Assessing the quality of repositories of volunteered geographical information

Antony K Cooper[1,2], Serena Coetzee[3] and Derrick G Kourie[2]
[1]Built Environment, CSIR, PO Box 395, Pretoria, 0001, South Africa
[2]Department of Computer Science, University of Pretoria, 0002, Pretoria, South Africa
[3]Centre for Geoinformation Science, University of Pretoria, 0002, Pretoria, South Africa

Abstract

In this article we assess two repositories of volunteered geographical information (VGI) (the 2nd South African Bird Atlas Project and OpenStreetMap) against the seven commonly-used dimensions of the quality of geographical information (positional accuracy, thematic accuracy, semantic accuracy, temporal accuracy, completeness, logical consistency and lineage) and in terms of five challenges we identified previously for the quality of VGI (dependence on the purpose and the context in which the data are used, lack of involvement by users in developing standards, anonymous VGI contributions, bias in VGI, and that not all aspects of data quality can be assessed quantitatively). Our research shows that these repositories have procedures in place to check the quality of the data and confirms that it is difficult to assess data quality. This work also contributes towards a new and improved understanding of VGI, its quality and also its usability.

Keywords

accuracy, data, GIS, quality, spatial data infrastructures, standards

Introduction

Motivation and aims

Perhaps the most common objections raised against volunteered geographical information (VGI), that is, user-generated content with a geographical dimension, are uncertainty over the quality of the VGI and poor documentation of the data (e.g. see [1]). Data are only of value to users when they have information that will allow them to determine the quality of the data. One problem is that for different uses, significantly different levels of quality for the same data are necessary or acceptable [2].

Unfortunately, there is a tendency for users to be unaware of the inaccuracies inherent in the data [3], being unaware of the issue of cartographic licence (generalization, aggregation, selective presentation, etc), for example.

In earlier work [4], we considered the different types of VGI from the perspective of quality, plotting them on a matrix of the type of data (ranging from base data to points of interest) against who determines the specifications for the VGI (ranging from custodians to users). This showed the diversity of VGI and that it is not possible to adopt a rigid and narrow perspective on VGI. Clearly, some VGI is definitely not suitable for a spatial data infrastructure (SDI), even with extensive post-processing and quality assurance. In this article we assess two repositories of VGI against the seven dimensions of quality, and against five challenges for VGI quality that were identified in [4].

This work contributes towards a new and improved understanding of VGI, its quality and also its usability. Other authors have also contributed in this regard, for example, [5] who aim "*to frame the crucial dimensions of VGI for geography and geographers, with an eye toward identifying its potential in our field, as well as the most pressing research needed to realize this potential*".

The remainder of the article is structured as follows: we discuss the seven dimensions of quality, VGI and the challenges for the quality of VGI. Then we assess two repositories against these dimensions and challenges. Finally, we conclude, looking at possible future work.

The dimensions of quality

Most, if not all, data and information products produced by government departments and related agencies, such as a statistics agency, have a geographical (spatial) dimension, and hence could be considered to be geospatial. When considering the quality of geospatial data, naïve users often consider only the positional accuracy of the data. However, there is more than just this aspect to the quality of geospatial data – the following dimensions of quality have been widely recognised and used for geospatial data for the last two decades [6,2,7,8,9,10], though the sub-dimensions are not well known and have been taken here from [11]. Unfortunately, the limitations of a conference paper mean that it is not possible to describe in detail the sub-dimensions and how to measure them. Typically, to be able to make a meaningful statement on quality, the sub-dimensions would be measured using a random, representative sample, because of the cost of assessing all the data.

- *Positional or spatial accuracy*: how closely the geometry of features in a digital data set corresponds to the true locations of the related phenomena in the real or imaginary world. The sub-dimensions of positional accuracy are planimetric and vertical accuracy, absolute and relative accuracy, and geometric fidelity. Ideally, positional accuracy should be calculated on the basis of standard error or circular error, and be expressed in terms of metres in the real world, to make it independent of the scale of the data [12].

