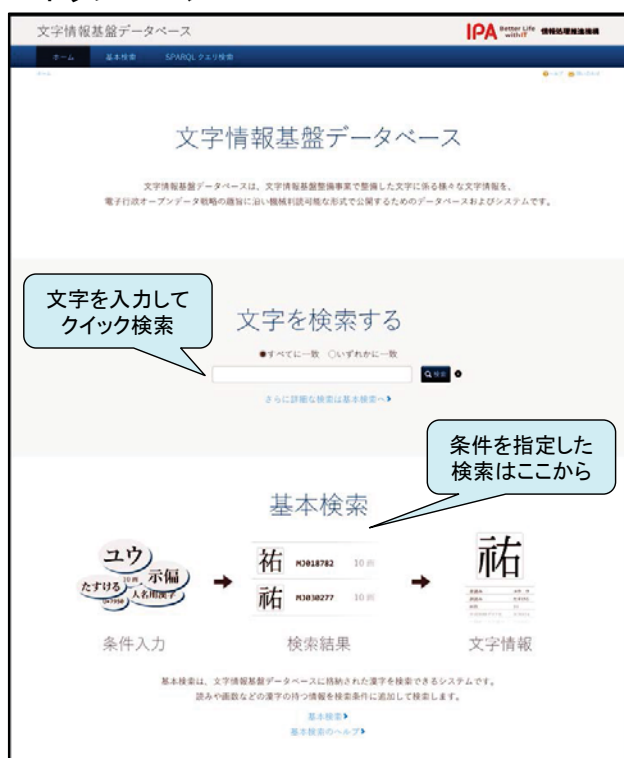
辞書における関係性の整理

代表字 異体字 関係

Subject	Predicate	Object
8351	本字	8352
8351	別体	8353
8351	別体	8354
8351	別体	8355
8351	別体	8356
8351	別体	8357
8351	別体	8358
8351	別体	8359
8351	別体	8360
8351	別体	8361
8351	別体	8362

検索機能	概要
1 基本検索	以下に示すような画数、読み、部首等の項目を指定した検索 ・ 常用漢字 ・ 音読み ・ 戸籍法上の人名漢字 ・ 訓読み ・ 総画数 ・ JIS X 0213水準 ・ 康熙字典 ・ 部首内画数
2 コード検索	以下に示すような各種コードを指定した検索 ・ MJ文字図形名 ・ 平成明朝グリフ名 ・ 戸籍統一文字番号 ・ 住基ネット統一コード ・ 入管正字コード ・ 登記統一文字番号 ・ 対応するJIS X 0213面区点位置 ・ 対応するJIS X 0212区点位置 ・ 対応するUCS ・ 大漢和辞典 検字番号 ・ 日本語漢字辞典 検字番号 ・ 新大字典 検字番号 ・ 大字源 検字番号
3 要素図形検索	利用者がMJ文字図形に含まれる要素図形を指定した検索
4 SPARQLクエリ検索	利用者が直接SPARQLクエリを指定する検索

トップページ

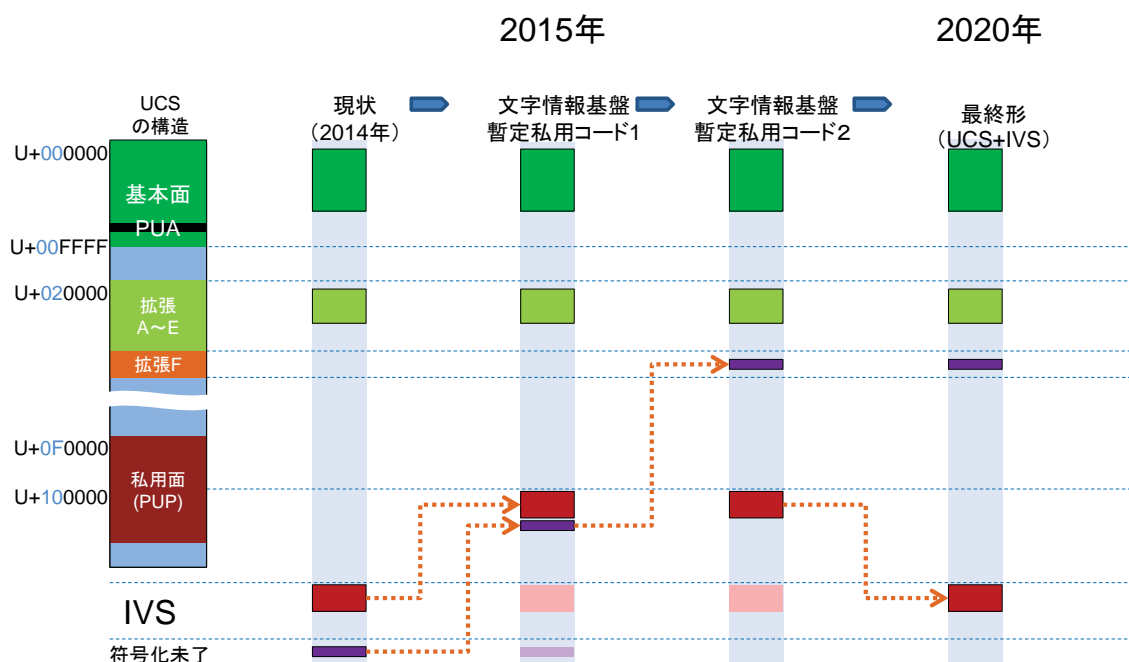


検索結果、文字詳細表示



暫定私用コードの策定と配布

- 暫定私用コードの必要性
 - 導入ガイドと同時公表のテクニカルスタディに明記
- 標準化作業と実装普及までの暫定的な施策
 - UCSにおいて未符号化のMJ文字図形約1900文字を暫定的に私用面にマッピング
 - IVDにMJコレクションとして登録済みのMJ文字図形約6000文字を暫定的に私用面にマッピング
- 文字情報技術促進協議会が策定・配布
 - 利用に当たっては、対象組織の限定等を行わない
 - 私用文字領域利用の際は、送受信間の相互了解が必須
 - 利用に当たっての副作用（コードを共有していない組織への送信制限の必要性、暫定私用コード利用期間終了時の移行処置の必要性等）については、利用者の自己責任を徹底



- 行政実務での必要性
 - 新戸籍法の施行（1948年）以前に生まれた人の名に使用例
 - 現行戸籍法下でも必要。戸籍法施行規則の改正で除籍謄本の保存期間が150年になったので、今後も相当期間必要
 - 現在の戸籍統一文字には168文字の変体仮名に固有IDが付与されている
- 学術用途での必要性
 - 高田智和氏（国立国語研究所）が中心となり、印刷業界、学術関係者からの聞き取り調査を実施
 - 日本語学、歴史学、書誌学、図書館情報学など、人文科学系の研究分野において、情報通信環境下での変体仮名利用への根強い要望があった
 - 研究者有志による討議検討を経て、必要変体仮名300文字弱を選定
- 情報規格調査会SC2専門委員会としての変体仮名レポートリーの選定と国際提案
 - 行政実務用と学術用を整理統合した300文字強を提案
 - 音価と字母の組に固有のIDを付与する方向で検討中



今後の計画(案)

日付	事象
2014年度中	標準提案へ向けた変体仮名文字セット、デザイン案の確定
2015年8月	MJ文字情報一覧表へ追加 (戸籍文字との対応関係確定)
2015年9月	ISO/IEC JTC1/SC2総会及び同WG2会議への提案
2017年度	標準化完了を目指す

- 変体仮名の国際規格化提案
- 提案済みの漢字2000文字標準化フォローアップ
- 文字情報データベースの構築と稼働
- 業界団体との連携
- 普及啓発

縮退マップについて

2015年2月

独立行政法人 情報処理推進機構

1

必要性



- インターネットでの情報公開や電子申請などの場面では、多くのデバイスで使用できるよう、使う文字をJISの範囲とすることが求められる。
- その他、使用する文字の範囲を不必要に大きくせず、業務に応じた適切な大きさとするのが、効率化の上で適切。



人名等で用いる6万文字をJIS範囲へ縮退する対応付けを、組織毎に異なる方式やポリシーで行うと、混乱を招く恐れ



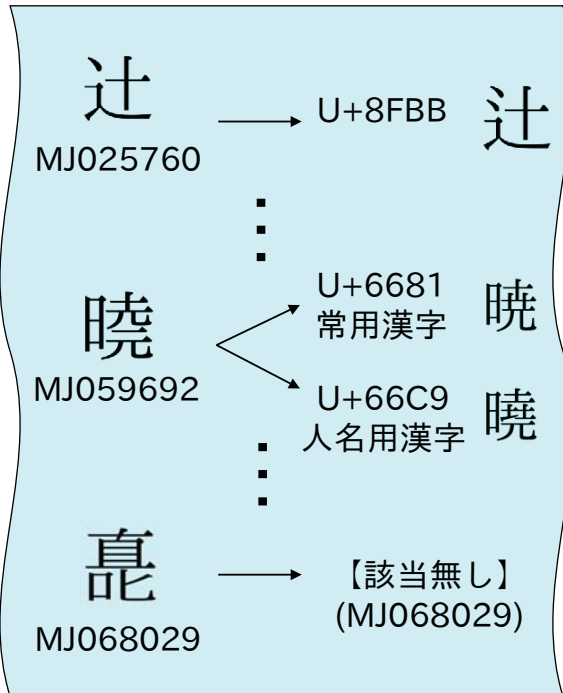
縮退対応にあたり、その規範となる情報が必要

2

- 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（平成26年4月25日 各府省情報化統括責任者（CIO）連絡会議決定）
 - 経済産業省は、各府省の円滑な導入を支援するため、文字情報基盤の文字（約6万文字）を、市販コンピュータで特別な設定無しで活用できるJIS範囲の文字への変換を行う際に参照する縮退マップの提供を、平成26年度中に実施する。

- 各情報システムの現場で、縮退変換処理を実施する装置等に設定するための変換テーブルを作成するために参照する情報である。
 - 直接マシンに設定することは想定しない
- 文字情報基盤文字セット→JIS文字セットの対応関係はn対1に限らず、
 - n対m（複数の候補を示す）、
 - n対0（IPAとしては候補を示さない）があり得る。
- n対m、n対0対応文字については、
 - 変換テーブル作成時に、当該システム関係者が選択
 - 変換テーブルによる変換作業時に、人の介入を求める等の対処が必要となる。

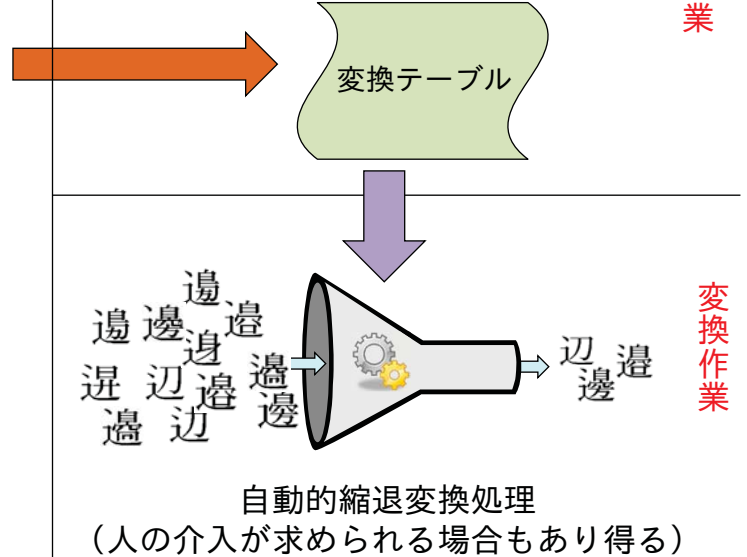
縮退マップ(IPA提供)



