Abstract—Organizations support their operations and decision making on the data they have at their disposal, so the quality of these data is remarkably important and Data Quality (DQ) is currently a relevant issue, the literature being unanimous in pointing out that poor DQ can result in large costs for organizations. The literature review identified and described 24 Critical Success Factors (CSF) for Data Quality Management (DQM) that were presented to a panel of experts, who ordered them according to their degree of importance, using the Delphi method with the Q-sort technique, based on an online questionnaire. The study shows that the five most important CSF for DQM are: definition of appropriate policies and standards, control of inputs, definition of a strategic plan for DQ, organizational culture focused on quality of the data and obtaining top management commitment and support.

Keywords—Critical success factors, data quality, data quality management, Delphi, Q-Sort.

I. INTRODUCTION

This paper describes a study that intends to answer the following research question: What are the CSF for DQM? To answer this question a literature review was performed, followed by a Delphi questionnaire with the Q-sort technique, supported by an online questionnaire. The CSFs identified in the literature review were presented to a panel of experts with the objective of obtaining consensus regarding their ordering by degree of importance.

Although data and information mean slightly different things, for reasons of simplicity the terms are often used interchangeably [4], [21]. Organizations are increasingly working with enormous volumes of information that is a fundamental asset and a differentiating factor if it has quality. The CSFs for DQM are a relatively unexplored issue that will be considered in this work [4].

In [5], Boynton & Zmud define CSFs as those aspects that must go well to ensure the success for a manager or an organization, and are vital for the company’s operational activities and its future success. In [23], Rockart categorized the CSFs in two groups: monitoring and building. The CSFs in the former group serve, as the name implies, to monitor the current results, and the CSFs in the latter one are oriented towards building for the future.

II. LITERATURE REVIEW

A. DQM

According to the Data Management Association [30, p.43], cited by Lucas [17], DQM consists of applying the concepts and practices of Total Quality Management (TQM) to improve the quality of data and information, involving the definition of policies and rules, assessment of DQ (including auditing and certification), data analysis, cleaning and correction, DQ improvement and education.

Reference [11] goes further and considers that DQM should incorporate the creation and implementation of roles, policies, responsibilities and processes related to the acquisition, maintenance, organization and distribution of data, requiring alignment between the technologies and the business. The Data Warehousing Institute estimates that DQ problems cost US companies more than $600 billion a year [7].

Table I summarizes the impact of poor DQ according to organizational levels.

<table>
<thead>
<tr>
<th>Organizational levels</th>
<th>Impact of poor DQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational</td>
<td>Reduces customer satisfaction</td>
</tr>
<tr>
<td></td>
<td>Increases costs</td>
</tr>
<tr>
<td></td>
<td>Decreases employee satisfaction</td>
</tr>
<tr>
<td></td>
<td>Makes decision making less accurate and more time-consuming</td>
</tr>
<tr>
<td>Tactical</td>
<td>Makes it difficult to implement data warehouses</td>
</tr>
<tr>
<td></td>
<td>More difficult to reengineer</td>
</tr>
<tr>
<td></td>
<td>Increased organizational mistrust</td>
</tr>
<tr>
<td></td>
<td>More difficult to set and to execute the strategy</td>
</tr>
<tr>
<td>Strategic</td>
<td>Contribute to issues of data ownership</td>
</tr>
<tr>
<td></td>
<td>Divergent management attention</td>
</tr>
</tbody>
</table>

B. CSF

Forster & Rockart [8] argue that the concept of CSF has decades and reported that Aristotle, almost two thousand years ago, expressed the idea that leaders should create a few simple things for their organizations and noted that those organizations that did so fared better than those that did not.

CSFs are essential elements for an organization or project successfully achieve its mission or objectives with success [15] and should be identified in a limited number of key sectors of the company, in order to ensure a competitive and successful performance [23].

Table II summarizes the CSF for DQM identified and defined through the literature review.

As the issue of CSFs to DQM is poorly treated in the literature, that was decided to search the CSFs in more mature areas, such as Knowledge Management Systems; Business
III. DATA COLLECTION AND METHODS

It was decided to use the Delphi method with Q-Sort, because according to Linstone [16] cited by Yousuf [40] there are two situations in which the Delphi method is appropriate:

- The question is not consistent with a precise technical analysis, but can benefit from subjective judgments on a collective basis;
• People who need to interact cannot meet face-to-face due to time or cost constraints.

A. Delphi Method

The original Delphi method was developed in the 50s by Dalkey of RAND Corporation for a military project and is defined by Gallego & Bueno [10] as a method that, through the availability of questionnaires, allows organizing and sharing opinions of experts, being one of the best known and used techniques in new areas of knowledge.

