2019年度 JUASビジネスデータ研究会
最終成果報告書

Copyright (C) 2019 JUAS All rights reserved

目次

1.研究概要
① ビジネスデータ研究会の紹介
② 研究テーマの設定
2. 研究成果
① データ連携
② 日本型データドリブン
③ データサイエンス
④ データビジネス

Copyright (C) 2019 JUAS All rights reserved

1－①ビジネスデータ研究会の紹介

～研究会活動内容～

《実施期間》

2019年5月22日 ～ 2020年1月23日

《主な活動実績》

・定例研究会 ： 8回
・講演会 ： 4回
・合宿 ： 1回

＊その他 各分科会を実施

《参加人数》

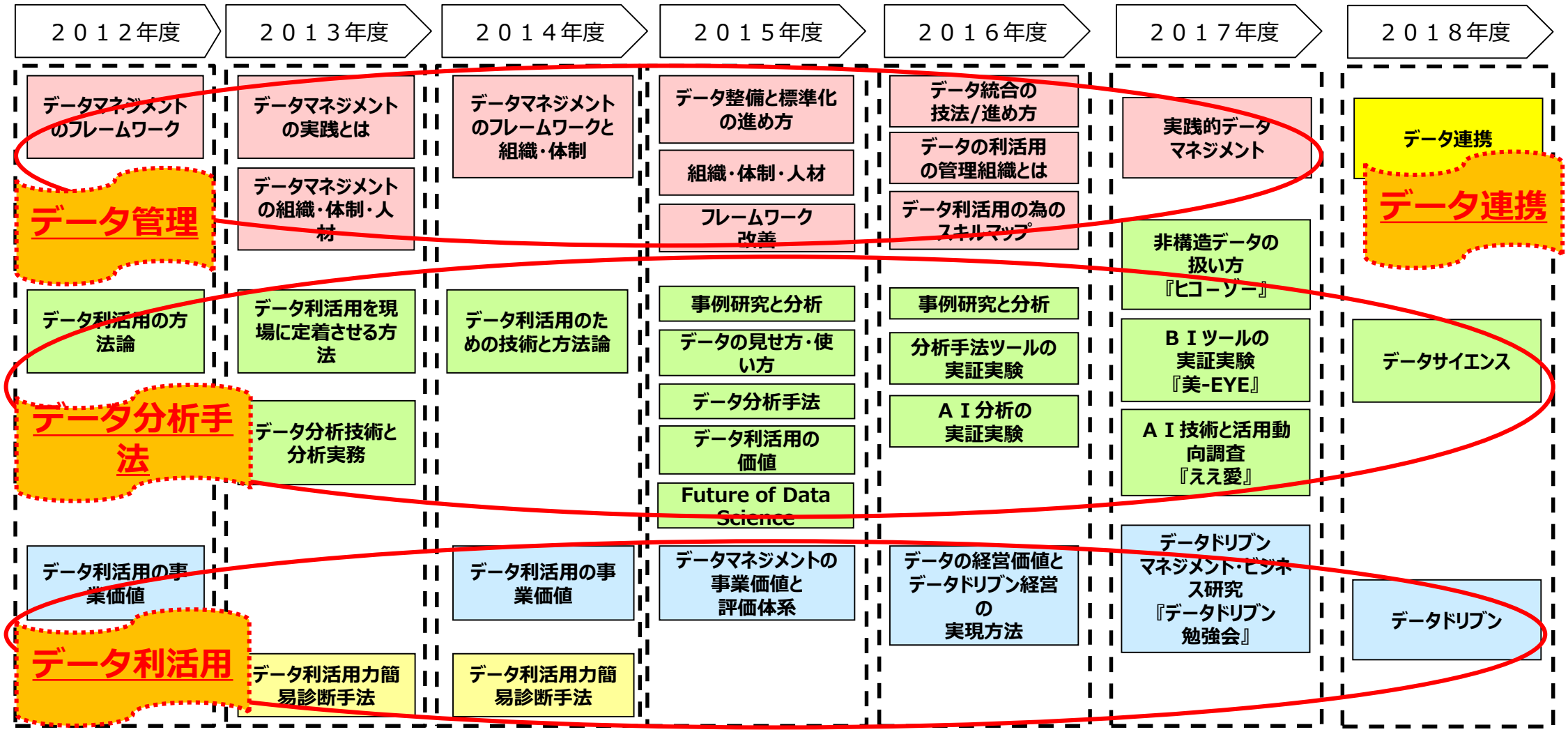
44名

Copyright (C) 2019 JUAS All rights reserved

3

1－②研究テーマの設定

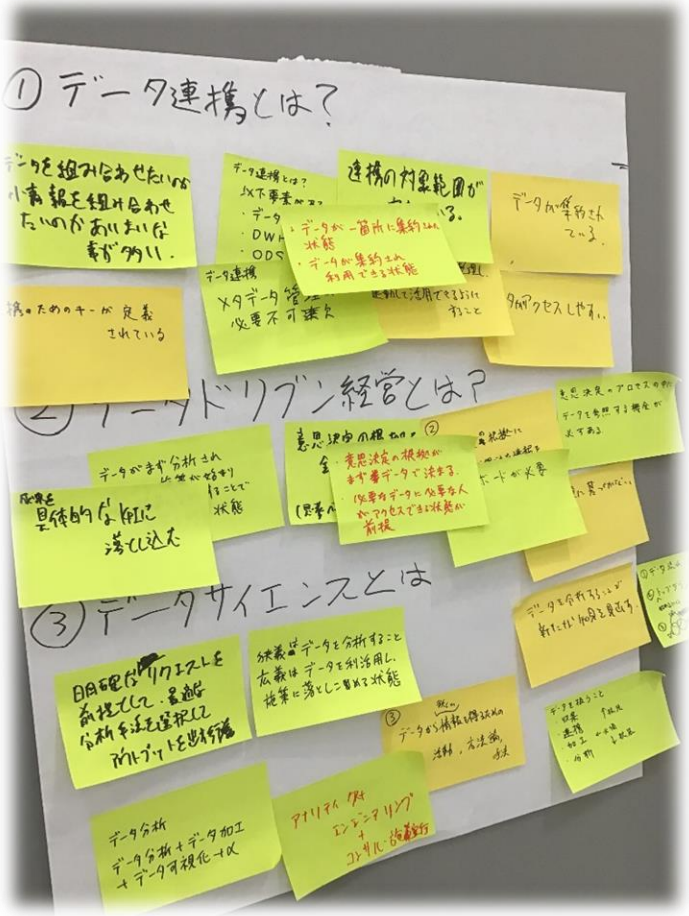
・2012年度にデータマネジメント研究会として発足。これまで、7年間データを中心に幅広く研究。

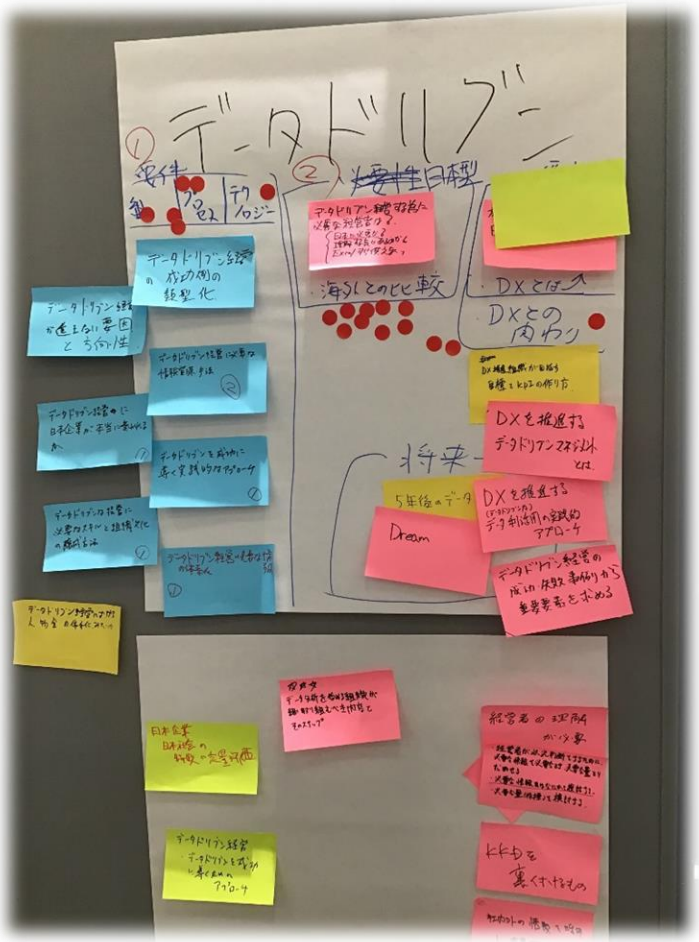


Copyright (C) 2019 JUAS All rights reserved

1 －②研究テーマの設定

・6月21日-22日　沼津プラザヴェルデで行われた合宿で、**全員参加でディスカッション**し、メンバーの課題認識を共有。議論を深め、研究テーマを決定。





Copyright (C) 2019　JUAS All rights reserved

5

1－②研究テーマの設定

4つの研究テーマ（分科会）に分かれて、研究実施

【第１分科会】；データ連携　9名
・データを社内外間で連携する場合の課題、社会動向について研究

【第２分科会】；日本型データドリブン　15名
・データ利活用組織の成功要因について研究

【第３分科会】；データサイエンス　12名
・データ分析の実践

【第４分科会】；データビジネス　8名
・ユーザ企業がデータ利活用する際の留意点（ルールと倫理）について研究

Copyright (C) 2019 JUAS All rights reserved

6

2 研究成果

Copyright (C) 2019 JUAS All rights reserved

2－①データ連携

Copyright (C) 2019 JUAS All rights reserved

目次

1　研究課題検討
　１－１　データ連携6つの課題　～2018年度課題～
　１－２　メンバーが抱える課題と2018年度課題の紐づけ
　１－３　2019年度に取り組む研究テーマ
　１－４　研究の進め方

2　ヒアリング
　２－１　ヒアリング内容
　２－２　ヒアリング先企業
　２－３　ヒアリング回答

3　文献分析
　３－１　対象文献
　３－２　文献の要旨まとめ

4　中間報告後の進め方見直し

5　ヒアリング結果の深掘り
　５－１　経営
　５－２　データガバナンス･品質
　５－３　データ流通
　５－４　教育･育成
　５－５　システム環境

6．考察･まとめ
　６－１　考察内容･提言
　６－２　今後の活動に向けて

Copyright (C) 2019　JUAS All rights reserved

1　研究課題検討
1－1　データ連携6つの課題　～2018年度課題～

■ 2018年度は、抽出された6つの課題から「①、②、⑥」について研究を実施

データ連携基盤に関する課題

課題①
データハブの構築
・マスタデータハブとトランザクションデータハブの構築
・マスタ登録機能、メタデータ管理機能の装備
・協業・関係会社向けAPIの提供　等

課題②
多様な連携方式の提供
・多様なデータ連携手段・タイミング・形式への対応
・それによる要件変化への柔軟な対応や利便性向上
・利用部門の所有するデータの活用　等

課題③
稼働環境設計手法の確立
・複数クラウド・自社DC間連携や分離ネットワーク対応
・大量データ処理や様々なアプリ連携・疎結合の実現
・データ活用と個人情報保護の両立を可能とする環境
等

データ構造や運用に関する課題

課題④
データ仕様の標準化
・コード体系の標準化・共通化、汎用コード体系の採用
・データ項目・オリジナル/コピーの明確な定義
・社外も含めた標準的なデータ仕様への準拠　等

課題⑤
データオーナーの明確化
・データオーナーによるコードの管理
・データオーナーによるデータ利用ニーズの取捨選択、
データ定義とその周知　等

課題⑥
データ流通の実現
・データ流通のための市場価格の設定
・標準的なデータ交換ルールの確立　等

Copyright (C) 2019 JUAS All rights reserved

1　研究課題検討
1－2　メンバーが抱える課題と2018年度課題の紐づけ

■ メンバーが抱える課題（問題点）を2018年度課題番号と紐づけ。
■ この結果、2018年度検討対象外の [課題④]データ仕様の標準化、 [課題⑤]データオーナの明確化に関する課題が多かった。

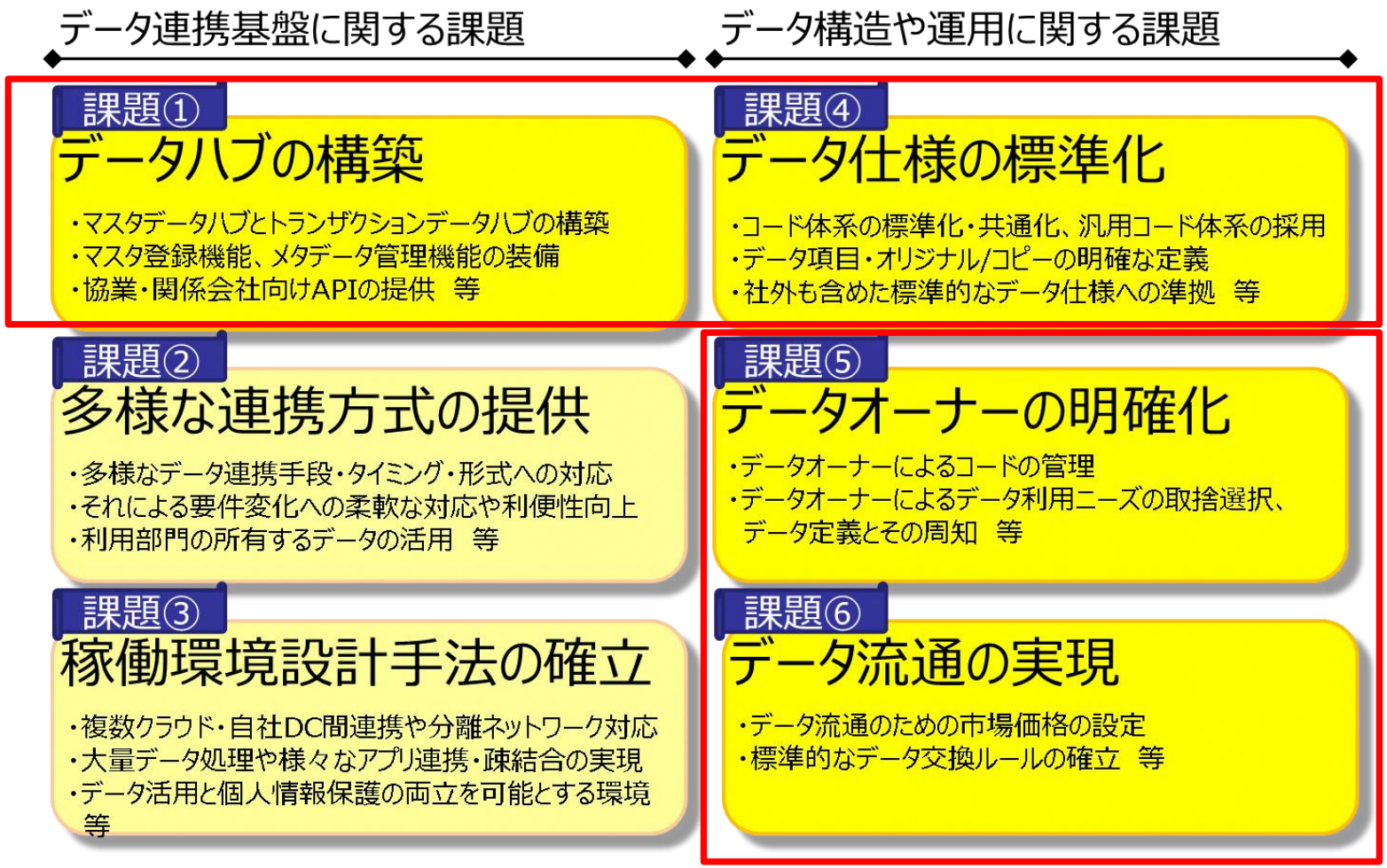
問題点	1	2	3	4	5	6	他
データの送受信が多く、データの関連性、データ処理の前後関係がわからなくなってくる				○			
社内とクラウド、マルチクラウドとなっていて物理的に離れたデータの連携・活用が困難、コスト増につながってしまう	○		○				
データ連携変換で、ETLツールや変換手法についてシステムごとに最適化されてしまっている		○					
データ連携元が同じでも、システム別にデータ連携ファイルが存在する。同じ項目をもらっていても別ファイルになっている				○			
社外データを購入・所有している部署がわからない					○		
欲しいデータを、ビジネスが求めるスピードで探して利用することができない。	○			○	○		
データを公開することに抵抗がある。					○		
社内データを外部から欲しい（購入したい）と言われた場合、即時に対応できない。					○	○	
社外から同じようなデータを購入しており、２重、３重投資となっている。					○		
社内データ形式等を変更する場合、影響調査及び関連システム対応が大規模となってしまう。	○			○			
社内連携HUBはあるが、連携に係る費用が単価化されておらず、見積が難しい。	○						
データ連携方式の最適解が無く、各システムごとにバラバラの接続方式で連携してしまう。（MQ/API等々）	○	○					
サブシステムごとにマスタデータを管理しており、同一情報であっても別々に管理している場合がある				○	○		

問題点	1	2	3	4	5	6	他
想定と異なるデータが格納されている				○	○		
ユーザがデータ項目を確認する手段がない。				○	○		
■データの点在化 データが各システムに点在化しており、集約管理されていない。	○	○					
■データ構造の不備 データが使いにくい				○			
■データ管理 構造化されたデータの管理ルール				○	○		
■社内セキュリティ制約 社内セキュリティ規定によるデータ活用の制約							○
他部門のマスターデータを使いたいが使えない					○		
社外のデータを使いたいとのニーズがあるがどうしていいかわからない					○		
部署ごとにパッケージ製品（クラウド）を利用しており、マスターデータの連携がタイムリーでない	○	○					
他部署のデータを使いたいがシステム的な影響が大きいとの理由で断られてしまった					○		
各アプリでどのようなデータを所有しているかわからない				○	○		
パッケージやスクラッチ等さまざまな仕様のアプリが存在しており、標準的な連携方式を定義できない	○	○					
アプリ更改や連携基盤の移行時に大規模な作業となり、関係者との調整や構築時に多大なコストがかかる		○			○		
	8	6	1	10	14	1	1

Copyright (C) 2019　JUAS All rights reserved

1　研究課題検討
1－3　2019年度に取り組む研究テーマ

■ “[課題④]データ仕様の標準化”、 “[課題⑤]データオーナの明確化”を中心にして研究に着手。
■ 合わせて、議論の中で関連性が出てきた課題①、課題⑥も研究で触れることにした。



Copyright (C) 2019 JUAS All rights reserved

1　研究課題検討
1－4　研究の進め方

■ データ活用に関する取り組みで先行している企業へのヒアリングと、関連文献の分析を行い、この結果を纏める

取り組み	調査の観点	想定する成果
ヒアリング［2章］	• 経営 • データガバナンス・品質 • データ流通 • 教育・育成 • システム環境	調査検討報告書
文献分析［3章］	• データ標準化 　• 文字コード、マスタデータ、コード（分類体系）、API導入 • データオーナーシップ 　• データの管理と使用、取引契約	

Copyright (C) 2019 JUAS All rights reserved

2　ヒアリング
2－1　ヒアリング内容

■ メンバーが聞きたいことを集約し、以下ヒアリングを実施。

No.	カテゴリ	ヒアリング内容
1	経営	データ分析担当者・組織が経営層の必要とするデータ（情報）を正確に理解し、提供することの難しさについて課題を持っている企業が多くあると考えております。貴社では、経営層の考え（必要とするデータ）をどの様に、データ分析担当者や部門に伝えておりますでしょうか。
2	データガバナンス・品質	他部門のデータやマスターを横断的に利用したいが、データ提供元部門が提供（≒公開）することに抵抗感があり、他部門データ使用が困難という企業が多くあると考えております。貴社において、横断的にデータ利用する場合に直面した課題や解決方法があればご教授願います。
3		マスターデータが複数のシステム（データベース）に散在しており、データの重複等の課題を持つ企業があると考えております。。貴社におけるマスターデータ管理について、一元化や統一のポリシーなどございましたらご教授願います。
4		データの品質維持やデータ辞書（意味）の運用について、事例がありましたらご教授願います。
5		データ利活用を進める上では、コード体系等のデータ標準化（ガバナンス）作成及び運用は必須ですが、実現できていない企業が多くあると考えております。貴社ではどのようにデータ標準化推進に向けた方策及び課題がありましたら、ご教授願います。
6	データ流通	貴社では社内データを販売するスキームが既に整備されていますでしょうか？ もし整備されていれば、適切な販売価格の設定方法に悩んでいる企業が多いのですが、貴社では、何をインプットにしてどういった観点で検討して販売価格を設定されているのかご教授願います。
7		社外からデータ購入を重複して実施している企業事例があります。購入したデータを会社全体で管理するスキームがございましたらご教授願います。
8	教育・育成	データ利活用を推進するには、社内における文化醸成も大事な要素ではないかと考えています。貴社でのデータ利活用推進に向けて、社内関係者に対してどのような働きかけを行ってきたのかご教授願います。
9		データ分析に必要なスキルとシステム開発・運用に必要なスキルは別物と認識しています。データ分析者の育成の取組みがございましたらご教授願います。
10	システム環境	貴社ではどのようなデータ分析基盤を構築しているのか、また、なぜそのような基盤構成にしたのか、ご教授願います。
11		業務部門がデータを利用する環境をご教授願います。 （例：全員にBIツールが提供されており、定期的な教育を実施している）

Copyright (C) 2019 JUAS All rights reserved

2　ヒアリング
2－2　ヒアリング先企業

■ データ活用に関する取り組みで先行している企業へのヒアリングを実施。

実施日	ヒアリング先業種	ご回答者
2019年10月	情報通信	
2019年10月	製造業	
2019年10月	情報通信	

Copyright (C) 2019　JUAS All rights reserved

3　文献分析
3－1　対象文献

■ 以下の文献を読み込み、述べている要旨を抽出。

文献名	概要
政府CIOポータル　標準ガイドライン群	・デジタル社会実現に向けた、対応方法やルールを記載した政府発行のガイドライン群
企業において発生するデータの管理と活用に関する研究	・企業向けアンケート調査結果のほか、工場データ、機械学習データ、個人情報の3事例に関する分析を収録
産業保安版のデータ契約ガイドライン及びセキュリティマニュアル	・経産省ガイドラインをもとに、複数者のデータ共有についても議論
Data Ownership	・カナダの非営利シンクタンクCIGIの研究者Teresa Scassaが発表したレポート ・データのオーナーシップについて、営業秘密、個人情報、著作権といった既存の法制度をもとに分析

Copyright (C) 2019 JUAS All rights reserved

3　文献分析
3－2　文献の要旨まとめ［1/2］～データ仕様標準化、データHUB～

文献名（詳細）	内容のエッセンス
[政府CIOポータル　標準ガイドライン群] 文字環境導入実践ガイドブック	・**文字コード**については**JIS X 0213**、**符号化方式**は**UTF-8**を基本とする ・**氏名**、**法人名**、**地名**の**漢字**の**外字**に対する対応方針や、**ヨミガナ**、**ローマ字**表記のルールを決める ・外字については、**文字情報基盤**が提供する縮退マッピング表を活用して**代替文字**を決めることが推奨。 ・JIS X 0213 の関連付け（マッピング表の作成）を行い、データ連携時には相互にマッピング表から外字を 代替する ・新規構築やシステム更新時には、外字の必要性を加味して、JIS X 0213と外字入りの両方の情報をもつようにする
[政府CIOポータル　標準ガイドライン群] マスターデータ等基本データ導入実践ガイドブック	・マスターデータの**データ構造**は、マスターデータとする対象ごとに矛盾が起きないよう項目を定める。（例：「事業者」や「利用者」といった違い） ・**データ項目**には意味の定義、記述形式を明確化。「必須データ項目」「推奨データ項目」「任意データ項目」を設定する。 ・**コード体系**には、汎用性の高い標準的なコードを優先し、独自のコードを使っている場合には、コードを説明したページを公開する。 ・**テンプレート**を使うことで高品質なマスターデータを設計することができる。 ・データ項目名は、わかりやすい名詞を基本とし、複数形で記載することを推奨。また、単語間をアンダースコアで繋げるスネイクケース（例：city_libraries）を推奨。 ・相互運用性を高めるために共通語彙基盤を活用する。各組織間で共通のコード体系、データ項目、データ構造とし、相互運用可能な仕組みを導入する。
[政府CIOポータル　標準ガイドライン群] コード（分類体系）導入実践ガイドブック	・コードは大きく「**識別のためのコード**（マイナンバー、法人番号、顧客コード、製品コード等）」と「**分類のためのコード**（標準産業分類や性別コード等）」に分けられる ・コードの付与方式には、**有意コード**（例：JP-CAS2017-003）と**無意コード**（コードの記号に乱数など）に分けられる。 ・ISO や JIS、業界団体等の**標準コード**を使うことで、他システムとの連携でコードの齟齬を避ける。 ・**階層化**の検討（上位コード、下位コード） ・独自のコードは、既存のコードと独自に設計したコードを示す**マッピング**表を作成 ・作成した**コードの公開**をおこない、可能な限りコード情報を API も提供する。
[政府CIOポータル　標準ガイドライン群] API導入実践ガイドブック	・API には提供又は利用することで新たな価値を生み出すことが期待されている「**API エコノミー**」 ・**データフォーマット**（データ項目名、形式、コード値、入力規則等）について行政連携データ、国際、国内、業界等の各種標準に準拠し、相互接続性、拡張性及び一意性を担保することが重要 ・当該分野にデータフォーマットの**国際標準**がある場合には、それを活用することが推奨 ・**デファクト標準**（日時など国際標準や業界固有）に沿ったデータフォーマットに準拠する ・**相互運用性**を確保するためには、データ項目を含めた API 仕様をドキュメント等の形で詳細に示す

※太文字は、ドキュメント内で指定されたキーワード

Copyright (C) 2019　JUAS All rights reserved

3　文献分析
3－2　文献の要旨まとめ［2/2］～データオーナー、データ流通～

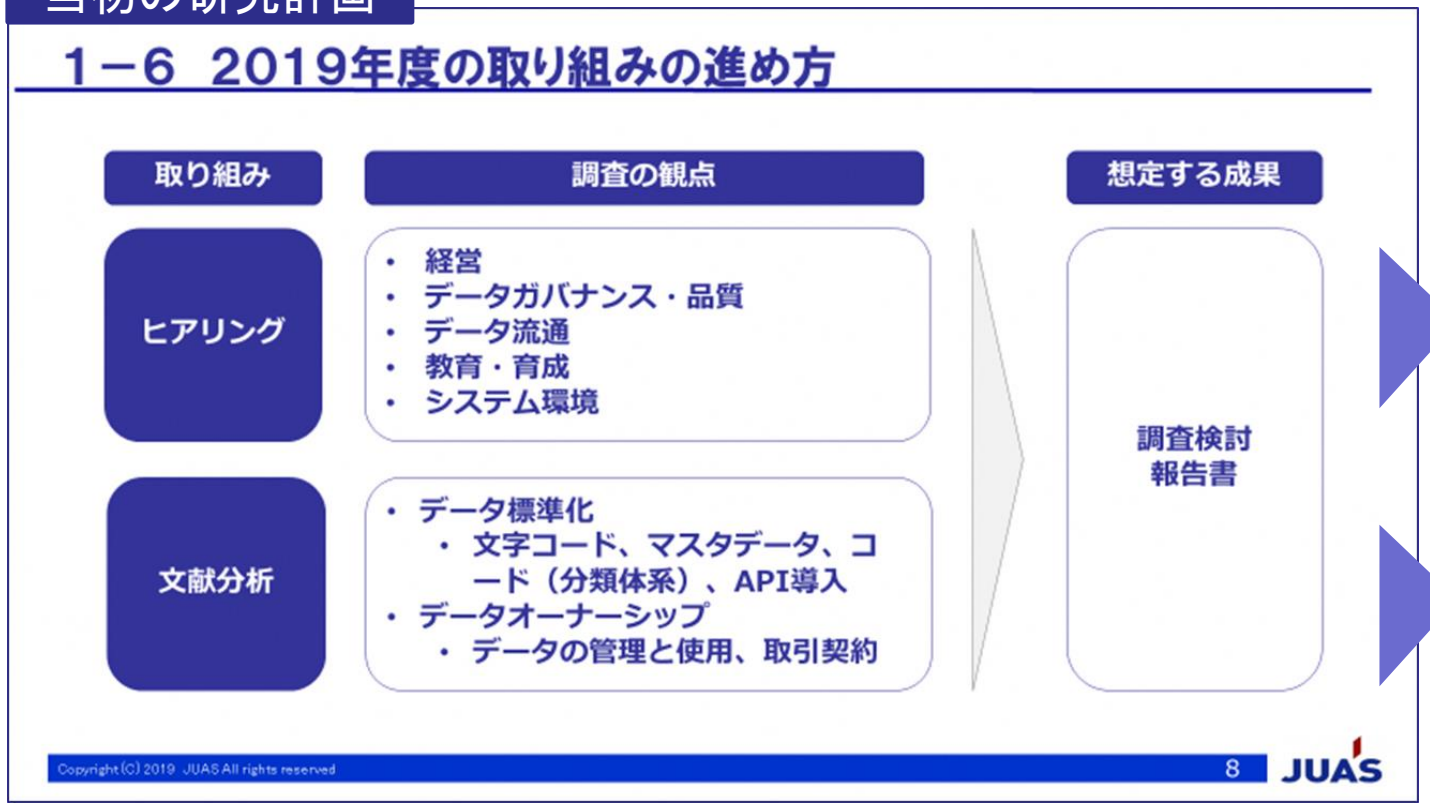
文献名（詳細）	内容のエッセンス
Data Ownership	・カナダの非営利シンクタンクCIGIに所属するTeresa Scassaが公表したレポート。 ・ヨーロッパでは、データに関する独自の所有権（オーナーシップ）を創設すべきとの議論が高まっていて、EU GDPRでは、データ主体（個人）はデータポータビリティ権を持つと規定された ・データの権利者（オーナー）の保護と、公益のためのデータ取得や再利用とのバランスをとる方法が必要
企業において発生するデータの管理と活用に関する研究（アンケート、事例検討）	・企業向けアンケートからデータ利活用を積極的に行っている企業は比較的多いものの、利活用することを期待しているがまだ実際に利活用できていないデータも相当量存在する ・本研究の分析結果ではデータ利活用においてデータ発生源への便益の提供が重要であることが示されたように、他組織や個人といったデータ発生源への配慮も必要
企業において発生するデータの管理と活用に関する研究（契約方法）	・データ利活用に関わる契約を類型化し、契約方法を提示。（「1. サービス利用規約等」、「2. 情報解析等業務委託契約」、「3. 業務提携契約」「4. 業務提携契約」） ・データ利活用に関わる契約には専門の人財が必要であり、育成の必要についても言及。
産業保安版のデータ契約ガイドライン及びセキュリティマニュアル	・産業用（プラント等）のデータについて、企業間をまたがって利活用する場合の契約ガイドライン (1) データ・オーナーシップの考え方 (2) データの保護に関する知的財産権等の法制度 (3) データの不正利用を防止する措置 (4) プラントデータの管理方法・開示方法に関する義務 (5) プラントデータを活用した結果得られる権利

Copyright (C) 2019　JUAS All rights reserved

4　中間報告後の進め方見直し［1/3］

■ 中間報告後、分科会で現状の取り組みを評価。データ利活用先行企業のヒアリングを題材にして、あるべき姿の深掘りに注力することを決定した。

当初の研究計画



先行企業のヒアリングからデータ連携、データ利活用に関して多くの学びがあった。これを題材にしてデータ連携のあるべき姿を深堀りすることへの興味が高くなった。

文献分析の結果、文献自体が完成されたものが多く、これを分析して新たな検討の実施が難しい状況とあった（文献自体のまとめ報告になる可能性）

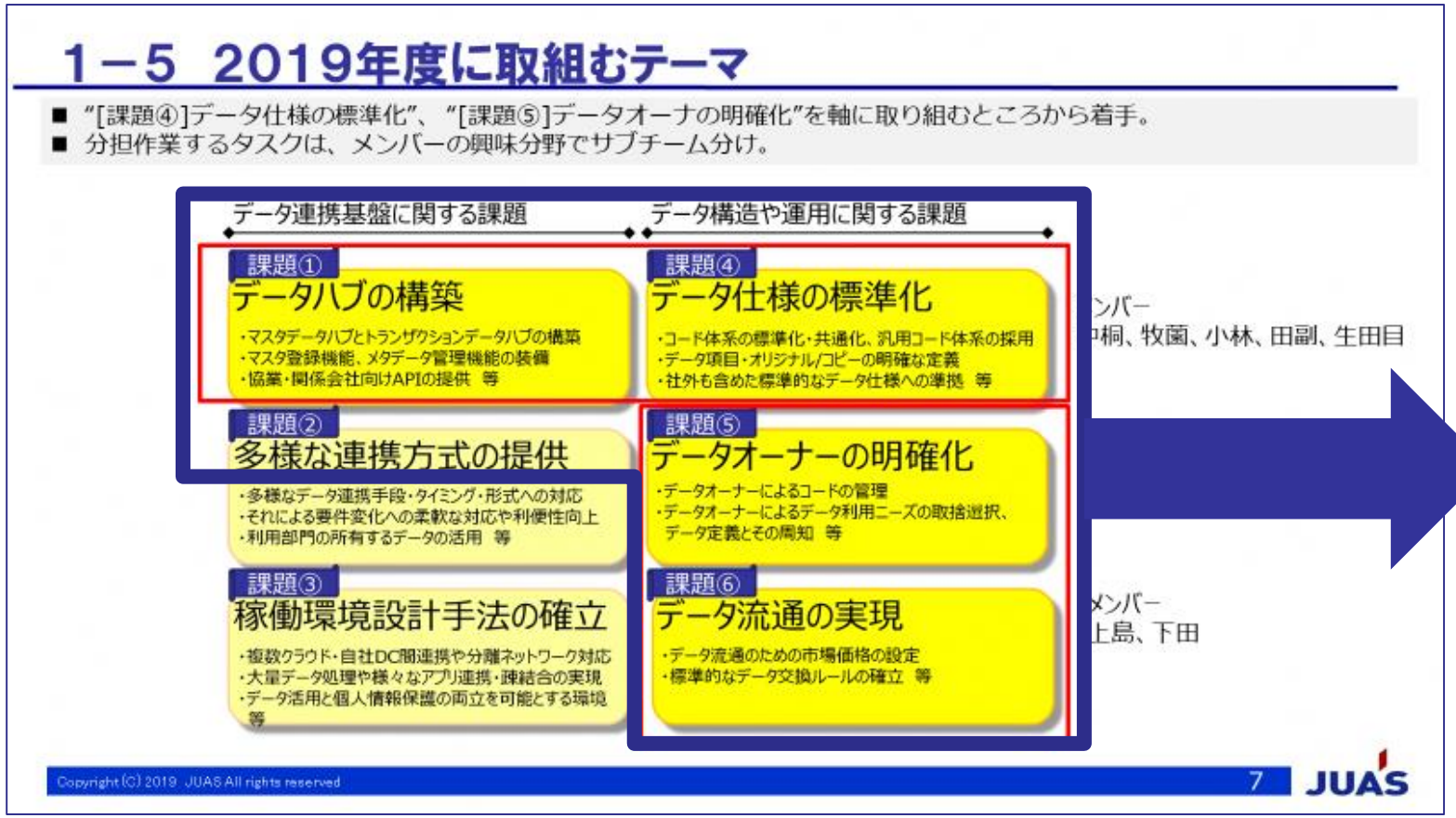
データ利活用先行企業を参考に
あるべき姿の深掘りに注力

Copyright (C) 2019 JUAS All rights reserved

JUAS

4　中間報告後の進め方見直し [2/3]

