

WEBINAR



MARCH 9, 2023 | 10 AM - 11 AM EDT
**SYNTHETIC DATA USE: EXPLORING USE
CASES TO OPTIMISE DATA UTILITY**



PRESENTER

STEF JAMES, SENIOR DIRECTOR,
CAPABILITY AND OPERATIONS, **ASTRAZENECA UK**

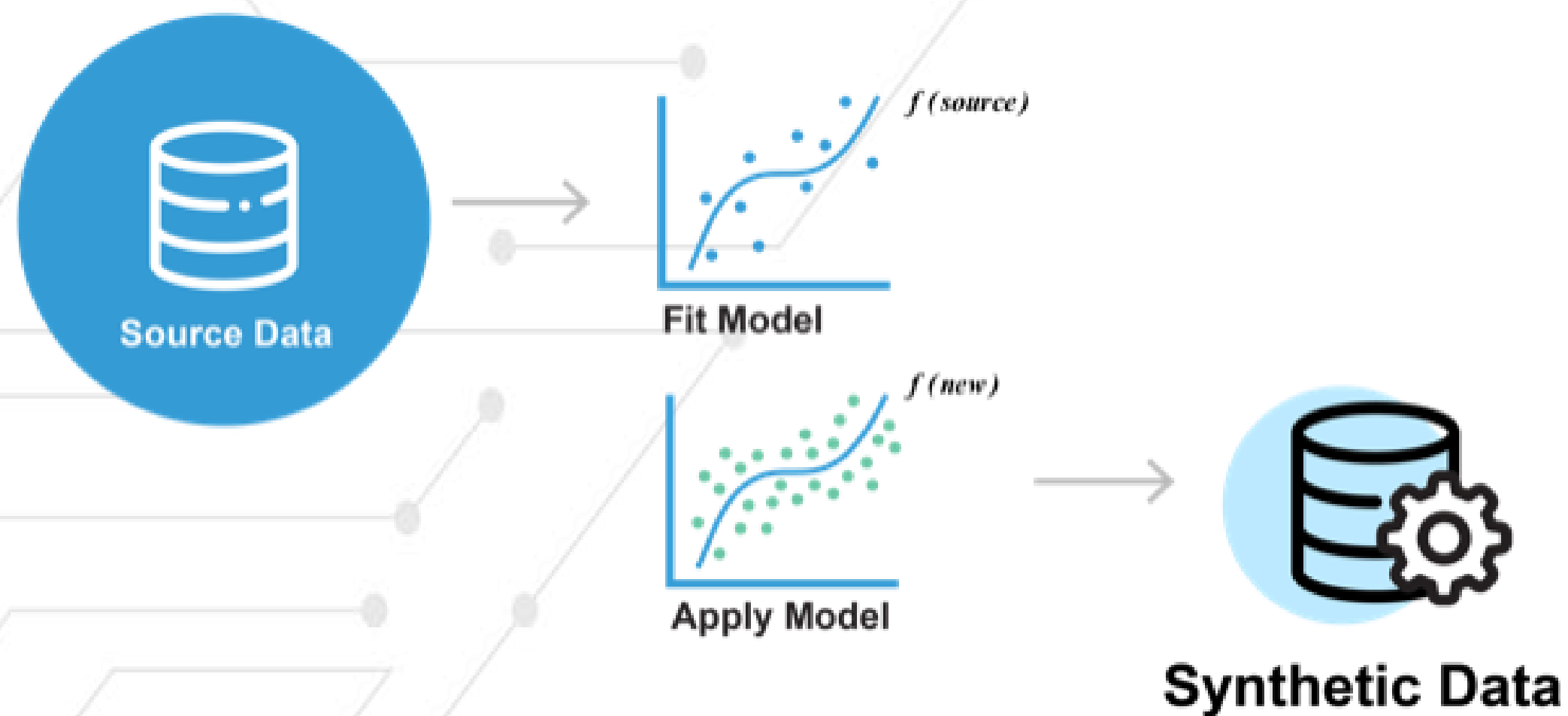
INTRODUCTION BY KHALED EL EMAM, UNIVERSITY OF OTTAWA



Methods for Synthetic Data Generation

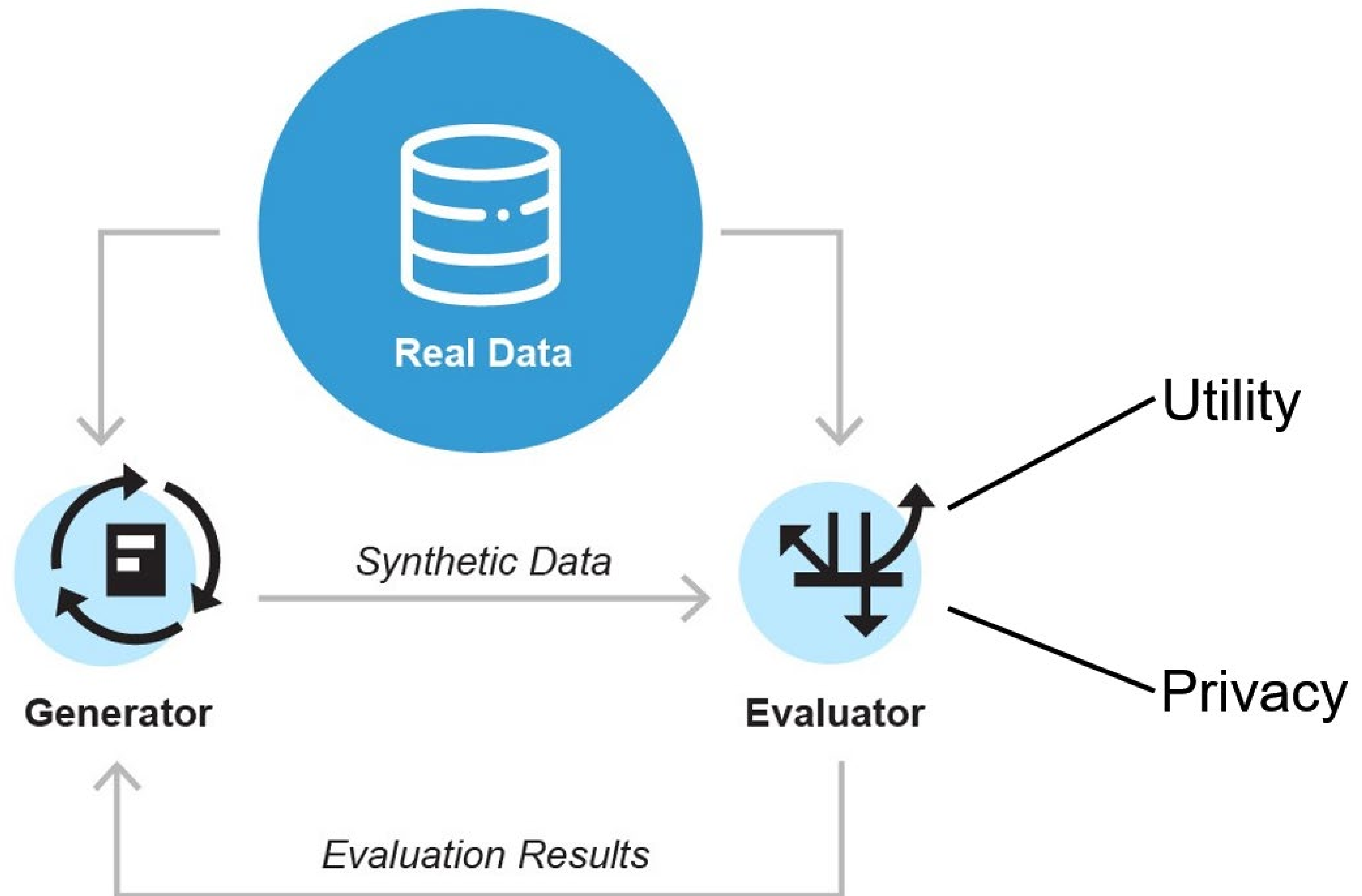
Khaled El Emam
kelemam@ehealthinformation.ca

