Benefits Of OGD For Providers Of Authoritative Data

Jürg H. LÜTHY, Christian KAUL and Markus LUGINBÜHL, Switzerland

Key words: Voluntary Geographic Information, GSDI, Open Government Data; Web-services

SUMMARY

In the last years the concept of Open Government Data (OGD) got increased relevance and attendance. For mapping agencies as the most important data provider it became evident that OGD is more than just data for free. With the open access to data and wider usage of data sets the position of the unique knowledge bearer due their particular access to data got lost. This shift impacted the position of the mapping organisation and posed challenges to their structure as discussed by Lüthy et. al. (2016). With the first practical experiences of OGD provision owners and providers of authoritative data realised that they perceived many benefits compared to the situation before. The first and most relevant change is the simple fact that data provided under OGD are utilised more often and more intense. This effect can be shown in all Cantons of Switzerland where Cadastral data are published under OGD: What are the reason for this increase? Due to the omission of the costs for data private constructors and architects update the base maps more frequently than before and do not work anymore with outdated, incorrect data. The regular updates lead to massive increased quality in all processes from initial concept to planning, approval and to the construction. Looking from a macroeconomic perspective the omission of the costs for data is more than compensated.

The broader use of the data can lead to an improved data product. If the data owner is open for criticism and feedback, the OGD infrastructure can be enriched with a simple to use module where data users have the possibility of providing error indications, improve timeliness (new objects or missing information) or comments to the overall content of the data set (data model). The one-way process should be transformed to a circuit thank to the inclusion of volunteered geographic information (VGI). First experiences show that aspects of data reliability, completeness and temporal accuracy (timeliness) are more relevant than spatial accuracy.

For a successful transition towards the combination of volunteered geographic information with established data maintenance processes through the data owner different challenges have to be solved. The tools for providing feedback must be aligned to the end users which are in many cases no spatial data specialists. Furthermore, the tool should use state-of-the-art technologies, should be simple to handle and should have some appeal to ensure that it is actually used. Acknowledging feedback from non-specialists may be difficult for some experts. They have to learn, how such reports are considered to improve the data quality. Once accepted, it is important the VGI is included appropriately in the processes.

The benefit of the incorporation of VGI in authoritative data is not limited to cadastral data. In Switzerland the Federal Office of Aviation (FOCA) is responsible for the publication of an

obstacle data set. This OGD data set is used in designing approaches, planning of flights but also in-flight. It can be considered as mission critical information, because missing or wrong information may lead to severe accidents. To improve the quality of the data the feedback loop is in use since several years.

Benefits Of OGD For Providers Of Authoritative Data

Jürg H. LÜTHY, Christian KAUL and Markus LUGINBÜHL, Switzerland

1. INTRODUCTION

1.1 Legal base

The Canton of Zurich amended its law last year to define its new data policy as Open Government Data (OGD) by default. This means that every data set which is publicly accessibly must be published under the rules of OGD (cf the principles published by the Sunlight Foundation[1]). Where the data owner wants to apply any restrictions, they have to justify the exception and the this must be declared in the metadata as such. The reversed justification increased the data owners' sensitivity for OGD and enlarge the number of available data sets.

1.2 Current data exchange platforms

Even under restricted access to governmental data there has been a large number of data sets which were accessible and used paying a royalty fee. For each product, i.e. for each data set and for each format in which it was made available, a tool to extract, transform and load (ETL-tool) had to be configured manually. For each data delivery the configuration has to be checked for actuality and then the process was triggered manually. The exchange data were the sent by mail to the customer. For the data set "Cadastral data" (meaning land ownership including topographic information) where the data distribution started earlier than for other data sets and for which the demand has always been much higher a separate platform with fully automated process was in use. For several years to different platforms with differentiating data sets, technologies and different level of automation have been in use. It's obvious that the redundancy of the system led to complicated use cases when data from both platform were ordered. Furthermore, the effort to support and improve the tools has been duplicated as well.

