

# 第3部 データマネジメント

長岡技術科学大学 システム安全専攻

張 坤

2020.9.20

kunzhang@vos.nagaokaut.ac.jp

# 内容

## 第3部 データマネジメント

1 3. 安全ビッグデータと安全スモールデータ (張)

1 4. データマネジメント戦略と実践 (張)

1 5. 実践事例学ぶ：品質管理とデータ活用 (佐藤)

概説：データマネジメントについて

# 1. なぜ？

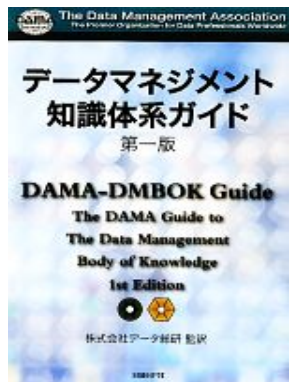
- ◆競争力を優位に保つため
- ◆優れた意思決定を下せないため
- ◆好機を見逃すことを妨げないため
- ◆データ資産から価値を得られるため

**ビジネス上の意義**

## 2. 定義：

データと情報資産の価値を  
獲得し、統制し、保護し、提供  
し、向上させるために、ポリシー、  
実践、及びプロジェクトについて  
計画し、実行し、監視する活動。

(出所：DAMA-DMBOK (Data Management Body of Knowledge) 第2版、2017)



(第1版, 2010)



(第2版, 英語版,2017 日本語版2018)

## 重点：

目的：データの利活用



活動：データの維持/管理

### 3. 原則：

#### 効果的なデータマネジメントにはリーダーのコミットが不可欠

データは価値を持つ

- データは特有の性質をもつ資産
- データの価値は経済的観点で評価可能であり評価されるべき

データマネジメント要件はビジネス要件そのものである

- データの管理はデータ品質の管理
- データ管理にはメタデータが必要
- データマネジメントには予め計画が必要
- データマネジメント上の要件がIT上の決定を左右する

データマネジメントには多様なスキルが求められる

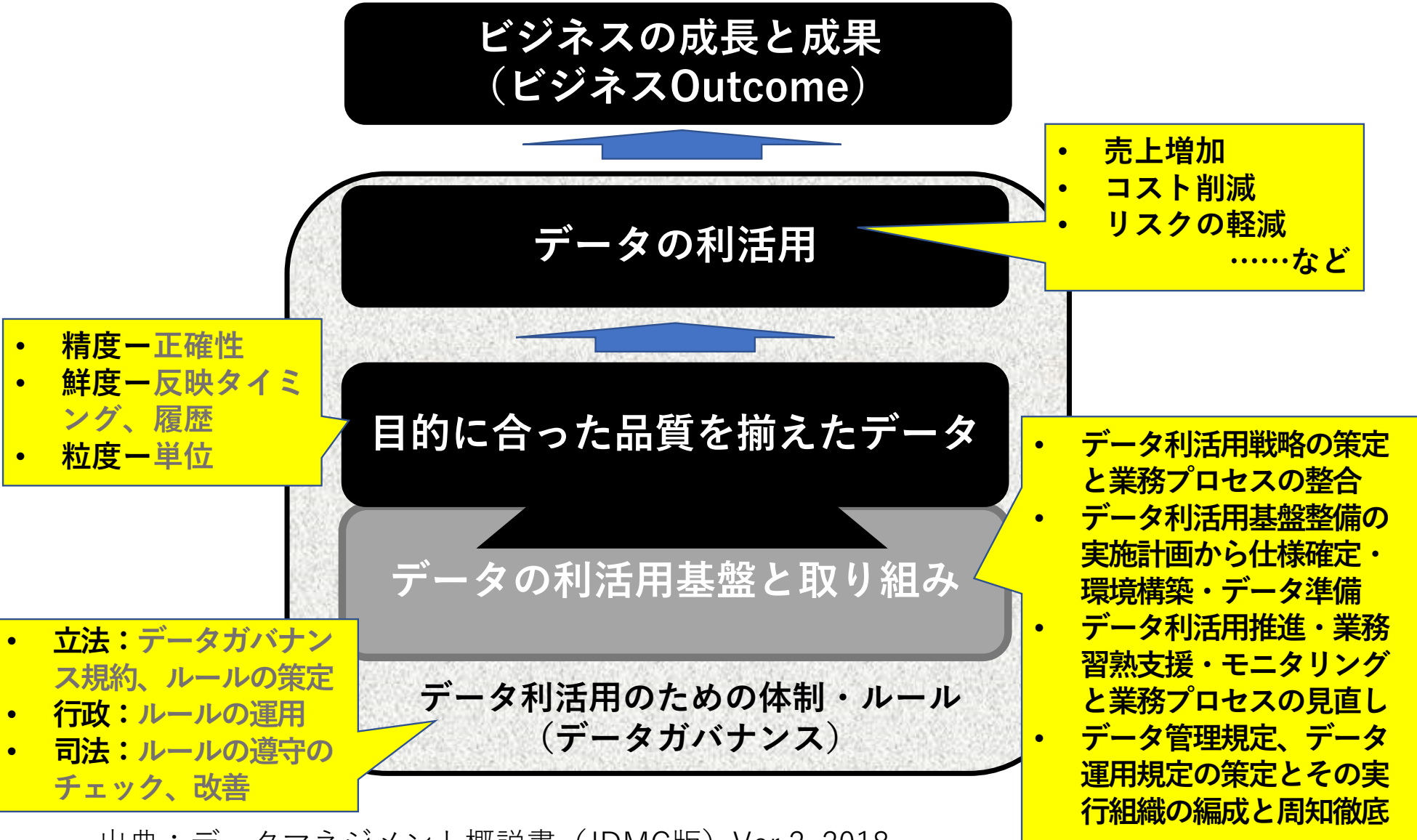
- データマネジメントは機能横断的な活動
- データマネジメントには全社的な視点が必要
- データマネジメントは幅広い視点を考慮することが必要

データマネジメントはライフサイクル管理である

- データは種類によってライフサイクルの性格がことなる
- データに伴うリスクを管理することもデータマネジメントに含まれる

# 4. データマネジメント

## 基本概念図

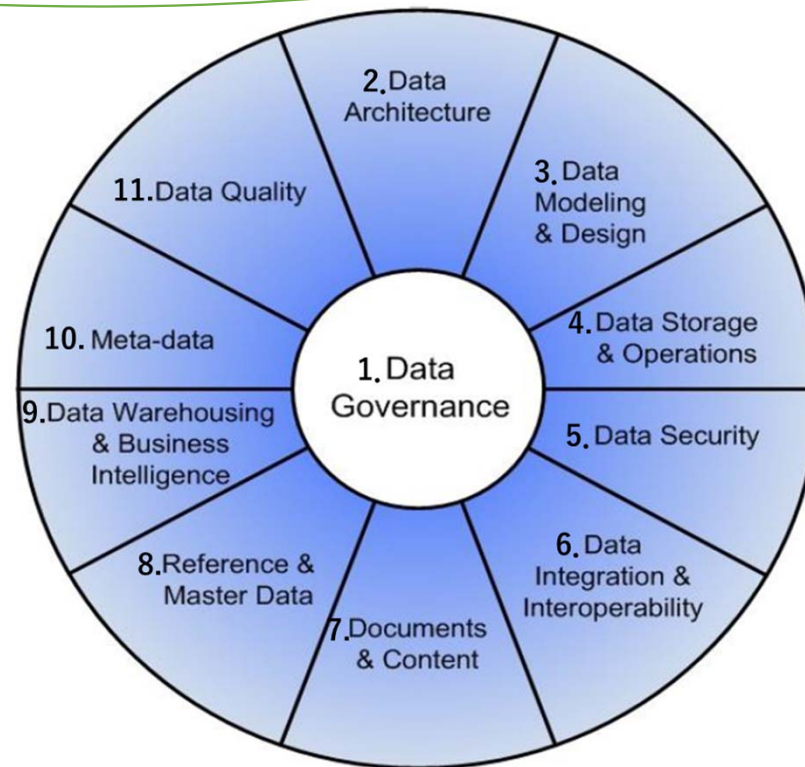


出典：データマネジメント概説書 (JDMC版) Ver.2. 2018

## 5. フレームワーク:

1. データガバナンス
2. データアーキテクチャ
3. データモデリングとデザイン
4. データストレージとオペレーション
5. データセキュリティ
6. データ統合と相互運用性
7. ドキュメントとコンテンツ管理
8. 参照データとマスターデータ
9. データウェアハウジング&ビジネスインテリジェンス
10. メタデータ
11. データ品質

