



オープン&ビッグデータ活用・地方創生推進機構

平成26年度 第1回技術委員会資料

技術委員会活動報告

2014.12.24

オープン&ビッグデータ活用・地方創生推進機構 事務局

技術委員会

■ 目的

- ▶ オープンデータ推進に必要な技術標準の在り方に関する検討

■ 体制

▶ 主査

◇越塚 登（東京大学大学院情報学環 教授）

▶ 副主査

◇武田 英明（国立情報学研究所 教授）

▶ 委員

◇中尾 彰宏（東京大学大学院情報学環 准教授）

◇平本 健二（経済産業省 CIO補佐官）

◇深見 嘉明（慶應義塾大学SFC研究所 上席所員(訪問))

■ 活動概要

- ▶ 種々のオープンデータを流通させるための技術仕様の検討
- ▶ 国際標準化のための作業検討
- ▶ オープンデータを長期運用できる体制に関する検討

これまでの活動成果

■ 平成24年度

- ▶ Call for Commentを求められる技術文書を作成
 1. オープンデータ化のためのデータ作成に関する技術ガイド
 2. オープンデータ化のためのCSV形式データ規格
 3. 情報流通連携基盤システム外部仕様書(平成24年度版)
- ▶ 上記のうち1.と2.については「電子行政オープンデータ実務者会議」にインプット
 - ◇インプットした文書をもとにしたガイドラインが、現在「数値(表)、文章、地理空間情報のデータ作成に当たっての留意事項(ガイドライン別添)」として公開された。

■ 平成25年度

- ▶ 情報流通連携基盤システム・外部仕様書の改訂
 - ◇仕様書の利用者や委員からのコメントを反映
 - ◇参照している技術規格の成立経緯から、データの表現形式を見直し(RDF/JSON → JSON-LD)
- ▶ オープンデータガイドの作成(データガバナンス委員会と連携)
 - ◇平成24年度に作成した「オープンデータ化のためのデータ作成に関する技術ガイド」「オープンデータ化のためのCSV形式データ規格」に、利用ルールに関する検討内容やオープンデータの作成・公開手順を追加し、ガイドとしてまとめた。
- ▶ オープンデータ・アプリコンテストの審査(技術賞)
 - ◇民間における公共データの活用を促進するため、平成25年度に実施した7つの実証実験でオープンデータ化された公共データを活用した、アプリケーションの開発を一般公募により行う「オープンデータ・アプリコンテスト」を開催
 - ◇応募されたアプリケーション(92件)を技術的観点から技術委員会にて審査し、3作品を表彰

オープンデータ化のためのデータ作成に関する技術ガイド

オープンデータ化のためのcsv形式データ規格

■ 概要

- ▶ 政府・自治体・企業等が保持するデータを、オープンデータとして公開するときに技術的に留意すべき点や、それを満たすための手順を説明した文書。
- ▶ 表形式データ、文書形式データ、リアルタイムデータ、地理空間データの4種類について、3段階のレベルを設定してまとめた。

■ 作成後の経緯

- ▶ 「電子行政オープンデータ実務者会議」にインプット
- ▶ 平成25年度に作成した「オープンデータガイド」に取り込み

オープンデータ・アプリコンテスト

オープンデータ・アプリコンテスト 一般公募によるアプリケーション開発

開発者サイト公開・データ提供開始	2014年1月上旬～1月下旬
応募受付開始	2014年2月3日(月)
応募締切	2014年2月17日(月) 正午(終了致しました)
主催	オープンデータ流通推進コンソーシアム、総務省





情報流通連携基盤システム 外部仕様書

「情報流通連携基盤システム外部仕様書」の位置づけ

■ 背景

- ▶ さまざまなデータを情報通信ネットワークを經由して提供可能
 - ◇文書や統計に関するデータ
 - ◇センサによって計測されたデータ など
- ▶ 機械判読に適したデータ形式+二次利用が可能な利用ルール（ライセンス）により公開し、流通させようとする、オープンデータ化の動きの広がり
- ▶ 対象とするデータは多岐にわたり、その流通方法も数多く存在する



