

Quality framework for combining survey, administrative and big data for official statistics

Yvonne A.P.M. Gootzen^{a,b,*}, Piet J.H. Daas^{a,b} and Arnout van Delden^a

^a*Statistics Netherlands, The Hague, The Netherlands*

^b*Eindhoven University of Technology, Eindhoven, The Netherlands*

Abstract. Creating statistics by combining data sources allows for the production of new, more timely and/or more detailed statistics. With an intended statistical output in mind, and various potentially useful data sources, there is a need to assess the potential of each source to contribute to the intended statistic. Quality frameworks provide tools for such tasks. This paper proposes a quality framework that includes dimensions applicable to survey, administrative and big data to support the assessment of the potential of each source to contribute to the intended statistic. The framework is applied to a case study of mobility data and a case study of virus particle detection in sewage data.

Keywords: Data integration, selectivity, derivation variables, validation

1. Introduction

The amount of multi-source statistics based on a combination of survey and administrative data is increasing. Besides survey and administrative data, the first examples of successful applications of big data in official statistics are appearing [1–3]. Big data comes with its own challenges, which are partly different from the challenges regarding the use of survey or administrative data. To create a (new) intended statistic, we must first assess whether the contents and quality of available data sources are sufficient. This can be approached in a systematic way using a quality framework. The experience at Statistics Netherlands reveals that quality frameworks created for survey and administrative data alone [4,5] cannot be completely applied to big data sources. In such a framework, the nature of a big data source is usually so different that the evaluation becomes very uninformative. Similarly, quality frameworks specifically developed for big data [6–8] are not designed to include the most relevant informa-

tion in survey and administrative data. Therefore, the need arises for a multi-source quality framework that is relevant and applicable to survey, administrative data, big data and combinations thereof.

Frameworks for two of the three types of sources have been developed for survey and administrative data [9], and for survey and big data [10]. Quality frameworks agnostic to the type of input data are often focused on statistical output rather than on the input data [11,12]. Though the intended application of such frameworks is not specifically to assess a single data source for the purpose of creating a multi-source statistic, the underlying quality dimensions are relevant to consider for specific sources [13,14].

The current paper repurposes similarities observed between existing quality frameworks and proposes a framework applicable when survey, administrative data and big data are combined. The framework is meant to be applied during the design phase of a (new) statistic. After application of the framework and selecting which sources will be used for the statistic, an exploratory analysis of the data sources is recommended to validate the assessments of the framework and to select which methods are most suitable. In terms of the hyperdimensions introduced by Karr, the proposed framework mainly focuses on the data hyperdimension of qual-

*Corresponding author: Yvonne A.P.M. Gootzen, Statistics Netherlands, P.O. Box 24500, 2490 HA Den Haag, The Netherlands. Tel.: +31 455706870; E-mail: yapm.gootzen@cbs.nl.

ity [15]. The framework takes into account the target variable and target population of the intended statistic. Additionally, it includes the intended aggregation level and any available accompanying data.

The remainder of this paper is organised as follows. In Section 2 we describe the dimensions and categories of the quality framework. In Section 3 the quality framework is applied to two case studies. Finally, Section 4 contains a conclusion as well as a discussion.

2. Dimensions for categorisation of data sets

The quality framework presented in this paper is applied to multiple data sets by individually assessing each data set and later combining the results. The quality of each individual data set is assessed by a set of dimensions. Each dimension consists of categories that were chosen such that they summarise the information relevant to the process of combining the data set with other data sets to create an intended statistic within a given context. The other data sets which are considered for combining with the current data set, are called accompanying data. More than one category may be applicable per dimension.

The starting point before application of the framework, is to define the context in which it will be applied. The context can be interpreted as the perspective from which we will look at the data set, and should be determined before the data set is categorised. The context of the intended statistic consists of a target variable, target population and aggregation level, as well as available accompanying data. The time dimension is considered part of the aggregation level.

Each of the sources that is considered for use with respect to the intended statistic is assessed individually. In the context of this assessment, the other sources that are not subject of this assessment, are considered as accompanying data. After this, the assessments per source are combined to judge whether the combination of sources can be used to produce the intended statistic.