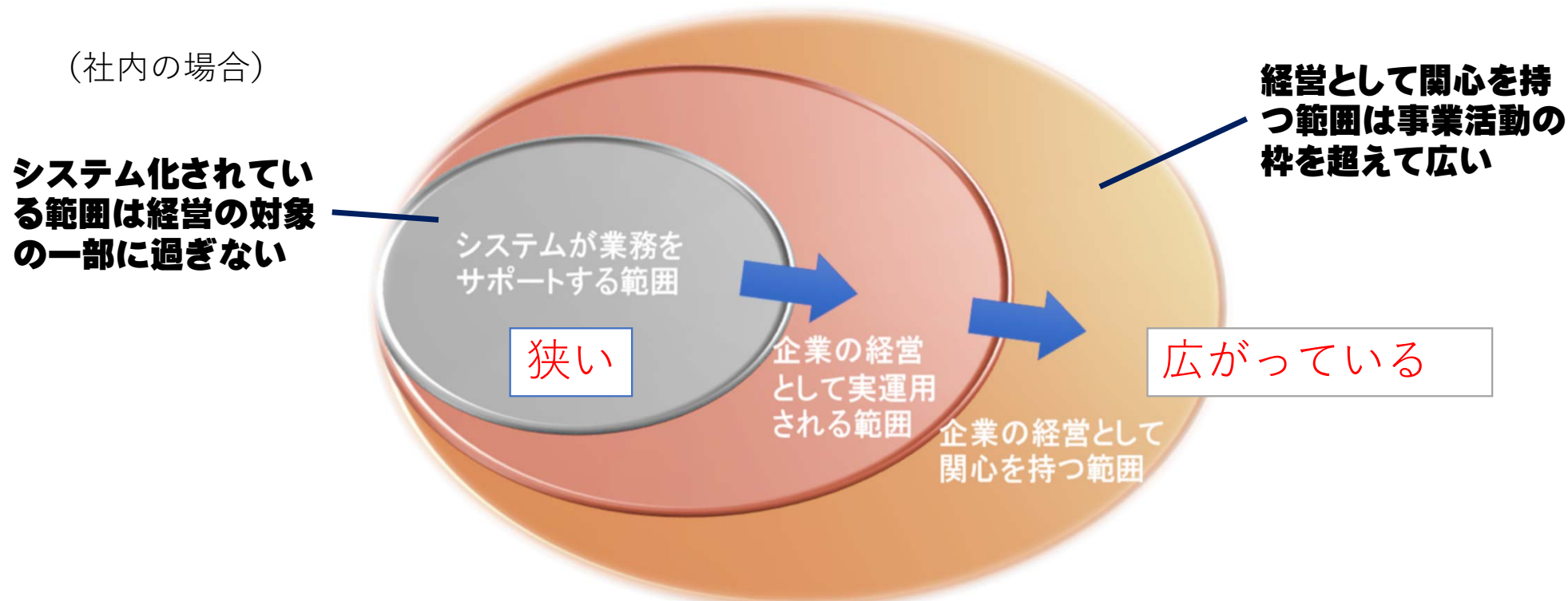
DAMA-DMBOKフレームワークはデータマネジメントの  
**全範囲の知識領域を定義**する。



(出所: DAMA-DMBOK (Data Management Body of Knowledge) 第2版、2017)



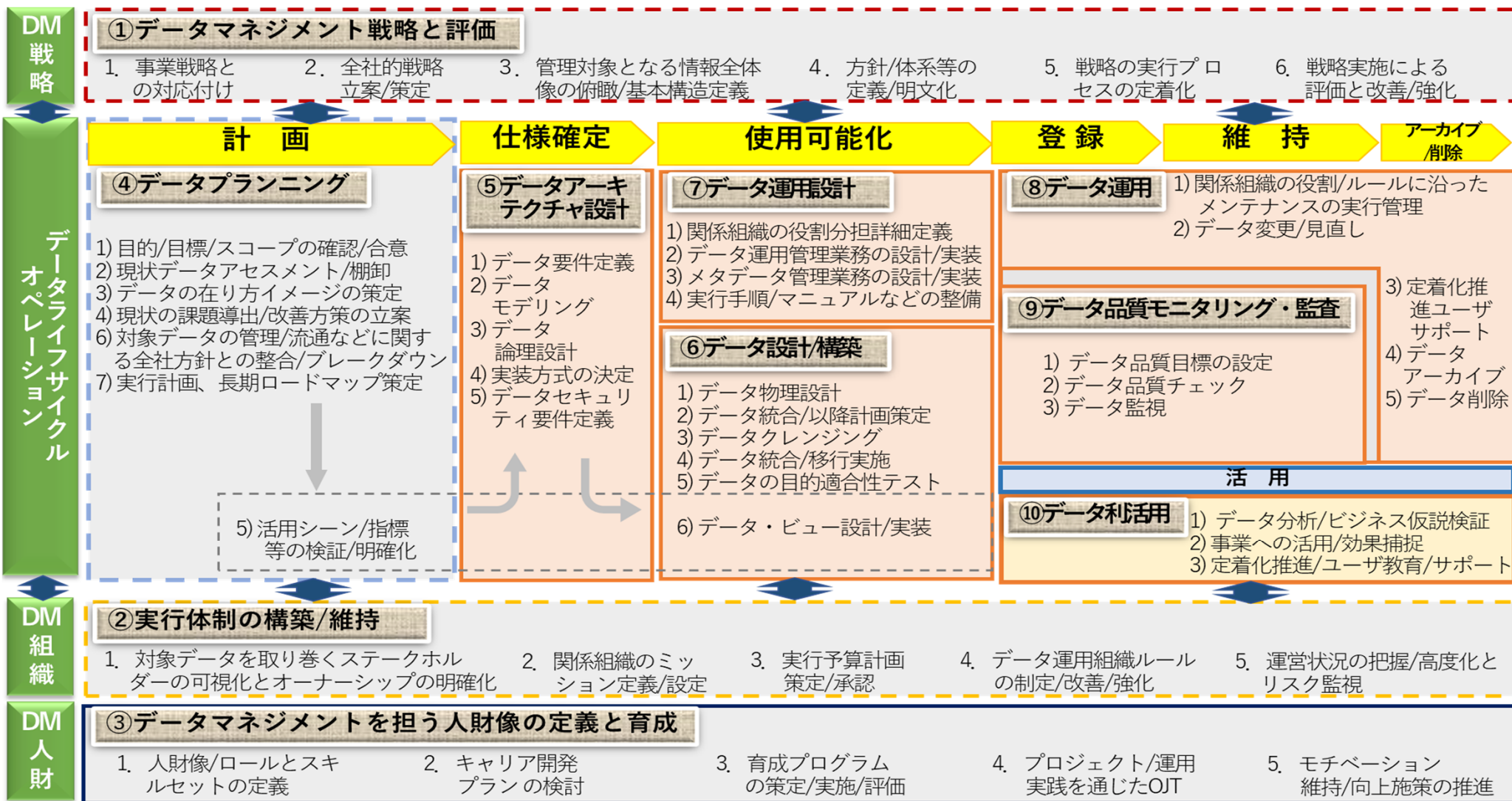
(参考：データマネジメントが対象とするデータの範囲 (一例) )



データマネジメントが対象とするデータの範囲

(出所：「データマネジメントの基礎と価値」研究会、データマネジメント概説書 (JDMC版 Ver.2) 、2011)

