

Article

Enhancing DSS Exploitation Based on VGI Quality Assessment: Conceptual Framework and Experimental Evaluation

Tarek Sboui ^{1,2,*}  and Saida Aissi ^{3,4} ¹ Department of Geology, Faculty of Science of Tunis, University of Tunis El-Manar, Tunis 1068, Tunisia² GREEN-TEAM Laboratory, INAT, Tunis 1082, Tunisia³ SMART Lab, Higher Institute of Management, University of Tunis, Tunis 2000, Tunisia⁴ Department of Computer Sciences, European University of Tunis, Les Berges du Lac III, Tunis 2015, Tunisia

* Correspondence: tarek.sboui@fst.utm.tn

Featured Application: Authors are encouraged to provide a concise description of the specific application or a potential application of the work. This section is not mandatory.

Abstract: The latest advances in spatial information technology have led to the emergence of Volunteered Geographic Information (VGI) as enrichment to existing spatial data sources. Additionally, Decision Support Systems (DSS) are among the fields that have seen major advances. Volunteered Geographic Information (VGI) has great potential as a valuable data source to decision support systems. Several studies have been proposed to integrate VGI data into DSS. However, as VGI data may have different levels of quality, integrating VGI data with poor quality may affect the decision-making process. In fact, VGI data with poor quality, that are obsolete or incomplete, could, if integrated into a spatial DSS, lead to inappropriate analysis results. This paper presents an approach that aims to enhance spatial DSS analysis and exploitation by integrating high quality VGI data that are appropriate to the user requirements, and that have a good indicator completeness and time relevance. The approach introduces a conceptual framework that evaluates VGI data quality and integrates only high quality VGI data into spatial DSS. The proposed approach is experimented on a road maintenance project in Grand-Tunis. We develop the *Map-Report* prototype, and we evaluate the efficiency of our approach in enhancing data analysis and exploitation in spatial DSS by reducing the error rate and providing accurate and precise analysis results.

Keywords: decision support system; datacube; data integration; VGI; data quality



Citation: Sboui, T.; Aissi, S. Enhancing DSS Exploitation Based on VGI Quality Assessment: Conceptual Framework and Experimental Evaluation. *Systems* **2023**, *11*, 393. <https://doi.org/10.3390/systems11080393>

Academic Editor: Oz Sahin

Received: 12 February 2023

Revised: 2 March 2023

Accepted: 8 March 2023

Published: 1 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The advances of data acquisition technologies (e.g., sensor networks and satellites) have led to the availability of large amounts of data in many domains, such as Earth observation, telecommunication, and navigation. Decision support systems (DSS) have been among the notable advancements [1–6]. They play a pivotal role in assisting decision-makers such as analysts, executives, and managers in tackling intricate strategic problems. Data warehouses are a fundamental component of these DSS, acting as a substantial repository of subject-oriented, integrated, time varying, non-volatile collection of data [7,8]. In a data warehouse, data are typically extracted from different data sources, transformed to meet the strategic needs, and finally integrated and loaded into a unified database [4]. Data warehouses provide decision makers with historical and aggregated data based on using various functions (e.g., sum, count, and average) according to different levels of details. This data plays a crucial role in equipping decision-makers with valuable insights, enabling them to analyze phenomena and arrive at well-informed strategic choices. Data warehouses are commonly organized following a multidimensional structure known as datacubes [9]. These datacubes allow decision-makers to explore aggregated and summarized data based

on various dimensions and their corresponding levels. Moreover, datacubes help strategic decision-makers in accessing diverse perspectives of the data through various tools like Online Analytical Processing (OLAP) tools [7]. OLAP tool allows for analyzing datacube's measures, such as 'agricultural production values', based on different dimensions, such as 'type of agricultural production'. A measure is calculated and aggregated according to one or several dimensions. Each dimension may contain one or more levels.

Volunteered Geographic Information (VGI) can serve as input data for spatial datacubes. VGI is a social and technological phenomenon that originates outside the realm of authoritative mechanism of spatial data acquisition [10,11]. VGI pertains to utilizing the internet to access, gather, and distribute spatial data contributed voluntarily by individual contributors [12–16]. It empowers individuals without specialized spatial skills or training to actively engage in acquiring spatial data. Recent advancements in spatial technology have further enriched the potential of VGI. Its application spans across various domains, including forestry, healthcare, disaster management, and road maintenance [17].

Recently, there have been some studies that combined Volunteered Geographic Information (VGI) and spatial DSS [18–20]. These studies showed that combining VGI and DSS is a valuable enhancement for decision making.

However, VGI data may be of poor quality which may—if integrated into a spatial datacube—result in a negative effect on decision-making. For example, integrating VGI data that are obsolete (e.g., a pothole in the street that has been repaired) or incomplete (e.g., a pothole wrongly located on the map due to a missing reference system) will lead to inaccurate decision-making, and produce inappropriate analysis results. Despite the risk of inappropriateness, it is still interesting to integrate VGI data into DSS [18–20]. However, we should pay attention to integrate only data with high quality to allow an appropriate decision-making analysis.

While several works showed promising results in integrating VGI data into a spatial datacube [18–20], to the best of our knowledge, there has been no research study that focuses on assessing the quality of VGI data with regard to user requirements prior to its integration. In fact, VGI data that are obsolete or incomplete could not be appropriate to the decision-making process and to the user requirements.

In this paper, we propose to enhance spatial DSS integrating VGI data with high quality; the data, which are appropriate for the requirements of a particular decision-making task, and that have a high quality indicator regarding completeness and time relevance. Hence, we focus on VGI quality assessment by proposing and evaluating a VGI quality based on set of indicators. The proposed approach is implemented and tested on a road maintenance project in Grand-Tunis, Tunisia. The implementation aims to support users in their decision-making about road maintenance, and to enhance spatial DSS analysis and exploitation. The proposal is described theoretically and validated by experiments.

The paper is organized as follows: Section 2 presents some related works that combined VGI and DSS. Section 3 introduces the approach for enhancing spatial DSS capabilities by integrating high quality VGI data. Section 4 describes a prototype we developed to implement our approach. Section 5 presents the experimental evaluation and the obtained results, and Section 6 concludes the paper and presents some perspectives for future work.

2. Related Works

Spatial DSS may include, besides authoritative spatial data, VGI data collected by people having little or no spatial knowledge [17]. Integrating VGI data into spatial DSS has been demonstrated in various domain applications, such as health care, transportation, managing events, and urban space management [20–22].

Sadeghi-Niarak [20] incorporated VGI into a decision support system, harnessing citizen-contributed data to efficiently manage waste pollution. They proceeded to create a web-based prototype of this decision support system, effectively showcasing its practical feasibility and relevance in conducting pollution-related decision analyses.

Omidipour et al. [23] integrated VGI into a spatial DSS to enhance urban management. They developed a spatial DSS that combines VGI with multi-criteria decision analysis to facilitate participatory renewal procedures in urban-blighted areas.

Horita et al. [19] used interoperable standards of the Open Geospatial Consortium (OGC) to integrate VGI into a spatial DSS. The developed DSS is implemented to improve flood risk management in the town of São Carlos, Brazil. The implementation results showed that interoperable standards can support the integration of VGI data in spatial DSS. Horita et al. [24] developed a framework that combines VGI data and conventional data in the spatial DSS in order to enhance decision making in disaster management.

Rajabifard et al. [25] integrated VGI and geospatial data from distributed sources in the Intelligent Disaster Decision Support System (IDDSS). The developed IDDSS proved its efficiency in managing floods and fires in Melbourne, Australia.

