DMN for Data Quality Measurement and Assessment

Álvaro Valencia-Parra[1(✉)], Luisa Parody[2], Ángel Jesús Varela-Vaca[1], Ismael Caballero[3], and María Teresa Gómez-López[1]

[1] Universidad de Sevilla, Seville, Spain
{avalencia,ajvarela,maytegomez}@us.es
[2] Universidad Loyola Andalucía, Córdoba, Spain
mlparody@uloyola.es
[3] Universidad Castilla-La Mancha, Ciudad Real, Spain
Ismael.Caballero@uclm.es
http://www.idea.us.es/

Abstract. Data Quality assessment is aimed at evaluating the suitability of a dataset for an intended task. The extensive literature on data quality describes the various methodologies for assessing data quality by means of data profiling techniques of the whole datasets. Our investigations are aimed to provide solutions to the need of automatically assessing the level of quality of the records of a dataset, where data profiling tools do not provide an adequate level of information. As most of the times, it is easier to describe when a record has quality enough than calculating a qualitative indicator, we propose a semi-automatically business rule-guided data quality assessment methodology for every record. This involves first listing the business rules that describe the data (data requirements), then those describing how to produce measures (business rules for data quality measurements), and finally, those defining how to assess the level of data quality of a data set (business rules for data quality assessment). The main contribution of this paper is the adoption of the OMG standard DMN (Decision Model and Notation) to support the data quality requirement description and their automatic assessment by using the existing DMN engines.

Keywords: Data quality · Decision Model and Notation · Data quality measurement · Data quality assessment

1 Introduction

Globalization and emerging technologies are bringing new challenges to companies in the contexts in which enterprises should use massive amounts of data, most of them provided by third parties. In order to guarantee the success of the various tasks, and to satisfy the necessities of the business processes that use the data, it is paramount to face up with the quality of these data [12]. Just to mention an example of a critic task, let us think about the integration of data coming from different and heterogeneous sources in complex scenarios [3], and the many problems related to data quality that can arise when it comes to build a new dataset.

To achieve the largest benefits of data, companies will need to find out ways to automatically manage the levels of data quality; this is even more important, in scenarios where it is required high efficiency in terms of computational cost of the operations per second. The assessment of data quality largely depends on the context of the use of data. Despite of the extensive literature around data quality (e.g., definition of methodologies for requirement definitions, selection of criteria to judge data quality- data quality dimensions [1]), assessment and measurement procedures have been typically developed ad-hoc [13]. The data quality context-awareness needs the description of the business rules representing the data quality requirements. In order to describe the data quality requirements, we propose the description of three sets of business rules: those describing the data requirements (BR), those describing how to measure the data quality dimensions (BR.DQM), and those describing how to assess the data quality (BR.DQA) in terms of the measures of the data quality.

The assessment follows a procedure based on the sequencing of the verification in cascade of the three previously-mentioned types of business rules in two phases: first, we conduct the verification of the BRs to estimate a measurement for every data quality dimension according to BR.DQM; and, then, we use the verification of the BR.DQM to produce an estimation of the assessment of the level of data quality, according to the stated BR.DQA. Based on the result of this assessment, the organization should make a decision on the use of the record based on its risk appetite. In this paper, we propose the use of DMN (Decision Model and Notation) [15], a declarative language proposed by OMG to facilitate the description of the business rules, as well as their evaluation [6]. DMN provides human readers with a more understandable and visual representation of business rules [8], being data quality requirements a new scenario where DMN can be used.

The remainder of this paper is organized as follows: Sect. 2 details the foundations involved in this paper; Sect. 3 introduces a case study to make the proposal accessible; Sect. 4 presents the proposal of application of DMN to data quality measurements and assessment respectively; Sect. 5 presents the related work and Sect. 6 concludes and remarks the lessons learned.

