Preliminary Study on Problem Identification, Cause Analysis and Countermeasures of Data Application in Engineering Testing Industry

Kun Qian and Jie Yao
Tongji University, 1239 Siping Road, Yangpu District, Shanghai, China
E-mail: 1941796@tongji.edu.cn

Abstract. As a key node of engineering quality assurance, the engineering testing industry is obviously lagging behind other industries in its data application aspect, and there are various problems in policy, technology, and resources. Such problems have obviously affected the development trend of the entire industry, and it is urgent to solve the data problems in the engineering testing industry. By combing and analysing the status quo of data application in the industry, the article found three common problems of data application, no data available, inability to apply and no system management. These common problems directly affect the data application level of the entire industry. Based on the identified problems, this article further analyze the causes of the problems and point out the general direction of the industry data governance strategies.

1. Research background
With the rapid development of Chinese economy, the number of construction projects has increased rapidly. According to the data published in the "China Construction Industry Statistical Yearbook-2018", at the end of 2017, there were 88,074 construction companies in China, with 55.296 million people in the construction industry, and a total output value of 21,394.456 billion yuan, a year-on-year increase of 10.52%.[1]

In order to further ensure the quality of the project, especially from the perspective of the quality of materials and structure function, the engineering testing industry has developed rapidly in recent years and has played an increasingly critical role in ensuring the project.

In engineering testing industry, data and its management is critical. The essence of testing is to "speak with data." Whether it is commissioned data, testing original records, or test report, it is "data" that can ultimately explain the problem. The data reflects the real state of the inspected object through different forms such as numbers, curves, images and video. In addition to the testing data, there are also various types of business data, such as the project information, the contract-signing documents, the status of the equipment, the personnel information, the project progress information, and the cost settlement information. Although these data with partial management functions are not directly related to the quality of testing, the management quality of such data directly affects the operational efficiency of the business process of the enterprise. Therefore, data is the core of a testing company, and business data and testing data have an equally important position in the company.

At present, the level of data applying in the engineering testing industry is lagging behind, and there have long been situations such as single expression forms of testing results, poor data application and processing capabilities, unclear division of responsibilities, lack of supervision by competent
authorities, and low data security levels. The industry needs to find a systematic method to better improve the overall data application level of the industry.

Data governance focuses on the management of the data assets of the enterprise. From the perspectives of strategic development, procedures and policies, corporate culture, and technology use, it uses a variety of means to ensure data quality, improve data security and availability, and maximize the value of data. Improve corporate data quality and business efficiency, and then serve the company's strategic development.

In recent years, the research on data governance has become a hot spot, and many theoretical and application results have also been produced[2]. There have been many successful cases in the fields of finance, IT[3], education, etc., but the research in the field of engineering testing is almost blank. Therefore, it is a good attempt to analyze the data application problems encountered by the engineering testing industry from the perspective of data governance.

2. Problem Identification and causes analysis of data application in the engineering testing industry

2.1. Identification of data application problems

From the perspective of implementation, the current main problems of data governance in the engineering testing industry include three aspects: no data available, inability to apply, and lack of system management.

2.1.1. No data available

At present, most of the business data in the industry, such as contract information and entrusted information, are electronicized, mainly in the form of electric tables. These tables are modified and maintained by different person according to the business process, and information is shared through network sharing. Data integration is more difficult.

The testing data, especially the original testing records, are still recorded manually, and the circulation of paper documents is the main form. Some original testing records are manually typed into the computer to generate testing reports. A large number of key data has not been digitized, and there are errors and omissions in the testing data. The phenomenon of writing, or even random alteration, makes poor data consistency level. In addition, due to the restrictions of administration, the test report should be submitted to the client as in paper form, which brought great inconvenience to data collection and data management.

In addition, the lack of correlation between the business process links, the one-way flow of business data, the lack of timely and effective feedback, and the lack of business status data result in low communication efficiency between business links and affect the overall business progress.

Through the above analysis and industry research, the current status of data availability in the engineering testing industry is summarized as follows:

| No. | Data(Business type)                        | Availability                       |
|-----|-------------------------------------------|-------------------------------------|
| 1   | Data Storage                              | Electric Tables                     |
| 2   | Data Usage                                | Local Area Network Sharing          |
| 3   | Data Collection                           | Manual                              |
| 4   | Data Integrity and consistency            | Average                             |
| 5   | Form of Test Report                       | Paper                               |
| 6   | Business Status Query                     | Artificial                          |
| 7   | Business Requirements Adjustment          | Lack of Communication               |

2.1.2. Inability to apply

Detection data includes structured data, semi-structured data and unstructured data, which require high data application capabilities. For existing data, due to the lack of technical ability to integrate all data
from the entire business level, all kinds of business data are scattered in company servers, or personal computers. Data query is still in the primitive stage of searching, filtering, and counting manually in different worksheets. Although some of the information platforms launched by administrations have made some attempts in data integration and utilization, they are still far from being practical.

