戦略的創造研究推進事業
－チーム型研究(CREST)－

研究領域「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進のための次世代アプリケーション技術の創出・高度化」

研究領域中間評価用資料

研究総括：田中　譲

2018 年 2 月

目　次

1. 研究領域の概要

(1) 戦略目標

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将来実現し得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利活用して，研究開発や実用化を推進することで，例えば

a. 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，オーダーメード医療や早期診断，効果的治療法の確立

b. 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

c. 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

(i)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と，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。

(ii)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。

a. 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

b. 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

c. 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・体系化を行う。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，a 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，b 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，c の次世代基盤技術の創出・高度化・体系化に当たっては，a 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，a と b 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, Stwart 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及び役割分担・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，強化するとしている。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。

(2) 研究領域

「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進のための次世代アプリケーション技術の創出・高度化」（平成２５年度発足）

ICT の社会浸透や，実世界から情報収集するセンサーや計測・観測機器の高度化と普及に伴い，様々な分野で得られるデータは指数関数的に増大し，多様化し続けている。これらのビッグデータの高度な統合利活用により,新しい科学的発見による知的価値の創造や,それらの知識の発展による社会的・経済的価値の創造やサービスの向上・最適化などにつながる科学技術イノベーションが期待されている。

本研究領域では，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，科学的発見および社会的・経済的な挑戦的課題の解決や革新的価値創造のために，個々の研究者や組織のみでは集積することが困難な大規模かつ多様な関連データを相互に関連付けて高度な統合的分析処理を行うことにより，これらのビッグデータに隠されている革新的知見や価値を抽出し創成することを実証的に研究開発する。そのために必要な次世代アプリケーション技術を実証的に創出・高度化することを目指す。

具体的には，生命，物質材料，健康・医療，社会・経済，都市基盤システム，防災・減災，農林水産業，宇宙地球環境などにおける様々な科学的発見および社会的・経済的な挑戦的課題の解決や革新的価値創造を，ビッグデータを高度統合利活用する革新的技術によって実証的に実現する。単に，既知の基盤技術の適用による知見や価値の創造を目指すのではなく，目的達成に必要な次世代アプリケーション技術を新たに実証的に創出・高度化し，適用分野の特性に応じた総合的かつ統合的なビッグデータ解析システム技術を確立することを目指す。

また，本研究領域では，関連領域の「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」で得られる次世代基盤技術を共有・活用するなどの連携を推進する。

(3) 研究総括

田中　譲（北海道大学　名誉教授）

(4)採択研究課題・研究費

（百万円）

採択年度	研究代表者	中間評価時 所属・役職	研究課題	研究費*
平成 25 年度	船津　公人	東京大学 大学院工学系研究科 教授	医薬品創薬から製造までのビッグデータからの知識創出基盤の確立	392
	三好　建正	国立研究開発法人理化学研究所 計算科学研究機構　チームリーダー	「ビッグデータ同化」の技術革新の創出によるゲリラ豪雨予測の実証	317
平成 26 年度	越村 俊一	東北大学 災害科学国際研究所 教授	大規模･高分解能数値シミュレーションの連携とデータ同化による革新的地震･津波減災ビッグデータ解析基盤の創出	336
	角田 達彦	東京医科歯科大学 難治疾患研究所 教授	医学・医療における臨床・全ゲノム･オミックスのビッグデータの解析に基づく疾患の原因探索･亜病態分類とリスク予測	325
	西浦 博	北海道大学 大学院医学研究科 教授	大規模生物情報を活用したパンデミックの予兆,予測と流行対策策定	281
	吉田 直紀	東京大学 大学院理学系研究科／カブリ数物連携宇宙研究機構 教授	広域撮像探査観測のビッグデータ分析による統計計算宇宙物理学	351
平成 27 年度	大浪 修一	国立研究開発法人理化学研究所 生命システム研究センター チームリーダー	データ駆動型解析による多細胞生物の発生メカニズムの解明	301
	平藤 雅之	東京大学 大学院農学生命科学研	フィールドセンシング時系列データを主体とした農業ビッグデータ	307

		究科　特任教授	の構築と新知見の発見	
	松本 裕治	奈良先端科学技術大学院大学 情報科学研究科 教授	構造理解に基づく大規模文献情報からの知識発見	330
			総研究費	**2940**

*研究費：2017 年度上期までの実績額に 2017 年度下期以降の計画額を加算した金額

2. 研究総括のねらい

ビッグデータを対象とする分析や可視化の個々のアルゴリズムや数理的手法，ソフトウェア・ツールは近年急速に研究開発が進んでおり，その種類も急速に増大している。しかし，科学的発見や社会的・経済的な実際の挑戦的課題とそれに関連する多様なビッグデータが与えられたとき，これらのツールや手法をどのように組み合わせ，どのような手順でどのような分析や可視化を行うことによって課題解決に繋がるのかについては，経験知すらも充分に蓄積されてなく，そのような方法論は科学的にも工学的にもほとんど研究されていない。データ・サイエンスと呼ばれるこの分野を，科学的，工学的に創成し発展させる必要がある。これらも考慮して本研究領域では以下の観点の研究を推進する。

a. 分野や組織を越えた大規模データの統合的分析処理で価値創成

大規模かつ多様な関連データを分野や組織を越えて集積し，相互に関連付けて高度な統合的分析処理を行うことにより，これらのビッグデータからそこに隠されている革新的新知見や価値を抽出し創成することを実証的に研究開発することを目的とする

b. 次世代アプリケーション技術やシステム技術を実証的に創出

既知のアルゴリズムや数理的手法を対象応用分野のビッグデータに適用して何らかの知見や価値の創造を目指すだけでは不充分で，そのような研究開発過程の中で，目的の達成に必要な次世代アプリケーション技術やシステム技術を新たに実証的に創出・高度化・体系化することを目指す必要がある

c. 各種要素技術を組み合わせての分析シナリオ作成

多様な種類のツールを自在に連携活用した試行錯誤的で探索的な分析可視化の繰り返しをどのような革新的技術で支援できるかが重要で，このためには各種要素技術を組み合わせての分析シナリオが必要となる

d. 国として注力すべき応用分野の掘り起しと国際連携

欧州や米国が先行している医療関連，持続可能な社会を構築するための地球環境分野関連，防災機能強化のための災害・事故関連のビッグデータ解析等，特に国として今後注力すべき応用分野の掘り起こしを期待する。そのため海外の研究者やプロジェクトとの連携を積極的に推進する。

e. 再利用可能なノウハウの知識化，データ・サイエンティストの育成

実証的研究を通して，データ・サイエンティストを育てると共に，ノウハウを科学的，工学的に抽出し再利用可能な知識に昇化する努力も望まれる。

f. 個人情報保護に関連するシステム機構の提案も期待

個人情報保護に抵触するデータの取り扱いに関しては法制度的な配慮とそれに整合したシステム機構の提案も含めることを期待する

以上の研究を推進するため以下の研究体制を想定した。

A. 科学的発見，社会的・経済的課題解決をねらう分野の研究者と，情報工学・コンピュータサイエンスの研究者または数学者のチームであること。

B. ビッグデータのオーナーはチームに含まなくてもよいが，実問題の最新のビッグデータが更新も含めて常に利用可能であることと，対象分野の専門家で実データとその分析結果の意味解釈ができる研究者をチームに含めること。

C. 課題解決に必要となる社会科学者や経済学者も積極的に取り入れたチームを期待する。特に個人情報保護に抵触する可能性のあるデータをチーム内で共有・流通して取り扱う場合には，法律の専門家と共同してデータ共有・流通のための制度設計やそれに整合したシステム設計を行うと共に，実施に当たっては特区の利用なども考慮すること。

D. 実ビッグデータを対象として，その処理分析の全過程にわたって総合的実証的に革新的技術を研究開発することを目標とするので，外部機関への外注は極力回避すること。

E. 研究成果の社会への速やかな波及を促進するために，民間企業をチームに組み込んだ共同研究体制をつくることも期待する。

以下に上記の条件を踏まえた研究実施体制の図を示す。

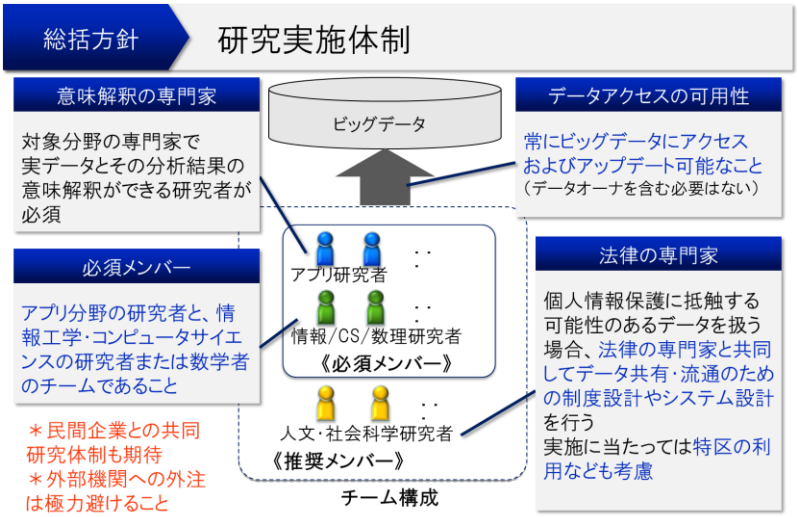

図１　本研究領域で想定する研究実施体制

本研究領域は，同じ戦略目標の下に同時に設定される CREST・さきがけ複合領域「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化（以下，ビッグデータ基盤研究領域)との連携・協同を重視し，二つの研究領域の相乗効果や国内外の研究者のマッチングを推進する。具体的には以下のように運営する。

○　本研究領域と次世代基盤技術研究領域とで領域会議やワークショップなどを共同して行い，多様な分野の研究者で密に情報共有する。

○　次世代基盤技術研究領域へ可能な限りデータや技術を共用・提供する。

○　次世代基盤技術研究領域で創出された共通基盤技術の活用を推進する。場合によっては，次世代基盤技術研究領域からの CREST 共同研究グループとしての参画を受ける。

3. 研究課題の選考について

本研究領域は,「研究総括のねらい」にも述べたように，ビッグデータの複数ドメインに共通する本質的課題を解決し，様々な分野のビッグデータの統合解析を可能にする次世代基盤技術の創出・高度化・体系化を目指した研究を対象として公募した。具体的には，様々な研究分野（アプリケーション分野）において科学的発見及び社会的・経済的な挑戦的課題の解決や革新的価値創造を，ビッグデータを高度統合利活用する革新的技術によって実証的に実現する提案を取り上げることとした。各分野に関しては我が国を代表する１～２のフラグシップ・プロジェクトを採択し，全体として国の重要課題分野をバランスよく含んだアプリケーション分野のポートフォリオを形成することを目指した。

2013 年度から募集を開始し,初年度は幅広いアプリケーション分野からの採択を目指し特に重点分野は設けずに募集を行なった。2014 年度は対象となるアプリケーション分野は限定せず，次の二つの分野を重点分野として公募した。（１）オーダーメイド医療（個人化医療ともいう）を目指したバイオメディカル・ビッグデータの分析技術,（２）防災，減災，災害対策，復興支援のためのビッグデータ応用技術の 2 分野である。2015 年度も対象となるアプリケーション分野は限定せずに,次の３つの分野を優先重点分野として公募した。（１）農業・漁業（生産・流通・販売）へのビッグデータ・アプローチ,（２）大規模な文献情報を推論可能な知識表現に変換して大規模な知識ベースを構築し，有益な知識を推論により発見する技術,（３）大規模システムの設計や機能材料物性におけるビッグデータ・アプローチの 3 分野である。

これらの重点領域における研究提案の掘り起こしのために，公募に先立ち，各重点領域における国内外の著名な研究者による研究セミナーを開催したり，関連の国際ワークショップなどへ総括が出向いて説明を行った。具体的には，2014 年度の公募に向けては，重点項目として挙げた「オーダーメイド医療を目指したバイオメディカル・ビッグデータの分析技術」に関して，欧州連合（EU）の第 7 期フレームワーク・プログラム（FP7）における VPH（Virtual Physiology Human）カテゴリーにおいて p-medicine (Personalized Medicine)プロジェクトの研究代表を務め，後に本研究領域の国際アドバイザに就任して頂いたザーランド大学医学部の Prof. Nobert Graf を招き，p-medicine プロジェクトの概要とビッグデータ解析に基づく個人化医療の研究について講演をいただき，同時に同プロジェクトにおいて，個人情報保護の観点から，技術と制度と運用を統合して，EU および参加国の法律と倫理に適合した解を提案したことで国際的に高い評価を受けているハノーバ大学ライプニッツ法学校の Prof. Nikolaus Forgo と，東大医科学研究所教授で本領域のアドバイザでもある宮野悟教授に講演をして頂いた。2015 年度の公募に向けては，重点項目として挙げた「大規模な文献情報を推論可能な知識表現に変換して大規模な知識ベースを構築し，有益な知識を推論により発見する技術」に関して，本領域の進捗報告会を兼ねた国際シンポジウムに，XML トピックマップの著者で，現在，IBM 社の Watson のような機能を持ったオープンソース版のシステム OpenSherlock を研究開発している Jack

Parc 氏をシリコン･バレーから招いて講演を行ってもらうと共に, 国立情報学研究所(NII)湘南セミナー・ハウスで開催されたテキストからの知識マイニングに関する国際セミナーに総括が参加して，参加者に研究提案を促した。また，ビッグデータ解析に基づく防災・減災・災害対応と農業に関しても，英国大使館とフランス大使館が開催したビッグデータ関連の 2 国間ワークショップや, 米国 Arlington で NSF と JST が共同開催した災害ビッグデータの共同研究立ち上げワークショップに総括が参加し，研究代表者の候補となり得る研究者に研究計画提案を促した。

その結果, 当初考えていた応用分野のポートフォリオに関しては, 2015 年度公募で求めた「大規模システムの設計や機能材料物性におけるビッグデータ・アプローチ」を除いては, 予定していた重点領域をほぼカバーするようなプロジェクトを採択することができた。カバーすることができなかった機能材料物性におけるビッグデータ・アプローチに関しては，本研究領域の公募終了後に関連する CREST 研究領域が複数立ち上がっており，今後は，これらの領域の関連する採択プロジェクトを領域シンポジウムに招待するなどの方策を考えている。

3 回にわたる公募に対して，生命，健康・医療，社会・経済，都市基盤システム，防災・減災，農林水産業などさまざまな範囲に渡る幅広い研究分野から各年度 60 件，38 件，20 件の延べ 118 件の応募があった。これらの研究提案を初年度は 6 名，残りの年度は 8 名の領域アドバイザと 1 名の外部評価者の協力を得て書類選考を行った。2014、2015 年度は面接選考会には 4 名の国際・領域運営アドバイザにも参加いただき発表・質疑応答，審査会議を全て英語で実施した。審査に当たっては，応募課題の利害関係者や，他制度の助成金などとの関係も留意し，公平・厳正に行なった。

