

NII Today

National Institute of Informatics News

特集 **グリッド×e-サイエンス**

—データをつなぐ、データを活かす—

競争と協調の大規模計算ネットワーク構築

グリッドコンピューティングが牽引する第4の科学

計算資源の連携が創出する
新たな研究・開発の世界

国際標準仕様の策定を目指して



データをつなぐ、データを活かす

複数のデータをつないだ、「グリッドコンピューティング」、これにより新たな科学的発見を目指す「e-サイエンス」が注目されて久しい。グリッドの実用的なサービス提供に向け、国内の研究開発も着実にステージアップしている。今、克服すべき課題とは、その先に広がる未来とは？

NII Interview

競争と協調の大規模計算ネットワーク構築

大流行の商用クラウドは 大規模科学計算にフィットしない

高橋 コンピュータの世界は本当に進み方が速くて、最近ではクラウドが大流行で、グリッドコンピューティングという言葉はすっかり聞かなくなった気がします。

三浦 クラウドとグリッドというのはそんなに違うものではないですよ。リソース(計算資源)がどこにあるかを知らなくても、情報を投げれば計算されて答えが返ってくるという理念は共通している。クラウドはまだ話す人によってイメージするものが違うので困るのですが、我々がやっているのは背骨になるような基幹系のグリッドで、クラウドというのはもう一つ上の、よりユーザに近い階層のサービスと位置づけられる。むしろ両者は相補的と考えるべきでしょうね。

高橋 グリッドは主に科学技術計算で利用されていますが、サイエンス分野のクラウドサービスはあるのですか？

三浦 いや、ようやく出始めたところです。今のところはクラウドを供給する側はほとんどが白紙です。要するに、使い道はご自由にと。商用のクラウドサービスは帳票計算や給料計算といった定型業務を中心に始まっていますが、現状のままでは大規模な科学計算にはフィットしないのです。使ってみたら全然性能が出なかったとか、安いと思ったらけど10倍くらい高いとわかったとか、世界中でいろいろな評価報告が出て、サイエンス向けに改良されつつあります。

高橋 三浦さんは長年富士通でコンピュータ事業にかかわっていらして、2003年に国内のスパコンをつなごうというNAREGI(National Research Grid Initiative)プロジェクト(※1)が始まる時にNIIに移られたんですね。

三浦 富士通にいたときグリッドコンピューティングが注目されるようになり、私がグリッドに関していろいろ調査して状況報告をしていたら、「NAREGIを始めるからNIIにこないか」という話になったのです。蛇口をひねれば水が出るのと同じように、あまり意識しなくても誰もが使えるグリッドを構築したいとずっと考えていましたので。

高橋 NAREGIプロジェクトは2007年度で終わったわけですが、そこで作り上げたミドルウェアはどのくらい使われているのですか？

三浦 今は、7つの全国共同利用情報基盤センターと筑波大学、東京工業大学がつながっています。若い人たちにも参加してもらって一部の計算資源を提供していただいています。ただ、本格運用の一つ手前の段階で、採用規模は大学によって大小があります。さらに高エネルギー加速器研究機

構(KEK)と国立天文台との連携、東北大学の金属材料研究所や分子科学研究所などでも使っていただいています。

高橋 海外への広がりはどうでしょう？

三浦 NAREGIのソフトをダウンロードしているところはいくつかあります。10数か所あったかな。ヨーロッパが中心だったと思います。

高橋 アジアではどうなんですか？

三浦 これは中国も韓国も台湾も、EGEEというヨーロッパ製のグリッドに席卷されているといっているですね。

「自前のものではないから 受け入れない」症候群を超えて

高橋 EGEEとNAREGIの違いは？

三浦 EGEEはCERN(※2)の加速器実験のデータを世界中の共同研究機関に配るということから始まったものです。ですからCERNと同じ分野の研究に取り組んでいるKEKは、EGEEとちゃんとつながらなくては行けない。一方で、日本の研究者とはNAREGIでつながらないといけなくて、その部分のインターオペレーションを一緒になってやっているところです。そういう意味では、みんなが一つのグリッドになるというのは無理なのです。我々の分野ではNIH(Not Invented Here)シンドローム(※3)という言葉があって、誰しも自分のところでつくったものをキープしたい。だからよそのものが来ても、それはうちでつくったものじゃないよと拒む。それは当然ではないと思うので、それを超えていかにつながるかという解決策を考えている。それがインターオペレーションなのです。

高橋 とところで、中国や韓国になぜNAREGIが入らないのでしょうか？

三浦 それはなかなか難しいというか。

高橋 マーケティングをしないと無理でしょうね。

三浦 まあ、きっとそういうことでしょうね。我々は開発に注力してよそに売り込む余裕はなかったという感じです。

高橋 マーケティングが手薄になるというのは、日本の研究開発の典型的なパターンですね。

三浦 残念ながら欧米とは層の厚さが違います。

高橋 マーケティングするとしたら、NAREGIのセールスポイントは何ですか？

三浦 縦型にきちんとインテグレートされた、異機種な計算機がつながるグリッドということでしょうね。それがほかにないかという、ヨーロッパにはUNICORE(ユニコア)というのがあります。現在ヨーロッパのスパコングリッドの中核になっていますが、そのグリッドでは可能です。ア



三浦 謙一
Kenichi Miura

国立情報学研究所
リサーチグリッド研究開発センター長
アーキテクチャ科学研究系教授

アメリカのテラグリッドという大規模なグリッドでも可能ですが、大きい割には中身がばらばらな面があります。

高橋 NAREGIはヨーロッパと似たタイプなのですか？

三浦 そうですね。ヨーロッパは2001年ぐらいから着々とやっていて、我々も一時は彼らの技術を取り入れた時期もあったんですよ。ただ、開発のフェーズが合わなくなって、ちょっと独自というか、違う路線に我々がいったこともあったので、完全に同じではありません。今はインターオペレーションという意味では、ヨーロッパのUNICOREとNAREGIはお互いに計算情報を投げ合うことが実験的にできる状況までできています。国際標準として完全に成熟した状態になるには、あと1~2年かかるかもしれませんが。

高橋 NAREGIが国際標準になるのですか。

三浦 いや、そうではなくて、国際標準としてみんなで提案しているものが、ということです。まだそれぞれのグリッドと必ずしも100%フィットしていないし、足りない機能があったりしまして。それをみんなで議論しながら煮詰めている。非常にいいことだと思います。今から4~5年前、それぞれが勝手にグリッドをつくっていた時期があったのですが、「それでは連携するのに具合悪いね」という意見がここ最近クローズアップされてきました。

