

A Conceptual Quality Framework for Volunteered Geographic Information

Andrea Ballatore¹ and Alexander Zipf²

¹ Center for Spatial Studies
University of California, Santa Barbara
aballatore@spatial.ucsb.edu

² Geoinformatics Research Group, Department of Geography
University of Heidelberg, Germany
zipf@uni-heidelberg.de

Abstract. The assessment of the quality of volunteered geographic information (VGI) is cornerstone to understand the fitness for purpose of datasets in many application domains. While most analyses focus on geometric and positional quality, only sporadic attention has been devoted to the interpretation of the data, i.e., the communication process through which consumers try to reconstruct the meaning of information intended by its producers. Interpretability is a notoriously ephemeral, culturally rooted, and context-dependent property of the data that concerns the *conceptual quality* of the vocabularies, schemas, ontologies, and documentation used to describe and annotate the geographic features of interest. To operationalize conceptual quality in VGI, we propose a multi-faceted framework that includes accuracy, granularity, completeness, consistency, compliance, and richness, proposing proxy measures for each dimension. The application of the framework is illustrated in a case study on a European sample of OpenStreetMap, focused specifically on conceptual compliance.

Keywords: data quality, interpretability, conceptual quality, volunteered geographic information

1 Introduction

The importance of data quality has been noted since the disciplinary inception of geographic information science (GIScience) [13]. The quality of geographic information has been framed along the spatial, temporal, and thematic dimension, in terms of accuracy, precision (or resolution), consistency, and completeness [32]. Because any discussion on data quality assumes the presence of a producer who encodes some information and a consumer who has to interpret it and use it, the *conceptual quality* of the data is crucial to enable the semantic decoding of data. For example, a dataset can contain highly accurate geometries, but if the description of the entities and their attributes is not clear, articulate, rich, and complete enough, the value of the data for consumers will be severely curtailed.

In past decades, the assurance of conceptual quality was facilitated by the fact that both producers and consumers tended to belong to professional circles, and shared to some degree a semantic ground, i.e., the conceptualization of the domain and its entities, and a common vocabulary to describe them. The advent of volunteered geographic information (VGI), with its less centralized production models, leads to a novel state of affairs, with important consequences for conceptual quality. In VGI, different actors generate data for a variety of purposes, interpreting and consuming data produced from other actors. These processes of informal and loosely constrained *prosumption* [10] usually result in data with higher heterogeneity and fragmentation than traditional datasets, creating new, unforeseen barriers to data interpretation. Despite the growth of interest in VGI in both academia and industry, recurring issues make the application of traditional approaches to data quality problematic. For instance, a crowdsourced set of points of interest might possess sufficient quality to enrich spatial social media, but could fail to capture the changes in businesses in rural areas studied by economists.

The problem of quality assessment is intimately linked to the quantification of several orthogonal or correlated dimensions. In fact, it is impossible to state anything about data quality without well-defined criteria to measure it. To date, many researchers have tackled the issue of quality in VGI [19, 20, 14, 23, 5]. Unlike the expert-controlled data generated by government agencies, crowdsourced resources do not come with systematic documentation about their production protocols, biases, and shortcomings [11]. In this sense, the context of VGI is open to many questions concerning conceptual quality, some of which are also relevant to traditional datasets, while others are novel and specific. In VGI, the data is often characterized by loose application of standards and a lack of thorough documentation. The traditional distinction between schema and records in databases does not always hold. Contributors produce data and then re-define and update its schema in an open process, which leads to uneven conceptualizations. In this sense, another peculiar difficulty lies in the associations between classes and instances—or, alternatively, universals and individuals.

How, then, is it possible to operationalize the conceptual quality of VGI, taking into account semantic aspects in the data that are so central to its production? To answer this question, it is important to take an ecological viewpoint on the environments in which VGI producers and consumers operate. Rather than designing the data production as a deterministic process with clear inputs and outputs, in VGI diverse actors use a combination of natural language, data sources, schemas, vocabularies, and software tools to generate the data they are interested in, through many feedback loops. In a semiotic sense, contributors need to develop shared conceptualizations that constrain the intended meaning of the symbols to enable the interpretation of the data, enabling the data as a medium for communication [22]. While the centrality of documentation and semantics is often acknowledged by researchers [e.g., 2], conceptual quality has not yet been reduced to tractable measures and deployed within the relevant communities. Effective conceptual quality assessment techniques would benefit

project owners, contributors, as well as end users, informing the production, evaluation, and consumption of data sources.

To fill this gap, in this article we present a conceptual quality framework that can be adapted and applied to any VGI source, tapping indirect and intrinsic proxy measures. The framework moves a first step towards the operationalization of the following difficult questions: Is the schema appropriate to describe the domain? How current and clear is the documentation? Is there consensus and consistency in the terms used in the data? Is the data prone to foster divergent interpretations? Are the descriptions of classes sufficiently detailed? What are missing elements that need to be described? To what degree are the users interpreting the terms correctly? How intuitive are the terms for the users? Is the data internally coherent? What geographic areas present differences in conceptual quality and why? Does the data conform to an external reference or standard? Is there conformance within specific groups of contributors or within geographic areas?

The remainder of this article is organized as follows. Section 2 surveys the literature on geographic information quality, focusing particularly on VGI. Section 3 illustrates the core ideas of our framework and proposes formal measures to operationalize it. Subsequently, Section 4 reports on a case study in which we illustrate the applicability of the framework on OpenStreetMap, a prominent example of VGI. Finally, Section 5 draws conclusions and indicates directions for future research.

2 Related work

This quality framework for VGI lies at the intersection of several research areas, including GIScience, conceptual modeling, and ontology engineering. This section provides an overview of the notions of quality discussed in these inter-related disciplines.

Geographic information quality. Because all geographic information is produced through measurement with some level of uncertainty, the debate on quality has been central to geography and GIScience for a long time [17, 27, 21]. As pointed out by Goodchild and Li [14], broad consensus was established in the 1980s along five dimensions: positional accuracy, attribute accuracy, logical consistency, completeness, and lineage (p. 111), embedded in the US Spatial Data Transfer Standard (SDTS).³ A pioneering theoretical discussion about the semantic dimension of geographic information quality was provided by Salgé [26]. Assuming that any description of reality is inevitably a reduction to a model, Salgé defined semantic accuracy as the “quality with which geographical objects are described in accordance with the selected model,” as well as the “pertinence of the meaning of the geographical object rather than to the geometrical representation” (p. 139).

