An Intelligent Data Broker for Virtual Integration of Heterogeneous Distributed Meteorological Databases

May 10, 2005

Yamakawa A., T. Kiura & S. Ninomiya

National Agricultural Research Center, National Agriculture and Bio-oriented Research Organization
## Potential for Data Sharing Between DSS for Agriculture

<table>
<thead>
<tr>
<th>Decisions (Clients)</th>
<th>Data Needed</th>
<th>Weather Data</th>
<th>Crop details</th>
<th>Soils</th>
<th>Topography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variety selection</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>To dam?</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Optimal land use</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Spray for disease</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Irrigate or not?</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Manage flooding</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Agricultural applications and databases

• Agricultural applications need several different types of data sets

• Databases needed for the applications are usually maintained and managed by different organizations and are located in different places
Concept of Agricultural Grid

- Neural Network
- Growth Model 1
- Weather Data 1
- Field Data Monitoring
- Weather Data 2
- Growth Model 2
- Farm Management
- Regression Model
- Meta Database
- Data Broker
- User who needs Decision

The Internet
Advantages of Grid

• More powerful DSS by dynamic and flexible integration of distributed resources
• No duplication of databases and programs
• Easy updates of databases and programs
• Highly efficient utilization of resources
• Reduction of development and maintenance cost
Challenges to Realize Grid

• Database heterogeneity
  – Those databases are heterogeneous in access method, data formats, available items, time resolution etc.

• Internationalization
• Security
• QoS
• Efficient field data collection
Example of Heterogeneity of Meteorological Database

• Item
  – Air Temperature, Wind, Rain, Radiation, etc.

• Sub Item
  – Rain Amount, Rain Intensity, Rainy Days, etc.

• Time Resolution
  – Sub Hourly, Hourly, Daily, Monthly, etc.

• Summarization
  – Mean, Maximum, Minimum, Total, etc.

• Equipments
Goal of Our Project

- To provide an environment where heterogeneous distributed databases can be flexibly, seamlessly and virtually integrated for Agricultural Grid
Our Initial Challenge

- Some Data Brokers
  - Meteorological DBs
    - MetBroker (15DB, >13000 stations)
  - Map DBs
    - ChizuBroker (3DB, Japan, NZ, World)
  - Digital Elevation DBS
    - DEMBroker (2DB, Japan 50m, World 1Km)
  - Soil DBs
    - SoilBroker

- [http://www.agmodel.net/](http://www.agmodel.net/)
## Coverage of MetBroker

<table>
<thead>
<tr>
<th>Country</th>
<th>Data Source</th>
<th>Start Year</th>
<th>End Year</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>MAFF’s database of AMeDAS data</td>
<td>1976</td>
<td>1997</td>
<td>hourly</td>
</tr>
<tr>
<td>Japan</td>
<td>Wakayama Rainfall data</td>
<td>1976</td>
<td>2001</td>
<td>hourly</td>
</tr>
<tr>
<td>USA</td>
<td>Oregon Integrated Plant Protection Center (NorthWest)</td>
<td>1996</td>
<td>1996</td>
<td>daily</td>
</tr>
<tr>
<td>USA</td>
<td>Long Term Ecological Research Network (ClimDB)</td>
<td>1964</td>
<td>1964</td>
<td>daily</td>
</tr>
<tr>
<td>USA</td>
<td>Georgia Automated Environmental Monitoring Network</td>
<td>1997</td>
<td>1997</td>
<td>daily</td>
</tr>
<tr>
<td>NOAA/WMO</td>
<td>Globa; Surface Summary of Day</td>
<td>1964</td>
<td>1964</td>
<td>daily</td>
</tr>
<tr>
<td>Norway</td>
<td>Planteforsk Crop Research Institute</td>
<td>1987</td>
<td>1987</td>
<td>hourly</td>
</tr>
<tr>
<td>New Zealand</td>
<td>HortPlus Ltd</td>
<td>1996</td>
<td>1996</td>
<td>hourly</td>
</tr>
<tr>
<td>USA</td>
<td>Florida Automated Weather Network</td>
<td>1996</td>
<td>1996</td>
<td>hourly</td>
</tr>
<tr>
<td>Japan</td>
<td>Kanagawa Prefecture Agricultural Research Station Network</td>
<td>1998</td>
<td>1998</td>
<td>hourly</td>
</tr>
<tr>
<td>South Africa</td>
<td>South African Sugar Association network</td>
<td>1997</td>
<td>1997</td>
<td>daily</td>
</tr>
<tr>
<td>Korea</td>
<td>Seoul National University Plant Disease and Epidemiology Lab</td>
<td>1993</td>
<td>1993</td>
<td>hourly</td>
</tr>
<tr>
<td>Japan</td>
<td>Hokkaido Memuro MAMEDAS</td>
<td>2000</td>
<td>2000</td>
<td>hourly</td>
</tr>
<tr>
<td>Japan</td>
<td>Chiba and Tottori Prefectural Research Stations</td>
<td>1986</td>
<td>1986</td>
<td>hourly</td>
</tr>
<tr>
<td>Japan</td>
<td>FieldServer project</td>
<td>2002</td>
<td>2002</td>
<td>hourly</td>
</tr>
<tr>
<td>New Zealand, Pacific Antarctica</td>
<td>National Climate Database</td>
<td>1853</td>
<td>1853</td>
<td>hourly</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>Horticulture Research International</td>
<td>1919</td>
<td>1919</td>
<td>daily</td>
</tr>
<tr>
<td>Taiwan</td>
<td>Taiwan Ecological Research Network</td>
<td>1995</td>
<td>1995</td>
<td>daily</td>
</tr>
<tr>
<td>USA</td>
<td>Washington State University Public Agricultural Weather System</td>
<td>1987</td>
<td>1987</td>
<td>hourly</td>
</tr>
</tbody>
</table>
Present Coverage
Database Broker Service

