



Data Scientist in Shionogi : Education, Skills & Training

11Nov2021

Ayaka Yamashita, Yoshitake Kitanishi

Data Science Department,

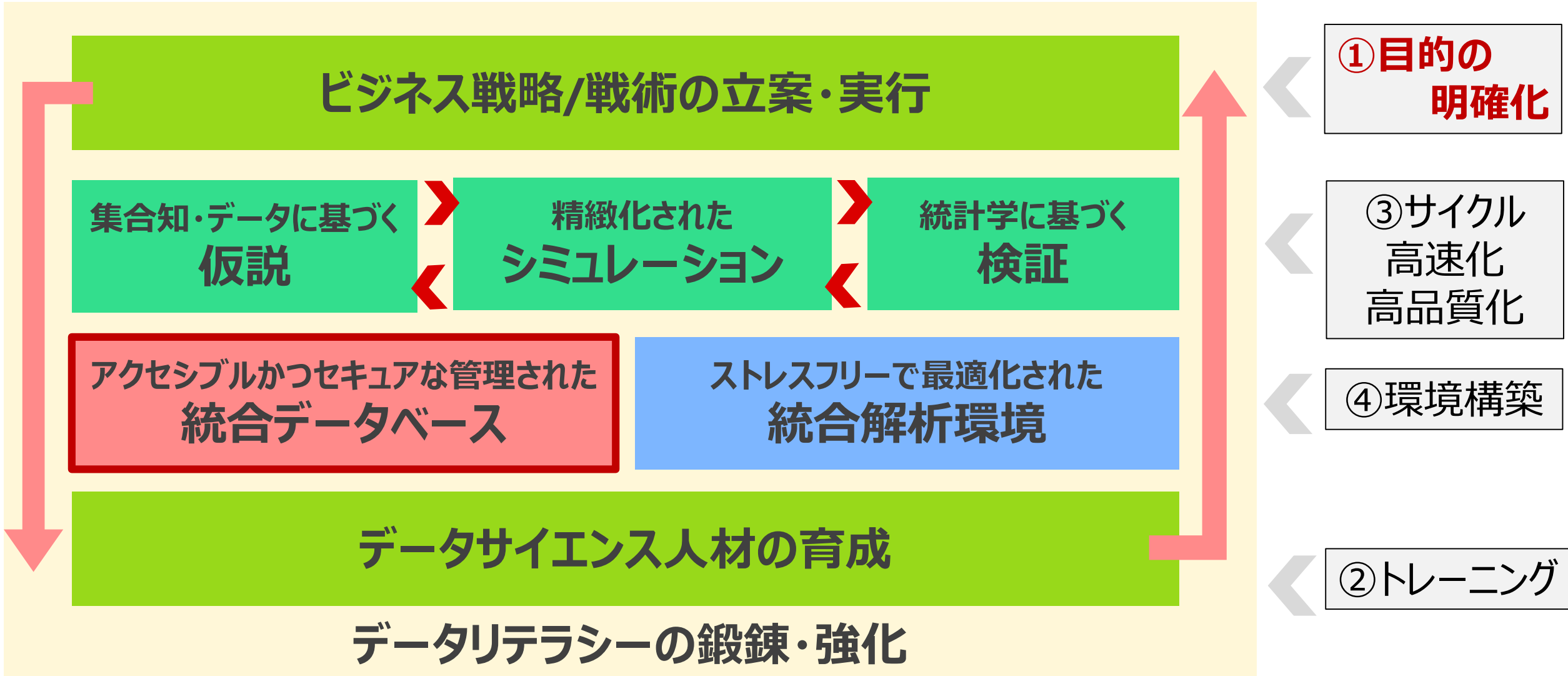
Shionogi & Co., Ltd.





