# BCS Open Source and Geospatial Specialist Groups Talk: 22nd January 2009

## Open Source Geospatial and the OSGeo Foundation

***Jo Cook, Senior IT Support and Development Officer, Oxford Archaeology/OA Digital***

Senior IT Support and Development Officer for Oxford Archaeology (http://thehumanjourney.net), and also a consultant for OA Digital (http://oadigital.thehumanjourney.net), our geospatial consultancy arm. I am also the representative of the UK local chapter of the Open Source Geospatial Foundation (http://www.osgeo.org), and I blog online about computing, GIS and archaeology at www.archaeogeek.com.

## Part 1: Introduction to Open Source GIS

### *Introduction to open source software*

Open source is an approach to design, development, and distribution offering practical accessibility to a product's source- goods and knowledge (http://en.wikipedia.org/wiki/Open_source). In software terms, this means that the human-readable source code for the programme has been made available under a copyright license that meets the Open Source definition, as set out by the Open Source Initiative (http://www.opensource.org/docs/osd). This definition, in broad terms, protects the intellectual property of the author, whilst allowing distribution, and even selling the software, as long as the source code is provided free of charge, and the original terms of the license continue to be observed.

Free and Open Source software are often confused. This is lamentable because people often believe there is a catch if something is free. Also, products such as google earth are free to consumers, but are not open source. It is definitely OK to sell open source software or products, as long as you abide by the terms of the license.

For users and consumers, this tends to mean that the software is given away for free, or for very low cost, without any vendor lock-in. As the user always has access to the source code for the software, there is never a possibility that they can be prevented from using the software (due to changes in licenses, increased costs, companies going bust). It also tends to mean that the software supports open standards for file formats, because there is no financial incentive to lock people in to your software platform. For developers, open source means the protection of your intellectual property, peer-review, transparency and a community of willing participants who want to improve and use your package.

There are established business models around open source solutions. Building solutions from open source components, providing enterprise-level support and documentation, training and accreditation, are all viable options.

### *The development of Open Source GIS (see included timeline for details)*

The development of open source GIS is usually thought to have begun with the GRASS project, initiated by the United States Military in 1982, and made public domain and then open source in 1999. However, in 1978 a vector-based GIS called MOSS was set up by the United States Dept of the Interior. This was based on, and was released as, public domain code (it is still available as open source), and has been cited as the world's first GIS designed to be used on personal computers (rather than main-frames as was the case at the time). Throughout the 1980's, as well as GRASS, the PostgreSQL database was

created, as a spin-off of the Ingres RDBMS, and in 1996 was released as open source. In the 1990's the University of Minnesota's mapserver was created, and important libraries and tool kits such as GDAL/OGR, which are now fundamental parts of most open source GIS packages were started. Also in the 1990's the Open Grass Foundation (which later became the Open Geospatial Consortium) was founded, with the OGC's first specification (simple features) released in 1997. Since then, other desktop GIS packages such as QGIS and GvSIG and other map servers (such as geoserver) and web-based mapping packages such as OpenLayers and Mapguide have been created or made open source. Further standards (WFS, WMS, GML) have been released, and the Open Source Geospatial Foundation was created in 2005.

The development over the last few years has continued apace, with more focus perhaps on web-based packages and desktop competitors. The type of developments that we have seen since 2000, such as all the standards that have been released, really show that the open source GIS world is becoming more mature and influential. OGC standards are followed by most software manufacturers (there seem to be less issues with geospatial standards than for, say, office documents). OSGeo is fast becoming the de facto place to go for sustainable software choices as well as providing an educational/public geodata remit. Their annual international conference, FOSS4G, attracts 800+ attendees from all over the world. There is a strenuous process for admitting projects to the foundation- projects must first go into incubation, and then demonstrate that they have a successfully operating open and collaborative development community, clear IP oversight of the code base of the project, and that they adopt the OSGeo principles and management procedures. They are mentored throughout the incubation process and finally, hopefully, graduate as full OSGeo projects. This helps ensure the sound development principles are upheld, and provides a measure of project sustainability for end-users.

