

# ビッグデータ時代における アカデミアの挑戦

～アカデミッククラウドに関する検討会 提言～

平成 24 年 7 月 4 日

文部科学省 アカデミッククラウドに関する検討会



## 目 次

	頁
I. はじめに	1
II. ビッグデータ時代におけるアカデミアの役割	3
1. ビッグデータに係る研究開発の推進等	3
(1) データ科学の高度化に関する研究開発	3
(2) アカデミッククラウド環境の構築	4
(3) ビッグデータ活用モデルの構築	5
2. 推進するにあたって留意すべき事項	6
(1) 分野間の連携	6
(2) 国際連携	6
(3) 人材育成	7
3. 我が国としてのビッグデータ基盤構築に向けて	8
III. 文部科学省が推進すべき研究開発課題	10
1. データ科学の高度化に関する研究開発	10
(1) 研究開発の方向性	10
(2) 研究開発を推進するにあたっての重要事項	11
2. アカデミッククラウドの構築に向けたシステム研究	15
(1) システム研究の方向性	15
(2) システム研究を推進するにあたっての重要事項	16
3. ビッグデータ活用モデルの構築	19
(1) 活用モデル構築の方向性	19
(2) 活用モデルを構築するにあたっての重要事項	20
4. 研究開発課題の推進によるイノベーション創出	20
図 分野を越えたビッグデータ利活用による科学技術イノベーション実現 のための共通基盤技術・活用基盤の創出	21
参考資料	22
・アカデミッククラウドに関する検討会における提言（概要）	23
・アカデミッククラウドに関する検討会の設置について	25
・アカデミッククラウドに関する検討会 委員名簿	27
・アカデミッククラウドに関する検討会 検討経緯	28



## I. はじめに

高度情報化社会の進展に伴い、デジタルデータが爆発的に増大するビッグデータ<sup>1</sup>（情報爆発）時代が到来した。世界のデジタルデータの量は、民間調査機関の推計によれば、2020年には、約35ゼタバイト（2010年度時の約35倍）へ拡大する見込みである。その質的・量的に膨大なデータ（ビッグデータ）の中には新たな知識や洞察が埋もれているが、現況においては、その多くの情報が整理・構造化されておらず、有効に活用できていない状況である。

このため、ビッグデータを効果的・効率的に収集・集約し、革新的な科学的手法により知識発見や新たな価値を創造することの重要性が、国際的に認識されてきている。第一の科学的手法である経験科学（実験）、第二の科学的手法である理論科学、第三の科学的手法である計算科学（シミュレーション）と並び、データ科学<sup>2</sup>（data centric science =e-サイエンス）は第四の科学的手法と言われ、ビッグデータ時代における科学の新たな地平を拓く方法論として注目されている。

第三の科学的手法である計算科学（シミュレーション）は、相対的に単位データあたりの計算量が多く、演算速度が大きな課題となる（例：スーパーコンピュータを用いた高度な理論的計算等）。一方、第四の科学的手法であるデータ科学（data centric science）は、相対的に単位データあたりの計算量が少ないことが多く、計算科学とは異なり、演算速度よりも、質的・量的に膨大なデータを連携させ、知的に処理、活用することがより重要な課題と

---

<sup>1</sup> 「ビッグデータ」：大量で、多種・多様なデータ、並びにそれらのデータを許容できる時間内に効率的に収集・蓄積・処理・分析し、活用するための技術のこと。

<sup>2</sup> 「データ科学」：ゲノムデータ、地球観測データ、人の活動データ等の大量で、多種・多様なデータを統合等することで、新たな知を創出し、他の学問分野の課題解決や新たな普遍的な方法論を開発する学問分野のこと。

なる。

情報科学技術分野では、質的・量的に膨大なデータを連携し、高度に処理・活用するためには、新たな方法論等が必要であり、これらを研究開発することにより、新たな知の創造や科学技術イノベーションの創出、社会的・科学的課題の解決が見込まれ、その重要性が認識されている。国際的にデータ科学に関する研究開発やビジネスへの活用が活発化しているなど諸外国も取組を進める中、一刻の猶予も許されない。

したがって、ビッグデータに関し、我が国が研究開発を進めるべき事項を早急に整理し、分野間連携、国際連携、人材育成を図りつつ、産業化につなげることも視野に入れて、戦略的に研究開発や環境構築に取り組む必要がある。その際には多様なデータの共有化を促進する観点から、全国の大学等の研究者が、科学的あるいは社会的意義のある研究成果を適宜適切に得ることのできるアカデミッククラウド<sup>3</sup>環境のあり方を検討する必要がある。

そこで、必要な検討等を進めるため、文部科学省研究振興局長の下に、アカデミッククラウドに関する検討会を設置した。本検討会では、平成24年4月～6月に、「データベース等の連携」、「システム環境の構築」「データ科学の高度化に資する研究開発」の3点を検討課題として議論を行い、本提言をとりまとめた。本提言では、平成25年度概算要求に向けて、これらの課題を解決するため研究開発及びシステム構築に関する具体的事業を提案している。

