

データ・プレパレーション が求められる理由とは？

セルフサービスでデータを価値に変えるアプローチ

株式会社アシスト
情報基盤技術統括部
山口晃司

超サポ
愉快カンパニー

アシスト

アシストについて

設立	1972年3月
代表取締役会長	ビル・トッテン
代表取締役社長	大塚 辰男
資本金	6,000万円
売上高	317億円（2018年度）
社員数	1,160名（2019年4月現在） ※グループ会社含む
事業内容	コンピュータ用パッケージ・ソフトウェアの販売、技術サポート、教育およびコンサルティング
本社所在地	東京都千代田区九段北4-2-1 市ヶ谷東急ビル
オフィス所在地	札幌、仙台、名古屋、金沢、大阪、広島、宇部、福岡、沖縄
取引会社数	6,300社（2018年度）
主要取扱製品数	60製品（2019年4月現在）
グループ会社	株式会社アシスト本舗 株式会社アシスト北海道 株式会社のれん



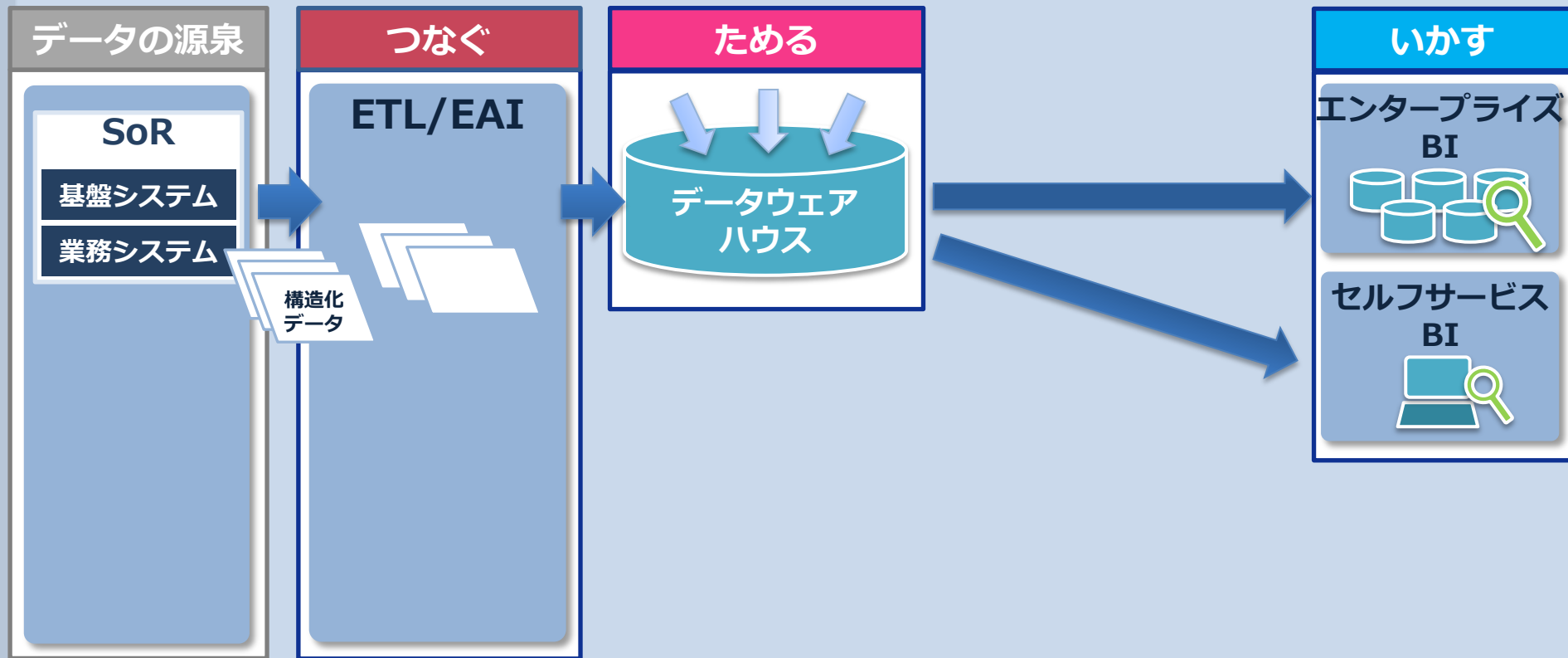


企業における データ活用の変化

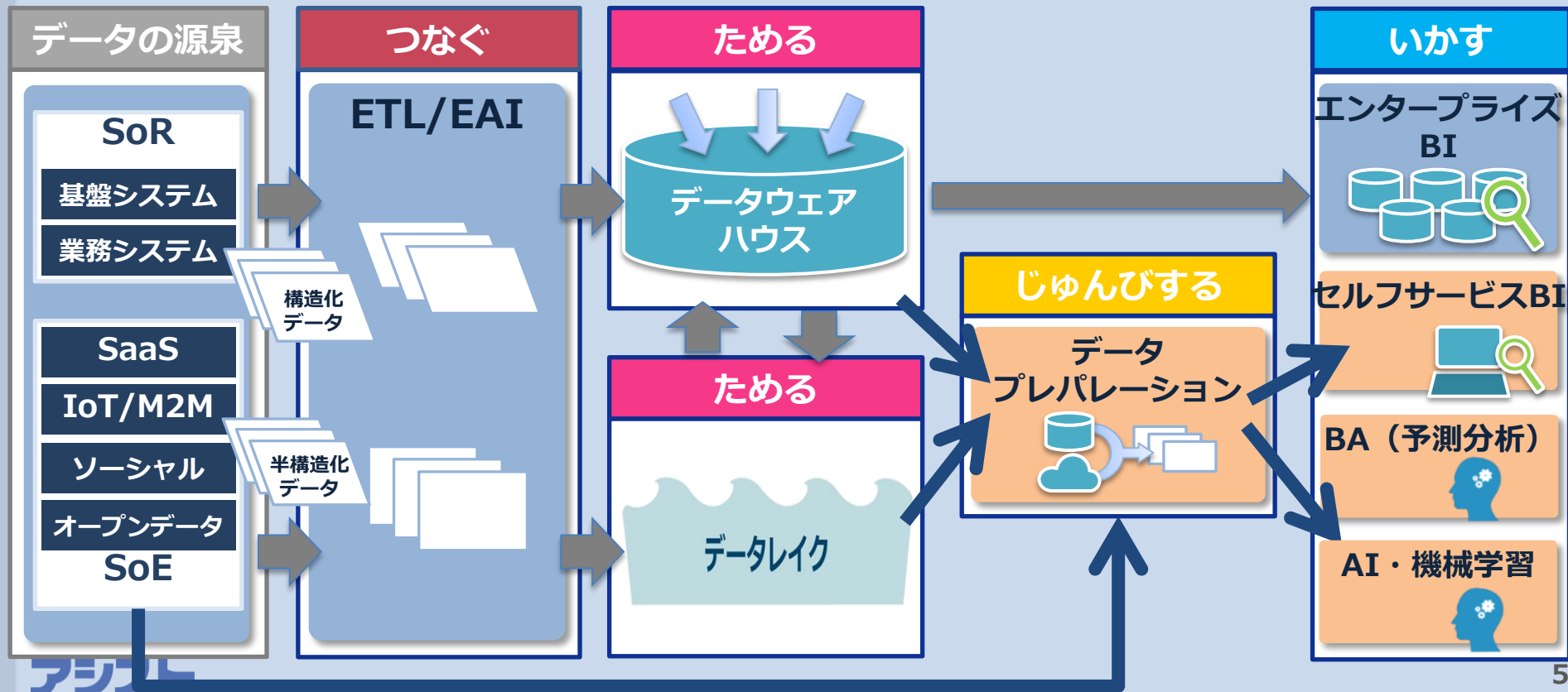
超サポ
愉快カンパニー

アシスト

これまでのデータ・マネジメント

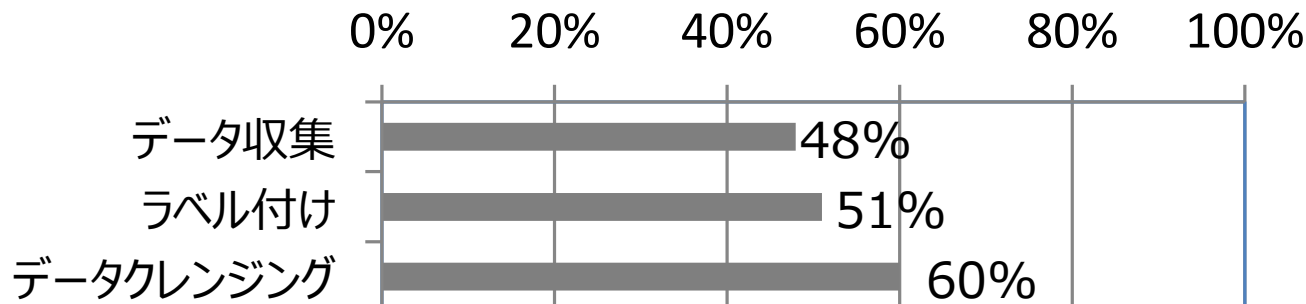


これからのデータ・マネジメント