高橋 その国際標準づくりの作業には日本も最初から入り込んでいるんですね。

三浦 そうです。主にNAREGI時代から関係のある研究者なんですから。そういう意味では、国際的に見てもグリッドの研究をいい方向に持っていくつつあるという気はしています。UNICOREやEGEEを日本に配ればそれで済むかということ、やはりそういうものではない。言語のこともあるし、他人のふんどしで相撲をとるといのは、どこに落とし穴があるかわからない。EGEEは今まではただで使ってくださいとやっていたのに、この春から急に国ごとにお金を準備しないと保守は知らないよと言い出して、どこの国も困っています。

日本の基礎科学のひ弱さにつながる 国のインフラ軽視

高橋 保守にはどれぐらいお金がかかるのですか。

三浦 我々のグリッドセンターとしては1億数千万円ぐらいの規模で、専門のスタッフを抱えてやっています。NAREGIの後続のe-サイエンスプロジェクトが2011年度まで続き、そこまではある程度の規模でできるのですが、その先はちょっと不透明です。国も、インフラ的なものになかなか目を向けないので、へたをするとせっかくできたものがつぶれていく。日本のプロジェクトはそういうのが多いですね。

高橋 まさにそこが日本の基礎科学のひ弱さにつながる問題点ですね。今後の見通しはいかがですか？

三浦 NAREGIのもう一つ上の、よりユーザに近い階層のところを今我々はe-サイエンスの中のRENKEI (REsources liNKage for E-science) プロジェクトでやっているのですが、そこで使い勝手のいいものをつくろうとしています。

高橋 グリッドのユーザとなる研究者にとっては、計算技術のことより、自分の研究がいかにか効率よくできるかが関心の的ですからね。

三浦 その通りです。NIIの別のグループが教育用のクラウドを研究しています。それとNAREGIの環境をつないで、前処理・本計算・後処理と考えたときに前処理と後処理は手元でやって、真ん中のヘビーな部分はグリッドに投げるといっている。ユーザから見ると、ワークフロー中のジョブ(※4)を処理する環境がローカルなのかグリッドなのかの違いがあるだけで、使い勝手は全然変わりません。グリッド的な技術がないと、使う計算機ごとに異なる作法というか、プロトコル(※5)やデータの扱い方、サブミッションの方法を覚えなければいけない。それだと使うのが嫌になっちゃう。そうならないための技術開発がRENKEIプロジェクトです。

高橋 それが2012年3月で終わってしまうのですね。

三浦 だから今、実際にエンドユーザに試してもらおうと話を進めています。2011年3月に公開シンポジウムをやって、我々の成果の紹介と、一部のエンドユーザに使ってもらった報告をし、さらに来年度に向けて使ってもらいたいという人を募る。やはり、ユーザというか、どういう研究コミュニティがどういう使い方をしてくれているかというのがプロジェクト存続の一番の鍵です。そういう意味では、KEKや国立天文台の人たちが先行して使ってくれているのはありがたい。今は計算処理よりもデータインテンシブコンピューティングという、大規模データの中から目的に合ったサイエンティフィックなデータを採り出すための利用法が注目されています。他にもさまざまな活用の道が広がるグリッド。本番はまだまだこれからだと思っています。

※1 NAREGI (National Research Grid Initiative) プロジェクト: 最先端の研究や教育活動を支援するための、学術グリッドを構築することを目的とした国家プロジェクト。

※2 CERN(セルン): 巨大な加速器を用いた素粒子の研究を行っている研究所。運営資金の約90%を、20の加盟国からの出資でまかなっている。現在の正式英語名称は、European Organization Nuclear Researchであるが、前身組織の仏語名称(Conseil Européen pour la Recherche Nucléaire)の略称を現在も使用している。

※3 NIHシンドローム (Not Invented Here syndrome): 他人や組織、国が発祥であることを理由に、技術やアイデア、製品などを採用したがるしないこと。『『自前のものではないから受け入れない』症候群』『自前主義』などと呼ばれることもある。

※4 ワークフロー、ジョブ: システムの一連の処理の流れを記述したものを「ワークフロー」、ワークフロー上で一つの処理として数えられるものを「ジョブ」と呼ぶ。

※5 プロトコル: コンピュータ同士が、ネットワークを介して通信をする際のルール。通信手順、通信規約などと呼ばれることもある。HTTP、IPなどもプロトコルの一種である。



高橋 真理子
Mariko Takahashi

朝日新聞東京本社報道局
科学医療グループ記者

インタビューの一言

グリッドといえば、最近では電力網のスマート化の話題ばかり。グリッドコンピューティングの陰はほとんど薄い感じがした。だが、国際標準化の動きがこの数年で軌道に乗るなど、大規模科学計算を支える縁の下の力持ちは着実に前進していることを知った。今後の方向性を決めるのは、ユーザからのラブコールの量だろう。しかし、本当に使いやすいとグリッドの存在はユーザに意識されない。パソコンに「intel inside」のシールが貼ってあるように、エンドユーザに縁の下の力持ちの存在を知らせる工夫がいるのではないか。

グリッドコンピューティングが 牽引する第4の科学

1990年代半ば、コンピュータサイエンスの分野に登場した

「グリッドコンピューティング(以下グリッド)」というキーワード。

グリッドとは複数のスーパーコンピュータやパソコンをネットワーク化し、

大規模かつ高性能なデータ処理などを行うことを言う。

最近では、計算機だけでなくストレージや実験装置、センサなどもグリッドでつないで

仮想的なコミュニティを形成し、新たなサイエンスの知見を切り拓く試みも始まっている。

その研究の最前線を探る。



Kento Aida

合田憲人

国立情報学研究所
リサーチグリッド
研究開発センター
(アーキテクチャ科学研究系)
教授

大規模計算だけでなく、 e-サイエンス実現の手段として

情報科学やITビジネス界でその概念が浸透しつつある「グリッド」。その概要について、NII リサーチグリッド研究開発センターで、グリッドの実用化を目指す合田憲人教授は、次のように語る。「インターネットという言葉もそうですが、情報科学の分野ではキーワードが先行し、後から中身がついてくるということがよくあります。まさにグリッドもそうで、当初は、複数台のスーパーコンピュータやパソコンをインターネットでつなげて大規模な計算を行うことを指していました。ちなみに、グリッドという言葉は、高圧送電線網(パワーグリッド)に由来し、コンセントに差し込めばいつでもどこでも電気を使えるように、必要なときに必要な情報やサービスが得られることを目指すものです。

