GridletGIS - a White Paper
First Draft (May 2004)

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

and

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

"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.

Introduction

GridletGIS was established as a result of a workshop held at the National Institute of Environmental e-Science entitled GridGIS (Cambridge, UK, 15th-16th December 2003). Like minded people came together during one of the break-out sessions in response to the perception that much, if not all, of the emerging field of distributed, and particularly grid, computing is concerned with manipulating large volumes of data using high-powered computing facilities connected by high-speed networks. Whilst the grid-computing model is useful, the focus on technical issues clouds the real issue of delivering useful spatial[1] information to as many people as possible. Consequently, GridletGIS felt grid computing had little relevance to the majority of the world's population and was potentially another example of knowledge being restricted to an elite group. Hence, GridletGIS felt that the grid community was missing a golden opportunity to empower those members of society who do not currently have access to geo-information. This opportunity has been given impetus with the advent of low cost hardware platforms offering cheap, 'out of the box' Internet-ready networks suitable for deployment in classrooms and public places.

Grid computing

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 is *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 as well as what happens

[1] Throughout the document, reference to data can be taken to mean spatial data unless stated otherwise. This reflects the areas of interest of GridletGIS, but the concepts described here are not limited to spatial data.

in the event of resource failure. The group of organisations and/or individuals sharing services[2] across computer networks in this way is called a **Virtual Organisation** (VO).

Virtual Organisations

Virtual Organisations (VOs) can vary greatly in their size, duration, structure, resource demands and scope; VOs arise out of mutual trust based on a sense of co-destiny in which the fate of each member is dependent on the other(s). VOs are **highly scaleable** and could be applied to a small cluster of computers in the 'jungle' as to the European Data Grid used by CERN, 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; 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.

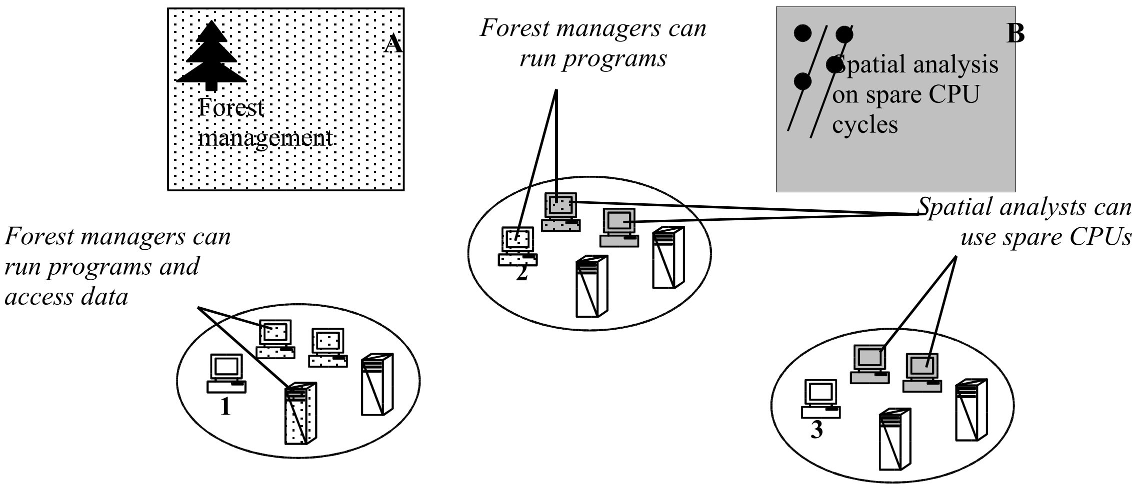

Figure 1. Three actual organisations (represented by the ovals 1, 2 and 3) may participate in one or more VOs (represented by the squares 'forest management' and 'spatial analysis' on spare CPU cycles) by sharing some or all of its resources. Organisation 1 participates solely in forest management, organisation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

Large-scale science

To date, grid computing has focused on large-scale science, technology and engineering applications. 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

[2] 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.

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**.

Grid Middleware

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.

