DATA ETHICS

PROF. DR. FLORIAN STAHL

Overview – Data Ethics

What are ethics?

Data ownership

Privacy

Anonymity

Data validity

Algorithmic fairness
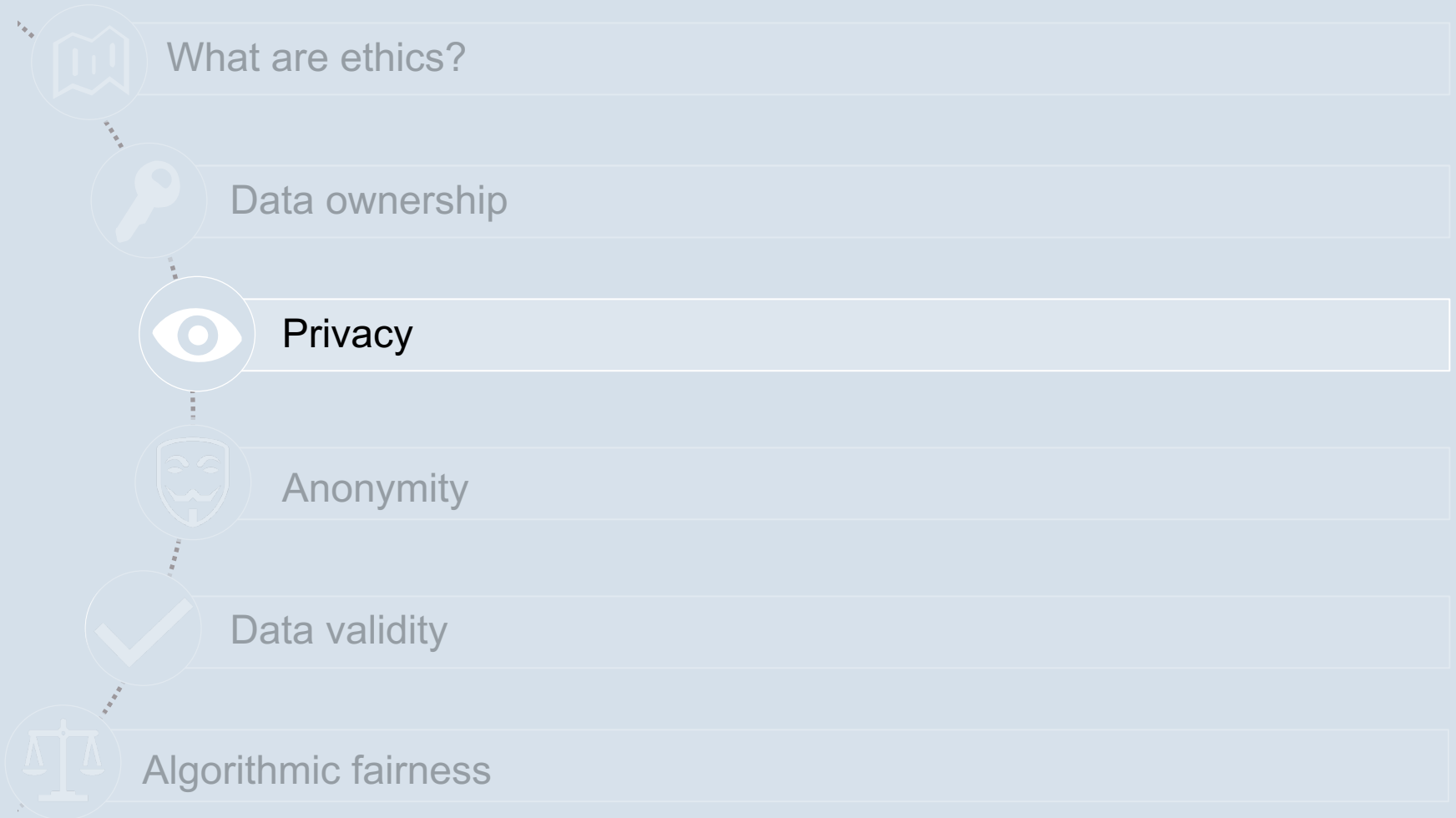
What is Privacy?
Argus Panoptes

