

戦略的創造研究推進事業  
－CREST(チーム型研究)－

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」

研究領域事後評価用資料

研究総括：田中 讓

2021年2月

## 目次

1. 研究領域の概要	1
(1) 戦略目標	1
(2) 研究領域	6
(3) 研究総括	7
(4) 採択研究課題・研究費	8
2. 研究領域および研究総括の設定について	9
(1) 研究領域設定の理由	10
(2) 研究総括指定の理由	11
3. 研究総括のねらい	12
4. 研究課題の選考について	14
5. 領域アドバイザーについて	17
(1) 領域アドバイザー	18
(2) 国際・領域運営アドバイザー	19
(3) 領域アドバイザー選定方針と国際アドバイザリ・ボードの設置	19
6. 研究領域のマネジメントについて	20
(1) 研究課題の進捗状況の把握と評価、それに基づく研究課題の指導	20
(2) チーム型のネットワーク型研究所として、研究課題間や他の研究領域、国内外の他の研究機関、異分野との連携・協力の推進	22
(3) 研究費配分上の工夫	32
(4) 研究領域中間評価結果への対応	32
(5) その他マネジメントに関する特記事項	38
7. 研究領域としての戦略目標の達成状況について	39
(1) 研究総括のねらいに対する研究成果の達成度	39
(2) 研究領域全体として見た場合の特筆すべき研究成果	50
(3) 研究成果の科学的・技術的な観点からの貢献	56
(4) 研究成果の社会的・経済的な観点からの貢献	59
(5) その他特記事項	62
8. 総合所見	68
(1) 研究領域としての戦略目標の達成状況	68
(2) 研究領域のマネジメント	68
(3) 本研究領域を設定したことの意義	70
(4) 科学技術イノベーション創出に向けた、今後への期待、展望、課題	70
(5) 所感、その他	71

## 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、オーダーメイド医療や早期診断、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（2011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2009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、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。

##### (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

る調査研究結果(平成 21 年度)-情報インデックスの計量-」 2011 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 2012 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

2012 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 2016 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、2012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、2012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）（2012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）（2012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」（2013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	所属・役職 研究終了時(採択時)	研究課題	研究費*
2013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	412
	三好 建正	理化学研究所 計算科学研究センター チームリーダー (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	316
2014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	360
	角田 達彦	東京大学 大学院理科学研究科教授 (理化学研究所 統合生命医科学研究センター グループディレクター)	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	359
	西浦 博	京都大学大学院 医学研究科 教授 (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	301
	吉田 直紀	東京大学 大学	広域撮像探査観測のビッグデータ	335

		院理学系研究科 ／カブリ数物連 携宇宙研究機構 教授	分析による統計計算宇宙物理学	
2015 年度	大浪 修一	理化学研究所 生命機能科学研究 センター チ ームリーダー (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	349
	平藤 雅之	東京大学 大学 院農学生命科学 研究科 特任教 授 (農業・食品産業 技術総合研究機 構北海道農業研 究センター 研 究領域長)	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	400
	松本 裕治	理化学研究所 革新知能統合研 究センター チ ームリーダー (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	365
			総研究費	3,197

\*各研究課題とも研究期間の総額，進行中の課題は予定を含む(2020年10月31日現在)

## 2. 研究領域および研究総括の設定について

本戦略目標の下，相互に関連する以下の二つの研究領域が設定された。

研究領域 1 「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」(CREST)

研究領域 2 「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」  
(CREST・さきがけ複合領域)

研究領域 1 が本研究領域である。

## (1) 研究領域設定の理由

本戦略目標は、分野を超えたビッグデータの利活用により、新しい科学的発見や社会的課題解決に資する新たな知識や洞察を得るための革新的な情報技術や数理的手法の創出・高度化・体系化を目指すものである。このためには、生命、健康・医療から地球環境、都市基盤システム、防災・減災までの多様な分野における科学的発見・社会的課題解決を目指す横断型アプローチと、ビッグデータを利活用する上で本質的となる技術課題を情報学・数理科学的見地から深く掘り下げブレークスルーをねらう深堀型アプローチによる重奏的な研究が行われる必要がある。そのため、本戦略目標の下には、2つの研究領域を選定することが適切である。

とくに、研究領域 1 における次世代アプリケーション技術の創出・高度化に当たっては、研究領域 2 の研究で得られる IT 基盤技術を取り込みながら推進し、その逆に研究領域 2 の IT 基盤技術の創出・高度化・体系化に当たっては、研究領域 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、研究領域 1 と研究領域 2 が相互連携することで、効果的に戦略目標が達成されるものと考えられる。

研究領域 1 においては、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のために、個々の研究者や研究室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。さらに、この目的の達成に必要な IT 基盤技術を実証的に創出・高度化・体系化することを目指す。

そのためには、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、宇宙地球環境など様々な IT 以外の研究者と IT 系研究者との連携が不可欠であるが、現状ではデータの活用についてはまだ端緒の段階にあり、このような連携はあまり多く見られない。本研究領域の設定により、このような連携による研究を促すことで、優れた研究提案が数多く見込まれるとともに幅広く数多くの研究分野の発展に寄与することが期待される。

研究推進体制としては、上記のように IT 系以外と IT 系の研究者がチームを組むことができる CREST を選定することが適切である。

以上のことから、本研究領域は戦略目標の達成に向けて適切に設定されている。

研究領域 2 においては、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・最適化などに資する、情報学を基盤とした IT 基盤技術の創出・高度化・体系化を目指す。このため、大規模データを圧縮・転送・保管する大規模管理システムの安定的運用技術や、多種多様な情報を横断して検索・比較・可視化して真に必要な知識を効率的に取り出す技術、これらを可能にする数理的手法やアルゴリズムなどの開発を対象としている。これらの研究の推進にあたり、ビッグデータから社会における価値創造に至る全体のメカニズムデザインやエコシステムを視野に入れ、個別要素技術を組み合わせ総合的に高度化・体系化する取り組みが必要とされている。そのため、関係する研究者がチームを組んで研究推進する CREST を選定することが適切である。

また、独創的なアプローチを有するポテンシャルの高い個人研究者について当該研究領域への参画を促すことで、ビッグデータ時代に必要となる IT 基盤技術を研究開発する研究者コミュニティが醸成されると見込まれる。

さらに、ビッグデータのドメインに依存しない本質的課題を抽出・解決する上で、個人研究者が研究チーム等の研究目標に捉われず、自由に IT 以外の分野との横断的交流を進めることは、将来的な IT 以外の分野との積極的な連携・融合に資すると期待される。したがって、さきがけを選定することは適切である。

本領域の設定により、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・体系化への取り組みが行われるとともに各アプリケーション分野における次世代アプリケーション基盤技術の創出・高度化へ資することから、本研究領域は戦略目標の達成に向けて適切に設定されている。また、ビッグデータから新たな知識等を得ようとする社会的ニーズが高まる中、そのための技術・手法の創出・高度化・体系化を促すものとして適切な研究領域であり、データマイニング・機械学習・統計学等の分野から優れた研究提案が多数見込まれる。

## (2) 研究総括指定の理由

田中譲氏は、知識の構築・編集・再利用のための知識メディア研究および抽出した様々な価値を人が評価・意思決定できるよう複数のプログラムモジュールを容易な操作で自在に組み替えて連携させる技術で、国際的・学際的な情報学の発展に寄与するなど優れた研究実績を有する研究者である。近年ではこれらの研究を元に EU（欧州連合）の推進する FP7（Framework Programme 7）の中の「p-medicine」に 2011 年から参画しており、臨床治験や遺伝子情報、患者のカルテ履歴といった多様で大量の情報から、患者ごとにオーダーメイド医療を提供することを最終目標としている。このプロジェクトには、EU 圏外の研究機関として田中譲氏のラボラトリーだけが唯一参加している。また文部科学省の 2011 年

度の委託研究にて「知識創成 HCPS スマート・フェデレーション統合環境と降雪地域における除排雪の効率化・最適化への適用」という研究テーマを実施。札幌市の除排雪システムの最適化やドライバーの適切な誘導などを目指しながら、データ分析可視化基盤を開発している。

これらのことから、ビッグデータ時代における科学的発見・社会的課題解決に向けた次世代アプリケーション技術を創出することを目的とする本研究領域の運営と推進に関し、分野を越えた俯瞰的な視野と豊富な人脈および優れた先見性と洞察力を有すると判断できる。

同氏はまた情報処理学会英文論文誌 JIP 編集委員長、21 世紀 COE リーダーなどを歴任、EU の推進する FP6 の研究開発プロジェクト ACGT(Advancing Clinico-Genomic Trials on Cancer)に正規メンバーとして参画するなど、国際的にも関連分野の多くの研究者から信頼され、適切な研究マネジメントを行う経験、能力を有するとともに、公平な評価を行いうると認められる。

以上より、同氏は本研究領域の研究総括として適任であると判断される。

### 3. 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、経験知すらも十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十分で、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た分析シナリオ作成

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。

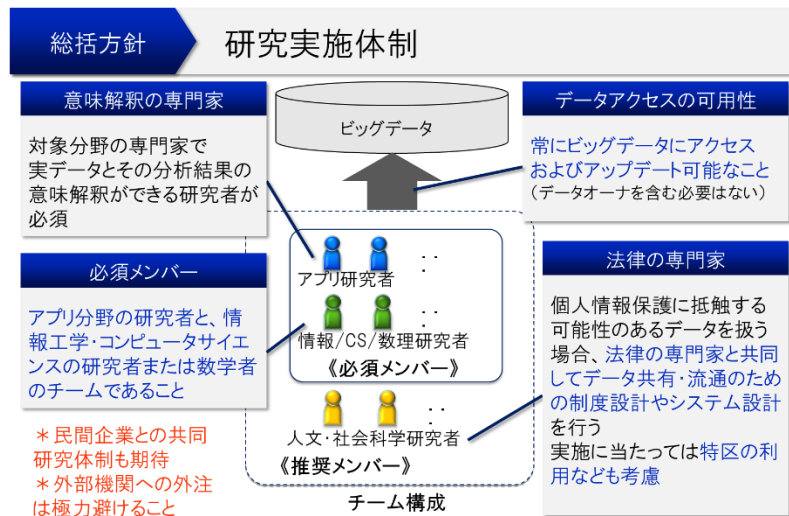
f. 個人情報保護に関連するシステム機構の提案も期待

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、対象分野の専門家を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。



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。

#### 4. 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ーに就任して頂いたザール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ー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おうと共に、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ーと 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ー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、十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に、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、十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。



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詳述する。

## 5. 領域アドバイザーについて

(1) 領域アドバイザー

アドバイザー名 (専門分野)	終了時の所属	役職	任期
天野 肇 (ITS)	特定非営利活動法人 ITS Japan	専務理事	2014年4月～2021年3月
岩野 和生 (クラウドコンピューティング, アルゴリズム, 組合せ最適化)	三菱ケミカルホールディングス	執行役員	2013年4月～2016年3月
柴崎 亮介 (GIS, 空間情報処理)	東京大学 空間情報科学研究センター	教授	2013年4月～2021年3月
下田 正文 (バイオメディカル情報)	未病社会の診断技術研究会	理事	2014年4月～2021年3月
鈴木 良介 (ビッグデータ・IoT・人工知能と社会)	(株) こゆるぎ総合研究所	コンサルタント・代表取締役	2013年4月～2021年3月
武田 浩一 (自然言語処理, 知識発見)	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	2017年4月～2021年3月
西浦 廉政 (非線形偏微分方程式, 反応拡散系, 数理モデリング)	北海道大学 電子科学研究所 名誉教授	特任教授	2013年4月～2021年3月
松井 知子 (知覚情報処理, 統計科学)	統計数理研究所 データ科学研究系	研究主幹・教授	2013年4月～2021年3月
宮野 悟 (システム生物学, バイオインフォマティクス)	東京医科歯科大学 M&D データ科学センター	センター長	2013年4月～2021年3月

## (2) 国際・領域運営アドバイザー

アドバイザー名	終了時の所属	役職	任期
Costantino Thanos (Research Cyber Infrastructure)	Italian National Research Council Research	Director	2013 年 4 月～2021 年 3 月
Norbert Graf (Personalized Medicine on Cancer)	Saarland University Hospital	Professor, Doctor, Director	2014 年 4 月～2021 年 3 月
Nicolas Spyrtatos (Database and Data Science)	University of Paris Sud 11	Professor Emeritus	2014 年 4 月～2021 年 3 月
Nigel Waters (Geographical Information System)	University of Calgary	Professor Emeritus	2013 年 4 月～2021 年 3 月
Randolph Goebel (AI and Machine Learning)	University of Alberta	Professor	2013 年 4 月～2021 年 3 月

## (3) 領域アドバイザー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ー

をお願いした。2013 年度は地理情報科学の専門家であった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、(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ー 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ーも、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 Spyrtos 教授の 2 名を増員した。Spyrtos 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ー 1 名 (岩野アドバイザー) 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ー及び 5 名の国際・領域運営アドバイザーで運営している。就任時、米国のジョージメイソン大学教授であった Waters 教授は、カナダのカルガリー大学に戻られた。

依頼した国際アドバイザーは、いずれも自身の専門分野に限らず、コンピュータ・サイエンス、特にデータベースやビッグデータ、機械学習の理論と応用に関して幅広く精通しており、このことが、採択課題の選考審査、中間および事後評価、領域主催国際シンポジウムにおける議論や助言を活性化し、国際的な視点に立った判定や評価に大きく貢献した。

