

A Practical Data Quality Assessment Method for Raw Data in Vessel Operations

Gang Chen¹, Jie Cai², Niels Gorm Maly Rytter² and Marie Lützen³

Received: 27 February 2022 / Accepted: 16 September 2022
© Harbin Engineering University and Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

With the current revolution in Shipping 4.0, a tremendous amount of data is accumulated during vessel operations. Data quality (DQ) is becoming more and more important for the further digitalization and effective decision-making in shipping industry. In this study, a practical DQ assessment method for raw data in vessel operations is proposed. In this method, specific data categories and data dimensions are developed based on engineering practice and existing literature. Concrete validation rules are then formed, which can be used to properly divide raw datasets. Afterwards, a scoring method is used for the assessment of the data quality. Three levels, namely good, warning and alarm, are adopted to reflect the final data quality. The root causes of bad data quality could be revealed once the internal dependency among rules has been built, which will facilitate the further improvement of DQ in practice. A case study based on the datasets from a Danish shipping company is conducted, where the DQ variation is monitored, assessed and compared. The results indicate that the proposed method is effective to help shipping industry improve the quality of raw data in practice. This innovation research can facilitate shipping industry to set a solid foundation at the early stage of their digitalization journeys.

Keywords Data quality; Vessel operations; Shipping; Validation rules; Noon reports

1 Introduction

With the continuously developing of ICT (Information and Communications Technology), the IOT (Internet of Things) is built as a new infrastructure for the next wave of innovation (Hermann et al., 2016). It has started to transform some traditional low-tech industries into high-tech ones, e. g.,

transportation logistics. Large volume of information is accumulated with high speed and/or high variety, which is normally defined as the concept of “big data” (De Mauro et al., 2015). The proper storage and management of big data are necessitated so that it can be properly analyzed to facilitate a better decision-making and the quality control of the data is a prerequisite for the correctness of analysis results (Blake and Mangiameli, 2011; Bates, 2019). Such a revolution in both “big data” and digitalization is called Industry 4.0 in the onshore manufacturing industry (Hermann et al., 2016).

Similarly, shipping is conducting the same transition, calling Shipping 4.0 or cyber-shipping (Røseth, 2016). The current digitalization of shipping covers a wide range of activities, such as the navigation at sea and cargo operations in port. The digitalization of ship operations implies that a large amount of data are collected on board and then are sent to offices onshore. Data from operations can be used to better understand, optimize and improve navigation, energy consumptions, emissions and maintenance of vessels. Data can be based on noon reports from vessels, semi-automatic data combined with data logged automatically on board, or fully automatic systems which allow for data to be sent ashore with short time intervals, such as Automatic Identification System (Cai et al., 2021). An overview of vessel performance at daily level can be presented

Article Highlights

- A data quality assessment method is proposed for raw data in vessel operations ;
- Concrete validation rules are developed to quantify the data quality of noon reports from vessel operations;
- A case study was conducted in a shipping company to monitor the data quality variation;
- The proposed practical method is demonstrated to be effective to improve raw data quality in shipping practice.

✉ Jie Cai
jjec@iti.sdu.dk

¹ World Maritime University, Malmö, Sweden

² Department of Technology and Innovation, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

³ Department of Mechanical and Electrical Engineering, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

via noon reports. A noon report is a manual report prepared by the crew on board on a daily basis, summarizing relevant sailing information, e.g., sailing time, sailed distance, course, engine running time, power produced and total fuel consumption. Auto-log data are automatically recorded and logged in real time on board via sensors and ICT technology, which will provide more sufficient and hopefully accurate data than noon reports. However, it is still infeasible in the short term due to practical reasons such as costs and lack of regulations.

These vessel data can be used for vessel performance monitoring and improvement of operations, which has recently become a focus in shipping due to tightening environmental regulations and fierce market competition. For instance, there are today a considerable number of commercial software applications for vessel performance available at market using vessel operational data, e.g. Vessel Performance Solutions (VPS, 2021), Kongsberg Vessel Performance (Kongsberg, 2021), and the performance monitoring system SeaTrend (FORCE Technology, 2021). These systems enable reporting and analysis of vessel performance on several key performance indicators (KPIs), but have a common challenge regarding varying quality of operational data, which impacts the validity of their recommendations. It is current an industry practice to improve DQ from datasets by pre-processing methods such as removing outliers and filling null values, introducing a large risk for further information processing. Therefore, there is a need to develop a DQ quantifying method tailored to the needs of monitoring vessel operations and performance. A detailed literature review about DQ will be conducted in Section 2.

To fill this gap, a new methodology for measuring and improving DQ so as to better monitor vessel performance is developed in this paper. With this method, a set of data quality metrics have been developed for ship operations, which can be applied to first categorize data from specific dimensions, and next to score the data quality of an entire dataset so that the data quality of the entire dataset will be revealed through different severity levels. The rest of this paper is structured as follows. In Section 2, a detailed literature review on data quality assessment is conducted. In Section 3, the proposed method for the evaluating of DQ in vessel operations is introduced, including the data categorizes, data dimension metrics and the scoring criterion. A case study is further conducted on a noon report dataset in Section 4, applying the proposed method. In Section 5, discussions and limitations of the proposed DQ assessment method are provided, and finally concluding remarks are drawn in Section 6.

2 Literature review on the data quality assessment

Data quality is a multidimensional, complex and morphing

concept and has evoked much interest in academia and industry since 1980s (Tejay et al., 2004). The effective collection, representation, and application of data are important to an organization because these activities facilitate business operations and business analytics. Errors in data items from organizational databases, data warehouses, and data streams may lead to costly errors in business decisions (Dey and Kumar, 2010). Redman (1998) argued that poor data quality would impact setting strategy, data execution, ability to derive values from data and ability to align the organization. With the booming of Artificial intelligence (AI) and machine learning application in different engineering fields (Cai et al., 2022), the data quality issue is increasingly concerned by researchers. Karagiannidis and Themelis (2021) indicated that the accuracy of his proposed ship fuel consumption model based on operational data has been significantly increased when a proper DQ control was conducted in advance.

