The ROI of privacy-preserving synthetic data

Evaluation guide

Statice

Content

Introduction

When data teams start evaluating privacy-preserving synthetic data, three factors will usually come into consideration:

- *Does privacy-preserving synthetic data address my challenges?*
- *Does the technology fit my requirements?*
- *What's the Return on Investment (ROI) of implementing privacy-preserving synthetic data?*

This presentation focuses on providing insights about the value of adding synthetic data generation capabilities to your team's toolbox.

The costs of data inertia and insufficient privacy-preservation mechanisms

Data is critical to the operations of data-driven companies. It is the key to gaining competitive advantages.

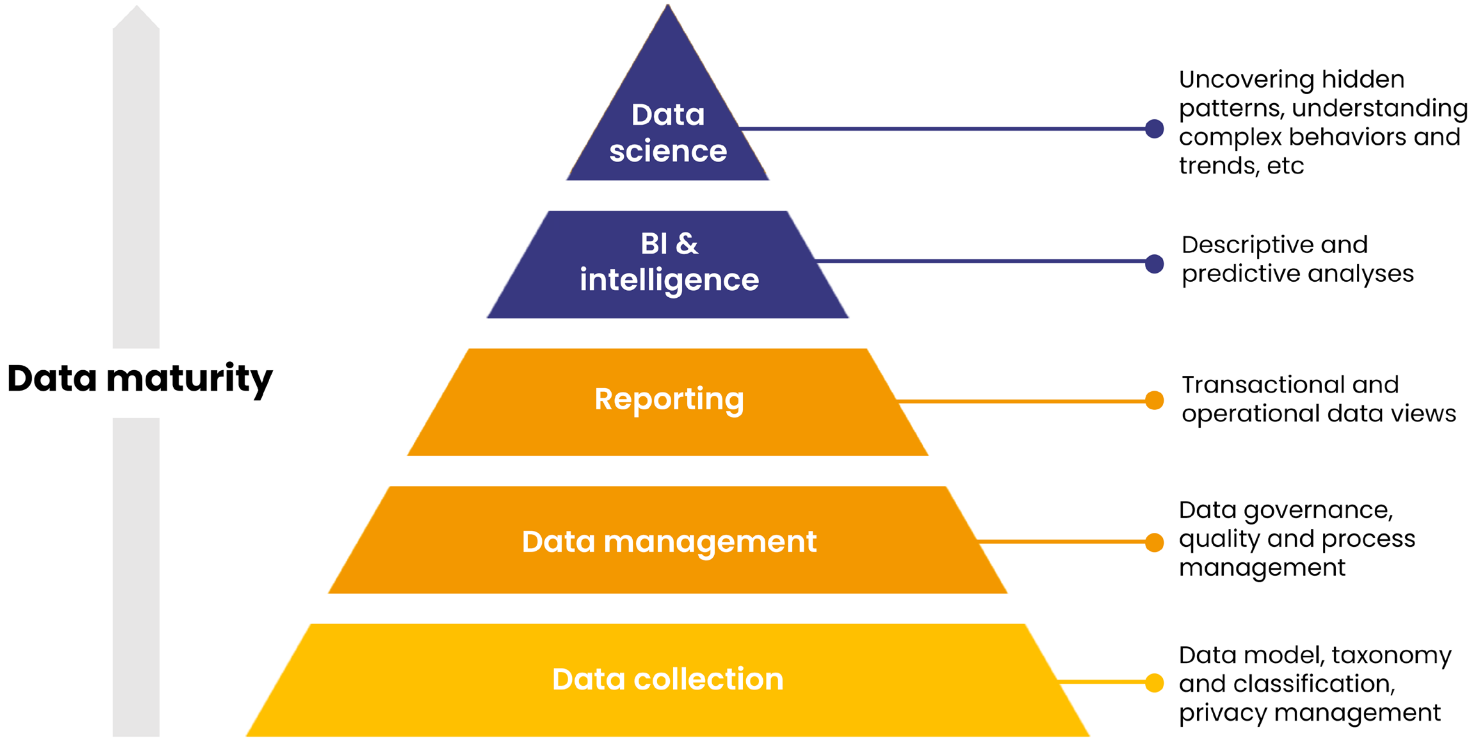
Forrester reported insights-driven businesses growing at an average of more than 30% annually[1].

However, many obstacles are preventing businesses from achieving the transition to being data-driven, costing them money and business opportunities.

The costs vary from one enterprise to another, but we usually see two cost types. The first one comes from the inability to leverage data and the second one from the compliance, brand, and social risks of using weak privacy-preservation mechanisms.

The inability to derive insights or drive innovation from data represents a high cost for businesses.

Data's value comes from our ability to transform it into insights to drive decisions and strategy, predict trends and revenue opportunities, build new products, and solve problems.