2 Foundations

2.1 Data Quality Management: Rules for Measurement and Assessment

The cornerstone of our proposal of data quality management is grounded in the difference between two important concepts typically used as synonymous: "measurement" and "assessment". This differentiation is based on the two definitions of quality: the "meeting requirements" by Crosby (or *"how well data is built"*) and the "fitness for use" by Juran (or *"how usable the data is"*) [19]. The basis

Table 1. Data quality dimensions from Wang [19]

Data quality category	Data quality dimension
Intrinsic	Accuracy, Objectivity, Believability, Reputation
Accessibility	Access, Security
Contextual	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational	Interpretability, Ease of understanding, Concise representation, Consistent representation

of our proposal is to describe by means of business rules when "data is well built" according to several data quality dimensions and when "data is usable" according to the assessment of the quality including all data quality dimensions. Data quality dimensions (criteria used to evaluate the quality of data) are at the core of data quality management [20] because they represent users' data quality requirements. Several authors have proposed their own set of data quality dimensions, both generic ones (like the ones proposed by [19] -see Table 1- or the introduced in ISO 25012 [10]) or for specific context.

When it comes to estimate the amount of data quality that a dataset has, data profiling tools are typically employed to produce some measures that data quality processes use as indicators [9]. But, without knowing how the semantics of data has been implemented in the data model, the results cannot be interpreted, and cannot be used to diagnose the root causes of a low level of data quality. However, it is more than enough to have a qualitative indications of whether data is usable or not. This strategy is specially recommendable when it comes to determine if a record should be used as part of the execution of an instance of a business process. More specifically, when the decision should be made according to the quality of the used data, desirably in an automatic way. At this point, let us recall that every record is a set of attributes a_i; every attribute a_i or every set of attributes $a_i, a_j, a_k, \ldots$ must meet some data requirements specified by means of several Business Rules (BR). Some typical statements of business rules look like:

– *BR.01. The attribute a_1 must meet the regular expression RE*
– *BR.02. The attribute a_2 (datatype numeric) should be lower than the attribute a_3 (datatype numeric)*

The measurement implies the verification of the business rules associated to every chosen data quality dimension (e.g. completeness, consistency, ...). To produce a value for measurement, we need a BR.DQM that describes the possible values that every data quality dimension could obtain. Depending on the granularity, a BR.DQM can be defined in terms of one or more attributes and one or more BRs. We propose to use Likert scales to define the possible values for the results of measurement. As an example of possible values of completeness is the set {"Complete Enough", "Not Enough Complete", "Dramatically Non-complete"}. Typically, BR.DQM sentences for completeness can look like:

- *BR.DQM.01. A record can be considered as "Complete Enough" if it meets BR.01 and BR.02*
- *BR.DQM.02. A record can be considered as "Not Enough Complete" if it only meets BR.01*
- *BR.DQM.03. A record can be considered as "Dramatically Non-complete" if it does not meet neither of BR.01 nor BR.02*

Finally, and after defining how to measure the data quality dimensions, it is time for aggregating the measures for the various data quality dimensions to produce an indication of the amount of data quality that a record has. The result of this aggregation will represent the level of usability of a record for a task in terms of the risk appetite of the organization for the underlying task. Once again, a Likert scale (e.g., "Usable", "Potentially usable but risky", "Non-usable") is proposed and the statement of the corresponding business rules for the assessment (BR.DQA) should be done in terms of the chosen data quality dimensions and the corresponding BR.DQM. Typically, BR.DQA can look like:

- *BR.DQA.01. A record can be considered as "Usable" if it meets BR.DQM.01*
- *BR.DQA.02. A record can be considered as "Potentially usable but Risky" if it meets BR.DQM.02*
- *BR.DQA.03. A record can be considered as "Non-usable" if it meets BR.DQA.03*

Please see Sect. 3 for a larger motivating example of the description of an assessment.

2.2 Decision Model and Notation

Decision Model and Notation (DMN) is a modelling language and notation defined to describe business rules [15]. DMN provides a simple way to define the decision logic model understandable by all users, in our case, from the business experts in charge of describing the processes to the data quality experts responsible for defining the quality requirements.

The DMN standard includes two components:

- Decision requirements diagram that defines the decisions to be made, their interrelationships, and their requirements for decision logic.
- Decision logic that defines the required decisions in sufficient details to allow validation and/or automation.

