


Article

Without Data Quality, There Is No Data Migration

Otmane Azeroual ^{1,*}  and Meena Jha ²

¹ German Center for Higher Education Research and Science Studies (DZHW), 10117 Berlin, Germany

² Centre for Intelligent Systems, School of Engineering and Technology, College of Information and Communication Technology (ICT), Central Queensland University, Sydney, NSW 2000, Australia; m.jha@cqu.edu.au

* Correspondence: azeroual@dzhw.eu; Tel.: +49-302064177-38

Abstract: Data migration is required to run data-intensive applications. Legacy data storage systems are not capable of accommodating the changing nature of data. In many companies, data migration projects fail because their importance and complexity are not taken seriously enough. Data migration strategies include storage migration, database migration, application migration, and business process migration. Regardless of which migration strategy a company chooses, there should always be a stronger focus on data cleansing. On the one hand, complete, correct, and clean data not only reduce the cost, complexity, and risk of the changeover, it also means a good basis for quick and strategic company decisions and is therefore an essential basis for today's dynamic business processes. Data quality is an important issue for companies looking for data migration these days and should not be overlooked. In order to determine the relationship between data quality and data migration, an empirical study with 25 large German and Swiss companies was carried out to find out the importance of data quality in companies for data migration. In this paper, we present our findings regarding how data quality plays an important role in a data migration plans and must not be ignored. Without acceptable data quality, data migration is impossible.

Keywords: data quality; cleansing; data migration; dependency; structural equation models (SEM); business enterprise success



Citation: Azeroual, O.; Jha, M. Without Data Quality, There Is No Data Migration. *Big Data Cogn. Comput.* **2021**, *5*, 24. <https://doi.org/10.3390/bdcc5020024>

Academic Editors: Min Chen and Sabri Pllana

Received: 12 April 2021
Accepted: 16 May 2021
Published: 18 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Companies today often use IT systems that are old and have been specially developed for the company. These systems are called legacy systems [1] and have high operating costs or the employees lack the know-how for maintenance work, since the system is based on old programming languages and mainframes, and the documentation has been lost [2]. These are often the triggers for the procurement of a modern, new system. When switching to the new system, the operationally relevant data must be transferred to the new system. A change to a new system is associated with a migration project. Data migration moves data from one location to another. This could be both a physical relocation and/or a logical relocation. They switch from one format to another or from one application to another. Usually, this happens after a new system is introduced or a new location for the data is introduced. The business backdrop is typically application migration or consolidation, where older systems are replaced or expanded with new applications that use the same dataset. Data migrations are rampant these days, as companies move from on-premise infrastructures and applications to cloud-based storage and cloud-based applications to optimize or transform their business.

Data migration is an important part of digitization in companies. Whenever they introduce new software systems, they have to migrate existing content and information from different data sources. Therefore, quality assurance aims to find errors in the data, data migration programs, and the underlying infrastructure [3]. In order for the data migration to take place, the data must first be cleaned, and the required data quality level must be

achieved. Data cleansing finds incorrect, duplicate, inconsistent, inaccurate, incorrectly formatted, or irrelevant data in a database and corrects it. The data cleansing process consists of several successive individual steps or methods (such as parsing, matching, standardization, consolidation, and enrichment), some of which have to be repeated [4]. Data cleansing offers a number of advantages, such as wrong decisions due to an inadequate database are avoided. Poor data quality can mean that a migration project is unsuccessful. Therefore, it is a prerequisite for the success of the data migration that measures must be taken to improve and secure the data quality.

Data migration is not just a process of moving data from an old data structure or database to a new one; it is also a process of correcting errors and improving overall data quality and functionality. In this paper, the research questions related to data quality and its migration plan are investigated. The research in this paper provides new insights into the issue of data quality in relation to data migration. The aim is to make an important contribution to understanding the dependency of data migration on data quality. In order to determine the relationship between data quality and data migration, an empirical study with 25 large German and Swiss companies was carried out to find out the importance of data quality in companies for data migration. The companies surveyed are innovative solution providers (software development houses) for IT software solutions based on the latest technologies and aiming for long-term market success. The empirical study is carried out through a quantitative analysis in the form of an online survey aimed at people who have already worked on one or more migration projects. A structural equation model is created to illustrate the results. The structural equation model makes it possible to measure the two not directly observable variables data quality and data migration. There are two research questions:

RQ1: How can data quality affect the success of a data migration project?

RQ2: What are the factors that influence the data quality's effect on the success of the migration project in order to be able to derive recommendations for a migration project?

The remainder of the paper is organized as follows. In Section 2, we present concepts of data migration materials, Section 3 presents data quality and its impact on data migration, and Section 4 highlights the relationships between data quality and data migration. Discussion and analysis of our survey and concluding points are given in Section 5.

