CHAPTER 5

Handling quality in crowdsourced geographic information

Laura Criscuolo*, Paola Carrara*, Gloria Bordogna*, Monica Pepe*, Francesco Zucca†, Roberto Seppi†, Alessandro Oggioni* and Anna Rampini*

*CNR-IREA, via Bassini 15, 20133 Milan, Italy, criscuolo.l@irea.cnr.it, carrara.p@irea.cnr.it, gloria.bordogna@idpa.cnr.it

†Department of Earth and Environmental Sciences, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy

Abstract

While spatial information quality is an established discipline in traditional scientific geographical information (GI), standards and protocols for representing and assessing the quality of geographic contributions generated by volunteers or by the generic 'web crowd' are still missing. This work offers an analysis of strategies for quality control and describes a simple representation of the components of the quality in crowdsourced GI. In this framework, and based on the research carried out in Criscuolo et al. (2014), we also introduce a methodology for quality assessment, based on the given representation, which goes beyond the limitations of previous methods in the literature defined for a specific purpose, being able to deal with many quality features, GI categories, and types of application. The method is designed as a decision making approach, so flexible as to take into account the purpose of GI analysis, and so transparent

How to cite this book chapter:

Criscuolo, L, Carrara, P, Bordogna, G, Pepe, M, Zucca, F, Seppi, R, Oggioni, A and Rampini, A. 2016. Handling quality in crowdsourced geographic information. In: Capineri, C, Haklay, M, Huang, H, Antoniou, V, Kettunen, J, Ostermann, F and Purves, R. (eds.) *European Handbook of Crowdsourced Geographic Information*, Pp. 57–74. London: Ubiquity Press. DOI: http://dx.doi.org/10.5334/bax.e. License: CC-BY 4.0.

as to make explicit the criteria driving to quality evaluation, namely the quality features (e.g. the credibility of the volunteers, or the accuracy of the spatial features, etc.) and their relevance.

Keywords

Crowdsourced Geographic Information, Volunteered Geographic Information (VGI), spatial information quality, quality assessment

Main issues in utilizing crowdsourced GI for scientific purposes

With the development of the geo-web and the increasing popularity of mobile devices and communication technologies, in the last decade many geographic information consumers have extended their role to the most active one of geographic content producers. The geographic information generated so far, is characterized by great heterogeneity – both in semantics, formats, contents, and quality.

In fact, crowdsourced GI is most frequently provided on the web – by both aware contributors within scientific initiatives (VGI) and unaware contributors within social networks – in the form of text commenting events, advice, warnings related with physical locations, geo-tagged photographs, and points of interest (POIs), corresponding for instance to historic, cultural, and naturalistic destinations valuable for touristic or commercial purposes. Contributors frequently provide also a geometric georeferenced representation of the footprints of the POIs in the form of points, polylines, or polygons (e.g. centroid of a building, a polyline for a road or a trail, a polygon for a park boundary). This geometric information can be acquired by GPS, or by sensors, or special equipment.

These contributions arise the interest of the scientific community, historically engaged in the creation and distribution of geographic information, together with concerns on the consequences that such new practices can lead to established scientific disciplines.

In fact, there are many problems related to the creation and use of geographic information coming from non-traditional sources for scientific purposes.

First of all, for a (spatial) dataset to be reusable, it should be coupled with its metadata, which define the domain within which its usage is recommended (temporal and geographic references, spatial resolution, quality and validity, constraints, etc.). Crowdsourced geographic information is often lacking, in whole or in part, meta-information allowing us both to locate it precisely in space and time, and to evaluate the basic parameters for its usage, such as acquisition procedure, measurement accuracy, instrumental precision, time stamps, contact details, etc. (Sui, Elwood & Goodchild, 2013). In some cases, some elements are available to enrich the meta-information of the crowdsourced contribution, but they are expressed in unusual forms (for instance authors' nicknames, tags and geotags, external links, attached Exif files, etc.).

This issue especially emerges when datasets of user generated content created by their authors for non-scientific purposes (social, promotional, documental, etc.) are retrieved, selected, and exploited in the framework of scientific projects, for public or governmental decision making purposes.

The second critical point arises when processing crowdsourced geographic information. In fact, while gathering large volumes of user generated geographic contributions is relatively easy (typically 15% of social media contents are georeferenced), to spatially overlay and thematically integrate this information could be extremely difficult. This is due to the different – or commonly undefined – instrumental precision, reference systems, spatial and temporal granularity, together with the absence of common attributes and conceptual schemas, which often make the spatial analysis of user-generated georeferenced data a burden.

A third issue is related to the trustworthiness of contributed data. The quality of a crowdsourced contribution indeed is not just a characteristic of the data: it is also related to the author's reliability and experience, i.e. knowledge of the domain and ability in using the tools for data creation. By taking into account these aspects, it is possible to state the trustworthiness of the information.