Greek myth giant with
"a hundred" eyes
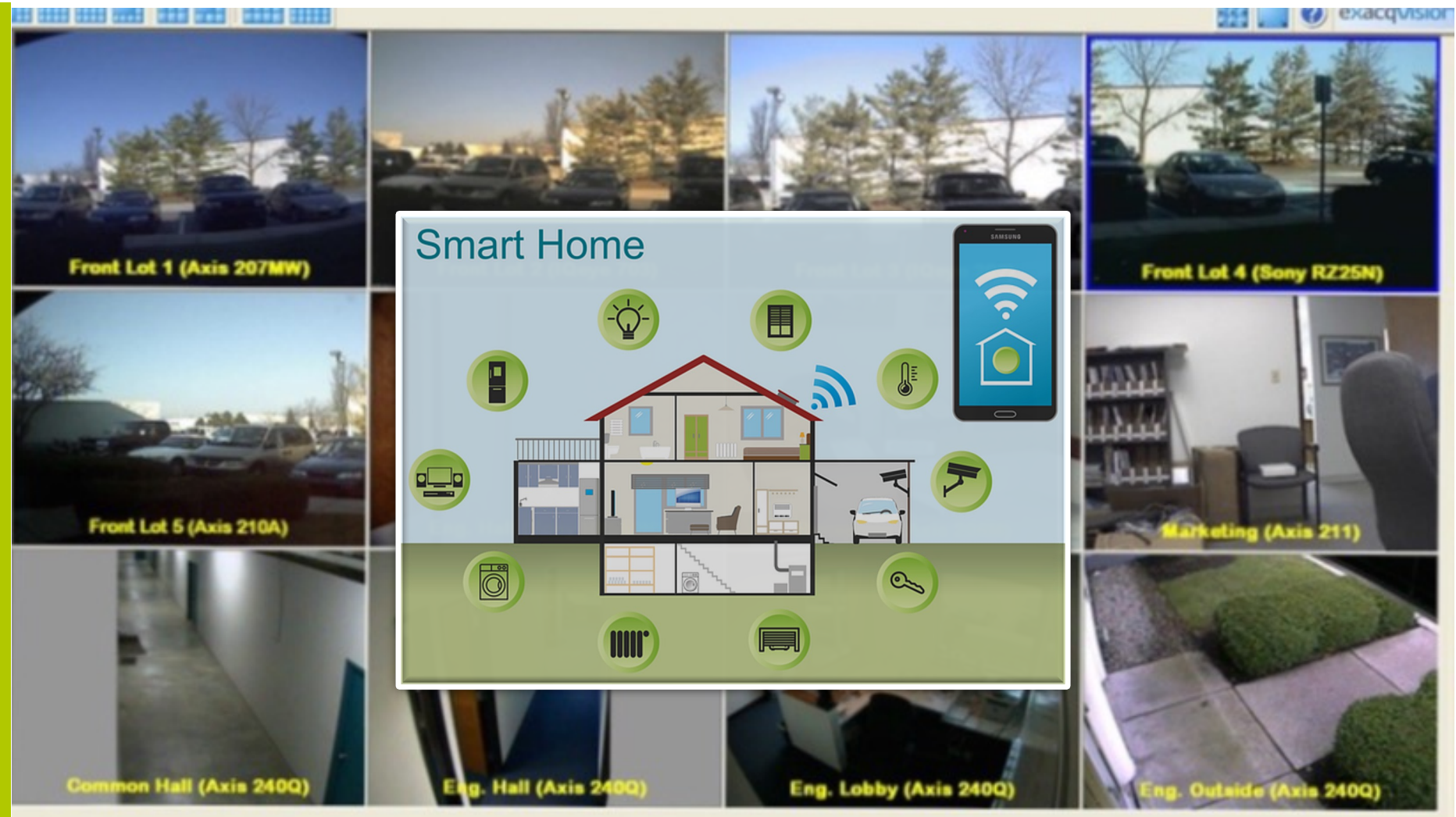is **always watchful.**

