

## ia generativa in medicina: i dati sintetici



## focus sul dato

paziente



dati socioeconomici  
anamnesi



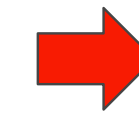
dati clinici



dati omici



immagini  
biomediche



db azienda  
sanitaria



#sanita2030



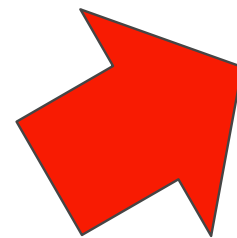
[www.sanita2030.it](http://www.sanita2030.it)



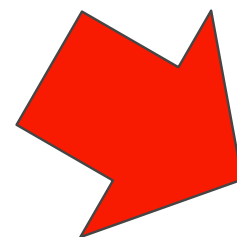
focus sul dato: criticità



db  
azienda  
sanitaria



privacy



scarsità



patologia #1

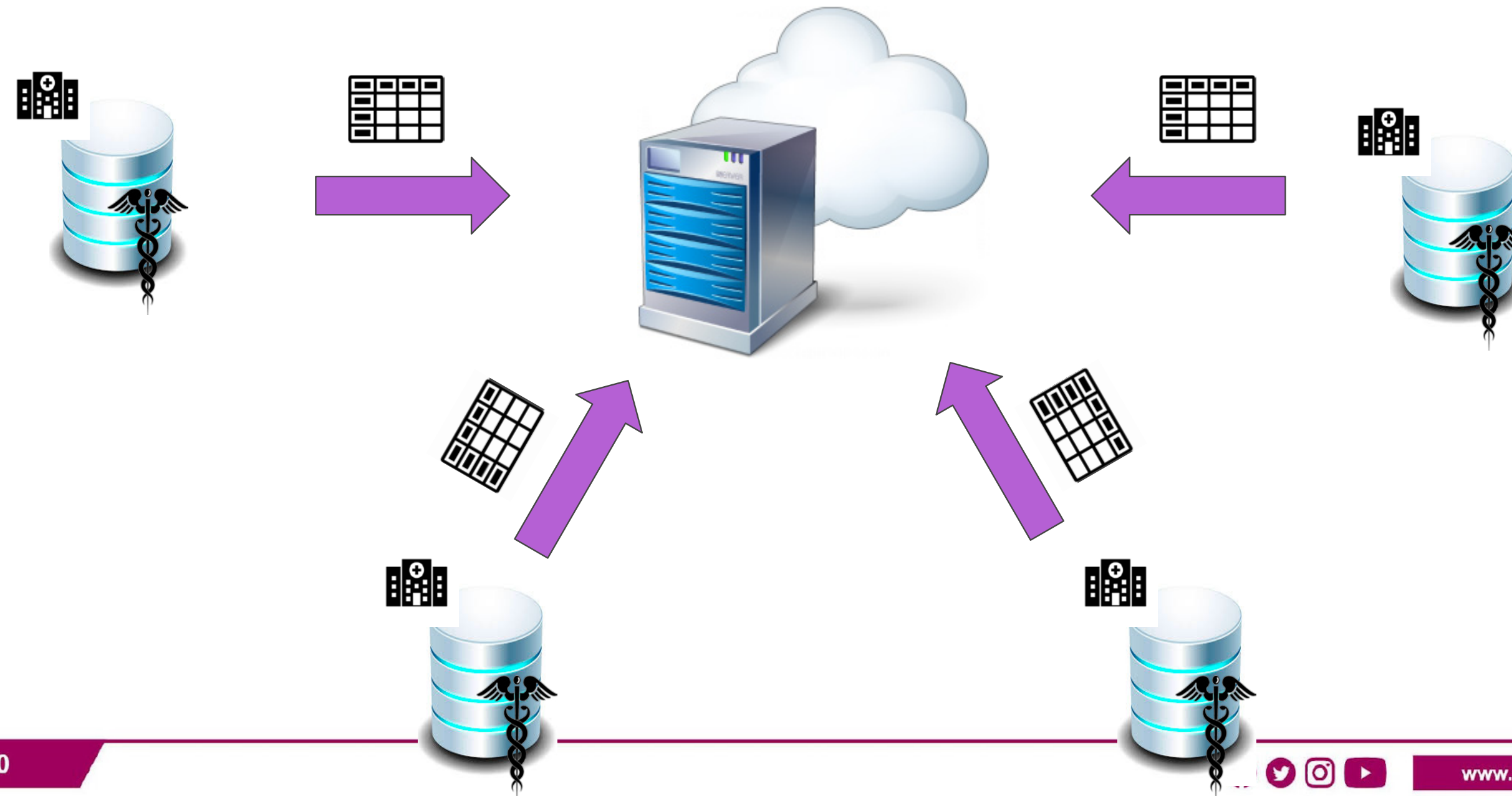
patologia #2

patologia #3

patologia #4



## studi multicentro



#sanita2030



[www.sanita2030.it](http://www.sanita2030.it)

