

---

# オープンデータ化ガイド

## 第1版(案)

---

2014年〇月〇日

オープンデータ流通推進コンソーシアム

The background of the lower half of the page is decorated with various colorful geometric shapes, including rectangles and parallelograms in shades of blue, green, and yellow, scattered across the white space.



# 目次

<b>第 I 部 Getting Started: オープンデータ化をはじめよう</b> .....	5
<b>第 1 章 はじめに</b> .....	6
1.1 本書の目的 .....	6
1.2 本書の対象読者 .....	6
1.3 本書の構成 .....	8
1.4 用語定義 .....	10
<b>第 2 章 オープンデータの動向と意義</b> .....	12
2.1 オープンデータに関する主な動向 .....	12
2.2 オープンデータの意義 .....	17
2.3 本書におけるオープンデータの定義 .....	18
<b>第 3 章 オープンデータ化の手順</b> .....	19
3.1 オープンデータ化推進組織の設立 .....	19
3.2 現状把握 .....	19
3.3 計画立案 .....	20
3.4 公開作業 .....	21
3.5 公開・運用 .....	21
3.6 改善点の洗い出し .....	21
<b>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</b> .....	23
<b>第 4 章 オープンデータで必要となる利用ルール</b> .....	24
4.1 オープンデータにおける利用ルールの重要性 .....	24
4.2 国際的なオープンデータの利用ルールの動向 .....	26
4.3 日本政府におけるオープンデータ利用ルールの検討状況 .....	28
<b>第 5 章 オープンデータ利用ルールの概要</b> .....	34
5.1 CC ライセンス .....	34
5.2 CC-BY ライセンス .....	37
5.3 CC0 .....	44
5.4 政府標準利用規約（第 1.0 版） .....	50
「政府標準利用規約（第 1.0 版）」（仮称）（案）の解説（案） .....	53
<全体の構成、基本的考え方について> .....	53
<各項目について> .....	54
<b>第 6 章 オープンデータ利用ルールの比較</b> .....	59
6.1 情報利用者の視点からの比較 .....	59
6.2 情報提供者の視点からの比較 .....	60
6.3 データの性質による比較 .....	62
6.4 オープンデータ化の際に望ましい利用ルール .....	66

<b>第 7 章</b>	<b>利用ルールの適用</b> .....	67
7.1	オープンデータ化の主な対象 .....	67
7.2	利用ルールの適用 .....	69
参考	個別データに利用ルールを適用するまでの手順 .....	71
<b>第 8 章</b>	<b>利用ルールに関する今後の検討について</b> .....	77
8.1	今後の方向性について .....	77
<b>第 III 部</b>	<b>技術編：機械判読に適したオープンデータにしよう</b> .....	79
<b>第 9 章</b>	<b>オープンデータ化の技術レベル</b> .....	80
9.1	機械判読に関する解説 .....	80
9.2	データカタログに関する解説 .....	81
9.3	オープンデータと識別子 .....	82
9.4	オープンデータ化の技術レベル .....	82
9.5	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 .....	84
<b>第 10 章</b>	<b>オープンデータ化のための技術的指針</b> .....	85
10.1	データの公開方式に関する指針 .....	85
10.2	識別子に関する指針 .....	87
10.3	ファイル形式に関する指針 .....	88
10.4	データに関する指針 .....	89
<b>第 11 章</b>	<b>(付録)オープンデータに関する規格・ツール</b> .....	106
11.1	データフォーマットに関する規格 .....	106
11.2	識別子に関する規格 .....	114
11.3	オープンデータ化に有用なツール .....	119
<b>第 12 章</b>	<b>(付録)CKAN 解説</b> .....	122
12.1	CKAN とは .....	122
12.2	CKAN の利用方法 .....	126
<b>参考文献</b> .....		136
<b>オープンデータ化ガイド 改正履歴</b> .....		138

## 第 I 部 Getting Started: オープンデータ化をはじめよう

---

# 第1章 はじめに

## 1.1 本書の目的

2012年7月4日に高度情報通信ネットワーク社会推進戦略本部で決定された「電子行政オープンデータ戦略」を契機として、国、地方公共団体、独立行政法人、公益企業等におけるオープンデータへの取り組みが活発になっている。国、地方公共団体、独立行政法人、公益企業等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れば、様々な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ものである。

本書は、国、地方公共団体、独立行政法人、公益企業等が、自身が保有している公共データをオープンデータ化する際の際の参考となるよう、オープンデータ化に当たっての留意事項等について、「利用ルール」と「技術」の2つの観点から、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）においてまとめたものである。

## 1.2 本書の対象読者

本書が対象とする読者は、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である。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、公益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。

オープンデータは、情報提供者が作成・公開し、情報利用者からアクセスされ利用される（図 1-1）。第I部と第II部では、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を対象としている。一方、第III部については、このうち機械判読性の高いデータを作成・整形しようとする人を対象としている。

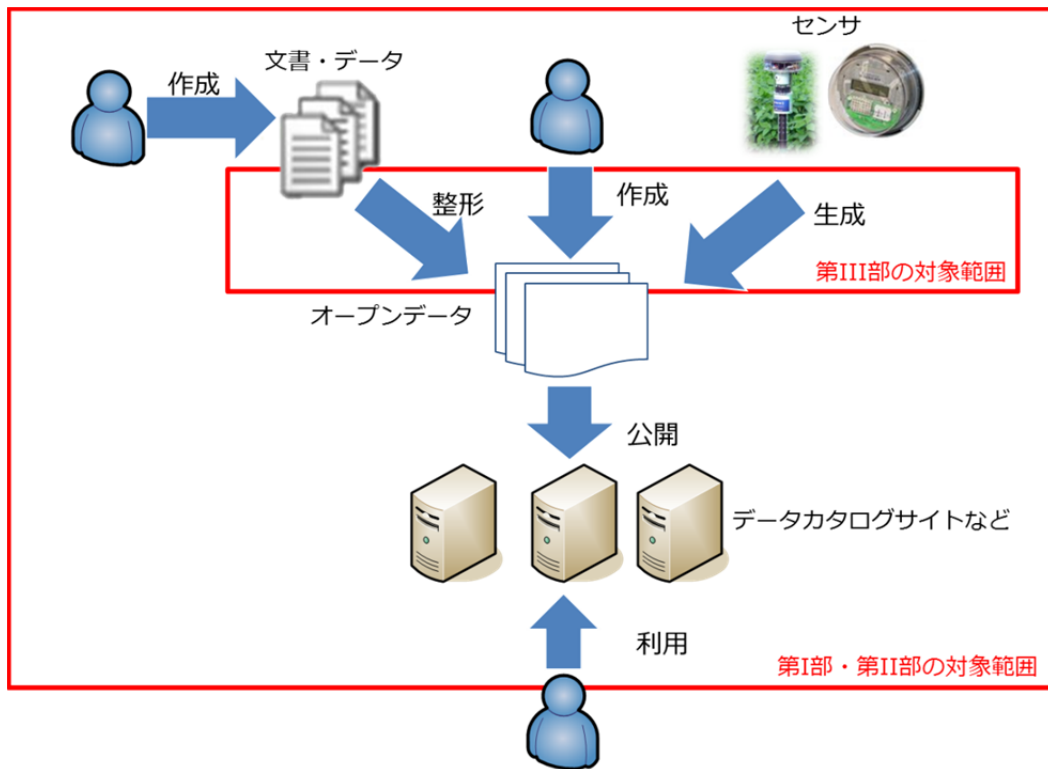


図 1-1 オープンデータの公開・利用までの流れ

### 1.3 本書の構成

本書の構成は、以下の表の通りである。

表 1-1 本書の構成

構成	内容
第 I 部 Getting Started : オープンデータ化をはじめよう	オープンデータの利用ルール・技術の解説に先立ち、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、オープンデータ化の意義について解説する。
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	オープンデータに関する利用ルールについて解説する。 第 4 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。 第 5 章では、諸外国政府で採用が進んでいる CC-BY 又は 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 章では、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、データの性質という 3 つの観点から比較を行う。 第 7 章では、上記を受けて具体的に利用ルールを利用する対象ごとに推奨される利用ルールについて解説する。 第 8 章では、政府標準利用規約（第 1.0 版）の今後の見直しにあたっての視点について紹介する。
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	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留意事項について解説する。 第 9 章では、第 3 章に記したオープンデータ化の手順のうち、技術的な事項に関して解説する。 第 10 章では、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、公開方式、識別子、ファイル形式、データの 4 項目について示す。 第 11 章は付録であり、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 12 章も付録であり、データカタログシステムの 1 つである CKAN とその使用方法について解説する。



また、表 1-2 に、読者が知りたいと考えられる内容ごとに記載箇所を示しているので、本書を読む際に活用いただきたい。

表 1-2 内容別対象章

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化するまでの手順を知りたい。	第 3 章
4. オープンデータ化の際には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 7 章
8. 利用ルールについて、政府における今後の検討の方向性について知りたい。	第 8 章
9. オープンデータ化のために、どのような技術レベルを目指すべきか知りたい。	第 9 章
10. 表形式データ、文書形式データ、地理空間データ等、様々なデータをオープンデータ化したいが、それらの作成に際して技術的に留意すべき事項を知りたい。	第 10 章
11. オープンデータ化に際して有用なファイル形式やデータ伝送プロトコル、識別子にどのようなものがあるか知りたい。	第 11 章
12. Web サービス、GIS ツール等オープンデータ化に有用なツールにどのようなものがあるか知りたい。	第 11 章
13. データカタログシステムの 1 つである CKAN と、その利用方法について知りたい。	第 12 章

## 1.4 用語定義

本書が使用する主な用語の定義は、以下の表の通りである。

表 1-3 主な用語の定義

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。著作権のある情報も無い情報も含む。
公共データ	国、地方公共団体、独立行政法人、公益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。詳しくは、第 2.3 節を参照。
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・改変等を行い、新たなデータを作成すること。また、情報提供者の提供したデータの単なる複製や再配布についても含む。
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。
機械判読	コンピュータプログラムが自動的にデータを再利用（加工・編集等）すること。
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ことを示す情報。
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、一部図や表などを含み、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。

用語	定義
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。
地理空間データ	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。

## 第2章 オープンデータの動向と意義

本章では、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、オープンデータ化の意義について解説する。

### 2.1 オープンデータに関する主な動向

#### 2.1.1 日本政府におけるオープンデータに関する動向

2012年7月4日に高度情報通信ネットワーク社会推進戦略本部で決定された「電子行政オープンデータ戦略」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り組みが急速に進んでいる（表 2-1）。

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、オープンデータが重要な政策の1つとして取り上げられている。

表 2-1 オープンデータに関する政府等の主な動向

年月	名称・URL	位置づけ
2012.07.04	電子行政オープンデータ戦略 <a href="http://www.kantei.go.jp/jp/singi/it2/denshigyousei.html">http://www.kantei.go.jp/jp/singi/it2/denshigyousei.html</a>	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 <a href="http://www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000047.html">http://www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000047.html</a>	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（情報流通連携基盤共通 API の開発等） <a href="http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/opendata03.html">http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/opendata03.html</a>	総務省
2013.01.18	「Open DATA METI」（β版）公開 <a href="http://datameti.go.jp/">http://datameti.go.jp/</a>	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 <a href="http://www.kantei.go.jp/jp/singi/it2/densi/">http://www.kantei.go.jp/jp/singi/it2/densi/</a>	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 <a href="http://www.soumu.go.jp/johotsusintokei/open.html">http://www.soumu.go.jp/johotsusintokei/open.html</a>	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） <a href="http://www.soumu.go.jp/menu_news/s-news/01toukei01_02000024.html">http://www.soumu.go.jp/menu_news/s-news/01toukei01_02000024.html</a>	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） <a href="http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/saikou_jpn.pdf">http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/saikou_jpn.pdf</a>	閣議決定

年月	名称・URL	位置づけ
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) <a href="http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf">http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf</a>	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 <a href="http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou3.pdf">http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou3.pdf</a>	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 (原文) <a href="https://www.gov.uk/government/publications/open-data-charter">https://www.gov.uk/government/publications/open-data-charter</a> (邦訳) <a href="http://www.mofa.go.jp/mofaj/gaiko/page23_000044.html">http://www.mofa.go.jp/mofaj/gaiko/page23_000044.html</a>	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) <a href="http://www.kantei.go.jp/jp/singi/it2/cio/dai52/kihon.pdf">http://www.kantei.go.jp/jp/singi/it2/cio/dai52/kihon.pdf</a>	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 <a href="http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan_jp.pdf">http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan_jp.pdf</a>	各府省情報化統括責任者 (CIO) 連絡会議決定
2013.12.20	データカタログサイト試行版「DATA.GO.JP」公開 <a href="http://data.go.jp/">http://data.go.jp/</a>	内閣官房
2014.04.●	府省等ホームページの標準利用規約 <a href="http://www.kantei.go.jp/jp/singi/it2/densi/dai6/gijisidai.html">http://www.kantei.go.jp/jp/singi/it2/densi/dai6/gijisidai.html</a>	各府省情報化統括責任者 (CIO) 連絡会議決定

## 2.1.2 地方公共団体におけるオープンデータの取組

地方公共団体においては、2012年7月の「電子行政オープンデータ戦略」の決定前から、一部で先行的な取り組みが行われており、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。専用のデータポータル等によるオープンデータの公開を行っている例が多いが、ホームページ全体をオープンデータ化したり（福井市）、県内市町村でデータ形式等を統一したりする取り組み（福井県）を行っている例もある（表 2-2）。

