



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2回利活用・普及委員会資料

オープンデータガイド第1版

～オープンデータのためのルール・技術の手引き～技術編

2015.2.13

オープン&ビッグデータ活用・地方創生推進機構 事務局

目次

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

第2章 オープンデータの動向と意義

第3章 オープンデータの作成・公開手順

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

第5章 オープンデータ利用ルールの概要

第6章 利用ルールの比較と望ましいルール

第7章 利用ルールに関する今後の見直しの方向性

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9章 オープンデータのための技術的指針

付録

第10章 オープンデータに関する規格・ツール

第11章 データカタログシステムCKAN



第I部 Getting Started: オープンデータをはじめよう

6. オープンデータの作成・公開手順

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。

6. オープンデータの作成・公開手順／Step2: 現状把握

■ このステップでの実施内容

- ▶ 各部署が管理しているデータをまとめる。

■ 現状把握に当たり注目すべき項目

1. データの形式

- ◇紙 → 電子データがない資料を公開するには、紙をスキャンする必要がある。
- ◇電子データ → ファイル形式を確認すべき。

2. データの管理者

- ◇管理者は設定されているか。統一されているか。

3. データの更新頻度

- ◇データはどれくらいの頻度で更新されるか？
- ◇年に1回更新／月に1回更新／適宜更新等

4. データの権利関係

- ◇第三者が著作権等の権利を有するデータはあるか？
- ◇法令上の制約があるか？ 等

5. ニーズ分析

- ◇以下のようなニーズの高いデータからオープンデータとしての公開に取り組むことも有用。
 - ◆ 情報利用者から多く問い合わせられるデータ
 - ◆ 他の同様の組織で公開されているデータ

6. オープンデータの作成・公開手順／Step 3: 計画立案

■ このステップでの実施内容

- ▶ オープンデータの対象とするデータやその作成・公開手法を明確にする。
- ▶ マイルストーンを作成し、それに基づきスケジュールを立てることが望ましい。

■ 留意事項

1. データ形式・システムの準備計画

◇どのレベルの「データ」と「データカタログ」を準備するか、方針を策定（→ 8.4節参照）

2. 運用ルールの策定

◇データの入手手順・頻度を明確にする。

◇適宜更新される場合は、更新方法をルール化。

3. 利用ルールの設定

◇第三者権利問題や法令上の制約がある場合は、それを踏まえ、利用ルールの内容や適用範囲を整理する。

4. スモール・スタートの原則

◇作業は段階的に行い、完了したのから順次公開できるように、マイルストーンを設定する。

◇年度ごとに目標・計画を立てることが望ましい。

6. オープンデータの作成・公開手順／Step 4: 公開作業

■ このステップでの実施内容

- ▶ 立案した計画に基づき、調達をかける等して必要なツールを揃え、オープンデータを作成・整形し、公開の準備作業を行う。

■ 留意事項

1. 公開時に明確にすべき情報

- ◇メタデータ（どんなデータか？）
- ◇アクセス方法（そのデータはどのようにして取得できるか？）
- ◇利用ルール（そのデータはどのような条件で取得・利用できるか？）

2. 公開による影響

- ◇オープンデータとして公開したデータは、全世界に対して公開される。
→ 海外からの問い合わせも想定される。

3. データを公開するサーバに関する留意事項

- ◇公開するサーバにどれくらいのアクセスがあると予想するか？
 - ◆ 予想外のアクセスが集中し、サーバの処理が追いつかなくなると、公開したデータに対するアクセス障害が発生する。
- ◇リアルタイムデータを扱う場合の留意点
 - ◆ サーバの記憶容量を動的に消費するため、サーバの記憶容量の枯渇によりアクセス障害が発生する可能性がある。
- ◇公開サービスを運用する業者や部署と、事前に協議しておくことが望ましい。

