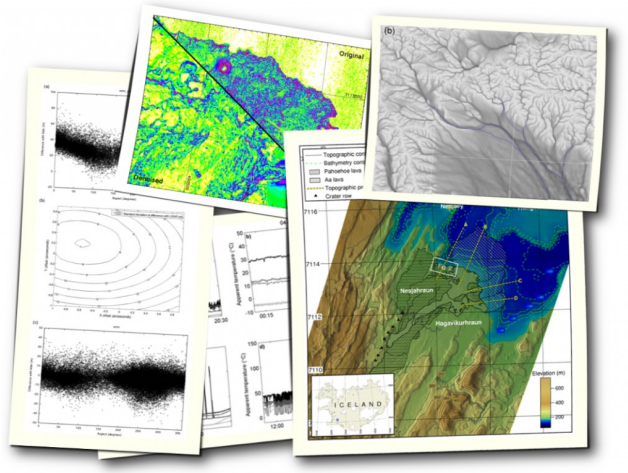
Case study

Title:	An Open Source software stack for geoscientists
Developed for:	Edinburgh University
Sector:	Geoscience Research
Date Completed:	November 2011

Geoscience research involves processing data, usually in a geospatial context, and presenting the results as journal articles or as conference talks or posters. This case study describes the software stack used to complete a number of projects including mapping Icelandic lava flows, testing smoothing algorithms for DEMs and investigating the products of explosive volcanic eruptions.



Operating System

The GNU/Linux operating system was chosen, mainly because of the command line interface (CLI), which can be used to carry out tasks quickly and precisely. Commands can be scripted in order to carry out repeated or automated tasks.

The Ubuntu 11.10 version of Linux was chosen because it has a wide range of software available in easily-installed packages. The names of the Ubuntu software packages for each program are given below. Windows and Mac versions exist for most.

Maps and Geographic Information Systems

- **GRASS** (*grass)*: Fully-featured and extremely-powerful GIS package with both GUI and command line interfaces. It handles raster and vector data in all formats and is easily scriptable to automate workflows. It was used to create new GIS datasets from raw data e.g. by processing LiDAR point clouds, digitising field maps, image analysis of multispectral remote-sensing data.
- **Quantum GIS** (*qgis, qgis-plugin-grass*): Easy-to-use GUI-based GIS package. It is ideal for making and printing maps from pre-existing datasets. It also has a nice georeferencing tool and can be used as an interface to GRASS GIS.
- **GDAL** (*gdal-bin*): A command-line swiss-army-knife for GIS files allowing you to convert formats, change projection, join, crop and alter the resolution of raster files and much more. Includes OGR, which does the same with vector files (e.g. shape files).
- **Proj4** (*proj-bin, proj-data*): Command line tools used to reproject data points in different map projections (*cs2cs*). This works behind-the-scenes of GDAL.
- **Generic Mapping Tools** (*gmt, gmt-coast-low, gmt-doc*): Command-line tools used for plotting publication-quality maps of geophysical data.
- **Google Earth** (*manually installed*): 3D globe for plotting data in global context. Not FLOSS.
- **GPS Babel** (*gpsbabel*): Used to communicate with handheld GPS units, and convert formats between gpx, kml, and garmin. The Windows version has a graphical user interface.
- **GPS Prune** (*gpsprune*): GUI-based tool for editing GPS point and track data. Used to geotag photos for viewing in Google Earth.

Data Processing and Plotting

- **Python** (*python*): Open source, cross-platform programming language. It is widely-used by scientists and is extremely versatile because it can be easily extended using addon modules such as these below. Used for most of the same purposes as Matlab, but more portable because anyone can install Python to run the scripts.
- **Spyder** (*spyder*): A development environment for Python, giving a Matlab-like appearance and with features such as code-checking, command completion and automatic display of documentation for the current command / object. Used to write python code.
- **Numpy and SciPy** (*python-numpy, python-numpy-doc, python-scipy*): Scientific and numerical computing modules for Python, allowing it to handle arrays of numbers.
- **Matplotlib** (*python-matplotlib, python-matplotlib-doc*): Plotting modules for Python are used to make all kinds of publication-quality 2D and 3D figures.
- **Basemap** (*manually installed*): Add-on for Matplotlib giving Python similar map-plotting functions to those of GMT e.g. plotting in different projections, adding coastlines or the Blue Marble image).

