

2014年12月10日

プレスリリース

報道関係各位

一般社団法人データサイエンティスト協会

## データサイエンティスト協会、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、スキルレベルを発表

一般社団法人データサイエンティスト協会（所在地：東京都港区、代表理事 草野 隆史、以下データサイエンティスト協会）およびスキル定義委員会（委員長 安宅 和人、副委員長 佐伯 諭）は、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」について公開いたしました。

本内容は、本年11月27日に開催された「データサイエンティスト協会 1st シンポジウム」において、スキル定義委員会より「データ社会に求められる新しい才能とスキル」として発表されたものです。

### データサイエンティスト協会の設立背景と目的

2013年5月に、データサイエンティスト協会が設立された背景には、以下のような課題がありました。

1. 「データサイエンティスト」という言葉の定義の欠落とバズワード（\*1）化
2. 当該人材の「雇用者側」の期待に対して、「被雇用者側」のスキルセットが一致しないケースが増加
3. 若い才能をもった当該人材の候補者が、どのように訓練し、スキルを身に付けていくべきか、わからないといった状況

上記の課題認識を踏まえ、データサイエンティスト協会は、広くデータ活用の担い手となる人材の育成を目指し、以下の3つの目的を果たすために設立されました。

1. 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こと
2. スキルの育成と評価のための軸、基準を作ること
3. データサイエンティストが相互に接し、意見交換等が行える、開かれた環境を提供すること

今回発表された、スキル定義委員会の「ミッション、スキルセット、定義、スキルレベル」では、上記の1、2に主眼をおいて作成されております。



## データサイエンティストのミッション

現在、情報技術革命が進む中であって、センサの数が激増し、人類の計算キャパシティは幾何級数的に拡大しております。また、ブロードバンド化に伴い、ネットワークの接続スピードは劇的に高まり、情報トラフィックも激増しております。携帯電話、Wi-Fi などのワイヤレスネットワークが広がることに伴い、これまで拾い上げられなかった粒度の情報もリアルタイムで拾い上げることが可能となりました。

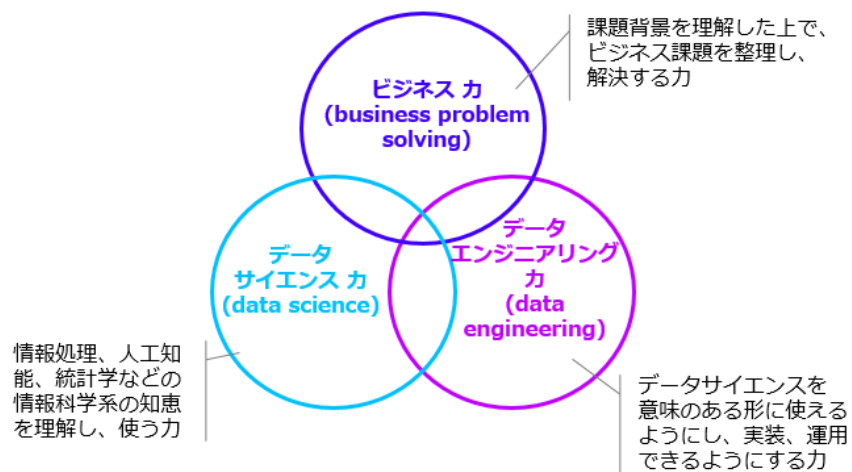
このような時代に求められるのは、人間を数字入力や情報処理の作業から開放するプロフェッショナル人材であり、「データの持つ力を解き放つ」ことが新しい時代におけるデータプロフェッショナル、すなわち「データサイエンティスト」のミッションだと考えます。

## データサイエンティストに求められるスキルセット

前述のミッションを踏まえ、データサイエンティストに求められるスキルセットは、以下の3つであると考えます（図1参照）。

1. ビジネス力 (business problem solving) : 課題背景を理解した上で、ビジネス課題を整理し、解決する力
2. データサイエンス力 (data science) : 情報処理、人工知能、統計学などの情報科学系の知恵を理解し、使う力
3. データエンジニアリング力 (data engineering) : データサイエンスを意味のある形に使えるようにし、実装、運用できるようにする力

