

企業におけるデータ活用の課題と、必要となる取り組みについて

佐々木 潔 (Z ホールディングス株式会社 常務執行役員 グループチーフデータオフィサー (GCDO))
萩原 元彦 (ヤフー株式会社 CDO 室)

Challenges in Data Utilization at Business Enterprises and Initiatives that Need to be Implemented

Kiyoshi Sasaki
Managing Corporate Officer, GCDO (Group Chief Data Officer), Z Holdings Corporation
Motobiko Hagiwara
CDO Office, Yahoo Japan Corporation

【要旨】 ヤフーの強みは数千万人のユーザーと約 100 もの幅広いサービスであり、そこから生み出されるデータである。しかし、2010 年代前半のヤフーは、全社のデータ総量は多いものの、弱いサービスが持つデータ量は少なく、全社のデータを横断的に利用しようとしても、個別最適やサイロ化がそれを阻み、プライバシーリスクや利益毀損も発生しうった状態であった。本来の強みを活かすべく、ヤフーが持つ多種多様なビッグデータを「マルチビッグデータ」と呼び、組織、技術、プライバシーの観点で改革に取り組むことで、マルチビッグデータを全社で横断利用し、各サービスを改善し伸ばせる状態を目指してきた。本文書では、それらの課題や取り組みを解説することにより、企業全般に通じるデータ活用の要点を示唆するものである。

【キーワード】 データ AI 体制整備 標準化

【Abstract】 Yahoo Japan Corporation's strength lies in the tens of millions of users, over 100 diverse services, and the data generated from them. However, in the early 2010's, while the company as a whole possessed a large amount of data, the amount of data held by the weaker services was small. Even if we tried to use the company's data across the board, individual optimization and data silos prevented us from doing so. Moreover, there was a potential privacy risk and damage to earnings. In order to make the best use of our true strengths, we have named the variety of big data existing in Yahoo Japan, "multi-big data," and have been working on reforms in terms of organization, technology, and privacy so that we will be able to use multi-big data across the entire company and improve and expand each service. By explaining the challenges encountered and initiatives taken, this paper aims to present the implications and essential points in data utilization that can be applied to business enterprises in general.

【Keywords】 Data AI Organizational Improvement Standardization

1. はじめに

本文書では、ヤフー株式会社（以下、ヤフーと記載）がこれまでに行ってきたデータ利活用や、その過程でぶつかった課題、その課題に対する解決策な

どの事例を解説することにより、企業のデータ活用における要点を示唆するものである。

ヤフーは比較的、データをビジネスに用いやすいウェブ企業であるが、1996 年の創業時から万全のデータ利活用ができていたわけではなく、外部環境の変化や社内ビジネスの需要などに合わせてデータ

に向き合う形を変え、少しずつ前進してきた。これは、ウェブ企業だけにとどまらず、他の業種・業態の企業、及び、企業以外の組織にも参考になるものと考えている。

2. データと AI の重要性の高まり

近年、ますますデータと AI の重要性が増しており、もはや企業のビジネスにデータを用いることは当たり前になっていると言える。

ヤフーにおける事例を説明する前に、本章では世界全体で起きている変化の背景について説明する。

2.1. データの増加

インターネットの世界ではデータの取得や蓄積が容易である。ユーザーが Web ページを見ればアクセスログが残り、クリック・購買などユーザーの行動ごとにログが蓄積される。ログインが必要なサービスであれば、氏名や住所などある程度の個人情報登録される。要するに、インターネットが発展すると、それに伴いデータも増加することになる。ヤフーが誕生した 1996 年のインターネット普及率¹は 3.3%。ここから 2000 年には 37.1%、2010 年には 78.2% まで増え、2019 年には 89.8% まで来ている。