While the aforementioned studies have shown the importance of integrating VGI data into DSS datacubes, to our knowledge, there has been no research study that focuses on assessing the quality of VGI data with regard to user requirements prior to the integration. Assessing VGI data is needed as VGI with potentially poor quality may—if integrated into a spatial datacube—result in a negative effect on decision-making process. In fact, VGI data which are obsolete or incomplete are not relevant to the user requirements and the decision-making needs, and may lead to produce inappropriate analysis results.

In this work, we propose an approach to enhance spatial DSS exploitation by integrating high VGI quality data that are adapted to the user needs and requirements and having a high quality indicator regarding its completeness and time relevance. That is, before integrating VGI data into spatial datacubes, we first evaluate its quality with regard to the requirements of a particular decision-making. The proposed approach is described theoretically and evaluated by experiments using a prototype we implemented called *Map-Report*.

The next section introduces the proposed approach for enhancing spatial DSS by integrating high quality VGI data into spatial datacubes. The approach involves a framework for integrating only high quality VGI data into DSS datacube, and an algorithm for recommending the integration or not of VGI data. In order to define high quality data, the proposed framework includes the evaluation of VGI quality with regard to the requirements of a particular decision-making task.

3. Integrating VGI Data in SDSS

In Spatial Decision Support Systems, spatial datacubes are broadly deployed as common repositories where data sources are integrated and stored. In order to integrate data into a spatial datacube, a complex Extract-Load-Transform (ETL) process is typically used. ETL process consists of cleaning and unifying data extracted from multiple sources. Then, spatial data are stored and managed mainly according to a multidimensional structure (i.e., datacube). Eventually, the data is used by different tools (e.g., spatial OLAP tools and mining tools) that allow decision-making analysis and reporting [8,11]. Figure 1 shows the typical architecture of a spatial DSS, which includes three major parts that consist of ETL process for extracting transforming and loading data into the datacube, organize data in a multidimensional structure, and exploiting data to generate interactive dashboards and analysis reports.

3.1. Conceptual Framework for Integrating VGI Data into Datacubes

In order to enhance spatial DSS exploitation, we propose a framework for integrating high quality VGI data. In fact, besides conventional data sources, we integrate VGI data into DSS datacubes. However, before integrating VGI data, we assess its quality with regard to (1) its completeness, and (2) its time relevance. Only VGI data that respect these criteria and that have a high-quality indicator will be integrated in the DSS.

Figure 2 illustrates the proposed framework which consists of three main layers: VGI data collection Layer, VGI quality Assessment Layer, and VGI integration Layer. The first

layer includes receiving VGI data from contributors using a VGI platform, and cleaning and transforming VGI data to make it into an effective and consistent form. Cleaning and transforming operations can be achieved automatically or manually.

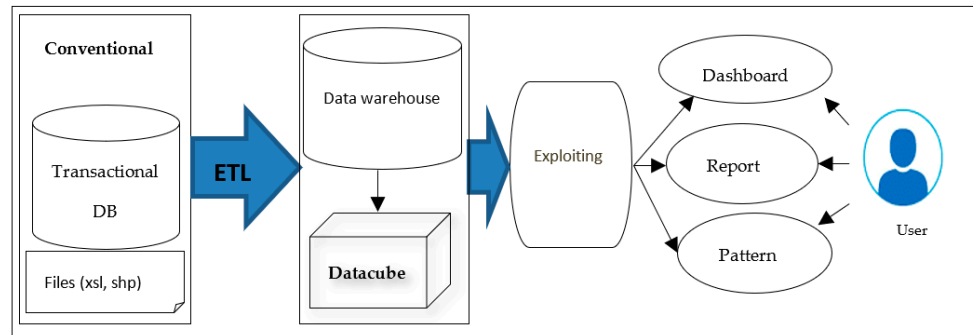


Figure 1. Typical architecture of a Spatial DSS.

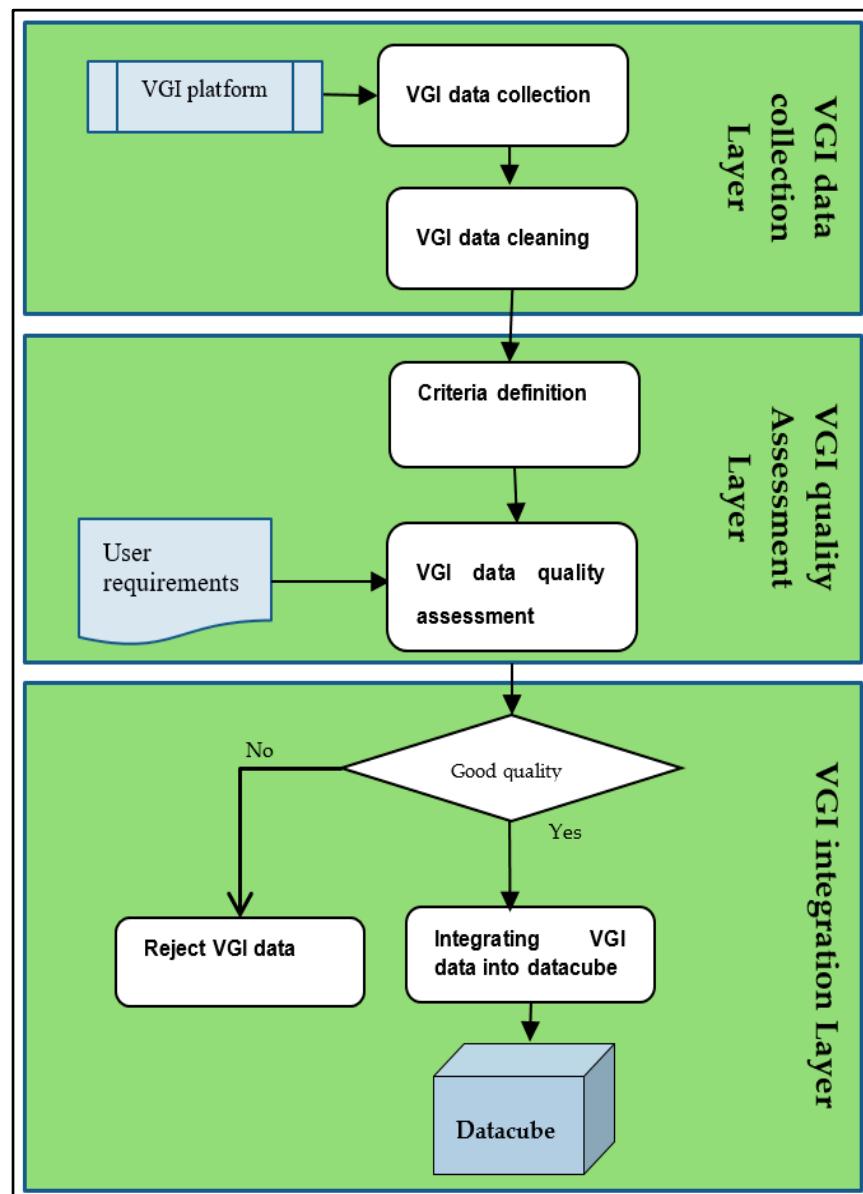


Figure 2. Framework for assessing VGI quality and integrating high quality VGI data into DSS datacube.

The second layer, ‘VGI quality Assessment’, allows for assessing the quality of VGI data based on two indicators, ‘completeness’ and ‘time relevance’. The proposed indicators perform a key role in the proposed framework, specifically, these indicators have two principal aims: First, to make users aware of the quality of the data they are dealing with. Second, to enhance spatial DSS by integrating only high quality VGI data, which are adapted to the user requirements, and to the decision-making needs. All details about the proposed VGI quality indicator are presented in Section 3.2.

The third layer ‘VGI integration’ aims to integrate only VGI data that is high quality to the DSS’s datacube. For that, in this layer, we test if the result of VGI quality assessment is satisfactory with regard to user requirements, and have a sufficient value of time relevance and completeness. If it is the case, VGI data is validated as a data source to be integrated into the DSS’s datacube. Otherwise, if the quality is not satisfactory with regard to the system requirements, VGI data is normally rejected.

