Boosting Data Quality by a Participatory Approach: The Success of Quality Dashboards and Crowdsourced Feedback

Anouk HUISMAN-VAN ZIJP and Jaap-Willem SJOUKEMA, The Netherlands

Key words: data quality, dashboards, quality indicators, crowdsourcing, volunteered geographic information, feedback, key registers.

SUMMARY

The Netherlands has a system of key registers, which are governmental registers of authentic data to be used by all public institutions in fulfilling their public tasks. Data quality is key for the users of these registers as poor data quality can have serious consequences such as wrong tax assessments or unjustified building permits.

The Netherlands' Cadastre, Land Registry and Mapping Agency – in short Kadaster – has successfully experimented with two participatory systems for improving the data quality of three key registers: quality dashboards and crowdsourced feedback. Both systems have been in place for over four years, resulting in a data quality boost for these datasets.

Kadaster has developed the quality indicator dashboard in close consultation with the key register producers to provide insight in the data quality of a key register. The dashboard assists a data producer in quality management. Alongside the quality dashboard for data owners, there is also a public version for data users.

Next to the quality indicator dashboards, Kadaster has made it possible for the public to get involved in the improvement of register data through crowdsourcing. At dedicated websites anyone can report a change or error in the data by clicking on the map, adding a comment and adding supporting information.

This paper will show the results of four years of quality dashboards and crowdsourced feedback on the Dutch topographic and addresses key registers.

Boosting Data Quality by a Participatory Approach: The Success of Quality Dashboards and Crowdsourced Feedback

Anouk HUISMAN-VAN ZIJP and Jaap-Willem SJOUKEMA, The Netherlands

1. INTRODUCTION

In the Netherlands, a system of key registers has been set up. The key registers are government registers of authentic data which are mandatory to use by all public institutions in fulfilling their public tasks. This system of key registers operates on the principle "collect data once, use it many times". This reduces the administrative burdens for citizens and businesses and at the same time quality improvements and cost savings for the government itself are realized. Figure 1 shows the system of key registers and the links between the different registers (Digitale Overheid, 2020).

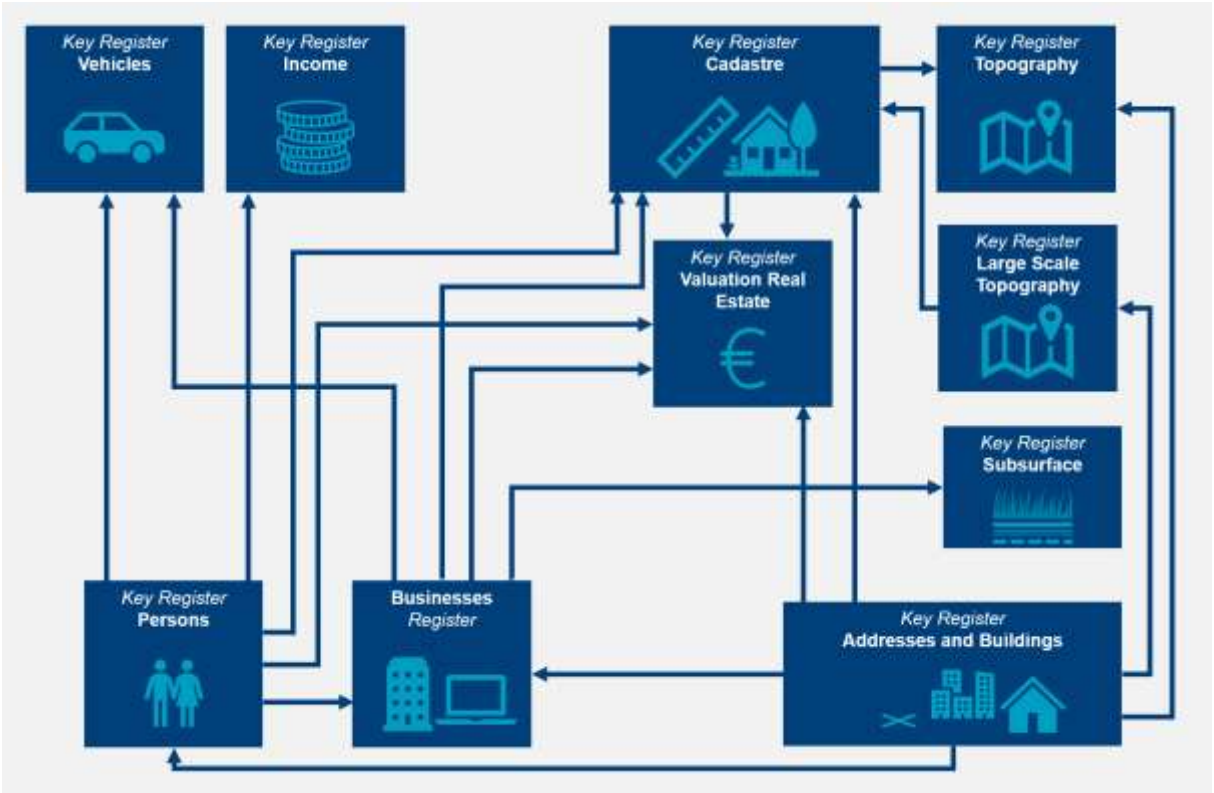


Figure 1: System of key registers in The Netherlands.

