

平成25年度技術委員会

オープンデータ流通推進コンソーシアム オープンデータ化のための技術ガ イド骨子案

2014.02.14

オープンデータ流通推進コンソーシアム 事務局

「オープンデータ化のための技術ガイド」の作成目的・作成方針

■ 作成目的

- ▶ これからオープンデータ化に取り組もうとする公的機関や民間組織の職員が、保管しているデータをオープンデータ化するために必要な技術的事項の解説を行う。
 - ◇オープンデータ化を始めるための手引き
 - ◇保管しているデータをオープンデータとして整備するための手引き

■ 作成方針

- ▶ 以下の資料をベースとし、関連する規格との整合性を考慮した改訂を加える。
 - ◇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）^(*1)
 - ◇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^(*1)
 - ◇オープンデータ化のためのデータ作成に関する技術ガイド^(*2)
- ▶ オープンデータ化を始めるための手引きを追加する。

■ 備考

- ▶ 本資料はガイドの骨子案である。最終的には文章化したドキュメントを作成する。
- ▶ 最終的には、データガバナンス委員会で別途検討している「行政職員向けの利用ルール案の解説・FAQ」と統合する。
- ▶ 必要な事項については、「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）」の改訂を提案する。

(*1) 電子行政オープンデータ実務者会議 <http://www.kantei.go.jp/jp/singi/it2/densi/>

(*2) オープンデータ流通推進コンソーシアム 技術委員会資料3-3, 3-4

目次

1. はじめに	4
2. オープンデータ化の背景と意義	6
3. Getting Started: データをオープン化する手法	12
4. オープンデータ化のための技術的指針	25
付録 オープンデータに関する技術・規格	53



1. はじめに

本書の背景と目的

■ 背景: オープンデータ化への関心の高まり

▶ オープンデータに関する政策

◇電子行政オープンデータ戦略（2012.07 IT総合戦略本部決定）

◇G8サミットでの「オープンデータ憲章」合意（2013.06）

◇「世界最先端IT国家創造」宣言（2013.06 閣議決定）

◇「日本再興戦略」（2013.06 閣議決定）

◇日本のオープンデータ憲章アクションプラン（2013.10 各府省情報化統括責任者（CIO）連絡会議決定）

▶ オープンデータへの取組

◇政府オープンデータポータルサイト（内閣官房）

◇Open Data METI（経済産業省）

◇情報通信白書のオープンデータ化（総務省）

◇自治体によるオープンデータ化への取組（鯖江市・横浜市・流山市・静岡県など）

■ オープンデータの2つの側面

▶ ガバナンスの側面から: オープンデータライセンス

▶ 技術面から: ファイル形式やデータ作成・管理手法


■ 本書の目的: オープンデータ化の実施に必要な技術の解説

▶ オープンデータ化を実施するうえで必要となる技術・規格や、それらの利用方法について解説する。

▶ ライセンスについては別冊「行政職員向けの利用ルール案の解説・FAQ」にて解説する。

本書の構成

1. はじめに
 - ▶ 本書の位置づけ・目的・構成を示す。
2. オープンデータ化の背景と意義
 - ▶ オープンデータに関する背景と意義を解説する。
3. Getting Start: データをオープン化手法
 - ▶ データをオープン化に関する手順を解説する。
4. オープンデータ化のための技術的指針
 - ▶ 表形式／文書／地理データ／リアルタイムデータのそれぞれの形式ごとに、オープンデータ化を行う上での留意事項や推奨事項を解説する。
 - ▶ オープンデータを利用しやすくするためのメタデータの記述手法や留意事項、推奨事項を解説する。
5. オープンデータに関する技術・規格
 - ▶ オープンデータ化にあたって参考になる技術や規格を解説する。



2.オープンデータ化の背景と 意義

オープンデータとは？

■ OpenDefinition.org による定義

- ▶ Open data is data that can be freely used, reused and redistributed by anyone.
 ✧オープンデータとは、誰でも自由に利用・再利用・再配布できるデータである。

■ 5★Open Dataによる指標

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などのデータでも可)
2. 構造化されたデータとして公開されている (ExcelやWordなどのデータ)
3. 非独占の (標準化された) 形式で公開されている (CSVなどのデータ)
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)



■ 「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 IT総合戦略本部決定) による定義

- ▶ 機械判読に適した形式のデータを
- ▶ 営利目的も含めた二次利用が可能な利用ルールで公開されたもの

■ 本書では「電子行政オープンデータ推進のためのロードマップ」による定義を採用する。つまり

1. オープンなライセンスで提供されている
2. 機械可読に適した形式のデータ

(*) 2013年6月 IT総合戦略本部決定 <http://www.kantei.go.jp/jp/singi/it2/densi/>

政府における最近のオープンデータ化の流れ

2012.07 「電子行政オープンデータ戦略」 IT総合戦略本部決定 (*1)

- オープンデータ戦略の意義・目的と方向性を明確にする
- 政府が率先してデータを公開することを宣言

2013.06 「世界最先端IT国家創造宣言」 閣議決定 (*2)

- 「目指すべき社会・姿」の中に「公共データの民間開放（オープンデータの推進）」が盛り込まれる

2013.06 「日本再興戦略」 閣議決定 (*3)

- 「ビッグデータやオープンデータの利活用が世界最高水準で実現するよう積極的に進める」ことを宣言

2013.06 「電子行政オープンデータ推進のためのロードマップ」 (*4)

- 平成25年度中にデータカタログを整備すること、平成27年度末において、他の先進国と同水準のオープンデータの公開と利用を実現することを宣言

2013.10 「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 (*5)

- G8のオープンデータ憲章に基づくアクションプラン
- 2013年に国のオープンデータポータルサイトを開設し、2014年度中に本格稼働を開始することを宣言

2013.01 経済産業省

データポータルサイト「Open Data METI」 (*6)を公開

2013.04 総務省

情報通信白書をオープンデータ形式で公開 (*7)

2013.05 気象庁

気象庁防災情報XMLフォーマットを公開、配信を開始 (*8)
気象観測データのダウンロードサービスを開始 (*9)

2013.06 総務省

次世代統計システムAPIを公開 (*10)

2013.12 内閣官房

政府のデータカタログサイト試行版「DATA.GO.JP」 (*11)を立ち上げ

(*1) <http://www.kantei.go.jp/jp/singi/it2/denshigyousei.html>

(*6) <http://datameti.go.jp/>

(*2) <http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf>

(*7) http://www.soumu.go.jp/menu_news/s-news/01tsushin02_02000053.html

(*3) http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/saikou_jpn.pdf

(*8) <http://xml.kishou.go.jp/>

(*4) <http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou3.pdf>

(*9) <http://www.data.jma.go.jp/gmd/risk/obsdl/>

(*5) http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan_jp.pdf

(*10) <http://statdb.nstac.go.jp/system-info/api/>

(*11) <http://data.go.jp/>

国際的なオープンデータの流れ

■ G8 summitにおいて「オープンデータ憲章」が合意された。(2013.06.18)

▶ その中で、5つの「オープンデータ原則」を宣言している

1. Open Data by Default
 - ◆ オープンデータを原則とする
2. Quality and Quantity
 - ◆ 質的／量的に十分なデータを提供する
3. Usable by All
 - ◆ すべての人々が利用できる
4. Releasing Data for Improved Governance
 - ◆ ガバナンス改善のため、データを公開する
5. Releasing Data for Innovation
 - ◆ イノベーションのため、データを公開する

■ データを公開するための利用しやすいライセンスが整備されつつある。

▶ Creative Commonsなど。

オープンデータ化によるメリット

1. 透明性・信頼性の向上

- ▶ 公共データが二次利用可能な形で提供されることにより、国民が自ら又は民間のサービスを通じて、政府の政策等に関して十分な分析、判断を行うことが可能になる。それにより、行政の透明性が高まり、行政への国民からの信頼を高めることができる。

2. 国民参加・官民協働の推進

- ▶ 広範な主体による公共データの活用が進展し、官民の情報共有が図られることにより、官民の協働による公共サービスの提供、さらには行政が提供した情報による民間サービスの創出が促進される。これにより、創意工夫を活かした多様な公共サービスが迅速かつ効率的に提供され、厳しい財政状況、諸活動におけるニーズや価値観の多様化、情報通信技術の高度化等我が国を取り巻く諸状況にも適切に対応することができる。

3. 経済の活性化・行政の効率化

- ▶ 公共データを二次利用可能な形で提供することにより、市場における編集、加工、分析等の各段階を通じて、様々な新ビジネスの創出や企業活動の効率化等が促され、我が国全体の経済活性化が図られる。
- ▶ また、国や地方公共団体においても、政策決定等において公共データを用いて分析等を行うことで、業務の効率化、高度化が図られる。

