Professionally Leveraging Open Source

Arnulf Christl, Germany

Key words: Open Source, Free Software, GIS, Spatial Data Infrastructure, Quality, Foundation, OSGeo, OpenStreetMap, STDM, Tenure, Land Management

SUMMARY

An introduction to Open Source, Free Software and the Open Source Geospatial Foundation. The first part of this paper details the inner workings of Open Source development and Free Software licensing. It gives a brief history of the development of software in general and explains why for a brief period of a few decades proprietary business models dominated on the market and how they are increasingly replaced by Open Source.

The second part of this paper introduces you to OSGeo's mission, goals and the organizational structure implemented to achieve them. One recent achievement of the educational sector of OSGeo is the Geo-For-All initiative which lead to the emergence of OSGeo-ICA labs all around the world. They provide access and facilities to Open Source geospatial software, infrastructure and curricula.

In the third part we will give an overview of the currently available Open Source tools, programs and software stacks around OSGeo, how and where to get them. The software is collected and bundled on the OSGeo Live distribution available for download and immediate use. OSGeo Live contains software ranging from a full blown spatial data infrastructure to highly specialized tools for data import, manipulation and export and is available completely free of licensing costs. All software is ready to go and is used in professional contexts all around the world. Explore OSGeo on you own by visiting the web site at http://osgeo.org.

Professionally Leveraging Open Source

Arnulf Christl, Germany

1. A Brief History of Computing

The term "software" describes algorithms which are encoded to run on a computer. The working format to develop software is called the source code. It contains all the functionality of a software in a human readable format (a programming language). Once the source code is ready it is compiled into binary code which can then be executed directly by a machine. The binary code of a software can usually not be modified. Instead, to make any changes (fix bugs or implement new functionality) one has to go back to the source code, do the modification there and compile the software into the machine readable binary format again.
The source code may be compared to the form into which concrete is poured. Once the concrete is hardened (has been compiled into binary code) it can be used as the foundation for other constructions (like the end users working with the software).

In early computer history software was considered to be a part of the hardware and every computer needed their very own software which would only work on exactly the hardware it was built for. But as computer hardware became standardized by IBM in the Personal Computer (PC) it became possible to develop software independently from the hardware and still be sure that it would run together. A new business model and market emerged with people starting to sell software separately from the hardware.

At that time it was awkward to install and maintain computer software systems and required special skills. As computers became more and more a commodity people started to expect to receive a fully functional device instead of buying hardware and software separately and having to install it by themselves. This was the dawn of the Microsoft age which lasted well into this millennium and shaped a whole industry – the proprietary software market.
Before the Internet became widely available it was necessary to ship software on a physical data medium like a diskette, tape or later CD ROM. Microsoft managed to convince most computer vendors to ship their computers with a pre-installed Microsoft operating system. Consequently most retail software also aimed for this platform, knowing that it will reach the majority of users. A monopoly developed.

1.1 The Nature of Hard- and Software

Due to the close link between computer hardware and software both were seen as a regular limited good – although they are so very different. When computer hardware is produced and sold something physical has to be shipped and changes hands and ownership. This is very different to software – which can be copied. Creating a copy of a software is almost free of cost and after the process there is more than before – two copies. Still, the ownership of a software typically stays with the creator and does not move to the buyer. So the owner of the software now has two copies
This intrinsic difference becomes more obvious when you try to copy a computer. It will not work. In order to replicate a computer hardware you need exactly the same amount of

physical resources as for the first "copy". You can also not do this at home but need a factory, component suppliers and so on. It is "hard" business and "soft" or virtual business is completely different: You can make as many copies of the same software as you like – anytime, without losing the original.

This posed a major problem to software retailers who were interested in selling each copy of their software separately even although the copy process itself was incredibly simple and cheap. So they tried to make it difficult to copy software. Billions were invested into keeping the source code a secret and into making it impossible to run one copy of a software on different machines. Every computer hardware was furnished with a unique fingerprint and the software checked that fingerprint before starting up. The result was that a software would only run when it was registered with the vendor. The whole point of standardized hardware – to give freedom of choice – was obliterated by this move.

