

ビッグデータ時代を開拓するデータサイエンス最前線

特集1

ゲリラ豪雨を「30分前に予測せよ！」

気候変動の影響で、従来の方法では予報が追いつかない気象災害が頻繁に起きている。その代表格が、狭い地域で短時間に強い雨をもたらす局地的大雨、通称「ゲリラ豪雨」だ。理化学研究所計算科学研究機構の三好建正チームリーダーらは、膨大な観測データを高速に処理する「ビッグデータ同化」と呼ばれる手法を用いて、荒れ狂う気象に挑んでいる。

観測データを増やして 予報精度を高める

私たちの生活に天気は切っても切り離せない。出かける際には天気予報を参考に服装を選び、傘を持つかどうかを決める。予報によっては、その日の行事を延期や中止することもある。観光施設やイベントの集客、弁当や飲み物の売り上げも大きく左右するなど、経済にも大きな影響を与えるために、その重要性が注目されている。

天気予報は、観測技術やコンピューターの発達に伴って急速に進歩し、近年は予報精度がより高まった。例えば、雨風

をもたらす低気圧の位置や進路については、30年前の24時間予報より現在の72時間予報の方が確度が高いといわれる。今では明日の台風の位置を誤差100キロメートル（東京都の東西の幅程度）以内で予測できるようになった。

精度向上に大きく貢献したのがスーパーコンピューターと人工衛星による観測である。気象庁は1959年に初めて大型コンピューターを導入し、天気のシミュレーション予測（数値予報）を始めた。初期のころはあまり精度が高くなかったが、70年代後半から気象観測衛星が次々に打ち上げられるようになり、気温や水蒸気量などの膨大な観測データが得られる

ようになったことで改善された。

予報で重要なのは、集めた観測データをどのようにシミュレーションに生かすかだ。ある時点の大気の状態を出発点として、数理モデルに基づいてシミュレーション解析し、将来の天気を予測する。

「しかし、モデルは完璧ではありません。時間の経過とともに、少しずつずれが大きくなっていきます。そこで予測の精度を高めるために、一定時間ごとに観測データと突き合わせて軌道修正することが必要になります。これをデータ同化といいます。現実の気象の動きと、コンピューターの中の仮想気象の動きに橋をかけるわけです」と三好さんは言う。



三好 建正

みよし・たけまさ

理化学研究所計算科学研究機構データ同化研究チーム チームリーダー

2000年、京大学理学部卒業。同年、気象庁入庁。03年より2年間、人事院行政官長期在外研究員として米国メリーランド大学に留学し、博士号を取得。その後、気象庁予報部数値予報課技術専門官、メリーランド大学助教授を経て、13年より現職。現在、メリーランド大学大気海洋科学部客員教授、海洋研究開発機構アプリケーションラボ招聘主任研究員を兼務。13年からCREST研究代表者。08年度日本気象学会山本・正野論文賞、14年度文部科学大臣表彰若手科学者賞受賞。気象予報士。



コンピュータ計算に用いられる数値予報モデルの例。気象庁の資料を用いて作成。

気象モデルの精度と計算力を補う努力

シミュレーションは、数理モデルを使ってコンピュータの中に現実世界を模倣的に再現した仮想空間を構築する。気圧の高低で、大気はどう動くか。風が吹けば、雲はどう流れていくか。モデルは、現実には起きていることを再現できるようにつくられているが、あらゆる現象をすべて基本原理に基づいて計算しているわけではない。それをやると、いくら計算能力があっても追いつかない。実際に予報に使える計算機には限界があるので、要領よくまとめて計算できるところは簡略化して、妥