The assessment of some dimensions requires information about the collection method of the data, especially if it was carried out by other organisations. If this information is unavailable, it may not be apparent from the data itself which categories apply. In these cases, the *undetermined* category is most applicable.

Each dimension is meant to answer a key question relevant to the usefulness of the data source for the purpose specified within the context. We propose the following dimensions for categorising a data set for the purpose of combining it with other sources within a specified context:

Relevance: Does the data contain information related to the intended statistic?

- *Directly relevant.* The data contains the target variable, or a variable that so closely resembles the target variable which implies that no accompanying data, variable or model is needed to extract the target variable at the intended aggregation level.
- *Indirectly relevant.* The data contains information that can be relevant to the intended statistic, but only in combination with an accompanying data set, variable or model. Or the aggregation of the target variable is only available for a unit type which is non-trivially linked to the intended aggregation level.
- *Irrelevant.* The data does not contain information that is relevant to the user. If this data was not available, it would not influence the final result. If new accompanying data become available in the future, the classification in this category should be reconsidered.

Population coverage: How complete is the population in the data compared to the target population?

- *Complete coverage.* Every unit in the target population occurs exactly once in the data.
- *Duplication.* Units of the target population are included more than once in the data.
- *Overcoverage.* The data contains units which are not part of the target population.
- *Undercoverage.* The data contains units of the target population. Some units of the target population are not present in the data.
- *Undetermined.* No direct link is available between the units in the data and the target population. No claims can be made about the coverage of the data set.
- *No unit-type coverage.* The unit type of the data is different from the unit type of the target population. Some accompanying data or modelling is needed to convert the unit type to that of the target population, before the population coverage can be assessed.

Population representativity: To what extent can we derive whether the set of units in the data represent the target population?

- *Known inclusion probabilities.* The inclusion probabilities of units in the target population are known. This includes cases with a probability sample or deterministic selection.

- *Unknown inclusion probabilities.* The inclusion probabilities of units in the target population are not known. This includes cases with a non-probability sample.
- *Non-zero inclusion probabilities.* All inclusion probabilities of units in the target population are larger than 0.
- *Zero inclusion probabilities.* Some inclusion probabilities in the target population are 0.
- *Undetermined.* Population representativity cannot be determined if no unique identifiers of units are available, or the unit type of the target population is not covered in the data. No accompanying variables are available to measure the representativity.

Variable validity: How well does the data set measure the target variable?

- *Perfect.* The definition of the target variable is identical to the definition used in the data and no measurement errors occur.
- *Definition inconsistency.* The definition of the target variable or unit type is different from the definitions in the data. Definition inconsistency is also known as concept (in)validity.
- *Measurement error.* A measurement error causes the values in the data to be different from the intended definition in the data set.
- *Modelling error.* The variable in the data set was previously derived from a different variable by an imperfect model or derivation.
- *Processing error.* The variable contains errors from a previous processing step, such as data entry or manual editing.
- *Causal error.* Errors have been introduced by disregarding causal connections between variables in a previous version of the data (or multiple data sets if the current set is a combination of data sets).
- *Undetermined.* The definition, measurement process or modelling process of the variable in the data is (partially) unknown.

Concept stability: Does the assessment of the data set in the variable validity dimension remain stable over time?

- *Stable.* The level of definition consistency, measurement error and modelling error of the data set with respect to the target variable are stable over time.
- *Concept drift.* The level of definition consistency of the data set compared to the target variable changes over time. Note that this can also be due to a change of definition in the target variable over time, when the change is not present in the data set.

- *Unstable.* Either the measurement error or the modelling error of the data set changes over time.
- *Not applicable.* For the purpose of this study, the concept stability is irrelevant. This may be the case for sources where the target variable is not included in the data source.

Correctability: Can inaccuracies (such as bias) in the data be corrected, for instance by modelling or by combining with other data sets?

- *Unnecessary.* No correction is needed because the data accurately measures the target variable.
- *Self-correctable.* The inaccuracies in the data set can be corrected using accompanying variables in the data set itself, without usage of other data sets.
- *Supplement-correctable.* The bias in the data set can be corrected using accompanying sets, possibly by linking them with variables from the current data set.
- *Uncorrectable.* The data cannot be corrected within the given context.
- *Undetermined.* It is unclear whether the data can be corrected within the given context.