高度情報通信ネットワーク社会推進戦略本部決定「電子行政オープンデータ戦略」による
http://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryou2.pdf



3. Getting Started:

データをオープン化する手法

データを公開する際に明らかにすべき項目

1. メタデータ: どんなデータか？

- ▶ メタデータとは、データに関する情報をいう。
- ▶ たとえば「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

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？

- ▶ Web上のアドレス(URL)やAPIを明記する。
- ▶ 複数の形式でデータを取得できることが望ましい。
◇理由: 用途によって、最適なデータ形式が異なる場合があるため。
(例) 機械可読な表形式データと人間可読な表形式データは、必ずしも一致しない。

3. ライセンス: そのデータはどのような条件で取得・利用できるか？

- ▶ 二次利用できるか? / 商用利用できるか? / 利用の際にデータ提供者に通知が必要か? など
- ▶ 詳しくは「行政職員向けの利用ルール案の解説・FAQ」を参照のこと

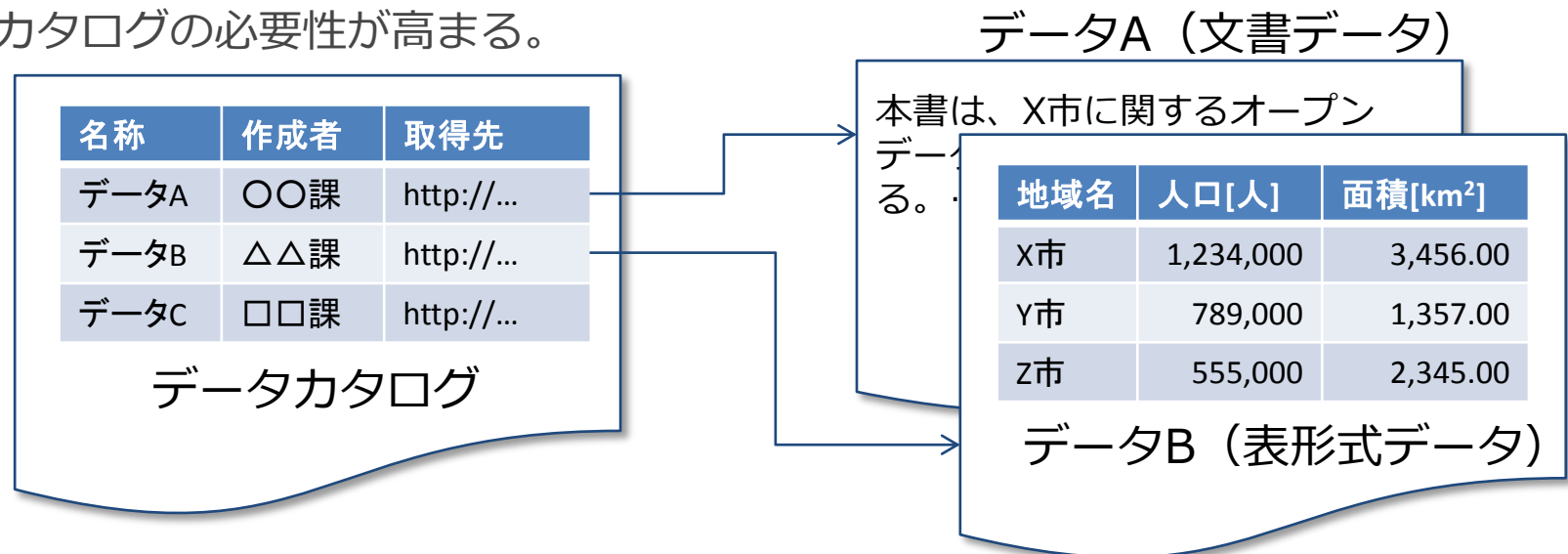
「データ」と「データカタログ」

■ 「データ」とは

- ▶ 公開する情報そのもの。
- ▶ 表形式のデータや文書データ、地理情報データ（地図データ）、リアルタイムデータなどがある。

■ 「データカタログ」とは

- ▶ データの所在、種類、名称など、公開しているデータに関する情報（これをメタデータという）をまとめたもの。いわば、データの目録・索引である。
- ▶ 公開するデータが増加してくるにつれて、それらのデータを検索・一覧する要求が高まる。
→ データカタログの必要性が高まる。



オープンデータ化の技術的レベル

	Level 0	Level 1	Level 2	Level 3
データ	PDFや画像ファイルをWebで公開する。	機械可読なデータを作成し、Webで公開する。 (CSV, XLS, DOCなど)	機械可読性の高いデータを作成し、公開する。 (第4章参照)	RDF、XML等の技術を導入したデータを作成し、公開する。
データのカタログ (目録)	存在しない	カタログを表形式データ (CSVなど)として作成し、公開する。	データカタログシステムを導入する。	RDFやSPARQLを利用したメタデータ検索機能を提供する。
必要なツール	Webサーバ	Webサーバ	Webサーバ+CKANなど	Webサーバ+CKAN+情報流通連携基盤など

■ Level 1に移行するメリット

- ▶ 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。
- ▶ データのありかを電子的に入手できる。

■ Level 2に移行するメリット

- ▶ 利用者のデータ利活用の効率が向上する。(データの再利用性や検索性の向上)

■ Level 3に移行するメリット

- ▶ 他のデータとのマッシュアップや他のデータとの横断検索などが可能になる。
→ 利用者によるデータ利用の幅を広げられる。

オープンデータにとっての識別子 (ID)

■ 識別子 (ID) とは？

- ▶ 一意に識別するためにデータやデータが対象とする実物や組織・場所等に付与する番号を、識別子 (ID) という。

■ オープンデータにとっての識別子とは？

- ▶ 識別子は、グローバルにユニークであるべきである。
 - ◇たとえば社員番号は、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。
- ▶ 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。
 - ◇ucode・DoI (Digital Object Identifiers) ・企業コード (ISO 6523) ・地方自治体コードなど。
 - ◇広く使われている識別子の一覧を付録のii.に掲載する。
- ▶ 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。

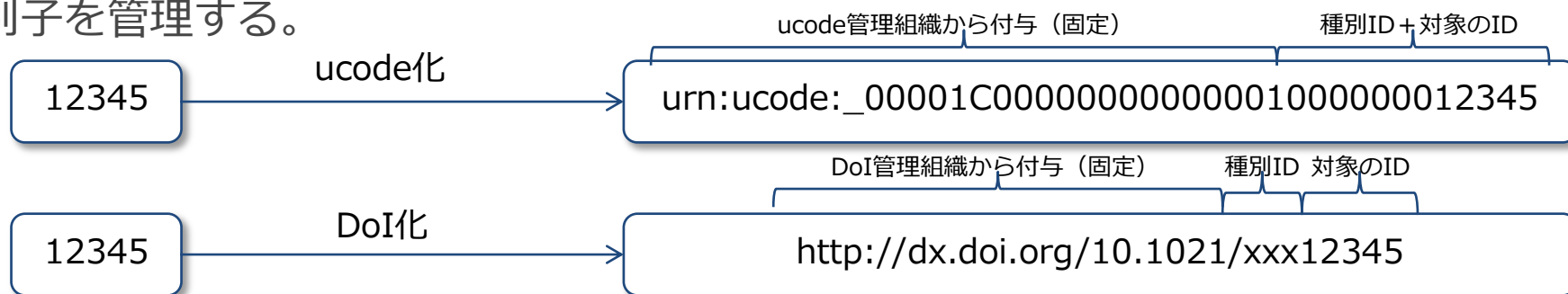
適切な識別子体系がない場合の対処法

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

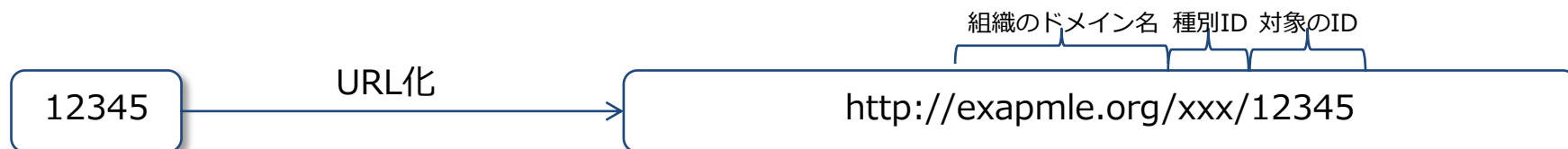
- ▶ 対象の実物や組織・場所に番号を付与する。

2. 付与した番号をグローバル化する。

- ▶ ucodeやDoIなど、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。



- ▶ 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。



オープンデータ化の手順

1. オープンデータ化推進組織の設立

オープンデータ化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

情報の棚卸し

データ形式

管理体制

権利・法律関連

3. 計画立案

オープンデータ化の対象・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、オープンデータ化の作業を行う。

