

A large, stylized, light blue 'N' logo is positioned in the background on the left side of the slide. It has a thick, rounded, and slightly irregular font style.

**N<sup>th</sup>**  
party

**Navigating Clean Room and  
Privacy Enhancing Solutions  
for AdTech and MarTech**

# Contents

An informal framework for navigating and evaluating emerging clean room and privacy-enhancing solutions

- **Why a Clean Room?**
- **Three Ways to Build a Clean Room Solution**
- **Two Distinct Purposes of Privacy Enhancing Technologies**
- **Comparison of PETs**
- **Other Considerations and Further Reading**

# Ad-/MarTech Workflow

## Data Enrichment

- Use two variants of common **enrichment** workflow to **illustrate concepts**
- Same principles will apply to many other workflows: **onboarding, identity resolution, segmentation, lookalikes, attribution, real-time ad delivery, etc.**

### 1<sup>st</sup> Party Publisher or Intermediary

Email	Address
ann@example.com	123 A St. 1234
bob@example.com	456 B St. 12345
⋮	

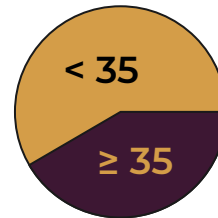
- Prefers to disclose **as little data as possible** to 2<sup>nd</sup> party
- Expects to receive **agreed-upon output** as part of transaction

### 2<sup>nd</sup> Party Identity Graph Vendor

Email	Age	Income
ann@example.com	48	> \$50K
bob@example.com	31	≤ \$50K
⋮		

- Prefers exposure to **as little data as possible** from 1<sup>st</sup> party
- Wants to disclose **only data that has business value** to 1<sup>st</sup> party

### Aggregate Output (for 1<sup>st</sup> Party)



### Row-Level Output (for 1<sup>st</sup> Party)

Email	Age
ann@example.com	48
bob@example.com	31
⋮	

# Clean Room Solutions

## What Kind of Implementation?

- Preferences/requirements presented on previous slide **can be fully satisfied** in multiple ways
- Solutions grouped by **extent of disclosure to a 3<sup>rd</sup> party**

### Trusted 3<sup>rd</sup> Party

SaaS platform and multi-sided data marketplace

Scalable, proprietary workflows enabled by platform

Secure data storage  
(fully accessible by 3<sup>rd</sup>  
party)

- Governance & access controls (primarily for protecting data from 1<sup>st</sup>/2<sup>nd</sup> party, not 3<sup>rd</sup> party)

### Untrusted 3<sup>rd</sup> Party

Secure data storage  
(limited or no access by 3<sup>rd</sup>  
party)

- Governance & access controls
- Non-cryptographic techniques
- Cryptographic techniques

### No 3<sup>rd</sup> Party

Full integration with 2<sup>nd</sup> party  
infrastructure

Support proprietary 2<sup>nd</sup> party  
workflows/capabilities

Secure joint computation  
(limited or no access by other  
party)

- ~~Governance & access controls~~
- Non-cryptographic techniques
- Cryptographic techniques

# Extent of Disclosure to 3<sup>rd</sup> Party

## What Informs Your Decision?

- One purpose of data protection technologies is to **limit disclosure of data to the (newly introduced) 3<sup>rd</sup> party**
- Your organization's goals w.r.t. this informs solution choice

- **Is it acceptable to disclose data to a 3<sup>rd</sup> party?**
  - If yes, trusted 3<sup>rd</sup> party may be sufficient
  - If not, costs of other two variants may be justified
- **What disclosure is acceptable when considering...**
  - Competitive advantage
  - Cybersecurity considerations
  - Regulatory compliance (current and future)
  - Internal policies
  - Legal liability
- **How does this interact with non-privacy solution/vendor features?**
  - Multi-sided marketplace/network participation
  - Scalability and cost
  - Standardization and stability
  - Proprietary capabilities offered by 3<sup>rd</sup> party

