

Accelerate your Business with synthetic data



Mattias Ripoll
CEO

SYNDATA

THE PROBLEM

Organizations today cannot make the best use of their data assets because of

- 1) They have too few data points or biases in the data**
- 2) GDPR and similar data protection regulations**
- 3) Data is considered to business sensitive to share**

mattias@syndata.co

+46 70 007 94 60

WHAT



Synthetic data is generated artificially to replicate the statistical properties of real data but without any identifiable information in the new data set.

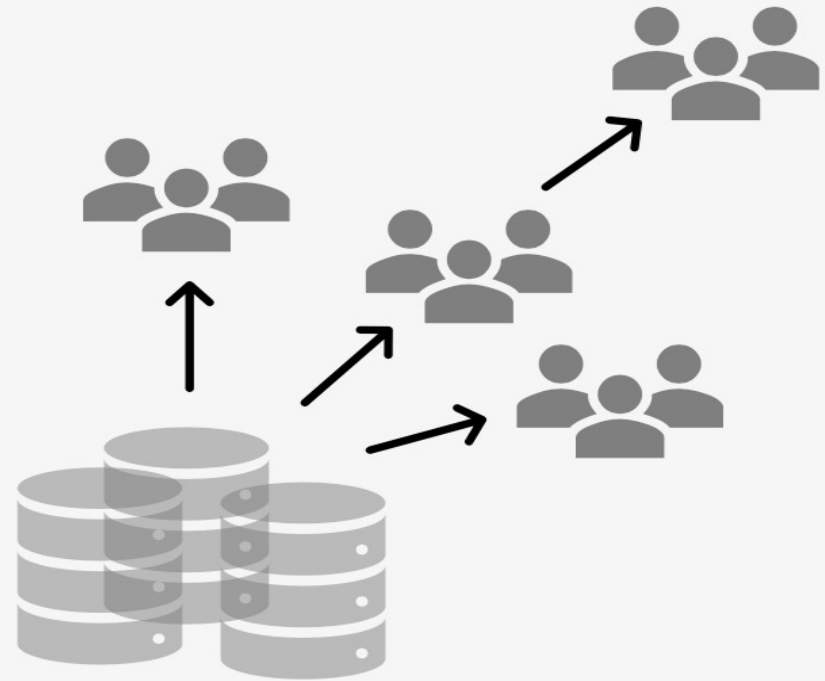
mattias@syndata.co

+46 70 007 94 60

Innovate, test & share your data in a simple & clever way



PRODUCTION DATA



SYNTHETIC DATA

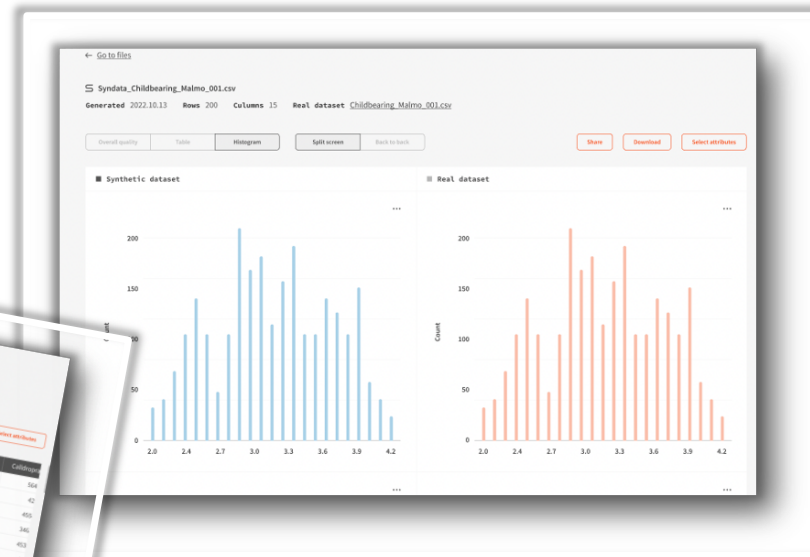
HOW

Syndata has built a synthetic engine that can create synthetic data of different quality depending on the purpose of the synthesis.