6. 改善点の洗い出し

利用者や作業担当者からのフィードバックを元に、改善点を洗い出す。

5. 公開・運用

オープンデータ管理のマイルストーンに基づき、ある程度の情報が登録された段階で公開し、システムの運用を開始する。

1. オープンデータ化推進組織の設立

■ なぜ「オープンデータ化推進組織」が必要か？

- ▶ オープンデータ化は、組織を横断する取組になる。
- ▶ オープンデータ化を進めて行くにあたり、データを保持している各組織との連携・調整が必要になる。
- ▶ このため、各組織から独立した、オープンデータ化の推進を目的とした組織を立ち上げることが望ましい。

2. 現状把握（情報の棚卸し）

■ 「情報の棚卸し」とは？

- ▶ 組織が管理している資料を、以下の観点でまとめる。
 - ◇担当している部署
 - ◇資料の種類（予算・各種報告・統計・広報など）
 - ◇分量

■ 棚卸しの着目点（実施すべき項目）

1. データの形式

- ◇それぞれの資料の形式を確認する。
 - ◆ 紙（同一情報の電子データがあるか要確認） → ない場合は、公開するにはスキャンする必要あり
 - ◆ 画像（写真・紙をスキャンした画像）
 - ◆ 電子データ（ファイル形式・レベルを含む）

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。（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

- ◇データがどのくらいの頻度で更新されるのか確認する。
- ◇年に1回更新／月に1回更新／適宜更新など

4. データの権利関係

- ◇それぞれの資料について下記を確認する。
 - ◆ 他者が著作権等の権利を有する素材
 - ◆ 法律やプライバシーの観点での課題
- ◇詳細は「行政職員向けの利用ルール案の解説・FAQ」参照のこと。

3. 計画立案

■ 実施すべき項目

- ▶ 現状把握（またはフィードバック）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。
- ▶ マイルストーンを作成し、計画を立案する。
- ▶ 大きな組織であるほど、計画立案が重要である。

■ 計画立案時の留意事項

1. データ形式・システムの準備計画

- ◇ p.15の表を参考に、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。
- ◇ 必要なツールを揃える。（調達にかけるなど）

2. 運用ルールの策定

- ◇ 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。

3. 権利関係の要検討項目を洗い出し

- ◇ 対象のデータをオープンデータとして公開する際に、解決すべき権利関係の問題を洗い出す。
- ◇ 問題がある場合は、それを解決するための計画を立案する。（権利者に確認するなど）

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。
- ◇ 「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」という、いわゆるスモール・スタートの考え方が基本原則とされている。

4. 公開作業

■ 実施すべき項目

- ▶ 立案した計画に基づき、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、そのライセンスを明確にすること。

5. 公開・運用

■ 実施すべき項目

- ▶ 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、オープンデータの提供を開始する。
- ▶ 運用中は、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。


6. 改善点の洗い出し

■ 実施すべき項目

- ▶ 一定の期間ごとに、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、要改善点を洗い出す。
- ▶ 新規のデータを公開するタイミングで、改善点の洗い出しすることが望ましい。
- ▶ 得られた要改善点を解決するための計画を立案する。
- ▶ 公開しているデータのレベルを上げる



3. 計画立案に戻る



4. オープンデータ化のための 技術的指針

指針の概要

■ 目的

- ▶ 機械可読性の高いオープンデータを作成するための、技術的な指針を示す。
 - ◇まず、機械可読性の高いファイル形式を示す。
 - ◇続いて、それらのファイル形式を利用して、機械可読性の高いデータを作成するための指針を記す。
 - ◇また、表形式データや文書データのプロパティを適切に設定することにより、データの作成者や作成日時などのメタデータを埋め込むことができる。これは、必要なデータを検索する時に有用であるだけでなく、データカタログ等の管理システムにオープンデータを登録する手間の軽減にもなる。このための方法について解説する。

■ 対象とするデータ

- ▶ 表形式データ
- ▶ 文書データ
- ▶ 地理情報データ
- ▶ リアルタイムデータ

指針の概要

■ 指針のレベル: 満たすべき指針の重要度にあわせて2つのレベルを設ける。

▶ レベル1

◇レベル1は、オープンデータが満たすことを強く推奨する指針であり、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できるプログラムが書けること。

▶ 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。

- ◆ データを取得した利用者が、そのデータの項目や構造を正しく解釈し、データを扱うプログラムを書けること。

■ 構成

i.	データの公開方法に関する指針	28
ii.	ファイル形式に関する指針	30
iii.	表形式データに関する指針	31
iv.	文書データに関する指針	42
v.	地理情報データに関する指針	45
vi.	リアルタイムデータに関する指針	48
vii.	メタデータの付与手法	50

i. データの公開方法に関する指針

■ データを公開する際には、以下の3項目を明記すべきである。

1. メタデータ: そのデータは何か？

◇たとえば「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

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？

◇Web上のアドレス(URL)やAPIを明記する。

◇複数の形式でデータを取得できることが望ましい。

- ◆ 理由: 用途によって、最適なデータ形式が異なる場合があるため。
(例) 機械可読な表形式データと人間可読な表形式データは、必ずしも一致しない。→ 後述

3. ライセンス: そのデータはどのような条件で取得・利用できるか？

◇詳しくは「行政職員向けの利用ルール案の解説・FAQ」を参照のこと

i. データの公開方法に関する指針

■ 機械可読性の高いデータを公開するうえでの留意点

- ▶ 機械可読な形式と人間可読な形式が、必ずしも一致するとは限らない。

◇本ガイドでは、機械可読性の高い形式を示す。

しかしその形式は、必ずしも人間に読みやすい形式であるとは限らない。



- ▶ 必要であれば、機械可読な形式と人間可読な形式の2種類のファイルを用意して公開することも、方法としてありえる。

◇「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。

ii. ファイル形式に関する指針

■ 基本方針

- ▶ 機械可読性、オープン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、5★ open dataの指標に基づいてまとめると、下記のようなになる。

推奨するファイル形式

	★2	★3	★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)	RDFa
地理空間データ	shape	KML GML	
リアルタイムデータ	(ファイルの形で交換しない)		

※プレインテキスト(TXT)も非独占の文書形式であるが、構造化されていない。

iii. 機械可読な表形式データに関する指針

■ 用語定義

- ▶ 表形式データの代表的なフォーマットであるCSVについて定めたRFC4180(*)の記述に合わせて、以下のように定義する。

◇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。

The diagram shows a table with 5 columns and 5 rows. The first row is the header, and the following four rows are records. The entire table is enclosed in a red box labeled 'ファイル'. The first row is labeled 'ヘッダ'. The second row is labeled 'レコード'. The cell containing '6.5' in the fourth row, fourth column is labeled 'フィールド'.

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

(*) Common Format and MIME Type for Comma-Separated Values (CSV) Files. RFC4180.
<http://www.ietf.org/rfc/rfc4180.txt>

iii. 機械可読な表形式データに関する指針

■ レベル1

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。
2. ヘッダは、1行で構成されるべきである。

} これを満たさないファイルは
RFC4180に準拠しない

■ レベル2

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。
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。
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。
6. 年の値には、西暦表記を備えることが望ましい。
7. フィールドの単位と記数単位（フィールド値の桁を示す数。たとえば、単位が「百万円」である場合、記数単位は「1,000,000」である）が明記されることが望ましい。
8. 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。
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。
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。

iii. 機械可読な表形式データに関する指針／解説

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。
2. ヘッダは、1行で構成されるべきである。

解説

■ RFC4180に、以下のような規定がある。

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もよい。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。

■ つまり…

- ▶ 指針2は、RFC4180の規定3「ヘッダが1行からなるべき」そのものである。
- ▶ 上記と、規定4にある「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」という条件から、1ファイルに複数の表を置くことはできない。これが、指針1の理由である。

iii. 機械可読な表形式データに関する指針／解説

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。
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。
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。
6. 年の値には、西暦表記を備えることが望ましい。

(指針3) このフィールドには、「1.6」という数値と「(*1)」という注釈が含まれている。ここで、注釈へのリンクである(*1)と、自然言語で書かれた注釈文は機械が解読できない。従って、機械可読性の観点からは、(*1)を除き、数値「1.6」のみを記載することが望ましい。

