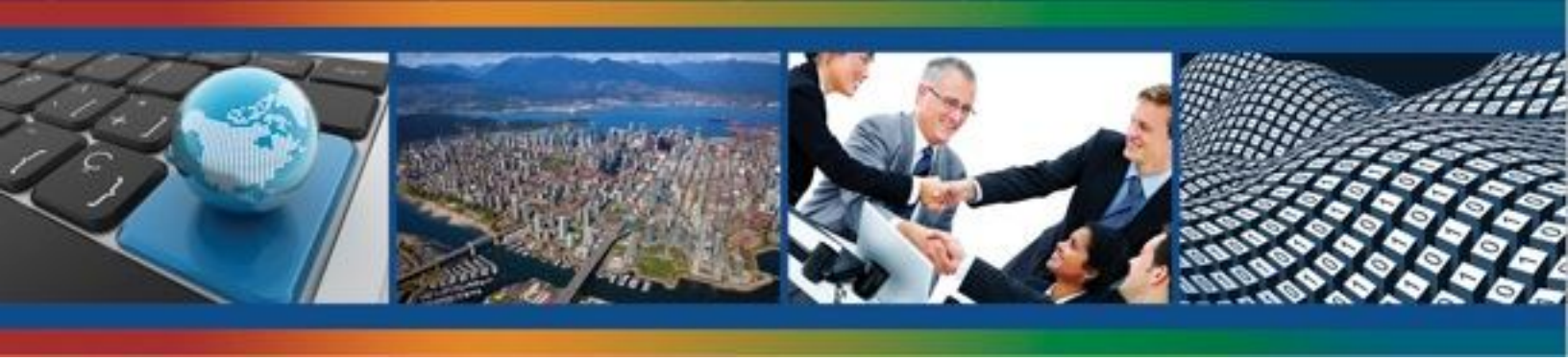


CANADIAN GEOSPATIAL DATA INFRASTRUCTURE
INFORMATION PRODUCT 21e

Volunteered Geographic Information (VGI) Primer

Hickling Arthurs Low
Science & Technology Policy Research and Analysis Resource team

2012

Natural Resources Canada | Ressources naturelles Canada

Canada

Table of Contents

1. Preamble

This guide is one in a series of Operational Policy documents being developed by GeoConnections. This guide is intended to inform CGDI stakeholders about the nature and scope of VGI and the realities, challenges and good practices of related operational policies.

Volunteered geographic information, or VGI, is "the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information" (Goodchild M. F., 2007). Using volunteered geographic information to help create or maintain geospatial datasets is a rapidly growing trend. This guide introduces key issues in geospatial operational policy, imperative to the success of any venture into VGI. Operational policies are the guidelines, directives and policies that an organization employs to address the life cycle of geospatial data (i.e., collection, management, dissemination and use).

This guide will be of interest to anyone seeking a better understanding of the emerging trend of VGI and areas of related operational policy, such as data quality, liability, privacy, security, licensing and copyright.

The **GeoConnections** *program is a national initiative led by Natural Resources Canada. GeoConnections supports the integration and use of the* **Canadian Geospatial Data Infrastructure (CGDI).**

The **CGDI** *is an on-line resource that improves the sharing, access and use of Canadian geospatial information – information tied to geographic locations in Canada. It helps decision makers from all levels of government, the private sector, non–government organizations and academia make better decisions on social, economic and environmental priorities.*

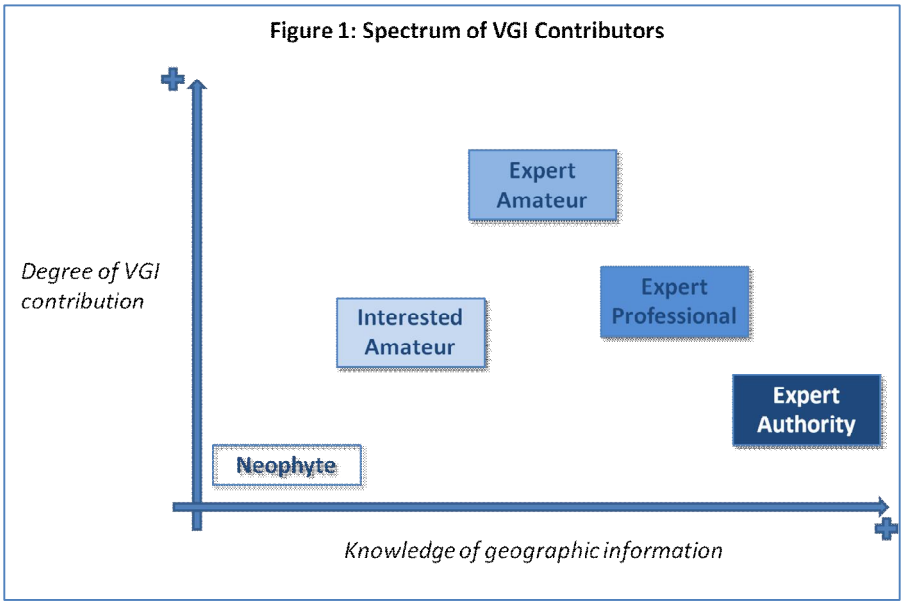
2. An Introduction to VGI and its Use

VGI is part of a broader trend of "user-generated content", or UGC (IAB, 2008). UGC has become commonplace in Web 2.0-based applications, but is based on a long history of user contribution systems in the marketplace. There has also been a history of interested individuals offering geographic input and feedback to authoritative geospatial data producers and communities of interest (e.g., environmentalists, land use planners, etc.) but VGI involves the community playing a much more organized and influential role. The term coined for such contributors, "produsers" (Produsage.org, 2007; Coleman et al, 2009), signifies that VGI participants are typically users of geographic information content on the Web, who want to improve that content by submitting notices of changes to and errors in the data.

VGI caught the attention of major companies with geographic information assets (e.g., Google, TomTom, Navteq and TeleAtlas) early on. These firms quickly leveraged growing interest and citizen

participation in VGI, combined with the availability of wireless technology and networks and growth of online applications, into a lower-cost means of updating mapping products. This trend has spurred the creation and growth of new sources of geographic information on the Web that are based solely on VGI contributions, such as OpenStreetMap, Wikimapia, and Ushahidi. Organizations that employ VGI provide simple tools for contributors to submit geographic information (e.g., Google Map Maker and TomTom Map Share). Contributions can be in a variety of forms, including GPS-derived coordinates of new or changed features, geotagged media such as digital images and videos, or digital map files.

