A Survey of Grid Knowledge and Grid Perception in the Public Sector

Xin ZHOU [a,1], Owen APPLETON [b], Henning MÜLLER [a,c]
[a] *Medical Informatics, Univ. Hospitals and University of Geneva, Switzerland*
[b] *Emergence Tech Limited, London, United Kingdom*
[c] *Business Information Systems, HES–SO, Sierre, Switzerland*

Abstract. Grids have been developed to enable computational or data intensive scientific applications and to share resources. Existing infrastructures can benefit other possible user groups. This qualitative survey focuses on the public sector with universities, governments and hospitals. Interviews were performed to collect information on the perception of Grids by decision makers, system administrators and high level users in the sector. Application scenarios and main barriers for Grid adoption are discussed based on the users' and institutional interests.

Keywords. Grid, Survey, User perceptions, Public sector

1. Introduction

A substantial increase of commodity computing and network performance in the last decade led to the birth of the concept of the "Grid". The word Grid is inspired by the electrical power grid, which provides a pervasive access to electrical power irrespective of its sources [1]. In the same manner, a Grid in computing provides pervasive access to computing power, data storage capacity, remote data management, web services, and potentially other computational capabilities. The Grid has the potential to bridge computer resources and can perform as a super computer with ever–increasing capacity [2]. Efforts have been made by the European Union to build large–scale Grid infrastructures [3,4], which can be used by scientific communities, such as high–energy physics [5], chemistry [6], biology [7], and healthcare [8,9]. These application domains are often named e–Science. One of the target user groups for Grids is the public sector, including universities, governments and hospitals as these groups usually have a large number of desktop computers available and centrally managed. In current commercial environments the concept of Grids has been somewhat superseded by the cloud concept where commercial actors build large computing centers and sell storage or computing power on demand instead of the resource sharing of Grid structures.

[1]Corresponding Author: Medical Informatics, Geneva Univ. Hospitals and University of Geneva, Rue Gabrielle–Perret–Gentil 4, 1211 Geneva 14, Switzerland, xin.zhou@sim.hcuge.ch

During the past few years, surveys have been made to gain a better understanding of Grid knowledge and perception in these areas. Work focussing on summarizing the status of Grid use for education on a national level is described in [10], others concentrate on business models of healthgrid adaptation [11,12]. The survey described in this paper focuses on Grid use in the public sector. Only decision makers were interviewed to limit the number of interviewees. A similar survey was produced in 2007 [13] and this survey is meant to be a qualitative follow–up. Besides obtaining information from the interviewed persons it was also the goal of the survey to inform decision makers on current technologies.

2. Methods

The survey is based on interviews taken during the summer of 2009 in London, UK and Geneva, Switzerland. Previous work [13] on the topic targeted mainly system experts. Decision makers including university professors, medical doctors and research group leaders were added to the list of interviewees. This change is based on Grid strategies moving from building an infrastructure to developing Grid application.

2.1. Survey design

The persons present various roles and possess differing professional interests:

- high level users are usually interested in short–term resource availability and benefits for projects;
- system experts (or administrators) may consider the flexibility of new technology and interoperability with existing systems.

Three types of questionnaires were prepared. The number of questions depends on the type of person interviewed. Inside the survey forms four groups of questions were used: (1) the background of interviewees, (2) the IT (Information Technology) strategy of organizations, (3) the technical knowledge of existing systems and (4) the perception of Grids. The questions related to *background* and *Grid perception* were standardized for all interviewees. *Strategy* and *technical* questions were only asked for decision makers and high level Grid users respectively. High level Grid users were divided into two groups: application programmers and system administrators. The number of questions in each type of questionnaire is shown in Table 1. Questionnaires were validated by a professional outreach staff from the KnowARC[2] project. All the questions were open questions and the goal was mainly qualitative and not a quantitative assessment.

2.2. Interviewees

Besides the standard questions, additional questions were added during surveys for an in–depth understanding. All responses given by interviewees were recorded on tape and the interviewer prepared this article by revising these records. The

[2] `http://www.knowarc.eu/`

Table 1. The number of questions of the various types that were asked.

