



経済産業省『デジタル時代の人材政策に関する検討会』

NTT DATA
Trusted Global Innovator

AI時代だからこそ不可欠なデータマネジメントの 重要性ならびに人材育成の必要性について

2024年6月6日

(一社)日本データマネジメント・コンソーシアム[JDMC] 発起人 兼 事務局長

(株)NTTデータバリュー・エンジニア 代表取締役社長

大西 浩史

Trusted Global Innovator
NTT DATA Group **NTT DATA**

INDEX

- 01 会社&団体の紹介
- 02 なぜ、いま、データマネジメントが求められるのか？
- 03 データマネジメントを牽引する人材育成の必要性



01

会社&団体の紹介

会社概要

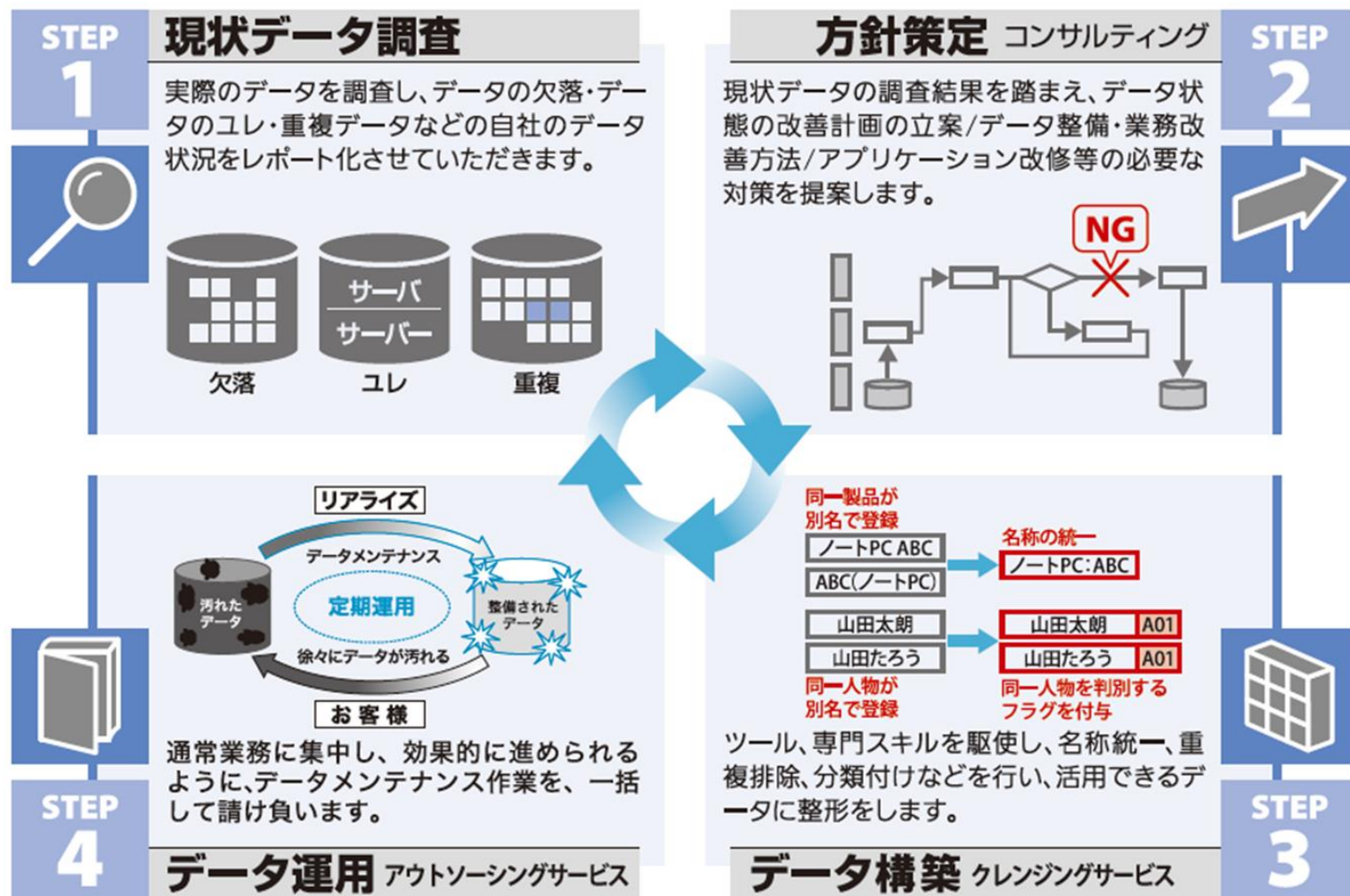
- ☑ 創業者自らが体験した“データの品質が悪いと活用できない”という純粋な問題意識からスタート
- ☑ NTTデータの社内ベンチャーとして1997年に創業(20年以上の事業活動実績)
- ☑ 幅広い領域のお客様に対して“データをビジネスで活用可能にする取り組み＝データマネジメント”を支援し、企業・組織が保有するデータの実態を把握、どうすればデータを活用可能にできるかを熟知していることを強みとしています。

社名	株式会社 NTTデータバリュー・エンジニア (英文名： NTT DATA Value Engineer Corporation)		
代表者	代表取締役社長 大西 浩史	株主構成	▶ 株式会社NTTデータ ▶ アステリア株式会社
設立	2001年(創業:1997年)	ミッション	NTTデータ バリュー・エンジニアは、 データを通じてお客様に寄り添い、 データ成功体験を共創します。
資本金	1億円	事業 コンセプト	遠い将来の絵空事ではなく、ビジネスの 現場の判断・意思決定をより良いものに する～『データで創る一歩先の未来』を 企業理念に掲げ、お客様がデータをビジ ネスに活かすことに貢献します。
所在地	東京都江東区木場5-5-2 CN-1ビル		
従業員数	約100名(協業者含む)		
HP URL	https://www.nttdata-value.co.jp/		

