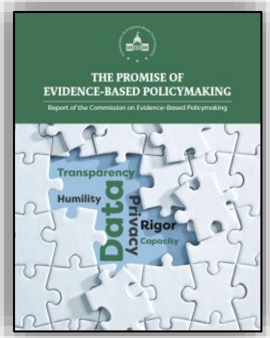
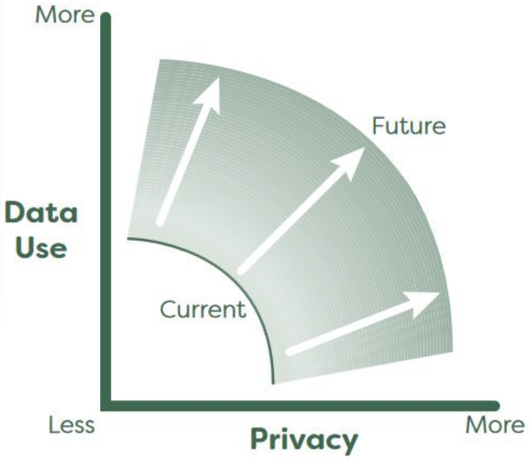
CHALLENGES TO PRIVACY AND CONFIDENTIALITY ACROSS THE BLENDED DATA LIFECYCLE

Mayank Varia
Boston University

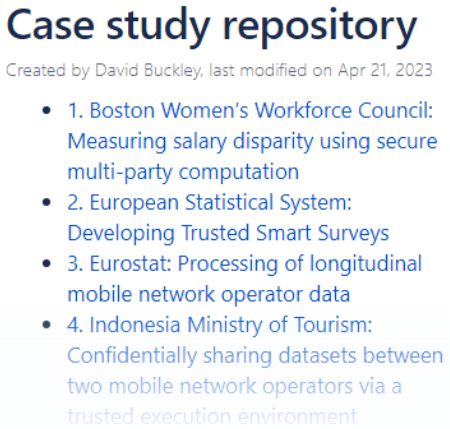
May 22, 2023

WHY USE PRIVACY TECH?


THE PROMISE OF
EVIDENCE-BASED POLICYMAKING
Report of the Commission on Evidence-Based Policymaking
Transparency
Humility
Data
Privacy
Rigor
Capacity



THE PET GUIDE
THE UNITED NATIONS GUIDE ON PRIVACY-ENHANCING TECHNOLOGIES FOR OFFICIAL STATISTICS
2023
United Nations BigData



Case study repository

Created by David Buckley, last modified on Apr 21, 2023

- 1. Boston Women's Workforce Council: Measuring salary disparity using secure multi-party computation
- 2. European Statistical System: Developing Trusted Smart Surveys
- 3. Eurostat: Processing of longitudinal mobile network operator data
- 4. Indonesia Ministry of Tourism: Confidentially sharing datasets between two mobile network operators via a trusted execution environment

"Privacy, when implemented correctly, fosters more information sharing, not less."

– Marc Groman

"The benefits of using data for official statistics can be realized while minimizing privacy risks to those entrusting sensitive data to National Statistics Offices."

Image source: The Promise of Evidence-Based Policymaking

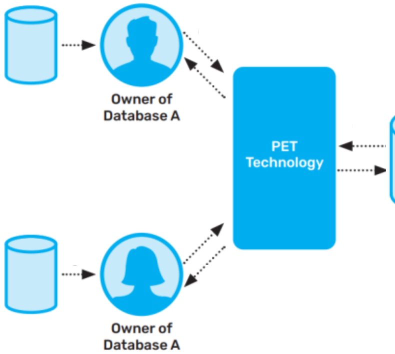
Image source: UN PET Guide

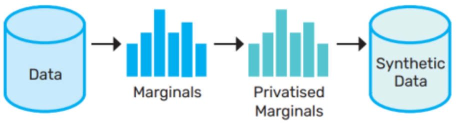
INPUT PRIVACY

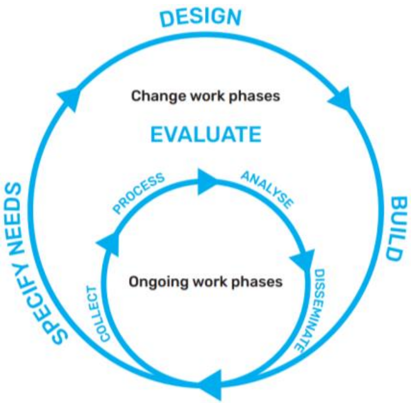
OUTPUT PRIVACY

“Allow two or more parties to submit data into a calculation without the other respective parties seeing data in clear.”