- *Attribute or thematic accuracy*: how closely the non-spatial attributes and the classification of features in a digital data set reflect the characteristics of the related phenomena in the real or imaginary world. Attribute values can be on nominal, ordinal, interval or ratio scales [13], can be free text, or even multimedia. The classification assigns features to feature types or classes. The sub-dimensions of thematic accuracy are qualitative and quantitative attribute correctness and classification correctness. For data on nominal and ordinal scales, the correctness of a value can only be Boolean (true or false), for data on an interval scale the error can be quantified as a distance from the true value and only data on a ratio scale can also be quantified as a proportional error. However, the aggregated accuracy rates can also be quantified as proportional error.

- *Semantic accuracy*: this measure links the way in which the object is captured and represented in the database to its meaning and the way in which it should be interpreted. The difference between semantic accuracy and thematic accuracy is that the former deals with the transformation of data (particularly for generalization) and the appropriateness of the resulting classification and geometry (particularly shapes), rather than the correctness of the recording of the classes and attributes [14]. It would appear that the interdependence of semantic and thematic accuracy has not yet been assessed. Semantic accuracy has also not been used much to date (it was not included in [8], for example), perhaps because it is considered to be esoteric or because few understand it. The sub-dimensions of semantic accuracy are the shape fidelity and type selection of the aggregated data.

- *Temporal accuracy or quality*: how closely the temporal data (time related attributes and relationships) in a digital data set correspond to the true values for the related phenomena in the real or imaginary world. As the measurement of time is on an interval scale (there is no true zero), the error for a value can be quantified as the difference in time from the true value, but not as a proportional error. One can also record aggregated accuracy rates for each for the five sub-dimensions of temporal accuracy: currency, accuracy of time measurements, temporal consistency, temporal validity and updating efficiency.

- *Completeness*: how well the data set represents all that it is meant to represent, and only what it is meant to represent, that is, the presence or absence of data and data elements. One key problem with completeness is having a common understanding of what is meant to be in the data set – for example, whether or not a data set of roads includes tracks or footpaths. Another problem is how to measure completeness: it is expensive to compare the whole data set to the real world, though it could be done on a sample basis, and it is difficult to find a data set of higher authority for comparison that is known to be complete. These errors can be assessed by considering the data set as a whole or by identifying individual errors. The sub-dimensions of completeness are missing and unexpected data.

- *Logical consistency*: degree of adherence of the data to logical rules of data structure, attribution and relationships – that is, the presence, absence or frequency of inconsistent data, such as inappropriate attributes for a feature, inappropriate relationships between features, values out of range, missing delimiters, mismatched XML tags, sliver polygons or mismatches across data set boundaries. These errors can be assessed by considering the data set as a whole or by identifying individual errors. The sub-dimensions of logical consistency are conceptual, domain, format and topological consistency.

- *Lineage*: the history of the data, including the people and organisations responsible for each stage. To the extent known, it should recount the life cycle of a data set, from collection and acquisition, through compilation and derivation, to its present form. Typically in the lineage, there will be 'chains' of the two sub-dimensions: sources separated by process steps, with some process steps combining multiple sources.

Volunteered geographical information

The Internet has spawned the development of *virtual communities* or *virtual social networks* which share data with one another, and with the public at large. This *user generated content* is most obvious in web sites such as Wikipedia [15], the free, online encyclopaedia in many languages, consisting of contributions mainly from the public at large, rather than from domain experts (though it does also include much content from encyclopaedias that are out of copyright and other expert sources). Similarly, virtual communities have also facilitated *folksonomies* or *collaborative tagging*, which are the classification and identification of content by the general public, rather than by domain experts.

Within *geographical information science (GISc)*, user generated content is also known as *volunteered geographical information* and is made available as base maps on public websites, such as Tracks4Africa [16] and OpenStreetMap [17], or as third party data overlaid on virtual globes, such as Google Earth [18]. VGI is also contributed as observations to *citizen-science* projects such as SABAP2, the Second South African Bird Atlas Project [19].

[20] introduced the term VGI without actually defining it, but suggested that it combined elements of *Web 2.0* (where the user becomes a creator of resources), *collective intelligence* (also termed the *wisdom of the crowd*: aiming for a better answer by involving more people in the process of understanding the problem and deriving the solution) and *neogeography* (new geography, going beyond the traditional scope of professionals). The emerging research on VGI is multifaceted, taking into account industry, technology, discipline, social, political, and other aspects [21]. Much has already been published on VGI: Google Scholar [22], for example, already lists over 1300 items containing the term "*volunteered geographic[al] information*", with about 400 from 2011 alone and over 200 from 2012 so far, showing the tempo of interest.

