

QUALITY IN SOFTWARE ENGINEERING ESSAY

Halil Ibrahim Uluoglu

University of Eastern Finland
312843

Topic: Discussion about data quality in software engineering

2020

Discussion about data quality in software engineering

This essay examines the relationship data and dataset quality in software engineering. As stated that data quality has been studied and researched for years and it also depends on use. Since the use of geodata becomes wider, QC is increasingly getting more importance [1]. As Prat and Madnick noted, data quality is crucial for operational efficiency and sound decision making [2]. Intelligibly, researchers are concerned that data quality is a fair indication of software engineering and data science.

This essay argues that in general, quality of data is importance factor for performance, although this is not evident in all situations, nor with all types of project. Further, it is argued that the main mechanism for this data quality assessment is a complex process through which we can obtain a complete assessment of an organization's data. Through this process we get not only a full image regarding the data quality issues the company is facing but also an accurate view of the time and effort required to fix those problems [3]. The essay also points out that although many impacts on data quality in software engineering in various claims.

To figure out data quality in software engineering, first of all, we should understand about which components creates quality in data. Low data quality is normally understood as a form of fetching data, although how it is fetched varies. Researcher indicates that there are two forms of affects on data quality: fetching and handling. Fetching data is understood as a basic and only affects for quality. For instance, Indulska and Sadiq say that "yet data consumers may not be aware of data quality problems or small nuances between how the data was initially collected versus its newly planned use [4]". Importance about fetching data is starts at very beginning. Because no matter what you are working you need some data to analyze, modify and process. In order to fetch data, it needs genuine, pure, not artificial and this kind of 200 elements like [5] resources. If you have not enough these qualifications, probably, you are not going to achieve good result end of the research. Furthermore, the research needs to start with well enough qualifications. As we understand from there, fetching data is not only affects quality and also handling data affects quality of data.

Açıklamalı [HİU1]: This is a claim of fact→ In this sentence, it is a fact center sentence but it debates about the sentences is true or not. Reader may argue with this fact and also she/he may do some empirical work on this sentence by researching about data quality history and how have it used for.

Açıklamalı [HİU2]: In first paragraph, the writer indicates that he will use Aristotelian Argument in his essay. In order to that he introduces to essay issue. At the end of his introduction, the idea is to present his readers with his main point and then dig into it.

Açıklamalı [HİU3]: This is a claim of cause→ In these sentences, the writer gives an references in first sentence. The first sentence makes "What let to do this?" question to reader. After sentence responses "What is the result of the reference?" this question.

Açıklamalı [HİU4]: This is a claim of value→ In this sentence, the writer says about something arguing over the years. However, it is been argued but the reader should decide should this statement is admissible or not.

Açıklamalı [HİU5]: In the second introduction paragraph, writer present his case by explaining the issue in detail. The author expresses the what is the main point of essay and what is the idea behind it. It considered as Aristotelian present case paragraph.

Açıklamalı [HİU6]: This is a claim of definition→ In these sentences, the author gives a definition about there are two forms of affects on data quality.

Açıklamalı [HİU7]: In here, the author makes for his introduction to address the opposition.

Açıklamalı [HİU8]: This is a claim of cause→ In these sentences, the author finalizes his paragraph and he ends with a cause for previous sentences.

Açıklamalı [HİU9]: In here, the writer provides the side of his thinking with by giving example. He provides his conclusion his claims. Also, He blends his positive and negative claims and make a conclusion from his claims. He finalize with a suggestion.

Although, fetching data is important in qualification, it is not the only element for qualification. Evidence points to fact that, many papers published that for analyzing the data [6]. This relationship has been studied for since computer era started. During that time, some studies have reported that only fetching data is important and also some of them have reported that only handling data is important. However, these foresights are not complete because these studies point out single part of the whole picture. So, these findings are not likely to be true of all cases. For example, In Italy, an e-Government project, called “Services to Businesses” has been carried on with the specific aim of reconciling conflicts among three different public administrations. The limit of this project was to adopt a centralized architecture and a considerable process re-engineering. Conversely, a more effective solution to address data quality problems in CISs requires to be distributed in order to be as less invasive as possible and to avoid bottlenecks [7]. As they mentioned in this example, the researcher have issued about handling the data in order to keep same as quality as fetching data. Processing data depends on processing techniques, methods, tools, technicians and variety of criteria. In order to keep quality stable researcher should take care of handling data as good as fetching data. Lastly, handling and fetching data are very important to keep well quality data.

Açıklamalı [HİU10]: This is a claim of fact→ In here, the author points out a fact. The reader may clarify or not clarify this fact by using empirical study.

