




Big Data Quality Models: A Systematic Mapping Study

Osbel Montero¹(✉) , Yania Crespo¹, and Mario Piatini²

¹ School of Computer Engineering, University of Valladolid, Valladolid, Spain

² Information Technologies and Systems Institute (ITSI), University of Castilla-La Mancha, 13071 Ciudad Real, Spain

Abstract. In the last decade, we have been witnesses of the considerable increment of projects based on big data applications and the evident growing interest in implementing these kind of systems. It has become a great challenge to assure the expected quality in Big Data contexts. In this paper, a Systematic Mapping Study (SMS) is conducted to reveal what quality models have been analyzed and proposed in the context of Big Data in the last decade, and which quality dimensions support those quality models. The results are exposed and analyzed for further research.

Keywords: Big data · Quality dimensions · Data quality models

1 Introduction

Big Data involves the management of large datasets that due to its size and structure exceed the capabilities of traditional programming tools for collecting, storing, and processing data in a reasonable time. In data generation the main big data sources are users, applications, services, systems, sensors, and technological devices, among others [18]. They all contribute to Big Data in the form of documents, images, videos, software, files with a multi-diverse format style. The huge volume and heterogeneity present in Big Data applications contribute to the complexity of any engineering process involved.

Currently, different kinds of Big Data applications can be identified such as Recommendations, Feature Prediction and Pattern Recognition [72]. The real-life domains of big data applications include smart cities, smart carts, healthcare systems, financial, business intelligence, environmental control and so on.

The importance and relevance that Big Data is acquiring these days and the promising future we can expect on this knowledge area has been discussed widely. The lack of research on the adequate test modelling and coverage analysis for big data application systems, and the clear practitioners' demand for having established a well-defined test coverage criterion is an important issue [62]. In addition, how to effectively ensure the quality of big data applications is still a hot research issue [72].

For defining a Big Data quality evaluation system, is necessary to know what quality models have been investigated and proposed. The reminder of this study is organized as follows: Sect. 2 exposes related papers where similar topics were investigated; Sect. 3 presents the research methodology used to conduct this study. In Sect. 4 we reveals answers to our research questions, and finally in Sect. 5, we discuss considerations based on our analysis and threats to validity our mapping study.

2 Related Work

In [52] a review on key non-functional requirements in the domain of Big Data systems is presented, finding more than 40 different quality attributes related to these systems and concluding that non-functional requirements play a vital role at software architecture in Big Data systems.

[71] presents another review that evaluates the state of the art of proposed Quality of Services (QoS) approaches on the Internet of Things (IoT) where one of the research questions mentions the quality factors that quality approaches consider when measuring performance.

The research that comes closest to the actual work is [53]. In this paper a SMS is presented involving concepts like “quality models”, “quality dimensions” and “machine learning”. A selection of 10 papers is done where some quality models are reviewed and a total of 16 quality attributes are presented that have some effects on machine learning systems. Finally, the review is evaluated by conducting a set of Interviews with other experts.

3 Methodology

By applying a SMS, we attempts to identify the quality models that have been proposed to evaluate Big Data applications in the last decade by making a distinction between the different types of quality models applied in the context of Big Data applications.

3.1 Definition of the Research Questions

- RQ1: What quality models related to Big Data have been proposed in the last 10 years?
- RQ2: For which Big Data context have these quality models been proposed?
- RQ3: What Big Data quality characteristics are proposed as part of these quality models?
- RQ4: Have Big Data quality models been proposed to be applied to any type of Big Data application?

3.2 Inclusion/Exclusion Criteria

In Table 1, pre-defined criteria for inclusion/exclusion of the literature are presented. Papers included are published between 1st January 2010 and 31st December 2020 whose main contribution is the presentation of new or adapted quality models for Big Data

applications or even the discussion of existing ones. A small database was prepared which contained all documents and a special column was defined to determine the expected level of correlation that the document might have with the topic being investigated.

Papers excluded were those duplicated in different databases or published in Journals and Conferences with the same topic. Papers which could not answer any of the research questions proposed were equally excluded. In addition, those that could not be accessed, or an additional payment was requested for access the full content, or with non-English redaction were similarly excluded.

Table 1. Inclusion and exclusion criteria

Inclusion criteria
Papers published between 2010 and 2020
Papers whose main topic is the presentation or discussion of a quality model for Big Data applications
Desirable: Papers published in a trustworthy and indexed source
Desirable: Papers with a peer-review
Exclusion criteria
Duplicate papers in different databases or under the same topic in different Conference papers and Journal articles
Papers that do not answer any of the research questions
Papers that require an additional payment to provide access
Papers with a low quality in the defined methodology
Papers with non-English redaction

3.3 Search and Selection Process

The following search string: “Big Data” AND “Quality Model” was defined to obtain papers which correlate these two concepts. After processing the results of these searches, other papers could be analyzed by the applying “snowballing”.

