VOSON: A Web Services Approach for Facilitating Research into Online Networks*

Robert Ackland[1], Mathieu O'Neil[2], Russell Standish[3], Markus Buchhorn[4]

[1]The Australian National University, Australia

[2]The Australian National University, Australia

[3]University of New South Wales, Australia

[4]The Australian National University, Australia

Email address of corresponding author: `robert.ackland@anu.edu.au`

Abstract. The Virtual Observatory for the Study of Online Networks (VOSON) is providing a web-based environment facilitating research into networks. While the initial focus of the VOSON Project is on empirical analysis of online networks by social scientists, we aim to eventually facilitate network research in other disciplines such as biology and applied physics. Via a VOSON web portal, researchers can access various storage and computational resources such as existing network data sets, tools for creating and storing new data sets, network analytical methods and network visualisation tools. While these resources are currently provided "in-house", we envisage the involvement of third-party resource providers (primarily, other researchers) who wish to share their data and methods to promote scientific endeavour in their particular areas of network research. Thus, VOSON is an example of using emerging e-Research technologies to promote "open-source" research practice and we see the web services model as a means of bringing together of disparate network research resources, many of which are already well-established. In this paper, we describe VOSON System and our plans (and progress) for incorporating web services into the system. New VOSON-facilitated research into online environmental activists is also presented.

Introduction

The formation and impact of networks on the World Wide Web (WWW) is a topic that has generated much inter-disciplinary research interest in recent years with major contributions from applied physics, library and information sciences (LIS) and the social sciences. Researchers are faced with major challenges relating to data collection (which often involves the use of web crawlers), data storage (web data sets can be potentially huge) and analysis (particularly if network visualisation tools are required). The Virtual Observatory for the Study of Online Networks (VOSON) is providing a collaborative research environment facilitating inter-disciplinary research.[1] The VOSON Project comprises researchers from both the computer and social sciences who are working on the development of new tools for

*Full paper submission to Second International e-Social Science Conference, 28-30 June 2006, Manchester, UK.

1 The VOSON Project was formally established in 2005 under the auspices of an Australian Research Council Special Research Initiative (e-Research Support) grant. See http://voson.anu.edu.au.

research into online networks and the use of these tools in empirical social science research. While the initial focus of the VOSON Project is on social science research, it is envisaged that the VOSON System will eventually facilitate other disciplinary research into networks. In this paper, we present information on the current status of the VOSON System and outline new VOSON-facilitated research into online environmental activist networks. We relate the VOSON System to grid portals that are being developed in e-Science and discuss the role of grid portals in e-Social Science more generally. Finally, we detail our progress in planning and developing the next version of the VOSON System, which incorporates web services to facilitate access to distributed resources such as datasets, methods and computational cycles.

The VOSON System – “production” version

The VOSON System presently incorporates the following open-source software components : linux operating system, PHP/javascript web interface, MySQL database, Perl-based web crawler, data manipulation/analysis in Perl and C++, third-party packages for text analysis and visualisation. Most VOSON System functionality is accessed via a web portal (Figure 1).

Web mining and text mining. A purpose-built web crawler and the Google API[2] are used to identify hyperlinks between web pages. Web datasets can be potentially huge (especially if researching the evolution of online networks) necessitating access to large-scale computational and storage resources. The processing of page text content is conducted using two methods: page meta-data keywords are extracted using the web crawler; page text content is also optionally stored and then parsed using the Lingua::LinkParser[3] Perl module which is a front-end to the Link Grammar Parser software.[4]

Data preparation. While data are collected at the page-level, analysis is generally conducted over aggregations of pages (we are interested in constructing networks where the nodes represent organisations, groups or people, not web pages) and the software provides a method for aggregating web pages into pagegroups or “sites”.[5] Via the DataBrowser (which is a “window” into the underlying relational database, featuring drop-down boxes and text edit fields), the user can manually classify sites according to the particular domain of study.

Data visualisation. Maps showing the shortest path between the root node and all other connected nodes in the database are constructed using the LGL (Large Graph Layout)[6] algorithm of Adai et al. (2004) – see Figure 2a. To visualise all nodes and all links simultaneously we have implemented the LinLogLayout[7] force-directed graphing (FDG) algorithm of Noack (2004, 2005) – see Figure 2b. In the context of our web graph, web sites are given initial random positions and modelled as electrostatic charges (repulsion forces that act to push nodes apart from one another). Hyperlinks between web sites are modelled as springs (attraction forces that act to pull together those sites that are connected to one another via hyperlinks). The algorithm shifts the position of nodes in an attempt to minimise the energy of the system leading to the identification of web clusters or communities - collections of sites that have more links to other members in the collection than to nodes outside the

2 http://www.google.com/apis

3 http://search.cpan.org/~dbrian/Lingua-LinkParser-1.09/LinkParser.pm

4 http://www.link.cs.cmu.edu/link/

5 See Thelwall (2004) on “alternative document models” based on directories, domains and multi-domain sites.

6 http://apropos.icmb.utexas.edu/lgl/

7 See http://www.informatik.tu-cottbus.de/~an/GD/ for more details.

collection. With both the LGL and FDG maps, node attributes can be displayed and edited (by selecting a node) and thus the maps can be used for both data preparation and analysis.

