QUALITY ASPECTS OF HIGH-DEFINITION MAPS

J. M. Lógó [1, *], N. Krausz [1], V. Potó [1], A. Barsi [1]

[1] Dept. Photogrammetry and Geoinformatics, Budapest University of Technology and Economics, Hungary -
(logo.janos.mate, krausz.nikol, poto.vivien, barsi.arpad)@emk.bme.hu

Commission IV, ICWG III/IVb

KEYWORDS: data quality, quality control, highly automated driving, HD map

ABSTRACT:

A self-driving vehicle is one of the most expected inventions in the near future. These vehicles are enabled by several technological developments, like artificial intelligence, robust control, vehicular sensors, and high-speed communication. But beyond all these elements, the essential component is the knowledge about reality. Our profession has answered that question with the development of high-definition (abbreviated as HD) maps. Fully automated driving (also called driverless transportation) must be reliable enough to entrust our lives to the car. This fact indicates that the applied technology and the used map must be of high quality. But how can the quality of such a map be expressed? We are looking for the answer in the current paper.
Following Carlo Batini's idea, the general approach is based on the triumvirate of data sources – quality dimensions – life cycle phases. Data sources cover aerial, terrestrial and mobile mapping products with the available highest technological care; furthermore, onboard vehicular sensing extends the corresponding data sets. Lifecycle phases focus on the production (data collection and processing technologies) expanded by conceptualization (pre-production) and data delivery and use (post-production). Quality dimensions are strongly related to the dimensionality of the data; they can be measured by dimension metrics.
The first part of the paper summarizes the applied data collection methodologies, emphasizing the output data. This description contains a summary of the processing mechanism – inevitably characterized by quality indicators. The paper aims to give a complete outline for the quality dimensions; we do not limit the resolution and accuracy dimensions, but other significant clusters like completeness or consistency are also discussed. Because the reality changes are enormous in transportation (vehicles, pedestrians, etc., are moving – even at higher speed) and the newly developing HD maps are expected to be live, actuality is a cardinal quality dimension as well. Vehicular technologies like SENSORIS give an excellent option to the equipped vehicles to download and use maps from the cloud and upload their field observations, opening a new way to maintain the map database. The so established crowd-sourced data collection intensely influences the map quality; therefore, this method generates quality-related issues that are also to be analyzed.
The second part of the paper is a case study, where a pilot site close to the university campus was selected. In this area, thousands of images were captured and uploaded into the Mapillary database. Artificial intelligence processes were applied for segmenting, classifying, and evaluating the content of the georeferenced imagery. The map database stores various object categories in the area, for example, pedestrian crossings, traffic signs, or trash cans. All extracted objects are available in georeferenced format, enabling spatial analyses to derive numeric quality indicators. The paper presents the complete results of this study.

1. INTRODUCTION

In 1888 Karl Benz' wife, Bertha took a vehicle named Model 3, supposedly without the knowledge of her husband, to visit her relatives and traveled from Mannheim via Heidelberg and Karlsruhe to Pforzheim (Herrtwich, 2018). In 2008 the Bertha Benz Memorial Route was officially approved, then in 2010, at the Mercedes-Benz research, the Bertha Drive Project was conceptualized (Ziegler et al., 2014). In 2013 an automated S 500 completed the route through the Black Forest. It needed a map used like a sensor, which level of details was significantly higher than of traditional navigation maps. The "high-definition map" was born (Herrtwich, 2018).
As a partner in the project, HERE Technologies participated in the mapping procedure. After the idea of HD map was created, researchers at the Karlsruhe Institute of Technology suggested extending the original concept with camera image storage to increase the vehicle positioning, and then a layer with lane descriptions was additionally merged. This three-layer map is based on real-time sensor input; therefore HD Live Map name was introduced. Now, the company operates the mapping workflow in the cloud, receiving observations from the vehicles

