

# データ利活用を支えるための データマネジメントへの取り組み

技術本部 システム研究開発センター

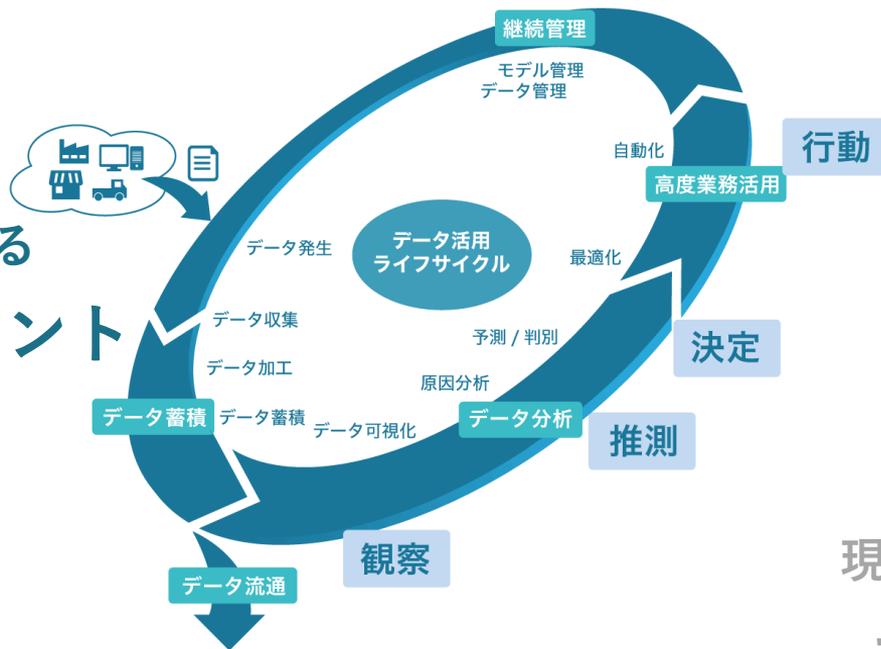
**福富 正弘**

DX&イノベーションセンター

**蔣 偉**

# 本セッションの位置付け

必要なデータを届ける  
データマネジメント



人と協調できるAI  
最適化技術

現場が見える、触れる  
デジタルツイン

# アジェンダ

1

データ活用とデータマネジメント

2

NSSOLのデータマネジメントプロセス

3

NSSOLのデータマネジメント技術

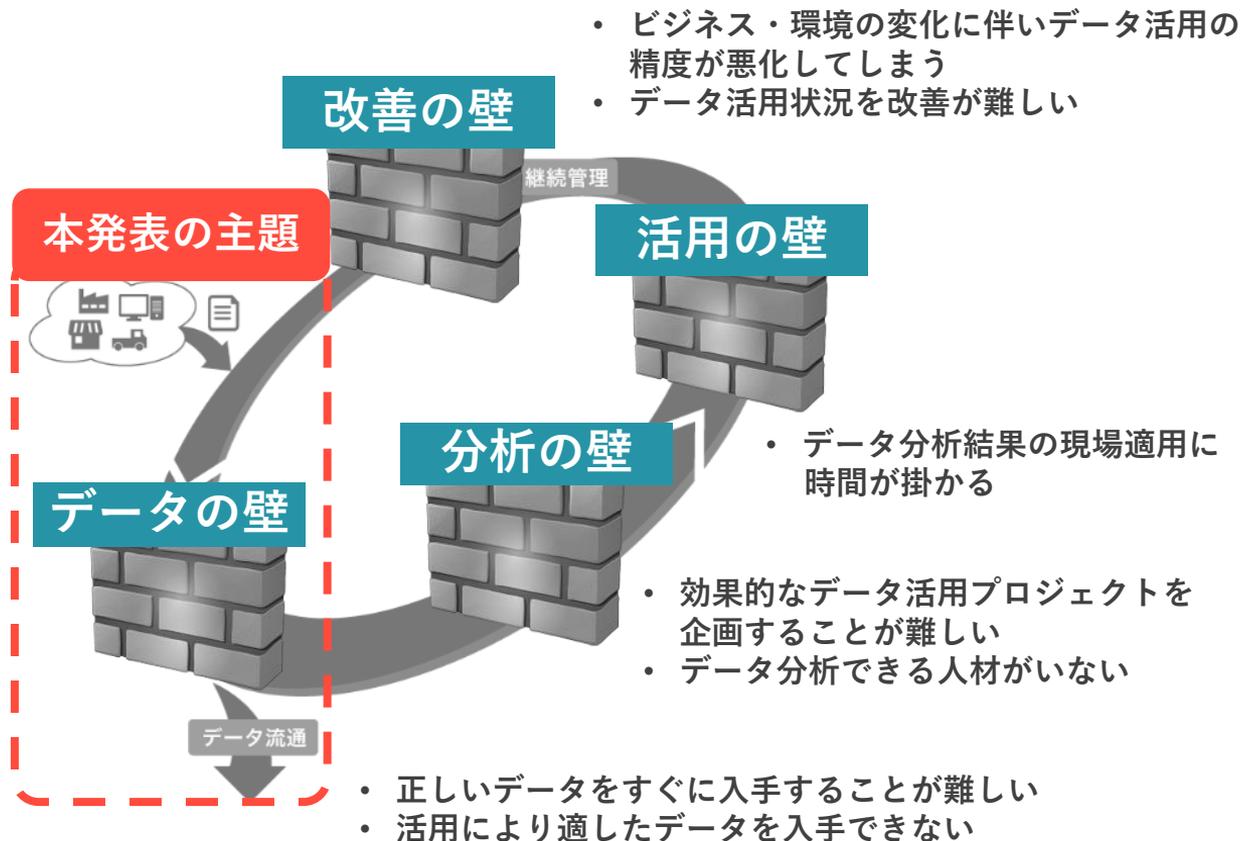
4

まとめ

# 1

## データ活用とデータマネジメント

# データ活用ライフサイクル実現に立ちはだかる4つの壁



# 「データの壁」とは

IoTデータや非構造データなど必要なデータが多様化

複数の組織やシステムにサイロ化されたデータ

