

分野を超えたビッグデータ利活用により新たな知識や洞察を得るための革新的な情報技術及びそれらを支える数理的手法の創出・高度化・体系化

(平成25年度設定)

1. 戦略目標名

分野を超えたビッグデータ利活用により新たな知識や洞察を得るための革新的な情報技術及びそれらを支える数理的手法の創出・高度化・体系化

2. 達成目標

情報科学・数理科学分野とビッグデータの利活用により大きな社会的インパクトを生むような様々な研究分野（アプリケーション分野）との協働により研究を進め、アプリケーション分野での課題解決を通じてビッグデータから新たな知識や洞察を得ることを可能とする次世代アプリケーション技術を創出し、高度化すると同時に、様々な分野のビッグデータを統合解析することを可能とする共通基盤技術の構築を目指す。そのため、以下の目標の達成を目指す。

- 各アプリケーション分野においてビッグデータの利活用を推進しつつ様々な分野に展開することを想定した次世代アプリケーション基盤技術の創出・高度化
- 様々な分野のビッグデータの統合解析を行うための次世代基盤技術の創出・高度化・体系化

3. 将来実現しうる重要課題の達成ビジョン

本戦略目標を実施し、「2. 達成目標」に記載した研究成果が得られることで、様々な分野のビッグデータを統合解析するための共通基盤技術を構築することができ、分野を超えたビッグデータの利活用を実現することができる。構築された技術を用いることで、ビッグデータの利活用が有効な研究分野の論文データ、実験・シミュレーションデータ、観測データ等の高度利用が可能となり、社会科学・人文科学等を含む複数の分野が連携した異分野融合領域のイノベーション創出を加速させることができる。

本事業終了後、アカデミア・企業等が様々な分野のビッグデータを統合解析できる共通基盤技術を活用して、研究開発や実用化を推進することで、例えば

- ・ライフサイエンス分野では、診療情報と関連づけられた10万人規模の全ゲノムデータ（30億塩基対）を活用した、疾患関連遺伝子の効率的な探索技術等による、オーダーメイド医療や早期診断、効果的治療法の確立
- ・地球環境分野では、様々な要因が複雑に絡み合う地球規模課題の解決に貢献し持続可能な社会を構築するため、地球温暖化、森林や水などの自然循環、生態系、地理空間等の異なるデータ間の関係性を高度につなぎ合わせる基盤的情報技術の確立
- ・防災分野では、災害・事故から得られた気象、地理空間等のデータを容易に分析可能な形に蓄積・構造化する技術等による精緻な災害の予測や防災機能強化の推進、都市の最適設計手法の高度化等

の実現を目指す。これらの実現によって、イノベーションによる新産業・新市場の創出や、国際競争力の強化を推進し、第4期科学技術基本計画（平成23年8月19日閣議決定）の「我が国の産業競争力の強化」、「研究情報基盤の整備」の達成に貢献することを目指す。

4. 具体的内容

(背景)

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（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※1によれば、2020年には、約40ゼタバイト（2010年度時の約50倍）へ拡大する見込みである。また、情報通信政策研究所の調査※2によると、日本における平成21年度の流通情報量は7.61E21ビット（一日あたりDVD約2.9億枚相当。例えば、E18ビットは10の18乗であることを示している。）であるが、消費情報量は2.87E17ビット（一日あたりDVD約1.1万枚相当）であり、流通に対して消費された情報量は0.004%にしかすぎない、と言われている。

その質的・量的に膨大なデータ（ビッグデータ）には新たな知識や洞察を得られる可能性があるが、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）を組み合わせて、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、現況においてはその多くのデータが整理・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（data centric science = e-サイエンス）は第四の科学的手法と言われ※3、ビッグデータ時代における科学の新たな地平を拓（ひら）く方法論として注目されている。

(研究内容)

本戦略目標では、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創出等のため、2つの達成目標の実現を目指す。具体的には以下の研究を想定する。

- (1) 各アプリケーション分野においてビッグデータの利活用を推進しつつ様々な分野に展開することを想定した次世代アプリケーション基盤技術の創出・高度化

個別のアプリケーション分野の課題解決とともに、固有技術の他分野展開や新規基盤要素技術の導入を強力に推進する。このため、情報科学・数理科学分野とアプリケーション分野の研究者等による協働研究チーム体制を構築することが期待される。具体的には、以下の研究を推進する。

