

GEO Grid のシステム構成とデータベース統合

美しい衛星画像や緻密な地質図は、それだけでも知的な刺激に満ちたものですが、これらを組み合わせることで、新たな知見が得られます。観測時期の異なる衛星画像からは地形の変化がわかりますし、地質情報や土地利用の様子（土地被覆）、雨量など他のセンサー情報と合わせて考察してみると、地形と地盤、植生、天候などとの因果関係など、社会にとって有益な科学的知見が得られます。一見して分かる画像に限らず、こうした情報の統合は知識の発見にきわめて重要で、GEO Gridにおけるデータベースの統合の技術も、いかに効果的に知識の発見・創出を支援するかが目的のひとつです。

異種、かつ分散した情報をどう統合するのか？

世界中に分散したさまざまな組織で作られた情報には、その価値の高さから、インターネットに自由に公開できないものも多々あります。公開されている場合でも、公開元のWebサイトに人がアクセスして検索しないと情報が得られないのでは、Webサイトにまたがった情報の統合は不可能です。データの形式や項目も作成組織の背景に応じて異なることも多く、情報の相互利用を妨げています。

これらを解決するため、地球観測などの分野では、標準化によってデータ形式とアクセス方法を共通化して相互利用を促進しようとしており、GEO Gridでも、国際コンソーシアムOGC (Open Geospatial Consortium) の定め

た標準のアクセス法やデータ形式などを支援しています。特に、「どこでどんなサービスや情報が提供されているか？」という情報のカタログについて、CSW (Catalog Service Web) というサービスを構築して、GEO Gridで連携する内外のサービスへの入口点としても機能させようとしています。

大規模・複雑化する要求にどう対応するのか？

さて、地球温暖化など複雑かつ地球規模で解決する必要がある問題には、やはり地球規模での大規模なデータ統合が必要になります。環境、防災分野の応用でも、大規模化するにつれて従来の地理情報システムでは扱えないほどデータベースの規模や個数が拡大します。広域分散に対応した高性能なシステムでなければ検索もままなりません。複雑な問題の解決には他の科学分野のデータとも統合する必要が生じるので、分野標準を超えた異種データ統合の問題も発生してしまいます。

そこで、グリッド技術で知られる高性能で規模の拡大が容易な分散技術を適用することがGEO Gridにおける特徴となっています。例えば、GEO Gridにおけるデータベースのアクセスは、OGSA-DAI (Open Grid Service Architecture Database Access and Integration) という英国で研究開発中の分散データベースミドルウェアを元来实现されており、私たちはこれを用いて、台湾と日本との衛星画像データベースを、インターネットを介して、しかもセキュリティを維持して統合する実験に成功しています。

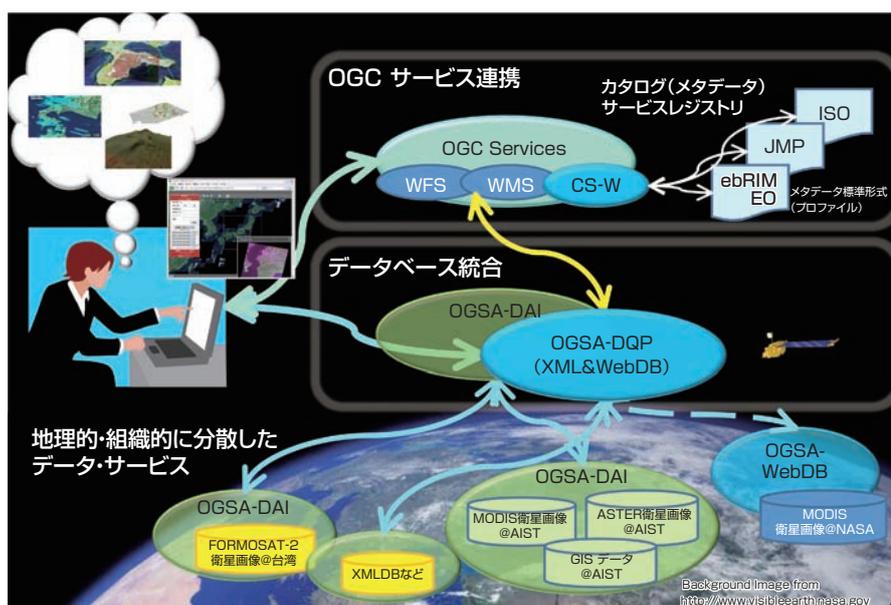


図1 GEO Gridにおけるデータベース統合のイメージ
水色の部分が産総研開発のミドルウェア。グリッド技術による異種の分散データベース統合環境と、異なるメタデータ形式のサービス定義を検索できるカタログサービス(CSW)を、Webインターフェースもあわせて開発している。

