1. Introduction
The current widespread adoption of free and open-source software (FOSS) geographical information system (GIS) has been well documented in the academic literature (Neteler and Mitasova, 2008; Steiniger and Hunter, 2012) and is evidenced by the large number of actively developed FOSS GIS and geospatial libraries (e.g. QGIS and PostGIS), vibrant GIS communities (e.g. the OSGeo Foundation and the GIS StackExchange), and large attendance and support of the annual FOSS for Geospatial (FOSS4G) international conferences. The increasing FOSS GIS usage has paralleled broader trends in the widening application of GIS within society as well as the increased prevalence of the FOSS model for software development (von Hippel and von Krogh, 2003). Many of the current FOSS GIS projects have achieved mature and stable platforms for geospatial analysis and visualization with large, international user communities. This paper introduces a FOSS GIS project called Whitebox Geospatial Analysis Tools (Whitebox GAT) and explores some of the project’s main design goals as it relates to the concept of open-access software.

2. History of the Whitebox GAT Project
The Whitebox GAT project began in 2009 through the development efforts of researchers at the University of Guelph, Canada (http://www.uoguelph.ca/~hydrogeo/Whitebox/). The project was conceived as a replacement for the Terrain Analysis System (Lindsay, 2005), a freeware software package with an emphasis on analysis of digital elevation data. Whitebox GAT was intended to have a broader focus than its predecessor, positioning it as a desktop GIS and remote sensing software package for general applications of geospatial analysis and data visualization. The project also adopted the GNU General Public License (GPL) and the source code was published in a public repository hosted by Google Code (https://code.google.com/p/whitebox-geospatial-analysis-tools/). The entire source code can be retrieved from the repository using the version control system Subversion.

Since Whitebox GAT’s inception, three major versions of the software have been released. The 1.0 series was developed using Microsoft’s .NET framework and the Visual Basic and C# programming languages. It was therefore only compatible with Microsoft Windows operating systems, which was seen as an impediment for wider adoption. The 2.0 series of Whitebox GAT was a complete re-write that was developed using a combination of programming languages targeting the Java runtime environment (JRE) including Java, Groovy, and Python. By switching development to Java, Whitebox GAT became cross-platform, targeting all major operating systems including Microsoft Windows, Mac OSX, Linux, and all other operating system with a Java runtime. The 3.0 series of Whitebox added
enhanced vector data analysis support based on the Shapefile data format. The 3.0 release series also featured significantly improved cartographic capabilities. Recently, efforts have been made to internationalize Whitebox GAT by translating much of the text that appears on the user-interface into 11 languages. This task was accomplished through the contributions of volunteers drawn from the user community.

3. Capabilities and Design

The current version of Whitebox GAT contains over 360 plug-in tools for the analysis of geospatial data. These tools include a wide range of geospatial functions for performing typical analysis tasks including vector overlay, distance analysis (e.g. buffering and cost-distance analysis), terrain analysis, spatial hydrological processing (e.g. watershed and stream network extraction), raster algebra, geostatistical analysis (e.g. variogram model fitting and kriging interpolation), and many other common GIS operations. A LiDAR toolbox contains tools for working with and interpolating LAS files. For example, these tools can be used to create bare-Earth digital elevation models (DEMs) and canopy models from LiDAR datasets. There is also a Patch Shape Analysis toolbox that contains tools for characterizing the shape and distribution of raster and vector polygon features. Additionally, Whitebox GAT possesses significant image processing capabilities including tools for image filtering and enhancement, multi-spectral data analysis, classification, and change detection. From the outset, Whitebox has been designed to process large raster data sets, recognizing the ongoing trend towards the application of increasingly extensive data coverage and finer image resolutions. In a recent application, a 3.5 GB digital elevation model containing more than one billion grid cells was successfully processed with Whitebox GAT to model surface flow paths. Additionally, several tools in Whitebox GAT have been designed with algorithms that perform parallel processing to improve the performance of long-running, computationally intensive tasks.

Whitebox GAT has been developed with an extensible design that allows users to integrate custom plug-ins that add new functionality. Plug-in tools can be developed using the Java programming language or Whitebox’s built-in scripting capabilities. Supported scripting languages include Python, Groovy, and Javascript. All of Whitebox’s plug-ins can be called from scripts as a means of carrying out advanced geoprocessing workflows and task automation.

