



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～

概要版

2014.07.31

オープンデータ流通推進コンソーシアム

「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: 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）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）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

◇政府標準利用規約（第1.0版）の今後の見直しに当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、地理空間情報、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 データカタログシステムCKAN

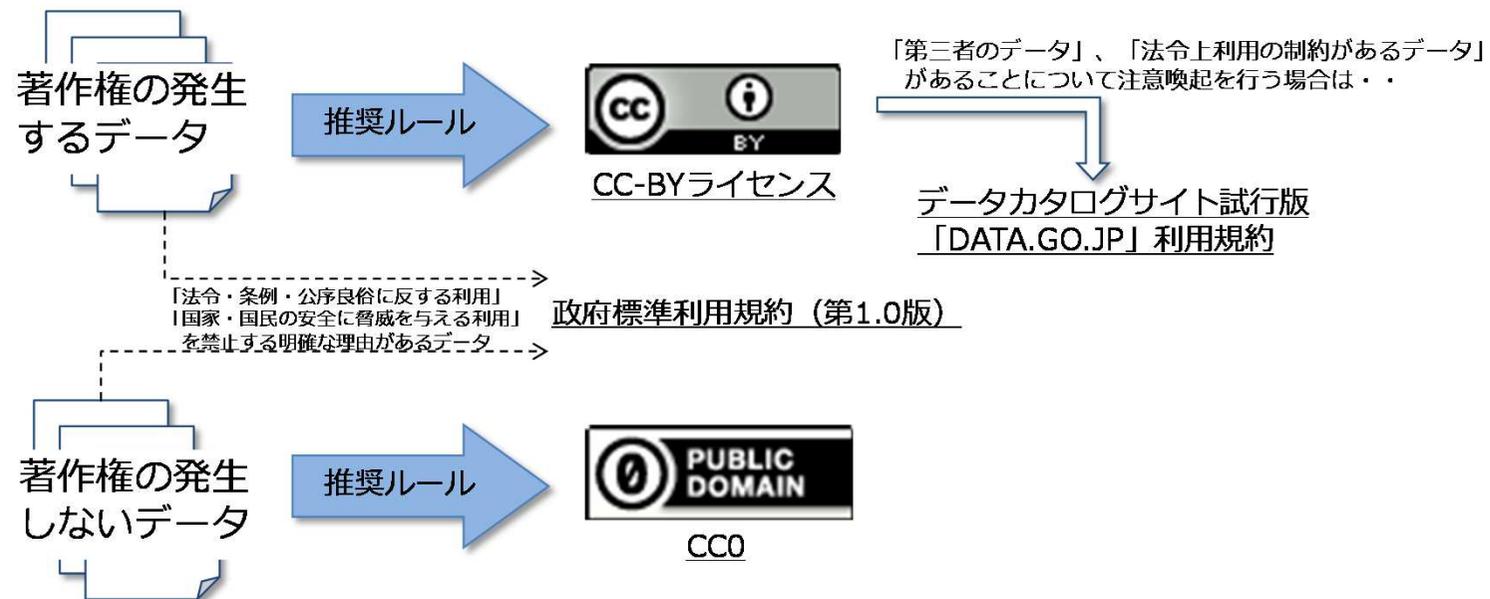
◇データカタログシステムであるCKANの概要とその使用方法を解説する。



エグゼグティブサマリ

利用ルールを選択

- 「オープンデータ」と言えるためには、提供する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、著作権の発生しないデータ（数値、簡単な表・グラフ等）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で、具体的注意喚起等を行う場合には、「データカタログサイト試行版（DATA.GO.JP）利用規約」を参考にすることが望ましい。



機械判読に適したデータの作成・公開①

■ 機械判読とは

- ▶ 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。
- ▶ 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。

■ オープンデータの技術レベル

- ▶ 機械判読性を高める観点から、作成・公開するデータ・データカタログ・識別子に関してまとめた指標。
- ▶ 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。

	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

機械判読に適したデータの作成・公開②

■ 機械判読に適したデータを作成するための指針

▶ 識別子に関する指針

◇ユニークであることと、共通に利用できる体系であることが望ましい。

- ◆ グローバルにユニークな識別体系／公的機関が定める識別子体系・コード体系／URIとして表現できる体系（候補は付録10.2節）

◇ファイル形式に関する指針

◇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), …
地理空間情報	KML, GML

▶ データに関する指針

◇2つのグレードを設けて、機械判読に適したデータの指針を提示する。

- ◆ 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であり、データ形式に関する標準的な規格に矛盾しないことと、データを取得した情報利用者が、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を目的とする。
- ◆ 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であり、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を目的とする。

◇例: 表形式データのグレード1指針

- ◆ 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。

◇例: 地理空間情報のグレード1指針

- ◆ 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。



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

■ 本章の概要

- ▶ 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。

■ 本章の構成

1. 本書の目的

◇オープンデータの特徴を示し、本書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ものであることを示す。

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、各部ごとの対象読者を示す。

3. 本書の構成

◇章ごとの構成と、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。

1.1 本書の目的

■ 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、透明性・信頼性の向上だけでなく、国民参加・官民協働の推進、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。特に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もの。

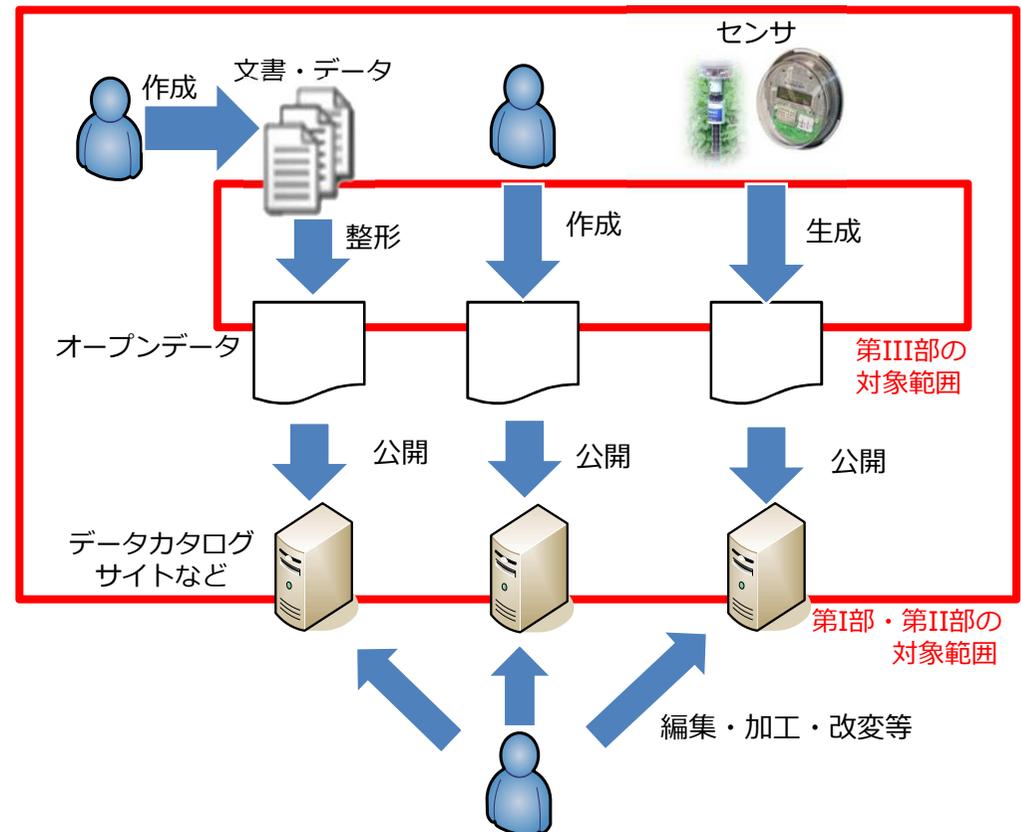
1.2 本書の対象読者

■ 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・改変等する。
- ▶ 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。