各自治体等での作業

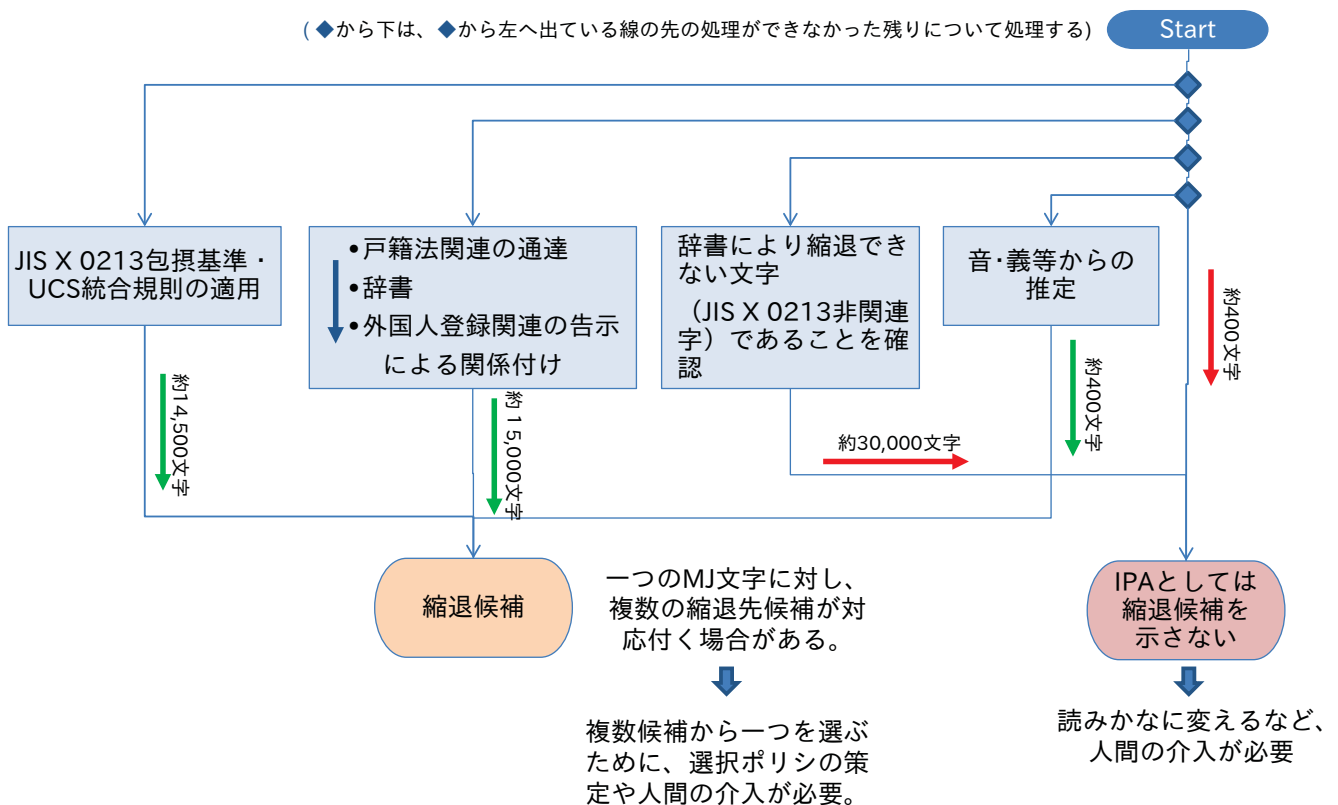
- 対応先の確定
- 変換作業時の人の介入有無等に係るポリシーを決め、変換テーブルを作成

準備作業



縮退マップ作成の手順

(◆から下は、◆から左へ出ている線の先の処理ができなかった残りについて処理する)



- 縮退できない文字で有ることを辞書で確認できた場合

例：早→日下(くさか)、𪗇→女形(おやま)、𪗇→忌寸(いみき)

MJ057651
大漢和辞典 補231

MJ057298
大字源国字一覧 2116頁4段目6個目

MJ057347
大字源国字一覧 2116頁4段目4個目

例：𪗇→四十雀(しじゅうから)、𪗇→牛の羹(うしのあつもの)

MJ059246
大漢和辞典 補786

MJ058405
大字源国字一覧 2112頁2段目9個目

- 音・義不詳／辞書に見いだせなかった場合

例：𪗇

MJ059986

𪗇

MJ059305

𪗇

MJ059350

𪗇

MJ059224

縮退マップのイメージ

実際のマップはCSV形式（図形無し）で公開予定

MJ文字図形	JIS包摂規準/ UCS統合規則	戸籍法関連通達	法務省告示584号	字辞典掲載 を確認	字辞典での関連字	音・義等からの推定
𪗇 MJ025760	𪗇 1-36-52 U+8FBB			○		
𪗇 MJ000259		左 1-26-24 U+5DE6	𪗇 1-53-87 U+5C22	○	左, 𪗇 1-26-24, 1-54-05 U+5DE6, U+5C6E	
𪗇 MJ059692		𪗇, 𪗇 1-22-39, 1-58-92 U+6681, U+66C9		×		
𪗇 MJ016424				×		𪗇 1-87-65 U+71FE
𪗇 MJ068029				×		

- 辞書典拠のある文字の調査
 - （典拠辞書：大漢和辞典、大字源、新大字典、日本語漢字辞典）
 - 典拠辞書とMJ文字図形との対応調査
 - 各典拠辞書内の関連字調査
- 辞書典拠の無い文字の調査
 - 援用資料：誤字俗字・正字一覧表（平成一六年一〇月一四日民一第二八四二号民事局長通達）
 - 援用資料：正字・俗字等対照表（平成二年一〇月二〇日民二第五二〇二号通知別表）

2014年度中

- 3月12日 文字情報基盤運用検討SWG
- 3月24日 文字情報基盤WG
- 3月末 「縮退マップ v.0.1」 および「利用の手引き」を公開
- 意見募集の実施
- 2015年夏 「縮退マップ v.1.0」を公開

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

平成 26 年 4 月 25 日

各府省情報化統括責任者（CIO）連絡会議決定

「世界最先端 IT 国家創造宣言」（平成 25 年 6 月 14 日閣議決定）に基づき、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を下記のとおり定める。

記

1. 基本的考え方

世界最先端 IT 国家創造宣言では、より便利で利用者負担の少ない行政サービスの提供を、国民と政府の相互協力関係を基軸として進めることにより、持続的に発展可能な新たな電子行政モデルを構築することとしている。

創造宣言の目指すべき社会を実現するためには、政府・自治体、官民間の連携を円滑にするためのデータ構造等の共通基盤を整備することが不可欠であり、このような共通基盤が整備されれば、行政の効率化やサービスの品質向上、社会における生産性向上も期待できる。

また、従来電子政府施策において重点的に取り組んできた、インターネットを活用した情報提供についても、提供情報の分かりやすさや最新化等についての意見・要望もみられるところであり、今後も引き続き、利用者視点に立った継続的な改善活動に取り組んでいく必要がある。

2. オープンな利用環境整備に向けたアクションプラン

本アクションプランでは、創造宣言の目標年次である 2020 年度（平成 32 年度）までの間、以下のとおり、**文字、用語、コードの標準化・共通化**等オープンな利用環境整備のための基盤整備に優先的に取り組みつつ、国民・企業等の利便性向上を図るための Web サイトの見直しを順次展開していくこととする。

なお、利用者の利便性向上のためには、国と地方公共団体のシームレスな連携が必要であり、地方公共団体においても、本アクションプランを参考に、オープンな利用環境の整備に向けた取組を推進することが望ましい。

2. 1 オープンな利用環境の実現

(1) 文字情報の標準化と活用について

今後整備する政府情報システムにおいては、国際標準に適合した文字情報基盤を活用することを原則とする。

各府省は、経済産業省を中心に策定した「文字情報基盤導入ガイド」を参照しつつ、導入を推進する。また、経済産業省は、各府省の円滑な導入を支援するため、文字情報基盤の文字（約6万文字）を、市販コンピュータで特別な設定無しで活用できる JIS 範囲の文字への変換を行う際に参照する縮退マップの提供を、平成 26 年度中に実施する。

(2) 用語の共通化等について

システム間の連携やデータの二次利用を円滑にするためには、氏名、住所、組織等の普遍的用語の記述に関する情報連携用語彙の早期整備を図るとともに、各組織が運用する既存の語彙との連携方法を整理することが必要である。

このため、経済産業省において、国際的な連携を踏まえた情報連携用語彙の整備を進め、整備された語彙から順次公開するとともに、平成 27 年度中に情報連携用語彙データベースの運用を開始する。また、総務省において、情報流通連携基盤共通 API（データモデル、共通ボキャブラリ、標準 API 規格等）等の開発・実証を平成 27 年度中に行う。

(3) コードの標準化等について

公開データの二次利用を促進するため、「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）」に沿って、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。

2. 2 利便性の高いサービスの実現

(1) 調達情報、制度情報の集約提供について

調達情報や制度等に関する情報は、国民・企業等が行う社会・経済活動に有益な情報であり、国や地方公共団体の情報を集約して提供することについての要望も高い。

このため、調達情報については、「電子行政オープンデータ推進のためのロードマップ」に沿って、より高度な利用が可能なデータ形式での公開を拡大する。

また、制度等に関する情報については、復旧復興支援制度データベースの

ように国と自治体が一体となったサービスを実現している例も見られるところである。内閣官房において、このような先行事例の取組も参考にしつつ、利用者視点でより一層迅速かつ容易に情報を入手することが可能となるよう、関連情報の集約的提供等に係る方策等を検討の上、平成 26 年度中に方針を策定し、推進する。

(2) 利便性向上のための Web サイトの見直し

各府省は、それぞれの保有する Web サイトについて、時宜を得た情報提供と提供内容の最新化、また、提供情報のわかりやすさと利便性の向上等の観点から点検するとともに、利用者のニーズを踏まえた継続的な改善活動を行うこととする。

このため、内閣官房において、平成 26 年度中に、以下のような基本的なガイド類の整備を行うこととし、各府省は、これを踏まえ Web サイトの改善活動を実施する。

ア. 組織別サイトの基本デザイン・メニュー項目等

組織別サイトについて、ユーザビリティ・アクセシビリティに配慮した基本デザイン、基本情報項目、見出しや案内方法も含めた共通コンテンツ、政府の共通的な開発環境の在り方 等

イ. 統一的タグの付与

各府省トップページに掲載している新着情報への、統一的タグ（例：[調達]、[会議]、[報告]、[発表]等）の付与の在り方

2. 3 統一的なガバナンスの確保

(1) 政府ドメインの運用の見直し

各府省は、別途定める要領に基づき、ドメイン管理のプロセスを明確にするとともに、国以外の機関が保持する政府ドメイン（go.jp）についての早期解消を含めたドメインの集約化等に係る計画を策定し、同計画に基づく見直しを実施する。

内閣官房は、各府省が策定した計画を取りまとめの上、公表する。

(2) 統一的なガバナンスの確保等

各府省は、上記に係る施策を府省内において統一的に推進するための体制を整備する。

内閣官房において、より利便性の高い公共サービスの実現に向け、官民協働で改革に取り組む体制の在り方等についての検討を行いつつ、本アクションプランを統一的に推進するために必要なガイドライン等を整備する。

