

# Improving the Privacy and Security of Online Survey Data Collection, Storage, and Processing

---

- Student Researcher: Richard Wartell
- Professor: Mikhail Atallah
- Other Members:
  - Marina Blanton
  - Prof. Juline Mills

## Abstract

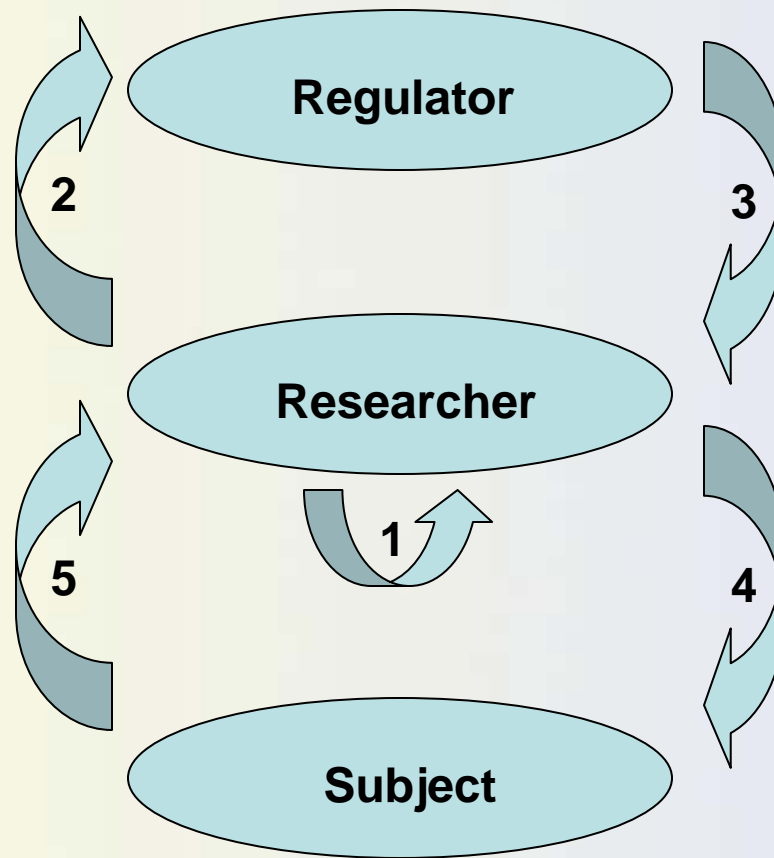
---

Important public policy conclusions are drawn from surveys, yet problems exist in the way such data is collected and handled. This results in low quality (biased) responses from participants. The reasons for biased responses include many (like social desirability bias, and acquiescence bias) that would be mitigated by a technology that hides the individual responses and reveals only the aggregate outcomes of the agreed-upon data analyses.

Our research attempts to implement such a technology, where individual responses are entirely unavailable once they are entered, but the aggregate outcomes can be retrieved.

