

データ利活用のための データマネジメント技法の研究 (クラス1)

－最初に取り組むべき5つの知識領域－

**この発表は
データ利活用が進まない
何から始めていいかわからない
そんな方たちが**



**データ利活用に向けて
取り組みの第一歩を踏み出せる
という話です！**



本日のメニュー

1 利活用の現状

なぜデータ利活用が進まないのか

2 仮説

データ利活用を進めるための取り組み

3 シナリオ

課題からシナリオを作成する

4 評価

検証の評価はどうだったか

5 まとめ

データ利活用へのスタートに向けて



本日のメニュー

1 利活用の現状

なぜデータ利活用が進まないのか

2 仮説

データ利活用を進めるための取り組み

3 シナリオ

課題からシナリオを作成する

4 評価

検証の評価はどうだったか

5 まとめ

データ利活用へのスタートに向けて



4つの躓くポイント

そもそも、**どんなデータがあるのか**分からない

自社でデータを活用するための**方針・ルール**がない

データ利活用に**適した環境**がない
(プラットフォーム・インフラ)

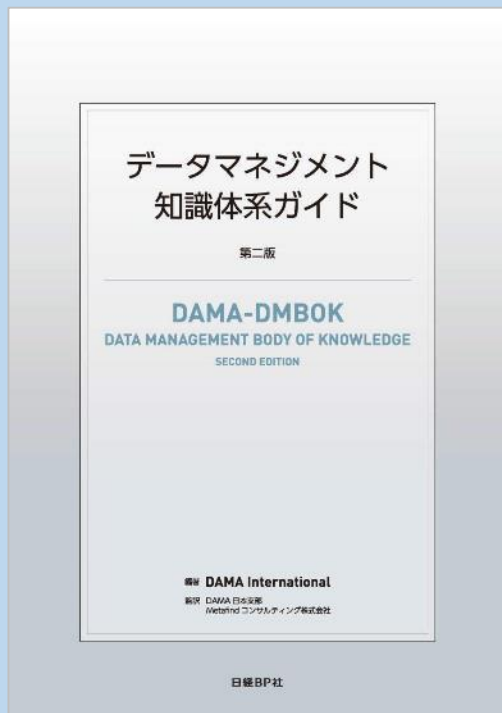
推進するための**組織**がない

これらを
総合的に解決するため
データマネジメントが重要
とされています



必要な情報は提供されている

DMBOK



17章 671ページ



11の知識領域

取り組み順序
記載が少ない・・・

やるべき事が
多い・・・

何からやれば
いいかわからない



本日のメニュー

1 利活用の現状

なぜデータ利活用が進まないのか

2 仮説

データ利活用を進めるための取り組み

3 シナリオ

課題からシナリオを作成する

4 評価

検証の評価はどうだったか

5 まとめ

データ利活用へのスタートに向けて



**自らの課題に対して
関連する知識領域に絞りこんで
スモールスタートすることができる**



本研究分科会の参加企業から抽出した課題

データスワンプ

データ取得していても
活用されていない

データ品質

各システムのデータを
紐付けできない

人材

BIツールがあっても
活用できる人がいない

データのサイロ化

複数システムのデータ分析
データ集約工数が多い

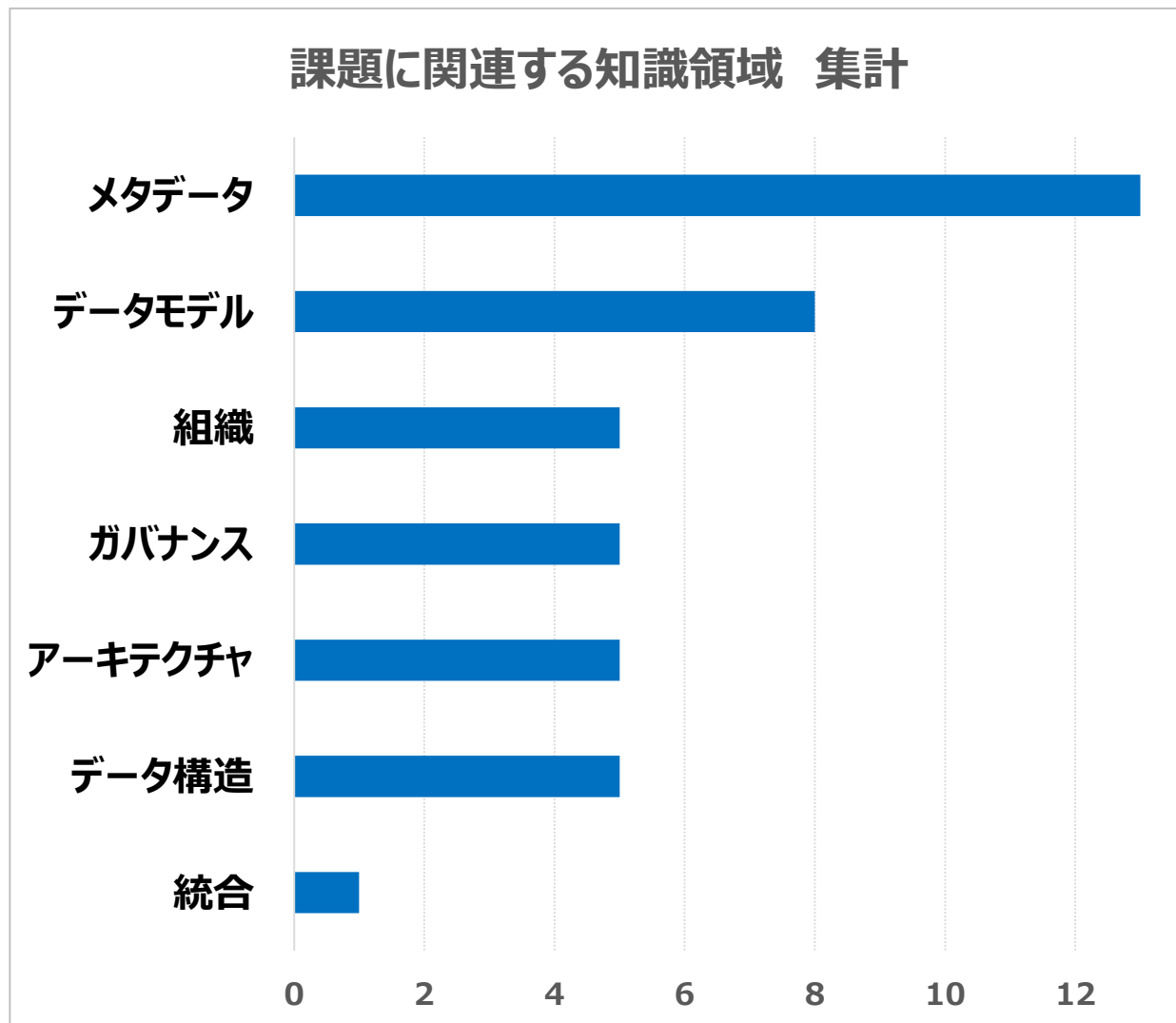
組織体制・ガバナンス

データの主管部署が不明瞭


セキュリティ


どのデータに顧客データが
存在するか分からない

着目すべき知識領域




 データモデリング

 メタデータ管理

 データ品質

 組織体制

 データ統合

5つの知識領域を選定

5つの知識領域の説明

データモデリング

データ要件を洗い出して分析し、取り扱いスコープを定めるプロセスである。そのデータ要件を伝えるため、明確に定義されたデータモデルと呼ばれる様式を用い、データの構造と関係を明らかにする。

