



# Leveraging Data Remediation for Better Data Quality

Whitepaper  
**2023**

Online:

[contact@opplane.com](mailto:contact@opplane.com)

[www.opplane.com](http://www.opplane.com)

Phone:

+1 833-OPPLANE

Date of Issue:

01 // 09 // 2023

# Table of contents

- 1. Introduction.....3
- 2. Understanding Data Quality.....5
  - Definition and key features.....7
  - Importance for business and decision-making.....7
- 3. The Business Costs of Poor Data Quality.....8
- 4. The Role of Data Remediation in Enhancing Data Quality.....10
  - What is Data Remediation?.....11
- 5. Opplane’s 6-Step Masterplan to Data Remediation.....12
  - A Primer on Data Remediation Tools and Technologies.....13
- 6. Best Practices for Data Remediation and Data Quality Improvement.....15
- 7. Conclusion.....16
- References.....18
- About Opplane.....19

# 1. Introduction

**For centuries, humanity's capacity to produce data was relatively scarce. Technological constraints meant collecting, storing, and processing data was generally time-consuming and expensive.**

However, with the introduction of an ever-increasing panoply of Digital Information and Communication Technologies, the scarcity of data has given way to an enormous abundance of data in the space of a few decades.

In these first decades of the 21st century, primarily due to the spread of broadband Internet and smartphones, few people worldwide go an entire day without creating and using data. In 2010, Eric Schmidt, the then-CEO of Google (now Alphabet), made the following estimate: "There were 5 Exabytes of information created between the dawn of civilization through 2003, but that much information is now created every two days."

Since then, the pace of growth has been so unstoppable that we are now living in the Zettabyte<sup>1</sup> era. According to IDC's Global DataSphere Forecast 2021-2025<sup>2</sup>, in 2020, the total amount of data created or replicated worldwide reached 64.2 Zettabytes. IDC projects that, at this rate, this number will increase to 181 Zettabytes by 2025.

Just as oil was the engine of the industrial economies of the 19th and 20th centuries, so data is the growth

factor for the digital economies of the 21st century. No wonder some, like mathematician Clive Humby did in 2006, claim that data is the new oil. Just as oil needs to be refined to create gas, plastic, and other materials, data must also be organized and analyzed to gain insights, make decisions, and detect problems<sup>3</sup>.

The upshot of this sudden transformation is that small and large organizations feel inundated with a deluge of data. Hence the need to ensure the quality of that data. Data is useless in its raw state. Yet, with the proper processing to increase its quality, it can contribute to the decision-making process.

For an organization to succeed in its business sector, ensure compliance with regulations and laws, and adequately manage risks, it needs access to quality data. Yet most companies still need to address the importance of good quality data. This neglect leads to losses in productivity, production disruptions, and higher maintenance costs. According to an IBM forecast from 2016, companies in the United States alone lose \$3.1 trillion annually due to poor data quality<sup>4</sup>.

This white paper explains how you can improve the quality of your organization's data through data remediation. This process aims to ensure that all available data is fit for purpose. In addition to data cleaning - i.e., identifying and resolving errors, inconsistencies, or inaccuracies -, data remediation includes data validation, normalization, enrichment, consolidation, and reconciliation.

At the same time, with this white paper, we aim to present to potential customers some of the data remediation techniques commonly applied by Opplane in its most significant projects. With experience building PayPal's first data warehouse, Opplane's leaders have years of leadership experience in other areas and top organizations. With its talented data scientists, engineers, and analysts, Opplane's engineering teams specialize in solving complex data problems and have experience in fintech, banking, and retail. At Opplane, data quality has been a critical focus since its inception. From that perspective, data remediation is the most important of the three steps comprising our approach to implementing a data quality program.

In addition to a *more in-depth explanation of what data quality is* and *why it matters to contemporary organizations*, you will find the following topics in this white paper:

- *The cost impact of poor data quality* on operational efficiency and business decision-making.
- *The role of data remediation in enhancing data quality*, including *key benefits and advantages*.
- A step-by-step explanation of *how to implement Opplane's data remediation methodology*.
- Presentation of some of the most successful *data remediation tools and technologies*.
- *How organizations can leverage the potential of data remediation* in their daily operations.