さらに最近では、コンピュータだけでなく、地理的に分散しているファイルサーバ、データベース、ストレージ、実験装置なども連携させて、仮想的なコミュニティをつくり、あたかも一つのシステムとして使えるような枠組みのこともグリッドと呼んでいます。すでに、膨大な実験データや高圧電子顕微鏡のような実験装置との連携も始まっています。

すなわちグリッドには、コンピュータ同士をつなげることで性能を高めることに加え、コンピュータだけでなく各地に分散しているデータベースや実験装置をつなぎ、データ解析をするという、二つのニーズがある。

「ただし、前者のニーズは減ってきているように思います。というのも、1台のコンピュータの性能が格段に向上したため、よほど計算数が多い限り、わざわざネットワー

ク化する必要がなくなっているからです。一方最近は、後者の、異なる運用環境上のサーバに分散して存在する情報や、各地で発生するデータを、ネットワーク化されたコンピュータで計算し、新たな知見を得るというニーズ、すなわち『e-サイエンス』実現の手段としてのグリッド技術が主流になりつつあります。これにより、分野横断的、学際的な新たな知が生み出されることに期待が集まっているのです」と合田教授は言う。

異なるデータを連携させ、 新たな知見を生み出す「e-サイエンス」

各研究機関がもつさまざまな研究データを束ね、仮想的に一つの巨大な知識として見せ、そこから新たな知の創造を目指す「e-サイエンス」。従来の実験科学、理論科学、計算科学に対して、膨大なデータを連携、高度に処理・活用するe-サイエンス(データ中心科学)は、第4の科学とも言われ、現在、科学の新潮流として注目されているのだ。

独立行政法人 産業技術総合研究所 情報技術研究部門でグリッド技術を研究している小島功主任研究員は、具体的な研究の成果を次のように語る。

「例えば、私たちが手がけているプロジェクトに、GEO Grid(Global Earth Observation Grid:地球観測グリッド)があります。これは、衛星からのリモートセンシングデータや地質図などの地質情報を統合した、地球観測に関する巨大なデータベースで、容易にオンラインでアクセスできるe-サイエンスの先駆けです(図1)。

このように異なるデータを一つのシステムの中で扱える



Osamu Tatebe

建部修見

筑波大学
システム情報工学研究科
准教授

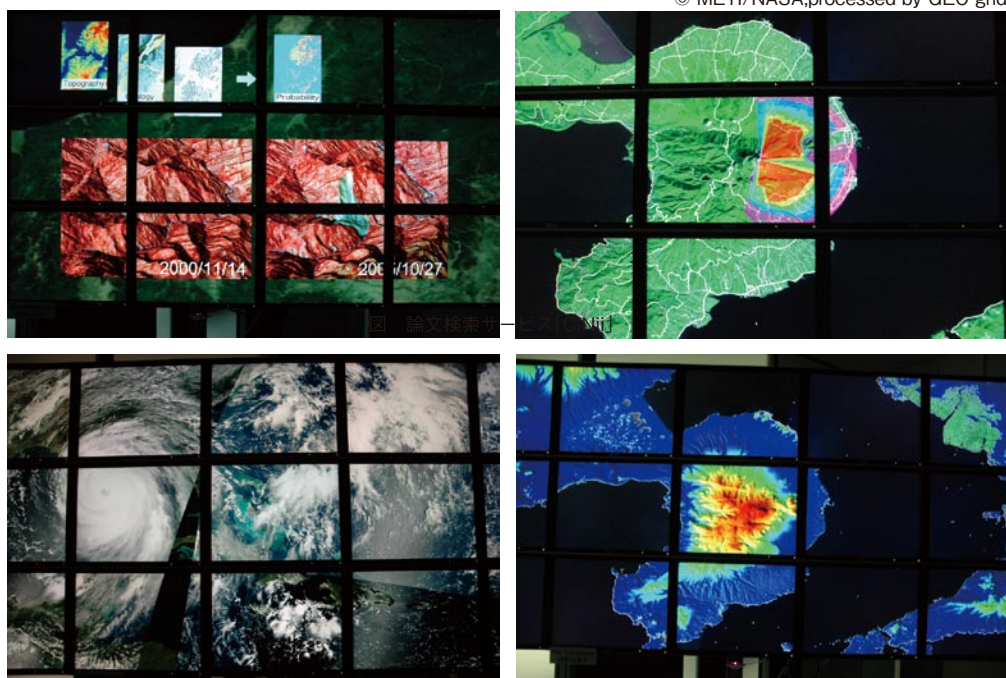


図1 GEO Gridで処理した結果を高精細ディスプレイに表示。左上:地震と地滑り、右上:雲仙普賢岳での火砕流シミュレーション、右下:雲仙普賢岳近辺の標高マップ、左下:ハリケーン

ようになると、さまざまな新たな知見を得ることができます。例えば我々は、東南アジアの熱帯雨林に立てられた観測用の鉄塔『フラックスタワー』のセンサから得られた高精度のCO₂データと、衛星で観測された広域のCO₂データを重ね合わせ、その相関関係を見出すことによりデータを補正、グローバルに高い精度でCO₂の吸収・放出量のマップをつくっています。あるいは、AMeDASによる雨量データと、衛星画像による地形の傾きや地質の硬さなどのデータを重ね、地滑りハザードマップをつくるなど、e-サイエンスが環境や防災などに活かされはじめています。

このように、膨大なデータを地球規模で連携、相補的に活用することによって、グローバルにもローカルにも、より解像度の高い情報を得られるというのが、グリッドが実現するe-サイエンスの成果の一つなのだ。

グリッドとクラウド、 コンセプトは同じ

ところで、このグリッドコンピューティング、最近脚光を浴びているクラウドコンピューティングと、どう違うのだろうか。小島主任研究員と同じく、産業技術総合研究所でグリッド研究に携わる田中良夫主幹研究員は、「クラウドは、2006年にGoogleのエリック・シュミットが使い始めた言葉ですが、実は、グリッドもクラウドも目指しているものは同じです」と言う。「いずれも、どこかにリソースがあり、なんらかのサービスが提供されるけれど、ユーザはそのリソースがどこにあるか、どういう形で提供されるか、いちいち気にすることなく使うことができます。ただ、グリッドはアカ

デミックな研究現場から生まれてきた発想で、コンピュータやサーバを束ねて高次の計算を行うことから、計算能力においても、セキュリティにおいても、技術面でクリアすべきハードルが高い。一方で、クラウドはビジネスから生まれてきた発想であり、GmailやGoogle マップのように提供するサービスや機能を絞ることで、誰もが簡単に使えるという特徴があります。つまり、グリッドとクラウドは、目指している方向は同じだけれど、発生の背景とアプローチが違うのです」と説明する。