Because the amount of information we are trying to collect, analyze and classify, we will focus on this paper the search on scientific databases like SCOPUS and ACM. Figure 1 resumes the steps conducted at this SMS following the review protocol and finally a filtered excel sheet has been obtained with the primary studies selected. SCOPUS database indexes also publications indexed in other databases such as IEEE and Springer. In such cases, a deuration was executed to eliminate duplicates.

After searching in Scopus and ACM databases additional papers were included using snowballing. From the total of 958 papers obtained, the inclusion/exclusion criteria were applied and finally a resumed of 67 papers were selected as the primary studies.

Reviewing the number of citations in the primary studies, it has been found that five papers stand out from the rest. The most cited with 121 citations, is related to

measuring the quality of Open Government Data using data quality dimensions [65]. With 76 citations, [42] proposes a Quality-in-use-model through their “3As model” which involves Contextual Adequacy, Operational Adequacy and Temporal Adequacy. The third most cited paper has 66 citations and explores the Quality-of-Service (QoS) approaches in the context of Internet of Things (IoT) [71]. In [29] is reviewed the quality of social media data in big data architecture and has 48 citations. Finally, [39] with 40 citations, proposes a framework to evaluate the quality of open data portals on a national level.

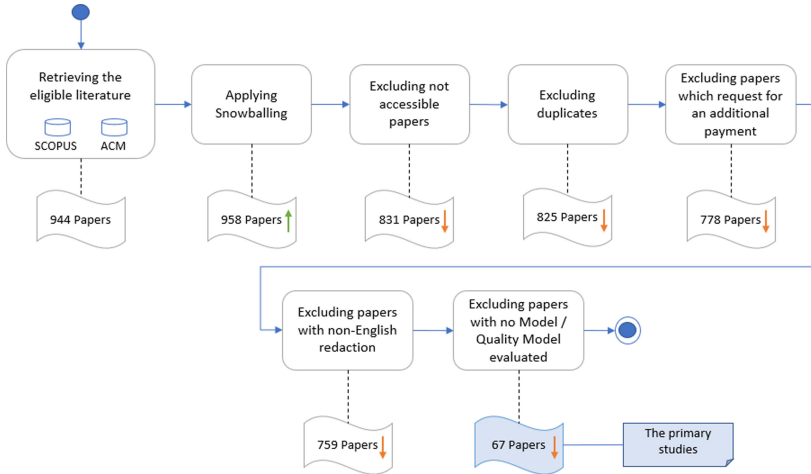


Fig. 1. Facets inside the search and filtering process.

3.4 Quality Assessment

To assess the quality of the chosen literature some parameters were defined as Quality Assessments (QA) such as:

- QA-1: Are the objectives and the scope clearly defined?
- QA-2: Do they proposes/discusses a quality model or related approaches? (if yes, the quality model is applied to a specific Big Data application?)
- QA-3: Do they discuss and present quality dimensions/characteristics for specific purpose?
- QA-4: Do they provide assessment metrics?
- QA-5: Where the results compared to other studies?
- QA-6: Where the results evaluated?
- QA-7: Do they present open themes for further searches?

At this point, the next step was assessing the quality to the selected primary studies which overall results are presented in Table 2. This is a process which complements

the inclusion/exclusion is assigned to answer on each paper the quality assessments described above. These primary studies were scored to determine how well the seven quality items defined were satisfied. The punctuation system used was basically a predefined scale with Y-P-N (Y: Yes, P: Partially, N: No), which was weighted as Y: 1 point, P: 0.5 points, N: 0 points.

Table 2. Quality assessment overall results

Quality assessment	Total score	Compliance ratio
QA-1	66	98,51%
QA-2	67	100,00%
QA-3	56	83,58%
QA-4	22,5	33,58%
QA-5	8,5	12,69%
QA-6	31,5	47,01%
QA-7	52,5	78,36%

4 Results

The overall results of this SMS are presented in the current section. A distribution per document type exposed that the largest number of documents obtained (95,52%) are distributed as conference papers and journal articles.

Regarding the year of publication inside the initial range of 2010 – 2020, a gradual increase can be seen starting from 2014 in the number of studies published on the related topics. The 67% of all selected studies were published in the last three years (2018–2020), the 88% of all selected studies were published in the last five years (2016–2020) which is indicating that the issue of quality models in the context of Big Data is receiving more attention among the researchers, and if this trend continues the theme could become in one of the hottest research topics.