Conceptually, the issues affecting the quality of volunteered geographical information should be the same as those for professionally generated geographical information. However, there are differences. The ready availability of cheap and reasonably accurate GNSS (global navigation satellite system) receivers means that the positional accuracy of VGI recorded using such a receiver should generally be accurate enough for most consumer-oriented purposes, such as navigation and recording points of interest. Typical errors that are likely to occur with amateur use of a GNSS receiver are transposing coordinates (quite easy to do in South Africa, because the coordinate values for latitude and longitude are similar for a large part of the country), or using the incorrect reference surface. As the GPS signal is relatively weak, it can also be jammed easily, intentionally or unintentionally [23].

The strengths of VGI include openness, market-orientation and interaction between stakeholders, while the weaknesses of VGI include heterogeneous data (e.g. VGI coverage mainly where young and well-educated people live – creating a digital divide within countries), lack of metadata (some contributors are anonymous) and uncertainty over the reliability of the data in comparison to official data [24]. [5] suggest that the abundance of data, geographical context and peer review by users and other contributors makes it difficult to produce incorrect VGI. Of course, these authors live in a developed country rich in data and peer reviewers. On the other hand, [25] considers that such peer review relies implicitly on a fallacy in classical logic, namely *proof by assertion*, or more narrowly in this case, *proof by repeated assertion*.

Challenges for the quality of VGI

Drawing on our observations about user-generated content, volunteered geographical information and data quality, we identified five challenges for assessing the quality of VGI against the seven dimensions of quality in practice [4]:

- *Dependence on purpose & context*: The quality of spatial data is subjective, i.e. data quality depends on the data user, purpose and the context in which it is used. Therefore, the contributor cannot assess the quality of their contribution in isolation. Rather, the user should assess the quality based on their intended purpose and context, using the information provided by the contributor about the data, i.e. the *metadata*. However, despite metadata software utilities and a widely-used international metadata standard [26], metadata is still not readily available for many datasets, and even less so for VGI.

- *Non-involvement in standards*: A further complication is that in general, users are not involved in the development of standards, such as for assessing quality or for documenting metadata. The result is that even if they are aware of the relevant standards, they do not necessarily "buy in" to the standards nor understand their context or utility.

- *Anonymous contributions*: VGI can be contributed anonymously, which makes it very difficult to verify without ground-truth or other data sets for comparison.

- *Bias*: In addition to "normal" errors, contributions could be made to promote a particular political, religious or social agenda; out of malice (e.g. to denigrate someone or some community); with criminal intent (e.g. to manipulate asset prices); or simply out of mischief [27]. Such malevolence can be in both commission and omission. Whereas poor data are likely to be poorly documented, malicious data might well have detailed metadata, albeit fraudulent!

- *Qualitative aspects*: Not all aspects of data quality can be assessed quantitatively, and there are important types that have qualitative aspects to them. While quantitative measures can be understood in many languages (e.g root mean square error for positional accuracy), qualitative assessment is language dependent (e.g. a statement about what should be included in the data set, for assessing completeness).

Assessing the Quality of Two Repositories

Methodology

It is difficult to characterise any particular set of user-generated content (UGC) without having insight into the data set and its creation, or having comprehensive metadata available for the data set. Unfortunately, metadata for UGC is invariably sparse — or even non-existent. Any repository containing UGC (whether made available over the Internet or not) could contain a wide variety of types of UGC, depending on the explicit and implicit policies pertinent to the repository and the diligence with which they are applied. Further, a web site is only a tool to access or display the data in a repository: one web site can access multiple repositories and one repository can be accessed through different web sites.

We conduct here exploratory research on VGI quality using two case studies, the Second South African Bird Atlas Project and OpenStreetMap. We have been able to obtain sufficient information from these repositories to assess the VGI contributions qualitatively (though not the post-processing done by the custodians of the repositories) against the seven dimensions of quality and against the five challenges for the quality of VGI.

