

価値創出のサイクルを回す データマネジメント

Data Management System that Facilitates the Value Creation Cycle

● 松原 正純 ● 中村 実 ● 佐藤 充 ● 吉田 英司
● 松岡 直樹

あらまし

官民データ活用推進基本法や改正個人情報保護法などに代表される法整備の後押しや、産官学連携によるデータ流通推進協議会の発足など、データ流通・利活用の準備が整いつつある。今後、企業のビジネスがデジタル化され流通するようになると、異業種の間でDigital Co-creation(共創)が今まで以上に加速され、新しいイノベーションが生まれる。富士通研究所では、異業種間のDigital Co-creationを実現可能とするConnected Digital Placeを掲げ、その中核としてデータをつながる形式に変更して管理する、データドリブンプラットフォームの研究開発を進めている。

本稿では、データドリブンプラットフォームを実現する最新の研究と、本プラットフォームを用いて価値創出のサイクルを回す実践への取り組みについて紹介する。

Abstract

Driven by legislation such as the Basic Act on the Advancement of Public and Private Sector Data Utilization and the Amended Act on the Protection of Personal Information and the establishment of the Data Trading Alliance through industry-government-academia collaboration, various preparations for data distribution and utilization are underway. As corporate business becomes increasingly digitalized and distributed in the future, digital co-creation among different industries will accelerate more than ever and more innovation will result. Fujitsu Laboratories has proposed the “Connected Digital Place” concept to enable digital co-creation among various industries. As a core part of the concept, the Laboratories conduct R&D on a data-driven platform designed to convert data to a connectable format and manage data. This paper introduces the latest research to realize the data-driven platform, and Fujitsu’s approach to promoting the value creation cycle through the use of this platform.

ま え が き

近年、複数の企業のデータをつなぐことでイノベーションを促進しようとする動きが活発化している。例えば、日本においては2016年末から2017年にかけて、官民が保有するデータの流通・活用を促す官民データ活用推進基本法の制定や、個人情報の安全な活用を盛り込んだ改正個人情報保護法の全面施行などの法整備が進んでいる。また、産官学連携によるデータ流通推進協議会が2017年11月に発足している。データの利活用にいち早く注目したGEやコマツなどの企業は、自社のデータ利活用プラットフォームを展開し、その上にエコシステムを構築することでデータ利活用ビジネスをリードしている。

富士通研究所においても、異業種のビジネスをつないで企業間のDigital Co-creationを実現する、Connected Digital Placeの研究開発に取り組んでいる。このConnected Digital Placeは、行政、教育、製造業、ヘルスケアなど様々な業種のお客様のビジネスをデジタル化してつなげることで、これまでになかった新しいビジネスが次々と生まれる場を提供する。同様に、ドイツのIndustrie 4.0、国際団体Industrial Internet ConsortiumのIndustrial Internet、そして日本のSociety 5.0なども、異業種のシステムをつなげることにより、経済的な発展や人が快適で豊かに過ごせる社会の実現を目指している。

Connected Digital Placeを含めて、前述したいずれのシステム連携においても、異業種の企業間で行われるDigital Co-creationはデータがつながることによって実現される。しかし、各企業のデータは、構造や意味が統一されておらず、同じ種類のデータであってもつなげることは困難である。データを利用する企業が望む形に整形し、素早く提供できるようにすることが、全てのデータ利活用プラットフォームにとっての喫緊の共通課題である。そこで、富士通研究所ではConnected Digital Placeの中核として、この課題を解決するデータドリブンプラットフォームを開発中である。

本稿では、まずデータドリブンプラットフォームの概要を紹介し、次に本プラットフォームを構成するData Bazaar（データバザール）とデータ

インフラストラクチャーについて説明する。そして最後に、本プラットフォームを用いて価値創出のサイクルを回す実践への取り組みについて紹介する。

データドリブンプラットフォーム

富士通研究所が目指すデータドリブンプラットフォームの概要を図-1に示す。Connected Digital Placeのキーファクタとなるのは、お客様のシステムによって扱いが異なる多種多様なデータである。データドリブンプラットフォームは、ビジネスの目的や契約に基づいてこれらのデータを組み合わせ、分析し、利用する意義のあるビジネス価値に昇華させる。その結果、新たな事業やサービス、すなわちデジタルイノベーションが創出され、経済的、産業的な成長につながる。