データ準備は楽しい？ 楽しくない？

データサイエンティストが最も楽しめない業務TOP3

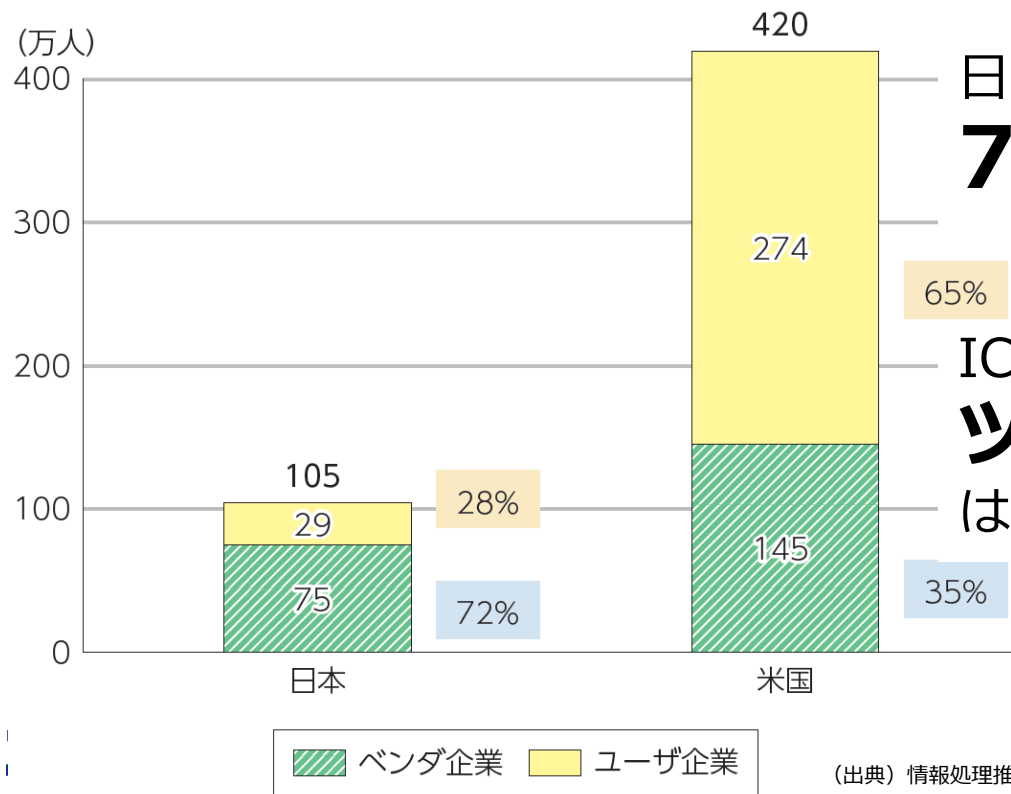


データのパターンの探索 3%
データ分析のモデル構築 3%
アルゴリズム（定式化）の洗練 5%
トレーニングデータの構築 14%
その他 10%

データ活用において、
データ準備に **80%** の労力が費やされている

日米のICT人材の比較

日本と米国の情報処理・通信に携わるICT人材



日本のICT人材は

70%以上がベンダ企業に在籍

ICT人材育成・獲得の推進と並行し
ツールを活用した民主化が