1.3 本書の構成

■ 知りたい内容と参照すべき章との対応表

知りたい内容	該当する章
1. オープンデータの定義や背景・意義が知りたい。	第2章
2. オープンデータに関する国内外の動向を知りたい。	第2章
3. 組織体制や準備・計画すべきこと等、データをオープンデータにするまでの手順を知りたい。	第3章
4. データをオープンデータにする際には二次利用を認める利用ルールをつけると聞いたが、その背景や考え方について知りたい。	第4章
5. 具体的にどのような利用ルールがあり、それはどのような特徴を持っているのか知りたい。	第5章
6. どの利用ルールを適用すべきかを検討するための視点や、その視点に基づく各利用ルールの評価について知りたい。	第6章
7. オープンデータにすることが決まったが、データにどのような利用ルールをつけるべきか知りたい。	第6章
8. 利用ルールについて、政府における今後の見直しの方向性について知りたい。	第7章
9. 機械判読性を高めるために有用なデータ形式や識別子体系、データ伝送プロトコルについて知りたい。	第8章 第10章
10. オープンデータを作成・編集する際に、どのような技術レベルを目指すべきか知りたい。	第8章
11. 表形式データ、文書形式データ、地理空間情報等、様々なデータをオープンデータにしたいが、それらの作成・編集に際して技術的に留意すべき事項を知りたい。	第9章
12. Webサービス、GISツール等オープンデータの作成・編集・公開に有用なツールにどのようなものがあるか知りたい。	第10章
13. 代表的なデータカタログシステムの一つであるCKANと、その利用方法について知りたい。	第11章

1.4 用語定義

■ 本書で利用する用語の定義

用語	定義
データ	オープンデータの対象となる情報一般のこと。著作権の発生する情報も発生しない情報も含む。
公共データ	国、地方公共団体、独立行政法人、公共企業等の保有しているデータ。
コンテンツ	データと同様の意味を持つ。本書では引用箇所以外では使用しない。
オープンデータ	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。
情報提供者	オープンデータとしてデータを提供する者又は機関。
情報利用者	オープンデータとして公開されているデータを二次利用する者又は機関。
二次利用	情報提供者の提供したデータをもとに、情報利用者が何らかの編集・加工・改変等を行い、新たなデータを作成することや、情報提供者の提供したデータの単なる複製や再配布を行うこと。
マッシュアップ	情報利用者が、複数の情報提供者の提供したデータ同士や、自らの保有するデータを組み合わせて、新たなデータを作成すること。
ライセンス	情報提供者がデータを提供する際に指定する利用条件。著作権に基づいて情報提供者と情報利用者が契約するという構成をとる。本書では引用箇所や固有名詞以外では使用しない。
利用ルール	情報提供者がデータを提供する際に指定する利用条件。著作権に基づかない契約や、情報提供者による一方的な宣言も含む。
改ざん	オリジナルデータを改変し、それをオリジナルデータだと偽ること。
機械判読	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	対象とするデータに対する機械判読の可能性。
メタデータ	公開するデータに関して、それがどのようなデータであるかを示す情報。
データカタログ	データの所在、種類、名称等、公開しているデータに関する情報（メタデータ）をまとめたもの。データの目録・索引。
表形式データ	行と列の、縦横2次元状に配列されたデータ。
文書形式データ	1次元状に配列された文字を主な構成要素とし、一部図や表等を含み、人間がそれを読むことによって人間に何らかの作用を与えることを目的としたデータ。
リアルタイムデータ	時刻に応じて、値が刻々と変化するデータ。
地理空間情報	空間上の特定の地点又は区域の位置を示す情報と、これに関連付けられた情報。例えば、2次元平面状の地図の地点や領域と関連づけられたデータ。
語彙(ボキャブラリ)	ある分野に属する物事やデータを記述するために、その分野で共通に理解されるべき属性や種別に関する意味定義の集合。

第2章 オープンデータの動向と意義

■ 本章の概要

- ▶ オープンデータの背景について理解するため、日本政府、地方公共団体、諸外国におけるオープンデータに関する主な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。

■ 本章の構成

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」の記述に基づき、オープンデータの定義を行う。

2.1 オープンデータに関する主な動向

1. 日本政府の取組

- ▶ オープンガバメントから取組が始まる。
- ▶ 「電子行政オープンデータ戦略」（2012年7月4日 高度情報通信ネットワーク社会推進戦略本部決定）を契機として、日本政府におけるオープンデータに関する取組が急速に進んでいる。
- ▶ 2013年6月14日に閣議決定された「日本再興戦略」や「世界最先端IT国家創造宣言」においても、オープンデータは重要な施策の一つとして取り上げられている。

年月	名称	位置づけ
2009.10.14	電子経済産業省アイデアボックス公開	経済産業省
2010.07.29	「オープン・ガバメント・ラボ」公開	経済産業省
2011.03.15～現在	東京電力の計画停電、電力データ公開	東京電力
2011.07.01	「データボックス」公開	経済産業省
2012.01.17	復旧・復興支援制度データベース（制度のオープン化）	内閣官房、復興庁、経済産業省
2012.07.04	電子行政オープンデータ戦略	高度情報通信ネットワーク社会推進戦略本部決定
2012.07.27	オープンデータ流通推進コンソーシアムの設立	オープンデータ流通推進コンソーシアム
2012.09～現在	オープンデータ実証実験（情報流通連携基盤の開発等）	総務省
2013.01.18	「Open DATA METI」（β版）公開	経済産業省
2013.03.28	電子行政オープンデータ実務者会議設置	高度情報通信ネットワーク社会推進戦略本部決定 （2012.11.30～2013.03.27は企画委員会の下に設置）
2013.04.19	情報通信白書及び情報通信統計データベースのオープンデータ化	総務省
2013.06.10～順次試行	統計におけるオープンデータの高度化（API機能の提供、統計GIS機能の強化等）	総務省統計局、独立行政法人統計センター
2013.06.14	日本再興戦略（公共データの民間開放と革新的電子行政サービスの構築）	閣議決定
2013.06.14	世界最先端 IT 国家創造宣言（オープンデータ・ビッグデータの活用の推進）	閣議決定
2013.06.14	電子行政オープンデータ推進のためのロードマップ	高度情報通信ネットワーク社会推進戦略本部決定
2013.06.18	オープンデータ憲章	G8サミット（英国ロック・アーン）での合意
2013.06.25	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）	各府省情報化統括責任者（CIO）連絡会議決定
2013.10.29	日本のオープンデータ憲章アクションプラン	各府省情報化統括責任者（CIO）連絡会議決定
2013.12.20	政府データカタログサイト試行版「DATA.GO.JP」公開	内閣官房
2013.04.25	電子行政分野におけるオープンな利用環境整備に向けたアクションプラン	各府省情報化統括責任者（CIO）連絡会議決定

2.1 オープンデータに関する主な動向

2. 地方自治体の取組

- ▶ 2012年7月の「電子行政オープンデータ戦略」の決定前から、一部で先行的な取組が行われており、同戦略の決定後は、オープンデータの動きが更に加速化している。
- ▶ データポータル等によるオープンデータでのデータ公開を行っている例が多い。
- ▶ ホームページ全体をオープンデータにしたり（福井市）、県内市町村でデータ形式などを統一する取組（福井県）を行っている例もある。