## studi multicentro

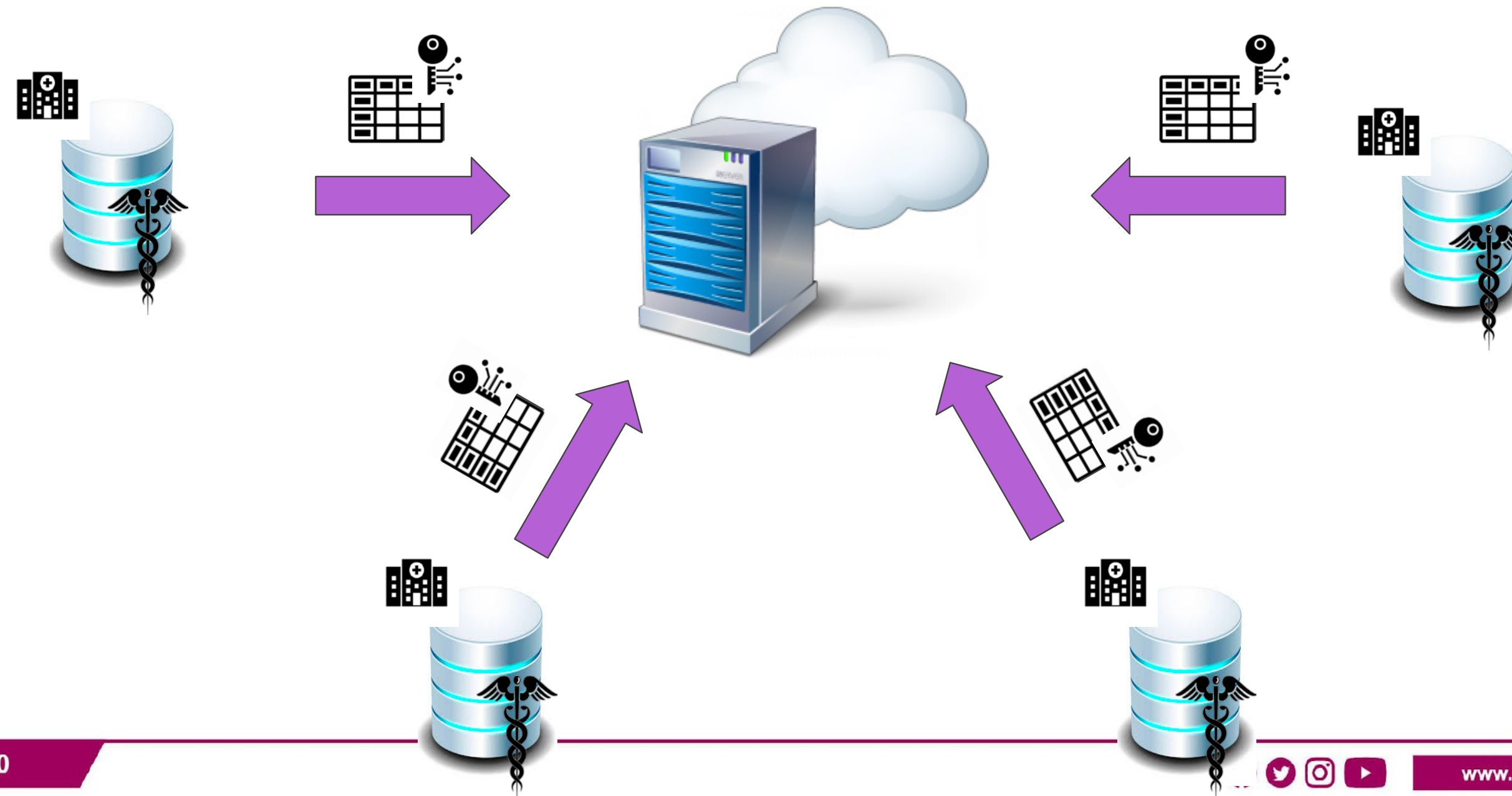


#sanita2030

www.sanita2030.it



## crittazione omomorfa

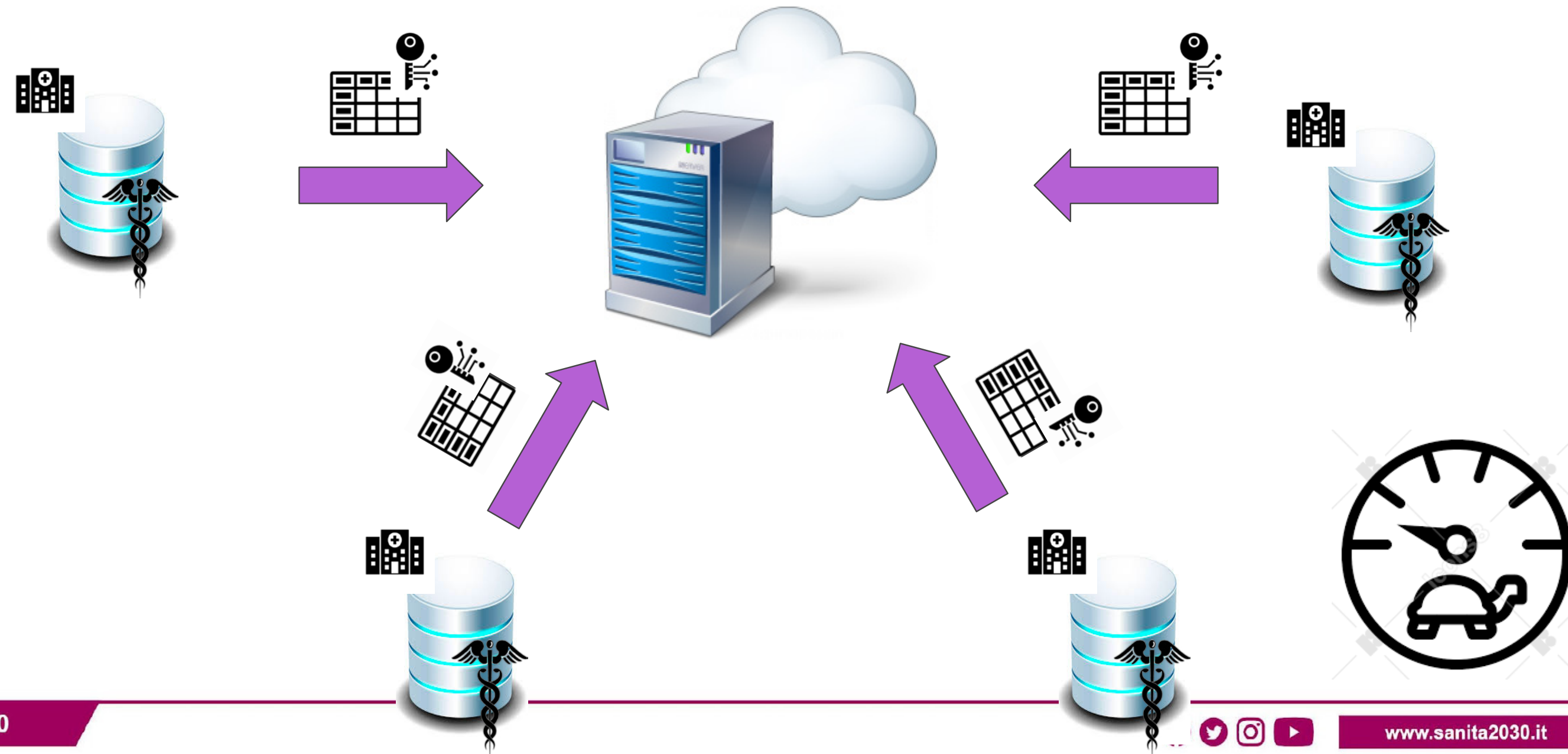


#sanita2030



www.sanita2030.it

## crittazione omomorfa



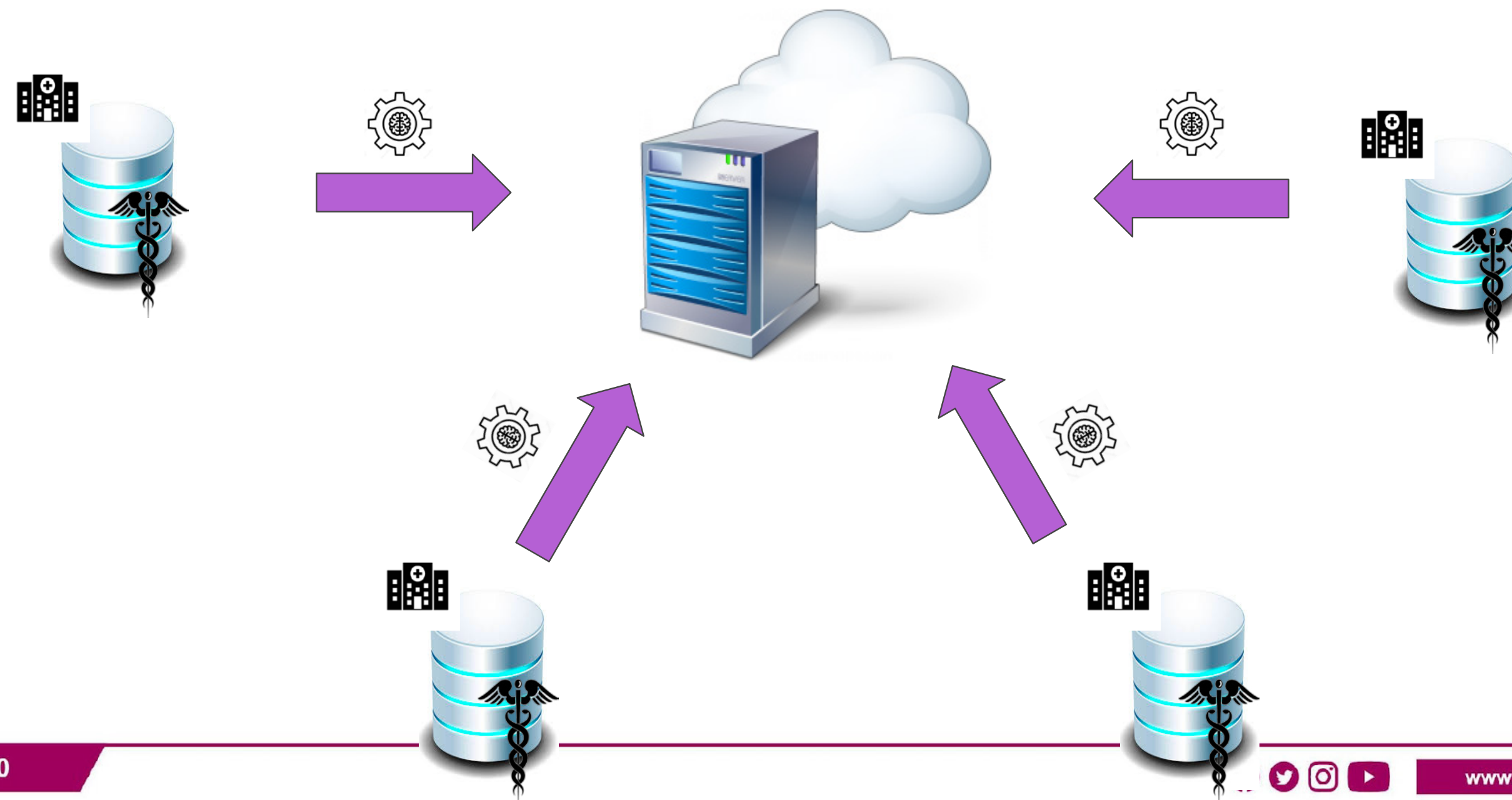
