

## *Applying Information Grid for Agricultural Applications*

Naiyana Sahavechaphan, Kamron Aroonrua, Sornthep Vannrat and Asanee Kawtrakul

National Electronics and Computer Technology Center, Thailand  
112 Thailand Science Park, Pahon Yothin Rd  
Klong 1, Klong Luang, Pathumthani 12120 Thailand  
naiyana.sah@nectec.or.th, kamron.aro@nectec.or.th

### ***Abstract***

Information is considered to be an essential resource towards the achievement of a specified goal. Today, many agricultural applications have been developed to facilitate farmers' decision-making and problem-solving by providing them the desired information or knowledge through computers, mobile phones and pocket PCs. Their underlying information is typically from different repositories either within the same organization or across organizations. Accordingly, there is the diversity of information, making the development of such applications themselves time-consuming and expensive. To simplify such application development, many approaches such as OGSA-DQP (Alpdemir et al. 2003), GDMS (Woehrer et al. 2004) and OGSA-WebDB (Pahlevi et al. 2004) have been developed. While these approaches take steps into the right direction, they still have some drawbacks: *limitation* to in-house repositories, *inefficiency* to any changes in repositories' structures and any to-be queried repositories, *inflexibility* to transient repositories and *non-scalability* to a large number of repositories.

To address the above drawbacks, in this paper, we thus propose *Information Grid* as a comprehensive framework that facilitates the development of distributed applications by providing them the *virtualization* of distributed and heterogeneous repositories. Specifically, it has taken into account not only information access and unification but also information query and discovery in a *distributed* manner. Information standard is also considered. Its capacity is to deliver applications that require information from distributed and heterogeneous repositories on time and within budget, and to stay still irrespective of any changes in repositories' structures as well as the addition and removal of repositories. Accordingly, our on-going work is thus to introduce Information Grid as an underlying framework for developing distributed applications in agriculture domain with the expectation that there would be various repositories in agricultural and related domains *ready* to be used to reflect farmers' needs.

**Keywords:** Information Unification, Information Access, Information Discovery, Information Integration

## 1. Introduction

Information is considered to be an essential resource towards the achievement of a specified goal (see Figure 1). For example, farmers' goal "*Making a satisfactory income*" (Robinson et al.) can be reached through the application of various information such as fertilizer utilization, disease control and market price. Essentially, information inferences knowledge that enables farmers to efficiently perform decision-making and problem-solving for their agricultural activities and hence to finally meet their goals. As defined by Trnková et al. (2004), knowledge can be: *orientation* knowledge (know-what) – what the topic is about; *reference* knowledge (know-where) – where to find additional or new information; *explanation* knowledge (know-why) – why something is the way it is; and *action* knowledge (know-how) – how to accomplish something.

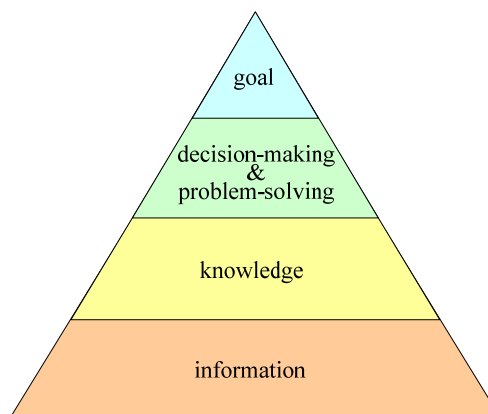


Figure 1: The relationship of information towards goal

Today, many agricultural applications have been developed to facilitate farmers' decision-making and problem-solving by providing them the desired information or knowledge through computers, mobile phones and pocket PCs. Examples of these applications include *Fertilizer Expert System* (Kawtrakul et al. 2008) – to answer farmers an appropriate fertilizer with respect to plant, soil type and current nutrient in soil; and *Disease/Pest Warning System* – to feed potential farmers news about disease/pest outbreak as well as the way to control and treat disease/pest. As illustrated in Figure 2, the underlying information of these applications, termed as *distributed applications*<sup>1</sup>, is typically from different repositories either within the same organization or across organizations. Accordingly, there is the diversity of information which encompasses: (i) *format* – information can be differently represented in term of database, XML, HTML and even text or image files; (ii) *structure (schema)* – information can be described based on different element sets that may define elements using different definitions, labels, types and hierarchies; (iii) *granularity level* – information can be instantiated at different compound levels. For example, region or province can be used to define cropping area; and (iv) *measurement unit/method* – information can be measured in different units and via different methods. Here, to simplify the development of these distributed applications, there is need for *information access* across multiple repositories irrespective of their formats as well as *information unification* across heterogeneous repositories that are different in term of the above four aspects. Clearly, the process of both information access and unification against

