

地域におけるデータ利活用の推進

Promotion of Utilization of Data in the Region

長谷川 拓

要約 日本ユニシスは、倉敷市におけるデータ利活用による地域課題の解決や新たなサービス創出に向けた事業に参画してきた。地域におけるデータ利活用を加速させるためには、データに関心を持ちデータから示唆を受けて行動をとれる地域の人材や組織を育てていく必要があり、それがデータの民主化につながり、地域のデータ利活用を押し上げていくことになる。ICTは、これらの人材や組織を対象にデータの収集から加工、分析、示唆提供といった一連のデータ利活用プロセスを支えることが望まれる。日本ユニシスは、ICTを通じて地域のデータ利活用推進、地域課題の解決に貢献していく。

Abstract Nihon Unisys has worked with Kurashiki City to promote the data utilization. In order to accelerate the data utilization in the region, it is necessary to cultivate human resources and organizations in the region who are interested in data and receive suggestions from the data and can act, although they do not have the capability specialized as experts, which leads to democratization of data and pushes up the utilization of data in the area. It is desirable for ICT to support a series of data utilization processes such as data collection, processing, analysis, and suggestion for these human resources and organizations. Nihon Unisys will contribute to promoting regional data utilization and solving regional issues through ICT.

1. はじめに

昨今、さまざまところでデータ活用による社会課題解決の重要性が示されている。「まち・ひと・しごと創生総合戦略（2017改訂版）」^[1]では、データを活用したまちづくりなど、それぞれの地域の特性に即した地域の課題解決と活性化に取り組むことの必要性が示され、「未来投資戦略2018」^[2]では、直面する人口減少、少子高齢化など様々な社会課題に対して、豊富なリアルデータを使い課題解決を図ることが不可欠ということが示されている。また、超少子高齢化社会を迎える日本では、多くの問題を解決するために国や自治体、民間事業者間でのデータ活用が欠かせないとの考えから、官民データ活用推進基本法^[3]が2016年12月に公布・施行され、官民データの適正かつ効果的な活用の推進に関し、基本理念が定められた。これらの背景のもと、岡山県倉敷市では、データを利活用することで地域の課題を解決していく「データ・ドリブン・シティ」という理念を掲げ、オープンデータ、ビッグデータの活用推進に取り組んできた。

本稿では、倉敷市のデータ利活用の事例を紹介するとともに、地域におけるデータ利活用のあり方について述べる。2章では、倉敷市のデータ利活用事業「高梁川インテリジェントICT実装事業」の事例とそこから得られた成果と課題について、3章では、データ利活用を地域の基盤産業へ発展させるための人材や組織のあり方について、4章では、データ利活用を支えるICTについて説明する。

2. 高梁川流域インテリジェント ICT 実装事業

倉敷市は、高梁川流域7市3町とともに2015年度よりデータ利活用事業を実施している。日本ユニシス株式会社（以降、日本ユニシス）は、オープンデータ、ビッグデータの利活用技術として主にIoTやAI技術を提供する外部連携機関という立場で本事業に参画している。本章では、この事業の活動事例を紹介しつつ、得られた成果と課題について説明する。

2.1 事業の背景と目的

倉敷市には日本有数の工業地域の一つである水島臨海工業地帯があり、また、繊維産業の年間出荷額は国内1位である。これらを中心とした製造業は倉敷市における雇用の大きな受け皿となっているが、IoTやAIの発展に伴うパラダイムシフトにより、製造業の雇用構造は大きな変革が起こることが予想される。一方、地域経済分析システム RESAS^{*1}で、倉敷市における情報サービス業の従業員数の特化係数（全国を1.0としたときの比率）を調べてみると、0.13であることが判明し、今後の成長が見込まれる分野での雇用水準が極めて低いことが分かった。

強みが弱みに移り変わることが予想されるのであれば、弱みとされているところを強化して強みに変えていけないか。そこで倉敷市では地方都市において最先端の「高度データ利活用圏域」となることを目指し、オープンデータの推進とその活用技術の向上と普及、地域データサイエンティストの育成、地域社会におけるデータ活用の調査研究、普及啓発といった活動を行い、官民協働でデータ活用推進により地域活性化を図ることを方針とした事業を立ち上げた。ここから、データの収集、加工、分析をビジネスとするデータ産業を、地域における新たな基盤産業とすることを理念とした事業展開を実施している。