#sanita2030



www.sanita2030.it



## apprendimento federato



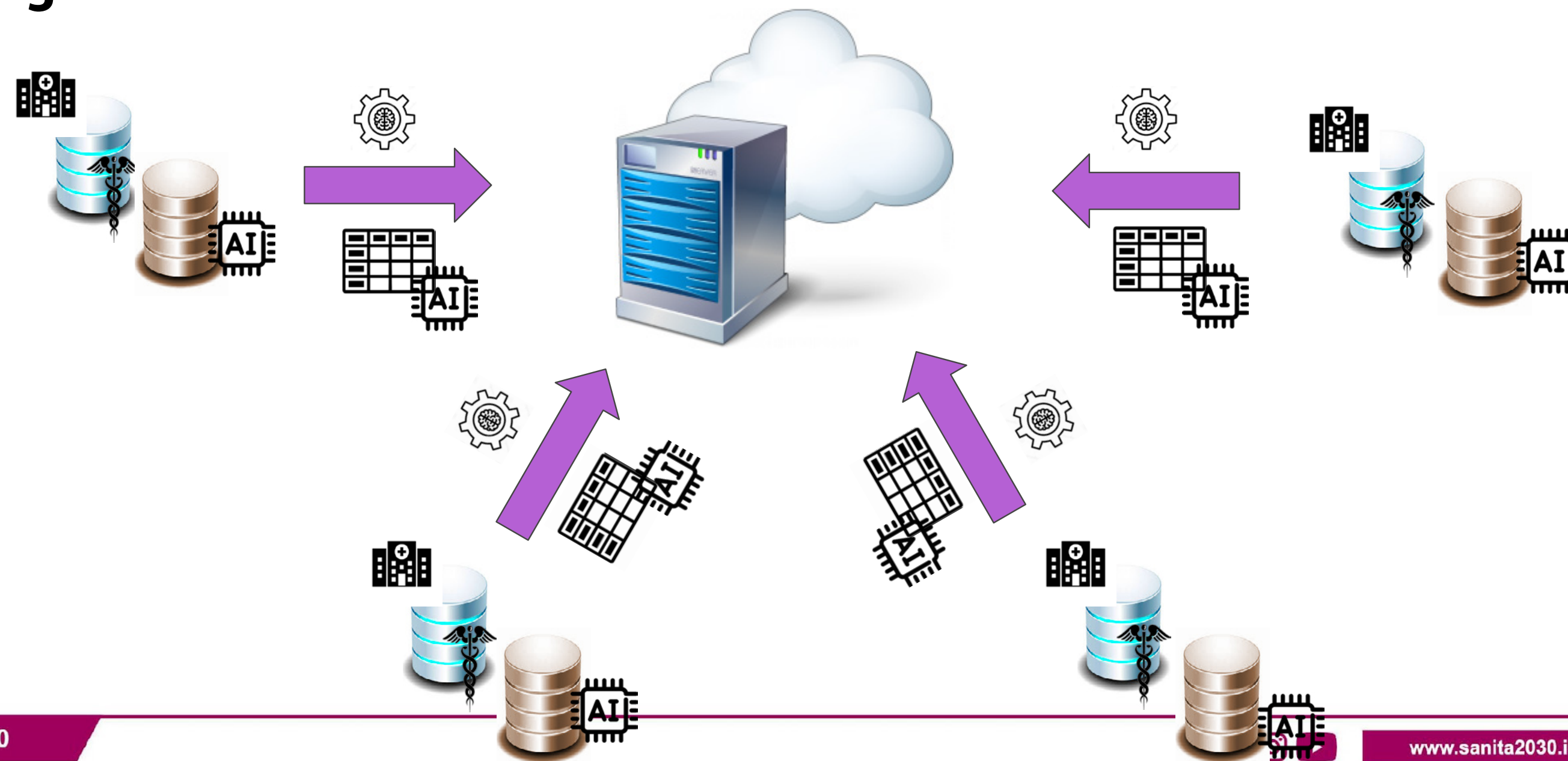
#sanita2030



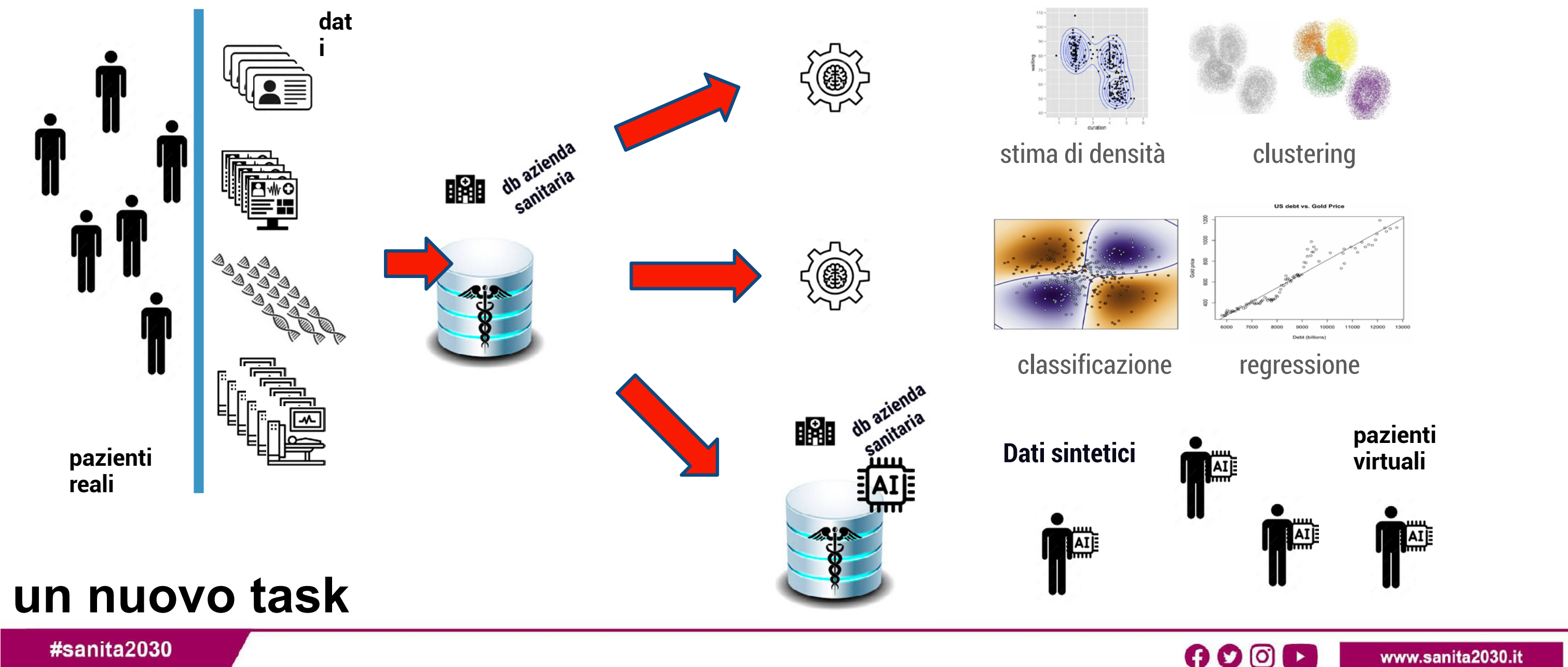
[www.sanita2030.it](http://www.sanita2030.it)