### *Overview of the open source geospatial stack and the key packages within it*

The geospatial stack can be seen as three or four layers, where each plays a separate role in data management and use. It is possible to have a single package or suite that covers all aspects of this stack, but I believe this is a dangerous position to get into, as if one element breaks, all of it stops working. The modular approach, where each component serves a single purpose, can take longer to set up, and perhaps to maintain, but has the advantage that single elements can be replaced or upgraded without danger to the rest of the stack.

The base layer is the data storage layer. The data can be either raster or vector format, and stored as files or in a database. The PostgreSQL database (http://www.postgresql.org), with the PostGIS spatial extension (http://postgis.refractions.net), is a fully-featured server-based relational database, similar to MySQL, with the same multi-user capabilities, but also the added bonuses of transactions which make it possible to "roll-back" a set of changes rather than commit them to the database, locking, which restricts access at row or table-level, preventing database corruption by multiple people updating a given record at any one time, and functions, which allow additional processing to happen on triggers such as when rows are updated or added. The PostGIS spatial extension for PostgreSQL extends it's native, rather limited, ability to handle spatial data, and builds in spatial queries and operators, geometry operations, and other types of spatial management functions so these can be handled at database level without the need for another software package.

The next level up is the servers that translate the spatial data into formats suitable for web use, such as Web Mapping Services (WMS and WFS), style the data, and re-project it into other spatial reference systems as required. The main options here are mapserver (http://mapserver.org) and geoserver (http://geoserver.org), which both have their strengths and limitations and are based on different underlying programmes (cgi and java respectively), but basically do the same thing.

The services layer is provided by the map server. There are two main services, WMS and WFS which respectively provide raster-based (tiles) and vector-based access (GML) to the data. They can be connected to from desk-based and web-based GIS packages. WMS is necessarily read-only access, so is good for providing base mapping or other data that you do not wish people to edit. WFS can be read-only or editable, though not all desktop packages have support for transactional (editable) WFS or WFS-T. WCS, or Web Coverage Service, is related to grid coverages, such as satellite data or digital elevation data. This is less commonly supported. CSW is a catalogue service, related to publishing metadata

about geospatial data.

The top layer is the visualisation layer, either in a web-based or desktop form. Desktop GIS packages can either access data from the web mapping service, or directly from the file or by connecting to the PostgreSQL database. It is key at this point that many proprietary packages can also connect to open source components, although some require you to purchase additional extensions to do this. The main open source desktop GIS packages (of which there are many) are GvSIG, Quantum GIS and GRASS.

GvSIG (http://www.gvsig.gva.es/index.php?id=gvsig&L=2) was created by the Valencia Regional Council for Infrastructures and Transportation in Spain in 2003, and entered OSGeo incubation in 2007. GvSIG was developed as a replacement for ArcView 3, but has considerable additional functionality such as spatial analysis and network analysis, in the form of plug-in extensions. It is currently on release 1.12, but version 2 is expected to be released some time in 2009. A major advantage to choosing GvSIG over other desktop packages is that it supports CAD (DWG files) whereas many open source packages do not, as the DWG format is not truly open. A weakness of the current release, slated to be fixed in the next release some time this year, is the production of high-quality cartographic output. Templates are basic at best, as is fine-grained control over labelling.

Quantum GIS, or QGIS (http://www.qgis.org), is a full OSGeo project, first released in 2002. Out of the box it is more limited in it's capabilities, but this is more than made up for by it's plug-ins, which are python-based, most importantly the GRASS plug-in. This gives access to the full functionality of GRASS, within the more familiar GUI-based environment of QGIS. GRASS (http://grass.osgeo.org/) is also now a full OSGeo project. Traditionally GRASS was purely a Linux/UNIX-based package, and was notoriously hard to install and use on Windows or for people with limited command-line experience. This has now changed, and there is a native windows version, however the learning curve for actually understanding how to do anything is still very steep. Many experienced GRASS users suggest using QGIS with the GRASS plug-in instead!