# Clean Room Solutions

## How/Where is Data Disclosed?

- Two disclosure points: to 3<sup>rd</sup> party and to intended recipient(s)
- Any individual technology might be used to limit disclosure to 3<sup>rd</sup> party, recipient(s), or both: **technology ≠ purpose in use case**

Not Disclosing  
Input and  
Intermediate Data  
(to 3<sup>rd</sup> or any party)

### Trusted 3<sup>rd</sup> Party

- Encrypted data (**with key**) stored by 3<sup>rd</sup> party
- Governance, access controls, & auditability

### Untrusted 3<sup>rd</sup> Party

- Encrypted data (**no key**) stored by 3<sup>rd</sup> party
- **Minimization & redaction**
- **De-identification**
- **Pseudonymization**
- **Noise introduction**

### No 3<sup>rd</sup> Party

- Encrypted data (**no key**) jointly processed by 1<sup>st</sup> and 2<sup>nd</sup> party

“Non-cryptographic”

- Traditional hashing & tokenization
- **Trusted Execution Environments (TEEs)**
- **Secure Multi-Party Computation (MPC)**
- **Homomorphic Encryption (HE)**

Disclosing as  
Little as Possible  
in Output Data  
(to 1<sup>st</sup> party)

- **Minimization & redaction**
- **De-identification**
- **Pseudonymization**
- **Noise introduction**

- K-anonymity
- Cohorts and clusters
- Data models
- **Differential Privacy (DP)**

“Cryptographic”

# Privacy-Enhancing Technologies

## Maturity, Characteristics, & Costs

- Brief comparison of a traditional technique and PETs
- Some dimensions to consider: **scalability/cost**, **maturity**, **longevity** ("future-proof"-ness), **accuracy/precision**

- Traditional **cryptographic hashing and tokenization**
  - Susceptible to guess-and-check, persistent, impossible to federate/convert
- Non-cryptographic techniques (**de-identification, anonymization, noise**)
  - **Highly scalable**: no significant costs compared to in-the-clear
  - **Not necessarily future-proof**: difficult to quantify privacy benefits
- Techniques intended for aggregate results (**cohorts, models, DP**)
  - **Highly scalable**: no significant costs compared to in-the-clear
  - **Not necessarily future-proof**: difficult to quantify privacy benefits
  - **Reduced Precision/accuracy** of results
  - **Sometimes (mis)used** as less costly substitute to cryptographic techniques...
- Cryptographic techniques
  - **TEEs**: Least costly, **most scalable** (appropriate for AI/ML); relies on hardware
  - **MPC**: Scalability depends on workflow; **future-proof**
  - **HE**: Least scalable; secure/general like MPC but **can run offline**; **future-proof**

# Example Scenario Variants

## Data Enrichment

- Depending on what kind of data disclosure is acceptable, different PET solutions can be employed
- Other features enabled/prevented by choices may be a factor, as well

**1.** Publisher and ID graph vendor send data to 3<sup>rd</sup> party; 3<sup>rd</sup> party delivers results to publisher

- Both sides trust 3<sup>rd</sup> party
- Publisher and ID graph owner pay for 3<sup>rd</sup> party platform use by volume of data
- Workflow runs very quickly

**3.** Publisher and ID graph vendor build **noisy data** via 3<sup>rd</sup> party; both receive the results

- 3<sup>rd</sup> party does not store raw data, **but briefly sees it** to build noisy model
- **Results have an error margin of 2-3%**
- Workflow takes ~1 minute for 300M rows

**2.** ID graph vendor offers **MPC**-based portal that lets publisher run enrichment workflow

- **No 3<sup>rd</sup> party involved**, so no 3<sup>rd</sup> party fees
- **Publisher's data never disclosed**, and only intended results are disclosed to publisher
- Workflow takes ~60 minutes for 300M rows

**4.** Publisher and ID graph vendor run workflow on 3<sup>rd</sup> party's **TEE** cloud infrastructure

- **3<sup>rd</sup> party sees no data**, charges by volume
- **Publisher's data never disclosed**, and only intended results are disclosed to publisher
- Workflow takes ~5 minutes for 300M rows



# Evaluating & Selecting Other Considerations

- Noisy landscape: vendors list every PET, but many cannot offer them all
- Important to ascertain for what purpose each feature is present: how does it help you avoid disclosure and to which parties?