Different roles can be defined in the establishment of these registers, e.g. collecting data, storing data, disseminating data and data quality management. These roles are allocated to different governmental organisations. The Netherlands' Cadastre, Land Registry and Mapping

Agency – in short Kadaster – fulfils multiple roles for the spatial key registers. This paper will focus on the role of quality management and its instruments.

The world is constantly changing. Therefore, key registers need to change as well to maintain data quality on a sufficient level. This is not an easy task and a continuous effort, because low data quality can have serious consequences such as wrong tax assessments or unjustified building permits.

Kadaster has successfully experimented with two participatory systems for improving the data quality of the topographic and addresses key registries: quality dashboards and crowdsourced feedback. Both systems have now been in place for over four years, resulting in a data quality boost for these three datasets. The next section will shortly explain these key registers.

1.1 Spatial key registers

This paragraph will give a short introduction to the topographic and addresses key registers. For these key registers, the Kadaster has developed the quality dashboards and a crowdsourced feedback system. All three registers are provided as open data, making them highly suitable for participatory systems. Figure 2 shows images of these three key registers.

Figure 2 a. Key Register Addresses and Buildings (BAG), *b. Key Register Large Scale Topography* (BGT), *c. Key Register Topography* (BRT).

Key Register Addresses and Buildings (BAG): Building and address data. Unique objects with address, area, contour and purpose. The data is produced and maintained by 352 municipalities. The register is updated by municipalities continuously. Kadaster is register keeper and responsible for quality management.

Key Register Large Scale Topography (BGT): Large scale topographic data, scale 1:500 – 1:5000. The data is produced and maintained by almost 400 local, regional and national authorities. The register is updated by data producers continuously. Kadaster is register keeper and from the start of 2021 responsible for quality management.

Key Register Topography (BRT): Small scale topographic data and maps, scales 1:10.000 – 1:1.000.000 produced and maintained by Kadaster. Fully updated yearly in 5 releases with uniform quality and automatically generalised small scales. Kadaster is producer, register keeper and responsible for quality management.

This contribution will show how Kadaster supports quality management for these three spatial key registers with quality indicator dashboards (chapter 2) and crowdsourced feedback (chapter 3) and the results of these quality management systems (chapter 4).

2. QUALITY DASHBOARDS

In chapter 1 the Dutch system of key registers was introduced. Quality management has been a key element of these registers from the start. Kadaster has created quality dashboards for three key registers as a part of an integral quality management approach commissioned by the Ministry of the Interior and Kingdom Relations.

2.1 Principles of quality dashboards

The goal of quality management is to improve register quality. For all three key registers in paragraph 1.1 Kadaster has a responsibility for quality management. The main principles are that the data producer is responsible for data quality and that Kadaster will support them with quality management. This is a continuous process in collaboration with data producers. The responsible Ministry demands and supports the quality management for key registers.

Our philosophy is to actively identify, help, stimulate, consult, register, measure, follow-up and address issues. For this purpose, quality managers actively address issues with data producers. The quality dashboards are an important instrument in this process. The dashboards provide information on data and data quality and they are a tool for communication towards data producers and quality managers.

One of the main benefits of the dashboards is that quality managers and data producers consider the same tool and therefore the same information. Exported lists with quality issues are easily outdated and a dashboard gives the opportunity to check numbers live.

2.2 Quality dashboards components

Figure 3 shows the main components of the Kadaster data producer quality dashboard. On top different themes are shown. These are aggregations of the scores on a specific quality theme. Clicking on a theme will show the specific reports of the underlying analyses as shown in the middle of figure 3. And finally, there is the option to click on a specific report to get to the details at the bottom. Here the IDs of an object are listed with additional relevant information. In this way, the data producer can refer to their own system for addressing the issues. There is also an option to visualise these objects on a map.

The quality indicators in the dashboard have been determined in close cooperation with the data producers and users. The indicators are a consideration between what can be queried automatically and what is useful for users and producers. When setting up the dashboards it was a deliberate choice to focus on data errors first in order to make data producers familiar with the dashboard and to ensure that they experience it as a useful tool in solving data issues.

Every data producer has a logon to access this producer dashboard. The most recent data is shown as default. However, in the top right it is possible to choose a different date and to go

to the dashboard of an earlier date. For every part of the dashboard there is an option to create an export in either excel or pdf.