In practical terms, producers can enforce minimum quality standards in their data collection process, and consumers can assess the quality of a dataset through

³ <http://mcmcweb.er.usgs.gov/sdts>

metadata standards [32]. Improvements in web technologies have tightened the feedback loop between consumers and producers, providing mechanisms to improve quality in a targeted way based on users' needs. In the early 2000s, aspects of quality were defined in the International Organisation for Standards (ISO) 19113:2002 for quality principles, and in ISO 19114:2003 for quality evaluation procedures, then superseded by ISO 19157:2013.⁴ While progress in theorization and standardization of data quality has been made, particularly in the context of public agencies, many challenges remain to be met. As Hunter et al. [21] pointed out, the communication, visualization, and description of data quality and its application to decision making are far from having satisfactory solutions for the many actors involved.

Conceptual and ontological quality. Conceptual modeling and ontology engineering are concerned with the quality of models, schemas, and ontologies [4]. The operationalization of such conceptual dimension of quality offers an important tool to facilitate the adoption, correct interpretation, and re-use of conceptualizations by practitioners. Agent-based approaches have been proposed to model the correctness of spatial information [12]. In recent years, applied ontologists have designed formal semantic approaches to assess the quality of an ontology, epitomized by the OntoClean method [15]. Other approaches are grounded in semiotics: Tartir et al. [30] outlined a triangular model where quality can be assessed in the mappings between the real world and the schema, between the real world and the data, and between the schema and data. In their formulation, metrics for schema quality include dimensions such as relationship, attribute, and inheritance richness, while instance metrics should reflect connectivity, cohesion, and readability of the data.

Along similar lines, Burton-Jones et al. [8] defined ontological quality from four facets: syntactic quality (richness of lexicon and correctness); semantic quality (interpretability, consistency and clarity); pragmatic quality (comprehensiveness, accuracy and relevance); and social quality (authoritativeness and history). An overall indicator of quality is obtained with a linear combination of these four dimensions. In the context of conceptual modeling, Cherfi et al. [9] defined a framework for conceptual quality, outlining metrics applicable to entity-relationship schemas, and to UML diagrams. While these methods inform the foundations of our framework, they are hard to apply directly to the semantically weak folksonomies and tagging models used in VGI [28, 31].

Quality in VGI and OpenStreetMap. The emergence of VGI deeply re-configured the geographic information landscape, raising immediate concerns about quality assurance [20]. Analogously to Wikipedia, crowdsourced geographic information can be often of higher quality than authoritative sources, but shows considerable spatial and temporal variability, and is affected by gender and socio-economic contribution inequalities [29]. Vandalism in open mapping platforms has also been identified as a multi-faceted, complex challenge for VGI [1]. More generally, VGI communities have several strategies to ensure data quality [14]. These

⁴ <http://www.iso.org>

include crowdsourcing, based on the assumption that more people working on the same area will tend to result in higher quality [19], social approaches that rely on surveillance and control, and geographic approaches that exploit knowledge from geography to detect unlikely or impossible configurations in the data. The quality of OpenStreetMap (OSM), one of the leading VGI projects, has been studied along different lines of investigation. Several studies compared a sample from the OSM vector dataset against the corresponding data from more traditional and authoritative sources [18, 33], showing high variability in the data quality, and identifying several geographical divides, particularly between rural and urban areas, and natural and man-made features.

The conceptual and semantic dimension of the data presents many specific challenges to ensuring quality. OSM has a lightweight semantic model that relies primarily on user-defined tags [2]. While the positional accuracy of features can be measured with standardized methods, the annotation process has no stable ground truth, as it is rooted in alternative conceptualizations of geographic world. Problems identified in OSM semantic set up are the flexibility of the tagging process and the lack of a strict mechanism for checking semantic compliance, even for core elements of the data, often resulting in tag wars [24, 23]. In the project’s forum and mailing lists, contributors often debate data quality, pointing out a strong need for “consistency in tagging, editor improvements, better documentation, better training materials.”⁵

To date, the most substantial attempt at quantifying quality in OSM has been carried out by Barron et al. [5]. Their *iOSMANalyzer* tool generates a range of intrinsic quality indicators, focusing on the spatio-temporal evolution of the data, including geometric and thematic quality. The framework adopts a fitness-for-purpose perspective, grouping indicators by application area, such as geocoding, routing, and point of interest (POI) search. Despite its comprehensiveness, this framework is narrowly focused on OSM, and could not be easily applied to other datasets. Moreover, the semantic aspects of the vector data are discussed only tangentially. Our proposed quality framework, outlined in the next section, aims at overcoming these limitations.

3 A framework for VGI conceptual quality

To establish a framework to operationalize the conceptual quality of VGI, we analyze and revise each dimension of data quality, comparing it with traditional views on geographic information quality and proposing indirect indicators. In VGI, heterogeneous communities produce information for a variety of purposes, relying on a combination of tools, vocabularies, and data sources [3]. The main purpose of this framework is to enable the measurement of conceptual quality of a VGI dataset. Understanding what information producers meant to express in their data is a crucial, and yet ephemeral aspect of spatial information quality [22]. In traditional database theory, the quality of a database includes the quality

⁵ <http://goo.gl/W0rmVU>

of its conceptual schema, metadata description, and provenance of data [6]. Similarly, in VGI, *conceptual quality* should answer questions about the conceptual schema and its relations with the data. Conceptual quality is intimately intertwined with *interpretability*, the fundamental communication problem between the data creators and consumers.

In the context of VGI, we frame the production and consumption of information semiotically as the interaction of semantic agents in an information community. Hence, we refer to symbols (e.g., words, icons, or images) pointing to concepts, psychological models used by semantic agents to produce and interpret information about domain entities, called referents. The mappings between symbols and concepts is dynamic, and are established through social agreements [3]. To clarify our notion of interpretability, we distinguish between interpretability of the *data format* (i.e., file formats, formal languages, conceptual schemas), and the interpretability of the *domain content* (i.e., the concepts prior to their encoding into data).

