

# Artificial Data Amplifier

ADA



# Why do our clients need **ADA**?

*Industry-wide, companies are finding it difficult to overcome data related challenges that hinders them from unlocking revolutionary AI solutions.*



01

## **GDPR Regulations**

- Strict on data usage
- Requires transparency



02

## **Sensitive Data**

- Customer data is sensitive
- Needs anonymisation
- Can't use production data



03

## **Financial Implications**

- Heavy fines for not complying with GDPR
- **4%** of annual global turnover or **€20M**



04

## **Data Scarcity**

- AI solutions require tons of historical data
- Quality data is hard to come by & difficult to make model robust

**ADA addresses data challenges and helps mitigating GDPR compliance risk**

# The problem with 'anonymous' data

*With machine learning methods, it is possible to re-identify **99.98%** of anonymized individuals in data sets*



Data anonymization is reversible

About 90% of anonymised data can be re-identified with just 4 attributes







Current approaches of subsetting and sampling data for anonymization is not enough

Only if data cannot be traced back to individuals can it be considered as 'anonymous' under GDPR

# Artificial Data Amplifier



## What is **ADA**?

-  A synthetic data generating solution using advanced deep learning
-  Generates all types (image, tabular, text etc.) of data
-  Data characteristics & relationships are preserved
-  Accelerates sophisticated AI models that need lots of data
-  Anonymizes data and helps in getting more GDPR compliant
-  ADA is not a generic data management tool, it is a custom method

# How does **ADA** work?

**Extract Data**



ETL to extract data




**ADA**



Upload to ADA docker on-premise or cloud




**Scrub/Mask**



Scrub data of PII (optional)




**Push Data**



ETL to push synthetic data back into applications/environment



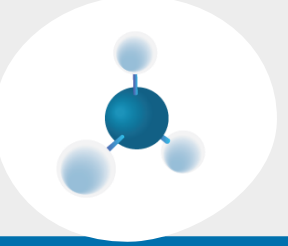
**Synthetic data**



Synthetic dataset that is scalable



**Generation**



Mapping relationships with advanced Neural Networks

# Applications of ADA



## Testing & Development

- Generate production data for testing
- Generate testing data for ML models and accelerate QA