表 2-2 地方公共団体における主なオープンデータの取り組み

地方公共団体名	取組名称 (URL)	概要
福井県 鯖江市	データシティ鯖江 <a href="http://www.city.sabae.fukui.jp/pageview.html?id=12765">http://www.city.sabae.fukui.jp/pageview.html?id=12765</a>	地方公共団体のオープンデータ化では先駆的な取り組み。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。

地方公共 団体名	取組名称 (URL)	概要
千葉県 流山市	流山市オープンデータトライアル <a href="http://www.city.nagareyama.chiba.jp/10763/">http://www.city.nagareyama.chiba.jp/10763/</a> 流山市議会オープンデータトライアル <a href="http://www.nagareyamagikai.jp/opendata/">http://www.nagareyamagikai.jp/opendata/</a>	ウェブサイト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化に取り組み。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 <a href="http://data.yokohamaopendata.jp/">http://data.yokohamaopendata.jp/</a>	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化。
静岡県	ふじのくにオープンデータカタログ <a href="http://www.pref.shizuoka.jp/kikaku/ki-330/opendata/">http://www.pref.shizuoka.jp/kikaku/ki-330/opendata/</a>	都道府県で始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 <a href="http://fugaku3776.okfn.jp/">http://fugaku3776.okfn.jp/</a>	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化される。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 <a href="http://www.pref.fukui.lg.jp/doc/toukei-jouhou/opendata/">http://www.pref.fukui.lg.jp/doc/toukei-jouhou/opendata/</a> 県内公共データの形式統一 <a href="http://www.fukuishimbun.co.jp/localnews/politics/46384.html">http://www.fukuishimbun.co.jp/localnews/politics/46384.html</a>	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り組みにも着手。
福井市	市のホームページ利用規約の改訂 <a href="http://www.city.fukui.lg.jp/sisei/kohou/hp/site-p.html">http://www.city.fukui.lg.jp/sisei/kohou/hp/site-p.html</a>	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 <a href="http://amcp-aomori.jp/">http://amcp-aomori.jp/</a>	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化
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 <a href="http://www.city.aizuwakamatsu.fukushima.jp/docs/2009122400048/">http://www.city.aizuwakamatsu.fukushima.jp/docs/2009122400048/</a>	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。

### 2.1.3 国際的な動向

諸外国においては、欧米を中心に 2000 年代後半から取り組みが開始されている。

米国では、2009年5月21日には、連邦政府のデータポータルサイト「Data.gov」が開設され、同年12月9日にオバマ大統領が発表した「透明性とオープンガバメントに関する覚書」では、オープンガバメントの3原則として、透明性、市民参加、官民連携を掲げた。2012年5月23日に発表した「21世紀のデジタル政府構築に関する覚書」では、連邦政府が保有するデータは原則オープンにすることとし、あらゆるデバイスでいつでもアクセスできることや、全てのデータにAPIを用意すること、主な政府機関のホームページには開発者向けページを用意すること等が示された。

一方、EUでは、2003年にEUが発表した「PSI（公共保有データ）の再利用に関する指令」を契機として各国のオープンデータへの取り組みが始まった。中でも英国政府が積極的に取り組んでおり、2005年7月には「PSI再利用に関する規則」を定め、2006年10月にはOffice of PSIがThe National Archives傘下に入り、オープンデータに関する推進体制が確立された。2009年には政府のデータポータルサイト「data.gov.uk」を開設、2010年5月にはキャメロン首相による「透明性アジェンダ」が発表され、「再利用可能かつ機械判読な形でのデータの公開」、「営利利用も可能とする同一のオープンライセンスでの公開」、「単一のオンラインアクセスポイントでデータ入手可能であること」の3原則が掲げられた。また、2010年9月には、オープンガバメント・ライセンスが制定された。そのほか、フランス、ドイツ、イタリア等、多くの欧州の国々でオープンデータへの取り組みが進められている。

このような欧州各国の動きを受けて、欧州委員会は2011年に「欧州オープンデータ戦略」を発表し、オープンデータの経済効果を見据えて、EUとしてデータポータルを立ち上げる等の計画を発表している。

2013年6月8日には、G8サミットにおいてG8首脳による「オープンデータ憲章」が合意された。この憲章では、5つのオープンデータの原則を示し（表2-4）、2013年10月末までにG8各国は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ている。日本もこれに従って、2013年10月29日に「日本のオープンデータ憲章アクションプラン」を公開している<sup>1</sup>。

---

<sup>1</sup> [http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan\\_jp.pdf](http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan_jp.pdf)

表 2-3 諸外国におけるオープンデータの主な動向

時期	実施事項	国名
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

表 2-4 オープンデータの原則

<p>1. Open Data by Default (原則としてのオープンデータ)</p> <ul style="list-style-type: none"> <li>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。</li> </ul> <p>2. Quality and Quantity (質と量)</p> <ul style="list-style-type: none"> <li>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。</li> <li>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。</li> <li>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。</li> <li>可能な限り早急に公表する。</li> </ul> <p>3. Usable by All (すべての人々が利用できる)</p> <ul style="list-style-type: none"> <li>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。</li> <li>可能な限り多くのデータを公表する。</li> </ul> <p>4. Releasing Data for Improved Governance (ガバナンス改善のためのデータの公表)</p> <ul style="list-style-type: none"> <li>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。</li> <li>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。</li> </ul> <p>5. Releasing Data for Innovation (イノベーションのためのデータを公表)</p> <ul style="list-style-type: none"> <li>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。</li> <li>将来世代のデータイノベーターの能力を強化する。</li> </ul>
--

出典：G8 サミット「オープンデータ憲章」<sup>2</sup>

<sup>2</sup> <http://www.kantei.go.jp/jp/singi/it2/densi/dai4/sankou8.pdf>



## 2.2 オープンデータの意義

以上のように、日本国内及び諸外国においてオープンデータが進められているが、オープンデータの意義としては、「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）において、「透明性・信頼性の向上」、「国民参加・官民共同の推進」、「経済の活性化・行政の効率化」の3点が挙げられる<sup>3</sup>。

また、これを受けて、「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）」（2013年6月25日 各府省情報化統括責任者（CIO）連絡会議決定）においては、以下の通り整理されている<sup>4</sup>。

### ア 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）

### イ 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）

### ウ 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）

出典：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）  
（2013年6月25日 各府省情報化統括責任者（CIO）連絡会議決定）

<sup>3</sup> [http://www.kantei.go.jp/jp/singi/it2/pdf/120704\\_siryu2.pdf](http://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryu2.pdf)

<sup>4</sup> <http://www.kantei.go.jp/jp/singi/it2/densi/>

## 2.3 本書におけるオープンデータの定義

前述の「電子行政オープンデータ戦略」では、我が国における公共データの活用の取組に当たり、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、という4つの基本原則を掲げている。

これを受けて作成された「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）では、オープンデータについて、「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」するとしている。

以上を踏まえ、本書においては、「電子行政オープンデータ推進のためのロードマップ」に沿って、「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 オープンデータ化の手順

本章では、オープンデータ化の手順について示す。

オープンデータ化は、以下に示すような手順で進めることが望ましい。なお、スモール・スタートの原則から、オープンデータ化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のではなく、例えば、個別の部局において、比較的オープンデータ化しやすいデータから着手するということがひとつの方法である。

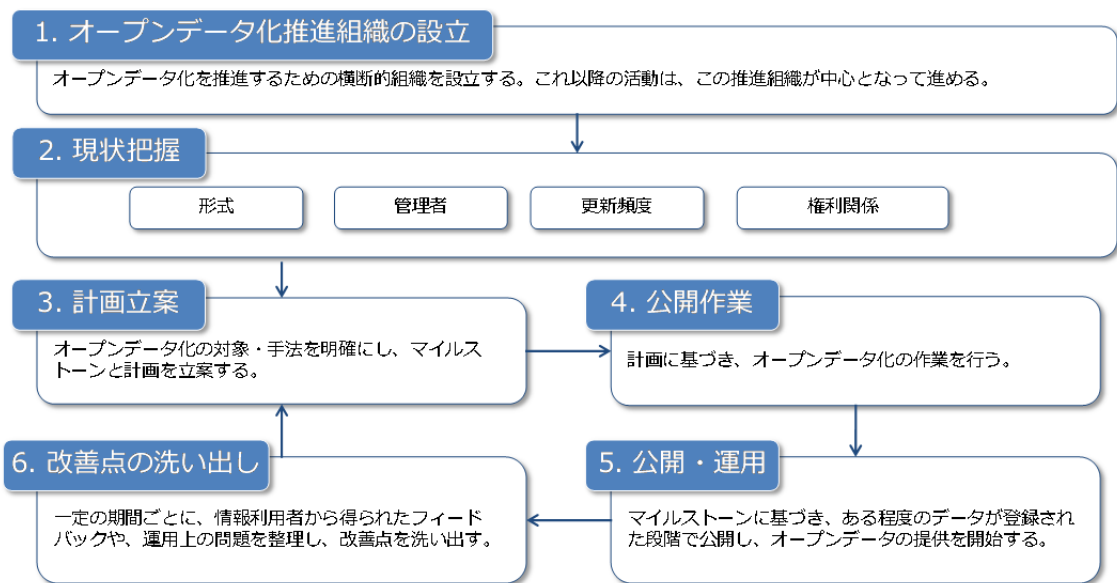


図 3-1 推奨されるオープンデータ化の手順

以下、それぞれのステップでの実施・検討内容について解説する。

### 3.1 オープンデータ化推進組織の設立

オープンデータ化は、各部署を横断する取組になる。オープンデータ化を進めるにあたり、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。

このために、オープンデータ化を進めるにあたって、オープンデータ化推進のための、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

現状把握として、各部署が管理しているデータを、以下の観点でまとめることが重要で

ある。(データの棚卸しのような形で実施することが望ましい。)

- データの管理担当部署
- データの種類 (予算・各種報告・統計・広報等)
- データの分量

現状把握にあたり注目すべき箇所を、以下に挙げる。

#### 1. データの形式

それぞれのデータの形式を確認する。

- 紙 (同一情報の電子データがあるか要確認)
  - ▶ 電子データがない資料を公開するにはスキャンする必要がある。
- 電子データ (ファイル形式を含む)

#### 2. データの管理者

データを管理する各部署の情報管理体制を確認する。

- 管理者が設定されているか。
- 管理者が統一されているか。

#### 3. データの更新頻度

データがどのくらいの頻度で更新されるか、確認する。

- 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

それぞれのデータについて、例えば下記を確認する。詳細は第 II 部を参照されたい。

- 第三者が著作権等の権利を有するデータ
- 法令上の制約 等

### 3.3 計画立案

計画立案のステップでは、現状把握又はフィードバックに基づき、オープンデータ化の作業対象・手法を明確にする。その際、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。

また、大きな組織であるほど、計画立案が重要である。

計画立案のステップで留意すべき項目を以下に挙げる。

#### 1. データ形式・システムの準備計画

- 9.4 節を参考に、どのレベルの「データ」と「データカタログ」を準備するか、方針を策定する。
- また、10.1.1 節、10.2 節を参照し、必要なメタデータや識別子体系を検討する。

2. 運用ルールの方策定
  - データを管理している組織からのデータの入手手順・頻度を明確にする。
  - 適宜更新される場合は、更新手法をルール化しておく必要がある。
  - 9.5 節を参照し、データの登録ポリシーを決定する。
3. 利用ルールの設定
  - オープンデータの利用ルールを設定する。
  - その設定にあたって、第三者権利問題や法令上の制約がある場合は、その解決を図る。
4. スモール・スタートの原則
  - 作業は段階的に行い、完了したも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。
  - 「電子行政オープンデータ戦略」においても、「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」と記載されており、いわゆるスモール・スタートの原則が採用されている。

### 3.4 公開作業

立案した計画に基づき、調達をかけるなどして、必要なツールを揃えオープンデータ化の作業を実施する。

データカタログシステム（CKAN 等）を利用する場合は、定められた運用ルールに基づき、対象のデータをデータカタログシステムに登録する。

データを公開する際には、その利用ルールを明確にする必要がある。

### 3.5 公開・運用

マイルストーンに基づき、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する。

運用中は、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。

### 3.6 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。

得られた改善点を解決するための計画を立案する。  
また、この際に、9.4 節に記すデータの技術レベルを上げること、又は利用ルールを見直すことの検討が推奨される。

## 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Ⅱ部 利用ルール編の資料本体は省略しています





### 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9章 オープンデータ化の技術レベル

本章では、第3章に記したオープンデータ化の手順のうち、技術的な事項に関して解説する。

まず、機械判読、データカタログ、識別子に関して解説する。続いて、これらを元に「オープンデータ化の技術レベル」を規定する。また、オープンデータの管理ポリシーとメタデータに関して補足する。

なお、オープンデータ化に関する技術的な指針の詳細は、第10章にまとめる。

### 9.1 機械判読に関する解説

5★Open Data は、Tim Berners-Lee が提唱したオープンデータの評価指標であり、以下の5段階からなる<sup>33</sup> (図 9-1)。これを元に機械判読のレベルを解説する。

1. *make your stuff available on the Web (whatever format) under an open license*  
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. *make it available as structured data (e.g., Excel instead of image scan of a table)*  
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. *use non-proprietary formats (e.g., CSV instead of Excel)*  
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. *use URIs to denote things, so that people can point at your stuff*  
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. *link your data to other data to provide context*  
他のデータにリンクしている (Linked Open Data)