2013 年度は予算の関係もあり採択課題は次の 2 件と少ない採択数となった。「創薬プロセスにおける，医薬品候補物質探索のための超大規模仮想化合物ライブラリ構築および複数タンパク質と膨大なリガンド間相互作用解析，製薬プラントの品質安定化のためのリアルタイム高精度ソフトセンサからの知識抽出」,「「ビッグデータ同化」という技術革新の創出により観測データを 30 秒ごとに更新しリードタイム 30 分のゲリラ豪雨や竜巻等の局地的天気予報を可能とするシステムの開発と実証実験」。2014 年度の採択課題は 3 つの重点分野を含む次の 4 件となった。「臨床・全ゲノム・オミックスのビッグデータの解析に基づく個人化医療」,「大規模・高分解能数値シミュレーションの連携とデータ同化による地震・津波減災」,「広域撮像探査観測のビッグデータ分析によりダークマターの正体に迫る統計計算宇宙物理学」,「大規模生物情報を活用したパンデミックの予兆・予測」。2015 年度の採択課題は 2 つの重点分野を含む次の 3 件となった。「フィールドセンシング時系列データを主体とした農業ビッグデータの構築と新知見の発見」,「構造理解に基づく大規模文献情報からの知識発見」,「データ駆動型解析による多細胞生物の発生メカニズムの解明」。いずれも，各分野の研究におけるフラグシップとなり得る，革新的な科学的ないし社会的価値創造を目指す研究提案を採択することができた。いずれの研究提案も，優れた実績を

持つ研究者チームにより研究が進められる計画になっており，大きな社会的インパクトに結び付く成果が期待される。

データ利用に関する法的，倫理的配慮が必要な課題においては，充分な配慮がなされている点も高く評価した。特に個人医療データの取り扱いには，インフォームド・コンセントの取得と，学内倫理委員会の許諾の取得を始め，極めて慎重な取り扱いを研究代表者に指示し，データの取り扱いの不備によって，研究途中でデータが利用できなくなるといったトラブルが生じないよう，十分な配慮をした。

初年度公募では，応募数も多く，採択には至らなかったものの，個人情報の取り扱いの不備の是正や，研究計画の明確化の努力をすることで，改善可能な提案もいくつかあり，それらの中でも優れた提案 3 件を特定課題調査として選んで予算をつけ，1 年間の予備研究を進めてもらった。その結果，この内の 1 件が第 2 年度に採択となった。第 2 年度の公募の審査においても，不採択となったものの，出口を明確にすることで改善可能と判断した 1 件に特定課題調査の予算をつけ予備研究を行ってもらったが, 充分な改善には至らず, 第 3 年度に採択とはならなかった。

以下にビッグデータ応用の研究分野ポートフォリオを示す。

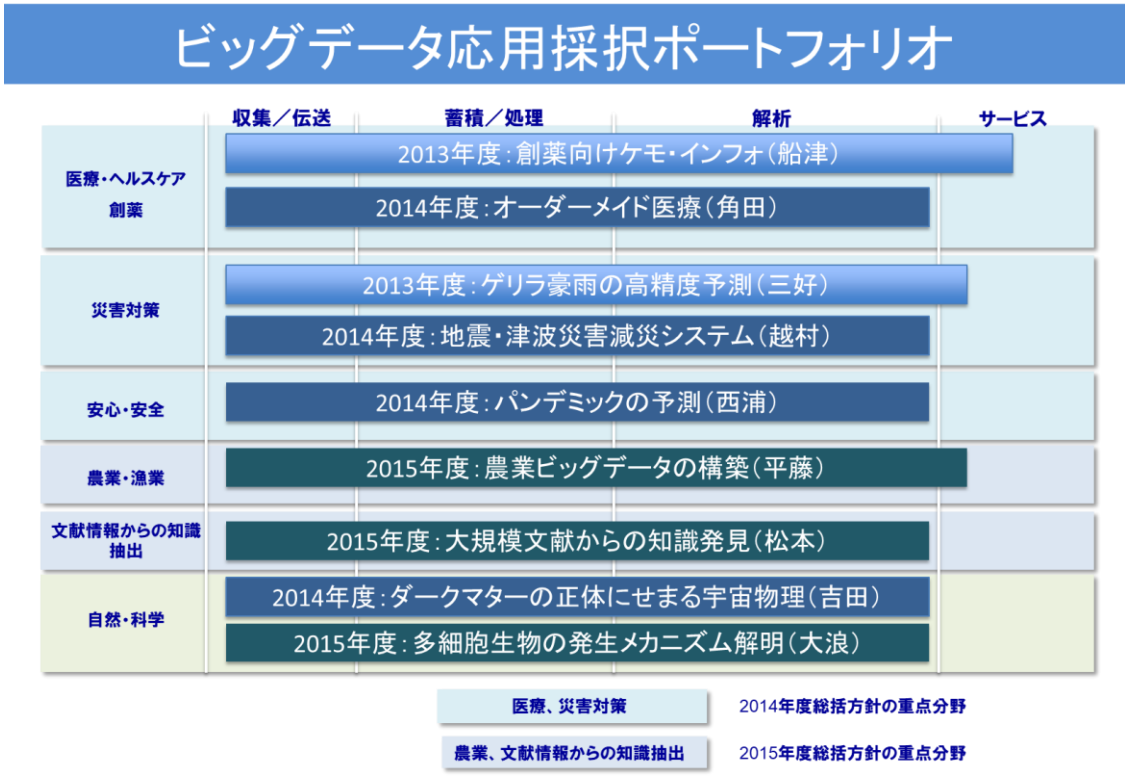

図 2　本研究領域がカバーするアプリケーション分野のポートフォリオ

各公募年度の面接選考においては，各提案に対し，計画の改善や拡張に関する要望を積極的に提示した。これらの計画改善や拡張に関する要望や指示と，その効果に関しては，「6 .研究の経過と所見」に各プロジェクトごとに記述する。

4. 領域アドバイザについて

(1) 領域アドバイザ

アドバイザ名	現在の所属	役職	任期
天野　肇	特定非営利活動法人 ITS Japan	専務理事	平成２６年４月～平成３３年３月
岩野 和生	三菱ケミカルホールディングス	執行役員	平成２５年４月～平成２８年３月
柴崎　亮介	東京大学　空間情報科学研究センター	教授	平成２５年４月～平成３３年３月
下田　正文	株式会社 DNA チップ研究所 事業企画	顧問	平成２６年４月～平成３３年３月
鈴木　良介	株式会社野村総合研究所 ICT・メディア産業コンサルティング部	主任コンサルタント	平成２５年４月～平成３３年３月
武田　浩一	名古屋大学　大学院情報学研究科／附属価値創造研究センター	教授／センター長	平成２９年４月～平成３３年３月
西浦　廉政	東北大学　材料科学高等研究所	特任教授	平成２５年４月～平成３３年３月
松井　知子	統計数理研究所　モデリング研究系	教授	平成２５年４月～平成３３年３月
宮野　悟	東京大学　医科学研究所	教授	平成２５年４月～平成３３年３月

(2) 国際・領域運営アドバイザ

アドバイザ名	現在の所属	役職	任期
Costantino Thanos	Italian National Research Council　Research	Director	平成２５年４月～平成３３年３月
Norbert Graf	Saarland University Hospital	Professor, Doctor, Director	平成２６年４月～平成２８年３月

Nicolas Spyratos	University of Paris Sud 11	Professor Emeritus	平成２６年４月～平成３３年３月
Nigel Waters	University of Calgary	Professor Emeritus	平成２５年４月～平成３３年３月
Randolph Goebel	University of Alberta	Professor	平成２５年４月～平成３３年３月

(3) 領域アドバイザ選定方針と国際アドバイザリ・ボードの設置

生命，健康・医療，社会・経済，都市基盤システム，防災・減災，農林水産業などさまざまな範囲に渡る幅広い研究分野にビッグデータの観点で評価やコメントができるように幅広い専門分野の人選をおこなった。2013 年度は２次以降の募集に向けてある程度専門分野の異なる領域アドバイザを増やす余力を残しておきたいとの考えで６名から始めた。産業への貢献と技術の両面からの評価・コメントをいただけることを念頭に産業界から２名，大学から３名，国立の研究機関から１名に就任いただいた。ビッグデータの応用観点から生体医学情報学の宮野悟教授，空間情報科学の柴崎亮介教授，ビッグデータ・ビジネスに造詣が深い鈴木良介氏に就任いただき，技術面からは，コンピューターサイエンスの岩野和生氏，統計的学習理論の松井知子氏，数学および数理モデリングの西浦廉政教授にお願いした。

また，国際的視野で研究を進めることができるように CREST 研究領域では初めて国際アドバイザリ・ボードを設置することとし，海外の著名な研究者に国際・領域運営アドバイザをお願いした。2013 年度は地理情報科学の専門家で当時ジョージメイソン大学地理情報 COE(Center of Excellence)の拠点長であった Nigel Waters 教授，機械学習と可視化分析，及びシステム生物学の研究者で，カナダのアルバータ州の研究助成機関副所長でもある Randy Goebel 教授，イタリアのピサにある国立研究所（CNR）の元所長で，欧州委員会（EC）の委員会委員を歴任し，EC がファンドする EU プロジェクトの多くの科学コーディネータも勤め，EU のサイバー・リサーチ・インフラストラクチャに関する白書の主たる著者でもある Costantino Thanos 教授の合計３名に就任いただいた。

2014 年度は公募で（１）オーダメイド医療を目指したバイオメディカル・ビッグデータの分析技術，（２）防災，減災，災害対策，復興支援のためのビッグデータ応用技術，の２つを重点分野としたため，この分野に関連する領域アドバイザ２名を新たに依頼した。具体的には，生体医学情報の下田正文氏，ITS 及びモビリティ情報の天野肇氏に就任頂いた。医療関連では書類審査段階で医学関連１名に外部識者の評価をお願いした。

研究の国際化を一層促進するために国際・領域運営アドバイザも，オーダーメイド医療分野の国際的パイオニアであり EU プロジェクト p- medicine（期間：2011～2015）

の研究代表者を務めた小児腎臓ガンの専門家である Norbert Graf 教授，データベース並びにビッグ・データの基礎理論の研究者で，これらに関わる欧米の研究者に強力な人的ネットワークを持つ Nicolas Spyratos 教授の２名を増員した。Spyratos 教授は，後の 2017 年に，ギリシャ研究イノベーション国家協議会委員コンピュータ・サイエンス担当責任者に任命されている。

2016 年度には産業界のアドバイザ１名（岩野アドバイザ）が退任したため，企業経験があり IBM Watson の開発への参画など人工知能の開発に携わってきた武田浩一教授に 2017 年度より新たに就任いただき，2018 年１月現在，８名の領域アドバイザ及び５名の国際・領域運営アドバイザで運営している。就任時，米国のジョージメイソン大学教授であった Waters 教授が，カナダのカルガリー大学に戻ったため，米国の研究者を新たに加えるべく，現在，NSF のビッグデータ担当責任者である Chaitan Baru 教授に人選をお願いしている。

国際アドバイザを設けたことにより，本研究領域の面接審査，判定会議，中間評価，シンポジウムの公用語はすべて英語としている。

5. 研究領域の運営について

研究領域の運営に当たっては，選考時に以下の３点

A. ポートフォリオの設計

我が国におけるデータ駆動型研究の推進を目指し，重要なビッグデータ応用分野を網羅するようなポートフォリオを形成する。

B. フラグシップ・プロジェクトの選定

各アプリケーション分野ごとに，その分野のフラッグシップ・プロジェクトとなるようなプロジェクトを採択し，研究を加速・拡充する。

C. 応用と基盤技術の両分野の研究者の緊密な共同研究

プロジェクト内で，アプリケーション分野の研究者と，計算機科学および統計学や数学を専門とする研究者が密に連携して課題解決にあたる研究計画であること。

を，重視した。プロジェクト採択後は，個々のプロジェクトがアプリケーション分野において顕著な研究成果を達成するだけでなく，以下の目標を領域の最重要共通課題として，全プロジェクトに周知徹底した。

D. 共通応用基盤技術の抽出と確立

異なるプロジェクトがターゲットとする多様なアプリケーション分野に共通する考え方や方法論，システム技術など，これまで，ビッグデータ基盤技術の研究においては充分に研究されてこなかったものの，ビッグデータ応用の共通分母となるような重要なテクノロジーを明確に抽出して，より広い応用範囲に適用可能なジェネリックな考え方や技術に育成し，共通応用基盤技術として確立する。

さらに，この目標に向けた研究活動の活性化のために，以下の４つの方策を実施した。

a. 若手研究者合宿ワークショップ

各プロジェクトに属する若手研究者を中心に，種々の応用分野を専門とする研究者と，計算科学および統計学や数学を専門とする研究者が，プロジェクト横断的に，分野横断的に，研究課題を相互に理解し，考え方や方法論，さらにはシステム技術を相互に再利用し，新しいソリューションの創出へとつなぐような密な議論を行う場を提供する。

具体的には，2 日間の若手合宿ワークショップを，当初は年 1 回，現在は年 2 回行っている。毎回，参加者の研究発表と招待講演を行い，これらに関して活発な議論ができるよう討論に充分な時間を割り当てている。参加者は研究代表者に推薦していただき，徐々にメンバーを増やし，30 名程度を上限に，同じメンバーが少なくとも 2 年間は継続して参加できるように運用している。毎回，活発な議論が続き，プログラム終了後も深夜まで議論や交流が続いている。招待講演者として招いた産業技術総合研究所 AIST の辻井潤一所長は，このような分野横断的な活発な研究討論の機会は他では得られないとの感想を述べられ，以降の合宿にも継続して参加して頂いている。

参加者からは，参加するごとに，共通応用基盤技術の内容が徐々に見えてきたとの感想が寄せられている。

この合宿は現在までは日本語で行っているが，参加者からは海外の研究者を招待講演に招いて，英語で議論することを望む声もある。現在，実施に関して総括が検討している。

b. 研究領域国際シンポジウム

領域の国際シンポジウムを年 2 回開催し，国内アドバイザのみならず，国際アドバイザにも可能な限り参加していただき，英語による活発な議論を行う。

9 月のシンポジウムは中間評価会議翌日に開催し，全プロジェクトが進捗報告を行うと共に，キーノート講演を海外より 1 件，国内より 1 件招待する。

1 月の国際シンポジウムは 2 日間とし，総括が計画するキーノートセッションでは海外から 2 件のキーノート講演者を招き，各プロジェクトには各々1 時間 20 分のセッションを海外からの招待講演 1 件とプロジェクト進捗状況報告で構成してもらう。これにより，国際アドバイザ 5 人と，キーノート 2 名，招待講演 9 名の計 16 名の国外の著名な研究者が議論に参加することになり，毎回，極めて活発な議論が交わされている。彼らが牽引役となって，応用分野に跨った議論が活発に行われるようになり，当初は自身の報告が終了次第会場を去ることが多かった研究代表者も，多くが最初から最後まで議論に参加するようになってきた。

本領域では，これら 9 月と 1 月の国際シンポジウムに加えて，ビッグデータ基盤領域と合同で，JST と NSF の共同主催になるシンポジウムも共催している。本年度の場合は，ここでも，9 プロジェクトがすべて発表している。

このように，領域が主催ないし共催する国際シンポジウムが年に 3 回あり，研究代

表者にはかなりの負担をかけているが，複数の研究代表者から，これらがプロジェクト間に跨る相互理解を深め，分野横断的な新しい共通基盤技術の明確化に大きく役立っているとの意見を頂いている。

c. 全プロジェクトの体験型ポータルの開発と公開

船津プロジェクトの中に，体験型ポータル研究開発グループを設置し，船津プロジェクトのみならず，9 つのすべてのプロジェクトの成果を，分析対象のデータと共に分析ツールやサービスも含めてウェブ上に公開し，プロジェクト内の研究者だけでなく，研究領域内の他のプロジェクトの研究者や，ビッグデータ基盤 CREST 研究領域の研究者，さらには一般の方々も，利用できるよう体験型ポータル環境を開発し公開することとした。利用者は体験型ポータル環境を介して，各プロジェクトの実際のデータ群を用いて，プロジェクトが開発した種々の分析ツールやサービスを実データに適用し，有益な知見の発見に至る分析過程をインタラクティブに体験することができる。