Second South African Bird Atlas Project (SABAP2)

For the *Second South African Bird Atlas Project (SABAP2)*, begun in July 2007, volunteer bird watchers (and professional ornithologists) are gathering data according to a detailed, published protocol (recording bird distribution, observer effort and an index of abundance) and submitting the data either directly to the SABAP2 web site, or by sending them through the post [19,28]. The data are gathered by geographical units, namely *pentads* (5' by 5'), and by temporal units, namely *pentades* (up to 5 days). As of 24 June 2012, 1052 observers have submitted 71512 field sheets. While the raw data are not made available on the web site, various processed data are available. SABAP2 builds on SABAP1 (run from 1986 to 1997 [29]), but with a more rigorous protocol with a finer spatial and temporal resolution that produces a better index of abundance. The same platform and similar protocols are being used for other atlassing projects, such as butterfly and reptile atlases.

SABAP2 was selected for this analysis as the first author is familiar with it, having contributed 70 field sheets to date.

- *Positional accuracy*: Field sheets need to be located in the correct pentad (5' x 5', about 9km x 9km), i.e. relative accuracy. The protocol is tolerant of errors greater than those of consumer GNSS receivers, because it accepts observations of birds heard or seen close to the pentad boundary, and because the data are not meant for use at a large scale. For incidental observations, coordinates are preferred (i.e. absolute accuracy), otherwise geographical identifiers with narrative. Vertical accuracy and geometric fidelity are not relevant here.

- *Thematic accuracy*: Probably the biggest problem (classification correctness), as it depends on the identification skills of the observer (jizz, behaviour, plumage, calls, etc) and their knowledge of recent splits and lumpings in the taxonomy. The SABAP2 Data Management System can help by identifying when one is entering confusing species. Species out of range are identified automatically and the observer is required to justify the observation, possibly submitting photos, videos or audio recordings. For example, Figure 1 shows the SABAP2 distribution for Southern Black Korhaan as at 24 June 2012, but it probably includes records of Northern Black Korhaan, as indicated (the two species were split after SABAP1). The field sheets also cater for attributes, though they are not mandatory, so qualitative attribute accuracy (e.g. number observed) and qualitative attribute correctness (e.g. breeding status) can be relevant for some records.

- *Semantic accuracy*: This is not an issue for the data capture (ie: for the VGI), but is for the analysis, which could involve aggregating, interpolating and smoothing data, by an expert. As the aggregation is not done by the observers, it is not relevant for the discussion onVGI quality here.

- *Temporal accuracy*: For full-protocol field sheets, observations need to be for the correct hour (accuracy of a time measurement) and provided in the order recorded (temporal consistency). For ad-hoc field sheets and for incidental observations, observations need to be for the correct day. Many records are submitted months, or even years, after they were recorded, which could have accuracy problems, depending on the quality of the observer's record keeping. Temporal validity is not relevant here, but updating efficiency is, as a measure of how often a pentad is re-surveyed.