コンピュータが★1の形式のファイルから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の形式のファイルは構造化されているため、特別な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の形式のデータの解析方法は、公開されている。このため、★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。さらに★4以上になると、データ間を相互に

<sup>33</sup> 5 ★ Open Data. (原文) <http://5stardata.info/> (邦訳) <http://5stardata.info/ja>

接続することができ、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。

特に★3 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、10.4 節で詳しく述べる。

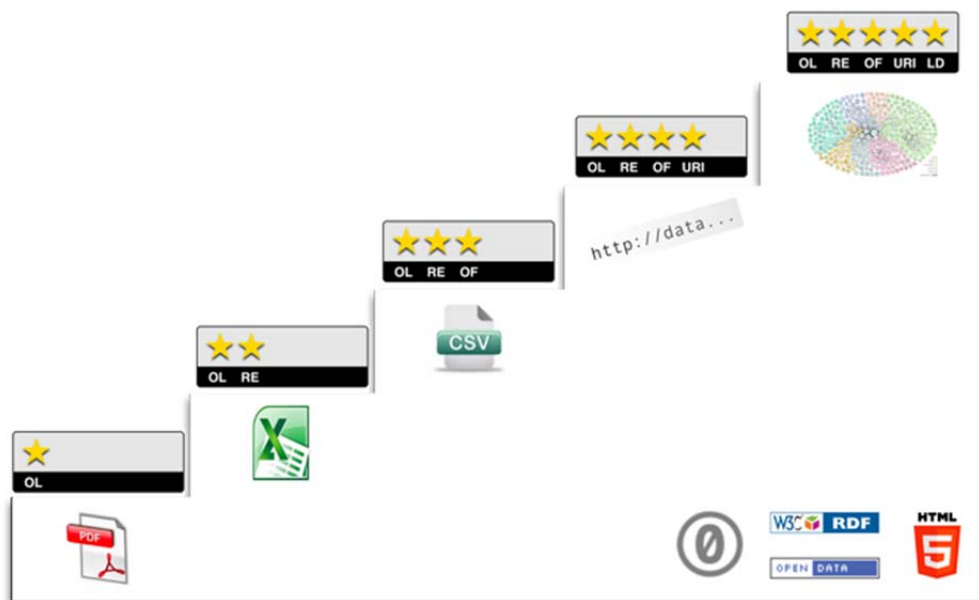


図 9-1 5★Open Data の指標

## 9.2 データカタログに関する解説

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。図 9-2 は、「データ」と「データカタログ」の関係を示したものである。

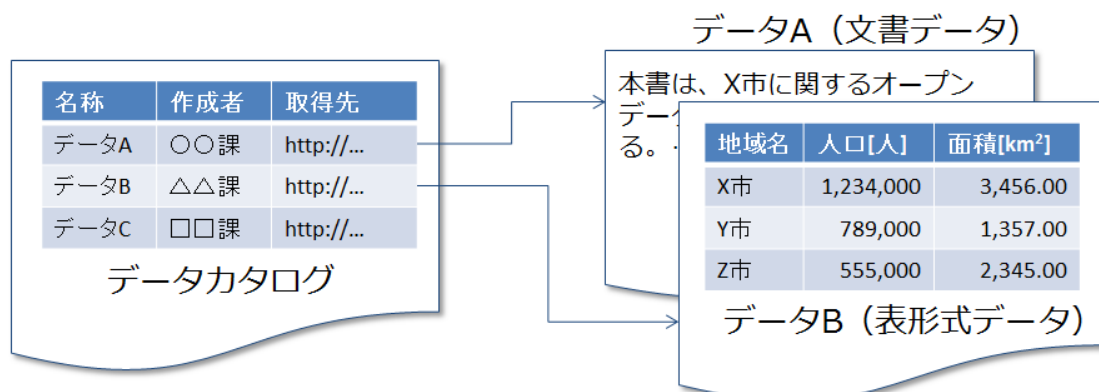


図 9-2 「データ」と「データカタログ」の関係

図 9-2 に例示したような、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。

### 9.3 オープンデータと識別子

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。また、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、次のような問題が発生する可能性があるためである。第一に、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。第二に、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。

これらの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が、識別子 (ID) である。

オープンデータの識別子が満たすべき性質については、10.2 節で述べる。

### 9.4 オープンデータ化の技術レベル

これまで述べてきた事項を、オープンデータ化の技術レベルとしてまとめると、表 9-1 のようになる。

技術レベルの適用を検討する際に、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。例えば、Level 1 のデータカタログ形式を利用して Level 2 のデータを提供してもよい。

Level 1 に掲げた技術を利用することにより、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。また、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。

Level 2 に移行すると、データに対する機械判読性が向上する。

続いて、Level 3 に移行すると、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。

さらに、Level 4 に移行すると、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。

表 9-1 オープンデータ化の技術レベル

	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 10 章参照)	RDF <sup>34</sup> 、XML <sup>35</sup> 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 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 <sup>36</sup> 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 <sup>37</sup> 等	Web サーバ+ CKAN+ 情報流通連携基盤等

<sup>34</sup> RDF (Resource Description Framework) とは、Web 上にある「リソース」(言及対象の事物)に関する情報を記述するための枠組みである。RDF は主語 (Subject)、述語 (Predicate)、目的語 (Object) の 3 つの要素でリソースに関する情報を記述する。

Fabien Gandon, Guus Schreiber, and Dave Beckett, RDF 1.1 XML Syntax. February 25, 2014. W3C Recommendation. <http://www.w3.org/TR/rdf-syntax-grammar/>

<sup>35</sup> XML (Extensible Markup Language) とは、データおよびそれを処理するコンピュータプログラムの振る舞いについて記述するための言語仕様、およびそれに基づいて書かれた文書である。

Tim Bray, et al. Extensible Markup Language (XML) 1.1 (Second Edition). August 16, 2006. W3C Recommendation. <http://www.w3.org/TR/xml11/>

<sup>36</sup> SPARQL (SPARQL Protocol and RDF Query Language) とは、RDF モデルに基づき記述されたデータを検索・操作するクエリ言語である。

Lee Feigenbaum, et al. SPARQL 1.1 Protocol. [Online] May 21, 2013. W3C Recommendation. <http://www.w3.org/TR/sparql11-protocol/>

<sup>37</sup> CKAN とは、Open Knowledge Foundation が無償で提供している、web ベースのデータ管理・配信システムである。オープンデータを配信する多くの政府系組織で利用されている。

<http://www.ckan.org/>

## 9.5 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。本節では、そのような手法について解説する。

メタデータの付与方法は、データの登録ポリシーによって変わる。例えば、システム管理者や、オープンデータ化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（これを「集中登録方式」と呼ぶ。）をとる場合は、システム管理者や独立した組織が各組織・部署から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。一方、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（これを「分散登録方式」と呼ぶ。）をとる場合は、担当の組織・部局が何らかの方法でメタデータを登録することが求められる。この場合は、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。

表形式データや文書形式データを編集するソフトウェアによっては、これらにメタデータを付与する方法を提供しているものがある。これらを利用すれば、ファイルの作成者や作成日時等のメタデータを、ファイル作成時に格納できる。例えば、Microsoft Office や OpenOffice、Acrobat 等のソフトウェアには、ファイルの「プロパティ」を編集する機能がある。これを利用して登録したメタデータを、Apache Tika<sup>38</sup>（無償）等のソフトウェアを利用してコンピュータが取得できる（図 9-3）。なお、メタデータをデータカタログシステムに効率的に登録するツールについては、技術委員会において今後検討予定である。

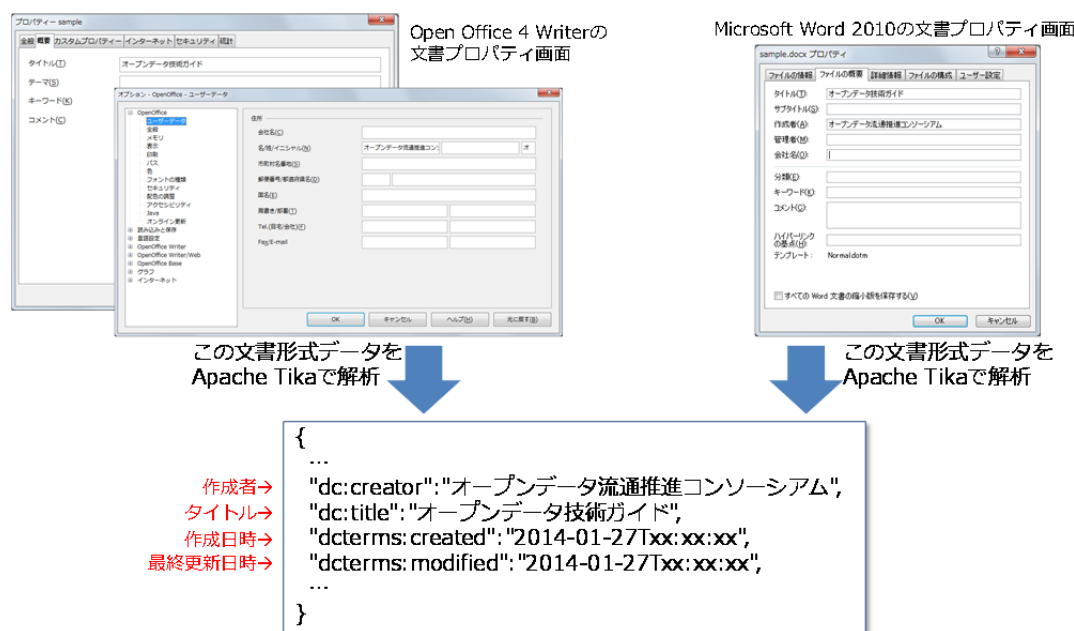


図 9-3 文書編集ソフトウェアからのメタデータの登録方法

<sup>38</sup> <http://tika.apache.org/>

## 第10章 オープンデータ化のための技術的指針

本章では、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、データの公開方式、識別子、ファイル形式及びデータの4項目に関して示す。

また、本書で扱うデータ形式は、以下の4種類である。

- 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

### 10.1 データの公開方式に関する指針

#### 10.1.1 公開時に明確にすべき情報

データを公開する際には、以下の3つを明らかにすべきである。

1. メタデータ: どんなデータか?
2. 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?
3. 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?

以下、それぞれについて解説する。

##### 10.1.1.1 メタデータ

例えば、政府データカタログサイト試行版「DATA.GO.JP」では、以下のようなメタデータが掲載されている。

- タイトル
- 組織名
- 公表者（部局）
- 作成者
- 更新頻度
- タグ
- リリース日
- URL
- ファイルサイズ
- 最終更新日
- 使用言語
- 補足

#### 10.1.1.2 アクセス方法

アクセス方法とは、そのデータを取得するための手法である。例えば、Web 上のアドレス（URL）やデータを取得するための API<sup>39</sup>を明記する。

用途によって最適なデータ形式が異なる場合は、複数の形式でデータを取得できることが望ましい。例えば、10.1.2 節に述べる場合がこれにあたる。

#### 10.1.1.3 利用ルール

例えば、利用ルールとして、以下のような事項がある。詳細は、第 II 部を参照されたい。

- 二次利用できるか。
  - データの二次利用が許可されていなければ、対象のデータをスマートフォンのアプリケーションに取り込んで提供できない。
- 商用利用できるか。
- 法令上の利用制約があるか。
- 二次利用にあたって、出典記載等の条件があるか。

#### 10.1.2 機械判読に適したデータの扱いに関する留意点

本書が示す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ことに留意すべきである。このため、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である。

このことは、「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）<sup>40</sup>にも、以下のように記載されている。

本ガイドライン策定後、各府省が新たに作成し、インターネットを通じて公開する数値（表）、文章、地理空間情報については、人間が読む、印刷することを念頭に置いた従来のデータ形式（代表的なものとして *pdf*）のほか、別添の留意事項に示す事項を踏まえて作成した（構造が整った）データを、機械判読に適した、特定のアプリケーションに依存しないデータ形式でも公開することに努めるものとする。

#### 10.1.3 データ公開時の留意事項

データを公開する際には、公開したデータにどれくらいのアクセスが予想されるか、検

---

<sup>39</sup> Application Programming Interface の略で、管理するデータ等を、外部の他のプログラムから呼び出して利用するための手順やデータ形式などを定めた規約をいう。

<sup>40</sup> <http://www.kantei.go.jp/jp/singi/it2/densi/>



討する必要がある。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った場合、公開したデータに対するアクセス障害が発生する。

本事項についての詳細は割愛する。データを公開する際に、公開サービスを運用する業者や部局と、事前に協議しておくことが望ましい。

## 10.2 識別子に関する指針

オープンデータを識別する識別子（ID）が満たすべき性質と、それを満たすための手法について解説する。

### 10.2.1 オープンデータにとっての識別子が満たすべき性質

オープンデータを利用するコンピュータは、複数の組織・団体が公開するオープンデータを横断的に利用して、処理を行う場合がある、つまり、マッシュアップされることが想定される。また、これらのオープンデータ同士がリンクによって相互に接続される可能性がある。

このため、オープンデータを識別する識別子は、グローバルにユニークであるべきである。例えば、社員番号は、その社内ではユニークであるが、社外では唯一性を保証できない。