メタデータ管理

データを定義し、説明するデータであり、主には定義として知られる。品質を測る基準にもなり、利用時に意味を知る拠り所となる。ビジネス上の意味から物理設計時の定義ともなり、関係を示すものまで含む。

データ品質

高い品質を維持するため、定期的なモニタリングを実施し、必要に応じて是正する。上記メタデータでの定義に沿った値なのか、必要とされる十分な量が確保されているかまで含める。

組織体制

データマネジメントは、IT部門と業務部門が共同で実施するものであり、そのための役割を定義し、アーキテクチャを定義するところから、データ品質を守るガバナンスの実施まで、それらを担う組織を要する。

データ統合

点在する社内のデータを一元的に扱える環境として、ETLのツールなどの支援を受け、共通にアクセスできる環境として用意する。統合するためには、上記の何れの要素が必要である。

仮説について次の**2つの検証方法**で証明する

内部検証

本研究分科会内の課題からシナリオを作成し
実行内容が5つの知識領域でカバーできるか

外部検証

内部検証結果を本研究分科会参加メンバー
以外第三者にて評価する

本日のメニュー

1 利活用の現状

なぜデータ利活用が進まないのか

2 仮説

データ利活用を進めるための取り組み

3 シナリオ

課題からシナリオを作成する

4 評価

検証の評価はどうだったか

5 まとめ

データ利活用へのスタートに向けて



抽出した課題に対してシナリオを作成



課題の特定

課題に対する事象の設定

(1)

事象から原因を特定

(2)

あるべき姿の設定

(3)

あるべき姿実現のためのアクティビティを選定

(4)

実行ステップの設定

参加企業から抽出した課題（抜粋）

データの保管はできていても
分析活用できない

分析データ用データ抽出作業
を都度頼まれる

人事異動で保守運用非効率化

今あるデータのままシステム
で加工して分析する仕組みを
作ろうとする

どの部門に、どのようなデー
タが存在しているのかがわか
らない

データを一つに集約すること
に開発工数がかかる

情報システム部門の役割と
ユーザー（利用部門）の役割
分担ができていない

メンバーの高齢化、人員不足

シンプル化したシステムを再
度複雑化しないための維持に
問題あり

データマネジメントできる社
員がいない

データマネジメントを知る人
が少ない

全社横断的なルール整備とそ
れをするための体制が弱い

各システムのデータを紐付け
できない

連携元システムの改修・デー
タ利用方法変更に伴う不具合
の発生

とりあえず収集したデータが
実際は使用されない

過去に要件定義をしたうえで
収集したデータの定義が、担
当部門でも共有されず今では
誰も意味がわからない



課題の選定

本研究分科会の参加企業から抽出した課題
その中から重複度が高い**4つの課題**を選定

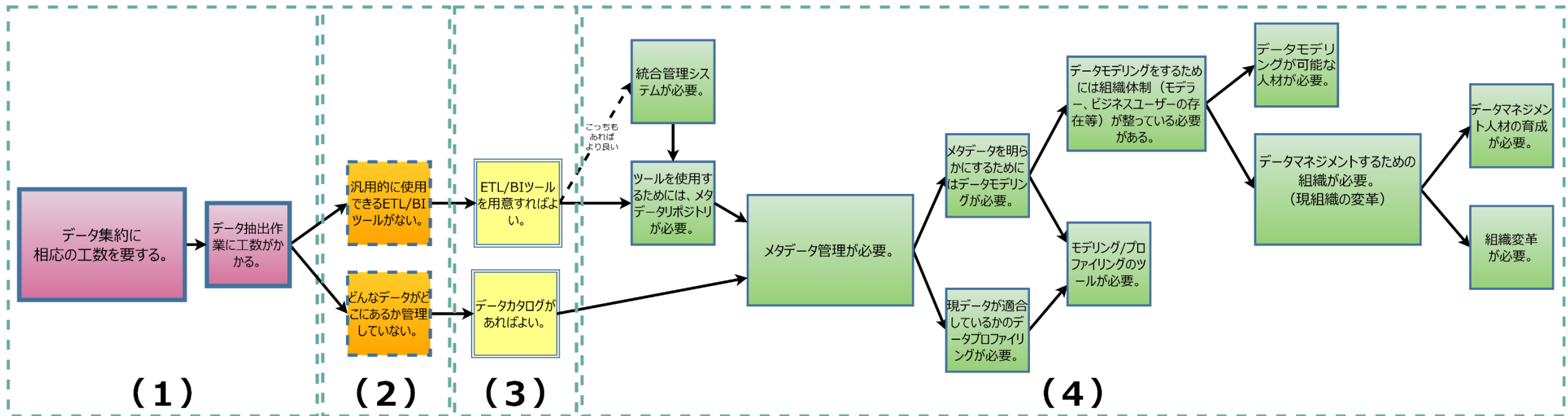
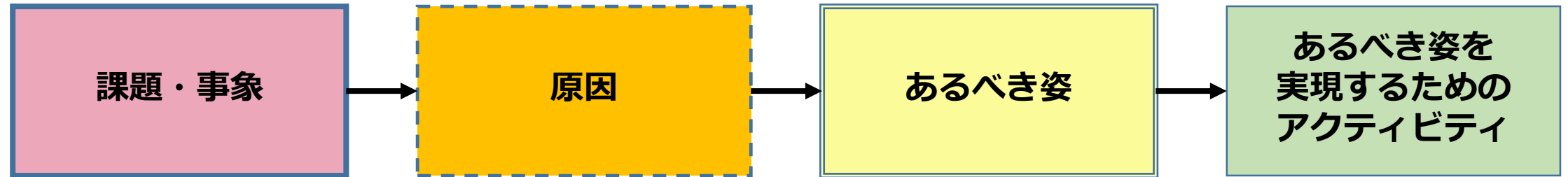
データ集約に相応の工数を要する