■ データ利活用先行企業を深堀りすることで、課題①④⑤⑥の課題解決が見えてくるのではないかとの仮説をベースに、社（情報通信）ヒアリング結果を中心に掘り下げていくことを決定



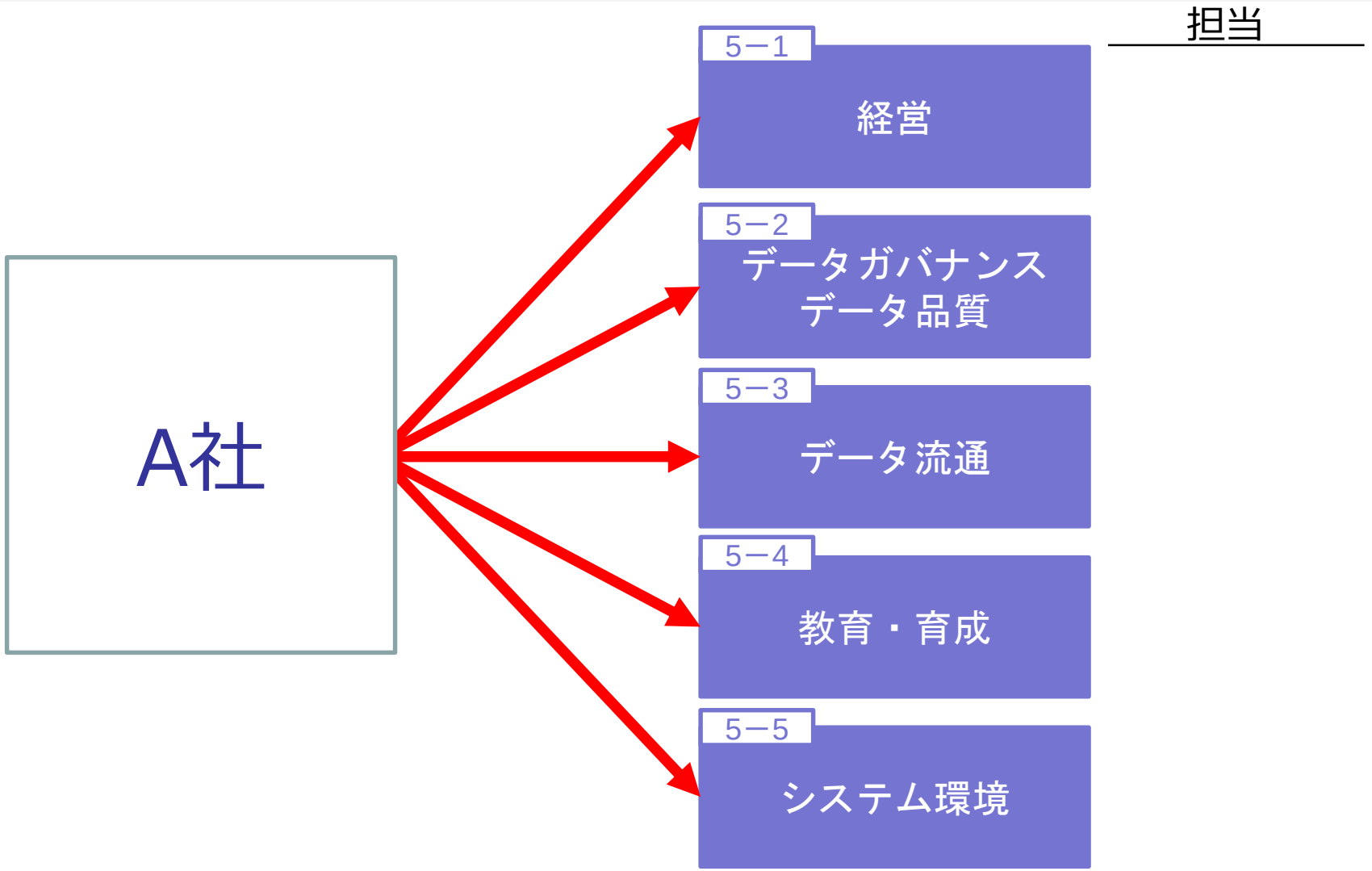
データ利活用先行企業の深堀りにより、課題①④⑤⑥の課題解決に導けるとの仮説立て

A社（情報通信）ヒアリング結果をベースに研究を開始

Copyright (C) 2019 JUAS All rights reserved

4　中間報告後の進め方見直し［3/3］

■ ヒアリングカテゴリを以下５つに分割し、メンバーで分担して深堀りを実施



Copyright (C) 2019　JUAS All rights reserved

5－1　経営［1/6］

■あるべき姿

社内のデータを利活用する基盤、仕組みを構築し、継続的に維持・浸透・発展させる推進役を設置すること

■ステップ

- **デジタル／データ最高責任者(CDO)の任命**
- **データマネージメント組織(DMO)の設置**
- **社内での利活用推進、浸透**
- **継続的な発展**

■手法

- **情報システム部と同等の組織・役割**
- **情報システム部内に設置することも可能**

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－1　経営　[2/6]　～CDO(Chief Digital/Data Officer: 最高データ責任者)の必要性～

データドリブン経営(DX)、データ利活用が進んでいる会社は、経営の三要素(ヒト、モノ、カネ)に追加し、第四の要素として「情報(データ)」を認識し、業務処理と外部コミュニケーションに活用する能力を維持し続けている。

①経営層が「情報(データ)」の価値を認識している
経営層が自らデータ操作(SQLを書く等)を行ったり、当該部門に必要なデータ取得や判断の検証を依頼している。

②経営層が「情報(データ)」を実務的に活用し、経営判断に利用している
KGI、KPIを定量的に判断したり、事象の背景を探ったり、KKD(勘、経験、度胸)の裏づけに使っている。

③「情報(データ)」の価値を維持し、品質向上を継続的に行っている
④上記を支えるための専門組織を確保している
データを"使える化"し、データ品質向上を継続的に行う組織を保有している。

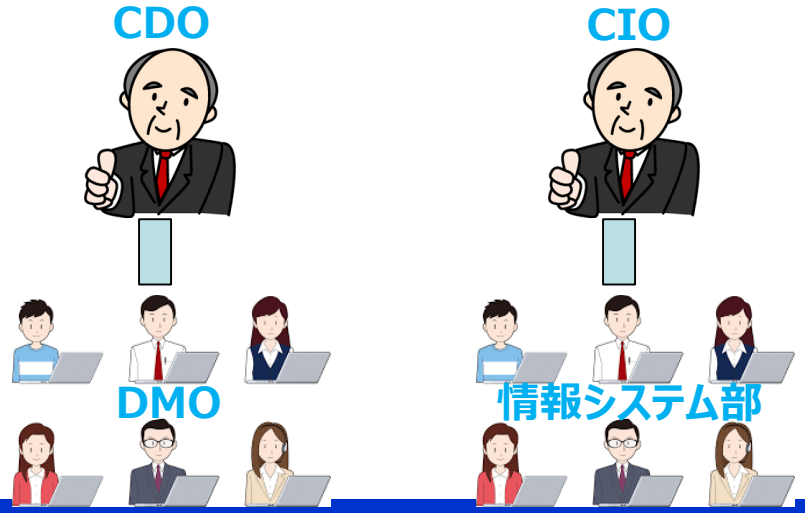
従来の多くの企業は、「情報(データ) 」を情報システム部門が主幹しているが、アプリケーションやハードウェアに趣を置いており、価値が維持されず、利活用に耐えうる品質ではないため、データドリブン経営やデータ利活用が進まない現実がある。

Copyright (C) 2019 JUAS All rights reserved

5　ヒアリング結果の深掘り
5－1　経営　[3/6]　～CDO(Chief Digital/Data Officer)の定義～

CEO（最高経営責任者）やCTO（最高技術責任者）という役職には馴染みがありますが、CDO（Chief Digital Officer＝最高デジタル責任者・Chief Data Officer＝最高データ責任者）という役職は日本ではあまり馴染みがありません。
今日、企業におけるデータの利活用、そしてそこにデジタルテクノロジーを組み込んだビジネス戦略や組織変革が、 競争力を左右し始めています。 昨今の急激なデジタル化が、企業に変革（デジタルトランフォーメーション）の潮流をもたらし、**経営陣・マネジメント層は、自社の企業戦略の中に『デジタル戦略』を組み込み、社内環境や風土を変え、顧客への新たな価値提供に寄与する変化を導くことが、必要不可欠な役割**になってきました。
CDOとは、AI、ロボティックス、IoT、デジタルマーケティング、ドローン、ビッグデータ等を有用に活用し、日々変化し続けるテクノロジーと消費者の行動に迅速に対応し、 幅広い**デジタル戦略を統括、組織を横断して改革を推進する統括責任者**の総称です。

出典： CDOとは何か？　CDO Club JAPAN　https://cdoclub.jp/cdo-club-japan/cdo/



Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－1　経営　[4/6]　～CDOとＤＭＯが主幹すべき機能～

CDOが主幹する機能は、DAMA(Data Management Association International)により提唱されている「DMBOK」(Data Management Body of Knowledge)のモデルを参考にし、データ・ガバナンスの機能を技術的な視点で、以下の17項目が構造的に整理されている。

管理対象	機能概要
①　情報方針管理	データ管理実務を当局と社内コンプライアンス、などの観点からの厳しい監視に従うための情報方針管理機能
②　データ管理者設定と適正運用監視	範囲内でのデータのセキュリティ、プライバシーのサポート、信用と信頼に値する特性に対する責任のある個人を特定するプロセス
③　データ・アーキテクチャ管理	データ・システム間に発生する相互に作用するモデル構造としてデータ・システムの関係を規定する統制管理
④　データ・モデリング管理	企業内データ・モデルの作成と組織内で柔軟性が高く安定的な業務環境を推進するための共有方法を現実的に精度高く流通する記述方法の管理と改善の機能
⑤　業務用語集	重要用語のリポジトリで業務とIT横断し共通の定義をまとめる辞書機能
⑥　データ統合管理	3種(大量データ移動/データ複製/データの見える化)の技術
⑦　データベース管理と運用	データ辞書(リポジトリー)の管理
⑧　データ・セキュリティとプライバシー管理	企業内データのセキュリティとプライバシーの確保、データ・マスキングとデータ暗号化とデータのトークン化とデータ監視機能
⑨　マスターデータ管理	データを理解するプロセスで、システム内の所在と他のデータとの関係の明確化する機能
⑩　参照データ管理	比較的変化の少ないデータでシステム共通に参照されるデータの管理
⑪　業務処理管理(BPM)	業務処理を調整する全体論的な管理機能
⑫　データウェアハウジング管理	レポーティングと解析に関する集中化されたリポジトリの作成・管理
⑬　重要データ・エレメント管理	規制管理、運用効率並びに業務インテリジェンスに重要な影響を与える属性管理
⑭　メタデータ管理	名前、場所、値、桁数、形式などの関係するデータの特徴を記述されたデータ管理機能
⑮　データ品質管理	データの品質と完全性を測定し、改善する方法並びに名寄せを含む統制機能
⑯　情報サイクル管理	情報の作成から廃棄に至るプロセスの一連を管理する機能
⑰　コンテンツ管理	デジタル化、収集と紙並びに電子化ドキュメントの分類のプロセス管理

出典：　～CDO(Chief Data Officer: 最高データ責任者)の主要役割と機能　JDMC　https://japan-dmc.org/?p=6999

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－1　経営　[5/6]　～CDOとＤＭＯの効果～

欧米では、社内に散在するデータ資源を統合し、一元管理する事で企業価値向上につながるＣＤＯおよびＤＭＯ設置が急速に普及している。設置の効果としては以下が見込める。

①規制対応
　当局からのコンプライアンス遵守要求が強まる中、CDO／ＤＭＯは当局からの問合せや監査に対して、正確かつ適正に回答を行える。

②一元管理
　物理的統合、仮想統合を問わず、一元管理を行い共通化、可視化することで、企業全体での利用が可能となり、社員が当たり前に使える環境となる。

③価値化
　データを企業内で利用し続けるために、継続的なデータの保守、改善を行うことで価値を維持・向上できる。セキュリティ視点、プライバシー配慮、法令順守等の統制が必須事項となる。

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－1　経営　[6/6]【参考】DMBOK2からのメッセージ

Data Management Association International（DAMA-I）が、刊行しているデータマネジメント知識体系ガイド（The DAMA Guide to The Data Management Body of Knowledge　：略称ＤＭＢＯＫ）。2nd Editionは、2017年に米国で刊行され、2018年には世界に先駆けてＤＡＭＡ日本支部が監訳し、日経BP社より日本語版を刊行しました。

DMBOK2からのメッセージ

章	英語オリジナル	日本語訳
第1章	Data is a **vital enterprise asset**.	データは企業にとって**重要な資産**である。
	Data and information have been called the '**currency**', the '**life blood**', and even the '**new oil**'.	データとインフォメーションは、情報経済にとって「**通貨**」であり、「**企業の生命線**」であり、さらに「**新しい原油**」とさえ呼ばれるようになった
	Long-standing definitions of data emphasize its role in **representing facts about the world**.	データの定義にあたり「**データは万物に関する事実を表現する役割を持つ**」ことが長い間強調されてきた。
第3章	…data **represents things other than itself**	データはそのデータ以外の何かを表す （**データは自分のことを語れない**）
第14章	Traditional Business Intelligence provides '**rear-view mirror**' reporting… Data science techniques are used to provide '**windshield**' view of the organization.	•伝統的なビジネスインテリジェンスは、過去の傾向が持つ特徴を「**後方確認ミラー**」として報告する・・・ •データサイエンス技法が組織の将来像「**フロントガラス**」を提供する
第16章	Data governance is about '**Doing the right things**' and data management is about '**Doing things right**'	•データマネジメントとは「**正しくことを行う**」こと •データガバナンスとは「**正しいことを行う**」こと
第17章	Organizations don't change, people change	組織が変わるのではなく、**人が変わる**
	People don't resist change. They resist being changed	人々は変わるのに抵抗するのではなく、**変えられることに抵抗する**

出典：ＤＭＢＯＫとは？　データマネジメント協会 日本支部　http://www.dama-japan.org/Introduction.html

Copyright (C) 2019 JUAS All rights reserved

5　ヒアリング結果の深掘り
5－2　データガバナンス・品質　[1/10]

■あるべき姿

データの一元管理＆信頼性の高いデータの管理（野良データがない）

■ステップ

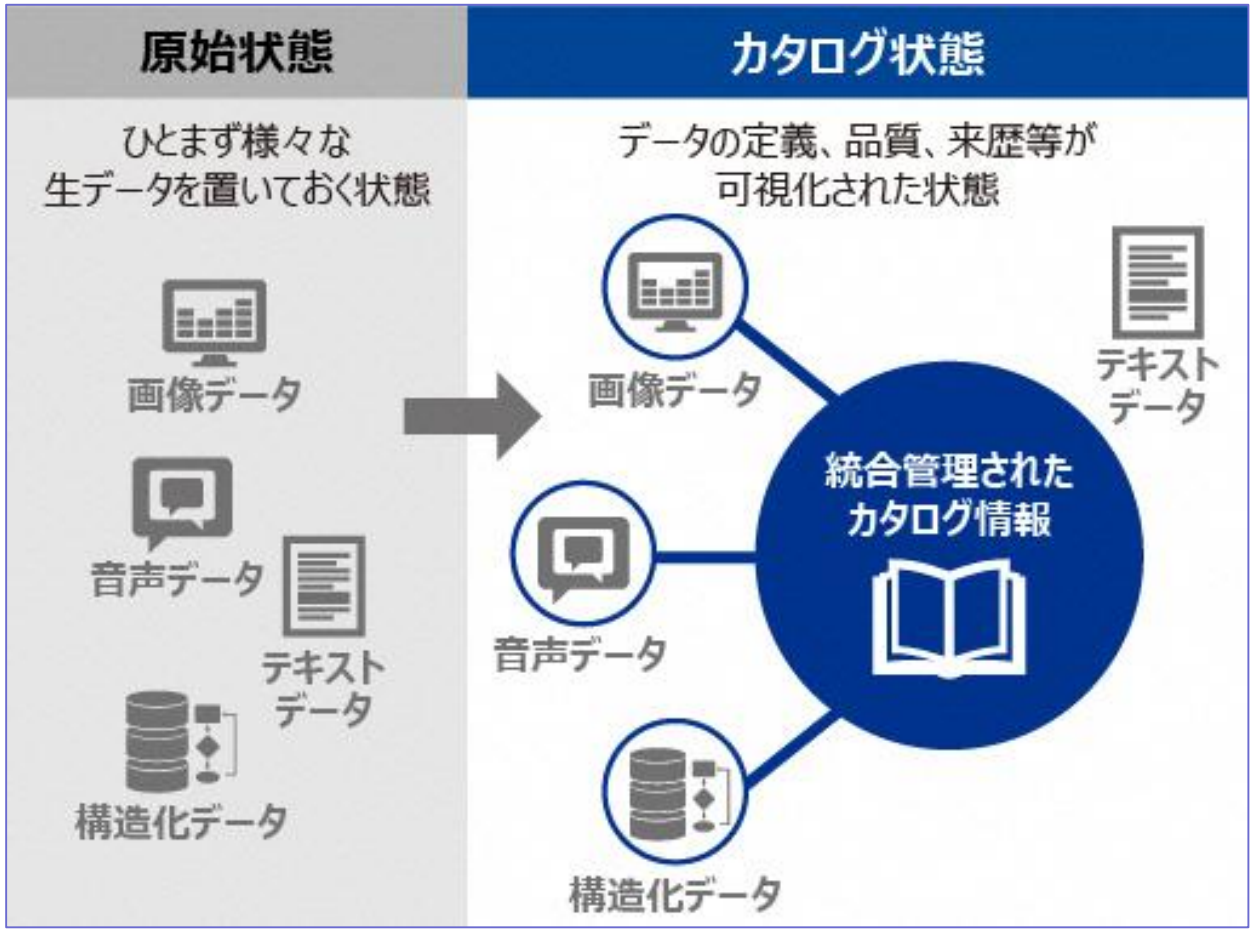
①データカタログ
①-1　データカタログ導入を進めるための効果算出を行う
①-2　自社に最適なソリューション選定を全社横断的視点で選び、データは出来る限り自動化登録する
①-3　カタログに付記するビジネスメタデータの管理も含めデータオーナーを明確化する
②ＩＤ共通化（データ利活用整備）
②-1　アプリケーションに依存しないマスターデータの管理
②-2　データ利用・活用環境の整備（データカタログ整備）　※②-1、②-2は順不同

■手法

①データカタログを導入する
②MDM

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－2　データガバナンス・品質　[2/10]　～①-1 データカタログとは～

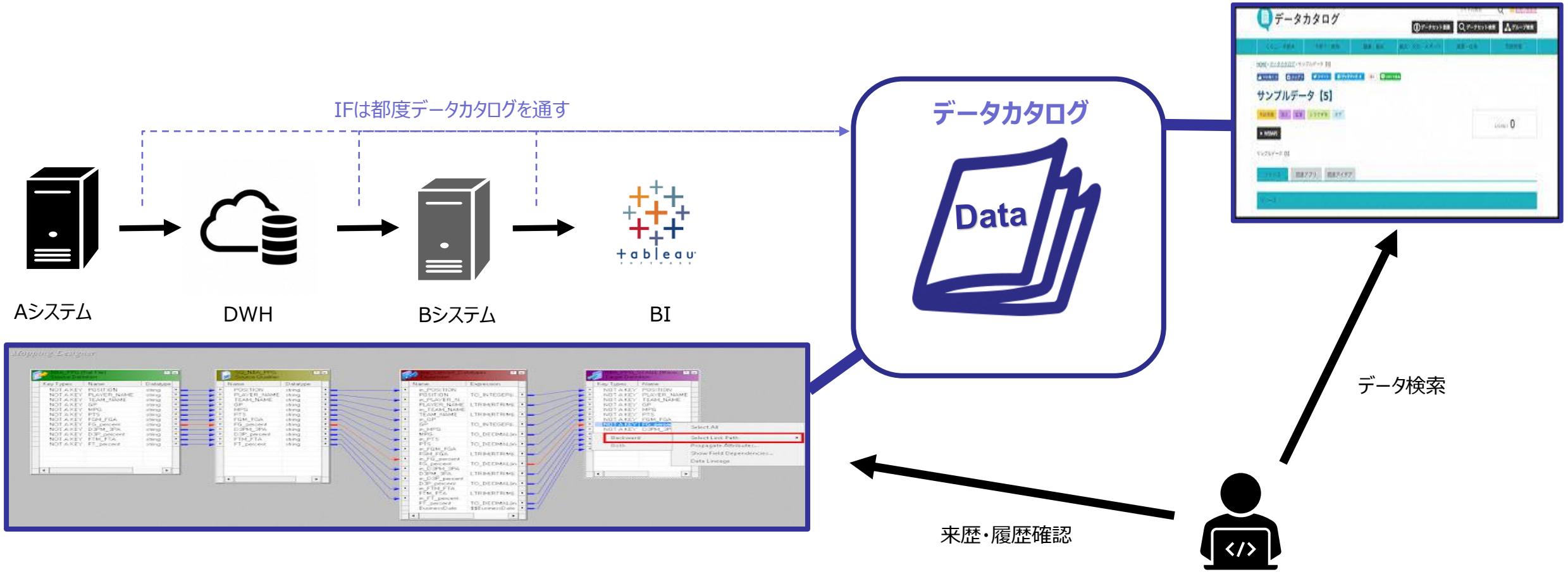


- **企業内の分散したデータを一元し、体系的にまとめた管理台帳**
- **データの“定義”、“品質”、“来歴”等を可視化する**
- **構造データ（メタデータ管理）に加え、非構造データも管理するものをデータカタログというとの説もある**

参考：https://home.kpmg/jp/ja/home/insights/2019/03/data-governance-risk.html

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－2　データガバナンス・品質　[3/10]　～①-1【参考】データカタログのイメージ～



Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－2　データガバナンス・品質　[4/10]　～①-2 データカタログの導入検討の背景～

背景①

データレイクの登場

これまでの構造データだけでなく、非構造データがリポジトリに保存され、データレイクの中にどんなデータが入っているか、煩雑な管理になってしまう
また、データレイクはメタデータを効率的に管理できる能力に乏しく、ツールの必要性も高い

背景②

データ活用の阻害

データサイエンティストはデータ収集や準備などの単純作業に作業の50%～80%を費やしているとも言われている。そのため、より環境を整備して分析や研究効率を上げることを求められている

背景③

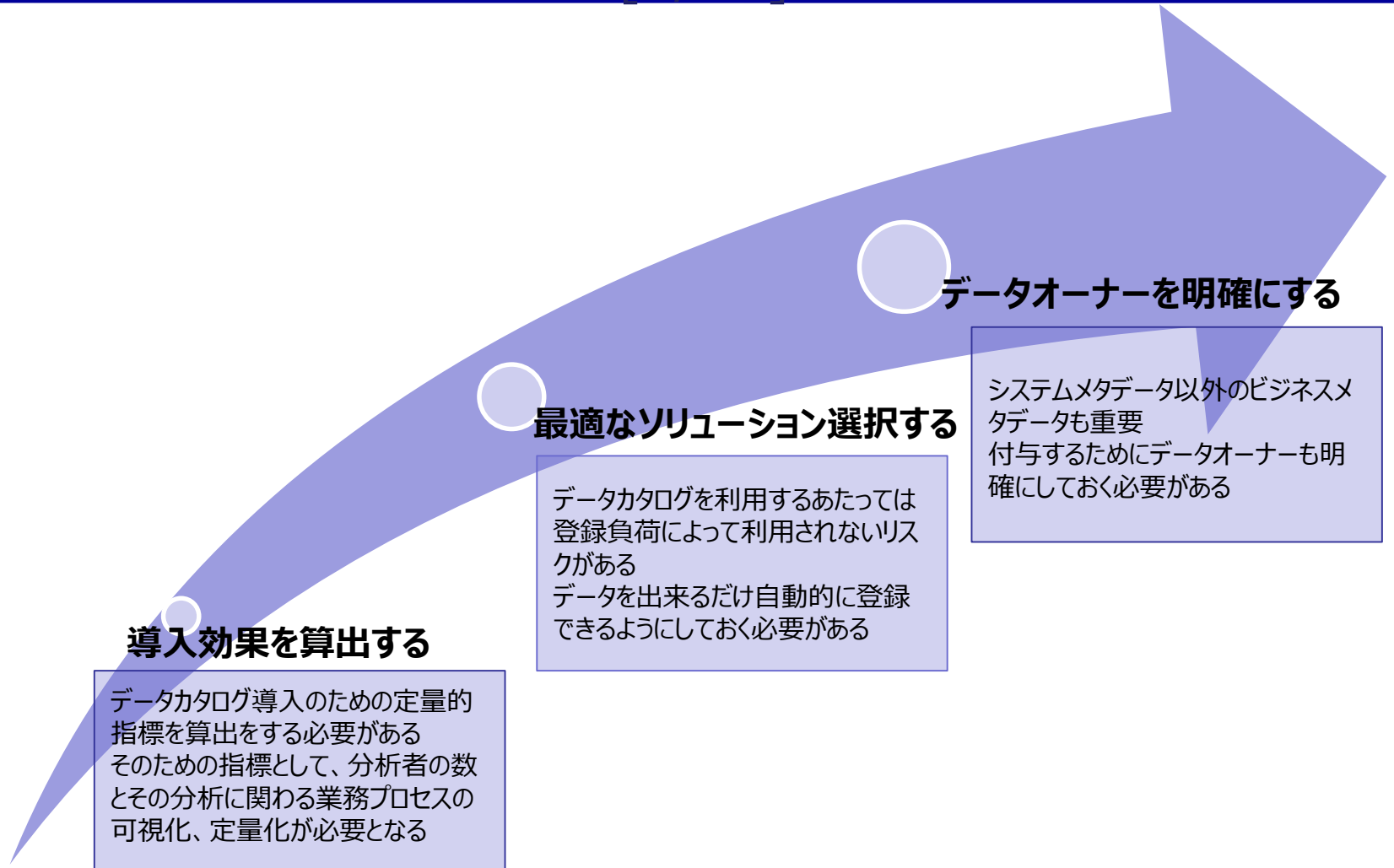
データガバナンス

GDPRや個人情報保護法等、ガバナンスを求められるシーンが増えてきており、どのデータがどのシステムから来てどのシステムで利用いるといった来歴や履歴の管理が重要になっている

参考：https://blogs.informatica.com/jp/2019/05/22/suitable-edc/

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5 － 2　データガバナンス・品質　[5/10]　～①-3 データカタログ導入のためのステップ～



← データをシステム個別最適でのソリューション選定ではなく、ビジネスや組織横断での最適化の観点でみる視点が必要 →

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－2　データガバナンス･品質　[7/10]　～② ゴールデンレコードの整備～

ゴールデンレコード：
企業もしくは組織が保有するすべてのデータ要素を「明確に定義できる単一のバージョン」。
全社システムへブロードキャストされても問題ないデータ品質が必要。

メタデータ的観点
レコードに必要十分なデータ項目を保有しているか
●検討事項
・全社アプリケーションに適用できる項目かどうか
・ドメイン※の関連付け

インスタンス的観点
各データ項目の値そのものが正しいかどうか

●検討事項
・各項目において、最も信頼できるソースはどれか
・結合のワークフロー、承認プロセスを整備
・重複削除

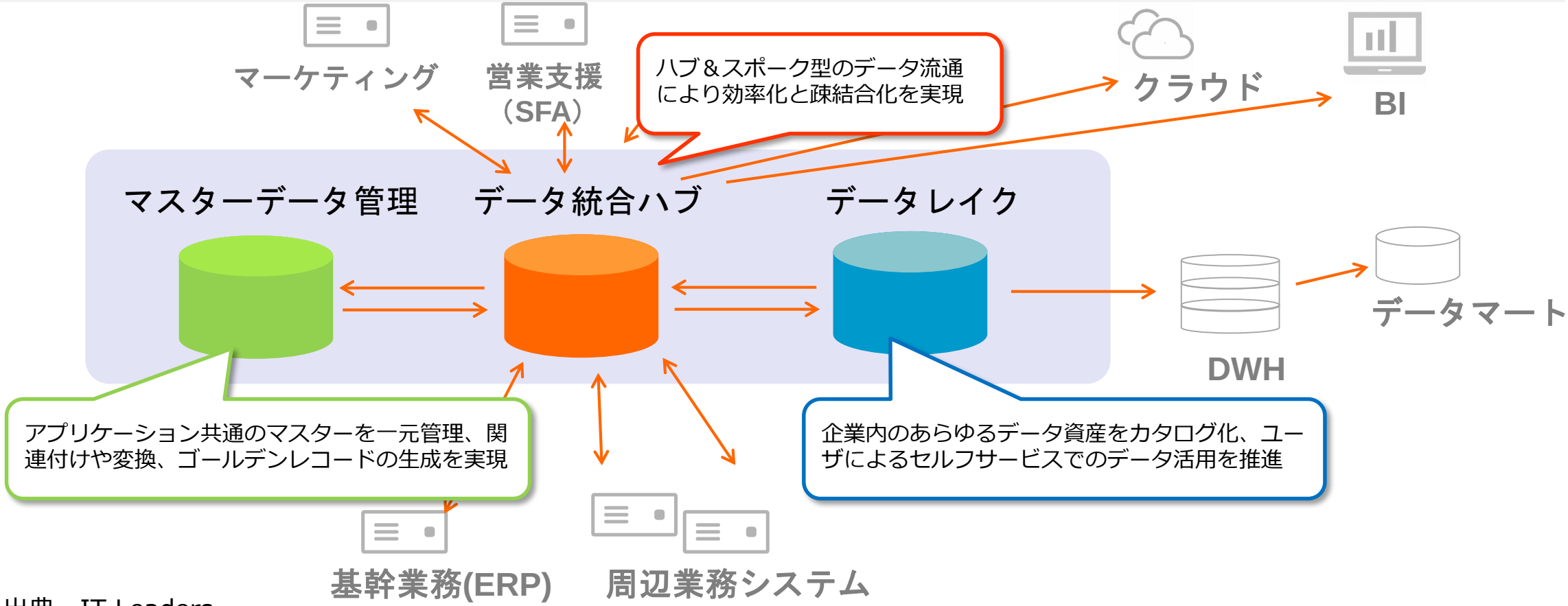
※ドメイン
MDMの世界ではマスターデータの種類。単純化すれば顧客名や商品名といったものが単一のドメイン。ゴールデンレコードによって、各ドメインにおいて最も純度・信頼性の高い完全なデータにアクセスすることができるようになる。

出典
中山 嘉之氏のブログ　https://www.it-innovation.co.jp/2014/12/18-114835/
Informaticaブログ　https://blogs.informatica.com/jp/2016/07/14/golden_record/

Copyright (C) 2019 JUAS All rights reserved

5　ヒアリング結果の深掘り
5－2　データガバナンス・品質　[8/10]　～② データ利用・活用のアーキテクチャ～

マスターデータを複数の仕組みで共有・活用するには、データハブが有効な解決策。
データハブで多種多様なシステム間のデータを一元的に収集、タイムリーな配信、データ品質の担保と粒度の均一化、信頼性、安全性の担保などをおこなう。

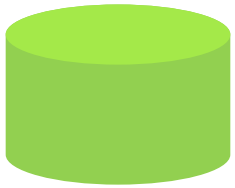


出典　IT Leaders
https://it.impressbm.co.jp/articles/-/15855

Copyright (C) 2019 JUAS All rights reserved

5　ヒアリング結果の深掘り
5－2　データガバナンス・品質　[9/10]　～② 各データベースの役割と機能配置～

マスターデータ管理

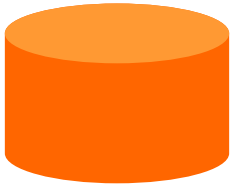


- 役割

特定アプリに依存しない、横断的なマスターデータの拠り所

- 機能要件
  - マスターメンテナンスとワークフロー
  - 名寄せとゴールデンレコード生成
  - 相互参照表の管理
  - 履歴とリレーション管理
  - ユーザーへの360°ビューの提供

データ統合ハブ

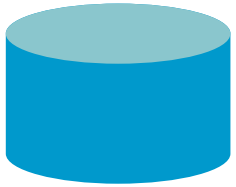


- 役割

オンプレ/クラウド、業務/分析問わずあらゆるシステム間のデータ集配信

- 機能要件
  - 各種システムへの接続性
  - Pub/Subモデルの集配信
  - さまざまなレイテンシー要件
  - ハブストアの自動データ管理
  - 運用監視、アラート

データレイク



- 役割

あらゆるデータを蓄積、ユーザーに供給し分析活用を促進

- 機能要件
  - 大規模データの統合処理
  - リアルタイムストリーミング処理
  - データの検索、把握
  - ビジネス用語の管理と紐づけ
  - データのセルフ抽出・加工

出典　IT Leaders
https://it.impressbm.co.jp/articles/-/15855

Copyright (C) 2019 JUAS All rights reserved

5　ヒアリング結果の深掘り
5－2　データガバナンス・品質　[10/10]　～② マスターデータ管理の重要な部分～

目的の設定
マスターデータ管理（MDM）の概念については、日本に入ってきたのは2007年の終わりから2008年頃。そこから何度かのブームがおきておりMDMツールも整備されてきているが、『とにかくMDMツールを導入して、データを一元管理してしまえばよい』ということではコストがかかるばかりになってしまう。MDMは手段であり、『何のためにMDMを構築するのか』かという目的を明確にすることが重要。

成功させるための必要条件
①目的を明確にする
　MDMを構築することで“こんなことを実現したい”ということを明確にする。 MDMのプロジェクトは個別業務アプリケーション開発とは異なり、様々なステークホルダーを巻き込む必要があり、そのためにも関係者を説得する熱量が重要

②実現したいイメージを明確にする
　MDMのプロジェクトを進めるうえでは、問題点を自分で考え、自分なりの解答を持つ必要がある。そのうえで、必要に応じ外部の専門家に問題を具体的に示して、意見を求め、それによって自分の頭の中を整理する。

③MDMの推進プロジェクトオーナーとの信頼関係
　熱い思いを持った担当者がプロジェクト推進し、その思いを理解しているオーナーが後押しする体制
　→CDOや部署横断的な組織の必要性

　出典　IT Leaders
　https://it.impressbm.co.jp/articles/-/17694

Copyright (C) 2019 JUAS All rights reserved

5－3　データ流通［1/10］

■あるべき姿

社内間及び社内と社外でデータ活用するために、必要なルールが社内に整備されている。

■ステップ

・データオーナの定義
・契約スキームの整備
・データ流通ポリシーの整備

■手法

①データごとにデータオーナを定義する。
②データ流通（主に社外との）に関するデータ契約雛形を準備する。
③データ流通のポリシーを決めるために、事前に課題等を整理する。

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－3　データ流通［2/10］～①-1省庁ガイドラインが社内ルールの整備に応用できる～

■ 経済産業省が、企業間のデータ取引を想定した作成した契約ガイドラインが参考になる

基本視点
データ流通・利活用の重要性と課題：データは保有するだけでは大きな価値がなく、利用する方法を開発することで価値が創出される。契約に際しては、データの利用権限および発生した利益を、適切に分配することが重要。他方、データの流出や不正利用のリスクへの配慮も必要。
契約の高度化：本ガイドラインは、あくまで契約で定めておくべき事項を示したにとどまる。したがって、契約当事者が協議し、本ガイドラインを参考としつつ、データの創出や利活用に対する寄与度等を考慮し、取引の実状に応じて契約を高度化させていくことが望ましい。
イノベーションの促進：多様な立場に配慮したデータ契約の考え方や契約条項例等を示すことにより、データ利用の促進を図り、オープン・イノベーションを促進する。
国際協調：クロス・ボーダー取引が一般化する状況下、データの越境に関する問題も考慮する。

