



OPEN DATA
オープンデータ流通推進コンソーシアム

第四回 利活用・普及委員会資料

オープンデータ流通推進コンソーシアム 技術委員会の検討状況報告

2013.3.13

オープンデータ流通推進コンソーシアム 事務局

技術委員会のミッションと実施内容

■ ミッション

(1) オープンデータを流通させるための技術仕様のあり方を検討

◇ データ規格

- ◆ オープンデータの表現モデル(オープンデータはメタデータとして表現される。それを表現するモデル)
- ◆ オープンデータを表現するためのボキャブラリ(メタデータを理解するための標準的な辞書)

◇ API規格

- ◆ オープンデータを取得・交換するための手法に関する規定

(2) 国際標準化のための作業検討

- ◇ ITU-T, W3C等の適切な組織での国際標準化

■ 実施内容

(1) 技術仕様のあり方に関する議論

- ◇ 仕様素案に関する意見交換
- ◇ 類似する既存規格・取組の整理と整合性検討
- ◇ 利活用・普及委員会などへの技術要件ヒアリングとその要件に対する技術仕様へのフィードバック検討
- ◇ データガバナンスのライセンス表現や、それに基づくデータ処理に関する技術的な検討(2年目以降)

(2) 国際標準化活動に関する意見交換

- ◇ 1年目は、標準化の動向調査と体制に関するヒアリング、意見交換

■ 想定アウトプット

- ▶ 活動報告書(標準仕様を含む)

技術委員会の論点

1. 本委員会が扱うオープンデータの全体像
2. オープンデータのデータ規格
 - ▶ 既存のデータ規格の調査
3. オープンデータアクセスのAPI規格
 - ▶ 既存のAPI規格の調査
4. データ規格・API規格のありかた(技術ガイド案)
 - ▶ 公共／産業界が保有する具体的なデータを事例とした、オープンデータ化を実施するための技術ガイド
5. コンソーシアム規格案
 - ▶ 情報流通連携基盤システム外部仕様書案(平成24年度版)
 - ▶ オープンデータ化のためのCSV形式データ規格案(平成24年度版)
6. ケーススタディ
 - ▶ 各実証実験からの評価(利用した技術・外部仕様書に追加した項目)
7. 国際標準化
 - ▶ 標準化の範囲と手順
8. 次年度以降の課題
 - ▶ 規格やサービスの維持・メンテナンスする組織のありかた
 - ▶ データ利用者・アプリケーション開発者向けツール(ライブラリやルーチンなど)、マニュアル等の整備
 - ▶ データホルダ向けツール(データ編集・変換ソフトウェアなど)、マニュアル等の整備
 - ▶ オープンデータライセンスをシステムが扱う(機械可読にする)手法
 - ▶ ヘルプデスク

今回の報告範囲

電子行政オープン
データ実務者会議
へのインプット

など

技術ガイドの概要

■ 目的 (Objective)

- ▶ 幅広いアプリケーションやサービスが有効に利活用するために、政府自治体、企業等、さまざまな組織が保持するデータをオープンデータ化するための技術的な要求事項、およびそれを実現するための手順を示す。

■ 対象 (Scope)

▶ 位置づけ

◇ 公開データの形式に関するガイド (本書)

- ◆ データを「どのような形式で」公開するかを規定

◇ データ公開ポリシー作成のためのガイド

- ◆ 「どんなデータを」公開するかを決めるための指針を規定

◇ データの信頼性確保のためのガイド

◇ データのライセンス策定のためのガイド

▶ 以下のデータを対象とする。

- ◇ 表形式データ
- ◇ 文書形式データ
- ◇ 地理空間データ
- ◇ リアルタイムデータ

技術ガイドの規定範囲

技術ガイドの規定範囲外 (別途定める)



技術ガイド案

技術ガイドの作成方針

■ 留意点(満たすべき条件)に「レベル」を設ける。

▶ レベル1(必須レベル)

◇オープンデータが必ず満たさなければならない要件。

◇レベル1の要件を満たすデータに対して、データ本体の中身を修正したり手を加えたりすることなく、かつ容易にそのデータを扱うプログラムが書ける。

▶ レベル2(推奨レベル)

◇オープンデータが満たすことを強く推奨する要件。

◇レベル2の要件を満たすデータを取得した利用者は、そのデータの項目(タイトル)・値・単位を正しく解釈した上で、データを扱うプログラムを書けることができる。

▶ レベル3(理想レベル)

