Analysing commons to improve the design of volunteered geographic information repositories

Heinrich van den Berg[1], Serena Coetzee[2] and Antony Cooper[1,3]

[1]Department of Computer Science, University of Pretoria, Pretoria, 0002, South Africa
hvdberg@tuks.co.za, [2]Centre for Geoinformation Science, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, 0002, South Africa, serena.coetzee@up.ac.za
[3]Built Environment, CSIR, PO Box 395, Pretoria, 0001, South Africa
acooper@csir.co.za

Abstract

Commons are resources that belong to or affect an entire community. In a Web 2.0 environment, users interact and collaborate with each other to produce user generated content that belongs to the community, i.e. a commons. Web 2.0 implies that everyone with Internet access can join the user community. Volunteered geographic information (VGI) is a special case of user generated content. Web 2.0 technologies have enabled user generated commons, such as open source projects and Wikipedia; the former have been around since the late nineties, the latter was launched in 2001. VGI commons followed a few years later, with the advent of Google Maps, Wikimapia and OpenStreetMap. The communities that create the content, as well as the type of content, are quite different for open source projects, Web 2.0 encyclopaedias and VGI repositories; nevertheless, they share the fact that they are commons. We asked ourselves whether there is anything to be learnt for VGI repositories from the approaches, methodologies and technologies applied in open source projects and Web 2.0 encyclopaedias, which have been around for longer and have a larger user base. In this research paper we present the results of an analysis of three user generated commons: open source projects in general, Wikipedia and OpenStreetMap. The findings suggest that there are indeed approaches and technologies that could be useful for VGI repositories, and surprisingly, also technologies from the other commons that can be useful to Wikipedia.

1 Introduction

The Internet has given us various methods of communicating better and faster in ways that we hadn't imagined before. Web 2.0 technologies have made it even easier for anyone to contribute information on the web, resulting in internet users that are not just consumers of information, but also producers. The technologies give an ordinary consumer the ability to communicate and influence many people in a community. Their contributions are known as *user generated content*. This content can be anything from adding features in open source software projects to participating on collaborative social media-like websites such as Wikipedia (www.wikipedia.org), Facebook (www.facebook.com), Flickr (www.flickr.com), YouTube (www.youtube.com) and OpenStreetMap (www.openstreetmap.org), to name only a few. User generated content in the spatial domain is referred to as *volunteered geographic information (VGI)* (Goodchild, 2007). Collaborative websites

are online communities where the information is no longer just provided by a small group of professionals, but by a whole community with common interests. The content that the users contribute in these communities forms a shared resource belonging to, or affecting, the whole of a community, and this is also known as a *commons* (Oxford Dictionary, 2010).

In these online virtual communities the data is shared with the public and in some cases also with other communities. The other communities can in turn provide services that would otherwise not be possible, such as a mashup of open source services. Community membership grows every day as the communities are discovered for their potential to use social networking to improve and distribute information (Albors et al., 2008).

The work presented in this paper is part of a research project at the University of Pretoria that aims to investigate how VGI contributions can be used to improve the institutional address datasets of local authorities. This would lead to improved planning, development and maintenance of services in their areas of jurisdiction, such as the supply of sanitation and safe drinking water (Coetzee & Cooper, 2007). As part of this project, we asked ourselves whether any of the approaches, methodologies and technologies applied in open source projects and Web 2.0 encyclopaedias can be transferred to VGI repositories. Although the term 'open source' was introduced in the late nineties, open source projects have been around since the 1950s. Wikipedia was launched in 2001, while VGI commons emerged a few years later: OpenStreetMap was inspired by Wikipedia and launched in 2004. The shared resource in each of these three examples is software code, facts and geographic information, respectively. The virtual communities that create these shared resources are quite different, but for all the shared resource is created and maintained by voluntary users.

A *spatial data infrastructure* (SDI) is an evolving concept about facilitating and coordinating the exchange and sharing of spatial data and services between various stakeholders (Hjelmager et al., 2008). Traditionally, SDI data is provided by a limited number of mandated authorities, such as mapping agencies or municipalities, who employ professionals to prepare the spatial data and can therefore be trusted to supply data of sufficient accuracy and quality. However, in a Web 2.0 world where citizens (thus amateurs) become the sources of data, this assumption does not hold anymore, User-generated content in general, and VGI in particular, is becoming more important as a source for official data bases, such as those used in national SDIs (Cooper et al., 2010).

