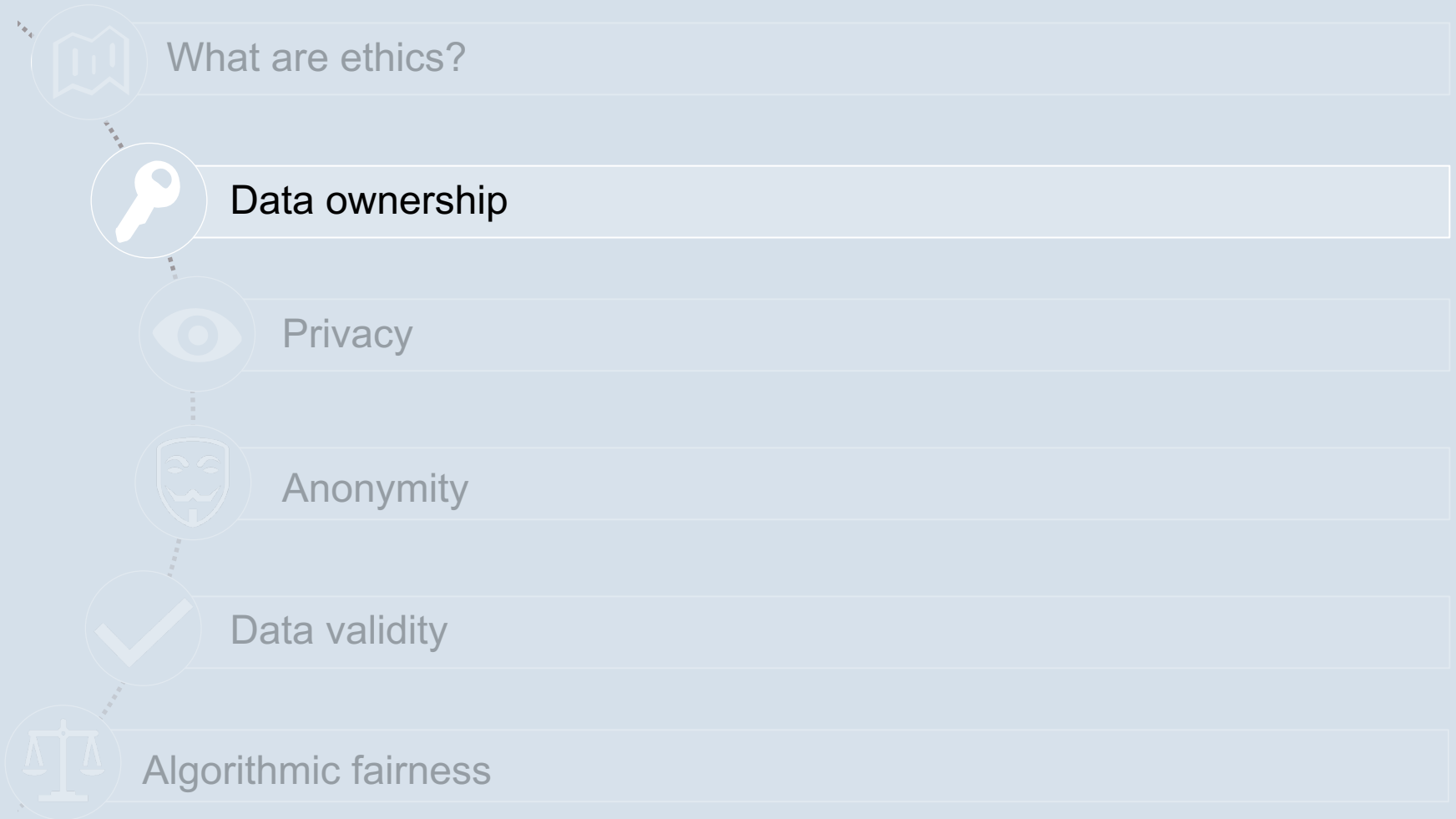




DATA ETHICS

PROF. DR. FLORIAN STAHL

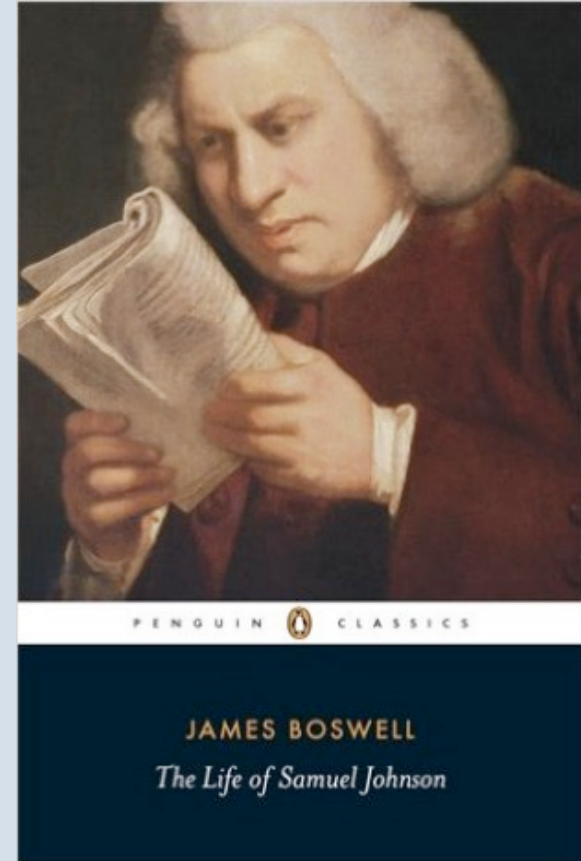
Overview – Data Ethics



Example: Biography Ownership

The biography **is about you**. But
is it yours?

If I write your biography, I own
the copyright. If you dislike
what I say, there is not much
you can do, except sue for libel
where I am inaccurate.



Example: „Nirvana Baby“

„**Baby model**“ sued the band:

- No **consent**?
 - Child pornography?
 - **Emotional harm**?
-
- „Nirvana Baby“
 - Did the baby model **benefit**?





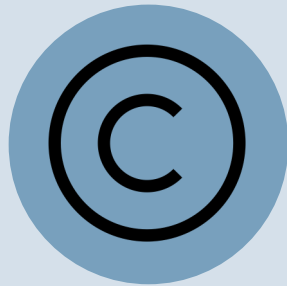
Similar limits on what data I can record about you and what I can do with it.



Free to record and free to use in other ways.



And we have done this forever: recommendation letters, gossip...



Copyright

Its an artistic expression. A rearrangement is a derivative work.



Patent

New idea for making or doing something, mostly of technical nature.



Trade Secret

I have it, but don't tell anyone.



Source: XY

If I use an image, I display it, and can **credit the owner**.



If I use your data, I am almost always **taking some piece** of what you know, **merging** it with what I know, and **expressing something new**.

- At best I can say I used some data from you.
- It is not easy to say exactly what and exactly how much

Data Collection and Curation

It can take a great deal of **effort to collect data** about something or someone.



Whoever does this work **deserves credit and has ownership** of the data “asset”.



It can often take even more **effort to clean, validate and standardize the collected data** to place it into a form that is of immediate value.



