

2017年度JUAS研究活動成果報告会

－ ビジネスデータ研究会 －

2018年4月26日

ビジネスデータ研究会

本日のアジェンダ

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

第二部 研究成果発表

ビジネスデータ研究会 活動概要

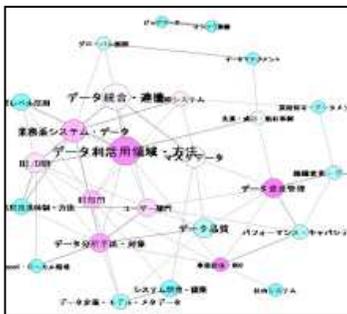
ビジネスデータ研究会のミッション

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。

研究会参加メンバーの課題認識の推移

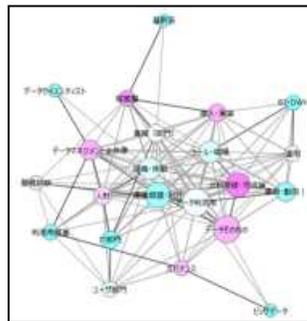
2013年度より研究会メンバーへの事前アンケート結果をテキストマイニング分析し、参加者の課題認識や研究会への期待を可視化した上で、毎年の研究計画を策定してきました。

【2013年度】



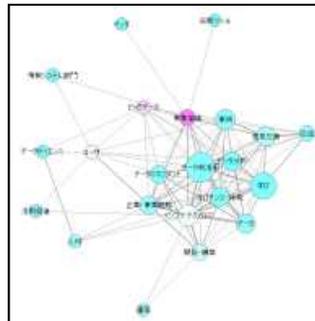
- ・データ統合・連携
- ・データ利活用の領域
- ・ビッグデータの事はまだあまり理解されていない

【2014年度】



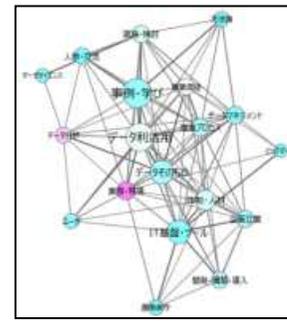
- ・データマネジメント
そのもの
- ・データの事業価値
- ・分析技法・技術
- ・データサイエンティストが出現

【2015年度】



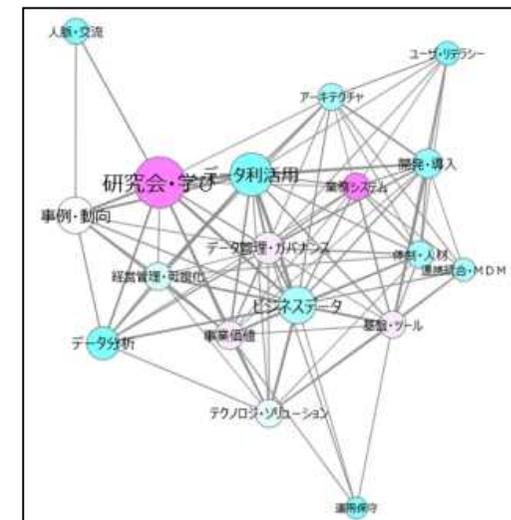
- ・データ利活用
- ・組織・体制・人材
- ・分析テクノロジー
- ・学び・交流といった要素が現れる

【2016年度】



- ・事例・学び
- ・データ利活用
- ・組織内での推進プロセスや運用など、より現実的な要素が増えてきた

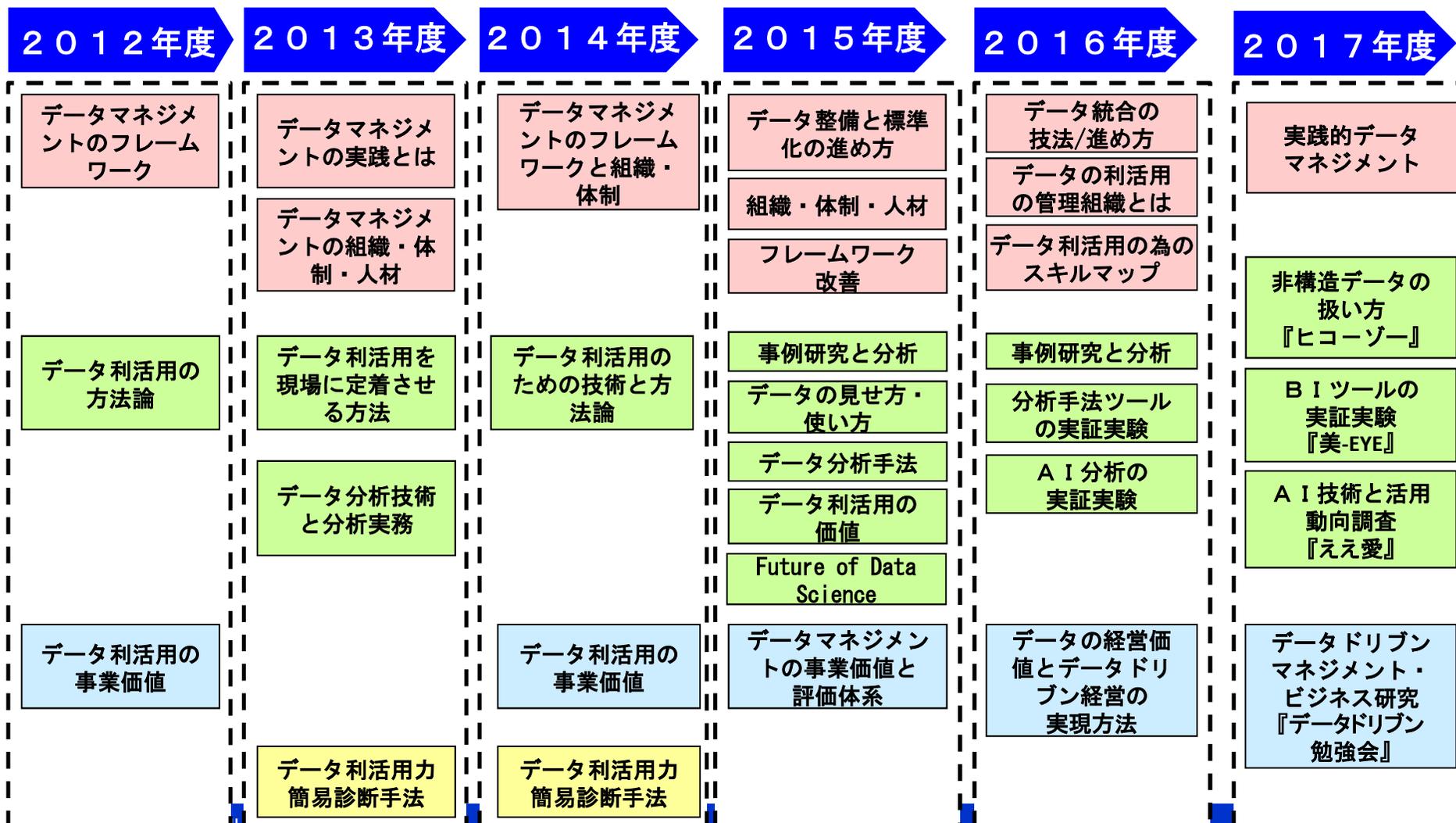
【2017年度】



- ・研究会・学び
- ・データ利活用
- ・データ管理・ガバナンス
- ・理論よりも、研究会でのコミュニケーションを通じた実践的な経験・知見の交換が求められている

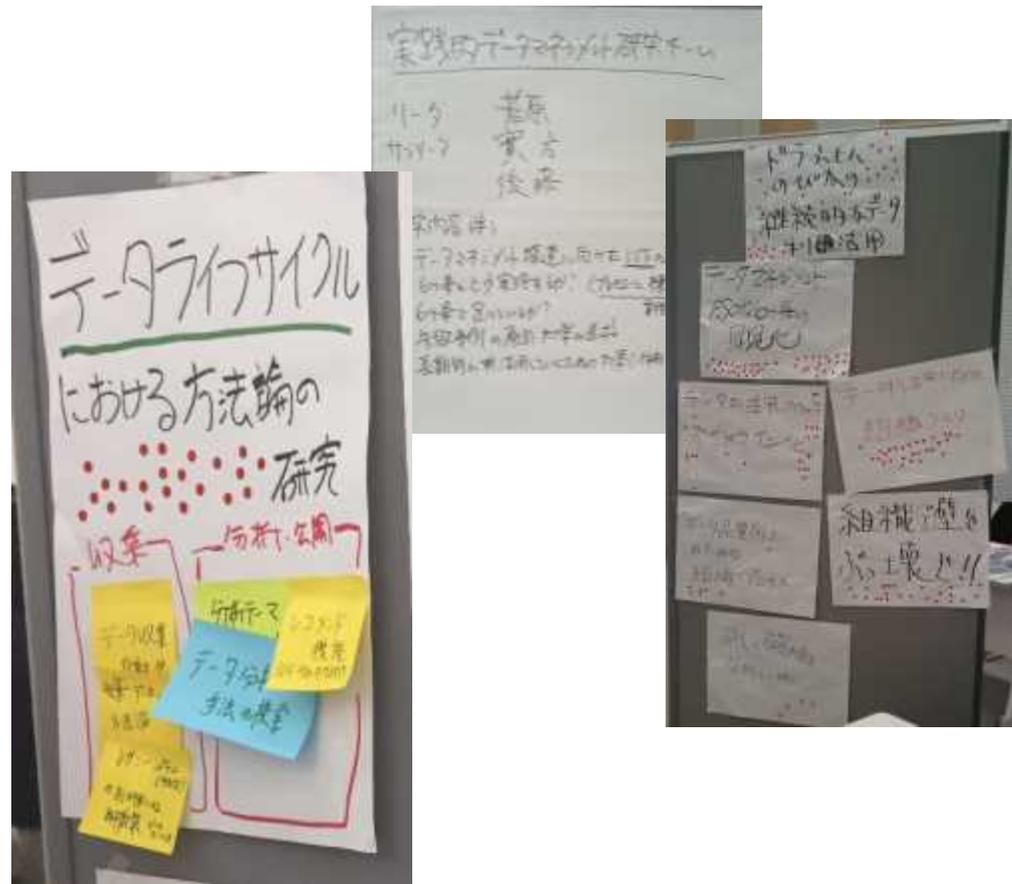
ビジネスデータ研究会の経緯

- ・2012～2016年度の研究成果を継承しつつ、5つの分科会に分かれて研究活動を推進。
- ・過去5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。

