

Your Personal Analysis Toolkit - An Open Source Solution

T. Mitchell

Open Source Geospatial Foundation, OSGeo, Williams Lake, BC, Canada.

ABSTRACT

Open source software is commonly known for its web browsers, word processors and programming languages. However, there is a vast array of open source software focused on geographic information management and geospatial application building in general.

As geo-professionals, having easy access to tools for our jobs is crucial. Open source software provides the opportunity to add a tool to your tool belt and carry it with you for your entire career - with no license fees, a supportive community and the opportunity to test, adopt and upgrade at your own pace.

OSGeo is a US registered non-profit representing more than a dozen mature geospatial data management applications and programming resources. Tools cover areas such as desktop GIS, web-based mapping frameworks, metadata cataloging, spatial database analysis, image processing and more.

Learn about some of these tools as they apply to AGU members, as well as how you can join OSGeo and its members in getting the job done with powerful open source tools. If you haven't heard of OSSIM, MapServer, OpenLayers, PostGIS, GRASS GIS or the many other projects under our umbrella - then you need to hear this talk.

Invest in yourself - use open source!

<http://osgeo.org>

Contact Information

Tyler Mitchell, Williams Lake, British Columbia, Canada

Greetings from British Columbia. Thank you for attending this session this morning and a special thanks to Drew for helping organise this opportunity for us to share about our work.

My talk is entitled "Your Personal Analysis Toolkit - An Open Source Solution".

Open source software is commonly known for its web browsers, word processors, operating systems and programming languages. However, there is a vast array of open source software focused on geographic information management and geospatial application building in general.

I'm going to talk about this software and how open source, freely available, software tools are the core of our future work. By 'our' I refer to those involved in the geo sciences, geography, GIS and more.

An array of mature tools is now available and has come with many great benefits proven over years. I will also **introduce some of these specific tools that you may find useful** and will also direct you to other global resources that you can tap into to for your work that will help you to achieve an efficient end result.

Furthermore, beyond being merely useful and accessible tools I posit that it is integral to our work, research, planning and even our public life that free and open source software be used. When the importance to our daily lives of the 'geo sciences' is appraised, it is incumbent upon us to have ready access to our tools and that all our tools become as natively obtainable to our daily practise as a hammer is to a carpenter - tools that are flexible, adaptable and supportive propel us into our future. A tool that is out of reach is of little use. One that cannot be shared becomes a burden for helping others and delivering solutions collaboratively.

We can all likely recognise the extraordinary value, and fundamental positive

outcomes, that have resulted from the past few years of increased access to mapping technology, generally speaking. Many people have learned how to use several new tools. I believe this trend will vigorously persevere and as such it would be beneficial to consider how the increased adoption will continue to influence our work and lives. **I come back and explain how all this ties into open source software in a few minutes.**

Personal Intro

My frame of reference for this talk is built on a decade of GIS work in forestry, natural resource management and related industrial data management while also regularly playing the data custodian of results of innovative research projects.

I was a relatively early adopter of various open source technologies almost as soon as I started in the industry. The proceeding exploration led me to write a book on the subject, "Web Mapping Illustrated – Using Open Source GIS Tools, published by O'Reilly Media in 2005". The book introduces, in detail, several data analysis tools as well as how to use a spatial database and also basic publishing of mapping data through web services.

I will admit my bias is obviously toward promoting the use of open source geospatial technologies, as part of my current role in working for the Open Source Geospatial Foundation - also know as OSGeo - a non-profit that helps support more than a dozen of these core applications. We are a global volunteer led organisation and invite you to become involved with our community.

I.T Already Knows Open Source

You don't have to look far to see open source software already used in the enterprise. In fact a majority web servers run open source technology and products like Linux are powerful platform for business and home users. Consider just how important and pervasive open source in I.T. has become - barely anything you do on the Internet today can avoid being touched by an open source programming language, web server, operating system, virus scanner, email server or even your web browser may already be open source. Open source in the enterprise is alive and well for all walks of life, from start-ups to the Fortune 500. The debates about open source are over, its strengths and weaknesses are well documented and ready for you to dig into if you have any doubts.

Geography on top of OSS I.T. Is Powerful and Important

I.T. trades in business of bits and bytes of data - email, web pages, messages, etc. Important parts of our daily lives, no doubt. They allow us to do a range of activities from managing banking, looking for jobs and even share our feelings through public forums. These fairly abstract tasks are well managed by open source applications, and likewise so may geographic information.