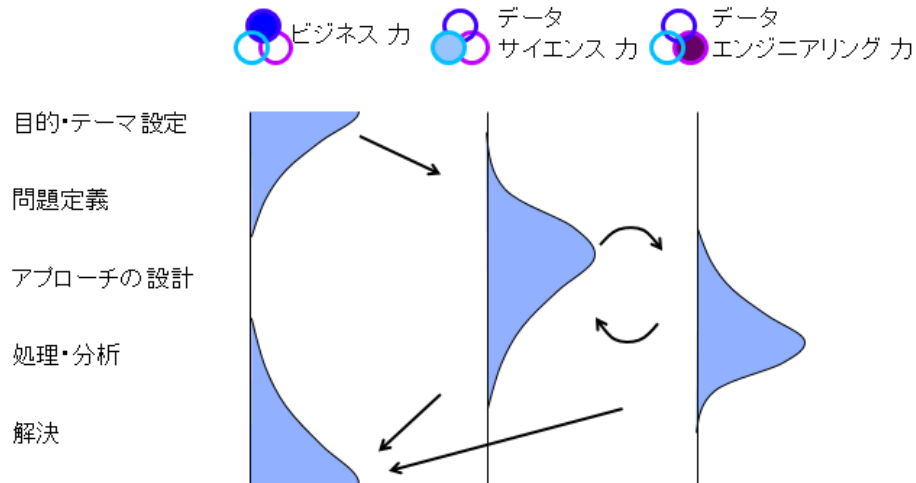
図1 : データサイエンティストに求められるスキルセット



上記の3つのスキルはどの一つが欠けてもいけません。また、この3つのスキルは課題解決のフェーズによって、中心となるスキルが変化します（図2参照）。



図 2 : 課題解決の各フェーズで要求されるスキルセットのイメージ



### データサイエンティストの定義

データサイエンティスト協会では、これからの時代に求められるデータサイエンティストを以下のように定義します。

「データサイエンティストとは、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」

ここでいう「ビジネス」とは、社会に役に立つ意味のある活動全般を指します。また、「プロフェッショナル」とは、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、その結果に対して認識された価値の対価として報酬を得る人を示します。

### データサイエンティストのスキルレベル

データサイエンティスト協会では、3つのスキルセットのレベルによって、大きくデータサイエンティストには4つのスキルレベルがあると考えます（図3参照）。

1. 業界を代表するレベル : Senior Data Scientist
2. 棟梁レベル : (full) Data Scientist
3. 独り立ちレベル : Associate Data Scientist
4. 見習いレベル : Assistant Data Scientist

\*前提条件として上位レベルは、その下位レベルのスキルも保有することとします。

\*スキルレベルで、対応できる課題も変わります。

\*Senior Data Scientist（業界を代表するレベル）は、一人である必要はないと考えます。一人で現実的に全て持てる多くの場合の目標点が、(full) Data Scientist（棟梁レベル）という見立てです。全体をコーディネートし、俯瞰できる人は必要ですが、加えて個別のスキルセットで秀でた人とのチームを作り、推進することも現実的には多いためです。