システム間でデータの紐付けができない

データの定義が共有されず、意味がわからない

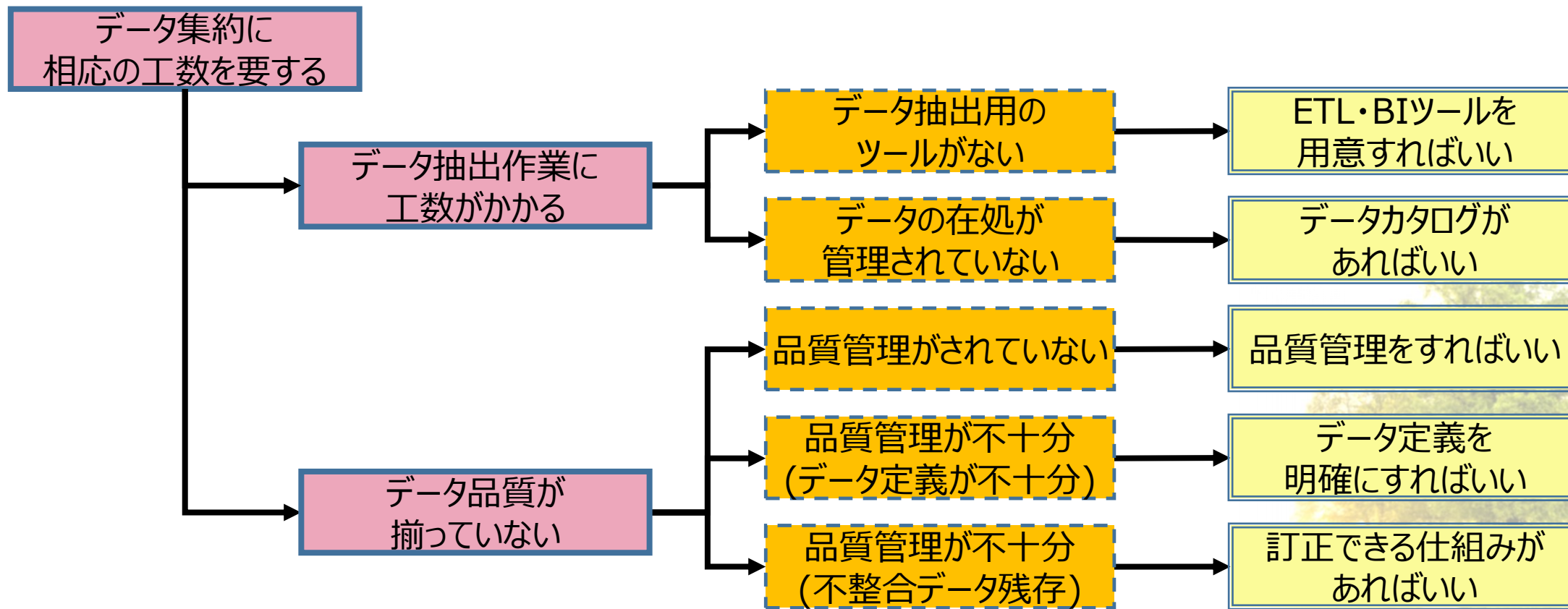
同一名称のデータで意味が異なるものがありユーザーが混乱する

シナリオの流れと表現

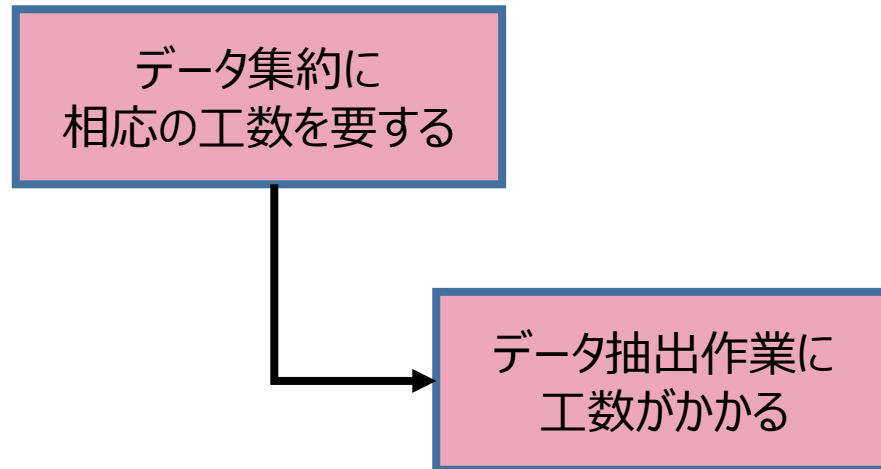


事象・原因・あるべき姿の考察

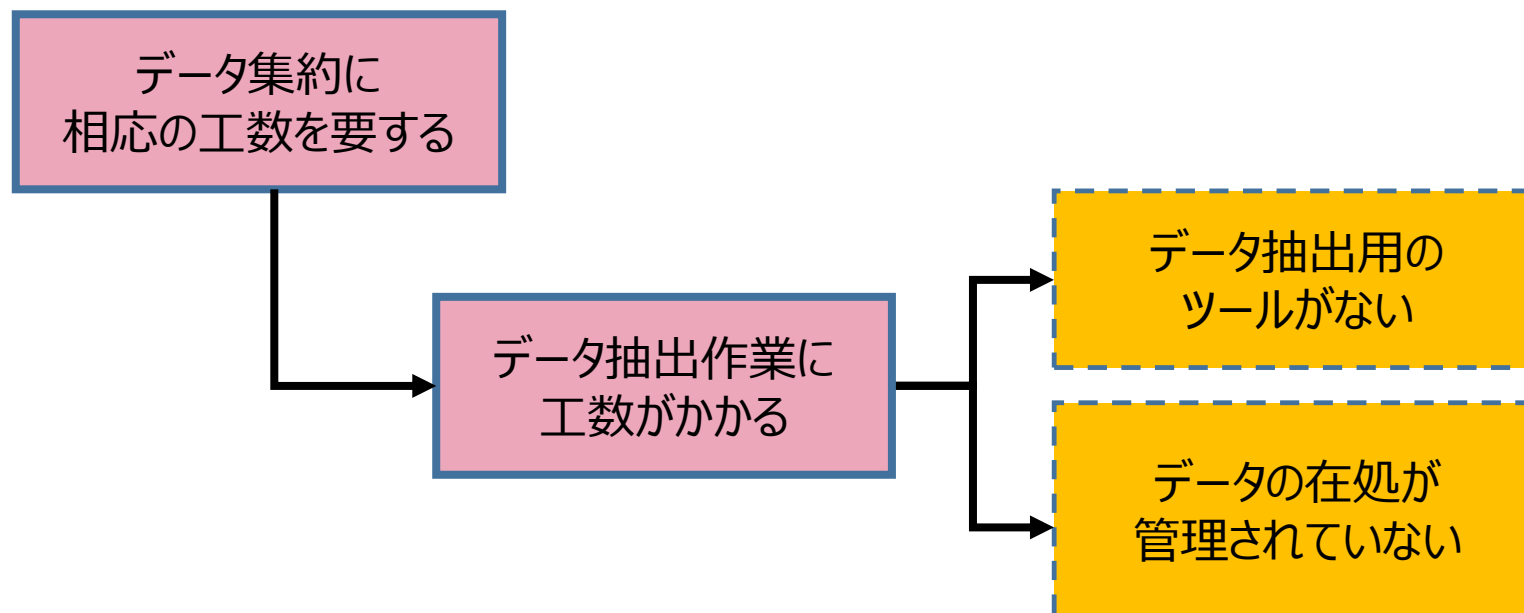
課題に対してまず**どのような事象**があるか、
そして事象に対する**原因**、**あるべき姿**を考える。



