異種データの利活用と行政システム間の情報連携を促進する共通語彙基盤

足立 和騎 Adachi Kazuki
高澤 亮典 Takazawa Ryosuke
柴田 博成 Shibata Hiroshige
矢戸 晃史 Yato Akifumi
山本 暖 Yamamoto Dan

近年，活動が活発化しているオープンデータの推進や行政システム間の情報連携の取り組みには，多様な意味や構造を持つ異種データの利活用が欠かせない。共通語彙基盤は，異種データの組み合わせや横断的なデータ利活用，扱えるデータ形式の異なる行政システム間のシームレスな情報連携を可能とする基盤である。日立グループは，共通語彙基盤を実現するため，データ形式の共通化に必要な語彙データを整備する技術，および語彙データを管理する語彙データベースの設計技術を開発した。この技術を，組織内のデータ統合，アプリケーション開発，オープンデータ作成などの場面に適用することで，組織の業務改善や社会に影響を及ぼす新サービスの創出に貢献する。

1. はじめに

近年，行政機関が保有する公共データの民間開放（オープンデータ）や，行政システム間の情報連携（データ統合・データ交換）を推進する動きが欧米を中心に活発化している。また，「世界最先端IT国家創造宣言」（2013年6月14日閣議決定，2015年6月30日改定）[1]では，オープンデータ推進および利便性の高い電子行政サービスの提供による行政の効率化や透明性の確保のほか，新ビジネスの創出・既存産業の生産性向上を目的に，データの組み合わせや横断的利用を容易とする共通の語彙（ボキャブラリ）の基盤構築に取り組むことが明記されている。この宣言に沿う形で，経済産業省および独立行政法人情報処理推進機構は，「共通語彙基盤（IMI：Infrastructure for Multilayer Interoperability）」[2]の構築プロジェクトを推進している。図1に共通語彙基盤の概要を示す。

ここでは，共通語彙基盤の概要，共通語彙基盤の実現に向けた日立グループの取り組み，および今後の展望について述べる。

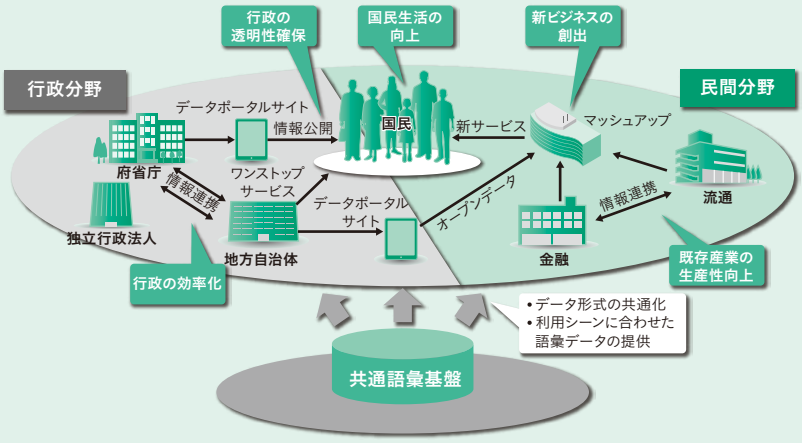


図1｜共通語彙基盤の概要
共通語彙基盤は，異種データの組み合わせや横断的なデータ利活用，扱えるデータ形式の異なる行政システム間のシームレスな情報連携を可能とする基盤である。

2. 共通語彙基盤の概要

2.1 国内外のオープンデータ・情報連携の動向

オープンデータについては，G8サミットで合意されたオープンデータ憲章を受け，政府はデータカタログサイト※1)「DATA.GO.JP」[3]にて，各行政機関が提供しているオープンデータを一括検索できるポータルサイトを公開した。2015年7月時点で，1万3,000を超えるデータセットが登録されている。

情報連携については，米国において，情報交換基盤であるNIEM（National Information Exchange Model）[4]の導入・活用により，国土安全保障省，司法省，保健福祉省を中心に，多数の行政システム間の情報連携が実現されている。また，国内においても，2016年1月から社会保障・税番号制度（マイナンバー制度）の利用が開始される予定であり，行政システム間の情報連携の推進が期待される。

2.2 異種データの利活用に向けた課題

昨今，オープンデータを提供する行政機関や，情報を連携しあう行政システムが増加しており，それに伴い，さまざまなデータ形式（用語の表記や構造など）で作成されたデータが利用されている。