The concept of trustworthiness suits both the conventional production of expert scientific information, and the crowdsourced contents, even if the latter is more complex, due to several reasons, among which the difficult traceability of authors, their unknown reputation, and the lack of standards and merit systems. In the last decade several studies have been focused on building credibility models (Metzger 2007; Keβler et al. 2013), analyzing quantitatively and qualitatively user generated content fluxes on the web by discussing their intrinsic characteristics, sources, subjects, drives (Eysenbach & Kohler 2002; Coleman et al. 2009; Van Dijck 2009), and currently the issue is still open and debated.

Because of the absence of a systematic procedure for amateurs' data production, it is commonly acknowledged that official data have a greater reliability and usefulness to science, while volunteered and non-specialist data are more affected by inaccuracies and contain less scientific value. Some authors have spent efforts to prove - or contradict - such a hypothesis by comparing datasets of crowdsourced and specialized observations. Dickinson et al. (2010) reports a series of studies in which variations in observer quality are correlated to the author's preparation. Among factors influencing such variations are background and experience (Galloway et al. 2006) together with the type of task (De Solla et al. 2005; Genet & Sargent 2003; Lotz & Allen 2007), the level of training, the company of a specialist in the field (Fitzpatrick et al. 2009), and the age and education of the author (Delaney et al. 2007). On the other hand, several studies have shown that the creative, aggregate use of non-expert contributions can generate new valuable information (De Longueville et al. 2010; Antoniou et al. 2010; Friedland & Choi 2011), and have documented situations in which local knowledge or expertise provide information of greater value than the expert knowledge alone (Fisher 2000). There is evidence of the high potential of crowdsourced geographic information when collected and managed in well-structured contexts, also in the results of the analysis conducted by authors such as Haklay (2010), Girres and Touya (2010), Ciepłuch et al. (2010), who have evaluated the accuracy of OpenStreetMap data against reference sources, and found that sometimes crowdsourced data have a better accuracy than the reference datasets.

Several other sensitive topics can be identified, related to the scientific usage of crowdsourced GI. Since an adequate treatment of these problems would be beyond the scope the paper, we just mention here the complex issue of the reproducibility in the procedures, the one of personal data protection, and those related to the distribution policy (establishment of intellectual property rights, copyrights, and related rights).

In order to make it possible a controlled use of crowdsourced contributions depending on the purpose of the reuse, the authors propose to establish a theoretical framework for a flexible and transparent quality representation and handling (further analysis on this can be found in Criscuolo et al. 2014, Bordogna et al. 2014a and in Bordogna et al. 2014b): flexibility is intended to offer the possibility to customize the criteria of the quality assessment to different purposes and needs; transparency is intended to offer the possibility for a user to know the criteria used for selecting the crowdsourced information.

In the next sections the topic of quality management for generic GI is addressed, firstly by describing a comprehensive model to represent GI quality, then by discussing the approaches for its control, finally by introducing a methodology for its assessment, suitable for both traditional and crowdsourced GI.

Representing quality in crowdsourced GI

In this work the types of multimedia geographic information are grouped in the following categories:

- **images**: photographs, video recordings and graphic objects;
- **annotations**: mostly textual reports;
- **features**: spatial entities, mono- or multi-dimensional, with associated attributes (such as Shapefiles or Geography Markup Language files);
- **measurements**: values derived from human or sensor's observations, mainly as numbers.

The contributions expressed in form of rating, i.e. the public evaluation of user-contributed geographic contents (e.g. thumbs up / down, star ratings…), are deliberately excluded; in fact, these expressions are certainly informative, and are widely used too, but are more similar to quality assessment tools than to stand alone geographic information. For this reason in the present work the contributions in the form of ratings will be discussed as mechanisms for quality control, i.e. a kind of quality indicators, and not as a type of crowdsourced GI.

Each GI item, i.e. each informative contribution, can consist of a single piece or be composed of multiple elements. In case of multiple elements, they may belong to the same category (for instance, they can report measurements of various physical parameters from a single measuring station) or be of different categories (for instance a photograph and its textual description).

Once the category and structure of a GI have been described, it is necessary to represent its quality.

The discussion on quality in GI has a long history, which starts from the last century, deepens with the advent of GIS technology (for a comprehensive review refer to Van Oort 2006), and finds a new flourishing in the last decade, with the advent of geo-web and the proliferation of collaborative mapping applications. In fact, although the quality of GI has been widely discussed and has its reference standard in ISO 19157:2013 (ISO/TC 211/2010 - Geographic information/Geomatics), the quality of crowdsourced GI presents some different features, such as to require new indicators to be adequately described and evaluated (Van Exel et al. 2010).

The quality of crowdsourced GI is actually a composite property: it includes not only some aspects dealing with the characteristics of the data, but also aspects dealing with the characteristics of the data producer and with the application context.

