Well-kept secrets in your business applications

Dan Bogdanov

Cybernetica / University of Tartu

dan@cyber.ee

Secrets

business plan

product specifications

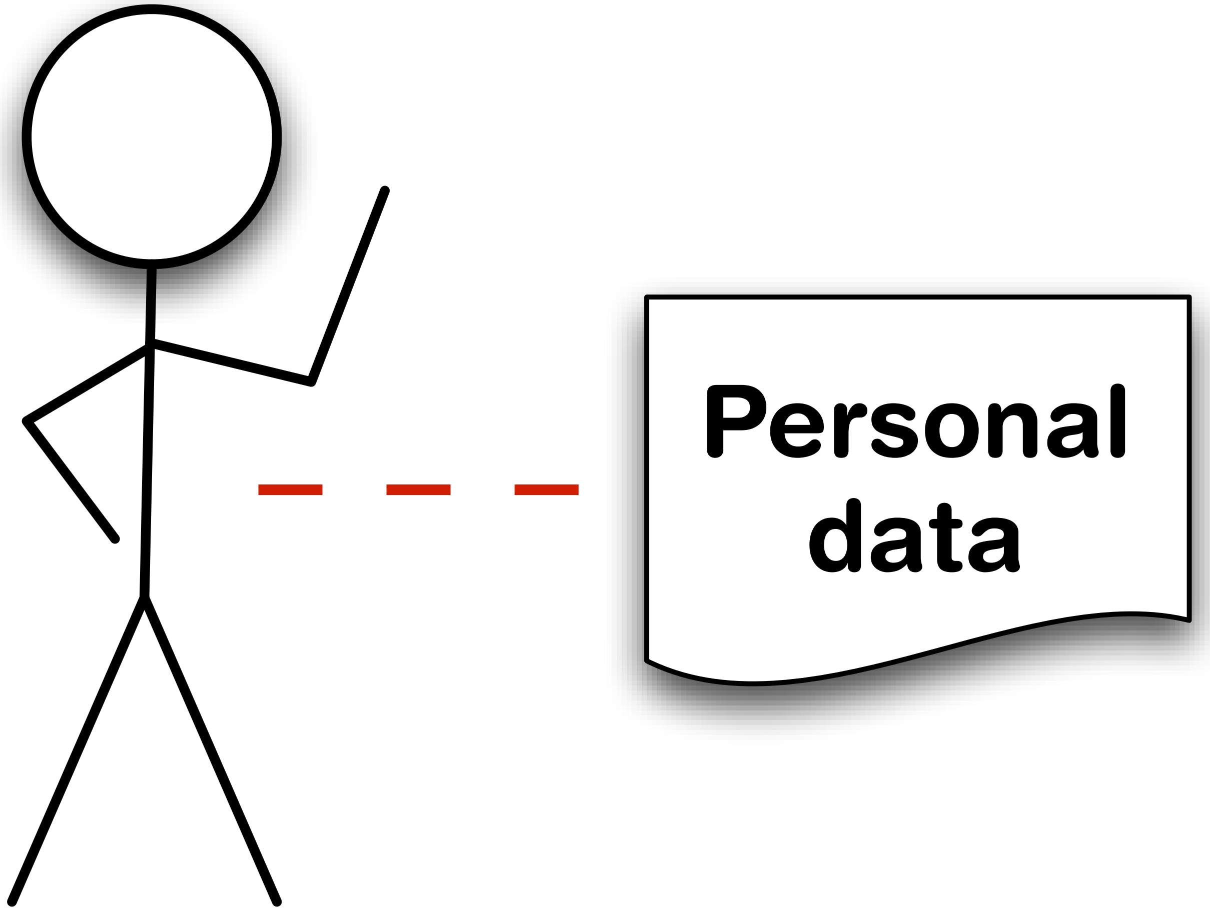
financial data

personal data

CEOs collection of kitten pictures

Privacy

Starring

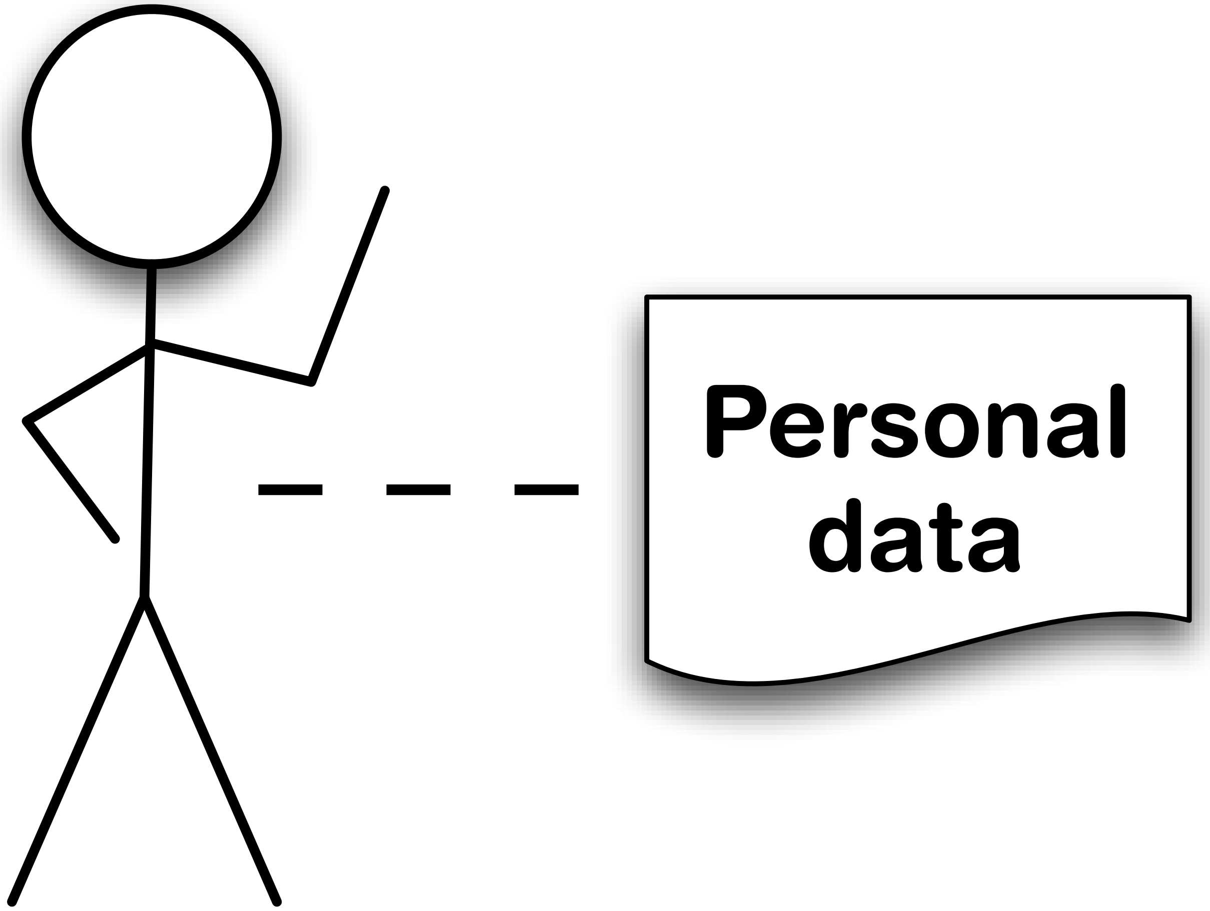


Personal data

Tom

Tom’s data

