



---

# Standardized **International Data**

To Normalize or Standardize – that is the question  
(and one that makes all the difference).

# Standardized International Data

To Normalize or Standardize – that is the question  
(and one that makes all the difference).



## The Cost of Bad Data

Garbage in, garbage out, or GIGO as it's often said. No where else does this ring more true than for data – in all its applications and use cases. According to Gartner's 2020 Magic Quadrant for Data Quality Solutions, marketers incur hard costs to their bottom line to the tune of \$12.9 million+ each year due to poor quality data. And, as digitization continues to accelerate, and more and more data is available and collected, losses from poor quality data will only increase.

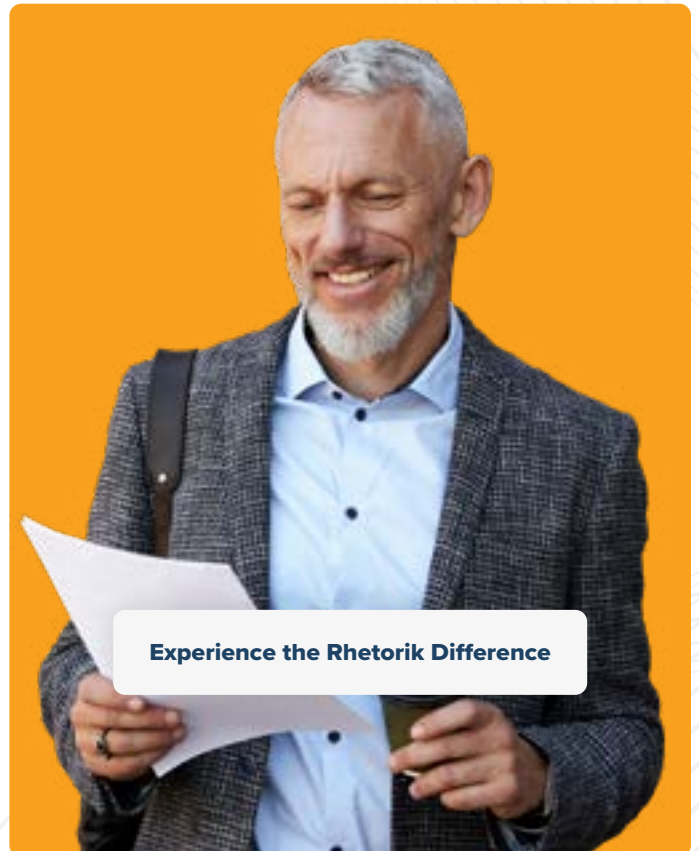
**According to CMSWire, a recent B2B revenue marketing study found that 82% of marketers are measured on pipeline initiated, which is a 12% increase over 2021. For revenue of any kind, you still need a high-velocity pipeline. And that still needs high-quality leads.**

**“For revenue of any kind, you still need a high-velocity pipeline... and that still needs high quality leads.”**

## The case for Data Standardization – why Normalization isn't enough.

Most organizations utilize data from a number of sources; this can include both first and third-party data from data warehouses, lakes, cloud storage, and databases from CRM and marketing automation systems. However, data from disparate sources can be problematic if it isn't uniform, leading to difficulties down the line for sales, marketing, and product development professionals.

Data normalization, in its simplest definition, refers to the process of data cleansing; Cleaning up data and removing the noise and mistakes to ensure format consistency. For example, it's very common for databases, especially those in CRM systems, to contain records that have different data formats. Let's look at titles, for example. One record may say “Sr. Engineer”, while others may contain alternative titles – such as “Sr Engineer” (no period), “Senior Engineer”, or “Senoir Engineer” (typo intended). Normalizing the data cleans this up for future processing – such as using “Senior Engineer” consistently in the examples above. Normalization is the first step in the data hygiene process, and is absolutely critical to sales, marketing and product development efforts.



**Experience the Rhetorik Difference**

# However, it's not enough. Introducing Standardization.

Normalization will only get you so far. What happens when a title, such as “Senior Engineer”, means something completely different in another language, geography, or industry? For example, “Senior Engineer” could have the same responsibilities as a “Senior Developer” or a “Staff Engineer”, also called “SDE 3” or “Software Developer Level 3” in some global companies. Would you want to exclude those titles in a critical marketing campaign targeted to individuals with certain roles, responsibilities, or decision-making authority? Of course not.

Let's look at another example.

