SYNTHO

Synthetic Data Guide

All Synthetic Data Generation approaches in one platform

- ✓ AI Generated Synthetic Data
- ✓ Smart De-Identification
- ✓ Test Data Management

Cedars Sinai · Centraal Bureau voor de Statistiek · Erasmus MC · KVK · Leiden University Medical Center · lifelines · umcg

Table of content

SYNTHO

About Syntho

Syntho is revolutionizing the tech industry with AI synthetic data. As a leading provider of synthetic data software, Syntho's mission is to empower businesses worldwide to generate and leverage high-quality Synthetic Data at scale.

About Syntho

Founded in 2020, **Syntho** is the Amsterdam-based scaleup that is revolutionizing the tech industry with AI-generated synthetic data for analytics and test data management.

As a leading provider of synthetic data software, Syntho's mission is to empower businesses worldwide to generate and leverage high-quality Synthetic Data at scale.

Through our innovative solution, **The Syntho Engine**, we are accelerating the data revolution by unlocking privacy-sensitive data. Dramatically reducing the time to data and providing high-quality test data.

By doing so, we aim to foster an open data economy where information can be freely shared and utilized without compromises on privacy.

Clients and Recognitions

Client references

Cedars Sinai

lifelines

Centraal Bureau voor de Statistiek

Leids Universitair Medisch Centrum

Erasmus Universiteit Rotterdam

umcg

Erasmus MC University Medical Center Rotterdam

Industries

- HealthCare
- Public Organizations
- Financial Services

Awards

Philips Innovation Award

SAS Global Hackathon in Healthcare & Lifescience

NVIDIA GTC

Selected by NVIDIA as leading Generative AI scale-up

UNESCO

UNESCO Gender Bias challenge

SYNTHO

Introduction synthetic data

Welcome at Syntho!

Synthetic data introduction

Synthetic data definition:

- ✓ Artificially generated data that mimics the statistical characteristics and patterns of real-world data. It is created using algorithms or models based on existing data, without containing any actual information from individuals or entities.
- ✓ Synthetic data is commonly used in various fields, including machine learning, data analysis, and software testing, to protect privacy, enhance data security, and overcome limitations in accessing or sharing real data.

Synthetic data is essential for addressing various challenges in data-driven fields:

Unlock data & valuable insights

Organizations today are collecting vast amounts of data. However, not all of it can be used as it is sensitive and contains personal information. Consequently, this data is "locked" and cannot simply be used. This is challenging as **data-driven tech is only as good as the data it can utilize**. This is where AI-generated synthetic data comes in.

Gain digital trust

Customers want to know that **their personal data is safe and secure** and that the organizations they do business with are transparent and honest. One way that companies can build digital trust is by using AI synthetic data.

Drive industry collaborations

Organizations are constantly looking for ways to collaborate and share data internally or maybe even externally to drive innovation and gain a competitive edge. However, **privacy concerns and data silos can make it difficult** to work with sensitive data across departments, companies, and industries.

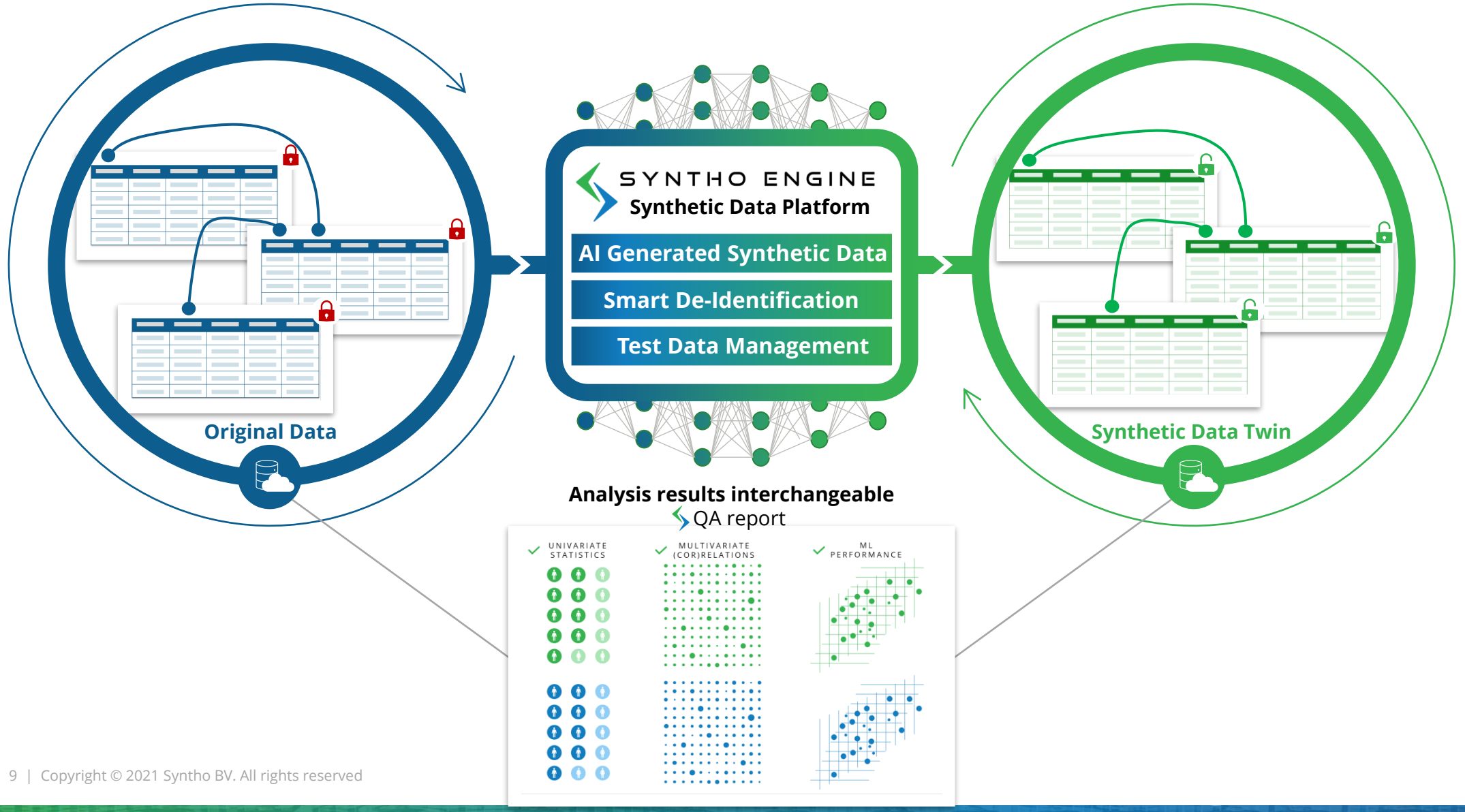
Resources: Grand View Research forecasts that the market for synthetic data generation with Generative AI will grow from $1.63 billion in 2022 to around $13.5 billion by 2030 at a CAGR of 35%. According to **Gartner**, 60% of data used for AI in 2024 will be synthetic — that's 60 times more than in 2021. Synthetic data platforms are on the rise, too. Market Statesville expects the global synthetic data platform market to grow from $218 million in 2022 to $3.7 billion by 2033.

SYNTHO

Syntho solutions

From AI-Generated Synthetic Data, Smart De-Identification to Test Data Management, Syntho supports all Synthetic Data Generation approaches in one platform

All Synthetic Data Generation approaches in one platform



All Synthetic Data Generation approaches in one platform

AI Generated Synthetic Data

Mimic statistical patterns of original data in synthetic data with the power of artificial intelligence

- ✓ *Quality assurance report:* Assess generated synthetic data on accuracy, privacy, and speed
- ✓ *External evaluation by SAS:* our synthetic data is assessed and approved by the data experts at SAS
- ✓ *Time series synthetic data:* Synthesize time-series data accurately with Syntho

Smart De-Identification

Protect sensitive information by removing or modifying personally identifiable information (PII)

- ✓ *PII Scanner:* Identify PII automatically with our AI-powered PII Scanner
- ✓ *Synthetic mock data:* Substitute sensitive PII, PHI, and other identifiers
- ✓ *Consistent mapping:* Preserve referential integrity in an entire relational data ecosystem

Test Data Management

Create, maintain, and control representative test data for non-production environments

- ✓ *De-identification and synthetization:* generate test data used for comprehensive testing and development in representative scenarios
- ✓ *Rule-based synthetic data:* Generate synthetic data to mimic real-world or targeted scenarios using predefined rules and constraints
- ✓ *Subsetting:* Reduce records to create a smaller, representative subset of a relational database while maintaining referential integrity

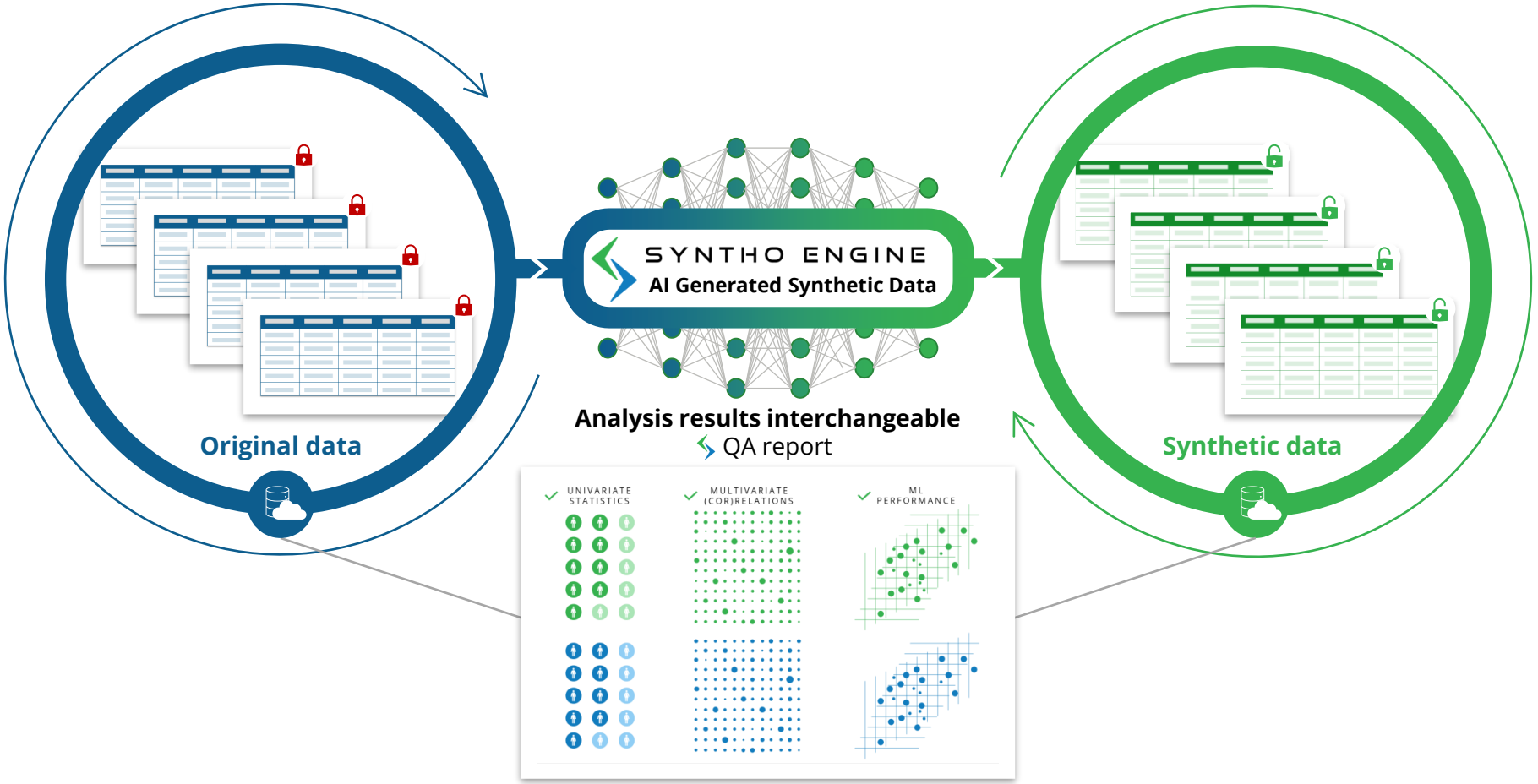
SYNTHO

AI-generated synthetic data

Mimic statistical patterns of original data in synthetic data with the power of artificial intelligence

