

# XQuery および OAI に基づくメタデータ統合検索システムの実現と その OGS I・Web サービス基盤への検討

小島 功<sup>†</sup> 花坂 元伸<sup>‡</sup>

<sup>†</sup>産業技術総合研究所グリッド研究センター 〒305-8568 つくば市梅園 1-1-1

<sup>‡</sup>産業技術総合研究所先端情報計算センター/三菱スペース・ソフトウェア 〒305-8561 つくば市東 1-1-4

E-mail: <sup>†</sup> kojima@ni,aist.go.jp <sup>‡</sup> r0614@aist.go.jp

**あらまし** . 本稿では、多様なデータベースのメタデータを統合するシステムについて述べる。多様な情報を統合するためのインターフェイスとして、OAI(Open Archives Initiative)プロトコルをベースとしたデータインターフェイスと、XQuery のサブセットに基づく問い合わせ機能を組み合わせている点で、検索結果自体もこのプロトコル互換のデータ集合となっている点に特徴がある。また、多様なメタデータプロファイルを統合的に扱うために、異なるXMLスキーマ間においての対応に基づく等価・変換関係を管理することができる。これを用いて問い合わせ変換を行うことができるので、ボトムアップ的に構築されたデータに対して統合的な検索が可能となる。さらに、グリッド・コンピューティングで用いられるOGS I基盤についての検討を行い、その結果を示す。

**キーワード** XML , OAI , メタデータ、プロファイル、XQuery、グリッド基盤

## Design and Implementation of Integrated Metadata Management System Based on XQuery/OAI and Its Application for OGS I/Web Services

Isao KOJIMA<sup>†</sup> Motonobu HANASAKA<sup>‡</sup>

<sup>†</sup> Grid Technology Research Center, A.I.S.T 1-1-1 Umezono, Tsukuba 305-8568 Japan

<sup>‡</sup> Tsukuba Advanced Computing Center/Mitsubishi Space Software , A.I.S.T. Higashi 1-1-4, Tsukuba, 305-8561, Japan.

**Abstract** This paper presents the design and implementation of a metadata management system to integrate various research databases.. The system is based on OAI(Open Archives Initiative) protocol and the subset of XQuery -based database functions are combined with OAI data access. Result set of XQuery is also a set of OAI data. In order to cover various XML schema, relationships which describes equivalence and conversion of XML schema components are managed. Querying over different XML schema is possible by query modification methods using this information. Application of this system to OGS I/Web service environment for grid computing is also discussed.

**Keyword** XML, OAI, Metadata, Profile, XQuery, Grid Computing

### 1. 背景

産業技術総合研究所(産総研)では、RIO-DB(研究情報公開データベース)をはじめとする、多分野の研究情報データベースを多数作成している[2]。これらは、研究プロジェクトやグループの成果や実験の結果として発生する情報であり、スキーマやシステム設計がボトムアップ的で、相互の連携による高度利用が難しいという課題がある。分野ごとの

共通的なスキーマ設計なども進んでいるが、研究所全体の分野横断的なデータ項目に関しては十分でない。そこで、既存の個別データベース作成・データ整備関連の活動を生かしながらデータの相互連携や統合を容易にする試みとして、XMLスキーマに基づくメタデータ統合・検索システムを設計開発し、プロトタイプ[1]を実現したので報告する。

## 2. 課題

システムの目的は、既存のデータベース群を統合するためのデータベースの構築で、メタデータ場合にはデータそのものを統合し、その上の統一的なインターフェイスを提供することでデータ統合を容易にするものである。この実現にあたっては、次のような課題がある。

