Linking eScience capabilities for eSocial Science communities: extending the UK-Australia INWA grid to the Chinese Academy of Sciences

A.D. Lloyd[1,2], K. Nan[3], D. Qian[4], T.M. Sloan[5], Y. Sun[2] and B. Yan[3]

[1]Management School, The University of Edinburgh, Scotland

[2]Curtin Business School, Curtin University of Technology, Western Australia

[3]Computer Network Information Centre, Chinese Academy of Sciences

[4]School of Computer Science & Engineering, Beihang University, Beijing, China

[5]EPCC, The University of Edinburgh, Scotland

Ashley@curtin.edu.au

Abstract. The INWA Grid project connects grid resources in EPCC (Edinburgh, Scotland), Curtin Business School (Perth, Western Australia) and the Chinese Academy of Sciences (Beijing, China). The project has demonstrated the application of grid computing power to the analysis of distributed data drawn from collaborating companies in each time zone. This data describes regional behaviour within the "e" economy and provides a focus for global eSocial Science collaborations. The infrastructure supporting these interactions however is designed for eScience collaborations and hence differences with the models for global collaborative research within the social sciences, as well as regional differences in access to this infrastructure by the social science community, determine the immediate extent to which global grid technologies can support global eSocial Science. We explore this issue in an examination of the recent evolution of eScience/Grid infrastructure in China as well as broader measures relating to Internet access within China. We observe that some divergence in the uptake of access technologies is already evident within China with potential consequences for future modalities of collaborative social science research in this region.

Introduction

The INWA project is one of the Grid Demonstrator projects supported by the UK Economic & Social Research Council (ESRC). In December 2003 it established grid interoperation between Edinburgh Parallel Computing Centre (EPCC) at The University of Edinburgh and Curtin Business School in Western Australia. A year later it had replicated the grid resource in BeiJing at the Computer Network Information Centre of the Chinese Academy of Sciences (CAS). In January 2005 it successfully linked these Grid resources in each country together.

An objective of the project was to evaluate the use of existing Grid technologies to re-aggregate data sampling real consumer behaviour in markets that globalisation has tended to disperse. The markets chosen were in the telecommunications and financial services sectors

because of their sensitivity to changes in consumer preferences, one expression of which in competitive markets is a high level of customer turnover or ‘churn’. Analysing this dynamic so that predictive models of behaviour can be built draws on the ability to apply high-performance levels of computing power to very large databases across the Grid, whilst the same infrastructure allows local market insights and expertise to be drawn into the analytical process, reflected in output models with a higher degree of predictive validity. Such models, that explain and predict the behaviour of consumers of a product or service, allow companies to improve the management of customer relationships for existing products, and innovate in response to the emergent demand for new products and services that the analysis also identifies. It is this "value-add" to the industry partner that is needed to access proprietary data samples of different perspectives on the same region that can be combined with publicly available data to build a much clearer picture of regional behaviour within the "e" economy.

The link between Australia and China provides an important insight into market dynamics in East Asia under conditions of high growth. Perth and BeiJing share a common time zone that accounted for 2/3 of world GDP growth in 2003 and had the highest proportional increase in investment of any region in the world. Since CAS manages the entire .cn domain and traffic between government, industry and academia across a common network, the project also provides an interesting insight into a unique information infrastructure for innovation.

Building global collaborations based on global standards

Standards directly support the creation of shared infrastructures to support collaboration and the exchange of ‘value’ that, in turn, drives further investment in and wider access to that infrastructure. In the global Grid world however, whilst standards have emerged they have yet to converge sufficiently to provide a “satisfactory basis to construct a robust, international Cyberinfrastructure” (Hey, 2005). In the case of the INWA Grid, the wide range of core grid technologies needed to achieve the required functionality meant that interoperability could only be guaranteed through a process of replicating the same operating environment at each node. The technical challenges of interoperation are described in the companion paper to this article by Sloan & Lloyd, however once these have been solved, the range of eSocial Science collaborations that can be supported depends on the match between the eScience/Grid architecture and the configuration of distributed communities of practice (Lave & Wenger, 1991) that have emerged as a result of existing models for collaboration and the allocation of research resources at a local, national and international level.

Forbes and Abrams (2004) highlight this in their analysis of an investigation into international social science research commissioned by the ESRC and conducted by the Academy of Social Sciences: “most social science research is done within national and local boundaries, and, most often by individual scholars rather than the research teams which populate medical and natural science research.” They also note that whilst innovative social science research often does make reference to global discourses, this kind of “cutting edge” research does not constitute the bulk of research reported in the social science major journals. They conclude, however, that ideal conditions currently exist for the proliferation of ‘big social science’, but that such scale does not require the major infrastructure of an organization like CERN, rather “a much more distributed human infrastructure of researchers in multiple locations…. responding to big research questions by providing data about how very local phenomena articulate with more global phenomena,” raising a question over the degree to which a global infrastructure for eScience can support international social science research. In the case of the INWA grid extension to the Chinese Academy of Sciences, this

issue was explored through a review of the evolution of Grid infrastructure and wider Internet access within China.

The ‘Scientific’ Grid and the ‘Enterprise’ Internet in China

Grid research in China started in late nineties. Between 1999 and 2001, more than ten institutions participated in the National High Performance Computing Environment (NHPCE) Project with funding from the National 863 Program. A Grid system that connects eight major super computing centres was established.

The network infrastructure around which these developments are being built is exemplified by CERNET (China Education and Research Network) and CSTNet (China Science and Technology Network). China’s first high speed next generation internet project NSFCNET was initiated in 2000 in Beijing. In 2001 CERNET proposed country wide next generation internet development plan – CERNET 2, which is now an important participant of CNGI (China Next Generation Internet) program, the world’s largest planned IPv6 network. CERNET already connects over 900 education and research institutions, 1.2 million PCs and 8 million users, and will be extended to incorporate a further 200,000 schools with over 175 million pupils.

