

第四回 技術委員会資料

オープンデータ流通推進コンソーシアム 技術委員会検討報告（案）

2013.06.03

オープンデータ流通推進コンソーシアム 事務局

平成24年度成果概要

- Call for Commentを求められる技術文書ができあがった。
 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」にインプットした。
 - ▶ 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、パブリックコメント募集中である。

目次

1.	検討の基本方針	4
2.	技術調査	6
i.	データ規格	8
ii.	API規格	14
iii.	国際的動向	17
3.	技術ガイド・規格案	28
i.	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	30
ii.	オープンデータ化のためのCSV形式データ規格	36
iii.	情報流通連携基盤システム外部仕様書(平成24年度版)	38
4.	次年度の課題	47



1. 検討の基本方針

オープンデータ技術検討に関する基本方針

■ 適用範囲

- ▶ 公共機関（政府・自治体）のデータだけでなく、民間事業主体におけるオープンデータにも適用可能
- ▶ 静的／集約データだけでなく、リアルタイムデータにも適用可能
- ▶ 数値パラメータデータや小型ファイルデータから、巨大なビッグデータまで、多様な規模のデータに適用可能
- ▶ 文書のようなデータにも適用可能

■ 目標

- ▶ 機械処理への適合性
 - ◇情報の精度・正確性の向上、曖昧性除去の実現
 - ◇高い処理性能・通信性能の実現
- ▶ 利用者の利便性向上
 - ◇機械可読性の達成
 - ◇プログラムが容易に書けることの実現
- ▶ 提供者の利便性向上
 - ◇データを出すことが容易に



2. 技術調査

技術的調査の概要

■ 目的

- ▶ オープンデータへの適用が考えられる規格等を洗い出し、それらの適用指針や課題を抽出する。

■ 調査の対象

- ▶ データ規格（ファイル形式規格）
- ▶ API規格
- ▶ 国際的な動向



i) データ規格

データ規格の例 (静的・集約データ)

種類	主なファイル形式	標準規格等
文字	テキスト(.txt)	オープン (文字コードとしてISO/IEC 646, 8859等)
ドキュメント 表計算 プレゼンテーション	PDF (.pdf)	(Adobe社) ISO 32000-1
	Microsoft Office Binary(.doc, .xls, .ppt)	(マイクロソフト社独自)
	Office Open XML(.docs, .xlsx, .pptx)	ISO/IEC 29500
	OpenDocument(.odt, .ods, .odp)	ISO/IEC 26300
	CSV (.csv)	RFC 4180
電子書籍	EPUB (.epub)	国際電子出版フォーラムが規格化
画像	JPEG (.jpg)	ISO/IEC 10198
	JPEG 2000 (.jpeg)	ISO/IEC 15444-1
	PNG (.png)	ISO/IEC 15948
	SVG (.svg)	W3Cにより制定
映像	MPEG-4 (.mp4)	ISO/IEC 14496-14
音声	MPEG Audio Layer-3 (.mp3)	ISO 11172-3
地図	GML (.gml)	ISO 19136
気象	GRIB2, BUFR	国際気象通報式として規格化
書誌情報	MARC21	米国議会図書館が規定
セマンティクスデータ	XML (.xml)	W3Cにより制定
	RDF (.rdf)	W3Cにより制定

(参考) Five Star Open Data

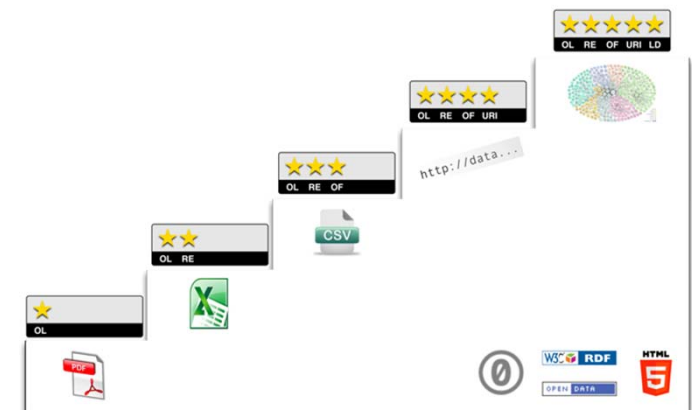
- Tim Berners-Lee氏がオープンデータに向けたデータ公開の段階を示したもの

▶ <http://5stardata.info>

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)

5 ★ Open Data

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5 star deployment scheme for Open Data. Here, we give examples for each step of the stars and explain costs and benefits that come along with it.