There is no single type of Delphi method because the method is modified to suit the circumstances and the research question [26] and because it is a flexible method being used by researchers in the area of Information Systems and Technologies (IS/IT).

In a Delphi study, an expert should be a specialist in the area of knowledge in which the study is conducted and each questionnaire should correspond to a round, making as many rounds as necessary to obtain a consensus or the confirmation that the consensus is not possible.


B. Q Method

The Q method was developed by Stephenson [28] cited by Santos [25] and provides grounds for the systematic study of subjectivity. The distinctive feature of the Q-Sort technique, a component of Q method, is that the panel members are required to order the questions provided under a predefined distribution, usually approximately normal.

The Q-Sort technique is usually preferred on a Likert scale when one wants an order and not just a weighting.

For the analysis and interpretation of the results, the e-Delphi tool with Q-Sort provided by the University of Minho was used. Developed specifically for the Delphi method, this tool automatically calculates some statistical data and orders the CSF in tables generated by the application itself [24].

In order to identify consensus, non-parametric statistical tests were performed. The IBM SPSS Statistics tool was used to calculate the Kendall’s concordance coefficient and the Spearman’s correlation coefficient.

C. Delphi Method and Q-Sort

24 CSFs identified and defined through the literature review were presented to the panel of experts using an online questionnaire.

Specialists with deep knowledge of DQ were invited to respond to the questionnaire, which is crucial for the final quality of the research. Fig. 1 illustrates the distribution by professional area of the 45 experts who accepted to participate in this study.

D. First Round

For the first round, 45 participants were invited to order the 24 factors that were presented in alphabetical order, according to the procedures of the Q-Sort methodology. 30 responses (66.67%) were received.

At the end of the first round, experts chose the five most important factors: organizational culture focused on DQ, the definition of appropriate policies and standards, the definition of a strategic plan for DQ, commitment and support of top management and input controls.

In addition to the mean, variance and standard deviation, the Kendall’s coefficient was calculated to determine the agreement within the ordering of factors. The value of this coefficient (0.178) shows that there is no strong correlation between the responses, so it was necessary to perform a second round, with the objective of improving the degree of convergence.

At the end of the first round, two of the experts presented 3 apparently new factors, but only one of them was added to the list in position 25. The other two factors were deemed to be covered by the already existing CSFs.

The new factor "data governance" is defined as the set of essential actions to ensure data compliance with organizational strategies.

E. Second Round

For the second round were only invited the respondents of the first round. Table III presents the results of the first and second rounds and in second round the first 5 CSFs are the definition of appropriate policies and standards, the input controls, the definition of a strategic plan for the DQ, the organizational culture with a focus on DQ, and top management commitment and support.

The Kendall’s W coefficient (0.227) rose slightly and according to Schmidt [27], this is still a low value that remained practically stabilized and indicates that it will be difficult to improve the agreement with a new round.

The Spearman’s coefficient that measures the convergence between the ordering of the first and second rounds was 0.425, which reflects a moderate agreement. In view of the result obtained, it was decided not to proceed with a new round for the following reasons:

• The Kendall’s coefficient remained virtually identical in the two rounds;
• The Spearman’s coefficient obtained reflects a moderate convergence between the answers of the first and second rounds;
• The research question has not yet reached a reasonable degree of maturity in the academic and professional communities;
• There were difficulties related to the time of the year in
which the survey was conducted;

- There were difficulties regarding the availability of experts to continue to participate;
- There was little time available for research.

**F. Final Results**

The results obtained after the two rounds are summarized in Table III, where the first CSF is "definition of appropriate policies and standards". This factor consists of the implementation of a standard methodology, standardization of codes, rules and definitions and reformulation of the data model, when necessary.

The second position is the "input controls" which means to ensure the quality of data loading, to certify external sources and existing data, to implement validation routines in data collection, and to control how data are generated.

It should also be noted that the five first factors maintained their positions in the two rounds, having only changed their order, which shows concordance of opinions.

