

# AI時代のデータ活用は「DataOps」で

～データ活用ライフサイクル管理により業務の改善・高度化を目指す～

データ分析の結果やAI(人工知能)が業務オペレーションに深く関与すればするほど、データ分析から業務活用までのプロセスを効率化して加速させることや、業務活用からフィードバックを得て継続的にデータ活用の取り組みを改善することが重要になる。その実現に向けたコンセプトが「DataOps」である。



瀧本 秀典

日鉄ソリューションズ株式会社  
技術本部 システム研究開発センター  
データ分析・基盤研究部  
データ分析グループリーダー 統括研究員



高木 健一

日鉄ソリューションズ株式会社  
ソリューション企画・  
コンサルティングセンター  
情報系グループリーダー (※所属、役職は取材当時)

## AIの業務活用が進展するにつれ データ活用に新たな課題

DataOpsは、図1のようにデータの蓄積、分析、活用、改善というデータ活用ライフサイクルを高速で回しながら、事業への貢献を持続的に高めていく。

主なポイントは、①データ分析から業務活用にかけての効率化により事業貢献のアジリティを高めること、②業務活用からのフィードバックによってデータ蓄積・分析プロセスを改善すること、③データ分析活動のトレーサビリティを確保して過去の分析資産を再利用可能とし、改善プロセスを効率化すること、である。

さらに、データ活用ライフサイクルの反復を通して事業部門や情報システム部門の協働を推進し、改善・成長のためのアイデアを創出する狙いもある。

これらのポイントはどれも重要だが、AIの活用が本格化する今後を見据えたときに注目してほしいのは②の「改善」プロセスである。DataOpsを推進すればするほど、データ分析の結果やAIの予測モデルを活用する業務が増えていく。そのような業務では、データ分析や予測モデルの質が向上すれば事業への貢献が増える半面、劣化すると思わぬ損失を生む原因になり得るからだ。

データ活用のプロセスを業務の側面から描いた図2を見てほしい。一般的な

データ活用の流れは、①市場や業務の現状を「観察」し、②そこから原因や将来動向を「推測」、③対処方法を「決定」し、④実際に業務の現場で「実行」に移す、という流れになる。

現在は「観察」をBIツールで効率化し、「推測」の一部をAIの予測モデルに担わせる、という状況だが、企業はより多くのデータを使って、観察、推測、決定、実行のすべてのプロセスをITで効率化することを期待している。さらには、今まで見えなかったものをデータで可視化し、単なる効率化を超えたデータ活用を求めている。

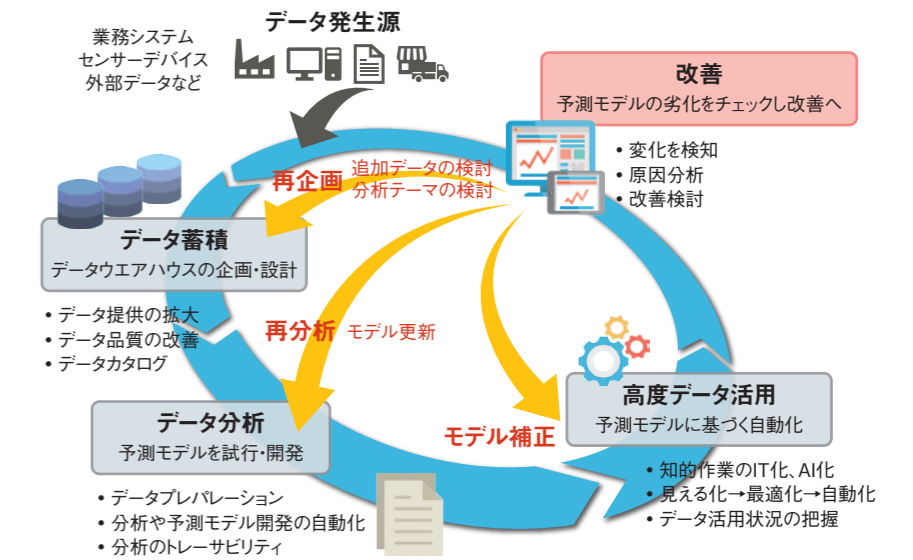
EC(電子商取引)サイトのレコメンデーション・エンジンに例えると分かりやすいだろう。サイト訪問者の閲覧行動を観察しながら求めているものを推測し、場面(閲覧しているページなど)に応じて最適な「おすすめ商品」を決定してサイト訪問者に提示(実行)する。このような全プロセスにわたる自動化(あるいは半自動化)を、幅広い業務に適用していこうというわけだ。今後、この動きはより活発になっていくだろう。