On the other hand, in some specific situations are not negatively impacted by quality of data. Despite the general agreement about the relationship between fetching data and handling data, the issue is, in fact, more complex. When the data is less considered on a project, the impact of qualification of data is lower or even absent. For example, in this research, they have examined to way different result affects on the project. They noted that, “We also analyze the key factors that affect its performance, based on testing with heterogeneous filtering requirements. [8]” Furthermore, some projects do not deal with quality of data. They need enough data to analyze Gade says that to get effective and more accurate answers for the tool generated questions we make use of crowd sourcing concepts and developed algorithms. This has helped in getting better results and filter out low quality links that got generated which are irrelevant to the search [9]. As we understand from here, some project not depends on data and moreover quality of data.

Açıklamalı [HİU11]: This is a claim of policy→ In here, the author tells the readers that you should accept this policy. It is that handling data and fetching data should use together in case of qualifying the data.

Açıklamalı [HİU12]: In here, the author makes for his introduction to address the opposition. The writer provides the side of his thinking with by giving example. He provides his conclusion his claims. Also, He blends his positive and negative claims and make a conclusion from his claims. He finalize with a suggestion.

Açıklamalı [HİU13]: This is a claim of cause→ In these sentences, the author gives and references to prove that they need enough data to analyze. After a couple facts based sentences, the author gives cause for that basement of opinion at the end of paragraph.

Açıklamalı [HİU14]: In here, the author give an opposite opinion to previous paragraph. He points that fetching and handling data is not as important as in some project. First, he introduces the opposite side of opinion and present his case then, he proofs a point and he present a conclusion.

The impact of quality on data has been very closely studied for a long time. There is general acceptance that debilitative quality of data positively impacts project or programs performance and that it does so via an interference mechanism in which task-irrelevant thoughts undermine a computers ability to recall previously learned material. However, this essay also shows that not all projects are impacted by quality of data. Quality of data appears to have its greatest impact in difficult projects and in earlier years of a computer history. As well, not all projects have equally importance. Some projects have needs higher quality of data than other projects, and while their human including capability are reduced, the needy for qualified data also reduce. While the question of data qualification has been studied for many years, will it still an important issue to consider as its presence does mean that a data qualification is not fairly assessed during a project [10].

Açıklamalı [HiU15]: This is a claim of value → In here, the author specify a claim that quality of data will gain importance in the future. Moreover, with this claim authors wants from us to make a decision for claim of author.

Açıklamalı [HiU16]: In the last of paragraph, he summarizes all previous paragraphs and he makes a conclusion. It is that while the question of data qualification has been studied for many years, will it still an important issue to consider as its presence does mean that a data qualification is not fairly assessed during an project. He provides a reference with that.

Reference List

- [1] Erdogan, M., Torun, A., & Boyaci, D., "REVISITING THE PROCEDURES FOR THE VECTOR DATA QUALITY ASSURANCE IN PRACTICE" , *Int. Arch. Photogram. Remote Sens. Spatial Inf. Sci.*, XXXIX-B4, 23–27, pp 23.
- [2] Prat, N., Madnick S., "Evaluating and Aggregating Data Believability across Quality Sub-Dimensions and Data Lineage" , *ESSEC Business School & Sloan School of Management Massachusetts Institute of Technology*, January 2008.
- [3] Andreescu, I., A., Belciu, A., Florea, A., Diaconita, V., "Measuring Data Quality in Analytical Projects", *University of Economic Studies, Bucharest, Romania* , *Database Systems Journal* vol. V, pp. 15, no. 1/2014.
- [4] Indulska, M., Sadiq, S., "In order to fully realise the value of open data researchers must first address the quality of the datasets", *blogs.lse.ac.uk* , pp. 2, April 2017.
- [5] Andreescu, A., Belciu, A., Florea, A., Diaconita, V., "Measuring Data Quality in Analytical Projects", *Database Systems Journal*, pp. 15, vol. V, no. 1/2014.
- [6] Liebchen, G., Shepperd, M., "Data Sets and Data Quality in Software Engineering", *PROMISE'08, ACM*, May 12–13, 2008, Leipzig, Germany.
- [7] Milano, D., Scannapieco, M., Catarci, T., "Peer-to-Peer Data Quality Improvement in the DaQuinCIS System", *Universita di Roma "La Sapienza" ` Dipartimento di Informatica e Sistemistica Via Salaria 113*, 00198 Rome, Italy
- [8] Li, M., Kotz, D., "Group-aware Stream Filtering for Bandwidth-efficient Data Dissemination", *Institute for Security Technology Studies Department of Computer Science, Dartmouth College Hanover, NH* 03755
- [9] Gade, S., "LinkWiper – A System For Data Quality in Linked Open Data", *Master of Science (Computer and Information Science) in the University of Michigan-* pp. 3, Dearborn 2016
- [10] Andreescu, I., A., Belciu, A., Florea, A., Diaconita, V., "Measuring Data Quality in Analytical Projects", *University of Economic Studies*,

Bucharest, Romania , Database Systems Journal vol. V, pp. 16, no.
1/2014.