2. Concept of Data Migration

This section explains a definition of data migration and discusses the requirements, goals, types, and strategies of data migration.

2.1. Definition of Data Migration

The concept of migration is complex and is derived from the Latin "*migrare*" and is one of the greatest concerns of the 21st century [5]. In the Information Technology (IT) area, it can denote a complete system changeover/renewal/modernization as well as any adaptation process of individual components of a system contained therein [6]. A partial or even complete change in the logical and/or physical representation of the data in an information system is called data migration [7]. With data migration, two problems should be addressed. "*First, it is necessary to decide which database system is the target and how data can be transferred to it, and second, how applications can be converted smoothly without affecting availability or performance. It's important not to forget the significant investments that have already been made in data and applications*" [8,9].

2.2. Requirements

The core area of the requirements analysis in the context of a migration is to clarify what the target system to be developed should achieve, and this is the initial phase of a migration project [10]. The migration of data from one system to another can have many reasons, e.g., the introduction of new software application or a change in technology. However, the data structure of a legacy system must be aligned to the requirements of the

new data structure for migration to be successful. Only a successful migration guarantees a conflict-free coexistence of old and new data. The data migration requirements can be divided into three phases:

Exporting and cleaning up old data: When exporting data, it must first be clarified which data should be reused at all. Basically, the data can be divided into two areas: master data and movement data. In order not to burden the new system unnecessarily, a point in time is normally defined for how far back movement data should be transmitted. Everything that goes back is archived separately. Once the amount of data is staked out, the content needs to be cleaned up.

Mapping of old and new data structures: Here, the data structures of the old and the new system have to be adapted. For this purpose, each field or each value in the source data is assigned to a corresponding field in the target system. For example, it is important that the data formats match in terms of field type (text, numeric, alphanumeric, etc.) and field length. A migration tool that supports data synchronization and import can be used for this [11].

Importing the data into the new system: The topic of merging the data shows how good the preliminary work was during the import. As a rule, errors often occur in the first test, so that the mapping must be adjusted. An auxiliary database with the same structure as the target system offers the possibility of checking and editing data content again [12]. It should not be forgotten to thoroughly check and, if necessary, edit all data transferred in the new system.

2.3. Goals of Data Migration

Companies are constantly confronted with the issue of data migration [13]. Data migration occurs whenever software or hardware is replaced because it is out of date. According to Jha et al. [2], business processes are required to be re-engineered for the integrating of structured data, unstructured data, and external data. The goal of data migration could be integrating all different types of data to fulfill the changing requirements of the organizations.

There are three reasons for data migration to fulfill the different requirements of the organizations. The three reasons for data migrations are update migration, ongoing migration, and replacement migration. The update migration is a migration that contributes and generates version change. Version change could be major, minor, or patch depending on the functionality required to be added to the existing legacy system [14]. The ongoing migration includes fundamental changes to the product and thus impacts the environment. It may also be necessary to use migration tools to transform the datasets. Some of the data migration tools are Centerprise Data Integrator, CloverDX, and IBM InfoSphere. The replacement migration includes a product change or the skipping of a product generation and is associated with considerable effort, since no suitable migration tools or documents are available. In addition to these three types, the migration can also take place in two ways, on the one hand by changing the key date, and on the other hand by gradual migration.

Poor data quality is just one of the many challenges that must be overcome as the data changes daily [15]. Data migration will sooner or later face challenges. As part of the migration preparation, test migrations should be carried out with real data. In this way, generic validations and checks can identify errors in data migration at an early stage and improve them. Based on one's own experience in a migration project, several source systems are often integrated into one system. When integrating several systems into one system, it is necessary that a master system is defined so that duplicate data in the target system can be avoided. The data from the source system must be transferred to the target system. In addition, the migration of data is a complex process in which special attention must be paid to the data quality of the master data. Therefore, the following goals are treated and pursued:

1. To analyze and clean up the existing data and documents (by the project and the core organization),

2. Correct automated, semi-automated, and manual migrations of the relevant data and documents, including linking the business objects with the documents,
3. Understand the migration and validate the results obtained. The data protection requirements must be observed.