※指針を満たした形式は次頁

解説

年	月	A市	B市	C市	D町
2013	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

(指針4) (指針5) これは第1レコードから第4レコードまでの値がすべて「2013」であることを示している。人間が見ればわかるが、機械は「2013」という値がないフィールドが、前のレコードと同じ値であることを理解できない。むしろ、すべてのフィールドに同じ値を記した方が可読性が高くなる。
 (指針6) 年が経過するごとに、年の値が単調増加する方が、機械は処理しやすい。和暦に比べて西暦の方が、この特徴を有しているため、西暦の記載を推奨する。和暦を併記してもよい。

iii. 機械可読な表形式データに関する指針／解説

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。
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。
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。
6. 年の値には、西暦表記を備えることが望ましい。

解説

(指針3～6を満たした形式)

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

iii. 機械可読な表形式データに関する指針／解説

指針

7. フィールドの単位と記数単位（フィールド値の桁を示す数。たとえば、単位が「百万円」である場合、記数単位は「1,000,000」である）が明記されることが望ましい。
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

■ 指針7について

- ▶ 単位や記数単位は、機械がデータを解読・解析する際に必要である。また、表のタイトル、作成者なども機械が認識できれば処理しやすい。
- ▶ しかし、指針2、指針3によりこれらをファイル中に記載することはできない。
→ 後述するsimple data formatを利用することにより、指針7を満たすことができる。

■ 指針8について

- ▶ 日本語を記述する文字コードには、JIS (ISO-2022-JP)、Shift-JIS、EUC、UTF-8など、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い場合は、機械が読み取ることは難しい。
- ▶ 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など）。

iii. 機械可読な表形式データに関する指針／解説

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。
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

■ 指針9について

- ▶ 後述するsimple data formatを利用するか、データカタログシステムを利用することにより、指針9を満たすことができる。

■ 指針10について

- ▶ データ本体を、XMLやRDFなど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、機械可読性がさらに高まる。

(参考) Simple Data Format

■ Simple Data Format^(*1)の概要

- ▶ Data Packages^(*2)やJSON Table Schema^(*3)等の規格を参照し、以下のようなフィールドを利用して、CSV形式データの(メタデータ)定義を、CSV形式データのファイル外でJSON形式で行う。
 - ◇ 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)
 - ◆ url (データのURL)
 - ◆ path (データのパス)
 - ◆ schema (urlまたは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)
- ▶ フィールド名にボキャブラリを割り当てれば、RDFによるメタデータ表記も可能である。
- ▶ Simple Data Formatにおける文字コード規定
 - ◇ Simple Data Formatに基づく情報は、UTF-8で記述されるべきである。
 - ◇ Simple Data Formatが参照するCSVデータも、UTF-8で記述されるべきである。

■ Simple Data Formatを採用する理由

- ▶ 指針7が求める単位や記数単位をCSVファイル内に記述すると、指針2、指針3に反する。このため、単位や記数単位に関する定義を、ファイル外で行う必要がある。
- ▶ W3CのCSV on the Web Working Group Charter^(*4)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。ここでも、Simple Data Formatが検討対象に挙がっている。
- ▶ このことから、ファイル外で単位や記数単位を記述する仕様として、本書ではSimple Data Formatを採用する。

