



#ds_symp2nd



DataScientist Society

データ社会に求められる 新しい才能とスキル

第2回シンポジウム

2015年11月13日

一般社団法人 データサイエンティスト協会

スキル委員会

設立時の状況

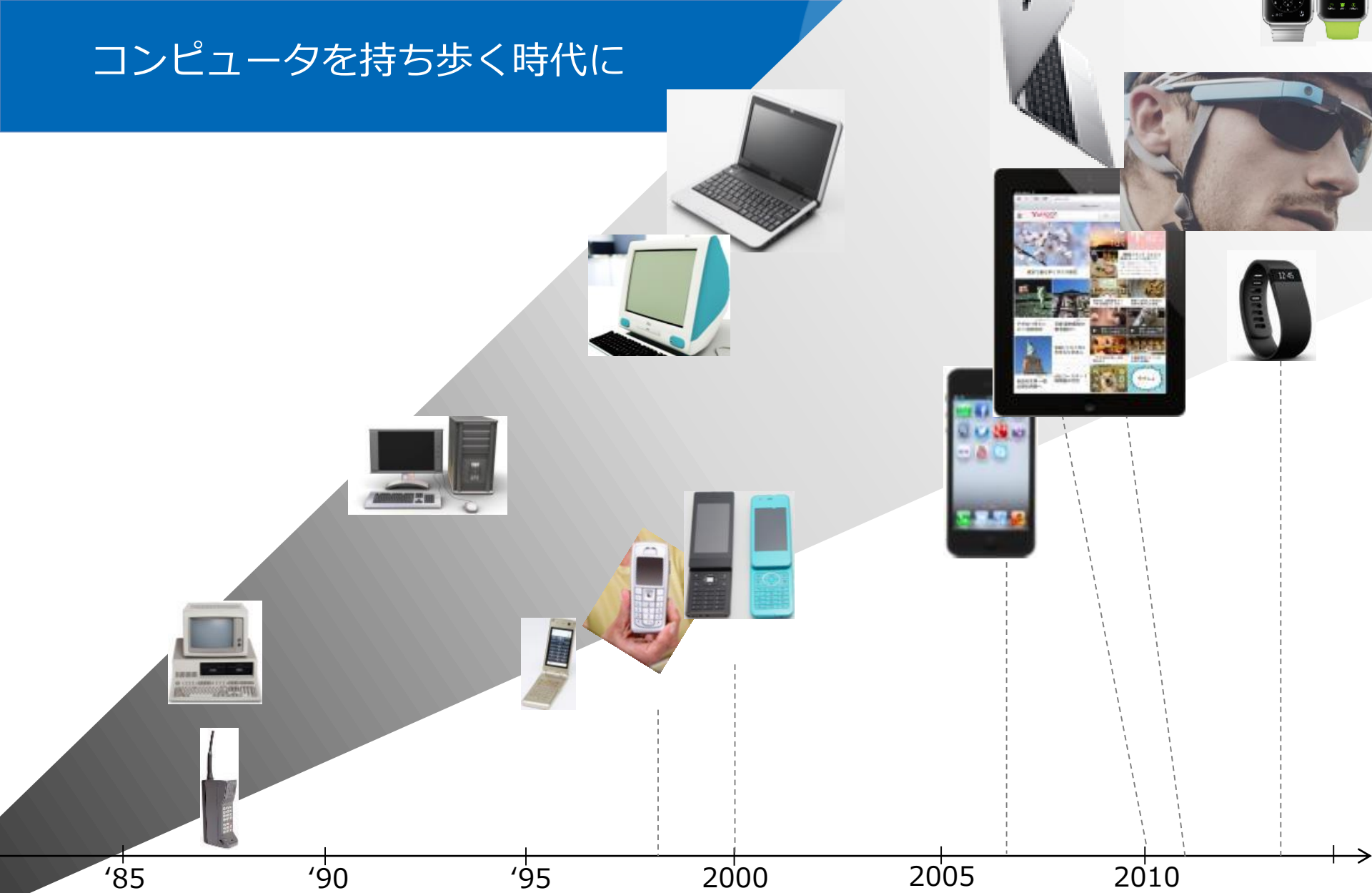
- バズワードである「データサイエンティスト」という言葉の定義が欠落
- 雇い主側の期待に雇われる側のスキルセットが合致しないケースが増加
- 若い才能たちが、自分達をどう訓練し、スキルを身に付けていくべきかわからない

設立の目的

- この新しいデータプロフェッショナル（「データサイエンティスト」）に必要とされるスキルセットを定義する
- スキル育成と評価のための軸・基準を作る
- このデータプロフェッショナルの皆さんが相互に接し、やりとりできるオープンな環境を提供する
- 社会との対話を行なう

データサイエンティスト とは何か？

コンピュータを持ち歩く時代に

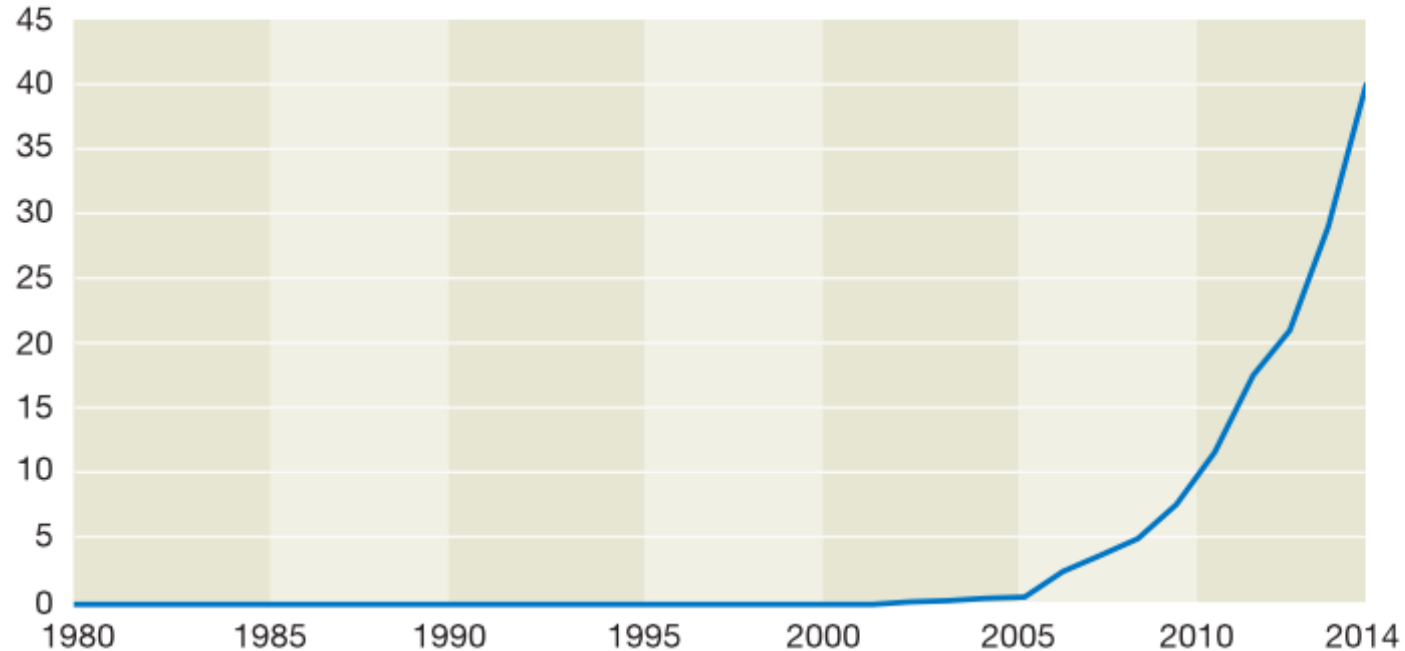


資料：安宅和人「データ時代に向けたビジネス課題とアカデミアに向けた期待」応用統計学セミナー2015.5.23
(<http://www.applstat.gr.jp/seminar/ataka.pdf>)

モノからも含め
多種多様なログデータが
今後、爆発的に発生
… ビッグデータ

計算キャパシティの爆増

Annual additions to global business and consumer computing power, exaflops¹

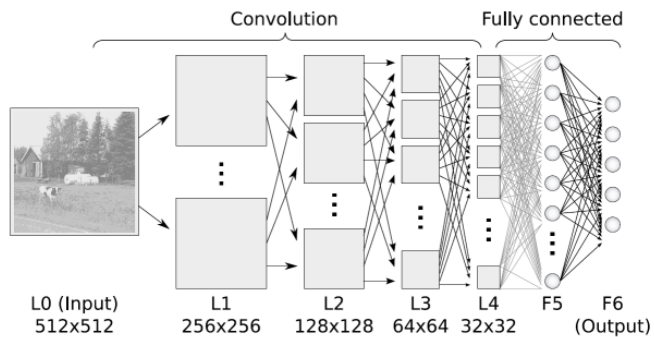


¹An exaflop is 1 quintillion (10 to the 18th power) floating-point operations per second.