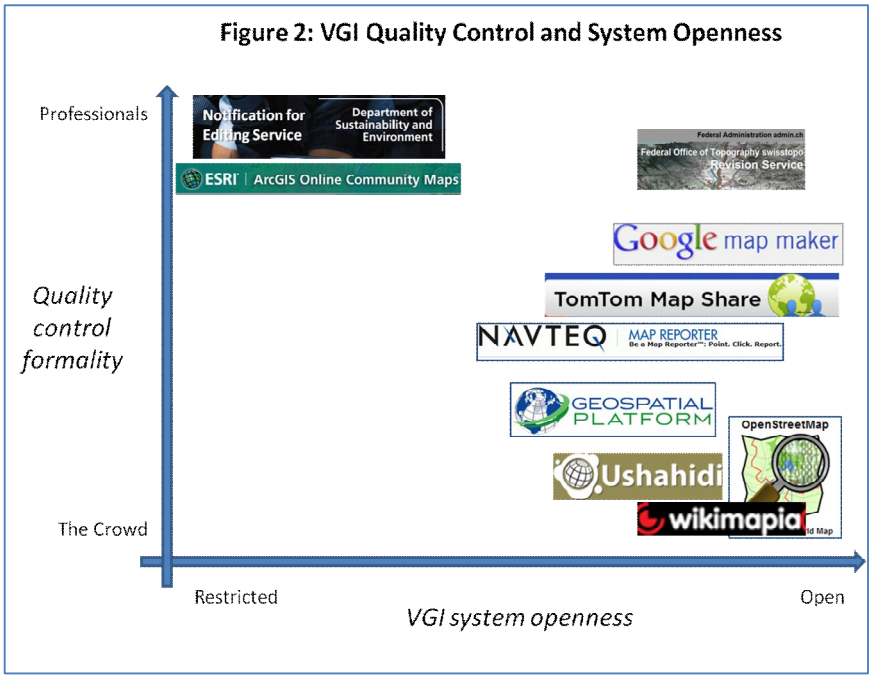
Figure 1: Spectrum of VGI Contributors



The range of VGI contributors is fairly broad. The overlapping categories into which they can be subdivided, based on relative knowledge of geographic information, can be described as; Neophyte, Interested Amateur, Expert Amateur, Expert Professional, and Expert Authority (Coleman, Georgiadou, & and Labonte, 2009). As illustrated in Figure 1, the patterns of contribution of VGI content to date have not necessarily been in relative alignment with levels of knowledge of the subject.

The range of contribution models for VGI input to Web-based geospatial databases is also broad. Figure 2 illustrates this spectrum, from systems that are restricted to authorized contributors who are typically in the expert professional or authority categories, to open systems that will accept content from anyone, with variations of these models in between. It also shows that the formality of quality control methods covers a spectrum, from iterative refinements made by individual citizen contributors (the "crowd") to quality assessment by trained geographic professionals. The approximate relative positions on these two axes are shown for a representative sample of VGI system operators (i.e., Victoria State, Australia Department of Sustainability and Environment,

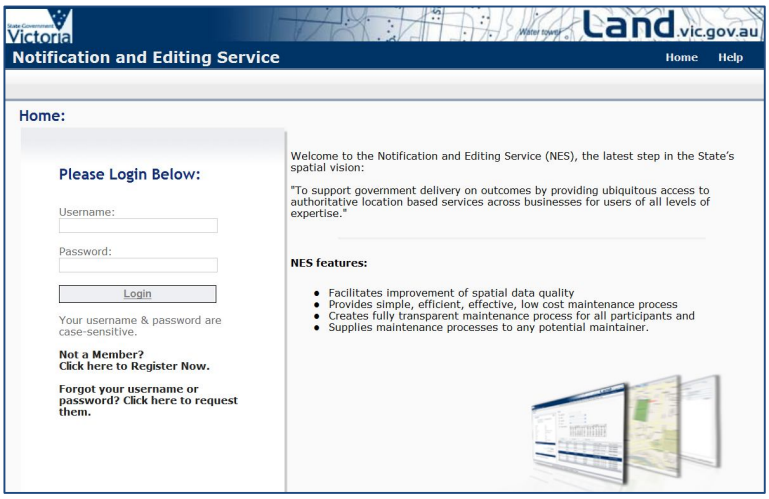
Figure 2: VGI Quality Control and System Openness



swisstopo, Esri, Google, TomTom, Navteq, Geospatial Platform, Ushahidi, OpenStreetMap and Wikimapia).

2.1 VGI in Use

To inform the preparation of this primer, the use of VGI was examined in detail through case studies of three organizations: the Department of Sustainability and Environment, Victoria State, Australia; OpenStreetMap; and Esri Canada.



The state government of Victoria, Australia launched the Notification and Editing Service (NES), an Internet-based application for distributed data maintenance, in 2007. The system was designed to allow local governments and interested organizations from across Victoria direct change request access to government data administrators of spatial information held in DSE's Victorian Spatial Data Library. The VGI contributors in this case are primarily mapping professionals or other significant geospatial data users in those organizations, and not the general public. In December 2011, there were 500 registered users of NES. The VGI model was adopted to meet the growing demand for higher quality (particularly more current) geospatial data within existing resource constraints. Change request contributors submit notices of new features or changes to existing features, which are then channelled to the organizations responsible for those features (i.e., data custodians). The custodians advise NES operations that the changes are approved or rejected and, if approved, NES edits the database. Of the approximately 100 feature types in their database, they had 85 under custodianship arrangements in December 2011.

OpenStreetMap (OSM) is perhaps the most globally recognized example of collecting and leveraging VGI. The project, initiated in 2004, is a worldwide mapping effort that allows any interested party to view and edit geographic data in a collaborative way. OSM was started by an entrepreneur in United Kingdom who was frustrated with the strict copyright laws that apply to the Ordnance Survey maps and geodata. The OSM Project sees daily feature additions, modifications of inaccurate features, and deletions of stale or invalid data from its contributor base of over 530,000 registered users. While OSM started with merely GPS

tracks to record roadways, it has expanded to include anything contributors desire to have mapped. In addition to individual contributions, organizations have also donated complete data sets to OSM, some of which have been incorporated wholly into their data base. OSM does not use quality control experts to vet contributions. The quality of their data is refined over time through iterative corrections of submitted data by subsequent contributors.

Launched in June 2010, the Esri Canada Community Maps Program is a response by the company to user demands for a multi-scale, cartographically uniform web basemap of Canada. The program aims to facilitate the distribution of geographic information from governments and other interested organizations to GIS professionals, application developers, and the general public. Contributions are presently restricted to geospatial data providers, but in the longer term, they want to allow the public to identify data errors and new features, and transfer those notifications back to the authoritative data sources for action. As of December 2011, Natural Resources Canada and some 20 municipalities were participating in Community Maps and approximately 80 other organizations were considering becoming involved.

This brief introduction to VGI and examples of its use demonstrates the potential benefits of this new model of data creation and improvement to geospatial data providers. Harnessing this potential involves the examination and resolution of a number of operational policy challenges, which are addressed in the next chapter.

3. Operational Policies and VGI Implementation

3.1 Data Quality and Authoritativeness

VGI offers enormous potential for improving the quality of geospatial data holdings. While definitions of data quality vary, a common definition is “fitness for use”, or how well the data satisfies particular needs or fulfills certain requirements to solve a problem. How well data fits a particular use is often assessed in terms of attributes such as completeness, currency, validity, consistency, timeliness and accuracy (BC Ministry of Forests and Range, 2006). However, quality is relative, and describing data as being of high

or low quality is meaningless without measuring it against a production specification or a user requirement (Coote & Rackham, 2008).