VARONIS  
DATAPRIVILEGE  
Data Access Governance

cinchy  
Platform Customers Resources  
The Data-Centric Platform That Makes Data Integration Open  
Learn more

snowflake  
PLATFORM SOLUTIONS  
SNOW THE CLOUD  
Trusted organization

Collibra

EGNYTE  
Platform Solutions Resources Partners Pricing  
Start Free Trial

Cornell University

arXiv.org > cs > arXiv:2011.01192

Computer Science > Databases

[Submitted on 2 Nov 2020 (v1), last revised 2 Jan 2021 (this version, v2)]

### Budget Sharing for Multi-Analyst Differential Privacy

David Pujol, Yikai Wu, ...

Large organizations that collect statistics that are used by these organizations are collecting data. Differential preserving privacy. How optimize the total accuracy release and have overlaid problem in DP where the In this work, we initiate capture the competing g mechanism should satisfy Adaptivity – while still of mechanisms in the multi DP algorithms that provide realistic tasks.

**Duality scores \$14M DARPA contract for hardware-accelerated homomorphic encryption**

Devin Coldewey @techrunch / 9:03 AM EST • February 3, 2021

Image Credits: Yulchiro Chino / Getty Images

Training AIs is essential to today's tech sector, but handling the amount of data needed to do so is intrinsically dangerous. DARPA hopes to change that by tapping the encryption experts at Duality to create a hardware-

IMMUTA

## Differential Privacy

Powering Private

Many data platforms don't offer a protection. Immuta allows you to inject noise into your data source, injecting noise into your data source, injecting noise into your data source. Combine differential based controls to meet even the most

- FEDERATED LEARNING
- DIFFERENTIAL PRIVACY
- HOMOMORPHIC ENCRYPTION
- ATTACK-BASED EVALUATION
- SECURE MULTI-PARTY COMPUTATION
- AND MORE...

WATCH THE DEMO

DATAFLEETS

Cryptoidentity uses advanced cryptography and information security best practices to enable privacy-compliant data sharing, including:

- Single-Use Pseudonymization  
Re-hashing all identifying data mixed with random noise for each activity to ensure pseudonyms cannot be leveraged for re-identification.
- Differential Privacy  
Adding noise to data elements to obscure individual details while maintaining aggregated metrics.
- Blind Facilitation  
Enlisting a neutral party to determine matches without any computational insight into the de-identification process. This gatekeeper approach blocks the flow of any unmatched data.
- Anonymization  
Completely remove all identifying attributes before working with personal data whenever insights can be gained at an aggregate level.

# Privacy-Enhancing Technologies

## Further Reading

- Resources below mostly focus on current/future state of PETs
- Vendors **cannot decide on their own** which PET solves your data non-disclosure needs; progress will involve collaboration

- “What Are Privacy-Enhancing Technologies (PETs) and How Will They Apply to Ads?”  
Facebook. <https://about.fb.com/news/2021/08/privacy-enhancing-technologies-and-ads/>
- “A guide to Privacy Enhancing Technologies (PETs) & how to adopt them!”  
Gilbert + Tobin via Lexology.  
<https://www.lexology.com/library/detail.aspx?g=79a577a0-a4b7-409c-aca6-cef77154ffa9>
- “Privacy Enhancing Technologies Adoption Guide”.  
Centre for Data Ethics and Innovation. <https://gov.uk/cdei>
- “The Age of Collaborative Computing”. Lunar Ventures via Medium.  
<https://medium.com/lunar-ventures/the-age-of-collaborative-computing-e73374b7aedc>