特集号
招待論文

データサイエンティスト育成と人材利活用のベスト・プラクティス

丸山 宏[†1]　神谷 直樹[†1]　樋口 知之[†1]　竹村 彰通[†2]　大西 立顕[†2]

[†1]統計数理研究所　[†2]東京大学

ビッグデータ利活用の主要なボトルネックの 1 つが人材不足だといわれている．我が国におけるデータサイエンティストの育成を加速するため．我々は文部科学省委託事業「データサイエンティスト育成ネットワークの形成」を 2013 年に開始した．本稿では，この事業を推進する上で見聞きした，さまざまなデータサイエンティスト育成の取り組みと，データサイエンティストを実際に組織の中で活かしていく取り組みについて，ベスト・プラクティスとして紹介する．

1. はじめに

本データ分析を行う上で，特に我が国において最大の問題の1つとして指摘されているのが，人材の不足である．Mckinsey Global InstituteのBig Dataに関するレポート[1]によれば，“deep analytical talent”が2018年までに米国で14万人から19万人不足するという．日本経済新聞は2013年7月17日付けの記事で，米調査会社ガートナーの数字として，将来的に国内ではデータサイエンティストが約25万人不足すると報じた．

我が国におけるデータ分析専門家の持続的な育成と効果的な活用を目指して，我々は文部科学省委託事業「データサイエンティスト育成ネットワークの形成」を2013年7月に開始した．この事業では，我が国におけるデータ分析人材の「あるべき姿」を明らかにするとともに，データ分析人材育成に熱意を持つ教育機関と，データ分析をビジネスに活かしたい企業・組織を広くネットワークし，それらの間で知識・経験を共有することで，多くのデータ分析人材が育成され，有効に活用されることを狙っている．

データサイエンティストに必要なスキルについてはコンセンサスが固まりつつあるが，そのスキルをどのように育成しどのように活用するかについては，いまだ手探りの状態である．本稿では，我々の事業を通して得られたさまざまな事例をもとに，データサイエンティスト育成とそれら人材利活用のベスト・プラクティスについて議論する．

2. データサイエンティストとは

2.1 日米のデータサイエンティスト像

Harvard Business Reviewの記事において，トーマス・ダベンポート（Thomas H. Davenport）は，「データ・サイエンティストとは，高度な数学的素養を持ち，プログラミングに長けていて，好奇心旺盛で企業の経営にも興味を持つ，スーパースターである」と述べている[2]．

だが，必ずしもすべてのデータサイエンティストが，ダベンポートの言うようなスーパースターではない．我々の事業においては，我が国におけるデータサイエンティスト像を調べるために，統計検定合格者313名に対するアンケート調査と，20名のデータサイエンティストに対する聞き取り調査を行い，我が国のデータサイエンティスト像について，以下のように分析した[3]．

1. データ分析人材のバックグラウンドはさまざまであること．現在，データサイエンティストとして活躍されている方々の出身は，理系も文系もあり，さまざまである．
2. データ分析は全人的な能力であること．データ分析人材の能力は，データ分析だけでなく，ビジネスの問題を発見してデータ分析の問題に落とし込む力，またデータ分析の結果を現場に適用してビジネスの成果につなげるためにコミュニケーション力も要求される．
3. データ分析は，個人の能力というよりは，組織の能力であること．「データサイエンティスト」という個人がデータ分析のすべての局面を担当するのではなく，チームとしてそれぞれの得意分野を担当している．また，社員全員が基本的なデータ分析スキル

を持つように教育している会社もある．

4. 発注側のリテラシーが大切なこと．どんなに良いデータ分析結果を持っていても，経営者がそれに基づいて意思決定できなければ役に立たない．

2.2　必要なスキル・セット

我が国でデータサイエンティストの育成や社会に対する普及啓蒙活動を行っているデータサイエンティスト協会は，およそ1年にわたる集中的な議論の結果，2014年12月に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」を発表した☆1．それによれば，データサイエンティストとは，「人間を数字入力や情報処理の作業から解放するプロフェッショナル人材」であり，そのミッションは「データの持つ力を解き放つ」ことである，としている．