AI-generated synthetic data

Mimic statistical patterns of original data in synthetic data with the power of artificial intelligence



- ✓ *Quality assurance report:* assess generated synthetic data on accuracy, privacy, and speed
- ✓ *External evaluation by SAS:* our synthetic data is assessed and approved by the data experts of SAS
- ✓ *Time series synthetic data*: Synthesize time-series data accurately with Syntho

Quality evaluation of the synthetic data

1. Syntho's Quality Assurance (QA) Report

We provide a comprehensive quality assurance report for every synthetic data run, that demonstrates the accuracy of the synthetic data compared to the original data.

2. External assessment by the data experts of SAS

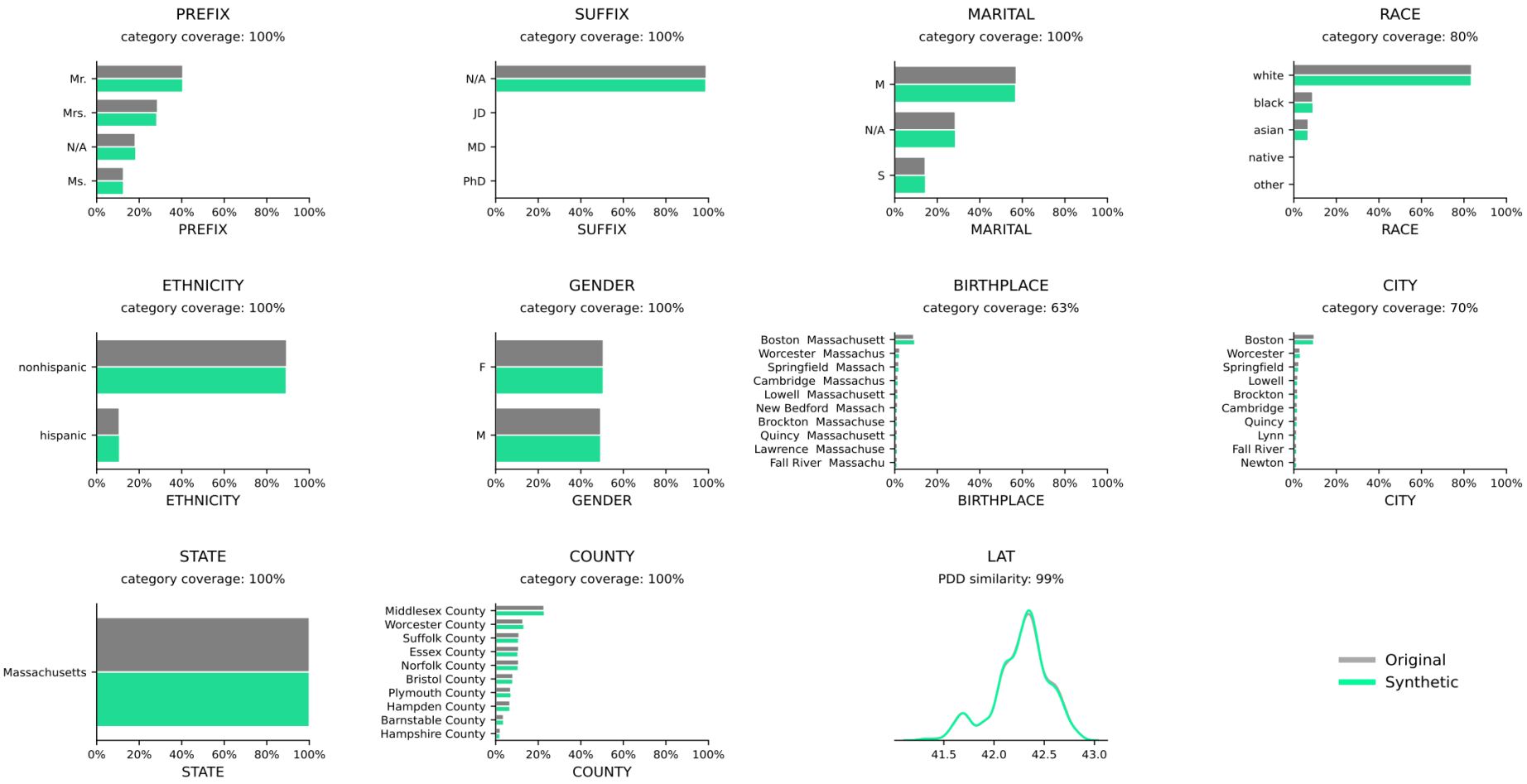
The data experts from SAS approved our AI-generated synthetic data for the use of model development as part of a real case study for a telecom customer.

Our synthetic data is approved by the data experts of SAS

Quality assurance report overview

Distributions

The frequency of variables within given categories or values and are accurately captured by the Syntho Engine.

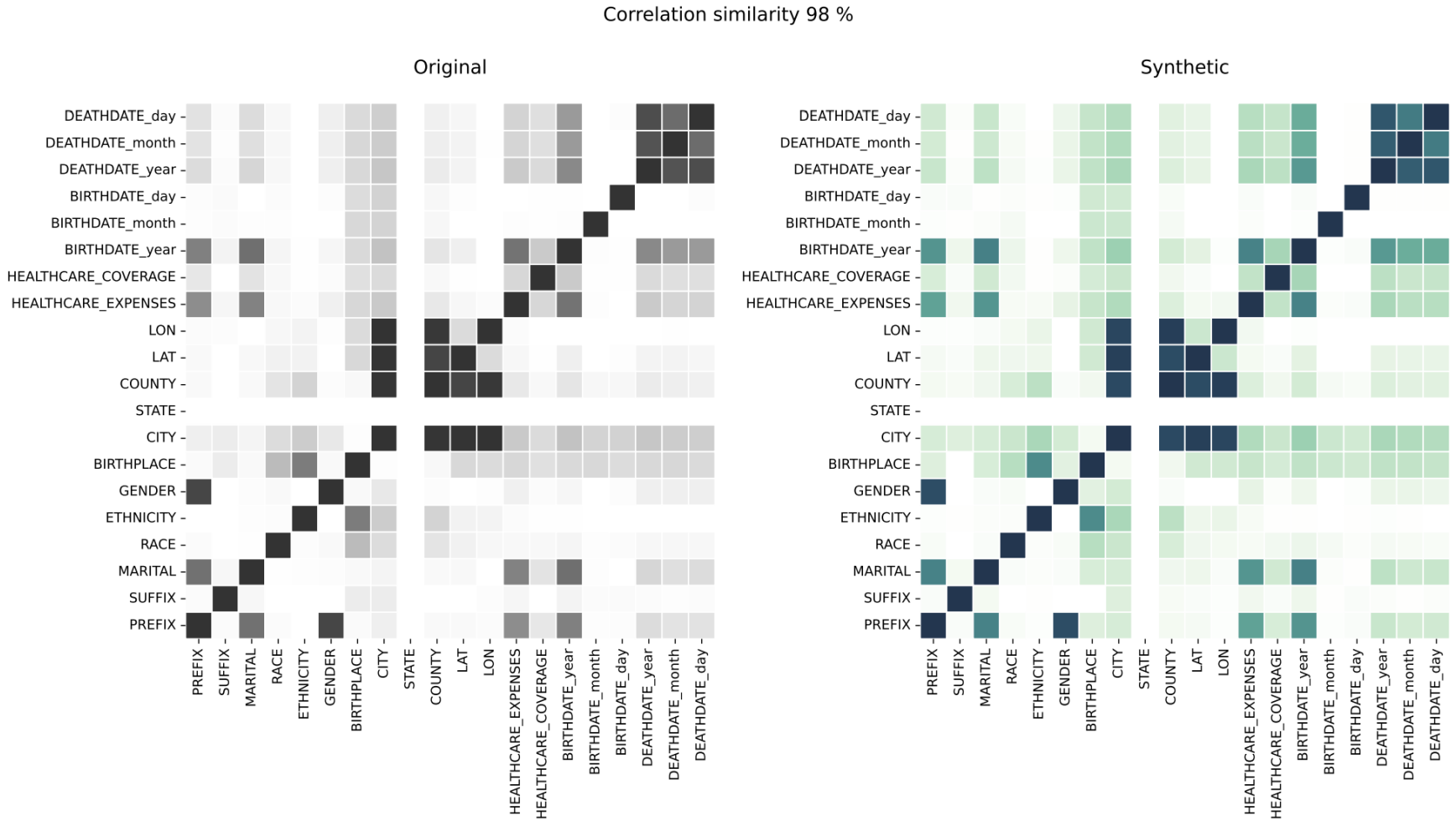


[Download the report]

Quality assurance report overview

Correlations

Shows the relationship between variables, illustrating the degree to which variables are related. The Syntho Engine accurately captures these relationships.