Interviewee	Application programmer	System administrator	Decision makers
Background	7	7	7
Strategy	0	0	6
Hardware	0	7	0
Network	3	6	0
Security	0	4	0
Sys/app	5	7	0
Grid perception	5	4	4
Total	18	33	17

Table 2. Background of interviewees and organizations.

Questions	Answers
Title/occupation	1 legal expert, 1 clinician, 3 research group leaders,
	3 system administrators (leader of maintenance group),
	1 IT unit header, 1 government consultant
IT qualification	3 IT PhDs, 1 PhD in physics, 1 master in IT, others are
	other domains (qualification in ethics, chemistry, or medicine)
IT experience	professional (years) : 0–5 (2), 5–10 (2), 10–20 (2), 20–30 (3), >30 (1)
Organization	hospitals (4), universities (5), government (1)
Organization size	all have >5000 people
IT department size	hospitals (200–350 IT engineers), government (>300),
	universities (∼120 IT staff and researchers)
Grid knowledge	expert (2, system admins at CERN), good (4), basic (3), none (1)

survey was anonymous but interviewees' titles and background were analyzed. In total 10 persons were interviewed by two different persons for this qualitative survey. Each interview lasted from 30 to 90 minutes.

3. Results

In this section we present the main results of the interviews regrouped by the type of the questions. The number in parentheses presents the number of interviewees with the same answers.

3.1. Background

The background of interviewees and their institutions are presented in Table 2. Half of the interviewed persons obtained an advanced IT qualification, 80% have long–term professional IT experience. One third are system administrators, another third application developers and the remaining persons either end users or consultants. All persons are group or department leaders. High level decision makers generally have a good global understanding, which ensures the quality of answers. It also shows that Grid related knowledge has increased over a previous survey.

All organizations chosen are large administrations with an IT department of over a hundred employees. Due to the priorities of each organization, the composition of the IT department is different: universities involve researchers as part of the IT staff, hospitals employ experienced developers to assure the stability of systems and infrastructures, and governments outsource much of the development and employ mainly project leaders and technical consultants. These orientations change the perception of new technologies in general as well as Grids.

3.2. Strategy

Strategy questions were only addressed to decision makers. The answers are classified by type of organization. The major strategy information is summarized in Table 3 and the end of the text, where Q = Question, U = University, H = Hospitals, G = Government.

3.3. Hardware and network

As all the system administrators chosen for the survey are from universities and hospitals, information concerning more detailed hardware constraints for governments was not acquired. Results are listed in Table 4. Responses concerning Internet and security are shown in Table 5. Universities generally have rather open networks. All PCs can directly connected to the Internet. Local firewalls are installed on desktop PCs as well as on servers. Due to security constraints not all PCs in hospitals and governments are able to access the Internet directly. A proxy and firewalls are used for Internet access.

3.4. Applications

The application questions were asked to the application developers, mainly researchers. Decision makers and system administrators often give global views of interest and constraints to set up Grids. However, one system administrator mentioned that even when such an infrastructure is available, research groups still prefer to use dedicated servers for their computing, as well as storage. Application developers are basically end users of this infrastructure. Answers are collected in Table 6.

3.5. Perception of Grids and Clouds

The general perception of Grids and clouds was inquired at the end of interview. The answers are shown in Table 7.

4. Discussions

4.1. Strategy

Strategy answers presented in Table 3 show that IT is the daily work for the interviewees' institutions. Strategies for validation of new technology can be classified

Table 3. IT Strategies in the organizations.

