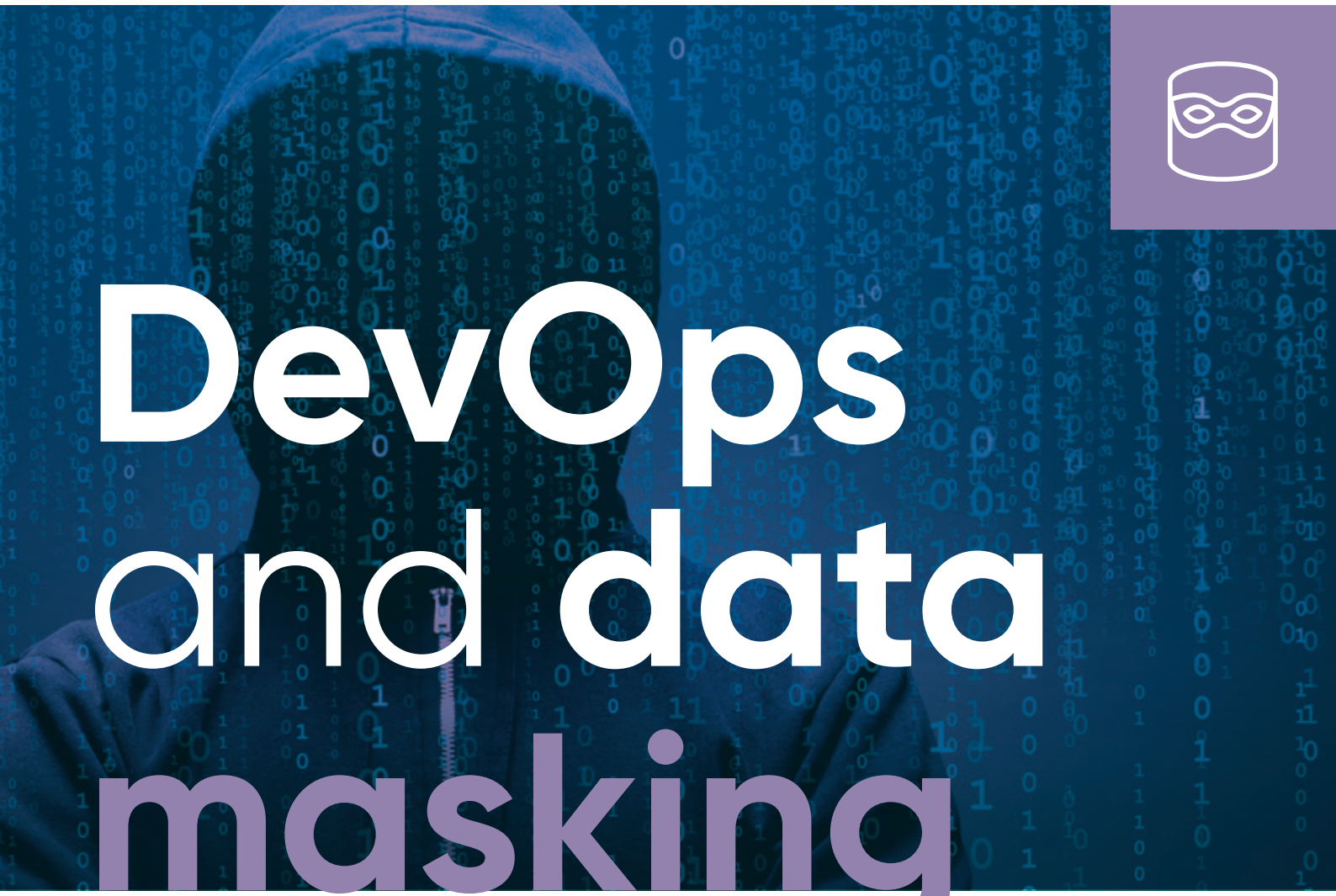