By far the greatest concern for organizations considering the use of volunteered geographic information as a means of creating or updating geospatial datasets is data quality. Typically, concerns include four quality aspects: positional and attribute accuracy; completeness of data; currency of data; and credibility of data sources (Coleman et al, 2010). In some organizations, these quality assessments are handled by experts, while in an increasing number of others, the community of users or social network are assuming that role.

ISSUES TO ADDRESS IN QUALITY BENCHMARKING (Coleman et al, 2009)

- **How to assess the credibility of a contributor**
- **How to assess the accuracy of VGI contributions (e.g., in-house quality assurance, a moderated on-line community, or the public)**
- **The best and quickest means of delivering credible input**
- **The control over content and quality given to contributors**
- **Decision-making on acceptability of updates**

Positional accuracy is highly dependent upon the technology employed by the contributors and their proficiency with its use. **Attribute accuracy** in VGI contributions is dependent upon such factors as how interested contributors are in completeness of their contributions, how detail oriented they are, etc. Experience has shown that there is significant variability in contribution accuracy, resulting in sometimes unpredictable data quality from region to region. **Data completeness** is also a concern, and this is often correlated with geographic location (e.g., densely populated urban areas have more complete VGI coverage than rural areas). Data can be incomplete in terms of gaps in the data (e.g., missing land use polygons), missing feature segments (e.g., sections of a street, or missed closures at intersections) and missing attributes (e.g., no street names) (Maué & Schade, 2008). While an organization's quality standards may include a number of aspects, the rapidity with which new features are added through VGI underscores the importance of data **currency** for most users. This fact illustrates the reality that "quality" is a subjective term that producers and users often view differently (Coote & Rackman, 2008).

One of the most important predictors of contribution quality is the **credibility of the VGI source**. Unlike the professional data production model, VGI presents the challenge of assessing the credibility of contributions from geospatial non-experts. As is the case with Wikipedia, factors that may be important indicators of a geospatial data contributor's credibility include: i) a contributor's frequency, type and degree of edit operations; ii) operations quality and veracity (in terms of whether or not a given edit is subsequently changed by someone else); and iii) reputation for reliability (in terms of past contributions and edits influencing the lifespan of subsequent contributions) (Goodchild, 2009). TomTom has recognized the viability of this approach, and uses a graduated means of sharing, assessing, and using the volunteer-provided updates. It allows contributors to decide whether to only use their updates themselves, within their own group, or share them with the general TomTom community. TomTom assigns a level of credibility based on the sources of independent confirmation of a given update (Coleman et al, 2010). Trust and relationship models can also be used to assess the quality of VGI (Bishr & Janowicz, 2010); geographical proximity to the features contributed and contribution timing are key factors in assessing contributors' trustworthiness.

PROFESSIONAL VS AMATEUR

Each category of VGI contributors has its strengths and weaknesses, and organizations need to carefully assess how their needs will be best met. Professionals will not necessarily make higher quality contributions than amateurs, since factors such as local knowledge and personal motivation play such an important role.

Both NES and Esri Community Maps opted for sourcing/quality controlling VGI contributions from authoritative data providers to meet their quality standards. Lessons learned include:

- **Using data custodians to vet VGI-notified changes can greatly enhance data quality**
- **Benchmarking VGI performance can improve throughput and help to isolate problem areas**
- **Diminishing VGI contributions over time can signal improved quality and user satisfaction**

OpenStreetMap (OSM) sources VGI contributions primarily from citizens, but helps to ensure quality by storing the entire history of all features, including every modification and deletion. Lessons learned:

- **In densely populated areas, contributions from amateurs produce data of equal quality to professionally produced data.**
- **Patterns of individual user behaviour can be accessed if necessary for investigating malicious users who are damaging the quality of data**

More formal means of assessing quality based on contributors' credibility are beginning to appear. VGI site operators can learn, for example, from the programmer's questions and answers system StackOverflow, started in 2008, which was among the first to define user reputation (i.e., as a rough measure of how much a community trusts a contributor, the contributor's communication skills, and the quality and relevance of their questions and answers) as a key criterion for qualifying information quality. The generic online questions and answers platform Quora also recently launched with the promise of tapping into and delivering a higher-trust level of collaborative knowledge (van Exel & Dias, 2011). The use of algorithms and automated methods is another technique, which is employed by Swift River, a project under development alongside Ushahidi (Meier, 2009; Hersman, 2009). It uses a two step process that consists of initially applying machine based algorithms to the incoming flow of data, followed by a filtering process that engages self-interested citizens who curate the information.

Interestingly, the contributions of a neophyte with extensive knowledge and experience of a local region has been shown to be more accurate than updates from a mapping professional who, potentially sitting in an office thousands of kilometers away, lacks the same level of local knowledge. Proven correct in an investigation in an urban area in western Kenya, the classification accuracy of road infrastructure from high resolution satellite imagery, achieved respectively by surveyors and non-surveyors, with and without local knowledge showed that, irrespective of surveying background, those with local knowledge classified roads with over 92% accuracy on average, and professional surveyors and laymen without local knowledge achieved accuracies of 67.7% and 42.9%, respectively (de Leeuw et al, 2011).

The issue of **data authoritativeness** is closely connected with that of the quality of contributions. There is a debate within the geospatial community about the relative merits of professional mapping data providers and VGI-based data providers as sources of the most "authoritative" data (see for example (Ball, 2010) (Vlugt, 2011)). Traditionally, data available from an "authoritative" source was considered to be of the highest quality. The advent of VGI has brought into question the meaning of authoritative geospatial data. VGI advocates point out that such traditional sources of authoritative geospatial data have neither the funding nor the mandate to keep their databases current to within specified timeframes. In addition, VGI sourced data has been shown to be of as high accuracy, completeness and of superior currency, at least in heavily populated areas.

Notwithstanding this debate, national and provincial or state mapping agencies in several public jurisdictions and some professional private data providers are using, or investigating the use of, VGI to help maintain their authoritative geospatial databases. In Canada, the Centre for Topographic Information of Natural Resources Canada is assessing the potential of a collaborative mapping model (i.e., contributions from provincial and municipal mapping organizations, crowdsourcing from citizens, etc.) for data maintenance. ESRI Canada's Community Maps Program sources geospatial information from a range of federal, provincial and local government mapping organizations (Esri Canada, 2011). International examples include: the Notification and Editing Service developed by the Department of Sustainability and Environment, State of Victoria in Australia; the OpenStreetMap Collaborative Prototype (OSMCP) project at the United States Geological Survey (USGS); and swisstopo Revision Service at the Switzerland Federal Office of Topography.