Q:	How would you characterize the level of IT competence of your staff?
U:	People in education often have good IT competence.
H:	Completely heterogeneous, young people are an IT generation whereas older people have difficulties to adapt, although basic equipment (screen, mouse) are used everywhere, fear of computers remains.
G:	Regular IT tutorials are given for education so that people understand IT concepts without going into technical details.
Q:	How would you rate the importance of IT for your organization?
U:	Materials and plans of courses, accounting, research, everything needs IT.
H:	Many applications such as nursing care plan and workflow management rely on IT. It serves also as internal information connection.
G:	2000 applications are permanently running and required to be online.
Q:	Rough percentage of your budget going to IT services and resources?
H:	Budget of Hospitals' IT services is $\sim$100 mio CHF, about 3% of the total budget.
G:	Budget of government IT services is $\sim$187 mio CHF, about 1.2% of the total budget.
Q:	What approach does your organization have for new technologies?
U:	Researchers have specific requirements, we leave them validate new technologies. Users take responsibility themselves.
H:	Neither early nor late adoption, final decision depends on the cross–validation with end users as a proof of maturity.
G:	Technology observatory regularly surveys new technologies and makes proposals, otherwise companies try to promote their products to the government.
Q:	What are the main barriers to adoption of new technology for your organization?
U:	Hard to find funding suited for the work, too often calls are essentially for a single application rather than being truly open.
H:	Heavy adaptation for users. If users find that new technology is cumbersome for their work, this technology is ignored.
G:	Lack of competent people who manipulate new techniques well, another problem is application adaptation. One example is that OpenOffice was used to replace MS Office where format adaptation was a big problem. Finally, this decision was taken back.
Q:	Do you have an opinion on the use of open source technology in your organization?
U:	Open source is widely used in the University (Moodle for teaching, webmail...). A software center is set up to provide and maintain open source software.
H:	Few open source tools are used and are only for small user groups (Osirix). Linux is used. only by IT staff. OpenOffice is being tested but most applications are based on Commercial solutions (Microsoft Windows XP, MS Office, Oracle, etc).
G:	Efforts were used to investigated Linux and OpenOffice for financial reasons, requirements are very specific and it is hard to find suitable open source solutions.

into three categories: bottom–up validation, top–down validation and cross validation. Universities mainly use bottom–up validation for their technology choices. Only PCs for students are fully under central supervision. End users such as professors and other researchers select the software based on their needs. The only mentioned barrier for researchers is funding that is required to buy hard– and

Table 4. Hardware and its configuration inside the institutions.

Questions	Universities	Hospitals
Number of desktop PCs	>5000	∼8000
Number of servers	∼150	>200
PC configuration	Dual–core 1GB RAM	Dual–core 2.8GHz, 2GB RAM
Server configuration	variable, standard: 2 CPU, 2–4GB RAM	variable, 1–32 CPU 1–32 GB RAM
OS for PCs	70% Windows XP, 30% MAC	100% Windows XP
OS for servers	Solaris	Suse Linux
Renewal cycle (PCs)	3–5 years	3 years
Renewal cycle (servers)	5–10 years	5–8 years
Large–scale data storage	∼100TB	8–50TB
Data protection	backup	RAID + backup

Table 5. Network and security inside institutions.

Questions	Universities	Hospitals	Government
Network access	direct Internet access (wired and wireless)	Intranet, firewall +IP phone	Intranet, firewall +IP phone
Network speed	1–10 GB	0.1–4 GB	10GB
Network Charge	variable, planned not to fully charge network	70–80% (20h/day) for outbound, only 20% for internal network	each service has its rush hour in different periods
Outbound protection	NAT, proxy	NAT, DMZ, proxy	NAT, DMZ, proxy
High priority applications	admin applications + research applications	clinical applications + admin applications	online services + admin applications

software. If funding is available purchasing new material or using open source software is allowed. So far funding is only for single research groups' requirements but not for the common infrastructure of the institution.

Government bodies are rather based on the top–down model as consultants survey new technologies and make proposals if there exists an interest. The interviewed persons state that the employees of the government have good IT competence but also face barriers regarding new technologies. Sometimes a lack of qualified personnel knowing well the new techniques was mentioned. This leads to the problem that even users with good IT competence can have problems with an adaptation of the infrastructures

Hospitals use a model in between the two above. Innovation is important and new technologies or ideas can be proposed as research projects. When going into the production stage, cross–validation with end users (usually clinicians) and particularly security constraints are taken into account for the decision making. Not all users are IT experts and mainly simple user interfaces and quick response times are expected.

Table 6. Responses on application development.