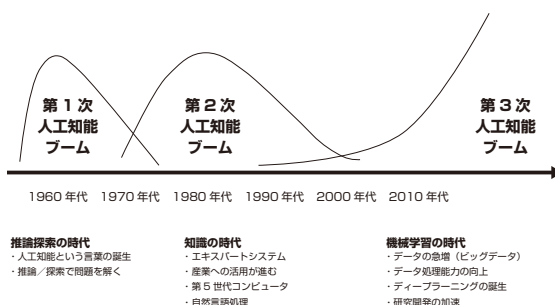
また、2000 年代後半からのスマートフォンの普及により、それまで時間的・空間的に限定されていた PC 中心のインターネットへのアクセスが、いつでもどこからでも可能になり、ユーザーとの接点に大きな変化が起きた。そして、2010 年代に入ると家電やビル管理システムなど、様々な「もの」がインターネットに繋がるようになった。スマートスピーカーやスマートウォッチ、各種センサーなどの IoT 機器の増加だけでなく、QR 決済などリアルとネットの融合が進むことにより、さらにユーザーとの接点が増加している。

このように、過去 20 年ほどの間に、インターネットの発展やスマートフォンの普及、IoT によるデバイス増加、リアルとネットの融合など、多くの劇的な変化が起こった。そして、それに合わせて生み出されるデータも非連続的に激増したと言える。

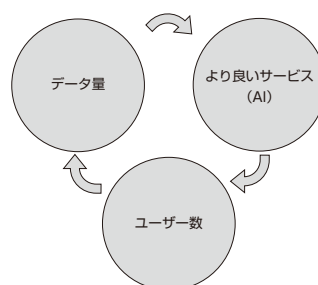
2.2. AI の進化

現在は「第 3 次人工知能ブーム」と言われている。2000 年代に入った頃から AI の研究が活発になり、画像分類などの AI が得意な分野で人間の精度を上回っているのはもちろんのこと、これまで難しいとされていた、より複雑で汎用的な分野でも大きな成果をあげている。これにはディープラーニングが大きく関わっているが、その原料こそがビッグデータである。

「第 3 次」の名の通り、人工知能のブームはこれまで 2 度(1960 年代と 1980 年代)来て、去っている。ブームが衰退した理由として、学習用データの準備が大変だったことやコンピュータの処理能力が足りていなかったことなどが挙げられるが、インターネットの普及などの理由により学習に使用できるデータは飛躍的に増加しているため、今回のブームは終わらないのではないかという見方もある。ちなみに、並列処理技術や GPU (Graphics Processing Unit) の利用により、処理能力の方も大きく向上している。



AI は実用レベルに十分進化した。企業が提供するサービスに AI を実装することで、サービスはどんどん便利になり、その便利を享受しようとユーザーが増え、さらに多くのデータが生み出される。そして、そのデータを用いて AI はさらに賢くなっていく。このサイクルは無限のループである。



が発生している企業は多いのではないだろうか。

3. ヤフーが直面したデータ利用の壁

ヤフーはインターネット黎明期の1996年に創業された。初期はごくシンプルなポータルサイトであったが、インターネットの発達やデータ量の激増、PCからスマートフォンへのデバイスシフトなど、世界的な潮流に合わせて事業やユーザー体験を変化させてきた。そして、常に便利なサービスを提供することでユーザー数を増やし、企業としての売上を作っている。例えば、ユーザーの興味関心に合ったニュースを掲載すれば、ユーザーの数や滞在時間が伸び、メディアとしての価値が上がる。そして、広告枠から多くの収益を得ることもできる。また、ユーザーの興味関心に合った商品を掲載すれば、ユーザーのクリック率が上がり、多くの商品が売れて収益が得られる。これらを可能にするのは、ユーザーの興味関心やニーズに対する深い理解であり、その元となるのはデータである。

ヤフーでは、2005年にデータ利活用の基盤としてデータウェアハウス（DWH）を導入したのが、最初の全社規模の活動である。その後も、ビジネスの可視化・分析やBI（ビジネスインテリジェンス）の用途でデータ利活用が行われていたが、サービス開発やサービス改善に本格的にデータサイエンスを用いたのは、2010年代初頭の広告サービスが最初となる。当時も現在も、広告サービスはヤフーの収益を支える屋台骨であり、ユーザーの興味関心に合わせて掲出する広告の精度が1%上がるだけで、年間の収益が億単位で変わってくる。当然、広告の精度向上を目指すわけだが、当時行っていた手法では頭打ちになっていたため、新しい試みとして機械学習などのデータサイエンスが行えそうなメンバーを全社から集めて部門を組成し、取り組むこととなった。数年間の試行錯誤を経て、結果、広告サービスは大きく成長することができた。

