

4. 第三者参加オープン型サービス開発

オープンデータアプリ開発の公募
情報流通連携基盤システムの評価

第三者参加オープン型サービス開発

- データをオープンするだけで良いのではない。
- 誰もが簡単にサービスを開発できる環境を提供することが重要。
 - ▶ 開発環境プラットフォーム=開発者サイト



- サービスのCrowd Sourcing型開発手法
- ICTボランティアの導入
 - ▶ “Code for Japan”, “Hack for Japan”

開発者サイトの構築

- データを出すだけでなく、アプリ開発を積極的に支援する
- APIや提供データに関する情報、FAQを公開
- 登録者には、情報が更新される度に通知するサービスを提供



公募によるアプリケーション構築

■ 実施概要

- ▶ 実施期間 約3週間
- ▶ 構築されたアプリケーション数 16
- ▶ 開発者に公開したオープンデータ（公共交通情報）は以下の通り。

データの種別	項目
静的な運行情報	全国の鉄道およびバス路線情報 JR山手線・東京メトロ東京駅・都営バスの時刻表情報
リアルタイムな運行情報	JR各線・東京メトロ・都営地下鉄の運行情報（遅延・運休・ダイヤ変更・臨時ダイヤ等） 都営バスの走行位置情報
静的な公共交通施設	全国の鉄道駅・バス停情報 東京駅構内の改札・窓口・売店・店舗に関する情報（250種類／位置情報・名称）
リアルタイムな公共施設情報	東京駅10構内箇所に設置したセンサの値（温度・湿度・花粉量）

応募アプリ (1/3)



OpenData+RailMapping

- 山手線の運行情報を可視化
- 東京駅構内の指定した施設まで案内する電子コンパス
- 路線図から駅の状況を表示



トレバサ!

- 最寄りの列車・バス・駅・バス停を集めるゲーム感覚のアプリ



駅停ナビ

- 駅やバス停を通過する路線、施設情報を提供するアプリ



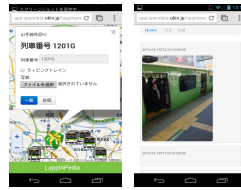
Busreg

- シルバースーツや車椅子スペースの空きを確認し、また確保をリクエストするアプリ



コントレ山手線

- 駅に停車している山手線車両のコレクションと、目覚ましタイマー機能を提供するアプリ



LappinPedia

- 山手線車両の中にあるラッピング電車を集めるアプリ

応募アプリ (2/3)



HyperTransfer

- 遅延情報を考慮した乗り換え案内



OriNavi

- 目的の降車バス停をナビゲートするアプリ



SpotNavi

- 観光情報とバス情報をマッシュアップしたアプリ



SuggestingSensor

- 東京駅構内環境を提案型で提示するアプリ



TokyoStationNavi

- 東京駅構内に設置されたNFCタグをタッチすると、周辺の施設を案内するアプリ

応募アプリ (3/3)



EkiSen

- センサによる温湿度情報を利用して、施設情報に環境情報のフィルタをかけるアプリ



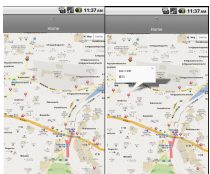
駅待ちアプリ

- 待ち時間にに応じて、構内の適切な立ち寄りスポットを紹介するアプリ



山手線及びバスの走行状況閲覧アプリ

- 山手線およびバスの走行状況を可視化するアプリ



BusMap

- バスの運行状況を可視化するアプリ



駅フィックスアプリ

- 駅構内の改善要求を出せるアプリ

5. 2013年度 ドコシル2.0

公共交通オープンデータ（ドコシル2.0）



一部の列車、バスのリアルタイムな位置情報の提供



「ドコシルなう」山手線

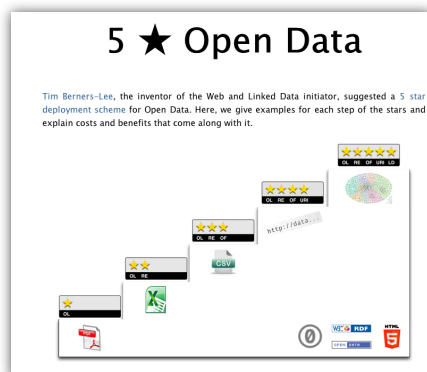


都営バス

PART 5 オープンデータを支える技術

Five Star Open Data (1)

- Tim Berners-Lee氏がオープンデータに向けたデータ公開の段階を示したもの (<http://5stardata.info>)



Five Star Open Data (2)