自治体名	取組名称 (URL)	概要
福井県鯖江市	データシティ鯖江	地方公共団体でのオープンデータに関する先駆的な取組。地元企業と連携して様々なアプリを開発。2013年度には、総務省のオープンデータ実証実験に協力して、オープンデータを拡充。
千葉県流山市	流山市オープンデータトライアル 流山市議会オープンデータトライアル	ホームページのリニューアルに併せて、市役所と市議会が同時にオープンデータの取組を開始。議案に対する議員毎の採決結果等も公開。
横浜市	横浜オープンデータポータル	2012年度から民間団体に対して図書館情報等の提供を支援。2013年度にオープンデータ推進プロジェクトを庁内に設置したほか、総務省のオープンデータ実証実験に協力して横浜市自身のデータをオープンデータとして公開。
静岡県	ふじのくにオープンデータカタログ	都道府県で初めてデータポータルを開設。県内市町村も利用可能（裾野市が利用）。
静岡県・山梨県	富岳3776景	富士山の写真を位置情報付きで誰でも投稿可能。投稿した写真はオープンデータとして公開される。災害（大雪）の際には災害情報共有ポータルとして活用された。
福井県	オープンデータライブラリ 県内公共データの形式統一	オープンデータと、オープンデータを活用したアプリを公開。県内市町村のデータ形式の統一に向けた取組にも着手。
福井市	市のホームページ利用規約の改訂	市のホームページ全体にCC-BY-SAライセンスを付与。
青森県	あおもり映像素材ライブラリー	県職員が撮影した県内の様々な映像素材をオープンデータとして公開。
福島県会津若松市	オープンデータライセンスによるデータ公開	オープンライセンス（CC-BY）に加え、オープンドキュメント形式（ODF）でも公開。
神奈川県横須賀市	防災情報のオープンデータ公開	横須賀市が公開する防災関連データを機械判読に適した形式に変換して公開し、これを利用したアイデアソン・ハッカソンを実施。

2.1 オープンデータに関する主な動向

3. 国際的な動向

- ▶ 欧米を中心に、2000年代後半からオープンデータに関する取組が進められている。
- ▶ 米国では、2009年1月21日にオープンガバメントの3原則（透明性・市民参加・官民連携）を掲げ、5月にポータルサイトData.gov開設。
- ▶ EUでは、2003年の「PSI（公共保有データ）の再利用に関する指令」を契機に各国がオープンデータへの取組を開始。英国が2007年に設置したPower of Informationタスクフォースの取組は、世界のオープンデータに関する取組の源流となった。
- ▶ 2013年のG8サミットでは「オープンデータ憲章」が合意された。
 - ◇5つのオープンデータの原則が示された。
 - ◇2013年10月末までにオープンデータ憲章履行のための行動計画を作成し、2014年10月と2015年に履行状況の報告を行うことが定められた。

時期	実施事項	国名
2003年	PSI（公共保有データ）の再利用に関する指令	欧州委員会
2005年7月	PSI再利用に関する規則	イギリス
2007年	Power of Informationタスクフォースの設置	イギリス
2009年1月	「透明性とオープンガバメントに関する覚書」	米国
2009年5月	データポータルサイトData.gov開設	米国
2009年9月	データポータルサイトdata.gov.uk開設	イギリス
2010年3月	政府サイトが初めてCC0を採用	オランダ
2010年5月	「透明性アジェンダ」発表	イギリス
2010年11月	Etalabの設立に関する閣議決定	フランス
2011年12月	データポータルサイトdata.gouv.fr開設	フランス
2011年12月	欧州オープンデータ戦略	欧州委員会
2013年6月	オープンデータ憲章	G8

2.1 オープンデータに関する主な動向

■ G8サミット「オープンデータ憲章」における「オープンデータの原則」

1. Open Data by Default（原則としてのオープンデータ）

◇データによっては、公表出来ないという合理的な理由があることを認識しつつ、この憲章で示されているように、政府のデータすべてが、原則として公表されるという期待を醸成する。

2. Quality and Quantity（質と量）

◇時宜を得た、包括的且つ正確な質の高いオープンデータを公表する。

◇データの情報は、多言語に訳される必要はないが、平易且つ明確な言語で記述されることを確保する。

◇データが、強みや弱みや分析の限界等、その特性がわかるように説明されることを確保する。

◇可能な限り早急に公表する。

3. Usable by All（すべての人々が利用できる）

◇幅広い用途のために、誰もが入手可能なオープンな形式でデータを公表する。

◇可能な限り多くのデータを公表する。

4. Releasing Data for Improved Governance（ガバナンス改善のためのデータの公表）

◇オープンデータの恩恵を世界中の誰もが享受出来るように、技術的専門性や経験を共有する。

◇データの収集、基準及び公表プロセスに関して透明性を確保する。

5. Releasing Data for Innovation（イノベーションのためのデータを公表）

◇オープンデータ・リテラシーを高め、オープンデータに携わる人々を育成する。

◇将来世代のデータイノベーターの能力を強化する。

2.2 オープンデータの意義

■ 「電子行政オープンデータ戦略」に挙げられた3つの意義

- ▶ 透明性・信頼性の向上
- ▶ 国民参加・官民協働の推進
- ▶ 経済の活性化・行政の効率化

■ 「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）」に整理された意義

▶ 経済の活性化、新事業の創出

◇データ収集や各種コードによるデータの横断的利用が機械で自動的に可能になることからコスト圧縮ができ、新しいサービスを提供するビジネスが可能となる。（例えば、気象、地質、交通その他の観測・調査データのような専門的データを収集・分析してビジネスに活用するなど）

▶ 官民協働による公共サービス（防災・減災を含む。）の実現

◇複数の行政機関や民間のデータを組み合わせることで、民間からも、生活利便を高めるサービスや災害時に有用なサービスを提供できる。（例えば、子育て、教育、医療、福祉等の身近な公共サービスの内容、品質等を利用者に分かりやすく示す、災害時に迅速に複数の情報を組み合わせた情報発信が可能となるなど）

▶ 行政の透明性・信頼性の向上

◇政策・事業に関する計画、決定過程、決定内容、結果等について、横断的に検索・集計・比較することで、政策の変化・特徴の把握や、政策の妥当性の理解・評価ができる。（例えば、補助金や政府支出について、府省、分野、地域、支出先等別に分析するなど）

2.3 本書におけるオープンデータの定義

■ 5★Open Dataによるオープンデータの整理

- ★ 1 オープンなライセンスで提供されている（データ形式は問わない／画像やPDF等のデータでも可）
- ★ 2 構造化されたデータとして公開されている（ExcelやWord等のデータ）
- ★ 3 非独占の（標準化された）形式で公開されている（CSV等のデータ）
- ★ 4 物事の識別にURIを利用している（他のデータから参照できる）
- ★ 5 他のデータにリンクしている（Linked Open Data）

■ 「オープンデータハンドブック」によるオープンデータの定義

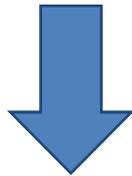
- ▶ 自由に使えて再利用もでき、かつ誰でも再配布できるようなデータのこと。従うべき決まりは、せいぜい『作者のクレジットを残す』あるいは『同じ条件で配布する』程度である。

■ 「電子行政オープンデータ戦略」による4つの基本原則

1. 政府自ら積極的に公共データを公開すること
2. 機械判読可能な形式で公開すること
3. 営利目的、非営利目的を問わず活用を促進すること
4. 取組可能な公共データから速やかに公開等の具体的な取組に着手し、成果を確実に蓄積していくこと

