3 e-サイエンス基盤構築のためのミドルウェア技術

宇佐見 仁英　玉川大学／国立情報学研究所
松田 秀雄　大阪大学／国立情報学研究所
小島 功　産業技術総合研究所
佐々木 節　高エネルギー加速器研究機構
東田 学　大阪大学

e-サイエンスについて

e-サイエンスは，高速ネットワーク上で計算資源（ソフトウェア含む），ストレージやデータベース（コンテンツ含む）等を活用した新たなる融合研究領域を作り出すための研究インフラであり，先進各国で整備が進められている．我が国でも国立情報学研究所を中心に最先端学術情報基盤（CSI：Cyber Science Infrastructure）の一環として整備が進められている．e-サイエンスの進展に伴って，科学技術分野の研究活動が国境を越えた国際連携や学際的な活動での知の融合へと変貌を遂げ始めている．グリッド技術は，この e-サイエンスを実現する主要な基盤技術ではあるが，日々の研究活動の身近な存在となるには，解決すべき課題が残されている．本稿では，これらの問題点とその解決に必要な新たなミドルウェア技術について解説する．

■グリッドの課題

欧米を中心にグリッドはすでに実運用フェーズに突入している．特に国際共同研究を実施するビッグサイエンスでは，IT インフラとして欠かせない存在となっている．しかしながら，必ずしも一般的な研究者が手軽に使える状況にはなっていない．英国サザンプトン大学の David De Roure 教授によるとグリッドの主な課題として 1）大規模で硬直したグリッドインフラであり，日々の研究室レベルの研究活動には適応できていない．2）計算指向のインフラであり，データ中心の設計となっていない．3）計算資源の提供サービスとの意識が強く，研究コミュニティへの参加，メンバ間の協調などの機能が弱い．などを挙げている．特に，国際共同研究での研究コミュニティでは，地理的に分散した組織が独立に生成した個々の研究室から発生する科学情報を連携・統合することで新しい知見を得ることが重要となる．そのためには，生データが格納されているファイルシステムや整理された情報であるデータベースなどの連携，それも貴重なデータ等の資源を安全に保護しつつ情報の連携を図り，それらの情報を研究コミュニティで共有できることが重要となる．データ共有とともに，仮想組織の柔軟な構成と仮想組織への研究室レベルの計算資源，あるいは他のグリッド環境などとの資源連携機能を実現し，従来のスパコンユーザだけではなく，広く一般研究者を取り込むような新しいグリッド環境での運用が必要となってきている．

■次世代 IT 基盤の構築

我が国の次世代 IT 基盤としてペタフロップス級の大規模データを扱う次世代スパコン（NLS：National Leadership System）を頂点に，7 大学の情報基盤センター，国研のセンター等の大規模データを扱う大学・研究機関のスパコン（NIS：National Infrastructure System），日々の研究業務での小規模データを扱う研究室レベルのクラスタ（LLS：Laboratory Level System）を高速ネットワークで接続し，研究室のサーバから次世代スパコンまで，利用者が計算資源を意識せずにシームレスに利活用できる環境の整備が計画されている．このような階層的な計算資源の整備は，すでに米国，欧州でも始まっており，米国では実行速度（LINPACK）で 1 ペタを超えるマシンが設置されている．このような環境整備が国際競争力を左右する時代となってきており，我が国でも一刻も早い整備が待ち望まれている．