As a labor-intensive industry, the engineering testing industry has low information literacy and poor data awareness. In fact, it has already had a certain impact on the production of the industry. According to the report issued by the Shanghai Construction Engineering Testing Industry Association, at the end of 2017, there were 8312 employees in Shanghai construction testing industry, with an average age of 36.2 years. Among them, 10.09% have a master's degree, and a bachelor's degree and a junior college degree each accounted for about 30%. Persons with higher academic qualifications and professional titles are mainly in management positions, and the level of overall culture and professional skills of frontline employees is relatively low.

From the perspective of testing, there are still a large number of equipment that still rely on manual operation and manual readings, and the automatic collection of detection raw data still has a lot of room for development. As of the end of 2016, the total number of instruments and equipment in Shanghai's construction testing industry was about 40,000, of which the automatic data collection rate was only about 15%, and a large amount of reading work was still done manually. In addition, software such as bridge monitoring systems requires secondary development and modeling for specific projects, which require higher theoretical foundations and professional knowledge capabilities of practitioners, and higher requirements for the computing capabilities of testing companies.

2.1.3. No system management
The management model of engineering testing companies is still relatively traditional, and the editing of management systems focuses on various process operating procedures, quality management requirements, employee code of conduct, equipment operating procedures, etc. The use of information technology and data is only used as a means to assist management, not as a tool to truly improve productivity. The internal control system of the leading companies in the industry has become systematic, but it has not yet been able to meet the requirements of corporate management in the data age.

2.2. Causes analysis of data application

2.2.1. Causes of data availability
Jianzhong Li and others subdivided data availability into five properties: consistency, accuracy, completeness, timeliness, and entity identity. The first four properties of them can be used to analyze the reasons for data availability in the engineering testing industry.

(1) Consistency: Consistency means there should be no obvious contradictions in the content of each field describing the same fact. Typical data consistency problems may include information mismatch, content not included and inconsistent agreements. The data consistency problems that usually exist in the engineering testing industry are summarized in the table.

| No. | Problem Description      | Example                                                                 |
|-----|--------------------------|-------------------------------------------------------------------------|
| 1   | Unmatched information    | The nominal construction unit is A, but the actual construction unit is B. |
| 2   | Content not included     | The client specified the testing parameters and standards, but the actual standards did not include the testing parameter. |
| 3   | Inconsistent agreement   | The discount rate agreed in the contract is inconsistent with the settlement discount rate. |
| 4   | Irregular name           | The difference between the standard name of the sample and the popular name. |
| 5   | Synchronization is not timely | The general contractor has registered a set of samples on the |
(2) Accuracy: It means that the content contained in information can accurately express its meaning. Accuracy can be divided into two aspects, descriptive accuracy and date accuracy. Descriptive accuracy refers to the fact that the test sample information expressed in written form and commissioned information is as consistent as possible with objective facts. For example, the name of the production unit is not abbreviated; the accuracy of the data mainly refers to the error of the test data. For example, the test value of the compressive strength 59.491MPa, although it is reasonable to record it as 59.5Mpa according to the requirements of the specification, but it is not accurate. Another example is the vertical permeability coefficient test of geotextile, and the water flow should be accurate to 0.1ml. As part of data availability, the more accurate the data availability, the better.

(3) Completeness: Refers to the data set containing enough information to support the query and calculation required by the business. For the testing industry, the data set contains two types information: basic information and additional information. Basic information refers to the information necessary to complete a certain test and issue a test report. Additional information refers to information that does not affect the progress of the testing work, but is helpful for a better understanding of the testing work. Meanwhile, it should be noted that paying attention to data integrity does not mean that more additional information is better. Overly redundant information not only does not help the quality of business, but will also increase the pressure of data storage, queries, calculations and other applications.

(4) Timeliness: Mainly refers to that each information in the collection is valid and the latest version, for example, the update of testing specifications and evaluation standards. For the testing industry, the use of outdated standards will have serious consequences.

2.2.2. Data application capability analysis
For the engineering testing industry, data application capabilities can be summarized into four aspects: the ability of data planning, data collection, data integration, and data analysis and decision-making [4].

(1) The ability of data planning
Data planning ability refers to the ability of the enterprise planning and designing its data strategy, including details such as basic data framework designing, data system establishment and the person responsible for the data, and the formulation of data coding. The level of data planning ability is the embodiment of the basic skills of an enterprise and the embodiment of the enterprise's data literacy. A well-planned data structure will bring great convenience to data applications. Since most of the engineering testing companies are traditional enterprises, the leadership has very few clear plans for data usage and development prospects. Data planning capabilities are often missing.