An example of a Decision Requirement Diagram is presented and detailed in Fig. 1. The *decision task* describes a specific task that includes a decision logic, that is, depending on some input values set output values as described in a decision table. The input data, as the name implies, is the necessary data that the decision logic needs to determine the output value. For instance, in the example,

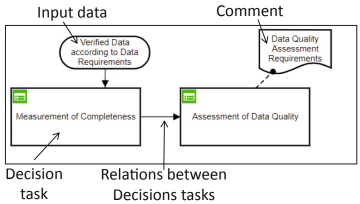

Fig. 1. Decision requirement diagram example

the *Verified Data according to Data Requirements* feeds *Measurement of Completeness* task. On the other side, the arrows that connect two decisions tasks indicates the relationships between these two tasks. In the example presented in Fig. 1, the output of the decision task *Measurement of Completeness* feeds the *Assessment of Data Quality* task, which means that the decision output of the first task is considered as input for the second decision.

Each decision task includes a decision table. The decision table used in this article has an horizontal orientation: the input and outputs are defined in columns and the rules as rows. An example is presented in Fig. 2, which also indicates the different components of the table.

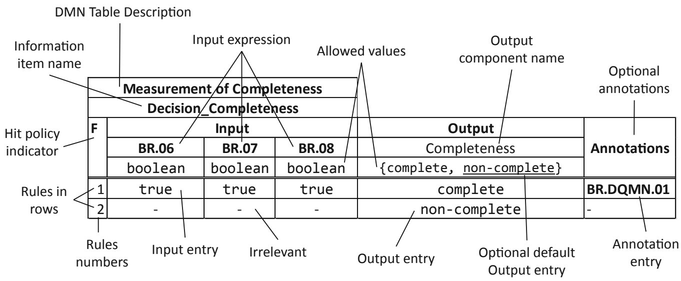

Fig. 2. Decision table example for the *Measurement of Completeness* task (see Fig. 1)

The *information item name* is the name of the variable (i.e. information item) for which the decision table provides the decision logic. The *hit policy indicator* indicates how to handle the multiple matches. In our case, the *F* means that although multiple rules can match, only the first hit by rule order is returned. There are other possible *hit policy indicators* such as Unique (U), Any (A), and Priority (P) (see [15] for further details). There is also a set of input clauses composed of an *input expression* and optional *allowed values* for the input entries, for instance *BR.06* is an input expression whose possible value is a *boolean* (true and false values). An input entry is contained in a rule: the value *true* corresponds to the input entry for the *BR.06* in the *rule number* 1. The input cell entry '-' means irrelevant, i.e. it can have any of the allowed values. Moreover, a set of output clauses are also included in a decision table. An output clause consists of an *output component name* and its *allowed values*. The allowed value that is underlined corresponds to the default value. Finally, a set of annotations clauses can also be included in the decision table. In our case, rule number 1 is annotated with the entry *DQ.DQMN.01*. The decision table displays the rules in an abbreviated notation organizing the entries in table cells. For example, the rule number 1 can be read as: ***If*** *(BR.06 and BR.07 and BR.08)* ***then*** *Completeness = 'complete'*.

3 Motivating Example

In order to illustrate how a semi-automatically business rule guided-data quality assessment can be done, we adapted a well-known example of a movie-database introduced in [2] and shown in Table 2. The adaptation just consist of changing some values to have a much better casuistry in the data quality assessment.

Table 2. Example of dataset with data quality problems.

Id	Title	Director	Year	#Remakes	LastRemakeYear
1	Casablanca	Weir	1942	3	1940
2	Dead Poets Society	Weir	1989	0	NULL
3	Rman Holiday	Wylder	1953	0	NULL
4	Sabrin	NULL	1964	0	1985

To illustrate the data quality assessment, we introduce several business rules (*BR*, *BR.DQM*, and *BR.DQA*):

- **Data Requirements**: Associated to the given dataset, some business rules (BR) describing some syntactic and/or semantic data requirements are listed:
  - *BR.01*. The attribute *Title* contains a string no longer than 256 characters
  - *BR.02*. The attribute *Title* must exists in the IMDB database
  - *BR.03*. The attribute *Director* contains an string no longer than 30 characters
  - *BR.04*. The attribute *Director* must appear in the IMDB database associated to the movie having the title specified in the attribute *Title*
  - *BR.05*. The attribute *Year* must be a positive number between 1895 and 2030
  - *BR.06*. The attribute *LastRemakeYear* must be always greater than *Year*
  - *BR.07, BR.08* and *BR.09*. The attributes *Title*, *Director*, and *Year* can not be null
  - *BR.10*. If the attribute *#Remakes* is zero, then the attribute *LastRemakeYear* must be null