Copyright (C) 2019 JUAS All rights reserved

5－3　データ流通［3/10］～①-2用語の確認：データオーナーシップとは～

■ オーナーシップ（Ownership）という語には、大別して2つのニュアンスがある。

1. 所有者であること（権利の帰属先）

- データオーナーシップといえば、一般にはこの用法
  - 1-1.個人のプライバシー保護と、公益とのバランス
  - 1-2.組織間のデータ流通における権利処理・標準ルール

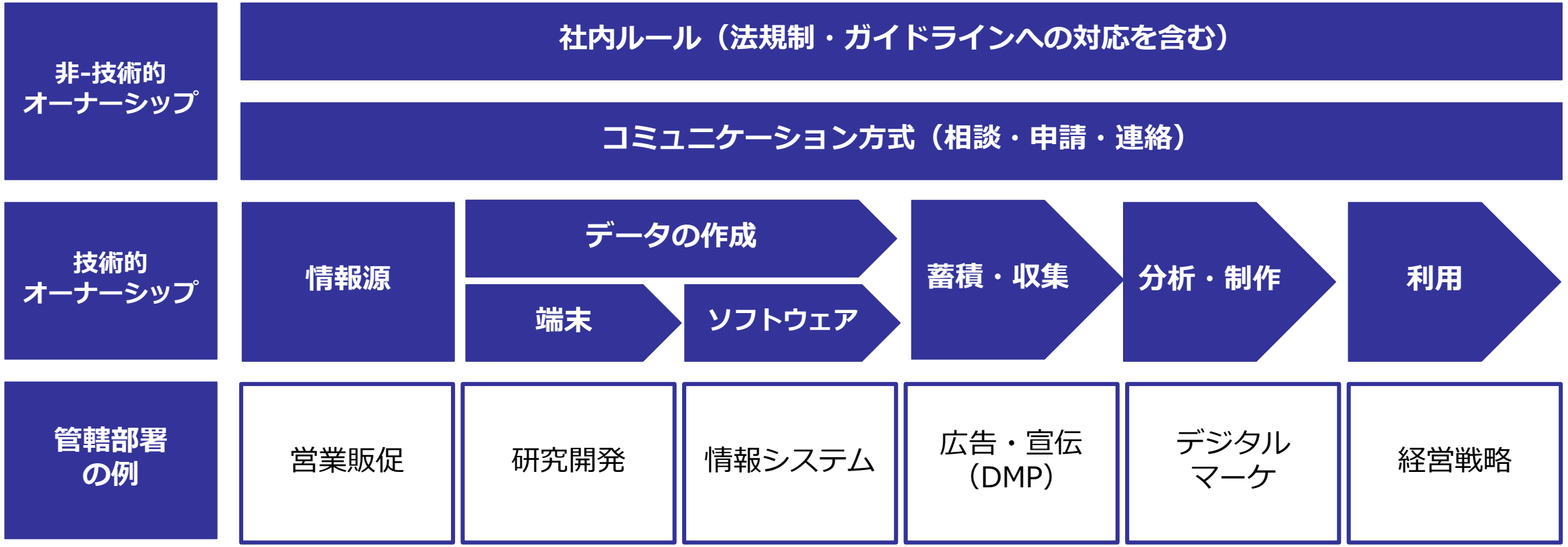
2. 当事者であること（責任の所在）

- 本分科会で挙がった論点では、この意味でのオーナーシップが問題
  - 2-1.組織内のニーズ、利用状況などの把握と対応
  - 2-2.誰がデータ（項目やコード）を定義・管理するのか

Copyright (C) 2019 JUAS All rights reserved

5　ヒアリング結果の深掘り
5－3　データ流通　[4/10]　～①-3社内で誰がデータオーナーシップを持つべきかを決定する～

■ オーナーシップの配置例を以下に示す。
■ 部署間で連携し、社内でデータ流通させる場合、誰が責任者となるべきかは企業ごとに異なる。



出典：CIGI「Data Ownership」図3をもとに和訳し、簡略化して加筆
https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－3　データ流通［5/10］～①-4よくある懸念・課題～

■ オーナーシップが曖昧だと、所有者／当事者の様々な懸念が重なり、業務上の課題につながりやすい。

1.「所有者」の懸念	2.「当事者」の懸念	業務上の課題
社外のデータを使いたいとのニーズがあるがどうしていいかわからない	データを公開することに抵抗がある。	欲しいデータを、ビジネスが求めるスピードで探して利用することができない
社内データを外部から欲しい（購入したい）と言われた場合、即時に対応できない	各アプリでどのようなデータを所有しているかわからない	［…］関係者との調整や構築時に多大なコストがかかる
社外データを購入・所有している部署がわからない	他部門のマスターデータを使いたいが使えない	サブシステムごとにマスタデータを管理しており、同一情報であっても別々に管理している場合がある
社外から同じようなデータを購入しており、２重、３重投資となっている	ユーザがデータ項目を確認する手段がない。想定と異なるデータが格納されている	他部署のデータを使いたいがシステム的な影響が大きいとの理由で断られてしまった

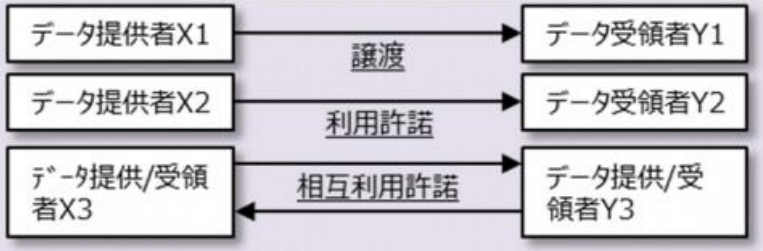
Copyright (C) 2019 JUAS All rights reserved

5　ヒアリング結果の深掘り
5－3　データ流通［6/10］～②類型ごとの契約形態～

データ提供型	一方当事者から他方当事者へ、すでに保持しているデータを提供する
データ創出型	まだ存在していない、新たなデータの創出を、複数の当事者間で行う
データ共用型　（プラットフォーム型）	データの集約・保管先を指定することにより、データを共同利用する

契約類型を３つに整理し、それぞれ、構造・法的性質、課題、法的論点、適切な取決め方法、モデル契約書案等を整理

「データ提供型」契約

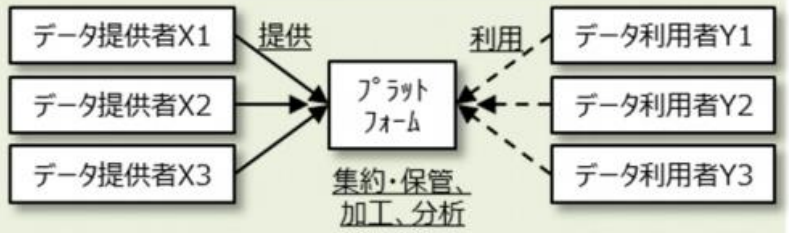


・データ提供者から他方当事者に対してデータを提供する際に、他方当事者のデータ利用権限等を取り決める契約(対象データをデータ提供者のみが保持しているという事実状態が明確である場合)

「データ創出型」契約

・複数当事者が関与することによりデータが新たに創出される場面において、データ創出に関与した当事者間で、当該データの利用権限を取り決める契約

「データ共用型」契約



・プラットフォームを利用したデータの共用を目的とする類型の契約

出典：経済産業省

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－3　データ流通［7/10］～③よくある懸念／課題も（抽象的ではあるが）洗い出されている～

■ 想定される懸念／課題についても、自社の事情に置き換えて読み込めば、参考にしやすい

懸念	課題	類似する社内的課題
データ取引を仲介するプラットフォームの管理・運営は誰が担うべきか？	中立性の確保、安定的な運営のための収入確保、セキュリティや透明性確保等の義務・責務、等	• どの部署がやるのか？ • どこの予算を使うのか？ • どの情報システムを導入するのか？
競合他社で同じプラットフォームを利用すると、独占禁止法に抵触しないか？	公取委とも協議の上、独禁法で問題となり得ると考えられる要素を整理（価格情報の共有、一方的なデータの提供規約等）	• 一方的に負担を押し付けていないか？ • グループ間取引における利益供与に当たらないか？
データの提供先や利用目的を限定したいが、どのようにすればいいか	データの性質や種類（秘匿性、提供頻度等）に応じて、利用範囲や制約条件を取り決める方法を複数パターン提示	• 営業秘密に当たらないか？ • プライバシーポリシーに違反する利用目的ではないか？（例：内定辞退率情報）
複数の事業者間で大小様々なデータのやり取りが行われるのに、どのような契約や規約で管理すればいいのか？	プラットフォーム型の特徴を踏まえ、個々の事業者間での契約ではなく、参加者全員に適用される統一的な利用規約を設けて全体を規律する考え方を提示	• 外注先／委託先にデータを開示していいのか？ • 海外拠点／地方拠点にデータを開示していいのか？

Copyright (C) 2019 JUAS All rights reserved

5　ヒアリング結果の深掘り
5－3　データ流通［8/10］～【参考】データオーナー、データ流通の参考文献～

ドキュメント	内容（エッセンス）
Data Ownership	・カナダの非営利シンクタンクCIGIに所属するTeresa Scassaが公表したレポート。 ・ヨーロッパでは、データに関する独自の所有権（オーナーシップ）を創設すべきとの議論が高まっていて、EU GDPRでは、データ主体（個人）はデータポータビリティ権を持つと規定された ・データの権利者（オーナー）の保護と、公益のためのデータ取得や再利用とのバランスをとる方法が必要
企業において発生するデータの管理と活用に関する研究（アンケート、事例検討）	・企業向けアンケートからデータ利活用を積極的に行っている企業は比較的多いものの、利活用することを期待しているがまだ実際に利活用できていないデータも相当量存在する ・本研究の分析結果ではデータ利活用においてデータ発生源への便益の提供が重要であることが示されたように、他組織や個人といったデータ発生源への配慮も必要
企業において発生するデータの管理と活用に関する研究（契約方法）	・データ利活用に関わる契約を類型化し、契約方法を提示。（「1. サービス利用規約等」、「2. 情報解析等業務委託契約」、「3. 業務提携契約」「4. 業務提携契約」） ・データ利活用に関わる契約には専門の人財が必要であり、育成の必要についても言及。
産業保安版のデータ契約ガイドライン及びセキュリティマニュアル	・産業用（プラント等）のデータについて、企業間をまたがって利活用する場合の契約ガイドライン (1) データ・オーナーシップの考え方 (2) データの保護に関する知的財産権等の法制度 (3) データの不正利用を防止する措置 (4) プラントデータの管理方法・開示方法に関する義務 (5) プラントデータを活用した結果得られる権利

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－3　データ流通　[9/10]　～【参考】参考文献URL～

政府CIOポータル　標準ガイドライン群
https://cio.go.jp/guides

CIGI「Data Ownership」
https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf

独立行政法人経済産業研究所「企業において発生するデータの管理と活用に関する研究」
https://www.rieti.go.jp/jp/publications/summary/18090010.html

経済産業省「データの利用に関する契約ガイドライン 産業保安版」
https://www.meti.go.jp/policy/safety_security/industrial_safety/oshirase/2019/4/20190425.html

Copyright (C) 2019 JUAS All rights reserved

5－4　教育・育成［1/4］

■あるべき姿（目指す姿）

データドリブン経営（データ利活用、…etc)を推進する営みが、社内の関係者に浸透・定着する。

■ステップ

あるべき姿に到達するためのステップ

- データ利活用の効果・価値を理解してもらう。
- データ利活用を、自分事として認知してもらう。

⇒ 研修だけでは浸透・定着しない

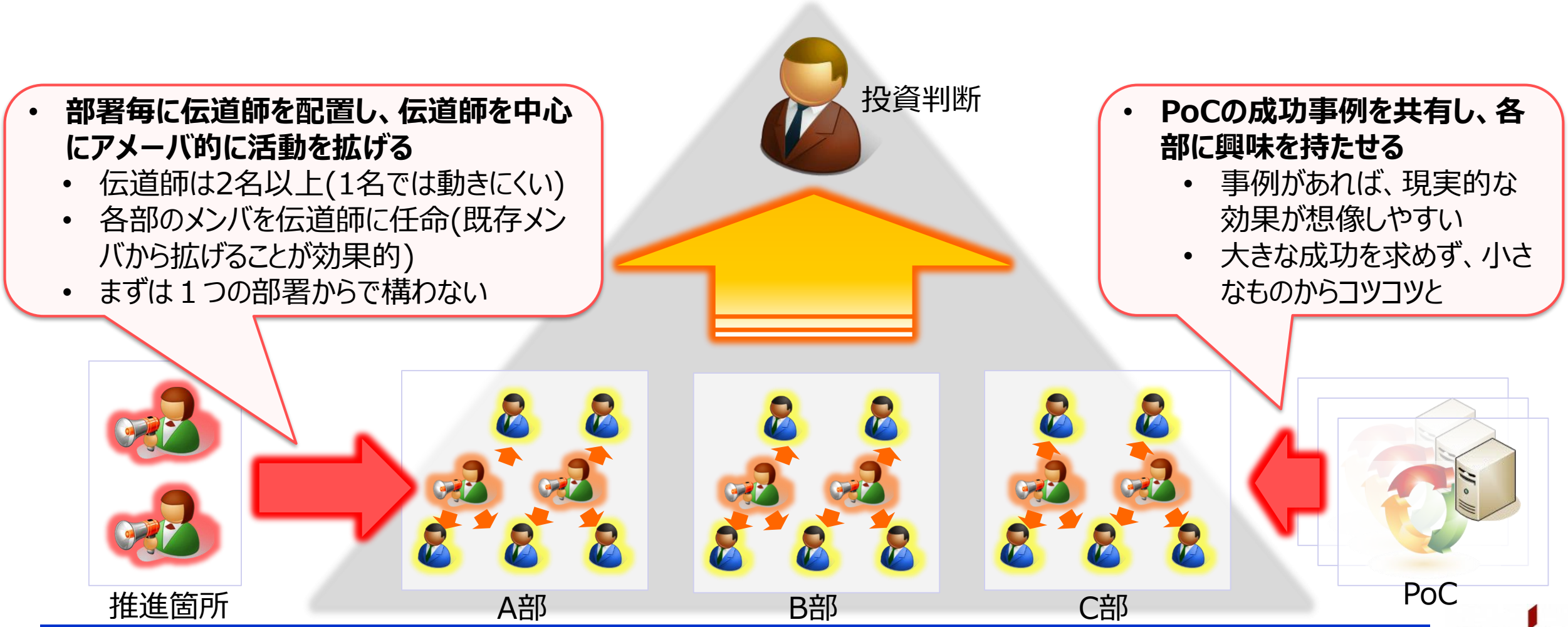
■手法

- 業務課題に対してPoCを回し、データ利活用での成功事例を積み重ねてこれを共有する。
- 各部に伝道師を配置して、伝道師からメンバーへ、知識・ノウハウの展開を図る。
- データ利活用を推進する専属体制を構築し、強力なリーダーシップで推進する。

Copyright (C) 2019 JUAS All rights reserved

5　ヒアリング結果の深掘り
5－4　教育・育成　[2/4]

■ 現場が熱を帯びれば経営層の判断材料が増えて、大きな投資判断も可能になる。
[例]専属体制構築、トップダウンの施策



Copyright (C) 2019　JUAS All rights reserved

5－5　システム環境［1/4］～データ分析基盤の整備～

■あるべき姿

データの活用を促進するデータ分析基盤（プラットフォーム）を整備する

■ステップ

1　自社の業務を踏まえ、データ分析の利用者や利用目的を定義する
2　定義に応じた分析環境を構築する

■手法

①　分析基盤として必要なセキュリティ方針やデータマネジメント等を定める
②　データ分析の利用者や利用目的を定義する
③　利用者／利用目的に応じた分析データ蓄積用のデータベースやデータウェアハウス、利用するデータ分析ツール（ＢＩツール）を選定し、構築する

Copyright (C) 2019 JUAS All rights reserved

5－5　システム環境［2/4］～データ分析基盤を構築する際に定めるべき事項～

■例えば以下のような事項を定める必要がある

- データ分析部門の権限（何をどこまで決めさせるか）
- 利用者、データ内容等に応じたセキュリティ方針（アクセス制御、物理制御等）
- データマネジメント方針
- クラウド等の外部サービスの利用方針
- 分析基盤として必要な信頼性、可用性、拡張性

Copyright (C) 2019 JUAS All rights reserved

5－5　システム環境［3/4］～利用者／利用目的に応じたデータ分析基盤の整備～

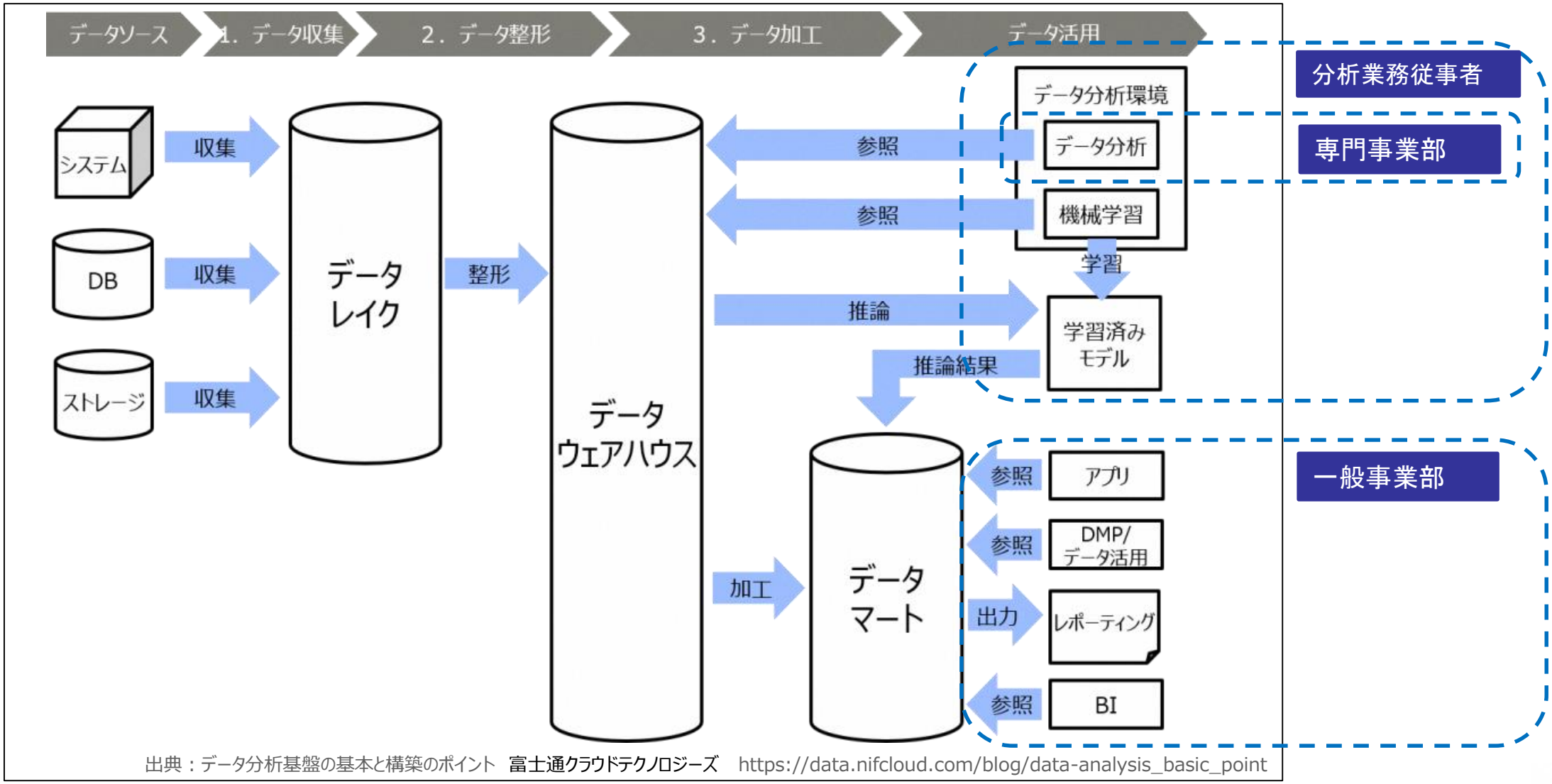
■例えば以下のとおり利用者・利用目的に応じてデータ分析基盤を整備する

分析業務従事者	● 利用者はデータサイエンティスト等、分析力が高いメンバー ● 分析アプローチは「探索型」の比重が高い ● ハイスペックな機器や分析専門言語が用いられる
専門事業部	● 利用者は特定部門でビジネススキルの高いメンバー、分析力も一定程度有す ● 分析アプローチは「仮説検証型」の比重が高い ● 機器にはそれほど高いスペックは求められない、ＢＩツールが中心
一般事業部	● 利用者は分析業務に携わることのないメンバー ● 基本的に分析結果のレポーティング資料を確認するのみ ● 分析環境は不要だが、ＢＩツールによるレポーティングやダッシュボードが必要

Copyright (C) 2019　JUAS All rights reserved

5　ヒアリング結果の深掘り
5－5　システム環境［4/4］～データ分析基盤イメージ～

■利用者、利用目的に応じたデータ分析基盤

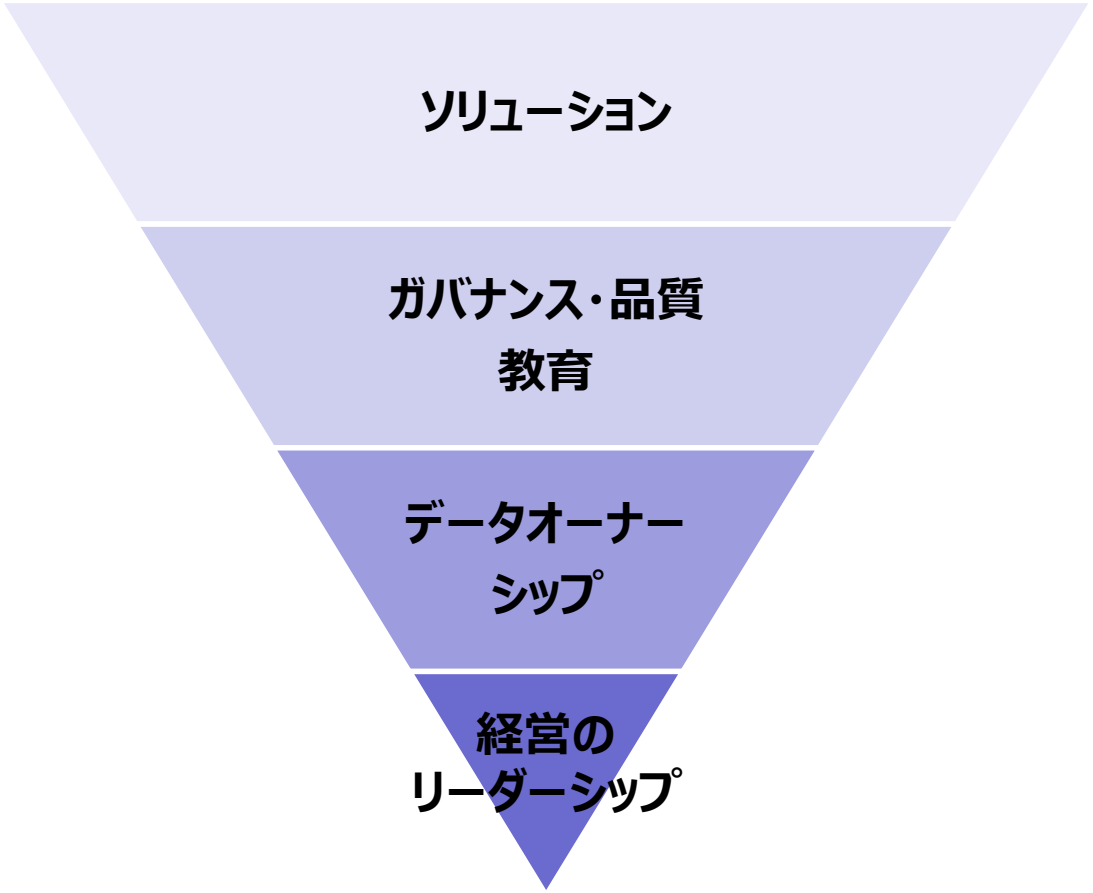


出典：データ分析基盤の基本と構築のポイント　富士通クラウドテクノロジーズ　https://data.nifcloud.com/blog/data-analysis_basic_point

Copyright (C) 2019　JUAS All rights reserved

6　考察・まとめ
6－1　考察内容・提言

■ 今回モデルとしたA社のようなデータ連携（利活用）を推進するためには、強力な経営層のリーダーシップを基礎に、経営意思に基づくデータのオーナーシップ明確化、ガバナンスや品質の担保、高レベルでデータ活用できる人財育成のための仕組みの整備、それを実現するためのソリューションの導入が必要となる。



■ **ソリューション**
データ利活用における課題を認識し、それを解決するための仕組みやツールを利用する。それによって得られた効果は測定し、継続的に改善、より透明性、精度の高いデータ利用と組織内での啓発に努める。

■ **ガバナンス・品質、教育**
データの不正利用をなくし、一定水準の信用性が担保された状態を構築する。教育プログラムを支援し、データリテラシーの向上や、データ利活用を現場レベルで活性化させ、データドリブンな組織文化を醸成する。

■ **データオーナーシップ**
法的要件に注意した上で、取り扱うデータの所掌(データの所在や管理先、利用方法など)を明確にする。

■ **経営のリーダーシップ**
データ利活用方法を理解した上で、企業のありたい姿を定義し、戦略を立案、社員に対して熱意をもって発信し続ける。現場から得られたデータやその分析結果をもとに、速やかに意思決定を行う。

Copyright (C) 2019 JUAS All rights reserved

６　考察・まとめ
６－２　今後の活動に向けて

■ 今年度活動の振り返りでメンバーから出された意見は、今後の活動の材料としたい。

- ■ データ流通でデータ価格の標準・指標がない。これを見える化したい。
  ⇒時代の変化に伴い、業界毎の価格標準体系が少しずつ出てきている
  ⇒データ流通のオープン化の動きも注視する必要がある

- ■ 今回はあまり踏み込まなかったシステム技術面の研究もしていきたい。
  ⇒大量データの連携ネットワーク、セキュリティ、他

- ■ 他社ヒアリングは、実経験に触れることができて非常に有益であった 。スタートアップ企業や外国・外資企業など、今後も様々な企業のヒアリングを実施していきたい。

Copyright (C) 2019 JUAS All rights reserved

2－②日本型データドリブンデータ連携

Copyright (C) 2019 JUAS All rights reserved

54

目次

Ⅰ．日本型データドリブンを選んだ理由
　１．データドリブン経営っていうけどみんなの会社は出来てるの？
　２．改めてデータドリブン経営を行う必要はあるのか!?
　３．深堀のためのアプローチ
　４．これが問題？　－　仮説
　（参考）一般的なデータドリブンの定義
Ⅱ．日本でデータドリブンが進まない理由
　１．日本企業の文化的背景
　２．日本固有の要素：文化的背景
　３．日本固有の要素：人
　４．日本固有の要素：組織
　５．日本固有の要素：テクノロジー
Ⅲ．日本での好取組事例（チャレンジ）
　2社実施
Ⅳ．こうすればデータドリブンが進む
　（日本型データドリブン）
　１．人から言えること
　２．組織から言えること
　３．好取組事例から言えること
　４．総評
Ⅴ．申し送り事項・メンバーから一言
　１．申し送り事項
　２．メンバーから一言

Copyright (C) 2020　JUAS All rights reserved

Ⅰ．日本型データドリブンを選んだ理由

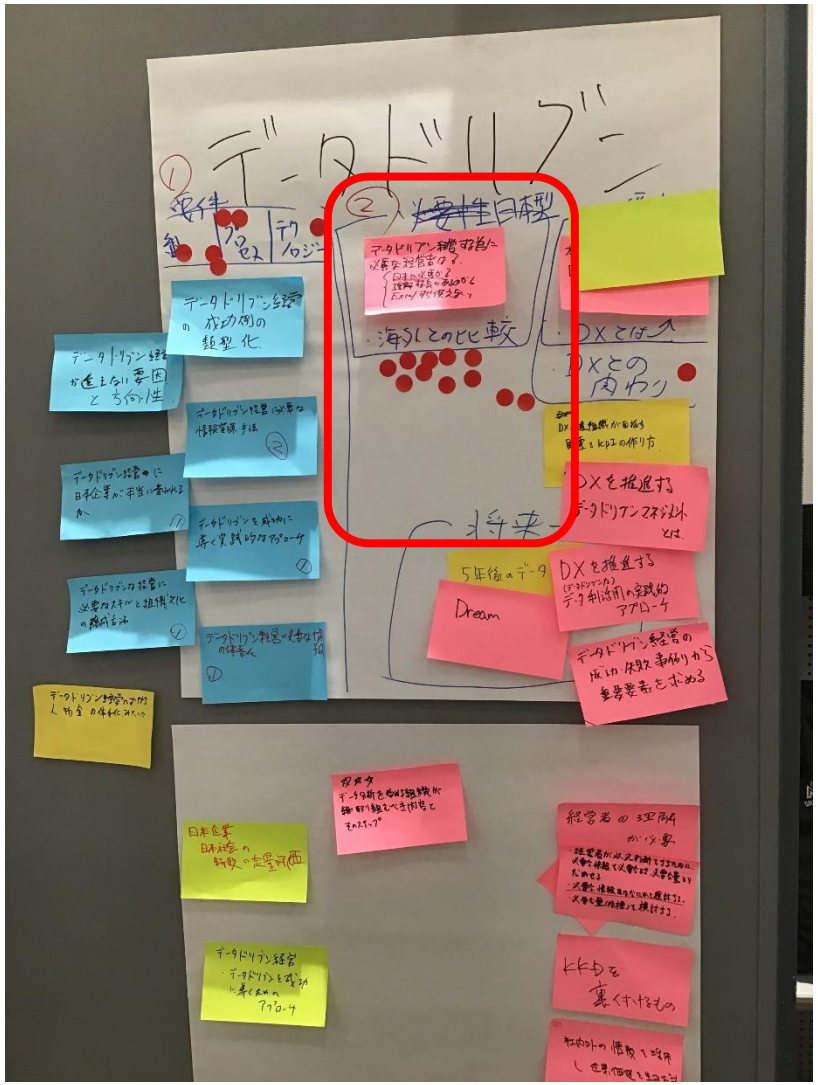
Copyright (C) 2020 JUAS All rights reserved

1．データドリブン経営っていうけどみんなの会社は出来てるの？

旧データマネージメント研究会からずーっと追いかけているデータドリブン経営。
ある程度体系的な整理は出来てきた。
しかし、いざ自社の状況を見てみると肌感覚としてなかなか出来ている感じがしないのは何故なのか？　そもそも自分の会社にデータドリブン経営が必要なのか!?

ネットや雑誌の見出しには沢山の成功事例を見つけることが出来る。
本当に日本企業で取り組みが成功して、運用できているところはあるのか？　と考えたときに出来ていない企業には何か日本固有の問題があるのではないかという疑念が湧きました。

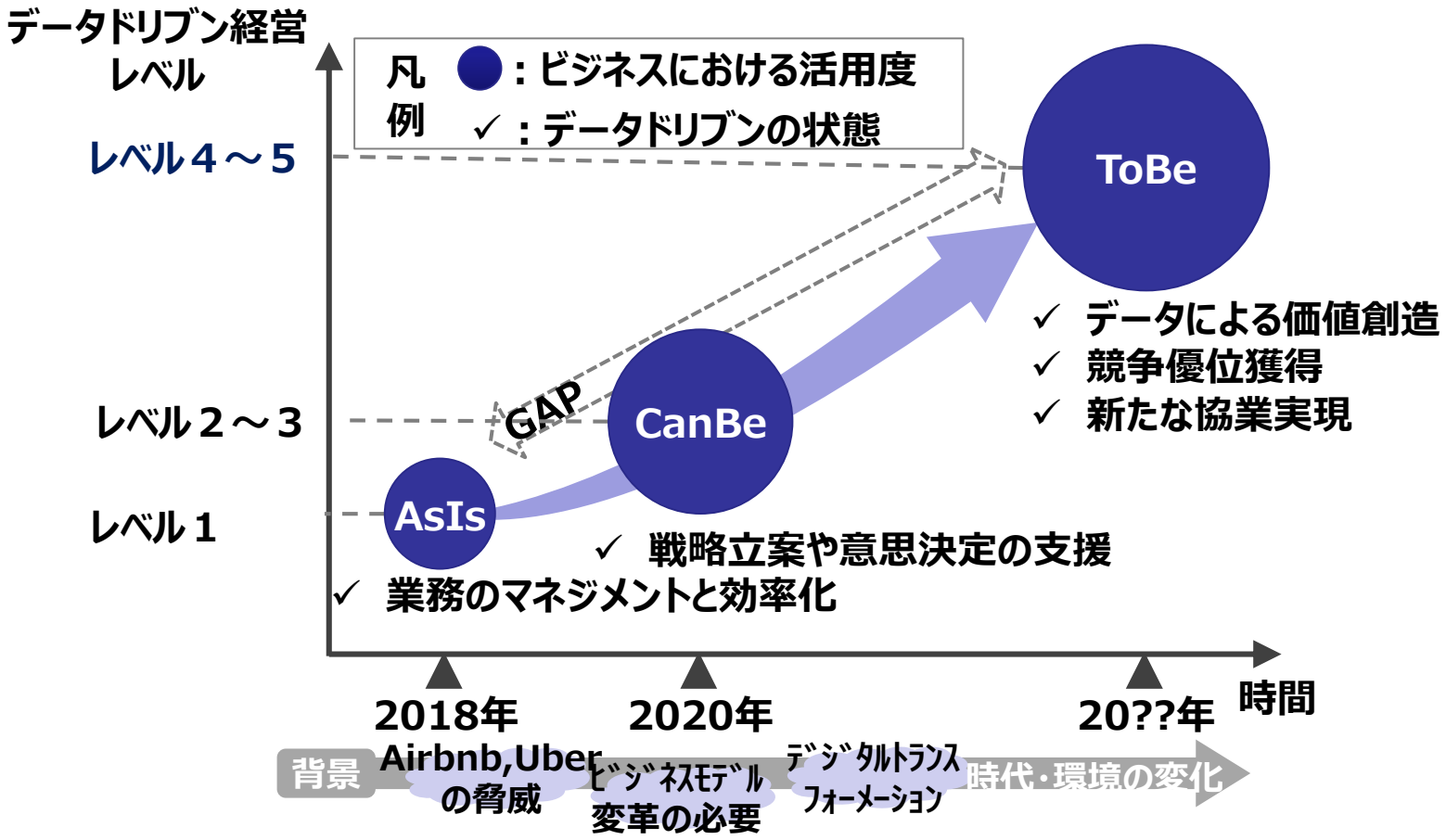
データドリブン分科会では、この疑念に対して調査を行い課題があるのであれば、これを克服した「日本型データドリブン経営」の姿を追いかけることとしました。



Copyright (C) 2020　JUAS All rights reserved

2．改めてデータドリブン経営を行う必要はあるのか!?