1.3 Open Government Philosophy Data for spatial data

In its strategy for the development of the OGD policies the Canton of Zurich differentiates between base products and value-added products. Within the group of base products, the realisation for the data sets is achieved according to the strict definition of OGD. For spatial data this means that a machine-readable Web-service compliant to the OGC standards and optionally a download of physical files in a standard exchange format is provided. These services should always be performed by the responsible governmental agency. All products which are more sophisticated are considered to be value-added products and may are set-up and marketed through any organisation. This is actually one of the basic goals of OGD that (commercial) organisation use the base service and refine the data or combine data sets to a new product and make money out of this difference. Therefore, all products offered by the

[1] https://sunlightfoundation.com/policy/documents/ten-open-data-principles/

governmental agencies going beyond basic products must be evaluated very carefully to not interfere with the business ideas. The Canton of Zurich decided to launch few added-value products with the intention to foster the use of spatial data and promote the concept of OGD. One of the core added-value services it a Web-API for data ordering (by topic, area) and to integrate this API in the public WebGIS infrastructure of the Canton.

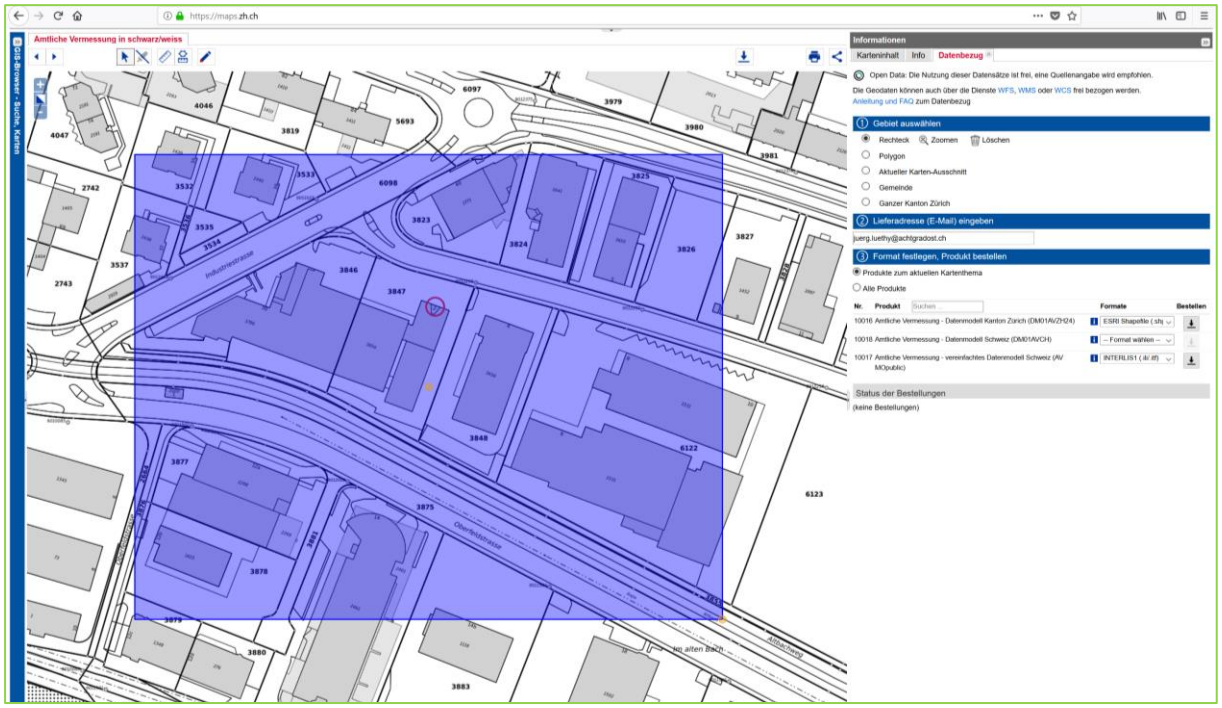

Figure 1 – Data ordering form integrated in the WebGIS browser

1.4 Voluntary Geographic Information for authoritative data

In the last decade many spatial datasets have been built-up by a community driven approach. Here users generate content in a web application either by on-screen digitalisation or by integration data from their mobile device. The most well-known data set created as voluntary geographic information (VGI) is OpenStreetMap (Demetriou, 2017). But besides base maps many other topics which are nowadays in wide use such as crisis and disaster management (Humanitarian OpenStreetMap, HOT). In recent years various mapping agencies realised the benefit of opening their data production environment to feedbacks from end users. Whilst authoritative data provides like National Mapping Agencies (NMA) focus on strict standards (for data interchange), reliability and continuity the data sets may are not updated as often as needed due to the high costs involved. The crowd sourcing approach on the other hand is very cheap, but may not achieve the same level of standardisation and can't be considered as official data. Combining the two methods is considered as beneficial for the authoritative data providers.

