The potential of the GRID for small scale GIS use: a proposal from the UK

J. Swindell, Royal Agricultural College; J. Lee, Oxford Brookes University and A. Berardi, Open University.

Correspondence address: Royal Agricultural College, GL7 6JS, UK; julian.swindell@rac.ac.uk

Abstract

The GRID is seen as the next order of magnitude in the development of the internet. Its purpose is to give participants access to bigger databases and more powerful software than could be run on a single computer. A recent meeting at Cambridge University in the UK sought to establish ways in which the GRID could be used by the GI Science community. Several of the participants, all teachers and researchers at UK universities and Colleges, were dismayed that the general approach seemed be to give even more data and power to those who already had enormous quantities of both. It was very much a “Big Science” approach. Whilst not condemning this desire, we felt that the real power of the GRID would be to provide access to these resources to those who would otherwise be excluded.. This might be small university departments, student researchers, developing countries and voluntary agencies. What we thought was really needed was access to the power of the GRID for the small scale player. We don’t need the whole GRID, just some of it. The term ‘GRIDlet’ was adopted to reflect this change of scale. This paper sets out the development of this group’s thoughts as they have evolved between that Cambridge meeting in Dec 2003 and the EUGISES conference in September 2004.

Introduction

“It is not enough for a handful of experts to attempt the solution of a problem, to solve it, and then apply it. The restriction of knowledge to an elite group destroys the spirit of society and leads to its intellectual impoverishment.”

Attributed to an address by Albert Einstein at Caltech, 1931.

Source: Alice Caprice (ed.) 2000. The expanded quotable Einstein, Princeton University Press, Princeton.

What is the GRID?

Grid computing is a relatively new field, which describes a distributed computing infrastructure. Whilst it incorporates other technology trends such as Internet, distributed and peer-to-peer computing, the specific focus of grid computing has been defined as *the coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organisations* (Foster *et al.*, 2001). The resources shared are not primarily files, but software, computer processors, data, storage elements and sensors as required by a group of individuals and organisations in a dynamic problem-solving environment. The resource owners control what is shared, how it is shared, and by whom and the group of organisations and/or individuals sharing services[1] across computer networks in this way is called a **Virtual Organisation** (VO).

These VOs can superficially look like early mainframe plus dumb-terminal computer systems, where a user booked time on the mainframe and then accessed it through a dumb terminal, which had little or no processing power of its own. The difference is that the mainframe and its linked terminals were all part of a single closed and managed system. A VO would be made up of a shifting partnership where members would come and go, supplying processing power, connectivity, data collections or system management responsibilities. Some may be permanent members, others may only join for specific tasks. Some

---

[1] Resources and services are used interchangeably, though services, especially grid services are the norm now that the Global Grid Forum has adopted the Open Grid Services Infrastructure.

members may belong to several VOs thereby growing the whole into a SuperVO and ultimately the GLOBAL-GRID (although this may be too limiting a term as the various devices now trundling about on Mars are very likely to want to join in).

The Virtual Organisation

Virtual Organisations (VOs) can vary greatly in their size, duration, structure, resource demands and scope; they are **highly scaleable**, yet there are commonalities in terms of their requirements. Specifically, there is a need for some or all of the following: highly flexible resource sharing structures; sophisticated levels of control over resource sharing; diverse usage modes and delivery of consistent and thus, predictable resources. Currently, there is no single distributed computing technology offering a complete solution; VOs are flexible, dynamic and heterogeneous, whereas existing stand-alone solutions are not. Grid computing aims to complement existing technologies to provide the flexibility and control of the diverse resource sharing needs within VOs in such a way as to guarantee a consistent quality of service.