■ 目的

- ▶ これら各種のオープンデータを登録・利用するアプリケーションやサーバの構築方法を示すことにより、これらの構築を容易にすること。

以下、「情報流通連携基盤システム外部仕様書」を単に「外部仕様書」と呼ぶ。

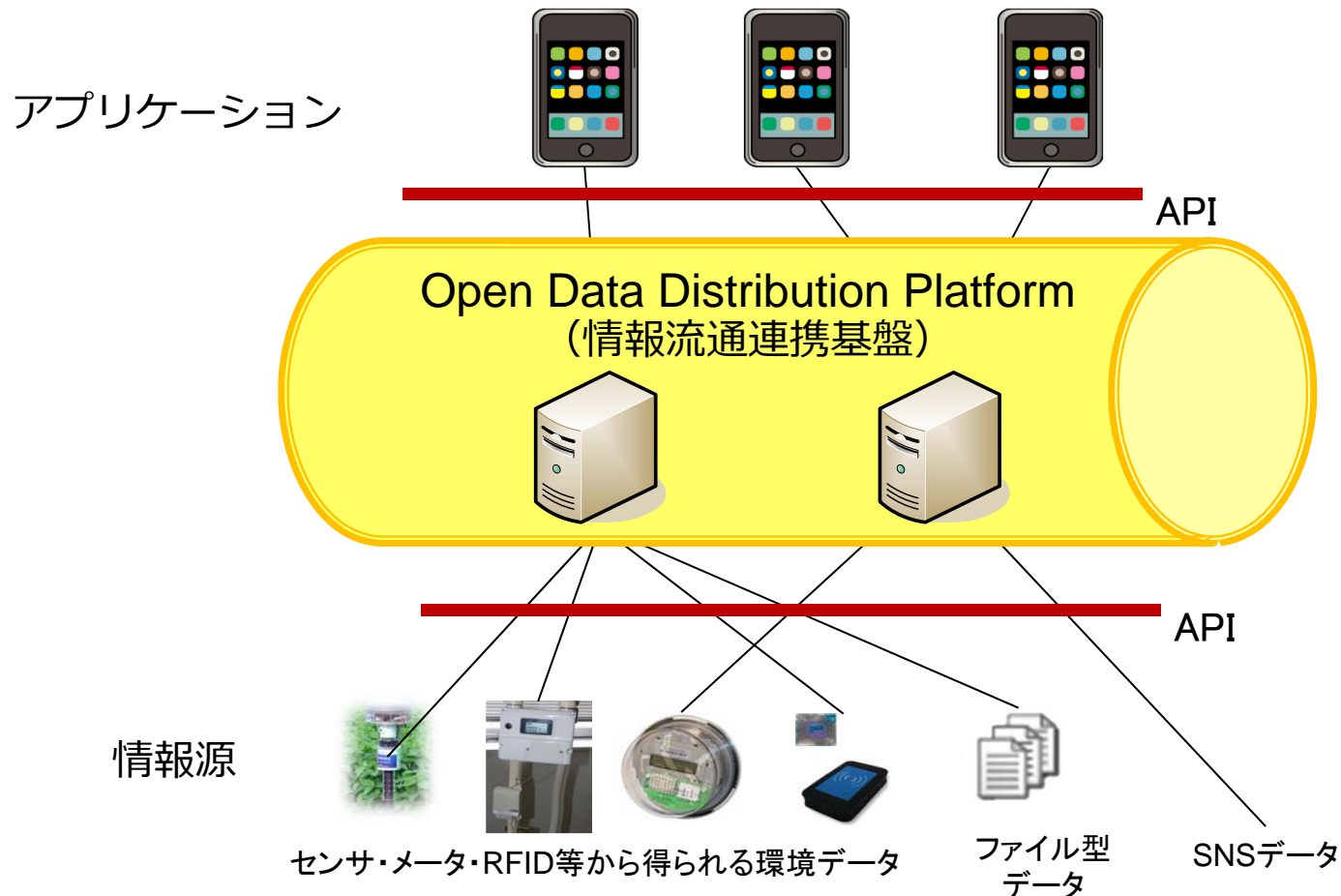
* 外部仕様書は、下記から取得できる。

<http://www.opendata.gr.jp/cfc/>

情報流通連携基盤の全体像

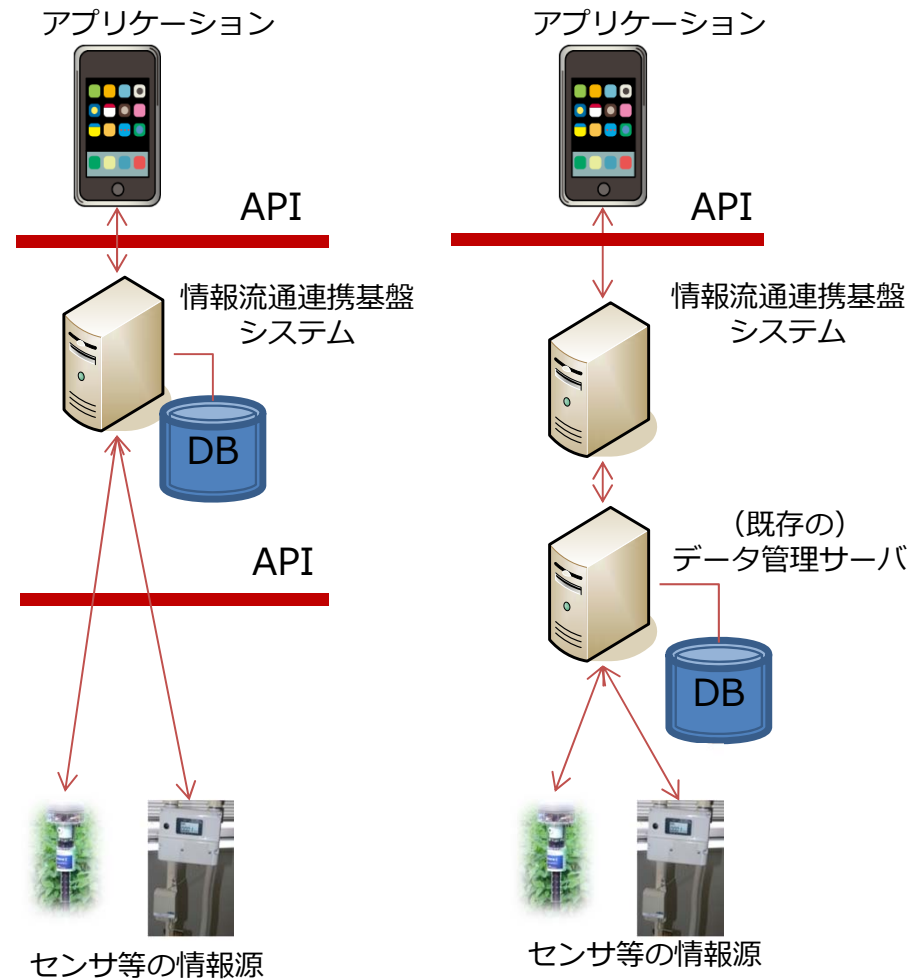
■ 情報流通連携基盤（Open Data Distribution Platform: ODDP）とは

- ▶ データを登録・利用するアプリケーションの構築を流通・連携させるための、汎用性を持つ技術・運用ルールが整った環境



外部仕様書の規定方針・特徴

- 外部仕様書が規定するもの
 - ▶ データ規格
 - ▶ API規格
- 外部仕様書が規定しないもの
 - ▶ データベースの実装方法
 - ▶ システムの実装方法
- 既存の規格との互換性を考慮
- 規格の選択利用・拡張の許容
 - ▶ 外部仕様書に記載された機能のうち必要なものを選択して、アプリケーションやサーバを実装してよい。
 - ▶ ユーザビリティの確保や性能向上のため、独自の拡張を行ってよい。
 - ▶ ただし、これらの選択・拡張に関する仕様を開発者に提示すべき。



外部仕様書の規定内容

■ データ規格（ODDPデータ規格）

- ▶ データモデルは、RDFデータモデルに準拠する。
- ▶ 表現形式は、RDF/XML、Notation3、N-Triples、Turtleなどを利用する。
- ▶ データの意味を共通に理解するための辞書である、ボキャブラリを規定する。

■ API規格（ODDP API規格）

- ▶ 2種類のAPI規格を規定する。
 1. SPARQLベースのAPI
 2. RESTベースのAPI

対象データとその識別

■ 対象とするデータ

- ▶ 文書・表形式データ・画像・動画・音声など、データの格納されたファイル
- ▶ 上記を解釈し、RDF 形式に変換したデータ
- ▶ センサによる計測データ
- ▶ SNS 等に利用者が書き込んだ情報
- ▶ その他アプリケーションに基づくデータ
- ▶ 上記データ群に関するメタデータ

■ データの識別

▶ 識別方針

◇ データ、およびデータが指し示す実物や組織・場所等は、一意に識別されるべきである。

◆ 他のデータ・実物・組織・場所等との混同を避けるため。

◇ データの識別子は、RDFのリソースを表現する識別子として規定されているURI (Uniform Resource Identifier) 形式で表現されるべきである。

▶ 識別方法

◇ すでに上記を満たす識別子体系がある分野においては、その識別子を利用できる。

◇ データやそれに関連する実物・組織・場所等を一意に識別する手法がない、あるいはその識別子をURIとして表現できない場合は、識別子としてucodeを利用できる。

ODDPデータ規格の全体概要

■ データモデル

- ▶ RDFモデルに準拠する。

■ 表現形式

- ▶ RDF/XML、Notation3、N-Triples、Turtleなどを利用する。

■ ボキャブラリ

- ▶ ボキャブラリとはデータの意味を共通に理解するための、辞書に相当する情報
- ▶ DCMI Metadata Termsに基づき、ボキャブラリを定義する際に明記することを推奨するメタデータを列記
 - ◇Name: A token appended to the URI of a DCMI namespace to create the URI of the term
 - ◇Label: The human-readable label assigned to the term
 - ◇URI: The Uniform Resource Identifier used to uniquely identify a term
 - ◇Definition: A statement that represents the concept and essential nature of the term
 - ◇Type of Term: The type of term as described in the DCMI Abstract Modelなど
- ▶ 参考となるボキャブラリー一覧をAppendixとして掲載

