

職員の疑問に答える

# オープンデータの基礎知識

**Q** 今もホームページに情報を公開していません。オープンデータとは何が違うのでしょうか？

オープンデータとは、単にデータを公開するという意味ではなく、情報を転用や加工など、様々な二次利用しやすい状態で提供するものです。

二次利用しやすいとは、技術的な条件として、情報を機械判読や処理・加工が容易なデータの構造・形式で公開することであり、利用条件として、著作権や商用利用の可否などを整理し、明示して公開することです。

**Q** 行政の情報を全てオープンデータとして公開するのでしょうか？

行政はすでに多くの情報を公開しています。しかし、利用しやすい状態で公開しているかという点、まだ不十分です。

既に公開している情報をオープンデータの形式にする

には手間がかかりますので、すぐに全てというのは現実的ではありません。まずは、ホームページに掲載している情報のうち、必要性の高い情報などから順次整備していくこととなります。

**Q** 機械で判読が可能にするとは、具体的にどのようなのですか？

コンピュータが機械的に読み取り、処理しやすくするには、データの構造を整え、さらに機械判読に適したデータ形式に変換する必要があります。データ構造に関しては、文書形式のデータの場合は、部、章、節、段落などを見出しを明示して文章の構造を判読しやすくすることなどであり、表形式のデータの場合は、数値に位取りのカンマなどを加えないことや、西暦で統一すること、セルの結合はしないなど、必要な整理を行います。データ形式に関しては、特定のアプリケーションに依存しない、仕様が公開さ

れ標準化された形式とします。(文書形式はXML、HTMLなど、表形式はCSV、XMLなど)

このような機械判読に適したデータ作成のために留意すべき内容は、25年6月のCIO連絡会議で決定された「二次利用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)」の別紙に詳細が示されています。

**Q** 個人情報の保護は大丈夫でしょうか？

個人情報の保護は大切なことであり、行政への信頼を維持するためにも重要です。これは、オープンデータに限らず、個人情報を取り扱う業務において常に求められていることです。横浜市のオープンデータの当面の取組は、ホームページに掲載する情報をオープンデータ化するものから個人情報流出する危険性はありますが、十分に注意して進めていくことが必要です。

**Q** オープンデータとして整備するのは手間がかかりませんか。通常のホームページで公開することに加えオープンデータとしても公開するのは、たいへんではないでしょうか。

ホームページ用のデータとオープンデータ用のデータをそれぞれ作成・管理するのは、確かに大変です。しかし、データの中身は同じであるので、情報は一つのデータベースに格納することとし、所管部署が、データベースの情報を更新すれば、ホームページにもオープンデータとしても反映する仕組みにするよう、平成26年度予算案に本市ウェブサイトのオープンデータ化に向けた再構築事業を計上しています。

情報は、探しやすいように分類・整理し、また、利用しやすいように整備しておくことで、容易に活用できるようになります。ウェブサイトの再構築に伴い、全庁共通のデータベースが整備され、庁内においても情報の活用が促進されると見込まれます。

## Q 各種地図（地理）情報は、既に市のホームページで公開しています。オープンデータとは何が違うのでしょうか？

本市では現在、都市計画、固定資産税路線価、道路台帳など、いろいろな地図をウェブ上で見ることが出来ます。これらの地図情報をオープンデータ化することにより、情報の二次利用が容易になり、他のGIS上でデータを使用することや、他の情報と組み合わせることなど、様々なデータの活用が広がります。

※施設的位置や土地の形状などの地理的な情報をコンピュータを用いて地図上に表示するシステムを地理情報システム（GIS）といいます。情報を地図上に視覚的にわかりやすく表示することや、様々な情報を組み合わせることで、地域情報の分析にも使えます。本市では次のようなインターネットを用いたGISによる情報提供を行っています。

「行政地図情報提供システム」  
都市計画情報（1マッピー）、道路台帳、下水道台帳、固定資産税路線価、地盤情報、わかりやすい防災マップ

「よこはまマップ」  
市内の施設などの位置や名称、写真などを閲覧できるシステム。区民生活マップのウェブ版として各区が公開しているほか小学校の授業での活用

防災マップ、子育てマップなど、市民による地図作りにも活用されている。

「統計GIS」  
国勢調査などの統計データを抽出、集計したり、表示した地図を作ることが出来るシステム。統計情報を地図を用いて視覚的にわかりやすく表示することが出来る。