## Health

- Generate medical records/claims
- Generate CT/MRI scans to boost ML models



## Financial Services

- Synthesise sensitive client data
- Boost scarce fraud detection data



## Insurance

- Synthesise claims data

### Top 3 needs in any industry:

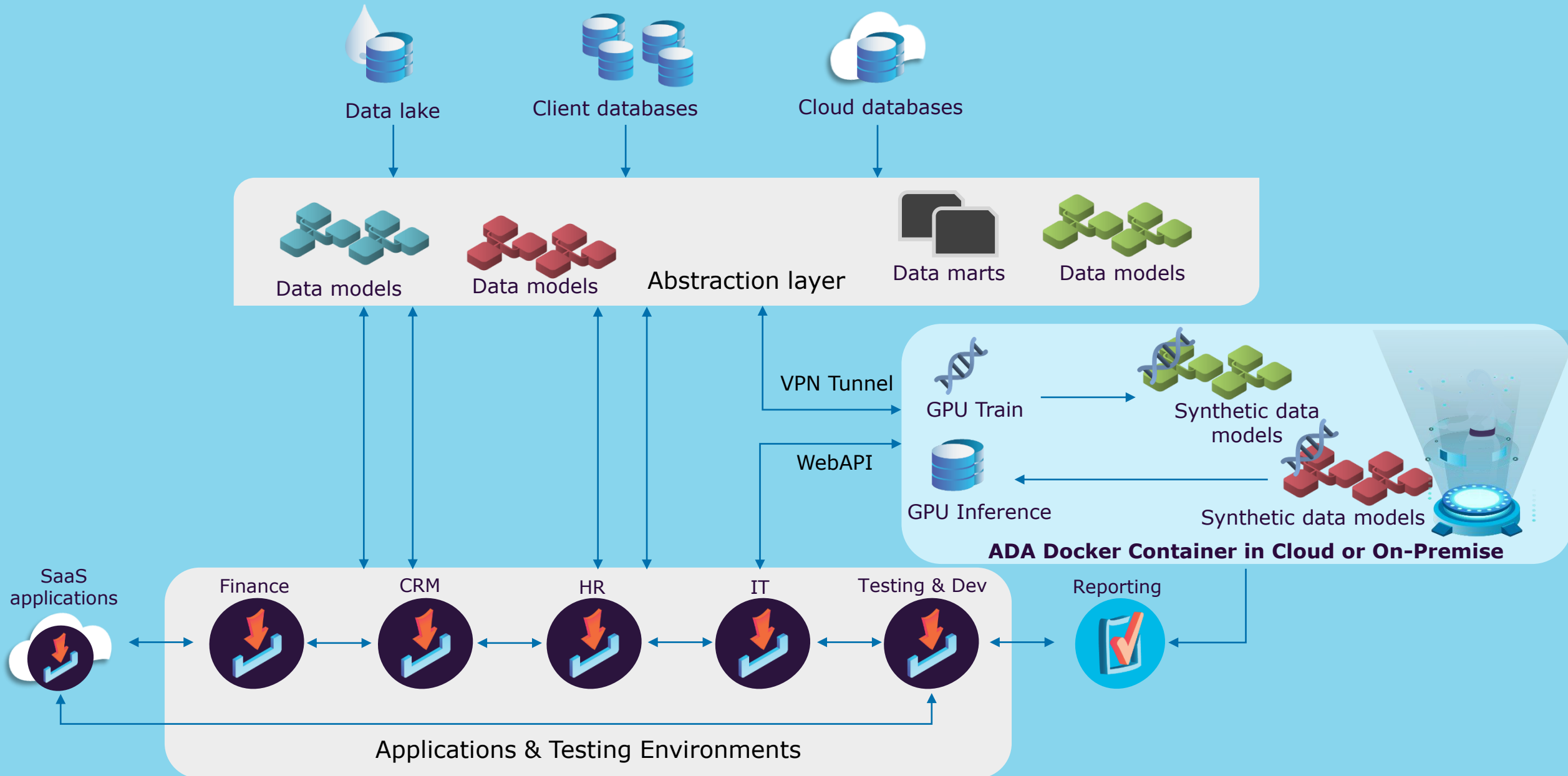
1. **GDPR compliant data**
2. **Anonymous, de-sensitised data**
3. **More data**



# ADA Solution Designs



# Agnostic Solution Design **with ADA**





# ADA Value Proposition



# How does **ADA** bring value to our clients?

## Risk-free, scalable dataset

GDPR compliant, synthetic, scalable data

## Exponential Opportunities

Dataset is interchangeable with real data and can unlock & accelerate many complex AI solutions

## High quality & speed

Fast delivery with an emphasis on quality

## ETL & DB agnostic

Can be easily implemented into Oracle, SQL, SAP etc.



# ADA Use Cases

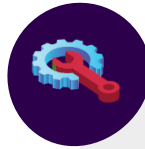


# ADA for a Dutch Financial Services



## Client Challenges

- Bank could not use production data for performance testing as it violates GDPR regulations
- Testing team was manually creating data for performance testing, which is a tedious and time consuming process
- Time taken to create data for 1 months worth of releases was 1 FTE



## Sogeti Solution

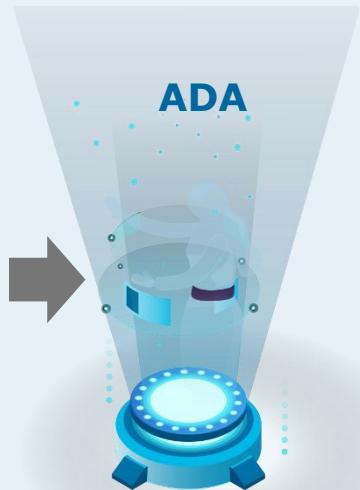
- Used ADA to generate synthetic testing data for performance testing
- From 1000 rows of real data, ADA generated 10K rows within 2 days
- Generated data was validated by testing the application



## Key Benefits delivered

- Can use generated data to test as generated data is scalable
- Testers can reallocate tasks
- More compliant with GDPR and does not pose security risk

