ISSUES PREVENTING THE MIGRATION OF THE AUSTRALIAN SCHIZOPHRENIA RESEARCH BANK TO THE CLOUD

D. Paul[1,2], F.A. Henskens[1,2], C.M. Loughland[1,2], K. McCabe[1,2], J. Bridge[1,2], L. Duffy[1], V.J. Carr[1,3], S.V. Catts[1,4], A. Jablensky[1,5], P.T. Michie[1,2], B.J. Mowry[1,4], C. Pantelis[1,6], U. Schall[1,2], R. J. Scott[1,2]

[1]*Schizophrenia Research Institute, Sydney, NSW, Australia*
[2]*University of Newcastle, Callaghan, NSW, Australia*
[3]*University of New South Wales, Sydney, NSW, Australia*
[4]*University of Queensland, Brisbane, QLD, Australia*
[5]*University of Western Australia, Perth, WA, Australia*
[6]*University of Melbourne, Melbourne, VIC, Australia*

ABSTRACT

The Australian Schizophrenia Research Bank (ASRB) is an eResearch project that aims to facilitate scientific research in one of the most perplexing mental disorders facing researchers today. There has been a recent push to move the ASRB systems into an Australian research cloud. This paper describes the issues that are currently preventing such a migration.

KEYWORDS

Cloud, Schizophrenia, ASRB.

1. INTRODUCTION

Schizophrenia is a mental disorder characterised by cognitive, perceptual, and affective dysfunctions (Os & Kapur, 2009). The cause is unknown, though research suggests that a combination of genetic, environmental, and social factors all contribute (Mueser & McGurk, 2004). The burden of the illness upon society in general, and upon sufferers and their carers in particular, is extremely high.

The Australian Schizophrenia Research Bank (ASRB) is an Australia-wide collaboration of scientists that aims to facilitate scientific research into schizophrenia by giving access to a large set of data from both people with schizophrenia and healthy controls. This includes the collection and storage of clinical interviews, brain scans, and blood samples that are then made available for approved researchers to use in their work. The project currently includes data from over 1500 participants.

The ASRB system is currently implemented as a centralised server available over the Web. Clinical interviews are collected "in the field", using laptops in locations convenient to the participants (where Internet access may not be available), and later securely uploaded to the central server. All other information is entered directly through the Web interface. Similarly, data are extracted from the system via the Web.

It has been suggested that the ASRB should be moved from the current central server to an Australian research cloud. Cloud computing is an emerging paradigm that provides computing resources as a service. Depending on the service provided, cloud computing can be categorised as either Infrastructure as a Service (IaaS), Platform as a Service (PaaS), or Software as a Service (SaaS). This paper will concentrate on IaaS, where the service provided is a set of virtual machines that can dynamically scale depending on the requirements of the system being run in the cloud.

There are currently a number of issues preventing the ASRB from moving to an IaaS cloud system. This paper examines some of those issues. Section 2 describes the ASRB as it is currently implemented. Section 3 then details the benefits that would be provided by moving the ASRB to the cloud. Section 4 identifies the blocks that currently make such a move impossible. Finally Section 5 concludes the paper, summarising the problems preventing this medical eResearch project from moving to the cloud.

2. THE AUSTRALIAN SCHIZOPHRENIA RESEARCH BANK

Schizophrenia affects approximately 0.3-0.7% of the population (Os & Kapur, 2009). Despite this apparently low prevalence, schizophrenia has a very high burden upon society when considered from the viewpoints of disability-adjusted life years, and the flow-on effect to family, friends and supporters of the sufferer. Patients with the illness often suffer through hallucinations, delusions, and social withdrawal. Most patients can live fairly independently outside of hospitals, thanks to advances in antipsychotic drugs and psychosocial treatments, though they often require some financial or daily living support.

Schizophrenia research suffers because of the high expense of collecting the data required for robust results. The cause of schizophrenia is unknown, but genetics, neurobiology, and psychological and social processes all seem to contribute (Mueser & McGurk, 2004). Schizophrenia research thus requires the collection of information such as DNA samples, brain scans, and clinical information, all of which have a cost for collection and storage. However, many schizophrenia researchers cannot afford to collect large amounts of data, limiting the significance of any findings discovered through their work.

The ASRB allows scientists to analyse vital genetic and brain information from schizophrenia sufferers on a scale never before possible. This is achieved by collecting a large set of inter-connected data from both people with schizophrenia and matched healthy controls. The data collected include blood samples, MRI brain scans, and intensive clinical interviews to provide genetic, neuropsychological, and clinical information about each participant. Researchers from all around the world are then able to apply for access to the data relevant to their area of study, rather than having to collect the data themselves.

