

補足：公開されたデータの悪用とその責任について

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

45

第7章 利用ルールに関する今後の見直しの方向性

- 公共データをオープンデータとして公開する場合、情報利用者視点に立ち、基本的には、国際的にオープンデータの利用ルールとして広く使用されているCC-BY又はCC0を適用することが望ましい。
- 情報提供者に配慮し、公序良俗に反する利用等の禁止事項を盛り込むことが、できることから速やかに着手するというスモール・スタートの原則にかなう場合や、できるだけ多くのデータをオープンデータにする対象としたいといった場合には、次善策として、政府標準利用規約（第1.0版）を適用することが考えられる。
 - ▶ 特に、「法令・条例・公序良俗に反する利用」と「国家・国民の安全に脅威を与える利用」という規定は、その具体的な利用態様が必ずしも明確ではなく、公開されたデータの利用に際して萎縮効果を生む可能性もあり、データガバナンス委員会としては、政府標準利用規約（第1.0版）の運用状況を注視し、これらの禁止条項を削除しても問題がないと判断できる場合には、当該禁止条項の削除又はCC-BYやCC0への移行を視野に入れて検討すべきと考える。
 - ▶ 政府標準利用規約（第1.0版）は、利用ルールの政府標準利用規約（第1.0版）への変更後のコンテンツの利用状況等を踏まえ、見直しの検討が行われる予定となっている。国以外において、政府標準利用規約（第1.0版）を適用する際には、今後見直しが行われる可能性があることを理解した上で、適用することが望ましい。
- オープンデータとして公開されたデータの不適切な利用によって第三者等に何らかの問題が起きた場合、その責任はデータを不適切に利用した情報利用者にあるのであって、情報提供者である国、地方公共団体等が責任を負うものではないということを啓発していく必要がある。
 - ▶ オープンデータによって公開された情報には、誤りがある場合等もあると想定されるが、その誤りが原因で、情報利用者や第三者等に何らかの問題が起きた場合についても、CCライセンス、政府標準利用規約（第1.0版）のいずれにおいても無保証で公開しているのであり、国、地方公共団体等が責任を負うものではない

46

第III部 技術編： 機械判読に適したデータにし よう

第8章 オープンデータの技術レベル

■ 本章の概要

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

■ 本章の構成

1. 機械判読性に関する解説
 - ◇機械判読に適したデータが必要である理由と、機械判読性に関する指標について解説する。
2. データカタログに関する解説
 - ◇データカタログの意義を解説する。
3. オープンデータと識別子
 - ◇オープンデータに識別子が必要である理由を解説する。
4. オープンデータの技術レベル
 - ◇上記3項目をもとに、オープンデータの技術レベルを定める。
5. オープンデータの管理ポリシーとメタデータの付与方法
 - ◇オープンデータを登録・管理する際に、メタデータを自動的に付与することができるならば、管理・登録のコストを軽減できる。そのような手法について解説する。

8.1 機械判読性に関する解説

■ 機械判読に適したデータとは？

- ▶ コンピュータプログラムがデータの論理的な構造を判読でき、構造中の値（表の中に入っている数値、テキスト等）を自動的に編集・加工・改変等できること。“Machine Readable”の日本語訳であり「機械可読」ともいう。
- ▶ 機械判読性の高いデータを提供することにより、コンピュータの解析に必要な情報利用者のコストを軽減できる。
- ▶ 機械判読に適さないデータ：PDFや画像データ
 - ◇このデータをコンピュータに与えて、解析させるためには、以下の作業が必要。
 - ◆ 事前に人間がその画像にあるデータを表計算ソフトウェアに入力して保存する。
 - ◆ 画像認識等の技術により公開されているデータから数値やテキストを得て、それをコンピュータに与える。
 - ◇これらは情報利用者に負担を求める方法であり、効率的でない。

■ 機械判読性に適したデータに関する留意点

- ▶ 機械判読に適したデータは、必ずしも人が読みやすいとは限らない。
 - ◇必要であれば、機械判読に適した形式と人に読みやすい形式の2種類のファイルを用意して公開することも考慮すべき。