	A	B	C	D	E	F
	Country	Salesperson	Order Date	OrderID	Units	Order Amount
1	USA	Fuller	1/01/2011	10392	13	1,440.00
2	UK	Gloucester	2/01/2011	10397	17	716.72
3	UK	Bromley	2/01/2011	10771	18	344.00
4	USA	Finchley	3/01/2011	10393	16	2,556.95
5	USA	Finchley	3/01/2011	10394	10	442.00
6	UK	Gillingham	3/01/2011	10395	9	2,122.92
7	USA	Finchley	6/01/2011	10396	7	1,903.80
8	USA	Callahan	8/01/2011	10399	17	1,765.60
9	USA	Fuller	8/01/2011	10404	7	1,591.25
10	USA	Fuller	9/01/2011	10398	11	2,505.60
11	USA	Coghill	9/01/2011	10403	18	855.01
12	USA	Finchley	10/01/2011	10401	7	3,868.60
13	USA	Callahan	10/01/2011	10402	11	2,713.50
14	UK	Rayleigh	13/01/2011	10406	15	1,830.78
15	USA	Callahan	14/01/2011	10408	10	1,622.40
16	USA	Farnham	14/01/2011	10409	19	319.20
17	USA	Farnham	15/01/2011	10410	16	802.00



	A	B	C	D	E	F
	Country	Salesperson	Order Date	OrderID	Units	Order Amount
1	USA	Fuller	1/01/2011	10392	13	1,440.00
2	UK	Gloucester	2/01/2011	10397	17	716.72
3	UK	Bromley	2/01/2011	10771	18	344.00
4	USA	Finchley	3/01/2011	10393	16	2,556.95
5	USA	Finchley	3/01/2011	10394	10	442.00
6	UK	Gillingham	3/01/2011	10395	9	2,122.92
7	USA	Finchley	6/01/2011	10396	7	1,903.80
8	USA	Callahan	8/01/2011	10399	17	1,765.60
9	USA	Fuller	8/01/2011	10404	7	1,591.25
10	USA	Fuller	9/01/2011	10398	11	2,505.60
11	USA	Coghill	9/01/2011	10403	18	855.01
12	USA	Finchley	10/01/2011	10401	7	3,868.60
13	USA	Callahan	10/01/2011	10402	11	2,713.50
14	UK	Rayleigh	13/01/2011	10406	15	1,830.78
15	USA	Callahan	14/01/2011	10408	10	1,622.40
16	USA	Farnham	14/01/2011	10409	19	319.20
17	USA	Farnham	15/01/2011	10410	16	802.00
18	USA	Farnham	15/01/2011	10410	16	802.00
4	UK	Bromley	2/01/2011	10771	18	344.00
5	USA	Finchley	3/01/2011	10393	16	2,556.95
6	USA	Finchley	3/01/2011	10394	10	442.00
7	UK	Gillingham	3/01/2011	10395	9	2,122.92
8	USA	Finchley	6/01/2011	10396	7	1,903.80
9	USA	Callahan	8/01/2011	10399	17	1,765.60
10	USA	Fuller	8/01/2011	10404	7	1,591.25
11	USA	Fuller	9/01/2011	10398	11	2,505.60
12	USA	Coghill	9/01/2011	10403	18	855.01
13	USA	Finchley	10/01/2011	10401	7	3,868.60
14	USA	Callahan	10/01/2011	10402	11	2,713.50
15	UK	Rayleigh	13/01/2011	10406	15	1,830.78
16	USA	Callahan	14/01/2011	10408	10	1,622.40
17	USA	Farnham	14/01/2011	10409	19	319.20
18	USA	Farnham	15/01/2011	10410	16	802.00

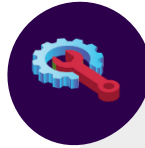


# ADA for US Accountancy Advisory



## Client Challenges

- Firm could not use production data for testing as it violates privacy regulations
- Production data could not be transferred intra-company due to sensitivity of information and privacy regulations



## Sogeti Solution

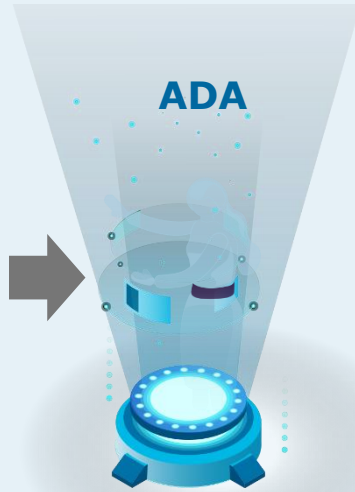
- Initiated as a Proof of Value and then scaled up for wider implementation
- Use ADA to generate synthetic testing data for 50+ applications
- Delivered as an API micro-service and deployed as an Azure Cloud solution



