

2016年度JUAS研究活動成果報告会

— データマネジメント研究会 —
— データサイエンス研究会 —

2017年4月19日

本日のAGENDA

第一部 活動概要

1. 研究会プロフィール
2. 合宿および研究テーマ設定
3. 事例講演 & 第二合宿

第二部 研究成果発表

1. データマネジメント研究会
2. データサイエンス研究会

2016年度 データマネジメント・データサイエンス研究会活動概要

《実施期間》

2016年5月26日 ～ 2017年3月25日

《主な活動実績》

- ・定例研究会 : 各研究会 各9回
(内、3回 DM、DS合同開催)
- ・事例発表 : 6事例
- ・合宿 : 2回 (沼津合宿＋JUAS日帰り合宿)
- * その他 分科会多数

《参加人数》

データマネジメント研究会 : 38名
データサイエンス研究会 : 39名

総勢 77名！

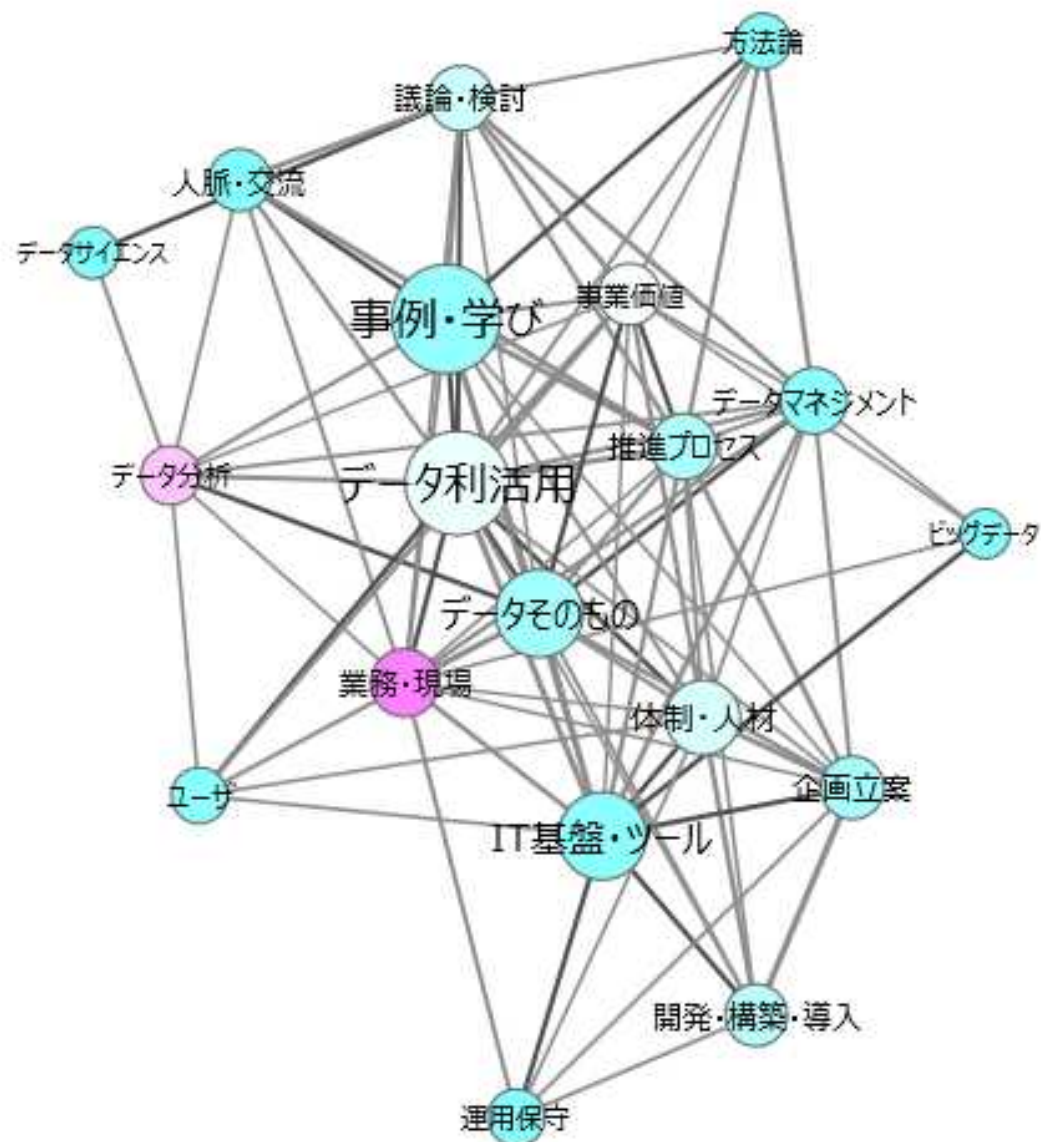
研究会のスタート地点 ―参加各社の課題認識―

データマネジメント研究会
＋
データサイエンス研究会

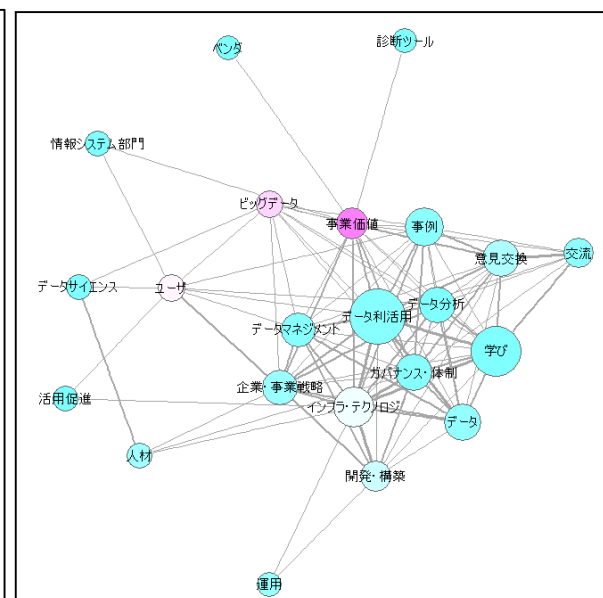
研究会参加申し込み時の
コメント欄



KHコーダーによる
テキストマイニング
を実施



【2015年度の分析結果】



研究会 活動実績

定例研究会 : DM研6回、DS研6回 合同3回 合宿 : 2回
事例発表 : 6事例

* 分科会の活動多数

日程	研究会(場所)	メインテーマ	事例共有	DM	DS	自主
5月26日 木	DM#1 定例研究会 (JUAS)	・今年度研究会の方針発表 ・参加メンバー自己紹介 ・情報交換会	なし	○		
6月2日 木	DS#1 定例研究会 (JUAS)	・今年度研究会の方針発表	なし		○	
7月22日 金	合宿(沼津)	・昨年までの成果共有	テラデータ 金井様	○	○	
7月23日 土	合宿(沼津)	・研究テーマ設定	経済産業省 平本様	○	○	
8月4日 木	DM#2 定例研究会 (NTTデータ)	個別研究会活動	なし	○		○
8月25日 木	DS#2 定例研究会 (日本テラデータ)	個別研究会活動			○	○
9月7日 水	JUAS FUTURE ASPECT	-				
9月8日 木	JUASスクエア(目黒雅叙園)	-		-	-	-
9月9日 金	JUASスクエア(目黒雅叙園)	-		-	-	-
9月15日 木	DM#3 定例研究会 (JUAS)	個別研究会活動	なし	○		
9月29日 木	DS#3 定例研究会 (JUAS)	個別研究会活動 & 分科会共有			○	
10月20日 木	DM&DS#1 定例研究会 (JUAS)	16:00～事例発表① 17:00～事例発表②	ライオン 皆川様 博報堂 酒井様	○	○	○
11月17日 木	DM#4 定例研究会 (JUAS)	個別研究会活動 & 分科会共有	なし	○		
11月24日 木	DS#4 定例研究会 (JUAS)	個別研究会活動 & 分科会共有	なし		○	
12月22日 木	DM&DS#2 定例研究会 (JUAS)	16:00～事例発表① 17:00～事例発表②	大阪学院大 喜田先生 日本航空 箕様、渋谷様	○	○	○
1月19日 木	DM#5 定例研究会 (JUAS)	個別研究会活動	なし	○		
1月26日 木	DS#5 定例研究会 (JUAS)	個別研究会活動	なし		○	
2月7日 火	DS&DM#3 交流会 (JUAS)	DS2(組織) & DM2 DS1 & DM3	なし	○	○	○
3月2日 木	DS#6 定例研究会 (JUAS)	16:00～DS研究会分科会成果発表 & 意見交換会	なし		○	
3月9日 木	DM#6 定例研究会 (JUAS)	16:00～DM研究会分科会成果発表 &	なし	○		
3月25日 土	第2合宿DM&DS (JUAS)	10:00～	なし	○	○	○
4月19日 水	JUAS研究活動成果報告会(Jフェス)	(詳細別途)				

研究会での事例発表



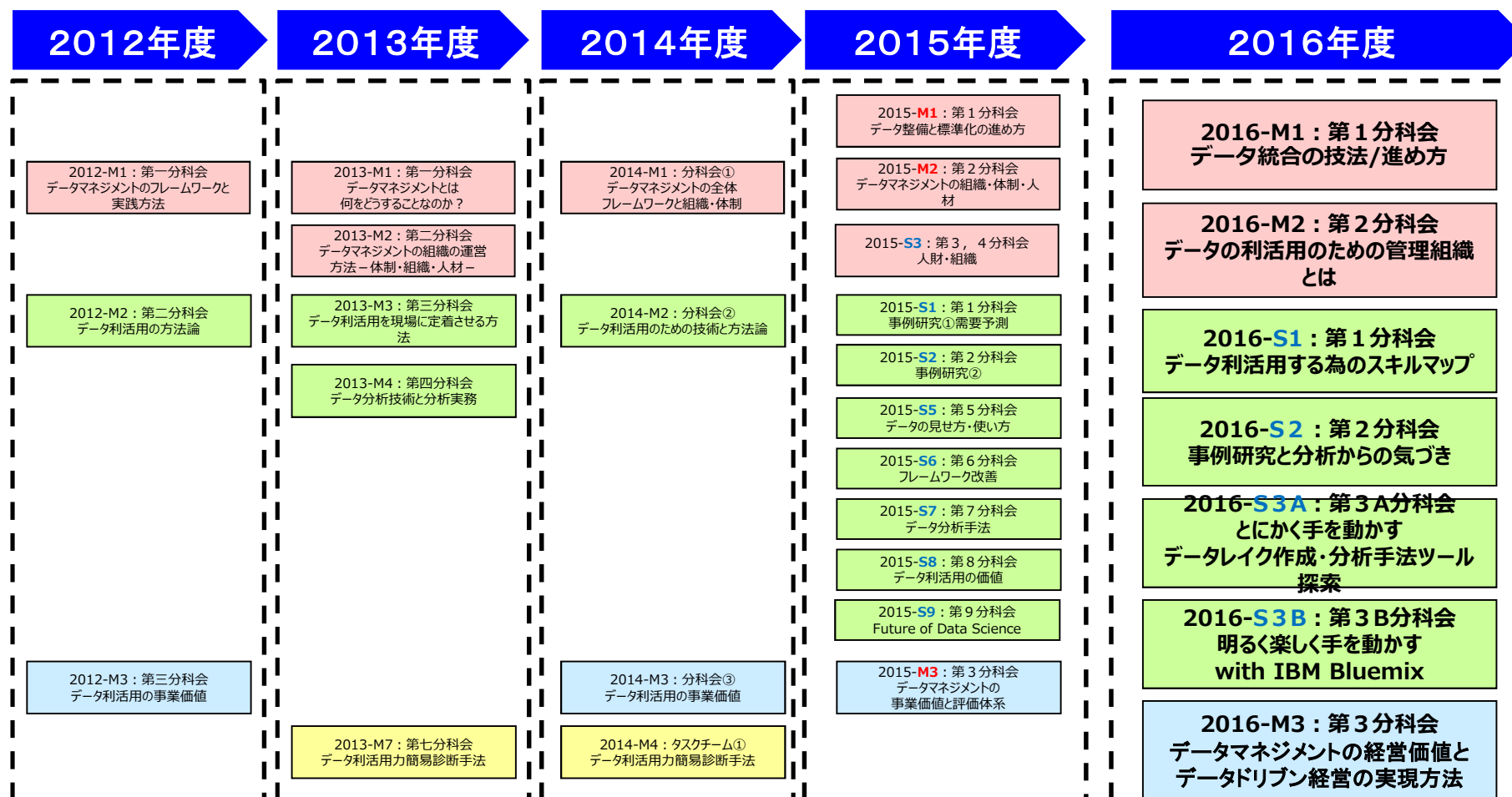
ご講演者	講演テーマ
日本テラデータ株式会社 コーポレート・エバンジェリスト／エグゼクティブ・コンサルタント 金井 啓一様	『データ活用の典型的な事例、および事例から学ぶデータマネジメントのあり方』 ～データ統合とアナリティクスの観点から～
経済産業省 CIO補佐官 平本健二様	『行政機関におけるデータマネジメントの取り組み』
ライオン株式会社 統合システム部 皆川 泉様	MDM基本のき ～データマネジメントの理想の姿～
博報堂生活総合研究所 上席研究員 酒井 崇匡様	マイビッグデータの時代へ テクノロジーによってあばかれる、誰も知らなかった「自分」
大坂学院大学 経営学部 喜田 昌樹教授	大学経営でのデータ活用 退学者問題を中心に
日本航空株式会社 渋谷 直正様 笥 朋子様	『わたしはこうしてデータサイエンティストになった』 ～日本航空のデータ活用を実践して学ぶ～

沼津合宿での研究会

7月22～23日 沼津プラザヴェルデ ふじのくに千本松フォーラムで行われた合宿で、DM研、DS研 77名が全員参加でディスカッションし、メンバーの課題認識の共有し一気に議論を深めました。

DM・DS研究会のあゆみ

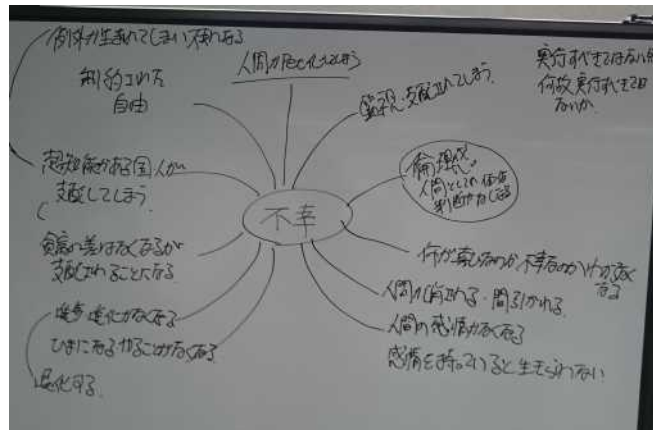
- ・2012～2015年度の研究成果を継承しつつ、2つの研究会に分かれて研究活動を推進しました。
- ・過去4年間の成果を共有し、更なる深耕・論考を進め、研究成果を成長させました。



第2合宿での研究会のしめくり

3月25日(土)、DM研 & DS研合同で日帰りの合宿を行い、お互いの研究成果を相互に発表し、議論を交わしました。

また、『データ利活用の未来予測』をテーマとしたディベートセッションや、ハッカーチームとセキュリティチームとに分かれての『シュミレーション・ウォーゲーム』など、日常業務では味わえない体験をしました。



役割2: 守備側

セラノゲニクス社は、8つの事業部から成っており、本社IT統括本部の傘下に、事業部ごとのIT部門をもち、それぞれSection-1〜Section-8と呼ばれていた。

製薬会社という性格上、臨床データを始めとする機微性の高いデータを大量に保有するセラノ社のIT統括本部の元には、事業部横断でセキュリティ管理を専任する、通称「Section-9」が存在していたが、その存在そのものが秘匿とされ、どの組織図にも明記されることはなく、セラノ社の社員ですら、一部の役員を除いてSection-9の存在は知られていない。

役割1: 攻撃側

セラノゲニクスによるイノベーションの買収は、セラノ社のマスコットにより、ありふれたM&A事業としてビジネス誌の小さな記事に取り上げられただけだった。

しかし、ネット上に書き込まれた元セラノ社員を名乗る内部告発文から、一連の事実が一部のネット民の知るところとなる。セラノ社は情報の抹消に乗り出し、「HACKON2018」が意図的に作り出したという種々な噂を流すことで、機微の正確性をとおしめるという手段を執行した。以後、ネット上に出る内部告発文は信用を失い、事態は沈静化したかに見えた。

それまでは、単なる愉快犯として活動していた「ハッカー集団『笑い男』」は、匿名アカウント上の情報交換の末にセラノ社が行ってきたことの全容を知り、彼ららの「独断的な正義感」によって、セラノ社へサイバー攻撃を行うことを決めた。

「笑い男」を構成するメンバーは、数名とも数百名とも言われており、次々とアカウントを乗り換える彼らの全容を知ることには極めて困難に思われた。事実、笑い男に参加しているメンバー自身でさえ、何人の仲間がいるかわかってはいなかった。

匿名な集団である笑い男ではあったが、ネット上に駆々と作られては消去される匿名アカウントで常にコミュニケーションを行い、密く密に、一貫性のある「組織」として機能していた。



JUAS

2016年度 研究成果報告

**データマネジメント研究会
データサイエンス研究会**

データマネジメント研究会

研究成果 —第1分科会— —データ統合の技法/進め方—

第1分科会の目指すゴール

データ統合のパターンと手段の確立

過去のデータマネジメント研究会では、データマネジメントを推進する方法等を検討されていた。今年度は、データマネジメントの成果としてデータ利活用を事業価値に結び付ける過程で必須となる『データ統合』にフォーカスし、具体的にどのような手順/ツールを利用すれば実現できるのかを実践的に論考するため、事例研究を主軸としながら、技術的な方法を検討・整理した。

「データを統合する」とは

「データ統合」という言葉の意味するところには、
大きく3つの「統合」が存在する。

1. データの地理的統合

- ・複数の場所、システムにあるデータを1箇所に集約すること。
※物理的には離れたままでも利用者が1箇所でデータを入手できれば良い。

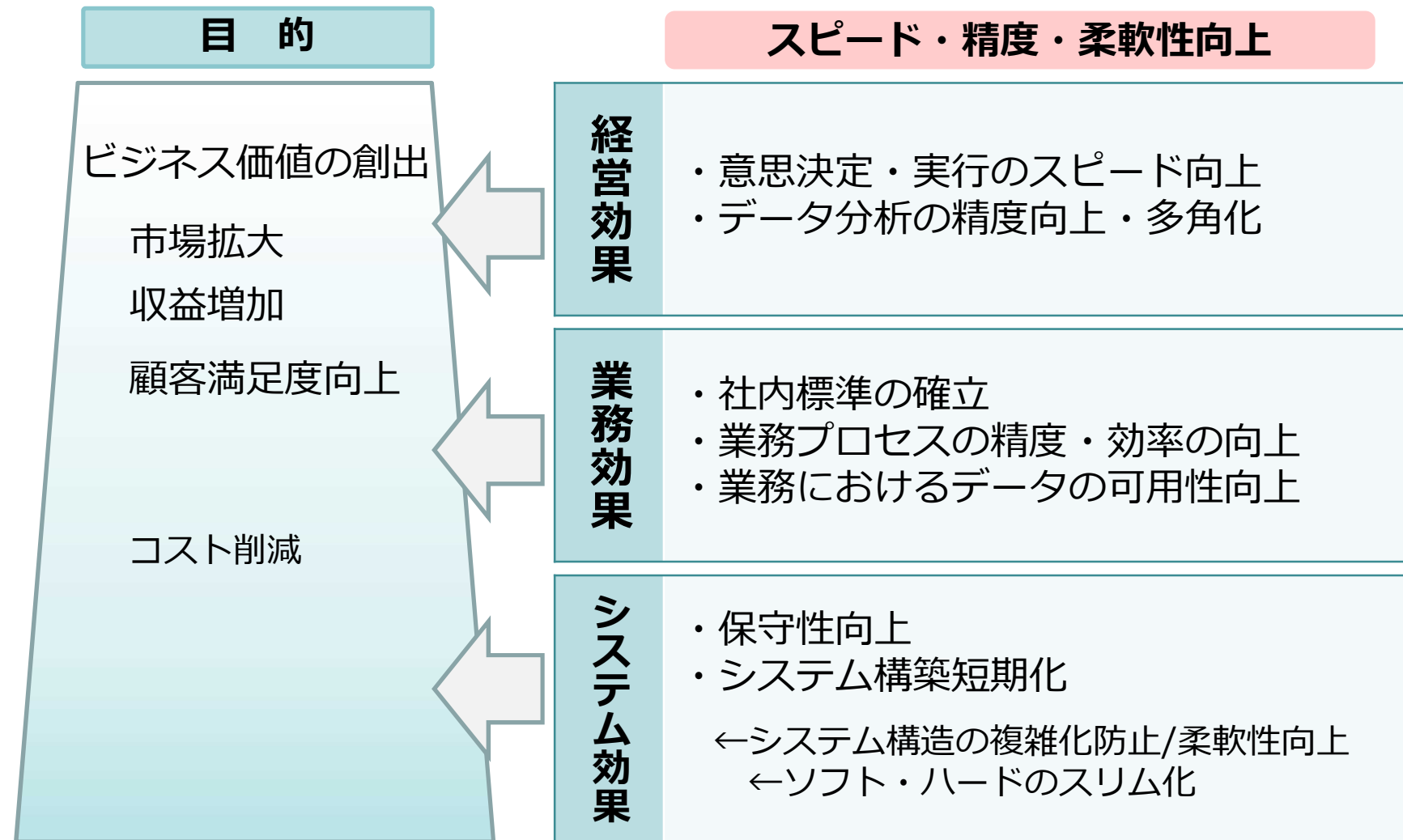
2. データの構造的統合

- ・様々なシステムが生み出す異なるフォーマットのデータを、所定のモデルに揃えること
※必ずしも正規化を意味するものではない。
構造的には冗長でも、定められた1つのモデルに収めることが肝要。