4. データの信頼性

- ◇データの流過程において、情報利用者による改ざん、情報提供者の意図しない編集・変更の可能性がある。

5. プライバシー・匿名化

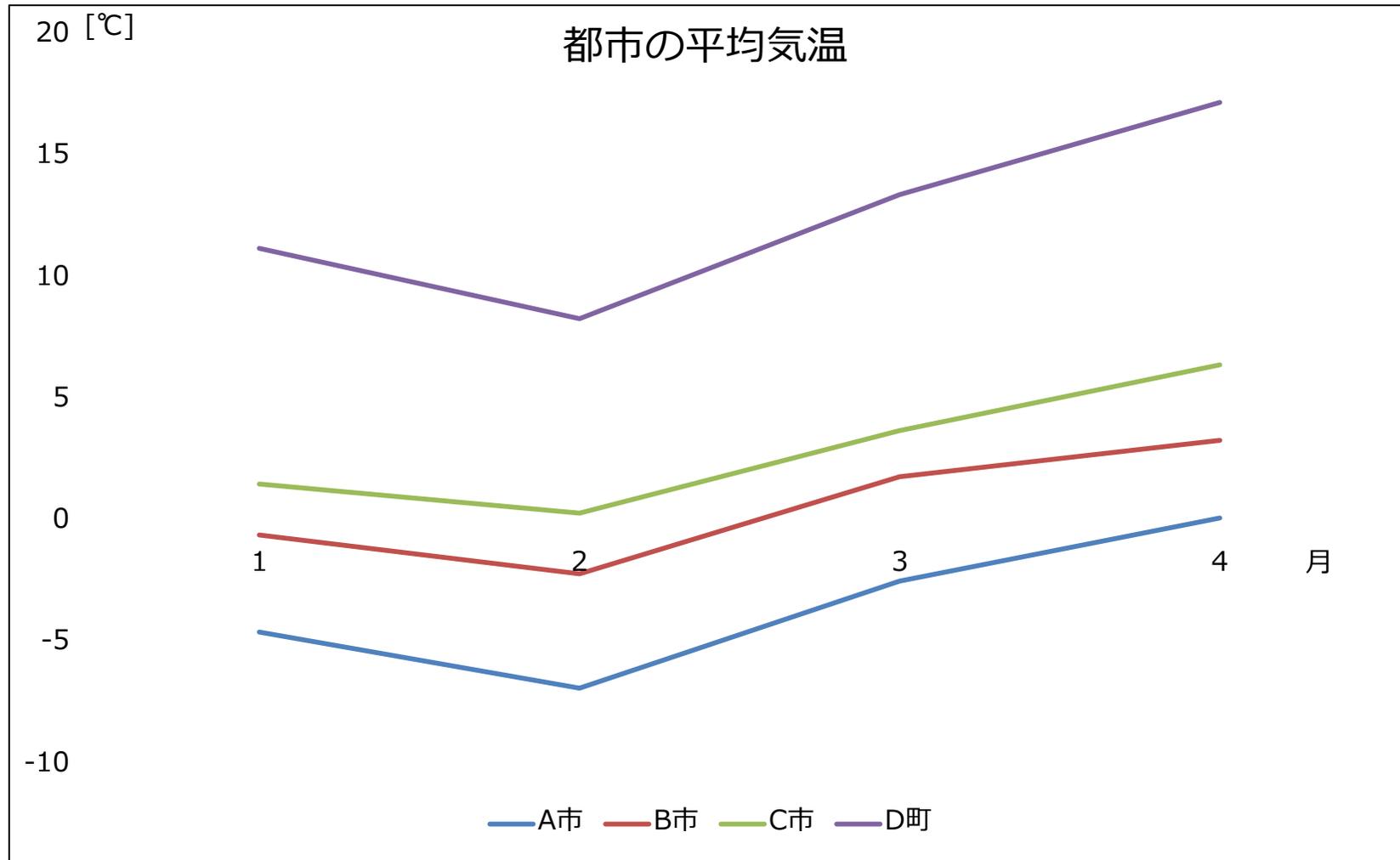
- ◇そのデータに個人を特定する情報が含まれていないか、確認する必要がある。
- ◇必要に応じて、匿名化の手法を利用して、プライバシーを考慮すべきである。