(2) The ability of data collection
Data collection ability means that testing companies obtain data directly related to the testing business from various channels. Generally include project background data, commissioned information, sample information, circulation information, original test records, etc. There are various methods to obtain data, including Internet collection, filling in forms, instrument collection and data collection.

(3) The ability of data integration
Data integration not only forms a new, unified data source through the process of sorting, cleaning, and converting data collected, which is convenient for further use. Data integration is generally divided into in-system data integration and out-system data integration. In-system integration mainly includes process integration and report integration. Integration outside the system is the integration among systems and external equipment data, such as attendance data (with attendance equipment), testing data (with non-fixed testing equipment), financial data (with financial system), mails, etc.
At present, the testing industry has various data sources, complex data formats, and poor integration and relevance of information. Although a certain amount of data has been accumulated through manual entry, there is almost no data storage management. Data query and access are extremely inconvenient. To obtain specific information, business personnel need to query multiple different data tables, and the data cannot be directly related.

As the detection business increases yearly, there are more data operators. If the scattered intermediate business processing links still rely on manual processing for circulation, they will not be able to keep up with the needs of business development. Therefore, it is necessary to use tools to integrate the data of various business links, integrate and refine the scattered data into useful information, and free the labor force from repeated inquiries. Thus lay the foundation for the next step to find out suitable tools (such as ERP or data mining algorithms) to use data effectively.

(4) The ability of data analysis and decision-making

Data analysis refers to the use of appropriate methods to analyze and refine the collected data to obtain information and conclusions.

Usually, data becomes cold data after the test report is issued. The data is stored in the computer and archives, almost never recalled. The industry generally lacks applications and motivation to analyze and reuse data. The fundamental reason is that at this stage, all kinds of data are still mainly stored in paper, supplemented by electronic form records. Non-electronic data makes data analysis impossible. In addition, data analysis and decision-making have higher requirements for analysts. Not only must they be familiar with the company's various businesses, but also have a certain degree of testing practices, so as to integrate business data and testing data to carry out data analysis.

Among the four abilities of the above data application, data planning belongs to the construction of the upper-level architecture; data collection and data integration take up more time and resources, are difficult to execute, and produce lower value; data analysis and data decision-making take up less resources, and operation is most difficult, also value is the highest. Therefore, in the application of data, tools should be used as much as possible to realize the automatic operation of data collection and integration. At the same time, it trains qualified data analysts and develops data analysis tools that meet business needs.

Through the above analysis, it can be seen that the four types of abilities required for data use are still generally lacking in the engineering testing industry. It is necessary to formulate countermeasures through software and hardware to improve data use abilities.

2.2.3. Analysis of data system management

The management system is the specification of certain management mechanisms, management principles, management methods, and the establishment of management institutions. Standardized management is an arduous task that requires continuous improvement. A good management system is the guarantee for the robust and sustainable development of an enterprise. As a part of the enterprise management system, the data application system mainly clarifies which data can be used, how to use the data, and who can use it.

For the engineering testing industry, data application system management can be divided into three categories: organization management, data quality management, and data life cycle management.

(1) Organization is the division of personnel within an enterprise. By dividing the organizational structure, the powers and obligations are allocated and determined, and through appropriate constraints, different departments can maintain relative independence and work together for the company's strategic goals. As data governance issues receive more attention, data use and management departments will also have an impact on traditional corporate organizations. Data governance is concerned with the setting of decision-making power and execution power on data-related issues, including data security, data quality, and data utilization. No matter how it is set up, the organizational structure of an enterprise must be a combination of enterprise strategy, business needs and objective capabilities.
(2) Data quality management refers to the activities of identifying, analyzing, monitoring and solving quality problems that may occur at each stage of the data life cycle. The value of data must be based on data quality. Data quality work will inevitably be affected by the internal political and business factors of the enterprise, and requires the participation of multiple departments. Therefore, when formulating a management system, data quality should be combined with the overall business of the organization, and the implementation of the organization's business strategy should be coordinated to achieve the company's strategic goals. The implementation of data quality management is usually carried out from several aspects such as business requirements, technical realization, work flow and job responsibilities.

(3) Data life cycle management is the control and management of each node during the whole process of data collection, storage, integration, use and deletion. At each node, there should be a corresponding management system for clarity. For example, data collection specifications should clarify the overall policy of data collection, whether to collect as much data as possible or to collect data for the purpose of business development, which data should be collected, what standards should be followed when collecting data, and what methods should be used for collection. How the data is stored on the platform, how the customer’s privacy and confidential information are protected, etc. To sum up, by formulating corresponding systems, establishing a data governance organizational structure, clarifying staffing and job responsibilities, so that the scope of powers and responsibilities between various business departments and testing departments can be clarified, and the work can be carried out organically; the data storage management and importation should be clarified through the system Relevant regulations on modification, extraction and distribution, data transmission management, data storage methods, and data security, and reminders of risks in the full life cycle of data are tasks that need to be completed in the process of establishing and improving the data management system in the engineering testing industry.