また、以下のような性質を有していることが望ましい。

1. 既に確立している、グローバルな識別子体系を利用することが望ましい。
  - このような識別子体系を、付録 11.2 にまとめている。例えば、`ucode`・DoI (Digital Object Identifiers) ・企業コード (ISO 6523) ・地方公共団体コード等がある。
2. URI (Uniform Resource Identifier) として表現できる体系が望ましい。
  - Web にアクセスするときに利用する URL (Uniform Resource Locator) は、URI の一部である。

### 10.2.2 適切な識別子体系がない場合の対処法

10.2.1 節に記した条件を満たす、適切な識別子体系がある場合は、そのような識別子を採用すべきである。

適切な識別子体系がない場合は、以下のように対処することが望ましい。

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. 付与した番号をグローバル化する。その手法は以下の通りである。
  - `ucode` や DoI 等、グローバルな体系に基づく識別子を取得し、その体系に基づき識別子を管理する (図 10-1)。
  - 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

(図 10-2)。ただし、この方法を採用した場合、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が必要である。

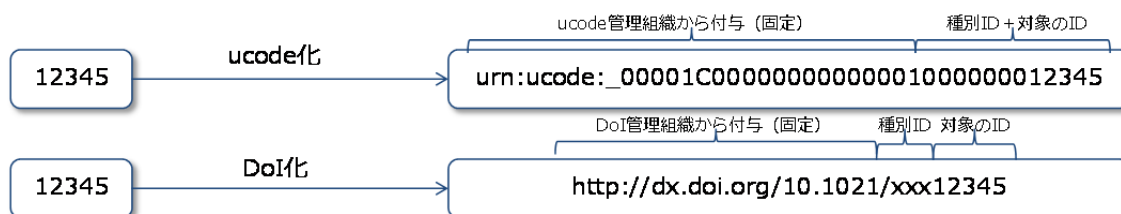


図 10-1 グローバルな識別子を利用した番号のグローバル化手法

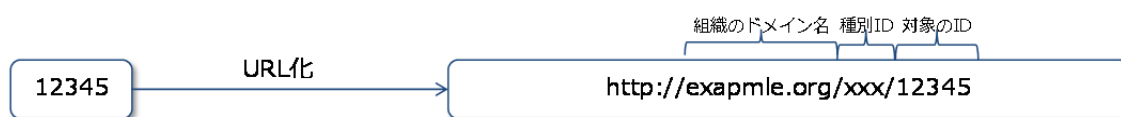


図 10-2 組織が持つ URL を利用した番号のグローバル化手法

### 10.3 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。代表的なファイル形式を、表 9-1 に示したデータの技術レベルに基づいてまとめると、表 10-1 のようになる。

表 10-1 機械判読可読性の観点で整理したファイル形式

	Level 1	Level 2 / Level 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	PDF (Acrobat 形式) doc (Microsoft Word 形式)	HTML, XML, docx (Office Open XML), odt (OpenDocument)	
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)		

## 10.4 データに関する指針

### 10.4.1 指針のグレード

データに関する指針として、以下の2つのグレードを設ける。

#### 1. 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。

#### 2. 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。

### 10.4.2 表形式データに関する指針

#### 10.4.2.1 用語の定義

まず、以下の用語を定義する（図 10-3）。

- フィールド (field)
  - 表の1行1列からなる要素。表計算ソフトの「セル」に相当。
- レコード (record)
  - 表の1行からなる要素。1個以上のフィールドからなる。
- ヘッダ (header)
  - 表の各列の名前を保持する行。1個以上のフィールドからなる。
- ファイル (file)
  - 表全体を指す。レコードとヘッダからなる。

月	A市	B市	C市	D市
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

Diagram labels: ヘッダ (Header) points to the top row; レコード (Record) points to a row; フィールド (Field) points to a cell; ファイル (File) points to the entire table.

図 10-3 「表形式データに関する指針」で利用する用語の定義

#### 10.4.2.2 グレード1の指針

(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。

#### 【解説】

図 10-4 に示すファイルは、複数の表を含んでいる。このようなファイルをコンピュータが判読するためには、表の切れ目を扱う必要があり、判読手順が複雑になる。このため、1つのファイルは、1種類の表からのみ構成されるべきである。

ファイルに含まれる複数の表を分割し、それぞれ別のファイルに格納すれば、本指針を満たす(図 10-5)。

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

図 10-4 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

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

図 10-5 図 10-4 を 2 つのファイルに分割 (指針 1 を満たす)

**【補足】**

本指針は、CSV (Comma-Separated Values) の書式を定めた RFC4180<sup>41</sup>に準拠するためのものである。RFC4180 が規定する規定 3、規定 4 は以下の通りである。

3. *There maybe an optional header line appearing as the first line of the file with the same format as normal record lines. This header will contain names corresponding to the fields in the file and should contain the same number of fields as the records in the rest of the file.*

ファイルの先頭に、各フィールドの名称を示す、1 行からなるヘッダを置いてよい。ただし、ヘッダのフィールド数は、他のレコードのフィールド数と一致しているべきである。

4. *Within the header and each record, there may be one or more fields, separated by commas. Each line should contain the same number of fields throughout the file.*

ヘッダと各レコードは、コンマで区切られた 1 以上のフィールドを含む。フィールド数は、ファイルを通して一致しているべきである。

<sup>41</sup> Shafranovich, Y. Common Format and MIME Type for Comma-Separated Values (CSV) Files. October 2005. RFC 4180. <http://www.ietf.org/rfc/rfc4180.txt>.

規定 3 と規定 4 の両方を満たすには、1 ファイルが 1 つの表から構成されていなければならない。

(指針 2) ヘッダは、1 行から構成されるべきである。

**【解説】**

図 10-6 に示すファイルのヘッダは、2 行からなっている。このようなファイルをコンピュータが判読するためには、ヘッダとデータの切れ目を解釈する必要があり、判読手順が複雑になる。このため、ヘッダを 1 行で構成するべきである。

ヘッダの内容を統合して 1 行にまとめれば、本指針を満たす (図 10-7)。

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

図 10-6 ヘッダが複数行からなる (指針 2 を満たさない) 例

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

図 10-7 ヘッダを 1 行に統合 (指針 2 を満たす)

**【補足】**

本指針は、指針 1 と同様、RFC4180 の規定 3 に準拠するためのものである。

### 10.4.2.3 グレード2の指針

(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。

#### 【解説】

図 10-8 に示すファイルの C 市 1 月の値には、「1.6」という数値と「(\*1)」という注釈へのリンクが含まれている。ここで、注釈へのリンクである(\*1)と、その先にある自然言語で書かれた注釈文は、コンピュータが判読できない。機械判読性を高めるためには、注釈へのリンクである「(\*1)」を除き、数値「1.6」のみとすることが望ましい(図 10-9)。

なお、図 10-8 のような、注釈を含むファイルは、人がデータを解釈するためには必要である。このため、10.1.2 節の指針に基づき、このようなファイルは、図 10-9 のような機械判読性の高いファイルとは別に提供することが望ましい。

月	A 市	B 市	C 市	D 町
1	-4.5	-0.5	1.6(*1)	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

図 10-8 データでない情報がレコードに含まれている(指針3を満たさない)例

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

図 10-9 データでない情報を除去する(指針3を満たす)



(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。

**【解説】**

図 10-10 に示すファイルの「年」の各フィールドが結合されている。人が見ればこの 4 ヶ月のデータが 2013 年のものであることはわかるが、コンピュータはそれを判読できない。機械判読性を高めるためには、フィールドの結合を解除し、それぞれ値を記載することが望ましい (図 10-11)。

年	月	A 市	B 市	C 市	D 町
2013	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

図 10-10 フィールドが結合されている (指針 4 を満たさない) 例

年	月	A 市	B 市	C 市	D 町
2013	1	-4.5	-0.5	1.6	11.3
2013	2	-6.8	-2.1	0.4	8.4
2013	3	-2.4	1.9	3.8	13.5
2013	4	0.2	3.4	6.5	17.3

図 10-11 フィールドの結合を解除する (指針 4 を満たす)

(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。

**【解説】**

図 10-12 に示すファイルでは、「年」フィールドの第 2 行目以降の記述が省略されている。人が見ればこの 4 ヶ月のデータが 2013 年のものであることはわかるが、コンピュータはそれを判読できない。機械判読性を高めるためには、省略されている値を補完することが望ましい(図 10-13)。

年	月	A 市	B 市	C 市	D 町
2013	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

図 10-12 フィールドの値が省略されている(指針 5 を満たさない)例

年	月	A 市	B 市	C 市	D 町
2013	1	-4.5	-0.5	1.6	11.3
2013	2	-6.8	-2.1	0.4	8.4
2013	3	-2.4	1.9	3.8	13.5
2013	4	0.2	3.4	6.5	17.3

図 10-13 省略されている値を補完する(指針 5 を満たす)

(指針 6) 年の値には、西暦表記を備えることが望ましい。

【解説】

図 10-14 に示すファイルの「年」の各フィールドは、和暦で記載されている。コンピュータは、数値の大小で年を比較できる方が処理しやすいため、年の値が単調増加する西暦の方が扱いやすい。このため、西暦値を追記することを推奨する (図 10-15)。

年	月	A 市	B 市	C 市	D 町
平成 25	1	-4.5	-0.5	1.6	11.3
平成 25	2	-6.8	-2.1	0.4	8.4
平成 25	3	-2.4	1.9	3.8	13.5
平成 25	4	0.2	3.4	6.5	17.3

図 10-14 和暦で年が記載されている (指針 6 を満たさない) 例

年	年	月	A 市	B 市	C 市	D 町
平成 25	2013	1	-4.5	-0.5	1.6	11.3
平成 25	2013	2	-6.8	-2.1	0.4	8.4
平成 25	2013	3	-2.4	1.9	3.8	13.5
平成 25	2013	4	0.2	3.4	6.5	17.3

図 10-15 西暦を付加する (指針 6 を満たす)

(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。

**【解説】**

図 10-16 に示すファイルには、値の単位が記載されていない。データの単位(物理単位、貨幣単位等)は、データ処理に必須であるので、単位が明記されていることが望ましい。

ヘッダに単位を付記する(図 10-17)か、Simple Data Format<sup>42</sup>という規格に基づいたこのファイルの説明を別途付記する(図 10-18)ことにより、フィールドの単位を明記できる。

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

図 10-16 フィールドの単位が記載されていない(指針 7 を満たさない) 例

月	A 市[°C]	B 市[°C]	C 市[°C]	D 町[°C]
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

図 10-17 ヘッダに単位を付記(指針 7 を満たす)

<sup>42</sup> Simple Data Format については、付録 11.1.3 にて解説する。

```

{
  "name": "各地域の気温", } データセット名 "各地域の気温"
  "resources": [
    {
      "path": "data.csv", } データファイルのパス情報 "data.csv"
      "schema": {
        "fields": [
          {
            "name": "年",
            "type": "integer"
          },
          {
            "name": "月",
            "type": "integer"
          },
          {
            "name": "A市",
            "type": "number",
            "unit": "deg_c"
          },
          ...
        ]
      }
    }
  ]
}

```

カラム定義  
 第1カラム: 「年」という名前の整数情報。  
 第2カラム: 「月」という名前の整数情報。  
 第3カラム: 「A市」という名前の数値情報。単位は「℃」

図 10-18 Simple Data Format 規格に基づき図 10-16 の定義ファイルを追加  
 (指針 7 を満たす)

(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。

**【解説】**

日本語を記述する文字コードには、JIS (ISO-2022-JP)、Shift-JIS、EUC、UTF-8 等、複数ある。このため、記述されている文字コードが明記されていなければ、コンピュータが読み取るとは難しい。さらに、データの国際的な展開や他の規格との整合を考慮すると、UTF-8 を利用して記載することが望ましい。

現在広く利用されている Microsoft Excel の日本語版は、Shift-JIS で CSV 形式のデータを出力する。これを UTF-8 に変換する代表的な方法を、以下に示す。

1. メモ帳で CSV データを開き、UTF-8 形式で保存する。
2. OpenOffice.org<sup>43</sup>で CSV データを開き、UTF-8 形式で保存する。
3. コマンドラインのツールを利用する (nkf<sup>44</sup>等)。

(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。

**【解説】**

メタデータがフォーマルに記述されていれば、コンピュータがデータを効率的に検索、判読できるようになる。

指針 7 で提示した Simple Data Format を利用するか、データカタログシステムにメタデータを登録することにより、本指針を満たせる。

(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。

**【解説】**

データ本体を、XML、RDF 等、セマンティクスを記述できる形式を利用してフォーマルに記述することにより、各フィールドの意味を含めてデータを記述でき、機械判読性がさらに高まる。

<sup>43</sup> <http://www.openoffice.org/>

<sup>44</sup> <http://sourceforge.jp/projects/nkf/>

### 10.4.3

#### 10.4.4 文書データに関する指針

##### 10.4.4.1 グレード1 指針

文書データに関する、グレード1 指針はない。

##### 10.4.4.2 グレード2 指針

(指針1) 文章に存在する部・章・節・図表等の構造が、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。

#### 【解説】

文章は、部、章、節、段落、図表等の構造を持っている。コンピュータがこれらの構造を抽出できるようにするために、文書編集ソフトを利用する場合、文章に存在する部・章・節・図表等の構造を、フォントや文字飾りで表現するのではなく、編集ソフトが提供するスタイル機能（見出し等）を利用して表現する。また、HTML で表記する場合は、スタイル表記だけでなく、<div>や<h3>等のタグを利用した構造を示す。