「データの持つ力を解き放つ」ためには，データを分析するだけでは足りないことに注意してほしい．ビッグデータを解析して「何か面白いことが見つかった」だけでは，データサイエンティストの仕事は半分しか終わっていない．分析の結果から，既存のビジネスプロセスを変え，その結果新しい価値を生んで初めて「データの持つ力を解き放った」ことになるのである．逆に，データの持つ力を解き放つことができるのであれば，必ずしもデータの分析は必須ではない．分析しなくても，データを整理・統合しうまく見せるだけで，データの力をビジネスに結びつけられる場合もある．このように，データサイエンティストとは，単にデータの分析にとどまらない，新たな種類のプロフェッショナルということができる．

データサイエンティストに必要なスキルはどのようなものであろうか．前述のデータサイエンティストのレポートにおいては，データサイエンティストのスキルを図1のように3つに分類している．

ビジネス力とは，課題背景を理解した上で，ビジネス課題を整理し，解決する能力である．データサイエンティストは，データの持つ力を解き放たなければならない．そのためにはビジネスを理解し，データをどのように価値創造につなげられるかを見通せる人材でなければならない．ビジネス力のもう1つの重要な側面は，コミュニケーション能力である．どんなに良いデータ分析結果が得られても，意思決定者が理解できなければ使ってもらえない．また，使ってもらえたとしてもデータ分析の結果は不確定要素が多く，うまくいかないときのリスクを正しく理解してもらえないこともある．したがって，データ分析結果を正しく意思決定者に理解してもらうためには，コミュニケーションに相応のスキルが必要なのである．

データサイエンティストに求められるスキルセットの2つ目はデータサイエンス力である．これは，統計学，情報処理，人工知能などデータ分析に必要な手法を理解し，使う能力である．統計的仮説検定や回帰分析など伝統的な統計手法だけでなく，機械学習など人工知能の分野で新たに開発された手法などについても使えるようにしておかなければならない．また，それらを効率よく実装するためのアルゴリズムや，プログラミングモデルについても，理解とともに実践が大切である．特に，ビッグデータの分析には，並列計算など計算機アーキテクチャやネットワークの構成など，計算資源の特質を熟知した上で実用的な実装を行わなければならない．これらのテクノロジは日進月歩であり，努力してフォローしていなければならないだろう．

データサイエンティストに求められるスキルセットの3つ目はデータエンジニアリング力と呼ばれる．データサイエンスを意味のある形で使えるようにシステムを設計，実装，運用する能力であり，ある意味これが一番難しい．ある程度以上の複雑なシステム構築を繰り返して身に付くスキルであり，体系的に学べる性質のものではないからである．

データサイエンティストになるには，上記の3つのスキルセットをすべて満遍なく身に付けなければならないのだろうか．このスキル定義を行ったデータサイエンティスト協会スキル定義委員会の委員長である安宅氏は，どれか得意分野があってよいが，それぞれのスキルセットについて最低限の知識・経験は持っていてほしい，と述べている．我々の調査からもいえることは，お互いにコミュニケーションのできる最低限の共通スキルを持った上で，ビジネス力の強い人，データサイエンス力の強い人，データエンジニアリング力の強い人をチームとし

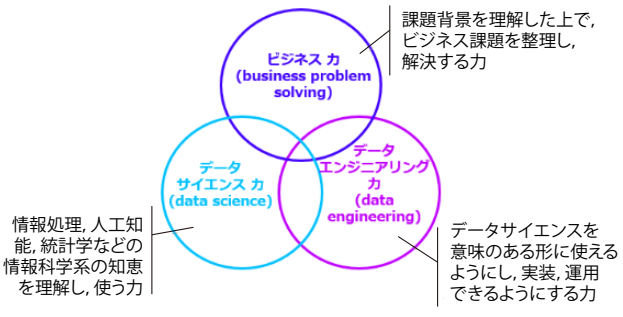

図1　データサイエンティストに求められるスキル

☆1　http://prtimes.jp/main/html/rd/p/000000005.000007312.html

て組み合わせて，組織能力としてデータ分析の力を発揮するのが，多くの会社がやっていることのようである．

3. 育成のベスト・プラクティス

前章で議論したデータサイエンティストのスキルセットを身に付けるには，どのような育成を行えばよいのだろうか．以下，カリキュラム，実習･インターンシップ，実習用データ，ツールに関して，現在行われているベスト・プラクティスを紹介する．