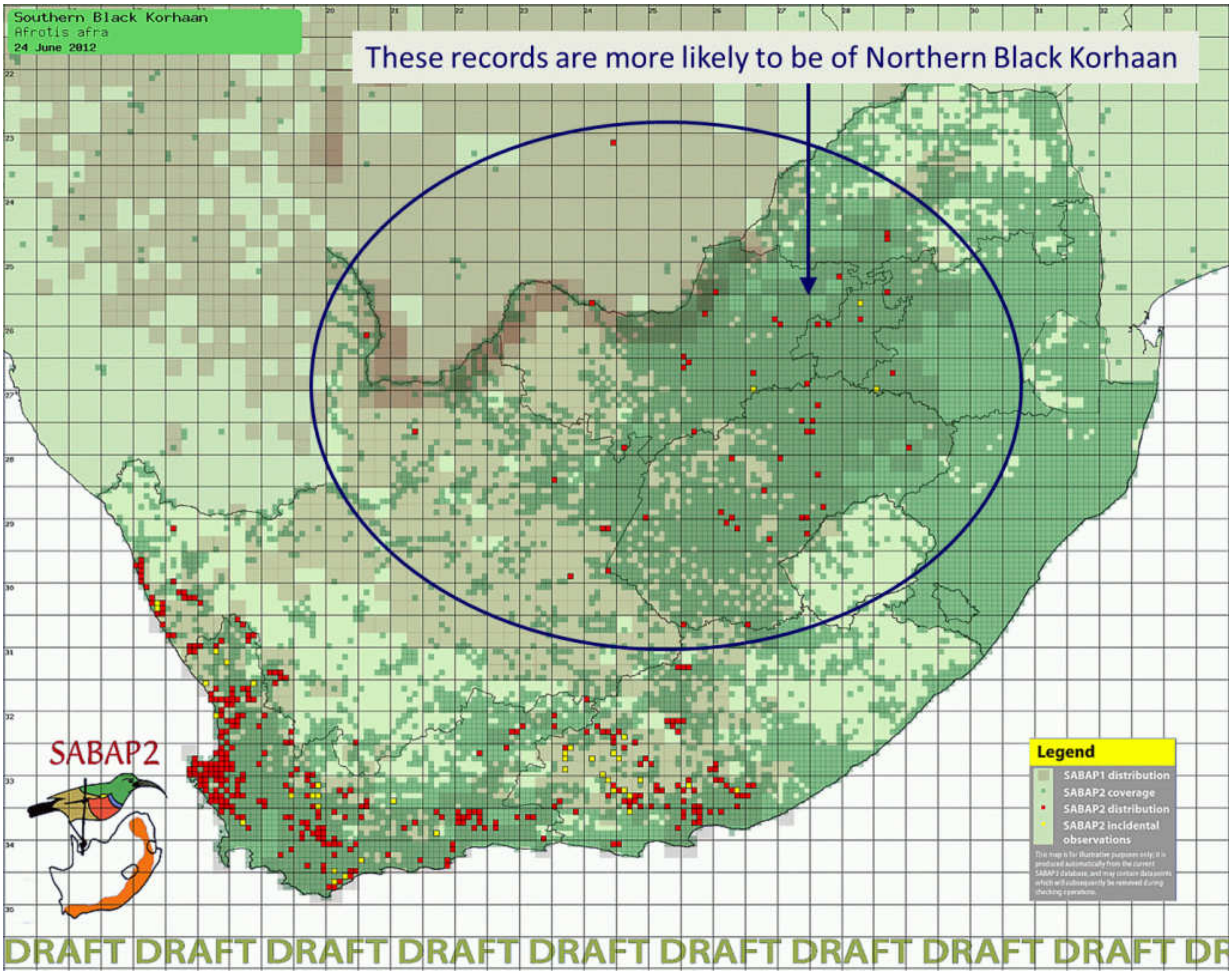

Figure 1: SABAP2 distribution of Southern Black Korhaan, as at 24 June 2012 [19].

- *Completeness*: The more skilled the observer, the better the completeness for a field sheet, in terms of both missing species and unexpected species. Completeness could be estimated by comparison to SABAP1 and other data. Spatial coverage is not even, with 61.7% of the pentads (June 2007 to 2012.06.24) having been atlassed at least once, and about 30% of the pentads being atlassed each year. On the other hand, there are over 50 pentads (~0.3%) that have been atlassed 100 or more times. Hence, it is easy to determine exactly what the spatial completeness is for them.

- *Logical consistency*: The SABAP2 Data Management System presents some fields as drop-down boxes and does some consistency checks before a record can be submitted, and both work well to ensure consistency. Given the nature of the data being captured, conceptual, domain and format consistency will be trapped by the SABAP2 software, but not when an observer submits a record as a spreadsheet file (which is permitted). Topological consistency is not relevant here, because the observers do not submit geometric data.

- *Lineage*: The observer for the submitted record is documented, together with any additional observers (source information). The process steps are not relevant here, as each field sheet or incidental observation is independent.

- *Dependence on purpose & context*: A record does not have metadata indicating the abilities of the observer. However, since the record does identify the observer, one could use all their records to assess their abilities in comparison to other observers with records from the same or similar pentads and pentades. The metadata does provide indications of effort.

- *Non-involvement in standards*: There was little participation by potential contributors in developing the protocol for SABAP2. However, both the protocol and the software were modified early on, to accommodate suggestions from observers. Many workshops have been arranged all over the country to explain the protocol and the SABAP2 tool and experienced atlassers often take novices with them when atlassing.

