

Using open standards to share spatial data using Mapserver, PostGIS and other open source GIS software.

Antti J. Roppola

Bureau of Rural Sciences, GPO Box 858, ACT Australia
Tel: + 61 2 62725980, Fax: +61 2 62724687 Email: antti.roppola@brs.gov.au

Abstract

Many Australian and International agencies already use open source applications to deliver spatial data for free over the Internet. A steady increase in the variety and capability of open source GIS technologies, and widespread adoption of open standards is greatly enhancing this capability. This paper will give an overview of selected open source GIS & mapping technologies such as Mapserver (<http://mapserver.gis.umn.edu>), PostGIS (<http://postgis.refrations.net>), GDAL (<http://www.remotesensing.org/gdal>) & MapLab (<http://www.dmsolutions.ca>) for online mapping as well as open standards such as those promoted by the OpenGIS consortium (<http://www.opengis.org>) and the Australian Spatial Dataset Directory (<http://www.auslig.gov.au/asdd/tech/>) distributed database. We will then examine how these tools have been applied to enable community access to spatial Natural Resource Management (NRM) data, using applications developed by the Bureau of Rural Sciences (<http://data.brs.gov.au/mapserv/plant>) and others as examples.

Background

This talk will cover:

- our experiences in using open source online software for online data access
- some detail on the technologies used and their capabilities

What is BRS and what does it do?

The BRS is a research Bureau for the Department of Agriculture, Fisheries and Forestry. As the mission statement "Science for Decision Makers" implies, key roles of the Bureau are to research issues affecting natural resources and just as importantly, communicate the implications of that research to the people who decide how those resources are managed.

Maps and mapping are important tools for the organisation. We are heavy users of Geographic Information Systems (GIS) and maps can convey huge amounts of information in a single glance. Over the last ten years, we have seen GIS become more affordable, but it still requires considerable resources and knowledge to use.

The Internet is an attractive alternative because:

- More people know how to use web browsers than GIS systems
- The cost of web publishing is relatively cheap, allowing greater access
- We can update data more frequently and manage versioning
- We can do things that are not possible with paper documents
- We are keeping digital data in digital form, maintaining its depth
- No need for users to acquire complex, expensive GIS software

Like most, we are a small work area (five people who also have other duties) with very limited resources. While we want to explore new and better ways of doing things, delivering more for less and minimisation of risks are the key principles that guide our thinking.

This interest in making our data available to everyone isn't new, or peculiar to the BRS, not even use the use of open source software. However, the number of quality OSS tools for GIS has grown enormously in number and scope (<<http://www.freegis.org>>). The modular nature of many of these developments makes it easy to combine them into new and exciting applications.

What is Mapserver?

Mapserver is the key online mapping package that we are using to give people wider access to mapping data. Mapserver (<<http://mapserver.gis.umn.edu>>), is an open source online mapping tool developed by the University of Minnesota and others. To quote the project home page:

"MapServer is an OpenSource development environment for building spatially enabled Internet applications. The software builds upon other popular OpenSource or freeware systems like Shapelib, FreeType, Proj.4, libTIFF, Perl and others. MapServer will run where most commercial systems won't or can't, on Linux/Apache platforms. MapServer is known to compile on most UNIXes and will run under Windows NT/98/95."