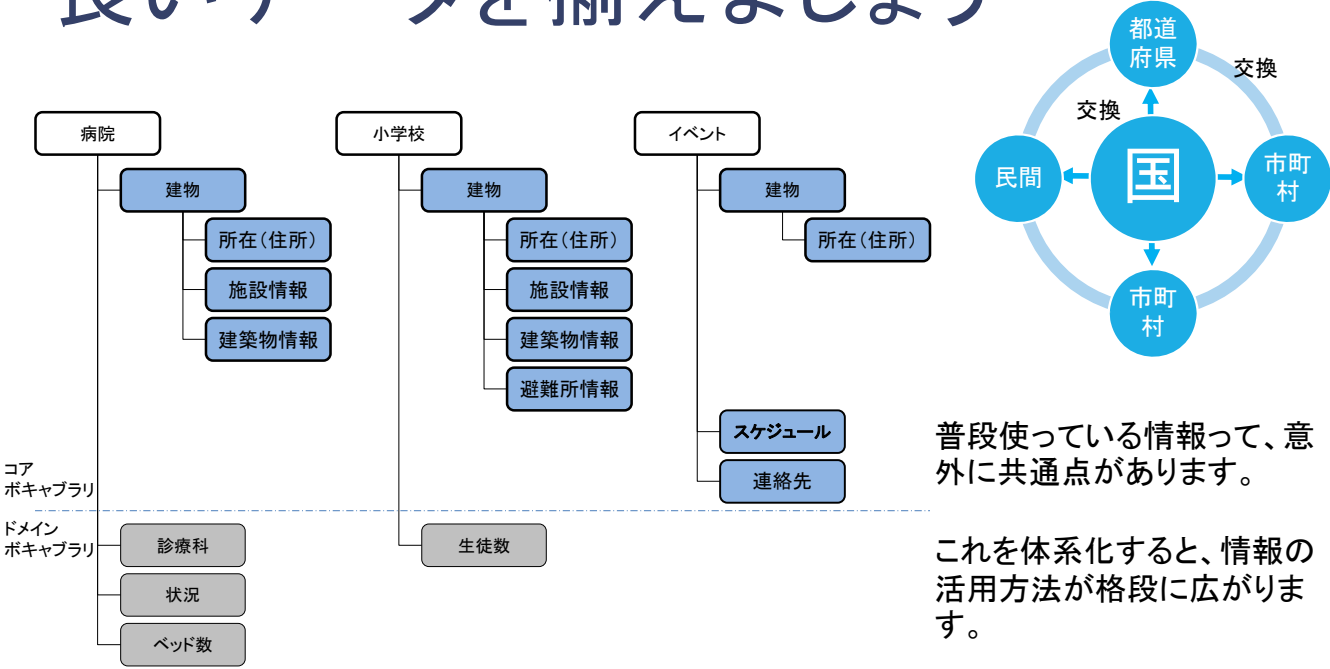
(3) 進捗状況のフォローアップ等

内閣官房は、本アクションプランに基づき推進する施策の進捗状況について、各府省から適宜報告・説明を求めつつフォローアップを実施するとともに、これを踏まえ、新戦略推進専門調査会電子行政分科会の意見も聴取しつつ、本アクションプランの必要な見直しを行うこととする。

共通語彙基盤

IMI Infrastructure for Multi-layer Interoperability

良いサービスを作るために、 良いデータを揃えましょう



この基盤を使うことで、サービスの検討や広範囲に使えるサービス構築が容易になります。

IT戦略「世界最先端IT国家創造宣言」(平成25年6月閣議決定)で、「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の基盤構築にも取り組む。」「国民がステークホルダーとして積極的に参加できるよう、(中略)、データ・フォーマット、用語、コード、文字等の標準化・共通化、APIの公開等を行いつつ整備する。」と今後の活用が国の方針として決定されています。

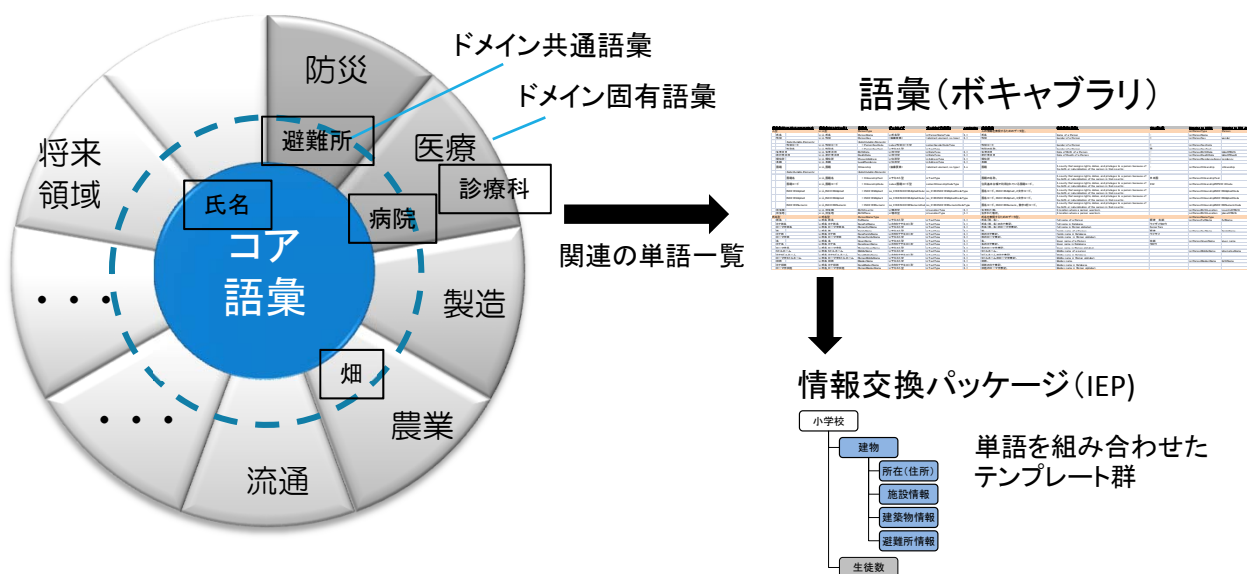
共通語彙基盤の内容

共通語彙基盤では、3つの語彙の階層を定義しています。

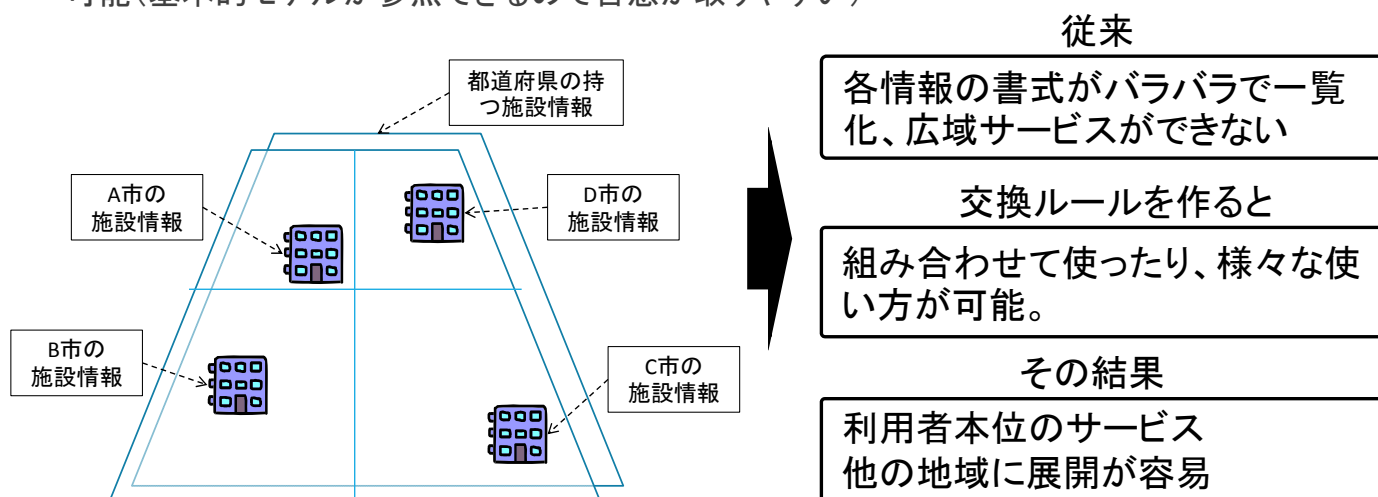
コア語彙 : 「氏名」「住所」等、どのようなサービスでも使う用語

ドメイン共通語彙 : 「病院」「避難所」など、複数の分野で使われる用語

ドメイン固有語彙 : 「診療科」「病床数」等、特定分野での専門用語



このように、情報交換するデータを定義、構造化することで、広域のサービスを迅速に連携可能(基本的モデルが参照できるので合意が取りやすい)



導入方法

アプリ開発や情報連携で活用すると、短い検討時間で、高品質なデータ設計をすることができます。

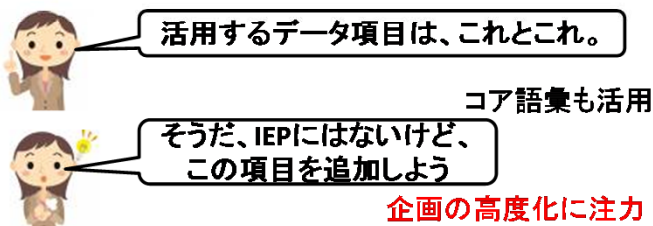
1. やりたいことを決定
2. 情報交換パッケージ(IEP)やコア語彙等の資料から見えそうな資料を選択



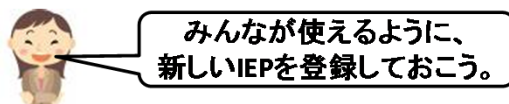
3. 利用するデータ項目を抜粋



4. 不足するデータを追加



5. 実装



6. (IEPを登録)

現在提供中の語彙・情報交換パッケージ(IEP)

氏名IEP	地物IEP	コア語彙
組織IEP	施設IEP	
住所IEP	避難施設IEP	
イベントIEP	医療機関IEP	
	設備IEP	

他に不足する語彙がある場合には、米国政府が整備するNIEMや検索業界が定義するSchema.orgも参照すると参考になる資料が手に入ることがあります。

<https://www.niem.gov/>

<http://www.schema.org/>

FAQ

Q:語彙って何ですか？

A:語彙は、特定範囲において使われる単語のグループの総体を言っており、システム間の情報連携をするための単語の総体として使っています。

Q:既存の用語集との関係はどうなっていますか？

A:専門分野や企業毎に用語集が整備されていることが多いです。それを変える必要はありません。外部と情報交換するときに、共通語彙基盤を活用してください。一度対応関係を整理すれば、その後は容易に変換できるようになります。

Q:業界内で用語を整備しているので必要性を感じないのだけれど？

A:最近のサービスでは業界横断や地域横断のサービスが増えてきています。そのような広範囲での情報連携の時に効果が生じます。

Q何を根拠に語彙を選び、情報交換パッケージを整理しているのですか？

A:行政や語彙の専門家が委員会で、既存の関連データ構造や実例を収集しデータの設計をしています。また、実データをサンプルとして入れて検証しています。

Q:コードの一覧は提供されないのですか？

A:主要なコードは情報交換パッケージに記述していますが、今後拡充していきます。

Q:国際的な情報交換はできるのですか？

A:諸外国の類似プロジェクトと協力しながらプロジェクトの推進をしています。対応表も整備していく予定です。翻訳も効率的にできるようにするひな形を整備する予定です。

Q:各種資料の更新頻度を教えてほしい？

A:現在のコア語彙はv2.2です。しばらく大規模なメジャーバージョンアップは予定していません。ドメインの追加等のマイナーバージョンアップは、適宜行う予定です。

Q:既存の語彙を当てはめようとすると、共通語彙基盤の語彙が細かすぎて、そこまで入力できません。どうしたらよいのでしょうか？

A:共通語彙基盤の用語は、用途に応じて取捨選択して活用するようになっていて、すべてを正確に合わせる必要はありません。従来データを活かして、大括りで情報を扱うことも可能です。現場に負担をかけず、徐々に移行していくことをお勧めします。

Q:データの入力や変換が大変なのですが、何か工夫はありますか？

A:データの入出力を支援するツールを整備中です。もうしばらくお待ちください。

文字情報基盤

IPAmj

行政機関で使う文字の問題は
これで解決

2015年4月以降有効



文字情報基盤を使うことで、あらゆる機関との文字情報の交換が簡単にできるようになります

IT戦略「世界最先端IT国家創造宣言」(平成25年6月閣議決定)の電子行政部分で、「文字の標準化・共通化に関しては、今後整備する情報システムにおいては、国際標準に適合した文字情報基盤を活用することを原則とする。」と今後の活用が国の方針として決定されています。