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

1. 機械判読性に適したデータとは？

■ このデータを、コンピュータが判読するには？



1. 機械判読性に適したデータとは？

■ 機械判読に適したデータとは？

- ▶ コンピュータプログラムがデータの論理的な構造を判読でき、構造中の値（表の中に入っている数値、テキスト等）を自動的に編集・加工・改変等できること。
“Machine Readable”の日本語訳であり「機械可読」ともいう。
- ▶ 機械判読性の高い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。

1. 機械判読性に適したデータとは？

■ 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ためには画像解析等の技術が必要であり、これは容易ではない。

★2:この形式のファイルは構造化されているため、対応するソフトウェアを用意すれば、コンピュータはこれ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: この形式のデータの解析方法は公開されているため、★3の形式のデータ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について、9.3節で詳しく述べる。

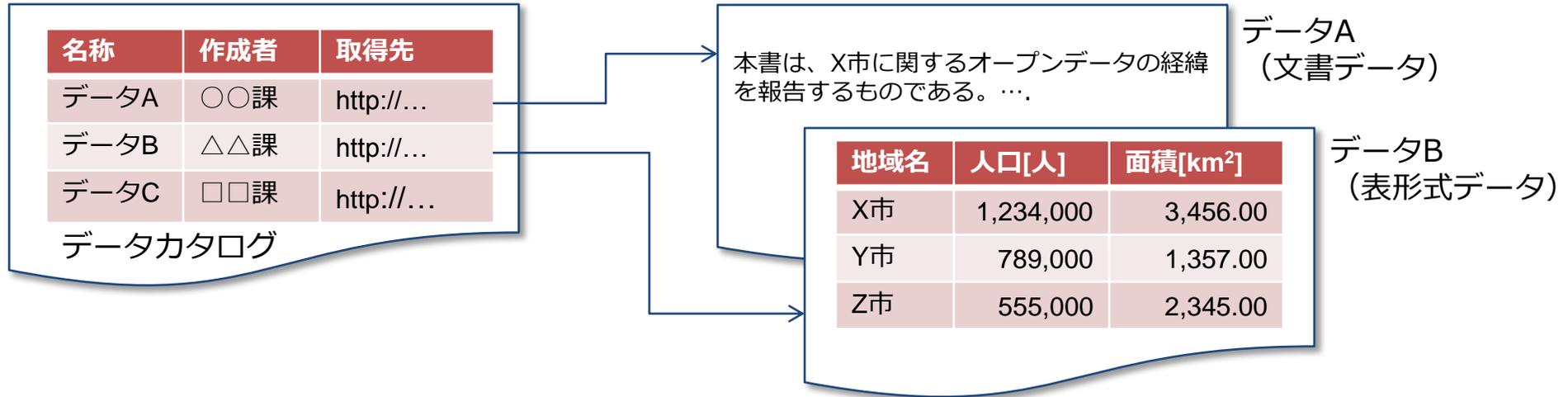
★4～: この形式のデータは、相互に接続でき、コンピュータによるデータのマッシュアップが容易になる。



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

2. データカタログとは？

■ 「データ」と「データカタログ」との関係



■ データカタログの必要性

- ▶ 公開するデータが増加するにつれて、それらのデータを整理し、検索、一覧する機能に対する要求が高まる。このような要求に応えるものが、データカタログである。

■ さまざまなデータカタログ

- ▶ データの名称、取得先等を表形式データとしてまとめたものも、一種のデータカタログである。
- ▶ 高機能なデータの管理・検索・一覧機能を提供するためには、データカタログシステムを導入することが望ましい。

3. オープンデータと識別子

■ 識別子とは

- ▶ 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。

■ 識別子が必要である理由

- ▶ 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る。これらのデータは、コンピュータが一意に識別できるべきである。
- ▶ オープンデータに含まれる実物や組織、場所等も、一意に識別されることが望ましい。これは、組織や場所が文字列で提供されている場合、以下のような問題が発生する可能性があるためである。
 1. 情報利用者のコンピュータは、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可能性がある。（例：「中央一丁目一番地一号」と「中央1-1-1」）
 2. 情報利用者のコンピュータは、同一名称だが違う意味である組織や場所を、文字列だけでは識別できない。（例：全国各所にある「中央一丁目」）