データ提供のスピードアップや頻度向上の要求



**データの収集・蓄積・提供がますます困難に**



# データマネジメントを進める上での重要なポイント

データマネジメントで予想通りの効果を上げられているケースは限定的

NSSOLに寄せられるお客様の声（一例）

べき論は分かるが、取り組むべき範囲が広すぎて、組織としてどこから着手してよいか分からない。

データマネジメントに必要な実作業を誰に任せるか、そのインセンティブはどうするかが整理できない。

データ統合ツールやメタデータ管理ツールなどを導入したが、実作業が大変すぎて継続してツールを活用できていない

ゴール設計や実行体制検討のプロセスが不十分

データマネジメントの実作業において手作業が多く、自動化・省力化が不十分

データマネジメント  
プロセス



データマネジメント  
技術

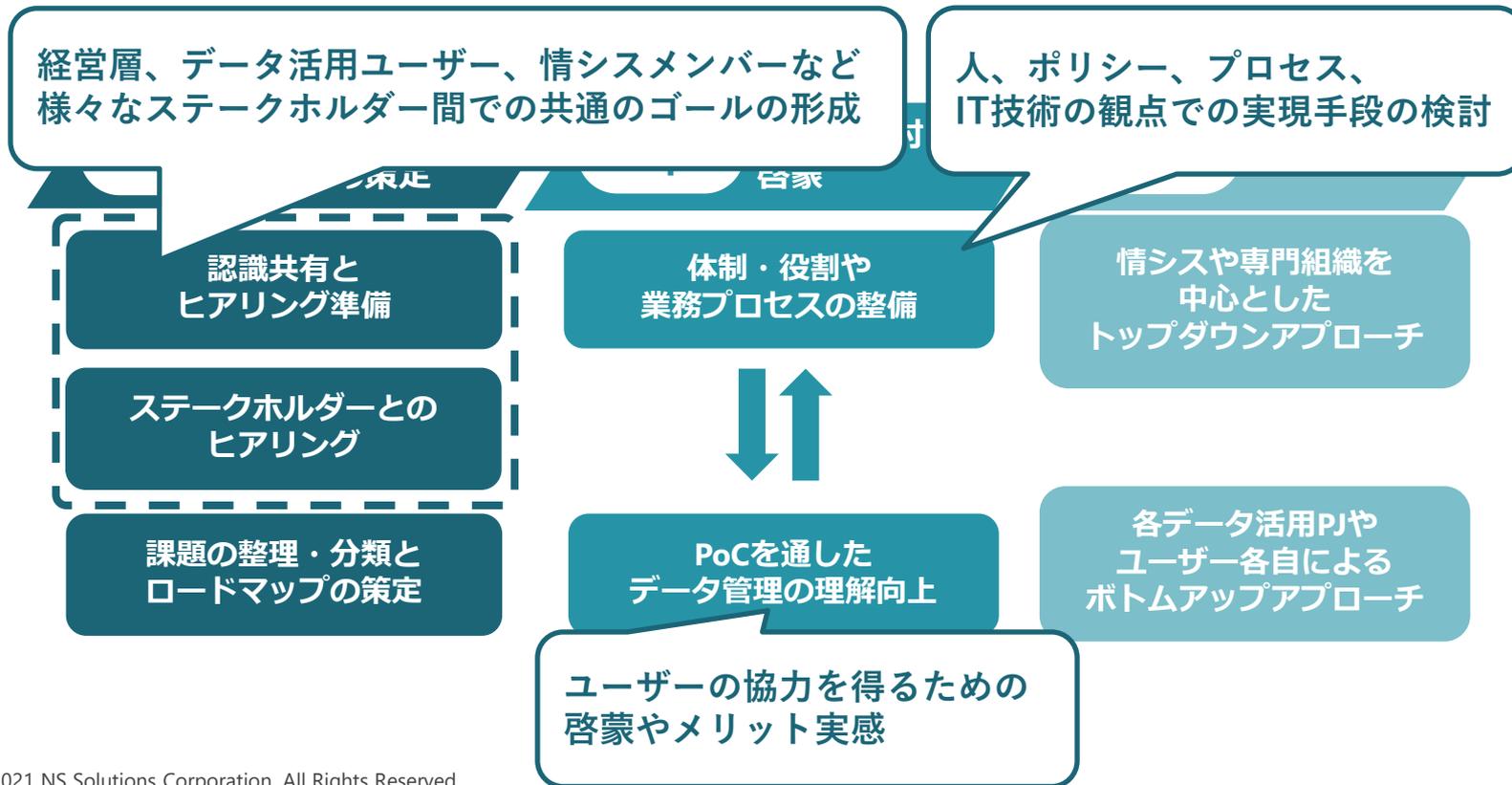
この両面の取り組みが重要