図 3 : データサイエンティストのスキルレベル

	ビジネスカ (business problem solving)	データサイエンスカ (data science)	データエンジニアリングカ (data engineering)
1. Senior Data Scientist 業界を代表するレベル	<ul style="list-style-type: none"> <li>組織や市場全体にインパクトを出せる。</li> <li>対象とする事業全体、産業領域における課題の切り分け、テーマ、論点の明確化ができる。</li> </ul>	<ul style="list-style-type: none"> <li>新しいアルゴリズムや分析手法の開発ができる。</li> <li>複数のパラメータやアルゴリズムの選択など、適切な分析アプローチの設定ができる。</li> </ul>	<ul style="list-style-type: none"> <li>複数のデータソースを統合したデータシステム、もしくはデータプロダクトの構築、全体最適化ができる。</li> </ul>
2. (full) Data Scientist 棟梁レベル	<ul style="list-style-type: none"> <li>分析を通じオペレーション上の革新が実現できる。</li> <li>仮説や可視化された問題がない中で（フレーミングされていなくても）、適切に問題を定義し、解き、価値を見出すことができる。</li> <li>特定の課題領域において、課題と取組のテーマを構造的に整理し、見極めるべき論点をクリアにできる。</li> <li>組織全体を見渡して、必要なデータの当たりをつけることができる。</li> </ul>	<ul style="list-style-type: none"> <li>多変量解析の概念を理解し、活用することができる。</li> <li>機械学習、自然言語、画像処理のアルゴリズムを理解し、適切に活用、問題解決することができる。</li> <li>モデルを構築できる。</li> </ul>	<ul style="list-style-type: none"> <li>分析に必要なデータフォーマット、取得蓄積仕様等を設計できる（分析のためのデータシステム設計ができる）</li> <li>問題設定に応じた新規データマート設計ができる。</li> <li>構造化データ/非構造化データを問わず、分析システムを設計できる</li> <li>構築したモデルを実装できる。</li> <li>データ分析を作ったシステムを自身で構築できる。</li> </ul>
3. Associate Data Scientist 独り立ちレベル	<ul style="list-style-type: none"> <li>仮説や既知の問題が与えられた中で、最適解・最大解を見出すことができる。</li> <li>扱っている課題領域で新規の課題を切り分け、構造化できる。</li> <li>当該プロジェクト・サービスを超えて、必要なデータの当たりをつけることができる。</li> </ul>	<ul style="list-style-type: none"> <li>SPSS/SAS/R 等が使える。指示されなくてもサンプル抽出ができる。とともに内容を確認できる。</li> <li>データクレンジング、分布、単回帰や P 値の概念を理解し、活用することができる（二次元の分析はできる）。</li> </ul>	<ul style="list-style-type: none"> <li>大規模のファイルや、データベースにアクセスし、大量の構造化データを処理することができる（一般的なスプレッドシートで処理不能な規模感への対応力）。</li> </ul>
4. Assistant Data Scientist 見習いレベル	<ul style="list-style-type: none"> <li>ビジネスにおける論理とデータの重要性を認識している。</li> <li>仮説や既知の問題が与えられた中で、必要なデータに当たりをつけて、データを用いて改善することができる。</li> <li>扱っている課題領域（例：配送の最適化）における基本的な課題の枠組みが理解できる。</li> </ul>	<ul style="list-style-type: none"> <li>基本統計量（平均、中央値など）の知識を有し、指示されればデータの抽出、グラフ作成を正しく行うことができる。</li> </ul>	<ul style="list-style-type: none"> <li>一般的なアクセス解析システムを使うことができる。</li> <li>抽出されたデータサブセットに対し、Excel や Access 等の統合環境を用い、目的に応じた処理をすることができる。</li> </ul>
Data Scientist 以前の方	<ul style="list-style-type: none"> <li>ビジネスは勘と経験だけで回すものだと思っている。</li> <li>課題を解決する際に、そもそも定量化する意識がない。</li> </ul>	<ul style="list-style-type: none"> <li>基本統計量の意味を正しく理解していない。</li> <li>指数を指数で割り算したりする。</li> <li>「平均年収」をそのまま鵜呑みにしたりする。</li> <li>グラフ・チャートの使い方が不適切。</li> </ul>	<ul style="list-style-type: none"> <li>レポートされてくる数値サマリーに目は通すが、特に記憶には残らない。</li> <li>アクセス解析システムを使っていない。</li> <li>Excel や Access は数字しか入れない。</li> </ul>



## 今後の展望と課題

当協会では、データサイエンティストのミッション、スキルセット、スキルレベルについては大枠が見えてきたと考えております。今後、スキル定義委員会としては、以下の育成、運用課題について、さらに検討を進めてまいります。

- データサイエンスの道に入る人がどこから始めたらいいか？
- どのように成長していったらいいか？
- どうやって本物のデータサイエンティストになるのか？
- どうやって人材を配置するのか？

また、社会がデータサイエンティストを育成し、データの持つ力をさらに解き放っていくためには、以下の課題にも答えていく必要があると考え、検討を進めていく予定です。

- 基本的なデータリテラシー（分析的、データ・ドリブンな思考力と基本的な知見）の社会的欠落。この新たな挑戦の幅と深さを理解できる人が少ないこと。
- 教える人と訓練するシステムが足りないこと。
- 純粋なデータサイエンスの過度の偏重とデータエンジニアリング力、ビジネス力の重要性への理解の低さ。
- データと現実のビジネスをつないで考えるアーキテクト的な人材、翻訳能力を持つ人材の欠落。
- 目的別のデータ保管、利活用のポリシー（政策）を揃えることが困難なこと。