2.4. Types of Data Migration

There are several types of migration that need to be considered before deciding on a migration strategy. The most complex type of migration is system migration, which affects the entire system. However, depending on the requirements of the migration project, it is possible to only migrate individual parts. The type of migration differs in surface, interface, program, and data migration [16]. During the program migration, all data remain in the old environment, and only the application logic is re-implemented. There are three variants: change the programming language in the same environment, change the environment in the same language, or migrate the language and environment. A pure surface migration leaves the application logic and the data in their old environment. Only the user interfaces are migrated. However, to do this, the user interface must be separated from the program logic. If this is not the case with an old system, the separation can take place through a renovation measure. At the same time, the program logic and the user interfaces can also be migrated. During an interface migration, the system interfaces that connect the system to other systems are migrated. This type of migration must be carried out whenever the external systems with which the system exchanges data change. The way in which the external system changes is unimportant, be it through migration or a new development with new interface protocols. If the legacy system has set up a data exchange with sequential export files, the migration is more complex than if modern XML files or SOAP messages have already been set up for the data exchange [17]. This is because the intervention in the program code when exchanging data via sequential files is much more complex, instead of just connecting the existing code to the new interface. During the data migration, only the data from the old system are transferred, the programs themselves remain unchanged. If the system relied on a relational database, the change is relatively easy and often completely automated. However, the evaluation of the data must be examined more closely to ensure that all data have been transferred correctly. Migrating from relational structures to an object-oriented structure is complicated [18]. This can only be automated to a limited extent, but many problems can be avoided here by suitable modeling. The worst case of data migration is the migration of data based on a non-relational, outdated database. Only the migration of the data is rarely successful in these cases, since both the structure and the access logic of the new database are fundamentally different from the old database. Therefore, the type of data migration is the most complex migration and can be a challenge for the developer.

2.5. Strategies of Data Migration

Most strategies differ in terms of the needs and goals of each organization. As Sarmah [10] said, “A well-defined data migration strategy should address the challenges of identifying source data, interacting with ever-changing goals, meeting data quality requirements, creating appropriate project methodologies, and developing general migration skills” [10]. Basically, there are two strategies to replace an old system, the gradual introduction or the big bang strategy, i.e., the introduction in one step. Which of the strategies is suitable for a particular case must be examined and defined in detail. With a big bang strategy, the old system is switched off, the new system installed, and system parts and data are migrated within a defined period of time—often over a weekend. With a step-by-step migration, the old system is migrated in several steps. Gradual migration is generally less critical than the big bang strategy [19]. Users can slowly get used to the new features. If the new system is not yet stable, the old system can be used in an emergency.

There are two types of step-by-step introduction to migration, and they are as follows:

- The new system offers full functionality but is only available to a limited group of users. New and old systems run in parallel. The group of users is expanded with each level. The problem here is the parallel use of the old and the new system and, in particular, the maintenance of data consistency.
- Another type of introduction is the provision of partial functions for all users. The users work in parallel on new and old systems. With each step, the functionality of the new system is expanded until the old system has been completely replaced.

The right strategies need to be included in a migration plan under different circumstances [20]. Data are the central part of the migration. Data from the old system may need to be transformed into a new format and loaded into the database(s) of the new system. The data migration must be planned in detail. The data flow from the source databases to the target databases is determined. In addition, all necessary data transformations are defined. The process of migrating from a source system to the target system almost always involves the same steps. Nevertheless, the status quo of the data quality in the source systems should be recorded. To this end, it is recommended that project managers work with the defined stakeholders to create a data quality set of rules for the business areas concerned. The next section discusses the impact of data quality on data migration.

3. Data Quality and Its Impact on Data Migration

Bad data quality has different causes. This is a challenge that should not be underestimated. With master data in particular, it can happen that the data formats of the fields in the source system and target system do not match. It can happen that the source data has the wrong format or is in the wrong range of values. To cope with this challenge, the source side must be cleaned up or a validation installed so that these constellations are cleaned up and no longer occur. For databases and information systems, a high data quality is not only something desirable but one of the main criteria that determine whether the project can come about and the statements obtained from it are correct. A higher quality and usability of the data has a direct and positive effect on the decisive business results. As (English, 1999) said, *“the best way to look at data quality is to examine what quality means in the general market and then translate what quality means to data and information”* [21]. There is no clear definition of the term data quality in the literature, which is why it is very individual and subjective. Therefore, according to Würthele [22], data quality is defined as a *“multi-dimensional measure of the suitability of data to fulfill the purpose associated with its acquisition/generation. This suitability can change over time as needs change”* [22]. This definition makes it clear that the quality of data depends on the point in time at which it is viewed and on the level of demand placed on the data at that point in time [23].

With the introduction of data quality methods and tools, the data will be analyzed, and the reports generated with the data quality tool can show where improvement is needed. The use of the data quality tool always makes sense if documentation and planning are currently taking place in several different systems and the data is to be migrated to a new, comprehensive system. The quality of the data is checked and improved so that it is optimally prepared for transfer to the new system. The most important criteria and the influence of data quality on successful data migration are explained below:

- The consolidation and quality improvement takes place before the project of introducing new software. The separation of these two projects is an important success factor.
- As part of the analysis of the existing data landscape, requirements for the new system are identified, which flow into the selection or the initial adjustment of this system.
- Data consolidation and data quality improvement is a project with some factors that cannot be precisely planned in terms of time. By separating the process from the actual migration project, it is easier to plan and be more successful.
- Since the project pressure of the implementation process is largely eliminated, the data can be better prepared with more time and brought to a significantly higher level.
- Data quality methods and tools rank the errors with the greatest impact on the overall result first. Therefore, the time available for error corrections can be used more efficiently.