将来的には、グリッド技術がクラウドに実装される、逆にクラウド技術がグリッドに応用されることで、一体化していくということも充分にあり得るだろう。

グリッド進展の鍵を握る グリッドミドルウェア

しかし、グリッドコンピューティングの実用化にはまだ多くの課題が残されている。グリッドの技術的課題の中でも、最重要課題といえるのが、「グリッドミドルウェア」と呼ばれるソフトウェアの開発だ。ネットワーク上の資源を連携させ、ネットワーク上に仮想組織をつくるためには、セキュリティの確保、ネットワーク上の情報の共有、計算の割り当て(スケジューリング)、ネットワーク上のデータの共有、プログラミング環境の提供など、さまざまなサービスが必要であり、これらのサービスを実現するのが、グリッドミドルウェアなのである。

「例えば、パソコンでもWindowsやMac、LinuxとさまざまなOS(オペレーティングシステム)がありますが、どのマシ



Isao Kojima

小島 功

独立行政法人
産業技術総合研究所
情報技術研究部門
サービスウェア
研究グループ長
主任研究員



Yoshio Tanaka

田中良夫

独立行政法人
産業技術総合研究所
情報技術研究部門
主幹研究員

ンからもe-mailを送ることができますね。これは、共通にやり取りできるようにインタフェースの設計がなされているからです。同様に、既存のスパコンやパソコン、データベースを連携して使えるようにするのがグリッドミドルウェアの役割。OSとアプリケーションの中間的な存在で、両者をつなぐ役割をしているためミドルウェアと呼びます。もちろんこのミドルウェアには、セキュリティや認証・許可の仕掛けも含まれます。スパコンごとにアカウントを取得しログインするような手間ひまをできるだけ削減することも、我々の研究課題の一つです」と合田教授は言う。

また、ネットワーク上の資源へのアクセスをどこまで可能にするのか、分散環境にある異種多様なデータベースをいかに簡便な統一のインタフェースで安全に連携するか、それもミドルウェアの大きな役割の一つだ。

こうした技術的な課題について、小島主任研究員は「それぞれ異なる方法で格納されている異種多様なデータを一覧できるようにするためには、統一した入れ物に収める必要があります。例えば、データによってはエクセルの表だったり、ワードのテキストデータだったりするわけですね。それぞれのデータ形式が違っていても、それを吸収し、使いやすいインタフェースを開発しなければならないのです」と語る。

「これに加え、データを扱う上での社会的な課題もあります」と指摘するのは、田中主幹研究員だ。「例えば、衛星画像データ一つをとっても、産総研、JAXA、NASAなど、研究を目的とした複数機関が保有するデータがある一方、企業がビジネス目的に保有するコマーシャルデータもあり、その連携は容易ではありません。データを束ねて使うためには、運用に関する政策や合意形成も必要になります。現状、グリッド技術では、認証・認可のセキュリティ技術は確立されてきていますが、持ち出したデータをどう使うかまでは追跡できません。あるいはバイオやライフサイエンス分野のように、プライバシーに配慮が必要なデータのセキュリティをどう守るかということも課題の一つです。データは検索できるけれど、どんなキーワードで検索したかは伏せて公開しないといったやり方など、まだ検討すべきことがあります」。

RENKEIプロジェクトで グリッドミドルウェアを開発する

こうした背景のもと、現在、NIIの合田教授、産総研の小島氏、田中氏、筑波大学システム情報工学研究科の建部修見准教授らが取り組んでいるのが、文部科学省の次世代IT基盤構築のための研究開発「e-サイエンス実現のためのシステム統合・連携ソフトウェアの研究開発」の中で進められている通称RENKEIプロジェクトである(図2)。平成20年度から4年間の研究としてスタートしたこのプロジェクトでは、研究室や計算機センター、国際的なグリッドなどの複数の組織に分散されたさまざまな資源を連携、または共有するためのミドルウェアの開発を目的としている。取りまとめ役をNIIの合田教授が担当し、分散したデータベースをいかにつなげるか(小島主任研究員)、つなげたデータベースをいかに安全な仕組みで使っていただくか(田中主幹研究員)、という二つの側面のミドルウェアの開発については産総研が担当。筑波大学の建部准教授は、ファイル単位でのデータの融合技術を手がけている。

「従来は、一つのミドルウェアですべてを動かそうという発想でしたが、実際にはアメリカ、ヨーロッパ、日本とそれぞれ別々のミドルウェアが動いていて、連携が難しかったんですね。そこで、違うミドルウェア間でもやり取りができるようにインタフェースの仕様を決めたり、スパコンと研究室にあるパソコンのように、レイヤーが違うシステムをつなげるミドルウェアを開発したり、さまざまなミドルウェアの開発を行っています。ただし、勝手に仕様を決めても技術を使ってもらうことはできませんから、OGF(Open Grid Forum)という国際標準化団体に加わり標準化活動を行いながら、オープン・スタンダードに開発を行っています(詳細は10-11P)。開発、実証実験、標準化活動を同時に進めることで、実際に使えるものを目指しています」(合田教授)。すでにプロトタイプ的设计を終えており、実証実験を経て、グリッドミドルウェアの実用化に向けて研究を加速させていくという。

