

サブテーマ 3

データベース連携技術に関する研究

産業技術研究所 情報技術研究部門

小島 功 田中良夫

データベース連携の背景

- 分散した大規模なコンテンツとその高度処理
 - 知見の創出や発見にきわめて重要

e-Science => Content Aware な応用

- 産総研 GEO Grid
- デジタル・アーカイブ、デジタル・レポジトリ
 - 科学の発展の基盤として全世界的に整備されつつある
 - データの**構造化・抽象化・知識化**による整理(データベース化とその高度化)



分散・大規模化 ー > 相互接続・連携が課題

地理的・組織的に分散して 生成・蓄積された情報の統合

- IVO(仮想天文台)に基づくデータベース天文学
- GEO Gridにおける広域のデータベース連携
- 統合遺伝子データベース上のデータ解析・マイニングなど

ボトムアップ的な情報統合

標準化

OGF DAIS-WG
提案者が主導中

- 各組織やグループが(勝手に)構築した異種の情報統合する。
- トップダウンの分散ポリシーや、統一したデータモデルに欠ける。

異種データベースの連携・統合の必要性

1. 異種データベースの統合・連携
2. 組織間のデータ管理ポリシーの違いの吸収

サブテーマ3 データベース連携技術に関する研究

- 背景

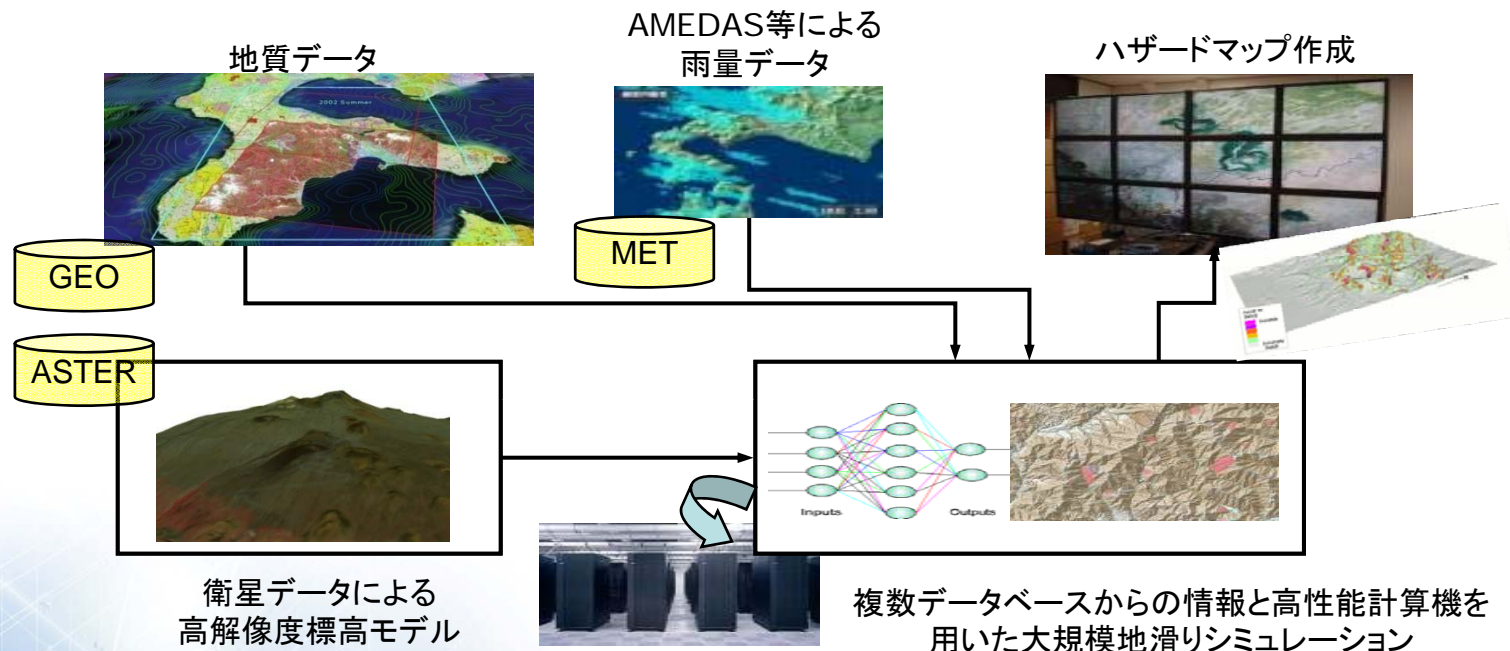
- 異なるデータ格納方法、アクセス方法、スキーマ、データポリシなどが混在したデータベース群を扱う技術が確立されていない。