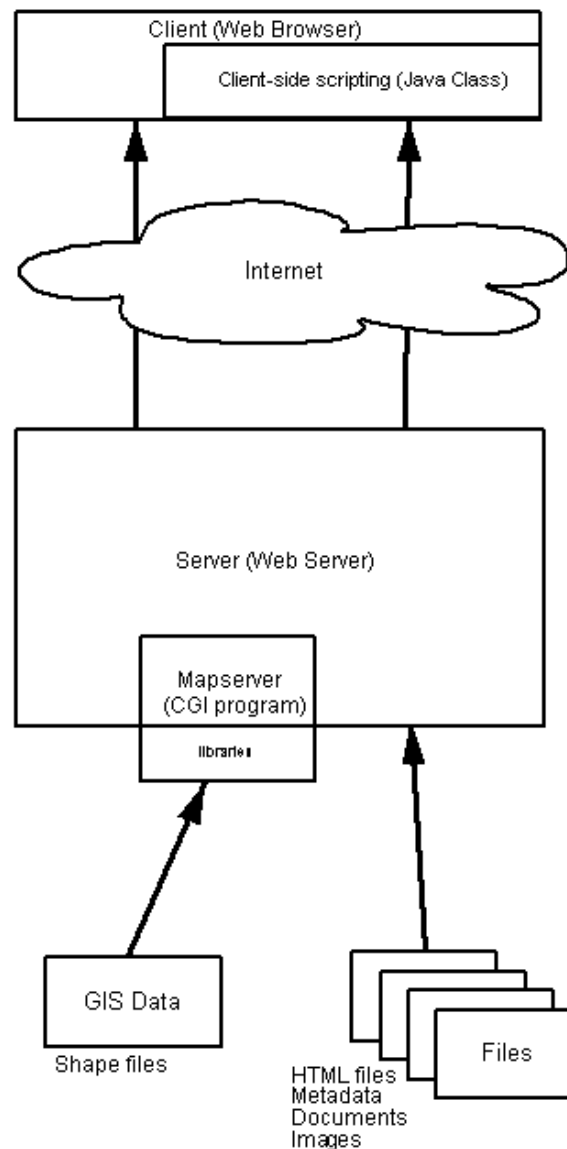


Figure 1. Mapserver deployed as a Common Gateway Interface (CGI) application.

Figure 1. shows how Mapserver is deployed as a Common Gateway Interface (CGI) application. Core functions are similar for other deployments, but the CGI is a simple example. Typically, a web client invokes the Mapserver CGI via a HTML form and nominates several arguments:

- A map file to interrogate
- Layers in the map that are required
- A location (either an extent or a location)
- A method (such as “draw” or “query”)

```
...
LAYER
  NAME 'Land Exclusion Mask'
  TYPE polygon
  GROUP "default"
  STATUS OFF
  DATA data/mask
  CLASSITEM 'mask'
  CLASS
    NAME 'Unavailable or Unsuitable'
    EXPRESSION 'yes'
    COLOR 204 204 204
  END
END
...
```

Figure 2. An example from a map file showing properties for a layer of data

In response, the Mapserver reads the map file, which explains where the GIS data is located on the filesystem and how it should be displayed (see figure 2). Mapserver then reads GIS data from files, which includes linework and corresponding attributes. For “draw” requests, data is rendered to a view and the URL returned to the client, while for “query” requests, retrieved attribute information is returned to the user. Because GIS data is typically very large, methods such as tiling, indices and pyramiding are used to greatly reduce the amount of data that must be read from disk. The views rendered from spatial data are much smaller than the data from which they are derived and therefore easier to send to via HTTP.

Initial steps

The Bureau’s initial involvement with Mapserver was with a Commonwealth Spatial Data Committee (CSDC) web mapping trial in 2001. The goal of the trial was to test the implementation of an Open GIS Consortium (OGC) Web Mapping Services (WMS) deployment across several Commonwealth agencies that would allow spatial data to be synthesised from many sources in “near real time”.

Mapserver was suggested as a tool since it was one of the few applications implementing the WMS specification (at the time through a Perl wrapper provided by Social Change Online). Of the available options, Mapserver was also the cheapest, and lowest risk option. The default build was deployed as a CGI on an existing server, and most time and effort went into preparing data to participate in the trial. No real knowledge of Mapserver was required, as we modified the example map files and used precompiled binaries.

Our experience in the trial was an essential first step. It demonstrated that the technology and standards were maturing, and that online delivery of spatial data was

potentially a reliable and cost effective way of giving widespread access to GIS data. It also demonstrated that Mapserver was a mature and reliable product.

First iteration

The online Bioenergy Atlas of Australia (BRS, 2002) (http://www.brs.gov.au/bioenergy_atlas/) was developed collaboratively by the University of Adelaide and the Bureau of Rural Sciences. It is also the first online application developed by the Bureau to incorporate mapping and a simple Decision Support System (DSS) capability. The anticipated users include anyone with an interest in bioenergy, such as farmers with waste products suitable for power generation, staff of power utilities and government advisors on energy policy. These users need to see what quantities of bioenergy resources existed in an area and how their distribution is related to necessary infrastructure such as transport, power generation facilities and transmission apparatus. The key driver for the creation of an online atlas was a need to make spatial information on potential bioenergy resources available to a user base that was widely spread; in terms of location, expertise, resources and interests (Bugg et al., 2002).

