



# 2020 International Association for Official Statistics Young Statistician Prize

Kenza Sallier and Claude Girard,  
Senior methodologists  
Statistics Canada



Presented virtually to the Office  
for National Statistics

June 26, 2020

Delivering insight through data for a better Canada

## Outline

- A little background about myself
- The Young Statistician Prize (YSP) competition
- The paper: overview and taking into account the competition's criteria
- How it feels to be the first prize winner?
- Perspective of a mentor
- What's next?

## A little background about myself

- Master degree in statistics – Université de Montréal
  - Focus on times series and data analysis
- Recruited as a mathematical statistician in 2017 – Statistics Canada (Ottawa)
  - **Centre on Access and Confidentiality** (2½ years)
  - Census Operations Section (7 months)



# The Young Statistician Prize competition

- **Goal:** “propose solutions to pressing methodological or strategic issues in the area of official statistics at the regional, national or international level.”
- **Format and criteria:** submit a paper (4,000 words maximum), selection by an international panel
  - Scientific and/or strategic merit;
  - Originality;
  - Applicability in the practice of statistical organizations;
  - Quality of the exposition.

Source: <https://officialstatistics.com/news-blog/2020-iaos-young-statisticians-prize>

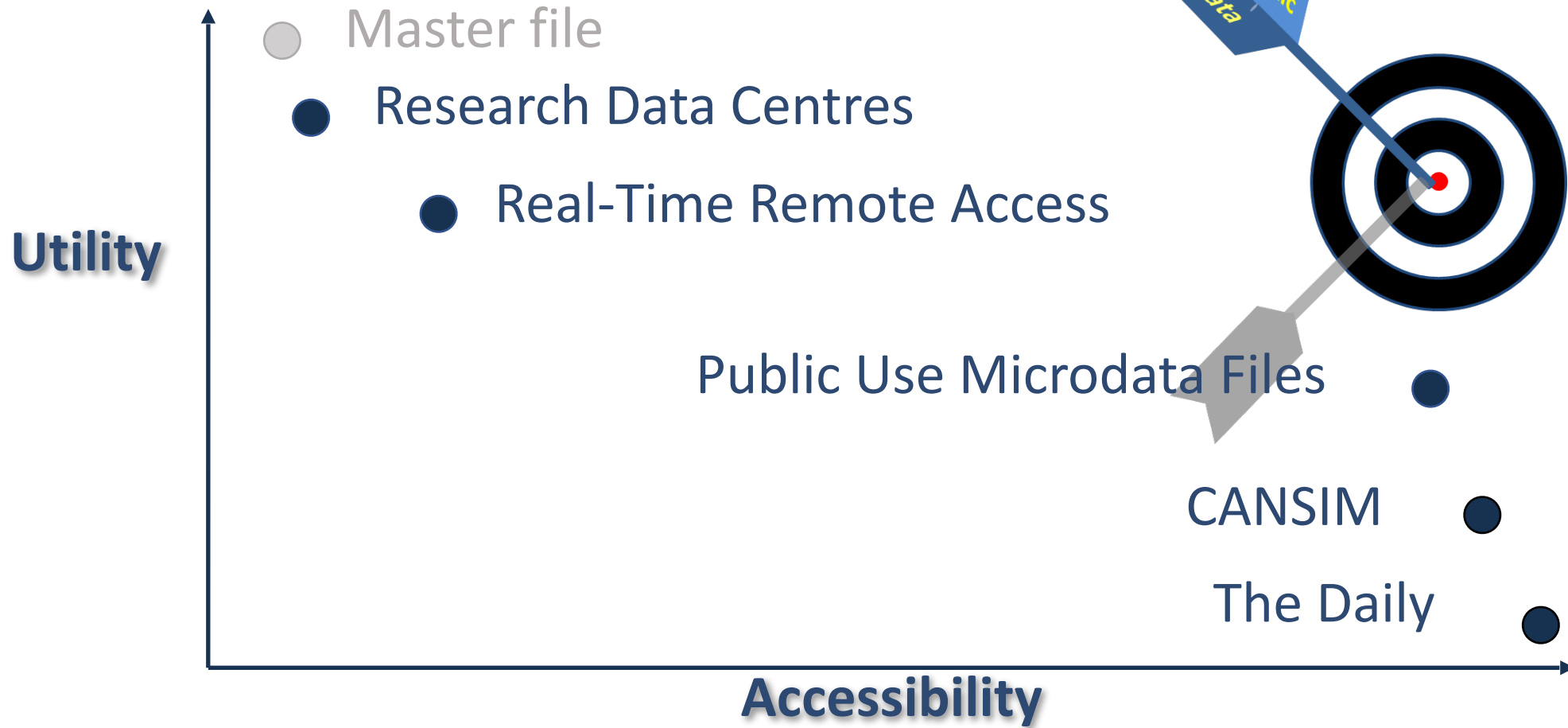


# The Young Statistician Prize competition – the paper

## Toward More User-Centric Data Access Solutions: Producing Synthetic Data of High Analytical Value by Data Synthesis

# Current Data Access Solutions at Statistics Canada

## Strategic merit and quality of the exposition



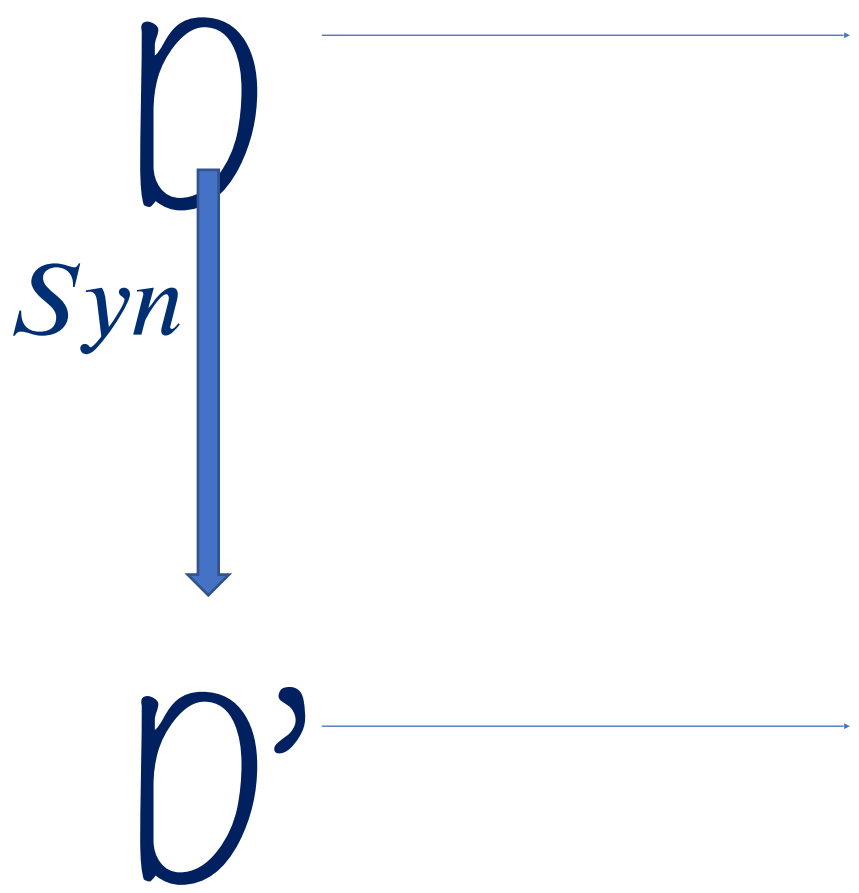
**The bigger picture**  
**Strategic merit and quality of the exposition**

- **Data Synthesis:** produce synthetic data of high analytical value that can be publicly released
- **Digital era and new data context:** governments and National Statistical Organizations need to adapt
- **Digital Agenda at the Government of Canada level and Modernization initiative at Statistics Canada**
- **User-centric and data driven:** we need to offer more flexible and useful data access solutions while keep protecting the confidentiality of respondents