2.2 取り組みの紹介

高梁川流域インテリジェント ICT 実装事業では、大きく分けてオープンデータ、IoT、AI、人材育成の四つの領域で取り組みを進めてきた。本節の各項で説明する。

2.2.1 オープンデータに関する取り組み

高梁川流域圏データポータルサイト data eye^{*2}は、高梁川流域圏7市3町に関するオープンデータカタログサイトであり、あわせて地域の特性を分析したビジュアルコンテンツを公開し



図1 高梁川流域圏データポータルサイト「data eye」のビジュアルコンテンツ

ている(図1)。圏域のデータ標準様式を設計し、地域オープンデータの統一化を推進してきた。

データビジュアライゼーションの価値はすでに論を待たないところであるが、例えば RESAS という地域デザインに広く使われているシステムもある中で、data eye で公開しているビジュアルコンテンツは地域密着型として特化した粒度を持っていることが特徴となっている。RESASでの分析は市町村単位であるのに対し、data eye は町丁目あるいは500mメッシュという単位での分析が可能であり、自治体内での施策立案に有効なものである点は、2017年10月のRESASフォーラムでも高い評価を得た。

2.2.2 IoT に関する取り組み

倉敷市における代表的な観光スポットである美観地区の数か所に設置したIoT機器の画像を解析し、歩行人の位置情報と属性情報(性別、年齢)をテキストデータとして抽出し、ほぼリアルタイムに通行量、性別、年代を可視化している(図2)。

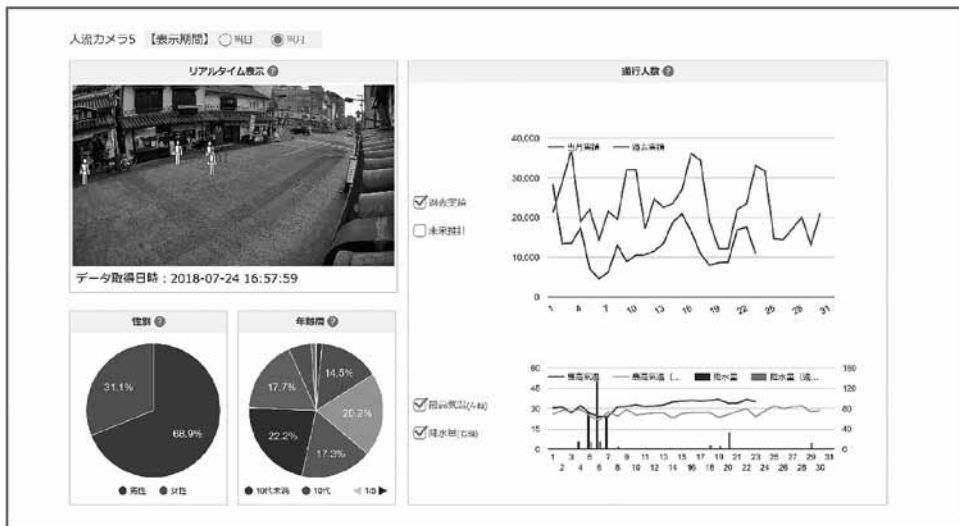


図2 人流解析コンテンツ

日本ユニシスが提供する人流解析サービスおよびデータ統合・分析共通PaaSで構成され、データ収集から可視化までの一連のプロセスをシームレスに連携することを可能にしている。データ統合・分析共通PaaSに蓄積されたデータは、オープンデータとしてdata eye上に公開されており、美観地区の小売店において店舗運営などに活用されている。

この人流解析のコンテンツは、各方面からの注目が高いことから、応用として地域イベントへの適用を試みた。倉敷市の水島地区で毎年夏に開かれている「水島港まつり」や「阿智神社秋季例大祭」において、開催期間の2日間限定で人流解析のシステムを出張構築し、地域イベントの動員状況把握を実施した。

2.2.3 AI に関する取り組み

地元の観光・文化に関する情報をRecommendする音声応答型のスマートフォンアプリとロボットを開発した(図3)。ユーザとの会話を通じて、高梁川流域の観光スポットのRecommend、