In ISO 19113-15 two main categories of quality are taken into account: Internal and External. The first one relates to intrinsic characteristics of information (spatial accuracy, temporal accuracy, semantic accuracy…), while the second one deals with the fitness for use of the information. These categories are certainly necessary to perform quality assessment on single pieces of information or on whole datasets, but they don't cover a third aspect of user generated information quality, which is important especially for crowdsourced resources: the trustworthiness of information.

To take into account this complexity, we choose to describe the quality of GI through three main categories, inspired by the ISO 19113-15 and by the thematic literature:

- **intrinsic quality**, corresponding to ISO internal quality, which depends on the characteristics of the informative content;
- **extrinsic quality**, which depends on the characteristics of the context, and responds to the needs of assessing the credibility both on the information

and on the author (Flanagin & Metzger 2008; Galloway et al. 2006; Genet & Sargent 2003);
- **pragmatic quality**, first described by English (1999) and similar to the ISO external quality, which measures the capability to meet the needs of a user or of a usage.

The features that contribute to determine the intrinsic, extrinsic, and pragmatic quality of a piece of geographic information can be broken down into elementary properties. These properties are many and varied, and can be updated under different project conditions. We list only a few of them, selected from the most important and most frequent in the specialist literature.

Intrinsic quality can be described, for instance, by the following elementary properties:

- **accuracy**, i.e. its conformity to the actual or expected value;
- **precision**, i.e. the repeatability of the observation or of the measurement;
- **correctness**, i.e. the absence of formal errors;
- **completeness**, i.e. the absence of significant omissions;
- **intelligibility**, i.e. the possibility of the contribution to be understood and examined.

The elementary properties relatable to the extrinsic quality can be:

- **reliability** of the information;
- **credibility** of the author.

Finally, the pragmatic quality can be described by the two following elementary properties:

- **pertinence** of the information;
- **fitness for** a particular **use**.

While the elementary properties contributing to define the intrinsic and extrinsic quality can be defined by evaluating elementary quality indicators associated to specific pieces of information constituting the VGI items, the last two properties, pertinence and fitness for use, may both be defined in terms of the extrinsic and intrinsic quality (Bordogna et al. 1914b).

The representation of quality sketched in Figure 1 is independent from the categories of contributions (images, annotations, measurements, features), from the information content and context, and can therefore be taken as a general framework to evaluate – and possibly compare – the quality in any crowdsourced or generic GI project.

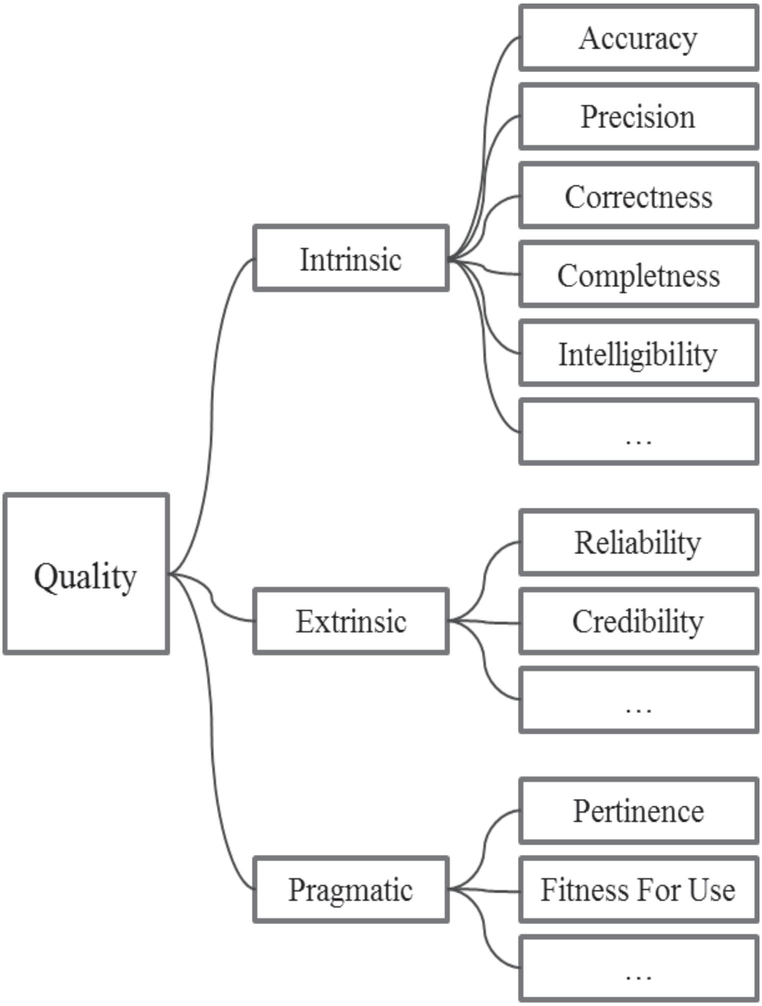

Figure 1: A sketched representation of the proposed categories and the main elementary properties of crowdsourced GI quality.

