

# グリッドが実現するE-サイエンス

## — 地球観測グリッド (GEO Grid) の設計と実装 —

田中 良夫

本稿では、新たな科学技術手法として注目されているE-サイエンスの実装例として、地球観測グリッド (Global Earth Observation Grid、GEO Grid) の情報処理基盤の設計と実装について報告する。GEO Gridは、グリッド技術を用いて複数の組織の有するデータや計算を統合し、仮想的な研究環境を構築して提供する。幅広い応用分野のコミュニティにグリッドを用いたE-サイエンスの実例を提示するとともに、多数のソフトウェアコンポーネントを組み合わせて大規模システムを構築する手法を示す。

キーワード: キーワード: グリッド、GEO Grid、E-サイエンス、ミドルウェア、システム構成論

### How Grid enables E-Science?

– Design and implementation of the GEO Grid –

Yoshio Tanaka

In this paper, we report on the design and implementation of the GEO (Global Earth Observation) Grid IT infrastructure as an example of E-Science. Using Grid technologies, the GEO Grid provides an IT infrastructure that integrates wide varieties of data sets and computing services for the Earth science community. This paper presents an example of Grid-based E-Science and its application to a wide range of scientific communities. A methodology for system development based on many different software components is also discussed.

Keywords: Grid, GEO Grid, E-Science, middleware, software design

#### 1 はじめに

近年、ネットワーク技術の進歩およびネットワーク基盤の普及にともない、高速ネットワークで接続された高性能計算機、データベース、大規模記憶装置、実験装置などの様々な資源を統括的に利用し、科学技術における新発見や融合研究領域などの新たな研究分野の創出を促進する科学技術研究手法であるE-サイエンスに関する研究が盛んに行われている。グリッド<sup>[1]</sup>は、高速ネットワークで接続された様々な資源を安全に、動的に、かつ柔軟に組み合わせて利用する技術であり、E-サイエンスを支える基盤技術として期待されている。グリッドという用語が初めて使われてから約10年が経ち、グリッドの要素技術の研究開発は進み、グリッドは実証実験の段階から実用化に移る時期を迎えている。

一方、近年は地球温暖化などの環境問題、地震や洪水などの災害予測・対策、および天然資源探索などの地球観測に関する問題が重要になっている。これらの分野では、衛星データ、センサーデータや地質図などの、複数の組織が保持する複数のデータを参照し、データ解析やシミュ

レーションを実行する必要がある。データ所有機関のポリシーに従いつつ、複数のデータベースを検索して解析を行う作業は簡単ではなく、それを容易に実現するシステムの開発が期待されていた。これは、まさしくグリッドを用いて実現できるものであり、GEO (Global Earth Observation) Gridの研究開発が進められることとなった。

GEO Grid<sup>[2][4]</sup>は、グリッド技術を用い、地球観測衛星データなどの大規模アーカイブおよびその高度処理を行い、分散環境下の各種観測データや地理情報システムデータと統合した処理・解析を、ユーザが手軽に扱えることを目指したシステムかつコンセプトであり、地球観測に関わる多種多様なデータや計算を研究コミュニティや事業者が安全・安心に利用できる環境の提供を目指している。産総研の有する地質情報と衛星情報との情報融合を進め、さらに広く地球観測情報との融合化を図り、また、国際連携を積極的に推進し、特にアジアにおける高度利用を重点的に展開する計画である。この際に国際的な標準動向に配慮し、情報システムとデータの国際的な相互利用性を確保することを目指している。

産業技術総合研究所 情報技術研究部門 〒305-8568 つくば市梅園 1-1-1 中央第2  
Information Technology Research Institute, AIST Tsukuba Central 2, Umezono 1-1-1, Tsukuba 305-8568, Japan E-mail: yoshio.tanaka@aist.go.jp

Received original manuscript October 16,2008, Revisions received December 21,2008, Accepted December 21,2008

GEO Grid はE-サイエンスの1例であるが、他にもタンパク質データベースなどバイオ情報データベースを用いて創薬を効率よく行うといったバイオ情報分野、大量の医療画像データベースを活用して癌診断を支援するシステムや医療データベースを利用して治療に役立てる次世代医療診断システムなどを開発する医療産業分野など、E-サイエンスにより技術開発を大幅に加速することができると期待される分野は多岐にわたる。

グリッドの要素技術の多くは実用に耐えるレベルに達しているが、実際にはいまだ広く利用される技術となっていないのは、以下の理由によると考える。

- ・もともと高性能計算の分野から派生したグリッドは、「スパコンを束ねて大規模計算を行う技術」というイメージが強いが、この計算グリッドにおいては、インターネットの性能がベストエフォート型であり性能が保証されないといった性能的な課題や、既存の並列プログラミング手法が耐障害性や計算機の同時確保などの問題によりグリッドには適していないこと、また、最適な計算資源を選び出すスケジューリングの技術がまだ研究段階の域を脱していないといった技術的課題など、解決すべき課題がある。
- ・「グリッドを使ったらこんなことができた」といった実例報告がまだ少なく、応用分野のコミュニティから見ると、依然として「使ってみたいがどう使うのか分からない」「どうせ使えないだろう」「何に使えるのか分からない」といった印象がある。

確かにグリッドの要素技術がすべて実用化のレベルに達しているというわけではないが、これまで開発されてきた技術を組み合わせれば多くの科学技術分野に対して新たな、かつ現実的な研究手法を提供できると考える。本研究の目的は、GEO Grid を題材としてグリッドを用いたE-サイエンス基盤を構築し、地球科学の研究者に研究環境を提供するとともに、他の応用分野も含めた幅広い科学技術分野における真の実用化に向けた課題を明確にし、その解決を図ることにある。これにより、科学技術分野におけるイノベーションの創出に寄与することを目指す。目標の達成を目指し、本研究ではGEO Gridの応用事例のシナリオに基づいて情報基盤に対する要求仕様を解析し、システムの設計と実装を行った。実際に衛星データを配信するシステムを構築し、地球科学の研究者にグリッドを用いた研究環境を提供しつつ、そこから得られる知見やフィードバックを通じて残された課題を明確にし、真の実用化に向けての道筋を立てる戦略をとった。