ORGANIZATIONAL CHANGE

Procedural and cultural changes that may be necessary with VGI use (Coleman et al, (2009):

- **Acceptance of and respect for rules imposed by contributor communities (e.g., quick acceptance and use of contributions, crediting of source, etc.)**
- **Toleration of contributor community's values taking precedence over traditional practices and policies (e.g., releasing some control to "the crowd" over decisions about whether or not to post a contribution)**
- **Acceptance of data produced through VGI as a perpetually unfinished artefact (i.e., authoritative geo-information in a state of constant imperfection and fluidity)**
- **Balancing the rights of individual contributors, the contributor community and the producer organization**
- **Shift of the planning and production focus from a "coverage-based" to a "feature-based" orientation**
- **Shift from production of data to filtering of data contributions**
- **Evolution to a mix of professional quality controllers and networks of informed data consumers for quality control**

Decisions will be required on how **quality control** is to be exercised over VGI contributions and who will determine if contributed data is acceptable. Organizations that are VGI site operators (relying primarily on data sourced from the public) employ the model of quality assessment by the community of "produsers". Quality assessment is continuous and improvements are made iteratively through multiple updates of the same features. However, a professional data provider needs to adapt internal quality control procedures to this new data updating mode. For example, some organizations decide to treat known high quality features as fixed and non-editable as a means of anchoring VGI contributions

(McDougall, 2009). Others use the ISO 19113: 2003 Geographic information – Quality principles for assessing VGI data quality (Coote & Rackham, 2008).

While there have been limited attempts to formally assess the quality of datasets created with VGI, an assessment of OpenStreetMap (OSM) was conducted in March 2008 by comparing the positional and attribute accuracy of major roads with the Ordnance Survey (OS) of Great Britain's Meridian 2 dataset (Haklay & Ellul, 2010). This project demonstrated that the quality of contributed geospatial data and attributes is much higher in densely populated areas due to the higher number of volunteers and the ability of those volunteers to focus on their local area, rather than travel long distances. It was also clear that less affluent urban areas and rural areas received much lower VGI contributions, presumably because of factors like lower population density and awareness of VGI, and more limited access to the requisite VGI technologies.

Further research was conducted in 2009 by extending the same analysis to secondary roads and using a higher level dataset for comparison: OS MasterMap (Ather, 2009). This project concluded that high levels of positional accuracy are achievable from OSM (there were still issues with regards to thematic accuracy and attribute completeness), and that organizations considering OSM for commercial purposes (e.g., selling GPS navigation products) will need to develop a strong set of quality assurance measures. Research in Germany, which involved a statistical analysis of the accuracy of OSM and TeleAtlas data compared with official survey data from a city also found that OSM is a suitably accurate alternative for use in urban areas, but recommended further research to compare urban with rural areas, which are mapped with significantly less completeness (Helbich et al, 2010).

UNDERSTAND THE CONTRIBUTOR

Tracking and analyzing the following factors can help in understanding VGI contributors and assessing their credibility:

- **Location of contributed data versus location of contributor's IP address**
- **Timing of data contributions versus independent information (e.g., timing of the contribution of a new road feature compared to independent road construction reports)**
- **The degree of conformity between the same data element or attribute that has been submitted by multiple contributors**

To help sustain contributions over time (Coleman, et al 2009) and (Esri, 2010), consider that:

- **All contributions should be welcome (e.g., attributes like gravel road now paved can be as valuable as geographic data)**
- **Contributors want to receive acknowledgement for their contributions and see quick evidence that they have been used**
- **The process to make contributions should be as easy and streamlined as possible**
- **Volunteers may not be strongly motivated to contribute to extensive feature classification and metadata requirements of public mapping programs**
- **Different user interfaces may be required for first-time or occasional contributors, and internal production people or external "power users"**
- **Tools like a "report as inappropriate content" link allows contributors some control over data quality**

An equally important challenge is assessing the **sustainability of VGI contributions**. Research has revealed that the key motivators that keep contributors engaged include: professional or personal interest, social reward, enhanced personal reputation, and pride of place (Coleman et al, 2010). However, while there is typically great enthusiasm among contributors about the creation of data in a new area, the level of engagement tends to decrease over time as the focus shifts to adding new features and correcting errors in the existing data (Feldman, 2009). Organizations need to assess the likelihood of such motivators being strong enough in a prospective contributor community to ensure the sustainability of their proposed VGI initiative.

3.2 Legal Concerns

Organizations considering the use of VGI will also need to consider a number of interrelated legal issues; such as copyright, licensing, privacy and liability.

It is important to note that, in Canada, information or raw data cannot be subject to **copyright**. Only an original work (i.e., the expression of an idea) can be copyrighted. The creation of the work must involve the exercise of skill and judgment that is not trivial or purely mechanical in nature (see *CCH Canadian Ltd. v. Law Society of Upper Canada*, [2004] 1 S.C.R. 339) such that certain compilations or datasets may be subject to copyright but others are not. In addition, copyright can only be infringed where the work itself (not only its contents) is used or reproduced without the authorization of the copyright holder or in a manner that is inconsistent with the use that was authorized. Merely using or reproducing the information or raw data contained in the work or using a portion of the work that is not substantial gives rise to copyright infringement. Copyright issues associated with VGI will vary, and will depend upon:

- The type of information being volunteered – Where the contributed information was produced by the contributor (e.g., is in the form of a street location mapped by the contributor using GPS), the contributor may have concerns about how their copyright interests will be treated by the host site or by users of the site. If the contributor is not the owner of copyright in the contributed information (e.g., information downloaded from a government web map

COPYRIGHT AND VGI

Take the following steps to help mitigate copyright problems (Australian Copyright Council, 2009):

- **Make contributors aware that copyright may be vested in the crown if a government organization is the first "publisher" of the VGI contributions. (NB: this rule does not apply where the Crown has only communicated the work to the public by telecommunication since that is not considered to be a "publication (s. 2.2(1)(c) Copyright Act)).**
- **Clearly specify this intent in terms and conditions of site use.**
- **Ensure that the permissions given to use the site content are no broader than the permissions received from the people who posted VGI content.**
- **Use a binding agreement with contributors stating that they will not infringe the rights including copyright) of any third parties and that they will indemnify the site operator for any damages arising from any infringing behaviour.**
- **Respond quickly to claims that VGI on the site infringes copyright.**
- **With links to other sites that may contain copyrighted material, take someone clicking on the link directly to the other URL, and respond quickly to any complaint from a linked website.**

service), this will raise issues for the contributor (who may be infringing the producer's copyright by uploading the information) and the host site (which will not want to be held liable for copyright infringement).

- The type of work to which it is contributed – Contributors will typically be asked to license any contributed content to the site operators to the extent necessary to permit the operation of the site, and any future developments. If the information is contributed to web content where other users may modify already contributed content, licence terms should require contributors to waive their moral rights. The licence should additionally include a right to use, reproduce, communicate, publish, translate, etc. (as the case may be) the work for the purpose of subsequent modification by other users of the site, not simply for the operation or future development of the site.