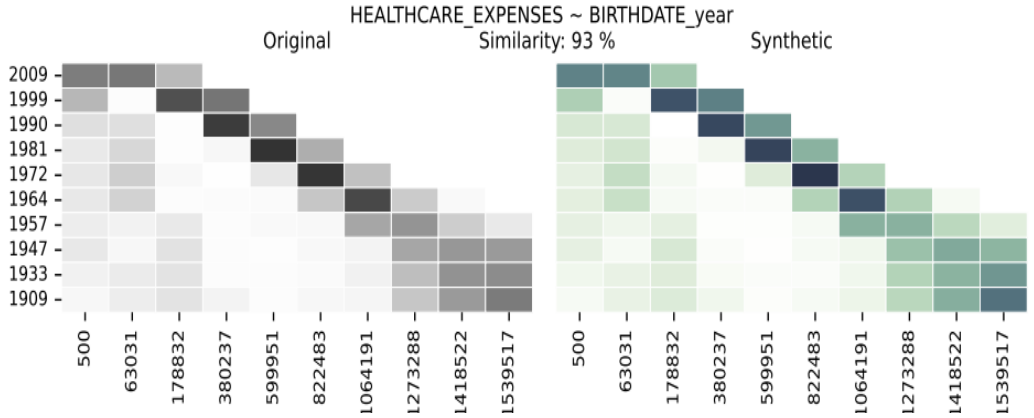


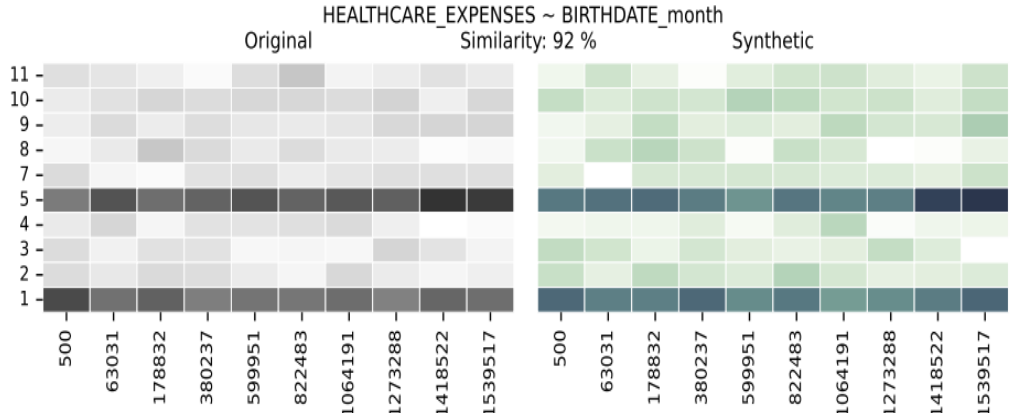
Download the report

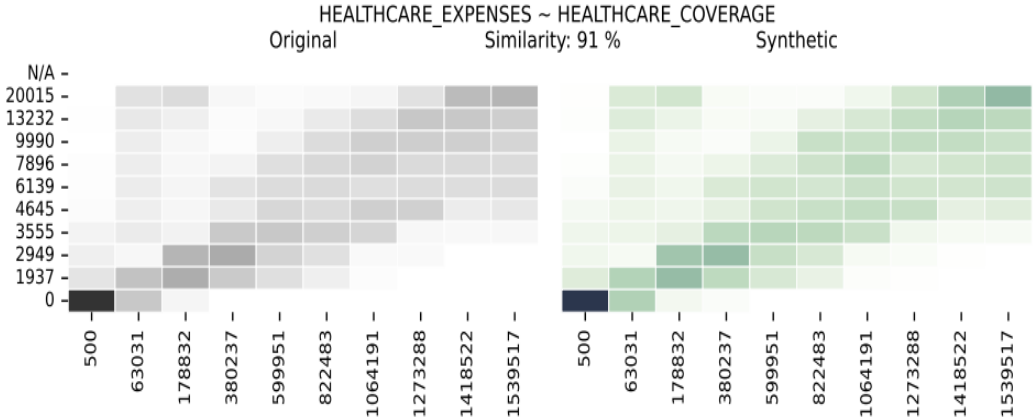
Quality assurance report overview

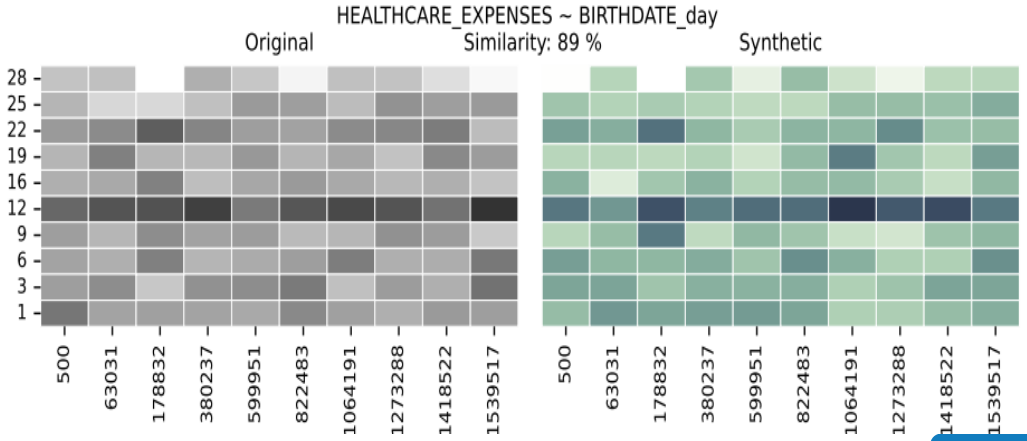
Multivariate distributions

Multivariate distributions and multivariate correlations take us beyond singular dimensions, providing a comprehensive view of how multiple variables are related. The Syntho Engine captures these relations.







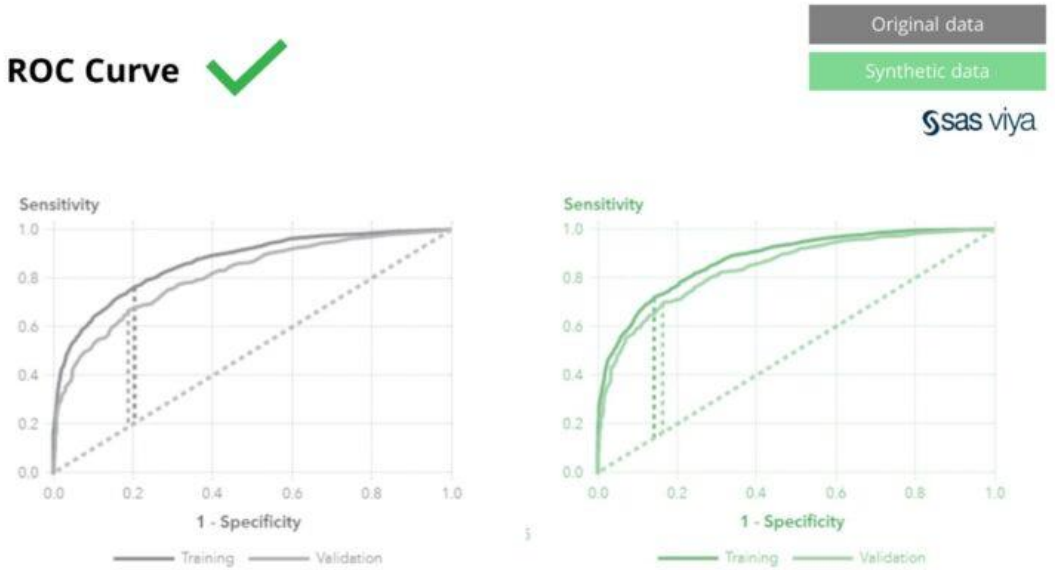


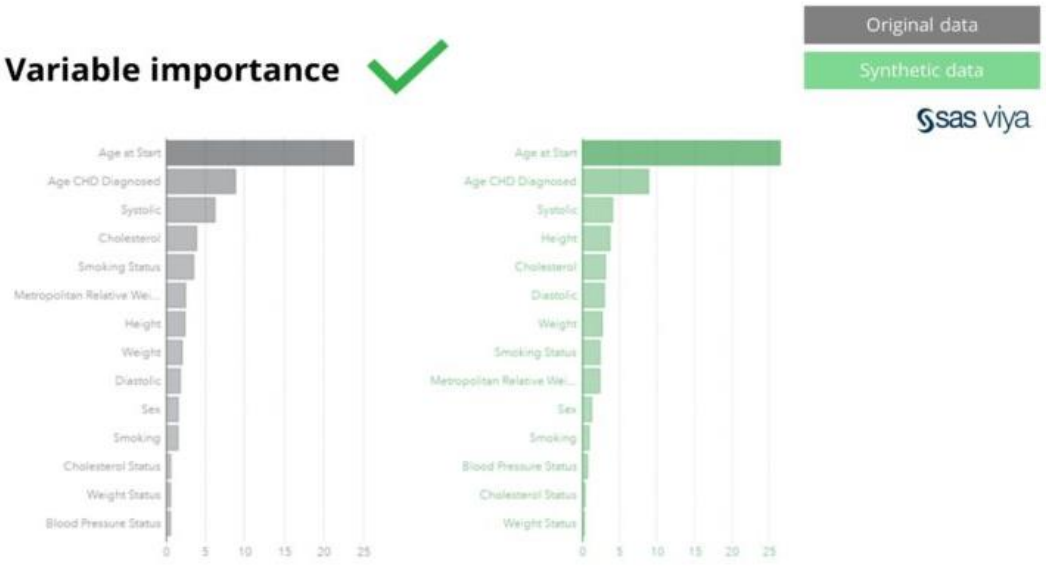
Download the report

Cancer Research using synthetic data with SAS Viya

Read more

The goal of this project was to predict deterioration and mortality as part of cancer research for a leading hospital. That is why Syntho and SAS collaborate for this hospital, where Syntho unlocks data with synthetic data and SAS realizes data insights with SAS Viya, the leading analytics platform.





Conclusions:

1. Synthetic data preserved: correlations, ROC Curve, variable importance.
2. Synthetic data generated by the Syntho Engine in SAS Viya is indeed as-good-as-real and we can use synthetic data for model development

Our synthetic data is assessed and approved by the data experts at SAS

Read more

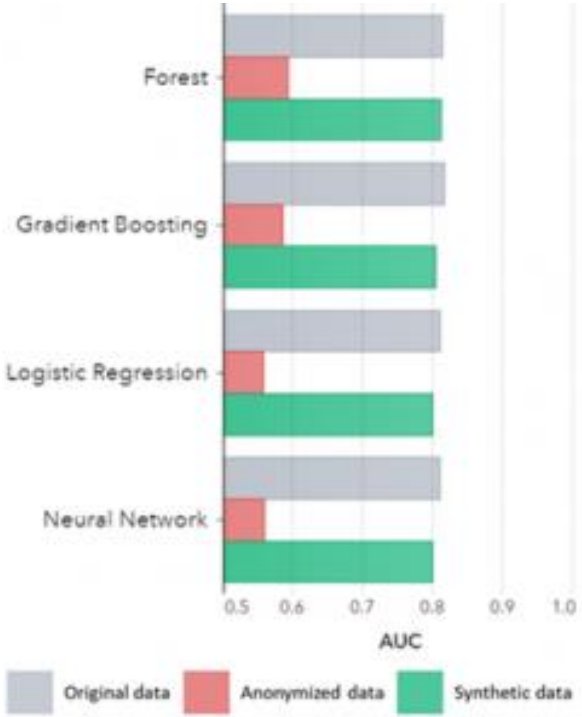
The project aimed to use synthetic data to train some models to predict customer churn and to evaluate the performance of those trained models on real data. The target dataset was a telecom dataset provided by SAS containing the data of 56.600 customers. The dataset contains 128 columns, including one column indicating whether a customer has left the company (i.e. 'churned') or not.