- Data Summarization
- Data Secondary Processing
- Data Brokage
- Data Standardization
- Whereabout
- Data Request translated to each DB
- Data acquisition

- Meta Database
  Where, How to use Data contents

- Database Driver (Wrapper Applications)
  - A-DB
  - B-DB
  - C-DB
  - D-DB

- Client
  - Data Request
  - Standardized Data
Existing Solutions

Web service can hide heterogeneity of access method by using XML technology
Limitations of Conventional Approach

- Needed to create a wrapper program which translates the meanings of the terms used in each database.
  - Web service can’t hide semantic heterogeneity
  - Hard to add new observation items
Problem of Static Translation

Any Daily Air Temperature is acceptable

Daily Mean Air Temperature
Daily Max Air Temperature
Daily Min Air Temperature

Only Daily Air Temperature based on 1-minute data is acceptable

For client A, Air Temperatures of DB1 and DB2 are regarded as the same item, but for client B, they are not. It is very hard for conventional solution to meet such requests.
Solutions by Ontology

• RDF/OWL technology can solve the limitations which the conventional solution has
• Ontology can realize seamlessly and virtually integrated database
• Intelligent Data Broker is a challenge to provide clients with access to this virtually integrated database
Intelligent Data Broker

• System Overview
• System Implementation
• How Intelligent Broker works?
  – Single station access
  – Spatial access
  – Access with data summarization
• Discussion
System Overview

**Decision-Making Support Services**
- Operational Products
- Simulation Models
- Detailed Digital Forecast

**Inference Engine**

**1. Register**
**2. Request**
**3. Request metadata**
**4. Request data**

**Metadata database**
- Item Definition OWL
- Station metadata RDF

**Meteorological databases**

**DB Wrapper**

**Broker**

**NARO**

**NARC**
System Implementation

- A prototype version
- Coded in Java language
- Main part of the inference engine
  - Jena2
- Composed of one basic vocabulary file and as many item definition files and station metadata files as the number of targeted databases
# Roles Of the RDF/OWL files

<table>
<thead>
<tr>
<th>Name</th>
<th>File type</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic vocabulary</td>
<td>OWL</td>
<td>• All standard weather items</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Vocabulary to describe weather stations</td>
</tr>
<tr>
<td>Item definition</td>
<td>OWL</td>
<td>Local vocabulary that is used in each database</td>
</tr>
<tr>
<td>Station metadata</td>
<td>RDF</td>
<td>Description about all the weather stations included in a particular database</td>
</tr>
</tbody>
</table>
How Intelligent Broker works?

• **Single station access**
  – Clients request some weather items from a specified weather station

• **Spatial access**
  – Clients request some weather items from the weather stations located in a specified geographical area

• **Access with data summarization**
  – Clients allow Intelligent Data Broker to summarize data
Single station access

1. Request
   Resolution → Daily
   Summary → Mean
   Item → Air Temperature

2. Find the basic vocabulary that is equivalent to the request

3. Find the local item equivalent to “DailyMeanAirTemperature”

4. Pass the local item “Air”

5. Query the database using the local item “Air”
Spatial Access

Client

Broker

Temperature
Rainfall
Spatial Access

- Complements of sparse observations
Spatial access

1. Request
   NW → latitude 35
   →longitude 134
   SE → latitude 25
   →longitude 154
   Item → Daily Mean Air Temperature

3. Find the local items equivalent to “DailyMeanAirTemperature” for each database that includes the stations in the requested area

2. Find the stations located in the requested geographical area

4. Pass the local items

5. Query the databases using the local items
Access with summarization

1. Request
   Resolution → Daily
   Summary → Mean
   Item → Air Temperature
   Summarization → allowed

2. Find the basic item that has the same or shorter time resolution than the request

3. Find the local item that is compatible with the basic item

4. Pass the local item

5. Query the database using the local item

6. Summarize the obtained data to requested time resolution

Basic Vocabulary
<owl:Class rdf:ID="DailyMeanAirTemperature"/>
<owl:Class rdf:ID="HourlySampleAirTemperature"/>