- The commercial or non-commercial nature of that work – The Government of Canada is contemplating a change to the Copyright Act that would permit an exception for user-generated content. This exception would allow an individual to use a work previously published or made available to the public to create a new work or other subject-matter in which copyright subsists, and to authorise an intermediary to disseminate that new work for only non-commercial purposes and with appropriate attribution, when reasonable. There must be reasonable grounds to believe that the existing work was not infringing copyright. As proposed in the current iteration of Bill C-11, this exception would not apply where the use has a "substantial adverse effect" on the exploitation or potential exploitation of the existing work, for example, if the new work acts as a substitute for the existing work. These proposed reforms to Bill C-11 are not yet legislation and do not yet represent the law in Canada. Bill C-11 wording may change before it is enacted.

WAYS TO MITIGATE THE RISKS OF LEGAL PROBLEMS

- **Require VGI contributors to confirm that they have the rights to contributed data and that they will indemnify the organization for any damages arising from law suits relating to the data**
- **Recognize contributions by posting names of contributors if they request the attribution or consent to it, while protecting privacy by not linking specific contributions to names**
  NB: attribution should be given to the contributor of copyrighted material in the manner s/he requests it. There may also be privacy concerns to automatically attributing contributions. Consents or authorizations should be properly documented.
- **Ensure that contributor and user licence terms are consistent**
- **Rapidly remove any content that may potentially infringe copyright or privacy**
- **Use disclaimers alerting users to the limitations of the data that is made available on the site.**

Where the information contributed is just raw data, no copyright issues should arise, as there is no copyright in facts. However, contributors of data that is compiled into *datasets* may assert some copyright in the original selection and arrangement of their data[1], and may insist on a licence for the use of these datasets. This may be the case even where the scope or extent of copyright in the dataset is questionable.

[1] Copyright protection hinges on the originality of the selection or arrangement of the data compiled into the dataset. As a result, the copyright status of some compilations of data is difficult to predict. The creator of a compilation of data will only know if the selection or arrangement meets the originality threshold following a

If datasets are contributed and the contributor is not the owner of copyright in the dataset, this will raise issues for the contributor (who may be infringing someone else's copyright by uploading the dataset) and the host site (which will not want to be held liable for copyright infringement) (HAL Corporation, 2011). A contributor is normally required to guarantee that they have all necessary rights in the works that they contribute, and to indemnify the host site for any law suits that may arise relating to the contributed materials.

Where contributions are coming from individuals, most VGI site operators handle copyright by requiring those individuals to accept agreements that **licence** the contributed content to the operator. A VGI site operator will want to ensure in their licence terms that users who contribute content have the legal right to do so, and that they have the authority to license that content according to the terms set out in the licence agreement. The contributor should also affirm that s/he has not infringed third party rights (IP, moral, privacy or other rights) in contributing the copyrighted content. Dealing with copyright can be more challenging for organizations that will be sourcing data contributions from government or commercial data providers, due to potentially conflicting licence terms. In such arrangements, when recipients of VGI contributions provide regular copies of their updated data to the contributing organizations, issues related to copyright are often avoided.

Where data is sourced from other data providers, conflicting terms in those organizations' data licences may need to be resolved. Adoption of open data licensing models, such as the Open Data Commons licence, can help to overcome this barrier. VGI sites also have to establish licences for users of the content hosted on their site. It can be challenging for organizations that combine data from a wide variety of sources to ensure that all of the data on their site is licensed for use under the same or compatible terms. This issue was encountered in connection with the potential use of VGI for maintenance of GeoBase data (Natural Resources Canada, 2011). The interim solution that was proposed in this case was to: only use VGI for change detection in cases where the contributor's licence was too restrictive; use the contributor's licence where data modifications and dissemination were permitted; and negotiate a common licence in the longer term.

It is important for any contributor of VGI to be aware of **privacy** concerns that may arise with VGI. Contributors of VGI content need to be aware that their submissions may pose a risk to their privacy (e.g., a pattern of submissions in a local geographic area over an extended period may permit their identity to be discovered) and to the privacy of others. In addition, individuals may inadvertently

court decision. Jurisprudence offers guidance as to the required elements and factors that will assist a creator in assessing whether or not a selection or arrangement meets the originality test. See a general enunciation of the originality test in the context of law reports: *CCH Canadian Ltd. v. Law Society of Upper Canada*, [2004] 1 S.C.R. 339; for a database of freelance articles: *Robertson v. Thomson Corp.* [2006] 2 S.C.R. 363); for a Yellow Pages phone book: *Tele-Direct (Publications) Inc. v. American Business Information, Inc.*, [1998] 2 FC 22; for a tax table based in part on a government publication: *Collins v. Rosenthal* (1974), 14 C.P.R. (2d) 143 (Fed. T.D.). Maps are considered to be an artistic work under the *Copyright Act* and must be original to be copyrightable: see *Fox on Canadian Law of Copyright and Industrial Designs*, 4th ed. (Carswell: Toronto, 2011), at 7:10, p. 7-13. If a map already exists, the second map maker must refer to common sources of information and start the map from scratch: *Robinson v. Sands & Macdougall Proprietary Ltd.* (1916), 22 C.L.R. 124 (Australia H.C.). If the map of a given area is compiled from other plans and surveys, it is original and attracts copyright protection: *Geographia Ltd. v. Bacon & Co.*, [1914] Macg. Cop. Cas. 179.

compromise their own privacy or that of third parties when they contribute other types of content that is geospatially tagged, such as location information attached to uploaded photos or videos (Jia, 2010). Additionally, government entities or organizations that collect, use and disclose personal information in the course of commercial activity may have certain special statutory obligations in respect of personal information. Moreover, government entities may be required to make certain information accessible to third parties, upon request. Site operators who receive VGI contributions containing personal information may have an obligation to protect the privacy of their contributors[2]. Government entities must comply with the *Privacy Act*, RSC, 1985, c. P-21, when they collect, use and disclose personal information. They must conduct these activities for the purpose of a specific governmental program or activity (s. 4). If the government entity does not directly collect personal information from individuals, e.g., receives it from a VGI contributor, it must ensure that the person whose information was collected has consented to the collection, as well as its subsequent use and (potential) disclosure by the Crown (s. 4). The Privacy Act limits for what purposes personal information can be disclosed (s. 8(2)).

One way to mitigate possible privacy infringements is to encourage contributors to use pseudonyms, rather than their real names when submitting data. In addition, organizations' VGI applications can provide tools to help safeguard and respect contributors privacy (e.g., transparent and easy ways of setting privacy parameters, restricted access, systems of user levels, and means of reporting abuses). There have been a number of cases reported of site operators sharing their contributors' location information with third parties without their permission (EPIC, 2012) and VGI site operators need to ensure that such practices are not adopted. The Privacy Commissioner has offered remarks respecting new trends in maps and geomatics, as well as legal risks associated with geospatial data. We encourage you to review these publications, among others: http://www.priv.gc.ca/media/sp-d/2009/sp-d_20091015_ed_e.asp & http://www.priv.gc.ca/media/sp-d/2009/sp-d_20090617_ed_e.asp).