今後、データ科学技術の高度化に資する戦略的な研究開発や環境構築を通じ、ビッグデータを体系化・構造化することにより、我が国発の科学技術イノベーションの創出を強く期待する。

---

<sup>3</sup> 「アカデミッククラウド」：大学等が有する研究や教育に関するデータを大学等間で共有するためのクラウドのこと。

## Ⅱ. ビッグデータ時代におけるアカデミアの役割

「情報を制する者は世界を制する」といわれるように、現在、世界の多くの分野（医療、防災、金融等）で情報の重要性が叫ばれている。このことは情報の宝庫ともいえる大学等の教育研究機関（アカデミア）においても同様である。しかし、アカデミアにおいては、研究対象が学際的になっているにもかかわらず、個々の研究分野で収集した膨大なデータを共有する場がほとんどなく、有効に利活用されているとは言えない。このため、アカデミアにある膨大なデータを連携し、高度に処理・活用する第四の科学的手法であるデータ科学を高度化する共通基盤技術の開発やアカデミッククラウド環境の構築により、新たな知の創造や科学技術イノベーションの創出、社会的・科学的課題の解決につなげる必要性が高まっている。

以下、文部科学省が施策を推進するにあたっての具体的観点について述べる。

### 1. ビッグデータに係る研究開発の推進等

#### (1) データ科学の高度化に関する研究開発

アカデミアにおいて蓄積されているデータ量が爆発的に増大している昨今、既存のデータ処理技術では対応が困難な場面が近い将来に生じることが予想される。例えば、ゲノム解析における遺伝子データ収集のスピードは、情報通信の進展の速度を超える勢いである。また、人間の活動データの収集が容易になる等により、人文社会学分野も含めた学際研究もデータ科学の対象となってきている。このような状況に対応し、医歯薬学分野、理工学分野や人文社会学分野において新たなブレイクスルーを起こすためには、データ科学の高度化に関する研究開発を進める必

要がある。

これまで我が国の情報科学技術は、時代時代の要請、あるいは変化を先取りする形で、データ科学に係る研究開発を進めてきた。しかし、データ量が指数関数的に増大しているビッグデータ時代においては、これまでの研究開発やその成果の延長線上だけでは対処できず、ブレイクスルーを起こすための研究開発を新たに推進しなければ対応できない。

なお、データ科学の高度化に関する研究開発においては、社会的・科学的課題の解決のための知識を創出するだけでなく、アカデミアと社会（政策決定の場等）をつなげる取組を併せて行い、社会からのフィードバックを受けてさらに研究開発の質を高めていくといった、社会のニーズにも応えられるイノベーションサイクルの構築も必要である。

## (2) アカデミッククラウド環境の構築

ビッグデータの利活用を推進するに当たっては、データベースの連携を推進し、必要な情報が適切に入手可能なクラウド環境を構築することが望まれる。民間では、クラウド技術の利活用は先行していたものの、データの権利問題等により、複数企業間のクラウド連携の環境を構築できないという現状がある。この点、大学等の研究機関は、研究開発するための技術を有し、クラウド環境構築に資する公共的で多種・多様なデータが存在することから、次世代のクラウド環境の構築に係るテストベッドとしての役割を果たすことが可能である。

アカデミアにおいて、クラウド基盤を構築することにより、大学等の研究者や民間企業等が研究開発に活用できるデータ、情報、研究資料等を容易に利用でき、科学的あるいは社会的に意義のある研究成果を出す



ことが可能となる。

また、アカデミアのうち、まずは大学等の高等教育機関においてクラウド環境を構築することにより、初等・中等教育機関等の他の環境にも適用できる可能性が十分にある。

アカデミッククラウド環境の構築に向けて、多数の研究分野が何らかの形でデータを持つ現状や、今後の第四の科学的手法をサポートする環境を構築する必要性、教育分野における質の向上、教育環境の活性化等、様々な観点を考慮し、例えば、複数の大学等による検討組織を設けて、大学等の立場から課題を精査した後、国主導で課題解決を遂行することが必要である。

なお、アカデミアは、学術利用等の観点で個人情報の扱いについて方向性を打ち出せる立場にあることから、社会に対し指標を提示する意味からも、データに付随する個人情報に係る検討に取り組むことが求められる。

### (3) ビッグデータ活用モデルの構築

近年、アカデミアの研究開発が複雑化する中で、組織や国を飛び越えて研究データを共有する動きが活発化している。

我が国でも、研究開発法人や大学共同利用機関法人等においては、膨大な海洋、天文科学、災害、材料等の専門データが存在する。これらのビッグデータによる研究開発を効率化・加速化するためには、研究開発法人等の所有する膨大なデータの掘り起こしを行い、研究に活用できるよう、ルールを含め整備することが必要となっている。併せて、ビッグデータ利活用に関するモデルの構築も進める必要がある。

## 2. 推進するにあたって留意すべき事項

### (1) 分野間連携

これまで、情報科学技術分野の研究者が他分野の研究者と協業する機会においては、情報科学技術分野の研究者が分析のための下支え的な役割となることが多く、相互に恩恵を得ることができる関係になっていなかったとの指摘もある。