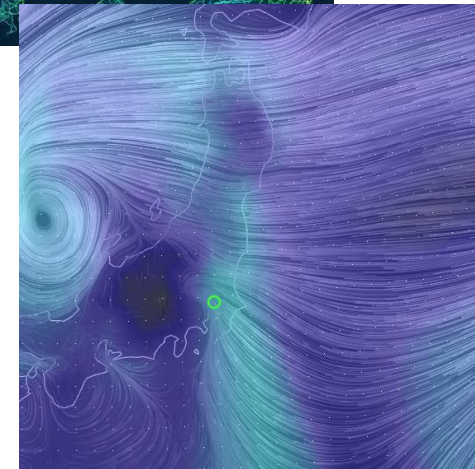
Source: William D. Nordhaus, "Two centuries of progress in computing," *Journal of Economic History*, 2007, Volume 67, Number 1, pp. 128–59; IDC; US Bureau of Economic Analysis; McKinsey analysis

資料 : McKinsey Quarterly 2014 Number3 p.16

大量データ処理技術 深層学習の実用化



可視化 (Data visualization) 技術の劇的な進展

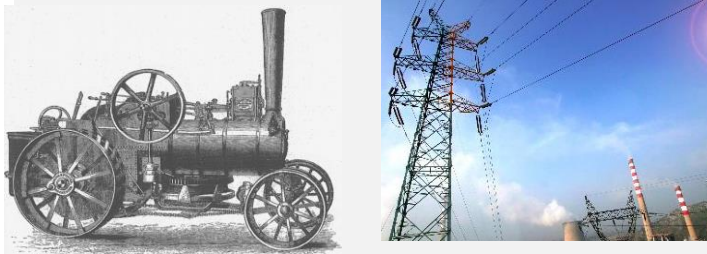


資料：東京大学 喜連川・豊田研HP、Universität Bonn, Institute for Computer Science HP、ヤフービッグデータレポートHP、Cameron Beccario「earth」WEBサイト

新しい
リソース

産業革命 (18~20世紀)

- 石炭と石油
- 蒸気機関
- 電気工学



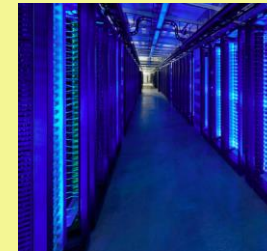
本質

- 人間と家畜を肉体労働、手作業から開放する



情報産業革命 (現在)

- ビッグデータ
- 高い計算能力
- 情報科学の進化

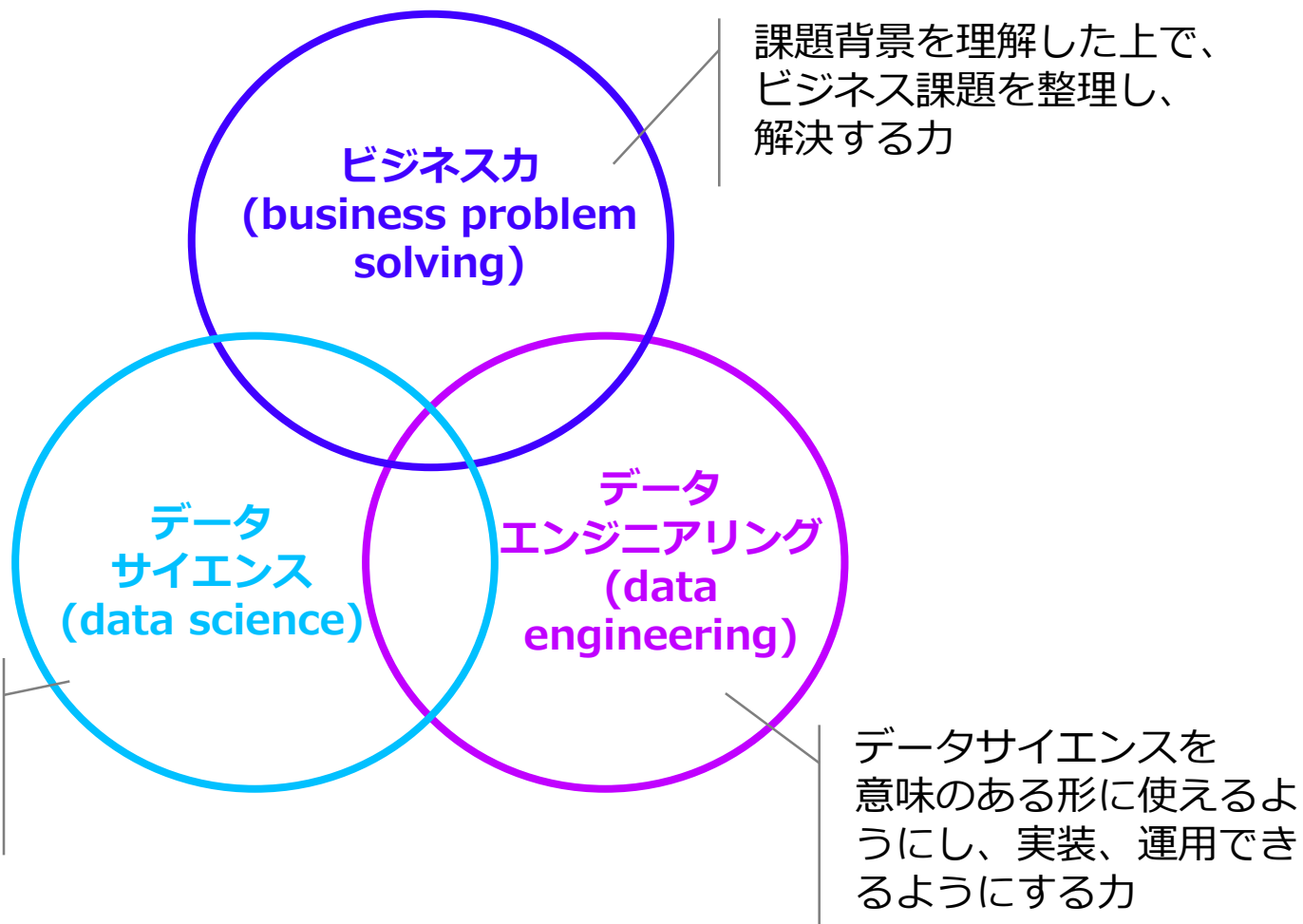


- 人間を退屈な数字入力、情報処理作業から開放する

資料：データサイエンティスト協会スキル委員会ディスカッション:安宅和人「データ時代に向けたビジネス課題とアカデミアに向けた期待」
応用統計学セミナー2015.5.23 (<http://www.applstat.gr.jp/seminar/ataka.pdf>)

「データの持つ力を解き放つ」

3つのスキルセット



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>

データサイエンス力、
データエンジニアリング力をベースに
データから価値を創出し、
ビジネス課題に答えを出す
プロフェッショナル

* ここで「ビジネス」とは社会に役に立つ意味のある活動全般を指す

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>

単なるスペシャリストでもエキスパートでもない

スペシャリスト

- 何らかの分野に特化した人

エキスパート

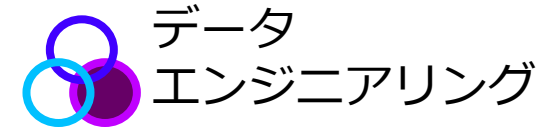
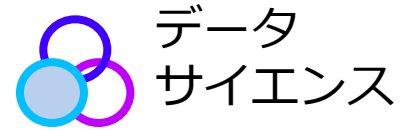
- 何らかの分野について体系的で秀でた知識とスキルを持っている人

プロフェッショナル

- 体系的にトレーニングされた専門性を持つスキルを持ち、
- それをベースに顧客（お客様、クライアント）にコミットした価値を提供し、
- その結果に対し、認識された価値の対価として報酬を得る人

資料：データサイエンティスト協会スキル委員会ディスカッション

課題解決のフェーズによって核となるスキルが変わる



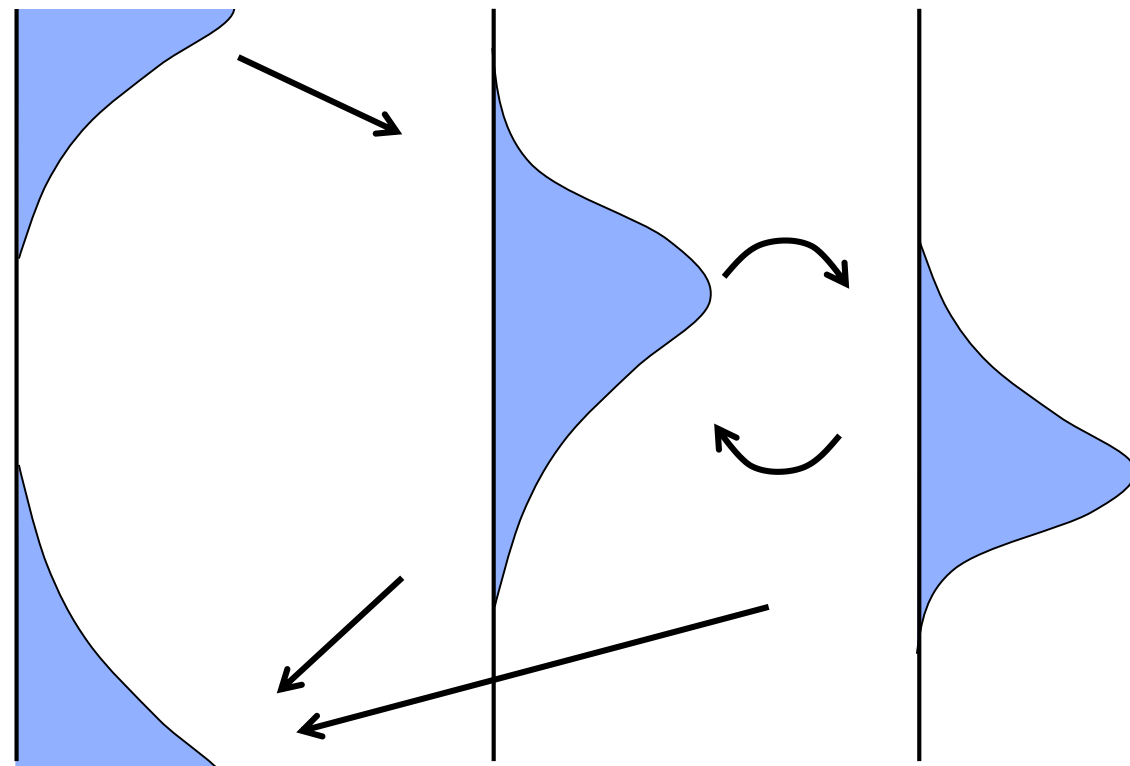
目的・テーマ設定

問題定義

アプローチの設計

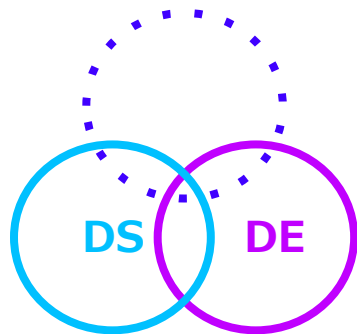
処理・分析

解決



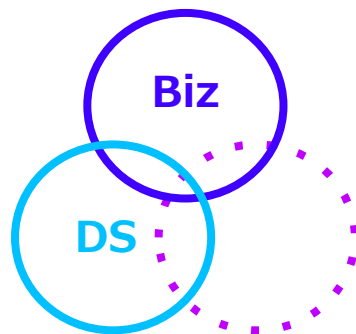
資料 : データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