Data Collection and Curation

Example: Statista

The screenshot shows the Statista website interface. At the top, the Statista logo is on the left, and a search bar with the text 'Search Statistics' is on the right. Below the logo, a navigation bar contains links: 'Prices & Access' (highlighted with a red box), 'Statistics', 'Reports', 'Outlooks', 'Tools', 'Infographics', 'Services', 'Global Survey', and a 'NEW' badge. On the far right of the navigation bar are icons for email, a user profile, and a 'Login' link.

Below the navigation bar, the breadcrumb trail reads 'Technology & Telecommunications > IT Services'. The main title of the page is 'Big data and business analytics market share worldwide in 2021, by country'. To the right of the title, a 'PREMIUM +' badge is highlighted with a red box.

The main content area features a large pie chart. Overlaid on the chart is a dark blue box containing a promotional offer. The offer text includes: 'Exclusive Premium statistic', 'You need a **Single Account** for unlimited access.', a list of benefits (Full access to 1m statistics, Incl. source references, Available to download in PNG, PDF, XLS format), and a price comparison: '\$39' (in a red circle) vs '\$59 per month * in the first 12 months'. A green button labeled 'Get discount now' is also present. Above the price, a small white box says '33% discount until Oct 31st'. Below the price box, a link says 'View for free'. At the bottom of the offer box, a small grey box states '*Duration: 12 months, billed annually, single license'.

