



DataScientist Society

8thシンポジウム

2021年度スキル定義委員会活動報告 2021年度版スキルチェック&タスクリスト公開

2021年11月16日

一般社団法人 データサイエンティスト協会 スキル定義委員会

パネリスト紹介

- | | |
|-------|--|
| 安宅 和人 | 慶應義塾大学SFC教授 / ヤフー株式会社CSO |
| 孝忠 大輔 | 日本電気株式会社 AI・アナリティクス事業部 事業部長代理 |
| 菅 由紀子 | 株式会社Rejoui 代表取締役社長 |
| 高橋 範光 | 株式会社デジタルグローースアカデミア 代表取締役社長 |
| 森谷 和弘 | データ設計事務所 代表 / データアナリティクスラボ株式会社CTO |
| 佐伯 諭 | 一般社団法人データサイエンティスト協会 スキル定義委員会副委員長 他 スキル定義委員会メンバー |

アジェンダ

1. スキル定義委員会の活動
 - スキル定義委員会とは？
 - 主な活動紹介
2. スキルチェックリスト ver.4 公開（本日公開）
3. 2021年度版データサイエンス業務のタスクリストアップデート
（IPAとの協働作業）
4. 第1回データサイエンティスト検定 実施報告
5. ディスカッション

1. スキル定義委員会の活動

データサイエンティスト協会
スキル定義委員会とは？

そもそもの「データサイエンティスト協会設立の目的」

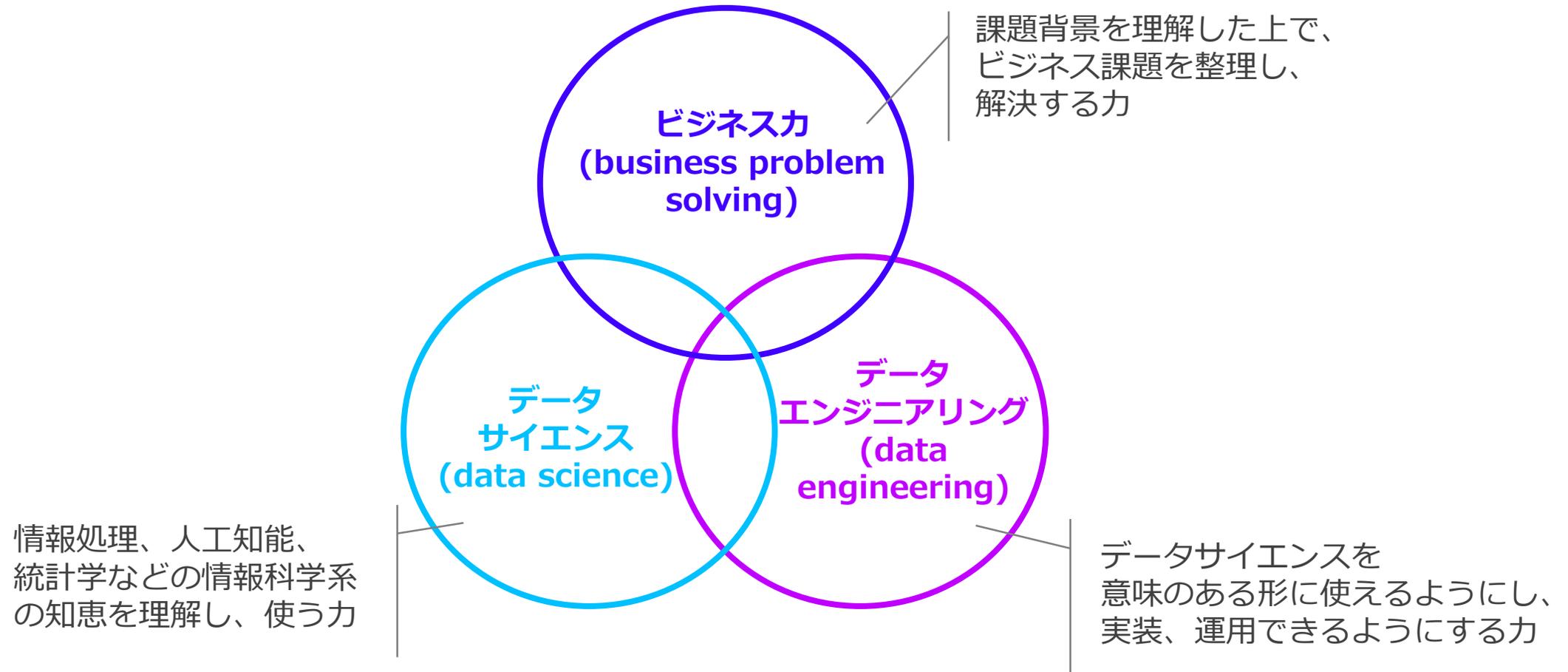
設立時の状況

- ・ 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」）に必要な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みなさんが相互に接し、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