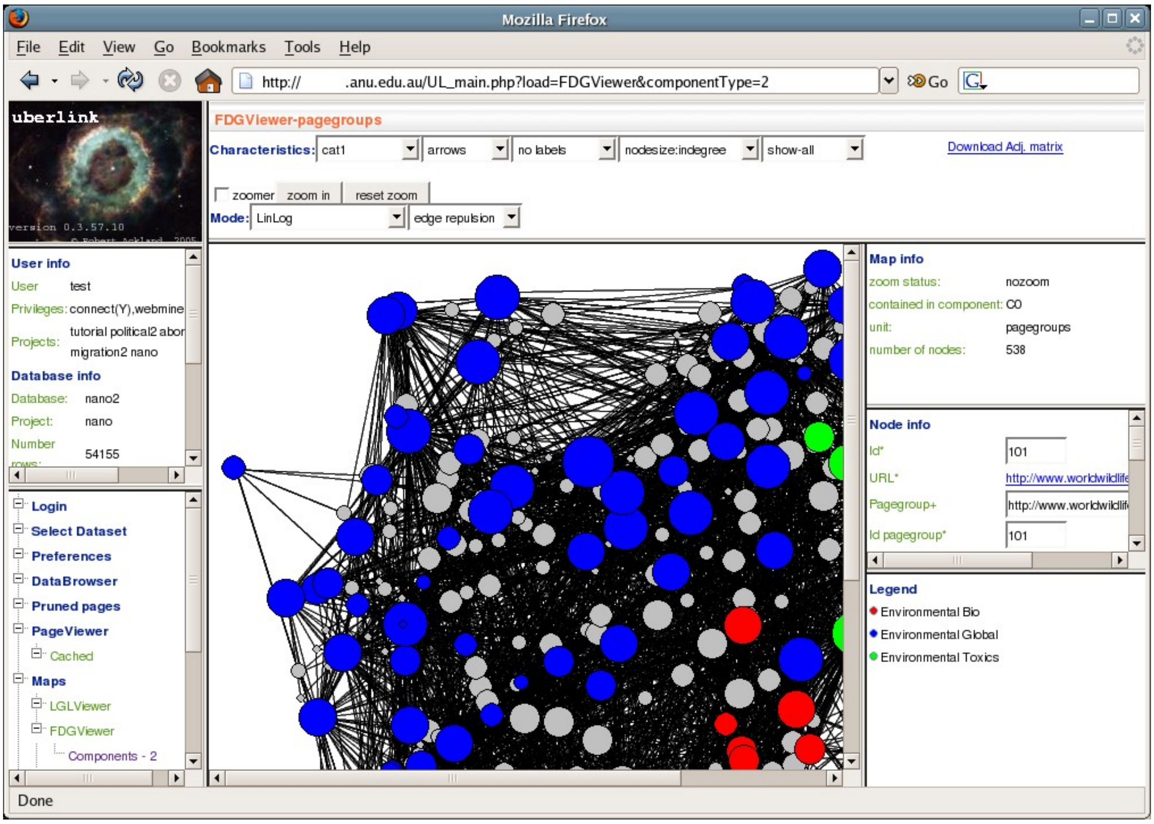

Figure 1: VOSON System web portal

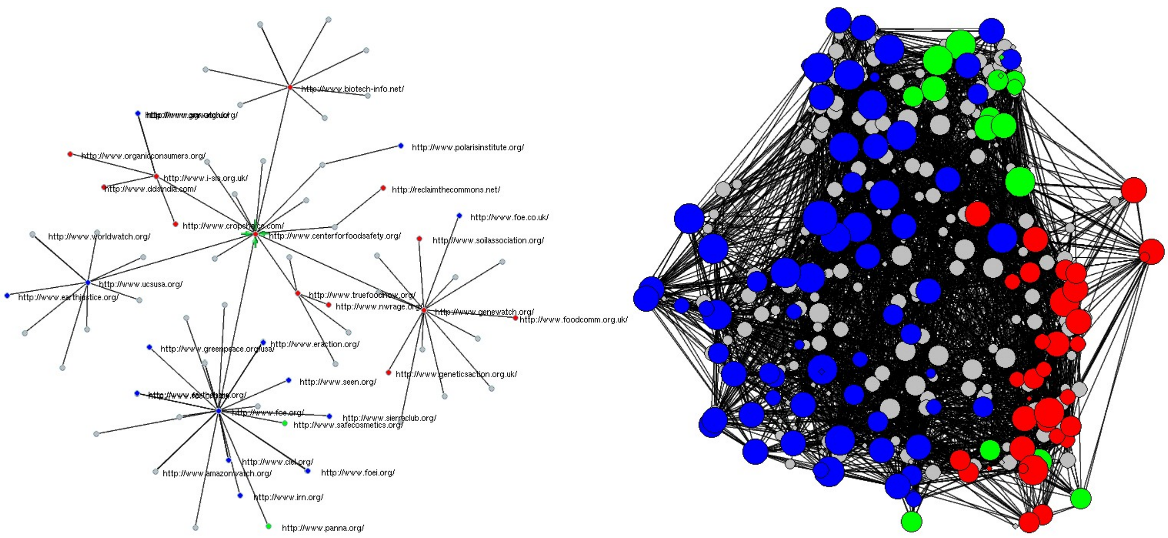
Figure 2: (a) Minimum paths from www.centreforfoodsafety.org, (b) LinLogLayout FDG