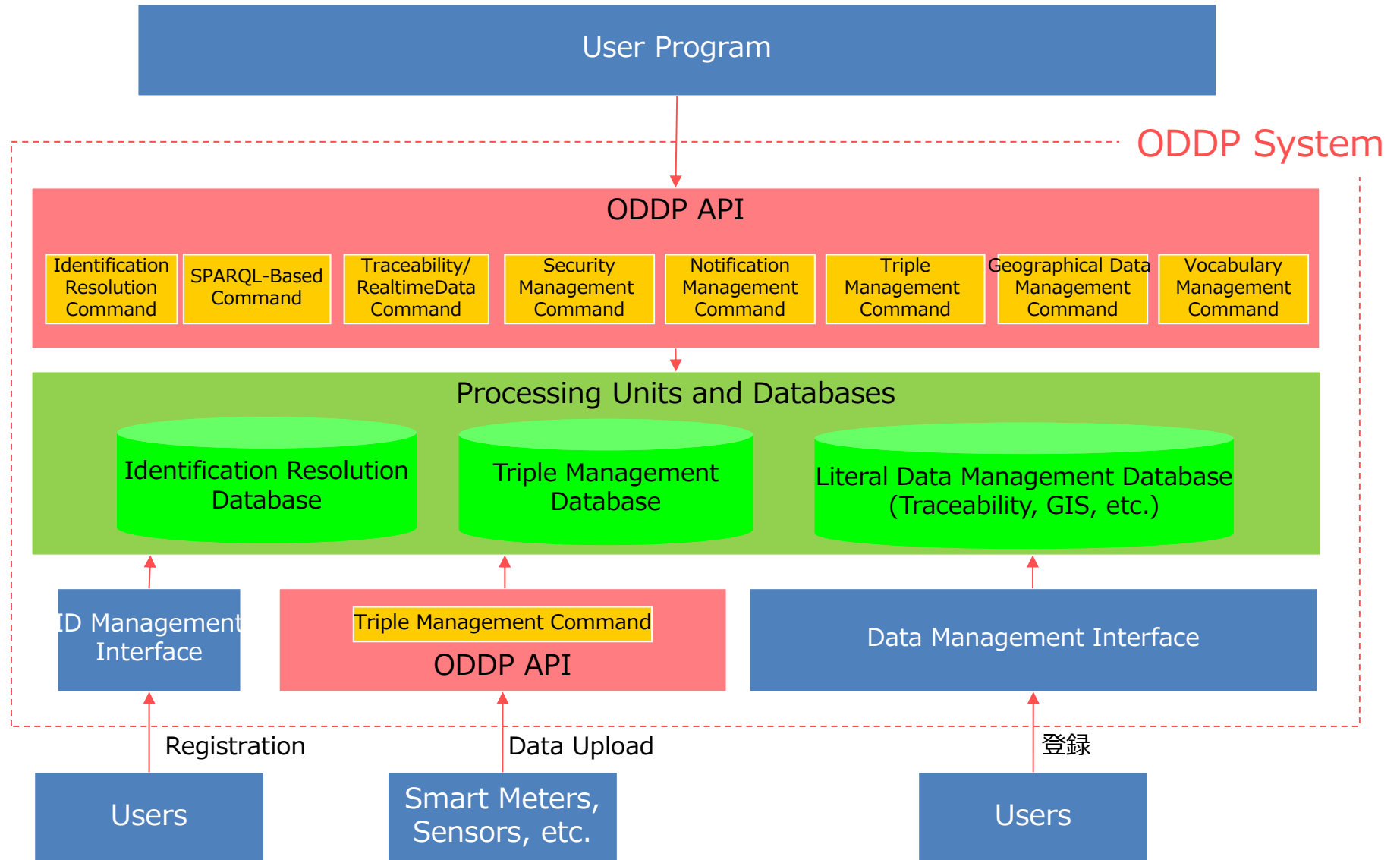
参考となるボキャブラリー一覧 (1)

名称	規定範囲	ネームスペース	ターム例
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)
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	geo:wktLiteral(Well-Known Text規格の地理情報), geo:gmlLiteral(GML規格の地理情報)
W3C Basic Geo	WGS84に基づく一点を表現するためのボキャブラリ。	http://www.w3.org/2003/01/geo/wgs84_pos#	wgs84_pos:lat(緯度)・ wgs84_pos:long(経度)
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)

参考となるボキャブラリー一覧 (2)

名称	規定範囲	ネームスペース	ターム例
事物の基本クラス・物理量	事物の基本クラス・物理量を扱う基本的なボキャブラリ。	http://uidcenter.org/ucr/vocab/uc#	uc:Entity(エンティティクラス), uc:length(長さ), uc:issued(ucode発行日)
単位系	物理量・貨幣単位を記述するボキャブラリ。		uc:Meter(メートル), uc:Seconds(秒)
地物	山・建物・移動体, 行政界や関心地点など, 場所に関するボキャブラリ。	http://uidcenter.org/ucr/vocab/ug#	ug:Facility(施設), ug:Railway(鉄道), ug:floor(階層), ug:consistsOf(含んでいる)
地理情報サービス	地物や施設に関するサービス情報を記述するボキャブラリ。	http://uidcenter.org/ucr/vocab/ugsrv#	ugsrv:keyword(キーワード), ugsrv:price(料金), ugsrv:lowerAge(利用可能な最低年齢)
地物アクセシビリティ	関心地点に関する通行可能性について記述するボキャブラリ。	http://uidcenter.org/ucr/vocab/spac#	spac:Walker(歩行者), spac:Bamp(段差)
製品・物品	製品や物品に関する基本的な情報を記述するボキャブラリ。	http://uidcenter.org/ucr/vocab/uobj#	uobj:InsuatrialProduct(工業製品), uc:owner(管理者), uc:producer(生産者)
イベント	生成・流通等のイベントを記述するボキャブラリ。	http://uidcenter.org/ucr/vocab/ev#	ev:IssuedEvent(発生), ev:DivisionEvent(分割), ev:target(対象物)
取引	取引に関するボキャブラリ	http://uidcenter.org/ucr/vocab/trans#	trans:Receipt(領収書), trans:creditor(販売者), trans:priceUnit(金額単位)

情報流通連携基盤システムの構成



ODDP 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でも、データ検索・取得コマンドのレスポンスにRDF/XML、RDF/JSON等を利用している。これは、RDFモデルに基づくデータとの互換性を保つためである。
- ▶ 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している。

機能名	概要
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。
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を提供する。
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・操作機能を提供する。
Notification Management Command	データの登録・更新をトリガとしてデータ利用者のシステムにコールバックする（Notification）仕組みを提供する。
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を提供する。
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・取得に関する機能を提供する。
Identification Resolution Command	IDをキーとしてデータを登録・検索する機能を提供する。

SPARQL-Based Command

■ 機能概要

- ▶ SPARQL 1.1プロトコルに準拠した、RDFモデルに基づくデータの登録・更新・削除・取得・検索機能を提供
- ▶ /api/v1/sparql/
 ✦ SPARQL 1.1が規定するクエリを発行する。
- ▶ /api/v1/rdf-graph-store
 ✦ 「SPARQL 1.1 Graph Store HTTP Protocol」に準拠した、グラフ単位での登録・取得・更新・削除を行う。