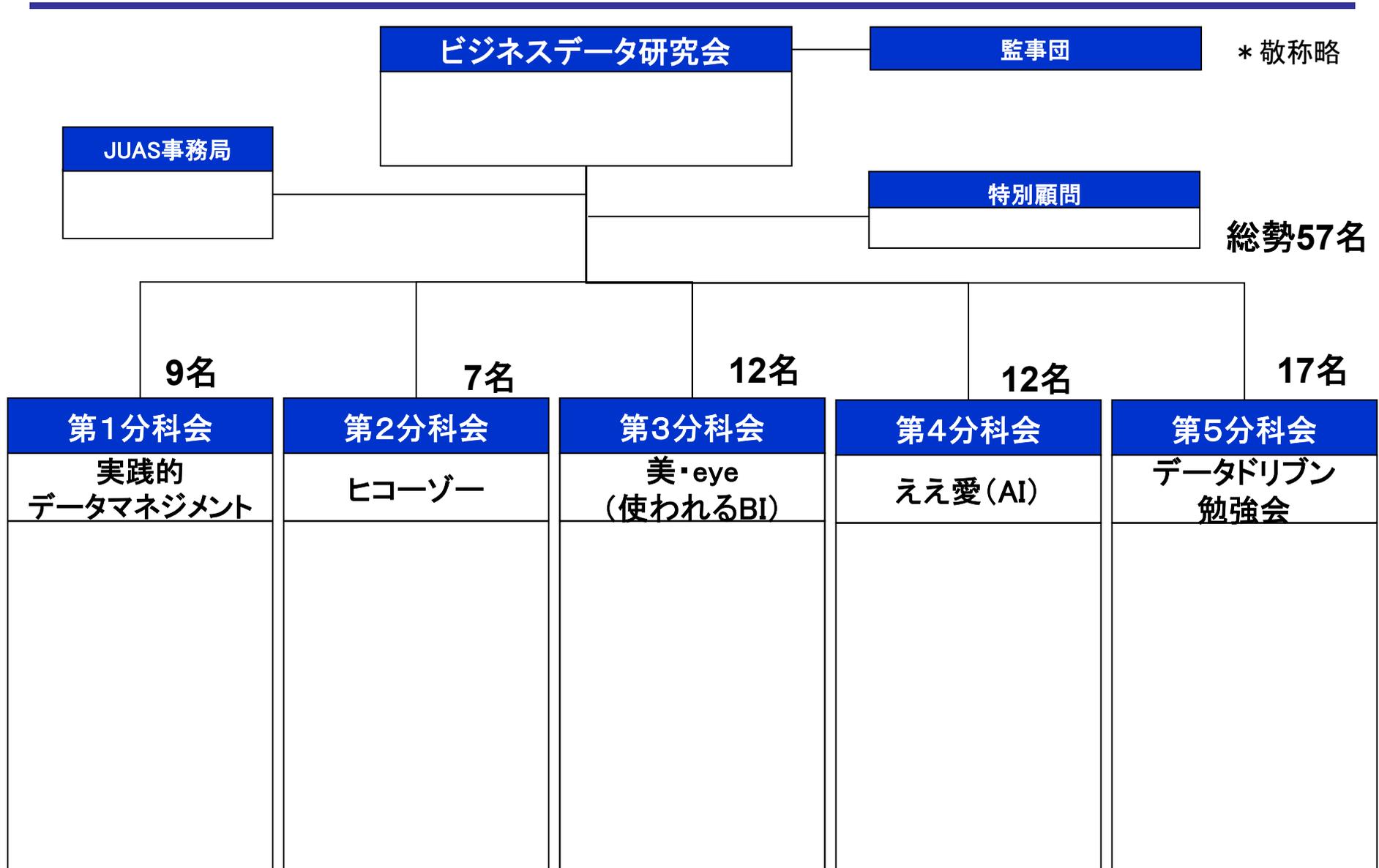


沼津合宿にて全員参加で研究テーマを検討

7月21～22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、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して一気に議論を深めました。



ビジネスデータ研究会の体制



2017年度 研究成果報告

…281ページありますので本日は「さわり」だけ。。

研究成果①『実践的データマネジメント研究チーム』

研究テーマ：「データの一元化と標準化に向けた具体策案」

成功要因が

成功企業からの成

其の一

其の二

其の三

其の四

其の五

其の六

Copyright (C) 2017 JUAS All rights reserved

其の三

データ基盤環境を整備せよ！
～マスターデータの一元管理、各システムへの自動配信が理想！

データ管理方針
データ管理方針にもりこむべき項目・観点、方針作成時に留意すべき事項を検討する。
※ 共通化、利便性、ガバナンス、明文化、継続性…

【データ一元化に関するテーマ】

其の四

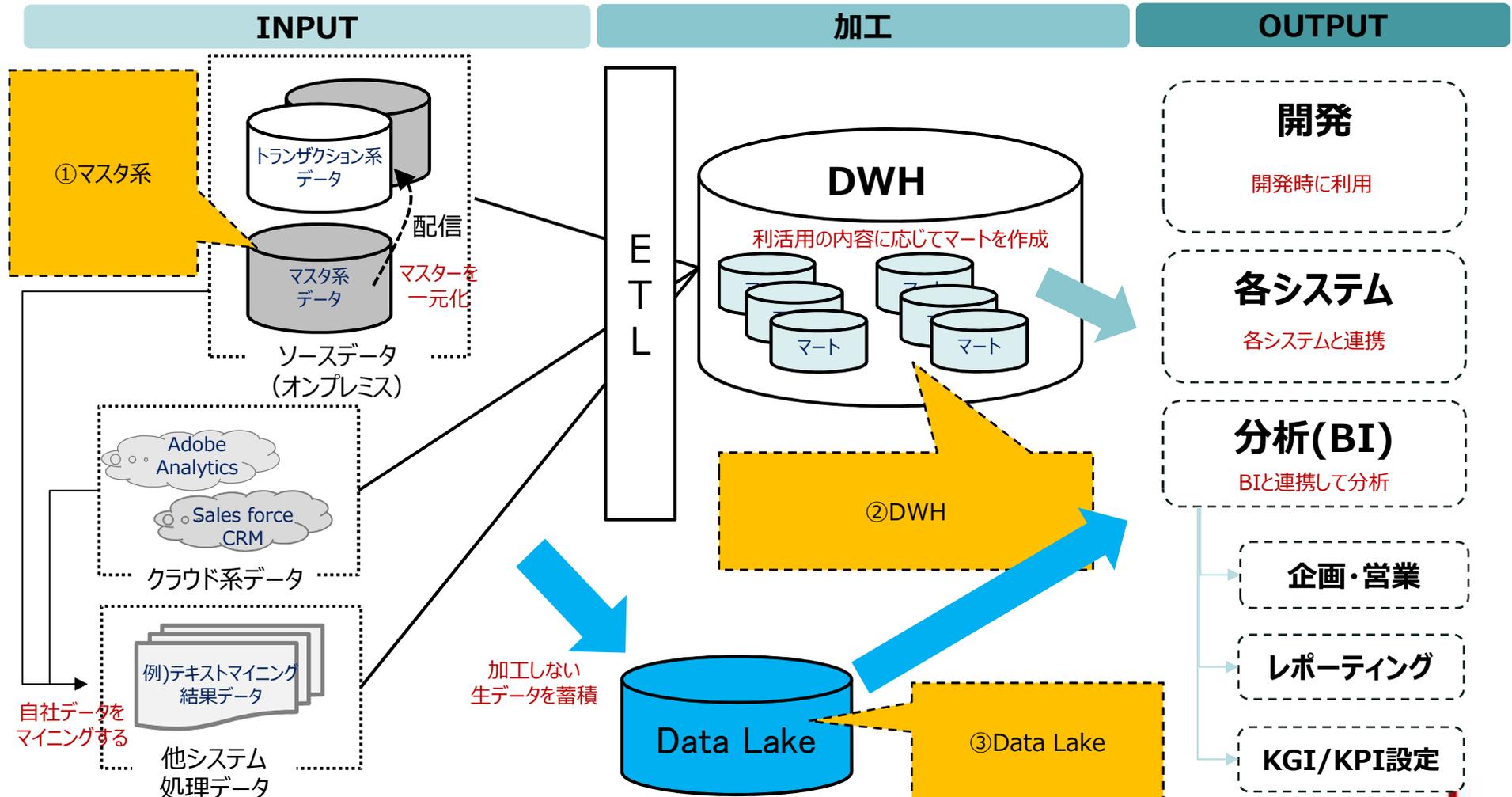
同音異義語、異音同義語が整理されたデータ辞書を整備せよ！
～ユーザーが自由に閲覧できる項目説明された画期的なもの。

データ辞書の運用プロセス/体制
データ辞書は、どのような体制で、どのようにメンテナンスし、全社的に浸透(普及)させていくのか、運用体制や運用プロセスを検討する。

【データ標準化に関するテーマ】

データ管理方針を考える上での3つの範囲

今回データ管理方針を考えるうえで、以下の様にINPUTフェーズの「マスタ系データ」、加工フェーズの「DWH」、「Data Lake」の3つの範囲にフォーカスして検討した。



データ管理方針で検討すべき観点とその達成方法について

データ管理方針として、データの品質や一元管理を意識した「設計」、共通化や利便性を考慮した「管理」、およびそれらを運用していく「組織」を重要観点として、それらを理想的な状態にするためのチェックリストを策定した。