Picture: Pinterest

What is Privacy?
Panopticon

Jeremy Bentham designed a prison where a single guard could **observe all the prisoners**. Prisoners would not know when they were being watched, so they would "**behave**" all the time.
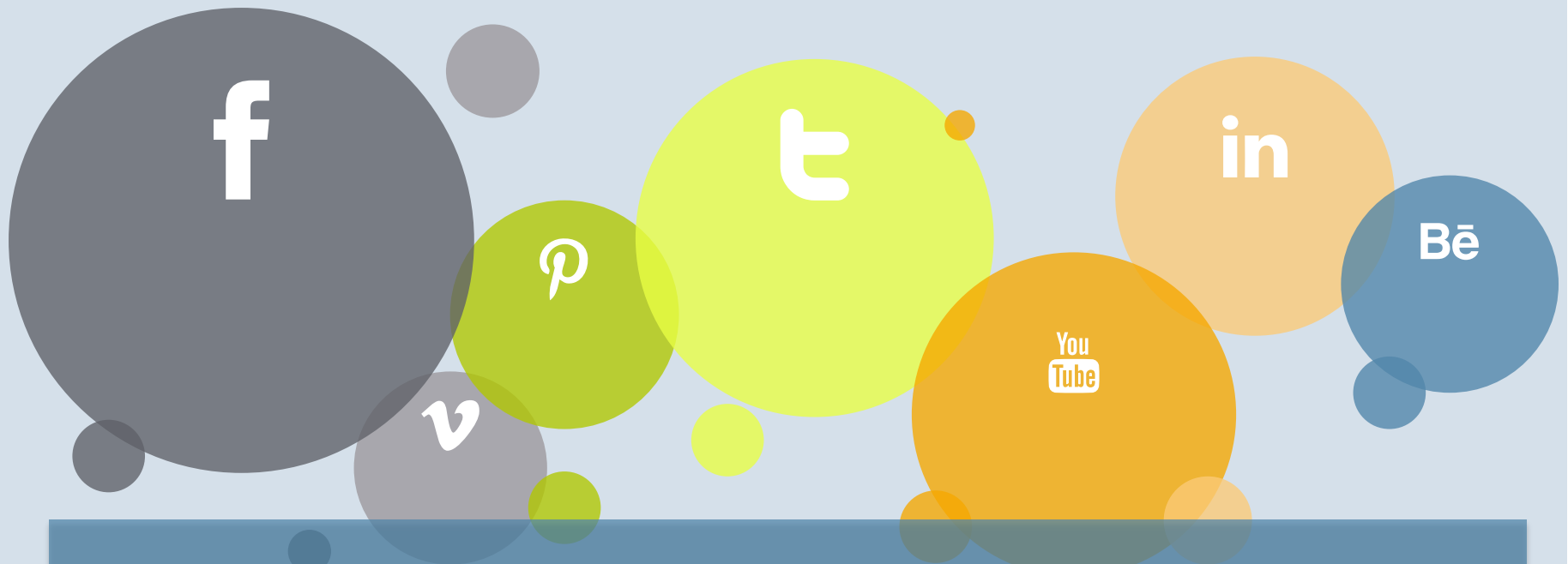

Picture: Mimisawhney

Modern-Day Panopticon

We Are Being Watched, Even if It Is Less Obvious

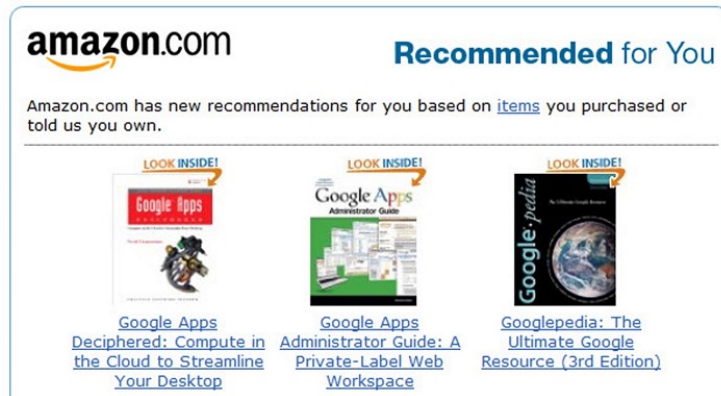**Is there no no need to care just because you don't do anything wrong?**

The Cambridge Analytica Scandal

- Huge amount of **data was stolen** from Facebook

- **Used by Cambridge Analytica** to target voters for Donald Trump

- Big drop in Facebook's share price plus **$5BN fine**

- **Responsibility** of Facebook?
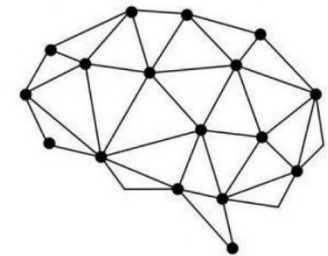
Personalized Advertising

Technology Is Eroding Privacy

" You have zero privacy any way …
Get over it. "

Scott McNealy, CEO of SUN, 1999

➡ Then everybody needs to be completely **honest**, right?

Changing Attitudes

As society adapts to new technologies, **attitudes will change**.

Teenagers freely post facts about their activities online that adults never would dream of doing. Their **privacy boundaries are different** from those of older people.