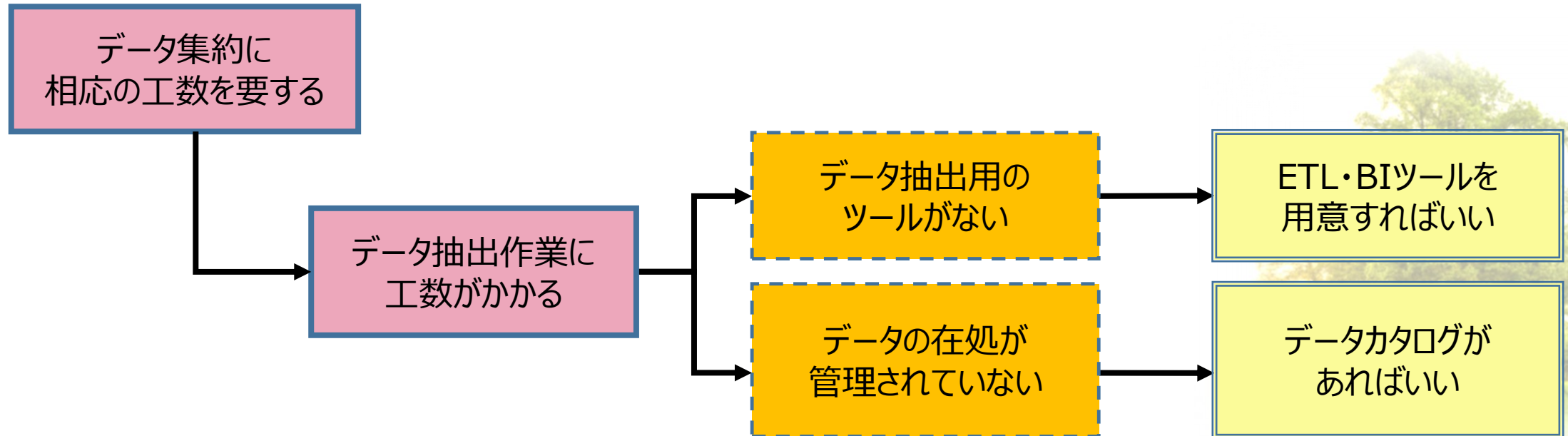
課題から対応策を検討できる粒度の事象に設定する。



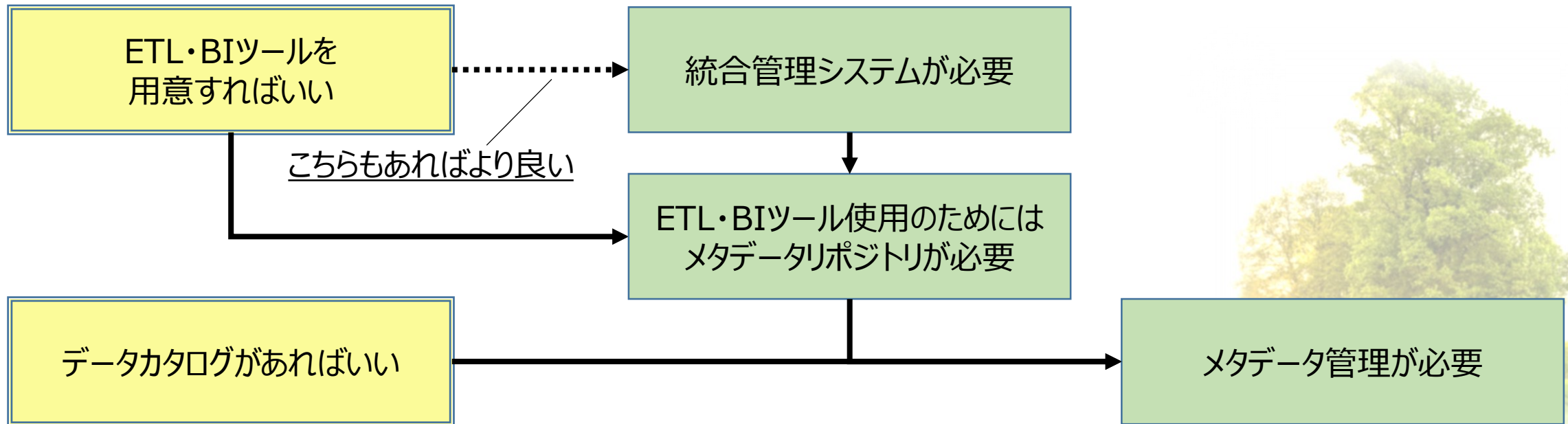
事象に対する原因を特定する。



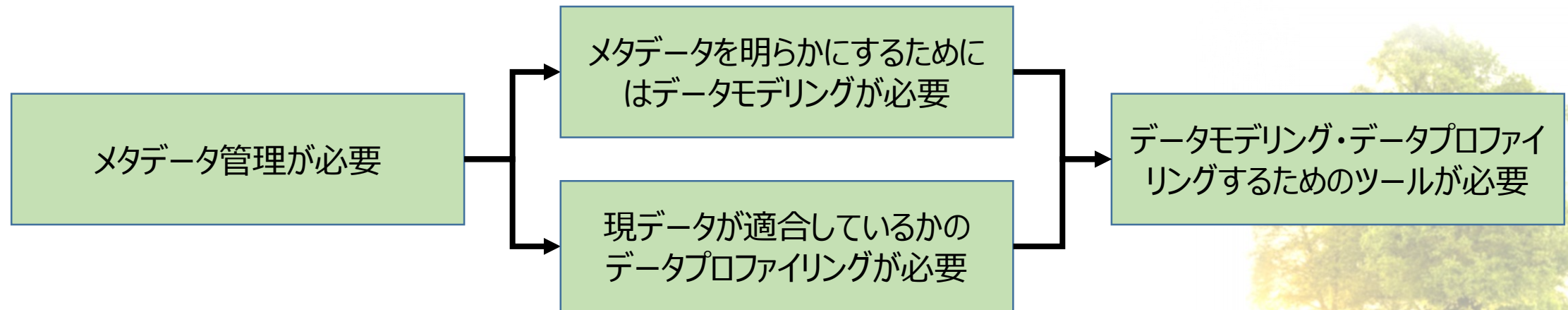
原因に対してあるべき姿を設定する。



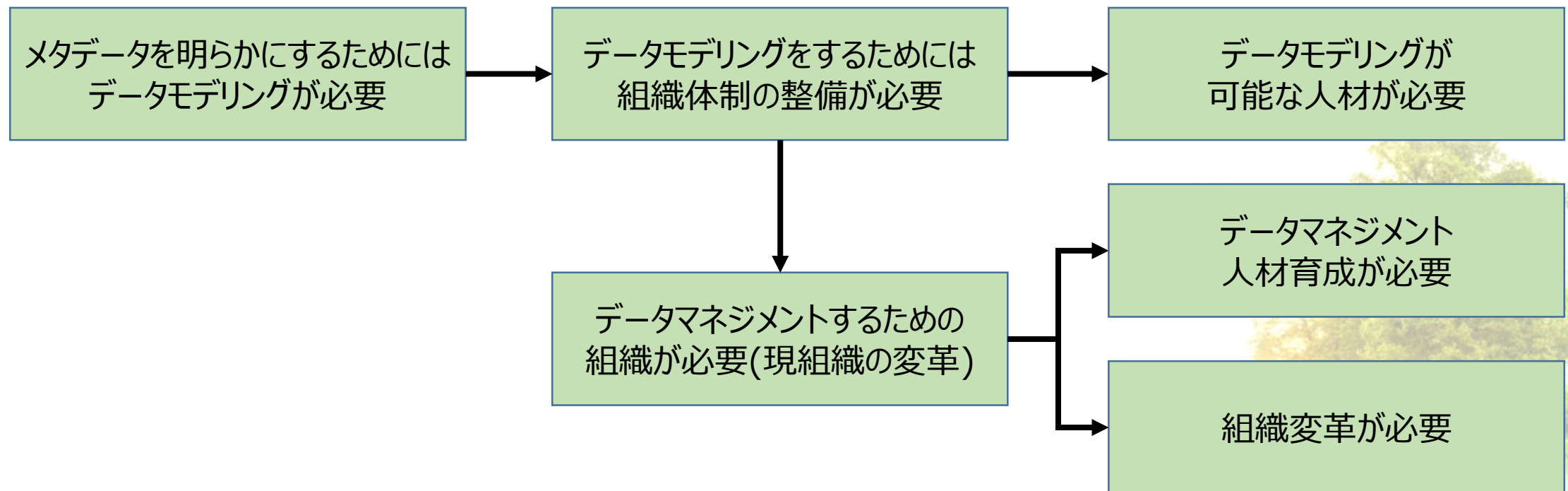