(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。

#### 【解説】

文章に含まれる空白、改行が有意であるか否かを、コンピュータは判断できない。文書の解析や読み上げを行う際に、これらの空白、改行が支障となる。このため、コンピュータが判読する必要のない空白や改行は、事前に除く。

(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。

#### 【解説】

文章が図表を含む場合は、グレード1 以上の表形式データのファイルをリンク先として文章中等で示すことにより、本指針を満たせる。

(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。

**【解説】**

メタデータがフォーマルに記述されていれば、コンピュータがデータを効率的に検索、判読できるようになる。

データカタログシステムにメタデータを登録することにより、本指針を満たせる。



## 10.4.5 地理空間データに関する指針

### 10.4.5.1 グレード1 指針

(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。

#### 【解説】

地理情報を表記するための測地系は、複数存在し、それぞれ値が違う。

例えば、国際地球基準座標系 (ITRF) による緯度・経度と日本測地系による緯度・経度では、東京付近の地表面において 400m 程度ずれる。従って、地理情報が準拠している測地系が明記されていないと、位置を特定できない。

地理情報システム (GIS システム) を利用することにより、地理空間データを簡単に編集でき、また、出力されるデータには測地系が明記されることが多い。

### 10.4.5.2 グレード2 指針

(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。

#### 【解説】

地図データを公開する際には、ラスタ形式と比較して、同一の情報を表すのに必要な容量の小さくなるベクタ形式や、GML 形式が望ましい。また、公開にあたり、準拠している座標参照系 (世界測地系等) を表記することで、データ利用の際の座標変換が容易になる。

JPGIS は、ISO の地理情報に関する専門委員会 (ISO/TC 211) で策定された国際規格を基にした国内実用標準であり、異なるシステム間で地理空間情報データを相互利用する際の互換性の確保を主な目的に、データの設計、品質、記述方法等のルールを定めたもので、GIS 関係省庁連絡会議では政府の技術的標準と位置づけられている。JPGIS 及び JMP に基づいて地理空間データ及びメタデータを整備・提供することで、データを相互利用しやすい環境が整備され、異なる整備主体で整備されたデータの共用、システム依存性の低下、重複投資の排除等の効果を期待することができる。

(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。

**【解説】**

メタデータがフォーマルに記述されていれば、コンピュータがデータを効率的に検索、判読できるようになる。

データカタログシステムにメタデータを登録することにより、本指針を満たせる。

## 10.4.6 リアルタイムデータに関する指針

### 10.4.6.1 グレード1 指針

(指針1) データの取得仕様が明記されているべきである。

#### 【解説】

リアルタイムデータの性質や要求されるリアルタイム性は、データを取得する機器や提供するシステムに依存する。ただし、それらデータをコンピュータが取得し、解釈するためには、データの取得方法やデータの表記仕様が明確になっている必要がある。

(指針2) 表形式データや地理空間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。

#### 【解説】

リアルタイムデータは通常コンピュータが取得・解釈するものである。従って、それが表形式データや地理空間データのファイル形式であるならば、少なくともそれぞれのグレード1指針を満たすべきである。

### 10.4.6.2 グレード2 指針

(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。

#### 【解説】

データ送受信のリアルタイム性を確保するためには、送受信時間が短いほどよく、そのためには、送受信データ量が少ないほど望ましい。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ることにより、送受信データ量を軽減できる。

このような手法を可能にするためには、Streams API を利用する、「情報流通連携基盤システム外部仕様書」に基づくシステムを利用してデータを提供する、リアルタイムでRDF データを提供する、等の方法がある。

## 第11章（付録）オープンデータに関する規格・ツール

本章では、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。

### 11.1 データフォーマットに関する規格

#### 11.1.1 ファイル形式に関する規格

表形式データ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、表 11-1 に示す。

表 11-1 代表的な表形式データの規格

規格名	策定・公開者	規格番号	概要	本形式を扱えるソフトウェア
★2				
Microsoft Office Binary (.xls)	Microsoft		Microsoft Excel 2003 までのファイル形式。2008 年 6 月に Microsoft Open Specification Promise の下で仕様が公開された。	Microsoft Excel OpenOffice 等
★3				
Office Open XML (.xlsx)	Microsoft	ISO/IEC 29500	XML をベースとしたオフィススイート用のファイルフォーマットの 1 つ。Microsoft Excel 2007 以降の標準のファイルフォーマットである。	Microsoft Excel OpenOffice(*) 等
OpenDocument (.ods)	構造化情報標準促進協会 (OASIS)	ISO/IEC 26300	XML をベースとしたオフィススイート用のファイルフォーマットの 1 つ。	Microsoft Excel OpenOffice 等
CSV (Comma-Separated Values) (.csv)	Internet Engineering Task Force (IETF)	RFC 4180	いくつかのフィールド (項目) をカンマ「,」で区切ったテキストデータ及びテキストファイル。長らく公式な仕様が存在しなかったが、2005 年 10 月に RFC 4180 として規格化された。	Microsoft Excel OpenOffice 等

文書形式データ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、表 11-2 に示す。

表 11-2 代表的な文書形式データの規格

規格名	策定・公開者	規格番号	概要	本形式を扱えるソフトウェア
★2				
Microsoft Office Binary (.doc)	Microsoft		Microsoft Excel 2003 までのファイル形式。2008 年 6 月に Microsoft Open Specification Promise の下で仕様が公開された。	Microsoft Word OpenOffice 等
Rich Text Format (.rtf)	Microsoft		プレーンテキストに装飾やレイアウトのための制御用の文字列を付加した形式である。 フォントの指定や、文字の色・大きさや太文字等の装飾指定、画像の表示や中央揃え・箇条書き、表等の簡易レイアウトを行える特徴がある。	Microsoft Word OpenOffice 等
Portable Document Format (.pdf)	Adobe Systems	ISO 32000-1	アドビシステムズが開発及び提唱する、電子文書に関するフォーマット。 特定の環境に左右されずに全ての環境でほぼ同様の状態で文章や画像等を閲覧できる特性を持っている。	Acrobat 等

規格名	策定・公開者	規格番号	概要	本形式を扱えるソフトウェア
★3				
Office Open XML (.docx)	Microsoft	ISO/IEC 29500	XMLをベースとしたオフィススイート用のファイルフォーマットの1つ。Microsoft Excel 2007以降の標準のファイルフォーマットである。	Microsoft Word OpenOffice(*) 等
OpenDocument (.odt)	構造化情報標準促進協会 (OASIS)	ISO/IEC 26300	XMLをベースとしたオフィススイート用のファイルフォーマットの1つ。	Microsoft Word OpenOffice 等
HTML (HyperText Markup Language)	World Wide Web Consortium (W3C)	ISO/IEC 15445	ウェブ上の文書を記述するためのマークアップ言語の1つ。ハイパーテキストを利用して、相互間の文書や図表等を参照できる。	Microsoft Word OpenOffice 等
XML (Extensible Markup Language)	World Wide Web Consortium (W3C)		個別の目的に応じたマークアップ言語群を創るために汎用的に使うことができる仕様。	
XHTML (Extensible HyperText Markup Language)	World Wide Web Consortium (W3C)		HTMLをXMLの文法で定義し直したマークアップ言語。	

地理情報形式データに関する代表的な規格を、表 11-3 に示す。これらの規格のデータを扱えるツールについては、11.3.3 節に列記する。

表 11-3 代表的な地理空間データの規格

規格名	策定・公開者	規格番号	概要
★2			
Shapefile	ESRI		他の地理情報システム(GIS)間でのデータの相互運用におけるオープン標準として用いられるファイル形式である。
★3			
GML (Geography Markup Language)	Open Geospatial Consortium (OGC)	ISO 19136	地理的特徴を表現する XML ベースのマークアップ言語。平成 20 年 4 月から国土地理院が提供している基盤地図情報は、この形式で提供されている。
KML	Open Geospatial Consortium (OGC)		アプリケーション・プログラムにおける三次元地理空間情報の表示を管理するために開発された、XML ベースのマークアップ言語。座標の前提となる測地基準系の定義をサポートしていない。
GeoJSON	なし (世界各地の開発者達が開発し管理)		JSON (JavaScript Object Notation)を用いて空間データをエンコードし、非空間属性を関連付けるファイル形式。属性にはポイント (住所や座標)、ライン (各種道路や境界線)、ポリゴン (国や地域) 等が含まれる。



### 11.1.2 データの伝送プロトコル・形式に関する規格

代表的なデータ伝送プロトコルや伝送形式に関する規格を、表 11-4 に示す。

表 11-4 代表的なデータ伝送プロトコルや伝送形式に関する規格

規格名	策定・公開者	規格番号	概要
FTP (File Transfer Protocol)	Internet Engineering Task Force (IETF)	RFC 959	端末とサーバの間でファイル（ドキュメントや画像・動画等）を転送するための、代表的なプロトコル。
HTTP (HyperText Transfer Protocol)	Internet Engineering Task Force (IETF)	RFC 2616	Web ブラウザと Web サーバの間で HTML 等のコンテンツの送受信に用いられる通信プロトコルである。
REST			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。
SOAP	World Wide Web Consortium (W3C)		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。交換メッセージは XML に準拠している。
SPARQL	World Wide Web Consortium (W3C)		RDF モデルに基づくデータを検索・操作するクエリ言語。
JSON (JavaScript Object Notation)	Internet Engineering Task Force (IETF)	RFC 4627	JavaScript におけるオブジェクトの表記法をベースとした軽量なデータ記述言語

また、リアルタイムデータの伝送に関する規格としては、以下のようなものがある。

- Streams API<sup>45</sup>
  - サーバ・クライアント間での HTTP コネクションを継続し、値が更新されるごとにその結果を返す仕組みである。
  - World Wide Web Consortium (W3C) が規格化している。
  - Twitter 等で利用されている。
- GTFS (General Transit Feed Spec) Realtime<sup>46</sup>
  - GTFS は、公共交通機関の時刻表とその地理的情報に使用される共通形式である。
  - GTFS Realtime は、公共交通機関が運行車両に関するリアルタイムの最新情報をアプリケーション デベロッパーに提供できるようにするためのフィードの仕様である。
  - Google 社が規格化している。
- 情報流通連携基盤・外部仕様書
  - Streams API を利用してリアルタイムデータの伝送ができる規格になっている。

### 11.1.3 表形式データの定義を記述するフォーマット: Simple Data Format

Simple Data Format<sup>47</sup>は、以下のようなフィールドからなる JSON 形式のフォーマットであり、CSV 形式のデータ定義をそのファイル外で行う規格の 1 つである。

- name (データ名)
- licenses (ライセンス)
- datapackages\_version (バージョン)
- resources (CSV ファイルの定義)

---

<sup>45</sup> Moussa, Feras and Yoshino, Takeshi. Streams API. [Online] November 5, 2013. W3C Working Draft. <http://www.w3.org/TR/streams-api/>.

<sup>46</sup> Google. GTFS-realtime. July 26, 2012. <https://developers.google.com/transit/gtfs-realtime/>

<sup>47</sup> Simple Data Format. March 16, 2014. <http://dataprotocols.org/simple-data-format/>

- url (データの URL)
- path (データのパス)
- schema (url 又は path が示す CSV データの定義)
- fields (CSV データのカラム定義)
- name (カラム名)
- type (データ型 / string, number, integer, date, time, datetime, boolean, binary, object, geopoint, geojson, array, any)
- description (カラムの説明)

Simple Data Format に基づく情報は、UTF-8 で記述されるべきであると規定されている。また、Simple Data Format が参照する CSV データも、UTF-8 で記述されるべきと規定されている。

W3C の CSV on the Web Working Group Charter<sup>48</sup>において、RFC 4180 の新しいバージョンの規格が検討されている。その検討の対象候補として、Simple Data Format が挙げられている。

---

<sup>48</sup> <http://www.w3.org/2013/05/lcsv-charter/>

## 11.2 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、代表的な識別子に関する規格を、表 11-5 に示す。

表 11-5 オープンデータの識別に利用できる識別子規格

種類	規格名	URI 表現例	運営主体	説明	長さ	永続性	個体識別
汎用	ucode [ITU-T H.642.1]	urn:ucode:_0123456789ABCDEF0123456789ABCDEF	ユビキタス ID センター	モノ・場所・概念等あらゆるものに付与できる ID である。ID の再利用を禁じているため、唯一性は永続的に保証される。	128bit	○	○
物流	EPC SGTIN (Serialized Global Trade Item Number)	urn:epc:id:sgtin:4512345.167890.2 urn:epc:tag:sgtin-96:2.4512345.167890.2	GS1	商品を識別するコード。96 ビットコードである SGTIN-96 では、ヘッダ (8 ビット)、流通形態を表すフィルタ (3 ビット)、パーティション (3 ビット)、企業コード (20~40 ビット)、アイテムコード (24~4 ビット)、シリアル番号 (38 ビット) と続く。企業コードとアイテムコードは合計 44 ビットである。	96bit	△	○