しかし、今後、異なる分野のデータを融合しつつビッグデータの利活用を推進し、新たな科学技術イノベーションにつなげるためには、異なる分野の研究者の積極的な連携・協力が必要となる。特に、アプリケーション側の研究者が増大するデータに独自で対処できなくなり、情報科学分野の専門的知見を必要とする事例が増えている。

そこで、情報科学技術分野の研究者が中心となり、グリーン・ライフサイエンス・防災等の各アプリケーションにおいて、関連する多くの分野の研究者たちと対等な立場で協業できるような仕組みを構築することが求められる。

### (2) 国際連携

世界では、地震、津波、洪水等の自然災害が各地で発生しており、それらによる被害は毎年甚大である。その中で我が国は、東日本大震災をはじめ、多数の自然災害の被害を受けている。

その際、気象データ、地球観測データ等の自然災害に関わるビッグデータの蓄積が行われており、これまでに大学等の研究機関において、地震、津波、洪水をはじめとする自然災害について研究開発が進められている。そのため、我が国固有の自然災害に関するビッグデータを活用し、

我が国が主導して防災・減災のための国際連携を行うことが可能である。

諸外国のうち、米国は、平成 24 年 3 月 29 日に、総額 2 億ドル（約 185 億円）を投ずるビッグデータイニシアチブ構想を発表した。本構想では、今後の重要な技術課題を強力に解決していく姿勢が示されている。特に、ビッグデータを、スーパーコンピュータやインターネットと並んだ重要な分野ととらえており、大量データの核となる技術を向上させることで、安全保障、教育の改革、人材養成を実施することとしており、NSF（全米科学財団）等、データサイエンティスト育成のための大学院コースや、研究開発課題として、機械学習、クラウドコンピューティング、クラウドソーシング等を挙げている。

我が国は、災害等に関するビッグデータを活用した研究開発を進めるなかで、国際連携を図り、人類共通の課題解決に向けた取組を進める必要がある。その際、アカデミアが有する人材、知的資料を最大限活用していくことが求められる。

### (3) 人材育成

我が国では、ビッグデータを有する学問分野（グリーン、ライフサイエンス、防災等）と情報科学技術分野の両方に関する知識を有し、学際領域で能力を発揮できる情報科学技術分野の人材が、他国に比して不足しているとの指摘がある。また、民間調査機関の統計によれば、統計学や機械学習に関する高等訓練の経験を有し、データの深い分析に係る才能を有する大学卒業生数（2008 年）については、米国の 2 万 4,730 人であるのに対し、日本は 3,400 人と著しく少ないのが現状である。

ビッグデータ時代においては、統計学、プログラミングや可視化とい

ったデータ分析に高度なスキルを有し、分析結果を新たな知に結びつけられる人材（データサイエンティスト）が求められるとともに、大量で、多種・多様なデータから、埋もれた知識を引き出し、異分野間をつなぐ専門家（データキュレーター、データアナリスト）や、特定のビッグデータに専門的に精通した人材（バイオインフォマティシャン等）などの人材を育成することが必要となっている。

現状の課題を踏まえ、データサイエンティスト等を適切に育成するためには、データが身近に使える環境の構築が必要である。例えば、オン・ザ・ジョブトレーニング（OJT）の形で、実際のデータを用いながら、日々の研究開発をする中で、情報科学技術の研究を推進できる人材を育てていくこと等、今後、ビッグデータの活用に長けた人材を育成する環境を意識的に構築していくプロジェクトが必要である。

なお、ビッグデータ時代においては、「データ管理システム」、「クラウドデータベース」についてシステム研究を行いつつ、当該技術者の育成が必要となっている。

また、ビッグデータ時代においては、論文の引用数による評価だけではなく、価値の高いデータを作り出し、社会に提供した研究者に対して、インセンティブを与えるようなシステム設計も重要になると言われている。

### **3. 我が国としてのビッグデータに関する基盤構築に向けて**

現在、ビッグデータについては、文部科学省に限らず、複数の政府機関においても、重要な検討課題として位置づけられている。

複数の政府機関において適切に連携し、役割分担をしてビッグデータに対

する施策を進めることにより、国全体として効果的に取り組むよう、文部科学省においても、科学技術全体を俯瞰して、ビッグデータに関する共通基盤技術の開発や、利活用を推進する環境構築の整備を進めるべきである。

以上、ビッグデータの持つ可能性を最大化するため、データ科学の高度化に資する情報科学技術分野の研究開発やアカデミッククラウド環境構築のためのシステム研究等のビッグデータに関する研究開発、研究開発法人等におけるビッグデータ活用モデルの構築に関する事業について、分野間連携、国際連携、人材育成の観点に十分留意しつつ、早急に開始する必要がある

### Ⅲ. 文部科学省が推進すべき研究開発課題

Ⅱ章「ビッグデータ時代におけるアカデミアの役割」を踏まえ、ビッグデータに関して今後文部科学省が推進すべき研究開発課題は以下のとおりである。

#### 1. データ科学の高度化に関する研究開発

##### (1) 研究開発の方向性

ビッグデータの活用が有効な研究分野（ライフサイエンス、地球環境等）においては、これまでとは量的・質的に異なる大規模で多様なデータを自在に操り、データ比較・解析・モデリングから有意義な情報を抽出し、新たな知識を創出する、データドリブン型発見的探索スタイルの研究アプローチ（第四の科学的手法）が進展してきている。