Within the VGI context, **liability** can arise from a number of sources. It is normal practice for VGI site operators to have users accept terms of use of their data, which include advice about data limitations. However, VGI user organizations could still be liable in several circumstances. For example, liability may arise from unauthorized use of copyrighted data that is contributed by another data supplier The inappropriate use of personal information by VGI user organizations may result in contravention of privacy legislation or regulations. As stated, government entities must comply with the Privacy Act, by directly collecting personal information (or ensuring that the individual has given their consent to the collection by a third party) only for the purpose of a specific program or activity (s. 4), and by obtaining the consent of individuals to use and disclose that information (s. 5). The information may only be

[2] Individuals may be held liable for the statutory tort of invasion of privacy, which is codified in British-Columbia, Manitoba, Saskatchewan and Newfoundland. To be found liable under that law, the person must have wilfully violated the privacy of another individual. The Ontario Court of Appeal in *Jones v. Tsige*, 2012 ONCA 32, recognized a similar tort at common law, but established a high threshold for proving a claim, i.e., the conduct must have been "highly offensive" and "caused anguish." Generally, there may be a risk that an individual who volunteers the personal information of another person is held liable for invasion of privacy but that risk is limited to instances where the individual has acted wilfully, negligently or recklessly.

disclosed for certain purposes (s. 8). There is a real risk of collecting too much information for unrelated purposes or of not obtaining the proper authorizations.

Organizations could also be liable for the negligent contribution of erroneous data that results in injury or damages to a person having relied upon that data; reinforcing the need for implementation of adequate quality control procedures. Site operator responsibility for user-contributed content hinges upon a number of factors, including the degree of editorial control or filtering asserted by the operator and the nature of the contributions made (opinion-based or purely factual in nature). The higher the degree of control, the greater the site operator's responsibility to rectify erroneous data in a timely manner: see *Carter v. B.C. Federation of Foster Parents Association*, 2005 BCCA 398. The decision to monitor and edit contributed content is a matter of business policy, and must take into account the nature of the services provided by the site operator, the level of legal risk it is willing to shoulder as well as its confidence in implemented quality control measures.

Some commentators who favour "professionals" as contributors to dataset creation and maintenance emphasize that professional geospatial data producers could be sued for providing incorrect, deliberately misleading or defamatory information. In contrast, it is problematic to take legal action against individual amateur contributors because anonymous contributing is widely accepted and Internet Service Providers are very reluctant to reveal details of site owners or contributors (Coleman et al, 2009).

3.3 Archiving and Preservation

One of the key challenges organizations that use VGI face is keeping track of contributions in a volatile and rapidly changing geospatial data environment. Research has shown that some high profile features in busy urban areas (known as "high edit" features) can experience millions of updates (Mooney & Corcoran, 2011). While the history of these massive amounts of contributions, edits, etc. is retained by organizations like OSM, accessing it can be rather difficult and tedious (Full Planet.OSM) history is one option), requiring considerable time and processing power. Organizations need to consider these potential future demands on their data in planning their VGI initiatives.

If an organization decides to embark on a VGI initiative, consideration must be given to the potential future use of the contributed data. Changes in mapped features over time may be of interest. For example, access to data for time series analysis is particularly important for scientific research, and citizen involvement is becoming increasingly important in such areas as biodiversity. While the term VGI is relatively new, the voluntary contribution of geographic information on species occurrences and population numbers and trends is a long-established tradition (Klinkenberg, 2010). Insights can be gained into what is needed today by looking at how VGI was collected and stored in the past, at what efforts are being made today to automatically georeference that information, and how it has been used. Government entities may be under a statutory obligation to retain data for a period of time, and, in some cases, to make it accessible to third parties upon request. In addition, the entity should ensure that proper retention policies are documented and implemented.

Identification of the state of an organization's datasets at a particular point in time may also be required for legal purposes. Electronic discovery (eDiscovery) processes require parties to ensure that proper preservation processes, and responsive search methodologies and selection processes exist to defend the value, reliability and credibility of any documents to be produced for legal proceedings (Selznick, 2011). Large amounts of VGI content can significantly increase the complexity of these undertakings; however, NES archived data has been used in legal proceedings.

PROPER DATA PRESERVATION AND ARCHIVAL METHODS (NES and OSM Case Studies)

- **In data model design, use persistent identifiers for all features, so that feature changes over time can be easily tracked.**
- **Store full details of each addition, deletion or change of features that is derived from VGI, including the identity of the contributor.**
- **Ensure that data is fully backed up, either in singular offsite facilities or across multiple site locations, and can be accessed in the long term.**

Some experts are predicting that applications requiring assessment of geospatial data over time will become more commonplace in Web 3.0, heightening the demand for properly archived data. As Web 3.0 emerges, it is possible that the field of geospatial science will face new challenges in dealing with questions of time in this new 'location-aware and moment-relevant Internet' (Batty et al, 2010). If this vision of the future emerges, all geospatial data providers, including those that employ VGI, will face much greater demands for archived data than exist today.

Good data preservation and archiving practices are important in all cases, and VGI initiatives present some unique challenges. The sheer volume of changes can be an issue, but **good records retention** practices are essential. OpenStreetMap stores full details of each addition, deletion or change of features that is submitted, including the identity of the contributor. While identification of individual contributors is not accessible from OSM under their privacy policy, change data is made available. OSM archived data has been used, for example, to create an animation of feature changes over time in the City of Toronto.

The Notification and Editing Service has learned that the use of **effective data models** will facilitate the identification of specific data at any point in time if required in future. In their data model, features have both "persistent feature identifiers" and "unique feature identifiers". Whenever a feature is updated, it is retired and a link is established with the new feature, so changes in features over time can be tracked with the persistent number. This has paid off for NES, since archived data has been used in a couple of court cases dealing with road changes.

REDUCING SECURITY RISKS

NES reduces the risk of security breaches by using authentication procedures and strictly limiting the system functions available to contributors.

OSM uses the services of volunteer system administrators, who have also employed good IT management practices to mitigate security risks.

Esri has successfully managed security issues by storing the contributed data on separate servers from the cache data that is published on their Community Maps website.

Finally, proper short and long term data storage facilities will be required to meet the need for future accessing and processing of large volumes of data. In the short term, the primary emphasis is on backup of data. Although OSM stores all of its backed up data on site, parts of all the data are replicated on servers all over the world, and this data of the local

areas is updated regularly, in many cases on an hourly basis. In the longer term, data storage can be challenging because organizations need to deal with technological changes over time that can impact the accessibility of very old records. Since data storage technology can remain reasonably usable for as little as five years, migration costs can be considerable. One way to manage this challenge is to use service organizations specializing in data archiving that undertake continual, asynchronous and ongoing migrations in their storage technology (Telepaxx, 2012). HP has published a tecnical white paper on data storage archiving that contains useful hints for dealing with technology change (HP, 2010).

3.4 Security

While procedures to reduce the risk of security breaches with VGI are not dissimilar from in-house production, opening geospatial applications to contributions through VGI can result in some unique security challenges. Malicious attacks such as SQL injection and cross site scripting are possible security issues with VGI if an organization's website is not well designed (Januszewski, 2009). Organizations can prevent such breaches by setting up procedures and employing software to validate data as it is entered, and security can also be enhanced by passing data leaving the system through an anti-cross site scripting library.