文字情報基盤の内容

IPAmj明朝フォント

- 拡大縮小が自由にできるOpentypeフォーマット
- 文字の画像ファイルも併せて提供

文字情報一覧表

- 各種文字コード間の関係や部首などの文字の基本情報を一覧で整理

文字情報基盤導入ガイド

- 業務によって使うべき文字範囲の例示など基本的方針

文字情報基盤導入テクニカルスタディ

- コード化完了までの暫定運用推奨案

文字縮退マップ

- 文字情報基盤からJIS第4水準までの縮退対応表

参考:変体仮名一覧(準備中)

- 変体仮名のフォントを提供し、文字情報基盤の運用を補助

導入事例

- 札幌市等での文字情報基盤導入実証の結果を整理

文字データベース(構築中)

- 文字をだれでも使いやすくするデータベース

導入によるメリット

- ・外字管理コストの削減
- ・幅広い業務での一貫した作業
- ・文字によるベンダロックインの回避
- ・目的別に文字を使い分けることで、行政サービスと行政効率化を同時に実現

文字

IPAmj明朝
フォント

変体仮名
フォント

文字
情報
一覧

一覧

ガイド

導入ガイド

テクニカル
スタディ

支援ツール

縮退マップ

導入事例

仕様例

文字データ
ベース

導入方法

文字に関する導入基本方針の決定

- 中長期にどのように管理していくかの方針を決める
- 導入は、組織内で一斉に行う必要はなく、順次実施してもよい

現在、保有している文字の棚卸

- 同定方針を決める(実施しながら整理していくのもよい)
- 各システムで持っている外字を文字情報基盤の文字に同定作業をする

導入ルールの整備

- どの業務で、文字をどの範囲(文字数)で提供するか方針を決める
- 縮退ができない文字をひらがな表記にするなどの方針の整理
- 外部連携にあたってのプライベート領域(PUA)利用方針の整理

仕様への記載

- 調達仕様文字情報基盤の活用を記載する。

適用業務と文字の範囲(参考)

		文字情報基盤で 整備された文字 図形 (IVS使用*)	文字情報基盤で 整備された文字 図形 (IVS不使用*)	JIS X 0213	JIS X 0208	常用漢字
組織内 に閉じた 業務	戸籍業務	○(*A)	△			
	住民記録業務	○(*A)	△			
	法人関連業務	○(*A)	△			
	生徒名簿等(字形に正確さが 求められる場合)	○(*A)	△			
	生徒名簿等(日常業務)			○	△	
	一般的な書類作成			△→○(*B)	○→△(*B)	
	災害等緊急時対応システム			△→○(*B)	○	
外部へ の 通知等	一般向け広報文書			△	○	○
	正確な人名表記を要する通 知文等	○(*C)	△(*C)			
	情報公開等	△(*C)	○(*C)	○	○	○
外部から の 申請等	一般的な申請等			△→○(*B)	○→△(*B)	
	正確な人名表記を要する申 請等	○(*D)	△(*D)			

(*A) 必要に応じ、文字情報基盤で整備された文字図形以外の文字図形を、「外字」として利用することが有り得るが、できる限り文字情報基盤の文字範囲に収まるよう、調整することが望ましい。

(*B) 市場にある一般的情報機器の対応状況に応じ、変化する。

(*C) JIS X 0213範囲以外の文字については、図形イメージによる表示、伝送を検討する。

(*D) JIS X 0213範囲以外の文字については、表示を図形イメージによるものとするほか、Web上に漢字入力を支援するサービスを設けることを検討する。

(*) IVS(字形選択子)を使用するシステムでは文字情報基盤で整備した全ての文字(約6万文字図形)を区別可能。使用しないシステムでは約5万文字図形を区別可能。

FAQ

Q:「文字情報基盤を活用することを原則とする」とは、6万文字を使える様にしろ、ということでしょうか？

A:いいえ、違います。業務に応じた適切な範囲を用いることが重要です(前ページの表を参照してください)。
また、文字情報基盤では、6万字近いフルセットからJISの範囲へ縮退するための情報(縮退マップ)を提供しています。
これを活用することで、用途に応じて、文字の利用範囲を自由に選択できます。
人名等の正確な表記が必要な業務では「外字」等を用いず、国際標準に適合した文字情報基盤の文字セットを活用することが重要です。

Q:縮退マップを使うと、6万文字のどの文字もJISの範囲に自動的に縮退できるのでしょうか？

A:いいえ。縮退候補が複数になる場合や、漢字一文字への縮退対応ができず、かなや複数漢字からなる単語に置き換える必要のある場合があります。そのような場合には、縮退変換に人間の介入が必要となるため、全自動変換はできません。

Q:外字って何ですか？

A:コンピュータに標準では入っていない文字を特別に追加した文字のことです。

Q:文字情報基盤を導入するのに条件はありますか？

A:文字情報基盤は、無料で誰でもご利用になれます。異体字をまとめて扱うことができるIVSの機能を使うには、windows7以降のパソコンを使う等の制限がありますが、その機能を使わなければ、制限はありません。

Q:ベンダから外字は自由に作れますといわれますが、そのほうが住民の要望に応えられるのではないですか？

A:独自の外字を作ると、運用にも費用が掛かることとなります。また、外部との情報連携ができないなど、運用に支障が生じます。新たな外字を作らずに、文字情報基盤に対応した文字を活用することで、これらの課題やベンダロックインを回避することができます。

Q:戸籍統一文字、住民基本台帳ネットワークシステム統一文字、登記統一文字との関係を教えてください？

A:戸籍統一文字、住民基本台帳ネットワークシステム統一文字は、文字情報基盤ですべて含んでおり、コード間の関係も一覧として提供しています。登記文字についても、必要な文字については、縮退マップを提供する予定です。また、文字情報基盤は、すべての業務に使える汎用文字です。

Q:文字情報基盤は、変体仮名を含んでいないのではないですか？

A:現在の文字情報基盤本体には変体仮名を含んでいません。それは、変体仮名には規範となる同定基準が存在しないためです。しかし、現場からの要望が強いため、国際標準化に着手するとともに、参考データとしての公開を開始いたします。

Q:ベンダが、文字情報基盤に対応できませんと言っていますが、どうしたらよいのでしょうか？

A:文字情報基盤は、技術的には導入は難しくありません。ベンダに対して、再度の確認をお勧めします。

Q:コード化されていない1900文字はどのように扱えばよろしいのでしょうか？

A:コード化が完了するまでの数年間は、暫定的にプライベートな外字領域(私用面:PUP)を使う方を業界団体が提示しており、これに対応したフォントも提供されています。ただし、これはあくまでも暫定的なもので、利用上の注意等について合意のとれた利用者の間だけで使用することが求められます。

共通語彙基盤

住所
情報交換パッケージ (IEP)
v1.0

平成 27 年 2 月 3 日

目 次

1. 本 IEP の目的	1
2. データ項目	2
3. IEP 活用の流れ	6
3. 1 新規システム構築の場合	6
3. 2 既存システムの場合	6
3. 3 自治体で中間標準レイアウトを使っている場合	7
4. コード等	9
5. サンプル	10

図 表 目 次

図 1 IEP を介したデータの連係	1
図 2 主なデータ項目	2
図 3 詳細データ項目	4
図 4 英語対応、schema.org 対応	5
図 5 新規システムでの導入イメージ	6
図 6 既存システムでの導入イメージ	7
図 7 サンプル	12

1. 本 IEP の目的

住所とは、行政区域に基づき表記される位置情報である。ビル名や階数等の方書情報を含む場合もある。住所には、住民票の住所、現住所、法人登記上の住所、本店所在地等がある。

多くの業務において住所の記述が求められるが、その形式は申請等のサービス毎に定義されている。例えば、「1 行で書かれる住所」、「都道府県・市区町村等の構造化した住所」が混在し統一されていない。そのため、システム内で住所データを保存していたとしても、申請を送る時に相手先毎に再入力もしくはデータ変換を行う必要がある。

そこで、住所を交換する時の共通的な交換方法が必要となる。既存のシステムのデータ構造を変更する必要はないが、相手先にあわせてデータ変換する時に参照すべきデータ一覧が必要となる。

本 IEP (Information Exchange Package:情報交換パッケージ) は、住所の情報を公開、交換、管理するために整備されたものである。実装に当たっては、共通語彙基盤の語彙セットを本 IEP と合わせて参照されたい。

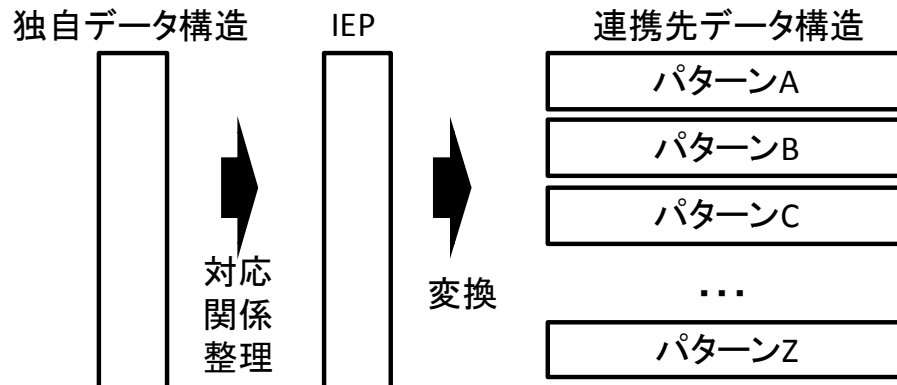


図 1 IEP を介したデータの連係

IEP 活用のメリット

- ・組織内外の多くのシステムと情報交換がしやすくなる
(調整やインターフェース改造が不要)
- ・情報項目に定義があるので、誰が見ても誤解なく情報が活用できる
- ・新規システム開発では、データ設計が標準的なデータでできる

世界的には通りを中心にした住所表示をする国が多いが、日本の面を中心とした住所特性を基本にして整備を行っている。

2. データ項目

データ項目と記述イメージは、以下の通りである。灰色項目はその下に記述された項目を組み合わせることで、アプリケーションで自動生成できる項目である。ただし、構造化した住所を使用しないで、直接記入することも可能である。

データ項目名	説明	備考
住所表記	都道府県、市区町村、町名を連続して記載。その後、丁目、番地、号をハイフンで接続して半角で記載。「東京都千代田区霞が関 1-3-1」	下記の構造化した住所をもとに自動生成することが望ましい。
郵便番号	郵便番号を半角、ハイフンなし、連番で記載。	
都道府県	都道府県を記載。	
市区町村	市区町村名を記載。郡の場合は町村と間を空けずに記載。「**郡&&町」	
区	政令指定都市の区を記載。	東京都の特別区は、上の市区町村に記載。
町名	地区の名称を記載。	字等、数字の前の漢字項目を記載。
丁目	半角で記載。	
番地補足	番地の前の文字列を記載。	丁目以降の数字の前に、「東」「北」「浜」「甲」「イ」等の文字が付く場合には記載。
番地	半角で記載。	
号	半角で記載。	町名の後ろに数字が1つしかない場合はここに記載。
ビル名	ビル名を記載「合同庁舎」	
ビル番号	半角数字で記載 4号棟は「4」。	ビル棟名が新館、〇〇棟数字以外で示される時にはビル名の後ろにスペースを空けずに記載。「〇〇省新館」等
部屋番号	半角数字で記載。「1234」	部屋名が文字の場合には、文字で記入。ただし、半角数字でない場合には、一部の処理でデータが処理されないことがある
方書	ビル名と部屋番号を連続して記載。「合同庁舎 4- 1234」	構造化したデータをもとに自動生成することが望ましい。