- **SQLite** (*sqlite, sqlite3, sqlitebrowser*): Open source database format that can be accessed via the same Structured Query Language used by cutting-edge data servers, but the data are stored in a single, portable file. Used to store sample data.
- **SQLiteManager** (*Firefox plugin*): A nice viewer used to edit and perform queries on SQLite databases.

Writing Journal Articles / Reports

- **Zotero** (*Firefox plugin*): Reference manager software that runs in Firefox and can add articles to the database directly from the journal website or the results page of a Web of Science query. Used to organise references, to insert references into Word or Writer documents or to prepare BibTex files.
- **LaTeX** (*texlive, texlive-latex-extra, texlive-fonts-extra, texlive-humanities + others*): LaTeX is an open source typesetting program. It is used to produce beautifully laid-out pdf documents from plain text files containing the text and some simple formatting codes e.g. *\section{Introduction}*. LaTeX does referencing, section numbering, figure captions and tables of contents automatically.
- **LibreOffice Writer** (*libreoffice-writer*): Open source word processor. This is an ideal substitute for Microsoft Word on all platforms, as it can read and write .doc and .docx files. Used for collaborative writing requiring features such as comments and track changes, which work perfectly.

Conference Presentations

- **Scribus** (*scribus*): Professional quality desktop publishing package used to make conference posters. It is very easy to create good-looking layouts, align images and set font-themes. The output is a pdf file that you can print anywhere.
- **Beamer** (*latex-beamer*): Used to make pdf-format conference slides in LaTeX. It has all the advantages of LaTeX e.g. beautiful results, no-fussing about layout, referencing and contents all taken care of. Plus the pdf files don't get messed up between Mac/Windows/Linux versions like Powerpoint slides can.

Images, Graphics and Photos

- **Gimp** (*gimp*): The Gnu Image Manipulation Programme is equivalent to Adobe Photoshop or Corel Photopaint. Used for photo editing.
- **Inkscape** (*inkscape*): Inkscape is a vector graphics package equivalent to Adobe Illustrator or Corel Draw. Used to draw diagrams.
- **Image Magick** (*imagemagick*): Command-line tools used for automatic or batch processing of image files: resize, rotate, label, crop, change format etc.
- **Shotwell** (*shotwell*): Photo viewing programme for viewing images using tags, ratings and events. Used to organise field photos.

Videos and Media

- **VideoLan Player** (*vlc*): Used to play video files in many formats.
- **Openshot** (*openshot, openshot-doc*): Used for simple video editing.
- **FFmpeg** (*ffmpeg*): Command-line tool to change the size, framerate, format etc. of videos. Used to extract sound from video files.
- **Audacity** (*audacity*): Used to edit mp3 and other sound files.

Computer Administration Tools

- **Open SSH** (*openssh-client, openssh-server*): Used to connect securely to machines across the internet without the fuss of a VPN, in order to monitoring running jobs or to use a secure FTP program such as WinSCP to copy files.
- **Rsync** (*rsync*): One-way synchronisation over SSH. Used to automatically back desktop machine to the department server. It knows which files have changed and only sends the differences, so it runs very quickly. **Unison** (*unison*) gives two-way synchronisation e.g. between laptop and desktop machines.
- **WINE** (*wine*): Environment used for running (most) Windows programs on Linux machines, e.g. the simple panorama-making software, Autostitch.

Further reading

Futher information on these software, including installation scripts, can be found at:
http://all-geo.org/volcan01010/2011/11/all-the-software-a-geoscientists-needs-for-free/

Developed by:	Dr John A Stevenson, Edinburgh University
Contact:	john.stevenson@ed.ac.uk