GIM Interviews Paul Ramsey, Director, Open Source Geospatial Foundation

Open Source Success

As open-source software, PostGIS has gained much interest and credit from researchers and companies big and small, all over the world. In the geospatial domain, open-source libraries are finding their way into more and more proprietary products. What is the secret of OS success? Is it idealism or pragmatism that drives users into the arms of PostGIS? How will the software develop further? We asked Paul Ramsey, open-source guru in the geospatial domain.

By Mathias Lemmens, editor-in-chief, GIM International

Practical considerations led Paul Ramsey to begin working with open-source geospatial software; in 2001 a client needed a spatial database for a project and the proprietary alternatives at the time were both expensive and bulky. The solution Ramsey and his company developed was PostGIS; it proved useful too in other projects for other clients, and was soon adopted as the standard spatial database for the open-source community. Today PostGIS is used globally by organisations differing in scale from graduate research project and community mapping to DigitalGlobe and NASA. Paul Ramsey is a director of the Open Source Geospatial Foundation (OSGeo), committed to several open-source projects and a popular speaker and teacher on open-source topics at geospatial conferences and workshops.

What is PostGIS? What does it do and who is using it?

PostGIS is an open-source spatial extender for the PostgreSQL relational database, which is also open source. PostGIS is to PostgreSQL what Oracle Spatial is to Oracle, and is being used by organisations requiring a consistent, intelligent and reliable database for spatial objects, and in general by people with very large seamless datasets using spatial SQL for query, or database features for concurrent access. For example, Institut Geographique National (IGN) uses PostGIS to store the BDUni topographic base-map for France (Figure 1). PostGIS is also being increasingly used for geospatial analysis of the sort that used to be done exclusively with Arc/INFO and other workstation tools. Spatial SQL, like that supported by PostGIS and other spatial databases, is an extremely flexible analytical tool, allowing overlay, spatial joining and spatial summaries large and small.

What are the differences between open-source and proprietary software?

The legal difference, well described at www.opensource.org, is that when you receive a piece of open-source code you have the right to alter and redistribute it if you wish to do so. Iced tea is open-source (everyone knows how to make iced tea, and some people even have variations on the recipe) but Coca-Cola is proprietary (the recipe is a trade secret). The cultural difference, which the legal rules help promote, is that open-source tends to have vigorous user communities because users are empowered to improve the software in all kinds of ways that they themselves, and not a vendor, decide. All users may not make use of their freedoms under open-source, but all have the option. In the same way, you might not make use of your right to free speech, but it's nice to know you have it.

Could you explain the success of open-source software and, more specifically, why established firms are becoming increasingly interested in using it?

The success of open-source has been built on a combination of enthusiasm and economics. The enthusiasm comes from the technology geeks who are asked to build systems; they like to use software components that are transparent, so that they can talk easily to its authors, get a defect fixed very quickly, rather than working around it. And open-source software provides all these things. The economics comes from the zero capital cost aspect of open-source. Of course, open-source isn't zero cost, because it takes staff time to integrate and manage it, but it *is* zero capital cost. It also incurs zero for things like managing licensing and royalties. Anyone who has built a system on top of a proprietary library knows the pain of getting to the rollout stage and trying to figure out how just how many licences they are going to need. Will we need four CPUs, or eight? When a

CPU licence costs $20,000, and the penalty for contravening licence terms is a lawsuit, these are not trivial questions. For new organisations in particular, or for those building a net-new system, the low start-up cost and lack of licensing restrictions are particularly tempting. Most of the truly heavy users of PostGIS are organisations that either did not exist ten years ago or did not have a substantial database infrastructure.

Is the open-source 'movement' driven by idealism, or are there also business opportunities?

There is a great deal of idealism in open-source, and the Free Software Foundation does a good job of channelling this. However, for most developers the appeal seems to be one of pragmatism. Open-source tools are often easier to find and evaluate than proprietary products because there is no marketing material to wade through, just technical information. Further, there is no business case to make for choosing one open-source product over another; price doesn't enter into the decision, just the relative technical merits of the products. Business opportunities fall into many categories. One such offers software companies cost reductions by developing only those bits of software that truly differentiate their product: why write a JPEG reader/writer when there is an open-source library, for example? JPEG read/write isn't a differentiating feature; it's a commodity. In the geospatial domain, open-source libraries are finding their way into more and more proprietary products; this generates an opportunity for the software company to make savings and a business opportunity for library developers.