ただし、全プロセスで自動化が進むと、新たな課題が生じる。市場や事業環境が変化した時に、予測モデルの精度が劣化することだ。劣化に気づかないと、期待通りの成果を出せないばかりか、知らぬ間に損失を出している可能性もある。

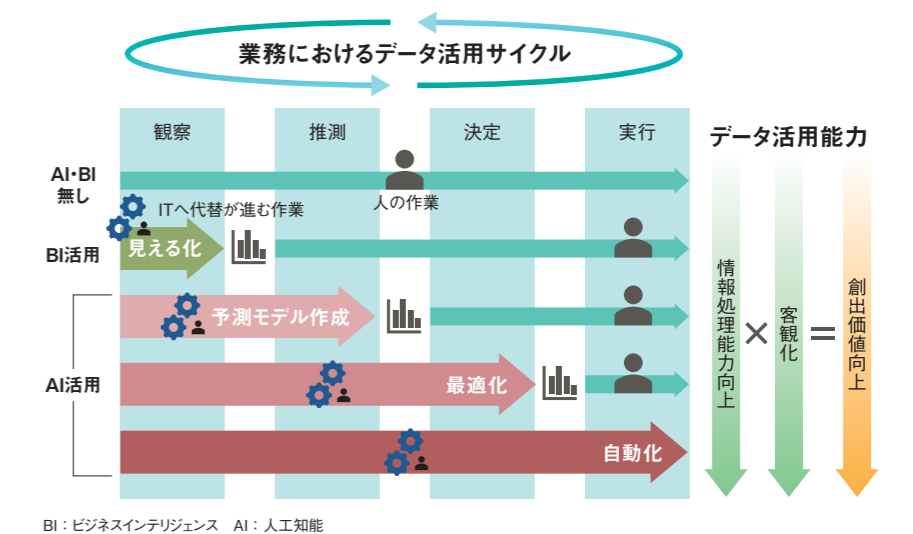
だからこそDataOpsでは、事業環境の変化に対応できるよう、改善プロセスが重要となる。予測値と実績値をモニタリングし、かい離を検知したら原因分析と改善(モデル補正、再分析、再企画)を始める必要がある。

システム研究開発センターでは、このようなモデルの改善プロセスを支援する仕組みを開発している(図3)。主な機能は、変化検知とアラート、予測値の精度を下げた原因の分析支援、機械学

■図1 データ活用ライフサイクル



■図2 AIの活用により、急ピッチで高度化する業務のデータ活用



習に基づく予測値の自動補正などだ。

## データ分析の課題を克服する セルフサービスの最新ツール群

ここまではデータ活用レベルを維持・向上させるための改善プロセスについて説明してきたが、その他のプロセスについてもDataOpsの実現に向けて効率化を図り、データ活用ライフサイクルを高速で回していくことが重要だ。

データ活用の範囲は拡大し、それに伴いデータ分析へのニーズも旺盛であるが、従来型のBIツールが出力するレ

ポートや一握りの専門スタッフによる分析結果に頼る現状では、柔軟性もスピード感も足りていない。そのため、業務現場のユーザー自身が柔軟かつ効率的にデータを分析することが期待されている。

しかし、従来のBIツールや分析ツールでは、これらのニーズにはフィットしづらく、より直感的なユーザーインターフェースやツールとしての使いやすさなどが必要である。

また、データ分析に使うデータは、情報システム部門が基幹システムなどか

■図3 予測モデルの運用状況を監視する仕組み



- モデルの精度やデータ傾向の変化を監視し、その原因の分析をサポート
- 上記に基づき予測値を自動補正

あり、「このデータは使える、使えない」といった判断を簡単にできる。また、このデータに詳しい人は誰かを調べられるとか、よく使われているデータをランキング表示してくれるなど、データ分析の入り口でユーザーが悩みそうなことを手助けしてくれる。

データの加工をユーザー自身が実行できる「データプレパレーション」のツールも出てきた。データ分析に使う情報系データベースは、加工なしでも使えるようにしっかり設計されていることが前提となるが、それでも分析内容によっては少しか加工したい場合がある。そういう時、分析ツール側で対処できることは少なく、情報システム部門に加工を依頼すると時間がかかる。この問題をユーザー自身の手で、かつその場で解決できるようにするのがデータプレパレーション・ツールだ。SQLの知識がなくても、クリック操作でデータを加工できる。