3. データの意味的統合

- ・用語の意味を揃え、一つの言葉が表すデータの内容を統一すること。
※用途の設定も含め、利用者への啓蒙も不可欠。

データ統合の主な効果



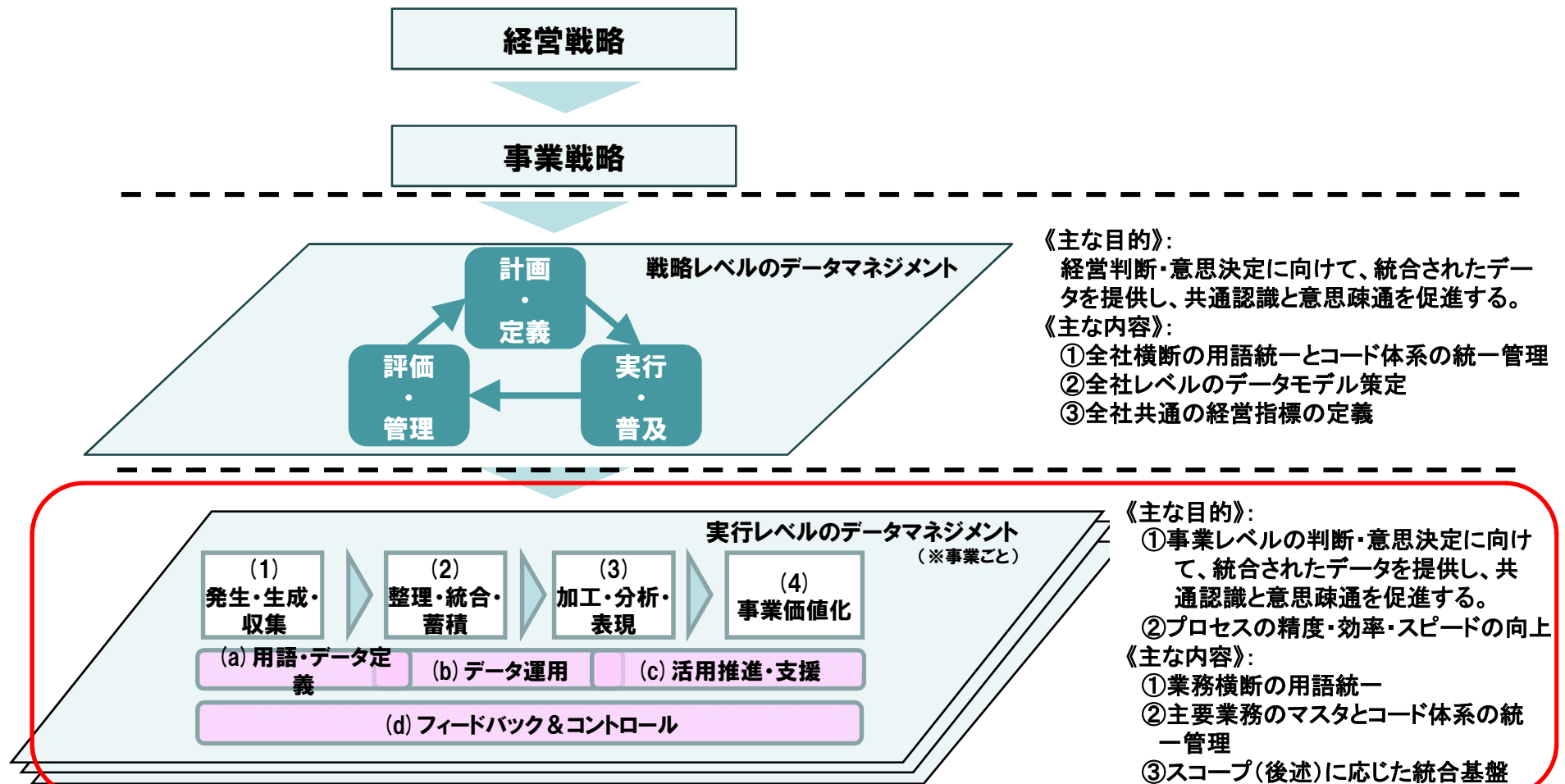
事例の収集と分析

- 23社27事例を収集・分類し、データ統合の「狙い」、「効果」、「手法」などの観点から整理・分析を進めた。

事例(会社)	Scope	業種	業種	改訂前のシステム課題	トピック(組織・主体等)	対象組織(事業・地域)	業務領域	パターン	手段	コスト	期間	効果	備考		
1 資生堂	S1	化粧品	1	海外各社等に個別システム	グローバル化、海外売上比率50% 主体：経営トップ	全事業 欧州、アジア、米国、日本	生産、販売、会計	2	基幹システムのDB統合	3	ERPパッケージ導入	高	トップダウンによる導入が前提		
2									手続APIの再構築	低		高			
4 匿名	S4	機械製造メーカー	2	故障予測の実現	工場内と顧客との連携、サービス稼働率の向上 高度なサービスレベル契約の実現と顧客満足度の向上 保守部品生産、在庫などの最適化とコスト削減 主体：？	日本国内	保守サービス	1	DWHシステムとの統合	中	1年	中	データを利用する際にデータ統合するため、柔軟性があるが、利用するには、ユーザ側にもデータに対する知識レベルが要求される http://www.fujitsu.com/jp/solutions/infrastructure/construction/dwh/cas02/		
7 オムロン	S1	電機メーカー	1	顧客情報の分散化	顧客情報CRMソフトウェア構築 主体：経営トップ(カンパニー系)	日本国内	販売、サービス(コールセン タ、Webサイト)	1	既存システムとの(仮設) 統合	2	1年	高	顧客マスタ統合保持のため、継続的な人的運用(名寄せ)が必要		
8 協和発酵キリン	S1	医薬品製造業	3	顧客情報の分散化・データサイニング	企業合併、情報統合、データサイニングの効率化 主体：経営トップ	日本国内	全業務	1	データHUBによる仮想統 合	小	2年	高	メタデータ、eレシジリ(社内Wiki)として社内公開		
9 セイコーエプソン株式会社	S1	精密機械メーカー	1	同一事業部内が工程ごとに分けて いたシステム	事業全体を完全に一元化(システム/システム/見 る化) 主体：？(経営層の事業理解への早期対応)	1 事業部 日本国内と海外	基幹系情報システム	2	ERPパッケージ導入	低	？	高	https://www.nttdata-gsi.co.jp/case/sap_erp_model/epson.html		
10 本田技研工業株式会社	S3	自動車メーカー	1	地域別に閉じた個別システムに存在するベ ーゼーション情報、製品情報	サブライティングの可視化による在庫の増減、リード タイムの短縮、在庫変化(リマニング)の即時把握、タイ ムの共有と見込みの把握 主体：経営	日本国内と海外	生産	2	データHUBによる統合	中	？	高	統合方法について引き続き検討中		
12 カシオ計算機	S2	電機メーカー	1	連結経営分析データ環境がない EUC普及の障害(不統一な分析、時間か かる)	ERP統一、連結での事業管理強化の必要性 リアルタイム分析、業務改善推進(営業スタッフ業務 減) 主体：？	グローバル	会計、販売	1	統合DWHの設置とETL 集約によるデータ 統合	中	10ヶ月程度(実装2ヶ月)	高	事業部ごとにサイロ化されたデータ連携運用を行っていたため、いざデータを集めても統一した データの見方が定まらない、変化の激しい、連結、グローバルへの集約の 意識が希薄化、独自運用や土着のやり方などへの抵抗		
14 三菱電機インフォメーション システムズ	S3	電機メーカー系IT	1	連携システム、異種連携システム、プロ シエールシステムなど分散	社員全員が最新のシステム状況を知ることができ、経 営情報基盤システムのインフラを刷新 主体：？	？	SCM全般	1	既存システムからのデータ 抽出と収集	1	ETLツールによる事例 データ統合	中	データを事前に統合するため、統合のための検討に時間がかかる →段階的統合が必要		
3 匿名	S3	大手小売業	2	加工現場に時間がかかり 売場分析が平日かかる	天候変化や交通状況などによる販売促進の実現 主体：？	日本国内	販売	1	DWHシステムとの統合	1	ETLツールによる事例 データ統合	中	データを事前に統合するため、統合のための検討に時間がかかる →段階的統合が必要		
15 株式会社シーアールイー	S3	不動産	4	M&A後の各社データの統合	データ経営意識を醸成して活用 主体：執行役員、経営企画本部	日本国内	会計・業務・営業	1	DWHシステムの統合	2	データ統合化による統 合	中	高	https://www.apr-japan.com/case/ore.html	
5 匿名	S3	石油	1	サイロ化されたデータ資産	収益の増大、コストの削減、およびリスクの軽減 リアルタイム分析 主体：？	メキシコ	生産	1	DWHシステムとの統合	2	データ統合化による統 合	中	高	https://www.cisco.com/web/JP/services/enterprise-it-services/data-visualization/pdf/Data_Virtualization_for_the_Energy_Industry_v2.pdf	
6 匿名	S3	エネルギー企業	1	マスターデータの管理の複雑化	マスターデータの品質向上、運用コスト削減 主体：？(IT部門に依存)	？	？	1	DWHシステムの統合	2	データ統合化による統 合	中	高	https://www.cisco.com/web/JP/services/enterprise-it-services/data-visualization/pdf/Data_Virtualization_for_the_Energy_Industry_v2.pdf	
11 紀陽銀行	S2	金融	1	各種データの分散と品質管理	各業務システムが連携して利用でき、統合データベース (物理統合)の構築 主体：情報システム部	日本国内	1 意思決定支援系	1	データ統合基盤による既 存DBの仮想統合	2	2.5ヶ月(1stリリース)	中	高	https://www.cisco.com/web/JP/services/enterprise-it-services/data-visualization/pdf/Data_Virtualization_for_the_Energy_Industry_v2.pdf	
13 紀陽銀行	S2	金融	1	業務単位で分散したDB	顧客関連情報の水増しと精度確保の必要性 主体：経営企画部	日本国内	顧客接点	1	仮想統合DWHの構築	2	2ヶ月で一度リリース	中	高	https://www.cisco.com/web/JP/services/enterprise-it-services/data-visualization/pdf/Data_Virtualization_for_the_Energy_Industry_v2.pdf	
16 紀陽情報システム株式会社	S2	金融	1	サイロ化されたデータ資産	業務効率化、マスターデータの品質向上 主体：経営企画部	日本国内	融資・営業・投資信託・収益	1	DWHシステムの統合	2	データ統合化による統 合	中	高	https://www.cisco.com/web/JP/services/enterprise-it-services/data-visualization/pdf/Data_Virtualization_for_the_Energy_Industry_v2.pdf	
17 XCB	S2	金融	1	システム分散	既存事業の増大と決済プロセスの提供や 業務品質の向上による業務改善の実現 主体：？	日本国内	クレジットカード業務 決済業務	1	DWHシステムの再構築	4	？	？	？	https://www.teradata.jp/pres/2009/20091022.html	
18 JFEスチール	S1	鉄鋼メーカー	1	1つのシステムの利用と2つのシステム	問題を迅速に解決し、統合のナレッジ構築を早期に出 主体：経営トップ	日本国内	販売、物流	2	基幹統合	4	手動でAPIの構築	高い	3年以上	高	http://www.jfe-systems.com/special/_j_smile.html
19 日本電産	S3	自動車部品メーカー	1	データ収集・集計業務の自動化(EXCEL ベース)	業務効率化と集計精度向上、セルフサービス 主体：本社営業部門	日本国内と海外	販売	1	DWHシステムの統合	1	ETLツールによる事例 データ統合	中	？	高	日経情報システムズ - september 2016
20 リクルートライフスタイル	S4	予約サービス	3	分析データ提供スピードの遅延 (IT部門依存によるデータ提供の遅延)	セルフサービス 主体：IT部門	日本国内	情報提供サービス	1	DWHシステムの統合	1	ETLツールによる事例 データ統合	中	？	高	日経情報システムズ - september 2016
21 アサヒ飲料	S3	飲料	1	分析データ提供スピードの遅延 (IT部門依存によるデータ提供の遅延)	データドリブン型経営のシフト 主体：経営トップ	日本国内	販売	1	DWHシステムの統合	1	ETLツールによる事例 データ統合	中	？	高	日経情報システムズ - september 2016
22 日本バーベ	S2	情報サービス (複合サービス)	3	経営に必要なデータが見えない (情報の非対称性)	データドリブン型経営のシフト 主体：経営トップ	日本国内	1 意思決定支援系	1	DWHシステムの統合	1	ETLツールによる事例 データ統合	中	？	高	日経情報システムズ - september 2016
23 資生堂	S3	化粧品	3	グローバル化によるデータ分散 (分析結果表示に20分かかる)	データドリブン型経営のシフト 主体：経営トップ	日本国内と海外	販売	1	DWHシステムの統合	1	ETLツールによる事例 データ統合	中	？	高	日経情報システムズ - september 2016
24 レノボ	S2	コンピュータメーカー	1	グローバルな計画・調達業務を支援する分 析環境が必要 ODMに依存した生産体制が前提	グローバルな計画・調達業務の支援 主体：調達部門	日本国内と海外	調達	1	DWHシステムの統合	2	データ統合化による統 合	中	？	高	CISCO データ統合化
25 匿名	S1	金融	3	合併によるシステム統合	早期なデータ統合による合併プロセスの加速 主体：経営トップ	海外	1 意思決定支援系	1	DWHシステムの統合	2	データ統合化による統 合	中	？	高	CISCO データ統合化
26 株式会社ユーキャン	S2	教育・出版	2	適度・拡張性	よりリアルタイムな顧客行動分析のマーケティング の実現 主体：？	日本国内	マーケティング	1	DWHシステムの統合	2	統合システム構築	？	？	？	http://www.teradata.jp/casestudy/cas_05.html
27 ロートンディー・株式 会社	S1	コンピュータ周辺機 器メーカー	1	異なるERPシステム使用 情報統合による大きな改善を要する	事業の世界的展開を加速させるため、 経営情報基盤の刷新と高度な経営分析の実現 主体：経営トップ	世界8拠点	1 意思決定支援系	2	基幹統合	1	？	？	？	？	？

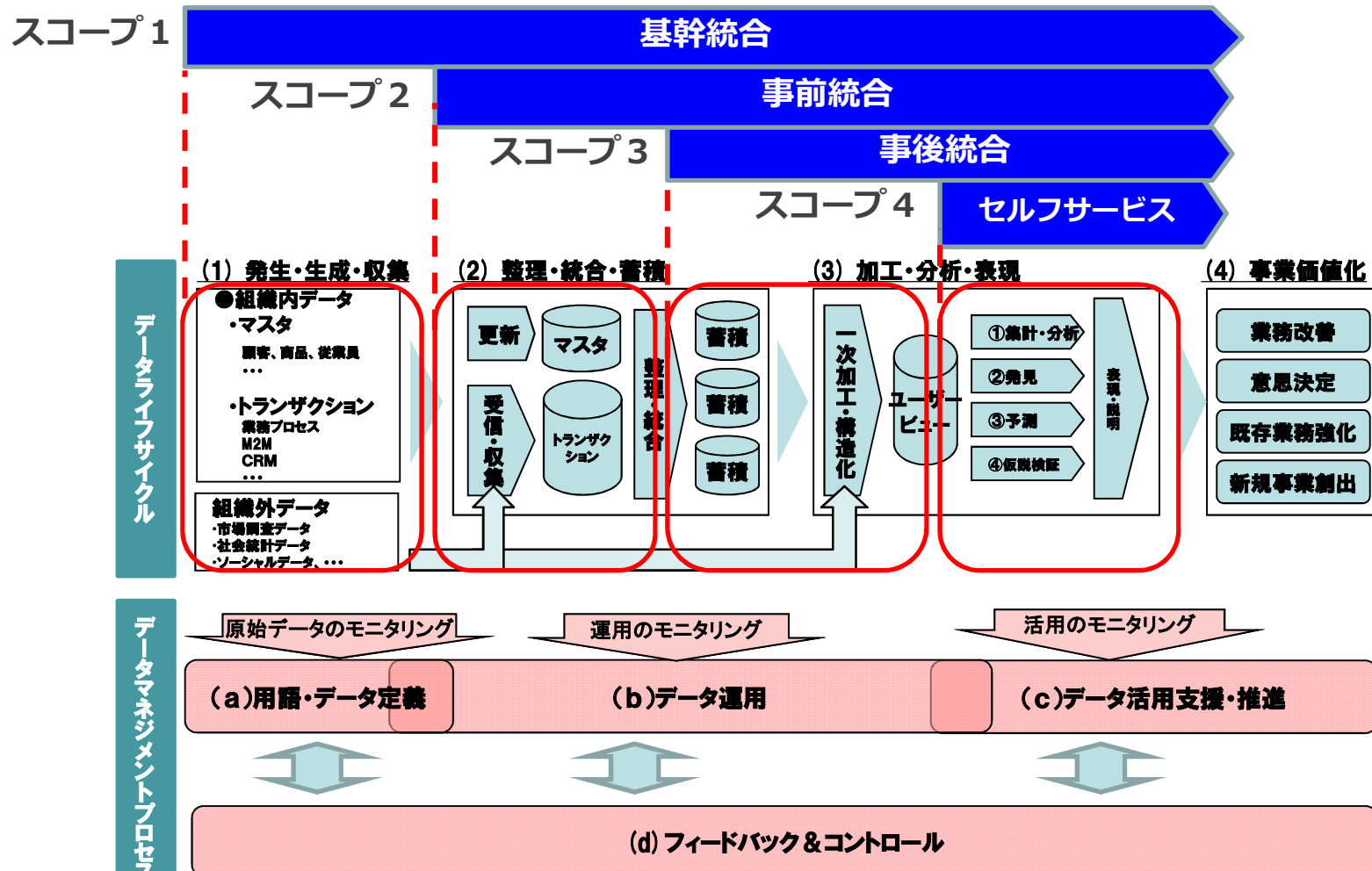
データマネジメント全体像におけるデータ統合の位置づけ

- データマネジメントの全体像におけるデータ統合は、「戦略レベルのデータ統合」と、「実行レベルのデータ統合」が考えられる。



実行レベルのデータマネジメントとデータ統合スコープ

■ 実行レベルのデータマネジメントにおけるデータ統合は、データライフサイクルのどの段階で行うかによって、4つのスコープが考えられる。



データ統合のパターンと手段

- データ統合にはデータ発生元の統合と、データ利用側での統合がある
- 統合の手段は、統合の目的/効果に応じて選択される

パターン	N O	手段	変更 柔軟 性	コス ト	期間	PJ 難 度	前提・リスク	主な目的
①データ発生元の統合 -基幹系システム側-	1-1	ERPパッケージ導入で基幹システムのDB統合	低	高 ～ 中	長期 ※ウォーターフォール	高	<u>トップダウン</u> による実行	経営方針に沿った、 <u>合併、海外子会社のグローバル統合等</u>
	1-2	手組アプリで基幹システムのDB統合	低 ～高	高	長期 ※ウォーターフォール	高		
②データ利用側での統合 -情報系システム側-	2	事前データ統合	低 ～中	小 ～中	スモールスタート可 ※アジャイル	小 ～中	統合レベルに応じたデータ設計の時間確保	データ収集・ <u>分析業務の効率化/精度向上</u> <u>データから新たな価値の創造等</u> (トライアル的なものも含む)
	3	事後データ統合	中 ～高	中	スモールスタート可 ※アジャイル	中	統合レベルに応じたデータ設計の時間確保 <u>利用側にデータに対する深い知識(データ構造等)が必要</u>	
	4	データ仮想化による統合(仮想DWH)	高	小	スモールスタート可 ※アジャイル	小	物理データを持たないため、 <u>柔軟性が高い</u> が、 <u>性能が課題</u> になる可能性がある	
	5	データ仮想化による統合(仮想DM)	中 ～高	小 ～中	スモールスタート可 ※アジャイル	小 ～中	統合レベルに応じたデータ設計の時間確保 <u>データマートの柔軟性が高い</u>	