Approaches to quality control

Once defined the categories of crowdsourced GI and reported its quality properties, we can describe the types of approach to quality control.

In temporal terms the approach to quality control may take place via:

prevention, if it takes place through procedures that precede or are contextual to the submission of information (e.g. learning materials, controlled vocabularies, web forms that guide data producers in compiling their contributions);

correction, if it occurs after the contributions are submitted to the system (e.g. selection of contributions, automatic or manual corrections).

From the point of view of the actors involved, the operations for a quality control can be carried out:

by the **administration team**, if they are performed manually by the project coordinators, a technical staff or a group of experts;
by the **community** of participants, if the group of volunteers itself assesses and validates the information entered;
automatically, if one or more IT components of the system operate the content selection or make some automated edits.

Finally, from the point of view of the remedial action performed, the contributions considered unsuitable may be subject to:

warning, and then be published with an appended message, an alerting symbolism, or a notice;
removal, being excluded from publication and successive processing.

The described approaches to quality control in crowdsourced GI are represented in Figure 2.

Figure 2: A representation of the approaches to quality control in crowdsourced GI.

For each use and each context of application the best fitting strategy must be carefully designed.

Here follows the discussion of the advantages and disadvantages associated with each option, presenting some possible implementations.

The preventive approach, aimed to facilitate the correct compilation of VGI before its submission, may consist, for example, in simple manuals and handbooks; in assisted completion procedures, with multiple choice fields, wordlists, or auto-completion; in automatic tools for the normalization of contents or for the automatic extraction of metadata; in ontologies and geographic gazetteers (Popescu et al. 2009; Kuhn 2001). Common preventive actions are also the selection and the training of volunteer contributors (Galloway et al. 2006; Crall et al. 2011). All these methods facilitate a uniform and formally correct data entry from contributors (intrinsic quality), but they do not ensure to control the reliability (extrinsic quality) or the fitness for use of the information entered (pragmatic quality).

The corrective methods act instead ex post, by amending or removing the weak VGI contributions. They may include the use of automatic algorithms or geostatistical filters (De Tré et al. 2010; Latonero & Shklovski 2010), but also can apply a human supervision to identify systematic errors and maintain the consistency of the dataset (Dickinson 2010; Huang et al. 2010), or still can monitor in real time the semantic integrity of the collected data (Pundt 2002). The corrective methods are suited to act on the reliability and effectiveness of the information provided (extrinsic and pragmatic quality), as well as on the intrinsic quality characteristics, but since they act for removing, merging, or reshaping inappropriate contributions, they may cause partial or even total loss of information.

A quality assessment performed by a team of experts, or supervisors, offers some guarantees, assuming they use competence, wisdom and fairness in the task. Yet even scientists or professionals cannot always enjoy a full mastery of all the variables, and their judgment can be subjective and approximate. It may happen that local citizens, or specialists in particular activities, or direct observers of phenomena make more detailed and reliable assessments than their scientific supervisors. On some occasions, however, it is hazardous to assign assessment tasks to volunteers. In fact, for lack of expertise, superficiality or bad faith, they could create confusion and even hamper the entire data collection.

The combination of the two methods – the traditional authoritative (or top-down), and the democratic one (or bottom-up), is not only possible, but can also produce significant results. In this context, in fact, the web can be used as a meeting point for a collective assessment: the ongoing access to a web item by a hybrid team of experts, local amateurs, occasional visitors, who are asked for evaluating the content, can give rise to a kind of participative evaluation of quality, with a high potential for selection and judgment (Flanagin & Metzger 2008; Connors et al. 2012).

Finally, the automatic control mechanisms can be extremely useful, especially when crowdsourced contributions form large volumes of data (Spinsanti & Ostermann 2013). At present, however, automatic control systems rarely reach an adequate degree of reliability, comparable to a human validation. They are likely to fail especially as regards to the relevance of the contribution, the pertinence (pragmatic quality), and the intelligibility/correctness of the textual content (intrinsic quality).

The control mechanisms that act for removing flawed contents are of great help to preserve the integrity and the consistency of the data collections. Nonetheless, since they discard information considered inadequate according to pre-set parameters, they lead to the exclusion or to the partial loss of information that, no matter how flawed, might be useful in other contexts. The control mechanisms, which keep the whole submitted information, even if not compliant, but report the flaws, do not lose any entered information and encourage the users to access the data consciously. These warning mechanisms, however, have two adverse consequences: on the one hand the storage process is non-effective, because it allocates some memory to data of doubtful relevance; on the other hand the usage is made more difficult, because the system lets the user decide on the data reliability.