開催中のイベント情報の紹介、また、国産ジーンズ発祥の地「児島」で生産される児島ジーンズに関する情報の紹介などを行う機能が備わっている。この機能を支えているのは、日本ユニシスの AI 関連技術を体系化した Rinza、それを活用した知的エージェントサービス RinzaTalk である。RinzaTalk の自然言語解析技術や意図推定技術、コモンセンスの活用により、ユーザの属性や会話の流れに応じた多様な受け答えを実現している。また、ロボットの開発は、日本ユニシスグループであるユニアデックス株式会社の未来サービス研究所が担当している。行政を含む様々な地域住民の知識と IoT で収集されるデータを知識ベースに蓄え、AI アプリケーションの応答力向上を進めている。



図3 高梁川たびコンシェルジュ Tabit (タビット)

2.2.4 人材育成に関する取り組み

データ分析手法などを学ぶセミナーを開催したり、データ分析サロンを開設したりすることで、データサイエンティストとして必要な知識を持つ人材を育成している（図4）。セミナー受講者の多くは、PCの基礎的な知識を持つ市民であり、データ分析サロンにはビジネスパーソンだけでなく、実践を通じて学びたいという学生なども多く訪れる。



図4 人材育成の風景

2.3 事業を通じて得られた成果と課題

オープンデータに関する取り組みの成果である高梁川流域圏データポータルサイト data eye を筆頭に、いくつかのサービスが社会実装された。日本ユニシスが提供する人流解析サービスや RinzaTalk などの技術によって構成された IoT や AI の取り組みは、これまで IoT や AI を単に流行りのワードとしか捉えていなかった地域の方々々に具体的な先進事例を示し、IoT や AI を地域で具体的に利活用する意識を一歩前進させた。また、セミナーやオンライン教育を通じてデータサイエンティストのスキルを身につけた後、テレワーカーとして仕事をしている人も少なくない。プログラムの修了者 40 人以上がすでにデータ関連業務を受注して働いた経験を持っているなど、人づくり、仕事づくりは確実に進み始めている。事業の成果に対して大きな反響をいただくことも多く、データ利活用の機運が高まっている。

一方で、事業の成果を地域に提供しつつ、事業を通じて得られたノウハウを活用した新たなデータ利活用を地域の基盤産業に成長させるという目論見は、今はその途上にある。

3. 地域における基盤産業への発展に向けて

本章では、データ利活用を地域の基盤産業に成長させていくために必要なことを、人材、組織、コミュニティの視点で述べる。

3.1 地域におけるデータ利活用の位置づけ

高梁川流域インテリジェント ICT 実装事業の成果について多くの反響をいただくことがある一方、データ利活用の推進を阻害する要因として、データをどのように利活用したら良いかわからない、費用対効果が不透明、データ利活用人材が不足している、といったことをよく耳にする。これは、総務省「情報通信白書」^[4](図5) や倉敷市が所在する中国地方で活動する中国経済連合会「地域デジタルイノベーションセンターの構築に係る方策検討調査」^[5]でも報告されている内容と一致する。

地域においてデータ利活用を普及させ基盤産業として成長させていくためには、データ利活用で何ができるのか、どうすればできるのか、私たちにもできるのか、という素朴な疑問に答え、データ利活用を専門家ほど特化した能力を持たない地域の方々のものとしていかなければならない。