データサイエンスの重要性と データサイエンティストの育成

データ駆動型でビジネスを行うためには？



データ活用とデータ駆動型ヘルスケア戦略

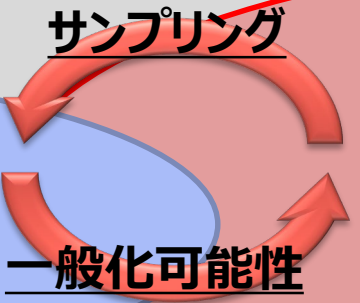


統合データベース

仮説を立てる
トレンドの把握
マクロ的解析

検証
因果の追求
ミクロ的解析

Clinical Trial Data
(Controlled Data)



Real World Data
(Un-controlled Data)

※Bigになればなるほど、
populationに近くなる

Simulation Data
(Prediction Data)
e.g. 疾患モデル

データ更新
予測
将来予測
仮想臨床試験

データ更新
予測

推定・検定

推測・ルール

疫学

Inferential Statistics

Machine Learning

Biostatistics
Multiple Comparison
Adaptive Design
Enrichment Design等

Data Science
各学問領域の融合は加速的に進む

Informatics
SVM (1990年代)
Random Forest (2000年代)
Deep Learning (2010年代)

(Bayesian)

データサイエンス業務の進め方の一例

例えば、社内のとある部署で抱える課題を解決していく場合・・・



各段階で様々なスキル・能力を駆使しながら、課題解決に向かっていきます

データベース (データ) を組み合わせよう

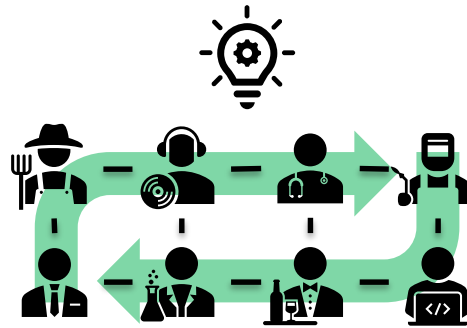
Why

集合知を強化して活用したいから

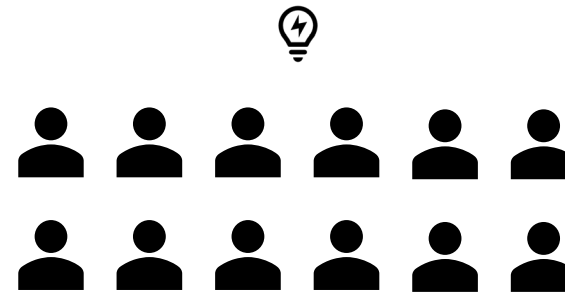


「個人の知より集団の知の方が勝る」という考え方で、知識を蓄積し
何らかの価値ある情報に体系付けしたもの

✓ 代表例: Webのオンライン百科事典「Wikipedia」



多様性 + 体系化

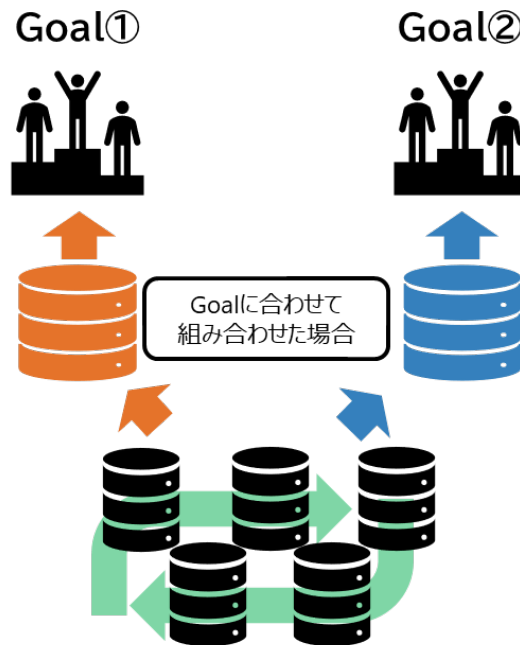


数

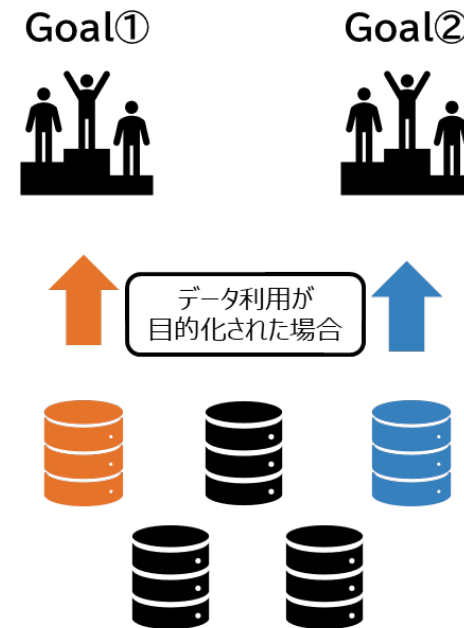
データベースの集合知 = 組み合わせたデータ

目的に合わせて組み合わせたデータでのData Driven型ビジネスが「データサイエンスビジネス」
データの利用が目的のData Driven型ビジネスは「データサイエンスビジネス」でない