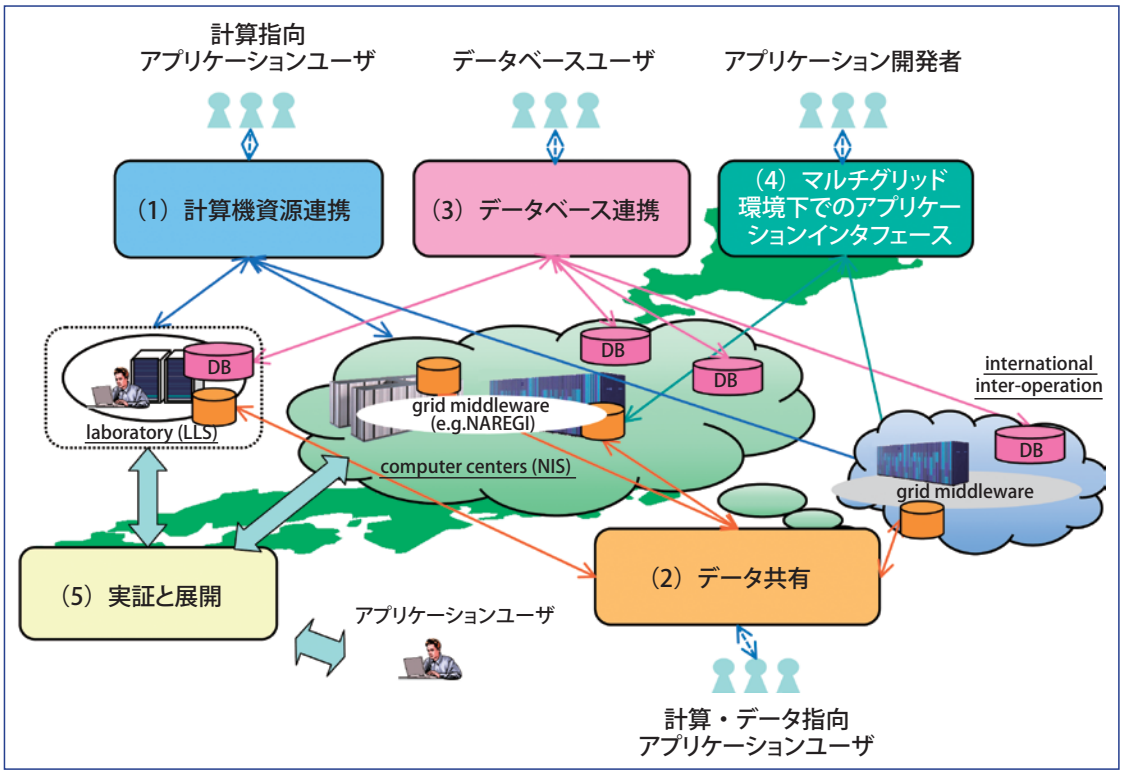


図 -1　RENKEI プロジェクトの全体像

RENKEI プロジェクト

科学技術分野での国際競争力維持・向上のために，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」～研究コミュニティ形成のための資源連携技術に関する研究～，通称 RENKEI（REsources liNKage for E-sclence）プロジェクトが平成 20 年度からスタートしている．RENKEI プロジェクトは，研究室，計算機センター，国際的なグリッド等の複数の組織に分散されたさまざまな資源（計算機，ストレージ，データベース，アプリケーション）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．具体的には，新たに LLS に対応する軽量級のグリッドミドルウェア（Light Weight Grid Middleware）を展開し，NAREGI との一体的な運営により，利用者から見たら NIS，LLS の区別を意識せずにシームレスにジョブ実行ができるとともに，実験データ等を LLS 間，あるいは LLS と NIS 間で共有できる仕組みとするものである．また，データベース連携機能の開発によりデータのハンドリングを強化するとともに，スクリプト言語からの複数グリッド環境へのジョブ投入監視など操作性の向上を図る予定である（**図 -1** 参照）．

計算機資源連携技術

e- サイエンス計算連携技術は，研究室と情報基盤センターの間，あるいは他のグリッドシステムとの間でのシームレスなジョブ実行環境を実現するもので，以下の 3 テーマに絞った研究開発を実施している（**図 -2** 参照）．

