

ビッグデータ・オープンデータ

ビッグデータが民間で、オープンデータは行政が提供するもの、というような図式で捉えられがちですが、正しいとは言えません。

オープンデータは「開放」を目的として利用しやすい形でデータを公開するものに対して、ビッグデータはデータ利用が難しい種類の大量・多種多様・変わりやすいデータを取り扱う仕組みですので、そもそも同列で比較することはおかしいのですが、データを活用するという意味では共通しています。

ビッグデータとは

ビッグデータの明確な定義はありません。文字通り巨大なデータではあるのですが、単に大容量というだけではビッグデータとは言えません。①データ量が多いこと、②データの種類が多いこと、③データの変化する頻度が多いこと、これらの3つの条件が重なることで、従来のシステムでは取り扱うことが困難であったデータとそれを扱うためのシステムがビッグデータと呼ばれる、インターネットの普及とIT技術の進化によって生まれた仕組みなのです。

特に、ビッグデータの特徴は、量だけでなく、扱うデータの種類が多いことにあります。扱うデータには、構造化データと呼ばれる、会計システムなどの基幹システムから発せられる数値や文字列といったデータだけでなく、非構造化データと呼ばれる、文章、音声、動画といったマルチメディア・データなどのデータが含まれます。さらに、電子メールのデータやXMLデータなどの半構造化データ、さらに、各種センサーや機器から発せられるデータや通信ログのように頻度が非常に多いようなデータも含まれます。さらに、これらのデータには社内だけでなく、インターネット上の社外にある場合も多くあります。

また、日々膨大に生成・記録される時系列性・リアルタイム性のあるようなデータが含まれていることが多いといえます。

ビッグデータの活用を先導している企業の多くは Google や Facebook といった Web サービス事業者です。そして彼らのデータ活用は、従来の売り上げデータや顧客データのような社内に存在するデータではなく、Web 上にある文章や画像といったデータが中心となっています。

さらに、これまで企業の多くはデータ活用をたとえ顧客データであっても、活用目的は個々の顧客の属性ではなく、集計することによって得られる傾向情報が中心でした。

しかし、これらの Web サービス事業者では、顧客の個々の属性を捉え、Amazon のようなリコメンデーションのように細分化された情報を利用しています。そして、大きく異なるのは、データ処理に対する精度よりもスピードを重視していることです。

ビッグデータを「事業に役立つ知見を導出するためのデータ」とする考え方もあります。今までは管理しきれないため見過ごされてきたデータ群を収集・蓄積し、即時的に解析を行うことで、異変の察知や近未来の予測を通じて、利用者個々のニーズに即したサービス提供、業務運営のいっそうの効率化、新産業の創出などが期待されます。

オープンデータとは

オープンデータとは、インターネットなどを通じて誰でも自由に入手し、許可されたルールの範囲内で自由に複製・加工や頒布などができるデータのことです。商用としても利用可能です。

オープンデータは「機械判読に適したデータ形式で、二次利用が可能な利用ルールで公開された」もので、「人手を多くかけずにデータの二次利用を可能とするもの」であることが求められます。図画や動画などのデジタルコンテンツも含まれます。

日本では、2012 年に政府が国家戦略として「電子行政オープンデータ戦略」を発表して、以下の 3 つを目的として、国と地方が一体となりオープンデータ推進の取り組みがスタートしました。

- 行政の透明性や信頼性の向上
- 官民協働による公共サービスの提供、民間サービス創出の促進
- 新ビジネスの創出や企業活動の効率化を通じた経済活性化

現時点ではほとんどのオープンデータは、国や政府、自治体、公共機関などが保有し公開しているもので、地理空間情報、防災・減災情報、調達情報、統計情報といった公共性の高いデータに限られています。民間企業としては、2014 年に初めて東京メトロが全線の列車位置や遅延時間などをオープンデータとして公開しました。また、東京電力などが電力需給データを公開しています。

政府は、2015 年に、独立行政法人や航空・空港・鉄道・乗合バス・道路・電力・ガス・郵便などの公益企業にオープンデータの推進を働きかけるとともに、民間企業でのオープンデータの一層の活用を呼びかけました。

さらに、2016年12月には「官民データ活用推進基本法」を制定し、官民データ活用推進基本計画において、電子行政、健康・医療・介護、観光、金融、農林水産、ものづくり、インフラ・防災・減災等、移動の8つを重点分野に指定し、将来的には分野横断的なデータ連携を見据えつつ、2020年を一つの区切りとした上で、分野ごとに重点的に構ずるべき施策を推進することとしています。

(出典:情報通信白書平成29年版 第1部第2章第2節「[データ流通・利活用による課題](#)」)