“Generally known as statistical disclosure control, [it] aims to conceal sensitive individual data from being identified or re-identified from the disseminated output.”







<u>Selected examples:</u>

- Secure multi-party computation
- Homomorphic encryption
- Zero knowledge proofs

<u>Examples:</u>

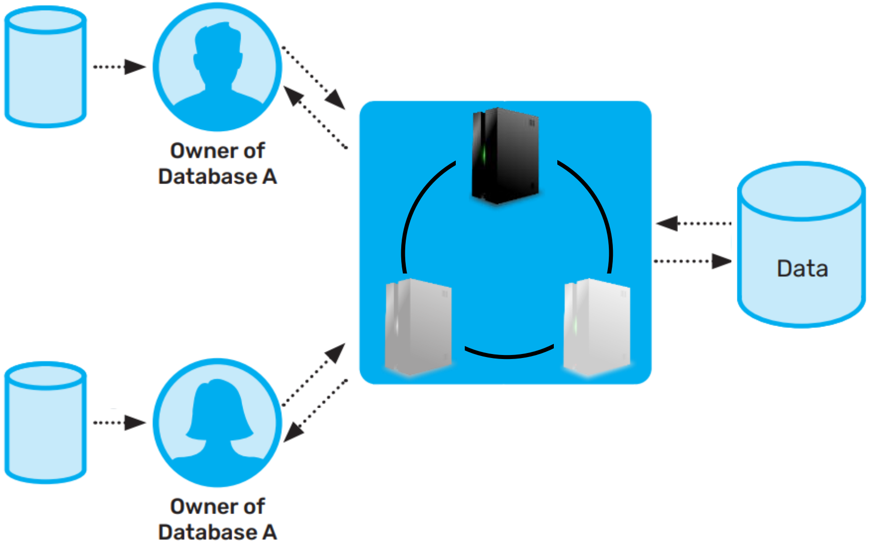
- Differential privacy
- Synthetic data generation

Image source: <u>UN PET Guide</u>

SECURE MULTI-PARTY COMPUTATION (SMC)

"Enables different participating entities in possession of private sets of data to link and aggregate their data sets … ***without transferring or otherwise revealing any private data*** to each other or anyone else."

– 2022 U.S. Senate bill S.3952



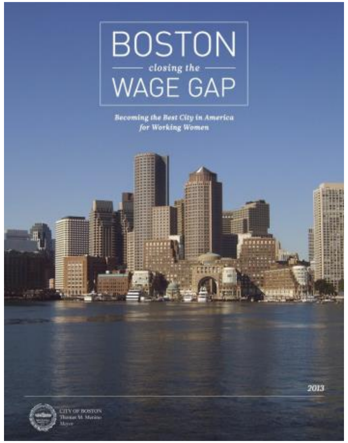
SMC DEPLOYMENTS IN THIS TALK

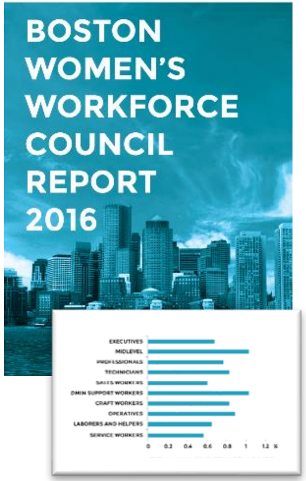
1. Boston Women's Workforce Council (non-blended data)
2. Massachusetts child advocacy study (blended data)

BWWC
BOSTON WOMEN'S WORKFORCE COUNCIL

OCA
MASSACHUSETTS
Office of the Child Advocate

BOSTON WAGE GAP STUDY



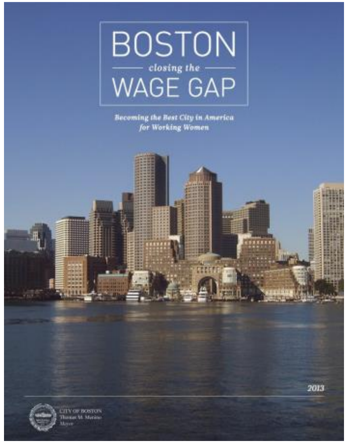


BOSTON WOMEN'S WORKFORCE COUNCIL REPORT 2017
■ ANNUAL COMPENSATION
■ CASH PERFORMANCE PAY
$108,661
$76,786
Men
Women

BOSTON WOMEN'S WORKFORCE COUNCIL REPORT 2019
In 2018, women in Boston earned*
70 cents
for every man's dollar.
Based on the 2019 BWWC Report
BW
EQUAL PAY. EQUAL OPPORTUNITY. 100% TALENT.