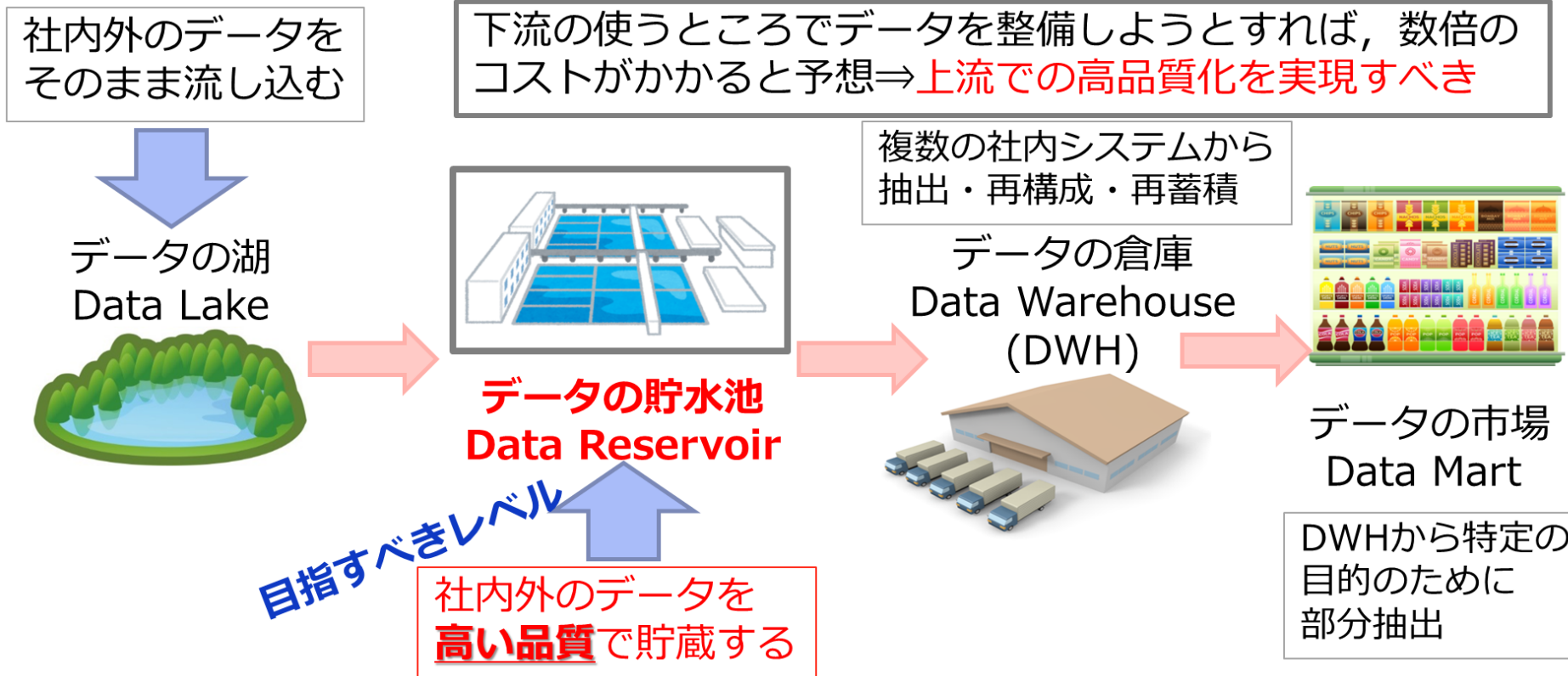
✓ 成功しやすいモデル



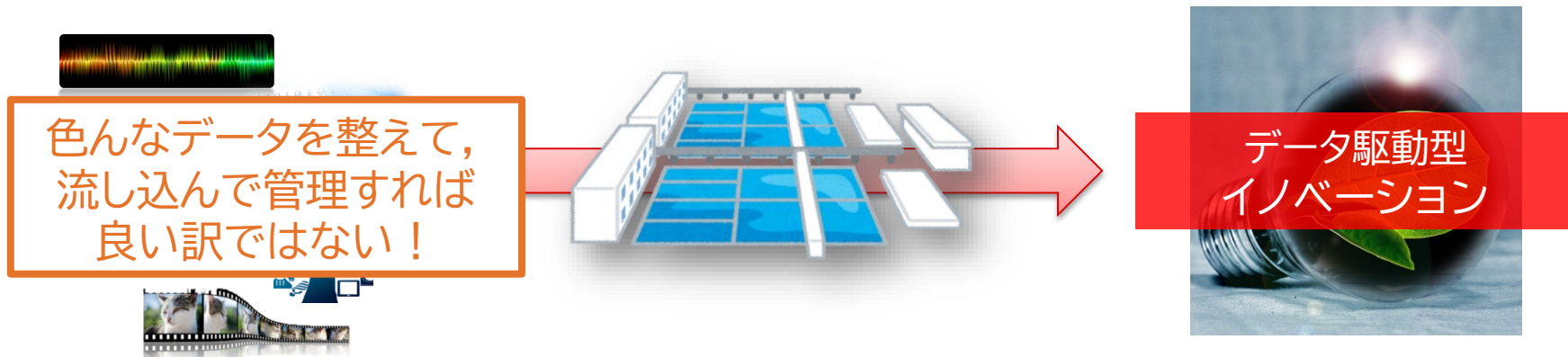
✗ 失敗しやすいモデル



データガバナンス戦略の一例



真のデータガバナンス戦略とは



真のデータガバナンス戦略では、ソースデータを手当たり次第に流し込むのではなく、“何らかのルール(規則)”を持ったデータを流し込むことがポイントである



Any Data×人工知能技術×シミュレーション

新たな解析技術を積極的に調査・習得し、
さらにその応用・実装を速やかに実施。
蓄積されたデータの集約と活用。

ヘルスケア領域のデータ駆動型イノベーションを促進

データサイエンスを駆使し、

- ① プロセスイノベーション（業務効率化）による生産性向上を図り、
- ② プロダクトイノベーション（開発品価値最大化）にタスクリソースを投下。

新たなデータサイエンス技術の導入と応用(事例)

プロセスイノベーション

人工知能技術を活用し、臨床試験の解析業務を準自動化(特許出願中)

- ①過去蓄積してきた解析計画とプログラムコードを整理し、ビッグデータ化。
- ②認識技術と機械学習を応用し、計画解析のプログラムコードを準自動生成。
 - ・計画解析のプログラムコード作成に費やす時間を**30%超短縮**。

蓄積されたデータと解析技術の融合・還流

プロダクトイノベーション

シミュレータの開発・疫学研究への応用、網羅解析の実施

- ・個人の特性、多様性を考慮したシミュレーションモデル*を応用
- ・臨床試験データに加え、リアルワールドデータ、気象/地理データなど
多様なデータを融合することでシミュレーションの予測精度向上を図る。