- *Anonymous contributions*: This is generally not possible, as an observer needs to register first, providing their contact details.

- *Bias*: Potentially, a contributor could deliberately and malevolently exclude a species from a record to reduce its reporting rate, or include it to increase its reporting rate. This could be done to influence environmental impact assessments (EIAs), for example. Specifically, one could omit endangered species to help an EIA get accepted for a proposed development or one could include endangered species to make an area appear to be an environmental hot-spot, to prevent a development. These might not happen in SABAP2, but there is concern that unethical developers might manipulate such data. Hence, while there is an emphasis on obtaining breadth of coverage in SABAP2 (getting records for as many pentads as possible), there is also an emphasis on obtaining depth of coverage (getting many records for each pentad), and both will reduce the vulnerability of any analysis to individual records.

- *Qualitative aspects*: Weather conditions are not recorded and are a big factor in the ease of identifying species in a pentad, but this could be obtained from other sources.

OpenStreetMap (OSM)

OpenStreetMap (OSM) is a repository and a web site providing a free, editable map of the whole world, initiated in 2004 as a repository of VGI, with Wikipedia [15] as its inspiration. OpenStreetMap has a wide variety of tools for detecting possible errors (especially topological errors and missing tags) and procedures for publishing and correcting detected errors. OpenStreetMap data are sometimes more up-to-date and of a higher quality than commercial or official data sets [17]. Several studies have been conducted to assess the quality of the data in OpenStreetMap, such as [30,31,32]. In collaboration with aid agencies, commercial satellite data providers and other organisations, OpenStreetMap made a significant contribution to mapping Port au Prince and other parts of Haiti for relief operations after the earthquake there on 14 January 2010, for example. OpenStreetMap data are widely used and are available through other web sites. OpenStreetMap also contains much data contributed by official mapping agencies, but only VGI is considered here.

OpenStreetMap was selected for this analysis as it is probably the best-known repository of topo-cadastral VGI.

- *Positional accuracy*: Most VGI is vector data captured using consumer-grade GNSS receivers. Most users probably expect to use the data at scales greater than 1:25 000 (ie: within urban areas), so all the sub-dimensions of positional accuracy are relevant here, for both absolute and relative accuracy: planimetric and vertical accuracy and geometric fidelity. Quality assurance is primarily done by other contributors and users.

- *Thematic accuracy*: This is likely to be a problem, particularly where users might not understand the taxonomy properly, especially for points of interest (i.e. classification correctness). Much attribute data are contributed as well, so qualitative attribute accuracy and qualitative attribute correctness are also important. Quality assurance is primarily done by other contributors and users. OSM also allows folksonomies (contributors can classify their data as they like), but it has a taxonomy of preferred classes.

- *Semantic accuracy*: This is not an issue for most of the data capture (ie: for the VGI), as that does not involve integrating different data sets to produce new ones, but is an issue for the analysis, which could involve aggregating, interpolating and smoothing data.

- *Temporal accuracy*: In areas where there are active contributors, OSM data can often be more up to date than official data. In other areas, OSM data are likely to be reasonably current, because OSM has been collecting VGI for less than a decade. All the sub-dimensions are relevant here.

- *Completeness*: Coverage is uneven, with better coverage where there are more observers: developed countries vs developing countries, urban areas vs rural areas, etc. It is difficult to determine where there are no data, except by comparison to other data (where available). OSM does arrange mapping parties to obtain data in unmapped areas, and/or to update or improve data.

- *Logical consistency*: This is not addressed explicitly in the OSM documentation, but such quality problems would be detected by the OSM peer review processes.

- *Lineage*: The history of the data and some details of the observers are recorded.

- *Dependence on purpose & context*: Most contributions are probably to provide data for generic use at reasonably large scales. A record does not have metadata indicating the abilities of the contributor, though as it does identify the contributor, one could use all their records to assess their abilities in comparison to other contributors.

- *Non-involvement in standards*: As OSM has evolved, its standards have been developed with extensive involvement from its contributors and others.

- *Anonymous contributions*: OSM no longer allows anonymous contributions. Contributors are identified by a user name, but have to provide a valid email address to OSM.