Security risks can be mitigated using a variety of techniques. For example, NES uses good data management practices such as user authentication procedures. Login authentication ensures that error reporting and change submission are only performed by authorized users. Certain users, such as the general public, only have the approval to submit change notifications. When a contributor submits a change request, the system generates a notification that is sent to the appropriate custodian(s) as catalogued in NES, as well as to the contributor. Data custodians are able to change and edit the datasets over which they have authority directly in the system.

Although OSM is a volunteer organization, which might leave it more vulnerable to security breaches, it has so far avoided malicious attacks and unauthorized access problems. OSM is fortunate to have a number of experienced systems administrators in their volunteer community who have worked together to employ best practices in systems and data management, including authentication procedures, to mitigate security risks.

Esri Canada has also not experienced any problems with security in the Community Maps Program. Their contributors submit data to a team of GIS professionals within Esri Canada that make the changes, so contributors do not have direct access to the database. In addition, Esri Canada publishes a cache version of the data on the Community Maps website, so users cannot interact with the database, which is stored on separate servers.

4. Conclusions

There is a growing trend of VGI use by public and private sector geospatial information providers. Organizations already using VGI have recognized that available technology and online public participation have converged to produce a powerful new production paradigm for online geospatial data content.

This primer was developed to highlight the key operational policy issues that organizations working with VGI may face – particularly data quality and authoritativeness, legal realities, archiving and preservation, and security concerns. Information provided on policies and practices currently in use and key lessons learned by VGI implementers during their experiences is intended to serve as guidance to anyone wishing to initiate or improve their own VGI system.

Appendix 1: Glossary

Acronym	Term	Definition
	Authentication	The act of confirming the truth of an attribute of a datum or entity (e.g., the identity of a person or software program)
	Authoritative data	Normally used to describe officially recognized data that can be certified and is provided by an authoritative source (i.e., an entity that is authorized by a legal authority to develop or manage data for a specific business purpose)
CGDI	Canadian Geospatial Data Infrastructure	The CGDI helps Canadians gain new perspectives into social, economic, and environmental issues, by providing an online network of resources that improve the sharing, use and integration of information tied to geographic locations in Canada.
	Copyright	The exclusive right to produce or reproduce the work or any substantial part thereof in any material form whatever, or to authorize such acts.
	Cross site scripting	A process used to attack the security of a website by finding ways of injecting malicious scripts into web pages, so that an attacker can gain elevated access-privileges to sensitive page content, session cookies, and a variety of other information maintained by the browser on behalf of the user
	Data custodian	An organization responsible for safeguarding corporate data, which includes: managing geospatial data to ensure it is accessible by the user community; and ensuring appropriate security and dissemination restrictions are in place, data meets data structure and quality standards, is properly managed with regard to accepting new datasets or revisions of existing content, protection, back-up, recovery and archiving.
	Data provider	An organization that is in the business of producing data and making it accessible to users
	Data quality	Fitness for purpose; how well the data satisfies particular needs or fulfills certain requirements to solve a problem
	Electronic discovery	Also called e-discovery or ediscovery, any process in which electronic data is sought, located, secured, and searched with the intent of using it as evidence in a civil or criminal legal case.
	Expert Amateur	Someone who may know a great deal about geographic information, practices it passionately on occasion, but still does not rely on it for a living
	Expert Authority	Someone who has widely studied and long practiced

Acronym	Term	Definition
		geographic information to the point where he or she is recognized to possess an established record of providing high-quality products and services and/or well-informed opinions, and stands to lose that reputation and perhaps their livelihood if that credibility is lost even temporarily
	Expert Professional	Someone who has studied and practices geographic information, relies on that knowledge for a living, and may be sued if their products, opinions and/or recommendations are proven inadequate, incorrect or libellous
	Geotagged media	Media submitted to websites such as photographs or video, other websites, or SMS messages to which geographical identification has been added, which can help users find a wide variety of location-specific information
GPS	Global Positioning System	A space-based satellite navigation system operated by the United States government that provides location and time information in all weather, anywhere on or near the Earth, where there is an unobstructed line of sight to four or more GPS satellites
	Infringement	Violation of a law or right (e.g., copyright or right to privacy)
	Interested Amateur	Someone who has "discovered" their interest in geographic information, begun reading the background literature, consulted with other colleagues and experts about specific issues, is experimenting with its application, and is gaining experience in appreciating the subject
	Liability	Legal responsibility for one's acts or omissions; failure to meet that responsibility leaves one open to a lawsuit for any resulting damages
	Licence	The grant by the owner of intangible or intellectual property, such as a dataset, trademark or software program, of the rights to make certain uses of the property.
	License	The act of granting a licence.
	Licensing	Authorizing by the licensor the use of the licensed material by the licensee
	Malice	A party's intention to do injury to another party
	Moral rights	In Canada, two moral rights are protected: i) the right of the author to be associated (by name or pseudonym), where reasonable in the circumstances, or not with the work and ii) the right not to have the work distorted, mutilated or modified, or used in association with a product, service or cause, without the author's consent and to the prejudice of the author's honour or reputation. This latter right includes the right not to have the work used in association with a product, service cause or institution. (see: *Copyright Act*, *supra* note 30, ss. 14.1, 14.2, 28.1 and 28.2).

Acronym	Term	Definition
	Neophyte	Someone with no formal background in geographic information, but possessing the interest, time, and willingness to offer an opinion on the subject
	Privacy	The ability of an individual or group to seclude themselves or information about themselves and thereby reveal themselves selectively
	Produsers	In the geospatial data context, individuals that are involved in both the production and use of the data, such as active VGI contributors
	Professional data production model	A production methodology that relies on geospatial information professionals creating and updating data in accordance with rigid standards and specifications
	Pseudonyms	Names that people use to hide their real identities
	Security breaches	The overcoming of the means of protecting information on computers from theft, corruption, or natural disaster, which allow the information to remain accessible and productive to its intended users
	SQL injection	A process used to attack the security of a website by inputting Structured Query Language (SQL) statements in a web form to get a badly designed website to perform operations on the database (e.g., to dump the database content to the attacker)
UGC	User-generated content	Any material created and uploaded to the Internet by non-media professionals
VGI	Volunteered geographic information	The widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information
	Web 2.0	A term associated with web applications that facilitate participatory information sharing, interoperability, user-centered design, and collaboration on the World Wide Web (e.g., social networking sites, blogs, wikis, video sharing sites, hosted services, etc.)
	Web 3.0	The first-generation Metaverse (convergence of the virtual and physical world), a web development layer that includes TV-quality open video, 3D simulations, augmented reality, human-constructed semantic standards, and pervasive broadband, wireless, and sensors. Definitions vary widely.
	Work	The original expression of an idea, including works consisting of data, like a memorandum, book, table, chart, plan, photograph, drawing or compilation, so long as the creation of the work required an exercise in skill and judgment that is not trivial or purely mechanical in nature.