■ 備考

- ▶ リクエストパラメータの形式をContent-Typeヘッダで指定する。
- ▶ レスポンスパラメータの形式をAcceptヘッダで指定する。

■ APIリスト

URL	HTTPメソッド	意味
/api/v2/sparql/	GET	SPARQL 1.1 準拠のクエリを発行する
/api/v2/sparql/	POST	SPARQL 1.1 準拠のクエリを発行する
/api/v2/rdf-graph-store	GET	RDFグラフを取得する
/api/v2/rdf-graph-store	POST	RDFグラフを追加する
/api/v2/rdf-graph-store	PUT	RDFグラフを更新する
/api/v2/rdf-graph-store	DELETE	RDFグラフを削除する

■ API例

▶ クエリ

```
POST /api/v1/sparql HTTP/1.1
Host: www.example.org
Accept: application/sparql-results+xml
Content-Type: application/x-www-form-urlencoded
Content-Length: xxx
```

```
query=PREFIX%20dc%3A%20%3Chttp%3A%2F%2Furl.org%2Fdc%2Felements%2F1.1%2F%3E%20%0ASELECT%20%3Fbook%20%3Fwho%20%0AWHERE%20%7B%20%3Fbook%20dc%3Acreator%20%3Fwho%20%7D%0A
```

▶ レスポンス

```
HTTP/1.1 200 OK
Content-Length: xxx
Content-Type: application/sparql-results+xml
```

```
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
  <head>
    <variable name="book"/>
    <variable name="name"/>
  </head>
  <results>
    <result>
      <binding
name="book"><uri>http://www.example/book/book5</uri><
/binding>
      <binding name="name">
        <literal>Alice</literal>
      </binding>
    </result>
  ...
</sparql>
```

Geographical Data Management Command

■ 機能概要

- ▶ GIS等地理情報処理に必要な登録・検索処理を提供

■ 備考

- ▶ 場所の地点・領域を表現するために、OpenGISのWell Known Text、GML、GeoJSON形式を利用できる。
- ▶ 検索時の条件として、領域の重なり・包含を指定できる。

■ APIリスト

URL	HTTPメソッド	意味
/api/v2/place	GET	場所情報を検索する
/api/v2/place	POST	場所情報を登録する
/api/v2/place/<target>	GET	場所情報を閲覧する
/api/v2/place/<target>/<property>	GET	場所情報を閲覧する
/api/v2/place/<target>	PUT	場所情報を更新する
/api/v2/place/<target>/<property>	PUT	場所情報を更新する
/api/v2/place/<target>	DELETE	場所情報を削除する
/api/v2/place/<target>/<property>	DELETE	場所情報を削除する
/api/v2/place/<target>/ug:consists Of	PUT	場所情報の包含関係を移設 する

■ API例

▶ クエリ

```
GET /api/v2/place?within=POLYGON((0 0) (2 0) (2 2) (0 2))
HTTP/1.1
Accept: application/json
Host: example.org
```

▶ レスポンス

```
HTTP/1.1 200 OK
Content-Length: xxx
Connection: close
Content-Type: application/json; charset=utf-8
```

```
{
  "@context": {
    "ug": "http://uidcenter.org/ucr/vocab/ug#",
    "ug:type": { "@type": "@id" }
  },
  "@graph": [
    {
      "@id": "urn:ucode:_00001C000000000000001000000100800",
      "ug:region": "POINT(1 1)",
      "ug:type": "urn:ucode:_0FFFDE000000000000000000000000001234567"
    },
    {
      "@id": "urn:ucode:_00001C000000000000001000000100801",
      "ug:region": "POINT(1.5 1.5)",
      "ug:type": "urn:ucode:_0FFFDE000000000000000000000000001234567"
    }
  ]
}
```

Triple Management Command

■ 機能概要

- ▶ RDFデータの入出力を簡素化して行うための機能を提供
 - ◇ センサやスマートメータなどの小型機器がRDFモデルの主語・述語・目的語からなるTripleの登録・利用を効率的に扱える

■ APIリスト

URL	HTTPメソッド	意味
/api/v1/datapoints	GET	公開データを検索する
/api/v1/datapoints	POST	公開データを登録する
/api/v1/datapoints/<target>	GET	公開データを取得する
/api/v1/datapoints/<target>/<property>	GET	公開データを取得する
/api/v1/datapoints/<target>	PUT	公開データを更新する
/api/v1/datapoints/<target>/<property>	PUT	公開データを更新する
/api/v1/datapoints/<target>	DELETE	公開データを削除する
/api/v1/datapoints/<target>/<property>	DELETE	公開データの属性値を削除する


■ API例

▶ クエリ

```
GET
/api/v2/datapoints?dc_title=ABC%20meter HTTP/1.1
Accept: application/json
Host: www.example.org
```

▶ レスポンス

```
HTTP/1.1 200 OK
Content-Length: xxx
Connection: close
Content-Type: application/json; charset=utf-8
{
  "@context": {
    "dc":
      "http://purl.org/dc/elements/1.1/",
  },
  "@id":
    "urn:ucode:_00001C00000000000000100000100124",
  "dc:title": "ABC Meter"
}
```



オープンデータガイド (第1版)

※これ以降の資料は、VLED「成果公開」ページ「オープンデータガイド第1版」概要版の抜粋です。

<http://www.vled.or.jp/results/>

オープンデータガイド作成の目的

■ 背景: 国・地方公共団体等によるオープンデータへの取組みの活発化

- ▶ これらの組織がもつ公共データをオープンデータとして公開すれば、情報利用者によってアプリケーション開発等の様々な形での利活用が促進され、経済活性化や行政の透明性の向上等が期待できる。

■ オープンデータの特徴

- ▶ 従来の情報公開制度とは異なり、公開したデータを利活用し、信頼性の向上、国民参加・官民協働の推進だけでなく、経済の活性化・行政の効率化等に役立てることを目的とする。 ビジネスでの利用についての期待が大きい。
- ▶ オープンデータの編集・加工・改変等はコンピュータによって行われる。



■ オープンデータを普及させるために重要な事項

- ▶ 利用ルールを定めてデータの二次利用を認めること
- ▶ データを利活用しやすい形式（機械判読に適した形式）で提供すること



■ このため、本書は…

- ▶ 国、地方公共団体、独立行政法人、公共企業等が、自身が保有している公共データをオープンデータとして公開するための参考となるよう、オープンデータ流通推進コンソーシアム（データガバナンス委員会・技術委員会）が、オープンデータの作成・整形・公開に当たっての留意事項等を、「利用ルール」と「技術」の2つの観点からまとめた。