これまで当研究会ではデータドリブン経営について研究を重ねてきました。
少なくとも企業は第四の経営資源と言われる情報を利活用して成長し続けるためには、「データドリブン経営」が重要であるというは明白です。



Copyright (C) 2020 JUAS All rights reserved

3．深堀のためのアプローチ

出来ていない企業には何か日本固有の問題があるのではないかという疑念に対して、では成功している欧米と比較を行うことで、その違いや課題が見えてくるのではないかと考えました。

そこで、 3つの観点で比較をしてみる事としました。

1．人と組織
2．業務プロセス
3．テクノロジー

Copyright (C) 2020 JUAS All rights reserved

4. これが問題？ － 仮説

3つの観点で比較をする前に、そもそも日本型の経営ってなんだ？ これが欧米と大きく違っていれば、そこに問題があるかもしれないという仮説のもと、日本型経営について調べてみました。
本編ではこの仮説のもと3つの観点で比較を行い、何が「日本型データドリブン経営」のあるべき姿なのかを追っていきます。

■ 従来の日本型経営とは？

日本型経営の特徴：3点

1．終身雇用

2．年功序列

3．企業別組合

*[参考文献]『日本の経営』 著者：ジェイムズ・アベグレン

■ 従来の日本型経営の強み・弱み

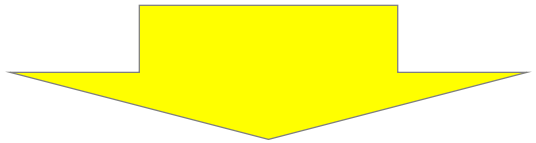
	強み	弱み	仮説
1．終身雇用	安定した労働？	変化が乏しい？	業務やお客様、提供サービスに変化がない場合には強みを発揮するが、業務に大きな変化が伴う場合には対応できないのではないか。
2．年功序列	業務と共に成長？	業務以外の成長が困難？	
3．企業別組合	労働者を守る？	新たな挑戦の弊害？	

Copyright (C) 2020 JUAS All rights reserved

参考：一般的なデータドリブンの定義

？？データドリブンとは？？

・直訳すると「駆動されたデータ」という意味。
　⇒「データが活かされた状態にある」というニュアンス。つまり・・・
・「**徹底的にデータに基づいて行動しよう**」ということ。



従来のような根拠に欠ける**投機的な行動・意思決定、直観**といった主観を排除し**客観的なデータによって導かれたものを徹底的に支持**する。

Copyright (C) 2020 JUAS All rights reserved

Ⅱ．日本でデータドリブンが進まない理由

Copyright (C) 2020 JUAS All rights reserved

1．日本企業の文化的背景（1/2）

日本企業は独特の文化・慣例に則り、経営されてきた。欧米型経営との比較をしながら、日本型経営の特徴について触れたいと思う。

欧米型経営

目標に対して直進的、他人との関係においては自己中心で弱者切り捨て、目的達成手段は実力行使かつ指揮統制重視である。
保守主義であり、独裁に走りかねない。自分にとって扱いやすい平均化された人物を好み、有能であっても自分と合わない人物は排除する。

トップダウン型 ＝ 父性文化

日本型経営

目標には進化と循環で臨み、他者とは調和・協調・対等・寛容・対話とボトムアップで目標達成を目指す。
革新主義であり、民主主義である。

ボトムアップ型 ＝ 母性文化

Copyright (C) 2020 JUAS All rights reserved

1．日本企業の文化的背景（2/2）

	父性文化	欧米型経営	母性文化	日本型経営
他者との関係	自己中心	個人主義=個の論理(個を優先する)	連帯	集団主義=場の論理(世間の空気を読む)
	弱者切り捨て	フェアプレイ(勝者が全てを得る)	弱者への配慮	フェアシェア(公平な分配)
	競争	一神教は神を唯一絶対とするため他の宗教を認めることができず、対立や衝突を招きやすい。	調和	多神教(八百万の神)は足し算の文化。共生・共存という概念と親和性が高い。
	対立		協調	
	排他性		開放性	
	ヒエラルキー		対等	
	厳格	法令順守。明確な責任範囲。	寛容	複合的な判断基準であり、時には人間関係を軸とした判断がなされる。
自己実現との関係	知性重視	形式知(知識・スキル・ノウハウ)	感性とのバランス	形式知＋暗黙知(人物的要素)
	強欲	富の分配は、トップが巨万の富を得る。	小欲知足	富の分配は「公平な分配」がベース。巨万の富を得るトップは少ない。
目的達成手段	実力行使	知識・ルール・ノウハウ重視(効率に重きを置いた労働姿勢)	対話	人間関係を重視した労働姿勢
	トップダウン	ウォーターフォール型組織	ボトムアップ	アジャイルスクラム的組織
	指揮統制		自発性	

日本は母性文化を生かした三種の神器（企業別組合・終身雇用・年功序列）で成功してきた。

Copyright (C) 2020 JUAS All rights reserved

2．日本固有の要素：文化的背景

日本人が古来から持つ特性や精神

儒教(「五常」:仁・義・礼・智・信)

仁：思いやり
義：利欲なしにやるべきことをやる
礼：「仁義」を行動化し、礼儀を重んじる
智：学問に励む
信：誠実で正直、約束を守る

年長者が得た経験を真摯に受け入れ、学び、受け継いでいくという姿勢。上下関係に縛られ、自己実現ができない傾向

敬語

儒教の「礼」や、和の心からつながる言語体系。
人を軽視せず、敬い、自らを譲る、秩序と調和の言葉。
現代では良好な関係・コミュニケーションを
構築・維持するために必要な要素。

ただし、敬語による関係性の創出によって下の意見が通りにくくなる場面が多々ある。通りやすくするためには上司の理解・寛容性が重要になり、上司の知識が重要となる

島国

「郷に入っては郷に従え」
「島国根性」

上記言葉に代表されるように、それぞれの土地・文化・規律・習慣に合わせ、それ以外の文化や考え方を排斥する思想が強い。

禅(ZEN)

・言葉では言い表せない、体験的に得る何か
・悟ること
言葉や文字で伝えられる以外の仏陀の心を
修行などによって感じるが禅の始まり。

言葉ではなく、相手の気持ちや自らの経験からものごとを察し、態度や表情で自己表現をすることを好む

Copyright (C) 2020 JUAS All rights reserved

3．日本固有の要素：人

**前項の日本人固有の特性や精神をデータドリブンに照らし合わせると、
データドリブンに向く面、向かない面の双方が見えてきた。**

データドリブンに関係する日本人の特性

- 代々受け継がれてきたものを守り抜くことは得意だが、
  データドリブンのような今までなかった手段を柔軟に取り入れることが苦手
- 相手の気持ちを察したり、経験や勘所を元に判断することに長けているが、
  ファクトや言語による意思表示は苦手

日本人がデータドリブンを成功させるには？

データドリブンな考え方を上の立場の人が知り、
今までの経験や勘所とファクトであるデータを組み合わせ、バランスを取りながら
成功事例を積み上げていく。
また、それを次の世代へ受け継いでいく。

Copyright (C) 2020 JUAS All rights reserved

4．日本固有の要素：組織

日本でデータドリブンが進まない理由として、日本固有の組織にも課題があると考える。

■ **課題①　ボトムアップ型の組織運営**

担当者の業務範囲で物事を考えるため、自組織の改善は得意だが部門をまたがる改革が苦手である。
個別最適に陥り全体最適が進まない。

■ **課題②　企業別労働組合の存在**

弱者への配慮が行われることにより、新たな挑戦が阻害される。（強者が走りすぎては弱者がついて来られない）

■ **課題③　年功序列制度**

能力を発揮しても役職が上がっていかないため、若者の労働意欲が低下する。事なかれ主義に陥りやすい。

■ **課題④　終身雇用制度**

生き生きと長く働くために調和を重視しリスクを冒さない働き方になる。

■ **課題⑤　組織横断でデータドリブンを推進する体制の構築が進みにくい**

上記①～④の課題とも関係するが、牽引者がいない、本業がある、現場が必要性を感じていない、などの理由で
推進体制の構築が進まない。

Copyright (C) 2020　JUAS All rights reserved

5．日本固有の要素：テクノロジー（1/2）

- テクノロジーにおいて、GAFA(米国)やBAT(中国)のメガプラットフォーマーを有さない国では固有要素は見受けられない。
- 欧州では独自のプラットフォームを構築するProject GAIA-Xが発表されており注視したい。

GAFAからデータ主権を取り戻す欧州の独自のプラットフォーム戦略　Project GAIA-X（ドイツ）

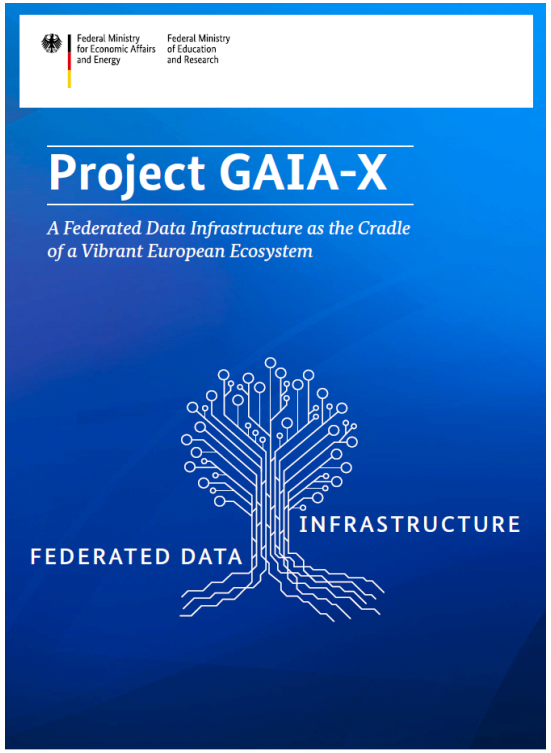
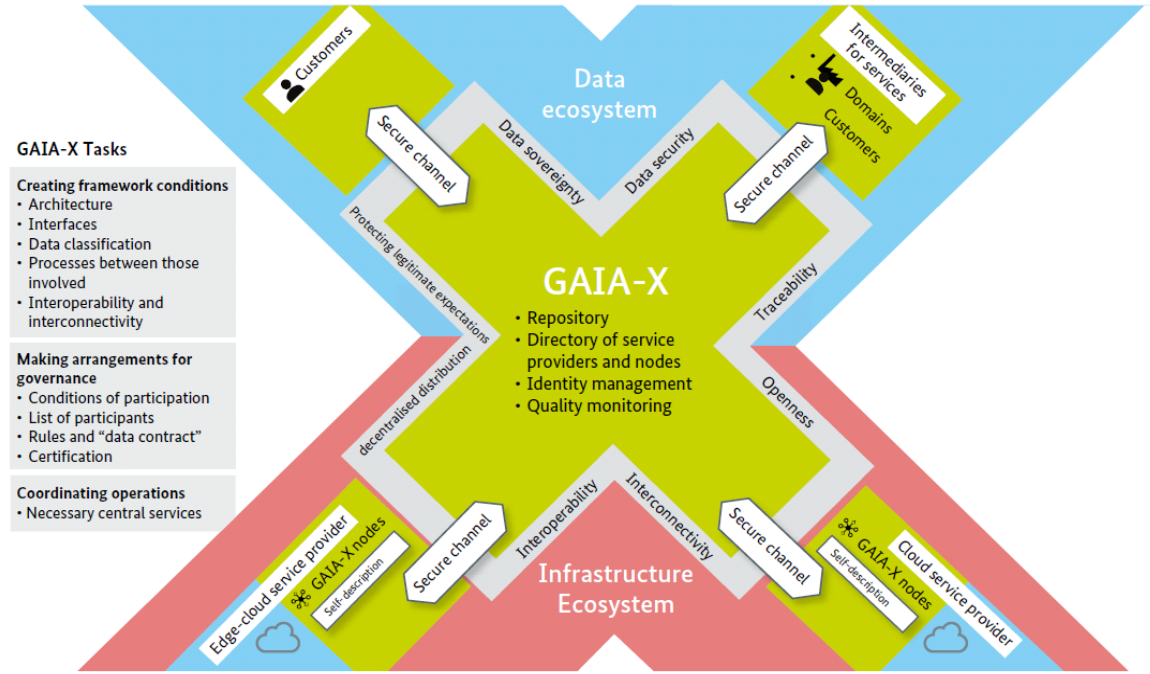


Figure 1: Overall picture of data infrastructure and ecosystem



Source: BMWi

Copyright (C) 2020 JUAS All rights reserved

5．日本固有の要素：テクノロジー（2/2） プラットフォームの考察

・世界的な第4次産業革命の流れの中、日本でも大企業を中心とする標準化コンソーシアムを通じ、規格化に携わっている。今後は日本企業の知財を守る基盤の構築がカギとなる。

第4次産業革命のデータドリブンのキーワード
- -CPS
- -スマートファクトリー
- -標準化プラットフォーム

日本の商習慣の特徴
- -元々資源がないことから、原材料を仕入れて作って売ることが主流となり、匠による高付加価値化を差別化につなげている
- -資本主義国家であるため、個社独自路線であり、業界で足並みをそろえていない

日本の第4次産業革命の迎え方
- -ノウハウ・プロセスのデータ化は進めるが、競争領域／協調領域を分ける必要があり
  ⇒知財を守るための仕組みを含むデータ連携基盤・ルールの構築

Copyright (C) 2020 JUAS All rights reserved

Ⅲ．日本での好取組事例（チャレンジ）
2社ヒアリング

＊詳細な内容は省略

Copyright (C) 2020 JUAS All rights reserved

Ⅳ. こうすればデータドリブンが進む
（日本型データドリブン）

Copyright (C) 2020 JUAS All rights reserved

1. 人から言えること

- 儒教や敬語などから成る「協調性」に優れた特性を活かし、他とのコラボレーションによる様々なナレッジを掛け合わせた課題発見につなげる
- 島国精神や終身雇用企業文化から成る「愛社精神」を活かし、会社をよくするためのモチベーションを有効な課題発見につなげる

→上記を目指すにあたり、「データドリブン推進部隊」の配置が有効なのではないか？

- データドリブン推進の責任を明確化する。日本人の特性の調和重視の裏側として、役割や責任を譲り合うことデータドリブン推進の阻害要因として考えられるが、これを除去することができる。
- 実際に事例ではデータドリブン推進する部隊の活動が有効であることがみえている。

Copyright (C) 2020 JUAS All rights reserved

2. 組織から言えること（1/4）

問題点からのアプローチ

様々な社会的環境や、日本企業固有の問題に対して、日本でのデータドリブンを加速させるには、以下のアプローチが必要であると考えている。

問題点

- 全体最適が進まない
- 新たな挑戦が阻害
- 事なかれ主義
- リスクを冒さない
- 組織横断で推進体制の構築が進まない

アプローチ

データドリブンの理解(啓蒙)

データドリブンには定量的な根拠がありビジネスに対し効果的で今後重要あることを、組織的な理解として深め、業務上のアクションに繋げていくことが重要

意思決定プロセスの迅速化

迅速な意思決定。短サイクル・高精度で分析し、評価の差分を小さくしていく事が重要

デジタルスキルのシフト

データ分析に関する、知識、スキル、ノウハウをもった人材の確保と育成が重要

Copyright (C) 2020 JUAS All rights reserved

2. 組織から言えること（2/4）

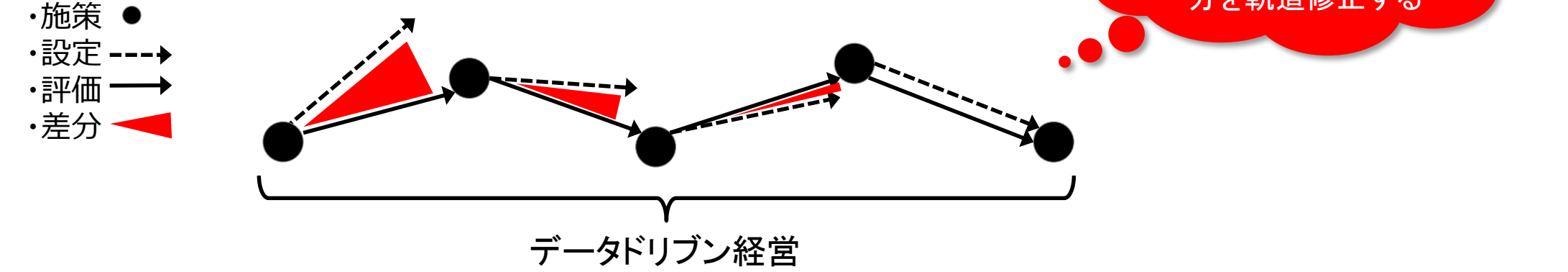
データドリブンの理解（啓蒙）

迅速かつ正確な経営判断によって新たなビジネス領域に挑戦する必要がある。そのためにはデータドリブンの手法を日本経営に取り入れて時代にあった経営を目指すことが重要である。

【データドリブン経営とは】

企業の成長に必要な施策・活動についてデータを用いて経営判断していく事。
- ・データを活用した施策／活動の設定
- ・データを活用した施策／活動の評価
- ・設定と評価の差分をデータ活用に反映

データドリブン経営のイメージ図



上記実現には、「**意思決定のプロセス迅速化**」と「**デジタルスキルへのシフト**」が重要となる。

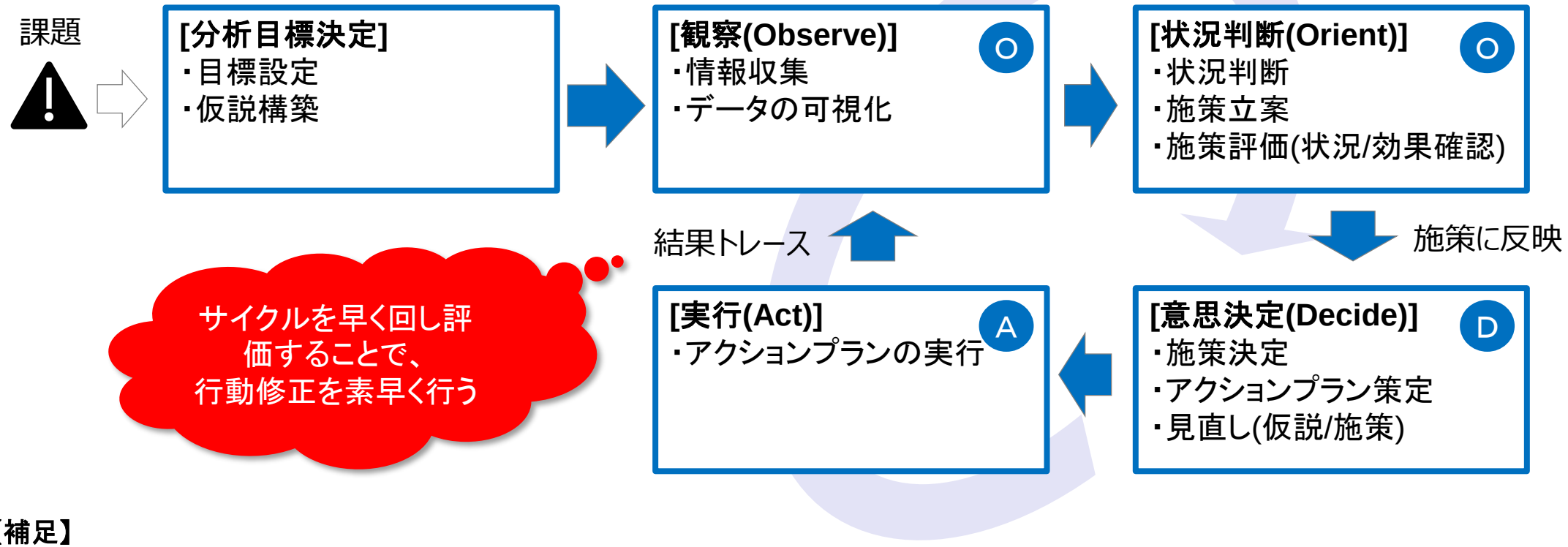
Copyright (C) 2020 JUAS All rights reserved

2. 組織から言えること（3/4）

意思決定プロセスの迅速化

意思決定のプロセスをいかにデジタルシフトさせ、判断のスピードを速めていくかが重要となる。
その為には、データ駆動型社会の実現と、データ利活用サイクルの迅速化が必要である。

【データ利活用サイクル（OODAループ）】



【補足】
PDCAとの違い：PDCAは不確実性に対する意思決定が弱い。OODAはより変化に柔軟性があり迅速な意志決定を志向している。

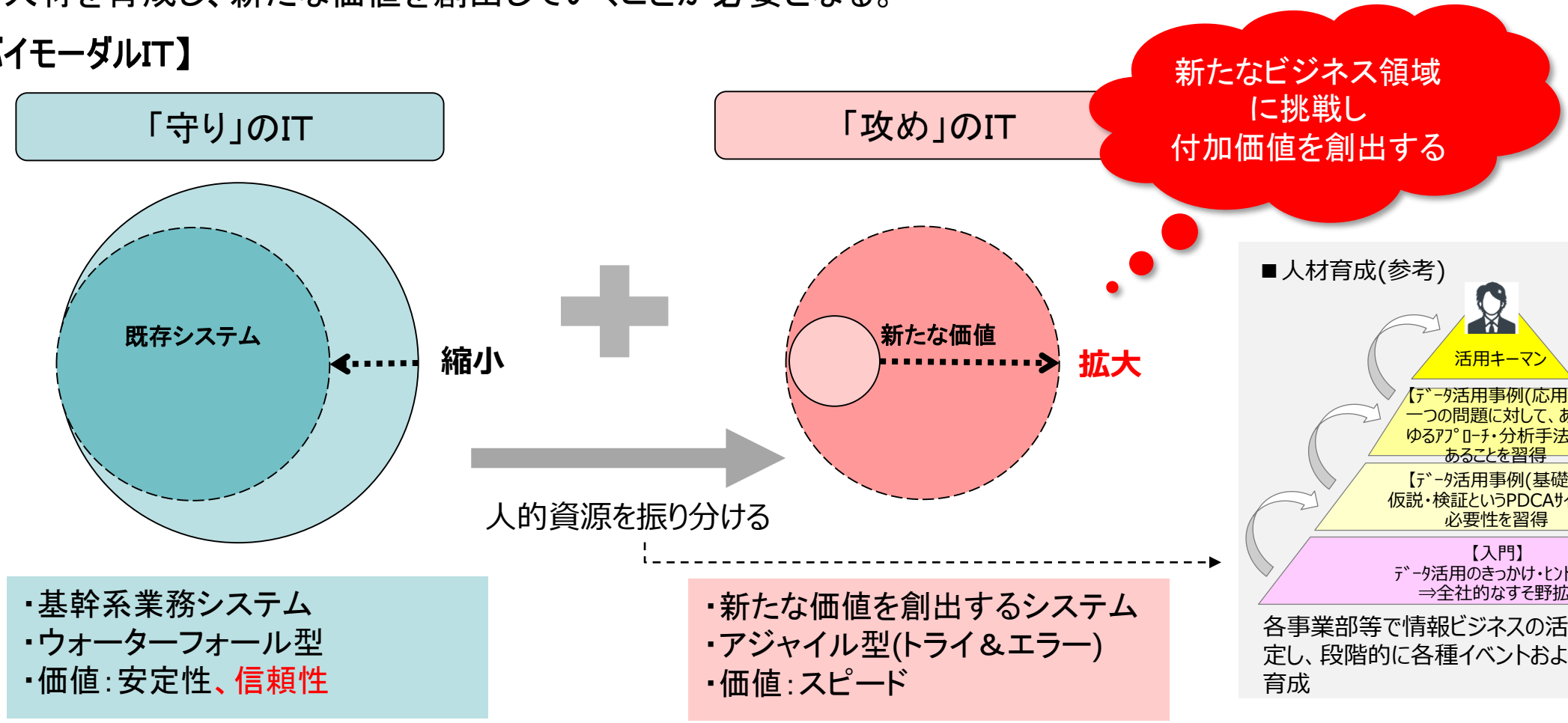
Copyright (C) 2020 JUAS All rights reserved

2. 組織から言えること（4/4）

デジタルスキルへのシフト

既存システムを中心とした「守り」のITシステムから、データ分析を含めた「攻め」のITにデジタルシフトさせ、活用人材を育成し、新たな価値を創出していくことが必要となる。

【バイモーダルIT】



■人材育成(参考)

各事業部等で情報ビジネスの活用キーマンを設定し、段階的に各種イベントおよび勉強会で適宜育成

Copyright (C) 2020 JUAS All rights reserved

3. 好取組事例から言えること

いきなりDXを絡めた大きな課題解決や価値創造に取り組むのではなく、実績を積むことで、段階的にデータドリブン経営を実現できる環境を整備していく必要がある

社員のデータリテラシーの向上

- データとは何か、ビジネスにどのように活用できるか、そのためにはどのようなデータが必要か等の基礎知識を会社全体に浸透させる。

データ活用推進の体制構築

- データ管理・分析組織だけでなく、経営や業務を巻き込んだ全社的な観点でプロジェクト化する。

現状課題や問題点を把握

- 業務やデータ活用における課題や問題点を全社的な目線で把握し、全体最適の観点でデータ活用に取り組む。

小さなデータ活用で実績を積む

- いきなり大きな課題(ＤＸを絡めて)に取り組むのではなく、身近(小さい)な課題から取り組み、社内にナレッジを蓄積・共有。

Copyright (C) 2020 JUAS All rights reserved

4. 総評

日本固有の要素（問題点）

＜人＞
- 代々受け継がれてきたものを守り抜くことは得意だが、データドリブンのような今までなかった手段を柔軟に取り入れることが苦手
- 相手の気持ちを察したり、経験や勘所を元に判断することに長けているが、ファクトや言語による意思表示は苦手

＜組織＞
- ボトムアップ型の組織運営で全体最適が進まない
- 企業別労働組合の存在が、過剰な弱者への配慮となり、新たな挑戦が阻害
- 年功序列制度により、事なかれ主義となる
- 終身雇用制度により、リスクを冒さない働き方となる
- 組織横断でデータドリブンを推進する体制の構築が進みにくい

とるべきアプローチ

「データドリブン推進部隊」の配置により、責任と役割を明確化

「意思決定プロセスの迅速化」、「デジタルスキルへのシフト」につなげるデータドリブンの理解(啓蒙) を行い、データドリブンの推進を加速

実績を積むことで、段階的にデータドリブン経営を実現できる環境を整備（データリテラシー向上、データ活用推進の体制構築、現状課題や問題点を把握、小さなデータ活用で実績）

Copyright (C) 2020 JUAS All rights reserved

Ⅴ．申し送り事項・メンバーから一言

Copyright (C) 2020 JUAS All rights reserved

1．申し送り事項

- ✓ データドリブンな状態を可視化できるようにしたい。
- ✓ 日本の文化的背景を踏まえ、DD経営実現に向けたガイドラインを作成したい。
- ✓ もう少し”経営”視点に寄った観点からデータドリブンを纏めたい。経営KPIやプロセスの変化、ITに関してはどのような経営データの活用が必要で、どうレベルを向上していくかなど。現状から「何が必要で、何を変えていく必要があるか」について今後考察を深めたいと思います。
- ✓ 日本型データドリブン経営の完成形をぜひ研究していってほしいと思います。
- ✓ もっと多くの実践事例をヒアリングしていってほしい。
- ✓ テーマは決まったものの、落とし込むまでに少し時間が掛かりましたが、予想以上に纏まった内容になりました.
  企業文化の対比による仮説検証というスタイルは面白かったので、企業文化の傾向や今回のような視点をさらに整理したチェックシートなどを作って、各企業が継続的にデータドリブンの状況を把握できる仕組みがあると面白いですね。

Copyright (C) 2020 JUAS All rights reserved

2. メンバーから一言

- ✓ データドリブン経営を実現するためには、ビジネスやデータに関するスキルセットを明確化し、必要なスキル人材を集めれば実現できるものだと思い込んでいましたが、研究を進めていく中で、スキル不足よりも根深い日本特有の文化が足枷となっている可能性に直目して、そこを深堀りできたのは、次につながる良い研究だったと思います。

- ✓ これまでデータドリブンを成功させるためには〇〇が重要、▲△へ取り組むべき、というようなことを多く聞いていた中で、どうにも腑に落ちずにいましたが、今回のテーマ決めにて「日本型」という言葉が出てきた時、とてもしっくり来たと感じ、研究会を通じてとても興味深く有意義な議論ができました。今回の研究内容を念頭に置き、他社事例を日本型または自社型へアレンジし、自社内で良い取り組みができるよう活かしていきたい。

- ✓ 「データドリブン」と一口にいっても、業界、立場によって観点やレベル感に差があると感じました。ただ検討が進むにつれ”日本型”をキーワードに意識を共有でき、なんとか形にできたと思います。多様なメンバと一緒にディスカッションできたことは発見も多くあり有意義でした。

- ✓ 日本企業はデータドリブンに乗り遅れているようなネガティブなイメージでしたが、日本企業の特性を活かした推進や、実際の事例など生で触れ、ポジティブに考えられたのがすごくよかったです。「日本人」ということに向き合い、特性や文化を学ぶのは、中学生以来でとても新鮮でした。

- ✓ なぜ海外はデータドリブン型経営の成功例が多いのに、日本ではあまり聞かないのだろうか？という疑問から、日本型データドリブンを研究するという事に興味を持ちました。一方、データドリブン経営と聞いて、自分自身が本質をとらえて理解できているのか疑問でもあり、研究と同時に自分自身の勉強としてとらえて研究会に参加しました。1年を通してメンバと議論を重ね、調べていくうちに、日本型データドリブンの入り口は何とか見えてきたかな？と感じました。今後も継続して参加できればより理解度が増すと思うので、次年度もぜひ参加したいと思います。

- ✓ 「データドリブン」というざっくりな言葉だけ認識しているところからのスタートだったが、皆と議論し具体的な事例に触れる中で自身の考えがまとまっていった。データ活用には課題解決型による身近な問題解決と、価値創造型となる現状では想像もつかない新たなビジネス創出の要素がある。いずれにしてもデータ活用の実践が必要不可欠であり、その手段として「データドリブン」は協力なツールとなる。

- ✓ ここ数年はDXの流れの中でバスワードを拾いながら、あるべき姿を追いかけてきましたが、今回は日本企業と欧米企業の文化を比較するという大胆なアプローチはとても意義があったと思います．　またあるべき論を世の中の流れから提示されるより仮説検証、インタビューという形で整理することですんなりと提言の内容が入ってきた感じがしました．

Copyright (C) 2020 JUAS All rights reserved

2－③データサイエンス

Copyright (C) 2020 JUAS All rights reserved

目次

1　分科会テーマの決定

2　ビジネス課題・実践のための知識
　①典型的な課題
　②データ分析手順と分担の事例
　③プレパレーション
　④BIツール
　⑤機械学習ツール
　⑥ベイズ統計
　⑦Kaggleとは
　⑧Signateとは

3　テーマ
　①Predict Future Sales
　②House Prices
　③IEEE-CIS Fraud Detection

4　「機械学習モデル解釈」 ハンズオン

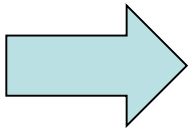
5　考察・まとめ

Copyright (C) 2019 JUAS All rights reserved

1.分科会テーマの決定

- 取り組む課題を決める。
  - 勉強フェーズ
    ・組織・スキルセット・人の育成
    ・分析手法の比較
    ・導入
  - 実践フェーズ
    ・kaggle
    ・SIGNATE
    ・BIツール

手を動かしたい ⇒ 実践フェーズを選択



- テーマの選定
  何をやるか。分析対象をどこから探すか。
  社内データは持ち出せない。
  オープンデータは分析に利用できるまでの工程が多い

kaggleのコンペを利用
分析データが用意されている。
コンペ参加者の分析コードが公開されている。

  - 全員で共通のテーマに取り組んでみた。
    選定テーマ：タイタニック
    初心者でもそれなりに実践できそう
  - 今後の取り組み
    複数テーマを選定し、チームで取り組む
    kaggleからアクティブなコンペから選定。
    選定基準：参加者の多いテーマから選ぶ。
    （やりやすいものと想定）

Copyright (C) 2019 JUAS All rights reserved

1.分科会テーマの決定

Kaggleから選択した課題

- Kaggleから選択した課題
  - Predict Future Sales
  - House Pricing
  - IEEE-CIS Fraud Detection

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　①典型的な課題

- **データ活用系**
  - システムログデータの活用（セキュリティ目的、活用度評価、労務管理など）
  - データ活用PoCの環境整備（クイックに分析ロジックや製品の検証をしたい　←IT統制とトレードオフ）
  - 営業（MR）の活動データの活用
  - 通報された安全性情報（副作用、誤用など）のテキスト化、AIによる自動分類・タグ付け

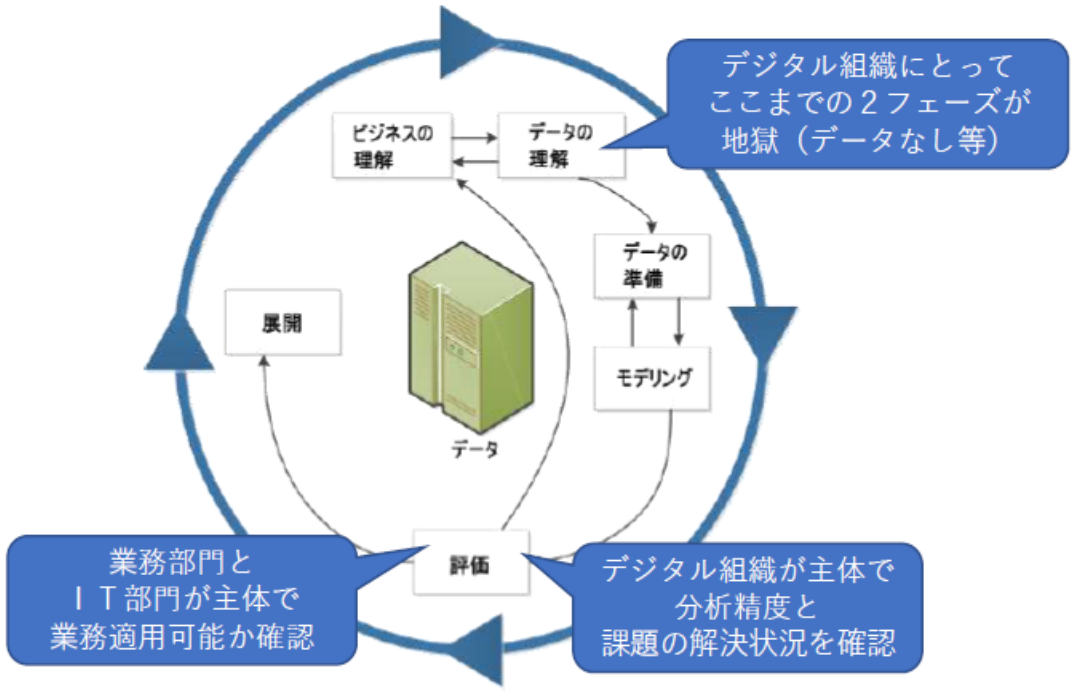
- **データ管理系**
  - 全社共通マスタデータ管理の基盤とルールの整備
  - サプライチェーン全体のデータ統合管理基盤整備

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　②データ分析手順と分担の事例

一般論は、CRISP-DMという手法が有名である。同名の団体にて定義。2001年から存在？

データ・マイニングのライフサイクル　※



- 業務部門の責任
  - ビジネス課題の種を出す
  - 分析に求める絶対基準を決める
  - 分析結果を業務に適用する
- デジタル組織の責任
  - ビジネス課題の明確化
  - データの前処理・モデリング・評価
  - 報告書(前提,分析手法,評価結果)の作成
- IT部門の責任
  - データを(主体的に)準備する
  - 分析方法を業務システムに反映
  - 反映に必要な分析知識の習得