## data augmentation











"Microdata records created by statistically modeling original data and then using those models to generate new data values that reproduce the original data's statistical properties"

United States  
**Census**  
Bureau



Real Data



Simulation/Generative  
Model



Synthetic Data

- ✓ incrementare la dimensione dei dati per migliorare il training dei modelli
- ✓ evitare le criticità legata alla privacy
- ✓ diversificare la struttura del dataset
- ✓ fornire un'alternativa didattica

### PLOS DIGITAL HEALTH

OPEN ACCESS PEER-REVIEWED  
RESEARCH ARTICLE

#### Synthetic data in health care: A narrative review

Aldren Gonzales, Guruprabha Guruswamy, Scott R. Smith

Published: January 6, 2023 • <https://doi.org/10.1371/journal.pdig.0000082>

### THE LANCET

CORRESPONDENCE | VOLUME 399, ISSUE 10335, P1601-1602, APRIL 23, 2022

#### Synthetic patient data in health care: a widening legal loophole

Anmol Arora, Ananya Arora

Published: March 28, 2022 • DOI: [https://doi.org/10.1016/S0140-6736\(22\)00232-X](https://doi.org/10.1016/S0140-6736(22)00232-X)

### nature

Explore content About the journal Publish with us Subscribe

nature > outlook > article

OUTLOOK | 27 April 2023

#### Synthetic data could be better than real data

Machine-generated data sets have the potential to improve privacy and representation in artificial intelligence, if researchers can find the right balance between accuracy and fakery.

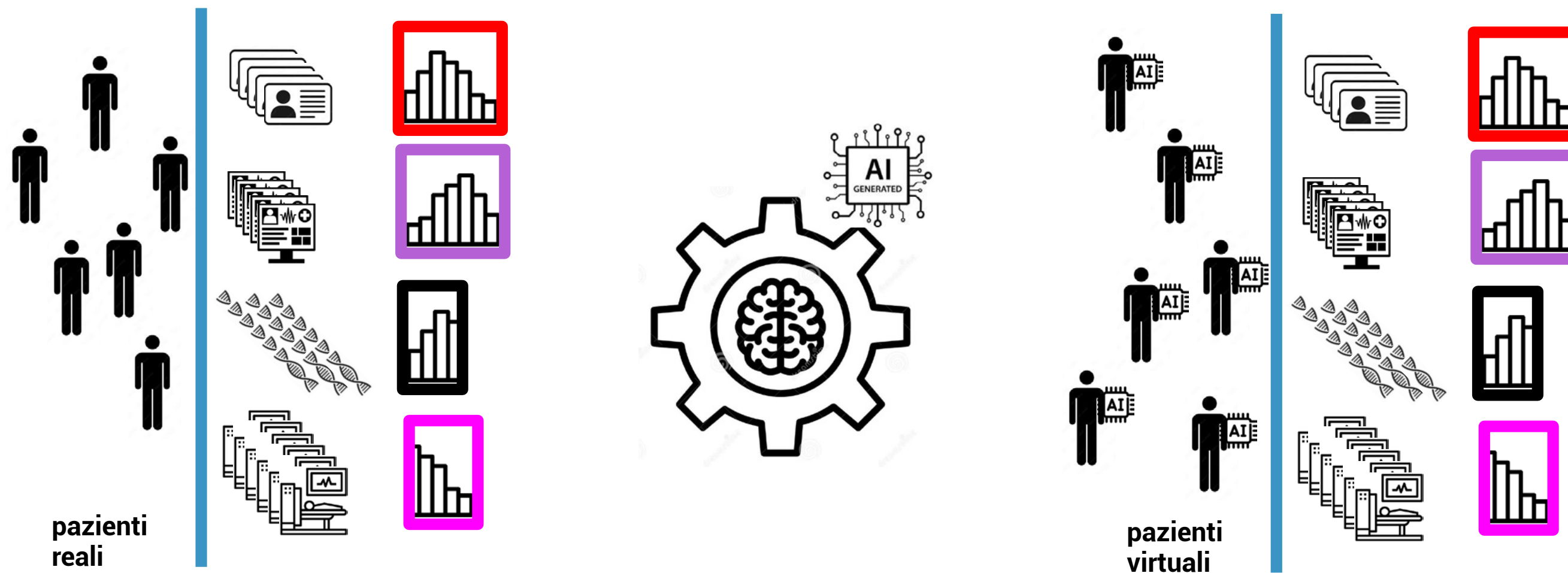
[Neil Savage](#)

#sanita2030



[www.sanita2030.it](http://www.sanita2030.it)





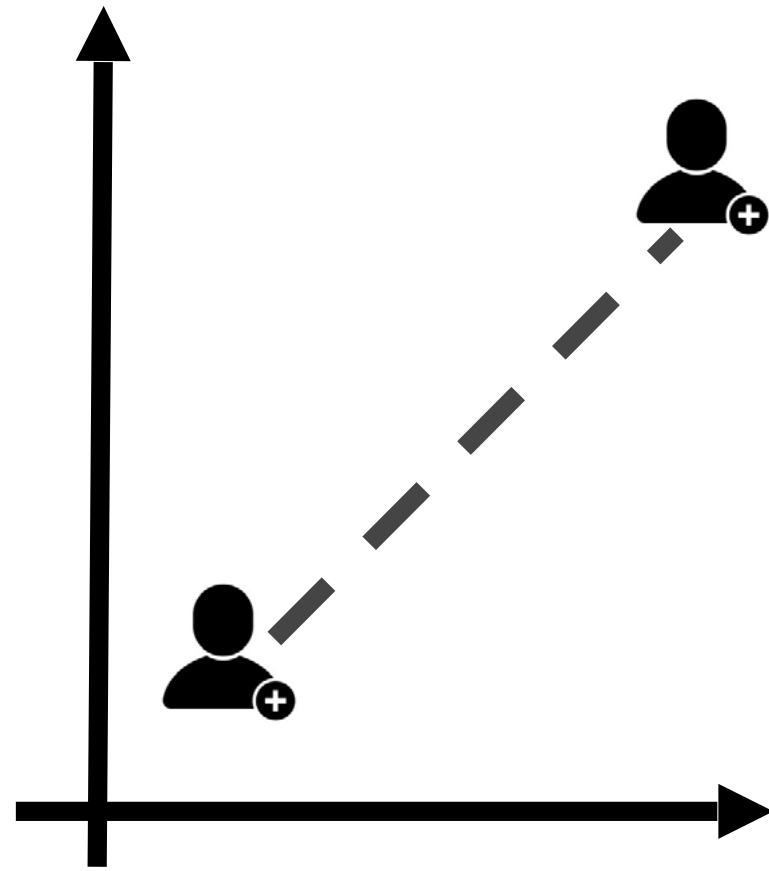
## Una popolazione di (non) pazienti

#sanita2030

I dati sintetici generati da AI e associati ai pazienti virtuali conservano la stessa distribuzione dei dati reali