8.1 機械判読性に関する解説

■ 機械判読性に関する指標：5★Open Data

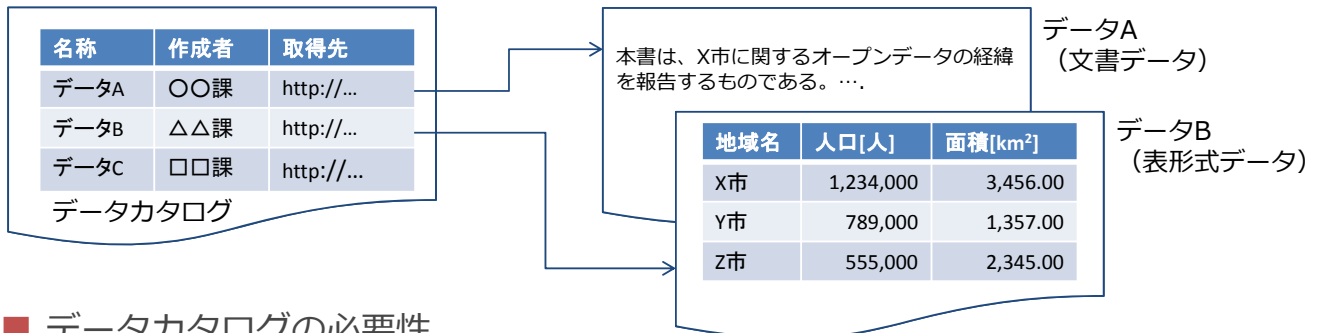
- ★1: この形式のファイルからコンピュータがデータを取り出すためには画像解析等の技術が必要であり、これは容易ではない。
- ★2: この形式のファイルは構造化されているため、対応するソフトウェアを用意すれば、コンピュータはこれからデータを抽出できる。
 - ◇一般に「機械判読性のあるデータ」とは★2以上のデータをいう。
- ★3: この形式のデータの解析方法は公開されているため、★3の形式のデータを解析するためのソフトウェアを構築することは、★2より容易である。
 - ◇この形式のデータに対する機械判読性を高めるための技術的指針について、9.3節で詳しく述べる。
- ★4~: この形式のデータは、相互に接続でき、コンピュータによるデータのマッシュアップが容易になる。



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

8.2 データカタログに関する解説

- 「データ」と「データカタログ」の関係は下図のとおり。



- データカタログの必要性

- ▶ 公開するデータが増加するにつれて、それらのデータを整理し、検索、一覧する機能に対する要求が高まる。このような要求に応えるものが、データカタログである。

- さまざまなデータカタログ

- ▶ データの名称、取得先等を表形式データとしてまとめたものも、一種のデータカタログである。
- ▶ 高機能なデータの管理・検索・一覧機能を提供するためには、データカタログシステムを導入する、あるいはRDF・SPARQL等の技術を利用したデータ・メタデータ検索機能を提供することが望ましい。

8.3 オープンデータと識別子

- 識別子とは

- ▶ データやデータが対象とする実物や組織・場所等をコンピュータに識別させるための番号。

- 識別子が必要である理由

- ▶ オープンデータは、コンピュータが読み取り解釈するためのデータである。これらのデータは、コンピュータが一意に識別できるべきである。
- ▶ オープンデータに含まれる実物や組織、場所等も、一意に識別されることが望ましい。これは、組織や場所が文字列で提供されている場合、以下のような問題が発生する可能性があるためである。
 1. 情報利用者のコンピュータは、表記の揺らぎにより同一の組織や場所を別物として解釈する可能性がある。(例：「中央一丁目一番地一号」と「中央1-1-1」)
 2. 情報利用者のコンピュータは、同一名称だが違う意味である組織や場所を、文字列だけでは識別できない。(例：全国各所にある「中央一丁目」)

- 識別子とコード

- ▶ コードとは
 - ◇カテゴリ化される概念や事物に対して付与される番号。
 - ◇対象とする概念や事物を短縮して符号化するために規定される。
- ▶ 識別子とコードは異なる。
 - ◇コードには意味が付与されているが、識別子に意味が付与されているとは限らない。
- ▶ ただし、多くの場合、コードは識別子として機能する。

8.4 オープンデータの技術レベル

	Level 0	Level 1	Level 2	Level 3	Level 4
データ	PDFや画像ファイルをWebで公開する。	構造化されたデータを作成し、Webで公開する。 (XLS, DOC等)	非独占の（標準化された）形式で公開する。 (CSV等)	機械判読に適したデータを作成し、公開する。 (第9章参照)	RDF、XML等の技術を導入したデータを作成し、APIを実装して公開する。
データカタログ	存在しない。	カタログを表形式データ (CSV等) として作成し、公開する。	Level 1と同じ。	データカタログシステムを導入する。	RDFやSPARQL等を利用したメタデータ検索機能を提供する。
識別子	何らかの手段で識別されている。	Level 0と同じ。	Level 0と同じ。	URLにより識別されている。	グローバルな体系に基づく識別子を利用する。
必要なツール	Webサーバ	Webサーバ	Webサーバ	Webサーバ+データカタログシステム等	Webサーバ+データカタログシステム+情報流通連携基盤等

- データ・データカタログ・識別子のレベルを合わせる必要はない。
- Level1のメリット
 - ▶ 情報利用者は、画像解析等の処理をすることなく、直接データを取得できる。
 - ▶ データのありか等のメタデータを電子的に入手できるようになる。
- Level2のメリット
 - ▶ データに対する機械判読性が高まる。
- Level3のメリット
 - ▶ データの解釈効率や検索性が向上し、情報利用者のデータ利活用の効率が向上する。
- Level4のメリット
 - ▶ 他のデータとの横断検索等も容易になり、情報利用者によるデータ利用の幅が広がる。

8.5 オープンデータの管理ポリシーとメタデータの付与方法

■ オープンデータの登録ポリシーには2通りある

1. 集中登録方式

- ◇システム管理者や、オープンデータにすることをを行う独立した組織が、各組織・部署からデータを集めて公開する手法。
- ◇この方式では、データを集める際にメタデータもまとめて収集することが望ましい。

2. 分散登録方式

- ◇各組織・部署が自ら、何らかのシステムを利用して直接オープンデータを登録・管理する手法。
- ◇この方式では、担当組織がオープンデータを作成する際に、メタデータもまとめて作成できることが望ましい。