

# OSGeo Infrastructure Service Migration Strategy

## *Concepts and Questions*

### **Introduction**

There is some board consensus that the current Collabnet hosting services do not meet all of OSGeo's needs or are inherently expensive for the value gained. In particular the lack of access to the back-end of all tools and the proprietary nature of some of the applications is making it difficult for some projects to justify moving to the platform. Also, the lack of a content management system (CMS) raises the barrier to entry for more contributors and increases the complexity of managing the main web site front-end. These services are also very expensive, which has brought the long-term viability of these services into question.

The largest concern does appear to be fear of vendor lock-in by projects that might otherwise consider moving to official OSGeo infrastructure. It is generally recognised that moving toward common infrastructure has some benefits; this is a desired but not required attribute for graduating from the incubator.

The other area that needs clarification is the availability of the Telascience hardware and Systems Administration Committee (SAC) commitment to supporting mission-critical services running on that system. In order to ensure some specific levels of quality/service, some paid staff/contractors are needed.

The current CollabNet contract ends at the end of the calendar year. This document outlines some of the higher level strategic issues that need to be dealt with if the migration is going to happen in a timely manner.

### **Overview of Current Situation**

Currently, OSGeo relies on four particular service providers as part of an overall infrastructure:

#### **Collabnet**

Mailing lists, main web site, issue tracking, code repository, project web sites, document and file management, support staff. MapGuide currently appears to be the heaviest user of the site's many services and will likely be the hardest hit by the migration off of Collabnet.

#### **Wiki**

The current OSGeo mediawiki instance is informally hosted by Terrestris (<http://terrestris.de>). It is seen as critical communication/collaboration tool.

#### **Other Project-specific Infrastructures**

Only MapGuide is doing all their main hosting through the Collabnet service. Mapbender moved their subversion and mailing list facilities – incurring the loss of their archives – but continue to host their own main web presence. All other projects have been managing their own hosting services and tools. Variations between them are significant.

#### **Telascience**

Various high-end servers and high bandwidth pipes are made available to OSGeo through an informal relationship with the members of telascience.org. Servers are hosted at San Diego State University and made available to OSGeo projects. There is interest in having OSGeo use these servers, particularly for processor heavy functions such as built tests and large scale data hosting. The Geodata Project hosting services are being built on one system. An LDAP authentication instance is also available. The utilisation of these servers for doing smoke test/building tests, etc. has been discussed though their use is minimal so far.

## ***Prototype Web Interface***

To help build comfort and momentum around an improved web page front-end, a prototype install of the Drupal CMS (<http://drupal.org>) was set up. It is currently running on a shared server at Open Source Labs (OSL) at Oregon State University, utilising OSL's MySQL cluster back-end. It is currently accessible at <http://community.osgeo.org/>.

This prototype does not yet address any of the other (non webpage) fundamental requirements for project hosting.

## ***Questions About Current Situation***

- Which projects require or desire to move to a new infrastructure?
- What internal (volunteer) capacity do we have for professional level support and maintenance of our infrastructure?
- What are the bounds between the Systems Administration and Web Committees?

## ***Target Migration Platform***

Once there was a recognised interest in migrating the infrastructure, the question of what components to migrate into started to be discussed. There are two aspects to the target migration platform: hosting provider and tool requirements.

## ***Hosting Provider***

Three main hosting options have been discussed:

The first is to remain with Collabnet for some or all of the current services. The issues around this have already been noted above, mainly: proprietary lock-in, high cost and questionable benefit in some areas. These issues have been discussed at length on various lists.

The second option is to migrate onto Telascience servers. These servers are very fast, appear reliable and have very high bandwidth – perfect for our needs in many ways. However, for many mission-critical applications, this may not be the optimal solution. For example, there is no formal agreement between the volunteers from Telascience or San Diego State University to use these systems. Telascience have been very generous, but it may not be feasible to get long term guarantees or more dedicated formal help when needed. Some services are already being built on this infrastructure and may likely be best suited to stay there.

The third option is to host with a third-party hosting provider. The main one discussed has been the Open Source Labs environment. OSL currently hosts various servers and services for dozens of recognised major open source projects. They are also interested in hosting us. In some cases they will host at no cost (i.e. our prototype web site), but ideally an organisation like OSGeo will have their own server set up on their site and have access to an engineer who can help as required.

## Proposed Provider

The proposed hosting solution is to use OSL for hosting most of the components listed in the Tool Requirements section below: web site, code repository, wiki, mailing lists, bug tracker. The additional items could be moved there as well, but some movement has already been made to host them on Telascience services which seems like a good use of those resources and fits with the goals of Telascience as a group. Telascience hardware could also be used as primary mirrors / offsite backups for OSL hosted services.

There are several benefits to going with OSL: they have staff available to aid in hosting administration, they support many other open source projects, they are very flexible, have good communication infrastructure and are interested in helping provide a solution for us.

We would purchase a server through them, for hosting in their data center. It is proposed that we hire a systems administrator for a few months to aid in setting up our services and help migrate content over to this new service. This will help address the tight timelines that we have for migrating off CN by the end of the calendar year. Afterward, we would have them on retainer (e.g. ¼ time) to help maintain our system over time. It will be critical to maintain a Service Level Agreement (SLA) with OSL to guarantee access to engineering resources and the resolution of issues in a timely manner. Much of the day-to-day work would be maintained by volunteers through WebCom and SAC.