本プラットフォームは、データの管理・流通・利活用を支援するData Bazaarと、データを保持するデータインフラストラクチャーの二つで構成される。現実世界の人、車、工場などから取得したデータは、データインフラストラクチャーに格納される。そして、Data Bazaarによって加工・流通された後、ビジネスの目的に応じて異業種のデータが組み合わせられ、新規デジタルビジネスの創出や既存デジタルビジネスの改善などに利活用される。

仮想世界で構築された新しいデジタルビジネスは、人の行動のナビゲーション、空き倉庫のシェアリング、工場の稼働率向上など、様々な形で現実世界に新しい価値をフィードバックする。一方で、デジタルビジネスを回す中で、データアクセ

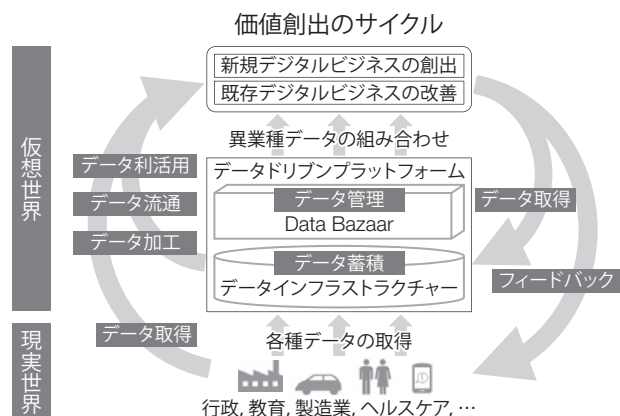


図-1 データドリブンプラットフォーム

スログやビジネストランザクションログなど、新しいデータが生成される。そのようなデータも本プラットフォームで管理し、適切に加工・分析することにより、当該ビジネスの改善や新しいビジネスへの利活用などが実現される。このように、本プラットフォームはデータ利活用が進むほど場が発展する仕組みを内包する。

以上を実現するためには、データをつなげる形にしなければならない。例えば、企業内のデータがその企業固有の商品IDの体系で記述されている場合、業界標準の商品IDの体系に付け替えたデータに変換する。富士通研究所ではこれを「つながる情報(Connectable Information)」と呼んでいる。

つながる情報として流通・利活用するためには、ほかの企業が使いやすいようにデータを加工したり、検索できるようにメタデータを抽出したりする必要がある。データの質・量の管理や、大規模データ処理、リアルタイムデータ処理など、データ処理基盤の充実も必須である。また、このようなデータ流通・利活用を支えるインフラとして、高速かつ大容量なデータインフラストラクチャーが求められる。

以降の章では、データ流通・利活用を支える技術群を提供するData Bazaarと下位層のデータインフラストラクチャーについて、最新の開発技術を紹介する。

Data Bazaar

Data Bazaarは、富士通研究所が開発しているデータ流通・利活用基盤技術群の総称である(図-2)。データ流通・利活用への参加、簡単・安全なデータ連携、データの質・量の管理の三つを自動化することで、データ利活用の障壁を下げる。この際に重要となるのが、データを「つながる情報」に整形し、流通させることである。

企業のデータや、国や地方公共団体が公開するオープンデータなどがつながる情報となり、企業から利用できるようにするためには、特に以下の二つのポイントを満たす必要がある。