Each one of the described options should be considered and evaluated carefully by the project coordinators; nevertheless, most of the times hybrid methods can help in achieving a proper management of quality, by balancing the pros and cons of the various strategies.

Suggesting a quality estimation method for crowdsourced GI

Whatever the strategies to address the control of quality in crowdsourced GI, subsequently it is useful to define a method to estimate the results. This estimation is important not only to establish the level of quality reached by the single contributions and by the whole dataset, but also to monitor the quality trends over time.

In recent years several efforts have been made to develop procedures for quality assessment. Some of them focus on the credibility issues (Metzger 2007), some others focus on the geographical accuracy (Keβler et al. 2013; Sabone 2009; Goodchild 2008), which is often calculated by comparing different datasets or by validating a data sample with a field survey (Haklay 2010). These proposals, while effective in assessing particular aspects of quality, are useful in their specific context, but do not offer a general or flexible method, nor include the different aspects that characterize the quality of non-traditional GI.

To overcome these limitations, we base on the representation of quality introduced in section 2 which makes it possible to define some elementary quality indices, to be associated with each component of the GI items, and then to aggregate the elementary indices into composite ones, until reaching an overall index of quality for the information item, and, in case, the quality index for a whole dataset. A similar method has been described in Bordogna et al. (2014a), and is proposed here in a simplified form, so as to make it easily applicable and customizable to any provided GI, i.e. traditional, non-expert, volunteered or even unaware.

First, we decompose a generic GI item into its elementary components, which may consist of one or more images, annotations, measurements, geographic features. The overall information of a contribution, which we name GI_{TOT}, is therefore achieved by aggregating the *n* informative elements GI_i, i=1,… n.

$GI_{TOT} = \oplus (GI_1, GI_2, GI_3,..., GI_n)$ $\quad \oplus$ being a mathematical aggregation operator

An overall quality index Q_{TOT} is then associated to the overall information GI_{TOT}. Q_{TOT} results from the aggregation of the *n* Q_i indices associated with the *n* components. In this aggregation step, each index Q_i is associated with a numerical weight K_i, which is properly chosen by the analyst depending on specific design requirements. Also the aggregation is chosen by the analyst, for example it may be a weighted average or a sum.

$$Q_{TOT} = \oplus (K_1Q_1, K_2Q_2, K_3Q_3, ..., K_nQ_n)$$

Each Q_i is in its turn the result of the aggregation of three quality indices – I_i, E_i, P_i – respectively connected to the intrinsic, extrinsic, and pragmatic properties of the GI quality.

I_i, E_i, and P_i, can be in their turn associated with three weights – K_I, K_E, and K_P – that are also set by the analyst, depending on the relevance stated for each property of GI quality.

The overall quality for a GI item results in:

$$Q_{TOT} = \oplus (K_1{*}\oplus (K_I I_1, K_E E_1, K_P P_1), K_2{*}\oplus (K_I I_2, K_E E_2, K_P P_2), ..., K_n{*}\oplus (K_I I_n, K_E E_n, K_P P_n))$$

I_i, E_i, and P_i can be finally decomposed in lower level indices, related to the elementary properties of GI quality: accuracy, precision, correctness, completeness, intelligibility, reliability, credibility, pertinence, and fitness for use.

Even at this level, the comprehensive evaluation of I_i, E_i, and P_i is performed by the aggregation of their lower components:

$$I_i = \oplus (\text{accuracy}_i, \text{precision}_i, \text{correctness}_i, \text{completeness}_i, \text{intelligibility}_i)$$
$$E_i = \oplus (\text{reliability}_i, \text{credibility}_i)$$
$$P_i = \oplus (\text{pertinence}_i, \text{fitness for use}_i)$$

This multi criteria assessment of GI quality depends on both the relevant quality indexes Q_i (those with weight K_i>0), the number of such relevant indexes, and the aggregation operator used to combine them.

The described indices and the progressive levels of aggregation are represented in the Figure 3a.

In order to clarify how the model applies to real cases, we can include as an example a real VGI project and make the quality indices explicit. Let's assume to work as analysts in the famous Wikimapia[1] project, and try to estimate the whole quality of a VGI item, consisting in a polygonal shape with an annexed photo. The quality index associated to the polygonal feature will be named Q_1, and the one associated to the photo will be Q_2.

We choose a sum function to perform the aggregation and set the weights for the two VGI components Q_1 and Q_2, and for the three quality indices I_i, E_i and P_i, depending on our project interests, in the following way:

$K_1 = 1{,}5 \quad K_2 = 1$	assuming more interest in preserving the quality of the polygonal feature than the quality of the photo;
$K_I = 1 \quad K_E = 1 \quad K_P = 0{,}5$	assuming more interest in controlling the intrinsic and extrinsic quality than the pragmatic one.