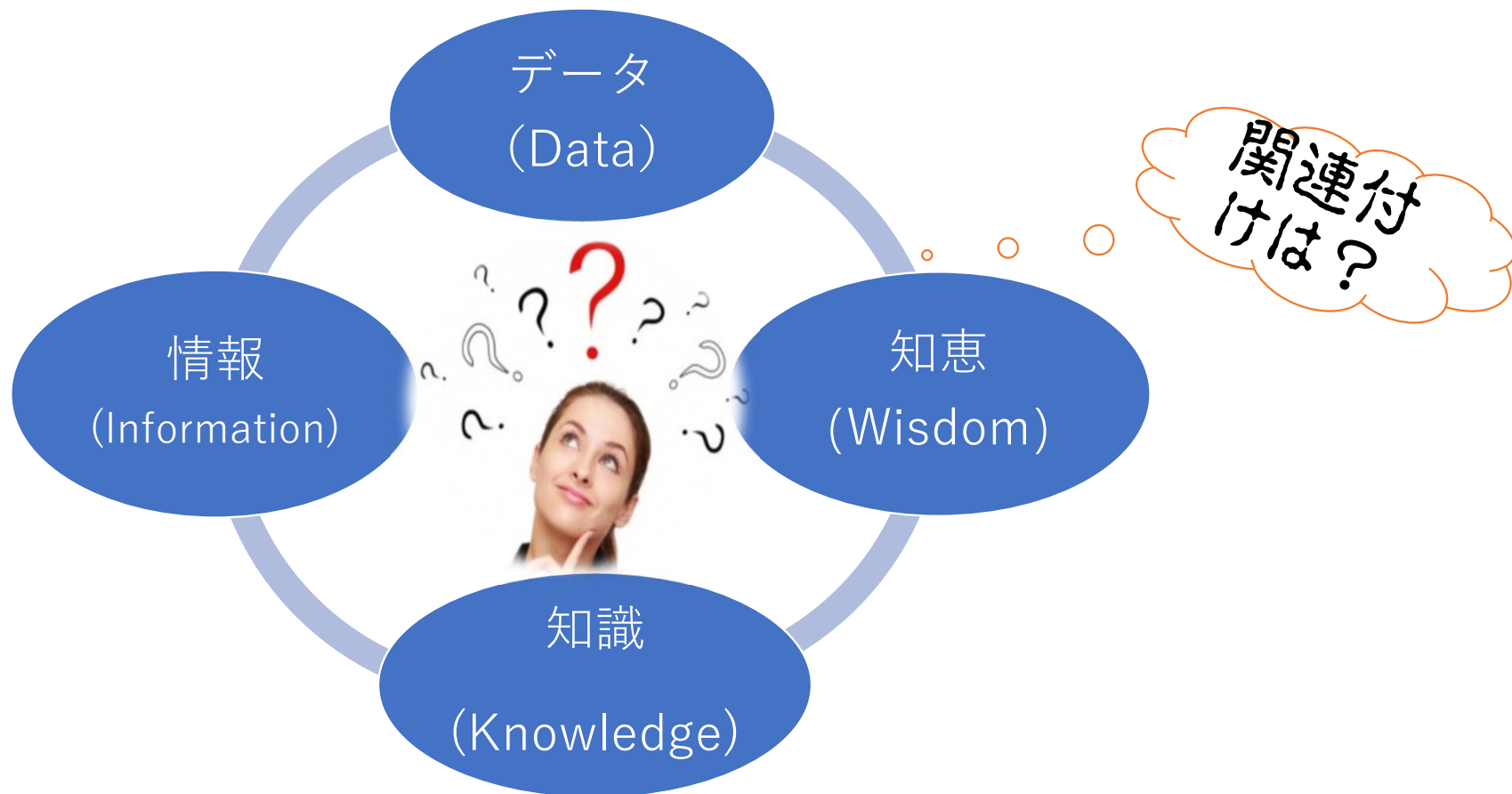
# (参考) データマネジメント構成要素



(出典: 「【データマネジメント・ケーススタディ トップダウン編】顧客満足度向上のための業務横断データ活用」, 一般社団法人 日本データマネジメント・コンソーシアム 『データマネジメントの基礎と価値』 研究会, Ver1.0, P11, 2018) 10

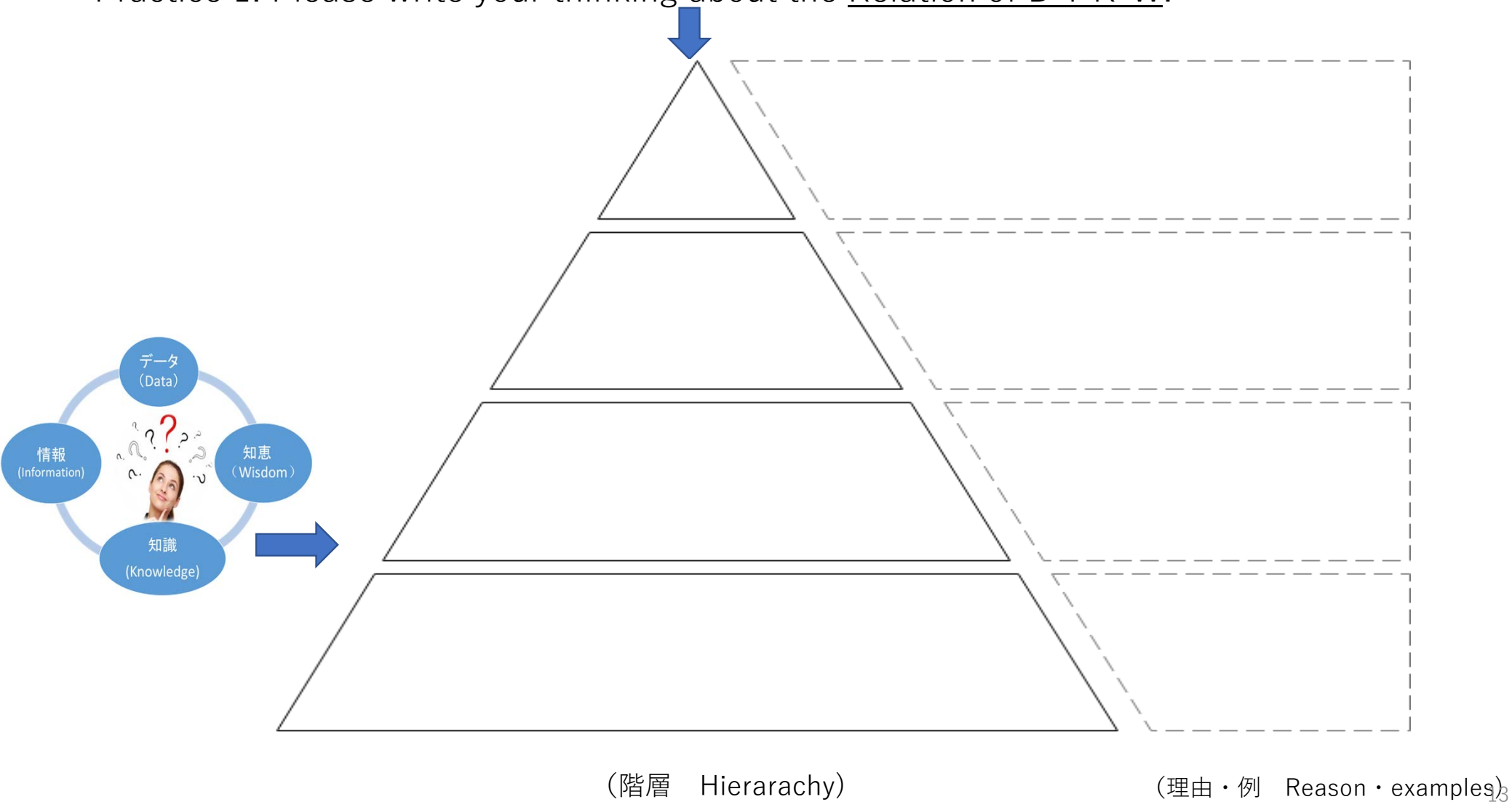
### 1 3. 安全ビッグデータ (SBD) と 安全スモールデータ (SSD)