段階	データの形式	ファイル形式例
1	オープンライセンスで公開(形式不問)	PDF, JPEG, ...
2	コンピュータで編集可能な形式	DOC, XLS, ...
3	オープンに利用できるフォーマット	TXT, CSV, XML, ...
4	Web標準のフォーマット	RDF, XML, ...
5	他のデータへのリンクを含む	RDF (linked data)

オープンデータを支える技術（一般論）

■ Semantic Web Tech. ≒ LOD Tech. (Linked Open Data)

- ▶ RDF (Resource Description Framework)
- ▶ SPARQL: RDFの検索言語
- ▶ Ontologyによる世界記述

■ Web API

- ▶ RESTful API

■ ビッグデータ

- ▶ NoSQL (Not Only SQL)
 - ✦ KVS型 (Key-Value Store)
 - ✦ Document 型
 - ✦ DHT/P2P (Distributed Hash Table)
- ▶ DCN (Data Centric Network)

これらの技術の相性が意外と悪い

識別子 (ID)

■ データが対象としている個体 (Object, Entity) を識別する識別子の運用が必要

- ▶ 基本的に政府が保有するリソースには、政府が定めた識別子が付与されているべきではないか
 - ✦ 公共オープンデータはその識別子を使ってデータを記述するのが筋。
- ▶ 既に確立して、利用されている識別子体系はそれを利用する
- ▶ RDFでは識別子はURI Formatを利用することが規定
 - ✦ URIは識別子の書き方を決めていても、識別子そのものの規定ではない
- ▶ RDF以外のデータ規格での識別子は、依然として課題
- ▶ URI以外の識別子形式の国際標準もある
 - ✦ ucode (ITU-T Rec. 642-1:2012)
 - ✦ PI: Place Identifier (ISO 19155:2012)

■ 識別子の永続性

- ▶ 唯一性が保証できるか、識別子が使い回されないか、など

■ RFIDやバーコードなどタグへの格納可能性

- ▶ この場合、固定長であることが望ましい

【参考】RDFから利用可能なURI形式化できるID規格

種類	規格名	URI表規則	運営主体	説明	長さ	永続性	個体識別	RFID格納
汎用	ucode [ITU-T H.642.1]	urn:ucode: 0123456789ABCDEF0123456789ABCDEF	ユビキタスIDセンター	モノ・場所・概念などあらゆるものに付与できるIDである。IDの再利用を禁じているため、唯一性は永続的に保証される。	128bit	○	○	○
物流	EPC SGTIN (Serialized Global Trade Item Number)	urn:epc:id:sgtin: 4512345.167890.2 urn:epc:tag:sgtin-9 6.2.4512345.167890.2	GS1	商品を識別するコード。96ビットコードであるSGTIN-96では、ヘッダ(8ビット)、流通形態を表すフィルタ(3ビット)、パーティション(3ビット)、企業コード(20~40ビット)、アイテムコード(24~4ビット)、シリアル番号(38ビット)と続く。企業コードとアイテムコードは合計44ビットである。	96bit	△	○	○
電子データ	DOI (Digital Object Identifiers) [ISO 26234]	http://dx.doi.org/ 10.1021/jo0349227	国際DOI財団 (The International DOI Foundation)	インターネット上のドキュメントに恒久的に与えられる識別子。サーバの移行によるリンク切れを回避するため、DOIディレクトリを経由させている。学術論文の分野で広く使われており、学術雑誌や論文誌の記事に付与されている。書籍のタイトルだけでなく、任意のページや図表、CDの1曲ごとに付与することもできる。	可変	○	○	△
	UUID (Universally Unique Identifier) [ISO/IEC 11578]	urn:uuid:f81d4fae-7dec-11d0-a765-00a0c91e6bfe	なし(乱数)	分散システムにおいて、どこかが統制を取らなくても一意に識別できることを目的としたコード。現在よく利用されているのは、乱数に基づくversion 4である。ブログ等のコンテンツIDとして使われることが多い。	128bit	x	○	△
企業・組織	企業コード [ISO 6523]	urn:oid: 1.3.170.201233049	ISOが定めたICD (International Code Designator)	組織(企業)を識別するコードの付与方法をISO (ISO/IEC JTC1 SC32) が定めたもので、複数の企業コードや組織コードを包含することの出来るマルチコード。先頭の4桁がICDを識別する。それ以降の表記は、ICDが決定する。現在、150ほどのICDが登録されている。	可変	x	x	△

文字コード (漢字の扱い)