But every lock in history of mankind eventually gets picked. The same is true for software locks. This was the rise of so called "pirated" software – copies which have been modified illegally to run without individual verification and therefore work on any hardware. When mass storage devices like CD writers became available at very low cost things became ever more difficult for vendors. With the advent of the Internet it became virtually impossible to lock down software in a way that would prevent ingenious hackers from picking the locks almost instantly.

The only measure left against lock picking software are legal steps. This was the rise of the proprietary license. Legal documents have a tendency to grow over time and became more and more complex. In their helpless attempt to close every possible loophole that would allow to copy or modify software, some licenses became so restrictive that the software they tried to "protect" could not be run legally at all. Even copying the software into main memory (RAM) of the computer would effectively violate the license conditions.

2. Free and Open Source Software

Free and Open Source Software software development was completely unheeded by this race of locking down software and cracking it up again. Open Source existed since the very first day of computing. Our whole scientific culture functions on a principle of openness: Publish a thesis so that it can be verified or rejected by peers. In the software context this became first know as "Free Software" in October 1985 when Richard Stallman and others started the Free Software Foundation (FSF). Thirteen years later in 1998 Eric S. Raymond and Bruce Perens coined the term "Open Source" and founded the not-for-profit Open Source Initiative (OSI) which tried to circumnavigate the ambiguity of the word free which can both mean free as in "gratis" or as in "freedom".

In most cases the terms "Free Software" and "Open Source" can now be used synonymously as in the acronym FOSS. For the sake of clarification this text differentiates between Free Software as a licensing model and Open Source as a development model. To emphasize the aspect of freedom sometimes the word "Libre" is included as an L to form the acronym FLOSS.

2.1 Free Software Licensing

The word "free" in Free Software refers to a degree of freedom and should not be confused with free as in gratis or in free beer. To make things a bit more complicated most software that comes with a Free Software license is available completely gratis or at a very marginal cost. The emphasis is on the freedom of the user which is why we need to emphasize this again and again. With a Free Software license you are free to:

- use the software anywhere and for any purpose
- take it apart, understand and improve it
- pass it on to anybody else in both the original or a modified version
- make money by using it for any purpose
- improve it in exchange for a monetary compensation or for any other reason
- provide all kinds of services around it including training, services, maintenance, etc.

These levels of freedom make up a Free Software license. For a comprehensive list of approved Free Software licenses please refer to the Free Software Foundation at http://www.gnu.org or the Open Source Initiative at http://www.opensource.org.

2.2 Open Source Development Model

In most cases "Open Source" can nowadays be used synonymously for "Free Software". For the sake of this introduction we will look at Open Source from a development model perspective. As we have learned above the source code of a software contains all the functionality in a human readable format. To change, enhance or extend the functionality of most software the source code has to be modified. Thus Open Source is a precondition to Free Software. End users will generally have no need to look into the source code and only work with the compiled, machine readable version. But it is still important to have the right to look into the software because only then can we fully understand what it is doing. Even if we are not capable of this ourselves, we can give it to someone else who does has the ability to read and understand source code. This will give the user the very freedom which is restricted by proprietary vendor licenses.

All scientific research is based on absolutely transparent reproducibility which is not given if there is no possibility to look into the source code. Therefore proprietary software which does not disclose the source code can also not be used for valid scientific research. This is a simple enough logic ascertainment but still violently opposed because it threatens proprietary business models. Software developers naturally tend to drift to open development models because it makes reusing code a lot easier and allows for collaboration across organizational boundaries. For many newcomers these basic facts are completely new concepts because they have never been communicated by proprietary software vendors. There are two important Open Source Development paradigms named "Publish Early" and "Release Often". They are so efficient that most proprietary vendors have also adapted them to a certain degree.