Starring

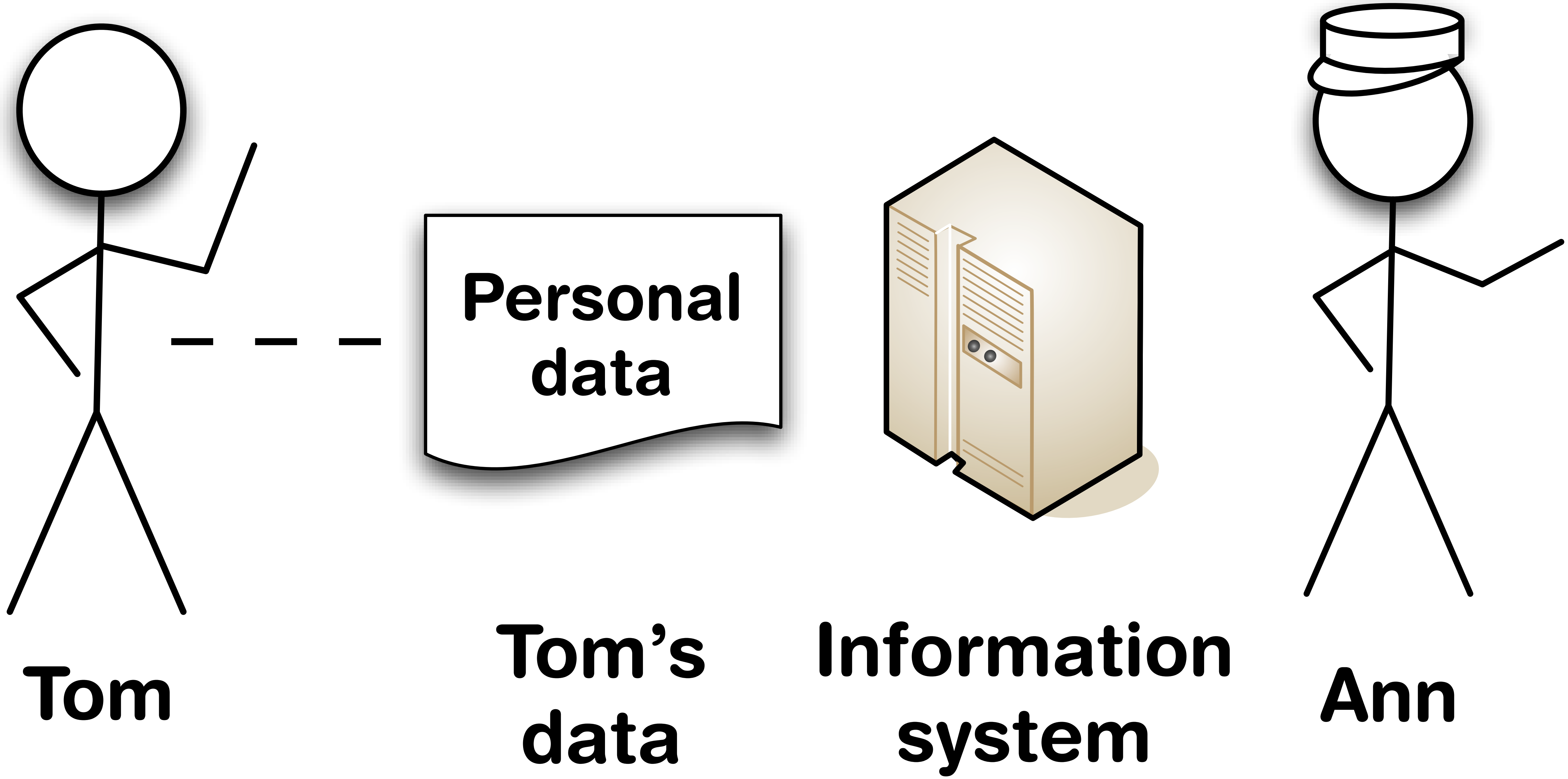


Personal data

Tom

Tom's data

Also starring



Personal data

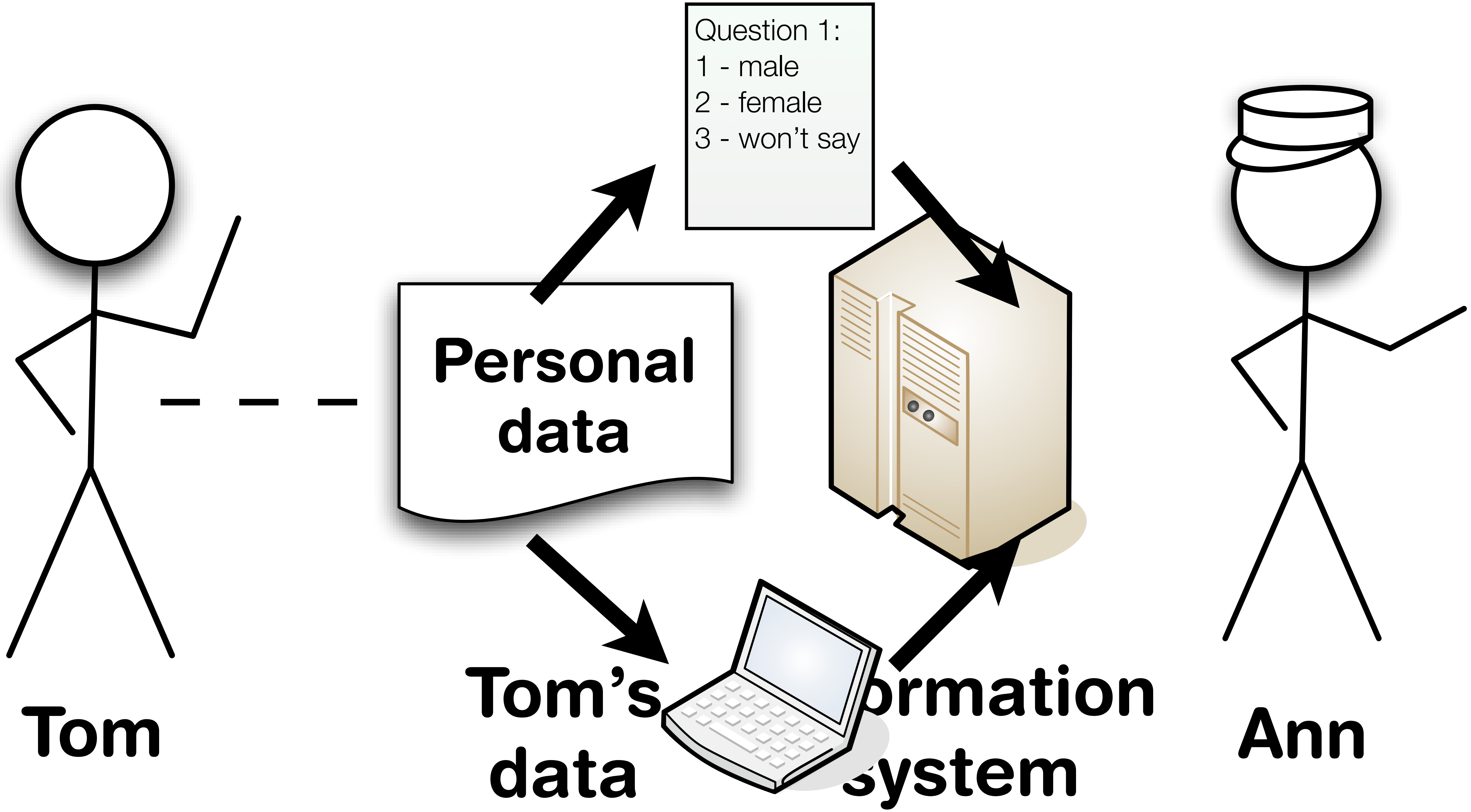
Tom

Tom’s data

Information system

Ann

People give data out quite freely



Question 1:
1 - male
2 - female
3 - won't say

Personal data

Tom

Tom's data

ormation system

Ann

Means of protection

