ATLANTIS PRESS
Advances in Intelligent Systems Research, volume 174
Proceedings of the 8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020)

Data Quality and Digital Twins in Decision Support Systems of Oil and Gas Companies

Tatyana Pavlovich*
Information Technoogies Department
Russian Technological UniversityMIREA
Moscow, Russia
tv_pavlovich@mail.ru

Elena Dron
Applied Informatics Department
Ufa State Aviation Technical University
Ufa, Russia
elena_dron@bk.ru

***Abstract*—In the modern world, most enterprises implement digital technologies to improve business performance and enhance their competitive advantages. All digital technologies are based on data. Global digitalization leads to huge amounts of data. It is impossible to talk about IT development without understanding the nature of data and its quality management technologies. The article presents the approach to data quality management in the Digital Twins design and use. Digital Twins or Smart Fields are decision support systems to manage field development. Such decision support systems are designed to address the issues of maximizing oil recovery ratio with optimal CAPEX and OPEX. Moreover, decision support systems should include data quality management algorithms. The lack of data quality control mechanisms can lead to incorrect decisions and large losses for the business. The approach presented in the article contains 7 steps of the field management cycle using Digital Twins. Some of these seven steps are related to data quality verification. It is also shown that data with minuses or disadvantages allow to identify problematic areas of field development. For the identified problematic areas, additional data is collected to eliminate disadvantages or minuses. As a result, it is possible to clarify the real "Why" or the reasons of the critical deviations from the target situation and KPIs. The work done showed that designing decision support systems such as Digital Doubles, Digital Fields and Smart Wells without data quality management is tantamount to creating an engine without taking into account the fuel on which it runs. The application of the developed approaches to data quality management in Digital Doubles almost by a third increased their economic efficiency in the development of oil births and, as a result, their investment attractiveness.**

Keywords—data quality management, Smart fields, Smart wells, Digital twins, Digital transformation

I. Introduction

According to Big explanatory dictionary of the Russian language, data is information and indicators that characterize someone or something as a basis for conclusions, decisions and assessments.

As known, the basis of any information system is data that is converted into graphs, reports. Based on the generated reports, key business decisions are made. At the same time, it is impossible to make the best decisions based on incorrect data. If the data is unreliable, contains errors or other disadvantages or minuses, then the decisions made can be disastrous for business [1].

Last few years Digital Twins have appeared. Digital Twins are digital copies of enterprises, business processes or individual physical objects. A Digital Twin is a tool for resolving business performance management issues, choosing the best ways for enterprise development to ensure the survival and well-being of companies in highly competitive markets.

To perform their functions, Digital Twins are equipped with the most advanced technologies - IoT, AI, machine learning, Big Data and much more. At the same time, each of the technologies listed above is based on data. Data quality directly determines the value of these technologies and, as a result, the quality of a Digital Twin, and business decisions quality.

Data quality is a key issue in decision support systems. The problem of data quality is not new, but still relevant. The relevance is due to the fact that the implementation of Corporate Information Systems, digital technologies and Digital twins does not lead the business to the target results. The costs that companies incur in connection with digitalization do not lead to the expected effects, such as increased operational efficiency, increased competitiveness of products and, as a result, achievement of strategic goals.

This article presents a comprehensive approach to data quality management in Digital Twins to create an effective decision support system for oil and gas companies.

II. Data Quality Management

According to the standard ISO 9000: 2015, the main criteria for data quality are their completeness, reliability, accuracy, consistency, accessibility and timeliness. First, it is necessary to formulate criteria for assessing the quality of data in order to understand which data are qualitative and which are not. To do this, you can identify 5 basic ways, the use of which reduces the quality of the data. The table below provides a list of these methods that were identified during the project to manage the quality of the source data and its impact on the investment attractiveness of the field of the ex. TNK-BP (see table 1).

.

