

DX推進に必要なデータマネジメントとは

～部門を横断したデータ活用の実現に向けて～

目 次

第1章 国内企業におけるDXの現状と課題	1
DXに向けた体制・プロセスの整備状況	1
部門横断的なデータ活用を阻む課題	2
第2章 DX推進に必要なデータマネジメントとは	4
データマネジメントによる部門横断的なデータ活用の実現	4
データマネジメント・ソリューションの構成要素	6
第3章 提言	10

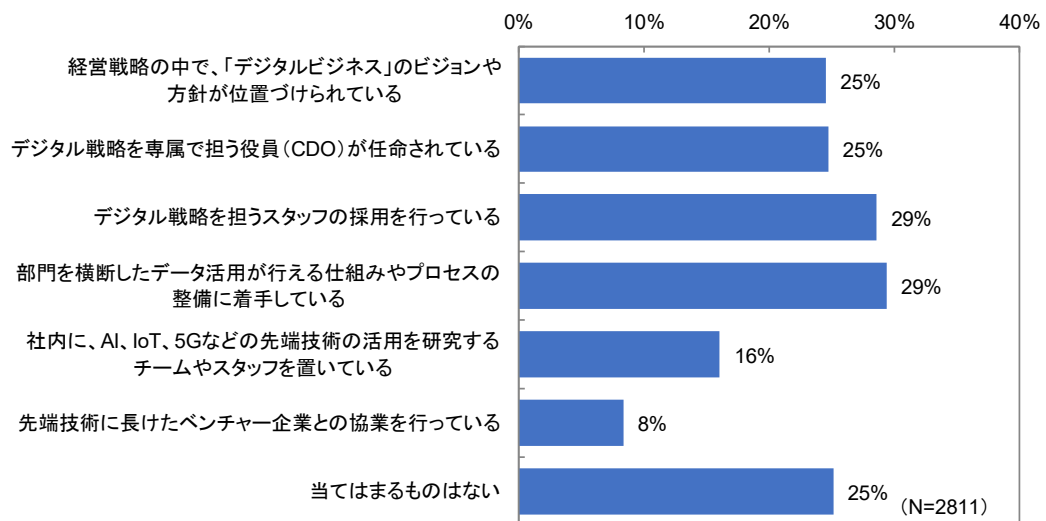
第1章 国内企業におけるDXの現状と課題

デジタルトランスフォーメーション（DX）を推進する際には、部門横断的なデータ活用が必須となる。しかし、現時点では、企業の多くは部門横断的な活用が困難な状況にある。この課題を解決するためには、分散しているデータを統合的に活用するためのIT基盤の整備が必要である。

DXに向けた体制・プロセスの整備状況

DXを推進するうえでは、単に技術を導入するだけでなく、そこから価値を引き出すための体制やプロセスの整備が不可欠である。ITRが2019年8～9月に実施した「IT投資動向調査2020」では、この点について7つの項目をあげ、回答者に自社の実施状況を問うた（図1）。

図1 デジタル変革に向けた体制・プロセスの整備状況（複数回答）



出典：ITR「IT投資動向調査2020」

7つの項目のなかで実施しているとの回答割合が最も高かったのは、「部門を横断したデータ活用が行える仕組みやプロセスの整備に着手している」と「デジタル戦略を担うスタッフの採用を行っている」であり、ともに3割に迫る水準となった。