## Q クリエイティブ・コモンズ・ライセンスとは、なんですか？

オープンデータを提供しても、そのデータが再利用できるのか、どこまで自由に使えるのか、利用条件が明らかでない、権利侵害を恐れて活用が進みません。現在、行政のホームページの多くが著作権を主張しており、ほとんどの情報について、私的使用のため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・転用をすることはできない状況です。その都度、所管部署に問い合わせ、利用条件を確認するのは活用が進みません。

クリエイティブ・コモンズ・ライセンス（CCライセンス）

（ス）とは、著作物の再利用に関する条件等の意思表示を手軽に行えるようにするため、国際的に利用されている仕組みです。

国際的非営利組織クリエイティブ・コモンズが運営・提供しており、インターネット時代のための新しい著作権ルールの普及を目指し、様々な作品の作者が自らの作品に対して、「この条件を守れば自由に使用可」という意思表示をするための仕組みであり、権利者は「表示」「非営利」「改変禁止」「継承」の4種類のマークで示される条件を取捨選択して使用します。

利用条件（ライセンス）の内容は、法律に詳しくない人でもライセンスの主な内容がすぐに理解できる簡潔な説明文「コモンズ証」と、同じ内容を法律の専門家が読むために法的に記述した「利用許諾」（ライセンス原文）とで明示されています。この仕組みを利用することで、作者は著作権を保持したまま作品を自由に流通させることができ、受け手はライセンス条件の範囲内で権利者に許可を得ずとも再配布やリミックスなどを行うことができます。

※参照 クリエイティブ・コモンズジャパン（CCJP）（活動母体・特定非営利活動法人コモンズフジア）  
<http://creativecommons.jp/>  
25年版情報通信白書（総務省）

## Q ビッグデータの活用はよく聞きます。オープンデータとビッグデータは、どのような関係があるのでしょうか？

ビッグデータとは、膨大な量のデータの集まりの呼称です。しかし、どの程度の量を超えるとビッグデータなのか、明確な定義はないのが現状です。技術の進歩により今まで扱えなかったような大量のデータを解析して有益な知見を得ることが可能となり、ビジネスなどへの活用が着目されるようになったものです。ビッグデータの代表的なものがITにより日々流通し、蓄積されている情報であり、ツイッターのつぶやきを分析してユーザーのコーホート特有の購買行動を予測し、対象を絞った効果的な情報発信をするなどの活用がされています。よって、公的データを民間へ利用しやすい形で提供していくオープンデータとは、切り口の異なる概念です。

◇ビッグデータの活用事例（本田技研工業の事例）24年版情報通信白書からの抜粋  
同社では、ドライバーの快適なカーライフを実現するため、より安全で環境にも配慮したドライブ情報ネットワークとして、安全・安心、防災、天気、省燃費ルート等の情報を提供する「Linkas」を2002年からサービス提供開始し、2012年5月現在、会員数は145万人に達している。

同サービスでは、会員から5分毎の間隔で収集した装着車の走行データの共有により、渋滞を回避し、目的地へより早いルート案内を行う「プロローディングカーシステム」を導入している。現在、毎月1億kmのデータがアップロードされ、2012年5月末現在、蓄積した走行データは25億kmに達している。

本システムによる主な効果は以下のとおり。

- ① 渋滞予測による効果検証結果では、約20%早いルートが案内され、CO<sub>2</sub>換算では約16%の削減効果があった。
  - ② 埼玉県の道路行政の取組において、急ブレーキポイントの多発地点を抽出し、街路樹の剪定や路面表示により、急ブレーキ回数が約7割減少した。
- その他、東日本大震災等の災害時に情報提供などに活用されたほか、日本気象協会からの凍結予測等の気象情報との連携により、路面凍結発生の予測時刻や予測時点等のカーナビ画面への表示や音声警告等の気象・減災情報を提供している。
- 加えて、車両内のセンサーから収集している燃料噴射量のデータ活用により、例えば、燃費のよいルートの探索・予測や車両の制御等が可能となっている。