どの一つが欠けてもダメ、、、優位はない



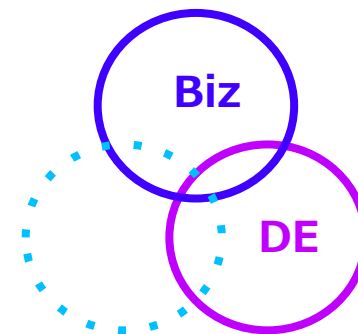
データスペシャリスト
ではあるが、そもそも
解決すべき問題が定義、
整理できない

> プロフェッショナル
ではない



ビジネス課題もわかり
そこでのサイエンスの
利活用もわかるが実装
できない

> 必要な変化を起こせ
ない



ビジネス課題の上で、
実装を用意できるが、
かなめとなるサイエン
スの知恵が足りない

> 賢いやり方を提供で
きない

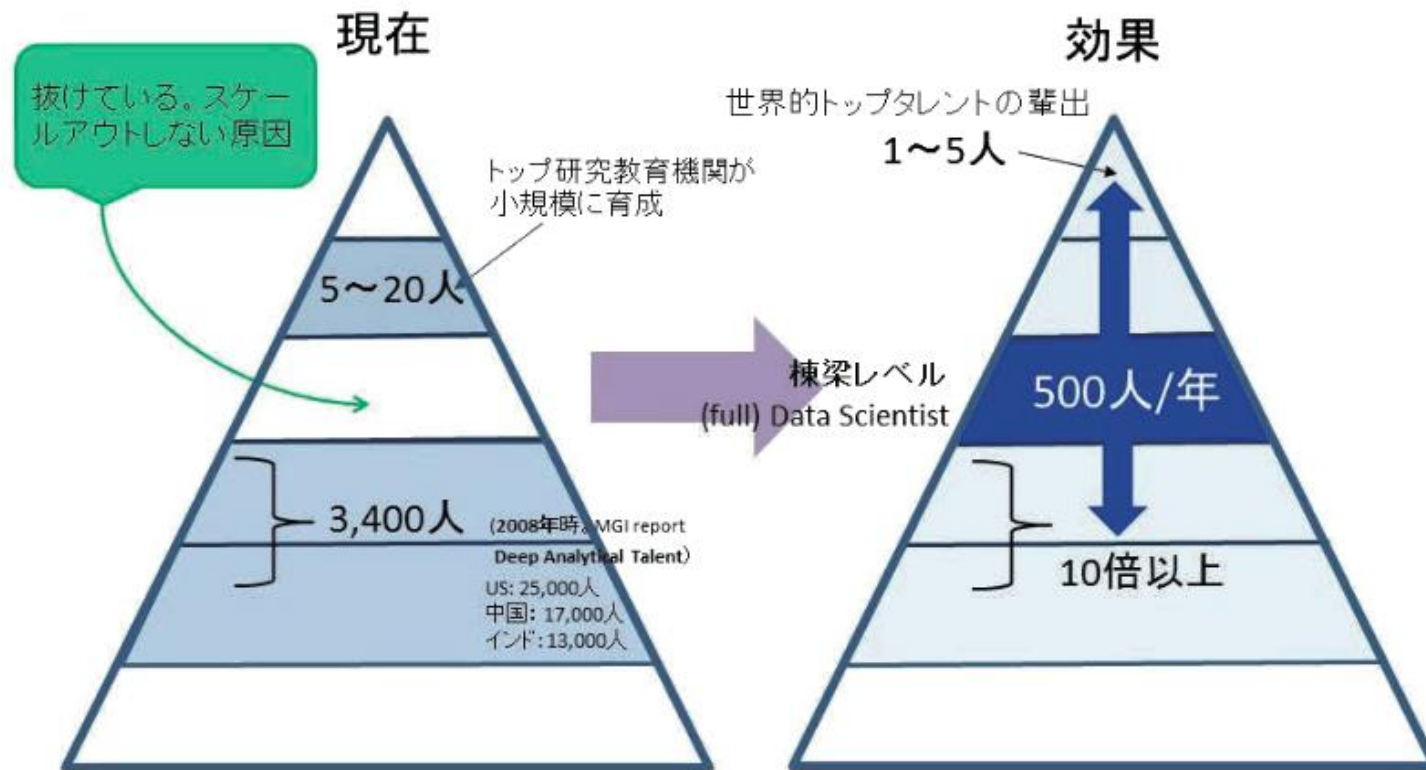
スキルのレベル感は？

スキルレベル

	スキルレベル	目安	対応できる課題
データサイエンティスト	シニア・データサイエンティスト	<ul style="list-style-type: none">業界を代表するレベル	<ul style="list-style-type: none">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	<ul style="list-style-type: none">棟梁レベル	<ul style="list-style-type: none">対象組織全体
	アソシエート・データサイエンティスト	<ul style="list-style-type: none">独り立ちレベル	<ul style="list-style-type: none">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	<ul style="list-style-type: none">見習いレベル	<ul style="list-style-type: none">プロジェクトの担当テーマ
一般人	データ使い	<ul style="list-style-type: none">賢くデータを器用にする人	<ul style="list-style-type: none">担当業務
	普通の人	<ul style="list-style-type: none">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>

とりわけ棟梁レベルが足りていない



資料：情報・システム研究機構 『ビッグデータの利活用のための専門人材育成について』（ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会）2015.7.15

1. 昨年のrecap

- ① 本協会設立の背景と目的
- ② データサイエンティストとは何か？
- ③ 3つのスキルとスキルレベル

2. スキルチェックリスト

- ① 全体として（昨年 vs 今年）
- ② 個別領域の広がり
- ③ レベル感の考え方

3. 実行に向けて

- ① チームをどう作るか？
- ② 人をどう育てるか？
- ③ 目指す人はどう着手すべきか？

気づき

ポイント

データ
サイエンス
力

- 情報科学も情報系、機械学習系、データ可視化系いずれかだけでは足りない

- 総合的な解決アプローチにつなげる力がより必要。分野別の専門性はnice to haveにすぎない

データ
エンジニア
リング力

- いわゆるプログラマー、Sier的なエンジニアの通常技術だけでは足りない

- 既存のエンジニアリング業務を超え、ビッグデータ処理を実現できる必要がある

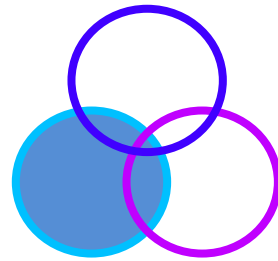
ビジネス力

- 問題解決力は単なる領域での知見やコンサル的なものを超える必要がある

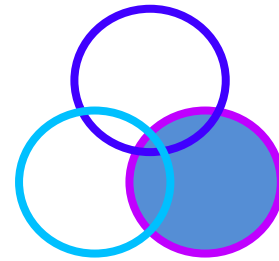
- 他のスキル領域と話せる力を含めたデータ視点での課題解決力を持つ必要がある

スキル表

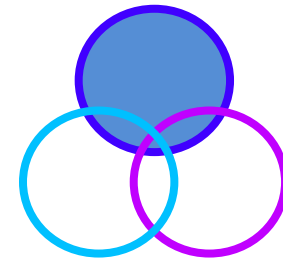
(項目数)



データ
サイエンス



データ
エンジニアリング



ビジネス

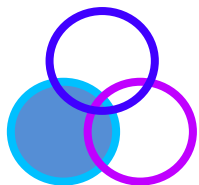
レベル
計

★★★★	棟梁レベル (full)	58	37	44	(139)
★★★	一人前レベル (Associate)	68	43	49	(160)
★	見習いレベル (Assistant)	54	39	30	(123)
	領域計	180	119	123	(422)

- データセキュリティはデータエンジニアリングとビジネスそれぞれの視点から扱っています
- 以下の項目は本チェックリストではカバーしていません
 - ✓ 特定のスキル項目ができると当然できると考えられるスキル
 - ✓ 特定の業界で重要であっても、一般性が低いスキル
 - ✓ 特定作業ができるために必要になるSAS、SPSSなど具体的ツールを使うスキル*
- 20名弱の実務家の協議による検討結果 (ver.1) であり、必要に応じ、適宜、見直していく位置づけです
- 個別の項目ができるできないの判断は、いまのところ各自のご判断で行っていただく前提です (後述)

* ただし具体名を入れないと理解が困難なケースでは固有名詞を入れてあります

領域の広がり (データサイエンス)