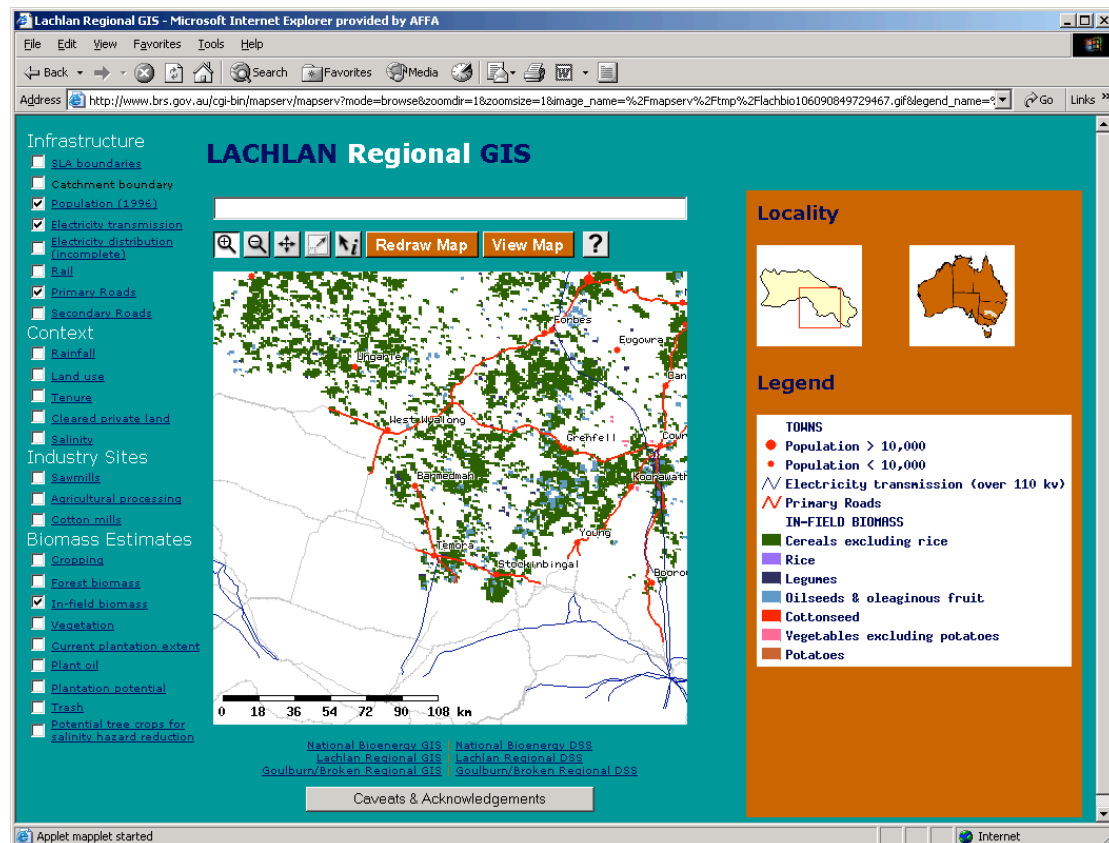


Figure 1. A screenshot of the Regional GIS in the Bioenergy Atlas, using the Minnesota Mapserver and static HTML pages.

Because of the wide range of interests, and the evolving nature of the data, an online delivery was also attractive in that an interactive application could be used to generate customised products. Manually generating maps and tabular summaries of all the possible combinations of areas and resources that might be possible would have been a costly exercise and many of these products might never be requested. Dynamically generating the products also meant that source data could be periodically updated without having to re-create a large number of derived products.

Mapserver was chosen as the basis for the atlas, mostly because it had been robust and trouble free in the CSDC trial. We also evaluated a number of other mapping products, but the capabilities and performance were very similar, and using most of them would have been too expensive. While one licence was reasonable, a requirement to purchase of multiple copies of the software (production, application development, system development) as well as several supporting applications (OS, scripting, etc.) rapidly inflated overall cost. Picking an open offering has also meant that we can easily transfer skills and technology.