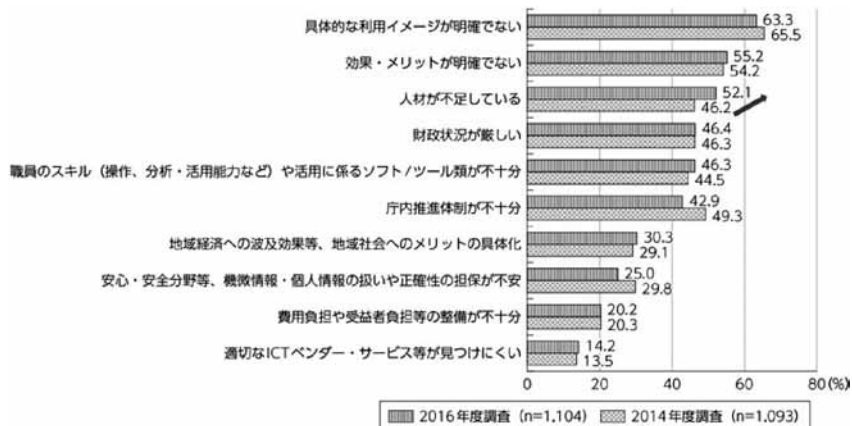


図5 ビッグデータに関する取組を進める上での課題（総務省「情報通信白書」）

3.2 地域に求められる人材

データ分析やICTのスキルに加えて、戦略を提案できる力をもったデータサイエンティストが、今まさに注目されている。しかし、「地方」や「地域」を考えたとき、すぐにプロフェッショナルなデータサイエンティストが必要かという点、必ずしもそうではない。地域や組織の中でデータに関心を持ち、データから示唆を受けて行動をとれる「地域データサイエンティスト」あるいは「市民データサイエンティスト」と呼ぶべき人材をまずは育て、その育成過程そのものがデータの民主化につながり地域のデータ利活用を押し上げていくことになる。データに対する意識のすそ野を広げるといっても良い。こういった「地域データサイエンティスト」あるいは「市民データサイエンティスト」がこれからの地域社会を支えていく。

3.3 データ利活用をけん引する組織

データは本来保有者の資産であることから、データ保有者はデータを外部に提供したくない。これはデータ流通、データ利活用を阻害する要因の一つである。その場合、ある一定の公共性、公平性を持った組織が、データ保有者とデータ活用者をつなぐ役割を担うことで地域におけるデータ流通、データ利活用をけん引していくことが有効である。倉敷市では、2015年に一般社団法人データクレイドルが設立され、オープンデータの推進とその活用技術の向上と普及、地域データサイエンティストの育成、地域社会におけるデータ活用の調査研究、普及啓発活動を行っており、地域におけるデータ利活用のけん引役として、認知されている。

3.4 データ利活用による地域コミュニティの形成

データ利活用を産業クラスターとして形成し推進していくことで、地域の方々に、それぞれの得意分野で積極的にデータ利活用に関わっていただくことが望ましい。地域社会に働きかけてデータを生産、収集する産業を第1次データ産業、第1次データ産業で生産、収集された生データを加工して付加価値の高いデータを生産する産業を第2次データ産業、第2次データ産業が加工したデータを利用してデータ分析、あるいはサービス提供する産業を第3次データ産業と捉え、需要が明らかになることによる活性化と地域コミュニティの形成を目指すものである(図6)。

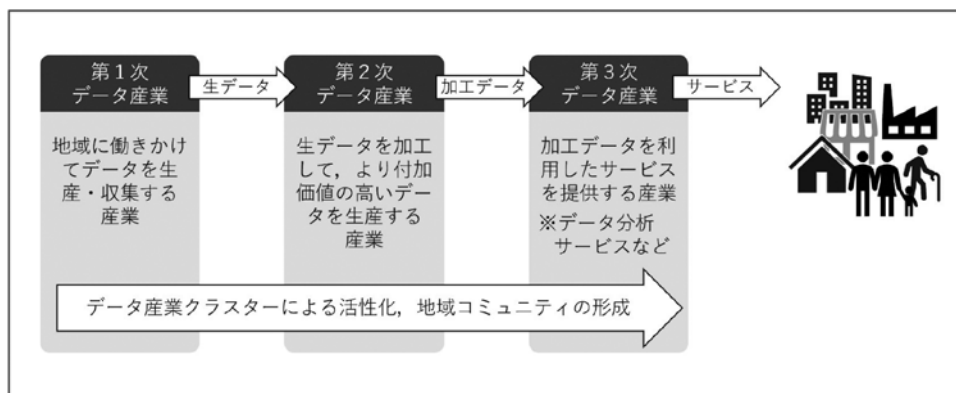


図6 データ産業クラスターのイメージ

4. データ利活用を支える ICT

専門家ほど特化した能力を持たない地域の方々を対象に、データの収集から加工、分析といった一連のデータ利活用プロセスが ICT によって支えられることが望まれる。

4.1 データの収集

データを活用する際は、データ利用者が自ら収集保有する一次データと、世の中に既に存在している二次データを重ねて利用することが多い。従って、二次データを容易に収集できる環境を ICT によって提供することが求められる。気象情報、人口動態情報など、利用頻度が高いと思われる二次データを、例えば 500 メートルメッシュなど地域の粒度で整備し、誰でもすぐに入手できる状態とする。システムとシステムがセキュアに接続し、産業データを共有するといった高度なデータ連携ではなく、まずは二次データがカタログ化されていて、データ利用者が容易にアクセスできるような環境である。