図 2 主なデータ項目

アプリケーションでは、図2のデータが利用者に見える。データベースでデータを管理する場合の IMI 対応は以下のとおりである。

ic:住所型					
ic:種別					現住所が基本であるが、「住民票上の住所」等住所項目を複数持つ時に使用する。
ic:表記					都道府県、市区町村、町名を連続して記載。その後、丁目、番地、号をハイフンで接続して半角で記載。
ic:国					国名を記載。
ic:国コード					国名コード（2文字）を記載。
ic:都道府県					都道府県名を記載。「東京都」「神奈川県」のように、都道府県まで記載。
ic:都道府県コード					全国地方公共団体コード（左2桁）を記載。
ic:市区町村					市町村名、および、東京都の場合は特別区名を記載。郡は、「〇〇郡〇〇村」と記入。北海道で総合振興局の記載が必要な場合には、市町村名の前に記載。
ic:市区町村コード					全国地方公共団体コード（右4桁）を記載。
ic:区					政令指定都市の場合は、区を記載。
ic:町名					町名、大字等、文字での住所記載部分を記載。
ic:丁目					半角数字
ic:番地補足					丁目以降の数字の前に、「東」「北」「浜」「甲」「イ」等の文字が付く場合には記載。
ic:番地					半角数字。
ic:号					半角数字。町名の後ろに数字が1つしかない場合はここに記載。
ic:ビル名					ビルに、「新館」等の日本語補足があるときには、「〇〇ビル新館」の要に記入
ic:ビル番号					半角英数字

ic:部屋番号				半角数字	
ic:方書				ビル名、ビル番号、部屋番号を連続して記載。	
ic:ID					
ic:住所コード					
	ic:コード種別				
		ic:名称		ID に名称がある場合に記載。	
		ic:発行者		ID を発行する組織名を記載。	
		ic:バージョン		ここで使用する ID のバージョンを記載。	
		ic:URI		ID に URI がある場合に記載。	
	ic:識別値			ID の実際の値を記載。この部分が一般には表示される。	
ic:郵便番号					

図 3 詳細データ項目

英語名は IMI の命名規則（NDR）に従い、単語の頭文字を大文字にして英単語を接続している。

また、検索エンジンでの用語の標準化している Schema.org との関係も下表に整理する。利用した Schema.org の分類は、以下の通りである。

ic:住所型	英語名	Schema.org
ic:表記	FullAddress	Thing>Place>address
ic:国	Country	Thing>Intangible>StructuredValue>ContactPoint>Address>addressCountry
ic:都道府県	Prefecture	Thing>Intangible>StructuredValue>ContactPoint>Address>addressRegion
ic:市区町村	City	Thing>Intangible>StructuredValue>ContactPoint>Address>addressLocality
ic:町名	AddressArea	(Thing>Intangible>StructuredValue>ContactPoint>Address>streetAddress)
ic:丁目	Thoroughfare	(Thing>Intangible>StructuredValue>ContactPoint>Address>streetAddress)
ic:番地	HouseNumber	(Thing>Intangible>StructuredValue>ContactPoint>Address>streetAddress)
ic:号	LotNumber	(Thing>Intangible>StructuredValue>ContactPoint>Address>streetAddress)
ic:ビル名	LocatorFull	
ic:ビル番号		
ic:部屋番号	LocatorName	
ic:郵便番号	LocatorDesignator	Thing>Intangible>StructuredValue>ContactPoint>Address>postalCode
ic:ID	PostCode	
ic:住所コード	AddressId	

図 4 英語対応、schema.org 対応

3. IEP 活用の流れ

3. 1 新規システム構築の場合

データ設計において、本 IEP の活用を検討する。本 IEP に対応することで、インターフェースのカスタマイズを最小限に抑えることができる。

データベース内のデータを詳細項目で記録して、入出力においては利用者に合わせた形で構築する。こうすることで、高い相互運用性と使いやすいインターフェースを両立させること。

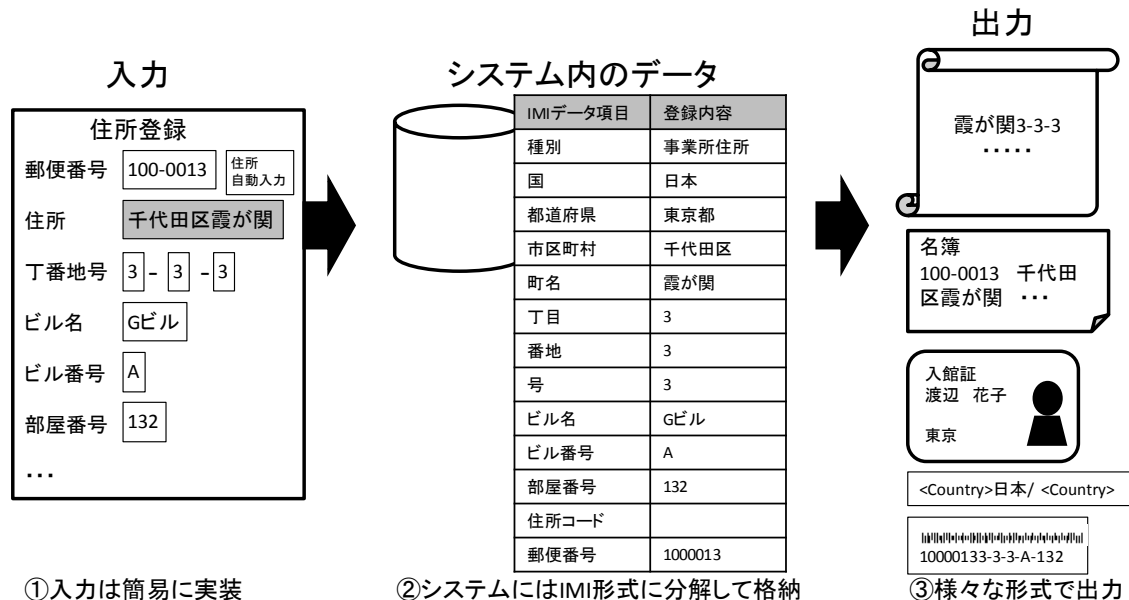


図 5 新規システムでの導入イメージ

(1) 手順1 データ洗い出し

新システムで必要なデータ項目の洗い出しを行う

(2) データ確認とサブセット作成

IMI の住所項目データを見て、過不足を検討する。そして、IMI のデータのうち使用する項目のサブセットを作成する。

(3) データ実装方式の検討

画面やデータベースでのデータ実装方式を検討する。郵便番号や町字コードから住所を自動入力するなどの工夫を行う。

3. 2 既存システムの場合

他システムとの情報交換や情報公開のためにデータのインターフェース設計 (API を含む) が必要な時に、本 IEP の活用を検討する。本 IEP を活用することで、2 回目以降の変換作業の負担を大幅に減らすことができる。

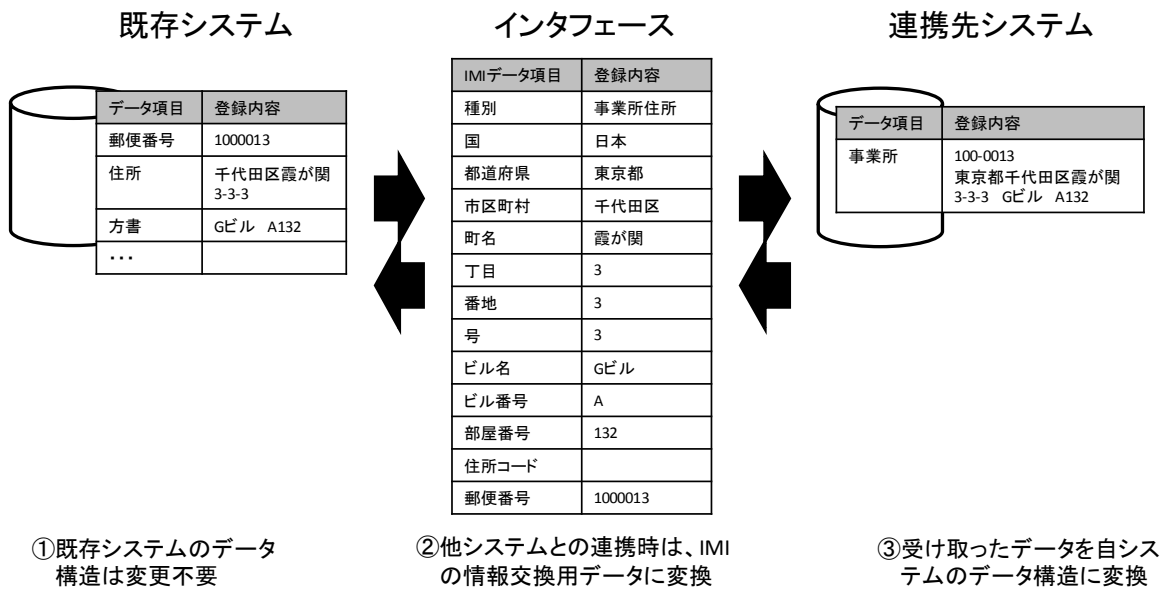


図 6 既存システムでの導入イメージ

(1) データ対応表の作成

自組織内で住所に関するデータベースを持っている時には、IMI のデータ項目に対して対応表を作成する。

(2) 情報交換用データセットに変換

情報交換先にあわせて、構造化された住所を、1行で記載する住所に変換したり、その逆に、1行で書かれた住所を分割したりする。

必要な場合には、漢字の住所データを元に、カナデータを生成する。

(3) IMI 形式データの受信と自組織のデータ形式への変換

データ連携先には、IMI 形式でのデータ送信を依頼する。その上で、IMI 形式で受信したデータを、自組織のデータ形式に変換する。

3. 3 自治体で中間標準レイアウトを使っている場合

自治体では、中間標準レイアウトの利用が推奨されている。その場合には、以下のルールで変換を実施する。

中間標準レイアウト	IMI	備考
住所区分	種別	以下から選択。 現住所、転入前住所、転出先、実定地、本籍地

住所コード	ic:住所_住所コード	町・字コードを記載。 ※町・字コードは有料であるため、IMI では任意使用項目である。
住所	ic:住所_表記	中間標準レイアウトでは、番地まですべて全角で記載。 IMI では、丁目以下は半角で記載。
方書	ic:定型住所_方書	
カナ方書		該当なし
郵便番号	ic:住所_郵便番号	半角数字7桁連番。

4. コード等

住所の表記には以下のコードが活用できる。

名称 : 国名コード (2文字)
 コード ID : ISO3166-1alpha-2,JIS X 0304-ISO3166-1
 発行者 : ISO
 発行者 ID :
 バージョン :
 URI (URL) : http://www.iso.org/iso/country_codes.htm

名称 : 国名コード (3文字)
 コード ID : ISO3166-1alpha-3,JIS X 0304-ISO3166-1
 発行者 : ISO
 発行者 ID :
 バージョン :
 URI (URL) : http://www.iso.org/iso/country_codes.htm

名称 : 全国地方公共団体コード
 コード ID :
 発行者 : 総務省
 発行者 ID :
 バージョン :
 URI (URL) : <http://www.soumu.go.jp/denshijiti/code.html>

名称 : 郵便番号
 コード ID : ISO3166-1alpha-2,JIS X 0304-ISO3166-1
 発行者 : 日本郵便株式会社
 発行者 ID :
 バージョン :
 URI (URL) : <http://www.post.japanpost.jp/zipcode/>

以下のコードは自治体での利用が行われているが、企業等の情報交換先が対応していない場合には使用しても交換に活用できない。

名称 : 全国町・字 (まちあざ) ファイル
 コード ID :
 発行者 : 地方公共団体情報システム機構
 発行者 ID :
 バージョン :
 URI (URL) : https://www.j-lis.go.jp/jyouhou-syori/machiaza/cms_11034.html

5. サンプル

サンプル1

データ項目名	説明	サンプル
種別		現住所
住所表記	都道府県、市区町村、町名を連続して記載。その後、丁目、番地、号をハイフンで接続して半角で記載。	東京都千代田区霞が関 1-3-1
郵便番号	郵便番号を半角、ハイフンなし、連番で記載。	1000013
都道府県	都道府県を記載。	東京都
市区町村	市区町村名を記載。郡の場合は町村と間を空けずに記載。「**郡&&町」	港区
町名	地区の名称を記載。	霞が関
丁目	半角で記載。	1
番地	半角で記載。	3
号	半角で記載。	1
ビル名	ビル名を記載「合同庁舎」	カスミビル本館
ビル番号	半角数字で記載 4号棟は「4」。	
部屋番号	半角数字で記載。「1234」	1208
方書	ビル名と部屋番号を連続して記載。「合同庁舎 4- 1234」	カスミビル 1208
種別		転入前住所
住所表記	都道府県、市区町村、町名を連続して記載。その後、丁目、番地、号をハイフンで接続して半角で記載。	京都府京都市下京区烏丸通高辻上る大政所町 682
郵便番号	郵便番号を半角、ハイフンなし、連番で記載。	6008413
都道府県	都道府県を記載。	京都府
市区町村	市区町村名を記載。郡の場合は町村と間を空けずに記載。「**郡&&町」	京都市
区	区名を記載。	下京区
町名	地区の名称を記載。	烏丸通高辻上る大政所町
丁目	半角で記載。	
番地	半角で記載。	
号	半角で記載。	682
ビル名	ビル名を記載「合同庁舎」	烏丸 A ビル
ビル番号	半角数字で記載 4号棟は「4」。	2
部屋番号	半角数字で記載。「1234」	201
方書	ビル名と部屋番号を連続して記載。「合同庁舎 4- 1234」	烏丸 A ビル 2-208
種別		転出先
住所表記	都道府県、市区町村、町名を連続して	1401 Constitution Ave., N.W.