Database: Item definition
<meta HourlySampleAirTemperature rdf:ID="ame_time.temperature">
</meta HourlySampleAirTemperature>
Sample of Basic Vocabulary

"DailyMaxAirTemperature" is a subclass of "MaxAirTemperature"

"DailyMaxAirTemperature" is recognized as maximum and daily data

Sample file:
http://www.agmodel.org/MetBroker.owl
## Available Basic Items (1)

<table>
<thead>
<tr>
<th>Item</th>
<th>Sub-Item</th>
<th>Number of Sub-sub-Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>AirTemperature</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>AtmosphericPressure</td>
<td>SeaLevelPressure</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>StationPressure</td>
<td>9</td>
</tr>
<tr>
<td>Cloud</td>
<td>CloudAmount</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>CloudForm</td>
<td>9</td>
</tr>
<tr>
<td>Evaporation</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>GasConcentration</td>
<td>CO2</td>
<td>2</td>
</tr>
<tr>
<td>Humidity</td>
<td>AbsoluteHumidity</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>DewPointTemperature</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>RelativeHumidity</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>SteamPressure</td>
<td>4</td>
</tr>
<tr>
<td>LeafWetness</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Pollutant</td>
<td>Aerosol</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Dust</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Pollen</td>
<td>0</td>
</tr>
</tbody>
</table>
## Available Basic Items (2)

<table>
<thead>
<tr>
<th>Item</th>
<th>Sub-Item</th>
<th>Number of Sub-sub-Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation</td>
<td>LongwaveRadiation</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>SolarRadiation</td>
<td>29</td>
</tr>
<tr>
<td>RainFall</td>
<td>RainAmount</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>RainDuration</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>RainIntensity</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>RainStatus</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>RainyDays</td>
<td>0</td>
</tr>
<tr>
<td>SnowFall</td>
<td>SnowAmount</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>SnowDepth</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>SnowDiff</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>SnowType</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SnowyDays</td>
<td>0</td>
</tr>
<tr>
<td>SoilMoisture</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>SoilTemperature</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Sunshine</td>
<td>PossibleDuration</td>
<td>0</td>
</tr>
</tbody>
</table>
# Available Basic Items (3)

<table>
<thead>
<tr>
<th>Item</th>
<th>Sub-Item</th>
<th>Number of Sub-sub-Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunshine</td>
<td>SunshineDuration</td>
<td>3</td>
</tr>
<tr>
<td>Weather</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Wind</td>
<td>HorizontalWind</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>VerticalWind</td>
<td>2</td>
</tr>
</tbody>
</table>
Sample of Item Definition

Local item name

```
<met:DailyMaxAirTemperature rdf:ID="ame_day.temp_max">
<met:summaryKind rdf:resource=
"http://www.agmodel.org/MetBroker.owl#DailyMaximumOfSampleEvery10Minutes"/>
</met:DailyMaxAirTemperature>

<met:HourlySampleAirTemperature rdf:ID="ame_time.temperature">
<met:summaryKind rdf:resource=
"http://www.agmodel.org/MetBroker.owl#SampleOnTheHour"/>
</met:HourlySampleAirTemperature>
```

“ame_day.temp_max” is recognized as maximum and daily data based on every 10 minutes data

A sample file is available on [http://www.agmodel.org/Aclima.owl](http://www.agmodel.org/Aclima.owl)
Sample of Station Metadata

<met:MetStation rdf:ID="01">
  <rdfs:label xml:lang="en">
    Ag. Res. Inst. Representative Observation Station
  </rdfs:label>
  <met:alt>64.0</met:alt>
  <met:log>139.2874298095703</met:log>
  <met:lat>35.34185791015625</met:lat>
  <met:belongTo>http://www.agmodel.org/Kanagawa.rdf</met:belongTo>
  <met:metCatalog>
    <met:MetCatalog>
      <met:metElement>&kngw;#DailyAverageWindVelocity</met:metElement>
      <met:catalogStart>1995-12-31T15:00:00+0000</met:catalogStart>
      <met:measurementHeight>2.0</met:measurementHeight>
      <met:MetCatalog>
    </met:MetCatalog>
  </met:metCatalog>
  .......
</met:MetStation>
Discussion

We realized two advantages of the RDF/OWL technology

• Secure scalability in building a large scale database by integrating heterogeneous databases
• Support to develop a system flexibly adjustable to various clients request
Future Plan

• Make this prototype more practical in agricultural decision making

• Bind various databases, e.g. crop growth databases and soil databases into our system in addition to meteorological databases
Thank you for your attention

http://www.agmodel.org/

Sample file:

http://www.agmodel.org/MetBroker.owl