2.2.1 Publish Early

Open Source software is often started as a solution to a concrete problem. The sooner this solution gets published the better for the project in the long run. This is crucial for all good Open Source software projects because it allows others to join the project at an early stage. If the solution is good then somebody else might also able to use it and also start to contribute. Contribution can come in many different forms, most obviously it can be actual software development but is can also be funding, documentation, testing, even publicity and recommendations. The more people who pick up on the idea, the more testing, enhancements and development the solution will experience. Good communication right from the start is one of the crucial aspects of any Open Source development and sets it apart from closed development.

It can also happen that there already is a solution or other groups who intend to do the same things or have already progressed farther. In those cases it can spare a lot of time and money to not reinvent the wheel and join efforts on a common project. It can also be an incentive to try and be better than the competing project. Diversity is good for any natural development. The traditional Open Source grass roots process is more spontaneous than planned projects that are top to bottom organized.

2.2.2 Release Often

Rapid prototyping and agile computing paradigms are followed by many Open Source projects. Publicly accessible code repositories allow to pick up the latest changes on a daily basis and keep up to date. Changes are documented in the repositories and distributed publicly on mailing lists so that anybody who is interested can follow edits in the code base very closely.

By nature software is always unfinished, there is no perfect software. All errors that are found in the software should be resolved as soon as possible and corrections to the software should be added as patches. A proper software release is more than just the latest code from the repositories, it should be a tested, approved and stable version of the software. Release cycles should not be dictated by commercial considerations or marketing dates. There are two basically different ways to determine release dates. One release type focuses on the availability of a defined set of new functionality that has been added to the Software or whenever a considerable amount of patches has been created and it becomes awkward to stay up to date by applying so many patches.

The other type defines an interval within which the next release is to be published. Depending on the amount of changes the version number will then be raised. For a collection of patches only the third version digit (the patch number) is increased. For changes or additions in functionality the second version digit is raised and the third set back to null. Only deep functional changes, a complete re-engineering of the software and broken backward compatibility will cause the first version number to be changed. Many of these development basics have also been picked up in slightly modified ways by proprietary vendors who release patches and intermediary versions in regular intervals.

2.3 **FOSS Governance – Taking Back Control**

A specialty of the geospatial realm is the high dependency of software on data. Without geodata the software is useless. Other than a text processing software which typically starts with a blank page most GIS software work nowadays starts by loading existing geospatial data. Spatial analysis often requires to process data from different domains and often across different software architectures. The need to be able to access datasets regardless of the hosting software environment is so strong that is is becoming more and more difficult for proprietary vendors to create typical vendor-lock-in situations by implementing closed formats. But the core of the proprietary business model is to maintain full control of the software and to explicitly limit the users' rights on the software. This means that users have no control of how the software is developed or when which new functionality is added or deprecated.

Free and Open Source Software governance models open up this development process. This does not mean that everybody can do anything but that control is managed by a group that takes decisions in an open and transparent way. This decision process can therefore also be influenced by the users, through features requests, funding and direct involvement. There are many different governance models in Open Source depending on the history and size of the project, the intended target audience and user groups. Small projects are often initiated by one single developer who is automatically in control of the code repository. As new developers join the project the initiator might chose to open up governance or keep it centralized. Quite a few projects can flourish perfectly well under a centralized model if the head is a good leader.

Whatever the model – it has to be transparent. With the proliferation of Open Source projects in the past 10 years the need for a central organization which addresses governance and maintains a professional level of quality became more and more pressing. As an answer to these questions the OSGeo foundation was born in 2006.

3. OSGeo: Professionalizing Geospatial Free and Open Source Software

The Open Source Geospatial Foundation, or OSGeo, is a not-for-profit organization whose mission is to support the collaborative development of open source geospatial software, and promote its widespread use. The foundation provides financial, organizational and legal support to the broader open source geospatial community. It also serves as an independent legal entity to which community members can contribute code, funding and other resources, secure in the knowledge that their contributions will be maintained for public benefit. OSGeo also serves as an outreach and advocacy organization for the open source geospatial community, and provides a common forum and shared infrastructure for improving cross-project collaboration. The foundation's projects are all freely available and usable under an OSI-certified open source license.

