

地方公共団体におけるデータの内部利活用に向けたデータ加工に関する考察

新井千乃・家中賢作

Consideration on data processing for utilization of data in local government.

Chino ARAI and Kensaku IENAKA

Abstract: In Tsukuba City, data utilization training is conducted to spread understanding of data utilization within the city hall. As a result of training to process the data, it was found that the data corresponding to “Shared data” was created in the data sharing model “Data spectrum”. In this paper, we examined data processing for data utilization. As a result, data processing includes 1) processing to convert the data into machine-readable data, and 2) processing to delete information related to personal information from the data. The above two points were found necessary.

Keywords: データ利活用 (data utilization), データ加工 (data processing), 地方公共団体 (local government), データスペクトラム (Data spectrum)

1. はじめに

平成 28 年 12 月に公布・施行された官民データ活用推進基本法に基づきさまざまな施策が行われ、平成 29 年 5 月には、国、地方公共団体、事業者が公共データの公開及び活用に取り組むうえでの指針としてオープンデータ基本指針が公開された。

その結果、データ利活用の取り組みが活発になり、オープンデータを提供する自治体が徐々に増えている。一方で、自治体が保有するデータは、個人情報をはじめとする公開に適さないデータも多く、オープンデータ数の種類と数が加速度的に増えることは困難な状況である。

つくば市は、EBPM (Evidence Based Policy Making, 証拠に基づく政策立案) を推進するために、2017 年度よりデータ利活用研修を実施している。研修

では、外部講師よりデータ利活用及びオープンデータ等の先進的な事例や具体の効果等について紹介することで、これらに対する職員の理解と行動を促している。その結果、2018 年 12 月にオープンデータサイトを構築、公開しているオープンデータ数は 2019 年 8 月末時点で 82 件となった。しかしながら、自治体内のデータ利活用については具体事例と効果は少なく、これを推進するための課題が多いものと考えられる

そこで、自治体内のデータ利活用を促進するための具体方法について、2018 年度の研修を基に、データ利活用を自治体内部で実現するためのデータ加工の在り方を検討し、研修の方向性を考察した。

2. データ加工研修

2.1 研修実施方針

つくば市では、データ利活用を確実に推進するために、人事研修でデータ利活用研修を取り入れ、促進を継続させる研修を実現している。データ利

新井千乃 〒305-1165 埼玉県川越市南台 3-14-4

朝日航洋株式会社

Phone: 049-244-6012

E-mail: arai.chino@live.jp

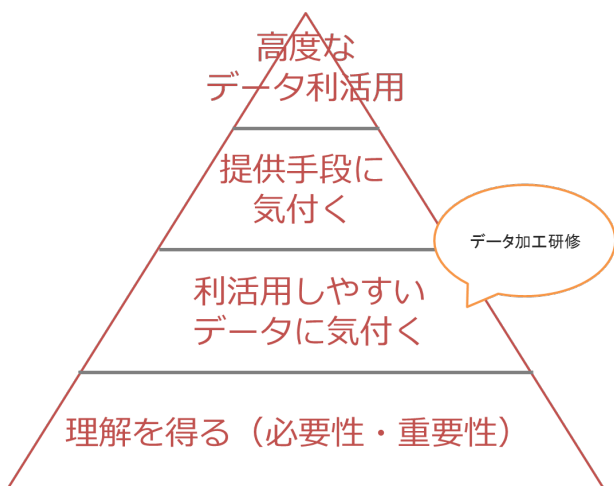


図1 つくば市におけるデータ利活用の考え方

活用研修では、図1のように、段階的に推進することを考え、第1段階として理解の浸透を目的にデータ利活用の重要性や国の指針等を説明し、第2段階として、データ利活用のためのデータ加工ができるようになることを目的として設定した。本稿ではこの第2段階の研修（以下、データ加工研修という）について述べる。

2.2 データ加工研修

データ加工研修に先立ち、つくば市が保有するデータのリストを確認したところ、GISで二次利用が可能と考えられるデータが多く存在することがわかった。

研修用として提供されたデータの一部をGISで活用できるような加工を事前に試行した結果、可視化や分析ができることがわかった。実際に行った加工作業と、必要と思われる加工作業を整理した結果を以下に示す。