データマネジメントサービス全体像

お客様の現状データ(=実際の値)の『健康状態』を診断・アセスメントし、「どう改善すれば、ビジネス上でうれしい活用ができるか」の『処方箋』を書きます。

さらにデータを健康な状態に『治療』し、良好な体調を維持する『運用』までトータルにサポートします。



主な実績

様々な業界/業種におけるトップクラスの企業様とともに、約1,500件ものデータマネジメント改善およびデータ活用促進のためのプロジェクト実績があり、お客様から高いご評価とご信頼をいただいております。

■Web・コマース事業者



■製造業



■官公庁・教育機関



■通信・インフラ事業者



■サービス業

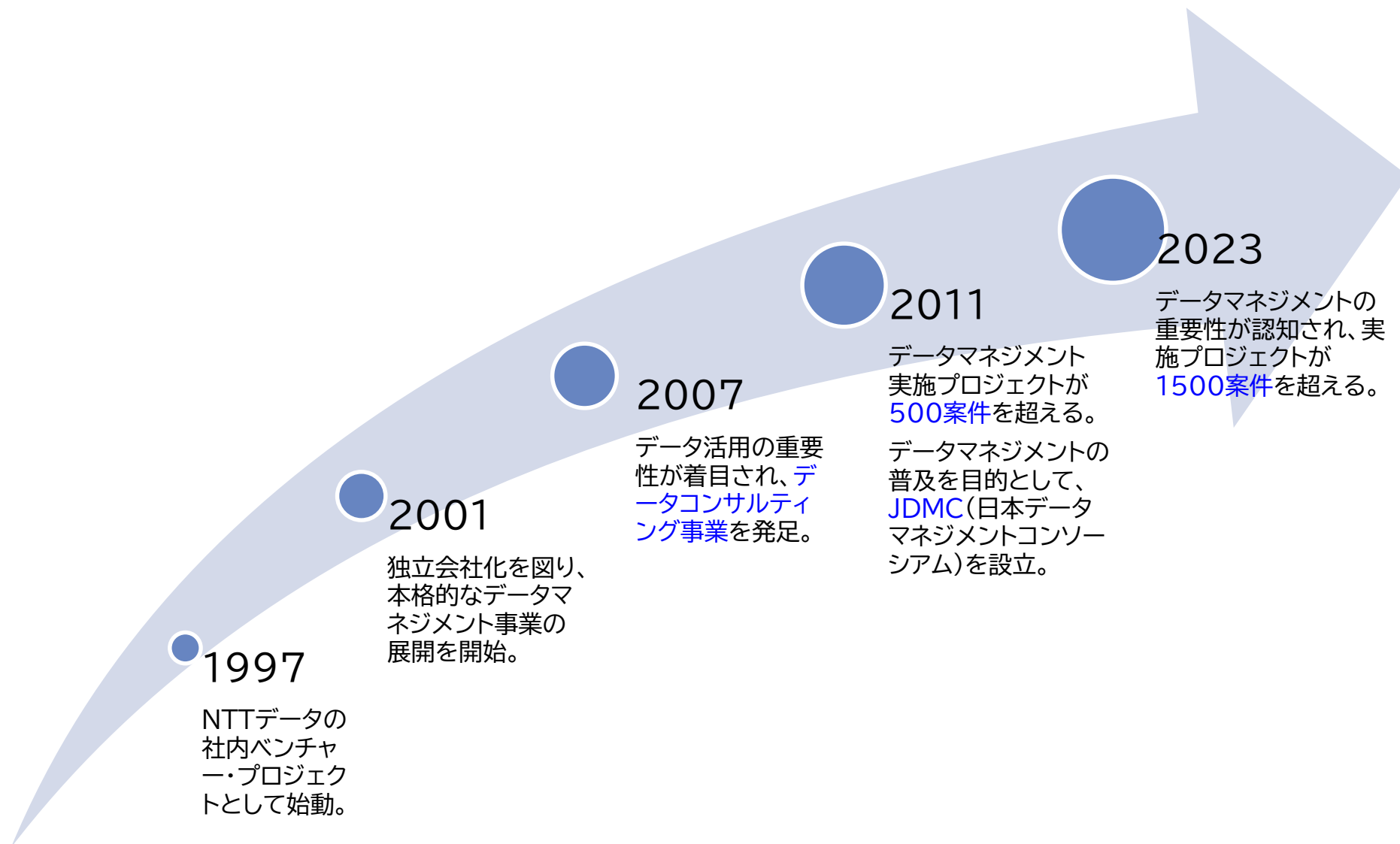


■金融機関



当社の歩み

株式会社NTTデータバリュー・エンジニアは、1997年に創業以来、データマネジメント事業を一貫して手掛け、企業・組織の情報活用を支援しています。



日本データマネジメント・コンソーシアム[JDMC]のご紹介

発起人としてデータマネジメントの普及・啓発・定着化団体の立ち上げに貢献しました。

正式 名称	一般社団法人 日本データマネジメント・コンソーシアム [Japan Data Management Consortium:JDMC]
設立	2011年4月18日(月)
設立 趣旨	情報システムがビジネス戦略の策定・実行に不可欠となっている中、 データマネジメントの問題により、情報システムが経営の羅針盤として機能していない という状況が多く見られる。 本コンソーシアムは、データマネジメントの重要性の普及展開と、実践的なデータマネジメント手法の確立を通じ、 日本企業・組織の国際競争力強化に寄与 する。
主な 活動	① 研究会／コミュニティ ② カンファレンス ③ データマネジメント表彰 ④ 書籍出版・育成動画配信等の各種情報発信

2011年設立時：22社



現在：300社の参加企業

02

なぜ、いま、データマネジメントが求められるのか？

急激な変化の時代～オフィス用品の法人営業の場合～

これまでのノーマル

客先回り



ニーズ発掘・提案

見積・受注



□ リピートオーダー受注のために、お客様オフィスに訪問し、隠れたニーズや消耗品切れを察知→次の商談時に提案・見積のうえで受注を獲得、といった一連の営業プロセスが確立