- 1) **ラッパ・ハーベスタによるメタデータ管理**：既存のデータベースやアプリケーションをそのまま利用し続けるため、ラッパあるいはハーベスタを用いてデータを収集・アクセス・管理することで、既存データベースとアプリケーションをそのまま稼働できるようにする。
- 2) **ラッパを含めた共通アクセスプロトコルの実装が簡便かつ関連ツールが多数存在すること**。一般にラッパを用いた実装では、ラッパの実装が簡便であることが必要であり、特に今回のように、データソースが多様な場合、ラッパやその生成・実装のコストが低いことが望ましい。さらに、分野の研究者はデータベースプログラミングの専門ではないため、ラッパを実現するプロトコル上で他の様々なツールやプログラムが稼働でき、DB応用以外の利用や実験に用いられることも重要である。
- 3) **多様な形式のメタデータプロファイルを扱えること**。メタデータには古くは文献のMARCや地理情報のGILS、JMPなどから、最近のXMLベースの情報交換形式に至るまで、分野ごとのさまざまなものが策定あるいは議論中であり、これらを比較的無理なく記述・格納できること。
- 4) **簡便な形でのスキーマ変換・統合を導入すること**。既に作られたデータベースでは、大量のデータやそれを使ったアプリケーションが既に多数存在するため、大規模な変更が発生しやすいトップダウン的なスキーマ統合は難しい。ところが現実には、異なるデータベースでの異なるデータ項目が同一の意味であったり、同じデータタイプが異なるデータ表現を持っていることが多々ある。そのため、何らかのスキーマ統合や型変換の機能を導入する必要がある。
- 5) **検索結果などの再利用や変換が可能なこと**。例えばRID-DBプロジェクトの多くのテーマでは、研究グループがデータベース上でのプログラミングに割ける労力はそれほど多くない。したがって、高度な検索機能を標準で提供すると共に、その結果に対する再利用や、データ形式の変換等が、可能な限り標準のプログラミング手法などを用いて実現されなければならない。

## 3 アプローチ

以上のような要求に対して、本実現におけるアプローチには、以下のような点で特徴がある。

- 1) **OAI (Open Archives Initiative) 互換による分散データベース基盤**：それぞれに構成されるデータベースおよびラッパの構築コストを下げるため、OAI Harvesting Protocol [3] とよばれる、分散デジタルアーカイブ構築とアクセスのためのプロトコルに基づく分散DB基盤とした。OAI 関連のツールをそのまま利用できるため、新規データベースの構築と接続が容易と考えられる。
- 2) **XMLスキーマベースのデータベース記述と、XQueryに基づくデータ操作機能の提供**：OAIは、HTTPベースで実装が容易であり、さまざまなツールが開発されているが、
  - (ア) 集合的な問い合わせ機能が存在しないため、クライアント側での機能実装の負荷が高い。
  - (イ) スキーマは OAI が規定するもののみが現時点で想定されている、という問題があるため、問い合わせ機能としてXQuery[5]を実装しOAIプロトコルに組み合わせた。また、対象スキーマをXMLスキーマ一般に拡張した。メタデータの扱いはRDF (Resource Definition Framework) を用いることも多いが、本システムでは問い合わせ機能との整合性と、周辺ツールの豊富さと記述力等のバランスからXMLスキーマを基礎とした。
- 3) **異なる複数のXMLスキーマの横断検索**：個別に作られたデータベースに対して、スキーマ間の関連付けを設定でき、複数のXMLスキーマにまたがった横断的な検索ができる。これはスキーマ統合の問題であり、スキーマ変換・統合からオントロジーの問題など、関係する研究が多数なされている。メタデータ統合の場合、意味的な統合はそれほど複雑ではないが、フォーマット同士の変換などは、かなり複雑・多様なものが設定されている場合が多く、この両者を簡便な形で実現する必要がある。
- 4) **XQueryの支援と、キーワード検索に基づくXMLデータベース検索**：XMLベースで管理するために、XQueryのサブセットを実装した。SQLの関係DB (Oracle) をサーバとするもので、XQuery/SQL変換のフロントエンドとしても利用可能である。一方で、Oracleが有する全文検索機能を用いたインターフェイスも設定されており、サーチエンジンのような単純なキーワード検索としても利用ができる。XQueryによるデータ加工も可能となっており、検索結果の加工ができる。検索結果は元のDB同様にOAIプロトコルに互換のデータ集合として生成するので、外部からの再利用も可能である。
- 5) **複数のプロトコルを支援する、データ収集・変換メカニズム (ハーベスタ)の実装**：すでに作られたデータベースから効果的にメタデータを収集するために、OAIだけでなく、WWW(HTTP)経由やSQLによるものなど、複数