本稿では、GEO Grid を例にE-サイエンスにおけるセキュリティやシステム構築の実現方法および残された課題を明

確に述べる。本稿の主たる目的は、応用分野の読者に対してグリッドの実例を提示し、「何ができて何ができていないのか」を明確にすることにより、グリッドの理解を深め、グリッドの普及を図ることにある。また、情報技術分野の研究者を対象に、多数のソフトウェアコンポーネントを組み合わせる1つのシステムを構築する方法論について論じる。

GEO Grid はアプリケーション、コンテンツ、および情報基盤により構成されるが、本稿においては情報基盤の設計と実装について報告する。はじめに情報技術分野におけるシステム構築の方法論について論じる。また、GEO Grid 情報基盤に対する要求およびそれに基づく設計方針を示し、実装方法および実システムの構築を通じて得られた知見、検証結果を述べる。

## 2 情報基盤に対する要件

GEO Grid 情報基盤に対する要件を以下にまとめる。

### (1) 大規模データの提供

衛星観測データはその運用期間を通じると数百テラバイトからペタバイト級の大きさになるが、そのような大規模データからユーザが求めるデータを迅速に見つけ出す、データサイズに対する高いスケーラビリティが要求される。

### (2) 多様なデータの取扱い

衛星観測データ以外にも、気候に関する温度、湿度、雲量などの異なる物理量や異なる時空間分解能から得られたデータなど、異なる組織により提供され、様々な書式で蓄積されている多様なデータを取り扱う機能が要求される。

### (3) データ提供ポリシーの尊重

データには利用に制限がかからないフリーなものも存在するが、一般的にはデータ所有者はデータアクセスの許可範囲や提供可能なデータ書式の制限など、利用許諾権とその条件を設定・変更する権限を有し、データ提供者の公開ポリシーに応じた柔軟なアクセス制御を実現する必要がある。

### (4) データと計算の統合

データの形式変更や事前処理などの簡便な計算およびデータに基づいた火砕流到達範囲の計算などの大規模シミュレーションなどの計算とデータとの統合が必要である。

### (5) 多様なコミュニティの支援

環境監視、災害監視、資源探査をはじめ、地球科学に関する多様なコミュニティや多数のプロジェクト研究を支援し、柔軟な構成変更および共通に利用できるデータ、計算およびツールやテンプレートになった処理フ

ローなどを共有する仕組みが要求される。

(6) 簡便性

ユーザ、データ提供者、プロジェクト管理者など、すべての参加者に対して「簡単に使える」ツール、インタフェースを提供する必要がある。また、数万規模を超えるユーザに対応しても、ユーザ管理を容易に行うことができるシステムである必要がある。

3 設計

前節において述べた要件に基づき、具体的な要素技術の選定および実装を行う前に、基本的な設計方針および利用モデルを定めた。本節では GEO Grid 情報基盤の設計方針および利用モデルについて述べる。

3.1 設計方針

要件 (2) にあるような多様なデータや計算を、要件 (3)、(4)、(5) にあるように、その提供ポリシーを尊重しつつ共有・統合し、研究コミュニティに対して提供するために、提供されるデータや計算を標準的なプロトコルやインタフェースを通じて提供される「サービス」として抽象化し、それらのサービスを動的に組み合わせることで利用環境を構築する、仮想組織 (Virtual Organization, VO) <sup>[5]</sup> の概念を GEO Grid 情報基盤の設計に導入する。世の中の多様なデータサービスや計算サービスの中で研究コミュニティが必要とするサービスが組み合わせられ、仮想的な 1 つの情報基盤として構成される研究環境が VO である。

3.2 利用モデル

図 1 に示す通り、GEO Grid 情報基盤においては、サー

ビス提供者、VO 管理者、エンドユーザ、および GEO Grid 管理者の 4 つの役割が存在する。サービス提供者はデータや計算の所有者であり、エンドユーザに対してそれらをサービスとして提供する。VO 管理者はコミュニティやプロジェクトの管理者と考えれば良く、VO の構築、VO に所属するユーザの管理、ユーザ向けポータル構築を行なう。GEO Grid 管理者は利用可能なサービスが登録されているレジストリの管理およびそれへのアクセス制御を行なう。エンドユーザは基本的には 1 つ以上の VO に所属し、実際にサービスを利用することにより研究・調査を行う。

サービス提供者は提供するサービスの情報を GEO Grid 管理者が運用するレジストリに登録する。VO 管理者はどのようなデータ、計算が利用可能であるかレジストリを通して検索し、利用を希望するサービスがあればサービス提供者と個別に交渉する。サービス提供者は VO へのサービスの提供を許可する場合はその VO に対するアクセスを許可するべくシステムの設定を変更する。前述のようにアクセス制御はサービス提供者のポリシーに応じて VO 単位、ユーザ単位、あるいはフリーアクセスなどに設定可能である。

4 要素技術の選定と実装

前節で述べた設計方針に基づき、我々はグリッド技術を用いて GEO Grid 情報基盤を実装する。複数の組織間でサービスを連携させることを考えれば、標準に従ったセキュリティの実現が必須である。また、実装コストの削減および他システムとの相互利用性を高めるため、既存のツール、ソフトウェアを有効に利用することも重要である。本節では