- The time span between replacing the old documentation and using the new system is optimized. The target system is filled from a data source. Errors due to different versions and errors in links were transparent and cleared up in advance.
- Direct and indirect costs are saved through good data quality in data migration projects (e.g., waste of budgets, costs due to wrong decisions and lost sales, etc.).

With the rapidly growing amounts of data, the importance of quality, accuracy, and timeliness of the internal data structure for corporate success increases even further. Since the data masses are managed and used automatically, inconsistent, incorrect, or incomplete information is often only recognized very late and lead to repeated errors. Data quality problems have drastic effects due to the strong networking of functions and central data storage. Once fed into the system, several departments and company-wide applications access the information and use it repeatedly. In this way, even a small data error can permeate the entire company, causing consequential errors and provoking wrong decisions. In order to counteract typical data quality problems, a systematic approach should be developed. The analysis and elimination of causes for identified problems is the decisive basis for sustainable success. There are many reasons for poor data quality. These can be different depending on the industry or company division. Various attempts at classification can be found in the literature research. Eppler [24] states: *“Literature on information quality problems categories these challenges according to their dimension (content, format, time), their view (information as product or process), or their phase in the information life cycle (production, storage, use)”*. The most important causes are identified below [23]:

- Data collection: Data collection is often the greatest source of errors in terms of data quality. This includes the incorrect use of input masks both by internal employees and by customers who enter information in incorrect input areas (e.g., confusion of first and last name fields). Typing errors, phonetically similar sounds (e.g., ai and ei in Maier or Meier) or inadequate inquiries from service employees are also potential sources of error. Many of these errors can only occur due to a poor design or poor mandatory field protection or plausibility checks in the input masks. However, the import of inadequate external data such as purchased address or customer data can also lead to deterioration in data quality.
- Processes: Processes become the cause of poor data quality if they are incorrect or incomplete (e.g., incorrect processing of existing data or missing check routines).
- Data architecture: Data architecture describes the data processing technologies (e.g., various application software) and the data flow between these technologies. Many of these programs require their own, special data representation, such as formatting or the order of the input and output arguments. Therefore, a conversion of the data is often necessary, which can lead to inconsistencies and thus poor data quality.
- Data definitions: In order for large companies to work effectively, there must be a common understanding of frequently used terms. For example, there is often no uniform scheme for calculating sales or different views as to which data are used. These heterogeneous interpretations can lead to inconsistent data descriptions, table definitions, and field formats.
- Use of data: Errors in application programs can give a user the impression of poor data quality, although the underlying operational system provides almost perfect data in terms of content. Such an impression can occur on the one hand through incorrect interpretation of the data by the user or the creator of the application program from the source system. A supposed correction resulting from this could, contrary to the original intention, introduce new errors into the information system. Apart from an incorrect interpretation, ready-made rules should be created and adhered to for the correction process, such as that data corrections are always made in the source systems and not in the application programs.
- Data expiration: This factor occurs automatically in some areas, as certain data can lose their validity after a certain period. This mainly includes address and telephone

data, but also bank details, price lists, and many other areas are affected by data deterioration, which obviously also limits the data quality.

The reasons for insufficient company data explained here illustrate the complexity of data quality deficiencies in the daily work of the company. However, they point to starting points where such causes can be traced in the organization, thus bringing treatment centers for data quality problems to the fore. Eliminating the causes and improving data quality are certainly the major challenges that companies have to face. Data quality cannot be installed in a company with a big bang; it can only be achieved, maintained, or increased with continuous work. Awareness of data and data quality must be created across the company. Projects and processes to improve quality must also be implemented accordingly. However, at the same time, automation also offers great potential for improving data quality. Measures to increase the data quality must be put into an organizational framework to define responsibilities and processes. The data quality team plays a decisive role in so-called data governance. It is home to the people who carry out data profiling, define the data quality rules, carry out an initial measurement and data quality statement, and define the evaluation with the data owners. The data owner is the person who is responsible for data quality in the department. The data owner receives the data quality statements from the data quality team, makes corrections to the data, and issues orders for new data quality rules or adjustments to rules to the data quality team. The data users are the users of the data in the company. They work with the data and are the ones who benefit most from high data quality. However, they too identify deficiencies in the data and make informal requests for a measurement of specific problems. There are powerful data quality methods and tools available today. With a manageable effort, they can identify the existing data quality problems and state the frequency of errors. Typically, redundant master data, incomplete data records, and incorrectly recorded data come to the surface as well as contradictions between different databases.