and evaluating the acquired data by artificial intelligence resulting near-real-time database. They called after that the HD Live map "self-healing" (Walker, 2018).
The Enhanced Digital Maps (EDMap) project initiated in 2001 together with map provider Navteq and the US Federal Highway Administration and National Highway Traffic Safety Administration, focused on map quality with respect to the various vehicular assistants like Curve Speed Warning, Lane Following Warning, etc. Their approach was extended for map quality issues, where statistics-based quality metrics were applied. The geometric quality of the database was determined by relative accuracy of sampled segments and confidence level was also computed. The spatial deviation was found as the most crucial feature. They also examined the attributes (e.g., number of lanes, speed limits). (EDMap, 2004)
After 2016 as HERE introduced the HD Live Map, its quality characterization has begun, and the Quality Index was introduced. It is a technology (being currently under patenting), which derives measures about the map quality considering "hundreds or thousands" of sensor reads performed by vehicles resulting high degree of certainty. Less traveled roads get a lower quality index (Bonetti, 2018).

---

* Corresponding author

Liu et al. (2020) have used in contrast not only relative but also absolute accuracy. As a literature survey, they collected some map providers' accuracy parameters (and.com, Sanborn.com, TomTom.com), indicating the need for standardization. Their understanding of HD map consists of 3D point cloud, which was analyzed against Iterative Closest Point (ICP) and Normal Distribution Transformation (NDT) transformations. (Liu et al., 2020)

The paper is organized as follows: Section 2 presents the general concept elaborated under the supervision by C. Batini. Section 3 gives the details about the concept adapted for HD mapping. This section is a general theoretical quality model, whereas Section 4 brings practical examples with two case studies. These examples illustrate the defined terms for geometry and thematic classification oriented. Finally, reference literature can be found.

2. THE BATINI-MODEL FOR DATA QUALITY EVALUATION

Several approaches exist to describe the data quality; the information science-based model developed by C. Batini was successfully implemented also in geographic information science and remote sensing (Batini et al., 2017), (Albrecht et al., 2018). Thus, the authors believe that it suits similarly to express the data quality aspects of high-definition mapping.

The Batini model contains three pillars: data source, life cycle phases, and quality dimensions. The **data source** pillar describes and analyses all the data acquisition moments; it covers all data collection technology. The captured data are processed, transformed with the second pillar elements: it is called **lifecycle phases**, as the data items are in modifications during their life phases. This quality approach can be characterized via different aspects – as the **quality dimensions**' pillar systematically discusses. If these dimensions are aimed to express by numeric and/or textual descriptors, **dimension metrics** must be defined. This philosophy can be seen in Fig. 1.
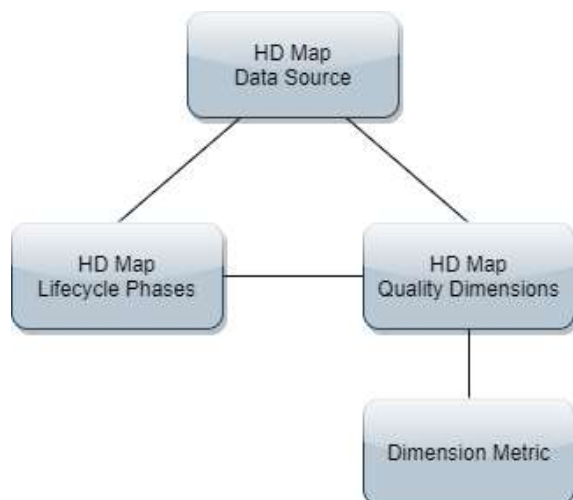

Figure 1. The pillars of Batini's data quality model

Using the data in an information system, the data quality approach extends to information quality. The demand against high-quality maps in self-driving has indicated a well-documented and quality-proved map database. This activity cannot ignore a sophisticated methodology covering regulations of capturing quality-oriented measures during data acquisition, processing, and map product delivery.