国際アドバイザー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。

## 6. 研究領域のマネジメントについて

### (1) 研究課題の進捗状況の把握と評価、それに基づく研究課題の指導

2014 年度は領域共催のビッグデータ基盤・応用合同領域会議と、領域主催の CREST「ビッ

「ビッグデータ応用」公開シンポジウムの際に、2015年度は、領域共催のビッグデータ基盤・応用合同領域会議とCREST・さきがけ「ビッグデータ基盤・ビッグデータ応用」効果成果報告会、さらには領域主催のCREST International Symposium on Big Data Applicationの際に、2016年度からは、課題中間評価とその翌日の夏のCREST「ビッグデータ応用」公開シンポジウムと、冬のCREST International Symposium on Big Data Applicationの2回のシンポジウムと、2回ないし3回の領域合同共催の会議やシンポジウムの際に、進捗状況の報告を受け、評価とアドバイスを行った。

領域主催の国際シンポジウム(2014~2015年度は年1回、2016年度以降は年2回)には、国内、国外のアドバイザーが原則全員出席し、全プロジェクトが進捗状況を報告し、総括ならびにアドバイザーからの活発な助言を受けた。これに加え、各研究課題の中間評価の際には、対象課題の中間報告がなされ、ほぼ全員の国内アドバイザーとすべての国際アドバイザーから、口頭でのアドバイスと共に、後日、詳細な評価と助言が各プロジェクトに送られた。

総括からも、採択時の面接審査の際、毎年1~2回の国際シンポジウムの際、課題中間評価の際、1年延長課題の延長前の事後評価の際に、研究課題ごとに、新規課題の追加や計画の一部の縮小に関し、積極的に介入し助言を行った。その詳細は、「7. 研究領域としての戦略目標の達成状況について」の「(1) 研究総括のねらいに対する研究成果の達成度」の各研究課題毎の報告の中の、「(vi) その他特記事項」(総括からの助言とそれへの対応)の欄に、研究課題ごとに詳しく報告している。具体的に概要を示すと、成果の統合化の強化の助言を、船津プロジェクト、越村プロジェクト、角田プロジェクト、平藤プロジェクト、松本プロジェクトに、新しい研究課題の追加を、三好プロジェクト、大浪プロジェクトに、当初計画の一部課題の進捗の遅れに対する加速を西浦プロジェクト、吉田プロジェクト、松本プロジェクトに、総括のねらいに沿った軌道修正を、角田プロジェクトと松本プロジェクトに助言し、いずれも対応がなされ、結果的として大きな成果につながった。

このような助言は、冬の領域主催国際シンポジウムの際に行うことも多く、各プロジェクトが企画セッションに招待したその専門分野の海外の著名研究者も助言に賛同の意見が述べられることも多々あり、研究代表者も納得の上で迅速に対応した。海外からの国際アドバイザー、キーノートスピーカー、プロジェクトの招待講演者がすべて参加する国際シンポジウムの場でのオープンな議論を通じて、助言や軌道修正の依頼を積極的に行ったことが、透明性、公平性の観点からも、関係者の理解を得るのに役立った。

これらの機会の他に、各プロジェクトに対し、研究期間中に1回以上のサイトビジットを行った。各サイトビジットには、全日ないし午後一杯を使って報告を受け議論を行い、助言や軌道修正の依頼を行った。毎回、数名の国内アドバイザーが同席した。

さらに、体験型ポータル開発チームが、各プロジェクトとの打ち合わせを行う際には、総括も現地に同行し議論に参加した。

船津プロジェクトに関しては、プロジェクトが国内で主催した国際会議にも総括が出席し議論に参加した。その際のドイツからのキーノート・スピーカーのJuergen Bajorath教授

と総括との議論の中で、船津教授を、Herman Skolnik Award 候補者として、共同で American Chemical Society に推薦することが決まり、後に、船津教授の同賞受賞につながった。

## (2) チーム型のネットワーク型研究所として、研究課題間や他の研究領域、国内外の他の研究機関、異分野との連携・協力の推進

研究領域の運営に当たっては、ビッグデータ応用に関するチーム型のネットワーク型研究所の組織化を目指し、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。更に、本研究領域全体の目標として以下を設定した。

- 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十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。

これらの目標に向けた研究活動の活性化のために、研究課題間や他の研究領域、国内外の他の研究機関、異分野との連携・協力の推進と、研究者間、特に若手研究者間の連携・協力の推進を目指して、以下の方策を実施した。

- a. 若手研究者合宿ワークショップ（若手研究者を介した研究課題間や他の研究領域との連携・協力の促進）



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した。他の領域の若手研究者との連携協力の推進を目指し、ビッグデータ基盤技術 CREST 領域の若手研究者である Adnan Sljoka 氏や、後に触媒インフォマティクス CREST 領域の研究代表者となる高橋啓介氏のような同世代の研究者を招待講演者として招く試みや、深層学習やデータ同化のような共通基盤アルゴリズムのチュートリアルを設ける試みも実施した。

具体的には、2 日間の若手合宿ワークショップを、国内アドバイザーにも参加していただき、2016 年度は 11 月に 1 回、2017 年度から 2019 年度は 6 月～9 月に 1 回と 11 月～1 月に 1 回の計 2 回行い、2020 年度は 10 月に 1 回行った。2018 年度と 2019 年度の 1 月の会は、国際アドバイザー全員と 1 月のシンポジウムに招いたキーノートスピーカー全員にも参加していただき、公用語を英語とした。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十分な時間を割り当てている。各回の招待講演者の選定と会の運営は若手研究者が幹事グループを組織して主体的に担当した。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た。何人かの研究代表者も積極的に参加されるようになった。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、以降の合宿にも参加された。

参加者からは、「参加を重ねるごとに、共通応用基盤技術の内容が徐々に見えてきた」との感想が多く寄せられた。

この合宿は当初は日本語のみで行っていたが、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った。そこで、2018 年度と 2019 年度の 1 月の合宿は、領域のシンポジウムのキーノート講演者と国際アドバイザーも招いて、議論や懇親会にも参加していただいた。

#### 若手合宿の開催歴

回	年度	日程	場所	若手数 (参加者数)	招待講演(敬称略) 備考
1	2016	2016/11/23-24	湘南国際村	9 (18)	高橋啓介, 辻井 潤一
2	2017	2017/6/2-3	舞子ビラ神 戸	7(27)	中垣俊之

3		2017/11/24-25	湘南国際村	19(32)	伊藤純至, 鹿島久嗣
4	2018	2018/6/12-13	湘南国際村	9(22)	近藤圭一
5		2019/1/16-17	湘南国際村	14(27)	Adnan Sljoka 国際AD参加
6	2019	2019/9/26-27	湘南国際村	6(21)	三好建正 国際AD参加
7		2020/1/16-17	湘南国際村	8(15)	大塚成徳, 今泉允聡
8	2020	2020/10/6	オンライン 開催	6(12)	松本裕治 半日の交流会

b. 研究領域主催国際シンポジウム（研究課題間や他の研究領域，国内外の他の研究機関，異分野との連携・協力の推進）

領域内の研究課題間の相互理解と連携・協力の推進と，国内外の研究機関との連携・協力の推進を目指し，領域の国際シンポジウムを年2回開催し，国内アドバイザーのみならず，国際アドバイザーにも可能な限り参加していただき，毎回，英語による活発な議論を行った。

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，海外から1～2名の著名な研究者を総括が選び，キーノート講演者として招待した。

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名の著名な研究者をキーノート講演者として招き，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った。これにより，国際アドバイザー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た。彼らが議論の牽引役となって，応用分野に跨った深い議論が活発に行われるようになり，2014年度～2015年度のシンポジウムで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，2016年度以降のシンポジウムでは多くが最初から最後まで議論に参加するようになった。これにより，応用分野が異なるために，他のプロジェクトの報告には参考になるものがないと考える傾向があった研究者の多くが，共通の課題や解決手法を他のプロジェクトの成果の中に見出すようになり，研究課題間での連携や共同研究が促進された。例として，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た。

領域主催および共催の国際シンポジウム等開催歴  
各シンポジウムの下段はキーノート講演者

太字のイベント名は領域主催

シンポジウム名	日時	場所	入場者数	特記事項
<b>医学分野ビッグデータの利活用に関するワークショップ</b> (領域公募時)	2013.7.6 13:00- 17:00	CIVI 北梅田研 修センター 5 06号室	20	次年度の研究領域募集に向けたワークショップ
	田中譲, Norbert Graf, Nikolaus Forgo, 宮野悟			
<b>公開ミニシンポジウム「ビッグデータ利活用に関する国際動向」</b> (領域キックオフ時)	2014.3.4 13:00- 18:00	JST 東京本部 地下1階ホール	80	
	Randy Goebel, Costantino Thanos, Nigel Waters			
ビッグデータ基盤・応用 合同領域会議 (非公開)	2014.11.5 ~11.6	神戸ポートピア ホテル	200	ビッグデータ基盤・応用
<b>CREST「ビッグデータ応用」公開シンポジウム</b>	2015.3.3 9:00-18:00	ベルサール九 段	100	
	<b>Dennis Tsichritzis</b> (President of the Council, Technical University of Crete, Greece) Title: Giga Data and Tera Data: Seven steps to Heaven <b>Jack Park</b> (Silicon Valley researcher and innovator) Title: HyperMembrane Structures for Open Source Cognitive Computing			
ビッグデータ基盤・応用 合同領域会議 (非公開)	2015.10.6 ~10.7	T K P ガーデ ンシティ仙台	200	ビッグデータ基盤・応用
CREST・さきがけ「ビッグデータ基盤・ビッグデータ応用」効果成果報告会	2016.2.4 10:00- 12:00	一橋大学一橋 講堂 201-203 会議室	50	情報処理学会 SOFTWARE JAPAN 2016の中で開催
	(キーノート無し)			
<b>CREST International Symposium on Big Data Application</b>	2016.3.4~ 3.5	アキバプラザ (秋葉原)	120	
	<b>Haridimos Kondylakis</b> (Collaborating Researcher, Institute of Computer Science Foundation of Research & Technology-Hellas (FORTH), Greece), on behalf of <b>Manolis Tsiknakis</b>			
JST・NSF 国際連携公開シンポジウム	2016.5.12	東京大学伊藤 謝恩ホール	180	ビッグデータ基盤・応用
	Peter Arzberger (Division Director, Computer and Network			

	Systems (CNS) in the Directorate for Computer and Information Science and Engineering (CISE) at the United States National Science Foundation (NSF))			
CREST 「ビッグデータ応用」 公開シンポジウム	2016. 8. 5 9:00-18:00	アキバホール (秋葉原)	120	
	<p><b>Peter Arzberger</b> (Division Director, Computer and Network Systems (CNS) in the Directorate for Computer and Information Science and Engineering at NSF),</p> <p><b>Chaitan Baru</b> (Senior Advisor for Data Science in the Directorate for Computer and Information Sciences and Engineering at NSF)</p> <p>“Big Data and Smart and Connected Communities: Opportunities and Challenges” (This lecture will be presented on TV conference.)</p> <p><b>Nigel Waters</b> (Professor of Emeritus of Geography, University of Calgary)</p> <p>“Using GIS in the global fight against the Aedes aegypti mosquito: recent case studies and some ongoing volume, value and velocity challenges”</p> <p><b>Satoru Miyano</b> (Director/Professor, Human Genome Center of The Institute of Medical Science, The University of Tokyo)</p> <p>“Breaking Cancer Big Data by Supercomputer and Artificial Intelligence”</p>			
ビッグデータ基盤・応用 合同領域会議 (非公開)	2016. 11. 28 ~11. 29	ベルサール六 本木	180	ビッグデータ基盤・応用
JST・NSF 国際連携公開シンポジウム	2016. 11. 30 10:00- 18:00	ベルサール六 本木	200	ビッグデータ基盤・応用
CREST International Symposium on Big Data Application	2017. 1. 11 ~1. 12	アキバホール (秋葉原)	140	
CREST 「ビッグデータ応用」 公開シンポジウム	2017. 9. 6 9:00-18:00	アキバホール (秋葉原)	120	

	<p><b>Michele Sebag</b> (Principal senior researcher (DR1) at CNRS (France) / deputy director of the Lab. of Computer Science at Université Paris-Saclay)</p> <p>“Algorithm selection and collaborative filtering”</p>			
ビッグデータ基盤・応用 合同領域会議（非公開）	2017.12.18 ～12.19	コクヨホール	120	ビッグデータ基盤・応用
JST・NSF 国際連携公開シンポジウム	2017.12.20 10:00- 18:00	コクヨホール	200	ビッグデータ基盤・応用
				<p>(<b>David Corman</b> (Program Director, CISE/CNS, NSF))</p> <p>“Research Trends in NSF and JST-NSF Collaboration Opportunities”</p>
CREST International Symposium on Big Data Application	2018.1.16 ～1.17	秋葉原コンベンションホール	120	
				<p><b>Christos Faloutsos</b> (Professor, Computer Science Department/Machine Learning Department, School of Computer Science, Carnegie Mellon University)</p> <p>“Anomaly detection in large graphs”</p> <p><b>Manolis Tsiknakis</b> (Professor of Biomedical Informatics and eHealth, Dep. of Informatics Engineering, School of Engineering, TEI CRETE)</p> <p>“Big Data Analytics: Promise and Potential for Individualized and Precision Medicine”</p>
CREST「ビッグデータ応用」公開シンポジウム	2018.9.30	アキバホール (秋葉原)	120	ビッグデータ応用
				<p><b>Christos Papadimitriou</b> (Professor of Computer Science, Columbia University)</p> <p>“A Computer Scientist Thinks about the Brain”</p>
CREST International Symposium on Big Data Application	2019.1.14 ～1.15	アキバホール (秋葉原)	120	ビッグデータ応用
				<p><b>Stuart Kauffman</b> (Theoretical Biologist, Complex Systems Researcher, The Institute for Systems Biology, Seattle)</p> <p>“The Shape of History”</p> <p><b>Rich Caruana</b> (Principal Researcher, Microsoft Research)</p>