4. Relationship between Data Quality and Data Migration

Data migration is a topic for practitioners, which means there are very few publications [25]. The literature [3,26,27] have focused on test, quality assurance, and data quality problems in data migration projects and discuss practical test and quality assurance techniques to reduce and eliminate data migration risks and propose a method for automation before data validation tests for data migrations for quality assurance and risk management in the migration process, which lead to a reduction in effort and costs with improved data quality parameters. The subject of the dependency of migration success on data quality was superficially discussed in a study by [28] and in a bachelor thesis by [29], and attention was given to it. This paper focuses on measuring the relationship between data quality and data migration. The topic is attracting great interest from both the private sector and research, as many of them are currently involved in a data migration project and are part of a data quality team.

To determine the relationship between data quality and data migration, an empirical study was carried out with companies using a questionnaire on SurveyMonkey (www.surveymonkey.de/, accessed on 4 November 2020). The following questions were asked:

1. Have you already worked on a data migration project? Or are you currently on a data migration project?
2. How important is the data quality in the context of the data migration project?
3. Was it a goal to improve and increase the data quality in the course of the data migration?
4. Which data quality criteria [correctness, completeness, consistency and timeliness] are taken into account in the data migration project? How do you rate the degree of fulfillment of the following criteria in relation to the quality of the data at the time or after the data migration project is completed?
5. Which data quality criteria do you take into account as part of the data migration project in order to control and improve the quality?

6. What methods and tools do you use in the data migration project to clean up the dirty data?
7. How do you rate the degree of fulfillment of the following success criteria [project budget, timing, top management, communication and involvement with the end user, training of employees, and employee satisfaction] with regard to the data migration project? Which success criteria do you also consider in the context of the data migration project?

These questions were asked in order to gather information about the data quality and the success of projects and in particular to examine the effects of the data quality from the source systems on the overall success of the data migration project.

A key advantage of the questionnaire approach is that it is cheaper than an oral survey. In addition, a larger number of study participants can be reached in a relatively short time. Data quality and data migration are two variables that cannot be measured directly. The variables can be made measurable via the data quality criteria and the success criteria. So that the data quality can be made measurable, data quality criteria must be defined. In the literature, there are different data quality criteria, but only the four data quality criteria are considered relevant for the survey:

- Correctness: The data must match the reality.
- Completeness: Attributes must contain all the necessary data.
- Consistency: A data record must not have any contradictions in itself or with other data records.
- Timeliness: All data records must correspond to the current state of the depicted reality.

These criteria were selected because, on the one hand, data security and data integrity must be guaranteed at all times during a data migration. Data integrity stands for consistency, completeness, and correctness. On the other hand, the authors are selected on the basis of personal experience (see the work by [30]).

In order for a migration to be carried out successfully, the critical success factors must be known. These are included as follows:

- A migration project always has a budget. The budget contains all the cost-effective resources necessary to achieve the goals.
- A migration project always has an end. It is often carried out under great time pressure.
- An efficient and successful migration can be difficult without the support of top management. Migration often changes processes and behavior. Top management must commit to change and be ready to take risks. If the decision to migrate to a new system is not made by top management, this is not very motivating for everyone involved. The complexity of exchanging a system is very high, and the advantages of exchanging only become apparent after its introduction.
- Open communication and involvement with the end user is an important factor right from the start, because the new system must be accepted in order to be successful. The decision to replace the old system may not be easy for everyone to understand. End users need to understand why the existing system is being replaced so there is no aversion to the new system. Therefore, the involvement of end users in the migration project is an important factor.
- In order for end users to be able to use the new software right from the start and to feel safe, it must be ensured that employees are trained in relation to the new system at an early stage.
- After the introduction of a new system, the satisfaction of the employees can be a criterion for the success of the migration project. The goal with a new system is to give employees a new system that makes their work easier.

In order to make a statement about the dependency between data quality and data migration, it is necessary to consider four data quality criteria and the success criteria of the data migration to one another. A structural equation model (SEM) is suitable for such a consideration, which was developed from our survey and its results. SEM can be defined

as “a comprehensive statistical approach to testing hypotheses about the relationships between observed and latent variables” [31]. It refers to concepts of standard statistical approaches in the social and behavioral sciences such as correlation, multiple regression, and analysis of variance [31]. A large number of statistical models can be combined with the SEM. SEM is a combination of a regression analysis and two factor analyses. With the help of SEM, causal relationships or dependencies can be checked. SEM consists of two parts: the structural and the measurement model. The structural model depicts the relationships between different latent variables in a path diagram, whereas the measurement models depict the relationship between the latent variables and their manifest variables. The SEM makes it possible to measure the two variables that cannot be directly observed. They represent the latent variables of the construct. Each latent variable is operationalized by directly observable variables. The number next to the arrow describes the relationship between the latent variable and the associated indicator. This number is to be interpreted as a factor load and indicates the strength of the reliability of the latent variable.