3つのスキルセット

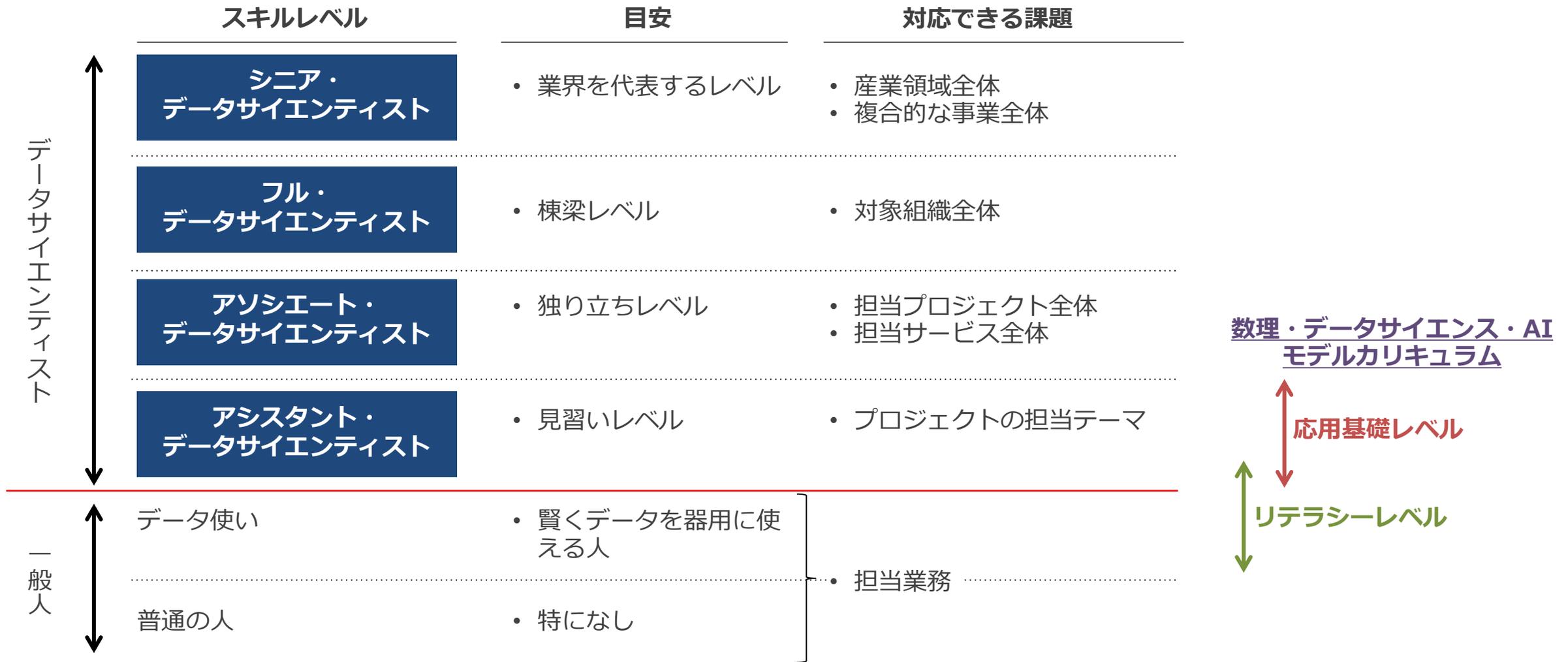


資料 : データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

データサイエンティストは、
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

スキルレベル

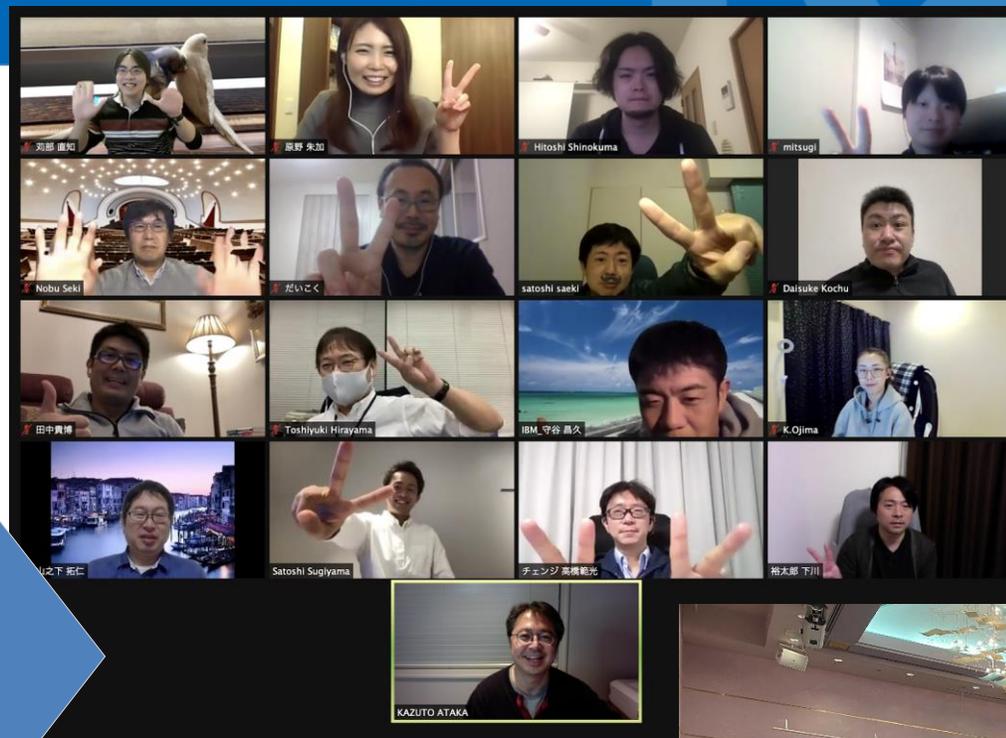


資料：データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

スキル定義委員会の活動

| | |
|------|--|
| 2014 | <ul style="list-style-type: none">● データサイエンティストの人材像、ミッション、スキルカテゴリ、スキルレベルの定義発表 |
| 2015 | <ul style="list-style-type: none">● データサイエンティストに必要とされるスキルを「<u>データサイエンティストスキルチェックリスト</u>」とし、ver.1を公開● 「データサイエンティストのミッション、スキルセット、定義、スキルレベル」について<u>2015年版</u>を発表 |
| 2016 | <ul style="list-style-type: none">● スキルチェックの普及活動、各種コンテンツの取材/発信 |
| 2017 | <ul style="list-style-type: none">● 2017年度版スキルチェックリスト作成ver.2、タスクリスト作成ver.1(IPA協働) |
| 2018 | <ul style="list-style-type: none">● スキルチェックの普及活動、各種コンテンツの取材/発信 |
| 2019 | <ul style="list-style-type: none">● DS活動普及のためのPRやスキルチェックの普及活動● 2019年度版スキルチェックリスト作成ver.3、タスクリスト作成ver.2 (IPA協働) |
| 2020 | <ul style="list-style-type: none">● データサイエンティスト検定のスコープ決め、問題策定などの準備● スキルチェックリスト/タスクリスト概説発表 (IPA協働) |
| 2021 | <ul style="list-style-type: none">● 2021年度版スキルチェックリスト作成ver.4、タスクリスト作成ver.3 (IPA協働)● 第1回データサイエンティスト検定 9月実施 |

スキル定義委員会の風景



2. 2021年度スキルチェックリスト

アップデート内容のハイライト

全体

- モデルカリキュラムとの整合性の調整。★1スキルやカテゴリ・サブカテゴリ名称の大幅見直し
- AI活用や機械学習のシステム運用、アルゴリズム、それに伴う倫理課題に関するスキル追加
- ライブラリ活用やクラウドサービスの一般化に伴うスキルの更新。特に、★1スキルの相対的増加

ビジネス

- AI活用時のビジネス視点を強化
「AI活用検討」、「AI-ready」サブカテゴリを追加
「データ倫理」を「データ・AI倫理」に変更
- プロジェクトの多様なフェーズに対応するスキル
データ入手、PoC、サービス維持/完了時スキル
- DSから「分析アプローチ」スキル移設
- マネジメントスキルをプロジェクトと組織に2分割

データ
エンジニア
リング

- 新規カテゴリ「AIシステム運用」を追加
「ソース管理」「AutoML」「MLOps」「AIOps」
- 環境構築技術等、技術トレンドへの対応
ノーコード・ローコード、クラウドマネージドサービス、
コンテナ技術、認証、ゼロトラストなど
- リテラシーレベルのプログラミングスキル追加
- HadoopやScala、GPUなど一般化技術の一部削除

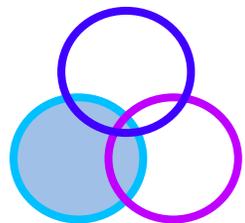
データ
サイエンス

- 基礎数学の重要性見直し
対数・指数や集合に関するスキル追加
- 学習と予測カテゴリの明確化、潮流への対応
機械学習や強化学習を「学習」スキルへ集約
敵対的サンプル、深層学習メリット、精度低下
学習と予測の★1を増加
- 近年、重要度の増したスキル項目追加
因果推論、標本抽出、自然言語、画像認識

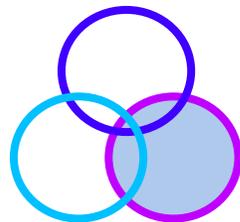
タスク
リスト

- 全体としては大きな変更はなし
- AI倫理やプライバシー課題に対応するタスク追加
「コンプライアンス・倫理・権利の確認」タスク追加
ユーザデータの扱いやアルゴリズム活用のモニタリング
- 開発、運用時に重要度が増したタスクを追加
UI/UX開発、分析モデルモニタリング、
プロジェクト要否や終了の判断タスク
- 文言更新: 自然言語処理、画像・映像認識、音声認識

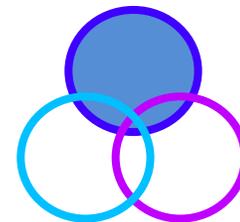
スキル表 項目数



データサイエンス



エンジニアリング



ビジネス



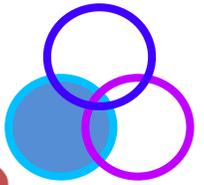
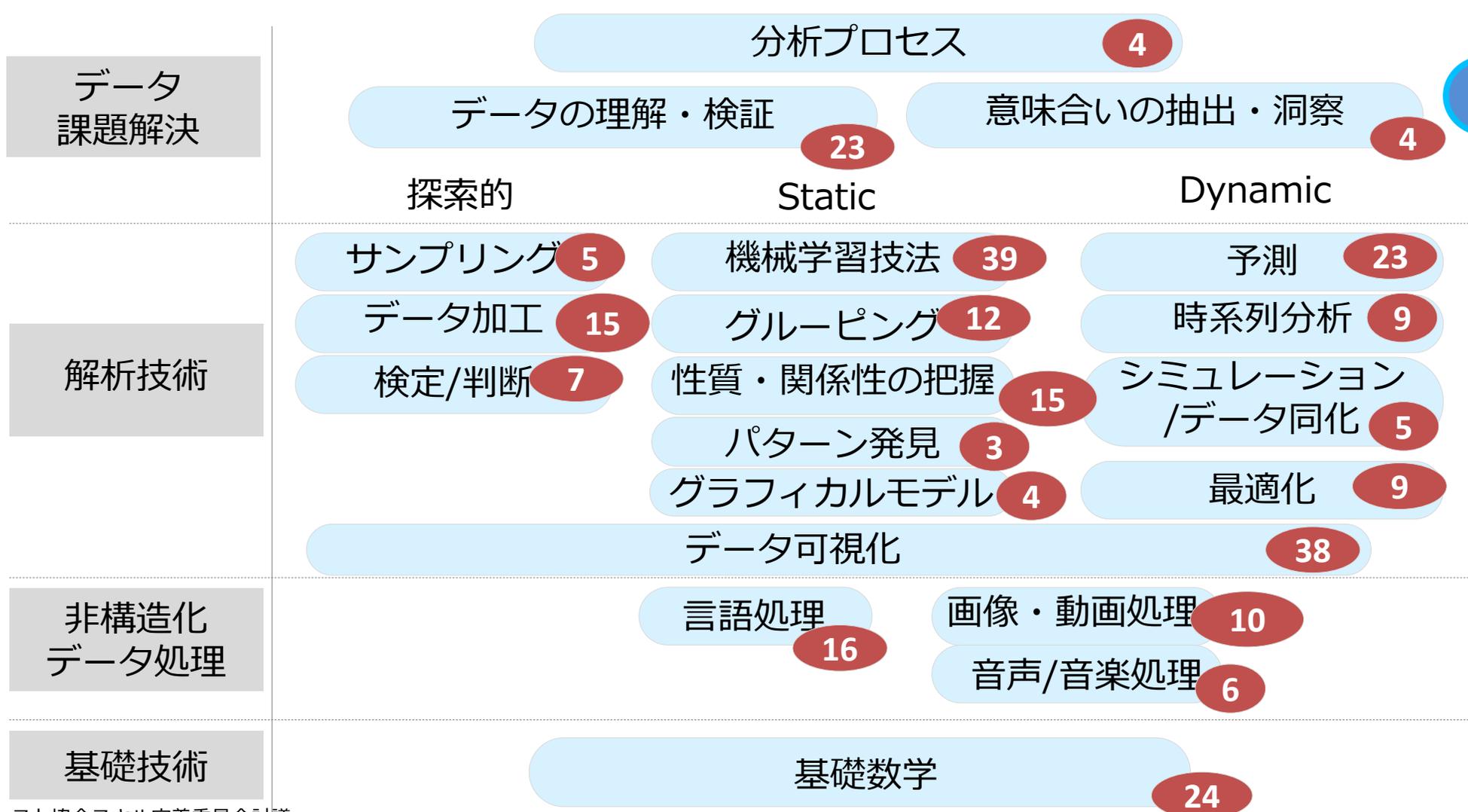
レベル計