Second iteration

Unlike the Bioenergy Atlas, which was primarily concerned with presenting a user-friendly front end to mapped data, the National Plantation Information Network (PIN) (BRS, 2003) (<<http://data.brs.gov.au/mapserv/plant>>) is a more generic application that ties together information from a variety of spatial and non-spatial sources, cross linking them where appropriate.

The need for the PIN is similar to the need that drove the development of the Bioenergy Atlas; over the years a variety of publications, maps, datasets and other sources of information in varying formats on the potential for plantation forestry had been completed by the Bureau and others. But over the years it has become difficult to locate complete sets of spare hardcopy reports for distribution to interested parties, and staff turnover had shown that corporate memory of completed work was not absolutely reliable. Rather than commission reprints of products that may or may not be requested, it is simpler to distribute them electronically and to provide an interface to allow users to locate information for themselves. The application also provided an opportunity to see how individual regional projects fitted into a bigger national picture.

By this stage, we were more confident with Mapserver and its capabilities. Mapserver was used as a PHP function library (Figure 5.) so that we could integrate map data with other data, and because we wanted to structure applications so that they were easier to maintain. For example, the map interface is entirely generated from the map file or by instantiating maps as objects (Figure 6.). Since this project, a number of third party interfaces for Mapserver have appeared, such as MapLab (<<http://www.dmsolutions.ca>>), if they had been available at the time, we would have modified an existing interface rather than maintaining our own. Building Mapserver as a PHP library also allowed us to integrate other functionality (such as the generation of PDF maps).

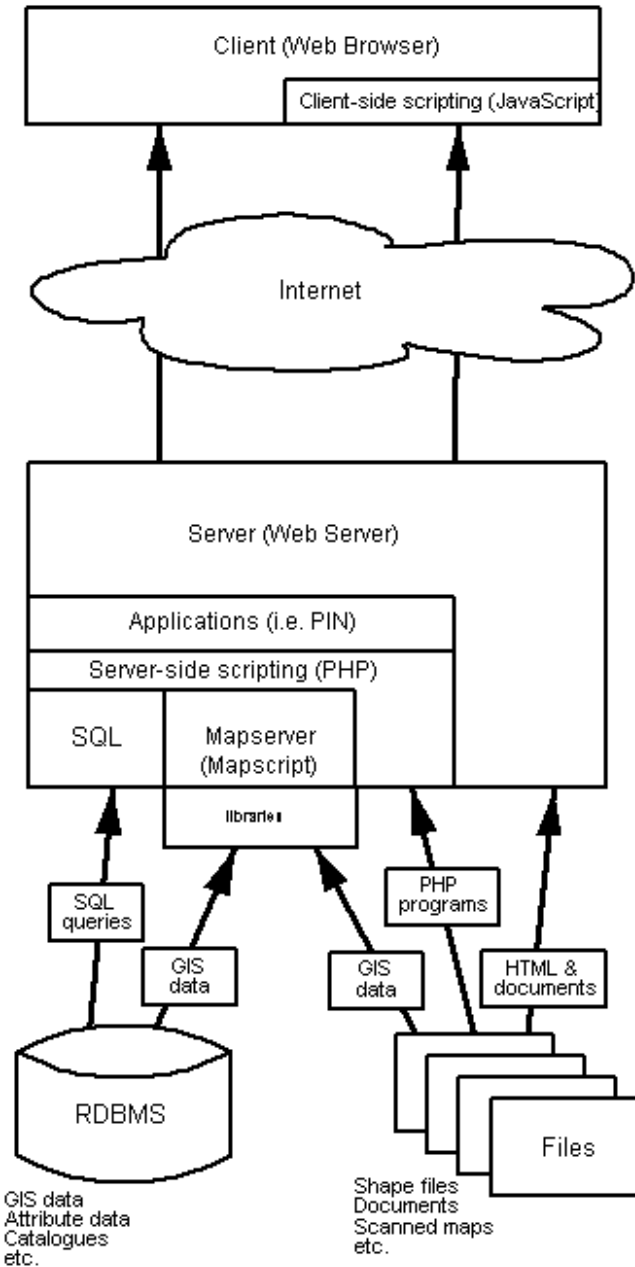


Figure 5. Mapserver's Mapscript deployed as a function library in PHP.

```

$layer->open ($map->shapepath) ;

$oShape=$layer->getShape ($oRes->{tileindex},
    $oRes->{shapeindex}) ;

$xmin=$oShape->bounds->minx;
$ymin=$oShape->bounds->miny;
$xmax=$oShape->bounds->maxx;
$ymax=$oShape->bounds->maxy;

$extent=$oShape->bounds;