In order to assess the quality of VGI data, we define a set of quality indicators, and we propose a quantitative approach to evaluate the quality based on these quality indicators.

3.2. Indicators for Assessing VGI Data Quality

VGI data quality can be described by guidelines (e.g., ISO 19157 [26]) and indicators (e.g., accuracy, consistency, and completeness) [27,28]. Quality indicators for VGI aim to provide a valuable support for weighting and selecting relevant VGI data to be integrated into spatial datacube. That is, based on quality indicators, one could decide what VGI data should be integrated into a spatial datacube. We propose two indicators, which are completeness and time relevance. These indicators are among the most widely used data quality indicators [29–33]. Measuring the completeness of VGI data is key for VGI to fit a particular usage [16,34,35]. Time relevance is critical to ensure that VGI data collected by contributors matches temporal requirements of decision makers [36].

The assessment of VGI data quality is based on the values of indicators, which range from 0 to 1, where 1 indicates perfect quality of VGI data with regard to decision-making needs, and 0 indicates the poorest quality. It is important to note that our goal is not to define an exhaustive list of indicators, but to create awareness about the quality of VGI data with regard to decision-making. A limited set of quality indicators can provide synthetic information about spatial data quality. Typical decision-making processes are based on a small number of indicators [34,37,38].

We illustrate the indicator-based assessment of VGI data quality with the following example: a VGI contributor provides information about a damage on the road in a given location (Ben Arous, Grand-Tunis), as shown in Figure 3. This VGI contributor sent information about a broken manhole on a road with no information about the type of the road where the damage was reported (i.e., incomplete VGI data). Due to the lack of information, users may misinterpret the VGI data, which may affect the decision-making process. That is, without more complete information, the quality of VGI may be insufficient and hence not appropriate to make a strategic decision. The proposed approach in this paper aims to make users aware of the quality of VGI data to be integrated into a spatial datacube.

3.2.1. Completeness

The completeness indicator illustrates the number of VGI data elements relevant to the specific requirements of a decision-making process. A data element is defined as {element nature, element type, element value}, where ‘element nature’ signifies whether it pertains to data or metadata, and ‘element type’ indicates whether the data represents a feature, an attribute, or a relationship. The ‘element value’ corresponds to the actual representation of the element. For instance, the data ‘broken manhole on a road’ has a value that comprises both text and an image. An example of a metadata element related to the data ‘broken manhole on a road’ is the reference system related to spatial coordinates. A missing information about the reference system leads to an incomplete VGI data, which may affect data analysis and exploitation.



Figure 3. Uncertainty about the quality of VGI data.

We measure the completeness with regard to (1) the datacube's measures (i.e., subjects of analysis), and to (2) the datacube's dimensions (i.e., axes of analysis).

The completeness with regard to the dimensions C_d is calculated as follows:

$$C_d = \begin{cases} 1; & \text{if } N_{med} \geq N_{mrd} \text{ and } N_{sed} \geq N_{srd} \\ (w_m * N_{med}/N_{mrd} + w_s * N_{sed}/N_{srd})/2; & \text{otherwise} \end{cases} \quad (1)$$

where N_{med} and N_{sed} are the numbers of thematic and spatial data elements related to a dimension provided by VGI contributor. N_{mrd} (N_{srd}) is the number of thematic and spatial, respectively, elements required for a particular decision-making. w_m and w_s are predefined weights for thematic and spatial, respectively, elements. These weights indicate the importance of each type in a particular decision-making. When the number of available thematic and spatial data elements equals or surpasses the required number of elements, the VGI dimension is considered complete, and its value is assigned as 1. In cases where the existing elements are fewer than the required elements, the ratio of available elements to the required elements represents the degree of completeness for each dimension of the datacube.

The quality of all dimensions C_{td} is calculated according the ratio of the sum of the dimensions' quality of a given datacube to the number of dimensions nd (as shown in Formula (2))

$$C_{td} = \sum C_{di} / nd \quad (2)$$

The completeness with regard to the measure C_m is calculated as follows:

$$C_m = \begin{cases} 1; & \text{if } N_{me} \geq N_{mr} \text{ and } N_{se} \geq N_{sr} \\ (w_m * N_{me}/N_{mr} + w_s * N_{se}/N_{sr})/2 * \sum C_{di}; & \text{otherwise} \end{cases} \quad (3)$$

where N_{me} and N_{se} are the numbers of thematic and spatial measure data elements provided by the VGI contributor. N_{mr} (N_{sr}) is the number of thematic (spatial and temporal) measure data elements required for a particular decision-making. w_m and w_s are predefined weights for thematic and spatial data elements. If the number of available elements is equal to or greater than the required number of elements, the measure data is complete, and its value is set to 1. Otherwise, the completeness of each measure is evaluated as a product of (1) the ratio of the VGI elements to the required elements, and (2) the quality of the dimensions related to that measure (i.e., its axe of analysis). The above Equation (3) indicates that the quality of each measure depends not only on its elements, but also on the quality of the dimensions' elements used as an axe of analysis for that measure. The quality of VGI measure increases and decreases with the increase and decrease in the quality of VGI

dimensions. The equation reflects that the measure is calculated and aggregated according to a set of dimensions.

The quality of all measures C_{tm} is calculated according the ratio of the sum of the measures' quality of a given datacube to the number of measures nm (as shown in Formula (4)).

$$C_{tm} = \sum C_{mi} / nm \quad (4)$$

The number of elements and the weights (w_m and w_s) can be predefined by VGI analysts in collaboration with users. We should note that VGI analysts and users may choose not to define these weights. In this case, the value of each weight is set to 1 to indicate that thematic and spatial elements have the same importance, otherwise the value of each of these weights should be between 0 and 1.

In the example presented in Figure 3, VGI data is about road maintenance measure. This measure is defined according to the following dimensions: 'type of road', 'type of damage', 'time', and 'geo-location'. In this case, decision-makers require, besides available data elements (time and location), another element specifying the reference system related to the location coordinates. With regard to dimensions, all elements required are provided by the VGI contributor. Thus, if we suppose that the weight of completeness is 1, then $C_m = (3/4 \times 3/3) = 0.75$.

3.2.2. Time Relevance

This quality indicator shows the degree of temporal relevance of VGI measure with regard to the decision-making requirements. The value of time relevance is evaluated based on the period of time needed for VGI data aggregation in a particular decision-making. The time relevance of the VGI data is evaluated as follows:

$$T_m = \begin{cases} 1; & \text{if } Taggrb \leq Tdef \leq Taggre \\ 1 - (\min(|Taggrb - Tdef|, |Taggre - Tdef|) / (Taggre - Taggrb)); & \text{otherwise} \end{cases} \quad (5)$$

where $Tdef$ is the time of definition of VGI data, and $Taggrb$ and $Taggre$ are the beginning and the end, respectively, of the period needed for data aggregation. The subtraction operation $Taggre - Taggrb$ is the period needed for aggregation.

If the time of definition is within the period of aggregation, the time relevance of VGI measure T_m is perfect, and its value is set to 1. Otherwise, T_m is measured as a ratio of (1) the minimum distance to the required period of aggregation, and (2) the period of time needed for aggregation in a particular decision-making.

The overall time of relevance quality for all datacube measures T_{tm} is calculated as follows:

$$T_{tm} = \sum T_{mi} / nf \quad (6)$$

The value of the times relevance decreases as the distance of its definition time to the aggregation period increases. A low value of time relevance of VGI data may have a negative impact on the quality of decision-making process. In the example presented in Figure 3, VGI data was defined in January 2017, the beginning (and the end) of period needed for data aggregation is January 2019 (and December 2021, respectively). Then, $T_m = 1 - (24/36) = 0.33$.