■ 漢字文化圏では、コンピュータ上での「漢字」の扱い方が大きな課題

- ▶ 特に、公共データにおいては、人名、地名などの漢字が、通常のJISコードやunicodeで表現できない部分がある。



■ 代表的な多漢字文字コード

- ▶ 日本における行政システムベンダ特有の漢字コード
 - ▶ unicodeの拡張コード
 - ▶ e漢字 (京都大学)
 - ▶ TRONコード (東京大学)
 - ▶ IPA漢字フォント (のコード)
 - ▶ 今昔文字鏡コード
- など

! データ規格によっては、文字コードを固定しているものがあり、日本の文化とコンフリクトを起すケースもある

(例) RDFはUnicodeを使うことになっているが、Unicodeの問題点は他方で指摘されている。

情報流通連携基盤

ボキャブラリ規定 (既存のボキャブラリ)

名称	規定範囲	ネームスペース	規定例
RDF基本構造	RDFでデータ構造を表現するための基本的なボキャブラリ。	http://www.w3.org/1999/02/22-rdf-syntax-ns#	rdf:subject(主語), rdf:predicate(述語)
RDFスキーマ	ボキャブラリを定義するためのボキャブラリ。	http://www.w3.org/2000/01/rdf-schema#	rdfs:subClassOf(サブクラス), rdf:range(値域), rdfs:subPropertyOf(サブプロパティ)
OWL	オントロジを記述するためのボキャブラリ。	http://www.w3.org/2002/07/owl#	owl:sameAs(同義), owl:inverseOf(反意)
ダブリンコア基本要素	書誌情報を記述するためのボキャブラリセットであるが、Webリソースの属性を記述するために広く用いられている。ISO 15836にて標準化。	http://purl.org/dc/elements/1.1/	dc:title(名前), dc:description(説明文), dc:creator(作者), dc:format(メディアタイプ)
DCMI語彙	ダブリンコア基本要素を拡張し、その意味を細分化したボキャブラリ。	http://purl.org/dc/terms/	dcterms:alternative(代替タイトル), dcterms:audience(対象としている利用者)
FoaF	人や組織に関する情報をRDFで記述するためのボキャブラリ。	http://xmlns.com/foaf/0.1/	foaf:familyName(姓), foaf:givenName(名), foaf:age(年齢)
SKOS	シソーラス、分類体系、件名標目表、タクソノミー、フォークソノミー、およびその他の同種の統制語彙のような概念体系の基本構造や内容を表現するためのモデルを提供するボキャブラリ体系。	http://www.w3.org/2008/05/skos-xl#	skos:definition(ボキャブラリの定義文), skos:broader(広義である), skos:note(ボキャブラリ定義に関するノート)
geoSPARQL	位置や形状に関するボキャブラリや、空間演算を行うための関数ボキャブラリが定義されている。	http://www.opengis.net/ont/geosparql# http://www.opengis.net/ont/sf# など	sf:wktLiteral(Well-Known Text規格の地理情報), sf:gmlLiteral(GML規格の地理情報)
DCAT	データセットを記述するためのボキャブラリが定義されている。	http://www.w3.org/ns/dcat#	dcat:theme(データセットのカテゴリ), dcat:accessURL(データにアクセスするためのリンク先情報)

(例) 情報流通連携基盤システム仕様

■ 目的: データをオープン化するためのモデル

- ▶ 自分のデータをオープン化するときの手法・システムの例示。
- ▶ さまざまな利用シーンを対象とする。
 - ◇データの直接取得・リアルタイムデータの取得・操作など。
 - ◇識別子を読み取ることをNotificationとしてデータを取得するような利用法にも対応する。

■ データ規格

- ▶ モデルはRDFに準拠する。
- ▶ 広く使われているボキャブラリは、そのまま利用する。 → 次頁
- ▶ 規定されていないが必要であるボキャブラリは、新規に定義する。
 - ◇物理量・単位系/地物関連/イベント関連/地理情報サービス関連/物品・製品関連/取引関連

■ API規格

- ▶ RESTベースのAPIとSPARQLベースのAPIを提供する。
- ▶ RESTベースのAPIでは、データ検索・取得コマンドのレスポンスにRDF/XML等を利用することにより、RDFモデルに基づくデータとの互換性を保つ。
- ▶ Streams APIに対応することにより、リアルタイムデータの送受信にも対応する。