A review undertaken by Olteanu-Raimond et. al (2017) among European NMA showed that only very few of them are using the collaborative approach to a higher degree than a simple feedback loop. The biggest concerns for NMA integrating VGI relate to data quality, reliability, legal issues and employment fears. The review with the NMA showed as well their interest in incorporating more VGI approaches in their production processes. According to the authors of the study the most developed crowd sourcing activity is in the field of change detection and error feedback loops for topographic maps. It can be regarded as selective improvement. In several countries VGI is used to generate a "citizen map layer" which is the used as one source for enhancing and updating maps. Some agencies are even using the approach for gather data of a new topic to enrich the content of the base maps (religious buildings in Netherlands). For all presented methods there is an asymmetry regarding the licensing. NMA ensure that the rights for provided information are with the NMA whilst still limiting the use of their data and maps by keeping licensing fees.

2. EVOLUTION OF OGD IN CADASTRAL DATA OF CANTON ZURICH

2.1 Developments for the provision of OGD data

For a fully automated handling of all the data orders and data downloads the existing spatial data infrastructure was extend with a new data extraction module. On the base of an ETL-tool processes were configured which can spatially extract the data according the order from the entire data set and transform it into the desired data delivery format. The pre-existing shop for the Cadastral data was integrated to this platform and processes. Taking into account the large number of data sets in the SDI and the various delivery formats which must be supported it is obvious that the process-control must be fully automated. To achieve this the existing metadatabase was expanded with the parameter-set for the ETL tool. The human-readable parameters provide the necessary intelligence to not only steer the processes but also supports a pure generical configuration of the data converter. The interaction between system and end user is achieved through a REST Web-API. The API supports a data order and delivery without media-discontinuity.

Besides the data download services all data sets they can also be accessed through OGC web-services. The data sources for the services are run independently from the other services to achieve higher stability and performance. The data bases are synchronised on a daily basis, triggered again by the metadatabase.

The OGC services are by default provided as (open) machine-readable interface but also the own API is accessible for everyone without restrictions. Since not every end-user is capable of handling API and web-services the decision was taken that a user interface integrated in the web-browser should be provided from begin on (see also fig 1).

2.2 Experiences after first month of operation

The release of the spatial data for use as OGD was promoted with intensive communication. During the development and testing phase of the new infrastructure the expert users and well-known data consumers were specifically informed through newsletters. The professional users were educated in-depth through article in journals. Some weeks before the roll-out the broad public was comprehensively informed at a media conference and a press release was published. The echo in newspaper and trade media was surprisingly high, considering that most citizen do not often get in touch with spatial information. The immediate release was accompanied with some tweets (https://twitter.com/GISZentrumZH) which were enthusiastically shared and commented.

In the first days under the OGD regime more than 400 data sets were ordered. The average number of orders is now in the area of 180 individual order per week day (see figure 2). During the weekend more or less no data-sets are requested. The interest of the users is approximately evenly distributed between Cadastral data, base data (addresses, municipal boundaries, elevation information and planning zones) and all other data (overall more than 100 data sets, see figure 3). Far the most users request data in the DXF format which is still the preferred format for architects (Cadastral data) and civil engineers. From the more structured data formats the Esri Shape Format and the File Geodata Base are widely requested. The modern standards like Geopackage attract very little interest. On the contrary the simple CSV format for point-based geometries has gained a brisk demand. Not surprisingly a drastic decrease of requests through the old ordering process can be observed. Only data sets which are not available under OGD are requested sporadically.

Overall it seems that the end-user can get the data and format they need. The help desk has to answer less than a hand-full request and questions per day. Most of the concerning the availability of data sets which are currently not published under OGD.