Classically, 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. However, GridletGIS believe that this middleware can successfully comprise **'liveware'** (a.k.a. wetware) or the human beings that operate the systems. Liveware as middleware is an important concept for two reasons: firstly, it reflects the nature of spatial problem solving, which is characterised by a high degree of human intervention and secondly, it is more appropriate to the needs of the vast majority of the world's population, particularly those in economically developing regions – this is the concept of Gridlet computing: distributed computer resources linked by human intervention and interaction. Human beings, rather than complex middleware, can satisfy the fundamental requirements of small scale VOs: (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.

In Gridlet computing VO would arise in response to a problem or a question that needed answering; the members of the VO would include the "affected" who require the resolution of the problem, individuals with access to computing capacity, individuals with access to data, individuals with access to software, and individuals that could facilitate communication between the interested parties. This community would communicate and exchange data via email and the web, and would undertake spatial data collation, analysis and output. Response would be via a personalised email, the web or a message broadcast to a wider audience (village PA, local radio, bulletin board, local meetings etc.).

Grid 'Liveware'

GridletGIS address 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. One of the most notable examples of this drive has been the development of the GRID workflow management system - "Kepler" (URL: http://kepler.ecoinformatics.org/).Kepler is an example of the current obsession with technical computing factors, with little or no regard for organisational issues. This is especially significant when one considers that VOs are at the core of GRID-based applications.

The Virtual Marketplace

The model of Gridlet computing encapsulates a network, or virtual marketplace (Paecke *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 GridletGIS, 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 GridletGIS umbrella. The case for merging is stronger, given recent advances in computer-computer communication via web/grid[3] 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.

Geo-information

Geo-information management and usage has evolved rapidly in the last two decades coinciding with the enormous increase in computer power and the rise of computerised Geographic Information Systems (GIS). This trend reflects humans' desire to categorise and map their surroundings and those of their neighbours to, for example: ensure fair trade and inheritance of land, animals and objects; regulate and enforce divisions of land and property; plan national and regional infrastructures and gain competitive or strategic advantage. Until recently, this powerful knowledge was held largely by the ruling elite, but the advent of GIS has, to some extent, lead to the 'democratisation of cartography' (Morrison, 1997[4]) in that technology has enabled immediate presentation of information as well as interactive query, analysis and visualisation.

[3] A grid service is essentially a stateful web service; i.e. it offers service lifetime management, inspection and notification of service state changes.

[4] Morrison's work referred to the look of the map (the symbology), which had been controlled by cartographer but is now often controlled by the user.

The Internet has played a major part in enabling this democratisation and has dramatically improved our ability to deliver information. In some cases this has lead to greater public participation in decision-making processes and the formulation of *ad hoc* VOs comprising stakeholders, Local and National Government etc. However, the public's ability to access, interpret and use digital information has remained the same or even declined. There are significant barriers to the access of this information, caused not only by limited availability of the necessary IT infrastructure, but also, for example, the issues of data ownership, IT expertise, data standardisation and willingness to succeed.

GridletGIS believe that the focus on liveware as middleware within distributed 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 overheads involved in their storage, transmission and analysis. By harnessing the power of distributed resources, these data can be brought to the people!

GridletGIS in Practice

We are planning to test our proposed model within two projects:

The Rupununi Gridlet
This will build upon a DEFRA funded Darwin Initiative project (£135,000 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 8 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. We are currently organising a consortium of service providers (Open University; Oxford-Brooks University; Royal Holloway, Univ. of London), hardware providers (WSP Environmental Ltd., Newnham Research Ltd.) and users (North Rupununi District Development Board, Iwokrama International Centre for Rainforest Conservation and Development) to establish a GridletGIS VO.

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 GridletGIS ideals. The Rupununi Gridlet project will provide most of the material for this project.

Conclusion

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. The e-Science approach exacerbates a society of the IT have and have nots. Our challenge is to provide a GridletGIS tool that delivers spatial problem solving to the IT have nots.

References

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

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.

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

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

Sodabot, (2004). http://www.ai.mit.edu/people/sodabot/slideshow/total/P001.html.

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