

OSGeo Infrastructure Service Migration Plan

Executive Summary

The OSGeo Foundation needs to provision essential services to keep itself and its projects online when the current contract with CollabNet terminates at the end of 2006. The Foundation, Committee and Local Chapter sites have a core set of requirements: Web CMS, mailing list management and log reporting, and a versioned repository for document management. In addition, OSGeo software projects require code management and bug tracking facilities. Both of these depend on critical infrastructural components: DNS management, which can be reliably outsourced, and LDAP for single signon across services.

The rest of this document elaborates infrastructure requirements and provides some detail about timescale for the proposed migration. A proposed budget for this project is in the range **\$52-85,000**. This provides for up to 2 servers owned outright by the foundation, full-time software integration work if necessary between October-December 2006, site design consultancy during this phase, and one year of maintenance assurance. It is anticipated that project management would be undertaken by OSGeo's Executive Director.

Ultimately this is a recommendation to pursue part-managed hosting at the Open Source Lab, an independent non-profit project at Oregon State University which currently hosts services for the Apache, Mozilla, Postgres and OpenOffice Foundations amongst others. Due to time constraints, the services that OSGeo builds and provides by December should be the minimum needed to keep the Foundation, and the MapGuide and Mapbender projects, running. The aim is to have infrastructure in place that does at least as much as the current service hosted on CollabNet, and to build up more ancillary project-oriented services according to their needs over the next year.

Overview

The Foundation has a set of core tool requirements to meet regardless of whether it is hosting collaboration and distribution infrastructure for its projects as well.

The Foundation overall needs Web content management with translation facilities, blogging and news aggregation facilities, access logging/reporting, mailing list management and document repository. This applies to the main www.osgeo.org site as well as the non-software committees (Geodata, Education, Conferences) and for each local chapter that needs web and mail management facilities.

In addition, the software projects have require code repository management and bug tracking software. Currently, many projects are hosted externally (on non-OSGeo services) and there are varying levels of interest in moving to OSGeo-hosted facilities. Most projects have been reluctant or uninterested in moving to the current Collabnet services.

There are three primary principles for the design of our target infrastructure: tools should not cause 'vendor lock-in', if projects leave OSGeo's infrastructure they should still be able to use the same tools without a license fee, and we aim to use tools that have adopted open standards.

Background

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 their long-term viability into question.

The largest concern appears to be fear of vendor lock-in by projects that might otherwise consider moving to official OSGeo infrastructure. Moving toward common infrastructure has been recognized as beneficial; 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 specific levels of quality/service, some paid staff/contractors are needed.

The current CollabNet contract ends at the end of the calendar year, Dec/06. 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 its overall technical infrastructure:

CollabNet

Provides mailing lists, the main web site, issue tracking, code repository, project web sites, document and file management, and technical support staff. MapGuide is the heaviest user of these 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 (external) hosting services and tools. Variations between them are significant.

Telascience

Various high-end servers and high bandwidth pipes have been 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 To Be Address 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 and where to host them 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 contracted 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. It is proposed to use Telascience hardware as the primary North American mirror for offsite backups of 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 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 by the end of the calendar year. Afterward, we would have an administrator 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.

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. Ten particular kinds of service needs were identified and are listed in Table 1. Some of them are currently offered as part of CollabNet, and other items would need to be developed regardless of the migration. A couple of areas for future consideration are also noted (1.11 & 1.12).

	Currently Available?	Provided By Collabnet?	
DNS Management	Yes	Yes	Built into CN system
Mailing Lists / Forums	Yes	Yes	ezMLM
Security & Authentication	Yes	Yes	CN hosted SSL certificate
SSL			
LDAP	No	No	LDAP integration prototype on Telascience host
Web Pages	Yes	Yes	Manual update via both CVS and SVN
Wiki	Yes	No	Hosted by Terrestris.de
Bug / Issue Tracking	Yes	Yes	Issue Tracker
Download Server (source, binary, data file hosting)	Yes	Yes	Project tool

	Currently Available?	Provided By Collabnet?	
Automated Build/Smoke Test System	No	No	Using Telascience has been suggested
Demo Site (for technology demonstrations)	No	No	Using Telascience has been suggested
IRC	Yes	No	Using Freenode.net service
Language translation tools	No	No	a translation portal for software user messages (e.g. rosetta/launchpad, KDE kartouche)
Communication servers	No	No	Web-based IRC client hosting. Voice conference server (Asterix, YATE, etc.)

