Data Protection & OSS in the Age of GDPR

by Cristina DeLisle
PerconaLive 2019
Amsterdam 2019

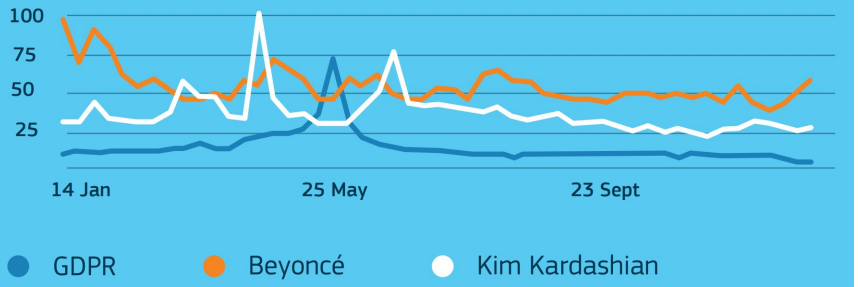
This talk is...

- ➔ **Not legal advice** for your particular situation
- ➔ **Sacrificing legal correctness** in order to be more common sense
- ➔ **Providing a basic understanding** of what you need to think about as a data controller who operates a database

How many of you didn’t hear about GDPR? What about Directive 95/46/EC, “Data protection directive”

Google searches

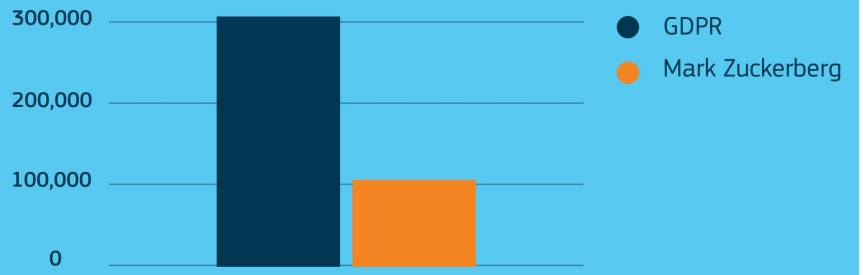
During the peak month of May 2018 GDPR was searched more often on Google than American superstars Beyoncé and Kim Kardashian.



Interest rated between 0-100, based on number of searches on Google.
Source: Google trends

Media coverage

GDPR received a lot of attention in **2018**. So much than even some celebrities had to stand in its shadow.



Annual worldwide mentions in the media
Source: Factiva

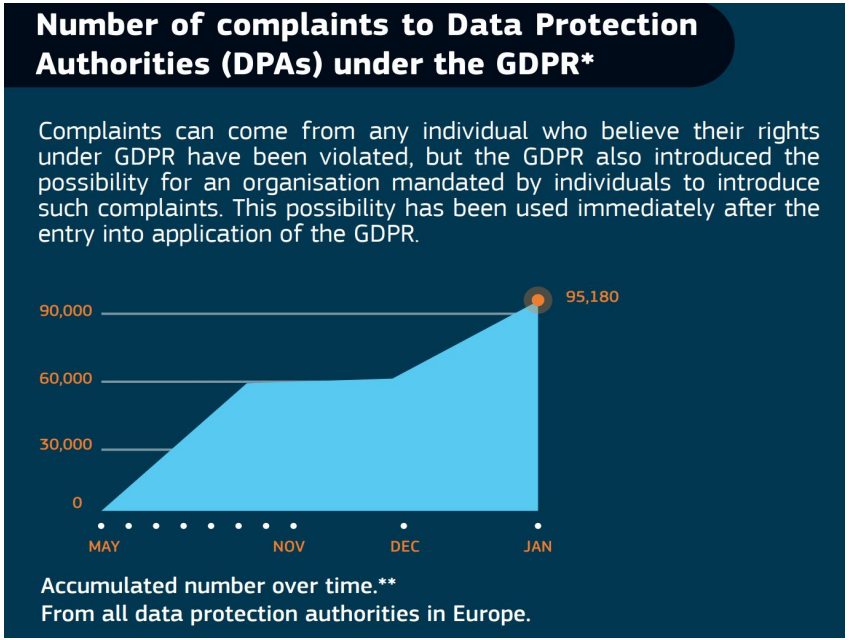
Transversal impacts of the GDPR

- **Legal and compliance governance:** privacy strategies, accountability, lawfulness, policy making, auditing
- **Data collection and lifecycle:** purpose limitation, data minimization, transparency
- **Tech:** data breaches handling, encryption solutions, privacy by design & default