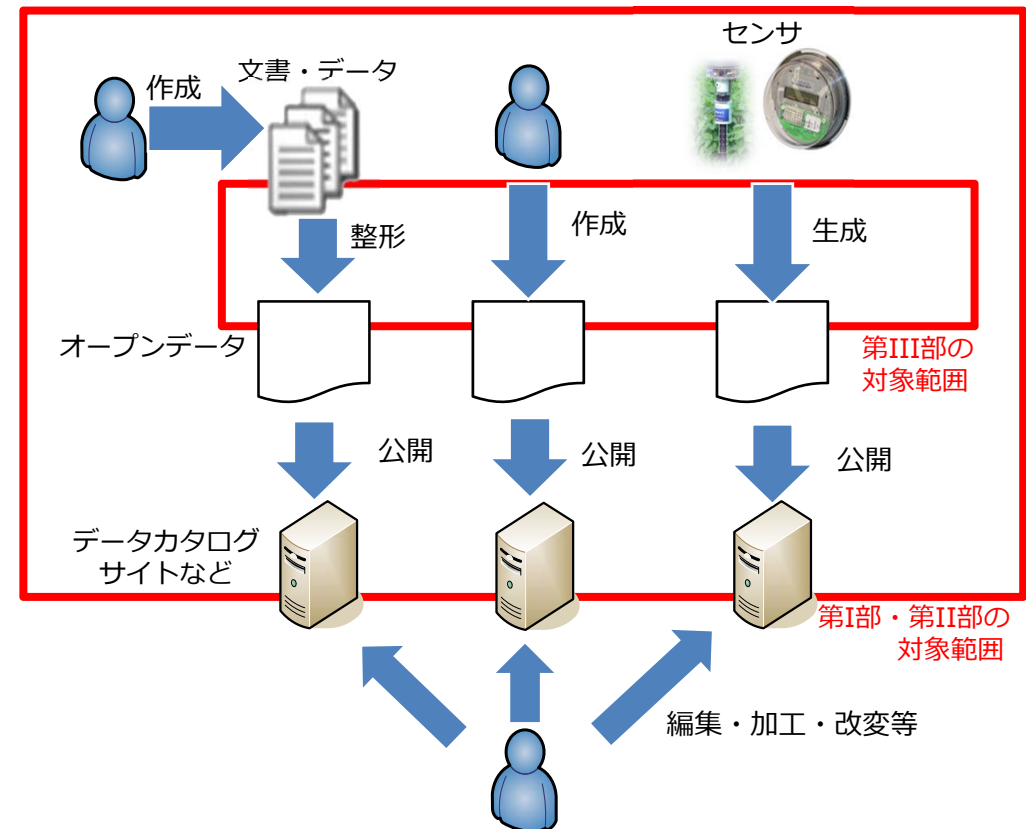
オープンデータの対象読者

■ 対象読者

- ▶ 現在保有しているデータや、これから作成するデータをオープンデータとして公開しようとする人。
 ☆主に国、地方公共団体、独立行政法人の職員を対象としているが、公共企業等の民間組織においても参考にできるものとして作成している。

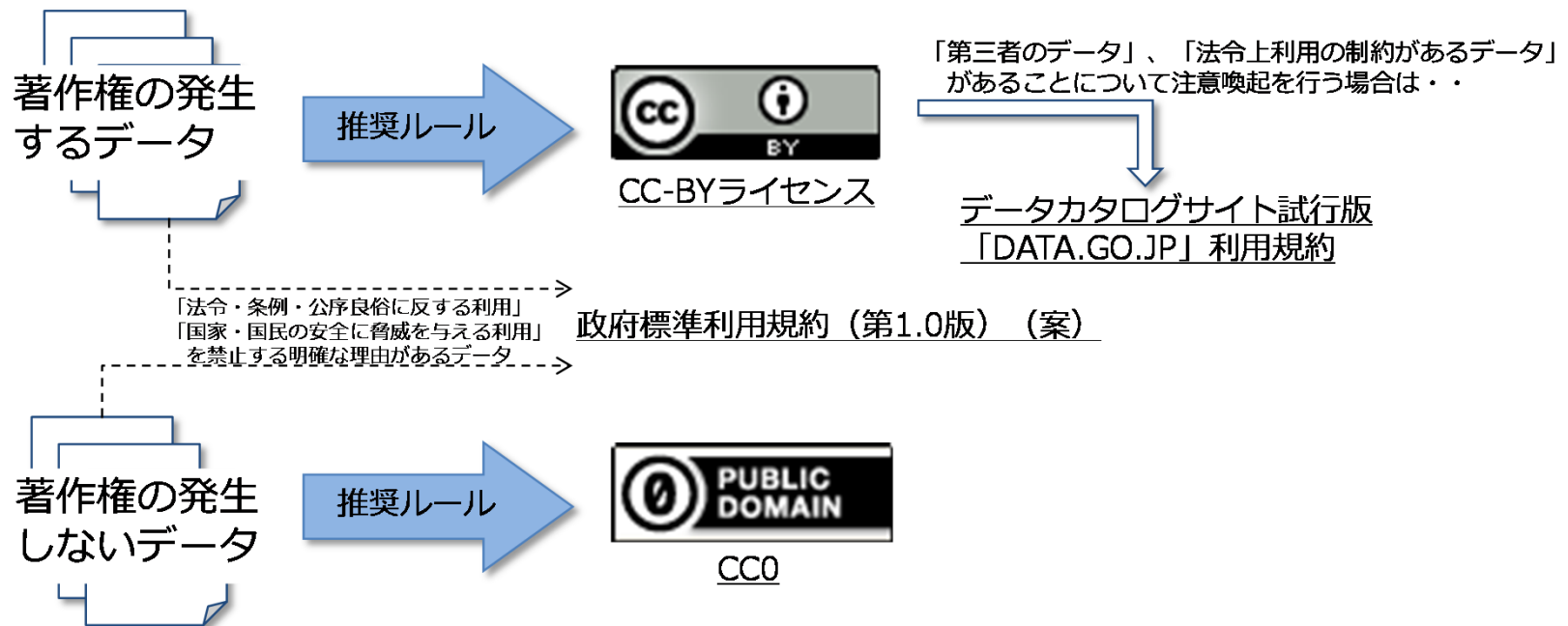
■ 部ごとの対象範囲

- ▶ オープンデータの流れ
 ☆情報提供者が作成・公開する。
 ☆これに情報利用者がアクセスし、編集・加工・変更等する。
- ▶ 第I部と第II部の対象
 ☆データの作成段階から公開段階に至るまでに関与する人。
- ▶ 第III部の対象
 ☆機械判読性の高いデータを作成・整形しようとする人。



エグゼクティブサマリ／利用ルールを選択

- オープンデータとする際には、データを自由に利用できることを示す利用ルールを付することが重要である。
- オープンデータでは、データを自由に利用してもらうことが重要であることから、情報利用者の視点（オープンデータの使い勝手）から見て、望ましい利用ルールを採用することが望ましい。
- そのため、基本的に著作権の発生するデータはCC-BYライセンス、著作権の発生しないデータはCC0を採用することが望ましい。
 - ▶ オープンデータとするデータの中に、第三者のデータや、法令上利用の制約があるデータが含まれている場合には、「データカタログサイト試行版（DATA.GO.JP）利用規約」を採用することが望ましい。



エグゼクティブサマリ／機械判読に適したオープンデータの作成・公開

- コンピュータにオープンデータを編集・加工・改変させ、それから新たな知見を得る作業を効率化するためには、コンピュータが数値やテキストを入手しやすい形式でオープンデータを作成・公開することが望ましい。
 - ▶ このようなデータを「機械判読に適したデータ」という。
 - ▶ このようなデータは、必ずしも人が読みやすいとは限らない。必要であれば、機械判読に適した形式と人に読みやすい形式の2種類を用意して公開することも考慮すべき。
- 機械判読に適したオープンデータを作成するには…

1. 非独占の（標準化された）ファイル形式で作成・公開することが望ましい。
 具体的には下記の形式を推奨する。

データの種類	望ましいファイル形式
表形式データ	CSV, xlsx (Office Open XML), ods (Open Document), JSON, RDF/XML, ...
文書形式データ	HTML, XML, docx (Office Open XML), odt (Open Document), ...
地理空間データ	KML, GML, ...