**Q** ハッカソンとかアイデアソンとは何を指すのですか？

ハッカソン(“hackathon”)とは、ハック(hack)とマラソン(marathon)を合わせた造語で、ソフトウェア開発者が集まり、一定期間集中的にプログラムの開発を競うイベントのことです。数日間寝泊まりして行うこともありま

す。アイデアソンは、アイデア(Idea)とマラソン(Marathon)を合わせた造語であり、ハッカソン同様に長時間集中的にソフトウェアのアイデアをまとめるイベントです。

アイデアソンは、専門知識を持っていない人も参加できますので、オープンデータの活用を多くの人に知っていただくためのイベントとしても開催されています。

**Q** データポータルサイトとかデータカタログサイトとは何のことですか？

ポータルサイトとは、インターネットの入り口となるウェブサイトのことであり、

ジャンルごとにウェブサイトにへのリンクを集めているものです。データポータルサイト(略してデータポータル)とは、オープンデータ化したデータセット(データの集まり)を集めて提供するポータルサイトのことであり、データセットを一覧で見せることからデータカタログサイトとも呼ばれます。

**Q** 以前、様々な行政手続きの電子化に取り組みましたが、一部の手続きしか利用が進んでいないと思います。手間をかけてオープンデータ化しても、多くは使われずに無駄になるおそれがありますか？

オープンデータは、行政が持つ情報の多くが対象となり、全てをオープンデータ化するには多額のコストや手間が必要となります。よって、本市のオープンデータ推進指針において、推進のための基本原則として次の2点を定めています。

「取組可能な公的データから速やかに着手し、実績を蓄積する。」  
これは言い換えますと、無

理に急ぐのではなく、できるところから着手し、成果を着実に増やしていきましようという趣旨です。  
「費用対効果について十分に考慮し、効率的に取組を進める。」

オープンデータ化しても使われなければ意味がありません。よって、推進指針においても、重点的に進める分野を定め、また、ニーズに応じて公開データを拡大していくこととしていきます。

ただし、オープンデータは、様々なデータを組み合わせることで新たなビジネスやサービスを生み出すことに意味があります。思いもつかないような分野で活用があるかもしれませんし、広い分野でオープンデータ化しないと活用が進展しないという考えもあります。

**Q** 統計表などの数値データは活用できると思いますが、文書データをオープンデータ化する意味があるのですか？

ホームページの内容の多くは、文章で記述されています。また、図や写真などの画

像も含まれることがありますが。このような文書形式データをオープンデータ化するには、文章の部、章、節、段落などの構造が、コンピュータが明快に認識できるように、見出しを追加します。(タグやマークアップ言語を使用します。)文書データをオープンデータ化することにより、データの検索性が高まり、必要な情報を探して活用することが容易になります。

**Q** データを更新しても利用者が気付かず、古いデータを利用しつづけてしまわないか心配です。

確かに普通にオープンデータをダウンロードして利用すると、データを最新のものに維持する作業が必要です。電車やバスの運行情報や道路の渋滞情報など、リアルタイムに更新されないと意味のない情報もあります。そのようなデータ更新を解決する仕組みが「リンクトデータ」(Linked Data)※です。リンクトデータとすることで、リンク元のデータが更新されると、リンク先のデータも同時に更新することができ、常に最新の

データを利用することができま

す。リンクトデータの仕組みを使うには、データをRDF形式に変換して公開する必要があります。

※リンクトデータとは、一言で言うとデータ版のウェブです。普通のウェブは、人が読む文書が、ハイパーリンクでつながってあり、常に最新の内容を開覧できます。リンクトデータは、文書でなくデータがハイパーリンクでつながったものです。ウェブがHTMLという標準言語で記述されるように、リンクトデータは、RDFという標準言語で記述されます。そして、RDFでデータを記述する際にURI(Uniform Resource Identifier)が使われます。ホームページのアドレスがURLとして指定できるように、URIは、ネット上にあるデータセットの中の特定の項目を指定することができます。この仕組みによってデータセットを超えて相互に参照し合うリンクトデータが可能になるのです。(参照…国立国会図書館「カレントアウェアネス」No.308(2011.6) 国立情報科学研究所 武田英明)