- **Data Quality Measurement Requirements**: The data quality dimensions along with possible values for expressing the results of measurements are the following:
  - **Completeness**: the possible values are {"Complete", "Non-Complete"}
    * *BR.DQM.01.* A record is *Complete* when meet the business rules *BR.06*, *BR.07* and *BR.08*
    * *BR.DQM.02.* A record is *Non-complete* when does not meet *BR.DQM.01*
  - **Accuracy**, having the values {"Very Accurate", "Accurate", "Inaccurate"}:
    * *BR.DQM.03.* A record is *Very accurate* when meets *BR.02* and *BR.04*
    * *BR.DQM.04.* A record is *Accurate* when meets *BR.02*
    * *BR.DQM.05.* A record is *Inaccurate* when does not meet neither *BR.DQM.03* nor *BR.DQM.04*
  - **Consistency**, having values from the set {"High Consistency", "Consistency", "Low consistency"}:
    * *BR.DQM.06.* A record is *High Consistency* when meets *BR.04* and *BR.09*
    * *BR.DQM.07.* A record is *Consistency* when meets BR.04
    * *BR.DQM.08.* A record is *Low consistency* when does not meet neither *BR.DQM.06* nor *BR.DQM.07*
- **Data Quality Assessment Requirements**: After measurement, it is the time to aggregate the results previously obtained, in order to generate a judgment about the usability of a record. For this example, we consider the following values: {"suitable quality", "enough-adequate quality", "non-usable"}
  - A record is said to be of *suitable quality* when meet *BR.DQM.01*, *BR.DQM.02*, *BR.DQM.04*, and *BR.DQM.05*.
  - *BR.DQA.02.* A record is said to have *adequate quality for use* when meet *BR.DQM.01, BR.DQM.04*, and *BR.DQM.05*.
  - BR.DQA.03. A record is said to be *non-usable* when does not meet any of *BR.DQM.01, BR.DQM.02, BR.DQM.03, BR.DQM.04* or *BR.DQM.05*.

4 Decision Model in DMN for Data Quality

The decision model presented in Fig. 3 establishes two hierarchical levels to reduce the DMN complexity [7]. The bottom level corresponds to the data quality dimensions to be measured. Following the motivating example, these dimensions are *Completeness*, *Accuracy*, and *Consistency*. The upper level is related to the assessment of the level of data quality.

4.1 Data Quality Measurement

Data quality measurement is at the bottom level in the Decision Model. As aforementioned, it is related to each dimension to be measured in accordance with the Data Quality Measure Requirements presented in Sect. 3.

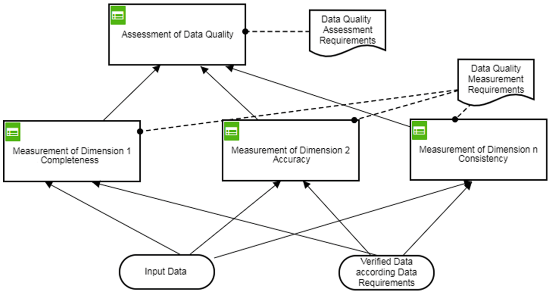

Fig. 3. Decision model diagram for data quality assessment

The first dimension modeled is *Completeness* (see Fig. 2). The input for building the DMN table for *Completeness* are: *BR.06*, *BR.07*, and *BR.08*. Each of these columns might take a *Boolean* value (i.e., true or false) after the verification of the corresponding rule. The different entries (i.e., rows) return an output assessment value for this dimension (cf., Output column). In this example, the completeness dimension has two entries (two rows). The first entry sets *true* to each one of the input value, associating the value *complete* to the output column. It indicates that, in order to consider that the data is *complete*, then it must fulfill the requirements *BR.06*, *BR.07*, and *BR.08*. The second entry basically indicates that if the data does not meet none of the aforementioned requirements, then it is considered *non-complete*.

The same logic applies for building the DMN tables of the other two dimensions. Figure 4 depicts the DMN table for *Accuracy*. It is composed of two input columns, that correspond with *BR.02* and *BR.04*. Three entries establish the data quality measurement requirements for *Accuracy*. The first entry indicates that both requirements must be fulfilled in order to consider the data as *very accurate*. The second entry specifies that only the requirement *BR.02* must be met in order to consider the data as *accurate*. The last entry indicates that the data is *inaccurate* if it does not fulfill none of both requirements.