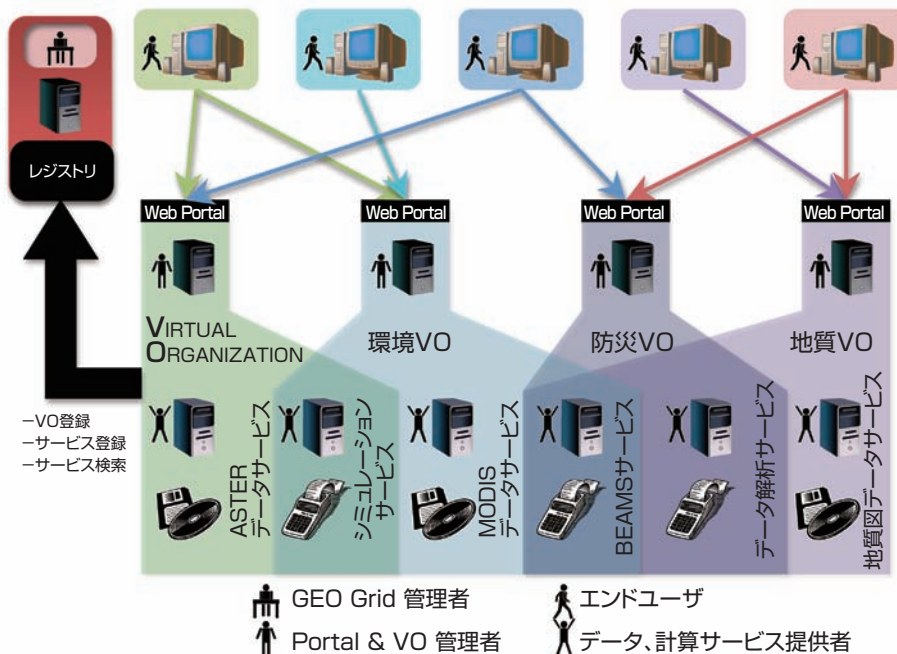


図 1 GEO Grid 情報基盤の利用モデル

GEO Grid 情報基盤の実装に際し、要素技術の選定およびその結合方法について述べる。

#### 4.1 セキュリティ

GEO Grid 情報基盤のセキュリティは、Grid Security Infrastructure (GSI)<sup>[6]</sup> および VO レベルの認可機構を基本とする。GSI は公開鍵暗号基盤 (Public Key Infrastructure, PKI) と X.509 証明書<sup>[7]</sup> を用いたグリッドにおける標準的な認証基盤であり、プロキシ証明書を用いてシングルサインオンと権限委譲を実現する。GSI はグリッドセキュリティの標準技術となっており、他のシステムとの互換性や多くのグリッドミドルウェアが GSI に対応していることを考慮し、セキュリティには GSI を採用することとした。

仮想組織の構成およびそれに基づくアクセス制御については、Virtual Organization Membership Service (VOMS)<sup>[8]</sup> を用いる。VOMS は欧州のプロジェクト Enabling Grid for E-Science in Europe (EGEE) において開発されたソフトウェアであり、VO に所属するメンバーを管理し、メンバーの登録、グループの作成、ユーザに対する役割の付加などを行なう。また、ユーザからの要求に応じて、ユーザのプロキシ証明書にユーザの VO における属性情報 (所属する VO 名、グループ名、与えられている役割など) を埋め込んだ VOMS プロキシ証明書を発行する。サービス提供者側では、ポリシーに応じた様々なアクセス制御が実現できる。

認証を受けたユーザは、通常サービス提供者側で UNIX アカウントにマップされ、UNIX アカウントの権限でアクセス制御が実現される。しかし、この方法ではすべてのユーザのエントリをサービス提供者側で管理しなければならず、サービス提供者に対する管理コストが高くなってしまふことや、ユーザ数に対してもスケラブルでないといった問題がある。そのため、VO レベルの認可機構を導入し、VO 単位での認可、VO においてユーザが所属するグループや与えられた権限に応じた認可などにより、サービス提供者の負担を軽減し、ユーザ数に対してスケラブルかつ柔軟なアクセス制御を実現する。

VO 単位のアクセス制御を実現するミドルウェアとしては、VOMS の他に PERMIS<sup>[9]</sup> や CAS<sup>[10]</sup> などが存在するが、プロキシ証明書に属性情報を埋め込む実装が 4.5 節で述べるアカウント管理システムとの親和性が良いこと、ユーザ管理を行うインタフェース等のツールが多々提供されていること、他のシステムに比べて広く普及しており、ソフトウェアの品質が高いことが期待されるなどの理由により VOMS を採用した。

#### 4.2 データ・計算資源のサービス化

データや計算を標準的なプロトコルを通じて利用可能な

サービスとして抽象化し、提供するために、データや計算をラップし、サービスとして提供するミドルウェアを利用する。データのサービス化については、英国 UK-eScience およびその後継プロジェクトである Open middleware Initiative-UK で開発されている OGSA-DAI (Open Grid Service Architecture - Data Access Integration)<sup>[11]</sup> を用い、計算のサービス化については米国 Globus Alliance により開発されている Globus Toolkit<sup>[12]</sup> の Grid Resource Allocation Manager (GRAM) を用いる。これらはいずれも GSI および VOMS を用いた認証認可に対応している。計算をサービスとして構築する方法としては、Apache Axis 上での Java Service として実装する方法などもあるが、GSI との親和性の良さなどを考慮し、GRAM を用いて計算サービスを提供している。

OGSA-DAI、Globus Toolkit のいずれも GSI に対応したグリッドのミドルウェアとして広く普及しており、他に選択肢はないと判断できる。

衛星データや地図情報の検索結果については、Web Map Service (WMS)、Web Feature Service (WFS)、および Web Coverage Service (WCS) など、Open Geospatial Consortium (OGC)<sup>[13]</sup> によって規定されている Web サービスによって提供されるのが一般的であるが、Apache についても VOMS を用いたアクセス制御を行うソフトウェア<sup>[14]</sup> が提供されており、GEO Grid のセキュリティに対応可能である。

#### 4.3 非均質データベース連携技術

OGSA-DAI を用いてデータベースをサービスとして抽象化し、適切な認証・認可に基づいて提供することが可能となるが、それだけでは非均質な複数のデータベースを統合することはできない。ユーザが必要とする機能は「複数の非均質データベースに対して一括問い合わせや分散結合を行うこと」であり、そのためのミドルウェアとして産総研で開発した Extended OGSA-DAI-DQP (Distributed Query Processing)<sup>[15][16]</sup> を用いる。