広告サービスの成功を受けて、その他のサービスでもデータやAIを用いた開発・改善に取り組んだが、やがて新たな壁にぶつかることになる。

以下では、この過程で直面した問題について具体的に説明する。程度の差はあれども、よく似た状況

3.1. コールドスタート問題

自社サービスへ新規のユーザーを誘導するためには、他のサービスやメディア面にリンクを貼ってもらい、そこからユーザーを送ることが一般的である。これはヤフーでも同様に行われるものだが、他社との違いは数千万人のユーザーが集まるメディア面を持っていることであり、その送客能力は非常に大きい。太い源流から幾筋もの細い水流が出る様を模して「シャワー効果」と呼び、各サービスの担当はヤフーのメディア面の良い誘導枠をこぞって求めていた。

一方、このシャワー効果の力は絶大だが、リンクで誘導する前時代的なWeb 1.0のアプローチでしかなく、ユーザーは興味のないリンクを結局クリックしないため、誘導効率の点でもUXの面でも、よいものではない。そのため、ビッグデータを利用して効率よくサービスを改善し成長させる方法が求められたが、ユーザー数が少ないサービスや新規サービスを成長させる場合、大きな問題が存在した。いわゆる「コールドスタート問題」である。

コールドスタート問題について一例を挙げる。ショッピングサイトでは、能動的にユーザーが自分のほしい商品を探す機能に加え、ユーザーが興味ありそうな商品を自動的にレコメンド（推薦）することによって、ユーザーが探す手間を減らし、ユーザーと商品の出会いを作る場合がある。これは、ユーザーが過去に閲覧した商品や購入した商品の履歴を分析し、それと類似の商品をレコメンドすることが一般的だが、サイトを初めて訪問したユーザーや、訪問頻度の低いユーザーの場合、過去に閲覧・購入した商品のデータが乏しいため、精度の高いレコメンドを行うことができないという問題がある。

解決策としては、個々の商品ではなく商品のカテゴリにマージして分析することでデータ不足を補ったり、ユーザー個別ではなくユーザー全体の集団としての興味関心を用いたりすることが考えられるが、これらの方法では個々のユーザーにマッチした精度の高いレコメンドを行うことは難しい。

2010年代前半のヤフーにおいては、ユーザー数が少なく、そこから得られるデータも少ないサービ

スはいくつもあり、その状況でビッグデータを用いてレコメンドを改善しようとしてもコールドスタート問題が発生してしまう。ウェブ検索やニュース、オークションなど、全社として豊富なビッグデータを保有している状況を横目に見て、自サービス以外のデータも使って精度を上げられないかという欲求は当然のように高まっていた。

3.2. データのサイロ化

先程の例のように、企業内で複数サービスを運営している場合、一つのサービスのデータだけでなく他のサービスのデータも合わせて利用することで、データの総量を増やすことができる。これが「データの横断利用」である。特に、ユーザー数が少ないサービスや新規サービスでは、データの横断利用の必要性はより大きいものとなる。しかし、企業が持つ重要なアセットを横断的に利用するという「ごく当たり前のこと」を実践するのは難しい場合が多い。この難しさを生み出す要因は、ビジネス的な成果を短期で求めることによる個別最適やデータのサイロ化である。

サービスを新規に立ち上げる場合、まず基本的な機能を開発し、世の中にリリースする必要がある。その際にもっとも早いのは、隣のサービスや隣の部署を気にすることなく、自分たちの作りやすいように作り、置きやすい場所にデータを置くことである。将来を見据えることが大事である点は理解しつつも、決まった期日までにリリースする、リリース後〇ヶ月後に〇円売り上げる、などの目標が連続することで、近視眼的になってしまうのは世の常であり、ある程度避けられないことだと考えている。

これによって発生する状態を説明するため、仮の例を一つ記載する。

ユーザーによる「クリック数」というのはウェブ企業では当然利用されるデータであるが、企業内のあるサービスにおいては「click」というキーでログを蓄積する一方、別のサービスでは「clk」、また別のサービスでは「c」というキーで蓄積している。この場合、一人のユーザーが複数のサービスを利用していても、サービスを跨いでクリック数を自動的に合算することはできず、各サービス担当者に聞き