1）異種グリッド環境間に跨るワークフローシステム
2）異種グリッド環境間でのアプリケーション共有
3）異種グリッド環境間の相互運用

■ ワークフローシステム

ワークフローシステム [1] は，NIS 対応アイコン，LLS 対応アイコンを定義し，利用者はそれぞれのアイコンを選択することで，ジョブ投入先の環境を指定することができる．ワークフロー内部に解釈エンジンを搭載することで，ワークフローが NIS あるいは LLS にジョブを投入するのかを判断し，指定されたそれぞれの環境にジョブを投入する．したがって，利用者は LLS や NIS の運用形態の違いを一々意識せずにジョブ投入，実行状態のモニタリングが可能となる．また，LLS の計算機の運用のためのミドルウェアとして，負荷が軽く手軽にインストールできる UK e-Science 開発の GridSAM や，Globus Tool Kit のジョブマネージャである WS-GRAM を用いることができる．NIS の計算機の運用のためのミドルウェアとして NAREGI を用いることで，NIS と LLS の両システムがユーザにとってシームレスに見えるワークフローシス

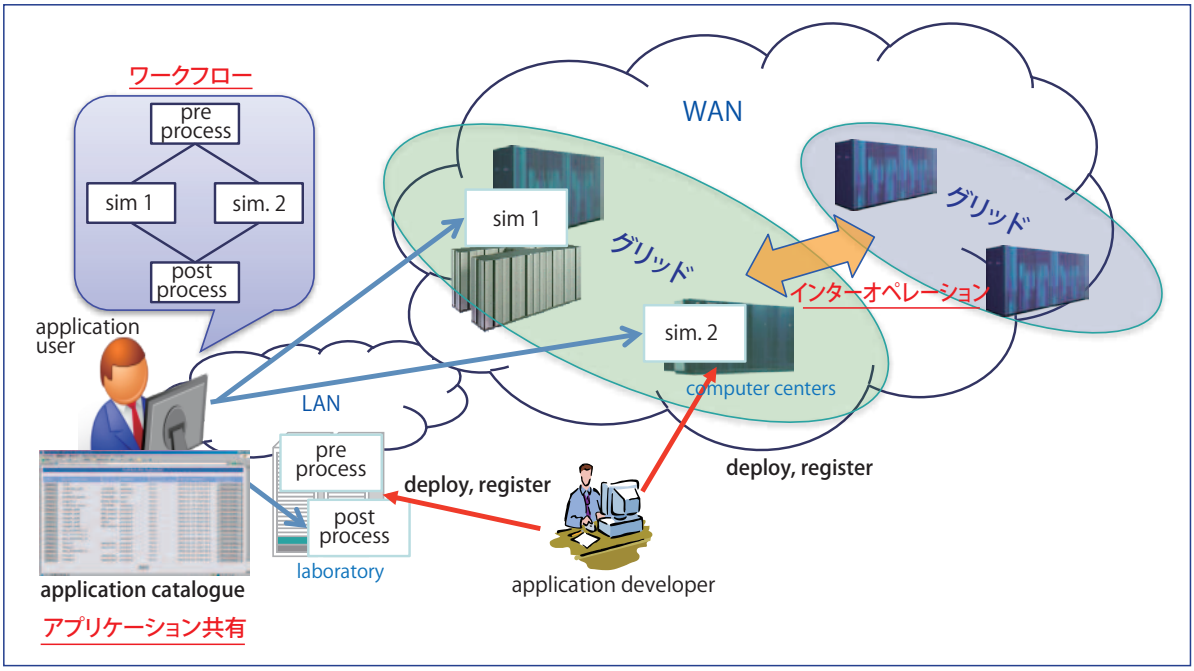


図 -2　計算機資源連携の概要

テムを実現した．アプリケーションレベルのワークフローの記述のために NAREGI ミドルウェアで導入されている NAREGI-WFML（NAREGI 独自仕様のワークフロー言語）をベースに，ジョブサービスサーブレットとして，各ミドルウェア対応のジョブの実行，モニタリングが可能な方式を実現している．

