



**HAL**  
open science

## **Data Quality Considerations for ERP Implementation: Techniques for Effective Data Management**

Mohamed Elbadri, Ahmed Altaher, Sharafedin Alkawan

### ► **To cite this version:**

Mohamed Elbadri, Ahmed Altaher, Sharafedin Alkawan. Data Quality Considerations for ERP Implementation: Techniques for Effective Data Management. Second International Libyan Conference for Information and Communications Technologies (ILCICT 2023, Libyan Telecommunication and Information Technology Society, Sep 2023, Tripoli, Libya. <hal-04342090>

**HAL Id: hal-04342090**

**<https://hal.science/hal-04342090v1>**

Submitted on 13 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

# Data Quality Considerations for ERP Implementation: Techniques for Effective Data Management

Mohamed Elbadri<sup>1</sup>, Ahmed Altaher<sup>2,3</sup> [0000-0001-7909-6444] and Sharafedin Alkawan<sup>2</sup>

<sup>1</sup> University of Grenoble Alpes, Grenoble, France

<sup>2</sup> S2A2I-Lab, Saint Martin d'Hères, France

<sup>3</sup> College of Electronic Technology, Tripoli, Libya  
ahmed.ah.altaher@gmail.com

**Abstract.** Most industrial companies implement enterprise resource planning systems to enrich their digital transformation. ERP systems are databased technologies that employ big data, machine learning, data science and automation. Data is important to deploy such centralized systems, and to improve business decisions. ERP system can merge data from warehouse, sales orders, marketing commands, finance, human resources and management offices. It helps enterprises to get advantages of digital ecosystems and competitive rank in the market. Data quality issues in ERP implementation can have significant consequences for businesses success and their competitive positions. To mitigate these consequences, enterprises must prioritize data quality during ERP implementation. Implementing robust data governance practices, conducting data cleansing and validation, ensuring data accuracy and completeness, and establishing data quality monitoring mechanisms are crucial steps to enhance data quality and minimize the negative impact on business operations.

This paper focuses on some data issues that may face implementers during data migration phase, in addition, it proposes some solutions. The work reviews current related studies that help companies to overcome lack of quality data and related problems that may cause delay, budget deficit and collapse of ERP projects. It discusses three concepts namely, data integration, ETL process and data integrity in order to help ERP project managers distinguish these activities. Additionally it provides technical recommendations to manage data quality for success of ERP implementations.

**Keywords:** Data Migration, ERP Integrity, Quality Data.

## 1 Introduction

During ERP implementation, old data migration, from old distributed resources to a novel centralized system, shapes an important step toward good execution of ERP projects. This phase requires suitable arrangements such as finding skilled developers, allocating data resources and preparing data backups. ERP developers, implementers and project managers can name many difficulties, although data migration is not appropriately covered. Some risks may occur such as extending project duration, ex-

ceeding a dedicated budget due to hiring outsourcing skills, and getting rough opinions from stakeholders. In this case, implementers face difficulties to avoid data quality issues, e.g. missed, repeated or altered data.

Large enterprises own distributed data in several places that can be in different data store formats such as database records or digital spreadsheets. This data cannot be migrated into the new ERP system without prior processing and handling. Low quality of data may therefore delay the project schedule due to the required processing and handling. Preparing or hiring effective data force implementers is important as well as taking data issues in their minds, i.e. data needs collecting, cleansing, and evaluating of its integrity, i.e. to be ready for the ERP system. For instance, as soon as the project staff start to collect the required ERP data, duplicate data can be found in several places. In this case, data should be processed and tuned through deleting, altering and reediting of this data. The ERP users shall notice data formats, adjustable fields size and meaningful values, i.e. according to their use with ERP data and business records. In small and medium enterprises (SME), ERPs are integrative systems that involve centralized version of the SME transactional operations. Today, all identified data redundancies and inconsistencies that can guarantee the transactional operation, and often the success of the company, although in the future can lead to the failure of the ERP implementation. Consultants should strive to understand, standardize and integrate deficient business processes before large-scale ERP implementation. Thus, no crucial aspect of the SME operation is forgotten or disregarded [1].