ながらデータのマッピングを行う必要が出てくる。これはフォーマットがバラバラという仮の例だが、近い状況はヤフーでも発生していた。

上記以外にも、サービスごとや担当者ごとにデータの置き場所（自社サーバー、外部クラウドなど）の選定を行うことでデータの横断的な結合が難しくなる例や、あるサービスのデータを別のサービスが利用している状況で、提供元サービスがデータの仕様を変えたために、提供先サービスが停止する例など、個別最適やサイロ化が生じさせる悪い状況は数多く挙げられる。

3.3. ガバナンスの不備

データガバナンスの観点でも、企業内でデータを横断利用する難しさがあった。

前述したように、ビジネスを伸ばすために自サービスに個別最適した判断を行うことは多い。例えば、個人の経済状況というのは非常にセンシティブであり、プライバシー性の高いデータだが、サービスの性質上、それを推定し得るデータに触る機会の多いコマース系サービスでは、他のデータと区別なく扱ってしまう傾向があった。サービス担当者は、法的要件やプライバシー影響を専門的に扱う役割やスキルを持っていないため、自サービスの常識で個別最適を行うと、ユーザー保護というサービス運営の大前提を揺るがす状態が発生する可能性がある。これは企業にとって大きなリスクである。

また、企業内の各サービスが個別最適で動いた結果、企業全体としての利益を毀損するケースもある。ここでも実例を元にした仮の例を挙げる。

企業内で複数サービス（ここではAとBという2つの事業とする）を運営する場合、それらのサービス間で競合関係は通常発生しない。しかし、サービスAが取引している外部企業とサービスBが競合である場合は存在する。企業内でデータ横断利用ができる状況では、サービスAがサービスBのデータを利用でき、サービスBのデータを元にした提案は、サービスAの取引企業（サービスBの競合）にとって魅力的な場合が多い。

さらにイメージを付けるため、サービスAは広告サービス、サービスBはコマースサービスとする。

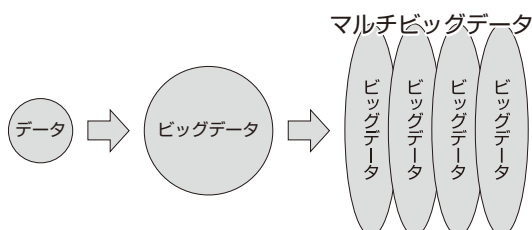
サービス A がサービス B の持つ「もう少しで買ってくれそうなユーザー」というデータを用いたターゲティング広告を社外に販売した場合、コマース系企業はこの広告を用いて潜在顧客よりも顕在顧客に近いユーザー層にアプローチでき、非常に有益である。サービス A の利益も上がる。一方、サービス B としては、自サービスで買ってくれそうだったユーザーを競合に奪われる状態になり、たまったものではない。これでは、データの横断利用に協力する気が失われる。また、一般的にサービス A が広告を売って得られる利益は、サービス B が商品売って得られる利益より小さいので、企業の全体最適としては利益を毀損している。

このように、ガバナンスが不十分な状態では、データ横断利用のメリットよりもデメリットが大きくなり、結局進まないという事態に陥る。

4. 壁を乗り越える取り組み

前章で記載したように、2010 年代前半のヤフーは、全社にあるデータの総量は多いものの、ユーザー数が少ないサービスや新規サービスは改善に資する十分なデータを持っておらず、また、全社のデータを横断的に利用しようとしても、ビジネスを優先することによる個別最適やデータのサイロ化がそれを難しくしており、個別最適の結果としてプライバシーリスクや利益毀損も発生しうる状態であった。

ヤフーの強みは数千万人のユーザーと約 100 もの幅広いサービスであり、そこから生み出されるデータである。この強みを活かすべく、ヤフーの様々なサービスが生む多種多様なビッグデータを「マルチビッグデータ」と呼び、サービスの枠を越えて全社的に利活用することで、各サービスを改善し伸ばすことを目指した。



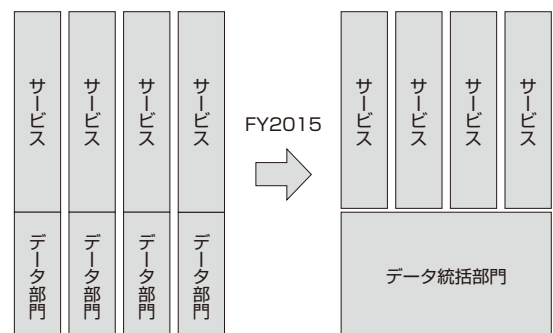
しかし、この大目標を実現するための道のりは簡単ではない。ヤフーでは、組織、技術、プライバシーの観点でこの課題に取り組み、強みであるマルチビッグデータを全社で横断利用できる状況を作ってきた。