◇オープンデータが満たすと望ましい要件。

◇レベル3の要件を満たすデータに対して、それを修正したり、手を加えたりすることなく、個別に作成された複数のデータセットを統合して利用(マッシュアップ)するプログラムが書ける。

◇そのために、データセットが含む個々のデータ本体の型や、データセット全体の構成や構造が、フォーマルに定義されていることを要件とする。

- ▶ レベル2・レベル3を満たすデータについては、機械が自動的にデータを取得・解読することにより幅広いアプリケーションでの利活用が期待される。

■ 公開されているデータを提示しながら、各レベルを満たすための要求項目をまとめた。

技術ガイド案(表形式データ)

■ レベル1(必須レベル)

▶ テーブル全体に対する要件

1. 1つのデータシートには、1種類の表のみを含むべきである。
2. セルに、整形のためのスペース・改行、位取りのカンマを含めるべきでない。
3. 年の値には、西暦表記を備えるべきである。
4. 数値やタイトル・単位以外の情報を、セルに含めるべきではない。

▶ セルに関する条件

5. すべてのセルが、他のセルと結合されているべきではない。
6. 値がない場合を除き、データセルが空白とすべきでない。

▶ タイトルに関する条件

7. データの内容を示すタイトルは、1行で構成されているべきである。
8. データの単位が明記されているべきである。
9. データセルの内容・単位・記数単位を示すタイトルが、それぞれ別の行に記載されているべきである。

▶ データの公開形式に関する要件

10. データセットは、オープンや標準データ形式で提供されるべきである。

技術ガイド案(表形式データ)

■ レベル2(推奨レベル)

▶ タイトルに関する要件

11. タイトルやデータ型は、一定の基準に従ったフォーマットで記述すべきである。→ 後述

■ レベル3(理想レベル)

▶ データの公開形式に関する要件

12. データセットの属性や説明を表すメタデータを、XMLやRDFの形式を使ってフォーマルに記述すべきである。そのメタデータからデータセット本体へリンクし、たどれるようにすべきである。
13. データセットに含まれるデータ本体を、XMLやRDFの形式を使ってフォーマルに記述すべきである。

技術ガイド案(文書形式データ)

■ 必須レベル(レベル1)

- ▶ 文章に存在する部・章・節・図表などの構造が、コンピュータが明快に認識できる形で記述されているべきである。
- ▶ 文章形式データに、整形のための符号や文字(空白・改行など)を含めるべきではない。
- ▶ 文書形式データは、オープンな標準データ形式で提供されるべきである。

■ 推奨レベル(レベル2)

- ▶ オープンに利用できるデータフォーマットで公開する。
 - ◇HTML、XML形式など。
- ▶ 文書形式データが図表を含む場合、それらを構成するレベル2以上の表形式データが添付されているべきである。

■ 理想レベル(レベル3)

- ▶ 図表やグラフとそのキャプションの対応関係がわかるように、文書形式データを構成すべきである。
- ▶ 文章形式データの属性や説明を表すメタデータを、XMLやRDF等の形式を使ってフォーマルに記述すべきである。そのメタデータは、文書形式データ本体へリンクすべきである。

技術ガイド案(地理空間データ)

■ 必須レベル(レベル1)

- ▶ 地理空間データは、測地系を明記すべきである。

■ 推奨レベル(レベル2)

- ▶ 地理空間データは、地理空間情報を記述するために広く使われているフォーマットで記述されるべきである。

■ 理想レベル(レベル3)

- ▶ 地理空間データの属性や説明を表すメタデータを、XMLやRDF等の形式を使ってフォーマルに記述すべきである。そのメタデータは地理空間データ本体へリンクするべきである。

技術ガイド案(リアルタイムデータ)

■ 必須レベル(レベル1)

- ▶ 表形式データ・地理空間データのレベル2以上に準拠した形式のデータを、ファイルとして取得できるべきである。

■ 推奨レベル(レベル2)

- ▶ リアルタイムデータの最新値・差分を取得する手法が提供されているべきである。

■ 理想レベル(レベル3)

- ▶ リアルタイムデータの最新値や差分を取得するための、メタデータ記述に対応したデータ取得規約が提供されているべきである。
または、メタデータ記述されたリアルタイムデータを取得する手法が提供されているべきである。