| TABLE III
<table>
<thead>
<tr>
<th>RESULTS OF THE TWO ROUNDS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st round</strong></td>
</tr>
<tr>
<td>sum of points</td>
</tr>
<tr>
<td>209</td>
</tr>
<tr>
<td>258</td>
</tr>
<tr>
<td>225</td>
</tr>
<tr>
<td>183</td>
</tr>
<tr>
<td>254</td>
</tr>
<tr>
<td>n/a</td>
</tr>
<tr>
<td>343</td>
</tr>
<tr>
<td>387</td>
</tr>
<tr>
<td>382</td>
</tr>
<tr>
<td>406</td>
</tr>
<tr>
<td>398</td>
</tr>
<tr>
<td>378</td>
</tr>
<tr>
<td>324</td>
</tr>
<tr>
<td>394</td>
</tr>
<tr>
<td>436</td>
</tr>
<tr>
<td>390</td>
</tr>
<tr>
<td>408</td>
</tr>
<tr>
<td>440</td>
</tr>
<tr>
<td>454</td>
</tr>
<tr>
<td>455</td>
</tr>
<tr>
<td>401</td>
</tr>
<tr>
<td>430</td>
</tr>
<tr>
<td>431</td>
</tr>
<tr>
<td>513</td>
</tr>
<tr>
<td>501</td>
</tr>
</tbody>
</table>

The two least important factors also maintained their positions in the two rounds. Thus it would seem that the workplace environment and all that is entailed by effective relationships with employees – empowerment and participation in the organization, satisfaction, job security and career development – are not considered relevant to DQM.

These results are in line with the literature review.

**IV. DISCUSSION**

DQM includes policies, processes and procedures that contribute to the improvement of DQ, so it is in the interest of the scientific and business communities to identify the CSFs that influence data quality management.

This study identified the five most important CSFs influencing DQM: the definition of appropriate policies and standards, the input controls, the definition of a strategic plan for DQ, the organizational culture focused on DQ and the top management commitment and support. These results are in line with previous studies. References [36], [33] and [2] are unanimous regarding the importance of top management support, as well as the input controls [36].

The involvement of the users and the focus on their needs and requirements pointed by [33] and [2], as well as the acquisition of new skills and initial and continuing training were classified in position 12 and 13 in last round, so they were not considered by the experts of this study as very important for DQM. The most important CSF in this work "definition of appropriate policies and standards” is not among the top 10 CSFs identified by [36] and [2] and ranks in the ninth position in [33]

The study also identified the importance of data governance, defined as a set of essential actions to ensure data compliance with organizational strategies. This CSF was proposed by one of the experts and although it is not mentioned in the literature, it was placed in the sixth position, thus demonstrating the convergence of opinions regarding its importance. The subjects Data Governance and DQM, though distinct, are related, for DQM’s pursuit of collaboration between business and IT implies cultural change. Such change can only be achieved with leadership, authority, control and resource allocation, which, in turn, implies data governance [17].

**V. CONCLUSION**

The present study used the Delphi method with Q-Sort technique to order the CSFs for the DQM identified in the literature review. The CSFs were evaluated and ordered by a panel of experts in the area.

Although there was no convergence of opinions, the study shows that DQM requires the implementation of a methodology, the definition of policies and the standardization of codes and rules, as well as the implementation of measures for data control and validation. The need for a strategic plan, the commitment to an organizational culture focused on DQ, and the top management commitment, control and support are also relevant, according to the panel of experts.

**VI. LIMITATIONS AND FUTURE WORK**

The study suffers from a few limitations concerning its implementation. First, due to the low degree of maturity of this subject in academic and professional communities, it was difficult to find experts willing to participate. Second, there...
were limitations related to the time of the year in which the investigation was carried out, and the available period for the investigation. Finally the literature notes several limitations associated with both the FCS and the Delphi method with Q-Sort.

Davis [6] cited by Guynes & Vanecek [12] identified several issues related to the approach of the CSFs, specially the difficulty of human beings in constructing simplified models that reflect real-world situations. There are other limitations related to the difficulties in using the method and in the ability of the researcher, Boynton & Zmud [5] indicate that the approach is feasible and useful with adequate training.

Regarding the application of the Delphi method there are some limitations like Santos & Amaral [25] mentioned, this method has been criticized for lack of guarantee of confidence in relation to the results obtained, being necessary, especially when using the Q-sort technique, to pay attention to the time spent collecting data and the effort that is requested from the expert panel, although these limitations can be reduced using a web tool.

Despite using one online tool, e-Delphi, it was considered unintuitive by some experts, requiring some effort to understand. Some CSFs, due to their complexity, have some overlap regarding its definition, that perhaps have hampered the response of the experts, since a good understanding of the CSFs could lead to a faster consensus.

This study suffers from the limitations of the qualitative research methodology used. As a result, it only allows analytical generalization, generalization to the theory, and not generalization to the population [39]. On the other hand, it is an exploratory study about a poorly studied subject, whose objective was to understand the phenomenon and to explore the research question: What are the CSF for DQM?

ACKNOWLEDGMENTS

This work was supported by FCT, I.P., the Portuguese national funding agency for science, research and technology, under the Project UID/SOC/04521/2019.

REFERENCES