のプロトコルを支援する。

このアプローチに基づく、システムの概要を図 1 に示す。

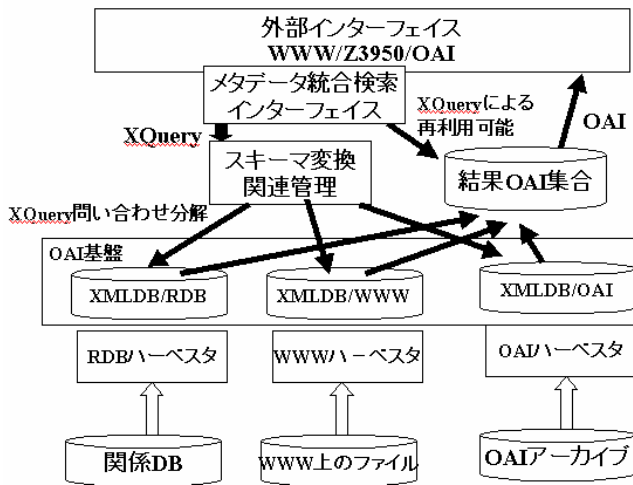


図 1 全体の概要 各データベースはハバスタによって OAI 互換の XML データベースとして記憶される)

#### 4. 比較

メタデータの統合には、さまざまな考え方やシステムがあり、またいくつものシステムが実装されている。これらのシステムと比較して、本システムの特徴を、次の3点で述べる。

4.1 ラッパ・ハバスタプロトコルの比較: 古くは、Z3950 などによる地理情報、図書情報の検索システムや、CORBA などによる分散データベース統合、SOAP ベースのデータベース統合[13]なども実装されている。

- 1) **OAI プロトコルの利用**: OAI は HTTP ベースのデータアーカイブへのアクセスプロトコルであり、分散資源に対するデータアクセスとして極めて単純、実装が容易である。また、OAI インターフェイスは集合データに対する標準的な HTTP アクセスと表現を規定するため、例えば SOAP でデータベースを接続した場合より仕様が厳密で互換性が高い。
- 2) **結果セットの作成と再利用**: 一般にラッパなどにより XML データベースを実現した場合、結果が XML 集合になるため、そのシステムの中で操作する必要がある。一方、本システムでは XQuery の仕様を制限する過程で、OAI のデータ集合を構築している。このことで、中間検索結果などすべてを OAI 互換のツールで再利用することができる。

4.2 スキーマ変換と統合: 分散データベースやマルチデータベースにおけるスキーマ変換やスキーマ統合、オントロジーなど、さまざまな研究が多様な形で[6,7,8]なされている。

- 1) **メタデータ向けの対応・変換の支援**: 本システムでは、意味的な対応・統合要素としては XML スキーマを基礎とした詳細化 (Qualification) と拡張 (Extension) をを扱うものとし、言語によらないものとした。次に、対応づけられた要素のフォーマット間には、プログラム処理や XSLT 変換など、かなり自由度の高い変換を結びつけるものとした。これは、メタデータははじめから意味的な類似性が高いため簡単な対応要素でも十分な能力を持つことと、現実のデータではフォーマット変換や型変換などかなりの自由度が必要とすることによる。
- 2) **拡張できる核スキーマの設定**: 対応関係が複雑化して意味的に矛盾(例えば、対応関係にループが生じて、同じデータベースの異なるフィールドに同一の視野が与えられるなど)が発生するのを防ぐため、後述の Dublin Core[4]と呼ばれる、基礎的メタデータプロファイルを核のスキーマとして設定し、新規に追加される XML スキーマは、この核スキーマとの対応関係等で記述する。

これは、オブジェクトデータベースにおけるスキーマ統合の一部の機能に過ぎないが、メタデータの統合にはこの程度の対応関係で十分な表現能力を持つと考えられる。また、スキーマ変換・統合は対応関係ベースであるため、言語によらずに簡単なウィンドウインターフェイスで管理することができる。また、問い合わせ言語の変換ロジックも容易となる。なお、このようなスキーマ間の対応づけは、応用や視野によって異なることが考えられる。したがって、本システムでは、将来的にこれらの情報自体をビューとして個別に環境を提供できるよう配慮されている。

4.3. 問い合わせ言語の機能: XQuery は W3C の規格であり、さまざまなメカ等の実現がなされつつある。ここでは、先に述べた考え方から、後述のように、XQuery 問い合わせを2つの機能部分に分離しており、中間結果集合の再利用とシステムの実現容易性を可能にした。なお、全体として本システムは次世代 Z3950 の目標検討機能に近い[9]が、OAI ベースである点と XQuery の機能を分けているところが異なる。

#### 5. システムの概要:

基本的には、本システムは OAI 互換のデータベース集合に対して、XQuery による検索と、対応関係の管理と横断検索を実現したものである。

5.1 **メタデータ統合検索機能**: 検索にはキーワード検索と XQuery によるものとの2つのインターフェイスがある。キーワード検索は利用者の簡便のため、内部的には共に XQuery ベースの処理として実現されている。