- ・ 見せる表資料として作成されたエクセルのデータベース化
- ・ 住所の結合（ジオコーディングするため）
- ・ 住所のジオコーディング
- ・ テーブル結合するためのキーの正規化
- ・ 共有できない情報の削除

- ・ 個人情報（氏名、位置）の匿名化

そこで、これらの加工作業がデータ加工研修で実践できるように、つくば市のオープンデータサイトや提供されたデータより選定し、研修用データとして用意した。

2.2 データ加工研修の実施

データ加工研修は、まずデータを2次利用するために望ましいデータの姿について理解するための座学を行い、次に事前に選定したデータを2次利用に適したデータに加工する作業を行った。

データを加工する作業は、2.2で用意した研修用データについて、まずは二次利用が可能な状態にする加工法を考え、ワークシートに記録する作業を行った。次に、記録した内容を5人で構成された班の中で共有し、データについての気づきをまとめた。最後に、班の中で加工するデータを重複しないように選び、実際に1人1種類以上のデータについて加工を施す作業を行った。まとめとして、班ごとにデータの加工についての気づきと、実際に加工をした時のギャップ等を発表した。

これまでデータ利活用研修では、オープンデータ基本指針に従い、オープンデータについて理解の普及を図ってきた。データ加工研修を実施したことにより、二次利用できるデータはオープンデータよりも、自治体内部で利用できるデータが多いことがわかった。また、所属部門で所有するデータをGISで二次利用したいと感じた受講者もいた。

3. データ共用モデル

オープンデータ基本指針においても、オープンデータとして扱われないデータについて、データの利用目的、範囲、提供先などを限定して公開し、その活用を図っていくことが有効であると示されている。さらに、信頼しうる関係者内において、インカメラ等のクローズドな環境で行われることが適当とし、つまり、自治体のデータは自治体