- 目的

- 分散環境での異種多様なデータベースを、簡便かつ統一したインターフェースで安全に連携させる技術を開発し、e-サイエンスの利用を推進し、科学技術の発展に貢献する。

- 応用例

- 複数のデータベース連携による地滑り予測



研究開発の内容:

1. 異種の分散データベースを連携・統合するミドルウェアの研究開発

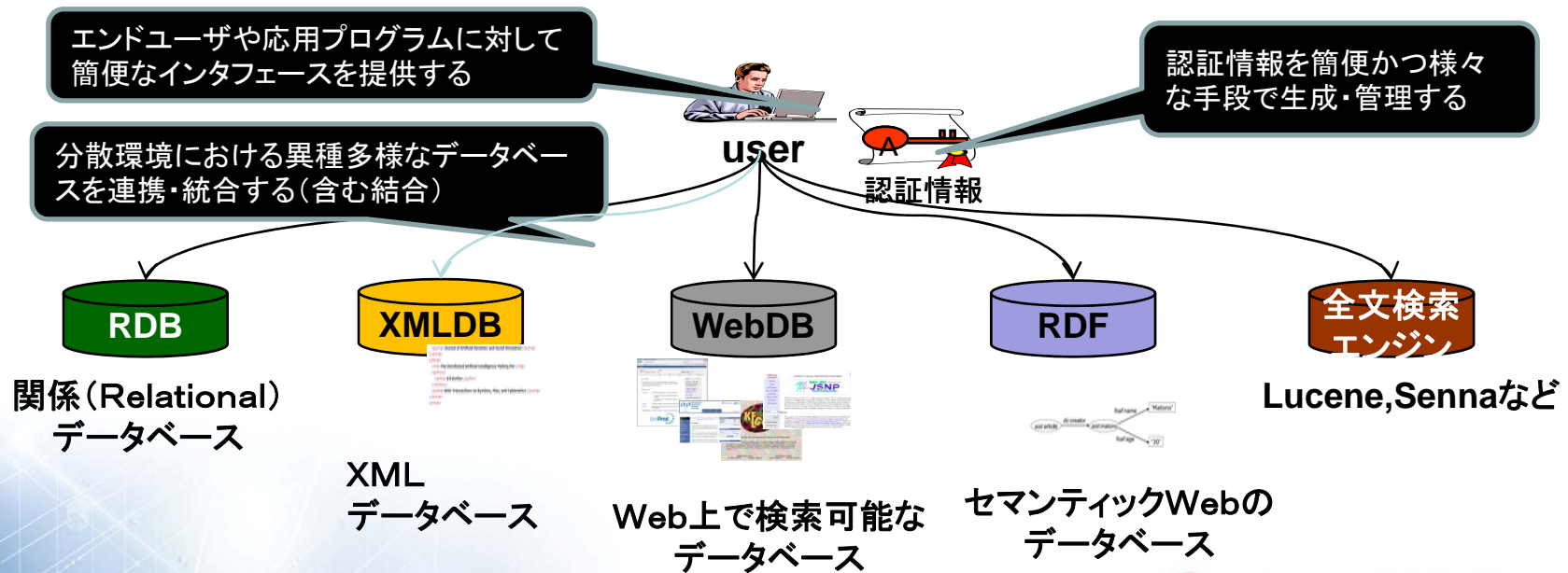
【特徴】

- ・ 分散統合を含むデータベース統合処理を可能とする。
- ・ WebDBやRDFに加え、全文検索エンジンもデータソースとして取り込むなど、異種多様なソースを結合する。