The overall quality Q is calculated as follows:

$$Q = \sum W_i * Q_i / N_q \quad (7)$$

where Q_i is the i th quality indicator of the VGI data. where W_i is the i th predefined weight of the i th indicator. Where N_q is the total number of indicators. We should notice that the value of the weight is predefined by analysts or users and should be between 0 and 1. If analysts and users choose not to be involved in the quality assessment process, the value of each weight is set to 1 (i.e., the indicators are given the same importance).

The threshold of the weight value depends on user's needs and preferences with regard to the level of completeness and time relevance. Users in collaboration with VGI analysts are in the best position to define the most appropriate threshold.

Consequently, if we assume that the weight of each indicator is 1, the overall quality $Q = (0.75 + 0.33)/2 = 0.54$.

We should remember that the defined indicators do not aim at being complete, but rather at helping users to make a decision about integrating—or not integrating—VGI data into spatial DSS.

In the next section, we propose an algorithm to help decisions makers to make appropriate decisions about integrating—or not—VGI data into spatial datacube.

3.3. Algorithm for Recommending VGI Data Integration

Based on the aforementioned indicator-based assessment of VGI data quality, we propose an algorithm that aims to recommend the integration—or not—of VGI data into spatial datacube.

The algorithm verifies the VGI quality for both measures and dimensions. Based on this quality, VGI analysts will be advised to:

- Integrate VGI, if VGI data that has high quality with regard to analytical requirements. High-quality VGI data have an overall quality value above a threshold that is defined by users in collaboration with VGI analysts (e.g., an overall quality value greater than or equal to a threshold of 0.75).
- Not to integrate VGI data that does not have high quality (e.g., less than a threshold of 0.75).

The proposed indicators perform a key role in the proposed algorithm. They increased awareness of the quality of VGI data, and allow to make appropriate decision about integrating—or not—VGI data into datacube.

Algorithm 1 VGI Quality for Spatial DSS

```

1: Input:
2:   Nq: number of quality indicators
3:   thr: threshold of acceptable data quality
4: Output: Rec: recommendation to the user
5: Begin
6: Rec ← ""
7: SQ ← 0
8: For i ← 1 to Nq do
9:   determine qi
10:  determine wi
11:  SQ ← SQ + wi * qi
12: End
13: Q ← SQ/Nq
14: if (Q ≥ thr) then
15:   Rec ← Recommending integrating VGI data into spatial datacube
16: else
17:   Rec ← Recommending not to integrate VGI data into spatial datacube
18: End

```

4. The Map-Report Prototype

We developed a prototype, called *Map-Report* (MR), which allows to collect data from VGI contributors, and to assess VGI data quality with regard to a particular decision-making (as shown in Figure 4).

The prototype implements three main functionalities: (1) the VGI data collection from contributors, (2) the assessment of VGI data quality based on the measurement of

indicators, and (3) the display of warnings to help users intuitively understand the value of data quality.

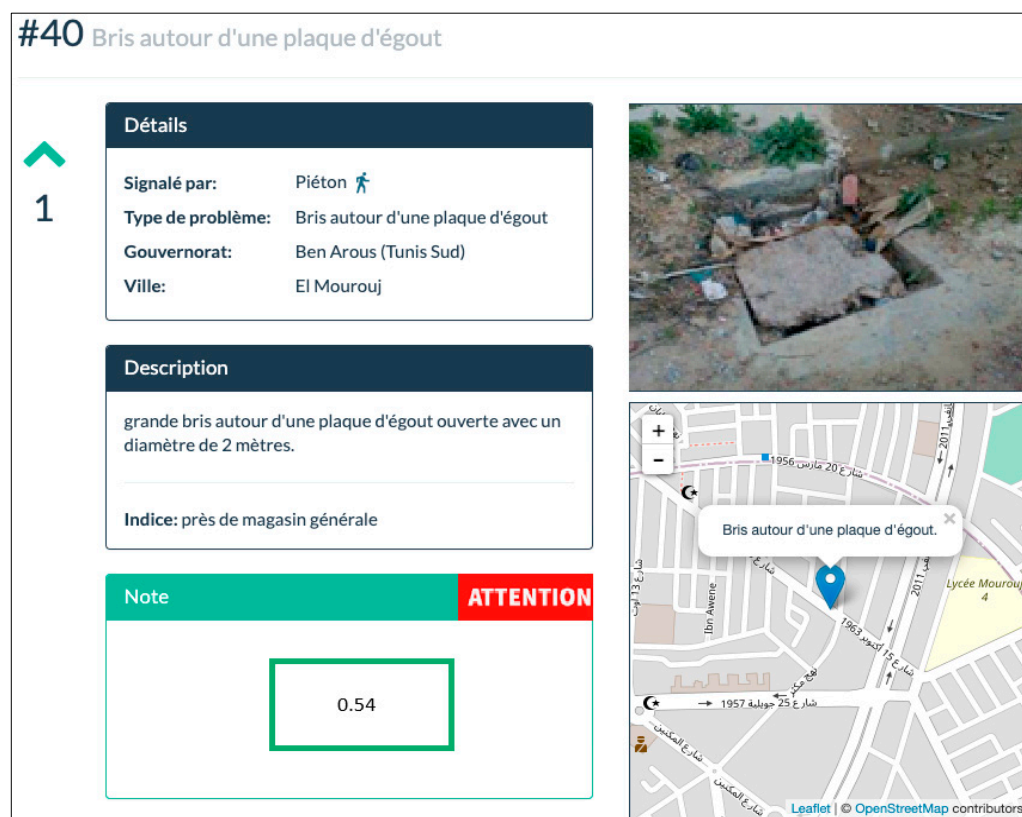


Figure 4. The *Map-Report* prototype.

The quality of the VGI data is assessed by our system (as detailed in Section 3), and a quality indicator value is shown in order to decide to include—or not—VGI data in the DSS datacube. Moreover, based on the quality value, a symbol is shown to the users to make them aware of the VGI quality. Figure 4 shows a VGI data provided by a pedestrian presenting a debris around a broken manhole. The diameter of the debris is observed to be 2 m. The data is collected in Ben Arous region located in El Mourouj city in Grand-Tunis. As shown in Figure 4, the quality of the VGI data is below 0.75, thus an attention symbol is shown to the users to make them aware of relatively bad quality of VGI data. We evaluate VGI data to be integrated in a road maintenance datacube. The project uses VGI data and other data resources to locate and manage road damage in Grand-Tunis. We used a set of georeferenced data sent by contributors as reports through the *Map-Report* prototype.

5. Experiments and Results

We used a set of VGI reports sent by 20 contributors. The reports contain images and textual information that are used to describe present manholes, potholes, bad roads, and other road deterioration aspects located in different areas of Tunisia. Using our *Map-Report* prototype, we evaluate the quality of each report. The evaluation process selects a set of reports based on the sample-building methodology of Miller and Charles [39]. Specifically, the evaluation involves 30 reports, divided into three categories: 10 reports with high-quality levels ($0.75 < Q \leq 1$), 10 reports with a medium-quality levels ($0.5 < Q \leq 0.75$), and 10 reports with a lower-quality levels ($Q < 0.5$); The quality assessment is carried out using our *Map-Report* prototype.

On the other hand, we asked 10 road maintenance experts to visit the roads where the VGI contributors reported damages using the *Map-Report* prototype, then to attribute to each VGI data a quality value that reflects its suitability to the requirements of the road

maintenance project. The value of the quality indicator attributed by experts to each VGI data should be between 0 and 1.