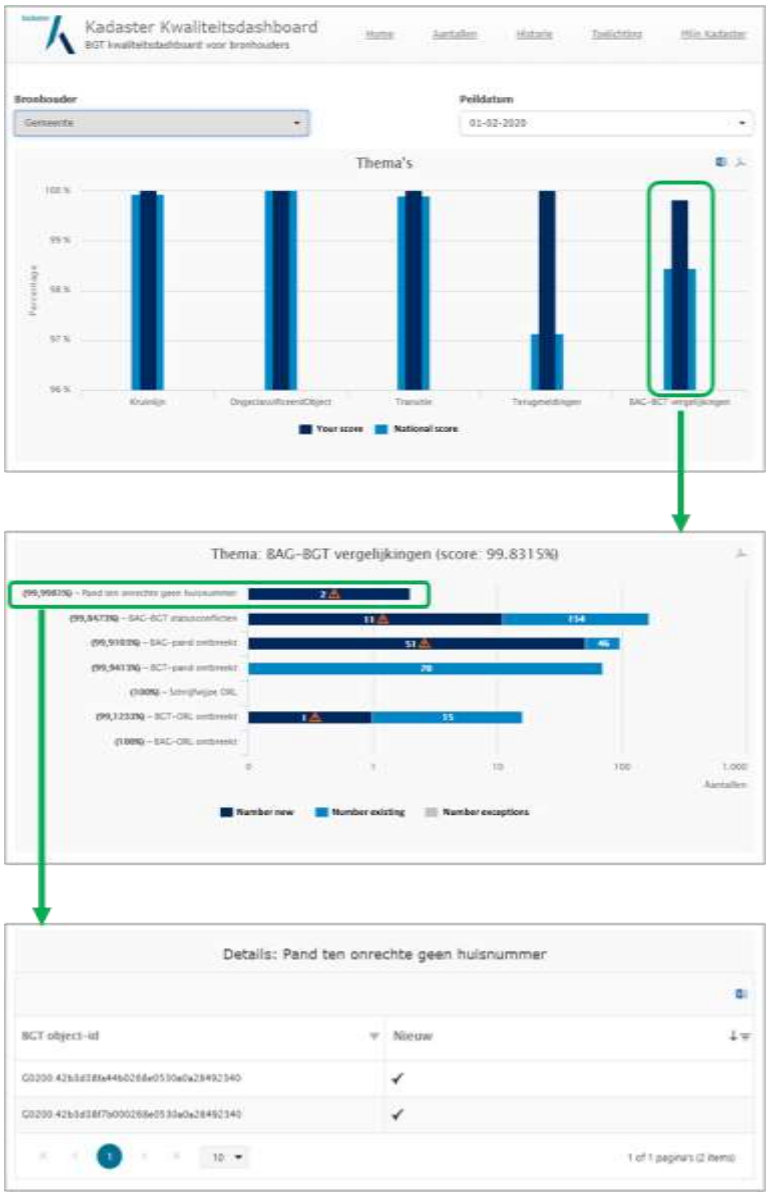


Figure 3: Kadaster quality dashboard for producers: themes – results – details.

Besides the producers scores on the themes in dark blue in figure 3 (top), the national average is also shown in light blue. This gives data producers a reference for their score on a theme. In the shown example, the producer scores above the national average on all themes. These numbers for the national average are also used by quality managers to determine which issue to focus their attention on. This extra attention can be given in many forms. It may be an article on the subject in a newsletter, extra visits to low-scoring producers, a presentation at a producers meeting or updating documentation.

It has been an explicit choice to work with positive numbers in the dashboard, so working towards a 100% score. Another choice was to use the terms ‘results’ and ‘signals’ instead of ‘errors’ because in some cases results are not necessarily errors. For example, a building area of 1 m^2 is remarkable and very likely to be incorrect, but not wrong by definition. It is up to the producer to determine if this result needs to be addressed.

Within the dashboard there are links to documentation for data producers explaining the results and how to solve them. Furthermore, there are tips in the documentation on how to prevent these results from appearing in the future.

A second page in the quality dashboard is called ‘Numbers’ and shows a variety of basic statistics for the specific producer. Some examples are shown in figure 4. This gives the producers a general insight in their data, for example in the number of changed objects in the last period.

