

横断する中核部隊として位置づけることが求められる。その部署を束ねる人材は企業における CAO<sup>3</sup>のような棟梁レベル（後述、3.4 節参照）が望ましいが、町村レベルだと、独立レベル（3.3 節参照）のデータサイエンティストが外的機関（組織）に指導を仰ぎながら実務を担当することも有効であろう。一方で、ビッグデータビジネスや、データサイエンティスト人材育成にかかわるイベントや機会が東京に集中しており、地方自治体は自ら人材育成することが大変困難となっている。従って、全国レベルの底上げの必要性は明らかであり、例えば、全国的に一定数の拠点を設置する取組を進めることも考えられる。

### 3. データサイエンス人材育成のあるべき姿と実現に向けた仮説

産業界とアカデミアからの要請をまとめると、我が国においてデータサイエンティスト育成のあるべき姿は、図2のようになると考えられる。スキルレベルとしては、対数スケールで6段階としたが、それと前出のデータサイエンティスト協会スキル定義委員会が発表した4つのスキルレベルとの対応を図に示した。これらの人材が、大学や企業において継

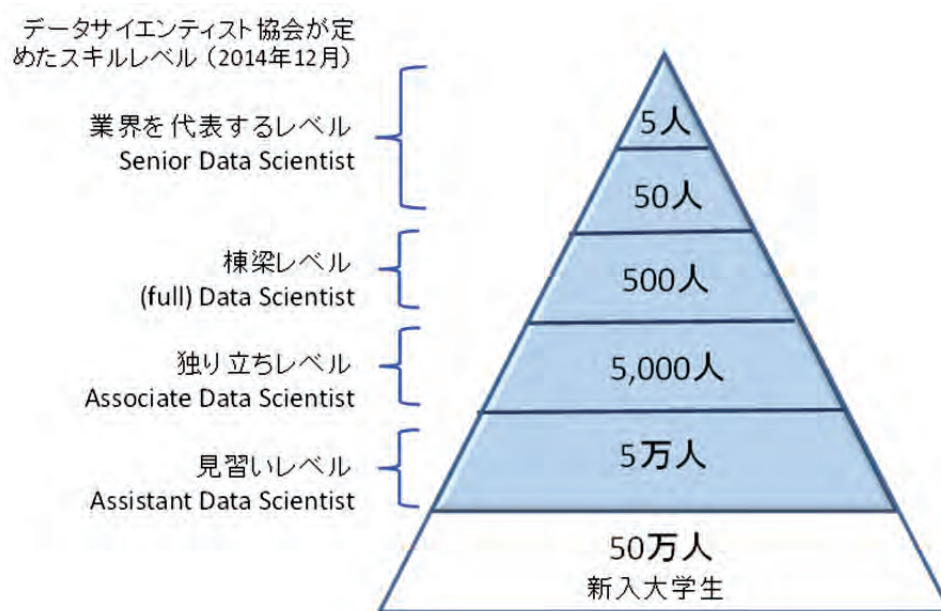


図2. 育成レベルと、毎年の育成目標人数

<sup>3</sup> 企業において複数の応用分野を俯瞰的にマネージすることができ、データサイエンスの観点から全体最適な戦略を策定し実行できるようなリーダーシップを持った人材を CAO（Chief Analytics Officer）と呼ぶこともある。

続的に育成され、かつ効果的に利活用されている姿を、あるべき姿と考える。なお、毎年の育成目標数は概数であり、桁としてこの程度必要であると解釈されたい<sup>4</sup>。

### 3.1. データリテラシー

高校から大学学部レベルの素養（リテラシー）で、文系・理系を問わず全ての学生が持つべきであり、ベースになる統計的概念、データに基づく思考や問題解決の基礎概念、ITリテラシー、データに関する研究倫理などの内容を含む。我が国の大学の入学者数はおおよそ60万人であり、目標育成数はほぼその全数と考える。

### 3.2. 見習い(基礎能力)レベル

全てのデータサイエンティスト（実務家・研究者を問わず）が持つべき共通能力。この能力は同時に、多くのミドルクラスマネジメントにも必須である。理系の修士は全て、また文系でも経済・経営などの社会科学系や言語学や心理学等の専攻ではすべからず身につけるべき能力であり、この能力を持つ者は、適切な指導の下でビッグデータ利活用プロジェクトの一部分を担当できる。データサイエンティスト協会のスキル定義では、見習いレベルに相当する。我が国の理系の修士入学者数は、年間およそ5万人であり、これが目標育成数の1つの目安となる。