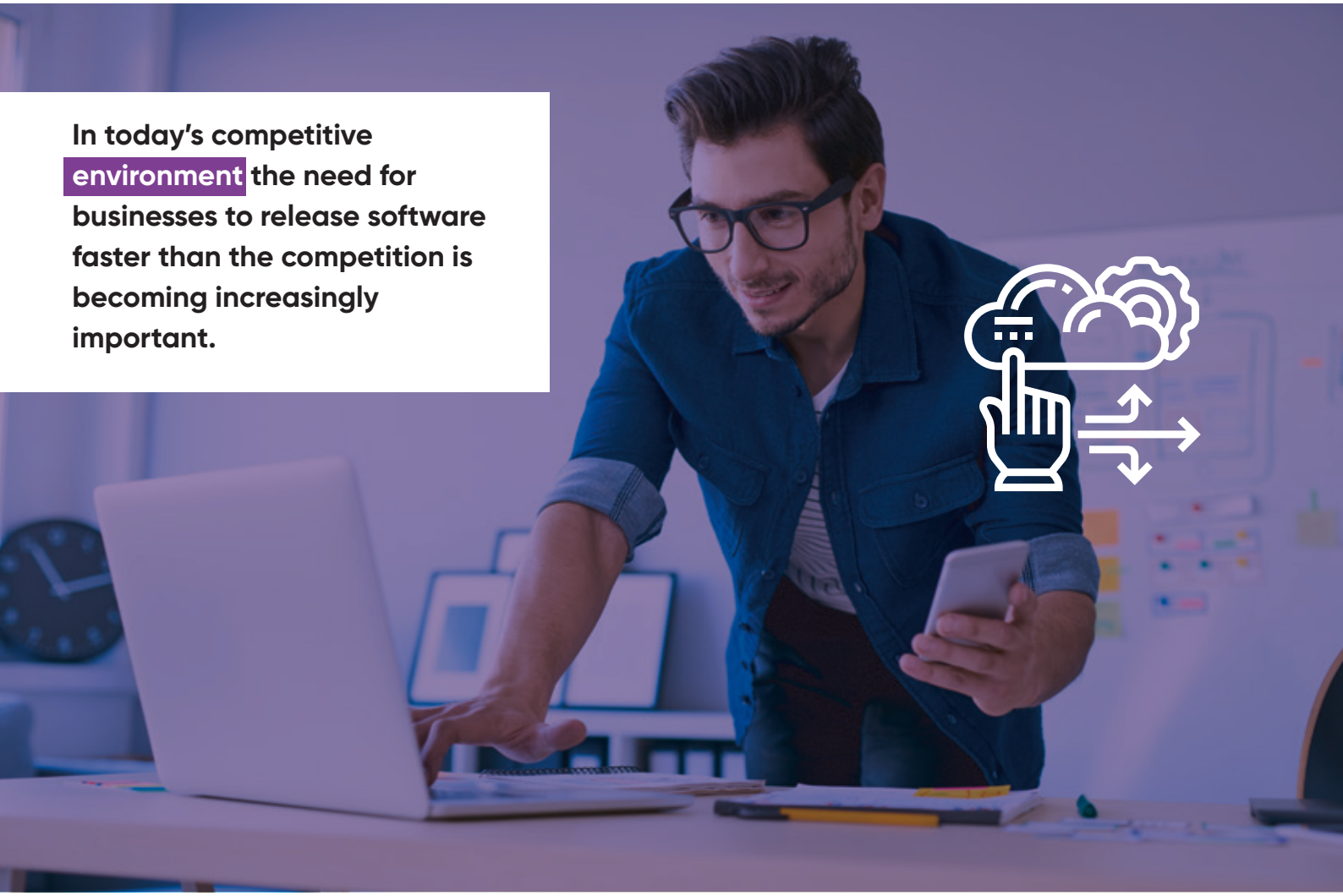
ACCELARIO