種類	規格名	URI 表現例	運営主体	説明	長さ	永続性	個体識別
電子データ	DOI (Digital Object Identifiers) [ISO 26234]	http://dx.doi.org/10.1021/jo0349227	国際 DOI 財団 (The International DOI Foundation)	インターネット上のドキュメントに恒久的に与えられる識別子。サーバの移転によるリンク切れを回避するため、DOI ディレクトリを経由させている。学术论文の分野で広く使われており、学術雑誌や論文誌の記事に付与されている。書籍のタイトルだけでなく、任意のページや図表、CD の 1 曲ごとに付与することもできる。	可変	○	○
	UUID (Universally Unique Identifier) [ISO/IEC 11578]	urn:uuid:f81d4fae-7dec-11d0-a765-00a0c91e6bf6	なし (乱数)	分散システムにおいて、どこかが統制を取らなくても一意に識別できることを目的としたコード。現在よく利用されているのは、乱数に基づく version 4 である。ブログ等のコンテンツ ID として使われることが多い。	128bit	×	○

種類	規格名	URI 表現例	運営主体	説明	長さ	永続性	個体識別
企業・組織	企業コード [ISO 6523]	urn:oid:1.3.170.201233049	ISO が定めた ICD (International Code Designator)	組織(企業)を識別するコードの付与方法を ISO (ISO/IEC JTC1 SC32) が定めたもので、複数の企業コードや組織コードを包含することの出来るマルチコード。先頭の 4 桁が ICD を識別する。それ以降の表記は、ICD が決定する。現在、150 ほどの ICD が登録されている。	可変	×	×
	帝国データバンクコード	urn:oid:1.3.170.201233049	帝国データバンク	ICD の 1 つ。帝国データバンクによる、企業信用調査の対象を識別するための ID。企業による電子証明書の取得等でも利用されている。約 175 万件登録されている。	10 進 9 桁	×	×
	組織コード	urn:oid:1.3.147.123456	一般財団法人日本情報経済社会推進協会(JIPDEC)	EDI や AIDC メディアによる企業間の情報共有・情報連携システムにおいて、情報の送り手あるいは受け取り手となる企業を一意に識別するための企業コード。約 25,000 件登録されている。	0-9、A-Z の 12 桁	×	×

種類	規格名	URI 表現例	運営主体	説明	長さ	永続性	個体識別
自治体	都道府県・市区町村コード（統計に用いる標準地域コード）	<a href="http://statdb.nstac.go.jp/lod/sac/13101">http://statdb.nstac.go.jp/lod/sac/13101</a>	総務省	都道府県及び市町村の区域を示す統計情報の表章及び当該情報の相互利用のための基準であり、統計審議会の答申を踏まえ、昭和45年4月（1970年4月）に定められたもの。	10進5桁	○ (*)	×
書籍	ISBN [ISO 2108]	<a href="urn:isbn:4-13-060800-2">urn:isbn:4-13-060800-2</a>	ISBN 国際本部 （ International ISBN Agency） 国内では日本図書コード管理センター	書籍を識別するための番号体系。 X-AAAA-BBBB-C という形の 10 桁コード（ISBN-10）と、その先頭に 978 を補った 13 桁コード（ISBN-13）がある。X は言語圏、A は出版社番号、B は書名番号、C はチェックディジットである。X, A, B の桁数は規模により変わる。 ISBN-13 は JAN/EAN コードと統合されている（書籍 JAN コード）。	10進10桁又は13桁	×	×

種類	規格名	URI 表現例	運営主体	説明	長さ	永続性	個体識別
	ISSN [ISO 3279]	urn:issn:1560-1560	ISSN ネットワーク／国内では国会図書館	学術雑誌等、逐次刊行物を識別する番号体系。8桁の数字からなり、通常4桁-4桁に分けて表記される。上位4桁が国ごとに割り当てられ、その次の3桁が追い番で付与される。最後の1文字はチェック用であり、モジュラス11で計算される。	10進8桁	×	×
その他	OpenID	http://<username>.openid.ne.jp/	OpenID 財団 (OpenID Foundation)	シングル・サイン・オン (複数のサイトに同じID・パスワードでログインする) のためのユーザ識別子。	可変長	×	×



## 11.3 オープンデータ化に有用なツール

### 11.3.1 Web サービス

Web サービスとは、PC やスマートフォンに搭載されているブラウザに対して、HTTP というプロトコルに則って情報を提供するサービスをいう。レンタルサーバサービスのほとんどで、web サービスの機能を提供している。代表的なツールとその入手先を表 11-6 に示す。

表 11-6 代表的な Web サーバとその入手先

ツール名	開発・提供元	入手先
Apache HTTP Server (無償)	Apache Foundation	<a href="http://www.apache.org/">http://www.apache.org/</a>
Microsoft Internet Information Services (IIS) (無償)	Microsoft Corporation	<a href="http://www.microsoft.com/ja-jp/server-cloud/windows-server/">http://www.microsoft.com/ja-jp/server-cloud/windows-server/</a>

### 11.3.2 データカタログサービス

データカタログサービスとは、データの登録・管理を行い、ポータルサイトとして公開するサービスを提供するソフトウェアである。

これを実現するツールとしては、CKAN が有名である。CKAN は、The Open Knowledge Foundation が開発し、無償で提供されているソフトウェアであり、<http://www.ckan.org/> から入手できる。

CKAN の画面遷移について、第 12 章で解説する。

### 11.3.3 GIS システム

GIS システムとは、地理情報形式データの作成・編集を行うソフトウェアである。

代表的なツールとその入手先を表 11-7 に示す。

表 11-7 代表的な GIS ツールとその入手先

ツール名	開発・提供元	入手先
QGIS（無償）	QGIS Development Team	<a href="http://qgis.org/">http://qgis.org/</a>
Google Earth（無償）	Google	<a href="http://www.google.co.jp/earth/">http://www.google.co.jp/earth/</a>
GRASS GIS（無償）	GRASAS Development Team	<a href="http://grass.osgeo.org/">http://grass.osgeo.org/</a>
ArcGIS（有償）	ESRI	<a href="http://www.esri.com/products/arcgis/">http://www.esri.com/products/arcgis/</a>

#### 11.3.4 情報流通連携基盤

情報流通連携基盤は、オープンデータを登録・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をいう。

このような環境を整備する際に、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ための規格として「情報流通連携基盤・外部仕様書<sup>49</sup>」が公開されている。

#### 11.3.5 RDF レポジトリ

RDF レポジトリとは、RDF データを格納し、SPARQL による検索を受け付けるデータベースシステムである。

<sup>49</sup> <http://www.opendata.gr.jp/cfc/>

代表的なレポジトリとその入手先を表 11-8 に示す。

表 11-8 代表的な RDF レポジトリの入手先

ツール名	開発・提供元	入手先
AllegroGraph (有償・無償)	Franz	<a href="http://www.franz.com/agraph/allegrograph/">http://www.franz.com/agraph/allegrograph/</a>
Apache Jena (無償)	Apache Foundation	<a href="http://jena.apache.org/">http://jena.apache.org/</a>
Neo4j (無償)	Neo Technology	<a href="http://www.neo4j.org/">http://www.neo4j.org/</a>
Sesame (有償・無償)	Aduna	<a href="http://www.openrdf.org/">http://www.openrdf.org/</a>
Virtuoso RDF (有償・無償)	OpenLink Software	<a href="http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSRDF">http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSRDF</a>

## 第12章 (付録)CKAN 解説

本章では、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。

### 12.1 CKAN とは

#### 12.1.1 CKAN 概説

CKAN<sup>50</sup> (図 12-1) は、web ベースのデータ管理・配信システムであり、Open Knowledge Foundation により無償で提供されている。

CKAN は、data.gov (米国)、data.go.uk (英国)、publicdata.eu (EU)、data.gov.au (オーストラリア)、DAT.GO.JP (政府データカタログサイト 試行版)、datameti.go.jp (Open DATA METI) など、オープンデータを配信する多くの政府系組織で利用されている。

下記サイトに、公式のインストール・設定方法がまとめられている。

<http://docs.ckan.org/en/latest/>

また、下記サイトで、環境ごとのインストール方法が紹介されている。

<https://github.com/okfn/ckan/wiki/How-to-Install-CKAN>

---

<sup>50</sup> <http://ckan.org/>

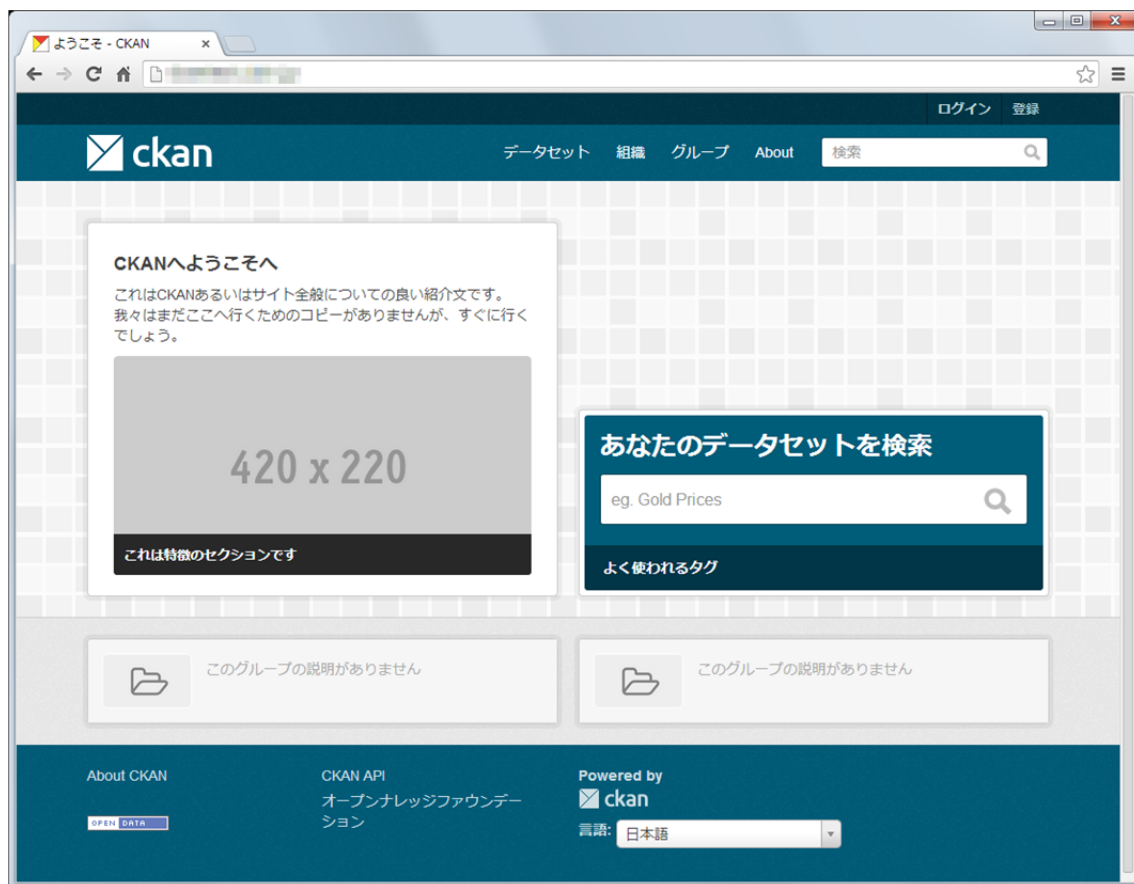


図 12-1 CKAN 初期画面

### 12.1.2 用語の解説

CKAN で利用されている用語の解説を以下に記す。

- ユーザ
  - CKAN にデータを登録する主体。
- データセット
  - 複数のデータをまとめたもの。
  - 例えば、「〇〇年統計データ」「××地区温度データ」など。
- 組織
  - データの公開・管理（アクセス制御）を行う主体。
  - 例えば、「〇〇省」「〇〇課」「〇〇局」など。
  - 組織単位でデータセットを管理できる。
  - 組織に追加したユーザに、データセットの追加・編集権限や閲覧権限を与えられる。
- グループ
  - データセットをコミュニティやトピック単位でまとめたもの。

- タグ
  - データの特徴を説明したもの。
  - 例えば、「財政」「測量」「交通」など。
  - データを検索するためのキーになる。

### 12.1.3 運用前に検討・準備すべき事項

CKAN を利用してデータカタログを運用する前に、以下に示す事項を検討・準備しておくことが望ましい。

#### 12.1.3.1 公開するオープンデータの洗い出し

公開対象のオープンデータをリストアップする。第 3 章の手順フローを参照のこと。

#### 12.1.3.2 オープンデータを管理するポリシーの策定

以下に示すポリシーを事前に策定することが望ましい。

##### (1) データセット・組織の決め方

公開・非公開の制御はデータセットごとになされる。非公開のデータセットは、データセットが属する組織に所属するユーザのみが閲覧できる。

上記を参考に組織とデータセットをリストアップし、公開対象のオープンデータを、データセットごとにまとめる。

##### (2) グループやタグの決定

グループやタグは、利便性を向上させるための項目である。グループやタグとして何を設定するか、また、各オープンデータをどのグループやタグに所属させるかを決定する。

##### (3) データ提供の利用ルールの選定

それぞれのオープンデータに対して適用すべき利用ルールを選定する。

##### (4) データの登録・管理規則の策定

CKAN にアクセスしてデータを登録する担当者とその手順を明確にし、規則として明文化する。アカウントの発行申請手順や、データの登録承認手続きなどがこれにあたる。

また、組織ごとに、管理者ユーザを定める。

#### 12.1.3.3 要求仕様の策定

CKAN システムのインストールには、サーバシステムの知識が必要である。また、利用形態によっては、コンソール操作や、システムのコード変更を要する設定もある。

これらの作業を外部委託するならば、少なくとも以下の項目を記した要求仕様を準備する。

- 対象とするオープンデータのリスト。
- 登録するデータセット・組織とそれらに所属させるオープンデータ・ユーザ。
- グループ・タグの設定。
- web ページからユーザ登録できないようにするためには、その指示。