倉敷市の事業では、データ利用者が容易にアクセスできる環境として、LOD^{*3}によるデータ共有環境を整備し、SPARQL^{*4}による外部公開 API も整備してきた (図 7)。しかし、使いこなすには SPARQL の深い知識が求められるなど課題が残る。専門家ほど特化した能力を持たない地域の方々を対象に普及を目指すのであれば、データや項目、対象期間、抽出条件などを直感的な GUI インタフェースで自由に設計定義できる仕組みを今後用意するべきである。

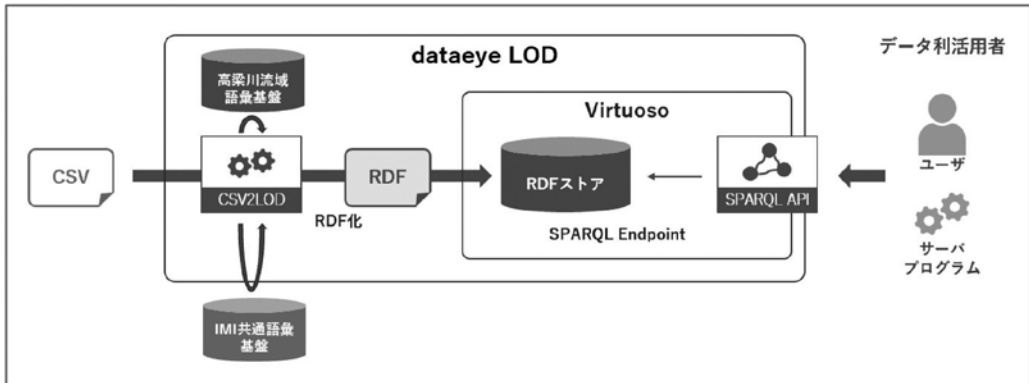


図 7 倉敷市事業における LOD 環境

4.2 データの加工

データ利用者が分析の準備をするための環境が求められる。例えば、データのフォーマット統一、クレンジング、データ特性の把握、データ結合などを、データ利用者自らが簡単な操作で実行できる環境である。直感的な GUI インタフェースを備え、データ準備のフローを定義して自動化でき、分析用途に合わせて簡単にデータを修正する。Excel で処理できる程度のデータであれば良いが、データ量が増大しデータが複雑化すると、人力でのデータ加工は非常に困難となり、過度な負担と膨大な時間がかかる。データ加工に時間を取られてしまい、本来行うべきデータ分析に時間をかけることができないという状況に陥らないために、データ利用者自ら簡単な操作、または自動でデータ加工できる環境が求められる。

収集したデータの表記、意味、データ構造を統一し、データ同士で互いに意味が通じるようにするための環境として、語彙基盤を提供することも求められる。倉敷市の事業では、IMI 共通語彙基盤^{*5} で用意されている情報はコア語彙を活用して RDF^{*6} 化し、IMI 共通語彙基盤で用意されていない情報については各自治体の用語を整理し、高梁川流域の語彙として用意した。

また、倉敷市の事業では、テレワーカーによってデータ加工が行われているが、今後、異なる分野から収集した多様な構造のデータを意味解析し、データのフォーマット統一やデータ結合を機械がサジェッションする機能の研究開発を進める予定である。意味解析は、単語の意味をベクトル表現化できる word2vec^{*7} を利用する。単語をベクトル化することができるため単語同士の意味の近さを計算したり、単語同士の意味を足したり引いたりすることができる。この機能を利用して、収集したデータの項目の和と共通語彙基盤データテンプレートの項目の和を比較し、類似度の高いデータテンプレートをサジェッションする。当該データテンプレートをもとにフォーマット統一やデータ結合を支援する機能である（図 8）。