- (1) 大量のデータの中から、自社のデータと組み合わせる価値を生み出せるものを探し出すことができる。
- (2) データをつなぎ合わせるために、一方に合わ

せてデータ形式の変換や、データ中の語彙を統一できる。

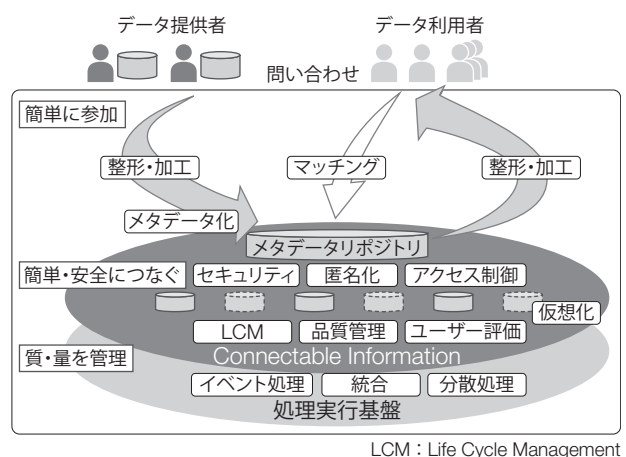
しかし、データを組み合わせる試行やデータ形式の変換、語彙の統一を行うためには、スキルのある技術者が膨大な時間をかけて作業する必要がある。これがデータ利活用を阻むボトルネックと言える。

富士通研究所のData Bazaarでは、スキルがある技術者に対しては作業を支援し作業時間を短縮できることを、またスキルがないユーザーでも自動解析により典型的な分析が行えることを狙っている。Data Bazaarの最終的なゴールは、データ利活用に関する一連の処理を自動化し、お客様にとって有効なデータの組み合わせ・活用方法を提示するコンシェルジュとなることである。

以下では、これらの処理の自動化および高速化を実現する技術として、メタデータ自動抽出・検索技術、データ準備自動化技術、データ分析処理高速化技術を紹介する。

● メタデータ自動抽出・検索技術

ポイント(1)で述べたデータを探索可能なつながる情報とするためには、そのデータについて人間が持っている構造・特徴・作成目的などの知見を、プログラムが読み込み可能なメタデータとして記録する必要がある。このようなメタデータとしては、データ型、データ列の意味、タグ、統計的特徴量がある。これらのメタデータをリポジトリに格納することで、検索や推薦が可能になる。例えば、富士通研究所で研究中の「対話型探索技術」では、



LCM : Life Cycle Management

図-2 Data Bazaar

メタデータ内の文章を自然言語処理技術を用いて解析し、適切なカテゴリを利用者に推薦する。これによって、あいまいな検索キーワードでも大量のデータセットの中から適合するデータを迅速に取得できる。

一方、メタデータを作成する作業にも膨大なコストがかかる。したがって、大量のデータを扱うためにはメタデータ生成の自動化が必須となる。富士通研究所では、この問題を解決するために登録済データとの類似度やデータ加工処理のフローの記録などをベースに、メタデータを自動生成する技術を開発中である。

● データ準備自動化技術

企業内のデータは、CSV/JSON/XMLのような様々なファイル形式が存在し、使用されるデータ形式もまちまちである。データそのもの（人名・企業名・住所など）も体系が異なったり、表記揺れが存在したりする。ポイント（2）で述べたようにデータを結合可能なつながる情報に変えるためには、データの形式や語彙を統一するデータ準備作業が必要となる。しかし、この作業は人手によって行われることが多く、データ分析処理の8割を占めるとされており、データ利活用を進める上での重大な障害となっている。

データの準備で問題となるのは、作成に時間がかかるという点である。一方、技術者は加工手順が不明であっても加工後のデータをイメージできる。富士通研究所では、加工後のデータのサンプルを数件記述してもらい、そこからグラフ探索と独自の枝刈りを行うAI技術を用いて、加工手順を逆合成する技術（PBE：Programming By Example）を開発した⁽¹⁾。この技術を、過去に行われた約8,000件のPOS購買データの準備作業に適用した結果、これまで5日かかっていた作業が約半日に短縮できた。