Whitebox tools are contained within a familiar toolbox ‘treeview’ structure located in the side pane (Figure 1), a design that allows for easy integration of new plug-in tools and toolboxes and the customization of functionality. Plug-in tools can also be accessed through a user search and a listing of recent and most-used plug-ins. This design for hosting tools permits more advanced functionality to be presented to the user in a consistent and easily accessed manner that allows for simplification of the toolbar and menu structures. To enhance the simplicity of the user interface design, the Whitebox GAT toolbar purposely contains a minimum number of icons, all of which are used for manipulating maps and data layers, vector digitizing, and standard or commonly accessed functionality (e.g. the raster calculator). All plugins have simple dialog-box-style user interfaces with a set of common components for specifying the necessary parameters for running a tool. Dialog boxes have a standard design (Figure 2) with a panel on the left for inputting parameters and a right panel that displays the help associated with the tool. A bottom pane contains buttons for running the tool, modifying and navigating the help documentation, and viewing the source of the tool. The View Code button is unique to Whitebox GAT and is central to the concept of open-
access software described below.

Figure 1. The Whitebox GAT user interface.

Figure 2. A typical tool dialog box including input parameters box, integrated help, and the ‘View Code’ button that is central to the concept of open-access software.

4. Whitebox GAT as Open-Access Software

Whitebox GAT derives its name from the concept of open-access software. Open access is
defined in the statement of the Budapest Open Access Initiative (Chan et al., 2002) as the publication of scholarly literature in a format that removes financial, legal, and technical access barriers to knowledge transfer. Although this original definition, and the subsequent Bethesda Statement on Open Access Publishing (Suber et al., 2003), focussed solely on the publication of research literature, we argue that the stated goals of reducing barriers associated with knowledge transfer can be equally applied to software. Open-access software can be viewed as a complimentary extension to the traditional open-source model of software development. All FOSS allow users the opportunity to download source-code, essentially giving users the ability to look inside the box. This is in contrast to proprietary software for which the user can only gain insight into the workings of a tool from the provided help documentation. The philosophy of the Whitebox GAT project is that the geospatial community as a whole benefits from the ability of users to examine the internal workings of specific algorithms or tools. Direct insight into the workings of algorithm design and implementation allows for educational opportunities, i.e. knowledge transfer, as well as the potential for innovation, improvements, and community-directed software development. Câmara and Fonseca (2007) recognized that adoption of open-source software is not only a choice of software, but also a means of acquiring knowledge. This is particularly important in the GIS field because many geospatial algorithms are highly complex and are impacted by implementation details. There are often multiple competing algorithms for accomplishing the same task and the choice of one method over another can greatly impact the outcome of a spatial analysis operation.

The concept of open-access software is based on the idea that software should be designed in a way that reduces the barriers that often discourage or disallow end-users from examining the algorithm design and implementation associated with specific geospatial tools. That is, open-access software encourages the educational opportunities gained by direct inspection of code. Câmara and Onsrud (2004) found that while some open-source GIS projects have large user-communities, most are developed by a relatively small number of individuals working closely in academic or commercial-sponsored settings. Thus, while many practitioners are taking up open-source GIS because they are free and often analytically powerful alternatives to proprietary geospatial software, it does not appear that all of the benefits of the open-source model are being realized in many cases. It is likely that this finding reflects a set of barriers that discourages user engagement and is inherent in the typical implementation of the open-source software model. An open-access software model, however, states that the reduction of these barriers should be a primary design goal that is taken into account at the inception of the project.

The main barriers that restrict the average user of an open-source GIS from engaging with the code include 1) the need to download source code from a project repository that is separate from the main software artefact (i.e. the executable), the common approach used by most open-source GIS projects, and 2) the required familiarity with the software structure. That is, an understanding of the organization of the source code is necessary to identify the code associated with a specific tool or algorithm of interest. Most desktop GIS projects consist of hundreds of thousands of lines of computer code that are contained within many hundreds of files. Large projects possess complex organizational structures that are only familiar to the core group of developers. The level of familiarity with a project’s organization that is needed to navigate to the code associated with a particular feature or tool presents a significant barrier to the casual end-user who may be interested in gaining a more in-depth understanding of how a specific feature operates. Additionally, project development is generally carried out by the core development team within a specialized software program.
called an integrated development environment (IDE). Again, a casual GIS end-user that may find themselves interested in how a particular tool works is less likely to install this additional IDE software, presenting yet another barrier between the user and the source code.