```

Figure 6. An example use of Mapscript to query properties of a map object

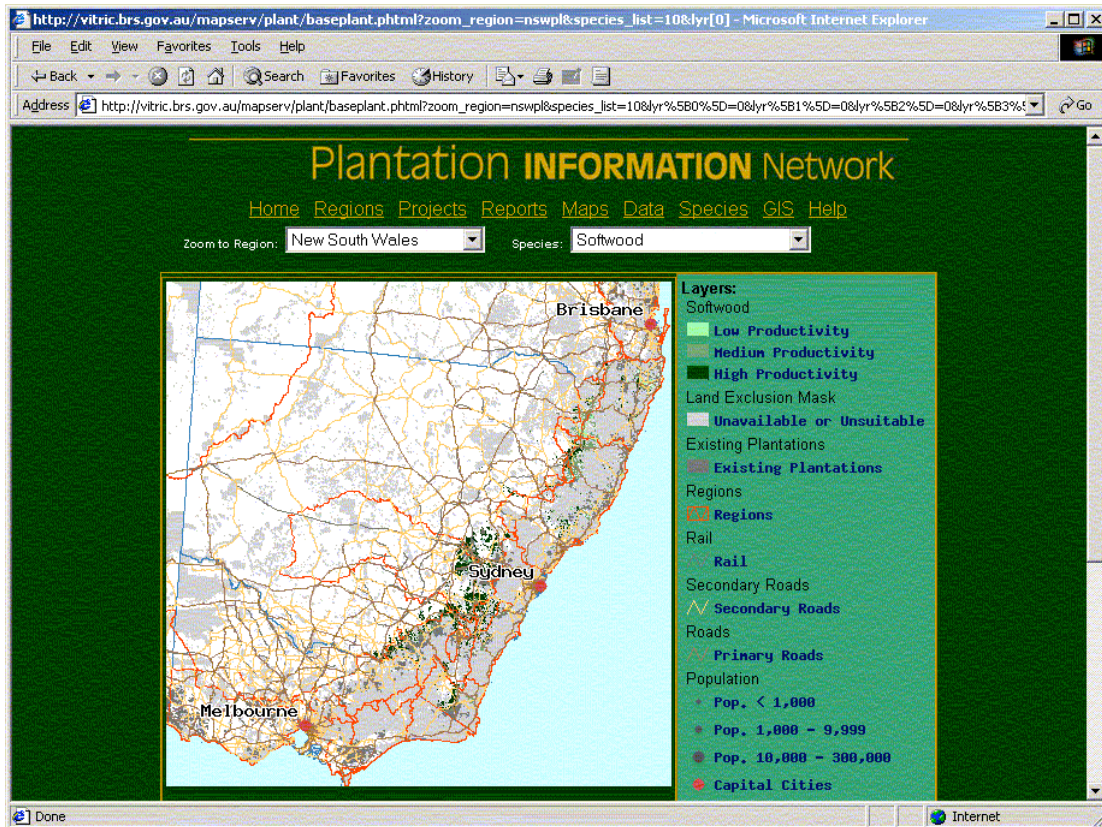


Figure 7. A screenshot of the Plantation Information Network illustrating the simple interface.

Project Outcomes

Mapserver has proven to be an effective online mapping tool, and the very low footprint (rendered pages are smaller than many banner ads) makes Mapserver especially suitable for delivery into rural and regional areas where Internet connection speeds are usually well below that available in urban areas. While significant investments in time and expertise are required to get a system operational, the system has been reliable and trouble free once deployed. New tools promise to reduce these setup costs. The very dynamic user community has meant that support has always been prompt, and improvements and enhancements arrive faster than we are able to consider them.

Client and interest groups have received applications built on Mapserver favorably. Formal usability testing has also demonstrated that online mapping is an effective way for “ordinary” people to locate spatial information.

Where to next

Open Standards Compliance

The Open GIS Consortium (OGC) (<<http://www.opengis.org>>) is a consortium of industry, government, individuals and academia that see inter-operability specifications as a great way to promote wider use of spatial information. The OGC has defined several interoperability standards, and we are looking at expanding our use of the Web Map Services (WMS) and Web Feature Services (WFS) specifications. These specifications allow data to be published as a service that can be used by any compatible client (Figure 8.).