# 1. 基本概念:



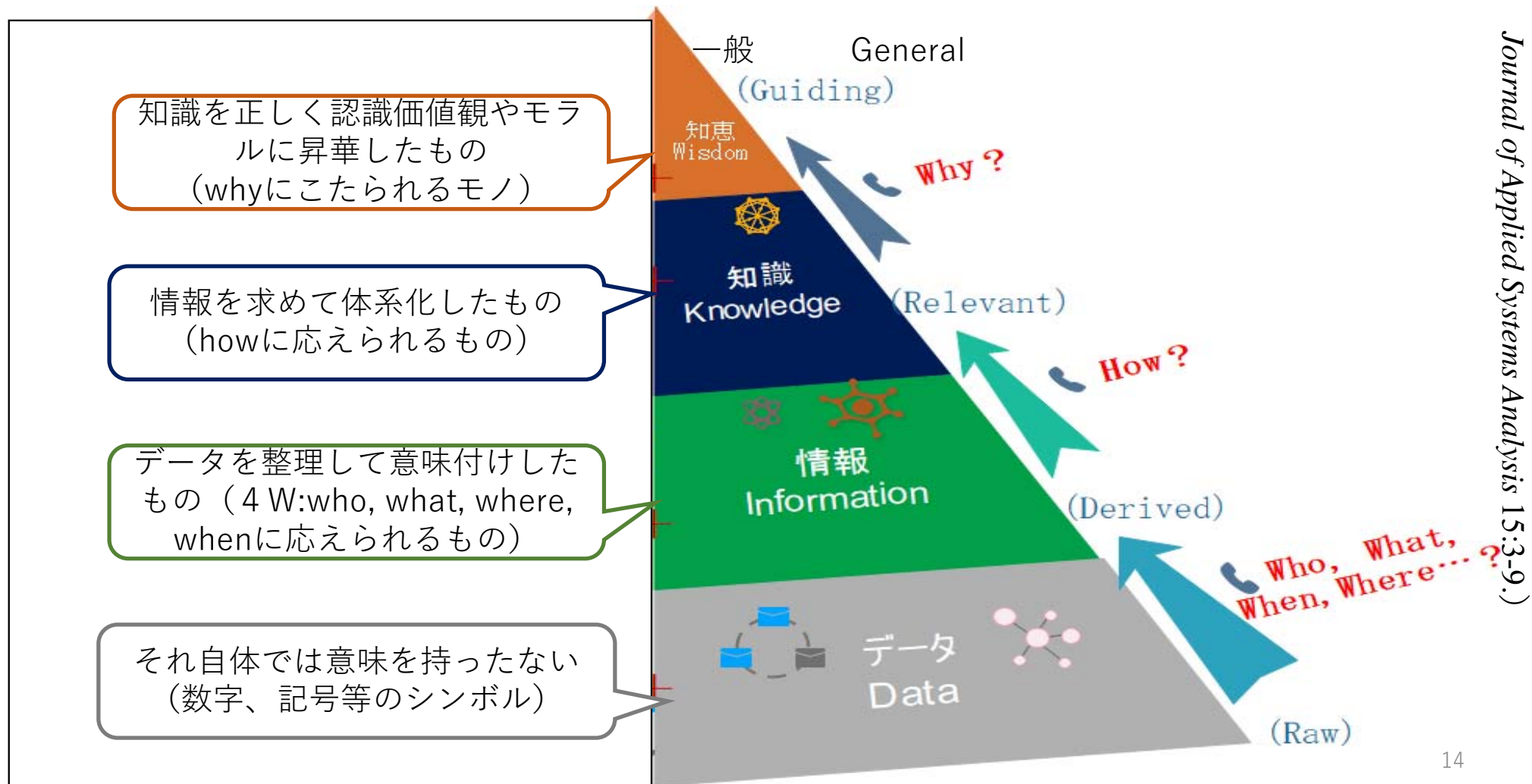
演習：D-I-K-Wの関連性についての考えを書いてください！

Practice 1: Please write your thinking about the Relation of D-I-K-W!



## 2. DIKWモデル (ピラミッド)

# 一 データの世界の概念



## (参考)

The Eagle soars in the summit of Heaven,  
The Hunter with his dogs pursues his circuit.  
O perpetual revolution of configured stars,  
O perpetual recurrence of determined seasons,  
O world of spring and autumn, birth and dying  
*The endless cycle of idea and action,  
Endless invention, endless experiment,  
Brings knowledge of motion, but not of stillness;  
Knowledge of speech, but not of silence;  
Knowledge of words, and ignorance of the Word.  
All our knowledge brings us nearer to our ignorance,  
All our ignorance brings us nearer to death,  
But nearness to death no nearer to GOD.*



Stories of the Human Spirit  
T. S. Eliot (1888-1965):  
*The Rock* (1934)

Edited by Peter Y. Chou  
WisdomPortal.com

ノーベル賞受賞者

受賞年：1948年

受賞部門：ノーベル文学賞

受賞理由：今日（こんにち）の  
詩文学への卓越した貢献に対し  
て

*Where is the Life we have lost in living?*

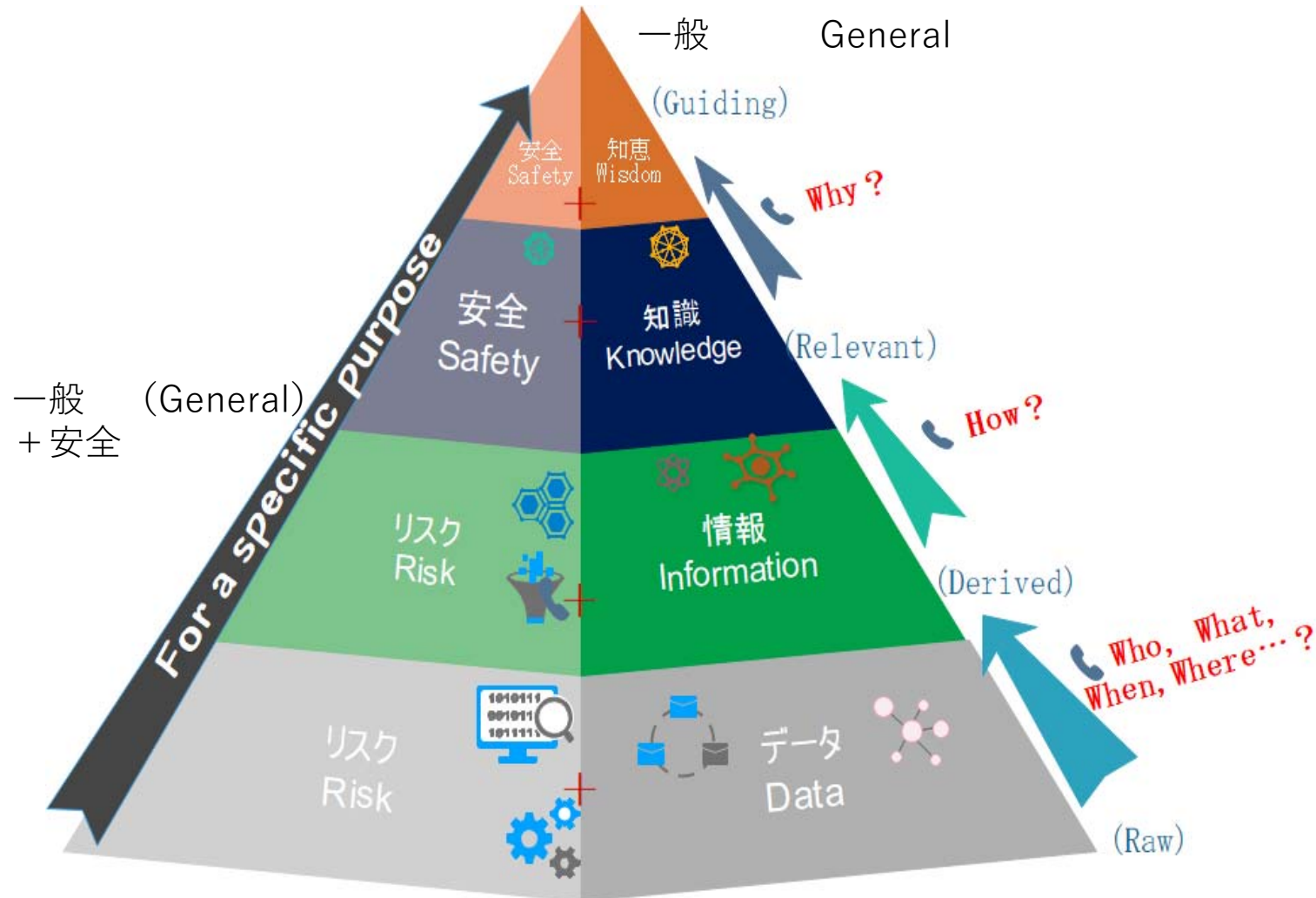
*Where is the wisdom we have lost in knowledge?*

