

# Federated Operation Model for the Language Grid

Toru Ishida<sup>1</sup>, Yohei Murakami<sup>2</sup>, Yoko Kubota<sup>1</sup>, Rieko Inaba<sup>1</sup>

<sup>1</sup>Department of Social Informatics, Kyoto University

<sup>2</sup>National Institute of Information and Communications Technology

ishida@i.kyoto-u.ac.jp, yohei@nict.go.jp,  
{yoko, inaba}@i.kyoto-u.ac.jp

## Abstract

The concept of collective intelligence is contributing significantly to knowledge creation on the Web. While current knowledge creation activities tend to be founded on the approach of assembling content such as texts, images and videos, we propose here the service-oriented approach. We use the term *service grid* to refer to a framework of collective intelligence based on Web services. This paper provides an institutional design mainly for non-profit service grids that are open to the public. In particular, we deepen the discussion of 1) intellectual property rights, 2) application systems, and 3) federated operations from the perspective of the following stakeholders: *service providers*, *service users* and *service grid operators* respectively. The Language Grid has been operating, based on the proposed institutional framework, since December 2007.

## 1 Introduction

Based on scalable computing environments, we propose a service-oriented approach to developing collective intelligence. This approach requires institutional design to share services among participants. In this paper, we call the infrastructure to form service-oriented collective intelligence the *service grid*<sup>1</sup>. The service grid has three stakeholders: *service providers*, *service users* and *service grid operators*. For the institutional design, we should consider the following issues related to each stakeholder:

---

<sup>1</sup> Service grid is a generic term meaning a framework where “services are composed to meet the requirements of a user community within constraints specified by the resource provider” (Furmento et al., 2002) (Krauter et al., 2002).

- How to protect intellectual property rights of service providers and to motivate them to provide services to the service grid. To this end, service providers should be allowed to define for what purpose or purposes their services can be used and to define usage rights accordingly.
- How to encourage a wide variety of activities of service users to increase their use of the provided services. To this end, service users should be allowed to run application systems that employ the services permitted for such use.
- How to reduce the load on service grid operators, while allowing them to globally extend their service grids. To this end, federated operation should be facilitated, where several operators collaboratively operate their service grids by connecting them in a peer-to-peer fashion.

In this paper, we describe our institutional design for a public service grid typically operated by non-profit organizations such as universities and research institutes. Based on this discussion, we have already developed the *service grid server software* and started the Language Grid that focuses on language services (Ishida, 2006). The rest of this paper describes the concept of service-oriented collective intelligence, the institutional design considering stakeholders including service providers, service users and service grid operators, and our experience in operating the Language Grid.

## 2 Stakeholders

To simplify the following discussions in this paper, the main stakeholders are classified into three groups:

- *Service provider* provides all kinds of services to the service grid.
- *Service user* invokes and uses the services provided to the service grid.
- *Service grid operator* is provided with services from the service providers, and allows the service users to invoke and use the provided services.

Service providers and service users are collectively called *service grid users*. A service grid user can act as a service provider as well as a service user. The role of the service grid operator is to stand between service grid users (typically between a service provider and a service user) and support their provision and use of the services. In the following sections, we discuss institutional design in terms of the contracts between a service grid operator and a service grid user.

Note that Web services are classified into *atomic services* and *composite services*. An atomic service means a Web service that enables service users to access the provided resources. Such provided resources include data, software, and human resources that are shared in the service grid as atomic services. On the other hand, a composite service means a Web service that is realized by a procedure called *workflow* that invokes atomic services.

To handle the intellectual properties present in the services and resources, the service grid operator may propose a unified license (GPL, Creative Commons etc.) to the service providers to register their services with the service grid. While a unified license will simplify the operation and promote the use of the service grid, it could cause the service providers to lose some or all of their incentives. Therefore, to better support the service providers, the institutional design of the service grid will not be based on the premise of a unified license.

The operation of the service grid discussed in the rest of this paper assumes that it is operated publicly mainly by non-profit organizations such as universities and research institutes. It does not assume the case of the service grid in a business firm, where service grid operators can completely or partially control the incentives of service grid users.