CRISP-DM(CRoss-Industry Standard Process for Data Mining)

※参考　IBM　CRISP-DM ヘルプの概要
https://www.ibm.com/support/knowledgecenter/ja/SS3RA7_sub/modeler_crispdm_ddita/clementine/crisp_help/crisp_overview.html

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　③プレパレーション

●データ・プレパレーションとは

データプレパレーション（Data Preparation）は「データの準備・整形」を意味する言葉です。
さまざまなデータソースから取得した生データを、データプレパレーションにより分析可能な状態に適正化することで、データ準備工程を短縮し、分析者は本来のデータ分析業務に集中することができます
※NECソリューションイノベータ（データプレパレーションツール「Trifacta」）より引用

●データ・プレパレーション登場の背景

モダンで使いやすいBI(Business Intelligence)ツールが多数登場してきていることで、ビジネス部門ユーザーがセルフサービスでデータ分析を行う「セルフサービスBI」を実施できる環境が整いつつある。

しかし、データ分析のためには良質なベースとなるデータが必要であり、BIツールの性能がいくら向上しても、分析の事前準備として、データの収集/結合/変換/整形/補完などの工程が必要であることは変わらない。

ある調査によれば、「データ分析者は80%以上の時間をデータ準備に費やしている」という報告がなされており、データ準備の生産性向上が新たな課題として認識され始めている。

これからますます増大し、多様化するデータを活用する機運が高まる一方で、従来型のアプローチだけではデータを準備することが困難になりつつある。このような背景から、最新のテクノロジーを駆使したデータ・プレパレーションという新たなアプローチが誕生したのです。

※アシスト（データ活用の民主化が始まった ～データ・プレパレーションとは何か？～）から一部抜粋

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　③プレパレーション

ツール	トライアル	URL
Paxata	無償トライアル（1か月間）あり	https://www.ashisuto.co.jp/paxata/product/
Trifacta	無償版（14日間）あり	https://www.trifacta.com/ https://www.nec-solutioninnovators.co.jp/sl/trifacta/index.html?
Talend Data Preparation	無償版（14日間）あり	https://jp.talend.com/products/data-preparation/
Tableau Prep	無料トライアル（期間不明）あり	https://www.tableau.com/ja-jp/products/prep
Alteryx	無料トライアル（14日間）あり	https://classmethod.jp/partner/alteryx/

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　④ＢＩツール

BI（ビジネス・インテリジェンス）ツール

企業の各部署がそれぞれに蓄積している膨大なデータを、収集・蓄積・分析・加工し、経営戦略のための意志決定を支援するツール
形式として自社のサーバーにインストールして使用する「ソフトウェア型（オンプレミス）」と、インターネットを通じてBIツール側のサーバーにアクセスして使用する「クラウド型」がある。

BIツールで期間限定のようなトライアルではなく、無料で使い続けられるもの

	Pentaho （ペンタホ）	Qlick Sense （クリックセンス）	Microsoft Power BI	Google データポータル （旧　Google データスタジオ）
形式	ソフトウェア型	ソフトウェア型/クラウド型	クラウド型	クラウド型
ドラッグ＆ドロップによる直感的操作	×	○	○	○
ノンプログラミングでのデータ取り込み	○	○	○	○
SQLクエリによるデータ取り込み	○	○	○	○
ガイド、テンプレート機能	×	○	○	○
データマイニング機能	×	○	○	
アクセス権などの管理コンソール機能	×	○	有償なら可	有償なら可
モバイルでの閲覧	×	○	○	○
コスト	0	0 ただしクラウド版の場合は以下の機能制限あり	0	0

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　⑤機械学習ツール

勝手な分類	名前	対応範囲				
		プレパレーション	モデル作成	チューニング	説明能力	デプロイ
予測特化	DataRobot	×	○	△	△	○
	Prediction One	×	○	×	×	○
	Findability Platform	×	○	×	×	○
ワークフロー型	Orange	△	○	△	△	？
	KNIME	○	○	△	△	○
	Alteryx	○	○	△	△	○
言語	Python	○	○	○	○	○
	R	○	○	○	○	○
	MATLAB (Octave)	○	○	○	○	○

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　⑥ベイズ統計：確率の概念

- ベイズ統計では、主観確率の概念を用いる。（vs客観確率）

- 客観確率（頻度主義）の考え方

サイコロを投げ続ければ、1の目が出る割合が1/6に近づく→1の目が出る確率は1/6

- 主観確率の考え方

サイコロの1の目が、他の目より出やすい、出にくい理由となる情報がない
→1～6の目が出る確率はひとしい→1の目が出る確率は1/6

主観確率は、サイコロのようになんども再現可能でなくても定義できる概念である。

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　⑥ベイズ統計：考え方の事例

ある人が、感染症にかかっているかの検査を受けるとする。
感染症罹患有無（原因）→陽性or陰性（結果）

- 普通の考え方
- 罹患者は全人口の1%（原因の確率）
- 罹患していれば95%で陽性が出て、罹患していない場合も5%で陽性が出る。（原因→結果の確率）

- 逆転した考え方
- 陽性が出た場合、どれくらいの確率で本当に罹患しているのか。（結果→原因の確率）

→これを算出できるのが、ベイズの定理。0.01×0.95/（0.99×0.05+0.01×0.95）＝16～17%

- 意義
- 陽性の結果をしることで、罹患確率が1%から16%となった。（原因の確率の洗い替え）

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　⑥ベイズ統計：モデリングへの応用

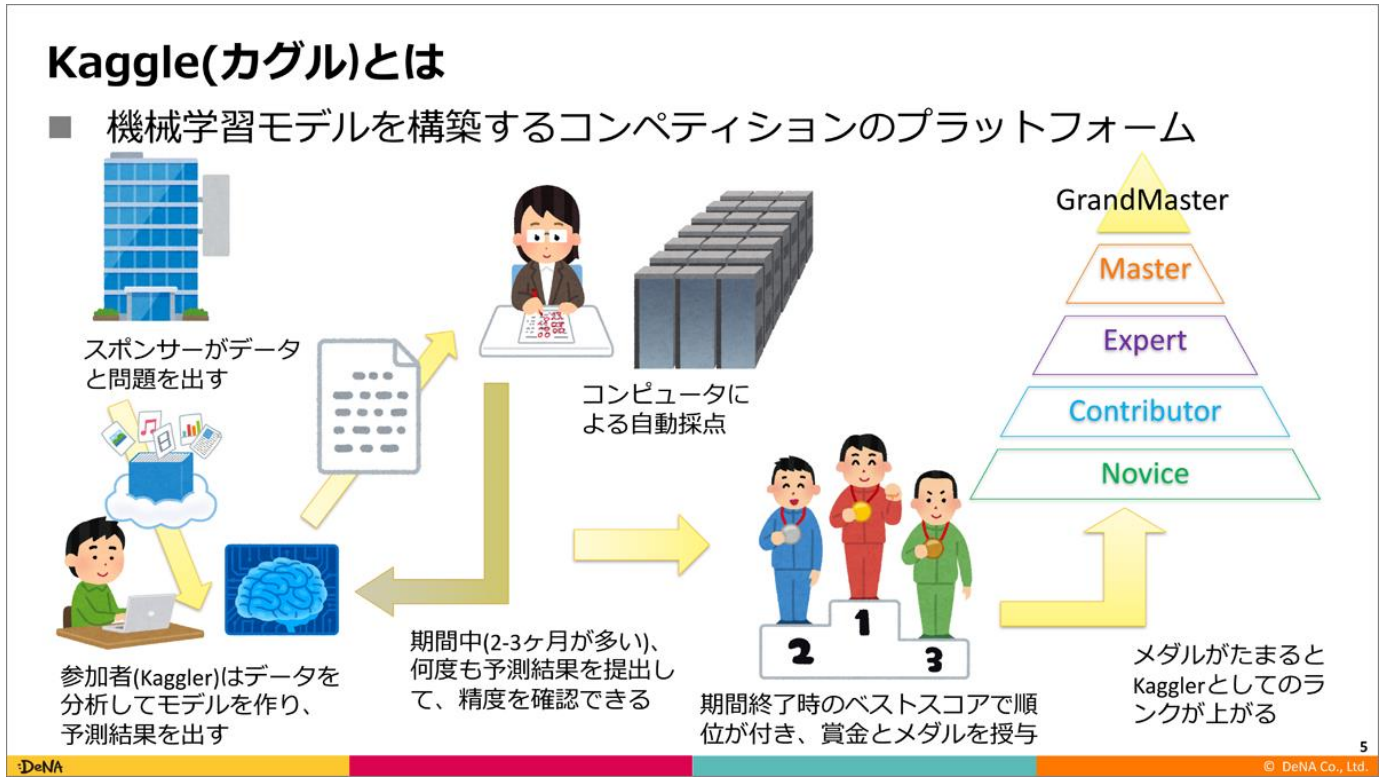
確率モデル：パラメータθ（例えば死亡率、原因）
　　→データ（例えば死亡者数、結果）の確率分布

①データを得る前のθ（死亡率）は、確率分布Lにしたがうとする。
②データD（死亡者数）を入手。
③データDの情報を踏まえた、θの確率分布L'が計算できる。（確率洗替）
④より正確な分布L'にもとづいて、結果（将来の死亡者数）を予測できる。

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　⑦Kaggleとは

・企業や政府などの組織とデータサイエンティストを繋ぐプラットフォームで誰でも無料で参加可能(会員登録は必要)



DNAから資料を拝借

【コンペティション例】

鋼の欠陥検出	**検知**
崩し字の認識	**識別**
家の価格予測	**予測**

初心者にも安心

コンペティションに参加しなくてもこんなことができる！

・Kaggle内のKernel(カーネル)と呼ばれるページで公開されているモデルや分析手法で勉強できる（日本語で記載されているものもある）

・ディスカッションページを見ることで、Kaggle参加者がモデル作成においてどのようなことを考えているかを知ることができる

⇒いろいろ試せる！

Copyright (C) 2019 JUAS All rights reserved

2.ビジネス課題・実践のための知識　⑧Ｓｉｇｎａｔｅとは

- 企業側の課題感
  - ビッグデータ、AIを利用したいが外注ではコストや時間がかかる
  - 満足するできになるか保証されない
  - 内製するにも、人材がたりない
- ビッグデータ、AIに関する課題をコンペティション方式で解決するプラットフォーム
- 提供データに基づいて優秀なモデルを開発したエンジニアを表彰する
- エンジニアへ企業の課題や学習機会を提供する
  - といいつつUdemyかも？

Kaggleとの比較

- ポジティブ
  - 日本語
  - 問題の提供も知ってる企業
- ネガティブ
  - 開催中のコンペ数が少ない → 現在、公開中なのは2コンペ
  - カーネルの提供がない
- その他
  - コミュニティ（情報交換の場）がない
  - 賞金額はあまりかわらないかも

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales

●概要

- ➢小売店の過去33か月分の売り上げデータをもとに、各商品が各店舗で、次の月にいくつ売れるかを予測する
- ➢予測のスコアは二乗平均平方根誤差を使用する（値が小さいほど制度が良い）

●URL

- ➢https://www.kaggle.com/c/competitive-data-science-predict-future-sales/overview

●データ

ファイル名	内容	カラム
items.csv	商品マスタ	商品名、商品ID、カテゴリID
shops.csv	店舗マスタ	店舗名、店舗ID
item_categories.csv	商品カテゴリマスタ	カテゴリ名、カテゴリID
sales_train_v2.csv	過去のPOSデータ（いつどの店舗で何が何個売れたか）	日付、店舗ID、商品ID、商品価格、売上個数
test.csv	次の月の売り上げを予測する対象	店舗ID、商品ID

⇒これらのデータから次月の売り上げを予測する

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales_事前準備

●まずはKaggleの該当コンペのNotebook（※）から、評価の高いものを1つ選定し、内容を確認した。
内容は以下の通り。

- Pythonのコードと、その内容説明が記載されている
- コードをそのまま実行すると、スコアは0.919（2019/10時点で4729人中1347位、1位は0.79）
- XGBoostという決定木系のアルゴリズムを使用している
- 前処理として、以下のような処理を実施している
  - ・外れ値を除外
  - ・異なる店舗IDなのに同一の店舗があるので、1つにまとめる
  - ・欠損値を中央値で埋める
  - ・店舗名には都市名が含まれているので、都市名を取り出して列として追加
  - ・カテゴリ名はカテゴリとサブカテゴリの2つからできているので、その2つを分離
  - ・1か月前、2か月前、3か月前・・・の売り上げデータ集計して列として追加
  - ・その月の日数を列として追加

（※）Kaggle上に公開されている、他の人が実施した分析情報

⇒このNotebookをベースラインとして、スコアの向上を目指した。

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales_前処理

●検証1：重要度の高い説明変数を抽出する

➢XGBoostの機能により、説明変数の重要度が計算できる。
学習に使う説明変数を重要度の高い変数に制限することで、スコアが向上するかどうか確認する。

説明変数の数	スコア
Top2（最低限）	1.13507
Top7	1.02953
Top12	0.96357
Top22	0.91272
Top32（ベースライン）	0.91029
38（全量）	0.90938

⇒今回のケースでは説明変数を多く含めた方が良いということが分かった。

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales_前処理

●検証2：感覚的に明らかに重要度の低い説明変数を除外する
　➢店舗が「どの地域か」は重要度低と判断し、除外。
　　スコアが向上するかどうか確認する。

説明変数の数	スコア
38（全量）	0.90938
31（感覚で選出）	0.90398

⇒感覚的に明らかに重要度の低い説明変数を除外することで、スコアが向上した。

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales_モデル選択

●検証1：予測アルゴリズムを変更する

➢NotebookではXGBoostを利用していたが、他のアルゴリズムではスコアはどうなるかを確認した。

アルゴリズム	説明	スコア
k-近傍法	最も特徴の近いデータをもとに値を算出するアルゴリズム	1.06
線形回帰（重回帰）	データから値を算出する数式を求めるアルゴリズム	0.96
ニューラルネットワーク	線形回帰を複数段重ねた数式を求めるアルゴリズム	0.95
ランダムフォレスト	複数の決定木を作成し、それぞれの結果の平均を取るアルゴリズム	1.13
勾配ブースティング	複数の決定木を繋げて高性能な決定木を作成するアルゴリズム	0.93
Light GBM	最適化された勾配ブースティング	0.92
XGBoost（ベースライン）	最適化された勾配ブースティング	0.91

⇒今回のテーマでは勾配ブースティング系のアルゴリズムが高いスコアが出ることが分かった。

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales_モデル選択

●検証2：アンサンブル学習

- ➢アンサンブル学習とは、いくつかの予測アルゴリズムがそれぞれ出力した予測結果をもとに、最終的な予測値を総合的に判断させるアルゴリズムのこと。
- ➢元々スコアの高かったアルゴリズムをいくつか選定し、アンサンブル学習によってスコアが向上するかどうか確認した。

アルゴリズム	元のスコア	アンサンブル学習スコア
Light GBM	0.92	0.96
XGBoost	0.91	
ニューラルネットワーク	0.95	
勾配ブースティング	0.93	
線形回帰（重回帰）	0.96	

⇒今回のテーマにおいては、アンサンブル学習ではスコア向上しなかった。

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales_チューニング

●ハイパーパラメータのチューニング

- ➢XGBoostについて、Notebookで指定してあるパラメータをベースとして、パラメータを変更し結果を確認する。
- ➢ほとんどのパターンにおいて、微調整してもスコアが変わらないか下落する結果となった。
- ➢変わらない場合は実行時間が短いパラメータを選択し、チューニングのサイクルを短くした。

	スコア	max_depth	n_estimators	min_child_weight	colsample_bytree	subsample	eta	seed	備考
1	0.91187	8	1000	300	0.8	0.8	0.3	42	ベース
2	0.94660	**3**	1000	300	0.8	0.8	0.3	42	Max_depth削減
3	0.91187	8	1000	300	0.8	0.8	**0.2**	42	Etaの調整
4	0.91339	**9**	1000	300	0.8	0.8	0.2	42	Max_depth追加
5	0.91187	8	**1500**	300	0.8	0.8	0.2	42	n_estimators追加
6	0.90873	10	1500	**0.5**	0.8	**1**	0.2	**1**	min_child_weight見直し
7	0.91463	**8**	1500	0.5	0.8	1	0.2	1	最高スコアからMax_depth削減

⇒項番6のパターンが最も良いスコアとなった。

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales_結果サマリ

●結果

➢最終的に最も高いスコアとなったのは、XGBoostでハイパーパラメータをチューニングしたパターンだった。

ベースラインの提出スコア	取り組み後の提出スコア	変更点	順位（2020/02時点）
0.919	0.90873	ハイパーパラメータのチューニング（min_child_weightを小さくする等）	5799チーム中1132位

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales_活動における学び・気づき_

●学び・気づき

◆ 全体

➢Kaggleは、データ分析手法を学ぶ上で非常に有効な手段であった

➢結果がスコアとしてわかる

➢お手本が多く、1からのコーディングは不要

➢ 分析の目的や、データ自体の意味を解釈できる知識は重要

➢ 一連のプロセスの方針をクリアに定めるには、目的やデータの意味を理解し、それに沿った前処理の実施し、適切な説明変数を与えて予測する必要がある

➢ Kaggleで与えられた情報だけでなく、その業界の特徴や良く使用される指標を理解すると有効かも

◆ モデル選択

➢モデルについては、単純に数多く知っておいた方が有利
とはいえ、まずは高性能で主要なものをいくつか使い分けられるレベルを目指すのが現実的か

➢予測アルゴリズムを変更するのは簡単（呼び出すPythonのライブラリを変えるだけ）なので、それぞれの特徴を深く理解し使い分ける必要は必ずしもなく、とりあえず一通りためしてみてスコアが高いものを採用する方針でもよさそう

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　①Predict Future Sales_活動における学び・気づき_

●学び・気づき（続き）

◆ チューニング

➢ 複数あるハイパーパラメータの組み合わせでスコアの上下動があり、今回はその分析パラメータを勘に頼るという結果となってしまった。自動化する方法や論理的にパラメータを決定する方法にする必要性を感じた

➢ モデル選択やチューニングについては、ある程度自動化することが可能なようなので、
このプロセスは自分の中でベストな流れを定型化しておき、あまり時間をかけずに行えるようにすべき
  ➢ 各モデルの仕組みや特徴を詳細に理解していればこの限りではないが、
  そこまでの理解度に到達するには、相当な学習コストと経験が必要になりそうなので

➢ パラメータチューニングは非常に時間がかかるため、同じ結果が短時間で出るパラメータにして、効率的に実行する必要性があると感じた

◆ 前処理

➢分析結果を出すまでの作業のほとんどは前処理であり、重要であると感じている

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　②House Prices：Advanced Regression Techniques

●概要

➢ボストンの一定期間（2006～2010年）における家の売買データから、同期間における成約価格を予測する。

●URL

➢https://www.kaggle.com/c/house-prices-advanced-regression-techniques

●データ

ファイル名	内容	カラム
train.csv	トレーニングセット	家の情報、売買に関する情報
test.csv	テストセット	上記に加え、成約価格
(data_description.txt)	(各項目内容の説明)	-

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　②House　Prices_事前準備①

●Notebook

➢Kaggleの公開情報を一切見ず、**自力**でモデルを構築することとした。

Copyright (C) 2019　JUAS All rights reserved

3.テーマ　②House　Prices_事前準備②

- KNIMEでデータ整備、モデル作成・分析はPrediction Oneにて実行

  - Prediction Oneの特徴
    - KNIME、Prediction One共に無料
    - 二値分類、多値分類、数値予測が選択可能
    - モデル作成時に寄与度として、説明変数ごとの目的変数への影響度を見ることができる
    - モデル作成時のアルゴリズムは表示される実行ログを見て推測するしかない（ブラックボックス）
    - モデル作成から予測実行まではGUIより、以下のプロセスで実行できる
      - Trainファイル選択
      - 予測分類、説明変数、目的変数、交差検証の有無などを指定し、モデル作成実行
      - Testファイルを選択し、予測実行。予測結果を出力

  - 「市民データサイエンス」なる言葉を聞きかじったので、モデル作成のロジック部分をすっとばし、超素人でもある程度の結果が出せないかと思い立ったため、当ツールを選定。

Copyright (C) 2019　JUAS All rights reserved

3.テーマ ②House Prices_事前準備③

- AWSのSagemakerのAutoPilotを使って予測を実行
  - できること
    - 自動的なデータクリーニングと前処理
    - 線形回帰、2値分類、多値分類の自動アルゴリズム選択
    - 自動ハイパーパラメータチューニング
    - 分散学習
    - インスタンス、クラスタサイズの自動選択
  - 特徴
    - Sagemaker上で少量コーディングのみでモデルを作成
    - 作成したモデルをAPIとして公開できる
    - 同様にバッチ処理のエンジンとしても使用できる
    - モデルの構築に使用したNoteBookを確認できる
      - データセットに関する情報
      - モデルの構築のために施行した内容など（EDAなど？）

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　②House　Prices_前処理

- ●前処理①
  - ●欠損値を埋める（数値項目は平均、カテゴリ項目は新カテゴリを代入）
  - ●順序データとしての意味を持つカテゴリ項目は、数値データに変換。
  - ●順序データとしての意味を持つカテゴリ項目は、0(非該当)と1(該当)に変換。
  - ●不要な変数（相関がほぼ0、ほぼ一つの値しかとらない等）は除去。
  - ●「販売年」と「販売月」から、「販売年月」の数値項目を作成
- ●前処理②
  - ●統計量（平均、中央値、標準偏差、ヒストグラムなど）の確認
  - ●散布図、相関係数の確認
  - ●散布図より外れ値の特定と除外
  - ●NoeteBook（主にEDA）を参考にして、項目の結合、欠損値の補完などを実施
- ●前処理③
  - ●AWSのSagemakerのAutoPilotによる自動的なデータクリーニングと前処理
- ●前処理④
  - ●「SalePrice」に相関のある項目を検討し、2項目を選択（OverallQual、SaleCondition）
  - ●SaleConditionは文字のため、数値に変換

Copyright (C) 2019　JUAS All rights reserved

3.テーマ　②House　Prices_前処理

●前処理⑤
- ●欠損が多いように見えるため、平均値で埋め、フォレストモデルで分析する（基本方針）
- ●データ項目の整理として、項目の意味を確認し、数値型・文字列型を定義。数値の欠損はすべて平均で補完
- ●欠損値の取り扱いの見直し
  ⇒改めて項目の意味を確認し、数値の”null”を原則として0に置き換える
  ⇒フォレストモデルで影響力の大きなカテゴリを探し、そのグルーピング範囲内で欠損値を埋める
- ●特殊な値がないか確認
  →ガレージの有無で、ガレージ関連項目がすべて記載なしとなる。そのまま？モデル分ける？
  　モデルを分けて誤差を見たが、ガレージ有だけでモデルを作ってもtrain内で誤差が大きかったのでそのまま。
  　プールも同様の問題があるが、件数が少ない（1463/1460がない）ためそのまま
- ●数値のnullへの対処法
  lotfrontageに空きが多い(200以上)
  MasVnrAreaにも空きがある(10未満)
  　→最も説明力のある(neighborhood)の平均値で埋め、再度フォレストモデルで分析実施
  GarageYrBltにも空きがある(80以上)

Copyright (C) 2019　JUAS All rights reserved

3.テーマ　②House_Prices_モデル選択

●モデル選択①
　●勾配ブースティングのライブラリ：LightGBMを使用

●モデル選択②
　●基本、Prediction Oneにまかせきり。実行ログを見る限り、ニューラルネットワーク、勾配ブースティング木(XGBoost)、線形回帰モデルを基にアンサンブル学習を実施していると思われる。

●モデル選択③
　●AWSのSagemakerのAutoPilotによる、線形回帰、2値分類、多値分類の自動アルゴリズム選択

●モデル選択④
　●LinearRegression　で以下の2パターンで実行
　①説明変数にOverallQualを設定して実行
　②説明変数にOverallQualとSaleConditionを設定して実行

●モデル選択⑤
　●線形回帰、フォレストモデルを選択
　⇒実データとの誤差を見ると、高価格帯を過小評価するモデルが出来上がった

Copyright (C) 2019　JUAS All rights reserved

3.テーマ　②House_Prices_チューニング

●チューニング①
　●モデルの誤差を最小化するために、ハイパーパラメータを調整。
　●方法としては、まずGridSearchCV（総当りで探索）を最初に試した。
　●次に、Optuna（ベイズ推定による探索）を用いて、より効率的に調整

●チューニング②、③
　●Prediction One　、AWSのSagemakerのAutoPilotによる、それぞれ自動チューニング

●チューニング⑤
　●誤差を平均値で埋め、予実にあまりにも乖離のあるデータ（7件）を外して検証
　　予測値が、実績値の1.4倍以上、または0.85倍未満

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　②House　Prices_結果サマリ

●結果

●やはり経験者が強かった。

No.	事前準備（ツール）	前処理	モデル	チューニング	RMSLE	順位
1	Python	欠損値埋め カテゴリ項目⇒数値変換 不要変数除外 項目追加	LightGBM	GridSerchCV	0.12758	1685位/5288人
	Python	同上	LightGBM	Optuna	0.1255	1507位/5459人
2	KNIME Prediction One	未加工	自動選択 （NN、XgBoost、線形回帰）	自動	0.13966	2849位/5654位
		外れ値除外 低相関除外 欠損値多項目除外		自動	0.12934	1861位/4960人
3	Python	未加工	LightGBM	なし	0.23899	4114位/4755人
	Sagemaker	自動	自動選択 （線形回帰、2値分類、多値分類）	自動	0.13558	2283位/4755人
4	Python	説明変数1つ	LinearRegression	−	1.16054	
		説明変数2つ	LinearRegression	−	1.16083	4490位/4616人
5	Alteryx	欠損値を平均で埋め データ項目の整理	線形回帰	−		
		同上	フォレストモデル	−		
		欠損の取り扱い見直し nullへの対処	フォレストモデル	−		
		乖離データの除外	フォレストモデル	−		

Copyright (C) 2019　JUAS All rights reserved

3.テーマ　②House　Prices_活動における学び・気づき_

- KNIMEを使用してのデータ加工、分析は初心者でもある程度まではお手軽にできるが、KaggleのNoteBook上はPythonが大正義のため、情報量が圧倒的に少なく、少しつっこんだ処理をしようとすると試行錯誤がとても多くなる（コピペができない）。
- Prediction One（ツール）を使用しての予測分析はお手軽で、それなりの精度が出ている（？）と思われるが、ロジックはブラックボックス。
- プロジェクトオーナーからしたら、どのツール、ロジックを使おうが、中身はブラックボックス。説明変数がどのように寄与しているかなど、「見える化」が説得力の強化に必要では。
- データ加工、統計量分析、結果の評価などは、どのツールを使うにしても勉強は必須。逆にモデル作成の部分はある程度ツールに頼れるのでは？
- RMSLEについて、どの程度であれば有意義な結果であると、偉い人たちに対して説得力を持って説明できるのか？
- Python初心者だったため、インストールや使い方から学ぶことで時間がかかった。
- コードの書き方なども試行錯誤で基本サンプル以上のことはできなかった。
- 分析方法の選択や結果の評価に時間を割く余裕がなかった。
- 他のメンバーの結果などから、分析についてはある程度ツールで結果が出せるように思った。
  実際の現場で使用することを想定すると、分析結果(説明変数がどう影響があるかなど)を
  説明できるかが重要だと思う。

Copyright (C) 2019　JUAS All rights reserved

3.テーマ　②House　Prices_活動における学び・気づき_

- 単純にデータ項目を削るだけだと、制度が悪くなる場合が多かった。
- 欠損値を埋めるには、その項目の意味を理解しないと誤った解釈を導く可能性がある
- PredictionOneなどのツールは、単体でそのままデータを投入しても精度の高い結果が出る
  ⇒データサイエンティストの確保だけでなく、機械的に判断不可能な「実務・実態に合わせて
  　欠損値を解釈できる人」の確保だけでも分析力のベースアップに繋がるのでは？
- 試行錯誤に加えて、検証時のデータの保全に時間を取られたことが反省
  ※試行錯誤に夢中になり、データ加工や分析断面の保存を忘れることがあった。

Copyright (C) 2019　JUAS All rights reserved

3.テーマ ②House Prices_追加でやりたかったこと

- 追加でやりたかったこと
  - 手法の洗練
    線形モデル、SVM、ベイズ推定など、他のモデルも用いて予測。
    アンサンブル学習（スタッキング）を用い、各モデルのいいとこどりをしたモデルを作成。

  - 特徴量加工
    寄与が線形[単調]でない特徴量は、変数変換を行い、できるだけ線形[単調]となるようにする。
    分布を正規分布でない特徴量は、変数変換を行い、正規分布に近づける。（線形モデルのため）
    交互作用項を作成。（線形モデルのため）

  - モデル可視化
    大域的な解釈性指標を用い、モデルの解釈を行う。得られた解釈が常識に反しないかを確認する。
    変数の重要度、Partial Dependence、ALE、決定木代替モデル等。

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　③IEEE-CIS Fraud Detection

●概要

➢クレジットカードの不正利用検知システムの改善

●URL

➢https://www.kaggle.com/c/ieee-fraud-detection

●データ

ファイル名	概要	内容
sample_submission.csv	提出ファイルのサンプル	トランザクションID、不正フラグ
test_identity.csv	テストセット	ネットワーク接続情報およびトランザクションに関連付けられたデジタル署名
test_transaction.csv	テストセット	タイムデルタ、支払額、製品、カード情報、位置情報、メールドメイン、意味が隠されたカウント
train_identity.csv	トレーニングセット	
train_transaction.csv	トレーニングセット	Testに正解データとして不正フラグが追加されたもの

⇒trainデータで学習させ、テストデータに対して不正検知を実施する

Copyright (C) 2019 JUAS All rights reserved

3.テーマ　③IEEE-CIS Fraud Detection

- Notebookなどをなぞった感想
  - 特徴
    - 大量データ
      - そのまま実行しても、データの初回読み込みでメモリエラーが発生
        - リッチな解析環境を持っていないとNotebookをなぞることも難しい
        - サンプル抽出により、データ量を絞り込んで分析する手法を取っているケースもある
    - データの意味、項目数の多さ
      - 項目数は400超で、そのうちの多くのデータの意味は秘匿されている
  - 簡単には試せないが、実業務での解析では遭遇するであろう課題が含まれている

- これから
  - このコンペ課題への取り組みは取り下げ、他の課題に人を振り分けることを検討
    - データ分析の一歩前の段階で多くの労力がかかってしまう
    - テーマ数が絞れていた方が濃い議論が期待できる

Copyright (C) 2019 JUAS All rights reserved

「機械学習モデル解釈」ハンズオン

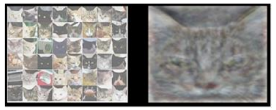
Copyright (C) 2019 JUAS All rights reserved

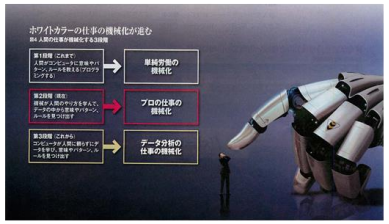
4.「機械学習モデル解釈」ハンズオン　背景

2013年度のデータマネジメント研究会では、機械学習はTopicとして取り上げられていた。

5－2－③. 技術動向
II　Topic　機械学習

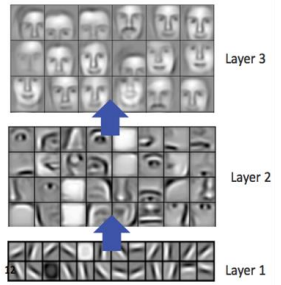
- 人工知能における研究課題の一つで、人間が自然に行っている学習能力と同様の機能をコンピュータで実現しようとする技術・手法のこと
- 大量のデータから従来にない知見を得るというビッグデータの時代では、特にその応用に期待が集まっている

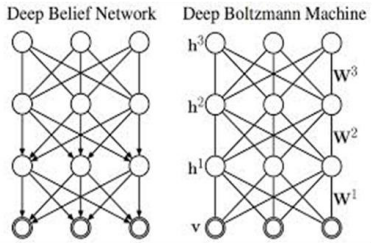




5－2－③. 技術動向
II　Topic　Deep Learning Evolution

- 近年「Deep Learning」が、機械学習分野を席巻している
  - 画像認識・音声認識・化合物の活性予測でDNNが最高精度を達成
  - Microsoftが音声認識アルゴリズムをDeep Learningベースに変更
  - メイン知識の無いチームが最高精度を獲得
- 従来よりも多くの層を持ったニューラルネットを用いる機械学習技術
  - 従来は研究者が手作りしていた特徴抽出器を、代わりに大量のデータから学習させ自己組織化させるアプローチ



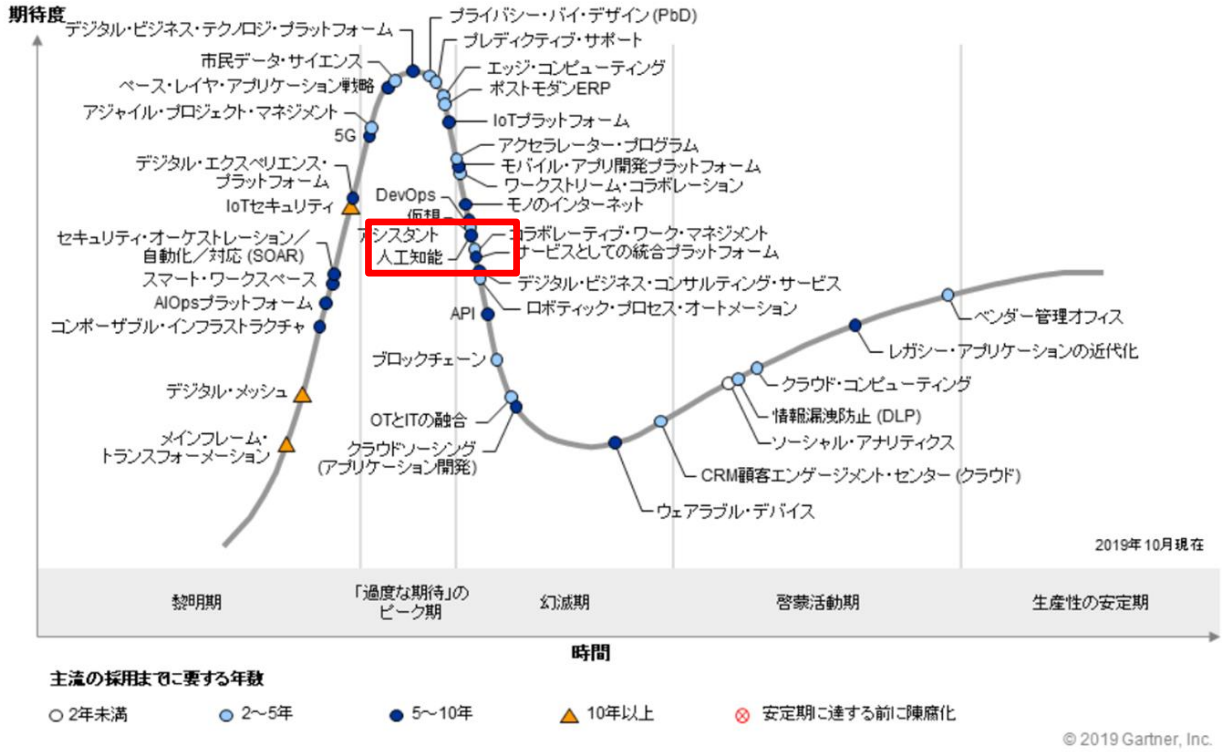


2013年度　データマネジメント研究会　「データ分析技術と実務」より

Copyright (C) 2019　JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　背景

2013年当時はガートナーのハイプサイクルには「AI」や「人工知能」は記載がなかったが、今や幻滅期に位置している。



「日本におけるテクノロジのハイプ・サイクル：2019年」
https://www.gartner.com/jp/newsroom/press-releases/pr-20191031
(2019年3月5日)

Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　背景

幻滅期から啓蒙活動期に移っていくためには、説明性が必要なのでは？と考えた。
※　分析対象にもよるが、我々の研究対象が「ビジネスデータ」のため

> 一口に「説明力」といってもいろいろな面があります。たとえば次の３点を挙げることができます。
>
> 1. 説得力：原理に説得力があり（関係者を説得しやすいという意味で）わかりやすい。
> 2. 単純性：（人が把握できる程度の個数の特徴量の観測値をもとにして複雑な場合分けのない算式によって予測値を算出できるなど）特徴量から予測値を求める算式が複雑すぎない。
> 3. 解釈性：各特徴量が予測値に与える影響がわかりやすい。

岩沢宏和・平松雄司　（2019），入門　Rによる予測モデリング　東京図書

我々はこの「説明性」ということについて、今、何が、どこまで明らかにすることができるかという観点で調べてみることとした。

Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　なぜ説明性が求められるか

[人間の意思決定に機械学習モデルを使うには説明が必要]

現在の機械学習モデルは高い精度での予測・認識は得意だが、その判断根拠の説明が苦手。そのことがAI（機械学習モデル）への不信/導入の阻害につながっている

例)
AIと医師で診断が食い違った
→ AIの判断根拠がわからないと**原因の特定ができない**。
→ 根拠不明と**AIの判断が無視される**なら、AI導入の意味がない

金融業者「あなたに融資できません」
→ 「なぜ？」と問いかけられたときに**理由が説明できないサービス業者は信頼されない**

⇒データの利活用に結びつかない

※このスライドのAIとは機械学習、ディープラーニングを用いた予測・認識を行う識別器のことを指します。

Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　説明性に関する要求の高まり

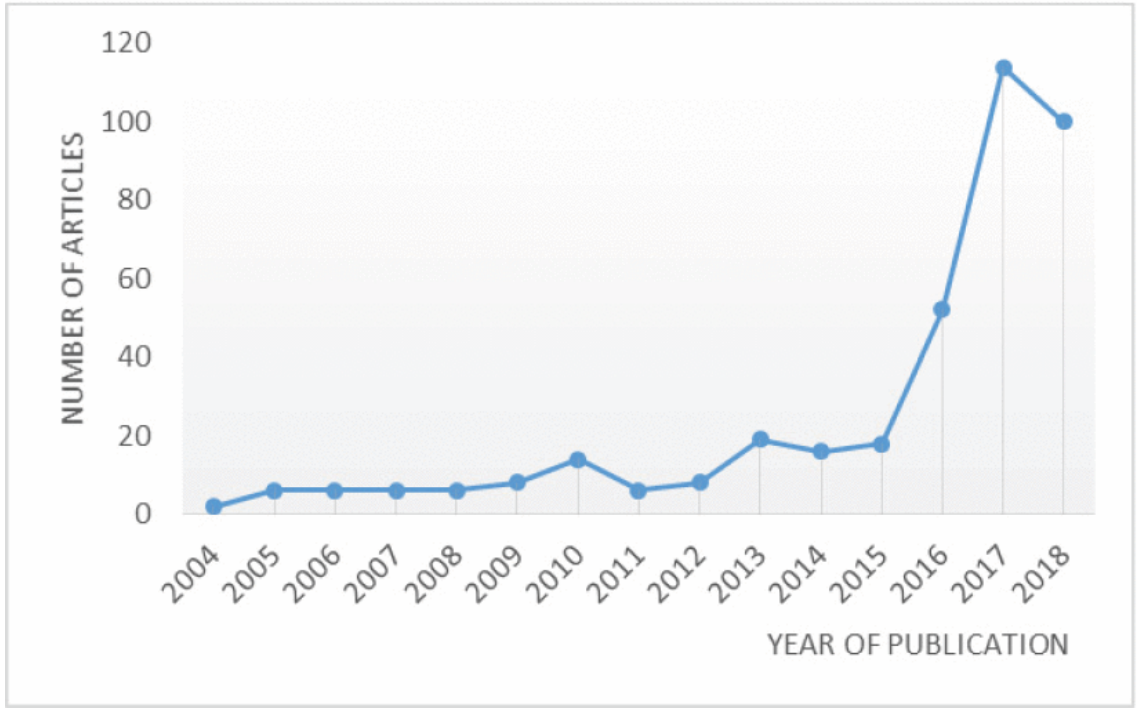
[ICML 2017 Tutorial]

“**Interpretable Machine Learning**”

As machine learning systems become ubiquitous, there has been a surge of interest in interpretable machine learning: systems that provide explanation for their outputs. These explanations are often used to qualitatively assess other criteria such as safety or non-discrimination. However, despite the interest in interpretability, there is little consensus on what interpretable machine learning is and how it should be measured. In this talk, we first suggest a definitions of interpretability and describe when interpretability is needed (and when it is not). Then we will review related work, all the way back from classical AI systems to recent efforts for interpretability in deep learning. Finally, we will talk about a taxonomy for rigorous evaluation, and recommendations for researchers. We will end with discussing open questions and concrete problems for new researchers.

ICML：International Conference on Machine Learning
機械学習に関する国際会議

機械学習モデルの説明性に関する発表論文数



[総務省　AI利活用原則案]

・透明性の原則

AIサービスプロバイダ及びビジネス利用者は、AIシステムまたはAIサービスの入出力の**検証可能性及び判断結果の説明可能性に留意する**

・アカウンタビリティの原則

AIサービスプロバイダ及びビジネス利用者は、消費者的利用者及び間接利用者を含む**ステークホルダに対しアカウンタビリティを果たすよう努める**

Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　説明性とは

Interpretability is the degree to which a human can understand the cause of a decision.

複雑な分類器の判断基準を人間にも解釈可能になるように提示すること

【説明性の方向性】

[挙動]
局所的な説明：特定の入力に対する予測の根拠を提示　→　今回はここを対象とする
大域的な説明：複雑なモデルを可読性の高い解釈可能なモデルで表現

[仕組み]
説明可能なモデルの設計：最初から可読性の高い解釈可能なモデルを設計

Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　挙動に関する理解

挙動について理解する手法として代表的なものにLIMEとSHAPがある

LIME（Local Interpretable Model-agnostic Explainations）

ー個別のデータに対する予測結果に対し、特徴量（テキストや画像の一部など）をハイライトして説明する

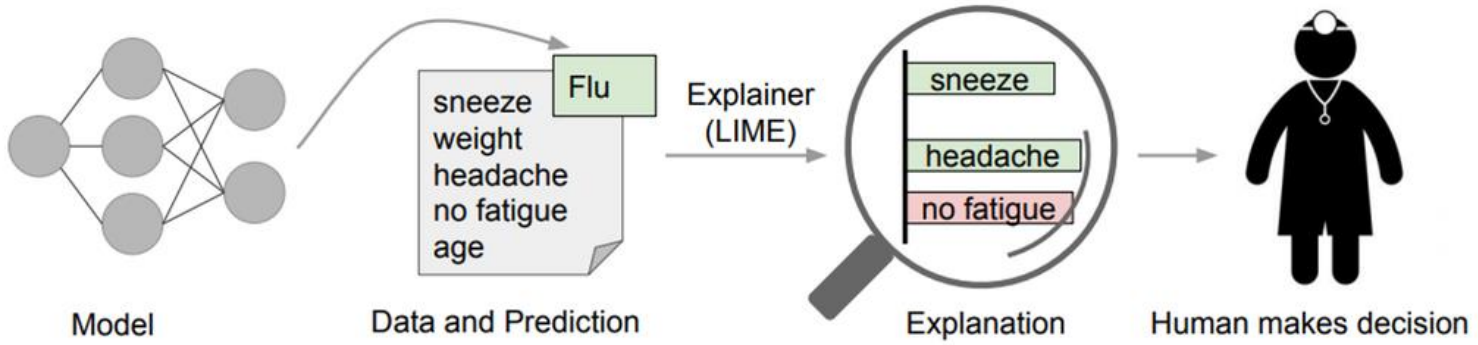


Figure 1: Explaining individual predictions. A model predicts that a patient has the flu, and LIME highlights the symptoms in the patient's history that led to the prediction. Sneeze and headache are portrayed as contributing to the "flu" prediction, while "no fatigue" is evidence against it. With these, a doctor can make an informed decision about whether to trust the model's prediction.

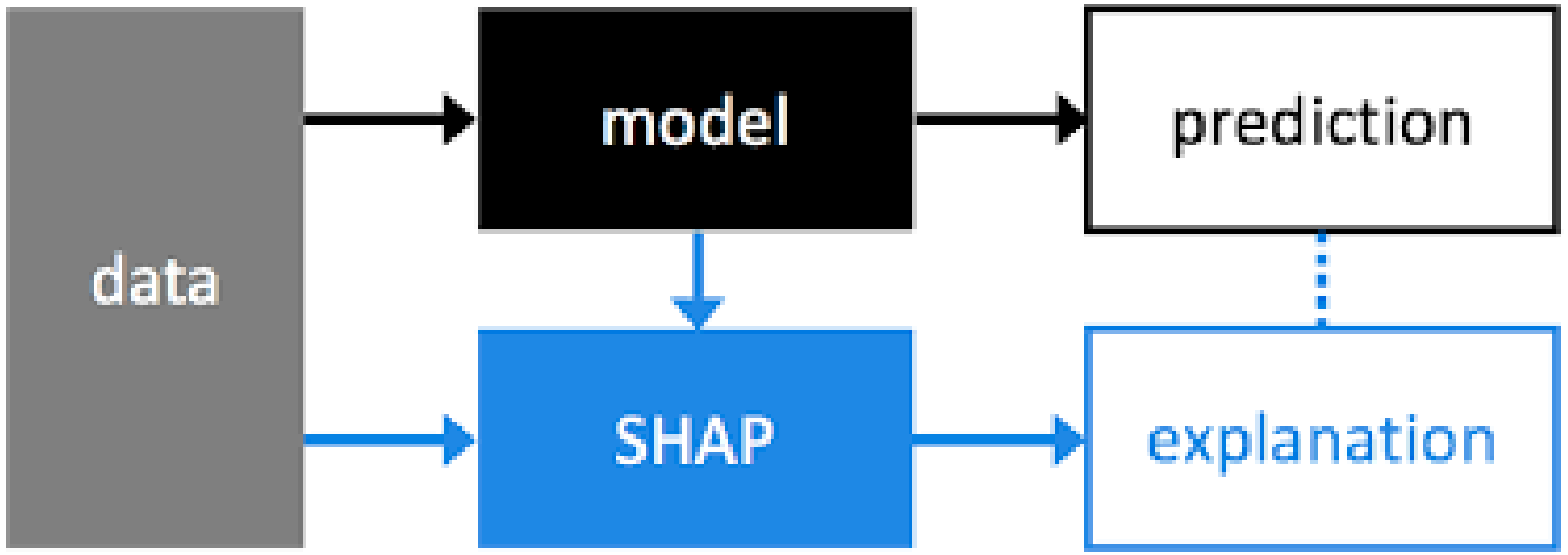
Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　挙動に関する理解

SHAP（Shapley Additive exPlanations）

ーLIMEを含むいくつかの局所的な説明法をゲーム理論の枠組みの下で統一的に記述

→よってLIMEはSHAPに包含される



Copyright (C) 2019 JUAS All rights reserved

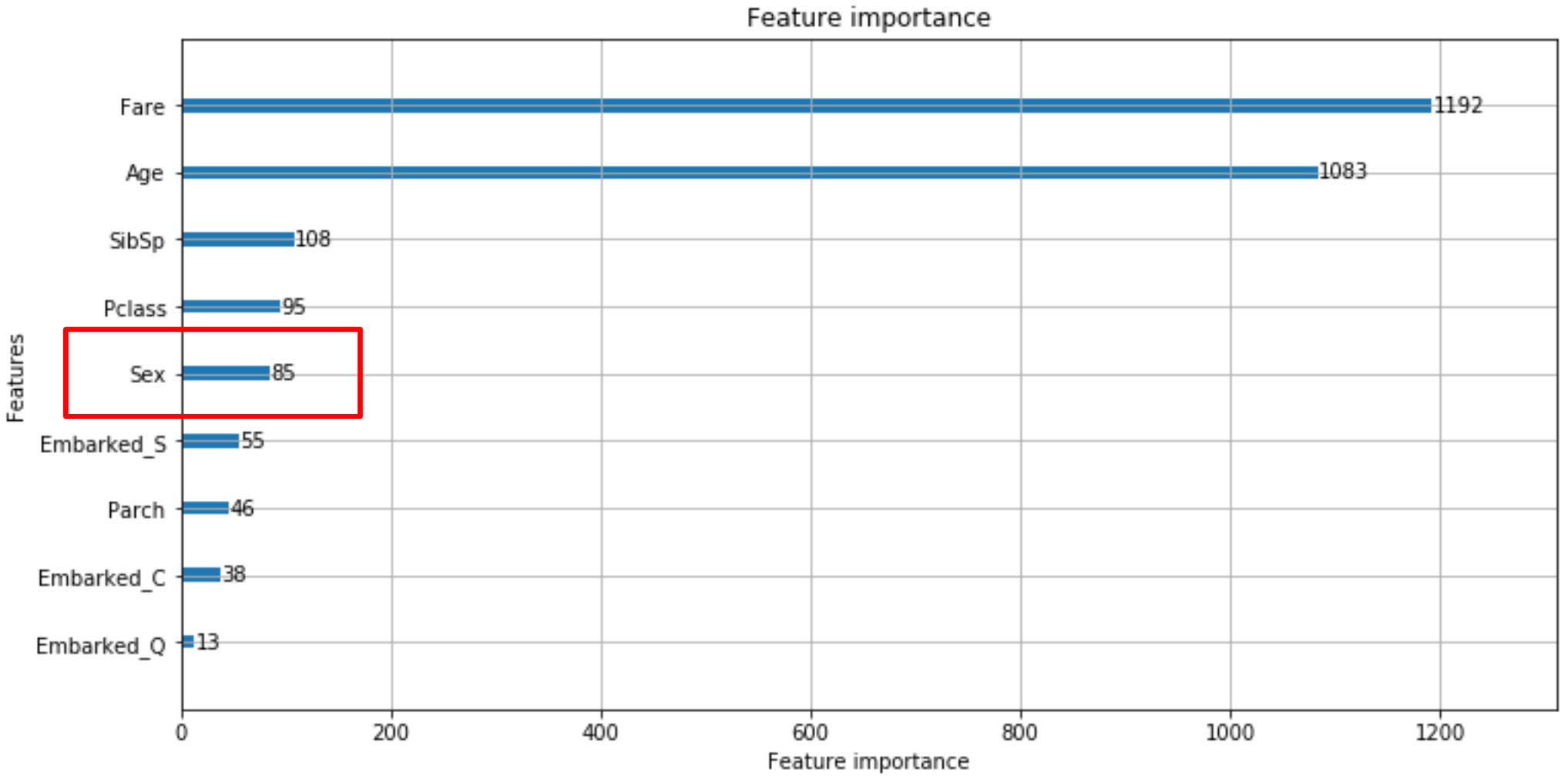
【検証内容】
モデルから俯瞰した重要度と、各々のデータから仰望した重要度が同じになるかを検証する

【手段】
　lightGBMのplot_importanceで表示される「重要な特徴量グラフ」とLIME,SHAPで重要と判断された特徴量を比較してみる

【用いたデータ】
　kaggleのtitanicデータ

Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　**結果(lightGBM)**

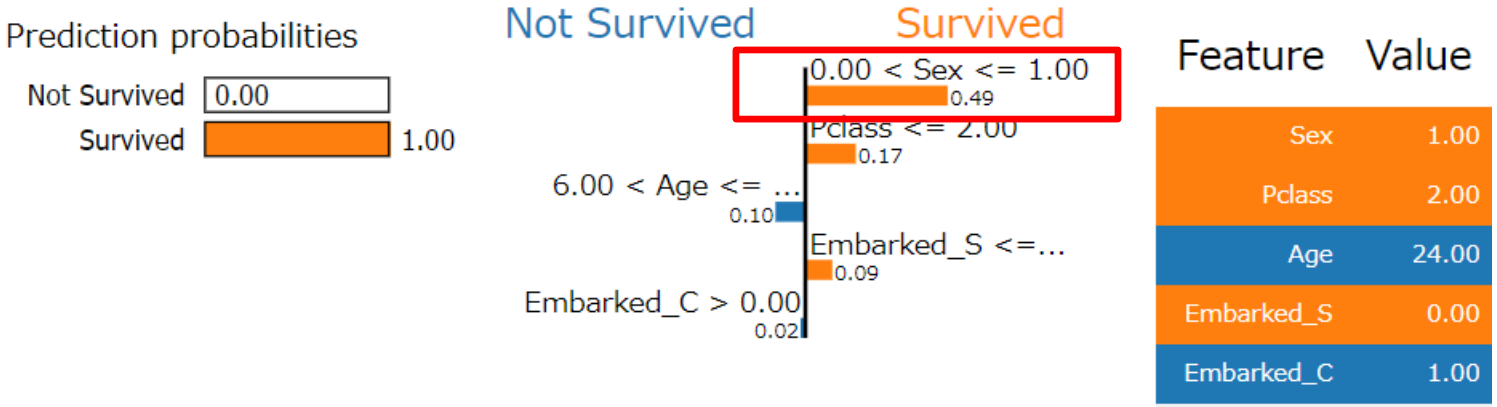


Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　**結果(LIME/SHAP)**

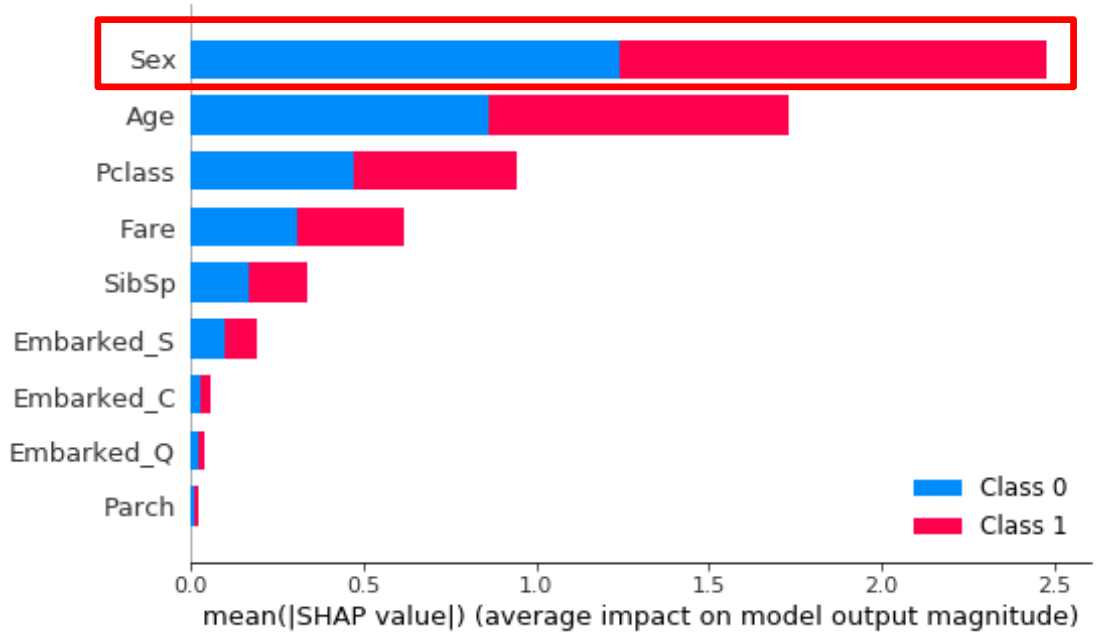
【LIME】

ランダムに一名を指定し、出力したもの



【SHAP】

個人が生前を判断するのに必要とされた項目の重要度を**全体で平均**したもの



Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　**評価**

【lightGBMとSHAPでは**結果が異なった**】

ligthGBMが**Fareが重要**といっているのに対し、SHAPでは**性別**。
　⇒feature importanceの計算式がカーディナリティが高い特徴量を重視する(※)傾向にあることを踏まえても圧倒的に重要度が低い

※特徴量がとる値の種類が多い（2値より3値、3値より4値）ため、数値を値に持つ特徴量のほうが重視されやすい

【導き出した仮説】

　lightGBMでも性別が重要視されるように識別器を作ったら精度が向上するのでは？

【改善内容】

①Titanicデータを男性/女性に分けてligthGBMに学習させる
②テストデータを男女それぞれの識別器にかけて予測を行う
③作成したモデルでの予測結果をkaggleにサブミットして正答率を比較

Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　**結果**

1. チュートリアルについてくるデータ（性別が女性だと生存と判断）

→　正答率：**0.77990**

2. ハンドメイドで作ったlightGBMによる予測

→　正答率：**0.78947**

3. 男女に分けて予測を行う（今回の改善結果）

→　正答率：**0.77033**

。。。だめでした(･ω<)

Copyright (C) 2019 JUAS All rights reserved

個別最適の積み上げは全体最適にならなかった。。。

可視化の手法をアレコレ使ってきたが、
モデルの改良につなげるには、まだまだこれからだ

データ分析の道は険しいが、
一体、どうすれば良いのだ!?
（次のチームに譲るぜ・・・。）

Copyright (C) 2019 JUAS All rights reserved

4.「機械学習モデル解釈」ハンズオン　参考文献

機械学習モデル解釈ナイト（エンジニア向け）聴講ノート
https://blog.masahiko.info/entry/2019/11/29/131309

機械学習モデルによる予測結果の説明性
https://speakerdeck.com/hightensan/lime-and-shap-ji-jie-xue-xi-moderuniyoruyu-ce-jie-guo-falseshuo-ming-xing

機械学習モデルの判断根拠の説明
https://www.slideshare.net/SatoshiHara3/ss-126157179

上記を参考とさせていただきました。

機械学習の解釈可能性を求めないほうがよい、という論文が最近出たなんて話もありますが。。。

Copyright (C) 2019 JUAS All rights reserved

5.考察・まとめ

【考察・まとめ】

- データサイエンスの勉強は、Kaggleの利用が有効
- 機械学習の実施にあたっては様々なツールがあり、初学者・初級者はこれらを利用するとよい
  ある程度の精度は出せるが、作成したモデルの説明性に課題
- Kaggle上ではPython、Rが最先端・大正義のため、情報量はツールよりこちらが上
- ツール利用にしろ、Pythonを操るにしろ、データの前処理の重要さは大前提
- 近年、AIの説明性、見える化にも注目が集まっており、LIME/SHAPなどの手法が出てきている

【今後の展望】

- **ツールを活用することで、「市民データサイエンス」を実現できないか？**
- **「データサイエンス」の利活用の成功は、説明性・見える化がカギを握っている？**

Copyright (C) 2019 JUAS All rights reserved

2－④データビジネス

Copyright (C) 2019 JUAS All rights reserved

データビジネスにおける
ツマヅキポイント

Copyright (C) 2019 JUAS All rights reserved

139

目次　最終報告

１．背景
研究テーマ選定の経緯と背景

２．倫理編
２－１　企業によるデータ活用に対して、社会が抱くイメージとは？
２－２「内定辞退率データ」に関する意識調査の方法
２－３　アンケート集計結果
２－４　アンケート回答におけるインサイト
２－５　データ活用検討時のチェックリスト

３．契約編
４－１　ベンダーとユーザー企業の対立構造・関心の違い
４－２　ベンダー側のコスト・収益構造・ビジネスモデル的な観察
４－３　契約上の落とし所、かわし方
４－４　ベンダー企業とユーザー企業それぞれの関心事の考察から得られるインサイト
４－５　ユーザー企業として考慮すべきポイントまとめ

４．考察・まとめ
４－１．考察内容
４－２．まとめ、今後の展望

参考資料・文献

Copyright (C) 2019 JUAS All rights reserved

1　研究テーマ選定の経緯と背景

- データ利活用については、GAFAMやベンチャーが積極的な姿勢を見せる一方、GDPRやCCPRなどデータ保護に関する動きも活発化している。
- また、データを活用したビジネス・サービスにおいて “炎上する”案件や、AI開発などのPoCでベンダーから高額なサービス費が見積として提示されるケースもあり、データ活用やAI活用に躊躇してしまうユーザー企業も多い。
- 本研究会ではユーザー企業がデータ活用する場合やAI開発をする場合に、どのようなポイントに気を付けるべきかを「事例研究をもとにした倫理面での考慮点」と「ベンダーとの契約上の留意点」を明らかにする。

Copyright (C) 2019 JUAS All rights reserved

2．倫理編

Copyright (C) 2019 JUAS All rights reserved

2-1　企業によるデータ活用に対して、社会が抱くイメージとは？

Copyright (C) 2019 JUAS All rights reserved

組織がデータの価値を活かしきれていない

主な阻害要因は何だろうか？

- データの整備がされていない
- 分析する技術がない
- 分析結果を行動に反映させる体制がない

など様々な要素が考えられる。

これまでのビジネスデータ研究会の活動でもそれらの問題は議論されてきた。
これらは主にケイパビリティーの問題であり、組織能力を向上させることで解決していくべき課題であると言える。

「できない」は「できるようにする」ことで解決する。

Copyright (C) 2019 JUAS All rights reserved

社会通念上「やってよい」ことなのか？

しかし、データ活用の事例が増えていくにつれ、そもそもその取り組みは社会通念上「やってよい」ことなのかどうかが自明ではない、というケースが見られる様になってきた。

個人情報、プライバシーの保護の課題は、GDPRのような既に具体化している規制に**場当たり的に対応するだけでは不十分**ではないだろうか。その**背後にある倫理的な枠組みを理解**しておかなければ、これらの規制が今後どのような方向に波及するのか、を読み解くことはできない。

また、既存のガイドラインが当てはまらないような新しい取り組みに対する判断も、難しくなってしまう。

Copyright (C) 2019 JUAS All rights reserved

「内定辞退率」を題材にした調査の実施

そこで本分科会では、社会的にインパクトが大きかった、就職活動中の大学生のデータを「内定辞退率」という形で顧客企業に提供するサービスを題材にアンケートを実施。

社会が企業によるデータ活用に対して、どのようなイメージを抱いているのかについて調査を行った。

Copyright (C) 2019 JUAS All rights reserved

2-2 「内定辞退率データ」に関する意識調査の方法

Copyright (C) 2019 JUAS All rights reserved

アンケート調査

- ビジネスデータ研究会参加者の直接の知人に限定し59人の回答者を対象にWebアンケートを取得した
- 回答を容易にするために設問形式はすべてYes/Noの2択に統一した
- 設問の内容は2019年の「リクナビ就活事件」を下敷きに、現実の事件の背景を少し改変した条件を設定した上で、そのビジネスモデルであれば許容可能であるかを問うた。（例「行動履歴等の生データだけを企業に提供し判断は各社が行うサービスであったとしたらどうでしょうか？」など）
- 匿名式だが各回答者の勤務先の業種業態や求人活動とのかかわり、回答者自身の年齢なども回答してもらった。

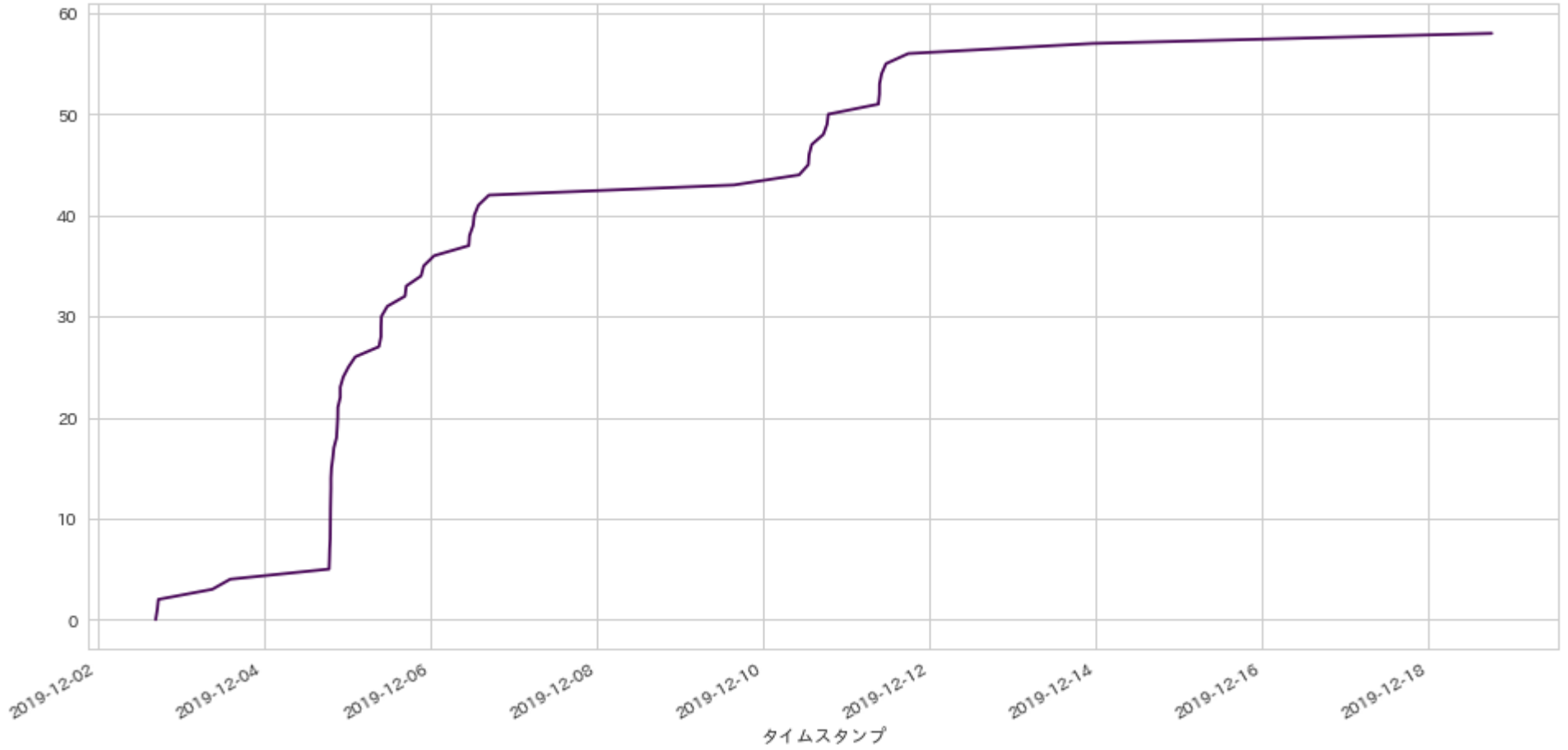
Copyright (C) 2019 JUAS All rights reserved