The semantic gulf. The role of conceptual quality is essential to overcome the semantic gulf that exists between producers and consumers. As shown in Figure 1, agent *A* describes a concept in his/her worldview with the symbol ‘mountain’ and encodes it into a dataset *D*. Because of cultural, linguistic, and individual variations, when agent *B* interprets and decodes the symbol *S*, his/her interpretation (Θ_B) overlaps with that of agent *A* (Θ_A) only to some degree. Following the notion of intended models by Guarino et al. [16], we refer to the overlap between the interpretations of symbol *S* for the two users ($\Theta_A \cap \Theta_B$). This overlap is an indicator of the quality of *D* in the sense that the highest quality would result in equivalent interpretations ($\Theta_A \equiv \Theta_B$). By contrast, the lowest quality leads to totally different interpretations of *D* ($\Theta_A \cap \Theta_B = \emptyset$). As low conceptual quality causes friction in the interpretation process, the operationalization of conceptual quality is essential to improve the interpretability of the data.

Conceptual quality questions include: how can the meaning of a term be assessed? To what degree is it possible to reconstruct the context and intentions of the producers from the data? How many alternative interpretations exist for a term? Does the compliance to external resources help the interpretation of the data? Is the usage of a term widespread or is it unusual? How easily can a consumer decode the terms? How clear are the constraints on terms? How ambiguous are the terms? To what degree are the agents interpreting the symbols correctly? How intuitive are the symbols for the agents? Are different agents mapping the symbols to different concepts? Are the symbols internally coherent?

Measuring conceptual quality. The measurement of conceptual quality can be carried out through some form of psychological testing in a controlled environment, asking human subjects to perform tasks on a piece of information, and measuring the cognitive load and other observable outcomes of the interpretation process. Less formally, ratings about the quality of resources can be collected

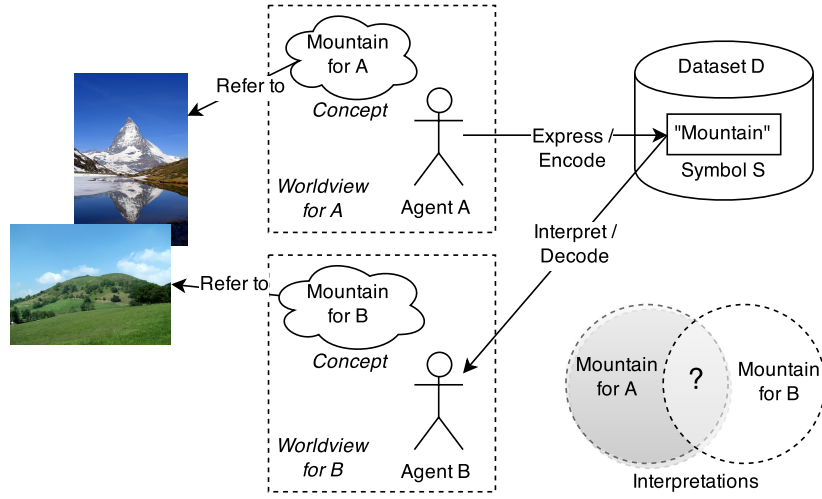


Fig. 1. The semantic gulf

directly from users of an online platform, identifying issues in the conceptualization. However, these approaches are impractical for large, decentralized projects such as OSM. On the other hand, measures of conceptual quality can rely on a number of indirect indicators, used as proxies to unobservable variables. For example, in ontology engineering, a measure of interpretability has been designed on the number of terms defined in external linguistic resources such as WordNet [8]. This extrinsic approach relies on the interlinking of a resource, assuming that connected resources are easier to interpret than isolated ones.

When measuring the conceptual quality of domain content, the documentation of a given term is crucially important. Indicators for this dimension of quality include the number of users who contributed to the definition, the stability of the definition over time, and the amount of discussions generated about it. Measures of VGI interpretability should be applied not only to schemas, as done traditionally in information systems [6], but also to the data itself, which might differ considerably from the documentation in local contexts. From the perspective of interpretability of data formats, different versions of the same piece of information can be evaluated, comparing traditional GIS formats like Esri's shapefile, and semantically richer formats such as RDF. A complementary issue lies in the interpretability of information about quality, which suffers from lack of standardization and from technical complexity [21].

Formalization and symbols. To formalize the core ideas in the framework, we adopt the following terminology, which we believe can describe the bulk of VGI. A geographic feature τ is an instance of a class C , and has a set A of attributes a . Attributes have values. Feature τ also has a geometric attribute g . Features can be aggregated in a set F . Our framework operationalizes conceptual quality

Dimension	Sym.	Purpose & indicators
Accuracy	I_{ac}	Distance between conceptualization and domain knowledge. It can be seen as the degree of correctness in the classification of features τ into classes C . <i>Indicators</i> : Number of features with multiple classifications; number of contributors.
Granularity	I_{gr}	Level of thematic description present in the data, moving from very abstract to very specific concepts. <i>Indicators</i> : Depth of classes in the class hierarchy (if applicable).
Completeness	I_{cl}	Coverage in the conceptualization of the features of interest. A distinction exists between class completeness and attribute completeness. <i>Indicators</i> : Number of classes; number of attributes
Consistency	I_{cn}	Degree of homogeneity in the descriptions of geographic features. <i>Indicators</i> : Number of features in a class described with the same attributes; ratio between consistent features or attributes to all others, weighted against the absolute number of attributes.
Compliance	I_{cm}	Degree of adherence of an attribute, a feature, or a set of features to a given source S , ranging from non compliance to full compliance. <i>Indicators</i> : Ratio between the number of classes and attributes defined in an external source S and the total number of classes and attributes.
Richness	I_{ri}	Amount and variety of dimensions that are included in the description of the real-world entity. <i>Indicators</i> : Number of attributes describing a feature.

Table 1. Dimensions of conceptual quality for VGI

with indicators either at the feature level $I(\tau)$, on a single τ , or at the aggregate level on a set of features $I(F)$.

For example, Lake Tahoe τ belongs to class C_{lake} , and has an attribute a_{name} whose value is set to “Lake Tahoe,” and a footprint g that is represented as a polygon. A crucial difference between VGI and the traditional geo-database approach lies in the flexibility and instability of the schema definition. Rather than a clear distinction between schema and records, VGI communities produce datasets and their schemas in an open-ended way. Classes, instances, and their attributes tend to be fluid and mutable, rather than centrally defined and controlled. The remainder of this section discusses the notion of interpretability, followed by several complementary dimensions that need to be considered, summarized in Table 1.