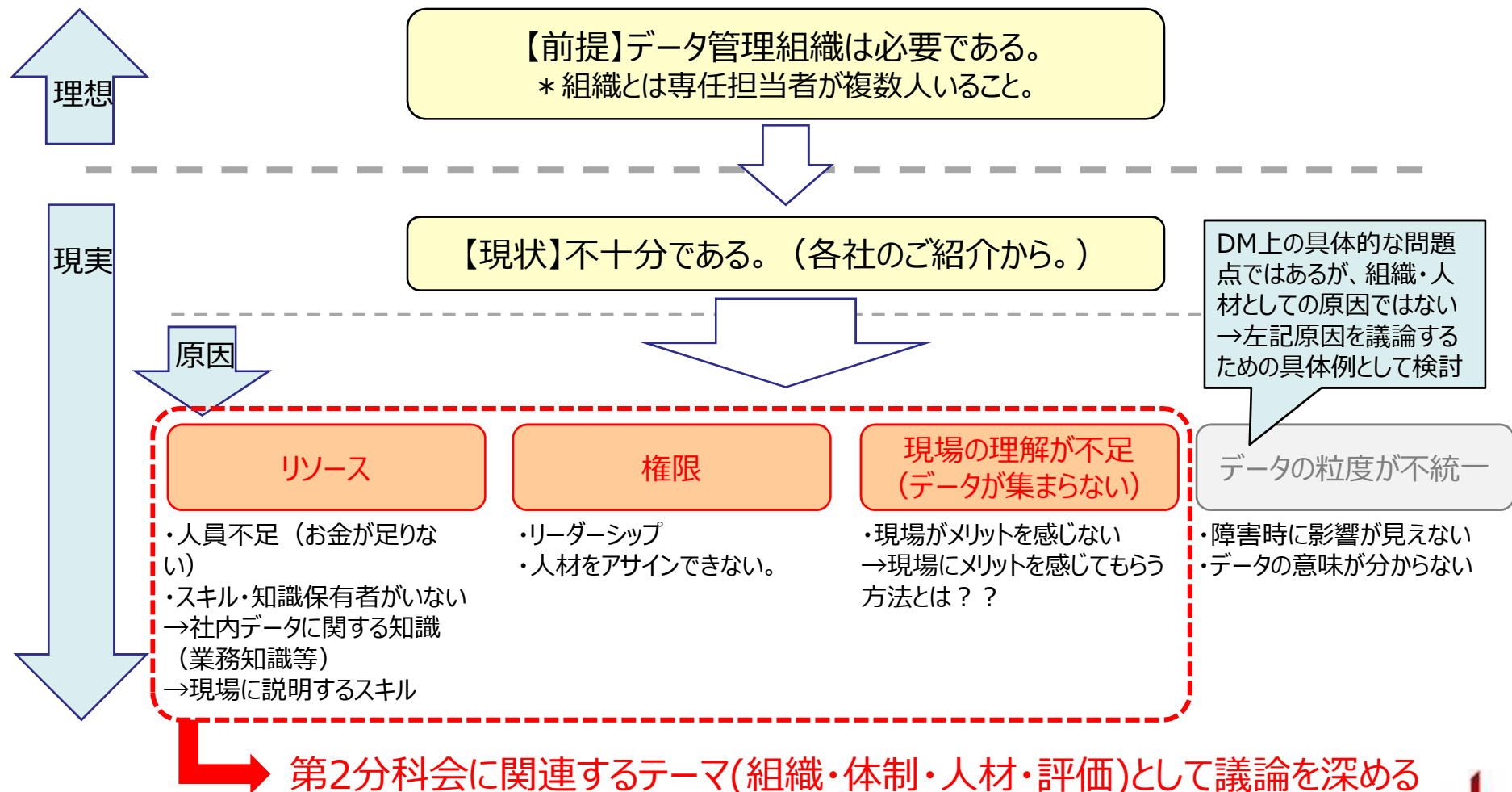
データマネジメント研究会

研究成果 ー第2分科会ー

ーデータの利活用のためのデータ管理組織とはー

第2分科会の課題認識

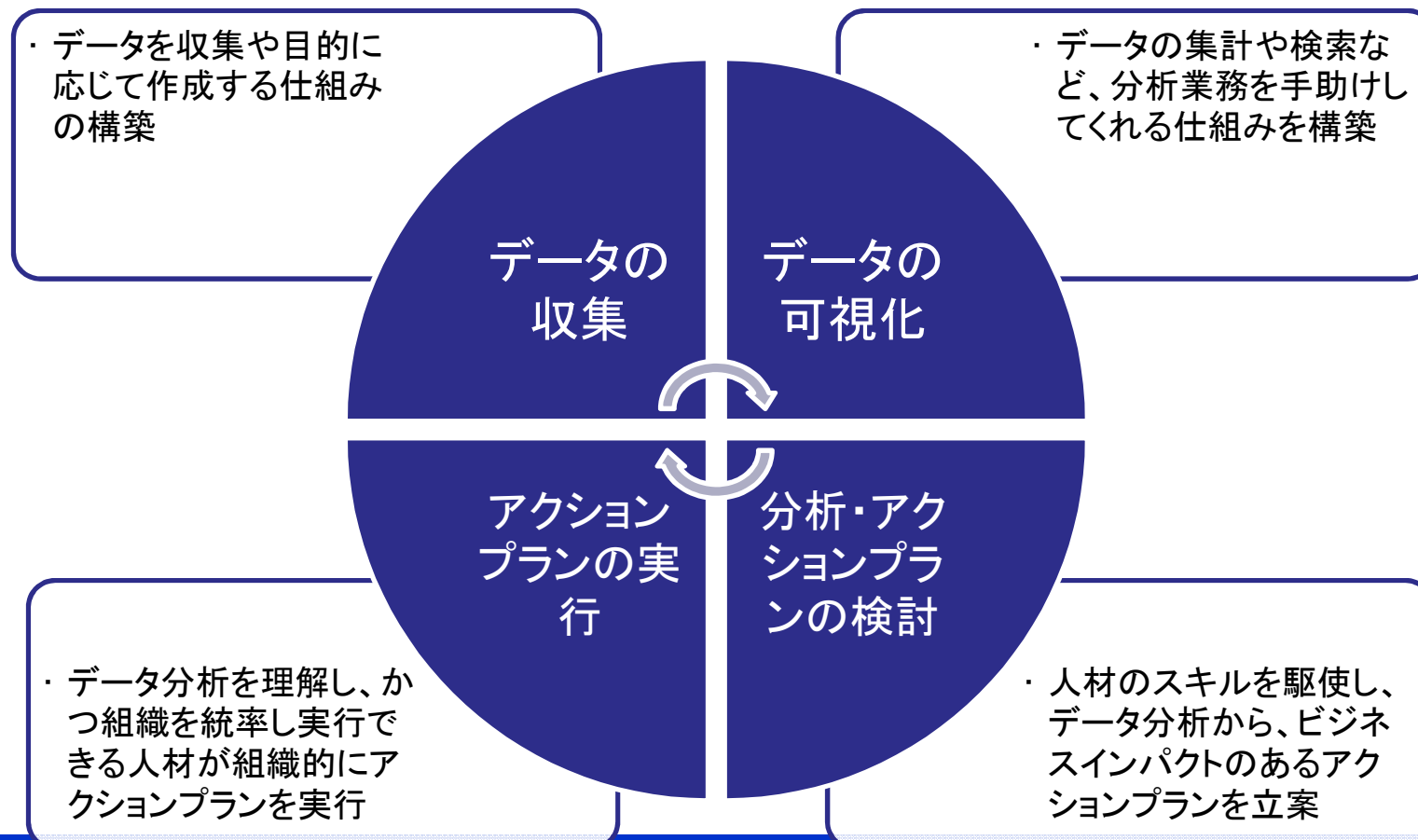
- 各社ご紹介頂いた事例を基に、DMおよびDM組織が不十分である原因について4点抽出した。



データ管理組織の必要性 –その1–

・データドリブン経営の定義

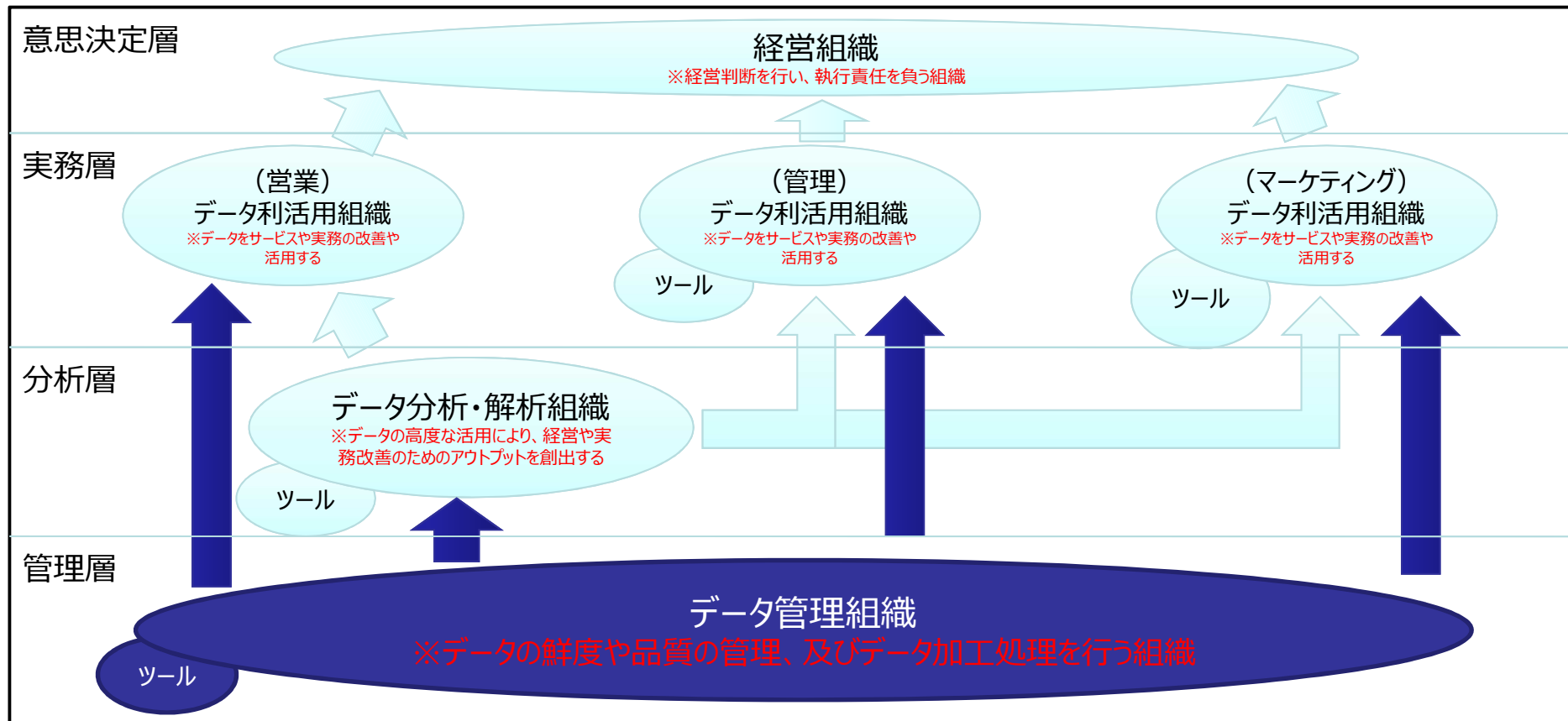
- データドリブンとは、「様々なデータに基づき判断・アクションする事」という意味であり、データドリブン経営とは以下にあるステップを実現・高度化していくことで、経営判断にデータを有効活用している経営を指すものとする



データ管理組織の必要性 –その2–

・データドリブン経営の構成要素

- 本書では前述のステップの実現・高度化によるデータドリブン経営を構成する要素として「データ利活用組織」、「データ分析・解析組織」、「データ管理組織」、「ツール」の4つの要素が存在すると考える

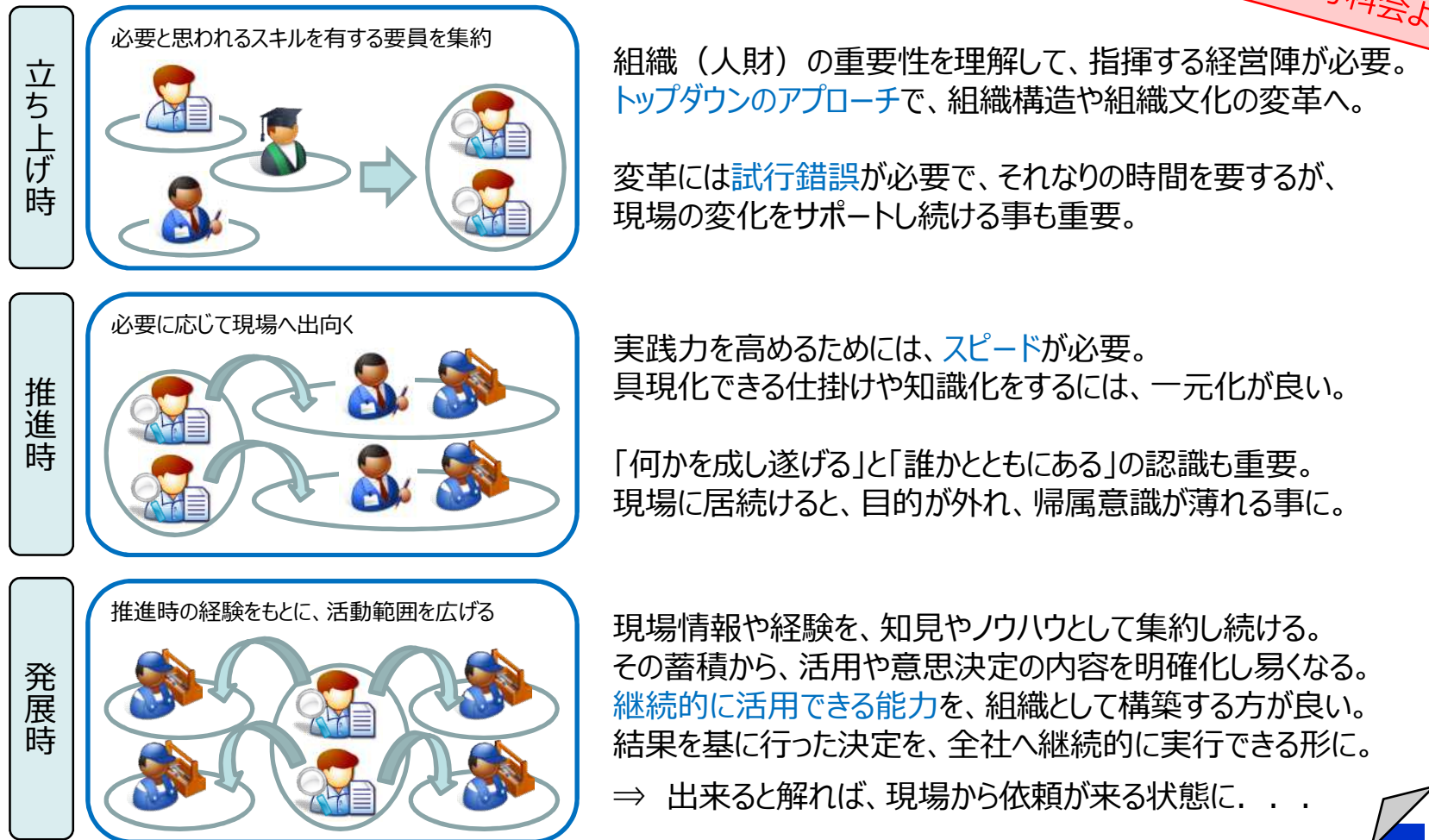


- 本書では「データ管理組織」の必要性やあるべき姿、実際の企業活動における実態等について言及していく

データ管理組織として想定される体制

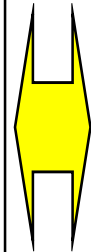
- データ利活用組織も、利活用の各段階によって組織パターンが異なるため、管理組織も各段階に合わせて最適なパターンがあるものと考えられる。

データ利活用各段階での良さそうな組織パターン



FY13第2分科会より

各段階のデータ管理組織のありかたとは？



データ管理組織として想定される体制 ― 立ち上げ時 ―

- データ利活用が試行錯誤の段階であるため、データ管理もシステム部門が都度対応する形で十分と想定される。(全社的・戦略的なデータ管理のニーズが出てきていない状態)

ニーズ・状態

組織・体制

データ利活用

- 経営層からのトップダウンニーズにより、一部メンバーがデータ利活用を推進するよう指示を受けた状態
- 現場からは具体的なデータ利活用のニーズはまだ出てきておらず、試行錯誤しながらニーズを喚起していく段階

- 有スキル者によるプロジェクトチーム？



データ管理

- データ利活用側からの散発的なデータ収集・整理などの依頼に対して、都度対応するような状態
- 全社的・戦略的なデータ管理の必要性がまだ明確でない状態

- データの格納先や状態を知っている情報システム部門が都度対応？

データ管理組織として想定される体制 ー推進時ー

- ・ 利活用に関しては、現場からのニーズに対し、一元的な組織で対応する段階。
- ・ この段階で明らかになる横串（全社横断型）のデータ管理課題に対し、データ管理組織の立ち上げが望ましい。

ニーズ・状態

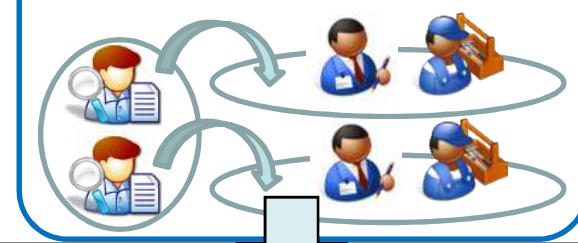
組織・体制

データ利活用

- ・ データ利活用の様々なニーズが現場から出てきている状態
- ・ データ利活用を推進する立場からすれば、この段階でいかに利活用の有効性を示せるかがポイント（＝実践力を高めることが必要）

- ・ 実践力・スピードを高めるため、一元的な利活用組織が望ましい？

必要に応じて現場へ出向く



データ管理

- ・ 一元的にデータ利活用に取り組むようになると、横断的・横串でのデータ管理の課題が明らかになってくるはず
- ・ スピード感ある利活用のためには、時間のかかるデータ整備の取組に長期的に考えていくべき、との声も出てくると推測

- ・ データ管理の社内有識者による組織を立ち上げることが有効と考えられる
- ・ 但し、専任組織を設けるほどの理解が得られない可能性があるため、現実的には、プロジェクトチーム型、或いは兼任者中心の組織になるものと考えられる

データ管理組織として想定される体制 ー発展時ー

- ・「立ち上げ」→「推進」を経て、専任者と中心とした組織として対応する体制となる。

ニーズ・状態

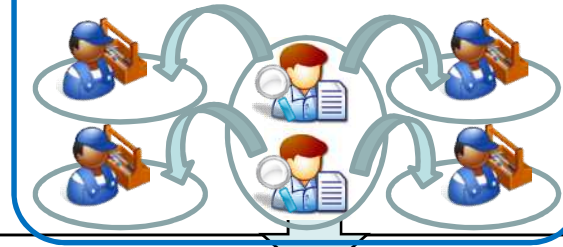
組織・体制

データ利活用

- ・ 全社的にデータ利活用が定着できている段階
- ・ いかにデータ利活用の社風を継続させられるかがカギ
→データ利活用が自分のビジネスにとって重要・効果のあること、と認識してもらうことが重要

