

# Synthetic Data

[QNS]



Unlock the value of your data, preserving its uniqueness, representativeness and confidentiality

## Main benefits

Enable data-intensive initiatives, in full compliance, even in the absence of data

Synthetic Data are artificially generated data that are realistic because they preserve the statistical and distribution properties of real data. The use of synthetic data is expected to far exceed that of real data in the coming years because synthetic data overcome the limitations of real data.

They adhere to privacy regulations as long as they do not contain sensitive information, they are reproducible because they are generated by an algorithm, and they can be used to study rare phenomena for which real data are scarce. Companies can enhance the use of their information assets by exploring new possibilities for growth and development and overcome challenges requiring high data quality.

## Scope

Synthetic Data find application in critical industries such as healthcare and finance, where accuracy, confidentiality and value of data is strategic to the success of new products.

For example in the field of testing they overcome obsolete masking techniques. In machine learning, they are the go-to choice for training models when data is scarce. Since synthetic data can be progressively improved over time with new real data, the value increases at a relatively low cost.

## Challenges facing companies

Companies face costs and complexities in collecting and preparing sufficiently representative, in the absence of which innovative initiatives are abandoned. The challenge has become increasingly daunting as regulation increases, restricting the use and monetization of data. These restrictions make it difficult for companies to fully exploit the value of their data, holding back the development of new business solutions and the pursuit of competitive advantages in the marketplace

## The advantages offered

The CIO enables innovative, data-intensive initiatives, meeting complex business needs in operational agility. The CDO unlocks the availability of representative data, reducing its management and acquisition costs. The CISO uses Synthetic Data to improve the protection of sensitive data, reducing security risks. Data scientists enjoy greater freedom in analysis and experimentation, accelerating innovation and value extracted from data



## S2E: approach and proposed methodology

Our company is the ideal partner to develop initiatives involving Synthetic Data due to our recognized expertise in big data management and advanced analytics. Our team of experts specializes in managing complex systems and addressing big data challenges. In partnership with Clearbox AI, an emerging leader in Synthetic Data generation, we offer innovative and secure solutions essential for analytics and data management suitable for highly regulated environments.

Our 'Data as a Code' approach enables end-to-end management of the Synthetic Data generation, management and deployment process, generating ready-to-use databases in a format similar to an 'executable code'. The collaboration between Clearbox AI and S2E is an effective pairing for comprehensive support in an area of Generative AI that has yet to be fully explored. Our expertise enables the rapid 'grounding' of the generated data, i.e., in its effective management and the ability to offer advice on both the architecture of the data and its practical management. This enables companies to enhance their initiatives with Synthetic Data, maximizing its effectiveness and efficiency.

### Future development benefits

The main benefits of Synthetic Data are:

1. Adherence to data privacy and security regulations.
2. Enable the study of rare or new phenomena of which there is little data available.
3. Democratization and availability of data within the organization by overcoming storage and reproducibility limitations.
4. Increased data quality and reduced risk of all data-driven monetization initiatives.
5. Transparency of synthetic data enables development of Explainable AI-based models, useful in applications that need documentation and understanding.
6. Data generation from algorithms: "Data as a Code" complements "Infrastructure as a Code" practices. In this way, the entire ICT asset is generated automatically, reducing setup costs.

### Offer model

Synthetic Data is offered based on project initiatives dedicated to generating synthetic data using specific models that replicate the characteristics of the original dataset.

We associate a managed service of qualified technical support for the end-to-end management of the data creation, management, integration, and implementation process within business initiatives, known as 'Data as a Code.'

The 'Data as a Code' managed service includes the generation, management, and grounding of regulated, optimized, and extended data on-demand, based on the needs of the client and its data-intensive initiatives.