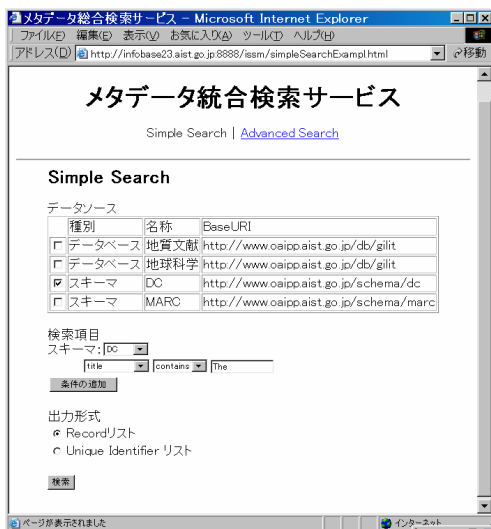
1) XQuery による検索 :本システムでは、XQuery を最終的に RDB 問い合わせに変換しているが、一般の XQuery の処理を、

- A) 問い合わせ記述部分 :FLWR 表現を中心とする検索条件式の記述
- B) 結果データ加工記述部分 :検索結果のデータを加工して、最終的に求めるデータ形式を得る部分。

に分けて記述するようになっている。これは XQuery の記法上は大きな制約となるが、実装の上からは、データベースの検索部分と結果加工部分のそれぞれの目的に応じた部分の変換処理を中心に実装すれば済む利点があり、FLWR 表現からの SQL 変換はかなり簡便となると期待される。問い合わせ処理の過程で、問い合わせはスキーマ間の対応関係情報に従って、個別のスキーマに対する XQuery に変換されて実行される。それぞれのデータベースからの結果は、マージされて、2) でひとつの集合として構築される。

2) OAI 結果集合の構築 :上記のように検索と結果データ加工に分離する過程で、検索結果のデータ集合をOAI 互換のデータセットとして作成することができる。この OAI データ集合には名前付けができるため、OAI 互換のクライアントで検索結果等を再利用することができる。一般に XQuery の最終的な加工結果は、結果を記述するXML スキーマがないため再検索が難しいが、本システムでは再利用が可能である。

検索インターフェイスと、検索結果を図 2 の 1 - 3 に示す。



(図 2 1 キーワード検索インターフェイス。スキーマやデータソースを選択して、キーワード検索を行う)



(図 2 2 検索結果の例。本システムでは、日本語を含むデータを扱う)



(図 2 3 検索インターフェイスの例：上記部分が問い合わせ部で、下の部分が結果の加工フィルタ部分である。中間段階で結果セットをどういう形式で作成するか指定できる)

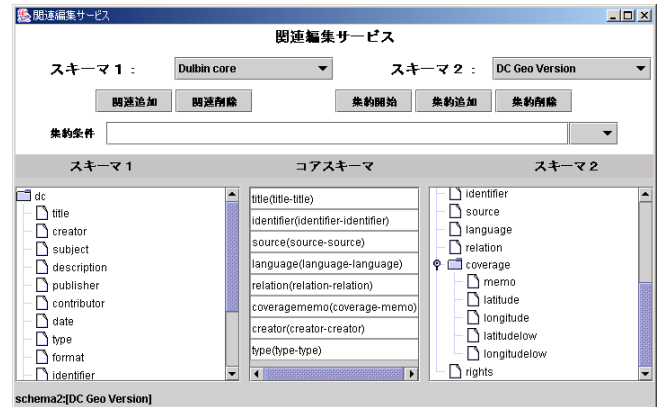
(注 :本論文の画面は、関連編集以外はすべて <http://www.aist.go.jp/infobase/issm/> におかれた例から実行可能である。関連の編集のみは、JDBC によるデータベース接続を使うので、インターネットからは実行不可能)

5.2 スキーマ管理機能 :対象とするデータベースのスキーマを XML Schema で与え、それに基づいてメタデータを管理する。スキーマ間の関連付けの機能があり、異なった XML スキーマ間での横断的な検索に用いる。