3.1 カリキュラム

データサイエンティスト育成のカリキュラムに関しては，前述の3つの領域を満遍なくカバーするコースは我が国においてはまだないようである．一番近いのは慶應義塾大学が行っている「スキルと実践を重視したビッグデータ人材育成プログラム」☆2であり，データサイエンス力とデータエンジニアリング力について，PBL（Project-Based Learning）による実践的な教育を含んだカリキュラムを策定している．

データサイエンス力のうち，伝統的な統計分野に関しては，「統計学の各分野における教育課程編成上の参照基準」[4]が参考になる．また，この参照基準では大きく取り上げられていないが，データに基づくビジネス課題の多くが最適化問題として定式化されることもあり☆3，さまざまな最適化手法についてもデータサイエンス力の主要なスキルとすべきであろう．

また，データサイエンス力のもう1つの柱である情報学についても参照基準[5]があり，こちらも参考になる．この中で，特にジェネリック・スキルとされている「モデル化・形式化・抽象化を行う能力」については，統計分野における考え方と通じるところも，若干ニュアンスの異なる点もあり，これらの類似点・相違点を俯瞰した形で理解できると望ましい．

カリキュラムの一例として，我々が作成したデータサイエンティスト・クラッシュコースの内容を図2に示す．このコースはおよそ20分のユニット8個からなり，現場で現れる主要なデータ分析手法のキーワードが一通り現れるように設計されている．それらの詳細については，必要になった時点で各ユニットの最後についている参考文献を参照すればよい．なお，このコースは無料ビデオコンテンツとして，ネット上で視聴できる．

また，大学・民間を問わず多くのデータサイエンティスト育成コースが開講されている．およそ180のコースについて一覧を作り，事業のWebサイトで公開している．

3.2 実習・インターンシップ

データサイエンティストのスキルは座学だけでは身に付かない．そのため，実習あるいはインターンシップなどを通して，実際のデータを分析してみる経験が必要である．我々の事業でも，インターンシッププログラムを通して，学生を企業のデータ分析の現場へ送り込み，育成につなげている．我々のインターンシッププログラムで学生が学んだことは，主に次のような点である．

- データ分析は予定通りいかない．期待していた結果が出ないことがほとんどであり，まったく予期しない結果が出ることもある．
- データ分析においては，前処理に大半の時間が取られる．
- チームで行う場合，それぞれのスキルを活かしたチームワークが大切である．

いずれも実習を通して初めて理解できることであり，インターンシップ・プログラムの効果が出ているといえよう．

なお，企業によってインターンシップ期間の組み方はさまざまであり，3～4週間でデータ分析からビジネス改善提案までを求めるものもあれば，1週間程度で特定の目的変数の最適化を求めるものもある．実習生にとってより効果があるのは，データ分析プロセスの全体像を把握できる前者のタイプであろう．

3.3 実習用のデータ

企業におけるOJTでは実際のデータを利用してデータ分析の訓練を行うことになるが，大学等における実習では，データの入手が問題になる．関連事業である慶應義

0. コース概要
1. データサイエンティストとは－ブレインパッド佐藤部長
2. データ解析基礎－統数研馬場特命教授
3. データ可視化とツール－統数研中野教授
4. 統計モデリングと機械学習－統数研松井教授
5. 統計的時系列モデリング－統数研川崎准教授
6. 最適化－統数研伊藤教授
7. データ分析と意思決定－統数研椿教授
8. データ分析の知的財産－統数研丸山教授

図2 データサイエンティスト・クラッシュコースの内容

☆2 http://bd.comp.ae.keio.ac.jp/
☆3 IBM社では，データ分析チームをBAO（Business Analytics & Optimization）と呼び，最適化の重要性を強調している．

塾大学の「スキルと実践を重視したビッグデータ人材育成プログラム」や，「分野・地域を越えた実践的情報教育協働ネットワーク」事業においても，教材用のデータ入手の困難さが指摘されている．個人情報保護意識の過剰な高まりのために，データ所有者がデータの流出に過敏になっていることと，データの重要性が認識されるにつれ，データのビジネス価値を保護したいという意図が，その背景にあると思われる．