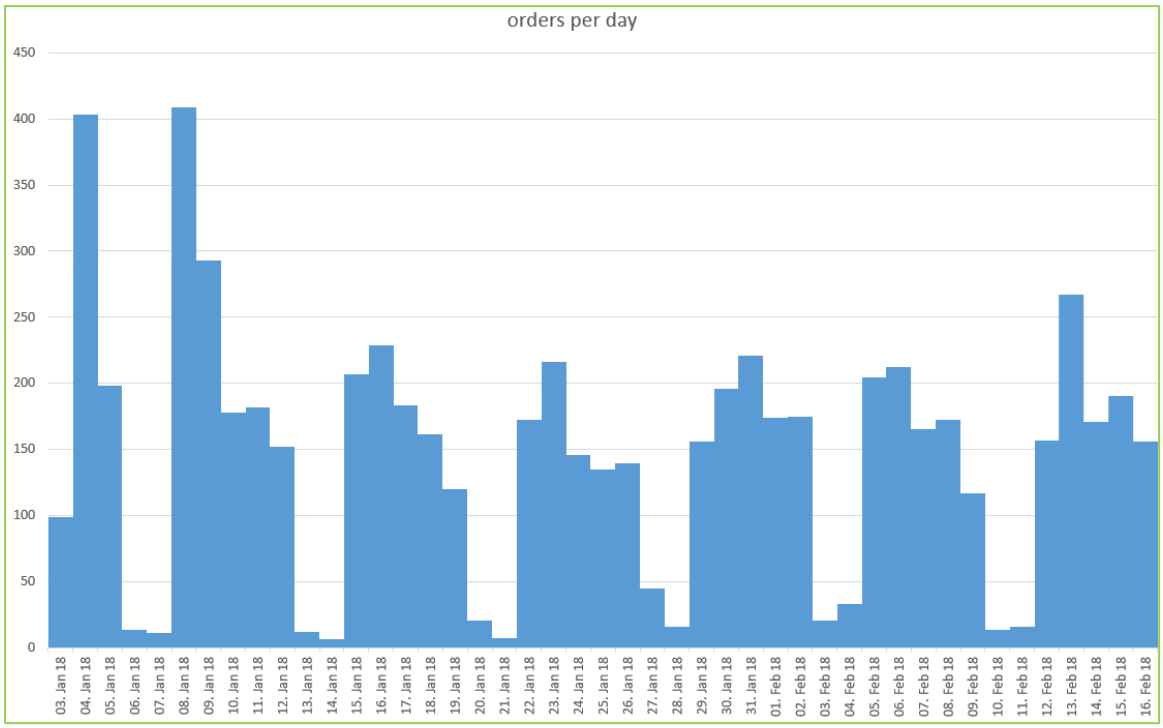

Figure 2 – Data orders per day

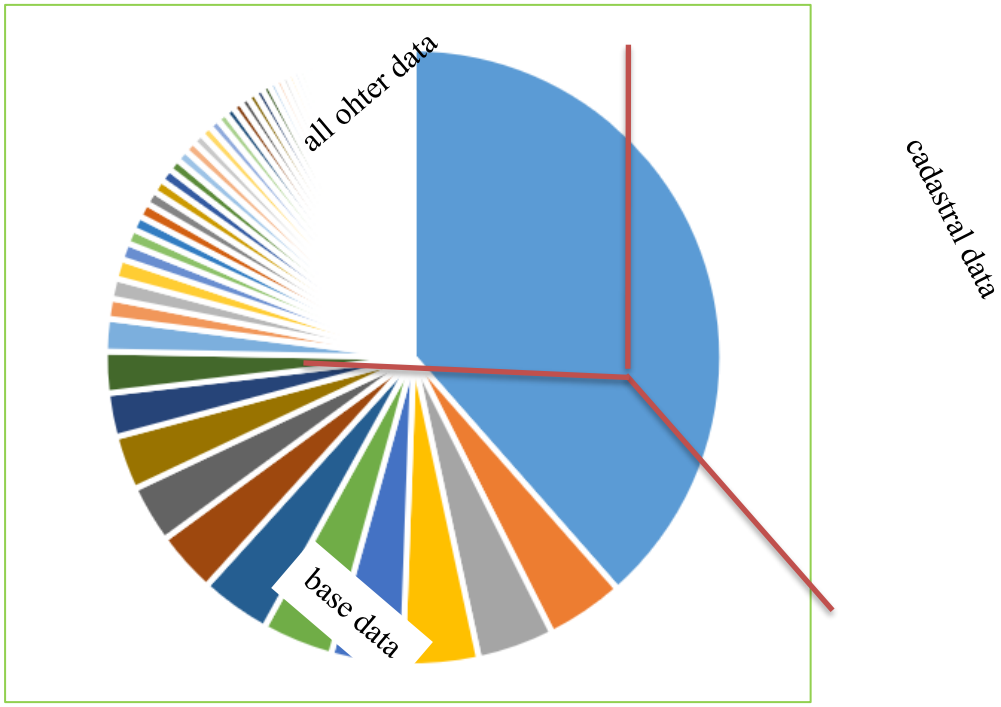

Figure 3 – Types of ordered data sets

2.3 Benefits

The effort to process manual data orders has fallen more than expected. This freed up human resources for more important tasks. The outdated web-portal for the acquisition of cadastral data has been discontinued. This eliminates its maintenance and operation.

The intensive use of the data from the official survey confirms that planners, architects, engineers and real estate developers now have up-to-date data for each task. This significantly increases the quality of their work.

Due to the great media response, the geoinformation is once again on everyone's lips.

2.4 Challenges

Due to the first experiences, the Web API will be technically optimized by mid-2018 and supplemented with other user formats such as CSV (which is currently available in the REST-API only). Subsequently, this machine interface is to be standardized as part of the national eGovernment activities. This process will lead to further optimizations.

The data owners are required to comprehensively check their spatial data sets and prepare them for use as OGD. The demands on the quality of content will increase with the use as OGD, since a wide circle of users works with the data. To avoid negative experiences on the part of data producers in the form of bad user feedback, a review of the data and an evaluation against the specifications is recommended. This realization also forms the basis for integrating VGI into the quality assurance processes.

3. INCORPORATION OF VGI IN OGD INFRASTRUCTURE

3.1 Feedback Loops

The broader use of the data can and should lead to an improved data product. If the data owner is open for criticism and feedback, the OGD infrastructure can be enriched with a simple to use module where data users have the possibility of providing error indications, improve timeliness (new objects or missing information) or comments to the overall content of the data set (data model). The one-way process should be transformed to a circuit thanks to the inclusion of volunteered geographic information (VGI).