- 1) Dublin Core を基礎とした、核スキーマの設定 :メタデータ関連のデータベースは、スキーマの相違はあるものの、ある程度共通な項目も潜在的に存在する。たとえば名前とか作成者などである。本システムでは、あらゆるメタデータに対する共通的なプロファイルとして提案されている、Dublin Core ElementSet を核スキーマとして設定し、新たに定義あるいは収集するデータベースのスキーマとの対応関係を管理する。
- 2) 対応関係管理 :対応関係は横断的な検索の基礎となるもので、原則として異なる名前の属性を結んで関連づけることで、同一のものとして扱うこととする。加えて、関連付けの過程で必要となるデータ変換を記述する。既述のように、メタデータの対応関係の設定は、1) エLEMENTの拡張、2) エLEMENTの詳細化 の2つなので、対応付けと集約の処理だけでも、メタデータ関連では十分な横断検索能力を持つと考えられ、記述言語を必要としない。なお、詳細化に対する集約処理は現時点では AND/OR/ と連結を実装している。

システムの実行例を図3に示す。

例えば緯度経度など、当初の核のスキーマにない要素を2つのスキーマ間で設定すると、それに対応した要素が核のスキーマに追加され、2つのスキーマでのそれぞれの要素は、核に追加された要素との間での対応関係として記憶される。これにより、対応関係の設定された任意のスキーマ要素は、必ず核となるXMLスキーマに含まれる形となり、大域的な一貫性の問題を避けている。逆に、この核をユーザ単位で設定すれば、個別の視野の設定が可能になると考えられる。



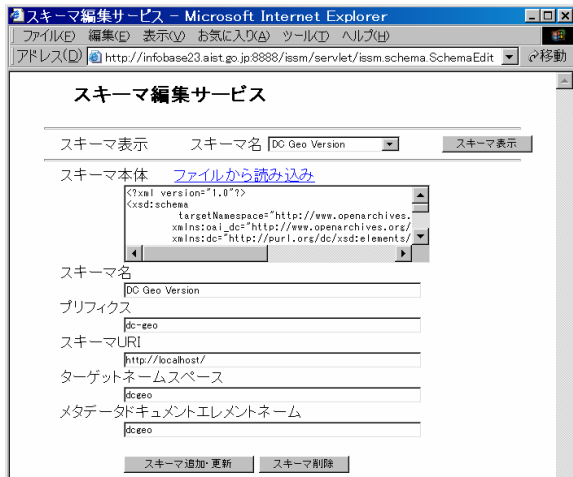
(図3 関連編集サービス :2つのスキーマを比較して、集約の設定を行う例で、対応関係の設定は必ず核となるコアスキーマとの二項関係の設定の繰り返しにより行う)

5.3 データベース ハーベスタ :分散して作られるデータベースに対してアクセスを行い、メタデータを収集するハーベスタ。対象とするデータベース実装は多数あるので、複数のハーベスタがある。

- 1) RDB ハーベスタ :Oracleなど、関係データベースのデータに対するデータ収集機能。SQLによる収集と、XMLスキーマ表現に沿った形の変換を行う
- 2) WWW ハーベスタ :WWW上のXMLドキュメントの形で格納されるデータに対する収集機能 :ロボットによりWWWページとしてXMLで書かれたデータを収集する
- 3) OAI ハーベスタ :Open Archive Initiative Metadata Harvesting Protocol 互換のハーベスタ。

現時点では、RDBハーベスタとOAIハーベスタが実装されている。OAIハーベスタやWWWハーベスタは、新規にデータベースを整備するためのもので、例えばOAI関連のツールでデータベースを構築すれば、OAI関連のツールによる検索以外に、本システム経由のXQueryによる検索やキーワード検索などができる。WWWハーベスタも同様で、XML文書のメタデータをディレクトリ内にそろえれば、DB化は本システムが行うので、DB作成者はデータの作成に集中できる。

5.4 スキーマ定義・データ投入ツール :XMLで書かれたデータを直接この統合検索システムに登録するツールと、そのためのXML Schemaを定義するためのツール。ここでは、単にテキストを処理するだけで、メタデータの作成やスキーマの作成そのものは、XMLSpyなどのツールによることを想定している。(図4参照)



(図 4 スキーマの表示と編集)