Therefore, significant research effort has been spent on estimation of the quality of data. The systematic study of data quality started in late 1990s in the information system field. It is found that the most effective way to evaluate data quality is through attributes or dimensions such as accuracy, reliability, timeliness, accessibility and consistency, representing a single aspect of data quality (Wang and Strong, 1996; Chengalur-Smith et al., 1999). Theoretically, Tejay et al. (2004) argued that the nature and scope of data quality dimensions changed with the changing of different semiotic levels (empirics, syntactics, semantics and pragmatics). Semiotics elucidates the intricacies associated with a sign as it moves from the physical world, where it is created, to the social world of norms. According to this research, people often mixed the dimensions of communication quality (empirics), information quality (semantics) and knowledge quality (pragmatics) into data quality (syntactic). Gradually, investigations of data quality were extended to data warehouses (Wang et al., 2006), e-commerce (Knight and Burn, 2005; Peltier et al., 2013), e-learning systems (Alkhatabi et al., 2011), collaborative business processes (Falge et al., 2012), web data (Caro et al., 2008; Yerva et al., 2012) and wireless sensor networks (Coen-Porisini and Sicari, 2012). The development of DQ assessment methods has been widely applied in various industries, e.g., finance (Pipino et al., 2002), manufacturing (Shankaranarayanan et al., 2003) and supply chain and logistics (Hazen et al., 2014).

In the transportation industry in general, data are increasingly becoming a valuable asset. The bad quality of data has caused problems in transportation operations, planning, traffic congestion information, transit and emergency vehicle management, and commercial truck operations (Ahn et al., 2008). To address data quality issues, Turner (2004) gave a definition of data quality as “the fitness of data for all purposes that required it”. Measuring

data quality requires an understanding of all intended purposes for the dataset. A framework for data quality management was developed and applied by Federal Highway Administration (US Department of Transportation, 2021), which was based on six fundamental measures of traffic data quality namely accuracy, completeness, validity, timeliness, coverage and accessibility. An important use of this framework was to estimate travel time, which is a key measure for the performance of transportation systems, therefore, receiving much attentions from transportation engineers (Eisele and Rilett, 2002). Richardson and Smith (2012) proposed a novel approach to measure errors so that the distribution of the errors can be quantified. Focusing on the automatic data collection systems, Liao and Davis (2012) explored several statistical analysis methods to detect measurement drifts. In their study, a mixture modelling technique using expectation maximization algorithm and cumulative sum (CUSUM) methods were explored for data quality. Jones-Farmer et al. (2014) further provided an overview on the analysis of data for use in process improvement and control charting.

In the specific shipping industry, a few studies have also been found. Røseth (2016) gave an overview of big data, data quality issues and possible solutions. Perera and Mo (2020) proposed a framework to handle high dimensional vessel data. Recently, the widely investigation of data-driven models for the prediction and monitoring of ship operational performance has been conducted on both noon reports and sensor data, such as Soner et al. (2018), Soner et al. (2019), Yan et al. (2020), and Karagiannidis and Themelis (2021). Among these applications, DQ was of great importance to determine the model accuracy and their generalization ability, which were pre-processed to some extent by these researchers and should be highly concerned.

According to such a literature review, it is found that visible efforts have been made to improve the data quality in different industries. However, few studies have focused on the quality dimension prioritization and quantitative quality assessment for vessel operational data, to the best knowledge of the authors. Therefore, the investigation to improve the DQ of operational data from vessels is urgently required. With a quantitative assessment method, the DQ of data in current shipping companies from operation can be constantly monitored and improved, which will further facilitate the development of data-driven models and eventually the operational decision-makings.

3 Methodology for assessing DQ of vessel operations

In this section, a specific method is proposed for the evaluation of DQ based on practice of vessel operations and the existing literature for managing DQ. In this method,

data are first categorized based on a categorical framework from practical operation functions. Then, a set of quality dimensions are applied on every data category. Afterwards, the quality of each data item after categorization is scored based on proposed validation rules.

3.1 Data categories

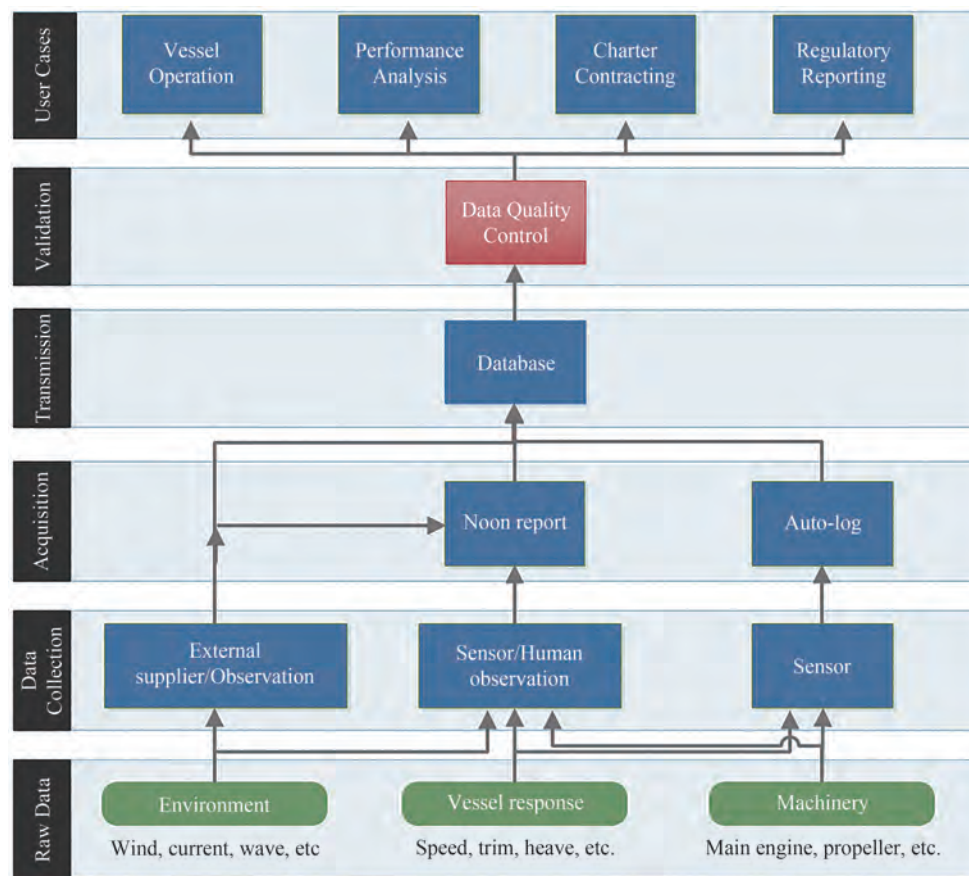
In vessel operations, vessel performance can be measured by processing data collected on board such as the noon report and autolog data. Three types of raw data are relevant for the domain of vessel performance: environmental data, vessel response data and machinery data, as seen in Figure 1. Environmental data consist of weather information such as wind, wave, current, temperature variation, and ice, which are normally recorded by crew on board. These data are, in some systems, also be provided by meteorological services.