ウェアラブルデバイス等による健康状態の見える化

バリューチェーンデータの統合化施策におけるデータ基盤のポイント



クラウドとオンプレのハイブリッド環境
DWHと高度解析環境は現状オンプレで構築

マスターデータ



配信/同期

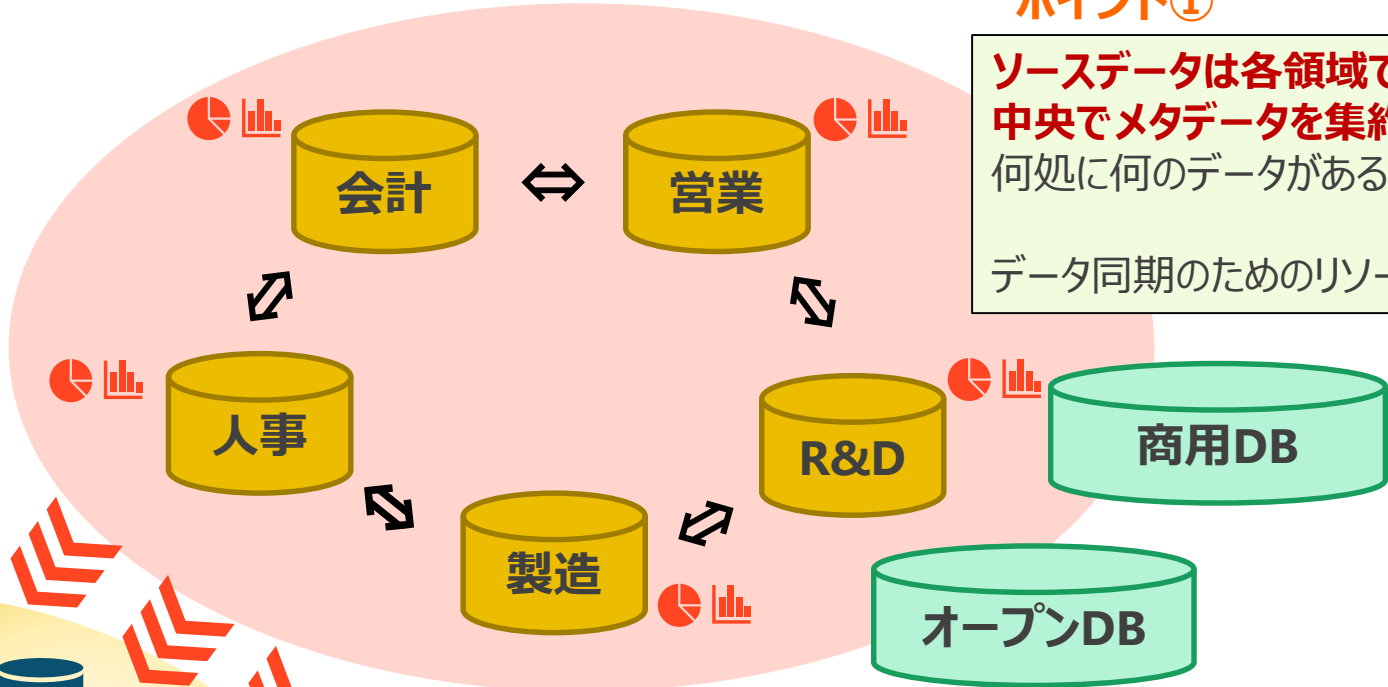
データリザーバー
(データウェアハウス含む)



統合DB

データカタログ
(メタデータ)

ルールに則った移管・蓄積
個人情報等のコンプライアンスにも対応



ポイント①

ソースデータは各領域で管理/利用。
中央でメタデータを集約管理することで、
何処に何のデータがあるかを把握。
(順次構築中)
データ同期のためのリソース負荷を軽減。