While remotely sensed data are generally available for developing regions such as Africa, key location information, such as addresses or house numbers, village and street names and points of interest, is often not available. Geographic information volunteered by citizens, who live out their daily lives in these streets and villages with no access to the Internet but with mobile phones at their disposal, can have a huge impact on the accuracy and availability of up-to-date spatial data.

In this paper we present the results of our case studies of the different ways in which open source projects, Wikipedia and OpenStreetMap handle user generated content. The approaches, methodologies and technologies applied in the three commons were compared and analysed in a qualitative manner. The following criteria were evaluated:

- The *parties* that are involved in the contribution process, the expertise needed to be able to contribute and the restrictions in place to prevent someone from contributing.
- The *process* of contribution, such as the procedure or steps followed.
- The *kind of contribution* made, the information stored that describes the contribution, such as any metadata that might be available, and the type of information contributed.
- The *tools* that are needed to facilitate the contribution process.
- The *reason* (motivation) for contributing the information.

The remainder of the paper is structured as follows: section 2 provides background information about the three commons; section 3 presents the results of the case studies; section 4 discusses these results; and section 5 concludes with a look into the future.

2 Background

2.1 Open source projects

In an open source project that produces *open source software*, the source code is made available to the public using a software license that either meets the *Open Source Definition* or that is in the public domain and allows anyone to change and redistribute the software. The Open Source Initiative (OSI) is a non-profit corporation that maintains the Open Source Definition, which creates a certain level of trust and understanding for organizing open source cooperation between the different parties involved. It is an important part of open source software projects and it specifies the distribution criteria required of open source software (The Open Source Definition, 2010).

Users contributing to open source software projects generally have to have some knowledge of the programming languages used in the projects to the extent whereby they can at least add features or fix bugs. It enables more flexible software to be developed at a lower cost and makes use of distributed peer review (Johnson, 2006). The contributing members of open source software projects are often developers that are also users of the software, because they have a great deal of experience in the domain where the software is used (Mockus et al., 2002). This indicates the people that are most likely to contribute data are the ones that are relatively the closest to the information and respond to a need of the specific community (Scacchi, 2007).

The typical path that a new member would follow in an open source project makes use of a centralized revision control system:

- They start as a passive member who mostly reads forums, message boards and posts on mailing lists.
- They then report bugs and suggest possible enhancements to the software.
- The community member writes or modifies source code, such as providing a patch for a reported bug. It usually takes a while to get to this step, because it requires some technical knowledge in understanding the system and the community.
- After making several quality contributions of source code, the member could get access to the source code repository and permission to commit code directly to it.

- If the member wants to become even more involved, the next step would be to take control of a module or section of the project and then move on to the project as a whole, to gather support and provide direction.

2.2 Wikipedia

Wikipedia is a collaboratively edited encyclopaedia available online, where readers can contribute to it. It has proven to be a very successful peer collaboration tool in that the content of the Wikipedia encyclopaedia is of comparable quality to traditional encyclopaedias, for example the Encyclopaedia Britannica (Giles, 2005). The articles can be edited by anyone and the changes can be seen immediately. There are, however, on-going reviews of articles by a community of editors, and each user can have a watch list of articles, which trigger alerts when the articles are edited. There are also version histories on all articles, to be able to revert back to a prior state of an article in the case of vandalism or inaccuracies. When such an event is detected, the content is usually reverted within minutes (Kittur & Kraut, 2008).

The process a new member goes through from novice or *non-Wikipedian* to expert or *Wikipedian* usually involves a change in their perspective of Wikipedia and inherently also of the goal of contributing to the community. The member starts to adopt a caretaker role and use watch lists to track certain articles. Over time, that watch list will grow and they will identify with the community as a whole and adopt its goals community to build a sound information resource. They are then not only focused on the quality of certain articles but also with improving the community. They start using more of the tools, which become more relevant to their goals, such as creating an account for contributing to the discussion pages, using the editing history of an article and tracking one's own contribution while maintaining a consistent identity (Bryant et al., 2010).