For a successful transition towards the combination of volunteered geographic information with established data maintenance processes through the data owner different challenges have to be solved. The tools for providing feedback must be aligned to the end users which are in many cases no spatial data specialists. Furthermore, the tool should use state-of-the-art technologies, should be simple to handle and should have some appeal to ensure that it is actually used. Acknowledging feedback from non-specialists may be difficult for some experts. They have to

learn, how such reports are considered to improve the data quality. Once accepted, it is important the VGI is included appropriately in the processes.

Particular attention must be paid to granting rights when using VGI to improve data quality. If a user notifies changes, he has to agree that the data controller may continue to use this information without royalties. These challenges and approaches have been extensively discussed by Brown (2017).

In the following sections three approaches with different integration level are presented in more detail.

3.1.1 Simple feedback loop and reporting system

In the case of a simple reporting system, the WebGIS application of the data owner provides the users a possibility of reporting a feedback on any object or on any location in the data set. The feedback loop is primarily achieved with an application form for factual data - what is not correct, why is something not correct.

If the WebGIS application is suitable for use on the smartphone, the capability should be added to post a picture to a message and the add the co-ordinates of the current position. With this simple reporting system, the barrier should be kept as low as possible for the users in order to gain access to a broad citizen segment. On the part of the data owner processes have to be established to process these feedbacks. Because not a lot of information is required by the user, it often happens that the messages are wrong (for example, a request is posted for a fact that is not part of the data set) or that further clarification is required. Therefore, it is important that active communication is maintained from the data owner to the community.

This form of feedback already exists for some time in the context of cadastral data. With the growing use of data released as OGD, it seems important to us that this capability be extended to all datasets.