2. ユーザ認証情報生成管理技術の研究

【特徴】

- ・ データベースへのアクセス制御の基礎となる、ユーザの認証情報を簡便かつ様々な手段で生成・管理する。
 - ・ ユーザ名・パスワードによる、OpenIDをベースにする、手持ちのユーザ証明書を用いる、など。



Tsukuba-GAMA

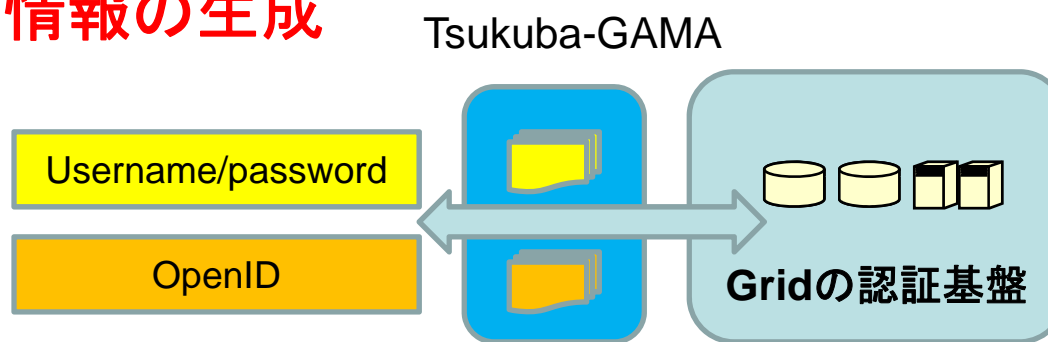
動機:

グリッドの認証で色々な既存のサービスを連携したい

- 今までの認証システムをできればそのまま使いたい、、
- 証明書ベースの認証システムは敷居が高い、、

多様なIDからグリッド認証情報の生成

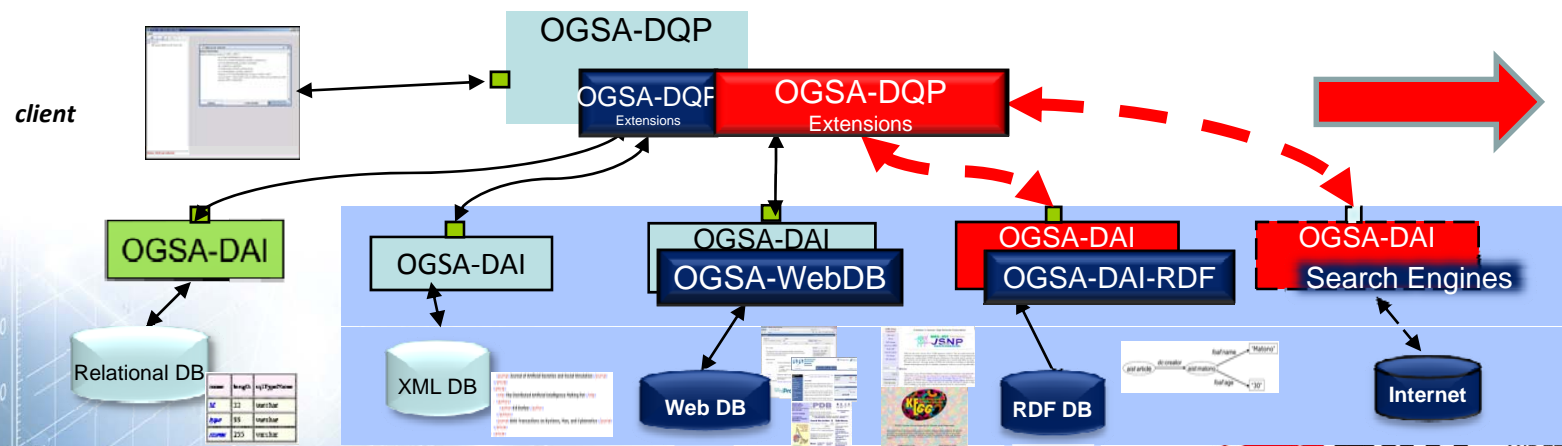
1. OpenID
 2. Shibboleth
 3. Kerberos
 4. Username/password
 5. PKI (X.509 certificates)
- プロトタイプを本年度に実装