## 2. Understanding Data Quality

### Definition and key features

While it's undeniably true that improving the performance of a business increasingly depends on the quality of the data this business collects and produces, it's also the case that the problems related to (poor) data quality are almost unavoidable in the life of an organization, given the growth in the amount of data consumed and produced by organizations. Nevertheless, companies that are conscious of the importance of data quality and implement data quality management policies tend to meet fewer hurdles.

Although data quality may seem intuitive, it can be complex and challenging to determine. Being contextual, it depends on the needs of the organization relying on the data. According to the Data Management Association (DAMA), data is high quality when it suits the business purpose and poor quality when it does not (DAMA, 2017, p. 454).

While data quality is always contextual and can vary widely across industries and departments within the same organization, there is a set of aspects that can be quantified and controlled with a relative degree of reliability. This measurable set of data characteristics is called dimensions. The term was initially introduced by Professors Diane Storm and Richard Wang in their 1996 paper "Beyond Accuracy: What Data Quality Means to Data Consumers" to describe data consumers' expectations of data attributes. The authors list 15 dimensions within four different categories of data quality (intrinsic, contextual, representational, and accessibility).



Subsequently, other authors and entities have come forward with different sets of data quality dimensions. Notwithstanding the diversity of themes and approaches included in the concept of data quality dimensions, it is possible to identify six frequently cited aspects:

#### 1. Accuracy:

Refers to the degree to which data correctly represent the described real-world entities. The closer a measured value is to a standard measure or an existing factual value, the more that measured value can be considered accurate. Thus, it is essential, for example...:

- » To ensure that names are spelled correctly.
- » That addresses correspond to real addresses.
- » That the total number of sales reflects the correct sales totals.

#### 2. Completeness:

This dimension assesses whether all the required data for a particular task or purpose is available. Completeness can be measured at a data set, record, or column level. Data is considered complete when there are no useless or missing elements, and it is possible to extract insights and make decisions from its analysis. Some questions that can be used to assess the degree of data completeness are:

- » Does a dataset have all the expected records?
- » Are records correctly filled out?
- » Are columns or attributes adequately populated?

#### 3. Consistency:

Primarily refers to the degree to which values for the same data are identical across all tables and databases. For example, if a company's sales data entered in a CRM shows a certain amount, this amount must be the same as the one entered as the sales data value in the accounting software.

In another sense, consistency also measures how uniform the data is in terms of size and composition at different points in time. For instance, an employee who loses access to the User and Permissions management system will also need to automatically lose access to a company's Enterprise Resource Planning (ERP) system.

#### 4. Timeliness:

Assesses how up to date the data is and how relevant it is to the decision-making process. The

timeliness of data thus depends on how often the data is changed and why. On the one hand, it's essential to measure how up to date the data is in the face of changes in the reality to which it refers (data currency). For example, if a supplier's payment terms have not been updated for years, the data would be considered outdated since there could be a potential opportunity to renegotiate contracts with the supplier to obtain better terms and conditions.

Another aspect of this dimension assesses whether the most current values for a given data set are readily available whenever needed. For example, imagine a retail company that receives daily sales data from its stores is affected by delays in data collection and processing, and the sales data is only available for analysis three days after the transaction takes place.

#### 5. Uniqueness:

Indicates whether a value of a data element is unique, i.e., that there are no duplicate or overlapping values. If there are, you need to remove them. Data uniqueness is assessed either at the level of all records in the same dataset or across multiple datasets. In the case of a telecommunications company's Customer Relationship Management (CRM) system, it is essential to ensure that each record represents a single customer or entity. Otherwise, it will be difficult to plan customized customer relationship strategies and do proper data analysis and reporting.

#### 6. Validity:

Measures the degree to which data values are consistent with rules, specifications, or standards. The conformity of data values may depend on characteristics such as data type, description, size, format, etc. For example, in SQL relational databases, a value of the CHAR data type cannot exceed 255 characters. Similarly, if the ZIP code of a record for a US address has more than five digits, we can conclude that the data is invalid since the US standard is for ZIP codes to have only five digits.





## Importance for business and decision-making

To better understand the significance of good-quality data, consider a few examples taken from our daily lives. Whether it's an inaccurate weather forecast, unreliable online product reviews, or outdated financial reports when buying shares in a company, all these situations show how poor data quality can negatively affect our life choices. Besides eroding our trust in the data sources, we can inadvertently introduce biases in your decision-making, leading to unfavorable consequences for the data providers.