- ・ 専任者を含む一元的な組織にて、利活用・定着を継続的に行う

推進時の経験をもとに、活動範囲を広げる



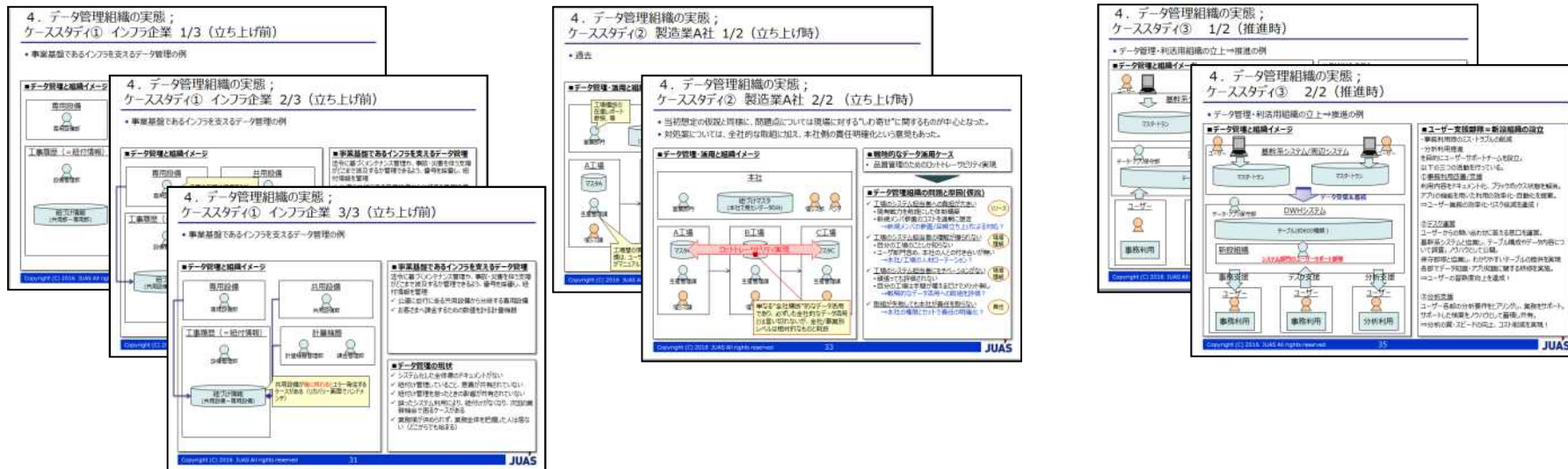
データ管理

- ・ データ管理に対するニーズも全社的なものとして明確になっている状態

- ・ データ管理組織も専任者を中心とすることが望ましい
- ・ 専任者のみ/専任者中心のどちらにするか、また、独立した組織/データ利活用組織内に設置のどちらが良いかは、各社状況に応じて判断

データ管理組織 ケーススタディー

分科会参加企業の組織紹介ヒアリングを実施し検証を行った。(計9社)

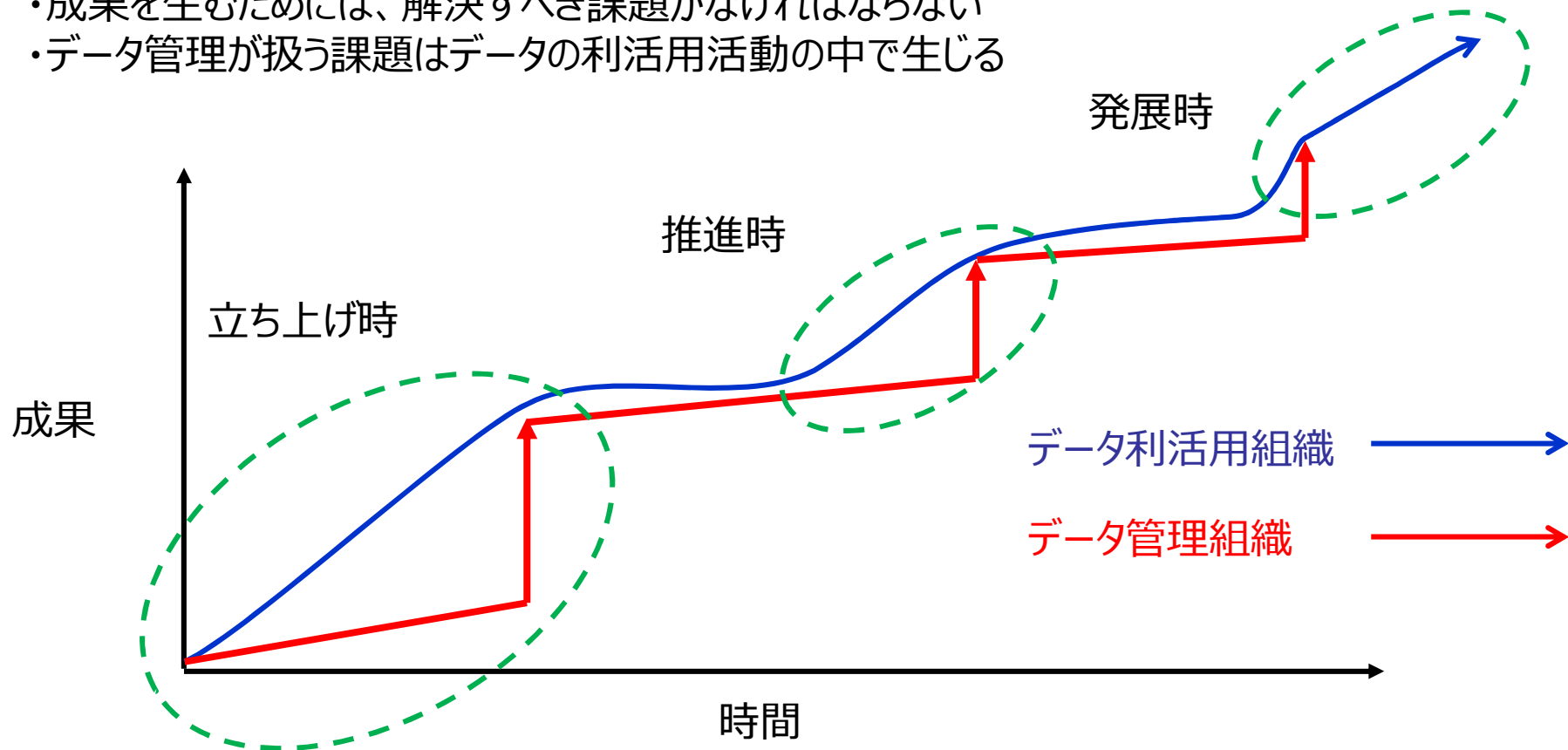


- データ利活用(組織)の各段階に応じてデータ管理組織を整備していくことによって、各段階のニーズに合わせた適切な役割を果たすことができる
- 全社的なデータ利活用が進み始めると、データ利活用のスピードアップ・効率化・更なる付加価値化のために、データ管理の重要性が増す。その段階になれば、全社横断型の専任組織を設置することがふさわしい。

継続してデータ管理組織が存在するためには

ー「全社」「専任」の難しさー

- 大きなリソースを要求する組織類型は、設立だけでなく維持ための条件も厳しい
- 部門横断の専任組織を存続させるためには、それに見合う成果を出し続ける必要がある
- 成果を生むためには、解決すべき課題がなければならない
- データ管理が扱う課題はデータの利活用活動の中で生じる



- データ管理組織は、データ利活用組織が発展するための土台作りが重要な役割となる。

データマネジメント研究会

研究成果 — 第3分科会 —

— データマネジメントの経営価値と
データドリブン経営の実現方法 —

第3分科会の研究テーマ

<<<データマネジメントの経営価値>>>

- ・経営レベル～全社へのデータマネジメント説明
- ・データを経営資源化するための施策を立案

<<<データドリブン経営の実現方法>>>

- ・業務やビジネスプロセスへデータ準備&利活用を組み込む
- ・データドリブン経営はどういった状態をいうのか

取組

3つの事例をもとに改良版VSを作成し、データマネジメント投資の説得力が上がるかを検証する。

「必要なデータ」や「データマネジメントの必要性」がうたえるか確認する。
そこから経営に昇っていくことができるか確認する。

データドリブン経営は、どういったことが出来ている状態をいうのかをまとめる。

世の中一般的な内容だけでなく、第3分科会としての検討結果がまとめれば良い。

◆3つの事例とVS改良版の作成

事例

- 事例1)SCMコクピット検討
- 事例2)宿泊(客室)稼働率向上
- 事例3)勤務時間の管理見える化

シナリオ整理(VS)



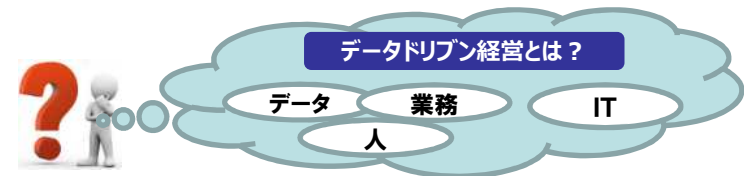
◆VSテンプレートを社内の説明資料に繋げる



- ・チームを分けて推進
- ・相互議論

◆データドリブン経営とはどんなもののかのイメージをまとめる

データドリブン経営イメージ



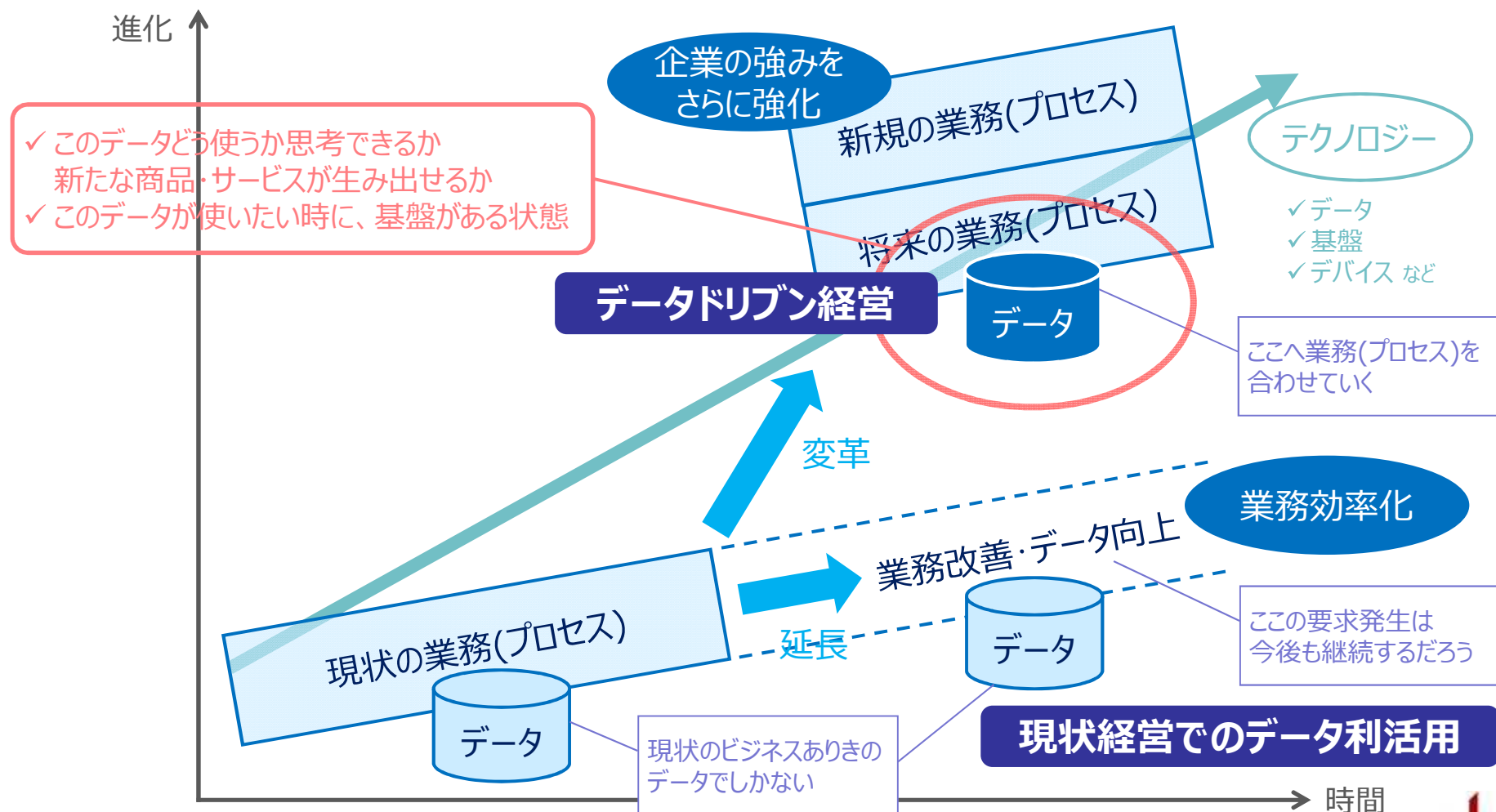
◆どのようにすれば実現することが出来るか



進め方(成果物)

データドリブン経営とは？ー①

これまでの業務改善では、現状の延長線上のデータしか得られない。現状のデータ価値のままである。テクノロジーの進化にあわせてデータをより活用し、業務を変革することで企業の強みにすることができる。



データドリブン経営とは？－②

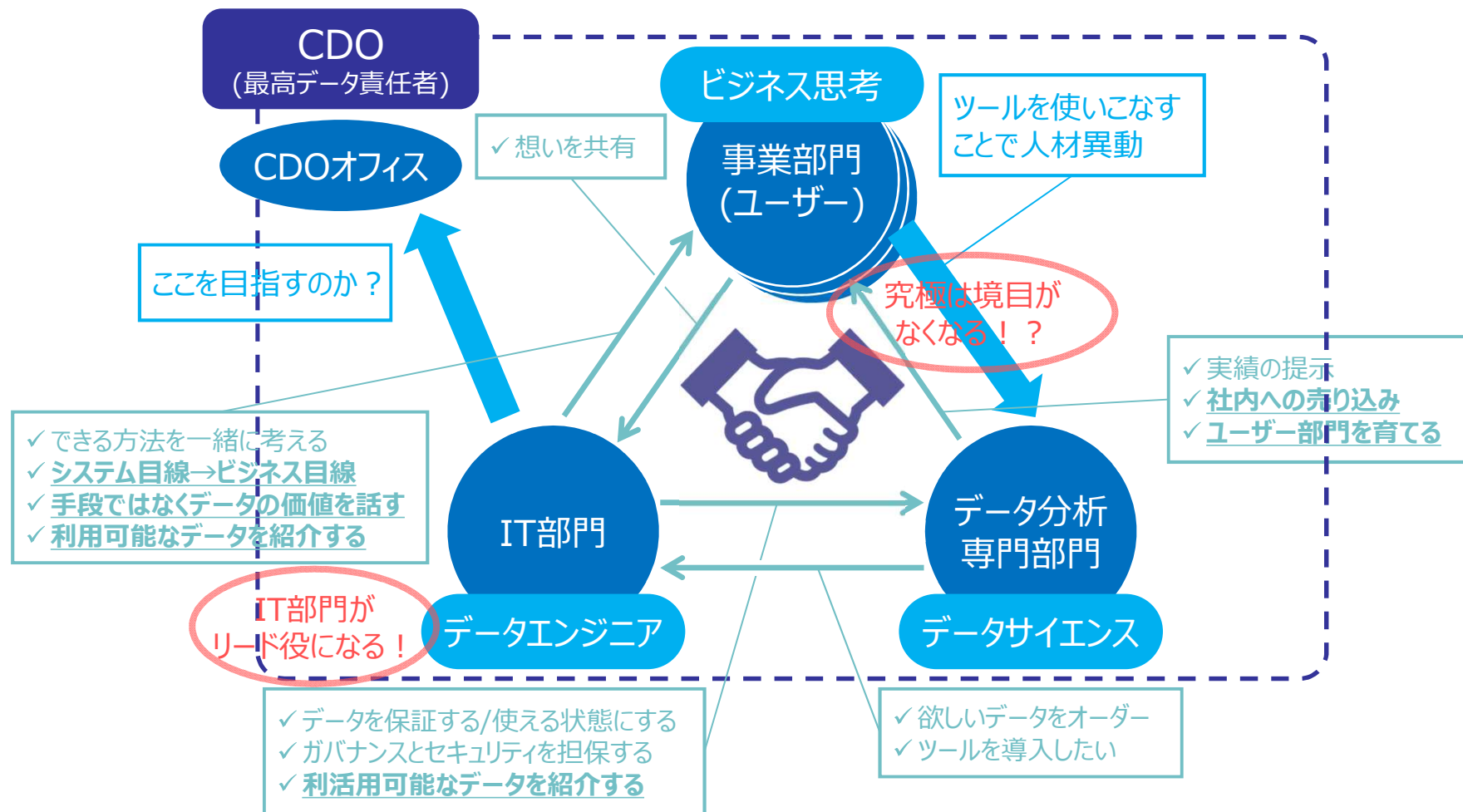
視点	今までの経営 (勘・経験・度胸 =KKDベースの経営)	現状経営での データ利活用	データドリブン経営
データドリブン経営 という状態① データ収集と利活用	- 現場の業務(プロセス)を変えない - データを使ってない	- 現場の業務(プロセス)を変えない - 既存業務で得られるデータ	- 現場の業務(プロセス)を変える - 業務を変えて得られるデータ - 外部から得られるデータ
データドリブン経営 という状態② 業務プロセスの進化	現状の業務(プロセス) →人海戦術 →気合いと根性	- 業務改善・データ向上 →業務効率化 - 既存のテクノロジーを活用	- 業務が大きく変革される →企業の強みをさらに強化 →新規事業・サービスを生み出す - 最新のテクノロジーを活用
データ利活用の 役割分担	- IT：分析・集計(Excelマクロ) - ユーザー：ITへ依頼	- IT：分析・集計(ツール) - ユーザー：一部セルフ&ITへ依頼	- IT：データ一元管理と基盤整備 - ユーザー：分析・集計(セルフ) - 分析専門：予兆分析 ⇒ユーザーと分析専門が組んで 経営層へ提言
経営者がデータ ドリブン経営を 受け入れられるか	期待：業務単位の業務効率化・コスト削減	期待：会社トータルの業務効率化・コスト削減	期待：ビジネスモデル変革/商品・サービス強化

SASが示す「データドリブン経営」

データを準備し、分析して洞察を引き出し、それを業務に適用してアクションを起こす。それがデータドリブンな経営であり今の時代に求められる姿だ。ITという枠組みを超えて、データ分析と業務改善が一体になってこそ、競争優位が確保できる。ビジネス課題から意思決定までの時間の約8割が準備に充てられている。この準備時間を短縮することが競争優位性を高める第一歩になります。