It's as simple as

1. Upload
2. Specify
3. Generate
4. Download

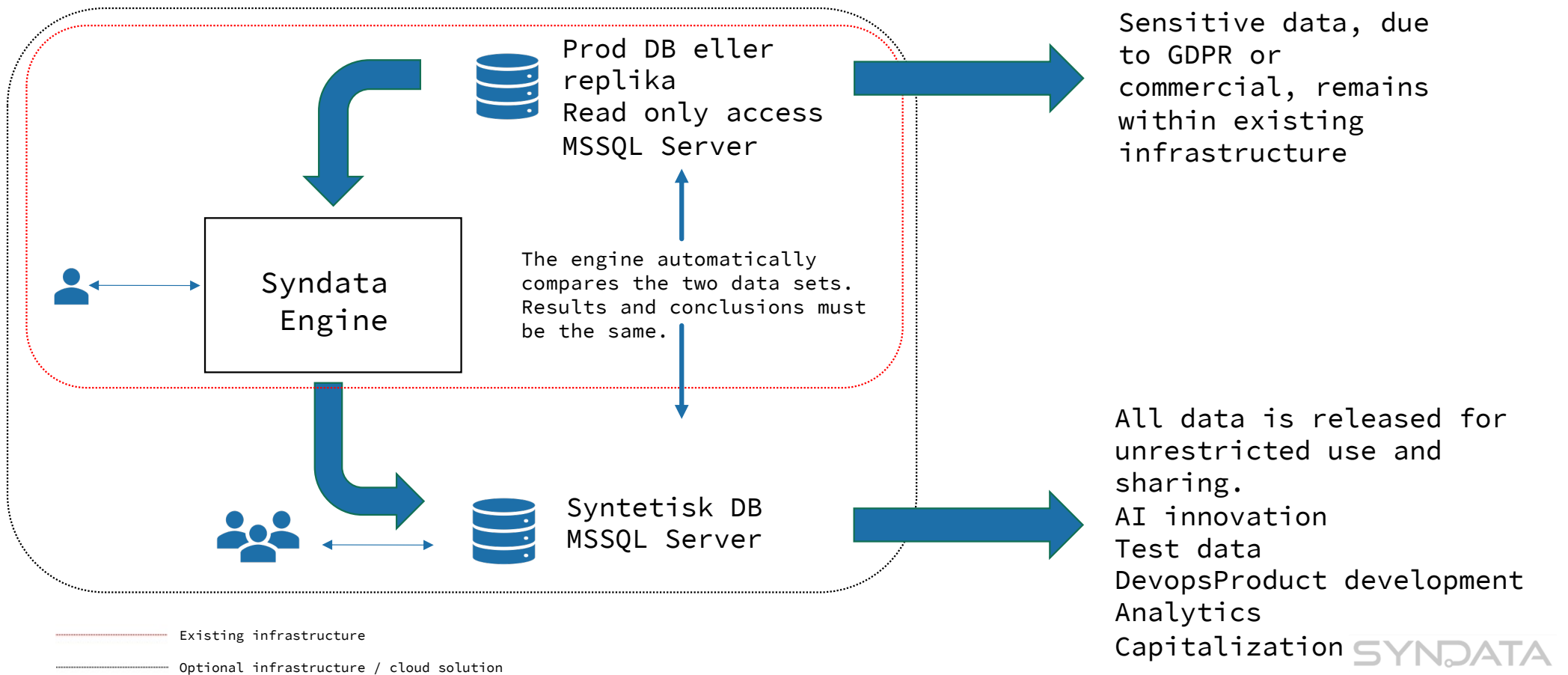
Age	Annuityscore	Calligraphy	Calligraphy2	Gender	Age	Annuityscore	Calligraphy	Calligraphy2	Gender	Age	Annuityscore	Calligraphy
32	136164	0.6	Phd	Male	34	235614	0.6	Phd	Male	32	129104	0.6
34	52342	0.3	Phd	Male	34	361365	0.3	Phd	Male	34	129104	0.6
45	345345	0.45	Phd	Male	34	279693	0.45	Phd	Male	65	345345	0.45
76	2345345	0.45	Phd	Male	65	52342	0.45	High school	Male	76	2345345	0.45
23	234534	0.45	Phd	Male	76	234534	0.45	High school	Male	23	234534	0.45
54	234534	0.14	High school	Male	54	52342	0.34	High school	Male	23	234534	0.45
54	52342	0.45	High school	Male	54	52342	0.45	High school	Male	54	52342	0.45
50	52342	0.34	High school	Male	54	52342	0.45	High school	Male	23	234534	0.45
12	12342	0.04	High school	Male	54	52342	0.45	High school	Male	23	234534	0.45
14	12342	0.04	Phd	Female	54	52342	0.45	High school	Male	23	234534	0.45
14	12342	0.04	Phd	Female	54	52342	0.45	High school	Male	23	234534	0.45
76	74238	0.35	Phd	Male	12	12342	0.34	High school	Male	54	52342	0.45
46	42345	0.45	Phd	Female	12	12342	0.34	High school	Male	54	52342	0.45
72	7423	0.45	Phd	Female	34	345345	0.34	Phd	Male	54	52342	0.45
33	7423	0.45	High school	Female	34	345345	0.34	Phd	Male	54	52342	0.45
89	82342	0.76	High school	Female	72	7423	0.45	Phd	Female	12	12342	0.34
23	2345	0.45	High school	Female	72	7423	0.45	Phd	Female	12	12342	0.34
23	2345	0.45	High school	Female	72	7423	0.45	Phd	Female	12	12342	0.34
13	1234	0.0056	Phd	Female	89	82342	0.76	High school	Female	66	62345	0.2345
13	1234	0.0056	Phd	Female	89	82342	0.76	High school	Female	66	62345	0.2345
23	2345	0.0056	Phd	Female	23	2345	0.0056	Phd	Female	13	1234	0.0056
23	2345	0.0056	Phd	Female	23	2345	0.0056	Phd	Female	13	1234	0.0056
23	2345	0.0056	Phd	Female	23	2345	0.0056	Phd	Female	13	1234	0.0056