我が国には、データサイエンスに関する学部・学科は存在しないが、そのような学部・学科が今後設置されたならば、学部生が卒業までの3年・4年次に習得するレベルがこの見習いレベルになる。

### 3.3. 独り立ちレベル

データサイエンティストとしての専門能力、すなわちビジネス、データサイエンス、データエンジニアリングのいずれかの分野で専門的な能力を持ち、自らのイニシアチブで高度なデータ分析・問題解決能力を発揮する。実務経験が必須であり、大学においては修士ま

---

<sup>4</sup>参考としてデータサイエンティストの必要数に関する推測(公表されたもの)を下記に記す。

- ・ McKinsey Global Institute Report on Bigdata (2011): "deep analytical talent"が2018年までにUSで14万人から19万人不足。また、"data-savvy manager and analyst"については150万人不足。
- ・ ガートナー(2011):「日本ではビッグデータ関連雇用が36万5000人分増える見込みだが、実際に雇用条件を満たせる人材は11万人程度しかない」
- ・ 日経(2013.7.7):「国内ではデータサイエンティストが約25万人不足する」

たは博士課程での具体的な PBL(Project-Based Learning)の経験が求められる。データサイエンティスト協会のスキル定義における独り立ちレベルに相当する。データサイエンスを主とする大学院の専攻においては、修士課程修了時までにはできれば習得しておいて欲しいレベルが独り立ちレベルになる。我が国における資本金 10 億円以上の会社はおよそ 6,000 社であり、その企業の中で毎年 1 名程度ずつ採用を検討するとすれば、産業界向けだけでも 5,000 名程度が目標育成数の目安となる。

### 3.4. 棟梁レベル

データサイエンティストのチームを率いて、組織におけるビッグデータ利活用を先導できる能力をもった人。複数の応用分野を俯瞰的にマネージすることができ、データサイエンスの観点から全体最適の戦略を策定し実行するリーダーシップが求められる<sup>3</sup>。主に実務を通して育成される能力であるが、専門知識については大学院における社会人コースを用いて集中的に育成(例:情報学研究所における高度 IT 人材育成「トップエスイー」プログラム)することも考えられる。毎年 5,000 名の「独り立ちレベル」を指導統括するこのようなリーダーは、組織論の観点から見れば、「独り立ちレベル」6~15 人につき 1 人程度必要と考えることができる<sup>5</sup>。したがって、目標育成数は、「独り立ちレベル」より 1 桁少ない年間 500 名程度であろう。

### 3.5. 指導的データサイエンティスト

学術においてはデータサイエンスの最先端を切り開くワールドクラスの研究者・開発者として指導的な能力を発揮する者、また産業界においては、業界におけるビッグデータ・データサイエンスに基づくイノベーションを牽引できる者。年間数名から数十名が現実的な目標育成数と考えられる。なお図での世界的トップタレントとは、グーグル CEO のラリー・ページ級の人物を指す。このようなトップタレントを系統的に育成することは困難であるが、棟梁レベルを対象に、そのための素地をつくる取組が必要である。