| | 2015 | 2017 | 2019 | 2021 | 2015 | 2017 | 2019 | 2021 | 2015 | 2017 | 2019 | 2021 | 2015 | 2017 | 2019 | 2021 |
|--------------------------|------|------|------|------------|------|------|------|------------|------|------|------|------------|------|------|------|------------|
| ★★★ 棟梁レベル (フル) | 58 | 71 | 64 | 62 | 37 | 39 | 42 | 37 | 44 | 38 | 42 | 47 | 139 | 148 | 149 | 146 |
| ★★ 一人前レベル (アソシエート) | 68 | 90 | 121 | 120 | 43 | 52 | 63 | 65 | 49 | 42 | 49 | 56 | 160 | 184 | 232 | 241 |
| ★ 見習いレベル (アシスタント) | 54 | 67 | 86 | 100 | 39 | 38 | 39 | 57 | 30 | 20 | 22 | 28 | 123 | 125 | 149 | 185 |
| 領域計 | 180 | 228 | 271 | 282 | 119 | 129 | 144 | 159 | 123 | 100 | 113 | 131 | 422 | 457 | 528 | 572 |

資料：データサイエンティスト協会スキル定義委員会討議

Ver.3



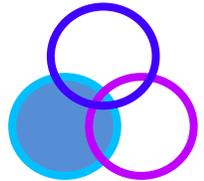
資料：データサイエンティスト協会スキル定義委員会討議

Ver.4

データ
課題解決

データの理解・検証 **25**

意味合いの抽出・洞察 **3**



探索的

Static

Dynamic

解析技術

サンプリング **5**

グルーピング **12**

予測 **20**

学習 **49**

データ加工 **14**

性質・関係性の把握 **22**

時系列分析 **8**

推定・検定 **7**

パターン発見 **4**

シミュレーション・データ同化 **5**

最適化 **9**

データ可視化 **38**

非構造化
データ処理

自然言語処理 **16**

画像・映像認識 **12**

音声認識 **6**

基礎技術

基礎数学 **27**

資料：データサイエンティスト協会スキル定義委員会討議

アップデートの詳細と例 ①データサイエンス

① 「基礎数学」の拡充

2019年版

基礎数学

- 統計数理基礎
- 線形代数基礎
- 微分・積分基礎

2021年版

基礎数学

- 統計数理基礎
- 線形代数基礎
- 微分・積分基礎
- 集合論基礎

② 「因果推論」のサブカテゴリ昇格

2019年版

性質・関係性の把握

性質関係性の把握

グラフィカルモデル

グラフィカルモデル

2021年版

性質・関係性の把握

性質関係性の把握

グラフィカルモデル

因果推論

因果推論スキルをサブカテゴリ化しつつスキルを追加し、関連の深いグラフィカルモデルを移動

③ 「予測」と「学習」のスキル配置

予測

回帰・分類

評価

学習

機械学習

深層学習

強化学習

評価のためのデータ分割やリークに関する理解、精度モニタリングとリモデルなど機械学習と関係の深いスキルを移動

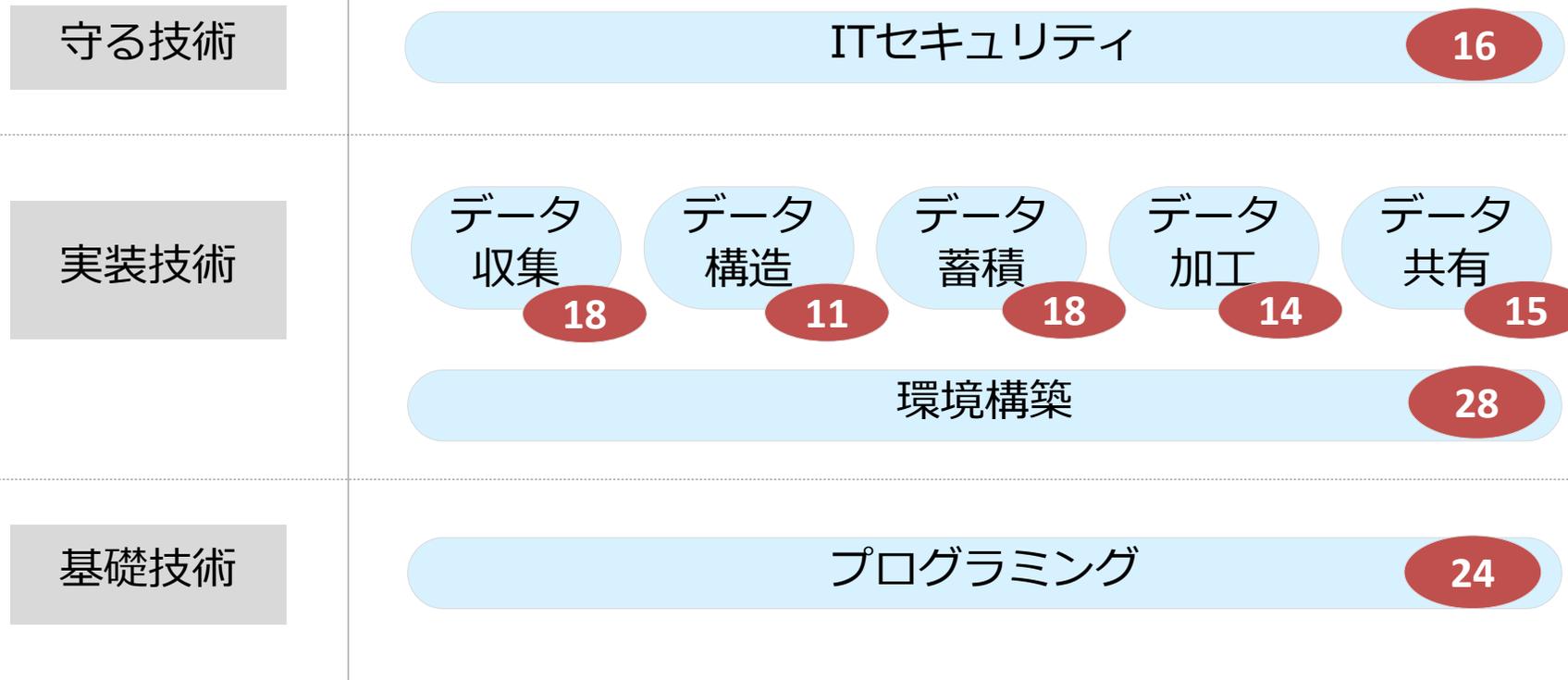
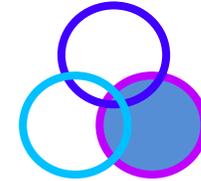
④ 近年の技術動向を反映