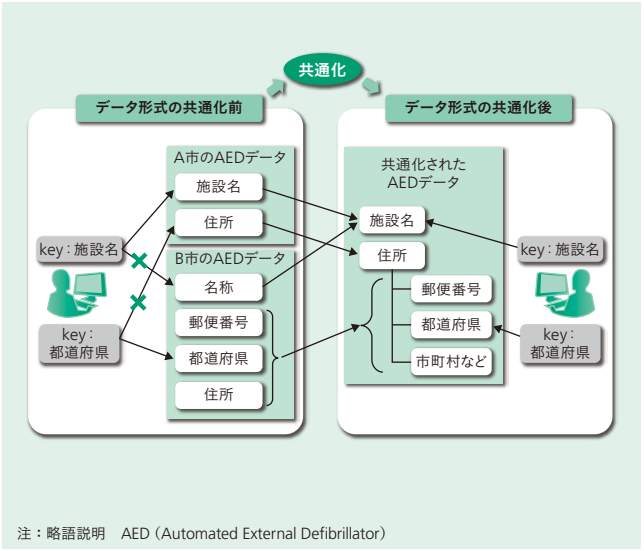
例えば，現状，各自治体が公開しているAED（Automated External Defibrillator：自動体外式除細動器）のデータを例にとると，AEDが設置されている施設の名称は，自治体ごとに「施設名」や「名称」といった異なる用語の表記で公開されている。

また，AEDが設置されている施設の住所は，郵便番号から番地までを一つにまとめた構造の住所情報もあれば，郵便番号，都道府県，市区町村，番地それぞれを細分化した構造の住所情報も存在する。

このように，同じ概念でも表記や構造がデータごとに異なる場合，人間は同じ意味だと理解可能であるが，コンピュータは異なる意味と解釈してしまう。そのため，多数の行政機関から公開された異種データの横断的な利活用や，行政システム間でのシームレスな情報連携は困難となる。

2.3 共通語彙基盤による異種データの利活用および行政システム間でのシームレスな情報連携の実現

前節の課題を解決するためには，用語に着目し，用語の表記や構造および意味，表記の制限，用語と用語との関係などを語彙データとして整理し，データ形式の共通化を図る必要がある。そのための仕組みとして，語彙データを管理する語彙データベースや，語彙データベースを利用するための各種API（Application Programming Interface）を備えた共通語彙基盤の整備が進められている。



注：略語説明　AED（Automated External Defibrillator）

図2｜データ形式の共通化
データの用語の表記や構造を共通化することで異種データの利活用を容易にする。

実際，前節のAEDの例では，共通語彙基盤が提供する語彙データにて，各自治体が公開しているAEDデータの形式を共通化し，全国に設置されているAEDを横断的に検索する「AED検索」アプリが開発されている[5]。図2に，AEDを例にしたデータ形式の共通化イメージを示す。

異種データの利活用および行政システム間でのシームレスな情報連携を実現するための準備として，データ形式の共通化に必要となる語彙データを整備するフェーズ，語彙データを利用し，各データ提供者が所有するデータから形式が共通化されたデータを作成するフェーズが必要である。

図3に，共通語彙基盤を活用した異種データの利活用および行政システム間でのシームレスな情報連携のイメージを示す。

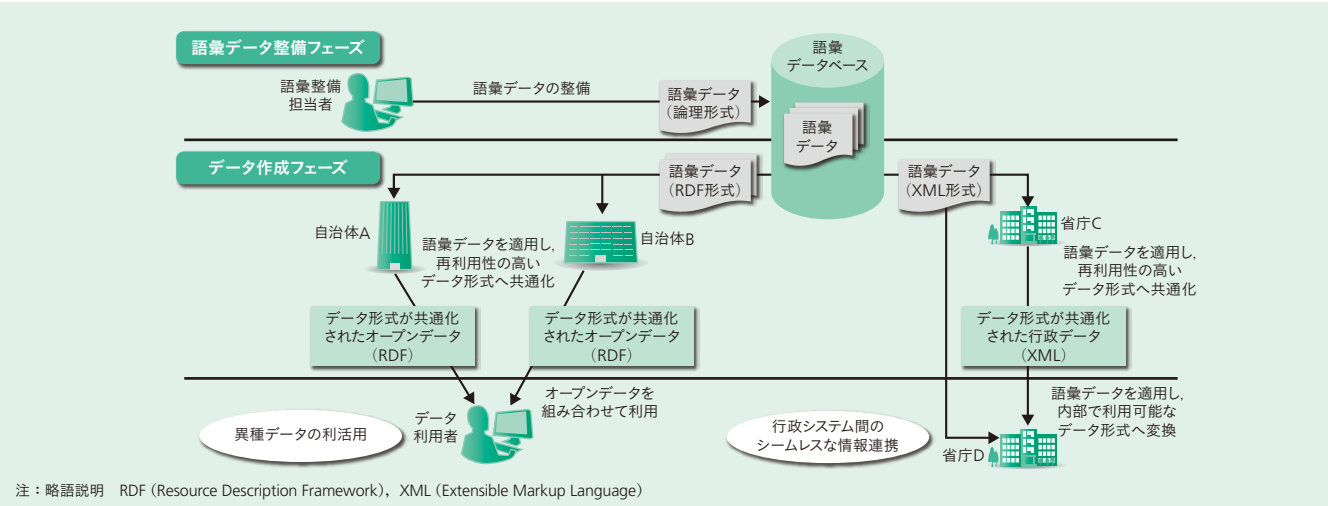
3. 共通語彙基盤の実現に向けた取り組み

3.1 共通語彙基盤の実現に向けた課題

日立グループは，情報処理推進機構の「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」プロジェクトに参画し，共通語彙基盤の在り方を概念モデルとしてまとめるとともに，実現に向けた課題を明らかにした[6]。以下に主な課題を2つ挙げる。

