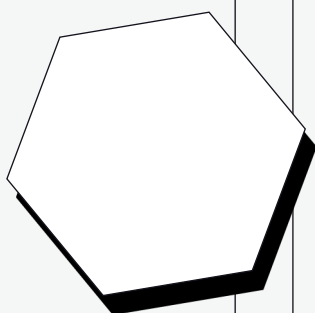


European Commission Digital Finance Platform

— Synthetic Data for Innovation



SYNTHESIZED

Overview

The European Commission's Digital Finance Strategy aims to foster a competitive and innovative European financial sector by leveraging emerging technologies while maintaining financial stability and consumer protection. As part of this initiative, the EU Digital Finance Platform was established to provide practical tools supporting the scaling of innovative fintech firms across the EU.

A key component of this platform was the Data Hub, designed to facilitate data exchange between national supervisors and financial firms, enabling participating companies to test innovative solutions and train AI/ML models. However, sharing confidential supervisory data presented a significant challenge: how to provide meaningful datasets for innovation while ensuring absolute confidentiality and compliance with data protection requirements.

Technical Challenge

National supervisors hold valuable supervisory datasets that could accelerate fintech innovation, but these datasets contain highly sensitive information that cannot be shared with third parties. Traditional anonymization techniques — such as removing identifying variables, adding random noise, or using deterministic transformations — proved inadequate because they can often be reversed if attackers know the anonymization procedure or possess the encryption key.



The Requirements

The European Commission established specific requirements for the synthetic data solution:

ABSOLUTE CONFIDENTIALITY

Enable national supervisors to participate without any real data leaving their premises.

STATISTICAL FIDELITY

Synthetic datasets must preserve the characteristics and structure of original data, maintaining analytical relevance for testing purposes.

ANONYMIZATION STRENGTH

Deliver a strong level of anonymization and confidentiality protection.

OPERATIONAL SIMPLICITY

The platform needed a straightforward implementation that national supervisors could deploy without extensive technical overhead.

STANDARDIZATION

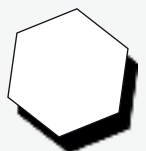
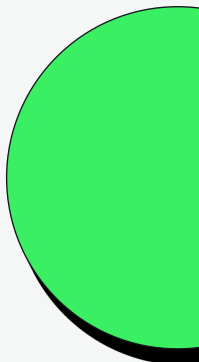
Produce synthetic datasets with a consistent format across participating authorities.

SUPPORT AUTOMATION AND REPEATABILITY

Through a programmatic workflow suitable for operational deployment.

ANALYTICAL VALIDITY

Synthetic data needed to produce reliable results when used in quantitative models and simulations.



The Solution

The European Commission acquired Synthesized's SDK through a public tender process, with the Joint Research Centre (JRC) and DG FISMA conducting comprehensive testing and validation:

Technical Implementation

Synthesized's SDK leverages TensorFlow, using deep learning algorithms to fundamentally transform how data confidentiality is achieved. Rather than anonymizing original observations, the synthesis process decomposes entire datasets into parameters — basic building blocks that characterize the data without revealing individual observations. These parameters train machine learning models involving neural networks, which then generate completely new datasets with similar statistical properties.

The platform proved operationally efficient, quick, and easy to use, requiring only two simple commands. None of the tested datasets took longer than a few minutes to synthesize using CPU processing.

Validation Results

JRC testing across three datasets — bank balance sheets (250 EU banks), retail transactions (105,316 observations), and interbank lending networks (800 Austrian banks) — demonstrated strong statistical fidelity:

JRC testing found that synthetic data generated with Synthesized successfully replicated many key patterns of the original datasets, particularly in univariate distributions. The solution demonstrated strong practical value in modelling: when synthetic balance sheet data (with rules applied) was used in the SYMBOL micro-simulation portfolio model, the estimated EU banking losses closely aligned with results from the original dataset.

Security Performance

Attack simulation testing confirmed exceptional privacy protection:

- For potential attackers attempting to infer hidden variables (such as Tier 1 capital ratios) using all other original variables, less than 10% of the original data could be correctly identified within a 5% accuracy threshold.
- When attackers used simple linear regression with known asset values to estimate gross loans, only 1-2% of observations fell within 1% of the original values.

Put simply, the synthesis approach used by Synthesized platform made re-identification impossible because original observations simply no longer existed in the synthetic version; there was no key, algorithm, or inverse function to retrieve them. Synthesized advanced PII masking capabilities not only guarantee anonymization but also a sophisticated level of data security.

“As opposed to traditional anonymization techniques, the synthesis process makes any potential attack virtually impossible. This is because instead of keeping an anonymized version of the original data, the synthesis strips it to a limited set of parameters which is then used to build up a completely new dataset with similar statistical properties. This makes any identification of individual original observations impossible, because they are simply no longer there.”

—
European Commission Joint
Research Centre (2024)

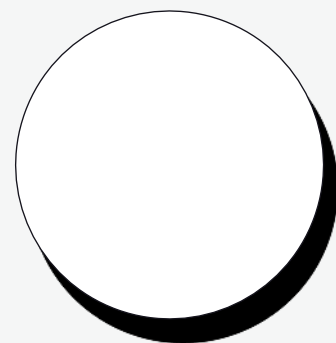
Conclusion

Synthesized's platform successfully enabled the European Commission to build a Data Hub that balances innovation enablement with uncompromising data protection. By ensuring that national supervisors can participate without exposing confidential information, the solution removes a critical barrier to fintech innovation across the EU single market.

The JRC validation confirmed that synthetic data provided sufficient analytical utility for training AI/ML models and testing innovative solutions while maintaining confidentiality standards that make traditional anonymization attempts “virtually impossible”. This approach represents a fundamental advancement over legacy anonymization techniques, allowing supervisors to gain insights into technologies used by innovative firms while providing fintech companies with the datasets they need to scale across Europe.

The Data Hub now serves as a collaborative space fostering closer relationships between innovative financial institutions and national supervisors, directly supporting the Commission's priorities of developing trustworthy data-sharing systems and enabling data-driven innovation in the European financial sector.

—
You can read the European Commission's findings directly in their independent report:
Synthetic data in the Data Hub of the Digital Finance Platform, Publications Office of the European Union, 2024, <https://op.europa.eu/s/AcLI>



With Synthesized, teams can instantly provision compliant test environments, clone, scale and refresh test datasets within minutes, and manage the entire test data lifecycle ensuring that test data is continuously available for automated and agentic workflows.