2.3 本書によるオープンデータの定義

- 「電子行政オープンデータ推進のためのロードマップ」によるオープンデータの定義
 - ▶ 「機械判読に適したデータ形式のデータ」を「営利目的も含めた二次利用が可能な利用ルールで公開」すること



- 以上を踏まえて、本書では、オープンデータを以下のとおり定義する。
 - ▶ 「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。

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。

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. 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。

第3章 オープンデータの作成・公開手順

■ 本章の構成: 6つのそれぞれのステップを解説

1. オープンデータ推進組織の設立
2. 現状把握
3. 計画立案
4. 公開作業
5. 公開・運用
6. 改善点の洗い出し

■ 留意点: スモールスタートの原則

- ▶ 上記に掲げたオープンデータ推進組織の設立や現状把握等、すべての準備を完了してから実施することは必須でない。
- ▶ 比較的オープンデータにしやすいデータから着手するというのも一つの方法である。
 - ◇例えば、個別の部署において既に公開されている情報から着手する等。

3.1 オープンデータ推進組織の設立

■ オープンデータ推進組織が必要になる理由

- ▶ オープンデータの作成・公開作業は、各部署を横断する取組になる。
 - ☆オープンデータを作成・公開するに当たり、データを保有している各部署との連携・調整が必要になる。



- ▶ オープンデータの作成・公開を進めるに当たって、オープンデータを推進するための、各部署から独立した組織を設立することが望ましい。

3.2 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 → 詳細は第II部参照

- ◇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

3.3 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。

3.4 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があることも想定される。
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
 - ◇データの流通過程において、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。
 - ◇その場合、情報提供者が正当性を主張するための方法は以下の2通り。
 - ◆ オープンデータの原則： 公開し続けることにより、原典を明示する。
 - ◆ 改ざんに関する技術的対処： 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）別添」より引用して解説する。
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。

3.5 公開・運用／3.6 改善点の洗い出し

■ 公開・運用

- ▶ ある程度のデータが登録された段階で公開し、オープンデータの提供を開始。
- ▶ 公開中は、情報利用者からのフィードバックが得られるように、アンケートページや問い合わせ窓口を用意することが望ましい。

■ 改善点の洗い出し

- ▶ 一定の期間ごとに、情報利用者から得られたフィードバックや、運用上の問題を整理し、改善点を洗い出す。
- ▶ 新規のデータを公開するタイミングで、改善点を洗い出すことが望ましい。
- ▶ その後、得られた改善点を解決するための計画を立案。
- ▶ 改善点を洗い出す際に、オープンデータの技術レベル（→ 8.4節）の向上、又は利用ルールの見直しを併せて行うことを推奨。



第II部 利用ルール編： 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

■ 本章の概要

- ▶ 「オープンデータ」と言えるためには、提供する公共データを情報利用者が自由に二次利用できることが重要であり、そのためには、二次利用を認める利用ルールを採用することが必要である。本章では、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。

■ 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
 - ◇日本の各府省のホームページに適用される予定の「政府標準利用規約（第1.0版）」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試行版「DATA.GO.JP」の利用ルールについて紹介する。

4.1 オープンデータにおける利用ルール的重要性

- 国、地方公共団体、独立行政法人等が保持する公共データにも著作権が発生する。著作権の発生した公共データを利用する場合は、作成機関に許諾を得なくてはならない。
- 利用ルールで二次利用が可能であることが明示されると、自由に利用できる。
- 公共データを二次利用可能な形で公開することを実現する手段は以下の3つがあるが、それぞれメリットとデメリットがある。

①公共データには原則、著作権は発生しないものとする	<ul style="list-style-type: none"> ○ 著作権者によって差し止めや損害賠償等の権利が行使されず、自由に利用できる。米国では国等が保有する公共データには著作権はないとすることによって利活用が活発化している。 × 著作権法の改正(政府が作成したデータは著作権法の対象外とする)が必要である。
②公共データに著作権は発生するが、これを放棄する	<ul style="list-style-type: none"> ○ 国や地方公共団体等が自ら権利を放棄することで、①と同等の効果を得ることができる。 × 著作権は、国や地方公共団体等の財産の一つであり、国有財産法、財政法、地方自治法、補助金等適正化法等との関係において、権利放棄を行うことが可能かどうか、十分に検討する必要がある。
③公共データを二次利用可能なルールで公開する	<ul style="list-style-type: none"> ○ 公共データについて著作者は著作権を保持したまま自由に二次利用を認める利用ルールを採用することによって、オープンデータを進めるという考え方である。①、②に比べて短期的に対応できる。

- 最も望ましいのは①だが著作権法の法改正が必要。次に望ましいのは②だが、著作権は、国や地方公共団体等の財産の一つであり、国有財産法、財政法、地方自治法、補助金等適正化法等との関係において、権利放棄を行うことが可能かどうか、十分に検討する必要がある

⇒ 本書では、短期的に対応可能な③の方法について具体的に解説する

4.2 国際的なオープンデータの利用ルールの動向

- 諸外国ではオープンデータに関する取組が進められている。
 - ▶ 利用ルールとしては、クリエイティブ・コモンズ（CC）の表示ライセンス（CC-BY）及び、その互換ライセンスを採用している国が多く、CC-BYは事実上の国際的な標準利用ルールとなりつつある。
 - ▶ 米国は法律上パブリックドメイン、オランダは「著作権の不在宣言」（CC0）を採用して、パブリックドメインとなっている。

	採用した利用ルール	国名
既存利用ルール 採用	著作権不在の宣言（CC0）	オランダ
	表示ライセンス（CC-BY）	ドイツ、オーストラリア、 ニュージーランド 他
独自利用ルール 採用	表示ライセンス（CC-BY）互換	イギリス、フランス、イタ リア（バージョン2.0） 他
	表示-継承ライセンス（CC-BY-SA）互換	イタリア（バージョン1.0）

4.3 日本政府におけるオープンデータ利用ルールの動向

- 日本政府は「政府標準利用規約（第1.0版）」を作成
 - ▶ 2012年7月4日「電子行政オープンデータ戦略」（高度情報通信ネットワーク社会推進戦略本部決定）と、2013年6月25日「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）」（各府省情報化統括責任者（CIO）連絡会議決定）を受けて、各府省のホームページの利用規約案の検討。
 - ▶ 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では、内閣官房IT総合戦略室からの依頼を受け、「各府省ホームページの利用ルール見直しひな形（素案）」を作成し、電子行政オープンデータ実務者会議のルール・普及WG（2014年1月17日開催）に提言。
 - ▶ 電子行政オープンデータ実務者会議では、この提言をもとに議論を行い、2014年4月1日に「政府標準利用規約（第1.0版）」を了承。
 - ▶ 2014年6月19日、各府省情報化統括責任者（CIO）連絡会議で決定。
- 各府省から示された意見も踏まえ、国のできるだけ多くのコンテンツに適用できるものとして検討された結果、CC-BYとは別の利用ルールとなっている。
- 政府標準利用規約（第1.0版）は、2015年度に見直しの検討を行う。その際には、利用ルールの「政府標準利用規約（第1.0版）」への変更後のコンテンツの利用状況等を踏まえ、禁止事項の必要性の見直しも含めて検討が行われる予定。
- 政府データカタログサイト試行版「DATA.GO.JP」ではCC-BYを採用

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）、政府データカタログサイト試行版「DATA.GO.JP」利用規約について、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. 政府データカタログサイト試行版「DATA.GO.JP」利用規約