	"Friends Don't Let Friends Deploy Black-Box Models: The Importance of Intelligibility in Machine Learning"			
JST・NSF 国際連携公開シンポジウム	2019. 3. 11	Annex Hall, Kyoto International Conference Center	200	ビッグデータ基盤・応用, 知的情報, 他
	<b>Erwin Gianchandani</b> (Deputy Assistant Director, NSF/CISE) "Scientific Priorities and Future Directions in NSF's CISE Directorate" <b>Nozha Boujema</b> (Previous Head of DATAIA / Chief Science & Innovation Officer, Median Technologies)			
ビッグデータ基盤・応用 合同領域会議 (非公開)	2019. 3. 12 ～13	Annex Hall, Kyoto International Conference Center	120	ビッグデータ基盤・応用
CREST「ビッグデータ応用」公開シンポジウム	2019. 9. 25	アキバホール (秋葉原)	100	ビッグデータ応用
	<b>Marc Schoenauer</b> (Principal Senior Researcher, INRIA) "When Big Data and Machine Learning meet Partial Differential Equations" <b>Michele Sebag</b> (Deputy director of Laboratoire de Recherche en Informatique, CNRS) "Learning Causal Models"			
ビッグデータ基盤・応用 合同領域会議 (非公開)	2019. 12. 18 ～19	TKP レクトー レ湯河原	200	ビッグデータ基盤・応用
CREST International Symposium on Big Data Application	2020. 1. 13 ～14	アキバホール (秋葉原)	120	ビッグデータ応用
	<b>Dimitris Plexousakis</b> (Director of the Institute of Computer Science, FORTH / Head of the Information Systems Lab / Professor of Computer Science, University of Crete) "Exploring Big, Dynamic RDF/S KBs Using Summaries" <b>Norbert Graf</b> (Professor, Medical Doctor, Director of the Department for Pediatric Hematology and Oncology at the			

	Saarland University Hospital) "Kidney cancer in childhood in the era of big data, in silico oncology and artificial intelligence"			
ビッグデータ基盤・応用 合同領域会議（非公開）	2020. 8. 24 ～25	オンライン開 催	70	ビッグデータ基 盤・応用
JST・NSF 国際連携シンポ ジウム	2020. 9. 30	オンライン開 催	230	NSF 講演は録画に より実施
	<b>Gurdip Singh</b> (Division Director, NSF CISE/CNS) "Perspectives on Systems Research and Data Analysis" <b>喜連川 優</b> (国立情報学研究所 所長/東京大学 教授/JST CREST 「ビッグデータ基盤」研究総括) 高等教育を止めなかった我が国の取り組みと、教育におけるビ ッグデータ			
CREST「ビッグデータ応 用」公開シンポジウム	2021. 1. 23	オンライン開 催		第1期～第3期の 研究代表者によ る発表
	(Keynote は無し)			
ERCIM-JST Joint Symposium	2021. 2. 18- 19(予定)	オンライン開 催		招待者限定(非公 開)
	TBD			

- c. 研究領域国際シンポジウムへの海外の著名な研究者の招待（国内外の他の研究機関との連携・共同研究の推進）

領域主催のシンポジウムでは、総括が企画して、毎回、国際的に著名な研究者をキーノートスピーカに招いた。元 GMD (German National Research Centre for Information Technology) の所長の Dennis Tsichritzis 博士, XML トピックマップの著者で, IBM の Watson と同様のオープンシステムを開発中の Jack Park 氏, フランスの人工知能研究を先導してきた Michele Sebag 教授と Marc Schoenauer 博士, データベースと情報検索に関する著名な研究者の Christos Faloutsos 教授, 計算の複雑性の著名な研究者で 2016 年の Von Neumann メダルを受賞した Christos Papadimitriou 教授, 複雑系の著名な研究者である Stuart Kauffman 博士を始め, 世界的に著名な研究者が講演を承諾し, 中間審査や事後評価の審査にもオブザーバとして参加し参考意見を述べた。シンポジウムに続けて企画された若手研究者の合宿にも参加していただいた。Christos Papadimitriou 教授からは, 「米国において NSF 関連の会議を始め, 多くのプログラムのシンポジウムに参加してきたが, 本領域シンポジウムのように内容が濃く質の高い会議は米国でも経

験したことがない」との感想を頂いた。彼等キーノート講演者が、滞在中に個々のプロジェクトの研究代表者に国際連携・共同研究のパートナー候補を紹介することもあった。

また、9プロジェクトの全てが、冬のシンポジウムにおけるプロジェクト企画セッションにおいて、海外からの関連分野の著名な研究者を招待講演に招き、その機会を利用して、従前から継続している国際連携の強化や、新しい国際共同研究を開始した。

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。

- d. ビッグデータ基盤領域を始め他の領域との合同会議と合同国際会議の共催（他の研究領域、国内外の他の研究機関、異分野との連携・協力の推進）

本領域では、領域主催の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、2領域の合同会議と、JSTとNSFの共同主催になる合同シンポジウムを共催し、ほぼ全プロジェクトが報告を行った。さらに、情報関連のCRESTプログラムとフランスのDATAIA研究所の合同シンポジウムにも選抜プロジェクトを送り込んで参加した。

他領域との合同会議により、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、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「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」と連携し、この技術を活用した。角田プロジェクトは、ビッグデータ基盤データさきがけの山田拓司氏と連携し、リウマチ患者の腸内細菌の解析の共同研究を行った。吉田プロジェクトは、ビッグデータ基盤さきがけの酒向重行氏と超広視野可視光カメラTomoe-gozenの新たなデータ処理法に関する共同研究を行った。平藤プロジェクトはビッグデータ基盤領域と共に開催した毎年のNSFとの合同シンポジウムでのマッチングが契機となりアイオワ大学との共同研究を遂行し大きな成果を上げた。

- e. 毎年2回の領域主催国際シンポジウムと毎年1回ビッグデータ基盤領域との合同会議への全プロジェクトの参加と、1～2回の共催国際シンポジウムへの一部プロジェクトの参加（研究課題間の連携・協力の推進）

このように、全プロジェクトが、毎年2回の領域主催国際シンポジウムと毎年1回ビッグデータ基盤領域との合同会議に参加し、加えて一部のプロジェクトは毎年1～2回の共催国際シンポジウムにも参加し研究報告を行った。

このような頻度で国際シンポジウムを開催したのは本領域の特徴である。各プロジェクトにはその都度、最新の進捗報告を行うことが要求され、研究代表者と主たる研究者にはかなりの負担であったが、複数の研究代表者から、①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と、②合同シンポジウムの機会を介して、他領域の研究者との共同研究の可能性



が検討でき、実際に共同研究が開始されたとの報告があった。さらに、個々の研究者、総括、アドバイザーの全員が、常に各プロジェクトの進捗状況を詳細に把握しているという状況が実現でき、新築状況に応じて、適切な助言や軌道修正を行うことができた。

f. ERCIM-JST 合同シンポジウムの開催（国際連携の推進）

領域の国際アドバイザーからの要望もあり、ERCIM(the European Research Consortium for Informatics and Mathematics)との合同シンポジウムを、総括と国際アドバイザーの Nicolas Spyrtatos 教授が中心になって企画をした。ERCIM は、数学並びに情報科学に関する EU の主だった国立研究所が全て参加する研究所コンソーシアムである。2019 年 5 月にメンバー研究所であるクレタ島の FORTH 研究所に総括が別件で訪問した際、FORTH のコンピュータ・サイエンス研究所長の Dimitris Plexousak 博士と会い、当初は 2020 年 5 月 27 日～28 日にクレタ島 FORTH 研究所で、第一回 ERCIM-JST 合同シンポジウム「Big Data & AI」として開催することを決め、プログラム案も作成し、2019 年 10 月の ERCIM の総会で ERCIM の同意を取り、JST と開催準備を行っていたが、その後の新型コロナの感染拡大で開催が延期となった。結局、2021 年 2 月 18 日～19 日に時差を考慮して 3 時間半ずつ 2 日間にわたってオンライン開催を行うことになった。領域の 9 プロジェクトの発表はフルバージョンを予め Web 上に登録して参加者に閲覧可能とし、開催時間内では短い発表と質疑応答を行うこととし、同時に日本側からは、AIP 所長の杉山将教授、江村克己 AIP ネットワークラボ長の他、ICT 関連 CREST プログラムの総括数名の出席を予定している。今後の、EU と日本とのビッグデータならびに AI 分野での連携強化の契機になることを期待している。先方からは、ERCIM メンバーの主たる国立研究所の所長や EC からの参加が予定されている。

g. AIP チャレンジへの若手研究者の参画と他領域の若手研究者との共同研究の推進

若手研究者の研究交流と共同研究の推進を目的とした AIP ネットワークラボ主催の AIP チャレンジに各プロジェクトから若手研究者を送り込み、他領域の若手研究者との共同研究が開始された。

平藤プロジェクトの郭威特任助教（東京大学大学院農業生命科学研究科）は、AIP チャレンジへの参加を通して、CREST 知的情報処理に参画している慶応義塾大学大学院理工学研究科の小篠裕子特任助教と共同で、AIP チャレンジ課題として 2018 年には「3 次元植物表現型解析用のベンチマークデータセットの構築とアノテーションツールの開発」を、2019 年には「農作物フェノタイピングのための葉の表現型形質抽出」を申請しいずれも採択された。これらのプロジェクトの成果は平藤プロジェクトの中で大きな役割を果たした。

h. 総括ならびに国際アドバイザーによる海外の研究拠点との共同研究の仲介

総括から角田プロジェクトに国際アドバイザーの Graf 教授との共同研究を、松本プロジェクトに国際アドバイザーの 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、国際アドバイザーの Spyrtos 教授、総括及び国際アドバイザーの Waters 教授がギリシャと米国の研究者を紹介した。

吉田プロジェクトは、ギリシャ クレタ大学の天文学者および FORTH のデータ科学者との研究交流を開始し、2018 年には Andreas Zesus 准教授を東大に招き天文観測データ解析に関する研究連携を強化し、シンポジウムでの講演を依頼した。また 2019 年には研究代表者の吉田教授がクレタ大学および FORTH を訪問して今後の継続的な研究協力の計画を議論した。

### (3) 研究費配分上の工夫

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や、毎年のシンポジウムでの進捗報告を踏まえ、特に総括のねらいや助言に沿った計画の拡張に関しては、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を追加し、大浪プロジェクトには、これも当初計画にはなかったが研究の進捗によって挑戦が可能になった生体細胞動態の 4 次元の観察・計測に必要な高速 4 次元多色共焦点撮影装置の購入費用を追加した。三好プロジェクトへの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。大浪プロジェクトへのこの追加は、線虫胚の細胞の細胞膜動態の 4 次元自動計測アルゴリズムの確立、大規模な遺伝子ノックダウン胚の細胞分裂動態の 4 次元計測データを用いた計算表現型解析法の確立、4 次元微分干渉顕微鏡データを用いた細胞核動態の 4 次元計測法の確立、ゲノム編集による蛍光標識を用い、遺伝子発現動態の 4 次元計測データを取得などの成果に結実し、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可能にする技術開発へと進んだ。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。

この 2 つのプロジェクト以外にも、総括裁量経費を用いて、毎年各プロジェクトに増額申請をしていただき、進捗状況と総括のねらいとの整合性を評価して、総括の判断で追加支援を行った。総括助言に逆行する計画に関しては追加支援を見送った。

### (4) 研究領域中間評価結果への対応

研究領域中間評価では高い評価を得たが、以下の点に関しての助言を得た。研究領域中間評価報告書から、関連個所のみ抜粋して、これらに対応して行った施策を述べる。

a. 総合所見に関する所見とそれへの対応

「これまで交流がほとんどない分野からビッグデータ応用をキーワードとして採択されたプロジェクト（研究代表者をリーダーとする研究チームであるが、目標達成の計画を含むため、本研究領域ではプロジェクトという）は、プロジェクト間の連携研究や、共通基盤構築などを通じて交流し、卓越した成果を生み出している。研究領域終了後も**体験型ポータルや分析シナリオ自動化パイプラインなどの共通基盤を維持し継続できるように、民間が所有するデータも含めて産業界の参画や異なる学界との協力連携を期待する。**」

対応策として、体験型ポータルに関しては、現在、商用のさくらクラウド上に実装し、将来的には、JSTのような適切な管理運用組織に移管したいと考えている。これに関しては、期間終了までにJSTと相談を進める予定でいる。民間が所有するデータも含めて産業界の参画や異なる学界との協力連携を進めることが望ましいが、個別プロジェクトでは、すべてのプロジェクトにおいて、これに向けた努力がなされ、それぞれのコミュニティにおける共有・公開と連携が進んでいるが、ビッグデータ応用全般として、分野を限定しないコンソーシアムを組織化することは難しいと判断した。

b. 領域アドバイザーの構成に関する所見とそれへの対応

「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、幅広い分野をカバーし、大学・国研・企業のバランスも考慮した人選を実施している。さらに海外の著名な研究者5名を国際アドバイザーとして迎え、選考や領域シンポジウムを英語で行うようにして国際化していることも特筆に値する。**分野を超えた共通基盤の専門家と、そこに対する実経験と課題意識の高い産業界の有識者を強化しても良いのではないか。**」