Since 2000, a series of strategic studies on Grid technology were conducted by the Ministry of Science and Technology. Experts were drawn from different industries to discuss possible implementations of Grid technology. These efforts resulted in the 863 High Performance and Grid Computing project and the CNGrid (China National Grid), to continue the NHPCE Project.

In December 2002, ShangHai Municipality initiated an E-Institute project to establish a ‘super virtual’ research institute based on Grid technology. In July 2003 an agreement was signed with IBM China to build ShanghaiGrid to connect all supercomputers in the metropolis.

In September 2003 the first conference of the China Grid Forum took place. In October 2003, the Ministry of Education (with IBM) launched the ChinaGrid Project, aiming to fully utilize the collective resources of CERNET in a Grid with 15000 GFLOPS computing power, connecting the top 100 higher education institutions in China.

These communities supported by these initiatives overlap significantly and develop cooperatively. Within CAS the focus is currently the Scientific Data Grid (SDG) being built to share data resources and enable collaboration. On top of the SDG will be the China Science Grid to integrate scientific instruments and computing resources and establish application grids that operate across this common infrastructure, such as the Virtual Observatory, Bioinformatics Integration, Resources and Environment Monitoring. These application grids parallel eScience developments in other parts of the Asia-Pacific region, the EU and US and provide obvious points for establishing wider international scientific cooperation.

Extensions of scale through links with TEIN2 and GLORIAD, and consequent improvements in links to the EU and US, offer new opportunities for collaboration between East Asia and the EU and US respectively. These will underpin major international collaborative projects

such as the Square Kilometre Array, the largest radio-telescope in the world requiring the worlds fastest computer and data storage facility to be located in the southern hemisphere.

Such international scientific instruments provide an important focus for developing the architecture of an international 'Cyberinfrastructure' of the type described by Hey (2005) for scientific computing, however access by "a much more distributed human infrastructure of researchers in multiple locations" (Forbes and Abrams, 2004) is more likely to be enabled by developments in the enterprise computing community, where over 70% of computing investment occurs. Here it is instructive to look at broader developments in the evolution of China's Internet and adoption of access technologies captured by CAS for a domestic market that is comparable in size to the whole of the 30 countries comprising the OECD.

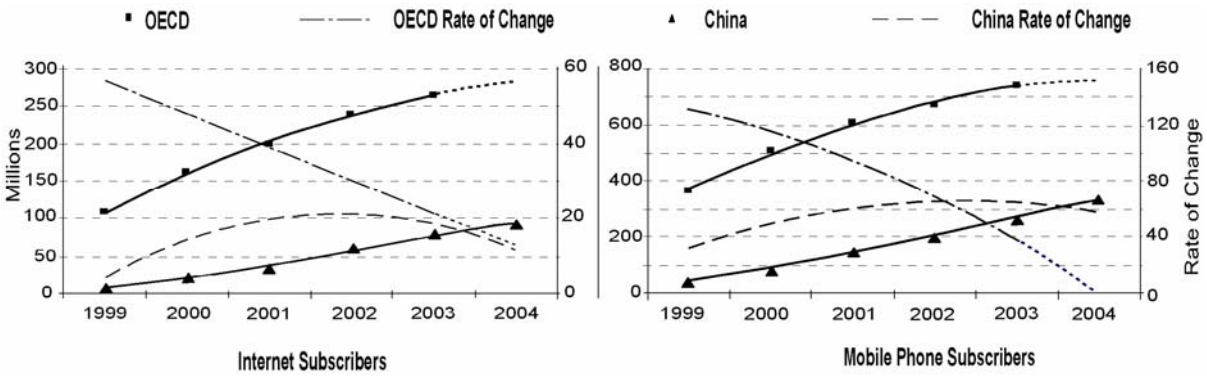

Figure 1. Comparison of the rate of change of Mobile and Internet subscriptions in OECD countries and China. Data sources: OECD Communications Outlook 2005 & Chinese Academy of Sciences (2005).

Figure 1 shows a strong decline in the rate of uptake of mobile telephone and Internet subscriptions in OECD countries, pointing to a maturing market. Though current uptake levels are approximately three times higher in OECD countries overall, the rate of growth in both product groups in China show the beginnings of a similar decline. In the case of mobile telephones this is compatible with a declining 'average revenue per user' reported by China Mobile in 2005 of its 203 million subscribers. When compared, these figures indicate that the relative level of mobile telephone adoption in China may become significantly higher than in OECD countries, with potential consequences for eSocial Science collaborative technologies.

Conclusion

This paper has abstracted from a larger review of the evolution of eScience/Grid capabilities and standards in China, and uptake of established Internet technologies. Whilst international cooperative research and development models established for eScience are already in evidence, the abstraction from the data recorded by CAS exemplifies one of a number of diverging measures of wider network access with potential consequences for the design of collaborative technologies supporting global research communities in the Social Sciences.

Acknowledgments

The authors gratefully acknowledge funding by the ESRC (award RES-149-25-0005), Sun Microsystems for hardware/support in Beijing and Perth and AARNet for network support.

References

Forbes, I. and Abrams, D. (2004): 'International social science research: craft industry or baby behemoth?', *International Social Science Journal*, vol. 56 no.180, pp. 189-192

Hey, T. (2005): 'Tony Hey on the need for Web services standards', *Grid Today*, vol. 4, no. 1, January 2005.

Lave, J. and Wenger, E. (1991): *Situated Learning, Legitimate Peripheral Participation*, Cambridge University Press, Cambridge.