Questions	Answers
Develop platform	Mac OS (2), Windows (2), Linux (1), FreeBSD (1)
Programming languages	C/C++ (2), Matlab (2), Java (2), Python/Perl/cgi (1)
Application domain	medical imaging (1), clinical simulation (2), data mining (1), medical coding (1), IT management (1), web services (1)
Interface adaptation	Oracle (3), Ansys (1), Matlab (2), smart card systems (1), patient data management (1)
Internet access	proxy (4), VPN (2), directly (1)
Data access	Local access (4), in–house access (2), via other applications (1)
Main end users	researchers (3), clinicians (4), system administrators (1), coders (1), public (1)
Configuration required	X–Grid (1), central infrastructure (1), dedicated servers (2), desktop PCs (3)
Available resources	X–Grid consists of 20 Apple Xserve G5 2.3GHz, 4GB RAM(1), dedicated servers (2 IBM servers, 4–core 12GB RAM + 8–core 16GB RAM) (1), dozens of desktop PCs (3)
Difficulty in development	user motivation (1), lack of computational resources (1), variable parsing (1), obtention of funding (1), lack of competence (1)

Table 7. Perception of Grid and clouds.

Q:	What aspects of Grid computing are most interesting for you?
A:	computing power (3), Grid web services (3), distributed storage (1), increased efficiency of resources (2), simple access to heterogeneous resources (2), data sharing (4), not interested at all (1)
Q:	If you have interest in Grid, cloud and distributed computing what barriers to adoption do you see?
A:	application migration (4), change of existing infrastructure (1), lack of expertise (2), not supported commercially (3), licenses for Grid computing(1), low motivation for sharing(4), legal restriction concerning confidential data (2),

To promote Grid infrastructures, choosing the right validation model is certainly important.

4.2. Technical

Answers for hardware in Table 4 show that the configuration of desktops is more homogeneous than the setup of servers. The desktop renewal cycle is shorter than that of servers as servers are generally dedicated to one application and expected to work permanently. Servers are usually reserved for priority applications and can not be targeted as available resources, for example for Grid applications. Increasingly virtualization is used to reduce the physical number of servers. Desktop PCs

are less critical and much larger in numbers. A Grid is considered a convenient choice under the condition that solutions do not cause problems for the desktop users. System administrators show interest in exploring these idle resources and making them available as a research infrastructure.

Current storage capacity is considered large enough and storage devices are not extremely expensive anymore. Data are confidential and a central storage with backup is preferred over fully distributed storage systems. Less interest was shown in distributed storage solutions.

Academic Grid solutions are often built based on the assumption of open networks. Network and security answers in Table 5 show that hospitals and governments do not always have open network, which can be a barrier for adoption. Both hospitals and governments do not allow data export, another constraint for Grid computing.

Besides security and policy constraints, network charge and application requirements differ between institutions. As government services are required to be online and accessible for the public a high network capacity is required. The government thus uses a high speed 10GB/s connection. Servers are not used at full capacity as each service has high traffic at a different time. Instead of using a server per application a better balance could be obtained by using a Grid to provide web services.

Hospitals usually do not foresee strong outbound network resources, which are sometimes at full capacity. On the Intranet the capacity is only used at around 20% on the other hand. Applications are mainly clinical, so for internal use. Hospitals thus attach more importance to a robust intranet inside the hospitals and to limit the outbound communication. Clinical applications using images can on the other hand be computationally intensive. Collecting resources from the hospital network can be a solution under the condition that the network charge does not increase strongly.

Applications can strongly differ in terms of platforms, languages, end users, etc. Developers have to encapsulate technical parts to provide interfaces. Lack of computational power was mentioned as one constraint for application development. Not all the applications required advanced hardware and some research groups have own high performance resources. Administrative applications are frequently mentioned as priority applications that require little computing power.

Except one person all are interested in using Grids either for getting access to more computing power, for optimizing the resource usage, or for accessing free resources from other institutions. Half of the interviewees have experience in using Grids, with some having set up a local Grid. The only one not interested in Grids emphasized that developing Grids beyond research generates overhead and publishing would be delayed.

4.3. Overall perception

The most frequently requested functionality of Grids is data sharing. However, interviewees were interested in data sharing in the sense of collaboration, which is more a political than a technical problem. Application developers and researchers were interested in computing power with an easy access, whereas system adminis-

trators were interested in distributed web service possibilities to increase the efficiency of resources. Only one researcher used distributed storage for very specific applications. All others prefer central storage with central backups.