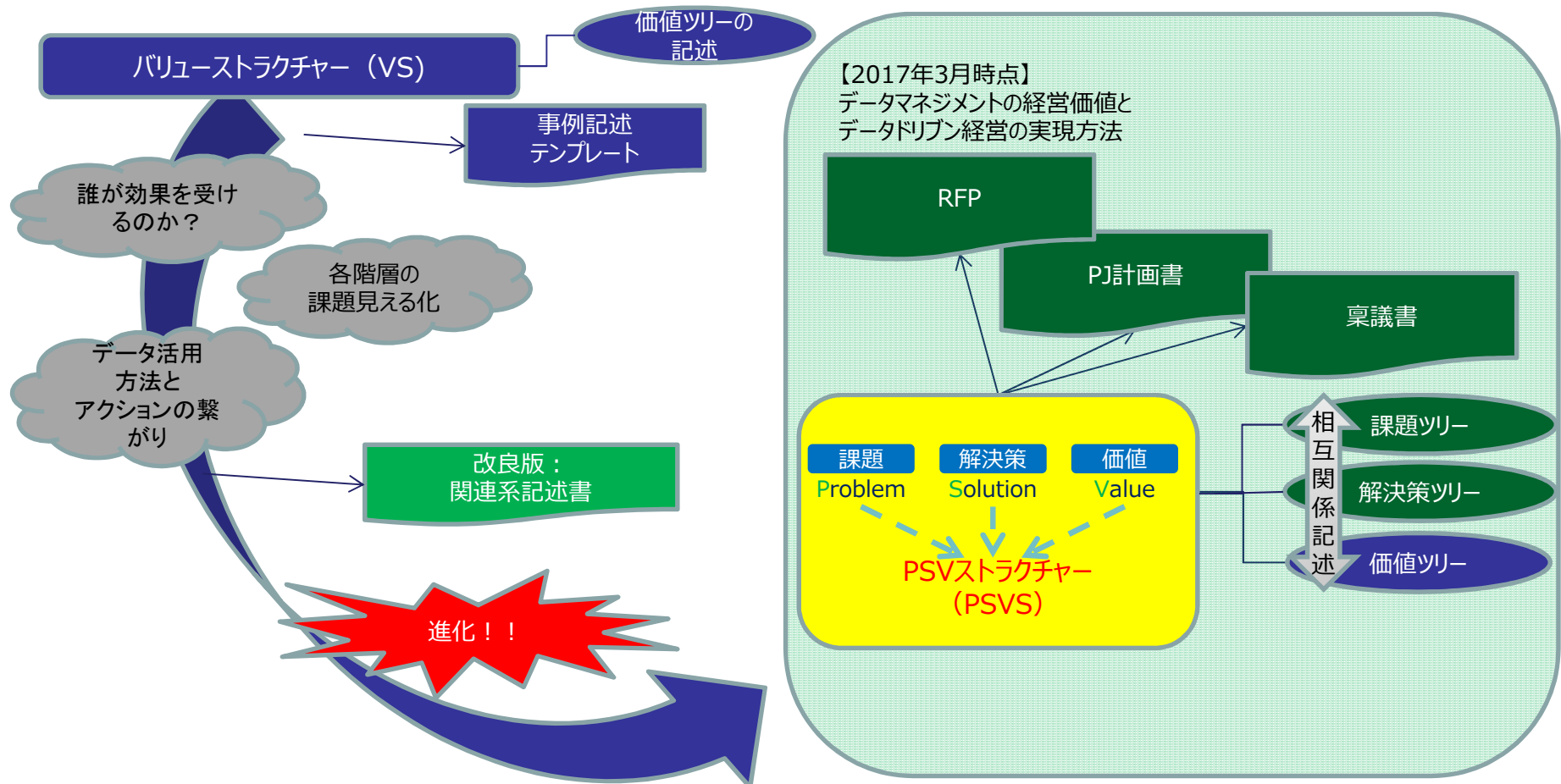
データドリブン経営とデータ利活用推進体制

IT部門・事業部門(ユーザー)・データ分析専門部門の3者で、どうコミュニケーションをとることが、データドリブン経営という状態に繋がっていくのか。



データマネジメントの事業価値説明

VSからPSVストラクチャーへの進化

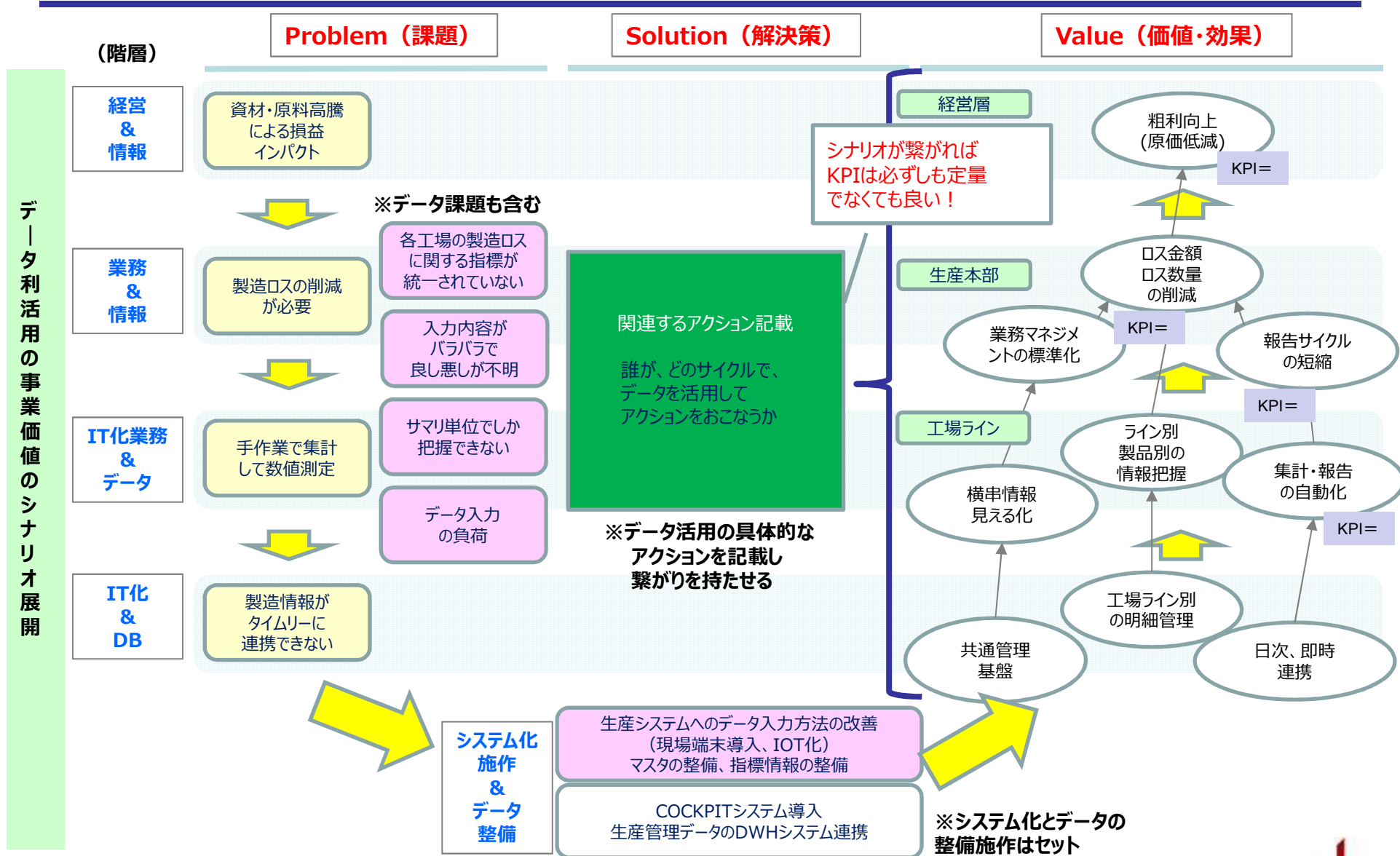


PSVストラクチャーは、IT化案件の企画・設計時の「シナリオと重要ポイントの整理」TOOLである

DM・DS案件も含む全てのIT化案件は、今後ますます「ビジネス価値」との関係性が問われます。

そのため、「経営視点」と「業務視点」に関係のある、どのようなテーマをIT化するか（DM対応するか、DSアプローチするか）を経営陣や事業責任者や業務部門関係者とコミュニケーションするためには、一度IT視点テーマ（業務APL、DB、各種基盤、利用クラウド、開発環境、DWH、分析TOOL・・・）を、そのIT化が最終的に狙う関係する一企業・業務活動視点から大きく捉え、【経営課題－業務テーマ－業務のIT化目標－IT化自体の中身】の関係性の説明シナリオを整理する必要があります。

PSVストラクチャーテンプレート(改良版:関連系記述書)



PSVストラクチャーの活用による(稟議)企画書 & 各種計画書作成

PSVストラクチャー

経営テーマ⇒
経営課題・価値・効果
(ROI・ROE・
KGI・・・)

業務テーマ⇒
業務課題・価値・効果
(KPI)
(業務情報価値)

業務APL機能テーマ⇒
IT化業務 (KPI)
課題・価値・効果
(データ価値&
APLの業務価値)

IT化テーマ⇒
IT化 (KPI)
課題・価値・効果
(APL機能価値と
データ・DB価値
それを支える
IT基盤価値)

テーマの繋がりとシナリオの流れ検討・確認

重要なポイント情報の提供

PSVストラクチャーの検討内容はRFPやシステム計画書や稟議書項目の赤字部分に指針を与えます。

1. <RFP目次>

2. 本文書の目的と提案依頼プロセス
 - 1.1 本文書の目的
 - 1.2 選定のタイムスケジュール
2. プロジェクトの目的と課題
 - 2.1 プロジェクトの目的
 - 2.2 システム化の課題
 - 2.3 システム化の方向性
 - 2.4 システム化に向けた問題点 (現状とギャップ)
3. ビジネスとシステムの現状
 - 3.1 現状業務の概要 (イメージ)
 - 3.2 現状システムの概要 (イメージ)
4. 想定するシステム像
 - 4.1 想定業務の概要 (イメージ)
 - 4.2 想定システムの関連図
 - 4.3 新業務フロー (オーナーズスタイリング業務)
 - 4.4 新システム想定機能一覧
 - 4.5 機能要件
- 5.1 プロジェクトの範囲 (システム化影響組織)
- 5.2 プロジェクト体制と役割
- 5.3 プロジェクトの範囲 (対象システム)
- 5.4 システム導入 (稼働) 時期
- 5.5 想定されるリソースとトランザクション
- 5.6 入出力に関する要件
- 5.7 データ要件
- 5.8 セキュリティ要件
- 5.9 ハードウェア要件
- 5.10 パフォーマンス要件
- 5.11 他システムとの連携要件
- 5.12 作業体制要件
- 5.13 設計・開発・導入要件
- 5.14 教育・トレーニング要件
- 5.15 保守運用・サポート要件
- 6.1 提案書の提出
- 6.2 提案依頼内容
- 6.3 その他留意事項

<システム化計画書 目次例>

1. 1 プロジェクトの目的 (背景・必要性・解決すべき業務課題)
1. 2 システム化の課題 (業務課題・IT課題・組織課題)
1. 3 システム化の方向性 (目的・狙い・課題解決方法)
1. 4 システム化に向けた問題点 (業務プロセス・データ・基盤)
2. 1 現状業務の概要 (イメージ)
2. 2 現状システムの関連図
3. 1 想定業務の概要と利用組織・部門・利用者数
3. 2 想定システムの関連図 (対象システムと関連システム)
3. 3 新業務フロー
3. 4 想定データフロー図
3. 5 新システム想定機能一覧
3. 6 個別機能概要と必要データ
3. 7 新システムの基本構造
4. 1 プロジェクトの範囲 (システム化影響組織)
4. 2 プロジェクト体制と役割
4. 3 プロジェクトの範囲 (対象システム)
4. 4 プロジェクトスケジュール
4. 5 プロジェクト予算概算
4. 6 システム化投資と効果
4. 7 システム保守・維持とコスト
4. 8 システム利用期間

<稟議書目次>

- 投資タイプ&対象テーマ (情報) レベル
- 投資タイプ:
- 主要情報&データ:
- 背景と必要性について
- 目的・狙いとKSF
- 導入規模・範囲/想定運用期間
- 導入&実施効果 (KSF実現の為の評価KPI)
- 導入費用 (初期・運用) と主幹部門
- 導入一時費用:
- ランニング費用
- 一月額費用
- 一年額費用
- ー随時費用
- 費用負担主幹部門
- 投資対効果についてー評価方法&回収期間
- プロジェクト関連ステークホルダーと 本プロジェクトへの協力依頼
- プロジェクト推進体制&責任者&担当者
- プロジェクトスケジュール&主要マイルストーン
- <WBS & スケジュール概要>
- 今回の対象システム&データと関連するシステム & DB、データ移行についてー注意事項

(稟議) 企画書にフォーカスを当てて、PSVストラクチャーの内容が落とし込み出来るかを各社で使われている項目やフォーマットを参考にして検証。

PSVストラクチャーの活用による(稟議)企画書フォーマット

■投資タイプ&対象テーマ(情報)レベル

投資タイプ:

主要情報&データ:

■背景・必要性・リスクについて

○

○

■目的・狙いとKSF

○

○

○

■導入規模・システム化範囲/想定運用期間

○

○

■導入&実施効果(KSF実現の為の評価KPI)

○

○

■導入費用(初期・運用)と主幹部門

○導入一時費用:

○ランニング費用

ー月額費用 XX費: XX千円(税別)、

ー年額費用 XX費: XX千円(税別)

ー随時費用 XX費: XX千円(税別)

○費用負担主幹部門: XX部

■投資対効果についてー評価方法&回収期間

※特に対象データの投資価値について記載する。

PSVSでシナリオ整理したことで業務への価値が

定量+定性での説明できるのではないか。

■プロジェクト関連ステークホルダーと 本プロジェクトへの協力依頼

■プロジェクト推進体制&責任者&担当者

■プロジェクトスケジュール&主要マイルストーン&プロジェクトリスク

プロジェクト期間:~.....

○マイルストーン1:

○マイルストーン2:

.....

○プロジェクトリスク1:..... 対策.....

<WBS & スケジュール概要>

■今回の対象システム&データと関連するシステム&DB、データ移行についてー注意点

データサイエンス研究会

研究成果

データサイエンス研究会の研究テーマ

1. 研究テーマ

データ分析・データ利活用を実践するための方法論、未来像の研究

2. 体制

部会長 庄司 稔（株式会社JALインフォテック）

副部会長 三好 克典氏（株式会社プライド）

清嶋 淑也氏（株式会社アイ・エム・ジェイ）

3. 活動方針

いまやデータ利活用によるICTが事業活動に不可欠であり、方式・手法・評価方法も多種多様です。本研究会は、今までの研究成果を引き継ぎ、兄弟研究会であるデータマネジメント研究会とのコラボレーション活動等を通じて、以下の3つのテーマで活動します。

①データ利活用に欠かせないプロセス・手法を突き詰める【DS1】

②事例研究・アンケート分析・ロールプレイからデータ利活用の気づきを得る【DS2】

③明るく楽しく元気よく手を動かしてデータ活用を実践してみる【DS3】

データサイエンス研究会 【DS1】

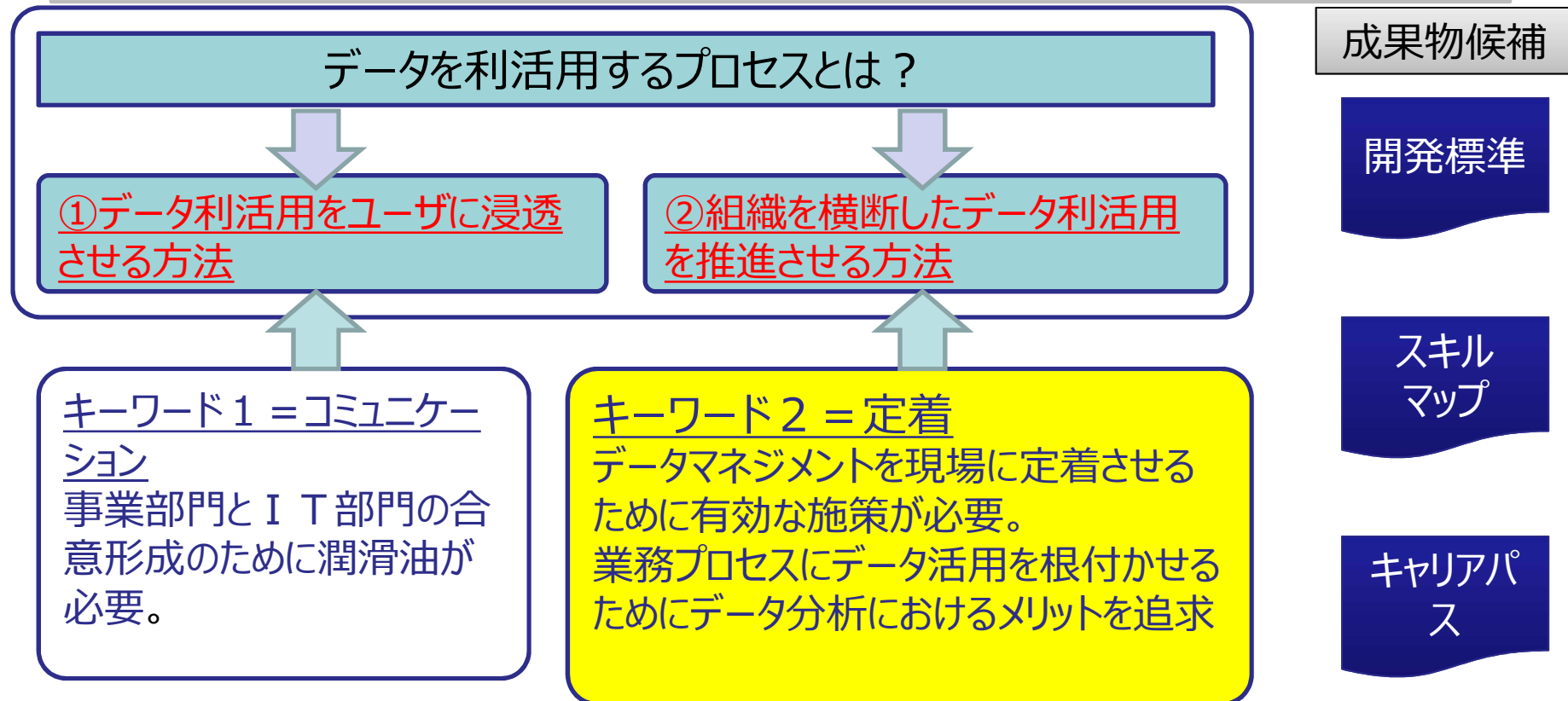
①データ利活用に欠かせない
プロセス・手法を突き詰める

【DS1】起点（合宿時の議論）

合宿において、「データ利活用の定着」をテーマに、その実現のために必要となるアクティビティをメンバーで持ち寄り議論しました。

その結果、データ利活用について①②の2つの方法に収れんし、その実現のためのキーワードとして、「コミュニケーション」と「定着」としました。

このうち、定着に必要なコンテンツ3つを成果物候補としました。



【DS1】テーマ決定

他にも、**2015年度DS第8分科会**において、データ利活用ができない理由が詳細に考察されており、それを踏まえた提言がなされているが、しかし・・・

結局、全ては“人”

データ利活用に向けたルール整備や提言は過去の分科会でもなされているものの、**データ分析を理解している人材が大幅に不足**しており、**データ分析を前提とした組織構築、人材育成が進まないことから、利活用が定着しない。**
そもそも業務のプロとITのプロであってデータ分析のプロではない。



データ分析する人財のスキルマップとそれに必要な事業部門、IT部門の役割を整理する



【DS1】INPUT1（縦軸）滋賀大データサイエンス学部における人材像＋データサイエンティスト協会における人材像



SHIGA UNIVERSITY

滋賀大データサイエンス学部およびデータサイエンス協会の資料を基に、人物像の役割分担（縦軸）を明確化

育成する人材像、卒業後の進路・キャリアパス



スキルレベル		
スキルレベル	目安	対応できる課題
シニア・データサイエンティスト	・ 業界を代表するレベル	・ 産業領域全体 ・ 複合的な事業全体
フル・データサイエンティスト	・ 棟梁レベル	・ 対象組織全体
アソシエイト・データサイエンティスト	・ 独り立ちレベル	・ 担当プロジェクト全体 ・ 担当サービス全体
アシスタント・データサイエンティスト	・ 見習いレベル	・ プロジェクトの担当テーマ
データ使い	・ 賢くデータを器用に使える人	・ 担当業務
一般人	・ 特になし	

資料：データサイエンティスト協会プレスリリース（2014.12.10）<http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

DataScientist Society Copyright © 2015 The Japan DataScientist Society. All Rights Reserved. #ds_symp2nd 17

↑データサイエンスを専門的に学ぶ大学教育の展望
（出典：滋賀大学HP 滋賀大学の目指すもの）

↑出典：データサイエンティスト協会資料

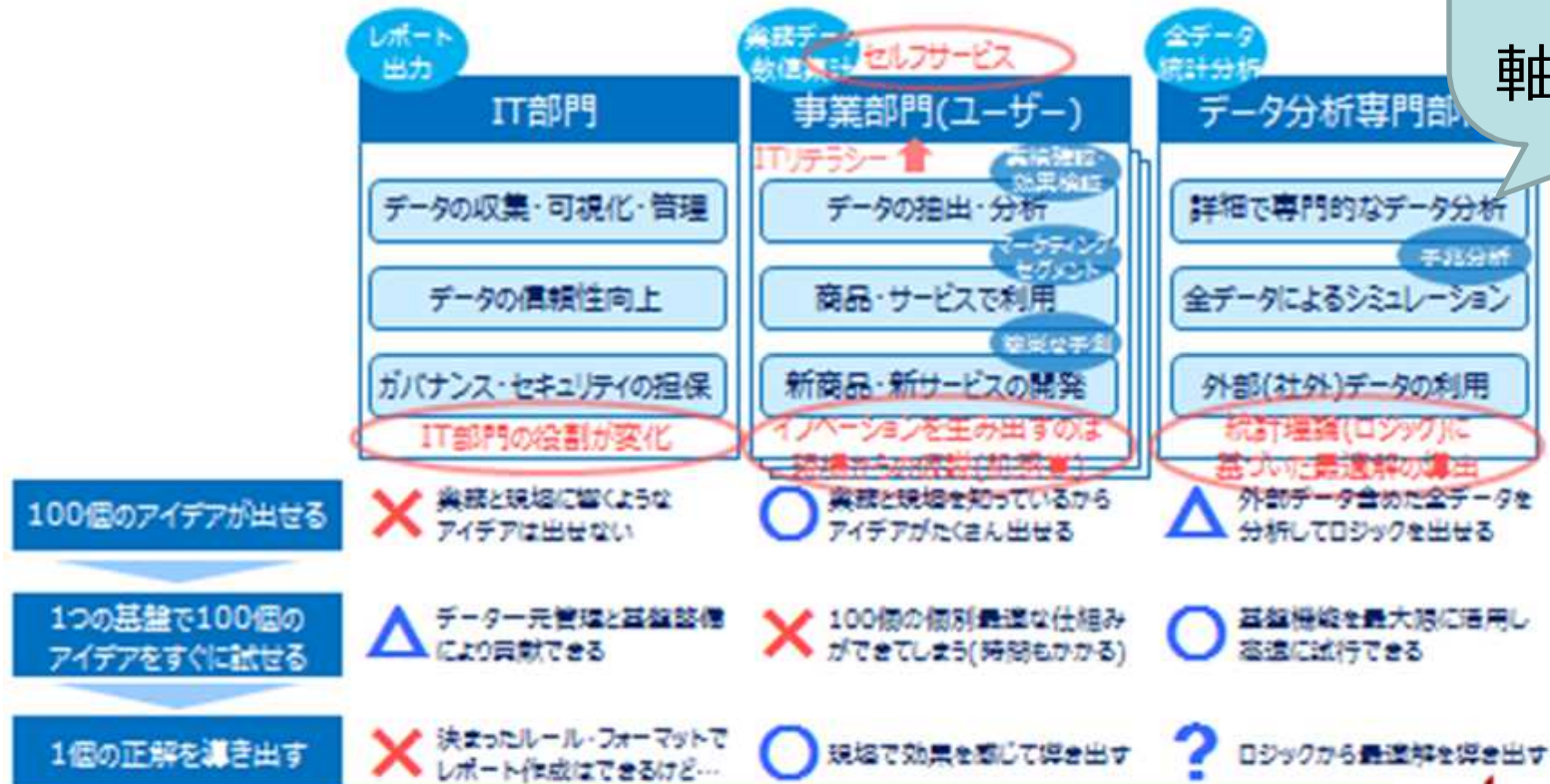
【DS1】INPUT2（横軸）

2月7日データサイエンス研究会データマネジメント研究会交流会にて

データドリブン経営におけるデータ利活用の役割分担

データドリブン経営という状態では、IT部門は裏方になる。事業部門(ユーザー)はセルフで担当業務の集約を実施し、データ分析専門部門はテクノロジーを活用し、新たな分析を実施する。
データ分析専門部門は、予兆分析などから経営に直結する方針を示す役割を担う。