legislation

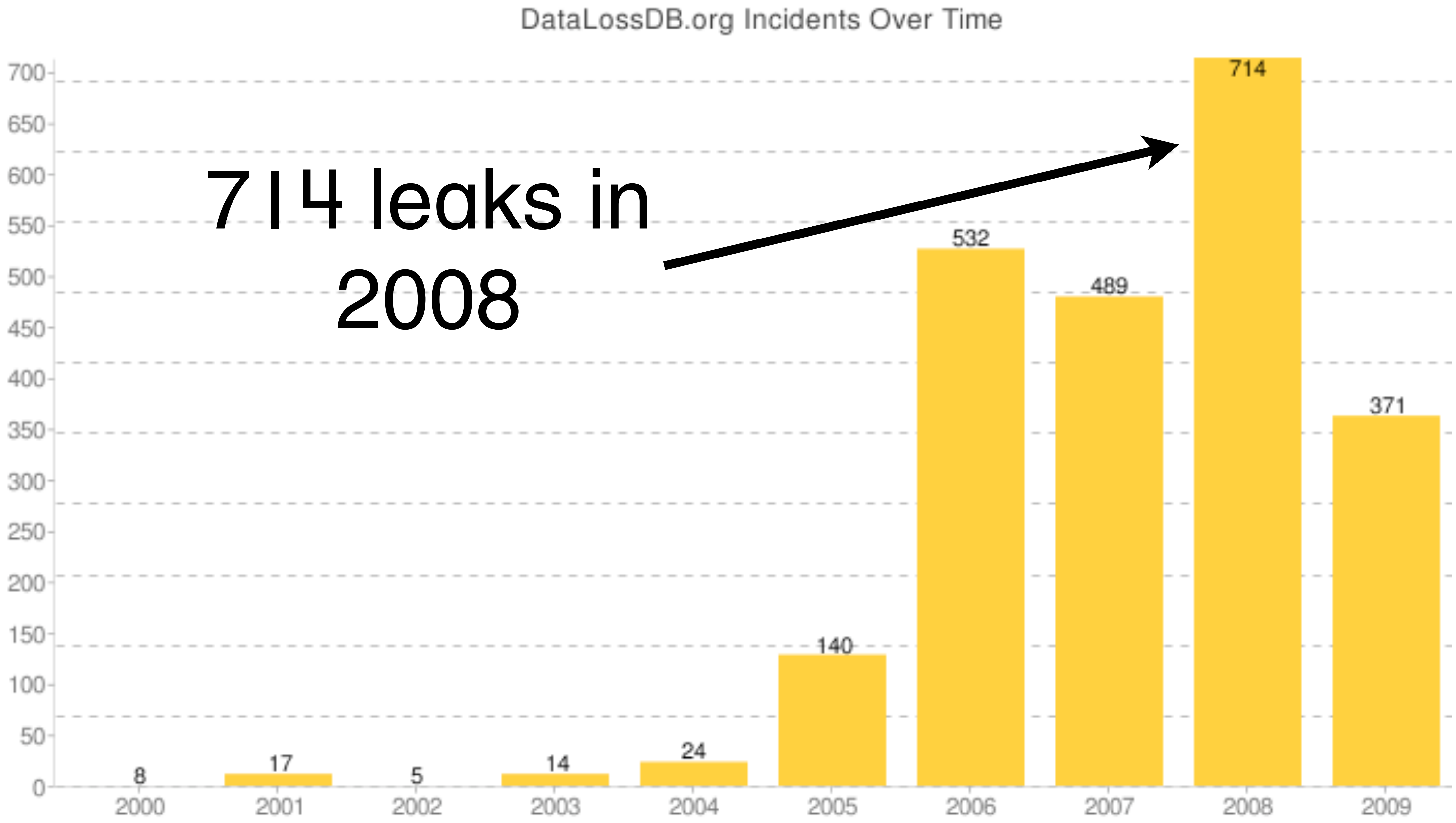
privacy policies

contracts

technical means

The Problem

Personal data leaks all the time



714 leaks in 2008

Source: www.datalossdb.org

How does this happen?

network break-ins

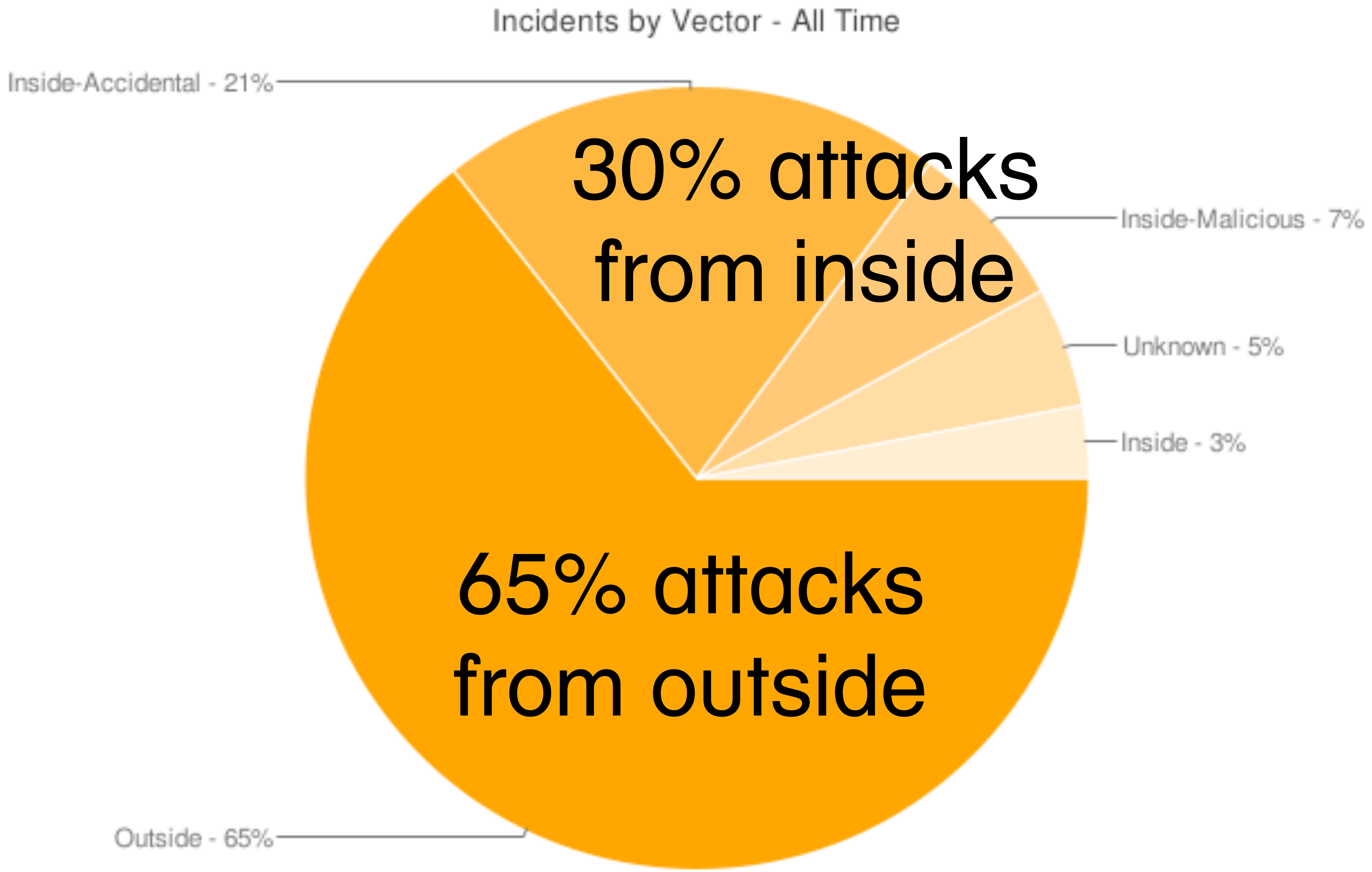
lost portable media

stolen backups

...

putting the data on the cloud?