# Current Design

---



# Current Design

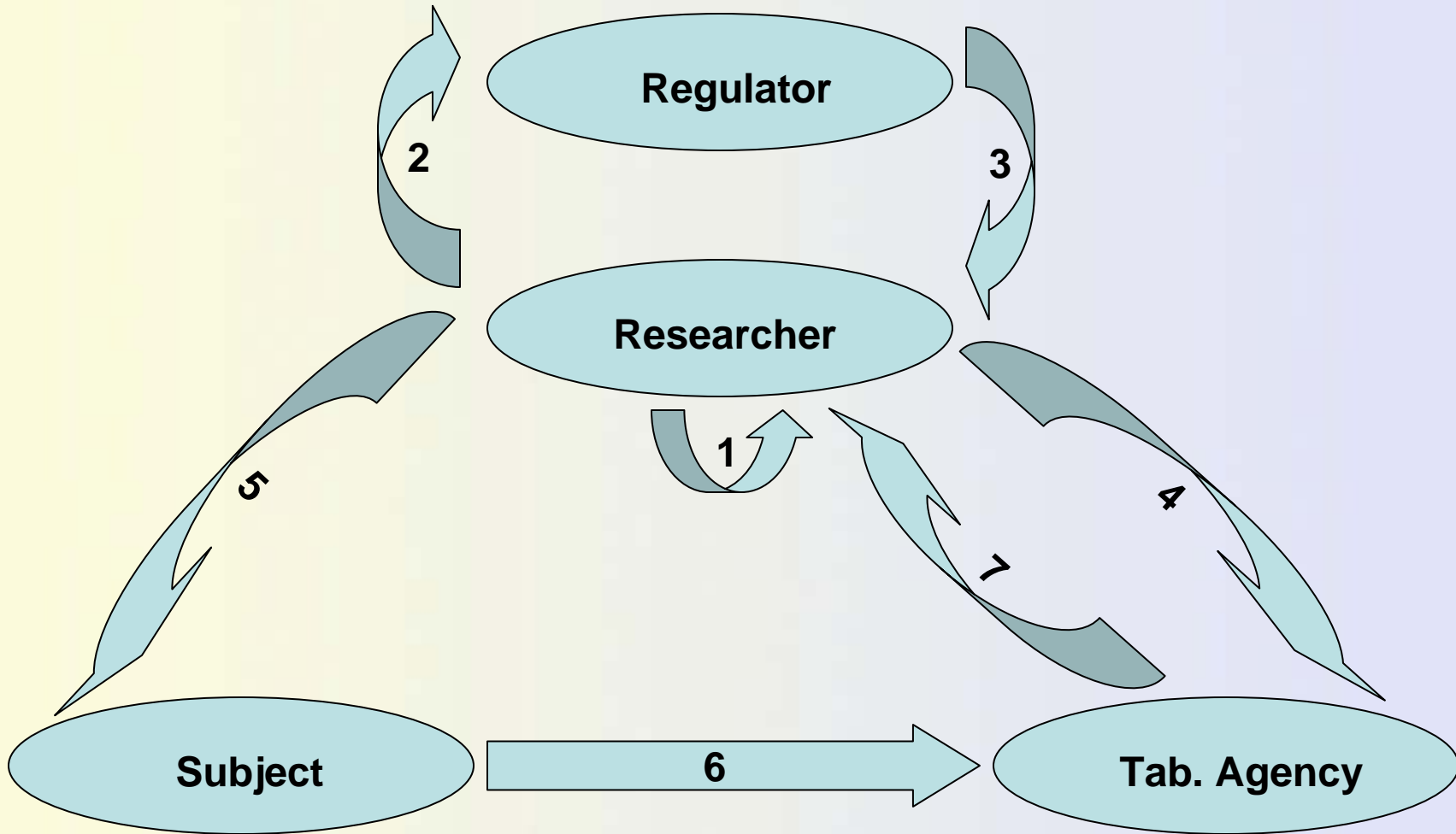
---

## Design Steps:

- 1) Researcher creates a survey.
- 2) Researcher submits survey to Regulator for approval.
- 3) Regulator approves survey.
- 4) Researcher gives survey to subjects to fill out.
- 5) Subjects return answers to researcher.

# Our Design

---



# Our Design

---

## Design Steps:

- 1) Researcher creates a survey.
- 2) Researcher submits survey to Regulator for approval.
- 3) Regulator approves survey.
- 4) Researcher creates user data and sends unencrypted data to the Tabulating Agency.
- 5) Researcher gives signature files to subjects.
- 6) Subject fills out survey and submits encrypted answers to the Tabulating Agency.
- 7) Tabulating agency uses homomorphic encryption to compile single and pair-wise counts of data and send them back to the researcher.

# Implementing Our Model

---

## Our model's ideal implementation:

- Fully online and streamlined process for survey creation, survey submission for approval, data generation for subjects, compiling the data, and viewing the results.
- Four separate web servers using Asp.net with C# code behind, one for each service of the design.
- Each server uses a MSSQL database to store data.
- All processes of our design can be done in real-time, including data retrieval and decryption by researcher.

# Implementation Status

---

## Our model's current status:

- Fully online and streamlined process for survey creation, survey submission for approval, data generation for subjects, compiling the data, and viewing the results is complete, though only being run locally.
- Four separate databases being stored locally on my machine, accessed by a single web site which is not currently online.
- Web server will go online today to enable online access to the project.
- Data can be retrieved by the researcher with short wait times for single counts, but pair-wise counts take time.

## Uses of This Research

---

- Testing may show that if subjects understand that their answers can never be connected to them, they will answer more truthfully than with traditional methods.
- Researchers may be less biased when approaching the data that they receive since they know nothing about who it comes from.
- When sensitive data is part of the survey, such as information that could potentially hurt someone's reputation or livelihood if released, this survey method might be a necessity in order to protect subjects from being damaged.
- Businesses may be interested in using this method of surveying since it protects them from being sued in the case of leaked information.

## Future Work

---

- Speed up encryption and decryption so as to push the project closer to running in real-time.
- Get the implementation online to ready it for testing.
- Implement different data analysis techniques (Outlier removal, Data sampling, t-tests, Chi square, Regression (simple, multiple), Correlations, ANOVA / MANOVA, Structural equation modeling, Factor analysis).
- Look into other surveys that may be done in order to determine how effective the design is.

Questions?

---

Q&A