ビジネスの変革が求められるDXにおいて必要な人材を新たに求めることは当然といえ、後者の項目が高い水準になったと考える。一方、前者の項目については、DX

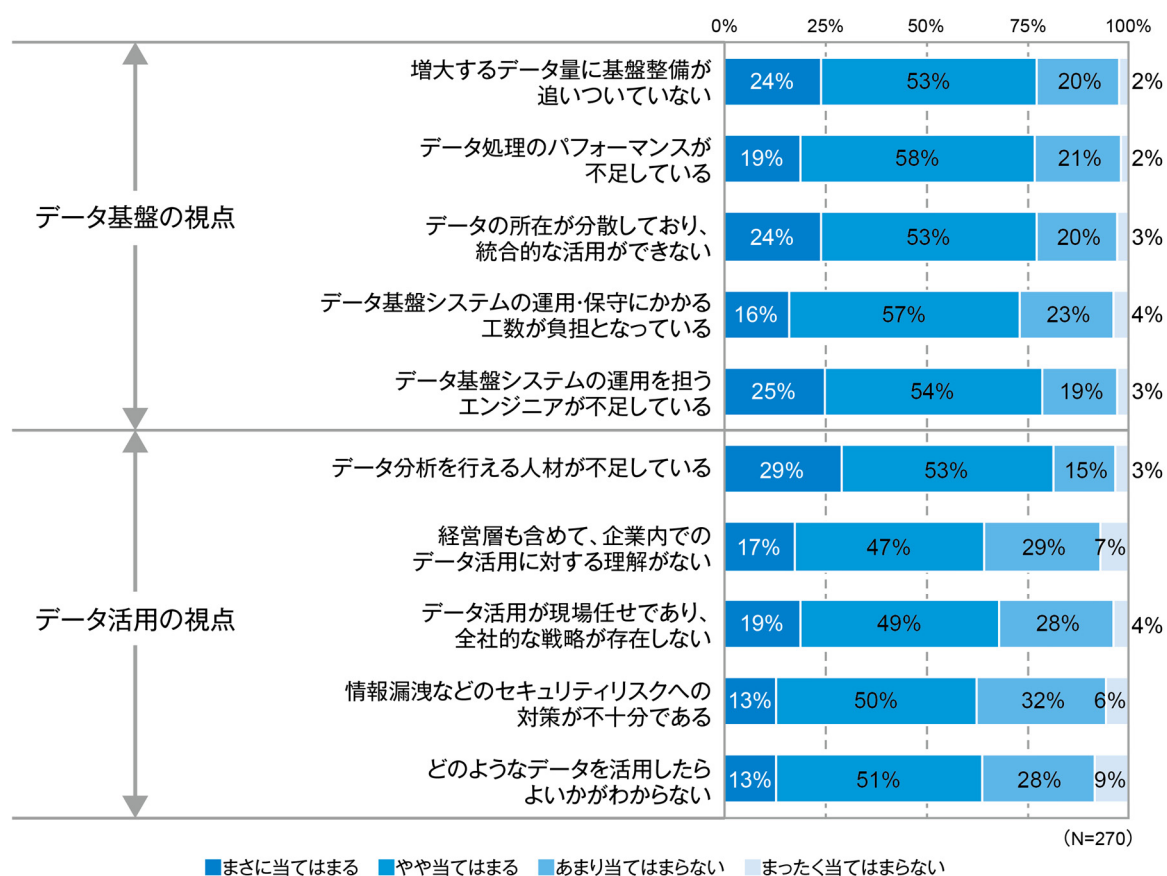
の推進では部門横断的なデータ活用が重要視されているものの、その仕組みやプロセスについては、まだ整備に着手した段階にあることが示された。

部門横断的なデータ活用を阻む課題

では、部門を横断したデータ活用が行える仕組みやプロセスの整備においては、どのような課題があるだろうか。

ITRでは、2018年に行ったデータ活用に関する意識調査のなかで10項目の典型的な課題を取り上げ、回答企業がそれらの課題にどの程度該当しているかについて確認した（有効回答270件）。データ基盤に関わる5項目と、データ活用そのものに関わる5項目から成る（図2）。

図2 データ活用に関わる課題認識



出典：ITR（2018年4月調査）

経営層の理解不足、活用戦略の欠如、セキュリティリスクへの対策不足といったデータ活用の視点での課題以上に、データ量の増大への対応、処理パフォーマンス不足、データの所在の分散といったデータ基盤に関わる課題で悩んでいる企業が多いという興味深い結果となった。データ基盤に関わる課題では、「増大するデータ量に基盤整備が追いついていない」「データの所在が分散しており、統合的な活用ができない」「データ基盤システムの運用を担うエンジニアが不足している」の3項目については、全体の約4分の1（24%ないし25%）の企業が「まさに当てはまる」としており、課題として特に強く認識されている。

「増大するデータ量に基盤整備が追いついていない」は、各企業がビッグデータへの対応を進める過程において、管理対象となるデータのボリュームが急激に増大することに起因すると考えられる。「データ基盤システムの運用を担うエンジニアが不足している」については、高度な専門性を持った人材の不足という全般的な傾向を反映していると考えられ、このことは、データ活用の視点での「データ分析を行える人材が不足している」について全体の約3割（29%）の企業が「まさに当てはまる」としたことにも表れている。

「データの所在が分散しており、統合的な活用ができていない」という課題は、前項で示した「部門を横断したデータ活用が行える仕組みやプロセスの整備に着手している」企業が3割近くに上っている点に直接関連しているといえよう。つまり、DX推進の一環として「部門を横断したデータ活用が行える仕組みやプロセスの整備」を行う過程において、「データの所在が分散しており、統合的な活用ができない」という課題を解決するIT基盤の構築が急務であるといえる。