Some more statistics



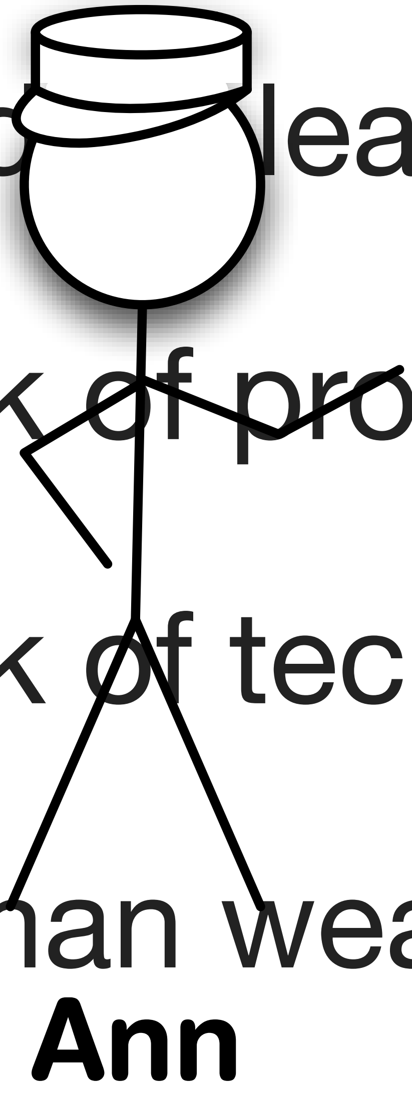
30% attacks from inside

65% attacks from outside

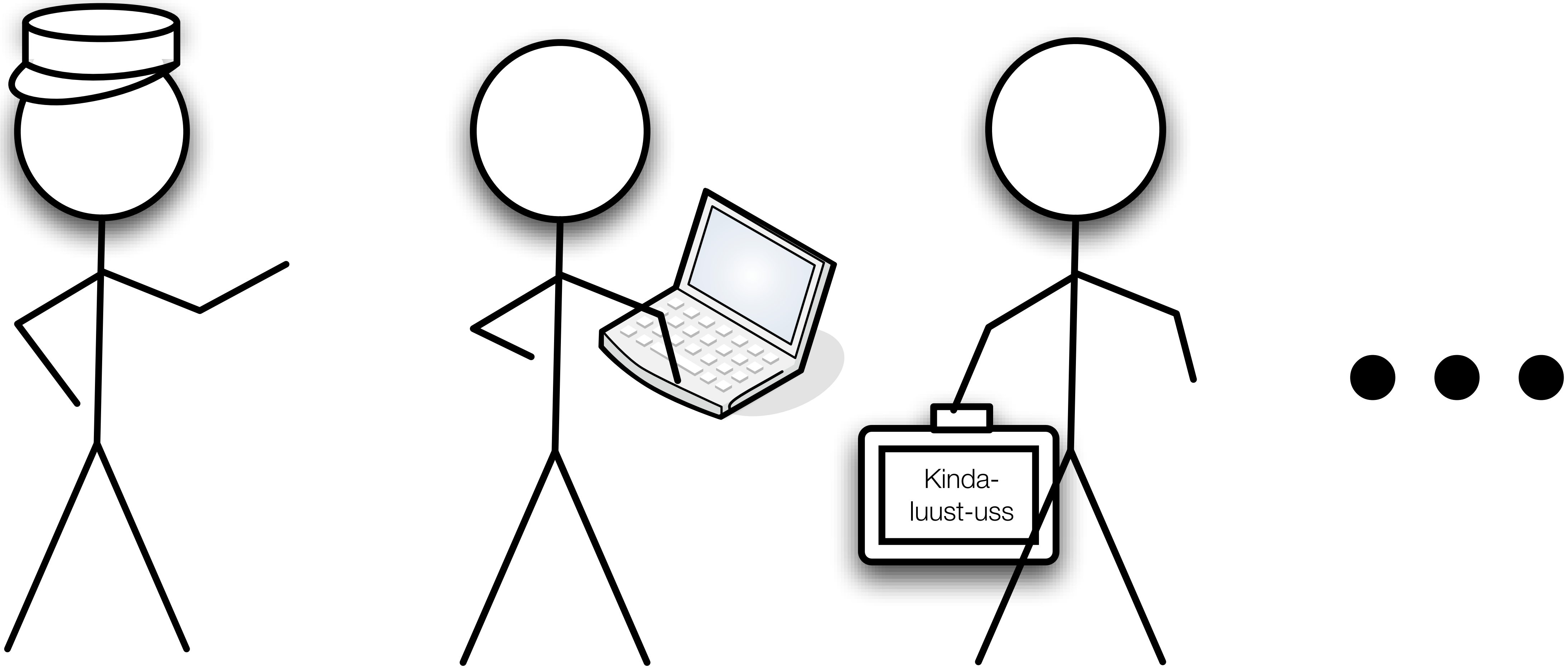
Source: www.datalossdb.org

One of the suspects

- Insiders leak data
- Lack of procedures
- Lack of technological means
- Human weakness



Who is interested in breaking privacy?



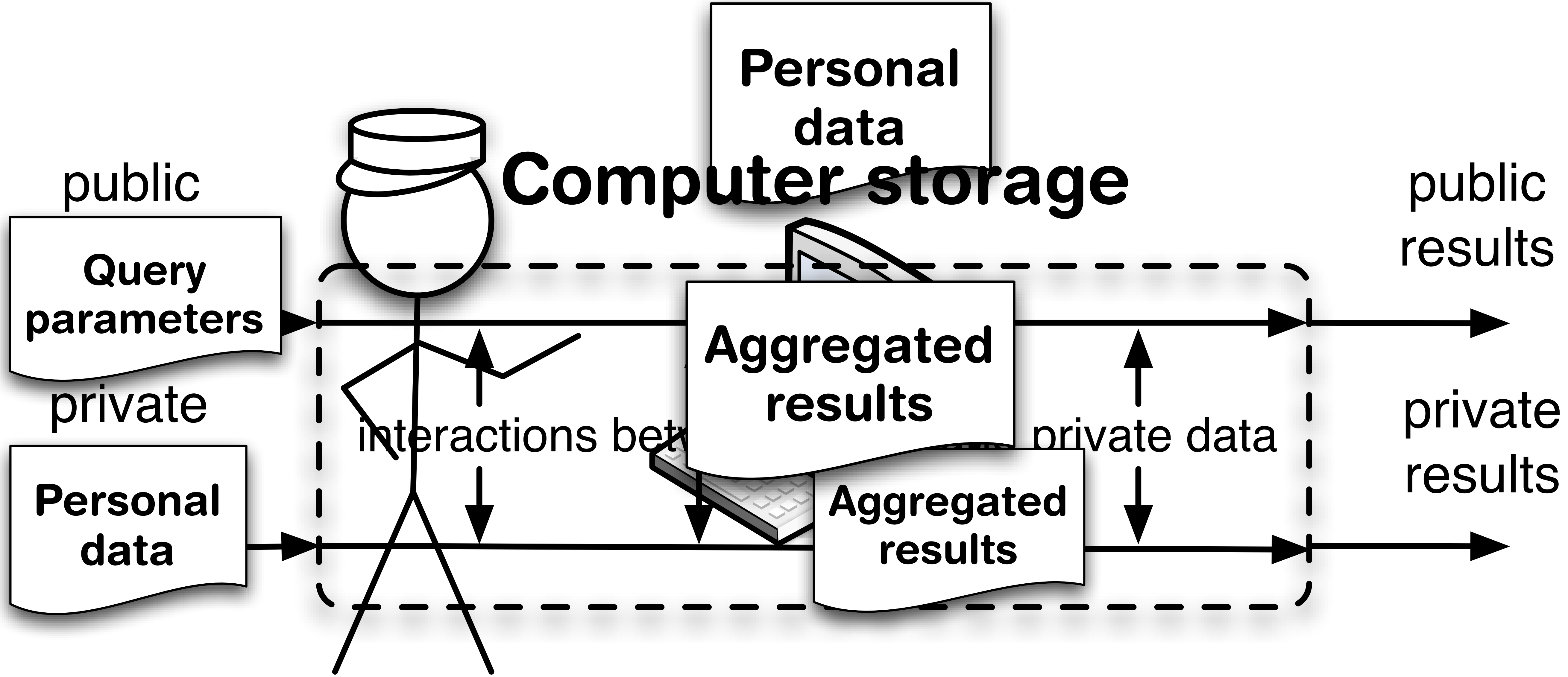
Ann

Mark (marketing)