- ・多様かつ大量のアプリケーションデータ（健康・医療データ、地球観測データ、防災関連データ、ソーシャルデータ等）の転送、圧縮、保管等を容易に実現するための研究
- ・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
- ・アプリケーションデータから新たな課題の発見や洞察をより正確に行うための研究（疾患要因の解明、気候変動予測、リアルタイム解析による減災、人のニーズの予測等）
- ・定量データから生体、自然現象等に係る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創出

- (2) 様々な分野のビッグデータの統合解析を行うための次世代基盤技術の創出・高度化・体系化
情報科学・数理科学分野や人文科学の研究者による、独自の新規基盤要素技術の創出や複数のアプリケーション分野に展開する新規要素技術の創出を行う。具体的には、以下の研究を推進する。

- ・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）や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
- ・高度な圧縮技術、圧縮したままで検索する技術、秘密性や匿名性を損なわないままマイニングす

る技術

- ・データマイニング技術や機械学習の高度化（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等）
- ・多様なアプリケーションデータの相関や関係性から新たな洞察を導くための可視化技術
- ・ビッグデータを共有・流通するためのシステム技術（データの加工，メタデータ管理，トレーサビリティ，匿名化，セキュリティ，課金等）
- ・課題の本質やビッグデータの構造を見いだすための数理的手法

なお，(1) の次世代アプリケーション基盤技術の創出・高度化に当たっては，(2) の研究で得られる次世代基盤技術を取り込みながら推進することが効果的であり，また，(2) の次世代基盤技術の創出・高度化・体系化に当たっては，(1) の研究で得られる次世代アプリケーション基盤技術やデータを共有，活用しながら研究を進めることが効果的であることから，(1) と (2) の研究が相互に連携することが求められる。

※1 IDC, “Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East”, 2012.12

※2 情報通信政策研究所調査部「我が国の情報通信市場の実態と情報流通量の計量に関する調査研究結果（平成 21 年度）－情報インデックスの計量－」，平成 23 年 8 月

※3 Tony Hey, Stewart Tansley, and Kristin Tolle, The Fourth Paradigm: Data-intensive Scientific Discovery, (Microsoft Research 2009)

5. 政策上の位置付け（政策体系における位置付け、政策上の必要性・緊急性等）

第 4 期科学技術基本計画では、「我が国が直面する重要課題への対応」において、「我が国の産業競争力の強化」として，電子デバイスや情報通信の利用，活用を支える基盤技術等，革新的な共通基盤技術に関する研究開発を推進するとともに，これらの技術の適切なオープン化戦略を促進すると掲げている。また，「科学技術の共通基盤の充実，強化」として，シミュレーションや e-サイエンス等の高度情報通信技術，数理科学等，複数領域に横断的に活用することが可能な科学技術や融合領域の科学技術に関する研究開発を推進すると掲げている。さらに，「国際水準の研究環境及び基盤の形成」において，「研究情報基盤の整備」として，研究情報基盤の強化に向けた取組を推進するため，研究情報全体を統合して検索，抽出することが可能な「知識インフラ」としてのシステムを構築し，展開すると掲げている。

文部科学省では，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，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，平成 24 年 4 月から 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，7 月に提言「ビッグデータ時代におけるアカデミアの挑戦」において，ビッグデータに関する共通基盤技術の研究開発として，ビッグデータ処理の各段階（データ収集，蓄積・構造化，分析・処理，可視化）における基盤技術の研究開発等が必要との方向性を取りまとめた。

6. 他の関連施策との連携および役割分担・政策効果の違い

平成 24 年 10 月に科学技術政策担当大臣及び総合科学技術会議有識者議員による「平成 25 年度科学技術関連予算重点施策パッケージ」の選定が行われ，総務省，文部科学省，経済産業省の 3 省合同で提案した「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た基盤整備」が資源配分

の重点化を行うべき重点施策パッケージとして特定された。この重点施策パッケージでは、3省が連携して平成28年頃までの実現を目指したある一定の分野におけるビッグデータの収集・伝送、処理、利活用・分析に関する基盤技術の研究開発及び人材育成を一体的に進めることとしている。

このうち、文部科学省は「次世代 IT 基盤構築のための研究開発」の一プログラム「ビッグデータ利活用のためのシステム研究等」を、重点施策パッケージの個別施策として位置付け、異分野融合型研究拠点によるデータサイエンティスト等の人材育成や国際連携を進めるとともに、データ連携技術等の技術開発課題やアカデミッククラウド環境（大学等間でクラウド基盤を連携・共有するための環境）構築の在り方に関する検討を行うこととしている。また、独立行政法人科学技術振興機構はビッグデータ活用モデルの構築のため、死蔵されている膨大なデータの掘り起こしやルールの整備を行い、研究機関のデータベース連携や民間等での利活用を推進することとしている。上記施策に加え、分野を超えたビッグデータの利活用を可能にするため、本戦略目標では、中長期的な視野で次世代の課題解決に向けた共通基盤技術の高度化・体系化のための研究を行う。