Regarding the publisher, Fig. 2 shows that most papers were published between Springer (26,87%), IEEE (25,37%), ACM (11,94%), and Elsevier (8,96%), most of them were indexed in SCOPUS.

Following, the research questions could be answered thanks to the review conducted, the findings are presented as follows.

RQ1: What quality models related to Big Data have been proposed in the last 10 years?

The study has revealed that 12 different quality model types has been published in the last 10 years, the most commons are those related with measuring Data Quality, Service Quality, Big Data Quality and Quality-In-Use. A complete distribution of these quality

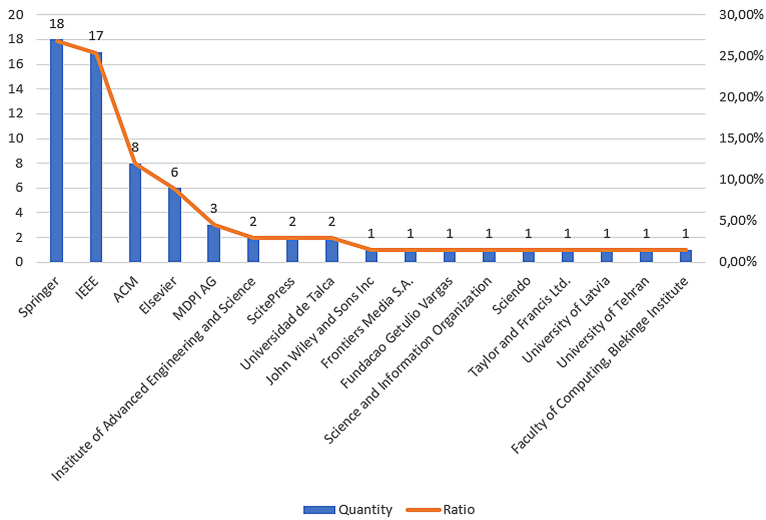


Fig. 2. Paper distribution per publisher

models can be viewed in Fig. 3. It is not a surprise that the largest number of quality models proposed are those related to measuring the quality of the data, representing almost half of all models found.

RQ2: In what Big Data context have these quality models been proposed?

The majority of quality models proposed can be applied to any Big Data project without distinguishing between the different types of Big Data applications. A number of 8 approaches have been identified as possible field of application which can be regarded in Table 3.

There is a differentiation between general Big Data projects and Open Data projects mainly because the dimensions presented for those Open Data are related such as free access, always available, data conciseness, data and source reputation, and objectivity among others, are specially required in Open Data projects. For Big Data Analytics, Decision Making and Machine Learning projects there is no such great differentiation with other Big Data projects, only in the case where non-functional requirements must be measured that are specific to the required purpose.

RQ3: What Big Data quality characteristics are proposed as part of these quality models?

In this case, is necessary to make a differentiation among the quality model types, because the authors have identified different quality dimensions depending upon the focus of the quality model inside Big Data context.

Data Quality Models: Are defined as a set of relevant attributes and relationships between them, which provides a framework for specifying data quality requirements

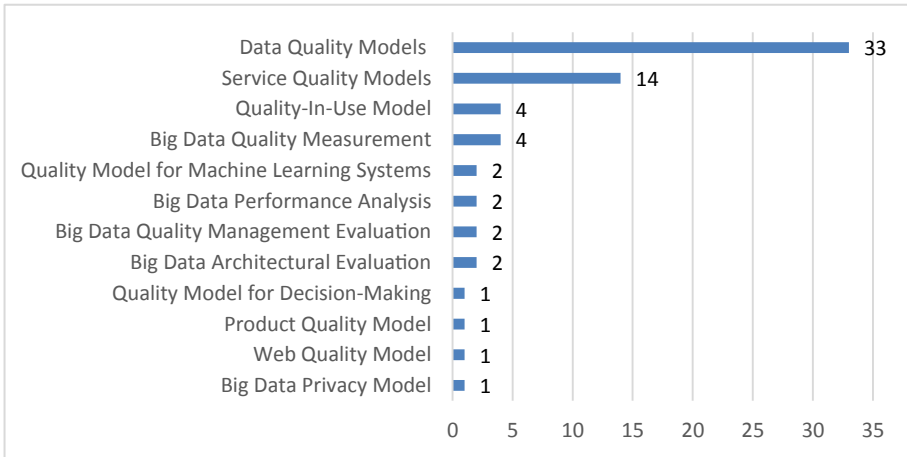


Fig. 3. Paper distribution per quality model type

Table 3. Quality model distribution per Big Data context

Context	Quantity	Ratio
Any Big Data Project	50	74,63%
Cloud Projects	6	8,96%
Social Information Services (Facebook, Twitter, etc.)	2	2,99%
Big Data Analytics	2	2,99%
Open Data Projects	2	2,99%
Decision Making process	2	2,99%
Machine Learning Projects	2	2,99%
Smart Cities Ecosystem Project	1	1,49%

and evaluating data quality. Represents data quality dimensions and the association of such dimensions to data. Good examples of those models are presented in [21, 29, 31, 48, 61, 65]. Figure 4 shows the categories that can be used to group the different data quality dimensions presented in the quality models.