Data maturity
Data science
BI & intelligence
Reporting
Data management
Data collection
Uncovering hidden patterns, understanding complex behaviors and trends, etc
Descriptive and predictive analyses
Transactional and operational data views
Data governance, quality and process management
Data model, taxonomy and classification, privacy management

Data starts bringing business value as data teams derive insights from it.

The cost of inertia

The inability to derive insights can be due to several factors. In some cases, the data is too sensitive to be moved, processed, or used.

Data teams are then unable to access it. In these cases, you are passing on 100% of the value you could be generating.

Depending on the business priorities, the current state of data access and the target stage, data inertia can cost businesses a large chunk of revenues.

In cases where it's possible to access parts of it, the data has been stripped of so much information to protect sensitive information that it is of no use for analysis.

In these scenarios, you can generate a smaller value but never reach the data's full potential.

In other cases, siloed architectures and cumbersome governance processes obstruct data sharing. Requesting data is a slow process, when possible at all, that impacts timely decision making, innovation, and analysis.

In these scenarios, data teams' time and resources might outweigh the value derived from the data.

> ***For Ted Friedman[2], vice president and distinguished analyst at Gartner, "as organizations accelerate their digital business efforts, poor data quality is a major contributor to a crisis in information trust and business value, negatively impacting financial performance."***

The cost of weak privacy protection mechanisms

Other costs can be expected from the impact weak privacy protection mechanisms have on companies.

To protect sensitive data and comply with personal data protection laws, many choose privacy mechanisms. However most of the traditional privacy mechanisms can not guarantee total anonymity[3]. These technologies are putting data and businesses at risk. And these risks are associated with costs.

Companies using data masking or pseudonymization approaches expose the data to leakage and reidentification. Whether it concerns customer data or sensitive business information, the cost of a breach isn't trivial. Researchers at IBM reported that the average cost of a data breach to a business could climb up to $3.92 million[4].
In addition to security risks, European enterprises are evolving within strictly regulated environments. The General Data Protection Regulation (GDPR) adds up to national and industry-specific regulations. The costs of non-compliance keep on rising, as they include not only fines settlements but also business disruption, productivity, and revenue loss.

In a 2018 report, research firm Ponemon Institute and security company GlobalScape published that the annual cost of non-compliance to businesses now runs an average of $14.8 million[5].

Privacy-preserving synthetic data offers an alternative to using sensitive data. It allows companies to overcome their inability to derive insights from sensitive data and the risks associated with traditional privacy protection mechanisms.

Where synthetic data brings value

Companies that implement synthetic data as an anonymization method can expect a return on investment along two axes.

The first one is the easiest to quantify. It is a set of new revenues or reduced costs directly derived from their newly-gained ability to leverage data.

The second axis is harder to quantify but equally essential. It encompasses the risks and costs mitigation associated with insufficient data protection.

Impact of privacy-preserving synthetic data based on data challenges			**Common business data challenges**			
			Impact from weak privacy protection mechanisms		Inability to derive insights	
			Security & privacy risks	Legal implications	Impaired data quality	Slow data access
Benefits of privacy-preserving synthetic data	Risks and costs mitigation	Compliance	**HIGH**	**HIGH**	LOW	**HIGH**
		Data security	**HIGH**	**HIGH**	MEDIUM	MEDIUM
	Internal data agility	Ability to monetize data	MEDIUM	**HIGH**	**HIGH**	**HIGH**
		Product development	MEDIUM	MEDIUM	**HIGH**	**HIGH**
		Minimizing time to data	MEDIUM	MEDIUM	**HIGH**	**HIGH**

Where does privacy-preserving synthetic data adds the most value based on data challenges

The ability to derive insight

Using synthetic data to fuel data analysis represents a critical opportunity for companies.

Theoretically, this approach offers companies an unlimited data pool to power any data-driven analytics use-case and unlock the ability to derive insights from sensitive data. It grants companies the ability to go from «I can't do anything with my data» to «I have a compliant, safe and usable data source».

This data source offers a significant analytics valuethanks to its non-sensitive nature. It holds the same statistical properties as the original data, without any risks associated with privacy violation. Correctly generated synthetic data will power any applications intended for the original data, with only minimal performance losses.

The fact that it offers a higher value is true whether we measure the use of privacy-preserving synthetic data against the original data itself or other protection mechanisms.

When compared with using sensitive data, privacy-preserving synthetic data will always win value-wise. The information is too sensitive to be used in the first place. Suppose we measure the value against other protection mechanisms. In that case, privacy-preserving synthetic data offers a higher value because it generates a more accurate, private, and easy-to-handle analytical resource. In most cases, alternative options induce a more extensive loss of utility of the resulting data or force compromises in privacy protection or data agility.

