Improving Data Value by Using Best of Worlds Approach From Government (Authoritative), Community and Commercial Data

A Case of Topographical Data in the Netherlands

Niels de Groot
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
n.e.degroot@student.utwente.nl

ABSTRACT

Topographical data is available through different data sources: authoritative data (public organizations), commercial data, and community data (Volunteered Geographical Information or VGI). VGI is user-generated and often also user-moderated. Previous research has already established that VGI has the potential to enrich authoritative topographical data, though numerous challenges are associated with their integration. This research provides a case of topographical data in The Netherlands and analyzes the current state of topographical data from different sources, looks into the feasibility of integration of these different data sets and executes a case study on the differences in data types across different sources. Both academics and NMAs have put forward concerns regarding legal issues, expertise and reliability of volunteers, continuity, and data quality. Nevertheless, the opportunities the use of volunteered geographic information provides are worth investing in mitigating these risks. A case study of data types in The Netherlands shows problems such as divergent data structures, varying frequencies of use and different levels of overlap, though these are not insurmountable given technological adaptations to data sources. Integration of or increased collaboration between community and authoritative data is possible and can lead to a well-functioning joint effort to enrich and improve topographic data in The Netherlands.

Keywords

Volunteered geographic information, VGI, authoritative data, commercial data, community data, open data, user-generated content, integration, land registration, mapping, national mapping agency, NMA, data value

1. INTRODUCTION

The Netherlands' Cadastre, Land Registry and Mapping Agency, hereafter *'Kadaster'* is data holder of the Key Register Topography (*Basisregistratie Topografie*, BRT) and Key Register Buildings & Addresses (*Basisregistratie Adressen & Gebouwen*, BAG), which encompass a wide variety of geo-information on The Netherlands, including official topographic maps and object-oriented data [14, 12]. The topographical data of the Kadaster can be considered authoritative data [4]. Additionally, there are two other sources for topographical data: commercial data (e.g. Google Maps) and community data (e.g. OpenStreetMap), also known in the scientific community as 'Volunteered Geographic Information' (VGI). [6].

Due to the introduction of the internet, a quickly growing group found interest in using it for geographic information, creating communities in which anyone can contribute to the creation or maintenance of data sets about the world around them. This, of course, does not merely hold for topographical data, but also for other forms of community-sourced data, such as Wikipedia, which have taken flight as part of the Digital Revolution. Data sets such as those of OpenStreetMap (OSM), a community-driven open data mapping platform [21], provide a relatively high accuracy [10] given that the data is not officially verified, and it may thus be of significant interest to government agencies or commercial parties.

Although previous research has already established VGI has the potential to enrich authoritative data of the BRT, numerous challenges are associated with the integration of different data sources [5]. Both academics and National Mapping Agencies (NMAs) have put forward concerns regarding legal issues, expertise and reliability of volunteers, continuity and data quality. Existing research broadly provides an overview of these problems and describes general ideas with regards to the integration of geographic data from different sources, though a perspective on The Netherlands has not been present so far. At this point, data sources are largely separated, though some integration has already been tried at various NMAs in recent years [20]. More about the theoretical background of different sources of geographic data and the possibilities and concerns of their usage can be found in section 4.

This research aims to analyze the current state of topographical data from different sources, look into the feasibility of integration of these different data sets, and execute a case study on the differences in data types across different data sources. In doing so, this research provides knowledge regarding topographical data in general, as well as its specific use in The Netherlands. Additionally, it can provide a framework from which to further explore a technologically feasible integration of or cooperation between different sources of geodata.

2. METHODOLOGY

As already briefly described in the introduction (section 1), this research can be divided into a literature review and case study. In sections 3 and 4, the current situation is discussed and linked to the theoretical background previous research provides. In section 5, several examples are investigated to find out what obstacles there are with respect to integrating data from the Kadaster and OpenStreetMap specifically. The findings of the case study and literature review are then discussed to come up with basic implementation recommendations in section 6. Conclusions of this research can then be found in section 7.

3. EXISTING SITUATION

The Kadaster is data owner to a variety of key registers within The Netherlands and is, among the general public, mostly known due to their function in registering property ownership. The Kadaster has a long history, dating back to 1820, when the first topographical map of The Netherlands was published. In 2007, when key registers are first introduced, the Kadaster was appointed as data owner for the aforementioned BRT, as well as the Key Register Kadaster (*Basisregistratie Kadaster*, BRK) [13]. Two years later, in 2009, the BAG is also added to their responsibilities. Large parts of their data have become available digitally over the years, and significant parts of it are even open data nowadays [17].

The data of the Kadaster has been aggregated and adapted over the years, which means the data set is large and the structure may have forcefully been adapted to contemporary needs. It is relevant to consider whether the Kadaster data can be included in the integration in its current form or if it should be made to fit first. Important to note as well is the way in which Kadaster operates their mapping process. There are two ways in which this can be done: by cyclical updates, also called sweep updates, which means that updates and revisions to subsets of the complete data set are done at regular intervals. The other possibility, which is often used in VGI and hence also by OSM, is continuous updating, in which updates to the complete data set are processed in real time as much as possible, often prioritizing updates to areas with a high rate of change [20].

For this research, the key registers of importance are the BRT and the BAG. These registers mainly focus on objects (buildings) and/or addresses. The BRT currently holds 81 different building types. These types are diverse, ranging from visitors centers to mosques and from sheepfolds to wind turbines [15]. The BAG divides buildings into 11 different amenity types which are more generic (e.g. living function, meeting function or office function) [11].