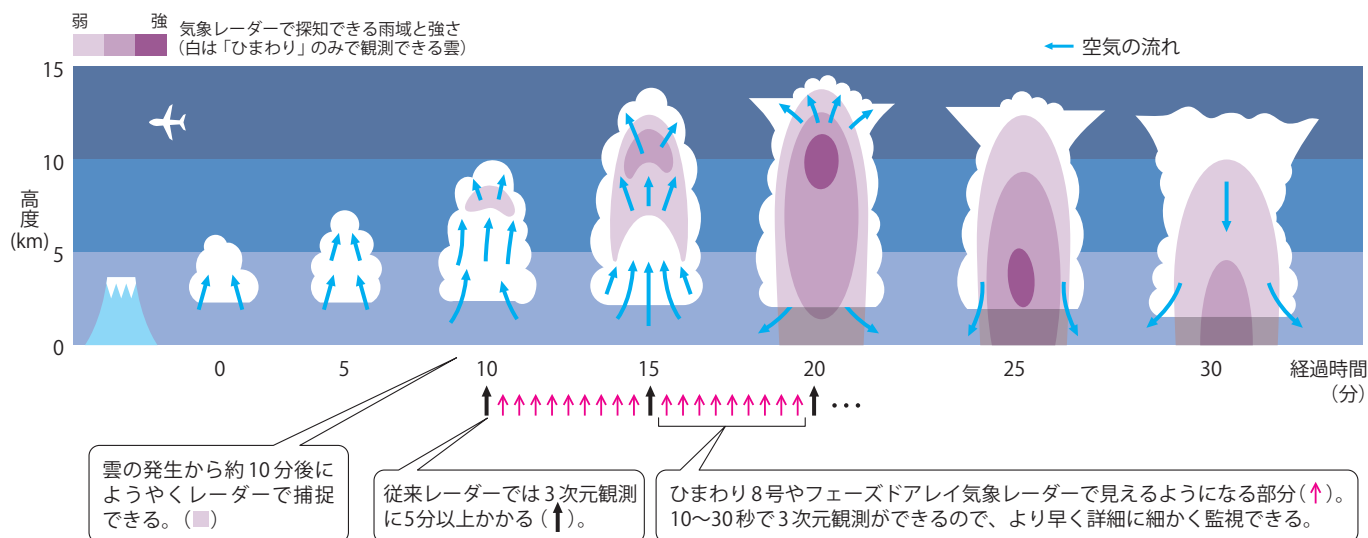
当なレベルの予測をはじき出す。簡略化するほど誤差は大きくなるため、データ同化は一層こまやかに行う必要がある。

一口に観測データといっても、人工衛星がとらえたデータもあれば、地上や定期航路の大型船上で測定したデータもある。国内だけでなく、各国から、場合によっては各地の漁業者などが測定した情報も集まる。仕様や質の違うデータをそろえてコンピュータに取り込まなければならない。

シミュレーションでは、上図中のイラストのように大気を縦、横、高さについて一定の間隔で区切り、それぞれの格子点について気温、湿度、気圧、風向き、風速な

どのデータを入力していく。その際、上図のようにシミュレーションしたい気象現象に応じた細かさの数値予報モデルが使われる。

近年、スーパーコンピュータの計算能力が飛躍的に向上し、日本全土の局地モデルでは2キロ四方で切っても計算できるレベルになった。そのためには実に数億個の数値データが必要になるが、実際の観測では、まだ何桁も少ない量のデータしか取ることができていない。“雲らしい雲”を描くには本来10メートル四方の格子で計算する必要があることから考えると、計算能力も、データ測定の緻密さも、まだまだ発展の余地がある。

