

# Understanding Data Collection, Minimization, Retention, Deletion & Necessity

## Did you know?

Currently, there are **over 147 zettabytes of data** in the digital universe, and more are created every day.<sup>1</sup>

Despite that, AI developers claim they are running out of data to train their models.<sup>2</sup>

*However, collecting massive amounts of data without a purpose or to maybe use it in the future comes with risks.*



## The risks of collecting too much data



### Data Accuracy & Quality

- Ensuring data accuracy at scale is challenging
- "Garbage in, garbage out" – poor quality data degrades over time



### Data Volume

- Storing large amounts can be expensive
- Requires scalable and cost-effective storage solutions



### Privacy Infringement

- Improper handling can infringe on privacy rights
- Risks increase with third-party sourced data



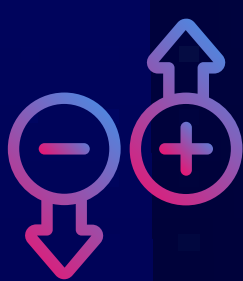
### Security

- Large amounts of data increase breach risk
- Breach costs include fines and loss of trust



### Data Integration

- Integrating data from multiple sources is complex
- Creating a complete business picture is challenging



## Why less is more

- **Reduce noise:** Focus on relevant information
- **Faster decision making:** Fewer errors, quicker insights
- **Cost reduction:** Lower storage and maintenance costs
- **Increased security:** Fewer records reduce breach risk
- **Environmental impact:** Lower digital carbon footprint

## 5 tips for implementing data minimization

1

### Review acceptable data use policies

- Limit data collection to necessary and relevant purposes
- Ensure data is deleted when no longer needed

2

### Embed privacy by design

- Integrate from the start with product/service development
- Raise awareness among employees about data minimization

3

### Leverage data minimization techniques

- **Anonymization:** Remove sensitive identifiers
- **Pseudonymization:** Replace personal info with unique codes

4

### Conduct a data system inventory

- Include third-party vendor systems
- Understand what data is collected and its purpose

5

### Use risk assessment processes

- Conduct PIAs and DPIAs
- Assess data collection risks and necessity

## Mitigating risks & enhancing trust

Save time and reduce privacy risk with TrustArc's Data Mapping & Risk Manager. Achieve continuous compliance with automated data flow mapping, risk analysis, and remediation.

[REQUEST A DEMO](#)

<sup>1</sup> Statista [https://www.statista.com/statistics/871513/worldwide-data-created/#--text=Data%20growth%20worldwide%202010%2D2025\\_replicated%20reached%20a%20new%20high](https://www.statista.com/statistics/871513/worldwide-data-created/#--text=Data%20growth%20worldwide%202010%2D2025_replicated%20reached%20a%20new%20high)

<sup>2</sup> The Week <https://theweek.com/tech/ai-running-out-of-data>