OpenStreetMap was founded in the United Kingdom in 2004 because the government of the UK would not make large data sets from government-run and tax-funded projects such as the Ordnance Survey freely available to the public. The OpenStreetMap Foundation has since used base maps supplied by commercial companies as well as governmental agencies such as the Kadaster to augment them with user-generated data. The result is an open source map that can be edited, but also checked by the public [22].

NMAs such as Kadaster have been curious and eager to involve the data gathered by the community in the past, and integration has been done in some instances. For example, the Kadaster featured their own VGI by collecting feedback on the BRT from users, which is then validated by surveyors from Kadaster [20]. The Kadaster's BRT and BAG do not, however, include any externally gathered data sourced from, for example, OpenStreetMap. Unfortunate, because research has suggested that VGI has serious potential to enrich, update and complement official mapping [5]. This is definitely very plausible, especially in this case, as OSM holds an entire augmented layer to the original BRT data that is currently not present in the Kadaster registers. Hence, it could be worthwhile feeding the valuable data gathered by volunteers back to the original source: the Kadaster. Nevertheless, there are also limitations in place. Both the possibilities and limitations are further described in the theoretical background under section 4.

3.1 Problem statement

If there is serious potential to be found in further integration of topographical data from different sources, the next step is to figure out how to properly implement this integration. For this, it should be established what the current state of this integration is - with some extra focus on the Kadaster in specific - and what has already been done to move towards enriching authoritative data. We would like to discover whether different data sources may use different best practices which could be useful to apply where they have not been applied before. A case study could hopefully bring these differences to light. Lastly, an implementation strategy should be discussed in which the possibilities and complications of the integration process are taken into account.

It is to be noted that the concrete implementation of the integration does not have to include merging data. There are many possibilities for linking data of one source to that of another without saving the data in duplicate. In doing so, there may also be opportunities for the Kadaster to no longer provide certain data to the public as they can reference OSM for it. However, due to concerns regarding uncertain continuity (as explained in section 4) this methodology is yet to be proven feasible.

3.2 Research question

The problem statement has led to the following research question:

> How can topographic data be enriched by the collaborative use of different data sources?

Additionally, to help answer the main research question, three subquestions have been defined:

1. What is the current state of topographic data throughout different data sources?
2. Can best practices from different data sources be shared to improve one another?
3. How can we identify overlap or discrepancies in the data models of different data sources to allow for easier integration?

4. THEORETICAL BACKGROUND

Volunteered geographical information (VGI) is a relatively new phenomenon, especially with respect to the profession of topography and cartography. The term 'VGI' was only coined by Goodchild in 2007 in response to the quickly growing interest in using the internet to "create, assemble and disseminate geographic information, provided voluntarily by individuals" [6]. OpenStreetMap and Wikimapia were good examples of such mapping communities at the time, though the former had a much larger success. Since then, many researchers from the fields of Information Management as well as Earth & Planetary Sciences have taken an interest in it.

A study by Fernandes et al (2020) forms a great summary of the developments, describing both the possibilities - in the exact same wording as Goodchild 13 years earlier - as well as the limitations that come forward [5]. One of those concerns is about data privacy and ethical use, which are of great significance in a society that is increasingly data-driven. What data can be gathered (through VGI) without permission? What is the data used for? In 2010, the German government insisted that Google gave their citizens a chance to opt out of their Street View program to protect their privacy [1]. This example goes to show that documenting potentially privacy infringing geographic information can sometimes be met with dissatisfaction.

Though VGI may be volunteered information, it is gathered by a host such as OpenStreetMap. Copyright ownership often stays with the volunteer due to the use of open source software licenses, though as part of such a license special rights may be deferred to the host platform. With the integration of multiple data sources, the ownership and, with that, the responsibility for the data may become unclear, which has consequences regarding liability if the data causes harm by being inaccurate, for example [20]. Other legal issues may also arise from the location and jurisdiction from which the host platform operates. In response to the United Kingdom leaving the European Union in 2020, OSM has been considering relocating to the EU due to, among others, mutual recognition of database rights between the UK and the EU [9].

Large VGI platforms have millions of users who contribute on a regular basis, however, they are not trained professionals, nor are they employed or do they have any legal agreement with the host platform that obliges them to contribute to their data. As VGI platforms are publicly available, it is almost impossible to control who makes edits to the data, which brings the question whether authoritative data owners such as NMAs can safely rely on using the data from these community sources. Presumably, making wrong edits to volunteered geographical data with malicious intent does not happen often, though mistakes can easily slip in if the editor lacks knowledge or expertise. As OSM users vary significantly in activity, interest and motivation [3, 19], edits may be made by professionals and experts just as well as complete laymen.

This effect, however, is minimized by the use of continuous peer reviewing [23]. Anyone may contribute to community data on OpenStreetMap, but likewise, anyone can also correct other contributions. In a study by Parker [24] using focus groups, participants indicated that they found VGI more accurate than PGI (professional geographic information), especially due to the inclusion of (more) real time information. Other studies have also found remarkably good positional accuracy scores with respect to OpenStreetMap [10].

So far, a large number of researches give context for the acceptance (or lack thereof) of community data or OpenStreetMap in particular for integration into authoritative data. Little attention is paid, however, towards commercial parties. Commercial mapping data is proprietary to large (internet) companies such as Bing or Google and is therefore a lot less accessible to the public. As these companies are driven by their own commercial interests and need of capital, the data is often only partially presented to the end users of platforms such as Google Maps and the possibilities to adjust it are limited [25]. Google Maps does allow VGI through their Local Guide program [8], but the possibilities for volunteers are more limited than those of OSM and Google retains legal rights, such as copyright, on their data. Due to the earlier mentioned restricted publication of data, it is hard to get insight into how, for example, Google Maps handles its information management. This is also the reason that commercial data is not used in the case study in section 5.