まず，語彙データ整備フェーズでは，さまざまな分野（交通，財務，法人など）における情報公開や情報連携の利用シーンを基に，用語の構造や意味，表記の制限，用語と用語との関係などを整理する必要がある。しかし，利用シーンの網羅や，その分野で使用される用語の意味などの整理

※1) データカタログサイト：データの取得を容易にするため，データの一覧表示，各項目の横断的検索，並べ替えなどの機能を備えたポータルサイト。



注：略語説明　RDF（Resource Description Framework），XML（Extensible Markup Language）

図3｜共通語彙基盤を活用した異種データの利活用および行政システム間でのシームレスな情報連携のイメージ

異種データの利活用および行政システム間でのシームレスな情報連携において，共通語彙基盤を活用することで異種データの形式が共通化され，データの利活用が容易になる。

には，その分野の専門知識が必要となる。

次に，データ作成フェーズでは，異種データの利活用や行政システム間のシームレスな情報連携など，さまざまな利用者や利用シーンに応じたデータフォーマットでのデータ作成が必要となるため，語彙データも各データフォーマットへの対応が必要となる。

3.2　情報公開・情報連携用語彙データ整備技術

語彙データ整備フェーズでの課題を解決するための技術として，日立グループは，情報公開・情報連携用語彙データ整備技術を開発した。

情報公開・情報連携用語彙データ整備技術とは，専門知識を持たなくても，情報公開・情報連携を目的とした大量の語彙データを高品質かつ効率的に整備する技術であり，その分野の有識者のサポートが十分に得られない場合でも，一定の品質を確保できる技術である。

日立グループは，この技術を活用し，独立行政法人中小企業基盤整備機構のプロジェクトに参画し，法人情報共有のための語彙データを，2014年11月から2015年2月の約4か月間で整備した。

具体的には，中小企業基盤整備機構内の業務システムに関する情報，中小企業に関連する申請書類，外部の企業情報サイトなどで取り扱っている企業情報の3分類で利用シーンを特定し，企業，株式会社などの用語を含む用語の収集，整理を行った。

以下に整備した語彙データの概要を示す。

・収集した利用シーン：2,504シーン
・抽出した用語数：6万1,337語
・語彙データに収録した用語数：1,304語

法人情報共有のための語彙データ整備において，情報公開・情報連携用語彙データ整備技術を適用した際のプロセスを**図4**に示す。

語彙データの整備にあたり，組織，名称，住所などの一般的な用語については，情報処理推進機構の「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」プロジェクトにて整備されたコア語彙[7]を参照・再利用することで，整備の効率化を図るとともに，他の語彙データとの相互運用性を確保することを可能とした。

3.3　語彙データベース設計技術

データ作成フェーズでの課題を解決するための技術として，日立グループは，語彙データベース設計技術を開発した。

語彙データベース設計技術とは，語彙データ整備フェーズにおいてデータフォーマットに依存しない論理形式で整備された語彙データを，Web上のデータの組み合わせや横断的な検索を容易にするRDF※2)（Resource Description Framework）フォーマットや，連携データの内容に不備や不整合がないことを簡単に検証可能なXML（Extensible Markup Language）フォーマットなど，利用者やその用途に応じて使いやすい物理形式のデータフォーマットで格納，提供するための技術である。

この技術を実証するため，情報処理推進機構の「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」プロジェクトでは，論理形式で整備された語彙データを格納し，APIを介して，物理形式で提供可能な語彙データベースのパイロットシステムを構