(\*1) 明確な定義や範囲が定まっておらず、人によって思い浮かべる内容がバラバラな新語や造語、フレーズのこと。

## ■ご参考情報

### ●スキル定義委員会 委員一覧（2014年11月現在、敬称略）

- 委員長 ヤフー株式会社 CSO（チーフストラテジーオフィサー） 安宅 和人  
副委員長 株式会社電通 統合データ・ソリューションセンター データ・マネジメント部長 佐伯 諭
- 委員
- ・株式会社 ALBERT 代表取締役会長 山川 義介
  - ・株式会社 インテージ MCA 事業本部 データサイエンス部 部長 橋本 正之
  - ・株式会社 インテリジェンス キャリア Div マーケティング企画統括部 DODA 編集部 データアナリティクスグループ マネジャー 大江 信明
  - ・ウイングアーク 1 s t 株式会社 営業企画部 小山 智久
  - ・ウイングアーク 1 s t 株式会社 サポート推進部 小林 香織
  - ・SAS Institute Japan 株式会社 ソリューションコンサルティング第一本部 エンタプライズアナリティクス推進グループ 担当部長 津田 高治
  - ・損害保険ジャパン日本興亜株式会社 経営企画部 ICT 企画グループ 課長 大原 克之
  - ・損害保険ジャパン日本興亜株式会社 経営企画部 ICT 企画グループ 課長代理 上西 優子
  - ・テクノスデータサイエンス・マーケティング株式会社 エンジニアリンググループ グループ長 執行役員 池田 拓史



- ・デジタル・アドバタイジング・コンソーシアム株式会社  
プロダクト開発本部 フェロー 遠矢 行史
- ・日本アイ・ビー・エム株式会社 クライアント IT 推進事業  
成長イニシアチブ推進 シニア IT アーキテクト 北山 浩透
- ・日本サード・パーティ株式会社 エデュケーション部 常務取締役  
古川 宏幸
- ・日本サード・パーティ株式会社 エデュケーション部 執行役員  
関口 大五郎
- ・日本サード・パーティ株式会社 エデュケーション部 マネージャ  
清水 怜美
- ・日本サード・パーティ株式会社 エデュケーション部 マネージャ  
森本 良照
- ・株式会社日立インフォメーションアカデミー ビジネス研修部 部長  
津村 利幸
- ・株式会社日立インフォメーションアカデミー ビジネス研修部 技師  
大黒 健一
- ・株式会社日立インフォメーションアカデミー ビジネス研修部 技師  
田中 貴博
- ・株式会社ブレインパッド 代表取締役社長 草野 隆史
- ・株式会社ブレインパッド アナリティクスサービス本部 橋本 武彦
- ・フュージョン株式会社 代表取締役社長 佐々木 卓也
- ・三菱電機インフォメーションシステムズ株式会社 生産技術本部 技術部  
ソフトウェア技術課 課長 山足 光義
- ・三菱電機インフォメーションシステムズ株式会社  
産業・サービス事業本部 ITコンサルティング部 尾崎 隆
- ・ヤフー株式会社 システム統括本部 データソリューション本部 本部長  
小間 基裕

- 一般社団法人データサイエンティスト協会について <http://www.datascientist.or.jp/>  
データサイエンティスト協会は、新しい職種であるデータサイエンティストに必要なスキル・知識を定義し、育成のカリキュラム作成、評価制度の構築など、高度 IT 人材の育成と業界の健全な発展への貢献、啓蒙活動を行っています。また、所属を超えてデータ分析に関わる人材が開かれた環境で交流や議論をし、自由に情報共有や意見発信ができる場を提供しています。2014年12月現在、28社2団体の法人会員と約1700名の一般（個人）会員が参画しています。代表理事：草野 隆史（株式会社ブレインパッド 代表取締役社長）、所在地：東京都港区、設立：2013年5月。

以上



<お問い合わせ先>

●本件に関するお問い合わせ

一般社団法人データサイエンティスト協会  
事務局

TEL: 03-6721-9001 e-mail: [info@datascientist.or.jp](mailto:info@datascientist.or.jp)

●報道関係の方からのお問い合わせ

一般社団法人データサイエンティスト協会  
事務局 広報担当

TEL: 03-6721-9001 e-mail: [pr@datascientist.or.jp](mailto:pr@datascientist.or.jp)

\*本プレスリリースに記載されている会社名・商品名は、それぞれの権利者の商標または登録商標  
です。

\*本プレスリリースに掲載されている情報は、発表日現在の情報です。