## 6. システム構成とプロトタイプ:

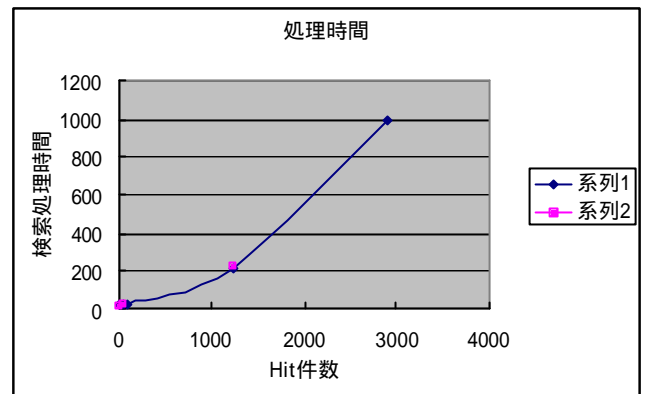
### 6.1. プロトタイプの構成

プロトタイプを Java (JDK1.4 以降) を用いて Linux (RedHat7.2) 上に構築した。データベースは JDBC 経由で Oracle (8.1.6) に接続されている。XML データや XQuery などの XML 表現は、DOM ベースで処理されている。OAI プロトコルは、バージョン 2 に準拠している。実現上の詳細としては以下のような実装を行っている。

- 1) データベースへの格納 :最終的にデータはすべて関係データとして格納する必要があるため、DOM 構造に基づくパス表現と、その先のノードの組でデータを表に記憶している。(テキストとしても別に保持している)
- 2) FLWR 表現の変換 :XQuery の FLWR 表現は、上記構造を持つデータ集合に対する SQL に原則として等価変換されて処理される。これは、商用システムにおける XQuery - SQL 変換とほぼ同等の機能である。なお、キーワードインターフェイスにおけるキーワード検索などを支援するために、XQuery から SQL に変換する過程で、Oracle における全文検索機能を前提としている。
- 3) データ加工部における処理で、特にデータベース検索を必要としない場合には、そのまま XSLT など XML のツールが呼び出され、その上で処理される。

### 6.2 性能評価

現在のプロトタイプは、表 1 のように性能が高くなく、3000 個の検索結果で約 1000秒程度の検索処理時間がかかっている。RDB の索引を用いている関係上、元データの個数は大きな要素となっていない。



(表 1 検索処理時間と結果 Hit 件数との関連)

これらの処理時間をさらに詳細に検討すると、表 2 のような内訳となっている。表 2 のとおり、処理時間のうちほぼ 8-9 割が、バックエンドの Oracle との個々のデータのやり取りと、中間結果のデータベースへのレコード単位での書き戻し起因するものであり、この部分の改善は十分可能であると考えられる。

次に、XQuery-SQL 変換のために約 20 秒前後かかっており、この部分は、DOM と同様の問い合わせ木構造への XQuery の展開と、SQL 相当に対する問い合わせ木の交換にかかる部分で、この部分でのパブリックドメインソフトの組み合わせなど XML 処理系のつなぎ合わせなどによるオーバーヘッドと考えられている。この 2 つは改良しやすいため、これを除けば、応答時間は現実的であると考えられる。

| 結果個数 | 応答時間 | XQuery SQL 問い合わせ変換時間 | SQL 実行時間 | 結果データ取得 | 対応関係別結果取得 | データ加工 |
|------|------|----------------------|----------|---------|-----------|-------|
| 0    | 24   | 23                   | 0        | 1       | 0         | 0     |
| 43   | 25   | 23                   | 1        | 1       | 0         | 0     |
| 10   | 26   | 24                   | 0        | 0       | 0         | 0     |
| 43   | 27   | 24                   | 1        | 1       | 0         | 1     |
| 98   | 28   | 22                   | 0        | 1       | 2         | 0     |
| 5    | 28   | 26                   | 0        | 0       | 1         | 0     |
| 1226 | 219  | 25                   | 1        | 89      | 93        | 8     |
| 1226 | 230  | 23                   | 1        | 99      | 103       | 1     |
| 2897 | 997  | 23                   | 1        | 556     | 317       | 0     |

(表 2 応答時間の内訳 :時間はすべて秒)

### 6.3 新規のデータベース構築と応用構築のアプローチ

以上のようなシステム環境を用いて新たにデータベースを構築あるいは統合する場合、以下ようになる。

1) 既存のRDBに存在するものなら、RDBハーベスタによりメタデータを収集させるか、RDBへのOAポータルウェイを実装して、本システムによりメタデータあるいはオリジナルのデータを収集させる。

2) 他のデータソースの場合、該当するハーベスタに対応する形式でデータを作成すれば、システムが収集を行う。

3) 本システム上ではメタデータはオリジナルの複製として扱われ、別に一意の識別子が与えられている。オリジナルのDBとの関連は、現実装ではURLがメタデータ内に埋め込まれるが、HTTPベースの他のアクセス手法についても可能なように(SOAPなど)想定しており、実データとあわせて統合的なデータベース処理へ対応している。