Training and Generation

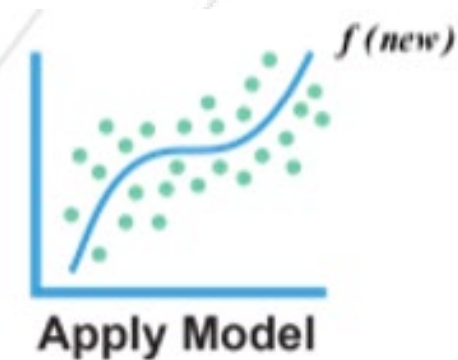
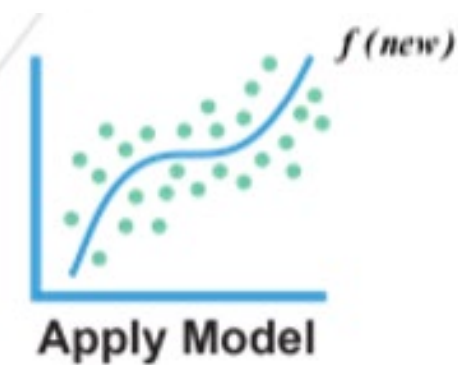
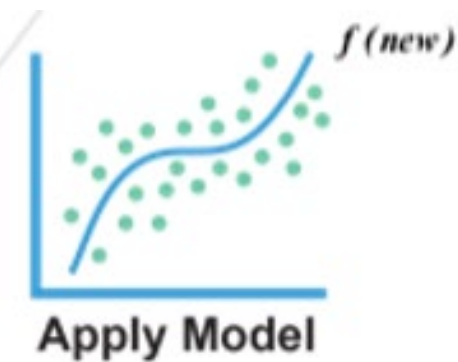


COU1A	AGECAT	AGELE70	WHITE	MALE	BMI
United States	2	1	1	1	33.75155
United States	2	1	1	0	39.24707
United States	1	1	1	0	26.5625
United States	4	1	1	1	40.58273
United States	5	0	0	1	24.42046
United States	5	0	1	0	19.07124
United States	3	1	1	1	26.04938
United States	4	1	1	1	25.46939

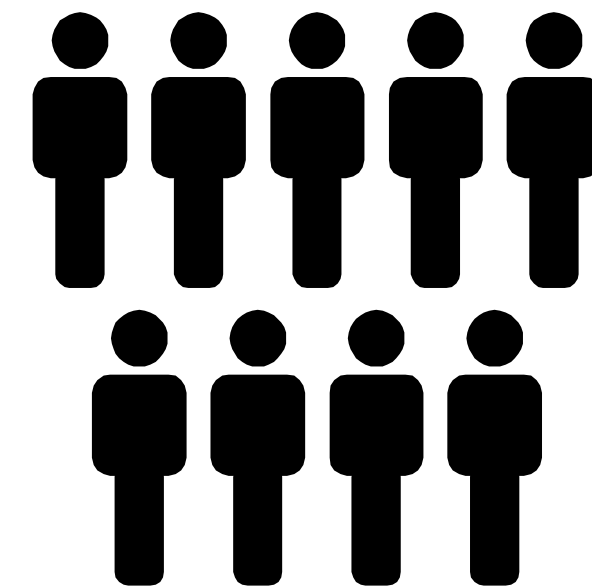
A combined loss of utility and privacy



A simulator catalogue allows data to be made available without sharing actual data



Data Consumers



SDG References

- K. El Emam, L. Mosquera, and X. Fang, “Validating A Membership Disclosure Metric For Synthetic Health Data,” *JAMIA Open*, vol. 5, no. 4, p. ooac083, Dec. 2022.
- Y. Jiang, L. Mosquera, B. Jiang, L. Kong, and K. El Emam, “Measuring re-identification risk using a synthetic estimator to enable data sharing,” *PLoS ONE*, vol. 17, no. 6, p. e0269097, Jun. 2022.
- K. El Emam, L. Mosquera, X. Fang, and A. El-Hussuna, “Utility Metrics for Evaluating Synthetic Health Data Generation Methods: Validation Study,” *JMIR Medical Informatics*, vol. 10, no. 4, p. e35734, Apr. 2022.
- Z. Azizi, C. Zheng, L. Mosquera, L. Pilote, K. El Emam: “Can synthetic data be a proxy for real clinical trial data? A validation study”, *BMJ Open*, 11:e043497, 2021.
- K. El Emam, L. Mosquera, E. Jonker, H. Sood: “Evaluating the Utility of Synthetic COVID-19 Case Data”, *JAMIA Open*, 14(1):ooab012, January 2021.
- K. El Emam, L. Mosquera, and C. Zheng, “Optimizing the synthesis of clinical trial data using sequential trees,” *JAMIA*, 28(1): 3-13, 2021.
- K. El Emam, L. Mosquera, and J. Bass, “Evaluating Identity Disclosure Risk in Fully Synthetic Health Data: Model Development and Validation,” *JMIR*, vol. 22, no. 11, Nov. 2020.
- K. El Emam, L. Mosquera, and R. Hoptroff, *Practical Synthetic Data Generation: Balancing Privacy and the Broad Availability of Data*. O'Reilly, 2020.
- K. El Emam, “Seven Ways to Evaluate the Utility of Synthetic Data,” *IEEE Security and Privacy*, July/August, 2020.