Measurement of Accuracy				
Decision_Accuracy				
F	Input		Output	Annotations
	BR.02	BR.04	Accuracy	
	boolean	boolean	{very accurate, accurate, inaccurate}	
1	true	true	very accurate	BR.DQMN.02
2	true	-	accurate	BR.DQMN.03
3	-	-	inaccurate	-

Fig. 4. DMN for measurement accuracy.

The last dimension is *Consistency*, shown in Fig. 5. In this case, *BR.04* and *BR.09* are the requirements used to evaluate this dimension. If both requirements are fulfilled, then the data is considered *highly consistent*. If only the requirement *BR.04* is met, then the data is considered *consistent*. In any other case, the consistency of the data is low.

Measurement of Consistency				
Decision_Consistency				
F	Input		Output	Annotations
	BR.04	BR.09	Consistency	
	boolean	boolean	{high consistency, consistency, low consistency}	
1	true	true	high consistency	BR.DQMN.04
2	true	-	consistency	BR.DQMN.05
3	-	-	low consistency	-

Fig. 5. DMN for measurement consistency.

4.2 Data Quality Assessment

Data quality assessment is at the top level in our decision model (cf, Fig. 6). In this level, there is only one DMN table, which takes as input the values returned by the previous DMN tables corresponding to data quality measurement of all dimensions.

Assessment of Data Quality					
Decision_Assessment					
F	Input			Output	Annotations
	Completeness	Accuracy	Consistency	Level of Data Quality	
	{complete, non-complete}	{very accurate, accurate, inaccurate}	{high consistency, consistency,low consistency}	{suitable, enough-adequate, non-usable}	
1	complete	not(inaccurate)	not(low consistency)	suitable	BR.DQA.01
2	complete	-	not(low consistency)	enough-adequate	BR.DQA.02
3	-	-	-	non-usable	BR.DQA.03

Fig. 6. DMN for assessment of data quality

The DMN table consists of three input columns, one per each data quality dimension measured. The value which each column might take depends on the value returned by the corresponding dimension. In this case: *Completeness* could take these values *complete* or *non-complete*; *Accuracy* can be *very accurate*, *accurate* or *inaccurate*; and *Consistency* could take *high consistency*, *consistency* or *low consistency*. The output represents the overall level of data quality which can be one of these values: *adequate*, *suitable*, or *enough*.

The table is composed of three rules. The first rule indicates that the overall level of data quality is *adequate* if *Completeness* is *complete*, *Accuracy* is not *inaccurate*, and *Consistency* is not *low consistency*. It must be highlighted that two of the entries of this rule support more than one valid value. For example, the entry for *Accuracy* indicates that must not be inaccurate, which means the value might be either *very accurate* or *accurate*. The same applies for *Consistency*. Thus, the second rule is similar to the first one, except for the *Accuracy* entry.

In this case, it establishes that this dimension can take any of the possible values. If this rule is met, then the overall level of data quality is *suitable*. The last rule establishes that, for any dimension value, the overall level of data quality is *enough*.

Remark that when the DMN table is analyzed, each decision rule must be unwound, by covering multiples cases for the input values. Thus, it might lead to conflicts between rules. In our example, it does happen since there are four cases covered by the first and the second rules. It means that data fulfilling any of these four conditions might be assessed as either *adequate* or *suitable*. Here is where the hit policy comes to play. In our case study, the policy is *First*, it means that rules might overlap between them and, in the case it occurs, the first hit by rule order is returned. Applied to this scenario, it means that any of the rules exposed would be assessed as *adequate*.

4.3 Results of the Data Quality Assessment

Table 3 shows the results obtained from applying the DMN decision table proposed in Sect. 4 to the tuples of Table 2 (i.e., the motivating example presented in Sect. 3).