産総研ではさらに、このOGSA-DAIを発展させたOGSA-DQP/XML, WebDBと呼ぶソフトウェア^[1]を国際的な協力の下で研究開発しており、XMLやWeb上のデータベースなど異種のデータベースにもアクセスできるよう拡張しています。また前述のCSWの実装においても、構造を意識しない全文検索の機能を効果的に使うことで、多様な構造(スキーマ)のデータに対しても統一的に検索できるシステムを研究開発しています。全体として、図1のような異種の大規模な分散データベースを統合できる基盤を構築しようとしているのです。

将来的には、これらのデータ統合基盤の上での分散データマイニングや分散ワークフローなどによる高度な計算処理や、メタデータの高度化などを可能にするよう発展させることで、計算機主体の科学(e-Science)基盤として知見・知識の創出に貢献することが期待されています。

知的資産のアーカイブとして

デジタル・レポジトリやデジタル・アーカイブとも呼ばれる知的資産のデータベースは、社会と科学の進歩にとって長く必要になる基盤であり、これを有益に使える形で維持管理・改良していくことはきわめて重要なものです。GEO Gridにおいても、産総研のイノベーション・ハブ機能の一環として引き続き技術開発に努め、社会に貢献していきたいと考えています。

GEO Grid システムの概要

以上のような統合の考え方に基

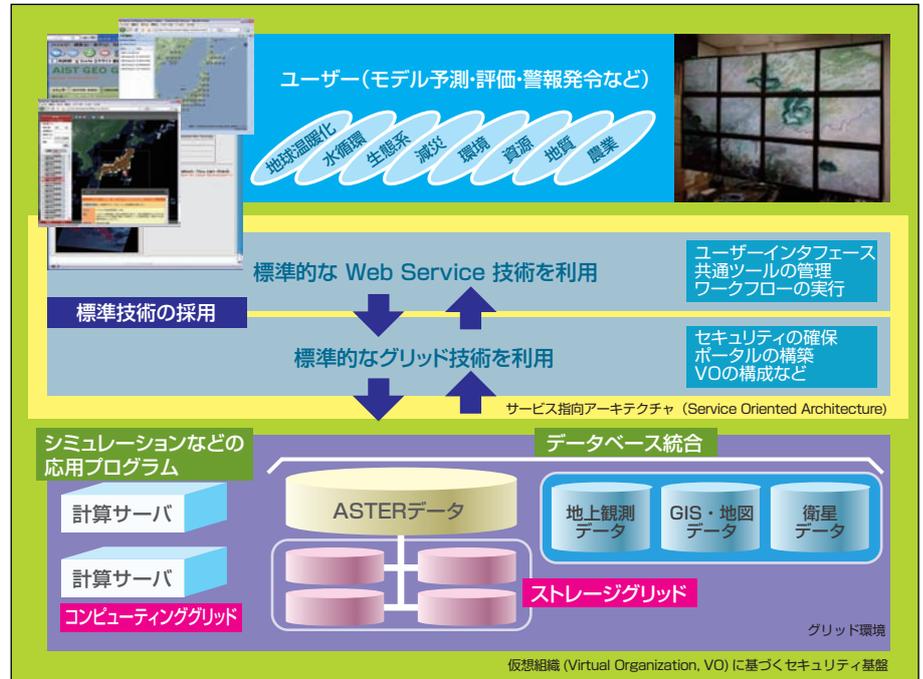


図2 GEO Gridのシステム構成

分散した高性能計算サーバや大規模ストレージがグリッド技術で統合されている。この上でOGC準拠のサービスやワークフローなど応用の支援環境を構築して、地球温暖化や水循環などの応用を支援する。また、大規模なマルチディスプレイを用いた可視化の環境も研究開発している。

き、GEO Gridでは図2のような全体構成^[2]の元で地理的、組織的に分散した計算機群やストレージなどの計算資源を統合(コンピューティング・グリッド、ストレージ・グリッド)しています。グリッド技術を用いることで、地理的分散と規模に対する拡張性を実現しながら、利用者の必要とする計算資源を柔軟に構成・提供することを可能にしています。さらに、ここで述べたデータベース統合のように、これらの資源上で構築されたデータベースや応用プログラム(ワークフロー)、その連携機能などの多様なサービスが実現されます。これらのサービスは、Webサービスに基づく標準のプロトコル

とインターフェースに基づいて構築されており、全体としてサービス指向アーキテクチャ「Service Oriented Architecture (SOA)」を実現しています。さらに次のページで詳述するように仮想組織(Virtual Organization, VO)の概念に基づく柔軟かつ安全なセキュリティ基盤を持つことが特徴です。これらのサービス全体にわたって、OGCが定める標準規格に準拠しており、地理情報にかかるさまざまな応用の支援をするとともに、他のシステムとの相互運用性を高めています。

情報技術研究部門
こじま いさお
小島 功

参考文献

- [1] S.Lynden et al: 9th IEEE/ACM International Conference on Grid Computing(2008.10).
- [2] S.Sekiguchi et al: IEEE Systems Journal, 2(3), 374-389(2008.09).