2-3　アンケート集計結果

Copyright (C) 2019 JUAS All rights reserved

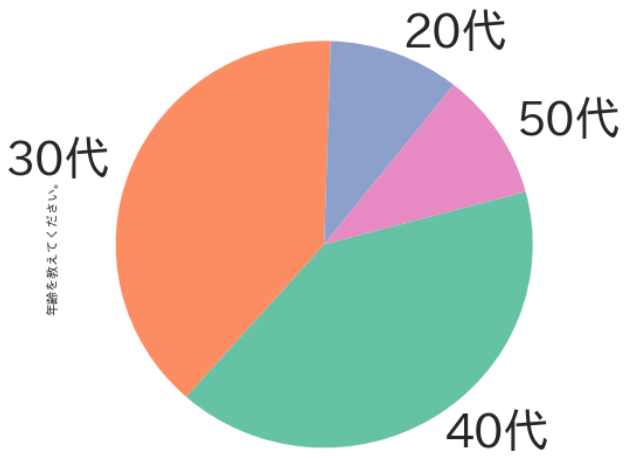
回答数、集計期間

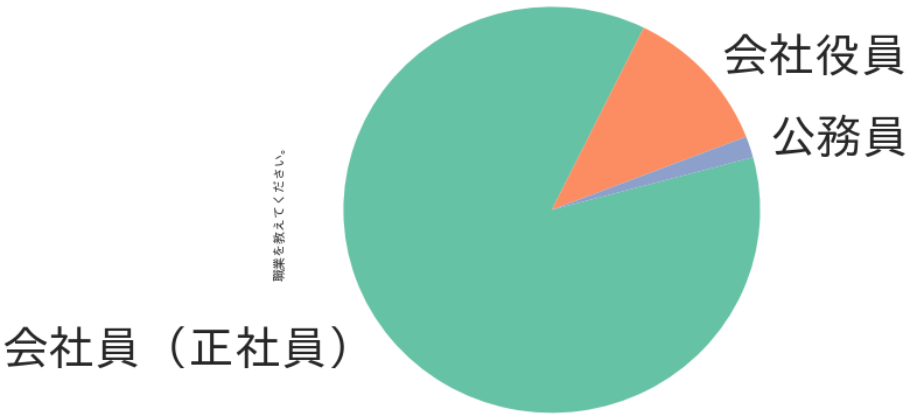
回答数：59人　集計期間 :2019/12/1～2020/12/27

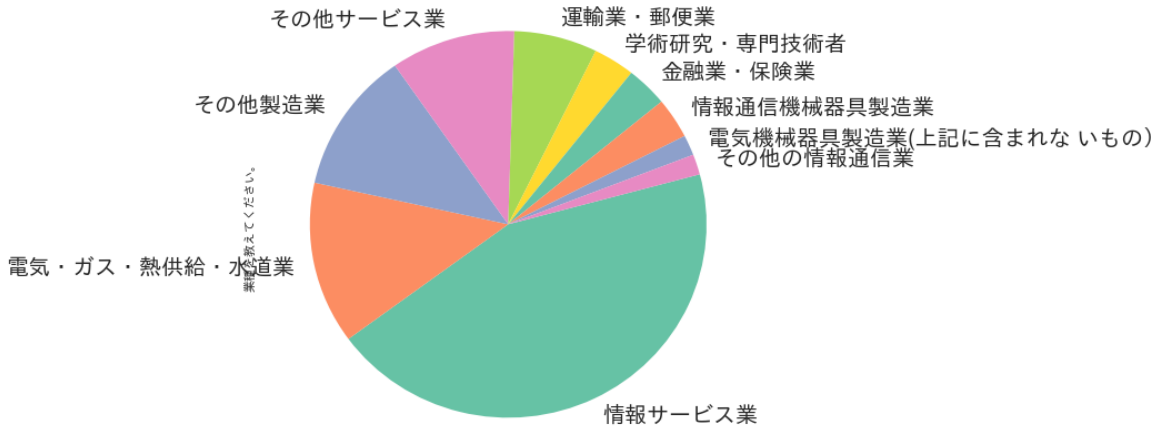


Copyright (C) 2019 JUAS All rights reserved

回答者の属性

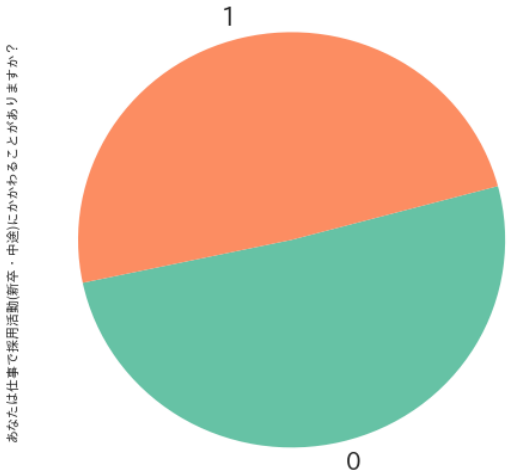


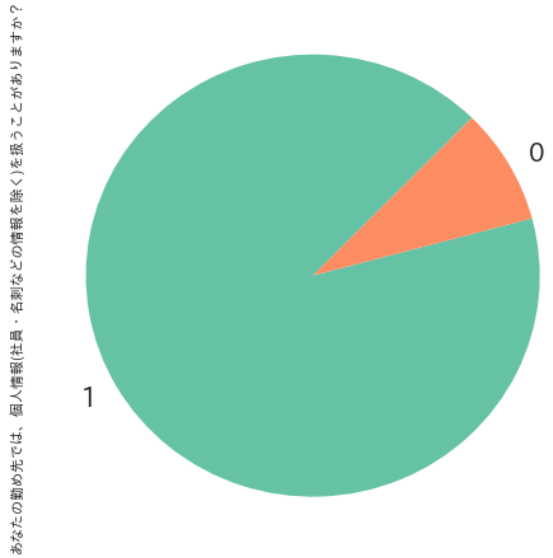




Copyright (C) 2019 JUAS All rights reserved

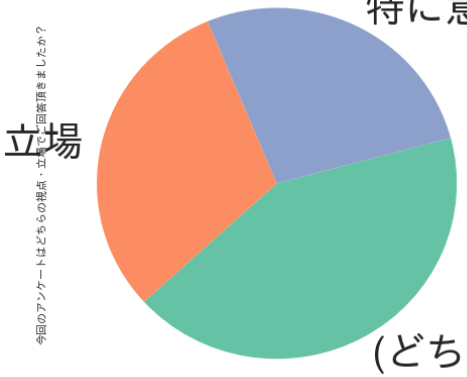
回答者の個人情報に対する意識





(どちらかといえば、)企業の採用担当者の立場

特に意識しなかった



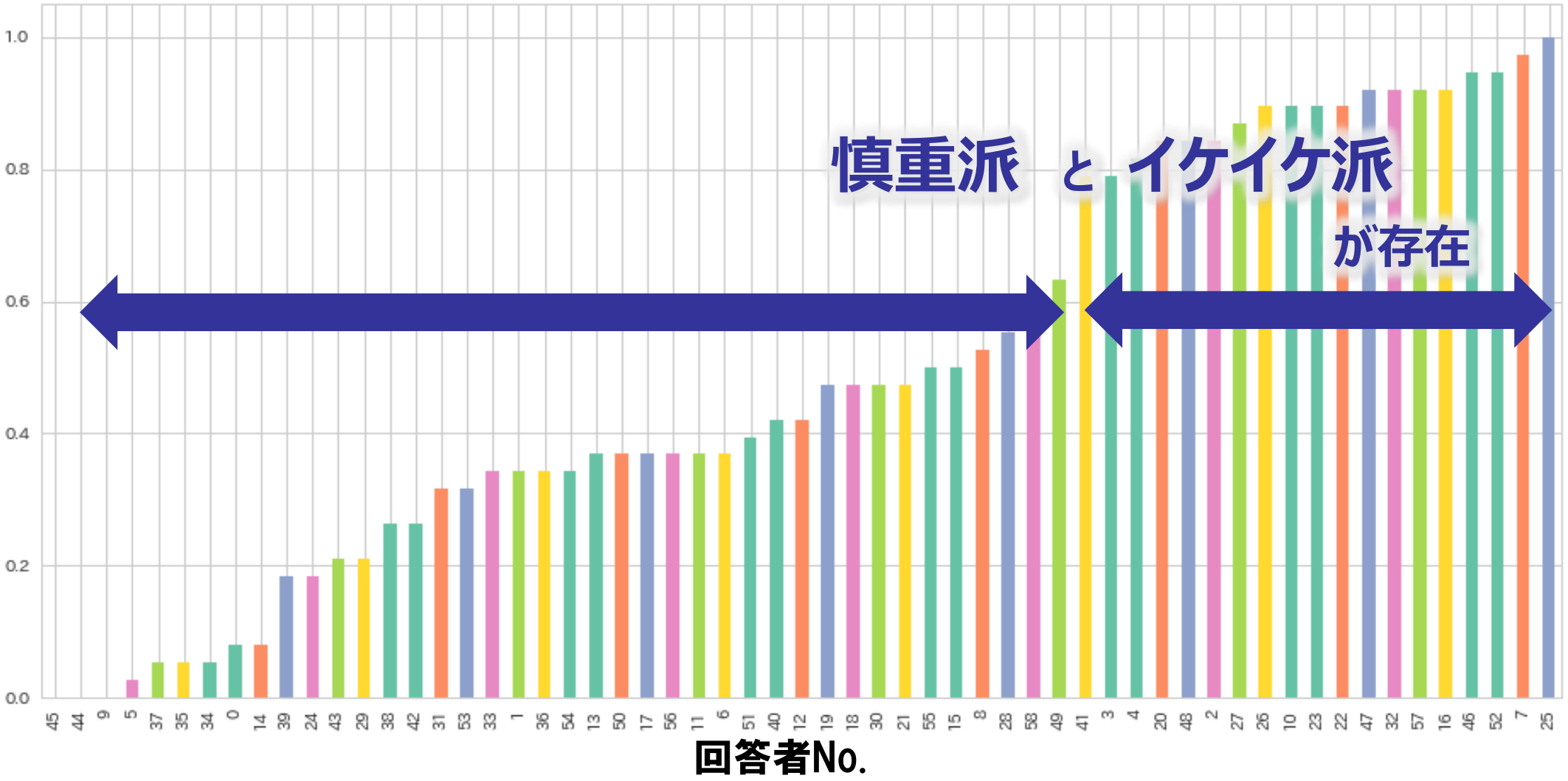
(どちらかといえば、)採用される側の立場

Copyright (C) 2019 JUAS All rights reserved

2-4　アンケート回答におけるインサイト

Copyright (C) 2019 JUAS All rights reserved

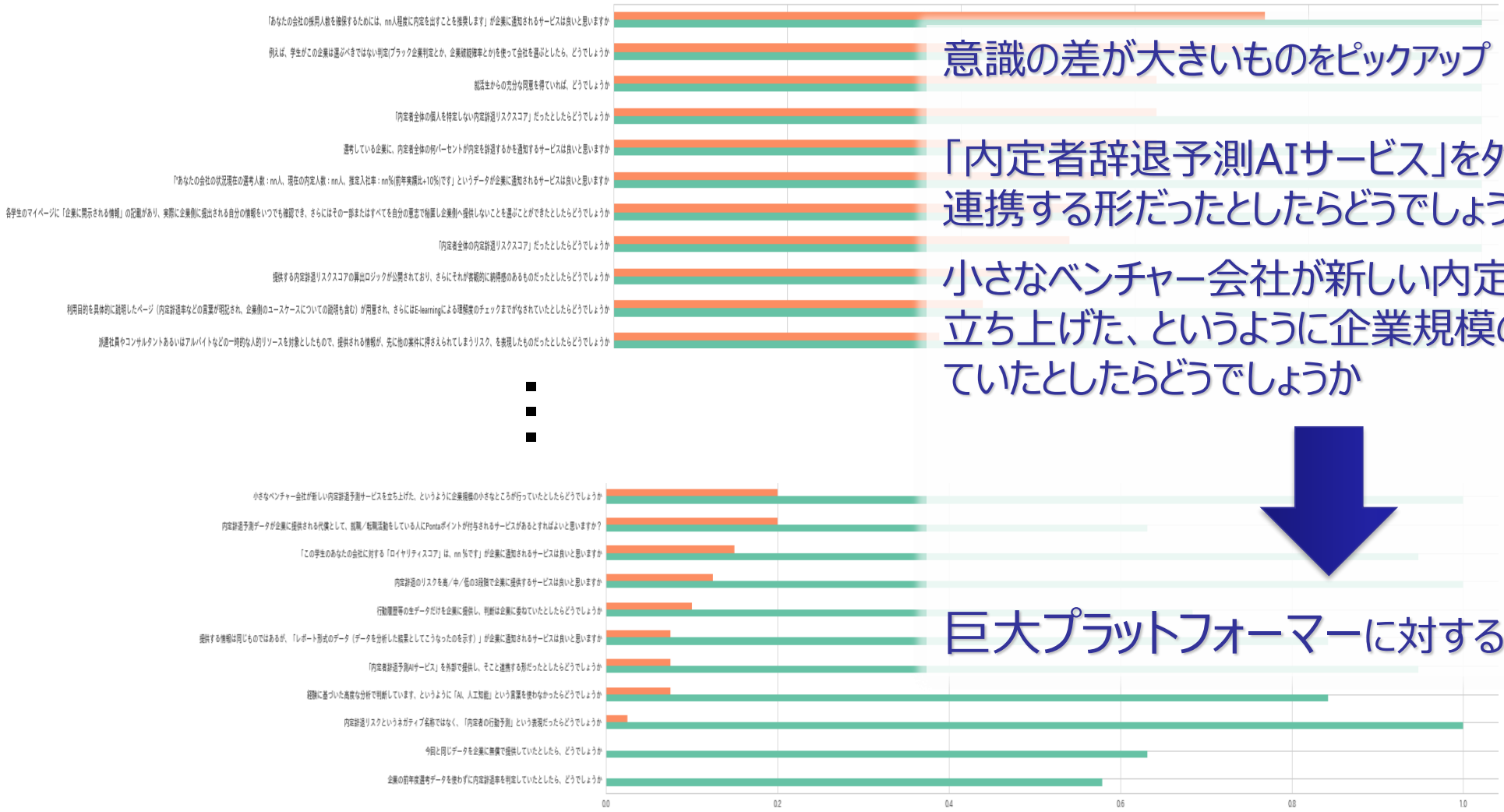
回答者毎のポジティブな回答の割合



Copyright (C) 2019 JUAS All rights reserved

慎重派とイケイケ派のポジティブ回答割合

(慎重派, イケイケ派)



意識の差が大きいものをピックアップ

「内定者辞退予測AIサービス」を外部で提供し、そこと連携する形だったとしたらどうでしょうか

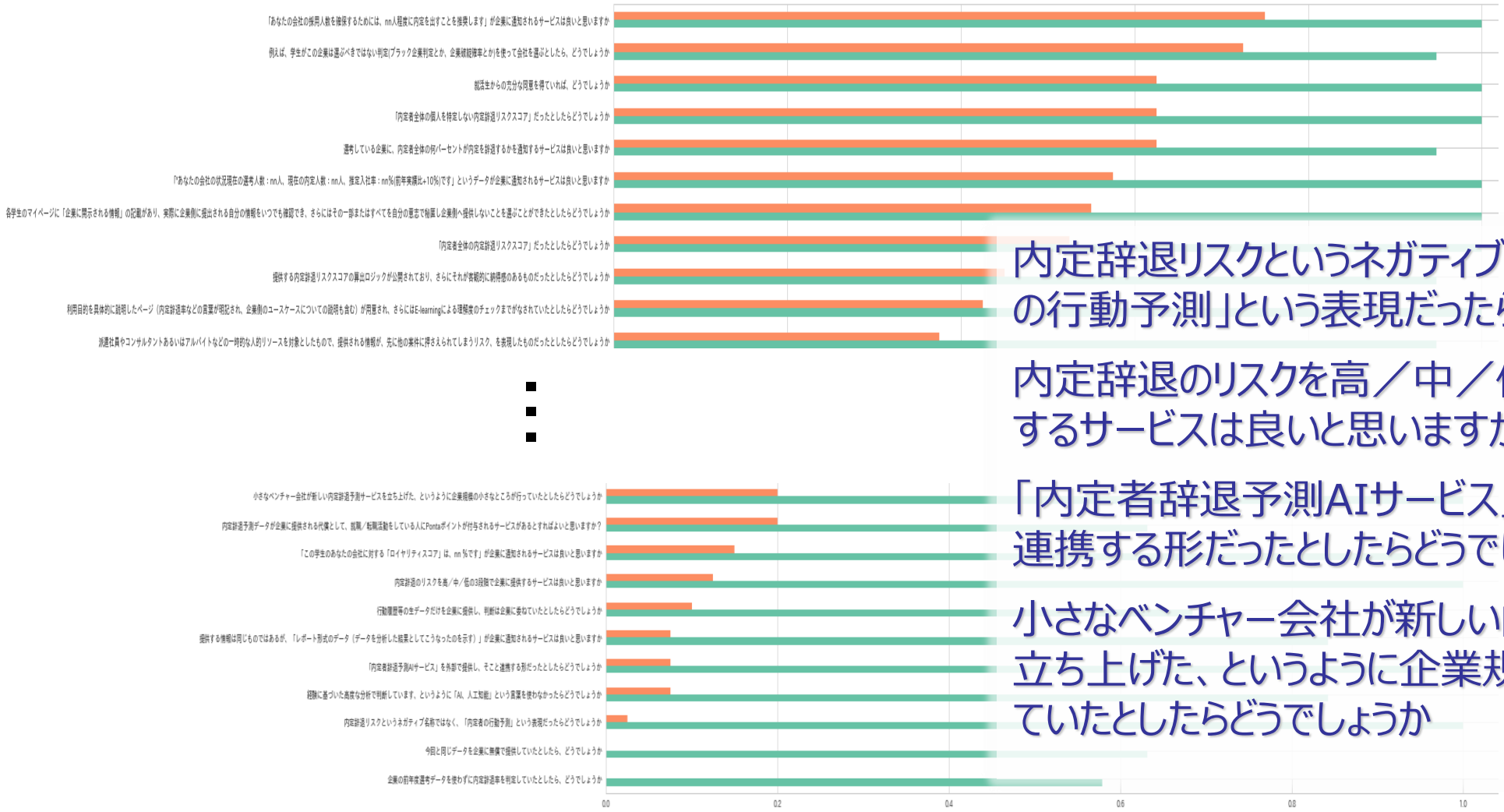
小さなベンチャー会社が新しい内定辞退予測サービスを立ち上げた、というように企業規模の小さなところが行っていたとしたらどうでしょうか

巨大プラットフォーマーに対する警戒

Copyright (C) 2019 JUAS All rights reserved

慎重派とイケイケ派のポジティブ回答割合

(慎重派, イケイケ派)



内定辞退リスクというネガティブ名称ではなく、「内定者の行動予測」という表現だったらどうでしょうか

内定辞退のリスクを高／中／低の3段階で企業に提供するサービスは良いと思いますか

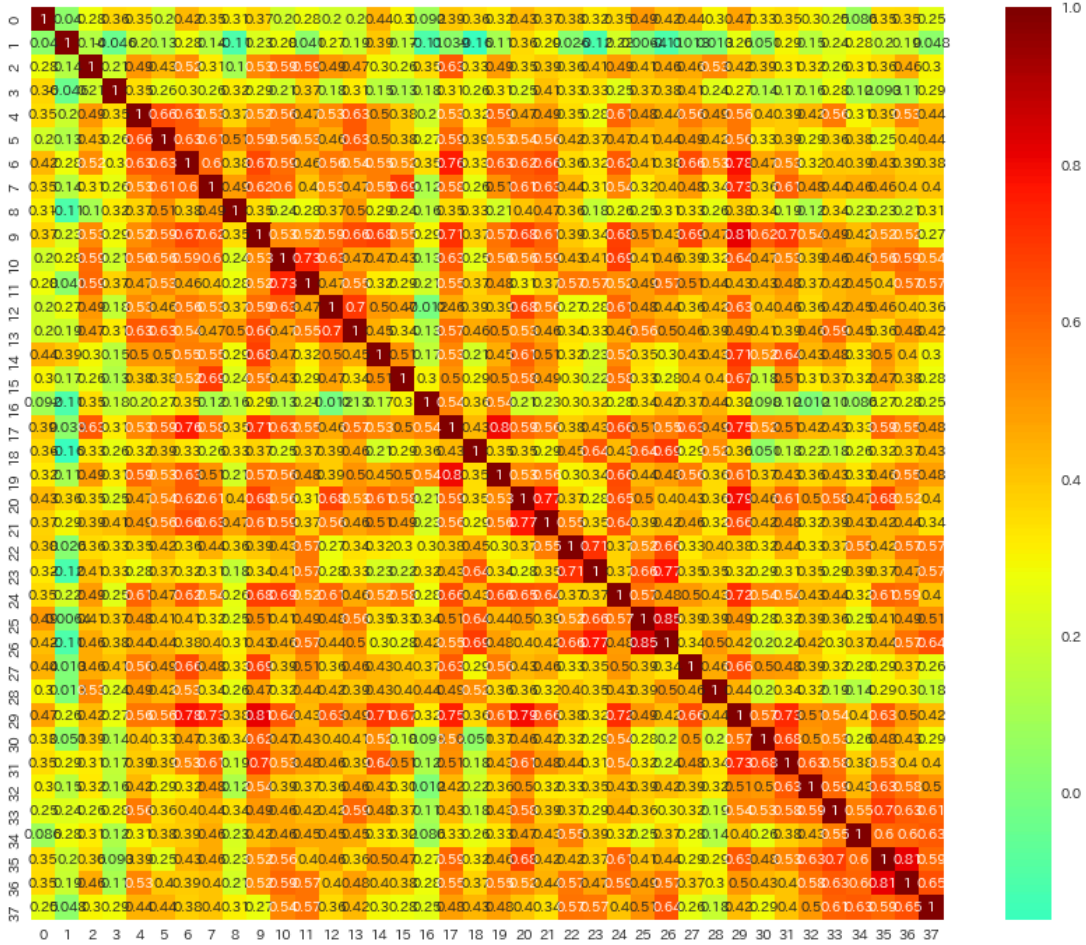
「内定者辞退予測AIサービス」を外部で提供し、そこと連携する形だったとしたらどうでしょうか

小さなベンチャー会社が新しい内定辞退予測サービスを立ち上げた、というように企業規模の小さなところが行っていたとしたらどうでしょうか

Copyright (C) 2019 JUAS All rights reserved

強者による搾取への拒否反応

質問間の相関行列



相関係数が高いものをピックアップ①

例えば、学生がこの企業は選ぶべきではない判定(ブラック企業判定とか、企業破綻確率とか)を使って会社を選ぶとしたら、どうでしょうか

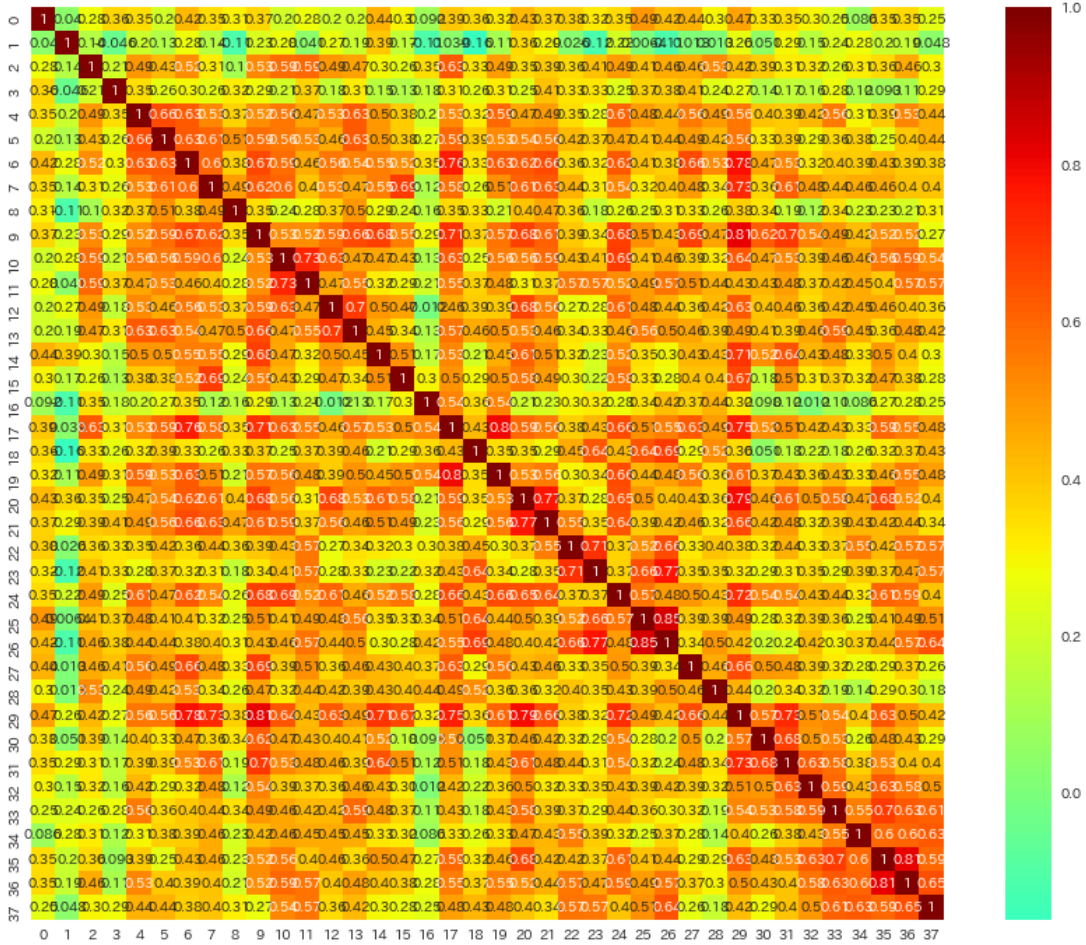
小さなベンチャー会社が新しい内定辞退予測サービスを立ち上げた、というように企業規模の小さなところが行っていたとしたらどうでしょうか

強者による搾取への拒否反応

Copyright (C) 2019 JUAS All rights reserved

影響の大きさを問題視

質問間の相関行列



相関係数が高いものをピックアップ②

リクルート社*AI*部署の新卒採用には内定辞退率を使っています、というように公共性の小さな市場で限定的に使われていたとしたらどうでしょうか

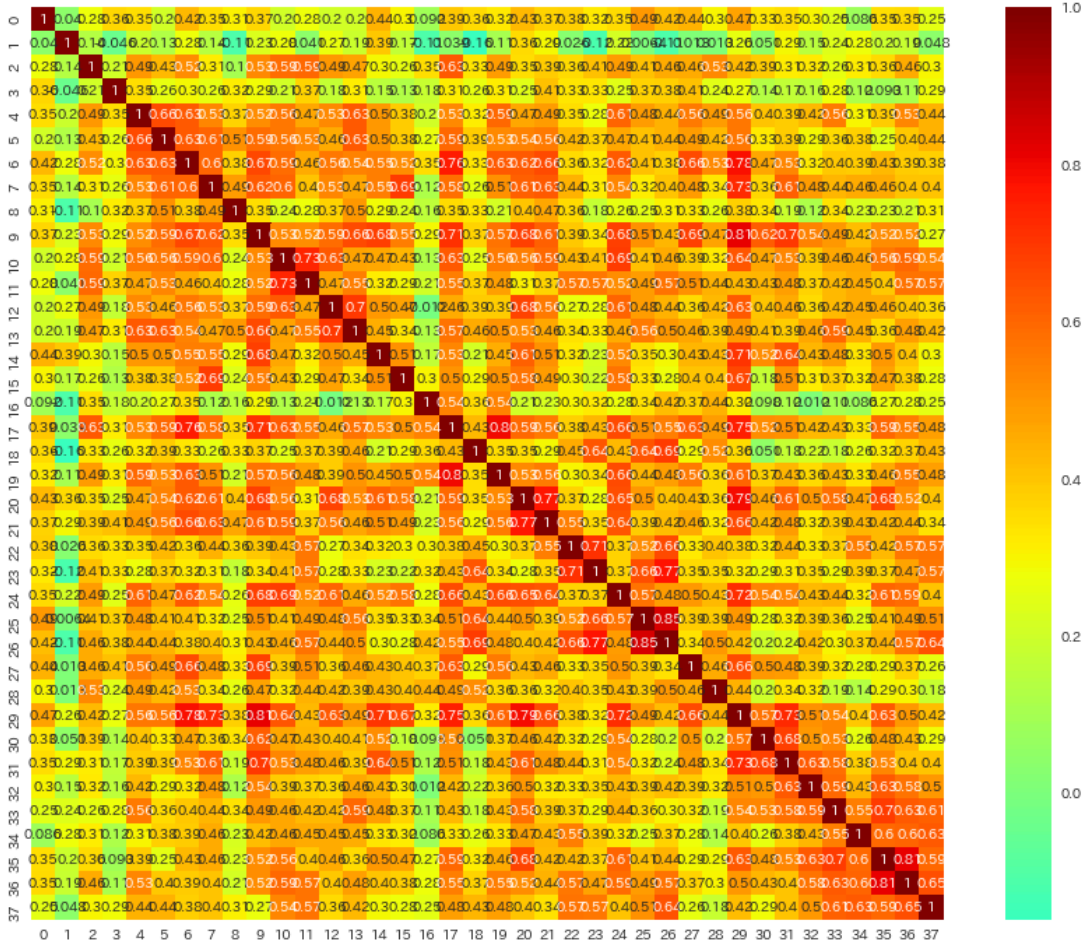
マクドナルドのアルバイトの採用可否に使っています、というように人生を左右しないようなところで使われていたとしたらどうでしょうか

影響の大きさを問題視

Copyright (C) 2019 JUAS All rights reserved

個人の利害を問題視する傾向

質問間の相関行列



相関係数が低いものをピックアップ②

就職／転職活動をしている人にPontaポイントが付与されるサービスがあるとすればよいと思いますか？

個人の人生にネガティブな影響を与える可能性が生じる時点で、社会に受け入れられない、と思いますか *(*1)*

「AI、人工知能」という言葉を使わなかったらどうでしょうか

↓

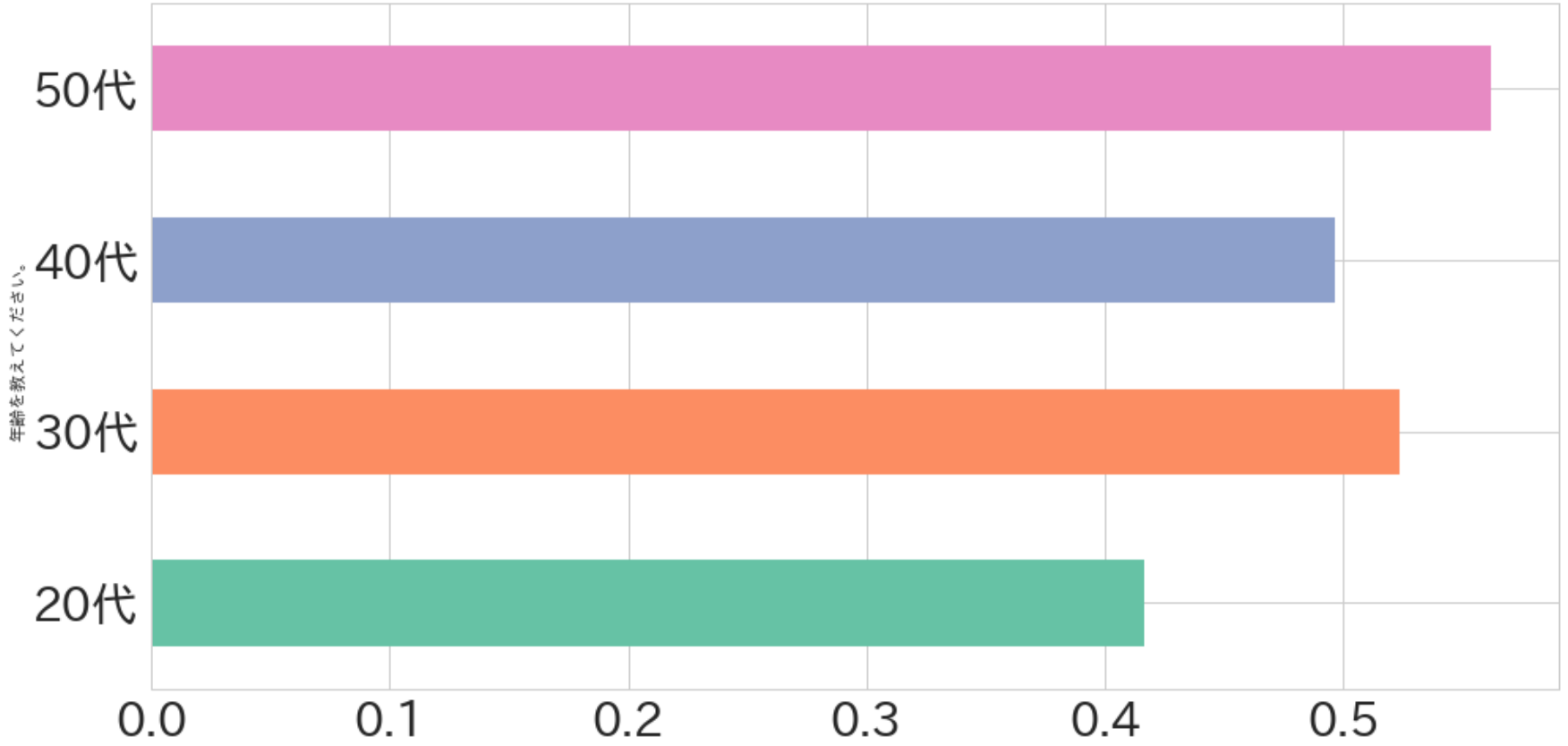
個人の利害を問題視する傾向

(*1) はい/いいえを逆にして集計

Copyright (C) 2019 JUAS All rights reserved

若い人がネガティブに捉える傾向

年齢によるポジティブ回答割合

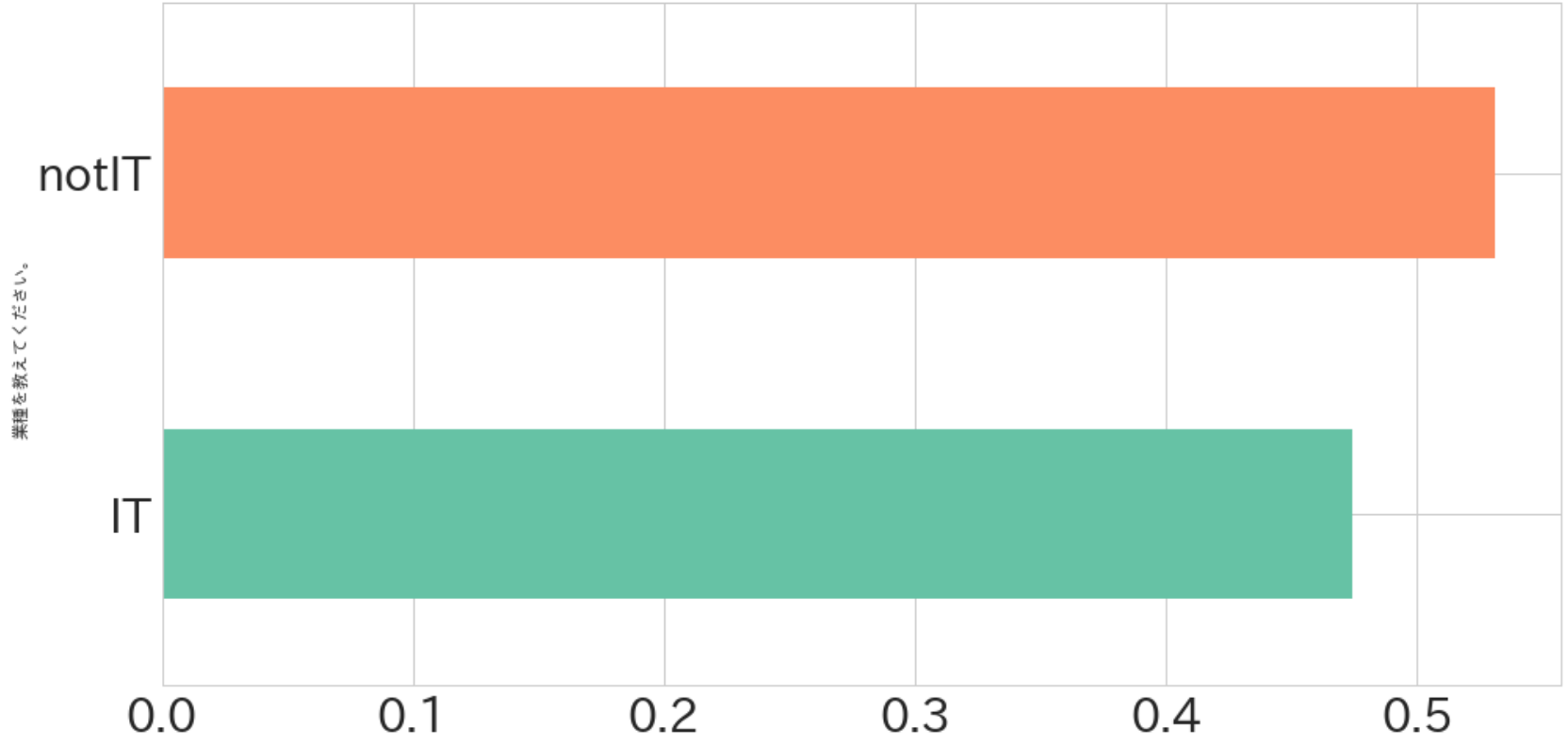


若い人がネガティブに捉える傾向がある

Copyright (C) 2019 JUAS All rights reserved

IT系の業種がネガティブに捉える傾向

業種によるポジティブ回答割合

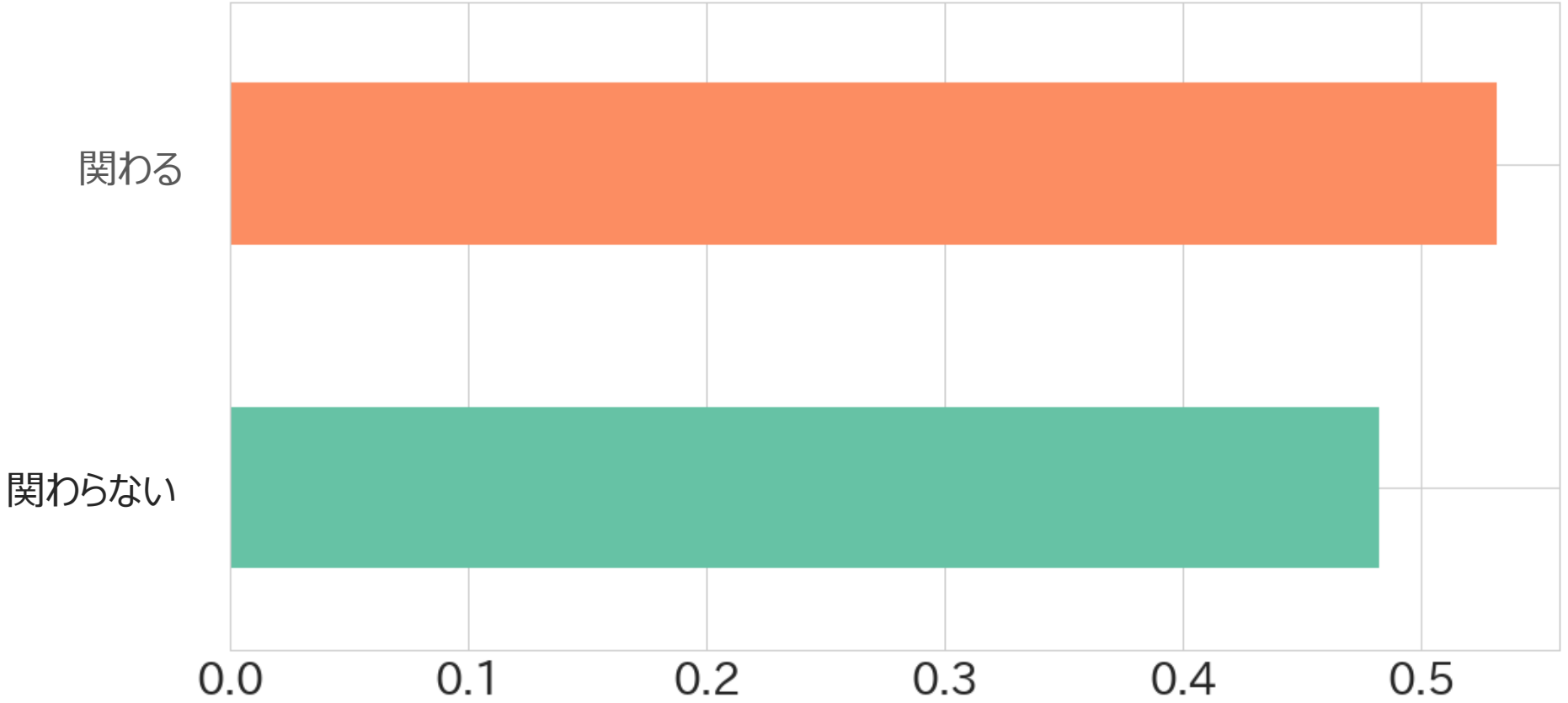


IT系の業種がネガティブに捉える傾向がある

Copyright (C) 2019 JUAS All rights reserved

採用活動に関わらない人がネガティブに捉える傾向

採用活動に関与するか否かによるポジティブ回答割合



採用活動に関わらない人がネガティブに捉える傾向がある

Copyright (C) 2019 JUAS All rights reserved

インサイトのまとめ

- 慎重派がいる
- データ所有そのものや手法はあまり問題ではなく、目的と個人の利害が大事
- 強者による搾取と思われると拒否反応がでる
- 意外に若い人がネガティブ