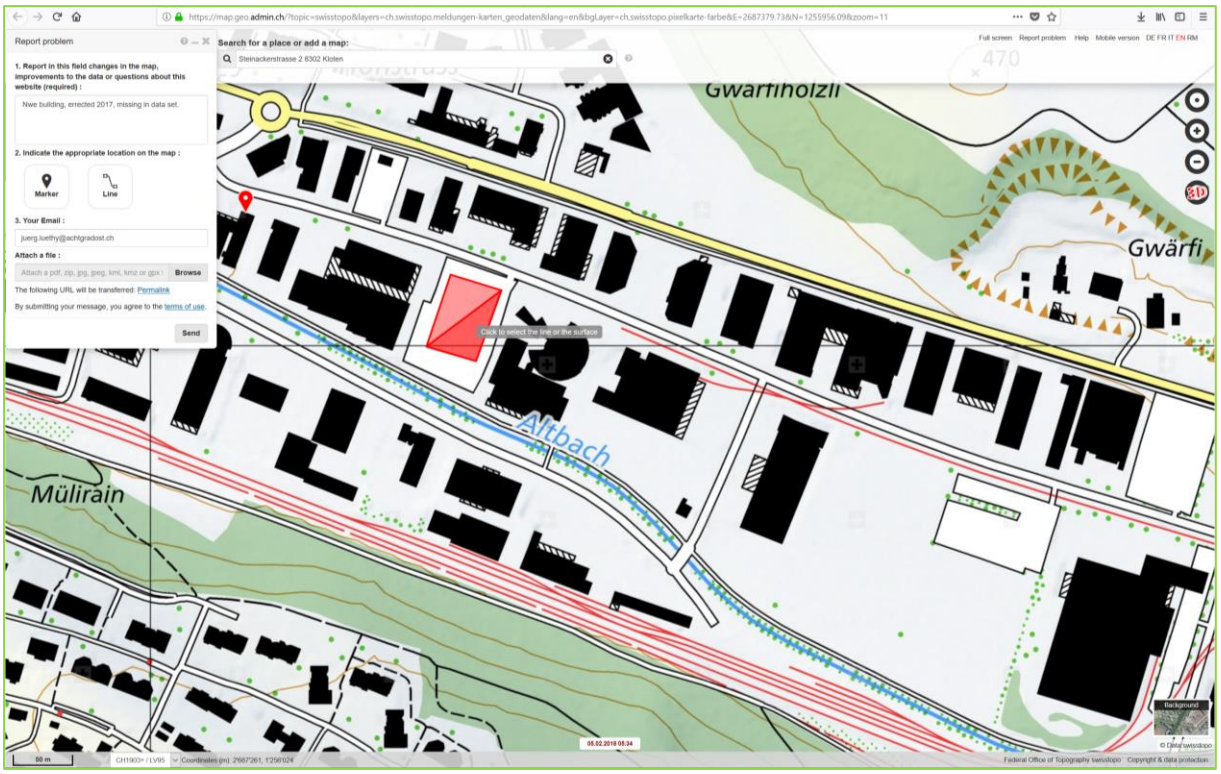

Figure 4 Simple feedback form of Federal Mapping Agency swisstopo

3.1.2 Feedback on object with extensive information

A higher-level feedback can be made if the users are more experience and have a good knowledge of the data catalog. In this case, not only can the fact that there is a difference between data and reality be reported, but also more precise information on geometry or factual information can be provided. For these cases, the form can become quite extensive depending on the size of the data model. It is therefore important that help tool tips are offered for each attribute, or that auxiliary information can be provided for mandatory fields. If the feedback is integrated into an app, the user can be well guided through a workflow so that as little information as possible is presented. The selection or value in one form then determines the layout of the next form. For example, if only the covering type of a path changes, no geometrical information needs to be captured by the user.

In the age of OGD, however, it is becoming increasingly standard that customers are using the data in their own system landscape. Therefore, it is not attractive for them to switch to another application for feedback. For a dataset published under OGD, the feedback loop should be offered as a web API, so that it can be called directly from the user's application context.

For the data owner, integrating this kind of feedback into business processes is usually easier because the quality of the data collected is higher. In order to be able to optimally exploit the

potential of VGI, the first feedbacks from a user must be checked more closely and constructive feedback on any incorrect information is important. With continuously consistent and error-free feedback, a user should receive a better rating, so that on the side of data owner the test can be differentiated.

3.1.3 Full inclusion of the crowd

For data sets that have a relatively high frequency of change (or are only partially available) and for which a larger group of users have distinguished themselves by a good rating in reporting differences, full participation is conceivable. The feedback should therefore be structured in such a way that the change messages are fed directly into the database. Since the responsibility for the content of an authoritative database cannot be ruled out as in OpenStreetMap, validating or authorizing a change by the data owner is in many cases indispensable.