### 3 Service Provider

#### 3.1 Purpose of the Service Use

From the service provider's standpoint, any discussion of the protection of their intellectual property must address the purpose intended in

using their services. In fact, many research institutes and public organizations clearly specify that their services are for *non-profit or research use only*. To reflect such service providers' concerns, we classify the purpose of service use into the following three categories and allow each service provider to permit one or more of the categories:

- *Non-profit use* means 1) use by public institutions and non-profit organizations for their main activities, or 2) use by companies and organizations other than public institutions and non-profit organizations for their *corporate social responsibility* activities.
- *Research use* means the use for research that does not directly contribute to commercial profit.
- *Commercial use* means the use for purposes intended to directly or indirectly contribute to commercial profit.

The above classification can be applied to organizational use as well as personal use. However, when personal use only means private use, personal use can be classified as non-profit use. Note that activities by public institutions and non-profit organizations other than their main activities are excluded, aiming to prohibit service use to obtain funding. Meanwhile, corporate social responsibility activities are included in non-profit use because such activities are often operated in collaboration with public institutions or non-profit organizations.

If a service provider is already selling its service to organizations like local governments, it may not wish to allow non-profit use through the service grid. If service users want to use services, the specified purpose of service use must comply with the terms of use specified by the service provider.

#### 3.2 Control of Service Use

When service providers register their services in the service grid, they are required to provide information on copyright and other intellectual property rights of the resources included in their services. In the event that the service provider has been granted a license to the resource by a third party, such information shall also be included. The service provider is required to own the resources or the authority to allow third parties to use the resources. This prevents the service users from accidentally violating the third party's intellectual property rights.

Now, who should register and manage the services in the service grid? If we stand on the

premise that the collective intelligence is autonomously formed by the service providers, the service providers should be responsible for the maintenance of their resources, and the process of developing the resources into an atomic service, which we call *wrapping*. The service providers also have to maintain their services and the connection between the services and the service grid. On the other hand, to guarantee the service's quality and safety, the registration and maintenance of services should be done by the operator or with the operator's approval. Therefore, the decision about who should register and manage the services needs to be made considering the trade-off between stimulating the autonomous activities of the service provider and ensuring the quality and safety of the service grid. Likewise, we need to consider whether to leave the service deregistration process to the service provider or the operator. When focusing on the quality and the safety of the service grid, at least to cover the case of emergencies, the operator needs to be able to deregister a service.

For the service provider, it is desirable that there be flexibility in setting out the terms of use of their services. For example, the possible conditions are as follows:

- Restrictions on the service users who may be licensed to use the services;
- Restrictions on the purpose for which the services may be used;
- Restrictions on the application systems that use the services;
- Restrictions on the number of times that the services may be accessed and the amount of data that may be downloaded from the services.

By setting out conditions of their services employing the same resource, the service provider can provide their services under dual license. For example, one is provided to every user under several restrictions on access counts and data transfer size without any charge. Meanwhile, the other is provided to the users who pay a fee without any restrictions.

In general, when the service grid allows the terms of use to be set in detail, it will increase the service provider's satisfaction, while forcing greater overhead on the service users to comply with the detailed terms of use. Moreover, when the service users use a composite service, they need to satisfy all terms of use of every atomic service in the composite service. If we try to assure that automatically, the operator must provide technical measures to ensure that the service

users will not violate the terms of use. Therefore, we must trade the service provider's flexibility off against the service user's convenience and the operator's cost.

## 4 Service User

### 4.1 Service Use through Application System

When service users use the service grid for purposes other than personal use, many of them provide an *application system* using services to other users. Here *application system* means, as shown in Fig. 1, a system that is provided by a service user and that allows users of the system to indirectly access the service grid without being personally authorized by the service grid. In this case, the service user is responsible for ensuring that the application system users comply with the terms of use of each service that is used through the application system.