Consultants that are hired, to carry out the ERP system, have an important role at early stages of the implementation and ongoing-live phases. In addition, the attention, from the top management and experts, is a significant element in the ERP project success [2]. ERP implementers may append timetable to have quality data besides hiring experts, in result, allocating more budget because special skills and knowledge are important to do so. To mention some skills, data dashboards require data science expertise at manufacturing and managements levels in order to track production capacity and quality. Companies normally hire dedicated personnel or train their adjacent IT staff to have these add-on skills. Many companies (around 20%) just starting to figure that data science and AI are among major skills gap and around 40% of these companies decided to enrich their staff skills [3].

In this paper, data quality is defined and related data issues are clarified. Additionally, authors survey some recent solutions that automate the process of data cleansing and help to reach data integrity. The paper defines the notion of the data quality issues and their characterization in the context of ERP implementation. The main goal is to provide concise recommendations, which help ERP project managers and implementers avoid data quality issues. The paper contributes to the data management of ERP projects by highlighting the differences between data integration, ETL process, and data integrity to help ERP project managers distinguish these activities. By presenting technical recommendations, the paper also contributes to the management of data quality in ERP projects to ensure successful implementations.

## 2 The Process of Data Migration

For business applications, data migration process aims to transfer/convert data originally founded in the old system into formats that are suitably conform to the new system purpose. This data can be in hard, e.g. catalogues or booklets, or soft records, e.g. databases. In modern systems, data is in digital formats, although it is distributed among enterprises departments or branches. In further complexity, modern enterprises switch on premise infrastructure and applications to cloud-computing models, i.e. cloud-based storage and apps, to allow users to store, access, and modify their data on the internet through a cloud service provider. Service providers are responsible for managing and operating this data storage as a service (SaaS). Enterprises subscribe to a specific data storage capacity, usually in a pay-as-you-go model [3]. In this model, data synchronization and downloading are necessary. Hence, data migration consist converting hard records and soft data into required digital formats in order to be ready for manipulation by computer software. The process also could be to convert from old distributed databases to new centralized ERP database as depicted by Fig.1.



**Fig. 1.** Data Migration principal steps with integrity after collecting and cleansing of data.

## 3 Importance of Data Quality

Poor data quality occurs when stakeholders discover inaccuracies and inconsistencies in a data collection or when data extraction is not as intended [4]. Incorrectly entered, duplicate, incomplete or irrelevant data cause poor quality. Hence, the use of poor quality data in an ERP can degrade the quality of transactional, master, analytical data resources. As a result, use of this data, for business analysis, shall lead to harmful, difficult, and unwise decisions.

Due to existence of non-proper data, the process of migrating data from physical catalogs or records, a single data resource (i.e. database), or distributed databases to an ERP database shall encounter some issues that lead to migration delays. In this case, data integrity is needed to increase the quality of data resources before bringing the ERP online. If the data resources are accurate, complete, and logically consistent, then one can said that these resources have data integrity (i.e. to guarantee quality data). The later therefore requires systematic inspection to verify storage, processing,

reporting and analysis of data. Additionally, checking the data integrity require good participation of the enterprise workforce to eliminate any information gaps between the implemented ERP system and reality. This participation facilities the communication from/to top management and enhances the integrity and efficiency of the system in order to increase data quality [2].

Indeed, the company's information system shall be improved by adopting good quality practices to ensure the data integrity. Moreover, through such practice improvements, enterprises can identify and overcome many barriers, e.g. inefficient cross-functional collaboration, and potential risks, e.g. users enter incorrect data into the ERP system, which can affect the success of the ERP project [1].

Quality issues could be related to data sources, data transfer or/and data entry. During data migration, i.e. to the ERP system, data transfer quality should be guaranteed through integrity inspection, e.g. incomplete migrated data, repeated, non-clean; entry mistakes, mismatch data and bad data mapping are the common factors that contribute to bad quality. Errors can happen throughout migration of a huge data transfer. Mismatch data usually is exist when dealing with big databases of tables from legacy systems. Indeed, testing is important step to evaluate the data integrity. This step lets the developers identify the scale of success. A good plan is to migrate a small sample of data individually first, to make sure everything in the system is working normally, then a larger dedicated sample of data is migrated afterwards, and so on.