Quality dimensions are presented in 28 from 33 papers related with data quality models. The most common dimensions for general data quality are:

- **Completeness:** characterizes the degree to which data have values for all attributes and instances required in a specific context of use. Also, data completeness is independent of other attributes (data may be complete but inaccurate).
- **Accuracy:** characterizes the degree to which data attributes represent the true value of the intended attributes in the real world, like a concept or event in a specific context of use.

- Consistency: characterizes the degree to which data attributes are not contradicted and are consistent with other data in a context of use.
- Timeliness: characterizes the latest state of a data attribute and its period of use.

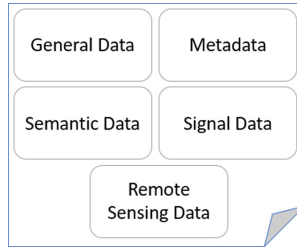


Fig. 4. Categories founded in the SMS that groups the data quality dimensions presented in the quality models.

In addition, for those quality models where the attention was focused on measuring the quality of metadata, in [29] the quality dimensions identified are believability, corroboration, coverage, validity, popularity, relevance, and verifiability. Other four dimensions are included apart from existing ones to Semantic Data [31] which are objectivity, reputation, value added and appropriate amount of data. For Signal Data [35], other dimensions were identified such as availability, noise, relevance, traceability, variance, and uniqueness. Finally, other two dimensions were included for Remote Sensing Data [7] which are resolution and readability.

It should be noted that quality dimensions proposed in each of these quality models refer to quality aspects that need to be verified by them in the specific context of use.

Service Quality Models: Are used to describe the way on how to achieve desired quality in services. This model measures the extent to which the service delivered meets the customer's expectations. Good examples of these models are presented in [8, 30, 32, 36, 41, 64]. The most common quality dimensions collected from those papers are: Reliability, Efficiency, Availability, Portability, Responsiveness, Real-time, Robustness, Scalability, Throughput.

Quality-In-Use Models: Defines the quality characteristics that the datasets that are used for a specific use must present to adapt to that use. In this research two papers were found that present such type of models [11] and [42], other papers discuss about them. These models are focused mainly in two dimensions: Consistency and Adequacy represented in Fig. 5.

It should be noted that, depending upon the quality characteristics that wants to be evaluated and the context of use, a different model should be applied. In those models, the two dimensions analyzed are presented as:

- Consistency: The capability of data and systems of keeping the uniformity of specific characteristics when datasets are transferred across the networks and shared by the various types of consistencies.

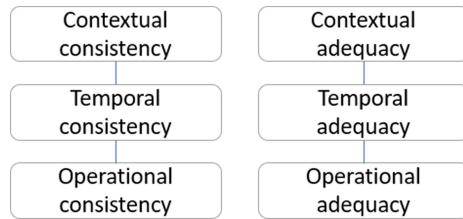


Fig. 5. Quality dimensions presented in Quality-In-Use Models

- Adequacy: The state or ability of being good enough or satisfactory for some requirement, purpose or need.

Big Data Systems Quality Models: There isn't a general definition for this types of models because the enormous number of different kinds. In this research will be defined as quality models applied to the context of Big Data viewed at a high level. Good examples of these models are presented in [28, 37, 50, 56]. The quality dimensions presented in these quality models are specified in Table 4, and can be separated in three groups:

- Dimensions for Big Data value chain
- Dimensions for Non-Functional requirements in Big Data Systems
- Dimensions for measuring Big Data characteristics.

Table 4. Quality dimensions for big data systems quality models

Big data value chain	Big data non-functional requirements (NFRs)	Big data characteristics
Timeliness	Scalability	Volume
Accuracy	High performance computing	Velocity
Completeness	Modularity	Variety
Consistency	Consistency	Veracity
	Security	Valence
	Real-time operations	Value
	Inter-operability	Volatility
	Availability	Vitality
		Vincularity

RQ4: Have Big Data quality models been proposed to be applied to any type of Big Data application or by considering the quality characteristics required in specific types of Big Data applications?

A Big Data application (BDA) processes a large amount of data by means of integrating platforms, tools, and mechanisms for parallel and distributed processing. As was presented in Table 3, the majority of quality models proposed (74,63%) can be applied to any Big Data project and only a few studies were developed specifically for Big Data Analytics such as [66], Decision Making process in [2] and Machine Learning presents in [54] and [55].