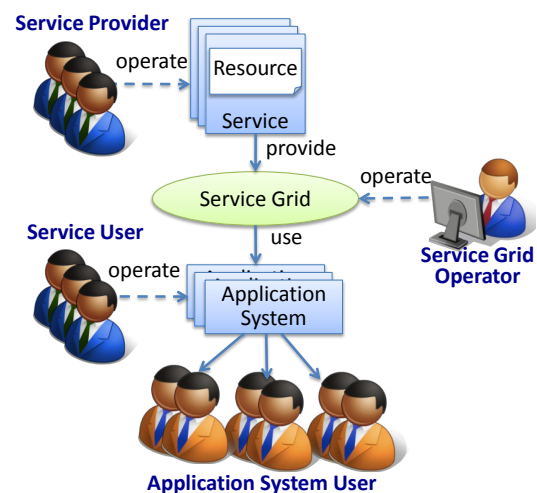


Fig. 1. Service Use through Application System

### 4.2 Control of Application System

A service user may operate different types of application systems; for example, one provides an application system to the general public through the Web, and another provides an application system through a particular terminal in a certain location like a reception counter. This paper focuses on how an application system can be controlled by the service user and classifies the control of application systems into two types: *under client control* and *under server control*.

- *Under client control* means the status where the users of an application system are under the control of the service user who provides the application system.

More specifically, it means the status where the terminals of application system users are under the control of the service user or where the service user is able to identify each application system user. In all cases, the service user who provides the application system must be able to fully grasp at any time the status of use of the application system at each terminal and/or by each user, and have the technical and legal authority to suspend use as necessary.

- *Under server control* means the status where the server on which the application system runs is under the control of the service user, while application system users are not under the control of the service user. In this case, the service user must be able to fully grasp at any time the status of use of the application system server and have the technical and legal authority to suspend the server as necessary.

Two examples of the operation of an application system are shown in Fig. 2. When an application system provided through the Web can be accessed by users from home without authentication, the status is not *under client control*; however, if the service user controls the Web server, the status is *under server control*. When an application system is provided through a terminal at a reception counter and the terminal is under the control of the service user, the operation is classified as *under client control*.

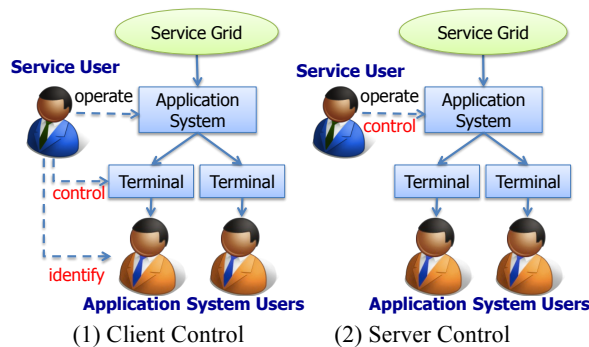


Fig. 2. Service Use through Application System

The classification aims to allow service users to develop their own application system and select properly the range of the application system to be offered. Furthermore, the service provider can limit the range of application system users by specifying which type of application system control the provided service must adopt. For example, when a service provider sells a service to

local governments, the service provider may agree to provide the service to patients at a reception counter in a hospital (*under client control*) but may refuse to provide the service to the public through a local government's Web server (*under server control*).

### 4.3 Return for Service Providers

Where is the service provider's incentive for providing their services? When the service providers provide their services for free, the service grid operator is required to provide statistical information on the use of the services to the service providers. The statistical information shows who used or is using which service and to what extent. Such information stimulates the interaction between the service providers and the service users. However, the statistical information should not include any transferred data or personal information regarding the senders of data. In case the service providers wish to obtain information on the use of the services other than statistics, the provider should conclude an agreement that establishes the provision of such information with the service user. The service grid operator is not involved in such an agreement.

When service providers provide their services for profit, they will receive fees from the service users by concluding a contract for the payment of such fees. Again, the operator is not involved in such contracts.

## 5 Service Grid Operator

To globally disseminate the service grid, which is centered on non-profit organizations like universities and research institutes, multiple operator organizations need to create/join an affiliation. We call this *federated operation*. The reasons driving federated operation include not only the limited number of users that a single operator can handle, but also the locality caused by geographical conditions and application domains.