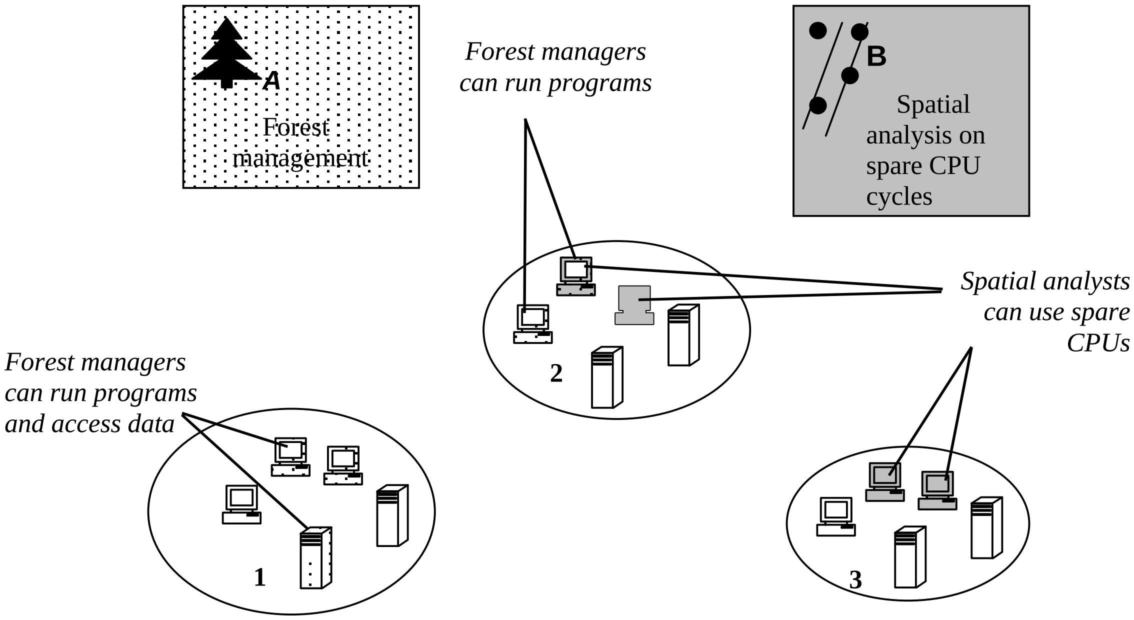

Figure 1 Three actual organisations (represented by the ovals 1, 2 and 3) may participate in one or more VOs (represented by the squares forest management and spatial analysis on spare CPU cycles) by sharing some or all of its resources. Organisation 1 participates solely in forest management, organisation 3 is involved only in leasing spare computing power, whilst organisation 2 in involved in both VOs. (Adapted from Foster et al., 2001).

The Role of Middleware in the GRID

Grid computing has been likened to an electricity grid; power is available at the flick of a switch yet the user neither needs to know nor cares where or how the electricity was generated, stored, transmitted and converted. The ultimate ideal of grid computing aims towards the same seamless delivery model. However, this requires new approaches and architectures, specifically in the area of *middleware*.

Middleware is used to describe separate products that serve as the glue between two applications; it can be thought of as the plumbing connecting both sides of an application, passing data between them. Grid middleware aims to support dynamic resource allocation to enable efficient sharing and reuse of IT infrastructure, whilst delivering consistent response times and high levels of resource availability. Fundamental to this is the need for: (i) interoperability to enable resource sharing across platforms, languages, organisational boundaries, operational policies and resource types and (ii) protocols to control the interaction between both VO members and the resources being used. Without these, VO formation and operation is all but impossible.

How can the GRID work with GIS?

GIS programs are powerful tools which can access, manipulate and analyse enormous spatial datasets. All of this work could be done, and is done, on single computers or local networks. The GRID, in its potential developed form, could allow the sharing and distribution of any and all of these resources and activities. A large, national scale dataset could remain at the location of its owners or managers, but be accessed by any authorised user for their own purposes via the GRID. Departments with occasional need for complex analysis of spatial data could use the resources and services of large scale users without having to become owners of such resources themselves. They might not even need GIS software, but would access spatial modelling resources supplied as web/grid services from a remote source.

Examples of proto-GRID GIS is the use of Internet map server systems, where a user can define a particular output they require from a central GIS system, and this is delivered to them in the required format. Typical examples would be simple place finder systems such as Multimap in the UK (www.multimap.com). A more sophisticated example is the UNEP Global Resource Information Database (also confusingly called GRID) Nordic/Baltic information resource, where the user can specify data layers, extents and combinations (www.grida.no). However, the end user does not have direct access to the original data, and no genuine analysis is possible. A developed GRID-GIS would permit both of these.