3.1 Conceptual accuracy

The notion of accuracy is central to the definition of geographic information quality. Accuracy answers questions about the correctness of the information with respect to a measurable phenomenon in the real world, for which there is a true value can that in principle be assessed. Accuracy is perhaps the best under-

stood dimension of quality [22]. Because of the strong spatiality of geographic data, positional accuracy is a core preoccupation, complemented by temporal accuracy (also known as currency) [27].

More relevant for conceptual accuracy, notions of thematic or attributional accuracy indicate the degree of correctness of attribute values grounded in a spatio-temporal region, typically in the context of classification, for example indicating an area as industrial where it is in fact residential. Thematic information relies on a conceptualization that defines salient domain entities, categories, and their attributes. In this sense, conceptual accuracy concerns the distance between the concepts and the real world entities that they are supposed to describe according to an observer. Low conceptual accuracy indicates that the instances encountered by contributors are not intuitively or easily described with the selected conceptualization, resulting in semantic noise.

The measurement of spatio-temporal accuracy relies on the assumption that a true value can be obtained at a higher accuracy using appropriate measurement techniques. This assumption cannot always be performed in thematic information, and can clash with the multi-authored, choral nature of VGI. Conceptual accuracy should answer several questions: To what degree are the classes and attributes capturing the underlying domain knowledge? In the case of categorical variables, are the categories reflecting the domain knowledge? Are there many observations that do not fit the categories? How good is the agreement on the classification of instances when performed by different actors?

For conceptual accuracy, the heterogeneity of VGI presents new and peculiar challenges. Contributors describe the objects of interest using loosely defined vocabularies that present high lexical and semantic variation [2]. The inconsistencies in the attributional data makes the measurement of the global conceptual accuracy very hard, prompting, again, local measures. New measures of accuracy for VGI should include a social dimension that plays a huge role in the production process. Intrinsic, local measures include that by Haklay et al. [19], who suggested that the number of active users in an area shows a non-linear relationship with positional accuracy. Along similar lines, Bishr and Kuhn [7] tapped the social dimension in VGI by using trust as a proxy measure of quality.

Given a set of classes C , we define conceptual accuracy I_{ac} as the degree of correctness in the classification of features τ into classes C . Although such an assessment of classification accuracy needs some form of extrinsic ground truth (i.e., a classification having higher accuracy), it is possible to devise indirect indicators of conceptual accuracy I_{ac} . The core impact of conceptual accuracy occurs in the definition of the schema and the application of the schema on the instances. When the classification of a feature is difficult, contributors tend to classify it in multiple, incompatible ways [23]. Hence, one indicator consists of the number of features that have been classified in different ways at different points in time:

$$I_{ac}(F) = 1 - \frac{|\exists \tau \in F : \tau \in C_1 \wedge \tau \in C_2|}{|F|}; \quad I_{ac} \in [0, 1] \quad (1)$$

High values of I_{ac} indicate low level of negotiation in the classification of features. While high $I_{ac}(F)$ might indicate that too few people worked on a classification to evaluate its quality, high $I_{ac}(F)$ signals that the contributors did not encounter problems in the classification of F .

3.2 Conceptual granularity

While accuracy generally concerns the distance between a measurement and the true value, granularity answers questions about the precision of information, i.e., the repeatability of measurements, regardless of their true value [22]. Accuracy and granularity are orthogonal: a piece of information can present high accuracy and low granularity, and vice-versa. The term resolution is a synonym of granularity. The notion of scale is indeed related to granularity, as different scales require higher or lower granularity. Geographic information quality standards prominently include granularity as a fundamental element to evaluate fitness for purpose. For example, satellite imagery can be described as having “10m resolution.”

In VGI, the heterogeneity in the production process results in varying granularity. As observed for accuracy, the assessment of granularity loses meaning when performed at the global level. Spatial granularity in VGI is bound by the technical apparatus available to mappers (e.g., GPS sensors), and by the pre-existing geospatial infrastructure, such as the quality of satellite imagery that OSM contributors rely on to draw roads [18]. While the notion of granularity is well understood at the spatial and temporal level, thematic information presents deeper challenges. The notion of thematic resolution has been defined as the precision of the scalar or nominal variables [32].

In an open process of negotiation, VGI contributors express quantitative and qualitative measurements about a wide range of phenomena, usually based on a loosely defined conceptualization. For this reason, the conceptual structure of information is rather fluid, and its granularity is hard to assess. As in the case of OSM, contributors define hierarchies of classes C and their attributes to describe the concepts of interest, such as *university*, *park*, and *river*. Using the categorization by Rosch [25], such taxonomical hierarchies span from the superordinate level (e.g., built environment), to the basic level (e.g., house), and to the subordinate level, which includes more specific concepts, rarely used in day-to-day language (e.g., detached single-unit house).

Given a VGI dataset, *conceptual granularity* should answer questions about the level of thematic description is present in the data, moving from very abstract to very specific concepts. Are the objects described simply as buildings or are they categorized in sub-types? Hence, for a feature τ in class C , we want to devise a measure that quantifies the thematic granularity in the data. Among all classes defined in the data, how specific is C ? To achieve this goal, measure of level of generality of a term in a hierarchy. This approach is meaningful only if the classes are organized in a subsumption hierarchy, which is not always the

case. An indicator is the depth of a class C in the class hierarchy, for features τ :

$$I_{gr}(F) = \sum_{i=1}^{|F|} \text{depth}(C) : \tau_i \in C \wedge \tau_i \in F; \quad I_{gr} \in [0, \infty) \quad (2)$$

As concepts can be organized in alternative ways in a taxonomy, caution is needed when comparing different datasets that adopt radically different approaches (such as OSM and the GeoNames gazetteer). The amount of details included in the description is captured by conceptual richness (see Section 3.6).

3.3 Conceptual completeness

Geographic data can be evaluated in terms of the coverage of the entities of interest in the real world. Given some mapping rules, completeness answers questions about how many objects are included or missing from the dataset. Completeness can be measured spatially (is the target space surveyed in its entirety?), temporally (how well is the target space covered at a given time?), and thematically (are all relevant types of features included?) [32]. To be assessed, completeness needs an external reference that can be used as ground truth, and for this reason its measurement tends to be extrinsic.