Web-based GIS packages come in two main flavours- those that require additional software to be installed on the web server, and those that don't. There are also many different options available, depending on your particular programming language skills, or your server setup (Apache or IIS). For many people the best option would seem to be something that does not require additional software to be installed, and doesn't mind which server software it is installed on. I would recommend OpenLayers (http://openlayers.org/). This is javascript-based, so does require some programming skills, but no server configuration. In it's most basic form this allows you to create "slippy maps" of your data on top of base mapping, including from commercial providers such as Google or Microsoft. The Mapfish framework (www.mapfish.org) extends the look and feel of OpenLayers. Modules written in php, java or python can be used to add additional functionality such as formatted printing, or even integration with the Google Earth browser plug-in.

Finally, it's possible to use standard database packages to connect to your spatial data if it's stored in PostgreSQL. Open Office Base and Microsoft Access both allow this if you download an additional driver. This can make data entry very simple, using standard forms, tables and queries that people are familiar with, but it's also possible to use PostgreSQL's user control to limit editing and access capabilities.

## Part 2: A Case-study of running a business using OSS, and making a business out of it

### *Short introduction into commercial archaeology (and why OSS is a good fit for us)*

Oxford Archaeology is the largest commercial archaeological unit in Europe, with over 300 staff across 3 UK offices and 2 in France. Commercial archaeological units exist mainly to fulfil the remit that archaeological and historical remains are not lost during development without first being recorded. Our software and hardware requirements can be quite demanding. We work within a development chain that includes many other contractors, and gather and use many different types of data. We are obliged to work to strict deadlines and produce work not only of the highest quality, but also that survives in

perpetuity often as the only record of the archaeology that we have examined. Whilst individual objects might go on display in a museum, without the data to back them up they provide very little information about the past or the people that used them.

Often commercial archaeological units are educational charities. We have educational and public access remits that we must meet in order to maintain that status, and a board of trustees to make sure that we do. Unfortunately commercial archaeology is very low paid, with most, if not all, of our staff having first and second degrees, yet archaeology graduates are amongst the lowest paid in the job market. Developers see us very much as a hindrance to their work, and it's true that an archaeological dig, whilst being about the cheapest part of the process, can cause massive delays and penalty charges if something unexpected is found. Consequently there is an over-riding force for prices for archaeological work to be reduced.

Recently there has been a move from some of the larger software companies to tighten up on their software licenses, and as educational charities archaeological units tend not to fit easily into a license-type. We are not educational establishments in the sense of universities or schools, yet we are not an attractive market for commercial pricing.

### *Oxford Archaeology's open ethos*

In order to fulfill our requirements as an educational charity and for data retention, we have decided upon an "open ethos", with three strands:

- Open access to our data
- Open standards for file formats
- Open source software

As we have a remit to preserve our data in perpetuity, we have two choices for how to sustain this. We can either devote considerable resources to ensuring that our data formats can be read by all the latest packages, transferring from one format to another as the software companies change formats, or we can choose open file formats and open source software to ensure that we will always be able to read the data.

Our "open ethos" an ongoing process and requires a change in mindset of some of our staff, as well as our clients, all of whom sometimes have an outdated idea about the "ownership" of data, and about the use of open source software (which they perceive as "free" or "dodgy"). Some aspects have worked well, such as being able to use money that would have gone on licensing software and using it for training people instead. Educating staff and clients about the use of standard open file formats is challenging, even when the software that they use is perfectly capable of delivering data in those formats. Some have,however, been very supportive, as they are tied to (for example) local authority purchasing agreements or with outdated versions of software because they cannot afford the upgrade.

### *What we're doing*

This open approach has coincided with (and has required) an investment in IT staff. Clearly transferring to new software packages has had a knock-on effect in terms of the skill-sets we require, and the support we provide our staff. It would be foolish to pretend that there is no knock-on effect to moving to different software, be that proprietary or open source. Our in-house staff have acquired expertise in a diverse selection of software packages and the hardware required to run them. Furthermore, since they are open, developments originally designed for internal use can easily be adapted and extended for external use. OA Digital (http://oadigital.net), our consultancy arm, was formed to market these developments, and our skills and expertise, to external clients.