「データベースをつなぐということに関しては、順調に進んでいます。たとえば、オーストリアを中心とした分散した

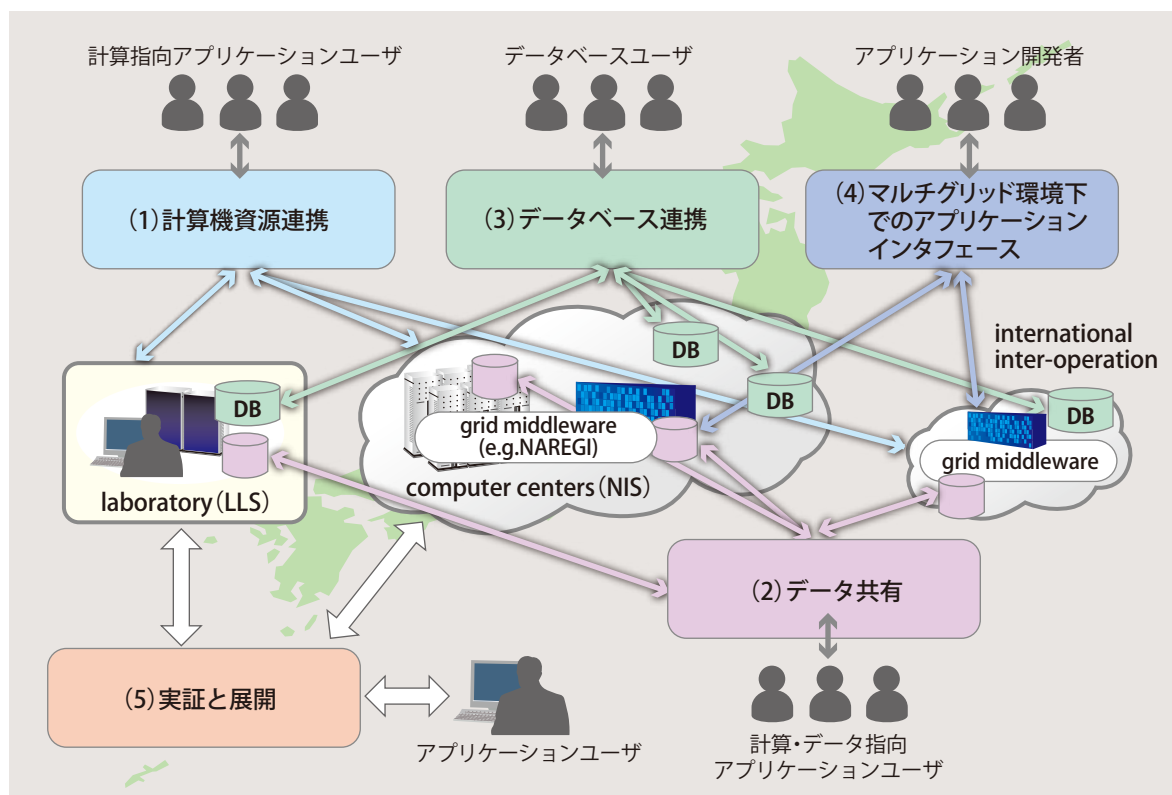


図2 RENKEIプロジェクトの全体像。計算機資源連携から実証と展開まで、各メンバーはそれぞれのミッションに従事している。

研究者が『呼気分析』に関する共同研究を進めているのですが、そのプロジェクトに私たちが開発したミドルウェアを提供して分散データベースを構築し、改善点や要望を聞き出し、フィードバックしています。今後は、データが膨大に増えても、扱いやすさや検索速度が変わらないようにすることが大きな課題ですね」と、小島主任研究員は言う。

また、建部准教授が開発した分散ファイルシステムは、すでに実用化されていて、ソフトウェアの公開・提供も始まっている。

「従来、研究室のマシン、計算センターのスパコンなど異なる組織のマシン間でデータ共有を行う場合、ファイルを転送しコピーするという手段が用いられてきました。ファイルのコピーを作成してしまうと、どのコピーが最新か、最新のコピーはどこにコピーされているかなど、人手で管理する必要があり煩雑になります。この問題を解決するため、異なる組織間で共有可能なGfarm広域ファイルシステム(※1)を研究開発しています。最近流行のDropboxを思い浮かべてもらえるとわかりやすいと思います。研究室のマシンからも計算センターのスパコンからも同一のフォルダが参照可能になりますので、どこにいても同一の環境でプログラムが実行でき

るようになります。これは素晴らしいことです。RENKEIにおいては特にシステムの高信頼化を行っています。一部システムが不意に障害を起こしてもファイルアクセスを続けられるようになりました。基本的には単一故障点を作らないこと、単一故障点の故障については復旧するまでブロックし、アプリケーションに故障を伝えないことが重要です。これまでも天文データ解析ワークフローや、MapReduce(※2)アプリケーションなどで性能評価を行ってきましたが、もっと多くの研究者のe-サイエンスに活用してもらえたらと思っています」。

まさに今が実用化に向けた正念場なのである。グリッドがより汎用的な技術となったとき、e-サイエンスを含め、人類がどのような新たな発見や知見を得るのか、私たちがどのようなサービスを受けられるのか、進展が楽しみだ。

(取材・構成 田井中麻都佳)

※1 Gfarm広域ファイルシステム: 複数のPC、クラスターなどで大規模、高性能な共有ファイルシステムとして利用することができるソフトウェア。建部准教授を中心に、オープンソースで研究開発が進められている。

※2 MapReduce: 大規模データを並列分散処理するためのフレームワークで、Googleによって考案された。

計算資源の連携が創出する 新たな研究・開発の世界

東田 学

Manabu Higashida
大阪大学
サイバーメディアセンター
応用情報システム研究部門
助教



小林泰三

Taizo Kobayashi
九州大学
情報基盤研究開発センター
特任准教授



全国の大学や研究機関にある計算機やデータベースなど、貴重な研究資源をネットワークで接続する「アカデミックグリッド」の構築が進行している。これが実現すると、膨大な計算データを用いる学問では様々な計算資源を連携させた計算処理ができるようになり、研究者の利便性を優先したフレキシブルなシステムの利用が可能になる。ユーザーが、より便利に、安全にシステムを利用できるために、現在取り組んでいる活動について、三人の研究者に話を聞いた。

国内初の アカデミックグリッド構築へ

あの大学の計算システムを利用できれば、もっと奥深い研究ができるのに…。そんな願いを持った研究者は少なくないはず。こうしたニーズを満たすために現在進められているのが、大学間連携グリッド基盤(CSI Grid)プロジェクトだ。NIIが運用支援業務を担い、スーパーコンピュータを所有する北海道大学、東北大学、東京大学、名古屋大学、京都大学、大阪大学、九州大学、筑波大学、東京工業大学の9つの大学が協力して研究開発をしている。

このような大規模なグリッドの構築による新たな可能性を、NIIリサーチグリッド研究開発センターの合田憲人教授は、次のように語る。

「計算システムを利用したいとき、ユーザはグリッドを介して他大学のシステムを利用できるようになります。さらに、あるシミュレーションを自らのシステムで実行し、グラフィックを用いた可視化処理はさらに異なるシステムで行うといった、フレキシブルな作業も実現します」。

また合田教授とともに、このプロジェクトに参加している大阪大学 サイバーメディアセンター 応用情報システム研究部門の東田学助教は、「計算システムは約4～5年で新しい設備に更新されます。更新された直後は多くのユーザが利用するため、過度に利用が集中する傾向があります。他の機関に新しい環境が

できたとき、ユーザはすぐに新たなシステムを使用したい、という欲求がわいてくるのは当然です。以前のように、フロッピーディスクやハードディスクなどに収まるような計算規模でしたら問題ありませんが、現在は研究・開発で扱うデータ量が膨大になってきたため、データを移動させるにも時間を要します。そういった背景からも、これまで使い慣れたシステムを用いながら、離れた場所にある別の計算システムを利用できるグリッド連携は必要になってくるのです」と開発の意義を語る。