TABLE I. WAYS TO REDUCE DATA QUALITY

№	Ways to reduce data quality	Examples
1	Miss a fact	The flow rate was decreased last week due to the shutdown of pumps. The fact that last time was problems with electricity has been missed. This omission gives the impression that the pumps are broken.
2	Distort a sequence of events	In order to intensify oil production and increase the interval between wells' repairs a number of geological and technical measures was taken. After carrying out these measures, their economic efficiency was calculated. Economic analysis of geological and technical measures showed a negative result. Obviously, by changing the sequence of events, the situation looks extremely illogical.
3	Do not indicate time	A number of wells failed due to breakdowns of submersible equipment. Such a message can create prerequisites for urgent repairs of submersible equipment if you do not know that this message dates from last year.
4	Add false information	The pressure of gas supply to the main gas pipeline was different from the established norms because the operator changed the operating mode of the booster compressor station (BCS). The information that the operator changed the BCS operating model is false. Because of it, this operator looks incompetent.
5	Change importance	The oil recovery coefficient (CIN) of the asset was decreased. Therefore, it is necessary to check the reliability of the ground infrastructure of this field. That is an order. Actually, this was just a recommendation.

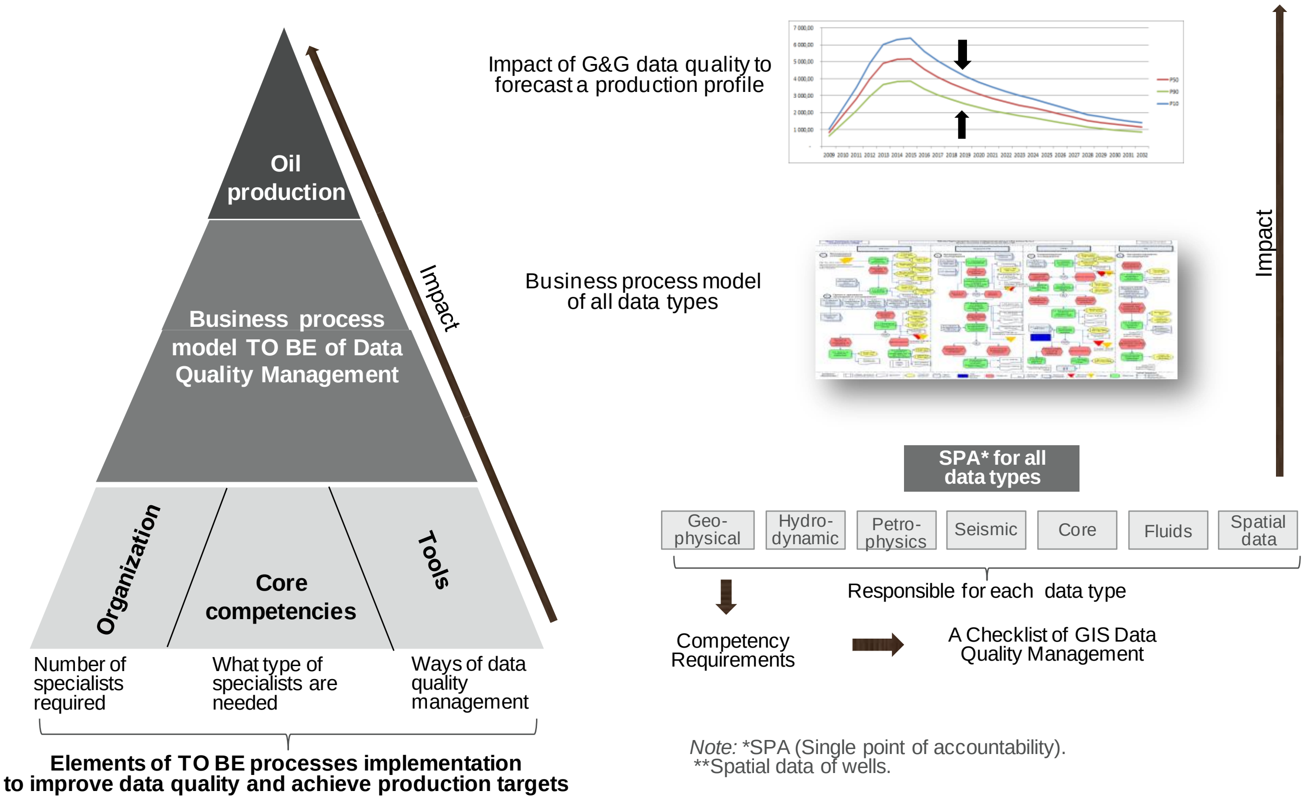


Fig. 1 - The key concept of the data quality management methodology

These are the main reasons why the manager has a misconception about the situation. There are hundreds of ways that these 5 types of data misuse can make up a completely false impression of a business situation. Decisions, actions or orders taken based on data containing any of the above in the table lead to errors and negative consequences for the business. There are a huge number of combinations of these methods or errors. The same message or report may contain several types of them or all 5 of them.