以下ではそれらの取り組みの概要を説明する。これらはヤフーにおける取り組みであるが、体制整備や標準化など、企業全般に应用可能なものもあると考える。

4.1. 組織体制のアプローチ

一般的に組織には機能別組織、事業部制組織、マトリクス組織の 3 つの編成方法がある。各組織について詳しくは記載しないが、それぞれにメリット・デメリットがあり、絶対的な正解はない。単体事業の成長速度や事業に対する当事者意識を優先する場合は事業部制組織を、全社最適や横連携、中期的な専門性向上を優先する場合は機能別組織を選ぶ場合が多い。中間のマトリクス組織は機能別組織と事業部制組織のいいとこ取りとなるような組織ではあるが、上司が 2 人いるなどの弊害もある。

ヤフーでは、サービスごとに日本の No. 1 になるべく事業部制組織の形をとっていたが、そのためデータのサイロ化が起き、全社でのデータの横断的な利用はなかなか進まない状況だった。そこで、2015 年に全社に分散していたデータや AI の関連部門を集め、数百人規模の全社横断の機能別組織を発足させた。これにより、全社最適となるデータ環境やガバナンス体制を整えるとともに、データや AI を専門的に扱うメンバーが学び合って専門性を高められる環境を作った。



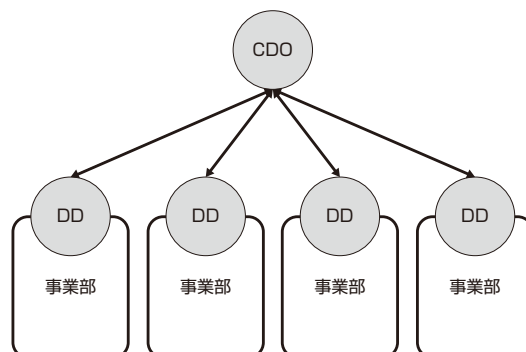
この変更は、これまで自分の事業のためだけに専任していたデータエンジニアやサイエンティストが中央に召し上げられることになるため、事業側とのハレーションが当然予想されたが、大きく2つの対策を行うことで乗り越えていった。経営陣から継続的なメッセージと、事業部との協力体制の構築である。

前者については、マルチビッグデータを全社で横断利用できる状況を作るという大目標に加え、中央の横断組織に入るメンバーに対しては、技術のための技術、AIのためのAIではなく、データとサイエンスを徹底的に活用してソリューションを提供し、その結果としてユーザーや会社の利益に貢献することが組織の役割であると、経営層から繰り返し伝えられた。これは、中央メンバーの意識の醸成に役立ったと考える。

また後者については、中央の横断組織にいるデータエンジニアやサイエンティストの担当事業を決めて事業部側に兼務を行うとともに、彼らの目標の中に担当事業部が追うビジネス目標を含めた。これは、中央メンバーの事業への貢献意識を高め、事業部側と一体となってサービス開発に取り組むことを意図している。もちろん、中央としてのミッションも同時に持ち、全社横断的な活動や専門性向上なども並行して行っていた。

また、事業部のスピードを低下させないための手当も行った。事業部がデータやAIに関連する事項をいちいち中央に相談するのでは意思決定が遅くなるため、事業部ごとにデータの責任者としてDD（データディレクター）という役職を置き、それを中央で統括する全社のデータ責任者であるCDO（チーフデータオフィサー）を設置した。このCDOとDDが連携する体制により、全社レベルで必要な判断と事業部レベルで行える判断を切り分け、全体最適と個別最適のバランスを取ることを目指した。