Recentness: What is the nature of the time lag between the occurrence of a phenomenon and the moment it is first reported in the data?

- *Event-based.* The data related to an event becomes available relatively soon after the event occurred, without grouping multiple events into a single "delivery" of data, resulting in a stream of data.
- *Periodically.* A system is in place that guarantees a periodical release of data.
- *Sporadically.* Availability of the data is dependent on individual actions that are hard to anticipate. Or there might not be any guarantee that a successor of a data set will become available in the future.

Processing timing: What is the nature of the time lag between obtaining access to the data and the intended statistic being ready for publication?

- *Instantly.* An automatic system is in place that ensures data can be processed virtually instantly, which is at least before the next instalment of data is available.
- *Automated.* Whenever a new instalment of data is available, it can be processed with few human interventions.
- *On request.* The data is processed manually and the process is started on request each time a new instance of the data becomes available.

Accessibility: To what extent are there limitations to access the data?

- *Full access*. Legal access is guaranteed for the foreseeable future and does not limit the options based on the technical availability of the data. Usage of the data is allowed for the publication of the intended statistic.
- *Paid access*. The data is accessible for a financial compensation.
- *Limited access*. Legal issues either prevent the user from accessing the full data or limit the scope of the intended statistic.
- *No access*. There is currently no access to the data, or usage of the data for the intended statistic is not allowed.

Meta-data: To what extent are the definitions and information of the data known?

- *Synergetic*. The metadata is complete and well defined and fits perfectly with metadata from accompanying data sets and the intended statistic.
- *Standard-compliant*. The metadata adhere to standards in the field of application. Standards may vary between different fields of application.
- *Well-defined*. The metadata is complete and well defined, but does not fit well with metadata from accompanying data sets or the intended statistic.
- *Ill-defined*. The metadata is largely available, but vague and allows for multiple interpretations.
- *Incomplete*. The metadata is largely unavailable, and exploratory data analysis or assumptions are necessary to interpret the data.

Comparability: To what extent can the data be compared to data from parallel research?

- *Fully comparable*. There is a general consensus of definitions used in the data and in parallel projects. In theory, the data can be interchanged with data from parallel research without major methodological complications.
- *Partly comparable*. Some discrepancies between the data and data in parallel studies can be expected, but a conversion allows for the outcome of the study to be compared to parallel studies.
- *Non-comparable*. The data is so unique that it is unlikely the results will be comparable to parallel studies.

The dimensions population coverage, population representativity, variable validity, concept stability and correctness are aspects of accuracy. The dimensions recency and processing timing both are aspects of timeliness and punctuality.

Note that some categories are not mutually exclusive. Though the categorisation process would be more

straightforward if all categories within a dimension were mutually exclusive, the required number of categories to achieve this would be so high that we decided against it. Let us discuss the *variable validity* dimension as an example. Here, the several types of error and inconsistency may occur simultaneously. The status of the three types of inconsistencies and errors can be “the issue is not present”; “the issue is present” and “unknown whether the issue is present or not”. This leads to a total of $7^3 = 343$ categories, whereas the current 7 categories provide enough definitions to accurately describe a data set in the variable validity dimension. The choice for non-mutual exclusive categories was made to keep the framework orderly, while not taking away the multi-layered real-world complexity of a data set.

2.1. A note on coherence

Readers familiar with quality dimensions in official statistics might wonder why the term *coherence* is not included in the list of dimensions in the framework. Eurostat states that *coherence is usually used when assessing the extent to which the outputs from different statistical processes have the potential to be reliably used in combination*, whereas *comparability is used when assessing the extent to which outputs from (nominally) the same statistical process but for different time periods, different countries/regions and/or different domains have the potential to be reliably used for comparisons* [16]. The key difference between *comparability* and *coherence* is that *comparability* is about the form of the intended output, and *coherence* is about the value of the output. The current framework is intended for the planning phase of a new statistic. During this time, the form of the intended output is known, but not the value of the output.