Therefore, management decisions depend on data. Using data containing one or more of the above errors can lead to incorrect decisions and large losses for the enterprise. An unreasonable or impracticable business decision or order is based on an illogical conclusion, which is illogical because it contains one or more of these errors (see the table above) [2]

A methodology for data quality assessment and management was developed for ex. TNK-BP. The key idea of this methodology is presented in the fig. 1.

First of all, it was decided to develop a data classification. This data create the impact to production profile. Seven types of data were identified that are obtained during the following operations:

- GIS - geophysical surveys of wells;
- well test - hydrodynamic studies of wells;
- petrophysical studies;
- seismic surveys;
- core research;
- fluid analysis;
- spatial data of wells.

In the future, business processes for collecting, storing and processing were formalized for each data type. A single responsible person (SPA) was identified for the relevant business processes. Competency requirements for single responsible persons were developed and a tool for data quality control was proposed.

This tool is a checklist for single responsible persons, which includes a check of 5 factors. These factors were defined as the opposite of data errors and data minuses. The fig. 2 below shows a checklist for checking data quality.

A single responsible person compares the data obtained with five factors as shown in the fig. 2. As a result of the audit, it may turn out that one or more factors are missing. In this case, the missing factors are clarified. Only after clarification of each factor does the single responsible person have the authority to transfer this verified data to the higher management for making business decisions. In this case, it is believed that the data correspond to the required level of quality and the manager will be able to clarify the situation and come to logical conclusions.

This approach was recorded in the regulations for data quality management. Using this approach increased the accuracy of the forecast of the production profile by 30%, and also ensured the achievement of the planned result when carrying out measures to reduce the cost of oil production.

The developed approach also increased the accuracy of exploratory drilling and, as a result, the effectiveness of capital investments in geological exploration. Within 2 years after this approach implementation the Return On Investments (ROI) increased by 15-20% for West Siberia assets of ex. TNK-BP.

Consequently, the quality of geological and geophysical data has a direct impact on the investment attractiveness of oil and gas assets.

III. Digital Twins as a decision support tool

The oil industry began to use Smart Fields technologies the 2000s. Ex. It has been actively implementing Smart fields for West Siberia assets to slow down the production decline, as well as to decrease capital costs and operating expenses.

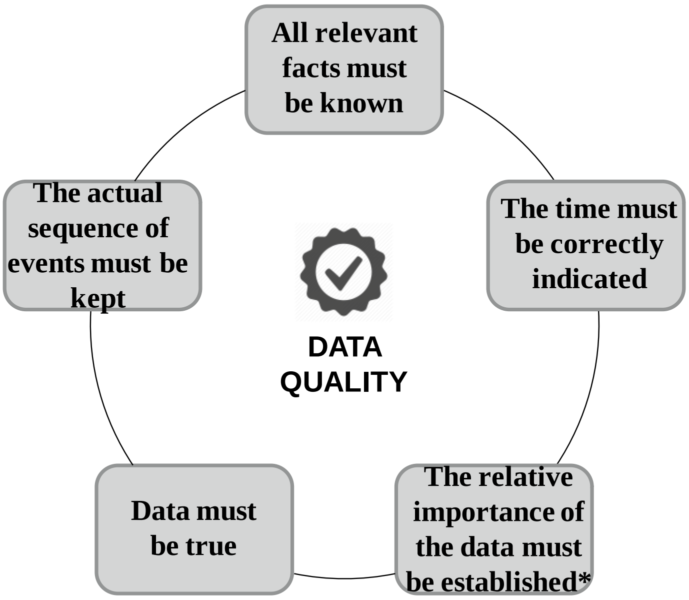
The main component of Smart Fields is Smart Wells. Smart Wells are designed to provide maximum of a flow rate at minimum cost. Smart Wells were equipped with telemetry systems that transmit real-time data on the operation of submersible equipment. Based on the data obtained, responsible specialists determined the best modes of reservoir development and the reliability of the submersible equipment.