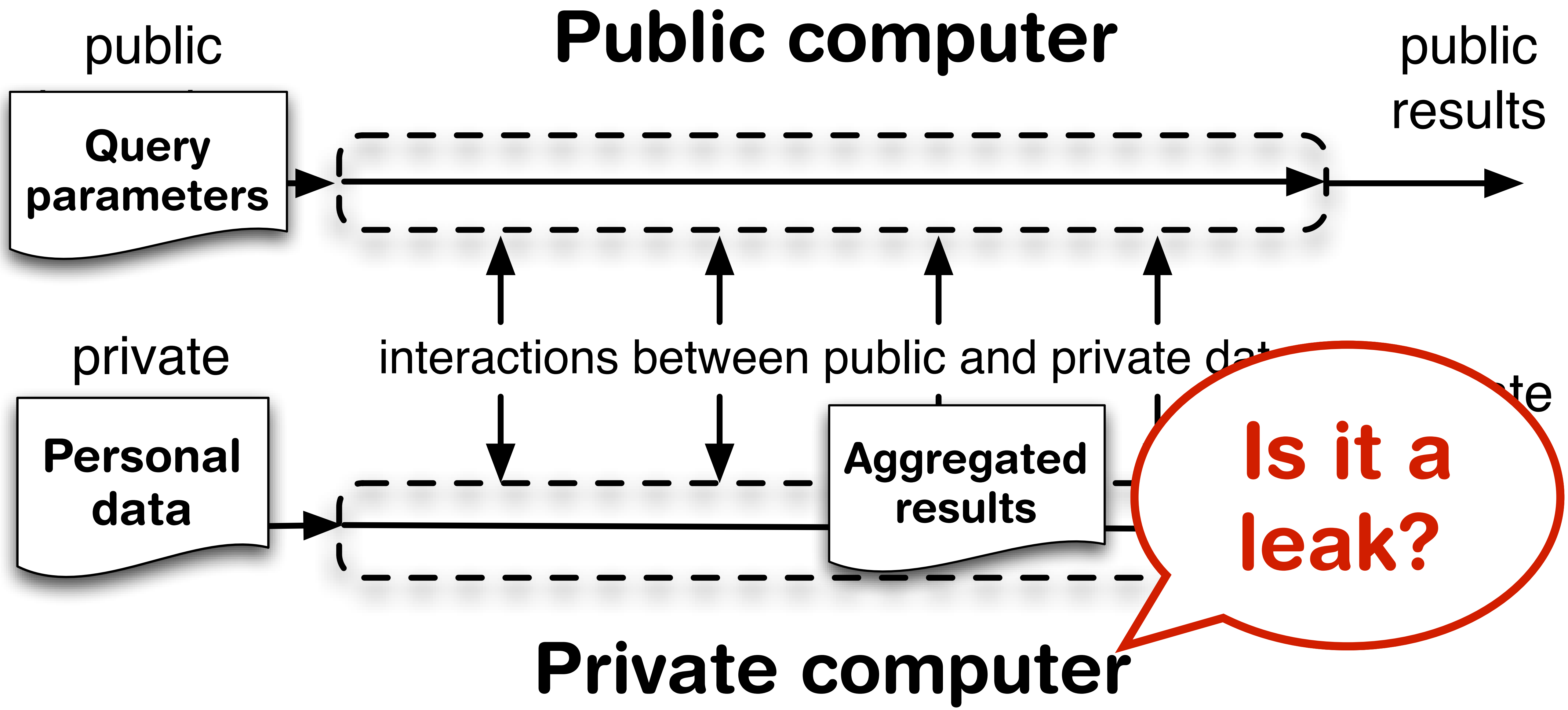
Ian (insurance)

Change the model!

How data is usually processed



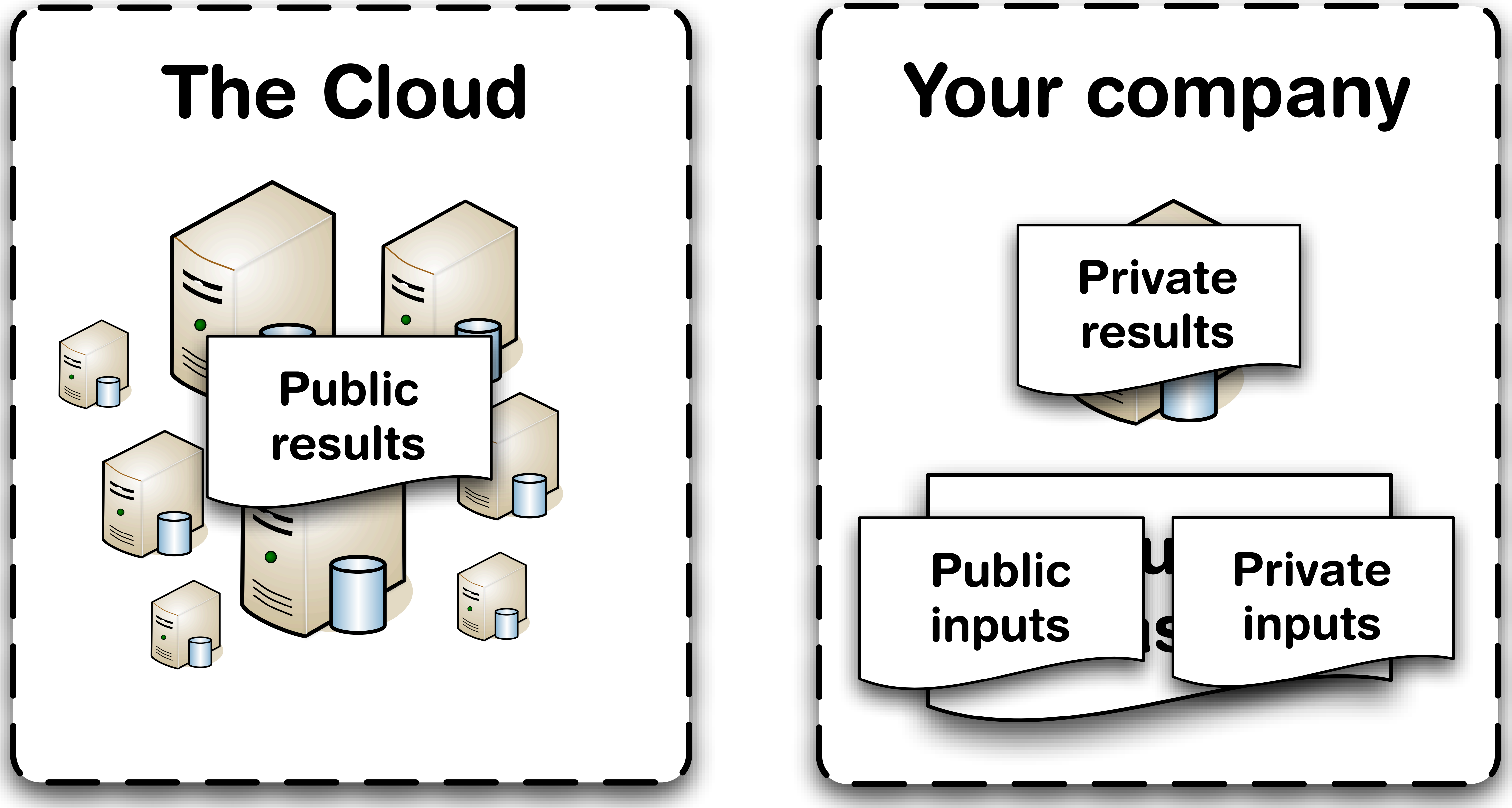
This is how we can protect private data



This is nice, but who will sell us a private computer?

First - note, that privacy depends on the context.