データ管理方針に必要な3つの重要観点

設計

データの品質や一元管理、セキュリティ、利便性の確保のために実施すべきことを検討

管理

データ管理方針に盛り込むべき項目・観点、方針作成時に留意すべき事項を検討 ※共通化、利便性、ガバナンス、明文化、継続性

組織・体制

データ基盤環境を構築・保守・運用していくための組織・体制について検討

理想的な状態にするための確認項目

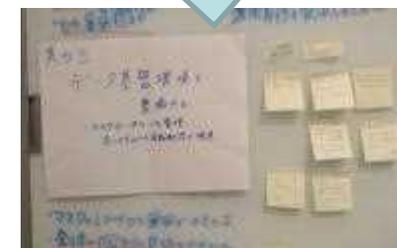
チェックリスト化

- ・データ抽出環境が統合され、容易にデータをシステム横断的に利用可能
- ・実際に必要となる形でデータが整備
- ・データマネジメントの実施単位が明確化

- ・マスタデータの利用ルールが明文化
- ・社内グループウェアが完全に移行され、データが一元化
- ・マスター運用業務の中で必要な部分が共通化
- ・メンテナンスや配信の仕組みが整備

- ・システム横断的に共有するマスタと、個別システムで管理するマスタが定義
- ・適切なセキュリティ対策も講じた、データ基盤環境が構築
- ・グループ会社を含めガバナンスの働くような仕組（基盤）・ルールが整備

このチェックリストで解決できる問題



メンバーの実際の問題をとりあげた



データ管理方針 チェックリストについて

データ管理方針として理想的な状態を構築するためのチェックリストを作成した。以下がその構成となる。



①		②		③			④		⑤	
1:設計 2:管理 3:組織	リスト No	チェックリスト (まとめ) 「～すること」		データ ソース	DWH	DATA LAKE	問題 関連No	このチェックで解消される問題		
1	1	<ul style="list-style-type: none"> 各システムのマスター運用業務の中で、共通化できる部分とその具体的な運用方法が明確になっていること。 各システムのマスター運用業務の中で、例外的に個別運用となる部分について、その理由と具体的な運用方法が明確にされ、適切な権限者の承認を受けていること。 		◎	○	-	28	<ul style="list-style-type: none"> システム毎にマスター運用業務が異なるため、同じデータを個々のシステム毎の運用ルールに沿って保持しなくてはならず、当該データを一元化できない。 		
							38	<ul style="list-style-type: none"> マスタデータの利用ルールが決まっておらず、システム連携がスパゲッティ化しているため、メンテナンスにより異常なデータが発生する。 マスタデータの利用ルールが明文化されておらず、一部明文化されていても秩序より開発費が優先される。 		
							39	<ul style="list-style-type: none"> システム連携のルールが明文化されていない。 マスタデータを配布する概念はあるが、PUSH（送る）、PULL（取りに来る）どちらがよいのか決めきれていない。 		
							40	<ul style="list-style-type: none"> 一元化できるものとできないものが柔軟に対応できるシステム設計になっていない。 一元化できるものとできないマスタの定義ができていないため、無駄なメンテナンス工数がかかる。 		
							44	<ul style="list-style-type: none"> 同じデータが各システムに散在し、システム毎で管理(更新)している。 正しいデータの姿を特定できず、ガバナンスが弱い。 		
							78	<ul style="list-style-type: none"> 個別システム毎に利用部門の現在のニーズに沿って個別にデータ抽出の仕組みを用意おり、不足分は個別依頼でIT部門が抽出している。 統合されたデータ利用環境がないため、システム横断的なデータ利活用が妨げられている可能性がある。 		



データ辞書の存在意義

目的

- 業務担当者、システム開発・保守担当者が業務を遂行する上で、**データを理解**するために必要な情報を提供する。

役割(データ辞書ツールの役割)

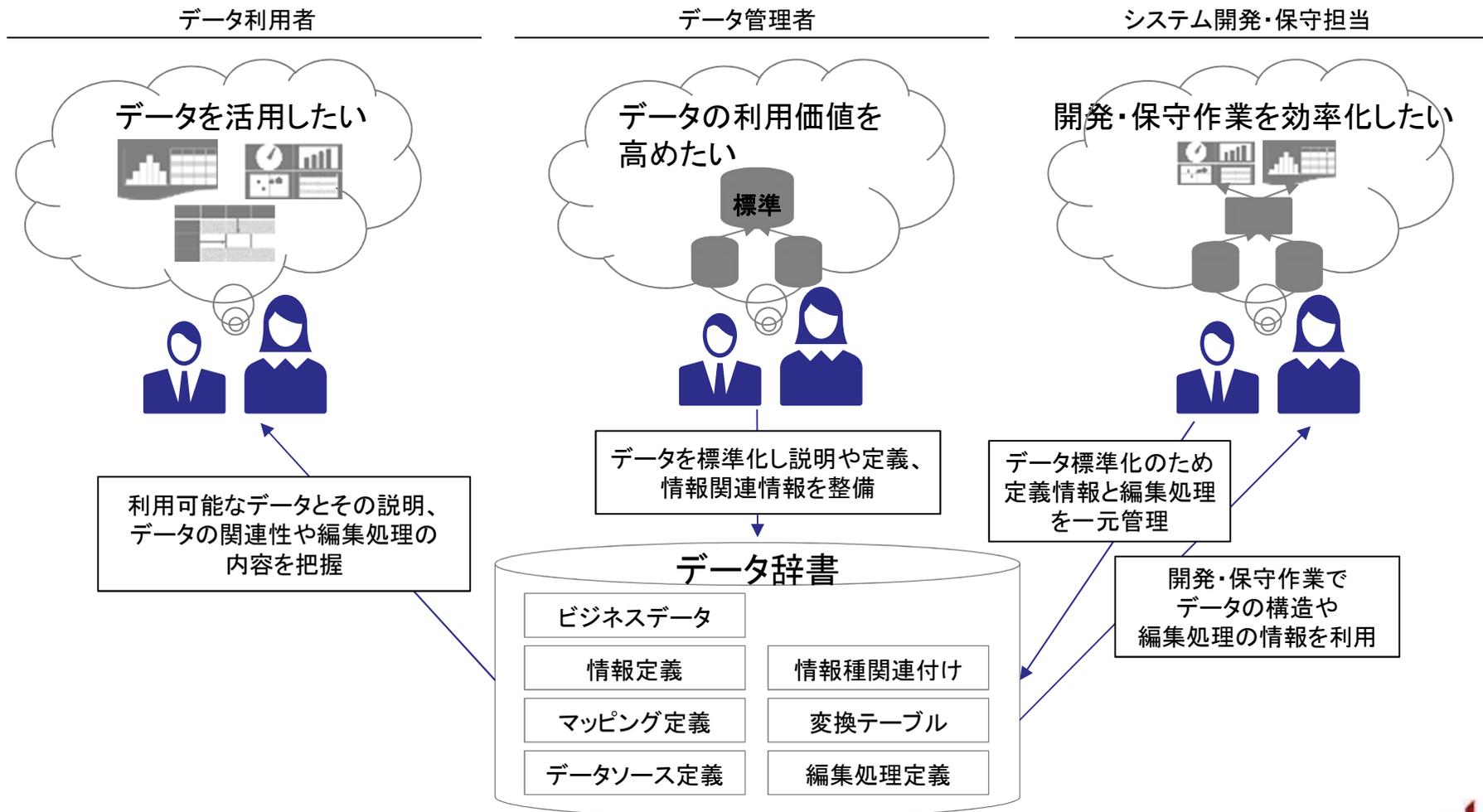
- データ・項目の説明情報とデータソースと情報(標準化データ)のデータ定義及び、テーブル及び項目の関連情報を保持し、情報の内容及び、情報間、情報とデータソース間の関連性を表す情報を提供する。

提供情報

- 情報(標準化データ)のデータ・項目の説明情報
- 情報のデータ定義情報
- データソースの定義情報
- 情報とデータソースのテーブル及び項目の関連情報
- 情報間の関連情報
- 情報とデータソース間の関連情報及びデータ変換情報

データ辞書の利用イメージ

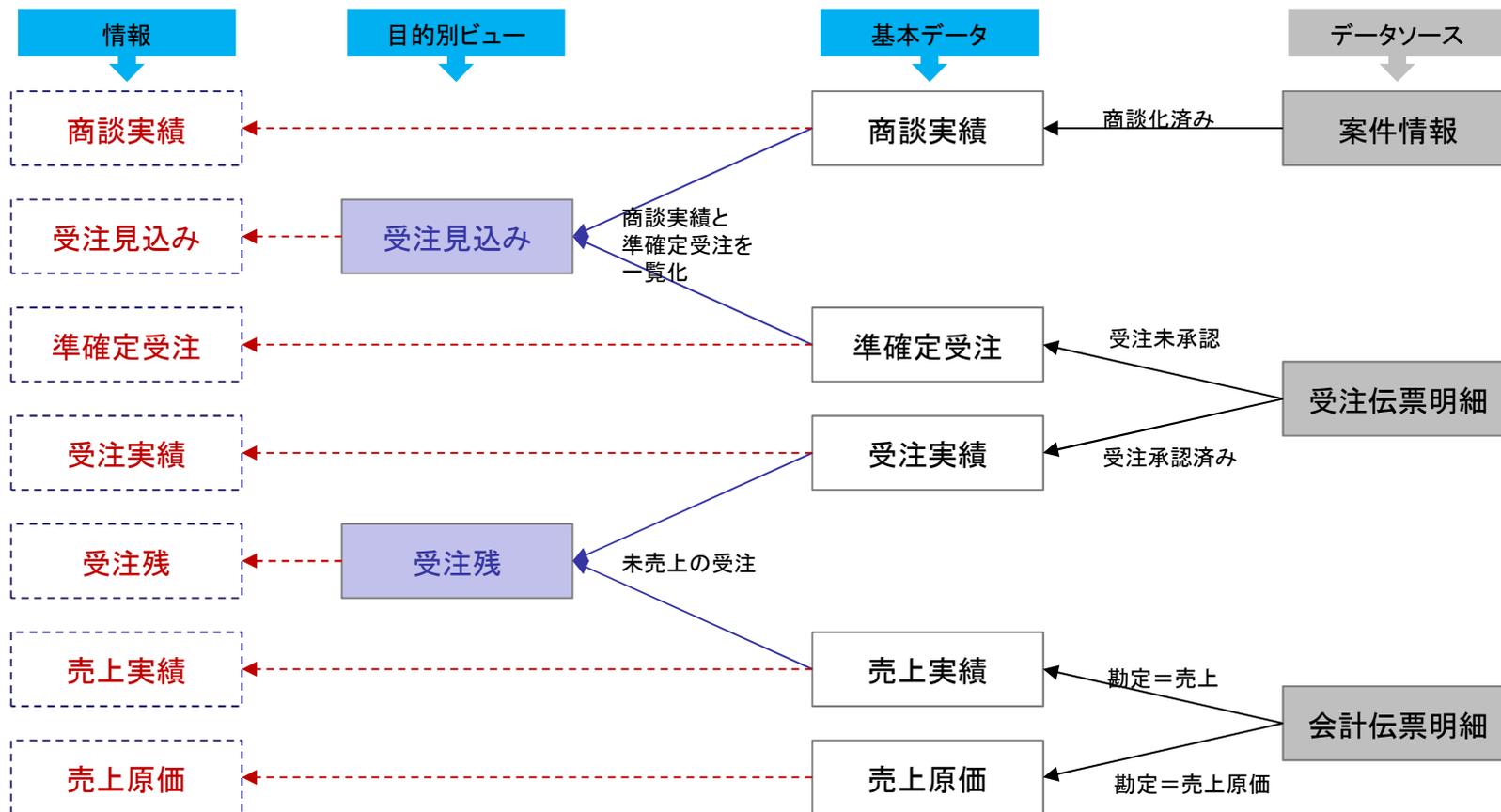
データ管理者、システム開発・保守担当者がデータの利用価値を高めるための情報を整備し、データ利用者へ提供する。