The case for the small scale user: GRIDlet and GRIDlet-GIS

When I first worked with GIS I used a program, *Idrisi,* which had an absolute limit on image resolution of about 1012 columns. The whole program fitted onto two floppy disks and the surprisingly powerful analytical routines could be run on an 8 megahertz DOS based PC. Happy days indeed! Now our programs arrive on multiple CD-ROM and datasets are being delivered routinely on DVD. This gives each user staggering power (provided you have upgraded your operating systems and hard disk capacity). But not all users of geo-information actually need this power or this quantity of data. All they really want are the answers to spatial questions. Their questions may be very brief, just one line of text. A suitable answer may be equally brief. The *Idrisi* 1012 column image may be more than adequate. But to arrive at a high quality answer, the best and most comprehensive geo-databases should be searched, queried and analysed by the most powerful and capable programs that exist. A student, for example, may wish to identify sampling sites for an ecology project. She could seek a suitable site on her own College campus, but why shouldn’t she be able to select the site that most suits her or her project, from the whole country? The data exist to answer her query and the costs of giving her access to them could be insignificant. A working GRID should make such databases and programs available to all potential beneficiaries, rather than allow them to be kept for the exclusive, and therefore cost-ineffective use of the, literally, powerful few.

To date, grid computing has focused on large-scale science, technology and engineering applications, rather than the single student with a research project to complete. For example, the Large Hadron Collider at CERN, which comes on-line in 2007, is expected to generate 14,000 Tb of data (equivalent to 20 million CDs), whilst the National Centre for Atmospheric Research in the United States currently holds 1 Pb (1,024 Tb) for the last 100 years of climatic recording effort, with this figure expected to double in one year. These projects come under the umbrella of **e-Science**, which refers to the large-scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists.

However, there is very little in the definition of Grid computing that *necessitates* the development of large scale, global, high performance systems. Indeed, the dominance of large-scale projects seeks to maintain the *status quo* in which the scientific establishment dominates the production and use of knowledge for the benefit of its members. However, we would suggest that whilst e-Science needs the Grid, the Grid does not need e-Science and could, if used judiciously, provide benefits to ordinary people.

This is the **basic tenet** of our group, *that computer Grids ought to empower ordinary people through the process of constructing and using their own knowledge*. With this in mind, we envisage public empowerment through harnessing small-scale computer grids, which we term **GRIDlets**.

We believe that grid(let) computing offers an unrivalled opportunity to further continue the Utopian ideal of democratic cartography. Large volumes of spatial data are currently in existence and are continually being updated. These data are often available for relatively little cost (e.g. satellite imagery) yet they are rarely used to empower ordinary people, often because of the computational and human overheads involved in their storage, transmission and analysis. By harnessing the power of distributed resources, these data can be brought to the people.

The Virtual Market Place

The model of GRIDlet computing encapsulates a network, or *virtual marketplace* (Paepcke et al., 1998), of service providers and service customers; the latter requiring a solution to a spatial problem, the former having a range of services, which may contribute wholly or partly to this solution. Virtual marketplaces (VMs) are characterised by the following: a high degree of provider autonomy; low infrastructure costs; ease of participation; ease of service and data access; scalability in the number of services provided and the breadth of tasks supported. This accurately characterises a spatial problem solving environment similar to that described above, which involves many loosely connected actors, providing and receiving what could be a low cost service, on low-cost hardware.

The VM has considerable parallels with the VO concept at the heart of Grid computing, particularly, the emphasis on collaboration and the provision and use of services. Crucially, there are further parallels with the specific tenets of GRIDlet-GIS, namely: scalability, provider autonomy, low cost and ease of participation. Both VO and VM concepts are applicable to problems involving spatial data in economically developing nations and there is a strong argument for merging these two significant concepts under the GRIDlet-GIS umbrella. The case for merging is stronger, given recent advances in computer-computer communication via web/grid services and software agents (Sodabot, 2004) used in the semantic web. Thus, transactions within the VO/VM could be partly via human interaction and partly via machine-machine communication, the complexity of the task reflecting the degree of machine-machine communication. For example, spatial problem-solving environments are generally highly heterogeneous, characterised by differences in, or a lack of, standards in data and metadata. There are likely to be considerable time benefits to be had from harnessing machine-machine interaction for data conversion using web services. Conversely, complex geographical analysis is a more iterative process, requiring considerable human-human dialogue and interaction.