Similarly, in the business world, the importance of data quality becomes even more critical. Organizations' decision-making relies heavily on data-driven insights and automation, where data is the foundation for strategic planning, operational efficiency, and growth.

Consider, for instance, a business that relies on automated systems to make decisions, such as loan approvals, customer segmentation, or pricing strategies. If the underlying data is flawed, biased, or lacks accuracy, it can lead to unfair outcomes, non-compliance with regulations, and ethical concerns. Furthermore, decisions based on inaccurate or unreliable data may lead to financial losses, missed opportunities, or damaged reputation.

# 3. The Business Costs of Poor Data Quality

To grasp the significance of data quality in daily business operations, we need to be aware of the negative consequences of poor-quality data.

While collecting and processing high-quality data demands attention and increased effort, mainly when dealing with data from different sources or domains, it's only through solid data governance practices consistent with the dimensions we just mentioned that organizations establish trust, make better-informed and ethical decisions, comply with regulations, and position themselves successfully in a data-driven world.

Based on Loshin's work (2011) , we can identify four significant impacts of poor data quality on business activities:

## 1. Financial impacts:

This category includes such issues as increased operating costs, decreased sales/revenues, missed opportunities (in terms of market trends, customer insights, and product improvements), overpayments, reductions or delays in cash flow, increased penalties, fines, or other charges, as well as low benefits realization.

## 2. Productivity impacts:

A decline in productivity ultimately affects the operational efficiency of an organization, which can lead to increased workloads, decreased throughput, extended cycle times, and various supply-chain constraints such as products going out of stock, delivery delays, missed deliveries, and duplicate costs for product deliveries.

## 3. Reputational impacts:

A company's reputation in the marketplace encompasses the confidence and satisfaction of not only its customers and employees but also its vendors and suppliers. When a company fails to meet the expectations of the different market stakeholders due to failures in its quality principles, this results in a decrease in organizational trust and credibility, as well as a low level of confidence in management reports, delayed or incorrect decisions, and an increase in the number of customer complaints.

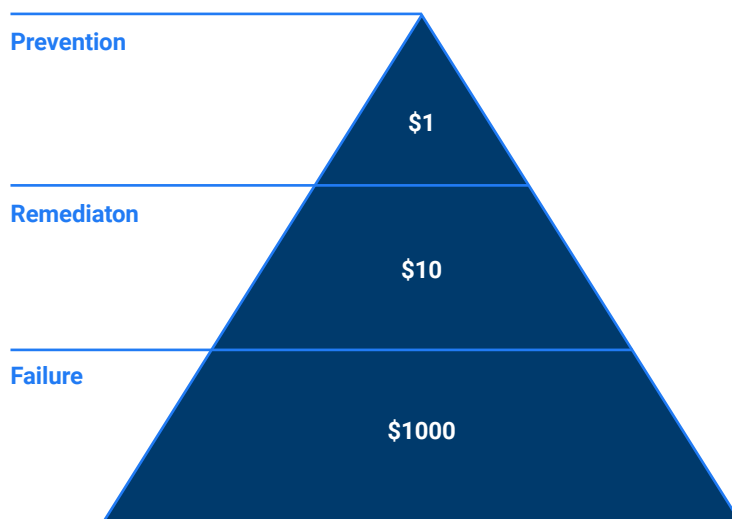
## 4. Risk and compliance impacts:

This refers to the exposure to several types of risks, including credit assessment, capital investment and development, fraud, data leakage, industry expectations, and non-compliance with regulations (such as the GDPR in Europe or the CCPA in California – USA).



The costs of poor data quality are so significant that it has been estimated that as a dataset progresses through the collection, storage, processing, and consumption phases, the costs of correcting it become higher and higher.

### 1-10-100 Rule of data issues management



This trend is captured by the 1-10-100 rule (Shoutekal, 2023, p. 54 and Mahanti, 2019, p. 36). According to this rule, the cost of verifying a set of records at the time of their insertion into the system is approximately one dollar. If the record is not verified, an organization will have to pay \$10 to remediate it. If no action is taken to clean up the records, the cost will escalate to \$100. This leads to a pyramid illustrating the increasing costs associated with poor data quality: Prevention cost, remediation cost, and failure cost in a ratio of 1:10:100.