- *VGI bias*: Potentially, a contributor could provide biased or false data to support their particular agenda. OSM's primary defence against this is the sheer number of active contributors providing peer review, though one does need to consider the danger of that fallacy in classical logic of proof by repeated assertion [25].

- *Qualitative aspects*: OSM does allow folksonomies, but these are dealt with by providing a mapping to the preferred OSM taxonomy.

Conclusions

We have assessed here two repositories of volunteered geographical information (2nd South African Bird Atlas Project and OpenStreetMap) against the seven dimensions of quality and in terms of five challenges for the quality of VGI that we identified previously.

It is clear that both repositories have procedures in place to check the quality of the data, showing that the implementers of VGI repositories are aware of the importance of quality and the related challenges. OSM's are the most extensive and transparent, while SABAP2 depends on its vetting committees. In both cases, uneven coverage is a challenge. Finally, this work confirms how difficult it is to assess data quality and that the quality assessment depends on the intended usage of the data.

While we have assessed only two repositories here, they are quite different and we feel that we can offer some suggestions and recommendations regarding the usability of VGI in general:

- The problem of *classification correctness* might well be more important than realised for the usefulness of VGI, as even experienced contributors mighty be using classification systems that are out of date, or their own peculiar folksonomies. Hence, educating contributors on the correct classification system is very important.
- The *updating efficiency* and *completeness* can be very uneven, depending on the availability of contributors and the volumes they can contribute.
- The availability of suitable and detailed *metadata* remains a problem, with no obvious solution.
- It is very important to involve contributors in the development of standards and protocols, to encourage participation, reduce opposition to the project and improve the data.
- The best defence against biased or false data would appear to be obtaining multiple records and/or peer review, but one must be aware of the danger of proof by repeated assertion.

It will be useful to apply this analysis to other VGI repositories and to spatial data infrastructures (SDIs). It will also be useful to conduct a more in-depth study of the quality of the data in these two repositories (and others), though perhaps with a specific application of the VGI in mind. We are using formal concept analysis (FCA) to understand the taxonomies used in VGI repositories [33] and intend using FCA to compare the quality dimensions of different repositories.

Acknowledgements

The research reported in this paper has been supported in part by the CSIR's project 59M2215 and by the South Africa/Poland Agreement on Cooperation in Science and Technologies, through the joint research project, *Volunteered Geographical Information (VGI) for Spatial Data Infrastructures (SDIs) and Geoportals*.

References

[1] A.K. Cooper, S. Coetzee and D.G. Kourie: "Perceptions of virtual globes, volunteered geographical information and spatial data infrastructures", *Geomatica*, Vol. 64 No. 1, pp. 333–348, 2010.

[2] A.K. Cooper: *Standards for exchanging digital geo-referenced information*, MSc, University of Pretoria, South Africa, 1993.

[3] A. Zargar and R. Devillers: "An operation-based communication of spatial data quality", *Proceedings of the 2009 International Conference on Advanced Geographic Information Systems & Web Services*, IEEE, pp. 140–145, 2009.

[4] A.K. Cooper, S. Coetzee, I. Kaczmarek, D.G. Kourie, A. Iwaniak and T. Kubik: "Challenges for quality in volunteered geographical information", *Proceedings of Africa GEO 2011*, Cape Town, South Africa, June 2011.

[5] S. Elwood, M.F. Goodchild and D.Z. Sui: "Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice", *Annals of the Association of American Geographers*, Vol. 102, No. 3, pp. 571-590 2012.

[6] H. Moellering (ed): *Digital cartographic data standards: an interim proposed standard*, National Committee for Digital Cartographic Data Standards, USA, No. 6, 164 pp., 1985.

[7] S.C. Guptill and J.L. Morrison (ed): *Elements of spatial data quality*, Elsevier and the International Cartographic Association, 1995.

[8] ISO 19113:2002, *Geographic information – Quality principles*, International Organization for Standardization (ISO), Geneva, Switzerland.

[9] P. Bolstad: *GIS Fundamentals*, Eider Press, White Bear Lake, MN, USA, 2005.

[10] P.A.J. van Oort: *Spatial data quality: From description to application*, PhD, Wageningen University, the Netherlands, 2006.

[11] A.K. Cooper, S. Coetzee and D.G. Kourie: "Data quality of geographical information", in preparation, 2012.