データ規格の例 (リアルタイムデータ)

種類	主な形式	標準規格等
バイナリデータ		Base64 [RFC3548], quoted-printable [RFC2045]等の規格によりテキスト化して伝送するケースが多い
文字データ	TXT	オープン(文字コードとしてISO/IEC 646, 8859等)
データ列	CSV	RFC 4180
構造化データ	XML	W3Cにより制定
	JSON	RFC 4627

データ規格に関する適用指針

■ 以下のような観点から、最適なデータ規格を利用すべき。

▶ データの種類・性質・サイズ・利用目的・利用環境

◇これらのターゲットから理想レベルのデータ規格と許容レベルのデータ規格を選択できる。

▶ プログラミングのしやすさ

◇データ処理の性能やデータ規格の複雑さ、データを扱うためのツールが普及しているかに依存する。

▶ データ作成のしやすさ

◇人間がデータを編集するためのツールが普及しているか。（例：文章や表形式のデータ編集ツールは広く普及している）

▶ 複数のデータ間の相互運用性

◇世界的な動向との一貫性・相互運用性があるか。

◆例：NIEM、ISAなど、政府／公共データのオープン化の枠組みでのデータ規格の動向

▶ データ規格の知的財産権

◇データ規格に付与されている特許・ライセンス条項はあるか。

◆画像を出力するツールが、データのヘッダ部分に著作権を明記させていたため、論文から画像を除かなければ公開できなかった、という事例もある。

データ規格を適用する際の留意事項

■ 上位層で整備すべき規定

1. ボキャブラリ、コード（共通の使われる意味情報のコード化）

◇RDFでのボキャブラリ

◇RDF以外のデータ規格の場合でも、ボキャブラリに相当する規定は必要とされる。

2. ID・識別子（データから参照される個体（Entity/Object）の識別）

◇RDFの場合、URI形式というIDの文法は定められているが、オープンデータを作成する時に、具体的なIDに関しては、そのデータホルダが定める必要がある。

◇RDF以外のデータの場合は（例えばCSV）、IDやコードに関する規約はない。

3. データカタログ

◇データ自体の識別

◇データのメタデータ（データの由来・説明など）の記述方法



ii) API規格

API規格

規格名	標準規格	概要
FTP (File Transfer Protocol)	RFC 959	端末とサーバの間でファイル(ドキュメントや画像・動画など)を転送するための、代表的なプロトコルである。 「ファイルの2行目を取得する」というような要求を出せない。ファイルサイズが大きい場合、通信帯域と端末側のリソース(メモリやディスク容量)を圧迫する。
HTTP (HyperText Transfer Protocol)	RFC 2616	URL(指定したリソース(通常はファイル)の内容を取得する、代表的なプロトコルである。 「ファイルの2行目を取得する」というような要求を出せない。ファイルサイズが大きい場合、通信帯域と端末側のリソース(メモリやディスク容量)を圧迫する。
CGI (Common Gateway Interface)	RFC 3875	ウェブサーバ上で任意のプログラムを動作させるための仕組みである。端末は、プログラムの位置とパラメータを指定した要求をサーバに送る。サーバは指定されたプログラムを実行し、その結果を端末に返す。 パラメータの与え方も結果の受け取り方も、プログラムに依存する。
REST (Representational State Transfer)		対象とするリソースをURLで指定し、HTTPの4つのメソッドGET, POST, PUT, DELETEを取得・登録・更新・削除の各操作に対応させてweb上のリソース(データ)を扱うスタイル。 パラメータの与え方も結果の受け取り方も、提供されるサービスに依存する。
SOAP	W3Cにより 規定	ソフトウェア同士がメッセージ(オブジェクト)を交換するためのプロトコルである。交換するメッセージはXMLに準拠している。従って、サーバへの要求もサーバからの応答も、XMLメッセージである。 汎用性を考慮してメッセージを抽象化しているため、交換するメッセージが複雑である。
SPARQL (SPARQL Protocol and RDF Query Language)	W3Cにより 規定	RDFモデルに基づくデータを対象とし、その検索・操作するクエリ言語である。