The majority of new members that contribute to the community are either one-time editors or casual users and the biggest challenge is to get these users to return and be more involved in the community, and on the other hand, to recruit more frequent contributors (Panceira et al., 2010).

2.3 OpenStreetMap

There are many communities where VGI is contributed. OpenStreetMap is one where VGI is used to provide free global geographic data, such as rivers, railways, streets, etc. OpenStreetMap allows anyone to contribute and all geographic information collected is uploaded to the project's main map. New geospatial technologies available on the web enable ordinary people to create digital spatial data and have the potential to represent aspects of everyday life (Elwood, 2008).

The community is based on the principle of giving everyone access to relevant tools so that they are able to converge on an accurate representation of the Earth, because the locals are closer to the spatial data than the experts and they benefit from an accurate and up-to-date representation of their relative locations (Crampton, 2009). A common implementation of providing objective information in the community might influence the majority of users to contribute their positive credibility assessments in the same community (Flanagin & Metzger, 2008).

There are various client applications available on the internet providing digital geographical information in the form of a virtual globe, such as Google Earth (earth.google.com). These applications make it easy for anyone to access the data and contribute data quickly and easily. The quality and reliability of VGI is a key concern and there is usually a lack of metadata available for this type of data (Cooper et al., 2010).

3 Results

In this section we present the analysis and comparison of the contributions to the three commons described in Section 2: open source projects, Wikipedia and OpenStreetMap. These commons were selected because they represent different kinds of virtual communities that contribute different types of user generated content: open source software programs, information or facts, and geographic information. We contributed information to these commons ourselves in order to understand how contributions are made and what processes are followed. Udig (udig.refractions.net) is an open source project where we contributed a feature implementation in the form of a patch; in Wikipedia an edit was made to correct an error in the article covering the University of Pretoria; and in OpenStreetmap the online editor, Potlatch, was used to add the Xanadu Eco Park point of interest, located in Hartbeespoort, to the map. Based on our experience, the commons are compared according to the criteria listed in the Introduction, summarized in the tables below.

3.1 The parties involved

Table 1 shows the people involved in contributions to the three different communities, which have the same scheme of contributors and reviewers. New members of open source projects are the most restricted as their contributions have to be reviewed before incorporated. This reduces the chances for vandalism considerably but increases the effort needed to make a contribution. There are not always enough people available to review the contributions and as a result, many contributions may never be used.

Table 1. Comparing people involved in contributions

	Open source projects	**Wikipedia**	**OpenStreetMap**
Contributor	Person registered with an account contributes a file indicating changes in the source code.	Any person contributes information or facts to the website that are immediately available to everyone.	Person registered with an account contributes geographical information that is immediately available to everyone.
Reviewer(s)	One or more people review the changes and decide whether they will be incorporated.	The information or facts are continuously reviewed by other people using the website.	The geographic information is continuously reviewed by other people using the website.

Wikipedia, on the other hand, follows the unrestrictive approach where new members do not have to be registered to contribute or edit most of the pages. The changes are reviewed only after they have been submitted and when they are already viewable to the masses. This allows vandals to

easily make unconstructive contributions and increases the amount of effort needed to maintain the quality of the information.