一方，前出のシリコンバレーにおけるInsight Data-science Fellows Programでは，実習生は「データ・プロダクト」と呼ばれる実働するシステムを構築することが求められるが，ここで使われるデータはすべて，Web上で得られるデータである．たとえばある実習生が作った“CouchTube.net”というサービスは，YouTubeにあるテレビドラマ映像が，どれが第何シーズンの第何話かを分析し，それをデータベース化して検索可能にするというものである．ここでは，TheTVDV.comというWebサイトで提供されている，テレビ番組に関するデータベースと，YouTubeから得られるそれぞれの映像のタイトル，長さ，説明，他ユーザからの評価やコメントなどを総合的に分析してしている．このように，Web上から得られるデータだけでも，工夫によってはデータ分析の教材となる．

政府・自治体等が公開しているオープンデータも，教育用の教材として使える．2014年に行われた第2回データビジネス創造コンテストでは，鯖江市[☆4]や流山市[☆5]などが提供しているオープンデータを利用した分析が行われた．日本の政府が公開しているオープンデータとしては，政府統計の総合窓口としてe-statがあり，各省庁の統計データ，たとえば人口動態や家計調査，経済指標などがCSV形式でダウンロードできる．また，政府のオープンデータの窓口data.go.jpからは，統計情報だけでなく，白書などのさまざまな文書も手に入れることができる．

政府のオープンデータは基本的には統計データであり，いわゆる個票データ（個人，個々の世帯，あるいは個別の企業等に関するデータ）は含まれていない．独立行政法人統計センターでは，高等教育・学術研究向けに匿名化された個票データの提供を行っていて[☆6]，指定された環境の中で国勢調査等の匿名化された個票データを分析することもできる．より手軽には，同センターが提供している家計調査の擬似ミクロデータがあり[☆7]，およそ32,000世帯の197にわたる項目についてのデータを利用することができる．ある程度のサイズのあるデータであり，実習用としては役立つと思われる．

事前にデータを指定した上で「このデータを分析してビジネス改善提案をしなさい」という実習を行うのは，分析手法の適用について学ぶのには役立つ．一方，実際のビジネス現場では，ビジネス課題が先にあり，それを解決するためのデータを自分で探してこなければならないこともある．第2回ビジネスデータ創造コンテスト[☆8]で総合2位になったチームは，課題として与えられた鯖江市のオープンデータに加えて，自分たちの発案でインターネット上で手に入るデータ（Foursquareのデータ）を組み合わせることで，新しい価値の提供に成功した．実習では，「手に入るデータは何を使っても良い」ということにすれば，実習の幅が大きく広がるといえよう．

3.4　ツールの利用

データ分析においてはさまざまなツールが使われる．

- **統計解析ツール**

  大学等で教育用の統計解析ツールとして最もよく使われるのがRである．Rは無料であり，また多くの統計分析アルゴリズムや，可視化ツールが実装されている．企業においては商用の，SAS，SPSS，MatLABなどが使われることが多い．

  いわゆる統計解析ツールではないが，エントリーレベルの教育や実務では，Excelが使われることもある．

- **データベース**

  企業のインターンシップで一番強く要求されるのが，SQLに関する知識である．いわゆる「ビッグデータ」はSQLデータベースでは処理できない規模のデータ，とされることが多いが，実際に企業等で分析の対象となるデータは，企業のデータウェアハウスにSQLテーブルとして格納されているものであることを反映していると考えられる．データベース管理システムとしては，無料のMySQLやPostgress，商用のOracleなどがあるが，どれもアクセスはSQLなので大きな違いはない．また，SQLは，最初のデータの抽出に使われるのみで，その後はCSVファイルに落として前処理を行うことが多いようである．

☆4 http://data.city.sabae.lg.jp/
☆5 http://www.city.nagareyama.chiba.jp/10763/
☆6 http://www.nstac.go.jp/services/archives.html
☆7 http://www.nstac.go.jp/services/giji-microdata.html
☆8 http://dmc-lab.sfc.keio.ac.jp/dig/

