OSMGB as a service - a proposal

Anthony Beck, Amir Pourabdollah, Jeremy Morley

Abstract

Please add in your name to the author list.

Written using CommonMark formatting: an unambiguous implementation of Markdown for scholarly writing.

Google doc - https://docs.google.com/document/d/1nBjBgxzuHPoCEWvFeNXZjt33xfx4BRyIedslOxiQZa0/edit

Can be rendered to other formats as using Pandoc. To render to PDF with a table of contents use the following:

cd /home/arb/ownCloud/Nottingham/OSMGB/Git/OSMGB/Documents/2015_ServiceRessurection

pandoc -f markdown -t latex -N -V geometry:margin=1in 2015_OSMGB_Action.md –filter pandoc-citeproc –latex-engine=xelatex –toc -o 2015_OSMGB_Action.pdf

Contents

1   TL/DR Executive summary

The good news is that Amir and myself would like to resurrect the OSMGB service. In terms of a future service we would propose the following:

- Deprecate the riskier components:
  - Noteably WikiMedia and Wordpress (migrate or archive the content as appropriate)
- Upgrade the 1Spatial component on *Server 1* thereby removing the Windows 2003 dependency
  - Verbally agreed with Mike Sanderson

The bad news is the infrastructure is at risk (in terms of UoN operating policy) and is out of warranty. We propose the following:

- Determine if UoN will allow current servers to be maintained (they are out of support and one runs an *at risk* operating system (windows 2003)).
  - Windows 2003 risk
    * update to Windows 2008
  - Warranty
    * Need to determine if this is possible within UoN
- The alternative is to migrate current services to:
  - new hardware
  - virtual server farm (which may be more appropriate for ongoing management)
- All of the above will come at a cost

In terms of on-going maintenance and management Amir and myself are not funded for our roles. Whilst we can provide time to manage the services we can not manage the environment. I have spoken to a number of people about moving the resposonsibility of the server environment (and their associated services) to others in the UoN (principally through Erin Snyder). This is, in theory, very possible. The University has recognised that supporting high impact services can go beyond project resposibility *and* can reflect positively on the university as an institution. We will need to provide a *case for support* which details evidence of: * External demand * Impact * etc.

General strategy would be as follows:

- Resolve the hardware/server issues
  - Secure funding to implemented the resolved approach
  - If necessary mothball current servers so they do not present a risk to UoN but do provide access to data and resources.
- Migrate or build new environments as appropriate
- Request on-going support for the service environment from UoN

Task list

- Resolve hardware:
  - Anthony Beck
  - Shaun Hindley

- Ian Hill
- Others?

- Funding requests

  - Ordnance Survey - Jeremy Morley
  - Mike Sanderson - 1Spatial
  - Others?

- Software requests

  - Mike Sanderson - 1Spatial (1Integrate)

The rest of this document provides background, details and thoughts.

2 OSMGB Background

The OSM_GB project was described as follows:

> The project is a co-operative research project between the University of Nottingham's Centre for Geospatial Science, and 1Spatial. We will be exploring means to apply country-wide data quality improvement rules to Open Street Map data of Great Britain with a transformation to the British National Grid projection to produce an "OSM-GB" product.
>
> Alongside production of an OSM-GB output we are interested in how 1Spatial's Radius Studio product, which we will be using for the rule-based quality improvement, can be used to adapt the output for different uses and to report locations of errors to feedback to the OSM community.
>
> The OSM-GB outputs are not intended to replace the main Open Street Map database but are intended to provide a version with specified data quality characteristics. The OSM-GB data will not be directly edited by users – updates will be carried out in the source OSM data and will then propagate through to OSM-GB via a regular update cycle.

The project itself was completed in 2013. The service was left intact but no resources were in place to manage, maintain and migrate the service. Inevitably the service was hacked and is now unavailable. Prior to the service going down it was in use by a range of different stakeholders from industry, research and the public sector. Whilst there was some goodwill surrounding the service this will evaporate the longer the service is down.

Jeremy Morley, Amir Pourabdollah and Anthony Beck had a brief meeting on 20150318 to discuss what could be done with the service. All three felt that the service should be resurrected. Ideally it should have the following characteristics

1. The original OSM_GB infrastructure should be pared to expose only the critical elements

   - Remove the wordpress which is an on-going security risk

2. The service should remain within an *academic* setting

   - This should also obviate any UoN Intellectual Property issues (if applicable - it would be nice to ensure that these were clearly stated up front)

3. Be free at the point of use and available to all

   - This does not entail any service level commitments

4. The 1Integrate service (Radius Studio has been re-branded as 1Integrate) will need licencing

   - Provided by 1Spatial?