*Where is the knowledge we have lost in information?*

The cycles of Heaven in twenty centuries  
Bring us farther from GOD and nearer to the Dust.

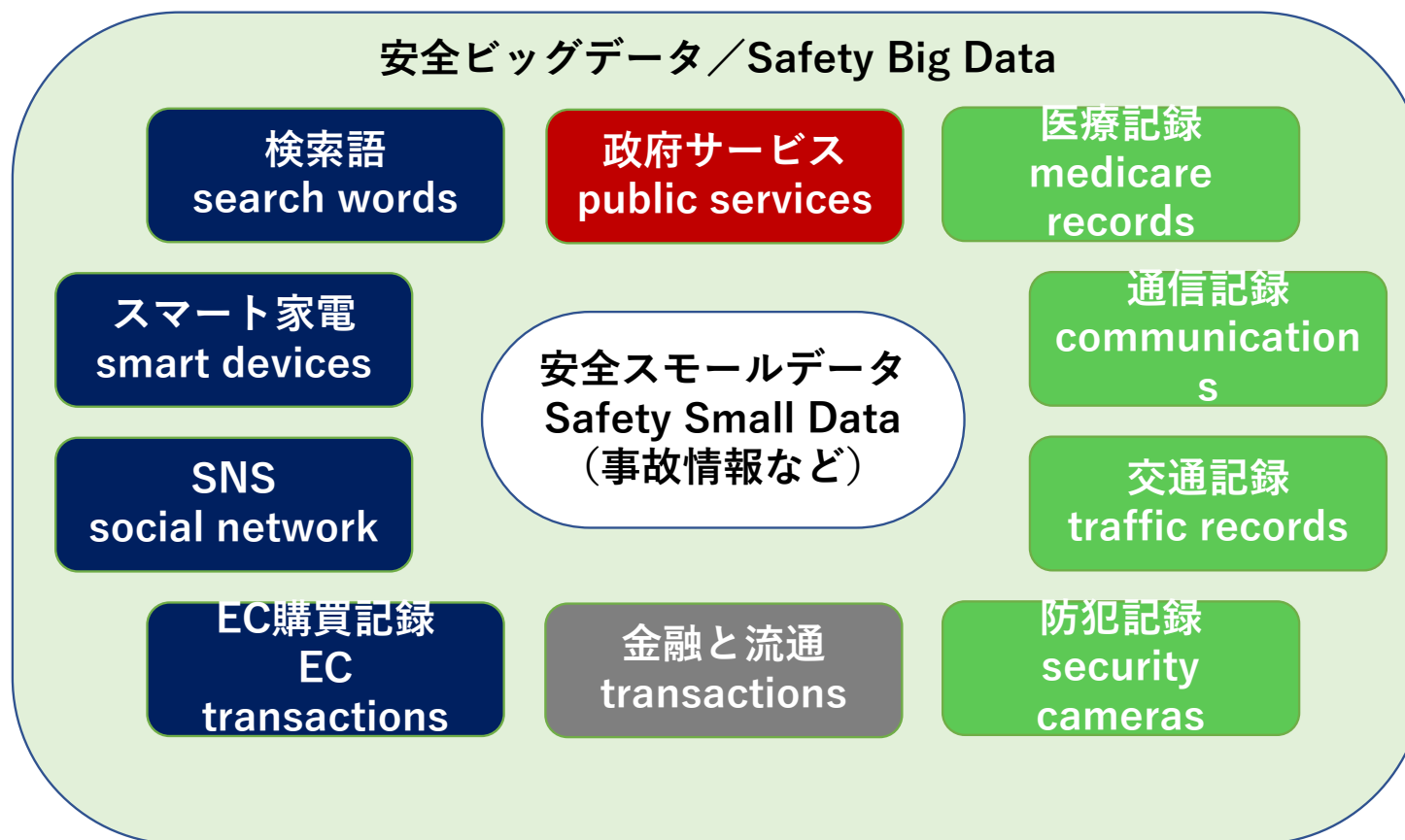
(出典：  
<http://www.wisdomportal.com/Technology/TSEliot-TheRock.html>)

## 2. DIKWモデル (ピラミッド) — データの世界の概念






# (参考) 安全ビッグデータと安全スモールデータ



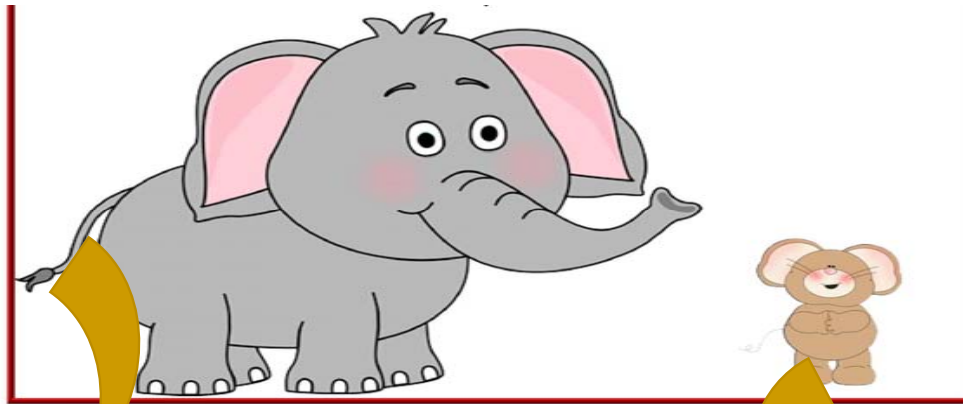
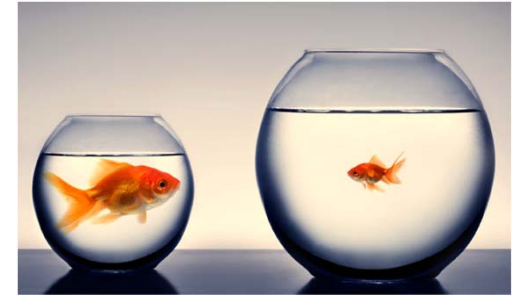
### 3. SBD VS SSD\_データの種類 (例)



カテゴリー	データの実例	データ発信源	volume & velocity	
(SBD)	IoTデータ	生活ライフログ ドライブデータ 携帯GPS情報 POSデータ モニター画像	スマート家電 自動車ITS機器 携帯端末 POS端末 街頭カメラ	<ul style="list-style-type: none"> <li>データ量 &gt;&gt; 人口</li> <li>発生速度 &gt;&gt;</li> <li>リアルタイム</li> </ul>
	非構造化データ	SNS上の書き込み 検索語	非専門家	<ul style="list-style-type: none"> <li>データ量 <math>\propto</math> 人口</li> <li>発生速度</li> <li>リアルタイム</li> </ul>
(SSD)	構造化データ	○各種事故記録 (死亡届, 労災, 交通事故, 火災, 製品事故等) ○救急搬送 ○診療保険支払記録	専門家  消防署 医療機関	<ul style="list-style-type: none"> <li>データ量 &lt;&lt; 人口</li> <li>発生速度</li> <li>時間遅れあり</li> </ul>