An example inspired by the cloud



The Cloud

Public results

Your company

Private results

Public inputs

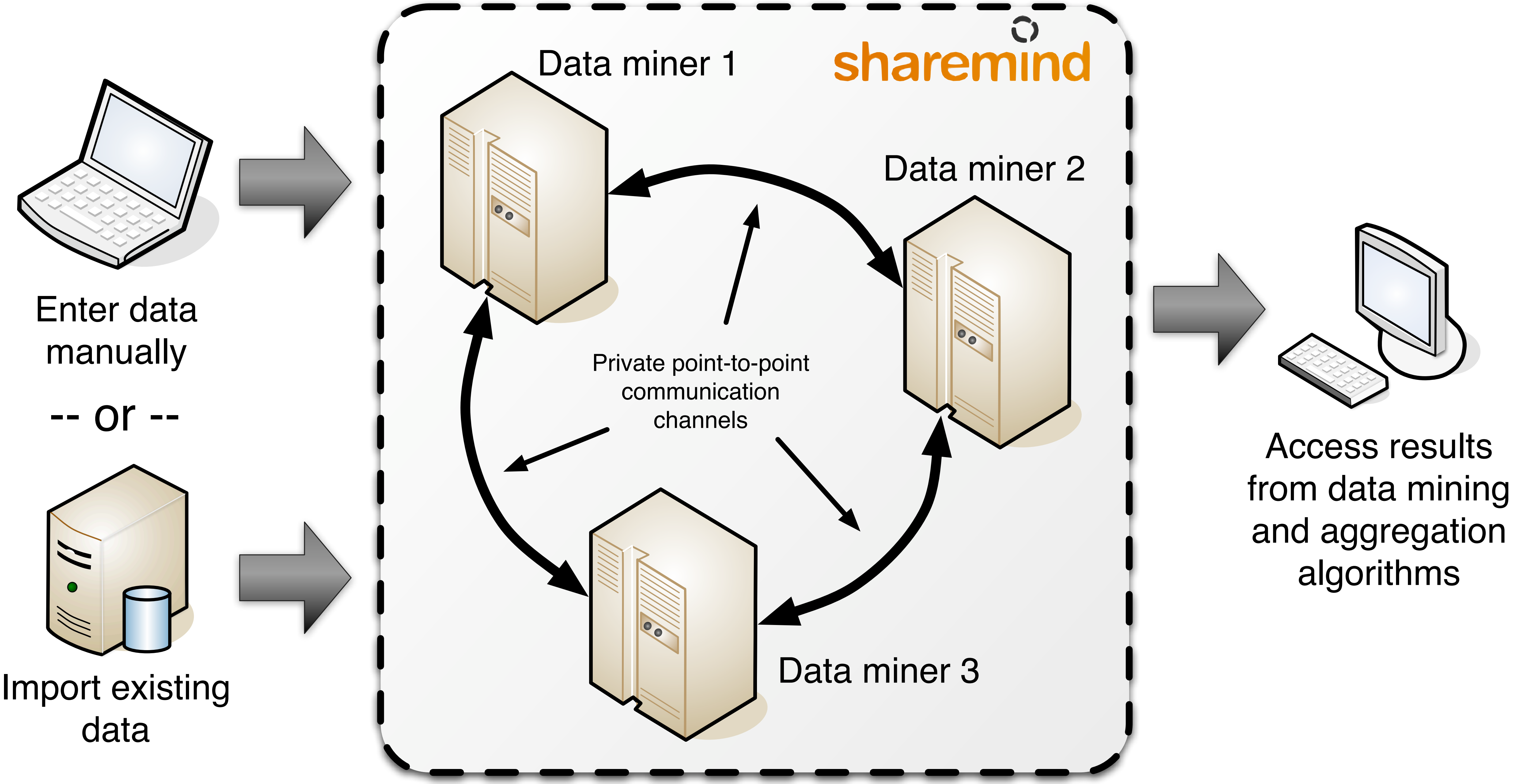
Private inputs

But what if even you yourself should not see the data?

Privacy-preserving computing!

- Researchers have been doing it for years.
- Several techniques - cryptocomputing, share computing, circuit evaluation...
- The possible privacy guarantees are better than with any existing technology.
- There are just a few implementations.

The Sharemind private computer


Enter data manually
-- or --
Import existing data
Data miner 1
sharemind
Data miner 2
Private point-to-point communication channels
Data miner 3
Access results from data mining and aggregation algorithms

Guarantees for the three-party case

- Given that no miner shares its data:
  - nobody can see the private inputs,
  - data can be processed privately,
  - only the final results are published.
- Note, that all the miners have to follow protocol for the system to work.

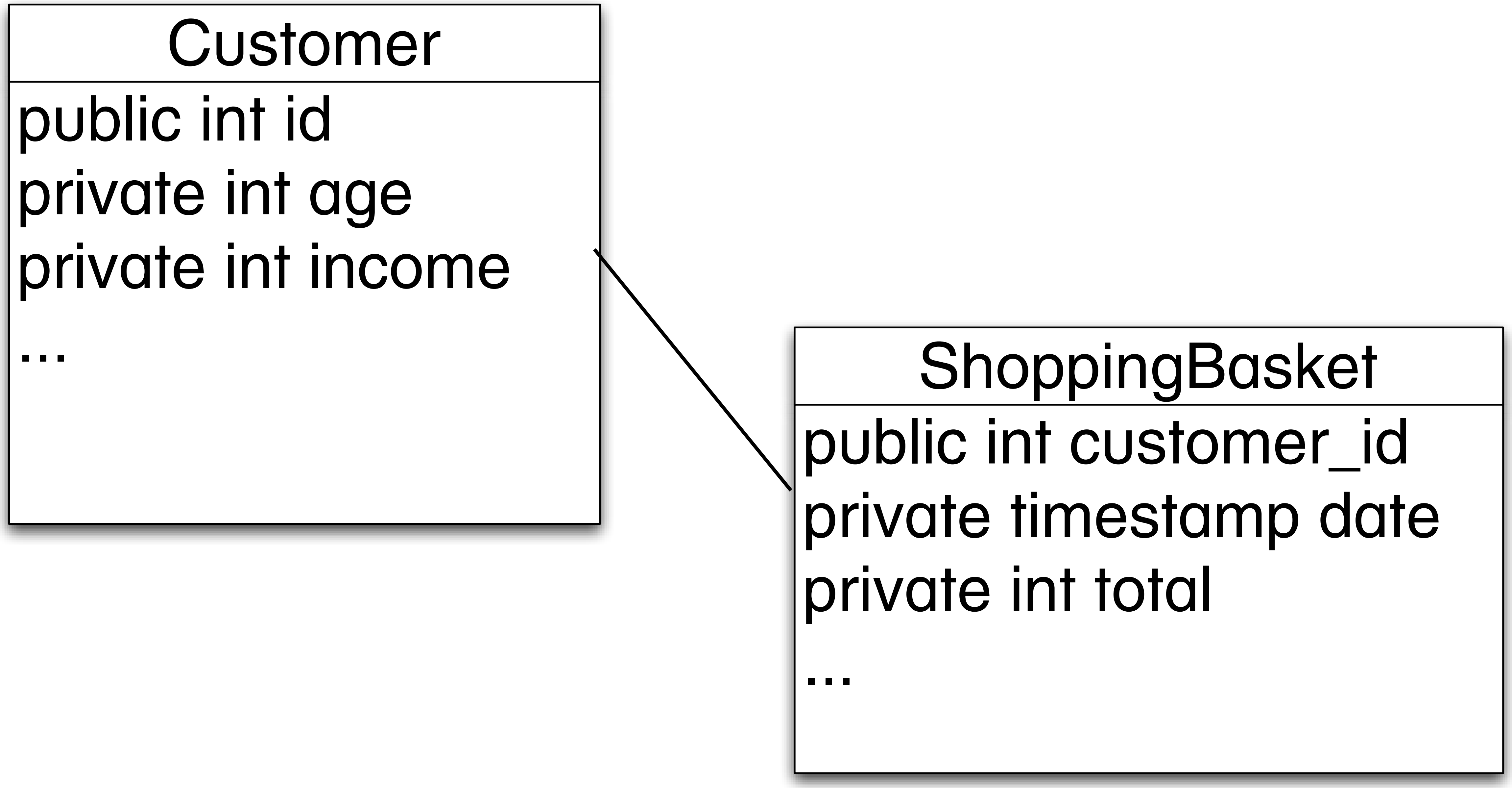
Software can be developed by non-cryptographers

Making an information system

1. Decide on the user roles.
2. Design a data model.
3. Choose data processing algorithms.
4. Implement user tools for entering data and running queries.
5. Convince users to use it.

Separate public and private in both data and processes.

Example: private data models



Customer

public int id
private int age
private int income
...

ShoppingBasket

public int customer_id
private timestamp date
private int total
...