学習

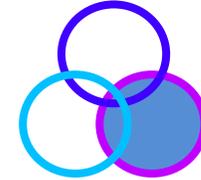
- ・ 深層学習におけるAttention/Transformer
- ・ 機械学習におけるデータドリフトとモニタリングなど

その他

- ・ 一般化線形モデル、一般化線形混合モデルに関するスキル追加
- ・ 分析プロセスのカテゴリをビジネスカに統合
- ・ 映像認識のトラッキングに関するスキル追加
- ・ 基礎的なカテゴリを上位に配置
- ・ 学習のサブカテゴリ名見直し（例：画像処理->画像認識） など



資料：データサイエンティスト協会スキル定義委員会討議



資料：データサイエンティスト協会スキル定義委員会討議

アップデートの詳細と例 ②エンジニアリング

① 環境構築

クラウド技術の拡充 → クラウド構築の簡易化

2019年版

オンプレミス

クラウド

2021年版

オンプレミス

クラウド

仮想化 (コンテナ)

マネージドサービス

② 新カテゴリ「AIシステム運用」追加

AIシステムを運用するために必要なスキル

MLOps

開発したAIモデルのライフサイクル管理

モデル作成 → モデルの評価 → デプロイ →
学習データの更新 → モデル再学習

③ 新技術

エンジニアとして注目すべき新技術

通信技術

5G、LPWA

設計手法

マイクロサービス

自動化技術

AutoML、AIOps、RPA

④ 重要スキル強化

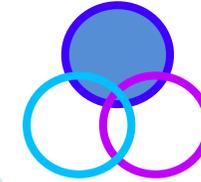
プログラム開発

オブジェクト指向言語
外部API活用
テスト技法 (ホワイトボックス・ブラックボックステスト)

セキュリティ

認証 (OAuth)
シングルサインオン (SAML)
ゼロトラスト

Ver.3



資料：データサイエンティスト協会スキル定義委員会討議

Ver.4



資料：データサイエンティスト協会スキル定義委員会討議

アップデートの詳細と例 ③ ビジネス

① 行動規範

データ・AI倫理として、意図的な悪用やディープフェイク、リバーズエンジニアリングなどへのリスク対処を追加

2019年版

行動規範

ビジネスマインド
データ倫理
コンプライアンス

2021年版

行動規範

ビジネスマインド
データ・AI倫理
コンプライアンス

② 着想・デザイン

プライバシーバイデザインなどを考慮したサービスデザインカ、学習済みモデルやライブラリの適用検討スキル追加

2019年版

着想・デザイン

着想
デザイン
開示・非開示の決定

2021年版

着想・デザイン

着想
デザイン
AI活用検討
開示・非開示の決定

資料：データサイエンティスト協会スキル定義委員会討議

③ アプローチ設計

DX課題などでの入手困難なデータを確保し、分析可能な環境を構築するスキルを追加。DS領域からモデル作成イメージのスキルを移管

2019年版

課題の定義

KPI
スコーピング
価値の見積もり
アプローチ設計

データ入手

データ入手

2021年版

課題の定義

KPI
スコーピング
価値の見積もり

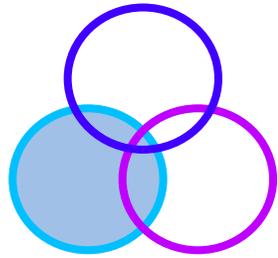
アプローチ設計

データ入手
AI-Ready
アプローチ設計

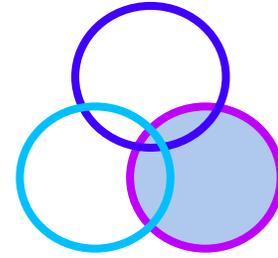
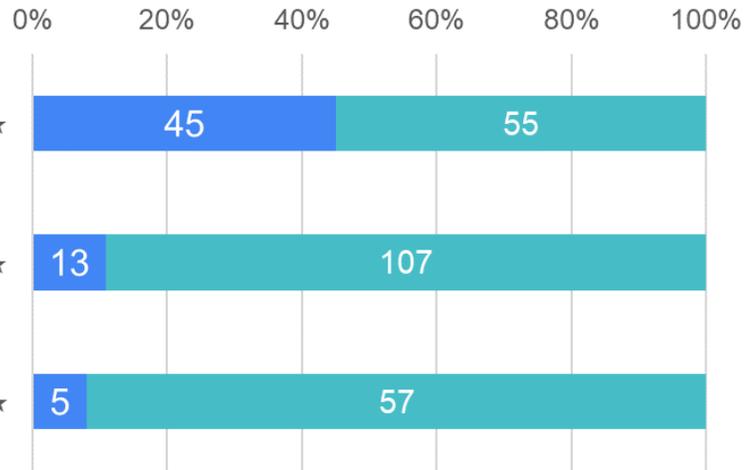
その他

- 活動マネジメントのスキルカテゴリをPJマネジメントと組織マネジメントに分割・変更
- ビジネスマインドとして、時代の変化に柔軟に対応し行動するスキルを追加
- ドキュメンテーションに出自や引用元についての適切な判断を追加
- 稼働管理や監視など、サービス維持スキルを追加
- 契約内容に応じたデータの削除や、完了処理スキルの追加 など

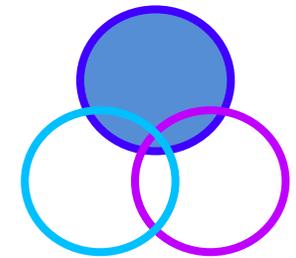
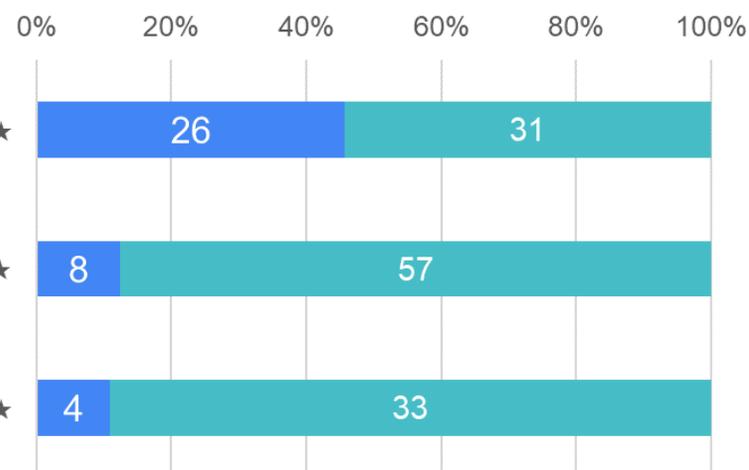
必須スキルの項目について



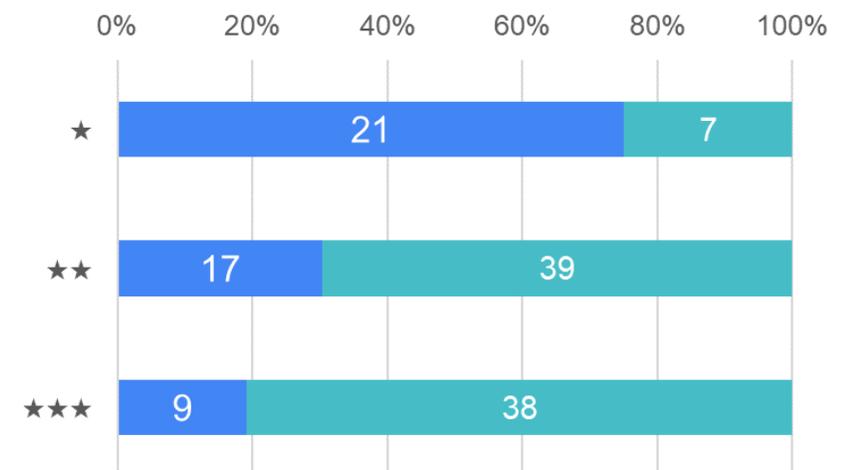
データサイエンス



エンジニアリング



ビジネス



■ 必須スキル ■ 通常スキル

資料：データサイエンティスト協会スキル定義委員会討議

委員会内での議論ポイント

■ 必須になりそうでなかったスキル

【DS 学習 機械学習 ★★】

予測モデルにおけるリークとは何か、それがもたらす問題について説明できる

【BZ データの理解・検証 俯瞰・メタ思考 ★】

データが生み出された背景を考え、鵜呑みにはしないことの重要性を理解している

【BZ データの理解・検証 意味合いの抽出、洞察 ★】

分析結果を元に、起きている事象の背景や意味合い(真実)を見抜くことができる

【DE AIシステム運用 ソース管理 ★】

GitやSubversionなどのバージョン管理ソフトウェアを活用して、開発した分析プログラムのソースをリポジトリに登録しチームメンバーと共有できる

■ 毎回議論して、ついに必須化

【DE データ蓄積 クラウド ★】

クラウド上のストレージサービス(Amazon S3、Google Cloud Storage、IBM Cloud Object Storageなど)に接続しデータを格納できる