Data analysis. To date, the main focus of VOSON System development has been on data collection/preparation, visualisation and basic cross tabulations rather than supporting sophisticated social network analysis techniques. However, VOSON datasets can be downloaded and analysed using standard social network analysis software. Additional analytical capabilities are being built into the VOSON System on an ongoing basis.

New research into online environmental networks

The VOSON System is being used in several research projects.[8] In this section, we briefly outline new research into cultural and organisational aspects of the online networking activities of environmental activist groups.[9]

The VOSON System was used to construct an activist connectivity database (ACD) which records the URLs of the web pages that either link to or are linked to by a given set of 162 activist group homepages (known as the "seed set"). Our primary manual classification of the seed sites placed organisations in one of three groups identifying their primary focus. Organisations dealing primarily with issues such as climate change, forest and wildlife preservation, nuclear weapons, and sustainable trade were grouped as "environmental-global". Organisations focused on pollutants and issues of environmental justice were grouped as "environmental-toxics". Finally, organisations dealing primarily with genetic engineering, organic farming, biopiracy and patenting issues were grouped as "environmental-bio". Of our 162 seed sites, 92 were classified as environmental-global, 47 were classified as environmental-bio and the remaining 23 are environmental-toxics.

Collective Action Frames. We argue that environmental activist groups use collective action frames to distinguish themselves from one another and communicate their vision of the world. We use content analysis of the homepages to identify "diagnostic frames" (used for problem identification and attribution) and "prognostic frames" (answering the question: what is to be done?).[10] Using frequency counts of word colocations, we determined that the diagnostic frame for the environmental-global group is articulated around the central notion of the global environment, which takes many forms such as climate/change (22), the impact on public/health (5), natural/resources (4), ancient/forests (4) and air/quality (3) and the existence of extinction/hotspots (3).[11] These problems stem from oil/addiction (5) and greenhouse/emissions (4). The prognostic frame is to take/action (10) by supporting local/communities (6), civil/society (5), protected/areas (5), indigenous/peoples (4) and to reduce/emissions (3) so as to improve air/quality (3). This will be achieved through soliciting/donations (3) so as to finance/programs (3), helping to find/solutions (3). The identified framing strategies for the environmental-bio and environmental-global groups were similarly both congruent and consistent with our *a priori* beliefs about these groups.

Outbound hyperlinks. By creating outgoing hyperlinks actors define who is part of the network; by receiving incoming hyperlinks actors are legitimated as part of the network. We assessed whether our three types of environmental organisations can be distinguished by their hyperlinking behaviour and their structural position within online networks. The LinLogLayout FDG in Figure 2b shows a clear division between an environmental-global cluster (blue) and an environmental-bio cluster (red), with the environmental-toxics subgroup (green) being situated in a fragmented and outlying position (four Pesticide Action Network sites act as brokers between this third group and the rest of the network). The manner in which the three subgroups define themselves by their choice of outlinks was also analysed numerically. Overall, sites in the environmental-bio and environmental-toxics groups show a

---

8 Ackland and Gray (2005) present a quantitative characterisation of the online information environment encountered by prospective migrants to Australia. Ackland and Gibson (2006) provide empirical analysis of the networking behaviour of over 100 parties from six countries. Ackland and Evans (2005) conduct a quantitative analysis of the visibility of abortion-related information on the WWW.

9 This section draws heavily from Ackland et al. (2006) and O'Neil and Ackland (2006).

10 See Benford and Snow (1986; 2000).

11 The numbers in brackets are the frequency counts for the word colocations.

greater tendency to link to the environmental-global group, while the environmental-global sites tend to link to one another.