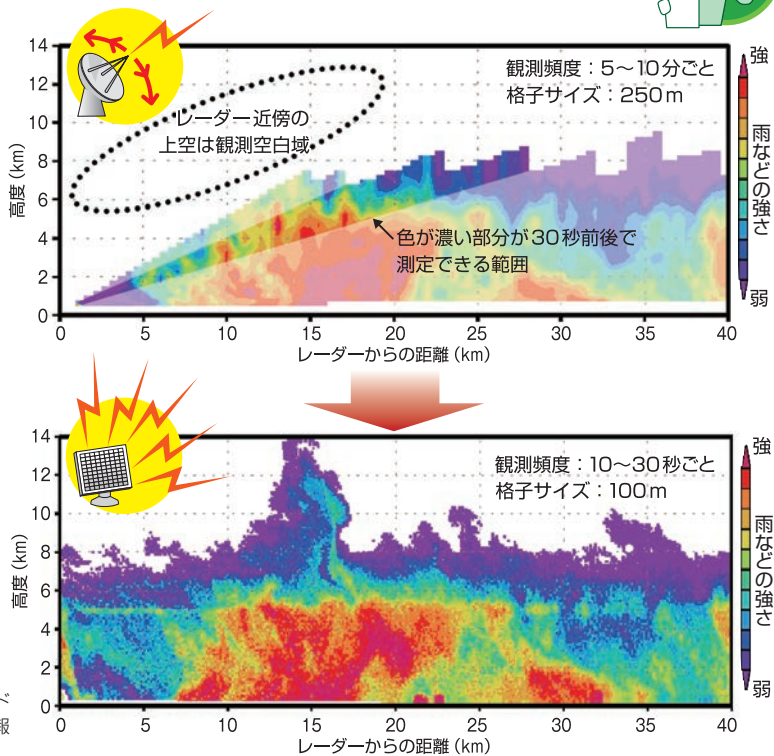


豪雨や雷雨をもたらす1つの積乱雲が発生して消滅するまでの寿命はわずか30分程度。気象レーダーが雨を探知しても約10分後には消えてしまう。衛星画像などで積乱雲の発生段階からいち早く検出・監視し、ゲリラ豪雨や突風を予報することが求められる。



大阪大学・吹田キャンパスに2012年に設置されたフェーズドアレイ気象レーダー（左、設置工事の様子）とそのアンテナ部分（上）。半径20～60km、高度14kmの詳細な降水分布が観測できる。従来5分以上かかっていた観測をわずか10～30秒に短縮した。
（写真：情報通信研究機構）

従来のパラボラ型レーダー（上）と新しいフェーズドアレイ気象レーダー（下）の性能比較イメージ（情報通信研究機構提供の別のレーダー画像より作成）



「最近、ビッグデータはいろいろな分野でクローズアップされ始めていますが、気象の分野は何十年も前からビッグデータと格闘してきました。しかも、不足しているデータをどのように補完すれば、精度のいいシミュレーションができるか、血のにじむような挑戦をしてきたのです」。

「京」と新型観測機器の登場

気象庁では国内でも有数の性能を持つスーパーコンピューターを導入し、毎時の

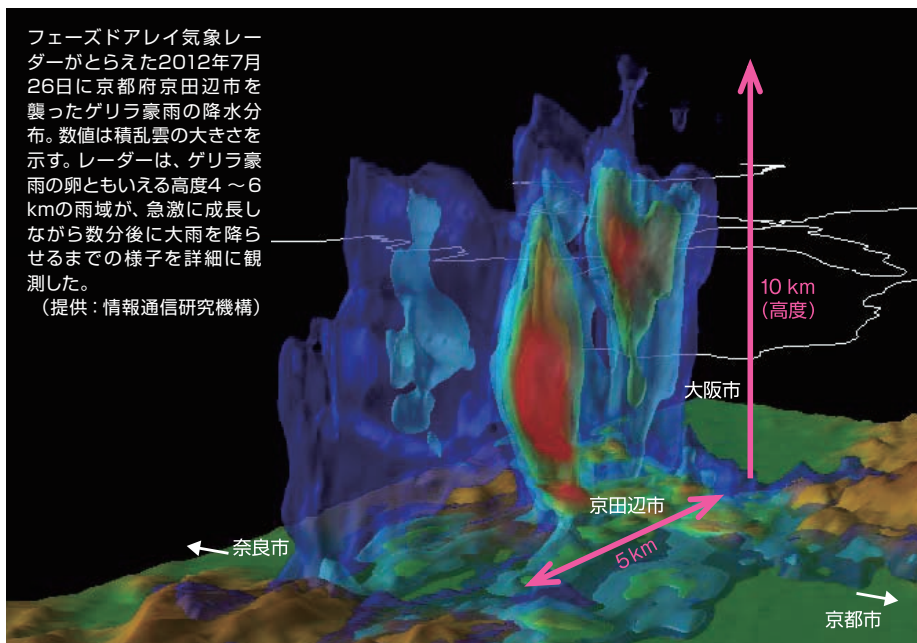
ように高度な数値予測を出している。衛星や陸上のレーダー網だけでなく、さまざまな気象観測機器から届く膨大なデータを活用しているが、まだ予測できない事象は多い。特に竜巻や雷雨などのように、影響範囲が気象モデルの格子より小さくて突発的な変化を、事前にとらえてピンポイントで予測することが何よりも難しい。