Liveware for GRIDlets as opposed to middleware for the GRID

GRIDlet-GIS addresses a fundamental issue with current e-Science approaches to the development and implementation of GRID-based technologies for resolving complex spatial problems. We believe that the structured development methods currently adopted by e-Science initiatives are insensitive to social, political and cultural factors. These initiatives assume that the problems are well-defined and objective, yet they concurrently state that these tools could be used for complex problem-solving within the social domain. There is little recognition that many problems, including current environmental crises, could be characterised as social constructs and thus require interpretative research approaches. It is interesting to note that many GRID-tool developers even have problems getting the “objective” scientific community to adopt their products, and there is now recognition that a move towards fluid approaches to tool development and implementation are necessary. Such approaches include Soft Systems Methodology (Checkland and Holwell, 1998) and its offshoots (Multiview: Avison and Wood-Harper 1990; FAOR: Shafer 1988; Client-Led Design: Stowell and West 1994). Yet, considerable investment has already taken place in rigid structures for GRID tool implementation, often referred to as *middleware*.

It has been argued that the development of suitable middleware is one of the key requirements of creating and running large scale grids. It would be unrealistic to propose that such middleware could be economically developed to serve the needs of a small scale user, and this can be seen as one of the prime excluding factors of the grid. We would argue that much of the interactivity between components of a GRIDlet can be mediated by real live people, or ‘liveware’, providing these people find it worth their while to act in this role. An example of this in practice is given below, in the language historian case study. In this case, none of the three people involved actually had to devote much time to the project, but the outcome achieved would have taken any one of them several weeks to complete.

Three examples of real and potential GRIDlet-GIS scenarios

These examples illustrate real and potential GRIDlet -GIS situations. One of them has actually been carried through, the others have not, due to the absence of a virtual market place where suitable data, liveware and applications could be bartered.

The language historian

A linguistics expert at Lancaster University, Jonathan Culpeper, is rewriting his textbook, *History of English* (Culpeper 1997). In it he has a chapter on English place names, which dealt specifically with the distribution of Danish and Norwegian place names in northern and central England. He needed good quality maps showing this distribution. He had found the web site of one of this paper's authors showing some work he had done on this problem (Swindell 2004) and asked for assistance. I could do this but was concerned that my database of names was very small (approximately 5,000 towns). He put me in touch with *Archaeology UK* (2004), who had a free-to-academia database of georeferenced placenames with about 50,000 entries. The contact person at *Archaeology UK* could provide the data, but said they would require about two weeks to write a program to calculate spatial distribution. Neither of these two people knew how to use GIS and one had never even heard of it. However, we developed a little, on-the-fly GRIDlet-GIS virtual market place where:

1. *Lancaster University* provided the details of significant name elements
2. *Archaeology UK* provided the raw geo-referenced data
3. *Royal Agricultural College* ran the data through *ArcGIS* using simple selection and density calculation to generate the required maps (figure 2 is an example)

The time commitment of each party to this VM was probably less than half an hour each, but the result was one that two members could not have produced at all, and one member, the data provider, estimated would have taken him about two weeks to create. The benefit gained by the linguist is clear. The benefit to the data provider was useful dissemination of a publicly funded database and benefit to the processor was acknowledgement in a high quality textbook and extension of a research study which has already generated additional research contacts.

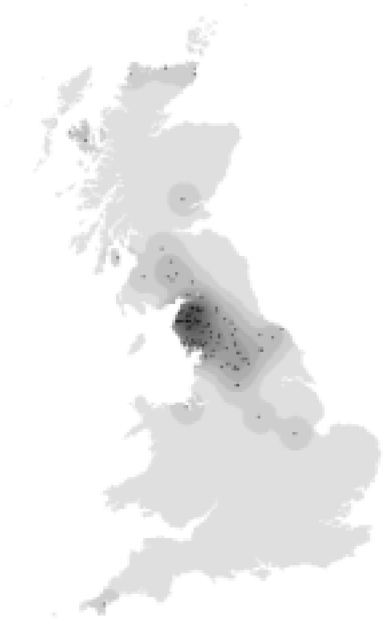

Figure 2 The distribution and *density* of British towns with "-thwaite, -scale, -slack or -gill " in their names (These are Norwegian derivations)