**DataOpsでは分析の反復が前提  
分析履歴を残して改善を容易に**

これら最新のツールを情報系システムに組み込んだデータ活用システムの構成例は、図4のようになる。既存の情報系データベースを基盤として、データを分析・活用する側の機能を強化する形だ。

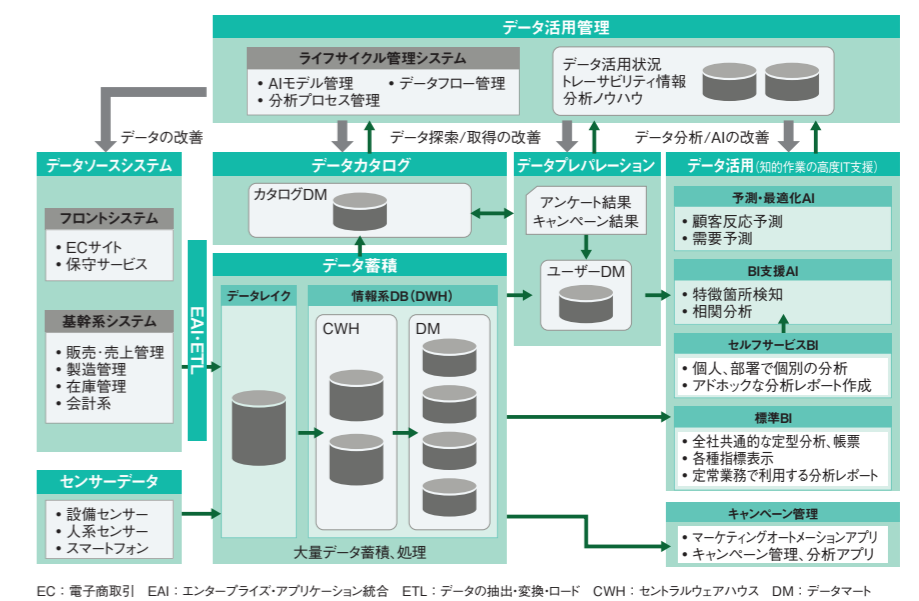
一方、DataOpsのマネジメント面では、データ分析をツールで自動化してだけでなく、改善のために「分析を繰り返しやすいようにすることも重要である。

データ分析はかなり属人的な作業なので、最初の分析から時間が経つと、担当者本人でも分析過程を忘れてしまいがち。これでは改善プロセスのスピードが上がらない。さらにその分析担当者が異動になってしまったら、次の担当者は分析をゼロからやり直す羽目になるかもしれない。

この問題を避けるには、各種ツールの履歴取得機能の活用を含め、データの提供から分析、活用までのトレーサビリティを管理することが重要である。データプレパレーション・ツールや機械学習自動化プラットフォームには、過去にどんなデータ変換・加工・機械学習をしたのか、履歴に残す機能がある。後から録画を再生するように同じ処理を再現できる。

当社のツールでは、統合データ分析環境のData Veraci(データヴェラチ)が分析時の試行錯誤(仮説検証)の履歴を管理し、分析の意図や経緯を残す仕組みを提供している。こうした履歴機能をベースに、分析の繰り返しに対応できる手順やルールを整備するべきだ。

■図4 データ活用システムの構成例



EC：電子商取引 EAI：エンタープライズ・アプリケーション統合 ETL：データの抽出・変換・ロード CWH：セントラルウェアハウス DM：データマート

ら抽出し、情報系データベースとして用意してきた。要件定義をして、設計、実装するという手順を踏むため「いろいろな分析のためにデータを少しだけ加工したくても、情報システム部門に依頼すると非常に時間がかかる」といった課題があり、分析プロセスのスピードを落としている。

この課題の解消に向けて、現在は「セルフサービス化」や「データ分析・AIの民主化」といった潮流が注目されており、現場ユーザー自身がさらに簡単かつ効率的にデータを準備・分析できるツールが各種登場してきている。

データ分析プロセスでは、「セルフサービスBI」と呼ばれるツールや、AIモデル構築のための機械学習自動化プラットフォーム「DataRobot」などが注目されている。

セルフサービスBIツールは、表計算ソフトではデータ量が足りないが、既存のBIツールでは難しいと感じる人でも使える。例えば、あるセルフサービスBIツールにはGoogleの検索窓のようなものがあり、そこに知りたい情報と集計