Additionally, the danger of including preexisting statistics in the *context* is that they could be interpreted as constraints. The seniority of an established statistic should not be misinterpreted as the ground truth, and should not limit the development of a statistic based on new data sets or methods. In the case of a new statistic that is not coherent with an established statistic, we prefer to see both existing alongside with an explanation of the differences in data sets or methods, rather than the new statistic being discarded because of dissimilarities to the established statistic.

3. Case studies

We illustrate the proposed quality framework by applying it to two studies in which multiple types of data sources were combined.

Table 1
Application of the proposed quality framework to the mobility case study

	Survey data	Admin. data	Infra. data	Sensor data
Relevance	Indirectly relevant (1)	Indirectly relevant	Indirectly relevant	Directly relevant
Population coverage	No unit-type coverage	No unit-type coverage (2)	Complete coverage	Undercoverage
Population representativity	Undetermined	Undetermined	Known inclusion probabilities	Known inclusion probabilities
Variable validity	Measurement error	Undetermined	Definition inconsistency, modelling error	Measurement error, definition inconsistency
Concept stability	Not applicable (3)	Not applicable (3)	Concept drift (4)	Stable
Correctability	Supplement-correctable	Supplement-correctable	Unnecessary	Self-correctable (5)
Recentness	Periodically	Periodically	Periodically	Event-based
Processing timing	Automated	Automated	Automated	Automated
Accessibility	Full access	Full access	Full access	Full access
Meta-data	Synergetic, Standard-compliant	Synergetic, Standard-compliant	Well-defined	Well-defined
Comparability	Partly comparable	Partly comparable	Fully comparable	Fully comparable (6)

3.1. Case study: mobility

In this case study, administrative data, survey data and big data were combined with Dutch road network data [17]. Traffic intensities on the road network in the Netherlands were studied by combining four different sources (discussed below). Before applying the framework to the available data sets, let us first formalise the context of the study. The *target variable* of the intended statistic is the number of passenger cars and motorcycles and the *target population* is the set of road segments in the Netherlands. The *aggregation level* is defined as the morning rush hour peak per road segment. The framework is applied four times: once to each source, with the other three remaining sources functioning as accompanying data sources. Table 1 shows the result of applying the quality framework to the data sets used in the case study, given this context.

The four data sources each have their own unit types and, at a first glance, cannot be combined. The first source is the national travel survey of the Netherlands, where people are asked to report on their transportation movements (including modality and motivation) during a particular day [18]. The unit type of this source is ‘person’. The survey data of persons that travel for work is used. The second source concerns a combination of different administrative data sets. It has the unit type ‘person’ and contains background characteristics. The third source is based on Open Street Map data, more specifically the road network of the Netherlands [19]. The unit type of this source is ‘road segment’. It includes the location and geometry of each segment, as well as connections to other segments. The last source is traffic loop sensor data which contains observations of traffic intensities for road segments per minute [3]. More details on each of the sources can be found in [17].

Let us discuss some of the assigned classifications that we consider to be the least straightforward. Each classification is marked by a corresponding number in Table 1.

1. *Relevance: survey data.* One of the reasons for assigning indirectly relevant for the relevance dimension is that survey data has the unit type ‘person’, whereas the target population of the intended statistic is on the level of road segments. The application of a route planner on the infrastructure data allows to bridge this gap and use the information from the survey on the road network.
2. *Population coverage: administrative data.* The target population is defined as the road segments in the Netherlands. The administrative data is person-based, a completely different unit-type.
3. *Concept stability: survey data and administrative data.* The target variable of the intended statistic is not present in the survey and administrative data. Therefore, the dimension concept stability is inapplicable to these sources.
4. *Concept stability: infrastructure data.* The available infrastructure for travel may change over time, which can affect the calculated routes from the infrastructure data. Using a data set that corresponds to the time stamp of the administrative data and the sensor data cancels out the effect of concept drift.
5. *Correctability: self-correctable.* The minute-based data was aggregated to a single average value for each road segment meant to resemble the morning rush hour (5 a.m. to 9 a.m.). To correct for measurement errors, the aggregate was averaged for all regular working days during a full month. Since these corrections were applied without using accompanying data, the sensor data was categorised as self-correctable.