# Data Synthesis: the process

## Scientific merit



$D$  the original dataset

$D'$  the synthetic dataset

*Syn* Process creating synthetic data

$\Theta$  Results of analyses

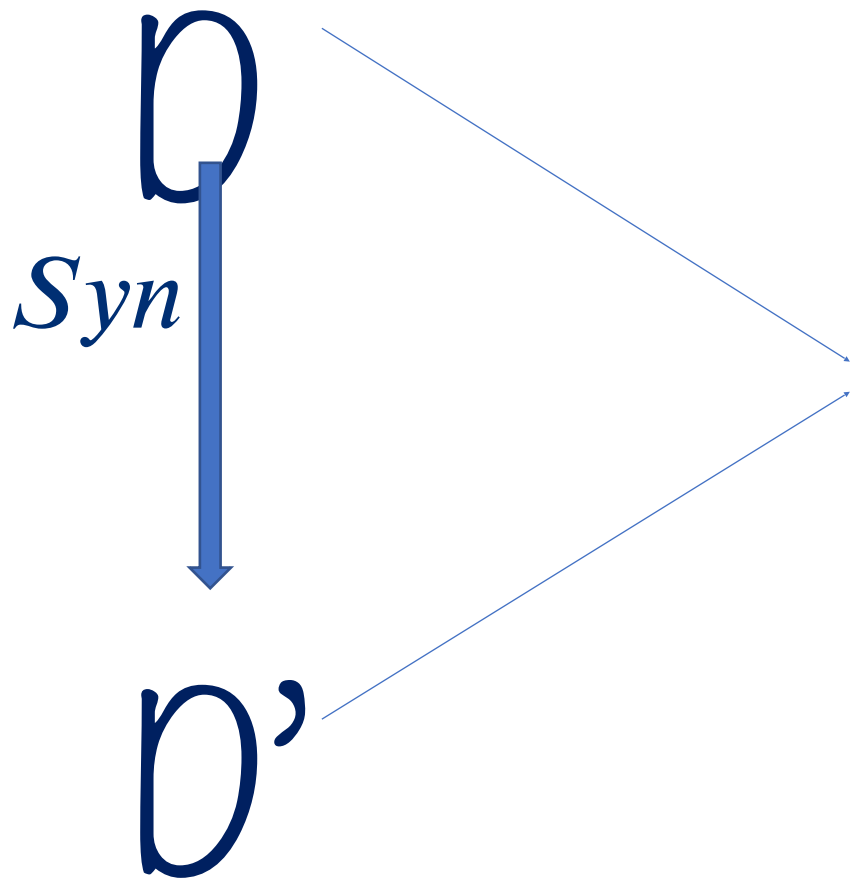
$U(D')$  the utility  
inverse function of  
the distance





# Data Synthesis: the process

## Scientific merit



$D$  the original dataset

$D'$  the synthetic dataset

*Syn* Process creating synthetic data

$\Theta$  Results of analyses

$M$  the maximum distance, maximum utility

# What underlies $\Theta$ ?

## Scientific merit

### I. Univariate statistics and tests

- ✓ Mean
- ✓ Variance
- ✓ Quantiles
- ✓ T-tests
- ✓ Kolmogorov-Smirnov tests
- ✓ Etc.

## The Joint Distribution

### II. Multivariate statistics and tests

- ✓ Covariance
- ✓ Correlation
- ✓ General linear models
  - ✓ Linear regression
  - ✓ Poisson regression
  - ✓ Logistic regression
  - ✓ Probit/Tobit
  - ✓ Etc.
- ✓ Survival analysis
- ✓ Etc.