Table 1: Tool requirements and current situation

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](#) is a 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. They are listed in order of priority for implementation.

Order		Tool	Location	
1	DNS Management	n/a	TBD	Outsource DNS services to third party, set up temp DNS name to work on during migration (e.g. osgeo.net)
2	Mailing Lists / Forums	Mailman	OSL	Need to ensure migration of archives. Use of forums needs to be assessed.
3a	Security & Authentication	SSL cert.	OSL	Currently held by CN
3b		OpenLDAP	OSL	If used extensively by OSL-hosted apps, then host LDAP on OSL.
4	Source Code Control	SVN	OSL	Need to ensure migration of history.
5	Web Pages	Drupal CMS	OSL	WebCom supports movement to this CMS and has experience maintaining it. Serves as a powerful base for other web reporting and membership management needs.
6	Wiki	mediaWiki	OSL	Move from Terrestris.de to OSL. Some types of content could be migrated into the CMS for more official management. Not high priority - is working well.
7	Bug / Issue Tracking	Trac	OSL	Trac is proposed as the bug/issue tracking tool. It has several methods for tying into

Order		Tool	Location	
				other parts of the infrastructure, e.g. SVN and other features. Unclear how easy to extract from current CN tracker.
8	Download Server (source, binary, data file hosting)		OSL & Telascience	Code on OSL. Data on Telascience. Will need to compute our required bandwidth needs.
9	Automated Build & Smoke Test System	Cruise Control or BuildBot	Telascience	Various processes currently in use, unclear on amount of work to migrate.
10	Demo Site (for technology demonstrations)	n/a	Telascience	Build demonstration apps to run.
11	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.
12	Language translation tools	TBD		
13	Communication servers	TBD		

Table 2: Proposed tools and locations

Budget Requirement Estimates

The rough estimates for costs show that we need about \$52-85,000 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.

Purchase of 2 servers to be located at OSL colocation site	\$7-\$10,000 (one time)
Short term systems administration engineer (Oct-Dec 2006)	\$15,000 (one time)
Design / Interaction development	\$5-10,000 (one time)
Ongoing part-time (25-50% time) systems administration engineer	\$25-50,000 (annually)
Total for next 15 months \$52-85,000	

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. Initially this will be done through the CMS as much as possible. However, in the longer term a management framework such as Gforge may need to be considered. Having multi-project management tools through one common set of services is an ideal end goal, providing 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 (external) 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.

Persistence of previous services (URLs, protocols, etc.) is an important feature to aim for, 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 wherever possible. Having a thoughtful plan for this will help to make migration into or out of OSGeo hosted infrastructure less painful.

Key Milestones

Timelines are highly dependent on the resources available. The table, below, is a very simple example based on using volunteers. The values are somewhat meaningless except that they show the general timelines required to meet the year-end migration deadline. At present, timelines depend ultimately on the capacity of SAC and WebCom. When the Executive Director (E.D.) begins, there will be more dedicated focus on successful implementation within these timelines.

The hiring of a dedicated systems administrator will help set up, test and migrate services more quickly. The following table outlines a rough potential timeline, assuming the help of a sysadmin and E.D.

		Delivery Date
1	Approve service provider	22-Sep-06
2	Contract set up for sys. admin	29-Sep-06
3	Finetune migration plan	29-Sep-06
4	Approve migration plan	29-Sep-06
5	Server purchase and set up	6-Oct-06
6	Install tools	13-Oct-06
7	Migrate content & services complete	30-Nov-06