Completeness in VGI is challenging as the mapping rules are either loosely defined or left implicit. As in traditional datasets, VGI completeness can be assessed extrinsically, using higher-quality data as ground truth [18]. In many instances, such as in the case of disaster management, the ground truth does not exist in the first place, and extrinsic measures might not be applicable. Therefore, intrinsic measures appear as particularly valuable. For example, Barron et al. [5] suggest that, if the growth of additions to the dataset is slowing down for a given feature type, in spite of general growth, that might indicate high completeness.

Depending on the degree of openness and structure of a data collection procedure, contributors decide what they want to include in the data from the potentially infinite knowledge about a geographic area. For this reason, *conceptual completeness* concerns the coverage in the conceptualization of the features of interest. Conceptual completeness can be further specialized in *class completeness* (e.g., how many building types are present in the dataset) and *attribute completeness* (e.g., how many streets have a name attribute).

To support the measurement of conceptual completeness in VGI, intrinsic measures can use various social and semantic signals as indirect indicators of completeness. Absolute and global completeness measures are doomed to be not very meaningful for VGI, because of the non-parametric distribution of features in the geographic space. By contrast, local and relative measures of completeness should answer questions about completeness with respect to a given type of features and attribute by comparing spatial, temporal, or thematic subsets. As simple indicators I_{cl} , we adopt the number of classes and the number of attributes in a set of features:

$$\begin{aligned} I_{cl\ C}(F) &= |C : \tau \in F \wedge \tau \in C|; \\ I_{cl\ A}(F) &= |A : a \in \tau \wedge \tau \in F|; \quad I_{cl} \in [0, \infty) \end{aligned} \quad (3)$$

Geo-statistical approaches can support the formulation of intrinsic measures of conceptual completeness based on these indicators. The automatic detection of missing attributes is a proxy to attributional completeness, rooted in the distribution of attributes over space, highlighting statistically anomalous regions. A similar approach can be applied to class completeness, exploiting geographic knowledge about an area to identify areas with unusually low number of classes being instantiated. From a social dimension, collective or individual activity patterns cannot be reliable indicators, but they might provide a crude proxy indicators to conceptual completeness.

3.4 Conceptual consistency

For each piece of geographic information, many alternative representations are possible. *Conceptual consistency* answers questions on the degree of homogeneity in the descriptions of geographic features. Are a set of features described with the same classes and attributes? Are synonyms used in the data? Are there multiple names for the same features? Are there individual and regional variations in the usage of terms or concepts? While in formal systems consistency usually refers to logical contradictions, VGI rarely relies on highly formal languages, favoring simple vocabularies, folksonomies, or tagging mechanisms.

Measures of consistency can focus on the use of classes and attributes in the data, with the advantage that no knowledge about the conceptual schema is needed. Given a set of features, pair-wise comparison can be used to identify clusters of features described similarly, both within the same spatial unit and between different spatial units. The ratio between consistent features or attributes to all others, weighted against the absolute number of attributes, can be used as a simple indicator of consistency, applicable at different granularities. As consistency is an intrinsic characteristic of the data, it is possible to devise an indicator I_{cn} based on a feature set F in class C containing attributes A :

$$I_{cn}(F, C) = \frac{|\forall(A_{\tau_i}, A_{\tau_j}) : A_{\tau_i} \equiv A_{\tau_j} \wedge \tau_i, \tau_j \in C|}{|\tau : \tau \in C \wedge \tau \in F|}; \quad I_{cn} \in [0, 1] \quad (4)$$

High (low) values of I_{cn} indicate that the description of class C tends to be (in)consistent. This measure captures the homogeneity of attributes across different features. Effective measures of consistency are useful to identify communities that adopt different representational conventions and terms, going beyond the global binary classification as correct or incorrect. The measures of consistency are also useful to analyze consistency over time, and not only in space, detecting regional trends.

3.5 Conceptual compliance

Compliance can be seen as an orthogonal dimension to consistency. Unlike consistency, compliance is extrinsic, as it refers to an external resource, such as

documentation, meta-data, standards, or guidelines. *Conceptual compliance* answers questions about the degree of adherence of an attribute, a feature, or a set of features to a given source S , ranging from non-compliance to full compliance. In VGI, contributors rely on a combination of sources to produce the data, intrinsically (using resources defined within the same project), and extrinsically (adopting external sources). Measuring conceptual compliance I_{cm} would increase the homogeneity of data, facilitating its interpretability. The quality of these resources S , such as the readability and completeness of the documentation, is out of the scope of conceptual compliance.

VGI projects define formats, schemas, vocabularies, and conventions to be used in the data, usually indicating a hierarchy of reference sources. For instance, OSM indicates its wiki website as the most authoritative source of documentation, and other sources such as the map editors as less reliable and possibly non-compliant. These pieces of documentation indicate at different levels of detail how to describe buildings, what spatial and temporal reference systems should be adopted, how street addresses should be encoded, etc. Indicators of conceptual compliance I_{cm} can be applied to attributes $a \in A$, to features τ , or to sets of features F , with respect to a given source S , such as a conceptual schema:

$$I_{cm}(A, S) = \frac{|A \in S|}{|A|}; \quad I_{cm}(F, S) = \frac{|\tau : \tau \in F \wedge \tau \in S|}{|F|}; \quad I_{cm} \in [0, 1] \quad (5)$$

These indicators enables the measurement of conceptual compliance, distinguishing it from conceptual consistency. A set of features F can be consistent and non compliant, and vice-versa. In projects like OSM, the detection of consistent and non compliant subset of the data can also help contributors identify suspect deficiencies in the documentation S .

3.6 Conceptual richness

A facet that is rarely mentioned in current quality frameworks concerns the richness of the data. By *conceptual richness*, we mean the amount and variety of dimensions that are included in the description of the real-world entity. For example, a building can be described as a simple point or footprint in space, and this description can be enriched by a unbounded set of observations about its architecture, usage, materials, infrastructure, ownership, functions, history, etc. A measure of richness I_{ri} therefore needs to quantify the dimensions of a feature τ , enabling comparison with other features (or sets of features).