また、総務省では、平成24年5月に情報通信審議会 ICT 基本戦略ボードにおいて、「ビッグデータの活用の在り方について」を取りまとめ、情報通信インフラの構築を進めているため、本戦略目標下の研究を推進する際には、当該インフラ（独立行政法人情報通信研究機構（NICT）が構築・運用するテストベッド（JGN-X））も必要に応じて活用する。

7. 科学的裏付け（国内外の研究動向を踏まえた必要性・緊急性・実現可能性等）

米国においては、2011年に科学技術に関する大統領諮問委員会（PCAST）が、連邦政府はビッグデータ技術への投資が少ないと結論づけたことに対応し、科学技術政策局（OSTP）が2012年3月29日にビッグデータイニシアチブに関する公告を発表した。このイニシアチブには6機関（NSF, NIH, DOD, DARPA, DOE, USGS）が総額2億ドルを投資し、データへのアクセス、体系化、知見を集める技術を改善、強化している。欧州、アジアにおいても、ビッグデータに対する研究投資を実施しており、今後、激しい国際競争が予想される。具体的には、欧州では2020年までにICTにおける研究開発への公共支出を55億ユーロから110億ユーロへと倍増させ、大規模なパイロットプロジェクトを実施し、公共に利益のある分野における革新的かつ相互運用可能なソリューション（エネルギーや資源を節約するためのICT、持続可能な保険医療、電子政府、インテリジェント輸送システム等）を開発することとしている。また、中国では情報資源を共有するためのセンターを設置し、収集したデータの相互の関係付けのためにメタデータの付与や自動分類等の技術開発を行っている。さらに、韓国ではビッグデータを含む研究データの共有とデータ科学を推進する National Scientific Data Center を2013年から構築することとなっている。このことから、官民の役割分担と省庁の枠を越えた連携のもと、科学技術分野におけるイノベーションの推進等に向け、分野を超えたビッグデータの利活用を促進するための研究開発が急務となっている。

我が国は、各種センサー情報が発達していること、ハイパフォーマンスコンピューティング、自然言語処理等、世界的に高い研究水準を有する関連研究領域があることや、遺伝子情報等の地域単位での研究が必要な大規模データを扱う領域にも取り組んでいる。このことから、大規模データの活用において、これらの強みが幅広い分野・領域に展開することで、科学技術における共通基盤の強化や産業競争力の強化が可能な環境である。

8. 検討の経緯

文部科学省の研究振興局長の下に設置したアカデミッククラウドに関する検討会においては、平成24年7月4日に提言「ビッグデータ時代におけるアカデミアの挑戦」を取りまとめ、ビッグデータに関する共通基盤技術の研究開発として、ビッグデータ処理の各段階（データ収集、蓄積・構造化、分析・処理、可視化）における基盤技術の研究開発等が必要との方向性や具体的な研究開発事

ERATO 川原万有情報網プロジェクト 戦略目標

項について取りまとめた。

これを踏まえ、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（第 77 回，第 78 回）（平成 24 年 7 月 5 日，8 月 2 日）においても，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により，新たな知的価値を創造する「データ科学」が重要との共通認識のもと，ビッグデータを利活用するための共通基盤技術の研究開発が必要との見解が示された。

また，科学技術・学術審議会先端研究基盤部会（第 5 回）（平成 24 年 8 月 7 日）で取りまとめられた「数学イノベーション戦略（中間報告）」においては，ビッグデータを有効に活用するための革新的な手法や技術を開発するには，数学研究者は情報科学分野の研究者や各アプリケーション側の研究者と積極的に連携を図るとともに，数学研究者の多様な知見とポテンシャルを最大限活用し，ビッグデータの有効活用において本質や構造を見いだすための共通基盤的技術の構築に向けて取り組むことが重要と述べられている。

本戦略目標は，これらの検討の結果を踏まえて作成したものである。

9. その他

本戦略目標を推進するに当たっては，情報科学・数理科学分野とビッグデータの利活用が有効な様々な研究分野の融合により，ビッグデータに関係する研究者に流動的なネットワークを生み出し，新たな人材育成スキームや，イノベーション創出サイクル（常にイノベーションを創出し続ける環境）の構築も目指すことを期待する。