3. DATA QUALITY FOR HD MAPS

3.1. Data sources

Surveying and mapping are fundamental techniques for collecting reliable data. The oldest mapping method is the "traditional" geodetic method. Discrete points are surveyed – now with a high-performance total station or GNSS receiver, possibly a combination of these. It is a time-consuming surveying method. A considerable advantage is that it does not store unnecessary data, the measurement results are extremely accurate. Satellite imagery can be used to cover large areas; its spatial resolution does not reach that of terrestrial or aerial methods. The orthophotos as a result of the aerial surveying, are widely used in online map systems (e.g., Google Maps, OpenStreetMap). The disadvantages also weather dependency, shadows, and overlays are the problems. Over the last few years, remote sensing technology has advanced in leaps and bounds, and UAV image acquisition technologies have been developed. Basically, they collect data from a relatively small area by capturing images or by scanning. However, laser scanners are used more and more as surveying instruments for various applications. Terrestrial laser scanning (TLS) can cover various areas and can be used at distances of up to 600-1000 m. In addition to the 3D point cloud, it also captures images used to color the point cloud. Its disadvantages include the need for visibility, the handling of obstructions, the disturbing effect of flying dust, humidity, and dependence on precipitation. Mobile Mapping Systems (MMS) is a dynamically evolving survey method. MMS combines various navigation and remote sensing technologies on a common moving platform; it allows large quantities and high-precision data to be collected and georeferenced on board, mainly using a positional subsystem, which consists of GNSS, inertial, and wheel rotation sensors. Mobile mapping data seems to be essential in the development of autonomous driving and automated vehicle systems.

3.2. Lifecycle phases

Creating HD maps is a complex activity, which requires cutting-edge technology with adequate sensors. The goal of HD maps is to deliver all details for the increasingly automated vehicles. Because the map is used now by the driving assistant functions, but in the future, within the vehicle's decision mechanism, the requests against quality are very high.

The data lifecycle phases contain three main groups. The first is **before production**, where the mapping workflow should be conceptualized. This is nothing else than the elaboration of the best available data capturing and processing work chain. The second group is the **production** with the field data collection followed by the data handling steps. Generally, this phase is dedicated to mobile mapping technologies. The end of these efforts is the available map database, delivered and used in the last phase (**after production**). Bearing in mind that map databases aren't static anymore and the reality changes are permanent, the HD map creation technology owns an obligatory recursion: the map update must be considered among these phases. Not least, the modern map update technology involves "additional" contributors: the map using vehicles can be equipped with data collection sensors, not merely to monitor their surroundings, but applying SENSORIS technology, they provide observations for the map update.

3.3. Quality dimensions and metrics

Our quality model for HD map – following the guideline of (Batini and Scannapieco, 2016) – can be summed up in the following table (Table 1):

Dimension name	Dimension metric
Resolution	
spatial	m (and related subunits)
radiometric	bit
spectral	μm
temporal	s
point density	point/m^3
Precision	
geometric	m
spatial	m
radiometric	%
spectral	μm (sometimes Hz)
temporal	s (and related subunits)
Accuracy	
spatial	m (and related subunits)
radiometric	%
temporal	s (and related subunits)
classification	confusion matrix and its derived metrics (e.g. TP, FP, TN, FN)
semantic	%
Completeness	%
Redundancy	%
Consistency	yes/no
Readability	yes/no
Trust of sources	yes/no or confidence level

Table 1. Quality dimensions and metrics
(TP – true positive, FP – false positive,
TN – true negative, FN – false negative)

The **Resolution** dimension characterizes the ability of the sensors to distinguish details during the measurement. It has a strong correlation to the sensing equipment, e.g., spatial resolution of a Lidar sensor or of a color camera have a very different meaning. While the Lidar resolution is for the distance measurement specific ability and its metric is m, the camera resolution is the differentiable pixel size. The latter can be measured on the sensor plane in μm or the captured reality in m. Point density is a dimension of a sensing instrument, where distance AND angle measurement influence the magnitude of captured points in the resulting point cloud.

Precision informs about the homogeneity of the captured data; the discretization of the distance values, or the distribution of the camera pixels. Many researchers agree that precision is an expression of random errors, e.g., reproducibility or simply a standard deviation.

The most used quality measure is **Accuracy**. It is usually understood as a deviation from a (sometimes only theoretically known) reference. The most frequently applied spatial accuracy is a measure to express, e.g., analyzed points' closeness to their ideal position.