#### 4.4 大規模ストレージシステム

数百テラバイトからペタバイト級の大規模データを格納するストレージシステムについて検討する必要がある。既存のシステムでは衛星データはテープに格納されているものがほとんどであるが、データ検索の実時間性や近年のハードディスクの価格低下を考慮すると、テープデバイスの利用や商用 Storage Area Network (SAN) の利用は適切ではなく、数テラバイト程度のハードディスクを搭載したノードをネットワークで接続し、大規模ストレージを実現するクラスタファイルシステムを利用することとした。クラスタファイルシステムは、分散した複数のディスクを仮想的なファイルシステムとして提供する技術であり、商用のものやフリーソ

ソフトウェアなどいくつか存在するが、並列 IO による高いスループットおよび柔軟な複製配置による高信頼性機能を実現する、産総研で開発を行った Grid Data Farm (Gfarm)<sup>[17]</sup>を採用した。

#### 4.5 アカウント管理

GSIはPKIに基づく認証技術であり、ユーザは秘密鍵とユーザ証明書の管理を行う必要が生じる。しかし、証明書を取得するための特別なソフトウェアのインストールや、秘密鍵の管理を適切に行うことはユーザにとって負担であり、より簡便なインターフェースを提供することが求められていた。そこで我々は、サーバ側でユーザのアカウントおよび証明書を管理する仕組みを San Diego Supercomputer Center で開発された GAMA (Grid Account Management Architecture)<sup>[18]</sup>を用いて実現した。GAMAはユーザに対してアカウントの作成依頼やログインなどの機能を、アカウント管理者に対してユーザ管理のための機能を、ポータルとして提供するソフトウェアである。ユーザのアカウントはGAMAサーバによって管理され、GAMAサーバはユーザに証明書を発行するための認証局の機能も備えている。GAMAを用いることにより、ユーザは自分で秘密鍵や証明書を取得・管理することなく、ユーザ名とパスワードによる認証で GEO Grid 情報基盤にアクセスすることができる。

また、ユーザに対するポータルとして GridSphere<sup>[19]</sup>を用いる。GridSphereはポータルアプリケーションで利用される「ポータル」と呼ばれる小型の Web コンポーネントを作成するための APIとして Java Community Process で標準化されている JSR168<sup>[20]</sup>に基づいてポータルを構築

するためのフレームワークであり、GAMAサーバからプロキシ証明書を作成するための認証モジュールと、ポータル管理者用のポータルが提供されている。オリジナルのGAMA認証モジュールはGAMAサーバからプロキシ証明書を取得するのみであり、VOMSとのインターフェースは提供されていないため、我々はGAMAサーバからプロキシ証明書を取得した後に、VOMSサーバに問い合わせしてVOMSプロキシ証明書を作成するようにGAMA認証モジュールを修正した。

#### 4.6 要素技術の統合

本節で述べた各要素技術はいずれもGSIに基づくセキュリティに対応しており、アカウント管理システムにおいてVOMSに対するインターフェースを実装した以外、各ミドルウェアが提供するインターフェースを通じた統合が可能であった。

GEO Gridのように大規模なシステムの構築に際しては、すべてを自力で開発することは現実的ではなく、コアコンピタンスをしっかりとおさえつつ、利用可能な技術は積極的に利用して開発コストを軽減することが重要である。

### 5 実システムの構築

我々は提案アーキテクチャに基づき、衛星データの1つであるASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer)<sup>[21]</sup>データを主要コンテンツとしたシステムを実装した(図2)。本システムは、ASTERデータ(Level 0データ)を保存、提供するGEO Grid クラスタおよびGRAMやGridFTPサーバを介してGEO Grid クラスタへのアクセスを提供するgatewayサーバ、ASTERデータのメタデータやカタログを提供す