コンソーシアム規格案

(1) オープンデータ化のためのCSV形式データ規格案

- 表形式データのガイド11「タイトルやデータ型は、一定の基準に従ったフォーマットで記述すべきである」を満たすための規格の1つ。
 - ▶ CSVデータについて規定。他の形式(TSV、Open Document Format、XML等)については、必要に応じて、追って規定。
- 表形式データのキャプション・タイトル・単位などのメタデータを、データセットのヘッダとして記述する。

```

@Caption:都道府県別人口と人口増加率.ja,.....
@Creator:総務省統計局.ja,.....
@Date:2010-10-01,.....
@Language:ja,.....
@@Title:ja,.....
都道府県,2000年の人口,2005年の人口,2005年の人口集中地区の人口,2000~2005年の人口増減率,2010年の人口,2010年の人口性比(女性100に対する男性),2010年の人口密度,2005~2010年の人口増減率
@@Baseval,.....
,1000 ,1000 ,1000 ,,1000 ...
@@Unit:ja,.....
,,,%/km2,%
@@Datatype,.....
xsd:string,xsd:integer,xsd:integer,xsd:integer,xsd:double,xsd:integer,xsd:double,xsd:double,xsd:double
.....
全国,126926 ,127768 ,84331 ,0.7,128057 ,94.8,343.4 ,0.2
北海道,5683 ,5628 ,4108 ,-1.0,5506 ,89.7,70.2 ,-2.2
青森,1476 ,1437 ,653 ,-2.6,1373 ,88.9,142.4 ,-4.4
岩手,1416 ,1385 ,407 ,-2.2,1330 ,91.3,87.1 ,-4.0
宮城,2365 ,2360 ,1371 ,-0.2,2348 ,94.3,322.3 ,-0.5
秋田,1189 ,1146 ,386 ,-3.7,1086 ,88.5,93.3 ,-5.2
山形,1244 ,1216 ,504 ,-2.2,1169 ,92.2,125.4 ,-3.9
福島,2127 ,2091 ,806 ,-1.7,2029 ,94.3,147.2 ,-3.0
茨城,2986 ,2975 ,1068 ,-0.4,2970 ,99.3,487.2 ,-0.2
栃木,2005 ,2017 ,860 ,0.6,2008 ,98.6,313.3 ,-0.4
群馬,2025 ,2024 ,801 ,-0.0,2008 ,96.9,315.6 ,-0.8
埼玉,6938 ,7054 ,5566 ,1.7,7195 ,100.6,1894.2 ,2.0
千葉,5926 ,6056 ,4342 ,2.2,6216 ,99.4,1205.5 ,2.6
東京,12064 ,12577 ,12329 ,4.2,13159 ,98.0,6015.7 ,4.6
神奈川,8490 ,8792 ,8250 ,3.6,9048 ,100.9,3745.4 ,2.9
新潟,2476 ,2431 ,1139 ,-1.8,2374 ,93.6,188.7 ,-2.3
富山,1121 ,1112 ,398 ,-0.8,1093 ,92.9,257.4 ,-1.7
石川,1181 ,1174 ,573 ,-0.6,1170 ,93.4,279.5 ,-0.4
福井,829 ,822 ,333 ,-0.9,806 ,93.5,192.4 ,-1.9
山梨,888 ,885 ,305 ,-0.4,863 ,95.9,193.3 ,-2.4
長野,2215 ,2196 ,764 ,-0.8,2152 ,94.6,158.7 ,-2.0
岐阜,2108 ,2107 ,822 ,-0.1,2081 ,93.6,195.9 ,-1.3
静岡,3767 ,3792 ,2216 ,0.7,3765 ,97.0,483.9 ,-0.7
愛知,7043 ,7255 ,5480 ,3.0,7411 ,99.9,1434.8 ,2.2
...
    
```

ヘッダ

データ本体

ヘッダ名	意味
@Caption	データセットのキャプション
@Creator	データセットの作成者
@Date	データセットの公開日
@Language	データセットの基本言語
@@Title	タイトル
@@Unit	カラムの単位(物理単位・貨幣単位)
@@Baseval	カラムの記数単位
@@Datatype	カラムのデータタイプ

付与するヘッダ

本規格に基づくデータ記述例

(2) 情報流通連携基盤システム外部仕様書案(平成24年度版)

■ 目的: データをオープン化するためのモデルの提示

- ▶ 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。
- ▶ SPARQLベースのAPIとRESTベースのAPIを提供する。
 - ! 利便性よりも、むしろ機能性や意味の正確性などに重点をおく。

■ 規定範囲

- ▶ 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

外部仕様書のボキャブラリ規定(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)

外部仕様書の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。

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繁に発生する、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バックする(Notification)仕組み。	センサ情報の共有を目指すcoms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、ユーザ管理は必要である。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。