データに関する意思決定の体制



本稿執筆時点でも、データやAIを扱う全社横断の機能別組織、及び、CDO-DD体制は改善を加えながら継続している。

4.2. 技術的なアプローチ

全社的なデータの横断利用を可能にするために、真っ先に思いつく技術的な対応は「標準化」である。しかし、これは私の持論であるが、人間はどんなに強いられたとしても、やりたくないことはやらず、一時的にはやったとしても継続せず止めてしまう。これは、標準化でも同様であり、その先にその人にとっての良い世界が待っていないと、いくら全社最適と謳っても進まない。なので、各人がやりたくなるところはどこかを考え、横断利用を可能にする状況を作った。この活動には終わりがなく、現在でも継続しているものである。

1) データが集まる「引力」の形成

2010年代前半、ヤフーが持つデータは劇的に増加し、各サービスにおけるデータ管理が難しくなっていた。サービスの開発・改善に集中するためにも、各サービスは可能なら誰かにデータ管理を任せたいという思いがあった。それに対応すべく全社として各サービスが利用できる共通のデータ基盤を準備してはいたが、なかなか安定しなかったため、技術者を多く抱える大きなサービスは自前でデータ基盤を運営することとなり、結局分散した状態が継続していた。

仮にデータが中央の一箇所に集まれば、中央組織の技術者がデータに触りやすくなり、データの整備

もしやすくなる。そうなれば、各サービスが他のサービスのデータと結合してデータを利用しやすくなる。そうすると、自前でデータを管理していたサービスも中央のデータ基盤にデータを入れてメリットを享受したくなり、よりデータが中央に集まってくる。このサイクルを引力に例え、「データグラビティ」と呼んだ。この引力を発生させる最初の一步が、各サービスのニーズを満たす強力なデータ基盤であると考え、その強化に注力していった。担当部門の努力によって、現在ではヤフーのデータ基盤は非常に強く安定したものになり、全社のほとんどのデータはその上で運用管理されている。

2) 標準化やエコシステムの整備

強力なデータ基盤の上に全社のデータが集まる状態ができれば、横断利用への道のりは大きく前進したと言える。

各サービスのデータ取得時にフォーマットを揃えるための共通ロギングシステム、各サービスが取得したデータを横断利用しやすい形に整備したデータウェアハウス（DWH）、全社にある大量のデータを一覧化し、データの件数やフォーマット、データオーナーなどを閲覧・検索できるデータカタログ、データ基盤に蓄積されたデータを用いた分析やモデリングを簡単にできるツール群など、必要な機能は無数にあり、現在もその開発は続いている。

ここでも重要なのは、例えば、フォーマット標準化のための共通ロギングシステムを各サービスに導入推進する際に、各サービスの都合（個別最適）と全社の都合（全体最適）をバランスさせることである。各サービス独特の要件をすべて含もうとすると標準化にならず、標準を広く厳格にし過ぎると各サービスの需要を満たさせず導入が進まない。ここまでは標準、ここからは個別という線引きをしながら、現実的な計画を行うのがよいと考える。

3) 横断データを用いたデータサイエンス

強力なデータ基盤の上に全社のデータが集まり、横断利用のために必要なエコシステムも整備されたら、成果を出す準備は整ったと言える。

Yahoo! ショッピングを例に、データの横断利用で実際に得られた成果を説明する。2010 年代前半の Yahoo! ショッピングはまだユーザー数が少なく、

そこから得られるデータも少なかったため、ビッグデータの力でレコメンドを改善しようとしてもコールドスタート問題が発生してしまうことは前述の通りである。一方、ウェブ検索やニュースなどはすでに数千万人のユーザーが頻度高く利用しており、初めて Yahoo! ショッピングに訪れたユーザーについても、過去のウェブ検索などの分析から興味関心を推定することが可能な状態であった。

そのコンセプトを元に開発を行い、Yahoo! ショッピングの利用回数が少ないユーザーがサイトに訪れた際に、ウェブ検索のデータで推定した興味関心を元にレコメンドを行った結果、4.5 倍の CTR を記録した。これはデータの横断利用によりユーザーのニーズにマッチした改善ができたことを証明している。

ヤフーではこれ以外にも数多くのデータの横断利用を行い、総じて大きな成果を上げている。

4.3. プライバシー保護のアプローチ

プライバシー保護の重要性は年々増してきている。EU においては、非常に厳しいと言われる一般データ保護規則（GDPR）が 2018 年に施行され、日本の個人情報保護法も 3 年ごとという短いサイクルで見直しがされるようになった。そして、世界では 100 億円、1000 億円という単位の罰金が課されたというニュースも枚挙に暇がない。