はじまっている

データ活用とその課題

AI/機械学習への挑戦

IoTやセンサーデータなどのデータを対象に、機械学習ツールやR、Pythonで予測モデル作成にトライ



データの加工作業に時間と手間が掛かるため予測モデルの精度が上がらない

セルフサービスBIの活用

データ・ドリブンな業務推進を実現するためにセルフサービスBIの利用を促進



インプットするデータの加工や整形作業に時間と手間が掛かる

データ基盤の構築

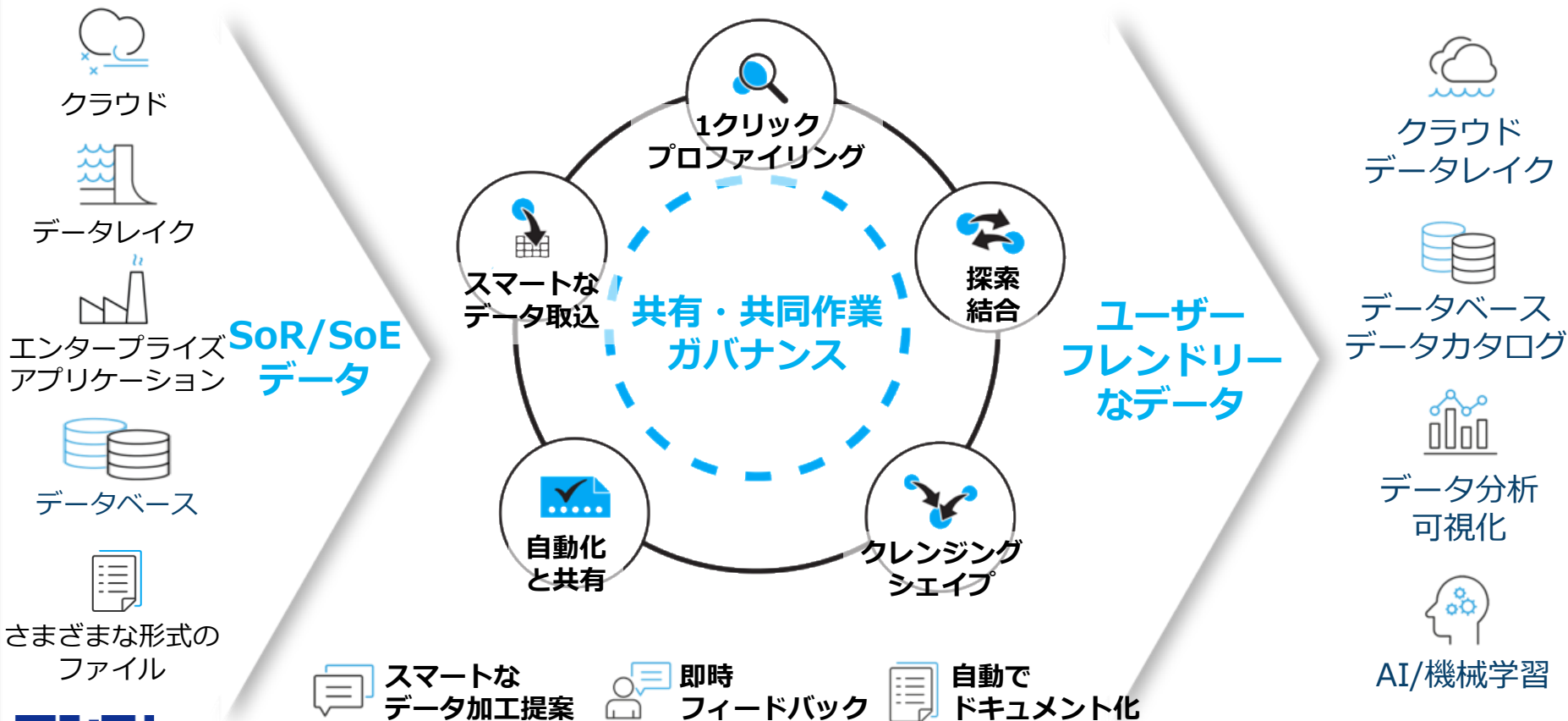
企業全体でのデータ活用を目的に、データレイクをはじめとしたデータ蓄積と活用基盤を構築