Appendix 2: References

Ather, A. (2009, May 29). *A Quality Analysis of OpenStreetMap Data.* Retrieved December 7, 2011, from Department of Civil, Environmental & Geomatic Engineering, University College London: http://homepages.ge.ucl.ac.uk/~mhaklay/pdf/Dissertation-OpenStreepMap-Quality-Aather-2009.pdf

Australian Copyright Council. (2009, March). *Websites: User-generated Content & Web 2.0.* Retrieved December 5, 2011, from Australian Copyright Council: http://www.copyright.org.au/find-an-answer/browse-by-keywords/

Ball, M. (2010, October 10). *What's the distinction between crowdsourcing, volunteered geographic information, and authoritative data?* Retrieved December 8, 2011, from V1 Magazine: http://www.vector1media.com/dialog/perspectives/16068-whats-the-distinction-between-crowdsourcing-vol

Batty, M., Hudson-Smith, A., Milton, R., & Crooks, A. (2010). Map mashups, Web 2.0 and the GIS revolution. *Annals of GIS, Vol. 16, No. 1* , 1-13.

Bishr, M., & Janowicz, K. (2010). Can we Trust Information? - The Case of Volunteered Geographic Information. *Towards Digital Earth: Search, Discover and Share Geospatial Data 2010. Workshop at Future Internet Symposium, September 2010.*

Coleman, D., Georgiadou, Y., & Labonte, J. (2009). Volunteered Geographic Information: the nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research, Vol 4* .

Coleman, D., Sabone, B., & Nkhwanana, N. (2010). Volunteering Geographic Information to Authoritative Databases: Linking Contributor Motivations to Program Characteristics. *Geomatica Vol. 64, No. 1 Special Issue on Volunteered Geographic Information. March* , 383-396.

Coote, A., & Rackham, L. (2008). *Neogeographic data quality – is it an issue?* Retrieved December 7, 2011, from ConsultingWhere Ltd: http://www.consultingwhere.com/resources/Neogeography+Data+Quality+-+is+it+an+issue+-+V1_1.pdf

Cunha, G. S., & Viola, d. A. (2010, August 25). *The Italian Google-Case: Privacy, Freedom of Speech and Responsibility of Providers for User-Generated Contents.* Retrieved December 5, 2011, from International Journal of Law and Information Technology: http://ijlit.oxfordjournals.org/content/18/4/356.full

de Leeuw, J., Said, M., Ortegah, L., Nagda, S., Georgiadou, Y., & DeBlois, M. (2011). An Assessment of the Accuracy of Volunteered Road Map Production in Western Kenya. *Remote Sensing* , 247-256.

EPIC. (2012). *Electornic Privacy Information Centre*. Retrieved January 18, 2012, from Locational Privacy: http://epic.org/privacy/location_privacy/default.html

Esri. (2010). Lessons learned developing a Web map for volunteered geographic information (VGI) and social media.

Esri Canada. (2011, February 16). *Community Maps Program: Fostering Openness & Innovation WHITE PAPER.* Retrieved December 12, 2011, from Esri Canada: https://www.esricanada.com/documents/EC5_122_1012_A_CommunityMaps_whitepaper.pdf

Feldman, S. (2009, September 15). *Is volunteered geographic information sustainable?* Retrieved December 6, 2011, from Vimeo: http://vimeo.com/6590739

Haklay, M. (., & Ellul, C. (2010). Completeness in volunteered geographical information: the evolution of OpenStreetMap coverage in England (2008-2009). *JOURNAL OF SPATIAL INFORMATION SCIENCE* .

HAL Corporation. (2011). *Final Report: Review of IP Law and Instruments (Copyright, Licensing) in the Context of Geospatial Data.* Ottawa.

Helbich, M., Amelunxen, C., Neis, P., & Zipf, A. (2010). Investigations on Locational Accuracy of Volunteered Geographic Information Using OpenStreetMap Data. Heidelberg, Germany.

Hersman, E. (2009, February 4). *Crisis Info: Crowdsourcing the Filter*. Retrieved December 16, 2011, from The Ushahidi Blog: http://blog.ushahidi.com/index.php/2009/02/04/crisis-info-crowdsourcing-the-filter/

HP. (2010, September 14). Solutions for mainframe data storage archiving: Technical white paper. Retrieved February 13, 2012, from HP: http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA0-3329ENW.pdf

Januszewski, K. (2009, Spetember 18). Design, Usability and Security Dilemmas With User *Generated Content*. Retrieved December 5, 2011, from Mix Online: http://visitmix.com/writings/design-usability-and-security-dilemmas-with-user-generated-content

Klinkenberg, B. (2010). *CITIZEN SCIENCE AND VOLUNTEERED GEOGRAPHIC INFORMATION: CAN THESE HELP IN BIODIVERSITY STUDIES?* Retrieved December 12, 2011, from Biodiversity of British Columbia : http://www.geog.ubc.ca/biodiversity/VGI--VolunteerGeographicInformation.html

Maué, P. and S. Schade (2008). "Quality of Geographic Information Patchworks". Proceedings of the 11th AGILE International Conference on Geographic Information Science, Girona, Spain. Retrieved February 10, 2012 from *http://plone.itc.nl/agile_old/Conference/2008-Girona/PDF/111_DOC.pdf*.

McDougall, K. (2009). The Potential of Citizen Volunteered Spatial Information for Building SDI. *GSDI 11 Conference.* Rotterdam: Global Spatial Data Infrastructure.

Meier, P. (2009, April 10). *Developing Swift River to Validate Crowdsourcing*. Retrieved December 16, 2100, from iRevolution: From innovation to Revolution: http://irevolution.net/2009/04/10/developing-swift-river-to-validate-crowdsourcing/

Mooney, P., & Corcoran, P. (2011, July 17). *Volatile Volunteered Geographic Information (VVGI): The OpenStreetMap Example.* Retrieved December 8, 2011, from 1st European State of the Map Conference of the OpenStreetMap project: https://sotm-eu.org/slides/22_PeterMooney_VVGI.pdf

Natural Resources Canada. (2011, November 16). Gestion des licences: Openstreetmap et autres licences. Sherbrooke, ON, Canada.

Produsers.org (2007, December 31). Produsage: A Working Definition. Retrieved January 23, 2012, from Produsage.org: http://produsage.org/node/9

Selznick, S. I. (2011). ADDRESSING E-DISCOVERY AND LITIGATION ISSUES. *Cloud Computing Law Workshop.* Toronto: Federated Press.

Telepaxx. (2012). *Security Concept*. Retrieved February 13, 2012, from Telepaxx: http://www.telepaxx.com/pacs-data

van Exel, M., & Dias, E. (2011). *Towards A Methodology For Trust Stratification in VGI.* Retrieved December 9, 2011, from vgi-net: http://vgi.spatial.ucsb.edu/sites/vgi.spatial.ucsb.edu/files/file/aag/van_Exel_abstract.pdf

Vlugt, M. v. (2011, January 4). *PSMA, Sensis or OpenStreetMap: what makes Spatial Data "Authoritative"?* Retrieved December 2, 2011, from Spatial Information in the 21st Century: http://spatial21.blogspot.com/2011/01/psma-sensis-or-openstreetmap-what-makes.html