Four popular classification models to make the predictions, including:

1. Random forest
2. Gradient boosting
3. Logistic regression
4. Neural network

Conclusions:

1. Synthetic data compared to the models trained on original data show highly similar performance
2. Anonymized data with 'classic anonymization techniques' show inferior performance
3. Synthetic data generation is easy and fast because the technique works exactly the same per dataset and per data type.



Time series synthetic data

Time series data is a datatype characterized by a sequence of events, observations, or measurements collected and ordered with date-time intervals, typically representing changes in a variable over time, and is supported by Syntho.

We support complex time series data

- Un-equally spaced intervals
- Equal length
- Un-equal length
- Univariate time series
- Multivariate time series

Examples of time series data:

- **Financial transactions:** payments with credit and/or debit cards for transaction monitoring
- **Health metrics:** heart rate, blood values, cholesterol level
- **Energy consumption:** smart meter data, electricity usage
- **Sensor readings:** time-stamped measurements from sensors, such as temperature, flow, etc.

Patients (Entity table)

ID	Name	Gender	Country	Birthdate	Blood type
3456	Julie	F	US	1955-12-09	O+
3457	Christine	F	NL	1998-03-22	B+
3458	Angela	F	UK	1982-02-16	A-
3459	Michael	M	US	1943-06-18	O-
3460	Cynthia	F	UK	1986-04-04	AB+

Patient Medications (Linked table)

ID	Medication	Reason	Datetime	Costs	Coverage	Patient ID
1	Albuterol	Diabetes	2021-10-28 03:10:00	569	139	3456
2	Albuterol	Diabetes	2022-11-21 17:21:00	526	112	3456
3	Metformin hydrochloride	Diabetes	2021-05-23 19:56:00	743	213	3456
4	Leronlimab	COVID-19	2018-01-23 09:11:00	850	850	3458
5	Insulin	Diabetes	2020-02-14 07:54:00	374	374	3458
6	Insulin	Diabetes	2020-07-05 12:02:00	374	374	3458
7	Insulin	Diabetes	2021-02-05 15:53:00	374	374	3458
8	Memantine hydrochloride	Alzheimer's disease	2018-12-18 13:25:00	4890	2759	3459
9	Donepezil hydrochloride	Alzheimer's disease	2019-01-31 14:26:00	4381	2459	3459

Multivariate time series information

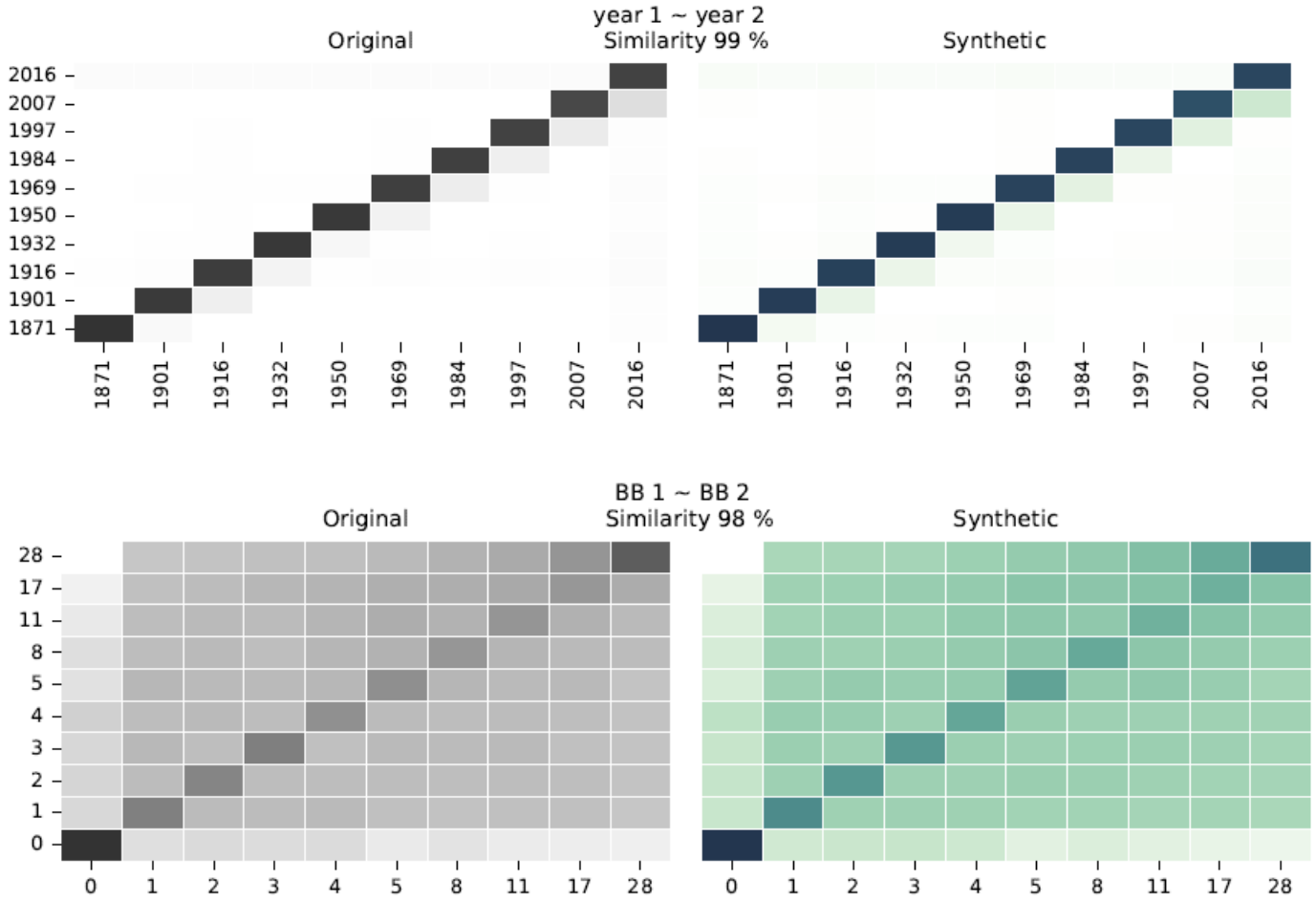
Syntho time series synthetic data. Autocorrelation plots

Support complex time-series data synthetization

Our Syntho Engine is optimized to synthesize the most complex time series data accurately. We have optimized our models in collaboration with leading organizations working with the most complex time series data.

Challenges related to time series data

- Time series data is more challenging to synthesize because it needs to capture the temporal dependencies and patterns inherent in real-world sequential observations.
- Unlike independent and identically distributed data, where each observation is unrelated to the others, time series data exhibits dependencies across time steps.



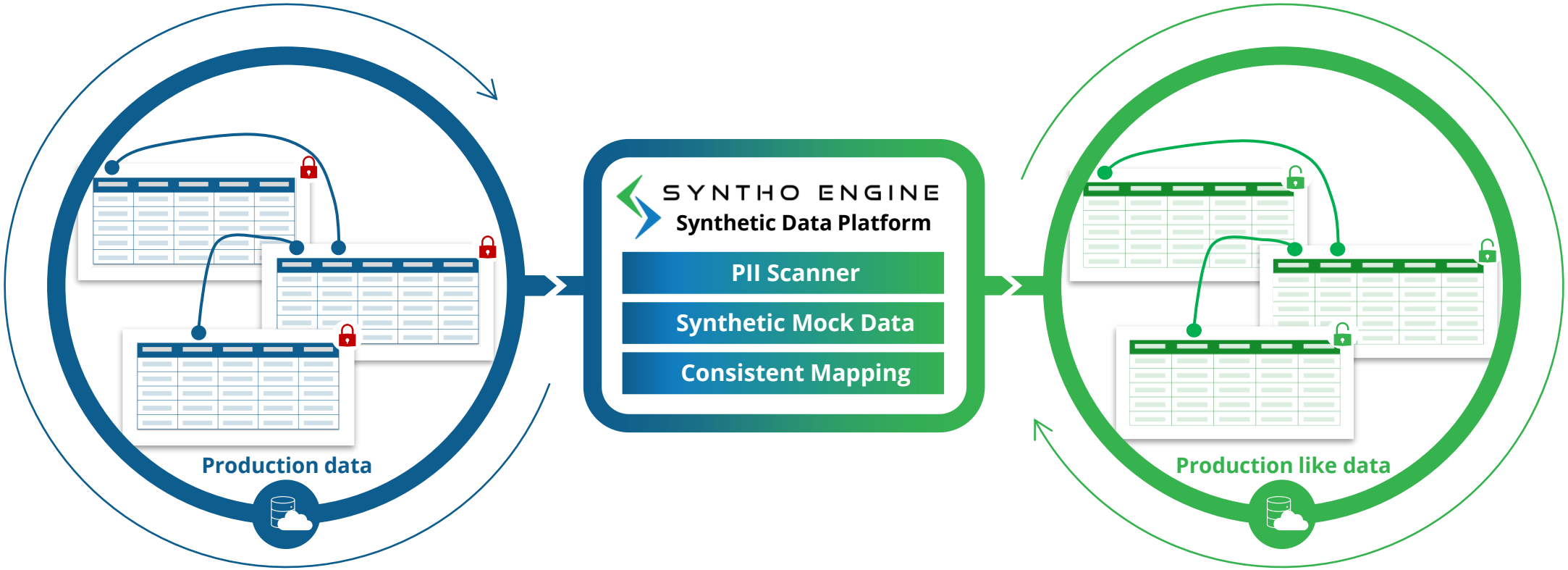
SYNTHO

Smart De-Identification

Protect sensitive information by removing or modifying personally identifiable information (PII)

Smart De-Identification

Protect sensitive information by removing or modifying personally identifiable information (PII)