Despite the concerns, the integration of open data and VGI has already had the attention of NMAs in recent years. Bégin (2012) describes how the Canadian NMA could receive updates from community data after providing contributors with a base data set of their own authoritative data [2], and in a survey and subsequent workshops with NMAs from across Europe and Greenland about half of the participants said to work with VGI in one or multiple ways. However, only three NMAs (Finland, Germany and Greece) used VGI for new data collection through, for example, OpenStreetMap [20]. The concerns of the NMAs about the use of VGI largely aligns with those highlighted by Fernandes et al.

5. CASE STUDY

As described in section 3.1, a case study could shed light on best practices used in one data source that may be worthwhile to apply for others as well. Additionally, when looking into the potential value of data from commercial and community sources to enrich authoritative data, it is important to establish whether this data has the same shape and form. If data from different sources has a different structure, uses different data types or uses different definitions, universalization could be necessary. This is not always possible, as in some data sets the data one needs to make it compatible with other sources may not be available. We will first look into the data types in general and then use some test cases to identify complications in aggregating data from different sources.

5.1 Available data & limitations

It should first be established what data is available. This is relatively easy for authoritative data, as most of it is open data and fairly transparent. On the website of the Kadaster, one can easily find all the different types that are used in the BRT. For community data - in this case, OpenStreetMap, the most used community mapping tool - the process is similarly easy, as all different types and their usages are publicly available. However, for commercial data, there are quite some limitations. Although the Google Maps UI is available for free to any internet user, the backbone of their system is not easily accessible for research purposes. Some documentation of their (paid) API can be found and, because of their Local Guide program, a list of data types for amenities can be found, but querying their database is an expensive business. Of course, commercial companies have a commercial interest at heart. With a view to their competitive advantage, they are less willing to share details about their data processing. Due to these limitations, however, it was eventually decided not to include Google Maps in this case study.

5.2 Data types, Buildings & Amenities

One of the first things that stand out when researching the different data types is not necessarily the differences between the types themselves, but between their usage and frequency. Whereas OpenStreetMap distinguishes a lot

Table 1: Overview of building and amenity types of objects in Kadaster and OpenStreetMap databases

Kadaster	Building (BRT)	81
	Amenity (BAG)	11
OSM	Building	78
	Amenity	120

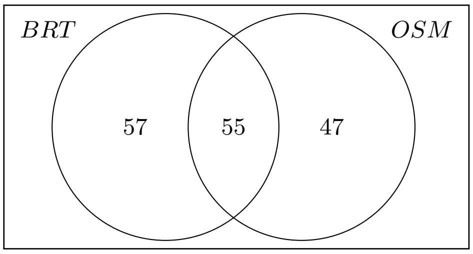


Figure 1: Number of unique and shared building types between BRT and OSM. $BRT \cap OSM$ refers to the items that can be mapped to the other data source. Note that types may have a one-to-many relationship. More information can be found in appendix B.

between the building type and the amenity of said building (e.g. a former church that is now being used as a restaurant has building type 'church' but amenity 'restaurant'[1]), the BRT and BAG do not always make such clear distinctions. This makes it difficult to properly compare types. If we also take into account that one data source provides subtypes or details that are missing in the other sources, the task becomes increasingly difficult.

The number of different available types throughout the different data sources was already briefly mentioned in 3, but can also be found in table 1. Though the OSM building types and the BRT building types overlap, some types are unique to just one data set. This is demonstrated in figure 1, which shows that there are 55 building types that can be found in both data sets. However, the BRT still includes 57 building types that could be not be matched to those of OSM. OSM also has 47 types that are unique to its data set, though some nuance must be added, namely that the abstraction levels do not seem comparative. This is discussed further in section 5.3.

OpenStreetMap provides a lot more amenity types than the BAG, from which we can see that the BAG data could be enriched by augmenting the data with that of OSM, which has a much lower abstraction level. To get an idea of the possibilities, we have sorted all OSM amenity types by BAG type, using the practical guidelines of the BAG [16]. The results can be found in appendix A. We notice that the OSM types are disproportionately distributed, with one third of the types being mapped to a meeting function. This can be interpreted as an indication that these types could be split into smaller and clearer categories. Furthermore, almost 30 percent of the data types from OSM could not be mapped to a category and are thus mapped to 'Other'. An explanation for this is that it is probable that the Kadaster and OSM map many of the objects in this category differently; Kadaster includes smaller points on a map that may not be considered objects in the Key Register Large-scale Topography (*Basisregistratie Grootschalige Topografie*, BGT), as opposed to the BRT or BAG.

[1]The example used concerns the *Jopenkerk* in Haarlem, The Netherlands.