データ辞書の情報定義

【情報の定義】(案)

- 基本データはデータソースから、同じ内容の情報を利用者が再利用可能なものに、重複なく統合したもの。
- 目的別ビューは基本データやデータソースから、利用目的に応じて抽出、一覧化、集計などをし、再利用可能な形で統合したもの。
- 目的別ビューと基本データを標準データの情報として利用者に提供する。

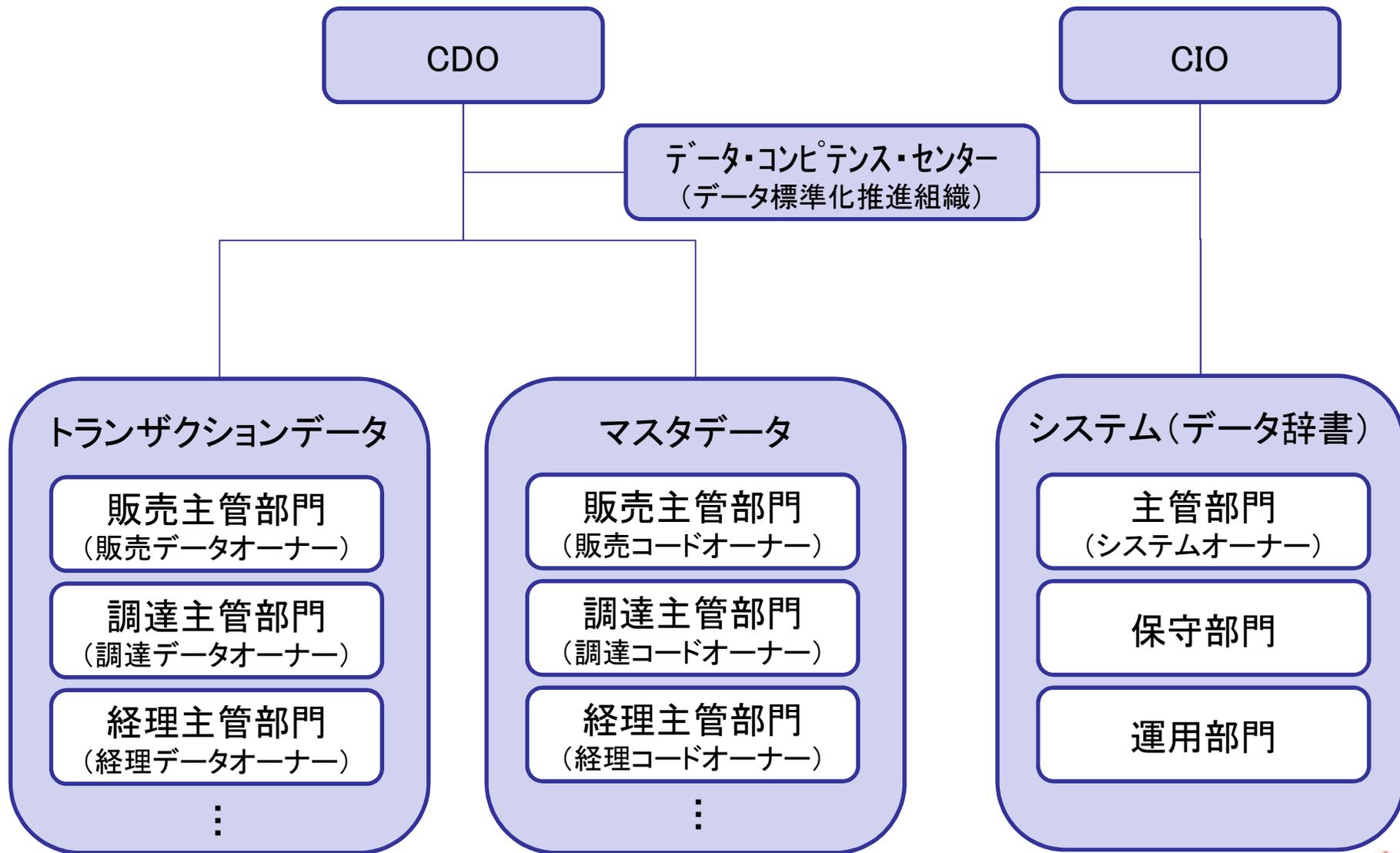


データ辞書で管理すべき項目

【データ辞書の項目】(案)

対象	項目	説明	例
情報	情報分類	情報が関連する業務や組織、品目等の共通する情報を分類するもの	受注
	名称(日本語)	情報の日本語名称	受注残
	名称(英語)	情報の英語名称	backorder
	テーブル分類	情報が属する情報のタイプ Fact=集計対象となる金額や数値をもつテーブル Master=品目、組織、取引先等の基礎情報をもつテーブル Mapping=情報の項目間の関連性を持つテーブル	Fact
	情報の説明	情報自体がどのようなものを記載	未売上の受注
	利用目的	情報が何を目的に作成されるか、 情報の利用される業務や情報の伝達方法など、どの様に利用するかを記載	営業部門が、当月の売上報告だけでなく、翌月や翌々月の売上見込や予測を報告する為に利用。
	利用対象者	情報を利用するユーザーについて記載(利用できる・できないユーザーやロール)	営業部門
	対象評価軸	情報の数値を分析、集計することができる評価軸	年月日、部門別、顧客別
	関連する情報	データ項目同士や変換テーブルで関連する情報を記載	受注実績、売上実績
	最小単位	情報が登録される最小単位	受注伝票明細単位
	更新サイクル	情報が更新されるサイクル	日次
	データ保管方法	データの保管方法	データ基盤、媒体
	オーナー組織	情報のオーナー組織	営業部門
情報区分	情報が”基本データ”か”目的別ビュー”かの区分	目的別ビュー	

データマネジメント体制案（データ辞書運用体制案）



研究成果②『「ヒコーゾー」』

研究テーマ：「非構造データに関する研究を実施する」

非構造データ利活用するためのアプローチ

- 非構造データを活用するポイントとしては、以下のような分野に研究が及んでいる

①データ基盤構築

- ✓ あらゆるデータを溜め込み、データ活用するための基盤が必要。
- ✓ データ種類（構造化、非構造化）、データ量（テラ、ペタ）、データ速度（データ鮮度）を問わずに蓄積可能。蓄積後はBIツールなどにて加工、分析を行う。
⇒スピードに直結するための技術が必要。たとえばHadoopなど

⇒今回はデータ基盤構築に着目して整理した

②データ統合

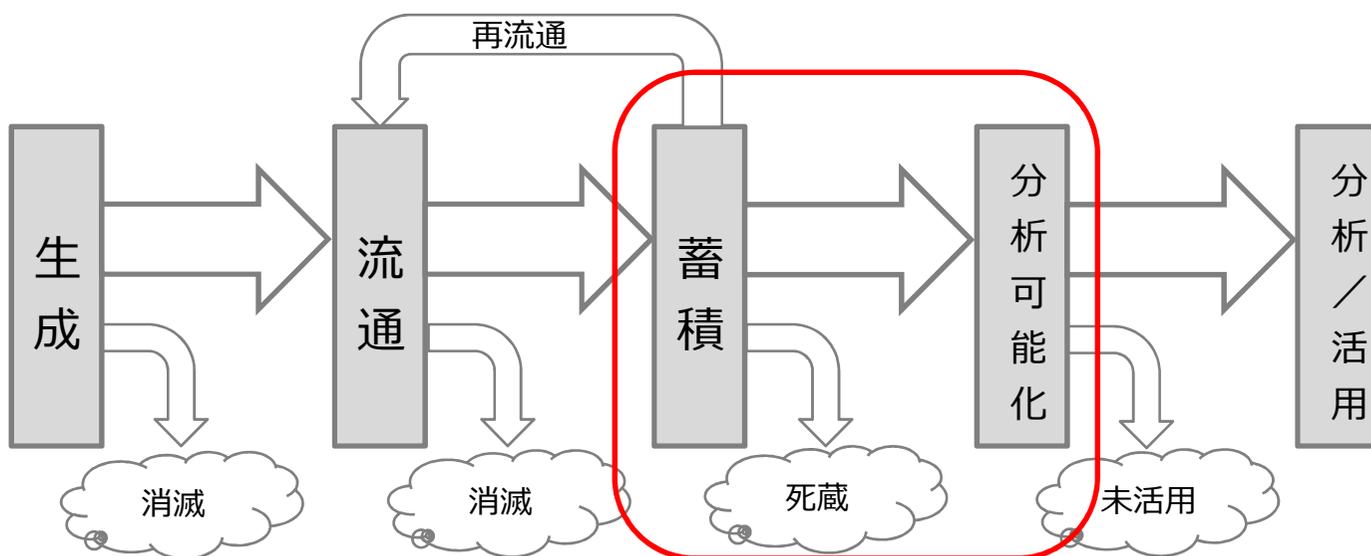
- ✓ データの地理的統合、データの構造的統合、データの意味的統合を推進する。
⇒分析効率を上げるために必要