解析技術

探索的

サンプリング

データ加工

検定/判断

Static

機械学習

グルーピング

性質・関係性の把握

パターン発見

グラフィカルモデル

Dynamic

予測

時系列分析

シミュレーション
/データ同化

Data Visualization

非構造化
データ処理

言語処理

画像処理

音声処理

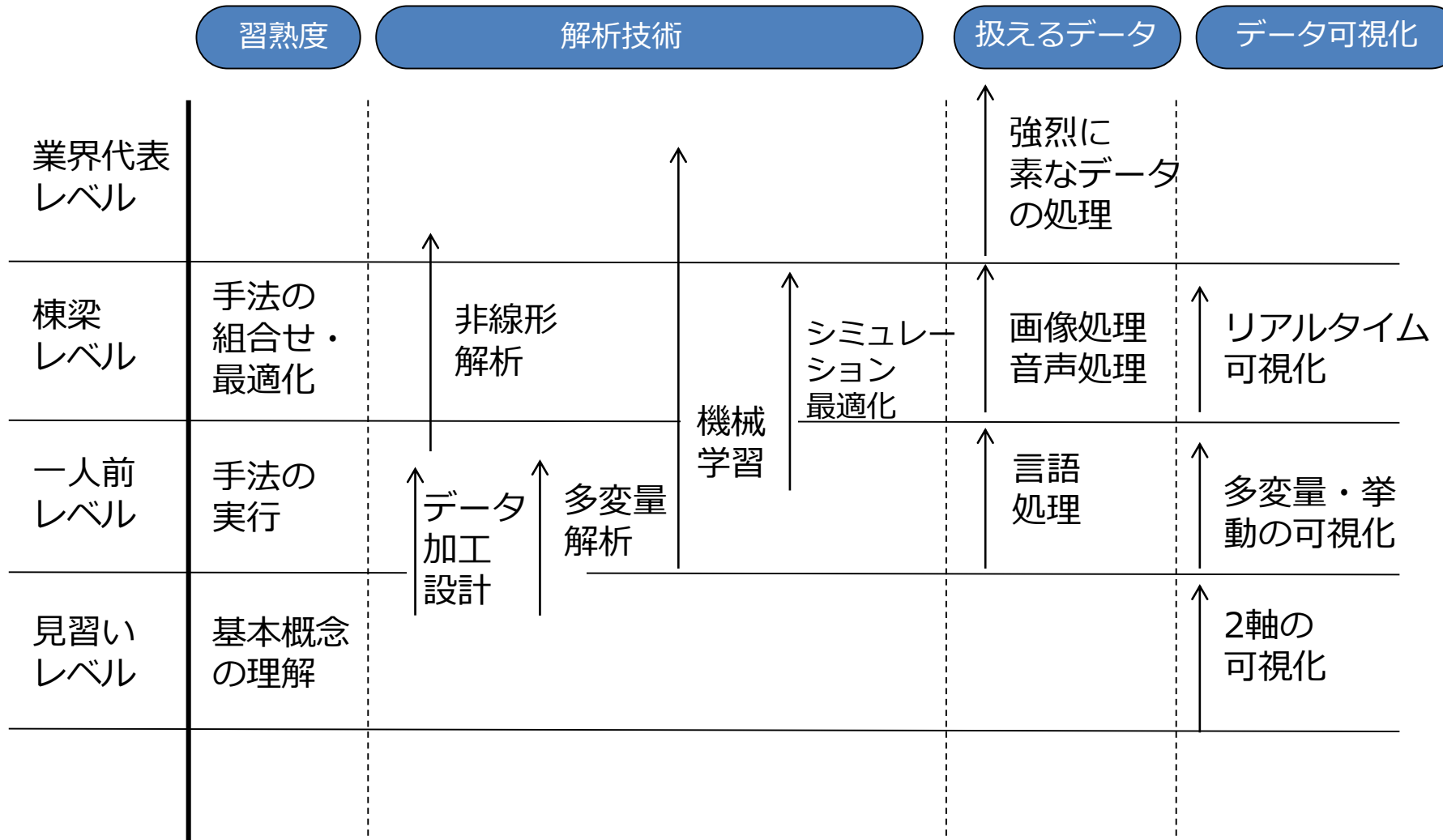
基礎技術

統計数理基礎

統計数理応用

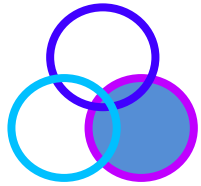
レベル感の目安 (データサイエンス)

2015.11.13 現在



資料：データサイエンティスト協会スキル委員会討議

領域の広がり (データエンジニアリング)



守る技術

ITセキュリティ

実装技術

データ
収集

データ
構造

データ
蓄積

データ
加工

データ
共有

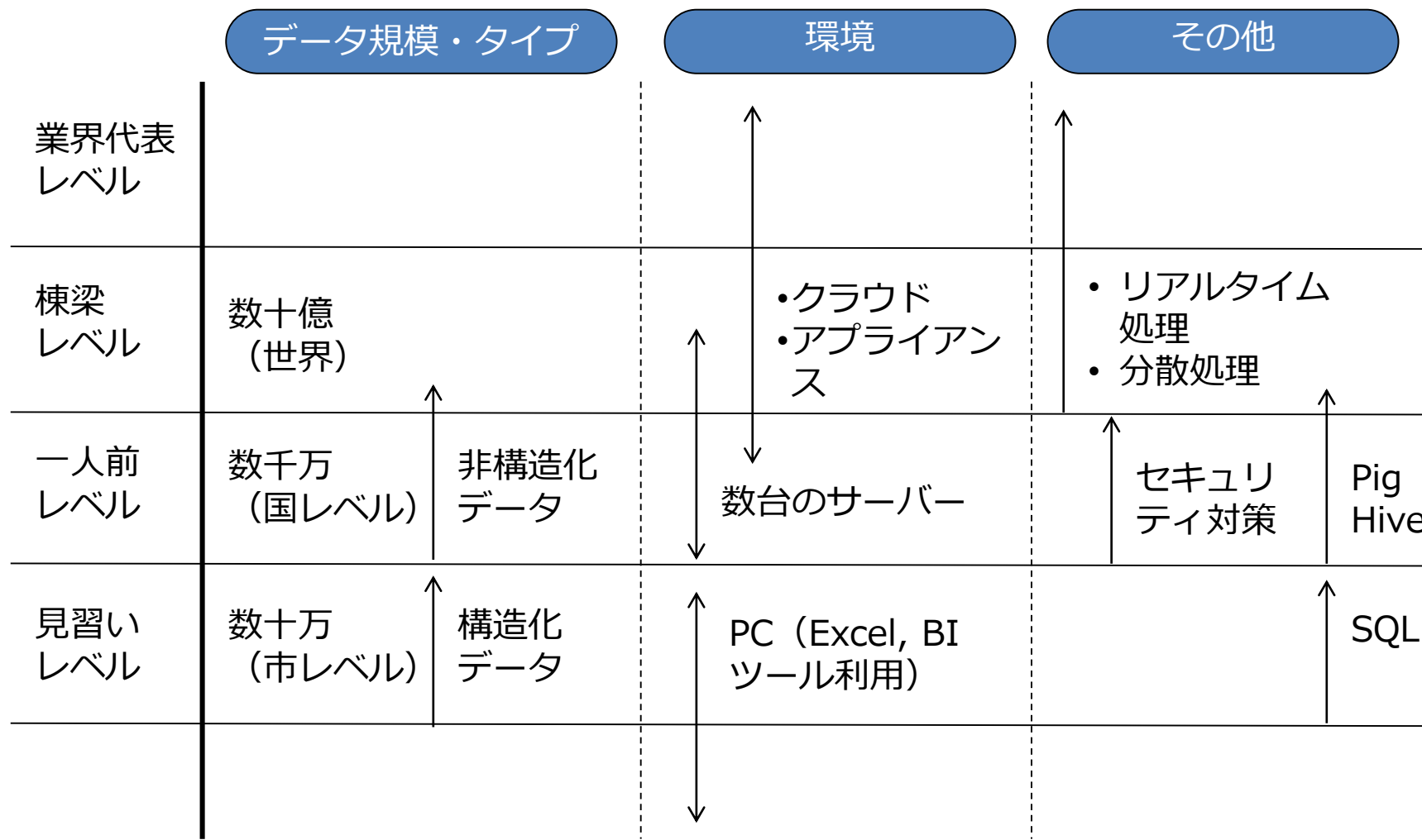
環境構築

基礎技術

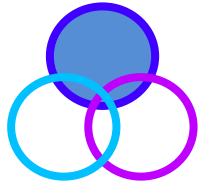
プログラミング

レベル感の目安 (データエンジニアリング)