あるべき姿を実現するための直接的なアクティビティを考える
「メタデータリポジトリ」・「データカタログ」ともに
「メタデータ管理」が関連アクティビティとなる。



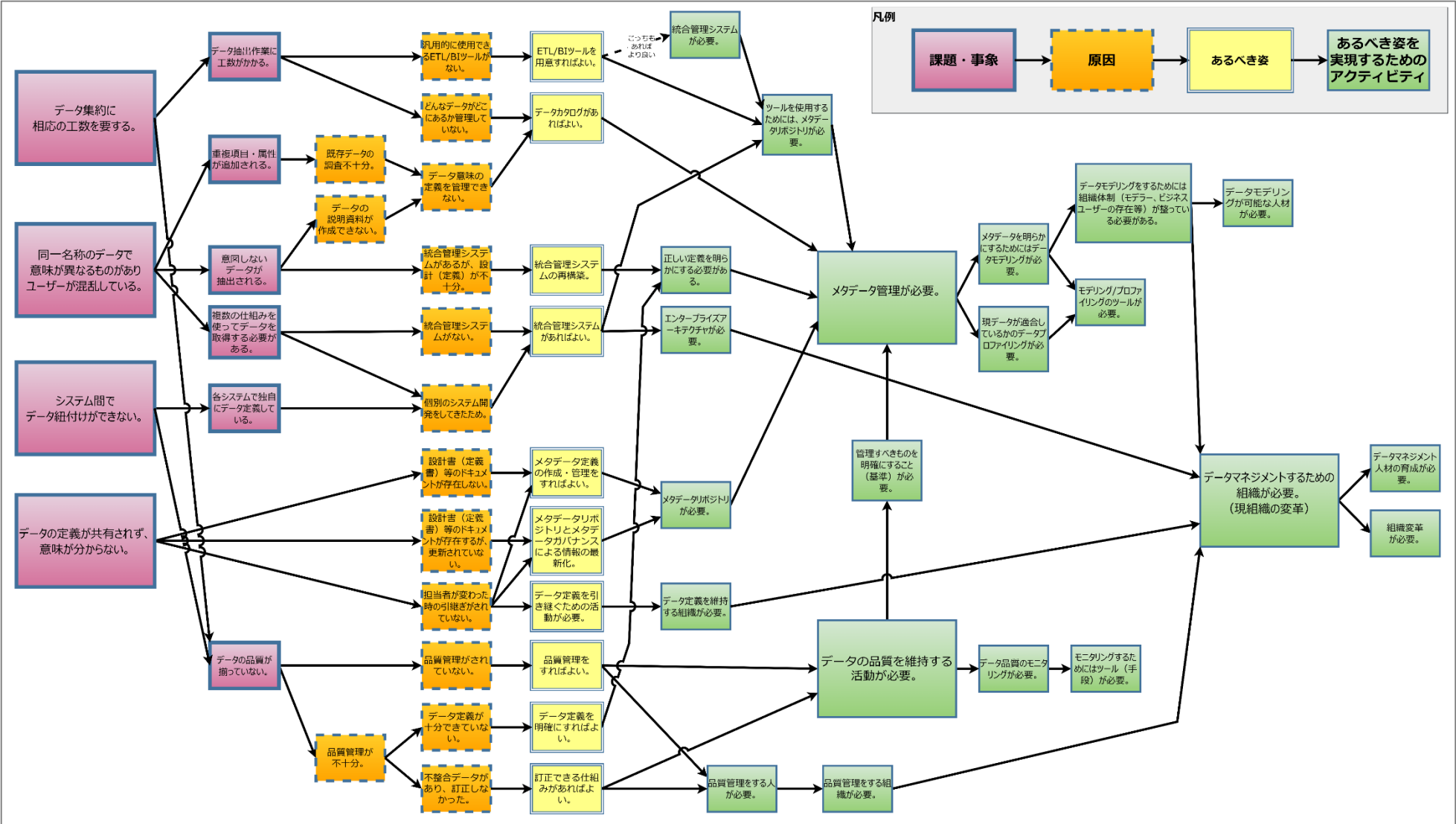
「メタデータ管理」に必要なアクティビティは
「データモデリング」・「データプロファイリング」となる。
そのためにはまず「ツール」が必要となる。