● *statistico/matematico*



**dati sintetici**

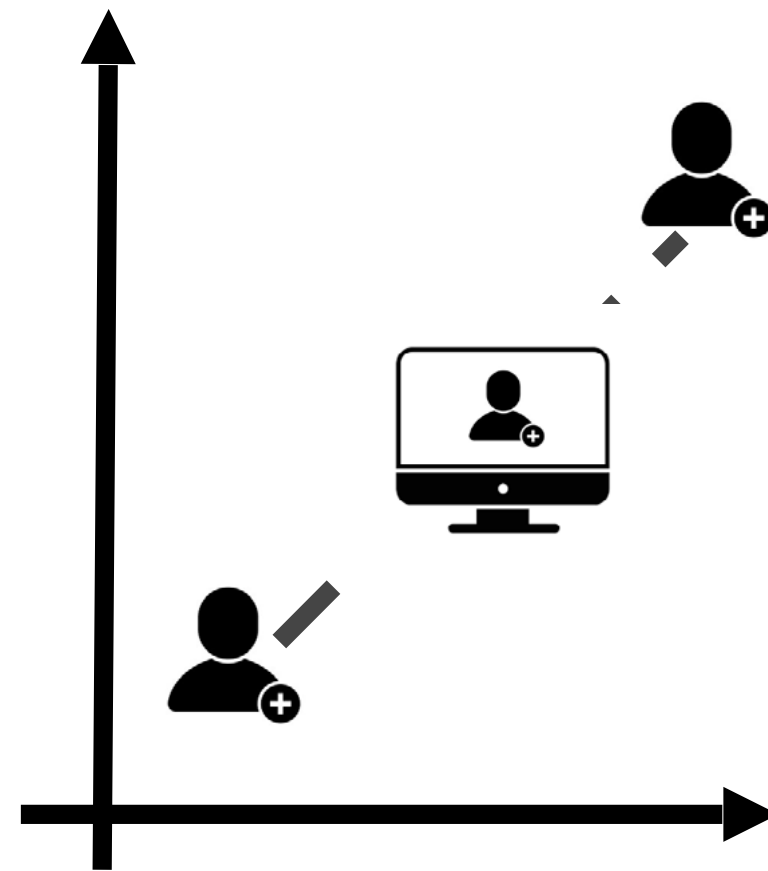
#sanita2030



[www.sanita2030.it](http://www.sanita2030.it)



● *statistico/matematico*



**dati sintetici**

#sanita2030



[www.sanita2030.it](http://www.sanita2030.it)