◇CC-BYの利用条件に加えて、第三者の権利が含まれている場合の対応等の注意喚起をしている利用ルールについて、概要と利用方法を紹介する。

4. 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。

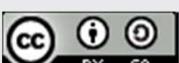
5. 政府標準利用規約（第1.0版）

◇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）について、概要と利用方法を紹介する。

5.1 CCライセンス

■ CCライセンスには以下の6種類がある。

- ▶ ①商業利用を許可するか（許可／不許可）、②改変を許可するか（許可／不許可／許可するが同一利用ルール利用）の2つの利用条件の組み合わせで生成される。

イメージ	ライセンス名称	利用の条件		
		出典表示	商業利用	改変
	表示 2.1 日本 (CC-BY 2.1 Japan)	必須 (タイトル、全ての著 作者、URLを表示)	許可	改変を許可する (※)
	表示-非営利 2.1 日本 (CC-BY-NC 2.1 Japan)	必須 (タイトル、全ての著 作者、URLを表示)	許可しない (改変されたものの商 業利用も許可しない)	改変を許可する (※)
	表示-改変禁止 2.1 日本 (CC-BY-ND 2.1 Japan)	必須 (タイトル、全ての著 作者、URLを表示)	許可	許可しない
	表示-非営利-改変禁止 2.1 日本 (CC-BY-NC-ND 2.1 Japan)	必須 (タイトル、全ての著 作者、URLを表示)	許可しない	許可しない
	表示-継承 2.1 日本 (CC-BY-SA 2.1 Japan)	必須 (タイトル、全ての著 作者、URLを表示)	許可	改変を許可するが、改変されてできた二次 的著作物は、このライセンスと同一のライセ ンスを採用すること。(※)
	表示-非営利-継承 2.1 日 本 (CC-NC-SA 2.1 Japan)	必須 (タイトル、全ての著 作者、URLを表示)	許可しない (改変されたものの商 業利用も許可しない)	改変を許可するが、改変されてできた二次 的著作物は、このライセンスと同一のライセ ンスを採用すること。(※)

※著作者の人格権を侵害する改変は許可しない

5.2 CC-BYライセンス

■ 概要

- ▶ CCライセンスの中で、最も利用の制約が少ない利用ルールで、出典を表示すれば自由に利用できる。
- ▶ 各国の法制度にあわせるために利用ルールの改訂が行われており、2014年3月時点で、国際的にはバージョン4.0が利用され始めている。日本でもバージョン4.0が利用される予定であるが、現在は翻訳が終わっていないため、バージョン2.1が利用されている。
- ▶ CC-BYは、政府の情報をオープン化する際の利用ルールとして、海外で多く利用されている。ドイツ、オーストラリア、ニュージーランド等のデータポータルサイトや、日本の政府データカタログサイト試行版、米国の省庁のホームページ等で利用されているほか、イギリス、フランス、イタリアでは、CC-BYと互換性のあるライセンスを政府のオープンデータの利用ルールとしている。

■ 特徴

- ▶ 出典を表示すれば、複製、翻案、頒布、上演、演奏、上映、公衆送信、口述、展示、録音・録画、放送、有線放送、送信可能化、伝達等などの自由な利用を許諾する。（商業的な利用も可能）。
- ▶ 出典を表示する際には、原作品の全ての著作権表示をそのままにして、原著作者・実演家のクレジットを合理的に表示し、原作品のタイトルを表示し、指定されたURIがある場合はそれを記載しなくてはならない。また、二次的著作物をつくった場合、原著作者の利用を示すクレジットを表示する必要がある。
- ▶ 許諾者からの通知があった場合、実行可能な範囲で、許諾者又は原著作者への言及を除去しなくてはならない。
- ▶ 利用が許諾されている範囲を狭めるような形でコピーコントロールを行ってはならない。

表 CC-BYライセンスのアイコン、利用状況等

項目	内容
名称	表示 2.1 日本 (CC-BY 2.1 Japan) (通称、「CC-BY」) (諸外国ではCC-BY 4.0等)
イメージ	
オープンデータで利用している国	<ul style="list-style-type: none"> • ドイツ • オーストラリア • ニュージーランド • 米国 • 日本（政府データカタログサイト試行版） • イギリス（互換ライセンス） • フランス（互換ライセンス） • イタリア（互換ライセンス） 等

【出典】 クリエイティブ・コモンズ・ジャパン ホームページ
(<http://creativecommons.jp/licenses/>) をもとに
データガバナンス委員会事務局作成

5.3 政府データカタログサイト試行版「DATA.GO.JP」利用規約

■ 概要

- ▶ 2013年12月20日に内閣官房IT総合戦略室で立ち上げた政府データカタログサイト試行版「DATA.GO.JP」においては、電子行政オープンデータ実務者会議での議論を受けて、CC-BYライセンスが採用されることとなった。
- ▶ 利用条件については、CC-BYと同じである。
- ▶ 第三者の権利が含まれている場合等について、注意喚起がなされている。
- ▶ 政府データカタログサイト試行版に特有の部分を削除した利用ルールひな形を別途掲載。

■ 特徴

- ▶ 出典を表示すれば、複製、翻案、頒布、上演、演奏、上映、公衆送信、口述、展示、録音・録画、放送、有線放送、送信可能化、伝達等などの自由な利用を許諾する。（商業的な利用も可能）。
- ▶ 出典を表示する際には、原作品の全ての著作権表示をそのままにして、原著作者・実演家のクレジットを合理的に表示し、原作品のタイトルを表示し、指定されたURIがある場合はそれを記載しなくてはならない。また、二次的著作物をつくった場合、原著物の利用を示すクレジットを表示する必要がある。
- ▶ 許諾者からの通知があった場合、実行可能な範囲で、許諾者又は原著作者への言及を除去しなくてはならない。
- ▶ 利用が許諾されている範囲を狭めるような形でコピーコントロールを行ってはならない。
- ▶ 第三者の権利が含まれている場合に注意を行うこと、対象データについて一切の保証がなされないこと、準拠法が日本法であること等について注意喚起を実施。
- ▶ 数値データ等についても出所表示を行うことについてお願いを記載。

表 政府データカタログサイト試行版「DATA.GO.JP」利用規約のアイコン、利用状況等

項目	内容
名称	政府データカタログサイト試行版「DATA.GO.JP」利用規約
イメージ	
オープンデータで利用している国	CC-BYと同様 <ul style="list-style-type: none"> • ドイツ • オーストラリア • ニュージーランド • 米国 • 日本（政府データカタログサイト試行版） • イギリス（互換ライセンス） • フランス（互換ライセンス） • イタリア（互換ライセンス） 等

【出典】クリエイティブ・コモンズ・ジャパン ホームページ（<http://creativecommons.jp/licenses/>）をもとにデータガバナンス委員会事務局作成

参考：政府データカタログサイト試行版「DATA.GO.JP」利用規約を元にした利用ルールひな形

利用規約

本サイトで公開しているデータ（以下「対象データ」といいます。）は、以下の条件の下、自由に利用できます。

第1条（国の著作権）

国が著作権を有する著作物の利用（複製、公衆送信、翻訳・変形等の翻案等）については、クリエイティブ・コモンズ・ライセンス（以下「CCライセンス」という。）の表示2.1日本（<http://creativecommons.org/licenses/by/2.1/jp/legalcode> に規定される著作権利用許諾条件を指す。）によるものとします。なお、数値データ、簡単な表・グラフ等のデータは著作権の対象ではありませんので、当該リソースのうち、これらの対象データについてはCCライセンスの適用はなく、自由に利用できます。

