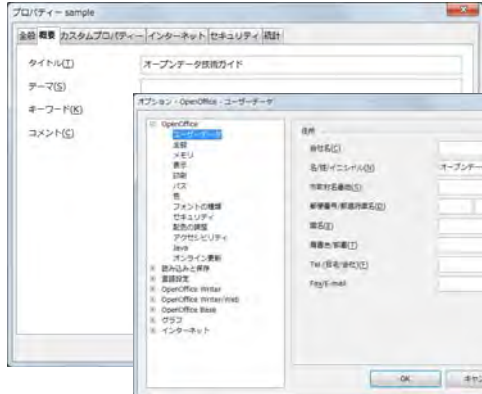
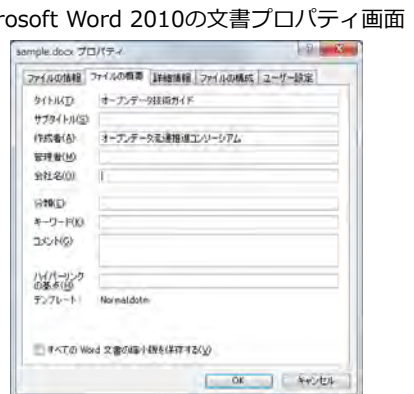


## 8.5 オープンデータの管理ポリシーとメタデータの付与方法

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




Open Office 4 Writerの  
文書プロパティ画面




Microsoft Word 2010の文書プロパティ画面

この文書形式データを  
Apache Tikaで解析



この文書形式データを  
Apache Tikaで解析



作成者→  
タイトル→  
作成日時→  
最終更新日時→

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

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

55

## 第9章 オープンデータのための技術的指針

### ■ 本章の概要

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

### ■ 本章の構成

1. 識別子に関する指針
  - ◇ オープンデータを識別する識別子が満たすべき性質と、それを満たすための手法を解説する。
2. ファイル形式に関する指針
  - ◇ 公開するデータのファイル形式は、機械判読性の高い形式を利用することが望ましい。
  - ◇ オープンデータの技術レベルに基づいて、代表的なファイル形式を整理して示す。
3. データに関する指針
  - ◇ 表形式データ・文書形式データ・地理空間情報・リアルタイムデータのそれぞれについて、機械判読性の高いデータを作成・編集する際の指針を示す。

56

## 9.1 識別子に関する指針

### ■ オープンデータにとっての識別子が満たすべき性質

1. ユニークであること。
2. 共通に利用できる体系であること。

### ■ 利用可能な識別子体系

- ▶ グローバルにユニークな識別子体系
- ▶ 公的機関が定める識別子体系・コード体系
- ▶ URI (Uniform Resource Identifier) として表現できる体系

### ■ 適切な識別子体系がない場合の対処法

1. 対象とする実物や組織・場所に番号が付与されていない場合は、まずそれらに番号を付与する。
2. 識別子のユニーク範囲を拡大する。
  - ◇ ucodeやDoIなどのグローバルな体系や、公的機関が定める識別子体系・コード体系に基づく識別子を取得し、管理する。
  - ◇ 付与した番号に組織が決めるURLを付与してグローバル化することもできる。
    - ◆ ただし、組織の統廃合等によりドメイン名が変わると、識別子も変わってしまうことに注意。

## 9.2 ファイル形式に関する指針

### ■ 基本方針

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

|           | Level 1                 | Level 2/3   | Level 4  |
|-----------|-------------------------|---|--|
| 表形式データ    | xls (Microsoft Excel形式) | CSV<br>xlsx (Office Open XML)<br>ods (OpenDocument)<br>JSON | RDF/XML<br>RDF/JSON, JSON-LD<br>Notation3<br>Turtle等のRDF形式 |
| 文書形式データ   | doc (Microsoft Word形式)  | HTML<br>XML<br>docx (Office Open XML)<br>odt (OpenDocument) |  |
| 地理空間情報    | shape                   | KML<br>GML  | 推奨するファイル形式   |
| リアルタイムデータ |                         | (ファイルの形で交換しない)  |  |

## 9.3 データに関する指針

### ■ 指針のグレード: 満たすべき指針の重要度にあわせて2つのグレードを設ける。

#### ▶ グレード1

◇グレード1は、オープンデータが満たすことを強く推奨する指針であり、以下を満たすことを目的とする。

- ◆ データ形式に関する標準的な規格がある場合は、それに矛盾しないこと。
- ◆ データを取得した利用者が、データ本体の中身を修正したり手を加えたりすることなく、そのデータの本質的内容を正しく解釈するためのプログラムを書けること。