The vessel response data include the motions due to external environment (e.g., heave and pitch.), and the vessel status (e.g., position, course, speed, draft, trim and speed). GPS and log systems, on-board sensors/devices, or even manual observation are commonly used to record vessel response data. The third main type of raw data source is machinery, which covers all equipment on board such as main engine, auxiliary engine(s), boiler, propeller, air-conditioning, hotel functions, cargo heating/cooling systems, fuel/water pumps and cargo handling equipment. The corresponding data from machinery include information such as engine speed, consumption of fuel and lubrication oil, shaft power, torque, temperatures and pressures.

Based on the data supply chain in vessel operations, a concrete categorical framework is proposed in this paper. The proposed framework consists of 10 general categories, as seen in Table 1. Examples for explanations are given in the following. For the “Main Engine” category, it consists of relevant data items such as the main engine running hours, the revolutions per minute (RPM), main engine power and the main engine fuel consumption. The category “Performance Indicators” includes vessel performance metrics such as SFOC and added resistance of vessels.

3.2 Data dimension metrics

In this section, the data are further assessed along a set of selected dimension metrics. The metrics are presented in Table 2, including the dimension names, concrete descriptions of dimensions and types. The two basic types of data quality metrics are intrinsic quality and contextual quality, respectively, which are originally from the concept of multi-dimensional nature of data quality (Wang and Strong, 1996; Lee et al., 2002). Intrinsic quality denotes the dimensions that are objective and native to the data, while contextual quality refers to the dimensions that are dependent on the context of the task at hand.

**Figure 1** Data supply chain of vessel operations**Table 1** Data categories related to vessel operations

No.	Category	Example data items
1	Report General	Report duration, reporting time
2	Operational aspect	ETA (estimated time of arrival), voyage instruction
3	Navigation	Ship position, speed, heading, sailed & remain distance
4	Environment	Air pressure, temperature, wind, wave, water depth
5	Loading Condition	draft fore & aft
6	Main Engine	Main engine running hours, RPM, power, fuel consumption
7	Production & Consumption	Fuel consumption of auxiliary engine(s), cargo plant & boiler
8	Performance Indicator	Added resistance, SFOC
9	Harbour Module	Fuel consumption of engines, cargo plant, cargo plant & boiler
10	Bunkering	Bunkering records and prices, stock tracking

Table 2 The six data quality dimension metrics and their definitions

No.	Dimension name	Description	Type
1	Completeness	Is every record field filled with a value	Intrinsic
2	Validity	Are all data values within the valid range with right format ?	Contextual
3	Timeliness	Are data reported at the time as it should ?	Intrinsic
4	Consistency	Are all data records stable over time ?	Contextual
5	Accuracy	Do data reflect the real-world objects or a trusted source ?	Contextual
6	Integrity	Are business rules on field and table relationships met ?	Contextual

Six different metrics are selected in our method for vessel data based on literature, as shown in Table 2. Among them, completeness is defined as to what degree the data records are complete in content without missing data. Validity metric is defined as the degree to which data are conforming to the related rules, including value formatting, data type and realistic range. These rules are often determined by data-owner or users' knowledge on a vessel's navigational, technical and operational characteristics. For instance, the speed of a cargo vessel is measured in knots and can not exceed the value of 30. Once relevant rules are set, validity dimension can be measured for received datasets. Timeliness is defined as whether data are reported at the desired time, according to the industry practice (e.g., noon reports should be sent to shore at noon every day). The fourth-dimension, consistency, can be considered as the presenting of large variety according to the comparison with some equivalent cases that share similar conditions. For instance, if a vessel suddenly consumes twice as much fuel as ordinary during sailing, then this data record is not consistent with the others. The accuracy metric is defined as to what degree the data are equivalent to their corresponding "real" values, which can be accessed via comparison with external values that are known to be correct. The last dimension metric, integrity, is defined as to what degree data of one variable conform to relevant relationship rules with other data variables. For instance, the torque meter on board is used to measure the main engine power value. If the registered values do not conform with, for example, registered values from RPMs, the integrity metric is violated.

3.3 DQ scoring

To quantify the quality of each data item, scoring methods should be finally used. Thus, a scoring method with three levels, namely *Good*, *Warning* and *Alarm* is proposed. Each level is further assigned with a numerical value, which is 0, 10, and 20, respectively, so a total score can be calculated for every data source. As shown in Table 3 for an example, the specific level of data quality (severity) is described. For a data item within validity metric, the level is considered as *Good* when the calculated values are within a valid range set by vessel operators. The scoring

method can be then applied on raw datasets of vessels after properly categorized with different metrics in Sections 3.1 and 3.2.

3.4 System of validation rules and root causes

Using the data categories and the data dimensional metrics from the previous sections, different validation rules are formed, which can be directly applied on the raw dataset for further analysis. Therefore, a concrete DQ validation rule system for vessel operations and performance is produced, which may consist of more than 100 validation rules (the number of rules depend on vessel and company requirements). Table 4 shows an example of the DQ rules for the vessel performance in terms of added resistance and data internal dependencies. A specific ID is assigned to each rule. For the sake of clarity, other validation rules are not presented here.