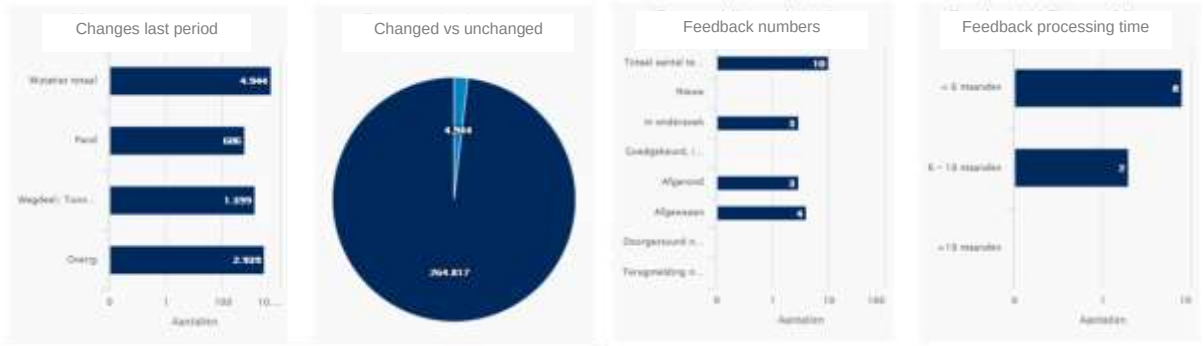


Figure 4: Kadaster quality dashboard for producers: selection of the Numbers page. Number of changed objects per object type (a) and compared to the total amount of objects (b), amount of crowdsourced feedback (c) and response time for processing the feedback (d).

The data in the dashboards are updated monthly. This period has been determined in consultation with data producers. This means the figures in the dashboards are always quite up-to-date, while it does give producers a chance to address issues. The progress of the results in the dashboard can be seen on the page ‘History’ as shown in figure 5. On this page a data producer can choose a specific quality indicator and a time period. A graph is then shown with the scores over time for the specific producer and for the national average.

The graph in figure 5 shows the numbers for one municipality showing if the crowdsourced feedback has been addressed on time by this data producer (within the legally defined period). The graph shows that in the beginning this municipality did not address the feedback in time, scoring 0%, but since late 2018 they have improved their processes and now they score 100%. The graph also shows improved scores over time for the national averages, from 88.8% to 97.1% in 20 months.

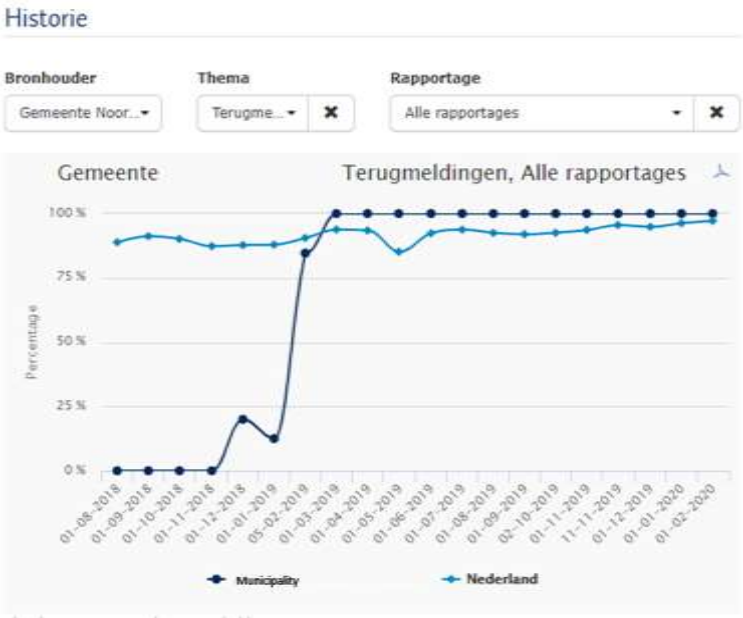


Figure 5: Kadaster quality dashboard for producers: History of addressing feedback on time. The score of the individual producer is shown in dark blue, the national average in light blue.

Currently, each mentioned key register has its own individual quality dashboard. However, although the data models of the Key Register Addresses and Buildings and the Key Register Large Scale Topography differ a lot, there is a strong connection between these two. Both key registers contain objects of buildings and although the geometries and attributes of these objects are different, there are issues that can be identified when comparing the two. An example of an indicator comparing these two registers is when a building that is present in one register, is not present in the other, while taking legal processing time allowance into account. Comparing data between different registers is a direction that is being taken more often and where the dashboards can really add value for data producers. This is also what future developments will focus on.

This section explained mainly the Kadaster data producer dashboards, which gives data producers and quality managers a tool to get insight in and improve on data quality. However, for users of the key register data it can also be interesting to get insights in the quality of the data. So, besides the producer dashboards that can only be accessed by producers through a logon, for the Key Register Addresses and Buildings (BAG) and the Key Register Topography (BRT) there is also a publicly available dashboard. It has a different interface, but the underlying data are the same. These dashboards can be found at the Kadaster website (BAG public dashboard: https://www.kadaster.nl/zakelijk/registraties/basisregistraties/bag/bag-voor-afnemers/bag-kwaliteitsdashboard-voor-afnemers, BRT public dashboard: https://www.kadaster.nl/zakelijk/registraties/basisregistraties/brt/brt-kwaliteitsdashboard).