コロナがもたらした
顧客行動の変容



必ずしも出社が前提でなくなった



重要な会議以外はリモート／オンラインで済ませる

これからのニュー・ノーマル

客先回り

ニーズ発掘・提案



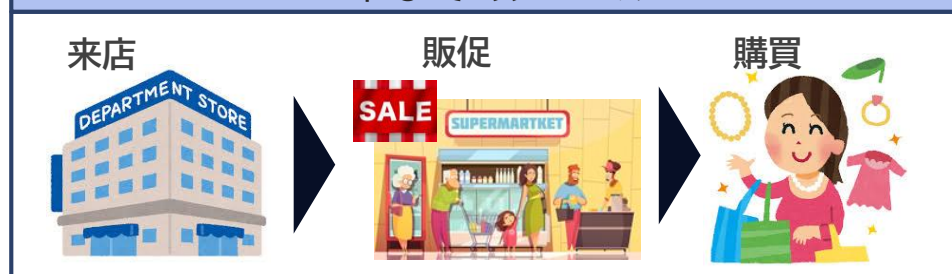
最も重要である『顧客
ニーズの把握』がオン
ライン越しには難しい



- ☑ 在宅ワークが一定範囲で定着化し、これまで定期的に発生していたオフィス消耗品等の需要が減少するため、なおさら既存顧客からの取りこぼしをなくしたい。
- ☑ 手厚い”御用聞き営業”がやりにくいため、顧客の購買特徴に合わせた営業タイミングを見極めなければならない。

急激な変化の時代～小売エリアマネージャーの場合～

これまでのノーマル



□「お客がいっぱい賑わっているお店でショッピングを楽しむ」という日常習慣があり、売り場での対面による接客やPOP等の販促により一定の需要喚起ができるプロセスが確立

コロナがもたらした
顧客行動の変容



一定層は“密”を避ける購買行動で元に戻らない



ネットで買えるものは店に行かずに済ませる購買層の発生

これからのニュー・ノーマル



- ☑ 自社のお客様との良好なリレーションシップをこれまでよりもオンラインで、つまりデータを通じて維持しなければならない。
- ☑ WithコロナでEC化率がさらに高まり、お客様のお住まいと店舗との距離が関係なくなるため、激化する競争の中で他社に奪われないようにしたい。

これまでの“勘と経験”が通用しない時代に

コロナウィルスによる
顧客の行動変容

エネルギーや原材料等の
急激な価格高騰

ウクライナや中東、各地域で
多発している国際紛争

歴史的な円安と反比例した
日本株価の上昇

労使協調した賃上げによる
国内デフレ脱却の機運

“Chat-GPT”に代表される
技術のブレークスルー

これまでの“経験則”や“現場の常識”が
まったく通用しない時代に

先の見通しにくいビジネス環境になればなるほど、『“勘と経験”といった
属人的な判断ではなく、データドリブンな意思決定、データ活用を全社的に
促進したい』という企業経営者のニーズが増えるのは必然といえる

経営者は“第二のオイル”で自社のビジネスを強くしたい

IoT、モバイル等の普及・浸透によって、これまでは把握できなかった最終顧客の購買前後の行動やモノの動きなどのログデータが獲れるようになった。

生活&ビジネスへのITの浸透により、把握可能なヒトやモノの“動き”のデータが爆発的に増大化している

正しいデータに基づいて、マネジメント層がグローバルでの戦略的な会話・対話を行う必要性が増えている

モノからコト消費へ～コモディティ化の時代に顧客から選び続けられるためには顧客をもっと知る必要がある

こうした膨大なヒトやモノの動きのデータを自社の業務で有効に活用したい、さらには、それを自社の商品やサービスに組み込んで価値を高めていきたい、経営者が強く意識するのは当然

多くの組織がデータ活用の上で直面する現状の課題

しかし、「いざ、活用したい」という場面で何が起こるか。

活用対象となる“顧客”や“商品”、“部品”等に関わる基軸データが社内のあちこちに散在し、統合が取れていない、粒度が合わない、精度が悪い、どのシステムにどのデータが存在するのかすらわからない、といった現状に直面する。

メインフレーム、C/S、Web・・・ITの進化とともにシステムが分散化し、無秩序なデータも拡散・大量化した

個別の業務部門ごとに紙処理を電子化するためのITが導入され、データのサイロ化・部分最適化が進行した

情報システム部門は「システム」を全社横断で見ているが、「データ」を全社的にケアする組織はこれまでなかった

各部門の業務上では「回っているデータ」だが、活用するには問題があるデータ、特に部門を横断した新しいインサイトを得ようとする重大な障害ー“Conflict Data(=矛盾したデータ)”が立ちふさがる

“Conflict Data”の実態

個別部門の業務上は支障が見えにくいが、
活用する際には“Conflict Data”が大きな阻害要因になる

問題タイプ		法人名称	生年月日	住所
仕様に合致せず	項目名と異なる値の混入	● “NTTデータ 豊洲太郎様” ⇒法人名称に個人の情報が混入 ● “(株)あいうえお” ⇒テスト時に使用したテストデータの残存	● “20歳” ⇒年齢表記の混入 ● “9999/99/99” ⇒ダミーデータの混入	● “Aさん担当” ⇒営業マンの担当区域のメモ書きの混入
	形式が違う値の混在	● “NTTデータ”⇔“N T Tデータ” ⇒全半角混在	● “2019/10/07” ⇔“20191007” ⇒“/”の有無の違い	● “NTTデータ築地ビル” ⇔ “N T Tデータ築地ビル” ⇒全半角混在
仕様に合致しても欠陥がある	欠損がある	● “株式会社エヌティティデータエン” ⇒“エンジニアリング”が欠損	● “10/07” ⇒“年”が欠損	● “コウトウトヨス 3 - ” ⇒番地以降が欠損
	不要な情報の混入	● “NTTデータ様” ⇒“様”が混入	● “2019年10月7日00時” ⇒“00時”が混入	● “ 江東区豊洲3-3-3内” ⇒“内”が混入
	異なる表現で重複	● “松下電器産業”⇔“P a n a ” ⇒旧称や略称の混在 ● “ユニクロ”⇔“ファーストリテイリング” ⇒ブランド名称の混在 ● “日本電気”⇔“日電”⇔“NEC” ⇔“日本電気(カブ)” ⇒社内略語や英名等の混在	● “2009年”⇔“平成21年” ⇒西暦、和暦の混在	● “埼玉県大宮市” ⇔ “埼玉 県さいたま市” ⇒旧称と新名称の混在 ● “京都府京都市中京区寺 町通御池上る” ⇒京都独特の俗称（通り 名）表記