[12] D.G. Clarke, A.K. Cooper, E.C. Liebenberg and M.H. van Rooyen: *A national standard for the exchange of digital geo-referenced information*. Special Report SWISK 45, CSIR, Pretoria, South Africa, Sep 1987.

[13] S.S. Stevens: "On the theory of scales of measurement", *Science*, Vol. 103 No 2684, pp. 677–680, 1946.

[14] J.-H. Haunert and M. Sester: "Assuring logical consistency and semantic accuracy in map generalization", *Photogrammetrie-Fernerkundung–Geoinformation (PFG)*, Vol. 2008 No. 3, pp. 165–173, 2008.

[15] Wikimedia Foundation, Inc: "Wikipedia", http://en.wikipedia.org/, 12 Jan. 2012.

[16] Tracks4Africa: "Tracks4Africa: Mapping Africa, one day at a time", http://www.tracks4africa.co.za/, 12 Jan. 2012.

[17] OpenStreetMap: "OpenStreetMap: The Free Wiki World Map", http://www.openstreetmap.org/, 12 Jan. 2012.

[18] Google: "Google Earth: Explore, Search, and Discover", http://earth.google.com/, 12 Jan. 2012.

[19] Animal Demography Unit: "Southern African Bird Atlas Project 2", http://sabap2.adu.org.za/, 12 Jan. 2012.

[20] M.F. Goodchild: "Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0, Editorial", *International Journal of Spatial Data Infrastructures Research*, Vol. 2, pp. 24-32, 2007.

[21] S. Elwood: "Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice", *GeoJournal*, Vol. 72, pp. 133–135, 2008.

[22] Google: "Google Scholar: Stand on the shoulders of giants", http://scholar.google.com/, 12 Jan. 2012.

[23] A.F. Van Niekerk and L. Combrinck: "The use of civilian-type GPS receivers by the military and their vulnerability to jamming", *South African Journal of Science*, Vol. 108, No. 5/6, 4 pp., 2012.

[24] A.K. Cooper P. Rapant, J. Hjelmager, D. Laurent, A. Iwaniak, S. Coetzee, H. Moellering and U Düren: "Extending the formal model of a spatial data infrastructure to include volunteered geographical information", *Proceedings of the 25th International Cartographic Conference*, Paris, France, July 2011.

[25] M. Keeler: "Crowdsourced Knowledge: Peril and Promise for Conceptual Structures Research", *Lecture Notes in Computer Science*, Springer, Vol. 6828/2011, pp. 131-144, 2011,

[26] ISO 19115:2002, *Geographic information – Metadata*, International Organization for Standardization (ISO), Geneva, Switzerland.

[27] D.J. Coleman, Y. Georgiadou and J. Labonte: "Volunteered geographic information: The nature and motivation of produsers, *International Journal of Spatial Data Infrastructures Research*, Special Issue on GSDI-11, Vol. 4, pp. 332-358, 2009.

[28] D. Wright: *Evaluating a citizen science research programme: Understanding the people who make it possible*, MSc, University of Cape Town, South Africa, 2010.

[29] J.A. Harrison, L.G. Underhill and P. Barnard: "The seminal legacy of the Southern African Bird Atlas Project", *South African Journal of Science*, Vol. 104 No. 3-4, pp. 82-84, 2008.

[30] M. Haklay: "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets", *Environment and Planning B: Planning and Design*, Vol. 37, pp. 682-703, 2010.

[31] P. Mooney, P. Corcoran and A.C. Winstanley: "Towards quality metrics for openstreetmap", *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, pp. 514-517, 2010.

[32] N. Govender: *A South African Analysis of the Quality of Volunteered Geographic Information in OpenStreetMap*, Honours project, University of Pretoria, South Africa, 2011.

[33] A.K. Cooper, D.G. Kourie and S. Coetzee: "Thoughts on exploiting instability in lattices for assessing the discrimination adequacy of a taxonomy", *Proceedings of Concept Lattices and Their Applications (CLA 2010)*, Seville, Spain, October 2010.

Contact Antony Cooper, Built Environment, CSIR, PO Box 395, Pretoria, 0001. 012 841 4121. acooper@csir.co.za