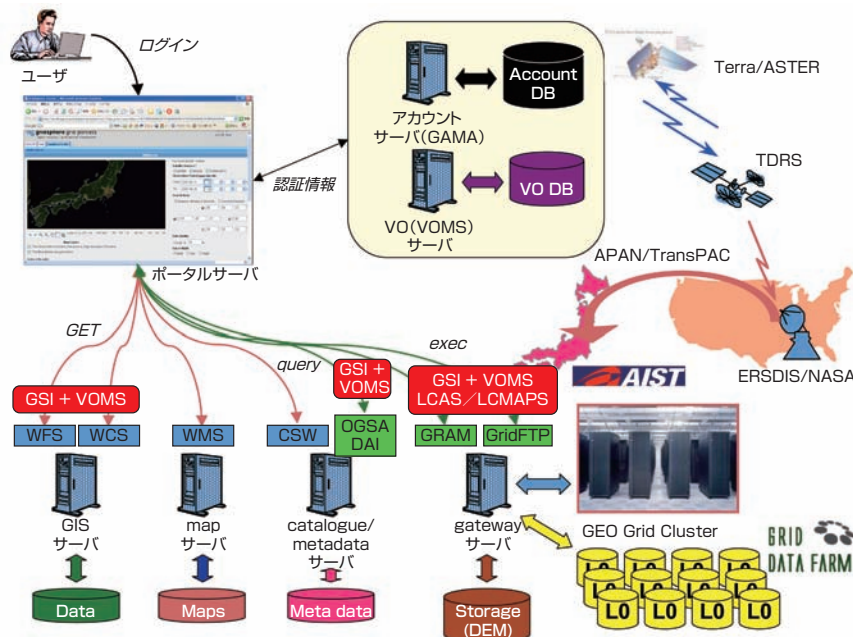


図2 GEO Grid システムの構成

るサーバ、画像データを WMS として提供する Map サーバ、WFS、WCS などの高次データを提供する GIS サーバ、アカウント管理を行う GAMA サーバおよび VOMS サーバにより構成される。本システムにおいては、現在、環境 VO、防災 VO、情報 VO の 3 つの VO が存在し、それぞれにおいて実際のユーザが利用する実運用を開始している。ASTER は NASA が打ち上げている Terra という衛星に搭載されたセンサーであるが、実際には 2 台のセンサーが備えられており、それらの観測結果をもとに地上の標高モデルを求めることができる。ASTER データは Terra の打ち上げ後 ERSDAC (Earth Remote Sensing Data Analysis Center, (財) 資源・環境観測解析センター) が管理するテープライブラリに保管され、有償データとしてユーザに提供されてきたが、昨年来産総研にもデータが提供され、産総研においては、ASTER データはテープライブラリではなくクラスタファイルシステムに保存されている。利用するクラスタ (GEO Grid クラスタ) は Giga-bit Ethernet で接続される 36 台の dual Xeon ノードにより構成され、全体で 264 TB の容量を備える。クラスタファイルシステムとしては Gfarm v1.4 を用いている。現時点で約 140 TB の全 ASTER データが格納され、日々 NASA から約 70 ~ 100 GB のデータが転送され、GEO Grid クラスタに保存されている。メタデータの管理には PostgreSQL に対して GIS の拡張を行なっている PostGIS を使い、メタデータは OGSA-DAI によりデータサービスとして提供される。また、データ処理には Giga-bit Ethernet で結合された 256 ノードの dual Xeon により構成される F32 クラスタを用いている。GEO Grid クラスタと F32 クラスタは 10 Giga-bit Ethernet で接続されている。

また、産総研の ASTER データおよび MODIS (Moderate Resolution Imaging Spectroradiometer)<sup>[22]</sup> データと、台湾の NSPO (National Space Organization) が保持する Formosat2 のデータを同時に検索する Database 連携アプリケーション「DB 連携アプリケーション」を構築して実際に複数の組織に管理される複数のデータベースを統合する GEO Grid の実証実験を行った。「データの検索」「数値地形モデルへの変換」および「結果の転送」のそれぞれが個別のサービスとして実装され、それらが GSI および VOMS による認証認可をベースに連携することにより上位のサービスとして提供され、統合されて VO を構成し、アプリケーションコミュニティに対して研究環境が提供されることを確認した。

また、この他にも ASTER データとフリーなセンサーデータを連携させるアプリケーションや、検索結果を WFS で返すアプリケーションなどを通じ、提案したセキュリティアーキテクチャがアプリケーションおよびサービス提供者の要求

に応じた適切なアクセス制御を実現することが確認された。

## 6 考察

ここでは、今回のシステム構築および実運用を通じて得られた知見、考察を記す。

各ソフトウェアコンポーネントの予備評価および ASTER Grid システムの開発・テストを通じ、我々は GEO Grid 情報基盤の設計および実装の妥当性を検証することができた。GAMA と VOMS を用いたセキュリティフレームワークにより、ユーザに対しては簡便なインタフェースを提供し、サービス提供者に対してはポリシーに応じた柔軟なアクセス制御が実現可能であることが確認された。また、VOMS を用いた VO レベルでのアクセス制御により、ユーザ数に対してスケーラブルなセキュリティ機構が実現される。GAMA、VOMS、OGSA-DAI、Globus Toolkit など、既存のソフトウェアおよびツール群を利用することにより、少ない開発コストで実システムを構築することができた。アプリケーション用ポートレット以外に、情報基盤側で行なった開発は VOMS と GAMA のインタフェースを GridSphere に組み込む部分のみであった。すべての要素技術が標準プロトコルおよび標準インタフェースに基づいて設計・実装されることにより、多数の独立な要素技術を連携させて上位のシステムを容易に構築することができた。

グリッドの実システムの構築においては、欧州の EGEE で開発されている gLite Grid Middleware<sup>[23]</sup> を用いた高エネルギー物理学における大規模加速器実験のデータ解析システムや、日本の超高速コンピュータ網形成プロジェクトで開発された NAREGI ミドルウェア<sup>[24]</sup> を用いて大学・研究機関上に大規模研究グリッドを構築する Cyber Science Infrastructure 計画など、巨大なミドルウェアパッケージを開発・利用するものが多い。米国 Earth System Grid<sup>[25]</sup> や GEON<sup>[26]</sup> などは GEO Grid と同様に地球観測データを統合した研究環境の構築を目指している。いずれも一部でグリッドセキュリティに基づく認証・認可を利用しているが、ほとんどは Web サービスに基づく非グリッド技術により構成される。GEO Grid のように様々なグリッドミドルウェアを用いてサービスとして提供されるデータや計算を組み合わせて VO を構成し、グリッドセキュリティに基づく柔軟なアクセス制御に基づく研究環境を構築する例は独自性が高く、本稿で述べたようにすべてが標準的なセキュリティおよびプロトコルに従って実装されていればグリッドミドルウェアを連携させ、容易に大規模な実システムを構築できることを示した本研究の意義は大きい。

今回のシステム構築を通じ、今回利用した要素技術については問題なく利用できることが確認された。今後真の実

用化に向けて解決すべき課題としては、以下の5つがあげられる。

#### (1) ツールキットの作成

今回利用したグリッドミドルウェアの多くは、インストールおよび設定が煩雑であり、誰でも容易にインストール、利用できるというレベルにない。ユーザに対しては、ユーザ名とパスワードのみで利用可能な簡便なインタフェースを提供しているが、今後様々な応用分野への展開を図るには、サービス提供者やVO管理者などすべての参加者に対して、必要なミドルウェアセットを容易にインストール、設計できるツールキットとして提供する必要がある。

#### (2) より柔軟な認証機能の実現

既存の応用コミュニティには、例えば一部のバイオ情報分野におけるOpenIDの利用など、すでに独自の認証機能を採用しているものがある。既存の研究環境からE-サイエンス環境へのシームレスな移行を実現するために、OpenIDやShibboleth、Kerberos認証などの、すでに利用されている認証機構からグリッドの認証情報を生成する、より柔軟な認証機能を実現すること必要である。

#### (3) ワークフローの構築

応用研究者の多くは、数多くのデータに対してある決まった手続き（処理の流れ）を適用する。地震発生時の地震振動解析や液化化予測、水位が上昇したときの洪水予測など、迅速に必要なデータを取得し、多数のシミュレーションを実行するには、それらの手続きをワークフローとして構築し、利用できることが望ましい。グリッドにおいてはワークフローの先行研究も多々行われているが、GEO Gridでもワークフローの導入、構築が必要である。