Lesson: Deploying SMC is a policy decision, promote from the top down

DATA CONTRIBUTORS



BOSTON
closing the
WAGE GAP

Becoming the Best City in America for Working Women

2013

CITY OF BOSTON
Thomas M. Menino
Mayor

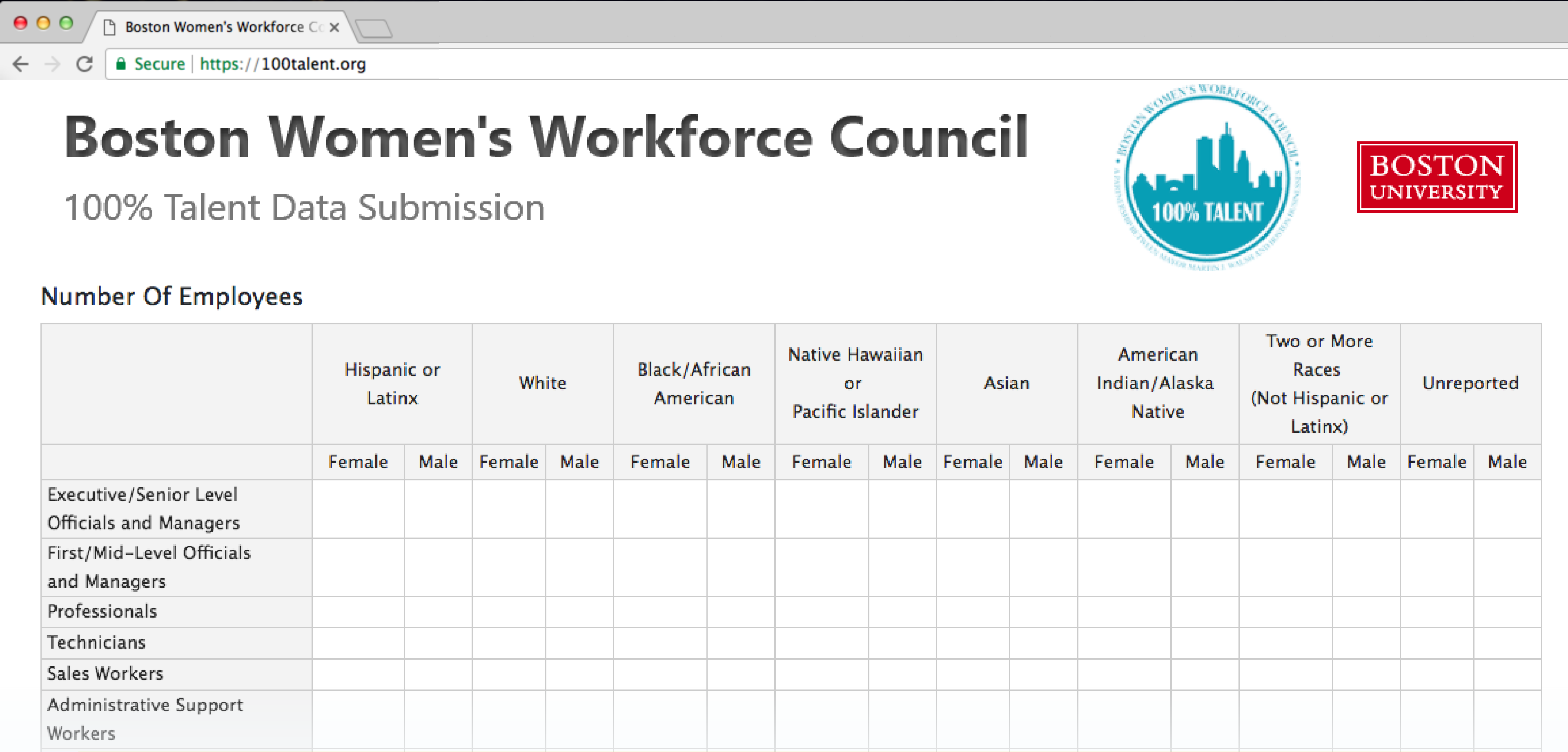
“Employers agree to anonymously contribute their wage data with the Council, which then creates a snapshot of what the wage gap looks like in our city.”