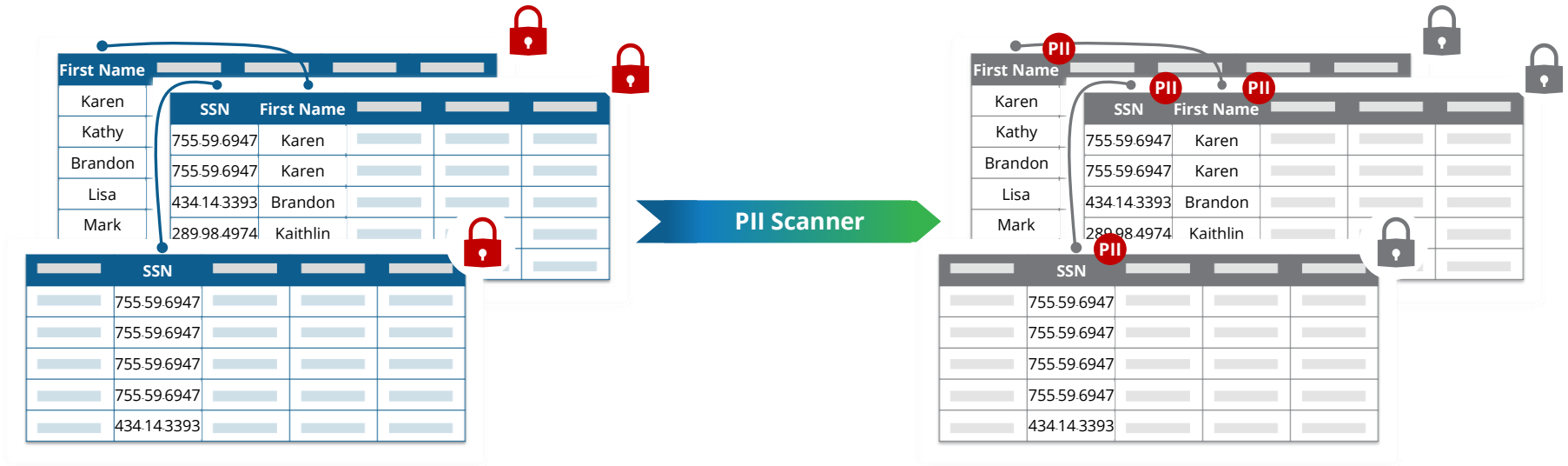
- ✓ *PII Scanner:* Identify PII automatically with our AI-powered PII Scanner
- ✓ *Synthetic Mock Data:* Substitute sensitive PII, PHI, and other identifiers
- ✓ *Consistent mapping:* Preserve referential integrity in an entire relational data ecosystem

PII scanner

Mitigate manual work and utilize our PII column scanner to identify columns in your database containing direct identifiers (PII/PHI) with the power of AI

Key benefits of our AI-powered scanner:

1. **Enhanced data privacy:** The AI-powered PII Scanner ensures enhanced data privacy by automatically identifying and flagging personally identifiable information (PII) within datasets
2. **Compliance with regulations:** By swiftly and accurately identifying PII elements, the AI-powered scanner helps organizations adhere to regulatory requirements such as GDPR, HIPAA, and CCPA
3. **Time and cost savings:** Automating the process of PII detection significantly reduces the time and resources required for manual data inspection, enabling organizations to achieve compliance more efficiently and cost-effectively.

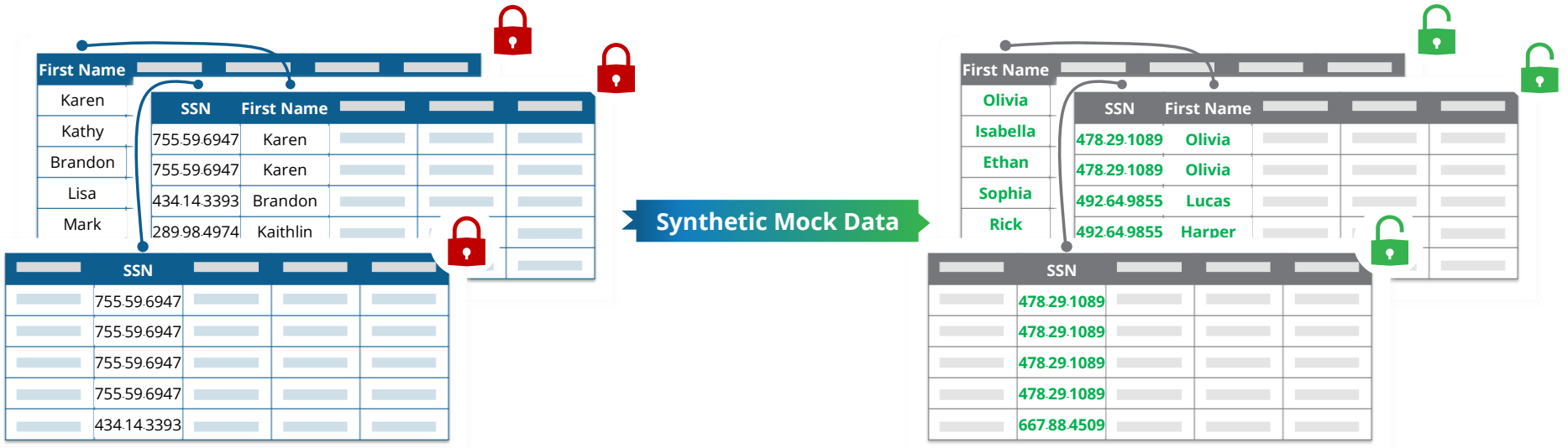


Synthetic Mock Data

Substitute sensitive PII, PHI, and other identifiers with representative Synthetic Mock Data that follow business logic and patterns.

Syntho supports +150 different mockers
The mockers are available in different languages and alphabets. Syntho supports default mockers like first name, last name, and phone numbers, but also more advanced mockers to generate mock data that could follow your defined business rules.

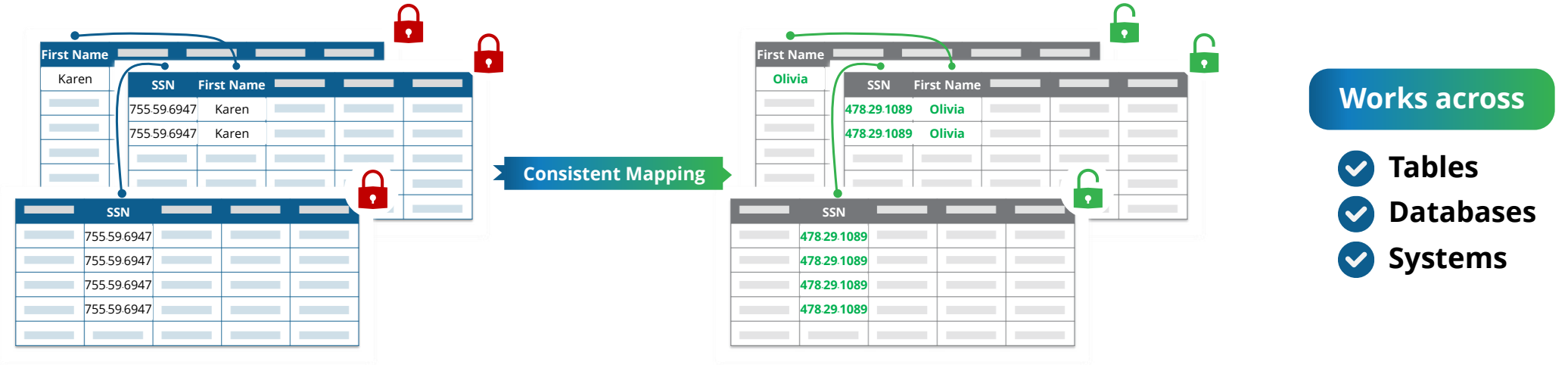
Advanced mockers
Our platform provides a wide variety of advanced mockers capable of producing synthetic data either from scratch or following predefined rules.



Consistent mapping

What is consistent mapping?
Preserve referential integrity with consistent mapping in an entire data ecosystem to match data across tables, databases, and systems.

What is referential integrity?
Referential integrity is a concept in database management that ensures consistency and accuracy between tables in a relational database. Enforcing referential integrity is crucial for maintaining the reliability of test data in a relational database as part of non-production environments. It would ensure that every value that corresponds to “Person 1” of “Table 1” corresponds to the correct value of “Person 1“ in “Table 2” and any other linked table.

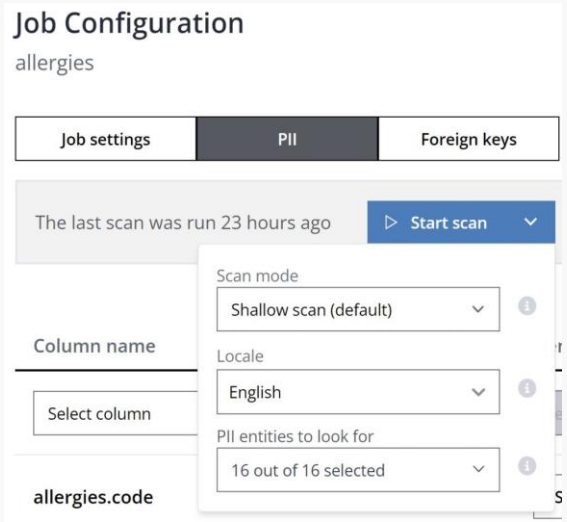


Smart De-Identification process

1. Automatically **scan** for personally identifiable information (PII) in your databases
2. Protect sensitive information by **removing or modifying** personally identifiable information (PII)
3. **Apply** the selected mocker to a column through either the PII or Job Configuration tab

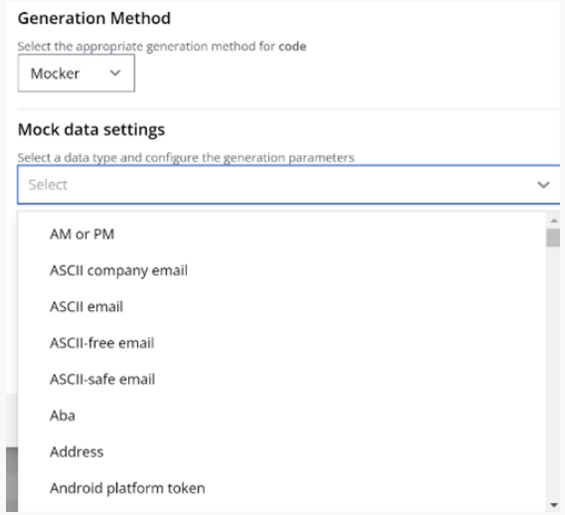
1.Identify PII

Scan PII automatically with our AI powered PII Scanner or identify columns that you would like to mock.



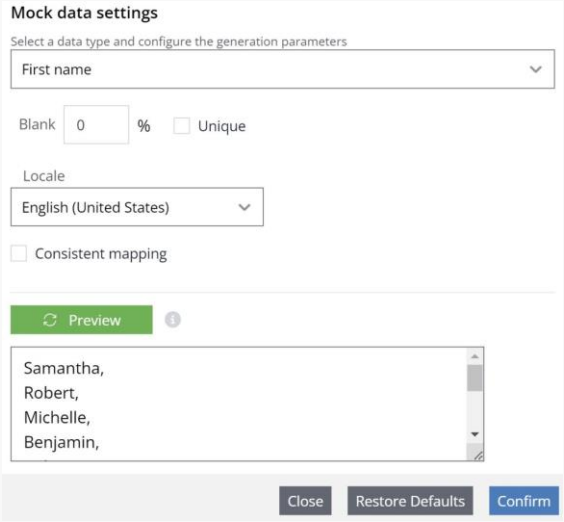
2. Select mockers

Select the suggested mocker by our PII scanner or configure mockers on column level.