#### (4) 高速化

大規模な画像データの処理には大規模な計算資源が必要なものが多いが、既存のソフトウェアでは画像処理に数分から数十分かかってしまう。インタラクティブなデータ配信を考えると、画像生成はせいぜい1~2分程度で終わることが望ましい。近年CELL/B.E.TMのように画像処理に適したマルチコアアーキテクチャが利用されるようになってきているが、GEO Gridにおいても最新のアーキテクチャを活用した画像処理やシミュレーションの高速化が期待される。

#### (5) メタスケジューラの開発

今回構築したシステムでは、シミュレーションや画像処理を行う計算サービスは単一のサイトから提供されているが、今後同一のサービスが複数のサイトから提供

されるようになった場合には、どこにどういったサービスやデータが存在するかを管理するレジストリ、サービスを提供する計算サーバの付加状況を監視するモニタリングシステム、それらの情報をもとに「最も良いと思われる」サービスを選択するメタスケジューラの開発が必要である。

これら5つの課題すべてについて研究を進めているが、(1) ~ (3) については1~2年程度を目途に実現できると考えている。(4) については個別のソフトウェアごとに高速化を行う必要があることと、商用ソフトウェアの場合はライセンスの問題でソースプログラムが提供されない場合が多いといった問題があるが、すでにCELL/B.E.TM上での画像処理ソフトウェアの高速化に関する予備評価を行い、実現可能性の目途は立てている。(5) に関してはグリッドが抱える最大の課題と考えている。グリッドの概念である「コンピュータをネットワークに接続すれば、どこかの資源を利用するかは気にせずにサービスを享受できる」世界を実現するためには、このメタスケジューラの機能が必須となるが、グリッドを構成する資源群（ネットワークや計算サーバなど）の構成や有用性が動的に変化し、「最適」の判断が計算の性質（通信量と計算量の比率など）に依存する複雑な環境上で「最適」な資源を選択することは非常に困難である。我々はGEO Gridのシナリオとしてユーザに影響を与えない範囲で制約を設けることにより、この課題を解決すべく研究を進めている。

## 7 研究開発の進め方

本研究は、融合領域研究として、産総研地質情報研究部門と、地質調査情報センター、環境管理技術研究部門及びグリッド研究センター（現在は情報技術研究部門）との間でニーズとシーズを持ち寄り、システムの実装方法の検討を行った。グリッド研究センターは、地質調査情報センターからグリッド研究センターへの人事異動や、防災科学研究所、宇宙航空研究開発機構、国立環境研究所など応用分野の研究者を採用するなどして、情報分野の研究者と応用分野の研究者が密に議論することによってGEO Gridの研究開発を推進する体制を強化した。

定期的にGEO Grid運営会議を開催し、GEO Gridの方針や対外機関との関係づけ、懸案事項の洗い出しおよび対処などについて議論を行うとともに、進捗管理を行った。また、グリッド研究センターだけでも、応用系と情報系を合わせると総勢20名程度となる大規模プロジェクトを円滑に推進するため、応用系と情報系のそれぞれが定期的に会議を開き、進捗管理や問題点の洗い出しおよび解決に向けた議論を行い、研究チームメンバー間での意識の共有に

努めた。両分野の研究者が同じ場所において、会議の他にも頻繁に議論、ブレイクアウトミーティングができる環境にあったことも重要であった。

また、今回は産総研だけではなく海外機関が開発したミドルウェアを多数利用した。基本的には各ミドルウェアが標準プロトコルやインタフェースを採用して実装されているが、実際にテストをするとミドルウェアの実装不備や機能不足による問題が生じたものもあった。我々は国際共同研究、グリッドの標準化団体である Open Grid Forum<sup>[27]</sup> や国際会議等の場を通じて各ミドルウェアの開発機関と密な協力態勢をとっているため、これらの問題が出た時に迅速に対応してもらうことができた。また、こちらの要求仕様を開発元に伝え、次期バージョンに新機能を組み込んでもらったり、具体的な実装方法について助言をもらうこともできた。グリッドのシステムは非常に多くのミドルウェアを連携させて実現する場合が多く、そのすべてを自分たちで実装するのは開発コスト的にも適切ではなく、自分たちのコアコンピタンスを押しえつつ、海外機関と協力して研究開発を進める体制づくりおよび日頃の活動が重要である。

## 8 まとめと今後の課題

本稿では GEO Grid 情報基盤の設計と実装、および実システムの運用を通じて得られた知見について報告した。E サイエンスにおける VO の考え方やアカウント管理、認証・認可の実現方法を明確に論じた。GEO Grid 情報基盤においては、計算およびデータリソースはすべて標準的なプロトコルにより利用可能なサービスとして提供される。研究コミュニティは VO を構成し、必要なサービスを組み合わせて VO 内のユーザに提供する。VOMS 属性を用いた認可機構により、サービス提供者のポリシーに応じた柔軟なアクセス制御および VO 単位での認可による、ユーザ数に対してスケーラブルなセキュリティ基盤が実現される。

6 章で述べた課題を今後解決すべく研究開発を進めつつ、運用を通じてシステムの評価およびブラッシュアップを進めていく予定である。GEO Grid のシステム開発を開始して約 1 年半が経過し、プロトタイプを超える実用に供することができるシステムを構築することができた。しかし、今後残る課題を解決し、真の実用化に至るにはまだ 2 ～ 3 年の期間を見て研究開発を進めているところである。

## 謝辞

本研究は、GEO Grid プロジェクトにおいて応用分野の研究者と情報分野の研究者がシステム要件や解決方法などについて活発に意見交換を行いながら進められた。GEO Grid プロジェクトに参画する全研究者に感謝の意を表す

る。UK e-Science Centre の Neil Chue Hong 博士および OMII Europe の Valerio Venturi 博士には、OGSA-DAI についてご助言等頂いた。産業技術総合研究所 Steven Lynden 博士には、OGSA-DQP について有益な議論をして頂いた。また、Dutch National Institute for Nuclear and High Energy Physics の David Groep 博士には、VOMS、LCAS/LCMAPS、Gridsite などについて様々なご助言を頂いた。ここに感謝の意を表する。