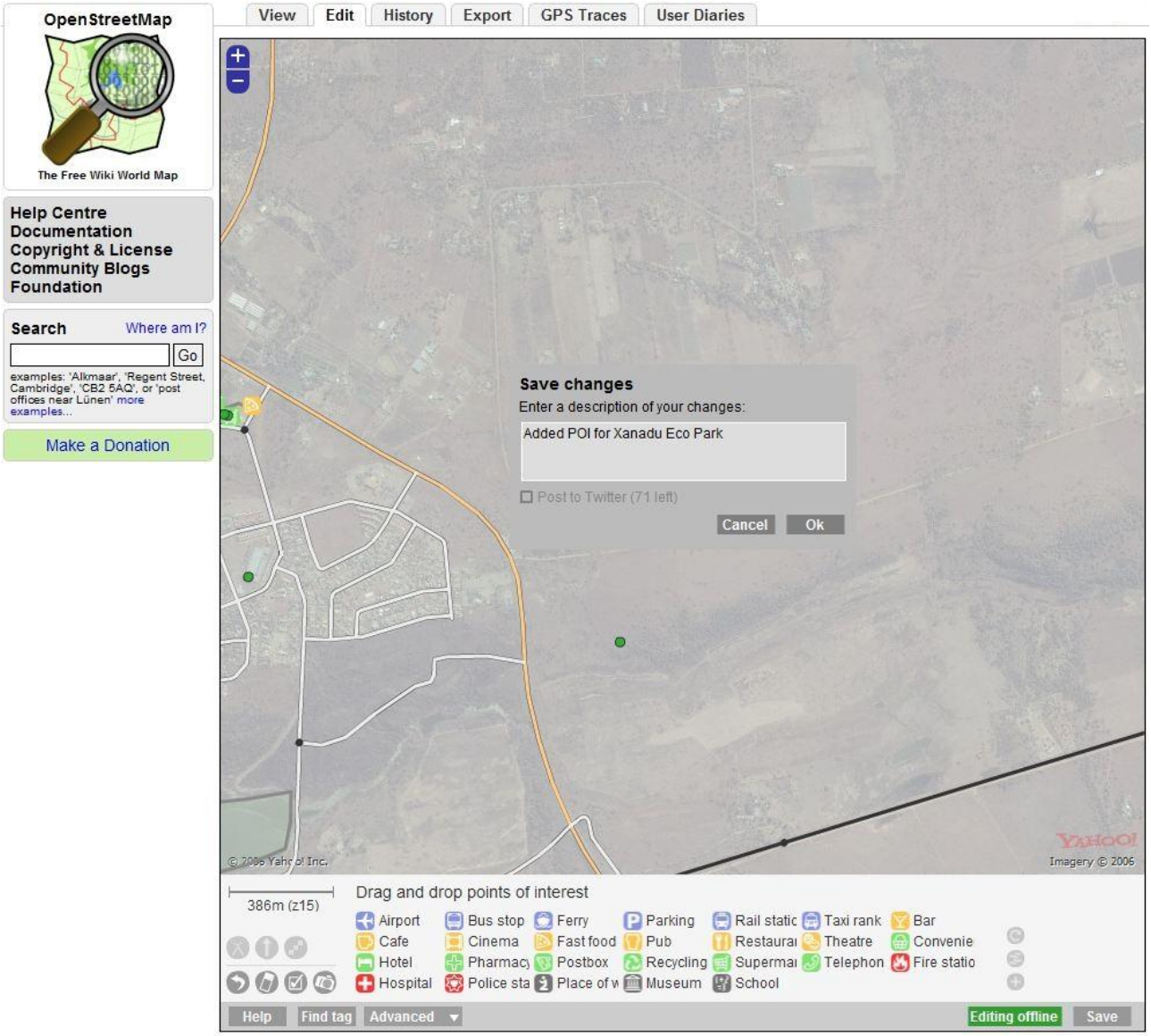

Figure 1. Screenshot of OpenStreetMap after adding a Point of Interest with Potlatch

OpenStreetMap allows only registered users to contribute geographic information. The changes are also immediately available to the masses, but vandals are easier to identify because OpenStreetMap users are identified by their usernames, which can be blocked. In Wikipedia, users are identified by username or IP-address and more than one user might use the same IP-address. When a vandal gets detected everyone using that same IP-address is blocked, not just the culprit.

3.2 The process of contribution

Contributing to a community generally starts with finding out more about the community and getting to know its structure. This helps to understand which contributions are regarded to be of high quality and great value. The process of contributing depends on the content that is contributed because the content needs to be prepared correctly before submission. For example, contributors to Wikipedia do not need specialized tools or processes to prepare the data in the correct format.

Therefore, the process of contributing to Wikipedia is less complicated than for open source projects and OpenStreetMap. In Wikipedia, there is also no facility for tracking bugs or prioritizing them, which is available in both the other communities.