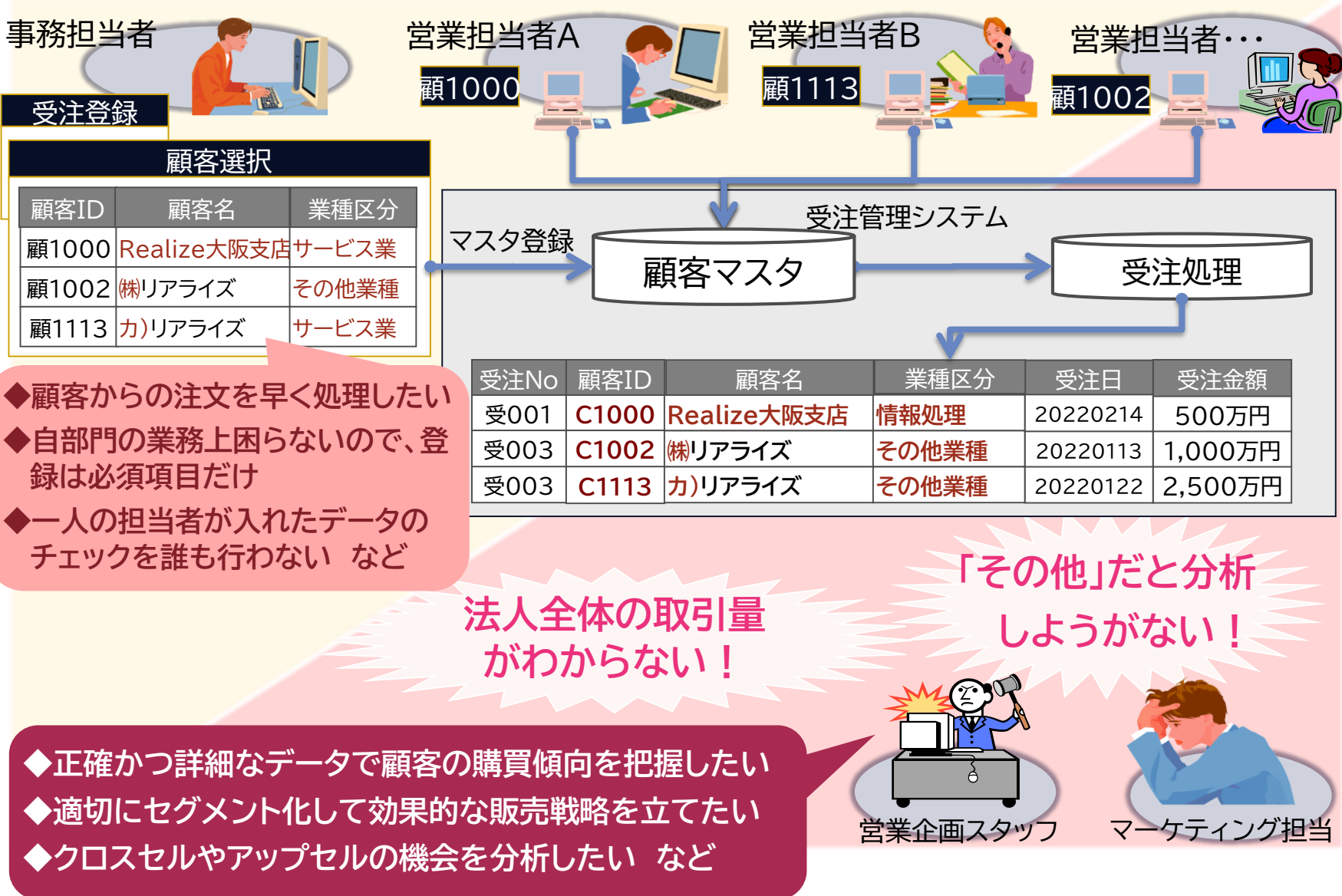
なぜ、“Conflict Data”が生まれてくるのか

- これらの原因は、全て「人がつくり出すデータ(中身)」。
- システムや機能の改善だけでは決して解決されない。不適切なデータを生産し、流通させてしまうマネジメントの問題。
- ITを経営に活かす、つまりはITの中の“血液”ーデータをどうビジネスに活用するか・・・同じく、マネジメントの問題。
- データが意味する商品・部材、顧客／仕入先の企業・個人、組織等の情報(値)が一意性と整合性を保ち、分析や仮説・検証が可能な状態であることが必須。



- フロントで業務処理を行うビジネス部門と、活用したい経営企画・商品開発部門やシステム利活用を促進したいIT部門では、本質的に異なるニーズが存在する。

本質的に異なるニーズとは～注文処理でよく見かける光景～



”Conflict Data”によってITは間違って動いてしまう

注文画面(システム)

最安値で買いたい

結果: 最高値のデータが出る

発注画面 -品目検索-

メーカー名	トラスコ中山
品名	六角ボルト
材質	SUS
型式	B23-1030
売価	760円

戻る 購入

メーカー【トラスコ中山】
品名【六角ボルト】
材質【SUS】
型式【B23-1030】

最安値を探したいのに！

発注画面 -品目検索-

メーカー名	TRUSCO
品名	BOLT (SUS/J1180)
材質	
型式	M10*30
売価	695円

戻る 購入

材料品目マスタ(データ)

すべて同一品目のデータ

No.	メーカー名	取引先名	品名	材質	型式	売価
1	トラスコ中山	正栄販売(株)	ロックク (1180)	ステンレス		760円
2	トラスコ中山	正栄販売	六角ボルト	SUS	B23-1030	760円
3	トラスコ中山	関東産業(カブ)	六角ボルト		B23-1030	740円
4	トラスコ中山	KANTO	ロックク (JIS1180)	SUS	M10*30	740円
5	TRUSCO	関東産業東京支社	BOLT	ステンレス	B23-1030	740円
6	TRUSCO中山	松葉商事(株)	六角ボルト	SUS	B23-1030	720円
7	TRUSCO	中部通商 関東支部	BOLT (SUS/J1180)		M10*30	695円

データマネジメントとは何か？

DAMA「DMBOK2」では、

データマネジメントとは、ライフサイクル全体を通じてデータおよび情報資産の価値を提供、管理、保護、強化する計画、ポリシー、プログラムおよびプラクティスの開発、実行、監督である。

【原文】 Data Management is the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles.