### 3. SBD VS SSD\_特徴

安全データの “大” と “小”



SBD :

- ◆全体性；
- ◆拡張の可能性；
- ◆採掘（マイニング）の可能性

SSD:

- ◆ターゲティング；
- ◆良質；
- ◆処理しやすい

### 3. SBD VS SSD\_\_一般的な特性

	SBD	SSD
哲学的基盤	異質性	均質性
操作モード	機械→人間	人間→機械
データの状態	動的 (Dynamic)	静的 (Static)
データの次元	多次元	一次元
収集の着眼点	シーンデータ	非シーンデータ
データモードの関係	相関関係	因果関係
メインデータ型	非構造化	構造化
データのオリエンテーション	サブジェクト指向	オブジェクト指向

(出所 : Qiumei Ouyang, Chao Wu, Lang Huang. Methodologies, principles and prospects of applying big in safety science research. Safety Science 101 (2018) 60-71)

### 3. SBD VS SSD\_\_データの統計処理

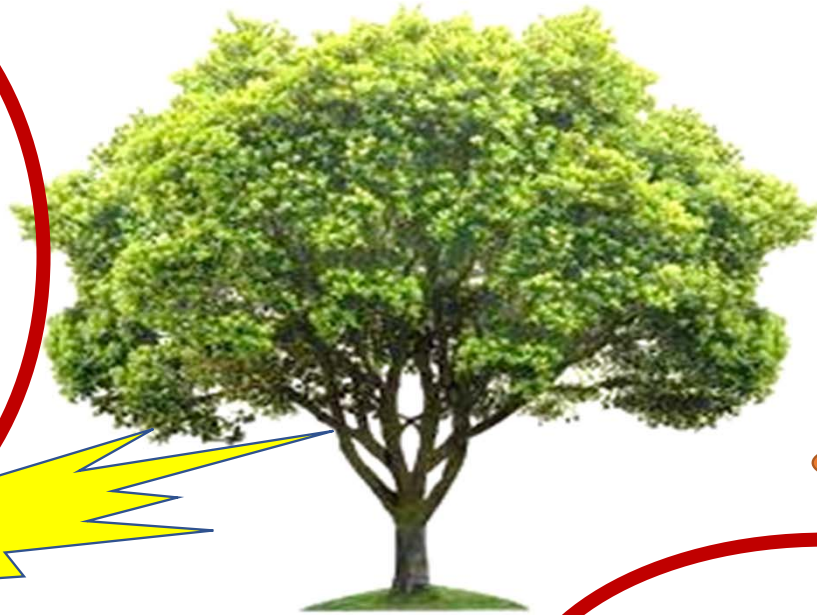
	SBD	SSD
統計的強調点	安全データの処理 (reducing)	安全データの収集 (adding)
統計の難点	有用な安全データを選ぶ (how)	安全データを収集する (how)
統計の機能	複雑なビッグデータを単純化し	スモールデータからの帰納推論
メイン統計方法	相関分析 (what)	因果分析 (why)
統計目的	全てのデータから有用なデータを発掘 発掘する	ランダム現象からの規則をまとめる
データ処理のプロセス	整理→分析-開発→応用	設計→収集→分析→開発→応用
データ処理の考え方	仮説←相関性	仮説→相関性
品質管理	事前予測	事後テスト
統計結果	近似値	正確な値
データ源	リアルタイム記録	ヒューマンインプット
処理コスト	低い (Lower)	より高い (Higher)

(出所：Qiumei Ouyang, Chao Wu, Lang Huang. Methodologies, principles and prospects of applying big in safety science research. Safety Science 101 (2018) 60-71)

### 3. SBD VS SSD\_\_相互利用

BIG  
DATA  
tells us  
**what's**  
happening

何が起っているの



何故こうなったの

small data tells us  
**why** it's  
happening.



## 4. ビッグデータ & データサイエンス

✓ 流行語として不用意に使われてきた

✓ ビッグ: 相対的である

✓ データサイエンス: 以前は「応用統計学」と呼ばれていた

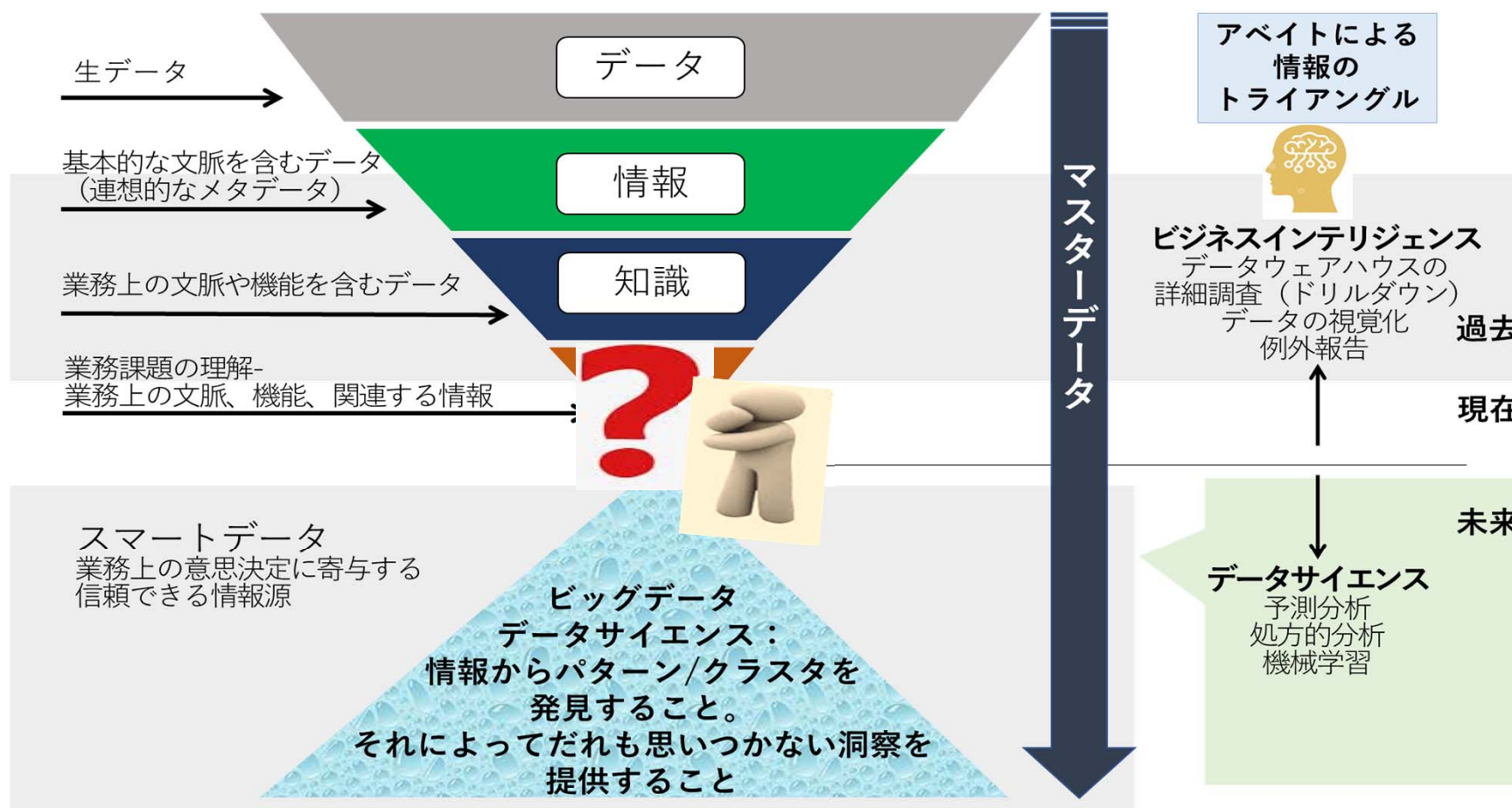
**ビッグデータ + データサイエンス:  
両者そろって重要な技術的変化を生み、**