To sum up, data quality means data integrity from a data analysis perspective and good decisions from a business perspective. In other words, having high quality and integrity data at the operational and tactical levels shall feed useful information for decision-making processes at the executive and strategic levels.

## **4 Consequences of Poor Quality Data**

In practice, the use of poor quality data leads to tangible penalties for any company. The industry 1-10-100 rule means; data entry verification costs \$1, while \$10 to clean this data once records are infected with impure data, and \$100 if the business does nothing [3]. Poor-quality data can lead to weak customers' relationships, imprecise data analytics, wrong decisions, and low business performance. Poor data quality may seem like a minor source of problems, but they can easily escalate as repeated data errors grow and accumulate [5][6]. Hence, any use/entry of poor quality data can lead to unsatisfied results that may include:

1. Imprecise reporting and decision-making: if users enter incorrect, incomplete or inconsistent data in the ERP system, results can be inaccurate financial statements, inventory records, customer or critical business information. This can lead to weak decisions based on unreliable information.
2. Process inefficiencies: ERP systems are planned to improve efficiency of business processes. However, poor quality data can disrupt these processes. Incorrect data, related to inventory/supply chain management, can increase orders delays or cause

stock outs. This might lead to operational inefficiencies, increased costs and dissatisfied customers.

3. Increased time and effort: poor quality data needs more time and effort than expected. During ERP implementation, data cleansing and data migration activities become complex and time-consuming while dealing with inaccurate data that leads to delay of implementation, rise of project cost and set a strain on resources.
4. User resistance challenges: users may resist adopting the ERP system if they face data quality issues. If employees experienced data related issues in the past, they may lack confidence in the system accuracy and resist using it. This resistance can hinder and undermine the overall success of the ERP project.
5. Customer dissatisfaction: poor quality data can have direct impact on customer satisfaction, e.g. incorrect billing, shipping errors or failed deliveries due to inaccurate customer records. Such issues can lead to customer dissatisfaction, trust loss, company's reputation damage.
6. Regulatory compliance issues: poor quality data can cause non-compliance with regulatory requirements, e.g. The ERP system fails to summarize and report accurate information required for regulatory purpose, leading to legal and financial consequences, i.e. fines, penalties or legal actions.

To mitigate these consequences, enterprises and stakeholders should prioritize data quality management as a crucial aspect.

## 5 Techniques to Handle Data Quality Issues

Many techniques, that can support ERP implementers, are exist since long time, although they are used for data warehousing and data management. ETL (Extracting, Transforming and Loading) and data integration are renovated tools in terms of ERP data management, i.e. managing data for ERP implementations. More insights on these techniques are followed.

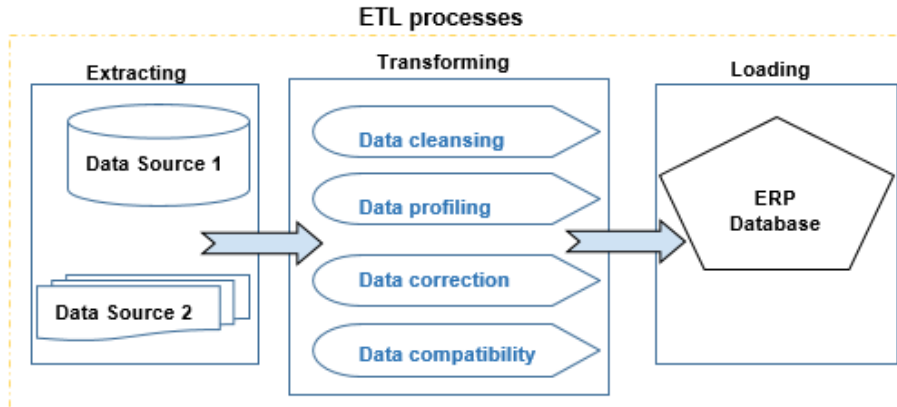
### 5.1 Extraction, Transforming and Loading

Many reasons justify the need for ETL as a data integration tool phase within the ERP systems implementation, this step therefore is needed to handle data issues such as existence of heterogeneous data formats (e.g. data formats might be ambiguous or difficult to interpret/transform), outdated databases in legacy systems, and structural changes of the data sources over time. These issues make decisions uncertain. Indeed, ETL is a crucial element in data warehousing process where most of the data cleaning and curation is done [7].