第2章 DX推進に必要なデータマネジメントとは

分散したデータを統合的に活用するためのIT基盤の構成を考える際に参考にすべきなのが、DMBOK（Data Management Body of Knowledge）に定義されているようなデータマネジメントの考え方である。本章では、DMBOKの概要と、分散したデータの統合的な活用に必要なIT基盤の構成を解説する。

データマネジメントによる部門横断的なデータ活用の実現

分散したデータを統合的に活用するための基盤を構築するには、データマネジメントの考え方が必要である。データマネジメントとは、データを適切に管理するために必要な作業の総称であるが、具体的な内容についてはさまざまな定義が存在する。なかでも、最も標準的かつ一般的に使用されているのが、DMBOK（Data Management Body of Knowledge）である。

DMBOKは、非営利団体DAMA-I（Data Management Association International）が作成したデータマネジメントに関する知識体系であり、その内容は「データマネジメント知識体系ガイド」として現在第二版が発刊されている。DMBOKでは、データマネジメントを実現するための必要な作業として11項目が定義されており、それぞれの名称と定義は図3の通りである。

DMBOKは、データマネジメントに必要な作業の全体像を捉えるうえで最適な体系であり、分散したデータを統合的に活用するためのIT基盤の構築を目指す企業は、DMBOKの内容を概略にせよ理解しておくことが有益であると考ええる。

しかし、DMBOKがカバーする範囲は非常に広く、一般の企業が短期間に全ての項目を実現することは現実的ではない。したがって、DX推進に必要な「部門を横断したデータ活用が行える仕組みやプロセスの整備」を実現することを目的として、図3にあげた11項目のなかで、特に必要な項目に集中して取り組むべきであろう。

部門横断的なデータ活用を実現するには、分散したデータを統合的に活用するための基盤を構築する必要があるが、「データ統合と相互運用性管理」によりシステム間のデータの移動と統合を実現することが第一歩となる。また、データの現状や活用要件の変化に対応させながらデータ統合を維持するためには「メタデータ管理」が必要である。さらに、データ統合の結果、データ品質が低下する場合に備えて、「データ

品質管理」も必要と考える。以下では、これらの3項目について個々に説明する。

図3 DMBOKにおけるデータマネジメントの11項目

	名称	定義
1	データガバナンス	データ管理の計画を立て、実行を監視し、徹底させることで、データ資産の管理を統制する。
2	データアーキテクチャ設計	企業のデータニーズを明確にし、それを満たすデータの構造と計画を設計する。
3	データモデリングとデザイン	データ要件を記述し伝えるために、概念、論理、物理それぞれのデータモデルを定義する。
4	データストレージとオペレーション管理	データモデルに基づいて、物理的なデータベースを構築し、運用管理する。
5	データセキュリティ管理	セキュリティポリシーや手順を定義、実行し、データと情報資産に対して適切な認証と権限付与を行い、アクセスを制御し、監査を行う。
6	データ統合と相互運用性管理	単一のアプリケーション・組織内および複数のアプリケーション・組織の間におけるデータの移動と統合を管理する。
7	ドキュメントとコンテンツ管理	データベースに保存されないデータや情報の取得、保存、アクセスと利用の制御を行う。
8	参照データとマスターデータ管理	業務遂行に必要な共通のマスターデータを作成し、組織間で共有する仕組みを提供する。
9	データウェアハウジングとビジネスインテリジェンス	意思決定を支えるデータを提供して、レポート作成、分析に携わるナレッジワーカーを支援するため、計画を立案し、実行し、統制する。
10	メタデータ管理	さまざまなデータソースのメタデータを収集、統合し、それにアクセスするための標準的な方法を提供する。
11	データ品質管理	データ品質の標準を設定し、それを維持するための仕組みを、データを生成、変換、保存するプロセスに組み込み、データ品質を監視する。

出典：「データマネジメント知識体系ガイド 第二版」DAMA International編著（日経BP発行）の内容を基にITRが作成

1. 「データ統合と相互運用性管理」

部門横断的なデータ活用の実現のために、「単一のアプリケーション・組織内および複数のアプリケーション・組織の間におけるデータの移動と統合」は必須といえる。しかし、その実現方法は多種多様であり、データの現状や活用の要件に応じて選択し、組み合わせる必要がある。ここでいうデータの現状とは、ソースとなるデータの保存形式や格納されているシステム環境のことであり、これが異なればそのデータへのアクセス方法も異なる。また、リアルタイムの更新が必要なのか、あるいは定期的なバッチ処理による更新でよいのかによっても、データの移動と統合には異なる方法が必要となる。