対応策として、2017年より、名古屋大学大学院情報科学研究科の武田浩一教授にアドバイザー就任をお願いした。武田教授は、名古屋大学教授就任の前には、日本アイ・ビー・エム株式会社東京基礎研究所にて自然言語処理研究に従事され、IBM Watson システムの開発に参画した日本チームのリーダーであり、自然言語処理以外のコンピュータ・サイエンス全般にも造詣が深く、産業界の視点からの助言も可能な研究者である。産業界からは、当初より、特定非営利活動法人 ITS Japan 専務理事の天野肇氏、未病社会の診断技術研究会理事で、日立ソフトウェアエンジニアリングで、バイオ情報関連やGIS関連のシステム開発と事業化に長くかかわってこられた下田正文氏、株式会社こゆるぎ総合研究所代表取締役で、野村総合研究所で長年、AI、人工知能、IoTの研究と産業化の橋渡しを行ってこられた鈴木良介氏にアドバイザーとして就任していただいております。鈴木氏には業界紙にて、領域のプロジェクトの紹介インタビュー記事の連載もしていただいた。国際アドバイザーの Costantino Thanos 教授はピサの CNR の元所長であり、Randy Goebel

教授は、カナダのアルバータ州の研究助成機関の責任者も務めた経歴の持ち主で、ドイツやカナダの民間企業の技術コンサルティングも行っている。Nicolas Spyrtos 教授は、パリ 11 大学名誉教授であるが、ギリシャの科学技術大臣へ科学技術政策助言を行う委員会の ICT 責任者でもある。毎年の 2 回の領域主催シンポジウムには国際アドバイザーのほとんど全員が出席して助言を行ってきた。

c. 研究成果の科学技術への貢献（研究成果の国際的水準）に関する所見とそれへの対応

「領域全体で論文 453 件、特許 10 件、多数の受賞、報道約 200 件など研究成果が上がっている。

特に、三好、角田、西浦、大浪、松本の各プロジェクトにおいては各分野でのトップに位置し高い h5-index 値の論文誌に論文を発表している。例えば、三好プロジェクトの *Geophys. Res. Lett.* の h5-index は 90、角田プロジェクトの *Nature Genetics* のは 182、西浦プロジェクトの *BMC Medicine* のは 84、大浪プロジェクトの *Bioinformatics* のは 110、松本プロジェクトの *ACL* のは 67 である。h5-index が唯一の評価値では必ずしもないが、著名な論文誌や国際会議での論文発表を行うことは、国際的に非常に高い研究成果として認められることとなる。

一方で、プロジェクトによって発表論文数に偏りがあるのが若干気になるところではある。大きな研究テーマの一つである共通応用基盤技術において、科学技術に貢献した内容は必ずしも明確ではない。論文文化が難しい分野ではあるが、産業界への展開を促進する根拠となるよう今後一層の尽力を期待する。」

この指摘に関して、まず、プロジェクトによる発表論文数の偏りに関しては、分野毎の業績評価基準の違いがあることを考慮して、総括としては、ある程度以上の質と量の論文発表があれば、他のプロジェクトとの差異は問わない方針を貫き、そのような場合には、社会実装や科学的ないし社会・経済的価値創成を評価するようにした。同時に、研究課題の成果を特定論文誌の特集号として定期的に発表している場合には、よりサーキュレーションの良い学術誌への投稿を重視するよう助言を行った。個々のプロジェクトの成果が広く認められるように、受賞に向けた推薦にも、総括ならびに国内・国際アドバイザーが尽力した。

共通基盤技術としては、深層学習、データ同化、可視化分析、スパースモデリングなどがあり、とくに最初の 2 つは多くのプロジェクトのコア技術の一つとなっている。深層学習は、9 つのプロジェクトの内、西浦プロジェクトと三好プロジェクト以外では、既にコアな基盤技術として活用されており、三好プロジェクトにおいてもナウキャストの誤差の補正に適用する研究が始まっている。データ同化に関しては、三好プロジェクト、越村プロジェクト、西浦プロジェクト、吉田プロジェクトでコア技術として用いられている。これらのプロジェクトはいずれもシミュレーションによる予測やデータ生成と、観測や計測のデータとの整合性をデータ同化で図り、シミュレーションの精度を図る必要がある。

そこで、深層学習とデータ同化に関しては、2020年1月の若手合宿でチュートリアル・セッションを設け、実習を伴う講習会を開催した。講師には、データ同化に関しては、三好プロジェクトのメンバーを務め、気象研究所に移った大塚成徳博士に依頼し、深層学習に関しては、東京大学先進科学研究機構の今泉允聡准教授に依頼をした。両者とも、それぞれの分野における理論から実応用までをカバーする最前線の研究者である。

これらの共通基盤技術は、応用分野の問題に即して適応するための工夫が必要であり、その部分を普遍的な知識として誰もがすぐに適用できるようにまとめ上げるには、まだ時間を要すると思われる。これらの産業界への展開に関しても、適用に際しての工夫に関する知識の普遍化が充分でない現状では応用分野ごとの努力にゆだねるしかない。データ同化に関しては、多様な問題に適用できるように、AIと連携したデータ同化の研究開発が、三好プロジェクトの成果を基に飛躍的挑戦を行うAIPネットワークラボ加速課題として提案され採択されたことが、このような普遍化の困難さを物語っている。

吉田プロジェクトにおいて研究開発された「高性能トランザクション技法の提案とメタデータ・サーバへの導入」と「STR-R木を用いた科学ビッグデータ処理技術の開発と応用」に関する技術も、ビッグデータの高速トランザクション処理と空間的希望探索効率化のための共通基盤技術である。これらに関しては、既に三好プロジェクトによって活用が進んでおり、開発者の慶応大学の川島英之准教授を中心に産学が連携したデータ基盤研究会が創設され、これらの成果の産業界への展開が促進されている。

d. 研究成果の科学技術イノベーション創出への貢献（事業化の可能性）に関する所見とそれへの対応

「各アプリケーション分野としての科学技術への貢献は、ほとんどのプロジェクトで認められる。たとえば、世界最大規模の高速・超並列バーチャルスクリーニング及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の実現、化合物・タンパク質相互作用予測手法、30秒毎に更新する100mメッシュの30分のゲリラ豪雨予測、ある種の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、感染症流行の予測、1日で10個の超新星爆発の発見、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などがある。

一方で、ビッグデータの利活用を行うことで国際競争に勝つことが求められている重要な産業がほかにも多くある。これを考えれば現時点では産業全体へのインパクトは非常に高いとは必ずしもいえない。いくつかの研究成果は既に特許が出願されていることは、最先端の優れた技術が開発された結果である。今後は、国際特許化や審査請求により国際的な産業競争に優位に立てるようにさらなる努力を期待する。産業への展開が可能なプロジェクトに関しては産業界との連携をより緊密にして実施していただきたい。」

最初の「ビッグデータの利活用を行うことで国際競争に勝つことが求められている重要な産業がほかにも多くある。これを考えれば現時点では産業全体へのインパクトは非

常に高いとは必ずしもいえない。」との指摘は、その通りであるが、プログラムの予算規模の範囲では、すべてをカバーすることは不可能であった。総括は当初、マテリアルズ・インフォマティクスをポートフォリオに組むことを計画していたが、その直後に、この分野と触媒インフォマティクスのCREST領域が新たに立ち上がったため、本領域の重点公募分野からは外した。他には、サービス・プラットフォームを介して収集される顧客や市場など、人々の社会的・経済的活動に関するビッグデータを用いた新価値創成や、漁業ビッグデータ、物流ビッグデータなども公募対象課題として考えてはいたが、採択に適う応募がなかった。次に、「今後は、国際特許化や審査請求により国際的な産業競争に優位に立てるようにさらなる努力を期待する。産業への展開が可能なプロジェクトに関しては産業界との連携をより緊密にして実施していただきたい。」との指摘に関しては、特許申請の促進を各プロジェクトに助言するとともに、若手研究者を対象に、JSTのSciFosプログラムへの参加を促し、産業界の第一線で活躍されている方々複数名からのインタビューを受け、産業化や社会実装に向けたメンタリングを受講していただいた。上述の川嶋准教授もこのプログラムへの参加者である。

e. その他に関する所見とそれへの対応

「若手研究者合宿ワークショップは、様々な観点で有効な手法と思われるので、さらに強化していただきたい。特に、分野を超えた研究者が、忌憚のなく議論を交わす環境から、共通基盤技術の内容が徐々に見えてきたとのポイントは重要である。この研究領域は、分野を超えた研究者、9プロジェクトが一堂に会することに意味がある。従って、分野を超えた議論やプロジェクト推進は、若手だけでなく、プロジェクトの研究代表者や研究マネジメント、中堅研究者も、同様の環境やアプローチが必要である。さらに、9プロジェクトがアプリケーション分野の単位で構成されるが、それを横串で支える次世代基盤技術（共通アプリケーション基盤と統合分析基盤）のプロジェクトがあってもよかったのではないか。

本領域では、幅広い応用分野を対象としているため、それぞれの分野ごとにばらばらに活動してしまいがちなところを、研究総括の強力なリーダーシップのもとに、ビッグデータ利活用の推進に向けた共通基盤技術の確立に向けて推進している点は非常に印象的である。ビッグデータの利活用は、欧米に比べて体制整備や人材育成などインフラ面での整備が十分でないため、本領域の成果を基に、それを発展する形で後継の研究領域を立ち上げるなど継続的な推進が強く望まれる。このためには、本領域のアドバイザーから、研究総括を補佐すると共に、将来的には新たな研究領域の研究総括になるような人材が輩出されるのが理想的と考える。残りの研究期間でこのような育成を行える枠組みの検討も必要と思われる。」

これらへの対応策として、先ず若手合宿の一層の強化に関しては、若手合宿にも国際アドバイザーならびに領域国際シンポジウムに招いた海外からのキーノート講演者全員

に参加をしていただき、議論の国際化と深化を図った。国際アドバイザー、国内アドバイザー、総括の研究内容や経緯の紹介も企画した。外部から著名な講演者や、同世代で活躍している研究者も招き、講演をお願いすると共に、交流の場を設けた。領域の重要な共通基盤技術となっているデータ同化と、深層学習に関しては、講師を招いて実習付きのチュートリアルを企画した。

次に、「従って、分野を超えた議論やプロジェクト推進は、若手だけでなく、プロジェクトの研究代表者や研究マネジメント、中堅研究者も、同様の環境やアプローチが必要である。」に関しては、対応策として全プロジェクトが参加する進捗報告を兼ねた領域主催の国際シンポジウムを年2回（夏は1日間、冬は2日間）開催し、昼食時間と懇親会には、十分な時間を取り、全研究代表者が出席するようにし、国際アドバイザー、国内アドバイザー、他の研究代表者との間で、情報交換や連携計画の相談ができるように配慮し、総括から具体的に特定課題間のマッチングを推奨した。懇親会には中堅研究者も出席を行い、彼らの間でも連携の計画が進んだ。具体例は後に示す。

次に、「さらに、9プロジェクトがアプリケーション分野の単位で構成されるが、それを横串で支える次世代基盤技術（共通アプリケーション基盤と統合分析基盤）のプロジェクトがあってもよかったのではないか。」との指摘に関しては、その通りであるが、公募の性格上、一つのプロジェクトとして領域内に設けることには困難があった。第3年時に採択した、松本プロジェクトは、「大規模文献情報からの知識発見」という分野に依存しない普遍的な基盤技術の一つであるが、それ自体がビッグデータ応用における大きな挑戦的課題として誰もが認識している課題であるので、単独のプロジェクトとして採択することが可能であった。この課題に関しては、事前に、総括が、この研究分野に関して国内と国外で開催された2つのワークショップに出席し、趣旨を説明して申請を促した。先にも述べたように、共通基盤技術としては、深層学習、データ同化、可視化分析、スパースモデリングなどが考えられるが、これらはビッグデータ基盤領域の研究対象であることから、本領域としては、ビッグデータ基盤領域との連携を強化するために、合同会議や国際シンポジウムの共催、若手を対象にしたAIチャレンジへの参画に努めた。次世代基盤技術（共通アプリケーション基盤と統合分析基盤）に関しても、クラウドベースのリサーチ・インフラスト、ラクチャの構築を目指すプロジェクトがビッグデータ基盤領域に設けられている。これらの基盤要素技術の研究開発と、本領域の各プロジェクトの研究開発は同時期に並行して進んでおり、前者の成果に乗っかって後者の成果をまとめ上げることは、依存関係から実施は困難である。このようなことから、領域内に体験型ポータル開発チームを設置し、可能な限り、共通のプラットフォーム上に全てのプロジェクトの成果の一部を構築し、領域内での共有を目指した。

さらに、将来、CREST領域の総括を担当しうる人材の育成に関しては、国内アドバイザーの岩野アドバイザー、柴崎教授、武田教授、松井教授、宮野教授は、本領域開始以前から、領域総括として相応しい方々である。