● データ分析処理高速化技術

データ分析作業が自動化できた場合であっても、現在は最終的な判断を人間が行う必要がある。しかし、1回のデータ分析処理を実行するのに一晩かかるのでは試行錯誤は敬遠されてしまう。そのため、データ分析処理を可能な限り高速化し、リアルタイムな分析を実現する必要がある。

富士通研究所では、データ分析処理の中でもよく

利用されるSQLなどの問い合わせ言語による分析の高速化を研究してきた。これまでに、PostgreSQLの並列クエリー実行機能・列指向格納機能の開発や、PostgreSQL/MongoDB/Shunsaku^(注)の問い合わせ処理を分散実行環境であるApache Sparkにオフロードするなど、データベースのコンポーネントを高速化する技術を開発している。

富士通研究所では、次世代のデータ処理高速化として、クラウドスケールの分散実行技術に着目している。従来のオンプレミスシステムでは、お客様ごとに3～10台程度のクラスタシステムを導入していた。一方、マルチテナント型分散実行基盤では、100～1,000台の計算機資源プールを複数のお客様が共有する。お客様によってデータ処理を行うタイミングは異なるため、お客様はプール内の許す限りの資源を占有した分散実行によって高速化する。このようなクラウドスケールの分散実行技術では、複数のお客様の問い合わせ間のスケジューリング・競合回避が重要である。富士通研究所は、Sparkをベースとしたマルチテナント型の分散実行基盤を研究中である。

データインフラストラクチャー

データドリブンプラットフォームにおいては、現実世界の人、車、工場などから取得したデータを蓄積し、流通に向けた加工を行うData Bazaarに提供するための基盤が必要不可欠である。このようなデータ流通を支える基盤には、大量のデータを効率的に蓄積する機能と、データを高速に処理する機能が必要である。これらの機能を備えたデータ基盤を、富士通研究所はデータインフラストラクチャーと呼び、その実現に向けた技術開発を行っている。

データインフラストラクチャーの構成を図-3に示す。データインフラストラクチャーは、様々な環境に分散して配置される様々な種類のデータを効率良く蓄積・提供するため、場所・用途に応じて柔軟に構成でき、それらを連携させる必要がある。また、更に高速なデータ分析を実現するために、従来のストレージ装置に比べて飛躍的に高いデータアクセス性能を備えている必要がある。これら

(注) XML型データベースエンジン FUJITSU Software Interstage Shunsaku Data Manager

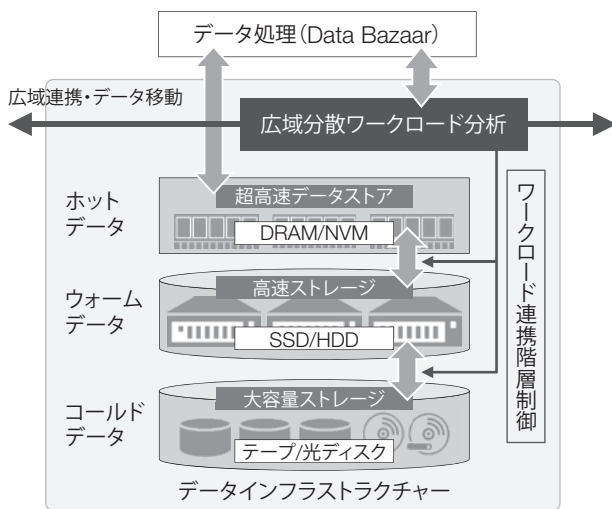


図-3 データインフラストラクチャー

の要求に応えるために、データインフラストラクチャーは高頻度で操作されるホットデータを扱う超高速データストア、それほど頻度は高くないが容量が必要とされるウォームデータを扱う高速ストレージ、および頻度は低いサイズの大きいコールドデータを効率的に扱える大容量ストレージの各層がシームレスに結合した構成を取る。データは必要に応じて必要な場所・階層に移動され、適切なタイミングで提供される。データの移動は、データを利用するワークロードの動作に応じて自動的に行われる。