Various barriers were identified. System administrators also had fears from a technical point of view:

- the goal of building an infrastructure is not to use the resources to the full capacity but to have more than enough to allow all applications to run correctly with the remaining capacity;
- from a debugging point of view system administrators prefer that each machine focuses on a single task.

Technical problems can be solved by virtualization, where applications are separated from hardware and resources used by applications can be adapted.

The most frequently mentioned barrier is not technical but the low motivation of sharing resources and the cost of application migration. Low motivation to share resources mainly comes from:

- people preferring to pay for their own project and their own group rather than investing in an infrastructure that everyone can use;
- legal restrictions using confidential data;
- IT infrastructures being often not funded in non–IT research projects;
- a required protection from competitors when sharing data.

The cost of application migration comes from:

- a lack of expertise in Grids making application migration time–consuming, lowering the potential gains;
- a lack of suitable tools for debugging.

The low level of Grid commercialization is also a problem as interviewees mentioned that widely used commercial solutions such as Oracle, SAP do not always foresee Grid use. Many applications are based on these and as a consequence end users will not use Grids if this is not properly integrated. Licensing is another problem as a large number of licenses is needed when distributing some commercial applications (one per potential computing node). New types of licenses for Grids are not available for all commercial solutions. So far, Grids are often a platform for executing self programmed or open source software.

4.4. Comparison

Compared with a previous survey [13] the interviews show that decision makers and IT staff have increased the understanding of Grids. Only the most important questions are cited in Table 8 to establish a brief comparison. More interviewees have tried Grid infrastructures in 2009. In 2007, possible application scenarios were discussed in a general way and problems concerned politics and security. In 2009, more specific needs and well formed application scenarios were mentioned during the interviews. such as using Grids to balance high capacity applications in governments. Moreover, the interviewees better identify the requirements. In 2007, many took the access to computing power and storage capacity as being totally

Table 8. Understanding of Grids between 2007 and 2009.

Q:	Percentage of interviewees having tested Grids?
A:	2007 : 30% 2009 : >50%
Q:	Which Grid services would you find most useful?
A:	2007 : Grid computing(8), Grid storage(6) 2009 : Grid computing(3), Grid storage(1), Grid web services(3), data sharing(4)
Q:	Main barriers for adoption?
A:	2007 : Security, lack of expertise for installation and maintenance; 2009 : motivation of sharing, funding, lack of expertise for application migration, lack of commercial support, etc.

free of charge. In 2009, the costs of application migration, debugging, maintenance are taken into account. Concrete problems in using Grids were described such as license problems when using commercial solutions. Funding policies and virtual organization policies were mentioned in the discussion.

5. Conclusions

A survey was performed collecting data on the perception of Grids in the public sector. All public sectors consider new IT as a key route to improving service quality. Barriers in strategies for Grid adoption are mainly the implied doubt of the stability of open source software and the low motivation to share resources. Perceptions from system administrators and application developers are related to their experience of using Grid infrastructures. One important message given by system administrators is that Grids are often based on the assumption of using all resources at their full potential, which is not the first priority for the system administrators who prefer having reserves. Idle resources are mainly desktop PCs and solutions exploring the use of these were welcomed by most. Sharing servers is more problematic as they are often dedicated to a particular project or application. Governments are interested in web–services and cloud technologies due to the requirement of putting all services online. Hospitals are rather interested in computing power and data sharing. Storage infrastructures are confidential and Grid solutions are not considered for storage. Another complaint is that Grid solutions are not supported well enough by commercial solutions, preventing their adoption. A common experience of time consuming application migration is reported by application developers, which is the biggest barrier to Grid adoption from the developers' point of view.

Grid and cloud usage are seen as a trend for administrations. It is critical to reduce the overhead for adapting software and respond better to the users' needs. Some technical barriers can be solved by virtualization. When these barriers can be resolved a higher acceptance of Grids and clouds can be expected.

6. Acknowledgments

This work was partially supported by the EU 6th Framework Program in the context of the KnowARC project (IST 032691) and by the Swiss National Science Foundation (FNS) in the context of Talisman–2 (project 200020 118638).