There are two types of federated operation. One is *centralized affiliation*, where the operators form a federal association to control the terms of affiliation based on mutual agreement. This yields flexibility in deciding affiliation style, but incurs a lot of cost in maintaining the federal association. The other is *decentralized affiliation*, which allows a service grid user to create and become the operator of a new service grid that reuses the agreements set by the first service grid. This type of operation promotes forming peer-to-

peer networks by the operators. The type of affiliation is defined by reuse of agreements, but the formation of the peer-to-peer network by the operators is flexible and no maintenance cost is necessary. In the following section, we further discuss decentralized affiliation since it suits non-profit organizations like universities and research institutes.

Let an *affiliated operator* be a service grid user who operates its own service grid that reuses the agreements of the original service grid. Let an *affiliated user* be a user who is licensed to use the affiliated operator's service grid. In such a case, as shown in Fig. 3, the affiliated user can use the original service grid, in which the affiliated operator takes the role of a service grid user. That is the key idea of the peer-to-peer federated operation. Even in such case, service providers still have the right to choose whether to allow the affiliated user to use their services or not.

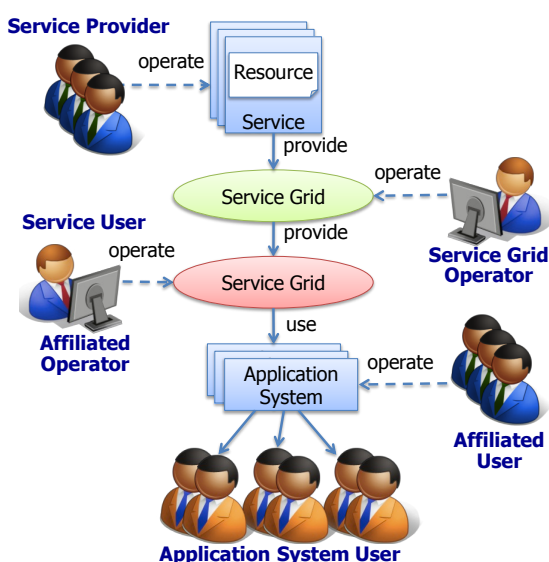


Fig. 3. Federated operation of service grid

Two service grids in equal partnership are likely to establish a *bidirectional affiliation*, where both operators become users of the other service grid. *Unidirectional affiliation* is also possible. For example, if one service grid provides only basic services and the other provides only applied services, the latter can be a user of the former service grid.

Sometimes it is impossible for different service grids to use exactly the same agreements. A typical problem is the governing law. For international affiliation, a possible idea is to adopt a common law like New York State law, but operators may wish to adopt the governing law of their own locations. In such a case, operators will

use the same agreements except for the governing law. In that case, the service providers would need to accept the use of the different governing law to handle the affiliated users in that location.

## 6 Operation of the Language Grid

### 6.1 Language Grid Service Manager

The Language Grid is a service grid for language resources. Its concept was developed in 2005, and the project was launched in April 2006 (Ishida, 2006). The fundamental software forming the service grid was developed and has been released by the National Institute of Information and Communications Technology (NICT).

In designing the Language Grid system, it was important to deal with service providers, who had various incentives. For example, some language services may already be sold for profit. If the service grid failed to allow the service provider to receive fees for their services, it would be hard to realize a service grid that truly satisfied service users. Furthermore, since each of the existing dictionaries and language processing software had various types of licenses, the operator could not unify those licenses. Many research institutes that develop language resources can provide their resources as long as they are used only for research. However, if they are used by non-profit organizations for their activities, the research institutes may need to know by who, when, and how much their resources are being used. Such various incentives and conditions form the background of our institutional design prioritizing the intellectual property rights of the service providers. In our operation model (Ishida et al., 2008), language service providers can fully control access to their language services using the Language Grid. Language service providers can select users, restrict the total number of accesses per year/month/day, and set the maximum volume of data transfer per access. Providers can set those conditions via the Language Grid Service Manager (see Fig. 4). This software provides the registration of services, measurement of service usage frequency, access control of services, and always monitors the Language Grid.