100% TALENT

The Boston Women’s Compact

SIMMONS COLLEGE BOSTON · MASSACHUSETTS
CITY OF BOSTON Office of the Mayor Martin J. Walsh
STATE STREET
EMC²
build smart
Raytheon
VERTEX
MassMutual FINANCIAL GROUP
SUFFOLK
STAPLES MAKE more HAPPEN
aim Associated Industries of Massachusetts
Putnam INVESTMENTS
Care.com
nationalgrid
MASSACHUSETTS TECHNOLOGY COLLABORATIVE
Eastern Bank
FINAGLE A BAGEL
EVERSOURCE ENERGY
Abt ASSOCIATES BOLD THINKERS DRIVING REAL-WORLD IMPACT
BBK WORLDWIDE
HILL HOLLIDAY
BJ's
MASSART MASSACHUSETTS COLLEGE OF ART AND DESIGN
WENTWORTH Institute of Technology
PARTNERS. HEALTHCARE FOUNDED BY BRIGHAM AND WOMEN'S HOSPITAL AND MASSACHUSETTS GENERAL HOSPITAL
Charlestown nursery school
Tech Networks of Boston We're better together.
TUFTS Health Plan
WGA WILLIAM GALLAGHER ASSOCIATES
WHEELOCK COLLEGE
tBf The Boston Foundation INNOVATION. INFORMATION. IMPACT.
Top it off
BOSTON MEDICAL CENTER EXCEPTIONAL CARE. WITHOUT EXCEPTION.
weber shandwick engaging, always.
BOSTON FINANCIAL DATA SERVICES
DANCING DEER BAKING CO.
the Y
Boston Children's Hospital Until every child is well
ROXBURY TECHNOLOGY
{Fiksu}
LANTERN FINANCIAL LLC
THE COMMONWEALTH INSTITUTE
BENTLEY UNIVERSITY
girl scouts of eastern massachusetts
JLL
MASSACHUSETTS
MASSACHUSETTS WOMEN'S POLITICAL CAUCUS MWPC
GREEN CITY GROWERS
CHARLES GROUP CONSULTING

Lesson: Easier to deploy SMC when all data contributors have similar interest in the output, and are similarly concerned about input privacy



Boston Women's Workforce Co
Secure | https://100talent.org
BOSTON WOMEN'S WORKFORCE COUNCIL
100% TALENT
BOSTON UNIVERSITY

Boston Women's Workforce Council

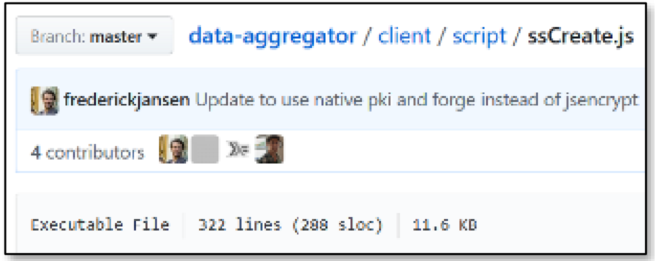
100% Talent Data Submission

Number Of Employees

	Hispanic or Latinx		White		Black/African American		Native Hawaiian or Pacific Islander		Asian		American Indian/Alaska Native		Two or More Races (Not Hispanic or Latinx)		Unreported	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Executive/Senior Level Officials and Managers																
First/Mid-Level Officials and Managers																
Professionals																
Technicians																
Sales Workers																
Administrative Support Workers																
Cra																
Ope																
Lab																
Ser																