蓄積されたデータの活用には、データ構造の理解やSQLなどの高度なスキルを要する

共通して課題となるのは
データをいかに準備するか

注目が高まる データ・プレパレーション



ETLとデータ・プレパレーションツールの違い

データを加工する、という目的では
ETLツールとデータ・プレパレーションツールに**大きな違いはない**

利用するユーザ

ETL : IT部門、システム担当者

データ・プレパレーション : ビジネスユーザー、データ分析者

利用するタイミング

ETL : 定常的なバッチ処理、システム間連携

データ・プレパレーション : 半定形・非定形で不定期

**ビジネスユーザが利用できるツールを
提供することでデータ活用の裾野を広げる**



データ・プレパレーション ツールの市場について

超サポ
愉快カンパニー

アシスト

代表的なデータ・プレパレーション製品

Alteryx	2014年からクラスメソッド社が代理店として展開 2018年に日本法人設立
Trifacta	2019年に三井住友銀行で採用されたというプレスリリースあり
Tableau Prep	2017年にProject Maestroとしてベータ公開 2018年にTableau Prepとして正式リリース
Talend	ETL製品とは別にTalend Data Preparationとして展開
Informatica	Informatica Enterprise Data Lakeの機能として提供

日本にも複数の製品が進出しつつあり、
市場としても認識され始めてきている状況

Paxata (パクサタ)



2012年	米シリコンバレーで設立
2013年	製品版をリリース
2016年4月	アシストと総代理店契約を締結
2016年10月	日本語版をリリース
2017年7月	日本国内でSaaS型クラウドサービスをリリース