第2条（第三者の権利）

対象データの中に第三者が著作権その他の権利を有している場合があります。第三者が著作権を有している箇所や、第三者が著作権以外の権利（例：写真につき肖像権・パブリシティ権等）を有している対象データについては、特に権利処理済であることが明示されているものを除き、利用者の責任で、当該第三者から利用の許諾を得るものとします。なお、対象データの中の第三者が権利を有している部分の特定・明示等は、原則として行っておりませんので御注意ください。

第3条（無保証）

対象データの内容については、その正確性・網羅性、特定の目的への適合性等一切の保証をしません。対象データを利用したことにより損害が生じても責任を負いません。

第4条（他のサイトの利用規約との関係）

対象データが、本サイトとは他のサイトにおいても公開されている場合において、当該他のサイトの利用規約（法令に定める利用条件とは別に、当該サイトにおいて独自に設けられた利用条件をいいます。）と本サイトの利用規約が異なるときは、本サイトの利用規約が優先するものとします。

第5条（準拠法と合意管轄）

本サイトの利用規約は日本法に基づいて解釈されます。本サイトの利用規約及び対象データの利用に関する紛争については、当該紛争に係る対象データ又は利用規約を公開している組織の所在地を管轄する地方裁判所を、第一審の専属的な合意管轄裁判所とします。

（利用に当たってのお願いと御注意）

御利用の際には、CCライセンスの適用外の数値データ、簡単な表・グラフ等のデータについても、できれば、出所（利用する対象データを含む 1. データセットのメタデータの作成者欄に記載する組織名、2. リソースの名称及び 3. リソースのURL）の表示をお願いします。また、利用状況等を把握したいと考えているため、できれば、本サイトの管理者あてに御一報をお願いします。

利用に当たっては、関連法令を遵守してください。

5.4 CC0

■ 概要

- ▶ CC0は、著作権が生じている著作物やデータについて、自発的に権利を放棄して、パブリックドメインにしようという試みである。
- ▶ 他のCCライセンスが著作権を前提として「利用の許諾を行う」のに対して、こちらは①著作権を放棄し、②放棄できない権利は無条件かつ永続的な利用許諾を行い、そして③利用許諾も無効な場合には権利行使をしないということを「確約する」という構成になっている。
- ▶ この宣言がなされたデータは、多くの人々が様々な利用を始めることから、途中で撤回することができないことに注意する必要がある。
- ▶ 2014年6月時点では日本語版はパブリックコメント対応中であり、近日中に正式版が公開される予定である。
- ▶ 諸外国ではオランダが採用している。また、複数の図書館等の施設で採用が進んでいる。

■ 特徴

- ▶ 当該作品・データに関する著作権、著作隣接権、肖像権等の権利を放棄することを表明し、無条件かつ自由な利用を許諾する。
- ▶ 同時に、当該作品・データに関するいかなる責任も負わず、いかなる表明・保証も行わないことを宣言している。

表 CC0ライセンスのアイコン、利用状況等

項目	内容
名称	CC0 1.0 Universal (通称「CC0」)
イメージ	
オープンデータで利用している国、機関	<ul style="list-style-type: none"> • オランダ • 米国ハーバード大学図書館 • ドイツ国立図書館 • Nature Publishing Group <p style="text-align: right;">等</p>

【出典】クリエイティブ・コモンズ・ジャパン ホームページ
(<http://creativecommons.jp/licenses/>) をもとに
データガバナンス委員会事務局作成

5.5 政府標準利用規約（第1.0版）

■ 概要

- ▶ 「電子行政オープンデータ推進のためのロードマップ」（2013年6月14日 高度情報通信ネットワーク社会推進戦略本部決定）に基づき、オープンデータ流通推進コンソーシアムのデータガバナンス委員会での検討成果を踏まえ、電子行政オープンデータ実務者会議において、各府省のホームページに適用する新しいオープンデータ対応の利用ルールとして検討。
- ▶ 電子行政オープンデータ実務者会議において、有識者や各府省の意見を踏まえて、検討・作成が行われ、2014年4月1日に了承。
- ▶ 2014年6月19日に各府省情報化統括責任者（CIO）連絡会議で決定。

■ 特徴

- ▶ 基本的には、出典を記載すれば、複製、公衆送信、翻訳・変形等の翻案等、自由な利用を認める。
- ▶ 出典の記載方法については各府省が定める（出典の記載方法を例示）。
- ▶ 著作物性のないデータも適用対象としている。
- ▶ 「法令・条例・公序良俗に反する利用」と「国家・国民の安全に脅威を与える利用」を禁止している。
- ▶ 編集・加工等した情報をあたかも国（又は府省等）が作成したかのような状態で公表・利用することを禁止している。
- ▶ データの利用に制約を課す個別法令については、主なものをわかりやすい形で紹介。
- ▶ 第三者が著作権を有しているデータや、第三者が著作権以外の権利を有しているデータについては、特に権利処理済であることが明示されているものを除き、利用者の責任で、当該第三者から利用の許諾を得る必要があることをわかりやすい形で説明。（出典の記載等によって第三者が権利を有していることを直接的又は間接的に表示・示唆しているものについては、その例を記載。）
- ▶ 各府省は、別の利用ルールを設ける具体的かつ合理的な根拠をホームページ上で明確に説明した上で、特定のコンテンツに、個別法令に根拠のない利用制約を課す利用ルールを適用することができる。

表 政府標準利用規約（第1.0版）の概要

項目	内容
名称	政府標準利用規約（第1.0版）
イメージ	なし
オープンデータで利用している国	<ul style="list-style-type: none"> • 日本

【出典】 電子行政オープンデータ実務者会議資料をもとにデータガバナンス委員会事務局作成

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）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

◇オープンデータとして公開されているデータを二次利用する者又は機関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提供する者又は機関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データの二次利用を行うことが前提であることから、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。

6.1 情報利用者の視点からの比較

■ 情報利用者の視点からは、CC0、CC-BYが利用しやすい

- ▶ CC0は著作権を放棄するため、情報利用者は何の制約もなく二次利用が可能である。また、諸外国のデータとのマッシュアップも容易である。
- ▶ CC-BYは二次利用の際に出典を記載するという条件がついているため、情報利用者はその条件を守る必要がある。マッシュアップに関しては、諸外国でCC-BYを採用している例が多いことから、同じ条件で組み合わせて利用できる場合が多い。
- ▶ 政府標準利用規約（第1.0版）は、禁止事項の追加により二次利用の範囲が必ずしも明確とは言えないため萎縮効果を生む可能性がある。また、CC-BYとの互換性がないため、諸外国のデータとのマッシュアップの際には注意が必要。

	CC0	CC-BY (※)	政府標準利用規約 (第1.0版)
①情報利用者が自由に二次利用できるか	可能	出典記載により可能	出典記載に加え、禁止事項がある
②諸外国のデータ（CC-BYのものが多い）とのマッシュアップが容易か	容易	数が多くなると出典記載が多くなる	CC-BYとの相違点を理解することが必要