JDMC「データマネジメント概説書」では、よりシンプルに、

データをビジネスに活かすことができる状態で継続的に維持、さらに進化させていくための組織的な営み

と「データマネジメント」を定義している。

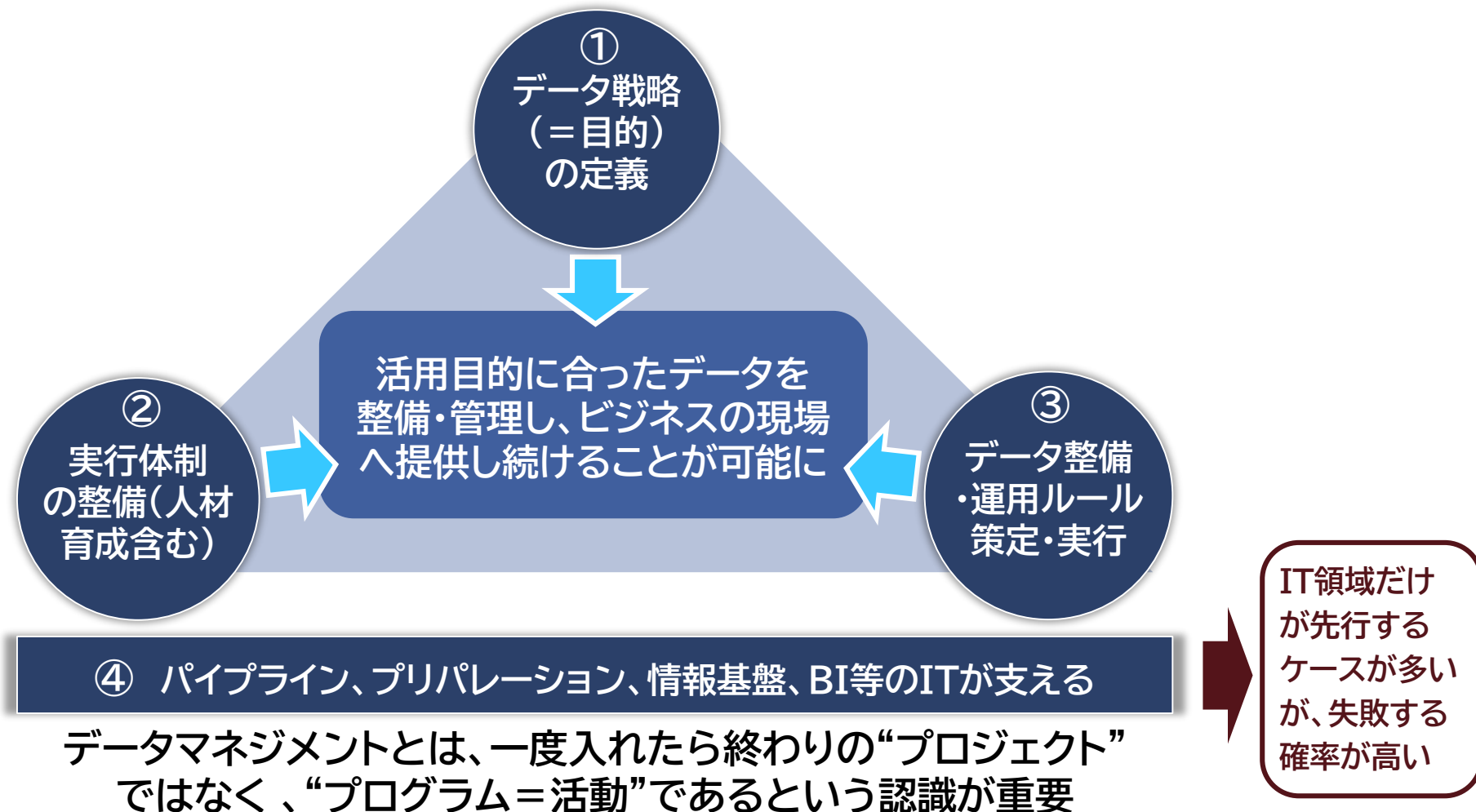
DAMAとJDMC、いずれの定義においても共通する本質としては、

ツールや基盤を導入すれば“終わり”ではなく、“これをやればよい”という都合の良い方法論でもない。それらはあくまで手段に過ぎない。

根絶が困難な”Conflict Data”と向き合い、『企業・組織にとって重要な資産であるデータの価値を適切かつ安全に統制・管理し、さらに持続的に向上させていくことを通じて目的であるビジネスでの成果を創出するための活動の総体』が「データマネジメント」であると捉えるべき。

データマネジメントの活動全体像～イメージ～

「情報(データ)は、ヒト、モノ、カネと並ぶ第4の経営資源」といわれて久しく、近年では「データは次のオイル」といわれる存在になったが、データを真のオイルに製錬するためには、①活用に足る状態のデータに整備・運用するための戦略(=目的)の策定、②実行体制の整備・人材育成・定着化、③ルール・プロセスなどの整備・実行・統制、つまり、データマネジメントの活動が不可欠となる。



【参考】AI時代においてより重要になるデータマネジメント

まったく同じAIのアルゴリズムを用いたとしても、インプットするデータの精度が高い場合と低い場合では予測結果の精度に大きな違いが出てくる、つまり、**AIに与えるデータ品質が高ければ高いほどアウトプットの精度が高くなる**、といった検証結果が出てきている。