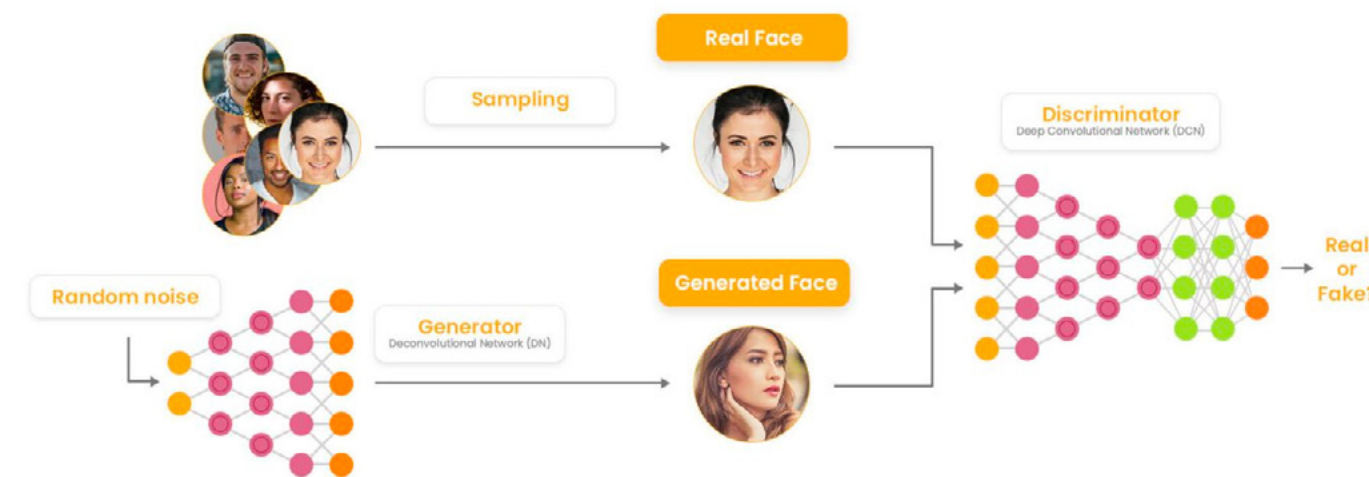




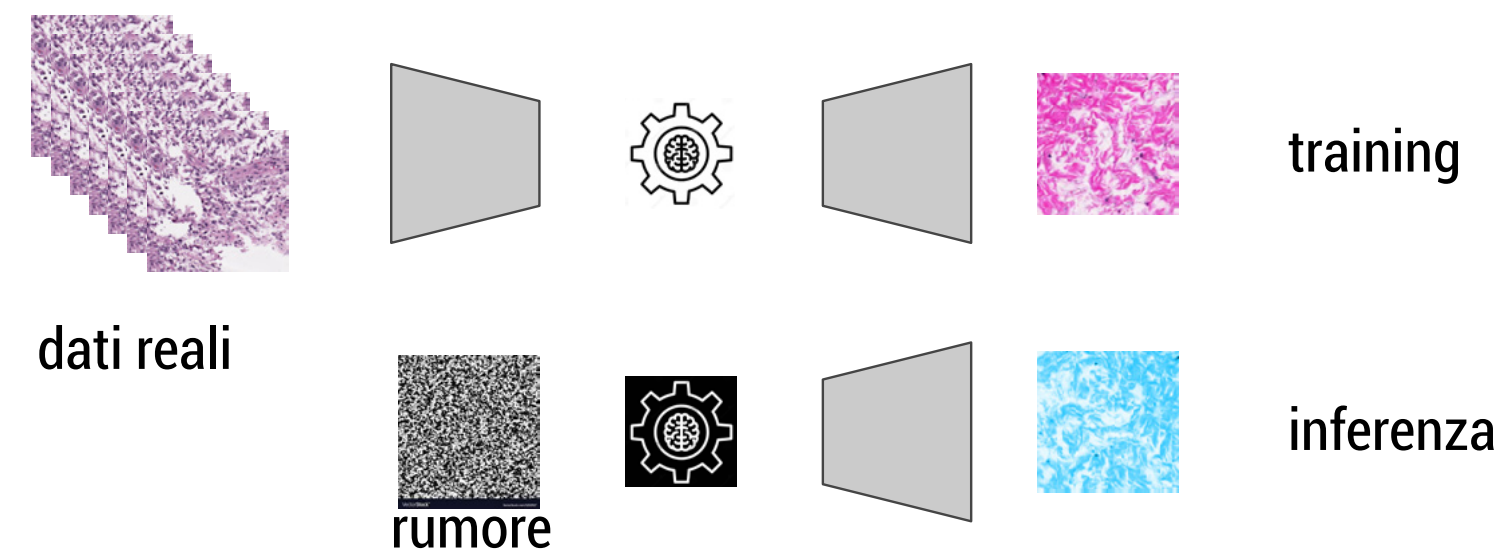
● **AI generativa**

L'INTELLIGENZA ARTIFICIALE IN SANITÀ  
TRA ETICA E APPLICAZIONE

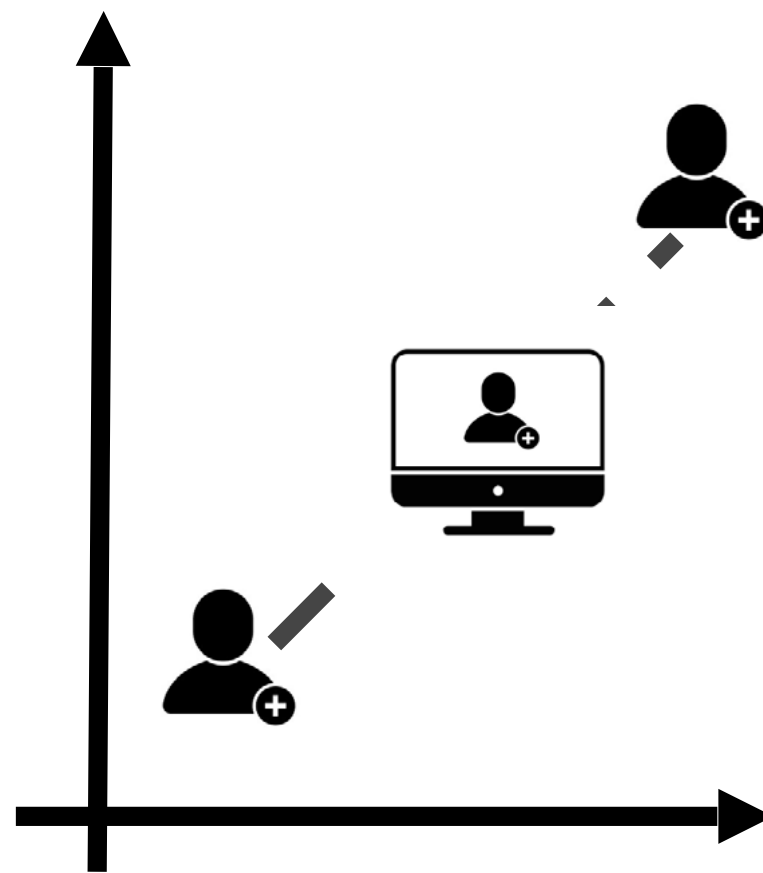
✓ generative adversarial networks (GAN)



✓ diffusion models



● **statistico/matematico**



dati sintetici

#sanita2030



www.sanita2030.it



nella pratica si usano differenti metodi e metriche diversi per valutare la qualità dei dati sintetici.

## PLOS DIGITAL HEALTH

OPEN ACCESS PEER-REVIEWED  
RESEARCH ARTICLE

### Synthetic data in health care: A narrative review

Aldren Gonzales, Guruprabha Guruswamy, Scott R. Smith

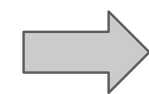
Published: January 6, 2023 • <https://doi.org/10.1371/journal.pdig.0000082>

*"Currently, there is no benchmark or standard for validating synthetic data"*

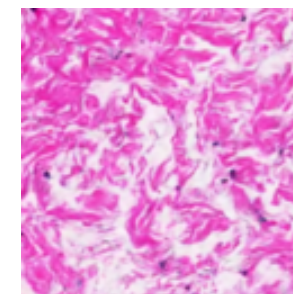
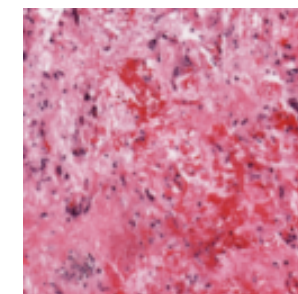
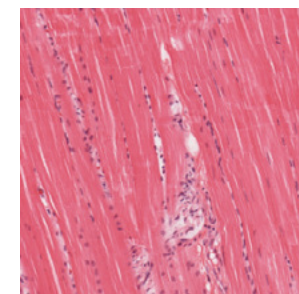
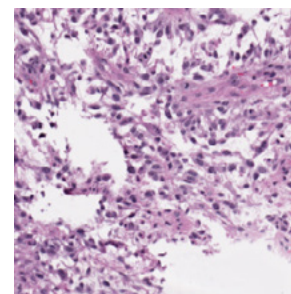


Diffusion Probabilistic Model per la generazione di vetrini sintetici (WSI) in anatomia patologica

150 WSI x 5 tessuti (TCIA)  
utero/polmone/cervello/rene/pancreas



49284 tasselli sintetici



## Validazione tramite valutazione di tre patologi

classificare 100 tasselli come reali o sintetici

		Patologi	
		S	R
Modello	S	104	46
	R	86	64

determinare il tessuto di origine di 100 tasselli  
accuratezza media 90%

## validazione

#sanita2030



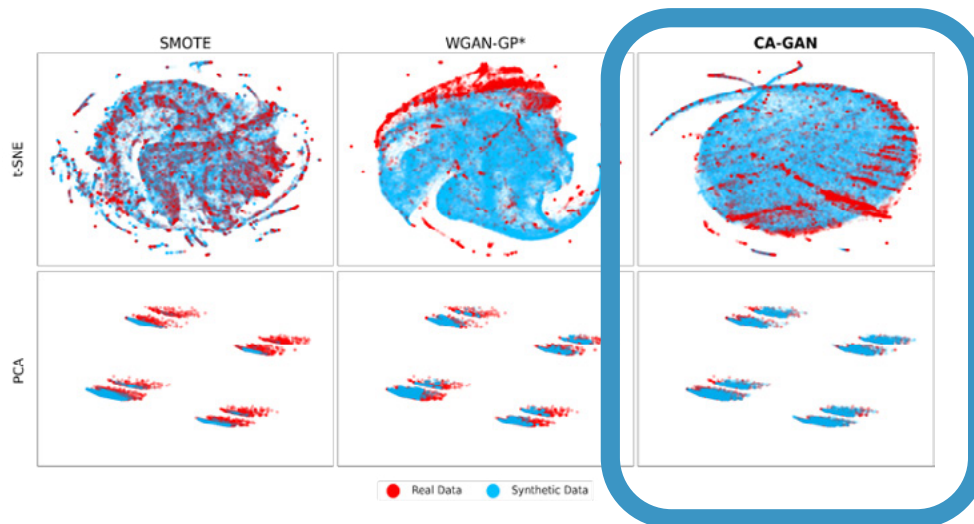
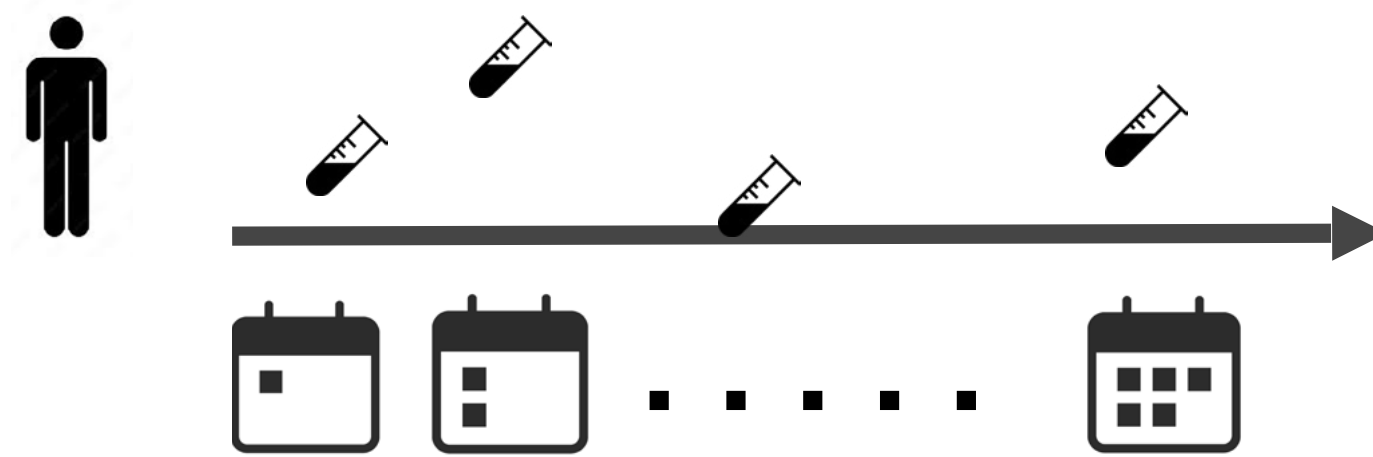
[www.sanita2030.it](http://www.sanita2030.it)