Table 2. Comparing the processes followed to make contributions

	Open source projects	**Wikipedia**	**OpenStreetMap**
Initial communication	Get involved in the community by viewing communications and communicating with members.	Get involved in the community by viewing communications and communicating with members.	Get involved in the community by viewing communications and communicating with members.
Active involvement	Report bugs or suggest features.	Contribute to the discussion on an entry.	Errors in the map data are reported in OpenStreetBugs.
Preparation for a contribution	Create a local copy of the source code available in the central repository. Fix a bug or implement a feature that is listed on the tracker in a local development environment.	Prepare the text for the contribution. No tools or specific skills are needed, e.g. hand written notes could be used.	GPS data or other geographic data like spatial images are collected.
Actual contribution	Upload the changes as a patch file so that it becomes available to the community.	Contribute by editing a page or an article using the edit tab on the website.	The data is uploaded, added to maps, and if necessary, the data on the maps is edited.
Review of the contribution	The changes in the patch are reviewed and committed to the source code.	Changes are reviewed by everyone that reads the page and it is reverted when it is not verifiable.	The maps are rendered and used by the community. All users act as reviewers of the changes.

In open source projects, contributions in the form of patches or changes to the source code are reviewed before being committed to the repository. This decreases the chances of errors or vandalism. In Wikipedia and OpenStreetMap, the probability of introducing errors and vandalism is bigger because contributions are reviewed only after they have been submitted. In June 2010, Wikipedia conducted a two-month trial of Pending Changes (PC), a tool that requires changes in protected and semi-protected articles to be reviewed before they are published. The tool aims to reduce the amount of unconstructive contributions in these articles. After the two-month trial, there was no clear consensus on how to proceed with its implementation (Wikipedia, 2011).

In OpenStreetMap, contributions can be made using Potlatch, an online editor that is accessed from the edit tab on the OpenStreetMap website. It is the preferred editor for beginners and one has the choice of uploading changes as they are made (‘live’) or only saving and uploading changes at the end of an edit.

The expertise needed for a user to make a contribution has a significant influence on the process followed to make the contribution. A user has to acquire the necessary skills and prepare the content in the required format before adding it to the commons. In an open source project, the user has to design, write and compile source code; in Wikipedia, the user has to be able to create Wikitext and

write coherently; and in OpenStreetMap, the user has to be able to use Potlatch and other tools and understand how spatial data is mapped. Of these, writing code probably requires the most skills.

3.3 The kind of contribution made

The contributions made to the communities differ in the content that is contributed and the size of the contributions. These can be anything from fixing a small bug or error to adding an entire software module to an open source project, a third party database to Wikipedia or an entire geographic dataset to OpenStreetMap.

Table 3. Comparing the kinds of contributions made

	Open source projects	**Wikipedia**	**OpenStreetMap**
Content	Contributions consist of changes in the source code of the program or documentation, that can be tested.	Contributions consist of factual information, that can be verified.	Contributions consist of factual geographical information, that can be verified.
Metadata	A description of the contribution is added with the contribution.	An edit summary is used to describe the contribution made.	An edit summary is used to describe the contribution made.
User metadata	The contributor is linked to the contribution by registered name.	The contributor is linked to the contribution by registered name or IP address.	The contributor is linked to the contribution by registered name.
Bug and error reports	Users can report bugs in the program.	Users can provide descriptive information or metadata, which is a source of information that can be incorporated into article at a later stage.	Users can report bugs or errors in the data.
Suggestions	Users can suggest new features to be added to the program.		

In open source projects and OpenStreetMap, the bug tracking systems that are implemented provide a way for people to inform the community of bugs or errors without the need for understanding the details to edit the data.

3.4 The tools needed

The tools used by the communities are related to the type of content that is contributed to a community. In many open source projects, the process of building or compiling the source code into binary code is automated using a script with a set of instructions to describe what needs to be done.

In Wikipedia, editing articles requires one to use Wikitext for formatting the text. This can be overwhelming to someone who without experience in using special characters to format text, and might even discourage them from contributing. Providing an alternative way of contributing to Wikipedia can help make the process more transparent, allowing more people to contribute.