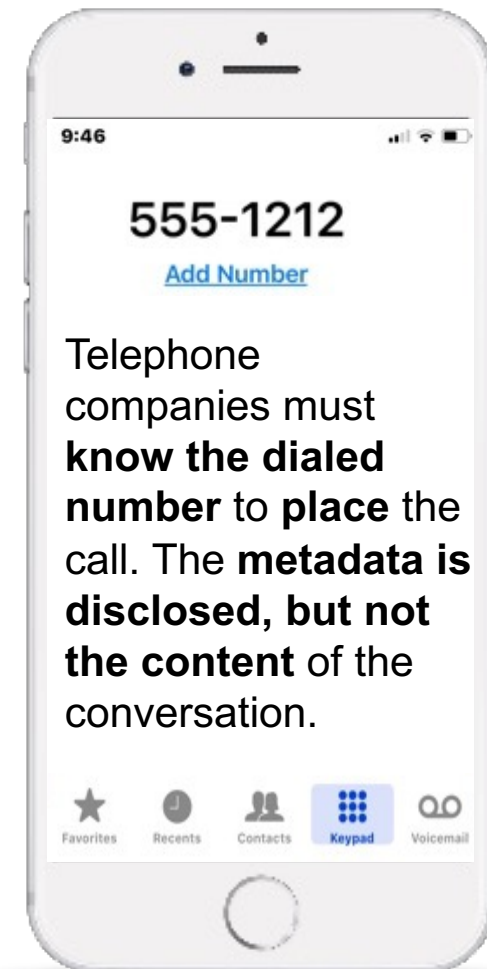But **different boundaries** does **not mean no boundaries**

Voluntary Disclosure

Anything you **voluntarily disclose** to others has much **less protection** than something you kept completely to yourself.

This quickly becomes problematic – due to **ubiquity**.

Example:

Telephone companies must **know the dialed number** to **place** the call. The **metadata is disclosed, but not the content** of the conversation.

It Takes a Village …

In small towns, there was **little privacy**. Everyone in town knew "everything" about everyone else.

Big cities provide **anonymity**.

Does **information technology** bring us **back** to the **"happy days"** in a small town?

No Option to Exit

In the **past**, one could get a fresh start by **moving to a new place** or by waiting till **memory fades**. **Reputation** could be **rebuilt** over time.

⬇

Big Data is universal and never forgets anything!

In a small town, **information** is (mostly) **symmetric**. You know as much about me as I know about you.

⬇

Data Science can result in major asymmetries.

Wayback Machine

It **archives** almost every **page on the web** that is **accessible** and (intends to) **retain for ever**. If you have an unflattering page written about you, it will survive for ever in an archive, even if it was taken down!

Right to be Forgotten

Laws are often written **to clear a person's record after some years**. For example, records about imprisonments or driving accidents are expunged after a defined number of years.

But what if an individuals' record is on the web?

Individuals' record

There are **laws in the EU and Argentina** since 2006. But in practice, these laws **primarily impact search engines**.

The World of Data Today

Every merchant, and many others, have **considerable data about me**. This includes at least my **interactions** with them, but often much more.

I am happy for them to **use** what they know to **serve my interests.**

But I really do not want them to **share the data with others** or **use** it **in other ways**.
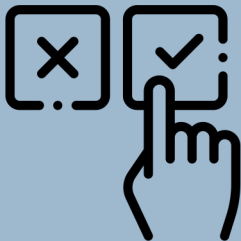
All Or Nothing Agreements

Many user agreements are **"all-or-nothing"**. Users must completely give up control on shared data to get any service benefits at all.

Users complain about **loss of privacy**.