### 3.6. 「中抜き」仮説

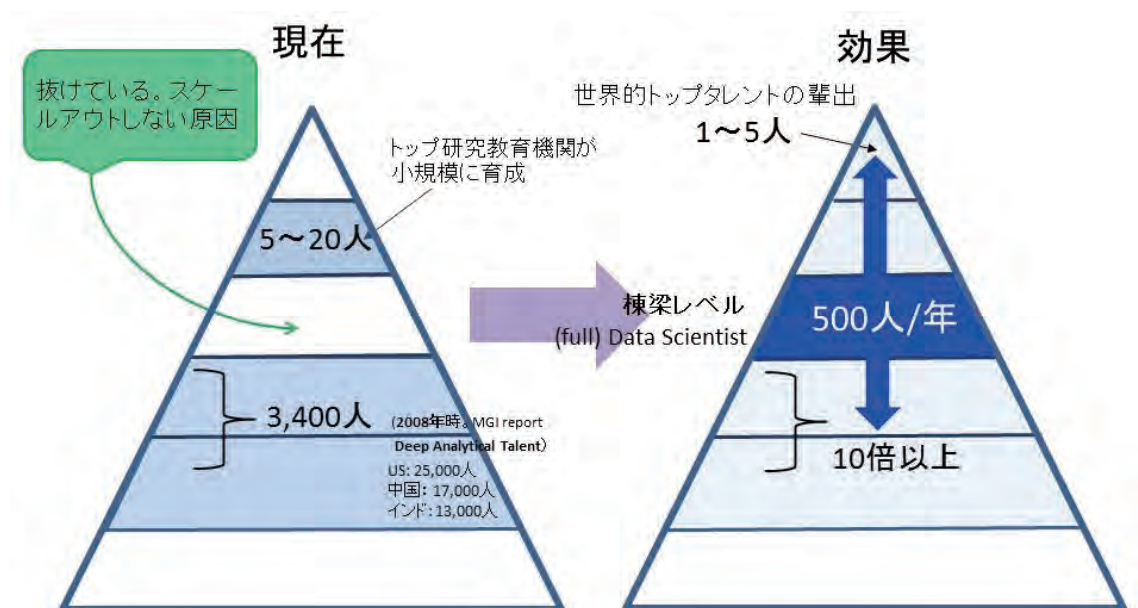
図 2 において、懇談会が最も深刻な問題と考えるのが、「棟梁レベル」の人材が育っていないことである。

---

<sup>5</sup> Span of Control (1 人の管理職がマネージする直接の部下の数) は業種や組織によって異なるが、平均 6-7 人と考えられている。GE 社のガイドラインは、10-15 人とされている。

その大きな原因は、棟梁レベルの人材が活躍できる場がこれまで我が国になかったことである。アカデミアにおいては最先端科学技術に関わる研究プロジェクト推進におけるデータサイエンスの重要性は、かねてからある程度認識されていたと思われるが、これまでの研究プロジェクトでは、データサイエンティストは分野を専門とするプロジェクトリーダーの指示に従いつつ研究開発を行うことが通常であり、棟梁レベルのデータサイエンティストがリーダーシップを発揮して、複数の分野の専門家とともに革新的な研究開発をすすめる機会が十分に存在しなかった。同様の問題は、部門縦割り主義の弊害が叫ばれている産業界においても多く見られ、データサイエンスを横串としたイノベーションを妨げる壁となっている。この状況は、企業におけるビッグデータ利活用に責任を持つ CAO (Chief Analytics Officer) が急速に普及しつつある米国<sup>6</sup>とは大きく異なる。

我が国におけるこの「中抜き」の状況をすみやかに解消しなければならない。棟梁レベルの人材が年 500 人規模で育てば、その中から世界的なトップタレントが現れてこの分野全体を引っ張っていくことが期待でき、また同時にこの棟梁レベル人材が、実務の傍ら随時大学や企業で人材育成を支援することで、独り立ちレベル以下の人材育成が促進され、スケールアウトが進むことを期待することができる(図 3)。



<sup>6</sup> 米国における CAO の現状については、例えば *Computerworld* の以下の記事を参照。  
<http://www.computerworld.com/article/2688352/chief-analytics-officer-the-ultimate-big-data-job.html>

## 4. 具体的施策

前章で述べた「あるべき姿」に向けて、懇談会が考える具体的な施策の例を示す。これらの施策は個別に実施すべき性質のものではなく、適切に組み合わせることで量と質の双方を担保した適切なポートフォリオを考えるべきである。

### 4.1. リテラシーの醸成（50万人規模）

- ・ 高校教育・大学教養での講義において、世界がデータとその利活用で大きく変わっている重大性を教える。たとえば、
  - 高校生・大学生がワクワクするような啓発書・教科書をつくる。
  - 社会でどう使われ役立っているかを示す事例のビデオ素材をつくる。
  - 活躍中のデータサイエンティストをプールし、大学や高校にこれらのデータサイエンティストを適宜派遣し講義を担当してもらう。
- ・ 大学基礎教育にデータサイエンスを取り入れる。
  - 大学 124 単位の内、共通教育で例えば 4 単位、専門教育では専門に応じて例えば 2 から 6 単位をデータサイエンス（統計）に割り当てる、と定める。この際、核となる週 1 時間の講義にコンピュータ実習や問題を解く演習もセットにする。
  - これに合わせて、基礎統計教育も見直す。
- ・ 社会一般の興味を惹くための施策を実施する。
  - 全国的なデータサイエンスに関するコンテストを実施する。スーパーグローバルハイスクール指定校等の先進的な取り組みを行っている高校にも積極的参加を呼びかける。
  - gacco における「社会人のためのデータサイエンス入門」のように、MOOC (Massive Open Online Course) を利用した一般向け教材を充実させる。