When we handle geographic information we are manipulating and inspecting the very physical properties of the world around us. Planners design roads, gold mines are excavated, humanitarian projects stand or fall depending on geographic information handling. Yet in many cases some researchers tools are subject to restrictive licensing and/or high costs. So much so that the modern student who graduates from college can rarely afford to be a GIS entrepreneur without dispersing large amounts of capital funds in software purchases alone. Unless, of course, he or she has learned about applications that exist on the open source side already.

Your Tools

The work you do is important and I believe you should have access to the tools you need wherever you are and whenever you need them. The agile researcher today cannot afford to be constrained by software licensing or costs - fortunately in many areas this is no longer a problem as many parts of the geospatial software stack are openly accessible.

Consider for a moment many of the other trades that work on aspects of the physical world. Would a carpenter be able to practise his or her trade without having open access to a toolbox? We currently have the option to choose from a perpetually open toolbox or used one that is more closed. I view it as being able to share your hammer with a colleague or having to lease it back for each job. I believe you need the option to share and use at your leisure.

Furthermore, once these newer tools are embraced you often stand a much better chance of expanding your skillset and offerings as a professional, individual, corporation or other. This diversification of tools can allow you to provide services that you were previously unable to due to cost, availability or long term future prospects.

OSGeo's Ecosystem

Having laid out my position on having open access to the tools that often undergird your professional skills, let's take a brief look at how the open source community, through the OSGeo foundation, operates, then come back to the technical specifics of its software offerings.

Committees

We rely on open public involvement in our various committees of which we have several to choose from. Of particular interest to this audience, we produce a digital **Journal** and we host an **Education and Curriculum Committee** that seeks to find or develop OSGeo-based training material. We

are interested in publishing your perspectives or research using open source geospatial tools and are currently in the midst of our first peer reviewed issue and will be seeking new articles in the new year.

next slide

We have also developed a worldwide network of local chapters or user groups in a dozen languages and dozens of countries. If you are looking to connect with a local regional group or language group, you are welcome to contact me for more information or goto osgeo.org/local. These are opportunities to connect with likeminded users, analysts and researchers.

next slide

Our international conference, called FOSS4G, stands alone as the foremost venue for both learning through hands-on workshops and by presenting hundreds of in-depth case study presentations - all relating to open source technologies being used today in the geospatial arena. I invite you to join us in Barcelona next autumn where our annual event is to be held.

next slide

OSGeo membership is open to all people without charge. There are projects and committees that suit all kinds of interests and personalities, finding a place to fit should not be hard for newcomers. We are overseen by a board of directors but really most work is done by volunteer members and myself as the sole staff person.

We depend on our annual sponsorship programs to fund our promotion and support activities. Additional sponsor positions are available.

Today's Toolkit

Today's most effective geospatial toolkits are fundamentally and structurally differentiated from traditional tools – tools used even as recently as 5 years

ago. The notion of a "traditional" toolkit is obviously relative to the situation and person, but it is no doubt a notion being challenged with great intensity today. In what ways is this challenge occurring? Primarily by the growing availability of tools that are powerful, complex, cross-platform, agile, reliable and also free and open source.

(Free and open source meaning they are developed in an open manner, supported by a community of users and the actual developers of the product and free of charge and with free licenses encouraging widespread adoption and further collaborative development.)

The **_modern_** geospatial toolkit is a **_collection_** of technology solutions - not a **monolithic package** to do all tasks, rather broken down into discrete functions including - scriptable command line processes, programming libraries, desktop applications, database services and web browser-based applications. This is similar to the diversity of tools we use in other aspects of our digital work lives - one tool for email, another for managing photo collections and others for music and word processing, etc. Finding the right tools for the job is becoming more important than remaining dedicated to a particular **_way_** of doing a task, for example.

Software Stack Components

Today's toolkit is a collection of applications designed for use at multiple levels. This diagram shows three technology tiers that are inherent with all projects. At the lowest level, all our projects involve some sort of data. This may be internal data, public data or purchased data for a particular project. OSGeo encourages the use of public data, but our primary interest is in the application of software that sits over top of the data.

The application access **layer** in the stack facilitates access to the source data by higher level applications. This in turn feeds into applications that provides various services. Services including running a task, processing data to be re-used in another context or sharing and interchanging data with others

and other applications.

These three layers in the stack make up an overall solution. They also mirror the functions of the software that OSGeo develops.