Example: writing private algorithms

```
  public int count (private int[] data,
                  public int needle)
{
  public int data_size = vecLength (data);
  private int matchcounter = 0;
  public int i = 0;
  for (i = 0; i < data_size; i = i + 1) {
    private bool match = (data[i] == needle);
    matchcounter = matchcounter + match;
  }
  return declassify (matchcounter);
}
```

Separation for public and private data

```
public int count (private int[] dat
                  public int needle
{
  public int data_size = vecLength (data);
  private int matchcounter = 0;
  public int i = 0;
  for (i = 0; i < data            i + 1) {
    private bool mat            == needle);
    matchcounter = m            match;
  }
  return declassify (matchcounter);
}
```

STOP!

STOP!

The Sharemind toolset

a runtime for three-party computations

a library for creating applications

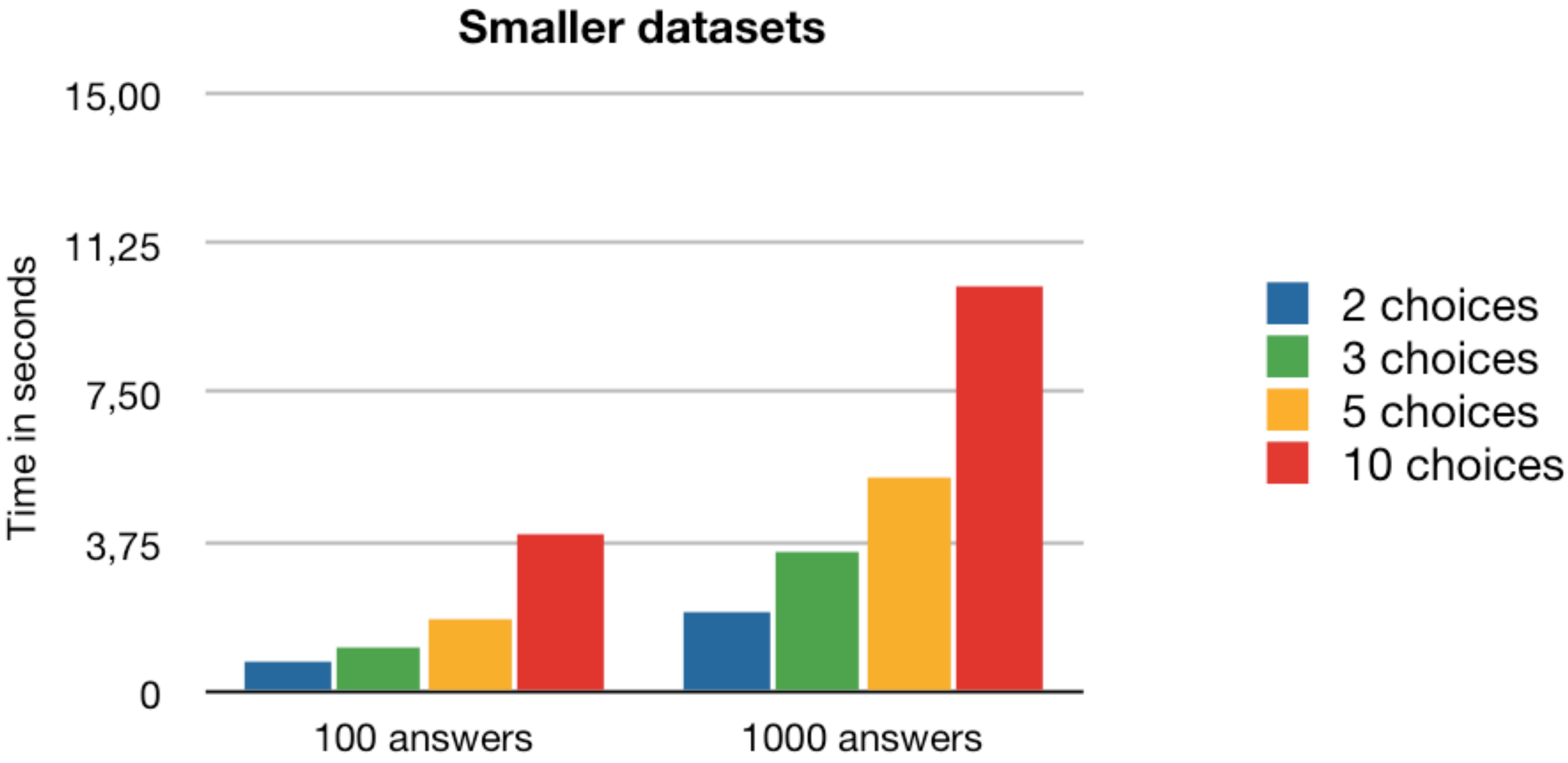
SecreC algorithm language

example applications

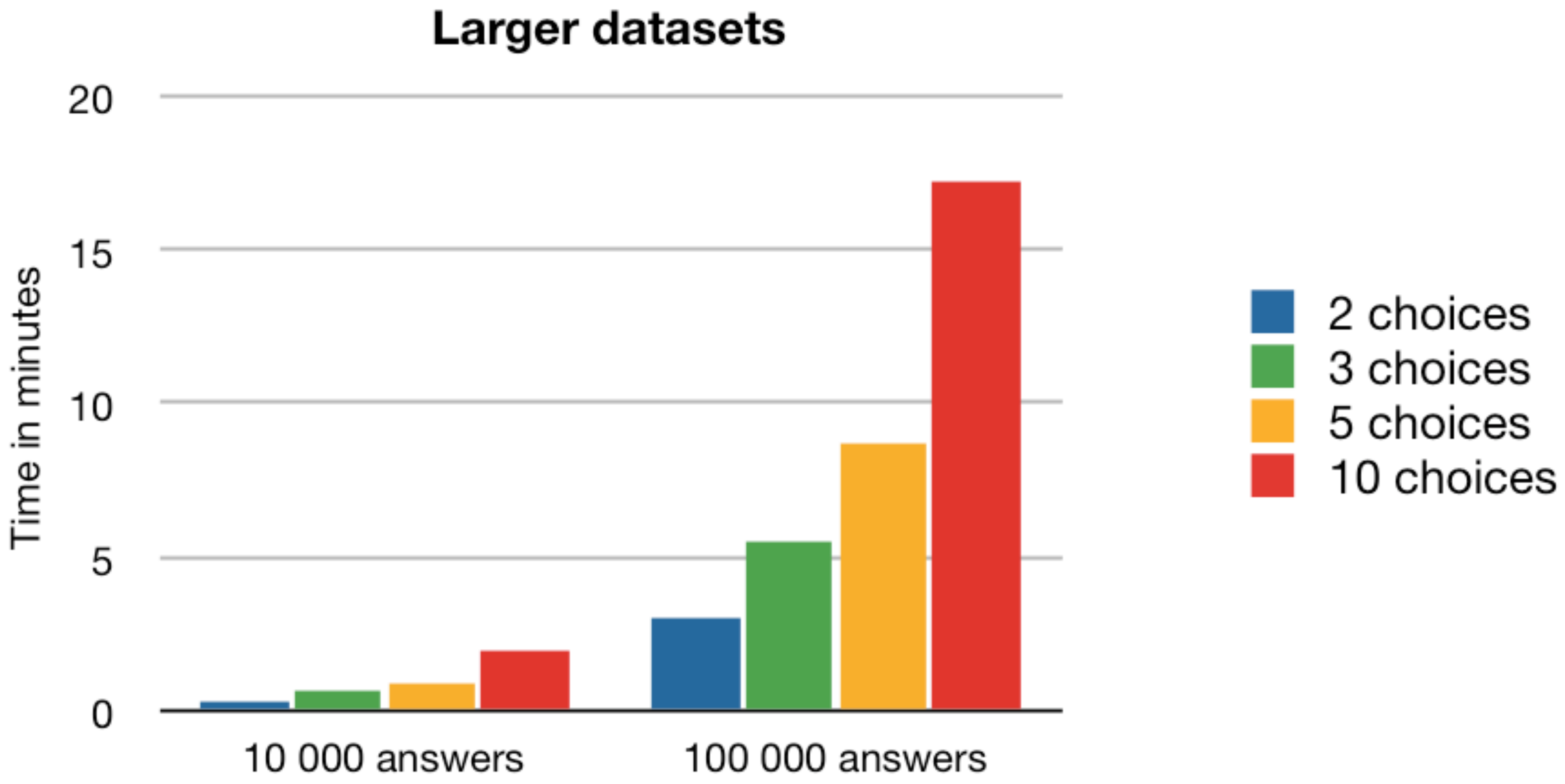
State of the art

- The three-miner case is almost done.
- Performance is best among competition.
- It is slower than normal computations.
- Currently, focus is on development tools.
- Considering extending to more platforms.