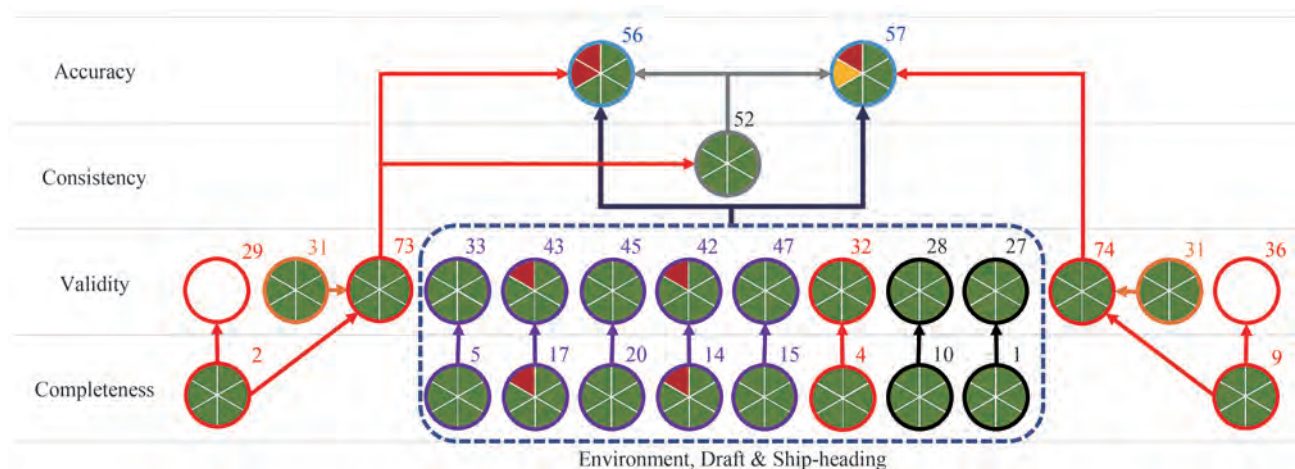
Dependency relations among rules can be also revealed based on their semiotic hierarchy, as indicated in the last column of Table 4. With the dependencies of validation rules, it is possible to drill down from high level performance metrics to identify lower level root causes of poor data quality. As shown in Figure 2, it is an example of the rule dependency for diagnosis when accounting for the high-level added resistance. In this figure, each circle pie with its ID denotes a specific rule, and the six sub-pieces of each pie denote the DQ severity for noon report in 6 consecutive days. The three different colors indicate the data quality scoring levels, namely red for alarm, yellow for warning and green for good. The added resistance rules (56 and 57), belonging to the category of performance indicator, are the high-level monitoring rules for vessel performance. For instance, when there is a DQ alarm in rule 56 (the top-left sub-piece), the root causes could be traced as the missing of wind speed (14) and/or the missing of wave height (17). Meanwhile, the corresponding validity rules are also alarmed (rules 42 and 43). Another simplified example about the rule dependency is illustrated in Figure 3 for SFOC diagnosis, where the range of SFOC (71) is determined by the range of power (65), the range of fuel consumption (52) and engine consistency (94). These rules are further determined by other low-level validation

Table 3 The proposed DQ (severity) levels

No.	Dimension	DQ(severity) level		
		Good (0)	Warning (10)	Alarm (20)
1	Completeness	Data are not null	Not relevant	Null data
2	Validity	Value within range	Slightly out of range	Significantly out of range
3	Timeliness	Report appears on time	Not relevant	Report delayed
4	Consistency	Volatile as previous	Slightly more volatile	Significantly more volatile
5	Accuracy	Same values	Slightly different values	Significantly different values
6	Integrity	Different data fit	Slightly deviate	Significantly deviate

Table 4 An example of DQ rules in vessel operations

Category	Dimension	Rule	Rule description	Rule dependency
(ID)	(ID)	(ID)		
Report General (1)	2	31	Range check field report duration	-
	1	2	Missing field observed miles	-
	1	4	Missing field ship heading	-
	1	9	Missing field logged miles	-
Navigation (3)	2	29	Range check field observed miles	2
	2	32	Range check field ship heading	4
	2	36	Range check field logged miles	9
	2	73	Range check field observed speed	2,31
	2	74	Range check field logged speed	9,31
	1	5	Missing field water temperature	-
	1	14	Missing field true wind speed	-
	1	15	Missing field true wind direction	-
	1	17	Missing field wave height	-
	1	20	Missing field wave direction	-
Environment (4)	2	20	Missing field wave direction	-
	2	33	Range check water temperature	5
	2	42	Range check true wind speed	14
	2	47	Range check true wind direction	15
	2	43	Range check wave height	17
	2	45	Range check wave direction	20
Loading condition (5)	1	1	Missing field draft-fore	-
	2	10	Missing field draft-aft	-
	2	27	Range check draft-fore	1
	2	28	Range check draft-aft	10
Main engine (6)	4	52	Consistency of main engine fuel consumption	73
	5	56	Added resistance based on observed speed	52,73,27,28,32,33,42,43,45,47
Performance indicator (8)	5	57	Added resistance based on logged speed	52,74,27,28,32,33,42,43,45,47

**Figure 2** An example of relevant rules for the vessel performance in terms of added resistance

rules, such as engine run hours (63 and 25), engine RPM (64 and 18) and the missing of engine power or not (26). Such diagnostic results will be directly shared with crew on board, which will facilitate them to improve their data recordings.

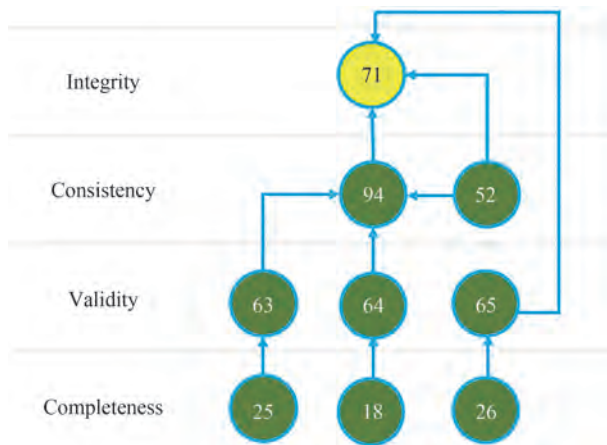


Figure 3 An example of relevant rules for the vessel performance in terms of SFOC

4 Case study and industry application

In this section, a case study is presented based on a real dataset from vessels in the Danish tramp shipping company TORM (TORM, 1889). The proposed method is first integrated into a cloud-based vessel performance analysis engine used by the company. The original noon report dataset will be first described, and then the implementation of the DQ method will be introduced, followed by comparison results.

4.1 The noon report dataset

The datasets consists of data between 2013 and 2017 from the fleet of TORM. Note that, due to the confidential reason, the original dataset will not be shared. Figure 4 shows an overview of the variation of fleet size and the number of noon reports each year. The number of noon

reports increase with the increase of fleet size. When old vessels are demolished, e.g., between 2016 and 2017, the corresponding noon report size is decreased. Only six vessels (Vessel A, Vessel B, ..., Vessel F) are selected for this case study and the vessel performance in terms of added resistance is used for DQ visualization and comparison for brevity.