## (5) その他マネージメントに関する特記事項

### a. 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。船津プロジェクトの終了後は、平藤プロジェクトの中に、研究開発グループを設置した。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あるか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。体験型ポータルを介しての成果の公開の範囲は、プロジェクト毎に公開可能なデータの範囲や取り扱いに違いがあることと、内部処理にプロジェクト自前の計算処理環境が必要なために、再利用可能なツールとして公開するのではなく、サービスとして公開するなどの違いが生じることはやむを得ないと判断した。結果として、すべてのプロジェクトの成果を実体験できるポータル・システムは構築できた。今後は一般公開に向け、運用主体とシステム管理を JST と相談する必要がある。

体験型ポータルの整備により、一般公開が可能になれば、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期待する。

### b. 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、国際アドバイザー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、先にも述べたように、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。



c. 総括ならびに国内・国際アドバイザーによる各プロジェクトの成果の広報活動

総括を始め、国内・国際アドバイザーが、各プロジェクトの成果の広報に努めた。例を挙げると、Springer Nature 社から出版予定の Encyclopedia of Big Data の Natural Hazard の項目において、アドバイザーの Nigel Waters 教授が越村プロジェクトを紹介している。その他にも、総括やアドバイザーが広報に努めたことが、いくつかの特筆すべき受賞に繋がった。

## 7. 研究領域としての戦略目標の達成状況について

### (1) 研究総括のねらいに対する研究成果の達成度

最初に、各プロジェクト毎の研究の経過と所見を述べ、その後、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新薬ターゲット発見とその合成・製造法の獲得と量子化学反応論に基づく実現性の検証、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じた知識創出基盤を確立することを目標とする。

(研究総括のねらいに対する研究成果の達成度)

研究総括のねらいは、新薬開発における候補化合物の大規模ライブラリを合成経路情報も含めて構築し、これらの膨大な数の化合物と標的タンパク質とのドッキング予測をシミュレーションではなく機械学習で高速に求め、新薬行為補として求めた化合物の合成実現性の検証を行ったうえで、これらの結果を製造プラントの最適制御につなげる *in silico*

統合新薬開発プラットフォームの要素技術と統合プロトタイプの研究開発を達成すること  
にあった。

本研究課題は、①大量のタンパク質と化合物の間のドッキング可能対を高速に得る機械  
学習手法の確立とこれを用いた創薬指針の抽出、②大規模仮想ライブラリ創出と、これにI  
を適用することによる新薬ターゲット発見とその合成・製造法の検証と獲得、③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運転監視・プロセス制御のための知識  
抽出を達成するとともに、④これらの成果を連携統合した実証統合プラットフォームを構  
築し、医薬品創薬から製造の段階を通じた知識創出基盤を確立する目標を達成した。

中間評価の時点では、④の達成度が充分ではなかったが、総括の助言を受けて、この研究  
開発に注力がなされ、実証プロトタイプ・システムが完成した。このような *in silico* 統合  
新薬開発プラットフォームで得られた候補化合物のその合成経路に関する情報を③の製造  
プラントの最適監視・制御に最大限に活用するには、今後更なる研究開発が必要である。

これらの成果は、産業界との密な連携によって、実用化が進んでいる。

## 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。

(研究総括のねらいに対する研究成果の達成度)

研究総括のねらいは、大規模な気象観測データと大規模高速気象シミュレーションの2つ  
を実時間でデータ同化することにより、急激に変化する局地集中豪雨のような気象現象の  
高精度予測に革新をもたらすことにあり、大規模実時間データ同化技術においても革新を  
もたらすことも期待した。

本研究課題は、シミュレーション、センサ双方からの大容量かつ高速なビッグデータに対  
応した「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 30 分予報」という革命的な天気予報を世界に先駆けて実証した。領域気象モデル  
にアンサンブルカルマンフィルタを適用して、フェーズドアレイ気象レーダ、静止気象衛星

ひまわり 8 号・9 号といった新しい観測ビッグデータを，スーパーコンピュータ「京」を使ってリアルタイムに同化するシステムを構築した．扱うデータの規模と「30 秒毎に更新する 30 分予測」を実現するデータ転送，処理速度，アルゴリズムの全てにおいて，世界に革新を起こすオンリー・ワンの成果を達成した．

### 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．

(研究総括のねらいに対する研究成果の達成度)

研究総括のねらいは，東日本大震災で大きな犠牲を出した津波災害のような，日本沿海部における海底地震に伴う大規模津波災害への対策を支援する革新的技術を，ビッグデータと高速シミュレーション，高度なデータ分析と可視化技術を連携統合することにより，防災指針を与える被害推定技術などと共に，災害発生時に減災，災害対応，避難指示，被害推定を実時間で支援する技術を研究開発することにあつた．

本研究課題はこれに答え，海底センサ群を用いた津波高推定技術，スーパーコンピュータによるリアルタイム津波浸水被害予測技術，リモートセンシングによる広域被害把握技術，大地震発生直後の初動対応を支援する大地震被害リアルタイム推計システム，大規模データ同化による人流推定技術，マルチエージェント人流シミュレーション技術などの革新的技術を確立した．これらの技術成果の多くは既に国や自治体等に導入され，実運用に向けて試験運用されている．

研究は当初計画通りに遂行され，当初計画以上の成果を達成した．各要素技術の研究のみならず，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がある程度まで達成できた．すべてのシミュレーション要素が，すべて連携統合された形で社会実装・実用化を果たすことは難しいと思われるが，多くの要素技術の成果が個別ないし一部連携して国や自治体等に導入され，実運用に向けて試験運用されていることは高く評価できる点である．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）も後半に積極的に進められ、国内でのベンチャー企業の設立や米国企業との連携も進んだ。

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、特に、国際アドバイザは極めて高く評価しており、実サービスの提供に向けて、継続的発展が可能なように継続支援すべきとの意見であった。

#### 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グデータの解析基盤を支える技術の開発研究を行う。

（研究総括のねらいに対する研究成果の達成度）

研究総括のねらいは、領域開始時に盛んになりつつあったバイオメディカル情報学の中でも、特にガンや成人病を対象に、ゲノム情報だけではなく、トランスクリプトーム、プロテオーム、メタボローム、インタラクトーム、セロームなどと呼ばれる様々な網羅的な分子情報も含めたオミックス・データを総合的に収集し、これらに関係づけて分析することにより、治療に関する新知見をビッグデータから発見する革新的技術を研究開発することであった。このようなアプローチは、計画された臨床治験に基づいて獲得される治験データを対象とするものと、取得可能な多様な患者のデータを区別することなく膨大に集積したコホートデータを対象にしたものに大きく分けられる。公募の際に総括が求めたのは、前者のアプローチによって、特定疾患の個人化治療に関する知見の発見を目指すことで、personalized medicine を目指す手法の確立であった。一方、国内で臨床治験と呼ばれているものは薬の効果を調べるもので製薬会社を中心に行っており、治療法全般を対象とするものは臨床試験と呼ばれ、小規模に行われるものが殆どであった。ある程度以上の数の臨床治験データを国内で用意することは難しいという状況と、コホートデータを用いて、特定疾患の患者を層別化して分析することにより、患者の特性ごとに適切な治療法を発見することを目指した精密医療（Precision Medicine）の考えが、Personalized Medicine と同様に普及してきたことから、このようなアプローチも認めることとした。後者の立場をとるにせ

よ、臨床試験と連携して、患者毎の治療に対する反応の違いを考慮する分析手法の確立を要望した。この点に関しては、公募の際の事前説明会、面接選考時、採択後は、サイトビジットや毎年の報告会、仲介評価の際に、繰り返し要望を出し、軌道修正をお願いした。

さらに、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委員会の許諾取得等、十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。

中間評価までは、数種類以上の特定疾患に関して、コホートデータに新たに独自の工夫を加えた統計分析や機械学習手法が適用され、その新手法と得られた多様な知見が国際トップジャーナルに個別に多数発表されるということが続いた。一方で、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、サイトビジットや毎年の国際シンポジウム、中間評価の際に、総括から繰り返し、「**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行うように**」と指示した。中間評価では、さらに、「対象疾患を絞り、その疾患の患者に対する個別化医療の確立に結び付く知見を得るような分析シナリオを確立すること」を要望した。

これらの助言を受け入れ、新たに臨床試験研究との連携が加えられるなど大きな軌道修正が行われ、最終的には以下のような顕著な成果を得た。①階層混合ベイズ・特徴抽出・特徴統合を備えたセミスーパーバイズド手法と、深層学習を基盤とした実装による、個別化医療推進のための一般化フレームワークを確立した。データのボトムアップな特徴抽出とトップダウンの特徴統合によるセミスーパーバイズドアプローチにより、超高次元でヘテロ性の高い特性を持つ医学ビッグデータが解析可能になった。②肝がん、多発性骨髄腫サリドマイド臨床試験、乳がん、関節リウマチ、臨床試験研究連携による大腸がん、肉腫の多層オミックスデータとセミスーパーバイズドの手法による解析を行い、③治療奏効・予後を決めるバイオマーカーセットを発見し、回帰モデルを用いた個人化した予後・治療奏効予測モデルを構築するなど、個別化医療の推進に大きく資する成果を得た。④国際がんゲノムコンソーシアムに参画し、全がん網羅解析により新たな創薬標的を発見した。⑤科学的発見として、治療・予後に、がんのヘテロ性、免疫・微小環境が大きく関わること、そのがん種・治療特異性、治療標的新規パスウェイとマーカーの相乗効果を発見した。

## 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。

#### (研究総括のねらいに対する研究成果の達成度)

研究総括のねらいは、公募当時、注目を浴びつつあった新規の人獣共通感染症の国を跨いだ急激な感染拡大や、既存感染症のウィルス変異による感染性や毒性の強化を、原因ウィルスのゲノム情報や過去の変異過程の情報と、航空機に等による国境を跨ぐ人流データなどのビッグデータを総合的に解析することにより、感染拡大の予測や、次の変異株の予測を行い、国や自治体の感染対策やワクチン開発と備蓄に有益な指針を与えることができる革新的技術を研究開発することにあつた。

実際、研究実施期間中に、2015年の勧告でのMERSの流行、海外で人への感染と重症化の例が多数報告された鳥インフルエンザのH5N1, H7N9の流行、ナイジェリアでの毎年のラッサ熱の流行、我が国における、2018年夏以降の風疹流行の再興、2015年以降のブラジルを起点としたジカ熱の大流行、2014年の西アフリカにおけるエボラ熱のパンデミックと、2018年以降の流行、2017年のマダガスカルにおける20世紀以降最大の肺ペスト流行などがあり、本研究課題のチームは、その都度、これらの流行に対して即座にリアルタイムで分析を行い知見を発信した。研究期間終了直前から、1年間延長の期間を通じて、2020年のCOVID-19パンデミックは全世界に拡がり、研究代表者の西浦博教授は、国の新型コロナウイルス感染症対策専門家会議のメンバーとして、厚生労働省新型コロナウイルスクラスター対策班の中で、数理疫学モデルの基づく感染拡大予測と対策に従事し、1年間延長の研究は中断せざるを得なくなった(2021年度に改めて1年間の延長予定)。

このような現実の脅威を踏まえて、その予測と対策に取り組みながら、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してきた。特に、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の3点を集中的に検討し、これらに対する基盤技術を確立した。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ための基盤技術と組織を構築し、期間中に流行した感染症に適用して、これらの拡大予測と対策に国内外で大きな貢献を行った。

基盤技術の面では、①感染症の患者発生データに加えて、病原体の遺伝子情報および実

験データを含む大規模生物情報の他、ヒト移動データや気象データなどあらゆる関連データを元に、精度が高い予兆、予測を可能にし、②ビッグデータ同化技術を深化させ、時々刻々と変化する状況のデータを高頻度に取り込むことを可能にし、数時間単位の短期予測も精度良く計算可能にした。

## 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グ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。

(研究総括のねらいに対する研究成果の達成度)

総括のねらいは、すばる望遠鏡に搭載された新型カメラHSCによる撮像データを解析し、遠方の超新星爆発を検出して宇宙進化の歴史を探るとともに、銀河画像の歪みを解析して宇宙の3次元的な物質分布を導くという最先端の宇宙物理学における貢献を目指し、超大規模な観測画像ビッグデータと、理論宇宙モデルに基づく大規模シミュレーション、最先端の統計分析ならびに機械学習の最先端技術を組み合わせることにより、データ駆動型の新しい研究手法によって、超新星などの時間変動天体の検出を自動化し、ダークマターの3次元分布地図を作成するといった挑戦的課題の解決に革新をもたらすことにあった。同時に、データの規模においても、適用可能な革新的統計処理法や機械学習法の開発の困難さにおいて、極めて挑戦的な手法の研究開発を進めることにより、大規模データの高速管理検索技術と、革新的な統計分析手法、機械学習手法の提案に繋がることを期待した。

本研究課題は、次の3つの主たる成果を達成した。

①宇宙画像の時間差分解析を行なって、2016年から行われた計52晩の変動天体観測画像から65000個の光度変動天体を検出し、そのうち1800個の遠方超新星を自動的に発見することを達成した。この中には、26等級という非常に暗い超新星もふくまれている。1800個のうち58個は赤方偏移が1以上、つまり、約80億光年より遠くにあることが明らかになった。これほど遠くのIa型超新星は、それまでは、主にハッブル宇宙望遠鏡が過去10年間に実行した観測で発見された50個弱が知られているのみであるであったことを考えると、革