Inbound hyperlinks. Counts of inbound hyperlinks are said to indicate visibility (Vreeland 2000, Hindman et al. 2003), trust (Davenport and Cronin 2000), or topic authority (Brin and Page 1998, Kleinberg 1999). Hyperlink affiliation networks operate as a function of the credibility among websites and the desire to augment this credibility: websites perceived as highly credible obtain more links than others (Park et al., 2004). We propose a principle: receiving hyperlinks is a form of endorsement, something to be sought out even when appearing to be disinterested.

We define organisations with high indegrees as central or successful and find that environmental-global sites received on average nearly 90 hyperlinks from other sites, compared with 70 links per site for the environmental-toxics group and only around 50 links per site for the environmental-bio group. The ratio of the number of inbound links to the number of outbound links is a measure of how hard a site is working to achieve its online visibility: a site that makes few links but receives many is more successful (with regards to online visibility) than a site that is pointing to many other sites, but receives few hyperlinks in return. The environmental-global and environmental-toxics groups are roughly equivalent, with each site on average making around two hyperlinks for each hyperlink it receives, while environmental-bio groups make on average three hyperlinks for each hyperlink they receive. The success of the environmental-global subgroup with regards to inbound hyperlink counts is further confirmed by the quintile distribution of indegrees; while environmental-global sites account for 56.2 percent of the sample, 71.2 percent of the top quintile of sites (ranked according to indegree) hail from this group. In contrast, the environmental-bio group is under-represented in the top quintile: only 12.5 percent of top-ranked sites are from this group, compared with their 29.6 percent representation in the entire sample.[12]

The contestation of nanotechnology. Nanotechnology – or the science of technology development at the atomic level – is increasingly being used in a range of industries including textiles and agriculture. While nano-proponents emphasise benefits such as increased food yields and novel consumer goods, nanotechnology has also generated concerns about implications for population health and safety. Online activist networks are structured as fields where incumbents and newcomers contest the primacy of issues (McAdam and Scott, 2005). Using the example of the diffusion of the contestation of nanotechnology, we show that online environmental organisations vary widely in their adoption of new issues, based upon their relative network prominence, length of presence in the field, and socio-cultural identity.

The nano-theme is mostly associated to the environmental-bio subgroup, a relative newcomer in the field which portrayed GM products as contaminants threatening nature and society, replacing, in the process, the environmental-toxics subgroup as primary challenger to the dominant environmental-global subgroup. Environmental-toxics groups are therefore understandably reluctant to endorse these successful competitors, or to deal with an issue which is closely associated to them: we found very few links from the environmental-toxics to the environmental-bio cluster, and we also observed that the nano-theme was least present in the toxics subgroup. This opposition also has ideological and sociological causes: first,

12 The environmental-global group has been present in the network longer than the other two groups; 43.5 percent of environmental-global sites have been operating for 7+ years (based on first date of archiving on the WayBack Machine, http://www.archive.org), compared with 30.4 percent of environmental-toxics and only 17 percent of environmental-bio sites. Thus, the higher visibility of the environmental-global group could be due to "preferential attachment" (e.g., Barabasi and Albert, 1999): the process whereby the "rich get richer", because newer entrants are inclined to link to already well-connected actors.

Environmental Justice activists are fearful that in the future genetics and genomics will be used to individualise the focus of environmental health analyses and interventions (Shostak 2004). Second, environmental Justice represents the attempts of urbanised poor people to resist being made the target of environmental discrimination. In contrast, the contestation of biotechnology allowed activists - generally held to issue from the ranks of the educated urban middle class (Crossley 2003) - to reconnect to their ancestral roots by forming an unusual alliance with farmers. The social and ideological divides separating these subgroups may make a connection impossible. However these divisions do not necessarily constitute an obstacle to the overall development of the environmental activist network: it is precisely the extreme diversity of the environmental movement which has made it hard to suppress (Gerlach 1999). New stakes on the network, such as the contestation of nanotechnology, aid in the generation and regeneration of divisions, thereby contributing to network robustness.

Web and grid portals in e-Social Science

Web and grid portals - background

The VOSON System is a *web portal* (a website that provides a gateway or interface to resources on the internet or an intranet) to enable social scientists to conduct empirical research into online networks. What distinguishes a web portal from a standard web site is that users maintain personalised accounts or identities that allow them to dynamically customise their configurational layout based on their needs. In the context of commercial web portals, the resources that are available for users are often information and links to other sites of interest. The resources that are available via the VOSON System are for use in academic research into online networks - CPU cycles, disk storage, data, and analytical methods. At present, these resources are located on the same machine as the web server, but as described below, we plan to distribute these resources across other machines and eventually, it is envisaged that VOSON will provide access to grid resources. It is therefore pertinent to provide some background information on grid portal development and articulate the development plans for the VOSON System in this context.