Maintaining compatibility of data formats is necessary process to have data integrity. Now, technology is more precise in data compatibility. In fact, ETL provides a solution of data sourcing (from original data), then transforming the data into a compatible format, finally loading the data, which is the part that solves the data migration issues [8]. This methodology enables implementers to use smart solutions regarding

data warehouse management. Some applications, e.g. the Talend software, are capable tools that can automatically manage this envisaged solution.

Intelligent platforms can centralize data processing and gives more automated, secured and smooth data migration. To name a method, ETL is one of the approaches that corresponds to data management (Fig.2).



**Fig. 2.** Extracting, Transforming and Loading (ETL) of data to new ERP central database

## 5.2 Data Integration

data integration share some ETL commonalities as they both deal with the process of combining data from various sources and making it accessible and usable in a unified manner. Although, data integration is more comprehensive process [9]. While ETL is a specific type of data integration process that focuses on the extraction, transformation, and loading of data into a target system, data integration is a broader concept that encompasses various techniques and approaches for integrating data from different sources and systems. ETL is a crucial part of data integration, and both are essential components in modern data management and analytics workflows [7-9]. Data integration ensures that data from different departments, systems, and sources are brought together and consolidated within the final ERP system. Additionally, Data integration enables many business functions (e.g. finance, inventory management, sales, procurement, human resources, and more) to seamlessly share and exchange data, facilitating cross-functional processes and ensuring consistency and accuracy across different areas of the enterprise [10]. Data integration in ERP systems often involves real-time or near-real-time data synchronization. This means that all changes are automatically propagated to other relevant modules or components. This real-time data integration allows for accurate and up-to-date information throughout the organization [11]. In this manner, integrating data within the ERP system eliminates data divisions, i.e. data silos. Enterprises can break down these silos and create a unified view of data, enabling better decision-making and analysis [12]. Finally, ERP systems also require data integration to interact with external systems, such as suppliers, cus-

tomers, banks, or third-party applications. Integration interfaces or APIs (Application Programming Interfaces) are often used to exchange data between the ERP system and these external entities, ensuring seamless data flow and process automation [13]. Generally, data integration is a critical aspect of ERP systems, enabling enterprises to streamline operations, improve data accuracy, and gain comprehensive insights into their business processes [14] [15]. Table 1 depicts some well-known integration software that have data ETL feature to assist ERP implementers to extract, transform and load quality data.

**Table 1.** Non-comprehensive list of well-known integration platforms with ETL capabilities

<b>Application tool</b>	<b>Type</b>	<b>Usage</b>	<b>Features</b>
Talend Data Integration	Data Integration & ETL	Data integration, data quality, data preparation, big data, and application integration.	Rapid development, reliable with different datasets, large users community, native support for Big Data Environment
Informatica PowerCenter	Integration platform	Integration of wide range of data	Integration capabilities that support ETL, data quality, and data governance.
Microsoft SQL Server Integration Services	Integration and ETL tool	Development environment to solve complex business problems by copying or downloading files, loading data warehouses, cleansing and mining data	Powerful visual development with Management Studio. It integrates well with other Microsoft products.
IBM InfoSphere DataStage	Integration and ETL tool	Designing, developing, and running jobs that move and transform data.	Visual interface for designing, developing, and managing ETL processes. Extensive connectivity options and supports large-scale data integration projects.
SAP Data Services	Integration services	Data integration, data quality, data profiling and	Integration capabilities, transform trusted data-