これにより，自身の専門とは異なるアプリケーション分野におけるデータ駆動型研究というものがどのように行われているかを体験することができるだけでなく，自身が開発した新しいツールやサービスを持ち込んで用意されているデータに適用したり，逆に，異なる分野で開発されたツールやサービスを，自身の持つデータ集合に自在に適用することが可能になる。

体験型ポータルの整備により，分野を超えた共通の必須応用基盤技術の理解と研究開発が促進できると考えている。

d. 国内外のプロジェクトとの共同研究の促進

AIP ネットワークラボの制度を活用して，領域内のプロジェクト間の共同研究，ビッグデータ基盤領域の CREST やさきがけプロジェクトとの共同研究，米国 NSF（National Science Foundation）の関連プロジェクトとの共同研究に関して，研究計画の提案を促し支援を行った。

同時に，国際アドバイザの人的ネットワークを活用して，国際共同研究のパートナーの紹介を行った。具体的には，総括から角田プロジェクトに Graf 教授との共同研究を，松本プロジェクトに Goebel 教授との共同研究を提案し，吉田プロジェクト，越村プロジェクトには各々，Spyratos 教授，総括及び Waters 教授がギリシャと米国の研究者を紹介した。吉田プロジェクトでは既に共同研究が開始されている。

三好プロジェクトはビッグデータ基盤 CREST 研究領域の松岡プロジェクトと連携しており，船津プロジェクトはビッグデータ基盤さきがけ研究領域の田部井プロジェクトと連携を行っている。吉田プロジェクトは独自開発の高性能データアクセス技術を三好プロジェクトに提供して連携を行っている。平藤プロジェクトはアイオワ大学との共同研究を進めている。大浪プロジェクトは領域シンポジウムのプロジェクト企画セッションの招待講演者として招いた UCSD の George Sugihara 教授と共同研究を

開始している。

また，個々のプロジェクトの研究加速や社会実装の研究開発強化のために，研究課題中間評価の結果を踏まえて必要と考える研究費を追加した。例えば，三好プロジェクトには当初計画にはなかった予測精度評価を継続的に行えるようにナウキャストシステム用のクラスタ計算機購入費用 16,000 千円を追加した。この追加は世界初となる 30 秒更新 10 分後までの降水予報のリアルタイム実証の開始につながっている。降水予報は気象庁の許可を得て，理研がインターネット上で発表している。
研究費追加やシンポジウム以外でもサイトビジットを通じて研究加速や社会実装の研究開発強化のために助言を行っている。

6. 研究の経過と所見

（1）各プロジェクトの研究の経過と所見

各プロジェクト毎の研究の経過と所見を述べ，その後に，領域全体としての研究の経過と所見を述べる。本研究領域では研究課題を推進する組織をプロジェクトと称する。プロジェクトは研究代表者グループと主たる共同研究者グループから構成される。

①船津プロジェクト

（研究課題）

医薬品創薬から製造までのビッグデータからの知識創出基盤の確立

（研究目標）

医薬品創薬から製造までの過程には蓄積された膨大な測定データ等が存在する。これまで異分野として個別にとらえられていた創薬の現場と製造の現場における知見および各種データを共有する仕組みを構築するとともに，創薬・製造を俯瞰的に見た医薬品開発のシステム全体の効率化および最適化を目指した研究を進める。具体的にはこれらのデータを活用することで，大量のタンパク質 対 化合物情報からの創薬指針の抽出，大規模仮想ライブラリ創出およびそこからの新薬ターゲット発見とその合成・製造法の獲得，製造プラントの安定運転・リスク事前管理・品質安定化のための知識抽出を達成し，医薬品創薬から製造の段階を通した知識創出基盤を確立することを目標とする。

（研究成果のポイント）

・化合物-タンパク質相互作用予測に DNN を初めて応用した化合物-タンパク質相互作用予測手法を開発した（Masatoshi Hamanaka, Kei Taneishi, Hiroaki Iwata, Jun Ye, Jianguo Pei, Jinlong Hou, and Yasushi Okuno. ″CGBVS‐DNN: Prediction of Compound‐protein Interactions Based on Deep Learning.″ Molecular Informatics, 2016）。

・スーパーコンピューター「京」を活用した世界最大規模の高速・超並列バーチャルスクリーニング，および大規模分子動力学シミュレーションを活用したタンパク質-化合物結合親和性の高精度予測に関し，190 億ペアの化合物-タンパク質相互作用計算を行い，実験結果と比較すると実際の結合パターンと類似するパターンを得た。分子動力学シミュレーション計算では，キナーゼや GPCR などのタンパク質とそれに結合するとわかっている複数の化合物での検証により，計算値が実験値と相関のある結果を得た(Brown, J.B., Nakatsui, M., Okuno, Y., “Constructing a foundational platform driven by Japan’s K supercomputer for next-generation drug design” Molecular Informatics, 2014）。
・化学品製造装置の運転状態をオンラインで連続監視する研究の一環として，異常状態の迅速な判断が可能となる手法を提案した。また，多次元ベクトルとして表現される運転の様子を可視化する仕組み（Generative Topographic Mapping(GTM)）の有効利用についても同時に提案し，その運用事例を示した（Matheus de Souza Escobar, Hiromasa Kaneko and Kimito Funatsu, ″On Generative Topographic Mapping and Graph Theory Combined Approach for Unsupervised Non-linear Data Visualization and Fault Identification″, Computers & Chemical Engineering, Volume 98, 113-127, 2017）。

（研究成果と総括助言）

奥野グループが研究開発した化合物-タンパク質相互作用予測手法（Masatoshi Hamanaka, Kei Taneishi, Hiroaki Iwata, Jun Ye, Jianguo Pei, Jinlong Hou, and Yasushi Okuno. ″CGBVS‐DNN: Prediction of Compound‐protein Interactions Based on Deep Learning.″ Molecular Informatics, 2016）は化合物-タンパク質相互作用予測に DNN を応用した初めての事例である。従前は，この予測にはリガンドとなる化合物とタンパク質のドッキングを第一原理計算に基づきシミュレーションするドッキング・シミュレータが用いられてきた。これには計算時間を要することから，多数の化合物と多数のタンパク質の膨大な組み合わせの中から，ドッキングの可能性がある組み合わせ候補を妥当な時間で篩い落とすことは不可能であった。そこで，タンパク質に対しては PROFEAT を用いて特徴空間にマッピングし，化合物に関しては DRAGON を用いて特徴空間にマッピングすることにより，これらの組み合わせを 2 つの特徴空間の直積空間にマッピングすることにより，200 万～300 万件のドッキングに関する既知のデータ集合を学習データとして用いて機械学習を行い，未知の組み合わせに対しては学習結果を用いてドッキングの可能性の有無を推定する手法を確立した。機械学習に当初はサポート・ベクトル・マシン（SVR）を用いていたが，その後，ディープ・ラーニングに変更して大きく性能を向上させた。

総括からは，複数のドッキングサイトが存在するような場合(promiscuity)の配慮も行うことと，毒性チェックの機能を盛り込むことを助言している。

創薬向け仮想化合物ライブラリとしては，その規模や指向性（ターゲット指向か多様性指向か）或いは構築方法（組み合わせ枚挙か反応準拠か）といった様々な切り口があり，国内外を問わず様々な取り組みが行われている。それらの初期の研究成果は発表されるも

のの恐らくは製薬会社内のプロプライエタリな資産として続報が途絶えてしまうものも見られる。学術的に公開されている多様性指向の大規模仮想化合物ライブラリの代表的な例としては，GDB-17 があり 1664 億以上の低分子化合物を含んでいる。これは規模において船津プロジェクトの仮想化合物ライブラリを上回っているが，個々の化合物のアクセシビリティについては特に示唆を与えるものではない。泰地グループが研究開発した仮想化合物ライブラリはスーパーコンピュータを用いた計算により，種となるリガンドの集合に対し，それらの間に可能な化学反応を仮想的に繰り返し適用することによって数十億以上の化合物を求めたもので，このように大規模でありながら個々の化合物の合成経路を示唆する情報を保持している点が異なり，創薬・製剤に向けた優位性があるといえる。

候補となった化合物の合成可能性に関しては，堀グループが計算科学の手法に基づく検証システムを開発している。

プロジェクト内各グループのビッグデータ技術を組み合わせ連携することにより，個々だけでは成し得ないスケールでの医薬品創薬・製造知識創出基盤の確立が期待できる。すなわち創薬・製造を俯瞰的に見た医薬品開発のシステム全体の効率化および最適化を達成することが期待できる。

一方，こうして得られた化合物の製造を行う化学プラントのオンライン監視は，製品製造および品質管理にとって必須の事項である。船津グループが担当するソフトセンサー研究は，製薬産業からも高い評価を受けており，すでに企業で実用化されている。医薬品連続生産プロセスはアメリカの FDA や我が国の PMDA の推奨もあり，今後製薬業界にとって大きな流れとなることが予想されているが，その要となるプロセスのオンライン連続監視技術への取り組みが遅れている。船津グループは，製薬プラントを含む化学プラントなどの化学品製造装置の運転状態をオンラインで連続監視する研究の一環として，異常状態の迅速な判断が可能となる手法を提案した。また，多次元ベクトルとして表現される運転の様子を可視化する仕組み（Generative Topographic Mapping(GTM)）の有効利用についても同時に提案し，その運用事例を示した（Matheus de Souza Escobar, Hiromasa Kaneko and Kimito Funatsu, ″On Generative Topographic Mapping and Graph Theory Combined Approach for Unsupervised Non-linear Data Visualization and Fault Identification″, Computers & Chemical Engineering, Volume 98, 2017）。

プロジェクト全体に対し，総括からは，開発技術の統合化の方針を明確にすること，特に，堀グループによる候補化合物の合成反応経路の精査検証と船津グループのプロセス監視の間の相互連携の明確化を指示している。これらの助言に対し，統合化が計画に新たに加えられ，検討が進んでいる。

（今後の展開・波及効果と総括の所見）

製薬企業のニーズと IT 企業の AI 技術を結び付け，ライフサイエンス分野における AI 技術を開発・活用するためのコンソーシアムの枠組みにより，本プロジェクトで開発した技術を産業や社会への展開・実装がスムーズに進むことが期待できる。合成経路情報の示

唆が可能な数十億規模の仮想化合物を含むライブラリを構築し公開することにより，製薬会社や研究者などの利用者がより高確率でリード化合物の探索に成功することが期待される。総括の助言に基づき，統合された創薬・製造知識を利用者がワンストップで獲得・利活用可能となるような仕組みを実装していくことが現在目指されており。連携プラットフォームを構築し公開することを計画している。

既知の化学反応は，量子化学計算を用いて詳細な解析が盛んにおこなわれている。これに対し，新規化合物の合成反応は，実際に合成を行ってみなければわからないとされている。船津プロジェクトでは，このような未知の合成反応の解析を，実験を行う前に実施する（*in silico* スクリーニングと呼ばれる）ことにより，合成反応の可能性を予測できることを明らかにしてきた。このような研究は，世界的にみても研究例は少なく，そのオリジナリティは高い。本プロジェクトでは，未知の合成反応の解析を短時間で行えるシステム（*in silico* スクリーニングシステム）の開発を進めている。

本プロジェクトで取り組み開発されてきたソフトセンサー技術に関しては，すでに多くの化学工業会社でオンライン実装の取り組みがスタートしている。また，すでに実装が完了し実運用に入っている企業もある。化学製品の品質管理を伴いながら，安定した生産を実施することは，化学企業の強い願いでもあるが，本研究成果がそれを実現する後押しとなっている。本ソフトセンサー技術は世界最高水準にある。船津グループが 2016 年 9 月の化学工学会秋季年会において発表したソフトセンサー構築オフラインツールの講演に対して，化学工学会 SIS 部会技術賞が授与されている。

将来的には，本プロジェクトの成果によって，創薬ターゲットが次々と明らかにされていくことが見込まれるが，その折に触れて仮想化合物ライブラリおよび連携プラットフォームの両者が適宜利用可能であることが重要であるため，継続的な利用を可能とする計算・運用資源の確保及び更新・増強を含めた内容物の維持管理が今後必要となる。候補化合物が薬として世に出るまでの臨床治験研究や毒性チェックとの連携も今後重要である。

②三好プロジェクト

（研究課題）

「ビッグデータ同化」の技術革新の創出によるゲリラ豪雨予測の実証

（研究目標）

本研究では，次世代の高精細シミュレーションと新型センサによる「ビッグデータ」を扱うための「ビッグデータ同化」の技術革新を創出し，ゲリラ豪雨予測に応用して，30 秒毎に更新するリードタイム 30 分の天気予報という従来では考えられない画期的なシステムを，フェーズドアレイ気象レーダ，次期気象衛星ひまわり，京コンピュータという我が国が世界に誇る次世代技術を駆使して実証実験する。これにより，ビッグデータ利用の基盤技術を確立し，ゲリラ豪雨や竜巻等の防災・減災に資するとともに，気象学的ブレークスルーをもたらす。

（研究成果のポイント）

・「30 秒毎に更新する 100m メッシュの 30 分予測」という天気予報に革命をもたらす「ビッグデータ同化」の技術基盤を確立し，現時点では 100 秒で処理の完了を実現。国内外の類似研究と比べて桁違いのスピードと解像度の天気予報を過去事例で実証（Takemasa Miyoshi, Masaru Kunii, Juan Ruiz, Guo-Yuan Lien, Shinsuke Satoh, Tomoo Ushio, Kotaro Bessho, Hiromu Seko, Hirofumi Tomita, and Yutaka Ishikawa, “″Big Data Assimilation″ Revolutionizing Severe Weather Prediction”, Bulletin of the American Meteorological Society, vol. 97, no. 8, 2016)。

・通常 100 個以下程度で行われる全球大気のアンサンブルデータ同化を，スーパーコンピュータ「京」を使って世界最大規模の 10240 個まで増やし，理想的なデータ同化実験に成功した。大気の誤差構造の非ガウス性を直接確認したほか，1 万 km を超える地球規模の誤差共分散を発見（T. Miyoshi, K. Kondo, and T. Imamura: The 10240-member ensemble Kalman filtering with an intermediate AGCM. Geophys. Res. Lett., 41, 2014)。

・フェーズ・アレイ・レーダ画像の時間変化をオプティカル・フローとして求め，これの機械学習により，30 秒毎に更新する 10 分後までの降水予測行う「３D 降水ナウキャスト手法」を開発（S. Otsuka, G. Tuerhong, R. Kikuchi, Y. Kitano, Y. Taniguchi, J. J. Ruiz, S. Satoh, T. Ushio and T. Miyoshi: Precipitation Nowcasting with Three-Dimensional Space-Time Extrapolation of Dense and Frequent Phased-Array Weather Radar Observations. Weather and Forecasting, 31, 2016)。

（研究成果と総括助言）

ゲリラ豪雨は局所的に起き，10 分も経たない間に急激に発生・発達する雨雲によって生じる豪雨で，これまでの気象予報技術では予測が極めて困難なことから，事前に避難警報を出すことが難しく，これまで度々痛ましい犠牲者を出している。