DevOps and data masking



In today's competitive environment the need for businesses to release software faster than the competition is becoming increasingly important.

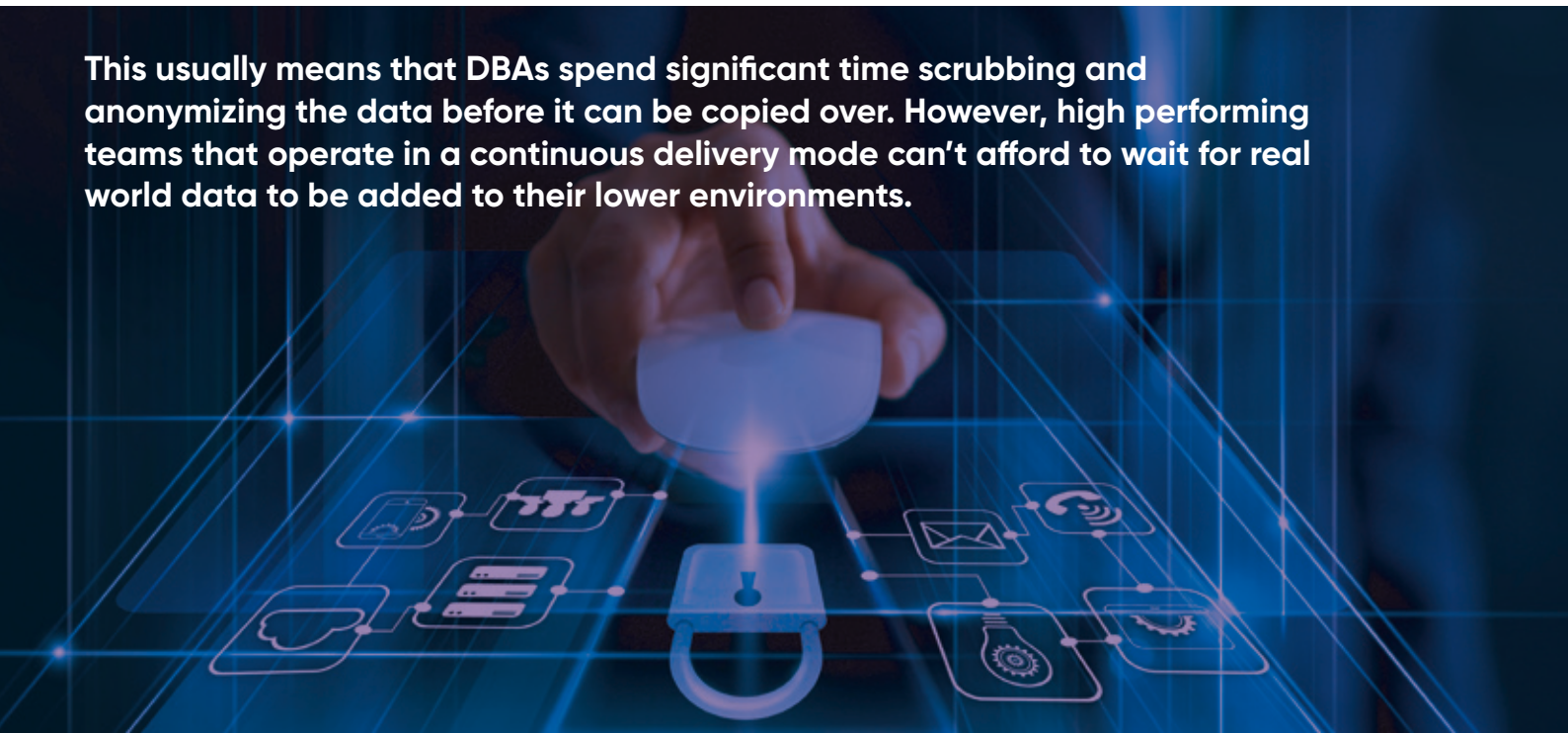
The requirement for ongoing innovation and customer-centricity is driving software companies to release features, updates, fixes and security patches on an almost daily basis

(just look at how frequently your phone applies app updates!).