Table 2
Application of the proposed quality framework to the sewage data case study

	Measurements data	Areas data	Register data
Relevance	Indirectly relevant	Indirectly relevant	Indirectly relevant
Population coverage	Complete coverage	Complete coverage	No unit-type coverage
Population representativity	Known inclusion probabilities	Known inclusion probabilities	Undetermined
Variable validity	Definition inconsistency, measurement error (1)	Definition inconsistency (2)	Undetermined
Concept stability	Concept drift (3)	Concept drift	Concept drift
Correctability	Supplement-correctable	Unnecessary	Unnecessary
Recentness	Event-based	Periodically	Sporadic
Processing timing	Automated (1)	Automated	On request
Accessibility	Full access (4)	Limited access (4)	Limited access (4)
Meta-data	Synergetic	Synergetic	Synergetic, Standard-compliant
Comparability	Partly comparable	Fully comparable	Fully comparable

6. *Comparability: sensor data.* We consider the sensor data to be fully comparable since it is likely that other countries have similar data that could be used in the same role if this project were to be carried out in another country.

The intended statistic is a result from the following approach. First, the survey data was used to train a transportation modality model that determines the probability of a certain modality, given the background characteristics of a person. Secondly, the transportation modality model was applied to the combined set of administrative data, which was subsequently aggregated into an origin-destination (OD) matrix. The OD-matrix is composed of pairs of neighbourhoods and the expected number of people that travel to work by car. Third, Open Trip Planner [20] was used to convert the OD-pairs to routes consisting of road segments which resulted in an expected intensity for each road segment. Essentially, the route planner acted as a converter between the two unit types (neighbourhoods and road segments). Finally, the minute-based traffic loop data was filtered based on vehicle length to include only short vehicles such as passenger cars and motorcycles while excluding longer vehicles such as trucks. The data was then aggregated to one intensity value for each sensor by taking the sum of all observations during the morning rush hour. This way, most travel from home to work is taken into account while minimising the inclusion of travel for leisure or travel from work to home, as they predominantly tend to travel outside of the morning rush hour. The intermediate result is a data set with two variables for a set of road segments: expected intensity and observed intensity. Assuming the observed intensity to be the ground truth, a model was trained that calibrates an expected intensity to be closer to the observed value. The model can then be applied to all road segments (even the ones where no observed intensities are available) to produce a calibrated value of the

expected intensity. The complete approach applied for combining the four sources mentioned is described in detail in the report [17].

3.2. Case study: Virus particles detection in sewage data

As a second case study, we include a project from the Dutch National Institute for Public Health and the Environment (RIVM) where coronavirus is monitored in sewage data and combined with administrative sources to create statistics about the number of virus particles per 100.000 people in local regions [21]. The context in this case study consists of the *target variable* number of virus particles per 100.000 people per day, the *target population* all inhabitants of the Netherlands and the *aggregation level* is administrative region.

The first source used in this case study contains *sewage measurements* for all sewage installations in the Netherlands [22]. The unit type of this source is sewage installation. Each installation is sampled about four times a week, resulting in measurements for the respective sewage area. The number of virus particles is indicative of COVID-19 in the area serviced by an installation. The unit types of geographical data published by Statistics Netherlands are commonly used geographical regions such as provinces and municipalities. The borders of sewage installation areas however, do not always align with the municipality based regions. The second source in this case study is *sewage areas*, which contains geographic information on the areas serviced by each sewage installation. These data are obtained from local water authorities and potentially have different reference dates. Changes in the sewage network might not be immediately included in the data. The third and last source in this case study was produced by Statistics Netherlands and contains information on the number of inhabitants in serviced areas in combination with municipality based regions [23]. This

source contains both sewage installation units as well as municipalities. Let us call this data the *registry*-based source which in itself is the result of combining multiple sources: detailed geographical address-level data (BAG), the person registry (BRP), and sewage areas.

Let us discuss some of the assigned classifications that we consider to be the least straightforward. These notes are marked by corresponding numbers in Table 2.

1. *Variable validity & Processing timing: measurements data.* A considerable amount of manual effort is required to process the samples and create measurements. The manual work could lead to measurement errors. This process has proven to be reliably fast and is therefore considered to be automated.
2. *Variable validity: areas data.* The areas data are provided by local water authorities and may have some minor inconsistencies in their definitions among different authorities.
3. *Concept stability: measurements data.* The measurement method might need updating when new virus variants arise.
4. *Accessibility: measurements data, areas data and register data.* RIVM was chosen as the point of view for assessing accessibility.