本プロジェクトでは，ゲリラ豪雨予測に関して，最先端の気象レーダ技術であるフェーズ・アレイ・レーダを用いた雨雲の 3 次元分布の実時間観測データと，世界最先端のスーパーコンピュータである京コンピュータを用いた高精度大規模アンサンブル・シミュレーションとを，研究代表者自身が独自に開発した局所アンサンブル変換カルマンフィルタ（LETKF）と呼ばれる高度な革新的手法でデータ同化することにより，「30 秒毎に更新する 100m メッシュの 30 分予測」という天気予報に革命をもたらす「ビッグデータ同化」の技術を確立し，実際の過去のゲリラ豪雨事例での実証結果を得た。国内外の類似研究と比べて桁違いのスピード，解像度の天気予報を過去事例で実証し，気象学の国際コミュニティで高く評価され，今後の科学技術に大きなインパクトを与えるものとして，米国気象学会のフラッグシップ誌（Bull. Amer. Meteor. Soc.）の 2016 年 8 月号に論文が掲載され（Takemasa Miyoshi, Masaru Kunii, Juan Ruiz, Guo-Yuan Lien, Shinsuke Satoh, Tomoo Ushio, Kotaro Bessho, Hiromu Seko, Hirofumi Tomita, and Yutaka Ishikawa, “″Big Data Assimilation″ Revolutionizing Severe Weather Prediction”, Bulletin of the American

Meteorological Society, vol. 97, no. 8, 2016)，また IEEE のフラッグシップ誌（Proc. of the IEEE）の 2016 年 11 月号ビッグデータ特集号の招待論文として掲載された（T. Miyoshi, G.Y. Lien, S. Satoh, T. Ushio, K. Bessho, H. Tomita, S. Nishizawa, R. Yoshida, S. A. Adachi, J. Liao, B. Gerofi, Y. Ishikawa, M. Kunii, J. Ruiz, Y. Maejima, S. Otsuka, M. Otsuka, K. Okamoto, H. Seko: "Big Data Assimilation" toward post-petascale severe weather prediction: an overview and progress. Proceedings of the IEEE, 2016)。

通常 100 個以下程度で行われる全球大気のアンサンブルデータ同化を，スーパーコンピュータ「京」を使って世界最大規模の 10240 個まで増やし，理想的なデータ同化実験に成功した。大気の誤差構造の非ガウス性を直接確認したほか，1 万 km を超える地球規模の誤差共分散を発見。例えば日本から 1 万 km 遠方の観測データが，瞬時に日本の大気状態の推定精度を向上する可能性が明らかとなり，天気予報シミュレーションの改善に貢献することが期待されている。

また，初年度サイトビジットの際の総括からの助言を受けて，アンサンブル・シミュレーションとのデータ同化の代わりに，フェーズ・アレイ・レーダ画像の時間変化をオプティカル・フローとして求め，これの機械学習により，30 秒毎に更新する 10 分後までの降水予測を行う「３D 降水ナウキャスト手法」の開発も研究計画に加えて研究開発を行った（Otsuka, S., G. Tuerhong, R. Kikuchi, Y. Kitano, Y. Taniguchi, J. J. Ruiz, S. Satoh, T. Ushio and T. Miyoshi: Precipitation Nowcasting with Three-Dimensional Space-Time Extrapolation of Dense and Frequent Phased-Array Weather Radar Observations. Weather and Forecasting, 31, 2016)。その結果，最新鋭気象レーダを生かした「3D 降水ナウキャスト手法」を開発し，30 秒毎に更新する 10 分後までの降水予報のリアルタイム実証を 7 月 3 日に開始した。30 秒更新の降水予報は世界でも例がない超高頻度であり，国内外の競合技術と比較して技術の創造性・先行性・優位性は他に例がない。スマホアプリ「3D 雨雲ウォッチ～フェーズドアレイレーダ～」を開発運用する株式会社エムティーアイ（エムティーアイ）と共同研究を実施しており，本研究で開発した 3D 降水ナウキャストによる 10 分後予測情報を，スマホアプリを通じてリアルタイム配信を開始した。

このような革新的，かつ科学的にも国民の生活や経済活動にも極めて有益な研究成果に対し，2014 年度科学技術分野の文部科学大臣表彰若手科学者賞，2014 年度地球惑星科学振興西田賞，2016 年度日本気象学会賞などを受賞し，その研究成果は資料編に示すように多くのメディアで取り上げられている。

（今後の展開・波及効果と総括の所見）

現状では，技術は確立したものの，30 秒で終えるべき計算に 100 秒程度かかっており，今後，4 倍程度の計算高速化が必要である。解像度を落とす，アンサンブル数を減らす，更新間隔を延ばす，という計算量低減策を検討し，予報精度を落とさず計算時間を 30 秒以内に抑える設定を現在探っている。これにより，「京」でリアルタイム実行可能なゲリラ豪雨予測手法の実現が期待できる。

本プロジェクトに対しては，アドバイザ，国際アドバイザの評価は極めて高く，国際競争力のあるオンリー・ワンの技術として，本プロジェクト終了後も継続的に一層の研究支援をすべきと総括は考えている。

③越村プロジェクト

（研究課題）

大規模・高分解能数値シミュレーションの連携とデータ同化による革新的地震・津波減災ビッグデータ解析基盤の創出

（研究目標）

世界最先端の災害シミュレーション，防災学，数理科学，情報科学の研究者が連携し，将来の国難となる地震・津波災害で一人でも多くの命を救うことを目標に，大規模・高分解能リアルタイム数値シミュレーションの連携とリアルタイム観測データ同化による，世界初のビッグデータ解析基盤・減災システムを創出することを目指している。これにより災害における最悪シナリオやそれを回避するための方策をリアルタイムで提示するとともに，政府・自治体等の防災システムへの実装を果たすことを目標としている。

（研究成果のポイント）

・海底センサ出力から沿岸部の津波高を即時推定の枠組みを，ガウスプロセスを用いた非線形回帰によって構築した。多様なシナリオの震源モデルによって海溝型巨大地震を対象とした数値シミュレーションしたデータを用いて本枠組みに適用することで，予測誤差を従来に比べて 30%程度減少させることに成功（Yasuhiko Igarashi, Takane Hori, Shin Murata, Kenichiro Sato, Toshitaka Baba and Masato Okada, ″Maximum tsunami height prediction using pressure gauge data by a Gaussian process at Owase in the Kii Peninsula, Japan″, Marine Geophysical Research, Vol. 37, No. 4, 2017)。

・災害被害シミュレータ（建物被害，道路閉塞，火災延焼）と人間行動シミュレータ（地域住民による救助・消火活動，広域避難行動）を連動させて，様々なシナリオのもとで精緻なマルチエージェント・シミュレーションを実行し，木造住宅密集地域に潜在する脆弱性を明らかにした（Takuya Oki, Toshihiro Osaragi,″Wide-area Evacuation Difficulty in Densely-built Wooden Residential Areas″, Proceedings of the ISCRAM 2016 Conference - Rio de Janeiro, Brazil, May 2016, Tapia, Antunes, Bañuls, Moore and Proto de Albuquerque, eds., 2016)。

（研究成果と総括助言）

研究代表者が従前より開発してきた津波シミュレータ技術を骨子に，災害発生時の人流シミュレーション,火災伝搬シミュレーション,これらのシミュレーションを用いた人的,物的損失の推定，リモートセンシング・データや，合成開口レーダデータからの建物の損壊状況の推定等，災害予測と災害時の損失推定に関する多岐にわたる要素技術を精力的に研究開発し，個々の要素技術としては後ほど論文，特許を引用して述べるように高い完成

度を達成している。

総括からは，2016 年度サイトビジットの際に，要素技術をバラバラに研究するだけではなく，対象地域を共通に設定し，これまで開発してきた要素技術がどのように連携適応されるのかに関して，具体的に統合システムを開発して明らかにし，その上で，このようなシステムが，災害の前後，際中，直後において，行政，救援組織，共同体，市民グループ，個々人の各々に対して，どのようなサービスを提供しうるのかを明確にするよう指示した。地域を決めた統合化に関しては即座に取り組みが行われ，統合化のデモも既に一部示されている。後者に関しても，災害に対する各時間フェーズに，誰を，あるいはどの組織を対象にどのようなサービスを提供するのかの議論が進み，チャートにまとめられている。

注目すべき成果として，海底センサ出力から沿岸部の津波高を即時推定する枠組みを，ガウスプロセスを用いた非線形回帰によって構築した。多様なシナリオの震源モデルによって海溝型巨大地震を対象とした数値シミュレーションを行って得たデータを用いて本枠組みに適用することにより，予測誤差を従来に比べて 30%程度減少させることに成功した。これに関して発表した論文（Yasuhiko Igarashi, Takane Hori, Shin Murata, Kenichiro Sato, Toshitaka Baba and Masato Okada, ″Maximum tsunami height prediction using pressure gauge data by a Gaussian process at Owase in the Kii Peninsula, Japan″, Marine Geophysical Research, Vol. 37, No. 4, 2017）は，Springer Nature 社の Change the World, One Article at a Time に，Earth and Environmental Sciences 分野の論文として唯一選出され，科学技術の進歩に資する革新的研究成果として高く評価されている。

研究代表者が従前より取り組んできたスーパーコンピュータによるリアルタイム津波浸水予測・被害推計の技術が内閣府の災害対応システムの機能として採用され，産学連携での構築を終えて 2017 年 11 月より運用を開始した。南海トラフ域で発生する地震の発生直後に総距離 6,000Km におよぶ太平洋沿岸地域における津波被害の推計を，約 30 分以内で行うものであり，短時間で津波浸水被害推計を行うシステムは世界初である。現在，ソースコードをスタンドアローンで動作するマシンに移植することが検討されており，これが完成すれば，多様なユーザニーズに対応する浸水予測システムの展開が世界的に可能となる。フォワード型（どこでも取得可能な地震情報のみを利用してリアルタイムで浸水計算を行う）技術は，全世界への展開が可能である。本技術は，特許第 6161130 号・「津波浸水予測システム，制御装置，津波浸水予測の提供方法及びプログラム」として平成 29 年 6 月 23 日に登録された。また，2 件の特許が現在審査中である（特願 2016-122638，特願 2016-122639）。

災害被害シミュレータ（建物被害，道路閉塞，火災延焼）と人間行動シミュレータ（地域住民による救助・消火活動，広域避難行動）を連動させて，様々なシナリオのもとで精緻なマルチエージェント・シミュレーションを実行し，木造住宅密集地域に潜在する脆弱性を明らかにした（Takuya Oki, Toshihiro Osaragi, ″Wide-area Evacuation Difficulty in Densely-built Wooden Residential Areas″, Proceedings of the ISCRAM 2016 Conference

- Rio de Janeiro, Brazil, May 2016, Tapia, Antunes, Bañuls, Moore and Proto de Albuquerque, eds., 2016)。大規模な都市圏レベルにおいて，データ同化手法によりリアルタイムに災害時の人々の移動を短時間予測する手法を研究開発している。本研究で考案した同化技術を用いた人流推定手法は，モバイルセンシング技術の発展に伴い，我が国のみならず特にアジア諸国でのインフラ開発や災害対策の推進に有効であり，社会的要請が極めて高い。この技術については，特許出願「特開 2017-49954 推定装置，推定方法及びプログラム」を行っている。

（今後の展開・波及効果と総括の所見）

研究は当初計画通り進んでおり，各要素研究について順調に成果が得られている。今後はシミュレーションとセンシングを融合し，地震および津波被災地の被害状況を予測・把握・開示するシミュレーション基盤の創出に取り組むことで，研究の最終目標を達成する計画である。一方，すべてのシミュレーション要素が，同時に社会実装・実用化を果たすことは難しい。自治体等で導入・運用となるには，技術が優れているだけでなく，ユーザ側の予算措置のタイミングなど難しい面がある。領域総括からのアドバイスを受けて，活用技術の研究(ユーザニーズやマーケットの要求に応える研究)が進められる予定である。

本プロジェクトも，特に国や地方自治体と連携した研究成果の社会実装に向けた貢献に対して国際アドバイザの評価が高く，実サービスの提供に向けて，継続的発展が可能なように総括として支援したいと考えている。

④角田プロジェクト

（研究課題）

医学・医療における臨床・全ゲノム・オミックスのビッグデータの解析に基づく疾患の原因探索・亜病態分類とリスク予測

（研究目標）

47 疾患 33 万症例のバイオバンク，患者由来の多層オミックス，リウマチと肝炎の前向き観察コホート等の国内屈指のゲノム・臨床データと外部の多種多様な分子データを加え国際的にもユニークなビッグデータを構成し，最先端の統計学・情報学を駆使した統合解析を戦略的に行い，個人ごとの疾患発症，薬剤効果，副作用の予測と予防が可能なオーダーメイド医療の根幹の確立をめざす。併せて近未来の新しい医療ビックデータの解析基盤を支える技術の開発研究を行う。

（研究成果のポイント）

・独自に開発したがん多層オミックス解析手法を国際がんゲノムコンソーシアム・全日本チームが産出した肝がん 300 症例の全ゲノム・多層オミックスデータに適用し，６種の亜病態の分類法を新たに発見した。さらに，再発・予後などの臨床情報と有意に相関する分子マーカーの発見に成功し，再発の起こりにくい症例と，それを特徴付ける変異のある，新規がん関連遺伝子 MACROD2，そして繊維化との関係を発見した。その研究成果を発表し，

実臨床での応用方法を示した（Fujimoto A+, Furuta M+, Totoki Y+, Tsunoda T+ et al. “Whole genome mutational landscape and characterization of non-coding and structural mutations in liver cancer”, Nature Genetics, 48(5), 2016)。
・健康者から得られた５種類の免疫細胞と未分化末梢血の eQTL データと公共エピゲノムデータを組み合わせることで，個体内の遺伝子発現を予測し，(従来のゲノムワイド関連解析とは異なり)遺伝子レベルでの関連解析と細胞特異的なパスウェイの予測が可能となった。これを間接リウマチに適用し，新たなリウマチ関連遺伝子とサイカインパスウェイの発見に成功した。多くのリスクアレルが細胞特異的な遺伝子発現を制御している生活習慣病に対して，細胞特異的なパスウェイへのポリジーン効果の解析に成功した(Ishigaki K, et al. “Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis”, Nature Genetics, 49, 2017)。
・多発性骨髄腫のランダム化臨床試験において，サリドマイド投与により延命する患者を分類する新しいマーカー解析法を，セミパラメトリック階層混合モデルを用いて開発し，効果予測遺伝子のスクリーニングに成功した。開発した階層混合モデル解析は，オミックスデータと臨床データの背後にある関連構造の推定を可能とし，がん以外にも多遺伝的疾患のゲノムワイド関連解析や二型糖尿病の薬剤効果予測マーカー解析など，既に多くの適用事例があり，オミックスデータの汎用的解析技術と治療効果予測モデルとしての確立が期待される（Matsui S, et al. “Multi-subgroup gene screening using semi-parametric hierarchical mixture models and the optimal discovery procedure: Application to a randomized clinical trial in multiple myeloma”, Biometrics, 2017.)。
（研究成果と総括助言）
本プロジェクトに関しては，個人情報保護の観点から，法律と倫理を順守することを総括が採択以前から研究代表者に強く求め，特定課題調査の予算をつけて，1 年間，インフォームド・コンセントの取得や倫理員会の許諾取得等，充分な準備を行ってもらった。これにより，プロジェクト遂行中にデータの利用に関してマスコミからの批判を受けるなどの問題が生じて利用ができなくなることがないように配慮した。
これまでの成果としては，肝がん 300 症例の全ゲノム・多層オミックスデータに適用し，6 種の亜病態の分類法を新たに発見し，再発の起こりにくい症例と，それを特徴付ける変異のある，新規がん関連遺伝子 MACROD2，そして繊維化との関係を発見した（Fujimoto A+, Furuta M+, Totoki Y+, Tsunoda T+ et al. “Whole genome mutational landscape and characterization of non-coding and structural mutations in liver cancer”, Nature Genetics, 48(5), 2016)。この成果は，メディアも大きく取り扱い（日本経済新聞（2016 年 4 月 12 日)，毎日新聞(2016 年 4 月 12 日)，読売新聞(2016 年 4 月 12 日)，朝日新聞(2016 年 4 月 12 日)，共同通信 47NEWS（2016 年 4 月 12 日)，CBnews（2016 年 4 月 12 日)，夕刊フジ（2016 年 4 月 12 日)，日刊工業新聞（2016 年 4 月 12 日)，朝日新聞（2016 年 4 月 13 日)，化学工業日報(2016 年 4 月 13 日)，産経新聞(2016 年 4 月 18 日)，日本経済新聞(2016