<sup>1</sup> A distributed application, in this paper, is made up of an application itself and distributed repositories that run in separate runtime environments and are owned by different owners.

heterogeneous repositories is time-consuming and expensive, affecting the delivery of applications in term of time and budget.

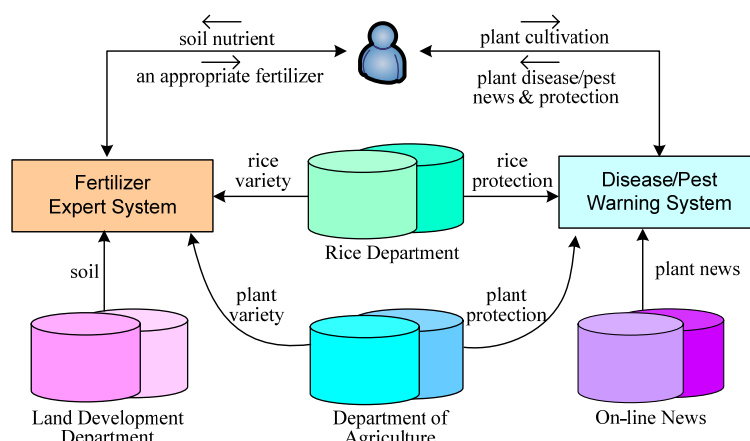


Figure 2: The Utilization of Information from Heterogeneous Repositories

Many approaches such as OGSA-DQP (Alpdemir et al. 2003), GDMS (Woehrer et al. 2004) and OGSA-WebDB (Pahlevi et al. 2004) have been developed to simplify such application development. In particular, they facilitate information access across multiple repositories wherein repository handles are varied by approaches<sup>2</sup>, while GDMS alone has taken into account information unification. While these approaches take steps into the right direction, they still have some drawbacks. First, the knowledge of data structures in repositories and a centralized effort is required, making these approaches potentially *limited* to in-house repositories. Second, information query is tightly coupled with underlying repositories, making the query itself *inefficient* to handle any changes of repositories' structures and to cover any repositories as preferred as well as *inflexible* to transient repositories and. Third, query transformation is typically done within a single component, making them *non-scalable* to a large number of repositories.

To address the above drawbacks, in this paper, we thus propose *Information Grid* as a comprehensive framework that facilitates the development of distributed applications by providing them the *virtualization* of distributed and heterogeneous repositories. Specifically, it has taken into account not only information access and unification but also information query and discovery in a *distributed* manner. Information standard is also considered. Its applicability has been evaluated, at the laboratory level, for implementing several distributed applications such as Thai-Research Web application and one-search library system via Excel. The result showed its capacity especially in delivering distributed applications on time and within the budget. In addition, these distributed applications do not need to be modified to reflect any changes in repositories' structures (as long as information standard stays still) as well as the addition and removal of repositories. However, it has a minor disadvantage in that its instructions must be strictly followed in order to successfully utilize it. Accordingly, our on-going work is thus to introduce Information Grid as an underlying framework for developing distributed applications in agriculture domain with the expectation that there would

<sup>2</sup> OGSA-DQP deals with relational databases. GDMS handles relational databases, XML databases and comma-separated value (csv) files. OGSA-WebDB supports data sources available on line on the Web.

be various repositories in agricultural and related domains *ready* to be used to serve farmers' needs.

This paper is organized as follow. Section 2 defines the key goals of Information Grid. Its details are described in Section 3. Section 4 illustrates the steps needed to apply Information Grid for developing distributed applications in agricultural domain. Section 5 concludes the paper.