2015.11.13 現在



資料：データサイエンティスト協会スキル委員会討議



ビジネス
課題解決

活動マネジメント

事業に実装する

データ
課題解決

データ
入手

データの
理解・検証

意味合いの
抽出・洞察

プロセス

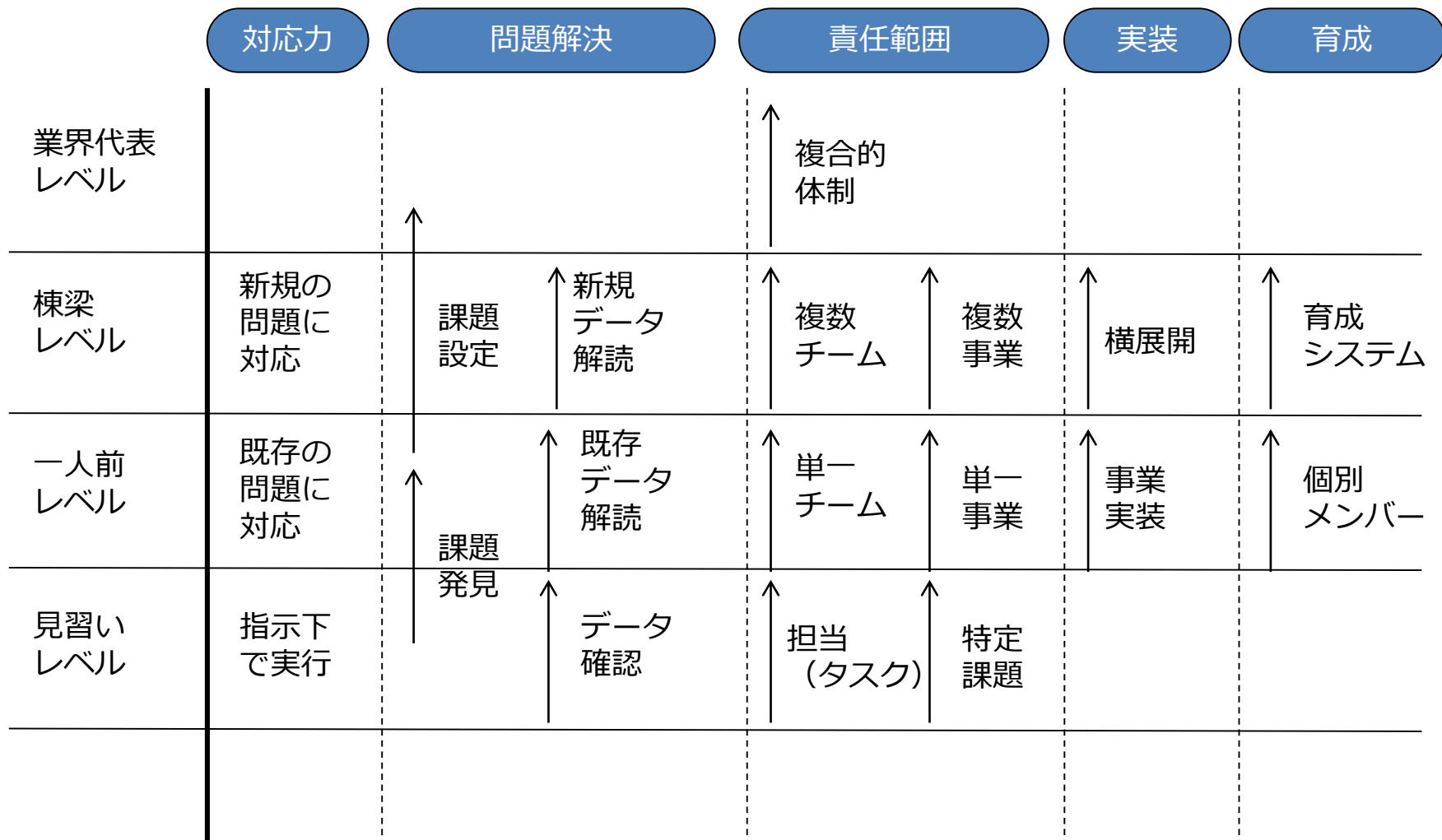
基礎能力

論理的思考

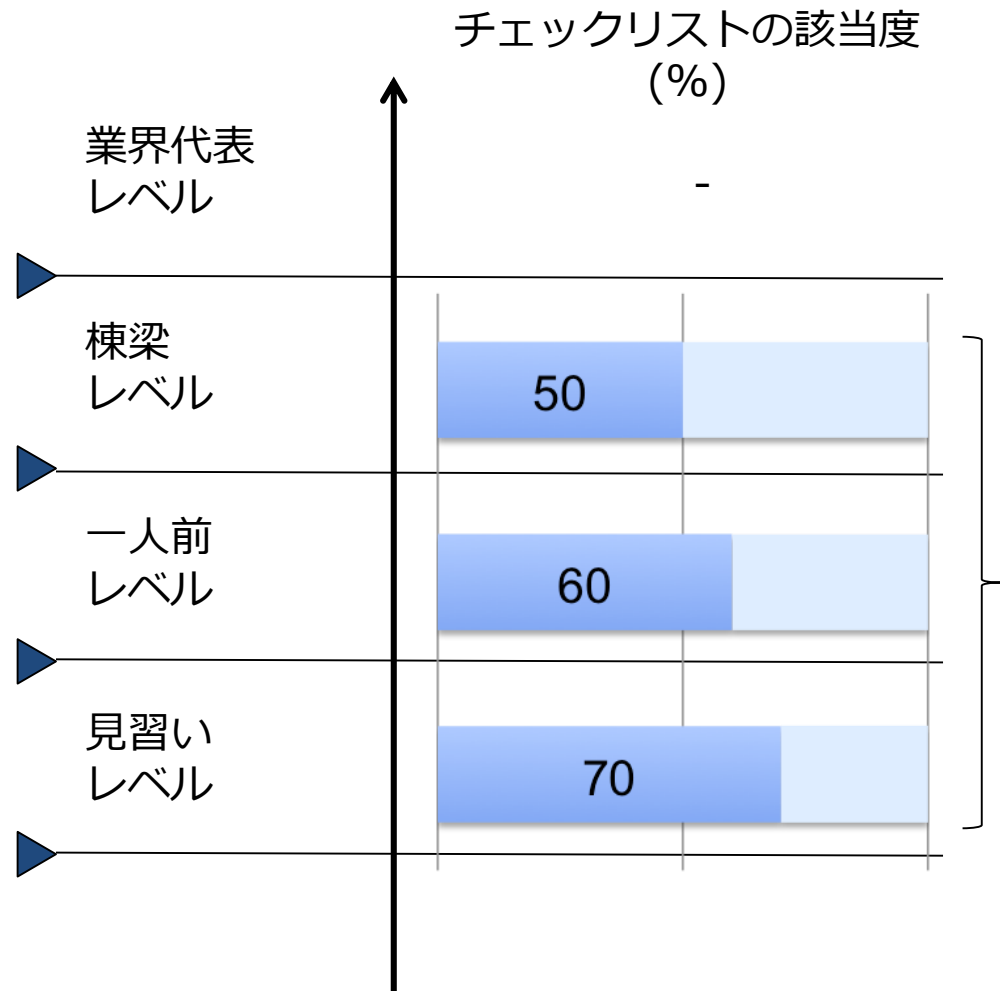
行動規範

レベル感の目安（ビジネス）

2015.11.13 現在



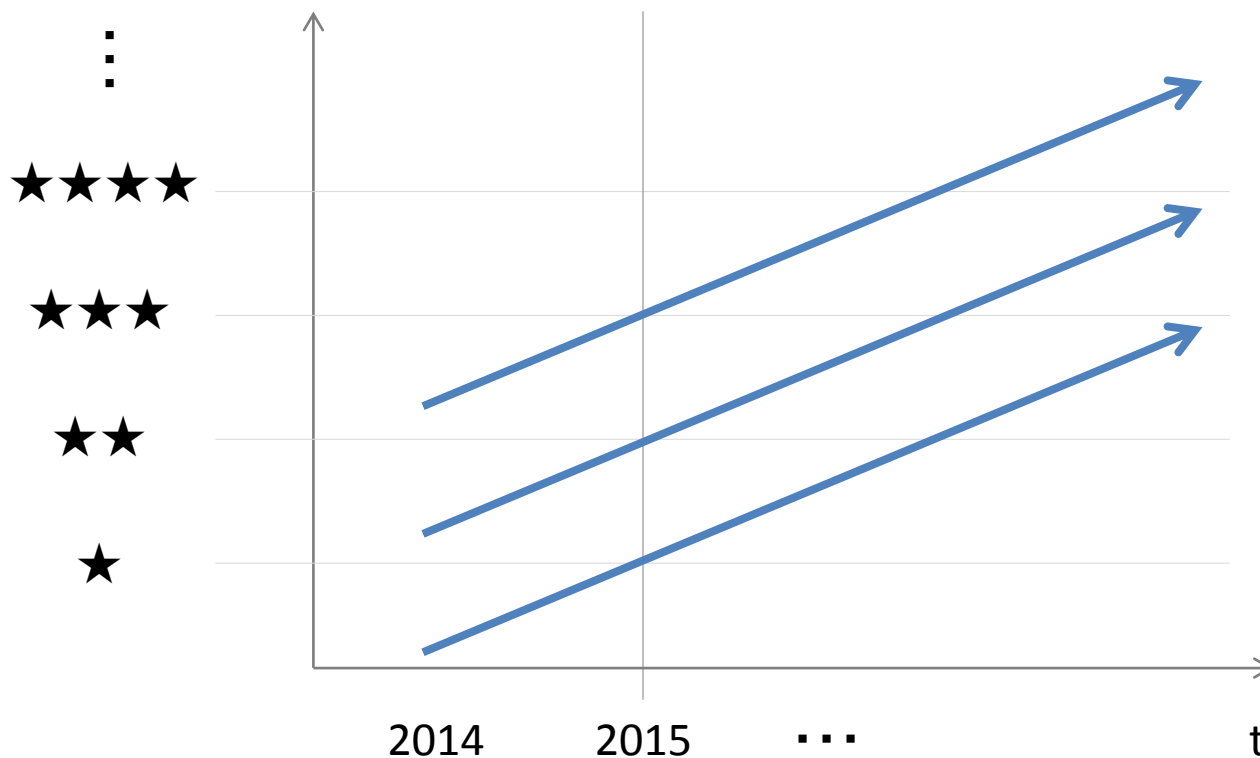
資料：データサイエンティスト協会スキル委員会討議



- 3つの領域別に判定
- 各レベルの必須項目を一つでも満たさないとアウト
- 下のレベルは満たしていることが前提

スキルは本質的にmoving target

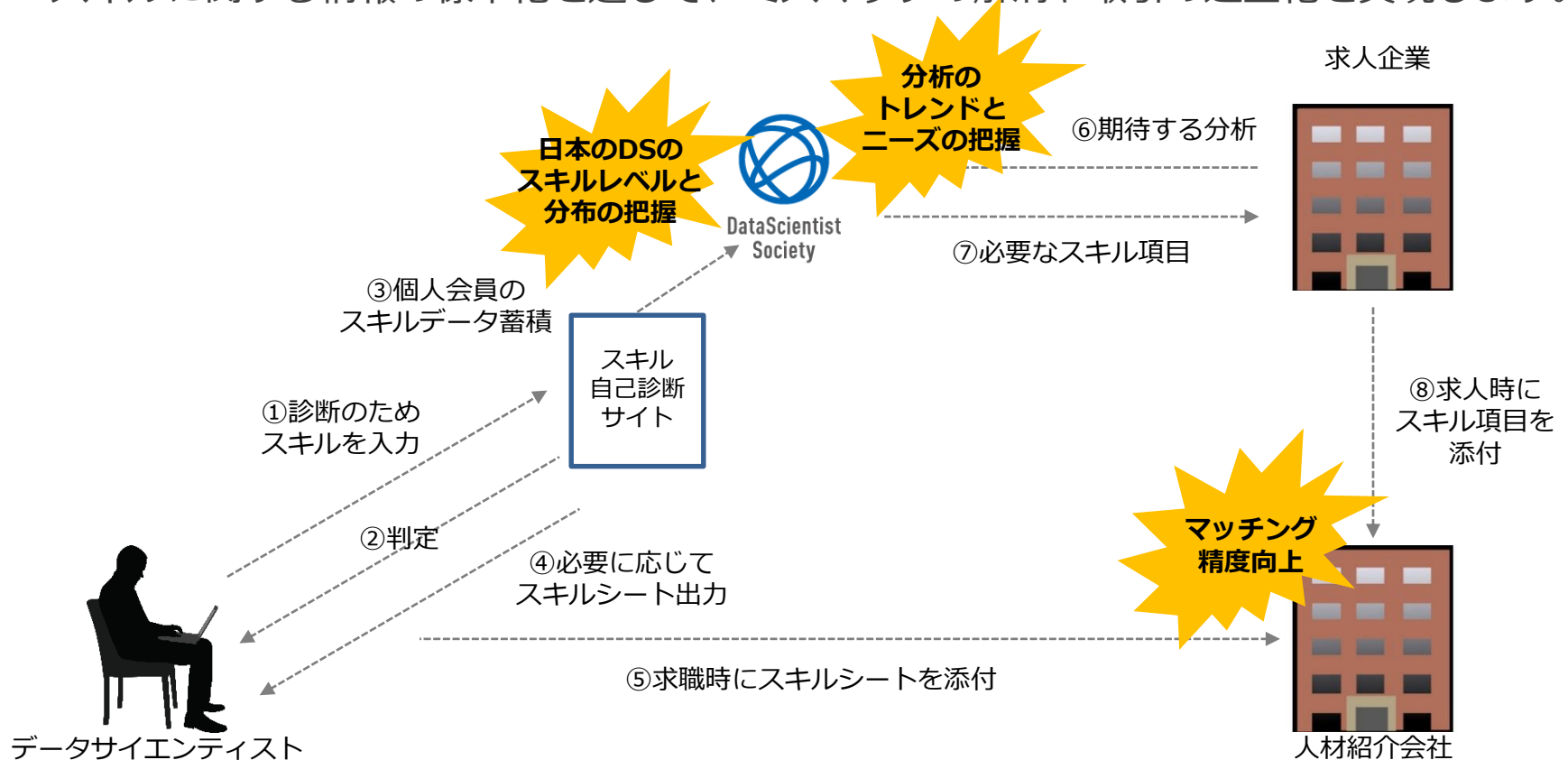
CONCEPTUAL



資料：データサイエンティスト協会スキル委員会討議

スキルチェックの仕組み（イメージ）

- 2016年度にデータサイエンティスト自身がスキルのチェックをサイト上で実施できる自己診断の仕組みを公開予定。
人材紹介業界でのスタンダードを目指し、データサイエンティストや企業間でのスキルに関する情報の標準化を通じて、ミスマッチの解消や取引の適正化を実現します。