年 4 月 24 日),「ライフライン 21 がんの先進医療」(2016 年 4 月 30 日),「がんサポート」(2016 年 6 月 16 日), NHK 総合 (2016 年 9 月 3 日)), 研究代表者は国内外で多くの招待講演に招かれている。

関節リウマチの研究では, 健康者から得られた 5 種類の免疫細胞と未分画末梢血の eQTL データと公共エピゲノムデータを組み合わせることで, 個体内の遺伝子発現を予測し, 遺伝子レベルでの関連解析と細胞特異的なパスウェイの予測が可能となり, 新たなリウマチ関連遺伝子とサイトカイン・パスウェイの発見に成功している (Ishigaki K, et al. “Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis”, Nature Genetics, 49, 2017)。この成果もメディアに大きく取り上げられた (日刊工業新聞 (2017 年 5 月 30 日), 日経産業新聞 (2017 年 6 月 1 日), 薬事日報 (2017 年 6 月 12 日))。

プロジェクト内連携や臨床医や病理医などとの連携により,疾患の原因探索,病態分類, 予後予測は, がん, リウマチなどで顕著な成果が得られ, 治療奏効も, サリドマイド, メトフォルミンなどで, 顕著な成果が得られている。

今後, 多くのがん多層オミックス解析から, 化学療法などの治療法の選択のためのマーカーセットおよび予測アルゴリズムの構築がなされ, 個別化医療の道筋が確立すると期待される。

本プロジェクトは, 個人化医療におけるビッグデータ・アプローチを想定して重点領域分野として公募したものであるが, 研究実施において, 当初は臨床試験と連携して, 化学療法に対する患者個々人の反応をも併せて分析対象とする取り組みがなく,このままでは, コホート・データを用いたゲノム解析を中心とするコホート研究に収束することになるのではないかと, 総括ならびに国際アドバイザからの憂慮があった。そこで, サイトビジットや国際シンポジウムの際に, 繰り返し臨床試験と連携し, 患者のオミックスデータといくつかの候補治療法の有効性とを関連付けた分析の取り組みを指示した。本プロジェクトが昨年度領域シンポジウムのプロジェクト企画セッションに招待した講演者からも, 化学療法に対する患者の反応データを結び付けた分析の必要性が指摘されたこともあり, 研究代表者の努力によって, 本年度よりほぼ 1 年間にわたって, 臨床試験と連携したデータ取得の環境を整え, 単なるコホート・スタディではない, 臨床試験研究におけるオミックス解析にも重点的に取り組む大きな軌道修正が行われた。今年度 1 月のシンポジウムで, 新たな取り組みとそのデータを用いた進捗状況が発表されたが, 総括と共に, 国際アドバイザでこの分野の専門家である Norbert Graf 教授を始め, 全員の国際アドバイザがその進展を高く評価した。本プロジェクトに対する憂慮事項が解消されたとの意見の一致を見た。もともと, 超一流の国際学術誌に多くの論文を発表しているグループであり, 唯一の課題が, コホート・スタディから臨床試験研究へのシフトであったので, わが国では実施が難しい後者の取り組みを真剣に始めたことに, 総括として大きな期待をしている。

多発性骨髄腫のランダム化臨床試験において, サリドマイド投与により延命する患者を

分類する新しいマーカー解析法を，セミパラメトリック階層混合モデルを用いて開発し，681 個の効果予測遺伝子のスクリーニングに成功した（Matsui S, et al. “Multi-subgroup gene screening using semi-parametric hierarchical mixture models and the optimal discovery procedure: Application to a randomized clinical trial in multiple myeloma”, Biometrics, 2017.）。これにより，サリドマイド治療が有効な患者の選別が可能になる。開発した階層混合モデル解析は，オミックスデータと臨床データの背後にある関連構造の柔軟な推定と視覚化を可能とし，多くの疾患のオミックスデータの汎用的解析技術としての確立が期待される。

乳がんに関し，本邦における 81 例の早期乳がん患者の 10 年以上の長期追跡データを用いて,予後予測のシステムを新たに開発し,日本人症例および欧米人症例の両方において,予後の予測精度の向上に多大な寄与が見られる成果が得られた（*Cancer Medicine*, 6, 1627-1638 (2017)）。現在，MammaPrint®を始めとする海外で開発された予後予測システムが国内で多く用いられているが，日本人における予測精度の限界が指摘されており，待望されている本邦発の予後予測システムの開発に繋がるものと期待される。

（今後の展開・波及効果と総括の所見）

成果の産業や社会への展開，実装として，開発したプログラム「VCMM」,「HML」,「SIML」,「SP-HMM」,「Arete」を公開している。本研究によって改良が進んだ臨床シークエンス用プログラム・パイプラインは，国立がん研究センターの臨床研究である臨床シークエンスプロジェクト TOPGEAR にて使用され，その研究成果を踏まえた実用化に向けて準備が進んでいる。一部のプログラムは三井情報開発に技術移転されている。総括の助言を真摯に受け入れ，コホートスタディから，日本では実施が困難な認証試験における個々の患者のオミックスとの候補治療に対する反応の解析を行う臨床試験研究にも注力するようになったことから，今後一層の研究の発展と大きな成果が期待できる。

⑤西浦プロジェクト

（研究課題）

大規模生物情報を活用したパンデミックの予兆，予測と流行対策策定

（研究目標）

本研究は病原体のゲノム情報や実験データを含む大規模な生物情報を利用したパンデミック予兆の捕捉と流行予測を実現し，それに基づいて最も望ましい感染症対策を明らかにすることを目指している。具体的には，(1) 大規模生物学的情報を取り込んだ流行予測モデルの構築，(2) パンデミックの予兆の探知，(3) これら 2 つのモデルに基づく感染症対策の改善を行う。大規模データを効率的に分析することで，パンデミックの予兆捕捉と流行拡大の予測を世界で初めて日常的に実現する。

（研究成果のポイント）

・2015 年,韓国の複数の医療機関において集団発生した中東呼吸器症候群 (MERS) に関し,

感染者の死亡リスク（致死率）と死亡のリスク因子（基礎疾患の有無や高齢であること等）をリアルタイムで推定・特定することが可能な統計モデルを開発し，それを韓国の観察データに適用した研究成果を発表した（Mizumoto K, Endo A, Chowell G, Miyamatsu Y, Saitoh M, Nishiura H. “Real-time characterization of risks of death associated with the Middle East respiratory syndrome (MERS) in the Republic of Korea”, 2015. BMC Medicine, 2015)。

・これまでに病原体の遺伝子配列を利用して系譜を構築することによって基本再生産数を推定する手法が提案されてきたのに対し，Tajima’s D という中立性のメトリックの時間変化を通じて基本再生産数を推定する新たな手法を提案した。事例研究として 2009 年のアルゼンチンにおけるインフルエンザ（H1N1）の観察データを分析し，基本再生産数を 1.31-2.05 と推定した（Kim K, Omori R, Ito K. “Inferring epidemiological dynamics of infectious diseases using Tajima’s D statistic on nucleotide sequences of pathogens”, Epidemics, pii: S1755-4365(17)30085-3, 2017)。

・2017 年のマダガスカルでの過去最大規模の肺ペストの流行に対し，リアルタイムで疫学データを分析し，一人あたりの感染者が生み出す 2 次感染者数の平均値を意味する基本再生産数を 1.73 と推定した。また，肺ペストの国外輸出リスクは，2017 年 8 月 1 日からの 92 日間，すべての国で 0.1 人未満程度と推定し，極めて限定的であることを証明した（Tsuzuki S, Lee H, Miura F, Chan YH, Jung SM, Akhmetzhanov AR, Nishiura H. “Dynamics of the pneumonic plague epidemic in Madagascar, August to October 2017”, Eurosurveillance)。

（研究成果と総括助言）

西浦グループが遂行している感染症流行の予測研究では，航空機を利用したヒトの移動データを使った感染症流行予測について，ジカ熱，中東呼吸器症候群（MERS）やエボラ出血熱など社会的に重要視されるタイムリーな題材を対象に新規性の高いモデルを実装し，その実装研究の成果をコンスタントに原著論文(Mizumoto K, Endo A, Chowell G, Miyamatsu Y, Saitoh M, Nishiura H. “Real-time characterization of risks of death associated with the Middle East respiratory syndrome (MERS) in the Republic of Korea”, 2015. BMC Medicine, 2015;13:228）として医学誌に発表してきている。また，国内を代表して保健医療関係の行政機関と情報の連携を実施し，世界保健機関（WHO）などの国連機関の専門家として密に連絡しつつ成果情報を発信している。プロジェクト後半では，このような，単なる流行のリアルタイム予測に関する時空間軸上での成功に留まらず，予兆研究に本格的に取り組むことを予定している。これまでに諸外国で気付かれなかった革新的な予兆モデルの構築が期待される。

感染症流行の予測においては，イエメンでのコレラ流行予測や肺ペストの流行動態分析（Tsuzuki S, Lee H, Miura F, Chan YH, Jung SM, Akhmetzhanov AR, Nishiura H. “Dynamics of the pneumonic plague epidemic in Madagascar, August to October 2017”,

Eurosurveillance, 22;pii=17-00710 (doi: 10.2807/1560-7917.ES.2017.22.46.17-00710)) を世界で最初に論文化して報告することに成功しており，感染症流行データのリアルタイムの利活用研究に関して，間もなく国際的に第一級の成果につながると期待している。

インフルエンザの流行予測は社会実装を進めており，地域での医療体制の整備（外来患者数予測はもとより，必要病床数や人工呼吸器数）の参考に利用される予定であり，保健医療行政と連携して進捗するよう工夫している。

今後，インフルエンザ予測に気象データも用いる予定であり，領域内の気象ビッグデータ研究グループ（三好研究代表）との共同研究の計画が進められている。

パンデミックのリアルタイム予測研究では，疫学モデルと遺伝学モデルを統合したかつてない研究成果の創出が検討されている。既にデータ同化に着手しており，統一理論を通じたハイインパクトの予測を目指して研究が進んでいる。

伊藤グループが遂行する病原体のゲノム情報を利用した流行ダイナミクスの推定に関しては，英米と異なる手法での評価に成功しつつあり（Kim K, Omori R, Ito K. “Inferring epidemiological dynamics of infectious diseases using Tajima’s D statistic on nucleotide sequences of pathogens”, Epidemics, pii: S1755-4365(17)30085-3, 2017），理論とアルゴリズムをより深化させることによって殻を破った斬新なハイインパクト研究が創出される可能性がある。

（今後の展開・波及効果と総括の所見）

西浦グループが進めている感染症流行の予測研究に関しては，順調に研究が進み，国際的にも高く評価されている。一方，伊藤グループが進める病原体のゲノム情報を利用した流行株予測などの流行ダイナミクスの推定に関しては，プロジェクト以前の顕著な成果に続く発展がまだ顕著になっていない。総括からは，ウィルスのゲノムのドリフトと共に，人や動物の免疫系の変化も考慮した共進化の観点から，分析を進めてはという助言をしている。

⑥吉田プロジェクト

（研究課題）

広域撮像探査観測のビッグデータ分析による統計計算宇宙物理学

（研究目標）

「統計計算宇宙物理学」というフロンティアを本研究で切り拓くことを目指している。地上大型望遠鏡を 5 年間用いて 25 兆ピクセルにおよぶ膨大な画像データを取得し,最新の機械学習と統計数理，大規模コンピューターシミュレーションを駆使して解析し，目には見えない宇宙のダークマター分布を 3 次元で明らかにする。その解析の過程において，大規模化する宇宙探査によるビックデータと情報統計学を融合させた新領域で次世代アプリケーション技術を開発する。

（研究成果のポイント）

・機械学習の手法を用いてすばる望遠鏡の Hyper Suprime-Cam の観測で得られた膨大な可視光変動天体の候補から，本物の突発天体を選別するアプリケーションを開発した。機械学習の手法として AUC Boosting，Random Forest 及び Deep Neural Network を用い，学習済の判別プログラムをすばる望遠鏡データ解析パイプラインに組みこんだ。いずれの分類器も True positive rate 90 パーセントの点で False positive rate 1 パーセント程度を達成するよう最適化したことを確かめた。2015 年 8 月の観測では，データ取得直後から選別を行い，1 日以内に 10 個の超新星爆発を発見し天文学コミュニティーに速報することに成功した（Mikio Morii, Shiro Ikeda, Masaomi Tanaka, Nozomu Tominaga, Tomoki Morokuma, Katsuhiko Ishiguro, Junji Yamato, Naonori Ueda, Naotaka Suzuki, Naoki Yasuda and Naoki Yoshida, ″Machine-learning Selection of Optical Transients in Subaru Hyper Suprime-Cam″, Publications of the Astronomical Society of Japan, 2016）。

・すばる望遠鏡の Hyper Supreme-Cam の観測で得られた突発天体候補から，Ia 型超新星のみを検出する方法を提案した。従来知られていた光度曲線フィッティングに基づく方法では，Ia 型超新星の明度の時間変化を表す光度曲線に実観測明度をフィッティングするアプローチを採用しており，必然的に多数の観測が必要であった。これに対し，各波長１回の観測で Ia 型超新星であるかどうかを判定する新手法を考案した。 Ia 型超新星の候補を観測の初期の段階で迅速に絞り込むことができるという利点がある（Akisato Kimura, Ichiro Takahashi, Masaomi Tanaka, Naoki Yasuda, Naonori Ueda, Naoki Yoshida, “Single-Epoch Supernova Classification with Deep Convolutional Neural Networks, ” 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), 2017）。