## 4. The Role of Data Remediation in Enhancing Data Quality

As we explained, any business needs high-quality data to work optimally and make more sound decisions. Too often, however, companies lack a data-driven culture fostering collecting and storing high-quality data at the source.

In addition to the absence of standardized data formats, data duplication is also common. Likewise, it sometimes happens that data is not complete at the time of its collection.

When this happens, data needs to be remediated. In practical terms, this means applying a set of techniques called data remediation. These techniques are essential to help organizations ensure the quality of their data.

# What is Data Remediation?

In broad terms, it consists in identifying, cleaning, and correcting inaccurate, incomplete, or redundant data within a dataset. The process may involve:

- » Removing duplicates.
- » Correcting spelling or formatting errors.
- » Filling in missing information.
- » Deleting irrelevant data.
- » Etc.

This set of activities aims to improve data quality in all six dimensions mentioned above. Data remediation is thus crucial to ensure that reliable and trustworthy data serves as a basis for decision-making.

Although you can perform data remediation either manually, through data migration, or via command-line scripts, it's now increasingly common for companies to use specialized user-friendly tools that handle most or even all the tasks in the process: data validation, normalization, cleaning, enrichment, consolidation, and reconciliation. In addition, some comprehensive data quality platforms encompass data remediation capabilities. Most often, however, what happens is that a combination of all the methods mentioned above is employed.

## Key Benefits of Data Remediation

As a business activity, data remediation is often ignored and unplanned by most companies. Many organizations are reluctant to invest in improving the quality of their data for a variety of reasons, including the inability to plan for all of the data remediation activities required, the lack of formal data ownership, the perceived complexity or cost of the process, and the fear of potential discoveries at the end of the process.

Nevertheless, the available empirical evidence shows that the lasting benefits of data remediation outweigh the effort:

### Enhanced data security and reduced risk:

Data is either securely stored or deleted after remediation. In addition, unstructured data is classified and secured. This way, the chances of your organization suffering a data loss or being the target of a data leak or cyber-attack are substantially reduced.

### Reduced data storage and reputational costs:

Not only is your organization less likely to suffer financial losses due to fines, lawsuits, and reputational damage from a data leak, but data storage costs tend to be lower due to the reduced overall volume of data.

### Compliance with privacy and data protection rules and laws:

Whether it's the GDPR, CCPA, or HIPAA, the number of norms and regulations on this subject has increased in recent years, and changes to these rules after their implementation are becoming increasingly frequent. When performed frequently, data remediation processes ensure that companies keep up with and comply with these regulations.

### Improved performance:

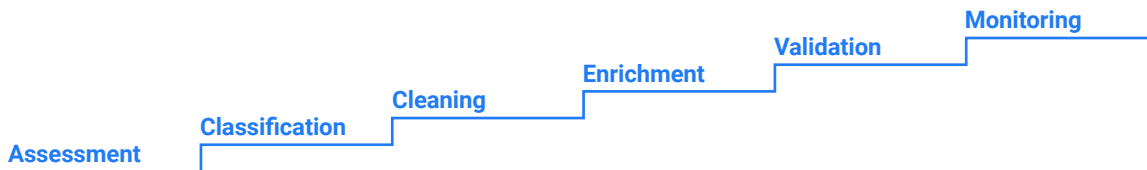
By adequately organizing data sets, employees will spend less time accessing and managing data, leading to substantial productivity gains.

In the next chapter, you will learn more about *Opplane's 6-Step masterplan for improving data quality through data remediation*, including...:

- A step-by-step guide on how to implement data remediation.
- *A brief presentation of the technologies and tools we commonly use.*

# 5. Improving Data Quality with Opplane's 6-Step Masterplan to Data Remediation

Although every data remediation project is unique, they all share the goal of identifying, analyzing, and correcting data quality problems to enhance data quality across its six dimensions. Opplane follows a consistent methodology for all remediation projects, which comprises of six steps: **Assessment, Classification, Cleaning, Enriching, Validation, and Monitoring**. Here is a brief description of each step.



## 1. Assessment:

The first step involves reviewing an organization's data, serving as the foundation for subsequent steps. The goal is to identify potential issues such as missing values, duplicates, inconsistent formats, or outliers. This step often includes a root cause analysis to identify the source or origin of data problems and understand the reasons for these errors.