③推進体制構築・人材育成

- ✓ 社内にシステム部門（企画部門、開発・保守）を保持。ITキャリアパス設定
- ✓ IT子会社では新領域への配置転換。

非構造化データ活用、利活用の流れ

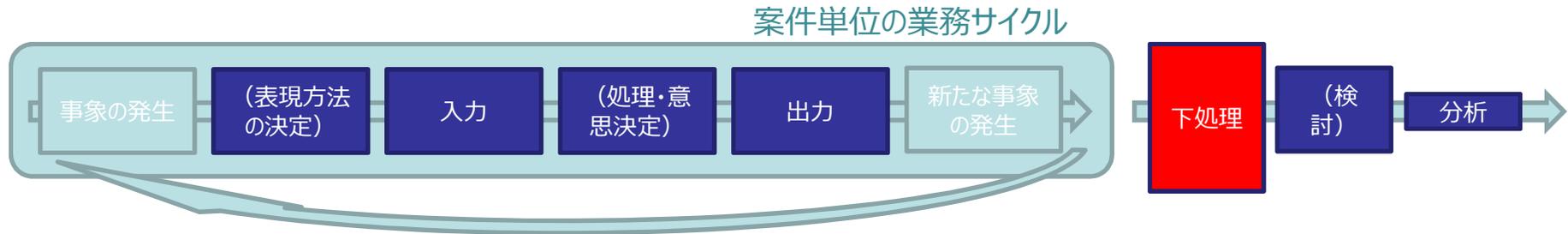
- ビッグデータは生成された後、流通、蓄積、分析可能化といった過程を経て、実際に分析・活用に供することとなるが、その過程で多くのデータが消滅、死蔵する。
- 蓄積状態から分析可能な状態に加工（データ統合、構造化）することで、データ活用できる



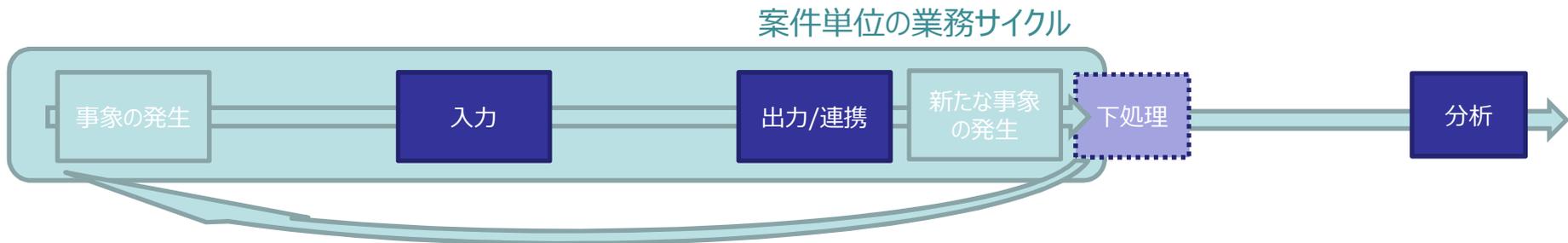
蓄積： 分析効率・活用性を上げるためには「データ統合」が必要
データの地理的統合、構造的統合、データの意味的統合

データ利活用のためには構造化が必要か？（分析=2次利用）

◆ Adhocな業務処理と構造化されていない記録の場合



◆ 定型的な業務処理と構造化された記録システムの場合



データの記録・発生の流れとは独立して分析・活用する際に、構造化されていないデータの取り扱いには困難が生じる。

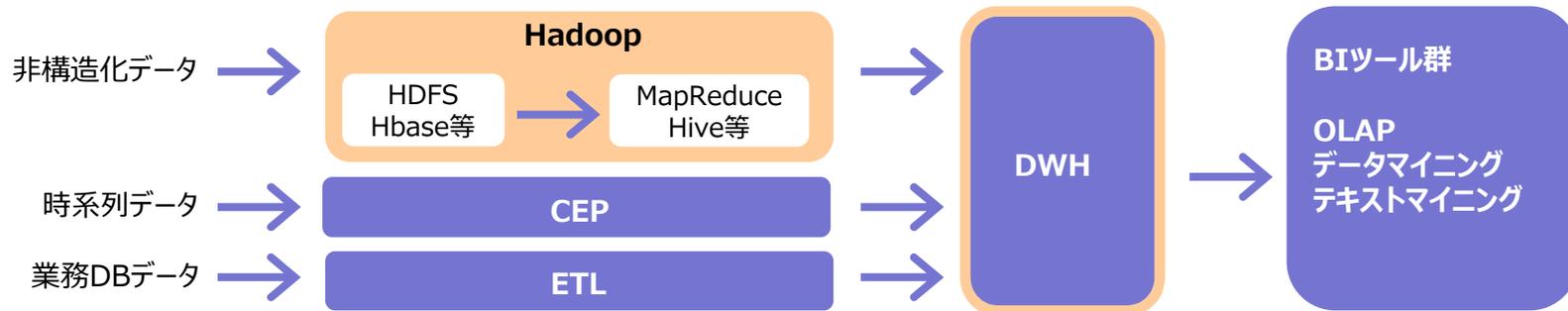
定量化のための下処理が技術的にも工数的にも難易度が高く、分析結果として得られる情報も部分的なものになってしまうことが多い。AI技術などにより下処理工程の自動化が実現することで非構造データの活用が推進される期待がある。

データ基盤のためのアーキテクチャ

■ Hadoopを利用した2つのタイプのデータ蓄積基盤

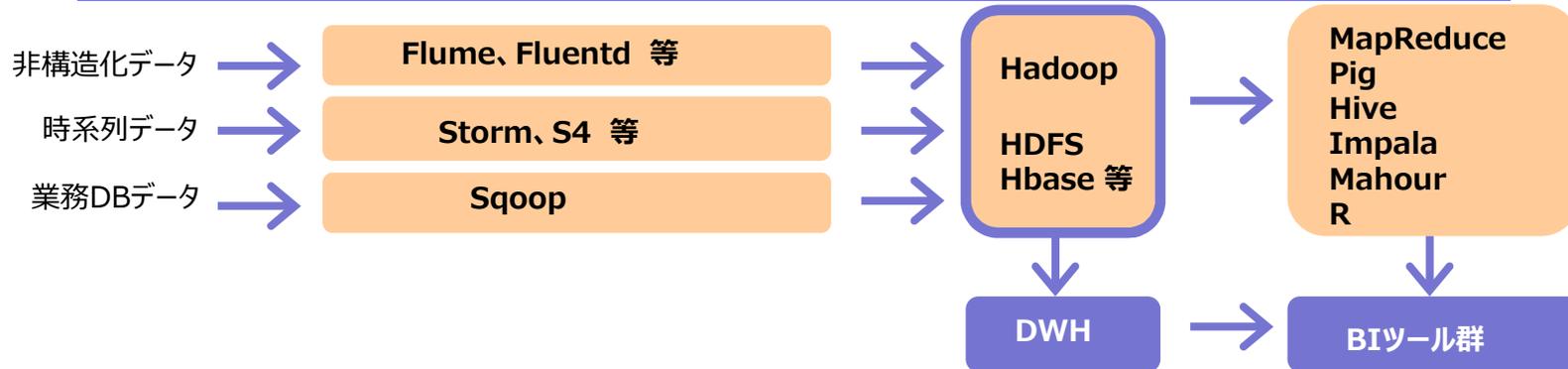
◆ データウェアハウス中心のアーキテクチャ

- ✓ 分析対象としてDWHに蓄積されるデータは、加工されて意味づけされたデータ
- ✓ 従来型のDWH/DIシステムに、非構造化データ対応のHadoopをETLとして追加した構成
- ✓ 従来型のBIツールを駆使する分析者に向いている。



◆ Hadoop中心のアーキテクチャ

- ✓ 分析対象としてHadoopに蓄積されるデータは、生データで、データ分析時に意味づけする
- ✓ 探索的データマイニングなど、意味づけの試行錯誤を繰り返すのに適している。
- ✓ 最新の統計理論をゼロからプログラミングしていけるデータサイエンティスト向けのアーキテクチャ



研究成果③『美・EYE(BI) ～使われるBI～』

研究テーマ：

「BIを導入したものの、当初企画とは異なり十分活用されていないケースが多々見受けられる。原因の追求とともに、より使われるBIのための方策を実際に手を動かして理解する」

- ① 海外・拠点データの統合・連携のコツ
- ② ETLツールの属人化の原因と対策
- ③ 使われるためのTableau（タブロー）とは

◆海外拠点データ統合のアンケート

№	項目	質問内容
1	データソース	データソースは？
2	データ統合業務	統合するデータの業務は？
3	データ利用目的	統合されたデータの活用目的は？
4	拠点数	海外拠点の数は？
5	拠点場所	海外拠点の場所は？
6	多言語	データソースは多言語でしょうか？
7	プラットフォーム	データソースのOSやDB、パッケージなどは統一は？
8	導入チーム	拠点へのシステム導入した際の主導チームは同じでしょうか？
9	データ定義	マスターのデータ定義は事前に定義されていますか？
10	運用ルール	各拠点でのマスターの運用ルールは事前に定義されていますか？
11	業務フロー	データ更新の業務フローは統一されていますか？
12	ツール	データ収集のツールは統一されていますか？
13	締め日	各拠点での会計締め日、入力締め期間などの違いはありますか？
14	IT/BIインフラ	IT/BIインフラが要因でデータ統合に問題ありませんか？