The covariance matrix widely used in surveying is an excellent feature to express the essential accuracy description. The covariance matrix is furthermore an input of Kalman-filtering and other computations of vehicle control procedure, so the derivation of this matrix and then its use in map accuracy is highly suggested. It enables the control mechanism to consider the variant accuracy circumstances during travel, remembering that environmental measurements can be executed even by a vehicle driving directly in front of the ego vehicle.

Temporal accuracy is also called validity, expressing the goodness of information in time. This phenomenon is again of high importance in self-driving applications. Classification accuracy brings information about the quality of the stored classes, about the correct interpretation (understanding) of the

sensed and perceived objects, like pedestrians or traffic signs. Classification quality is often used in remote sensing data processing, so the already elaborated metrics, confusion matrix, and its derivatives (e.g. true or false positive, true or false negative) and overall accuracy, recall, and many more have meaning in HD mapping. Semantic accuracy is a very complex and hard-to-interpret dimension. It formally contains the quality of the semantic information, as relations between objects. This dimension should be analyzed and described precisely, because scene perception phase of vehicle control mechanism has strong connection to this measure! Automated vehicles do the environmental sensing and evaluation automatically, and the quality of the onboard and map provided content must be close to each other.

Completeness characterizes how the map database covers the reality with every object. The removal of the obsolete objects influences this dimension, so the metric is time-dependent, dynamically changing. Beyond this simple theoretical definition, the expression in numeric value indicates problems in practice. Redundant data storage was a meaningless term in the mapping world. With the appearance of HD map with its huge information content and its role in the automatic vehicle control, this phenomenon has got quickly serious weight. **Redundancy** describes in this context the controllability of the objects in the database: it indicates e.g., for automatically recognized objects, whether the decision can consider the relevant one.

Consistency means "agreement or harmony of parts or features" (Merriam-Webster, 2002). Harmony in HD map context means, as an example, that road markings and traffic signs do not contradict, so these database elements express the same traffic regulations. This dimension refers to a higher meaning level of the map database, so correct computer formalization is still expected. This dimension has a significant influence on the right decision similarly. Table 1 shows therefore only a simple human approximation taken from explicitly executed analyses.

Readability covers schema, instance, and thematic characterizations. In practice, all these aspects do not arise because of the strong standardization in the automotive industry. In HD mapping, such standards are forming, or currently available formats are widely accepted.

Trust of sources measure has enormous importance in the automotive industry. Potential suppliers must be checked before they can support any production phase. Considering HD maps, this dimension revives the "traditions" in car making and map use business, where the result is primarily an approval for the providers. Our quality approach wants to point to the potential numeric evaluation of such databases.

4. CASE STUDIES

The presented theory can be illustrated by two case studies. The first one is focusing on the geometry, where the analyzed data were obtained by downloading OpenStreetMap GPS tracks. The second case study demonstrates the importance of other quality features in HD map content, emphasizing traffic signs.

4.1. OpenStreetMap case study

The crowd-sourced OpenStreetMap is excellent navigation aimed map database, which also contains some registered and published GPS tracks. With these tracks, we wanted to analyze how the traveling vehicles measured their position – similarly to the expected vehicular observations – and then some quality measures were determined.

The pilot site is a four-leaf clover formed motorway junction near Hannover ("Kreuz Hannover-Ost", where motorways A2 and A7 intersect). The necessary GPS tracks were downloaded with the

Java OpenStreetMap editor (JOSM version 17702). The archived gpx format tracks could then be processed with MathWorks Matlab R2021a.

Five test points were selected on the different parts of the intersection, where some statistics of the crossing tracks were derived. These tracks were in variable distance from the test points; the longitude and latitude differences were calculated. Fig. 2 shows the positions of the test points.

Figure 2. The geometric case study pilot site near Hannover with the test points

The variations near the test points were collected into the observations matrix, then a covariance matrix was calculated. Fig. 3 shows the longitude and latitude errors (bar chart) and the variety of observations (plots).