DM第3からの
資料を基に、
人物像の役割
分担（横
軸）を明確化



【DS1】成果物①スキルマップ「3 × 3」

人物像と役割の資料をもとに、それぞれのレベル・役割で求められるスキルマップを作成しました。9つに分類し、個々の業務レベルとキーとなるスキルについて深堀しています。

	I T 部門	事業部門（ユーザー）	データ分析専門部門
役割	・データ収集・可視化・管理 ・データ信頼性向上 ・ガバナンス・セキュリティの担保	・データの抽出、分析 ・商品、サービスで利用 ・新商品、新サービスの開発	・詳細で専門的なデータ分析 ・全データによるシミュレーション ・外部（社外）データの利用
棟梁	①	④	⑦
一人前	②	⑤	⑧
見習い	③	⑥	⑨

【DS1】スキルマップ⑨『データ分析部門 × 見習い』



◆業務レベル	◆キーとなるスキル
・統計数理の基礎知識を有している （代表値、分散、標準偏差、正規分布、条件付き確率、母集団、相関など） ・データ分析の基礎知識を有している -検定（帰無仮説、対立仮説・・・） -分類（教師あり分類、教師なし分類・・・） -予測（回帰係数、標準誤差・・・） - など ・適切な指示のもとに、以下を実施できる -基本統計量や分布の確認、および前処理（外れ値・異常値・欠損値の除去・変換や標準化など） -前処理後のデータに対し、抽出、集計、可視化（ヒストグラム、散布図など）	統計数値基礎(分散、正規分布、相関など)、単回帰分析、帰無仮説・対立仮説、教師あり分類・教師なし分類の説明、ヒストグラム、クロス集計表、散布図、指示の元クレンジング処理 

※一般社団法人データサイエンティスト協会スキルチェックリストより引用

【DS1】成果物②スキルマップによる各社の現状認識と 目指す姿のGAP分析

現状				おおよそ5年後に目指す姿			
	IT部門	事業部門	分析部門		IT部門	事業部門	分析部門
棟梁				棟梁			
一人前				一人前			
見習い				見習い			

① As-Is

② To-Be

③ GAP分析と実現プランの策定

- A : 質・人数・組織とも業務ニーズにマッチしており、複数の領域および継続的に価値創造をもたらしている
- B : 質・人数・組織のいずれかが業務ニーズに不足している、限られた領域で一時的に価値創造をもたらしている
- C : 質・人数・組織のすべてが業務ニーズに不足している、限られた領域で一時的に価値創造をもたらしている
- D : 質・人数・組織において欠落がある。データ活用を通しての価値創造の必要性を認識している。

【DS1】スキルマップ活用事例（ユーザ系IT子会社）

メンバー9社
分を作成

現状	IT部門	事業部門	分析部門	目指す姿	IT部門	事業部門	分析部門
棟梁	A	C	D	棟梁	A	B	A
一人前	A	B	C	一人前	A	A	A
見習い	A	A	B	見習い	A	A	A

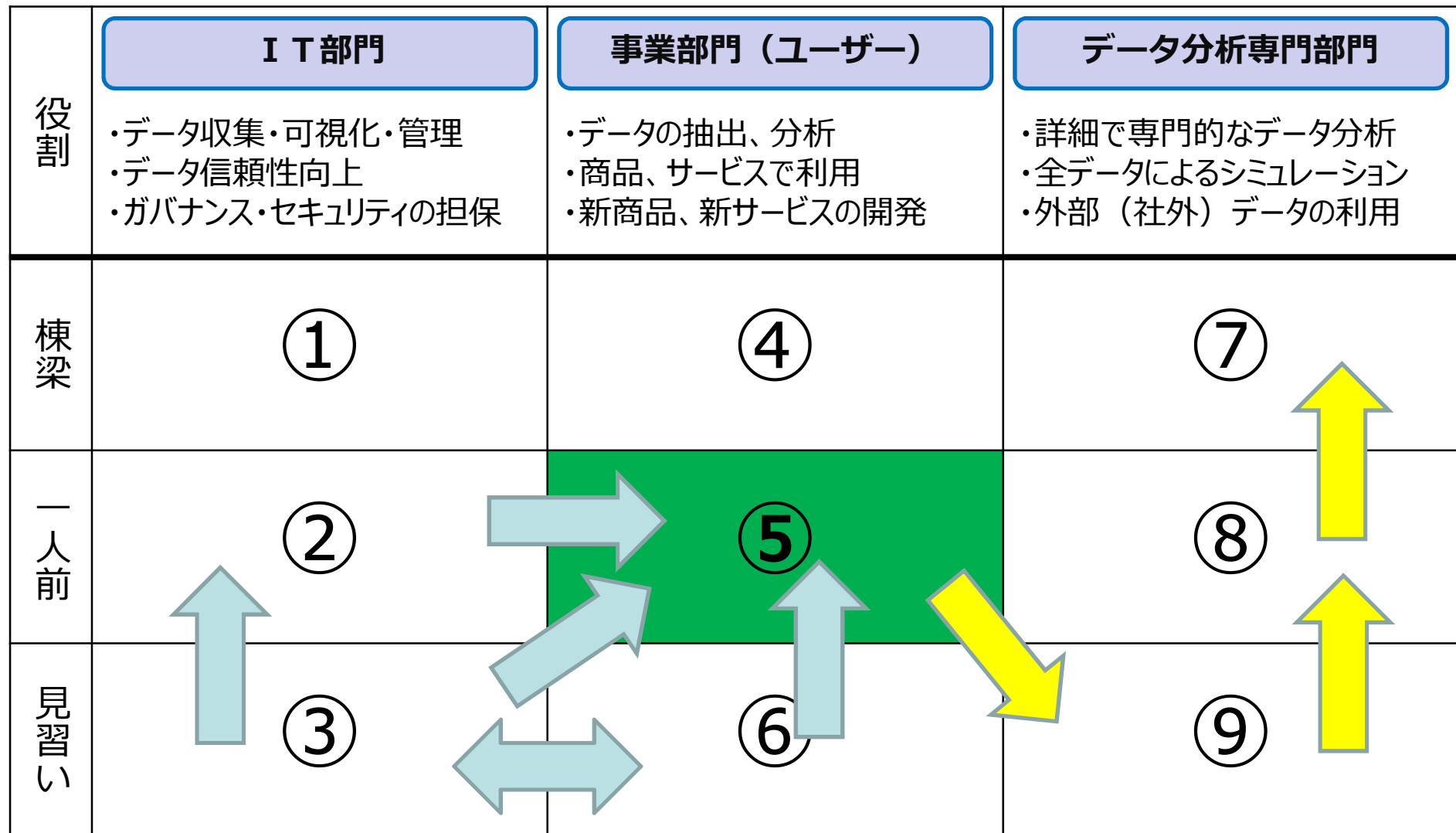
取り組むべき領域とその実現に向けた施策案 = 分析部門が目指す姿に向けて必要となるスキル

棟梁 D→A 見習い→一人前→棟梁と成長することができる事業領域の設定と、機械学習、AIの知識を習得

一人前 C→A 業務システムの一部としてのデータ活用業務を専門組織化しつつ、多変量解析などの標準的な分析手法を習得。顧客データ中心の領域を生産性向上に寄与する領域へ拡大。

見習い B→A 若手社員から選抜し、統計数値基礎や単回帰分析を習得しつつ、既存データ活用業務に従事する

【DS1】成果物③データ分析要員を育成するキャリアパス



【DS1】スキルマップを充足させるキャリアパスの考察①

3案を作成

おススメのキャリアパス

★★★③ ② ⑤ ⑨

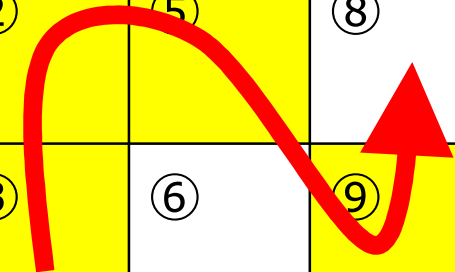
メリット

I T 部門を出発し事業部門で一人前になり、データ分析部門となるパターン
情報子会社人材の育成に

デメリット

事業部門で一人前になるとなると異動が困難か？

	IT部門	事業部門	分析部門
棟梁	①	④	⑦
一人前	②	⑤	⑧
見習い	③	⑥	⑨



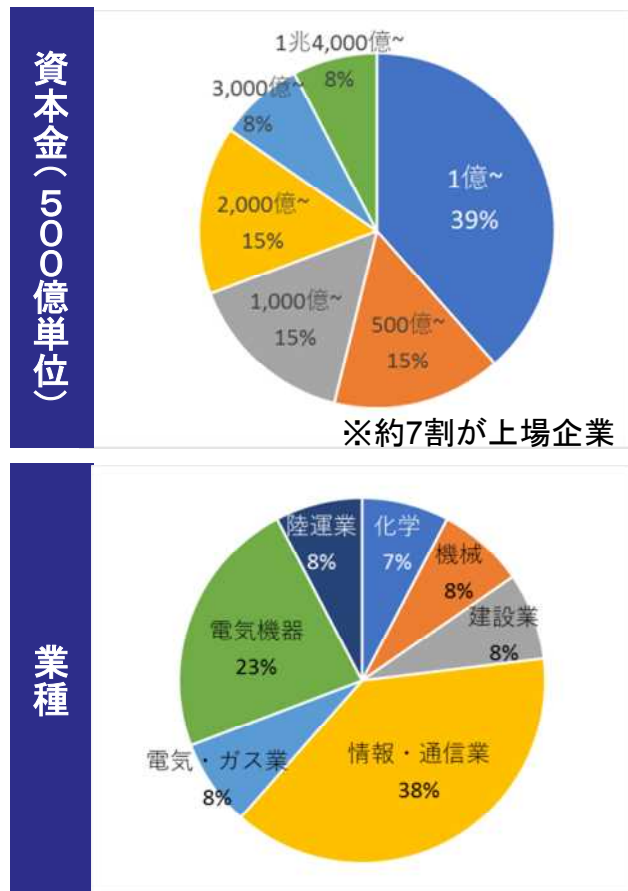
データサイエンス研究会 【DS2】

②事例研究・アンケート分析・ロールプレイからデータ利活用の気づきを得る

【DS2】 アンケートの実施

- アンケート内容：データ分析組織とデータ分析状況について
- 回答者数：17名 16社（DM/DS両研究会参加社） 分析組織有：13社
- 上場企業が多く、業種「情報・通信」が多い

■ 回答企業情報（全13社）

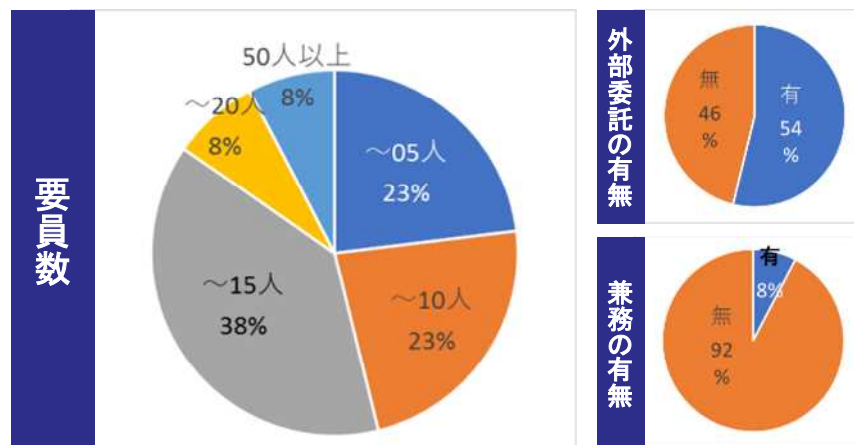
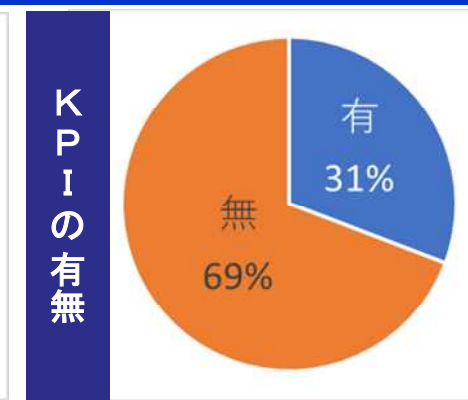
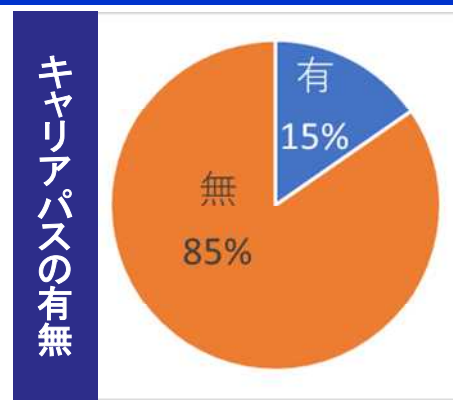
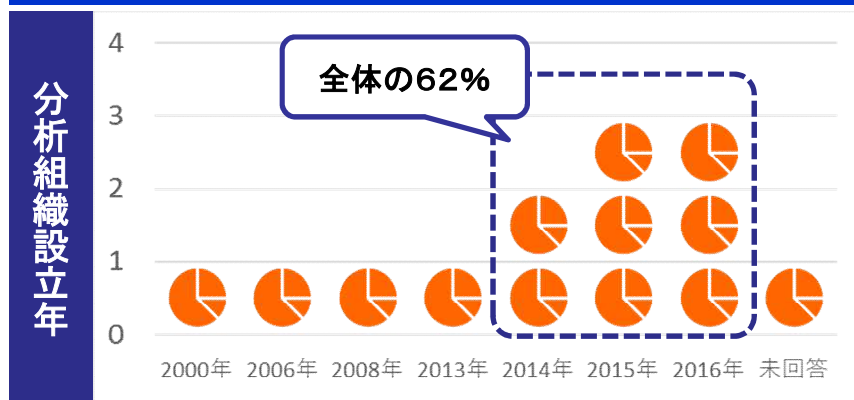


■ 主な質問内容

- データ分析組織に関する主な質問内容
 - ✓ データ利活用組織の名称
 - ✓ 設立年
 - ✓ 所属人数・兼務の有無
 - ✓ 組織の位置付け（全社横断・特定業務）
 - ✓ 育成・キャリアパスの有無
- データ分析実施に関する主な質問内容
 - ✓ データ分析業務の範囲
 - ✓ データ分析で利用しているデータ
 - ✓ データ分析で利用しているツール
 - ✓ データ分析で利用する分析手法
 - ✓ 苦労した点・困っている点

【DS2】 アンケート 分析組織

- 2014年から2016年で約60%を占めており、所属人数は10名以下が全体の約45%となっていることから、比較的新しく小規模な組織が多いといえる。
- 育成・キャリアパスを持つ企業は全体の15%であり、人材育成方針は確立されていない場合が多い



【研修実施内容】

部内輪講、部外セミナーへの参加
統計検定受験
外部研修（データ分析スキル習得研修、ツールベンダ研修など）

【KPIの内容】

収入向上額、コストダウン額
P o C 実施件数
稼働率
ツール獲得金額
一定規模の事案はユーザ部門協業型のプロジェクトとし、その推進管理（各タイミングでのオーナーレビュー含む）を仕組み化している。またその成果をスケジュールと合わせて評価している。

【DS2】 アンケート 分析状況

- データ分析がメインで、ビジネスに繋がる施策実施までは行っていない。
- 利用ツールは標準的もしくは無料のものが多く、スモールスタートで分析を行っている。

TOP3

■業務範囲



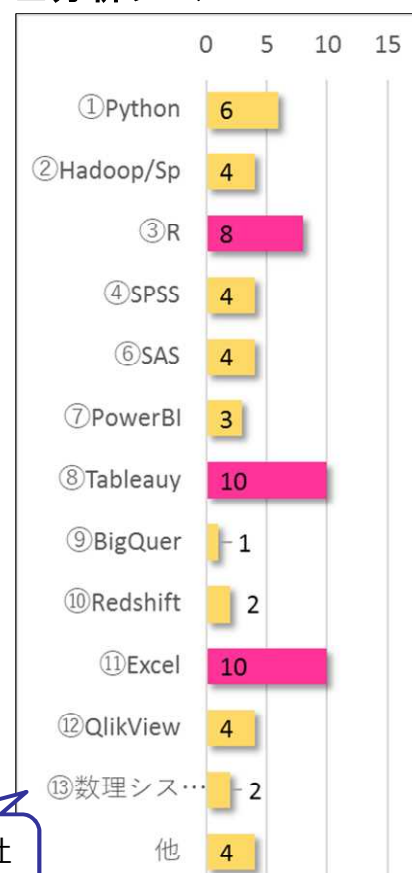
MAX:7(⑥を除く) 3社
MIN: 2 (③④) 1社

■分析対象データ



MAX:7(⑥を除く) 3社
MIN: 2 (③④) 1社

■分析ツール



MAX:10 1社
MIN:2 2社

■分析手法



MAX:13 1社
MIN:1(他 独自) 1社

【DS2】まとめ

データ整備の重要性

- 「データポリシーが統一されていない」「データが部署ごとに個別管理されている」など、データ整備に関する課題を抱えている企業が多いことが分かった。
- データ分析で効果を上げるためにはデータ整備が必須となる。その為、データ管理を担うことも社内横断組織として取り組むべきである。

データ分析推進の為に

- データ分析は結果が見えづらいことを考慮すると、必ずしも計画時に費用対効果を示すことが難しい。
- 事例でもあるように、成功を約束出来なくても多くの分析を手がけている。案件を数こなすことにより、組織力強化が進み、その中で成果が出て来るものと思われる。

データ分析組織に求められる人材像

- データ分析力だけではなく、データ分析組織にはビジネス力なども含めた総合力が必要となる。
- 人材を育成する為には現状と目標のギャップを正しく把握し、育成計画を立てる必要がある。その為には、スキルレベルの定量評価が有効である。
(参考) データサイエンティスト協会のスキルチェックを活用している企業が多い