With privacy-preserving synthetic data, companies can drive internal data agility, saving costs by cutting their time-to-data.

Because privacy-preserving synthetic data is private by design, companies reduce their teams' time on internal governance and clearance processes when it comes to data sharing.

Another direct ROI is the possibility of creating new revenue streamsfrom data monetization or product development.

According to Accenture[6], 40% of financial institutions worldwide are now investing in data insight offerings.

You won't be able to monetize individual personal data without consent and several layers of privacy protection. But what you can derive value from is the underlying statistical patterns, which are of course preserved in the synthetic data.
In this regard, privacy-preserving synthetic data represents a high-potential opportunity for companies storing large volumes of personal data.

Being able to share customer or market data without privacy concerns is an essential building block to being able to capitalize on this trend.

Risk mitigation

By nature, privacy-preserving synthetic data is completely anonymous and has no one-to-one relationship with the original data.

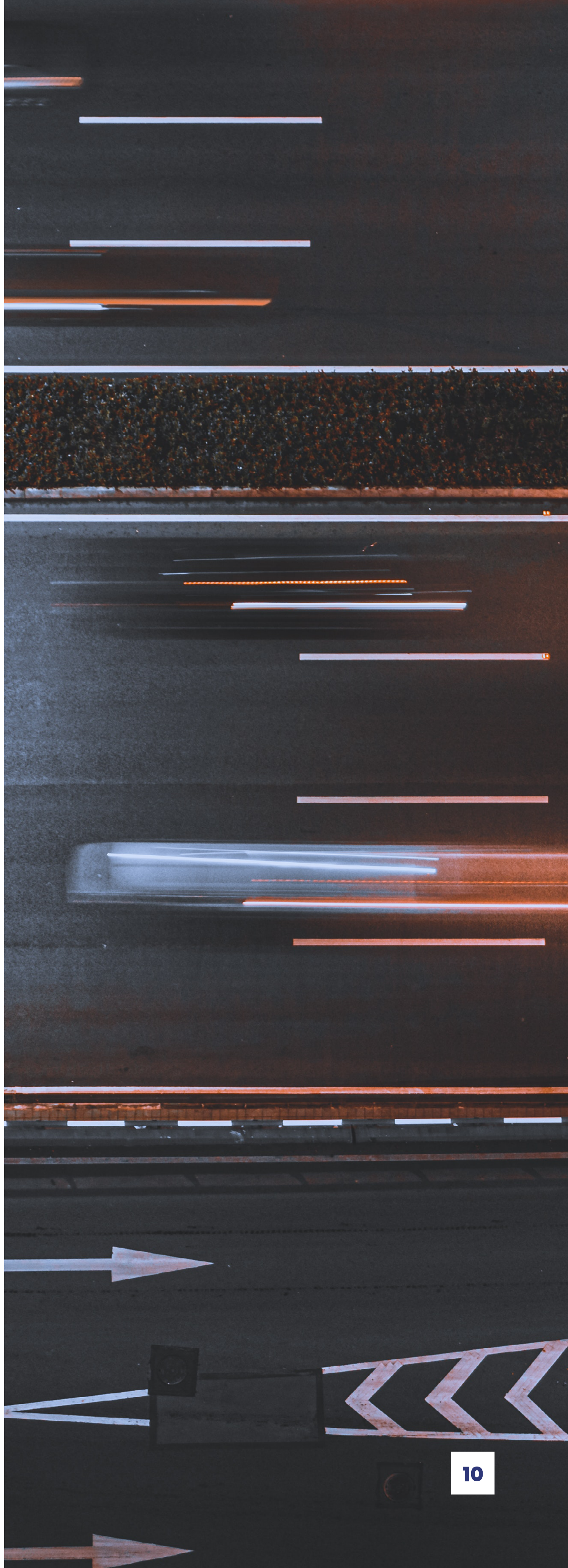
Companies using this technology can generate data that falls out of the scope of personal data regulations, limiting non-compliance risks. They also get stronger privacy guarantees than traditional masking and de-identification methods, mitigating the potential damages if breached.

A recent study conducted by the Ponemon Institute estimated that customers' PII was the most expensive data to lose in a breach, estimated at a recovery cost of about $150 per record[7]. That cost rose to $175 if a malicious attack caused the breach.

While working with synthetic data doesn't directly prevent hack attempts on personal data, it's a way of limiting unnecessary exposure of sensitive data.

Additionally, companies using privacy-preserving synthetic data can rule out any legal violation of privacy laws, at least for the associated use-cases.

Processing data that underwent pseudonymization is, by contrast, more heavily regulated and can be non-compliant with privacy regulations.



The gain of customer trust also represents a significant value for companies, although it's hard to quantify. In a world plagued by data breaches and scandals, customers are less inclined to trust businesses with their money and confidential information.