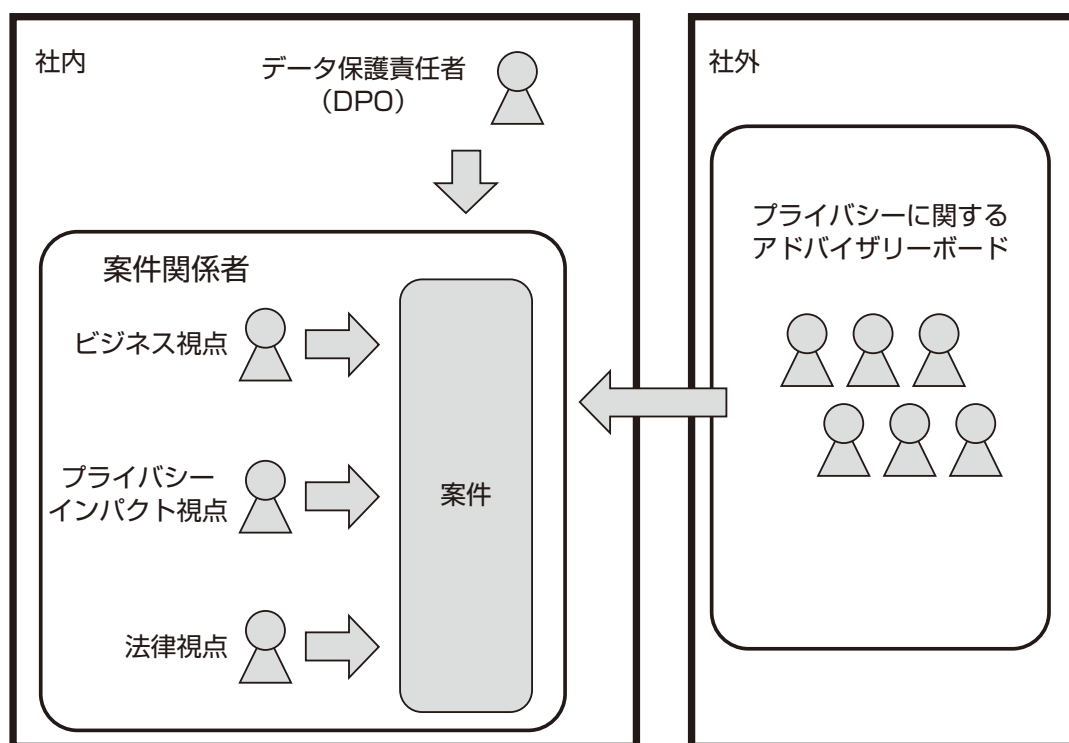
この背景として、データが世の中に及ぼす影響が「質」「量」とともに桁違いになったことが一因であることは間違いない。データはほぼすべての経済活動の裏側に存在し、国民がより便利な生活を送るための燃料となる一方、プライバシーの漏洩や差別の助長などにも繋がる危険な一面を持っている。しかし、危険だと言ってデータを使わない世界に戻ることはなく、不可逆である。企業はユーザーからデータを使わせてもらっていることを強く認識した上で、プライバシー保護を大前提として、ビジネスを遂行する必要がある。

ヤフーにおいても、ユーザーの安心安全を最重要と位置づけ、プライバシー保護やセキュリティ向上に取り組んでいる。ここで大事なのは、バランスと客観性だと考える。

まずバランスとしては、ユーザーの安心安全が第一なことは揺るがないが、プライバシー保護のもとにデータを一切使わないなどの過度な内部規制を行うと、ユーザーが享受する便利なサービスを阻害する可能性がある。そのため、ビジネスと法律、プライバシーインパクトを担当する各部門が観点を持ち寄って議論し、最終的に CDO が総合判断を行うことでバランスを取ることを目指している。

次に客観性としては、2020 年からデータ保護責任者（DPO）を新たに設置した。DPO はより専門的にプライバシーを始めとしたユーザーの個人情報

保護に関する活動を行い、社内の事業部門や経営陣との利害関係がなく、独立した客観的な立場からデータ保護について助言・監視、評価を行うものである。また、さらに客観的に自社の取り組みをチェックすべく、2019 年から外部有識者で構成された「プライバシーに関するアドバイザリーボード」を設置している。アドバイザリーボードでは、ヤフーのプライバシーにかかる施策やプロセスが、ユーザーや社会の期待に応えるか客観的かつ厳格に評価し、より良いものにしていくための助言を行っている。