(*1) <http://dataprotocols.org/simple-data-format/>

(*2) <http://dataprotocols.org/data-packages/>

(*3) <http://dataprotocols.org/json-table-schema/>

(*4) <http://www.w3.org/2013/05/lcsv-charter/>

(参考) Simple Data Format

■ Simple Data Formatによるp.31の表定義記述例

```
{
  "name": "各地域の気温", }   データセット名 "my-dataset"
  "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「℃」

(補足) 「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)」の別添との比較

ガイドライン別添の留意事項	今年度指針
(2) 表形式データにおけるデータの構造の留意事項	
1. 1つのデータセットには、1つのテーブルのみを含める。(複数個のテーブルを含めない)	→ 指針1
2. データセルに、整形や位取りのための文字(スペース、改行、カンマ等)を含めない。	→ 指針3
3. 年の値には、西暦表記とし、和暦を併記する。	→ 指針6
4. 数値等のデータの値やタイトル、単位以外の情報を、セルに含めない。	→ 指針3
5. すべてのセルは、他のセルと結合しない。	→ 指針4
6. 値が存在しない場合を除き、データセルを空白にしない。(データ値を省略しない)	→ 指針5
7. データセルの内容を示すタイトルは、1行で構成する。	→ 指針1
8. データの単位を明記する。	→ 指針7

(補足) 昨年度技術委員会で作成したガイドとの差分

昨年度指針	今年度指針
1. 1つのデータシートには、1種類の表のみを含むべきである。	→ 指針1
2. セルに、整形のためのスペース・改行、位取りのカンマを含めるべきでない。	→ 指針3
3. 年の値には、西暦表記を備えるべきである。	→ 指針6
4. 数値やタイトル・単位以外の情報を、セルに含めるべきではない。	→ 指針3
5. すべてのセルが、他のセルと結合されているべきではない。	→ 指針4
6. 値がない場合を除き、データセルが空白とすべきでない。	→ 指針5
7. データの内容を示すタイトルは、1行で構成されているべきである。	→ 指針1
8. データの単位が明記されているべきである。	→ 指針7
9. データセルの内容・単位・記数単位を示すタイトルが、それぞれ別の行に記載されているべきである。	→ Simple Data Formatを適用
10. データセットは、オープンな標準データ形式で提供されるべきである。	→ ファイル形式に関する指針
11. タイトルやデータ型は、一定の基準に従ったフォーマットで記述すべきである。	→ 指針9
12. データセットの属性や説明を表すメタデータを、XMLやRDFの形式を使ってフォーマルに記述すべきである。 そのメタデータからデータセット本体へリンクし、たどれるようにすべきである。	→ 指針9
13. データセットに含まれるデータ本体を、XMLやRDFの形式を使ってフォーマルに記述すべきである。	→ 指針10

iv. 機械可読な文書形式データに関する指針

指針

- レベル1
 - ▶ なし
- レベル2
 1. 文章に存在する部・章・節・図表などの構造が、機械可読なフォーマットで記述されていることが望ましい。
 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。
 3. 文書形式データが表形式データを含む場合、レベル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。
 4. 文章に対する、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

- 指針1を満たすには…
 - ▶ 文書編集ソフトを利用する場合、文章に存在する部・章・節・図表などの構造を、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（見出しなど）を利用して表現する。
 - ▶ 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を満たすには…
 - ▶ 文章に含まれる余分な空白を除去する。
- 指針3を満たすには…
 - ▶ 文章が図表を含む場合は、指針レベル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など示す。
- 指針43を満たすには…
 - ▶ データカタログシステムを利用する。

章

1. はじめに

本書は、オープンデータ化を実施するうえで必要となる技術・規格や、それらの利用方法について解説する。

本文

1.1 オープンデータとは

OpenDefinition.orgによると、オープンデータとは、誰でも自由に利用・再利用・再配布できるデータである。表1は、オープンデータを提供している各国のサイトを記したものである。

節

表1: 国内外のデータカタログサイト

国名	カタログサイト
米国	http://data.gov/
英国	http://data.go.uk/
日本	http://data.go.jp/

図表

表形式データへのリンク

<http://example.org/.../.csv>にデータあり

(補足) 「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)」の別添との比較

ガイドライン別添の留意事項	今年度指針
(2) 文書形式データにおけるデータの構造の留意事項	
1. 文章に存在する部、章、節、図表等の構造が、コンピュータが明快に認識できる形で記述する。	→ 指針1
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。	→ 指針2
(3) 文書データにおけるデータ形式の留意事項	
1. 文書データは、オープンな標準データ形式で提供する	→ ファイル形式の指針
2. 文書形式データが図表を含む場合、それらを構成する表形式データが添付されているべきである。	→ 指針3
3. 公開前におけるファイル内容、プロパティには十分注意して公開する。	

(補足) 昨年度技術委員会で作成したガイドとの差分

昨年度の指針	今年度指針との対応
1. 文章に存在する部・章・節・図表などの構造が、明快に分かる形で記述されるべきである。	→ 指針1
2. 文章内に、整形のための空白・改行などを含めない。	→ 指針2
3. オープンに利用できるデータフォーマットで公開する。	→ ファイル形式に関する指針
4. 文書形式データが図表を含む場合、それらを構成するレベル1以上の表形式データが添付されている。	→ 指針3
5. 図表やグラフとそのキャプションが結びつくように、文書形式データが構成されている。	→ 指針3
6. 文章の構造を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、文書にリンクされている。	→ 指針4

v. 機械可読な地理情報データに関する指針

指針

■ レベル1

1. 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。

■ レベル2

2. 地理情報に対する、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、当該地理情報にリンクされていることが望ましい。

解説

■ 指針1を満たすには…

- ▶ 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。
たとえば、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、地理情報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。

■ 指針2を満たすには…

- ▶ データカタログシステムを利用する。

(補足) 「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)」の別添との比較

ガイドライン別添の留意事項	今年度指針
(2) 地理空間情報におけるデータの構造の留意事項	
1.地理空間情報のうち、位置情報に関するデータを付与する場合は、緯度経度座標を付与する。付与する際、準拠している座標参照系(世界測地系等)を明記する。文章に存在する部、章、節、図表等の構造が、コンピュータが明快に認識できる形で記述する。	→ 指針1
(3) 文書データにおけるデータ形式の留意事項	
1.データの提供に当たっては、機械判読が可能なベクタ形式に依るものとする。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。	→ ファイル形式の指針

(補足) 昨年度技術委員会で作成したガイドとの差分

昨年度の指針	今年度指針との対応
1. 測地系が明記されている。	→ 指針1
2. 広く利用されているフォーマットで記述されている。	→ ファイル形式に関する指針
3. データに対する利用者が理解できる説明が、XMLやRDFによりメタデータ記述されており、かつその説明がデータをリンクしている。	→ 指針2

vi. リアルタイムデータに関する指針

指針

- レベル1
 1. データの取得仕様を明記するべきである。
 2. 表形式データや地理情報データをファイル形式で取得させる場合は、それぞれのレベル1の指針を満たすべきである。
- レベル2
 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

- 指針1,2について
 - ▶ 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、それらデータを機械が取得し、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。
 - ▶ リアルタイムデータは明らかに機械が取得・解釈するものである。従って、それが表形式データや地理情報データをファイル形式であるならば、少なくともそれぞれのレベル1指針を満たすべきである。
- 指針3を満たすには…
 - ▶ Streams APIを利用する。
 - ▶ 「情報流通連携基盤システム外部仕様書」によりデータを提供する。
 - ▶ リアルタイムでRDFデータを提供する。…などの方法がある。

(補足) 昨年度技術委員会で作成したガイドとの差分

昨年度の指針	今年度指針との対応
1.表形式データ・地理空間データのレベル2以上に準拠した形式のデータを、ファイルとして取得できる。	→ 指針2
2.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。	→ 指針3
3.リアルタイムデータの最新値や差分を取得するための、メタデータ記述に対応したデータ取得規約が提供されている。またはメタデータ記述されたリアルタイムデータを取得する手法が提供されている。	→ 指針3

vii. メタデータの付与方法

■ データの登録ポリシーとメタデータ付与方法

▶ 集中登録方式

- ◇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・部局からデータを集めて、公開する方式。
- ◇この方式の場合は、データを集める際にメタデータを一緒に集める

▶ 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）直接データを公開する方式。
- ◇この場合は、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。

■ ファイルにメタデータを登録する手法

- ▶ 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（次頁にて解説）

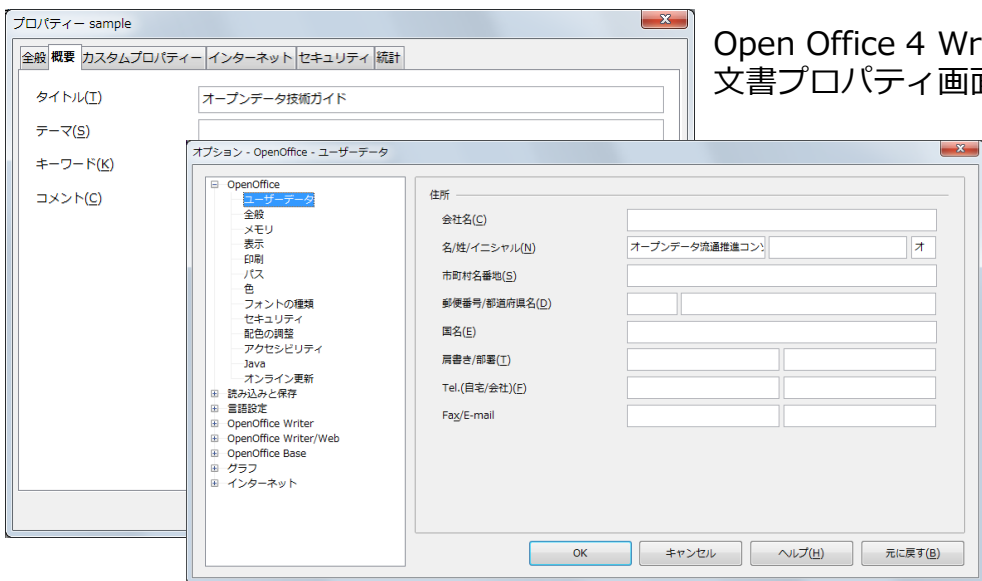
- ◇Microsoft OfficeやOpenOffice、Acrobat等のソフトウェアには、ファイルの「プロパティ」を編集する機能がある。
- ◇これを利用して登録したメタデータを、Apache Tika^(*)（無償）等のソフトウェアを利用して機械が取得できる。

■ 分散登録方式の際の留意点

- ▶ 登録コスト軽減のためには、上記のような方法で取得したメタデータを自動的にデータカタログに追加する手法の検討が必要。

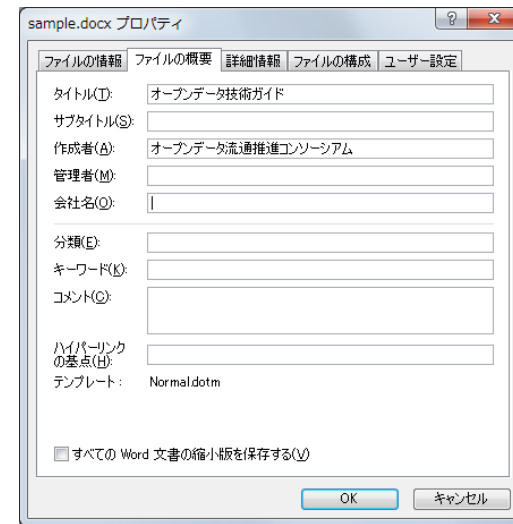
(*) <http://tika.apache.org/>

メタデータの付与例



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

```

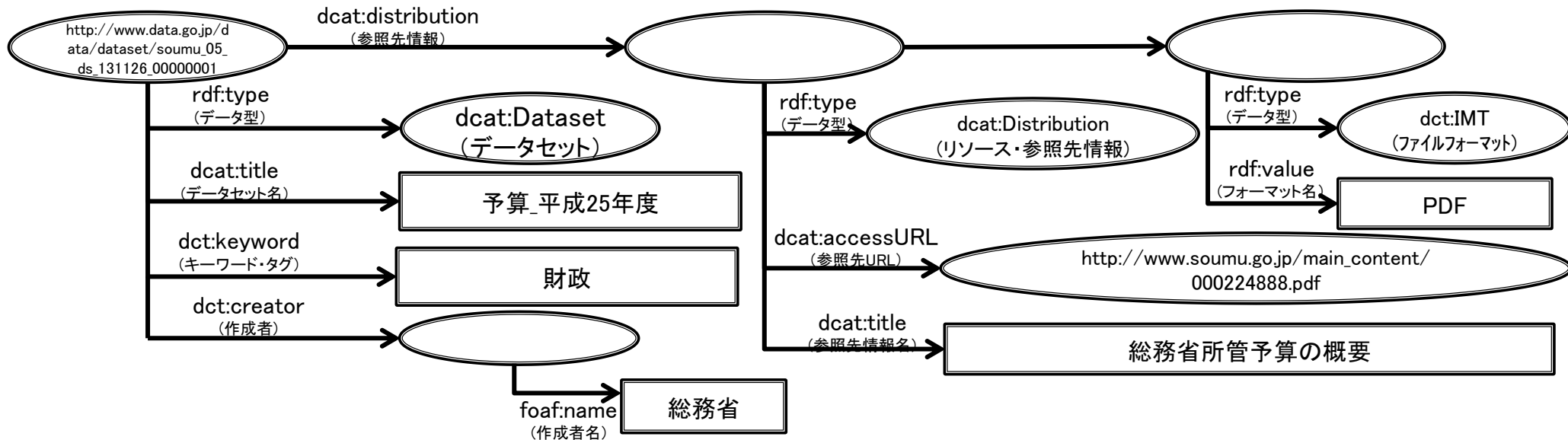
{
  ...
  "dc:creator": "オープンデータ流通推進コンソーシアム",
  "dc:title": "オープンデータ技術ガイド",
  "dcterms:created": "2014-01-27Txx:xx:xx",
  "dcterms:modified": "2014-01-27Txx:xx:xx",
  ...
}
    