「データモデリング」を実施するためには、「ツール」の他に
「人材確保・育成」・「組織体制の整備」が必要。



検証によって作成されたシナリオの全体像



本日のメニュー

1 利活用の現状

なぜデータ利活用が進まないのか

2 仮説

データ利活用を進めるための取り組み

3 シナリオ

課題からシナリオを作成する

4 評価

検証の評価はどうだったか

5 まとめ

データ利活用へのスタートに向けて



あるべき姿に対するアクティビティ

No.	あるべき姿	アクティビティ	知識領域
1	ETL/BIツールを用意すれば良い	統合管理システムが必要	データ統合
		ツールを使用するためには、メタデータリポジトリが必要	メタデータ
2	データカタログがあれば良い	メタデータ管理が必要	メタデータ
3	品質管理をすれば良い	品質管理をする人が必要	データ品質 組織体制
		データの品質を維持する活動が必要	データ品質
4	データ定義を明確にすれば良い	正しい定義を明らかにする必要がある	メタデータ
5	訂正できる仕組みがあればいい	No.3と同様	データ品質
6	データカタログがあれば良い	メタデータ管理が必要	メタデータ
7	統合管理システムがあれば良い	エンタープライズアーキテクチャが必要	データ統合
8	メタデータ定義の作成・管理をすれば良い	メタデータリポジトリが必要	メタデータ
9	メタデータリポジトリ+メタデータガバナンスによる情報の最新化		メタデータ
10	データ定義を引き継ぐための活動が必要	データ定義を維持する組織が必要	組織体制

内部検証の結論

各課題のアクティビティは
5つの知識領域に密接に関係していた

データ集約に相応の工
数を要する

システム間でデータの
紐付けができない

データの定義が共有さ
れず、意味がわからな
い

同一名称のデータで意
味が異なるものがあり
ユーザーが混乱する

データモデリング

メタデータ管理

データ品質

組織体制

データ統合

「内部検証」では課題の**解決につながる**と評価

「内部検証」で検証した**シナリオの評価**は
本研究分科会メンバー以外（第三者）からも
同様の評価になるかを**アンケート**で確認



アンケートの観点（選択式）

択一 設問

- ・シナリオとチェックリストの内容が理解できるか
- ・チェックリスト※の内容の理解度、活用の可能性

4段階中の3段階以上であれば理解・役に立つと評価

複数回答 設問

- ・4つの課題のうち、貴社の課題と類似するか
- ・シナリオの妥当性
「原因」⇒「あるべき姿」⇒「実行ステップ」を経て
「評価」となる論理展開についての違和感の有無

課題の類似性・シナリオについて納得性を評価

※チェックリスト：5つの知識領域への取り組み状況と評価結果の関連を分析するために作成

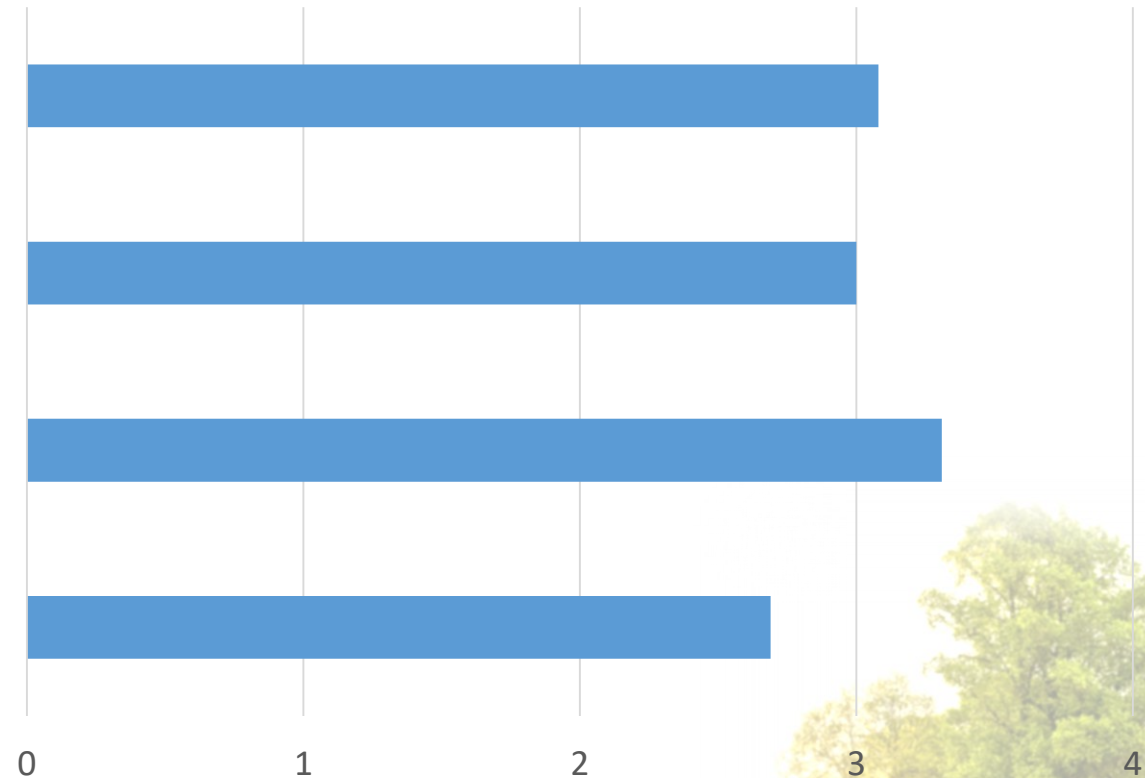
アンケートのまとめ

概説によりシナリオを理解できたか

概説によりチェックリストの基本項目・詳細項目
を理解できるようになったか

チェックリストの基本項目・詳細項目の意味は
理解できたか

チェックリストを使用することで、最初に取り
組むべき活動の実行度の評価に役立ったか



**シナリオ・チェックリストについて
理解・役に立つと評価**

アンケート抜粋（自社の類似性）

データ集約に
相応の工数を要する

- ・データオーナーが異なるデータについて特に集約が困難
- ・データ集約化するためのデータ転送に開発が必要

システム間で
データの紐付けができない

- ・データはシステムごとに定義されることが多い
- ・システムごとに定義が異なり、一つ一つ確認する必要がある

データの定義が共有されず
意味がわからない

- ・ドキュメントが最新化されていない
- ・相手システムのデータ定義が不明

同一名称のデータで
意味が異なるものがあり
ユーザーが混乱する

- ・各システムごとにデータが定義されている

4つの課題に対しての類似点について
具体的な意見が寄せられた

アンケート抜粋（シナリオの妥当性）

シナリオに対して「違和感がない」の回答者は半数以上
「違和感がある」の指摘は7か所

理想論と現実解

- ・あるべき姿に向かうまでにかかる労力やコストが示されていない
- ・組織が存在すべき理由が示されていない
- ・組織に持たせる権限と情報システム部とのかかわり方の想定が不明