「ゲリラ豪雨による災害を減らしたいのです」と三好さんは意気込む。p.4下図のように積乱雲が急激に成長して局所的に大量の雨を降らせ、わずか10分で河川敷を

水没させる恐れもある現象だが、正確な地域や時間帯を予測することはできていない。豪雨が発生する30分前に予報して防災無線で知らせることができれば、川の増水に伴う事故が確実に減らせるはず。大量データの処理が当たり前となったビッグデータ時代だからこそ、ごく短時間で予測するシステムをつくれると考えた。

昨年よりJST CRESTで、高精細なシミュレーションと高精細な観測データを橋渡しする「ビッグデータ同化」によるゲリラ豪雨予測の実証に取り組んでいる。そのカギを握るのが、2012年に日本で初めて稼働した「フェーズドアレイ気象レーダー」と、近く打ち上げられる次世代静止気象衛星「ひまわり」、そして日本最速のスーパーコンピューターである「京」の3種の神器である。最新設備で世界にさきがける研究をしたいと考え、昨年神戸に拠点を移した。

フェーズドアレイ気象レーダー（左上写真）は、たくさんのセンサーが整然と並べられた板状のアンテナを持つレーダーだ。向いている方角の雨粒の動きを一瞬でとらえ、30秒ごとに1回転して全方位を観測する。今主流のパラボラ型レーダーが水平から上空まで何周も回るのに対し、上図のように時間的にも空間的にも何倍も細かく降水データを取り、データセンターに次々と送ることができる高性能レーダーだ。現在、



大阪大学吹田キャンパス（大阪府吹田市）と情報通信研究機構未来ICT研究所（兵庫県神戸市岩岡町）に設置されている。

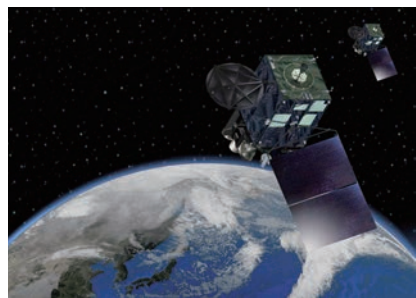
次世代静止気象衛星「ひまわり」（下図）は、現行後継の8号が今年の夏頃、またそのバックアップである9号が2016年に打ち上げられる。これらが稼働すれば、これまで30分ごとにしか撮れなかった地球全体の気象画像を10分ごとに地上へ送ることができる。情報量も段違いだ。さらに観測地域を絞り込めば、新型レーダーと同じ30秒ごとの気象画像を取得する機能も持っている。

データ量と計算力を生かす新しい天気予報システム

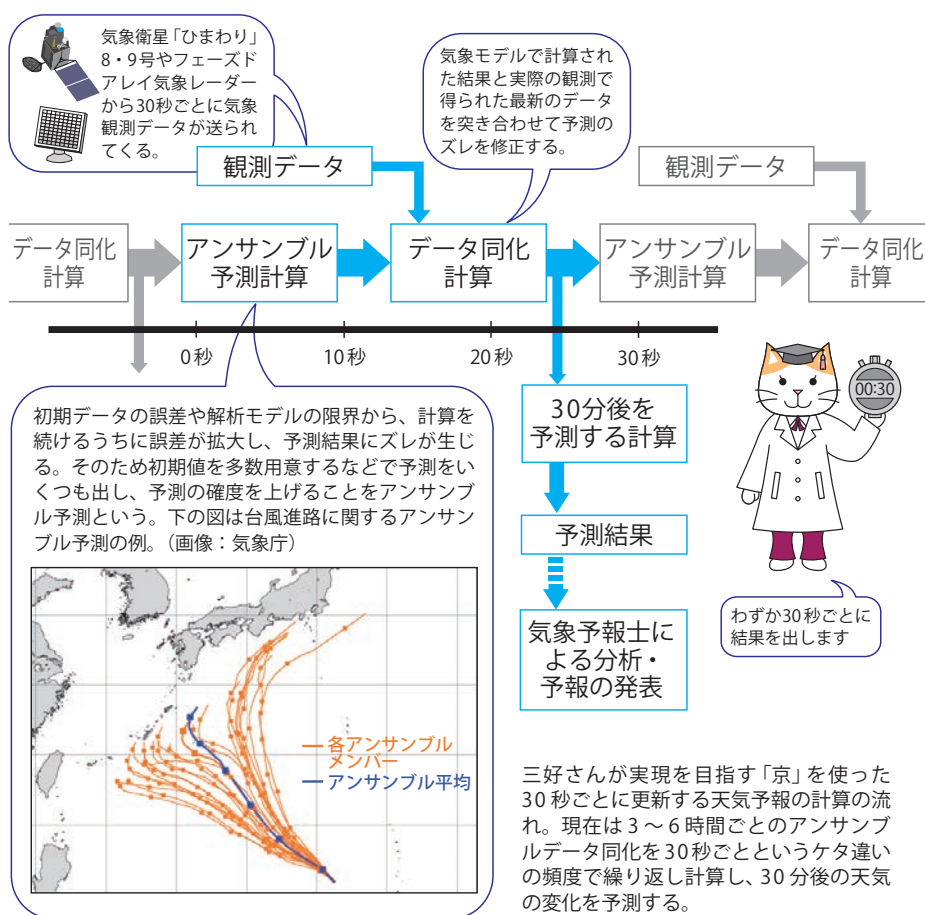
新しい観測装置が本格的に稼働すれば、観測できるデータ量に応じた詳細な予報計算が可能になるはずだ。しかし、観測データが細かくなるほど、乗り越えなければならない大きな問題が生まれてくる。