委員会内での議論ポイント

■新規追加したスキル

【BZ 論理的思考 ドキュメンテーション ★】

データの出自や情報の引用元に対する信頼性を適切に判断し、レポートに記載できる

【BZ 着想・デザイン AI活用検討 ★★】

直近の世界的なAI倫理やプライバシー保護情勢を考慮し、担当事業における適切なデータ・AI活用を検討できる

【BZ 着想・デザイン AI活用検討 ★】

弱いAI・強いAI、特化型AI・汎用型AIの違いを説明できる

■削除したスキル

【DE データ蓄積 分散技術分散処理 ★★】

環境のディストリビューションを導入できる(Hortonworks、CDHなど)

【BZ コンプライアンス ★】

仮名化と匿名化の違いを理解しており、適切な方法で匿名加工情報を扱うことができる

今後の公開予定

シンポジウム直後

本日の講演資料を協会HPからダウンロード可能に

11月19日
(予定)

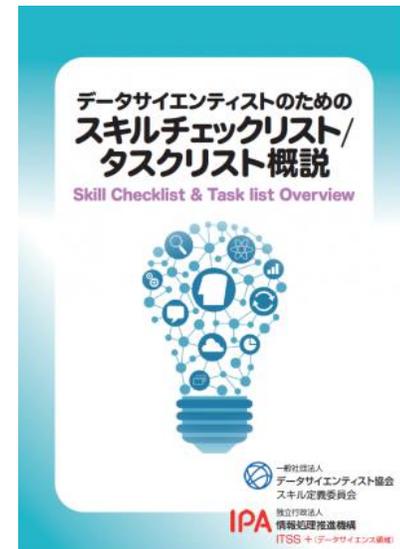
データサイエンティスト協会HP、IPAサイトで公開

- ・スキルチェックリスト ver.4
- ・タスクリスト ver.3

2022年1月目途

・概説書のアップデート IPAサイト、協会HPからお知らせ

<https://www.ipa.go.jp/files/000083733.pdf>



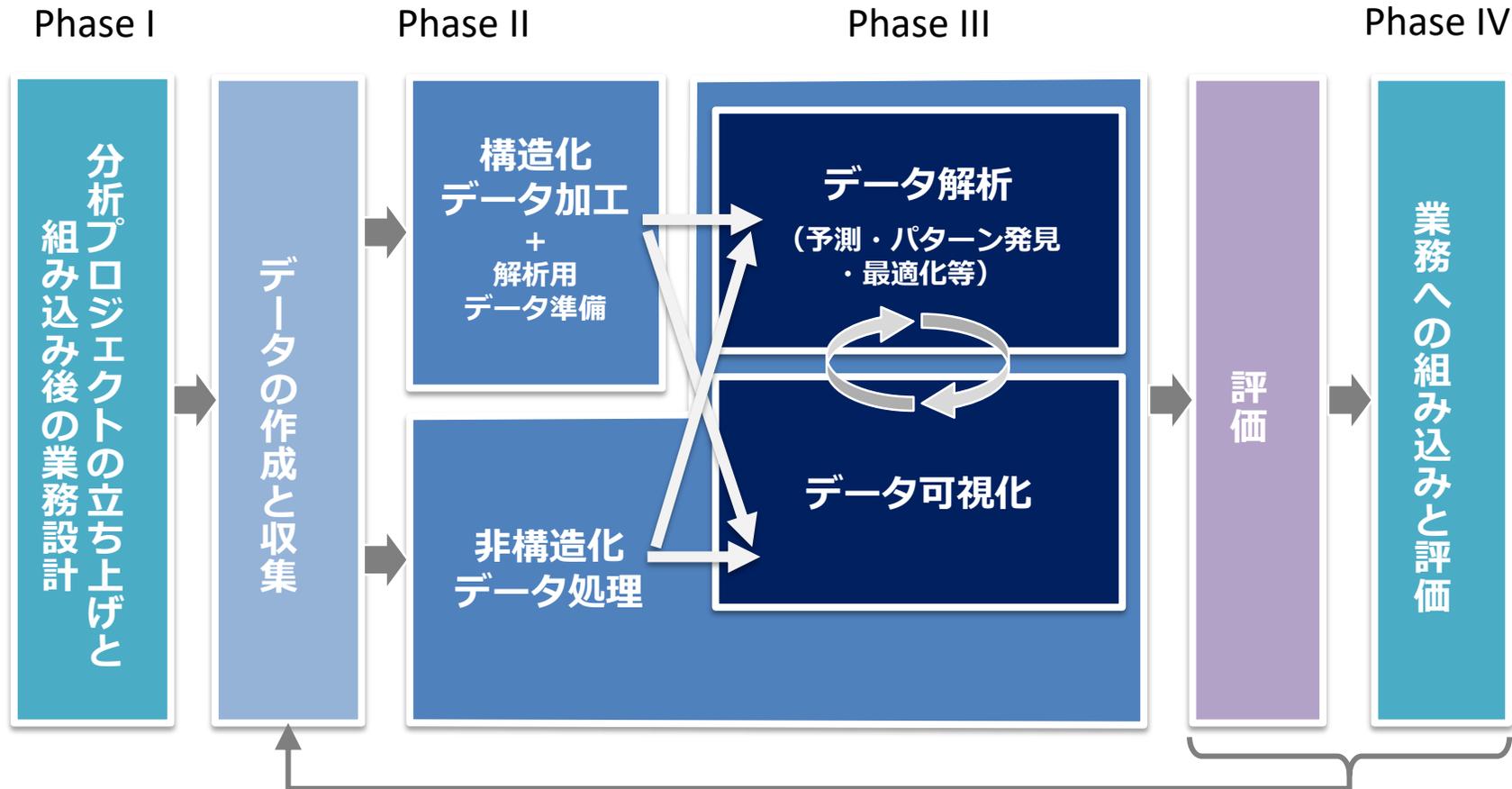
公開予定のExcel：色々変わりましたのでどのスキルカテゴリに どんなスキルの増減があるかをまとめたシートを作りました

| Ver4 | | | | スキル数 | | | 必須スキル数 | | | |
|------------|-------------|---|--------------|--------------|----|----|--------|----|----|-----|
| スキルカテゴリ | サブカテゴリ | カテゴリ・サブカテ レベルでの主な変更内容 | カテゴリ スキル数 | サブカテ スキル数 | ★ | ★★ | ★★★ | ★ | ★★ | ★★★ |
| 基礎数学 | 統計数理基礎 | ・対数関数と指数関数に関する知識追加 (★1) | 27 | 18 | 14 | 4 | 0 | 14 | 0 | 0 |
| 基礎数学 | 線形代数基礎 | | | 4 | 4 | 0 | 0 | 2 | 0 | 0 |
| 基礎数学 | 微分・積分基礎 | | | 3 | 3 | 0 | 0 | 2 | 0 | 0 |
| 基礎数学 | 集合論基礎 | ・新設 ・集合とベン図のスキル追加 (★1) ・論理演算と集合演算の対応に関するスキル追加 (★1) | | 2 | 2 | 0 | 0 | 1 | 0 | 0 |
| データの理解・検証 | 統計情報への正しい理解 | ・カテゴリの位置変更 (上に移動) ・母集団の比率的な指標も見べきことの重要性に関する知識追加 (★1) | 25 | 4 | 2 | 1 | 1 | 2 | 0 | 0 |
| データの理解・検証 | データ確認 | ・カテゴリの位置変更 (上に移動) | | 6 | 2 | 2 | 2 | 1 | 1 | 1 |
| データの理解・検証 | 俯瞰・メタ思考 | ・カテゴリの位置変更 (上に移動) | | 3 | 1 | 1 | 1 | 0 | 1 | 1 |
| データの理解・検証 | データ理解 | ・カテゴリの位置変更 (上に移動) | | 10 | 4 | 3 | 3 | 2 | 1 | 0 |
| データの理解・検証 | データ粒度 | ・カテゴリの位置変更 (上に移動) | | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| 意味合いの抽出、洞察 | 洞察 | ・カテゴリの位置変更 (上に移動) | 3 | 3 | 2 | 1 | 0 | 1 | 1 | 0 |
| 予測 | 回帰・分類 | ・一般化線形回帰と一般化線形混合モデルに関するスキル追加 (★2) ・Cox回帰に関するスキル追加 (★2) | 20 | 14 | 3 | 9 | 2 | 1 | 1 | 1 |
| 予測 | 評価 | ・機械学習寄りのスキルを学習・機械学習に移動 ・macro平均、micro平均、重み付き平均といった具体例追加 ・予測値・誤差の可視化、データの部分集 | | 6 | 3 | 1 | 2 | 1 | 0 | 0 |