To the right of the offer box, there is a 'DOWNLOAD' section with icons for PDF, XLS, PNG, and PPT. Below this is a 'Source' section with links to 'Show sources information' and 'Show publisher information'. Further down are sections for 'Release date' (August 2021), 'Region' (Worldwide), and 'Survey time period' (2021).

Crowd-Sourced Reputation Sites



Limits on Recording

Recording is **wrong** when there is **reasonable expectation of privacy**, e.g., no cameras in clothing store fitting rooms.

Similarly: Phone companies must not record the (content of) phone calls.



We Could Agree Otherwise



- You could pose for a photograph under an agreement that I will give you **ownership of the photo**.
- You could agree to participate in a research experiment under an agreement that has **informed consent**.
- A funding agency could require that **data** collected from your funded research be **made public**.

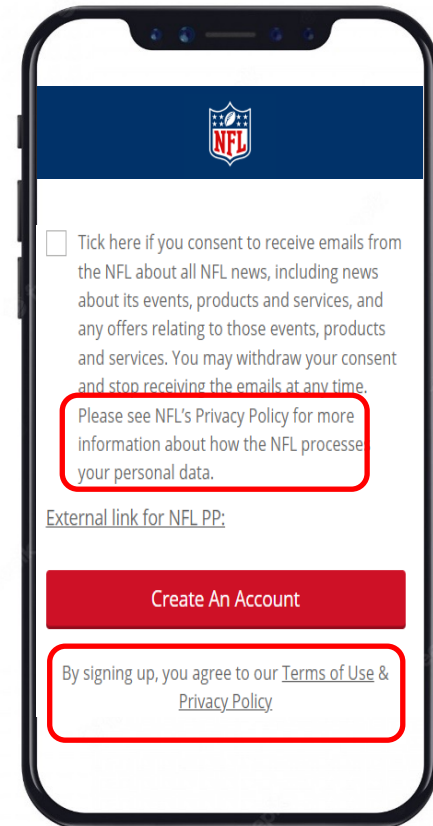
Example: Creating an Account



Your **choice/consent**.



Contract-bound.



Video Cameras in Stores



Can provide **security**.



Can even be used to **improve placement**.



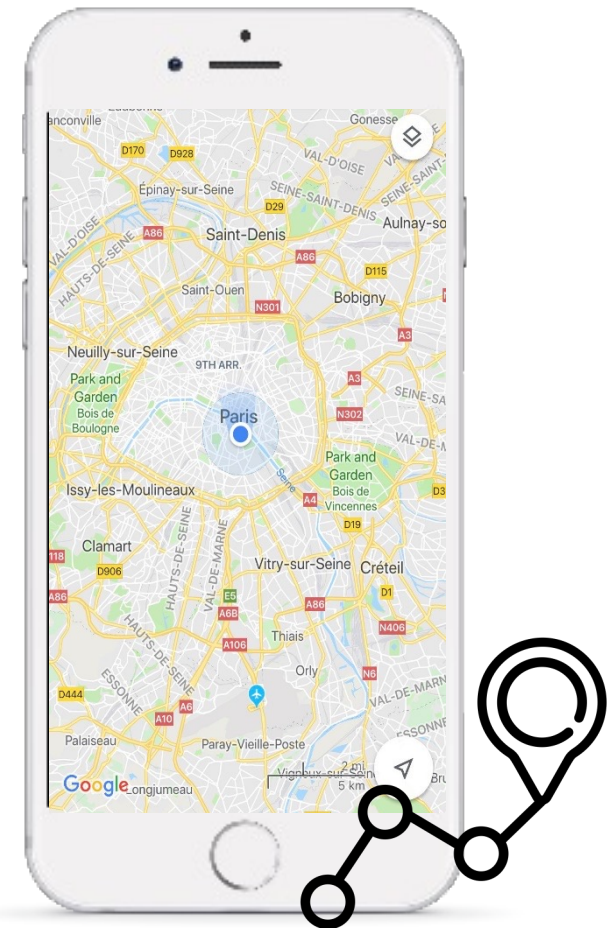
Recordings should **not** be **published**.