The small farmer in Central America

At the end of the last century (doesn't that sound like a long, long time ago now) one of the authors, Julian Swindell, was a participant in a rural development project in Honduras (Swindell, 1999). The premise of the project was that stakeholders in the Honduran rural economy should be able to benefit from access to an enormous database of spatial information relating to their country. The hope was that anyone, from the Minister of Agriculture down to individual farmers, could access this information through a simple to use interface, to help them in decision making at their scale of interest (figure 3). The data were available and of good quality. The decision support system failed to materialise because there was never a viable virtual market place, as we describe for GRIDlet-GIS. No one was prepared to host the data in an accessible location. The potential end users did not have the skills to process those data, although they could ask piercing and pertinent questions. None of the academics involved had the time to devote to acting as liveware on such a large scale. A great opportunity to benefit a very poor country was lost (although not entirely, as the database proved immensely helpful in planning relief operations after Hurricane Mitch devastated the country in 1998.) The GRIDlet-GIS virtual market place may have offered a solution to this:

1. The database could have been hosted by the Central government as a benefit to their citizens
2. International aid funds could have been used to build a basic, working interface, driven by GIS web servers, acting as the GRID middleware
3. Academics from any university with suitable IT resources could have been funded from the same aid agencies to maintain the interface, as the liveware, which would respond to queries the middleware could not manage.
4. The end users could finally ask the real questions that need to be answered, but are so seldom heard in development projects against the white noise of bickering NGOs, funding bodies and academics trying to fill their publication quotas.

Figure 3 Yorito watershed in Honduras, the site of key case studies, showing relief, drainage and settlements extracted from the national database.

The master’s research student in the UK

This is not a specific case study but an illustration of a growing need amongst non-doctoral researchers. The research carried out by masters students is growing in importance, certainly within a largely vocational environment such as the Royal Agricultural College. Their work is seen as timely and more near-market than that carried out by doctoral students. The latter take many years to produce detailed results which then may only be published, after a long peer review process, in what can only be described as obscure, small circulation journals. At best the results are often out of date, at worse, they have been superseded by developments from non-published commercial research. Masters research differs in that it is carried out over a short time span, usually only a few weeks or months. It is assessed very quickly and can then be published in the professional press almost immediately, if it is deemed of good enough quality. It has not been formally peer reviewed in the academic research definition, but it has been reviewed by the students’ supervisors, and it can then be subjected to criticism by real world end users within, for example, the farming or property management press.

The problems faced by masters researchers are lack of resources, funds and time. What would benefit their research would be easy access to comprehensive databases, powerful analytical software and the skills to use that data and software. They cannot justify purchase of such resources for a masters project, but everyone could benefit from their access to them. The GRIDlet model could be relevant here. Within the UK there are developments to create geo-data portals for use by researchers, but they fall short of access to original data. The *Geo-data Portal Project* (UK Data Archive 2004) aims to give access to meta-data relating to geo-data collections in the UK, but it does not give access to the actual data. The *Landmap* satellite image collection in the UK goes much further in this direction and provides an example of how data can be made directly accessible to end users at very low cost (MIMAS 2004). We have used this resource within the Royal Agricultural College both for teaching and research data. What is needed within this virtual marketplace are service providers who can help the masters researchers analyse these data for their own needs.

GRIDlet-GIS in Practice

Two of the authors, Andrea Berardi and John Lee, are planning to test our proposed model within two further projects:

The Rupununi GRIDlet

This will build upon a DEFRA funded Darwin Initiative project (running from September 2003 to August 2007) which is currently developing socio-ecological models of the Rupununi River catchment, Guyana. This region, the size of S.E. England, is home to the Makushi Amerindian tribe and contains one of the highest diversity of fish-species in the world. The Darwin Initiative project is funding the training in “eco-informatics” of eight community members, some to a post-graduate level. The project is also developing geospatial and socio-ecological databases and models that will incorporate remotely sensed satellite image analysis and GIS. A GRIDlet-based tool is the only realistic option for the implementation of the computation component of the adaptive management plan. The project is currently organising a consortium of service providers, hardware providers and users to establish a GRIDlet-GIS VO. The project concept is a radical one in that it aims to provide Amerindian communities access to a GRIDlet-GIS tool combining internet-enabled, low cost, low energy and rugged computing platforms based on the *Ndiyo!* ultra-thin terminal (www.ndiyo.org) with the computing capacity and spatial problem solving expertise of British universities.