データサイエンス研究会 【DS3A】

③ 明るく楽しく元気よく手を動かして
データ活用を実践してみる

【DS 3 A】 研究テーマについて

【目的】

ツールや手法を学ぶことをモチベーションにして分科会に参加している人が多い。

➡ 自業務に持ち帰って役に立つ知識を得て、意味ある活動にすること

役立つ知識とは？ → 分析ツールや手法を知ること

意味ある活動にするには？ → みんなで調べて、みんなで共有すること

【実施事項】

① データレイク作成

自社のデータを整理する。

自社データについて、データ名や項目例を列挙する。

サンプルが提示できるものはアップロードする。

② 分析手法・ツール探索

世の中にある分析ツールや分析手法を整理する。Google検索や書籍を使用。

実施事項①と組み合わせて、自分の手を動かして分析できそうなものを見出す。

【DS3A】実施事項① D a t aの一覧

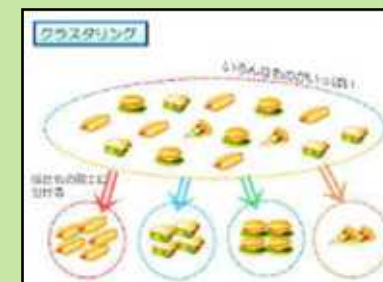
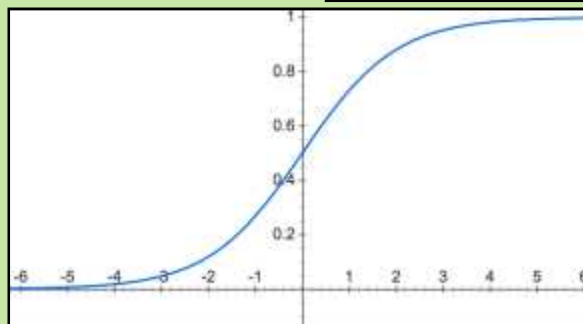
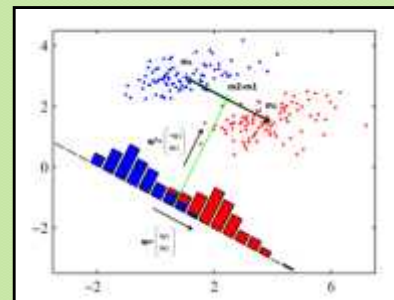
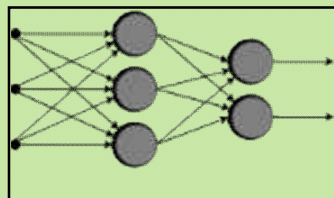
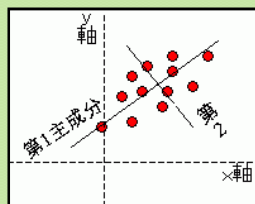
Data

機密データをマスクすることで、手を動かし分析することができる情報を集めた

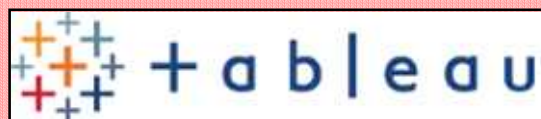
No	データ名称	内容	データ項目数	件数
1	施設予約記録明細	施設予約情報、キャンセル情報	53項目	50,886件
2	システム・アクセスログ	社内システムへのアクセス情報	5項目	584,345件
3	イベント発生明細	イベント情報明細	20項目	892件
4	製造工程品質明細情報	製造工程の品質OK/NG結果明細情報	470項目	6956件

【DS3A】実施事項② 分析手法・ツール探索

Methods



Tools



【DS3A】実施事項② 分析手法・ツール探索 成果物

領域	方法	概要
時系列系	線形近似(最小2乗法により)	線形に近い単純のデータに線形近似する方法
	移動平均法	一定の区間(期間)を定め範囲をずらしながら平均をとって
	ARIMA	自己回帰和分移動平均モデル
	Step Wise ARIMA	AICを最小化するように自動的に最適ARIMAモデルを選択
	1次指数平滑	単純指数平滑
	2次指数平滑	1次指数平滑ベースで上下をトレンドを加味した指数平滑
	Holt Winters法(3次平滑)	2次指数平滑ベースで周期性を加味した指数平滑
	Croston法	欠損が断続的に発生するデータに特化した補正付き平滑法
	Gray Model	灰色理論に基づいたモデル
	最近隣法	個体間の距離の最小のものをクラスター間の距離と定義す
判別系	線形判別	線形モデルに基づいた判別モデル
	ロジスティック判別	ロジスティックモデルに基づいた判別モデル
	記憶ベース推論法(K近傍)	多数決とも呼ばれる。
	決定木	決定木学習はデータマイニングで良く用いられる方法でも
	ランダムフォレスト	決定木の集合、複数の決定木に合成されるため、目的変数
	ニューラルネットワーク	脳の中には多数のニューロン(神経細胞)が存在しています
	サポートベクターマシン	サポートベクターマシンは、線形入力素子を利用して2クラス
	射影追跡回帰	実際の現象を統計的に解析する目的の一つは、種々の特
生存時間分析	ベイジアンネットワーク	因果関係を確率により記述するグラフィカルモデルの1つで
	ハザードモデル	このモデルは、生存時間を目的変数とした回帰モデルであ
バスケット系	アソシエーション分析	バスケットの中の商品間の関連性について分析を行う方法
次元集約	主成分分析	多変数解析の際に次元を集約するときに使う手法
クラスターリング	最短距離法	2つのクラスターのサンプル同士で最も小さいサンプル間距離
	最長距離法	最短距離法の逆で各クラスター中、最大のサンプル間距離
	群平均法	各クラスター同士で、全ての組み合わせのサンプル間距離
	ワード法	2つのクラスターP,Qを結合したと仮定したとき、それにより利
	k-means法	1. 各点にランダムにクラスターを割り当てる 2. クラスターの重
予測&要因分析	数量化理論I類	数値では表現しない事象を利用した商品の売上予測など
	数量化理論II類	数量化2類は判別分析と非常によく似た手法です。目的変
	数量化理論III類	数量化3類は数多くある観察変数からいくつかの潜在変数
	重回帰分析	単回帰分析が、1つの目的変数を1つの説明変数で予測し
	一般化線形モデル	一般化線形モデルには線形回帰、ポアソン回帰、ロジステ
	ロジスティック回帰	ロジスティック回帰分析とは、YESとNOを明確に定義して発

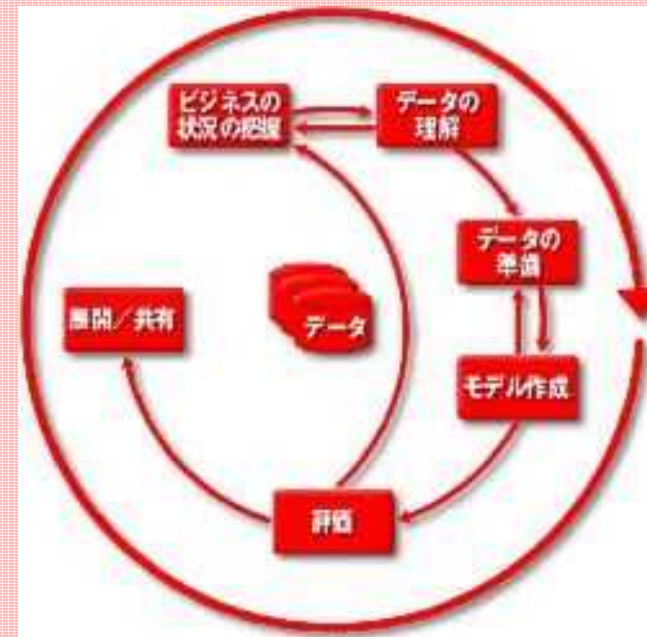
【DS3A】実施事項② 分析手法・ツール探索 成果物

分析手法	概要	使用例、使用目的	分析を実現できるツール			
			R	python	Bluemix	Azure
判別分析	個体(対象者)の特性(回答データ)から、その個体(対象者)がどの群に属するかを判別する手法	- 各種の検査項目から、ある病気であるかないかを判別する。 - 各種アンケート結果から各々の消費者が、製品Aを選ぶか製品Bを選ぶかを予測する。 - 筆記試験、面接、作文、適性検査など多面的な試験項目からなる入社試験、入学試験において合格群と不合格群を判別する。 - 文学作品における名詞や助動詞などの使用頻度から作者が分からない作品の作者を推定する。 - 重金属の含有比から、出土品の金属がどの鉱山のものか推定する。	○			
バスケット分析 (アソシエーション分析)	複数の事象の関連性を見つけ出す手法	ある商品を買う人は、こういう商品を合わせて買うという傾向を見つけ出す	○			
相関分析	2変数間の関係を数値で記述する分析方法	Aという事象が発生するならばBという事象も発生する、といった事象の相関を見つけ出す	○			
クロス集計分析	2つの項目に注目して同時に集計するデータ分析手法	アンケート調査における設問に対し、回答者の属性などの項目を交えて、集計結果を作成し、傾向を分析する	○			
決定木分析	樹木状のモデルを使って要因を分析し、その分析結果から予測をおこなう	年齢や趣向などの分析要因から、マーケティングのターゲットを抽出する	○			
因子分析	複数の変数間の共通因子を見つけ出すことで関連性を知る分析手法	アンケート調査で定義した数多くの説明変数の中から、潜在的な共通性を見出し、同じ意味合いをもつ説明変数(因子)のグループに分類・集約	○	○		
ABC分析	「重点分析」とも呼ばれ、主に商品や顧客などをその重要度によってランク分けするデータ分析手法	- 数多の商品に対して、それぞれどれくらい手間をかけて管理すれば良いのかの優先度をつける - 売上高の順に取扱商品を並べ、累積売上高割合が70%を占める商品グループをA、70%~90%の商品グループをB、90%~100%の商品グループをCという様に分類	○			
クラスタリング分析	異なる性質を持つ複数のものが混ざりあっている集団から、似た者同士でグループを作り分類するためのデータ分析手法	- ポジショニング確認を目的としたブランドの分類 - イメージワードの分類	○			
線形回帰分析	クロス集計で得られるグラフに、論理的に考え出された直線を引くというデータ分析手法	ある変数の変動をいくつかの変数の変動によって説明・予測する	○			
ロジスティック回帰分析	目的変数が0・1、または確率の係数に使用する。予測値や係数(効果)を、確率として直接表現する分析手法。		○			
グレイモデル	データのうち、明白なものを「白色」、不明なものを「黒色」、あいまいな状態を「灰色」と定義し、「白色」「黒色」のデータをもとにして「灰色データ(今後)」を予測するデータ分析手法	過去の数値データを基にグレイ法(灰色理論)により数値予測を行う				
時系列分析	時間の経過に沿って記録されたデータの、変化の傾向を分析する		○			
サポートベクターマシン	教師あり学習を用いるパターン認識モデル。回帰や分類に使用できる	パターン認識	○			
異常値検出	外れ値を検出する					
ソーシャルネットワーク分析	個人やグループ、あるいはその他の情報処理実体において、それぞれの間にある情報フローと関係を調査する手法					

【DS3A】CRISP-DMの実践

Data × **Methods** × **Tools**

要素は揃った！ → 方法論を探す！ → どうやらあるらしい！



各自、自分の扱いたいデータ・方法・ツールを選択し、
取り組むという方針が決まる

【DS3A】分析実践例 キャンセル要因分析(ロジスティック)

Data

施設予約データ

Methods

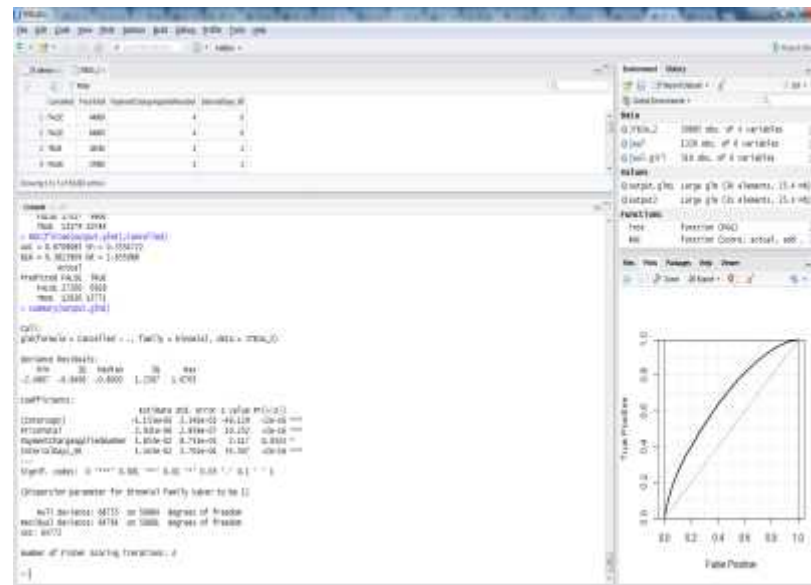
予約がキャンセルされる要因について、ロジスティック回帰分析を使って分析

Tools

R

Result

「ロジスティック回帰分析」とは何か、手順や雰囲気进行学习することはできた。Rが大変強力なツールであることは理解できた。



Consideration

業務で分析を行う場合は、分析結果を検証しながらデータのチューニングとモデリングを繰り返して精度を上げるステップが必要で、その部分はまだ体験できていない。

【DS3A】分析まとめ

No	使用データ	分析内容	手法	結果	ツール
1	施設予約記録明細	キャンセル 要因分析	決定木分析	キャンセル率は 申込間隔とは 正の相関関係。 現地払いが キャンセルの可 能性が大きい	R
2			ロジスティック回帰 分析		R
3			-	予約の傾向、 取消しの要因、 傾向を可視化	PowerBI、R
4	製造工程品質明 細情報	品質OK・N G 要因分析	決定木分析	要因分析はで きたものの、結 果の解釈でき ず	Python
5	イベントデータ	発生位置分析	テキストマイニング	発生する具体 箇所が分かっ た	R

【DS3A】全員でデータサイエンス

データサイエンス力(統計力)

データエンジニアリング力(ITスキル)



ビジネス力
(業務知識)

※データサイエンス協会より

有識者からのコメント



分析によって得られた取消を左右する要因は以下となった。

元々認識していた要因

- ① (割愛)
- ② (割愛)

あまり意識していなかった要因

- ① (割愛)
- ② (割愛)

データサイエンス研究会 【DS3B】

③ 明るく楽しく元気よく手を動かして
データ活用を実践してみる

【DS3B】活動目標

活動目標 “分析の実践”

データ収集、モデリング、分析結果解釈までの一連の分析業務をメンバーそれぞれが実践すること

チーム内の期待効果

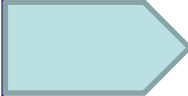
- 限られたリソース（時間、人数、データ）の中でどのような分析ができるか検証する
- 実際に手を動かすことで、ノウハウの獲得
- 先進的なツールを使ってみたい！

研究会内の期待効果

- チーム内でのノウハウを展開することで、分析のノウハウを獲得
- 先進的な分析ツールのできること / できないことを知ることができる

【DS3B】活動スケジュール

▼ 分科会、研究会

	2016年				2017年		
	9月	10月	11月	12月	1月	2月	3月
		Bluemix説明会 ▼	▼	▼	▼	▼	
分析シナリオ検討							
分析ツール選定							
データ準備							
テーマ1 分析作業							
テーマ2 分析作業							
最終報告準備							

【DS3B】たどり着いたテーマ

テーマ①

銀行顧客のキャンペーン反応予測



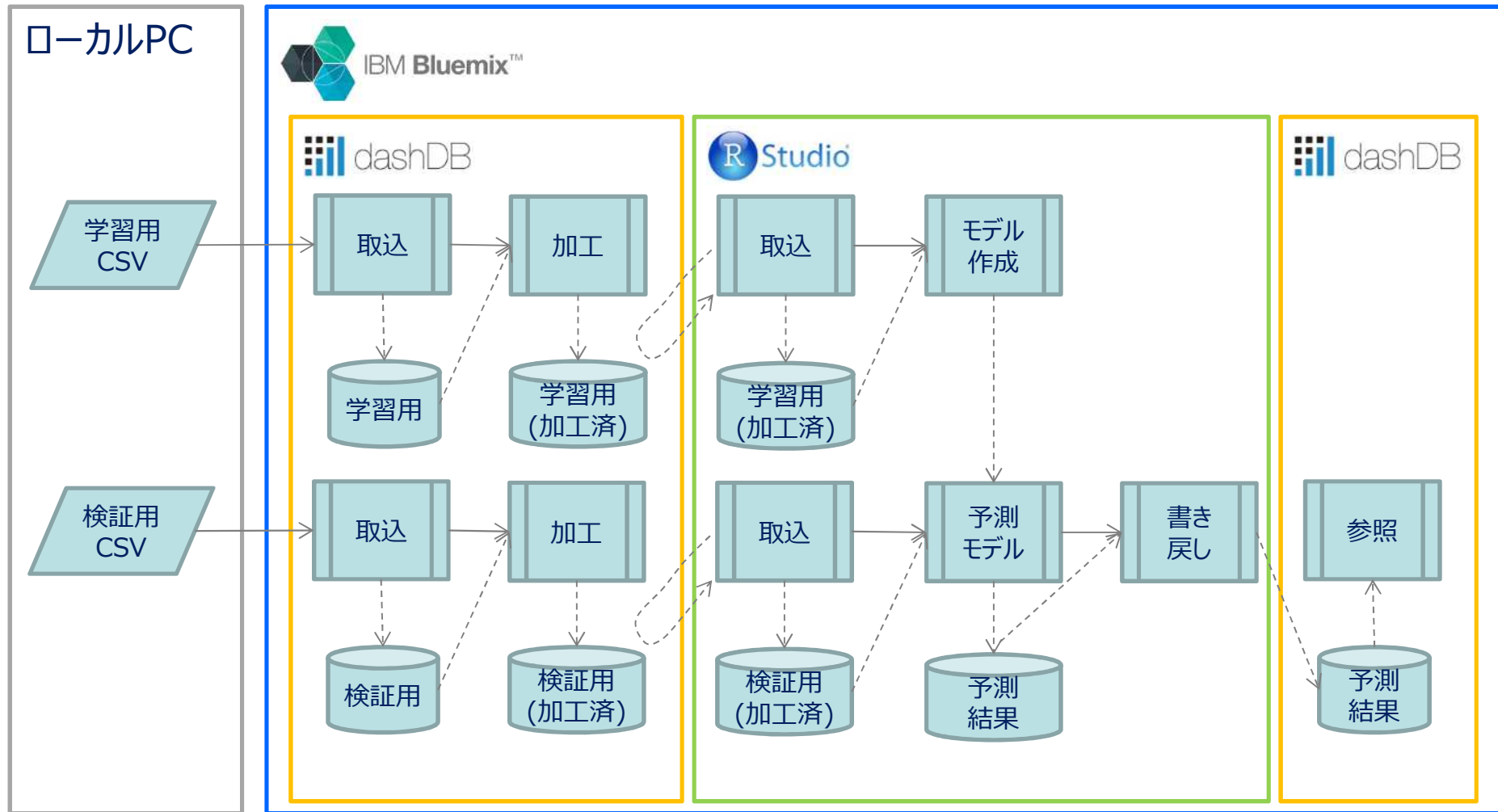
テーマ②

顔写真による年代判別（年代は0代、10代…80代）



【DS3B】銀行顧客のキャンペーン反応予測 処理フロー

本トライアルにおける処理フローは以下の通り。基本的にBluemix上で完結しています。



【DS3B】銀行顧客のキャンペーン反応予測 トライアル結果 – Dash-DB CSVからのロード

IBM dashDB™

0% dashDB-io tuozaip@gmail.com

Load from Desktop

Load one time from an Excel file or a delimited text file, such as a comma-separated value (CSV) file. [Learn more](#)

1. Specify source file 2. Choose the target 3. Select a table 4. Load complete

Transfer mechanism: Standard Aspera (BETA) [Install Aspera Connect](#)

Supported file types: CSV, XLS, XLSX

Drag a file here or [Choose a file](#)

File characteristics:

Does row one contain the column names? Yes No

Code page: 1208

Separator character:

☒ comma ☐ tab ☐ colon ☐ other:

【DS3B】銀行顧客のキャンペーン反応予測 トライアル結果 – R-Studio DashDBのテーブル取込

The screenshot shows the RStudio interface with the following components:

- Console:** Displays R session output, including package loading messages (RODBC, ibmdbR, MASS, grDevices, graphics, stats, utils, Matrix, arules) and the execution of R code to connect to a database and load data. The code includes comments and variable assignments for a data frame.
- Environment:** Shows the current environment with a data object 'dfJuas' containing 27128 observations and 18 variables. It also lists values for 'mycon' and functions like 'getCurrentBLUUserC... function ()'.
- Files:** Displays a file explorer view showing the current directory structure, including files like '.Rprofile', 'loadlogs', and 'RSamples'.