In order to evaluate our VGI quality assessment proposal, we employed the commonly used Spearman correlation method, designed to gauge the strength and direction of a relationship between two ranked variables [40]. In our evaluation, we assessed the correlation between two assessments: (1) the VGI quality assessment from the *Map-Report* prototype and (2) the quality assessment by road maintenance experts. The correlation analysis results are depicted in Figure 5, indicating a Spearman's coefficient value of 0.76. This value signifies a strong positive correlation between the quality evaluations conducted by the experts and our VGI quality assessment, which demonstrates the efficiency of our approach in presenting good VGI quality assessment.

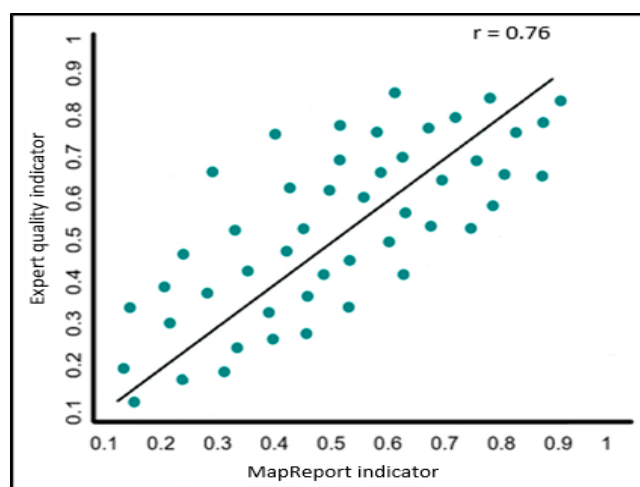


Figure 5. Correlation between quality assessment of *Map-Report* and quality assessment provided by road maintenance experts.

In order to evaluate the impact of integrating high VGI data quality in DSS, we implement a spatial datacube that integrates VGI data for a road maintenance project in Tunisia. The next section presents the implemented datacube and the integration process.

5.1. Integrating Assessed VGI Data into DSS Datacube

Based on the quality assessment of VGI data, we suggest integrating only high quality VGI data into DSS datacubes to enhance data analysis and exploitation.

5.1.1. The Spatial Datacube

In order to demonstrate the utility of our approach, we created a spatial datacube with which we intend to integrate only VGI data that is high quality. This datacube aims to enhance decision-making based on the number of damaged roads and number of accidents. The datacube stores the number of damaged roads according to the following dimensions: type of damage, location, time, and type of road. Spatial dimension 'geo-location' contains the governorate-region hierarchy; a governorate can contain one or more regions. The spatial datacube is modeled based on Malinowski and Zimányi's conceptualization of spatial data warehouses conceptualization of geospatial data warehouses [41]. Figure 6 shows the multidimensional schema of the developed datacube. This schema allows for the storing and analyzing of data regarding damaged roads in different areas of Grand-Tunis, for different type of roads (e.g., primary, secondary, and residential), and during different periods of time.

The datacube allows for exploring information about road maintenance, and allows answering queries, such as: "What is the total number of damaged roads in Grand-Tunis in 2020?", or "What is the total number of damaged secondary roads in Grand-Tunis in the first semester of 2020?"

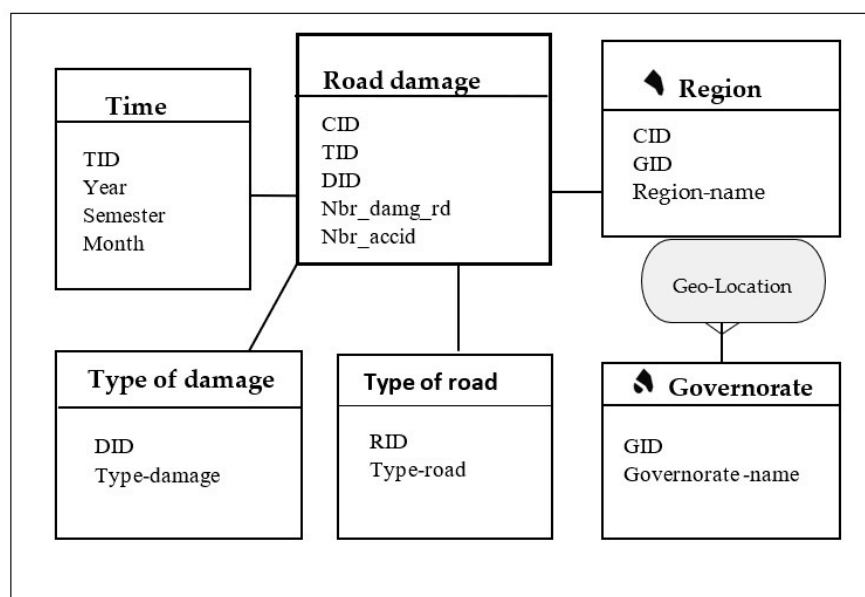


Figure 6. The damaged road datacube schema.

5.1.2. Integrating VGI into Spatial Datacube

After being assessed, high quality VGI data are manually extracted and stored in shape files and csv files. Then, these VGI data files are integrated into the road maintenance datacube using the Pentaho data integration (PDI) tool. Figure 7 shows the ETL process to integrate VGI data into the datacube. In this process, selected high quality VGI data are read and compared to data from other sources, before being loaded into the datacube.

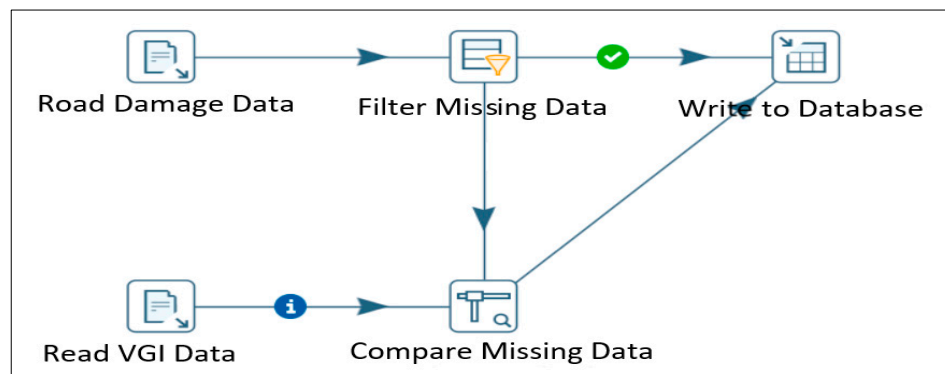


Figure 7. ETL integration of VGI data into road maintenance datacube.

In order to exploit datacube, we used MDX (Multi-Dimensional Expression) language to define queries which are used to select desired data from datacubes. The queries' extracted data is used to perform analytical analysis regarding the maintenance of damaged roads in Grand-Tunis, and to create reports, such as diagrams, pie graphs, map visualization, and so on. Figure 8 shows a bar chart displaying the total number of damaged roads and accidents from 2017 to 2020 in Grand-Tunis.

5.2. Experiments and Results

The adopted methodology for proposal evaluation is based on a comparative study between the relevance of the experimental results in two situations: (1) when the quality assessment approach is not applied, and (2) when the quality assessment approach is used. Accordingly, we create two datacubes: The first datacube (datacube 1) integrates VGI data without quality assessment. The second datacube integrates only high-quality VGI data (datacube 2).