しかし、様々なデータ（バイオ、天体観測等の自然科学のデータから社会科学的な人の観測データまで多様）を組み合わせ、大規模な処理を実行しようとする、想定外のデータや正常に分析できないデータが大きくなることが多く、必ずしもビッグデータが有効に活用できていないのが現状である。そこで、ビッグデータの解析を円滑に実行するための革新的な方法論等の研究開発を行うことが必要となっている。

情報科学技術分野においては、ビッグデータの可能性を開拓するため、データの収集、蓄積・構造化、データ処理・分析、処理結果の可視化・検証、モデリング、情報統合等の各段階における研究開発を推進することが必要である。その際、実効的なシステムを構築するため、実際のビッグデータを用いて研究開発を進めることが必要である。

したがって、ビッグデータの蓄積・構造化等を目指す情報科学・統計科学研究者と、ビッグデータの活用が有効な研究分野（ライフサイエン

ス、地球環境等)の研究者が対等に能力を発揮しうる「場」を形成し、融合研究体制により研究開発を推進することで、新たな科学技術イノベーションを創出する必要がある。

## (2) 研究開発を推進するにあたっての重要事項

### ①分野を越えた融合研究体制による研究開発

研究開発を推進するにあたっては、ビッグデータの蓄積・構造化等を目指す情報科学・統計科学研究者と、ビッグデータの活用が有効な研究分野(ライフサイエンス、地球環境等)の研究者がアプリケーションデータを共有し、異なる方向性からビッグデータに迫っていくといった、分野を越えた融合研究体制が必要である。なお、融合研究体制は、課題ごとに設定された複数の融合研究チームとそれらを全体的に統合する中核的拠点(プラットフォーム)により構成され、中核的拠点は、俯瞰的な立場でアプリケーションデータの共有や研究開発の連携・共有を図りつつ、体制全体の推進を図っていくことが考えられる。

融合研究体制によって研究開発すべき具体的事項は、以下のとおりである。

#### a) アプリケーション側と連携した新たな知の創出

- ・多様かつ大量のアプリケーションデータ(患者等のゲノムデータ、診療情報、地球観測衛星等によって得られる観測データ等)の転送、圧縮、保管等を容易に実現するための研究開発
- ・画像データや3次元データ等の多様なデータを検索、比較、解析等することで有意な情報を抽出するための研究開発

- ・ アプリケーションデータからの新たな課題の発見や洞察（疾患要因の解明、疾患予測・診断、気候変動予測等）をより正確に行うための研究開発
- ・ 定量データから生体レベル、自然現象レベル等の多様な数理モデルを構築し、実測データと組み合わせることで新たな知見を得るような、発見的探索スタイルの研究アプローチ推進のための研究基盤開発

#### b) 今後取り組むべき研究開発課題例

今後、分野を越えたビッグデータの利活用による科学技術イノベーションを実現していくために、集中して取り組むべき研究開発テーマとして、例えば、以下のような研究開発が挙げられる。

#### (ライフサイエンス分野)

ゲノム解析の分野では、シーケンサー技術の向上により、短時間に全ゲノムデータ（30 億塩基対）を解析することが可能となっている。これまで、様々な疾患は特定の塩基配列によるものと考えられてきたが、最近の研究では、これまで注目されていなかった複数の塩基配列が疾患に関連していることが分かってきた。このため、全ゲノムデータを活用して、効率的に、疾患に関連する複数の塩基配列を探索する必要があり、ライフ・情報・数学の協働により革新的アルゴリズムの研究開発を行う必要がある。

#### (地球環境分野)

地球観測のデータは、地球温暖化、水循環、生態系、地理空間等に



関するデータが存在するが、大量で、多種・多様なデータを管理し、新たな知見を見いだすには、異なるデータ間の関係性を高度につなぎ合わせる技術、特異点を高度に自動検出する技術、必要なデータのみを効率的に蓄積する技術等が必要である。また、データから有用な情報を容易に取得するために、可視化技術の高度化が必要である。

### (防災分野)

東日本大震災のような災害や事故の経験を踏まえ、今後、防災・減災に取り組むに当たっては、災害に対する堅牢性（ロバストネス）および回復力（レジリエンス）の強化を図るため、これまでの災害・事故から得られた、大量で、多種・多様なデータを容易に分析可能な形に蓄積・構造化する技術の開発や、リアルタイム分析、モデリング等の技術を高度化することが必要である。

なお、これらの技術は、将来的にはアカデミッククラウド環境において、大学をまたいで研究者や教員が研究・学習コンテンツを共有する際にも有効である。

## ②ビッグデータ利活用のための共通基盤技術の開発

ビッグデータの持つ可能性をより広げるため、特定のアプリケーション側のデータを活用しつつ、データの収集、データの蓄積・構造化、データ処理・分析、処理結果の可視化・検証、モデリング、情報統合等の各段階での基礎研究を総合的に推進する。