#### ▶ グレード2

◇グレード2は、オープンデータが満たすことを推奨する指針であり、以下を満たすことを目的とする。

- ◆ データを取得したプログラムが、そのデータの項目や構造を正しく解釈できること。

### ■ 対象とするデータ

- ▶ 表形式データ
- ▶ 文書データ
- ▶ 地理空間情報
- ▶ リアルタイムデータ

### ■ 各指針に関する記述内容

- ▶ 表形式データを中心に指針を満たさない例と満たす例を明記し、それに対して解説する。
- ▶ 以下、各データに関する指針のみを記す。

## 9.3 データに関する指針／表形式データ

| グレード | 指針   |  |
|------|------|--|
| 1    | 指針1  | 1つのファイルは、1種類の表から構成されるべきである。  |
|      | 指針2  | ヘッダは、1行から構成されるべきである。   |
| 2    | 指針3  | データでない情報を、レコードに含めないことが望ましい。  |
|      | 指針4  | 全てのフィールドは、他のフィールドと結合されないことが望ましい。   |
|      | 指針5  | 値がない場合を除き、フィールドを空白にしない（省略しない）ことが望ましい。  |
|      | 指針6  | 年の値には、西暦表記を備えることが望ましい。   |
|      | 指針7  | フィールドの単位が明記されていることが望ましい。   |
|      | 指針8  | 利用している文字コードを明記することが望ましい。また、国際的に広く利用されている文字コードを利用することが望ましい。                           |
|      | 指針9  | ファイルの属性や説明を表すメタデータが、フォーマルに記述されていることが望ましい。また、そのメタデータからデータセット本体へリンクし、たどれるようにすることが望ましい。 |
|      | 指針10 | データ本体を、XMLやRDFの形式を使ってフォーマルに記述することが望ましい。  |

## 9.3 データに関する指針／文書データ

| グレード | 指針  |   |
|------|-----|---|
| 1    |     | (なし)  |
| 2    | 指針1 | 文章に存在する部・章・節・図表などの構造が、機械判読性の高いフォーマットで記述されていることが望ましい。                                |
|      | 指針2 | 文章内に、整形のための符号や文字（空白、改行等）を含めないことが望ましい。   |
|      | 指針3 | 文書形式データが表形式データを含む場合、グレード1以上の表形式データが添付されていることが望ましい。                                  |
|      | 指針4 | テキスト形式の文書形式データを利用している場合は、利用している文字コードを明記することが望ましい。また、国際的に広く利用されている文字コードを利用することが望ましい。 |
|      | 指針5 | 文章に対する、情報利用者が理解できるような説明が、メタデータとして記述され、当該文書にリンクされていることが望ましい。                         |

## 9.3 データに関する指針／地理空間情報

| グレード | 指針  |  |
|------|-----|--|
| 1    | 指針1 | 位置情報に関するデータを付与する場合は、緯度・経度等の位置情報に加えて、測地系が明記されるべきである。屋外であれば、世界測地系を利用することが望ましい。屋内であれば、座標系と描画縮尺（入力精度）を示すべきである。 |
| 2    | 指針2 | 地理空間情報は、ベクタ形式に依るものが望ましい。ベクタ形式のデータの作成に当たっては、最新の ISO 規格及び JIS 規格に基づいた地理空間情報標準プロファイル (JPGIS) を用いる。            |
|      | 指針3 | 地理空間情報に対する、情報利用者が理解できるような説明が、メタデータとして記述され、当該文書にリンクされていることが望ましい。  |

## 9.3 データに関する指針／リアルタイムデータ

| グレード | 指針  |   |
|------|-----|---|
| 1    | 指針1 | データの取得仕様が明記されているべきである。                                |
|      | 指針2 | 表形式データや地理空間情報をファイル形式で取得させる場合は、それぞれのグレード1の指針を満たすべきである。 |
| 2    | 指針3 | リアルタイムデータの最新値・差分を取得する手法が提供されていることが望ましい。               |



# 付録

## 第10章（付録）オープンデータに関する規格・ツール

### ■ 本章の概要

- ▶ 機械判読に適したオープンデータを作成・編集する上で参考となる規格やツールをまとめる。

### ■ 本章の構成

1. データフォーマットに関する規格
  - ◇ 表形式データ・文書データ・地理空間情報のそれぞれのファイル形式に関する代表的な規格と、その形式を扱える代表的なソフトウェアを挙げる。
2. 識別子に関する規格
  - ◇ オープンデータの識別に利用できる識別子に関する規格を挙げる。
3. オープンデータに有用なツール
  - ◇ オープンデータを作成・編集・公開するために有用なツールを解説する。
    - ◆ Webサービス
    - ◆ データカタログシステム
    - ◆ GISシステム
    - ◆ 情報流通連携基盤
    - ◆ RDFレポジトリ

## 第11章（付録）データカタログシステムCKAN

### ■ 本章の概要

- ▶ オープンデータの技術レベルのうち、データカタログのLevel3を満たすためのデータカタログシステムとして広く利用されている、CKANを解説する。

### ■ 本章の構成

1. CKAN概説
2. CKANの運用前に検討・準備すべき事項
  - ◇ 公開するオープンデータの洗い出し
  - ◇ オープンデータを管理するポリシーの策定
  - ◇ 要求仕様の策定
  - ◇ データの整備計画
3. CKANを用いたオープンデータ登録例
  - ◇ アカウント登録
  - ◇ 組織の登録
  - ◇ 組織へのメンバ追加
  - ◇ データセットの作成
  - ◇ データの登録
  - ◇ CKANの管理ページ