The use of Smart wells increased the accuracy of forecasting well production by 15-20%, taking into account the changing production conditions. Additionally, virtual testing of well operation modes allows avoiding experiments with real objects. It is also allows specialists to create several scenarios and choose the best one regarding the given criteria. At the same time, in a number of cases, serious errors of oil production rate forecasts and the choice of submersible equipment operating modes were observed. These errors increased operating costs and lifting costs, costs of geological and technical measures and for repair and maintenance of equipment.

An analysis of the causes of these errors indicated problems with data quality. Data that does not correspond to at least one of the five elements of data quality is substandard or we can say that these data contain minuses. Algorithms for verifying data received from Smart wells have become part of the Smart Fields information system. In other words, Smart Fields support the data quality management methodology and compare the data obtained from the wells with 5 elements [3].

At the same time, automated data estimation algorithms made it possible to identify data minuses. For example, 70 indicators or data were obtained regarding the overall picture of the production rate of the producing well. As a result of the implementation of quality control algorithms, it is found that 21 indicators or data contain minuses or errors, i.e. not consistent with all 5 elements. In some cases, these poor-quality data cannot be supplemented and clarified to bring them into line with 5 quality elements. The managers responsible for the development of the field in these situations could not assess the current situation and make the right decisions regarding the well operating modes.

All this created the prerequisites for the refinement of the methodology of data quality management in relation to digital counterparts and smart fields in order to increase the effectiveness of decision support.



Note: * the relative importance of the data is established by comparing the facts with the goal that the person is trying to achieve, or with the problem that he is trying to solve.

Fig. 2 - Five elements of data quality

The revised methodology was based on the following idea: the quality of data analysis depends on knowledge and understanding of the ideal picture of field development, as well as target goals (for example, to reduce the rate of production decline). This means that the responsible manager needs to know what the area of activity should look like, in this case, the development of the field from a rational and logical point of view [4].

If this condition is met, then the following two steps must be completed:

- data analysis;
- analysis of the current situation.

In the methodology of data quality management, a situation is a general picture with a certain amount of data. Data is facts, charts, statements, decisions, actions, descriptions that are offered as correct. Minuses are any individual data that is offered as true, but turns out to be illogical as a result of its comparison with 5 elements. Pluses are true data that turn out to be true as a result of comparing them with the same 5 elements.

The fig. 3 below shows the main steps of the developed methodology for data quality management.

At the first step, the data quality is evaluated by comparing it with 5 elements. As a result of this step, data containing minuses can be detected, i.e. substandard data quality.

The analysis of the current situation at the step 2 is that the data with minuses are sorted by field sections or by well components. The area with the most data with minuses becomes the main goal of the correction. For example, if a flow rate decreases at any well at an unplanned pace and most of the data with minuses are related to the operation of pumping equipment, then management should first pay attention to the pumps operation. In other words, situation analysis refers to the distribution of data with minuses among the units of the whole. As a result of the situation analysis, it becomes clear which area needs to be fixed and what is the reason for the failure to achieve the planned indicators and goals [5].

These steps (data analysis and analysis of the current situation) were integrated into the work of Smart Field. Thus, the created decision support system, at the first step, issued to the responsible of the field development person a list of data with minuses or errors.

As a result of the second step implementation, the responsible person is informed of the problematic area, which corresponds to the largest amount of data with minuses.

In terms of time saving, a one-day analysis of data and the current situation in the described way is equivalent to 3 months of work experience in the absence of this methodology [6].

Thus, the developed methodology of data quality management provides the result instantly. Achievement of the same result by gaining experience takes a long time.