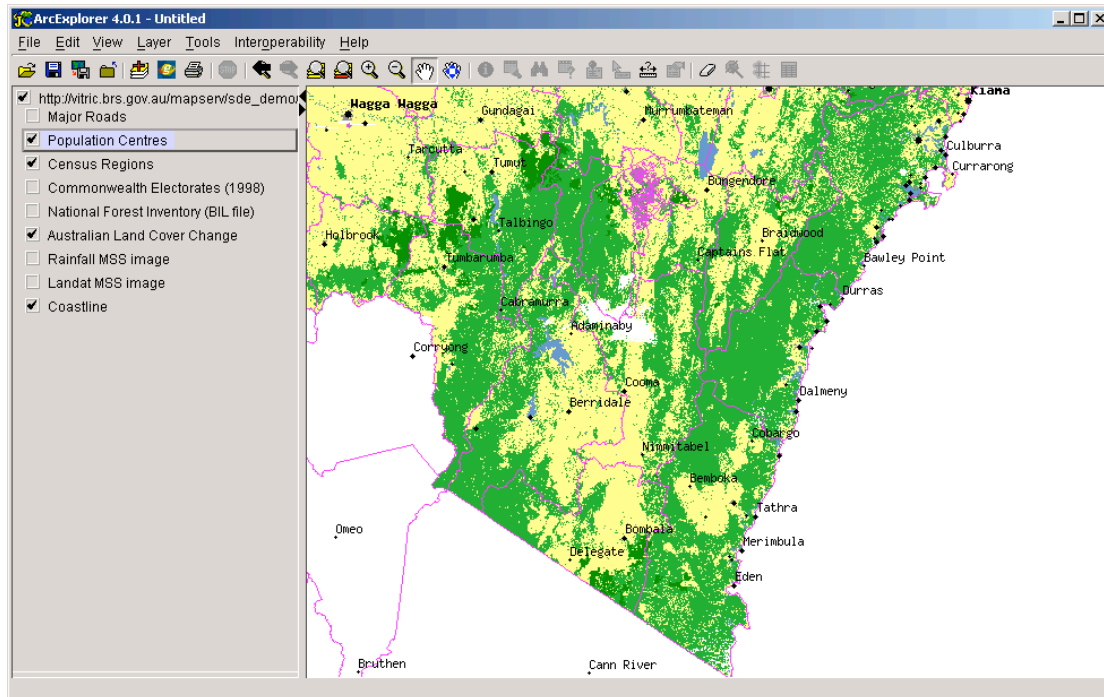


Figure 8. Bureau of Rural Sciences spatial data being integrated into ESRI's ArcExplorer using Open GIS Consortium Web Mapping Services implemented in Mapserver.

While we have proof of concept and prototype implementations, we have decided to postpone publishing large volumes of data until we implement mechanisms for management and discovery. As an interim measure for external discovery, the Commonwealth Office of Spatial Data Management (OSDM) (<http://www.osdm.gov.au>) has set up an interim register of OGC compliant nodes.

Mapserver also continues to grow as a product, and changes to the output abstraction are allowing Mapserver to serve data in an even greater variety of open and proprietary formats, including Geographic Markup Language (GML) and Scalable Vector Graphics (SVG).

Metadata

Greatly expanded libraries of spatial data are not much good unless the right data can be found. The Australian Spatial Data Directory (ASDD) (<http://www.ga.gov.au/asdd/tech/>) provides an online directory that allows people to search for GIS data held by participating agencies. The ASDD uses the Library of Congress Z39.50 (ISO 23950) search and retrieval protocol (<http://lcweb.loc.gov/z23950/agency/>) to coordinate searching of metadata records held on agency nodes. Many nodes (including the BRS) use the open source Isite Z39.50 server (<http://www.awcubed.com/Isite/>), which is also recommended by the US Federal Geospatial Data Committee (<http://www.fgdc.gov>). Currently, the core of the ASDD is an XML DTD metadata specification that provides metadata, but not the data itself.