API規格に対する適用指針

■ 以下のような観点から、適切なAPIを選択すべき。

▶ API規格の特質、想定されるデータのターゲット

- ◇ FTP / HTTP GETは、ファイル単位のデータ取得を対象としている。
- ◇ CGI / REST / SOAP / SPARQLは、データを直接取得することをターゲットとしている。
- ◇ SPARQLは、RDFモデルに基づくデータがターゲットである。

▶ 利用した際にネックとなる知財上の問題の有無

- ◇ APIのライセンス条項・制限条項など。

▶ API規格の普及状況、ツール等の整備状況

- ◇ 今回例示したFTP、HTTP GET、REST、SOAP、SPARQLについては、ツールが普及している。

▶ API規格の世界的動向との一貫性、相互運用性

- ◇ 今日では、RESTベースのAPIが広く使われている。
 - ◆ 復興・復興支援制度データベース（経済産業省）
 - ◆ Socrata Open Data API (SODA)
 - ◆ cosm (旧pachube)
 - ◆ Twitter API

■ データ規格と同様、更に上位層で整備すべき規定がある。

▶ 例: ボキャブラリ・意味コード・IDなど



iii) 國際的動向

国際的動向

■ 調査対象: 公共データ (Public Sector Information : PSI) の扱いに関する事例

1. Engage, Getting on with Government 2.0, Report of the Government 2.0 Taskforce
2. Government Data and the Invisible Hand
3. ベストプラクティス/ガイドラインの事例
4. オープンデータフォーマット仕様を提案している団体
5. メタデータ記述
6. データポータル事例

1. Engage, Getting on with Government 2.0, Report of the Government 2.0 Taskforce (*2)

■ オーストラリア政府に2009年末に提出された委員会レポート。主な論点は以下の通り。

▶ ライセンスの重要性

☆ライセンスも機械可読な標準として扱うことが重要である。

▶ メタデータの重要性

☆検索発見できないデータは、存在しないに等しい。(If it can't be spidered or indexed, it doesn't exist.)

☆オーストラリアでは、政府情報の検索に使うシソーラスに使う目的でAGLS (Australian Government Locator Service) メタデータを規定している。これは、Dublin Coreを拡張したボキャブラリ規格である。

2. Government Data and the Invisible Hand (*3)

■ Yale Journal of Law and Technology (vol. 160,2009) に掲載された論文。論点は以下の通り。

- ▶ Obama大統領のもとで、政府が Internet を利用したアカウントビリティのある政府をめざすとしても、政府がデータをきれいにみせるように自らウェブを運営するのではなく、再利用できるデータを公開すること、そしてそれを可能にする技術インフラに主眼をおくべき。
- ▶ 変化の速いウェブ技術においつくのを目的とはせず、簡単に使える信頼性の高い政府のもつデータを容易にだれもがアクセスできるインフラ整備することが重要である。
- ▶ 個々のエンドユーザの必要とするような表示、対話機能をもつウェブは政府ではなく民間が自発的につくるのがベストである。

3. ベストプラクティス／ガイドラインの事例

■ アメリカでの事例

- ▶ Requirements and Best Practices Checklist(*4)
 ✦具体的な Web 公開についてのベストプラクティスをまとめたチェックリスト。
- ▶ Draft Concept of Operation(*5)
 ✦data.govの運用に対する方針をまとめた文書。

■ オーストラリアの事例

- ▶ Enhancing the discoverability and accessibility of government information(*6)
 ✦Accessibility を向上することと、政府全体で共通のメタデータをつかうなど、検索を容易にするなどに関する提言。
- ▶ Early leadership in semantic web(*7)
 ✦政府機関のウェブやデータセットに“semantic tag” をつけるためのガイドライン。
- ▶ Whole of government information publication scheme(*8)
 ✦政府によるPSI公開にあたってのベストプラクティスの方針の議論。
 ✦集約した検索サイトの必要性を言及している。これが元となり<http://www.data.gov.au/> が誕生したと考えられる。