（研究成果と総括助言）

我が国の天文学および宇宙物理学コミュニティが主導してすすめる最大級の宇宙観測プロジェクト HSC サーベイに対し，そのデータ解析にいち早く統計的手法や機械学習をとりいれ， 2016 年からは実際に観測データ解析を行った（Mikio Morii, Shiro Ikeda, Masaomi Tanaka, Nozomu Tominaga, Tomoki Morokuma, Katsuhiko Ishiguro, Junji Yamato, Naonori Ueda, Naotaka Suzuki, Naoki Yasuda and Naoki Yoshida, ″Machine-learning Selection of Optical Transients in Subaru Hyper Suprime-Cam″, Publications of the Astronomical Society of Japan, (2016) 68 (6): 104)。プロジェクト開始から３年で，計画通りデータ処理パイプラインの最適化とともに，統計解析に必要なアプリ開発を拡充してきた。本プロジェクトが開発したデータ処理パイプラインは，口径 8.10 メートル級の巨大望遠鏡を用いる深宇宙探索としては世界最先端のデータ解析技術を有しており，HSC サーベイが 2019 年に終了しその全データを解析する際には，宇宙の膨張史や宇宙の構造形成成長率など主要な宇宙論パラメータの測定に大きく貢献できると期待できる。さらに，観測データ統計解析に必要な理論シミュレーションデータベースの構築にも機械学

習を導入し, 世界に先駆けて高精度統計解析ツールを開発した。これらのツール群により，宇宙画像取得から統計解析結果出力まで一連の作業を行うことができる。

本プロジェクトでは，超新星の発見と，ダークマターの 3 次元地図作成を目標にしている。観測画像からの超新星の発見には，新たに開発された統計的機械学習と，ディープ・ラーニングが用いられている。これらにより，超新星のタイプ識別が可能になり，特にタイプ 1a 型の超新星に関しては，理論モデルが存在することから，これを用いて画像生成をシミュレーションで行い，ディープ・ラーニングの学習データの不足を補うことで，高い識別力を確立することに成功している（Akisato Kimura， Ichiro Takahashi，Masaomi Tanaka，Naoki Yasuda，Naonori Ueda，Naoki Yoshida，“Single-Epoch Supernova Classification with Deep Convolutional Neural Networks，” 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), pp. 354-359, 2017)。

ダークマターの 3 次元分布地図の作成に向けては，強力な重力場の存在に起因する重力レンズによる天体画像の歪みを捉え，これらの情報を総合的に分析処理することによりダークマターの空間分布を推定する計画である。3 次元分布地図の作成まではまだ道が遠いが，計画に基づいて着実に研究が遂行されている。

データ基盤グループ(川島グループ)の研究成果である高性能トランザクション技法は，今後分散ファイルシステム Gfarm のメタデータ・サーバに導入される予定である。2017 年現在，メタデータ・サーバにおけるトランザクション処理の導入は世界的にも未開拓である。データ基盤グループの開発した並列ログ先行書込法 P-WAL と並行制御法 TicToc の結合方式の性能は，500 万トランザクション/秒(TPC-C ベンチマーク)ほどの性能に達しており，分散ファイルシステムの性能を革新的に向上させる可能性を有している。川島英之はこの研究で，2016 年に情報処理学会 山下記念研究賞を受賞している。

川島グループは，宇宙観測データの統計解析に必要な理論シミュレーションデータ分析を支援するため，並列 STR-R-木を用いた新たな手法と計算コードを開発した。この近傍探索技術は，本研究領域の三好プロジェクトが行っているゲリラ豪雨リアルタイム解析の前処理にも活用できる。データ同化のための Local Ensemble Transform Kalman Filter と川島の並列 STR-R 木コードを結合し，動作性能が確かめられた。 「京」コンピュータ上の SPARC で動作するよう，三好プロジェクトと共に調整を行っており，データサイズを現実的な規模にした場合に「京」上では 35 パーセントもの性能向上が見込まれることが明らかになっている。

（今後の展開・波及効果と総括の所見）

研究構想の宏大さと，着実な研究遂行が評価され，特に国際アドバイザの評価が極めて高い。このような基礎科学の分野における我が国の国際研究貢献の代表例として発展することを総括として期待している。本成果を用いて, 今後, 超新星の発見が加速すると共に，ダークマターの分布に関して大きな知見を得るものと期待できる。これらの成果は宇宙創

成モデルに関する論争にも大きな貢献をすると期待できる。

⑦大浪プロジェクト

（研究課題）

データ駆動型解析による多細胞生物の発生メカニズムの解明

（研究目標）

本研究では，世界最大の遺伝子ノックダウン胚の時空間動態計測データと，公共のゲノム塩基配列情報，エピゲノム情報，mRNA，タンパク質，代謝物等の生体分子の発現と相互作用の情報を統合し，最先端の統計解析技術とデータ可視化技術を活用して，多細胞生物の発生のメカニズムの全貌を解明するデータ駆動型の研究手法を開発することを目指している。これにより，データ駆動型の生命科学研究の基盤を構築し，生命科学を情報科学へ転換することを目指す。

（研究成果のポイント）

・国内外の生命動態の定量計測データと計測に利用した画像データ，および生命動態のシミュレーションデータを集積し共有する世界初の統合データベースを構築した（Yukako Tohsato, Kenneth Hunglit Ho, Koji Kyoda, Shuichi Onami, ″SSBD: a database of quantitative data of spatiotemporal dynamics of biological phenomena″, Bioinformatics vol.33, 2016)。

・表現型特徴因果関係ネットワークは辺数が多い密グラフであるため，これを階層グラフとして可視化した場合に多くの辺交差が現れる。辺交差の発生はグラフの視認性に悪影響を与えることが知られており，これを解消することが課題であった。本論文では，辺交差を削減するために新たな辺集中化アルゴリズムを開発し，表現型特徴因果関係ネットワークへの適用を行った。その結果，53 パーセント以上の辺交差が削減され，大幅な視認性の向上が達成された（Onoue, Y., Kukimoto, N., Sakamoto, N., and Koyamada, K., ″Minimizing the number of edges via edge concentration in dense layered graphs″ IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 6, 2016.)。

・近年のバイオイメージングおよび自動画像認識技術の発展により，様々な生物の胚発生における細胞核の動態に注目した発生解析が可能となった。しかし，多細胞の３次元構造の構築に直接的に関連する個々の細胞の形の動態については，光軸方向の解像度の不足や深部での画質の低下等のため自動計測が困難であった。本研究では，生物学的な条件に基づく最適化問題として画像認識と評価を自動化することにより，線虫 C. elegans 胚の細胞の形の動態を自動計測するアルゴリズムを開発し，手作業による計測との認識の差異を既存手法の約 57%（5.6%）に低減することに成功した（Yusuke Azuma and Shuichi Onami “Biologically constrained optimization based cell membrane segmentation in *C. elegans* embryos”, BMC Bioinformatics vol.18, 2017)。

（研究成果と総括助言）

生物におけるゲノム情報を始めとするオミックス情報の取り扱いは急速に進歩し，データ取得の効率も著しく向上してきた。これらのジェノタイプ（遺伝子型）のデータに対し，細胞レベルのフェノタイプ（表現型）のデータ取得にはこれまで大きな変革はなかった。生物の発生や形態形成の研究においては，卵割や各種器官を形成する細胞分裂がどのように行われるかを観測し，そのプロセスを明らかにすることが重要である。大浪プロジェクトでは，顕微鏡下で観測される生体の細胞分裂の過程のビデオ画像に対し，細胞核と細胞壁を認識する機能を開発し，細胞分裂の過程において，各細胞がどのように分裂し，個々の細胞内で核の位置がどのように変化するかなどを，機械可読な形で数値化して自動的に得られる技術を研究開発してきた。これを用いて，線虫の単細胞から成虫までの細胞分裂過程の全体像を解明することを目指している。すでに卵割プロセスの自動認識と数値化は実現できており，各細胞内での核の位置に関するモーメントなどの指標も自動的に得られるようになっている。種々の遺伝子をノックダウンしたサンプルにこの技術を適用して，遺伝子と形態形成の関係を調べる実験も行っている。この基盤技術の改良にも成功し，細胞壁の明確な認識を可能にした（Yusuke Azuma and Shuichi Onami “Biologically constrained optimization based cell membrane segmentation in *C. elegans* embryos”, BMC Bioinformatics vol.18, 2017)。

線虫の発生プロセスは，多細胞生物の発生プロセスの中で最も理解が進んだプロセスであり,これを対象にビッグデータから推定した因果関係は,実験等による検証が可能である。この推定–検証のサイクルを高速に回すことにより，他の研究開発では実現が困難な，因果推論のアルゴリズムの高度化が期待される。

面接審査の際に総括からは，フェノタイプとジェノタイプの相関分析を行うために，細胞内物質の細胞内や細胞間に跨る拡散の様子も数値的に計測できないかとの要望を出したが，この要望に応え，カルシウムの拡散や遺伝子発現の拡散に関して研究を行い，前者に関しては既に成果を上げている。このような技術が確立すると，生物発生学，形態形成学に大きな貢献をするだけでなく，腫瘍や iPS 細胞に適用することにより，腫瘍学やガンの治療法，再生医療の研究に極めて大きな貢献をすると期待できる。

この期待を裏付けるように，本研究は，バイオイメージデータを活用したデータ駆動型の生命科学研究として国内はもとより，国際的にも最先端の研究開発と評価されており，国内外の様々な学会等から招待講演に招かれている（IEEE PacificVis，日本学術会議，ConBio2017（2017 年度生命科学系学会合同年次大会），日本発生生物学会，2018 年の予定：IEEE PacificVis，日本生物物理学会)。2017 年度は，欧州のバイオイメージ・インフォマティクス技術開発のリーダーである OME プロジェクトの年次ミーティングや，欧米を中心としたバイオイメージング関連技術の世界連携機構である Global BioImaging プロジェクトの年次ミーティングでの招待講演を行っている。細胞の形の動態計測データ，および時空間遺伝子発現データの取得方法も既に確立しており，データの大規模な集積を開始している。

プロジェクト内のデータ可視化グループと密に連携し様々な可視化ソフトウェア等の開発を進めている。このようなソフトウェアの開発者とユーザーの密接なコラボレーションにより，可視化分野においても先導的な成果が出始めている（Onoue, Y., Kukimoto, N., Sakamoto, N., and Koyamada, K.,″Minimizing the number of edges via edge concentration in dense layered graphs″
IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 6, pp.1652-1661, 2016.）。

2017 年度に領域国際シンポジウムのプロジェクト企画セッションの招待講演者として米国より招待した，データからの因果推論やモデル化の世界的第一人者である George Sugihara 博士（UCSD）も，本プロジェクトの目的やデータ等の準備状況に大変興味を持ち，密接な共同研究の実施が計画されている。

データから推定された因果関係の検証や，データと既存の知識を組み合わせたより高度な因果推論の実現のためには，論文等のテキストデータからの知識の抽出が必須であることから，プロジェクト開始当初からテキストデータからの知識抽出の研究開発を実行してきているが，2017 年度より，産総研 AI センターの辻井潤一センター長との共同研究を開始している。線虫の研究関連のテキストは国際データベースとして非常に使いやすくまとまっているため，この線虫を対象として技術開発を行うことで，知識抽出の技術も大幅に向上することが期待される。

バイオイメージ・インフォマティクス技術の発展により，様々な生命現象に関して，生命分子や細胞，組織などの時空間動態の大規模な計測が可能になったが，これらの計測データは研究室や雑誌のウェブサイトから個々に公開されており，第三者からのデータの利用性が低かった。本研究では，国内外の生命動態の定量計測データと計測に利用した画像データ，および生命動態のシミュレーションデータを集積し共有する世界初の統合データベースを構築している（Yukako Tohsato, Kenneth Hunglit Ho, Koji Kyoda, Shuichi Onami, ″SSBD: a database of quantitative data of spatiotemporal dynamics of biological phenomena″, Bioinformatics vol.33, pp.3471-3479, 2016）

（今後の展開・波及効果と総括の所見）

線虫研究，および線虫研究コミュニティは，その蓄積されたデータと知識の量，大すぎないコミュニティのサイズ，メジャーな研究対象としての生命科学における位置取り等を考えても，このような生命科学の研究スタイルの刷新を先導するのに最適なコミュニティと考えられ,総括としてもその研究の発展と波及効果の大きさに大きな期待を寄せている。

このプロジェクトに連携して，理研の別予算で，線虫で取得してきたデータと同種のデータをマウス胚で取得できるようにする技術開発が進められている。マウス胚の結果が出れば，より大きなインパクトを学会に与えることになる。

本研究の成果が，腫瘍や iPS 細胞などにも適用可能になると，がん研究や再生医療研究にも大きなインパクトを与えることになると期待している。

⑧平藤プロジェクト

（研究課題）

フィールドセンシング時系列データを主体とした農業ビッグデータの構築と新知見の発見

（研究目標）

作物は環境の影響を受けて成長しているため，栽培管理，品種改良等では成長量，土壌水分などのデータを長期連続的に収集し，複雑系現象として解析・評価・制御を行う必要がある。本研究では，これまでに蓄積してきたフィールド用センサネットワーク，ドローン等の技術を用いてデータを時系列的に収集し農業ビッグデータを自動構築するシステム及び最適栽培条件等新知見を探索するビッグデータ解析手法を開発する。

（研究成果のポイント）

・ドローンや定点設置型カメラ（フィ―ルドサーバ）で撮影した画像データから作物の糖分の計測を行うための手法として，通常のＲＧＢ画像の他に，夜間，近赤外線 LED を用いて照明することで通常のカメラで近赤外画像を収集する方法を考案した。この４バンド画像と果実の形状に関するデータから果実の糖度を推定する手法を開発した（Xuefeng Wang, Chunyan Wu, Masayuki Hirafuji, Visible Light Image-Based Method for Sugar Content Classification of Citrus, January 2016）。

・撮影時に風，光などの影響によって画像が揺らぐと，画像の合成や 3 次元再構成でエラーが発生し，植被率の推定誤差が大きくなることが分かった。そこで，3 次元再構成データと元画像を組み合わせて植被率の推定精度を上げる手法を提案し，これを実圃場で収集した画像に適用したところ，育種研究者に大量の正確なデータを提供することに成功した（T. Duan, B. Zheng, W. Guo, S. Ninomiya, Y. Guo, S. C. Chapman. (2016) Comparison of ground cover estimates from experiment plots in cotton, sorghum and sugarcane based on images and ortho-mosaics captured by UAV, Functional Plant Biology, 2016.）。

（研究成果と総括助言）

データをやみくもに集めて統合してからマイニングによって得られる新知見は，単に相関が高いだけの無意味な知見ばかりになることが予想されるため，本プロジェクトでは，１）雑種強勢の解明，２）植物共生微生物相，３）小麦の収穫適期判定とタンパク推定を目標としている。１）２）は未解明で，３）は比較的簡単な課題である。

１）と２）は親と Ｆ１の共生微生物相の実験を行うという方向で統合され，順調に成果を挙げている。３）は，農林水産省の AI 関連プロジェクトの課題の一部として実施されて，収穫適期判定については期待通りの成果が出ているが，画像によるタンパク推定は，植物生長，代謝系，環境要因の影響が関わる複雑系型問題であることが明らかになり進捗が滞っている。本プロジェクトではビッグデータの観点からこの問題に取り組み農林水産省プロジェクトを支援する。

機械学習による画像データからの数値的形質データの抽出，機械学習のための真値データの収集，大規模機械学習のためのメタコンピューティング，ドローンの協調制御技術開発のためのテストベッド，ドローン撮影プランの最適化，データビュアー，データ統合ツール等の要素技術開発はいずれも順調に進んでおり（Xuefeng Wang, Chunyan Wu, Masayuki Hirafuji, Visible Light Image-Based Method for Sugar Content Classification of Citrus, PLoS ONE 11(1):e0147419, January 2016），特許申請，国際会議・論文発表が活発に行われている。