Distributed computing refers to interconnected computers sharing computing tasks. *Grid computing* can be distinguished from *cluster computing* in that the resources which comprise the grid are not all within the same administrative domain. That is, a grid may be comprised of any number of unrelated organisations who make their resources available via specific technologies that allow for making connections between computer systems. A *web service* is a software system identified by a URI[13] that can be interacted with by other software systems using XML-based messages. An example of a web service is the Google Web API which enables software systems to directly query the Google database. A grid service is a web service that further conforms to a set of conventions (interfaces and behaviour) that define how a computer system interacts with services available across a grid. If there were other organisations offering exactly the same search services as Google then one could imagine using a grid service (rather than web service) to make these sorts of queries - grid protocols would provide an automatic brokering mechanism (e.g. based on cost or activity levels) and the client software would pick up the necessary data from the appropriate organisation.

Grid Portals are “a class of WWW application servers that provide a secure online environment for gathering information about Grid services and resources as well as provide tools for utilising these Grid services and resources to perform useful tasks” (Russell, 2000). Grid portals can be broadly classed as either user portals which allow users to access grid

13 Uniform Resource Identifier - the addressing technology for identifying resources on the Internet (or intranets). A URL for a web page is an example of a URI.

services such as job submission and status tracking from any web browser (e.g. the PACI HotPage[14]), and application portals that cater to specific computational tasks (e.g. the Cactus Portal[15]). Grid portals are a means to provide a single access point to grid resources and help to lower the technology barriers that researchers encounter when trying to utilise the Grid.

Various toolkits have been developed to aid the development of Grid portals and a significant trend in e-Research tool development has been towards the development of portals consisting of portlets that comply with the JSR168 portlet definition standard. A portlet is a java servlet[16] that operates inside a portal and is able to interact with web and grid services. Portlets appear as small windows on a portal page (portals are typically comprised of several portlets). Adherence to the JSR168 portlet definition standard allows developers to share components and reuse software developed by other projects. The standard also facilitates interoperability with the various portal development frameworks/portal containers of which the main open source projects are GridSphere[17], and Jetspeed[18]. The portal framework itself manages and presents the contents of portlets (thus it is relatively easy to establish a "look and feel" that is appropriate for a particular portal), which helps to cut the development time needed when adding new features. The advantage of the portlet model is that users are able to dynamically change the user interface, e.g. which portlets are displayed and the location of the portlets within the portal. A portal development framework such as GridSphere provides a core set of portlets that provide the base functionality of the grid portal (e.g. user login, user and group management, portal configuration). Portal developers are then able to build upon these base portlets, adding higher-level interfaces and services without having to reimplement back-end functionality, such as resource discovery, job submission, and tracking. An example of a grid portal is the Cactus Portal (Kelley et al., 2005) which is being built upon GridSphere to provide customisable application portals for users of the Cactus Computational Toolkit.

Grid portals in e-Social Science

Grid portals are "a way of the future" for e-Research, and it would appear wise for empirical social scientists embarking on e-Research to follow communities in the physical and biological sciences and make use of the available portal development frameworks. However we contend that current grid portals are not well suited to the type of research that social scientists wish to conduct. The Cactus Portal, for example, supports research teams of scientists and engineers who develop simulation code and submit and monitor simulations to be run on high-performance computers. This type of research activity (submission of batch files to remote computers) is foreign to the majority of empirical social scientists.

It is not feasible (nor reasonable) for social scientists to completely adapt their research practices and culture in order to make use of currently available grid portal toolkits. At the same time, social scientists should not sit on the sidelines and wait for grid portal toolkits to be adapted to suit social science research practice; it is imperative for social scientists to actively participate in the development of e-Research technologies, otherwise these will continue to be developed solely to support "traditional" e-Research communities - namely the physical and biological sciences.

---

14 http://hotpage.paci.org

15 https://www.cactuscode.org

16 Servlets are a standard way of extending the functionality of a web browser and have largely replaced the use of CGI scripts.

17 http://www.gridsphere.org. GridSphere grew out of two early grid portal projects - the Astrophysics Simulation Collaboratory Portal and the Grid Portal Development Kit.

18 http://portals.apache.org/jetspeed

The question facing the VOSON Project is: do we take an existing grid portal development framework and build a grid portal that can be used to support social science research of online networks, or do we first build a web portal that supports our research activities with the view of later turning it into a grid portal using technology such as GridSphere? We have decided to take the latter approach, and we now briefly outline why we feel that current portal development frameworks are not yet suitable for building a social science grid portal.[19]