「30秒ごとにとれる観測データを生かすには、30秒ごとに天気予報の計算をしなければなりません。これは従来の天気予報では考えられないほど高い頻度です」と三好さん（右上図）。現在、気象庁で稼働させているシステムでは、1時間ごとに予報計算をしている。

新しい観測機器から得られる情報は、現在予報に使われているデータよりもはるかに多く、何十倍も細かいモデルをつくることができる。三好さんのように、格子サイズが100メートルの超高精細モデルを使って、100倍以上の頻度で予報計算をするには強力な計算力が求められる。



次世代静止気象衛星「ひまわり」8・9号。カラー画像が実現するだけでなく、赤外領域でもさらに多くの波長域で画像を取得することで、より詳しい雲分布・水蒸気分布・地表や海面等の温度などの情報が得られる。これまで約30分かかっていた観測を2倍の解像度で10分ごとに行いながら、日本周辺などの局地に限れば、30秒ごとという高頻度で観測することも可能になる。（画像：気象庁）



三好さんが実現を目指す「京」を使った30秒ごとに更新する天気予報の計算の流れ。現在は3～6時間ごとのアンサンブルデータ同化を30秒ごとというケタ違いの頻度で繰り返し計算し、30分後の天気の変化を予測する。

「これを担うのが『京』で、30秒ごとに予報計算ができる見通しが出てきました。京の計算能力の約3分1にあたる3万個のCPU（中央演算処理装置）を使えば、大量のデータをリアルタイムに処理できるシステムをつくることが可能と考えています」と自信のほどを見せた。

実際にはそこまで「京」を独占できないので、例えば1割の能力を使って1時間で行える処理量を試し、30秒で予報可能かどうかを実証する。このような換算は、同時に多数の計算をさせる並列処理ならではの。

計画では、「京」のために開発された次世代型気象モデル「SCALE（スケール）」を使って、30分後の局地的な天気の変化を予測するシステムをつくる。同時にSCALEが完成するまでのつなぎとして、現在気象庁で使われている数値予報モデルでも同様にゲリラ豪雨を予測するシステムを開発していく。

三好さんたちは、シミュレーションの確度を上げるため、「アンサンブル予測」という手法を取り入れた。観測されたデー

タの誤差を加味しながら、同時に100通りのシミュレーションを行うのだ。上図の台風進路予測のように初期データのわずかなばらつきが、まったく異なる予測結果をもたらす。はじき出されたたくさんの結果を組み合わせるデータ同化をし、最も現実に起こりうる予測をはじき出す。

アンサンブル予測やデータ同化だけでなく、データの転送やノイズ除去などの工程もばかにならない。「30秒で天気予報をしようと考えたら、それぞれの行程を数秒ずつで終わらせないとはいけません。世界でもトップクラスの『京』の能力を最大限に引き出すためにも、気象モデルや観測システムなどの幅広い研究者に参加してもらい、知恵で勝負しているところです」と言う。