2.3 Results of dashboards in quality management

The graph in figure 5 shows that the national average is going up for addressing crowdsourced feedback in time. But when aggregating all quality themes, the trend is also up as shown in figure 6. It can happen that the overall score drops a little when introducing a new indicator, but generally the trend for correct features is up, which shows the quality improvements of the key register data.

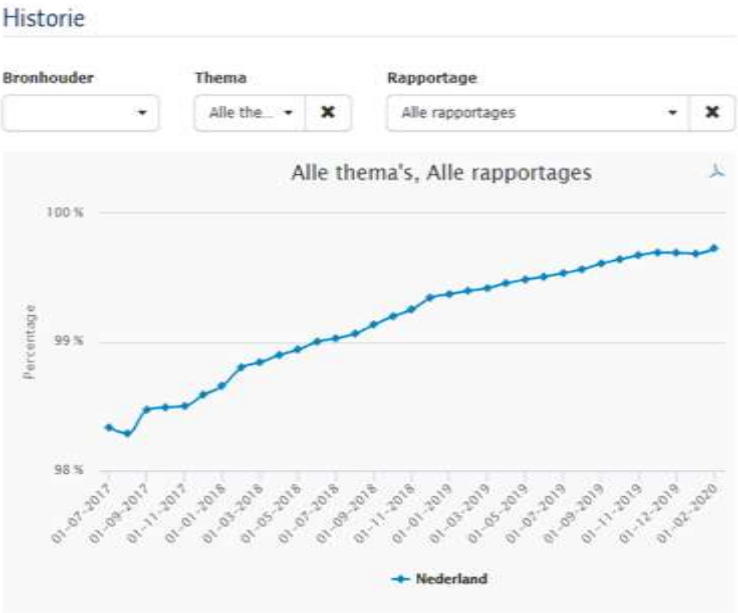


Figure 6: Development over time of national average of correct results for the Key Register Large Scale Topography (BGT).

The main drawback of these dashboards is that they only show figures that can be measured automatically. This means that there is a limit to the quality indicators that can be defined. Because there is no ground-truth available to compare the data with, valid features that do not exist in real life can still exist in the data without being identified. At the moment, it is a deliberate choice to focus on quality indicators that can be measured automatically. Therefore, the current focus is on comparing different key registers. With the help of crowdsourcing, quality issues which are not easy to measure automatically, can be addressed. Chapter 3 will elaborate on the crowdsourced feedback system.

3. CROWDSOURCED FEEDBACK

3.1 Motivation for crowdsourced feedback

Legally, every key register in the Netherlands has the requirement for governmental users to give feedback when they suspect an error in a specific key register. However, in practice, the amount of feedback from these users was very limited, especially for the spatial key registers.

Research from 2013 shows that there were five barriers for users to report feedback:

1. **Not using** a key register;

2. **Not knowing** about the key register or the possibility to report feedback;
3. **Not recognising** an error in a key register;
4. **Not willing** to report an error;
5. **Not being able** to report an error (Mies et al., 2013).

Remarkably, from these five barriers, 'not willing' appeared to be the most important barrier. Users did not report feedback, because they expected that the citizen or company involved would report the error. Also, users found it very demotivating that they would not receive any feedback when they reported an error (Mies et al., 2013). As the spatial key registers are mainly used indirectly, the direct consequences of an error in these registers for a company or citizen are less severe and visible than in the administrative key registers. This also resulted into a poor feedback flow from them. Furthermore, many registers did not even have a possibility for citizens or companies to give feedback, as the law specified only an obligation for governmental users. Because the spatial key registers are open data, this practice seemed a missed opportunity to gather feedback from everyone to improve the key registers.

3.2 Development of crowdsourced feedback

In 2013 the Dutch Kadaster launched a pilot for the Key Register Topography (BRT) to collect feedback through crowdsourcing. During this pilot, a group of volunteers could provide feedback by putting markers with comments on the map. This feedback was directly assessed by experts from the Kadaster, the data producer of the BRT. The choice was made to keep the markers and their status visible on the map, to give volunteers the possibility to track and trace the status of their own feedback and the feedback of other volunteers (Grus & Te Winkel, 2015).

The results of this pilot were impressing: the pilot lasted for two months and in this period the Kadaster received 369 cases of feedback. Compared to the amount of feedback in previous years via the old system, which was only 8 in 2011, 10 in 2012 and 28 in 2013, this was an enormous increase. Furthermore, 90% of the feedback was deemed correct and only 10% of the feedback was disapproved. From an evaluating survey it turned out that the users were very enthusiastic and content with this new approach (Grus & Te Winkel, 2015).