Cellphone Location Tracking

Necessary to **provide service**. Required for many valuable **applications**.

But can result in a **huge loss of privacy**.





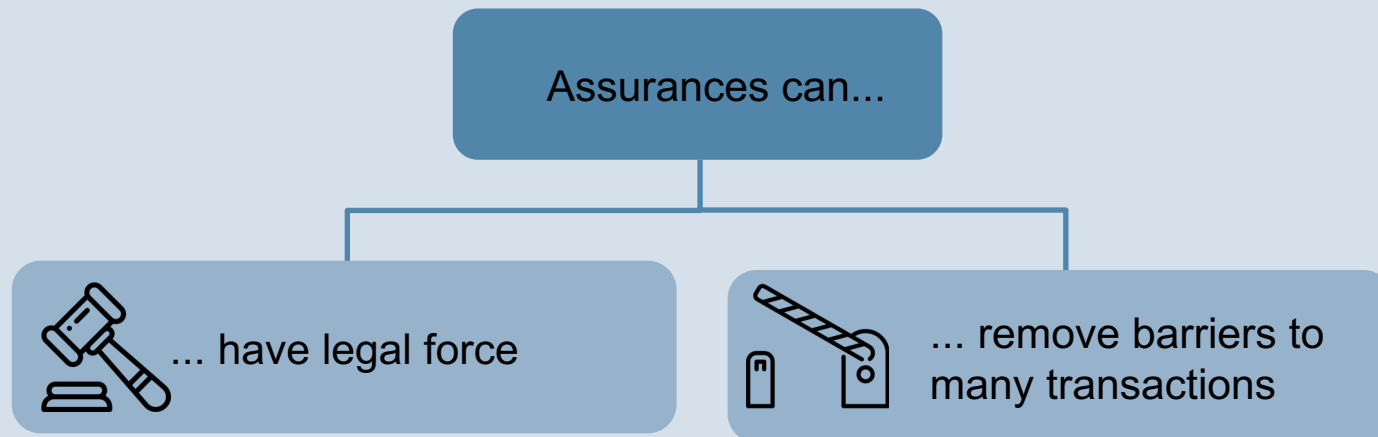
Often, there is a **strong**
reason to **record data...**



... but also **potential for**
misuse.

Prefer to **limit the use** rather than recording.
Allow desired **legitimate use** while disallowing other (undesirable) use.

Police reassure citizens that body-cam video will not be posted on the web. In the same way, **businesses** can **reassure** customers that **data collected for one purpose will not be used for another.**



Data Ownership Dispute Example: LinkedIn vs. hiQ Labs



LinkedIn users enter their personal data into their profile. The company **believes that it owns this data.**



In 2017, the startup “**hiQ Labs**” used data on public LinkedIn profiles to **predict whether people were likely to leave their jobs.** According to LinkedIn, this **violates privacy.**



US court ruled in favor of hiQ Labs because it considered scraping of publicly available data a **lawful business purpose.**



In June 2021, the **US Supreme Court vacated the decision** for a later review based on the Computer Fraud and Abuse Act. The **final decision is still pending.**



Companies legitimately collect data as part of doing business.



Companies need to retain goodwill of customers, and thus will try not to do shockingly bad things with the data.

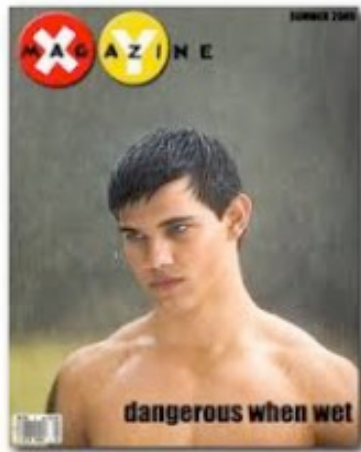


Once the company ceases to do business (e.g., because of bankruptcy), this data is an asset that is likely sold to a third party who may intend to misuse it.



Collected data must be destroyed, not sold.

Bankruptcy Law has Partial Protection



BORDERS®



Data ownership is **complex**, even for things you do want to share.



For the most part you do **not own data** about yourself. Therefore, there is a need to create **principles to control what isn't yours** and ensure the **rights to privacy**.