要求仕様に、運用マニュアルの作成を含めることが望ましい。

#### 12.1.3.4 データの整備計画

本書第 3 章や第 9 章に基づき、リストアップしたデータを、機械判読性の高いデータに変換するための計画を立て、それに基づき実施する。

データは、たとえ機械判読性が低くてもオープンな利用ルールでなくても、公開されることが望ましい。機械判読性の高いデータやオープンな利用ルールが整備でき次第、追加・更新すればよい。

計画は年度などある程度の期間ごとに、実施状況と比較して見直すことが望ましい。

## 12.2 CKAN の利用方法

本節では、CKAN の利用方法を、実際の画面を交えて解説する。

### 12.2.1 アカウント登録

まず、CKAN にデータを登録するユーザ（アカウント）を登録する。その登録方法は、以下の通りである。

1. 初期画面の右上にある「登録」リンクをクリックする（図 12-2 左）と、ユーザ名やメールアドレスを入力する画面が表示される（図 12-2 右）。
2. ユーザ名・メールアドレス・パスワードなど、必要事項を入力する。
3. 「アカウントの作成」ボタンを押すと、登録が完了する。このとき、ログインされた状態になっている。



図 12-2 アカウントの登録



## 12.2.1 データセットの作成

アカウントを作成すると、そのユーザでデータセットを作成できるようになる。以下にその手順を示す。

1. 初期画面（ログイン後の画面／図 12-3 左上）の上にある「データセット」をクリックする。
2. 「データセットを追加」ボタンを押す（図 12-3 右上）。
3. タイトル欄の下にある「編集」ボタンを押す（図 12-3 左下）。
4. タイトル・データの説明・タグ・ライセンスを入力する（図 12-3 右下）。
  - URL のみ必須である。
  - タイトルが英数字のみの場合は、3.の処理は不要である。
  - 「編集」ボタンを押して現れる URL 欄には、タイトルとほぼ同じ意味の英数字を入力する。
  - ライセンスは、オープンな利用ルールを選択することを強く推奨する。
5. これでデータセットの登録は完了である。  
ここで「Next: データの追加」ボタンを押すと、続いてデータを登録できる。

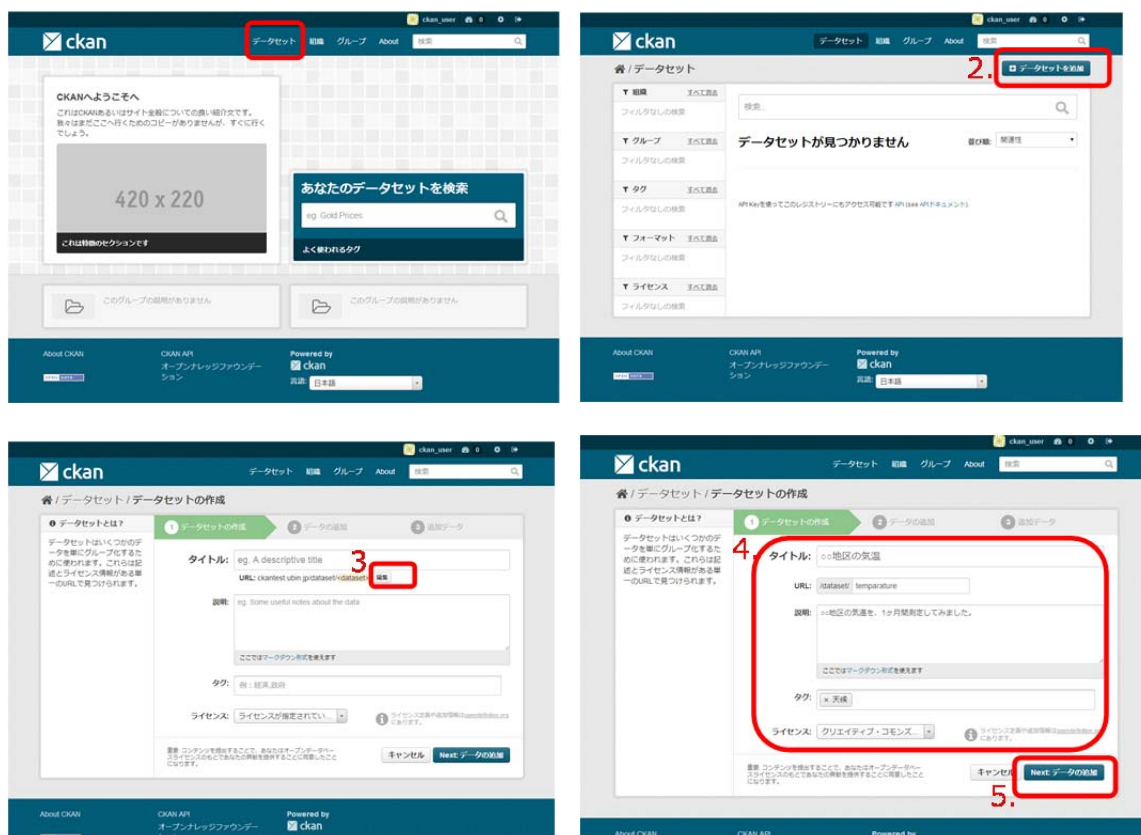


図 12-3 データセットの作成

## 12.2.2 データの登録

データセットの登録が完了すると、そのデータセットに属するデータを登録できる。以下にその手順を示す。

6. 登録するデータと、それを説明する情報を登録する (図 12-4 左)。
  - リソースのみ必須である。
  - データは「ファイルへのリンク」「API へのリンク」「ファイルのアップロード」から選択する。
  - 「ファイルのアップロード」をするには、サーバに **datastore** の設定が必要である。
7. 「Next: 追加情報」ボタンを押す (図 12-4 左)。
  - ここで「保存して別を追加」ボタンを押すと、再度 6 の入力画面が表示される。同じデータの別フォーマットファイルなどを追加できる。
8. 必要な追加情報を入力する (図 12-4 右)。
  - データの作成者やその **email**、メンテナ、所属するグループなどを入力する。
9. 「完了」ボタンを押す (図 12-4 右)。これでデータの登録が完了である。

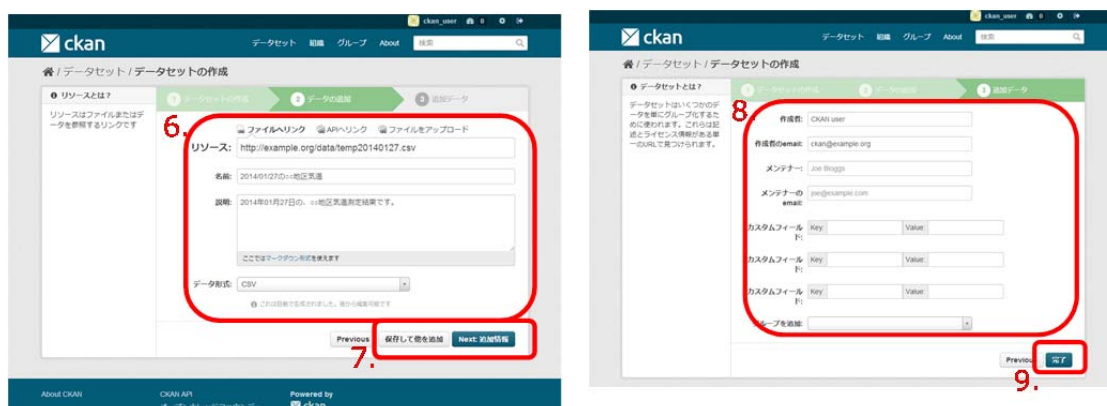


図 12-4 データの登録

データの登録が完了すると、図 12-5 のような画面が表示される。

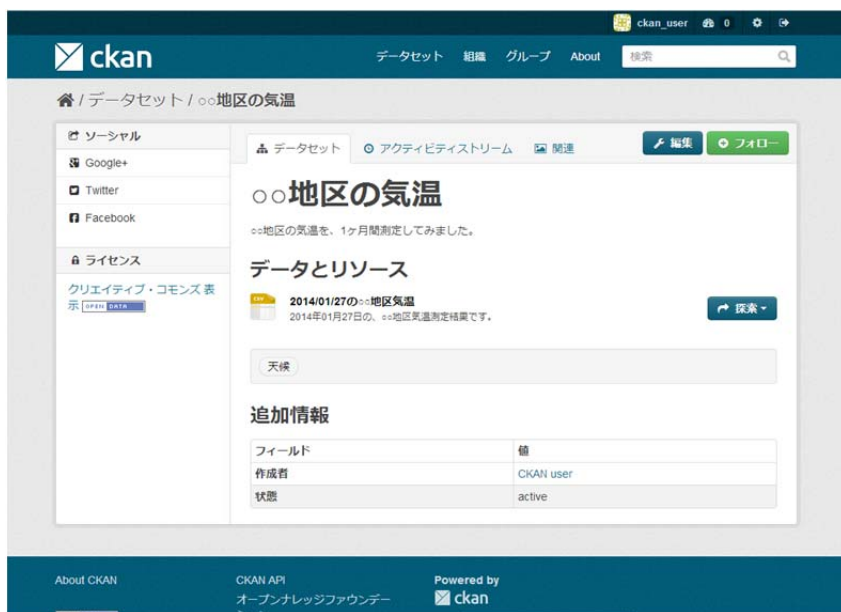


図 12-5 データ登録後の画面

### 12.2.3 組織の登録

組織を登録することにより、データのアクセス制御ができるようになる。以下にその手順を示す。

1. 初期画面（ログイン後の画面／左上）の上にある「組織」をクリックする。
2. 「組織を追加」ボタンを押す（右上）。
3. タイトル欄の下にある「編集」ボタンを押す（左下）。
  - タイトル・URL・説明・画像 URL を入力する。
  - URL のみ必須である。
  - タイトルが英数字のみの場合は、3.の処理は不要である
4. URL 欄には、タイトルとほぼ同じ意味の英数字を入力する（右下）。
5. 「組織の作成」ボタンを押す。これで組織の登録が完了である。

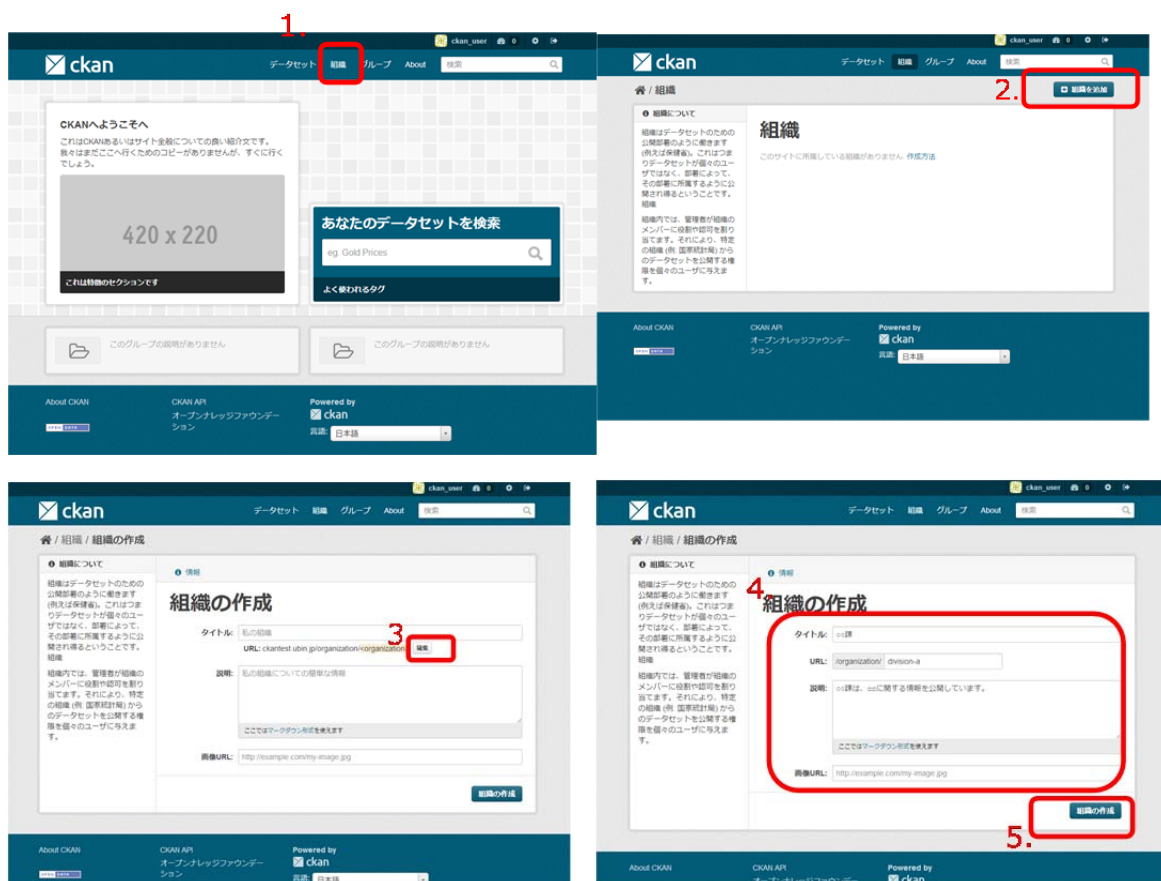


図 12-6 組織の作成

組織の作成後に「データセットの追加」を行うと、その組織に属するデータセットを追加できる（図 12-7 左）。その手順は節と同じである。ただし、データセットを追加する際に、所属する組織を指定できるようになっている（図 12-7 右）。