Richness can concern either the conceptual schema or the data, bearing in mind that in VGI the alignment between classes and instances cannot be taken for granted. This facet of conceptual quality is orthogonal to conceptual completeness, in the sense that a dataset can possess high richness but low completeness, and vice-versa. To measure richness of the conceptual schema, we can rely on number of classes and attributes defined in a dataset. At the feature level, richness $I_{ri}(\tau)$ can be quantified as the number of attributes. The richness of a set of features F can be computed as the mean of number of attributes defined

in the features:

$$I_{ri}(\tau) = |\{a \in \tau\}|; \quad I_{ri}(C) = \sum_{n=1}^{|C|} |\{a \in C_n\}|; \quad I_{ri}(F) = \frac{\sum_{n=1}^{|F|} I_{ri}(\tau)}{|F|} \quad (6)$$

These measures enable the comparison of different datasets and regions with respect to their richness, highlighting disparities in the data as well as in the conceptualization. However, the measurement of richness faces many challenges. In heterogeneous datasets, different attributes can be describing the same dimensions inconsistently, making it challenging to distinguish between emergent richness and noise. Moreover, the assumption that a higher number of attributes leads to better conceptual quality does not always hold true, for example in the case of machine-generated default attributes in OSM. In such cases, measures of information content might be helpful.

4 A case study on conceptual compliance

To illustrate our quality framework, we choose the measurement of conceptual compliance on real crowdsourced data as a case study. As a data source, we selected OpenStreetMap (OSM), the collaborative mapping project. For reasons of space, we focus on conceptual compliance I_{cm} , one of the most critical dimensions for OSM, leaving a more thorough and comprehensive evaluation of the approach as future work. In OSM, geographic features are encoded in the form of vector data, with geometries (points, polylines, and polygons) described with attributes called tags (e.g., *place=city*, *name=Berlin*).

The intended meaning of the attributes are documented in the OSM Wiki website,⁶ which hosts the definitions of the intended meaning and usage of tags. As OSM contains a wide range of feature types, we restrict the analysis to road-related features, described with the *highway* tag. The rationale for this choice lies in the centrality of the road network in the project: producers and consumers alike are particularly concerned about its quality for routing applications, where conceptual compliance is particularly important.

OSM contributors choose the attributes to describe a feature based on a number of compliance sources S . The official documentation is hosted on the OSM Wiki website, but the map editing tools, such as JOSM and iD,⁷ are particularly central to the tagging process. Hence, our case study aims at answering the following questions: How compliant is the road network with the attributes defined in the OSM Wiki website? What is the compliance of data with respect to the most popular map editing tools? What is the spatial variation in conceptual compliance?

Selection of regions. To explore conceptual compliance, we identified a sample of areas in Germany and the UK, which are expected to present geographic and

⁶ <http://wiki.openstreetmap.org>

⁷ http://wiki.openstreetmap.org/wiki/Comparison_of_editors

Region	Pop. (M)	Area (km ² K)	Highway objects	Wiki I_{cm}	JSOM I_{cm}	iD I_{cm}
<i>Germany</i>						
Upper Bavaria	4.47	17.5	59,531	0.92	0.96	0.88
Mecklenburg-West Pomerania	1.61	23.2	40,775	0.96	0.92	0.88
<i>England</i>						
South East	8.64	19.1	13,796	0.92	0.9	0.92
North East	2.59	8.6	831	0.81	1	1
<i>Average</i>	—	—	—	0.9	0.96	0.92

Table 2. The European regions included in the study, with estimated population, area (source: *freebase.com*), OSM highway objects, conceptual compliance I_{cm} for three sources, and averages.

cultural variability in the European context. A densely populated and highly developed region was selected for each country (respectively Upper Bavaria and the South East region of England), contrasted with regions characterized by relatively low population density and economic development (Mecklenburg-West Pomerania and the North East of England). Because the size of administrative units varies considerably between these countries in the Nomenclature of Territorial Units for Statistics (NUTS), we selected regions from the NUTS2 for Germany (*Regierungsbezirk*) and regions from NUTS1 for England, resulting in comparable units, summarized in Table 4. These regions provide a small sample of European OSM data with respect to population, size, and culture. The OSM data was downloaded in January 2015.

In this study, we restricted the analysis to the objects tagged with at least one *highway* tag. These tags are used to describe not only highways, as the name would suggest, but all road-related information, which is of particular importance to the OSM community and users.⁸ The selected regions are summarized in Table 4, including an estimate of the current population, area, and their total number of highway objects in OSM.

Computation of conceptual compliance. After having extracted the OSM data, we calculated the conceptual compliance I_{cm} as defined in Section 3.5. The indicator was computed at the aggregated level on the regions, as well as on a 10 km grid, in order to be able to observe the spatial variation at a higher granularity. As compliance sources S , we included the OSM Wiki website, and the two most popular editors that have a set of predefined tags (*JOSM* and *iD*). For each source, we consider compliant a tag that is explicitly defined and documented, and non-compliant all the others. A distinction was made between keys that should only accept a set of values (*highway=residential*, *highway=primary*, etc.), and open-value tags that accept any value (*name=**, *ref=**), relaxing the compliance definition for the latter cases.

⁸ <http://wiki.openstreetmap.org/wiki/Key:highway>

Results and discussion. The conceptual compliance for OSM Wiki website and the two editors for each of the four regions is displayed in Table 4. The conceptual compliance ranges from 0.81 to 1, for an average of 0.93, indicating that 7% of tags are not compliant, and their interpretation is problematic. The compliance average for the Wiki (0.9) is lower than for editors JOSM (0.96) and iD (0.92), confirming the misalignment between the different sources of compliance that OSM users complain about (see Section 2).

Non-compliant tags include for example *highway=no* and *highway:historic=primary*. Some tags appear to be deprecated (e.g., *highway=byway*), and their status with respect to compliance is hard to assess. The conceptual compliance was then calculated on a 10 km² grid. Figure 2 shows choropleth maps of the four selected regions, as well as the locations of the highway objects, highlighting the high spatial variability of conceptual compliance. This simple measure already enables contributors and users to quantify the amount of non-compliant tags, and localize them spatially. This information can be used for fitness-for-purpose by consumers, and for quality assurance by producers. Moreover, the incongruities between the OSM Wiki website and the map editors can be identified and resolved systematically with our approach.