Next, we explain how these results have been obtained:

- *Completeness*. The first and fourth tuple are *non-complete*. The first one violates *BR.06* while the fourth one violates *BR.08*.
- *Accuracy*. The first tuple is *accurate* due to the violation of *BR.04*, while the second tuple is *very accurate* because it fulfill all rules. The third and fourth tuples violate *BR.02*, causing them to be valuated as *Inaccurate*.
- *Consistency*. All tuples except the second one are labeled as *low consistency* because they violate *BR.04*.
- *Data Quality*. The first, third and fourth tuples are labeled as *non-usable* because their consistency is low. However, the second tuple is labeled as *Suitable* since its *Completeness* is *complete*, its *Accuracy* is *inaccurate*, and its *Consistency* is *low consistency*.

Table 3. Results of data quality assessment.

Id	Completeness	Accuracy	Consistency	Data quality assessment
1	non-Complete	accurate	low consistency	non-usable
2	complete	very accurate	high consistency	suitable
3	complete	inaccurate	low consistency	non-usable
4	non-Complete	inaccurate	low consistency	non-usable

5 Related Work

Data quality has been considered a key topic in many contexts, what makes that many researchers and practitioners have developed their own data quality models and the underlying assessment methods [1]. Although the idea of data quality dimensions have been widely studied and proposed through literature, only few authors as [9] or [18] have published specific implementable measurement methods for the various data quality management initiative. Many practitioners claims guides for a sound interpretation of the results of data profiling tools to better identify root cause of the problems. In fact, as [21] stated, data profiling are not explicitly presented associated to the idea of data quality dimensions but to the idea of data quality errors.

On the other hand, it is necessary to understand the necessity of having available mechanisms to determine almost in execution-time of the instance of a business process, if a given record has quality enough to be usable for the task at hand. So profiling the whole dataset as a way to assess the level of data quality is not useful for us. Therefore, we need to integrate the data quality assessment into the running instance of the business processes, and desirably, enable this assessment to be done automatically. We observed that many data quality analysts knew how to describe by means of business rules whether a record of data was usable or not.

However, to the best of our knowledge, there are not published similar approaches to the one we presented in this paper. Our proposal establishes a semi-automatically rules-guided waterfall cycle for assessing the usability of individual records in a datasets during the execution of the instance of business processes. The very nature of this approach suggest the use of DMN to implement the measurement methods to make a decision on the use of data. DMN has been used to represent business rules facilitating the understanding and description of business rules [4,5]. DMN is frequently used into BPMN [14] models by means of decision tasks that incorporate the set of rules that must be evaluated during instantiation time [11]. Moreover, some DMN extensions let the integration of the decision making process non-only incorporating dataflow [17]. The necessity to incorporate data quality measurement in business processes was early identified in [16]. However, to the best of our knowledge, there is no solution that use DMN as a mechanism to model and evaluate data quality assessment and measurement requirements and it is still a challenge how bring the gap between the human description of data quality requirement description, with and automatic data quality assessment and measurement.

6 Conclusions

The inclusion of data quality assessment requirements in business process helps organizations to make more reliable decisions on the use of data. We have introduced in the paper the application of DMN (Decision Model and Notation) standard with the aim of facilitating the assessment of data quality requirements.

Thanks to the use of DMN, the automation of the evaluation of the data quality level ceases to be a theoretical contribution and becomes a reality. On the one hand, business experts can easily include their knowledge in DMN since is a common notation readily understandable. On the other hand, DMN facilitates the inclusion of the business rules for data quality measurement and assessment as part of the decision process and feeds the process with the information related to data quality.

Acknowledge. This work has been partially funded by the Ministry of Science and Technology of Spain ECLIPSE (RTI2018-094283-B-C33) and (RTI2018-094283-B-C31) projects, the Junta de Andalucía via the PIRAMIDE and METAMORFOSIS projects, the European Regional Development Fund (ERDF/FEDER), GEMA: Generation and Evaluation of Models for dAta Quality (Ref.: SBPLY/17/180501/000293), and the Cátedra de Telefónica "Inteligencia en la Red" of the Universidad de Sevilla.