■アプリケーション共有環境

複数グリッド環境でアプリケーションを共有するために NIS および LLS のそれぞれのグリッド環境で，NAREGI-PSE（NAREGI で開発されたコミュニティのための問題解決環境）と同様なサービスを提供するアプリケーションホスティングサービス（AHS：Application Hosting Service）[2] のプロトタイプを開発した．AHS の開発にあたっては，NAREGI の設計コンセプトに合わせ，LLS 環境と NAREGI 環境とのシームレスな連携を実現した．特に，シームレスな連携のために NAREGI の仮想組織（VO：Virtual Organization），およびグループ間でのアプリケーション共有の枠組みを LLS レベルにも展開した．NIS および LLS を含んだ仮想組織，および仮想組織内でのアプリケーションの実行環境への配置（Deployment），およびアプリケーション共有が実現できている．アプリケーション共有シナリオは，アプリ開発者が研究コミュニティで利用可能な NIS，LLS 資源にアプリケーションを配置しておく．アプリ利用者は，たとえば，小規模解析なら LLS で，大規模解析は NIS で，とのように適時使い分けて同じアプリケーションを同じ操作環境で実行することができる．解析のための実験データ等は，データ共有機構によりコミュニティの中で仮想的に統合管理されているので，特に意識する必要がなく計算処理実験ができる．LLS レベルで実行，検証したのち，入力データ等の環境を特に変更することなく，NLS レベルの計算環境で大規模な解析が実行できる．将来的には，NLS の超大規模計算も同様に実行先を変更するだけでシームレスな実行ができるようになる予定である．

■グリッド間相互運用

運用形態の異なる NIS 間，すなわち使用するグリッドミドルウェアが異なる NIS 間でのシームレスなジョブの連携実行を実現する．そのためには，ユーザが普段使用している NIS のグリッド環境から他の NIS のグリッド環境にジョブを投入，管理する機能が必要となる．さまざまなグリッドミドルの NIS 間でジョブ連携をとるには，国際的なジョブ投入・管理のための標準インタフェース仕様である OGF（Open Grid Forum）の BES（Basic Execution Service）が重要となる．しかしながら，現時点では全体の仕様が未確定なので，現在確定している範囲でのジョブ連携機能についてプロトタイプを開発した．すべてのグリッドミドルウェアが BES をサポートすれば，世界中の計算資源を連携したジョブ実行が可能となる．また，相互運用のためには計算資源等の情報の共有が重要となる．NAREGI の情報サービスはグリッドの国際標準である CIM（Common Information Model）に準拠しているが，LLS や NIS が GLUE2（Grid Laboratory for a Unified Environment）スキーマを採用するグリッド環境を含む場合にも，それらの運用形態の異なる計算資