Lesson: Use standardized data formats & easily accessible applications

SIMPLICITY & TRANSPARENCY DRIVE ADOPTION

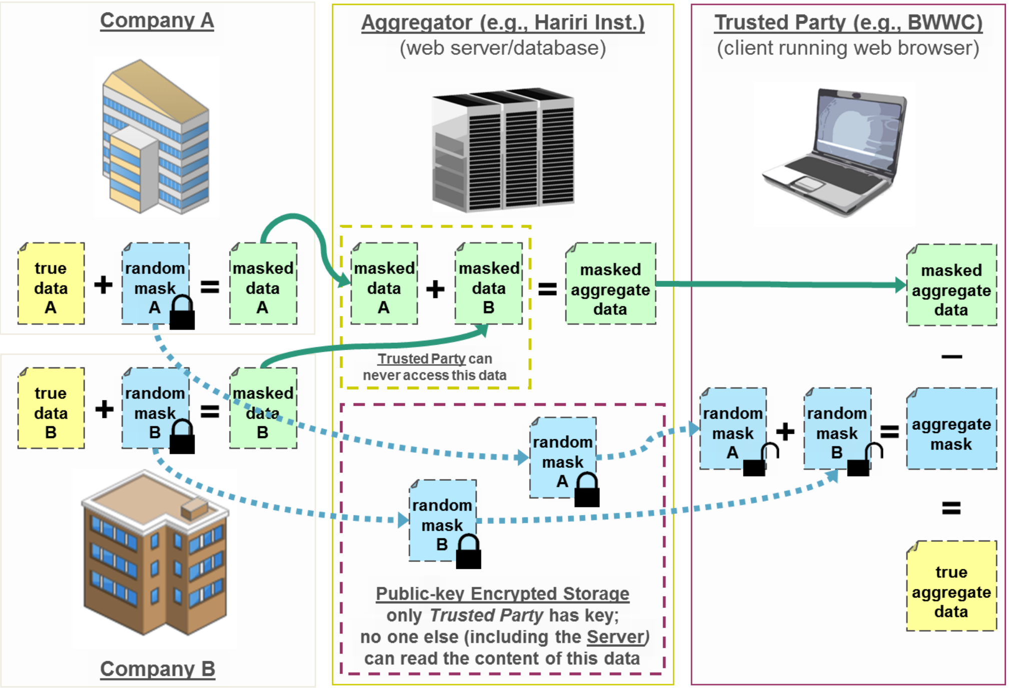


Branch: master ▾ data-aggregator / client / script / ssCreate.js

frederickjansen Update to use native pki and forge instead of jsencrypt

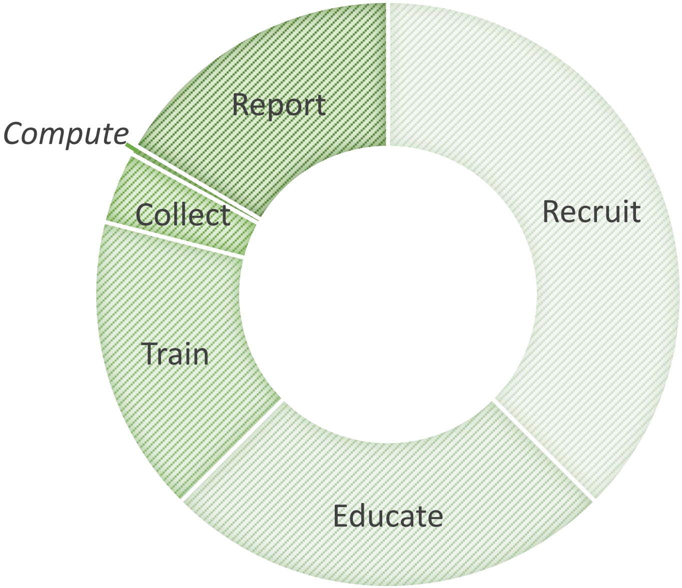
4 contributors

```
Executable File | 322 lines (288 sloc) | 11.6 KB
```