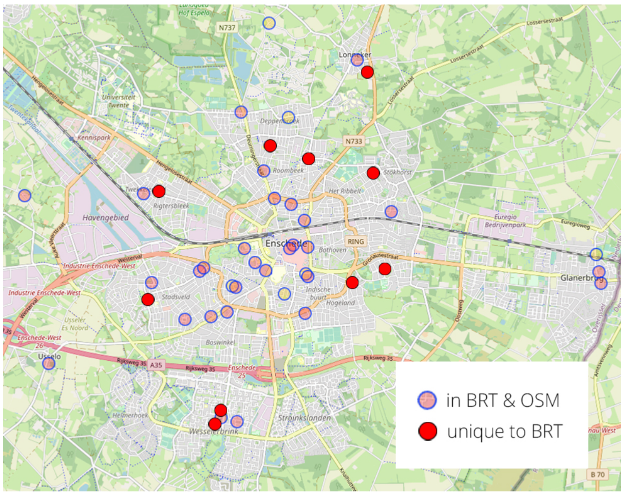


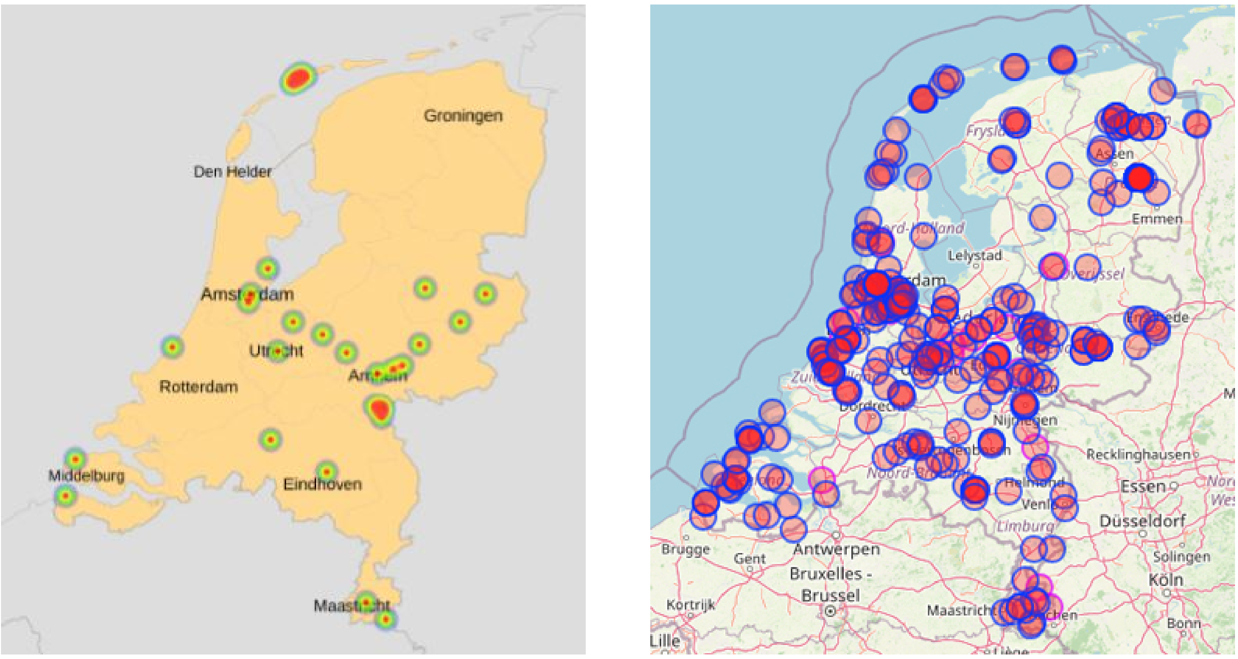
Figure 2: Overview of objects with building type 'Church' in Enschede.

Another case concerns churches (see figure 2). To keep it rather simple, we limit the search area to the city of Enschede and its neighboring town Glanerbrug. Using a query, one can find that the BRT includes 45 buildings of this type in its register. After a quick manual rundown, it can be concluded that this includes various religions, as the mosque and synagogue are also included. OpenStreetMap displays a mere ten churches in the same area if we search for building types, and these only include Christian churches. If we search for amenity, 36 churches in the area pop up, this time, however, also including other religions. If we look into the nine objects that differ between BRT and OSM, almost all of them are no longer acting as a church - though they were originally built with that intention and have often also been used as a church for a long time - but are now, for example, schools or offices.

Additionally, problems may arise because of the frequency of use of different data types. For a comparison, we look into hotels in The Netherlands. The BRT includes a building type 'hotel', however, this only returns 34 hotels in the entire country (see figure 3a). OSM on the other hand identifies 529 buildings identified as hotels (see figure 3b) which, according to their definition, are either still in use as a hotel or were originally used as or intended as hotel. When looking into some local cases, most hotels that are not included in the BRT search simply have type 'Building' (default type) where the BAG provides a meeting and/or lodging function as amenity. The fact that some hotels do not have building type hotel can be explained, as they were built with another purpose in mind. A hotel in a former farmhouse should not be registered as building type hotel. However, it is hard to believe only 34 buildings in The Netherlands were constructed specifically as hotel. In fact, even some newly built buildings with no other function than hotel are not correctly registered in the BRT and several test cases return the default type.

5.3 Findings

This subsection is concerned with the findings of the case study in particular. It includes interpretation of the results of the investigated cases and discusses their relevance to the (potential of) integration of topographical data from different data sets. Though these findings may already hint towards conclusions, an overview of conclusions can be found in section 7.



(a) BRT, $n = 34$ (b) OSM, $n = 529$

Figure 3: Comparison of objects with building type 'Hotel' in The Netherlands

An analysis of the data types of Kadaster (BRT & BAG) and OSM shows that there are still some practical hurdles to overcome. Though theoretically the division of building types in the BRT and amenity types in the BAG makes sense, it does not entirely match how OSM has defined its types, which makes for difficult matching. For building types, we most notably experience a disconnect given the strict designation rules of the BRT (only label a building according to their original purpose) and the fact that many buildings are only labeled with the default type. In fact, according to a query result, over 20,000 objects in The Netherlands are categorized with a default or non-specific building type. For amenities, we notice that many amenities from OSM can not be properly mapped to the restrictive framework of only 11 amenity types in the BAG.

Through test cases we identified further problems which both substantiated and strengthened the earlier findings and added new conclusions. Most notably: though there is a slight difference in definition of building types between the two data sources, the number of objects found (34 vs 529) are too far apart to only be a consequence of this definition. A clear connection between the objects where there was a match and where there was not has not yet been discovered. Possible reasons for this dissimilarity could include outdated information at the Kadaster, missing information from branch organizations or other government bodies or minor details in definition.