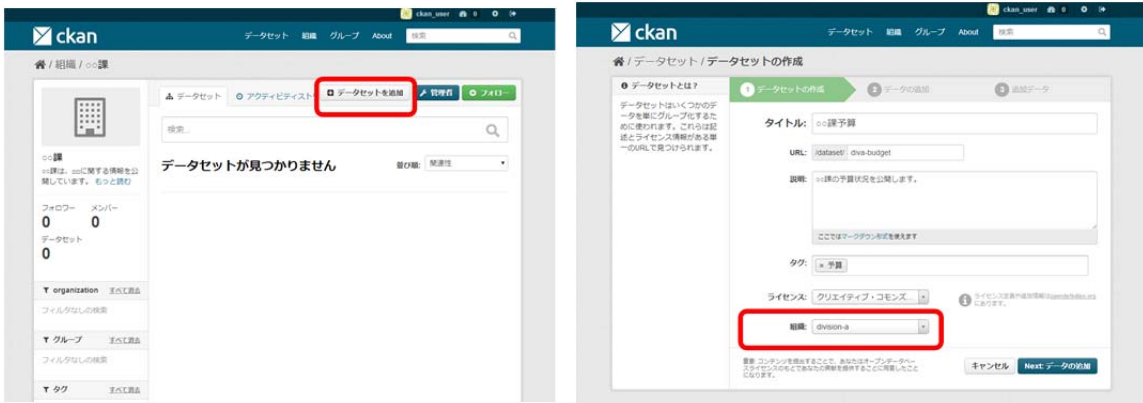


図 12-7 組織登録後のデータセット追加

#### 12.2.4 組織へのメンバ追加

組織を作成したユーザが、組織の管理者となる。組織へのメンバの追加手順を以下に示す。

1. 組織ページにある「管理者」ボタンを押す (図 12-8 左上)。
2. 「メンバ」タブを押す (図 12-8 右上)。
3. 「メンバの追加」ボタンを押す。
4. ユーザの入力欄をクリックしてユーザ名を入力する (図 12-8 左下)。
  - 途中まで入力すると、候補が表示されるので、そこから選択できる。
5. このユーザに与えるロールを選択する (図 12-8 右下)。  
ロールは「管理者」「編集者」「メンバ」の3種類であり、画面左側に説明がある。
6. 「追加」ボタンを押す。

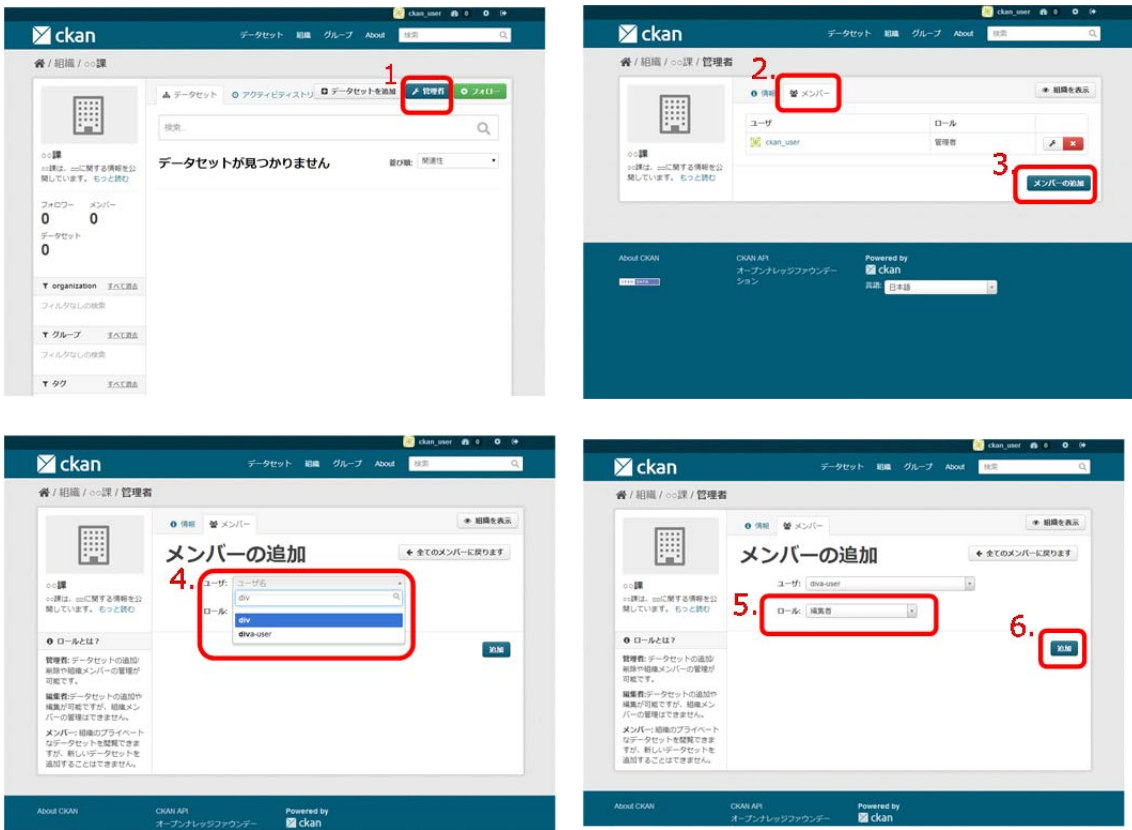


図 12-8 組織へのメンバ追加

## 12.2.5 CKAN によるアクセス制御

組織を登録すると、データセットの登録・編集画面に「公開・非公開」の選択肢が追加され、パブリック（公開）又はプライベート（非公開）を選択できるようになる（図 12-9 左）。

プライベート（非公開）のデータセットは、組織内ユーザのみ閲覧できる（図 12-9 右）。この機能は、データの公開前準備等の用途で利用できる。



図 12-9 CKAN によるアクセス制御



## 12.2.6 CKAN の管理ページ

CKAN の管理者は、データカタログシステムの管理ページにアクセスできる<sup>51</sup>。

管理者は、データカタログ内の全ユーザ・データセット・データを閲覧・編集・削除できる。

Web ページからは、以下のような機能を利用できる。

- サイトの見栄え (Look & Feel) の変更 (図 12-10 左上)
- データセットが所属する組織の変更 (図 12-10 右上)
- データセットの削除
- ユーザ管理 (図 12-10 下)

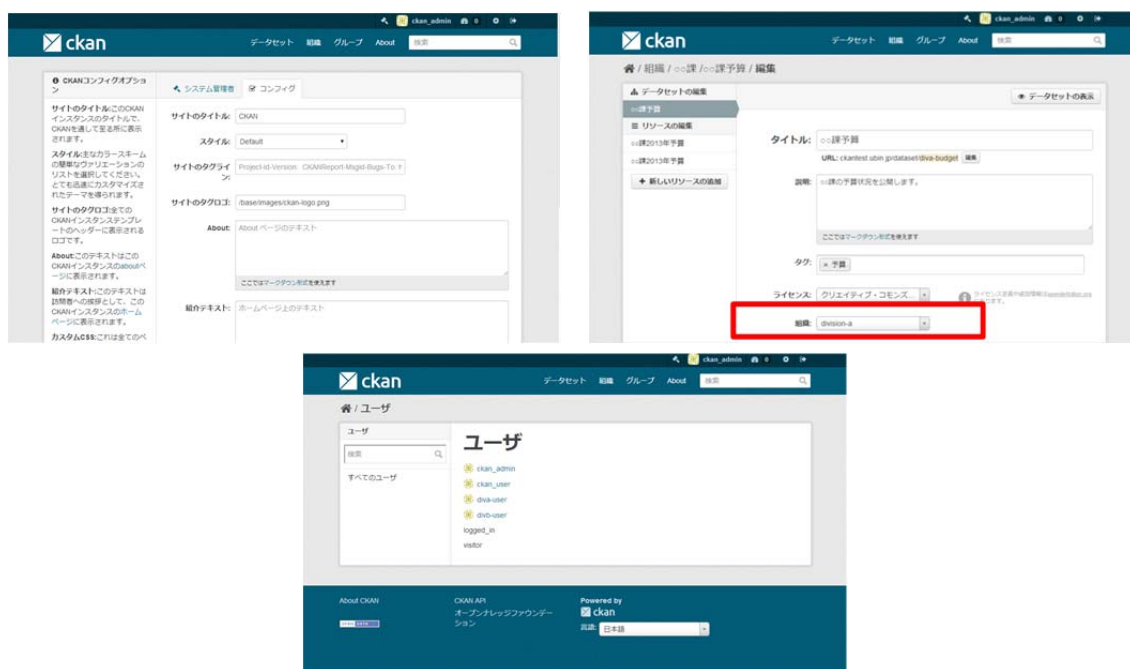


図 12-10 CKAN の管理画面

<sup>51</sup> ただし、管理者の設定は、サーバのコンソールからコマンドを発行する必要がある。

## 参考文献

1. **Shafraanovich, Y.** Common Format and MIME Type for Comma-Separated Values (CSV) Files. [Online] October 2005. RFC 4180. <http://www.ietf.org/rfc/rfc4180.txt>.
2. **Obama, Barack.** Transparency and Open Government. [Online] 2009. [http://www.whitehouse.gov/the\\_press\\_office/TransparencyandOpenGovernment](http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment).
3. —. Building a 21st Century Digital Government. [Online] 2012. <http://www.whitehouse.gov/the-press-office/2012/05/23/presidential-memorandum-building-21st-century-digital-government>.
4. **Moussa, Feras and Yoshino, Takeshi.** *Streams API*. [Online] November 5, 2013. W3C Working Draft. <http://www.w3.org/TR/streams-api/>.
5. **Gandon, Fabien, Schreiber, Guus and Beckett, Dave.** *RDF 1.1 XML Syntax*. [Online] February 25, 2014. W3C Recommendation. <http://www.w3.org/TR/rdf-syntax-grammar/>.
6. **Feigenbaum, Lee, et al., et al.** *SPARQL 1.1 Protocol*. [Online] May 21, 2013. W3C Recommendation. <http://www.w3.org/TR/sparql11-protocol/>.
7. **Bray, Tim, et al., et al.** Extensible Markup Language (XML) 1.1 (Second Edition). [Online] August 16, 2006. W3C Recommendation. <http://www.w3.org/TR/xml11/>.
8. 日本再興戦略. (オンライン) 2013年6月14日. 閣議決定. [http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/saikou\\_jpn.pdf](http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/saikou_jpn.pdf).
9. 総務省. 統計におけるオープンデータの高度化. (オンライン) 2013年5月28日. [http://www.soumu.go.jp/menu\\_news/s-news/01toukei01\\_02000024.html](http://www.soumu.go.jp/menu_news/s-news/01toukei01_02000024.html).
10. 世界最先端 IT 国家創造宣言. (オンライン) 2013年6月14日. 閣議決定. <http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf>.
11. 総務省. 情報通信白書のオープンデータ化の実施. (オンライン) 2013年4月19日. [http://www.soumu.go.jp/menu\\_news/s-news/01tsushin02\\_02000053.html](http://www.soumu.go.jp/menu_news/s-news/01tsushin02_02000053.html).
12. Simple Data Format. [Online] March 16, 2014. <http://dataprotocols.org/simple-data-format/>.
13. **Open Knowledge Foundation.** Open Data Handbook. *What is Open Data?* [Online] <http://opendatahandbook.org/en/what-is-open-data/>.
14. **G8.** Open Data Charter. [Online] 2013. (原文) <https://www.gov.uk/government/publications/open-data-charter>、(邦訳) [http://www.mofa.go.jp/mofaj/gaiko/page23\\_000044.html](http://www.mofa.go.jp/mofaj/gaiko/page23_000044.html).
15. **Google.** *GTFS-realtime*. [Online] July 26, 2012. <https://developers.google.com/transit/gtfs-realtime/>.
16. **European Commission.** Digital Agenda for Europe. [Online] 2011.

<http://ec.europa.eu/digital-agenda/digital-agenda-europe>.

17. 5 ★ Open Data. [Online] (原文) <http://5stardata.info/>、(邦訳)  
<http://5stardata.info/ja>.

18. 高度情報通信ネットワーク社会推進戦略本部. 電子行政オープンデータ戦略. (オンライン) 2012年7月14日. <http://www.kantei.go.jp/jp/singi/it2/denshigyousei.html>.

19. -. 電子行政オープンデータ推進のためのロードマップ. (オンライン) 2013年6月14日.  
<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryous3.pdf>.

20. 各府省情報化統括責任者(CIO)連絡会議. 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). (オンライン) 2013年6月14日.  
<http://www.kantei.go.jp/jp/singi/it2/densi/>.

21. -. 日本のオープンデータ憲章アクションプラン. (オンライン) 2013年10月29日.  
[http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan\\_jp.pdf](http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan_jp.pdf).

## オープンデータ化ガイド 改正履歴

版	発行日	主な改正内容
第1版	2014年〇月〇日	初版を発行。

# オープンデータ化ガイド

第 1 版

2014年〇月〇日

発行

オープンデータ流通推進コンソーシアム



**OPEN DATA**  
オープンデータ流通推進コンソーシアム