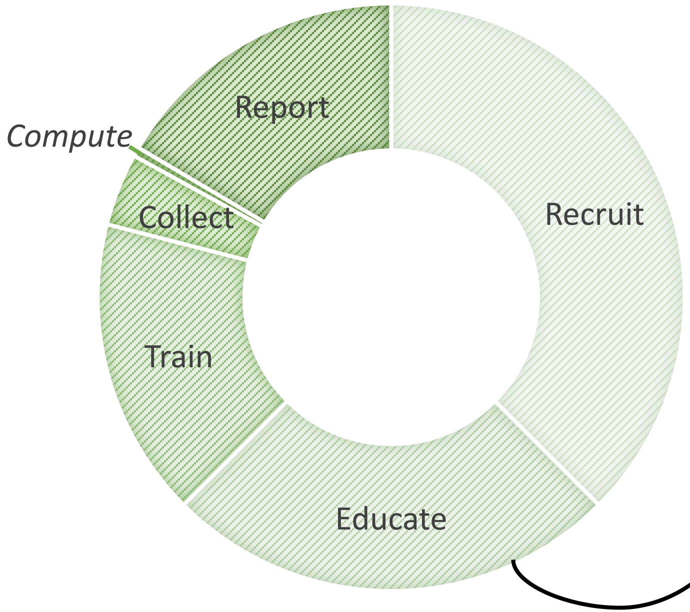
Lesson: SMC is simple to explain, and doing so improves trust + adoption

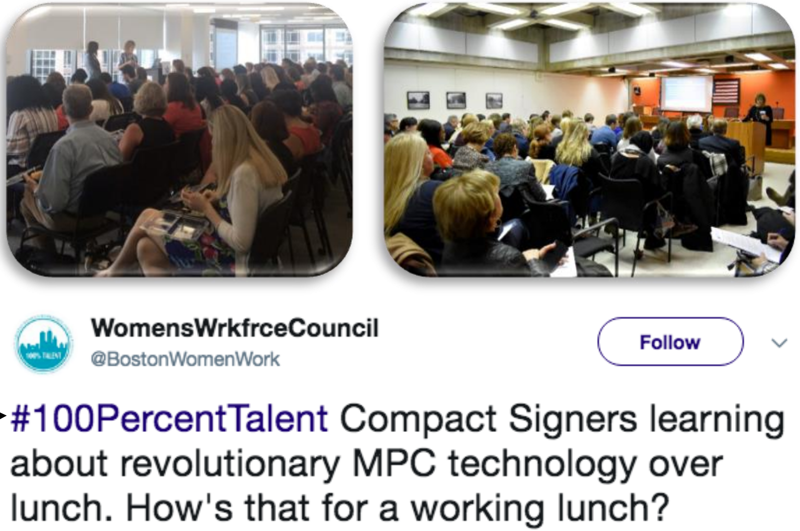
TIMELINE



Lesson: Time is spent on finding and helping people to work with data, not computer science metrics like computational efficiency

EDUCATION & TRAINING



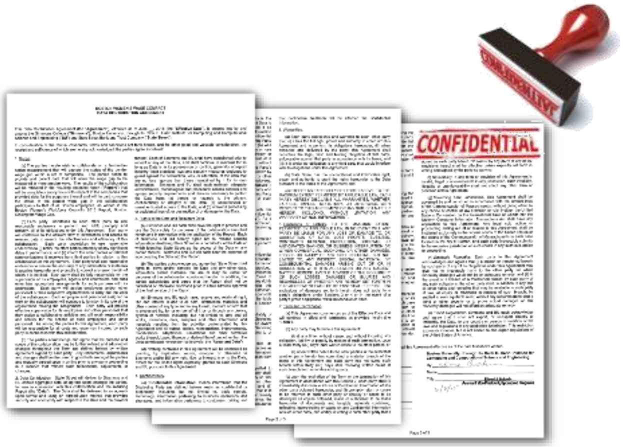


WomensWrkfrceCouncil
@BostonWomenWork

#100PercentTalent Compact Signers learning about revolutionary MPC technology over lunch. How's that for a working lunch?

Lesson: Identify key people to convince, together with domain experts

NDAs & LEGAL CONSIDERATIONS



Definitional question
Do encodings count as personal data?

Process question
Does computing constitute disclosure?

Liability question
Who should be blamed for an error?