■ 識別子とコード

- ▶ コードとは
 - ◇カテゴリ化される概念や事物に対して付与される番号。
 - ◇対象とする概念や事物を短縮して符号化するために規定される。
- ▶ 識別子とコードは異なる。
 - ◇コードには意味が付与されているが、識別子に意味が付与されているとは限らない。
- ▶ ただし、多くの場合、コードは識別子として機能する。

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等)	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、APIを実装して公開する。
データカタログ	存在しない。	カタログを表形式データ（CSV等）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等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+データカタログシステム等	Webサーバ+データカタログシステム+情報流通連携基盤等

5. 機械判読に適したデータを作成するための指針

- 指針のグレード: 満たすべき指針の重要度にあわせて2つのグレードを設ける。
 - ▶ グレード1
 - ◇ グレード1は、オープンデータが満たすことを強く推奨する指針であり、以下を満たすことを目的とする。
 - ◆ データ形式に関する標準的な規格がある場合は、それに矛盾しないこと。
 - ◆ データを取得した利用者が、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
 - ◇ グレード2は、オープンデータが満たすことを推奨する指針であり、以下を満たすことを目的とする。
 - ◆ 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。
- 対象とするデータ
 - ▶ 表形式データ
 - ▶ 文書データ
 - ▶ 地理空間情報
 - ▶ リアルタイムデータ
- 各指針に関する記述内容
 - ▶ 表形式データを中心に指針を満たさない例と満たす例を明記し、それに対して解説する。
 - ▶ 以下、各データに関する指針のみを記す。

5. 機械判読に適したデータを作成するための指針／ファイル形式

■ 基本方針

- ▶ 機械判読性の高い形式を利用することが望ましい。
- ▶ 代表的なファイル形式を、オープンデータの技術レベルに基づいてまとめると、下記のようなになる。

	Level 1	Level 2/3	Level 4
表形式データ	xls (Microsoft Excel形式)	CSV xlsx (Office Open XML) ods (OpenDocument) JSON	RDF/XML RDF/JSON, JSON-LD Notation3 Turtle等のRDF形式
文書形式データ	doc (Microsoft Word形式)	HTML XML docx (Office Open XML) odt (OpenDocument)	
地理空間情報	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)	

5. 機械判読に適したデータを作成するための指針／表形式データ

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。

5. 機械判読に適したデータを作成するための指針／表形式データ

1ファイルに複数の表がある（指針1を満たさない）例

ファイルX					
月	A市	B市	C市	D町	
1	-4.5	-0.5	1.6	11.3	
2	-6.8	-2.1	0.4	8.4	
3	-2.4	1.9	3.8	13.5	
4	0.2	3.4	6.5	17.3	
月	A市	B市	C市	D町	
1	230	58	377	103	
2	169	43	422	122	
3	144	54	322	144	
4	232	102	145	133	

ファイルを2つに分割（指針1を満たす）

ファイルX					
月	A市	B市	C市	D町	
1	-4.5	-0.5	1.6	11.3	
2	-6.8	-2.1	0.4	8.4	
3	-2.4	1.9	3.8	13.5	
4	0.2	3.4	6.5	17.3	

ファイルY					
月	A市	B市	C市	D町	
1	230	58	377	103	
2	169	43	422	122	
3	144	54	322	144	
4	232	102	145	133	



ヘッダが複数行からなる（指針2を満たさない）例

月	気温			
	A市	B市	C市	D町
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

ヘッダが1行に統合（指針2を満たす）

月	A市の気温	B市の気温	C市の気温	D町の気温
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3



5. 機械判読に適したデータを作成するための指針／文書データ

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	テキスト形式の文書形式データを利用している場合は、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5	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。

5. 機械判読に適したデータを作成するための指針／地理空間情報

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、座標系と描画縮尺（入力精度）を示すべきである。
2	指針2	地理空間情報は、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当たっては、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）を用いる。
	指針3	地理空間情報に対する、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。

5. 機械判読に適したデータを作成するための指針 ／リアルタイムデータ

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空間情報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。