アンケートは5/12社が海外データ統合対象

◆ETLツール属人化についてのアンケート

№	項目	質問内容
1	ETLツール	使用しているETLツールを教えてください
2	ETLツール担当	ETLツールのメンテナンスは実施していますか？
3		メンテナンスのヘルプデスクは実施していますか？
4		開発した内容は仕様書、マニュアルなどで共有されていますか？
5		開発経験、設計思想など記載はありますか？
6	組織体制	ETLの仕様を把握し、対応出来るメンバーを常駐のローテーション、勤務先などの情報をお知らせいただけますか？
7		ETLの運用、改善の担当者も教えてください
8		改善の発注はどの程度でしょうか？
9		
10	属人化	ETLツールの運用について属人化していますか？
11		属人化のあなたのお考えをコメントください
12		属人化解消のため、どのようにすればいいと思いますか？

◆Tableau分析のお題と方向性

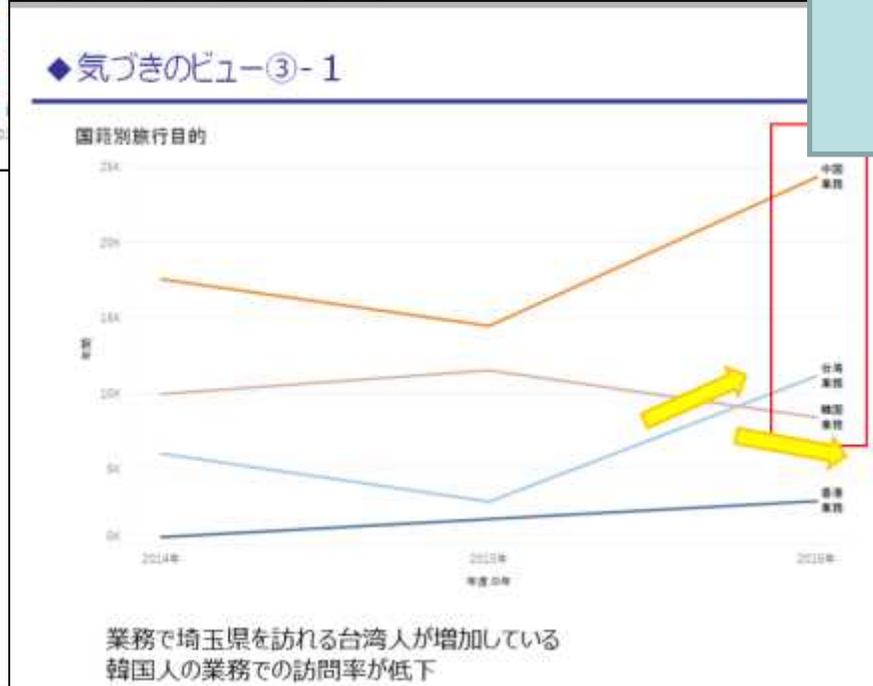
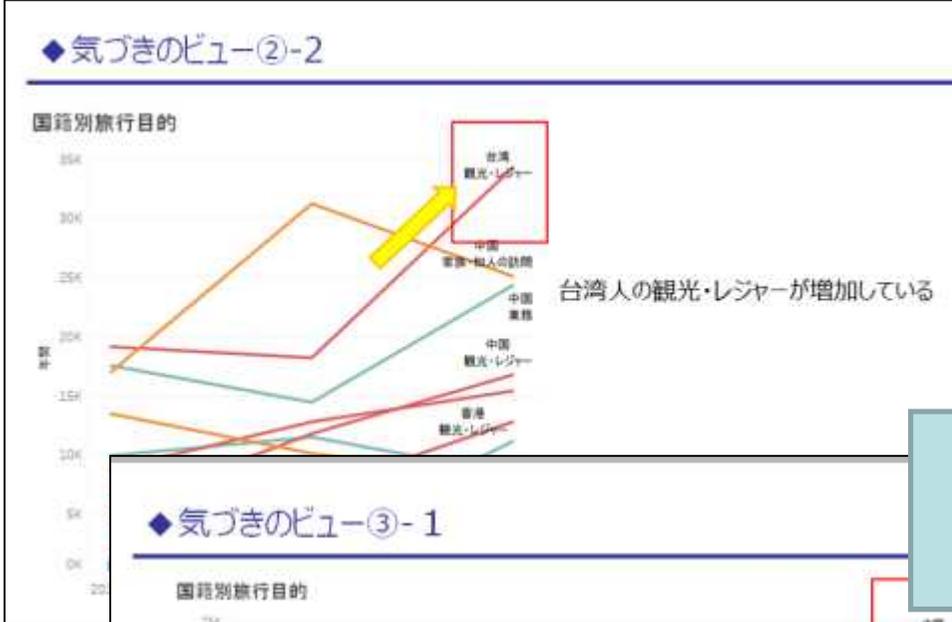
【お題】 埼玉に海外からの観光客を呼ぶ

- ・分析テーマは「埼玉に海外からの観光客を呼ぶ」をテーマとし、観光客の動向を可視化・分析を行い、訪問者増加の要因を分析する
- ・最終的には観光客増加の施策の立案を目指す

【使用データ】

- <観光庁>
 - ・訪日外国人客数ベース（2011年～2017年）
 - ・訪日外国人消費動向調査（2016年）
 - ・都道府県間流動表（2014年～2017年）
- <聖地巡礼サイト>
 - ・アニメ聖地スポット数
- <埼玉県HP>
 - ・観光課 統計情報

Tableauを用いた分析・原因調査の実践



2016年に埼玉国際観光コンシェルジュが台湾 (JTB台湾内) に設置

埼玉国際観光コンシェルジュを台湾に設置一県が初めて海外に観光プロモート拠点を設置

埼玉県は、近頃から本県への観光旅行や、観光旅行会社のツアーを拡大するとともに、本県観光の魅力を積極的に広げて観光客を増やすために、「埼玉国際観光コンシェルジュ」を平成28年8月1日（月曜日）に設置いたします。

台湾の技術者交流会をさかんに実施

台湾技術交流会、経済事情視察一参加者募集について

韓国企業の日本撤退

韓国企業は、日本市場からの撤退を加速している。日本市場からの撤退を加速している。日本市場からの撤退を加速している。

◆まとめ(Tableauを使ってみて)

- Tableau Japan様より使用ライセンスを借用してメンバーで実際に各自で検討してみた
- 各自でTableauのViewを作成してみたが、あまり進まなかった

なぜか？

- 日々の業務が多忙なため (Tableauだけが仕事ではない)
- やはり使い方が難しい (ハンズオンを受けてみたい)
- 既存のツール類 (Excelなど) である程度のことが出ている
- Tableauで見せる目的を探しながらだったので、試行錯誤だった。

しかし！

- チャンピオン(マスターしている人)を中心にViewを作ると新たな気づきをその場で得られ、モチベーションが上がる
- チームで「ワイワイ」Viewを作り上げると良い！
- D社では**Tableauを使うのが「楽しい」**という意見が出ている

「R」を使ったテキストマイニングの実証実験

「Rを使ったテキストマイニング勉強会」を開催@IMJ



お題「すももももももものうち」

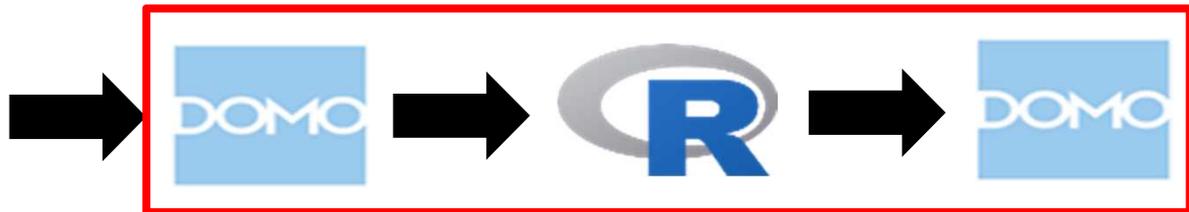
MeCabによる
形態素解析

```
> library(RMeCab)
> unlist(RMeCabC("すももももももものうち"))
  名詞      助詞      名詞      助詞      名詞      助詞      名詞
"すもも"  "も"    "もも"  "も"    "もも"  "の"    "うち"
> |
```

なるほど、このくらいは瞬殺で出てくるのですね・・・

実際のビジネスデータを使ったテキストマイニング実験

顧客訪問履歴データ



トレーニングで得た知識と
Google先生をフル活用し
実装

```
#ライブラリーの読み込み
library(RMeCab) library(sqldf)

#Domoへの接続認証
DomoR::init('xxx','xxx')

#特定データの読み込み
df <- DomoR::fetch('xxx')

#項目・件数の絞り込み
df2 <- sqldf("SELECT ID,CONTENTS__C from df
             where
             CONTENTS__C not like '%<NA>' and
             CONTENTS__C not like '<br>' and
             CONTENTS__C is not null")

#タグの除去関数
replaceTag <- function(x){
  r<- gsub("<br>","") r<- gsub("<u>","")
  r<- gsub("</u>","") r<- gsub("<p>","")
  r<- gsub("</p>","") r<- gsub("<U+00A0>","")
  r<- gsub("&#n","") return(r)}

#変換関数
#dfx : RMeCabDFで形態素解析したリストの1行分データ
#id : 元データのID
exData <- function(listdata,id){
  #データフレーム化
  type <- matrix(unlist(attributes(listdata[[1]])))
  morpheme <- matrix(unlist(listdata))
  dtf <- data.frame(type,morpheme,
                   stringsAsFactors=FALSE)

  #出現数カウント
  dtf_noun <- dtf[type=="名詞",]
  dtf_nounList <- table(dtf_noun[2])
  dtf_summary <- data.frame(dtf_nounList)

  #IDとマージ
  idCol <- c(rep(id,nrow(dtf_summary)))
  r <- cbind(idCol,dtf_summary)
  return (r)}
```