This could be means that researchers are not interested on develop a quality model for a specific Big Data application, instead a general quality model is proposed focusing on such topics like assuring the quality of data, the Quality-in-use, the quality of services involved, etc.

5 Threats to Validity of Our Mapping Study

The main threats to validity of our mapping study are:

- Selection of search terms and digital libraries. We search in two digital libraries and to complete our study other libraries should be included such as: IEEE, Springer, and Google scholar. In addition, because Big Data is an industrial issue it is recommended to include gray literature search [24] and achieve a Multivocal Literature Review (MLR).
- Selection of studies. Could be a better solution to apply other exclusion criteria such as the quality of papers. For example, if the results have been validated and compared to other studies.
- Quality model categorization. As a result of the small sample of papers in which quality models are not related to data quality, it is an arduous task to obtain sample quality metrics and quality dimensions for those Big Data quality models. With the amplification of the current study more samples could be obtained to support this task.
- Data categorization. We included all the categories identified in the primary papers, the extraction and categorization process was carried out by the first author, a MSc student with over five years of work experience in software and data engineering. The other two coauthors provided input to resolve ambiguities during the process. In this respect, the extraction and categorization process is partially validated.

6 Conclusions and Future Work

A SMS have been conducted to analyze and visualize the different quality models that have been proposed in the context of Big Data and the quality dimensions presented on each type of quality model. It has been found that different from what would have been thought, there is a considerable number of papers which do not present or partially discuss the quality metrics to evaluate the quality dimensions proposed in the model. Also, in the majority of studies the results of their research was not analyzed and compared with other similar studies.

This research have revealed that proposing, discussing, and evaluating new quality models in the context of Big Data is a topic that is currently receiving more attention

from researchers and with the actual tendency we should expect an increase of papers related with quality models in Big Data context in the coming years.

As first topic for future work we will consider an in-depth review of the analyzed papers where common metrics, quality dimensions and quality models evaluations could be obtained for further analysis on each Big Data quality model type.

In the context of Big Data, most of the proposed quality models are designed for any Big Data application and they are not explicit in evaluating a specific type of Big Data application such as Feature Prediction Systems or Recommenders. Considering their different specificities to assess the expected quality in the final result when using these Big Data applications, we consider this as an open research topic.

References

1. Ali, K., Hamilton, M., Thevathayan, C., Zhang, X.: Big social data as a service: a service composition framework for social information service analysis. In: Jin, H., Wang, Q., Zhang, L.-J. (eds.) ICWS 2018. LNCS, vol. 10966, pp. 487–503. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94289-6_31
2. Alkatheeri, Y., Ameen, A., Isaac, O., Nusari, M., Duraisamy, B., Khalifa, G.S.A.: The effect of big data on the quality of decision-making in Abu Dhabi government organisations. In: Sharma, N., Chakrabarti, A., Balas, V.E. (eds.) Data Management, Analytics and Innovation. AISC, vol. 1016, pp. 231–248. Springer, Singapore (2020). https://doi.org/10.1007/978-981-13-9364-8_18
3. Asif, M.: Are QM models aligned with Industry 4.0? A perspective on current practices. *J. Clean. Prod.* **256**, 1–11 (2020)
4. Baillie, C., Edwards, P., Pignotti, E.: Qual: A provenance-aware quality model. *J. Data Inf. Qual.* **5**, 1–22 (2015)
5. Baldassarre, M.T., Caballero, I., Caivano, D., Garcia, B.R., Piattini, M.: From big data to smart data: a data quality perspective. In: ACM SIGSOFT International Workshop on Ensemble-Based Software Engineering, pp. 19–24 (2018)
6. Barbara Kitchenham, S.C.: Guidelines for performing systematic reviews in software engineering. Durham, UK: EBSE Technical report. EBSE-2007-01 Version 2.3 (2007)
7. Barsi, Á., et al.: Remote sensing data quality model: from data sources to lifecycle phases. *Int. J. Image Data Fusion* **10**, 280–299 (2019)
8. Basso, T., Silva, H., Moraes, R.: On the use of quality models to characterize trustworthiness properties. In: Calinescu, R., Di Giandomenico, F. (eds.) SERENE 2019. LNCS, vol. 11732, pp. 147–155. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30856-8_11
9. Behkamal, B., Kahani, M., Bagheri, E., Jeremic, Z.: A metrics-driven approach for quality assessment of linked open data. *J. Theoret. Appl. Electron. Commer. Res.* **9**, 64–79 (2014)
10. Bhutani, P., Saha, A., Gosain, A.: WSEMQT: a novel approach for quality-based evaluation of web data sources for a data warehouse. *IET Softw.* **14**, 806–815 (2020)
11. Caballero, I., Serrano, M., Piattini, M.: A data quality in use model for big data. In: Indulska, M., Purao, S. (eds.) ER 2014. LNCS, vol. 8823, pp. 65–74. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12256-4_7
12. Cappiello, C., et al.: Improving health monitoring with adaptive data movement in fog computing. *Front. Robot. AI* **7**, 1–17 (2020)
13. Cappiello, C., Samá, W., Vitali, M.: Quality awareness for a successful big data exploitation. In: International Database Engineering & Applications Symposium, pp. 37–44 (2018)
14. Castillo, R.P., et al.: DAQUA-MASS: an ISO 8000-61 based data quality management methodology for sensor data. *Sensors (Switzerland)* **18**, 1–24 (2018)

