

Data Securit(ies)

Gianluca Iannaccone
Intel Research Berkeley

Disclaimer

- I don't do any data anonymization (research or otherwise)...
... because I don't collect/share data anymore
- Current policies are a strong disincentive
 - Opt-in for data collection
 - No sharing without prior written agreement
 - Disclose in advance all possible uses of the data
- and “not that there is anything wrong with that!”

Data Anonymization

- Allows data exchange among untrusted parties
 - Users (data producers/owners) provide data to *trusted* party (data collector)
 - Collector is trusted *if* private data is anonymized
 - Collector can share anonymized data with 3rd parties that may not be trusted by the users (or the collector).

Properties of Data Anonymization

- **Privacy:** are you really collecting just what you need?
 - “if I screw up, your private information is safe”
- **Security:** is the data collection/storage secure?
 - “if my IT dept screws up, your private information is safe”
- **Trust:** are you sharing more than I agreed to?
 - “if my company screws up, your private information is safe”

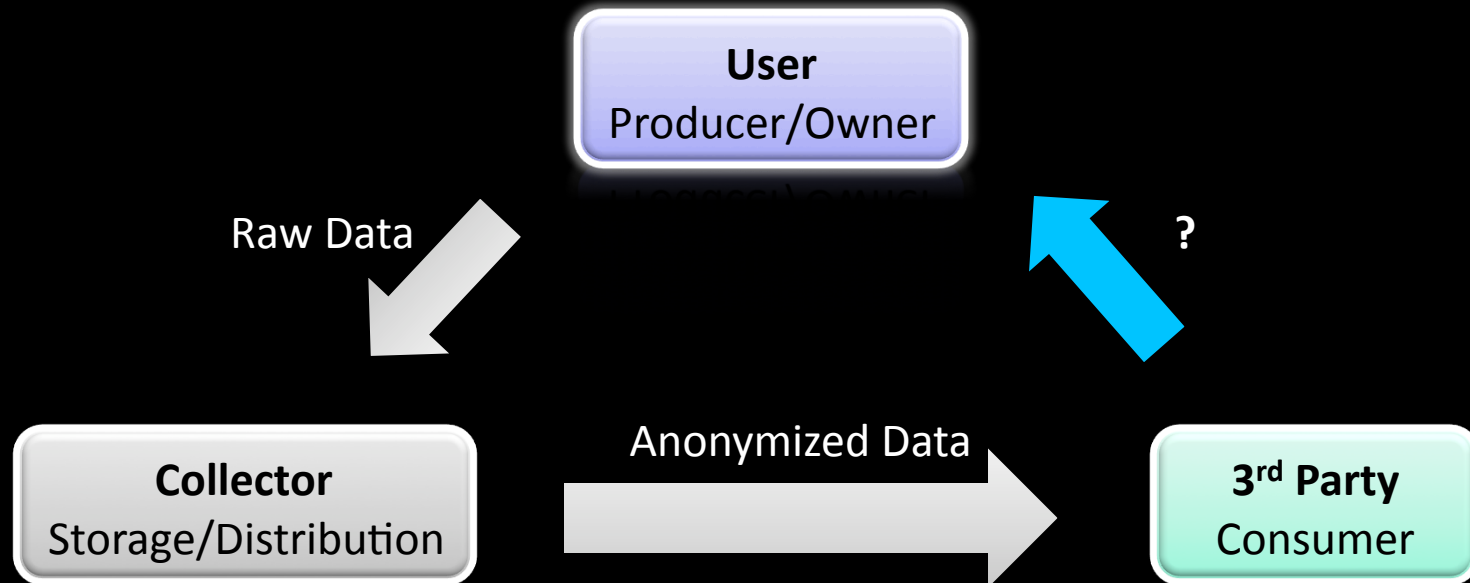
- Can you see the patterns?

So which one is it?

- Is data anonymization just a form of insurance for the data collectors?
 - Well... if that is the case, it's going to be a really hard sell...
- Does data anonymization have value for the users, the data owners?
 - It needs to be tied to the actual information to be protected and its uses.
 - Data is private if it allows to identify the user, its location, etc. but it may have a much broader definition
- ... But how does one give value to data?

Value in the data

Data anonymization is (part of) a multi-party transaction



But what is in it for the user (producer)?

How can users give value to data?

- One would need a market first
 - Not pretty but it's already happening (and it always had)
 - Privacy concerns do scale with the discount coupon
- Should we try to formalize the process?
 - Expose all private data transactions
 - Make it clear that they carry a value
 - Return some of the value back to the users
- Could also give value to properties (privacy/trust/security)
 - e.g., what is the cost of keeping the data safe?
 - e.g., how many other parties to share data with?