3. Apply mocker

Confirm to apply the selected mocker on a column via the PII or Job Configuration tab.



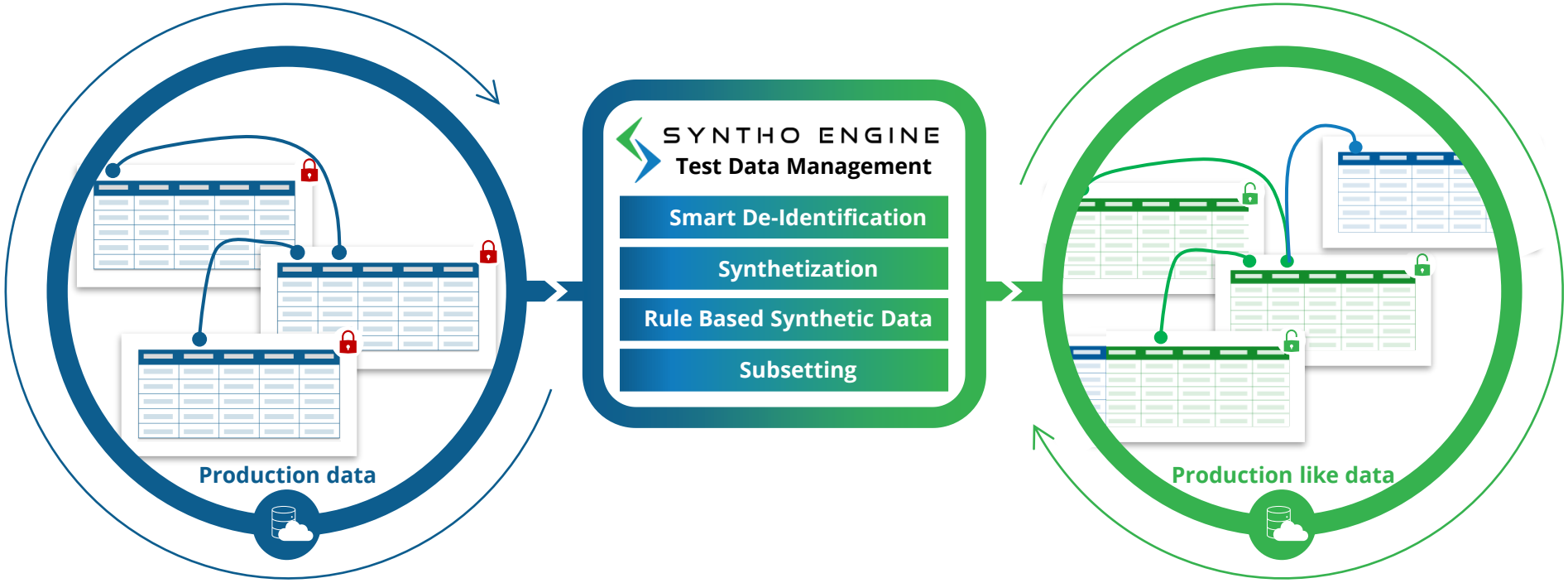
SYNTHO

Test Data Management

Create, maintain, and control representative test data for non-production environments

Test Data Management

Create, maintain, and control representative test data for non-production environments.



- ✓ *De-identification and synthetization:* generate test data that for comprehensive testing and development in representative scenarios.
- ✓ *Rule-based Synthetic Data:* Generate synthetic data to mimic real-world or targeted scenarios using predefined rules and constraints
- ✓ *Subsetting:* Reduce records to create a smaller, representative subset of a relational database while maintaining referential integrity

Test Data Management process facilitated by Syntho

Streamline connectivity to the production environment, ensuring seamless integration. Enable realistic testing scenarios by accurately mimicking the data types and structure of production data

Synthesis or de-identification techniques

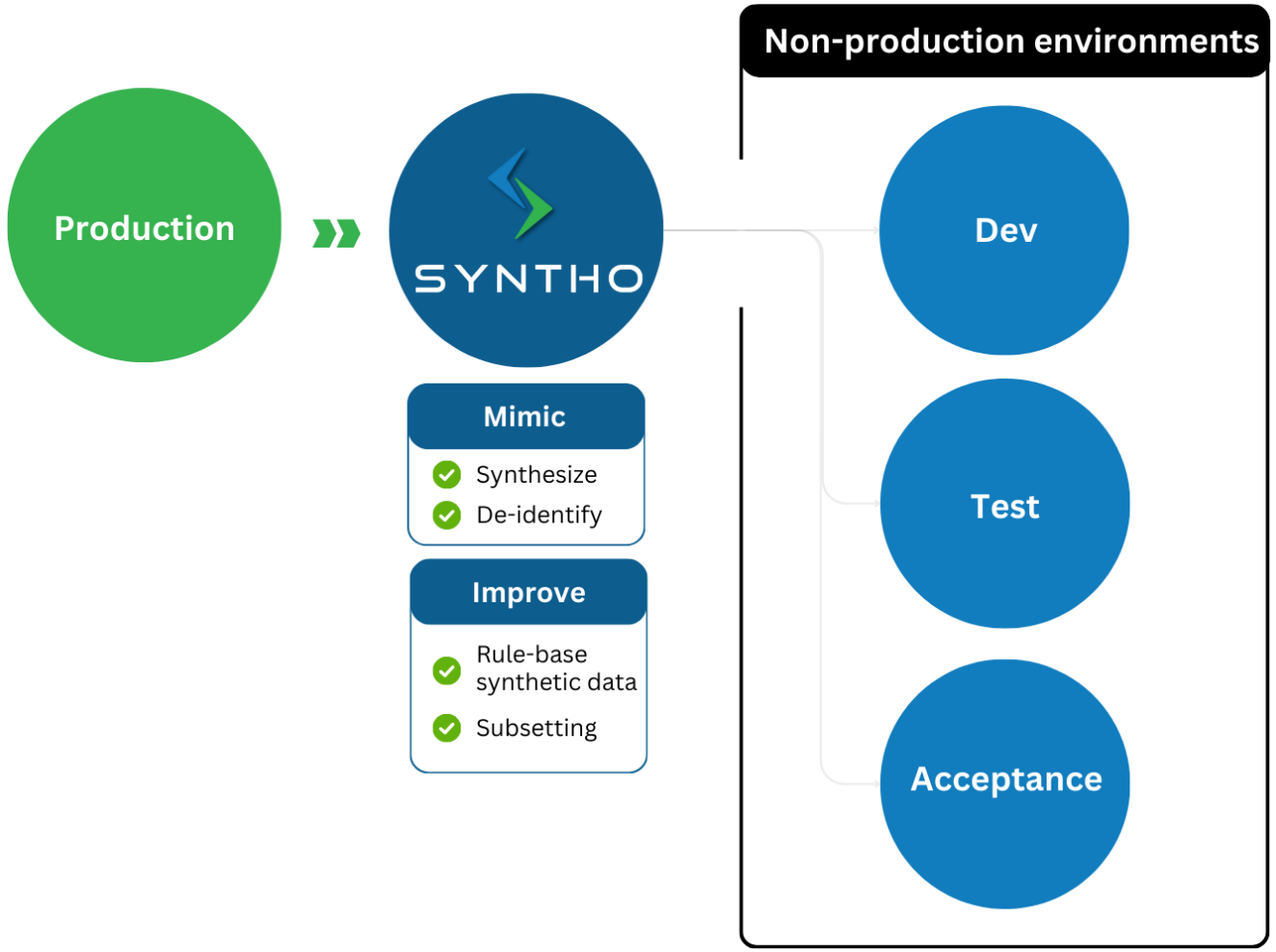
- Sensitive data is protected while maintaining utility

Rule-based synthetic data

- Enhances data quality and relevance for test, development, and acceptance environments, fostering efficient and effective testing processes

Subsetting

- Reduce records to create a smaller, representative subset of a relational database while maintaining referential integrity



De-identification and synthetization

Testing and development with representative test data is essential to deliver state-of-the-art solutions. Using original production data seems obvious, but is often challenging due to privacy regulations as it cannot simply be used as it:

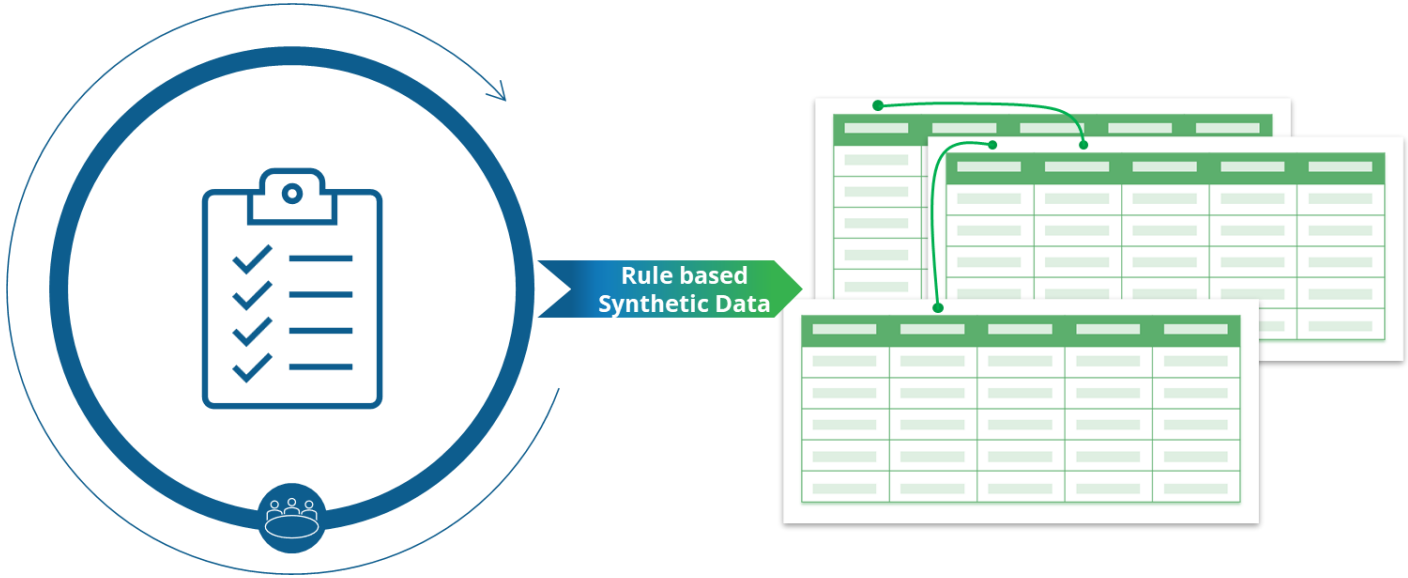
1. Contains (privacy) sensitive information
2. Is limited, scarce, or misses data
3. Or does not exist at all

This introduces challenges for many organizations in getting the test data right. Hence, Syntho supports all best practice solutions to establish your test data right to have test data that mimics production data as closely as possible.