4.2 Implementation of DQ assessment method

The implementation of the DQ method was realized via an industry application. First, the assessment framework with all the validation rules for vessel performance is built. In this step, experienced domain knowledge is needed so as to have proper range checking values for data validation rules. Second, the assessment framework is integrated into the existing vessel performance analysis engine and then connects to its cloud database, which enables the analysis of DQ of all historical noon reports. Lastly, an automatic assessment procedure is built to handle new incoming datasets. Whenever a noon report arrived, the preliminary diagnosis of DQ will be conducted, displaying the quality levels by colors in time-series.

4.3 Comparison results of DQ

Figure 5 illustrates the DQ severity (Good, Warning and Alarm) levels for the noon reports from the selected TORM's vessels between April 10 to April 16 in 2017. Only the data categorized by validation rules related to added resistance are presented. The different color bars denote different severity levels. As observed, the overall data quality in this week is quite good, indicated by the majority of green color. Specifically, there are not alarms or warnings for data in terms of loading condition (category 5), whereas warnings and alarms start to appear for the performance indicator (category 8). There are a few alarms for data in categories 3 and 4. Accounting for both data dimensions and categories, it is found that the noon report data from Vessel A has a perfect quality, while the data quality is getting worse for Vessels C, D, E and F. The root causes of low data quality can be reflected from the diagram. For instance, due to the violations of rules 14, 17, 42 and 43, the

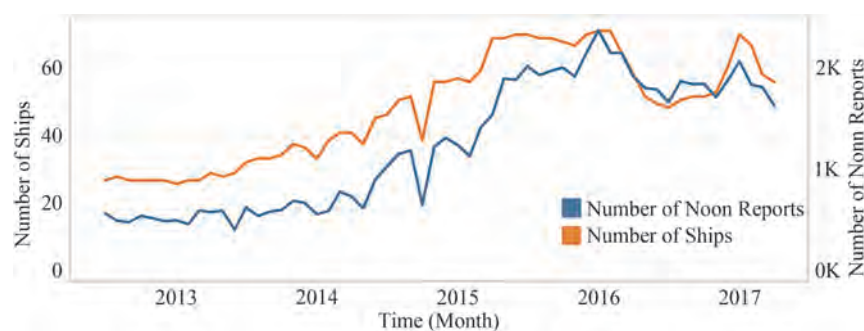


Figure 4 Overview of fleet size and noon reports

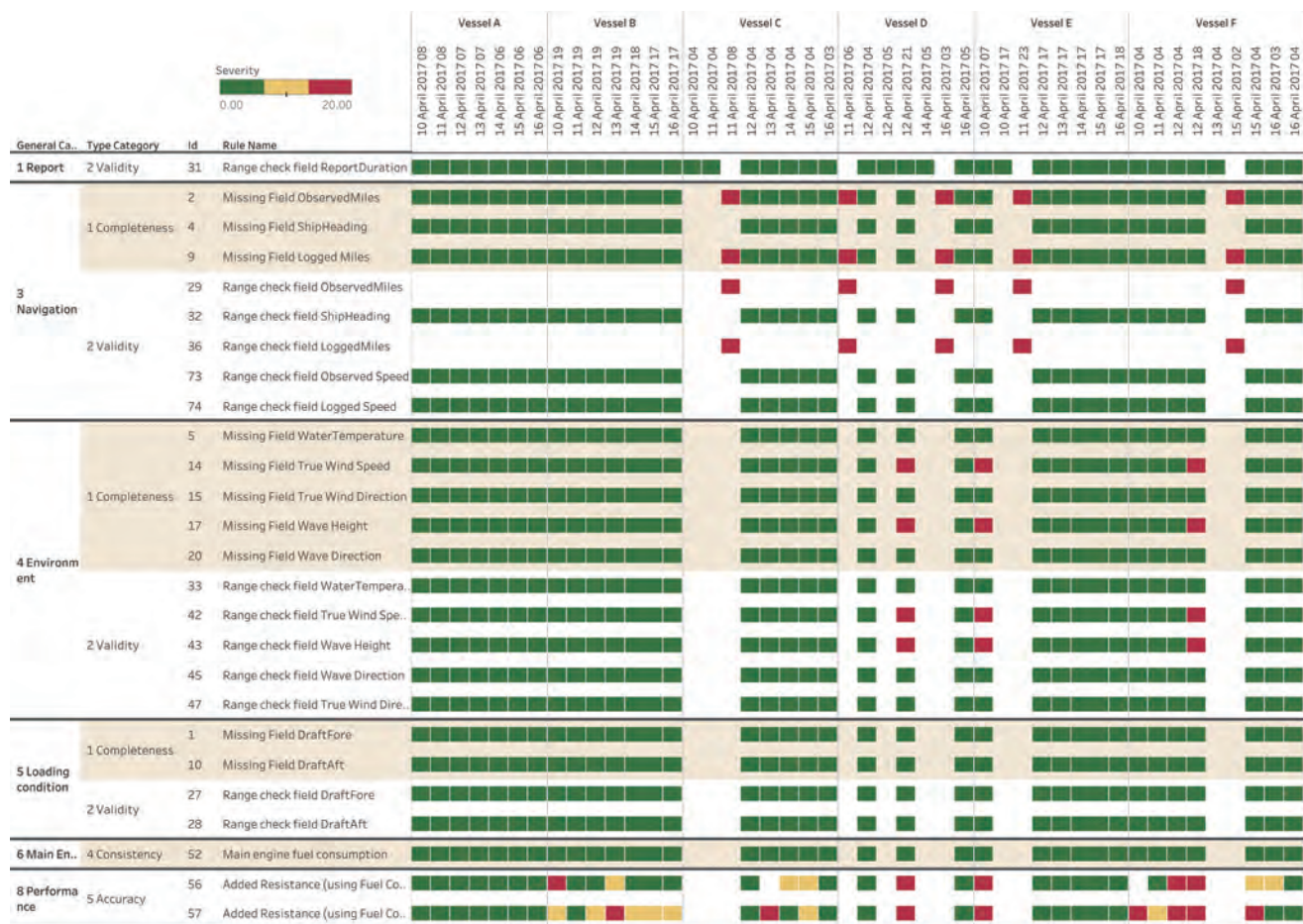


Figure 5 DQ (severity) levels for the noon reports from the selected TORM's vessels between April 10 to April 16 in 2017

performance rules (56 or 57) are displayed with a warning or an alarm. The method can not only instantly signal potential validity issues for vessel performance, but also points towards potential root causes of detected problems. The method thereby enables crew on board and onshore staff to work with continuous improvement of source data and KPI data quality for any (shipping) companies or operators.