Data quality is measured using the variables correctness, completeness, consistency, and timeliness. To measure data migration, variables such as project budget, timing, top management, communication, and end-user involvement, training, and employee satisfaction are used as indicators. The SEM, which was created with SmartPLS (www.smartpls.com/, accessed on 20 January 2021), can be seen in Figure 1.

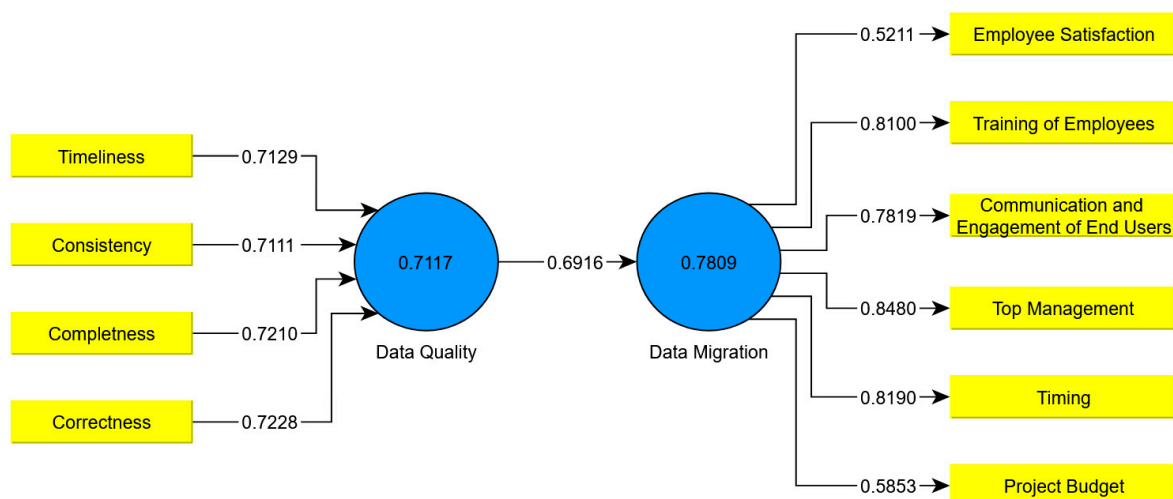


Figure 1. Relationship between data quality and data migration.

Table 1 shows the analysis results for assessing the reliability and validity of the indicators for data quality and data migration. The Cronbach alpha analysis, factor analysis, and principal component analysis (PCA) were calculated with the help of SPSS statistics software (www.ibm.com/products/spss-statistics, accessed on 20 January 2021).

Data quality is measured by four indicators. The four indicators correctness, consistency, completeness, and timeliness show a high load of between 71.11% and 72.28%. This means that these four factors are the most likely to explain the data quality in relation to a migration project. The second latent variable in data migration is described by six indicators. The three indicators top management, timing, employee training, and the communication and involvement of end users are the indicators that best explain the success of the migration between 78.19% and 84.80%. The three indicators budget and employee satisfaction explain the success of the migration being the worst between 52.11% and 58.53%. The results show that the budget does not play the most important role in a migration project. Interestingly, this indicator is the one that least explains the success of the migration. This shows that there is often no going back in a migration project. It is much more important to carry out the project in the agreed timing and to ensure good top

management. The training of the employees and the communication and involvement of the end users is also seen as an important criterion by the participants.

Table 1. Results of the reliability and validity.

Indicators	Cronbach Alpha	Factor Loading	Eigenvalues	(%) of Variance
Correctness	0.9105	0.7228	6.997	36.751
Completeness	0.9244	0.7210		
Consistency	0.9160	0.7111		
Timeliness	0.9090	0.7129		
Project Budget	0.9174	0.5853	11.012	63.249
Timing	0.9064	0.8190		
Top Management	0.9183	0.8480		
Communication and	0.9152	0.7819		
Engagement of End Users				
Training of Employees				
Employee Satisfaction			0.9044	0.5211
Total Cronbachs Alpha	0.9120	Kaiser–Meyer–Olkin Criterion (KMO)		1.000