本プロジェクトでは，アウトリーチ活動として，「基礎研究でも将来役に立つかもしれない」というスタンスでセミナーを開催している。地元での関心は極めて高い。昨年末に地元自治体と企業が主体となったセミナーが自発的に開催され，招待講演者として研究代表者が講演を行い，JA，農業者，地元企業等が多数，参集し，活発な質疑応答が行われた。

農業ビッグデータの構築に関して最も関係の深い研究分野はフィールドフェノタイピングである。施設内におけるフェノタイピングでは欧米先進農業諸国が施設・予算・人員面で圧倒的に先行しているが，野外におけるフィールドフェノタイピングでは，本プロジェクトが世界のトップを走っており，フェノタイピング研究で最も重要な国際会議であるIPPN2016に基調講演者として招聘されている。また，本プロジェクトの採択により東大に国際フィールドフェノタイピング研究拠点が設立された。

海外では，昨年から中国がフェノタイピング研究に予算を付け始めている。南京農業大学では10億元（160億円）の予算でフェノタイピング研究がスタートした。

本プロジェクトは，INRA（フランス），CSIRO（オーストラリア），アイオワ州立大学等の世界トップクラスの研究グループと連携しながら国際的研究活動を先導している。機械学習のための教師データ（画像に，識別したい植物や病徴をラベリングしたデータ）の共有が国際連携の具体的な足掛かりとなることが明らかになり，次年度以降，国際的な農業ビッグデータの構築に関する研究を進めることが予定されている。

本プロジェクトによって，「時系列データを主体としたビッグデータから新知見を発見できる」というコンセプトが実証された。具体的には，ドローンで収集した圃場の３D時系列データ（4Dデータ）から，収量（テンサイの地下部重量）を予測する新しい指標（４D Score）が見出された（T. Duan, B. Zheng, W. Guo, S. Ninomiya, Y. Guo, S. C. Chapman. (2016) Comparison of ground cover estimates from experiment plots in cotton, sorghum and sugarcane based on images and ortho-mosaics captured by UAV, Functional Plant Biology, 2016.）。従来はNDVI（植生指数）という近赤外線と可視光の比からなる指標しかなく，アプリの多くはNDVIで病気診断から収量までを予測しようとしてきたが無理があった。４D Scoreはドローンとセットで実用技術として広く普及する可能性がある。

ドローンによるデータ収集は簡単に見えるが，大規模圃場で時系列データのための高頻度撮影を行うのは非常に難しく，様々な問題を解決する必要がある。これに向けて，①複数のドローンを協調制御するための計測制御技術，②ドローンによる飛行ルート，撮影場

所，飛行高度，飛行速度などを最適化するミッション・プランナー，③異なる日時の圃場データを比較できるビュアー等を開発した。これらにより，農家オペレータが試験データを収集し，育種等の研究者が膨大なデータから新知見を発見しやすくなった。

多様なツール（市販及び自作の 3D 再構成ツール）と多様なパイプライン（撮影高度，埼栄頻度，撮影目的，カメラ／ドローン／定点撮影カメラ）に対応し，多様な目的に利用するためのフレームワークを開発し，特許を取得している。

本プロジェクトでは，農業ビッグデータを構築するための手法の研究開発を行っているが，実際に大規模な農業ビッグデータを構築し現場で利用されるようにするには民間企業のサービスとして社会実装される必要があると考え，大学発ベンチャー支援プログラム（JST SCORE）に応募し採択されている（2017 年 10 月）。これによって，具体的なニーズ（ユーザーペイン）の把握，アプリケーション（ユースケース）が明確となり，国内外の企業と急速に連携が進んでいる。

また，北海道更別村の農家圃場で実験を開始したところ，本研究プロジェクトのコンセプトが地元北海道にインパクトを与え，これを契機に，更別村はスマート農業特区の申請を行った。特区申請の内容は，１）オペレータなしでのドローンの完全無人飛行，２）夜間の飛行，３）ロボットの無人公道走行などであり，これらの規制が緩和されると本プロジェクトの研究が一層加速されると期待する。

（今後の展開・波及効果と総括の所見）

本プロジェクトに対して総括からは，今年の１月の領域国際シンポジウムの際に，多様な要素技術に関する研究成果を，種々の立場の異なる組織や個々の農業従事者に対する目的と内容を異にする多様な情報サービスにどのようにつなげるのかに関して，明確にすることを指示している。

農業関係の組織や自治体とも密に連携して，研究成果を社会実装しており，現場からの期待も大きく，今後もビッグデータを用いた農業改革において国内を代表する研究拠点の役割を果たすものと期待する。

⑨松本プロジェクト

（研究課題）

構造理解に基づく大規模文献情報からの知識発見

（研究目標）

多くの分野で科学技術論文の発行数が急速に増加しているため，個々の研究者が関連するすべての論文に目を通すことが不可能になっている。同時に，個々の研究者が査読を行う論文数も年々増加している。このような状況に対応するためには，言語解析技術や文書解析技術を深化させ，大規模な専門分野文献の記述内容の解析技術を実現し，類似論文の検索や論文内容の把握を支援する統合的な支援環境の構築が重要である。生命科学，物質科学，脳神経科学，法律などの専門家との協働により，内容理解に基づいた文献の検索や

内容の要約,新たな知識の発見や研究動向の把握を支援する総合的な技術環境を構築する。本プロジェクトは，特定の分野だけを対象とせず，広い分野に適用可能な統合環境の構築を通して，科学技術の進歩に貢献することを目指す。

本プロジェクトのもう一つの目標は，専門分野の知識ベースの構築支援である。現在，生物医学系や材料系の分野で大規模な知識ベースの構築が進められているが，そのほとんどは，知識ベースに登録すべき内容を専門家が最新の論文を読むことによって抽出している。文書解析および言語解析等の技術を利用することにより，知識獲得の半自動化を実現することにより，知識ベース構築の飛躍的な効率化，精密化を目指している。

（研究成果のポイント）

・論文テキスト中の各文の意味を解析する際に重要な役割を果たす「句」の意味表現を計算する新しい手法を提案した。提案手法では，句の意味を計算する際に，その構成要素の単語の意味からボトムアップに計算するモデルと，句をそのまま利用して非構成的にその意味を計算するモデルを適応的に組み合わせて意味表現の最適化を行う。イディオム表現の認識や動詞句の類似度判定タスクにおいて世界最高精度を達成した（Kazuma Hashimoto and Yoshimasa Tsuruoka, ″Adaptive Joint Learning of Compositional and Non-Compositional Phrase Embeddings″, Proceedings of ACL, 2016)。

・文レベルの類似度を単語の意味からどのように計算するかに関して，カーネル関数を利用し，高次元空間で文の類似性学習を通して単語の表現を学習する方法を提案し，特別な素性を設計することなしに従来手法と同等あるいはそれらを上回る文の類似度判定法を達成した（Masashi Tsubaki, Kevin Duh, Masashi Shimbo, Yuji Matsumoto, ″Non-Linear Similarity Learning for Compositionality,″ Proceeding of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 2016.)。

・参照要約が少量の場合においても文書要約に有効な重要文抽出手法を提案した。具体的にはEncoder-Decoderモデルに基づく重要文抽出において，要約対象の文書分類とのマルチタスク学習とそれを効率的に進めるためのカリキュラム学習を導入することで，要約の観点に基づいた文抽出を可能にした。その結果，重要文抽出による要約について，従来手法を上回る精度を達成することができた（Masaru Isonuma, Toru Fujino, Junichiro Mori, Yutaka Matsuo and Ichiro Sakata, “Extractive Summarization Using Multi-Task Learning with Document Classification,” Proceeding of EMNLP, 2017)。

（研究成果と総括助言）

松本グループが担当している論文検索については，PubMed や Google Scholar などの検索サービスが存在するが，著者名やタイトル，あるいは，概要や本文からのキーワード検索にとどまっている。本プロジェクトに最も近いプロジェクトとして，Allen Institute for Artificial Intelligence の Semantic Scholar project があり，現在，計算機科学分野の論文を対象に類似論文や重要な引用論文の検索，PDFファイルからの図表の抽出など，論文解析の基盤技術の開発を行っている。統合的なシステム構築では遅れを取っているが，

PDF から XML への変換，図表の抽出，論文中のグラフや数式の自動読み取り等，個々の技術では，本プロジェクトで開発したツールが彼らのツールの精度を上回る結果を挙げている（グラフ読み取り，数式読み取りについては，人工知能分野のトップ会議 IJCAI-2018 に投稿予定)。論文の類似度について，「目的」「手法」などの観点を考慮した類似性／非類似性の定義（自然言語処理分野のトップ会議 NAACL に投稿中)，論文共通の構造に基づいた類似度学習の研究開発を進めており，独自視点での検索法を提案している。

乾グループでは論文中の深い議論構造の解析を行っており，科学技術論文中の議論をここまで深く理解しようとする研究は世界的にも類を見ない。最近大きくなりつつある Argument Mining 分野でも，ほとんどは関係解析などの浅い解析に留まっている。これまで，何を抽出したら深い理解をしたことになるか，及び計算機で扱える形にするにはどうすればよいかといった問題設計，実際のコーパスの構築等を行ってきている（アノテーション済みコーパスの公開，自然言語処理のトップ国際会議 NAACL に投稿中)。世界的にも新しい研究分野を開くとともに，論文解析への実用の暁には，陽には書かれていない研究の前提などを考慮した高度な検索等が可能になり，論文検索技術の高度化に大きく寄与することが期待できる。

論文からの知識ベース構築支援については，植物関係，酵素関係，熱電材料関係の研究者と協働を実施しており，既存の知識ベースからの訓練データの自動構築とそれを利用した概念および概念間の関係抽出の枠組みの実装を行っており，評価データのアノテーションおよび自動獲得した概念や概念関係の表示を目的として，PDF ファイルに直接アノテーションや関係表示を行うことのできるツール PDFanno が開発されている（自然言語資源とツールに関する最大の国際会議 LREC に採択済。また，国内学会で言語資源賞を受賞)。

論文の PDF からのテキスト抽出，図表抽出といった基礎的な解析から，関係抽出，深い議論の理解まで幅広く研究を実施しており，実用的なシステムを構築する基盤が整いつつある。構築中のシステムは特定の分野に特化しないものだが，最初の実用システムとして，自然言語処理分野の重要な論文のほとんどを網羅している ACL Anthology の論文を対象に，いくつかの視点を考慮した類似論文検索，検索結果の 2 次元配置，引用情報に基づくトレンド解析など各グループが開発している種々の機能を統合する準備を整えている。当該分野の研究者の研究効率に資するとともに，大規模論文データを扱うためのテストベッドとして用い，他分野の論文に対象を拡げていくことを計画している。

論文の内容解析については，文書・言語解析の基礎技術と意味関係抽出に基づく知識獲得インターフェイスを実現し，分野知識の全体像や関連分野との差異の表示など高度な機能を取り入れていくことにより，科学技術の進展や革新に貢献することを目指している。

論文テキスト中の各文の意味の解析，文レベルの意味的類似性，論文要約に関しては基盤となる以下のような技術を研究成果として確立している。「句」の意味表現を計算する新しい手法を提案し，イディオム表現の認識や動詞句の類似度判定タスクにおいて世界最高精度を達成した（Kazuma Hashimoto and Yoshimasa Tsuruoka, ″Adaptive Joint Learning

of Compositional and Non-Compositional Phrase Embeddings", Proceedings of ACL, 2016, pp. 205-215.)。文レベルの類似度の計算に，カーネル関数を利用し，高次元空間で文の類似性学習を通して単語の表現を学習する方法を提案し，特別な素性を設計することなしに従来手法と同等あるいはそれらを上回る文の類似度判定法を達成した（Masashi Tsubaki, Kevin Duh, Masashi Shimbo, Yuji Matsumoto, "Non-Linear Similarity Learning for Compositionality," Proceeding of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 2016.)。参照要約が少量の場合においても文書要約に有効な重要文抽出手法を提案した（Masaru Isonuma, Toru Fujino, Junichiro Mori, Yutaka Matsuo and Ichiro Sakata, "Extractive Summarization Using Multi-Task Learning with Document Classification," Proceeding of EMNLP, 2017, pp.2101-2110)。

（今後の展開・波及効果と総括の所見）

研究代表者は，これまでに多くのプロジェクトを遂行し，いずれにおいても，基盤となる要素技術をおろそかにせず，研究の土台をしっかりと固めたうえで，壮麗な建造物を構築することで定評のある研究者であり，今回も，土台となる要素技術の開発に多くの労力と時間をかけている。一方，乾グループは，深い推論の研究に焦点を当てており，現状では，この間に大きな空隙が生じている。今後，この空隙を埋める研究が加速するものと期待しているが，この空隙に位置する中核技術の基盤となるアイデアやモデルがまだ明確になっていないことに憂慮している。このことに関しては，既に研究代表者に総括から助言しており，中間層のフレームワークの論理モデルを早急に明確にしていただくことになっている。

大規模な文献データベースからの知識発見は，あらゆる科学技術分野でニーズが高まっており，松本プロジェクトに対する期待は大きい。

（２） 研究領域全体に関わる研究の経過と所見

① 研究総括のねらいに対する研究の状況

個々のプロジェクトは，以上に述べたように順調に研究成果を上げている。研究計画から大きく遅れるプロジェクトはなく，むしろ，総括からの助言を受け入れて，殆どのプロジェクトが当初計画以上の成果を上げている。

総括のねらいは以下の通りであった。

a. 分野や組織を越えた大規模データの統合的分析処理で価値創成
b. 次世代アプリケーション技術やシステム技術を実証的に創出
c. 各種要素技術を組み合わせての分析シナリオ作成
d. 国として注力すべき応用分野の掘り起しと国際連携
e. 再利用可能なノウハウの知識化，データ・サイエンティストの育成
f. 個人情報保護に関連するシステム機構の提案も期待

a「分野や組織を越えた大規模データの統合的分析処理で価値創成」

各プロジェクト内において,異なるサイロに入っていた大規模データを相互に関連付け,統合的分析処理を行うことで価値創成が行われている。これは，ビッグデータ応用を対象とする本研究領域の特徴で，アプリケーション分野の課題解決のミッションを持たずにビッグデータ基盤技術の研究を行う場合には，単一データセットを対象にした分析処理の研究に留まってしまう傾向が強い。ビッグデータの活用をさらに進めるには，本研究領域のような哲学に基づく研究を我が国全体で一層強化する必要があると考える。

b「次世代アプリケーション技術やシステム技術を実証的に創出」

個々のプロジェクト内では，計算機科学や統計学の研究者がアプリケーション分野の研究者と緊密に連携して研究が遂行されている。CREST ビッグデータ基盤研究領域では，主としてジェネリックな機械学習や統計的学習のアルゴリズムやシステム技術が研究されているが,本研究領域では,このような研究に留まることなく,アプリケーション分野の個々の課題解決のために，これらのジェネリックな手法をどう組み合わせたり，どのように拡張して適合すればよいかといった観点から研究が行われている。必要な場合は，その課題に特化した独自の手法の開発も行っている。「必要は発明の母」と言われるように，目の前の重要課題が，計算機科学や統計学，数学の若手研究者に大きな駆動力を与えている。このような中から生まれた技術が，一般化され汎用の技術として発展している。吉田プロジェクトにおける上田グループや，川島グループの研究はその例で，川島グループの研究は三好プロジェクトへ適用を目指して共同研究が始まっている。計算機科学や統計学の研究者とアプリケーション分野の研究者の緊密な連携の例としては、船津プロジェクトにおける泰地グループによる大規模化合物ライブラリのシミュレーションによる構築の研究，奥野グループにおける若手計算機科学研究者による DNN の採用への貢献，ビッグデータ基盤さきがけ研究領域の田部井プロジェクト「透過的データ圧縮による高速かつ省メモリなビッグデータ活用技術の創出」との共同研究もこの良い例である。三好プロジェクトにおける石川グループによる京コンピュータのアーキテクチャを活かしたデータ転送の高速化の研究や，越村プロジェクトにおける岡田グループの混合モデルを用いたスパースデータの機械学習の研究，大浪プロジェクトにおける小山田グループの可視化技術に関する研究などもこの良い例である。

c「各種要素技術を組み合わせての分析シナリオ作成」

プロジェクトが目標とする課題解決に向けて，分析シナリオを創出し，さらにはその処理を自動化したパイプラインとして構築することが進められている。吉田プロジェクトの超新星発見パイプラインや，船津プロジェクトにおける候補化合物発見パイプライン，三好プロジェクトのゲリラ豪雨の 30 分後までの予測の 30 秒毎更新のためのパイプラインなどがその例である。

d「国として注力すべき応用分野の掘り起しと国際連携」

5 章に述べたように，課題選考にあたって，A. 「ポートフォリオの設計」，B. 「フラグシップ・プロジェクトの選定」の 2 点に注意し，重点領域を公募に明記し，関連分野の著名な研究者を招いて招待講演セミナーを開催し，総括が関連学会やワークショップに出向いて応募の促進を行った。その結果，ほぼこの方針を満たす研究課題を揃えることができた。