The majority of social scientists wishing to conduct e-Research will expect a user interface that supports a relatively high level of user interactivity, in particular, a menu-driven user interface. It is notable that just a few years ago, portlet development experts were explicitly warning that “Highly interactive user interfaces do not translate well to Web applications in general, or portlets in particular. If you want your interface to change automatically when a user takes some action, like selecting an entry in a drop-down list, you can either submit the form and reload the entire page (annoying), or use a scripting language to re-draw the portlet (very difficult)” (Fred and Lindesmith, 2003). The recent impact of Google Maps and Google Earth has highlighted the potential of the Asynchronous JavaScript and XML (AJAX) web development technique for creating interactive web applications. While the AJAX development approach has been available since the late 1990s, portlet developers are only now beginning to use AJAX to improve the level of interactivity of grid portals.[20] That grid portal developers are only now incorporating AJAX technologies to improve the responsiveness of user interfaces relates to the traditional audience of grid portal developers (e.g. computational scientists) primarily working with batch file submission. For e-Research to have an impact in the social sciences (and for social scientists to want to invest in using new e-Research tools) we contend that a high level of interactivity in the user interface is a priority and for this reason, the AJAX approach is already featured in the VOSON System.

The VOSON System – “development” version and plans

The VOSON Project faces challenges that are shared with other areas of e-Science: data set discovery and replication; access to novel analysis and visualisation techniques; authentication and access to datasets, methods and portals; resource hungry (in bandwidth and processing capacity) data collection and processing. To achieve our goal of making the VOSON System available for use by researchers outside of the VOSON Project (and used for network research in general, not just research into online networks) it will be important to automate and distribute VOSON functionality. The web services framework is a standard means of doing exactly this and in this section we detail our plans (and progress) for extending the VOSON System in this manner.

Planned services that will be accessible within the VOSON System

Data set discovery and replication. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[21] can be used at several levels — both as a means of capturing interesting data from the internet, and also as a standardized way of publishing datasets for later discovery by VOSON users. A centralised registration service can be provided which allows for services to be registered as VOSON-compatible.

---

19 Note that our decision to use PHP rather than java will not necessarily hinder our planned eventual use of a grid portal environment as it is possible for PHP applications to be run within JSR168 compliant portal frameworks (see http://www.itgroundwork.com/resources/php-setup.html).

20 A recent presentation from a GridSphere developer promised more responsive user interfaces using AJAX technologies in future releases (Novotny, 2005).

21 http://www.openarchives.org/

Access to methods. Web services framework provides a key opportunity for linking analysis and visualisation techniques created and maintained by other people (and written in various programming languages) into the VOSON capability. The existing VOSON capabilities will be recast into web services, exploiting wherever possible existing standards and toolkits such as gSOAP[22] and SOAP::Lite[23] - see below for details of preliminary work in this area.

Data collection. Data acquisition can also be offered on a web services model. There is no particular need for web mining to be hardwired to the VOSON portal; indeed, it makes sense for data mining to be offloaded to a separate server to prevent data traffic from interfering with the portal's interactive functionality. By adopting standardised interfaces, the services may well be available for other (non-VOSON) applications, and there may well emerge an economy of different webmining applications that VOSON can draw upon. Although webmining is not typically a CPU intensive task, and so perhaps not suitable for computational grid approaches, if the data has automatic classification and cleanup algorithms applied to it then computational resources do become a consideration.[24] With the Open Grid Services Architecture[25] computational grids will interoperate with web services, so it will be feasible to use grids to handle this computation aspects of web mining.

Statistical analysis. In the near future, social network analysis routines available in the "sna" package[26] of the open-source R statistical software will be accessible from within the VOSON System. We plan to achieve this by establishing an R service, possibly building on existing frameworks such as that offered by RServe.[27]

Authentication

Authentication is presently controlled manually by the VOSON administrator. Shibboleth technology (http://shibboleth.internet2.edu/) shows promise in solving the authentication and access problem - it is clear that it, or something like it, will control access to data and services in a web service world. Being distributed and trusted, it obviates the need for a VOSON administrator to manage individual user passwords, with all the associated security risks. Instead policy rules will be used to grant access rights to particular groups of users, and Shibboleth will match a particular user request to these rules to allow or deny the request.

An initial exploration of web services in VOSON

As an initial foray into the world of web services, the VOSON LGL network visualisation tool has been converted to a web service. LGL is normally available as an independent executable - for the web service, we converted the code into a library, to which the web service server is linked. Under the current "production" version of the VOSON System, the LGL service exists as a program callable from the CGI script implementing the VOSON portal user interface. This program has been split into a client/server pair, (see Figure 3) with the LGL algorithm and graph rendering implemented as a remote procedure call over SOAP, using the gSOAP framework. gSOAP can generate the WSDL definition file directly from a C++ specification of the remote procedure call, which is how we've implemented the web

22 http://www.cs.fsu.edu/engelen/soap.html

23 http://soaplite.com/

24 In this context, we are investigating the use of machine learning techniques (specifically, Support Vector Machines - see, for e.g., Vapnick 1995) for automatic categorisation of pages based on content – this will help with controlling the problem of topic drift, thus allowing focused crawling.