←ヘッダ

2. 表形式データでは、以下に留意することを強く推奨する。
 - ◇1つのファイルに、1つの表のみを格納する。
 - ◇ファイルのヘッダ（表の各列の名前を保持する行）を1行にする。
3. 地理空間データの場合は、以下に留意することを強く推奨する。
 - ◇位置情報に関するデータを付与する場合は、緯度・経度等の位置情報に加えて、測地系を明記する。

「オープンデータガイド」の全体構成

■ 第I部 Getting Started: オープンデータをはじめよう

▶ 第1章 はじめに

◇本書の目的・対象読者・構成を示すとともに、本書が利用する用語の定義を行う。

▶ 第2章 オープンデータの動向と意義

◇国内外のオープンデータに関する動向を紹介するとともに、オープンデータの意義について解説する。

▶ 第3章 オープンデータの作成・公開手順

◇オープンデータの作成・公開手順を、6つのステップに分けて解説する。

■ 第II部 利用ルール編: オープンデータに利用ルールを設定しよう

▶ 第4章 オープンデータで必要となる利用ルール

◇オープンデータにおける利用ルールの重要性について解説するとともに、利用ルールに関する国際的な動向、日本政府における動向について紹介する。

▶ 第5章 オープンデータ利用ルールの概要

◇諸外国政府で採用が進んでいるCC-BYとCC0、日本政府で採用される政府標準利用規約（第1.0版）（案）の3つの利用ルールの特徴等について解説する。

▶ 第6章 利用ルールの比較と望ましいルール

◇CC-BY、CC0、政府標準利用規約（第1.0版）（案）の3つの利用ルールについて、情報利用者の視点、情報提供者の視点から比較を行う。また、比較結果を踏まえ、データをオープンデータとして公開する際に望ましい利用ルールについて解説する。

▶ 第7章 利用ルールに関する今後の見直しの方向性について

◇政府標準利用規約（第1.0版）（案）の今後の見直しにあたっての方向性について述べる。

「オープンデータガイド」の全体構成

■ 第III部 技術編：機械判読に適したオープンデータにしよう

▶ 第8章 オープンデータの技術レベル

◇第3章に記したオープンデータの作成・公開手順のうち、技術的な事項について解説する。

◇機械判読性、データカタログ、識別子について解説し、機械判読性に関する指標を「オープンデータの技術レベル」として示す。

▶ 第9章 オープンデータのための技術的指針

◇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のそれぞれについて、機械判読に適したオープンデータを作成するための技術的な指針を、識別子、ファイル形式、データの3項目について示す。

■ 付録

▶ 第10章 オープンデータに関する規格・ツール

◇機械判読に適したオープンデータを作成・編集する上で参考となる規格やツールをまとめる。

▶ 第11章 CKAN解説

◇データカタログシステムであるCKANの概要とその使用方法を解説する。

第1章 はじめに

■ 本書の概要

1. 本書の目的

◇前頁の通り。

2. 本書の対象読者

◇オープンデータを公開するまでの流れから、各部ごとの対象読書を示す。

3. 本書の構成

◇章ごとの構成（前述の通り）と、知りたい内容ごとに参照すべき章を示す。

4. 用語定義

◇本書が利用する用語を定義する。

第2章 オープンデータの動向と意義

■ 本章の概要

1. オープンデータに関する主な動向

◇ 日本政府・地方公共団体・海外でのオープンデータに関する取組を、それぞれ紹介する。

2. オープンデータの意義

◇ 「電子行政オープンデータ戦略」及び「二次利用の促進のための府省のデータ公開に関する基本的考え方（ガイドライン）」の記述より、オープンデータの意義を示す。

3. 本書におけるオープンデータの定義

◇ 「5★Open Data」や「電子行政オープンデータ戦略」、「電子行政オープンデータ推進のためのロードマップ」に記述に基づき、オープンデータを以下のように定義する。

- ◆ 「オープンデータ」とは、「営利目的も含めた二次利用が可能な利用ルールで公開」された、「機械判読に適したデータ形式のデータ」である。

第3章 オープンデータの作成・公開手順

■ 本章の概要

- ▶ オープンデータの作成・公開手順を、以下の6つのステップに分けて解説する。
 - ◇サーバの処理能力、データの信頼性、プライバシー・匿名化等、それぞれのステップにおける留意点も示す。

1. オープンデータ推進組織の設立

オープンデータの作成・公開を推進するための横断的組織を設立する。これ以降の活動は、この推進組織が中心となって進める。

2. 現状把握

形式

管理者

更新頻度

権利関係

ニーズ分析

3. 計画立案

オープンデータとする対象のデータと手法を明確にし、マイルストーンと計画を立案する。

4. 公開作業

計画に基づき、データを作成・整形し、公開の準備をする。

6. 改善点の洗い出し

利用者や作業担当者からのフィードバックを元に、改善点を洗い出す。

5. 公開・運用

オープンデータ管理のマイルストーンに基づき、ある程度の情報が登録された段階で公開し、システムの運用を開始する。

第4章 オープンデータで必要となる利用ルール

■ 本章の概要

- ▶ オープンデータとするためには、利用ルールを付して自由に利用できることを示す必要がある。利用ルールとしてどのようなものが望ましいかを理解するため、日本及び諸外国において、利用ルールについてどのような検討が行われているかについて紹介する。

■ 本章の構成

1. オープンデータにおける利用ルールの重要性
◇公共データをオープンデータとする際に、なぜ利用ルールが重要なのかを解説する。
2. 国際的なオープンデータの利用ルールの動向
◇諸外国においてオープンデータについてどのような法制度となっているか、またどのような利用ルールを採用しているかについて紹介する。
3. 日本政府におけるオープンデータ利用ルールの動向
◇日本の各省庁のウェブサイトにも適用される予定の「政府標準利用規約（第1.0版）（案）」の策定経緯について紹介する。また、日本のデータカタログサイトであるData.go.jpの利用ルールについて紹介する。

第5章 オープンデータ利用ルールの概要

■ 本章の概要

- ▶ オープンデータの利用ルールとして主に利用されている（利用される予定の）CC-BY、CC0、政府標準利用規約（第1.0版）（案）について、その内容と利用方法について紹介する。