Figure 6 illustrates the number of alarms triggered for all 6 vessels over 15 months. Only the data categorized by validation rules related to added resistance are presented. A monthly interval is adopted in this diagram and only DQ breaches with “alarm” is included. It is observed that there is 414 alarms for rule 57 in Feb 2016 (specific number is measured by the bar height), at the time when the DQ assessment method starts to apply. Then, there is a constant improvement of the overall data quality. For instance, the percentage of data severity in terms of vessel performance (category 8, calculated by overall counts of all DQ levels) drops from 33.4% to 25%, and again drops from 9.0% to 5.8% over one year time for all vessels in general. The DQ assessment thus makes it visible that the DQ of the 6 vessels of the shipping company has been improved signifi-

cantly from 2016 to 2017. A major reason is the implementation and integration of the DQ assessment method into the vessel performance monitoring platform. It can therefore be concluded that this method is an effective instrument to drive ongoing improvements of quality of source data. Meanwhile, joint efforts from crew and personnel are also needed for such an improvement. It should be mentioned that, due to project closure, we do not have further access to the database any more. Hence, only data before 2017 are presented in this paper.

5 Discussion

DQ is of great importance in the ongoing digitalization process of shipping industry. Poor DQ costs extra time and energy, which largely compromises proper decision-making. With the exponential growth of data nowadays in shipping companies, the need of advanced DQ monitoring methods, such as the one proposed in this paper, is becoming urgent. The structured DQ assessment in this paper provides a mechanism for continuous monitoring and improvement, being able to capture the data value chain

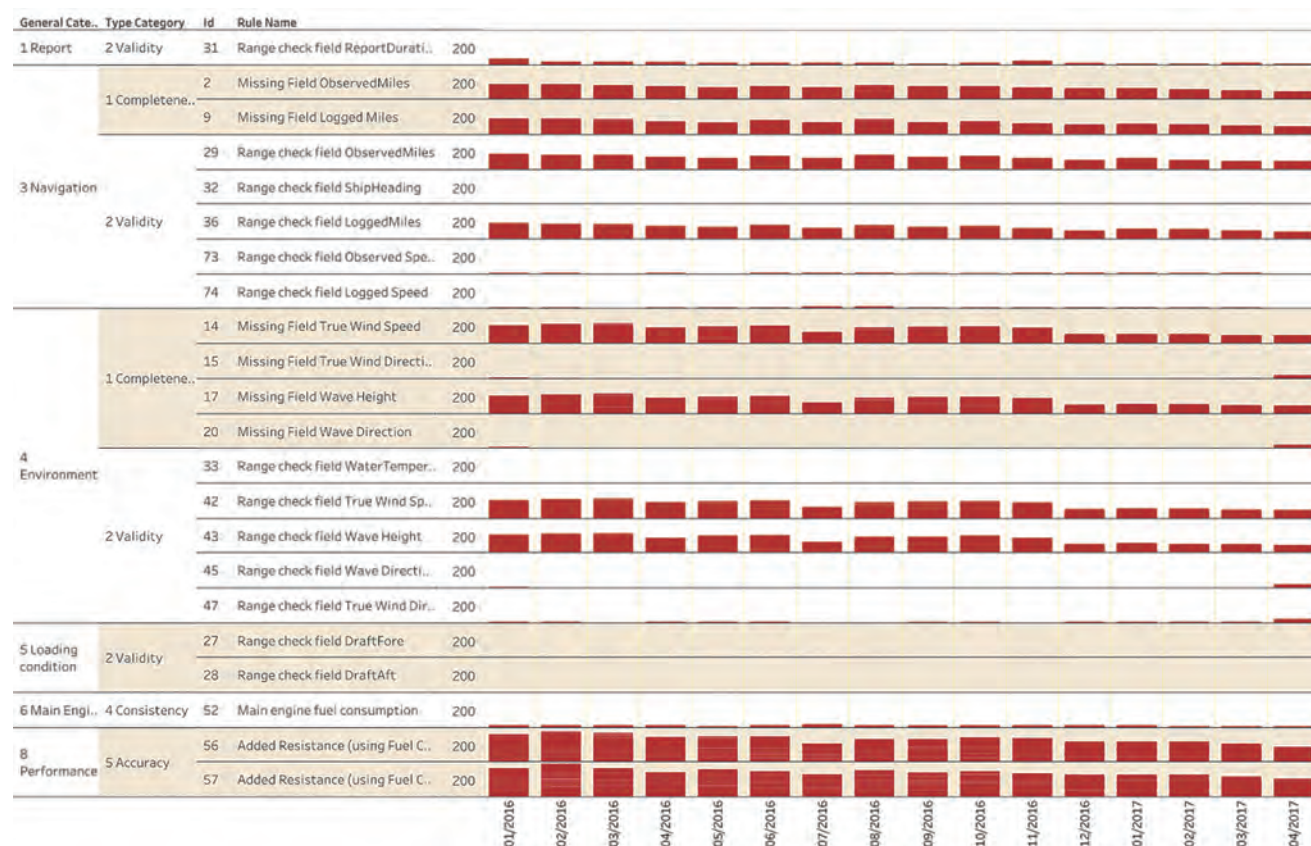


Figure 6 DQ variation after the assessment system applied in one year

deficiencies to some extent in real time.

5.1 Limitations of the proposed method

In order to make the proposed DQ assessment method work, as we have described in the previous section, the shipping company made a thorough effort to deploy a practical system and keep it running in reality accounting for proper thresholds for validation rules. There is a risk that the method provides inaccurate signals on data quality if there is no mastering of vessel knowledge and operation, which makes technical staff and crew take either wrong or no actions. Therefore, one limitation of this method is that sufficient domain knowledge is always needed for a proper implementation.