The results of the investigations show that the total alpha value of Cronbach with 0.9120 is regarded as a very good value. Therefore, Cronbach's alpha reliability coefficients show that all indicators are suitable as a measure for the latent variable (data quality criteria and success criteria of data migration) and have a relative consistency for it. The factor analysis and principal component analysis (PCA) were performed to determine the content and construct the validity of indicators. In factor analysis, the factors should fully explain and interpret the relationships between observed variables. When interpreting the results of a factor analysis of indicators, the number of factors, the number of communalities, and the number of loads are taken into account. Using PCA will reduce the data and extract the factors. This is based on the determination of a covariance or correlation matrix. The PCA can be determined with the help of the Kaiser–Meyer–Olkin criterion (KMO). The total KMO value for all indicators is 1.000. The KMO scores for all indicators were above 0.5, indicating that the sample size was appropriate and that there was a sufficient indicator for each factor (data quality criteria and data migration success criteria). All indicators had a factor loading greater than 0.5, which means that all indicators can be loaded with the same factor. With the help of factor loadings, it can be determined which indicators correlate to a high degree with which factor and which indicators can be assigned to this factor. For the first factor, data quality criteria, the extracted variance was 36.75% for one, while for the second factor, the data migration success criteria was 63.25%. The eigenvalue for both factors is greater than 1.000.

In the context of the SEM, as reported in Figure 1, it can be summarized that data quality is an important factor in data migrations and influences the success of data migration projects. The results show that data quality accounts for 69.16% of the success of the data migration. This means that the dependency between data migration and data quality is significant. A successful data migration can only be carried out if the data quality is observed and verified. In this sense, one can rightly say that a successful data migration does not depend on the data quality but on the measures taken to improve and secure the data quality.

5. Conclusions

The results of the dependency between data quality and data migration have shown that data quality is particularly relevant in data migration and that counteracting this is an essential prerequisite for success. A data governance provides a framework to define the business rules and the data quality level. If no criteria are defined, it is difficult to make

a statement about the status of the data. Then, corrections must be made continuously during the data migration. Such corrections can be time-consuming and delay the project. When migrating data, it is imperative to ensure that all relevant data can be transferred to the target system. Therefore, data quality plays an important role. The empirical study shows, measured against the examined four data quality criteria and six success criteria, that a clear dependence of the migration success on the data quality and data quality has a strong influence on the success of a data migration project. Without optimal data quality, there will be deviations or errors in central business processes and strategic migration projects are doomed to failure.

Author Contributions: O.A.: Conceptualization, Methodology, Formal analysis, Investigation, Supervision, Writing—original draft, Writing—review & editing. M.J.: Writing—review & editing. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request from the authors.

Acknowledgments: The authors would like to thank the 25 large German and Swiss companies who want to remain anonymous in order to participate in my empirical study. The answers provided valuable insights and the results provided an important insight into data quality and data migration and their future.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jha, S.; Jha, M.; O'Brien, L.; Wells, M. Integrating legacy system into big data solutions: Time to make the change. In Proceedings of the Asia-Pacific World Congress on Computer Science and Engineering, Nadi, Fiji, 4–5 November 2014; pp. 1–10. [CrossRef]
2. Jha, S.; Jha, M.; O'Brien, L.; Cowling, M.; Wells, M. Leveraging the Organisational Legacy: Understanding How Businesses Integrate Legacy Data into Their Big Data Plans. *Big Data Cogn. Comput.* **2020**, *4*, 15. [CrossRef]
3. Matthes, F.; Schulz, C.; Haller, K. Testing & quality assurance in data migration projects. In Proceedings of the 27th IEEE International Conference on Software Maintenance (ICSM'11), Williamsburg, VA, USA, 25–30 September 2011; pp. 438–447.
4. Azeroual, O.; Saake, G.; Abuosba, M. Data quality measures and data cleansing for research information systems. *J. Digit. Inf. Manag.* **2018**, *16*, 12–21.
5. Verhulst, S.G.; Young, A. The Potential and Practice of Data Collaboratives for Migration. In *Guide to Mobile Data Analytics in Refugee Scenarios*; Salah, A., Pentland, A., Lepri, B., Letouzé, E., Eds.; Springer: Cham, Switzerland, 2019; pp. 465–476.
6. Leloup, F. Migration, a complex phenomenon. *Int. J. Anthropol.* **1996**, *11*, 101–115. [CrossRef]
7. Stahlknecht, P.; Hasenkamp, U. *Einführung in die Wirtschaftsinformatik*; Springer: Berlin/Heidelberg, Germany, 1999.
8. Meier, A.; Mercerat, J.; Muriset, A.; Untersinger, J.; Eckerlin, R.; Ferrara, F. Hierarchical to Relational Database Migration. *IEEE Softw.* **1994**, *11*, 21–27. [CrossRef]
9. Meier, A. Providing Database Migration Tools—A Practitioner's Approach. In Proceedings of the 21st International Conference on Very Large Data Bases (VLDB'95), Zürich, Switzerland, 11–15 September 1995; pp. 635–641.
10. Sarmah, S.S. Data Migration. *Sci. Technol.* **2018**, *8*, 1–10.
11. Khajeh-Hosseini, A.; Sommerville, I.; Bogaerts, J.; Teregowda, P. Decision Support Tools for Cloud Migration in the Enterprise. In Proceedings of the 2011 IEEE 4th International Conference on Cloud Computing, Washington, DC, USA, 4–9 July 2011; pp. 541–548. [CrossRef]
12. McAdam, J. The concept of crisis migration. *Forced Migr. Rev.* **2014**, *45*, 10–11.
13. Morris, J. *Practical Data Migration*, 3rd ed.; British Informatics Society Ltd.: Swindon, UK, 2006.
14. Derr, E.; Bugiel, S.; Fahl, S.; Acar, Y.; Backes, M. Keep me Updated: An Empirical Study of Third-Party Library Updatability on Android. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17); Association for Computing Machinery: New York, NY, USA, 2017; pp. 2187–2200.
15. Laranjeiro, N.; Soydemir, S.N.; Bernardino, J. A Survey on Data Quality: Classifying Poor Data. In Proceedings of the 21st IEEE Pacific Rim International Symposium on Dependable Computing, (PRDC 2015), Zhangjiajie, China, 18–20 November 2015; pp. 179–188.
16. Morris, J. *Practical Data Migration*; BCS, The Chartered Institute: Swindon, UK, 2012. Available online: <https://ws1.nbni.co.uk/fusion/v2.0/supplement/5d6e240d646eb18c10cb4e84.pdf> (accessed on 17 May 2021).