Whitebox GAT attempts to address these issues by allowing users to view the computer code associated with each plug-in directly from the tool’s dialog. Thus, just as a detailed description of a tool's working is provided in the help documentation, which appears within the tool's dialog, so to can the user choose to view the actual algorithm implementation simply by selecting the 'View Code' button on the dialog. This removes the need to download separate, and often large, project source code files and it eliminates the requisite familiarity with the project to identify the lines of code related to the operation of the tool of interest. Furthermore the tool’s code will appear within an embedded window that provides syntax highlighting to enhance the viewer’s ability to interpret the information. This model has the potential to encourage further community involvement and feedback. Among the group of users that are comfortable with GIS programming and development, the ability to readily view the code associated with a tool can allow rapid transfer of knowledge and best-practices for enhancing performance. This model also encourages more rapid development because new functionality can be added simply by modifying existing code. The 1.0 series of Whitebox, developed using the .NET framework, had the ability to automatically translate code written in one programming language into several other languages, thereby increasing the potential for knowledge transfer. Unfortunately this feature could not be replicated when the project migrated to the Java platform although there are on-going efforts to implement a similar feature.

5. Usage and Applications
A survey was carried out to gain a basic census of Whitebox GAT usage. The survey consisted of 795 users who downloaded the software 1138 times over a 13-week period starting in December 2013. The survey excluded downloads originating from the University of Guelph (the institution in which Whitebox is developed), of which there were 109. The analysis revealed that Whitebox is being used in at least 77 countries with the most widespread application in North America and parts of Europe (Figure 3). There is moderate uptake of the software within South America (particularly Brazil), Asia, and Oceania. One of the most striking findings however was the relatively few downloads within much of Africa and the Arabian Peninsula compared with the rest of the world. It is difficult to speculate on the cause for this observation based on the results of the survey alone. Nonetheless, it is possible that the sparse use of Whitebox GAT within these regions could reflect difficulties with Internet access or the lack of an edition of the software that has been translated in an appropriate language. There is currently no available Arabic version of Whitebox and this represents an area of interest for future development. In addition to the geographical data collected, the study also revealed that 82.4% of the downloads were from computers using Microsoft Windows as an operating system, 9.9% were based on Mac OS X, and 7.7% used Linux. It is interesting to note that compared with the general population of computer users, this distribution has a substantially higher proportion of Linux based users, which likely reflects the open-source nature of both Linux and Whitebox, as well as a slightly lower proportion of Microsoft Windows users.

Although Whitebox GAT is being used extensively in government organizations and the private sector (mainly consulting and resource management), it has found greatest application in academia both for education and research purposes. A review of academic literature
demonstrated that Whitebox GAT is being used in research applications involving the extraction of terrain parameters from DEM data (Cho et al., 2011; Lindsay and Seibert, 2013; Schwanghart and Scherler, 2014), soils mapping and erosion modelling (Gutiérrez et al., 2011; Omuto and Paron, 2011; Hales et al., 2012), wetlands research (Clare and Creed, 2013; Rampi et al., 2014), ecological studies (Poulos et al., 2012a, 2012b; Abdi, 2013), and for geoprocessing (Cao and Ames 2011, 2012). Whitebox GAT is frequently used in these studies to perform DEM-based analysis, which likely reflects the fact that this functionality is particularly well developed, owing to the projects origins in the Terrain Analysis System.

Figure 3. Downloads of Whitebox GAT over a 13-week period starting December 15, 2013.

6. Conclusion
This paper briefly introduced the open-source GIS Whitebox Geospatial Analysis Tools, highlighting some of its capabilities and design goals. One of the unique characteristics of this FOSS GIS is the ease with which users are able to interrogate the algorithms for individual geoprocessing tools. It does so by attempting to remove or lessen some of the barriers that are often imposed on typical users when they attempt to gain deeper understanding of how a specific tool operates. We argue that this innovative ‘open-access’ software development model may encourage greater knowledge transfer to typical end-users and lead to rapid innovation.

7. Acknowledgements
Funding for this research project has been provided by a grant of the Natural Sciences and Engineering Research Council of Canada.

8. References


Biography

*John Lindsay is an Associate Professor of Geography at the University of Guelph, Canada. His research areas include open-source GIS, algorithm design for spatial analysis, digital terrain analysis, and spatial hydrogeomorphology.*