幅広い業種で**100社以上**の採用



デモンストレーション

超サポ
愉快カンパニー

アシスト

キャンペーン申込みデータ加工デモ

加工する元データ



Webキャンペーンのログデータ

お客様向けのWebサイトで公開しているキャンペーンの利用実績や利用したユーザーの情報が日々格納されている。



Salesforceの顧客情報

社内の営業がお客様にアプローチした場合に登録される。営業が手入力で登録したデータであるため値に統一性がない。



要件

- ・ユーザー自身でキャンペーンデータと顧客情報を結合し、顧客へのアプローチ状況を確認したい



- ・Webキャンペーンの結果から、Salesforceのリードスコアの高いユーザーを抽出し、新たなキャンペーンの検討を実施したい





データ・プレパレーション 導入事例

超サポ
愉快カンパニー

アシスト

事例 2 : 某大手製造業

ユーザーの要求
に追いつけない

ビジネス領域
の知識不足



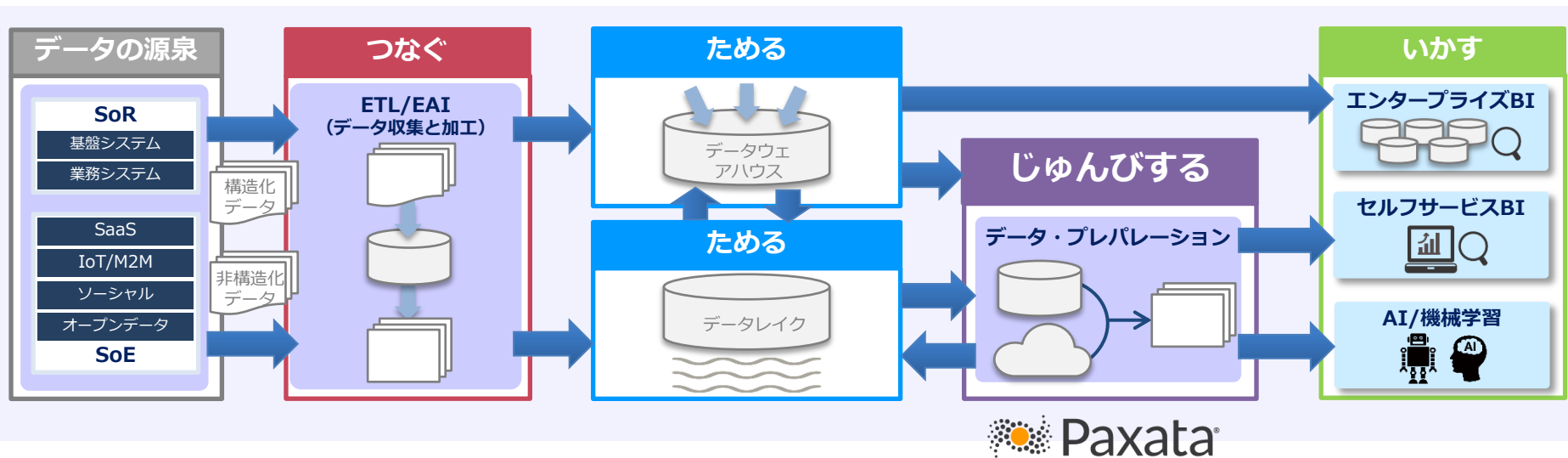
データの品質
に問題がある

AI/機械学習
に対する
過大な期待

<IT部門の方針>

ユーザー自身で可視化・分析できるプラットフォームを
構築し、データの民主化を進めることを決断

事例 2 : 某大手製造業



ビジネス部門が親しみやすいツール
IT部門として必要なガバナンス
エンタープライズレベルの共有や再利用性

事例に学ぶ：ツールによるデータ活用の民主化ステップ^o

事例2

STEP1：立ち上げ

社内の情報共有会で様々な部署のユーザーにツールともに事例を紹介

STEP2：社内案件の発掘

関心を持ったユーザー向けに実データを使った個別デモを実施

STEP3：効果の実証

- ・ ユースケースに関する個別コンサルティング
- ・ ユーザーの質問に対する個別サポート（チャット等も活用）
- ・ 社内案件の管理と共有

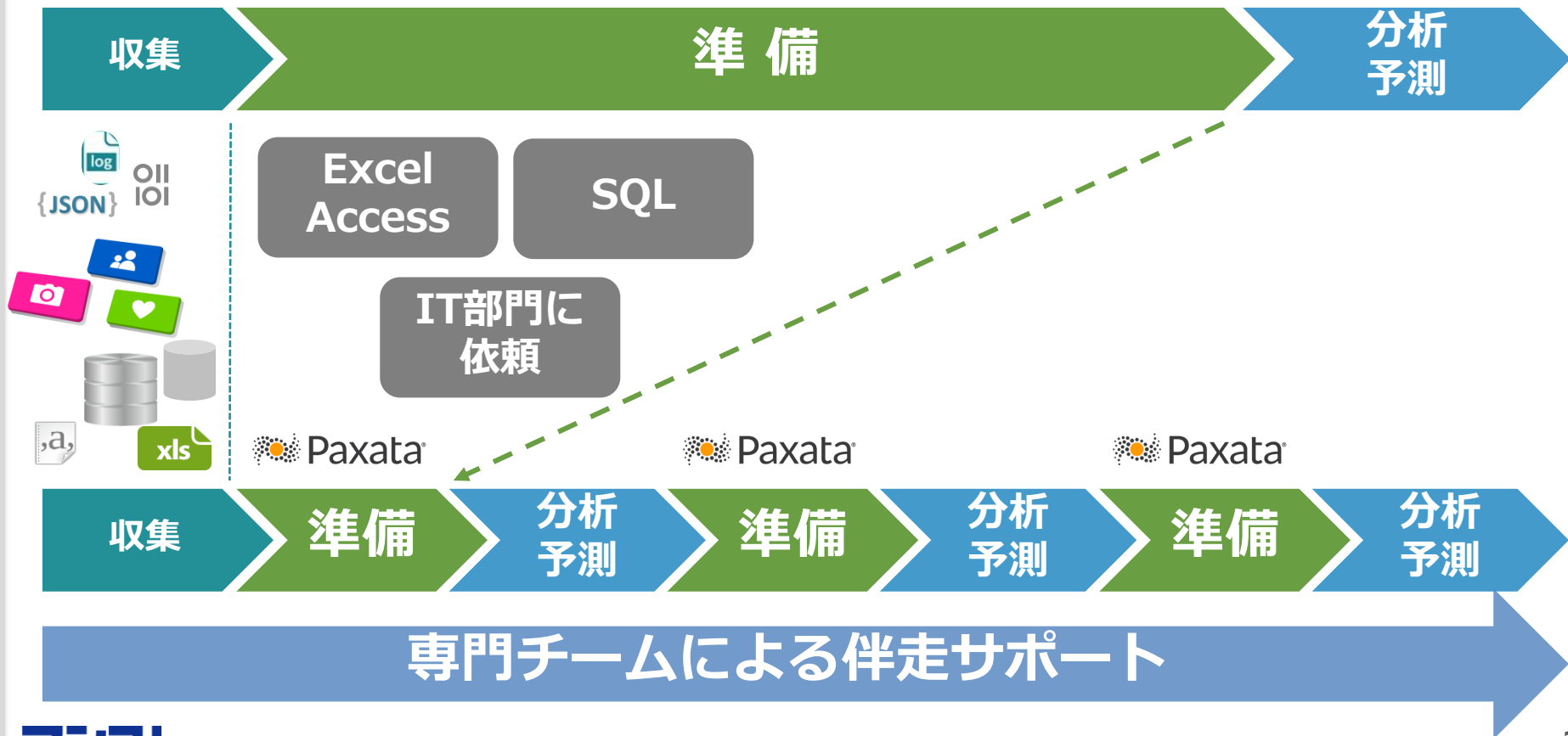
事例1

STEP4：自律的な運用・風土の形成