1. 昨年のrecap

- ① 本協会設立の背景と目的
- ② データサイエンティストとは何か？
- ③ 3つのスキルとスキルレベル

2. スキルチェックリスト

- ① 全体として（昨年 vs 今年）
- ② 個別領域の広がり
- ③ レベル感の考え方

3. 実行に向けて

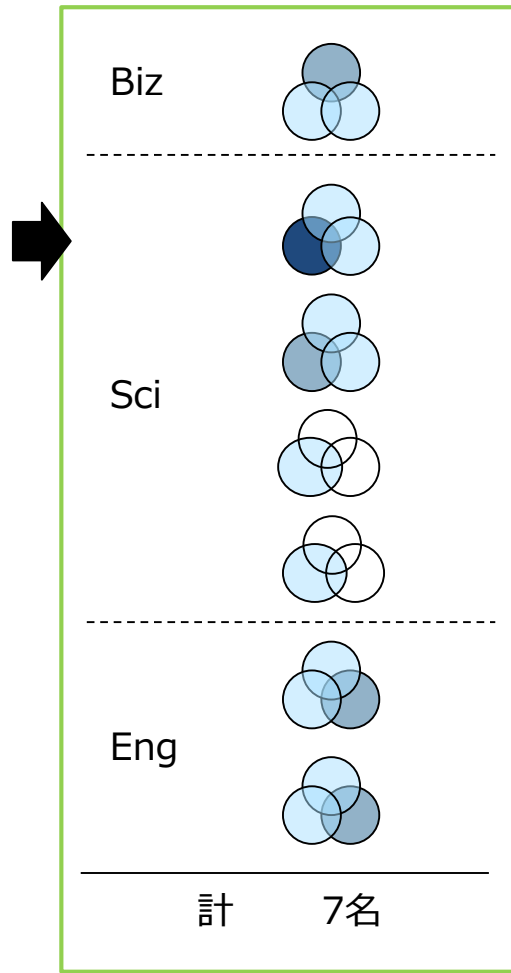
- ① チームをどう作るか？
- ② 人をどう育てるか？
- ③ 目指す人はどう着手すべきか？

- 具体的にどう回すか？
- 具体的にどうやって育成していくべきか？
- 目指す人は何をどうしていったらいいのか？

実際のチームは
どう組んだらいいのか？

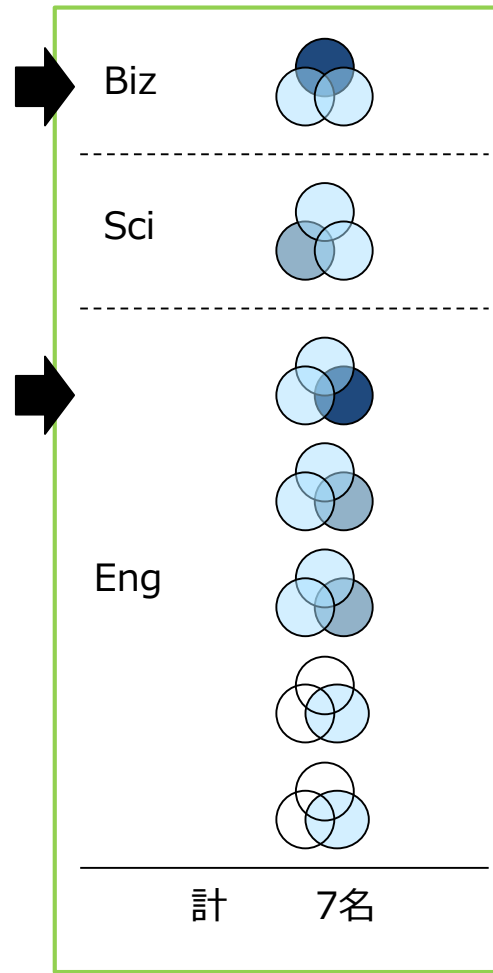
チーム編成事例 (1/3) : ウェブサービス提供

例1 : レコメンド (Sci寄り)

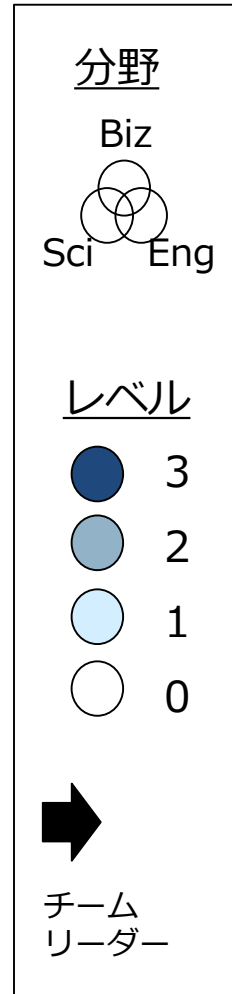


3ヶ月(改善)~12ヶ月(新規)

例2 : アクセス解析 (Eng寄り)

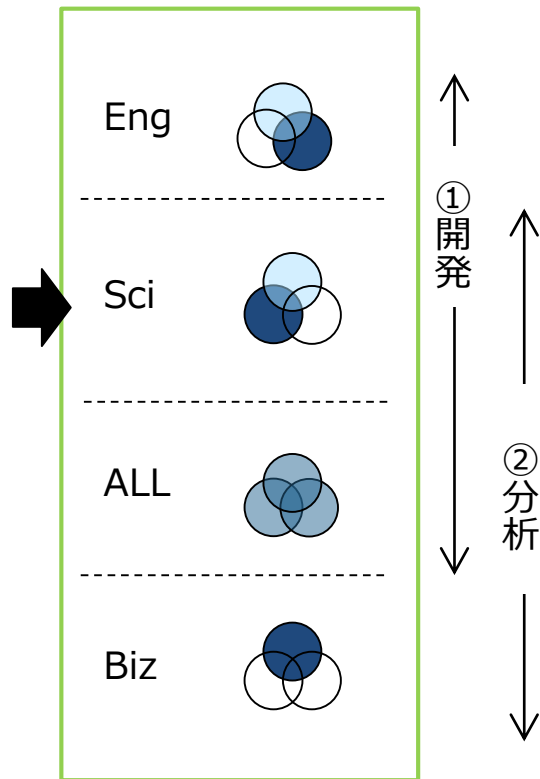


1ヶ月(改善)~6ヶ月(新規)

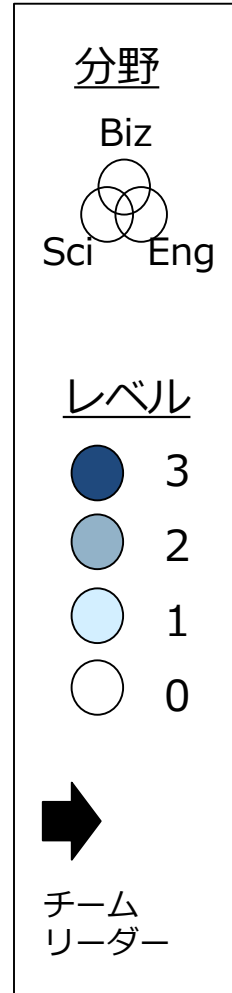
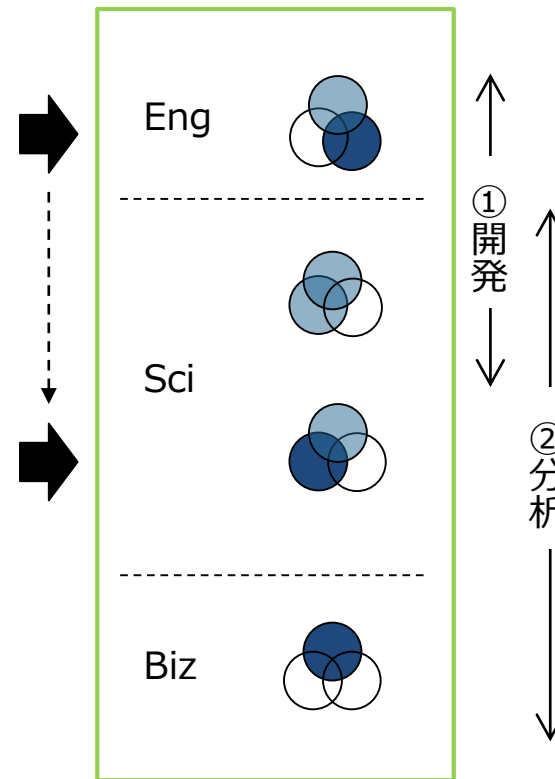