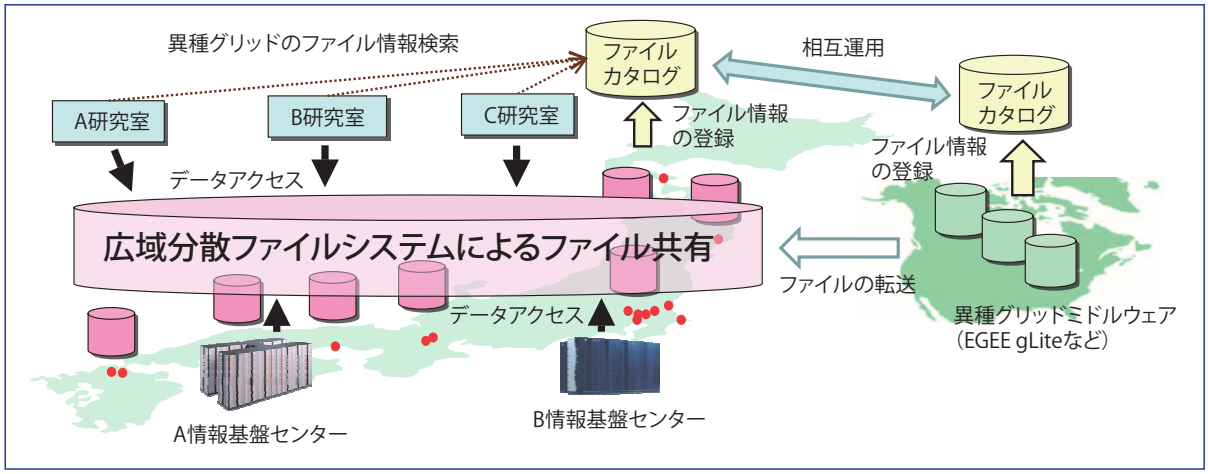


図-3　データ共有技術の概要

源間でジョブをシームレスに実行する必要がある．そのために必要な資源情報の収集・管理が可能となる分散情報サービスの機能を開発している．

データ共有技術

高エネルギー物理学における巨大加速器実験のデータ解析，天文学における仮想天文台の構築，バイオ情報分野におけるゲノムデータ解析など，e-サイエンス分野での大量データの共有の必要性が高まっている．このような背景から，データ共有の研究では，LLS から NIS，さらには海外の研究機関が持つデータを必要性に応じて柔軟に共有または連携するための研究開発として，広域分散ファイルシステムとファイルカタログシステムの2つのアプローチから研究が進められている（**図-3** 参照）．

■広域分散ファイルシステム

広域分散ファイルシステムでは，LLS や NIS 間での効率的なファイル共有と分散並列処理を実現するための技術が開発されている．RENKEI プロジェクトでは Gfarm をベースにして，ファイルの複製の分散配置による効率的なアクセス技術や耐故障性技術を開発し，最適な複製の自動配置を実現することで広域に分散したどの場所からでも手元のファイルサーバと同等のアクセス性能でのファイル共有の実現を目指している．

効率的なデータアクセスのためには，必要な数のファイルの複製を分散配置することが有効であり，ファイル複製の管理をファイルシステムのレベルで行うことが求められている．これにより，アクセス性能の向上だけではなく，耐故障性が向上する．すでに広域分散ファイルシステムの基本的な機能を実現する技術として，多数のクライアント（300 名程度）からの接続，遠隔からの多数のファイル情報取得の高速化，シンボリックリンク機能などを実現し，ファイルサーバとしては国内 13 拠点の 239 ノードからなる大規模環境でのテストを実施済みである．現在，耐故障性技術を開発中であり，今後，ファイル複製の自動配置に関する技術の開発を行う予定である．

■ファイルカタログシステム

データ共有の対象を海外の研究機関が持つデータにまで拡大すると，ネットワークの通信遅延の増加などから広域分散ファイルシステムよりも疎なデータ共有方式が必要となる．このような方式の例としては，ファイルカタログによるデータ共有がある．ファイルカタログは，ファイルシステムとは異なりファイルの実体は持たずに，ファイルの場所を指す URL 等のリファレンス情報のみを階層的なディレクトリの下で管理する．実際のファイルの実体は各研究機関のファイルサーバ上にあるため，実体が必要な場合は，そのリファレンス情報をもとにファイル転送ツールで手元に転送してアクセスする．たとえば，EGEE や TeraGrid では，それぞれ LFC や MCAT 等のファイルカタログシステムが利用されているが，これらのシステムは，それぞれのグリッドミドルウェアへの依存性が高く，ミドルウェア間での相互運用が困難という問題があった．

そこで RENKEI プロジェクトでは，OGF の標準である RNS（Resource Namespace Service）を基にしてファイルカタログを構築している．RNS は，リファレンス情報を管理する Web サービスのインタフェースを規定しており，仕様が公開されているため，誰でもこの仕様に基づいて実装することで相互運用が可能である．また，リファレンス情報をディレクトリごとに別々のサーバに分散配置して，相互に連携する機能も規定されており，大規模ファイルカタログの複数サーバへの負荷分散が可能である．すでにファイルのリファレンス情報の登録・検索