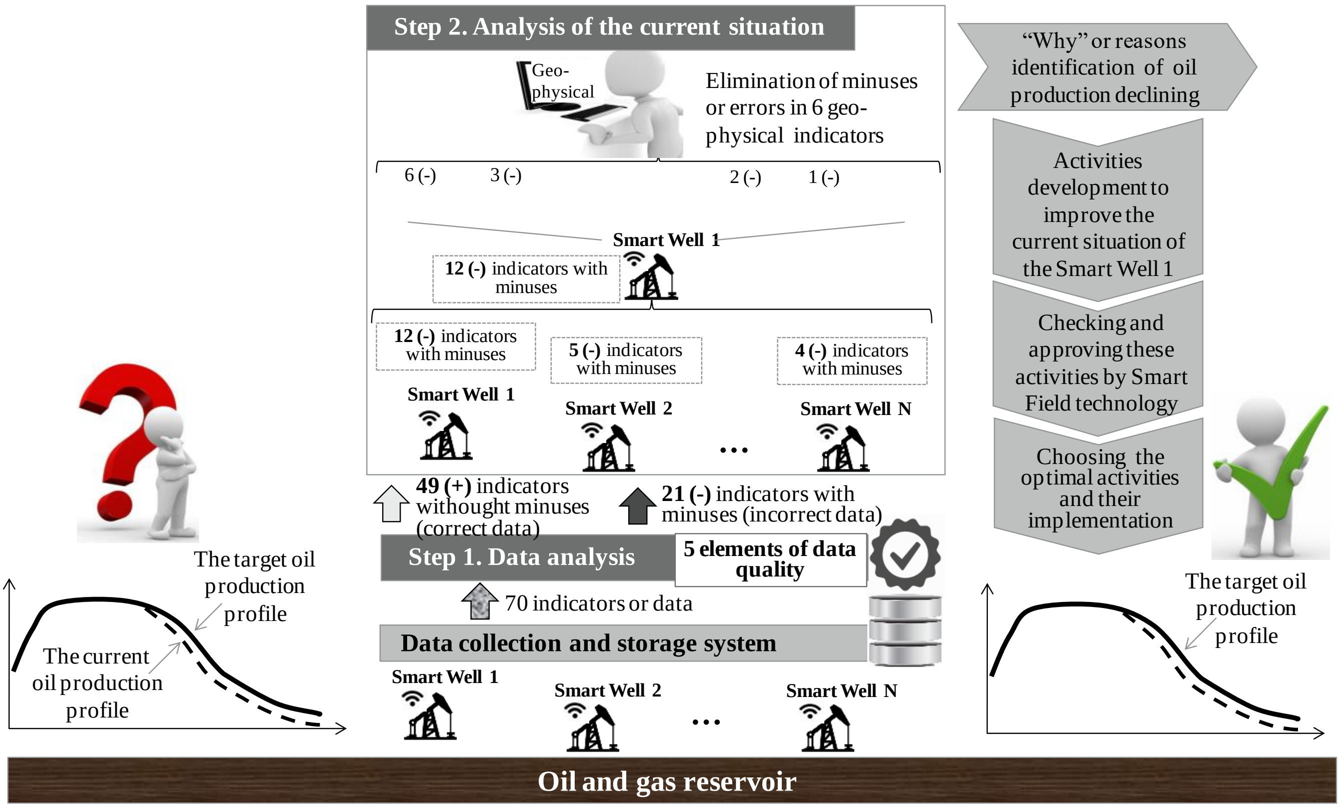


Fig. 3 - The key steps of data quality management methodology

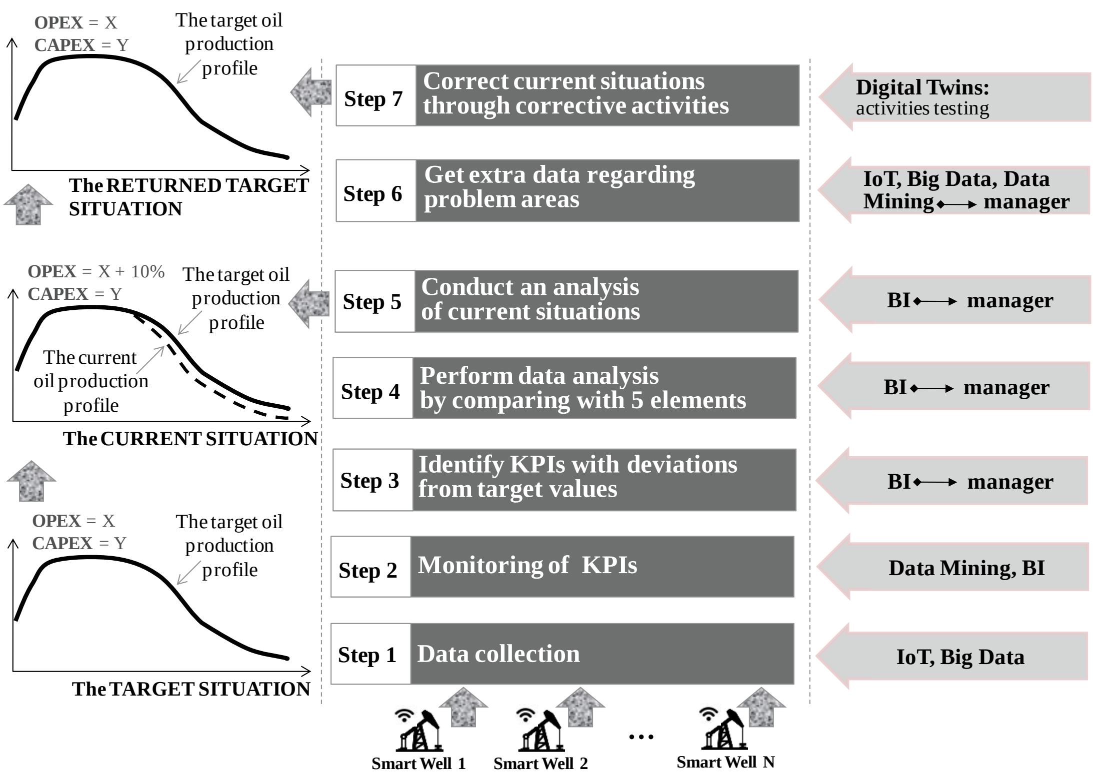


Fig. 4 - The field development cycle by Digital Twins using

IV. The Impact of Data Quality and Digital Twins on The Investment Attractiveness of Fields

As known, cars run on gasoline, electric motors run on electricity, and decision support systems like Digital Twins run on data. Using any mathematical formula, the answer obtained will never be more accurate than the data substituted in this formula.

No matter what digital technology is used - Big data, Data Mining, Artificial intelligence, neural networks, Business Intelligence, etc. - the quality of conclusions and the optimality of business decisions will directly depend on the quality of the data [7,8].

The developed methodology convincingly proved that the creation of Digital Twins, without taking into account the data quality management, is equivalent to an attempt to develop an engine without consideration of the required fuel.

The methodology for data quality management in Digital Twins includes 7 steps, which can be called the decision support management cycle [9]. These seven steps are shown in fig. 4.

In step 1, the target situations and goals of the field development are analyzed. At this step, data is also collected from Smart Wells by using IoT and Big Data technologies. Collected data are used to calculate key performance indicators (KPIs) of an oil and gas field development.

Step 2 is design to monitor the KPIs. This step is implemented by using Data Mining and Business Intelligence technologies.

The third step is determined by KPIs with deviations from the target values. This step uses Business Intelligence technology, which helps the field manager quickly identify KPIs.

Step 4 is to compare KPIs with deviations from target values with 5 data quality elements. This step is also implemented by using Business Intelligence technology, which compares data with 5 factors and determines KPIs with the largest number of minuses. The final choice is made by the manager, who selects KPIs to conduct a situation analysis.

At step 5, KPIs with the largest number of minuses, are distributed by field development areas. It is possible to identify the area with the biggest amount of KPIs with minuses. For this area is necessary to get additional data. This step is also performed by using BI technology.

The task of step 6 is to obtain data on problem areas of field development in order to correct the minuses in KPIs and obtain reliable data on the current situation. The objective of this step is to find the real "Why" or the reasons for the deviations from the target values of the indicators. Step 6 uses IoT, Big Data, and Data Mining technologies.

At the next stage, corrective measures are determined to correct the current situation and return to the target picture. Corrective measures are tested on the Digital Double of the field. As a result of testing, the optimal set and sequence of these activities is determined. The result of this stage is to achieve the target KPIs and return to the target state of affairs in the most optimal way. Also at step 7, a number of rules are developed to prevent similar situations in the future.