可視化と高度解析



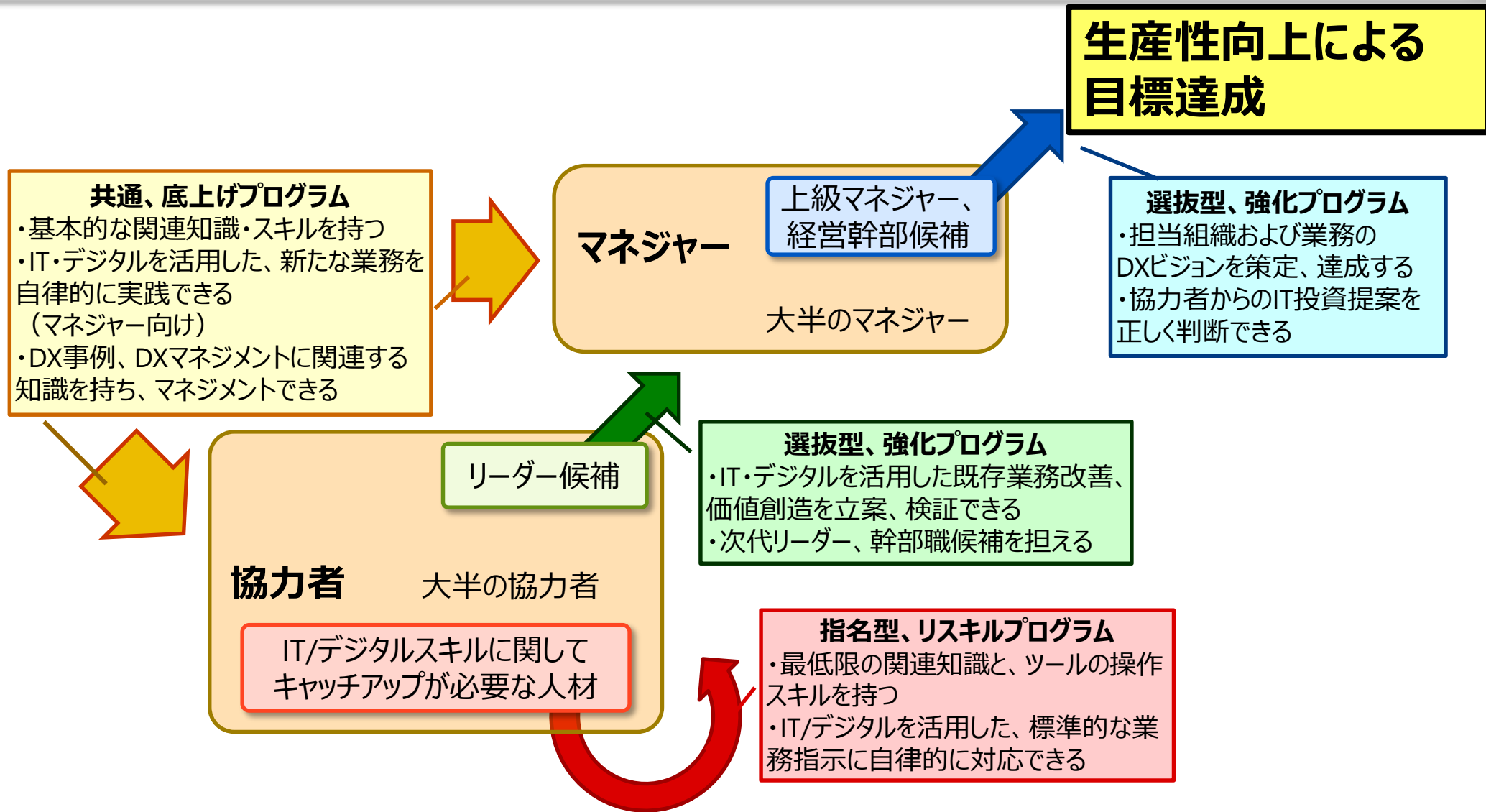
In-Database機能により統合DB内で抽出と一次加工
することにより、解析と思考のサイクル高速化。
BIツールで直接に可視化。