5 Conclusions

In this paper, we outlined a multi-dimensional framework for the assessment of conceptual quality, tailored for the context of VGI. Conceptual quality answers questions about the quality of conceptualization and its relationship with the data. This notion is strongly related to interpretability, the communication problem between the data creators who encode information according to their explicit and implicit knowledge, and consumers who need to interpret the data, reconstructing its intended meaning. Conceptual quality is essential to facilitate the communication over the semantic gulf that separates producers and consumers.

As conceptual quality is a complex, multi-faceted notion, six dimensions were identified: accuracy, granularity, completeness, consistency, compliance, and richness. Each dimension of conceptual quality was defined as complementary to traditional notions of quality developed in GIScience, proposing indicators to compute it and operationalize it. As an initial illustration of the framework, we explored a case study on four regions in Europe in OSM, focusing on the conceptual compliance of the tags. The case study highlights the wide applicability of conceptual quality to real data, and its potential to identify semantic and modeling issues in VGI.

Operationalizing conceptual quality is essential to increase the usability of VGI, adding a semantic facet to traditional notions of spatial, temporal, and thematic quality. The current state of the framework has several limitations that need addressing before deployment in realistic settings. The indicators described in this article need to be applied to OSM and other datasets in order to assess their strengths and weaknesses. Some dimensions of conceptual quality,

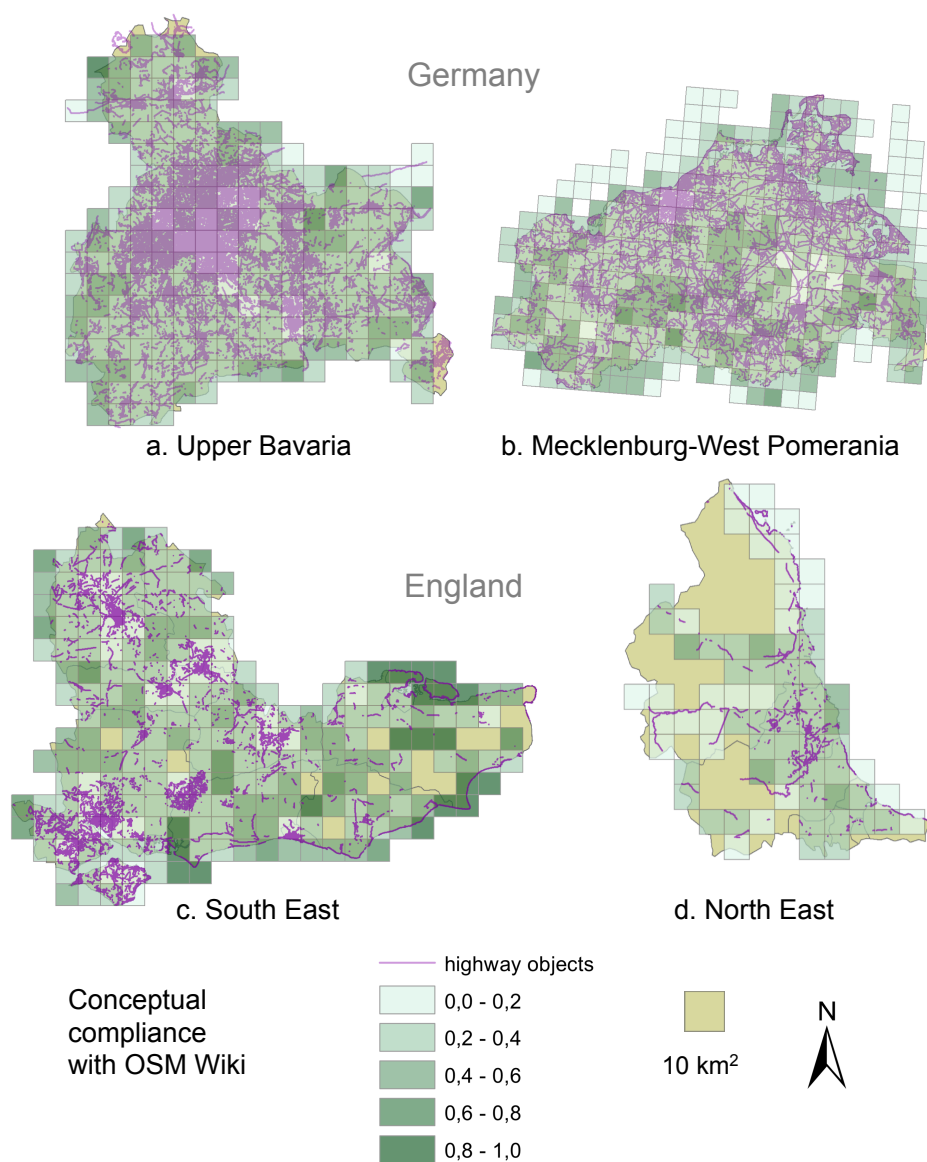


Fig. 2. Conceptual compliance I_{cm} with OSM Wiki on four European regions, calculated on a 10 km² grid.

such as conceptual granularity, will certainly prove harder than others to operationalize in different contexts. Without doubt, much empirical work is needed to deploy the framework effectively, with the goal of increasing the value and interpretability of VGI.

The core future direction for this work involves the application of the six dimensions to different datasets, comparing and contrasting the results, and tailoring more sophisticated and alternative indicators. For example, the relationship between how many contributors work on a region and its conceptual completeness needs further investigation [19]. A more mature version of the framework will be implemented into actual tools for VGI contributors and users, particularly for OSM. Conceptual quality, in its many empirically unexplored facets, will play an important role in overcoming the barriers to the usage of data as communication medium, mitigating the friction encountered when crossing the semantic gulf.

Acknowledgments

The authors thank Sophie Crommelinck and Sarah Labusga (University of Heidelberg) for the implementation of the case study, and the OpenStreetMap community for supplying the data.