	記載。その後、丁目、番地、号をハイフンで接続して半角で記載。	Washington,DC
郵便番号	郵便番号を半角、ハイフンなし、連番で記載。	20230
国	国を記載。	United States
都道府県	都道府県を記載。	DC
市区町村	市区町村名を記載。郡の場合は町村と間を空けずに記載。「**郡&&町」	Washington
町名	地区の名称を記載。	Constitution Ave., N.W.
丁目	半角で記載。	
番地	半角で記載。	
号	半角で記載。	1401
種別		実定地
住所表記	都道府県、市区町村、町名を連続して記載。その後、丁目、番地、号をハイフンで接続して半角で記載。	群馬県吾妻郡中之条町大字 小雨 577-1
郵便番号	郵便番号を半角、ハイフンなし、連番で記載。	3771795
都道府県	都道府県を記載。	群馬県
市区町村	市区町村名を記載。郡の場合は町村と間を空けずに記載。「**郡&&町」	吾妻郡中之条町
町名	地区の名称を記載。	大字小雨
丁目	半角で記載。	
番地	半角で記載。	577
号	半角で記載。	1
方書	ビル名と部屋番号を連続して記載。「合同庁舎 4-1234」	田中様方
種別		本籍地
住所表記	都道府県、市区町村、町名を連続して記載。その後、丁目、番地、号をハイフンで接続して半角で記載。	北海道札幌市北区北 24 条西 6-2-12
郵便番号	郵便番号を半角、ハイフンなし、連番で記載。	0018585
都道府県	都道府県を記載。	北海道
市区町村	市区町村名を記載。郡の場合は町村と間を空けずに記載。「**郡&&町」	札幌市
区	区名を記載。	北区
町名	地区の名称を記載。	北 24 条西
丁目	半角で記載。	6
番地	半角で記載。	2
号	半角で記載。	12
種別		本社所在地
住所表記	都道府県、市区町村、町名を連続して	石川県金沢市磯部町ニ 4 5

郵便番号	記載。その後、丁目、番地、号をハイフンで接続して半角で記載。 郵便番号を半角、ハイフンなし、連番で記載。	9200012
都道府県 市区町村	都道府県を記載。 市区町村名を記載。郡の場合は町村と間を空けずに記載。「**郡&&町」	石川県 金沢市
区	区名を記載。	
町名	地区の名称を記載。	磯部町
丁目	半角で記載。	6
番地補足	丁目以降の数字の前に、「東」「北」「浜」「甲」「イ」等の文字が付く場合には記載。	ニ
番地 号	半角で記載。	45

図 7 サンプル

共通語彙基盤
コア語彙テクニカルガイドライン

V1.0
2015/2/3

目次

1. 共通語彙基盤の概要.....	1
2. 語彙.....	2
2.1. 語彙の種類.....	3
2.2. 語彙の構造.....	4
3. 用語.....	5
3.1. クラス用語.....	5
3.2. プロパティ用語.....	5
3.3. 用語の構造.....	5
3.3.1. 項目名.....	5
3.3.2. データ型.....	5
3.3.3. カーディナリティ.....	5
4. コア語彙の設計と利用.....	6
4.1. コア語彙の設計方針.....	6
4.2. コア語彙の構成.....	7
4.2.1. 用語の階層構造.....	7
4.2.2. コア語彙の名前空間 URI.....	8
4.3. コア語彙の利用方法.....	9
4.4. コア語彙の物理形式表現.....	10
4.4.1. XML 形式.....	11
4.4.2. RDF 形式.....	12
4.4.3. 自然言語形式.....	13

1. 共通語彙基盤の概要

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ります。それら多くの用語を、どのような統一フォーマット（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の枠組みが「共通語彙基盤」（英語名称：Infrastructure for Multilayer Interoperability (IMI)）である。

「共通語彙基盤」は、用語における類義語や表記の揺れ、背景によって異なる意味を持つ同一用語など、「オープンデータ」の目的である電子的情報共有や、情報連携の効率化を進める際の障害を取り除き、データの組み合わせや横断的利用を容易にする。

語彙基盤の利用によって以下のようなメリットが得られる

- 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可能
- 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
- 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が発展

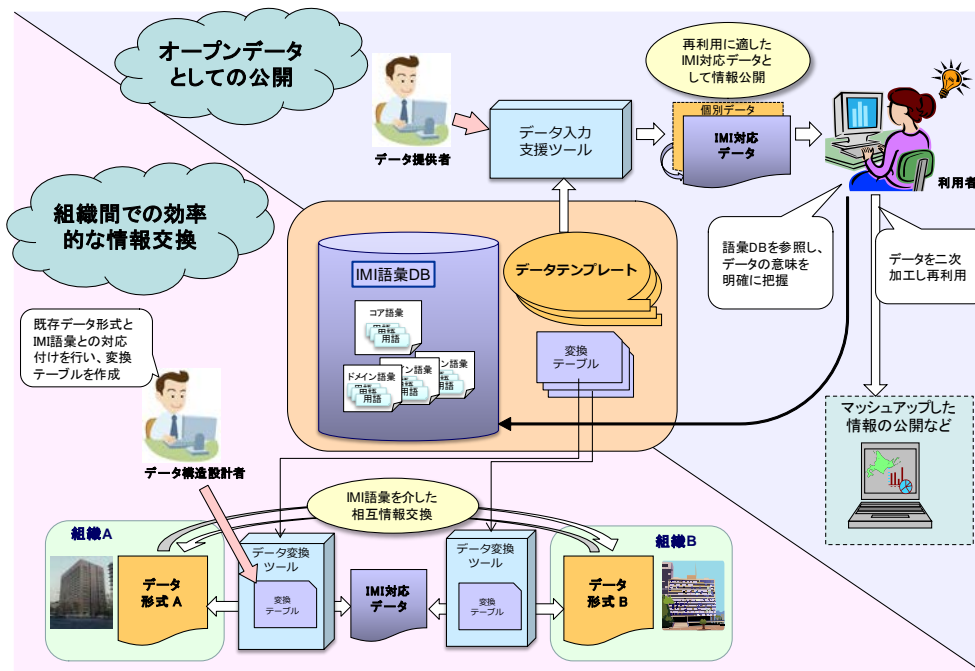


図 1. 「共通語彙基盤」の利用イメージ

2. 語彙

「語彙」とは、一般には言葉の集合を意味するが、共通語彙基盤における語彙は単に言葉を集めたものではなく、ファイルやインターネットなどを介して行われるコンピュータ間のデータ通信を円滑かつ確実にを行うことを目的とした、意味や構造、言葉と言葉の関係などを明確にした言葉の集合である。共通語彙基盤では、それぞれの言葉を「用語」と呼ぶ。

「用語」は、ある特定の意味あるいは概念を示すものであり、文字を組み合わせた表記を伴う。また、用語は、その意味あるいは概念の一部として、他の用語の関係（構造）を持つ場合がある。たとえば、「時刻」という用語を「時」、「分」、「秒」という下位の用語から構成すると定義することができる。

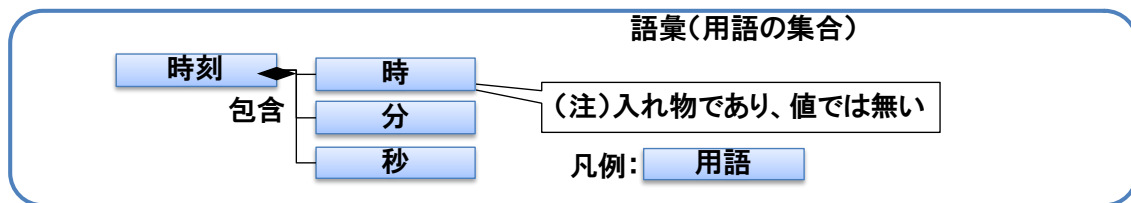


図 2. 語彙と用語

このような用語の集合が「語彙」である。用語の構造や意味、表記の制限、用語と用語との関係等、用語に係る様々な情報を総称してここでは「語彙データ」と呼ぶこととする。

2.1. 語彙の種類

語彙は、その示す内容によって「コア語彙」、「ドメイン語彙」に分類される。

ここで「ドメイン」とは、語彙を利用する目的・対象別に分割された領域の概念であり、その例に「地理空間・施設」「移動・交通」「防災」「財務」などがある。

【コア語彙】

どのドメインでも広く利用される普遍的な語彙。「人」「名称」「住所」「連絡先」「組織」「場所」「日時」「日付」「建物」「製品」「座標」など、特定のものを表現する語彙から、「数量」「面積」「重量」「長さ」「時間」「金額」のような計量のための基礎概念を表現する語彙まで含まれる。

【ドメイン固有語彙】

各ドメインでの利用に特化した語彙。ドメイン内の各種標準や既存システムの語彙などを含む。

ドメイン語彙には、領域固有の用語としてコア語彙とは完全に独立したものが入ることもあるが、コア語彙をプロパティとして用いたクラスとして領域に必要な用語を新たに作ることもある。

なお、ドメイン固有語彙の中で、他のドメインでも参照することのできる語彙で、それがコア語彙に含まれないものはドメイン共通語彙と呼ぶ。

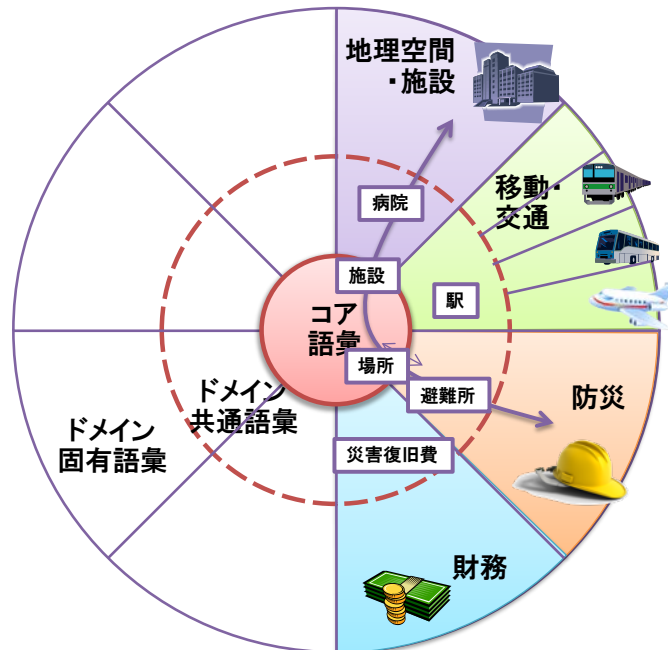


図 3. コア語彙とドメイン語彙

2.2. 語彙の構造

語彙は以下のような情報と共に用いられる。

1 名前空間

用語の名前が一意であることが保障される範囲。URI で示される。全ての用語はただ一つの名前空間に所属する。

2 名前空間接頭辞

用語が含まれる名前空間名前空間を識別するため、用語の前にコロン（:）を挟んで置く文字列。

名前空間接頭辞は、名前空間の URI にマッピングするための文字列に過ぎず、語彙を XML 形式や RDF 形式でデータ化する際に名前空間と接頭辞のマッピングを宣言して使う。つまり、名前空間の本質は URI で表されており、名前空間接頭辞はその一時的なエイリアスなので、データ化の都度名前空間 URI に対して異なる名前空間接頭辞を与えることも可能である。