ポイント③

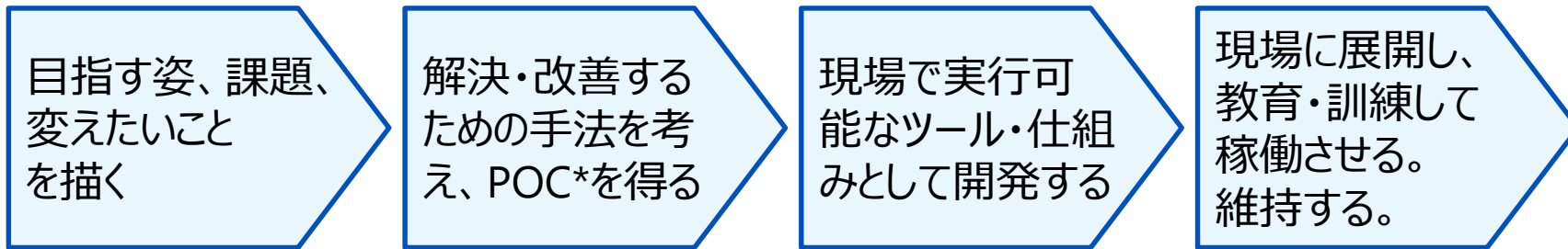
ポイント②

ビッグデータの特徴である4つのV*に
対応した統合DBにニーズドリブンかつ
ルールに則ったデータを蓄積することで、
解析担当者のデータへの
アクセシビリティを向上させ
仮説検証サイクルを加速させる