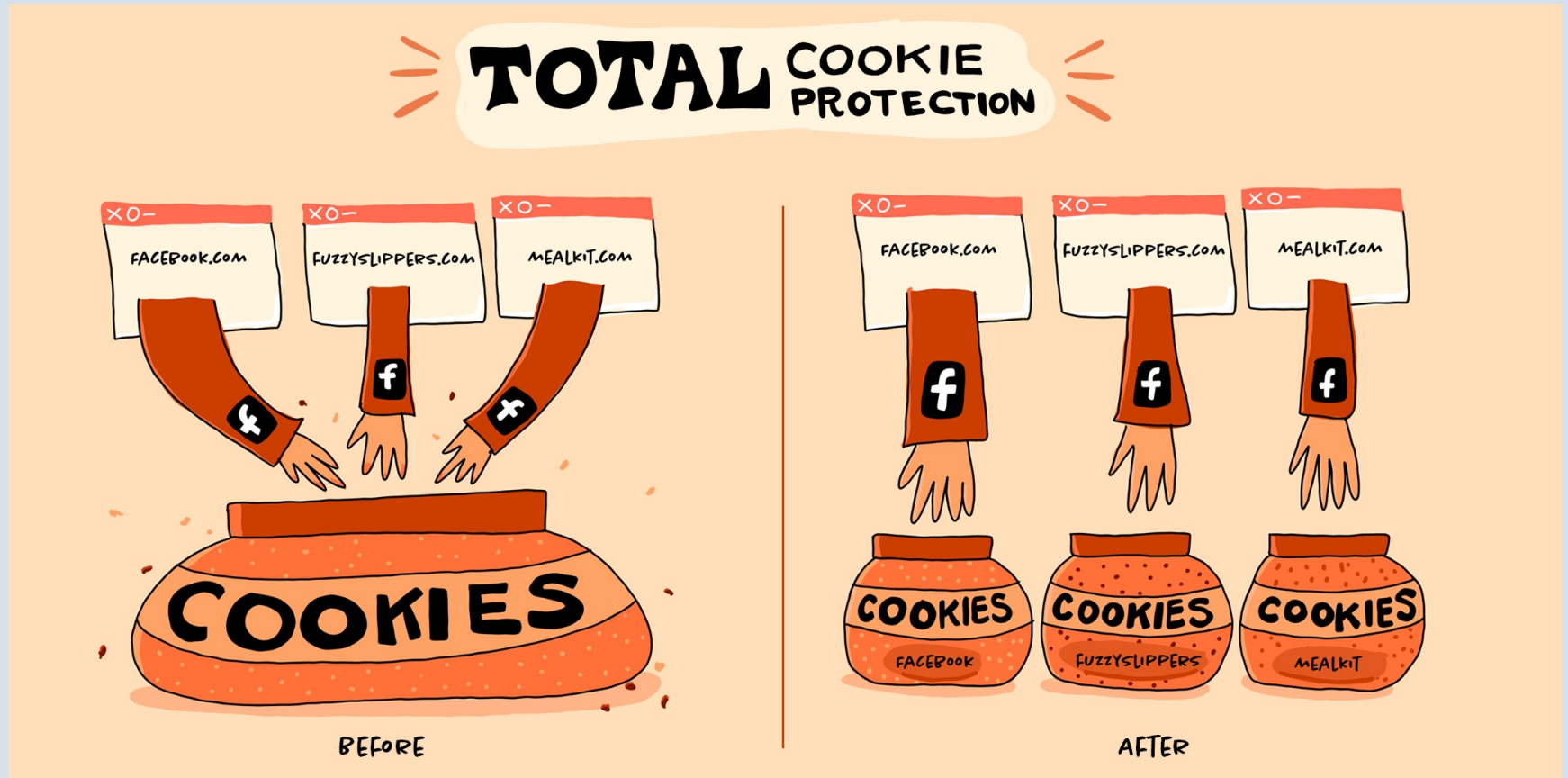It is better to **provide gradual choices**, so users can **make tradeoffs to their liking**, e.g., incognito browsing

Privacy
Example: Cookies

Picture: https://blog.mozilla.org/security/2021/06/01/total-cookie-protection-in-private-browsing/