25 http://www.globus.org/ogsa/

26 http://cran.r-project.org/src/contrib/Descriptions/sna.html

27 http://stats.math.uni-augsburg.de/Rserve/

service here. Alternatively, it can parse an existing WSDL file to generate a C++ specification. In either case, gSOAP generates C++ stub code that allows the web service call to be performed by means of a simple member call on a proxy object, and the server implementation to be provided as a standard C++ function.

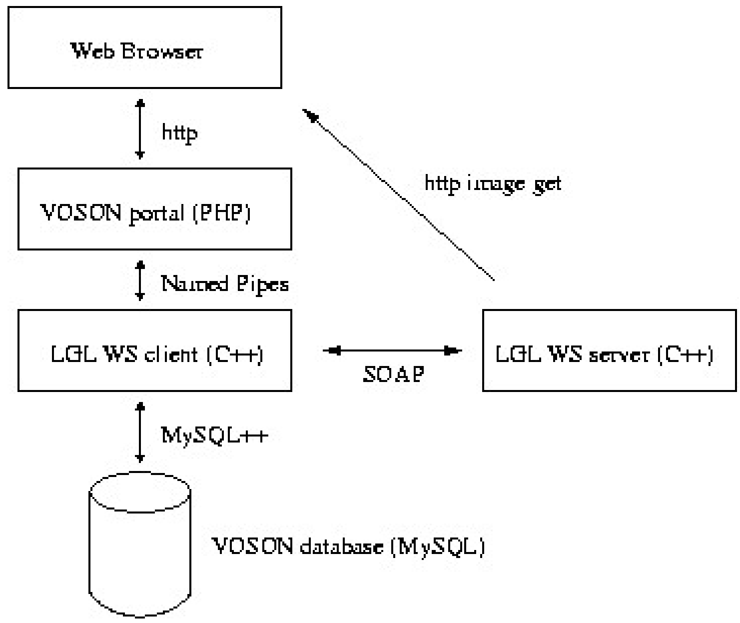

Figure 3: LGL mapping implemented as a web service

For displaying the graph on the browser, the server renders the graph to a PNG image file using Cairo library, an open source drawing library used by a number of well-known open source applications including Mozilla. To return the rendered image to the browser, the web service server returns a URL to the web service client which can be later used by the browser to fetch the image.

Presently, graphs are described by means of gSOAPs rendering of a C++ vector of pairs of integers describing the graph's link list. A typical graph description might be:

```
<edgelist>
<list><u>0</u><v>1</v></list>
<list><u>0</u><v>2</v></list>
<list><u>1</u><v>3</v></list>
<list><u>1</u><v>4</v></list>
<list><u>3</u><v>1</v></list>
<list><u>2</u><v>0</v></list>
</edgelist>
```

For a moderate sized graph (up to 65536 nodes), each link can optimally be represented by 4 bytes. By using XML encoding, this expands to 28 bytes, a factor of 7 times the optimal representation. This suffers from the dual problems of being non-standard and verbose and so we are looking at various ways of providing an optional binary channel to efficiently supply the graph specification. The problem with standards is that there are so many to choose from. However, from the XML world, we are most likely to support the XGMML standard[28], which has the same verbosity as the gSOAP generated language, and has a typical link:

```
<edge source="1" target="2"/>
```

28 http://www.cs.rpi.edu/~puninj/XGMML/

Conclusion

In closing, we feel that at this early stage of e-Social Science there are two key requisites for the successful integration of e-Research infrastructures and tools into the working practices of social scientists. First, the research, by its nature, must be unachievable - or perhaps even unimaginable - without these tools, because of the scale/complexity of the data involved or the need for remote collaboration (over and above what can be offered by email and the telephone). Otherwise, social scientists will not bother to re-skill. Second, the new tools need to (as far as possible) closely resemble tools that empirical social scientists are already familiar with. A simple application of current grid portals that primarily support submission of batch files to remote computing resources is therefore ruled out, as this approach would not reflect the working practice of the vast majority of social scientists and cannot be expected to have an impact. However, when appropriate, social science e-Research projects can (and obviously should) build upon the tools that have been developed for supporting research in the physical and biological sciences.

References

Ackland, R. and Evans, A. (2005): 'The visibility of abortion-related information on the World Wide Web', Presentation to the Annual Conference of The Australian Sociological Association, 5-8 December, Hobart.

Ackland, R. and Gibson, R. (2006): 'Hyperlinks and horizontal political communication on the WWW: The untold story of parties online', mimeo., The Australian National University. Under review.

Ackland, R. and Gray, E. (2005): 'What can potential migrants find out about Australia from the WWW?' *People and Place*, vol. 13, no. 4, pp. 12-22.

Ackland, R., O'Neil, M, Bimber, B., Gibson, R.K., and Ward, S. (2006): 'New methods for studying online environmental activist networks', Paper presented to 26th International Sunbelt Social Network Conference, 24-30 April, Vancouver.