■ ニュージーランドの事例

- ▶ Government ICT Direction and Priorities(*9)
 ✦政府関連の PSI データについての情報をあつめるときに誰もが参照するページ。W3CのOpen Data five-star Modelが解説されており、PSIはレベル3以上（non-proprietaryで、マシン可読、ウェブ経由でアクセスできて、NZGOAL に応じてライセンスされていること）であるべきと規定している。
 ✦実際のPSI データ公開サイトの運用の際のチェックリスト(*10)も掲載されている。

(*4) <http://www.howto.gov/web-content/requirements-and-best-practices/checklists/long>
 (*5) <http://www.ideascale.com/userimages/sub-1/736312/ConOpsFinal.doc>
 (*6) <http://gov2.net.au/projects/project-1/>

(*7) <http://gov2.net.au/projects/project-5/>

(*8) <http://gov2.net.au/projects/project-7/>

(*9) <http://ict.govt.nz/>

(*10) <http://ict.govt.nz/programme/opening-government-data-and-information/toolkit-agencies>

3. ベストプラクティス／ガイドラインの事例

■ イギリスでの事例

▶ Public sector information Directive and Regulations^(*11)

◇イギリスがPSI をオープンデータとして公開する政策や、ベストプラクティスをまとめたサイト。

◇以下のようなレポートが掲載されている。

- ◆ United Kingdom Report on the Re-Use of Public Sector Information 2010
- ◆ Guide to the Regulations and Best Practice
- ◆ United Kingdom Report on the Re-Use of Public Sector Information 2009
- ◆ United Kingdom Report on the Re-Use of Public Sector Information 2008 UK implementation of the EU Directive on the Re-use of Public Sector Information - the first two years

▶ オープンデータ政策が主流になる前に、特定の政府機関などに独占的にデータを利用、販売させる契約がなされた事例も紹介されている。

■ IATIの事例

▶ IATI: International Aid Transparency Initiative^(*12)

◇国外から経済支援をうけるような支援開発途上国が、その必要性を自ら評価し、実際に手に入りそうな経済支援を探すのを助け、さらに支援をする側の国の市民が、税金が有効に使われているかを助けるためのNPOである。

◇その活動の中で、世界中の経済支援の資金の流れ（目的、実際の使い道など）を比較できるように、共通のオープンデータフォーマット形式を提案し、利用を促している。

◆ 世界中の団体が標準形式で公開したデータは、IATI レジストリへリンクを登録することで公開、周知、共有ができる。

◆ IATI レジストリはデータそのものもホストするが、リンクを収集することも主要な目的である。

◇データ提供者への「入力補助、データの検証、データビューワー」の3点セットを提供することで、その利用促進を促している。あるルールに従ってCSV データを用意すれば、CSV - XML 変換ツールにより、自動的に IATI 標準のXML データが作成できる^(*13)。

(*11) <http://www.nationalarchives.gov.uk/information-management/legislation/directive-and-regulations.htm>

(*12) <http://www.aidtransparency.net/>

(*13) <http://iatistandard.org/getting-started/tools-to-help/>

3. ベストプラクティス／ガイドラインの事例

■ XBRL (eXtensible Business Reporting Language)

- ▶ XBRLは、財務諸表などのビジネスレポートを電子化して、ビジネスレポートの作成の効率化、比較・分析などの二次利用が行えることを目的として、XMLの規格をベースに作られた言語である。
- ▶ これは、データの主なユーザである会計士の団体が提案のきっかけの活動を担い、さらにその利用を政府機関が促したことから標準になるという、標準化に対する成功事例である。

◇web の初期の段階でアメリカのSEC (Securities and Exchange Commissions, www.sec.gov/) は、自らが監督する株式市場、証券会社のデータのWEBでの公表に躊躇した。

◇しかし 個人の努力で一部のデータがInternet に流れる、その情報源へのアクセスが大量に殺到した。

◇その人気をみた結果SEC も公開に前向きになった。

◇この過程で、XMLを財務データ向きにした表現を使って財務情報を交換するアイデアが民間から提案され、検討の末XBPLという標準規格になった。

(*11) <http://www.nationalarchives.gov.uk/information-management/legislation/directive-and-regulations.htm>

(*12) <http://www.aidtransparency.net/>

(*13) <http://iatistandard.org/getting-started/tools-to-help/>

4. オープンデータフォーマット仕様を提案している団体

■ FGDC (Federal Geographical Data Committee)