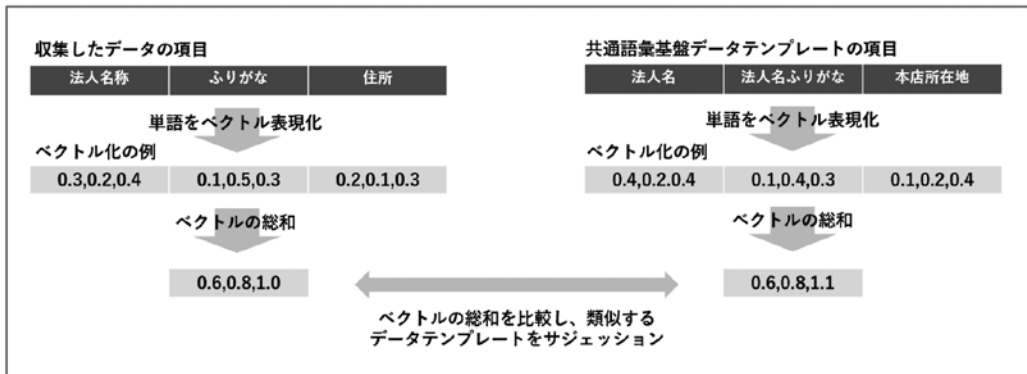


図 8 word2vec による意味解析のイメージ

4.3 データの分析

専門家ほど特化した能力を持たない地域の方々が、データ分析のメニューを選択できる環境が求められる。例えば、来店者の属性別購買行動を分析する、日配品の売上を予測する、といったことが選択できる環境である。備えるべき分析メニューは、現状を分析するものと未来を予測するものがある。これまでのデータ利活用は、過去データを分析し現状を把握することが主流とされてきたが、機械学習やディープラーニングに代表されるような AI 技術の発展により、未来を予測し意思決定することもできるようになってきている。

現在、倉敷市をフィールドに生鮮ベース加工食品の需要を予測する実証を行っている。需要の変動に影響を与える要因から重回帰分析により予測するモデルである（図 9）。このようなモデルを特定商品に依存しないように汎化し、メニューとして提供することを検討している。

また、分析した結果はグラフとして可視化したり、地図上に重ね合わせて表示するなど、視覚的にわかりやすく伝えることで、データ利活用者の今後の行動にヒントを与えるものでなければならない。例えば、e-stat^{*8} を通じて各政府機関が作成する主な統計データを表形式で提供してきた総務省も、2017 年 5 月には統計ダッシュボード^{*9} という形で視覚的にわかりやすく情報提供するようにしている。

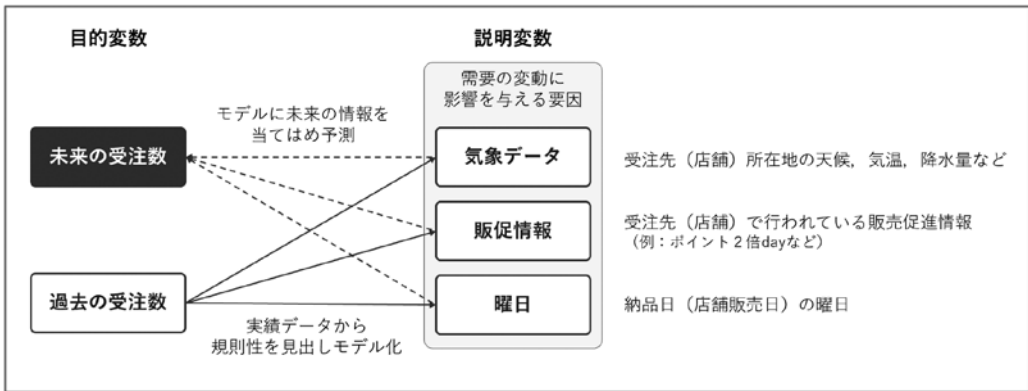


図9 生鮮ベース加工食品の需要予測モデル

現在、倉敷市の事業成果である data eye に対して、ダッシュボード機能を開発している。収集されたデータは、棒グラフ、散布図、地図などの形で視覚的にわかりやすく表現されるため、データの特徴を把握できるようになる(図10)。



図10 data eye ダッシュボード

5. おわりに

ガートナーによると、2019年までにはシチズンデータサイエンティストによるデータ分析の量がデータサイエンティストのデータ分析量を上回るだろうと予測されている^[6]。シチズンデータサイエンティストとは、数学や社会科学などを専攻し、ある程度のデータ分析スキルを有するものの、データサイエンティストほど特化した能力はもたない人々のことをいう。倉敷市の事業を通じて、データ分析を専門としない地域の方々が、データを活用して地域課題の解決、あるいは地域の活性化を目指す方向に向かっていくことを体感しているところであり、この方向性は、ICTによっても後押しされていこう。データ利活用を専門家のもことから、地域の方々のものへ変えていくことにより、地域の方々が当たり前のようにデータを活用する社会、データ利活用によって人と人がつながり、コミュニティが形成されていく社会が到来