On the hardware side, we have been trying to identify a suitable platform for on-site digital recording. This has been the aim of technically inclined archaeologists for many years, as it reduces our reliance on paper-based data entry that we later transfer into a database in the relative calm of the site hut or back at the office. The devices need to be robust, and small enough for us to give to all staff, which

discounts many tablet-based pcs. We already give all of our staff a phone, so how about a smart phone? Our choice is the openmoko (www.openmoko.org). This is an open source phone, with a built-in GPS (with the option of plugging in a more powerful aerial), WIFI, bluetooth and accelerometer, on a Linux operating system. It also has USB host capabilities, which means it can have a keyboard plugged in, or a camera to download photos. It will run a mobile version of GvSIG, and the SpatialLite spatially enabled database (http://www.gaia-gis.it/spatialite/). This can be synchronised to our main database servers whenever a phone signal or internet connection is available, using a cross-platform database-agnostic synchronisation package called Sqlsync (http://silvercoders.com/index.php?page=sqlsync).

On the software side, we use the stack described above and are looking at ways of improving the work-flow for getting data in and out of the stack. This is important for staff with little or no experience of server-based databases, but with considerable experience in products like Microsoft Access. We are also helping with the development of the GvSIG package, and have been instrumental in polishing the english translation and producing an improved installer.

More generally, it is clear that many people are put off moving towards an open source solution because they can appear less polished than their proprietary counterparts. Installation and configuration can be difficult, and sometimes the documentation and support forums are geared towards people with considerable experience already. While products like live cds and dvds have existed for some time, they do not deal with the problems of installation and cannot be used in conjunction with the user's more familiar operating system and data- they are simply a sandbox for demonstration and teaching purposes. I personally have experienced the frustration of this, and have come up with the Portable GIS concept (http://www.archaeogeek.com/blog/portable-gis/) that allows users to load a pre-configured software stack onto a USB stick or their hard drive, and have it "just work" with the bare minimum of setup (you simply have to tell it which drive letter you are using).

As well as being popular with new users, other uses for this have been suggested, from disaster response kits where no internet connection is available (the kit could be pre-loaded with data) to demonstration setups for people trying to persuade their bosses that open source is a good way to go without needing to invest time and money setting it up. Version 2 of the package will be out shortly, with the addition of paid-for support and customisation for people wishing to use it in a business environment.

## Part 3: OSGeo

### *Overview of the foundation and it's aims*

The Open Source Geospatial Foundation (OSGeo: http://www.osgeo.org) is a not-for-profit organization whose mission is to support and promote the collaborative development of open geospatial technologies and data.

For developers this manifests itself as providing support for projects such as infrastructure, funding, and legal information, ensuring high quality by mandating an open and sustainable development process, and encouraging communication between the different programming language (python, C++, java) and operating system platforms. For users OSGeo promotes the use of open source software in the geospatial industry, makes software more accessible to end users via binary stacks (like portable GIS), promoting freely available geodata (since free software is useless without data) and providing support for support for the use of OSGeo software in education via curriculum development, support, and outreach.

The big focus each year is the international FOSS4G conference. This takes place in a different country each year, and tries to cover all geographic regions over time, to ensure that it is not focused entirely on the US or Europe. In previous years it has been in Lausanne Switzerland, Victoria Canada, Cape Town South Africa. This year it will be in Sydney Australia and next year in Barcelona, Spain.

## *Local Chapter introduction*

OSgeo local chapters (http://www.osgeo.org/content/chapters/index.html) are set up to provide a geographic or language-specific focus, to support local users and developers. Chapters must prove that they have an established network, and terms of reference, and adhere to OSGeo's main mission statement before they are officially recognised. Once recognised they can use OSGeo branding, but are not financially supported in any way and are subject to the organisational laws of whatever country they are based in.