To arrive at the number of particles per 100.000 people per day in local regions, the sources were combined in the following way. The number of inhabitants in the register data was used to calculate a weight for each combination of sewage area and municipality. This data set has been created for multiple years, with weights (slightly) changing each year and the first of January as date of reference. The weights were used to convert the measurements data per installation to measurements per municipality based region. The resulting statistic can be updated for every new measurement, which happens about four times a week in practice.

4. Conclusion and discussion

In this paper, we propose a quality framework that can be used when combining survey, administrative and big data for official statistics. The framework consists of dimensions and categories and focuses on the input quality of an individual source. The selected categories depend on the context: the intended statistic and the accompanying data. The framework enables the user to find the strengths and weaknesses of each source and the combination thereof. It is particularly useful to gain insights in the possible ways by which data

sources can be combined. The quality framework has been applied to two case studies where all three types of sources were combined. Applying the framework to each source separately and comparing the outcome as a whole allowed for an overview of the challenges that were encountered during each case study.

The framework was designed for cases with one or more target variables for which the process of obtaining the intended output is (nearly) the same for all target variables. In cases where many target variables are intended and each requires a distinctly unique approach of combining sources, the framework should be applied separately for each target variable.

We note that the quality of the final statistical output not only depends on the quality of the input sources, but also on the choices about the process and models used to combine these sources. After assessment of the candidate data sources with the framework, additional design work is needed in selecting which methods are most suitable. Additionally, a positive assessment of a source by this framework should not be taken as a guarantee that future iterations of the source are also fit for the purpose of creating the intended statistic. Any changes in a data source or the context should be reassessed using the framework. This framework systematically reviews the potential quality issues when combining sources. It does not, however, propose solutions to these quality issues.

The case studies illustrated a number of key points, particularly when non-trivial modelling steps were included to obtain the intended statistic. In the mobility case study, the network data source served as a link between the two populations. Here, it became clear that even though the populations differed in a number of sources, they could still be combined in a useful way. A similar situation was observed in the sewage case study, where conversion weights between two non-congruent area definitions were the key to combining measurement data and administrative data. The combination of categorisations for each source helps to identify the key points for data integration beforehand and helps to identify similar situations, which allows for re-use of the solution.

Acknowledgments

We gratefully acknowledge Marko Roos and Jeldrik Bakker for their valued knowledge about the case studies. We thank the two anonymous peer-reviewers who improved the quality of this paper by their insightful comments and constructive criticism. The map data used in this study is copyrighted by all OpenStreetMap

contributors and available from <https://www.openstreetmap.org>.