		data processing.	to-data warehouse system for analytical reporting.
Oracle Data Integrator	Integration platform	Data transformation, movement, and synchronization capabilities	Declarative design approach and tight integration with Oracle databases.
Pentaho Data Integration	open-source integration & ETL	Data integration, ETL processes	Visual interface, supports a wide range of data sources, and offers features like data profiling and cleansing.
Apache NiFi	open-source data integration	Data flow management.	Web-based interface and allows users to design and execute data integration workflows with a focus on data routing, transformation, and system mediation.
Matillion ETL	cloud-native ETL platform	Modern data integration in cloud environments.	Support for popular cloud services such as Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure.
Mage	open-source ETL	Integration and transforming data	Integration and synchronization of data from third-party sources. Supports building real-time and batch pipelines.
Airbyte	open-source ELT and integration platform	Integration and synchronization of data sources	Offers pre-built data connectors from an API or

			User Interfaces.
--	--	--	------------------

ERP project managers and implementers can choose an integration or ETL tool. However, better to have team skills and to decide with their teammates. The applications that are mentioned in table.1 can assist teams to migrate data and to manage its quality.

## 6 Discussions

Considering data quality, when it comes to ERP implementations, all the three aspects ETL, Data Integration, and Data Integrity—are important and play different roles in ensuring high-quality data. Focusing on data quality in new ERP projects, all these three aspects are interconnected and essential for success. ETL ensures that data is appropriately extracted, transformed, and loaded into the ERP system, addressing quality issues along the way. Data integration establishes a unified view and consistent data flow, reducing data discrepancies and enhancing quality, while data integrity ensures the overall accuracy, completeness, and reliability of data within the ERP system.

By addressing data quality issues during the extraction and transformation stages, ETL helps maintain accurate and reliable data within the ERP system. Data integration supports data quality by ensuring consistency, coherence, and completeness of data across the ERP system. It helps prevent redundancy, data inconsistencies, and conflicting information, ultimately improving the overall data quality and reliability. Data integrity measures, such as data validation rules, referential integrity constraints, and data governance policies, help enforce data quality standards and prevent issues like duplicate records, missing values, or incorrect data. Maintaining data integrity is crucial for ERP implementations as it safeguards the reliability of information and supports effective decision-making processes.

Technically, to achieve optimal data quality in ERP implementations, it is crucial to select one or to combine all these three aspects as they work together to ensure accurate, consistent, and reliable data within the system.

## 7 Conclusions

Data quality plays a crucial role in ERP implementations, especially with the advancements in business analytics and artificial intelligence. When migrating data from diverse sources, whether they are outdated or modern systems, it is essential to prioritize integration, ETL processes, and data integrity. This approach ensures that the data maintains a high level of quality and becomes a fundamental factor in the success of the ERP project. In this regard, the present work aims to provide practical recommendations specifically tailored for ERP project managers and implementers. Data quality can be improved, and potential risks during the migration phase can be mitigated, ultimately resulting in a successful ERP implementation, by following these recommendations:

- **Develop a Data Strategy:** Create a comprehensive plan that outlines the standards and processes for data quality throughout the project. Define data quality metrics, establish data cleansing and validation procedures, and identify responsibilities for data quality management.
- **Engage Data Stewards and Data Governance:** Assign dedicated data stewards or a data governance team responsible for overseeing data quality, resolving data-related issues, and ensuring adherence to data management policies and practices.
- **Foster Collaboration and Communication:** Encourage collaboration and open communication among project stakeholders, including business users, IT teams, and data experts. Regularly review and discuss data quality metrics, address concerns, and incorporate feedback to improve the overall success of the ERP project.
- **Perform Data Profiling and Analysis:** achieve a thorough analysis of the data to be migrated, including identifying inconsistencies, duplicates, missing values, and data quality issues. This analysis will help in determining the scope and complexity of the migration process and identify appropriate correction actions.
- **Define Data Migration Requirements:** Clearly define the data migration requirements and objectives, including data mapping, transformation rules, and data validation criteria. Collaborate with stakeholders to ensure that all necessary data elements are captured and migrated accurately.
- **Implement Data Cleansing and Transformation:** Prior to the migration, cleanse and transform the data to ensure consistency, accuracy, and completeness. Remove duplicate records, standardize formats, and resolve any data anomalies or errors to improve data quality.
- **Ensure Data Integrity and Validation:** Establish data integrity checks and validation mechanisms to ensure the accuracy and reliability of the migrated data. Implement data validation rules and perform thorough testing to identify and rectify any discrepancies or inconsistencies.
- **Conduct Thorough Testing:** Perform comprehensive testing of the ERP system and the migrated data to validate functionality, performance, and data integrity. Test various scenarios and perform end-to-end testing to ensure that the system operates as intended and that data is accurately processed and displayed.
- **Train Users and Provide Documentation:** Educate and train users on the importance of data quality and the correct usage of the ERP system. Provide detailed documentation and resources that explain the data migration processes, data entry guidelines, and best practices to maintain data quality.
- **Monitor and Maintain Data Quality:** Establish ongoing monitoring and maintenance processes to continually assess and improve data quality within the ERP system. Implement data governance practices, regular data audits, and data quality checks to address any issues that may arise post-migration.