References

1. Batini, C., Cappiello, C., Francalanci, C., Maurino, A.: Methodologies for data quality assessment and improvement. ACM Comput. Surv. (CSUR) **41**(3), 16 (2009)
2. Batini, C., Scannapieco, M.: Data Quality: Concepts, Methodologies and Tea niques. Data-Centric Systems and Applications. Springer, New York (2006). https://doi.org/10.1007/3-540-33173-5
3. Ceravolo, P., et al.: Big data semantics. J. Data Semant. **7**(2), 65–85 (2018)
4. Dangarska, Z., Figl, K., Mendling, J.: An explorative analysis of the notational characteristics of the decision model and notation (DMN). In: 20th IEEE International Enterprise Distributed Object Computing Workshop, EDOC Workshops 2016, Vienna, Austria, 5–9 September 2016, pp. 1–9 (2016)
5. Dasseville, I., Janssens, L., Janssens, G., Vanthienen, J., Denecker, M.: Combining DMN and the knowledge base paradigm for flexible decision enactment. In: Supplementary Proceedings of the RuleML 2016 Challenge, 10th International Web Rule Symposium, RuleML 2016, New York, USA, 6–9 July 2016 (2016). http://ceur-ws.org/Vol-1620/paper3.pdf
6. Figl, K., Mendling, J., Tokdemir, G., Vanthienen, J.: What we know and what we do not know about DMN. Enterp. Model. Inf. Syst. Arch. **13**(2), 1–16 (2018)
7. Hasic, F., Craemer, A.D., Hegge, T., Magala, G., Vanthienen, J.: Measuring the complexity of DMN decision models. In: Business Process Management Workshops - BPM 2018 International Workshops, Sydney, NSW, Australia, 9–14 September 2018, Revised Papers, pp. 514–526 (2018)
8. Hasić, F., De Smedt, J., Vanthienen, J.: Towards assessing the theoretical complexity of the decision model and notation (DMN). In: CEUR Workshop Proceedings, vol. 1859, pp. 64–71. CEUR Workshop Proceedings (2017). https://lirias.kuleuven.be/1548429?limo=0
9. Heinrich, B., Klier, M.: Metric-based data quality assessment - developing and evaluating a probability-based currency metric. Decis. Support. Syst. **72**, 82–96 (2015)
10. ISO-25012: Iso/IEC 25012: Software engineering-software product quality requirements and evaluation (square)-data quality model (2008)

11. Janssens, L., Bazhenova, E., Smedt, J.D., Vanthienen, J., Denecker, M.: Consistent integration of decision (DMN) and process (BPMN) models. In: Proceedings of the CAiSE 2016 Forum, at the 28th International Conference on Advanced Information Systems Engineering (CAiSE 2016), Ljubljana, Slovenia, 13–17 June 2016, pp. 121–128 (2016). http://ceur-ws.org/Vol-1612/paper16.pdf
12. Lee, S., Ludäscher, B., Glavic, B.: PUG: a framework and practical implementation for why and why-not provenance. VLDB J. **28**(1), 47–71 (2019)
13. Loshin, D.: The Practitioner's Guide to Data Quality Improvement. Elsevier, Amsterdam (2010)
14. OMG: Business process model and notation (2017). http://www.omg.org/spec/BPMN/2.0
15. OMG: Decision Model and Notation (DMN), Version 1.2, January 2019. https://www.omg.org/spec/DMN
16. Parody, L., Gómez-López, M.T., Bermejo, I., Caballero, I., Gasca, R.M., Piattini, M.: PAIS-DQ: extending process-aware information systems to support data quality in PAIS life-cycle. In: Tenth IEEE International Conference on Research Challenges in Information Science, RCIS 2016, Grenoble, France, 1–3 June 2016, pp. 1–12 (2016)
17. Pérez-Álvarez, J.M., Gómez-López, M.T., Parody, L., Gasca, R.M.: Process instance query language to include process performance indicators in DMN. In: 20th IEEE International Enterprise Distributed Object Computing Workshop, EDOC Workshops 2016, Vienna, Austria, 5–9 September 2016, pp. 1–8 (2016)
18. Sebastian-Coleman, L.: Measuring data quality for ongoing improvement: a dataquality assessment framework. Newnes (2012)
19. Wang, R.Y.: A product perspective on total data quality management. Commun. ACM **41**(2), 58–65 (1998)
20. Wang, R.Y., Reddy, M.P., Kon, H.B.: Toward quality data: an attribute-based approach. Decis. Support. Syst. **13**(3–4), 349–372 (1995)
21. Woodall, P., Oberhofer, M., Borek, A.: A classification of data quality assessment and improvement methods. Int. J. Inf. Qual. **3**(4), 298–321 (2014)