On the other hand, service users wish to use the provided language resources in their various activities. At a school with multi-national students, teachers and parents as well as students will use language services. To allow a large number of people to use the services, the school is required to identify their registered users properly. At the reception counter of a hospital, however, it is difficult to ask patients to register themselves to the reception support system. It is more real-



Fig. 4. Language Grid Service Manager

istic to identify the terminals to permit service access. In this way, the system must be designed to allow many application system users to use language services in their different environments. To avoid the fraudulent usage of language services, however, service users should not allow the application system users to discover the ID and password of the Language Grid. For example, in the case of an NPO offering medical interpreter services to foreign patients, the NPO is required to enter their Language Grid ID and password in such a way that they do not become public; one solution is to embed the ID and password in their patient support systems.

## 6.2 Centralized Operation

The *service grid server software* has been developed and released as open source software. Using this source code, universities and research institutes can operate any kind of service grid. The Department of Social Informatics of Kyoto University started operation of the Language Grid for nonprofit purposes in December 2007. As of June 2011, 139 groups in 17 countries had joined the Language Grid: research institutes include Chinese Academy of Sciences, the National Research Council (CNR), German Research Center for Artificial Intelligence (DFKI), and National Institute of Informatics (NII), universities include Stuttgart University, Princeton University, Tsinghua University and a number of Japanese universities, NPO/NGOs and public sector bodies. Companies have also joined: Nippon Telegraph and Telephone Corporation (NTT), Toshiba, Oki and Google are providing their services without any charge.

We first expected that NPO, NGO and public sectors would become the major users, but uni-

versities are using the Language Grid more intensively at this moment; researchers and students who are working on Web analyses, CSCW, and multicultural issues are using language services for attaining their research goals. This trend is natural in the early stage of introducing a new Internet technology. Fig. 5 shows the recent statistics of member organizations.

Research institutes, universities, and companies are providing atomic language services such as dictionaries and machine translators. The number of shared language resources now totals 67. Organizations that provided language resources include Chinese Academy of Sciences, Stuttgart, Princeton, Kookmin, and Kyoto Universities, NICT, NII, NTT, Google, Toshiba, Oki, Kodensha, Asian Disaster Reduction Center and a number of public sector groups and NPO/NGOs. When providing atomic language services, providers specify copyright notices and license information in the profiles of the resources. To create composite services that involve the combination of atomic services, many workflows are being written and released. Currently more than 100 services are registered in the Language Grid.

The operation model designed by the authors reflects the intentions of user groups around the world like research institutes and non-profit organizations (Ishida et al. 2008). We were only able to attract such participants because we developed the Language Grid with a strong bias towards formalizing the obligations of all parties. Design of the operation model was conducted in parallel with development of the service grid server software. It took more than six months to achieve consensus on the model. It is probably fair to say that the software was written to realize the operation model.

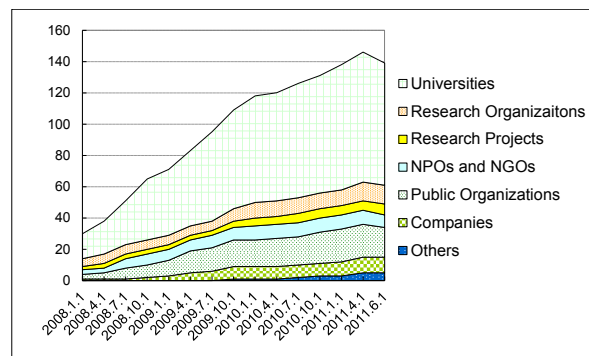


Fig. 5. Number of participant organizations

### 6.3 Federated Operation

From operating the Language Grid over two years, we have gained many insights. One of them is the importance of federated operation. Since the operation center in Kyoto cannot reach local organizations in other countries, over 70 percent of participating organizations are in Japan. Since we need global collaboration, even for solving language issues in local communities, this imbalance should be overcome: the Language Grid operators need to be dispersed into different organizations globally and to collaborate with each other. The federated operation model was invented to realize such collaboration. In fact, the National Electronics and Computer Technology Center (NECTEC) in Thailand launched the Bangkok Operation Center in October 2010, and is now federated with the Kyoto Operation Center. The Bangkok Operation Center has a plan to provide a collection of atomic services for language processing i.e. LEXiTRON for a Thai-English dictionary, Parsit for English to Thai machine translation, Vaja for Thai text to speech conversion, and morphological analysis utilities. Those services can be accessed by users of the Kyoto Operation Center.