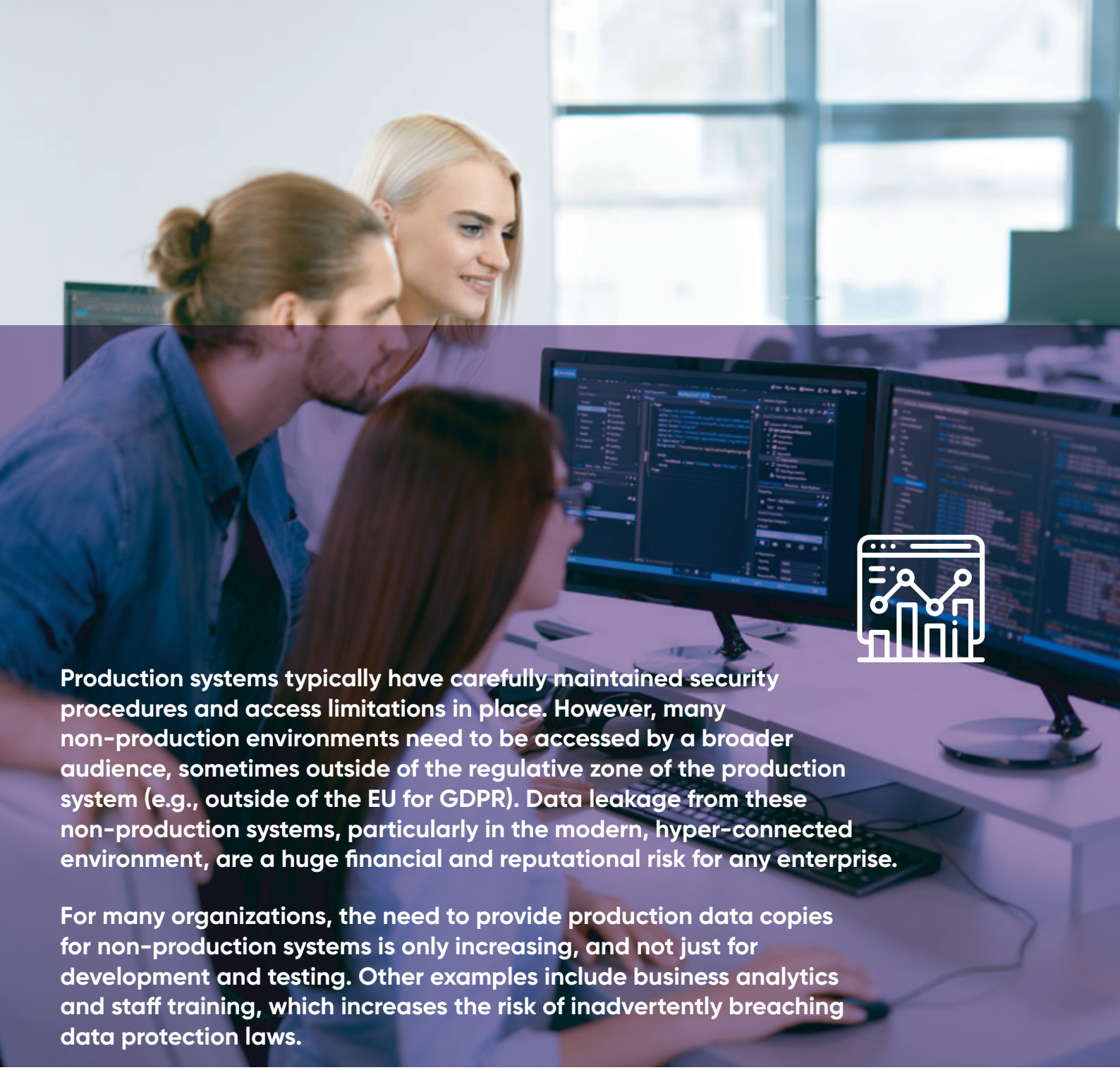
To ensure the delivery of a high quality product, engineering and quality assurance teams need access to production-like data to validate that new features are optimally developed in a customer-centric way.
Unfortunately, production data is often sensitive and private, requiring special handling in order to comply with regulatory guidelines. These regulations, including GDPR, CCPA, HIPAA and PCI-DSS, are designed to protect against the exposure of sensitive data.