具体的には、以下のような研究開発事項が挙げられる。

#### a) データの収集

- ・ データクレンジング技術（ノイズ除去、データの正規化、不要なデータ変動の吸収等）
- ・ データに対して、自動的に意味や内容に係る注釈を付与する技術
- ・ データの持つあいまい性に対処する技術

#### b) データの蓄積・構造化

- ・ ストリーミングデータをオンライン処理し、構造化する技術
- ・ 複数のストリームデータから関連性を発見する技術
- ・ 必要なデータのみを蓄積する技術
- ・ 高度な圧縮技術、圧縮したままで検索する技術
- ・ 大規模データを容易に検索可能な形に構造化、組織化するデータベース設計技術
- ・ 大規模データの不確実性を定量化し、有効性、完全性、一貫性等を改善する技術
- ・ 国益上重要なデータに対する高度なアクセス制御機構に関する技術

#### c) データの分析・処理

- ・ データの意味理解と対話技術
- ・ (ストリーミング) データのクラスタリング技術の高度化・高速化
- ・ 大量・多様なデータからのモデリング技術
- ・ 異種データから関連性を探索する技術
- ・ データを匿名化する技術
- ・ 秘密性や匿名性を損なわないままマイニングする技術

- ・機械学習の高度化
- ・映像、画像等からのコンテキスト理解や異メディア融合によるデータマイニング技術の高度化
- ・ノイズに埋もれた大量データから有為な情報（特異点）を自動検出する技術
- ・大量データに対してリアルタイムに推定等を行う等の革新的アルゴリズム

#### d) 処理結果の可視化・検証

- ・多次元データや広域データの可視化技術
- ・実世界と関係のない抽象的なデータ処理結果の可視化技術
- ・多様なデータの相関や関係性を描き出し、新たな洞察を導く技術
- ・処理結果の信頼性を検証する技術
- ・分散ファイルシステムで高度にビッグデータをやりとりする際のシステム技術

## 2. アカデミッククラウドの構築に向けたシステム研究

### (1) システム研究の方向性

我が国においては、諸外国に比べて、パブリッククラウド（インターネットで不特定多数の利用者に対して提供されるクラウドサービス）の導入が遅れている現状がある。一方、大学等においては、耐災害性の強化やIT投資の合理化の観点から、大学間でクラウド基盤を連携・共有することの必要性が生じている。

このため、アカデミッククラウド環境構築のあり方について、大学間

のクラウド環境の連携、ビッグデータの管理・運営、教育クラウド構築のあり方、設置形態、対象範囲等についてシステム研究を行い、例えば、以下のような具体的事項について検討する必要がある。

## (2) システム研究を推進するにあたっての重要事項

### a) 大学間のクラウド環境の連携

- ・各大学等の組織毎に策定されている情報セキュリティポリシーについて、運用方針のあり方の検討
- ・学術認証フェデレーション等と密接に連携した認証システムのあり方の検討
- ・アカデミッククラウド環境で整備された情報を二次的な加工等を通じて知識として抽出できるよう、データの取扱いに関するルールの検討
- ・クラウドソースを利用した大量データの収集方法の検討

### b) 研究環境におけるビッグデータの管理・運営等

- ・ビッグデータの利活用が有効な研究分野で生成される大規模な解析データ、シミュレーションデータ等について、効率的な蓄積・管理・運用を実現するための検討
- ・データの量や利用数の変化に柔軟に対応するため、クラウド環境に登録するデータの標準化と共有化を図るためのあり方の検討
- ・データの永続的な保存についての仕組みや制度についての検討
- ・分散的にビッグデータを扱うためのシステムのあり方の検討

#### c) 教育クラウド構築のあり方

- ・ 学生のポートフォリオの管理・活用等のあり方の検討
- ・ 医歯薬学分野、理工学分野や人文社会学分野において、システムを活用できるような設計のあり方の検討
- ・ 商業クラウドサービスの活用の有無について、メリット・デメリットを整理した上でのあり方の検討
- ・ 教育コンテンツのライセンス、利用許諾・条件の集中的な管理機能とその条件を満たしうる利用者認証機能の構築のあり方の検討
- ・ 高等教育機関の大学間単位互換等の課題への適用のあり方の検討

#### d) 個人情報の保護等を考慮したシステム構築のあり方

- ・ データに付随する個人情報の保護やセキュリティの確保について、有効利用と個人情報の保護等の両面を法制度等から検証し、望ましいシステム構築や社会制度のあり方の検討
- ・ 異種のデータを連携した結果、想定されていなかった複数のデータが結合することにより、特定個人のデータが抽出されてしまう可能性があるため、その対策についての検討
- ・ 学習コンテンツに限らず、利用許諾・条件の発生するデータの管理・運営方法の検討

#### e) 設置形態

- ・ アカデミッククラウド上に構築した大量のデータについて、利用者が安定的にデータを活用できるような設置形態のあり方についての検討
- ・ 特定の大学等間に限定して実現するものではなく、汎用的に、全国的