OpenStreetMap uses software programs to perform automated edits on the data and a wide variety of other tasks, partly because of the complexity of geographical data. The automated

programs used in open source projects and Wikipedia might be useful in creating automated programs for OpenStreetMap.

Table 4. Comparing the tools needed to make contributions

	Open source projects	Wikipedia	OpenStreetMap
Connectivity	Internet access is needed to access the source code repositories.	Internet access is needed to access the Wikipedia site.	Internet access is needed to access the OpenStreetMap site.
Access tools	A client program to access the source code in the central repository, a standard text editor to view the source code, and a standard web browser to access the bug tracker.	A standard web browser to access the website.	A standard web browser to access the website.
Preparation tools	A client program to check out the source code from the central repository, a text editor in which the code changes are made and a compiler.	Wikitext is used to add content to pages with the proper formatting.	A GPS device can be used to collect geographic data, but the user can also point-and-click to add features or attributes in one of the editors supplied by OpenStreetMap.
Automated tools	Build and compile source code.	Detect and correct some errors or send a notification.	Automated edits on data, send notification and a variety of other tasks.

WikiScanner used in Wikipedia provides additional information about anonymous edits made by linking the edits to their originating organizations. A similar tool might be useful in OpenStreetMap where edits can be linked to the general location of the user who made the changes. This information is mostly used in disputes.

Wikipedia users use watchlists to track edits made in selected articles, making the process of reviewing articles much easier. OpenStreetMap users can make use of the service OSM Mapper (www.itoworld.com), which keeps track of edits in a certain geographical area and alerts the user if changes were made. OSM Mapper is then comparable to Wikipedia's watchlist.

3.5 The reason for contributing

Table 5. Comparing the reasons for contributing.

	Open source projects	Wikipedia	OpenStreetMap
Continuous user	Uses the software and would like to improve it.	Uses Wikipedia as a source of information and would like to improve it. Promotes their interests by making information available.	Uses OpenStreetMap data and would like to improve it. Promotes their area by making data available.
Once-off user	Is able to fulfil a request made by another user.	Has the expertise to identify and correct an error in an article.	Has extensive knowledge about a certain area and would like to add or correct the information.

Overall, there are two reasons for contributing to these commons: firstly, because one uses the product or service and would like to improve it. Effort should be focused on getting such people involved so they can continue to contribute, resulting in a sustainable community in the long run.

Secondly, one contributes because one has the knowledge or expertise to make a valuable contribution, though usually making only a few contributions and then not doing it again. The best way to get more people contributing is to make the process as seamless and simple as possible.

4 Discussion of the results

All the communities that were analysed use peer-review to improve their quality, but the reviewing process followed differs between the communities. The system for tracking and prioritizing bugs works well in both open source projects and OpenStreetMap and could improve the detection of errors in Wikipedia. The ability to prioritize errors in articles would encourage people to contribute to specific Wikipedia articles and encourage more people to contribute (by also correcting errors). Further, Wikipedia users could benefit from a watchlist that monitors not only selected articles, but also related articles.

In open source projects, vandalism is generally not a problem because contributions are reviewed before they are accepted. This creates a lot of overhead and contributions might never be incorporated because there are not enough people to review and implement them. Wikipedia has launched a trial for reviewing contributions to selected articles before they are published or viewable, but a decision on how to go forward after the trial has not yet been reached. OpenStreetMap could benefit from implementing a similar process of reviewing selected features only. For example, changes to suburb boundaries could be reviewed, while changes to points of interest are not. In this way the changes would be better regulated, without too much overhead.

The automated tools developed for Wikipedia have made it easier and faster to detect and revert certain kinds of unconstructive edits. For VGI repositories such as OpenStreetMap, similar tools could be developed to detect and notify users or automatically revert unconstructive changes, for example changes that introduce logical inconsistency errors.

Further research can be done on these three commons relating to vandalism and how it is handled. For example, a tool could be developed for OpenStreetMap that is similar to Huggle for Wikipedia, which automatically reverts incorrect changes or sends notifications about possible vandalism attempts. This research was limited to only three types of commons and there are others that could provide additional insight if added to the comparison. A more thorough investigation of equivalent proprietary repositories might reveal approaches, methodologies and technologies that prove useful to the commons. In addition, the use of standards in commons could be investigated to propose open standards for further streamlining contributions and enabling use of the commons.