The comparison between churches was done at a more local level as to more easily identify *why* certain objects differ in type. Though the dissimilarity between the two data sets was not striking here (about 80 percent match), we conclude that those cases without a match were the result of outdated or incorrect information. This may be due to the cyclical nature of the Kadaster process, whereas OSM uses a continuous process with the help of their volunteers.

We can conclude from this small case study that there are starting points for an integration. However, extra care is required when comparing and connecting the data sets, given their different structure. The next section will go into a basic implementation plan and discuss considerations for this implementation.

6. IMPLEMENTATION & DISCUSSION

With the help of the aforementioned cases we have established ways in which community data (OSM) can be used to enrich the existing authoritative data. To make sure the data from different sources can add mutual value, we need to match the correct objects to each other. How can we aggregate the data and verify that it is correct, and what process could be used to keep a continuous link between these data sources as opposed to using cyclical updates?

6.1 Aggregation & verification

Given the different data structure, it is hard to match the different points of interest (POIs) by type. After all, the methodologies for assigning types are very different, so one cannot always easily find overlap. A different approach, in which one overarching characteristic can be used, is therefore needed. In cartography, determining items on a map relies heavily on three different concepts: points, lines or ways, and polygons or objects. A point may display something small at a concrete position: a postbox, an electricity pole or a fire hydrant. A line or way can represent a road, train rails or waterway, and these may be connected to form a network. A polygon can represent areas, such as buildings, parks or larger regions such as counties, cities, provinces or nations. As we are largely interested in mapping buildings at this time, the focus is on objects and we leave points and ways outside the scope of this research.

In the ideal case, every object that exists in the real world is present in both data sets, though it may be categorized or labeled differently. Returning to the example of the churches: every single one of the 45 objects identified by the BRT can be found in OSM as well, though for some of them OSM may provide different or additional information. If it is additional information, e.g. in the BRT an object is known to be a church building, but the BRT is unaware of its current use as office building (amenity), this information can be verified and added to the authoritative data. The same holds for the example of the hotels: if OSM can provide more objects with building type hotel, these can be verified and added. The situation becomes more complex if both sources provide information on the same variables, but the information differs. If OSM claims a building has amenity 'school' and the BAG claims it has an office function, only one can be correct. There is a multitude of possibilities for this discrepancy: one claim can be outdated, the amenity of the object may be complex and debatable, or controversial, or there may be a different level of abstraction due to which subtypes are not easily aggregated. Naturally, if an object cannot be found in both data sets, one would first need to identify why this is the case. Could the object have been demolished recently, which either one of the sources has not yet processed, for example?

The aforementioned verification steps are not to be taken lightly. One of the main concerns regarding the aggregation of volunteered geographic information into authoritative data is accuracy and validity [10, 7]. However, as already mentioned in section 4, the accuracy of volunteered geographic information is generally quite high due to - among others - community verification. The verification process relies on two factors: the Kadaster and the community. Kadaster itself of course already has a process for verifying information, and OpenStreetMap relies heavily on (multiple) other users to objectively check and verify information.

This is where the distinction between a cyclical and continuous process (as already introduced in section 3) comes up again. The Kadaster relies on a periodical review of their data by field and office workers. OSM on the other hand can depend on millions of users, of which a relatively high percentage from The Netherlands [18] to add, edit and correct information in their data set, at any moment in time. With the integration of community data into Kadaster data, it would be worthwhile for the Kadaster to increasingly adapt to a continuous process with respect to object data. This recommendation is also in line with research conclusions (see section 4).

In addition to the adaptation of a more continuous process, Kadaster would do well to adopt other strategic choices from OpenStreetMap as well. The focus on volunteers and peer reviewing in particular can save Kadaster time and costs, although the consequences for their dependency and continuity, as well as the personnel policy of Kadaster, must of course be kept in mind. Additionally, legal steps would have to be undertaken to protect the position of the Kadaster and to ensure that data gathered or reviewed by volunteers is responsibly and appropriately dealt with. OpenStreetMap already has significant experience in this regard, and could provide a proven framework if necessary. To take the step towards extension of VGI towards data gathering, the existing feedback system can be used, which can be expanded, or an even more in-depth integration - or collaboration (in line with the arguments of section 3.1) - with OSM data may be considered. Although the exact technical implementation is outside the scope of this research, it is probable technical adjustments to the Kadaster IT are a necessity.

However, the most important factor in a transition may actually be that of trust. As already discussed in section 4: research among actors in the field of topography has indicated that the data quality of community data is a great concern which has prevented government organizations from swiftly implementing it in the past [20]. On the other hand, research has also shown that the quality of VGI is not substandard either [10]. It is vital that NMAs and other bodies trust that VGI has added value and can be a valuable addition to their topographic data.

6.2 Synchronization

When the integration of the current information from OSM is first completed, it would be worthwhile to the Kadaster to make sure updates continue to be received when new insights are provided by OSM. This does not prevent the Kadaster from also doing periodic checks on the data themselves. In fact, it has added value that any errors noticed by the Kadaster can also be easily forwarded to OSM. As mentioned previously, the technical implementation of the implementation as well as synchronization of the data is outside the scope of this research. It could, however, be part of future research, either academically or by a private research firm.

7. CONCLUSION

From the the theoretical background (section 4) and the executed case study (section 5), we can now answer the research questions posed in section 3.2.

RQ1: What is the current state of topographic data throughout different data sources?