Because the pilot was very successful, it was decided to make the pilot system the new feedback system for the BRT. Meanwhile, the Key Register Large Scale Topography (BGT) was being developed. For this key register, at that time more than 400 governmental organisations (currently 389) cooperated to create this large-scale map of the Netherlands. Research showed that both (potential) users of the BGT and data producers supported the idea to use volunteered geographic information (VGI) as feedback system (Sjoukema, 2015). As the distribution of feedback among these 389 data producers would be more complex, a new crowdsource feedback application was developed: 'Improve the Map' (Verbeter de Kaart: https://www.verbeterdekaart.nl/) which was launched in June 2016.

Later in 2016, the BRT was also included in the 'Improve the Map' application. For the Key Register Addresses and Buildings (BAG) it was decided to adapt the highly used 'BAG viewer' to make it suitable for crowdsourced feedback (https://bagviewer.kadaster.nl/). In

2017 the BAG viewer connected to the backend of the 'Improve the Map' feedback system through a REST API. In 2019, this API was also released for external applications and map viewers.

3.3 Feedback process

'Improve the Map' is an easy-to-use map viewer. Everyone can contribute feedback by clicking on the map and adding a comment, an attachment and an e-mail address (see figure 7). Based on the geometry of the click, the feedback is automatically distributed to the correct data producer. Data producers receive an e-mail notification when they are assigned to feedback. It is their job to investigate and assess the feedback in a timely manner.

Users can track and trace their feedback and that of other contributors. Also, they will get notifications at their e-mail when their report is registered and assessed by the data producer. Furthermore, the feedback data is disseminated as real-time webservices (WMS and WFS) via the Dutch National Spatial Data Infrastructure: PDOK (https://www.pdok.nl/) and through a REST API. Only the comment, the location and the comment of the data producer to the feedback are made public. Personal information and attachments stay secure.

Figure 7: Screenshot of 'Improve the Map' (https://www.verbeterdekaart.nl/) showing the form for giving feedback on the Key Register Topography.

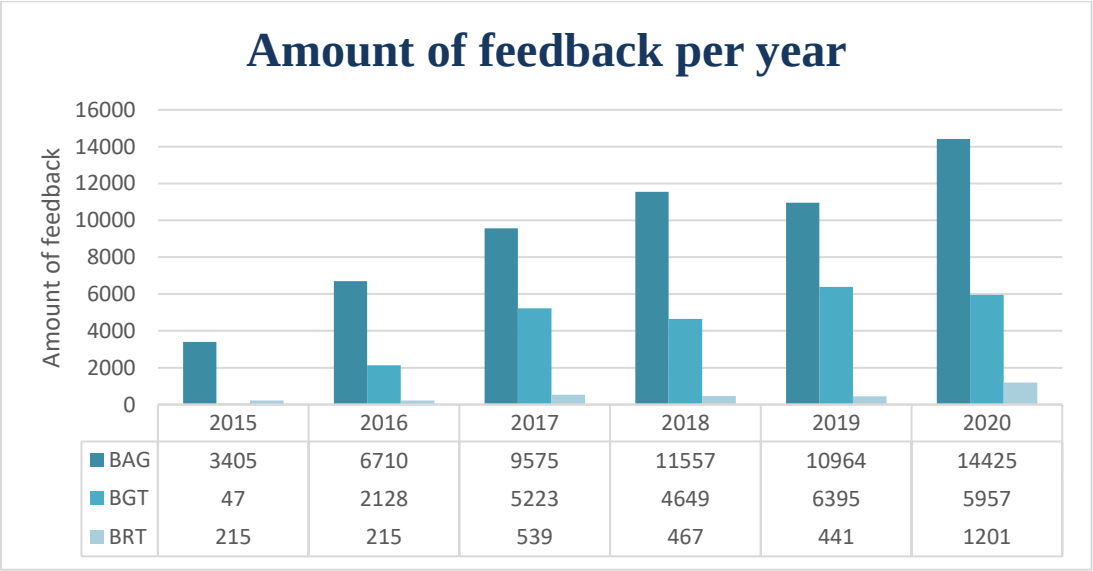
The feedback can get different statuses which is made visible on the map. These statuses give an indication of the assessment process by data producers. When data producers change a status, the feedback provider will get an e-mail notification. The following statuses can be given in the feedback management system for data producers:

- **New:** For newly registered feedback.

- **Under investigation:** The data producer has the feedback under investigation.
- **Approved, planned by data producer**: The data producer has approved the feedback, but not yet corrected in the key register. For example, because measurements by a surveyor are necessary.
- **Parked:** The data producer investigated the feedback but could not decide yet. For example, because newer aerial imagery is needed to assess the feedback. This status can only be set by the BRT.
- **Forwarded to other key register:** The source of the error lies in another key register. The data producer forwarded the feedback to this register. The system will automatically register a new feedback for this register.
- **Disapproved:** The feedback is not correct, or the feedback is correct, but changing the key register is not necessary according to the key register standards.
- **Finished:** The feedback is approved and corrected in the key register.
- **Spam:** The feedback is double, it is a test feedback or it is an obviously fake feedback.