Lesson: Consider the legal ramifications of data disclosure & processing

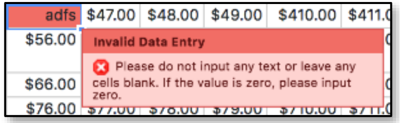
VALUE PROPOSITION

BWWC
Data quality

HR Personnel
Accessibility

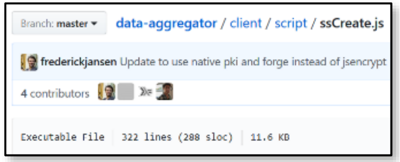
Lawyers
Liability

IT Personnel
Comprehension



adfs $47.00 $48.00 $49.00 $410.00 $411.0
$56.00
Invalid Data Entry
Please do not input any text or leave any cells blank. If the value is zero, please input zero.
$66.00
$76.00 $77.00 $78.00 $79.00 $710.00 $711.0

CONFIDENTIAL



Branch: master ▾ data-aggregator / client / script / ssCreate.js

frederickjansen Update to use native pki and forge instead of jsencrypt

4 contributors

```
Executable File | 322 lines (288 sloc) | 11.6 KB
```

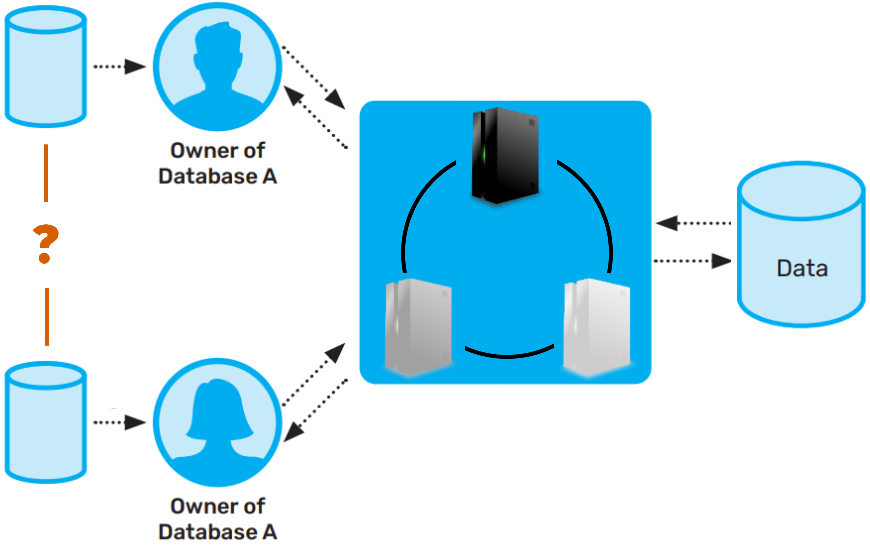
Lesson: Provide value to all stakeholders involved in the data analysis

CHILD ADVOCACY STUDY

Goal: make data-driven recommendations for reforming the juvenile justice system

Blend datasets containing data on:

- Vulnerable groups
- Educational success
- Judicial outcomes
- …



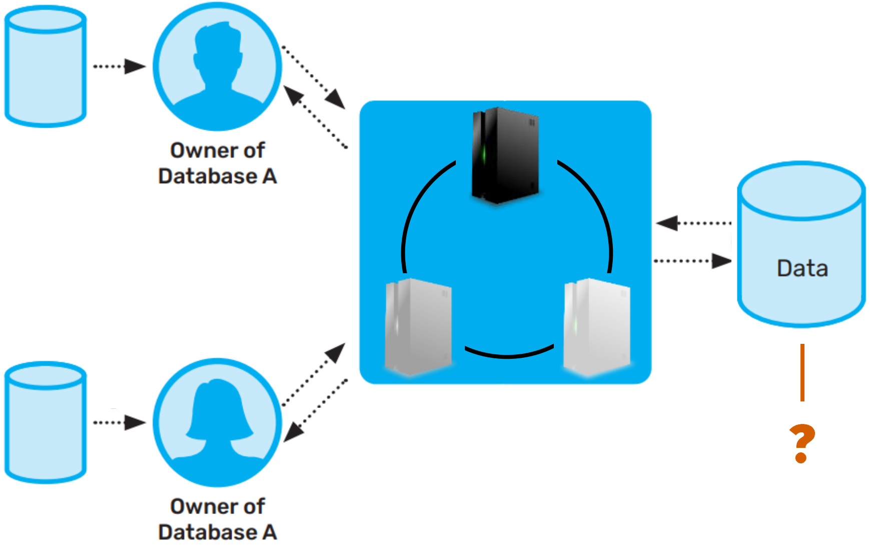
Lesson: Challenging to conduct any data analysis (with or without input privacy) in the absence of standardized data formats and common fields

CHILD ADVOCACY STUDY

Goal: make data-driven recommendations for reforming the juvenile justice system

Blend datasets containing data on:

- Vulnerable groups
- Educational success
- Judicial outcomes
- …



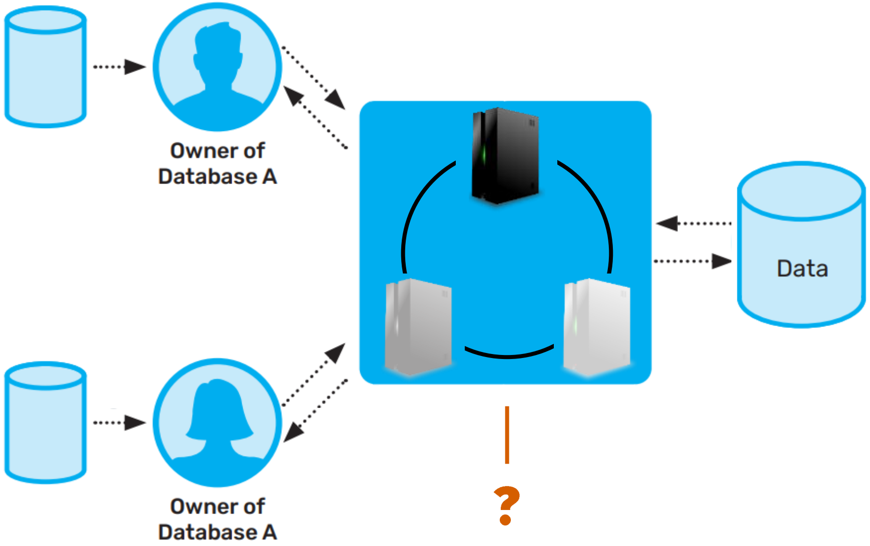
Lesson: Challenging to build consensus for data analyses when each potential contributor has its own policy goals and interest in the output

CHILD ADVOCACY STUDY

Goal: make data-driven recommendations for reforming the juvenile justice system

Blend datasets containing data on:

- Vulnerable groups
- Educational success
- Judicial outcomes
- …



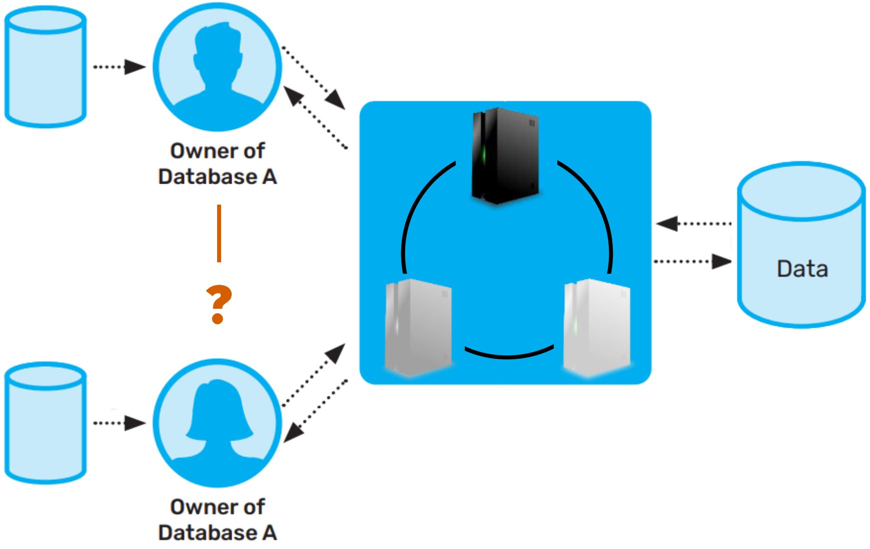
Lesson: Sometimes privacy questions are masking deeper concerns

CHILD ADVOCACY STUDY

Goal: make data-driven recommendations for reforming the juvenile justice system

Blend datasets containing data on:

- Vulnerable groups
- Educational success
- Judicial outcomes
- ...



Lesson: Ultimately, can only move forward with a data analysis if the result provides value to all stakeholders involved

SUMMARY: LESSONS LEARNED

WHO

- Promote from the top down
- Find orgs with interest in the result
- People, not CPU, dominate runtime

WHAT

- Use standardized data, schemas, apps
- Identify key people to convince
- Consider legal ramifications

HOW

- Provide value to all stakeholders
- Simple, transparent tech drives adoption
- Improve both data privacy and utility