Topographic data has improved a lot over the past decades as a consequence of increasing technical possibilities. Due to the introduction of the internet, a quickly growing community found interest in using it for geographic information. OpenStreetMap has become one of the largest sources of volunteered geographic information. With its wide-reaching augmented layer of user-generated object information, it is a valuable data source for National Mapping Agencies to improve their data sets. However, both academics and agencies themselves have put forward concerns regarding legal issues, expertise and reliability of volunteers, continuity, and data quality. Nevertheless, the opportunities the use of volunteered geographic information provides make it worth it to try and overcome these limitations. In addition to community data, large commercial parties such as Google also provide incredible amounts of geographic data through their products (such as Google Maps). This data is often gathered using methodologies similar to authoritative and community data and may use VGI as well. Unfortunately, the data is proprietary and therefore legal and financial limitations make it less viable data for integration with NMAs.
Community data uses a variety of best practices that are different from a traditional authoritative approach, but are valuable to introduce at NMAs. Additionally, authoritative data organizations such as NMAs offer knowledge that may help increase the stability and reliability of community data platforms. RQ2 explains how sharing best practices can improve data value.

RQ2: Can best practices from different data sources be shared to improve one another?

Due to the different zeitgeist and function of authoritative data and community data, the two sources are misaligned in some part. Whereas authoritative data is often not meant for the general public to have direct interaction with and is mostly used as source for other applications, data of OpenStreetMap is built with that idea at the very center of it. Though NMAs such as the Kadaster have started to use forms of VGI in the past, only very few make use of VGI for actual data gathering. The best practices gathered by community data platforms are therefore very valuable to NMAs and it is recommendable for them to seek active communication with people from the world of VGI to facilitate a smooth integration process. A selection of conclusions for the implementation of best practices can be found below:

- **Time and costs can be saved by usage of a peer reviewing process such as OSM has.** Dividing the workload of reviewing over a large number of volunteers saves time for NMA personnel. This personnel could potentially be used for less frequent but more in-depth quality control.

- **The use of volunteered geographic information allows for a more direct connection to the real-life area being mapped**, as the many volunteers are spread out throughout the geographical area being mapped and may observe it on a daily basis in their lives.

- **Understandable data architecture** as used by open data platforms allows for easier management of topographical data and reflects the real-life situation in a more accurate way.

- **Possible limitations of volunteered geographical information can be mitigated by linking with authoritative data**, as opposed to a complete integration of data. This is easy to achieve, as in essence authoritative data is always used as base data for additions and edits made on open data platforms.

As one may notice, the easiest way of achieving a smooth integration is foreseen by letting authoritative data adapt to community data, instead of the other way around. Nevertheless, the pillars of validity and accuracy that belong to the authoritative nature of NMAs should not be should not be affected by this process. Fortunately, few problems are expected, because open data platforms make mutual use of the data made available by NMAs. As ultimately both want to provide the most accurate and up-to-date data, this is advantageous to the cause.
Though possibilities have been identified to improve data value through sharing best practices, practical overlap and discrepancies between data sets is key to a working strategy. RQ3 elaborates on identifying overlap and discrepancies and discusses their use in integration of data sources.

RQ3: How can we identify overlap or discrepancies in the data models of different data sources to allow for easier integration?

As demonstrated in the brief case study in section 5, there are practical burdens because of different approaches to data gathering and information management. A shortlist of the identified problems can be found below.

- **The data structure of the Kadaster is significantly different from OpenStreetMap.** As the Kadaster is split in multiple registers, different parts of OSM are hard to align with specific registers. A model to transfer data between the two sources is currently missing and hard to create without some adaptations on either side, though most notably at the Kadaster.

- **The frequency of data types differs significantly through largely unknown reasons.** Further research into numerical statistics of data types could be advantageous and might proof essential to a working model to transfer data between these sources.

- **The levels of overlap differ immensely throughout the data sets.** Some data types can be easily matched appropriately, though others depend on technological implementations to be altered in some way. This makes it a case with very divergent problem levels, for which a general approach may not be suitable.

These burdens are fortunately not insuperable and can often be overcome through minor adaptations of the traditional NMA process. It does require some confidence in the far less traditional and - some may even say - ‘anarchistic’ approach of letting any stranger on the internet contribute to the quality of the world’s cartography. The processes in place at community data platforms for data verification have been increasingly successful over the years and are a promising sign of the potential of VGI in modern cartography.
Unfortunately, a wider analysis of the (statistics of) data types of both sources is outside the scope of this research. However, the belief is that extending this kind of research to a larger scale will easily address discrepancies in the data structures of the different data sets. This, however, does not mean that there is no overlap. In fact, the core principles of the two sources align perfectly. As said, these discrepancies are not impossible to overcome, though may require adaptation, mostly from the side of the NMAs. In the future, the Kadaster data may be largely sources by means of VGI, which makes it probable that at some point large discrepancies will be resolved and, unless the two data sources are significantly disengaged again, will not easily return as topographic information will fuse together more and more.

Main RQ: How can topographic data be enriched by the collaborative use of different data sources?

As already concluded by previous research, the potential of collaborating with topographic data from different sources is present. Many of the known concerns also apply to the situation in The Netherlands specifically. However, they are not insurmountable. Given technological adaptations, integration of or increased collaboration between community and authoritative data is possible and can lead to a well-functioning joint effort to enrich and improve topographic data in The Netherlands.

8. ACKNOWLEDGMENTS

I want to thank my supervisor, dr.ir. Erwin Folmer for his friendly and helpful guidance and his colleagues at Kadaster for providing their knowledge and expertise where necessary. Special thanks also goes out to my girlfriend, my family, and friends, who have acted as sparring partners and motivators whenever I was stuck while working on this research project.

9. REFERENCES

[1] Thousands of Germans opt out of Google Street View. *BBC*, Oct 2010. Accessed: 2022-01-13.