When a feedback gets the status 'forwarded to other key register', 'disapproved' or 'finished' it will stay visible for two months in the feedback system and then it will be automatically removed and archived. The reason to keep this feedback on the map is to show it to the original contributor and let other contributors learn how a good or bad feedback may look like. When a feedback gets the status 'spam' it will be directly removed from the map. All other feedback remains on the map until it gets one of the close statuses. There is no required sequence in statuses for data producers, as this proved to hinder the handling of feedback instead of facilitating.

3.4 Results of crowdsourced feedback



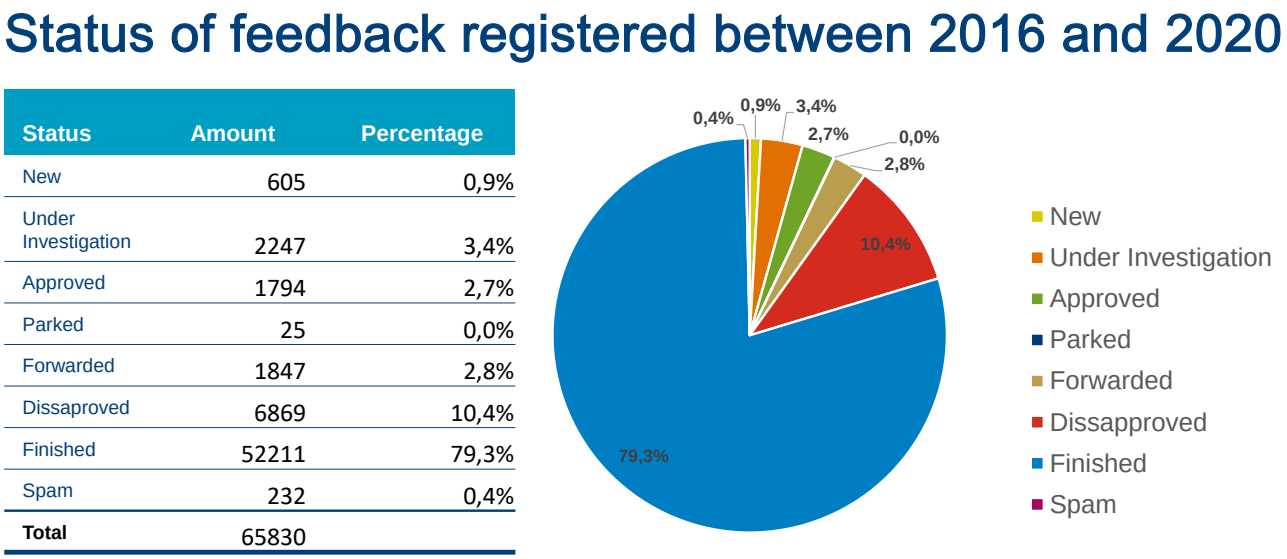
	2015	2016	2017	2018	2019	2020
BAG	3405	6710	9575	11557	10964	14425
BGT	47	2128	5223	4649	6395	5957
BRT	215	215	539	467	441	1201

Figure 8: Annual amount of feedback in the period 2015-2020 per key register: Key Register Addresses and Buildings (BAG), Key Register Large Scale Topography (BGT), Key Register Topography (BRT).

Since the introduction of the crowdsourced feedback system mid 2016, the amount of feedback has grown tremendously (see figure 8). When we compare the amount of feedback in 2015 to the amount of feedback in 2020, 5.6 times more feedback on the Key Register Topography (BRT) was registered in 2020 and 4.2 times more feedback on the Key Register Addresses and Buildings (BAG). As the Key Register Large Scale Topography (BGT) was not yet completed in 2015, it is not fair to compare the number of feedback to 2015, but it is clear that also for this register the amount of feedback has grown since the introduction in June 2016.

In 2020, especially the feedback of the BRT and BAG was boosted considerably. There are multiple factors which may explain this increase. First, several external applications connected to the REST API and provided feedback. In total 1320 were registered in 2020 through these applications. Second, multiple external factors played a part. For example, new regulations on energy labels of buildings confronted many citizens with the building age registered in the BAG, which might be incorrect. Also, we found out that multiple recreational BRT users made a hobby of reporting errors.

Besides the amount of feedback, the way how feedback was provided also changed. In 2015 it was possible to provide feedback through phone, e-mail or a publicly accessible online webform. However, these channels largely disappeared in 2017 and in 2018 almost all feedback was provided through the crowdsourced feedback system. This makes it easier for data producers to handle, process and assess the feedback.