5. The 1Integrate service will be re-usable as a SAAS research and teaching platform within Nottingham University

   - Managed by Anthony Beck and 1Spatial

6. The workflow framework for OSM_GB will be managed by Anthony Beck with support from Amir Pourabdollah

7. The service could be enhanced by undergraduate or masters research projects

   - There is no expectation that the service shall be developed or enhanced

8. Server management should be institutionally managed by Nottingham University

The most important issue is the final point. The underlying workflow and analytics for OSM_GB are still valid - the problem is the resources to manage the server environments. Whilst this is viewed from a project perspective it will always remain a problem. However, the underlying landscape of academic infrastructure provision is changing. The government, through the funding councils, is starting to change the way that academia publishes and shares research outputs and is encouraging universities to maximise research impact within public and industrial settings. This is most obviously seen in the *open access* initiative to improve availability to research papers. Improved access to data, under more open licences, has been recommended by RCUK and is mandated by EPSRC. The underlying corrollary is that data analytics and data services will be part of the future of university research outputs. This will require the management and maintenance of services like OSM_GB. It is hoped that in the long-term OSM_GB will become part of a suite of managed services offered by the University of Nottingham in perpetuity (or at least as long as they can demonstrate there is a demand for the service and that the service is providing impact which is beneficial to UK funding agencies).

3   Infrastructure summary

Figure 1 describes the relationships between infrastructure, data software, and services used in the OSMGB project.

The data workflow represents a cycle of change which is repeated on a daily basis. The data processing workflow is controlled by a *Windows Schedular* process which orchestrates the web processes exposed throughout the framework.

3.1   Data summary

OSMGB used the following data:

- Open Street Map (OSM): OpenStreetMap country dump files taken from GeoFabrik in *.osm* format.
- OSM Updates: The daily updates of OSM taken from planet.openstreetmap in *.osc* format.
- Ordnance Survey OpenData: Ordnance Survey's OpenData in ESRI shapefiel format. Including:
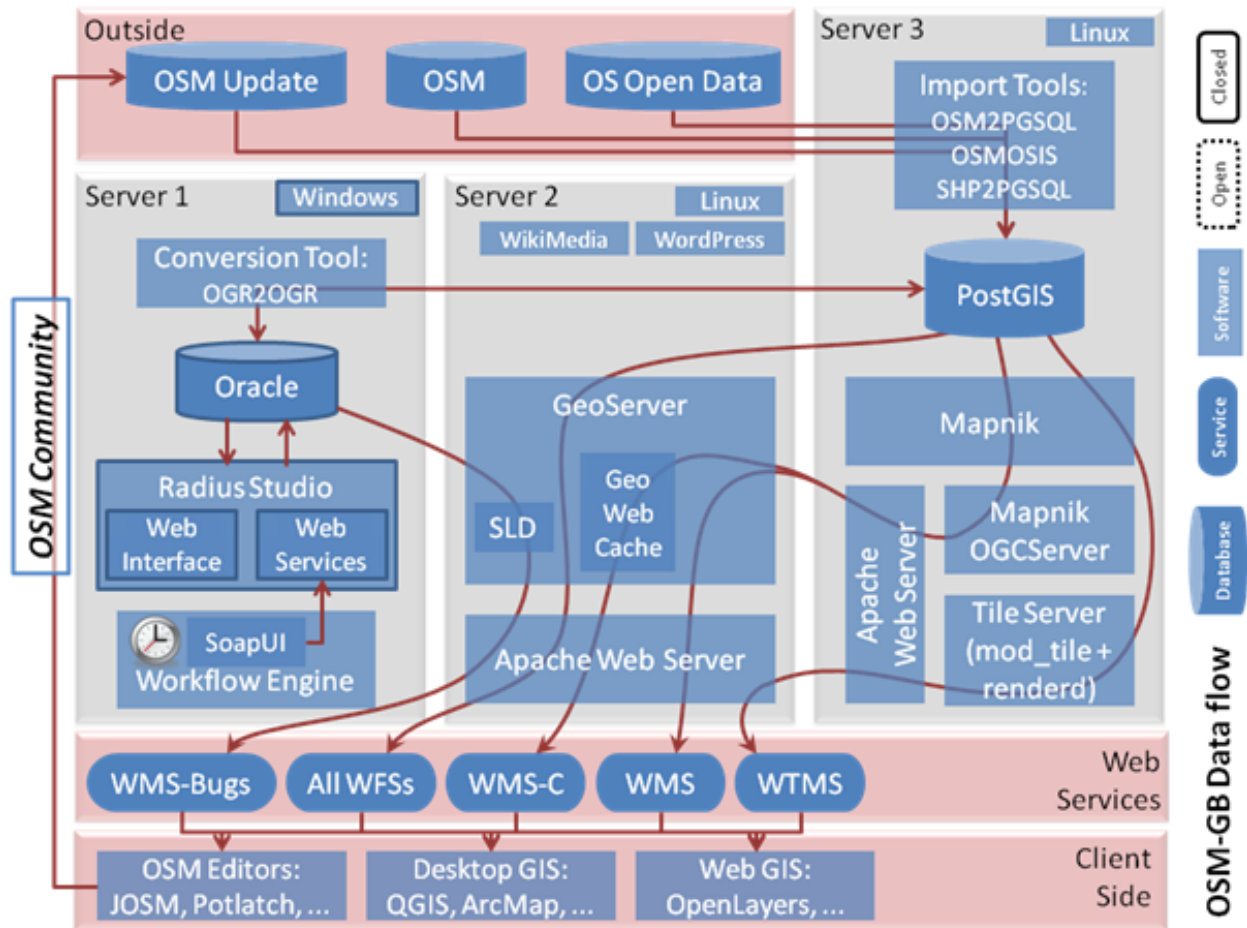 - Vector Map District (VMD)
  - Meridian

Figure 1: The original OSMGB diagram detailing the relationships between infrastructure, data, software and services.