## 2. The Key Goals of Information Grid

Recall that each distributed application requires information from distributed and heterogeneous repositories. It is thus likely that each distributed application may need information from 20 or more repositories. On the other hand, each repository alone may be needed by one or more distributed applications. Therefore, it should provide standard protocol that allows any distributed applications to access its underlying information, once its accessibility is previously made available. Accordingly, there would be plenty of repositories available for distributed applications. To facilitate the development of distributed applications that share repositories, we have thus proposed *Information Grid* as a middleware between distributed applications and heterogeneous repositories. In particular, its goals is defined as following:

- (i) *general* – in that it facilitates an access of information from multiple repositories irrespective of repository formats and repository owners;
- (ii) *efficient* – in that it allows information to be easily acquired from repositories with either specified repositories to get desired information, or with any repositories to obtain potentially rich information;
- (iii) *flexible* – in that it enables the addition and removal of repositories at all time without affecting information query; and
- (iv) *scalable* – in that it supports information query from a large number of repositories without affecting the overall system performance. In addition, it supports the transport of large-scale information.

## 3. The Conceptual Model of Information Grid

Figure 3 illustrates the conceptual model of Information Grid. Here, Information Grid consists of five essential components: *Standard Metadata Schema*, *Information Service*, *Information Broker*, *Metadata Schema Service* and *GiSTool* (Generic Information Service Tool). Each component has played an essential role to reflect the achievement of the above goals as indicated in Table 1. (Its development is based on XML, Web Services and P2P technologies)

Information Grid's Components/Tools	Key Goals			
	General	Efficient	Flexible	Scalable
Standard Metadata Schema	√			
Information Service	√			√
Information Broker		√	√	√
Metadata Schema Service	√			
GiSTool	√			

Table 1: The Relationship between Information Grid's Components/Tools and Its Key Goal

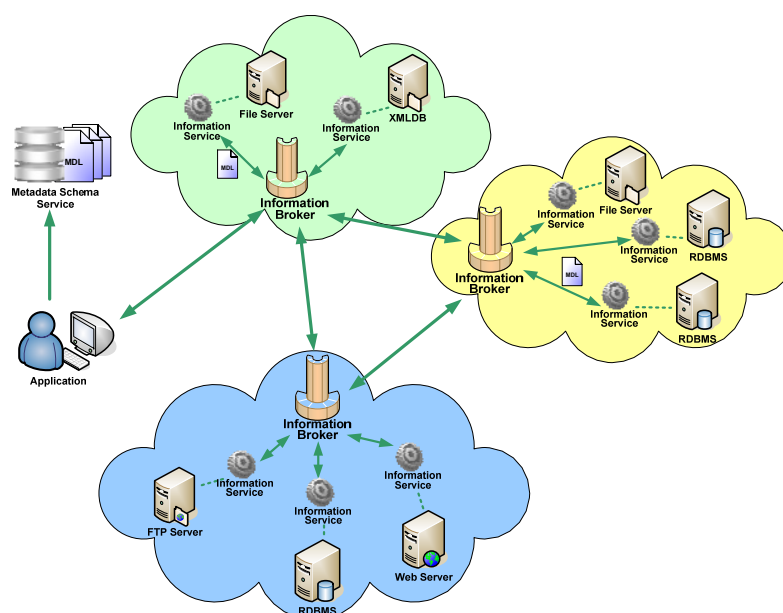


Figure 3: The Conceptual Model of Information Grid

### 3.1 Standard Metadata Schema

Standard Metadata Schema is a schema that defines the standard element set for describing specified information. For example, AgMES, defined by FAO, is a standard schema for describing agricultural document-like object. Essentially, the purpose of standard metadata schemas is to enable information-sharing and interoperability across repositories.

In Information Grid, standard metadata schemas are defined in term of XML schema and are available via Metadata Schema Service (see Section 3.4). Moreover, each standard metadata schema is taken into account while querying for information across repositories (see Section 3.3), or publishing information of a specified repository via Information Service (see Section 3.2).