- ✓ *PII Scanner:* Identify PII automatically with our AI-powered PII Scanner
- ✓ *Synthetic Mock Data:* Substitute sensitive PII, PHI, and other identifiers
- ✓ *Consistent mapping:* Preserve referential integrity in an entire relational data ecosystem

Rule-based synthetic data

Create synthetic data based on pre-defined rules and constraints, aiming to mimic real-world data or simulate specific scenarios.



Generate Data from scratch

In cases where data is either limited or where you do not have data at all, the need for representative data becomes crucial when developing new functionalities.

Enrich data

Rule-based synthetic data could enrich data by generating extended rows and/or columns. It can be used to produce extra rows to create larger datasets easily and efficiently.

Flexibility and customization

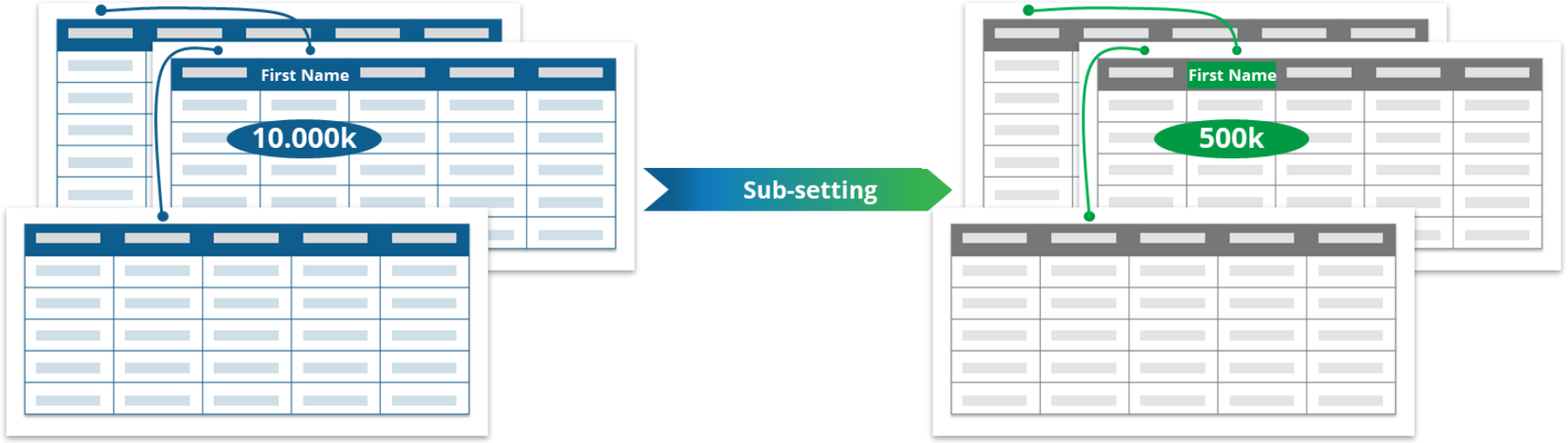
The rule-based approach provides flexibility and customization to adapt to diverse data formats and structures, enabling the full tailoring of synthetic data according to specific needs.

Data cleansing

Rule-based synthetic data facilitates data cleansing by generating data adhering to predefined rules, correcting inconsistencies, filling missing values, and removing errors,

Subsetting

Decrease the number of records to create a smaller representative subset of a relational database with preserved referential integrity



Reduce infrastructure and computational costs

Excessive data volumes can lead to high infrastructure and computation costs, which are unnecessary for test data in non-production environments.

Manageable test data by testers and developers

Managing huge data volumes in non-production environments poses challenges for testers and developers.

Faster test data setup and maintenance

Smaller data volumes facilitate faster and more straightforward setup and maintenance of non-production test environments.

SYNTHO

Use cases

This section showcases the real-world applications of Syntho's solutions through client testimonials and use cases with reference articles.

Main use cases

Synthetic Data as Test Data

Challenge

Using personal or original production data as test data is not allowed.

Our solution

Using AI-generated synthetic test data, we can provide faster delivery and quality of data.

- ✓ Production-like data
- ✓ Privacy by design
- ✓ Easy, fast and agile

Read more

Synthetic Data for Analytics

Challenge

For many organizations, data cannot simply be used and shared.

Our solution

We help you gain easy and fast access to AI-generated synthetic data that is as good as real data.

- ✓ Unlock (sensitive) data
- ✓ As-good-as-real data
- ✓ Easy, fast and scalable

Read more

Synthetic Data as Data Sharing

Challenge

Data-sharing issues (i.e., legal delays, untapped valuable data, lack of a solid framework) cause project setbacks.

Our solution

Using synthetic data instead of real data removes data-sharing obstacles, enabling you to:

- ✓ Get faster access to data
- ✓ Share data with different parties without privacy concerns
- ✓ Speed up innovation processes
- ✓ Increase customer retention and acquisition

Read more

Main use cases

Synthetic Data for Product demo

Challenge

Your demo data may be suboptimal, leading to missed opportunities during product demonstrations.

Our solution

Impress your prospects with exceptional product demos featuring AI-generated synthetic data tailored to their needs.

- ✓ Error-free, high-quality demo data
- ✓ Tailor product demos
- ✓ Easy, fast and agile

Read more

Synthetic Data for Data monetization

Challenge

Data monetization faces significant challenges, including ensuring data privacy and compliance, maintaining data quality and integrity, and implementing robust data governance practices.

Our solution

Synthetic data allows you to create a data marketplace with a data catalog and share it among internal and external stakeholders.

- ✓ Sandbox with data preview
- ✓ Data marketplace with data catalog
- ✓ Share data for life-crucial research

Read more

Synthetic Data for AI modelling

Challenge

Ensuring equal access to high-quality data and mitigating biases to influence the accuracy and fairness of predictive outcomes.

Our solution

A synthetic data platform allows you to create high-quality synthetic data, which will capture all variables from the original data and allow you to upsample underrepresented data.

- ✓ Access to a full dataset with all valuables
- ✓ Upsample data for better predictions
- ✓ Reduce data access to 1 day

Read more

SYNTHO

Case study

This use case section highlights Syntho's current clients and their success stories.

HealthCare

Synthetic healthcare data with Cedars Sinai Medical Center

The situation
The leading US hospital Cedars Sinai aims to conduct research and analysis without compromising sensitive information (Leading hospital (#1 California, 2#Nation).

The solution
Cedars-Sinai has announced the implementation of synthetic data in its research and clinical data science initiatives. The move marks a significant shift in the way healthcare organizations can securely and ethically conduct research and analysis without compromising sensitive information.

Organization:
Cedars Sinai Medical Center

Location:
The United States

Size:
12000+ employees

Target data:
Patient data, data from the electronic health record system

Read more

Synthetic patient EHR data for advanced analytics with Erasmus MC

The situation
Healthcare data is the most privacy-sensitive data and is therefore locked. This privacy-sensitive data: is time-consuming to access, requires extensive paperwork, and cannot simply be used. This is challenging as data-driven insights could improve patient care and efficiency with the potential to save lives.

The solution
The Research Suite and datahub from the Erasmus MC are responsible for data distribution. Now they offer the opportunity to request synthetic data as an alternative to real-world data.

The benefits
- Analytics with synthetic data
- Faster access to data
- Enlarge data for testing purposes

Organization:
Erasmus Medical Centre

Location:
The Netherlands

Size:
16000+ employees

Target data:
Patient data, data from the electronic health record system

Read more

HealthCare

Realistic synthetic patient data for clinical research with Lifelines

The situation
Lifelines is a large, multigenerational data and biobank that provides insight into the population-based cohort study. The company sells the data of 167 000 participants followed by 30 years of research to insurance, hospitals, and research institutions. The participant's data contains private information, so data sharing is a time-consuming process. Additionally, data buyers request the opportunity to preview the data in the catalogue.

The solution
Lifelines partners with Syntho to synthesize data, thereby enhancing its accessibility and preserving the privacy of participants. As an alternative to using real data, everyone has now the possibility to work with synthetic data. Unlike open-source and other commercial solutions, Syntho's platform has consistently delivered better results in terms of accuracy, privacy, and usability, particularly when handling geographical location and longitudinal data, effectively demonstrating that synthetic data is as good as real.

The benefits
- Faster access to data
- Preserve the privacy of participants
- Increased accessibility of data
- Preview data before buying with a data catalogue

lifelines

Organization:
Lifelines

Location:
Netherlands

Industry:
HealthCare

Size:
100+ employees

Target data:
HealthCare historical data

Read more

Public organizations

Synthetic statistical data for data exchange with CBS

The situation
The CBS (Dutch Central Bureau of Statistics) is the statistical agency in the Netherlands, responsible for collecting, analyzing, and distributing a wide range of data. This results in actionable insights into societal trends, economic developments, demographics and more. Privacy is key, as it serves as role-model in the way it utilizes data.

The solution
Synthetic data holds significant potential for the Dutch Central Bureau of Statistics (CBS). With a growing demand for data, synthetic data offers a solution for improved data sharing and collaboration with the scientific community and private sector. Moreover, synthetic data serves as a vital tool for testing IT systems, refining algorithms, and simulating real-life scenarios, enhancing data-driven decision-making while preserving individual privacy.

The results
- Improve and facilitate data sharing
- Promotes CBS's role as a data partner and hub
- Facilitate the utilization of data smarter

Centraal Bureau voor de Statistiek

Organization:
Centraal Bureau voor de Statistiek (CBS)

Location:
The Netherlands

Industry:
Public sector

Size:
2000+ employees

Target data:
Data related to the Dutch population

Read more

Synthetic data for The Netherlands Chamber of Commerce (KVK)

The situation
In pursuit of leveraging data for organizational advancement, a 2-day hackathon was organized to foster new initiatives within the company. Utilizing internal data sources as the foundation, the challenge lay in balancing the accessibility of business information with the imperative of safeguarding sensitive data and adhering to privacy regulations.

The solution
The solution entailed the utilization of synthetic data for the internal hackathon. This synthetic dataset, mirroring real business register data, ensured privacy protection while enabling participants to develop and test innovative solutions without compromising sensitive information. Furthermore, synthetic data served as test data in various development stages.

The results
- Privacy-by-design hackathon with representative and actionable data
- Innovative hackathon initiatives on relevant data
- Fast access to data

KVK

Organization:
The Netherlands Chamber of Commerce (KVK)

Location:
The Netherlands

Industry:
Governmental

Size:
1500+ employees

Target data:
Business register data

Read more

Research institutions

Synthetic data for scientific research with Erasmus University Rotterdam

The situation
The Erasmus University of Rotterdam is a leading Dutch university known for its academic excellence and research across various disciplines. The university places a crucial emphasis on data, integrating data analysis and research methodologies into its programs and the execution of academic research, including the publication of papers. However, the evolving landscape of data utilization raises important privacy implications, prompting the university to navigate the balance between utilizing the full data potential and safeguarding individual privacy rights.