Since the original data has to be processed in this expansion stage, implementation of the platform requires the data owner to supplement the platform. Although the technical basis for distributed processing in various systems is doable with WFS-T, data integrity must be given great consideration so that different users cannot simultaneously change the same object. The implementation of the third stage is therefore not in the foreground for the canton of Zurich.

3.2 Example Improve obstacle data set for civil aviation

The benefit of the incorporation of VGI in authoritative data is not limited to cadastral data. In Switzerland the Federal Office of Aviation (FOCA) is responsible for the publication of an obstacle data set. This OGD data set is used in designing approaches, planning of flights but also in-flight. It can be considered as mission critical information, because missing or wrong information may lead to severe accidents. To improve the quality of the data the feedback loop is in use since several years.

The form is used for various types of messages: approval of an owner who wants to set up a new obstacle, changes to an obstacle approved by the FOCA, but also the announcement of pilots to previously unknown or not properly documented obstacles (see Figure 5). A variety of factual information is to be included in a notification, so that the obstacle can not only be mapped, but also the assessment of an authorization capability can take place.

After the form has been sent to the FOCA, it will be checked by a specialist. If required, additional information is requested and the object is supplemented with it. If the position and height information is not plausible or in sufficient accuracy, an on-site inspection and survey will be initiated. The data will be published in the on-line dataset after the examination. By including the users, FOCA was able to considerably improve the timeliness of the database and significantly reduce the effort required to process the applications.

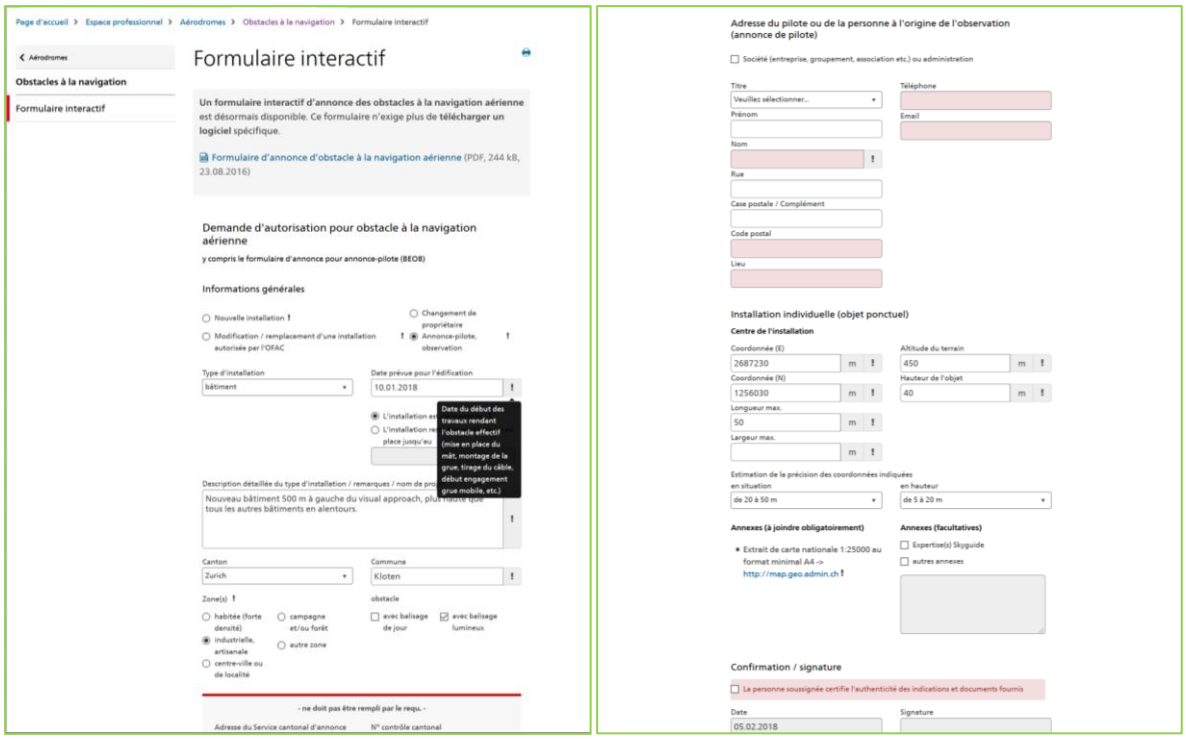

Figure 5 Reporting form for recording a change indication for flight obstacles

4. SUMMARY AND OUTLOOK

After the experience in the development and first month of operating the OGD infrastructure in the Canton of Zurich it can be postulated that the benefits clearly outweigh and the use of spatial data has entered a new era. VGI can be used in particular for authoritative data as a useful addition to the classic processes of maintaining data. Especially for cost reasons, the classic processes never deliver a complete quality. With VGI, the data quality can be significantly increased again, which benefits all users.