分析結果と考察



というわけで実践してみました編

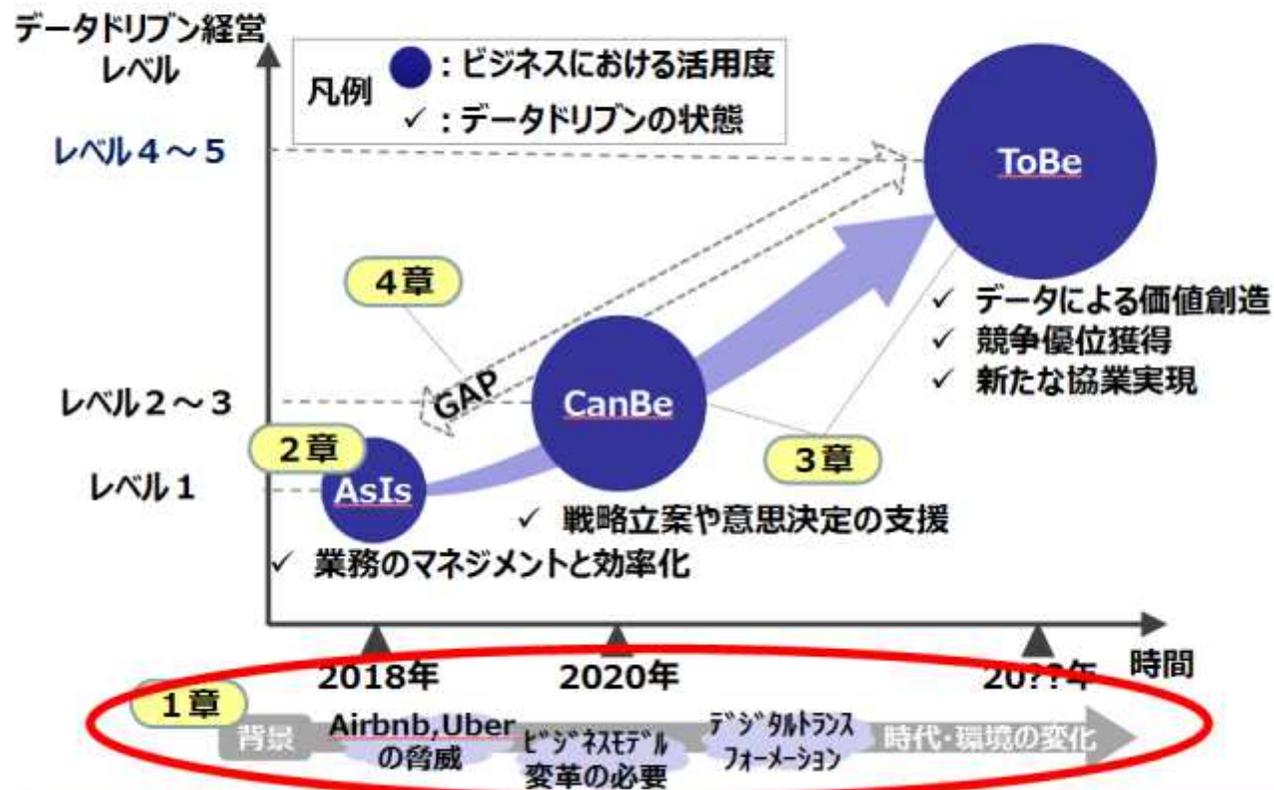
という問題はありつつも、本取組をまとめます。

- ✓ 簡単なトレーニングを受けて、あとはGoogleの検索さえできれば、オープンソースのRでも最低限の形態素解析は出来る。
- ✓ 既存のツールと連携して新たな可視化の方向性が図れそう。
- ✓ 実践してみたことで、Rのこと、MeCabのこと、形態素解析のことが理解できるようになった。

研究成果⑤『データドリブン勉強会』

研究テーマ：「データドリブン経営のあり方と実現方法」

研究成果エグゼクティブサマリ



総括 データドリブン経営を進めるうえでの具体的なアクションプランのヒントを得た！

データドリブン経営の整理

データドリブンができている企業は大きく以下の整理ができる

事業戦略やビジネスモデル論
になってしまう

視点	過去の経営 (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
経営	<p><期待> 余りデータは重視しない</p> <ul style="list-style-type: none"> ざっくりとした結果把握、部分的・業務単位の業務効率化・コスト削減 	<p><期待> 結果&原因把握重視</p> <ul style="list-style-type: none"> 詳細な結果把握、定型的な原因分析、会社トータルの業務効率化・コスト削減 	<p><期待> データは経営資源</p> <ul style="list-style-type: none"> ビジネスモデル変革/商品・サービス強化。 大幅かつ革命的な生産性改革
業務	<ul style="list-style-type: none"> 現場の業務(プロセス)を変えない 業務に対してデータの活用度が低い。 	<ul style="list-style-type: none"> 現場の業務(プロセス)を変えない 業務の部分最適・予実管理でのデータ活用 (戦略レベル) 	<ul style="list-style-type: none"> 個別業務(プロセス)が有機的に結合 戦術レベルの情報を享受できるデータ活用
IT・情報システム	<ul style="list-style-type: none"> 業務効率化・生産性向上が目的 →レガシー →低柔軟性 	<ul style="list-style-type: none"> PDCA運用が目的 →データ量に応じた適正なシステム →個別部分最適 (個別乱立) 	<ul style="list-style-type: none"> 業務変革が目的 →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し、データ質・量・鮮度に最適化されたシステム
組織・人材	<ul style="list-style-type: none"> IT部門：分析・集計(Excelマクロ) ユーザー：IT部門へ依頼 	<ul style="list-style-type: none"> IT部門：DB/DHW ユーザー：セルフ&ITへ依頼 	<ul style="list-style-type: none"> IT部門：データプロバイダー ユーザー：分析・集計(セルフ) 分析専門：予兆分析 経営層としてのCIO/CDO