Is interest in business opportunities offered by open-source software coming from smaller companies, or from others too?

Projects like Linux, Apache, Eclipse, Hibernate, and various system libraries, are all funded by technology companies which maintain staff experts to help them integrate their offerings and move the projects forward. The cost of the expert is an order of magnitude lower than that of building a piece of equivalent in-house software. Smaller companies are taking advantage of the business opportunity in providing expert services with the open-source stack. WhereGroup and Camptocamp in Europe are examples of small companies with unique open-source geospatial expertise. As a differentiator, open-source allows them to provide solutions at lower capital cost, and an easier development path using tools that they have distinct expertise in using.

ESRI is working on a spatial extension to PostgreSQL. In doing so, is it not violating the open-source philosophy?

By building its own ST_GEOMETRY type for PostgreSQL, rather than using the GEOMETRY type already available from PostGIS, ESRI has violated the commonsense rule of not reinventing the wheel, but not any open-source philosophy per se. PostgreSQL is itself designed to allow run-time extensions and there are numerous proprietary companies, now including ESRI, that sell proprietary extensions to PostgreSQL for specialised database features.

PostGIS has been around since the turn of the millennium. Looking to the future, what additions to the software would it be wise to pursue?

The rise of amateur geography has put a real premium on the transparent handling of geographic (latitude/longitude) coordinates. Amateur geographers expect geographic coordinates 'just to work', and that requires a proper spherical storage model, index, and analysis functions that respect the poles and international dateline. So the next big extension to PostGIS will be a 'geography' type based on spherical coordinates, to complement the existing Cartesian 'geometry' type. This will, incidentally, benefit some major PostGIS users in NASA and the European Space Agency, who are using PostGIS to index their imagery collections. I would like to encourage readers who are interested in sponsoring development of the geography type to get in touch with me.

Is there is any other matter about which you feel strongly?

It may be that we are experiencing the last decade in which the relational database server is an important piece of infrastructure. The rise of cloud services for data storage and simple indexing schemes using hash keys for spatial indexes are probably going to be 'good enough' for a large proportion of developers of spatial

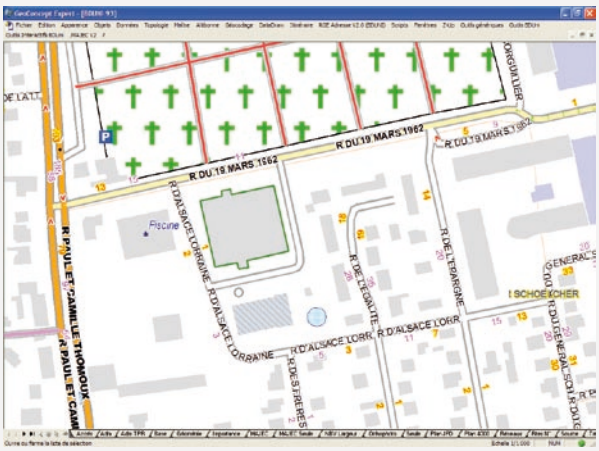

***Figure 1**, Institut Geographique National (IGN) uses PostGIS to store the BDUni topographic base-map for France.*

systems. Already open-source tools like FeatureServer (www.featureserver.org) allow spatial features to be persisted directly into the Google BigTable data store and retrieved in formats like GML, KML, or GeoRSS. At the same time that databases are potentially being made redundant by the 'cloud computing' services, a wave of users is learning about GIS analysis in SQL. A surprising number of users on the PostGIS mailing list are using the database to perform operations (overlays, spatial joins, clustering, summaries) that were previously the sole domain of desktop GIS. As more developers get their first taste of spatial processing via a SQL database, whether it be PostGIS, Microsoft SQL Server or Oracle Spatial, the **desktop** GIS will fade in importance as an analysis tool.◆

Paul Ramsey, OSGeo, Victoria, British Columbia, Canada, email: pramsey@cleverelephant.ca

Desktop GIS will fade in importance as an analysis tool