



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

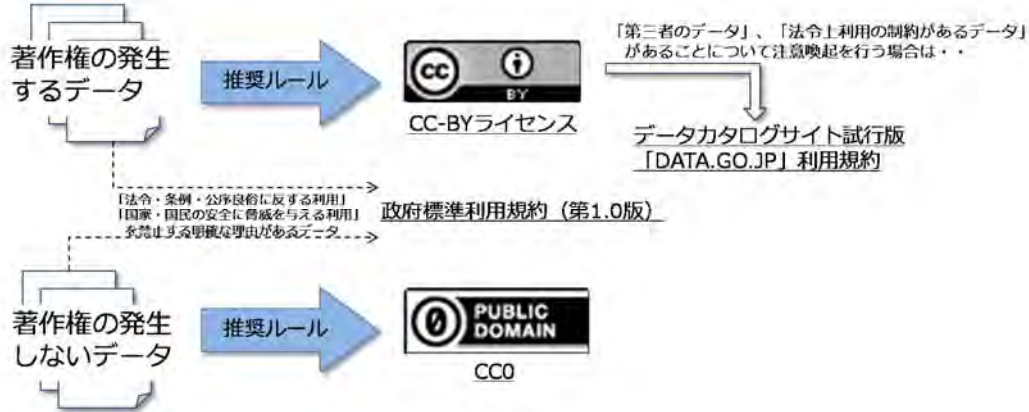
◇データカタログシステムであるCKANの概要とその使用方法を解説する。



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