チーム編成事例(2/3) : マーケティングソリューション

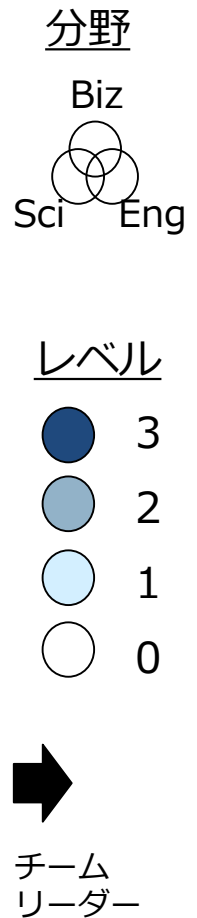
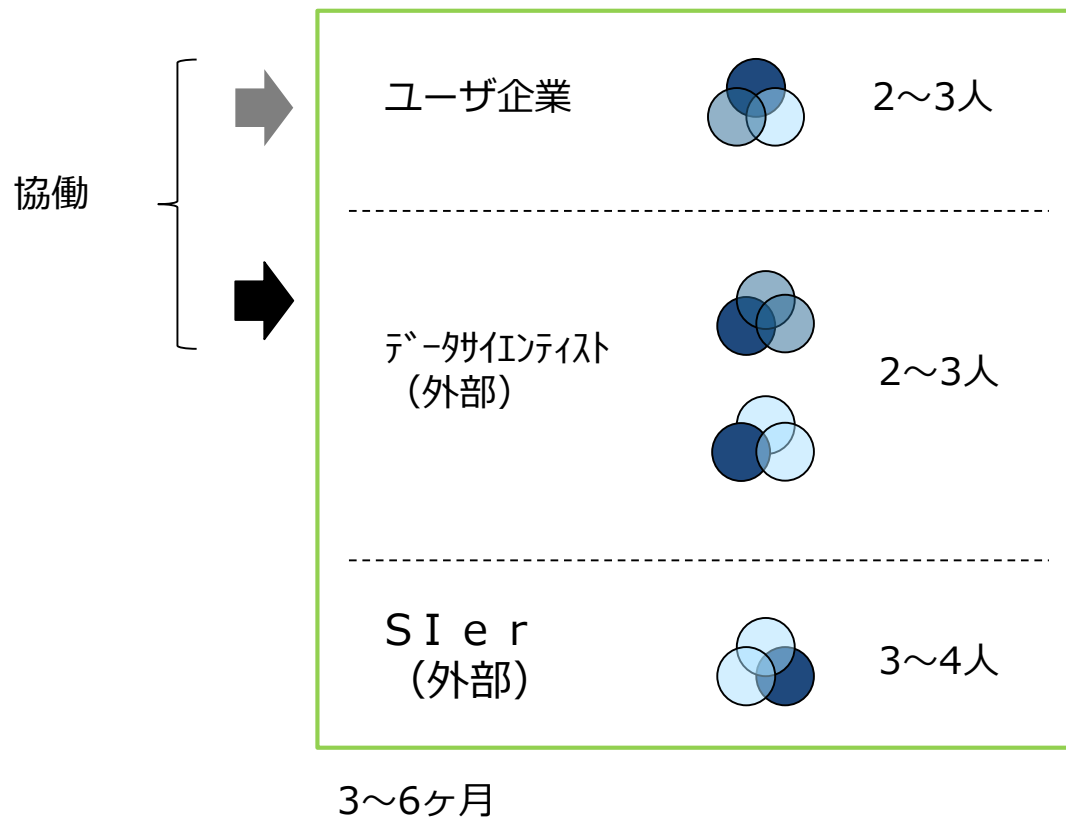
例3 : 行動データ解析



例4 : マーケティング投資最適化



例5：マーケティングオートメーション導入

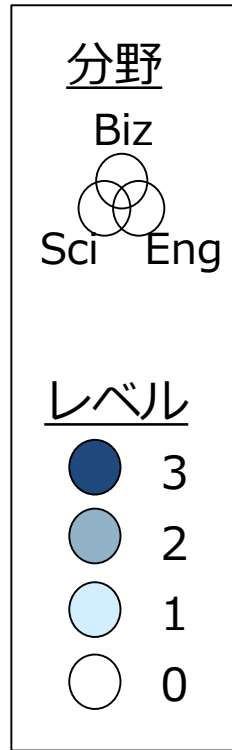
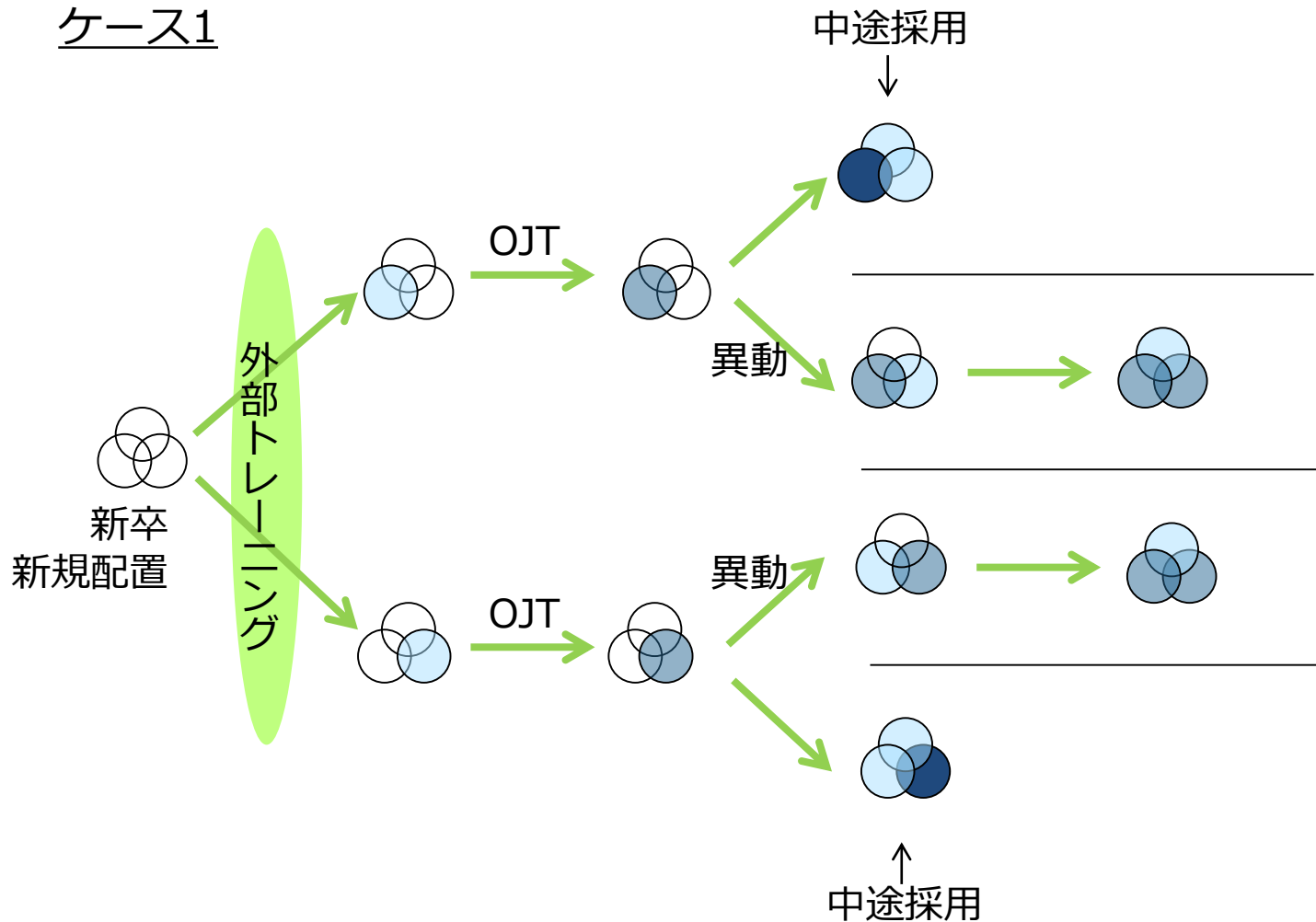


- 単独の一人ではなく、チームとして必要なスキルセットを実現するのが現実解
- 課題解決の重さがどこにあるかで組成も変化
(サイエンスより、エンジニアよりなど)
- 開発、運用などのフェーズによっても異なる
- チームリーダーはコアになる課題解決と俯瞰できる力の有無できまる

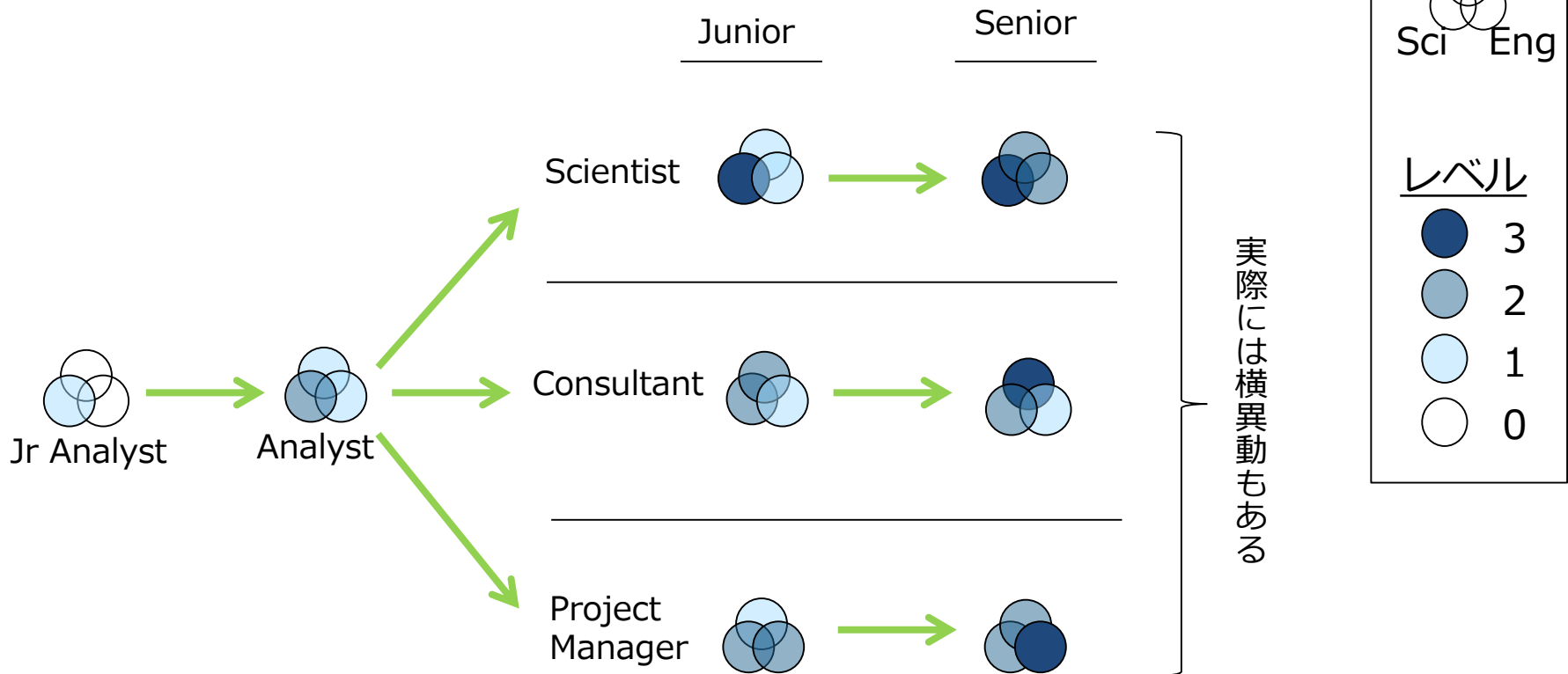
- そもそもこのような体制を作れない会社はどうしたらいいのか？
- どういうヒトの組み合わせ、体制にしたらいいのかわからない会社はどうしたらいいのか？