**In the United States, the title of “CEO” is well understood. By normalizing the data we will ensure we capture the “Chief Executive Officer” equivalents within the database. However, what normalization will not catch are the titles of “PDG”, or “Président Directeur Général”, which are the CEO equivalents in France. “CFO” is another well-known title in the US, and normalization will ensure to recognize “Chief Financial Officer” as the same thing. But once again normalization will fall short, as the title of “Director of Finance” is the CFO equivalent in many other parts of the world.**

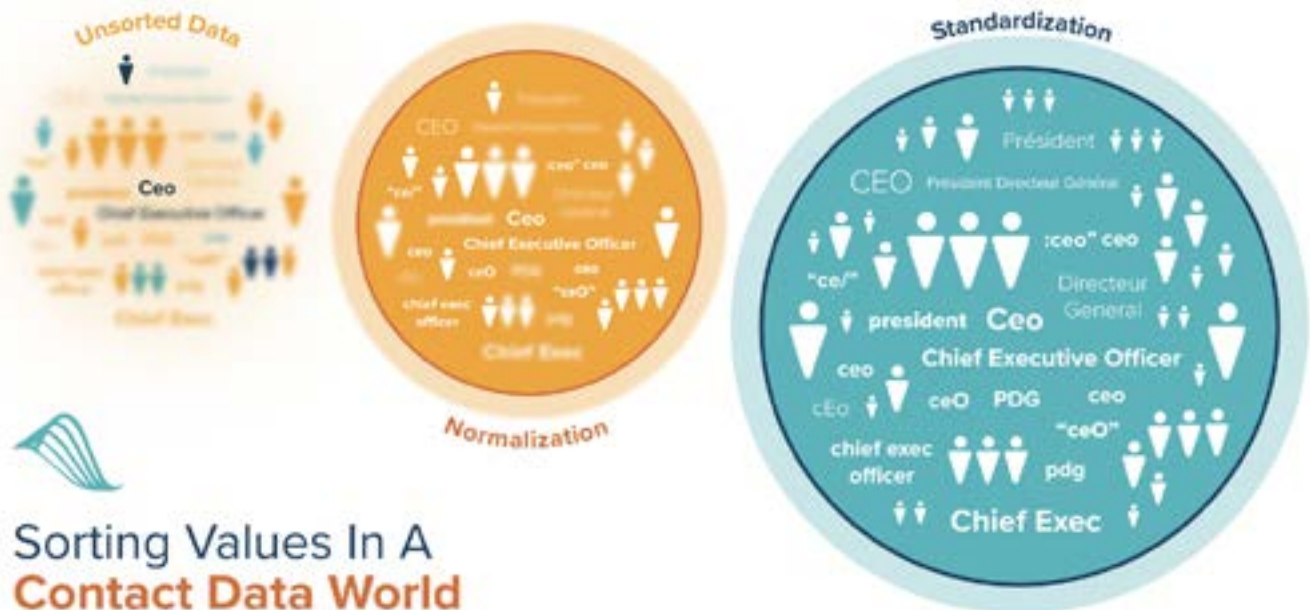
In layman's terms, standardization is the process of matching data signals within the database records, such as roles, responsibilities, and experience, to ensure that marketers and product developers receive

- 1 The most accurate data datasets
- 2 Increased reach

Standardization provides a longer tail of critical job titles, and after all, isn't that what sales, marketing, and product leaders are ultimately after?

The effectiveness of your marketing and product efforts hinges on the usefulness of your underlying data. And, unless it is in a consistent format that addresses the needs of all users, you will never be able to effectively analyze customer behavior, predict trends, nurture, target, and drive conversions.

Data standardization is the key, and the secret-sauce, to valuable and high impact third-party data. It establishes clear, consistently defined data elements and attributes, providing a comprehensive catalog of your data. Whatever insights you're trying to get, or problems you're attempting to solve, data standardization enables you to properly understand your data across languages and geographies.



## First-Party Data and Data Partners

Marketing and Sales Leaders have prioritized first-party data collection in the quest for deeper direct customer knowledge. However, more often than not, as in the case of data normalization, first-party data isn't enough. While this data goldmine can help companies get closer to customers and improve products and services, the vast majority of companies are discovering that their databases and datasets are too fragmented to deliver the promised results. Augmenting the company database with third-party data, from a trusted source, is critical to marketing and product development success.

## Data Partners: How to choose

The heart of the problem with accurate data lies in the quality of the data input into AI-ML algorithms. Marketing datasets and databases are created with different data standards, or no data standards at all. Despite the best of intentions, AI-ML investments (and marketers) are being set up to fail.

**In 2020, O'Reilly's The State of Data Quality found that more than 60% of AI-ML projects failed due to siloed and inconsistent data.**

The most advanced, and hence most effective, providers of third-party data use both normalization and standardization within their AI and ML engines.

"Poor data quality is enemy number one to the widespread, profitable use of machine learning." — Harvard Business Review

## The Rhetorik Difference

Rhetorik's Occupation and Classification (ROC) process automatically normalizes and standardizes jobs from 47 different languages in 164 different countries. Truly a global solution, the ROC was designed for flexibility and scale, continuously recognizing new job titles and updating the Rhetorik ML model in real-time. The ROC also contains over 286M unique job titles and 1.33Bn "experiences", which are parsed excerpts from profile resumes that contain relevant keywords. What's the result? Rhetorik surpassed Amazon Web Service's AI/ML model by over 5%.

## Experience the Rhetorik Difference for Yourself

Visit us Online to See the Difference

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

“

“Poor data quality is enemy number one to the widespread, profitable use of machine learning.”

- Harvard Business Review

## Trusted By



## Seamlessly Integrates With All Your Existing Tools



Rhetorik provides enterprise grade solutions for all users of global profile data.  
Get started at [rhetorik.com](http://rhetorik.com) or +44 (0)118 989 8580