Collection versus Use

Data Use

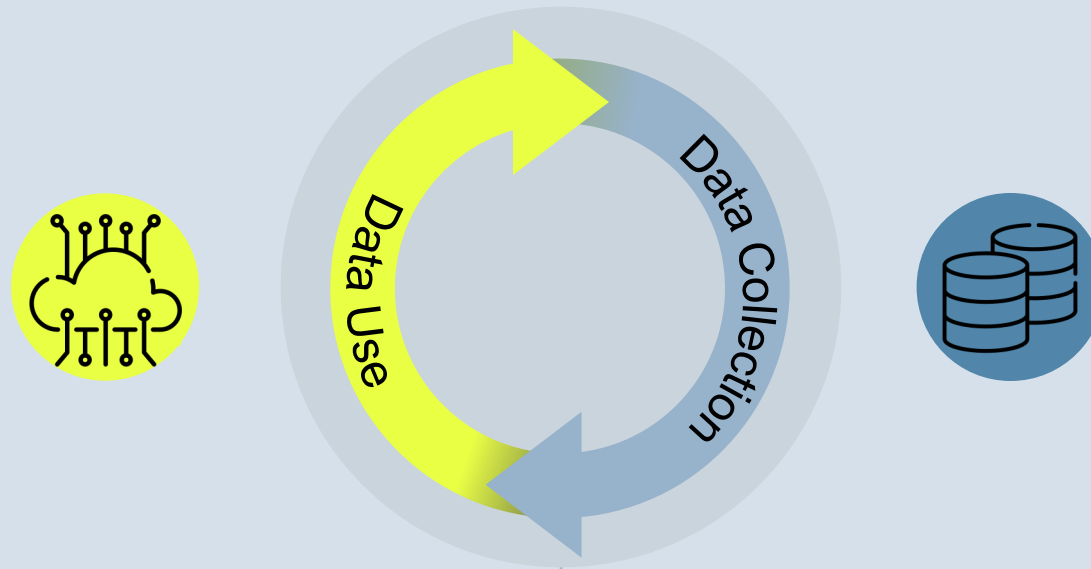Data Collection

Privacy is usually harmed only **upon use of data**.

Collection is a **necessary** first step **before use**.

But **collection without use** may sometimes be **right**, e.g., surveillance

Loss of Privacy

We have a **loss of privacy** due to a **loss of control over personal data.**

I am **OK** with there being **certain data** about me that I have chosen to share or that is public, but I really **do not want others to share** my data in ways that I **do not approve of**.

Main Drivers of Privacy Violation

Government agencies for national security, but also private enterprises, private investigators, …

Advertising

Prospective employee, prospective borrower, prospective date, …

Surveillance

If a company can show you focused, person-alized, and relevant ads everyone wins.

Introduction

Main Sources of Information

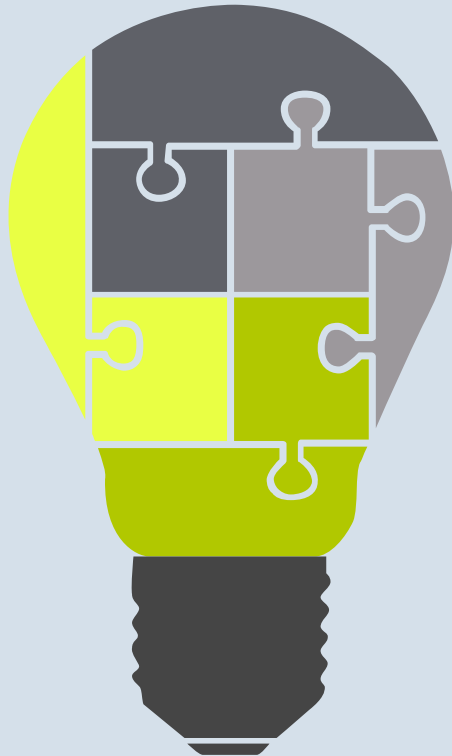Data collected by **merchants** and **service providers.**

Activity tracking – on and off the web.

Sensors in devices around us
- Personal devices
- Infrastructure
- Third party devices

Data Brokers

Companies that **aggregate and link information from multiple sources** to create more complete, and hence more valuable, information products.

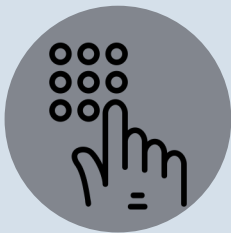Many don't realize **what can be learned about them** by linking multiple sources.

"Waste" Data Collection

Your ID is scanned at the bar to confirm **proof of age**. The scanning computer at the bar captures your **name, address, date-of-birth**, etc.

Metadata

Metadata is data about the data. It is often distinguished from data content.

E.g., for a phone call, metadata includes:
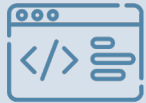
| Caller | Callee | Time and date of call | Duration |