配備と運用、 2方向からの取り組み

2008年8月、NIIと9つの大学は、各大学が所有する計算システムのグリッド構築のさらなる加速に向けて、「グリッド配備・運用タスクフォース」を立ち上げた。タスクフォースでは大きく「配備技術」と「運用技術」の2つの課題に対して各大学で研究開発が進んでいる。配備技術面での主たる課題はシステム上の管理方法、運用技術面ではユーザアカウント連携の仕組みづくりだ。

まずは、配備技術面の課題について見ていこう。現在CSI Gridのミドルウェアには「NAREGI」(※1)が採用されている。しかし、当初のNAREGIは、グリッド構築にあたって、導入が煩雑で配備が困難であると評価され、パッケージやインストーラの改善が急務となった。特に問題になったのは、グリッドを管

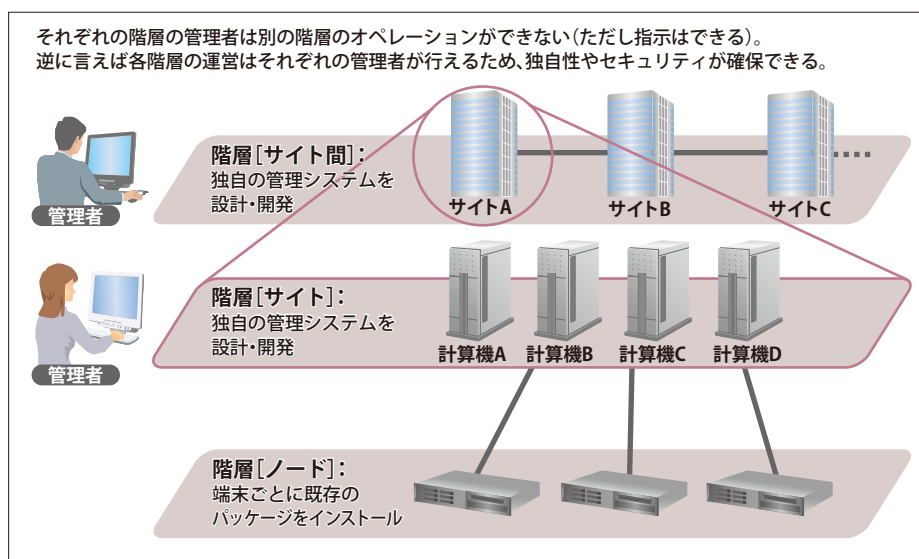


図 ネットワークを3階層に分けて管理する新たな試み

理する一つの組織がグリッド全体を管理することを前提にしていた点だ。この問題に取り組んでいるのが、九州大学情報基盤研究開発センターの小林泰三特任准教授だ。「グリッドはそれぞれの大学が管理する貴重な計算資源を出し合って構築されています。当然、各大学の計算資源の管理は、大学ごとに行っています。そうした状況にあって、計算機内部の情報まで一組織が管理するのは、到底不可能なのです」。

そこでグリッド配備・運用タスクフォースでは、管理の対象をネットワークの構成要素である「ノード」を最小単位に、ノード内／ノード間＝サイト内／サイト間の3階層に明確に分離、それぞれに管理するシステムを設計、構築することで、課題を克服しようとしている(図)。「これはCSI Gridにおける完全にオリジナルの研究成果です。その内容をごく大まかに説明すると、ノード内にある各ソフト間の整合性管理をパッケージシステムが担い、サイト内の各ノード間の整合性、さらにその上部に位置するサイト同士の管理は、それぞれ別のシステムが担当することにしたのです。こうすることで、階層ごとに管理すべき情報が整理され、各階層での整合性確保が可能になり、グリッド全体を管理できるようにしたのです」。

一方で運用技術面での改善も徐々に進んでいる。1980～1990年代にかけて、日本国内には専用のネットワーク回線をつないで計算資源の利用アカウントを共有しあう仕組みがすでにあった。当時は共通アカウントによる管理が実現されていたが、1990年代に入ってイン

ターネットが普及し始めたころから、「ユーザが主体的にインターネットにオープンに接続された計算資源を選択する時代になる」との見方が広まり、共通アカウントによる管理手法はいったん廃止された。しかし、インターネットのセキュリティの問題から期待するようなオープン化は実現しなかった。「こうした状況を改善するために私たちが導入しているのが、ミドルウェア「シボレス」(※2)を用いたサービスです」と東田助教は、新たな試みを紹介する。シボレスを利用したシステムでは、一度ID・パスワードを取得すれば、グリッド上でIDが連携され、容易に電子証明書の取得が可能となり、各大学のグリッドサービス利用が促進する。「約1万人の利用者に対し、NII一ツしか証明書の申請・発行窓口がないのではサービスになりません。その意味でも、ネットワーク上で9つの各大学のIDを連携するメリットはあります」(東田助教)。

以上のような活動に加えて、グリッド配備・運用タスクフォースでは、セキュリティ技術や課金制度についての検討も行われている。実用化に向けて一つひとつ課題をクリアしつつある状況といえよう。

大学共同利用機関法人として、NIIが進むべき道とは

現在、CSI Gridプロジェクトは3年目に入っている。今後こうしたプロジェクトを確実に運営し、実用化を実現するために、NIIにはどのよう



合田 憲人

Kento Aida

国立情報学研究所
リサーチグリッド
研究開発センター
(アーキテクチャ
科学研究系)教授

な役割が求められるのだろうか。

「NIIは大学共同利用機関法人という名称が示す通り、各大学と連携をとりながら、グリッド基盤の構築を進めていきたいと考えています。また世界に目を向けると、International Grid Trust Federation (IGTF) という、グリッドの認証局が行うべき運用基準を定めている世界機関があります。IGTFで認証を受けたグリッド認証局から発行される証明書は世界共通となります。私たちはIGTFに積極的に働きかけ、世界共通の証明書を発行できるようになりました。このように、NIIでは、大学単位ではできないグローバルな取り組みに対しても積極的に取り組んでいきたいと考えています」と合田教授は抱負を語る。東田助教も「これまで日本では、国際的な認証局を一元的に設立しようという試みは、道半ばで終わってしまいました。しかし、国際協調しながらグリッドの仕組みを作らないと、せっかく完成しても世界で認められないものになってしまいます。NIIにはぜひ世界に目を向けてほしいですね」と、海外でのプレゼンス向上へ期待を寄せる。また、小林特任准教授は、「グリッドが、膨大なデータを駆使する物理学や天文学などの研究領域で有用であることはさまざまな場面で言われてきたことです。もうそろそろ、こうした学問以外での利用方法を模索しても良い時期だろうと思います」と多分野への水平展開の可能性を示唆する。