【DS3B】銀行顧客のキャンペーン反応予測 トライアル結果 – Dash-DB 結果テーブル

Create, drop, and work with tables

For existing tables, you can view details, browse data, and export data. [Learn more](#)

[Add Table](#) [Delete Table](#) Schema: Table Name: [Refresh](#)

[Table Definition](#) [Browse Data](#)

**無事予測値は算出できましたが
精度はイマイチでした。。。**

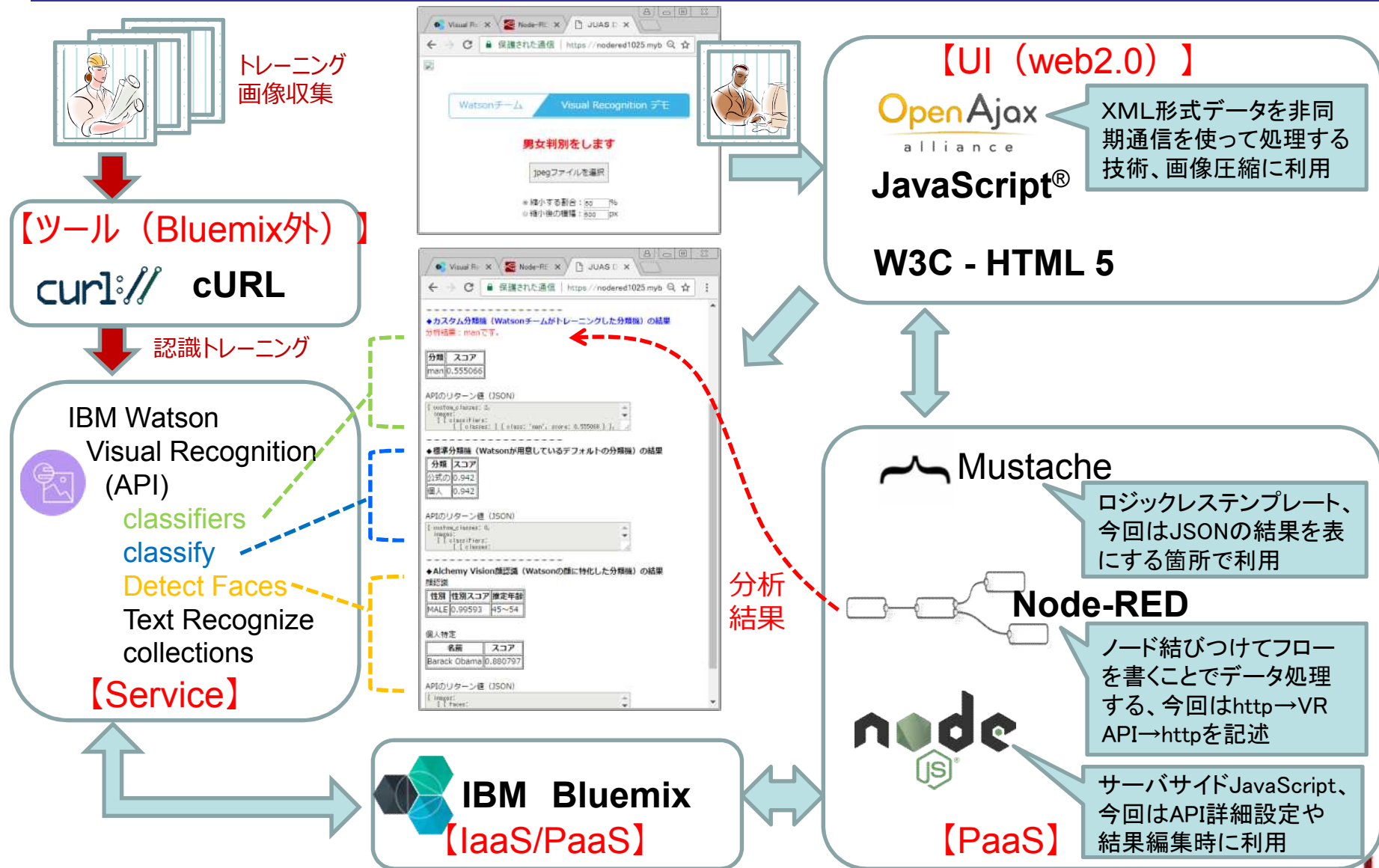
0	249	1	-1	0	0	0	0	1	0.009077320 505837361
0	328	1	-1	0	0	0	0	1	0.285270922 934156
0	445	2	-1	0	0	0	0	1	0.035365211 07933391
0	768	2	-1	0	0	0	0	1	0.467968910 09188673

合計: 1000 選択済み: 0

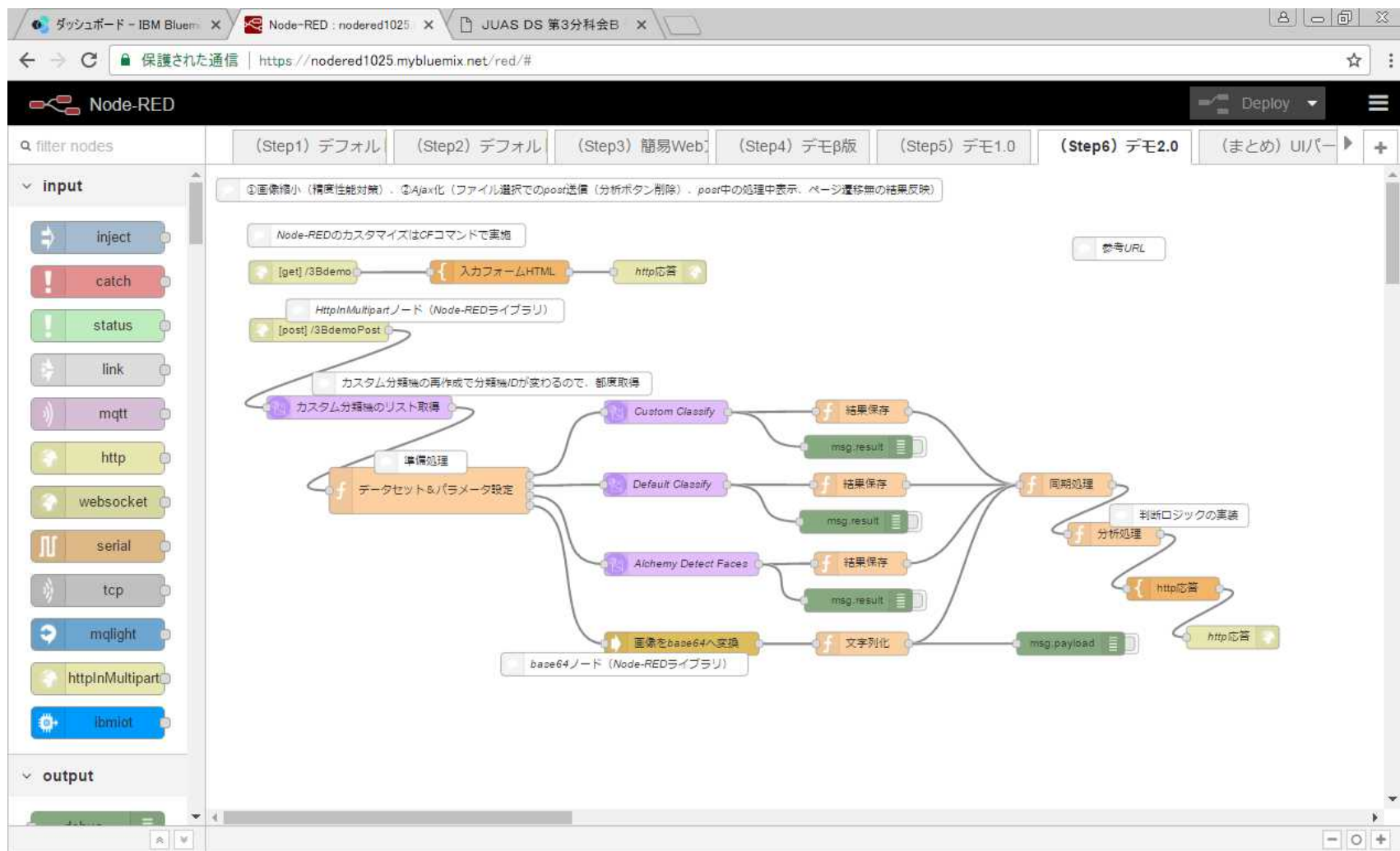
1 2 3 ... 40

10 | 25 | 5

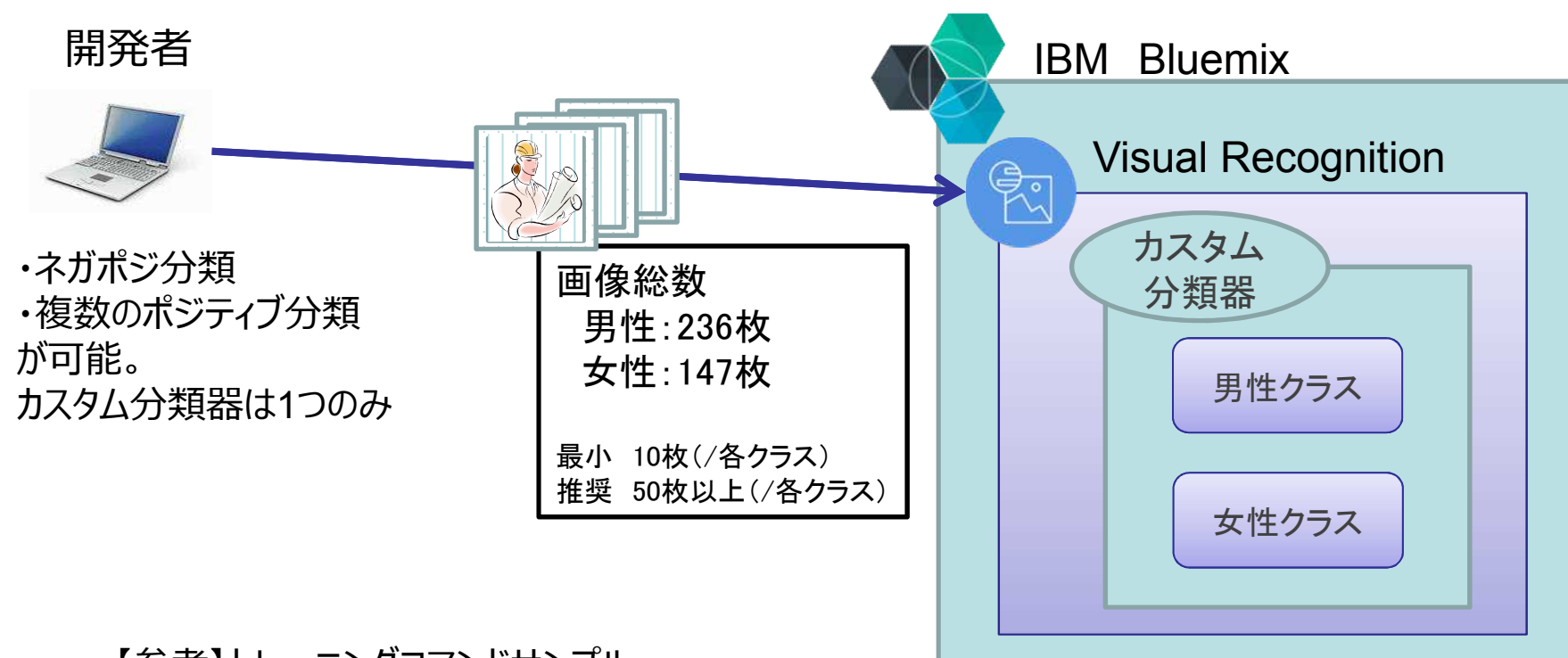
【DS3B】顔写真による性別判定 実践概要 – technology overview



【DS3B】顔写真による性別判定 UI



【DS3B】顔写真による性別判定 カスタム分類器の使用



【参考】トレーニングコマンドサンプル

```
コマンドプロンプト
C:\work\juas2016\man_woman>
C:\work\juas2016\man_woman>
C:\work\juas2016\man_woman>C:\app\curl\curl -k -X POST -H "Accept: application/json" -F "man_positive_examples=@man.zip" -F "woman_positive_examples=@woman.zip" -F "name=juas2016_3B2" "https://watson-api-explorer.mybluemix.net/visual-recognition/api/v3/classifiers?api_key=...&verbose=true&version=2016-05-20"
```

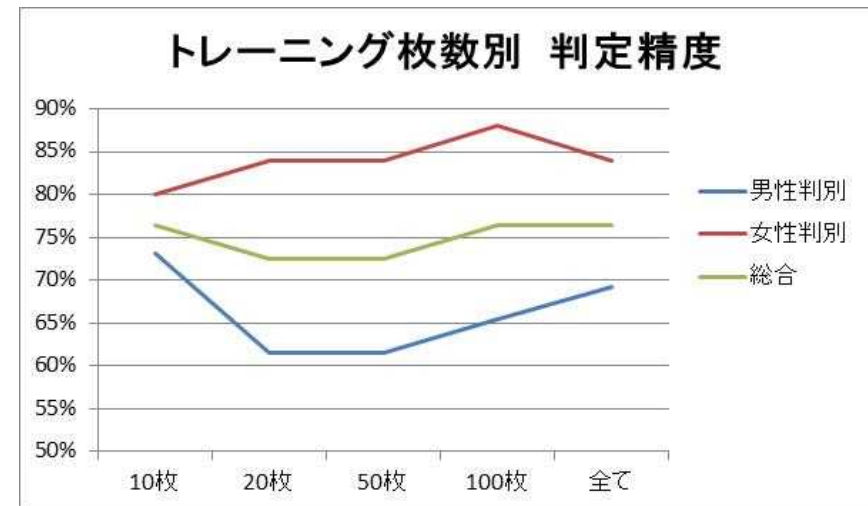

【DS3B】顔写真による性別判定 カスタム分類器 実行結果評価

学習画像の枚数に応じた判定精度向上を狙う

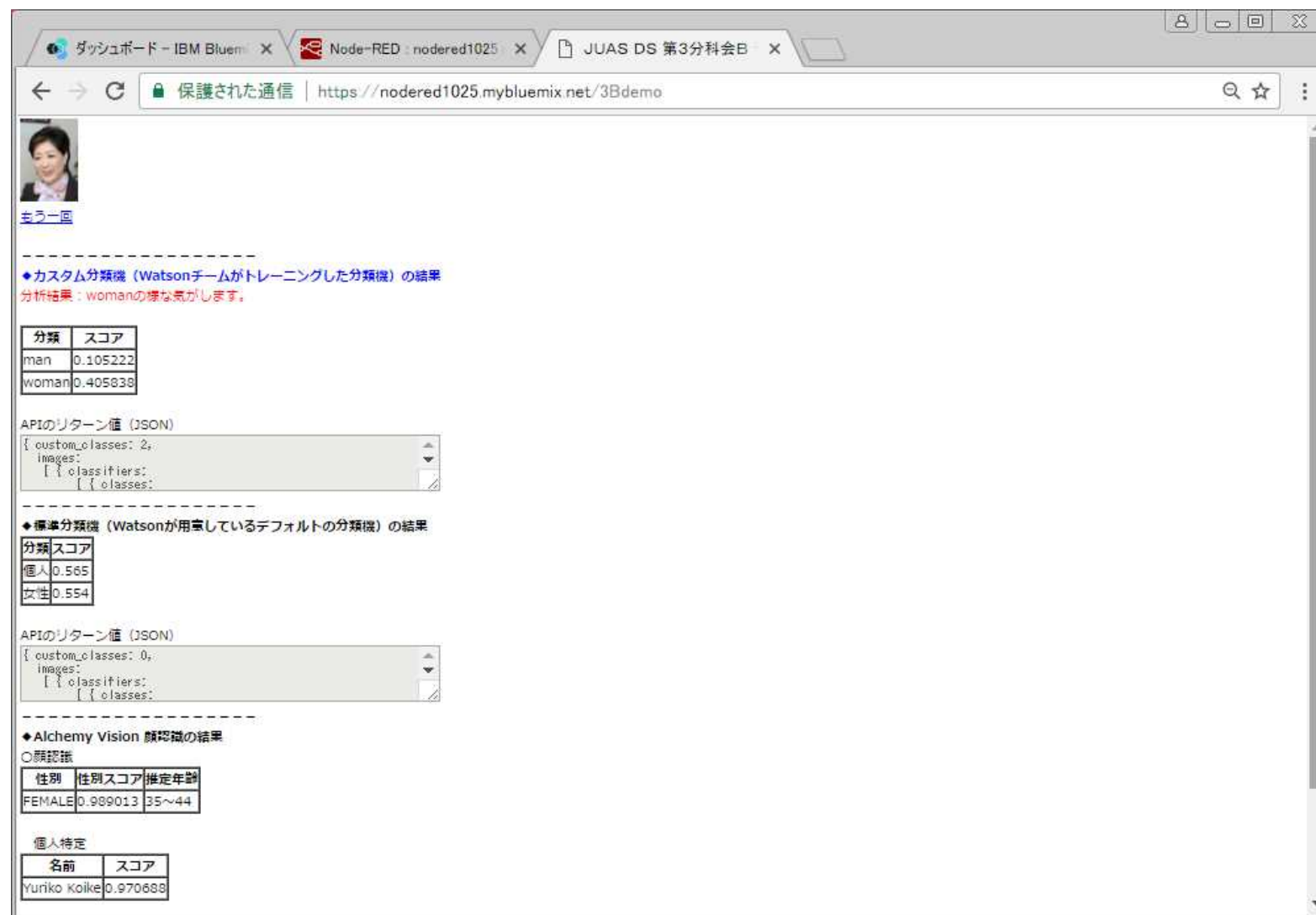
- ・学習画像を10枚、20枚、50枚、100枚、全ての5段階とし、男女判定精度が変化するかを検証
- ・学習画像のサンプリングは無作為に必要な枚数を抽出
(20枚実施時には10枚のサンプル画像を含む、50枚にも20枚のサンプルを含むという形で学習画像を設定)
- ・検証画像は学習画像とは異なる男女25枚の画像を判定する。
- ・検証画像も無作為にインターネット上の画像を利用

検証結果

- ・残念ながら本検証では学習画像増加による精度向上は見られなかった。
- ・画像全体での特徴点が抽出されるため、サンプル画像においても不要部の削除が必要？
- ・誤判定となった画像は中性的な顔立ちのイケメン男性や女性アスリート等、人が判定する際にも若干難しい(?)画像が多い。



【DS3B】顔写真による性別判定 実践概要 – アプリケーション実行結果



ダッシュボード - IBM Bluemix x Node-RED : nodered1025 x JUAS DS 第3分科会B x

保護された通信 | https://nodered1025.mybluemix.net/3Bdemo

もう一回

◆カスタム分類機 (Watsonチームがトレーニングした分類機) の結果
分析結果: womanの様な気がします。

分類	スコア
man	0.105222
woman	0.405838

APIのリターン値 (JSON)

```
{ "custom_classes": 2, "images": [ { "classifiers": [ { "classes":
```

◆標準分類機 (Watsonが用意しているデフォルトの分類機) の結果

分類	スコア
個人	0.565
女性	0.554

APIのリターン値 (JSON)

```
{ "custom_classes": 0, "images": [ { "classifiers": [ { "classes:
```

◆Alchemy Vision 顔認識の結果

☐顔認識

性別	性別スコア	推定年齢
FEMALE	0.989013	35~44

個人特定

名前	スコア
Yuriko Koike	0.970688

【DS3B】次年度に向けた提言

- 各サービスの無料枠で実践するのが非常に大変でした！

※今回はメンバーの井上さんの協力により、Bluemixの利用に制限がなかったので助かりましたが。。

→JUAS様のご協力を仰いでスポンサーを募って進めた方がよい！

- 解析や各ソリューションを実践していく上で、知見のある人がいなかったため、Googleなどで調べて進めるのが非常に大変でした！

※Bluemixについては、IBMのエバンジェリストの方に勉強会して頂き理解が進みました。。

→有識者にスポット勉強会を開催して頂けるような進め方がよい！

- 話題のPaaSで実際に構築できて非常にエキサイティングでした！

→次年度も是非ホットなソリューションを実践してほしい！

1年間、明るく楽しく、
手を動かすことができました！（^ ^）

2016年度研究会 参加メンバーの声

研究会参加アンケートから

- ◎ 他業種の方と協業でき、文化の違いを知ることができ、モチベーション向上につながった。
- ◎ データサイエンスの他社の取組状況を社内で発信し、今から始めても間に合うと判った

- ◎ 他社事例や講演で有益な話を聞くことができた。
- ◎ 昨年の研究内容から1歩進むことができた。

- ◎ データ活用について議論できる仲間がたくさん出来た。
- ◎ こんなにもフレンドリーで開かれた集まりだとは思っていませんでした。

- ▲ 講演を聞くだけで終わってしまう会が複数回あり、もったいないと感じた。
- ▲ 過去に何度も参加した方との温度感を感じ、少し馴染みづかった。

研究会
第六年度
予告

2017年度

JUAS

ビジネスデータ研究会



ビジネスデータ研究会のミッション

私たちは、データを事業価値につなげるために必要な
“考え方”、“技法・実行プロセス”、“組織体制・人材スキル”を
探求し、あらゆる事業従事者へ提案することによって、
データに携わる全ての人々が豊かな未来を描き
実現できる世界を目指します。

ビジネスデータ研究会の研究テーマ(案)

いまやデータ利活用は事業活動に不可欠であり、これからの事業経営もデータを基軸としたデータドリブン経営が提唱されています。本研究会では、これまで兄弟研究会として活動してきた、データマネジメント/データサイエンス両研修会を統合して、更なるデータ価値の探求に向けて、以下の3つのテーマで活動します。

データマネジメントのフレームワークと、そのために必要な組織・体制・人材

データサイエンスのプロセス・手法の調査と実証検証、未来のデータ利活用の探索

データドリブン経営に必要な要素と、実現に至るステップ

2016年度 データマネジメント／データサイエンス研究会