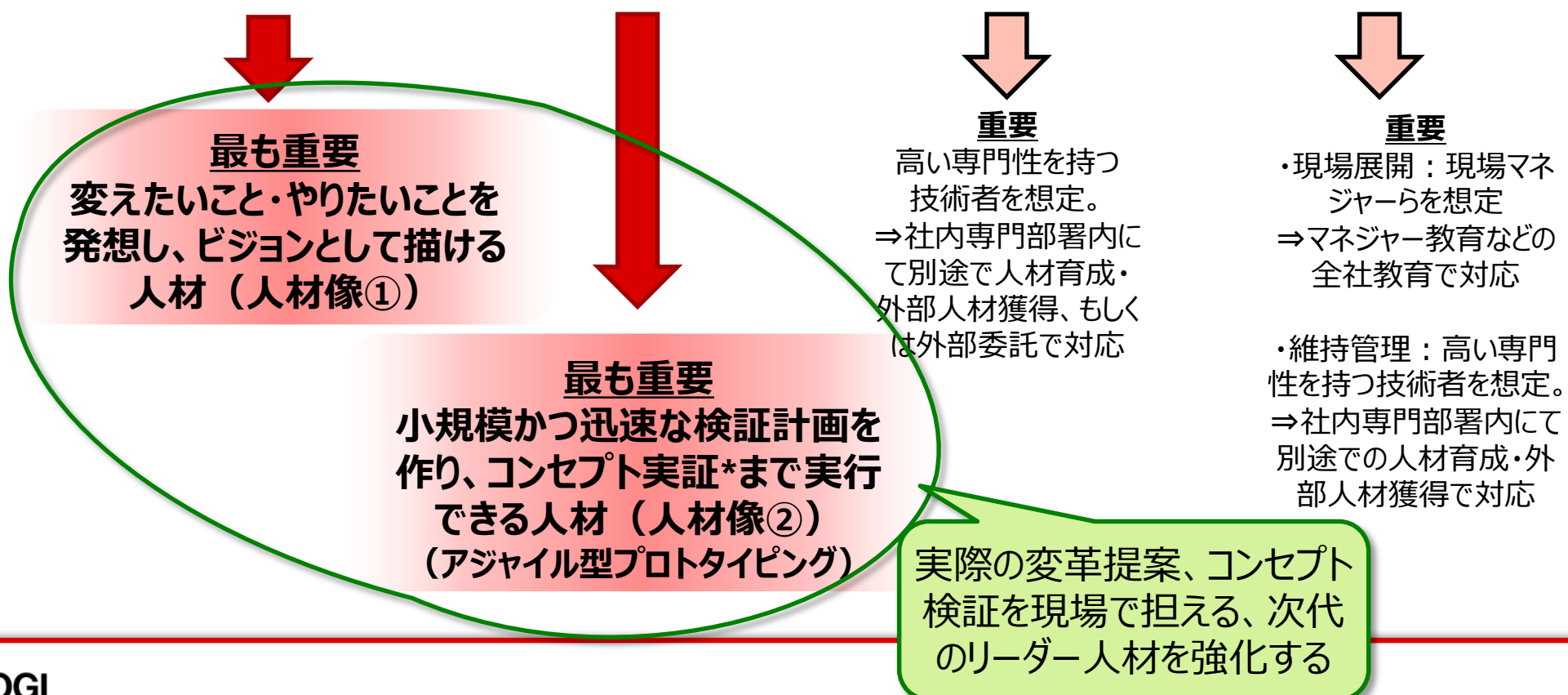
DX実現に向けて各層別プログラムを強力に全社展開



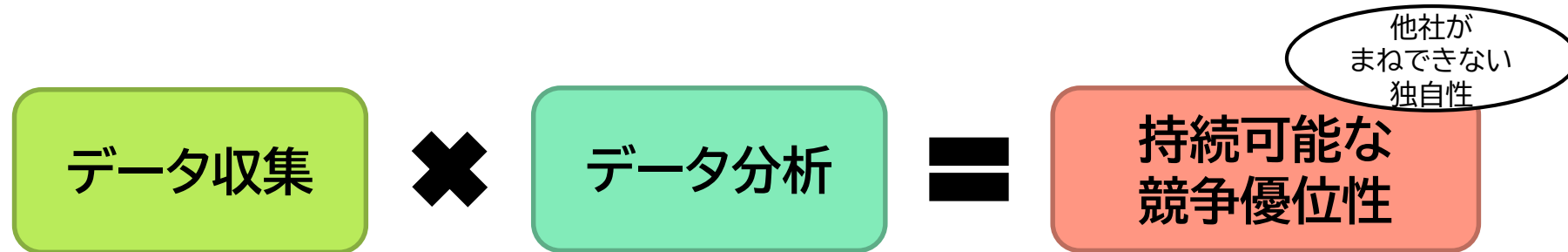
想定されるDXプロセスにおける本プログラム強化人材の位置づけ



* Proof of concept



ヒト・モノ・カネに次ぐ第4の経営資源であるデータ(情報)を収集・分析し、アクションを起こすことのできる人材を開発・育成する。



1. 経営課題やビジネス状況の明確な理解
2. 仮説を導き出すためのデータ収集・作成
3. 正しい分析手法の調査, 選択
4. 発見した知識の評価と共有化

データに基づく
仮説⇔検証サイクルを
体得する

ビジネス課題を解決するための分析
テーマを設定する。
分析に必要なデータを把握し、その
価値を上層部に説得し、入手できる。

**課題
設定力**

3つの力を駆使して、データに基づき
課題解決や意思決定につなげる

**データ
分析力**

仮説を立てて必要なデータを手
し、適切な分析手法を選択できる。
分析結果を解釈し、データの意味
を正しく伝える。

**データ
整理力**

データを収集して複数のデータソースを
組み合わせて分析することができる。
データベースから分析に必要なデータを
抽出し、加工することができる。

データ活用人材育成プログラム



プログラムの目的