3.2   Hardware summary

The project uses three servers:

- Server 1 (Linux) - lgosmgb1.nottingham.ac.uk: PostGIS and geo-tile server
- Server 2 (Linux) - lgosmgb2.nottingham.ac.uk: Web Server
- Server 3 (Windows 2003) - lgosmgb3.nottingham.ac.uk: Oracle and RadiusStudio data processing server

each has the same hardware specification:

- 24 core
- 33 Gig Ram

Server 3 is *at risk* as the operating system *Windows Server 2003* has no future support path. Server 2003 was a dependency of the 1Spatial *RadiusStudio* software. RadiusStudio has been rebranded as *1Integrate* and can be run on Linux and Windows 2008 server platforms.

ALL servers are out of warranty.

3.3   Software and services summary

The original infrastructure ran the following software:

- Server 1 (data server):
    - Linux O/S - update and retain
    - PostGIS - update and retain
    - Geo Tools - update and retain
        * OGR
        * OSM2PGSQL
        * OSMOSIS
        * SHP2PGSQL
    - Mapnik - update and retain
- Server 2 (web server):
    - Linux O/S - update and retain
    - GeoServer - update and retain
    - WikiMedia - deprecate and migrate to *Wikipedia*
    - Wordpress - deprecate, host as static files and archive with the British Library UK Web Archive
- Server 3 (data processing server):
    - Windows 2003 0/S - upgrade to Windows 2008
    - Oracle 11g with Spatial extension - update and retain
    - Radius Studio - upgrade to 1Integrate

We have included proposed changes to the environment to ensure compliance with UoN operating procedures and to reduce the risk of hacking. In terms of the latter it is recommended that WikiMedia and WordPress components are deprecated with their content migrated to either Wikipedia or archived at the British Library UK Web Archive. This was the most at risk element of the previous service. This means that the only publically exposed element is GeoServer via Apache on Server 2.

4   Proposed infrastructure needs

Fill in at a later date

5   Proposed Tasks

5.1   Resurrect the service (kick off - end of April 2015 - completion May 2015)

The aim of this task is to get the services back up and running

Tasks include:

- Knowledge transfer from Amir to Ant
- Server patching and potential consolidation (i.e. moving services from the linux server to one of the windows servers)
- Paring down OSM_GB content
    - Removing the WordPress component (or migrating this to an external host via a pointer)
    - This may include scraping and then archiving the project web-pages with the British Library
- Configuring the service based upon server configuration
- Testing the service
- Publicly exposing the service

5.2   Securing and maintaining the service platforms (post May 2015)

The aim of this task is to provide long-term hosting and associated server/DBA management.

By removing the at-risk components (such as Wordpress) the service will become more stable. However, it will still need managing in terms of server and database maintenance. In terms of the latter there is DBA level database maintenance and operational level maintenance.

Ant and Amir will look after the operational maintenance. However, the server and DBA components need resourcing. This can be done in one of two ways (assuming it stays within the University of Nottingham):

1. The UoN takes ownership of the server and DBA maintenance at zero cost (as part of building a data analytics environment)

    - This has potential but may take some time to implement

2. The UoN is paid to manage this environment

- Find out how much this would cost?

5.3   Determining credit and impact (on-going)

Impact and credit may be key in terms of getting buy-in from key stakeholders. As already discussed the government is redefining the metrics surrounding the value of University outputs. Cross sector/societal impact will be key to this. This service will provide such impact. However, in the first instance this may need demonstrating to the university. This can be demonstrated through letters of support[1]. We could ask for letter of support from the following (with suggested themes)[2]:

---

[1] any other approaches that are quick and easy?

[2] first cut - please add more

- University of Nottingham (Jeremy, Muki, etc.) on the theme of:

  - General societal impact
  - Teaching and Learning resources
  - Provision of stable data analytics/web services that are a bridge between research and industry
  - Ways of building on and exploiting OSM

- Industry (Mike Sanderson, Harry Wood, Mike Saunt, Steven Feldman, AGI, etc.)

  - Quality and provenance
  - Platform to build on
  - Pathways to exploiting research
  - Use of OSM in UK projection
  - Ways of building on and exploiting OSM

- Public Sector (whom?)

  - Quality and provenance
  - Platform to build on
  - Pathways to exploiting research
  - Use of OSM in UK projection

- GDI (Jeremy)

  - OS

In addition to such letters of support those who contribute to the project should be able to exploit this in their own PR.

5.4 Longer term developments

A place where I've just put some thoughts

5.4.1 Change management

How do we institute and manage change, provide service stability (i.e. if products are built on this then changes to the way the service is exposed can break downstream