### 4.2. 見習いレベルの育成（5万人規模）

- ・ データサイエンス力の基本的スキルとしての、統計学、機械学習、最適化およびビッグデータ解析に必要なプログラミングやデータ可視化の基礎を学習させる。学部・大学院におけるデータサイエンスの教育カリキュラムに関しては、統計教育参照基準・情報教育参照基準に相当するものを早急に作成する<sup>7</sup>。

---

<sup>7</sup> HP の一覧で確認できるだけでも現在 279 のデータサイエンス教育プログラムが存在する（巻末の表 1）。そのうち、修士レベルが 214。

- 大学院に、ダブルディグリー・ジョイントディグリー<sup>8</sup>の考え方を導入し、専門科目と共にデータサイエンスを副専攻などの形で学べるようにする。逆に、データサイエンスを主専攻とする学生が、データサイエンスの適用分野を副専攻として学べる機会も提供する。データサイエンスのコースを持たない大学では、他大学の大学院とのダブルディグリーも視野に入れる。
- スケールアウトの方策として、MOOC(Massive Open Online Course)を積極的に利用すると共に、コンテンツの充実を図る。
  - gacco における「統計学 I, II, III」を参考に、データサイエンス向けのコンテンツの充実を図る。
  - また、Coursera における Data Science の証明書の発行などを参考に、MOOC 修了者に対して修了書を発行する、あるいは、大学における単位認定を行うなど、MOOC の位置付けを社会的に認知させる。
- 社会人に対しては、特に機械学習など新技術に関して、ミドルマネジメント層を含む広い対象に、再教育のプログラムを提供する。統計数理研究所の公開講座(主に初級)の取組を参考に、地方を含む各大学に展開する。この際、必要に応じて産業界からの寄付を募ることも検討する。社会人の学び直しを推進するため、文科省において検討が進められている「職業実践力育成プログラム」認定制度(大学等における社会人や企業等のニーズに応じた実践的・専門的なプログラムを国が認定する制度)の活用(見習いレベルの育成に限ったものではなく、各レベルでも活用しうる)も効果的と考えられる。

#### 4.3. 独り立ちレベルの育成 (5000 人規模)

- 問題設定能力、問題解決のための戦略立案能力、データの収集・キュレーション能力、データ分析結果の業務や事業への実装能力、異分野研究者や事業者との連携・コミュニケーション能力、研究倫理、情報セキュリティの能力を備え、独立してデータサイエンスを推進できるレベルを目指す<sup>9</sup>。
- データ解析スキルとしては、MCMC (マルコフ連鎖モンテカルロ) 法、データ同化、インピュテーション技術、高次元空間の構造探索とモデル化、異種情報統合による個人化技術、隠れた関係の検出、特異性の発見、因果推論など、またデータ可視化技術

<sup>8</sup> Double Degree は 2 つの大学が 1 つの学位論文に対して異なる学位記を発行すること、Joint Degree は、1 つの学位記に対して 2 つの専門分野を認定すること、である。これらに限らず、学位記とは別に、修了書を発行する形でもよい。

<sup>9</sup> 日本学術会議 提言「ビッグデータ時代に対応する人材の育成」。米国の修士プログラムのカリキュラム例を巻末の表 2・4 に掲載。

としては、次元圧縮、特徴抽出、パターン認識などが重要とされているが<sup>9</sup>、データサイエンスを巡る状況は急速に変化しているので、不断の検討が必要である。

- 大学院において PBL に基づく専門育成プログラムを推進する。
  - 「分野・地域を越えた実践的情報教育協働ネットワーク(enPiT)<sup>10</sup>」、「先導的 IT スペシャリスト育成推進プログラム<sup>11</sup>」などにおける PBL の取組が参考になる。
- 企業との連携を通して、企業の事例を PBL 化したり、インターンシップ・プログラムを通して実務を学べる機会を与える。
  - 産学連携では経団連主体の CeFIL（高度情報通信人材育成支援センター）<sup>12</sup>が参考になる。