データリテラシー・データサイエンス力の向上を図り、データ活用を促進することで各組織が抱える課題に対し、データ駆動型の業務変革を行う。

レベル1
課題認識

レベル2
データ把握と可視化

レベル3
データに基づく
仮説立案と検証

レベル4
データ駆動型
価値創造

アセスメント

【基礎編】

データベース活用講座

【応用編】

データサイエンス講座

アセスメント

✓データ収集～分析を体感し、分析結果を伝える。

(例) 疾病と気候との関連を調査するために、必要なDBを選定し、データを抽出・加工、分析を行い、グループ毎に結果をプレゼン。

✓課題抽出～解析～企画提案の流れを実践する

(例) 在庫管理のため、機械学習を用いた最適化モデルを構築する。課題設定は生産部門と協力し、得られた結果は実際に提案。

内製化のポイント

講座を設計・パッケージ化し、更には講師として参画することで、多様な視点や幅広い気付き等の多くの学びを得る。

⇒教えることは教わること

- 全社施策の中でDX/IT人材，そしてデータ活用人材の育成プログラムを設計すべき。
- 教育を内製化することで，各組織の課題を把握することができる。
- データサイエンス専門組織が講師（コンサルタント）を担い，業務改革を受講者が担うことで，効率的に生産性向上を実現できる。
- 「教えることは教わること」であり，データサイエンス専門組織のスキルアップにもつなげることができる。

- 地道にDX/IT，データサイエンストレーニングを行うことで，風土改革にもつなげたい。