フラグシップ・プロジェクトの選定に関しては，採択時から，それ自体が将来に独立の CREST 研究領域や，一層大型のプロジェクトの立ち上げへと発展するような研究課題を選ぶように心掛けたが，既に，いくつかのプロジェクトには，その可能性が高まるような顕著な成果が得られている。特に，三好プロジェクト，船津プロジェクト，越村プロジェクト，吉田プロジェクト，角田プロジェクト，大浪プロジェクト，平藤プロジェクトに関しては，計画以上の成果を上げており，三好プロジェクト，船津プロジェクト，越村プロジェクト，平藤プロジェクトに関しては，行政，産業界，メディアからの期待も大きい。吉田プロジェクトと大浪プロジェクトは，今後，関連学問分野への波及効果が非常に大きくなると期待している。角田プロジェクトは，臨床試験研究の進展によって，国内外，特に，臨床研究が遅れている国内において先駆的道を開く可能性に期待している。西浦プロジェクト，松本プロジェクトに関しても，社会や学界に対するインパクトは大きく，研究の一層の発展を期待している。

国際連携に関しては，米国 NSF のビッグデータ関連プログラムとの連携に関して，責任者の Chaitan Baru 教授，David Corman 教授と連携の相談を進めている。フランスに新しく設立される DataIA Institute とは，ディレクターの Nozha Boujemaa 教授，研究担当ボードメンバーの Michele Sebag 教授（2017 年度本研究領域国際シンポジウムのキーノート講演者）から連携の誘いを頂いている。また，国際アドバイザを通じて，個別のプロジェクトに国際連携のパートナー候補を紹介してきた。CREST ビッグデータ基盤研究領域と共に，毎年，NSF の関連プログラムとの合同シンポジウムも開催している。また，1 月の領域国際シンポジウムでは，各プロジェクト企画セッションに海外から招待講演者を招いてもらい，国際連携の促進を図っている。これらの試みはいずれも好ましい効果を上げている。

e「再利用可能なノウハウの知識化，データ・サイエンティストの育成」

体験型ポータル開発グループを船津プロジェクトの中に設け，全プロジェクトの体験型ポータルを開発し，再利用可能なノウハウの知識化に役立てようとしている。体験型ポータルを公開することにより，データ・サイエンティストの育成にも貢献できると考えている。また，若手合宿ワークショップを年 2 回開催しており，若手を中心に，再利用可能なノウハウの知識化とデータ・サイエンティストの育成を進めている。

f「個人情報保護に関連するシステム機構の提案も期待」

公募に先立ってこの分野の著名な研究者である Nikolaus Forgo 教授を招いて招待講演セミナーを行った。角田プロジェクトでは，インフォームド・コンセントの取得の徹底と倫理委員会の許諾取得を徹底していただいた。しかし，個人情報保護に関連するシステム

機構の提案は，本研究領域への研究申請の中には見いだせなかった。今後，若手合宿ワークショップで検討を進めたい。

以上に加えて，1 章（1）節②に述べたように，「共通基盤技術の構築」が，本研究領域の重要な目標である。異なるプロジェクトがターゲットとする多様なアプリケーション分野に共通する考え方や方法論，システム技術など，これまで，ビッグデータ基盤技術の研究においては充分に研究されてこなかったものの，ビッグデータ応用の共通分母となるような重要なテクノロジーを明確に抽出して，より広い応用範囲に適用可能なジェネリックな考え方や技術に育成し，共通応用基盤技術として確立することが目標である。

これには，

a. 若手研究者合宿ワークショップ
b. 研究領域国際シンポジウム
c. 全プロジェクトの体験型ポータルの開発と公開
d. 国内外のプロジェクトとの共同研究の促進

の 4 つの方策が，有効に機能している。若手合宿は年 2 回，2 日間にわたって開催し現在，約 30 名の若手研究者が登録されている。研究領域国際シンポジウムは，9 月に 1 日と，1 月に 2 日間の予定で開催している。研究代表者も全日程参加する方が急速に増え，プロジェクトに跨った相互理解が深まり，共通基盤技術が何であるかを議論するための理解も深まりつつある。国内外のプロジェクトとの共同研究の促進にかんしては，既に述べたように順調に進んでいる。

体験型ポータルに関しては，2018 年 1 月現在，船津プロジェクト，吉田プロジェクト，三好プロジェクト，越村プロジェクトのプロトタイプが開発されており，今年度中に西浦プロジェクトのポータルの開発も始まる予定である。今後も，毎年 2 プロジェクトの体験型ポータルの新規開発を行うと共に，既に開発しているポータルの内容に更新を行い，各プロジェクトの終了時に完成させる予定である。

共通応用基盤技術に関しては，以下の項目が，共通の応用基盤技術としての重要性が高いことが明らかになってきた。

A　シミュレーションと観測データのデータ同化
B　探索的可視化分析
C　分析シナリオのインタラクティブな定義・実行支援
D　サイバー・リサーチ・インフラストラクチャ
E　オントロジーに基づくリソース管理
F　オープン・サイエンス
G　文献からの知識発見

これらに関しては，今後，若手合宿ワークショップで集中議論のテーマとして取り上げて検討を進める。

7. 総合所見

（１） 研究領域のマネジメント（研究課題選考，研究領域運営）

米国 NSF のビッグデータ関連プログラムでは，以下の 2 点が応用の観点から重視されている。

A. 21 世紀の科学・工学のためのデータ活用 (Harnesing data for 21st Century science and engineering)
B. スマートで繋がったコミュニティ (Smart and connected community)

A は科学・工学におけるビッグデータ応用により，データ駆動型の科学・工学へとパラダイム・シフトを図ることを目指しており，B は，都市やネットワーク上の大規模なコミュニティにおける安心・安全で快適な社会基盤サービスの効率化，最適化を図ることを目指しており，NSF が他の研究助成機関と連携して進めてきた CPS（Cyber Physical System）の考えをコミュニティや社会にまで拡大したものである。

本研究領域が先行した研究課題では，このいずれかに対応しており，以下のように分類できる。

A. 船津プロジェクト（創薬／製薬），三好プロジェクト（気象学），越村プロジェクト（津波工学），角田プロジェクト（生体医学），西浦プロジェクト（感染症学），吉田プロジェクト（統計計算宇宙学），大浪プロジェクト（生物発生学），平藤プロジェクト（農学），松本プロジェクト（文献知識工学）
B. 三好プロジェクト（ゲリラ豪雨災害予防），越村プロジェクト（津波災害予防），西浦プロジェクト（感染拡大防止，感染予防），平藤プロジェクト(農業支援サービス)

B に含まれる最初の 3 プロジェクトは，防災の観点から安全・安心社会の基盤技術として重要であると共に，先端科学研究の簡単からも重要であり，A にも含まれている。最後の平藤プロジェクトも今後，農業コミュニティにおける作業支援情報サービスへと発展すると，B の観点からも重要になる。

A に関しては，創薬，生体医学，感染症，生物発生学，農学，気象学，天文宇宙学などの，ビッグデータ応用の観点から重要と考えられている主要分野をほぼカバーしている。当初重点領域として挙げていた物質材料科学の分野からは課題選考に至らなかったが，前にも述べたように，その後に関連 CREST 研究領域が立ち上がっていることから，これらとの連携を今後計画している。B に関しては，モビリティやスマート・シティ，物流，インフラ維持管理などが本研究領域ではカバーできていないが，これらに関しては SIP を始め，

別の施策で促進が行われている。ネットワーク上のコミュニティのスマート化も本研究領域ではカバーしていないが，CREST ビッグデータ基盤研究領域では，関連プロジェクトが含まれている。

本研究領域は，もともと 1 つのビッグデータ研究領域を，基盤技術を担当するビッグデータ基盤研究領域と，応用後術を担当するビッグデータ応用研究領域に 2 分したもので，採択可能な課題数も通常の領域に対して半数程度少ないことから，すべてをカバーすることは難しいが，主要な重要分野はある程度カバーできたと考えている。

領域の運営に関しては，領域アドバイザ 8 人に加え，国際アドバイザリ・ボードを設置して 5 人の海外の著名な研究者に就任していただき，面接審査，進捗報告会を兼ねた年 2 回の国際シンポジウム，そのうちの 1 回の前日に開催される中間評価会議に出席を頂いている。これにより，各プロジェクトに対し，適切かつ合意的な評価と，根幹にかかわる重要な助言が行われている。

年 2 回の若手研究者合宿ワークショップはプロジェクト間の壁を越えた活発な議論の場を提供し，共通基盤技術の明確に役立つと共に，若手育成に効果を発揮している。

国際アドバイザが出席するイベントはすべて英語を公用語としたことにより，海外からのキーノート講演者やプロジェクト企画招待講演者との分野を超えた議論が深まり，新しい国際連携協力関係が育っている。

体験型ポータルの開発は，各プロジェクトにも好評で，プロジェクト内で拡張し活用する例も出始めている。

（２） 研究領域としての研究成果の見通し

個々の研究課題プロジェクトが，それぞれの分野でフラグシップ・プロジェクトにふさわしい成果を上げるであろうことには疑念がない。既に，国際的にもオンリー・ワンの技術を確立したと高く評価されているプロジェクトもある。特に，三好プロジェクトの京コンピュータを用いたアンサンブルシミュレーションとフェーズド・アレイ・レーダからの実時間の雨雲分布データとの実時間データ同化によるゲリラ豪雨の 30 分後までの予測を 30 秒ごとに 100m メッシュの精度で行う技術，船津グループの巨大化合物ライブラリの構築と，化合物とタンパク質のドッキング推定ディープラーニングで高速に行う技術，越村プロジェクトの津波シミュレーションと海底センサ出力から沿岸部の津波高をデータ同化とガウスプロセスを用いた非線形回帰で即時推定する枠組み，理論モデルを用いたシミュレーション画像データを学習データとして用い，ディープラーニングで実現した，吉田プロジェクトの超新星の自動検出パイプラインの構築，これらの成果は世界を大きく先導している。残りのプロジェクトも，世界を大きく先導する技術開発を目標としており，その完成は間近に迫っている。

（３） 本研究領域を設定したことの意義

ビッグデータの分野では，アルゴリズム研究などの基盤技術の研究者は，ともすれば自身の研究に都合よく適合しているデータ集合を対象に研究することが多い。これらのデータ集合は，研究コミュニティにおいてベンチマーク・テスト用データ集合として用意されているものを用いることが多い。そのため，ビッグデータ基盤技術とビッグデータ応用技術との間にはまだ非常に大きなギャップが存在する。前者の立場からは，これを単に，データのキューレーションの問題と見做す研究者も多く，このギャップはこのままでは当面埋まりそうにもない状況であった。

本研究領域で採択された研究課題は，いずれもそれぞれのアプリケーション分野のサイエンスを先導する研究者が，ビッグデータ基盤技術の分野において活躍をしている研究者と密に連携し，共同で，X―サイエンスをX―インフォマティクスへとパラダイム・シフトさせることを目指す計画になっている。X―サイエンスの中で，ビッグデータ基盤技術を適用する手伝いをするということでもなく，ビッグデータ基盤技術の研究者がX―サイエンスのデータをつまみ食いするということでもないという点が，極めて重要である。

しかも，異なる多様なXに対して，1 つの研究領域の中で一緒に情報共有し議論しながら，このパラダイム・シフトを進める点に，この研究領域の大きな意義がある。上述のように，その意義を反映するように，大きな成果が出ており，その相乗効果が，プロジェクト間の連携として具体的に出始めている。

（４） 今後への期待，展望

三好プロジェクトのゲリラ豪雨予測技術，船津プロジェクトの創薬のための巨大化合物ライブラリとドッキング推定技術，プロセスのソフトセンサ技術，越村プロジェクトの津波シミュレーションと津波高の事前予測技術，吉田プロジェクトの超新星検出パイプライン技術，西浦プロジェクトの感染症流行予測技術，角田プロジェクトのオミックス解析技術，大浪プロジェクトの顕微鏡画像からの細胞分裂のフェノタイプデータに自動抽出，平藤プロジェクトの自動編隊飛行ドローンを用いた作物の育成状況データの自動抽出技術，松本プロジェクトの PDF 形式の文献からの表データの自動抽出技術，これらの技術は即座に利用できる待ち望まれていた技術である。これらは，各プロジェクトの成果の一部であり，研究はさらに進行中である。今後も，このような重要な技術成果が確立され，社会実装が進み，利用者が増大し，社会や学界に大きなインパクトを与えると期待する。

（５） 所感，その他

本研究領域は，ターゲットとするアプリケーション分野の多様性故に，領域運営には困難が予想された。総括はすべての研究課題の詳細内容と国内外の研究動向を理解する必要があり，研究領域全体として，共通応用基盤技術を明らかにしてこれを確立するには，研究代表者と参加する研究者が，他の研究課題の研究内容にも興味を持ち，相互理解を進めることが必須であった。前者に関しては，アドバイザを適切に依頼し，国際アドバイザリ・

ボードを設置することで，充分に対応ができたと考えている。後者に関しては，個別のサイトビジットに加え，研究領域国際シンポジウムを年 2 回，合計 3 日間行い，国際アドバイザも含めて議論するようにしたことと，1 月のシンポジウムではプロジェクト企画セッションを９つ設け，それぞれ海外から招待講演者を招いていただくことで，プロジェクトに跨る議論が活発になった。加えて，NSF との合同国際会議も年 1 回開催している。年 2 回，各 2 日間の若手研究者合宿ワークショップは，議論と情報交換により時間を取り，プロジェクトに跨る情報共有と活発な研究討論の場となっている。これらによって，当初の懸念事項を解消しつつある。

このような研究体制は，国内外を通じて他に例がなく，NSF の担当者や，国際アドバイザ，キーノート講演者，招待講演者の多くが驚き，議論に参加できたことに感謝をしている。

今後，本領域が掲げるような哲学に則った研究体制が，更に増えることを強く望む。

以上