※政府データカタログサイト試行版「DATA.GO.JP」の利用規約に採用されている。

6.2 情報提供者の視点からの比較

■ 情報提供者の視点では政府標準利用規約（第1.0版）が親切に見えるが、実効性には課題がある。

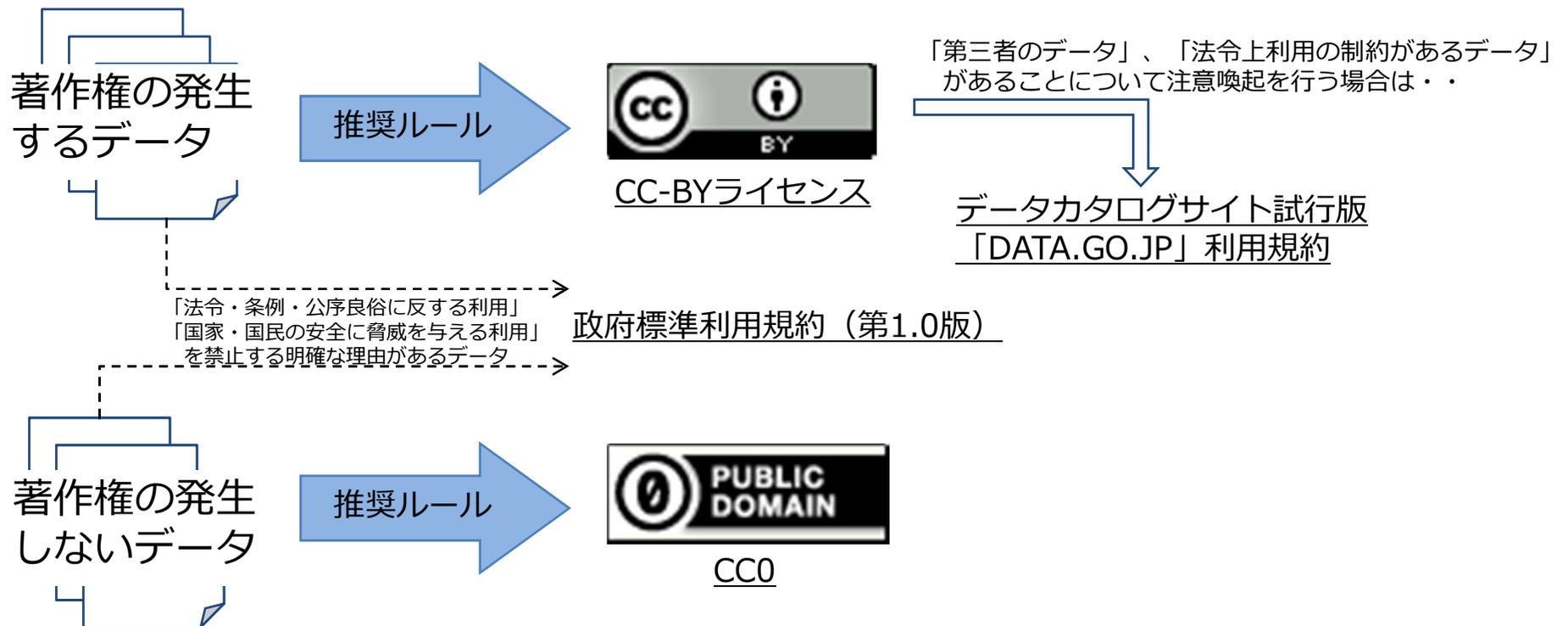
- ▶ ①提供したデータについて保証する必要が無いこと（無保証）、②情報提供者の名前を騙って改ざんしたデータが公開されるのを防ぐこと、③情報提供者が一般的に望ましくないと考える利用の態様を示すことができること、の3点が重要である。
- ▶ ②については、規定があったとしても、是正を求めた際に情報提供者が対応しない場合や、転々流通した先での改ざんの場合は是正を求めることができないなど、実効性に課題がある。
- ▶ ③については、規定があったとしても、情報提供者が禁止しているつもりであった行為が裁判所で認められない場合や、認められても国外に対しては執行が困難な場合がある等、実効性に課題がある。

	CC0	CC-BY	政府標準利用規約 (第1.0版)
①提供したデータについて保証する必要が無い（無保証）	無保証規定あり	無保証規定あり	無保証規定あり
②情報提供者の名前を騙って改ざんしたデータが公開されるのを防ぐこと	規定なし	規定あり 実効性に課題	規定あり 実効性に課題
③情報提供者が一般的に望ましくないと考える利用の態様を示すことができること	規定なし	規定なし	規定あり (「法令・条例・公序良俗に反する利用」と「国家・国民の安全に脅威を与える利用」を禁止) 実効性に課題

6.3 オープンデータにする際に望ましい利用ルール

■ 著作物が発生するデータについてはCC-BY、著作権が発生しないデータ（数値、簡単な表・グラフ等）についてはCC0を適用することが望ましい。

- ▶ 政府標準利用規約（第1.0版）は、第三者の権利が含まれているデータや、利用に法令上の制約があるデータに関する注意喚起が予め盛り込まれているが、それを理由にCC-BYをとりやめて、政府標準利用規約（第1.0版）を採用することは情報利用の際の萎縮効果による悪影響の方が大きいと考えられ、望ましくない。
- ▶ 第三者の権利が含まれているデータや、利用に法令上の制約があるデータに関する注意喚起を行う場合は、政府データカタログサイト試行版「DATA.GO.JP」の利用規約を参考にすることが望ましい。



補足：公開されたデータの悪用とその責任について

- 公開されたデータが増えれば、その悪用も当然に増加する。従って、オープンデータを促進することは、当然にデータの悪用の機会を増やすことにつながる。しかしながら、諸外国でオープンデータ政策が進められてきたのは、当然のことではあるが、公開されたデータの悪用によるデメリットを遥かに上回るメリットがオープンデータにあるからである。
- 基本的に、一旦公開されたデータについては、情報提供者がその利用をコントロールし悪用を防ぐことは不可能である。利用ルールにおいて情報提供者が望ましくないとする利用を禁止することは可能だが、仮にそのような禁止条項を設けたとしても、データを悪用しようとする者が当該利用ルールを読んだことによって悪用を思い留まるような状況は、現実的には想定し難い。
- 公開対象となるデータについては、それが個人の権利を侵害するものではないか、危険な結果（危険物の製造等）を生じるものではないか等の合理的なスクリーニングがなされるべきであり、それがなされないことによってリスクが現実化したのであれば、情報提供者は批判を受けることもあるであろう。しかしながら、そのような事情もないのに、事実上の因果関係のみから「悪用されるようなデータを公開した情報提供者に責任がある」という評価をすることは適切ではない。オープンデータにおいては、情報提供者は、自身の利益のためにデータを公開するのではなく、広く情報利用者一般の利益のためにデータを公開するのであるから、それによって責められるのであれば、情報提供者としては防御的な行動を取るだろう。データ悪用のリスクを低減する最も（そしておそらくは唯一の）効果的な方法は、データを公開しないことであるから、このような責任評価に接すれば、情報提供者は当然にデータの公開を抑制することとなる。
- 以上のとおり、データが悪用された場合の責任の評価については、オープンデータの趣旨を踏まえた合理的・非情緒的な判断が強く期待されるところである。

第7章 利用ルールに関する今後の見直しの方向性

- 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、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や、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）を適用することが考えられる。
 - ▶ 特に、「法令・条例・公序良俗に反する利用」と「国家・国民の安全に脅威を与える利用」という規定は、その具体的な利用態様が必ずしも明確ではなく、公開されたデータの利用に際して萎縮効果を生む可能性もあり、データガバナンス委員会としては、政府標準利用規約（第1.0版）の運用状況を注視し、これらの禁止条項を削除しても問題がないと判断できる場合には、当該禁止条項の削除又はCC-BYやCC0への移行を視野に入れて検討すべきと考える。
 - ▶ 政府標準利用規約（第1.0版）は、利用ルールの政府標準利用規約（第1.0版）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）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）のいずれにおいても無保証で公開しているものであり、国、地方公共団体等が責任を負うものではない