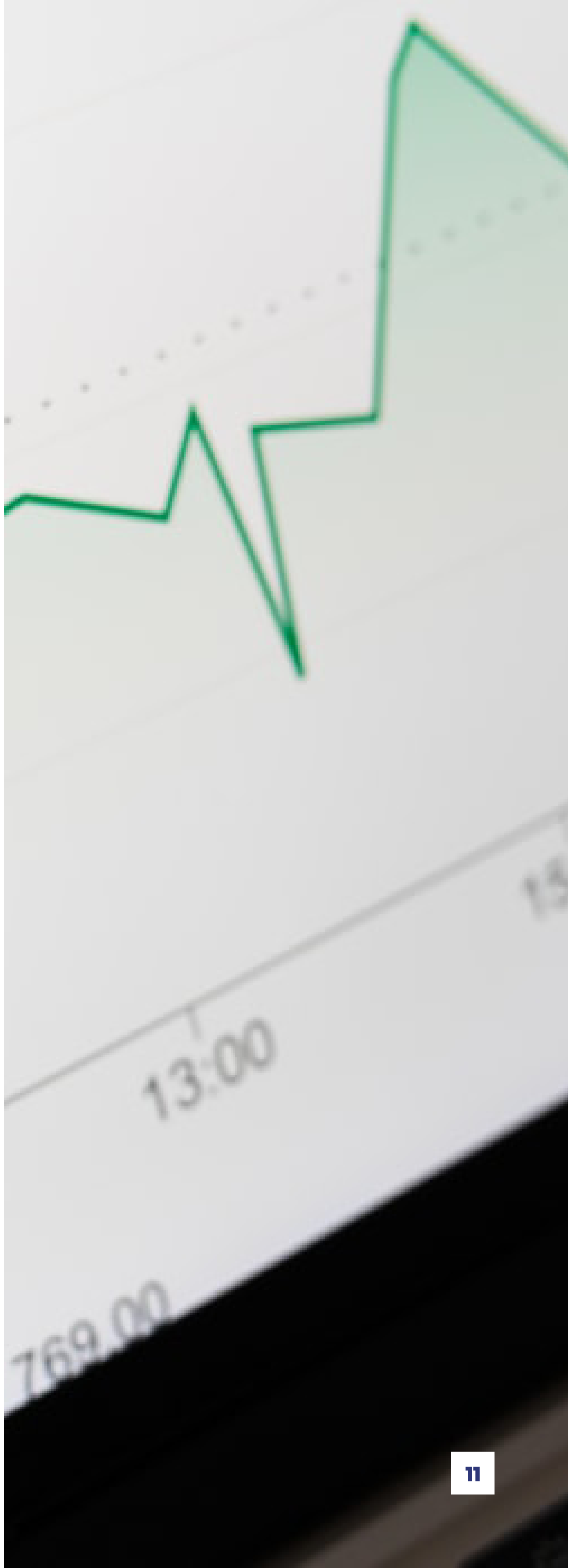
The EU Agency's Fundamental Rights survey reported that *"41% [of Europeans] do not want to share any personal data with private companies, almost double the number compared to public bodies"*[8].

And forrester's analysis showed that better privacy practices directly impact customers' purchases and willingness to provide personal data.

There are, of course, costs associated with implementing new technology. However, solutions like Statice's solution are designed to be as straightforward as possible regarding integration and usage. The costs and benefits are also highly tied to multiple factors specific to each industry and company.

From our experience, privacy-preserving synthetic data remains an opportunity for all the enterprises faced with data inertia and data privacy concerns.

Want to assess the ROI of synthetic data for your company? Talk to our team!



About
Statice

Statice develops state-of-the-art data privacy technology that helps companies double-down on data-driven innovation while safeguarding the privacy of individuals.

With Statice, companies generate privacy-preserving synthetic data compliant for any type of data integration, processing, and dissemination. Enterprises from the financial, insurance, and healthcare industries drive data agility and unlock the creation of value along their data lifecycle.

Start training your machine learning models, finally process your data in the cloud or easily share it with partners with Statice.

Contact us

www.statice.ai/contact/
hello@statice.ai

Find us

1. «Insights-Driven Businesses Set The Pace For Global Growth», Forrester, October 2018 [Online]
2. Ted Friedman's Recent Gartner Activity [Online]
3. «Which data protection techniques do you need to guarantee privacy?», Statice, September 2020 [Online]
4. «Cost of a Data Breach Report 2020», IBM & Ponemon Institute [Online]
5. «The cost of compliance with data protection regulations», Globalscape [Online]
6. «Global distribution and marketing consumer study: banking report», Accenture Financial Services [Online]
7. «Data protection and privacy - Fundamental Rights Survey», EUAFR [Online]