それぞれの認証認可システムに対応した
動的な証明書の生成と
統合的な管理

OGSA-DAIに基づいた 異種データベースの連携

- OGSA-DAI: OMII-UKと産総研は協力関係。(AIST=OMII-UK Product/Area Liaison)
 - 関係DBとXMLDBをサポート:
異種データベースの扱いや分散処理に課題
 - グリッドの認証をサポート
既存のグリッドのツールや環境に容易に接続できる
- **異種データベースのアクセスの実現**
 - 従来型のWeb: **OGSA-WebDB**
 - インターネットに現在膨大に存在するWebデータベースを連携可能としている。
 - セマンティックWeb: **OGSA-DAI-RDF**
 - RDFデータベース(Semantic Webに不可欠なデータ形式)を扱えるようにしている。
- **統合・分散処理環境の実現**
 - **OGSA-DQP拡張:**
XMLDBやWebDBについて、最適化を含む異種分散を実現している。



データベース統合・連携

設計&実装の目標:

産総研開発のOGSA-WebDB/XMLextension/RDFを基礎 (Grid2008で発表)

として、以下の機能の取り込みを行う。

1. 全文検索エンジンの取り込み

- 属性やデータ型ごとの索引の構築と組み込み
- ポリゴン型など独自の索引アルゴリズムとランキング手法を含む

2. 問い合わせ処理の動的な(適合型の)最適化

- 中間結果の個数やネットワーク状況に応じた結合順序の順序変更など、処理手順の動的な変更
- サイトの故障など、予期しない事態に対する信頼性の実現

3. REST系呼び出しの仕様確定と実装

- OGC Webサービスなどとの連携の実現

OGSA-DQP Extension

広域のDBをWebサービス基盤上で宣言的に結合

- Integration of WebDB, relational, XML data

```
<authors>
  <author>
    <name>Steven Lynden</name>
    <field>Grid Computing</field>
  </author>
  <author>
    <name>...</name>
    <field>...</field>
  </author>
  ...
</authors>
```

column name	column type
author	string
title	string

The screenshot shows the CiteSeer.IST website with a search bar and a table of columns. The table has the following columns: column name, column type, recid, title, authors, snapshot, url, and string.

AUTHOR

XML

PUBLICATION

関係データベース

CITeseer

Webデータベース

OpenXML

```
select xauthor.name, publication.title,  
       citeseer.url, xauthor.field  
from publication, citeseer,  
     OpenXML(  
         author,  
         '//author',  
         '//name/text() name, //field/text() field`  
     ) as xauthor  
where publication.title=citeseer.title  
and xauthor.name=publication.author;
```

name	title	url	field
...
...

```
<authors>  
  <author>  
    <name>Steven Lynden</name>  
    <field>Grid Computing</field>  
  </author>  
  <author>  
    <name>...</name>  
    <field>...</field>  
  </author>  
  ...  
</authors>
```

```
<author>  
  <name>Steven Lynden</name>  
  <field>Grid Computing</field>  
</author>  
  
<author>  
  <name>...</name>  
  <field>...</field>  
</author>
```

Distributed Query Processing over Heterogeneous Databases

The screenshot displays the OGSA-DQP Client interface in User Mode. The main window shows a query plan for a query executed on 2007-11-13-04:06:22. The query expression is: `select XMLGen('<title>{'$c.title}</title>', author_author.name from webdbResource_csb as c, author_author where c.authors like '%paton%' and author_author.name=c.authors;`

The query plan visualizes the execution flow, starting with a `TABLE_SCAN` operation on the `author` table, followed by an `EXCHANGE` operation. The plan then branches into `HASH_LOOPS` and `OPERATION_CALL` operations, which connect to the `webdbResource_csb` table and the `XMLGen` function. The final output is a `PROJECT` operation that filters the results based on the `XMLGen` function's output.