### 3.2 Information Service

Information Service is a Web Service that is responsible for transforming the query and result back and forth between Information Broker and its underlying repository. In particular, it receives a *standard query* from Information Broker (see Section 3.3) that it connects to. It then transforms a standard query to a query understood by its underlying repository. Next, it submits such query to its underlying repository and waits for the result. On the return of result from its underlying repository, it transforms such result into a *standard result*. Finally, it returns a standard result to Information Broker. It should be noted that a standard query is a query that is defined as per a given standard metadata schema. Similarly, a standard result is a result that describes information in term of standard metadata schema.

Clearly, there is a one-to-one relationship between information service and repository. Information access and unification are thus accomplished in a distributed manner. Accordingly, Information Grid is scalable to a large number of repositories.

### **3.3 Information Broker**

Information Broker is a Web Service that is responsible for discovering and integrating potential information from any repositories, connecting to Information Grid, as per a given query. Particularly, an initial Information Broker  $iB_i$  receives a standard query from a client. It then distributes a query to its potential Information Services as well as other Information Brokers  $iB_j$  connecting to it. Other Information Brokers  $iB_j$  will recursively perform in a similar manner. Next, each Information Broker  $iB_i$  recursively waits for results on return from its potential Information Services and other Information Brokers  $iB_j$ . Once the initial Information Broker receives all standard results, it integrates and returns results to a client. In particular, the result can be sent to a client in synchronous, asynchronous or file transfer manner as per request.

Specifically, each information service should be registered to a specified information broker along with standard metadata schemas to indicate that its information can be sharing in term of these schemas. Upon the register of information services, each information broker updates its lookup registry. Additionally, information brokers are connected to each others based on our P2P system, WinChimer (Phengsuwan et al. 2008). An information broker thus propagates a query message to other information brokers with respect to this P2P system.

### **3.4 Metadata Schema Service**

Metadata Schema Service (Lousangfa et al. 2008) is a Metadata Schema Portal that is responsible for (i) enabling organizations or committees<sup>3</sup> to define, register and publish standard metadata schemas. This eases their discovery via a single portal; and (ii) allowing developers, who design information repository to support information-sharing and interoperability, to choose the desired metadata schema as per information in their repositories. Moreover, it eases information seekers to discover for information across repositories by specifying conditions based on the standard metadata schemas of choice.

### **3.5 Generic Information Service Tool (GiSTool)**

Generic Information Service Tool (GiSTool) is a tool that facilitates the development of an information service as per a standard metadata schema and a specified repository. In general, it allows developers to perform mapping between elements defined in a standard metadata schema and elements defined in a repository, and then automatically develop an information service based on such mapping information. As Information Grid aims for information in any repository formats, GiSTool thus consists of (i) GiSTool-RDB for information available in relational databases; (ii) GiSTool-WebDB for information available on-line on the internet; (iii) GiSTool-File for information available in text or even image files. Currently, GiSTool-RDB has been developed with its basic functionalities, while the rest is under development.

## **4. Agricultural Applications with Information Grid**

There are three essential steps to develop agricultural applications based on Information Grid. Their details are described in the following sections.

---

<sup>3</sup> To reflect information-sharing and interoperability, many organizations and committees have developed standard metadata schema with respect to their need and usage. These schemas, however, are distributed throughout different web sites.

**Step 1: Specify or define a standard metadata schema.** In this step, a developer needs to choose the desired metadata schema, via metadata schema service (see Section 3.4), applicable to the underlying information in his repository. Otherwise, he needs to define new standard metadata schema and registers it into metadata schema service such that any developers can recognize it to facilitate information-sharing and interoperability. In particular, each metadata schema is identified by its URI (Uniform Resource Identification).

**Step 2: Make repositories available via Information Grid network.** In this step, a developer needs to make all needed repositories available via Information Grid network. As per each repository, this can be done by (i) using an appropriate GiSTool to develop the corresponding information service with respect to a chosen metadata schema; (ii) deploying such information service on the desired server using Information Service Installation Kit; and (iii) registering such information service to either a preferred information broker available in the network or his own information broker that is newly deployed on a specified server and is registered to the Information Grid network. It should be noted that this step can be omitted if all needed repositories are available in Information Grid.