Now we set some numerical values to the elementary quality properties, simulating a likely situation in the Wikimapia project. Let us define the numerical values in the domain $[-1, 0, 1]$:

we set the feature accuracy and the feature precision = 0, assuming, in this example, that is not possible to determine them directly in Wikimapia;[2]

we set the feature correctness, completeness and intelligibility = 1, assuming they are completely fulfilled;

we set reliability = −1 and credibility = 0, assuming that some users from the Wikimapia community commented negatively the entered feature, and assuming the author is a neophyte (corresponding to *user level 0*, or *Unregistered* in Wikimapia);

we set pertinence and fitness for use = −1, assuming the polygonal feature entered is not belonging to the categories requested in the project (for example it could figure out the area in which a temporary event takes place);

we set similarly the values for the elementary quality properties of the photographic component of the VGI item.

---

[1] http://wikimapia.org is a multilingual open-content collaborative map, where volunteers are asked to mark places, add descriptions provided with proof links, give them appropriate categories and upload photos.

[2] In the literature some procedures have been developed and adopted to calculate geometrical accuracy and/or precision of VGI polygonal contributions, usually based on a comparison with the base map images. Nevertheless these procedures can be sometimes challenging or not applicable. This could happen for various reasons: for instance the user generated polygon could refer to a physical element that is not completely visible, or not updated, in the base map images; sometimes it could be difficult to determine which data is more accurate (the polygon from the volunteer contributor or the base image provided by the map application).

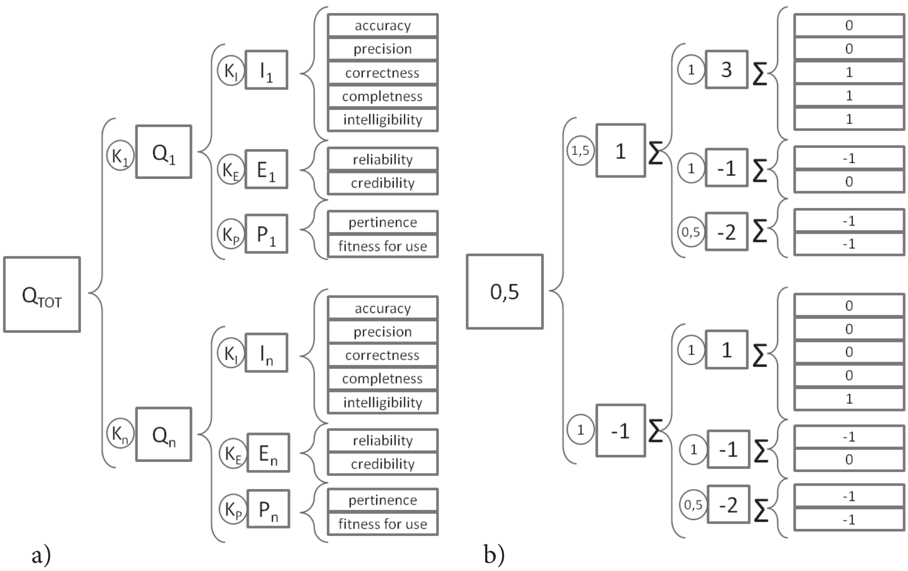

Figure 3: The theoretical model for estimating quality in a generic GI item (a) and its enactment for a plausible VGI case study (b).

The situation described above is represented in Figure 3b, as homologous to the theoretical model in Figure 3a. The numerical score resulting from the simulation (Figure 3b) is a direct consequence of the initial decision – taken by the imaginary analyst – to assign equal weights to the intrinsic and to the extrinsic components of the VGI item, and it is directly connected with the numerical domain stated ([1, 0, −1]). These choices lead to a slightly positive total score (0.5), which represents the overall quality index for the volunteered contribution. Some alternative decision – for instance to assign a higher value to the extrinsic quality weight K_E – would lead to different and even negative results. The result, whatever it would be, is meaningful only if compared with analogous ones, belonging to the same dataset, or to different datasets. The procedure indeed doesn't assess itself the quality of the VGI items, but allows normalizing the quality components in order to facilitate their comparison and evaluation. This procedure can be carried out manually, automatically or semi-automatically. The processed items could be finally ranked and possibly filtered, depending on the accomplishment of a minimum quality threshold.

Discussion and conclusion

The issue faced in this work – i.e. to flexibly represent and assess the quality in crowdsourced GI – has led to define a methodology and a set of quality indices suitable to represent and quantify quality.

A synthetic representation of quality control strategies in crowdsourced GI activities has been proposed, as a guide for project design and comparison of existing ones.