The IT systems of the ASRB are used to manage all electronically storable components of the collected data. This includes answers to each question in the clinical interview, copies of the MRI brain scans, and meta-data describing the collected physical samples. The system also tracks each participant in the system, allowing researchers to request existing participants' involvement in new studies rather than having to directly recruit new participants on their own.

The ASRB server uses an enterprise open source portal and provides portlets for each component of the system. The system includes components for: managing participant intake; conducting clinical interviews; storing brain scans and blood samples; and searching for available data based on particular criteria. Most of the data collected for the ASRB are entered directly into the server portal. A notable exception is the clinical assessment information, which is typically collected on a laptop running a cut-down version of the ASRB portlets, and later uploaded to the central server. The reason this information is collected offline is because Internet access is not always available in the locations where the interviews are conducted.

Security is paramount to the ASRB system. The information stored includes private medical information, personal details from throughout the lives of the participants, and identifying information including each participant's contact details. It is vital that all users of the system have access only to the data they require and for which they have ethics approval (from both the user's own institution and the ASRB Access Committee). This high level of security is achieved by assigning roles to each user, and strictly defining and enforcing the permissions afforded to each role.

3. BENEFITS OF THE CLOUD

Cloud computing is the evolution of previous Internet technologies to support the dynamic provisioning of services to clients via a network (Wallis, Henskens, & Hannaford, 2010). IaaS clouds deliver hardware and associated software as a service, allowing client applications to execute on hardware managed by the cloud host (Bhardwaj, Jain, & Jain, 2010). An application deployed to an IaaS cloud can scale dynamically, only using the hardware resources required to meet the system's current needs. Such systems provide many potential benefits for projects like the ASRB, including the removal of a central point of failure, and potential performance increases and cost improvements (Armbrust, et al., 2010).

Currently, it is vital that the central ASRB server is functioning correctly. The entire system is reliant on that server being available, and any failure results in all users being unable to access or modify their data. Cloud computing can reduce the risk of downtime by monitoring the status of the system and automatically creating a new instance of the ASRB server whenever a problem is detected. The newly created system can then handle all responsibilities of the main server, ensuring that the overall system remains accessible.

Further, in an IaaS cloud, multiple instances of the central server can exist simultaneously. These instances could be created whenever the load of the system is too high to be handled by the current instances, and any new requests could then be directed to the least busy instance. Instances could also be destroyed whenever the system's load falls below a certain threshold, ensuring that only resources required by the system are being used. This dynamic scaling would improve the accessibility of the site, ensuring it remains responsive regardless of the number of concurrent users requesting services from the system.

While the location of cloud services is meant to be invisible to clients, many current offerings include settings that specify some geographic information. This allows clients to suggest where the servers in their IaaS cloud should be instantiated. By utilising this information, it is possible to locate the dynamic servers closer to the source of client requests. For example, positioning an instance of the ASRB server on the east coast and another on the west coast of Australia may improve access time for researchers around the country, rather than requiring all requests to go through a single location.

Another benefit of cloud computing is that it can be used to reduce running costs. Since the system can create and destroy instances based on current load, only the cloud provider resources that are actually required need to be paid for. This is in contrast to a central server, where the system must be able to handle peak load, even if the average load is far lower. Further, in cloud systems, much of the maintenance of the system is the responsibility of the cloud provider, rather than local IT services, freeing local IT time for other issues.

4. PROBLEMS WITH THE CLOUD

The ASRB stores medical information and contact details of participants, as well as other information that must be kept private. The storage and provision of access to this data is governed by very strict ethics approvals from each of the institutions that are a part of the project. The current approvals allow storage on a centralised system under the control of ASRB staff. To permit migration of ASRB systems to a cloud, amendments to the current approvals would be required, and would need to be agreed upon by all involved ethics bodies. The need for agreement between multiple independent ethics bodies would make any alteration of approvals a lengthy process, and it has been indicated that movement of ASRB systems so they fall under the control of a cloud provider would prove particularly difficult.

One of the main issues that would make ethics approval difficult is a lack of transparency in regards to data storage. When a system is moved to the cloud, while the client may give some indication of where the systems should be stored, it is up to the cloud provider to determine the physical location of the virtual servers for the system. Further, the system could dynamically migrate from one physical location to (potentially) multiple other physical locations depending on use of the system and discretion of the cloud provider. Such uncertainty of the location makes it extremely difficult to specify exactly where the data is stored and who has access to the system at any particular time. Guarantees could be provided by using techniques such as homomorphic encryption, but current implementations are impractical (Naehrig, Lauter, & Vaikuntanathan, 2011), negating the performance benefits that moving to the cloud should provide.