世界へ、そして学問の新たな分野へ。日本発のアカデミックグリッドは、そのフィールドの範囲を着実に広げようとしている。重要な時期にある今日、その推進役としてのNIIの存在意義はますます大きなものとなるだろう。

(取材・構成 森本淳一)

※1 NAREGI: 前述の通りNAREGIとは国家プロジェクトの名称であるが、本文にあるように、狭義では「NAREGI」でそのプロジェクトが手がけたミドルウェアをさす。

※2 シボレス (Shibboleth) 認証: 組織をまたいだアクセス制御や認証情報の交換などを実現するミドルウェア。コンピュータネットワーク分野を研究するアメリカの共同事業体「Internet 2」のもとで開発された。

国際標準仕様の 策定を目指して

ITを用いて、世界中に蓄積された

膨大な知のストックやインフラなどをいかに共有し、有効活用するか。

それが、現代においてサイエンスを進展させる大きな鍵となる。

しかし、基盤となるグリッドの仕様がそれぞれ異なるため、

その連携によりリソースを共有するのは容易ではない。

グリッド間の連携を目的とした国際標準仕様の策定に向けた動きを追った。

共通の仕様なくして グリッドの連携は不可能

現在、サイエンスを進展させる上で、コンピュータによる計算やシミュレーションは欠かすことができない。しかし、大規模な研究・開発となると、一研究者のコンピュータや、一組織が所有する設備だけでは間に合わず、また特殊な実験装置は高額で設置できない場合もある。

そのような問題を解決する技術が、グリッドコンピューティングだ。複数の大学や研究センターをネットワークでつなぎ、さまざまな研究リソースを共有するのである。

すでに世界各地で、大規模なグリッドが構築され、日本でも、9つの大学とNIIによるサイエンスグリッド(CSI Grid)を形成し相互利用が可能になっている(2003～2008年、NAREGIプロジェクトの成果。詳細はP8-9)。

国際的な共同研究が盛んな現代、国を越えたリソース連携のニーズも必然的に高まっている。現在、日本ではNAREGIに続き、e-サイエンスを推進するためのRENKEIプロジェクトが進行中だ(詳細はP6-7)。NIIでは、研究室レベルのリソースと情報基盤センターレベルの大規模なリソースを連携させるためのグリッドミドルウェアな

どの研究開発を担当している。その中で、NIIリサーチグリッド研究開発センターの佐賀一繁特任研究員が手がけるのが、国際連携も含めた異なるグリッド間の連携だ。

「グリッドのミドルウェアの仕様はそれぞれの国や地域で異なり、ネットワークでつなげばすぐに連携可能なわけではありません。ハードウェアやソフトウェアなどのリソースを共有し、連携するためには、標準の仕様を決める必要があるのです。

NAREGIプロジェクトは、その発足当時から将来の研究リソースの国際連携を目指して研究開発に取り組んできました。開発するミドルウェアの仕様には、まだ未成熟であった国際標準を採用し、また2005年には、NAREGIのサブリーダーのひとりである東工大の松岡教授が米国のTeraGridのリーダーと共同で各国のグリッドプロジェクトに呼びかけ、異なるグリッド間の相互運用について会議を持つなど積極的に活動してきました。この会議以後、その重要性が国際的に認識されるようになり、OGFでグリッドの国際連携の活動が始まったのです」

OGFは、グリッドコンピューティングの国際標準化組織で、年3回の会議には多くの国から学術機関や企業の研究者が集まり、国際標準仕様を策定するための議論を繰り広げている。

学術分野が クラウドに移行しないのは？

かつてグリッドコンピューティングはビジネスの世界からも注目されていたが、現在取り組んでいるのは、ほとんどが学術分野だ。これはなぜなのだろうか。

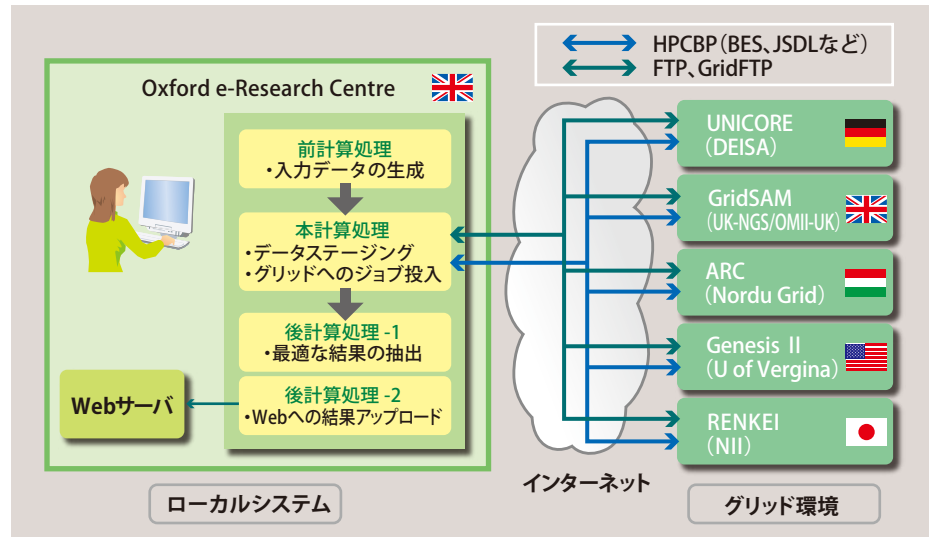
「一つは、ビジネスにおける利用がクラウドコンピューティングに移行しているからでしょう」

ビジネス用途の処理では入出力処理などCPUが空いている時間が比較的多く、1つの物理CPUに複数の仮想的なCPUを割り当て、その上

佐賀一繁

Kazushige Saga
リサーチグリッド
研究開発センター
特任研究員





で実行しても実質的な処理時間の低下はさほど大きくない。つまり、一台のコンピュータを複数のコンピュータに見せることができるので、クラウドのサービスを用いると運用コストを抑えることができるというわけだ。

「一方で学術分野、特に科学分野の処理は、常にCPUを使い続け、空き時間がほとんどない計算が多くあります。このような計算では、1つの物理CPUに複数の仮想的なCPUを割り当て、その上で実行させると処理速度が著しく低下してしまいます。このため、一台のコンピュータを複数のコンピュータに見せかけることができず、コストは自分でシステムを保有するのとさほど変わらなくなります。これでは、ビジネスとして成り立ちにくいと思います。また、最近解決されつつありますが、物理インタフェースと仮想インタフェースの橋渡しに時間がかかり、計算機間のデータ転送速度がボトルネックになるような大規模な科学計算では高い性能を出しにくいといった大きな問題が存在しています。これらのため、現在のところ、科学計算ではクラウドがあまり使われていません」

