
オープンデータ研修予習用

オープンデータの定義

オープンデータは、機械判読に適した形で、二次利用可能なルールで公開される公共データです。

オープンデータの定義としては、さまざまなものがありますが、政府が出している「オープンデータ基本指針」では以下のように定義されています。

国、地方公共団体及び事業者が保有する官民データのうち、国民誰もがインターネット等を通じて容易に利用（加工、編集、再配布等）できるよう、次のいずれの項目にも該当する形で公開されたデータ。
営利目的、非営利目的を問わず**二次利用可能なルール**が適用されたもの
機械判読に適したもの
無償で利用できるもの

出典:オープンデータ基本指針

<https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20170530/kihonsisin.pdf>

※ 公益企業など民間事業者や個人が保有し、二次利用可能な形で公開されるものもオープンデータに含まれます。

1

講義ノート:

オープンデータとは、機械判読に適した形で、二次利用可能なルールの下で公開される公共データです。

オープンデータの定義としては、さまざまなものがありますが、政府が公開している「オープンデータ基本指針」では、以下のすべてを満たす形で公開されたデータ、と定義されています。

第1に、営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの。

第2に、機械判読に適したもの。

第3に、無償で利用できるもの。

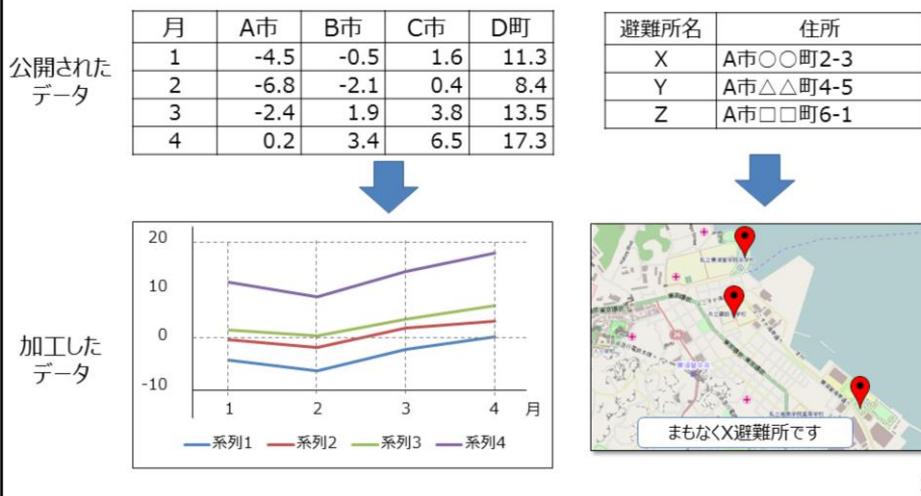
公益企業など民間事業者や個人が保有し、二次利用可能な形で公開されるものも、オープンデータに含まれます。

ここで、「二次利用可能なルール」「機械判読に適した」という言葉が出てきます。

これがどういうことなのか、説明します。

二次利用とは？

二次利用とは、公開されたデータをコピー・加工して利用することをいいます。



講義ノート：

二次利用とは、公開されたデータをコピー・加工して利用することをいいます。

たとえば、公開されたデータをグラフにしたり、分析して傾向をつかんだり、公開されたデータから利用者に通知をするアプリケーションを構築したりします。

二次利用とは？

アプリケーションで利用するためには、**二次利用できる利用ルール**が必要です。

二次利用できない利用ルールの例

当ホームページの内容について、「私的使用のための複製」や「引用」など著作権法上認められた場合を除き、無断で複製・転用することはできません。

二次利用できる利用ルールの例（政府標準利用規約より抜粋）

…カタログ掲載実データ…は、クリエイティブ・コモンズ・ライセンス（以下「CCライセンス」といいます。）の表示4.0 国際…により利用できます。
 なお、数値データ、簡単な表・グラフ等のデータは著作権の対象ではありませんので、…、自由に利用できます。

出典：<http://www.data.go.jp/terms-of-use/terms-of-use/>

3

講義ノート：

このように、公開されたデータから分析結果を得たり、公開されたデータからアプリケーションを構築したりするためには、そのデータが二次利用できる利用ルールで公開されている必要があります。

たとえば、公開されたデータに対して

「私的使用のための複製」や「引用」など著作権法上認められた場合を除き、無断で複製・転用することはできません
 というような制限がある場合、そのデータを利用することができません。

一方、政府標準利用規約にあるように

『データはクリエイティブ・コモンズ・ライセンス（以下「CCライセンス」といいます。）の表示4.0 国際により利用できます。
 数値データ、簡単な表・グラフ等のデータは著作権の対象ではありませんので、自由に利用できます。』
 というような表示がされていれば、そのデータを利用して分析をしたり、アプリケーションを構築したりできます。

機械判読とは？

機械判読に適した形とは、コンピュータが扱いやすい形式です。

機械判読に適した形のデータは、アプリケーションから加工・利用しやすくなります。

機械判読の難しい例

年	月	A市	B市	C市	D町
2018	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

人間は、この表をみて、2018年の4ヶ月分のデータが掲載されていることが分かりますが、これをコンピュータは簡単に解釈できません。

機械判読に適した例

```
年,月,A市,B市,C市,D町
2018,1,-4.5,-0.5,1.6,11.3
2018,2,-6.8,-2.1,0.4,8.4
2018,3,-2.4,1.9,3.8,13.5
2018,4,0.2,3.4,6.5,17.3
```

表を構成するすべての箇所にデータがあり、そのデータはカンマで区切られています。このようなデータは、コンピュータが簡単に解釈できます。

4

講義ノート：

機械判読に適した形、とは、コンピュータが扱いやすい形式です。

オープンデータは、アプリケーションが利用するなど、コンピュータが分析・解析して利用するケースが多いです。

このため、データがコンピュータが分析・解析しやすい形であることが望ましいです。

機械判読に適した形は、必ずしも人間が見やすい形と限りません。

たとえば、我々が左側の表を見れば、2018年の4ヶ月分のデータが掲載されていることが分かります。

しかし、これをコンピュータが解読するためには、左端の2018というデータがどこまでかかっているのかを理解する必要があります。

一方、右側の表は、人間が見やすい形ではありませんが、表を構成するすべての箇所にデータがあり、そのデータはカンマで区切られています。

このようなデータは、コンピュータが簡単に解釈できます。

END

オープンデータ研修予習用