In the proposed DQ method, each data item has been scored by three different levels. However, the boundaries of the defined levels are blurry due to changing environment and operation states. The defined boundary values for the validity check rules are somewhat overlapping depending on different vessels or different shipping companies. As seen in Figure 7 for the validity ranges, the boundaries of each DQ level are overlapping due to the differences among physical extremes, realistic extremes and typical extremes. There is a small possibility that the identified alarms (Errors in Figure 7) could be still within the warning

domain (Inaccuracy in Figure 7) due to overlapping. Such limitation can't be avoided. Normally, the setting of accurate boundary or threshold values should be done by industry professionals. Based on such reasons, a dedicated ongoing effort is required for a robust vessel performance monitoring system to implement the proposed method, which might be a barrier for shipping industry.

5.2 Remaining challenges for DQ assessment in vessel operations

Challenges in further developing and refining of the DQ assessment methods exist. The current method was applied on the noon report datasets from vessel operation. With the digitalization of shipping industry, autolog datasets in high-frequency from sensors on board or datasets from AIS become more and more popular. The direct application of this method on such datasets may be challenging. Further development of the current framework is needed, for instance, the update of DQ dimensions and corresponding validation rules. Besides, how to properly implement advanced mathematical models, such as statistical process control (SPC) methods or other control chart methods, is another challenge that we have to figure out in the following research. The root causes of the identified data problems are crucial. How to smartly diagnose root causes

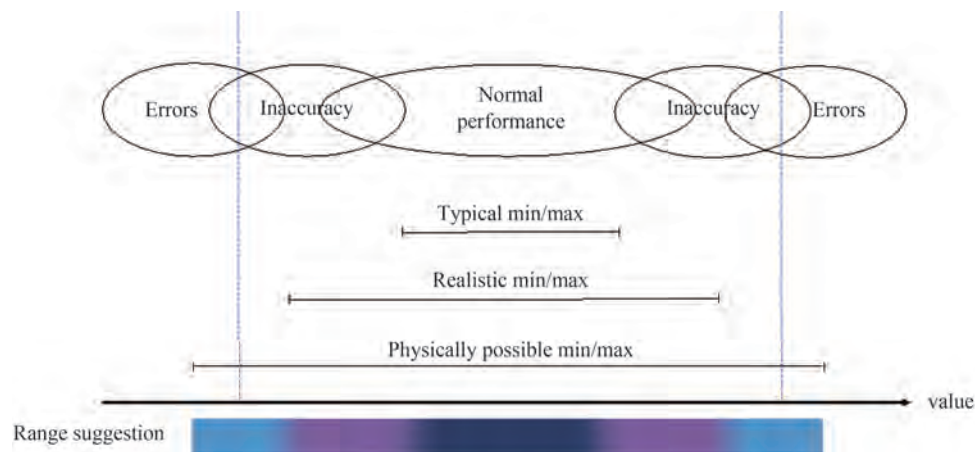


Figure 7 The overlapping of validity ranges

based on machine learning algorithms and AI is a big challenge as well. Last but not the least, vessel performance by nature is a very complex phenomenon due to different factors from mechanical responses, environmental effects and operational uncertainties. Therefore, sufficient domain knowledge is needed to have a better assessment. A reliable vessel performance analysis requires not only knowledge from naval architecture but also rich experience from sailing practice. How to properly define the boundaries of the data quality levels, as we have mentioned above, needs extra concerns.

6 Conclusions

In this paper, a well-structured DQ assessment methodology has been proposed. The data quality of the raw dataset from vessel operations can be properly analysed. In this method, 10 different data categories and 6 different data dimension metrics are proposed for the dividing of raw data. More than 100 validation rules are produced within the framework, which can be directly used to assess the DQ of raw datasets via scoring. Three quality levels, namely good, warning, and critical alarm are adopted.

The proposed DQ method has been successfully applied on a real noon report dataset from the Danish shipping company TORM. The DQ of the dataset is assessed and visualized. The comparison results are analysed and results indicate that this method is effective to monitor and thus improve the data quality of noon reports from shipping operations. Root causes could be identified once the internal dependence among validation rules have been clarified, which facilitate actions to improve operations in practice.

This research will help shipping industry reap full benefits by the improvement of DQ in digital ship operations in the near future. Further research is needed to be done in order to address new challenges. For instance, advanced diagnostics should be investigated on the monitoring of unstable periods. Smart diagnostics of the root causes of abnormal

alarms should be studied based on new technology such as machine learning algorithms.

Acknowledgement The work of this paper has been conducted as part of two projects. The first one is the Danish Blue Innoship project which is partly funded by the Innovation Fund Denmark under File No. 155-2014-10, the Danish Maritime Fund and the Orient's Fund. The second one belongs to ShippingLab project "Digital vessel operation". The funding of the project is provided by the Innovation Fund Denmark (IFD) under File No. 8090-00063B, the Danish Maritime Fund, the Lauritzen Fonden and the Orient's Fond. Gratitude is also expressed to the shipping company TORM for the test of this DQ method and to VPS for the application of the DQ assessment method in its VESPER system.

Competing interest The authors have no competing interests to declare that are relevant to the content of this article.

References

- Ahn K, Rakha H, Hill D (2008) Data quality white paper. Technical Report. United States. Federal Highway Administration. Office of Operations
- Alkhatabi M, Neagu D, Cullen A (2011) Assessing information quality of e-learning systems: a web mining approach. *Computers in Human Behavior* 27: 862-873. <https://doi.org/10.1016/j.chb.2010.11.011>
- Bates MJ (2019) Understanding information retrieval systems: management, types, and standards. Auerbach Publications
- Blake R, Mangiameli P (2011) The effects and interactions of data quality and problem complexity on classification. *Journal of Data and Information Quality (JDIQ)* 2: 1-28. <https://doi.org/10.1145/1891879.1891881>
- Cai J, Chen G, L/tzen M, Rytter NGM (2021) A practical ais-based route library for voyage planning at the pre-fixtue stage. *Ocean Engineering* 236: 109478. <https://doi.org/10.1016/j.oceaneng.2021.109478>
- Cai J, Jiang, X, Yang Y, Lodewijks G, Wang M (2022) Data-driven methods to predict the burst strength of corroded line pipelines subjected to internal pressure. *Journal of Marine Science and Application* 21: 115-132. <https://doi.org/10.1007/s11804-022-00263-0>
- Caro A, Calero C, Caballero I, Piattini M (2008) A proposal for a set of attributes relevant for web portal data quality. *Software Quality*