# 2

## NSSOLの データマネジメントプロセス

# データの壁の乗り越え方 = データマネジメントプロセス

NSSOLは以下のプロセスに沿って顧客のデータマネジメントを支援



# データマネジメント支援の中で遭遇した技術面のニーズ

データの前処理に  
時間と手間が掛かっている

個人情報や機微情報が含まれているため  
他の組織と共有できない

データの意味が分からない  
データ辞書のメンテに手間が掛かり過ぎる

大量データをセルフサービスで  
分析する環境がない

大量の文書が内容に基づいて  
整理されていない



これらの技術的課題に対して  
シス研において技術開発に取り組む

# 3

## NSSOLのデータマネジメント技術

# シス研で取り組んでいるデータマネジメント技術

データの前処理に  
時間と手間が掛かっている

個人情報や機微情報が含まれているため  
他の組織と共有できない

データの意味が分からない  
データ辞書のメンテに手間が掛かり過ぎる

大量データをセルフサービスで  
分析する環境がない

大量の文書が内容に基づいて  
整理されていない

## 本発表のご紹介内容

自然言語処理&機械学習による  
データの前処理の自動化・省力化

安全なデータ流通のための  
データセキュリティー

データカタログ

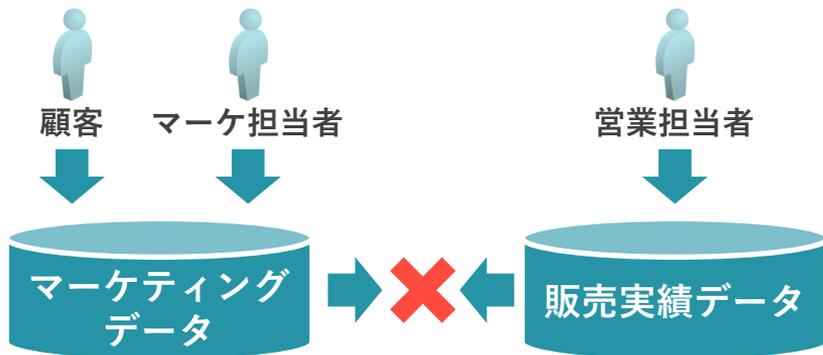
クラウドDWH

文書・コンテンツに関する  
自動分類や自動タグ付け

# 事例：サイロ化したデータの自動マッチング

自然言語処理&機械学習による  
データの前処理の自動化・省力化

## お客様の課題



それぞれのデータが個別に管理（サイロ化）  
されており紐付けが困難

データの紐付けを行い、マーケティング施策の  
効果可視化や改善検討に活かしたい。

## PoCにおける取り組み内容

自然言語処理技術と機械学習を用いて、  
マーケデータと販売実績データの紐付けの推定

### 類似文字列マッチングによる会社名の名寄せ

ABC Engineer

ABC Eng. Inc.