Areas of biggest fines so far

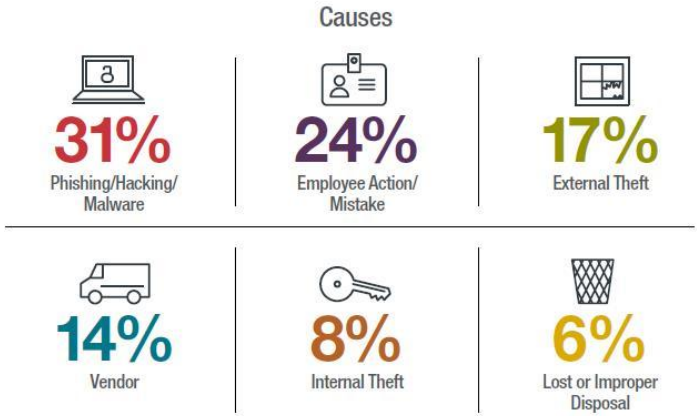
- **Coerced consent** from data subjects - most common complaints:
  - Telemarketing
  - Promotional emails
- **Data security areas:**
  - leaks, breaches of confidentiality, availability, integrity
- **Video surveillance/ CCTV**

European Commission infographics



Some oldie but goodie statistics

BakerHostetler 2016 Data Security Incident Response Report



Verizon 2014 Data Breach Investigations Report

63,437 security incidents[1]

1,367 confirmed data breaches[1]

1 in **3** documented data breaches occurred in businesses with under 100 employees[2]

60% of small businesses close their doors within half -a-year of being victimized by a "cybercrime"[3]

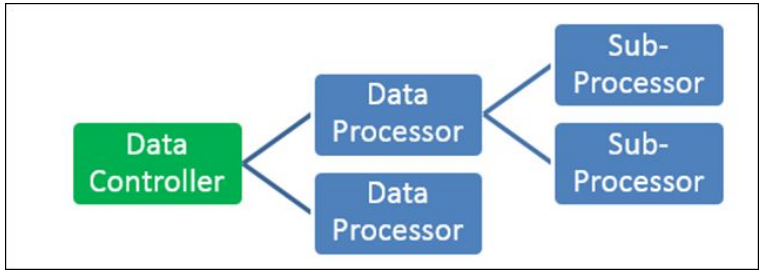
What is the GDPR in practice

- PEOPLE CAN’T JUST SUE YOU - it's investigation based
- Vaguely written law (that's intentional)
  - Meant for general purpose, all sectors of businesses
  - Establishes supervisory authorities who investigate and issue guidance

You can talk to your supervisory authority, their objective is to help you protect personal data! (not customer service)

The model of controllers & processors

- **Controller:**
  - determines the purpose and means of processing



- **Processor:**
  - third party that processes it on a controller's behalf
- **Data processor agreement (DPA)**
  - You can act as a controller & processor at the same time, depending on how the personal data gets handled

Data controllers & processors

- **2012**: Google Inc. as a controller, under Directive 95/46/EC, “Data protection directive”

ECJ on Google Sp & Google Inc vs. Mr. Gonzales

- **By 2016**: Google received 347,533 separate requests to remove aprox. 1.2 million websites

Google - responsible for the processing that it carries out of personal information which appears on web pages published by third parties

The OSS model

- **The OSS community**
  - Data subjects
  - Enforced rights on their personal data

- **The “infrastructure providers”**
  - Controllers & Processors
  - Ex.: Github
    - Controller of the PD from your free private user account
    - Processor of your invoices

General obligations of a data controller

- You have to report "serious" data breaches
- When you collect a piece of data, you need to keep track of why you did that
  - consent -> the data-subject is ok with you collecting it
  - contract -> you have a contract with the data-subject
  - legal obligation -> AML/KYC, invoices
  - legitimate interest -> technical logs, IP addresses
- You need to have a privacy policy where you specify the data lifecycle for different types of data
- When a piece of data is no longer needed and will be removed

5 major requests a data-subject can legally make*

- ➔ What data do you have on me ?
- ➔ Who else did you give my data to ?
- ➔ Please delete what you have on me
- ➔ This thing about me is incorrect, please correct it
- ➔ Let me download my data

*not exhaustive

GDPR as it applies to a database

- Need to **know** how you came to have a particular piece of data
- Ability to **delete** things
- Ability to **find** all of the things related to particular person
- **Automate** deletion in order to fulfill data lifecycle

Tips for schema design

- ➔ When you collect personal data, you should create a data_collection_event with
  - ◆ The date it happened
  - ◆ Some way to identify the data-subject (if you know)
  - ◆ The reason for collection: (consent, contract, legal obligation, legitimate interest)
- ➔ Every piece of related data should contain the ID of the related data_collection_event
- ➔ When you copy data into another database, or another table or whatever, copy the data_collection_event ID

What about the backups?

- ➔ Supervisory authorities understand technical limitations:
  - They're not going to throw the book at you for being unable to delete everything immediately
  - But this is not a free pass, you have to be trying as hard as you can
  - You have to be clear to the data-subjects exactly what is happening
  - Put the backup data ‘beyond use’, even if it cannot be immediately overwritten (ICO)
- ➔ One option is simply to rotate backups often
- ➔ Another (interesting) option would be to encrypt the individual rows in the backup using a per-data_collection_event key
  - When you have a deletion request OR when that data_collection_event ends its life cycle, you can delete the key

Feel free to contact me!

- @cristina.r:matrix.org
- @redchrision@mastodon.social
- https://www.linkedin.com/in/cristina-delisle-10848029/