5 Conclusion

In this paper we presented the results of our case studies of the different ways in which open source projects, Wikipedia and OpenStreetMap handle user generated content. The approaches, methodologies and technologies applied in the three commons were compared and analysed in a

qualitative manner, comparing e.g. the parties involved, the process, the kind of contribution and the tools that are available.

We found that similar approaches are followed in the various commons, varying to allow for the different kinds of contributions. We wanted to establish if there are any approaches, methodologies and technologies applied in these commons that can be transferred to VGI repositories. We found that firstly, OpenStreetMap does not yet distinguish between restricted (feature changes needing review before acceptance) and non-restricted features. Secondly, OpenStreetMap could benefit from automated tools to detect and possibly revert incorrect contributions. For our project on VGI contributions and address datasets, restricted features are required to ensure that fraudulent addresses are not added, while automated tools are required for similar reasons, as well as to limit vandalism. This implies that OpenStreetMap cannot be used for this project, or at least that add-on tools will have to be developed.

Acknowledgements

This research was supported in part by a grant from the Research Development Programme of the University of Pretoria.

References

Albors, J., Ramos, J.C. & Hervas, J.L., 2008. New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source. *International Journal of Information Management*, 28, pp.194-202.

Bryant, S.L., Forte, A. & Bruckman, A., 2010. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work.*, pp.1-10.

Coetzee, S. & Cooper, A.K., 2007. What is an address in South Africa? *South African Journal of Science (SAJS)*, 103(11/12), pp.449-58.

Cooper, A.K., Coetzee, S. & Kourie, D.G., 2010. Perceptions of virtual globes, Volunteered Geographical Information and Spatial Data Infrastructures. *Geomatica.*, 64(1), pp.73-88.

Crampton, J.W., 2009. Cartography: maps 2.0. *Progress in Human Geography*, 33(1), pp.91-100.

Elwood, S., 2008. Geographic Information Science: new geovisualization technologies – emerging questions and linkages with GIScience research. *Progress in Human Geography*, pp.1-8.

Flanagin, A.J. & Metzger, M.J., 2008. The credibility of volunteered geographic information. *GeoJournal*, 72, pp.137-48.

Giles, J., 2005. Internet encyclopaedias go head to head. *Nature*, 438, pp.900-01.

Goodchild, M.F., 2007. Citizens as censors: The world of volunteered geography. *GeoJournal*, 69(4), pp.211-21.

Hjelmager, J. et al., 2008. An initial formal model for spatial data infrastructures. *International Journal of Geographical Information Science*, 22(11), pp.1295-309.

Johnson, J.P., 2006. Collaboration, peer review and open source software. *Information Economics and Policy*, 18, pp.477-97.

Kittur, A. & Kraut, R.E., 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. *Proceedings of the 2008 ACM conference on Computer supported cooperative work,ACM*, pp.37-46.

Mockus, A., Fielding, R.T. & Herbsleb, J.D., 2002. Two Case Studies of Open Source Software Development: Apache and Mozilla. *Transactions on Software Engineering and Methodology (TOSEM),ACM*, 11(3), pp.309-46.

Oxford Dictionary, 2010. *Oxford Dictionaries*. Oxford University Press.

Panceira, K., Halfaker, A. & Terveen, L., 2010. Wikipedians are born, not made: a study of power editors on Wikipedia. *Proceedings of the ACM 2009 international conference on Supporting group work.*, pp.51-60.

Scacchi, W., 2007. Free/Open Source Software Development: Recent Research Results and Emerging Opportunities. *Proceedings of the the 6th joint meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on The Foundations of Software Engineering,ACM*, pp.459-68.

The Open Source Definition, 2010. *The Open Source Definition*. [Online] Available at http://www.opensource.org/docs/osd viewed October 2010.

Wikipedia, 2011. *Pending Changes*. [Online] Available at: http://en.wikipedia.org/wiki/Wikipedia:Pending_changes viewed January 2011.