There are test points with lower and higher deviations. The former belong to the places lying on the main courses, while the more significant deviations relate to the driveways. The relatively high differences in the variances and covariances emphasize their importance.

4.2. Mapillary case study

Mapillary, a Swedish founded startup company, started in 2013 to make street-level imagery and map data available to everyone. ("About | Mapillary," 2019, "Mapillary - Terms of Use.," 2019) In 2018 they had a database with 350 million images, several million investment money, and in 2020 April, they owned 1147 million images. This year Facebook acquired the company, but the services remain unchanged.

Because Mapillary's technology is created on crowd-sourced street view imagery with AI-based image evaluation, we have selected it as a classification test case. The image processing workflow starts with segmentation, followed by object detection and database content creation ends the procedure. Object detection is feasible to extract 1500 traffic sign classes, 42 point features (e.g., fire hydrants, street lamps), and 19 line features (e.g., lanes, guardrails). The mapping work chain is fully automatic, so it looks like onboard vehicular environmental sensing. We wanted to use a pilot site for quality evaluation with respect to Section 3.
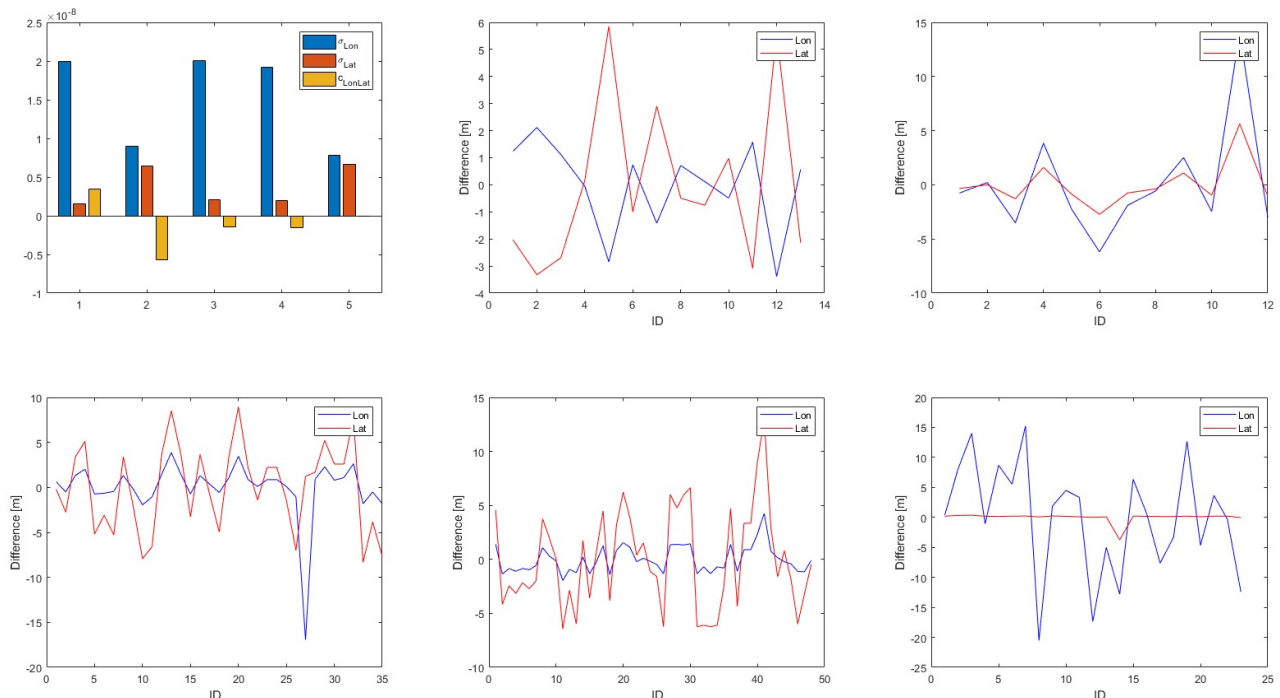

Figure 3. The obtained differences at the test points (OpenStreetMap case study).