Founded in 2006 OSGeo is now the leading voice for Open Source in the geo spatial domain. It is broadly inclusive because it is not driven by a single business with corresponding commercial interests but by a broad and diverse community of users, businesses, scientist and

universities. Therefore OSGeo can be the home for projects implemented in different programing languages, different user audiences and a variety of interests.

3.1 **The Structure of OSGeo**

OSGeo's organizational structure has been gleaned from the Apache model and adopted to cater for the special character of the communities that have evolved in the geospatial domain. OSGeo is based on volunteer work and funded through sponsorship. Its main purpose is to provide resources for FOSS4G Projects including infrastructure, legal advice and financial backing.

An additional effort focuses on the promotion of free and open spatial data and access, the creation of a free curriculum and the use of open standards. Another major task focuses on outreach by creating and maintaining a quality brand. As a global organization OSGeo supports FOSS4G (the acronym for Free and Open Source Software for Geospatial) on a global scale and organizes outreach and promotes. The main tasks include supporting local capacities for conferences, appearance at trade fairs and to facilitate inter-project communication to build a solid market for business and the creation of educational material and curriculum.

OSGeo has adopted a do-ocratic way of getting things done. The OSGeo Do-ocracy is implemented following some simple rules:

- If you want to do anything, first make sure the respective community knows about it.
- If in doubt, ask on the OSGeo Discuss mailing list, someone there will know.
- Wait for an appropriate moment (depending on what you want to do this can be one day up to two weeks).
- If there are no complaints, forge ahead.
- If there are complaints, kindly address them, get into a dialog, be inclusive.
- If your plan creates an unsolvable deadlock, push it up to the next level of the OSGeo hierarchy ideally a committee. If that does not work ask the Board of Directors. The ultimate arbitrator is the OSGeo President.

The organization is organized in committees, for example the System Administration committee which maintains the servers, web site, Wiki, bug trackers, mailing list, etc. The conference, finance, curriculum and data committees work independently and people attach to roles as they go. Where formal processes are required members have to be nominated and then get elected into a position, this mainly applies to the board of directors, the president and committee chairs.

3.1.1 OSGeo Local Chapters

One very important aspect of OSGeo are it's Local Chapters which in some cases even predate the global organization (like FOSSGIS in Germany which was founded in the year 2001). To date (March 2014) there are more than 50 local and regional groups associated with OSGeo, ranging from full blown legal entities with their own board of directors, funding and sponsorship schemes to informal groups of enthusiasts.

Local Chapters are often vibrant communities who in many cases also organize local events at existing trade shows and conferences – or even organize their own conferences specifically focused on Open Source Geospatial.

3.1.2 The FOSS4G Conference

The main OSGeo event of the year is the annual FOSS4G conference. Since its first edition in 2006 the conference has grown to an international event with now more than 1000 visitors. In previous years the conference has covered the whole world starting in 2006 in Lausanne Switzerland, then moving to Victoria in Canada, Cape Town in South Africa, Sydney in Australia, Barcelona in Spain, Denver in the USA, Nottingham in Great Britain, and in 2014 to Portland, Oregon, USA.

FOSS4G is a mix of technical presentations, workshops, community meetings, introductions, the open source way of doing things and use cases of prominent Open Source adoptions throughout the world. FOSS4G gives an insight to the bleeding edge of the most innovative software packages, allows for networking and is a must-go for anybody seriously involved with Open Source Geospatial.

3.2 **The Core of OSGeo**

At the core of OSGeo are it's projects. Picking up on the need for quality assurance the first operational committee within OSGeo was the Incubator. It is formed by a broad group of developers from different projects and development languages including the C/C++ and Java tribes as well as more web based technologies like PHP, JavaScript and Python. In many cases several projects focus on the same domain resulting in a competitive situation. When cooperation and competition merge into productive co-opetition the result is highest quality, performance and stability.

3.2.1 The Incubation Process

Projects that are interested in joining OSGeo first have to apply for the official incubation process. After acceptance into the incubation process and finding a mentor the project's code is checked for license and copyright consistency and the governance structure is evaluated. Basic requirements for professional development have to be met including the use of code repositories, bug tracking systems and so on. This assures high quality on a technical level.