More and more data sets are published continuously under OGD. The available data formats will in future be supplemented with a publication on the Atom Publishing Protocol (APP, AtomPub). The AtomPub messages are in turn controlled directly from changes in the metadatabase. In addition, a combination with OpenSearch technology is sought in Switzerland for further simplifying the access to spatial data.

After initial experience, VGI feedback loops are upgraded to level 2 and later towards level 3.

REFERENCES

Demetriou, D, Campagna, M, Racetin, I, Konecny, M. 2017. Integrating Spatial Data Infrastructures (SDIs) with Volunteered Geographic Information (VGI) for creating a Global GIS platform. In: Foody, G, See, L, Fritz, S, Mooney, P, Olteanu-Raimond, A-M, Fonte, C C and Antoniou, V. (eds.) *Mapping and the Citizen Sensor*. Pp. 273–297. London: Ubiquity Press.

HOT – Humanitarian OpenStreetMap Team – Web portal: https://www.hotosm.org/ (accessed 2018/01/21).

Lüthy J. H. (2016). Expanded data quality model for increased reliability in mashed-up environments. FIG Working week 2016, *Recovery from Disaster*, Christchurch, New Zealand, 2-6 May 2016.

Olteanu-Raimond, A-M, Laakso, M, Antoniou, V, Fonte, C C, Fonseca, A, Grus, M, Harding, J, Kellenberger, T, Minghini, M, Skopeliti, A. 2017. VGI in National Mapping Agencies: Experiences and Recommendations. In: Foody, G, See, L, Fritz, S, Mooney, P, Olteanu-Raimond, A-M, Fonte, C C and Antoniou, V. (eds.) *Mapping and the Citizen Sensor*. Pp. 299–326. London: Ubiquity Press.

BIOGRAPHICAL NOTES

Jürg H. Lüthy is member of the Management Board at Acht Grad Ost AG, one of the largest geomatics companies in Switzerland. He obtained a master's degree in 1996 from Federal Institute of Technology Zurich (Switzerland) in Rural Engineering and Survey. From the same institution he holds a PhD (2007). He has many years of experience in spatial data management, transition from paper maps to data centric systems and the operation of Spatial Data Infrastructures. His current focus lies in the provision of holistic information using modern web-technologies like designing information management platforms or building the technical infrastructure for Cadastre of Public-law Restrictions on landownership. He is the Swiss delegate to FIG Commission 3. Since 2016 he is president of SLM Swiss Landmanagement Foundation.

Christian Kaul is head of Department for Geoinformation at the Office for Spatial Development in Canton of Zurich (Switzerland). He obtained a master's degree in 1992 from Federal Institute of Technology Zurich (Switzerland) in Rural Engineering and Survey. After ten years of experience in different domains like communal infrastructure, land management and SDI-Projects he worked a consultant in cadastral issues and procurement processes. Back in an engineering company he completed his experience in land use planning and spatial development. Since 2013 he focuses as head for Department on building modern cadastre systems and holistic spatial information infrastructure.

CONTACTS

Jürg H. Lüthy

Acht Grad Ost AG

Wagistrasse 6

8952 Schlieren

SWITZERLAND

Tel. +41 43 500 44 48

Email: juerg.luethy@achtgradost.com

Web site: www.achtgradost.com

Christian Kaul

Office for Spatial Development

Canton of Zurich

Stampfenbachstrasse 12

CH-8090 Zürich

SWITZERLAND

Tel. +41 43 259 27 67

Email: christian.kaul@bd.zh.ch

Web site: www.are.zh.ch / www.geoinformation.zh.ch

Markus Luginbühl

Federal Office of Civil Aviation

Mühlestrasse 2

CH-3063 Ittingen

SWITZERLAND

Email: markus.luginbuehl@bazl.admin.ch

Web site: www.foca.admin.ch