意見は社内承認を得るための障害になるのではといった内容
⇒ 納得性を得られる**導入プラン**や**説明**の提示が必要

論理展開の網羅性

- ・課題の原因が全て挙げられていない
- ・あるべき姿の検証が唐突で、シナリオで取り上げられている内容がメタデータ管理に偏っている

各社固有の問題や個別の事象からくる原因が存在する
⇒ 原因分析が**いずれの企業でも一様になるとは限らない**

シナリオとチェックリストは内容を理解でき、
アクティビティ実行度の評価に利用可能で
妥当な内容になっていると判断できた。

一方で

組織や業務によって実用上適用しにくい部分
あらゆる組織の課題全てに対応できない点が
課題と考えられる。



本日のメニュー

1 利活用の現状

なぜデータ利活用が進まないのか

2 仮説

データ利活用を進めるための取り組み

3 シナリオ

課題からシナリオを作成する

4 評価

検証の評価はどうだったか

5 まとめ

データ利活用へのスタートに向けて



本研究分科会で作成した シナリオとチェックリスト の活用方法



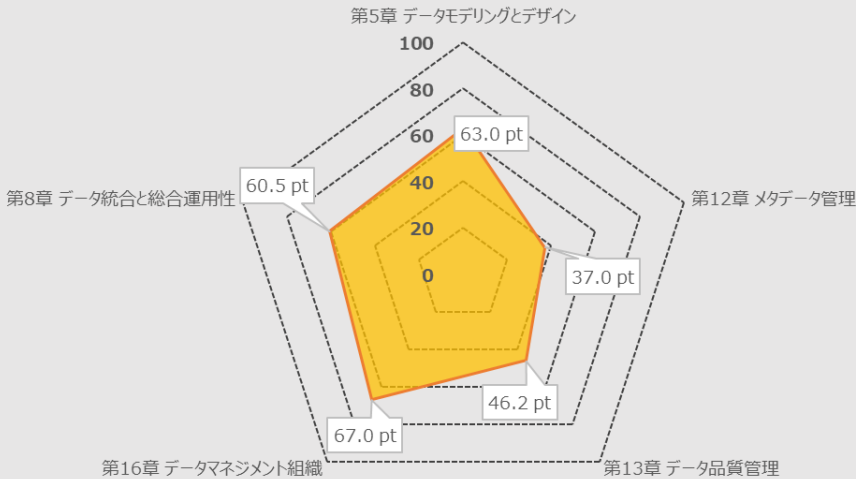
本研究分科会で作成した知識領域チェックリスト

データ利活用のためのデータマネジメント技法の研究～最初に取り組むべき5つの知識領域のレベルチェックリスト～

本チェックリストでは、データ利活用において最初に取り組むべき5要素について、DMBOKのポイントを押さえた現状確認ができます。
「**チェックリスト(解答例)**」シートを参考に、各項目についてご自身の状況をご回答ください。
ご回答結果は、「**チェック結果**」シートでご確認ください。

章	基本項目	詳細項目	回答
《第5章 データモデリングとデザイン》	組織としてデータモデリングを進めていくための体制や指針、変更する際のルールがありますか？		
	組織の中にデータモデルを作成するモデラーが確保されていますか？		
	あるデータモデルに対応する業務側のキーマン(データスチュアート、オーナー)が明らかになっていますか？		
	データモデルの記載ルール(エンティティの色分けなど)がありますか？		
	データモデリングに関する課題一覧を作成するルールがありますか？		
	何らかのデータが変更になった際にデータモデルから追加するような変更ルールになっていますか？		
	データのFROM-TOを示す表を作成するルールがありますか？		
	データモデルの構築時のプロセスが確立していますか、またそのプロセスではオーナー(業務側)と段階的に関与していますか？		
	各分野のデータのオーナー(業務担当者)を決めていますか？		
	モデラーがモデルを作成した後、業務要件と合致しているかオーナーと確認する体制となっていますか？		
	精確にするために、要件抽出と文書化が反復的なプロセスになっていますか？		
	要件と其中で扱う重要な用語を把握するために、概念データモデルを作成していますか？		
	組織やプロジェクトの標準に合ったスキーム(リレーショナル、ディメンショナル、ファクトベースなど)を採用していますか？		
	スキームに紐づく表記法(EやORMなど)は組織の標準に合っていますか？		
	同じ概念の用語に対してユーザー側と会社側で使う用語が異なっていた場合、両方で調整し一貫性を取るようにしていますか？		
	概念データモデル作成後に要件を満たしているのかレビューする体制となっていますか？		
	概念データモデルの詳細なデータ要件を表現するために、論理データモデルを作成していますか？		
	既存のデータモデルやデータベースからの作業結果が正確かつ最新かどうか分析をしていますか？		
	各分野ごとに業務データ要件をエンティティ、リレーションシップ等を明らかにしたデータモデルを作成していますか？		
	その業界として理解しやすい表現(表や属性の文言など)ができていますか？		
	組織の環境に合わせた環境を設計するために、物理データモデルを作成していますか？		
	各テーブルと列、ファイルとフィールド、スキーマと要素などに付ける体系的な名称やNULL値のような制約など詳細を盛り込めていますか？		
	データベース内のデータにアクセスするとにパフォーマンス向上のために適切なインデックスを追加もしくは検討していますか？		
	パフォーマンス向上のためにパーティション分割やビューの作成を導入もしくは検討していますか？		
	データモデルの品質を定期的に見直す管理基準(PDCAサイクルなど)がうまく機能していますか？		
	データモデルの品質(正確性、完全性、一貫性)を評価する基準がありますか？		
	例		