The chosen test site is in the vicinity of the campus of Budapest University of Technology and Economics (BME). After downloading the entire database content for the test site (in October 2019), some elementary statistics were calculated: the size of the area is 0.7 km^2, it covers approximately 18 km long street network. There were – just some examples – 1520 traffic signs detections in 99 categories, 1097 poles, 641 street lights, 650 signage, 237 utility poles, 188 trash cans, 133 manholes, 29 fire hydrants, 22 bike racks, and one phone booth. Traffic lights were in 273 (upright), 183 (pedestrian), 16 (general), 3 (cyclist) and 1 (other) cases. An overview of the detected objects in the pilot sites can be seen in Fig. 4.
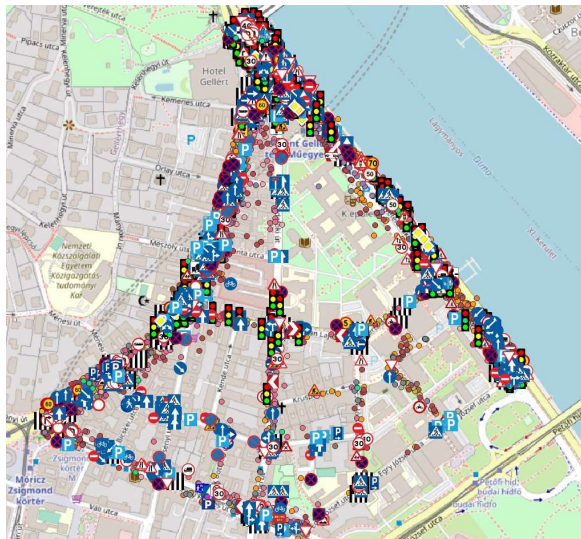

Figure 4. The Mapillary-detected objects in the pilot site

The geospatial analyses were executed in QGIS v3.18 Zürich. The downloaded object data are interpreted as point GIS objects having beyond the geometry also an attribute table. The original attribute table was extended by additional fields: sign and object codes, groupings, and a Boolean field for quality checking.
The classification accuracy of traffic signs means their correct recognition. The traffic sign database is available, which was cross-checked with valid Hungarian traffic regulations. For example, some signs were found and stored in the database with yellow background, incompatible with the Hungarian rules. Such a problem is not always a recognition problem: the "road construction" sign has the exact figure, just the background color differs. Speed limitation signs are with white (Hungary), yellow (Sweden!), and orange (USA) backgrounds; the shape varies between circular (Europe) and rectangular (USA) forms. (South Africa has both white and yellow signs differentiating between permanent and temporary limitations.) There were three wrong and 68 correct detected speed limit signs (96% correct). Furthermore, three wrong and 74 correct detected one-way signs (96%). There were 25 road work signs with the only available yellow background (evaluated them as correct) (Fig. 5).

Figure 5. The road work sign in the Mapillary database and in the Hungarian traffic regulation

Close to these road work sites, chevron signs were also detected: in 47 cases correctly (98%) and once wrongly.
There were two "wrong way" sign detections. There is no official such sign in Hungary, and funnily both are considering the same object: a red billboard containing the name of a Chinese fast-food restaurant.
The further analyses required a smaller test site, which was selected as a street intersection between two nearby streets.
Fig. 6 shows the crossing of the Budafoki Road (north-south directed street) with 1+1 lanes and Bertalan Lajos Street (east-west direction). The latter is a one-way street, as the arrow shows. The last captured images were taken as road works were observed; this is the reason for the found chevron signs. The position of the signs shows its geometric uncertainty.
In this crossing, four zebra crossings serve a safe place for pedestrians to cross these streets. The crossing is equipped with a traffic light, too. As a data quality question, the consistency can be studied here. Surface markings, like zebra crossings must be in harmony with adequate traffic signs (pedestrian crossing warning signs). It became clear that the signs are consistently placed with the road marking. The illustration also shows yield, don't stop, do-not-enter, some driving direction signs, and the traffic lights.
Humans performed these investigations; an automated technology is highly appreciated, but the complexity of the problem makes it hardly realizable.