Or through an automated process

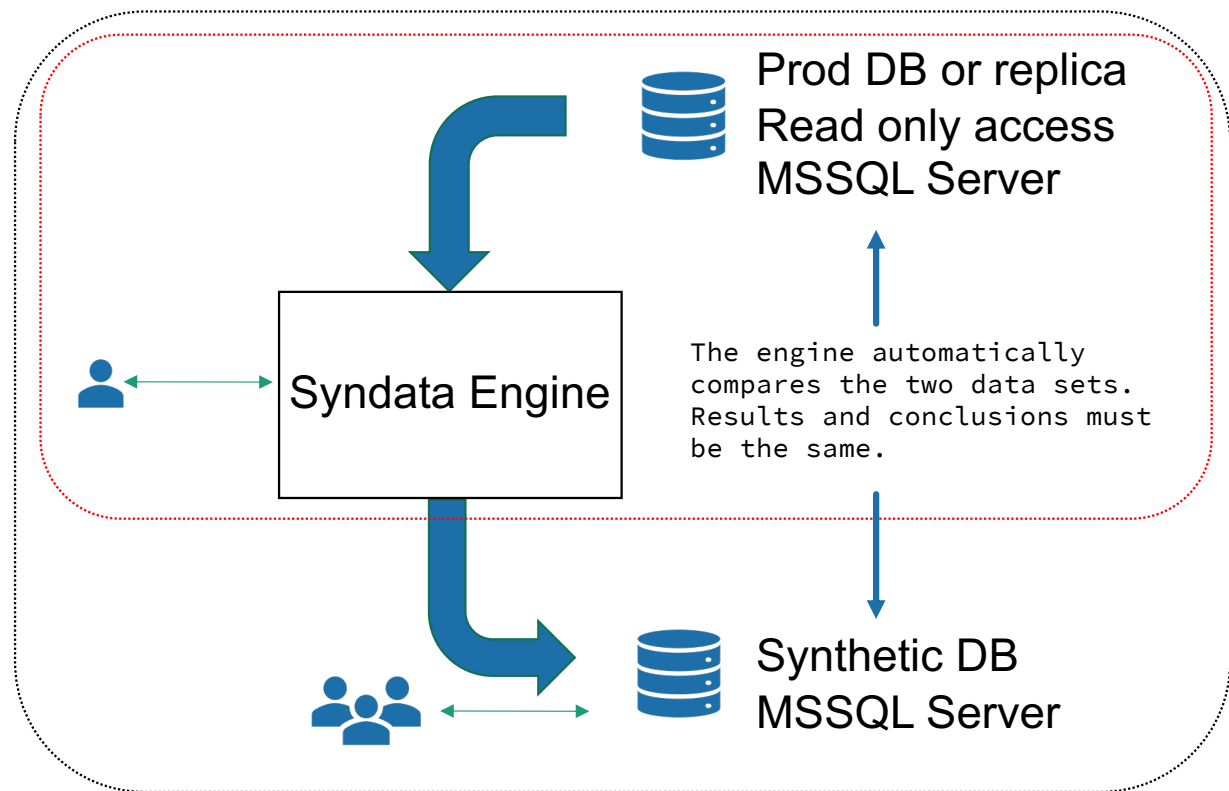
mattias@syndata.co
+46 70 007 94 60

HIGH LEVEL ARCHITECTURE



HIGH LEVEL ARCHITECTURE

- Syndata engine reads from prod DB, synthesize the data, and saves results to a synthetic DB.
- This can be done via a batch job (ETL ML pipelines) or on demand.
- Users with enough clearance interact with engine (webapp or via API)
- DB Infrastructure and other provided by client



INDUSTRY
Health-care

PROJECT OWNER

AI
S W E D E N

DATA OWNER

 region
västerbotten

Predicting Patient Recovery

Region Västerbotten want utilize AI modules for predicting time to discharge patients in intensive care (recovery/other) from their Patient Data Management System.

DATA COLLECTION & AI CHALLENGE

- Source: < 2000 patient observations with multiple parameters.
- Sensitive data (not possible to share between regions).

SYNDATA ENGINE ENABLED

- Generated a synthetic copy of the data sets. into scalable data to train the AI models
- Reviewed: Predictivity, Robustness, Evaluation

LINK

- <https://www.ai.se/en/projects-9/information-driven-healthcare/synthetic-data-in-intensive-care>

“Especially impressive is how the Syndata models have low error rate, i.e., they are capable of ranking patients according to the care length.”

SYNDATA

GDPR

CREATING TEST DATA
WITH WHILE
MAINTAINNING THE
STATISTICAL PROPERTIES

LACK OF
CONSISTNESS
IN TESTDATA

DATA GOVERNENCE

LACK OF DATA

MOVE FASTER
WITH
DEVELOPMENT

LESS COST AND LESS
RISK WHEN
PRODUCING TEST
DATA

HARD MOVING DATA
TO THE CLOUD

DATA SET IS TO
SMALL TO TRAIN
ALGORITM

QUITE COSTLY TO
ANONYMIZE AND
MOVE DATA

mattias@syndata.co

+46 70 007 94 60