Adai, A. , Date. S., Wieland, S., and E. Marcotte. (2004): 'LGL: Creating a map of protein function with an algorithm for visualizing very large biological networks', *Journal of Molecular Biology*, no. 340, pp. 179-190.

Barabasi, A.-L. and Albert, R. (1999): 'Emergence of scaling in random networks', *Science* no. 286, pp. 509–512.

Benford, R.D., Snow, D.A., Rochford, E.B., and Worden, S.K. (1986): 'Frame alignment processes, micromobilization, and movement participation', *American Sociological Review* no. 51, pp. 464-481.

Benford, R.D., and Snow, D.A. (2000): 'Framing processes and social movements: An overview and assessment', *Annual Review of Sociology* no. 26, pp. 611-39.

Brin, S. and Page, L. (1998): 'The anatomy of a large scale hypertextual Web search engine', available at: http://www-db.stanford.edu/~backrub/google.html (accessed 27/11/2005).

Crossley, N. (2003): 'From reproduction to transformation: social movement fields and the radical habitus', *Theory and Society* no. 20, vol. 6, pp. 43-68.

Davenport, E. and Cronin, B. (2000): 'The Citation Network as a Prototype for Representing Trust in Virtual Environments', in B. Cronin and H.B. Atkins (eds.): *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, Information Today, Metford (NJ), pp. 517-534.

Fred, A. and S. Lindesmith (2003): 'The case for portlets: How to decide if portlets are your best option', IBM Staff Paper, available at: http://www-128.ibm.com/developerworks/ibm/library/i-portletintro/ (accessed 27/11/2005).

Gerlach, L. (1999): 'The Structure of Social Movements: Environmental Activism and its Opponents', in J. Freeman and V. Johnson (eds) *Waves of Protest. Social Movements Since the Sixties*, Rowman & Littlefield, Lanham, MD, pp. 85-98.

Hindman, M., Tsioutsiouliklis, K., and Johnson, J. (2003): '"Googlearchy": How a few heavily-linked sites dominate politics on the Web', Mimeo., Princeton University, 2003.

Kelley, I., Russell, M., Novotny, J and O. Wehrens (2005): 'The Cactus Portal', available at www.vpac.org/ocs/viewpaper.php?id=20&cf=3 (accessed 27/11/2005).

Kleinberg, J. (1999): 'Authoritative sources in a hyperlinked world', *Journal of the ACM* vol. 46, no. 5, pp. 604-642.

McAdam D., and W.R. Scott (2005): 'Organisations and Movements', in G.F. Davis, D McAdam., W.R. Scott, M.N. Zald (eds) *Social Movements and Organisation Theory*, Cambridge University Press, Cambridge and New York, pp. 4-40.

Noack, A. (2005): 'Energy-Based Clustering of Graphs with Nonuniform Degrees,' Accepted for publication in: Proceedings of the 13th International Symposium on Graph Drawing (GD 2005, Limerick, Ireland, Sep. 12-14).

Noack, A. (2004): 'An Energy Model for Visual Graph Clustering,' Proceedings of the 11th International Symposium on Graph Drawing (GD 2003, Perugia, Italy, Sep. 21-24), LNCS 2912. pp. 425-436.

Novotny, J. (2005): 'Grid Portals: Bridging the gap between users and services', Keynote presentation, Australian Partnership for Advanced Computing Conference on Advanced Computing, Grid Applications and e-Research, available at: www.gridsphere.org/gridsphere/html/apac05/07_APAC_05_keynote.pdf (accessed 27/11/2005).

O'Neil, M., and R. Ackland (2006): 'The structural role of nanotechnology-opposition in online environmental networks', Paper presented to 26th International Sunbelt Social Network Conference, 24-30 April, Vancouver.

Park, H.W., Kim, C., and Barnett, G.A. (2004): 'Socio-communicational structure among political actors on the Web in South Korea. The dynamics of digital presence in cyberspace', *New Media & Society*, vol. 6, no. 3, pp. 403-423.

Russell, M. (2000): 'The Cactus Portal: A case study in grid portal development', available at: http://www.cactuscode.org/old/Presentations/GridPortals_2000.ppt (accessed: 25/11/2005).

Shostak, S. (2004): 'Environmental justice and genomics: Acting on the futures of environmental health', *Science as Culture*, vol. 13, no. 4, pp. 539-562.

Thelwall, M. (2004): *Link Analysis: An Information Science Approach*, Academic Press, San Diego.

Vreeland, R.C. (2000): 'Law libraries in cyberspace: A citation analysis of World Wide Web Sites', *Law Library Journal*, vol, 92, no. 1, pp. 49-56.

Vapnick, V. (1995): *The Nature of Statistical Learning,* Springer, New York.