15. Cedillo, P., Valdez, W., Cárdenas-Delgado, P., Prado-Cabrera, D.: A data as a service metamodel for managing information of healthcare and internet of things applications. In: Rodríguez Morales, G., Fonseca C., E.R., Salgado, J.P., Pérez-Gosende, P., Orellana Cordero, M., Berrezueta, S. (eds.) TICEC 2020. CCIS, vol. 1307, pp. 272–286. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62833-8_21
16. Ciancarini, P., Poggi, F., Russo, D.: Big data quality: a roadmap for open data. In: International Conference on Big Data Computing Service and Applications, BigDataService, pp. 210–215 (2016)
17. Cichy, C., Rass, S.: An overview of data quality frameworks. *IEEE Access* **7**, 24634–24648 (2019)
18. Davoudian, A., Liu, M.: Big data systems: a software engineering perspective. *ACM Comput. Surv.* **53**, 1–39 (2020)
19. Demchenko, Y., Grosso, P., Laat, C.D., Membrey, P.: Addressing big data issues in Scientific data infrastructure. In: International Conference on Collaboration Technologies and Systems, CTS, pp. 48–55 (2013)
20. Fagúndez, S., Fleitas, J., Marotta, A.: Data streams quality evaluation for the generation of alarms in health domain. In: International Workshops on Web Information Systems Engineering, IWCSN, pp. 204–210 (2015)
21. Fernández, S.M., Jedlitschka, A., Guzmán, L., Vollmer, A.M.: A quality model for actionable analytics in rapid software development. In: Euromicro Conference on Software Engineering and Advanced Applications, SEAA, pp. 370–377 (2018)
22. Gao, T., Li, T., Jiang, R., Duan, R., Zhu, R., Yang, M.: A research about trustworthiness metric method of SaaS services based on AHP. In: Sun, X., Pan, Z., Bertino, E. (eds.) ICCCS 2018. LNCS, vol. 11063, pp. 207–218. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00006-6_18
23. Garises, V., Quenum, J.G.: An evaluation of big data architectures. In: 8th International Conference on Data Science, Technology and Applications, DATA, pp. 152–159 (2019)
24. Garousi, V., Felderer, M., Mäntylä, M.V.: Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Inf. Softw. Technol.* **106**, 101–121 (2019). <https://doi.org/10.1016/j.infsof.2018.09.006>. ISSN 0950-5849
25. Ge, M., Lewoniewski, W.: Developing the quality model for collaborative open data. In: International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, KES, pp. 1883–1892 (2020)
26. Gong, X., Yin, C., Li, X.: A grey correlation based supply–demand matching of machine tools with multiple quality factors in cloud manufacturing environment. *J. Ambient. Intell. Humaniz. Comput.* **10**(3), 1025–1038 (2018). <https://doi.org/10.1007/s12652-018-0945-6>
27. Gyulgyulyan, E., Aligon, J., Ravat, F., Astsatryan, H.: Data quality alerting model for big data analytics. In: Welzer, T., et al. (eds.) ADBIS 2019. CCIS, vol. 1064, pp. 489–500. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30278-8_47
28. Helfert, M., Ge, M.: Perspectives of big data quality in smart service ecosystems (quality of design and quality of conformance). *J. Inf. Technol. Manag.* **10**, 72–83 (2018)
29. Immonen, A., Paakkonen, P., Ovaska, E.: Evaluating the quality of social media data in big data architecture. *IEEE Access* **3**, 2028–2043 (2015)
30. Jagli, D., Seema Purohit, N., Chandra, S.: Saasqual: a quality model for evaluating SAAS on the cloud computing environment. In: Aggarwal, V.B., Bhatnagar, V., Mishra, D.K. (eds.) Big Data Analytics. AISC, vol. 654, pp. 429–437. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-6620-7_41
31. Jarwar, M.A., Chong, I.: Web objects based contextual data quality assessment model for semantic data application. *Appl. Sci. (Switzerland)* **10**, 1–33 (2020)

