

GEO Grid における標準とメタデータ管理 Standard and Metadata Management in GEO Grid

小島 功^{1*}, 的野 晃整¹

Isao Kojima^{1*}, Akiyoshi Matono¹

¹ 産業技術総合研究所 情報技術研究部門

¹ITRI, AIST

GEO Grid(<http://www.geogrid.org>) は産総研における研究開発プロジェクトであり、ペタバイトクラスの大規模な衛星データアーカイブを中心とした分散情報処理・利用基盤を提供することで、科学研究の促進や新たなビジネスの創出に貢献しようとするものである。GEO Grid の基盤技術には以下のような特徴がある。

1) グリッド・クラウド技術に基づく情報基盤の実現: グリッドに基づく安全な認証や分散処理技術を使い、分散したデータベースの連携や高性能計算機の連携等を安全かつ容易に実現している。これにより異なる組織が保有するデータベースや大規模なサービスの安全な相互利用を促進できる。

2) OGC 標準の支援: この基盤上に OGC(open geospatial consortium,<http://www.opengeospatial.org>) が規定する標準、例えば地理空間情報のメタデータの検索 (Catalog Service Web,CSW) や地図検索 (WMS), 処理サービス (WPS) などの一連の規格を支援し、Web サービスとして提供している。大規模なデータに対する広範囲な利用や他の OGC サービスとの連携・相互利用を促進できる。

3) 独自コンテンツや高性能サービスの提供: これらの OGC サービス群を用いて独自のコンテンツや高度な応用を提供している。例えば、経産省の衛星センサ (ASTER) の画像データの並列処理に基づく、3D モデル (digital elevation model,DEM) の高速生成サービスや、複数組織の持つデータベースを連携することで独自の知見を提供する地震動マップ公開サービスなどである。これらを広く一般の利用者に提供することで独自のコンテンツや技術の普及をはかることができる。

今後データや処理が大規模化・複雑化するにつれてグリッドやクラウドといった大規模・分散基盤の重要性は高まる一方で、大規模分散基盤上で地理空間標準を支援してサービス提供するというアプローチは重要と考える。

さて GEO Grid においては、分散したデータやサービスの連携のためのメタデータの検索・管理が一つの重要な問題である。

1) 標準データフォーマット: 地球観測、地理空間関連のメタデータフォーマットは、ISO を代表として多くの分野標準があり、標準同士の連携が課題である。また、地物と衛星、衛星でも光学やレーダなど、対象によってメタデータの形式が微妙に異なる点も解消する必要がある。

2) 標準アクセス規格: アクセスの規格も OGC だけでなく OAI(Open Archive Initiative) を代表として多くの種類がある。また、OGC のメタデータ検索 (CSW) についても、分散メタデータ処理の実例が少ないために仕様が十分でなく、実装によって機能が固定されるといった問題がある。

これらの問題に対して GEO Grid では、OGC CSW を支援する独自のソフトウェア実装 (AIST-CSW) を研究開発している。最大の特徴は、従来一般的であった関係データベースによる実装でなく、全文検索ソフトウェアを基礎としている点であり、これにより以下のような特徴を有している。

1) 異種データフォーマットの統合検索: テキスト検索に基づく全文検索ソフトウェアに基づくことで、フォーマットを意識しない検索と日付や数値などによる比較検索、特定のフィールドに対する条件検索などを両立でき、異種のフォーマットのデータを容易に統合検索できる。AIST-CSW では XML の衛星プロファイルを基礎として、メタデータフォーマットとして ebRIM のモデルを採用することで、拡張性とメタデータ検索の柔軟性を実現した。

2) HTTP に基づく簡便な A P I の提供: 多くの全文検索サービスで提供されている、REST によるインターフェイス (API) を策定し、この上に CSW の仕様を実装した。これにより他のサービスや他のアクセス規格、例えば OpenSearch や検討中の CSW 規格との連携を容易にするとともに、データが大規模化した場合の並列検索を可能にした。

3) 答の個数に依存しない応答性能: グラフィックインターフェイスを用いた検索では、利用者が意図せず膨大な個数の検索を行うことがあるが、関係データベースに基づく実装の多くは、応答時間が個数に比例しているものが多い。全文検索エンジンを用いることで、検索結果の個数に依存しにくい性能が実現でき、大規模なデータベースに対する検索を容易にした。

GEO Grid では、AIST-CSW を用いて約 180 万件の ASTER 衛星のシーンカタログを上記 ebRIM の XML 形式で管理しており、ブラウザに基づくインターフェイスを介してメタデータの検索とそのメタデータに基づく衛星画像を WMS で提供するサービスを実現している。また、CSW の分散検索規格を改良して複数組織の保有する CSW サーバの統合検索

機能を実現しており、数十程度のサイト数での実用性を確認している。

参考文献：I.Kojima et al. "OGC Catalog Service for Heterogeneous Earth Observation Metadata using Extensible Search Indices", ACM GIR(Geographic Information Retrieval) 10, 2010.02.

キーワード: 地球観測, メタデータ, OGC 標準, GEO Grid

Keywords: Earth Observation, Metadata, OGC Standards, GEO Grid