The results window shows the following data:

XMLGen ()	author_author.name (VARCHAR)
<title>Summary</title>	Norman W. Paton
<title>Active Rules in Database Systems</title>	Norman W. Paton

The interface also displays a schema for the `webdbResource_csb` table, which includes columns for `name` (string), `recid` (string), `authors` (string), `title` (string), `booktitle` (string), and `year` (string).

Query Optimization over Multiple Heterogeneous Data Services

OGSA-WebDB Medical Application Search for Cancer Prevention



GUI environment

OGSA-WebDB

OGSA-DAI

The Web

JMS-JST from Japanese
JSNP
DATABASE

OMIM
Online Mendelian Inheritance in Man

PubMed
www.pubmed.gov

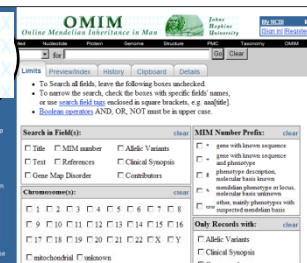
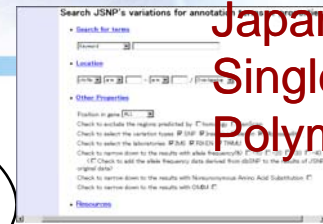
PharmGKB
The Pharmacogenetics and
Pharmacogenomics Knowledge Base

Japanese Database of
Single Nucleotide
Polymorphisms

Online Mendelian
Inheritance in Man

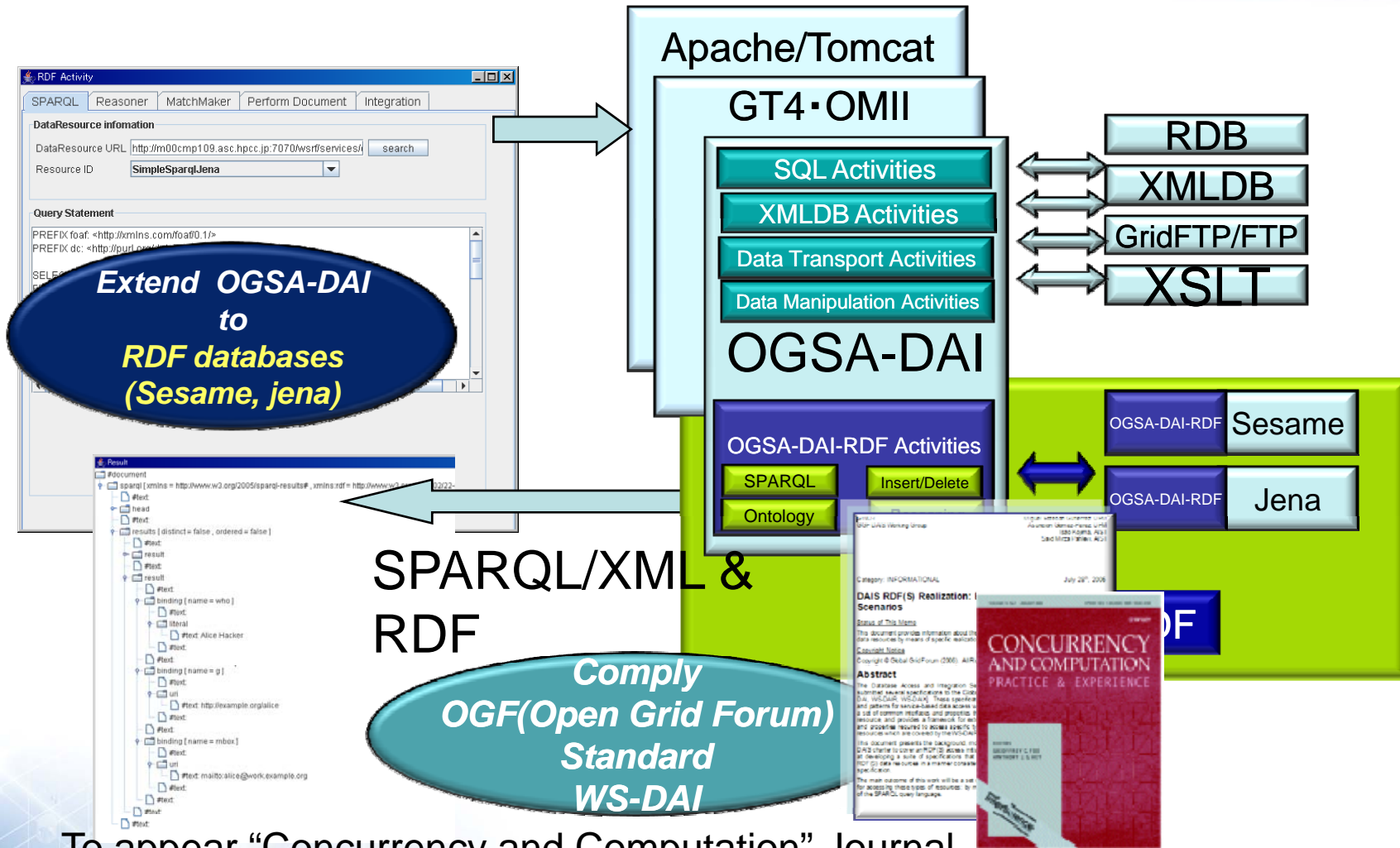
Medical
Bibliographic/
Citation Database