データサイエンスという学問の神髄：統計的、計算的、人間的の3視点から俯瞰すること  
（Blei and Smyth, "[Science and data science](#)"  
(英語). *Proceedings of the National Academy of Sciences* **114** (33): 8689–8692, 2017)

**データサイエンス:  
データを分析し価値を得る新しい方法を発見したのである**

# (参考)

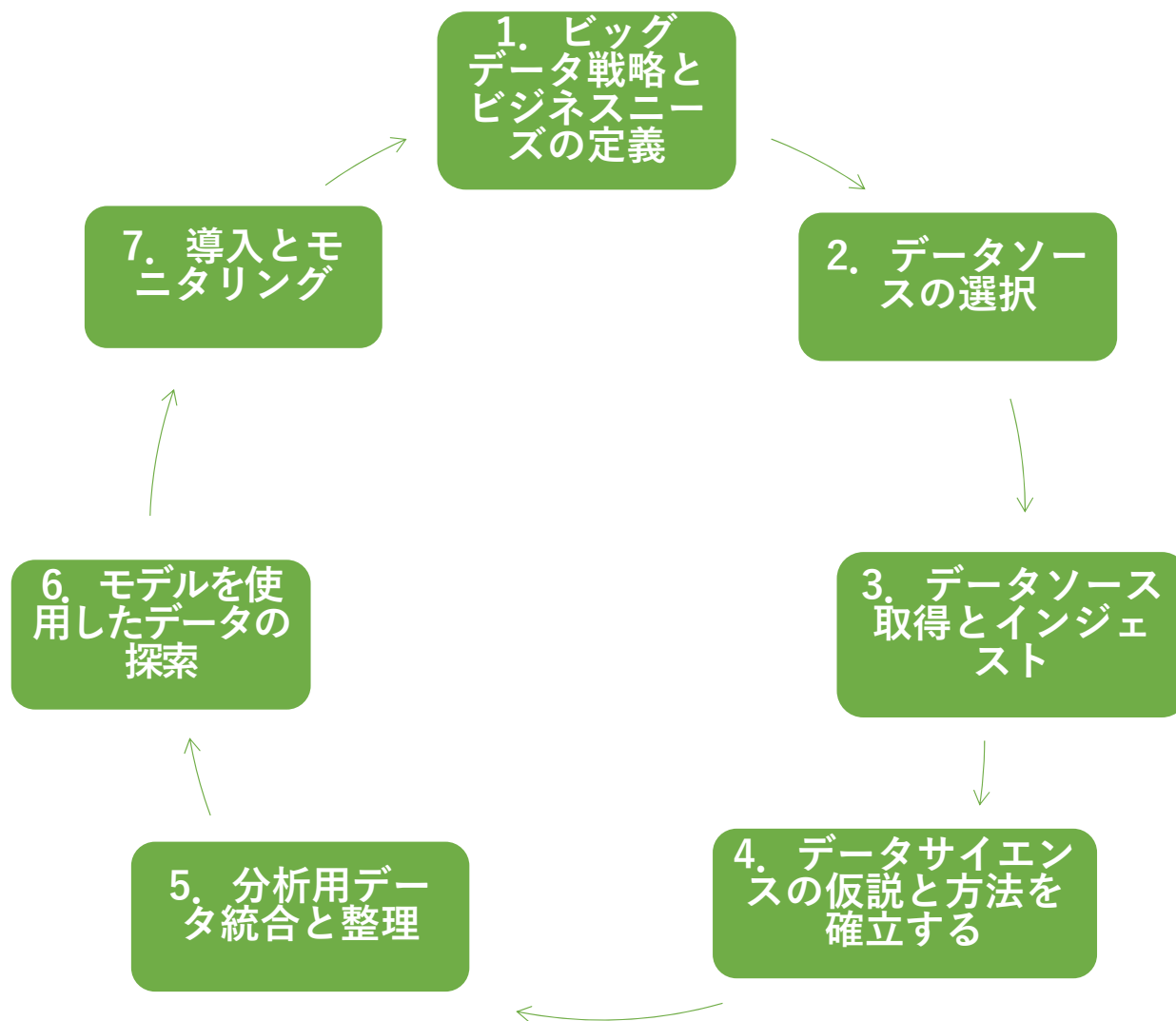
## (参考)アベイトによる情報のトライアングル (Abate Information Triangle)



(出所：DAMA-DMBOK (Data Management Body of Knowledge) 第2版、2017)



# (参考)



(出所：DAMA-DMBOK (Data Management Body of Knowledge) 第2版、2017)

(参考) データサイエンスプロセス

## 14. データマネジメント戦略と実践

# 1. データマネジメント戦略の定義：

**戦略**：大きな目標を達成するための概略行動指針が描けるように、一連の選択肢を考慮し意思決定を下すことである。

**データ戦略**：情報を利用することで競争上の優位性を確保し、企業の目標を達成するための事業計画が含まれる。

**データマネジメント戦略**：データ戦略をサポートするためデータマネジメント戦略が必要となる。一般的には、COD(チーフ・データオフィサー)が主幹となって維持し、データガバナンス委員会がサポートするデータガバナンスチームによって制定されることが多い。

## 2. データマネジメント戦略の要素：

- データマネジメントに対する説得力のあるビジョン
- 具体例を含んだデータマネジメントの業務提案概要
- 基本理念、価値観、マネジメントの視点
- データマネジメントの使命と長期的な目標
- データマネジメントの成功度合を評価する基準
- 短期的（12-24ヶ月）SMART（Specific：具体的、Measurable:測定可能、Actionable:実行可能、Realistic:現実的、Time-bound有期的）データマネジメント・プログラムの達成目標
- データマネジメントの役割と組織の説明、責任と決定権の概要
- データマネジメント・プログラムの要素と構想の説明
- 実施範囲を定めた優先プログラム
- プロジェクトやアクション項目を含む実施ロードマップの草案

### 3. データマネジメントの戦略的計画案の成果物

#### • データマネジメントチャーター：

□ 全体的なビジョン、ビジネスケース、目標、基本理念、成功の尺度、重要な成功要因、錦されたリスク、オペレーティングモデルなど

#### • データマネジメントスコープの記述：

□ 計画期間（通常3年）の目標と目的、それを達成するために責任を負う役割、組織と個々のリーダー

#### • データマネジメントの実施ロードマップ：

□ 特定のプログラム、プロジェクト、タスクの割り当てと導入マイルストーンの提示

(注：データマネジメントの戦略は、組織が取り組むべき、全てのデータマネジメント・フレームワークの知識領域に対処する必要がある)



(出所：DAMA-DMBOK (Data Management Body of Knowledge) 第2版、2017)