UK chapter history and terms of reference: The UK chapter was first conceived of at FOSS4G 2006 in Switzerland, but took a while to gain a figure-head and some momentum. The tipping point was in May 2007 where we were allowed to have an informal meet-up at the 1-Spatial Conference at Stansted, and since then we have gained approx 50 interested users, and have gained a profile by attendance at conferences and seminars. We were officially recognised in early January (2009) and have some specific objectives outside of the general OSGeo mission statement, which can be found on our wiki page (http://wiki.osgeo.org/wiki/United_Kingdom):

- Provide a forum for discussion and promotion of Open Source Geospatial Software in the UK, and provide networking opportunities for developers and users
- Help more UK organisations discover the opportunity of open source geospatial tools, and collate business studies of successful transitions
- Raise awareness of the benefits of public access to geodata in the UK by collating links to sources of legitimate free data

In addition we would like to work towards the following:

- A fully-featured open access UK SDI
- Someday hosting the FOSS4G conference in the UK

Our main focus is currently on the UK's first Open Source GIS conference that we are co-organising with the Centre for Geospatial Science at the University of Nottingham in June 2009 (http://www.opensourcegis.org.uk/). This will be a great chance for UK developers and users to get together, and will hopefully give us an idea as to whether we can look towards hosting FOSS4G itself in a couple of years time.

## *Thanks and where to go for more information*

 Please don't hesitate to get in touch if you have any further questions, or visit the UK local chapter of OSGeo and sign up for our mailing list or express your interest on the wiki (sign up for both if you can).

Joanne Cook
Senior IT Support and Development Officer
Oxford Archaeology/OA Digital
j.cook@thehumanjourney.net
+44 (0)1524 880212

## Development of Open Source GIS

## (mostly from http://wiki.osgeo.org/wiki/Open_Source_GIS_History)

**1978** Map Overlay and Statistical System (MOSS): a vector-based geographic information system (GIS) developed by the U.S. Department of Interior

**1982** GRASS (Geographical Resources Analysis Support System): a raster-based GIS developed from 1982 to 1995 by the U.S. Army Corps of Engineers, published as public domain software

**1983** PROJ4 library development started

**1986** Postgresql developed as a spin-off of the Ingres RDBMS

**1992** Open GRASS Foundation (OGF) founded

**1994** OGF was re-structured as the Open Geospatial Consortium (OGC)

**1995** UMN MapServer project started

**1995** International GRASS development team takes over GRASS development

**1996** PostgreSQL goes open source

**1996** Geotools java toolkit started by the University of Leeds

**1997** OGC release the simple features specification that enables diverse systems to communicate in terms of "simple features" based on 2D geometry and associates them with a Spatial Reference System

**1998** deegree/JaGo development started with an OGC Simple Features implementation

**1998** GDAL/OGR development started

**1999** GRASS license changed to GNU GPL

**2000** The OpenGIS Coordinate Transformation Services Specification and the OpenGIS Web Map Server Specification were released

**2001** PostGIS started

**2002** Quantum GIS initial revision in CVS

**2002** GEOS initial revision in CVS

**2002** OpenGIS Web Feature Server (WFS), Geography Markup Language (GML) and Styled Layer Descriptor (SLD) standards were released

**2003** Version 1.0 of Geoserver (web mapping server including Geotools toolkit) released by the Open Planning Project

**2003** GvSIG started by the Valencia Regional Council for Infrastructures and Transportation

**2004** ISO approved an International Standard based on the OpenGIS(R) Web Map Service (WMS) Interface Specification

**2005** Autodesk fork development of their Mapguide Project into an open source fork (Mapguide Open Source) and a proprietary fork (Mapguide Enterprise)

**2006** Open Source Geospatial Foundation established

**2006** OpenLayers started

**2006** Tilecache protocol for tiling web-based maps discussed at FOSS4G 2006

**2007** The mapfish framework for openlayers was announced at FOSS4G 2007