## 参考文献

- [1] I. Foster and C. Kesselman Ed.: *The Grid: Blueprint for a New Computing Infrastructure*, 2nd edition, Morgan Kaufmann Publishers (2004).
- [2] S.Sekiguchi, Y. Tanaka, I. Kojima, N. Yamamoto, S. Yokoyama, Y. Tanimura, R. Nakamura, K. Iwao and S. Tsuchida: Design principle and IT overview of GEO Grid, *IEEE Systems Journal*, 2 (3), 374-389 (2008).
- [3] 田中良夫, 山本直孝, 関口智嗣: 地球観測グリッドにおけるセキュリティ基盤の設計と実装, *情報処理学会論文誌コンピュータシステム*, 1 (2), 169-179 (2008).
- [4] 田中良夫, 小島功, 山本直孝, 横山昌平, 谷村勇輔, 関口智嗣: GEO Grid: 地球観測グリッドの設計と実装, *情報処理学会HPC研究会研究報告*, 2007 (88), 37-42 (2007).
- [5] I. Foster, C. Kesselman, J. Nick and S. Tuecke: The physiology of the Grid, *Grid Computing: Making the Global Infrastructure a Reality*, Wiley, 217-249 (2003).
- [6] B. Sundaram: Introducing GT4 Security, *IBM developer works* (2005).
- [7] [ITU-T Recommendation X.509]: Information technology - Open systems interconnection - The directory: Authentication framework, 08/05, (2005). <http://www.itu.int/rec/T-REC-X.509-200508-I>
- [8] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell' Agnello, Á. Frohner, K. Lórentey and F. Spataro: From gridmapfile to VOMS: Managing authorization in a Grid environment, *Future Generation Computer Systems*, 21 (4), 549-558 (2005).
- [9] PERMIS Project, <http://sec.cs.kent.ac.uk/permis/>
- [10] Community authorization system, <http://www.globus.org/security/CAS/>
- [11] OGSA-DAI project, <http://www.ogsdai.org.uk/>
- [12] I. Foster and C. Kesselman: Globus: A metacomputing infrastructure toolkit, *Supercomputing Applications and High Performance Computing*, 11 (2), 115-128 (1997).
- [13] Open geospatial consortium (OGC), <http://www.opengeospatial.org/>
- [14] A. McNab: The GridSite Web/Grid security system, *Software: Practice and Experience*, 35(9), 827-834 (2005).
- [15] S. Lynden, A. Mukherjee, A.C. Hume, Alvaro A.A. Fernandes, N.W. Paton, R. Sakellariou and P. Watson: The design and implementation of OGSA-DQP: A service-based distributed query processor, *Future Generation Computing System*, Elsevier Science (to appear).
- [16] S. Mirza and I. Kojima: OGSA-WebDB: Enabling web database access and integration in the Grid, *in Proc. 1st SIIK workshop*, 215-224 (2006).
- [17] O. Tabebe, Y. Morita, S. Matsuoka, N. Soda and S. Sekiguchi: Grid datafarm architecture for petascale data



- intensive computing, in *Proc. International Symposium on Cluster Computing and the Grid*, 102-110 (2002).
- [18] K. Bhatia, S. Chandra, and K. Mueller: GAMA: Grid account management architecture, in *Proc. IEEE Int. Conf. EScience Grid Computing*, 413-420, (2005).
- [19] GeidSphere,  
<http://www.gridisphere.org/>
- [20] JSR168,  
<http://jcp.org/en/jsr/detail?id=168>.
- [21] Y. Yamaguchi, B. A. Kahle, M. Pniel, H. Tsu and T. Kawakami: Overview of advanced spaceborne thermal emission and reflection radiometer (ASTER), *IEEE Trans. Geosci. Remote Sensing*, 36 (4), 1062-1071 (1998).
- [21] C. Justice, E. Vermote, J.R.G. Townshend, R. Defries, D.P. Roy, D.K. Hall, V.V. Salomonson, J. Privette, G. Riggs, A. Strahler, W. Lucht, R. Myneni, Y. Kniazihhin, S. Running, R. Nemani, Z. Wan, A. Huete, W. van Leeuwen and R. Wolfe: The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research, *IEEE Trans. on Geoscience and Remote Sensing*, 36 (4), 1228-1249 (1998).
- [23] F. Gagliardi, B. Jones, F. Grey, M. Begin and M. Heikkurinen: Building an infrastructure for scientific Grid computing: Status and goals of the EGEE project, *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 363 (1833), 1729-1742 (2005).
- [24] National Research Grid Initiative,  
<http://www.naregi.org/>
- [25] Earth System Grid,  
<http://www.earthsystemgrid.org/>
- [26] GEON,  
<http://www.geogrid.org/>
- [27] Open Grid Forum,  
<http://www.ogf.org/>

## 執筆者略歴

田中 良夫 (たなか よしお)

1995年慶應義塾大学後期博士課程単位取得退学。博士(工学)。1996年4月技術研究組合新情報処理開発機構。2000年4月工業技術院電子技術総合研究所。2001年4月改組により産業技術総合研究所。2002年1月同所グリッド研究センター。2008年4月同所情報技術研究部門主幹研究員。グリッドプログラミングおよびグリッドセキュリティに興味を持つ。アジア太平洋グリッドポリシー策定委員会議長、Open Grid Forum Certificate Operations Working Group 共同議長。

## 査読者との議論

### 議論1 論文の読者について

質問・コメント (大蒔 和仁)

誰に(どんな種類の人に)読ませたいのか、論文の目的が多少あいまいな気がします。すでに世の中にはすぐれた機能のソフトがPDS的に落ちていたのでそれを拾ってきて組み合わせるだけでもかなりのことができ、それにコアコンピタンスを示すだけでかなりの高機能が実現できる時代である、といて多少広めのソフトウェア技術者相手に安心感を与えサンプルとしてグリッド技術を示したいのか、あるいは、応用範囲を、今は地質だが、天文や地球環境へもっと広がるはず、と訴えたいのか、ソフトウェアあるいはデータを「サービス」として前面に押し出すべき時代に入っていると主張したいのか、あるいはそれから全部なのか、「はじめに」とか「考察と結論」とかで長い文書を使って目的をもっと明確に述べてはどうか、と思います。