```

作成者→
 タイトル→
 作成日時→
 最終更新日時→

(参考) DATA.GO.JPで運用されているCKANのメタデータについて

- CKANに登録されたメタデータは、国際標準化されたボキャブラリに沿って記述されている。

- ▶ 「http://www.soumu.go.jp/main_content/000224888.pdf」に掲載されている、「総務省」が作成した「総務省所管予算の概要」というPDF資料に関する、データカタログサイトに格納されているメタデータは、以下のとおり（下記は、説明のために一部のメタデータを割愛している）。



- メタデータの値についてはブレがあるので、一定の基準が必要。



(付録)オープンデータに関する技術・規格

概要

■ 本章の目的

- ▶ オープンデータを作成・編集・閲覧するうえで、参考になる規約やツール群を解説する。

■ 構成

- i. ファイル形式・データフォーマットに関する規格 55
 - ◇ 表形式・文書形式・地理情報に関する国際的に広く利用されているファイル形式やデータフォーマットを列記する。
 - ◇ データの取得プロトコルやAPIについても列記する。
- ii. 識別子に関する規格 63
 - ◇ オープンデータが増加するにしたがって、それを識別する仕組みが重要になる。
 - ◇ 国際的に広く利用されている識別子のうち、URI化できる規格を列記する。
- iii. オープンデータ化に有用なツール 67
 - ◇ オープンデータを扱う上で有用なツールについて、主要な機能と取得先を示す。



i. ファイル形式・データフォーマットに関する規格

表形式データに関する規格

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 など
★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(*) など
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 など
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およ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 など

(*) 現バージョンでは、加工はできるがこの形式で保存できない。

文書形式データに関する規格

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 など
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・大きさや太文字などの装飾指定、画像の表示や中央揃え・箇条書き、表などの簡易レイアウトを行える特徴がある。	Microsoft Word Openoffice など
Portable Document Format (.pdf)	Adobe Systems	ISO 32000-1	アドビシステムズが開発および提唱する、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 など

文書形式データに関する規格

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
★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(*) など
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 など
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 など
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。	

文書形式データに関する規格

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
★4				
RDFa (Resource Description Framework in Attributes)	World Wide Web Consortium (W3C)		メタデータをXHTMLで書かれた構造化文書に埋め込む仕様。	

地理情報データに関する規格

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。

■ これらの規格を扱えるツールについては、p.65に列記する。

データの伝送プロトコル・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・動画など）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など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

リアルタイムデータの伝送に関する規格

■ Streams API(*)

- ▶ 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。
- ▶ World Wide Web Consortium (W3C) が規格化している。
- ▶ Twitterなどで利用されている。

■ GTFS (General Transit Feed Spec) Realtime

- ▶ 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。
- ▶ 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。
- ▶ 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。

(*) Feras Moussa. Streams API. 2013. <http://www.w3.org/TR/streams-api/>



ii. 識別子に関する規格

識別子に関する規格

種類	規格名	URI表現例	運営主体	説明	長さ	永続性	個体識別	RFID格納
汎用	ucode [ITU-T H.642.1]	urn:ucode:_ 0123456789 ABCDEF012 3456789AB CDEF	ユビキタスID センター	モノ・場所・概念など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:s gtin:451234 5.167890.2 urn:epc:tag: sgtin- 96:2.45123 45.167890. 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	△	○	○
電子 データ	DoI (Digital Object Identifiers) [ISO 26234]	http://dx.do i.org/10.102 1/jo034922 7	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。学術論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:f81 d4fae-7dec- 11d0-a765- 00a0c91e6b f6	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	×	○	△

識別子に関する規格

種類ゆ	規格名	URI表現例	運営主体	説明	長さ	永続性	個体識別	RFID格納
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など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	×	×	×
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)	http://statd.b.nstac.go.jp/lod/sac/13101	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	○(*)	×	△

(*) コード自体は自治体の統廃合により変化することがあるが、その変化を追跡する仕組みが提供されている。

識別子に関する規格

種類	規格名	URI表現例	運営主体	説明	長さ	永続性	個体識別	RFID格納
書籍	ISBN [ISO 2108]	urn:isbn:4-13-060800-2	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桁または13桁	×	×	△
	ISSN [ISO 3279]	urn:issn:1560-1560	ISSNネットワーク/国内では国会図書館	学術雑誌など、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ためのユーザ識別子)。	可変長	×	×	×



iii. オープンデータ化に有用なツール

Webサーバ

■ 概要

- ▶ 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、およびそれを提供するコンピュータをいう。

■ 代表的なツールとその入手先

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

■ 補足

- ▶ レンタルサーバサービスのほとんどで、Webサーバの機能を提供している。

CKAN

■ 概要

- ▶ 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

■ ツールの入手先

ツール名	開発・提供元	入手先
CKAN (無償)	The Open Knowledge Foundation	http://www.ckan.org/

GISシステム

■ 概要

- ▶ 地理情報形式データの作成・編集を行うソフトウェア

■ 代表的な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	http://qgis.org/
Google Earth（無償）	Google	http://www.google.co.jp/earth/