第III部 技術編：
機械判読に適したデータにし
よう

第8章 オープンデータの技術レベル

■ 本章の概要

- ▶ 第3章に記したオープンデータの作成・公開手順のうち、技術的な事項に関して解説する。

■ 本章の構成

1. 機械判読性に関する解説

◇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

◇オープンデータに識別子が必要である理由を解説する。

4. オープンデータの技術レベル

◇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。

8.1 機械判読性に関する解説

■ 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。

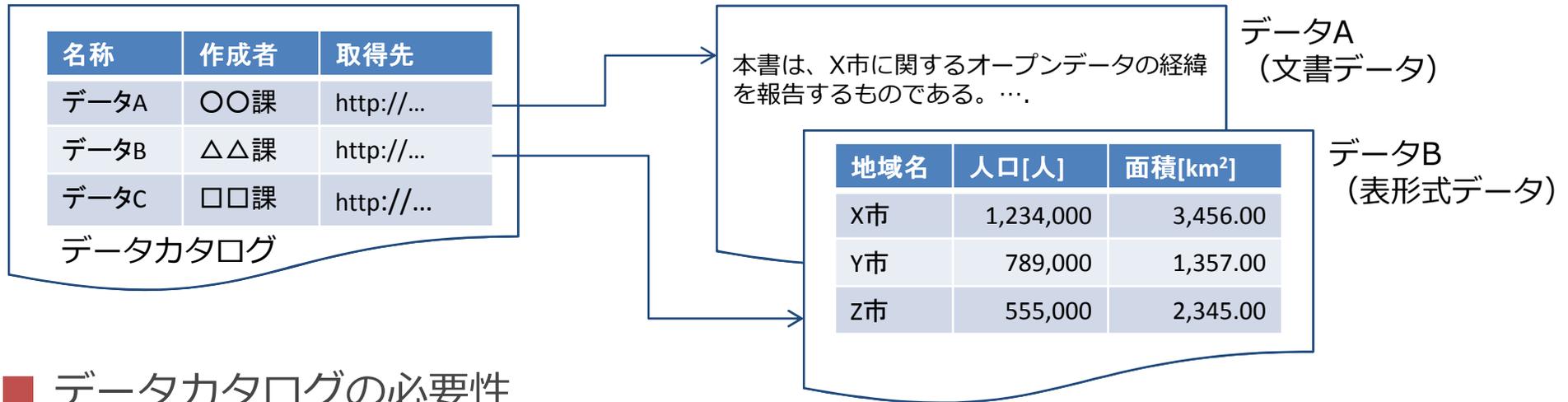
★4~: この形式のデータは、相互に接続でき、コンピュータによるデータのマッシュアップが容易になる。



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

8.2 データカタログに関する解説

- 「データ」と「データカタログ」の関係は下図のとおり。



- 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、あるいはRDF・SPARQL等の技術を利用したデータ・メタデータ検索機能を提供することが望ましい。

8.3 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。

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

- 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。

8.5 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

■ オープンデータの登録ポリシーには2通りある

1. 集中登録方式

◇システム管理者や、オープンデータにすることをを行う独立した組織が、各組織・部署からデータを集めて公開する手法。

◇この方式では、データを集める際にメタデータもまとめて収集することが望ましい。

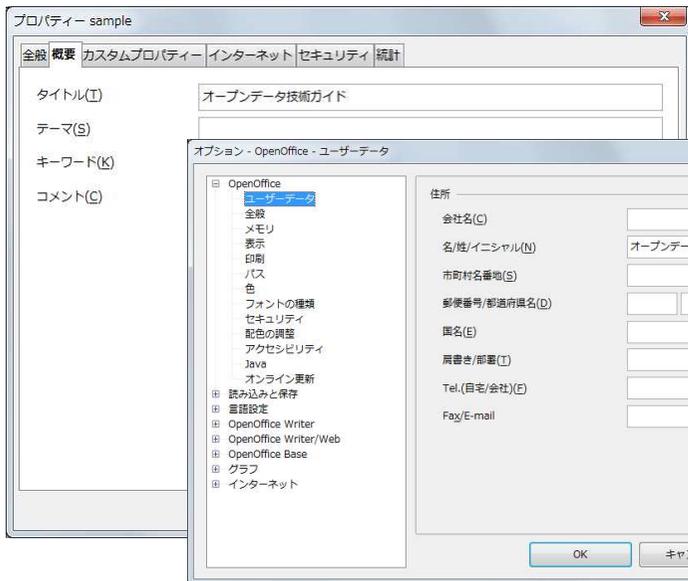
2. 分散登録方式

◇各組織・部局が自ら、何らかのシステムを利用して直接オープンデータを登録・管理する手法。

◇この方式では、担当組織がオープンデータを作成する際に、メタデータもまとめて作成することが望ましい。

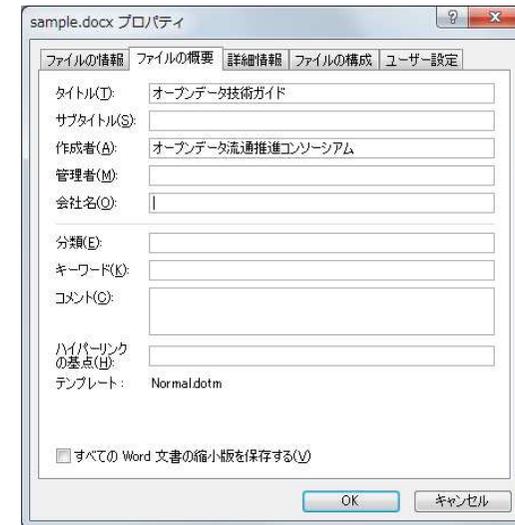
8.5 オープンデータの管理ポリシーとメタデータの付与方法

■ Apache Tika(*)を利用して、ファイルのメタデータを自動収集する例



Open Office 4 Writerの
文書プロパティ画面

Microsoft Word 2010の文書プロパティ画面



この文書形式データを
Apache Tikaで解析



この文書形式データを
Apache Tikaで解析



作成者→
タイトル→
作成日時→
最終更新日時→

```
{
  ...
  "dc:creator": "オープンデータ流通推進コンソーシアム",
  "dc:title": "オープンデータ技術ガイド",
  "dcterms:created": "2014-01-27Txx:xx:xx",
  "dcterms:modified": "2014-01-27Txx:xx:xx",
  ...
}
```

(*) <http://tika.apache.org>

第9章 オープンデータのための技術的指針

■ 本章の概要

- ▶ 機械判読に適したオープンデータを作成・編集するための技術的な指針を、識別子、ファイル形式及びデータの3項目に関して示す。

■ 本章の構成

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と、それを満たすための手法を解説する。

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して示す。

3. データに関する指針

◇表形式データ・文書形式データ・地理空間情報・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示す。

9.1 識別子に関する指針

■ オープンデータにとっての識別子が満たすべき性質

1. ユニークであること。
2. 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

1. 対象とする実物や組織・場所に番号が付与されていない場合は、まずそれらに番号を付与する。
2. 識別子のユニーク範囲を拡大する。
 - ◇ ucodeやDoIなど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取得し、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できる。
 - ◆ ただし、組織の統廃合等によりドメイン名が変わると、識別子も変わってしまうことに注意。

9.2 ファイル形式に関する指針

■ 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)	

9.3 データに関する指針

■ 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。

9.3 データに関する指針／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。

9.3 データに関する指針／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。

9.3 データに関する指針／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。

9.3 データに関する指針／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。



付録

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・地理空間情報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に関する規格を挙げる。

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

第11章（付録） データカタログシステムCKAN

■ 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



OPEN DATA
オープンデータ流通推進コンソーシアム