ノイズが混入していたにしても高い結果精度を導き出せるようなビッグデータを保持している“GAFAM”ではない日本企業が勝つためには、「スモールでもグッドデータ」を整備し、AIに活用していくことが肝要ではないか。

“Garbage In, Garbage Out～ゴミを入れたら、ゴミが出てくる～”はAI時代においても変わらない原理原則であることを忘れてならない。

データ中心のAI（DCAI：Data-Centric AI）とは、従来通りのモデルやアルゴリズムを偏重するアプローチよりも、データに焦点を当てたアプローチの方が大切であるとする、AIの開発方法に関する考え方である。

この考え方の根拠として、たとえ比較的少ないデータ量であったとしてもクリーン（Clean）で高品質なグッドデータ（Good Data）の方が、ノイズの多い（Noisy）ビッグデータ（Big Data）よりも良い性能を発揮することが示されている（図1）。

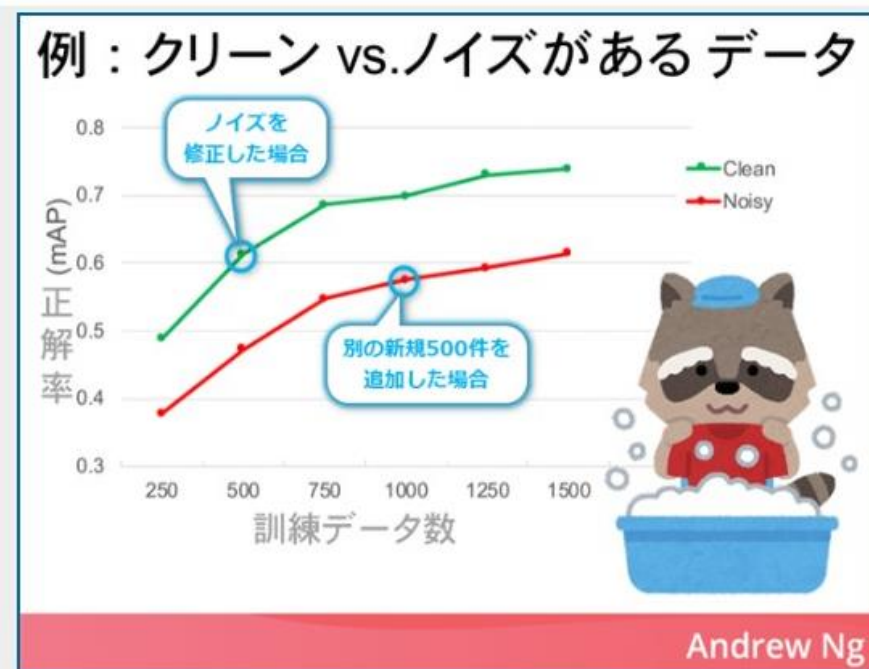


図1 クリーンなGood DataがノイズありのBig Dataに勝る事例（Andrew Ng氏のスライドから引用、一部を加工）

500件のデータ中に12%のノイズ（＝間違った／矛盾した教師ラベル）がある。このとき、ノイズをクリーンアップ（修正）することは、このグラフの正解率を見ると、別の500件の新規データを集めること以上の効果を持つことが示されている。

～@IT記事(2022.2.10)より～

<https://atmarkit.itmedia.co.jp/ait/articles/2202/10/news037.htm>

03

データマネジメントを牽引する人材育成の必要性

データマネジメント活動推進人材の育成に関する日本の現状

“IT(器)の整備士”を育成・普及するための国家試験はあるが、“データ(中身)の整備士”の知識・リテラシーとスキルをもつプロフェッショナル育成のための国家試験は存在しない。

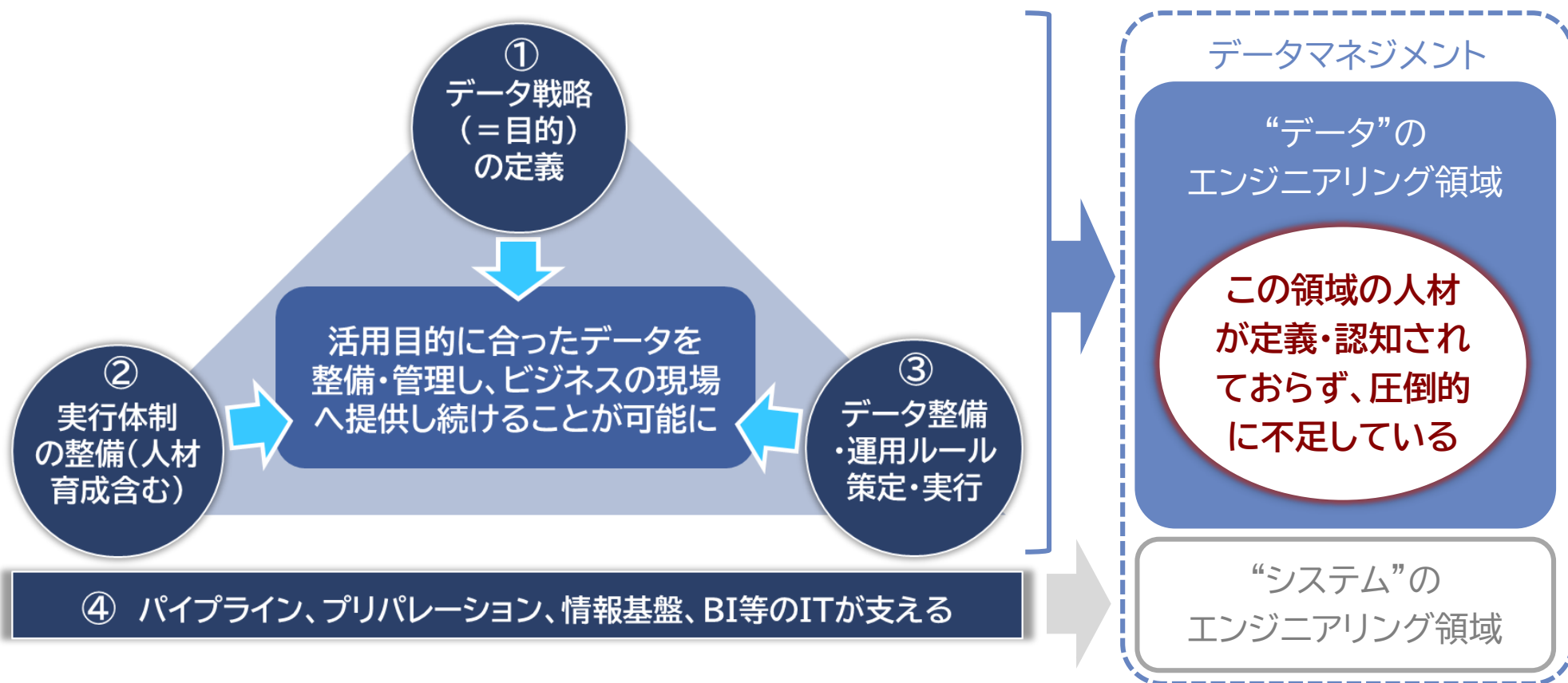
民間でも統計・解析の専門知識・スキルをもつデータサイエンティストの教育機関や検定制度等はあるが、データマネジメントを推進できる人材の育成プログラムや資格制度等は現状存在しない。