OSGeo Project Stack

This next diagram shows how each of the OSGeo applications fits into this stack concept. From the bottom up we have the lowest level programming interfaces providing access to the source data. The programmers among you will find data access drivers to all the major data formats (and many minor ones too) through libraries in C, C++, Java, Python and more. Several proprietary software packages use some of these libraries behind the scenes and they power many online applications as well.

For the programmers in the room, I hope you will go away with at least one clear message: please don't write your own data access code! The odds are that some of our lowest level data access APIs support the format you need. GDAL/OGR, for example, supports dozens and dozens of both vector and raster data formats – I encourage you to make use them instead of going it alone.

Some of these API projects also include command line tools. For example the GDAL/OGR project includes easy to use command line tools for converting and reprojecting raster and vector data from and to these dozens of different formats - all without starting up a new graphical based program or requiring payment of a license fee. These tools are for the power user or in-depth analyst who wants to interrogate their data quickly, make summaries and even filter contents.

On the right side of the Libraries you see two projects that I believe are particularly relevant to this audience. OSSIM – is a remote sensing image processing library. From understanding camera models to mosaicing, the functions run quite deep. It is not meant to analysis as much as it is for processing large quantities of imagery in an efficient manner.

The GEOS libraries support topological operations on your vector datasets. Using a handful of Python bindings you can effectively open and process your geometric datasets using this library. It sits firmly behind the PostGIS spatial database as well, giving it the same functions through an SQL environment.

Application levels here are broken up between desktop and web delivery platforms. Some of the desktop applications you may be familiar with already, such as GRASS GIS. But there are also some other up and coming applications from gvSIG and Quantum GIS in particular – which started as simple viewing applications but that now incorporate methods for GIS analysis. OSSIM is in this list again since it includes several graphical desktop applications for building your processing chain. It also includes the OSSIMPlanet desktop tool for visualising your data in a spheroid-based 3D digital earth.

The web world at the top here is an area of explosive change. From the Geonetwork metadata catalogue up to the OpenLayers web mapping client frameworks – and all the map serving platforms in between – there are many ways to get your spatial data served up through a network and serving them using OGC standard specifications is becoming easier every day.

Stepping back from the details, this stack diagram is a specific instance of a set of tools that you can choose from to build a particular customised solution. From the tools at the bottom end that support data storage, through the lower layer of data access through programming libraries and finally on top the web-based mapping clients, servers and desktop applications.

You have choice.

And for the most part it is up to you to pull together the tools you are interested in, however we also offer a **bootable Live DVD** and various cross

platform package installers. Once you get a taste of this kind of flexibility, it's hard to ever go back to one of the black box approaches. The black box approach is certainly known for keeping a tight, cohesive package. Whereas the modular and diverse open source approach is more 'loosely coupled' together. Open standards and specifications provide the glue across these vertical layers. This is a particular strength of open source toolkits which oftentimes act as the reference platform for standards and specifications. So, when you adopt one of these tools, you also inherit the ability to interact using standards-based approaches with other modules - allowing you to later remove/insert specific tools at will.

Development Statistics

I don't really have time to dive into this slide for you, so just to summarise, we can measure the amount of work being done in a product by observing how many developers submit code to the project. In summary across our 18 projects we have over 10M lines of code, provided by over 500 contributors around the world. Half of those contributors committed code back in the last 12 months – so it's an active ecosystem and is continually growing.

Beyond Technology

When I speak about toolkits I'm referring not only to technology but also to the unparalleled ecosystem of the users of the technology, their support network and the diversity of software development scenarios supporting them all. As with any software, free or otherwise, there are support systems, user groups and discussion forums. Users in our communities have diverse interests and support OSGeo work in different ways – there is a place for everyone who is interested.

This desire to be part of a larger movement or networked ecosystem of users is also pushing us away from the monolithic black box approach and more toward coordinating custom software toolkits with others.

Support Pyramid

To become connected to these 'communities' is to tap in the very heart of the software you are using. Proprietary, or non-free, non-open source, software cannot compete on this angle. In their case, customers and developers largely never meet. Users rarely know any of the plan for software development cycles or features that may be available (or dropped) in the future. And perhaps worst of all, proprietary software can change and likewise **force you to change with it** - as opposed to the open source models that allow you to continue using the software regardless of license payments or subsequent upgrades of the software core. You are in much more control of your implementation.

In fact, because open source development models actively engage the essence of the user community, developers are more reluctant to make massive egregious changes to the core product - thus providing a more stable product over the longer term.

These are only a few of the strengths of open source communities, but the point is that these communities of users and developers play a key part in the software cycle and help make a better project in the end.