**Q** 二次利用が可能になり、商業利用も構わないとなると、事業者が無料で入手した行政の情報を有料で提供することになりますか、良いのでしょうか？

オープンデータは活用され

ないと効果が生じません。幅広い活用を促進していくためには原則として無料とすることが必要です。例えば、気象情報は国民の安全・安心に関わるものとして気象庁から無料で提供されていますが、気象情報を活用して民間事業者が様々な有料サービスを提供しています。サービスは有料であっても、様々な活用されることで、情報が活かされ、国民の安全や利便性の向上に役立っています。

EUにおけるオープンデータの経済効果分析によると、オープンデータ化による税収増とオープンデータ化のコストを比較した経済効率率は、無料としたアメリカが有料で始めたEUの5・5倍と高い効果を挙げられており、EUがオープンデータ化コストを税収で賄うには経済効果を2倍とすれば足りるとしています。\*

なお、25年6月のG8サミットにおいて合意したオープンデータ憲章においても、無料で利用できるようにすべきとされています。

しかし、情報によっては、受益者が限定されるなどにより、公益性の観点から無料とすることがなじまないものがあるかもしれません。国での検討を参考としつつ、今後、

範囲を拡大していく中で具体的な事例を基に研究していく必要があると考えます。

※IPA International 2000/野口裕子「オープンデータの課題と展望」経済産業省IT融合フォーラム公共データWGにおける紹介

## Q

総務省の実証実験のために公開した横浜市

市のオープンデータを見ましたが、プログラムがないと開くことのできないRDFというファイルや、データが羅列しているcsvというファイルで、見てもよくわかりません。これで市民との情報共有が進むのでしょうか？

オープンデータとは、機械的に判読・処理が容易なデータの構造・形式で公開することであり、人が目でみてわかりやすく情報を提供するものではありません。データをそのまま個人が利用するよりも、ある程度の知識や技術を持った人がデータを活用しやすく提供するものです。行政のデータをわかりやすく加工して一般の人へ提供したり、分析して得た知見を広めたりすることで情報共有が進むと

考えています。

現在でも行政は様々な情報を公開していますが、市民はどんな有益な情報があるのか、全ては把握していません。行政側としても必要な情報を市民へ届ける努力はしていますが、オープンデータを進めることで、さらに民間の創意工夫を加えて、より多くの市民へ有益な情報をわかりやすく提供していくことが可能となります。

## Q

オープンデータを活用した新ビジネスや

新サービスが生まれた話はあまり聞きませんが、オープンデータにより経済の活性化が期待できるのでしょうか？

オープンデータの取組は、日本では始まってまだ日が浅く、オープンデータを活用したビジネスの成功事例として紹介されているものは少ないようです。しかし、例えば気象情報を活用した民間サービスは、既に当たり前のものになっていきますし、交通情報やGPSの位置情報も公共データを活用しているものです。オープンデータの活用が最も進んでいる米国において、

代表的な成功事例として紹介されるものは、Total Weather Insuranceという農業保険です。国立気象サービス局が提供する地域ごとの気象データや、農務省が提供する過去60年間の収穫量や土壌

などに関するデータを解析し、地域や作物ごとの作物被害の発生確率を予測する農家向けの保険サービスです。2006年に会社を設立し、2011年から農家対象のサービスを開始したところ急成長し、取扱高は3兆円に及び、海外にも事業を展開しています。

また、米国のMRSという不動産情報サービスは、公的機関が提供する人口統計、公共交通機関、教育等の情報を不動産情報と組み合わせ、不動産に関わる様々な情報を、不動産業者だけでなく一般消費者へも分かりやすく提供し、住む前に全てが分かる不動産情報サービスとして売り上げを伸ばしており、推定年間売上高が約50億円に達しています。具体的な情報の内容としては、教育関連では先生一人あたりの生徒数や生徒一人あたりの教育予算額、健康医療関連では、住民一人当たりの医師数などがあります。

日本においてもオープンデータとして提供されるデータが増え、さらに生活に直結した地方自治体のデータがより多く提供されることで、ビジネスへの活用が増えていくことが期待されます。

《編集部》