[2] D. Bégin. Towards integrating VGI and national mapping agency operations: A Canadian case study. In *Proceedings of the GIScience Conference, Columbus, OH, USA*, pages 18–21, 2012.

[3] D. Coleman, Y. Georgiadou, and J. Labonte. Volunteered Geographic Information: The nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research*, 4(4):332–358, 2009.

[4] J. Crompvoets, S. Wouters, M. Chantillon, D. Kopczewski, M. Cory, C. Agius, and S. Grimmelikhuijsen. Authoritative data in a European context. *European Spatial Data Research. Official Publication*, 72:1–31, 2019.

[5] V. O. Fernandes, E. N. Elias, and A. Zipf. Integration of authoritative and Volunteered Geographic Information for updating urban mapping: Challenges and potentials. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B4-2020:261–268, 2020.

[6] M. F. Goodchild. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.

[7] M. F. Goodchild and L. Li. Assuring the quality of volunteered geographic information. *Spatial statistics*, 1:110–120, 2012.

[8] Google. Overview - Local Guides Help. `https://support.google.com/local-guides/answer/6225846?hl=en&ref_topic=6225845#zippy=`. Accessed: 2021-11-25.

[9] OpenStreetMap looks to relocate to EU due to Brexit limitations. *The Guardian*, Jun 2021. Accessed: 2022-01-13.

[10] M. Haklay. How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance survey datasets. *Environment and Planning B: Planning and Design*, 37(4):682–703, 2010.

[11] Kadaster. BAG datamodel. `https://bag.basisregistraties.overheid.nl/datamodel`. Accessed: 2022-01-11.

[12] Kadaster. Basisregistratie Adressen en Gebouwen - BAG. `https://www.kadaster.nl/zakelijk/registraties/basisregistraties/bag`. Accessed: 2021-11-25.

[13] Kadaster. Basisregistratie Kadaster - BRK. `https://www.kadaster.nl/zakelijk/registraties/basisregistraties/brk`. Accessed: 2022-01-12.

[14] Kadaster. Basisregistratie Topografie (BRT). `https://www.kadaster.nl/zakelijk/registraties/basisregistraties/brt`. Accessed: 2021-11-19.

[15] Kadaster. BRT datamodel. `https://brt.basisregistraties.overheid.nl/datamodel`. Accessed: 2021-11-25.

[16] Kadaster. Catalogus Basisregistratie Adressen en Gebouwen - 2018. `https://www.geobasisregistraties.nl/binaries/basisregistraties-ienm/documenten/publicatie/2018/03/12/catalogus-2018/Catalogus-BAG-2018.pdf`, Jul 2018. Accessed: 2022-01-15.

[17] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Digitale Kadastrale kaart beschikbaar als open data! `https://www.geobasisregistraties.nl/actueel/nieuws/2016/01/19/digitale-kadastrale-kaart-beschikbaar-als-open-data`, Jan 2016. Accessed: 2022-01-12.

[18] P. Neis. Osmstats - edits per country. `https://osmstats.neis-one.org/?item=countries`. Accessed: 2022-01-16.

[19] P. Neis and A. Zipf. Analyzing the contributor activity of a Volunteered Geographic Information project—The case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1(2):146–165, 2012.

[20] A.-M. Olteanu-Raimond, G. Hart, G. M. Foody, G. Touya, T. Kellenberger, and D. Demetriou. The scale of VGI in map production: A perspective on European National Mapping Agencies. *Transactions in GIS*, 21(1):74–90, 2016.

[21] OpenStreetMap. About OpenStreetMap. `https://wiki.openstreetmap.org/wiki/About_OpenStreetMap`. Accessed: 2021-11-25.

[22] OpenStreetMap. History of openstreetmap. `https://wiki.openstreetmap.org/wiki/History_of_OpenStreetMap`. Accessed: 2022-01-21.

[23] OpenStreetMap. Verifiability. `https://wiki.openstreetmap.org/wiki/Verifiability`. Accessed: 2021-11-25.

[24] C. J. Parker, A. May, and V. Mitchell. Understanding design with VGI using an information relevance framework. *Transactions in GIS*, 16(4):545–560, 2012.

[25] S. Quinn and L. Yapa. OpenStreetMap and food security: A case study in the city of Philadelphia. *The Professional Geographer*, 68(2):271–280, 2016.

APPENDIX

A. AMENITY TYPES

Table 2: Overview of OSM amenity types sorted by BAG type. Note that the 'Other' category may also contain items that would have otherwise remained unmatched.

Function (NL)	Function (EN)	Subcategories
Wonen	*Living*	refugee_site
Bijeenkomst	*Meeting*	bar, biergarten, cafe, fast_food, food_court, ice_cream, pub, restaurant, library, toy_library, social_facility, arts_centre, brothel, casino, cinema, community_centre, conference_centre, events_venue, gambling, nightclub, planetarium, public_bookcase, social_centre, swingerclub, theatre, stripclub, music_school, courthouse, bbq, bench, shelter, childcare, funeral_hall, grave_yard, internet_cafe, monastery, place_of_mourning, place_of_worship, public_bath, public_building
Cel	*Cell*	prison
Gezondheidszorg	*Healthcare*	baby_hatch, clinic, dentist, doctors, hospital, nursing_home, pharmacy, veterinary
Industrie	*Industry*	ferry_terminal, post_depot, post_box
Kantoor	*Office*	studio, ranger_station, townhall, embassy
Logies	*Lodging*	love_hotel
Onderwijs	*Education*	college, kindergarten, language_school, school, university
Sport	*Sports*	dive_centre, gym, hunting_stand
Winkel	*Commerce*	kick-scooter_rental, bicycle_repair_station, bicycle_rental, boat_rental, car_wash, vehicle_inspection, fuel, atm, bank, bureau_de_change, driving_school, taxi, freeshop, crematorium, marketplace, photo_booth, vending_machine, car_rental, post_office
Overig	*Other*	bus_station, car_sharing, boat_sharing, charging_station, grit_bin, motorcycle_parking, bicycle_parking, parking, parking_entrance, parking_space, fountain, fire_station, police, dog_toilet, drinking_water, give_box, shower, telephone, toilets, water_point, watering_place, sanitary_dump_station, recycling, waste_basket, waste_disposal, waste_transfer_station, animal_boarding, animal_breeding, animal_shelter, baking_oven, clock, kitchen, kneipp_water_cure, lounger, user_defined