Pharmacogenetics &
Pharmacogenomics
Knowledge Base



Analyze relationships between
diseases, drugs & Side-Effects,
SNPs etc. from many aspects

OGSA-DAI-RDF: Service based Access to the Semantic Web(RDF) Databases



To appear "Concurrency and Computation" Journal

全文検索エンジンの連携

OGSA-DQP

Form検索インターフェイス

OGSA-DAI-FullText

OGSA-DAI-XML/We

OGSA-DAI-RDF

XML文書など

索引定義・
拡張の
フレームワーク

テキスト型索引

テキスト抽出

数値型索引

数値型データ抽出

R木型索引

ライブラリ
ランキング
ダ

独自型
データ抽出

各索引に対する条件の
AND/OR

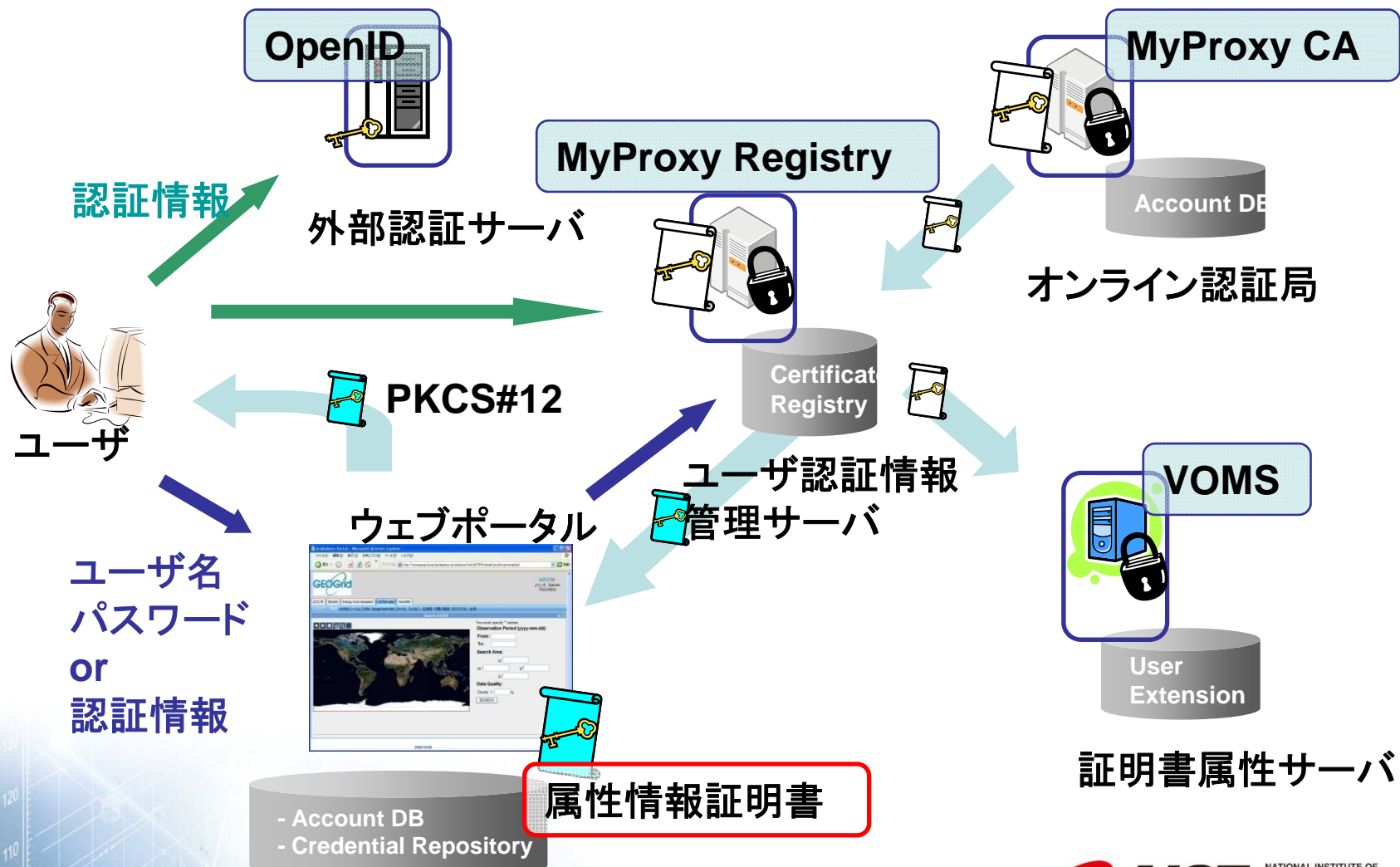
```
<?xml version="1.0" encoding="UTF-8" ?>
<getRecordsResponseType xmlns="http://www.opengis.net/cat/csw/2.0.2" xmlns:ns2="http://www.opengis.net/ogc"
xmlns:ns3="http://www.opengis.net/gml" xmlns:ns4="http://www.w3.org/1999/xlink" xmlns:ns5="http://www.opengis.net/ows"
xmlns:ns6="http://purl.org/dc/elements/1.1/" xmlns:ns7="http://purl.org/dc/terms" xmlns:ns8="http://www.isotc211.org/2005/gco"
xmlns:ns9="http://www.isotc211.org/2005/gmd" xmlns:ns10="http://www.opengis.net/gml/3.2"
xmlns:ns11="http://www.isotc211.org/2005/gts" xmlns:jmp20="http://zgate.gsi.go.jp/ch/jmp/">
<SearchResults numberOfRecordsReturned="100" numberOfRecordsMatched="98"> <jmp20:MD_Metadata>
<jmp20:identificationInfo>
<jmp20:MD_DataIdentification> <jmp20:citation>
<jmp20:title>20万分の1数値地質図幅「鹿児島」</jmp20:title> <jmp20:date>
<jmp20:date>2004-12-01</jmp20:date> <jmp20:dateType>002</jmp20:dateType>
</jmp20:date> </jmp20:citation>
<jmp20:abstract>20万分の1地質図幅は、既存の地質資料に基づいて、国土地理院発行の20万分の1地勢図のそれぞれの区画毎に編集