#### 4.4. 棟梁レベルの育成（500 人規模）

- 分野を俯瞰し戦略の立案・実行ができる棟梁レベルの人材は、常にデータサイエンスの最新の手法群(現時点でいえば例えばノンパラメトリックベイズ、スパースモデリング、カーネル法、グラフィカルモデル、深層学習、分散リアルタイム計算フレームワーク Storm<sup>13</sup>など)に加えて、新しい応用分野（例えば広告におけるリアルタイムビディング、マーケティングにおける協調フィルタリング、ロボティクスにおける画像・映像認識、質問応答システムの医療応用など）に精通している必要がある。このような「ホットな」情報は、実際にビッグデータを扱い、データサイエンスを研究開発している専門家が集中しているところに集まってくる。このため、ビッグデータ・データサイエンスの最先端の手法・応用の研究・開発及びそれらに精通した人材育成を行う国家的な拠点を設置する必要がある。
- データサイエンティストとしての実務経験のある社会人を、棟梁レベルに育成する集中的プログラムを上記拠点で開講する。ここでは、最先端の手法を PBL を通して実地で経験し、各応用領域での最新の成果をケーススタディとして学ぶと共に、拠点の特質を活かして棟梁レベルデータサイエンティスト間の人的ネットワークも形成する。
  - 社会人に 1 年間の集中的な教育をほどこす情報学研究所「トップエスイープログラム<sup>14</sup>」を参考にし、社会人を受け入れてデータサイエンスにおける経験と人脈を形成させる統計数理研究所「データサイエンス・リサーチプラザ<sup>15</sup>」の取組を量的に拡充する。

---

<sup>10</sup> <http://www.enpit.jp/>

<sup>11</sup> [http://www.mext.go.jp/a\\_menu/koutou/it/](http://www.mext.go.jp/a_menu/koutou/it/)

<sup>12</sup> <http://www.cefil.jp/>

<sup>13</sup> <https://storm.apache.org/>

<sup>14</sup> <http://www.topse.jp/>

<sup>15</sup> [http://www.ism.ac.jp/shikoin/overview/pdf/DSRP\\_lo\\_0323.pdf](http://www.ism.ac.jp/shikoin/overview/pdf/DSRP_lo_0323.pdf)

- ・ 科学の諸分野において、既にビッグデータを活用した研究方法論をある程度身につけた人材（博士号取得者等）に対して、産業界やアカデミアにおいて活躍できるキャリアパスを見据えて再教育する。
  - 米国の Insight Data Science Fellow Program<sup>16</sup>では、物理や天文学など諸科学における博士号を持つ科学者をデータサイエンティストとして訓練して産業界に送り込んでいる。この取組を参考にし、情報・システム研究機構「データ中心科学リサーチコモンズ<sup>17</sup>」、統計数理研究所「統計思考院<sup>18</sup>」で行われている、データサイエンス研究者の育成プログラムを量的に拡充し、アカデミア及び産業界で活躍できる棟梁レベルのデータサイエンティストを育成する<sup>19</sup>。

#### 4.5. 指導的データサイエンティストの育成（50人規模）

- ・ 3.5節で述べたように、指導的データサイエンティストは、ワールドクラスの人材であり、系統的に育成するのは難しい。むしろ施策としては、上記ビッグデータ・データサイエンス拠点において、世界最先端の手法・応用の研究・開発を推進することによって、このような指導的データサイエンティストが生まれてくる土壌を醸成することが大切である。
  - このため、国レベルで、ビッグデータのフラッグシップ・プロジェクトを実施し、その中で指導的データサイエンティストが活躍できる場を提供する。
- ・ 同時に、このような才能のポテンシャルを持つ者を若いうちに、発掘し、十分な機会を与える必要がある。このための施策として、以下の2点を提案する。
  - 定期的なデータサイエンス・ハッカソンを実施する。ITセキュリティの分野では、同様の試みが IPA「セキュリティ・キャンプ<sup>20</sup>」として行われている。これを参考に、データサイエンスにおけるプログラムを企画する。
  - 産業界で活躍する指導的データサイエンティストになれる人材を育成するために、才能のある若い者にメンターをつけ、資金等の援助を与えると共に人脈を形成する機会を提供する。例えば IPA「未踏 IT 人材発掘・育成事業<sup>21</sup>」や JST「さきがけ」を参考に、指導的データサイエンティスト候補者向けのプログラムをつくる。