Such approaches, in practical cases, are often combined into hybrid strategies. To break them down into their atomic properties helps to describe and normalize such strategies, even in the more complex operational cases. It seems unrealistic to point out one single optimal solution. On the contrary, several effective configurations can coexist, offering suitable solutions for specific use cases. The choice is usually determined by the objective of the crowdsourced GI activity (which can be recreational, social, scientific, professional, experimental, etc.), by the type and the amount of information expected (images, annotations, features, measurements), by the characteristics of the contributors addressed (citizens, unaware web users, trained volunteers, experts, etc.), and by the infrastructure and technologies on which the project leans (geographic databases, web and mobile clients for services, sensors, etc.).

The representation introduced in Figure 2 can be used not only *a posteriori*, i.e. to describe the control strategy performed, but can also be helpful during the design phase, to configure the most effective solution for quality management.

Besides the analysis of strategies for quality control, a simple representation of the components of the quality in crowdsourced GI has been depicted too. It helps in focusing on different aspects of quality and individually evaluating them.

A flexible methodology for quality assessment has been introduced on the basis of the given representation. It can be applied manually or automatically on a wide range of volunteered contributions, and differently weighted according to the needs of the analysts.

The estimation of the quality of crowdsourced GI is a challenge that has been addressed by several authors with different methods. The methods found in literature, however, are usually designed to respond to specific needs, and therefore, as far as valuable and useful in particular cases, appear to suffer from one or more of the following major constraints:

- they aim to quantify the uncertainty of a single quality feature (e.g. the credibility of the volunteers, or the accuracy of the spatial features, etc.);
- they deal with a single category of GI (images, annotations, features, measurements);
- they are suitable only for a specific application (depending on a given technology, or presuming the participation of a certain amount of volunteers, etc.).

The contribution brought by this work is the proposal of a generalized method for estimating the quality that goes beyond these limitations. It is formally defined in terms sufficiently operational to ensure their applicability, but also general enough to ensure its transferability to different application cases.

The proposed method is based on some choices and can be designed as a decision making approach including all the major quality features and allowing the description of each type of GI contribution, under all possible aspects, using different aggregation operators to get to a final decision. Its strength lies in its flexibility: the aggregation operations can be chosen to suit the purpose of the user's analysis, and the decision-making approach allows dealing with any specific case. Even if the analyst does not wish to join to the proposed representation of quality (Figure 1), the method is still applicable, by replacing the suggested indices with alternative properties. Finally, the method can provide a guide to systematize and make explicit the criteria for assessing the quality that are used in an application of crowdsourced or even heterogeneous GI.

Acknowledgments

The present work is based on the research carried out in the framework of a PhD programme of the University of Pavia, Dept. of Earth and Environmental Sciences, and has been supported by the Institute for the Electromagnetic Sensing of the Environment CNR-IREA. The research has been partially founded by the Italian Flagship Project RITMARE.

References

Antoniou, V., Morley, J., & Haklay, M. 2010. Web 2.0 geotagged photos: Assessing the spatial dimension of the phenomenon. *Geomatica, 64*(1): 99–110.

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., & Rampini, A. 2014a. A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. *Information Sciences, 258*: 312–327.

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., & Rampini, A. 2014b (November). On predicting and improving the quality of Volunteer Geographic Information projects. *International Journal of Digital Earth*: 1–22. DOI: http://dx.doi.org/10.1080/17538947.2014.976774

Ciepłuch, B., Jacob, R., Mooney, P., and Winstanley, A. 2010. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In: *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Leicester, UK on 20–23rd July 2010: p. 337.

Coleman, D. J., Georgiadou, Y., Labonte, J., Observation, E., & Canada, N. R. 2009. Volunteered Geographic Information: The Nature and Motivation of Produsers. *International Journal of Spatial Data Infrastructures Research, 4*: 332–358.

Connors, J. P., Lei, S., & Kelly, M. 2012. Citizen Science in the Age of Neogeography: Utilizing Volunteered Geographic Information for Environmental

Monitoring. *Annals of the Association of American Geographers, 102*(6): 1267–1289.

Crall, A. W., Newman, G. J., Stohlgren, T. J., Holfelder, K. A., Graham, J., & Waller, D. M. (2011). Assessing citizen science data quality: an invasive species case study. *Conservation Letters, 4*(6): 433–442.

Criscuolo, L. 2014. Monitoring of Italian Glaciers: Official, Volunteered and Incidental Information. Unpublished Thesis (PhD), University of Pavia, Dept. of Earth and Environmental Sciences. Advisors: Zucca F., Seppi R., Pepe M.

Delaney, D. G., Sperling, C. D., Adams, C. S., & Leung, B. 2007. Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions, 10*(1): 117–128.

De Longueville, B., Luraschi, G., Smits, P., Peedell, S., & Groeve, T. D. 2010. Citizens as sensors for natural hazards: A VGI integration workflow. *Geomatica, 64*: 41–59.