した地質図です。この数値地質図は、原資料である地質図を数値化したものです。</jmp20:abstract>
<jmp20:pointOfContact>
<jmp20:individualName>CD-ROM編集 総括：鹿野和彦・長谷川 功、表示システム作成： 鹿谷敏光・川畑 晶、データ編集：宮崎純
一 鹿谷敏光、ベクトルデータ作成：宮崎純一・中島和敏、ラスター画像データ作成：川畑 晶、メタデータ作成：渡辺和明、装丁：中島和敏
</jmp20:individualName>
<jmp20:organisationName>産業技術総合研究所地質調査総合センター</jmp20:organisationName> <jmp20:contactInfo>
<jmp20:onlineResource> <jmp20:linkage>http://www.gsj.jp/HomePageJP.html</jmp20:linkage> </jmp20:onlineResource>
</jmp20:contactInfo> <jmp20:role>009</jmp20:role> </jmp20:pointOfContact>
<jmp20:descriptiveKeywords> <jmp20:MD_Keywords>
<jmp20:keyword>20万分の1、ベクトル、数値地質図、鹿児島、中国西部、九州及び南西諸島、日本</jmp20:keyword>
</jmp20:MD_Keywords> </jmp20:descriptiveKeywords>
<jmp20:language> <jmp20:isoCode>jpn</jmp20:isoCode> </jmp20:language>
<jmp20:characterSet>023</jmp20:characterSet>
<jmp20:topicCategory>008</jmp20:topicCategory> <jmp20:extent> <jmp20:geographicElement>
<jmp20:EX_CoordinateBoundingBox> <jmp20:extentReferenceSystem> <jmp20:authority> <jmp20:title>測量法</jmp20:title>
<jmp20:date> <jmp20:date>1960-07-01</jmp20:date> <jmp20:dateType>003</jmp20:dateType> </jmp20:date> </jmp20:authority>
<jmp20:code>TD / (B, L)</jmp20:code> </jmp20:extentReferenceSystem>
<jmp20:westBoundCoordinate>130</jmp20:westBoundCoordinate>
<jmp20:eastBoundCoordinate>131</jmp20:eastBoundCoordinate>
<jmp20:southBoundCoordinate>31.333333</jmp20:southBoundCoordinate>
<jmp20:northBoundCoordinate>32</jmp20:northBoundCoordinate>
</jmp20:EX_CoordinateBoundingBox>
```