データインフラストラクチャーを柔軟に構成し、階層間の自由なデータ移動を実現するために、データインフラストラクチャー全体はSDS（Software Defined Storage）で構成される。SDSを用いることで、DRAMやNVM（不揮発性メモリ：Non-Volatile Memory）で構成される超高速データストアのように、処理エンジンと一体化したデータストアを提供できる。更に、ワークロードの動作をリアルタイムに分析し、シームレスなデータアクセスを可能にするための広域分散分析エンジンを備えている。これによって、ワークロードの動作に応じて異なる階層や場所に自動的にデータを移動できる。

富士通研究所では、データインフラストラクチャーを実現するための要素技術の開発を進めている。以下では、その中の超高速データストアを実現するための大容量メモリ技術と、自動的なデー

タ階層移動を実現するためのワークロード分析技術を紹介する。

● 大容量メモリ技術

CPUが高速化し、データ分析能力が向上するに従って、分析のためのデータ入出力もより高速性が求められるようになってきている。インメモリ処理は、その中でも特に高速処理が必要な場合に使われる方式である。これは、全てのデータをメモリ上に置くことで、ストレージの入出力オーバーヘッドを削減し、高速な処理を実現するものである。しかし、従来のインメモリ処理は、分析できるデータ量がサーバのメモリ量に制限されるため、扱えるデータサイズが小さいという問題があった。

富士通研究所では、これを解決するために次世代NVMを大容量の主記憶メモリとして扱うことで、インメモリ処理で扱えるデータサイズを拡大する方式を研究している⁽²⁾。この方式を用いると、ほとんど性能を劣化させることなくインメモリ処理で扱えるデータ量を10倍にでき、より巨大なデータを扱う処理を高速化できるようになる。

● ワークロード分析技術

メモリエースの超高速データストアから大容量データストアまでをシームレスにつなぐためには、ワークロードの利用状況に応じたデータ配置が必要となる。そのためには、まずデータインフラストラクチャー側でデータの動きを分析し、実際の使われ方に応じて制御する必要がある。富士通研究所では、このワークロードのデータアクセス分析技術に関しても研究開発を行っている。この技術は既存システムにも適用でき、例えば仮想デスクトップ向けのストレージボトルネック分析などに使うことができる⁽³⁾。

データインフラストラクチャーでは、この技術に応用し、データの使われ方を分析した上で自動的に必要なデータを最適なデータ階層に配置する。これによって、低コストで大容量のデータを格納し、高速に分析エンジンに提供できるシステムの実現を目指している。

価値創出のサイクルを回す事例

前章までに述べたデータドリブンプラットフォームの開発と並行して、ビジネス検証のための実証実験をお客様とともに進めている。本章で

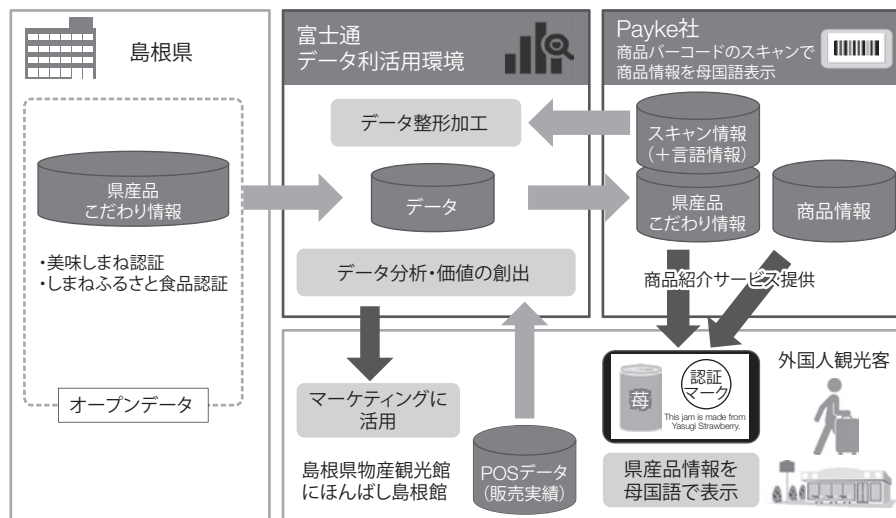


図-4 地域行政分野の実証実験

は、その中の一つである島根県様（以下、島根県）、株式会社Payke様（以下、Payke社）、および富士通とのDigital Co-creationによる地域行政分野の実証実験（図-4）について紹介する。