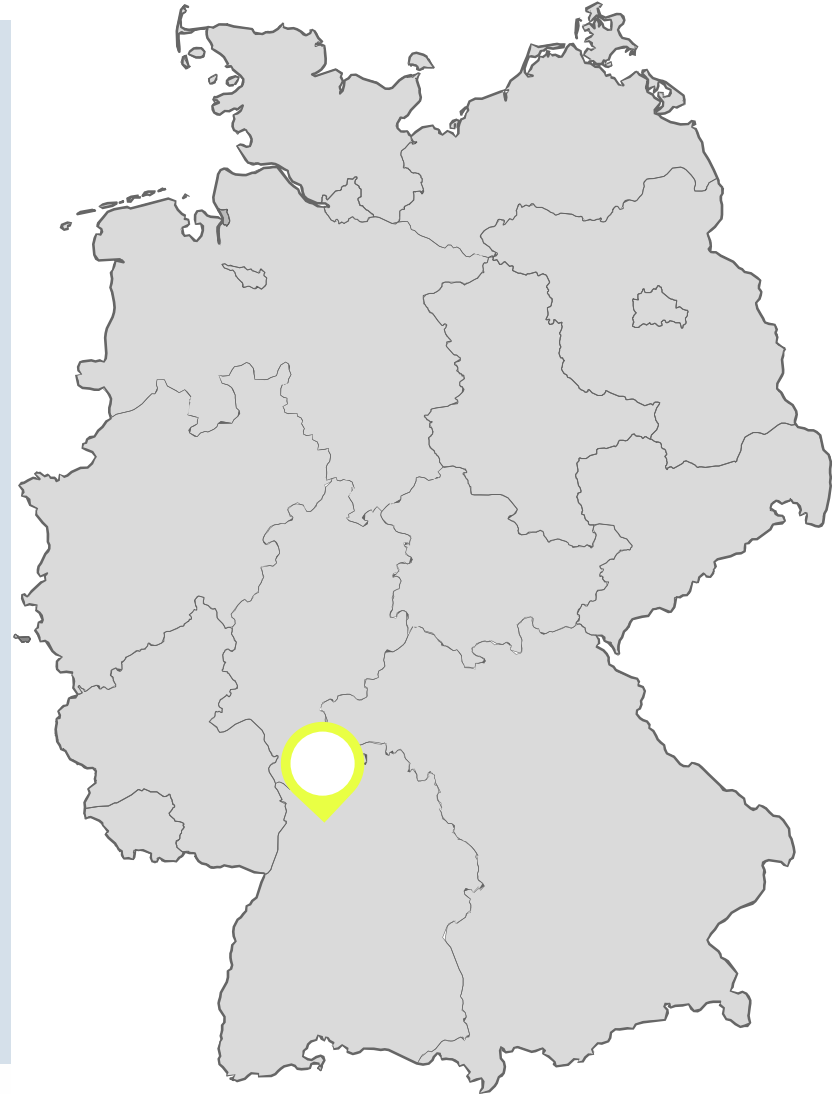# Metadata May Carry Much Information

Location information may be metadata for a cellphone call.

Knowing the location does not reveal the content of a call.

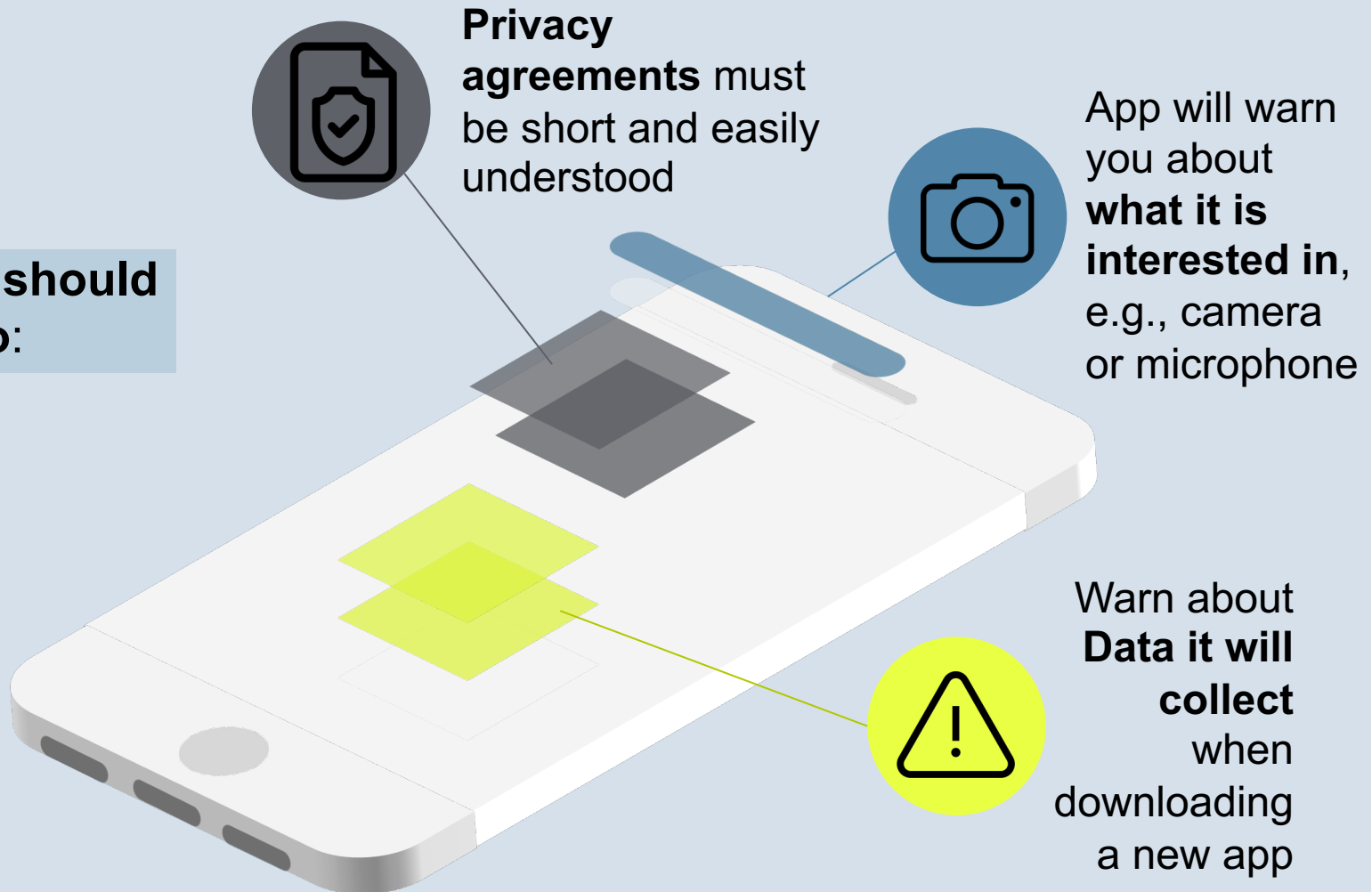But location tracking can reveal a great deal about the person.