ではスキル育成は
どうしたらいいのか？

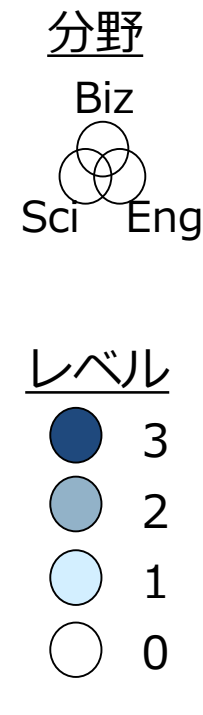
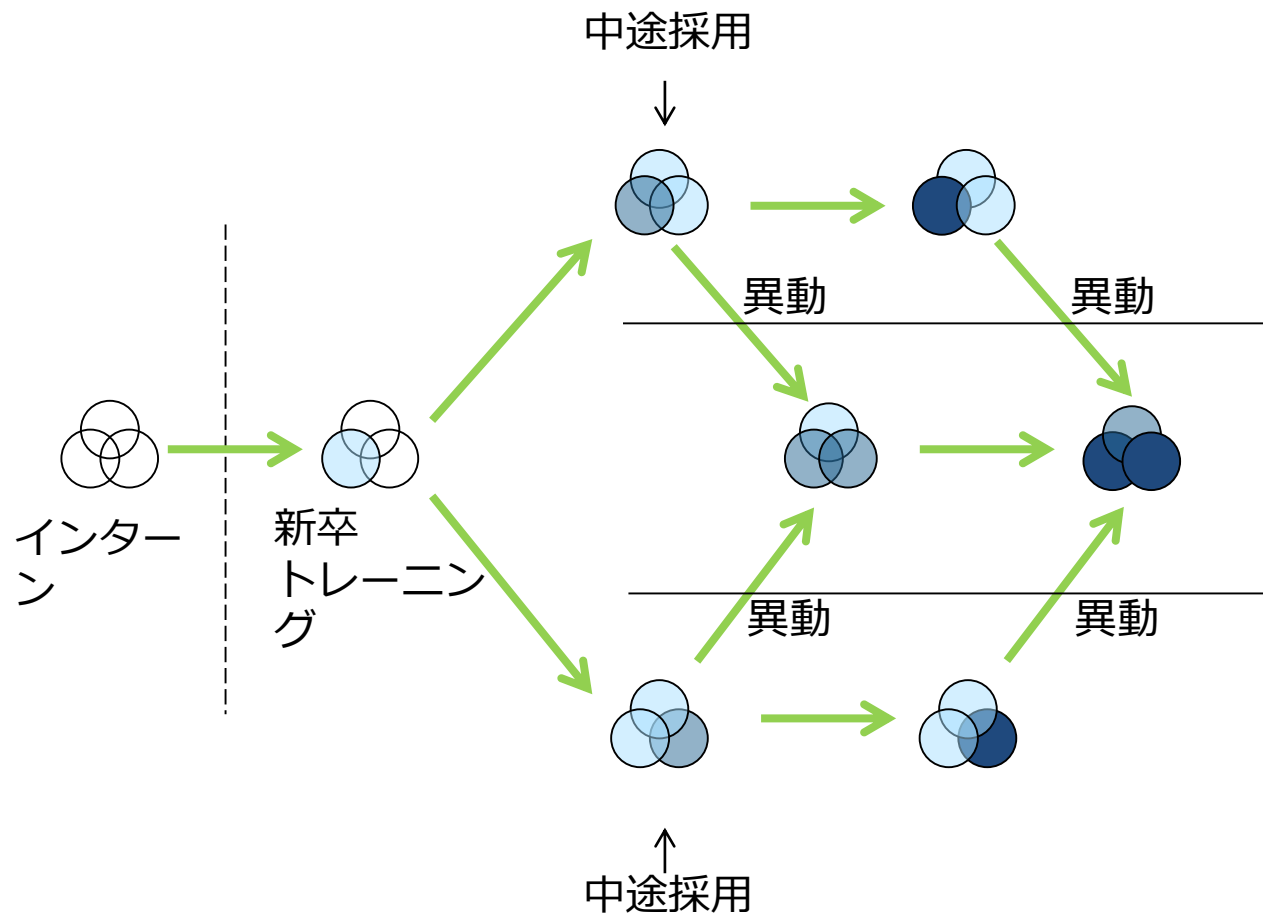
ケース1



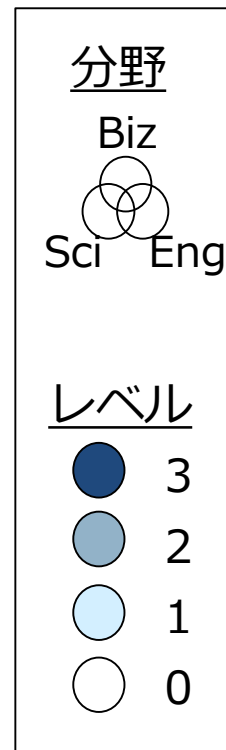
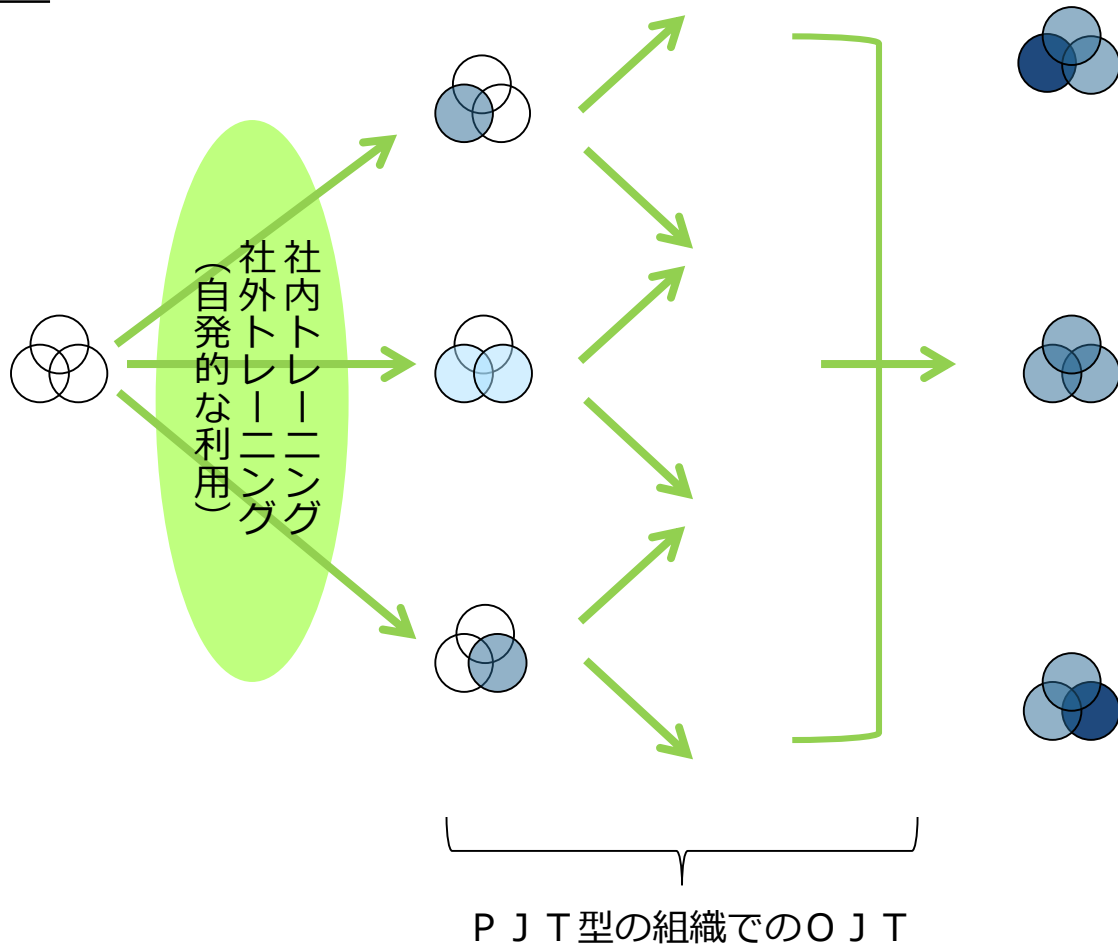
ケース2



ケース3



ケース4



- 見習いレベルまでは共通的なトレーニングで、そこから先は実践を通じてが基本
- サイエンス、エンジニア両方のスキルを同時というより、プロジェクトや部門異動によって育成していくのが現実解
- ビジネス力は複数の領域を超えて仕事を始め、全体を俯瞰してマネージする中で育成するケースが多い

- 特定の領域の職人的な環境にいるひと（特にエンジニア上がりの人）はどうやってデータサイエンティストになるのか？
- より高度化していくスキルにどのように対応していくべきなのか？
- 異動を自分でコントロールしようがない時にどのように新しいスキルを身につけていくべきなのか？
- 高等教育機関（特に大学）はどう活用したら良いのか？

「データの持つ力を解き放つ」