At the next step, special activities are determined to correct the current situation and return to the target picture. Corrective activities are tested by Digital Twin. As a result of testing, the optimal set and sequence of these activities is determined. The result of the 7th step is to achieve the target KPIs and return to the target state of the field development by the most optimal way. Also at step 7, a number of rules are developed to prevent similar situations in the future.

The using of the developed methodology by 20-30% reduced the uncertainty of the oil production forecast. Capital and operating expenses were reduced by 15-20%. The accuracy of the forecast of the production profile and the optimization of CAPEX and OPEX allowed us to reduce the discount rate, which contributed to an increase in the net present value (NPV) of field development. The increased NPV size contributed to increasing the investment attractiveness of the field development project, which was the purpose of the data quality management methodology to design and use Digital Twins.

The conducted work showed that designing decision support systems such as Digital Twins, Smart Fields and Smart Wells without data quality management is equal to creating an engine without taking into account the required fuel.

The practice of the developed approaches to data quality management in Digital Twins almost by a third increased their economic efficiency and, as a result, investment attractiveness.

V. Conclusion

Data quality is the cornerstone of creation and implementation of Digital Twins and decision support systems. Modern experts believe that data is the new oil of the 21st century. Like oil, data is the raw material for the production of a new technological product - knowledge. Only with the knowledge a manager can make the best business decisions [10, 11].

In modern realities, a transition from the generation of a huge amount of useless data to decision support systems is necessary. Digital Twins and decision support systems require high quality of the initial digital data. Unreliable data leads to incorrect decisions and nullifies the effect of using intelligent systems such as data mining, machine learning, predictive analytics, etc. Data quality management ensures the usefulness of digital data and the value of each bit.

The key success factor of companies' digital transformation is approaches that ensure the quality of corporate data. For this, companies are implementing new business processes, appointing responsible people and training personnel in data quality management, since data reliability, their availability and consistency in various IT systems require an integrated approach.

The project for ex. TNK-BP company convincingly proved that the initial data quality is the key success factor of Smart Fields and Digital Twins technologies, which ensures the effectiveness of decision-making systems and, as a result, the investment attractiveness of oil assets.

REFERENCES

[1] R. H. Bonczek, C. Holsapple, “Whinston Foundations of Decision Support Systems,“ New York: Academic Press, 1981.

[2] S. Sadiq, “Handbook of Data Quality. Research and Practice,” Springer-Verlag Berlin Heidelberg, 2013, pp. 15-73.

[3] O. I. Larichev, A. V. Petrovsky, “Decision support systems. Current state and prospects of their development,” Results of science and technology. Technical cybernetics, vol. 21. Moscow, 1987 pp. 131—164. http://www.raai.org/library/papers/Larichev/Larichev_ Petrovsky _1987.pdf

[4] A. D. Saraev, O. A. Scherbina, “System analysis and modern information technology,“ SONAT, 2006, pp. 47-59. http://matmodelling.pbnet.ru/Statya_Saraev_Shcherbina.pdf

[5] P. V. Terelyansky, “Decision Support Systems. Design experience,“ Volgograd State Technical University, Volgograd, p.127, 2009.

[6] S. L. Alter, “Decision support systems: current practice and continuing challenges,“ Mass.: Addison-Wesley Pub., 1980.

[7] G. G. Kulikov, E. A. Dron, “Formalization of planning models for organizational management in an enterprise,“ vol.16.№2, Chelyabinsk: Vestnik UGU, 2016, pp. 91-100.

[8] G. G. Kulikov, E. A. Dron, J. O. Bagaeva, M. A. Shilina, “Automated information economic systems,“ Ufa: USATU, 2013. pp. 180.

[9] C. Batini, M. Scannapieco, “Data and Information Quality. Dimensions, Principles and Techniques,” Springer International Publishing, Switzerland, 2016, pp. 309-402.

[10] D. Loshin, “The Practitioner's Guide to Data Quality Improvement,” Morgan Kaufmann, the 1st edition, 2010, pp.73-165.

[11] T.C. Redman, “Data Driven: Profiting from Your Most Important Business Asset,” Harvard Business Press, 2008, pp. 35-89.