Trust ≥ Design

Traditional social norms dealt with **privacy by trust**: You tell me private things because you trust me not to use what you told me in ways you would not approve of.
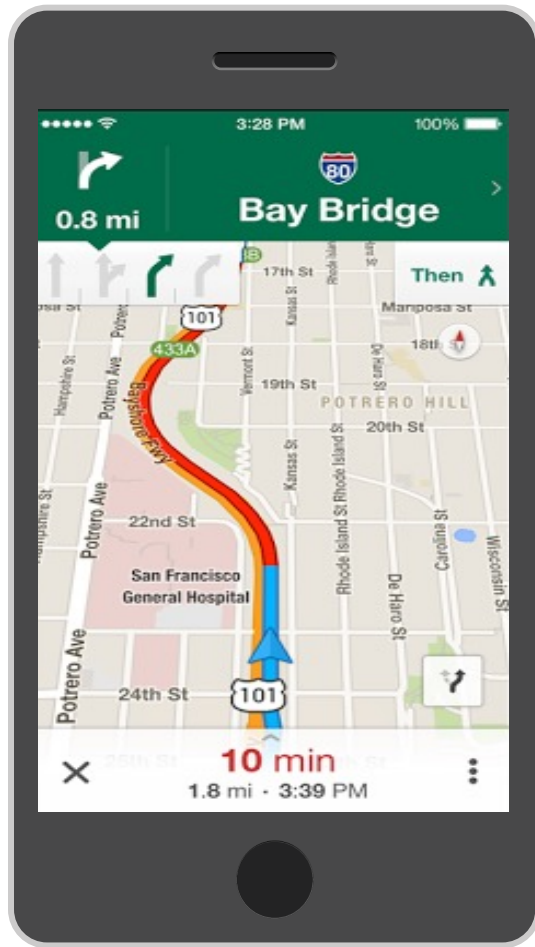
Modern data systems must deal with **privacy by design**. There are too many players in a complex system. Data sharing is contractual and not based on trust.

Sneaky Mobile App

Privacy agreements must be short and easily understood

App will warn you about **what it is interested in**, e.g., camera or microphone

What apps **should actually do**:

Warn about **Data it will collect** when downloading a new app

"Map App"

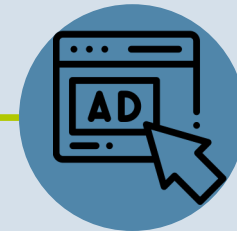Will **still function** even if you **do not share all necessary data** with them.

If you do not share your location, it still gives you the **simple route** or a **map of the neighborhood**.

Apps Ask For More Data Than Needed

Reasons why apps ask for more permissions than they actually need

Intentions for **future enhancements**

For **adware**

Tradeoff between getting an app for free and sharing the data with the app.

Conclusion

Privacy is the **ability to control sharing of information about the self**. It is a **basic human need**, even for people who have "nothing to hide".

Privacy is **easily eroded by thoughtless actions**, or in some cases, by intention.

Stakeholders have begun a **conversation around privacy**.

But there **is no consensus** on where the lines should be drawn just **yet**.