- **プログラミング言語**
  Rでも簡単なプログラミングはできるが，主にインタラクティブに探索的な分析に用いられることを想定して設計されているため，ETLなどビジネスプロセスに組み込む目的には向かない．また，分析も全データをメモリに読み込むことが前提なので，主記憶に入らないサイズのビッグデータの前処理には不適切である．このため，Javaなどの汎用言語，あるいはPythonなどのスクリプト言語の利用が欠かせない．
- **並列分散ミドルウェア**
  ビッグデータの分析にはHadoopやMahoutなどが必要だといわれる．しかし，実習で使われるデータセットについて，これらの分散ミドルウェアが必要になる場面はそれほど多くない．また，必要となったとしても，分析段階ではなく，特徴抽出など前処理の段階で使われるのが一般的である．分散ミドルウェアが効果を発揮するのはスケールする計算であるが，そのためにはふんだんな計算資源が必要であることもあり，実習に組み込むには慎重に行ったほうがよい．
- **クラウドコンピューティング**
  各大学が個別にITインフラを抱える非効率性への反省から，教育においてもクラウドコンピューティングを用いるのが一般的になりつつあるが，一方で，使用時間によって課金される商用のクラウドコンピューティングサービスにおいては，特定の学生がクォータ（計算時間，容量などの割り当て資源）を使いきってしまうことがあり調整が難しい，という指摘もある．

これらのツールにすべて習熟しなければならない，というものではない．実習においても，また実際の企業のデータ分析においても，データ分析はチームで行うのが一般的であり，前処理を行う者，分析や可視化を行う者，システム構築を行う者など分担するのが当然だからである．

特にプログラミング言語やミドルウェアなどIT系のツールは栄枯盛衰が激しく，あるツールに習熟してもその知識はすぐに陳腐化してしまうおそれがある．また，就職先，あるいは客先で使われるツールは自分がよく知っているツールとは別のものである可能性もある．したがって，教育においてはあまり深くツールにコミットせず，あくまでも使用体験を積む程度で良いと考えられる．

4. 人材利活用のベスト・プラクティス

組織においてデータ分析を行うためには，データ分析スキルを持つ人材，すなわちデータサイエンティストを調達しなければならない．我々が事業の実施の中で見聞きした範囲では，

- 組織内部で育成する
- 外部から採用する
- 外部のサービスを利用する
- クラウドソーシングを利用する

という戦略がありそうである．それらについて，個々の事例を紹介する．組織名は伏せてある．

4.1 組織内部での育成・転用

4.1.1 OJTによる育成

A社では，社内のさまざまなデータ分析を行うために，情報システム部の中におよそ10名程度のデータ分析専門チームがいる．ここでは現場力を重視していて，現場の目線で使ってもらえるデータ分析を行うにはどうしたらよいかを常に考えている．人材はたとえば新卒社員を一からOJTで育成する．育成の内容は主にビジネス力，データサイエンス力であり，データエンジニアリング力については，IT関連子会社とチームを組む．

4.1.2 社内研究部門からの転用

総合ITベンダのB社では，社外に対してデータ分析サービスを提供している．このデータ分析チームの立ち上げにおいては，社内の研究部門のチームの1つをまるごとサービス部門に異動した．研究者からサービス・プロフェッショナルになった後も，研究所との人的交流は活発で，人材育成のためにサービス部門と研究部門の間で積極的なローテーションを行っている．

4.1.3 社員全員の底上げ

データサイエンティストという専門家を育てるのではなく，社員全員のデータリテラシーを向上させる取り組みをしている企業もある．食品流通のC社では，社長以下全員がExcelによる基本的なデータ分析を習得している．これによって，社内に常にデータに基づく意思決定を行う文化が定着している．

4.2 外部からの採用

4.2.1 エージェント等の利用

シリコンバレーのソーシャルネット大手D社では，200 ～ 300名のデータサイエンティストがいて，全社の各部署に分散配置されている．彼らはそれぞれの部署の

「プロダクト」（サービスを構成するコンポーネント）の機能・性能をデータ分析を用いて向上させている．シリコンバレーにおいては，データサイエンティストは即戦力であり，データサイエンス力とともに，データ分析の結果をすぐにシステム構築につなげられるデータエンジニアリング力，プログラミング力を同時に求められる．この即戦力を得るために，必須スキルを細かく指定して，求人エージェントを使ったり個人のネットワークを使って積極的に採用を行っている．

4.2.2　学会・インターンシップの利用

国内ソフトウェアベンチャーのE社では，トップノッチのデータサイエンティストを採用するために，社員が著名な国際会議に参加し，優秀な学生・研究者に声をかけている．また，毎年インターンシップを実施し，有望な学生を採用している．