- ▶ 地理データを国家機関でのユーザ、一般市民に提供する地理情報のインフラ提供を行なうための調整を行う機関。
 - ◇FGDCの活動は、National Spatial Data Infrastructure (NSDI) として知られている。
- ▶ FGDC はNSDI向けに自分たちでアメリカ国内の規格を提案制定すると同時に、NSDIにおいて利用すべき外部で制定された規格 (de jure, de facto の両方の標準) の推薦リストを作成している。
 - ◇内部標準 (27件) : <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/fgdc-endorsed-standards>
 - ◇外部標準 (65件) : http://www.fgdc.gov/standards/fgdc-endorsed-external-standards/index_html
- ▶ FGDCはさらに、地理データを扱える既存のソフトウェア製品の推奨リストを作成している。(割引値段の交渉済み)

■ Big Data Initiative

- ▶ アメリカでは、ホワイトハウスが2012年3月に、2億ドルの予算をつけた Big Data Initiative という R&D プロジェクトをアナウンスした。目的は以下の通り。
 - ◇Advance state-of-the-art core technologies needed to collect, store, preserve, manage, analyze, and share huge quantities of data.
 - ◇Harness these technologies to accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning;
 - ◇and Expand the workforce needed to develop and use Big Data technology
- ▶ 第1の目的から、ファイルフォーマットのより厳密な定義、拡張、ポータビリティをあげていろいろなプラットフォームで動作するようにするなどの標準活動がなされると考えられる。

5. メタデータ記述用ボキャブラリ例

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、タクソミー、フォークソミー、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)
NIEM	政府内での情報交換を円滑にするためのボキャブラリ体系。	http://niem.gov/niem/niem-core/2.0 など	nc:Pereson(人), nc:PersonBirthDate(誕生日)
ISA	EU加盟国間で公共データの相互運用をはかるためのボキャブラリ	http://w3.org/ns/person# , http://w3.org/ns/legal# , http://w3.org/ns/locn#	person:birthName(出生時の名称), person:placeOfBirth(出生地), legal:name(ビジネス上の氏名)

6. データポータルに関する事例

■ 以下のデータポータルを調査し、公開されているデータのフォーマット、メタデータ記述例や利用上の課題を抽出した。

▶ 政府が公開するデータポータル

1. アメリカ: www.data.gov
2. イギリス: www.data.gov.uk
3. オーストラリア: www.data.gov.au
4. ニュージーランド: data.govt.nz
5. カナダ: data.gc.ca
6. インド: data.gov.in

▶ 市のレベルでコンテストを開きアプリケーションを集めたことで知られるデータポータル

7. Washington DC: data.octo.dc.gov/
8. New York City: nyc.gov/data

国際化動向調査のまとめ

- データを検索・発見・交換するために、メタデータ記述標準を定めている事例がある。
 - ▶ Australia AGLS
 - ▶ NIEM
 - ▶ ISA など

- 実際にデータを登録・提供する際には、データを容易にアクセスできるための具体的な指針・ベストプラクティスが必要であるが、それは確立していない。
 - ▶ データのカラム名をわかり易いものにする
 - ▶ 具体的な意味の厳密な説明をする など



3. 技術ガイド・規格案

技術ガイド・規格に関する基本方針

■ 目的

- ▶ データをオープンデータとして公開する、またはそれらのデータを利用する人が、その方法を検討する際の参考・推奨文書として例示する。

■ 24年度の対象

- ▶ 実務的なデータのオープン化を目的とした規格・ガイド
 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
- ▶ メタデータやリアルタイムデータ等を流通させる規格
 3. 情報流通連携基盤システム外部仕様書(平成24年度版)



i) 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データ作成に関する技術ガイド

■ 目的

- ▶ 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

- ▶ 公開データの形式に関するガイドである。
 - ◇ データを「どのような形式で」公開するか、指南するものである。
- ▶ 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

技術ガイドのレベル設定

■ レベル1

- ▶ レベル1は、オープンデータが満たすことを強く推奨する要件である。
- ▶ レベル1の要件は、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ことを目的とする。

■ レベル2

- ▶ レベル2は、オープンデータが満たすことを推奨する要件である。
- ▶ レベル2の要件は、データを取得した利用者が、そのデータの項目（タイトル）・値・単位、あるいは構造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を目的とする。