3. 2021年度版タスクリスト (ver.3)

タスクリストとは

ITSS+ データサイエンス領域のタスクリストとして2017年4月に公開

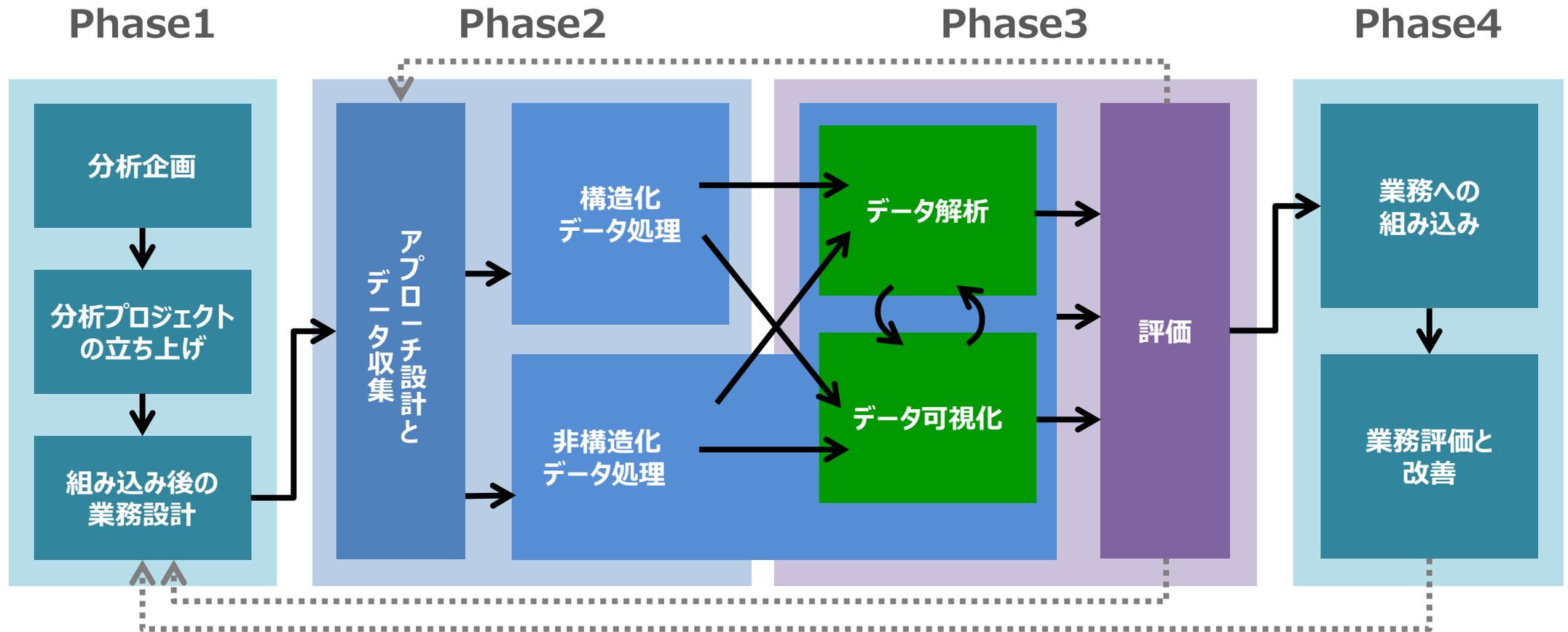


<https://www.ipa.go.jp/jinzai/itss/itssplus.html>

2021年度版タスクリスト (ver.3)

ITSS+ データサイエンス領域のタスクリスト

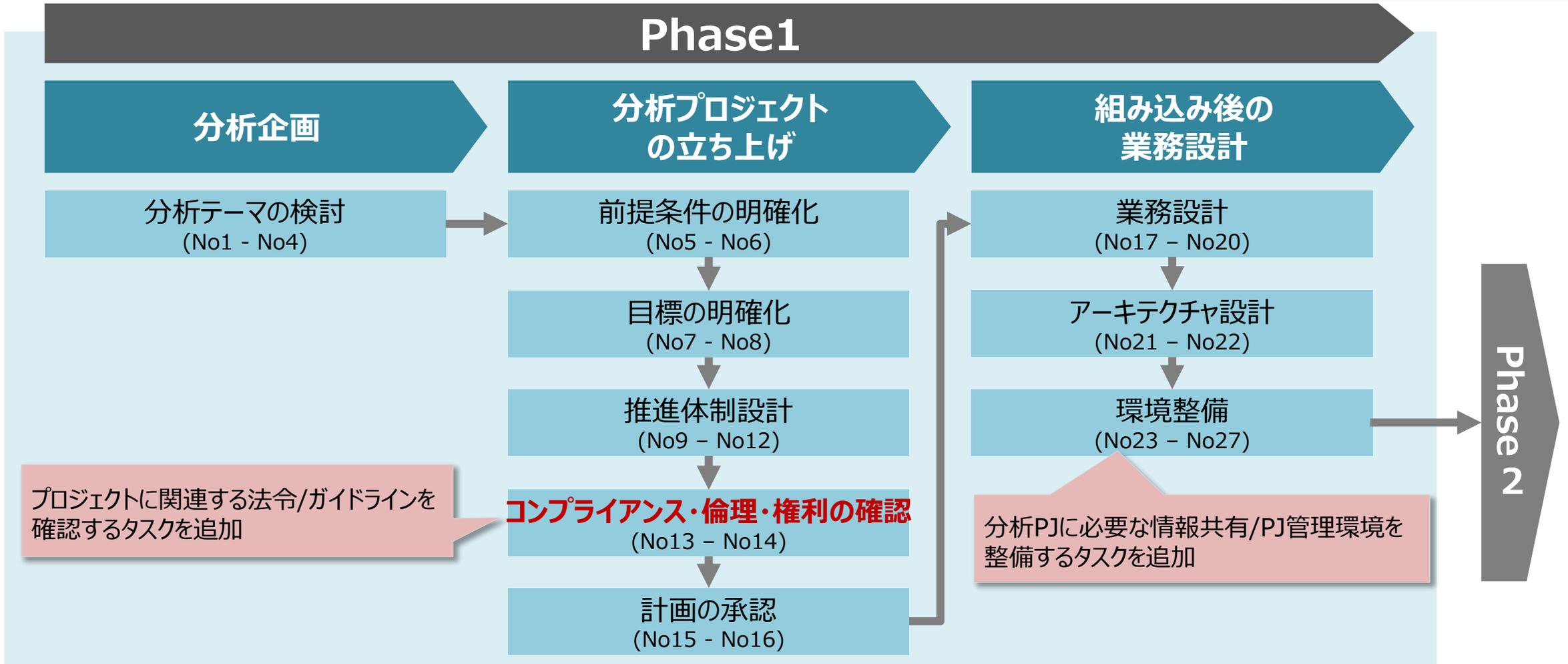
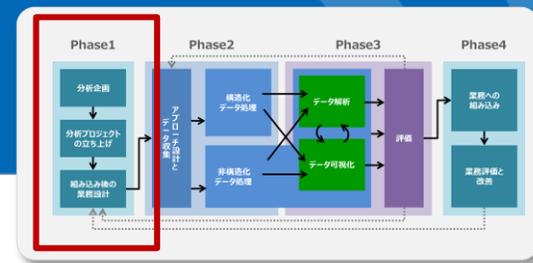
→ 主な流れ
.....> フィードバック