同一

会社名の表記揺れや  
タイプミスに対応

### フリーテキストから推定に必要な情報を自動抽出

Tell me the  
price of RT246



RT246

抽出

人が記述した文章から  
重要な情報（型番など）を自動抽出

### 機械学習を用いたデータの紐付け推定

- ユーザーによる一部データに対して紐付けを学習
- 継続活用のための再学習や自動パラメータチューニング

## データ活用は、前処理に8割の時間が掛かる

前処理の中でも特に人手・時間が必要な  
データの意味や背景に基づいて行う  
データ加工を自動化・省力化

自然言語処理技術



機械学習・AI



### データ修正・クレンジング

- ・ 意味的な誤りデータ検知
- ・ 名寄せ など

### データエンハンス

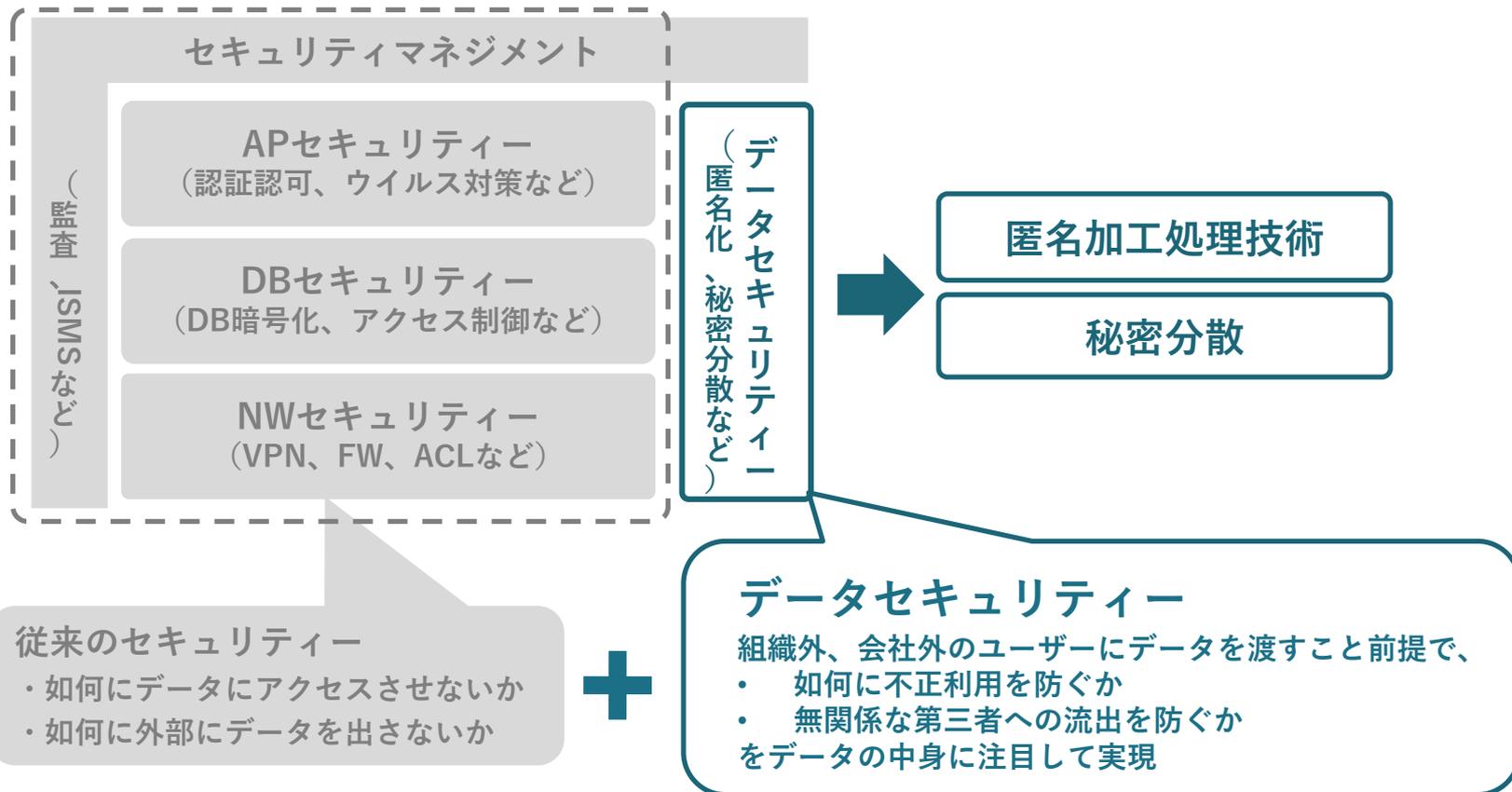
- ・ カテゴリ分類（タグ/ラベル付け）
- ・ 断片データのマッチング など

### データサマライズ

- ・ ネガポジ分析等の定量化、集計・要約
- ・ 非構造データの構造化 など

# データ流通に必要な新たなセキュリティ

安全なデータ流通のための  
データセキュリティ



# 事例：次世代医療基盤への「匿名丸」の導入

安全なデータ流通のための  
データセキュリティー

次世代医療基盤法に基づき、医療データなどを安全に研究機関に提供し活用

診療所, 介護施設など



提供



次世代医療基盤 (ICI様)