## serie temporali

Vincoli aggiuntivi sui dati reali rendono la generazione dei dati sintetici più complessi: e.g. i dati (clinici) longitudinali

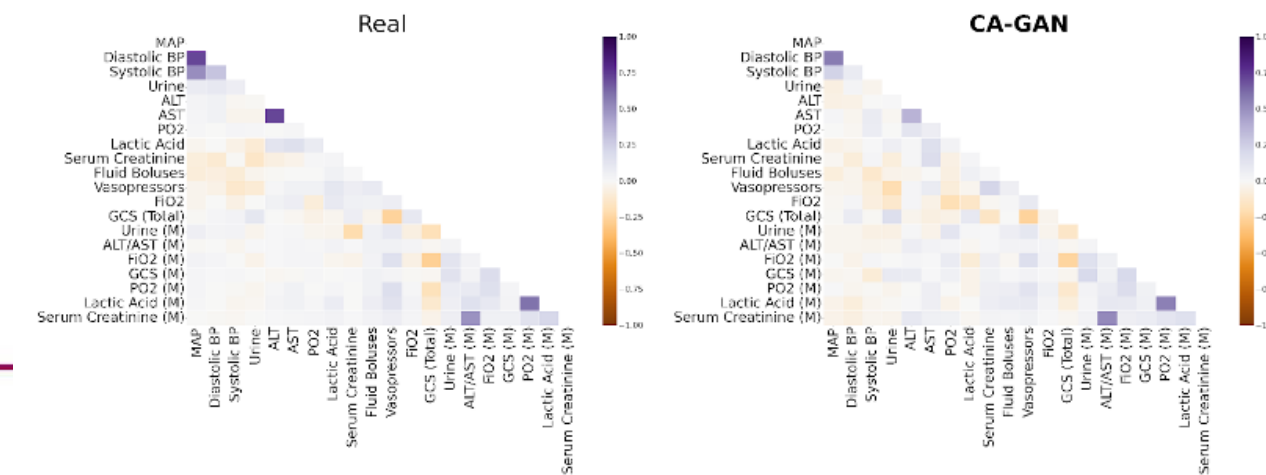


migliore conformità dei dati sintetici alla distribuzione reale rispetto a metodi di generazione alternativi



## GAN per EHR longitudinali

conservazione della struttura di correlazione dei dati reali



#sanita2030



## integrazione multiomica

IEEE Int C Bioinform, 2022

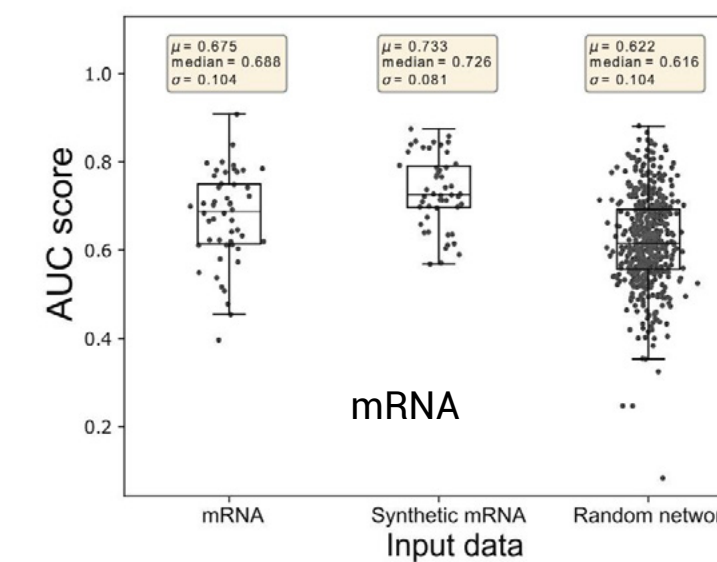
PROBLEM SOLVING PROTOCOL

### OmiTrans: generative adversarial networks based omics-to-omics translation framework

Xiaoyu Zhang<sup>1</sup> and Yike Guo<sup>1,2</sup>

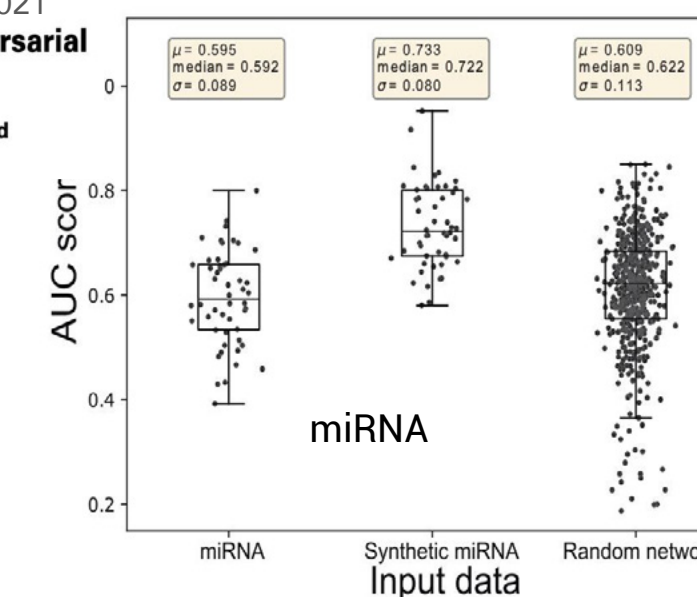
<sup>1</sup>Data Science Institute, Imperial College London, SW7 2AZ, London, UK and <sup>2</sup>Department of Computer Science, Hong Kong Baptist University, Hong Kong, China  
\*Corresponding author. x.zhang18@imperial.ac.uk

- omics2omics translation via gan
- reconstruct gene expression profiles from dna methylation data
- trained on a large-scale multiomics datasets & applied to small-scale datasets



Systems biology  
Bioinformatics, 2021  
**Multi-omics data integration by generative adversarial network**

Khandakar Tanvir Ahmed<sup>1,2</sup>, Jiao Sun<sup>1,2</sup>, Sze Cheng<sup>3</sup>, Jeongsik Yong<sup>3</sup> and Wei Zhang<sup>1,2,\*</sup>