So far, we have described the federated operation of the same kind of service grids. In fact, we had an opportunity to realize the collaboration of different kinds of service grids. The joint research between Tsinghua University's Smart Classroom and the Language Grid is a typical achievement (Suo *et al.*, 2009). We rebuilt Tsinghua University's Smart Classroom as a collection of pervasive computing services. That allowed easier connection between the Smart Classroom and the Language Grid to develop Open Smart Classroom, which connects classrooms in different countries. NECTEC also needs the collaboration of different kinds of service grids provided by neighboring interest groups. These services will soon be extended to cover other media resulting from NECTEC's initiative called the Digitized Thailand Project.

## 7 Conclusion

In this paper, we named an infrastructure that forms collective intelligence based on Web services a *service grid*, and designed an institutional framework for a public service grid operated by non-profit organizations such as universities and research institutes. From a consideration of the different standpoints of *service providers*, *service users* and *service grid operators*, which consti-

tute the service grid, we proposed the following framework:

- To protect the intellectual property rights of service providers, the purposes of service use are classified into *non-profit use*, *research use*, and *commercial use*. The service providers can set the terms of service use for each purpose.
- The type of control employed by application systems are classified into *client control* and *server control*. This flexibility allows service users to employ different types of application systems to support their activities.
- To decrease the cost of service grid operators and extend service grid operation globally, the framework allows service grid operators to conduct *federated operation*. The collaboration is realized in a peer-to-peer fashion by introducing the concepts of *affiliated operators* and *affiliated users*.

The institutional design discussed in this paper is based on our three-year experience of operating the Language Grid. We hope that our experiences will promote the accumulation of knowledge about designing institutional frameworks and contribute to the development of service-oriented collective intelligence.

### Acknowledgments

We acknowledge the considerable support of National Institute of Information and Communications Technology, and Department of Social Informatics, Kyoto University. A part of this work was supported by Strategic Information and Communications R&D Promotion Programme from Ministry of Internal Affairs, a Grant-in-Aid for Scientific Research (A) (21240014, 2009-2011) from Japan Society for the Promotion of Science (JSPS), and Kyoto University Global COE Program: Informatics Education and Research Center for Knowledge-Circulating Society.

### References

- Nathalie Furmento, William Lee, Anthony Mayer, Steven Newhouse, John Darlington. 2002. ICENI: an open grid service architecture implemented with Jini. *International Conference on High Performance Networking and Computing*:1-10.
- Toru Ishida. 2006. Language Grid: an infrastructure for intercultural collaboration. *2006 IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*: 96-100.

- Toru Ishida, Akiyo Nadamoto, Yohei Murakami, Rieko Inaba, Tomohiro Shigenobu, Shigeo Matsubara, Hiromitsu Hattori, Yoko Kubota, Takao Nakaguchi, Eri Tsunokawa. 2008. A non-profit operation model for the language grid. *International Conference on Global Interoperability for Language Resources*: 114-121.
- Klaus Krauter, Rajkumar Buyya, Muthucumaru Maheswaran. 2002. A taxonomy and survey of grid resource management systems for distributed computing. *Software: Practice & Experience* 32(2): 135-164.
- Paul P. Maglio, Savitha Srinivasan, Jeffrey T. Kruelen, Jim Spohrer. 2006. Service systems, service scientists, SSME, and innovation. *Communications of the ACM* 49(7): 81-85.
- Yue Suo, Naoki Miyata, Hiroki Morikawa, Toru Ishida, Yuanchun Shi. 2009. Open smart classroom: extensible and scalable learning system in smart space using web service technology. *IEEE Transactions on Knowledge and Data Engineering* 21(6): 814-828