医療情報収集基盤の構築

名寄せ



蓄積



匿名加工エンジンの導入

提供

匿名化



匿名加工者



研究機関など



提供

NSSOL開発の

匿名加工エンジン「匿名丸」

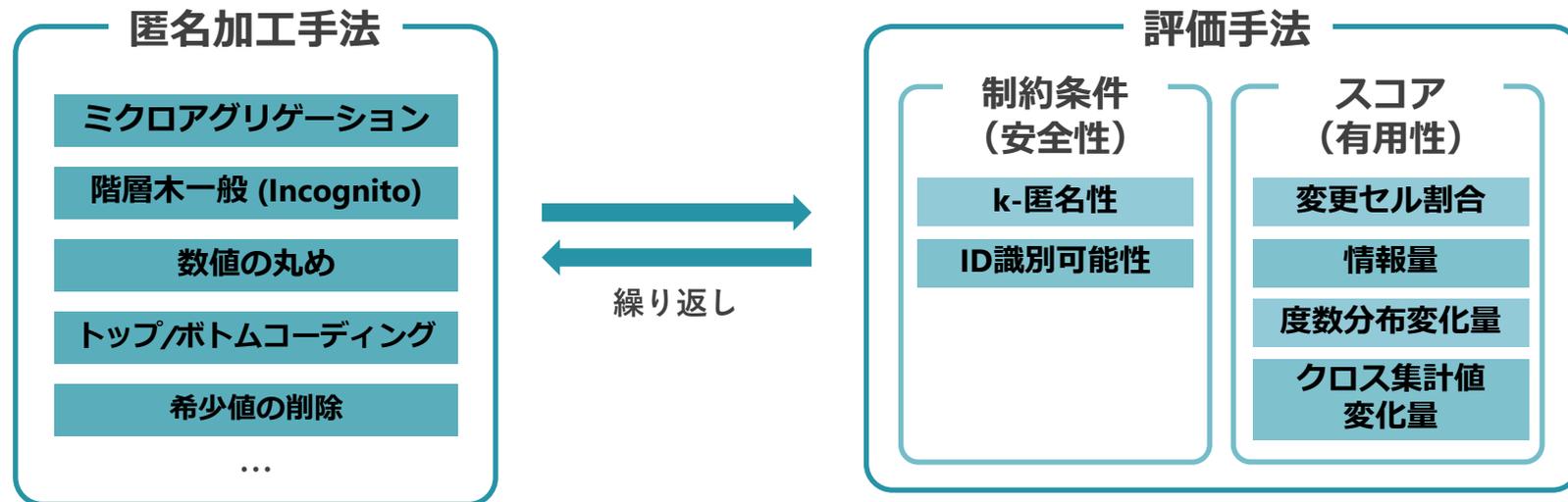


- さまざまな業界のガイドラインに準拠した匿名加工処理と評価手法を実装
- 過去案件で培った高度なデータ分析と最適化の技術を投入
- 匿名加工コンテストPWS Cupへの参加・運営を通して得られたノウハウを注入

実導入で遭遇した「匿名丸」の課題

- 多くの匿名加工/評価手法から適切な手法の選択が大変
- 匿名加工の専門家でないといと使いこなせない

➡ 匿名加工の自動化



最適な匿名加工手法を自動で効率的に探索

4

まとめ

# データマネジメントに取り組むためには プロセスの整備と技術適用の両輪が必要不可欠

## データマネジメント プロセス



## データマネジメント 技術

- データマネジメントで解決すべき課題の特定
- ロードマップ策定
- データ管理の体制や業務プロセスの整備
- ユーザーの理解度向上/啓蒙

- データ管理業務を効率的に進めるためのメタデータ管理やデータ品質向上などの技術
- 全社横断のデータ活用を可能にするデータアーキテクチャ技術
- 俊敏性/安定性/管理性を両立したデータ連携/統合技術

事業部・DXICがお客様支援を通して  
培ったコンサルティングノウハウ

システム研究開発センターにおける  
研究開発や現場適用で得られた技術

## 商標について

X Integrator、hifive、Geminant、Geminant（ロゴ）、安全見守りくん、IoX、IoXプラットフォーム、匿丸、匿丸（ロゴ）、OptBrain、OptBrain（ロゴ）、Tetralink、Tetralink（ロゴ）、日鉄ソリューションズ、NSSOL、NS(ロゴ)は日鉄ソリューションズ株式会社の登録商標です。

その他本文記載の会社名及び製品名は、それぞれ各社の商標又は登録商標です。