IPA「現行の試験制度(平成29年度春期から)」より
(<https://www.ipa.go.jp/shiken/kubun/list.html>)

データマネジメント活動推進人材の圧倒的な不足

前述のとおり、“データ(中身)の整備士”を育てる国家試験や教育プログラムもなければ、**データマネジメントの実践を担う人材はまだ社会から認知されていない**。昨今でこそ「DMO(Data Management Office)」といったデータマネジメントの全社的推進組織を設置する先進的企業が現れてきているが、**その推進を担う人材の育成をどうすればよいかはまったくの手探り、といった企業がほとんど**である。一過性の「システム開発のためのプロジェクト」ではなく、**「データの整備と活用を社内に普及・定着化させていくための持続的プログラム(=活動)」であるデータマネジメントを推進することができる知識・リテラシーとスキルをもった人材の育成が日本において急務**と考える。



“データ整備士”－データエンジニアリング人材をどう定義すべきか？

巷間で言われはじめた“データエンジニア”は『データの収集・整理・管理等を行うための基盤を構築するITエンジニア』だが、従来のSEの仕事の範疇に留まるためそう定義すべきではない。

データマネジメント実践の中核を担うデータエンジニアリングには、『データの実態を掌握し、その価値を高めることを通じて、ビジネスでのデータ活用を牽引・活性化・定着化させるコンピテンシー・スキルを有するプロフェッショナル人材』が必要であり、その役割は組織的なデータマネジメント活動のリード役として上流工程を担う『データマネージャー』とデータの整備・価値向上の実務者として下流工程を担う『データエンジニア』の二種類と定義すべきである。

非ITのビジネス
パーソンを社会
全体でもっと
巻き込んでいく
必要がある



- 業務・ドメイン知識を持つ
ビジネスユーザー

≡ データをビジネスでの意思
決定に役立てる活用人材

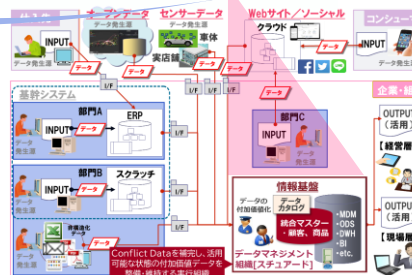
● データエンジニアリング人材 (データマネージャー & データエンジニア)

≡ データを共通言語とした“橋渡し役”、ビジネスユーザーに
よるデータ活用を支援するプロフェッショナル人材



- データサイエンティスト
- データアーキテクト
- 基盤エンジニア
- ITパートナー企業

≡ 高度な予測・解析やデータ活用
基盤の構築等を支えるIT人材



【背景①】データエンジニアリングのプロフェッショナルを育てる必要性

これまで日本企業でデータサイエンティストという肩書の人は存在しなかったが、各大学にデータサイエンス学部が設置され、そうした人材が企業へ徐々に輩出され、データ分析の実務に関わっていく中で、「データサイエンティストの仕事の8割はデータの前処理」という定説が実感を籠めて語られるようになってきた。データサイエンティストは本来の2割の仕事に集中したいはずであり、ボリュームの大きい8割の方の仕事をどう的確かつスピーディに裁いていくことができるかを真剣に考えるべき時期に来ている。(=日本企業のアジリティや”Time to Market”をそれだけ落としているともいえる)

データを活用したいビジネスユーザーと高度な統計的技術をもつデータサイエンティスト等の橋渡し役として、活用可能な状態のデータを整備・管理・高度化するデータエンジニアリングのプロフェッショナル人材が今後より必要とされる。

日経 XTECH

「データ分析は前処理の時間が8割」ともいわれています。自動化するためのツールも多く出ていますが、完璧にはできません。最終的にはデータサイエンティストの確認を必要とする泥臭い作業です。

データが欠損している場合は、平均値や中央値で埋め合わせる作業を行います。どういうデータで欠損を補うかにより分析結果が変わってきってしまうため、分析手法のことを理解した専門知識も必要です。前処理のための作業時間を軽視したり、データ提供者の「データは完璧にそろっている」などの発言をうのみにしたりすると、データ分析を始める前に3日徹夜するような事態に陥ります。

前処理で時間がかかった具体例

- ・「時間」を表す変数のフォーマットがそろっておらず、一つ一つの変数を探索しながらクレンジングする必要があった
- ・エリアを表す変数で、「都道府県」「都道府県 2」「地域」「県」など、似たような項目が乱立。入力値も、漢字、かな、末尾の県の有無、都道府県に「関東」などバラバラだった
- ・欠損値だらけで、平均値で穴埋めしたら、データ（変数・サンプル）間の差が見られにくく、まともな分析ができなかった