GRASS GIS（無償）	GRASAS Development Team	http://grass.osgeo.org/
ArcGIS（有償）	ESRI	http://www.esri.com/products/arcgis/

情報流通連携基盤

■ 概要

- ▶ 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。

■ 外部仕様の取得先

- ▶ 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」が公開されている。
◇<http://www.opendata.gr.jp/cfc/>

RDFレポジトリ

■ 概要

- ▶ 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

■ 代表的なツールとその入手先

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

[補足] CKAN (Comprehensive Knowledge Archive Network)

■ 概要

- ▶ webベースのデータ管理・配信システム
- ▶ Open Knowledge Foundationが提供（無償）

■ 利用事例

- ▶ 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
 - ◇ data.gov (米国) / data.go.uk (英国) / publicdata.eu (EU) / data.gov.au (オーストラリア)
 - ◇ data.go.jp (政府オープンデータポータルサイト) / datameti.go.jp (Open DATA METI)

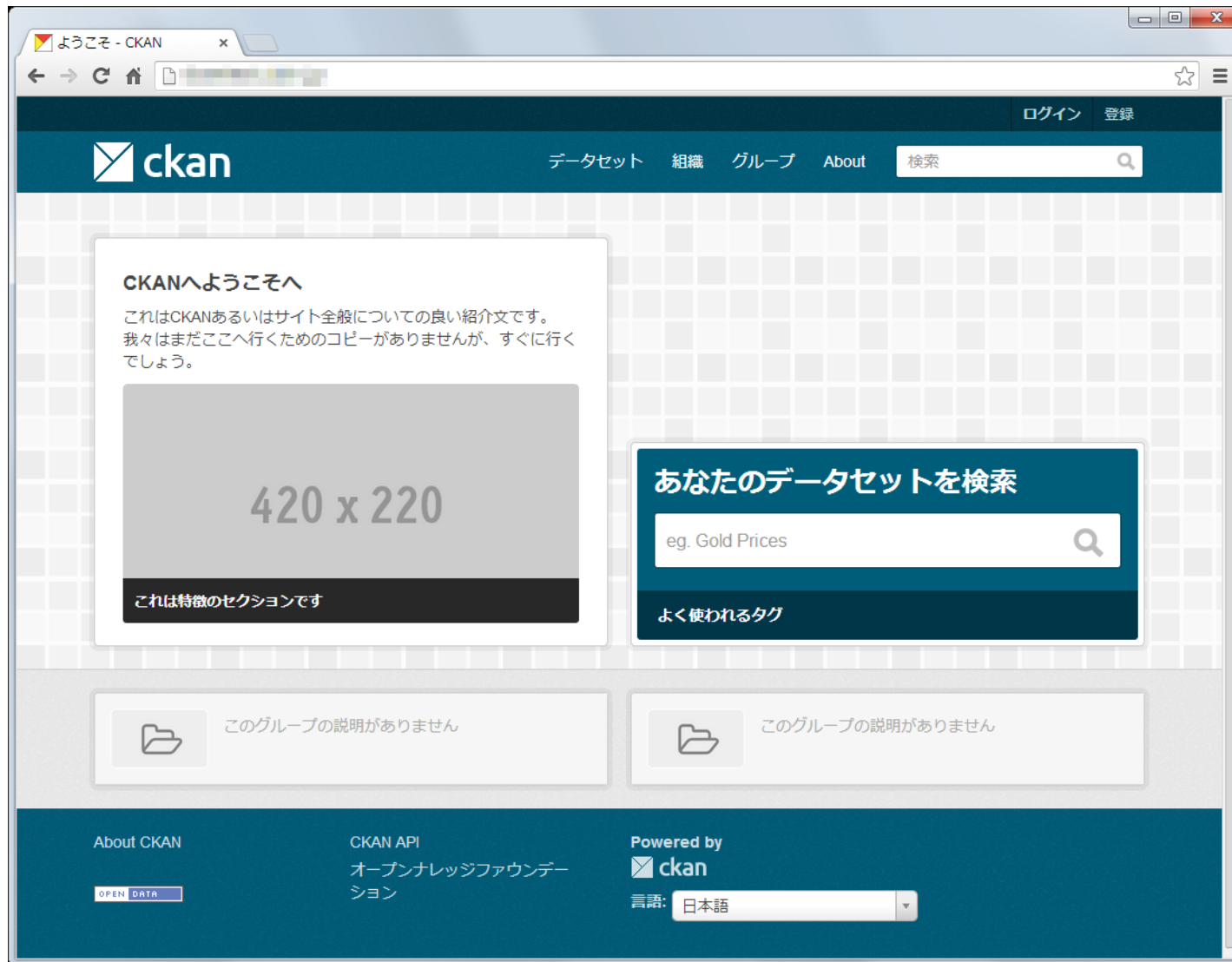
など

■ 入手先・インストール方法

- ▶ 公式サイト <http://ckan.org/>
- ▶ 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/>
- ▶ 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>

※以下CKAN 2.0をベースに解説

初期画面



CKANで使われている用語の解説

■ 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。
- ▶ たとえば「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。
- ▶ たとえば「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。
- ▶ たとえば「財政」「測量」「交通」など。
- ▶ データを検索するためのキーになる。

CKANの運用前に検討・準備すべき事項

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2章末の手順フローを参照のこと。

2. オープンデータを管理するポリシーの策定

▶ データセット・組織の決め方

◇データセットと組織の関係

◆ 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する。

◇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。

▶ グループ・タグの決め方

◇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、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。

▶ データ提供ライセンスの選定

◇それぞれのオープンデータに対して適用すべきライセンスを選定する。

◆ 詳細は「行政職員向けの利用ルール案の解説・FAQ」を参照されたい。

▶ 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。

◆ 組織ごとに、管理者ユーザを定める。

▶ CKANシステムに関する留意点

◇CKAN 2.0の初期状態では、誰でもデータを登録するためのユーザ登録ができる。

◇一方、data.govやdata.go.uk、data.go.jp等ではwebページからユーザ登録できないようにしている。

◇このように、webページからユーザ登録できないようにするためには、システムのソースコードを書き換える必要がある。

CKANの運用前に検討・準備すべき事項

3. (必要であれば) 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。
- ▶ 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ておいた方がよい。

4. データの整備計画

- ▶ 本書第4章「オープンデータ化のための技術的指針」に基づき、リストアップしたデータを、機械可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、たとえ機械可読性が低くてもオープンなライセンスでなくても、公開されることが望ましい。
 - ◆「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」という、いわゆるスモール・スタートの考え方が基本原則とされている。
 - ◇機械可読性の高いデータやオープンなライセンスが整備でき次第、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。

CKANにデータを登録してみよう

■ CKANにデータを登録するまでの流れ

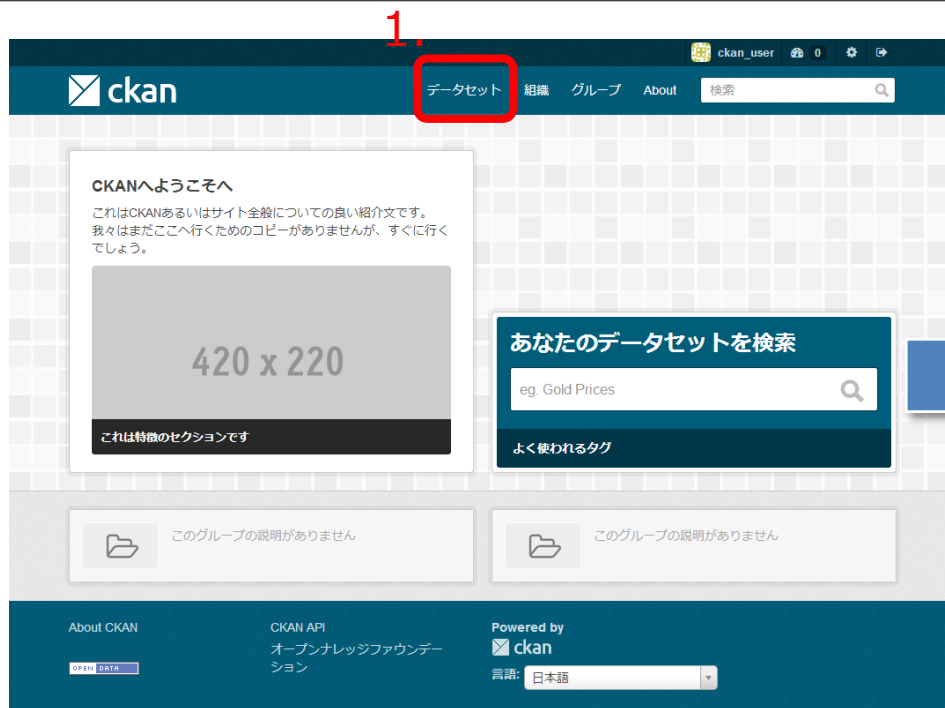
1. アカウントを登録する
2. データセットを作成する
3. 作成したデータセットにデータを登録する
◇最初のデータは、データセットを作成する時に登録できる。

Step 1. アカウント登録



1. 初期画面の右上にある「登録」をクリックする。
 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。
- 登録が完了すると、ログインされた状態になる。

Step 2. データセットの作成



1. 初期画面（ログイン後の画面）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。

Step 2. データセットの作成

ckan / データセット / データセットの作成

1 データセットとは? 2 データの追加 3 追加データ

データセットとは?
データセットはいくつかのデータを単にグループ化するために使われます。これらは記述とライセンス情報がある単一のURLで見つけられます。

1 データセットの作成

タイトル: eg. A descriptive title

URL: ckantest.ubin.jp/dataset/<dataset> **3** **編集**

説明: eg. Some useful notes about the data

ここではマークダウン形式を使えます

タグ: 例: 経済, 政府

ライセンス: ライセンスが指定されてい...

重要: コンテンツを提出することで、あなたはオープンデータベースライセンスのもとであなたの貢献を提供することに同意したことになります。

キャンセル Next: データの追加

ckan / データセット / データセットの作成

1 データセットとは? 2 データの追加 3 追加データ

データセットとは?
データセットはいくつかのデータを単にグループ化するために使われます。これらは記述とライセンス情報がある単一のURLで見つけられます。

1 データセットの作成

4 **タイトル:** 〇〇地区の気温

URL: /dataset/ temperature

説明: 〇〇地区の気温を、1ヶ月間測定してみました。

ここではマークダウン形式を使えます

タグ: x 天候

ライセンス: クリエイティブ・コモンズ...

重要: コンテンツを提出することで、あなたはオープンデータベースライセンスのもとであなたの貢献を提供することに同意したことになります。

5. **Next: データの追加**

キャンセル

3. タイトル欄の下にある「編集」ボタンを押す。
4. タイトル・データの説明・タグ・ライセンスを入力する。
 - ▶ URLのみ必須
 - ▶ タイトルが英数字のみの場合は、3.の処理は不要
 - ▶ 「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
 - ▶ ライセンスは、オープンなライセンスを選択することを強く推奨する。
5. 「Next: データの追加」ボタンを押す。

Step 3. データの登録

ckan_user 0

データセット 組織 グループ About 検索

ホーム / データセット / データセットの作成

1 データセットの作成 2 データの追加 3 追加データ

6. リソースとは?
リソースはファイルまたはデータを参照するリンクです

ファイルへのリンク APIへのリンク ファイルをアップロード

リソース:

名前:

説明:

ここではマークダウン形式を使えます

データ形式:

これは自動で生成されました。後から編集可能です

Previous **保存して他を追加** Next: 追加情報

About CKAN CKAN API オープンナレッジファウンダー Powered by ckan

ckan_user 0

データセット 組織 グループ About 検索

ホーム / データセット / データセットの作成

1 データセットの作成 2 データの追加 3 追加データ

8. データセットとは?
データセットはいくつかのデータを単にグループ化するために使われます。これらは記述とライセンス情報がある単一のURLで見つけられます。

作成者:

作成者のemail:

メンテナ:

メンテナのemail:

カスタムフィールド:

カスタムフィールド:

カスタムフィールド:

グループを追加:

Previous **完了**

9.

6. 登録するデータと、それを説明する情報を登録する。
 - ▶ リソースのみ必須