4.2.3　コンテストの利用

トップノッチのデータサイエンティストを探すまた別の方法として，コンテストの利用がある．金融機関F社では，Kaggle.com などのコンテスト型のクラウドソーシング（発注者はデータ分析問題を提示し，最も精度の高い解を出したものが受注する）を用いて，優秀なデータサイエンティストを発掘し，採用につなげている．

4.3　外部サービスの利用

4.3.1　コンサルティングサービスの利用

社内にデータサイエンティストを置いたとしても，その人のキャリアパスを描きにくい，あるいはデータ分析をどのようにビジネスに活かしていくかのビジョンがまだ明確でない，などの理由で，社内にデータサイエンティストを抱えることに二の足を踏む企業も多い．データ分析に力を入れ始めているG自治体の場合，最初はデータサイエンティストの採用に向けて動いたが，最終的にはコンサルティングサービスの利用に決定した．コンサルティング会社が持つ幅広いデータ分析ノウハウを期待してのことである．この取り組みがうまくいっている理由の1つは，発注者側に戦略コンサルティング経験者がいることで，このためにサービス供給側（コンサルティング会社）とサービス発注側（自治体）とのコミュニケーションがスムースなのであろう．

4.3.2　大学等の利用

研究対象としてもビッグデータが注目されているが，大学の研究者には，現場に直結したデータが得られないことも多い．このため，データを持っているが分析のスキルのない企業と，スキルはあるがデータのない大学研究者とのコラボレーションはうまくいくことがある．建設機械メーカのH社は大学との共同研究を通して社内データの解析を行っている．生のデータは大学へは開示せず，分析自体は社内で行って結果のみを開示し，それに対するアドバイスをもらう，という形で共同研究を進めている．このようにして，データの流出などのリスクを最小限におさえている．コンサルタントを雇うのに比べて大学との共同研究は安くつくが，その分目的意識が希薄になりがちな点には注意が必要である（「このデータを分析して何か面白いことを見つけてください」など）．

4.4　クラウドソーシングの利用

コンテスト形式で予測モデルのクラウドソーシングを行うWebサイト kaggle.com には情報科学，統計学，経済学，数学などの分野から全世界で約95,000人のデータサイエンティストが登録していて，多くの企業がビジネスに直結するデータ分析課題を投げかけている．GE社は，気象状況や航空路の制約などの条件下で，最適な航空ルート，高度，速度を計算する課題に総計50万ドルの賞金を用意した．

我が国でもクラウドソーシングによるデータ分析が可能かどうか，予備実験を行った[6]．具体的には，クラウドソーシングサイトであるクラウドワークス上で架空のアンケート結果の分析を，応募してきた10名のワーカにそれぞれ独立に発注した．その結果，応募時にスキル記述に嘘はなかったが，分析結果の品質は大きく分かれることが分かった．クラウドソーシングの利用には，kaggleのようなコンテスト形式のものを使うか（「最も高い予測精度」という結果によって最善のワーカを選択できる），あるいはワーカに対して，何らかのスキル認証が必要だと考えられる．データサイエンティスト協会等における，スキル認証の活動の進展が望まれる．

4.5　プロジェクトチームの編成と運用

2.2節で述べたように，データサイエンティストのスキルは個人によって得意分野と不得意分野がある．したがって，データ分析を行うには，必要なスキルセットがカバーできるチーム編成を行う必要がある．本特集の論文[7]においては，ビジネス・バリュー，アナリティクス，データ整備の3活動領域にわけてスキルを組み合わせることを提唱している．これは，2.2節におけるビジネス力，データサイエンス力，データエンジニアリング力に対応している．

データ分析プロジェクトは，やってみないと結果が出

るかどうか分からない，探索的な側面が大きい．プロジェクトが進むうちに，最初は想定していなかったスキルが必要になってくる場合もある．したがって，プロジェクトチームの編成は柔軟に行えるよう，また必要に応じて他プロジェクトの専門家が応援に入れるように手当しておく必要がある．データ分析スキルをコンサルティング会社など外部から調達する場合には，プロジェクトの局面に応じてそれぞれの分野専門家が対応できるというメリットもある．