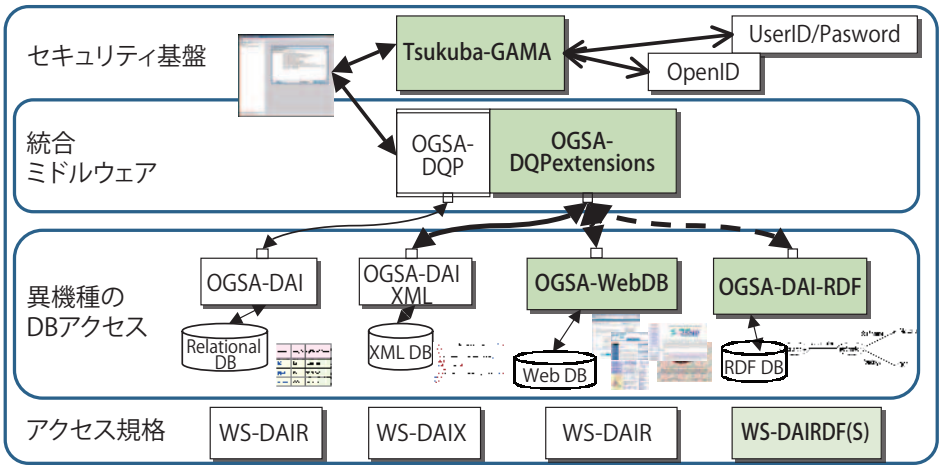


図 -4　データベース連携の概要

などの基本機能や，複数サーバに分散配置されたファイルカタログの連携機能を実装しており，現在，既存ファイルカタログとの相互運用技術を開発中である．今後は，ファイルカタログの分散管理による性能向上技術の実現等を行う予定である．

異種分散データベース連携技術

■ 異種データベース連携の現状と課題

e- サイエンスでは，地理的に分散した組織が独立に生成した科学情報，特に，整理・構造化された情報であるデータベースの連携が知見の創出に有効であり，天文分野の SkyServer やセマンティック Web における Linking Open Data，バイオ情報学や環境など，さまざまなプロジェクトにおいて重要な課題となっている．RENKEI プロジェクトにおいても，地球観測のための GEO Grid を対象に異種データベース連携の実現に取り組んでいる．

異種データベース統合や分散データベース処理といった課題に対しては，従来から多くの研究開発が行われているが，個別の分野向け，商用あるいは特定のシステムやモデルに基づいたものが多く，オープンで互換性・汎用性の高いソフトウェアや，応用システム開発のための実用的なツールに乏しい状況が広く知られている．

また，データベース連携は貴重なデータや異なる組織のユーザを安全に管理する技術と結びつける必要があるが，グリッドの分野で普及している堅牢な認証体系は，ユーザが証明書を管理する手間が多く，また既存の認証体系とも連携できないため他のアプリケーションとの連携が困難といった実用性の課題があると考えられる．

■ RENKEI におけるアプローチ

そこで RENKEI プロジェクトでは，仮想組織を支援する実用性の高いセキュリティ基盤のもとで，関係データベースや XML データベースなど，異種のデータベースを連携できるミドルウェアを研究開発している（**図 -4** 参照）．英国 OMII-UK 開発の OGSA-DAI と呼ばれるソフトウェアを基礎としている点が特徴で，標準化も含めて OMII-UK と密接に協力しつつ研究開発を行っている．

まず，OGSA-DAI を拡張して扱うデータベースの異種性を高め，インターネットに多数存在する Web データベースを扱う拡張（OGSA-WebDB）や，Resource Description Framework {RDF} と呼ばれる，セマンティック Web で一般的なデータ形式への拡張（OGSA-DAI-RDF）[3] を行っている．さらにこれらを統合できるミドルウェア（OGSA-DQPextension）[4] を開発しつつあり，広域に分散した Web データベースや Web サービス，XML データベース，関係データベースを SQL に基づいて宣言的に統合する．ソフトウェアの互換性向上のため，ミドルウェアが支援するアクセス手順を標準化すべく，OGF のデータベースワーキンググループに提案している（WS-DAI 規格）[3]．

認証基盤としては，ユーザ ID ／パスワード認証や OpenID 認証などインターネット等で広く普及している認証とグリッドの認証を結び付け，既存の認証に馴れたユーザが簡便にこの基盤を利用できるシステム Tsukuba-GAMA[5] を開発中である．動的な証明書生成とその一元管理を行うことで，他の認証体系とグリッドの認証体系を動的に連携し，あわせて証明書管理の手間を軽減する．