## Hosting Questions

1. Should we hire: a) staff directly through OSL, b) a local person recommended by OSL on contract or c) a 3<sup>rd</sup>-party person from the OSGeo community?
2. How much time/money will be required to maintain the solution in the short term (3-6 months)? In the longer term (6-18 months)?
3. What SLAs need to be in place?

## Tool Requirements

A wiki page was started to help enumerate the tool/service requirement options and ideas: [http://wiki.osgeo.org/index.php/Project\\_Infrastructure\\_Migration](http://wiki.osgeo.org/index.php/Project_Infrastructure_Migration). Ten particular kinds of service needs were identified. Some of them are currently offered as part of Collabnet, others are items that would need to be developed regardless of the migration:

	Currently Available?	Provided By Collabnet?	
1.1 Web Pages	Yes	Yes	Manual update via both CVS and SVN
1.2 Source Code Control	Yes	Yes	SVN
1.3 Bug / Issue Tracking	Yes	Yes	Issue Tracker
1.4 Mailing Lists / Forums	Yes	Yes	ezMLM
1.5 Download Server (source, binary, data file hosting)	Yes	Yes	Project tool
1.6 Wiki	Yes	No	Hosted by Terrestris.de
1.7 Automated Build/Smoke Test System	No	No	Using Telascience has been suggested

	Currently Available?	Provided By Collabnet?	
1.8 Demo Site (for technology demonstrations)	No	No	Using Telascience has been suggested
1.9 IRC	Yes	No	Using Freenode.net service
1.10 Security/Authentication (SSL cert. & LDAP)	Yes	SSL	Telascience hosting LDAP but no management interface

## Proposed Tools

The wiki page has some discussion on the topic of which tools would be preferred or are actively used by various OSGeo projects. This discussion ([http://wiki.osgeo.org/index.php/Project\\_Infrastructure\\_Migration](http://wiki.osgeo.org/index.php/Project_Infrastructure_Migration)) is a good summary of which projects are using various tools already in their workflow. The following summary is a proposed set of tools and where they would be hosted.

	Tool	Location	
1.1 Web Pages	Drupal CMS	OSL	WebCom supports movement to this CMS and has experience maintaining it
1.2 Source Code Control	SVN	OSL	Unclear how many projects would see SVN as a barrier
1.3 Bug / Issue Tracking	Bugzilla or Trac	OSL	Bugzilla and then Jira are the top two most used tracking tools by OSGeo projects. Trac looks worth considering as well and is popular with other open source projects. Consider which one ties in to other infrastructure pieces.
1.4 Mailing Lists / Forums	Mailman	OSL	
1.5 Download Server (source, binary, data file hosting)		OSL & Telascience	Code on OSL. Data on Telascience. Will eventually need to compute our required bandwidth needs.
1.6 Wiki	mediaWiki	OSL	Move from Terrestris.de to OSL. Some types of content could be migrated into the CMS for better management. Not highest priority - is working well.
1.7 Automated Build/Smoke Test System	CruiseControl or BuildBot	Telascience	Various processes currently in use, unclear on amount of work to migrate.
1.8 Demo Site (for technology demonstrations)	n/a	Telascience	Build demonstration apps to run.
1.9 IRC	n/a	Freenode.net	Become an official Freenode project and make donation for use of services. Move logging of IRC from QGIS host to OSL.
1.10 Security/Authentication	LDAP	OSL & Telascience	If used extensively by OSL-hosted apps, then host LDAP on OSL.

	Tool	Location	
	SSL cert.	OSL	Currently held by CN

## ***Budget Requirement Estimates***

The rough estimates for costs show that we need about \$85,000 for the next 15 months, from Sep/06 to Dec/07. The costs for hosting at OSL, with one of their engineers on our case, is roughly known. There are likely additional costs that have not been considered, particularly around any graphic design or website customisation that a systems administration would likely not do.

Service to be hosted at OSL	\$10,000 (one time)
Short term systems administration engineer at OSL (Sep-Dec 2006)	\$25,000 (one time)
Ongoing part-time (25-50% time) systems administration engineer	\$50,000 (annually)
<b>Total for next 15 months</b>	<b>\$85,000</b>

## ***Integration***

One question about these services is how tightly we will be able to draw them together. For example, it will be ideal to bring together CMS, project issue tracking and mailing lists. We will also want to have them all be searchable and feeding into each other easily – e.g. have an easy way to view the status of various parts of the infrastructure through a CMS page. This multi-project management through one common set of services is an ideal end goal; provision of an infrastructure that makes code easier to track, documentation easier to contribute, people easier to communicate with, and software easier to repackage.

The other side to the migration is that of bringing more of the OSGeo projects under one roof; to provide a common presence that will enhance “branding” and co-distribution. To date most projects have chosen to stay with their current infrastructure because of effort required or comfort with their stack of tools. It is hoped that the proposed ideas can be somewhat debated and a happy medium for all projects can be found. It is critical for other projects (that will come on board later) to have the option of moving to a well supported infrastructure as in the proposal. The collective volunteer effort in maintaining their project's systems could be reduced by introducing further cross-project efficiencies.

Mapbender and MapGuide Open Source are two projects with very specific needs if they are to migrate. If Mapbender is happy to be treated as a “guinea pig” for the process, it will make sense to use their services for testing migration during setup

Persistence of previous services is another key feature; particularly for documentation, list archives and distribution facilities; anything that is indexed by a search engine. A mapping between projects' existing services and new ones needs to be maintained; this will also provide comfort for any future migration out of OSGeo hosted infrastructure, if this proves an ongoing concern.

TODO: Timing Summary / Key Milestones & Targets