方法を入力するとBIツールがビジュアル化してくれる。「Total 売上 パイ 月単位および店舗地域」と入力すれば、毎月の売上高の推移を地域別にグラフ表示する、といった具合だ。

また、DataRobotのようなAIモデルの自動構築プラットフォームを用いると、専門家でなければつくれなかった高度な予測モデルを、誰でも機械学習用のデータを入力するだけで開発できるようになる。専門家でも数週間を要するような予測モデルの開発を30分ほどで完了できるため、データ分析・活用プロセスを劇的に効率化することができる。

セルフサービスBIやDataRobotを使いこなすには「データカタログ」も必要になる。どんなデータがあるかを知らなければ、ユーザーは分析を始められないからだ。データカタログツールは、情報系のデータベースをクローリングしてカタログをつくるだけでなく、データ分析に役立つ各種の機能を備えている。例えば、データの数や分布などの統計情報を調べるプロファイリング機能が

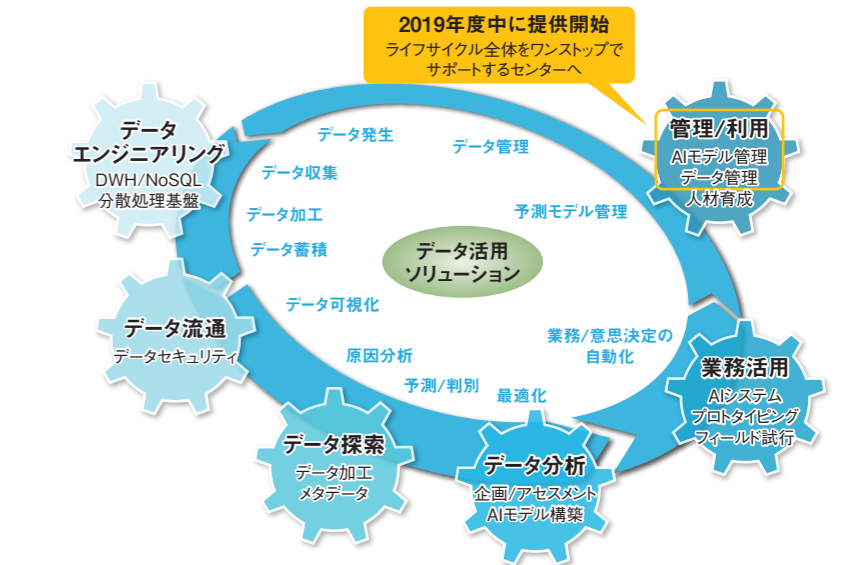
**Data Leverage Centerを開設  
DataOpsをワンストップで支援**

冒頭で述べたように、DataOpsに厳密な定義はない。「データ活用プロセスの自動化・高速化によって事業貢献や業務改善の効果を最大化する」という目的のもとで推進する取り組みなので、ツール、仕組みづくりだけでなく、データ分析教育や風土・文化の醸成も必要となる。

そこで当社は2019年4月に、データ活用ライフサイクルの推進を目指すお客様への相談窓口として、システム研究開発センター内に「Data Leverage Center (DLC)」を開設した。これまで個々の領域で提供してきたサービスを体系化し、データ活用の検討段階から実現までをDLCがワンストップで支援するとともに、プロセス推進支援を通してお客様の風土・文化の醸成にも寄与したいと考えている(図5)。

本稿では詳しく説明するスペースがなかったが、システム研究開発センターはデータ活用ライフサイクルの各局面で発生する課題を解決するためのソ

■図5 Data Leverage Centerが提供するデータ活用に向けた支援メニュー



リユースやサービスを開発、提供している。現実世界の人やモノの動きを仮想環境で再現するデジタルツインのプラットフォームや、データから暗黙知のプラットフォームや、データから暗黙知を学びベテランの分身となるスマートマシン(人と共に働くAI)、計画立案業務を最適化するソリューションなどがある。いずれも従来は自動化しにくかった業務を、AIなどの活用により大幅に効率化するものだ。

既にデータエンジニアリングの領域

からデータ分析、業務活用までの支援メニューは確立しており、ここ数年で30件以上の実績がある。今後さらに、DataOpsの実現に必要なマネジメント面の支援メニューを拡充していく。

2019年度中には、ライフサイクル全体を支援するプラットフォームも提供予定である。個別に開発してきたソフトウェア群の組み合わせ活用も含め、関係者間の協働や改善・高度化を促す基盤となる。