A GRID-based Virtual Practice Environment for Spatial Decision Support.

This will develop a “virtual” case-study that will be used as an open-source teaching resource in the Higher Education sector. The teaching material will use problem-based learning as the focus for guiding groups of students in collaborative on-line problem-solving along the GRIDlet-GIS ideals. The Rupununi GRIDlet project will provide most of the material for this project.

Conclusions

Along with the technical aspects of GRID technology development, we also have to focus on the humanisation of the technology so that the people that really need it can benefit from it directly. A core principle of our approach is that we are fighting against the e-Science tide to divide society by who is and who is not able to interface with IT tools, which will be even more pronounced given the very limited access to large-scale computer resources. The e-Science approach exacerbates a society of the IT have and have-nots. Our challenge is to provide a GRIDlet-GIS tool that delivers spatial problem solving to all. We propose a change of focus from one that is obsessed with technology to one that places human interactions first, including those that are not technologically proficient. At the core of our approach is the development of participative techniques to guide the establishment of GRIDlet-GIS Virtual Organisations and Virtual Marketplaces. This will be achieved through a series of applied projects involving students, teachers, experts and developing world users.

References

Archaeology UK (2004) Placename finder URL http://www.digital-documents.co.uk/archi/uk-gaz.htm (accessed June 1, 2004)

Avison, D.E. and Wood-Harper, A.T. (1990) Multiview: an Exploration in Information Systems Development, McGraw-Hill, Maidenhead.

Checkland, P.B. and Holwell, S. (1998) Information, Systems, and Information Systems, Wiley, Chichester

Culpeper J. (1997) History of English, Routledge, London

Foster, I., Kesselman, C. and Tuecke, S. (2001). The anatomy of the grid: enabling scalable virtual organizations. International Journal of Supercomputer Applications, 15(3) pp. 200-222.

MIMAS (2004) The Landmap Project URL: http://www.landmap.ac.uk accessed June 1, 2004

Morrison, J. L. (1997). Topographic mapping for the twenty-first century. In: Rhind, D. (ed.) Framework of the World. Geo-information International, Cambridge.

Paepcke, A., C.C. Chang, H. Garcia-Molina, and T. Winograd (1998). Interoperability for Digital Libraries Worldwide. Communications of the ACM, (41):4, 33-43.

Shafer, G. (1988) Functional Analysis of Office Requirements: A multiperspective approach, John Wiley & Sons, Chichester

Sodabot, (2004). The SodaBot program MIT
URL: http://www.ai.mit.edu/people/sodabot/slideshow/total/P001.html . Accessed June 1 2004

Stowell, F.A. and West, D. (1994) Client-Led Design: A Systemic Approach to Information Systems Definition, McGraw-Hill, Maidenhead

Swindell J. (1999) Scenario based decision support systems: their use in sustainable development of rural land in Central American hillside regions, proceedings of *ROOTS 99*, RICS; London pp 93-101

Swindell J (2004) English Placename distribution
URL: http://www.royagcol.ac.uk/~julian_swindell/projects/english-placenames/placenames.htm
(Accessed June,1,2004)

UK data archive (2004) The Geo-data Portal Project URL: http://go-geo.data-archive.ac.uk/ (accessed June 1, 2004)

Contacts

Julian Swindell (corresponding author)
School of Rural Economy and Land Management
Royal Agricultural College
Cirencester
GL7 6JS, UK
julian.swindell@rac.ac.uk

John Lee
Spatial Ecology & Landuse Unit (SELU)
School of Biological and Molecular Sciences
Oxford Brookes University
jlee@brookes.ac.uk

Andrea Berardi
Open Systems Research Group (OSRG)
Faculty of Technology
The Open University
a.berardi@open.ac.uk