17. Karnitis, G.; Arnicans, G. Migration of Relational Database to Document-Oriented Database: Structure Denormalization and Data Transformation. In Proceedings of the 7th International Conference on Computational Intelligence, Communication Systems and Networks, Riga, Latvia, 3–5 June 2015; pp. 113–118. [CrossRef]
18. Hudicka, J.R. An Overview of Data Migration Methodology. 1998. Available online: https://dulcian.com/articles/overview_data_migration_methodology.htm (accessed on 22 March 2021).
19. Latt, W.Z. Data Migration Process Strategies. Available online: <https://onlinerresource.ucsy.edu.mm/handle/123456789/1226> (accessed on 17 May 2021).
20. Lin, C.Y. Migrating to Relational Systems: Problems, Methods, and Strategies. *Contemp. Manag. Res.* **2008**, *4*, 369–380. [CrossRef]
21. English, L.P. *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*; Wiley: New York, NY, USA, 1999.
22. Würthele, V. *Datenqualitätsmetrik für Informationsprozesse: Datenqualitätsmanagement Mittels Ganzheitlicher Messung der Datenqualität*; ETH Zurich: Zurich, Switzerland, 2003.
23. Apel, D.; Behme, W.; Eberlein, R.; Merighi, C. *Datenqualität Erfolgreich Steuern: Praxislösungen für Business-Intelligence-Projekte*, 3rd Revised and Extended Edition; dpunkt.verlag: Heidelberg, Germany, 2015.
24. Eppler, M.J. *Managing Information Quality: Increasing the Value of Information in Knowledge-Intensive Products and Processes*; Springer: Berlin/Heidelberg, Germany, 2006.
25. Haller, K. Towards the industrialization of data migration: Concepts and patterns for standard software implementation projects. In Proceedings of the 21st International Conference on Advanced Information Systems Engineering (CAISE), Amsterdam, The Netherlands, 8–12 June 2009; pp. 63–78.
26. Manjunath, T.N.; Hegadi, R.S.; Archana, R.A. A study on sampling techniques for data testing. *Int. J. Comput. Sci. Commun.* **2012**, *3*, 13–16.
27. Paygude, P.; Devale, P.R. Automated data validation testing tool for data migration quality assurance. *Int. J. Mod. Eng. Res.* **2013**, *3*, 599–603.
28. Clément, D.; Ben Hassine-Guetari, S.; Laboisie, B. Data Quality as a Key Success Factor for Migration Projects. In Proceedings of the 15th International Conference on Information Quality (ICIQ) 2010, Little Rock, AR, USA, 12–14 November 2010.
29. Kreis, L. Datenqualität als kritischer Erfolgsfaktor bei Datenmigrationen. Bachelor's Thesis, Zurich University of Applied Sciences, Zurich, Switzerland, 2017.
30. Azeroual, O.; Saake, G.; Abuosba, M.; Schöpfel, J. Data Quality as a Critical Success Factor for User Acceptance of Research Information Systems. *Data* **2020**, *5*, 35. [CrossRef]
31. Hoyle, R.H. The structural equation modeling approach: Basic concepts and fundamental issues. In *Structural Equation Modeling: Concepts, Issues, and Applications*; Hoyle, R.H., Ed.; Sage Publications, Inc.: Washington, DC, USA, 1995; pp. 1–15. Available online: <https://psycnet.apa.org/record/1995-97753-001> (accessed on 17 May 2021).