Examples of sensitive data types include:

Payment card details	Health Information	Personally Identifiable Information (PII)



This usually means that DBAs spend significant time scrubbing and anonymizing the data before it can be copied over. However, high performing teams that operate in a continuous delivery mode can't afford to wait for real world data to be added to their lower environments.



Production systems typically have carefully maintained security procedures and access limitations in place. However, many non-production environments need to be accessed by a broader audience, sometimes outside of the regulative zone of the production system (e.g., outside of the EU for GDPR). Data leakage from these non-production systems, particularly in the modern, hyper-connected environment, are a huge financial and reputational risk for any enterprise.

For many organizations, the need to provide production data copies for non-production systems is only increasing, and not just for development and testing. Other examples include business analytics and staff training, which increases the risk of inadvertently breaching data protection laws.

Data masking automation represents a highly effective way for organizations to quickly and continuously provide access to representative data in non-production systems without risking the security of sensitive data.

ACCELARIO

What is data masking

What is Data Masking

Data masking is a way to change data so that it is representative of production data. This is achieved by fabricating data of the same type and format as the production data, while maintaining the structure mandated by the application. An example of this can be seen below:

Production

SSN	Last Name	First Name	Credit Card	Phone	Email
704-992-6441	Florence	Rachel	7609763516593819	311-294-7651	Aethaw6733@yahoo.io

Test

SSN	Last Name	First Name	Credit Card	Phone	Email
939-832-4119	Smith	Rachel	7609763296865614	311-219-7282	Hotkts4265@tnhcv.io

In this simple example you can see that the Social Security Number (SSN), Last Name, Credit Card, Phone and Email changed, while the First Name remained constant.

Many organizations create this masked data manually or via basic scripts. Unfortunately, these methods typically take up significant time and resources to produce high-quality, fabricated data. What's more, these methods are normally performed on a copy of the production data that has already been placed in a non-production environment. By extension this means that before the script is run there was a copy of the sensitive data in an environment that could infringe on regulations. Given the nature of non-production environments, this data could easily be vulnerable to a malicious attack from an internal or external actor.

Other than the security aspect of basic data masking there is the challenge of making sure the masking is done correctly and has meaningful context. Some data types are relatively easy to discover and mask in a rudimentary fashion. For example, a column labeled 'credit cards' in a table called 'customers' with a 16-digit number is likely to need masking. However, in many cases the data may not be labeled correctly, or is ambiguous, or the database is large and complex.

Let's take another look at the previous example. The second column is labeled 'Last Name' and the entry is Florence; Florence could be a city or a last name. We now need to understand the table to work out if we need to change it; and if we do, does it need changing as a city or as a last name in order to be meaningful?

If you then look at the credit card; we need to understand the type of credit card to make sure that the data in the non-production environment statistically represents production. This means that we need to make sure that the credit card network is analyzed. Replacing the data with a meaningless 16-digit number removes the context of the data.

ACCELARIO

DevOps

CI/CD

Automated process

Self service

Versions/ Platforms

Masking on the fly

Conclusion

The need to ensure that organizations are compliant with their legal obligations while providing valid and useful data to those that require it is one of today's key organizational challenges. With legislation such as GDPR resulting in penalties of €20 million or 4% of annual global turnover (whichever is greater) for infringements, the implications of bad data governance have never been more astute. EU businesses have already been fined €33.62 million ($40.56m) in Q1 2021 -- can your business afford not to mask data?

Accelario offers several ways to reduce the risks of data leaks with data masking being only one of them. For more information on how Accelario can help you and your business contact: sales@accelario.com