また、語彙の管理のために以下の情報も必要である。

3 バージョン番号

コア語彙の用語の追加・削除などのメンテナンスに対応し、どの用語の集合を扱っているかを識別するための番号。

4 作成日

コア語彙の当該バージョンを作成（公開）した日付。

5 作成者

コア語彙の開発主体。

3. 用語

用語には以下のようにクラス用語とプロパティ用語がある。

3.1. クラス用語

用語の内、後述のプロパティ用語を包含するような構造を持った用語。他のクラス用語を拡張することで、拡張先のクラス用語が包含するプロパティ用語義することも可能

3.2. プロパティ用語

用語の内、クラス用語と他のクラス用語とを関連付けるもの。コア語彙においては、どのプロパティ用語もそれぞれ一つのクラス用語に包含される。

3.3. 用語の構造

それぞれの用語は、項目名、データ型、カーディナリティで表現される。

3.3.1. 項目名

用語の名称を日本語で表記したもの。コア語彙においては、用語の「項目名」は、UTF-8でエンコード可能な文字であって、以下に挙げた文字を使って記載されることと規定している。

- ・ 英数半角文字（記号を除く）
- ・ 全角文字（全角英数文字、及び、全角括弧文字'(',')'以外の全角記号を除く）
- ・ 半角アンダースコア ('_')（単語、語句、値のセパレータとしてのみ使用されなければならない）

3.3.2. データ型

プロパティのデータの形式。コア語彙で定義されたクラスに対応するクラスが指定される場合と、W3Cの「XML Schema Part 2: Datatypes」で定義されたデータ型など、既存のデータ型が指定される場合がある。

3.3.3. カーディナリティ

プロパティがクラスの中で何回出現できるかを最小出現回数、最大出現回数で定義したもの。

4. コア語彙の設計と利用

共通語彙基盤の語彙は、政府機関等から公開されるデータの 2 次利用や組織間でのデータ交換・連携を容易にするために、以下のような設計方針に基づいて作られている。

4.1. コア語彙の設計方針

データ交換やデータ連携を行う際に、「人」と「人物」のように異なる言葉が使われていると、コンピュータで処理する際に不都合が生じる。そこで、共通語彙基盤では、言葉の意味が明確になっている必要が生じる。

しかし、一つ一つの言葉の意味を明確にただけでは、機械処理に適したものとはいえない。「人」という言葉の意味が記述されているだけでは、「人」が何かということは分かるが、ある特定の人に関するデータを機械処理に適した形式で記述することができないためである。そこで、言葉の意味だけではなく、その言葉で表現すものを、関連するデータの集まりとして表現する必要がある。

このように、共通語彙基盤の用語は構造をもつが、実際にデータ公開やデータ交換を行う際に利用されるデータ形式が同じような構造をもっているとは限らない。データは種類や用途によって、固有な階層構造やフラットな形式をもったデータなど、様々な異なる形式で公開される。そのため、用語は、一般的に利用されているどのようなデータ形式においても利用できるものとなっている必要がある。

コア語彙では、クラス用語がプロパティ用語の集合を包含するとう構造を持たせることにより、まずは用語の関連性を階層的（XML など）あるいはネットワーク的（RDF など）な形式で表現することを可能とした。さらに、それをフラットに展開したデータ項目を生成することによって CSV のような形式にも対応することができる。

人			
説明: 人の情報を表現するためのデータ型			
継承: ic:実体型			
プロパティ:			
項目名	データ型	Cardinality	説明
ID	ic:ID型	0..n	ID
氏名	ic:氏名型	0..n	氏名
性別	xsd:string	0..1	性別の表記
性別コード	ic:コード型	0..1	性別コード
生年月日	ic:日付型	0..1	生年月日
死亡年月日	ic:日付型	0..1	死亡年月日
住所	ic:住所型	0..n	現住所
本籍	ic:住所型	0..1	本籍
連絡先	ic:連絡先型	0..n	連絡先
国籍	xsd:string	0..n	国籍の表記
国籍コード	ic:コード型	0..n	住民基本台帳で利用されている国籍コード
出生国	xsd:string	0..1	生まれた国名
出生国コード	ic:コード型	0..1	生まれた国のコード
出生地	ic:住所型	0..1	生まれた場所

図 4. コア語彙のクラスとプロパティによる論理設計

4.2. コア語彙の構成

以下にコア語彙に含まれる用語の相互関係、またコア語彙活用する際の識別子となる名前空間 URI について説明する。

4.2.1. 用語の階層構造

コア語彙のクラス用語はプロパティ用語の集合である。そのプロパティのデータ型がコア語彙の他のクラスを指すことにより、クラス用語の階層関係が生じる。

例えば、次の図の「人」クラスに含まれる「氏名」プロパティはデータ型が「氏名型」となっているが、「氏名型」は「氏名」クラスに対応したデータ型である。したがって、ここに「人」クラスと「氏名」クラスとの階層関係がみられる。同様に、「住所」プロパティのデータ型「住所型」によって「人」クラスと「住所」クラスとの間にも階層関係ができる。

※上記の例ではプロパティとそれに対応するクラスの名称が同じであるが、プロパティ名は必ずしもクラス名と同じである訳ではない。例えば、「人」クラスの「出生地」プロパティのデータ型は「住所型」であり「出生地」が「住所」クラスと結び付けられていることが分かる。

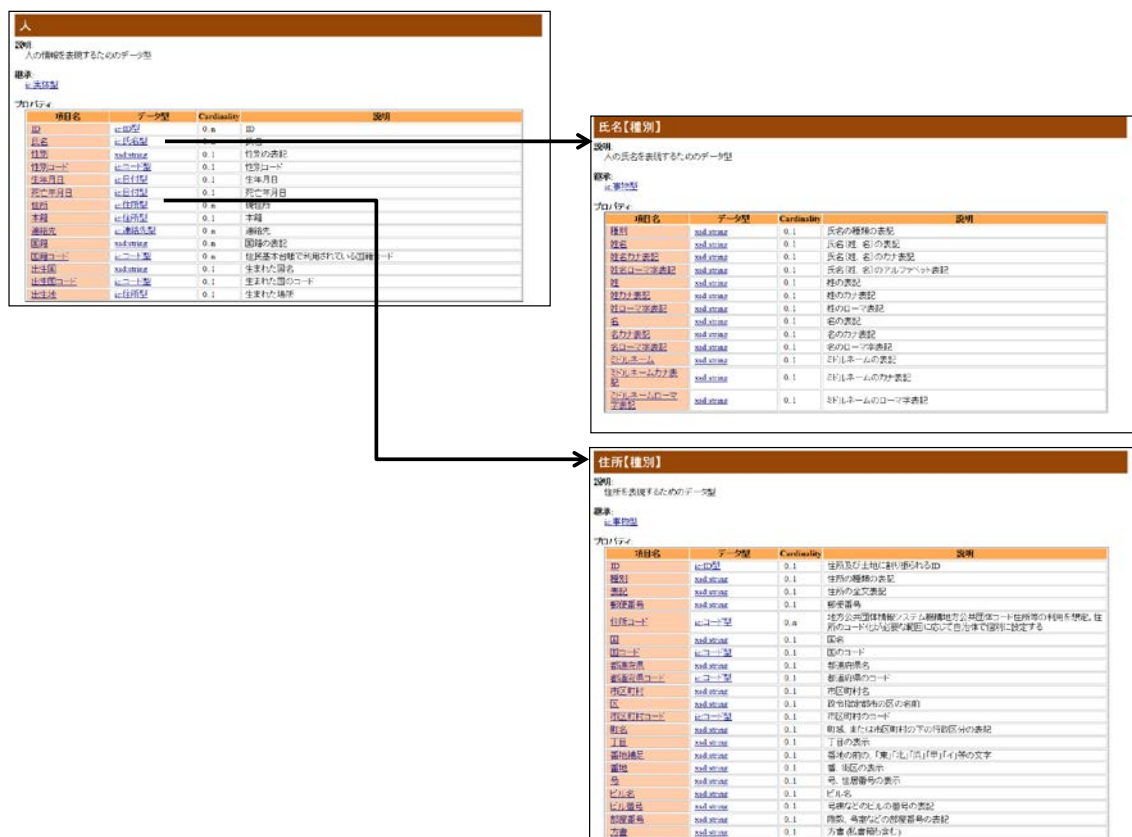


図 5. プロパティを介したクラスの階層構造

4.2.2. コア語彙の名前空間 URI

コア語彙 v2.2 には、それを識別するための名前空間 URI が付されている。ここで XML で使用する名前空間と RDF で使用する名前空間は、用途の違いから別のものを用いる。

XML : "http://imi.ipa.go.jp/ns/core/2"

RDF : "http://imi.ipa.go.jp/ns/core/rdf#"

RDF では、この名前空間 URI は、「ic:氏名」のような“接頭辞：項目名”というデータに対して接頭辞を置きかえることにより、実際にコンピュータ処理の対象となる"http://imi.ipa.go.jp/ns/core/rdf#氏名"という項目名が生成される。

このため、RDF では、名前空間 URI が変わるとデータ項目そのものが変わることになり、データのマッシュアップなどを行った場合に、データ項目が別のものになってしまう。このため、RDF では極力名前空間 URI は変えないことが望ましい。

一方、XML においては、名前空間 URI は、XML の要素（データ項目）の集合を識別するために用いられる。したがって、要素（データ項目）の追加、削除、などの変更があれば、集合の構成が変わるので、それが名前空間によって区別できることがコンピュータのアプリケーションにとって重要である。

このため、コア語彙の物理形式表現において、XML と RDF で異なる名前空間 URI を用いることとした。

4.3. コア語彙の利用方法

様々な業務領域では、それぞれのデータ項目およびその構造の設計が行われるが、コア語彙（およびドメイン語彙）は、それらの情報交換の要として活用される。

具体的な業務におけるデータ構造設計において、データベースに格納されたコア語彙（およびドメイン語彙）を参照して対応関係を付ける。このようにして定義された特定業務用のデータ項目が、後述するXMLやRDFなどのデータとして表現され、公開、交換された場合、コア語彙との対応関係をキーとして、他者にその意味（Semantics）を伝えることが用意になり、2次利用や組織間でのデータ交換・連携が可能になる。