 - ▶ 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
 - ▶ 「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
7. 「Next: 追加情報」ボタンを押す。
 - ▶ 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
 - ▶ 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可能
9. 「完了」ボタンを押す

データ登録の結果

ckan_user 0

ckan データセット 組織 グループ About 検索

ホーム / データセット / ○○地区の気温

ソーシャル

- Google+
- Twitter
- Facebook

ライセンス

クリエイティブ・コモンズ表示 [OPEN DATA](#)

データセット アクティビティストリーム 関連 [編集](#) [フォロー](#)

○○地区の気温

○○地区の気温を、1ヶ月間測定してみました。

データとリソース

[CSV](#) **2014/01/27の○○地区気温** [探索](#)

2014年01月27日の、○○地区気温測定結果です。

天候

追加情報

フィールド	値
作成者	CKAN user
状態	active

About CKAN CKAN API Powered by [ckan](#)
 オープンナレッジファウンダー

組織の登録

The image shows two screenshots of the CKAN website interface. The first screenshot, labeled '1.', shows the main navigation menu with '組織' (Organizations) highlighted in a red box. Below the menu is a search bar and a 'あなたのデータセットを検索' (Search your datasets) section. The second screenshot, labeled '2.', shows the '組織' (Organizations) page with the '組織を追加' (Add Organization) button highlighted in a red box. A blue arrow points from the first screenshot to the second, and another blue arrow labeled '続く' (Continue) points to the right.

1. 初期画面（ログイン後の画面）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。

組織の登録

ckan_user 0

データセット 組織 グループ About 検索

組織の作成

組織について

組織はデータセットのための公開部署のように働きます (例えば保健省)。これはつまりデータセットが個々のユーザではなく、部署によって、その部署に所属するように公開され得るということです。

組織内では、管理者が組織のメンバーに役割や認可を割り当てます。それにより、特定の組織 (例: 国家統計局) からのデータセットを公開する権限を個々のユーザに与えます。

情報

組織の作成

タイトル: 私の組織

URL: ckantest.ubin.jp/organization/<input type='text' value='organization' /> **3**

説明: 私の組織についての簡単な情報

ここではマークダウン形式を使えます

画像URL: http://example.com/my-image.jpg

About CKAN CKAN API オープンナレッジファウンデーション Powered by ckan 言語: 日本語

ckan_user 0

データセット 組織 グループ About 検索

組織の作成

組織について

組織はデータセットのための公開部署のように働きます (例えば保健省)。これはつまりデータセットが個々のユーザではなく、部署によって、その部署に所属するように公開され得るということです。

組織内では、管理者が組織のメンバーに役割や認可を割り当てます。それにより、特定の組織 (例: 国家統計局) からのデータセットを公開する権限を個々のユーザに与えます。

情報

4 組織の作成

タイトル: oo課

URL: /organization/division-a

説明: oo課は、ooに関する情報を公開しています。

ここではマークダウン形式を使えます

画像URL: http://example.com/my-image.jpg

5

About CKAN CKAN API オープンナレッジファウンデーション Powered by ckan

3. タイトル欄の下にある「編集」ボタンを押す。
4. タイトル・URL・説明・画像URLを入力する。
 - ▶ URLのみ必須
 - ▶ タイトルが英数字のみの場合は、3.の処理は不要
 - ▶ URL欄には、タイトルとほぼ同じ意味の英数字を入力
5. 「組織の作成」ボタンを押す。

組織の登録結果



- このページにある「データセットの追加」ボタンを押すと、この組織に属するデータセットを追加できる。
- データセットの追加手順は、前述のStep 2～Step 3と同じ。
 - ▶ 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。

組織へのメンバ追加

The image shows two screenshots of the CKAN web interface. The first screenshot shows the organization page with the '管理者' (Admin) button highlighted by a red box and labeled '1'. A blue arrow points to the second screenshot, which shows the 'メンバー' (Members) tab highlighted by a red box and labeled '2'. In the second screenshot, a table lists the user 'ckan_user' with the role '管理者'. A red box highlights the 'メンバーの追加' (Add member) button, labeled '3'. A blue arrow labeled '続く' (Continue) points to the right.

■ 組織を作成したユーザが、組織の管理者になる。追加手順は以下の通り。

1. 組織ページにある「管理者」ボタンを押す。
2. 「メンバ」タブを押す。
3. 「メンバの追加」ボタンを押す。

組織へのメンバ追加

ckan ユーザー 0

データセット 組織 グループ About 検索

組織 / 〇〇課 / 管理者

メンバーの追加

4. ユーザ: ユーザ名
div
div
diva-user

追加

ckan ユーザー 0

データセット 組織 グループ About 検索

組織 / 〇〇課 / 管理者

メンバーの追加

5. ユーザ: diva-user
6. ロール: 編集者

追加

4. ユーザの入力欄をクリックしてユーザ名を入力
 - ▶ 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。
 - ▶ ロールは「管理者」「編集者」「メンバ」の3種類。画面左側に説明あり。
6. 「追加」ボタンを押す。

CKANの組織機能を利用したアクセス制御

このスクリーンショットは、CKANのデータセット編集画面を示しています。左側のナビゲーションメニューには「データセットの編集」が選択されています。中央のフォームには、タイトル「〇〇課予算」、URL「ckantest.ubin.jp/dataset/diva-budget」、説明「〇〇課の予算状況を公開します。」、タグ「x 予算」、ライセンス「クリエイティブ・コモンズ」、公開・非公開「プライベート」（赤い枠で囲まれている）、作成者「ckan_user」、作成者のemail「ckan@example.org」が設定されています。

このスクリーンショットは、CKANの組織内のデータセット検索結果を示しています。検索条件は「〇〇課」で、結果として「1件のデータセットが見つかりました」と表示されています。検索結果には「〇〇課」のデータセットがリストアップされています。

←組織内ユーザ

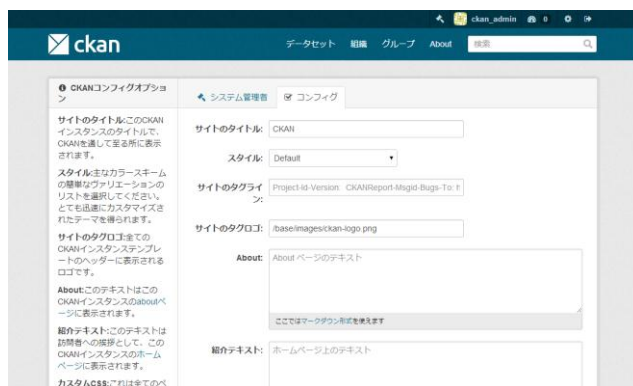
このスクリーンショットは、CKANの組織内のデータセット検索結果を示しています。検索条件は「〇〇課」で、結果として「データセットが見つかりません」と表示されています。これは、組織外からの検索ではプライベートデータセットは見つからないことを示しています。

組織外ユーザ→

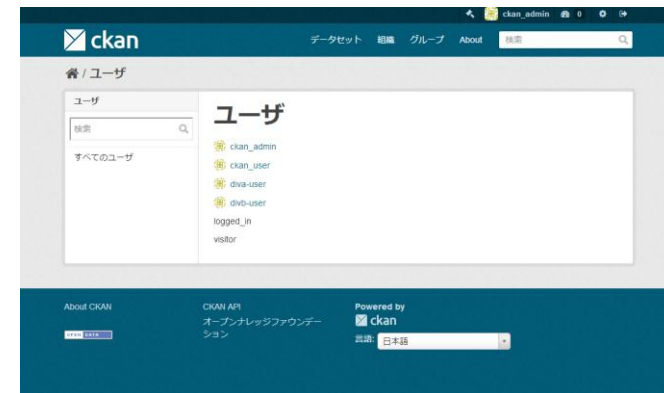
- データセットの登録・編集画面に「公開・非公開」の選択肢がある。
 - ▶ パブリック（公開）またはプライベート（非公開）を選択できる。
- プライベート（非公開）のデータセットは、組織内ユーザのみ閲覧できる。
 - ▶ この機能は、データの公開前準備等の用途で利用できる。

CKANの管理ページ

- ポータルサイトの管理を行うページで、管理者のみが利用できる。
 - ▶ 管理者の設定は、サーバ上でコンソールからコマンドを発行する必要がある。
 - ▶ 管理者は、ポータル内全ユーザ・データセット・データを閲覧・編集・削除できる。
- webページから利用できる機能
 - ▶ サイトの見栄え（Look & Feel）の変更
 - ▶ データセットが所属する組織の変更
 - ▶ データセットの削除
 - ▶ ユーザ管理



見栄えの変更画面

データセットの編集画面
(ここで所属組織を変更できる)

ユーザ管理画面



OPEN DATA

オープンデータ流通推進コンソーシアム