(出所：書籍『ビジュアル データサイエンティスト 基本スキル84』)

[画像のクリックで拡大表示]

～日経クロステック記事(2023.05.11)～

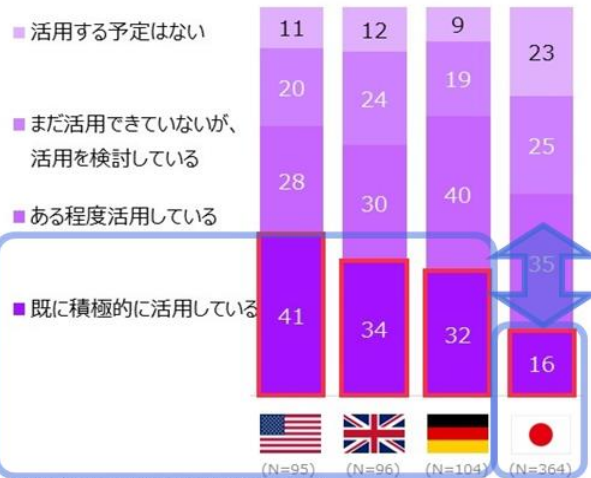
<https://xtech.nikkei.com/atcl/nxt/column/18/02408/042000002/>

【背景②】データエンジニアリングのプロフェッショナルを育てる必要性

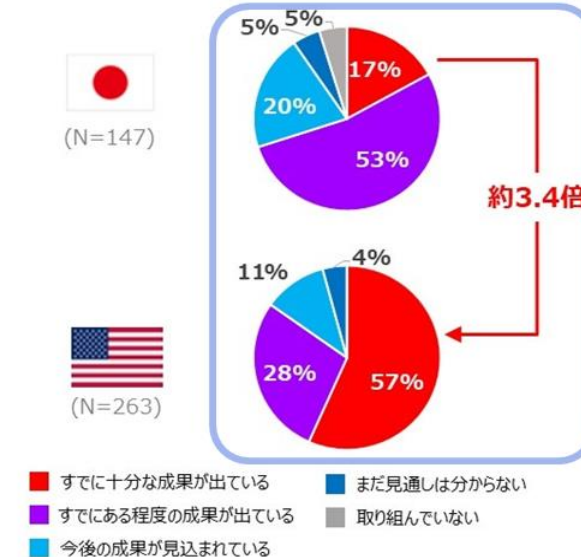
日本におけるデータ利活用の浸透

日本企業は先進各国と比べてデータ活用に遅れており、成果実感を十分に得られていないと感じている。

各国別の企業におけるデータ活用状況



日本企業におけるデータ活用によるビジネス成果



日本と海外におけるデータ活用の差異
(2022/8/5 アクセンチュア調べ)

https://financialservicesblog.accenture.com/why-do-we-not-realize-the-benefits-of-data-utilization-the-essentials-of-transforming-data-into-value-and-the-path-to-change?lang=ja_JP

効果実感はアメリカの3分の1以下

→日本の控えめなお国柄もあるだろうが、データ活用による効果を日本企業が実感できているかというと、かなり低いという結果は実感値と合っていると思料

“ERP”で業務変革を実現する

“ビッグデータ”で経営を変革する

“DX”で会社をデジタルに刷新する

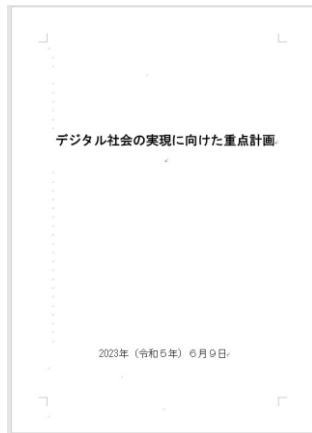
“AI”で。。。

このままでよいのか



“ITの整備士”を増やすよりも、これからはビジネスの現場で意思決定に役立つ“データの整備士”を増やす方が、諸外国から劣後したデータ活用によるビジネス成果をより早く日本企業が享受できるようになるのではないかと

【背景③】データエンジニアリングのプロフェッショナルを育てる必要性



2023年6月9日に閣議決定された「デジタル社会の実現に向けた重点計画」では、“データ連携基盤”というキーワードが45件、“プラットフォーム”が35件といった状態で頻出しているが、“データマネジメント”については以下の記述しかない。基盤やプラットフォームに流通する活用対象のデータが目的に適合した品質を保っているのか等についても対策が必要ではないか。

(3) 当面重点的に取り組むべき事項

⑥ データマネジメント

生活や企業のあらゆる活動でデータを活用するデジタル社会において、円滑なデータ連携において参照可能な標準群として政府相互運用性フレームワーク（GIF）を整備・公表している。AIでの活用も念頭に置き、データ連携における設計・変換コストの低減、環境整備による、データ間の相互運用性を高めるため、デジタル庁で先行してGIFの導入を検討できるシステムを選定し、具体的なGIFの導入支援、利用促進を図る。

総務省「情報信託機能の認定スキームの在り方に関する検討会（第9回）配布資料」より
(https://www.soumu.go.jp/main_sosiki/kenkyu/information_trust_function/02tsushin01_040_00571.html)

データ取引市場を構成する3つの機能



データが市場で取引される社会を政府は後押ししているが、企業は信頼性のないデータを購入したいと思わない。流通させるデータの品質を評価・担保できるデータエンジニアリング人材が不在だと、データを対価性のある資産として取引を活性化させることが難しいのではないか。

国が公表、つまり国民に約束している計画や政策のいずれも、基盤やプラットフォームといったIT(器)にばかり視線が集まり、データ(中身)の品質や価値を高めることに着目していないと感じられる。

ITの“血液”であるデータが安心・安全に流通できること、相互運用性が確保されていることが重要なことは論を待たないが、“血液”そのものの信頼性(精度、鮮度、整合性、等)を確保し、向上させるためのデータマネジメント活動を並行して計画・実行しないと、「画に描いた餅」に終わることを危惧する。