32. Jich-Yan, T., Wen, Y.X., Chien-Hua, W.: A framework for big data analytics on service quality evaluation of online bookstore. In: Deng, D.-J., Pang, A.-C., Lin, C.-C. (eds.) *WiCON 2019*. LNICSSITE, vol. 317, pp. 294–301. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-52988-8_26
33. Jung, Y., Hur, C., Kim, M.: Sustainable situation-aware recommendation services with collective intelligence. *Sustainability (Switzerland)* **10**, 1–11 (2018)
34. Khurana, R., Bawa, R.K.: QoS based cloud service selection paradigms. In: *International Conference on Cloud System and Big Data Engineering, Confluence*, pp. 174–179 (2016)
35. Kirchen, I., Schutz, D., Folmer, J., Vogel-Heuser, B.: Metrics for the evaluation of data quality of signal data in industrial processes. In: *International Conference on Industrial Informatics, INDIN*, pp. 819–826 (2017)
36. Kiruthika, J., Khaddaj, S.: Software quality issues and challenges of internet of things. In: *International Symposium on Distributed Computing and Applications for Business, Engineering and Science, DCABES*, pp. 176–179 (2015)
37. Kläs, M., Putz, W., Lutz, T.: Quality evaluation for big data: a scalable assessment approach and first evaluation results. In: *Joint Conference of the Int'l Workshop on and International Conference on Software Process and Product Measurement Software Measurement*, pp. 115–124 (2017)
38. Liu, Z., Chen, Q., Cai, L.: Application of requirement-oriented data quality evaluation method. In: *International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPDP*, pp. 407–412 (2018)
39. Máchová, R., Lněnička, M.: Evaluating the quality of open data portals on the national level. *J. Theor. Appl. Electron. Commer. Res.* **12**, 21–41 (2017)
40. Manikam, S., Sahibudin, S., Kasinathan, V.: Business intelligence addressing service quality for big data analytics in public sector. *Indonesian J. Electr. Eng. Comput. Sci.* **16**, 491–499 (2019)
41. Mbyone, V., Price, C.S.: A model to evaluate the quality of Wi-Fi performance: case study at UKZN Westville campus. In: *International Conference on Advances in Big Data, Computing and Data Communication Systems, icABCD*, pp. 1–8 (2019)
42. Merino, J., Caballero, I., Rivas, B., Serrano, M., Piattini, M.: A data quality in use model for Big Data. *Futur. Gener. Comput. Syst.* **63**(1), 123–130 (2016)
43. Micic, N., Neagu, D., Campean, F., Zadeh, E.H.: Towards a data quality framework for heterogeneous data. In: *Cyber, Physical and Social Computing, IEEE Smart Data, iThings-GreenCom-CPSCoM-SmartDat*, pp. 155–162 (2018)
44. Musto, J., Dahanayake, A.: Integrating data quality requirements to citizen science application design. In: *International Conference on Management of Digital EcoSystems, MEDES*, pp. 166–173 (2019)
45. Nadal, S., et al.: A software reference architecture for semantic-aware big data systems. *Inf. Softw. Technol.* **90**, 75–92 (2017)
46. Nakamichi, K., Ohashi, K., Aoyama, M., Joeckel, L., Siebert, J., Heidrich, J.: Requirements-driven method to determine quality characteristics and measurements for machine learning software and its evaluation. In: *International Requirements Engineering Conference, RE*, pp. 260–270 (2020)
47. Nikiforova, A.: Definition and evaluation of data quality: User-oriented data object-driven approach to data quality assessment. *Baltic J. Mod. Comput.* **8**, 391–432 (2020)
48. Oliveira, M.I., Oliveira, L.E., Batista, M.G., Lóscio, B.F.: Towards a meta-model for data ecosystems. In: *Annual International Conference on Digital Government Research: Governance in the Data Age*, pp. 1–10 (2018)
49. Olsina, L., Lew, P.: Specifying mobileapp quality characteristics that may influence trust. In: *Central & Eastern European Software Engineering Conference in Russia, CEE-SECR*, pp. 1–9 (2017)