## 2. Classification:

After identifying issues, Opplane's data analysts assign a classification scheme to organize the data based on its purpose and nature, making it ready for analysis. Although customizations may be needed for each project, Opplane uses a set of standard categories to classify data based on its sensitivity and value:

- » Essential business data for daily staff access.
- » Confidential data that needs to be securely stored.
- » Data containing business records suitable for archiving.
- » Redundant, Outdated, and Trivial (ROT) data eligible for deletion.

## 3. Cleaning:

During this phase, Opplane's analysts aim to find and correct errors in the raw data that may bias subsequent analysis. Tasks may include standardizing data formats, correcting misspellings, or typographical errors, removing duplicate records, and resolving conflicts or discrepancies between data sources.

## 4. Enriching:

Once the data is error-free, it can be enhanced with data from internal and external sources. By adding contextual information to existing data, new insights can be extracted from it. Data becomes more meaningful and valuable, enabling informed decision-making. This phase often requires specialized data science and machine learning skills and Opplane's teams have extensive experience in data transaction enrichment. Contact Opplane to learn more about our outstanding experience in data transaction enrichment.

## 5. Validation:

In this step, Opplane's teams implement validation rules and check if the remediated data complies with various quality metrics. It is also common to test the performance of the remediated data across the six data quality dimensions: accuracy, completeness, consistency, timeliness, uniqueness, and validity.

## 6. Monitoring:

Even after validating all the previously cleaned and enriched data, the remediation process is not complete. Since new data is continuously created and consumed, regular evaluations of existing data are necessary to ensure it serves its intended purpose.





## A Primer on Data Remediation Tools and Technologies

Before broadband internet and smartphones, data remediation was manually performed. Nowadays, however, organizations deal with too much data for it to be manually remediated. Empirical data has shown that business users can spend up to 50% of their time fixing data quality errors (Mahanti, 2019). Hence the enormous market growth for scalable platforms equipped with rich graphical interfaces that can be easily deployed across an organization. Despite their sometimes initial high cost, in the medium to long run, these solutions can help save time and money on data remediation efforts. However, not all organizations have complex and intricate data quality requirements to justify the initial investment.

Furthermore, even large organizations can sometimes exclusively rely on a “quick and dirty” command-line script written in Python or R. These scripts can automate tasks such as removing duplicate data, validation, and formatting based on predefined rules and logic. Although the cost is not an issue, sometimes organizations lack the required skills and expertise. If that is your case, please don’t hesitate to contact Opplane.



The bottom-line message is that you need to be aware of the requirements of your entire organization when choosing a data remediation technology. Throughout its existence, Opplane has worked with organizations of different sizes, each with its distinctive set of business practices, data quality requirements, and levels of data complexity. This awareness has led to the adoption of different tools, each with unique capabilities. Here is a brief description of the technologies we have used in straight collaboration with clients:

**Master Data Management (MDM) tools** aim to manage and maintain a trusted single source of truth across an organization so that there are no multiple inconsistent versions of the same data in different systems. Besides data consolidation and deduplication features, they also assist in data remediation efforts by helping to manage data hierarchies and improving data consistency and integrity.

**Data Integration and ETL (Extract, Transform, Load) tools** help fix data inconsistencies in different IT systems. They facilitate data extraction, transformation, and loading from several sources into a central repository.

**Data cleaning tools** can identify and remove duplicates, standardize and validate data against predefined rules, and correct various other errors. Even when not using more specialized tools, more organizations rely on SQL statements that can be reused and shared - stored procedures - to complement manual and highly contextual data cleaning efforts.

**Data Governance tools** support establishing and enforcing data quality policies, standards, and processes. By helping monitor data quality, tracking data lineage, and ensuring compliance with data governance frameworks, data governance tools enable organizations to manage data quality proactively.

**Data Quality Dashboards and reporting tools** provide visual representations and reports on data quality metrics, issues, and remediation progress. These tools offer insights into the overall data quality landscape.

**Machine Learning and Artificial Intelligence-based tools** can be leveraged by data remediation initiatives to automate the identification of data quality issues, predict potential errors, and suggest remediation strategies.



# 6. Best Practices for Data Remediation and Data Quality Improvement

Data remediation can be highly complex. In addition, it requires qualities such as teamwork and an understanding of data governance principles. However, with the appropriate planning and forecasting skills for the success of your data remediation project, your organization can significantly reduce the complexity. It is, therefore, crucial to implement a set of best practices for data remediation to improve data quality. Companies that follow the strategies outlined below will be better positioned to obtain more accurate, reliable, and consistent data to support informed decision-making and business.