データ利活用の発展段階とデータ分析レベル

経営ニーズとデータ利活用の発展レベル

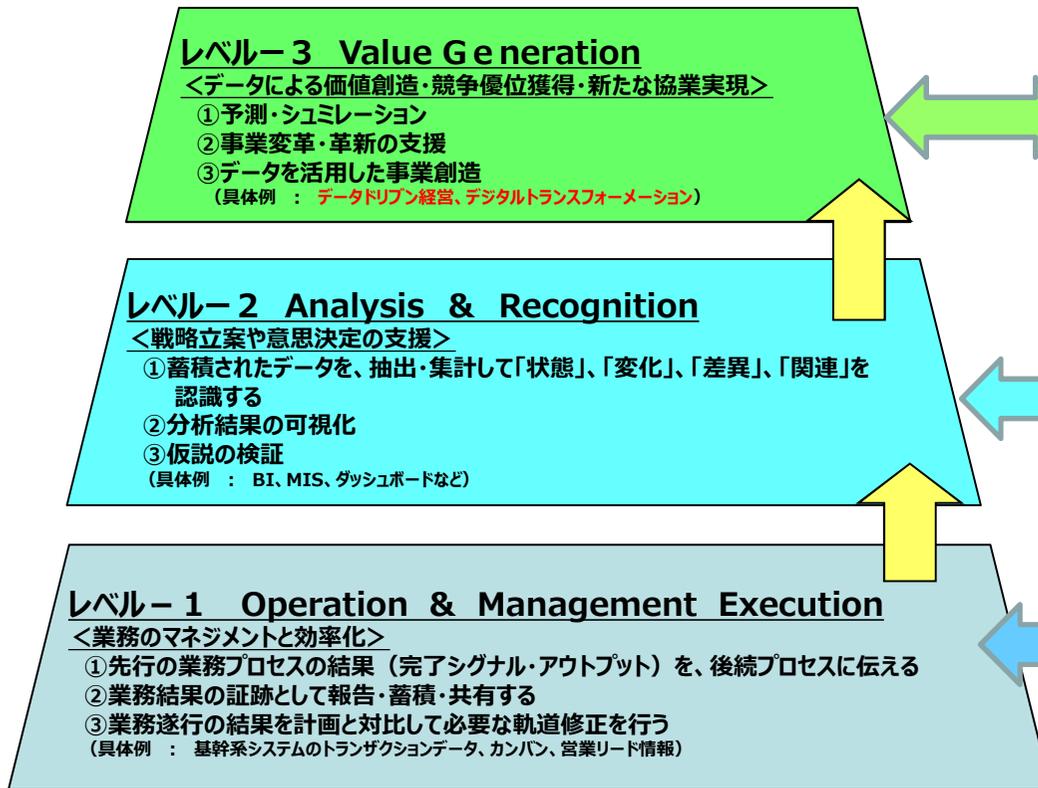


図 : 経営ニーズとデータ利活用の発展レベル

データ分析レベル

出典 : インテル分析プランニングガイドを参考に作成

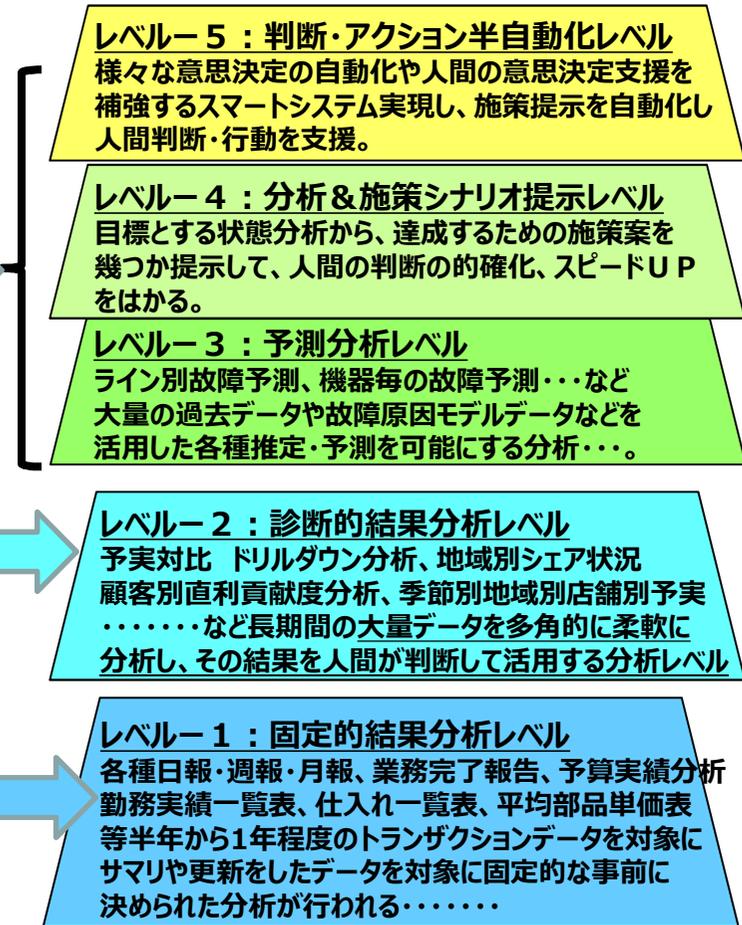
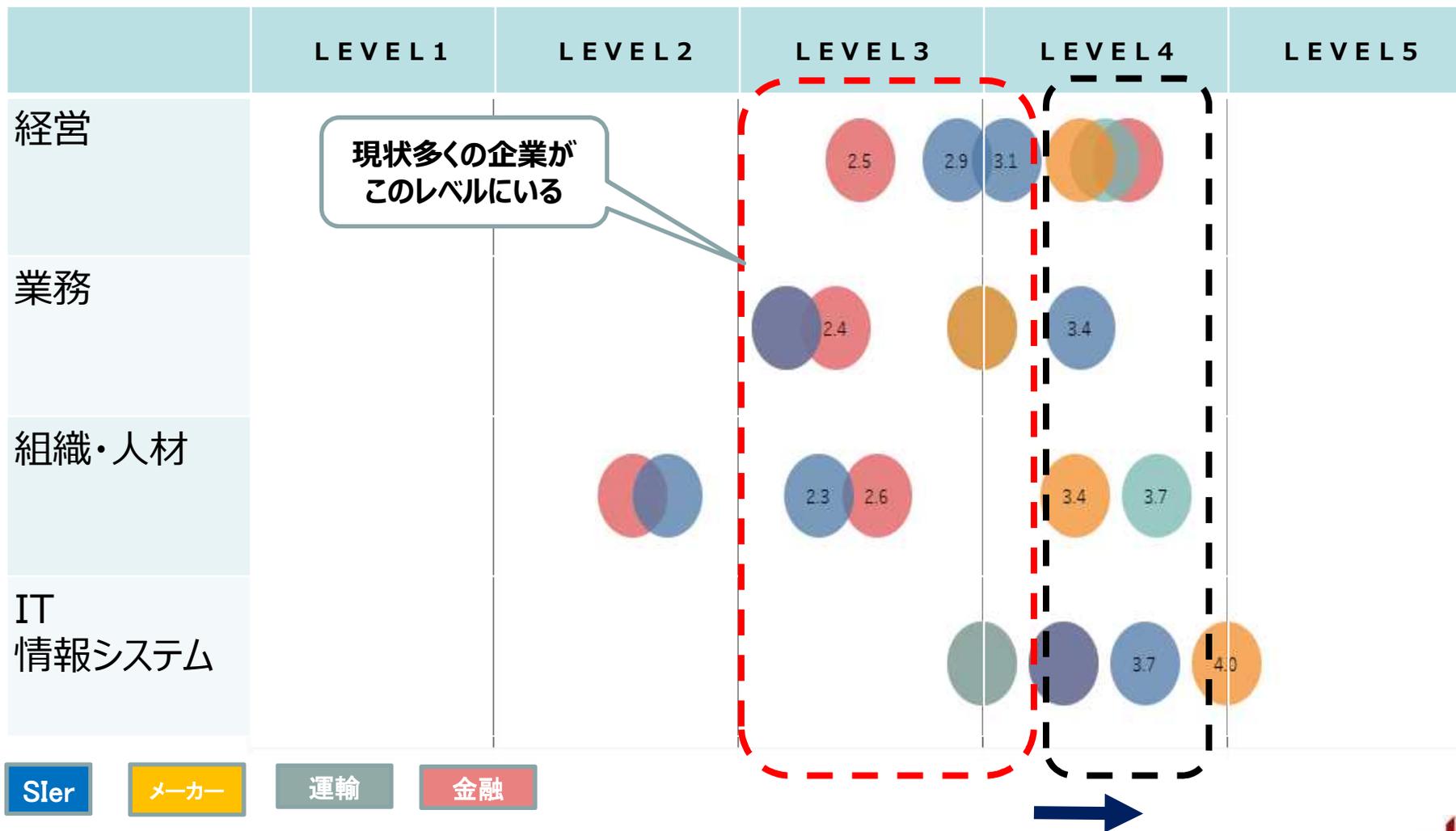


図 : データ分析の発展レベル

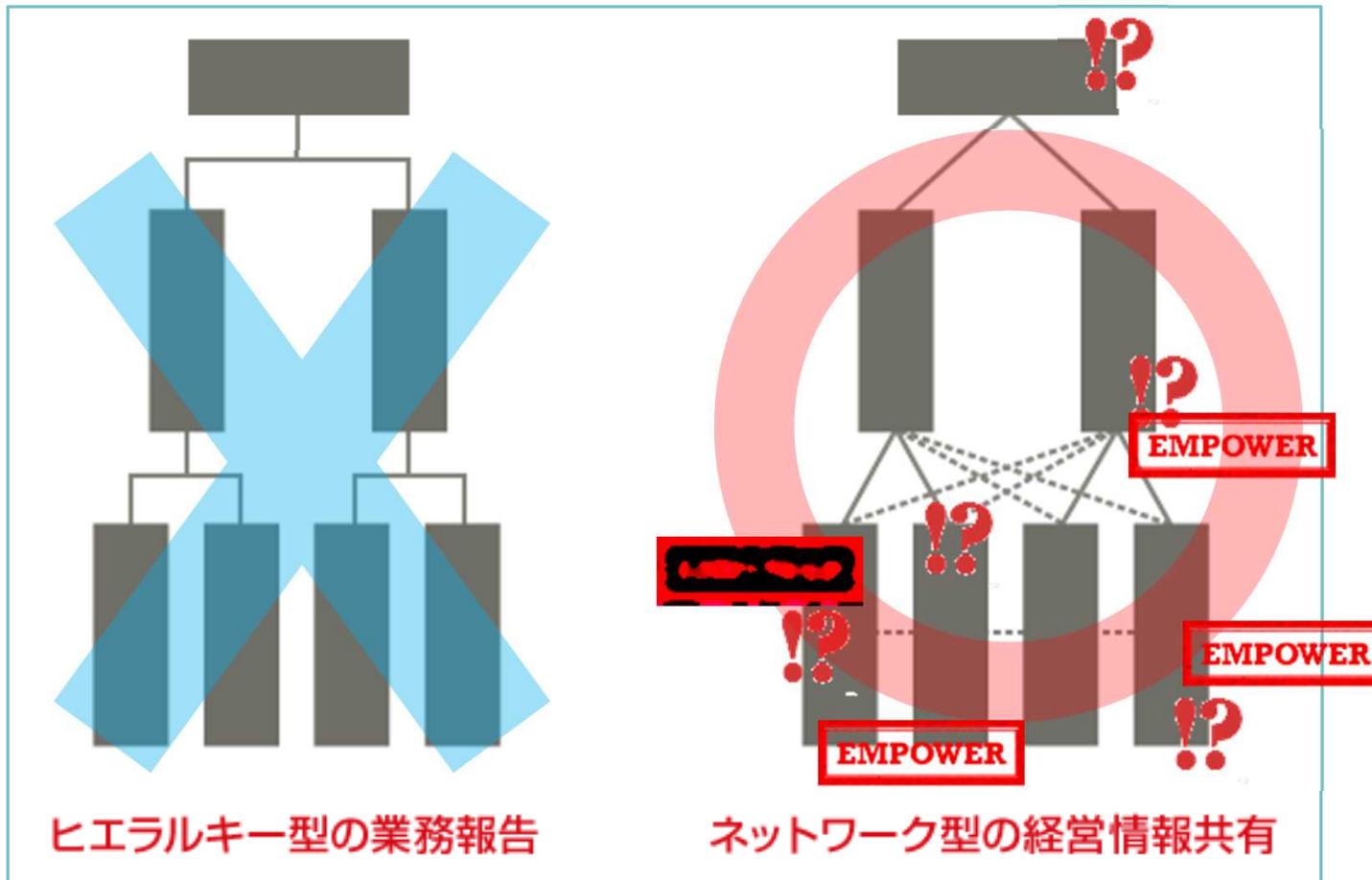
現状(As-Is)アンケート調査の結果

当初想定よりもレベルが高めにプロットされているが想定外であった。質問と回答内容について精査が必要で現在ブラッシュアップを行っております。



データドリブンな組織運営：権限や判断はエッジに任せる

BIの民主化によるOODA的エッジ型業務判断の推進（P2E:PowerToTheEdge）

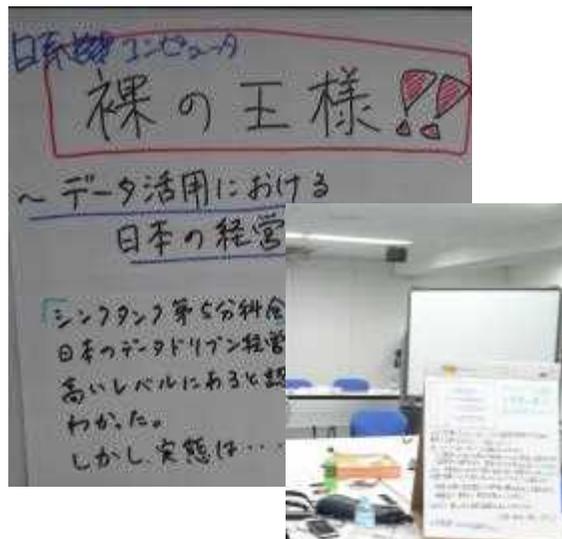


権限を与えて、判断とアクションをスピードアップ！！

第2合宿で研究会しめくり

3月24日(土)、JUAS会議室にて日帰り合宿を行い、分科会での研究成果に基づいた『体験型イベント』を行い、日常業務では味わえないエクスペリエンスを共有しました。

【日本のデータドリブン
経営の実態を報道しよう！】



【バイモーダルITを自分の
言葉で語れるようになるろう！】



【MODE-2開発を
体験しよう！】