■ レベル3

- ▶ レベル3は、オープンデータ満たすと望ましい要件である。
- ▶ レベル3の要件は、データ本体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ことを目的とする。
- ▶ 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、レベル3の要件とする。

表形式データの技術ガイド

レベル1	ガイド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。
	ガイド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。
	ガイド3	年の値には、西暦表記を備えるべきである。
	ガイド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。
	ガイド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。
	ガイド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。
	ガイド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。
	ガイド8	データの単位が明記されているべきである。
	ガイド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。
	ガイド10	データセットは、オープンな標準データ形式で提供されるべきである。
2	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。
3	ガイド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。
	ガイド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。

表形式データ・レベル1を満たす形式

2010年の人口,2012年の面積
 単位,,km2
 記数単位,1000,
 北海道,5506,83457.06
 青森,1373,9644.7
 岩手,1330,15278.89
 宮城,2348,6862.12
 秋田,1086,11636.3
 ...
 熊本,1817,7267.89
 大分,1197,5099.58
 宮崎,1135,6794.69
 鹿児島,1706,9044.66
 沖縄,1393,2276.64

例1

ガイド7: タイトルは1行
 ガイド8: 単位を明記
 ガイド9: 内容(1行目)・単位(2行目)・記数単位(3行目)を示すタイトルを別々の行に記載

ガイド3: 年の値に西暦表記を備える

年,年(和暦),東京都の人口
 記数単位,,1000
 1920,大正9年,3699
 1925,大正14年,4485
 1930,昭和5年,5409
 1935,昭和10年,6370
 ...
 1985,昭和60年,11829
 1990,平成2年,11856
 1995,平成7年,11774
 2000,平成12年,12064
 2005,平成17年,12576

例2

ガイド1: 1つのデータシートに1種類の表
 ガイド2: セルに整形のための改行・空白・カンマを含めない
 ガイド4: 数値やタイトル・単位以外の情報を、セルに含めない
 ガイド5: すべてのセルが、他のセルと結合されない
 ガイド6: 値がない場合を除き、データセルが空白でない
 ガイド10: オープンな標準データ形式（CSV形式）で提供

参考／表形式データ・レベル1を満たす形式(表計算ソフトで表示)

	2010年の人口	2012年の面積1135
単位		km2
記数単位	1000	
北海道	5506	83457.06
青森	1373	9644.7
岩手	1330	15278.89
宮城	2348	6862.12
秋田	1086	11636.3
熊本	1817	7267.89
大分	1197	5099.58
宮崎	1135	6794.69
鹿児島	1706	9044.66
沖縄	1393	2276.64

年	年(和暦)	東京都の人口
記数単位		1000
1920	大正9年	3699
1925	大正14年	4485
1930	昭和5年	5409
1935	昭和10年	6370
1985	昭和60年	11829
1990	平成2年	11856
1995	平成7年	11774
2000	平成102年	12064
2005	平成17年	12576



ii) オープンデータ化のための
のCSV形式データ規格

オープンデータ化のためのCSV形式データ規格

- 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、追って規定する。
- 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、データセットの冒頭に記述する。

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

```

@Caption,都道府県の人口と面積,
@Creator,オープンデータ流通推進コンソーシアム,
@Date,2013-03-19,
@Language,ja,
@@Title,ja,
,2010年の人口,2012年の面積
@@Unit,,
,,km2
@@Baseval,,
,1000,
@@Datatype,,
,xsd:integer,xsd:double
北海道,5506,83457.06
青森,1373,9644.7
岩手,1330,15278.89
宮城,2348,6862.12
秋田,1086,11636.3
山形,1169,6652.11
福島,2029,13782.76
茨城,2970,6095.72
栃木,2008,6408.28
群馬,2008,6362.33
埼玉,7195,3767.92
千葉,6216,5081.93
東京,13159,2103.97
神奈川,9048,2415.86
"
福岡,5072,4847.2
佐賀,850,2439.65
長崎,1427,4105.75
熊本,1817,7267.89
大分,1197,5099.58
宮崎,1135,6794.69
鹿児島,1706,9044.66
沖縄,1393,2276.64
    
```