### 3. データマネジメント成熟度アセスメント (Data Management Maturity Assessment) :

#### <定義>

組織内で実施されているデータ関連業務に対してランク付けする方法。  
データマネジメントの現状とそれが組織に与える影響を明らかにする。

#### <ゴール>

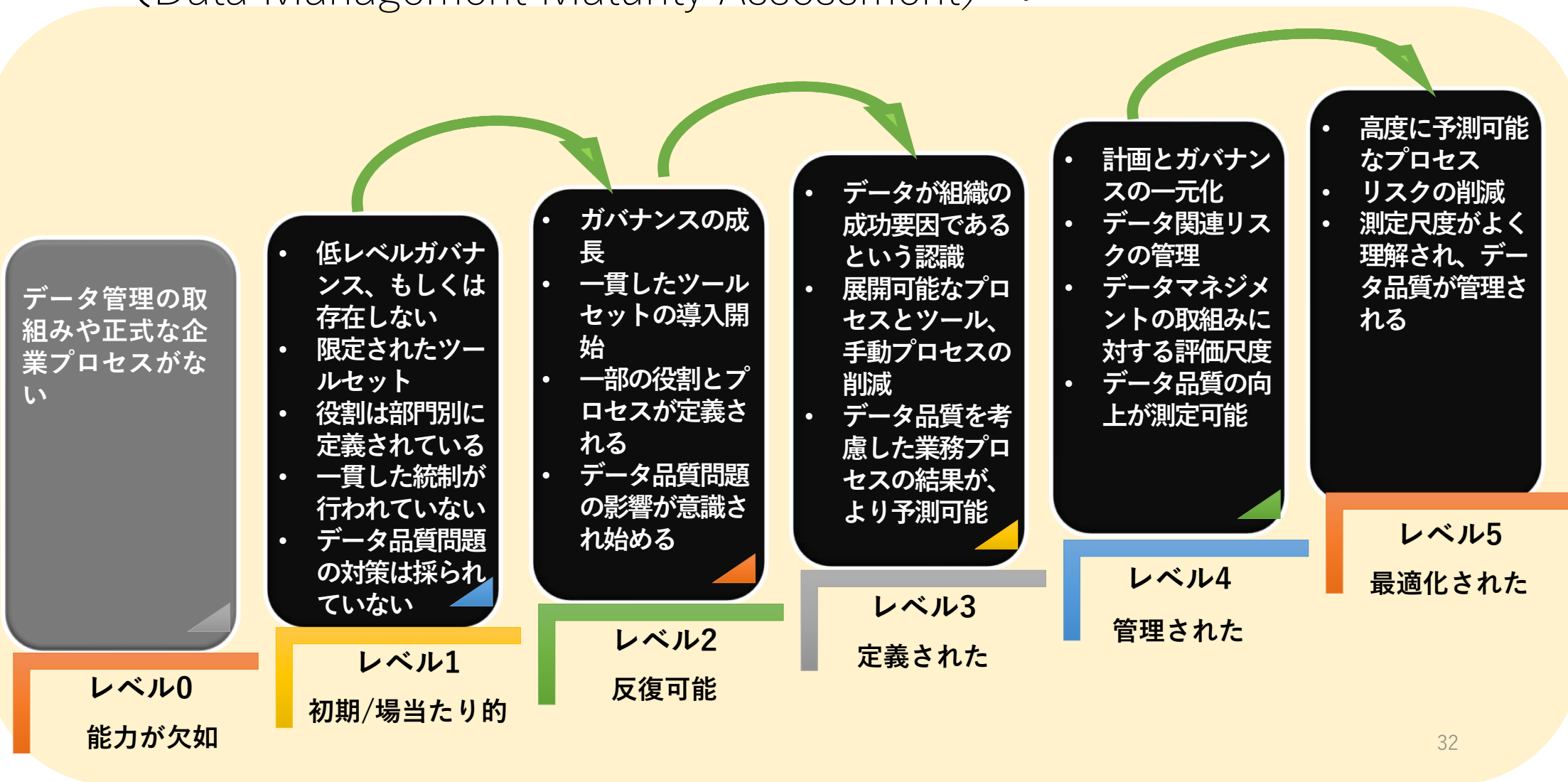
1. 組織を横断して重要なデータマネジメント・アクティビティを見つけ、総合的に評価する。
2. データマネジメントの概念、原則、実践についてステークホルダーに教育を行う。データの作成者から管理者までを含む広範囲な領域の中で、それぞれの役割と責任も明らかにする。
3. 企業全体で組織の業務や戦略的ゴールを支える継続可能なデータマネジメント・プログラムを確立して強化する。

## 4. データマネジメント成熟度アセスメント (Data Management Maturity Assessment) :

<評価基準>

レベル	社内/業務 関連実態
0	遂行能力が欠如している
1	初期段階か場当たりの：成功は個人の能力に依存している
2	反復可能:プロセスに最低限の統制が行われている
3	定義された：基準が設定され守られている
4	管理された：プロセスが数値化され統制されている
5	最適化された：プロセス改善のゴールが数値化されている

#### 4. データマネジメント成熟度アセスメント < 評価基準の特性 > (Data Management Maturity Assessment) :





## 4. データマネジメント成熟度アセスメント (Data Management Maturity Assessment) :

### <ビジネス上の意義>

- **規制** : 規制対応状況の監督のために最低限のデータマネジメント成熟度アセスメントを必要としている
- **データガバナンス** : データガバナンス機能は計画を立案し規制を遵守するために、成熟度アセスメントを必要としている
- **プロセス改善のための組織的準備** : 組織が業務を改善する必要性を認識し現状を評価することから始める。例えば、マスターデータマネジメントに取り組む場合にはMDMプロセスツールを導入する準備ができているかを評価する必要がある
- **チェンジマネジメント (組織と文化の変革)** : 合併などのチェンジマネジメントはデータマネジメントの課題を提起する。DMMAはこの課題に対応する計画立案のための情報を提供する。
- **新技術** : ITの進歩によりデータを管理し使用する新しい方法が出現する。組織は新技術をうまく採用できるかどうかを把握したい。
- **データマネジメントの問題** : データ品質の問題やその他のデータマネジメント関連課題に対処する必要がある。組織は現状レベルを基準に、そこからどのように変革を行うかについてより良い決定を下ろしたい。

(出所 : DAMA-DMBOK (Data Management Body of Knowledge) 第2版、2017)

実践：

**【課題】 自社または業務関連のデータマネジメント実態について評価してみてください。**

**(専用シートを確認してください。)**

## (参考) データマネジメントで解決が可能な課題の例示(1)

#	データ問題例	典型的な事例
1	データの内容面の食い違いの問題	データの意味が異なるため活用や連携ができない
2	データのタイミングにおける問題	鮮度が悪いため必要な時にデータが使えない
3	データ量の管理の問題	監視していないため夜間バッチが途中で落ちる事象が発生
4	データの一意性の問題	コード体系などが異なるため他システムのデータとつながらない
5	データの正当性の問題	変更の証跡が残らないシステムのつくり及び運用になっている
6	設計ドキュメントでデータの関連性が可視化されていない問題	運用が属人化している
7	有効なログデータが取得できない問題	データを取得して改善する意識や仕組みがない
8	システム運用における発生データ管理画面での問題	個別システムごとの管理になっている
9	調達仕様書作成段階でのユーザー不在の問題	利用に関する業務部門の関与が不足
10	「紙に印刷して確認する」という文化的問題	「原本は紙」発想で、データが使えなくなる
11	データの問題を解決する主体の問題	品質の維持・改善の責任者等が曖昧になりがち
12	簡単にデータ抽出できないシステム構造の問題	データと機能が非分離、高コスト体質に
13	データ取り扱い運用ルール未徹底の問題	必要なデータが登録されず、利用率が低調に
14	データのセキュリティ運用未徹底の問題	消去すべきデータが残存し、漏洩の温床に

## (練習課題) データマネジメントで解決が可能な課題の例示(2)

①データ問題例	#
データの内容面の食い違いの問題	1
データのタイミングにおける問題	2
データ量の管理の問題	3
データの一意性の問題	4
データの正当性の問題	5
設計ドキュメントでデータの関連性が可視化されていない問題	6
有効なログデータが取得できない問題	7
システム運用における発生データ管理画面での問題	8
調達仕様書作成段階でのユーザー不在の問題	9
「紙に印刷して確認する」という文化的問題	10
データの問題を解決する主体の問題	11
簡単にデータ抽出できないシステム構造の問題	12
データ取り扱い運用ルール未徹底の問題	13
データのセキュリティ運用未徹底の問題	14

### ②取り組みが必要な領域

「データマネジメント」領域

「制度」領域

「ITリテラシー・意識改革」領域

「システムのアーキテクチャ」領域

「システム運用・オペレーション」領域

### ③解決の方向性

データマネジメントの取り組みの実施により解決が可能

現行制度（規定類）やルールの見直し、条件緩和など

- 一般職員のITに関する基本的な知識やスキルの向上
- 手続きではなく結果（アウトプット）に対する意識教育など

抜本的な既存システムのアーキテクチャ見直し（構造の再構築、府省横断コード統一化など）など

連携システム全体としてのオペレーション運用の最適化、役割分担に明確化など