Here is the list:

## 1. Prioritize:

You should explain to other team members the importance of prioritizing data quality issues. This prioritizing should consider the impact of each of these obstacles on business activity, compliance requirements, and the organization's strategic goals.

## 2. Invest in data:

Being an activity that occurs during the classification phase of the data remediation process, segmentation is essential as it is the basis for the remediation methods you will apply to different data. Investing in segmentation training for your staff will help increase the consistency of your data and avoid disorganization.

## 3. Focus on Root Cause Analysis:

The best way to prevent the recurrence of data quality issues is to identify and address its root causes.

## 4. Collaborate with other stakeholders:

As data remediation is a team effort, you need to try to engage with business stakeholders, data owners, data custodians, and IT teams. Only by adopting a comprehensive and coordinated approach to data remediation can you solve all your organization's data quality challenges.

## 5. Ensure comprehensive documentation:

To avoid the recurrence of any data quality challenges, your team must document every step of the data remediation process thoroughly. The final document should clearly outline the decisions made, the team members responsible for each decision, the rationale behind those decisions, the methodology used for classifying data records, and the specific data remediation techniques employed. In addition to providing transparency and the ability to replicate successful methods, documentation protects your organization from any eventual litigation or regulatory challenges (defensibility).

The efficiency level of these strategies will rely on correctly implementing data governance policies and processes. We will cover these and other data governance aspects in a future Opplane white paper. For now, it's enough to be aware that the end goal of data governance is to ensure an organization can effectively manage its data. To reach this objective, it is essential to designate specific team members as responsible for data ownership and stewardship. This will ensure clear accountability for maintaining data quality.

# 7. Conclusion: Remediate Before Bad Data Ruins Your Business

If there is one message you should take away from this white paper, it is this one: With the continuous surge of data availability in recent decades, there is no indication of it slowing down. To stay ahead, your organization must prioritize data quality. Ensuring your data is of high quality will make it valuable for decision-making, preventing it from becoming a liability.

We have started this white paper by *highlighting the value of data in the 21st century* and showing that while there is plenty of data, much of it is raw data requiring proper processing and management for it to be useful.

In the *second chapter*, we presented the concept of data quality while focusing on its importance for businesses. Even though data quality issues are almost unavoidable in the life of a contemporary organization, those organizations that prioritize data quality management face fewer obstacles. Next, we stressed that while data quality is contextual and depends on an organization's needs, you can always measure some of its aspects. We call these aspects data quality dimensions: accuracy, completeness, consistency, timeliness, uniqueness, and validity. We then provided examples highlighting the significance of each of these dimensions. Finally, we showed how flawed data can lead to adverse outcomes such as non-compliance, ethical concerns, financial losses, missed opportunities, and reputation damage.

We then looked more in-depth at the business costs associated with poor data quality in the *third chapter*. Specifically, we identified four significant impacts of poor data quality: financial impacts, productivity impacts, reputational impacts, and risk and compliance impacts. Finally, we ended this chapter by mentioning the 1-10-100 rule (according to which verifying records during insertion costs approximately \$1, remediation costs \$10, and failure to address poor data quality costs \$100) to illustrate the escalating costs of poor data quality.

Data remediation and its role in enhancing data quality was the focus of the *fourth chapter*. The process involves identifying, cleaning, and correcting inaccurate, incomplete, or redundant data within a dataset. Activities like removing duplicates, correcting errors, filling in missing information, and deleting irrelevant data are essential for improving data quality across its six dimensions.

Despite the common reluctance by organizations to invest in data remediation, empirical evidence suggests that data remediation plays a vital role in the life of an organization by improving decision-making, reducing risks, ensuring compliance, minimizing costs, and enhancing overall performance.