Histogram performance



Histogram performance



We can find the average income of people in this room without looking at the individual wages.

So could your data warehouse.

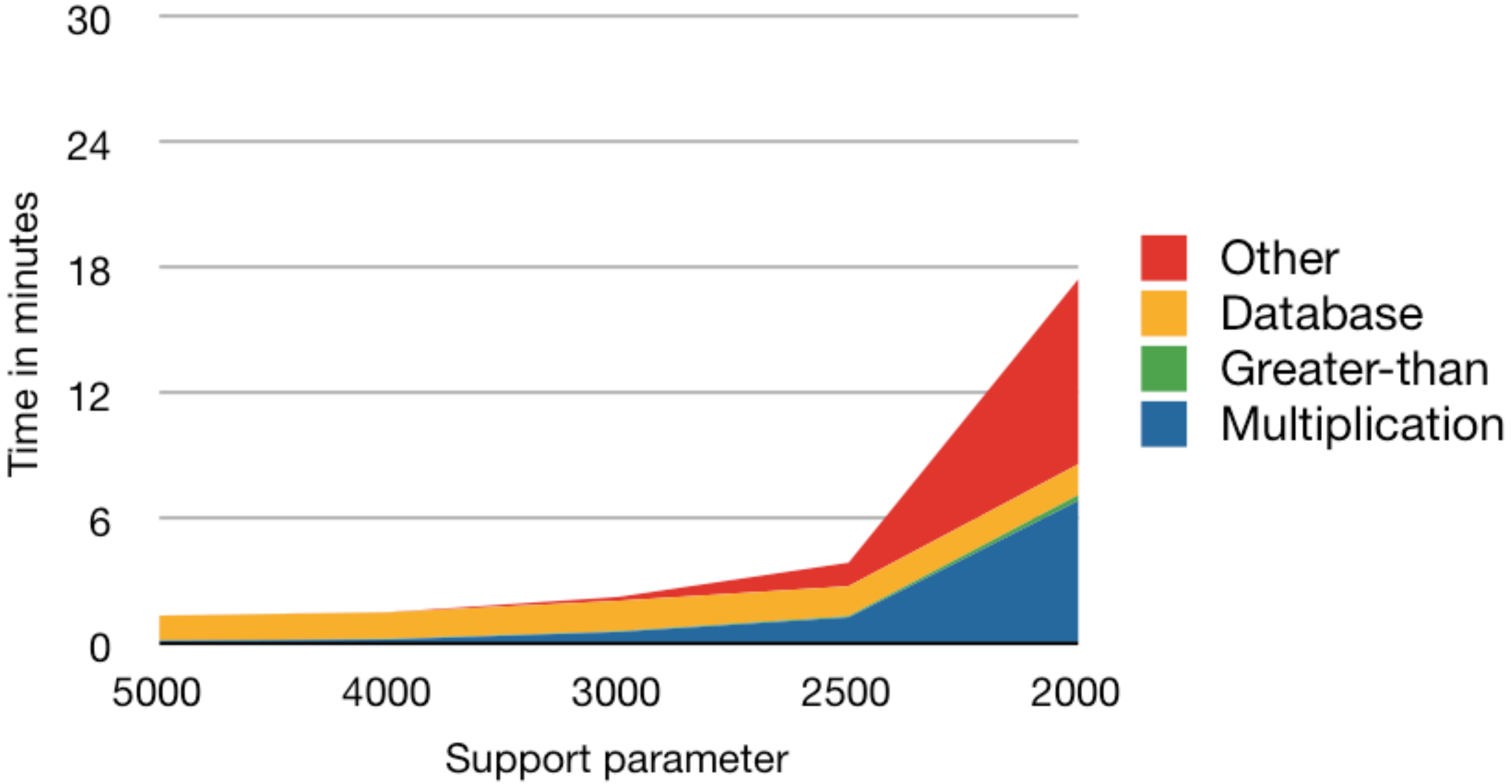
Thank you!

sharemind

http://research.cyber.ee/sharemind/

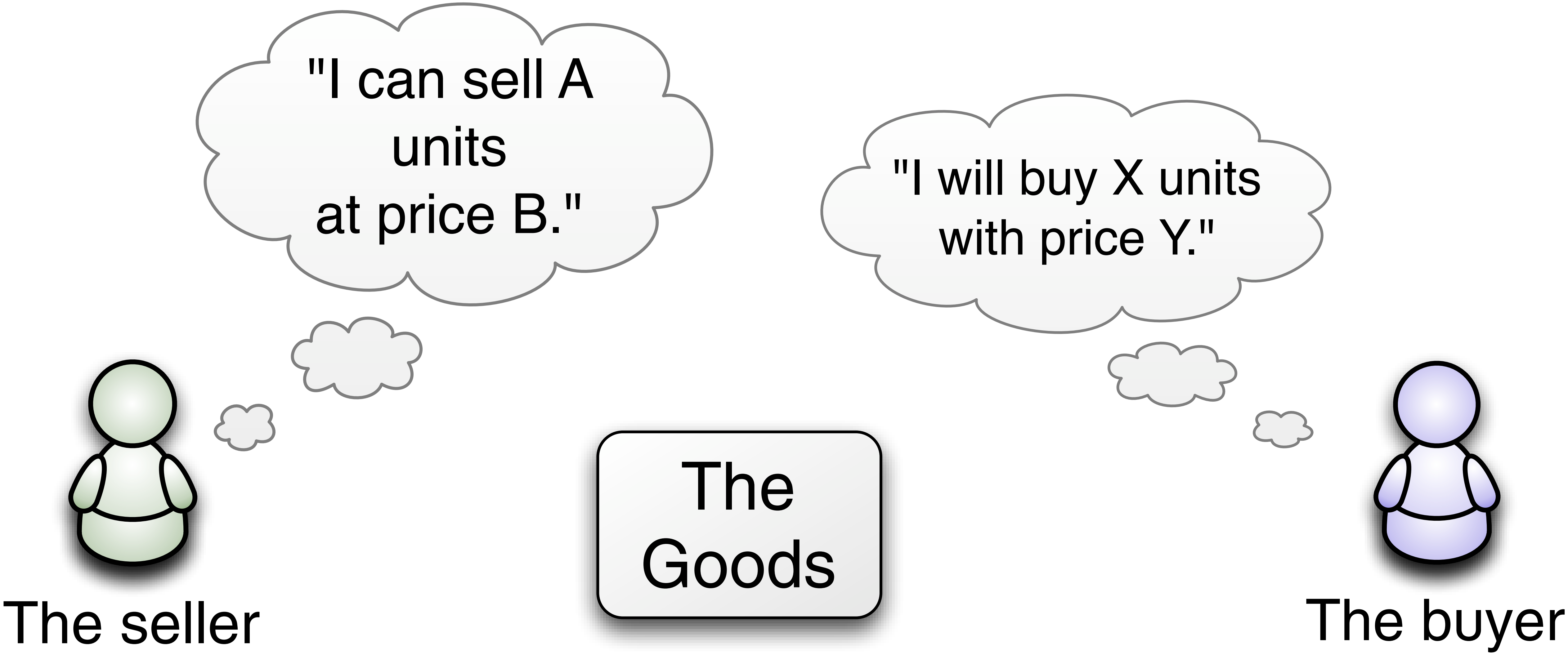
dan@cyber.ee

Frequent itemset mining performance



Experiments on the ‘mushroom’ dataset,
8124 transactions, 120 columns.

A real-life question



"I can sell A units at price B."

"I will buy X units with price Y."

The seller

The Goods

The buyer

Is there a price point *p* that would clear the market?