ヘッダ

データ本体

@caption	都道府県の人口と面積	
@Creator	オープンデータ流通推進コンソーシアム	
@Date	2013-03-19	
@Language	ja	
@@Title	ja	
@@Unit	2010年の人口	2012年の面積
@@Baseval		km2
@@Datatype	1000	
@@Datatype	xsd:integer	xsd:double
北海道	5506	83457.06
青森	1373	9644.7
鹿児島	1706	9044.66
沖縄	1393	2276.64

付与するヘッダ

本規格に基づくデータ記述例

←を表形式ソフトで表示



iii) 情報流通連携基盤システム
外部仕様書(平成24年度版)

情報流通連携基盤システム外部仕様書案(平成24年度版)

※以下本書の名称を「外部仕様書」と略する。

■ 本書の位置づけ

- ▶ 公開された／公開されるデータを利用したシステム構築手法の1つ。
 - ◇ 保持するデータを公開するときの手法・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トリガとしてデータを取得するような利用法も対象とする。
 - ◇ いわゆるオープンデータ（再配布可能なライセンスにより公開されるデータ）でない、条件付きで閲覧・改編・流用等の利用が許可されるデータも対象とする。
- ▶ 仕様に記載された機能から、必要なものを選択して利用できる。
 - ◇ 全機能の提供を求めているものではない。
 - ◇ 平成24年度の情報流通連携基盤事業では、上記の方針でシステムが実装された。

■ 規定範囲

- ▶ API規格
 - ◇ SPARQL規格に基づくAPI
 - ◇ RESTベースのAPI
- ▶ データ規格
 - ◇ データモデルはRDFに準拠する。
 - ◇ ボキャブラリとして、平成24年度実証で利用したものを例示する。
 - ◆ 既存のボキャブラリ+実証において追加定義したボキャブラリ

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	ボキャブラリ例
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)

外部仕様書のボキャブラリ規定 (各実証で利用したボキャブラリ)

名称	規定範囲	ネームスペース	ボキャブラリ例
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:InsutrialProduct(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)

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	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。	RDF Schemaの知識がなくてもボキャブラリの情報を取得できるようにするため。
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。	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。
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トリガとしてデータを取得する利用法に対応するため。

外部仕様書に関するケーススタディ

■ 背景

- ▶ 東日本大震災における、情報の横連携（流通・共有）の重要性が顕在化から、組織や業界内で利用されているデータを社会でオープンに利用できる環境の整備が重視されている。
 - ◇ これを実現する方法の1つとして、総務省では、情報流通連携基盤を整備している。
 - ◆ 情報流通連携基盤 = 主体、分野・領域に閉じない情報流通・利活用のための共通基盤としての、情報・知識やサービスの連携・共有環境の整備のための汎用性ある技術・運用ルール等が整った環境。
 - ◇ 平成24年度は、情報流通連携基盤のシステム外部仕様書（ドラフト版・以下「外部仕様書」と呼ぶ）の設計を行った。外部仕様書の規定範囲は以下の通り。
 - ◆ データモデル（RDFに準拠）
 - ◆ ボキャブラリ（広く流通しているもの+実証のために追加したもの）
 - ◆ API（データ交換のための外部インタフェース。SPARQLベース+RESTベース）