De Solla, S. R., Shirose, L. J., Fernie, K. J., Barrett, G. C., Brousseau, C. S., & Bishop, C. A. 2005. Effect of sampling effort and species detectability on volunteer based anuran monitoring programs. *Biological Conservation, 121*(4): 585–594.

De Tre, G., Bronselaer, A., Matthe, T., Van de Weghe, N., & De Maeyer, P. 2010. Consistently Handling Geographical User Data Context-Dependent Detection of Co-located POIs. *Communications in Computer and Information Science*, Springer*, 81*(1): 85–94.

Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics, 41*(1): 149–172.

English, L. P. 1999. Improving data warehouse and business information quality: methods for reducing costs and increasing profits. J. Wiley & Sons.

Eysenbach, G., & Kohler, C. 2002. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal, 324*: 573–577.

Fischer, F. 2000. *Citizens, experts, and the environment: The politics of local knowledge*. Duke University Press.

Fitzpatrick, M. C., Preisser, E. L., Ellison, A. M., & Elkinton, J. S. (2009). Observer bias and the detection of low-density populations. *Ecological Applications, 19*(7): 1673–1679.

Flanagin, A. J., & Metzger, M. J. 2008. The credibility of volunteered geographic information. *GeoJournal, 72*(3–4): 137–148.

Friedland, G., & Choi, J. 2011. Semantic computing and privacy: A case study using inferred geo-location. *International Journal of Semantic Computing, 5*(1): 79–93.

Galloway, A. W. E., Tudor, M. T., & Haegen, W. M. V. 2006. The Reliability of Citizen Science: A Case Study of Oregon White Oak Stand Surveys. *Wildlife Society Bulletin, 34*(5): 1425–1429.

Genet, K. S., & Sargent, L. G. 2003. Evaluation of methods and data quality from a volunteer-based amphibian call survey. *Wildlife Society Bulletin*: 703–714.

Girres, J. F., & Touya, G. 2010. Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS, 14*(4): 435–459.

Goodchild, M. F. 2008. Spatial Accuracy 2.0. In: *Proceedings of the eighth international symposium on spatial accuracy assessment in natural resources and environmental sciences*, Shanghai: pp. 1–7.

Haklay, M. 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning. B, Planning & design, 37*(4): 682.

Huang, K. L., Kanhere, S. S., & Hu, W. 2010. Are You Contributing Trustworthy Data? The Case for a Reputation System in Participatory Sensing. In: *MSWIM '10 Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*. ACM New York, NY, USA c2010, pp. 14–22.

Keßler, C., Theodore, R., & De Groot, A. 2013. Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap. In: Vandenbroucke, D., Bucher, B., & Crompvoets, J. (Eds.) *Geographic Information Science at the Heart of Europe, Lecture Notes in Geoinformation and Cartography*. Springer International Publishing, Cham: pp. 21–37.

Kuhn, W. 2001. Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science, 15*(7): 613–631.

Latonero, M., and Shklovski, I 2010 "Respectfully Yours in Safety and Service": Emergency Management & Social Media Evangelism. In: *Proceedings of the 7th International ISCRAM Conference*. Seattle, USA: pp. 1–10.

Lotz, A., & Allen, C. R. 2007. Observer bias in anuran call surveys. *The Journal of wildlife management, 71*(2): 675–679.

Metzger, M. J. 2007. Making sense of credibility on the Web: Models for evaluating online information and recommendation for future research. *Journal of the American Society for Information Science and Technology,* 13: 2078–2091.

Popescu, A, Grefenstette, G, and Bouamor, H 2009 Mining a Multilingual Geographical Gazetteer from the Web. In: *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Ieee*: 58–65.

Pundt, H. 2002. Field Data Collection with Mobile GIS: Dependencies Between Semantics and Data Quality. *GeoInformatica, 6*: 363–380.

Sabone, B. 2009. *Assessing Alternative Technologies for Use of Volunteered Geographic Information in Authoritative Databases*. Technical Report, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Canada.

Spinsanti, L., & Ostermann, F. 2013. Automated geographic context analysis for volunteered information. *Applied Geography, 43*: 36–44.

Van Dijck, J. 2009. Users like you? Theorizing agency in user-generated content. *Media, culture, and society, 31*(1): 41–58.

Van Exel, M., Dias, E., & Fruijtier, S. 2010. The impact of crowdsourcing on spatial data quality indicators. In: *Proceedings of the 6th GIScience international conference on geographic information science, 2010*: pp. 213–217.

Van Oort, P. A. J. 2006. Spatial Data Quality: From Description to Application, Thesis (PhD), Wageningen Universiteit.

Sui, D., Elwood, S., & Goodchild, M. 2013. Crowdsourcing geographic knowledge, Volunteered Geographic Information in Theory and Practice, Springer Verlag.