- ・ 動画やチャットボット等、セルフサービスヘルプツールを導入
- ・ QA対応をコミュニティベースに移行
- ・ 快適なシステム環境の維持と成功事例のさらなるチューニング
- ・ 口コミで新しいユーザー・案件を発掘

データの課題解決を支援する専門チームをIT・ユーザー合同で形成

データ活用のスピード・精度を飛躍的に向上





まとめ

超サポ
愉快カンパニー

アシスト

データ・プレパレーション が求められる理由とは？

利用できるデータや適用範囲の広がり

- SoRだけでなく、SoEデータの活用
- SoR、SoE両方のデータを組み合わせた分析により、新たな知見を得る

データを活用するユーザの広がり

- セルフサービスBIの浸透
- AI・機械学習を活用する市民データサイエンティストの登場
- ビジネス部門でのニーズの高まり、データ活用の民主化

データの分析や活用までにかかる時間の短縮化

- 定形的だがシステム化が難しいデータ加工業務の効率化
- 半定形・非定形で不定期なデータ加工処理の短時間化
- データ加工の時間を短縮化し、素早い意思決定の実現や分析時間を確保

シチズン・データサイエンスを実現する!

DataRobot × Paxata 紹介セミナー

AI・機械学習の活用がますます加速するなか、
シチズン・データサイエンティストの活躍が注目されています。

業務に精通するビジネスユーザが、ツールの力を借りて
データを準備し、高度な予測を行って事業に活用していくには
どのような方法があるのでしょうか?

本セミナーでは、シチズン・データサイエンスをパワフルに
支援するDataRobotとPaxataをご紹介します。

このような方におすすめ!

- ・機械学習のツール選定にお悩みの方
- ・DataRobotにご興味のある方
- ・Paxataにご興味のある方

Paxataの体験セミナーを開催しています



東京	: 8月8日AM/PM	8月21日AM/PM
名古屋	: 8月6日	9月10日
大阪	: 8月26日	9月30日
福岡	: 8月21日	

会場：株式会社アシスト



超サポ
愉快カンパニー

アシスト