API規格

機能名	概要	提供する理由
SPARQLベースのAPI		
SPARQL-based Command	SPARQL 1.1準拠のデータ操作APIを提供する。	RDFモデルに基づくデータに対するアクセスを提供するため。(既存の技術)
RESTベースのAPI		
Traceability/Realtime Data Management Command	トレースフォワード・トレースバックを含む、トレーサビリティに代表されるイベントを管理する機能。	対象応用分野の1つであるトレーサビリティで頻発に発生する、トレーサビリティイベントを効率的に扱うため。
Geographical Data Management Command	GIS等地理情報処理を必要とするデータ検索・取得・操作機能。	実物や実環境を扱う上で、地理情報演算は重要であるが、演算が複雑であるため。
Notification Management Command	データの登録・更新をNotificationとしてデータ利用者のシステムにコールドバックする(Notification)仕組み。	センサ情報の共有を目指すcomslにも同様の機能が提供されているため。
Security Management Command	ユーザ・グループの管理と、データのアクセスルールに関する機能。	民間データでは課金処理を要するケースがあり、ユーザ管理が必要である。comsやSODAにもユーザ管理機能が提供されている。
Vocabulary Management Command	ボキャブラリ情報の登録・検索・取得に関する機能。	モバイル環境や組み込み機器に対応するため、データの一部を登録・取得できる機能を提供する。
Triple Management Command	RDFモデルの主語・述語・目的語からなる基本データの登録・検索・取得に関する機能。	
Identification Resolution Command	IDをキーとしてデータを登録・検索する機能。	識別子を読み取ることをNotificationとしてデータを取得する利用法に対応するため。

オープンデータを推進した知見

オープンデータ推進から得られた知見

■ 様々な場面で「多様性」に直面

- ▶ 多様な技術/データモデル
 - ◇ RDF、Key-value Store、Geographical Data、.....
- ▶ 主体となる組織
 - ◇ 政府、地方自治体、公益企業、民間企業、学校等教育施設...
- ▶ 関係者の技術水準
 - ◇ Semantic Webの高い知識 ~ 全くのIT/ICT素人...
- ▶ ライセンスモデル
 - ◇ 改変への考え方、責任の考えかた
- ▶ ビジネスモデル
 - ◇ 完全無償 ~ 有償提供

■ 技術はまだ進展している

- ▶ データもAPIも今も新しい技術への転換が進んでいる。



■ 求められる技術、プラットフォームは何か？

「ビジネス」をリアルに推進するためには...

■ “Open and Free”は最終ゴールであるとしても...



■ そこに至るまでの過渡的には異なるモデルが必要

- ▶ 最終形と、明日何をやるべきかは、区別して考える必要がある。

■ “Open and Free” と “Proprietary”のマッシュアップが必要

- ▶ オープンソースでも同じ。



u2: uID 2.0アーキテクチャ

必要とされる技術要件

■ Semantic Webや語彙の技術はあたりまえ...



■ 多様なデータへの多様なアクセス

▶ 複数のデータベースモデルの連携 (Federation) 技術が不可欠

■ 多様なライセンス、ビジネスモデルへの対応

▶ パワフルなアクセス制御機構が不可欠

■ 規模、性能

▶ 大規模データを高速に扱える
→ 分散化を可能にするスケラビリティが不可欠

■ リアルタイムデータの扱い

▶ 検索に時間パラメータを扱えることが不可欠

■ 語彙間の不整合を解決する仕組み

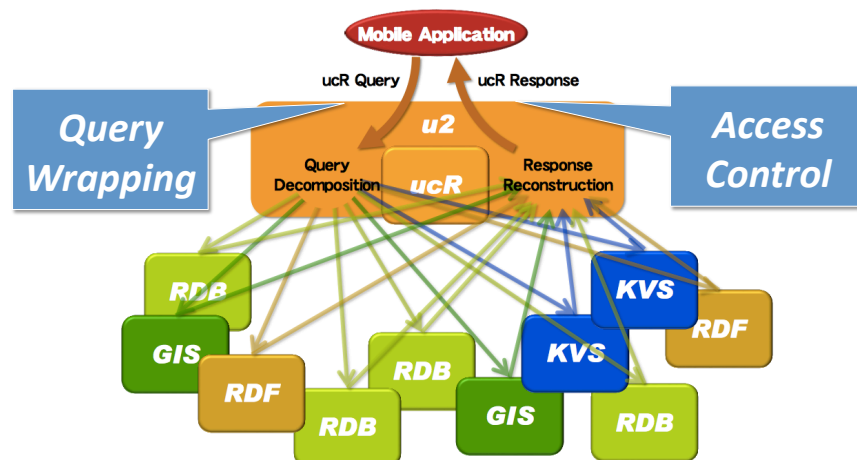
▶ 多様な品質のデータの整合をとる仕組みが不可欠

性能上の解決策 RDF/SPARQL の高速処理実装



応用毎に適したDBをFederationする
RDF/SPARQL I/FはWrapperとする

u2アーキテクチャ



for more information
koshizuka@sakamura-lab.org
<http://www.utacs.org/koshizuka/>