50. Omidbakhsh, M., Ormandjieva, O.: Toward a new quality measurement model for big data. In: 9th International Conference on Data Science, Technology and Applications, pp. 193–199 (2020)
51. Valencia-Parra, Á., Parody, L., Varela-Vaca, Á.J., Caballero, I., Gómez-López, M.T.: DMN for data quality measurement and assessment. In: Di Francescomarino, C., Dijkman, R., Zdun, U. (eds.) BPM 2019. LNBP, vol. 362, pp. 362–374. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37453-2_30
52. Pereira, J.D., et al.: A platform to enable self-adaptive cloud applications using trustworthiness properties. In: International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS, pp. 71–77 (2020)
53. Rahman, M.S., Reza, H.: Systematic mapping study of non-functional requirements in big data system. In: IEEE International Conference on Electro Information Technology, pp. 25–31 (2020)
54. Rudraraju, N.V., Boyanapally, V.: Data quality model for machine learning. Faculty of Computing, Blekinge Institute of Technology, pp. 1–107 (2019)
55. Santhanam, P.: Quality management of machine learning systems. In: Shehory, O., Farchi, E., Barash, G. (eds.) EDSMLS 2020. CCIS, vol. 1272, pp. 1–13. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62144-5_1
56. Serhani, M.A., Kassabi, H.T., Taleb, I., Nujum, A.: An hybrid approach to quality evaluation across big data value chain. In: IEEE International Congress on Big Data, pp. 418–425 (2016)
57. Surendro, O.K.: Academic cloud ERP quality assessment model. *Int. J. Electr. Comput. Eng.* **6**, 1038–1047 (2016)
58. Taleb, I., Serhani, M.A., Dssouli, R.: Big data quality assessment model for unstructured data. In: International Conference on Innovations in Information Technology (IIT), pp. 69–74 (2018)
59. Taleb, I., Serhani, M.A., Dssouli, R.: Big data quality: a survey. In: 7th IEEE International Congress on Big Data, pp. 166–173 (2018)
60. Taleb, I., Serhani, M.A., Dssouli, R.: Big data quality: a data quality profiling model. In: Xia, Y., Zhang, L.-J. (eds.) SERVICES 2019. LNCS, vol. 11517, pp. 61–77. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23381-5_5
61. Talha, M., Elmarzouqi, N., Kalam, A.A.: Towards a powerful solution for data accuracy assessment in the big data context. *Int. J. Adv. Comput. Sci. Appl.* **11**, 419–429 (2020)
62. Tao, C., Gao, J.: Quality assurance for big data application - issues, challenges, and needs. In: International Conference on Software Engineering and Knowledge Engineering, SEKE, pp. 375–381 (2016)
63. Tepandi, J., et al.: The data quality framework for the Estonian public sector and its evaluation: Establishing a systematic process-oriented viewpoint on cross-organizational data quality. In: Hameurlain, A., Küng, J., Wagner, R., Sakr, S., Razzak, I., Riyad, A. (eds.) Transactions on Large-Scale Data- and Knowledge-Centered Systems XXXV. LNCS, pp. 1–26. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-662-56121-8_1
64. Vale, L.R., Sincorá, L.A., Milhomem, L.D.: The moderate effect of analytics capabilities on the service quality. *J. Oper. Supp. Chain Manag.* **11**, 101–113 (2018)
65. Vetrò, A., Canova, L., Torchiano, M., Minotas, C.O.: Open data quality measurement framework: definition and application to open government data. *Gov. Inf. Q.* **33**, 325–337 (2016)
66. Bautista Villalpando, L.E., April, A., Abran, A.: Performance analysis model for big data applications in cloud computing. *J. Cloud Comput.* **3**(1), 1–20 (2014). <https://doi.org/10.1186/s13677-014-0019-z>
67. Vostrovsky, V., Tyrychtr, J.: Consistency of Open data as prerequisite for usability in agriculture. *Sci. Agric. Bohem.* **49**, 333–339 (2018)

68. Wan, Y., Shi, W., Gao, L., Chen, P., Hua, Y.: A general framework for spatial data inspection and assessment. *Earth Sci. Inf.* **8**(4), 919–935 (2015). <https://doi.org/10.1007/s12145-014-0196-9>
69. Wang, B., Wen, J., Zheng, J.: Research on assessment and comparison of the forestry open government data quality between China and the United States. In: He, J., et al. (eds.) *ICDS 2019*. CCIS, vol. 1179, pp. 370–385. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-2810-1_36
70. Wang, C., Lu, Z., Wu, Z., Wu, J., Huang, S.: Optimizing multi-cloud CDN deployment and scheduling strategies using big data analysis. In: *International Conference on Services Computing, SCC*, pp. 273–280 (2017)
71. White, G., Nallur, V., Clarke, S.: Quality of service approaches in IoT: a systematic mapping. *J. Syst. Softw.* **132**, 186–203 (2017)
72. Zhang, P., Zhou, X., Li, W., Gao, J.: A survey on quality assurance techniques for big data applications. In: *IEEE Third International Conference on Big Data Computing Service and Applications*, pp. 313–319 (2017)