このように、多様な形式でデータを作成する作成者側に可能な限りシステム機能の実装の負荷をかけないように配慮している。

## 7. アプリケーションの構築およびグリッド基盤の検討:

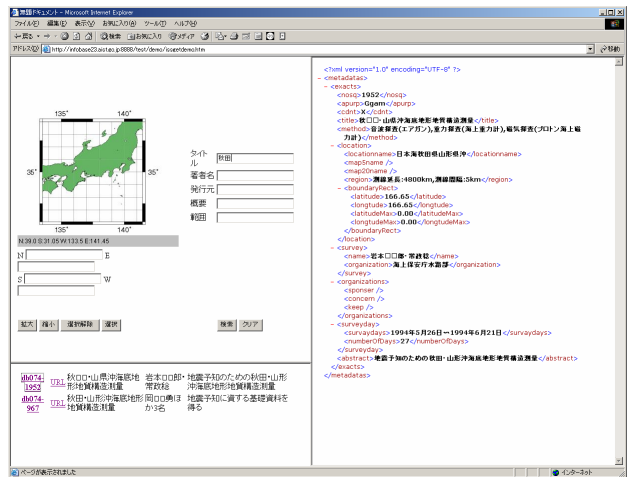
本章では、本システムのグリッド環境への適応を目的として、応用の構築と、その過程においてグリッドサービス化におけるいくつかの検討を行った。

WWW サービス・グリッド基盤への拡張については、グリッドサービス(Open Grid Service Architecture)としてこの応用を提供するものであり、1月中旬にOGSAの核部分のAlpha版が公開されている。

### 7.1 地理アプリケーションサービスの構築:

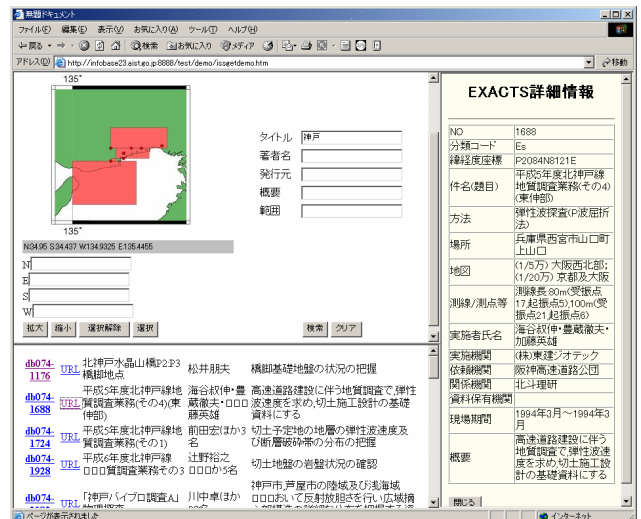
本システムにおける、具体的な応用例として、RIODBにおける、地質・地理情報関連の統合検索サービスを実装した。地理関係のデータベースは、多くが緯度経度情報を有するため、それぞれの個別のスキーマに含まれる緯度経度を相互に対応付けている。あわせて、ハーベスティング(元データからデータを収集する)過程で、現データに対するURLを生成してデータベースに格納した。図4-1に検索の結果、4-2は、メタデータを用いて、オリジナルのデータベースに詳細情報をアクセスした場合を示す。画面上で入力した検索条件や、カーソルで範囲指定した緯度・経度にかかわる情報を含めた検索条件は、すべてXQueryに変換されてシステム上で検索されている。

さらに、結果として返ってくるデータベースのリストは、これもOAICatなど、OAI互換のツールで別個にアクセスすることも可能である(図4-3)。このように、メタデータを介したデータベース統合を行うことで、複数の情報源を有機的に連携可能と考える。

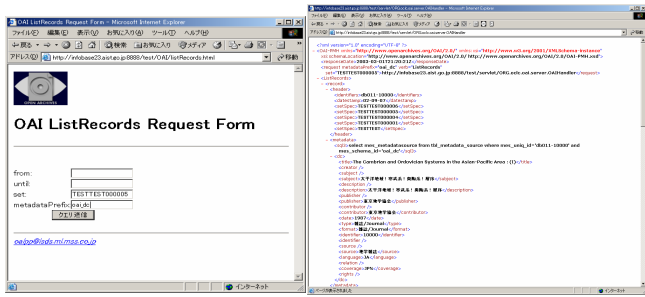


(図 4-1 地図インターフェイスと検索結果リスト、個々のメタデータの表示画面。緯度・経度は地図上で領域選択し、この項目情報とあわせて検索する。結果は左下に表示され、右側に個々のメタデータが表示される)