CONCLUSION

The high-definition map is a promising new map database containing massive information about the roads and infrastructure, as well as their neighborhood. The HD map is expected intensively in self-driving use. The map providers and users, the automotive world requires a specific technology, which describes the quality issues of the database by objective features. To reach that goal, different efforts were executed, but primarily the analyses concentrated merely on the geometry.
We have proposed that the information science base of Batini's quality model suits this purpose. In the theoretical overview of the quality model, the three pillars were presented. The focus was set on the dimensions and their metrics: beyond the widely used resolution, precision, and accuracy, new features are introduced. We formulated not only coarse definitions for these terms but pointed on the crucial potential role of the covariance matrix. The freshly introduced features are demonstrated, too.
As practical support, we have conducted two examinations. These studies demonstrate the importance of the geometric investigations (OpenStreetMap case study), whereas the research with Mapillary-detected traffic signs and road markings aims to underscore the increasing potential of the newly defined quality dimensions.

Figure 6. The smaller traffic sign test site

REFERENCES

About | Mapillary [WWW Document], 2019. URL https://www.mapillary.com/about (accessed 6.19.19).

Albrecht, F., Blaschke, T., Lang, S., Abdulmutalib, H.M., Szabó, G., Barsi, Batini, C., Bartsch, A., Kugler, Z., Tiede, D., Huang, G., 2018. Providing data quality information for remote sensing applications, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. pp. 15–22. https://doi.org/10.5194/isprs-archives-XLII-3-15-2018

Batini, C., Blaschke, T., Lang, S., Albrecht, F., Abdulmutalib, H.M., Barsi, Á., Szabó, G., Kugler, Z., 2017. Data quality in remote sensing, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. pp. 447–453. https://doi.org/10.5194/isprs-archives-XLII-2-W7-447-2017

Batini, C., Scannapieco, M., 2016. Data and Information Quality, Data-Centric Systems and Applications, Data-Centric Systems and Applications. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-24106-7

Bonetti, P., 2018. High five to 1 million kilometers of HERE HD Live Map! [WWW Document]. Here360. URL https://360.here.com/1-million-kilometers-here-hd-live-map (accessed 5.15.19).

EDMap, 2004. Enhanced Digital Mapping Project Final Report. Washington, DC.

Herrtwich, R., 2018. The evolution of the HERE HD Live Map at Daimler [WWW Document]. Here360. URL https://360.here.com/the-evolution-of-the-hd-live-map (accessed 6.5.19).

Liu, R., Wang, J., Zhang, B., 2020. High Definition Map for Automated Driving: Overview and Analysis. J. Navig. 73, 324–341. https://doi.org/10.1017/S0373463319000638

Mapillary - Terms of Use. [WWW Document], 2019. URL https://www.mapillary.com/terms (accessed 9.13.14).

Merriam-Webster, 2002. Dictionary by Merriam-Webster: America's most-trusted online dictionary [WWW Document]. Merriam Webster Dict. URL https://www.merriam-webster.com/ (accessed 4.15.21).

Walker, B., 2018. Daimler to bring HERE HD Live Map to the roadway [WWW Document]. Here360. URL https://360.here.com/daimler-and-here-to-bring-hd-live-map-to-the-roadway (accessed 5.15.19).

Ziegler, J., Bender, P., Schreiber, M., Lategahn, H., Strauss, T., Stiller, C., Dang, T., Franke, U., Appenrodt, N., Keller, C.G., Kaus, E., Herrtwich, R.G., Rabe, C., Pfeiffer, D., Lindner, F., Stein, F., Erbs, F., Enzweiler, M., Knoppel, C., Hipp, J., Haueis, M., Trepte, M., Brenk, C., Tamke, A., Ghanaat, M., Braun, M., Joos, A., Fritz, H., Mock, H., Hein, M., Zeeb, E., 2014. Making Bertha drive-an autonomous journey on a historic route. IEEE Intell. Transp. Syst. Mag. 6, 8–20. https://doi.org/10.1109/MITS.2014.2306552