# Data Synthesis: the process

Scientific merit

Joint  
Distribution

Respondents

$D$

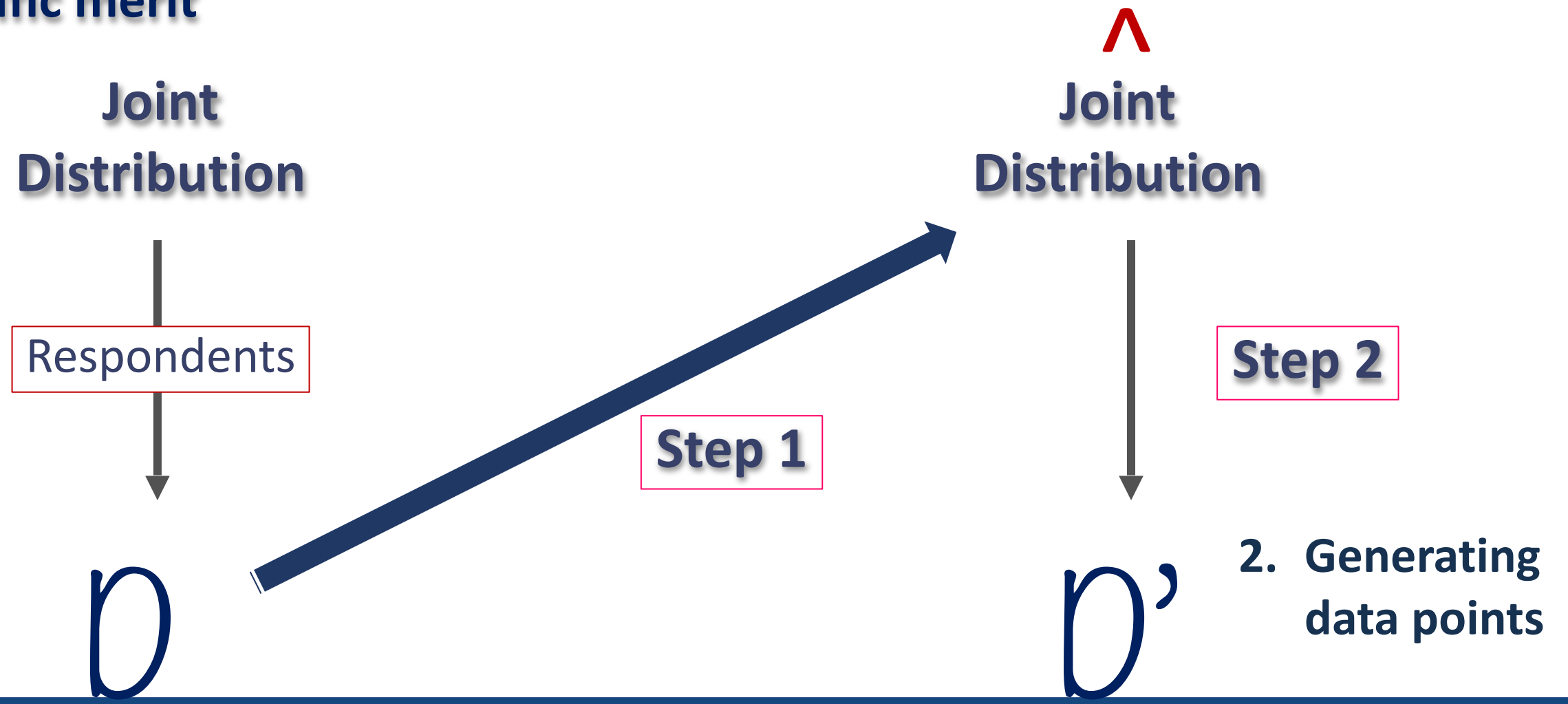
$\hat{\Delta}$   
Joint  
Distribution

Step 1

1. Careful modeling to capture the joint distribution of the original data.

# Data Synthesis: the process

## Scientific merit



## Real-use cases and implementation

### Applicability in the practice of statistical organizations

- Statistics Canada has **publicly released** synthetic versions of two different datasets related to census, mortality and cancer information (2018 and 2019). Produced for hackathons
- Same statistical conclusions as the ones found in the original data
- In both cases, synthetic data were generated using the **R package synthpop**
- One dataset currently used by the Canadian School of Public Service



## Real-use cases and implementation

### Applicability in the practice of statistical organizations

- Data Synthesis has limitations : “Analysis of synthetic data will never yield exactly the same results as would be found from the observed data.” (Raab et al.)
- Example: while waiting to be granted access to the real data, a researcher could explore a synthetic version of it.
- Data Synthesis is a proof of concept of Modernization

## How it feels to be the first prize winner?

- Surprise, joy and pride 😊
- Ideas of young statisticians can be heard and recognized!!
- Motivation to take initiative
- Rethink projects in a much broader, general way: How does this help the agency as a whole, and therefore the society?

## How it feels to be the first prize winner?

- The whole Agency was proud and happy
- Lessons learned:
  - Try new things and think outside the box
  - Importance of mentorship: “The quality of young statisticians has a lot to do with the quality of their supervisors and the support given to them during their career.” – John Pullinger



## Perspective of a mentor

- Producing and releasing 'smart' synthetic datasets outside of StatCan  
→ Breaking new ground (new techniques and tools, and no priors)
- Despite talented, enthusiastic and resolute Kenza on board, navigating uncharted seas is challenging: without direction → nowhere fast!
- Plan: Use my experience as compass, channeling in one direction all of Kenza's skills and energy to get us through safe and sound
- Boldly think ~~big~~ small: too tall an order can chew resources (and confidence) big time!  
→ Manage data processing (# variables, # records) within feasible time

## Perspective of a mentor

- Rapport: Moving in the one and same direction ≠ Being unidirectional
  - Assert my experience while keeping an open mind
  - Influence, but not intrude
  - Insist, and yet not impose
  - In the face of adversity show confidence, not undermine confidence
- Journey with Kenza (→ successful release and she winning IAOS Prize) brought me:
  - Immense satisfaction and pride in the breakthroughs we achieved together
  - Amazement as to what young talents can manage to reap from what was sown
  - Further validation as mentor



## What's next?

- Attend the next IAOS conference to present the article and receive the award
- Publication and interview in the Statistical Journal of the IAOS
- United Nation Economic Commission for Europe (UNECE) working group on Synthetic Data
- Encourage young statisticians to participate in the competition



100

**THANK YOU!**  
**MERCI !**

---

**Questions?**

[kenza.sallier@canada.ca](mailto:kenza.sallier@canada.ca)

[claudio.girard@canada.ca](mailto:claudio.girard@canada.ca)



Statistics  
Canada

Statistique  
Canada

Delivering insight through data for a better Canada

Canada