## References

1. Christofi, M., Nunes, M., Chao Peng, G., & Lin, A.: Towards ERP success in SMEs through business process review prior to implementation. *Journal of Systems and Information Technology*, 15(4), 304-323 (2013).
2. Mahar, F., Ali, S. I., Jumani, A. K., & Khan, M. O.: ERP system implementation: planning, management, and administrative issues. *Indian J. Sci. Technol*, 13(01), 1-22 (2020).
3. David, J.: The AI Talent Shortage is not Over Yet, *THE WALL STREET JOURNAL*, 27/10/2020, last accessed 2023/06/20.
4. Vosburg, J., & Kumar, A.: Managing dirty data in organizations using ERP: lessons from a case study. *Industrial Management & Data Systems*, 101(1), 21-31 (2001).
5. Keith, D.: The Impact of Poor Data Quality (and How to Fix It), *DATAVERSITY*, Foote on March 1, 2023, last accessed 2023/06/23.
6. Gudivada, V., Apon, A., & Ding, J.: Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software*, 10(1), 1-20 (2017).
7. Souibgui, M., Atigui, F., Zammali, S., Cherfi, S., & Yahia, S. B.: Data quality in ETL process: A preliminary study. *Procedia Computer Science*, 159, 676-687 (2019).
8. Nguyen, T.: Improving the ETL process for a case company. Thesis of Master in IT 2023, Metropolia University of Applied Science, Finland, last accessed 2023/06/20.
9. Jovanovic, P., Nadal, S., Romero, O., Abelló, A., & Bilalli, B.: Quarry: a user-centered big data integration platform. *Information Systems Frontiers*, 23, 9-33 (2021).
10. Chopra, R., Sawant, L., Kodi, D., & Terkar, R.: Utilization of ERP systems in manufacturing industry for productivity improvement. *Materials today: proceedings*, 62, 1238-1245 (2022).
11. Sarferaz, S.: Compendium on Enterprise Resource Planning: Market, Functional and Conceptual View Based on SAP S/4HANA. Springer Nature, (pp. 29-50). Cham: Springer International Publishing (2022).
12. Muniswamaiah, M., Agerwala, T., & Tappert, C.: Data virtualization for decision making in big data. *Int. J. Softw. Eng. Appl*, 10(5), 45-53 (2019).
13. Budiman, K., Putra, A. T., Sugiharti, E., Muslim, M. A., & Arifudin, R.: Implementation of ERP system functionalities for data acquisition based on API at the study program of Universities. In *Journal of Physics: Conference Series* (Vol. 1918, No. 4, p. 042151). IOP Publishing (2021).
14. M. Beheshti, H., K. Blaylock, B., A. Henderson, D., & G. Lollar, J.: Selection and critical success factors in successful ERP implementation. *Competitiveness Review*, 24(4), 357-375 (2014).
15. Watson III, E. F., & Schwarz, A. H.: Enterprise and Business Process Automation. In *Springer Handbook of Automation* (pp. 1385-1400). Cham: Springer International Publishing (2023).