2. 「メタデータ管理」

分散したデータの所在が多数に及ぶ場合、「メタデータ管理」は、上述の「データ統合と相互運用性管理」実現での前提条件となる。データの移動と統合は、継続的に行われる一方で、データの現状や活用の要件は時間の経過とともに変化する。このような変化に対応しながら、「データ統合と相互運用性管理」を継続するためには、「さまざまなデータソースのメタデータを収集、統合し、それにアクセスするための標準的な方法を提供」する「メタデータ管理」が必要となる。

3. 「データ品質管理」

分散したデータを統合した結果、データの品質が低下する場合がある。例えば、異なる場所にある顧客データを統合する際に、同一の顧客であっても、姓と名の間の空白の有無や住所における番地の表記方法などのデータの入力形式が異なり別の顧客として判定された結果、同一顧客のデータが重複して存在するようなケースがそれにあたる。データ統合によって品質の低下が予想される場合、「データ品質の標準を設定し、それを維持するための仕組みを、データを生成、変換、保存するプロセスに組み込み、データ品質を監視する」ことが重要である。

データマネジメント・ソリューションの構成要素

次に、分散したデータを統合して活用するために必要なIT基盤の構成を考える。

ここでは、前項であげた「データ統合と相互運用性管理」「メタデータ管理」「データ品質管理」の3項目を実現するために必要な機能をいくつかのコンポーネントとして整理する。

「データ統合と相互運用性管理」を実現するためには、異なるデータソースへのアクセスが必須であり、この機能を提供するのがデータの収集を行う「Data Entry」コンポーネントである。また、データを統合する方法には、物理的なデータの移動を行う方法と、仮想的にデータを連携させる方法の2つがあり、前者の機能を提供するのがデータの準備を行う「Data Prep」コンポーネント、後者の機能を提供するのがデー

タの連携を行う「Data Integration」コンポーネントである。これに、「メタデータ管理」を実現するためのデータの特典機能を提供する「Data Identification」コンポーネントと、「データ品質管理」を実現するためのデータ品質の維持機能を提供する「Data Quality」コンポーネントを加えた計5つのコンポーネントによって、部門横断的なデータ活用の実現に必要な機能を満足するデータマネジメント・ソリューションが構成される。各コンポーネントの内容を以下に順に説明する。

1. Data Entry

「Data Entry」は、データの保存形式や格納されているシステム環境が異なる、さまざまなデータへのアクセスを可能にする機能を指す。

DMBOKの「データ統合と相互運用性管理」を実現するためには、データの現状が異なるものに対するアクセスは必須であり、「Data Entry」はその機能を提供する。

「Data Entry」の機能を持つパッケージ製品やクラウドサービス（以下、製品・サービスと略す）は、ファイル転送ツール、ODBC（Open Database Connectivity）のようなアダプタ、異なる文字コードを相互に変換するコード変換ツールなど多岐にわたり、データの現状に応じて複数を組み合わせる必要がある。

2. Data Identification

「Data Identification」は、各所に分散したデータの名称、格納場所、バージョンといったメタ情報を集中的に管理することで、必要なデータの特典を容易にする機能を指す。

「Data Identification」は、データの現状や活用要件の変化に対応しながら、「データ統合と相互運用性管理」を継続するための機能を提供し、DMBOKの「メタデータ管理」を実現する。

「Data Identification」の機能を持つ製品・サービスは、IT市場において一般に「メタデータ管理ツール」や「データカタログ」と呼ばれている。

3. Data Integration

「Data Integration」は、ソースデータを持つシステムと活用する側のシステムを

連携させることで、仮想的にデータを統合する機能を指す。

データ活用の対象となるソースデータのデータ量が膨大だったり、リアルタイム性が要求されたりするような場合、DMBOKの「データ統合と相互運用性管理」を実現する方式として物理的なデータの移動を行うことは得策ではなく、「Data Integration」による統合が効率的である。

「Data Integration」の機能を持つ製品・サービスは、IT市場においては一般に「EAIツール」や「データ連携ツール」と呼ばれている。

4. Data Quality

「Data Quality」は、データの品質を向上させるためのデータ加工を行う機能を指す。例えば、名前や住所を標準化するジオコーディング、微妙な差異のあるデータの関連づけを行うマッチングなどがある。

「Data Quality」は、標準として設定されたデータの品質を維持するための仕組みを提供し、DMBOKの「データ品質管理」を実現する。

「Data Quality」の機能を持つ製品・サービスは、IT市場においては一般に「データ品質ツール」や「データクレンジング・ツール」と呼ばれている。顧客データの重複を排除する機能に特化している場合、「名寄せツール」と呼ばれるものもある。