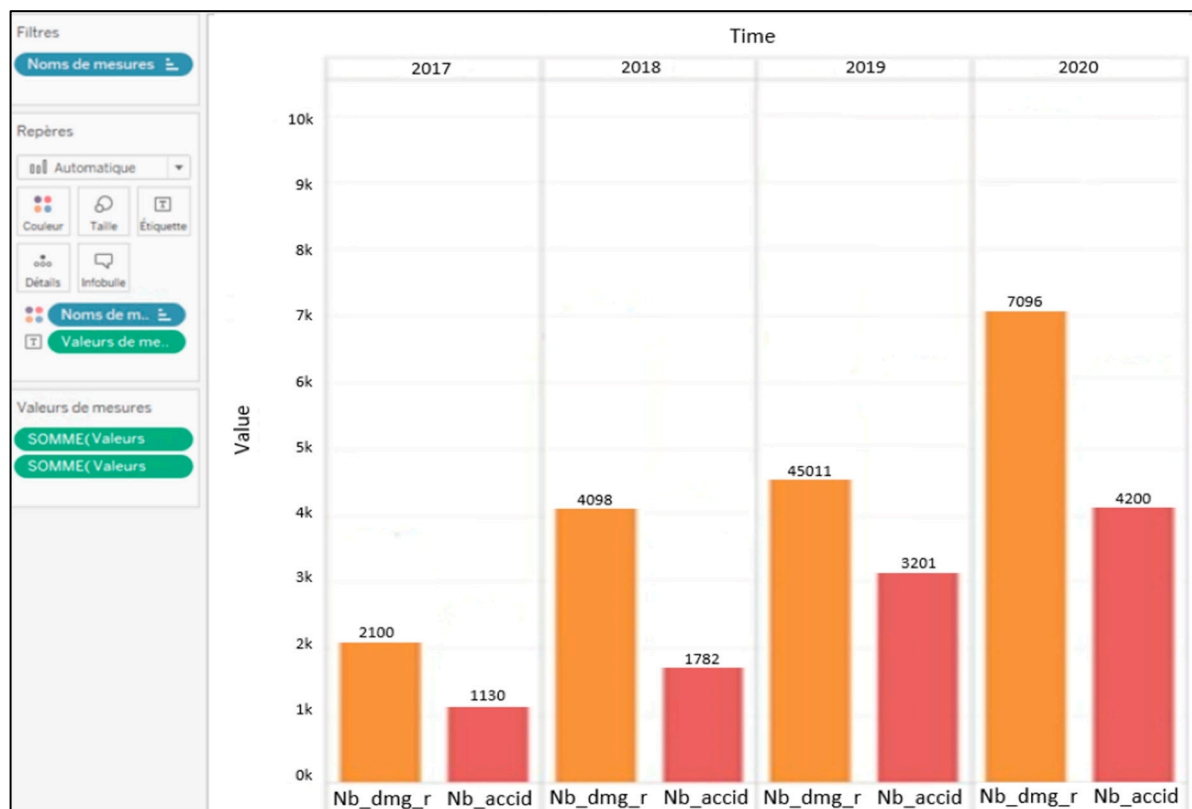


Figure 8. The total number of damaged roads and accidents from 2017 to 2020 in Grand-Tunis.

The relevance of the experimental results is evaluated using the mean absolute error (MAE), the mean quadratic error (MQE), and the root mean square error (RFSE). These indicators are computed based on the comparison of query results obtained from the two datacubes (datacube 1 and datacube 2) with real values presenting the answers to these queries. All these indicators are used in many domains to quantify the error generated by the system and in assessing system performance [42–45].

We ran 25 queries on the two datacubes (datacube 1 and datacube 2). For each query, we obtained different results for the two datacubes. For example, Figure 9 presents the obtained results for the query “What is the number of damaged secondary roads in the governorate of Ben Arous?” Figure 9a presents the results using VGI data without quality assessment (datacube 1), and Figure 9b presents results for the same query for the datacube that contains only high-quality VGI data (datacube 2). Likewise, Figure 10 presents the obtained results to the following query “What is the number of damaged roads and the number of accidents from 2017 to 2020 in Grand-Tunis?” Figure 10a presents results to this query on the datacube 1, and Figure 10b presents results to the same query on the datacube 2.

After identifying the real values presenting the answers to these queries, we calculate the mean absolute error (MAE), the mean quadratic error (MQE), and the root mean square error (RFSE). These error measures are compared with regard to the real expected values presenting the true answers to those queries and are computed in the two situations: (1) responses to queries without VGI data quality assessment (i.e., datacube integrating all VGI data into DSS datacube), and (2) responses to queries with VGI quality assessment (i.e., datacube integrating only high-quality VGI data). We use the error measures to evaluate which datacube allows for producing more accurate and precise query results with lower error rates. The lower are the error measures, the more efficient the system is.

Having:

n : the number of queries launched on the DSS; $n = 25$

R_i : the real expected value of the result of *query i*

S_i : the value presenting the response to *query i* on *datacube 1* (the DSS does not take VGI quality assessment into consideration)

S'_i : the value presenting the response to *query i* on *datacube 2* (the DSS integrates VGI quality assessment)

T_i : the absolute error relating to the *query i* without quality assessment

T'_i : the absolute error relating to *query i* with quality assessment

T_i and T'_i are obtained as follows

$$T_i = |S_i - R_i|$$

$$T'_i = |S'_i - R_i|$$

MAE_{cube1} : the mean absolute error without VGI quality assessment

MAE_{cube2} : the mean absolute error with VGI quality assessment

$$MAE_{cube1} = \sum_{i=1}^n T_i / n$$

$$MAE'_{cube2} = \sum_{i=1}^n T'_i / n$$

MQE_{cube1} : The mean quadratic error without VGI quality assessment

MQE_{cube2} : The mean quadratic error with VGI quality assessment

$$MQE_{cube1} = \sum_{i=0}^n (S_i - R_i)^2 / n,$$

$$MQE_{cube2} = \sum_{i=0}^n (S'_i - R_i)^2 / n$$

$RMAE_{cube1}$ = the root mean squared error without VGI quality assessment

$RMAE_{cube2}$ = the root mean squared error with VGI quality assessment

$$RMAE = \sqrt{\sum_{i=0}^n (S_i - R_i)^2 / n}$$

$$RMAE' = \sqrt{\sum_{i=0}^n (S'_i - R_i)^2 / n}$$

Table 1 presents the obtained results for MAE, MQE and RFSE relative to the analysis of query results.

Table 1. The obtained values of MAE, MQE, and RFSE relative to the results query analysis.

| | Datacube 1 | Datacube 2 |
|------|------------|------------|
| MAE | 78.545 | 17.818 |
| MQE | 12,085.181 | 663.473 |
| RFSE | 109.932 | 25.757 |

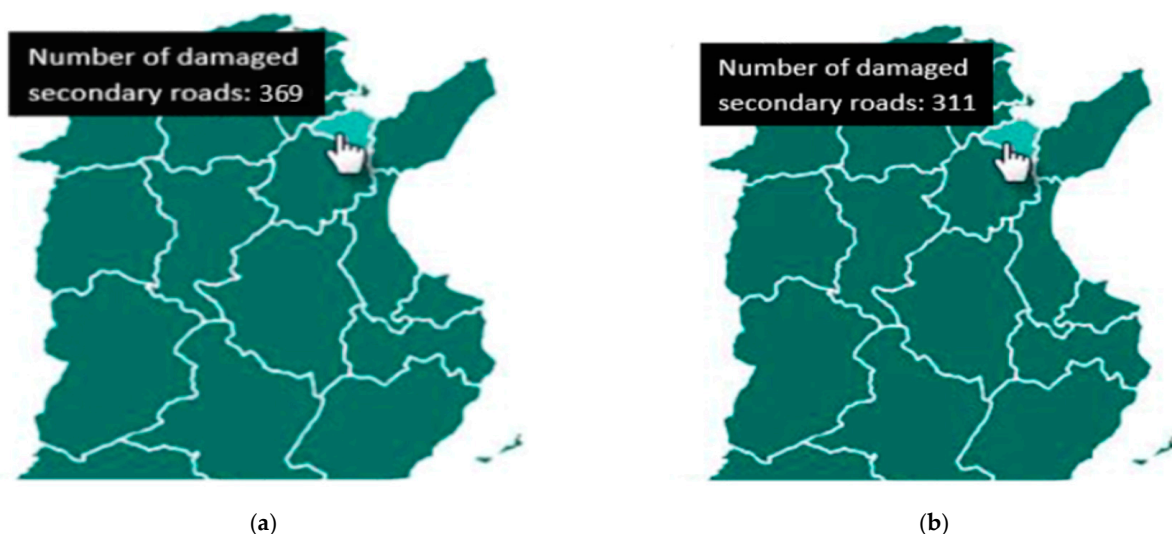


Figure 9. (a,b). The number o secondary damaged roads in the governorate of Ben Arous.