地域行政の課題の一つに地域活性化がある。行政と企業のデータをつなぐことで、これまで可視化されていなかった地域の人やモノの行動が明らかにできる。更に、この結果を地域企業のビジネスに活用することで、地域の産業を活性化できる。これを実証するために富士通研究所は、島根県が保有する美味しまね認証やしまねふるさと食品認証などの県産品認証データと、Payke社が保有するデータを相互に利活用できるデータ活用環境を、データドリブンプラットフォームを用いて実現した。

このデータドリブンプラットフォームを活用して、2017年9月から島根県松江市の島根観光物産館と東京都のにほんばし島根館において、外国人観光客向けの県産品紹介サービスを提供している。外国人観光客は、店内商品に付与されているバーコードを店内に設置されたカメラ付き端末にかざすことで、商品の説明とともに、県産品認証マークや生産者のこだわり情報を母国語で確認できるようになった。これにより、外国人観光客に対して県産品の情報をより詳細に伝えることができ、島根県ならびに県産品の認知度が向上するとともに県産品の購買が拡大し、地域産業の活性化が可能となる。

一方、データドリブンプラットフォームは、どの国のどんな人がどの県産品に関心を示したかをひも付けることを可能とした。更に、店内商品の参照履歴、県産品認証データ、そして島根県物産館および都内のアンテナショップにおける販売データをひも付けることで、そもそも関心を示されなかったもの、関心を示したが売り上げにつながらなかったものなど、お客様の行動にひも付けたきめ細かな分析ができた。その結果、よりお客様に寄り添ったマーケティング施策を打つことが可能となり、価値創出のサイクルが生まれた。

今後は、地域ビジネス全体の活性化を狙って、観光客の人流データや他小売り企業の販売データとの組み合わせや、異業種企業へのデータ提供なども検討中である。

む す び

本稿では、異業種をまたぐ企業間のDigital Co-creationを実現するConnected Digital Placeの中核であるデータドリブンプラットフォームの開発、および価値創出のサイクルを回す事例に関する富士通研究所の取り組みを紹介した。

今後も、引き続きデータを中心としてDigital Co-creationを実現する技術の開発を進めるとともに、お客様と連携して実証実験を行うことで、お客様のビジネスを真に支えることのできる技術を素早く提供していく。

参考文献

- (1) 富士通研究所：データ分析を加速するデータ準備作業の自動化技術を開発。

<http://pr.fujitsu.com/jp/news/2017/09/15.html>

- (2) 富士通研究所：ビッグデータ処理に最適なソフトウェア制御型SSDを開発。

<http://pr.fujitsu.com/jp/news/2015/11/19.html>

- (3) 富士通研究所：仮想デスクトップシステムの性能劣化要因を特定する自動分析技術を開発。

<http://pr.fujitsu.com/jp/news/2017/05/23.html>



松岡 直樹 (まつおか なおき)

(株) 富士通研究所
デジタル共創プロジェクト
デジタルビジネス創出の研究開発に従事。

著者紹介



松原 正純 (まつばら まさずみ)

(株) 富士通研究所
ソフトウェア研究所
デジタルサービスアーキテクチャーの研究開発に従事。



中村 実 (なかむら みのる)

富士通 (株)
プラットフォームソフトウェア事業本部
データベースの研究開発に従事。



佐藤 充 (さとう みつる)

(株) 富士通研究所
コンピュータシステム研究所
ストレージソフトウェアおよびストレージサーバアーキテクチャーの研究開発に従事。



吉田 英司 (よしだ えいじ)

(株) 富士通研究所
コンピュータシステム研究所
データシステム技術の研究開発に従事。