最初に取り組むべき5つの知識領域 チェック結果



チェックリストの内容

設問

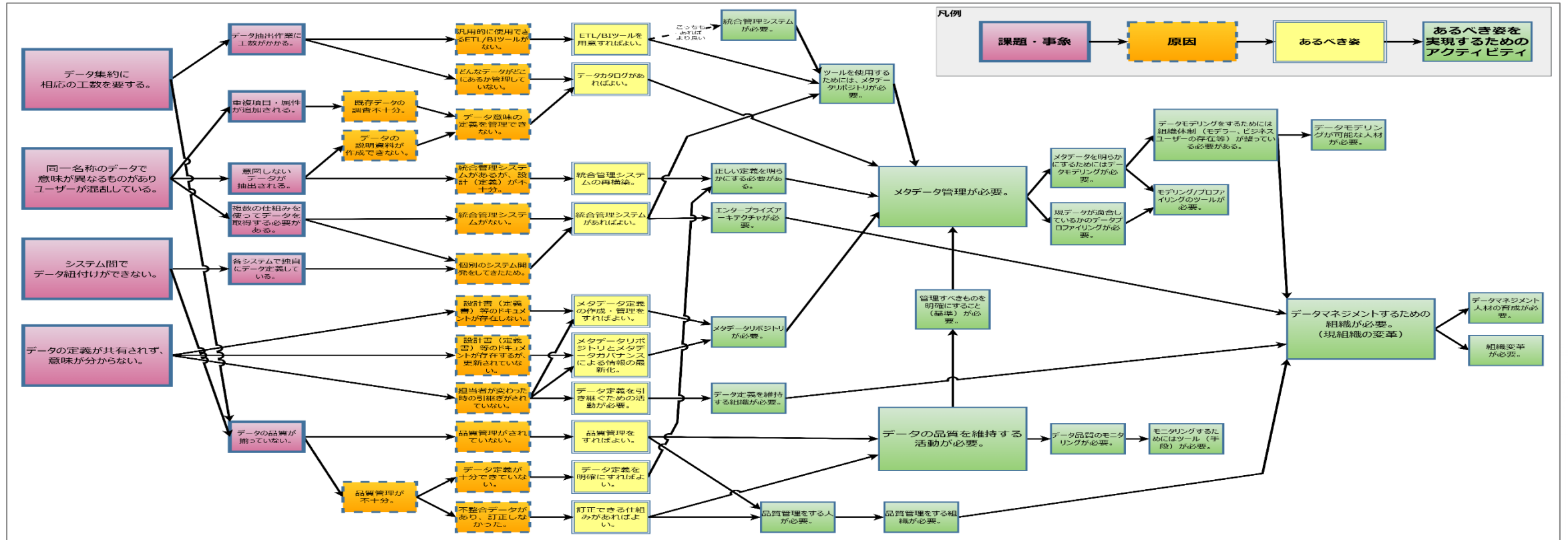
35の基本項目・122の詳細項目の
取り組みの実行度を5段階で評価

スコア

5つの知識領域ごとにアクティビティ
実行度が100点満点でスコア化

①現在の自社の**取り組み状況**を把握

本研究分科会で作成したシナリオ



②シナリオの展開方法を学び、
自組織にあったシナリオ作成に繋げる

自組織に合わせて
テラリングして
利用することで



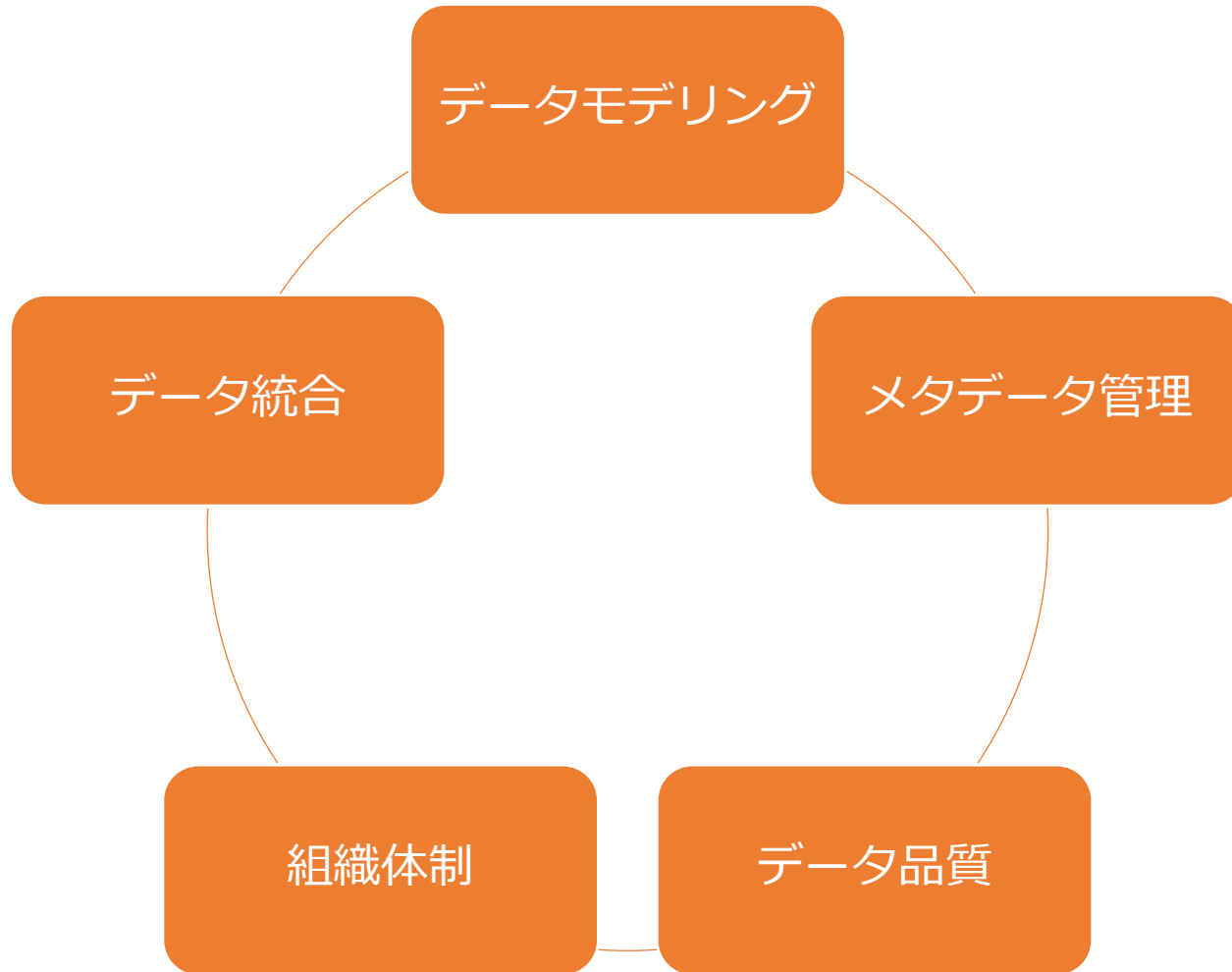
効果的にデータ利活用への
スモールスタートを推進できます



最後に

データ利活用が進まない
といっても
なにからはじめてよいか
わからない方は





まず・・・
論文で5つの知識領域の
概説を読むことから
始めてください

次に・・・
チェックリストを活用して
自社の取り組み状況を
把握してください





そして・・・
この発表で聞いていただいた
シナリオの展開方法で
自社のシナリオを作成
行動計画に落とし込み
実践してください

課題解決への
第一歩を
踏み出すことに
寄与できると
確信しています