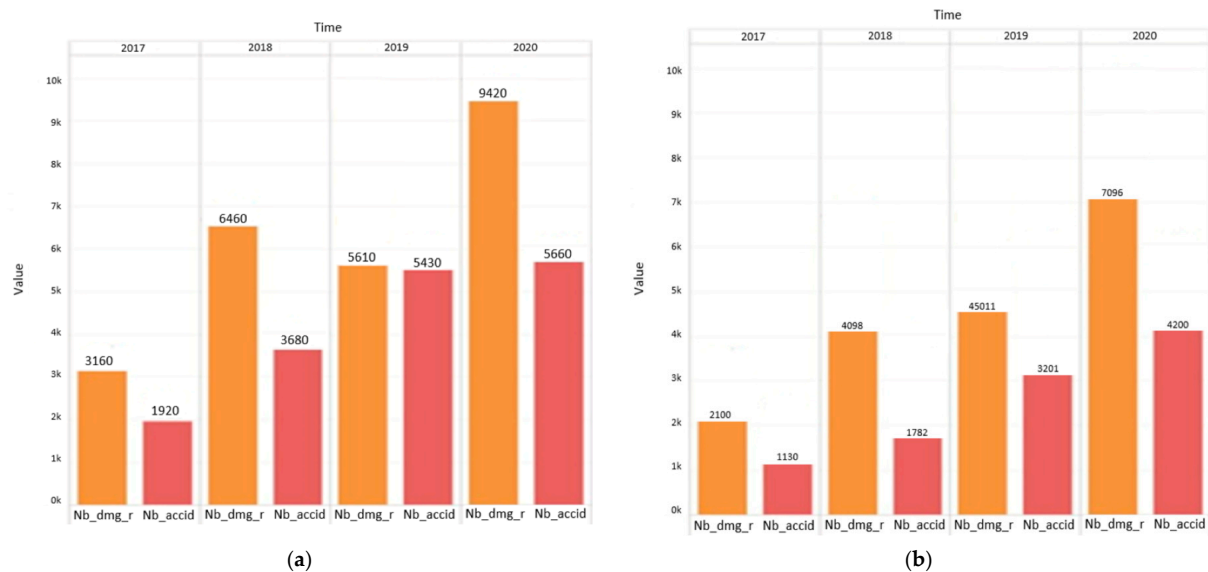


Figure 10. (a,b). Number of damaged roads and accidents from 2017 to 2020 in Grand-Tunis.

The obtained results show that all the error measures: the MAE, the MQE, and the RFSE are much lower when integrating only high quality VGI data into the DSS datacube. The DSS is more efficient and provides more accurate analysis results when it integrates only high-quality VGI data based on the proposed quality assessment. Thus, integrating VGI quality assessment in the DSS allows to considerably reduce the error rate.

The results show the ability of our system to enhance spatial DSS exploitation by providing more precise and accurate analysis results, and eliminating VGI data that are irrelevant, inappropriate, or even obsolete which, if integrated into the datacube, they may distort the query results.

However, we should notice that despite the VGI quality assessment, the error is not completely eliminated. In fact, our approach does not aim to completely eliminate the error related to the integration of VGI data into DSS, but rather, to reduce the error rate and to obtain query results and reports which are more accurate and closer to the reality.

6. Conclusions

Decision support systems and VGI are among the fields that have seen major advances. Integrating VGI data into DSS datacubes could enhance decision-making, as demonstrated by previous works. However, VGI data may be of poor quality, and may—if integrated into a spatial datacube—result in a negative effect on decision-making. For example, integrating VGI data that are obsolete or not adapted to the analytical requirements of end-users will lead to inaccurate decision-making results.

In this work, we presented an approach for enhancing spatial DSS by identifying high quality VGI data and integrating it into DSS datacube. The paper proposes a VGI quality assessment approach that evaluates VGI data regarding its completeness and its time relevance. Only high quality VGI data are integrated in DSS system. We also proposed an algorithm that presents a systematic process to increase awareness of the quality of VGI data and allows the appropriate decision about integrating—or not—VGI data into a spatial datacube to be made. The proposed approach was implemented and tested on a road maintenance project in Grand-Tunis, Tunisia: we developed the *Map-Report* prototype, we evaluated the proposed VGI quality indicator to assess VGI quality, and we evaluated the impact of integrating high VGI quality into DSS on data analysis and exploitation. The experimentation of our approach demonstrated that integrating VGI data into spatial datacubes taking into account VGI quality assessment allows to reduce considerably the error rate, and to obtain more relevant and reliable query results.

We should remember that the defined indicators do not aim at being complete, but rather making users aware of poor quality that may affects the analytical decision-making. It is recommended to consider other quality indicators, such as consistency, to enhance the assessment of VGI data quality.

Although the implementation of the proposed approach shows that selecting good VGI data based on its quality assessment is an important asset to strategic decision-making, there are still some limitations.

First, VGI quality assessment and ETL process are handled separately by two different systems, *Map-Report* and ETL system. Additionally, once assessed, VGI data are manually extracted before being loaded into datacubes. This may lead to error, and may affect the efficiency of the integrating process.

Second, decision-making requirements are manually predefined by VGI analysts in collaboration with users. However, there is neither uniformity, nor a standard way for representing these requirements. This may lead to the information heterogeneity which requires an additional time-consuming process of homogenization before being considered in the assessment.

Further research is required to combine both VGI quality assessment and ETL process within the same framework to further facilitate the integration of good VGI data into spatial datacubes. In addition, for future work, it is recommended to consider a uniform way of representing the requirements of decision-making. In order to have a uniform way of requirements representation, it is suggested to adopt spatial data representation standards, such as ISO 19109 (Rules for application schema) and 19110 (Methodology for feature cataloguing).

Author Contributions: Conceptualization, T.S. and S.A.; methodology, T.S. and S.A.; software, T.S.; validation T.S. and S.A.; formal analysis, T.S. and S.A.; investigation, T.S.; resources, T.S. and S.A.; writing—original draft preparation, T.S. and S.A.; visualization, T.S. and S.A.; supervision, T.S.; project administration, T.S. and S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: The authors thank Mohamed Radhi Guennichi for his assistance in the development of the prototype.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Crossland, M.D.; Wynne, B.E.; Perkins, W.C. Spatial decision support systems: An overview of technology and a test of efficacy. *Decis. Support Syst.* **1995**, *14*, 219–235. [\[CrossRef\]](#)
2. Turban, E.; Aronson, J.E. *Expert Systems and Intelligent Systems*; Prentice Hall: Hoboken, NJ, USA, 2001.
3. Malczewski, J. GIS-based multicriteria decision analysis: A survey of the literature. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 703–726. [\[CrossRef\]](#)
4. Rivest, S.; Bédard, Y.; Proulx, M.J.; Nadeau, M.; Hubert, F.; Pastor, J. SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data. *ISPRS J. Photogramm. Remote Sens.* **2005**, *60*, 17–33. [\[CrossRef\]](#)
5. Habibie, M.I.; Noguchi, R.; Shusuke, M.; Ahamed, T. Land suitability analysis for maize production in Indonesia using satellite remote sensing and GIS-based multicriteria decision support system. *GeoJournal* **2019**, *86*, 777–807. [\[CrossRef\]](#)
6. Svoray, T. Spatial Decision Support Systems. In *A Geoinformatics Approach to Water Erosion*; Springer: Cham, Switzerland, 2022. [\[CrossRef\]](#)
7. Chaudhuri, S.; Dayal, U. An Overview of Data Warehousing and OLAP Technology. *ACM SIGMOD Rec.* **1997**, *26*, 65–74. [\[CrossRef\]](#)