3.2.2 Licensing and Governance

OSGeo does not impose a specific license on projects. They can choose their own flavor – as long as it has been vetted and confirmed by either the Open Source Initiative (http://opensource.org) or the Free Software Foundation (http://www.fsf.org/). This gives projects the freedom to choose the license which best fits their requirements. An end user desktop project may want to use a different license than a highly specialized library software which is designed to be used by many other software maybe also including proprietary

packages. This is perfectly allowed, given that the project itself aims for real free and open source governance.

Very similarly governance can come in different styles. Some projects have a documented history of working without a formal structure, others prefer to set up an official project steering committee and hold official elections. However, OSGeo will check the governance model and recommend changes if the current way of doing things appears to contradict the openness which is core to any healthy Open Source project.

No matter which governance model has been chosen for a FOSS project, the license terms allow to use the software in any context by anyone and for any purpose. This also allows to create a so called "fork" giving any user the power to take control over future development of this "branch". This can happen at any point in time and means that any individual or group can take the existing code base and start to develop it in a different direction than the group initially in control had planned. This can have disruptive effects and dilutes the resources so that most projects will strive hard to not let it happen by being as open and collaborative as possible.

3.3 **Organizational Partners**

Openness a key factor in standards development. Education thrives on Open Access, and Open Data is key to innovative work, evolving businesses, e-government and inclusion. The core interest of OSGeo and it's members is software - but software is nothing without knowing how to use it. Therefore an Education and Curriculum committee has formed within OSGeo which contributes tutorials, publications and actively supports other initiatives. All publications through the EloGeo repository (http://elogeo.nottingham.ac.uk) are freely available under a Creative Commons license and can be used freely in any contexts.

Schuyler Erle (http://iconocla.st/cv.html), an early OSGeo activist once claimed that "Without data your software is useless". This is especially true in the geospatial sector where a lot of the really interesting work only starts once the data has been collected and can be used to overlay, intersect and combine with other data.

Therefore OSGeo supports all kinds of Open Data initiatives, has signed official letters to heads of government to support Open Data initiatives, etc. But software being at the core of OSGeo's mind and heart there is little leeway for a broader approach to geospatial. Therefore OSGeo partners with other organizations and initiatives.

Lastly, geospatial data has an intrinsic need to overlap, intersect and be combined with other geospatial data. One important factor to achieve this is interoperability and one way to interoperability is the use of Open Standards. Instead or trying to reinvent the wheel or compete with perfectly well functioning organizations OSGeo seeks to connect with other organization focusing on the above mentioned domains.

Since its inception in 2006 OSGeo has made formal contact with several key partners in openness related geospatial fields of work, three of which are described in more detail here.

3.3.1 Open Geospatial Consortium (OGC)

The first Memorandum of Understanding was set up with the Open Geospatil Consortium (OGC, http://opengeospatial.org). It details a special membership agreement between the two organizations and has lead to significant improvement in the communication between standards and Open Source software development.

It was signed by the OGC's president Mark Reichardt who said: "Openness benefits markets. Vendors of proprietary software have found that today's more open and complex "business ecosystem," which includes both Open Source software and open standards, is good for their businesses. It's also good for technology users. It makes sense for the OGC to work with the OSGeo." Arnulf Christl, then president of OSGeo said that: "We look forward to collaborating with the OGC to identify open source technologies that can be used as reference implementations for OGC standards and to identify standards requirements that result from our open source geospatial software development programs."

One recent result of this fruitful cooperation is the GeoPackage standard of the OCG which is based on the Open Source software SQLite and defines a format to efficiently store and transfer geographic vector features and image tiles.