Elsewhere in Australia, initiatives such as Community Access to Natural Resources Information (CANRI) (<http://canri.nsw.gov.au>) are already demonstrating how standards can be used to create open systems where open and proprietary solutions can be used to give end-to-end access to data. CANRI works by providing online resource extensions to the ANZLIC metadata specification so that data from many agencies can be discovered and viewed using a Z39.50 and WMS based broker.

The Bureau has (with contributions from others) developed a new metadata specification that is not only compliant with the ANZLIC standard, and the international standard, ISO 19115, but also allows metadata documents to provide links to the data that they are describing (Figure 9.). The specification is available for free as an XML schema and supporting XSLT transformations. The prototype (also freely available) is built from open source software, including Apache Cocoon, Tomcat and ISite2. The Bureau and the ASDD are also looking at other open source software such as the Zap Z39.50 HTTP gateway (<<http://www.indexdata.dk>>) to further improve access to spatial data.



Figure 9. XML metadata record including links to online and downloadable data rendered as HTML

PostGIS

As well as supporting proprietary spatial RDBMS solutions such as ESRI's ArcSDE and Oracle Spatial, Mapserver also supports the PostGIS spatial extensions to Postgres (<<http://postgis.refrains.net>>). Putting spatial data into an RDBMS is attractive for several reasons:

- Scalability – let the RDBMS handle threading, clustering
- Performance – let the RDBMS optimise retrieval of geometric data
- Maintenance – Let the RDBMS handle backups, updates and versions
- Interaction – Users can be allowed to add data in a controlled way
- Linkages – For example to live data logs
- Topology – Enables the creation of new layers by logically combining others

We have been testing PostGIS and created some example applications, but PostGIS is still in development. So far, we have chosen not to deploy to PostGIS because the

data model is still settling down, and installation involves patching your Postgres source.

Conclusions

Mapserver has proved to be an effective way of giving wider access to our spatial data assets, and an important demonstration of the capabilities of open source software (OSS) products in general. Small, low risk deployments have served as important precedents, paving the way for more ambitious projects and wider organisational support.

The modular nature of OSS projects has allowed us to integrate “black boxes” in confidence, without having to acquire detailed knowledge of their inner workings. Conversely, it is also important to build enough in-house skill to maintain strategic capabilities (this is equally applicable to proprietary systems). To a large degree, the success or otherwise of a project is determined by how well the team members understand the opportunity and the strengths and weaknesses of the tools available.

More generally, the range of open source GIS tools and solutions has increased remarkably in a very short time, and the quality of many open source GIS projects is excellent. This has been recognised by many Australian (and other) government agencies, which have been keen to utilise and promote the use of open source software.

Finally on a personal level, it has been very satisfying to see GIS evolve from an expensive and elite technology, to a resource that can be easily integrated into readily available web applications. Far from diminishing GIS, such activities open up even more opportunities for the application of spatial data.

Acknowledgements

The mapping applications were developed in consultation with a range of stakeholders and clients. The Bioenergy Atlas was a collaborative project with partial funding provided by the Joint Venture Agroforestry Program (JVAP) and a Renewable Energy Industry development grant. The Plantation Information Network was funded by the Fisheries and Forestry Industry Division, Department of Agriculture Fisheries and Forestry under the Vision 2020 initiative. Rod Keenan, Ian Nuberg, Lyndon Zimmermann, Stuart Row and Nick Stephens are particularly acknowledged.

Development of the tools described is a community effort, involving many people and organisations, within Australian and Internationally. David Crossley from Index Geo Pty. Ltd. is particularly acknowledged for the development of and assistance with Z39.50 and XML systems. Evert Bleys and Dale Young from the Bureau are acknowledged for the development of XML schema and transformations. Stuart Row, also from the Bureau, is acknowledged for the development and deployment of applications.

References

BRS, 2002, *The Bioenergy Atlas of Australia*, Bureau of Rural Sciences, Canberra, viewed 13 June 2003 <http://www.brs.gov.au/bioenergy_atlas/>.

BRS, 2003, *The National Plantation Information Network*, Bureau of Rural Sciences, Canberra, viewed 13 June 2003, <<http://data.brs.gov.au/mapserv/plant>>.

Bugg, A.L., Nuberg, I., Keenan, R., Zimmermann, L. 2002. *Bioenergy Atlas of Australia, An online Geographic Information System and Decision Support System*. The JVAP Research Update Series No. 5. Rural Industries Research and Development Corporation, Canberra.