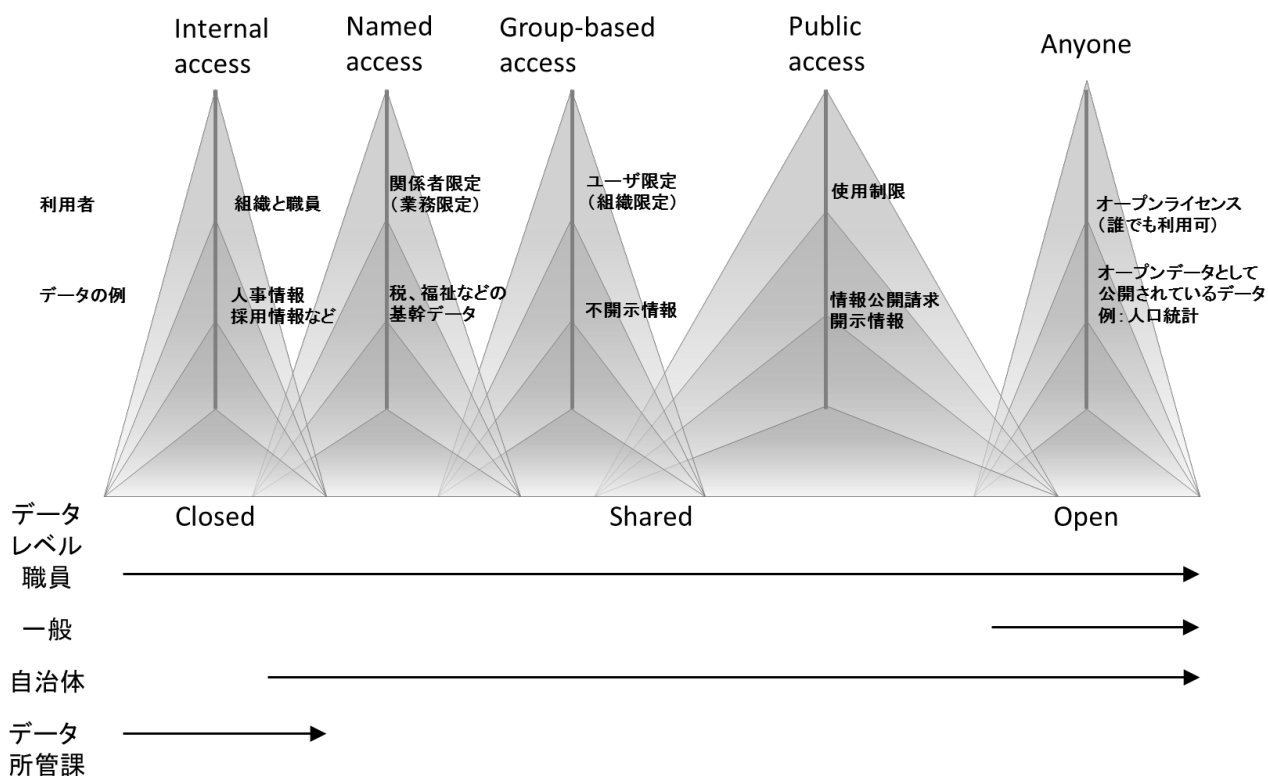


図2 自治体データのデータスペクトラム
及び、データ利用者ごとの使用可能データ範囲
(OpenDataInstitute より引用、一部加工)

内部での利用推進を示唆していると考えられる。この考えには、データスペクトラムが合致すると考えられる。

データスペクトラムは、オープンデータの価値や利用を推進するための基幹である ODI (OpenDataInstitute) において提唱されている概念である。データスペクトラムでは、オープンデータの他に、シェアードデータとクローズドデータが存在することを示している (図2)。

データの利活用の幅を広げるためには、この内のシェアードデータの理解が重要であると考えられる。

シェアードデータには Public access, Group-based access, Named access の3種類があり、使用者制限をかけて公開し、共有するデータを指す。

自治体データの使用者を一般利用者と職員に

置き換えると、図2のようになる。この図2から、一般利用者は自治体のデータをオープンデータの外に、シェアードデータの Public access データも共有が可能であることがわかる。つまり、自治体の役目として一般利用者向けに自治体のデータをオープンデータ及びシェアードデータ (Public access) で用意することで、データの公開数を増やすことに繋がり、ひいてはデータ利活用を促す効果が得られることが考えられる。

2種類のシェアードデータ (Group-based access, Named access) は、自治体職員が利用可能であることを示している。自治体内部でも、所管部門や担当者レベルのみの閲覧、利用となるデータは多い。多くの自治体データの現状が、オープンデータとクローズドデータのみであるとする、職員が利用できる Group-based access, Named access を用意することで、自治体内部利用を促進できる

と考えられる。

3. 考察

シェアードデータでは、使用制限を付けることが必要であり、そのためにはデータに対して共有できない情報の削除等の加工が重要となる。

例えば、現在の自治体内部データであるクローズドデータの内、**Public access** にしたいデータについては、個人情報等の削除や、匿名化をすることで、利用制限が広がり、シェアードデータにすることが可能となる。

国では、平成 28 年の行政機関個人情報保護法等の改正により、公的部門のデータの利活用の対象や範囲を適切に定め、非識別加工情報の仕組みが導入されている。このことから、シェアードデータの需要が高まっていることがわかる。

このように、共有できない情報に対して削除や匿名化等の加工をすることで、データ使用者が利用できるデータが拡充されることが期待される。

オープンデータ基本指針には、機械判読に適した構造及びデータ形式で掲載することを原則とされている。例えば、データベース化することは、機械判読に適したデータと言える。

文字情報は、OCR 化（デジタルの文字コード化）した上で PDF 化することで、機械可読性のあるデータとなる。しかし、数値データは可視化して見せるための資料として公開する他に、データベース化して公開することで、より一層データを分析、可視化しやすくなり、一般利用者にも利用されやすいデータとなる他、自治体内の既存のシステムにデータを連携させるデータになり得ることが考えられる。情報資産のデータベース化は、データ利活用の推進を後押しする重要な加工と言える。

以上のことから、加工したデータに適切な二次利用可能なルールを付与することで、オープンデータとして公開できる種類が増えることが見込まれる。

4. おわりに

本稿では、自治体のデータをシェアードデータに加工することで、自治体内部利用の推進に繋がる可能性を示唆した。また、データ利活用研修において、データスペクトラムの考えに基づいて加工する研修を取り入れることで、自治体職員が加工技術を習得し、作成データの内部利活用を促進できることを示唆した。