もう一つ、学術分野では広い範囲から数多くの組織がリソースを提供して一つの仮想組織として共有するというグリッドの体制面の特性も、ビジネスになりにくい要因と言えるだろう。

グリッド進展の鍵は行政が握っている

では、国や地域ごとのグリッド環境の整備状況をみていこう。インターネットのインフラ整備では、日本は世界でトップクラスだが――。

「グリッドが強いのは欧米です。特にEUでは、54か国280拠点、16,000人が参加するEGEEと呼ばれる世界最大のグリッド環境が構築されています。この他にもDEISAやNordurGridといった大規模なグリッド環境があり、この3つのグリッドを連携させるための研究開発も行っています。また、EUは自分たちのグリッドを他国でも使ってもらおうと、アジア各国やオーストラリアとの共同プロジェクトにも積極的です。

アメリカでも、主要な大学や国立研究所を結んだグリッドが構築されています。CERNに比べて拠点数は少ないですが、一拠点の規模が大きく、多くの研究者が利用しているのが特徴です」

日本では、前述したように9つの大学とNIIによってグリッドが構築されているが、欧米には大きく遅れをとっている。整備には、基盤セン

ターやユーザの意識だけではなく、行政による指導や支援が不可欠なところがある。

「EUのグリッド整備が進んでいる最大の要因は、明確なビジョンを持って政策として取り組んできたことです。アメリカでも全米科学財団(NSF)をはじめとしてエネルギー省(DOE)や国防省(DOD)などが大規模なグリッド構築や、そこで使用される計算リソースの研究開発や運用に資金を提供しています」

それに対して日本では、裾野を広げる方向には進むが、大規模な投資に見合うように有効活用するための方向性が弱いという。日本の国際競争力を高めるという観点から、行政がグリッドへの認識を深めてくれることに期待したい。

国際連携に向けた標準仕様策定の動き

それぞれの国でグリッドの規模や取り組みへの意欲は異なるが、国際連携の重要性はもはや共通認識だ。国際連携は現在、どこまで進んでいるのだろうか。

OGFにおいて、グリッド間の国際連携の実現に向けたワーキンググループが発足したのが2006年。昨年には国際標準仕様(HPCBP仕様)のプロトタイプを用いて、英国のグリッドから、日本を含む5か国の異なるグリッドにジョブを投入する実証実験が行われ、投入先のグリッドでジョブを処理し、計算結果を英国に戻すことに成功した。連携の構造は次のようになる。英国側のHPCBP仕様のクライアントを使い、インターネット経由で他国のグリッドのHPCBP仕様をサポートする計算リソースにジョブを投入する。それぞれのリソースで計算が行われ、その結果が同じ経路で英国のクライアントに返される(図)。

「しかし、現行のHPCBP仕様には認証やデータステージング(※1)に問題があることがわかりました。運用には、新たに策定されるPGI仕様

移行する必要があることが確認できました」

OGFでは現在、実験の結果を踏まえてPGI仕様の策定を進めている。本来はすでに決定している予定だったが、検討事項が多岐にわたるため、11月中旬現在、まだ議論は続行中である。

なお、佐賀特任研究員はこれまでも積極的に仕様策定の議論に参加し、NAREGIやRENKEIの要件やユースケースの提示、検出した問題点の報告などを行ってきた経験がある。本年11月にはOGFでインターオペレーションの実験などを推進するGIN-CG(Grid Interoperation Now Community Group)の共同議長に就任した。今後、HPCBP仕様のプロトタイプを用いて、RENKEIから他のグリッドの計算リソースへのジョブ投入実験を行っていく予定である。

「それに加えて、私自身は引き続き運用レベルのミドルウェアの開発に取り組んでいきます。そのためにはPGI仕様を決めなくてはならないので、積極的に議論に加わり、早期の策定に協力していきたいですね」

現在、各国ではエクサスケール(※2)のコンピュータの開発が話題になっている。もし、グリッドの国際連携が実現し、それらのスーパーコンピュータも相互利用ができるようになれば、高エネルギーやバイオサイエンス分野などのような大規模な計算が必要な研究についてもいっそうの進展が期待できる。私たちの未来を変える発見の種子は、国際連携プロジェクトの中に眠っているのだ。

(取材・構成 桜井裕子)

NII Today バックナンバーのご紹介



情報学に関する最先端の研究動向、他の教育機関や企業との連携プロジェクト、
さらには第一線で活躍する研究者の熱い思いなど、
国立情報学研究所の「今」を多彩な切り口で紹介するNII Today。
前身の「国立情報学研究所ニュース」から数え、今号で50号を迎えることができました。

号	発行年月	特 集
39	2008年3月	研究のライフライン SINET 3 (国立情報学研究所ニュースからのリニューアル第1号)
40	2008年6月	ICT社会のガバナンス—技術VS.法制度—
41	2008年9月	量子コンピュータへの道
42	2008年12月	コンピュータビジョン
43	2009年3月	知のインフラを創る
44	2009年6月	インタラクションデザイン
45	2009年9月	学術活動をいかにして計るか
46	2009年12月	礎としての数理・論理
47	2010年3月	ソフトウェア工学の新しい潮流
48	2010年6月	言語—言語を「知」として生かすために—
49	2010年9月	人工知能—個人の知から社会の知へ—

これからも読者の知的好奇心をかきたてる、質の高い広報誌を制作してまいります。

バックナンバーは、以下のアドレスから閲覧・ダウンロードが可能です。
<http://www.nii.ac.jp/publication/kouhou/today/>

今月の表紙イラスト：グリッド博士とクラウド氏。活躍のフィールドは違うが、元は血を分けた兄弟のようなもの。ネットワーク上のリソースを自由に利用できることを目指すコンセプトは同じだ。この二者が協力し合うと、コンピュータ利用はどのように深化するか。楽しみだ。

情報から知を紡ぎだす。

NII

国立情報学研究所 ニュース(NII Today) 第50号 平成22年12月

発行：大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 <http://www.nii.ac.jp/>

〒101-8430 東京都千代田区一ツ橋2丁目1番2号 学術総合センター

編集長：東倉洋一 表紙画：小森 誠 写真撮影：谷口弘幸 他 制作：株式会社 商業デザインセンター

本誌についてのお問合せ：企画推進本部広報普及チーム TEL:03-4212-2131 FAX:03-4212-2150 e-mail:kouhou@nii.ac.jp