## Key Expected Benefits

- Can use generated data to test applications while being regulatory compliant
- Can use scalable generated data for performance testing
- Can generate data on demand

	A	B	C	D	E	F
	Country	Salesperson	Order Date	OrderID	Units	Order Amount
1	USA	Fuller	1/01/2011	10392	13	1,440.00
2	UK	Gloucester	2/01/2011	10397	17	716.72
3	UK	Bromley	2/01/2011	10771	18	344.00
4	USA	Finchley	3/01/2011	10393	16	2,556.95
5	USA	Finchley	3/01/2011	10394	10	442.00
6	UK	Gillingham	3/01/2011	10395	9	2,122.92
7	USA	Finchley	6/01/2011	10396	7	1,903.80
8	USA	Callahan	8/01/2011	10399	17	1,765.60
9	USA	Fuller	8/01/2011	10404	7	1,591.25
10	USA	Fuller	9/01/2011	10398	11	2,505.60
11	USA	Coghill	9/01/2011	10403	18	855.01
12	USA	Finchley	10/01/2011	10401	7	3,868.60
13	USA	Callahan	10/01/2011	10402	11	2,713.50
14	UK	Rayleigh	13/01/2011	10406	15	1,830.78
15	USA	Callahan	14/01/2011	10408	10	1,622.40
16	USA	Farnham	14/01/2011	10409	19	319.20
17	USA	Farnham	15/01/2011	10410	16	802.00
18	USA	Farnham	15/01/2011	10410	16	802.00



	A	B	C	D	E	F
	Country	Salesperson	Order Date	OrderID	Units	Order Amount
1	USA	Fuller	1/01/2011	10392	13	1,440.00
2	USA	Fuller	1/01/2011	10392	13	1,440.00
3	UK	Gloucester	2/01/2011	10397	17	716.72
4	UK	Bromley	2/01/2011	10771	18	344.00
5	USA	Finchley	3/01/2011	10393	16	2,556.95
6	USA	Finchley	3/01/2011	10394	10	442.00
7	UK	Gillingham	3/01/2011	10395	9	2,122.92
8	USA	Finchley	6/01/2011	10396	7	1,903.80
9	USA	Callahan	8/01/2011	10399	17	1,765.60
10	USA	Fuller	8/01/2011	10404	7	1,591.25
11	USA	Fuller	9/01/2011	10398	11	2,505.60
12	USA	Coghill	9/01/2011	10403	18	855.01
13	USA	Finchley	10/01/2011	10401	7	3,868.60
14	USA	Callahan	10/01/2011	10402	11	2,713.50
15	UK	Rayleigh	13/01/2011	10406	15	1,830.78
16	USA	Callahan	14/01/2011	10408	10	1,622.40
17	USA	Farnham	14/01/2011	10409	19	319.20
18	USA	Farnham	15/01/2011	10410	16	802.00
19	UK	Bromley	2/01/2011	10771	18	344.00
20	USA	Finchley	3/01/2011	10393	16	2,556.95
21	USA	Finchley	3/01/2011	10394	10	442.00
22	UK	Gillingham	3/01/2011	10395	9	2,122.92
23	USA	Finchley	6/01/2011	10396	7	1,903.80
24	USA	Callahan	8/01/2011	10399	17	1,765.60
25	USA	Fuller	8/01/2011	10404	7	1,591.25
26	USA	Fuller	9/01/2011	10398	11	2,505.60
27	USA	Coghill	9/01/2011	10403	18	855.01
28	USA	Finchley	10/01/2011	10401	7	3,868.60
29	USA	Callahan	10/01/2011	10402	11	2,713.50
30	UK	Rayleigh	13/01/2011	10406	15	1,830.78
31	USA	Callahan	14/01/2011	10408	10	1,622.40
32	USA	Farnham	14/01/2011	10409	19	319.20
33	USA	Farnham	15/01/2011	10410	16	802.00

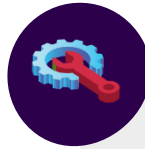


# ADA for Swedish Social Insurance Agency



## Client Challenges

- Medical records, x-rays and medical text contain highly sensitive data and cannot be analysed because of confidentiality
- This poses a security threat in terms of privacy violation and GDPR/ other regulatory non-compliance.
- Processes containing medical records are manual and tedious.