**Step 3: Develop an application to retrieve information from different repositories.** In this step, as per the aspect of information retrieval, a developer needs to write code to connect to and retrieve information from Information Grid. The sample code is as shown in Figure 4.

```
// 1. specify URI of metadata schema the describes the desired information
String URI = "http://.../project";

// 2. connect to Information Grid via a specified Information Broker's access point
Endpoint ep = new Endpoint("http://127.0.1.1/.../InformationBrokerFactoryService");
BrokerCommandFactory factory = new BrokerCommandFactory(ep);

// 3. create a command object for such Information Broker and specify it a query
Command cmd = factory.getBrokerCommand();
cmd.start();
cmd.setInformationQuery (new Query(URI).WHERE("Project/name>").LIKE("rice").AND
("Project/name>").LIKE("disease">));

// 4. retrieve information in which the result would be in term of corresponding information object
as
// per a specified URI
cmd.executeQuery();
String result = cmd.getResult ();
Project[] project = MetadataSchemaAnalyzer.convert(URI, result);
```

Figure 4: Code Example for Connecting to and Retrieving Information from Information Grid

## 5. Conclusion

In this paper, we have introduced Information Grid as a framework of choice to facilitate the development of agricultural applications that requires information from distributed and heterogeneous repositories. Essentially, it is in the middle of applications and repositories. Its key goals is to virtually provide information access to any repositories irrespective of their formats and owners; to query information from any repositories as preferred irrespective of any changes of repositories' structures; to adapt to the addition and removal of repositories; and to

handle a large number of repositories. Its development is based on XML, Web Services and P2P technologies.

According to Information Grid's capacity, our on-going work is thus to introduce Information Grid as an underlying framework for developing distributed applications in agriculture domain with the expectation that there would be various repositories in agricultural and related domains *ready* to be used to reflect farmers' needs.

## 6. References

- Alpdemir, M. N., Mukherjee, A., Gounaris, A., Paton, W. N., Watson, P. Fernandes, A. A., and Smith, J. (2003) OGSA-DQP: A Service-Based Distributed Query Processor For The Grid. Proceedings of the Second e-Science All Hands Meeting.
- Lousangfa, P., Sahavechaphan, N., Buranarach, M., Vannrat, S. and Kawtrakul, A. (2008) Survey and Modeling Metadata Schema Relationship in Agriculture Domain for Better Metadata Schema Service. Proceeding of World Conference on Agriculture Information and IT: *IAALD AFITA and WCCA. Tokyo, Japan. August 24-27, 2008.*
- Kawtrakul, A., (2008) A Smart Mobilized Fertilizing Expert System: 123 Personalized Fertilizer. Proceeding of World Conference on Agriculture Information and IT: *IAALD AFITA and WCCA. Tokyo, Japan. August 24-27, 2008.*
- Pahlevi, M., and Kojima, I. (2004) OGSA-WebDB: An OGSA-Based System for Bringing Web Databases into the Grid. Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC).
- Pengsuwan, J., and Sahavechaphan, N. (2008) WindChimer: The Partially Centralized and Controlled P2P System for Better Discovery. Workshop Proceeding of the International Conference on Advanced Information Networking and Applications. GinoWan, Okinawa, Japan. March 25-28, 2008
- Robinson, B., Freebairn, D., Bell, K., and Huda, S. Surveying farmers' goals and values reveals their diverse motivations and belief. Farming systems researchers have similar values. Available at [http://www.apsru.gov.au/apsru/Projects/WFS/pdf/Surveying\\_farmers.pdf](http://www.apsru.gov.au/apsru/Projects/WFS/pdf/Surveying_farmers.pdf).
- Trnková, J., Theilmann, W. (2004) Authoring processes for Advanced Learning Strategies. Available at <http://elara.tk.informatik.tu-darmstadt.de/publications/2004/StrategieAuthoring.pdf>
- Woehrer, A., Brezany, P., Janciak, I. (2004) Virtualization of Heterogeneous Data Sources for Grid Information Systems. Available at [citeseer.ist.psu.edu/woehrer04virtualization.html](http://citeseer.ist.psu.edu/woehrer04virtualization.html)