にアカデミッククラウドが展開できるような設置形態の検討

- ・ 研究開発法人や大学共同利用機関法人等が有するプラットフォームとの連携・活用について検討

#### f) 対象範囲

- ・ アカデミッククラウド環境において、競争領域のデータと非競争領域のデータの切り分け方の検討

#### g) システム研究の検討体制、実施計画

- ・ 複数の大学等で検討プロジェクトを設置し、大学等の立場からアカデミッククラウド環境構築に係る各種課題について検討
- ・ 実施計画について、中長期的視点で優先的に取り組むべき課題を抽出し、順次システム構築を進め、成功事例を作り続けることにより、成長型の実質的に役立つシステムの検討

#### h) データ量の調査

- ・ アカデミッククラウド環境に適応するデータが分野ごとにどの程度が存在するかの調査や、メタデータ付与の有無の調査

#### i) 効率的な運用

- ・ アカデミッククラウド環境を構築する場合、IT投資の合理化を促進するという視点も考慮し、アカデミッククラウドの効率的な運用（利用負担金等）の検討

### 3. ビッグデータ活用モデルの構築

#### (1) 活用モデル構築の方向性

近年、科学技術の研究開発が複雑化する中で、組織や国を飛び越えて、研究データを共有する動きが活発化している。

米国NSFやNIH（米国国立衛生研究所）がファンドする研究では、研究により生み出されたデータの供出を義務づけられており、また2009年5月には、欧州委員会で、「e-infrastructure（欧州委員会FP7の研究用データ共有基盤）」が提案され、現在構築に向けて各種プログラムが実施されているなど、欧米諸国でも国を挙げての取組が進められている。

我が国においては、研究開発に関するデータについて、研究開発法人ごとにさらに分野ごとに細分化されてデータベースとして整備されている場合が多いが、データベース間の連携が進んでいるとは言い難い。

そこで、まずはビッグデータの活用モデルを構築することにより、データ連携やオープン化を促す必要がある。また、研究開発法人等が所有する膨大な専門データの掘り起こしを行い、研究に活用できるよう、ルールを含め整備する必要がある。さらに、ライフサイエンス分野におけるバイオサイエンスデータベースセンター（NBDC）やライフサイエンス統合データベースセンター（DBCLS）のように複数の研究機関が有するデータベース間の連携の推進、これらのデータの民間等での利活用を促進することが重要である。ビッグデータ時代の到来に伴い、近い将来、需要の増大が見込まれるビッグデータを活用・分析できる人材の育成も重要である。

将来的には、ビッグデータ活用基盤を構築し、知識インフラを整備することで、イノベーションの創出に寄与することが望まれる。

## (2) 活用モデルを構築するにあたっての重要事項

- a) 研究開発法人等が所有する専門データや膨大なデータの掘り起こしを行い、研究に活用できるよう、ルールを含め整備
- b) 複数の研究機関が有するデータに関するデータベース連携の推進、これらのデータの民間等での利活用の促進
- c) データサイエンティストやバイオインフォマティシャン等、学際領域で能力を発揮できる情報科学技術分野の人材の育成
- d) (将来的には) ビッグデータ活用基盤の構築を推進
- e) ビッグデータの研究成果を社会で活用する方策の立案と実行

## 4. 研究開発課題の推進によるイノベーション創出

ビッグデータ時代において、1～3の研究開発課題を各省連携、産学官連携で実施することにより、成長のプラットフォームとなるビッグデータの利活用を促進し、新たな知の創造、新たな科学技術イノベーションの創出や社会的・科学的課題の解決を図ることが、我が国が持続的な成長を遂げるために重要である。これらの研究開発課題を積極的に推進することにより、イノベーションによる新産業・新市場の創出や国際競争力の強化につながることを強く期待する。



# 分野を越えたビッグデータの活用による科学技術イノベーション実現のための共通基盤技術・活用基盤の創出

研究開発

**大規模・多様なデータの存在**

大量に散在する多様な科学技術・学術情報、実世界データやセンシングデータ、WEB上のデータ等

グリーン ライフ 復興 地球観測 防災

大きな可能性があるが、整理・構造化されておらず、社会の課題解決等に有効に活用できていない状況

**1. ビッグデータに関する共通基盤技術の研究開発**

① ビッグデータ処理の各段階(a.データ収集、b.蓄積・構造化、c.分析・処理、d.可視化)における基盤技術の開発  
(※基盤技術を開発する上では、実データが必須であり、以下の具体的な分野やカウンタートハートと連携・協業することによる研究開発を実施)

② アカデミッククラウド環境構築に関するシステム研究

**分野間連携**

- アプリケーション側(ライフ、地球環境等)との協働によるプロジェクト研究
  - <ライフ分野>
    - ・効率的な相同性検索、ゲノム等情報から疾患の要因同定
    - ・臓器等の数理モデルと実測の融合(データ同化)
  - <地球観測分野>
    - ・複数パラメータの処理技術、高度圧縮技術
    - ・異なるデータ間の連携技術