新的成果であるといえる。

②およそ 1200 万個の銀河の重力レンズ現象の画像解析にスパースモデリングの手法を適用しシグナル/ノイズ比の高い物質分布マップを自動作成することに成功した。

③宇宙の物質分布の大規模シミュレーションデータベースを構築し、その出力を機械学習し、密度場のサマリ統計量を高速に計算する「エミュレータ」を開発した。

本研究課題は、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ー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しつつあり総括として大きく期待している。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。終了後に AIP ネットワークラボ加速課題として採択され、時間変動が極めて速い変動天体の自動検知も課題に加え、研究が発展的に継続されている（その成果は本報告書には含めない）。

## 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。

（研究総括のねらいに対する研究成果の達成度）

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が急速に進歩し、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。本研究課題は、提案時には、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の研究開発を目指していた。

研究総括のねらいは、バイオインフォマティクスにおいて、データ駆動型研究を支えるゲノムデータ、トランスクリプトーム・データ、プロテオーム・データ、メタボローム・データなどのオミックスデータの中で、計測の自動化とデジタル化が最も遅れているフェノーム・データに関して、光学顕微鏡の工夫に加えて、高度な画像処理とAIや機械学習を用いた解析手法の適用と、遺伝子ノックダウンやゲノム編集による特定タンパク質の蛍光標識技術を駆使することにより、生物の発生過程、形態形成過程における表現型の動的変化に関して、その他のオミックス情報の相違や動的変化と関係づけて自動的にデジタル情報としてデータ取得を可能にすることにあつた。このようなバイオ・イメージング技術が確立すると、線虫などのモデル生物の発生、形態形成の全過程に渡る機序を明確にすることが可能になると考えられ、発生学、再生医療、腫瘍学などの分野に革新的な新しい計測分析ツールを提供することになる。

成果として、本研究課題は、上述のような革新的バイオ・イメージング技術を確立し、これを用いて、線虫の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を構築し、これらのデータと、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とを統合し、独自に開発した統計解析/機械学習技術とデータ可視化技術を活用して、多細胞生物(線虫)の発生のメカニズムの全貌を64細胞期まで解明することに成功し、近い将来、発生学のみならず、再生医療、腫瘍学におけるデータ駆動型研究の重要な基盤技術となるバイオ・イメージング技術の基礎を世界に先駆けオンリー・ワンの技術として確立した。

## 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。

(研究総括のねらいに対する研究成果の達成度)

研究総括のねらいは、本研究課題を設定することにより、我が国の食料供給を支える重要

な農業を、ビッグデータ応用の重要な課題として位置づけることにあった。このため、公募段階から重点課題の一つとして掲げると共に、関係者の応募を奨励した。農業の ICT 化は、大きく分けて生育環境を制御可能なファクトリ型農業での展開と、屋外の圃場での展開に分けられるが、後者に重点を置いた。圃場では、環境の変化と生育状況の変化をデータとして取得し、関連付けて分析することを可能にすることが重要である。そのため、圃場に設置可能なセンサネットワークと、ドローンを使った空中からの撮像画像から生育状況を把握できるような新しい技術の開発が望まれる。

本研究課題は、農業における栽培技術の向上、育種の効率化のため、農業ビッグデータの構築手法の開発及びビッグデータによる有用な新知見の発見を目的として、①センサネットワーク、ドローン等による作物・環境・植物共生微生物に関する時系列データの収集手法、②安全性を考慮した複数ドローン協調飛行技術、③画像によるフェノタイピング手法及び④各種ソフトウェア群を統合したツールを開発し、得られたビッグデータから農業・農学上の新知見を多数発見できることを目標とし、これらのすべてに解を与えた。

具体的には、①画像から多様な作物形質情報を自動抽出する手法、②超解像による精度向上手法、③知識発見を支援するビューア等からなるビッグデータ構築手法を開発し、④作物群落の植被率及び地上部体積と収量との関係、⑤小麦収穫適期と作物の質的形質との関係、⑥環境と共生微生物動態との関係等の有用知見を多数発見した。

## 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。

(研究総括のねらいに対する研究成果の達成度)

IBM社のWatsonが、2011年2月に米国のクイズ番組*Jeopardy!*で、連戦連勝であった2人のチャンピオンに打ち勝って以来、百科事典やウェブ上の自然言語で書かれた膨大な文書を自然言語処理で読み取り内部表現に変換して膨大な知識ベースを構築し、質問に対する回答の候補を統計的有意度と共に提示することができるシステムへの関心は、その基盤技術の拡張・一般化と、医療診断への応用を始め、科学技術全般における知識発見への応用において急速に高まり、AIの分野において、同時期より急速に技術革新と応用展開の発展を見せた深層学習と共に、現在のAI研究の2本柱の一つとなっている。米国では、DARPAがBig Dataの次に来るものとして、Big Mechanismに注目し、Big Mechanism Programを2014年からの3年間に2400万ドルの予算で遂行した。例として、代謝メカニズムのような複雑なメカニズムを考えてみると、その全貌はまだ完全には解明されていないが、論文発表を通じて、かなりの部分が解明されてきている。個々の論文には、代謝ネットワークの中のごく断片的な因果関係が発表されているにすぎないが、これらの断片的知識をネットワーク上にアセンブルすることにより、次第に大規模な代謝ネットワークが構築できる。そこには相互に矛盾するようなパスも存在している。このような矛盾は、断片的な知識の何らかの前提となる条件がまだ知られていないために、アセンブルの際に、前提条件を付加することなく統合されたことによる。その時点での統合された代謝ネットワークを用いて、代謝の相互関係に関する質問に、ネットワークを辿る推論によって回答することができる。このようなアプローチを、代謝ネットワークに限らず、あらゆる分野の知識に適用し、出版された膨大な文書情報から断片的知識を収集し、内部表現を用いて知識ネットワークに統合し、このネットワーク上での推論によって、質問に答え、新知見の発見を支援する基盤技術を、膨大な文献からの知識発見(Literature-based Knowledge Discovery)という。

総括のねらいは、日々大量に増大し続ける科学技術分野の公表論文を対象に、上述の「膨大な文献からの知識発見」の基盤技術を確立することにあつた。先端科学技術分野において、専門性が増し、分野が細かく細分化される一方で、関連分野に跨る知識や、思いもかけない他分野の知識が、革新的研究開発にとって必須になるケースが増大しつつある。直近の例でいうと、新型コロナウイルス感染症に対する治療法やワクチン開発、予防対策などの研究がまさに良い例である。感染症学、疫学、ウィルス学、ワクチン学、分子細胞学等々の学問分野に跨る知識を必要とし、採録審査前の論文が、膨大にウェブ上に速報として公開されている。個々の研究者や、個々のグループが、人力でこれらすべてに目を通して、有益な情報を抜き出すことは不可能な状況になっている。

この課題は、当初より本領域の重点課題の一つとして挙げており、そのために関連コミュニティの国内外のワークショップに総括が参加して、趣旨を説明し、応募を奨励した。

本研究課題は、このような結果として申請され採択になった。研究代表者の従前からの方

針もあり、先ずは、PDF 文書の構造解析による情報抽出から、概念および関係知識の抽出、抽出された知識と既存知識の融合による推論へいたる総合的な手法開発が行われた。PDF 文書の構造解析による情報抽出では、機械可読な形式で、本文のみならず、表やチャートの個々の数値も取り出す基盤要素技術の研究開発が行われた。抽出された知識と既存知識の融合による推論においては、当初は、個々の論文内での内容に関する推論と、複数論文間の引用関係などに関する推論に重点が置かれ、個々の論文から得られる断片的知識を複数論文に跨ってたどることによる推論に関しては研究開発が充分には進展しなかったため、総括から重ねて助言を行い、結果として、複数論文に跨る内容に関する推論と検索が可能なシステムの開発を達成した。その対象領域として当初は言語処理分野が想定されていたが、最終年度における新型コロナウイルス感染症の拡大を受けて、COVID-19 関連の論文を対象に文書検索、新知識発見、分野の俯瞰情報提示などの具体的な応用システム構築を達成した。

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。中間評価時点では、統合的なシステム構築では遅れを取っていたが、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。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、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た。事後評価時点では統合的なシステム構築においても引けを取らないシステムが構築された。

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。このような取り組みにより、複数論文に跨る推論によって初めて可能になる新知識の発見に関しても基本機能に関して目標を達成した。知識ネットワーク全体における矛盾の発見や、それを解消する前提条件の付与の支援といった高度な機能に関しては、今後の研究に期待する。

## (2) 研究領域全体として見た場合の特筆すべき研究成果

個々のプロジェクトは、国際的に比較をしても顕著な研究成果を上げた。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た。

ここでは、研究領域全体としてみた場合の特筆すべき研究成果に関して報告する。

## ①研究総括のねらいの達成度

当初の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

これらに関して，以下のようにねらいを達成した。

- 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

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。本領域の遂行により，9つの応用分野において，相互に関連する研究チームが組織化され，それまでは異なるサイロに入っていた大規模データが相互に関連付けられ，統合的に分析処理されたことにより，新しい価値創成が達成された。更には，西浦プロジェクトにおける，インフルエンザ流行予測において，三好プロジェクトの気象予測を活用することを計画している。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，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「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」と連携し，この技術を活用した。角田プロジェクトは，ビッグデータ基盤データさきがけの山田拓司氏と連携し，リウマチ患者の腸内細菌の解析の共同研究を行った。吉田プロジェクトは，ビッグデータ基盤さきがけの酒向重行氏と超広視野可視光カメラ Tomoe-gozen の新たなデータ処理法に関する共同研究を行った。

大浪プロジェクトは成果のデータライブラリやツールを国際的なコミュニティに公開して連携の促進を図っており，松本プロジェクトは多様な先端科学技術研究分野に開発の成果を提供して，連携を促進している。

- 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た。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遂行された。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われた。「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での活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、吉田プロジェクトにおけるビッグデータ基盤さがけの酒向プロジェクト「タイムドメイン宇宙観測用動画データの高速逐次処理法の開発」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。応用分野の挑戦的課題が明確に設定されているがゆえに、課題解決のための機械学習や統計解析の手法の開発が、基盤技術における革新的成果を生み出すことにもつながった。

c. 各種要素技術を組み合わせた分析シナリオ作成

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た。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統合パイプライン、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、越村プロジェクトにおけるデータ同化手法により災害被害シミュレータと人間行動シミュレータを連動させたマルチエージェント・シミュレーションによる災害対応評価技術、角田グループにおける臨床シーケンス用プログラム・パイプラインの構築と実適用、西浦プロジェクトにおける疫学データ・移動データを用いた疫学モデルの定式化によるリアルタイム予測と感染症対策の有効性分析技術の確立と、遺伝情報およびヒト移動情報を加味したリアルタイム予測のためのデータ同化基盤の確立、大浪プロジ

プロジェクトにおけるゲノム編集による蛍光標識を用いた遺伝子発現動態の 4 次元計測のパイプラインの構築, 平藤プロジェクトにおける表現型分析プラットフォーム PPHAP を用いた農業フェノタイピングのパイプラインの構築, 松本プロジェクトにおける文献からの知識発見の全体システムがその例である。

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, 結果として, すべてのプロジェクトにおいて, この選択の正しさが証明された。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た。中でも, 三好プロジェクト, 吉田プロジェクト, 大浪プロジェクトの成果は, 国際的にもオンリー・ワンの技術である。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きつつある。免疫細胞との相互作用まで分析対象に組み込んだことは更なる発展に繋がると期待できる。西浦プロジェクトは, 期間中に世界で生じた感染症の拡大に迅速に対応し, その感染拡大の危険性, 国内への侵入の危機予測を迅速に行った。2020 年の COVID-19 パンデミックに対しては, 国内における感染拡大を, シミュレーションとデータ同化による分析結果に基づいて, 最低でも 70%, 可能ならば 80% の接触削減を提言することによって, 第一波の感染拡大の危機を見事に防いだ。松本プロジェクトの目指す文献からの知識発見技術は, 今後のあらゆる科学技術分野, 更には文系の学問研究においても, 必須の研究ツールとなることは明白である。その成果の社会や学界に対するインパクトは大きく, NEDO の複数プロジェクトや物質・材料研究機構との共同研究が既に始まっている。

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 教授とも連携の相談をし, 合同シンポジウムをビッグデータ基盤領域と共催した。フランスに新しく設立された DataIA Institute とは, ディレクターの Nozha Boujemaa 教授(現在は移動), 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、国際アドバイザーを通じて、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た。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た。これらの試みはいずれも好ましい効果を上げた（詳細は各プロジェクトの成果を参照）。

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ることを目指した。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今後の育成に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、共通基盤技術に関するチュートリアルも開催し、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た。

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。

## ②共通基盤技術の構築に関する達成度

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十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を目標とした。