B. BUILDING TYPES

A numerical overview can be found in table 5 and figure 1 (page 4). Lists of unmatched and matched building types can be found in tables 3, 4 and 6. Translations of Dutch names to English were done by Google Translate with manual error correction *ex post*.

Table 3: Overview of unmatched BRT building types.

BRT type (NL)	BRT type (EN)
Bezoekerscentrum, Boortoren, Brandtoren, Crematorium, Dok, Elektriciteitscentrale, Fabriek, Fort, Gemaal, Gemeentehuis, Gevangenis, Kasteel, KasWarenhuis, Kerncentrale Kernreactor, Kliniek Inrichting Sanatorium, Klokkentoren, Koeltoren, Koepel, Kunstijsbaan, Lichttoren, Luchtwachttoren, MarkantGebouw, Museum, Observatorium, OverigGebouw, Paleis, Peilmeetstation, Politiebureau, Pompstation, Postkantoor, PsychiatrischZiekenhuis PsychiatrischCentrum, Radarpost, Radartoren, Radiotoren Televisietoren, Recreatiecentrum, Reddingboothuisje, Remise, Schoorsteen, Stadskantoor Hulpsecretarie, Tank, Tankstation, Telecommunicatietoren, TolGebouw, Toren, Uitzichttoren, Veiling, Verkeerstoren, Vuurtoren, Waterradmolen, Wegrestaurant, Werf, Windmolen, Windmolen Korenmolen, Windmolen Watermolen, Windturbine, Zendtoren, Zwembad	Visitor center, Derrick, Fire tower, Crematorium, Dock, Power station, Factory, Fort, Pumping station, City Hall, Jail, Castle, Warehouse, NuclearPowerGenerator NuclearReactor, Clinic Facility Sanatorium, Bell tower, Cooling tower, Dome, Artificial ice skating rink, Light tower, Air guard tower, Striking building, Museum, Observatory, OtherBuilding, Palace, Toll measurement station, Police station, Pumping station, Post office, PsychiatricHospital PsychiatricCenter, Radarpost, Radiator, Radiotower TelevisionTower, Recreation center, Lifeboat house, Remise, Chimney, CityOffice AssistanceSecretary, Tank, Gas station, Telecommunication tower, Toll, Tower, View tower, Auction, Traffic tower, Lighthouse, Water wheel mill, Roadside restaurant, Yard, Windmill, Windmill Flourmill, Windmill Watermill, Wind turbine, Tower, Pool

Table 4: Overview of unmatched OSM building types.

OSM type
bakehouse, bridge, bungalow, cabin, carport, civic, commercial, conservatory, construction, container, cowshed, digester, dormitory, farm, farm_auxiliary, garage, garages, gatehouse, ger, government, grandstand, greenhouse, houseboat, hut, industrial, kindergarten, kiosk, hangar, office, pavilion, presbytery, public, residential, retail, roof, service, shed, slurry_tank, static_caravan, sty, supermarket, tent, terrace, toilets, transportation, tree_house, warehouse

Table 5: Numerical overview of building types in BRT & OSM.

	BRT	OSM	Total
Shared	24	31	$\boxed{55}$
Unique	57	47	
Total	81	78	159

Table 6: Overview of OSM building types sorted by BRT type.

BRT types (NL)	**BRT types (EN)**	**Associated OSM types**
Brandweerkazerne	*FireStation*	fire_station
Bunker	*Bunker*	bunker
Hotel	*Hotel*	hotel
Huizenblok	*HouseBlock*	apartments, house, residential, semidetached_house, detached
Kapel	*Chapel*[1]	chapel, religious
Kerk	*Church*[1]	church, religious, temple, shrine, cathedral
Klooster Abdij	*MonasticAbbey*	monastery
Manege	*RidingSchool*	riding_hall
Militair Gebouw	*MilitaryBuilding*	military
Moskee	*Mosque*[1]	mosque, religious, temple, shrine
Overig Religieus Gebouw[1]	*OtherReligiousBuilding*[1]	cathedral, religious, temple, shrine
Parkeerdak Parkeerdek Parkeergarage	*ParkingRoof ParkingPlace Parking-Garage*	parking
Rune	*Rune*	ruins
Schaapskooi	*Sheepfold*	stable
School	*School*	school
Silo	*Silo*	barn
Sporthal	*Sports hall*	sports_hall
Stadion	*Stadium*	stadium
Stationsgebouw	*Station building*	train_station
Synagoge	*Synagogue*[1]	synagogue, religious, temple, shrine
Transformatorstation	*TransformerStation*	transformer_tower
Universiteit	*University*	university, college
Watertoren	*Watertower*	water_tower
Ziekenhuis	*Hospital*	hospital

[1]Religious OSM building types are matched to a multitude of BRT types, as OSM generally gives most places of worship the default building type '*Building* $= *$' and specifies its religion by using the appropriate amenity. The building types for places of worship are therefore rarely used in OSM.