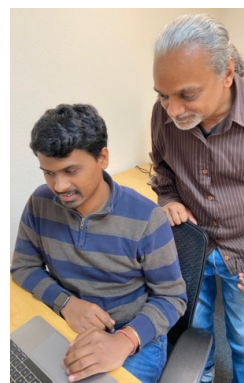
In the *fifth chapter*, we discussed Opplane's proven approach to improving data quality through data remediation. This process involves six stages: assessment, classification, cleaning, enriching, validation, and monitoring. While organizations can use manual or automated tools for these tasks, many now prefer scalable platforms with user-friendly interfaces to ensure efficiency. Although they may require high initial costs, such solutions ultimately save time and money. If you feel your organization needs more skills and expertise, feel free to contact Opplane for assistance.

Finally, in the *sixth chapter*, we highlighted how organizations can improve their data's accuracy, reliability, and consistency by implementing best practices for data remediation. These practices include prioritizing data quality issues, segmenting data, analyzing root causes, collaborating with stakeholders, and thoroughly documenting each step of the process. Implementing these strategies effectively relies on proper implementation of data governance policies and processes.

As we conclude this white paper, we want to emphasize the importance of dedicating more resources to data remediation. Doing so will significantly improve your organization's chances of success. By prioritizing data quality, you can prevent poorly managed data from limiting your organization's potential. We urge you to address any data issues promptly to avoid a negative impact on your business.

# References

- Arthur, Charles (2013, August 23). "Tech giants may be huge, but nothing matches big data.", The Guardian. <https://www.theguardian.com/technology/2013/aug/23/tech-giants-data>.
- DAMA International (2017). DAMA-DMBOOK: Data Management Body of Knowledge. Basking Ridge, NJ, Technics Publications.
- IBM (2016). "The Four V's of Big Data", IBM Big Data & Analytics Hub. <https://web.archive.org/web/20201012170526/https://www.ibmbigdatahub.com/infographic/four-vs-big-data>.
- IDC (2021). "Data Creation and Replication Will Grow at a Faster Rate Than Installed Storage Capacity, According to the IDC Global DataSphere and StorageSphere Forecasts", Business Wire. <https://www.businesswire.com/news/home/20210324005175/en/Data-Creation-and-Replication-Will-Grow-at-a-Faster-Rate-Than-Installed-Storage-Capacity-According-to-the-IDC-Global-DataSphere-and-StorageSphere-Forecasts>.
- Loshin, David (2010). The Practitioner's Guide to Data Quality Improvement. Burlington, MA: USA. Morgan Kaufmann.
- Mahanti, Rupa (2019). Data Quality: Dimensions, Measurement, Strategy, Management, and Governance. Milwaukee, Wisconsin, ASQ Quality Press.
- Southeikal, Prashanth H. (2023). Data Quality: Empowering Businesses with Analytics and AI. Hoboken, NJ. John Wiley & Sons.
- Strong, Diane M. and Wang, Richard Y. (1996). "Beyond Accuracy: What Data Quality Means to Data Consumers" in Journal of Management Information Systems, Vol. 12, No. 4, pp. 5-33. [http://mitiq.mit.edu/documents/publications/tdqmpub/14\\_beyond\\_accuracy.pdf](http://mitiq.mit.edu/documents/publications/tdqmpub/14_beyond_accuracy.pdf).



## About Opplane

Opplane is a company led by former PayPal leaders who bring years of leadership experience from various domains and top organizations. Our experienced team specializes in data modernization programs, covering areas such as privacy management, banking, fintech, security, machine learning, risk management, and cloud infrastructure. This expertise ensures comprehensive solutions for data modernization initiatives.

Our company operates globally with offices in Silicon Valley (USA), Western Europe, India, and Singapore. This wide geographic presence brings numerous advantages, including access to diverse talent, continuous support, and customized solutions for regional markets. Leveraging the expertise of our international team, we excel in delivering exceptional results at scale.

We assist organizations with data transformation and digitization, providing comprehensive solutions including end-to-end implementations of cloud-native platforms, data lakes, and privacy tools. Additionally, we utilize machine learning algorithms to gain consumer insights and enhance consumer experiences through personalization.



**Address:**

**US office:**

Suite 300, 6200 StoneRidge  
Mall Pleasanton,  
CA 94588

**Email:**

[contact@opplane.com](mailto:contact@opplane.com)

**Website:**

[www.opplane.com](http://www.opplane.com)

**Portugal office:**

Rua Camilo Castelo Branco 2  
3º esquerdo  
1150 - 084, Lisboa

**India office:**

204, Express Arcade  
Netaji Subhash Place  
Plot No. H10, Pitampura Delhi  
North West, Delhi, 110034