■ 本章の構成

1. CCライセンス

◇国際的に広く利用されている利用ルールであるCCとは何かについて概要を紹介する。

2. CC-BY

◇CCライセンスの中でも特に諸外国において利用されているCC-BYライセンスについて、概要と利用方法を紹介する。

3. CC0

◇CCが発行している著作権者が著作権を放棄するという宣言であるCC0について、概要と利用方法を紹介する。

4. 政府標準利用規約（第1.0版）（案）

◇日本の各省庁で採用される予定の政府標準利用規約（第1.0版）（案）について、概要と利用方法を紹介する。

第6章 利用ルールの比較と望ましい利用ルール

■ 本章の概要

- ▶ 前章で紹介したCC-BY、CC0、政府標準利用規約（第1.0版）（案）について、情報利用者、情報提供者のそれぞれの視点で比較を行い、どのような利用ルールを採用することが推奨されるかについて記載する。

■ 本章の構成

1. 情報利用者の視点からの比較

◇オープンデータとなった情報を利用する人の視点から利用ルールを比較分析する。

2. 情報提供者の視点からの比較

◇保有するデータをオープンデータとして公開する人の視点から利用ルールを比較分析する。

3. オープンデータにする際に望ましい利用ルール

◇オープンデータでは情報が利用されることを前提とすることから、情報利用者の視点を重視して、望ましい利用ルールとは何かについて検討し、推奨する。

補足：公開されたデータの悪用とその責任について

◇オープンデータで公開した情報については、悪用された際に誰が責任を負うのかということが議論となることから、責任の所在について検討する。

第7章 利用ルールに関する今後の見直しの方向性について

- 公共データをオープンデータとして公開する場合、情報利用者視点に立ち、基本的には、国際的にオープンデータの利用ルールとして広く使用されているCC-BY又はCC0を適用することが望ましい。
- 情報提供者に配慮し、公序良俗に反する利用等の禁止事項を盛り込むことが、例えば、できることから速やかに着手するというスモール・スタートの原則にかなう場合、できるだけ多くのデータをオープンデータにする対象としたいといった場合には、次善策として、政府標準利用規約（第1.0版）（案）を適用することが考えられる。
 - ▶ 政府標準利用規約（第1.0版）（案）は、利用ルールの政府標準利用規約（第1.0版）（案）への変更後のコンテンツの利用状況等を踏まえ、見直しの検討が行われる予定となっている。国以外において、政府標準利用規約（第1.0版）（案）を適用する際には、今後見直しが行われる可能性があることを理解した上で、適用することが望ましい。
- オープンデータとして公開されたデータの不適切な利用によって第三者等に何らかの問題が起きた場合、その責任はデータを不適切に利用した情報利用者にあるのであって、情報提供者である国、地方公共団体等が責任を負うものではないということを啓発していく必要がある。
 - ▶ オープンデータによって公開された情報には、誤りがある場合等もあると想定されるが、その誤りが原因で、情報利用者や第三者等に何らかの問題が起きた場合についても、CCライセンス、政府標準利用規約（第1.0版）（案）のいずれにおいても無保証で公開しているのであり、国、地方公共団体等が責任を負うものではない

第8章 オープンデータの技術レベル

■ 本章の概要

1. 機械判読性に関する解説（→ 後述）

◇機械判読に適したデータが必要である理由と、機械判読性に関する指標について解説する。

2. データカタログに関する解説

◇データカタログの意義を解説する。

3. オープンデータと識別子

◇識別子 = データやデータが対象とする実物や組織・場所等をコンピュータに識別させるための番号。

◇識別子の付与対象: オープンデータとそれに含まれる実物や組織、場所等。

◆ オープンデータは、コンピュータが読み取り解釈するためのデータであり、コンピュータが一意に識別できるべきである。

◆ 表記の揺らぎにより同一の組織や場所を別物として解釈する、同一名称だが違う意味である組織や場所を区別できない、という問題が発生するため、オープンデータに含まれる実物や組織、場所等も、一意に識別できることが望ましい。

4. オープンデータの技術レベル（→ 後述）

◇上記3項目をもとに、オープンデータの技術レベルを定める。

5. オープンデータの管理ポリシーとメタデータの付与方法

◇オープンデータを登録・管理する際に、メタデータを自動的に付与することができるならば、管理・登録のコストを軽減できる。そのような手法について解説する。

8.1 機械判読性に関する解説

■ 機械判読に適したデータとは？

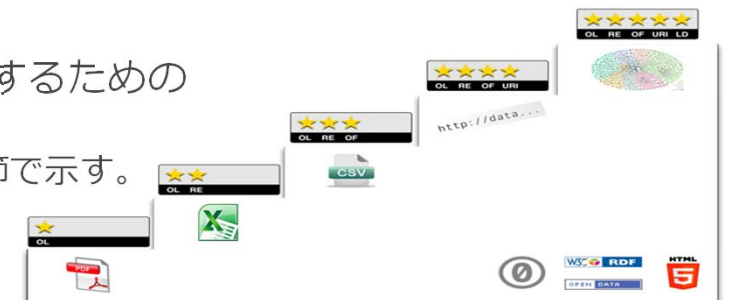
- ▶ コンピュータが数値やテキストを抽出しやすい形式のデータ
 - ◇このようなデータを提供することにより、コンピュータの解析に必要な情報利用者のコストを軽減できる。
- ▶ 機械判読に適さないデータ: PDFや画像データ
 - ◇このデータをコンピュータに与えて、解析させるためには、以下の作業が必要。
 - ◆ 事前に人間がその画像にあるデータを表計算ソフトウェアに入力して保存する。
 - ◆ 画像認識等の技術により公開されているデータから数値やテキストを得て、それをコンピュータに与える。
 - ◇これらは情報利用者に負担を求める方法であり、効率的でない。

■ 機械判読性に適したデータに関する留意点

- ▶ 機械判読に適したデータは、必ずしも人が読みやすいとは限らない。
 - ◇必要であれば、機械判読に適した形式と人に読みやすい形式の2種類のファイルを用意して公開することも考慮すべき。

■ 機械判読性に関する指標: 5★Open Data

- ★1:この形式のファイルからコンピュータがデータを取り出すことは容易ではない。
- ★2:対応するソフトウェアを用意すれば、コンピュータはこの形式のファイルからデータを抽出できる。
 - ◇一般に「機械判読性のあるデータ」とは★2以上のデータをいう。
- ★3: 解析方法が公開されているため、この形式のファイルを解析するためのソフトウェアを構築することは、★2より容易である。
 - ◇この形式のデータに対する機械判読性を高めるための技術的指針を9.3節で示す。
- ★4~: この形式のデータは、相互に接続でき、コンピュータによるデータのマッシュアップが容易になる。



5★Open Dataの指標 (<http://5stardata.info/>)

8.4 オープンデータの技術レベル

	Level 0	Level 1	Level 2	Level 3	Level 4
データ	PDFや画像ファイルをWebで公開する。	構造化されたデータを作成し、Webで公開する。 (XLS, DOCなど)	非独占の（標準化された）形式で公開する。 (CSV等)	機械判読に適したデータを作成し、公開する。 (第9章参照)	RDF、XML等の技術を導入したデータを作成し、公開する。
データカタログ	存在しない	カタログを表形式データ（CSVなど）として作成し、公開する。	Level 1と同じ。	データカタログシステムを導入する。	RDFやSPARQLを利用したメタデータ検索機能を提供する。
識別子	何らかの手段で識別されている。	Level 0と同じ。	Level 0と同じ。	URLにより識別されている。	グローバルな体系に基づく識別子を利用する。
必要なツール	Webサーバ	Webサーバ	Webサーバ	Webサーバ+CKANなど	Webサーバ+CKAN+情報流通連携基盤など