マルチグリッド環境下でのアプリケーションインタフェース

グリッド技術を利用するアプリケーションの中には，ワークフローツールで処理の過程を記述することは容易ではなく，スクリプト言語により一連の処理を記述する

方が容易な場合もある．また，アプリケーション開発者にとっては，ジョブのワークフローを標準的なスクリプト言語で記述しておいた方が複数のミドルウェアに対応することが容易となる．

ミドルウェア間の相互運用性が確立されれば，アプリケーションの実行に異なるミドルウェア配下の計算資源を利用可能となり，より多くの計算資源を同時に利用できるという意味で大きな意味がある．しかし，たとえばNAREGIのワークフローシステムに対応しているアプリケーションを実行しようとすると，NAREGIの環境なしには利用できないことになる．このような観点から，グリッドアプリケーションをミドルウェア間で可搬にし，ワークフローをスクリプト言語で記述するために必要な下位レイヤーとしてのアプリケーションインタフェースの開発を行うことにした[6]．

OGFでは，SAGA（Simple API for GRID Applications）の標準化が進められており，各ミドルウェアに対する実装が並行して行われている．SAGAは，ローカル，グリッド，クラウドなど対象を問わず，ジョブのバッチ処理とファイルアクセスに必要な機能がAPIとして整備されつつある．すでに，Globus Toolkit，Platform社のLSFジョブスケジューラ等への対応が進められており，我々はNAREGIおよびTorqueジョブスケジューラへの対応を行っている．ヨーロッパで広く利用されているgLite（Lightweight middleware for grid computing）を含むそのほかのミドルウェアに関する対応も別々のグループが分散して開発を行っている．ファイルアクセスに関しては，ファイルカタログとしてRNSを利用する計画で研究を進めている．SAGAには，C++とJAVAの実装が並行して行われているが，我々は，C++による実装のみを行っている．利用者には，利便性を考慮し，Pythonスクリプト言語のライブラリを提供する予定である．C++言語とPythonはBoostを介して，JAVA言語とPythonはJythonを通して，相互にAPIが利用可能であり，Pythonを利用することでSAGAの双方の実装を利用できる利点もある．

我々は，OGFにおけるSAGAの標準化に関して寄与しつつ，プロトタイプの開発を推進している．プロトタイプを試験するために，癌の粒子線治療のシミュレーションを例題として用い，NAREGI環境下でSAGAを介してジョブの実行を行うことに成功している．さらにGlobus Toolkitを採用している米国のTeraGridの環境下においてもジョブの投入実行を行うために必要な研究を行っている．gLiteのSAGA対応も進められており，完成すれば，同様に試験を行う予定である．

本成果とグリッドミドルウェアの相互運用性の研究成果とを合わせることで，ミドルウェアを透過的に利用できるアプリケーションの開発が可能となる．アプリケーション開発者は，シングルコードを保守すればよく，利用者は，アプリケーション利用のために新たなミドルウェアの導入を避けることが可能となることが期待できる．

e-サイエンス，実証と展開

情報基盤センター等が運用するNISに位置づけられる計算資源の相互提供を目的として，グリッドミドルウェアの配備が進みつつある．これまでの共同利用型のサービス提供から，グリッド型サービス提供を支援するための業務フローの検討が進んでいるが，これに伴って，e-サイエンスの課題であるコミュニティ形成を支援するための業務構築が行われなくてはならない．

また，実利用に供している計算資源をグリッド計算資源に振り替えていくためには，利用者サービスの継続性が損なわれることは許されない．既存の共同利用とグリッド利用を併存させる移行手段の提供が不可欠である．それと同時に，LLSとの連携によるNISさらにはNLSへのスケールアウトが求められている．

RENKEIでの研究開発成果物を迅速に配備することによって，これらの課題を解決し得るかを実証的に検証するために，RENKEI-POP（Points of Presence）と命名したサービス・ノードの拠点配備を進めている．特に，コミュニティ形成を支援する認証基盤の普及と，共同利用型計算資源からの漸次的な移行手段の提供，そして，データ中心サービス機能の提示によるグリッド型サービスの啓蒙という目的を掲げている．