(以下略)



NATIONAL INSTITUTE OF
ADVANCED INDUSTRIAL SCIENCE
AND TECHNOLOGY (AIST)

Tsukuba-GAMA: プロトタイプ評価と本システムの実装



Login Flow for OpenID user (cont'd)

ログイン

ユーザー名

パスワード

OpenID Login

OpenID.ne.jp

1回の登録と1つのアカウントでOpenIDを認証するすべてのサイトに、すぐ登録することができます。

OpenID.ne.jp

1回の登録と1つのアカウントでOpenIDを認証するすべてのサイトに、すぐ登録することができます。

<http://naotaka.openid.ne.jp/>

[ページ](#) | [アカウント管理](#) | [ログアウト](#)

OpenID 確認

**Accept Once
for GEO Grid Portal**

現在

<https://gfm41.apgrid.org:>

あなたのOpenid (<http://naot>

あなたの個人情報と一致しているという証明を要求しています。

https://gfm41.apgrid.org:13443/gridsphere/gridsphere?cid=openidlogin&gs_action=verifyRequest に

証明するための必要な個人情報を追加チェックし、認証してください。

ifyRequestで、



Login Flow for OpenID user (cont'd)

ログイン

ユーザー名

パスワード

OpenID Login

OpenID login VOMS Credential Portlet

Identifier

VOMS Credential Portlet

```
--- Credential ---
subject : /O=Grid/OU=GlobusTest/OU=simpleCA-gfm37.apgrid.org/CN=Naotaka Yamamoto
time left : 11 hours 59 minutes 52 seconds.
--- VOMS Extension Information ---
attribute : /testvo.geogrid.org/Role=NULL/Capability=NULL
```

--- Credential ---

subject : /O=Grid/OU=GlobusTest/OU=simpleCA-gfm37.apgrid.org/CN=Naotaka Yamamoto
time left : 11 hours 59 minutes 52 seconds.

--- VOMS Extension Information ---

attribute : attribute : /testvo.geogrid.org/Role=NULL/Capability=NULL

no ASTER attribute

まとめ

- データベース連携の研究開発
 - 異種のデータベースシステムを統合
 - RDB,XML,RDF,Web,,etc.
 - OGSA-DAIベースのインターフェイス
 - 既存のグリッド認証のシステムと容易に接続可
 - 異なった管理ポリシーのDBを連携
 - グリッドの認証と連携できるシステム
 - 他の応用にも適用可能

データベース連携技術に関する研究 (3/3)

年次計画