3.3.2 International Cartographic Association (ICA)

The purpose of the Memorandum of Understanding between OSGeo and the International Cartographic Association (ICA) is to establish a collaborative relationship. Both parties to this MoU share the goal of developing on a global basis collaboration opportunities for academia, industry and government organizations in open source GIS software and data. The MoU covers the following proposals and topics:

- To provide expertise and support for the establishment of Open Source Geospatial Laboratories and Research Centers across the world for supporting development of open-source geospatial software technologies, training and expertise. In the first phase five laboratories will be established, one each in Asia, Europe, Africa, the Americas and Australasia which will act as nodes for future expansion.
- To provide support for building-up and supporting development of open source GIS training materials. For example through the ELOGeo eLearning platform for the Open Geospatial Community.
- Development of collaboration opportunities for academia, industry and government organizations in open source GIS for the purpose of creating a sustainable ecosystem for open source GIS globally.
- Joint organization of open source GIS events, workshops through the ICA network for wider participation globally.
- ICA Commission on Open Source Geospatial Technologies to help OSGeo to establish framework for publications for the academic track of FOSS4G conferences.

One of the most prominent results of this MoU is the Geo for All initiative. The motto of ICA-OSGeo Labs initiative is "Geo For All." The goal of the initiative is to promote and

enhance education, research and service activities carried out by stakeholders in the area of Open Geospatial Science & Applications all over the world. By combining the potential of free and open geospatial software, open data, open standards, open access to research publications, open education resources in Geospatial education and research enables the creation of sustainable innovation ecosystems to advance the discipline. This is key for widening geospatial education opportunities, accelerating new discoveries and helping solving global cross disciplinary societal challenges from Climate change mitigation to sustainable cities.

3.3.3 Global Land Tool Network (GLTN)

The Global Land Tool Network (GLTN) is an alliance of global regional and national partners contributing to poverty alleviation through land reform, improved land management and security of tenure particularly through the development and dissemination of pro-poor and gender-sensitive land tools.

The purpose of the partnership is to establish a collaborative relationship between the Open Source Geospatial Foundation (OSGeo) and the Global Land Tool Network (GLTN). One of the geospatial tools developed by GLTN under the lead of the United Nations Habitat office is the Social Tenure Domain Model (STDM). It is based on a data model implemented using Postgres and PostGIS and a front end plugin based on QGIS. The partnership allows GLTN partners to fund individuals through OSGeo to join forces on projects, collaborate on the software, take part in meetings, conferences, workshops, etc.

4. OSGeo Live: The Geospatial Software Distribution

OSGeo-Live (http://live.osgeo.org/) is a collection of more than 60 Open Source software packages from the geospatial domain. It contains Desktop GIS, Browser Facing GIS, Web Services, Data Stores, Navigation and Maps, Spatial Tools, Domain Specific GIS, Geospatial Libraries and data. All applications can be trialled without installing anything on the computer, simply by booting the computer from a DVD or USB drive, or running in a Virtual Machine environment. Each featured package is accompanied by both a publication quality one page descriptive summary and a short tutorial on how to get started using it.

OSGeo-Live has been created and is maintained by more then 100 individuals from OSGeo and around. It comes as a self-contained bootable DVD, USB thumb drive or Virtual Machine based on Xubuntu, that allows the user to try a wide variety of open source geospatial software without installing anything. It is composed entirely of free software, allowing it to be freely distributed, duplicated and passed around. OSGeo Live provides pre-configured applications for a range of geospatial use cases, including storage, publishing, viewing, analysis and manipulation of data. It also contains sample datasets and documentation.

5. Summary

OSGeo has emerged from the need to professionalize the Open Source geospatial community. This was achieved through diligently working out a model how to develop open source

geospatial software, how to organize the project and help a community to build around it in a sustainable way. Projects are vetted against proven ways of doing open source and are legally sound once they have passed OSGeo graduation.

The future work for OSGeo will focus on broadening the adoption of geospatial and bridge the gap to general IT. GIS has evolved from being a highly specialized standalone software to becoming an integral part of many digital work flows. And the world of IT has recognized that even more openness in education, data and standards will be the key to the future.

CONTACTS

Arnulf Christl
metaspatial, OSGeo
Heerstr. 162
Bonn
GERMANY
Tel. +491722958004
Email: arnulf.christl@metaspatial.net
Web site: http://metaspatial.net