個々のRENKEI-POPシステムは，いわゆる「Linux Box」であるが，広帯域広域網に収容するための10GbEインタフェースと，データグリッド・サービスをホスティングするためのストレージを持つ．システムは，ある拠点から他の拠点へ1GB/sのデータ転送を維持できるよう，I/Oボトルネックを排除して注意深くデザインされている．開発されたミドルウェアの性能実証を行うだけでなく，ネットワーク計測技術を導入し，ソフトウェアのボトルネックを検出して開発者にフィードバックできる．また，この計測機能を活かした統計取得によって，開発者や利用者に対してネットワークウェザーマップを提示することも企画している．もう1つのRENKEI-POPの重要な機能は，仮想OSホスティングである．研究開発基盤となっているNAREGIグリッドミドルウェアの標準的な導入が行われた仮想OSを用意し，そのクローニングや世代管理を実現している．安定版が得られた段階でスナップショットを取得し，迅速な利用者サービス導入と提供を行うとともに，開発者へのフィードバックを取り持つ．一方で，開発者はクローニングした環境を引き続

き開発に利用できる．

以上のように，これまでのグリッドプロジェクトでは，研究・開発，配備・運用，利用・評価というエコシステムを立ち上げるに至らなかった反省を踏まえ，既存の計算機資源をグリッド型サービス資源へマイグレートする「グリッド・アプライアンス」としてのプロダクション化を視野に入れた RENKEI-POP を実現することによって，かつてのインターネット普及の成功体験を再現するための戦術的な実証環境の実現を進めている．

参考文献

1）田中他 : 異種ミドルウェア間に跨るワークフロージョブの実行方式と実装 ,Vol.2009-HPC-121, No.20 (Aug. 2009).
2）宇佐見他 : 異種グリッドミドルウェアに跨るアプリケーションホスティングサービス（AHS）の設計と実装，Vol.2009-HPC-121, No.21 (Aug. 2009).
3）Esteban, M. et al. : Accessing RDF(S) Data Resources in Service-based Grid Infrastructures, Journal of Concurrency & Computaion, Vol.21, Issue8, pp.1029-1051 (July 2009).
4）Lynden, S. et al. : Service-based Data Integration using OGSA-DQP and OGSA-WebDB., Proc. of 9th IEEE/ACM GRID 2008 Conference, pp.160-167 (Oct. 2008).
5）Yamamoto, N. et al. : VO-enabled Service Harmonization in the GEO Grid, Proc. of IEEE 4th eScience2008 conference, pp.174-181 (Dec. 2008).
6）Iwai, G. et al. : A Prototyping of Web Interface for Treatment Planning in Radiotherapy in the Multi Grid Infrastructure, The Asian Simulation Conference 2009, 7-9 October 2009, Shiga, Japan.

（平成 21 年 10 月 26 日受付）

宇佐見 仁英（正会員）
usami@nii.ac.jp
1978 年東北大学大学院工学研究科博士課程修了（工博）．同年富士通入社．2003 年国立情報学研究所客員教授，2008 年玉川大学学術研究所教授．

松田 秀雄（正会員）
matsuda@ist.osaka-u.ac.jp
1987 年神戸大学大学院自然科学研究科博士課程修了（学博）．同年同大学工学部助手，1994 年大阪大学基礎工学部助教授，2002 年同大学院情報科学研究科教授，2008 年国立情報学研究所客員教授．

小島 功（正会員）
kojima@ni.aist.go.jp
1984 年京都大学大学院修了 . 同年電総研入所．現在産総研情報技術研究部門サービスウェア研究グループ長．OGF DAIS-WG 共同議長等．

佐々木 節（正会員）
Takashi.Sasaki@kek.jp
1993 年新潟大学大学院自然科学研究科博士課程修了（理博）．同年高エネルギー物理学研究所研究員，1994 年同助手，2003 年同助教授，2007 年高エネルギー加速器研究機構教授．

東田 学
manabu@cmc.osaka-u.ac.jp
1994 年大阪大学大学院基礎工学研究科博士課程修了（工博）．同年大阪大学大型計算機センター助手，2007 年大阪大学サイバーメディアセンター助教．