※2）RDF：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。

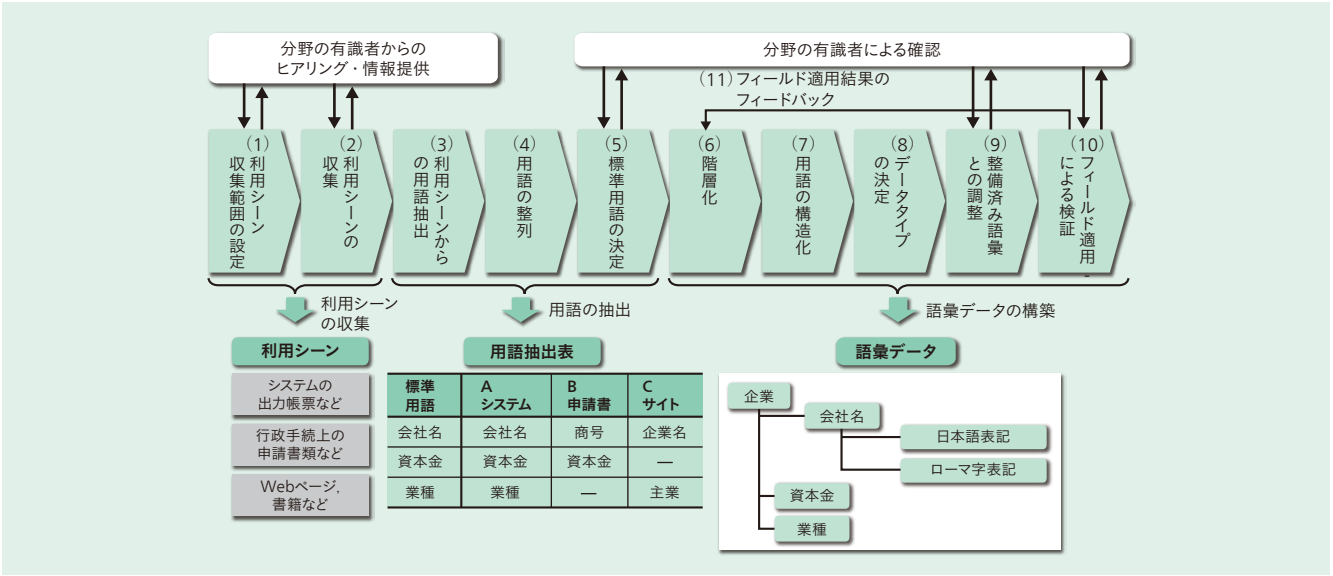


図4 | 語彙データ整備のプロセス

Webページなど実データのサンプル（利用シーン）の収集，用語の抽出，語彙データの構築の各作業を，分野の有識者と協調しながら実施する。

築・運用した。

パイロットシステムでは，施設・サービス，道路，観光，イベント，制度の5つの分野で試行的に整備された語彙データを格納し，各分野で試作されたツール群と連携して動作させることで，語彙データおよびAPIを介した物理形式での語彙データの提供の妥当性を検証した。

4. おわりに

ここでは，異種データの利活用や，行政システム間のシームレスな情報連携を容易化する共通語彙基盤の概要と，共通語彙基盤の実現に向けた日立グループの取り組みについて述べた。

日立グループでは，今後も共通語彙基盤の整備と活用のための技術開発を継続し，異種データの利活用と行政システム間の情報連携を促進する。今後，この技術の活用により，官公庁を中心とする各組織で実施が見込まれるデータ統合，アプリケーション開発，オープンデータやIoT（Internet of Things）データを含むビッグデータの利活用などを容易にし，組織の業務改善や社会に影響を及ぼす新サービスの創出を促進することで，社会イノベーションの実現に貢献していく。

参考文献など

1) 世界最先端IT国家創造宣言（2015.6），
https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20150630/siryou1.pdf
2) 共通語彙基盤概要，独立行政法人情報処理推進機構，
http://imi.ipa.go.jp/doc/IMI_Overview_v2.pdf
3) DATA.GO.JP，
http://www.data.go.jp/
4) NIEM，
https://www.niem.gov/Pages/default.aspx
5) AEDオープンデータプラットフォーム，
http://hatsunejournal.jp/w8/AEDOpendata/
6) 独立行政法人情報処理推進機構：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（2015.2），
http://goikiban.ipa.go.jp/contents/20150203_1_1.pdf
7) 独立行政法人情報処理推進機構：共通語彙基盤コア語彙，
http://imi.ipa.go.jp/ns/core/2

執筆者紹介

足立 和騎
日立製作所 情報・通信システム社 公共システム事業部
官公ソリューション第一本部 官公システム第四部 所属
現在，データカタログサイト「DATA.GO.JP」など，オープンデータ関連システムの設計・構築に従事

高澤 亮典
日立製作所 情報・通信システム社 公共システム事業部
官公ソリューション第一本部 官公システム第四部 所属
現在，公共システムの設計・構築に従事

柴田 博成
日立製作所 情報・通信システム社 公共システム事業部
官公ソリューション第一本部 官公システム第四部 所属
現在，公共システムの設計・構築に従事

矢戸 晃史
日立製作所 研究開発グループ システムイノベーションセンタ
セキュリティ研究部 所属
現在，公共システムの研究開発に従事
人工知能学会会員

山本 暖
日立製作所 研究開発グループ システムイノベーションセンタ
セキュリティ研究部 所属
現在，公共システムの研究開発に従事
電子情報通信学会会員