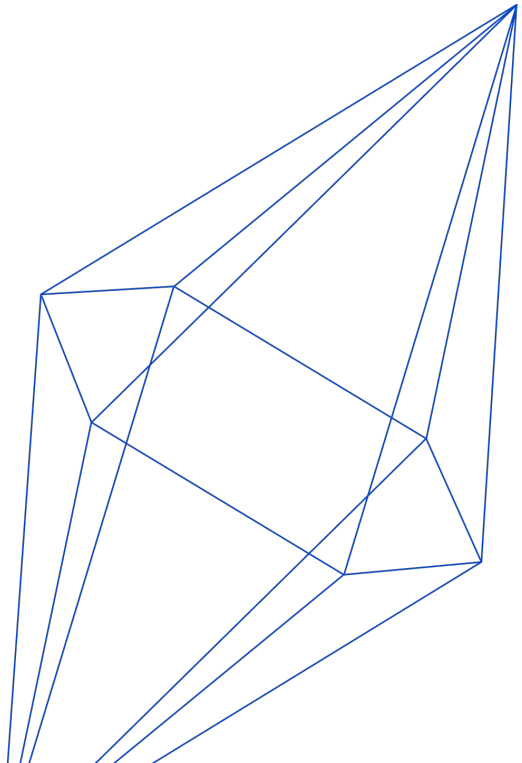
TUMULT LABS

How We Can Save Anonymization



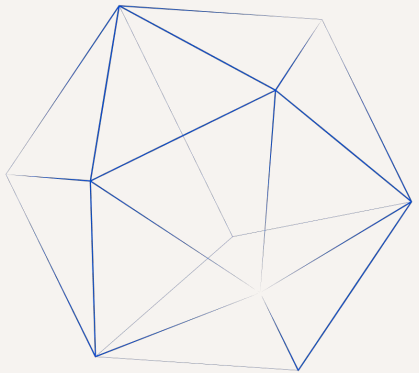
A ~~rant~~ talk by Daniel Simmons-Marengo

How do users understand anonymization?

→ “Data is only retained in anonymized form.”

→ “We only share data that has been anonymized ang aggregated.”

→ “Only anonymous data leaves your device.”



FEDERAL TRADE COMMISSION

Claims that data is “anonymous” or “has been anonymized” are often deceptive. Companies may try to placate consumers’ privacy concerns by claiming they anonymize or aggregate data. Firms making claims about anonymization should be on guard that these claims can be a deceptive trade practice and violate the FTC Act when untrue. Significant research has shown that “anonymized” data can often be re-identified, especially in the context of location data. One set of researchers demonstrated that, in some instances, it was possible to uniquely identify 95% of a dataset of 1.5 million individuals using four location points with timestamps. Companies that make false claims about anonymization can expect to hear from the FTC.

The Guardian

'Anonymised' data can never be totally anonymous, says study

Findings say it is impossible for researchers to fully protect real identities in datasets

UCLA LAW REVIEW

BROKEN PROMISES OF PRIVACY: RESPONDING TO THE SURPRISING FAILURE OF ANONYMIZATION

New dataset uncovers Wikipedia browsing habits while protecting users

21 June 2023 by Hal Triedman, Isaac Johnson and Nuria Ruiz

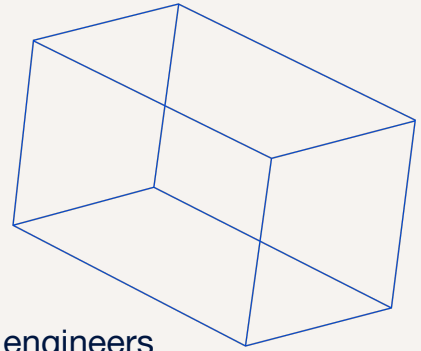
Translate this post

Isolation and Community, a CC BY-SA 4.0-licensed work from Wikimedia Commons.

Projects supported by the Wikimedia Foundation, such as Wikipedia, are among the most used online resources in the world, garnering hundreds of billions of visits each year from around the world. As such, the Foundation has access to terabytes of

Some anonymization techniques

- → k-anonymity
- → differential privacy
- → remove all the identifiers
- → vibes
- → cell suppression
- → expert determination
- → hashing
- → swapping
- → a new technique your engineers just thought up
- → generalize values
- → threaten to sue attackers
- → l-diversity



FCC:

Don’t mess it up.

CJEU

Only consider those intended to have access.

Article 29 Working Party:

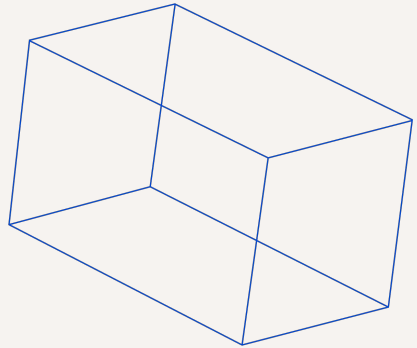
Consider anyone who might have access.

CPPA

It’s called deidentification.

How do we know which anonymization strategies work?

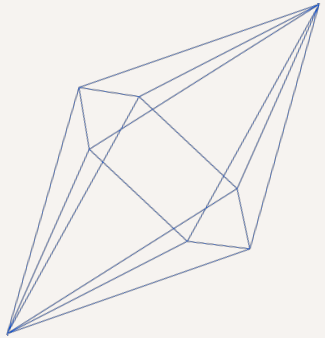
1. Nothing is obviously anonymization



2. Minimize assumptions

# Library Checkouts	Website Visited
0-10	www.slate.com
51-100	www.localdomesticviolenceshelter.org
25-50	www.youtube.com
10-25	www.netflix.com
25-50	www.nytimes.com

3. Governance is not anonymization

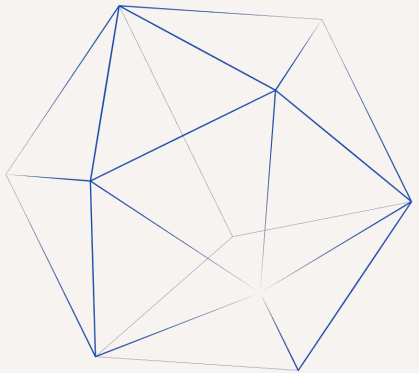


4. The next release shouldn’t break the last one

ZIP Code	Count
95050	101

ZIP Code	Cancer Diagnosis	Count
95050	Negative	100
95050	Positive	Redacted

5. Your technique should stand up under transparency




JAMES MONTGOMERY FLAGG
I WANT YOU
TO ANONYMIZE
WELL

TUMULT LABS

Questions?