**国際連携**

- 国際共同研究
  - ・災害に対する堅牢性(ロバストネス)及び回復力(レジリエンス)の強化に関し、ビッグデータを通じた研究協力
  - ・それぞれの国の災害経験及びその対処等を踏まえた共同研究

**人材育成**

- 人材育成
  - ・データ分析に高度なスキルを有し、分析結果を新たな知に結びつける人材(データサイエンティスト等)
  - ・ビッグデータから埋もれた知を引き出し異分野間をつなぐ人材(データキュレーター、データアナリスト等)
  - ・特定のビッグデータに精通し、分析等に適した人材(バイオインフォマティシャン等)

**2. 研究開発法人等におけるビッグデータ活用モデルの構築**

- ・研究開発法人等が所有する専門データや膨大なデータの掘り起こしを行い、研究に活用できるよう、ルールを含め整備
- ・複数の研究機関が有するデータに関するデータベース連携の推進、これらのデータの民間等での利活用の推進

ビッグデータの構造化・体系化による知識インフラの構築

ビッグデータ活用基盤の構築

成長のプラットフォームとなるビッグデータの利活用の促進

- 新たな知の創造
- 第四の科学としてのデータの確立
- 科学技術イノベーション創出

**社会的・科学的課題の解決**  
(具体的な例)

- ・消費エネルギー、CO2排出量削減
- ・復興の高速化、必要な支援の見える化
- ・病気や疾患の早期発見・予防
- ・地震・津波等災害の予測・防災

省庁の枠を超えた連携  
産学官連携

イノベーションによる新産業・新市場創出、国際競争力強化

未来 ← 現在

# 参 考 资 料

# ビッグデータ時代におけるアカデミアの挑戦 アカデミッククラウドに関する検討会 提言（概要）

## 1. 現状と課題

- 世界のデジタルデータの量は、民間調査機関の推計によれば、2020年には、約35 ゼタバイト（2010年度時の約35倍）へ拡大する見込みである。その質的・量的に膨大なデータ（ビッグデータ）の中には新たな知識や洞察が埋もれているが、現況においては、その多くの情報が整理・構造化されておらず、有効に活用できていない状況である。
- また、ビッグデータを効果的・効率的に収集、集約し、革新的な科学的手法により情報処理を行うことで、新たな知的価値を創造する、第四の科学的手法としての「データ科学」の確立が重要である。
- 情報科学技術分野においては、質的・量的に膨大なデータを連携し、高度に処理・活用するためには、新たな方法論等が必要であり、これらを研究開発することにより、新たな知の創造や科学技術イノベーションの創出、社会的・科学的課題の解決が見込まれるとともに、その重要性が認識されている。国際的にデータ科学に関する研究開発やビジネスへの活用が活発化しているなど諸外国も取組を進める中、一刻の猶予も許されない。

## 2. 今後推進すべき取組

上記の課題に対し、大学等の教育研究機関（アカデミア）においては、研究対象が学際的になっているにもかかわらず、個々の研究分野で収集した膨大なデータを共有する場がほとんどなく、有効に利活用されているとは言えない。このため、アカデミアにある膨大なデータを連携し、高度に処理・活用することを可能とする共通基盤技術の開発やアカデミッククラウド環境の構築により、新たな知の創造や科学技術イノベーションの創出、社会的・科学的課題の解決につながる必要性が高まっている。

文部科学省が施策を推進するにあたっての具体的観点は以下のとおりである。

### （1）ビッグデータ利活用のための共通基盤技術の研究開発

- 各分野（ライフサイエンス、地球環境、防災等）の取組との融合研究体制により、各分野に共通するビッグデータ処理の各段階（a. データ収集、b. 蓄積・構造化、c. 分析・処理、d. 処理結果の可視化）における基盤技術の開発を推進する。

### （2）アカデミッククラウド環境構築に関するシステム研究

- 大学等の研究機関は、研究開発するための技術を有し、クラウド環境構築に資する公共的で多種・多様なデータが存在することから、次世代のクラウド

環境の構築に係るテストベッドとしての役割を果たすことが可能である。

- このため、アカデミッククラウド環境構築のあり方について、大学間のクラウド環境の連携、ビッグデータの管理・運営、教育クラウド構築のあり方、設置形態、対象範囲等についてシステム研究を行う。

### **(3) 研究開発法人等によるビッグデータ活用モデルの構築**

- 研究開発法人や大学共同利用機関法人等にある、膨大な海洋、天文科学、災害、材料等の専門データや死蔵されている膨大なデータの掘り起こしを行い、研究に活用できるよう、ルールを含め整備する必要がある。
- 複数の研究機関が有するデータに関するデータベース連携の推進、これらのデータの民間等での利活用の促進を図る。
- 将来的には、ビッグデータ活用基盤を構築し、知識インフラを整備することで、イノベーションの創出に寄与する。

### **※研究開発を推進するに当たって留意すべき事項**

- 情報科学技術分野の研究者が中心となり、グリーン・ライフサイエンス・防災等に関連する多くの分野の研究者たちと対等な立場で協業できるような分野間連携の構築。
- 災害等に関するビッグデータを活用した防災・減災のための国際共同研究の推進。
- ビッグデータの利活用に関する高度なスキルを有するデータサイエンティスト、データキュレーター、データアナリスト等の人材育成。