<sup>16</sup> <http://insightdatascience.com/>

<sup>17</sup> <http://rc.rois.ac.jp/>

<sup>18</sup> <http://www.ism.ac.jp/shikoin/>

<sup>19</sup> 米国では Insight Data Science プログラムを含め 5 週・13 週程度の bootcamp が 17 プログラム実施されている。<http://datascience.community/bootcamps>

<sup>20</sup> <https://www.ipa.go.jp/jinzai/camp/index.html>

<sup>21</sup> <https://www.ipa.go.jp/jinzai/mitou/>



#### 4.6. トリクルダウン効果とスケールアウト

上記の具体的な施策の中で、最も重要なのが、現在わが国には決定的に不足し、ボトルネックとなっている棟梁レベルの育成であり、このレベルの人材が育ってくれば、独り立ちレベルのための PBL の指導者、見習いレベルのための専門科目の講義、リテラシーのための講師派遣などを通して、本報告書で述べたデータサイエンティスト育成の全体像がスケールしていくものとする。このためには産業界とアカデミアの密接な連携が欠かせない。

なお、これらの棟梁レベルデータサイエンティストについては、それらの間で情報交換、相互研鑽のためのコミュニティを作り、このコミュニティが、次世代のデータサイエンティストを育成する指導的役割（高校・大学等での講義も含む）を与える。

十分な数の「棟梁レベル」が存在しない現時点では、まずは国家的拠点の設立を急ぎ、あわせて全国主要 10 大学程度で他のレベルの人材育成をスタートすると共に、MOOC などのオンライン教材の利用を積極的に勧め、地方を含めた全国的な波及効果を狙うべきである。

#### 4.7. 出口戦略

育成された人材は、民間・アカデミアを問わず広く活用されなければならない。このためには、ビッグデータ利活用とデータサイエンティストの重要性を広く社会に発信していくことが重要である。また、データサイエンティスト教育を受けた者、あるいはある一定のスキル基準を満たす者に対してスキル認定を行い、雇用する側とデータサイエンティストのスキルのミスマッチが起こらないようにすることも重要と考えられる。

### 5. まとめ

我が国のビッグデータ利活用専門人材育成の課題の解決に向けた取組みの考え方を下記の 3 点にまとめる。

- 我が国の問題の根源は、棟梁レベルの決定的不足にある。この解決のために国家レベルの拠点を設置して、年 500 名規模の「棟梁レベル」の人材育成をめざし、上層への成長や下層へのトリクルダウン効果も狙う。
- リテラシーレベルや独り立ちレベルの大学教育を加速させるために、主要 10 大学

程度で本報告書の提案に基づく人材育成をスタートすると共に、MOOC などのオンライン教材を整備し、全国への波及効果を狙う。

- 社会全体のリテラシーやアウェアネスを向上させるために、全学的教養教育の実施、国家レベルのフラッグシップ・プロジェクトの推進、コンテストの開催、映像素材の充実などの取組を行う。

これらの方策を実現するにあたっては、データサイエンスを副専攻とするダブルディグリー制人材育成の推進やスキル認定制度も有効と考えられる。

## 参考資料

### 公開情報

日本学術会議 情報学委員会 E-サイエンス・データ中心科学分科会提言「ビッグデータ時代に対応する人材の育成」平成 26 年 9 月

日本学術会議 数理科学委員会 数理統計学提言「ビッグデータ時代における統計科学教育・研究の推進について」平成 26 年 8 月

統計検定出題範囲表 (4 級 - 1 級)

J. Manyika, M. Chui, et al., Big Data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, (2011).

### 各委員からの提出資料 (内部資料)

安宅委員提出資料 「データ社会に求められる新しい才能とスキル」  
「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」

岡本委員提出資料 「データキュレーター育成と大学教育への期待」

北川委員提出資料 「日本学術会議提言について」  
「ビッグデータ解析技術」  
「海外大学におけるデータサイエンスプログラムについて」

佐藤委員提出資料 「統計教育に関する取組」  
「統計家の行動基準」

樋口委員提出資料 「分野の壁を突き崩すビッグデータ」  
「データに関連した数理分野の俯瞰図」

丸山 (宏) 委員提出資料 「データサイエンティスト - 現状と課題」

渡辺委員提出資料 「統計科学・データサイエンス教育の体系化」  
「ヘルスデータサイエンス」