■ 目的

- ▶ 情報流通連携基盤の適用可能性を検証
- ▶ データのオープン化に対するメリットの可視化
- ▶ 外部仕様書に関する評価と課題の抽出

■ 平成24年度に実施した適用可能性検証対象分野

- ▶ 公共交通／災害関連／ボーリング（地盤）／生鮮農産物トレーサビリティ／水産物トレーサビリティ

各実証にて追加されたボキャブラリ

名称	規定範囲	ネームスペース	ボキャブラリ例
地盤	地盤情報の記述に関するボキャブラリ	http://www.jibaninfo.jp/vocab/ucr/gs#	gs:hasBoringData(ボーリングデータ), gs:stratumName(地層・岩体名)
水産物トレーサビリティ	水産物トレーサビリティ実証で利用したボキャブラリ	http://fishery.suisancloud.net/ns/fishery#	fishery:BrandName(ブランド名), fishery:qualityGrade(等級)
生鮮農産物トレーサビリティ	生鮮農産物トレーサビリティ実証で利用したボキャブラリ	http://www.agri-info.jp/vocab/ag#	ag:cultivationArea(栽培面積), ag:fieldCount(圃場枚数)
交通	公共交通実証で利用したボキャブラリ	http://opendata.ubin.jp/vocab/puti#	puti:StationMasterOffice(駅長室), puti:operator(路線の運営会社), puti:delay(遅延時間)
気象	気象情報に関するボキャブラリ	http://opendatafordisasters.jp/jmaxml1/eb/ など	jmr_eb:Pressure(気圧), jmr_eb:WindDirection(風向), jmr_eb:Precipitation(降水量)
防災	防災関係のボキャブラリ	http://opendatafordisasters.jp/cao/caor/ など	caor:gasServiceFailureHouse(ガス供給停止数), caor:waterSuspensionBuilding(断水世帯数)

実証事業者から挙げられた、APIやボキャブラリに対する課題

■ 全文検索に関する要求

- ▶ たとえば地盤情報の事業では、全文検索機能に対する要求が高かった。
 - ◇ プロパティに関係なく値に「〇〇岩」を含むデータが欲しい、というようなクエリ。

■ 地理空間情報の扱い方に関する要求

- ▶ 交通実証の事業では、地理空間情報をGMLやGeoJSONの形で得たいという要求があった。
 - ◇ 地理空間情報のフォーマットとして、GML・KML・Shape・GeoJSONなどが広く使われている。
- ▶ 住所で検索したいという要求もあった。

■ 単位系（特に物理単位系）に関する問題

- ▶ 地盤情報の事業では、測定時期によって、値の単位系が違うという問題がある。
 - ◇ 単位系が途中でSI単位系に変わったことに起因する。たとえば[kg重]→[N]など。
 - ◆ このため、単純に値の大小で比較できない。
 - ◇ 今回は、データが基づいている規格書のバージョンをつけることで、値の単位解釈をした。
- ▶ 生鮮食料品の事業では、放射能値などの単位を新規に利用した。

■ SPARQLに関して

- ▶ クエリ、データ量、サーバの実装によっては実用的な応答時間が得られない場合がある。
- ▶ 仕様だけではカバーしきれない。分散化技術やクエリの利用制限ポリシー等を含めた検討が必要。

■ その他

- ▶ 交通実証の公募案件では、(APIが提供する)RDF/JSON形式に慣れないという意見があった。
- ▶ 外部仕様書の記述例をもっと充実させてほしい、という要望があった。
 - ◇ ボキャブラリだけでなく、データの構造や問い合わせ方に関する解説(ベストプラクティスの提示)も必要。
 - ◇ 開発者サイトによる情報提供・意見交換などが必要。

(参考) 外部仕様書に基づくシステムを構築した実証事業者からの意見

■ SPARQL部分のAPIに関して

- ▶ RDF/SPARQLおよびlinked dataを使うという前提であれば有用である。
 - ◇RDF化しやすい（RDF化することによるメリットが大きい）データとそうでないデータがある。
- ▶ クエリ、データ量、サーバの実装によっては実用的な応答時間が得られない場合がある。分散化技術やクエリの利用制限ポリシー等を含めた検討が必要。

■ REST部分のAPIに関して

- ▶ 国が規定する場合は、外部仕様書レベルの仕様が必要。
- ▶ 民間が自由に作るという観点では、RESTベースの仕様まで規定する必要はない。
- ▶ 必須とオプションの少なくとも2段階に分けた方がよい。

■ ボキャブラリに関して

- ▶ ボキャブラリはトップダウンに定義できない。
- ▶ いろいろなボキャブラリを公開することが必要。そのうち、使われるものが残る。
- ▶ 情報処理の範疇でない分類もある。（自治体が定義した防災に関する分類など）



4. 次年度の課題

次年度の課題

■ ガイド・規格の精査

- ▶ Call for Commentsを実施し、得られたコメントを反映
- ▶ 技術ガイドの整理・拡充
- ▶ ボキャブラリを共有（登録・参照）する仕組み
- ▶ 外部仕様書のプロファイル化

■ ガイド・規格を普及させるための体制や周辺ツール整備

- ▶ 規格やサービスを維持・メンテナンスする組織体制



OPEN DATA

オープンデータ流通推進コンソーシアム