References

[1] Buyya, R., Venugopal, S.: A gentle introduction to grid computing and technologies. CSI Communications **29** (2005) 9–19

[2] Kesselman, C., Foster, I.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Publishers (1998)

[3] Gagliardi, F., Jones, B., François, G., Bégin, M.E., Heikkurinen, M.: Building an infrastructure for scientific grid computing: status and goals of the egee project. Philosophical Transactions of the Royal Society A **363** (2005) 1729–1742

[4] Eerola, P., Kónya, B., Smirnova, O., Ekelöf, T., Ellert, M., Hansen, J.R., Nielsen, J.L., Wäänänen, A., Konstantinov, A., Herrala, J., Tuisku, M., Myklebust, T., Saada, O.F., Vinter, B.: The nordugrid production grid infrastructure, status and plans. In: GRID '03: Proceedings of the 4th International Workshop on Grid Computing, Washington, DC, USA, IEEE Computer Society (2003)

[5] Read, A., Taga, A., O-Saada, F., Pajchel, K., Samset, B.H., Cameron, D.: Complete distributed computing environment for a hep experiment: experience with arc-connected infrastructure for atlas. Journal of Physics: Conference Series **119** (2008) 062041+

[6] Zhou, Z., Wang, F., Todd, B.D.: Development of chemistry portal for grid-enabled molecular science. In: E-SCIENCE '05: Proceedings of the First International Conference on e-Science and Grid Computing, Washington, DC, USA, IEEE Computer Society (2005) 48–55

[7] Krabbenhöft, H.N., Möller, S., Bayer, D.: Integrating ARC Grid middleware with Taverna workflows. Bioinformatics **24** (2008) 1221–1222

[8] Müller, H., Pitkanen, M., Zhou, X., Depeursinge, A., Iavindrasana, J., Geissbuhler, A.: Knowarc: Enabling Grid networks for the biomedical research community. In: Healthgrid 2007, Geneva, Switzerland (2007) 261–268

[9] Krefting, D., Bart, J., Beronov, K., Dzhimova, O., J., F., Hartung, M., A., H., Knoch, T.A., Lingner, T., Mohammed, Y., Peter, K., Rahm, E., Sax, U., Sommerfeld, D., Steinke, T., Tolsdorff, T., Vossberg, M., Viezens, F., Weisbecker, A.: Medigrid: Towards a user friendly secured grid infrastructure. Future Generation Computer Systems **25** (2009) 326–336

[10] Soni, D., Sharma, J.: Role of grid computing in indian education. In: MATH'07: Proceedings of the 12th WSEAS International Conference on Applied Mathematics, Stevens Point, Wisconsin, USA, World Scientific and Engineering Academy and Society (WSEAS) (2007) 417–423

[11] Scholz, S., Semler, S.C., Breitner, M.H.: Business aspects and sustainability for healthgrids - an expert survey. In Solomonides, T., Hofmann-Apitius, M., Freudigmann, M., Semler, S.C., Legre, Y., Kratz, M., eds.: Proceedings of 7th European Healthgrid Conference, Berlin, Germany, EU DG Information Society (2009) 163–172

[12] Dobrev, A., Scholz, S., Zegners, D., Stroetmann, K.A., Semler, S.C.: Economic performance and sustainability of healthgrids: Evidence from two case studies. In Solomonides, T., Hofmann-Apitius, M., Freudigmann, M., Semler, S.C., Legre, Y., Kratz, M., eds.: Proceedings of 7th European Healthgrid Conference, Berlin, Germany, EU DG Information Society (2009) 163–172

[13] Pitkanen, M., Zhou, X., Tuisku, M., Niemi, T., Ryynänen, V., Müller, H.: How Grids are perceived in healthcare and the public service sector. In Solomonides, T., Silverstein, J.C., Saltz, J., Legre, Y., Krantz, M., Foster, I., Breton, V., Beck, K.R., eds.: Global HealthGrid: e-Science Meets Biomedical Informatics - Proceedings of HealthGrid 2008, Chigaco, USA (2008) 61–69