5. Data Prep

「Data Prep」は、データソースからデータを移動し、活用目的に応じたデータ加工を施すことで、物理的にデータを統合する機能を指す。

仮想的なデータが必要となるデータ量やリアルタイム性に関わる要件がない場合は、「Data Prep」を利用した物理的なデータ統合が、DMBOKの「データ統合と相互運用性管理」を実現する一般的な方式である。

「Data Prep」の機能を持つ製品・サービスは、IT市場において、一般に「ETLツール」や「データプレパレーション・ツール」と呼ばれている。

これらの内容を表にまとめると図4のようになる。

図4 データマネジメント・ソリューションを構成する5つのコンポーネント

	名称	機能	DMBOKの項目	IT市場での呼称
1	Data Entry	データの保存形式や格納されているシステム環境が異なる、さまざまなデータへのアクセスを可能にする	データ統合と相互運用性管理	・ファイル転送ツール ・アダプタ ・コード変換ツール など
2	Data Identification	各所に分散したデータの名称、格納場所、バージョンといったメタ情報を集中的に管理することで、必要なデータの特定を容易にする	メタデータ管理	・メタデータ管理ツール ・データカタログ など
3	Data Integration	ソースデータを持つシステムと活用する側のシステムを連携させることで、仮想的にデータを統合する	データ統合と相互運用性管理	・EAIツール ・データ連携ツール など
4	Data Quality	データ品質を向上させるためのデータ加工（ジオコーディング、マッチングなど）を行う	データ品質管理	・データ品質ツール ・データクレンジング・ツール ・名寄せツール など
5	Data Prep	データソースからデータを移動し、活用目的に応じたデータ加工を施すことで、物理的にデータを統合する	データ統合と相互運用性管理	・ETLツール ・データプレパレーション・ツール など

出典：ITR

データマネジメントに関連する機能を提供するベンダーや製品・サービスは数多く存在し、それぞれが提供する機能の範囲や呼称は多岐にわたるため、各ベンダーから提供される情報は簡単に比較できない場合が多い。ベンダー選定を行う際には、本稿で解説した5つのコンポーネントごとに要件を整理し、それに対応した情報提供をベンダーに求める方法が効率的である。例えば、「Data Entry」のコンポーネントでは、実績が豊富で安定性が担保できるベンダーは限られていることから、比較選定に要する時間を短縮することが可能であろう。

第3章 提言

DXを推進する際には、部門横断的なデータ活用が必須となる。しかし、現在、企業の多くではそのようなデータ活用が実現できていない状況にある。この課題の解決に向けては、データマネジメントの考え方を理解し、分散したデータを統合的に活用するためのデータ基盤を構築する必要がある。

データマネジメントの考え方を理解するためにはDMBOKが参考になる。DMBOKは、データマネジメントに必要な作業の全体像を捉えるうえで最適であることから、DXを推進する企業は、その準備段階においてDMBOKの内容を理解することを推奨する。

しかしながら、DMBOKで定義されている項目は11項目と多岐に及ぶため、企業が実際にデータマネジメントを実現するための基盤を構築する段階においては、目的に合わせ優先度の高い項目に絞り込む必要がある。DXに必要な「部門を横断したデータ活用が行える仕組みやプロセスの整備」を実現するには、11項目の中でも、「データ統合と相互運用性管理」「メタデータ管理」「データ品質管理」の3項目に集中して取り組むことが求められる。

さらに、これらの3項目を実現するうえで必要なデータ基盤の構成を考えると、「Data Entry」「Data Identification」「Data Integration」「Data Quality」「Data Prep」の5つのコンポーネントに整理できる。データマネジメントに関連する機能を提供するベンダーや製品・サービスは数多く存在し、全機能を一括して比較することは容易ではないが、「Data Entry」のコンポーネントのように、実績が豊富で安定性が担保できるベンダーが限られる場合もある。DXを推進する企業は、これら5つのコンポーネントごとに要件を整理し、ベンダー選定を効率的に行うことが求められる。

分析／執筆：平井 明夫
text by Akio Hirai

ITR White Paper

DX推進に必要なデータマネジメントとは ～部門を横断したデータ活用の実現に向けて～

C20040127

発行 2020年4月20日

発行所 株式会社アイ・ティ・アール

〒160-0023

東京都新宿区西新宿3-8-3 新都心丸善ビル 3F

TEL：03-5304-1301（代）

FAX：03-5304-1320

本書に記載された全ての内容については株式会社アイ・ティ・アールが著作権を含めた一切の権利を所有します。無断転載、無断複製、無許可による電子媒体等への入力を禁じます。

本書に記載されている会社名、商品名等は各社の商標または登録商標です。