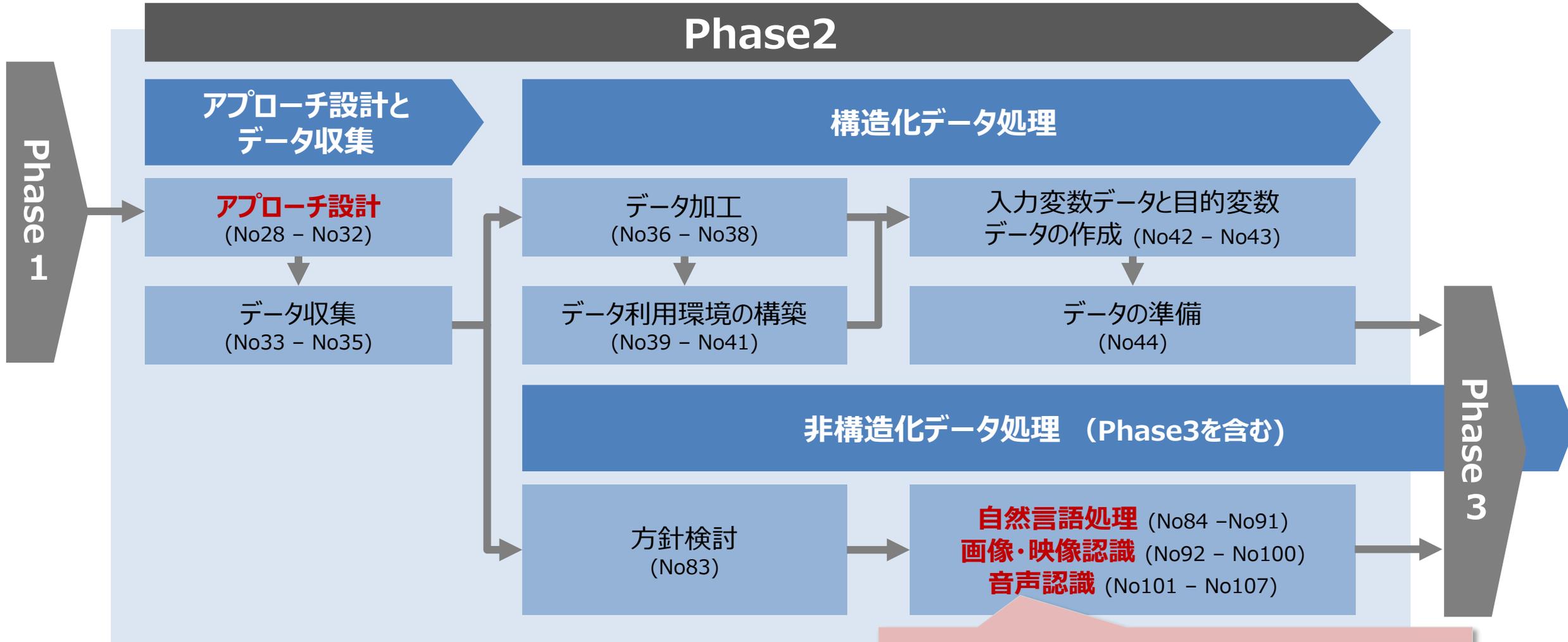
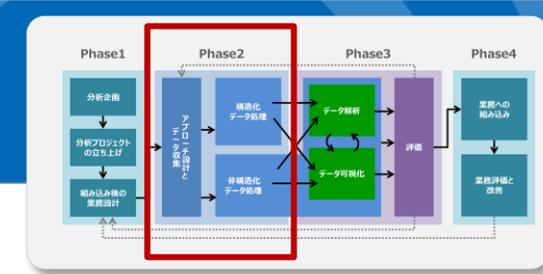
改定の要点整理と背景、理由

- データサイエンティスト スキルチェックリスト ver.4に合わせて文言反映
(**自然言語処理、画像・映像認識、音声認識**、など)
- AI倫理や個人情報、プライバシーの重要性を踏まえ、
**法令/ガイドラインを確認するタスク、ユーザデータの取り扱いに関するタスク、
データ運用およびアルゴリズム活用をモニタリングするタスク**などを追加
- UI/UXの重要性を踏まえ、分析結果を業務で活用するための
ユーザインタフェースを開発するタスクを追加
- MLOpsなど分析モデルを運用するシステムが増えている背景を踏まえ、
分析モデルをモニタリングする仕組みを構築するタスクを追加
- Phase4に**新プロジェクトの要否を判断するタスク**および**プロジェクト終了時のタスク**を追加

タスクリスト詳細 Phase 1

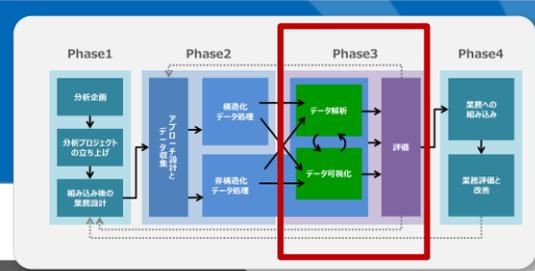


タスクリスト詳細 Phase 2



スキルチェックリストVer4対応
画像・映像認識のタスクの中に画像分類、画像生成を追加

タスクリスト詳細 Phase 3



Phase 3

データ解析

データ集計、抽出(No45 - No46) 性質・関係性の把握(No47 - No51)
検定・推定(No52 - No54) 予測(No55 - No59) グループング(No60 - No61)
パターン発見(No62 - No65) 最適化(No66 - No68)
シミュレーション(No69 - No70)

性質・関係性の把握のタスクの中に因果推論を追加
予測のタスクの中に判別分析を追加

データ可視化

可視化の企画とデータ収集 (No71 - No74) → 2軸チャート化(No75) 多次元の可視化(No76)
関係性の可視化(No77) 地図上の可視化(No78)
挙動・軌跡の可視化(No79 - No80)
ダイナミックな可視化(No81) リアルタイム可視化(No82)

評価

モデル評価 (No108 - No110)

分析評価 (No111 - No114)

分析評価のタスクの中に効果検証を追加

非構造化データ処理 (Phase2を含む)

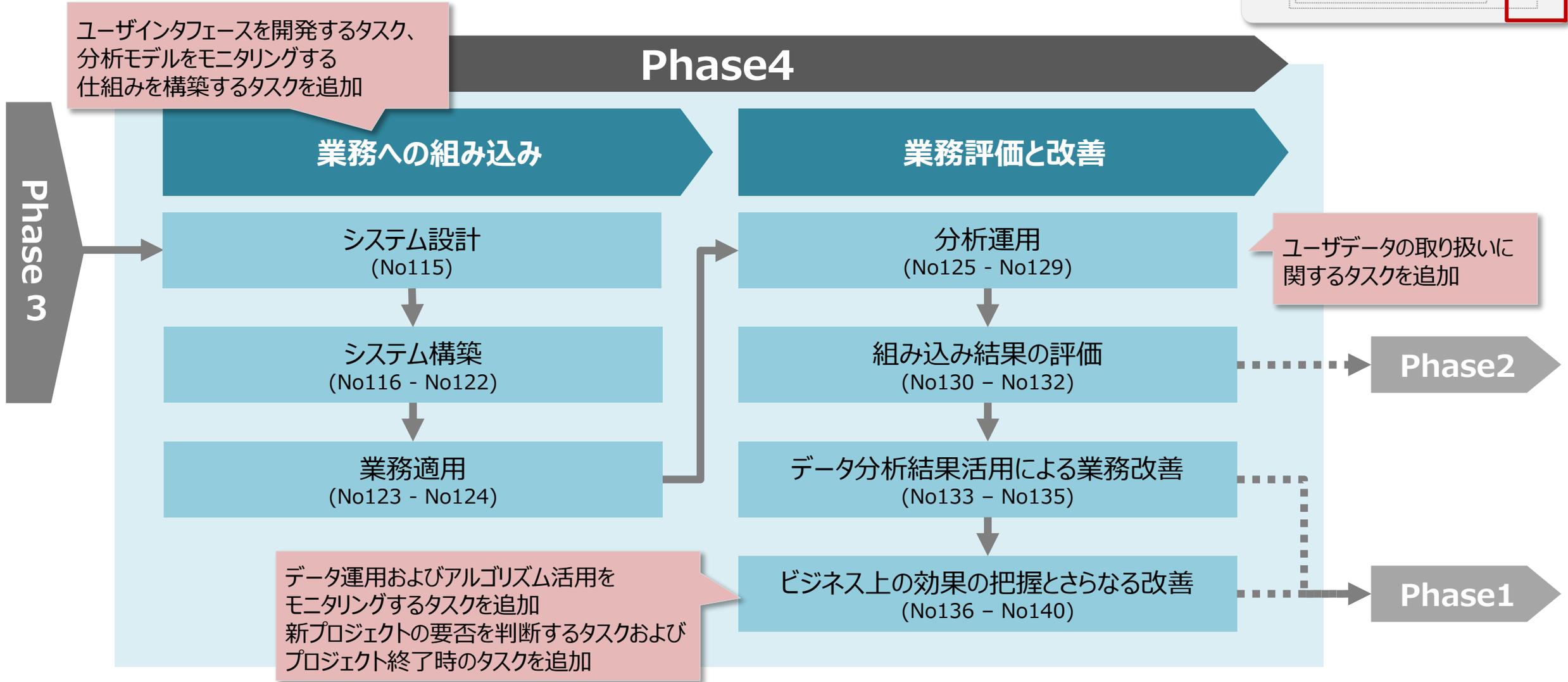
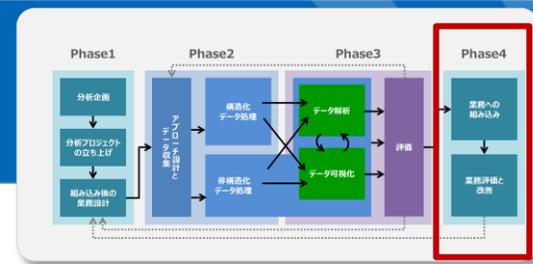
Phase 2