4.6 継続的な研鑽とコミュニティの利用

機械学習のアルゴリズムや，分散処理のプログラミングモデルなど，データサイエンティストが使う手法は日進月歩であり，常に新しい技術を習得し続ける必要がある．

社内の複数の部門に多くのデータサイエンティストをかかえるI社においては，社内の技術コミュニティがあり，このコミュニティの中で最新の技術動向や適用事例の情報交換をすることで，継続的な研鑽に役立てている．また，データマイニング+WEB@東京[☆9]，データサイエンティスト協会木曜勉強会[☆10]など，組織をまたがった勉強会が多く開催されていて，最新の動向に触れることができる．本会のビッグデータ活用実務フォーラム[☆11]もこのような勉強会の1つであり，無料で参加できる．

今後，データサイエンティスト協会が検討しているデータサイエンティスト資格認定制度ができれば，この資格を取得することも人材の継続的な育成に寄与することとなろう．

5. おわりに

データサイエンティストに対する効果的な育成と．長期的なキャリアパスを見据えた組織における利活用はまだ緒についたばかりであり，まだ十分に体系化されているとはいえない．しかし，本稿で述べたようなベスト・プラクティスを収集し共有していくことで，我が国において社会の需要と期待に応えるデータサイエンティストの育成が進み，また彼らが意義のあるキャリアパスを歩んでいくことを期待したい．

☆9 https://groups.google.com/forum/#!forum/webmining-tokyo
☆10 http://www.datascientist.or.jp/activity/thu.html
☆11 https://www.ipsj.or.jp/it-forum/big_data.html

参考文献

1) McKinsey Global Institute: Big Data: The Next Frontier for Innovation, Competition, and Productivity (2011). http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation
2) Davenport, T. H. and Patil, D. J.: Data Scientist: The Sexiest Job of the 21st Century, Harvard Business Review, pp.70-76, (Oct. 2012). 日本版 データサイエンティストほど素敵な仕事はない，DIAMOND ハーバード・ビジネス・レビュー，pp.84-95，2013 年 2 月号 .
3) 丸山　宏 他：我が国におけるデータ分析人材の育成と活用，第 5 回横幹連合総合シンポジウム (2014).
4) 統計学の各分野における教育課程編成上の参照基準, http://www.jfssa.jp/ReferenceStandard2.pdf (2014).
5) 萩谷昌己：情報学を定義する－情報学分野の参照基準，情報処理, Vol.55, No.7 (July 2014).
6) 井川甲作 他：クラウドソーシングにおけるデータサイエンティスト活用に関する初期的調査，第 16 回日本テレワーク学会研究発表大会 (2014).
7) 山田　敦：アナリティクスで継続して成果を生み出す仕組み，情報処理学会デジタルプラクティス，Vol.6, No.3 (July 2015).

丸山　宏（正会員）hm2@ism.ac.jp
1983 年東京工業大学情報科学専攻修士課程修了後，日本アイ・ビー・エム（株）入社．東京基礎研究所で自然言語処理，XML，セキュリティ等の研究．2006 ～ 2009 年同研究所所長．2011 年より統計数理研究所教授．工学博士．

神谷 直樹（非会員）nkamiya@ism.ac.jp
2003 年立教大学大学院文学研究科心理学専攻博士課程単位取得後退学．国立長寿医療研究センター研究員を経て，統計数理研究所統計思考院特任研究員．専門は計量心理学，行動分析学．博士（文学）早稲田大学．

樋口 知之（非会員）higuchi@ism.ac.jp
1989 年東京大学大学院理学系研究科地球物理学専攻博士課程修了後，文部省統計数理研究所に着任．専門はベイジアンモデリング．2011 年より統計数理研究所所長，ならびに情報・システム研究機構理事．理学博士．

竹村 彰通（非会員）takemura@stat.t.u-tokyo.ac.jp
1982 年スタンフォード大学統計学科 Ph.D.．スタンフォード大学客員助教授等を経て東京大学大学院情報理工学系研究科教授．数理統計学の理論，特に多変量解析の理論を中心に研究を行っている．

大西 立顕（非会員）ohnishi.takaaki@i.u-tokyo.ac.jp
2004 年東京大学大学院新領域創成科学研究科複雑理工学専攻博士課程修了．東京大学大学院情報理工学系研究科ソーシャル ICT 研究センター准教授．ビッグデータを実証科学の視点から研究している．博士（科学）．

採録決定：2015 年 3 月 16 日
編集担当：黒橋禎夫（京都大学）