- Journal 16: 513-542
- Chengalur-Smith IN, Ballou DP, Pazer HL (1999) The impact of data quality information on decision making: an exploratory analysis. *IEEE Transactions on Knowledge and Data Engineering* 11: 853-864. <https://doi.org/10.1109/69.824597>
- Coen-Portisini A, Sicari S (2012) Improving data quality using a cross layer protocol in wireless sensor networks. *Computer Networks* 56: 3655-3665. <https://doi.org/10.1016/j.comnet.2012.08.001>
- De Mauro A, Greco M, Grimaldi M (2015) What is big data? A consensual definition and a review of key research topics, in: AIP Conference Proceedings, American Institute of Physics, 97-104. <https://doi.org/10.1063/1.4907823>
- Dey D, Kumar S (2010) Reassessing data quality for information products. *Management science* 56: 2316-2322. <https://doi.org/10.1287/mnsc.1100.1261>
- Eisele WL, Rilett LR (2002) Travel-time estimates obtained from intelligent transportation systems and instrumented test vehicles: Statistical comparison. *Transportation research record* 1804: 8-16. <https://doi.org/10.3141/1804-02>
- Falge C, Otto B, Österle H (2012) Data quality requirements of collaborative business processes, in: 2012 IEEE 45th Hawaii International Conference on System Sciences, 4316-4325. <https://doi.org/10.1109/HICSS.2012.8>
- FORCE Technology (2021) Onboard decision support system. URL: <https://forcetechnology.com/en/services/onboard-decision-support-system>
- Hazen BT, Boone CA, Ezell JD, Jones-Farmer LA (2014) Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics* 154: 72-80. <https://doi.org/10.1016/j.ijpe.2014.04.018>
- Hermann M, Pentek T, Otto B (2016) Design principles for industrie 4.0 scenarios, in: 2016 49th Hawaii international conference on system sciences (HICSS), IEEE. pp. 3928-3937. <https://doi.org/10.1109/HICSS.2016.488>
- Jones-Farmer LA, Woodall WH, Steiner SH, Champ CW (2014) An overview of phase i analysis for process improvement and monitoring. *Journal of Quality Technology* 46: 265-280. <https://doi.org/10.1080/00224065.2014.11917969>
- Karagiannidis P, Themelis N (2021) Data-driven modelling of ship propulsion and the effect of data pre-processing on the prediction of ship fuel consumption and speed loss. *Ocean Engineering* 222, 108616. <https://doi.org/10.1016/j.oceaneng.2021.108616>
- Knight Sa, Burn J (2005) Developing a framework for assessing information quality on the world wide web. *Informing Science* 8
- KONGSBERG (2021) KONGSBERG Vessel Performance. URL: <https://www.kongsberg.com/digital/kognifaecosystem/kognifa-marketplace/maritime/vessel-performance/>
- Lee YW, Strong DM, Kahn BK, Wang RY (2002) Aimq: a methodology for information quality assessment. *Information & management* 40: 133-146. [https://doi.org/10.1016/S0378-7206\(02\)00043-5](https://doi.org/10.1016/S0378-7206(02)00043-5)
- Liao CF, Davis GA (2012) Traffic data quality verification and sensor calibration for weigh-in-motion (wim) systems
- Peltier JW, Zahay D, Lehmann DR (2013) Organizational learning and crm success: a model for linking organizational practices, customer data quality, and performance. *Journal of interactive marketing* 27: 1-13. <https://doi.org/10.1016/j.intmar.2012.05.001>
- Perera LP, Mo B (2020) Ship performance and navigation information under high-dimensional digital models. *Journal of Marine Science and Technology* 25: 81-92
- Pipino LL, Lee YW, Wang RY (2002) Data quality assessment. *Communications of the ACM* 45: 211-218. <https://doi.org/10.1145/505248.506010>
- Redman TC (1998) The impact of poor data quality on the typical enterprise. *Communications of the ACM* 41: 79-82. <https://doi.org/10.1145/269012.269025>
- Richardson JK, Smith BL (2012) Development of hypothesis test for travel time data quality. *Transportation research record* 2308: 103-109. <https://doi.org/10.3141/2308-11>
- Røseth ØJ (2016) Integrating iec and iso information models into the s-100 common maritime data structure
- Shankaranarayan G, Ziad M, Wang RY (2003) Managing data quality in dynamic decision environments: An information product approach. *Journal of Database Management (JDM)* 14: 14-32. <https://doi.org/10.4018/jdm.2003100102>
- Soner O, Akyuz E, Celik M (2018) Use of tree based methods in ship performance monitoring under operating conditions. *Ocean Engineering* 166: 302-310. <https://doi.org/10.1016/j.oceaneng.2018.07.061>
- Soner O, Akyuz E, Celik M (2019) Statistical modelling of ship operational performance monitoring problem. *Journal of Marine Science and Technology* 24: 543-552. <https://doi.org/10.1007/s00773-018-0574-y>
- Tejay G, Dhillon G, Chin AG (2004) Data quality dimensions for information systems security: A theoretical exposition, in: Working Conference on Integrity and Internal Control in Information Systems, Springer. pp. 21-39
- TORM (1889) TORM SHIPPING. URL: <https://torm.com/>
- Turner S (2004) Defining and measuring traffic data quality: White paper on recommended approaches. *Transportation research record* 1870: 62-69. <https://doi.org/10.3141/1870-08>
- US Department of Transportation (2021) Bureau of Transportation Statistics. URL: <http://ntl.bts.gov/lib/jpodocs/reptste/14058files/chap3.htm>
- VPS (2021) Vessel Performance Solutions. URL: <https://www.vpsolutions.dk/>
- Wang RY, Strong DM (1996) Beyond accuracy: What data quality means to data consumers. *Journal of management information systems* 12: 5-33. <https://doi.org/10.1080/07421222.1996.11518099>
- Wang RY, Ziad M, Lee YW (2006) Data quality. volume 23. Springer Science & Business Media
- Yan R, Wang S, Du Y (2020) Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship. *Transportation Research Part E: Logistics and Transportation Review* 138: 101930. <https://doi.org/10.1016/j.tre.2020.101930>
- Yerva SR, Miklós Z, Aberer K (2012) Quality-aware similarity assessment for entity matching in web data. *Information Systems* 37: 336-351. <https://doi.org/10.1016/j.is.2011.09.007>