- データ・データカタログ・識別子のレベルを合わせる必要はない。
- Level1のメリット
 - ▶ 情報利用者は、画像解析等の処理をすることなく、直接データを取得できる。
 - ▶ データのありか等のメタデータを電子的に入手できるようになる。
- Level2のメリット
 - ▶ データに対する機械判読性が高まる。
- Level3のメリット
 - ▶ データの解釈効率や検索性が向上し、情報利用者のデータ利活用の効率が向上する。
- Level4のメリット
 - ▶ 他のデータとの横断検索等も容易になり、情報利用者によるデータ利用の幅が広がる。

第9章 オープンデータのための技術的指針

■ 本章の概要

1. 識別子に関する指針

◇オープンデータを識別する識別子が満たすべき性質

- ◆ ユニークであること。
- ◆ 共通に利用できる体系であること。

◇利用可能な識別子体系

- ◆ グローバルにユニークな識別子体系
- ◆ 公的機関が定める識別子体系・コード体系
- ◆ URI (Uniform Resource Identifier) として表現できる体系

◇適切な識別子体系がない場合の対処法

- ◆ 対象に番号を付与し、ucodeやDOI等のグローバルな体系や、公的機関が定める識別子体系・コード体系に基づく識別子をして、管理する。
- ◆ 付与した番号に組織が決めるURLを付与してグローバル化することも可能。ただし、組織の統廃合等によりドメイン名が変わると、識別子も変わることにも注意。

2. ファイル形式に関する指針

◇公開するデータのファイル形式は、機械判読性の高い形式を利用することが望ましい。

◇オープンデータの技術レベルに基づいて、代表的なファイル形式を整理すると上表の通り。

3. データに関する指針 (→ 後述)

◇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のそれぞれについて、機械判読性の高いデータを作成・編集する際の指針を、以下の2つのグレードに分けて示す。

- ◆ グレード1:オープンデータが満たすことを強く推奨する指針。
目的は、標準的な規格に矛盾せず、情報利用者が、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、そのデータの本質的内容を正しく解釈するためのプログラムを書けること。
- ◆ グレード2:オープンデータが満たすことを推奨する指針。
目的は、データを取得したプログラムが、そのデータの項目や構造を正しく解釈できること。

	Level 1	Level 2/3	Level 4
表形式データ	xls	CSV xlsx ods JSON	RDF/XML RDF/JSON, JSON-LD Notation3 Turtle等のRDF形式
文書形式データ	doc	HTML XML docx odt	
地理空間データ	shape	KML GML	推奨するファイル形式
リアルタイムデータ	(ファイルの形で交換しない)		

9.3 データに関する指針

■ 表形式データに関する指針

グレード	指針	
1	指針1	1つのファイルは、1種類の表から構成されるべきである。
	指針2	ヘッダは、1行から構成されるべきである。
2	指針3	データでない情報を、レコードに含めないことが望ましい。
	指針4	全てのフィールドは、他のフィールドと結合されないことが望ましい。
	指針5	値がない場合を除き、フィールドを空白にしない（省略しない）ことが望ましい。
	指針6	年の値には、西暦表記を備えることが望ましい。
	指針7	フィールドの単位が明記されていることが望ましい。
	指針8	利用している文字コードを明記することが望ましい。また、国際的に広く利用されている文字コードを利用することが望ましい。
	指針9	ファイルの属性や説明を表すメタデータが、フォーマルに記述されていることが望ましい。また、そのメタデータからデータセット本体へリンクし、たどれるようにすることが望ましい。
	指針10	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ることが望ましい。

9.3 データに関する指針

■ 文書形式データに関する指針

グレード	指針	
1		(なし)
2	指針1	文章に存在する部・章・節・図表などの構造が、機械判読性の高いフォーマットで記述されていることが望ましい。
	指針2	文章内に、整形のための符号や文字（空白、改行等）を含めないことが望ましい。
	指針3	文書形式データが表形式データを含む場合、グレード1以上の表形式データが添付されていることが望ましい。
	指針4	文章に対する、情報利用者が理解できるような説明が、メタデータとして記述され、当該文書にリンクされていることが望ましい。

9.3 データに関する指針

■ 地理情報形式データに関する指針

グレード	指針	
1	指針1	位置情報に関するデータを付与する場合は、緯度・経度等の位置情報に加えて、測地系が明記されるべきである。屋外であれば、世界測地系を利用することが望ましい。屋内であれば、座標系と縮尺を示すべきである。
2	指針2	地図データは、ベクタ形式に依るものが望ましい。ベクタ形式のデータの作成にあたっては、最新の ISO 規格及び JIS 規格に基づいた地理空間情報標準プロファイル (JPGIS)、地理空間情報のメタデータの共通仕様を規定する日本版メタデータプロファイル (JMP) を用いる。
	指針3	地理情報に対する、情報利用者が理解できるような説明が、メタデータとして記述され、当該文書にリンクされていることが望ましい。

■ リアルタイムデータに関する指針

グレード	指針	
1	指針1	データの取得仕様が明記されているべきである。
	指針2	表形式データや地理情報データをファイル形式で取得させる場合は、それぞれのグレード1の指針を満たすべきである。
2	指針3	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ことが望ましい。

第10章（付録）オープンデータに関する規格・ツール

■ 本章の概要

- ▶ 機械判読に適したオープンデータを作成・編集する上で参考となる規格やツールをまとめる。

■ 本章の構成

1. データフォーマットに関する規格

◇表形式データ・文書データ・地理空間データのそれぞれのファイル形式に関する代表的な規格と、その形式を扱える代表的なソフトウェアを挙げる。

2. 識別子に関する規格

◇オープンデータの識別に利用できる識別子を列記する。

3. オープンデータに有用なツール

◇オープンデータを作成・編集・公開するために有用なツールを解説する。

- ◆ Webサービス
- ◆ データカタログサービス
- ◆ GISシステム
- ◆ 情報流通連携基盤
- ◆ RDFレポジトリ

第11章（付録）CKAN解説

■ 本章の概要

- ▶ オープンデータの技術レベルのうち、データカタログのLevel3を満たすためのデータカタログシステムとして広く利用されている、CKAN（Comprehensive Knowledge Archive Network）を解説する。

■ 本章の構成

1. CKAN概説
2. CKANの運用前に検討・準備すべき事項
 - ◇公開するオープンデータの洗い出し
 - ◇オープンデータを管理するポリシーの策定
 - ◇要求仕様の策定
 - ◇データの整備計画
3. CKANを用いたオープンデータ登録例
 - ◇アカウント登録
 - ◇組織の登録
 - ◇組織へのメンバ追加
 - ◇データセットの作成
 - ◇データの登録
 - ◇CKANの管理ページ