The solution
As part of the research integrity solutions of the EUR, the Syntho Engine is positioned as a data management tool & service for synthetic data generation. All EUR researchers have access and are encouraged to use the platform as much as possible.

The benefits
- Improved data & privacy to enhance research integrity
- Access to more and extended datasets
- Enhanced reproducibility of research by making it easier to access synthetic data
- More data to test and validate the hypothesis

ERASMUS UNIVERSITEIT ROTTERDAM

Organization:
Erasmus University Rotterdam (EUR)

Location:
The Netherlands

Industry:
Education and research

Size:
12000+ employees

Target data:
Academic research data

Read more

Finance

Synthetic test data for online banking products for a leading Dutch bank

The situation
One of the largest retail banks in the Netherlands constantly developing and testing new online products and services. It is crucial for staying competitive and providing a seamless customer experience. However, using real customer data for testing purposes can be risky, time-consuming, and costly, especially given the growing concerns around data privacy and security.

The solution
To overcome these challenges, our customer has adopted using AI-generated synthetic data for testing purposes and provides access to the Syntho Engine to each “agile Scrum Team” in the tribe.

The benefits
- Release faster and shorten the time-to-market
- Reduce bugs in pre-production tests and reduce bugs in production
- Cost saving of test data maintenance
- Enhanced data privacy and security
- Flexibility and scalability

Under NDA

Organization:
Top 3 banks in the Netherlands

Location:
The Netherlands

Industry:
Banking

Size:
40000+ employees

Target data:
Financial transactions

On Request

Synthetic data for advanced analytics and testing with a leading international bank

The situation
The banking sector faces challenges in data management due to fragmented storage and compliance regulations. Anonymization of data leads to loss of contextual information, hindering the machine learning model, while strict data privacy measures hinder seamless collaboration and innovation.

The solution
Syntho's AI-synthetic data generation platform addresses these challenges by offering privacy-compliant realistic datasets. This facilitates accurate machine learning model training, enhances fraud detection, and enables secure data collaboration among institutions, ultimately accelerating development cycles.

The benefits
- Upsampling minority groups
- Privacy-by-design
- KYC: combating fraud, anti-money laundering and anti-terrorist financing
- Keeping data value and quality

Under NDA

Organization:
Leading international Dutch bank

Location:
The Netherlands

Industry:
Banking

Size:
60000+ employees

Target data:
Financial transactions

On Request

SYNTHO

Product overview

This product section provides an overview of Syntho's solutions and services. This section includes the features and benefits of the Syntho Engine platform, as well as some initial technical insights related to deployment and integration.

SYNTHO ENGINE 2.0

REVOLUTIONIZING AI GENERATED SYNTHETIC DATA

SYNTHO

Why Syntho?

Maximized accuracy
Synthetic data is generated with the highest accuracy, assessed and approved by the data experts of SAS

Fair pricing model
Generate unlimitedly for a fixed price. Our monthly license is tailored to the features you need, not the volume of data you produce

Full data coverage
We seamlessly handle all data types and is optimized for supporting the most complex structures, such as time series data

All-in-one solution
From AI-Generated Synthetic Data, De-Identification, and Test Data Management. We have all solutions in one easy-to-use platform

Awards

PHILIPS Innovation Award

SAS WINNER HACKATHON AI Generated Synthetic Data in Healthcare Syntho, Netherlands HEALTH CARE & LIFE SCIENCES TRACK 2023

NVIDIA GTC

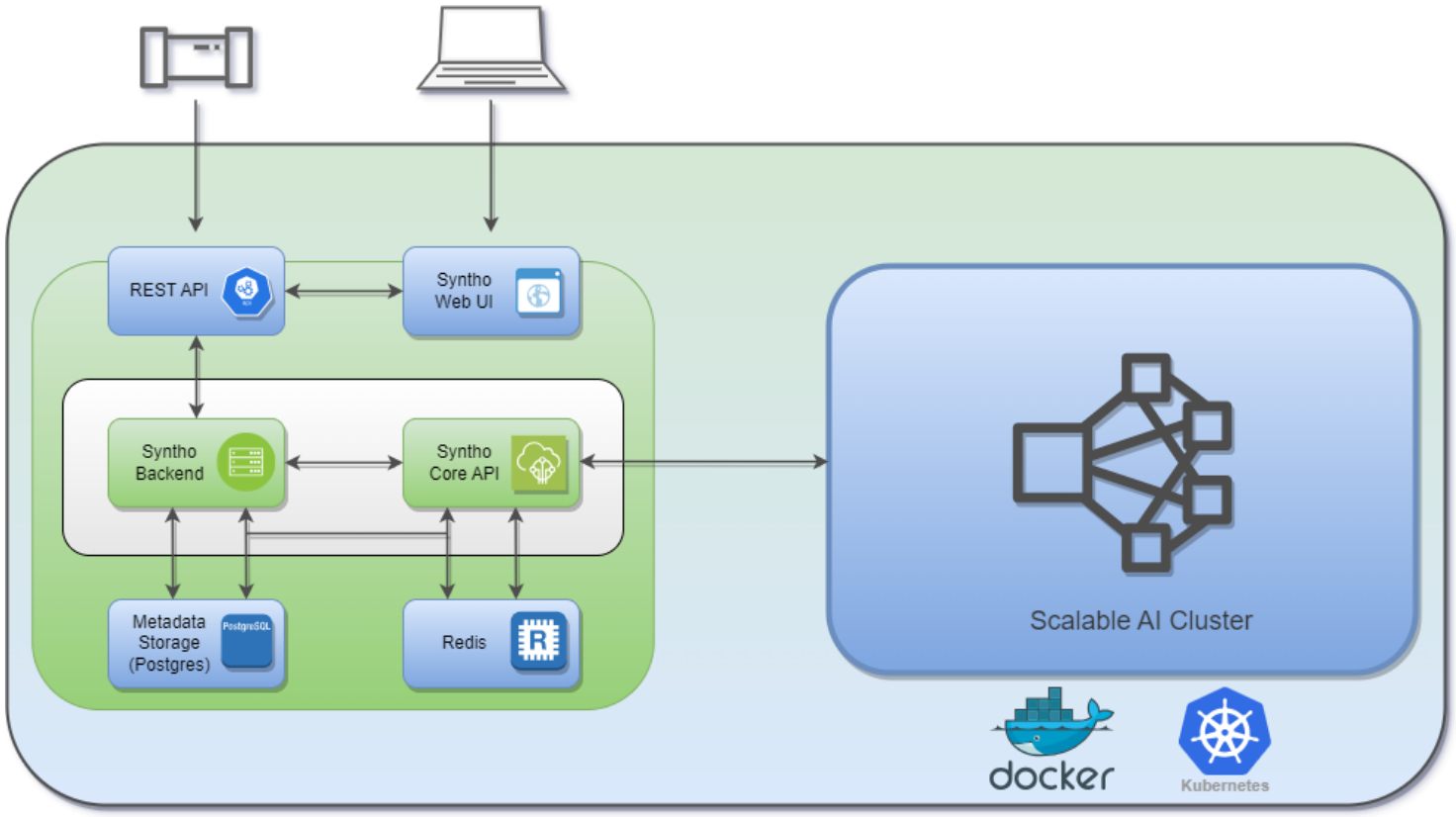
UNESCO

High-level deployment architecture via Docker and/or Kubernetes

Our platform is flexible, and deployable in any environment via docker-compose or Kubernetes. Within our Syntho Engine, we offer seamless integration options: a user-friendly interface (“Syntho Web UI”) or integration into your pipeline through our “Rest API”. This flexibility ensures easy deployment in your preferred environment, without external connections.

Possible deployment options:

- On-premise
- Any (private) cloud (Your AWS, Azure, Google Cloud, etc.)
- Syntho cloud
- Any other environment



Connectors for all leading databases

Easily connect to all the most popular databases, or easily integrate with your existing test automation and CI/CD tools.

Connect in 3 steps:

1. Connect to source data
2. Connect to destination data
3. Synthesize data

Read & write
Read only
Write only

Features

AI generated synthetic data

Time series synthetic data

Synthesize time-series data accurately with Syntho

Quality assurance report

Assess generated synthetic data on accuracy, privacy and speed

Connectors

Connect easily with the source and target data for an end-to-end integrated approach

Smart de-identification

PII scanner

Identify PII automatically with our AI-powered PII scanner

Synthetic mock data

Substitute sensitive PII, PHI and other identifiers

Consistent mapping

Preserve referential integrity in an entire relational data ecosystem

PII scanner in open text

Identify PII automatically in open text with our AI-powered PII Scanner

Test data management

De-identification and synthetization

Generate test data the reflects production data

Rule-based synthetic data

Generate synthetic data based on rules and constraints

Subsetting

Create a smaller, representative subset of a relational database

Check out all features

SYNTHO

Syntho's user documentation

[Access User Documentation]

SYNTHO

How to start

From exploring our solutions to mastering data generation, our team will guide you through the process of becoming a data generation expert.

Steps towards becoming a data generation expert

Customer Support & Success

Deepdive & decision making

Agenda

- Meet Syntho
- Alignment of client needs & requirements
- Demo
- Deepdive into Syntho's solutions
- Decision making

Outcome

- Go / No go

Syntho
Provide information

Client
Go / No go

Kick-Off

Agenda

- Alignment on requirements
- Define suggested Infrastructure
- Define way of working
- Planning

Outcome

- You are ready to start

Syntho
Share input & resources

Client
Confirm requirements

Deployment

Agenda

- Verify client infrastructure
- Deploy Syntho Engine
- Conduct tests
- Go live

Outcome

- Syntho Engine ready to use

Syntho
Software, user docs, support

Client
Prepare infra & deploy

Syntho Bootcamp

Agenda

- Syntho Engine user training
- Synthesize sample datasets together
- Synthesize client (sample) datasets

Outcome

- You are ready to use the Syntho Engine

Syntho
Host training sessions

Client
Prepare datasets

Data Generation

Agenda

- Client utilizes the Syntho Engine
- Ongoing customer support & success

Outcome

- Utilize synthetic data at scale!

Syntho
Support & check-in

Client
Data generation

SYNTHO

Pricing

Unlock transparent pricing tailored to your needs with Syntho. Explore our flexible plans designed to meet your data generation requirements efficiently and affordably

Pricing Plans

- Featured-based
- No consumption-based charges
- Free deployment

License

- Syntho Engine license
- Number of users
- Connectors

Features

- PII column scanner
- PII text scanner
- Mockers
- Consistent mapping
- Time-series
- Up sampling

Support

- Documentation
- Ticket system
- Dedicated communication channel

Contact us

Wim Kees Janssen

CEO

Contact: kees@syntho.ai

Schedule meeting: syntho.ai/kees

Uliana Krainska

Synthetic Data Consultant

Contact: uliana@syntho.ai

Schedule meeting: syntho.ai/uliana