## **3. ビッグデータによる新産業・イノベーションの創出に向けて**

ビッグデータ時代において、上記の取組を各省連携、産学官連携で実施することにより、成長のプラットフォームとなるビッグデータの利活用を促進し、新たな知の創造、第四の科学としてのデータ科学の確立、新たな科学技術イノベーションの創出や社会的・科学的課題の解決を図ることが、我が国が持続的な成長を遂げるために重要である。これらの取組を積極的に推進することにより、イノベーションによる新産業・新市場の創出や国際競争力の強化につながることを強く期待する。

# アカデミッククラウドに関する検討会の設置について

平成24年4月16日  
文 部 科 学 省  
研 究 振 興 局

## 1. 設置の趣旨

第4期科学技術基本計画においては、科学技術振興のための基盤として、「知識インフラ」の形成が盛り込まれている。

情報化社会の進展に伴い、「情報爆発」と呼ばれるように、大量のデータ、情報がインターネットを通じて流通している今日、様々な分野における知的活動の成果として生み出されている大量データを、効果的・効率的に収集、集約し、革新的な科学的手法により情報処理を行うことで、新たな知的価値を創造する、「データ科学」の重要性が認識されてきている。「データ科学」は、第四の科学的手法として注目されている。

また、社会問題の解決に向けた基礎研究からのイノベーションへの貢献という観点からも、その重要性が認識されている。

そこで、全国の大学等の研究者が、サイエンスに活用できる多分野にわたるデータ、情報、研究資料等を、オンラインにより、手軽に利用でき、最新の「データ科学」の手法を用いて、科学的あるいは社会的意義のある研究成果を得ることのできる「アカデミッククラウド環境」について、必要な議論、検討等を進めるための検討会を設置する。

## 2. 調査事項

- (1) データベース等の連携について
- (2) システム環境の構築について
- (3) データ科学の高度化に資する研究開発について
- (4) その他アカデミッククラウドに関すること

## 3. 構成及び運営

- ・ アカデミッククラウドに関する検討会は、研究振興局長の私的諮問機関として設置し、科学技術・学術審議会学術分科会研究環境基盤部会学術情報基盤作業部会及び科学技術・学術審議会研究計画・評価分科会情報科学技術委員会と連携しながら、議論、検討を進める。
- ・ アカデミッククラウドに関する検討会の構成員は、別紙のとおりとする。
- ・ アカデミッククラウドに関する検討会の運営に係る事項は検討会において

定める。

#### 4. 設置期間

平成24年4月27日～調査事項の終了までとする。

#### 5. その他

- ・ アカデミッククラウドに関する検討会の庶務は、研究振興局情報課が処理する。
- ・ 委員の委嘱は1年ごとに行うものとする。



## アカデミッククラウドに関する検討会 委員名簿

相原 玲二	広島大学情報メディア教育研究センター長
安達 淳	国立情報学研究所 学術基盤推進部長
安達 文夫	国立歴史民俗博物館情報資料研究系教授
北川 源四郎	情報・システム研究機構長
主査代理 喜連川 優	東京大学生産技術研究所教授
五條堀 孝	国立遺伝学研究所副所長
竹内 比呂也	千葉大学附属図書館長兼アカデミック・リンク・センター長
主査 西尾 章治郎	大阪大学大学院情報科学研究科教授
美濃 導彦	京都大学情報環境機構長
門田 博文	科学技術振興機構執行役
安浦 寛人	九州大学理事・副学長
山口 しのぶ	東京工業大学学術国際情報センター教授
山名 早人	早稲田大学理工学術院教授

(50音順)

アドバイザー (科学技術・学術審議会 学術分科会 研究環境基盤部会部会長)  
(学術情報基盤作業部会 及び 情報科学技術委員会 主査)

有川 節夫 九州大学総長

## アカデミッククラウドに関する検討会 検討経緯

### ◆ 第1回アカデミッククラウドに関する検討会

日時：平成24年4月27日（金） 14：30～16：30

- (1) アカデミッククラウドに関する検討会について
- (2) アカデミッククラウドに関する検討会に係る検討の進め方について
- (3) その他

### ◆ 第2回アカデミッククラウドに関する検討会

日時：平成24年6月4日（月） 15：00～17：00

- (1) アカデミッククラウドに関する検討の方向性について
- (2) アカデミッククラウドに関する具体的検討について
  - ①システム環境の構築について
  - ②データ科学の高度化に資する研究開発について
- (3) その他

### (第3回アカデミッククラウドに関する検討会前の有識者打合せ)

日時：平成24年6月21日（木） 14：00～16：00

- (1) 研究開発事項に関する具体的検討
- (2) その他

### ◆ 第3回アカデミッククラウドに関する検討会

日時：平成24年6月29日（金） 10：00～12：00

- (1) 概算要求に向けた提言について
- (2) その他

### ■ 提言「ビッグデータ時代におけるアカデミアの挑戦」取りまとめ

平成24年7月4日（水）