5. まとめ

ここまで、ヤフーにおけるデータ利活用や、その過程でぶつかった課題、その課題に対する解決策などの事例を解説してきた。これらはヤフーの事例であるが、企業全般に当てはまるものも多いと考える。

私が日々データを扱う中で強く感じるのは、データが持つポテンシャルの大きさだ。世界全体でデータはすでに大きな力を発揮している一方、まだそのポテンシャルは発揮しきっていない。さらに言うと、データを用いて成果を出している企業はあるが、絶対的に成功している企業はなく、まだ優勝は決まっていなと感じる。日本の企業にもチャンスはある。

ヤフーも、まだ理想には程遠い。例えば、2021年春からヤフーを含むZホールディングスグループにLINE株式会社が加わり、ヤフー・LINE・PayPayを中核とした日常生活に欠かせないサービスグループとなった。本来、IDやデータをシームレスに連携することで、ユーザーの生活をもっと豊かに楽しくできるはずであるが、2021年秋の本稿執筆時点では、まだ十分なシナジーは出せていない。

しかし、私は悲観していない。企業を超えたデータの横断利用は、本稿で説明した社内のサービス横断よりも技術的にもプライバシー的にも難易度が高いが、本質的には同じことであり、達成できたときのインパクトは疑いがないものなので、愚直に邁進するだけである。

まとめに代えて、私の思う「データ五箇条」を記載する。

その3以降はここまでに記載した内容と重なるものである。その1とその2は本稿では触れていないが、そもそもの前提として重要なので、ここで記載した。

その1—データやAIを正しく理解する

データやAIには魔法的な響きがあるらしい。メディアの取り上げ方のせいなのか、AIを入れるだけで何かすごいことが起こると誤解している人は一定数いる。工作機械や一般のソフトウェアと同様、データやAIは道具でしかないが、うまく使えば非常に強力な道具となり、ビジネスを一段高いレベル

に引き上げられる。

その2—適切な問題設定を行う

何をしたいか何を解決したいかを固めず、データやAIの導入が先行するというのは、よくある失敗パターンである。問題によってはAIではなく手作業でやった方がよいものもあり、データやAIでも解けない問題は、解けない。まずは問題設定をした上で、それに適した手法を選び、データを選ぶという順番がよい。

その3—リーダーの思いが重要

データ利活用を成功させるためには、組織再編や大きな投資が必要となる場合が多い。しかも、短期的に効果が出ることは稀であり、ある程度必要なものが揃うまで、地道かつ地味な作業が続く。トップマネジメントがデータやAIを正しく理解した上で、我慢して応援し続けることが、成功の必要条件となる。

その4—集約と分散のバランスを取る

組織編成や標準化などデータに関する様々な活動で、個別最適と全体最適の衝突が発生する。最適な形は業種・業態・企業規模・成熟度などによって変わるものであり、しかも文脈で変わることもある。要するに正解はない。高い視点を持ちつつ、利害関係者の意見を聞き、柔軟にバランスを取ることが重要である。

その5—データはユーザーのために

ヤフーを含め、企業側はユーザーにデータを使わせてもらう立場である。企業は、預かっているデータやプライバシーを安心安全にしっかり守り、ユーザー側が選択権を持つ状態を大前提とした上で、データ利活用はユーザーの便利で楽しい生活のためであることを常に忘れず、ビジネスを行う必要がある。

ヤフーはデータに関するコンサルティングも行っているので、お気軽に相談いただけると幸いです。
(<https://ds.yahoo.co.jp>)

注

1 1996年はインターネット世帯普及率、2000年以降はインターネット人口普及率

参考文献

首相官邸 ペンチマーク集(2001年秋 改訂版) (<https://www.kantei.go.jp/jp/singi/it2/others/0111benchmark.html>, 2021年10月29日最終閲覧)

総務省 情報通信白書 令和2年版 (<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/html/nd252120.html>, 2021年10月29日最終閲覧)