8. Bédard, Y.; Rivest, S.; Proulx, M.J. Chapter 13—Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective. In *Data Warehouses and OLAP: Concepts, Architectures and Solutions*; Wrembel, R., Koncilia, C., Eds.; IIRM Press: Hershey, PA, USA, 2007; pp. 298–319.
9. Gray, J.; Chaudhuri, S.; Bosworth, A.; Layman, A.; Reichart, D.; Venkatrao, M.; Pellow, F.; Pirahesh, H. Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals. *Data Min. Knowl. Discov.* **1997**, *1*, 29–53. [\[CrossRef\]](#)
10. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221. [\[CrossRef\]](#)
11. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plan. B Plan. Des.* **2010**, *37*, 682–703. [\[CrossRef\]](#)
12. Blaschke, T.; Merschdorf, H. Geographic information science as a multidisciplinary and multiparadigmatic field. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 196–213. [\[CrossRef\]](#)
13. Brown, G. A review of sampling effects and response bias in internet participatory mapping (PPGIS/PGIS/VGI). *Trans. GIS* **2017**, *21*, 39–56. [\[CrossRef\]](#)
14. Elwood, S.; Goodchild, M.F.; Sui, D.Z. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 571–590. [\[CrossRef\]](#)
15. Brunsdon, C.; Comber, L. Assessing the changing flowering date of the common lilac in North America: A random coefficient model approach. *GeoInformatica* **2012**, *16*, 675–690. [\[CrossRef\]](#)
16. Zhou, Q.; Zhang, Y.; Chang, K.; Brovelli, M.A. Assessing OSM building completeness for almost 13,000 cities globally. *Int. J. Digit. Earth* **2022**, *15*, 2400–2421. [\[CrossRef\]](#)
17. Sultan, J.; Ben-Haim, G.; Haunert, J.H.; Dalyot, S. Using crowdsourced volunteered geographic information for analyzing bicycle road networks. *Int. Fed. Surv.* **2015**, *14*, 1–14.
18. Keenan, P.B.; Jankowski, P. Spatial decision support systems: Three decades on. *Decis. Support Syst.* **2019**, *116*, 64–76. [\[CrossRef\]](#)
19. Horita, F.E.A.; Albuquerque, J.P.; de Degrossi, L.C.; Mendiondo, E.M.; Ueyama, J. Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks. *Comput. Geosci.* **2015**, *80*, 84–94. [\[CrossRef\]](#)
20. Sadeghi-Niaraki, A.; Jelokhani-Niaraki, M.; Choi, S.M. A Volunteered Geographic Information-Based Environmental Decision Support System for Waste Management and Decision Making. *Sustainability* **2020**, *12*, 6012. [\[CrossRef\]](#)
21. Chang, R.M.; Kauffman, R.J.; Kwon, Y. Understanding the paradigm shift to computational social science in the presence of big data. *Decis. Support Syst.* **2014**, *63*, 67–80. [\[CrossRef\]](#)
22. Andrienko, N.; Andrienko, G.; Fuchs, G.; Jankowski, P. Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. *Inf. Vis.* **2016**, *15*, 117–153. [\[CrossRef\]](#)
23. Omidipoor, M.; Jelokhani-Niaraki, M.; Moeinmehr, A.; Sadeghi-Niaraki, A.; Choi, S.M. A GIS-based decision support system for facilitating participatory urban renewal process. *Land Use Policy* **2019**, *88*, 104150. [\[CrossRef\]](#)
24. Horita, F.; de Albuquerque, J.P. An approach to support decision-making in disaster management based on volunteer geographic information (VGI) and spatial decision support systems (SDSS). In Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management, Baden-Baden, Germany, 12–15 May 2013.
25. Rajabifard, A.; Thompson, R.G.; Chen, Y. An intelligent disaster decision support system for increasing the sustainability of transport networks. *Nat. Resour. Forum* **2015**, *39*, 83–96. [\[CrossRef\]](#)
26. ISO 19157:2013; Geographic Information—Data Quality. International Organization of Standardization (ISO): Geneva, Switzerland, 2013.
27. Antoniou, V.; Skopeliti, A. Measures and indicators of VGI quality: An overview. In Proceeding of the ISPRS Annals of the Photogrammetry, Remote sensing and spatial information Science, La Grande Motte, France, 28 September–3 October 2015.
28. Senaratne, H.; Mobasher, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 139–167. [\[CrossRef\]](#)
29. Costabel, P.; del Carmen, V. *Data Freshness and Data Accuracy: A State of the Art*; Technical Report; Instituto de Computacion, Facultad de Ingenieria, Universidad de la Republica: Montevideo, Uruguay, 2006; pp. 1–36.
30. Wang, R.; Strong, D. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst.* **1996**, *12*, 5–34. [\[CrossRef\]](#)
31. Shin, B. An exploratory Investigation of System Success Factors in Data Warehousing. *J. Assoc. Inf. Syst.* **2003**, *4*, 141–170. [\[CrossRef\]](#)
32. Jackson, S.P.; Mullen, W.; Agouris, P.; Crooks, A.; Croitoru, A.; Stefanidis, A. Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 507–530. [\[CrossRef\]](#)
33. Chen, H.; Hailey, D.; Wang, N.; Yu, P. A review of data quality assessment methods for public health information systems. *Int. J. Environ. Res. Public Health* **2014**, *11*, 5170–5207. [\[CrossRef\]](#)
34. Sboui, T.; Aissi, S. A Risk-Based Approach for Enhancing the Fitness of Use of VGI. *IEEE Access* **2022**, *10*, 90995–91005. [\[CrossRef\]](#)
35. Hagenauer, J.; Helbich, M. Mining urban land use patterns from Volunteered Geographic Information by means of genetic algorithms and artificial neural networks. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 963–982. [\[CrossRef\]](#)
36. Savosin, S.; Teslya, N. Estimation and Aggregation Method of Open Data Sources for Road Accident Analysis. In *Intelligent Systems Design and Application; Lecture Notes in Networks and Systems Series*; Springer: Berlin/Heidelberg, Germany, 2022; Volume 418.

37. Devillers, R.; Bédard, Y.; Jeansoulin, R.; Moulin, B. Towards spatial data quality information analysis tools for experts assessing the fitness-for-use of spatial data. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 261–282. [\[CrossRef\]](#)
38. Pôças, I.; Gonçalves, J.; Marcos, B.; Alonso, J.; Castro, P.; Honrado, J.P. Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2356–2371. [\[CrossRef\]](#)
39. Miller, G.A.; Charles, W.G. Contextual correlates of semantic similarity. *Lang. Cogn. Process.* **1991**, *6*, 1–28. [\[CrossRef\]](#)
40. Spearman, C.E. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [\[CrossRef\]](#)
41. Malinowski, E.; Zimányi, E. *Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications*; Springer: Berlin/Heidelberg, Germany, 2008.
42. Cort, J.W.; Keni, M. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82.
43. Li, C.; Miklau, G. An adaptive mechanism for accurate query answering under differential privacy. *Proc. VLDB Endow.* **2012**, *5*, 514–525. [\[CrossRef\]](#)
44. Herlocker, J.L.; Konstan, J.A.; Terveen, L.G.; Riedl, J.T. Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Inf. Syst.* **2004**, *22*, 5–53. [\[CrossRef\]](#)
45. Hodson, T.O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.