Phase 2

Phase 1

Phase 4

タスクリスト詳細 Phase 4



4. データサイエンティスト検定 実施報告

データサイエンティスト検定始動！その理由とは・・・

アカデミアからの要望

大学でのデータサイエンス
の必修化

AIやDXに代表される社会的潮流
と実践できる人材の見極め

産業界でDSが活躍する場の
広がり&役割の広がり

実務家が集うDS協会自身が
感じている課題と責務

DSに求められるスキル定義
の一層の明確化とその普及

データサイエンティストのスキル・知識を身につけていることを証明する一定の基準が必要



2021

データサイエンティスト検定
リテラシーレベル★

データサイエンティスト検定とは？



2021

データサイエンティスト検定
リテラシーレベル★

アシスタント・データサイエンティスト（見習いレベル：★）と数理・データサイエンス教育強化拠点コンソーシアムが公開している数理・データサイエンス・AI（リテラシーレベル）におけるモデルカリキュラムを総合し、実務能力と知識を有することを証明する試験

検定の範囲と対象

- ・データサイエンティスト初学者やこれから目指すビジネスパーソン
- ・データサイエンティストに興味を持つ学生

| スキルレベル | | 目安 | 対応できる課題 |
|--|------|------------|--------------------------|
| Senior Data Scientist 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 |
| Full Data Scientist フル データサイエンティスト | ★★★ | 棟梁レベル | ・対象組織全体 |
| Associate Data Scientist 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 |
| Assistant Data Scientist 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 |

+

リテラシーレベルのモデルカリキュラムの内容

検定概要

- 受験資格 特になし
- 実施概要 選択式問題、CBT方式（全国）
- 試験期間 2021年9月11日～30日
- 申込期間 ～8月29日（日）
- 参考書籍や対策講座
9月上旬に順次出版、公開

DS協会HPより

データサイエンティスト検定実施報告

実施報告 第一回データサイエンティスト検定

- 受検者数 約1400人
- 合格者数 927人
- 合格率 約66%
- 合格ラインの目安 正答率80%

合格者の方には
・ デジタルバッジ
・ 認定証
・ 特製ステッカー
送付予定

来年の試験は2022年春、秋2回を予定
今回発表したスキルチェックリストVer.4に基づき、
試験問題を刷新しますので、上記合格率、合格ラインは参考値です

デジタルリテラシー協議会の発足とDi-Lite

デジタルリテラシー協議会

より良い社会(Society5.0)の創出に向け、「デジタルを作る人材」だけでなく「**デジタルを使う人材**」の育成が重要になってきています。デジタルリテラシー協議会は、現代におけるビジネスパーソンが持つべきデジタル時代の共通リテラシー「**Di-Lite(ディーライト)**」の整備と社会標準実装を目指し、官民連携の会議体として設立されました。当協議会は、情報発信や啓発活動を通じて、社会全体のリテラシーレベルを向上し、日本の産業力強化に貢献します。

参画団体

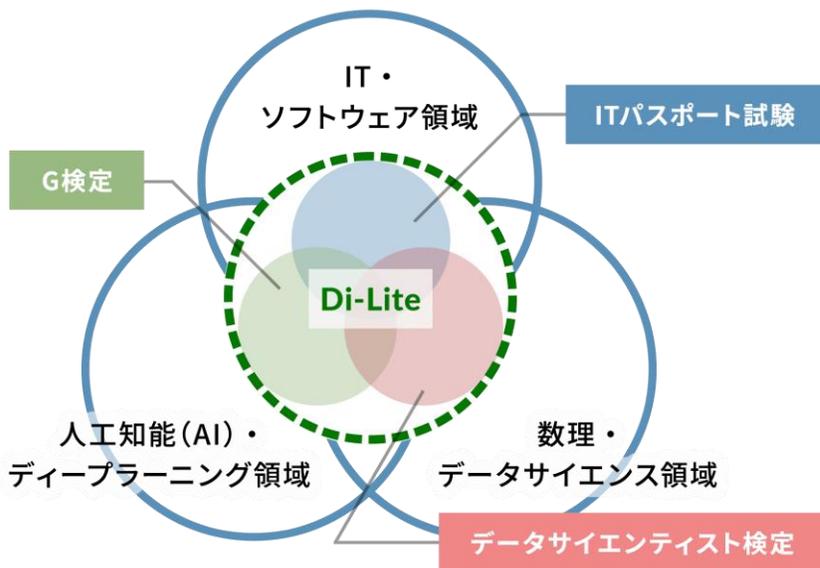
 DataScientist Society
一般社団法人
データサイエンティスト協会(DSS)

 Japan Deep Learning Association
一般社団法人
日本ディープラーニング協会(JDLA)

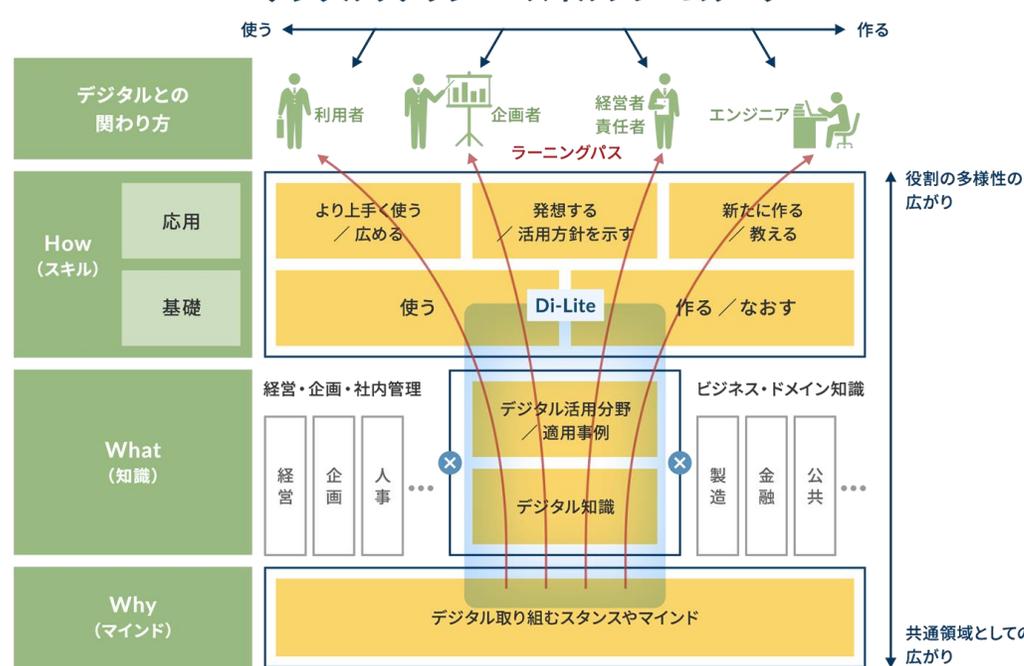
 IPA
独立行政法人
情報処理推進機構(IPA)

【オブザーバー】 経済産業省

Di-Lite デジタルリテラシー協議会が定義する、全てのビジネスパーソンが持つべきデジタル時代の共通リテラシー



デジタルリテラシー・スキルフレームワーク



ご清聴ありがとうございました！
来年もよろしくお願ひいたします