References

- [1] Eurostat. Report about possible new statistical output based on (European) AIS data. Eurostat; 2018. Available from: https://ec.europa.eu/eurostat/cros/sites/default/files/WP4_Deliverable_4.7_Possible_new_statistics_using_AIS_2018_03_31.pdf.
- [2] ESSnet Big Data II. Work package K: Methodology and quality. Deliverable K10: Report describing the methodological steps of using big data in official statistics with a section on the most important research questions for the future including guidelines. Eurostat; 2020. Available from: https://ec.europa.eu/eurostat/cros/sites/default/files/WPK_Deliverable_K10_Report_describing_the_methodological_steps_..._2020_11_20_Final.pdf.
- [3] Puts MJ, Daas PJ, Tennekes M, de Blois C. Using huge amounts of road sensor data for official statistics. *AIMS Math*. 2018; 4(1): 12-25. doi: 10.3934/Math.2019.1.12.
- [4] Groves RM, Lyberg L. Total survey error: Past, present, and future. *Public Opin Q*. 2010; 74(5): 849-79. doi: 10.1093/poq/nfq065.
- [5] Bakker BF. FOCUS: Quality evaluation of register-based statistics. In: *Proceedings of European Conference on Quality in Official Statistics 2018*. Krakow; 2018. pp. 1-10.
- [6] Batini C, Rula A, Scannapieco M, Viscusi G. From data quality to big data quality. *J Database Manag*. 2015; 26(1): 60-82. doi: 10.4018/JDM.2015010103.
- [7] ESSnet Big Data II. Work package K: Methodology and quality. Deliverable K9: Revised version of the methodological report. Eurostat; 2020. Available from: https://ec.europa.eu/eurostat/cros/sites/default/files/WPK_Deliverable_K9_Revised_version_of_the_methodological_report_2020_11_17_Final.pdf.
- [8] UNECE. A Suggested Framework for the Quality of Big Data. UNECE; 2014. Available from: <https://statswiki.unece.org/download/attachments/108102944/Big%20Data%20Quality%20Framework%20-%20final-%20Jan08-2015.pdf>.
- [9] de Waal T, van Delden A, Scholtus S. Commonly used methods for measuring output quality of multisource statistics. *Span J Stat*. 2020; 2(1): 79-107. doi: 10.37830/sjs.2020.1.05.
- [10] Amaya A, Biemer PP, Kinyon D. Total error in a big data world: adapting the TSE framework to big data. *J Surv Stat Methodol*. 2020; 8(1): 89-119. doi: 10.1093/jssam/smz056.
- [11] OECD. Quality Framework for OECD Statistical Activities. Organisation for Economic Co-operation and Development; 2011. Available from: <https://www.oecd.org/sdd/qualityframeworkforoeecdstatisticalactivities.htm>.
- [12] United Nations. UN Statistics Quality Assurance Framework. Committee of the Chief Statisticians of the United Nations System; 2018. Available from: <https://unstats.un.org/unsd/unsystem/documents/UNSQAF-2018.pdf>.
- [13] Puts MJ, Daas PJ. Machine Learning from the Perspective of Official Statistics. *Surv Stat*. 2021; 84: 12-7.
- [14] Schober MF, Pasek J, Guggenheim L, Lampe C, Conrad FG. Social media analyses for social measurement. *Public Opin Q*. 2016; 80(1): 180-211. doi: 10.1093/poq/nfv048.
- [15] Karr AF, Sanil AP, Banks DL. Data quality: A statistical perspective. *Stat Methodol*. 2006; 3(2): 137-73. doi: 10.1016/j.stamet.2005.08.005.
- [16] European Statistical System (ESS). Handbook for quality and metadata reports – 2020 edition. Eurostat; 2020. Available from: <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-19-006>.
- [17] Gootzen YA, Roos MR, Bostanci I. Data Collection for City and Subnational Statistics – Milestone: comparison between patterns from register- and big data sources in the mobility network. Den Haag: Statistics Netherlands; 2022.
- [18] Centraal Bureau voor de Statistiek. *Onderweg in Nederland (ODiN) 2018–2020* (in Dutch): Report describing the final recalculations for the years 2018, 2019 and 2020 for the Dutch mobility survey. Statistics Netherlands; 2022. Available from: <https://www.cbs.nl/nl-nl/longread/rapportages/2022/onderweg-in-nederland--odin---2018-2020>.
- [19] OpenStreetMap contributors. OpenStreetMap, editor. Planet dump retrieved from <https://planet.osm.org>; 2022. Available from: <https://www.openstreetmap.org>.
- [20] OpenTripPlanner contributors. OpenTripPlanner; 2022. Available from: <https://www.opentripplanner.org>.
- [21] Rijksinstituut voor Volksgezondheid en Milieu. Berekening cijfers rioolwatermetingen COVID-19 (versie 6). RIVM; 2022. Available from: <https://www.rivm.nl/documenten/berekening-cijfers-rioolwatermetingen-covid-19>.
- [22] Rijksinstituut voor Volksgezondheid en Milieu. COVID-19 Nationale SARS-CoV-2 Afvalwatersurveillance. 2020. Available from: <https://data.rivm.nl/meta/srv/dut/catalog.search#/metadata/a2960b68-9d3f-4dc3-9485-600570cd52b9?tab=general>.
- [23] Statistics Netherlands. Population per sewage installation. Statistics Netherlands; 2021. Available from: <https://www.cbs.nl/nl-nl/maatwerk/2021/06/inwoners-per-rioolwaterzuivering-sinstallatie-1-1-2021>.