#sanita2030

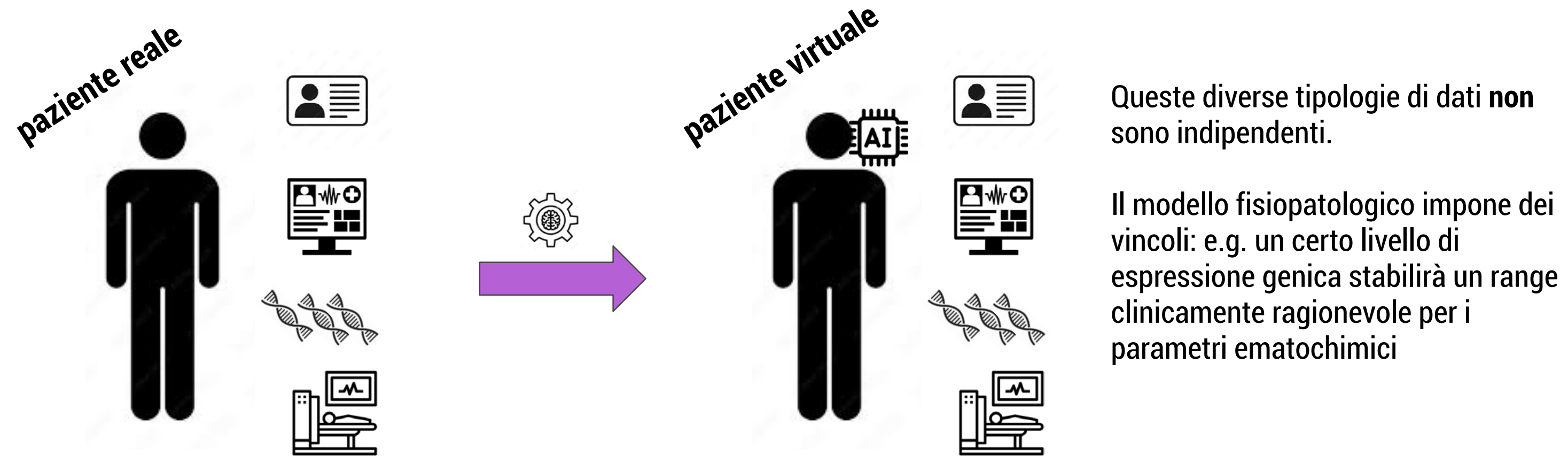


www.sanita2030.it



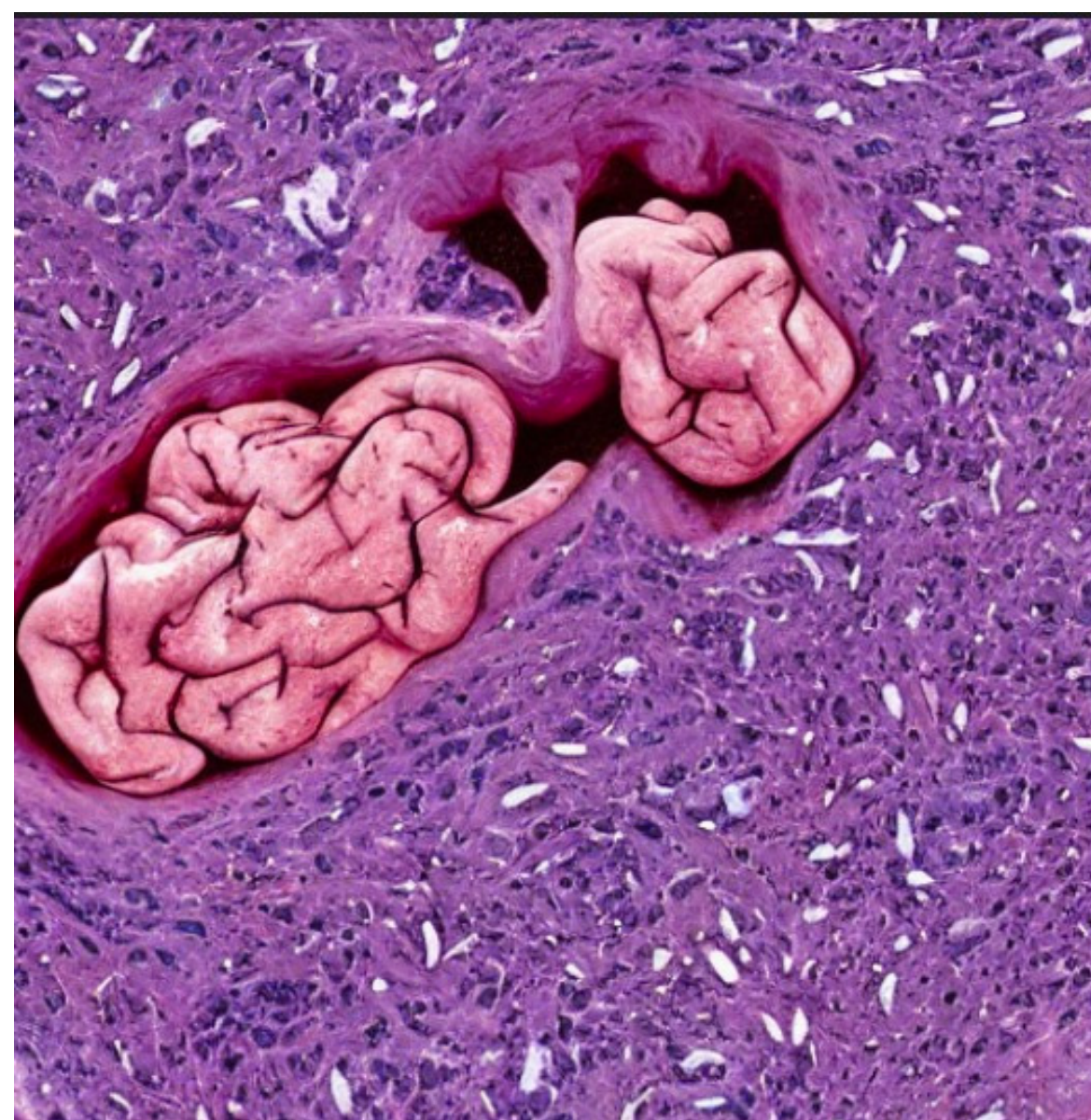


## la sfida



**Necessario introdurre la causalità (nonlineare) nella generazione dell'intero dataset di un paziente virtuale.**





「grazie!」



### **Delitti in materia di violazione del diritto d'autore (Art. 25-novies, D.Lgs. n. 231/2001) [articolo aggiunto dalla L. n. 99/2009]**

- Messa a disposizione del pubblico, in un sistema di reti telematiche, mediante connessioni di qualsiasi genere, di un'opera dell'ingegno protetta, o di parte di essa (art. 171, legge n.633/1941 comma 1 lett. a) bis)
- Reati di cui al punto precedente commessi su opere altrui non destinate alla pubblicazione qualora ne risulti offeso l'onore o la reputazione (art. 171, legge n.633/1941 comma 3)
- Abusiva duplicazione, per trarne profitto, di programmi per elaboratore; importazione, distribuzione, vendita o detenzione a scopo commerciale o imprenditoriale o concessione in locazione di programmi contenuti in supporti non contrassegnati dalla SIAE; predisposizione di mezzi per rimuovere o eludere i dispositivi di protezione di programmi per elaboratori (art. 171-bis legge n.633/1941 comma 1)
- Riproduzione, trasferimento su altro supporto, distribuzione, comunicazione, presentazione o dimostrazione in pubblico, del contenuto di una banca dati; estrazione o reimpiego della banca dati; distribuzione, vendita o concessione in locazione di banche di dati (art. 171-bis legge n.633/1941 comma 2)
- Abusiva duplicazione, riproduzione, trasmissione o diffusione in pubblico con qualsiasi procedimento, in tutto o in parte, di opere dell'ingegno destinate al circuito televisivo, cinematografico, della vendita o del noleggio di dischi, nastri o supporti analoghi o ogni altro supporto contenente fonogrammi o videogrammi di opere musicali, cinematografiche o audiovisive assimilate o sequenze di immagini in movimento; opere letterarie, drammatiche, scientifiche o didattiche, musicali o drammatico musicali, multimediali, anche se inserite in opere collettive o composite o banche dati; riproduzione, duplicazione, trasmissione o diffusione abusiva, vendita o commercio, cessione a qualsiasi titolo o importazione abusiva di oltre cinquanta copie o esemplari di opere tutelate dal diritto d'autore e da diritti connessi; immissione in un sistema di reti telematiche, mediante connessioni di qualsiasi genere, di un'opera dell'ingegno protetta dal diritto d'autore, o parte di essa (art. 171-ter legge n.633/1941)
- Mancata comunicazione alla SIAE dei dati di identificazione dei supporti non soggetti al contrassegno o falsa dichiarazione (art. 171-septies legge n.633/1941)
- Fraudolenta produzione, vendita, importazione, promozione, installazione, modifica, utilizzo per uso pubblico e privato di apparati o parti di apparati atti alla decodificazione di trasmissioni audiovisive ad accesso condizionato effettuate via etere, via satellite, via cavo, in forma sia analogica sia digitale (art. 171-octies legge n.633/1941).

**[Torna all'inizio](#)**