Copyright (C) 2019 JUAS All rights reserved

２－５　データ活用検討時のチェックリスト

Copyright (C) 2019 JUAS All rights reserved

チェックリスト

- ✓ 検討の幅：慎重派を含んだチーム
  - ✓ 多様な立場・来歴のメンバーを含む多様性のあるチーム構成か
  - ✓ 常に満場一致の全員賛成ではなく意見の対立が存在したか

- ✓ 搾取構造：Evilな搾取者イメージの回避
  - ✓ プラットフォーマーとして認識されているか
  - ✓ ユーザが自社よりも小さいあるいは立ち場が弱い取引先ではないか
  - ✓ 個人ユーザの情報を自社と顧客の法人同士の取引に使う構造になっていないか

- ✓ 利用者利害：当事者のメリットリスクの把握
  - ✓ 就職・教育・医療などの個人ユーザの人生を左右するデータを提供する仕組みになっていないか
  - ✓ ユーザにとって自分が与えた情報とサービスから提供される情報の対応関係が理解できる形になっているか

Copyright (C) 2019 JUAS All rights reserved

3.契約編

Copyright (C) 2019 JUAS All rights reserved

3-1　ベンダーとユーザー企業の対立構造・関心の違い

Copyright (C) 2019 JUAS All rights reserved

背景

- AI開発は、従来のSI開発とは開発形態が異なり、契約形態も異なる。
  ⇒　データからプログラムを作る、性能保証が難しい、等

- これに対し経産省から「AI・データの利用に関する契約ガイドライン」（2018年6月1.0版、2019年12月1.1版）が策定された。
  ⇒　具体的な契約内容は、当事者間で協議　⇒　ではどうすればよいのか？

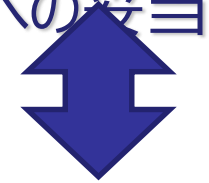
- 日本ディープラーニング協会から「ディープラーニング契約標準契約書」（2019年9月）が公開された。
  ⇒「リーガル対応力の低いAIベンチャーを支援する」という趣旨で作成されたものであるため、デフォルトは受注側優位の内容になっている

Copyright (C) 2019 JUAS All rights reserved

課題設定

- JUAS参加企業は、基本的に発注側である。
- 発注側のユーザ企業は、AI開発の契約に関して、以下の相反するリスクを抱えている。

受注側に過度に優位な契約（ベンダの言うなりなど）で、
自社の利益（競争優位、成果への妥当な権利等）を損なう



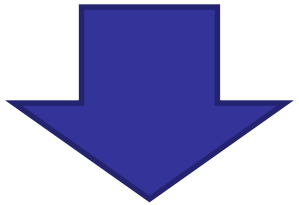
発注側に過度に優位な契約（ SIの延長など）で、
なんらかの成果質（※）を損なう

※　品質（ベンダ側意欲喪失など）、コスト（費用増）、期間（機械損失による遅延）などを想定

Copyright (C) 2019 JUAS All rights reserved

課題設定

では適切な落としどころを考える上で具体的にどういう観点があるのか？



分科会に、発注側（ユーザ企業）と受注側（AIベンダ）のメンバがいたので、
ディスカッションした結果を整理した。

Copyright (C) 2019 JUAS All rights reserved

ディスカッションで出た仮説

- ケースバイケースであるが、AI活用は、実施目的がおおむね「攻めのIT」であり、競争優位獲得に向いている。
  - 著しい優位性のあるサービスの開発
  - 抜本的な業務効率化
  - etc

- ゆえに、契約でもめる要素の一つは、通常のSIと違って、「同業他社に模倣されることの警戒心が強いこと」なのでないか

従来のシステム開発「守りのIT」

自分たちの業務を洗練するもの
⇒「自分たちのやり方」がベースなので、他社が模倣しにくい

AI開発「攻めのIT」

今までの業務を刷新するもの
⇒過去を捨てて新しく始めるものなので、他社が模倣しやすい

Copyright (C) 2019 JUAS All rights reserved

ディスカッションでの議論

- データは、発注側のものである。それは異論ない

- そこから実用レベルのモデルを作るには、ノウハウがある
  「データ渡すからやってごらん」で誰がやっても精度が出るものではない。そこには創造的側面がある。それゆえに経験に大きな価値がある

- しかしノウハウ（ナレッジ）は、一般的な知的財産保護を超えて保護する枠組みはない
  アメリカをGDPRで牽制するEUが、データと違って、ナレッジにはかけられていない

次ページから、図を使って整理

Copyright (C) 2019 JUAS All rights reserved

対立構造を考えるための図

【凡例】
青枠：ユーザ企業の既有・作成
緑枠：AIベンダーの既有・作成
青緑枠：曖昧

AI
ベンダ

獲得した領域知識

獲得したノウハウ

・KPI、前処理方法、特徴量設計、モデリング、検証方法などのAI開発の要点
・ビジネスアイディアの実現レベルの具体像とその実現性の知識
etc

既有
ノウハウ

領域知識

AI開発

データ前処理
特徴量設計

学習

検証

ユーザ
企業

ﾋﾞｼﾞﾈｽｱｲﾃﾞｨｱ
（構想レベル）

生データ

インプット

前処理PG

学習データ

学習PG
（アルゴリズム、パラメータ）

モデル

アウトプット

検証結果
（レポート）

プロダクト
開発

（省略）

プロダクト

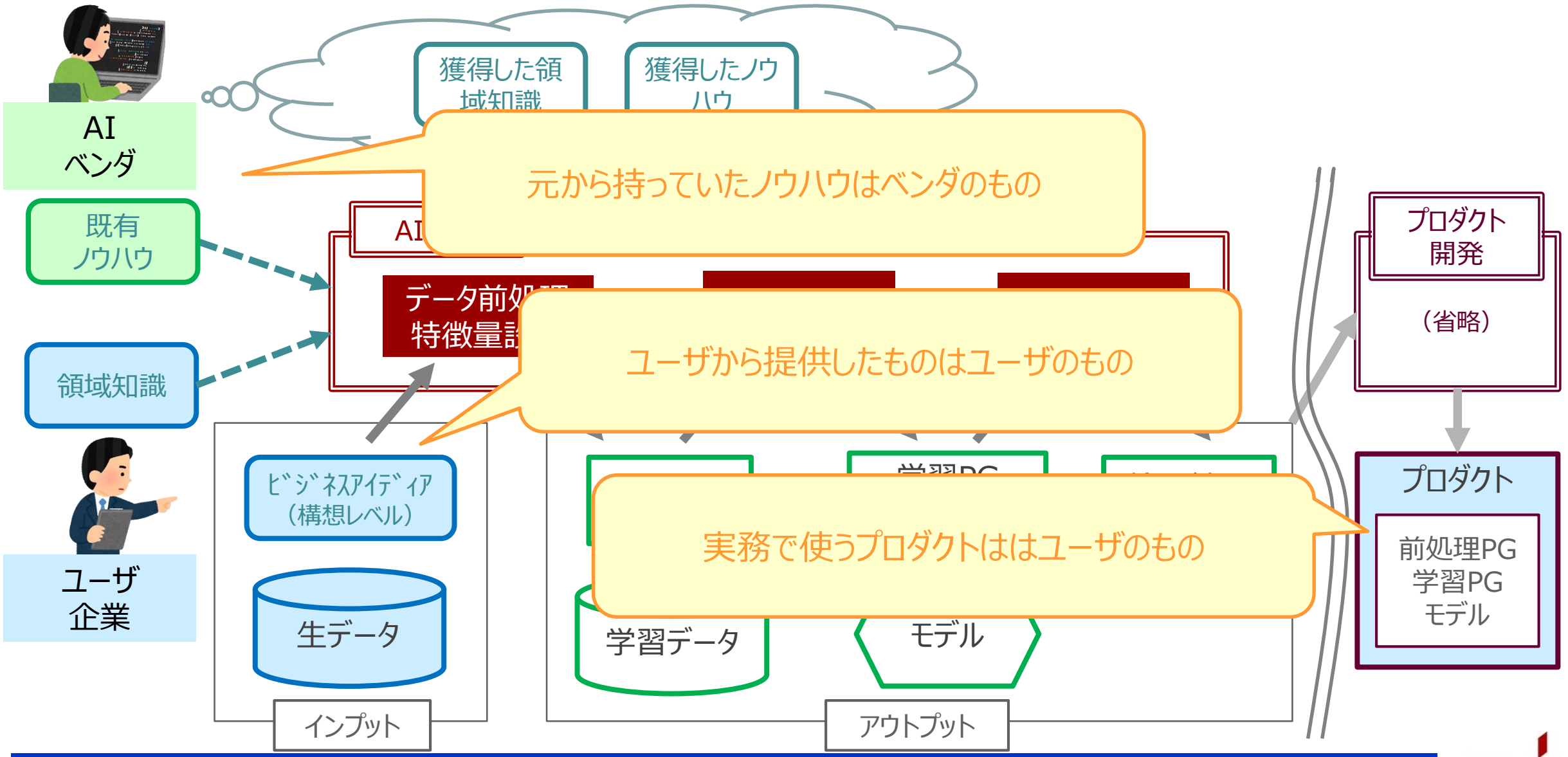
前処理PG
学習PG
モデル

AI開発の
アウトプット
が組込まれる

Copyright (C) 2019 JUAS All rights reserved

JUAS

まず、文句の出ないところ



Copyright (C) 2019 JUAS All rights reserved

ユーザ側の想い

AI
ベンダ

獲得した領域知識

獲得したノウハウ

「ユーザのもの」ではないが、縛りたい
∵競合他社のビジネス課題を解決する知見に他ならない。自社の苦労・コストの末の成果を、競合に楽に利用されてはたまらない

既有
ノウハウ

領域知識

ユーザのもの
∵ユーザのデータ（とアイディアと領域知識）があってこそ、生まれたものだから
∵自社の費用負担で生まれたものだから
→ 創造性？よくわかんないけど、お金払って作ってもらったものを受取るだけでは

プロダクト
開発

（省略）

ユーザ
企業

ﾋﾞｼﾞﾈｽｱｲﾃﾞｨｱ
（構想レベル）

生データ

インプット

前処理PG

学習データ

学習PG
（アルゴリズム、パラメータ）

モデル

検証結果
（レポート）

アウトプット

プロダクト

前処理PG
学習PG
モデル

Copyright (C) 2019 JUAS All rights reserved

ベンダ側の想い

AI
ベンダ

既有
ノウハウ

領域知識

ユーザ
企業

獲得した領域知識

獲得したノウハウ

ベンダのもの
∵これこそが欲しい物
横展開して、新たなビジネスにつなげたい

・学習データ、検証結果レポートは、ユーザでいい　∵ 生データから作ったので
・検証結果のレポートは、ユーザでいい。∵書き物なので
・前処理PG・学習PG・モデルは、著作権をベンダで持って使用許諾するのが理想。せめて共有にして、自分たちの横展開で自由に使いたい
∵ベンダの創造力があってこそ、生まれたものだから（データがあれば誰でも、ではない）

ﾋﾞｼﾞﾈｽｱｲﾃﾞｨｱ
（構想レベル）

生データ

インプット

前処理PG

学習データ

学習PG
（アルゴリズム、パラメータ）

モデル

アウトプット

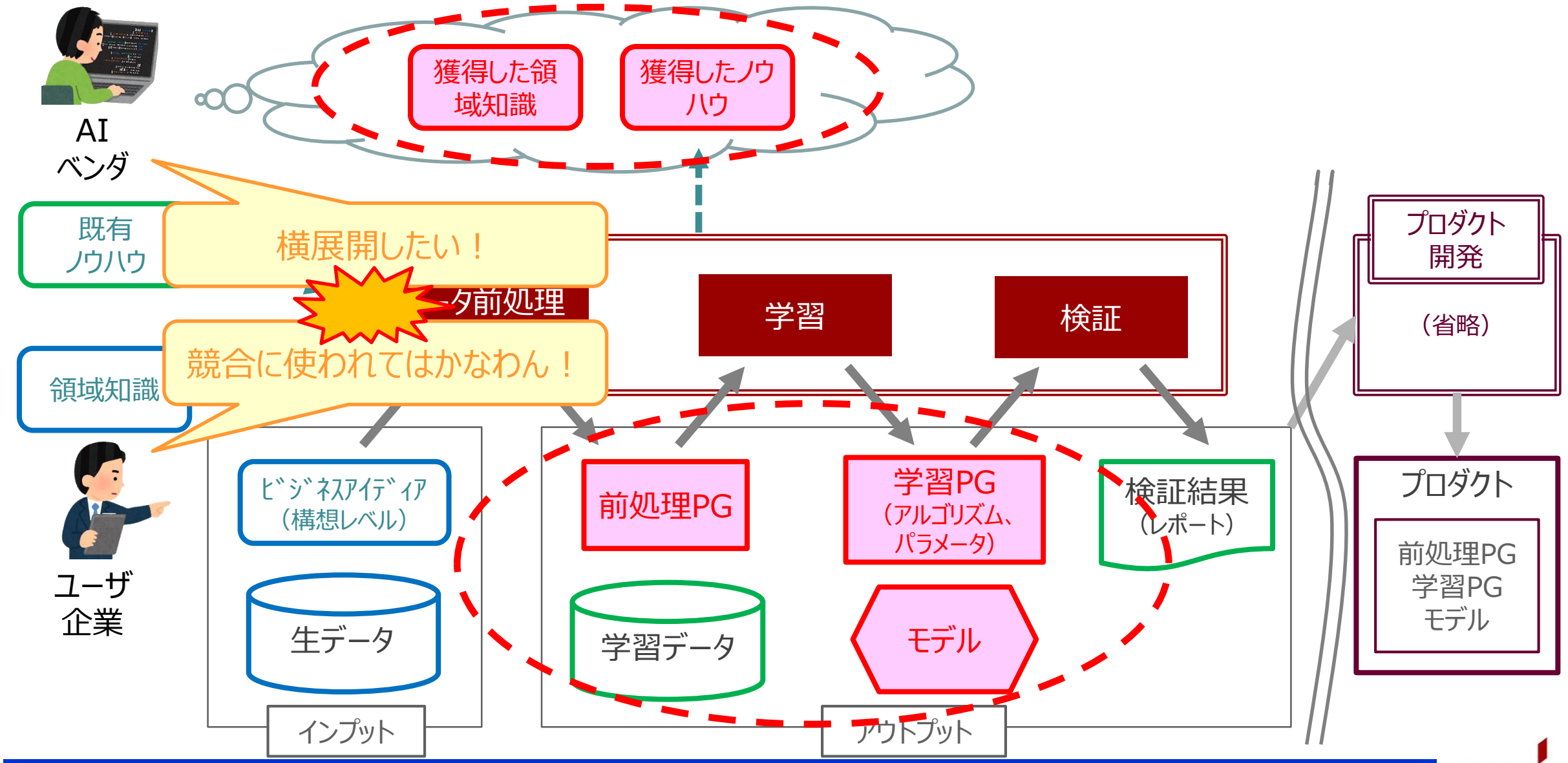
検証結果
（レポート）

プロダクト

前処理PG
学習PG
モデル

Copyright (C) 2019 JUAS All rights reserved

明らかになった対立要素



Copyright (C) 2019 JUAS All rights reserved

契約の記述内容として考えた場合

AI
ベンダ

獲得した領域知識

獲得したノウハウ

契約上の扱いが難しいのはこちら

下側は実体がある＝知財なので、原則として、
・そもそも納品物に含めるか
・権利移転するか
を契約に明記すればよいもの。（権利主張の調整はいるが）
※ 再学習の扱いが気になるが、今回は検討外

既有ノウハウ

領域知識

ユーザ
企業

ﾋﾞｼﾞﾈｽｱｲﾃﾞｨｱ（構想レベル）

生データ

インプット

前処理PG 知財

学習PG（アルゴリズム、パラメータ） 知財

検証結果（レポート） 知財

学習データ

モデル 知財

アウトプット

プロダクト開発

（省略）

プロダクト

前処理PG
学習PG
モデル

Copyright (C) 2019 JUAS All rights reserved

3-2　ベンダー側のコスト・収益構造・ビジネスモデル的な観察

Copyright (C) 2019 JUAS All rights reserved

契約の落としどころの作り方

最重要ポイントは「ベンダーの損益」

ベンダの見積には、「投資・リスクテイク」の箇所と、「工数・コストベース」の箇所が混在している

投資になりやすい項目

・成果物への契約(役務の量が関係ない)

・AI・機械学習モデルの開発

・共通SaaS化できる範囲の業務機能

工数ベースになりやすい項目

・物理的な役務費用
(サーバコスト、IoTセンサー施工費など)

・発注企業に特化したUI

スタートアップ段階の企業であるほど、投資型で見積提示を行う
大手企業は開発の受託という形で契約が可能
⇒大手の見積もりが高い理由の一因(多重請負は論外)

Copyright (C) 2019 JUAS All rights reserved

３－３　契約上の落とし所、かわし方

Copyright (C) 2019 JUAS All rights reserved

実際に着地させたケース

5年十数億のビジネスを取りに行くために、データ分析と業務手法開発を最小限の価格で先行投資して行った。

発注側の要求を通した点

・非知的財産の範囲まで流用に制約をかける

・同業他社を流用NGリストに盛り込む

受託側の要求を通した点

・NGリストはブラックリスト形式とする

・事業会社単位で最低限の会社のみを指定する
(今回は3大コンビニの残り2社のみ)

・成果が出た場合にしくみ(システム)の展開を
ベンダ指名で行う

3~5年で数億～十数億のPJ＆十分な粗利

以下の点で、過去に広告業界で行われてきた「1業種1社」のルールに比較的近い
・1社へのビジネスで事業が成立する取引規模である
・多業種への展開への制約はかけない
・業界下位の企業への展開に縛りを設けない

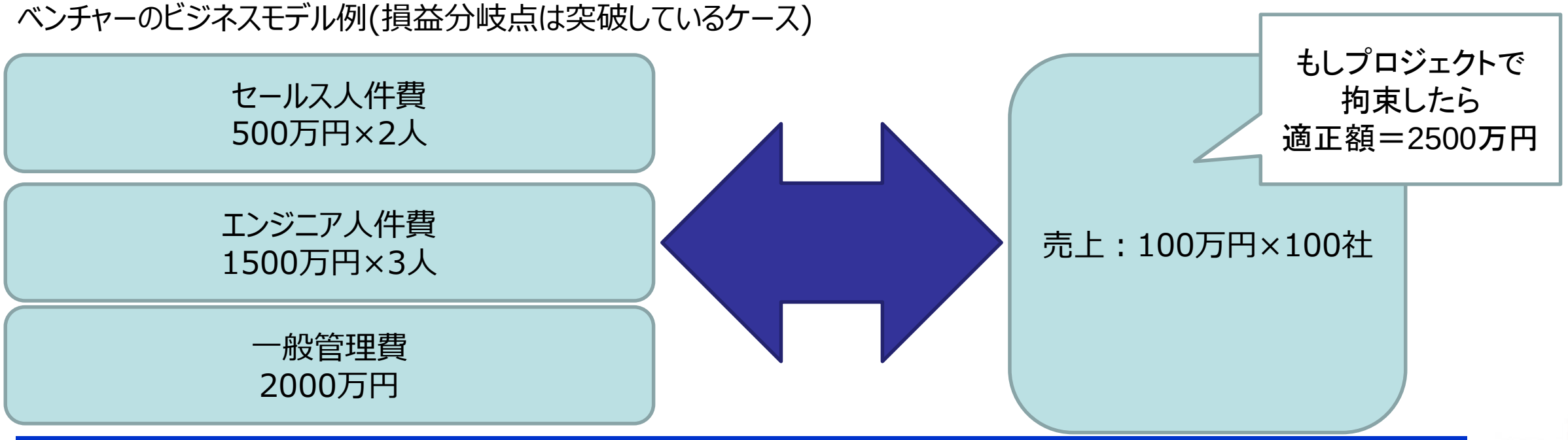
Copyright (C) 2019 JUAS All rights reserved

何故ベンチャーは投資型でビジネスを行うか

競合より安い価格で、回転の速いビジネスを行う必要があるから

・同価格、同等仕様の場合、顧客は大手企業を選択する
・資金繰りの体力が低い為、売上回収が早くなるよう商談を獲得しなくてはならない。
・ビジネスサイクルは商談～受注までの期間(原則売上げゼロ)も含まれるため、最短で営業をクローズできるよう
　標準機能を横展開するモデルを選択する
　⇒一つのプロジェクトにエンジニアを拘束されるとアウト。セールスの人件費が無駄になる

ベンチャーのビジネスモデル例(損益分岐点は突破しているケース)



Copyright (C) 2019 JUAS All rights reserved

クライアント1社で回収できる契約条件が必要

・市場価値の高いエンジニアのリソースを一定期間占有するのに妥当な金額での契約

・自社に対する売上のみで投資回収が成立するような、付帯する取引の実施
例：AIの開発を投資で行う代わりに、AIで使用するデータを取得するためのセンサーを大量に購入する

知見の流用を制限する場合は「最小限のブラックリスト方式」

⇒市場全体の成長性がエンジニアの質、量の充実に影響するため、委託先及び業界の成長を考慮した
ビジネスモデルが重要。

Copyright (C) 2019 JUAS All rights reserved

3-4　ベンダー企業とユーザー企業それぞれの関心事の考察から得られるインサイト

Copyright (C) 2019 JUAS All rights reserved

何が難しくしているのか？何が不安なのか？

データでビジネスに貢献するためにAI開発に取り組む。

「そもそもできるのか？」という不安がある？
できたとしてもどれだけの効果があるのか？
そこにどれだけの投資をするのが妥当かが分からない

ベンダーの作業内容、技術的な難易度が分からない。
普通のSI開発と何が違うのか？
作業範囲・成果物があいまい。

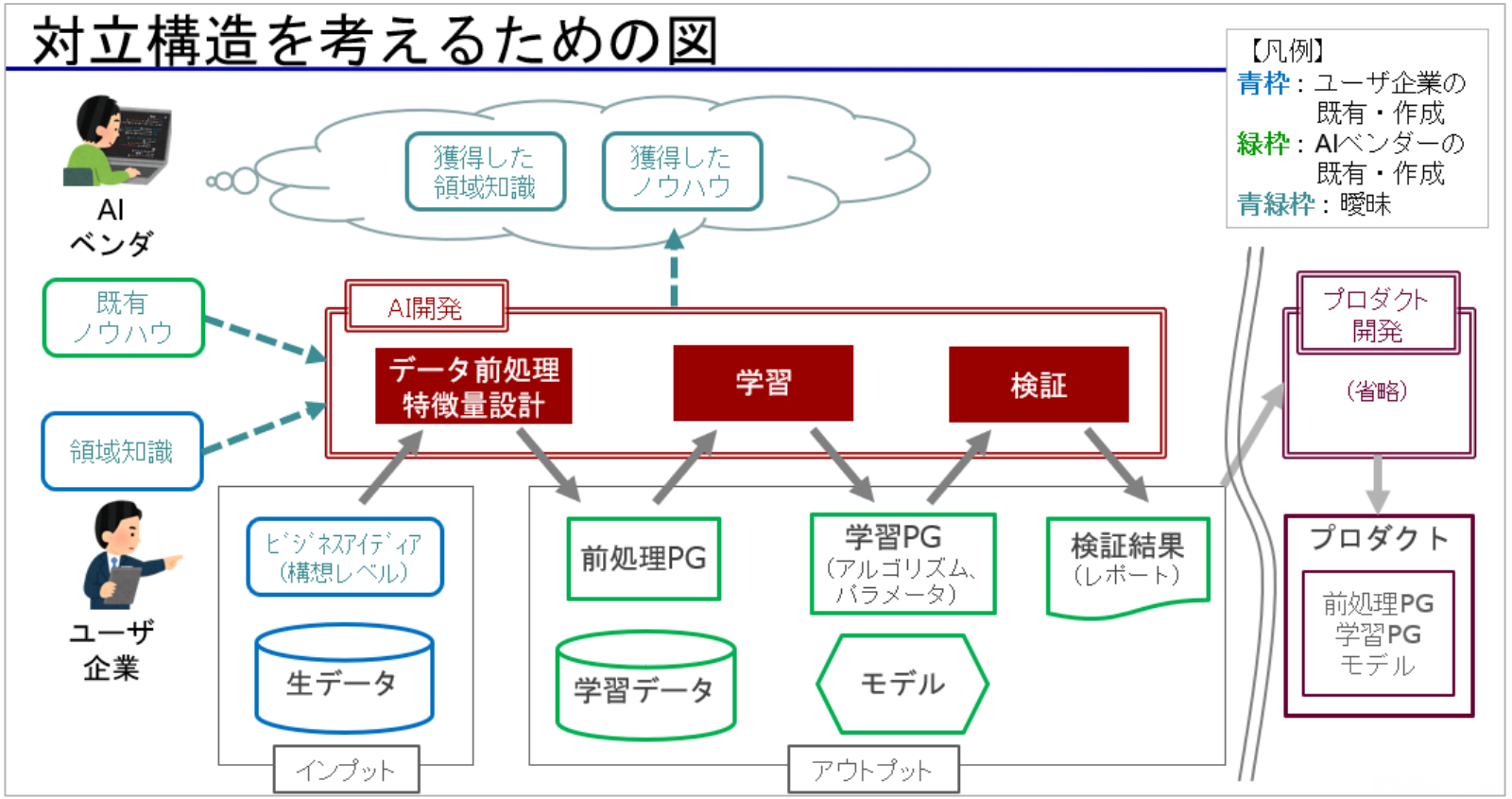
我々(主に情シス)にとっては未知の領域だが、
製品開発・サービス開発部門や法務部門は経験済みでは？

費用対効果　**知的財産権**　**瑕疵担保**　**NDA**　**契約形態**

Copyright (C) 2019 JUAS All rights reserved

AI開発の契約前段階で最初にやることは？

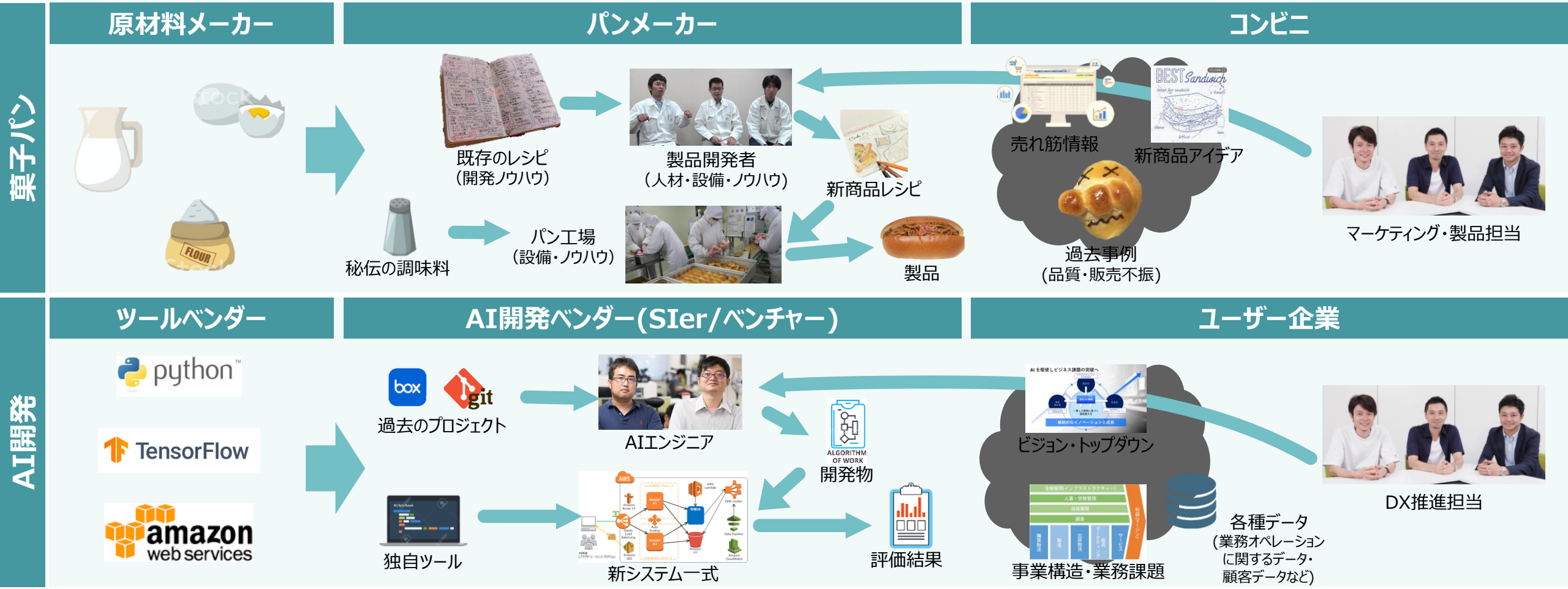
製品企画部門、製品開発部門、法務部門に相談してみる。
難しい場合は、下の図や、わかり易い例に当てはめてみて、
何を得らえるのか、何を失わないようにすべきか、その対応策を明確化する。



Copyright (C) 2019 JUAS All rights reserved

製品開発・法務部門がない場合の検討例

- 例えば、コンビニがプライベートブランドのパンをパンメーカーに委託するときには何を気しているか？を考えてみる。



Copyright (C) 2019 JUAS All rights reserved

ベンダーは何をしているか？どこまでが我々の権利なのか？

ベンダー側の負担を想像することで、相互理解を深める。

どこまでを合理的に権利主張できるか、落としどころを探ってみる。

目的外使用不可　開発のタイミングでメーカー・ベンダーに提供する コンビニ・ユーザー企業側に帰属する機密で特に注意が必要なもの

メーカー　メーカー側に帰属させるべきもの

コンビニ ユーザ企業　コンビニ・ユーザー企業側に帰属させるべきもの

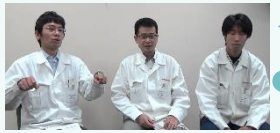
どっちもアリ　販売・運用後の運営ノウハウ 有無等により変わるもの

菓子パン

原材料メーカー

パンメーカー

既存のレシピ（開発ノウハウ）



製品開発者（人材・設備・ノウハウ）

秘伝の調味料

パン工場（設備・ノウハウ）

新商品レシピ

製品

コンビニ

売れ筋情報

新商品アイデア

過去事例（品質・販売不振）

マーケティング・製品担当

AI開発

ツールベンダー

python

TensorFlow

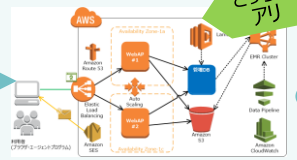
amazon web services

AI開発ベンダー(SIer/ベンチャー)

過去のプロジェクト

AIエンジニア

独自ツール



新システム一式

開発物

評価結果

ユーザー企業

ビジョン・戦略

事業構造・業務課題

各種データ（業務オペレーションに関するデータ・顧客データなど）

DX推進担当

3-5　ユーザー企業として考慮すべきポイントまとめ

Copyright (C) 2019　JUAS All rights reserved

考慮ポイントまとめ

観点	考慮すべき項目	←ができていない場合に発生する問題点
成果物	• 成果物一覧、品質が、ベンダーに期待しているものになっているか？ • ゴールイメージを共有できているか？ ※発注側は契約の範囲でゴールに到達できなかった場合のバックアッププランが必要。	• 委託期間が終了したときに、期待する成果がえられない。 • 動かないシステム、動いても意味のないシステムが納品される。
稼働後の体制等	• システム一式、ソフトウェアが納品される場合は、稼働後の保守について合意ができているか？	• 納品時点では稼働するものの、OS/ミドルなどの更新時に不具合・システム停止が生じる or 脆弱性を抱える。 • データが増えた際のパフォーマンス低下、不具合発生時の動作不良を早期に解決できない。作り直しになる。
プロジェクト管理	• 当初合意したゴールイメージ・QCDに向けたプロジェクト運営が可能な契約となっているか？ (善管注意義務、課題管理・リスク管理、変更管理、TODO進捗管理などの役割分担の明確化)	• 当初想定とは異なるシステムが出来上がったり、そもそもシステムが出来上がらない。
知的財産権	• 発注側の開示する機密情報の扱い、プロジェクト独自の成果物の扱い(自社に著作権が帰属するもの)、共同著作物として扱うものの分類が合意可能か？ • 競合排除条件の要否や範囲は合意可能か？	• 競合他社に自社のノウハウやデータが流出する。 • 独自の成果物をベンダーの他の顧客に不当に流用される。
費用妥当性	• リスク、期待効果を考慮しても受容可能なコスト化？	• 赤字案件を苦労して実行することになる。

Copyright (C) 2019 JUAS All rights reserved

4.考察・まとめ

Copyright (C) 2019 JUAS All rights reserved

4-1　考察内容

Copyright (C) 2019 JUAS All rights reserved

企画段階で、以下の2点を十分に精査することが重要

1.同調バイアスでリスクを見落としていないか

2.立場的に強者となる要素が重なりすぎていないか
採用側(労使関係)×プラットフォーマー(データ強者)

3.利用者のメリット／リスクを、提供側が正しく把握し
説明ができているか

Copyright (C) 2019 JUAS All rights reserved

発注元とベンダ双方の、ビジネスモデルへの理解が十分な状態で契約を行うことが重要

1.客観的に実現性のある成果物をゴールに設定できているか

2.知的財産権に関する要求の根拠を相互に開示する

3.ベンダの価格の根拠が、発注側の求める前提条件に合致しているか

Copyright (C) 2019 JUAS All rights reserved

4-2　まとめ、今後の展望

Copyright (C) 2019 JUAS All rights reserved

今後の展望

倫理

調査範囲を拡大し、設問に対する立場ごとの差異をより鮮明にする

契約

実際の契約締結事例をより集積し、具体的な要点を網羅する

相互の着地点につながる文章表現の精査

Copyright (C) 2019 JUAS All rights reserved