As well as a lack of transparency, cloud computing is not well defined, making it difficult for the various ethics boards to understand. Even if the members of a particular board believe they understand cloud computing, that understanding may be different to the cloud solution being suggested for the ASRB. Thus a very clear definition of how cloud computing would be used by the ASRB would be required to ensure that all approvers had a similar vision of the system without any incorrect preconceived ideas. However, if the definition were too precise, any cloud-based solution would not be able to evolve alongside cloud technologies without requiring another change to ethics approvals. Creating a balance that ensures understanding while allowing future development, and having those changes agreed upon by multiple approval bodies would prove difficult.

This difficulty is exacerbated by the fact that cloud computing is relatively new, and thus providers do not have a proven track record that could be used as proof of suitability for use with the ASRB. In contrast, the current system is working correctly and has continuously evolved to meet the ASRB's requirements since the Bank's inception. A cloud-based solution would be more flexible than the current central server, but as long as the current system is able to perform the required operations, many involved with the ASRB see little reason to change.

This is particularly true because the ASRB system has not been designed to run in the cloud. The system would require modification to support the dynamic scaling and location-based performance improvements that cloud computing could provide. The underlying database, for example, would require replication to gain any benefit from supporting multiple instances of the server running on different sides of the country. Replication then has a host of problems of its own (Wiesmann, Pedone, Schiper, Kemme, & Alonso, 2000), such as ensuring data is current, and handling the case where different data copies are modified concurrently.

Further, even if the system were redesigned to support cloud technologies, it would not completely replace the need for local IT systems. It is possible, no matter how unlikely, that a cloud provider could go out of business and stop supporting the system. While de facto standards, such as Amazon's API (Grossman, 2009), suggest that migration from one cloud provider to another would be possible, such a move would require a local copy of the system to be available. Thus, local infrastructure to store backups would still be required. Since the primary purpose of the ASRB is the collection, storage, and dissemination of data, the resources required to store such backups provide most of the requirements for hosting the remainder of the ASRB system. Thus the cost savings promised by the cloud would be greatly reduced, especially considering that the server running the system has already been purchased and is currently in use.

5. CONCLUSION

Cloud computing could offer many benefits to the ASRB project, such as the removal of a central point of failure, and the potential for reduced cost and more responsive systems. However, there are numerous issues preventing a migration to the cloud. Primarily among these is that current ethics approvals do not allow such a migration. Any change to ethics approval would require agreement from multiple ethics boards. Since cloud computing is new and relatively unknown, and details such as where data are physically located may be impossible to obtain for a cloud environment, eliciting agreement would be difficult. Further, since the hardware to run the current system has already been purchased, and the system has not been designed to run on the cloud, many of the benefits promised by the cloud are unobtainable, or would require extensive redevelopment to utilise. Thus migration of the ASRB system to the cloud is currently unfeasible.

ACKNOWLEDGEMENT

This work was supported by Schizophrenia Research Institute, utilising infrastructure funding from the NSW Ministry of Health.

REFERENCES

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., et al. (2010). A view of cloud computing. *Communications of the ACM, 53*(4), 50-58.

Bhardwaj, S., Jain, L., & Jain, S. (2010). Cloud computing: A study of infrastructure as a service (IAAS). *International Journal of engineering and information Technology, 2*(1), 60-63.

Grossman, R. L. (2009). The Case for Cloud Computing. *IT Professional, 11*(2), 23-27.

Mueser, K. T., & McGurk, S. R. (2004). Schizophrenia. *Lancet, 363*(9426), 2063-2072.

Naehrig, M., Lauter, K., & Vaikuntanathan, V. (2011). *Can homomorphic encryption be practical?* Paper presented at the 3rd ACM workshop on Cloud computing security.

Os, J. v., & Kapur, S. (2009). Schizophrenia. *Lancet, 374*, 635-645.

Wallis, M., Henskens, F. A., & Hannaford, M. R. (2010). *Component Based Runtime Environment for Internet Applications*. Paper presented at the IADIS International Conference on Internet Technologies & Society (ITS 2010).

Wiesmann, M., Pedone, F., Schiper, A., Kemme, B., & Alonso, G. (2000). *Database replication techniques: A three parameter classification*. Paper presented at the 19th IEEE Symposium on Reliable Distributed Systems.