し、社会課題の解決と地域の活性化が実現されることを期待する。

日本ユニシスは、データ利活用による地域課題の解決を事業と捉え、地域のデータ利活用に資するICTを積極的に強化しつつ、それらICTの提供を通じて地域の方々と共にデータ利活用を推進し、地域課題の解決、地域の活性化に貢献していく。

最後に、本稿の執筆にあたりご協力いただいた関係者各位に心より感謝いたします。

-
- * 1 RESAS：地方自治体の様々な取り組みを情報面から支援するために、まち・ひと・しごと創生本部事務局が提供する、産業構造や人口動態、人の流れなどの官民ビッグデータを集約し、可視化するシステム。
 - * 2 data eye：高梁川流域圏7市3町の発展に活かせるデータを公開・共有・発見・活用することを可能にするポータルサイト。オープンデータの管理、カタログ公開、ビジュアルの機能を有する。(https://dataeye.jp/)
 - * 3 LOD：ウェブ上でコンピュータ処理に適したデータを公開・共有するための技術の総称。インターネット上のオープンな場へLODの形式で発信することで、情報を多くの人々へ広くかつ迅速に伝えることが可能となる。
 - * 4 SPARQL：RDFクエリ言語の一つ。RDFで記述されたRDFデータのリソースを取り扱うためのコンピュータ言語である。
 - * 5 IMI共通語彙基盤：個々の単語について表記・意味・データ構造を統一し、互いに意味が通じるようにする仕組み。分野や地域を超えてデータの検索向上やシステム連携強化を可能とする。
 - * 6 RDF：ウェブ上にあるリソースのメタデータを記述するための枠組み。RDFのメタデータのモデルでは、主語 (subject) 述語 (predicate) 目的語 (object) の三つの要素でリソースに関する関係情報が表現される。
 - * 7 word2vec：テキストデータを解析し、各単語の意味をベクトル表現化する手法。単語をベクトル化することで単語同士の意味の近さを計算したり、単語同士の意味を足したり引いたりすることができる。
 - * 8 e-stat：各府省等が公表する統計データの一つにまとめ、統計データを検索したり、地図上に表示できたりするなど、統計を利用する上で必要な機能を備えた政府統計のポータルサイト。
 - * 9 統計ダッシュボード：各府省等が作成する主要統計をグラフなどに加工して一覧表示し、利用者にとって視覚的に分かりやすく、簡易なデータ利活用を実現するためのシステム。

- 参考文献**
- [1] まち・ひと・しごと創生総合戦略 (2017改訂版), 首相官邸, 2017年12月22日 <https://www.kantei.go.jp/jp/singi/sousei/meeting/honbukaigou/h29-12-22-shiryoul.pdf>
 - [2] 未来投資戦略2018 — 「Society 5.0」「データ駆動型社会」への変革 —, 首相官邸, 2018年6月15日 https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/miraitousi2018_zentai.pdf
 - [3] 世界最先端デジタル国家創造宣言・官民データ活用推進基本計画, 首相官邸, 2018年6月15日 <https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20180615/siryous5.pdf>
 - [4] 情報通信白書 ICT白書2017, 総務省, 2017年7月 <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/pdf/n2200000.pdf>
 - [5] 中国地域デジタルイノベーションセンターの構築に係る方策検討調査 報告書, 公益財団法人ちゅうごく産業創造センター, 2018年3月 <https://crirc.jp/jigyonaizou/research/jishu/pdf/project/h29-2.pdf>
 - [6] Gartner, Press Releases Gartner Says More Than 40 Percent of Data Science Tasks Will Be Automated by 2020, 2017.1 <https://www.gartner.com/en/newsroom/press-releases/2017-01-16-gartner-says-more-than-40-percent-of-data-science-tasks-will-be-automated-by-2020>

※上記参考文献に記載のURLは、2018年10月3日時点でのリンク先の存在を確認。

執筆者紹介 長谷川 拓 (Taku Hasegawa)

1999年日本ユニシス・ソフトウェア(株)入社。2007年日本ユニシス(株)へ転籍。公共事業部門にて官公庁、自治体のシステム開発に従事。現在は、オープンデータを中心としたデータ利活用に関する事業企画、システム構築に携わる。