Status	Amount	Percentage
New	605	0,9%
Under Investigation	2247	3,4%
Approved	1794	2,7%
Parked	25	0,0%
Forwarded	1847	2,8%
Dissaproved	6869	10,4%
Finished	52211	79,3%
Spam	232	0,4%
Total	65830	

Figure 9: Graph and table showing the current status on 31-3-2021 for all registered feedback between 2016 and 2020 via the crowdsourced feedback system.

Figure 9 shows the assessment of all feedback registered in the period from 2016 till 2020 and their status on 31 March 2021. The largest share of the feedback has been finished (79.3%). 10.4% percent of all feedback is disapproved and only 0.4% has been marked as spam. In many cases, feedback becomes disapproved not because it is not true, but because the

feedback does not fit the register guidelines of the key register. The number of 'parked' statuses is low because this status is only used by the BRT.

A challenge lies in the timely assessment of the feedback by the data producer. For example, 0.9% of the feedback is still registered as new in figure 9, while the feedback is more than three months ago reported. With the addition of an indicator in the BGT quality dashboard (see figure 5), pro-active quality management and easier handling of feedback, the number of BGT feedback which was not timely assessed dropped from an average of 734 (six months prior to the introduction of the indicator on 1 August 2018) to an average of 198 in 2020.

4. CONCLUSION

The two participatory instruments, quality dashboards and crowdsourced feedback, are very useful and successful instruments for the continuous improvement of key register data quality in a transparent way. Through the quality dashboards, data producers and quality managers gain a shared understanding on the results and where to improve on data quality. With the history function, progress of the indicators is traced, making it a stimulating instrument for data producers to improve. With the help of crowdsourced feedback, everyone can contribute which establishes a continuous flow of feedback for the key registers. With this approach, the process of providing feedback has been turned from a frustrating experience to a very motivating one, resulting in an enormous growth of high quality feedback.

However, the role of pro-active quality managers should not be underestimated. These instruments only work well when data producers are addressed when potential quality issues arise. When nobody looks at the quality dashboard, the dashboard becomes obsolete. And when feedback is neglected by the data producers, even a user-friendly crowdsourced feedback system becomes a frustrating experience. To some extent, the role of quality managers proved to be more helpful for improving quality and assessing feedback, than the requirements made by the standards or law.

Both participatory quality instruments are now used on three key registers within the Netherlands, but their potential reaches further. Other key registers and domains have shown interest in the positive and motivating approach of one or both instruments, for example the cadastral register, real estate valuation register, subsurface register, utility networks, road networks and spatial planning. Therefore, both applications are developed further in a generic way in order to add new registers and themes easily.

REFERENCES

Digitale Overheid, 2020, Basisregistraties, https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/basisregistraties-en-afsprakenstelsels/inhoud-basisregistraties/

Grus, M. & D. te Winkel, 2015, Assessment of the Volunteered Geographic Information Feedback System for the Dutch Topographical Key Register, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-3/W3, ISPRS Geospatial Week 2015, 28 Sep – 03 Oct 2015, La Grande Motte, France

Mies, K., Van den Berg, T., Kronenburg, T. & T. Monasso, 2013, Terugmelden, Waarom niet? Eindrapport, Den Haag, PBLQ ZENC.

Sjoukema, J.R., 2015, Keeping the BGT up to date: Volunteered geographic information, a serious option? GIMA Master thesis, Utrecht, Utrecht University.

BIOGRAPHICAL NOTES

Anouk Huisman - van Zijp is a member of the production and innovation team of the Key Register Topography at the Dutch Cadastre (Kadaster) and as such a representative at the Eurogeographics Quality Knowledge Exchange Network (QKEN). She earned her Master of Science in Geomatics at Delft University of Technology and has worked both nationally and internationally in the geo-information field.

Jaap-Willem Sjoukema is product owner of the crowdsourced feedback system at the Dutch Cadastre (Kadaster). During his Master Programme ‘Geographical Information Management and Applications’ at four Dutch universities (Delft University, Wageningen University, Utrecht University and Twente University) he studied the possibility for using a crowdsourced feedback system for the BGT. Next to his work at the Cadastre, he is also a PhD-candidate at Wageningen University and KU Leuven researching governance for Spatial Data Infrastructures (SDIs).

CONTACTS

Anouk Huisman-van Zijp MSc.
Kadaster
Hofstraat 110
7311 KZ Apeldoorn
THE NETHERLANDS
+31 88 1833840
anouk.huisman-vanzijp@kadaster.nl
www.kadaster.nl

Jaap-Willem Sjoukema MSc.
Kadaster
Hofstraat 110
7311 KZ Apeldoorn
THE NETHERLANDS
+31 6 15462942
Jaap-willem.sjoukema@kadaster.nl

www.kadaster.nl