2.3. Formulation of Data Governance Strategy

At different stages of management practice, there will be corresponding informatization measures to support what kind of data governance content there is. Carrying out data governance purely in terms of policies and systems is tantamount to talking on paper. The specific content of data governance can be explained from different aspects in the governance system.

Data governance strategies at different stages can be formulated from the following aspects:

1) Identification and cause analysis: Before implementing data governance, the existing data problems and business problems in the company or industry should be identified, and the identified problems should be analyzed to obtain the causes of the problems and the key crux of the problem. And then specify short-term governance goals and long-term governance plans. In the formulation of goals, for different governance goals, you can consider giving quantitative evaluation indicators, or benchmarking similar cases for supervision and evaluation. The completion and effect of data governance should be continuously optimized.

2) Organizational structure: Data governance cannot be done solely by network information departments or information specialists. Data governance is a strategic action that should also include the company's administration level, business departments, and third-party software and hardware service providers. The organizational structure of data governance participants can be designed from the three levels of strategy, strategy, and practice. Also, the job responsibilities of each post can be clarified to improve the efficiency and effectiveness of data governance.

3) Infrastructure: The physical aspect is the foundation of data governance. The implementation of data governance will greatly increase the requirements for hardware infrastructure, such as computer processing capabilities, data storage equipment, network access equipment, network security equipment, etc.

4) Data management: Data management mainly includes the organization and management of data, including data standard management, data quality evaluation, data flow design, data authority setting,
data security assurance, master data and metadata management, etc. These elements are related to each other. Centralized and unified management makes the work of data governance clear and rules-based.

5) Data governance technology and tools: Data governance technology supports the entire data governance process and directly affects the overall strategy of data governance. The goal of data governance is to obtain higher data quality and convenient data usage, and ultimately contribute to the improvement of corporate performance. Therefore, it is necessary to take into account factors such as data characteristics, business characteristics, and personnel composition. Choose appropriate data cleaning technology, data storage technology, data information standards, and data security technology.

6) Evaluation of results: In the process of implementing data governance, after a period of time or after completing a certain stage of governance, a summary and evaluation of the work of the previous stage should be carried out in a timely manner. While summing up the results of governance, we found the deficiencies in the work, and guided and adjusted the governance work in the next stage. Check whether the effect of governance is consistent with the strategy envisaged. The methods of outcome evaluation include maturity model, governance audit, etc.

3. Conclusion and outlook
The article analyzes the data application problems existing in the engineering testing industry, and identified are three main problems: no data available, inability to apply, and no system management. The lack of data is mainly manifested in the lack of consistency, accuracy, completeness, and timeliness of the data. The lack of the above attributes of data makes business personnel face the practical problem that the data cannot be obtained and the data obtained is not easy to use. Data application capabilities are mainly reflected in the lack of data planning capabilities, data collection capabilities, data integration capabilities, and data analysis and decision-making capabilities. Even if the data exists objectively, one does not know how to use it. The data system can be subdivided into three types of systems: organization management, data quality management, and data life cycle management. Through the construction and improvement of the institutional system, the integrity of the governance system can be guaranteed from the theoretical level.

The article suggests that we can learn from the mature model of data governance \cite{5}, and formulate specific data governance strategies from six perspectives: problem cause orientation, organizational structure, infrastructure, data management, data governance technology and tools, and outcome evaluation, so as to improve the industry data application as a whole level.

This article only conducts a preliminary identification and analysis of the problems in the application of data in the engineering testing industry, and gives the direction of governance. However, how to establish a targeted data governance model and how effective the model is for empirical testing remains to be further studied.

Reference:
[1] China Construction Industry Statistical Yearbook-2018, Fixed Assets Investment Statistics Department of National Bureau of Statistics, 2018.
[2] Al-Ruithe M, Benkhelifa E, Hameed K. A systematic literature review of data governance and cloud data governance. Personal and Ubiquitous Computing, 2018(23) 839.
[3] Saed K A, Aziz N, Ramadhan A W. Data Governance Cloud Security Assessment at Data Center, 2018(4) 4.
[4] Jianzhong Li, Xianmin Liu, 2013 Journal of Computer Research and Development. 50(06) 1147
[5] Dama I, 2009 The DAMA Guide to the Data Management Body of Knowledge 1st Edition. USA: Technics Publications. 2009.