これには、①若手研究者合宿ワークショップ、②研究領域国際シンポジウム、③全プロジェクトの体験型ポータルの開発と公開、④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、有効に機能した。若手合宿は年 1～2 回、1～2 日間にわたって開催し、約 10 名から 30 名の若手研究者が出席した（6. a 参照）。研究領域国際シンポジウムは、8 月～10 月に 1 日と、1 月に 2 日間の予定で開催している（6. b 参照）。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った。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。

a. 領域の全てのプロジェクトの研究開発成果の体験型ポータルの開発

体験型ポータルに関しては、すべてのプロジェクトの主たる成果の一部が体験できるポータルを開発した。データに関しては、プロジェクトによっては公開が難しいものもあり、一部やダミーのデータを対象としている。また、三好プロジェクト、越村プロジェクト、吉田プロジェクトでは、スーパーコンピュータを用いてシミュレーションを行うが、この種の計算資源を体験型ポータルに公開はできないので、シミュレーションやそのデータ同化、大規模計算に関しては、体験型ポータルを介してこれらの処理を行うのではなく、計算結果のデータをパラメータを変えて可視化する機能のみを提供することとした。角田プロジェクトに関しては、ダミーデータを用いてツールを公開した。大浪プロジェクトと松本プロジェクトに関しては、すでに研究コミュニティに対して成果をインタラクティブなウェブアプリケーションとして公開しているので、これらを体験型ポータルに組み込んでいる。その他のプロジェクトに関しては、ウェブ上で公開された視覚的機能部品を組み合わせて機能連携を定義することにより複雑なアプリケーションが構築できるウェブベースの知識メディア技術の WebbleWorld を用いて、体験型ポータルを構築している。これらに関しては、視覚的部品単位で、プロジェクトを跨って再利用したり、利用者が独自に視覚的機能部品を定義導入して、これらのプロジェクトの環境で組み合わせて用いることも可能である。システムは構築され、コードと内容を含め公開は可能であるが、これをサステイナブルに維持、管理運用するには、JST の今後の判断を待たねばならない。

体験型ポータルには、各プロジェクトの対象分野の概要、研究内容、主たる論文などの情報も記載されており、研究成果の一部をインタラクティブに体験し、分析や発見を行う体験が可能なアプリケーションが埋め込まれている。アプリケーションの画面コピーをクリックすると、ウェブページ上でインタラクティブにツールの操作を体験することができる環境に移動する。

b. 共通基盤要素技術の明確化とノウハウの共有

共通応用基盤技術に関しては、当初から、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と考えていた。

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

これらの内、AとBは若手合宿ワークショップでチュートリアルを実習付きで開催し、各プロジェクトの若手研究開発担当で、ノウハウの共有を図った。DとEに関しては、総括助言を通じて、各プロジェクトの成果をパイプラインとして統合システム化することを求め、ほぼすべてのプロジェクトで、ユースケース・シナリオを明らかにし統合パイプライン化を達成した。本来、このようなシステムの構築を目指して開発されたウェブ上の知識メディア技術 WebbleWorld を用いての各プロジェクトの体験型ポータルの開発にあたって、体験型ポータルチームと各プロジェクトとの密な打ち合わせも、プロジェクト毎の可視化統合環境の構築を促進するのに役立った。Fに関しては、各プロジェクトがもともと共通のオントロジーを共有している小規模の組織で遂行されたために、オントロジーに基づくリソース管理の必要性は、本研究領域の中では顕在化しなかった。Gに関しては、体験型ポータル開発に全てのプロジェクトが協力することを通して基板技術としてのあるべき姿が議論できた。Hに関しては、これが今後のあらゆる科学技術の研究開発や文系の研究における必須の基盤技術と考えられることから、第3期の公募で、重点課題と位置づけ、対応プロジェクトを立ち上げて対応した。

(3) 研究成果の科学的・技術的な観点からの貢献

領域全体としては、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組み立て、その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・拡充し、7章(1)の①～⑨に述べたような、顕著な成果を達成した。

中でも、三好プロジェクトの30秒毎に更新する100mメッシュのリードタイム30分でのゲリラ

豪雨予測と、全球大気の大規模アンサンブルデータ同化による天気予報シミュレーション、吉田プロジェクトの超新星発見のパイプライン構築と、ダークマターの3次元分布地図の作成、大浪プロジェクトの革新的バイオ・イメージング技術の開発と発生過程解明への適用は、国際的にみてもオンリー・ワンの革新的成果である。船津プロジェクトの深層学習による標的タンパク質と新薬候補化合物のドッキング可能性の高速判断技術と、シミュレーションによる合成反応情報を含んだ大規模仮想化合物ライブラリの構築、およびソフトセンサー技術も、世界的に先駆的な成果である。越村プロジェクトの津波シミュレーションと被害推定技術も世界のトップレベルであり、角田プロジェクトのデータ分析、機械学習手法で得られた多様かつ膨大な新知見の多くは、世界のトップジャーナルに掲載されている。西浦プロジェクトは、研究期間中に流行した数多くの感染症に国際チームとも連携して、迅速に数理モデルに基づいた多くの知見を提供すると共に、予測モデルの精度向上を目指した革新的手法を構築した。平藤プロジェクトは、我が国における比較的大規模な圃場における作物の生育や糖質、タンパク質の含有量の推定や、育種の最適化をドローンやマルチセンサを用いた画像のビッグデータ解析で行う基盤技術を構築した。松本プロジェクトは、膨大な文献から父式の断片を抽出し、これらを知識ベースに構築し、新知見の発見を支援する基盤技術を構築し、すでに多様な科学技術分野での応用を目指した共同研究へと発展させている。

各プロジェクト内は勿論のこと、ビッグデータ基盤領域を始めとする他領域の研究課題プロジェクトや、その研究者との間で、応用と基盤技術の両分野の研究者の緊密な共同研究が行われたことによって、双方に大きな成果が得られた。基盤技術側の研究者にとっても、先端応用分野のデータ分析・解析に関わることにより、常に挑戦的な課題設定がなされることで、基盤要素技術自体の革新的改良や新しい手法の開発を要求されることになり、これが良い成果につながった。

以下、個々のプロジェクトに関して、特に重要な成果に限って、項目のみ記載する。

#### ① 船津プロジェクト

- a. **ドッキング・シミュレーションの代わりに深層学習を用いることで大量のタンパク質と化合物の間のドッキング可能対を高速に得る機械学習手法の確立とこれを用いた創薬指針の抽出**
- b. **50億の化合物とその合成経路情報も伴った世界最大規模の大規模仮想化合物ライブラリの構築と、これにaを適用することによる新薬ターゲット発見とその合成・製造法の獲得**

#### ② 三好プロジェクト

- a. データ同化による 30 秒毎に更新する 100m メッシュのリードタイム 30 分でのゲリラ豪雨予測
- b. 10240 個の全球大気シミュレーションの観測データとのアンサンブルデータ同化

### ③越村プロジェクト

- a. ガウスプロセスを用いた非線形回帰による海底センサ群からの津波高推定
- b. スーパーコンピュータによるリアルタイム津波浸水被害予測技術
- c. 大地震発生直後の初動対応を支援する大地震被害リアルタイム推計システムの構築
- d. データ同化手法により災害被害シミュレータと人間行動シミュレータを連動したマルチエージェント・シミュレーション

### ④角田プロジェクト

- a. 肝がん 300 症例の全ゲノム・多層オミックスデータの分析と新知見の発見
- b. 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
- c. サリドマイド投与により延命する多発性骨髄腫患者を分類するマーカー解析法の開発
- d. 早期乳がん患者の予後予測システムの開発
- e. ガン臨床試験研究との連携：FFPE 組織検体の臨床シーケンス用高精度 SNV/indel, fusion, CNA 検出アルゴリズム

### ⑤西浦プロジェクト

- a. 多変量疫学データのリアルタイム分析を通じた致死率の推定と死亡リスクの特定
- b. 病原体の遺伝子配列を利用した感染性（基本再生産数）の推定
- c. 環境変動とそれに伴う動物生態の動的変化を加味したラッサ熱の予測システムの開発
- d. わが国における風疹の時空間伝播の捕捉と空間的予防接種の効果検討
- e. ヒトの国際移動データを用いた世界的なジカ熱流行の予測の実装
- f. 肺ペスト流行データとヒト移動データのリアルタイム分析に基づく世界的流行のリスク推定

### ⑥吉田プロジェクト

- a. 観測的宇宙論への機械学習の導入
- b. 101 個の理論宇宙モデルに対する計 300 試行の数値シミュレーションを実施し、宇宙の物質分布の理論データベースの構築
- c. 深層学習を用いた超新星分類器により遠方超新星を自動分類

- d. 重力レンズ現象の統計解析により宇宙の物質密度場を再構築
- e. STR-R 木を用いた科学ビッグデータ処理技術の開発と応用
- f. 高性能トランザクション技法の提案とメタデータ・サーバへの導入
- g. 超新星分類器のマルチクラス分類への拡張

#### ⑦大浪プロジェクト

- a. 生命科学イメージングデータと生命動態定量データの統合データベース SSBD の構築
- b. 表現型特徴と遺伝子ネットワークの横断的探索を可能にする Web ベース可視化システム (PheGeNet) の構築
- c. ゲノム 3 次元構造データを 4 次元高分子ダイナミクスに変換する PHi-C 法の開発

#### ⑧平藤プロジェクト

- a. 多様な作物の形質情報を自動抽出する手法を開発
- b. 画像によるフェノタイピングのための超解像手法及び転移学習最適化手法を開発
- c. サンプル数の少ない教師データによる転移学習手法の開発
- d. 実用性の高いフェノタイピングツール PPHAP の開発
- e. 画像による知識発見を支援するビューアの開発
- f. 撮影条件に頑健な植物領域抽出と植被率の推定
- g. 能動学習の概念を利用した学習データの効率的収集手法の開発 (アイオワ州立大学との共同研究)
- h. テンサイ育種圃場における大規模実験と雑種強勢の量的評価

#### ⑨松本プロジェクト

- a. 構造化知識とテキスト知識の統合的仮説推論機構
- b. 複雑な構造をもった文を高精度で解析する手法の開発
- c. 関連判例検索手法の提案

#### (4) 研究成果の社会的・経済的な観点からの貢献

領域全体としては、公募時に申請の条件として、その成果が科学技術における新知識の創成や、社会・経済における新しい価値の創成に繋がることを条件とした。前者を目的とするのは、吉田プロジェクトと大浪プロジェクトで、残りの7つのプロジェクトは、当初より、社会的・経済的な価値創成や貢献を出口目標として計画されている。

船津プロジェクトの成果は、新薬開発の速度を革新的に速めると共に、生産プロセスの安定操業に大きく貢献するものである。三好プロジェクトの成果は、10分間といった短い時間に静かな川が急に激流になって尊い命を奪ってしまう局地的なゲリラ豪雨の発生を30分前予測を30秒ごとに更新することを可能にしたもので、東京オリンピックの2020年の開催予定時期には社会実証実験も成功裏に行われた。越村プロジェクトは東日本大震災の際の津波災害を受けて、今後の発生が危惧されている南海トラフ大地震などの日本近海での大地震発生による津波による浸水予測を、海底に設置したセンサ群からのデータとシミュレーションをリアルタイムでデータ同化することで、リアルタイムで行い、被害推定や避難誘導に繋げたもので、地震国である我が国での生活にとって大変重要なシステム技術である。既に国や自治体で採用されつつある。角田プロジェクトは、ガンを始めとする成人病に対し、コホートデータや臨床試験データを用いてオミックス解析を行い、そのメカニズムや治療法、治療指針を与えるバイオ・マーカに関して、多くの新知見を発見し、個人化医療に向けて大きな貢献を行い、健康な社会の実現に大きく貢献した。平藤プロジェクトは、北海道で行われているような大規模な農業において、農業フェノタイピングによって、育種、生育の最適化に関する有益な知見を農家に提供しうる総合支援システムを構築し、農業生産者や協同組合に提供した。我が国の食料をさせる農業の効率化と合理化に重要な貢献をした。松本プロジェクトは、科学技術分野、社会・経済学分野における大規模文献情報からの新知見の発見に限らず、将来的には、SNSやウェブ上に集積された自然言語情報から、有益な新知見や、危機をもたらすような知見を発見することも可能な基盤技術の基礎を構築しており、学術面への貢献のみならず、将来的には社会的・経済的な観点からも大きな貢献を果たしうると期待する。現に、DARPAにおけるビッグ・メカニズム・イニシアティブの計画立案の背景には、SNS上で膨大な自然言語情報から、テロ計画を未然に検知することが目的としてあった。

以下、個々のプロジェクトに関して、特に重要な成果に限って、項目のみ記載する。

#### ①船津プロジェクト

- c. 新薬候補の探索から合成可能性の検証、製造プラントまでを繋いだ統合手法の構築
- d. ソフトセンサによる化学プラントのオンライン監視技術
- e. 産業界との連携と技術移転

#### ②三好プロジェクト

- c. 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」の実現とサービス公開

d. 2020年東京オリンピック当初開催予定期間中の東京での実証デモンストレーション

③越村プロジェクト

- e. 要素技術の統合化と国や自治体での実証運用
- f. リアルタイム津波シミュレータの内閣府の災害対応システムにおける採用
- g. 国際連携の強化と、発展途上の災害多発国との共同研究
- h. 成果の普及と社会実装の促進

④角田プロジェクト

- f. 成果の産業や社会への展開，実装
- g. 成果である臨床シーケンス用プログラム・パイプラインの実適用
- h. 我が国における臨床試験研究と連携して個人化医療を目指す本格的オミックス解析研究の実施と促進
- i. 国際がんゲノムコンソーシアムに参画し，全がん種横断的な解析結果の公開

⑤西浦プロジェクト

- g. 感染症流行に即応したリアルタイムの流行予測や流行動態分析と情報発信
- h. 疫学モデルと遺伝学モデルを統合した初の予測技術の構築
- i. インフルエンザの集団免疫を組み込んだ疫学数理モデルの構築とリアルタイム予測
- j. インフルエンザの流行予測の社会実装
- k. 予防接種歴と接種率データを用いた集団免疫度の推定

⑥吉田プロジェクト

- h. 欧米が進める次世代宇宙望遠鏡計画および地上大型望遠鏡計画の立案への参画
- i. データ基盤研究会創設

⑦大浪プロジェクト

- d. 研究成果の生命動態定量データと生命科学画像データの公開と国際協力に関する活動
- e. RNAi ノックダウン線虫 *C. elegans* 胚の細胞分裂ダイナミクス定量計測データと4次元画像データの公開
- f. HDF5を基盤としたオープンフォーマットBD5の情報公開と普及活動

- g. 研究開発成果の手法, ツール, データの公開
- h. 受精卵の品質判定への応用
- i. ニコンと ROI データのデータフォーマットに関する共同研究

⑧平藤プロジェクト

- i. 農業現場及びスマート農業への寄与
- j. 北海道十勝管区及び北海道全域への拡大
- k. 自治体と企業と連携したアウトリーチ活動
- l. 成果の普及と民間活用

⑨松本プロジェクト

- d. 成果の先端科学技術分野への応用
  - a.1 先端材料開発への応用
  - a.2 医科学系への応用
  - a.3 自然言語処理研究分野への応用
- e. 成果の公開と民間との連携

(5) その他特記事項

①学術論文誌のカバーストーリーや特集号招待論文として選ばれた論文の一例

a. 三好プロジェクト

IEEE のフラッグシップ誌の 2016 年 9 月号ビッグデータ特集号の招待論文

Miyoshi, T., G.-Y. Lien, S. Satoh, T. Ushio, K. Bessho, H. Tomita, S. Nishizawa, R. Yoshida, S. A. Adachi, J. Liao, B. Geroffi, Y. Ishikawa, M. Kunii, J. Ruiz, Y. Maejima, S. Otsuka, M. Otsuka, K. Okamoto, and H. Seko, 2016: “Big Data Assimilation” toward Post-peta-scale Severe Weather Prediction: An Overview and Progress. Proceedings of the IEEE, 104, 2155–2179.