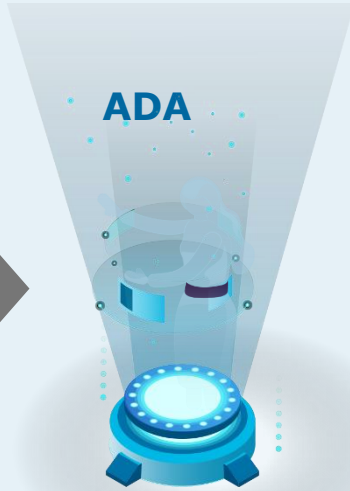
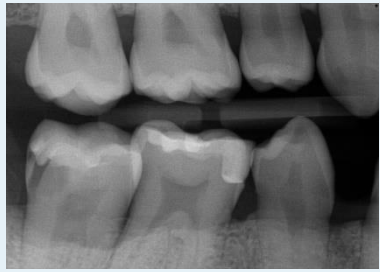
## Sogeti Solution

- Used ADA to generate synthetic testing data (hospital discharge data), medical x-rays and unstructured medical symptoms from medical records
- Used ADA to generate claims documents to boost OCR initiative

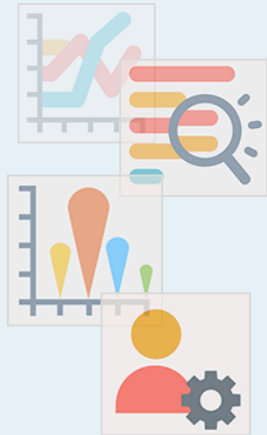
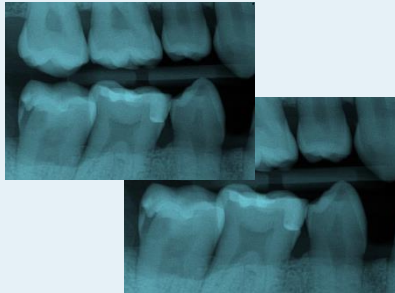


## Key Benefits delivered

- Can leverage data company-wide and deploy solutions to analyze the data
- Can automate tasks
- More compliant with GDPR and does not pose on-premise data security risk



History	Date December 10, 2001
ID	Ms. XY is a 23-year-old single female college student
CC	"Headache"
HPI	Ms. XY presents with a one-month history of headaches. The headaches are located in a band like distribution around her forehead and occipital area.
Physical Examination	
Vital Signs	BP 110/68 P 72 RR 14 T 98.4F Wt 160lb Ht 68 in
CV	No jugular venous distension. No carotid bruits. Apical impulse mid-clavicular line, 5th intercostal space. Regular rhythm. Normal S1, S2 with physiological split. No S3 or S4. No murmur or rub. No femoral bruit. Pulses: dorsalis pedis, posterior tibialis, femoral and radial 2+. Abdomen Flat.



History	Date December 10, 2001
ID	Ms. XY is a 23-year-old single female college student
CC	"Headache"
HPI	Ms. XY presents with a one-month history of headaches. The headaches are located in a band like distribution around her forehead and occipital area.
Physical Examination	
Vital Signs	BP 110/68 P 72 RR 14 T 98.4F Wt 160lb Ht 68 in
CV	No jugular venous distension. No carotid bruits. Apical impulse mid-clavicular line, 5th intercostal space. Regular rhythm. Normal S1, S2 with physiological split. No S3 or S4. No murmur or rub. No femoral bruit. Pulses: dorsalis pedis, posterior tibialis, femoral and radial 2+. Abdomen Flat.

## About Sogeti

Sogeti is a leading provider of technology and engineering services. Sogeti delivers solutions that enable digital transformation and offers cutting-edge expertise in Cloud, Cybersecurity, Digital Manufacturing, Digital Assurance & Testing, and emerging technologies. Sogeti combines agility and speed of implementation with strong technology supplier partnerships, world class methodologies and its global delivery model, Rightshore®. Sogeti brings together more than 25,000 professionals in 15 countries, based in over 100 locations in Europe, USA and India. Sogeti is a wholly-owned subsidiary of Capgemini SE, listed on the Paris Stock Exchange.

Learn more about us at

[www.sogeti.com](http://www.sogeti.com)



This message contains information that may be privileged or confidential and is the property of the Capgemini Group.

Copyright© 2019 Sogeti. All rights reserved.