本システムではメタデータ統合を想定しているが、実際に投入できるXML Schemaには制限がほとんどなく(現時点では、カンマ仕切りのリスト型が扱えない)、また、XQueryの問い合わせ能力が高いため、一般のデータベース検索系としても用いることができる。



(図 4-2 メタデータを介してオリジナルデータにアクセスした例。右側のデータは別のDBからのアクセス結果で、メタデータに入っているURLによりアクセスしている。なお、検索結果オブジェクトの緯度経度領域を矩形で表示している)



(図 4-3 OAI 関連ツールによる結果セットのアクセス)

この実装の過程で、以下のような知見が得られている。

- 1) OAI の HTTP ベースのプロトコルは、SOAP との親和性が高いが、結果のデータは XML をいったんテキストとして落として転送しているため、そのまま XML で転送したほうが、クライアント側での処理がたやすい。
- 2) OAI プロトコルは SOAP に比べて簡便な実装を狙っているため、SOAP による本プロトコルの実装は利点の 1 つを失いかねない。
- 3) 中間検索結果を一時サービスとして提供することは非常に有益であると考えられる。例えば、現時点では中間検索結果として作られた OAI データ集合は蓄積される一方であるが、利用されない結果のデータサービスを回収するには、OGSI におけるファクトリ機能が有益である。

8. まとめ：

本稿では、OAI プロトコルと XQuery に基づく、データ統合検索システムについて述べた。述べたとおり、本プロトタイプはいくつもの検討課題をかかえており、特に性能上の問題が大きい。しかし、例えば専門の研究者はデータベースシステムの構築面では労力をかけられない一方で、自らの持つデータを容易にデータベース化し、かつ高度な検索や加工の機能を実現し、また他人のデータと連携することへの要求はとても高い。ここでのアプローチは、OAI などの簡潔で整備ツールのそろったプロトコルなどうまく組み合わせることで、データベース作成コストを下げるのが期待されている。Web サービス基盤・グリッドサービス基盤への応用は、OAI やメタデータ統合の考え方と合わないために効果的な結果が得られていないが、データ統合は必要な応用であるので、さらなる改良を行う予定である。

**謝辞：** 産総研の諸氏、特に RIO - DB にかかわる各テーマ担当者のご協力と、塚本啓一氏、村山清司氏をはじめとする先端情報計算センターのスタッフに感謝いたします。

文献：

- [1] メタデータ統合検索プロトタイプシステム <http://www.aist.go.jp/infobase/issm/>
- [2] 産業技術総合研究所研究情報公開データベース RIODB(Research Information Data Bases) <http://www.aist.go.jp/RIODB/>
- [3] The Open Archives Initiative Protocols for Metadata Harvesting Protocol V2.0 <http://www.openarchives.org/OAI/openarchivesprotocol.html>, 2002.06.
- [4] Dublin Core Metadata element Set, Version 1.1 Reference Description, <http://www.dublincore.org/documents/1999/07/02/dces/> 1999.07.
- [5] XQuery:An XML Query Language W3C Working Draft, <http://www.w3.org/TR/2002/WD-xquery-20021115/> 2002.11.
- [6] A.Apps, R.MacIntyre L.Morris, "Exposing Cross-Domain Resources for Researchers and Learners" DC2002.2002.10.
- [7] S.Collins S.Navathe, L.Mark "XML Schema mappings for heterogeneous database access" OOPSLA,2001.10
- [8] E.Ralm, P.Bernstein,"A survey of approaches to automatic schema matching" VLDB J 2001.
- [9] ZING - z39.50 International:Next Generation Search/Retrieve Web Service Definition <http://www.loc.gov/z3950/agency/zing/zing-home.html> 2002.11.
- [10] Web Ontology Language Reference V1.0 W3C Working Draft, <http://www.w3.org/TR/2003/WD-owl-ref-20030221/> 2003.02.
- [11] S.Tueke et al, "Open Grid Service Infrastructure (OGSI)", GGF working draft, 2003.02.
- [12] N.Paton et al. "Database Access and Integration Services on the Grid" GGF Working Draft, 2002.02.
- [13] T.Malik et al. "SkyQuery: A Web Service Approach to Federate Databases" CIDR 2003, 2003.01.