ゲリラ豪雨の被害を減らすために

「いずれはライブカメラの映像やSNSを使った各地の利用者の投稿情報なども試していきたいですが、まずは新しい観測データを使いこなすこと。使える情報



が増えれば、雨の短時間予測も雲の発生からシミュレーションできるようになって、今までと質も量もまるで違う予報が可能になるのです。

現在の雨雲レーダーからの予測は、今見えている雲の動きの延長に過ぎず、新しく生まれる雲のことはわからない。ゲリラ豪雨は、積乱雲が急激に発生、成長

することで起こるので、三好さんたちの開発するシステムへの期待は高い。

「積乱雲の発生をつかんでゲリラ豪雨を予測することを目指して、30分先を詳細に予報する画期的なシステムを開発していきます。はじめは事後のデータで検証しますが、5年後にはリアルタイムの実証実験もしたいと考えています」と三好さ

んは自信を見せる。

ちょうど東京オリンピックが開催される頃だ。オリンピックの舞台で、ゲリラ豪雨予報が活用されれば、競技を実施するか一時中断するか微妙な判断も素早く行えるはずだ。世界のアスリートたちが競技に集中できる環境づくりにも、一役買うことになるだろう。

社会に役立つ日本発のビッグデータ解析技術を

田中 譲 たなか・ゆずる

CREST「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」領域 研究総括
北海道大学大学院情報科学研究科情報理工学専攻 特任教授



近年、ビッグデータは社会的にも注目を集め、ちょっとしたブームになっています。ビジネス分野では消費者の購買行動分析などで活用され始めました。しかし、その大多数はデータ量こそ多くても扱う情報の種類は限られており、分析手法を駆使して革新的な価値を生み出す段階にはまだ遠い状況です。

ビッグデータを対象とする分析や可視化、高速処理の技術は、日本でも急速に研究開発が進んでいます。しかし、科学的な発見や社会的・経済的な課題解決を目指す応用研究はまだ不十分です。ヘルスケアや災害、農業などの分野では、多種多様な大規模データを結びつけ、最も適した解析ツールや手法を組み合わせさせて分析し、実用的な成果につなげることが求められています。

こうした中、新たな知識や洞察を得るための革新的な情報技術の開発を目指して、昨年度よりJST戦略的創造研究推進事業で2つの研究領域がスタートしました（下図）。

1つは、私が担当するCRESTビッグデータ応用研究領域です。多様な情報を分析シナリオに基づいて統合的に処理し、新たな知見や価値を取り出すための横展開可能な次世代アプリケーションの開発を目指しています。そのため、大規模なデータが生み出される分野の研究者と、情報工学・コンピューターサイエンスの専門家が協力して研究を進めます。トップと中堅の組み合

わせより、双方の先頭を走る研究者同士にチームを組んでももらいたいと考えています。初年度は創業の船津公人先生（東京大学）に気象学の三好建正先生と、日本を代表するスーパーコンピューターの「京」を駆使するチームが2つ誕生しました。

もう1つは、個人型研究「さきがけ」とCRESTの複合領域であるビッグデータ基盤研究領域*です。こちらは、ビッグデータの統合解析を支えるために、大量データの高速処理・管理・検索技術などの次世代基盤技術の創出を目指しています。

領域こそ応用と基盤に分かれていますが、合同の交流会やお互いに連携するハイブリッド型の研究を通して、双方の研究者を積極的に引き合わせていきます。

日本ではあまり認識されていませんが、情報科学は大規模に展開される「ビッグサイエンス」の仲間入りをしています。世界的に見ると、日本のビッグデータ研究は遅れているのです。すでにEUの医療分野では、個人情報保護の仕組みを構築した上で先進的な研究が進められています。治験を受けた患者さんの遺伝情報、病理・免疫・代謝に関するデータや画像など、あらゆる情報を国の枠を超えて集めて分析できる環境が整っているのです。日本でも同様の動きはありますが、欧米に比べると規模も予算も小さ過ぎるといって過言ではありません。

こうした状況を打開するため、新しい2つの領域では、欧米から第一線で活躍する国際アドバイザーを招き、世界的な視野で助言を受けるとともに、これから生まれる成果を世界に発信してもらおうとしています。

今後、医療や地球環境、防災など応用研究を担う若い研究者の登場に期待します。その成果を通して日本の情報科学の重要性が広く認知され、世界に名をはせる発展につながることを願っています。

*CREST・さきがけ「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」領域（研究総括：国立情報学研究所・喜連川優 所長）