回答 (田中 良夫)

本論文の主題は、GEO Gridを例としてグリッド技術を用いたE-サイエンス基盤の構築例を科学技術の広い分野の研究者に対して示すことにより、グリッド技術およびそれを用いたE-サイエンス基盤の普及を促進し、科学技術におけるイノベーションの創出に寄与することにあります。また、情報技術に従事する研究者に対して、要素技術が標準的なプロトコルやAPIに従って実装されていれば、大規模なシステムを容易に実現できることを実際に示すとともに、各要素技術の開発に際しては標準化や海外機関との連携が重要であること、大規模なシステムの構築に際してはすべてを自力で開発することは現実的ではなく、コアコンピタンスをしっかりとおさえつつ、利用可能な技術は積極的に利用して開発コストを軽減することが重要であることを主張したいと考えています。

これらの主題が明確になるよう、1章に記述を追記しました。また、アブストラクトも修正しました。

### 議論2 「分散計算」について

質問・コメント (大蒔 和仁)

グリッド技術あるいは高速計算機をつないで、という発想は貴重で重要だと思いますが、素人から見てグリッドという概念が提唱され始めた10年ぐらい前と比べて身近になったような気があまりしていません。いやいや、それは査読者の認識不足で、当時とは雲泥の差で発展していてすぐ手が届くところまできている、というような書き方はできるものなのでしょうか。

回答 (田中 良夫)

グリッドの多くの要素技術がマチュアになっている反面、依然として身近になっていない印象を与えるのは、特に計算グリッドにおいていくつか技術的な課題が残っていること、グリッドを用いた実例、サクセスストーリーがあまり報告されていないことに原因があると考えています。しかし、使える技術を使うだけでもかなりのことができるレベルに達しており、実際にGEO Gridは既存の技術を組み合わせるだけで地球科学の研究者のかなりの要求に応えることができます。この点について、1章に記述を追加しました。この記述に続いて、議論1への回答に述べた、本研究の目的が続くようになっています。

### 議論3 セキュリティについて

質問・コメント (大蒔 和仁)

VO的発想で問題となるのは今も昔もセキュリティの問題であると思います。それが唯一残っている問題で、それが解決されれば曲がりなりにも動くのでしょうか。あるいは何か別のファクタが残っているのでしょうか。あるいはデジュール標準を取ると解決されるのでしょうか。特にグリッド技術の「問題点」の言及が欲しいと思います。

回答 (田中 良夫)

セキュリティについては多くの科学技術分野に対しては実用に耐えるレベルの技術が確立されていると考えています。グリッド技術の問題点は、「グリッドミドルウェアの多くが容易にインストール、設定できないこと」、「証明書の取り扱いなど、ユーザにとって使い勝手が良くないこと」、「メタスケジューラというカギとなる技術が依然として研究段階にあること」の3つが要因であると考えています。これらの問題点については、6章で「解決すべき課題」として詳しい説明を追記しました。

### 議論4 戦略やシナリオについて

質問・コメント (小林 直人)

グリッド技術の応用としてのGEO Gridと言うのは、地球温暖化防止、省資源、災害予測・防止、有効土地利用など持続的発展可能な社会をめざす21世紀の多くの課題解決のために、非常に有効な技術だと思います。

それを踏まえた上で、シンセシオロジーの論文では、研究目標と社会とのつながりやシナリオの重要性を強調しています。特に本論文では大きな目標概念は分かりますが具体的な研究目標が明確ではないように思います。最終的には、グリッドはグリッドと言う技術を意識せずに、多数の場所に分散している CPU やデータベース等をあたかも自分のコンピュータの内部やごく周辺にあるものとして、それを自在に駆使して情報処理を行うことにその真髄の1つがある、と理解しています。そうであるとする本研究の GEO Grid では最終目標がどこにあり、本論文ではその中でその大きな目標のどのあたりまでを狙ったのかと言う記述があるとよいと思います。またそのために取った戦略やシナリオについても是非記述を期待します。

回答 (田中 良夫)

ご指摘の通り、GEO Grid の最終目標と現状での達成度が明確ではありませんでしたので、1章に最終目標を以下の通り明記いたしました。「本研究の目的は、GEO Grid を題材としてグリッドを用いたE-サイエンス基盤を構築し、地球科学の研究者に研究環境を提供するとともに、他の応用分野も含めた幅広い科学技術分野における真の実用化に向けた課題を明確にし、その解決をはかることにある。これにより、科学技術分野におけるイノベーションの創出に寄与することを目指す。」

また、今回のシステム構築を通じて明らかにした解決すべき課題を6章で述べるとともに、8章に記述を追加いたしました。最終目標の達成に向けての戦略については、1章に記述を追加しました。

#### 議論5 要素技術の構成方法および専門用語の説明について 質問・コメント (小林 直人)

個々の要素技術の内容とその選択理由は非常に分かりやすく書かれています。ただし、専門用語が多いので、本文中で分かりやすく説明することが必要です。一方、課題は要素技術の構成方法だと思います。要素技術は、標準プロトコルおよび標準インターフェースに基づいて設計・実装と述べられていますが、その構成におけるユニーク性・革新性・優位性が何なのかを是非記述してください。もちろんシステム構成の容易さと言う点もユニーク性、優位性の一つであると思います。また、この構成方法により確保されたセキュリティやサービスの質が GEO Grid から見て十分なものなのか、まだ多くの改良が必要なのかについても言及していただけるとよいと思います。

回答 (田中 良夫)

要素技術が標準的なプロトコルや API に従って実装されていれば、大規模なシステムを容易に実現できることを実際に示すとともに、各要素技術の開発に際しては標準化や海外機関との連携が重要であることと、大規模なシステムの構築に際してはすべてを自力で開発することは現実的ではなく、コアコンピタンスをしっかりとおさえつつ、利用可能な技術は積極的に利用して開発コストを軽減することが重要であることの主張が本研究の優位性、重要性および本論文の主題と考えています。6章にこれらのことは述べてありますが、これとは別に1章にも記述を追加しました。