References

- [1] Ballatore, A. (2014). Defacing the map: Cartographic vandalism in the digital commons. *The Cartographic Journal*, 51(3), 214–224.
- [2] Ballatore, A., Bertolotto, M. (2011). Semantically Enriching VGI in Support of Implicit Feedback Analysis. In: Tanaka, K., Fröhlich, P., Kim, K. S. (Eds) *Proceedings of the Web and Wireless Geographical Information Systems International Symposium*, LNCS, vol. 6574, Berlin: Springer, pp. 78–93.
- [3] Ballatore, A., Bertolotto, M., Wilson, D. (2013). Computing the Semantic Similarity of Geographic Terms Using Volunteered Lexical Definitions. *International Journal of Geographical Information Science*, 27(10), 2099–2118.
- [4] Ballatore, A., Wilson, D., Bertolotto, M. (2013). A Survey of Volunteered Open Geo-Knowledge Bases in the Semantic Web. In: Pasi, G., Bordogna, G., Jain, L. (Eds) *Quality Issues in the Management of Web Information*, Intelligent Systems Reference Library, vol. 50, Springer, pp. 93–120.
- [5] Barron, C., Neis, P., Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6), 877–895.
- [6] Batini, C., Scannapieco, M. (2006). *Data Quality: Concepts, Methodologies and Techniques*. Berlin: Springer.
- [7] Bishr, M., Kuhn, W. (2013). Trust and Reputation Models for Quality Assessment of Human Sensor Observations. In: Tenbrink, T., Stell, J., Galton, A., Wood, Z. (Eds) *Spatial Information Theory - 11th International Conference, COSIT 2013, Scarborough, UK, September 2-6, 2013. Proceedings* (pp. 53–73), LNCS, vol. 8116.

- [8] Burton-Jones, A., Storey, V., Sugumaran, V., Ahluwalia, P. (2005). A semi-otic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering*, 55(1), 84–102.
- [9] Cherfi, S., Akoka, J., Comyn-Wattiau, I. (2003). Conceptual Modeling Quality – From EER to UML Schemas Evaluation. In: Spaccapietra, S., March, S., Kambayashi, Y. (Eds) *Conceptual Modeling – ER 2002*, LNCS, vol. 2503, Berlin: Springer, pp. 414–428.
- [10] Dodge, M., Kitchin, R. (2013). Crowdsourced cartography: mapping experience and knowledge. *Environment and Planning A*, 45(1), 19–36.
- [11] Flanagan, A. J., Metzger, M. J. (2008). The Credibility of Volunteered Geographic Information. *GeoJournal*, 72(3-4), 137–148.
- [12] Frank, A. U. (2000). Spatial communication with maps: Defining the correctness of maps using a multi-agent simulation. In: Freksa, C., Habel, C., Brauer, W., Wender, K. (Eds) *Spatial Cognition II*, LNCS, vol. 1849, Springer, pp. 80–99.
- [13] Goodchild, M. F., Gopal, S. (1989). *The Accuracy of Spatial Databases*. Boca Raton, FL: CRC Press.
- [14] Goodchild, M. F., Li, L. (2012). Assuring the Quality of Volunteered Geographic Information. *Spatial Statistics*, 1, 110–120.
- [15] Guarino, N., Welty, C. A. (2009). An overview of OntoClean. In: Staab, S., Studer, R. (Eds) *Handbook on Ontologies*, 2nd edn, Berlin: Springer, pp. 201–220.
- [16] Guarino, N., Oberle, D., Staab, S. (2009). What is an Ontology? In: Staab, S., Studer, R. (Eds) *Handbook on Ontologies*, 2nd edn, Berlin: Springer, pp. 1–17.
- [17] Guptill, S., Morrison, J. (Eds) (1995). *Elements of Spatial Data Quality*. Oxford, UK: Elsevier.
- [18] Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B Planning and Design*, (37), 682–703.
- [19] Haklay, M., Basiouka, S., Antoniou, V., Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus’ law to volunteered geographic information. *The Cartographic Journal*, 47(4), 315–322.
- [20] Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), 550–557.
- [21] Hunter, G., Bregt, A., Heuvelink, G., Bruin, S., Virrantaus, K. (2009). Spatial data quality: Problems and Prospects. In: *Research Trends in Geographic Information Science*, LNGC, Berlin: Springer, pp. 101–121.
- [22] Kuhn, W. (2012). Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science*, 26(12), 2267–2276.
- [23] Mooney, P., Corcoran, P. (2012). Characteristics of heavily edited objects in OpenStreetMap. *Future Internet*, 4(1), 285–305.
- [24] Mooney, P., Corcoran, P. (2012). The annotation process in OpenStreetMap. *Transactions in GIS*, 16(4), 561–579.

- [25] Rosch, E. (1999). Principles of Categorization. In: Margolis, E., Laurence, S. (Eds) *Concepts: Core Readings*, Cambridge, MA: MIT Press, pp. 189–206.
- [26] Salgé, F. (1995). Semantic accuracy. In: Guptill, S., Morrison, J. (Eds) *Elements of Spatial Data Quality*, Oxford, UK: Elsevier, pp. 139–151.
- [27] Shi, W., Fisher, P., Goodchild, M. F. (Eds) (2003). *Spatial Data Quality*. Boca Raton, FL: CRC Press.
- [28] Solskinnsbakk, G., Gulla, J. A., Haderlein, V., Myrseth, P., Cerrato, O. (2012). Quality of hierarchies in ontologies and folksonomies. *Data & Knowledge Engineering*, 74, 13–25.
- [29] Stephens, M. (2013). Gender and the GeoWeb: divisions in the production of user-generated cartographic information. *GeoJournal*, 78(6), 981–996.
- [30] Tartir, S., Arpinar, I., Moore, M., Sheth, A., Aleman-Meza, B. (2005). OntoQA: Metric-based ontology quality analysis. In: *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, at the 5th IEEE International Conference on Data Mining 2005, ICDM'05* (pp. 1–9), IEEE.
- [31] Van Damme, C., Hepp, M., Coenen, T. (2008). Quality Metrics for Tags of Broad Folksonomies. In: *Proceedings of International Conference on Semantic Systems (I-SEMANTICS)*, Graz, Austria (pp. 118–125).
- [32] Veregin, H. (1998). Data quality measurement and assessment. *NCGIA Core Curriculum in Geographic Information Science*, URL http://www.ncgia.ucsb.edu/giscc/units/u100/u100_f.html.
- [33] Zielstra, D., Zipf, A. (2010). A comparative study of proprietary geodata and volunteered geographic information for Germany. In: Painho, M., Santos, M. Y., Pundt, H. (Eds) *Proceedings of the 13th AGILE International Conference on Geographic Information Science*, pp. 1–15.