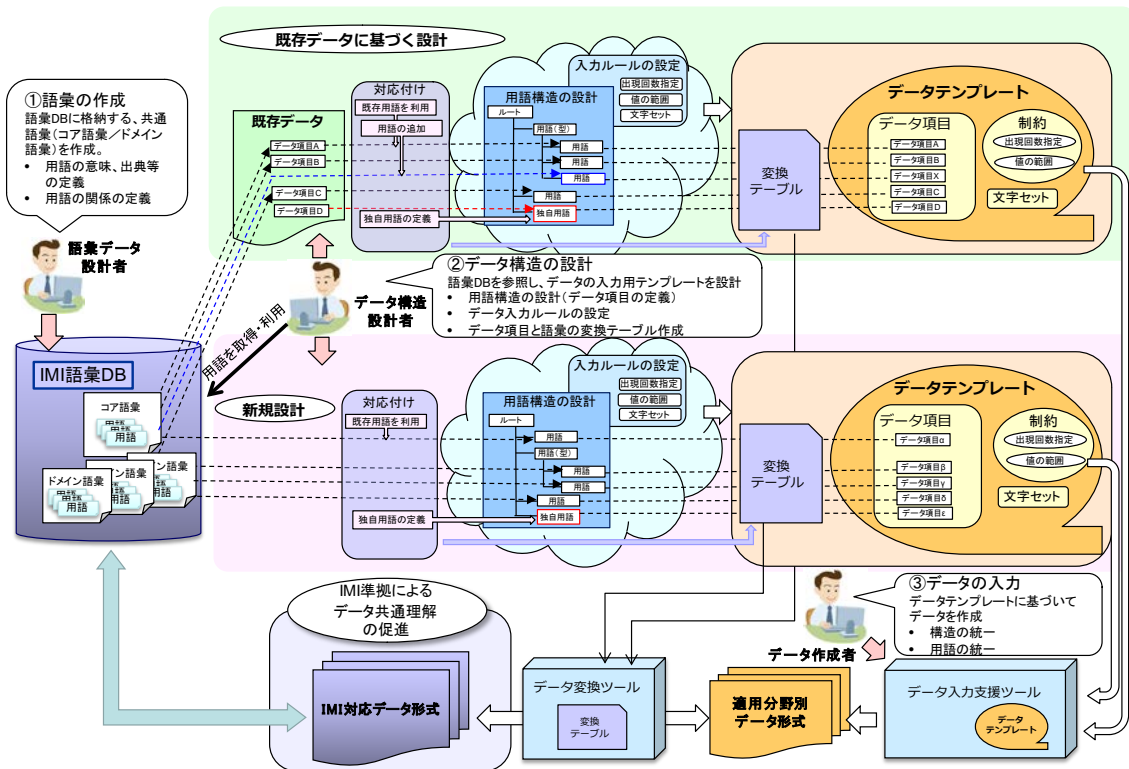


図6. コア語彙に基づくデータ活用の方法

4.4. コア語彙の物理形式表現

「用語」は概念を定義するものであり、抽象的な論理形式で表されているが、実際に使用する際には、XML や RDF など目的に合った物理的なデータ表現形式を利用する。これによって、データベースなどの XML や RDF の処理アプリケーションによる用語のコンピュータ処理が可能となる。

現在、語彙を物理的（機械可読）な形式で記述するための表現形式として以下の 3 種類がサポートされている。

①XML Schema 形式

XML 仕様や XML Schema 仕様に沿って、構造化データを定義するためのスキーマで語彙を表現した形式。

構造化情報を含めたデータ送受信や、可読性向上に利用される。

②RDF Schema 形式

RDF 仕様や RDF Schema 仕様に沿って、リソースのプロパティやタイプを定義するためのスキーマで語彙を表現した形式。

データ間の関係性記述や、オープンデータ、ビッグデータの表現形式に利用される。

③自然言語形式

XML や RDF 等の構造化仕様を用いずに、各用語を単純な（構造化されていない）自然言語として表現した形式。CSV データのヘッダ等として利用される。

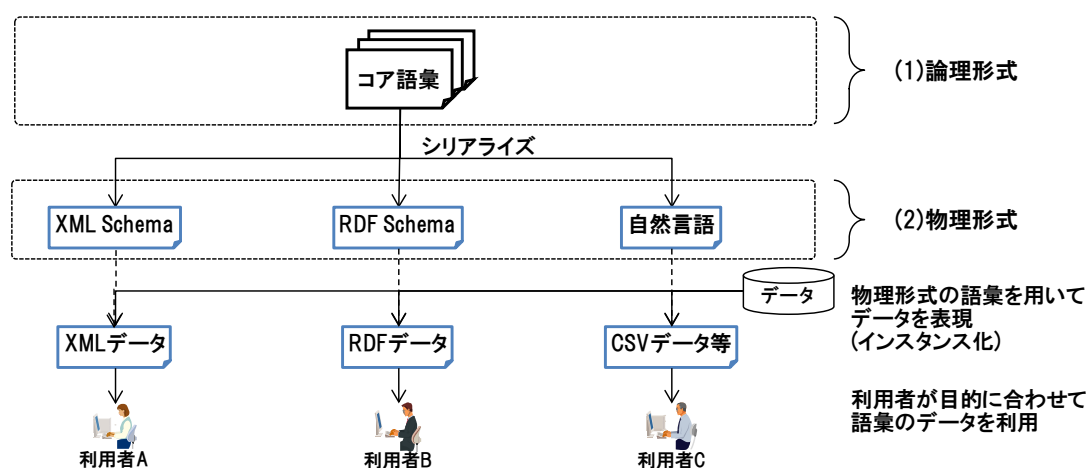


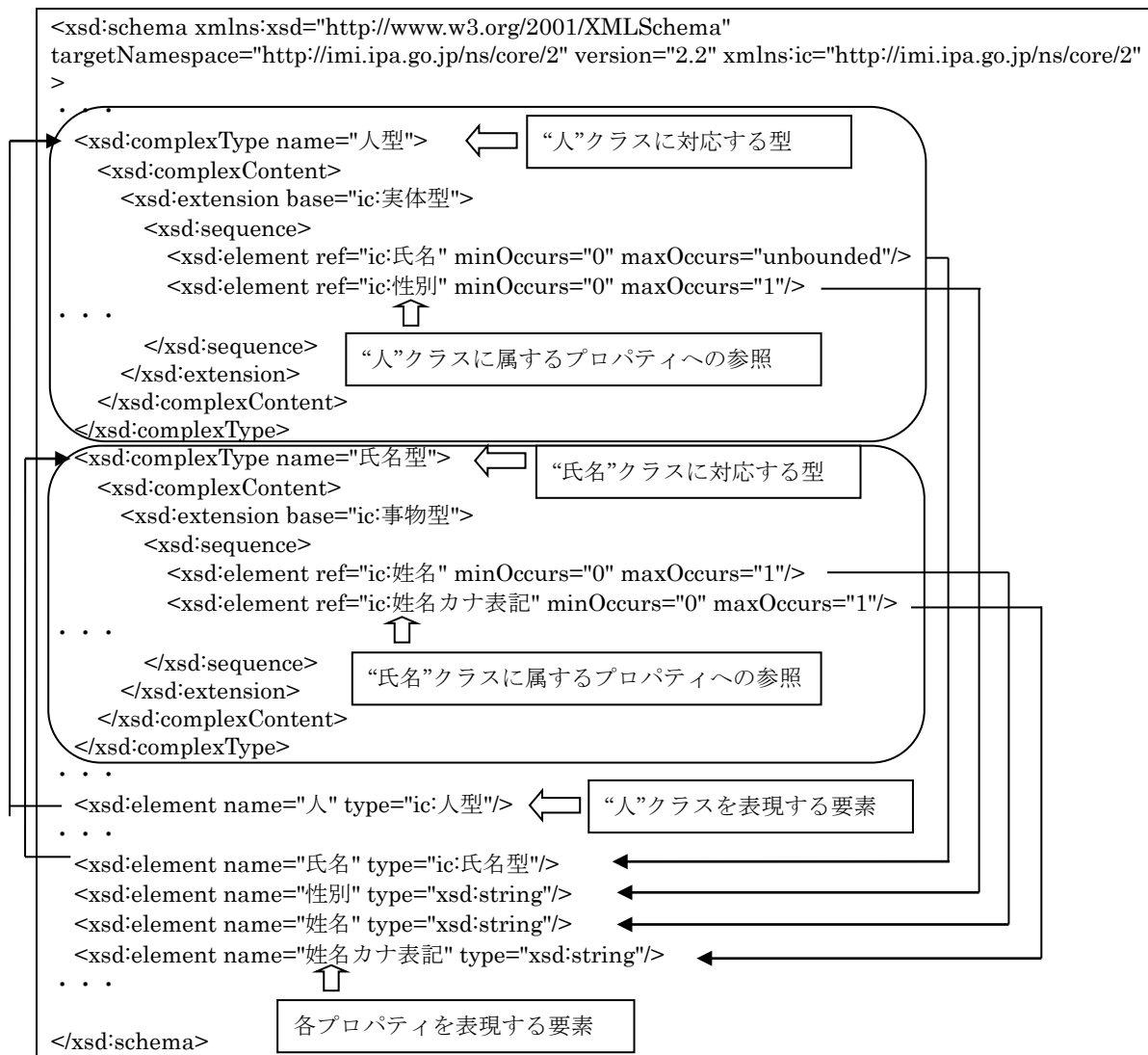
図 7. コア語彙の 3 つの表現形式

以下に 3 つの形式の事例をそれぞれ示す。

4.4.1. XML 形式

W3C XML Schema によるコア語彙の定義では、クラスはスキーマの「型」に対応させ、プロパティは「要素」として表す。

【XML Schema の例】



【XML インスタンスの例】

```

<ic:人>
  <ic:氏名>
    <ic:姓名>情報 太郎</ic:姓名>
    <ic:姓名カナ表記>ジョウホウ タロウ</ic:姓名カナ表記>
  </ic:氏名>
  <ic:性別>男性</ic:性別>
</ic:人>
  
```


4.4.2. RDF 形式

コア語彙の RDF 形式は OWL (Web Ontology Language) を使った定義を行う。

【RDF スキーマ (Turtle) の例】

```
@prefix ic: <http://imi.ipa.go.jp/ns/core/rdf#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix schema: <http://schema.org/> .
## プロパティ定義 #####
ic:氏名 a owl:ObjectProperty ;
    rdfs:label "氏名"@ja ;
    rdfs:comment "人の氏名"@ja ;
    schema:domainIncludes ic:人型 ;
    rdfs:range ic:氏名型 .
ic:性別 a owl:DatatypeProperty ;
    rdfs:label "性別"@ja ;
    rdfs:comment "性別の名称"@ja ;
    schema:domainIncludes ic:人型 ;
    rdfs:range xsd:string .
ic:姓名 a owl:DatatypeProperty ;
    rdfs:label "姓名"@ja ;
    rdfs:comment "氏名 (姓、名) "@ja ;
    schema:domainIncludes ic:氏名型 ;
    rdfs:range xsd:string .
ic:姓名カナ表記 a owl:DatatypeProperty ;
    rdfs:label "姓名カナ表記"@ja ;
    rdfs:comment "氏名 (姓、名) のカナ表記"@ja ;
    schema:domainIncludes ic:氏名型 ;
    rdfs:range xsd:string .
## クラス定義 #####
ic:人型 a owl:Class ;
    rdfs:label "人型"@ja ;
    rdfs:comment "人の情報を表現するためのクラス"@ja ;
    rdfs:subClassOf ic:実体型 .
ic:氏名型 a owl:Class ;
    rdfs:label "氏名型"@ja ;
    rdfs:comment "人の氏名を表現するためのクラス"@ja ;
    rdfs:subClassOf ic:事物型 .
```

【RDF インスタンスの例】

```
@prefix ic: <http://imi.ipa.go.jp/ns/core/rdf#> .
<#Person1> a ic:人型 ;
    ic:氏名 [
        ic:姓名 "情報 太郎";
        ic:姓名カナ表記 "ジョウホウ タロウ";
    ];
    ic:性別 "男性".
```

4.4.3. 自然言語形式

自然言語形式では、クラス名とプロパティ名を各助詞「の」で連結することによりデータ項目を表現する。インスタンスはデータ項目の並びに対応させた CSV 形式などを用いる。

【語彙の自然言語形式の例】

・・・
人の氏名の姓名
人の氏名の姓名カナ表記
人の性別
・・・

【自然言語形式に基づく CSV インスタンスの例】

"Person1","情報 太郎","ジョウホウ タロウ","男性"