IEEE Computer のカバーストーリー論文に選定

Takemasa Miyoshi, Keiichi Kondo and Koji Terasaki, “Big Ensemble Data Assimilation in Numerical Weather Prediction”, Computer, vol.48, no. 11, pp. 15–21, 2015



b. 越村プロジェクト

毎年 Journal of Disaster Research の特集号”Disaster and Bigdata”を企画して、4 冊、約 40 編の論文を発表。

c. 角田プロジェクト

研究代表者は、2020 年の 1 年間をとっても、*Nature* 6 報 (*Nature*, 578, 102-111 (2020); *Nature*, 578, 82-93 (2020); *Nature*, 578, 129-136 (2020); *Nature*, 578, 94-101 (2020); *Nature*, 578, 122-128 (2020); *Nature*, 578, 112-121 (2020);) を含むトップジャーナル 22 報の論文上で成果を発表するなど、極めて顕著な業績を上げた。

② メディアによる報道

a. 三好プロジェクト

(ゲリラ豪雨予測技術に関する報道の例)

神戸新聞 2016 年 8 月 9 日夕刊

ゲリラ豪雨 30 分前に予測/積乱雲の成長を再現

日刊工業新聞 2016 年 3 月 10 日朝刊

ゲリラ豪雨 30 分先予測/解像度 100m 30 秒ごとに更新

朝日新聞 (大阪府) 2016 年 8 月 9 日夕刊

「京」、ゲリラ豪雨予測へ一歩/ 理研などシステム開発

日経産業新聞 2016 年 8 月 10 日朝刊

ゲリラ豪雨、正確に予測/ 理研など「京」と最新レーダーで

(ひまわり 8 号のデータとアンサンブルシミュレーションのデータ同化による気象予測に関する報道の例)

産経新聞 2018 年 1 月 18 日朝刊

スパコン「京」+ひまわり 8 号/台風・豪雨 雲とらえ高精度予測

日刊工業新聞 2018 年 1 月 18 日朝刊

気象予測 10分ごと更新/理研・気象研が新手法

日経新聞 2018年1月18日朝刊

台風・豪雨 高精度に予測/衛星データを「京」で処理

東京新聞 2018年1月18日朝刊

豪雨や台風 より正確に/スパコン「京」x衛星「ひまわり」

b. 越村プロジェクト

(津波被害推計に関する報道の例)

2017年5月26日 読売新聞

津波被害 30分以内に推計/南海トラフ 東北大など新システム

c. 角田プロジェクト

(肝臓がんのゲノム解析による分類と生存率に関する報道の例)

2016年4月12日 日本経済新聞

肝臓がん 遺伝子で6分類 - 個々に効く薬期待

2016年4月12日 毎日新聞

肝臓がん 生存率に大差 - 0~80% ゲノム変異で6分類

2016年4月12日 夕刊フジ

肝臓がん 再発や転移に差 - 遺伝子変異で6つに分類

2016年4月12日 日刊工業新聞

肝臓がんのゲノム解説 - 日本人300例 関連遺伝子10個超発見

(角田プロジェクトの成果を用いたがんゲノム検査システムの承認に関する報道の例)

2018年12月15日 毎日新聞

がんゲノム初承認へ/検査システム普及へ一歩

d. 吉田プロジェクト

(吉田プロジェクトが開発したパイプラインによる1800個の超新星の発見に関する報道の例)

2019年5月30日 日本経済新聞

東大・東北大など、すばる望遠鏡超広視野主焦点カメラ Hyper Suprime-Cam で約1,800

個の超新星を発見

### ③ プロジェクト独自の成果公開ウェブページの一例

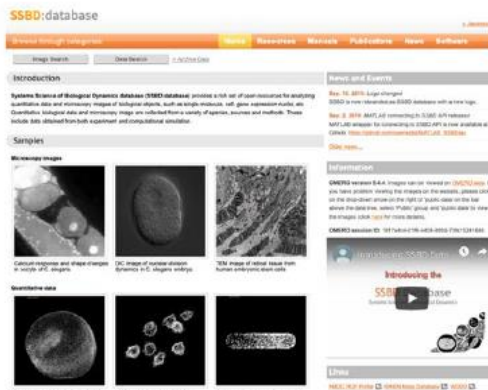
#### a. 三好プロジェクトの 3D ナウキャストによる理研天気予報



(出典: <https://wether.riken.jp>)

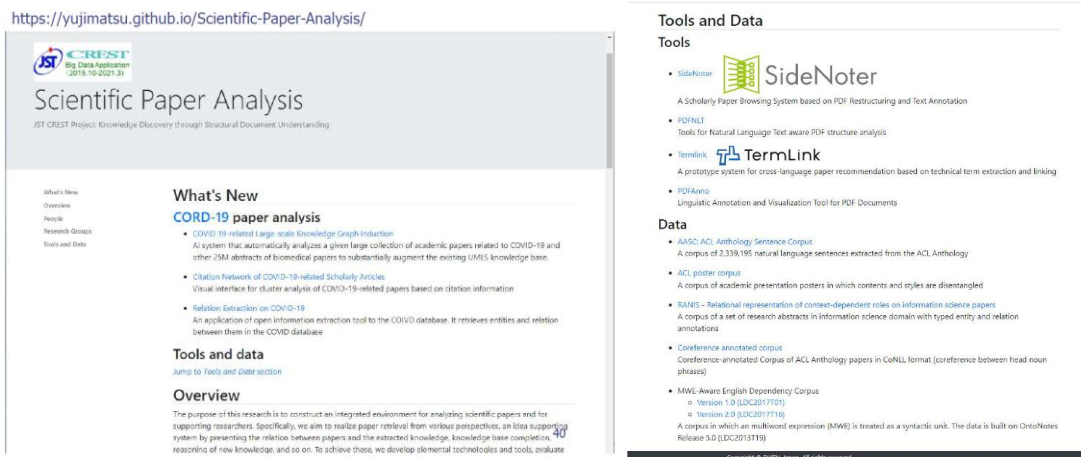
#### b. 大浪プロジェクトの SSBD の公開





(出典: <http://ssbd.qbic.riken.jp/>)

c. 松本プロジェクトの成果を科学文献解析サービスとして公開



(出典: <https://yujimatsu.github.io/Scientific-Paper-Analysis/>)

④新型コロナ感染症対策での西浦教授の活躍の例

東洋経済 2020年5月22日

東洋経済が新型コロナ「実行再生産数」を公開/感染状況を示す指標、西浦北大教授が監修

文芸春秋 digital 2020年6月23日

ドキュメント 感染症「専門家会議」 国家の命運を託された3人の研究者

数学セミナー 2020年9月号

特集1 新型コロナウイルスと戦うために数学にできること

日本循環器学会の企画 2020年7月

山中伸弥・京大教授と西浦教授のオンライン対談

## 8. 総合所見

### (1) 研究領域としての戦略目標の達成状況

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た。国際的にもオンリー・ワンの技術を確認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とダークマターの3D分布地図の作成、大浪プロジェクトの革新的バイオ・イメージング技術の構築と発生過程解明への適用など、これらの成果は世界を大きく先導している。

角田プロジェクトのデータ分析、機械学習手法で得られた多様かつ膨大な新知見の多くは、世界のトップジャーナルに多数掲載されている。西浦プロジェクトは、研究期間中に流行した数多くの感染症に国際チームとも連携して、迅速に数理モデルに基づいた多くの知見を提供すると共に、予測モデルの精度向上を目指した革新的手法を構築した。平藤プロジェクトは、我が国における比較的大規模な圃場における作物の生育や糖質、タンパク質の含有量の推定や、育種の最適化をドローンやマルチセンサを用いた画像のビッグデータ解析で行う基盤技術を構築した。松本プロジェクトは、膨大な文献から父式の断片を抽出し、これらを知識ベースに構築し、新知見の発見を支援する基盤技術を構築し、すでに多様な科学技術分野での応用を目指した共同研究へと発展させている。

### (2) 研究領域のマネジメント

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 (Harnessing data for 21<sup>st</sup> 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た。これらとの連携は、例えば、若手合宿の招待講演者として、後に触媒インフォマティクスCREST領域の研究代表者となった高橋啓介准教授を招くなど企画した。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おり、本領域の研究者にも合同会議を介してこれらの情報は提供できたが、共同研究には発展していない。

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、領域アドバイザー8人に加え、国際アドバイザー・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の全てに出席を頂

いた。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た。

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。

国際アドバイザー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た。

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

### (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点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・シフトを進めた点に、この研究領域の大きな意義があった。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共同研究がいくつか始まった。

### (4) 科学技術イノベーション創出に向けた、今後への期待、展望、課題

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。これらのプロジェクトでは、既に国や自治体、民間企業への技術提供、成果の社会実装と実運用、国内ならびに国際コンソーシアムの設立と成果の普及、他分野の大型プロジェクトへの協力などが、活発に行われている。

## (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、アドバイザーを適切に依頼し、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、国際アドバイザー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、ビッグデータ基盤領域との合同会議、DATAIA との合同会議も年2回から3回共催した。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、当初の懸念事項は解消した。

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、国際アドバイザー、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謝意を表された。驚くべきことに、国際アドバイザーは、夏のシンポジウムを欧州では休暇期間の8月に行った際に Costantino Thanos 教授が家族サービスのために欠席されたのを除くと、全員が年2回のシンポジウムすべてに出席し、若手合宿にも参加された。Costantino Thanos 教授も8月開催をずらした後はすべてに参加された。多忙な彼らが常に出席した理由を問われて答えたのは、世界中で、これほど多岐にわたる内容で深い議論がなされる機会は経験したことがなく、参加できることが自分にとってもとても貴重であるからということであった。

国際アドバイザーを数人設け、面接審査も引き受けてもらったことにより、国内アドバイザーだけでの審査結果とは異なる採択結果を得ることができた。具体的には、国際アドバイザーによる評価においては、本領域の趣旨に照らし、提案内容の革新性とその実施能力のみが書類と面接時の質疑応答で厳しく審査され、申請者の役職、年齢、経歴はさほど重視しなかった。結果的には、このような審査が、透明性のある評価を可能にし、年齢は若くても能力の際立った研究代表者を積極的に採択する結果を生み、結果として数々の革新的成果を生んだと

考えている。本領域の総括はこれまでに国内外の数多くの審査と評価に携わり、批評家者になることも数多く経験してきたが、透明性、合理性、公平性と、評価側の哲学を兼ね備えた評価会は数少ない。本領域の審査と評価は、本総括が理想と考える国際基準の方針で実施することができた。CREST で初の国際アドバイザリボードの設置を承認していただいた JST に感謝をしたい。

今後、本領域が掲げるような哲学に則った研究体制が、我が国において一層増えることを強く望む。

最後に、今後の CREST プログラムに関する要望事項として、第 3 期採択研究課題に対しても 1 年延長の機会を与えることを求めたい。現在は、1 年延長はプログラム期間内の実施とされているため、第 1 期と第 3 期の採択課題にのみ機会が与えられ、第 3 期の採択課題にはこの制度が適用されていない。本領域の場合、船津プロジェクト、角田プロジェクト、西浦プロジェクトがこれを認められた。西浦プロジェクトに関しては、1 年延長期間に新型コロナウイルスのパンデミックが生じ、西浦教授が国の新型コロナウイルス感染症対策専門家会議のメンバーとして、厚生労働省新型コロナウイルスクラスター対策班の中で、数理疫学モデルの基づく感染拡大予測と対策に従事し、1 年間延長の研究を中断せざるを得なくなり、2021 年における実施を認めたため、この間の成果の評価は行えないが、船津プロジェクトと角田プロジェクトは、5 年半の研究成果を発展的に展開し、総括の当初のねらいに沿って、成果の完成度を著しく高めた。1 年延長制度は極めて有効であり、第 3 期採択課題への適用の検討を強く要望する。

以上