GDPR for Data People

PyData Bristol Meetup
17 May 2018

Ed Boal
Technology & Data Protection Lawyer

@edboal

What we're going to cover

- Brief introduction to the GDPR

- Why do we need the GDPR *now*?

- Do people *really* care about privacy?

- What is the GDPR *really* about?

- Big Data v. GDPR

- What does the GDPR mean for data science?

- Towards a new Data Ethics

Brief introduction to the GDPR

- It's a <u>Regulation</u> not a Directive
- Central themes of accountability and transparency
- Greater territorial reach to "level the playing field"
- Calls out location data, online identifiers, biometric and genetic data as types of personal data
- Greater penalties for non-compliance
- But it's an *evolution* rather than a *revolution*
- Brexit won't make a difference

Why do we need the GDPR *now*?

1995

- Facebook? Google?
- 16 million internet users
- 23,500 sites

Why do we need the GDPR *now*?

2018

- Facebook: 1.45bn D.A.U.

- 3.5bn Google searches/day

- 4.02bn internet users

- 1.87bn sites (<200m active)

1,508,594 floppy disks!

Why do we need the GDPR *now*?

2025?

163 Zettabytes (1tn Gigabytes)

10x data created in 2016

25% of data will be *real-time*

4,800 device interactions per day

Source: Data Age 2025, IDC White Paper (2017)

Why do we need the GDPR *now*?

Data is a valuable commodity…

$$\frac{\$19,000,000,000}{450,000,000 \text{ MAU*}} = \$42/\text{user}$$

* 1.3bn MAU as at the end of Q4 2017

Why do we need the GDPR *now*?

And there's this…

Credit: Book Catalog via flickr.com

Credit: Anthony Quintano via flickr.com

Do people *really* care about privacy?

- Privacy is a fundamental and inviolable human right

- But there is a 'privacy paradox'

- <20% UK adults trust organisations with their data with only 1/10 trusting social media platforms*

* Source: ICO/ComRes Trust & Confidence in Data Survey of 2,153 adults (Nov 2017)

Do people *really* care about privacy?

"Nobody needs to just why they 'need' a right: the burden of justification falls on the one seeking to infringe upon the right….Arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say."

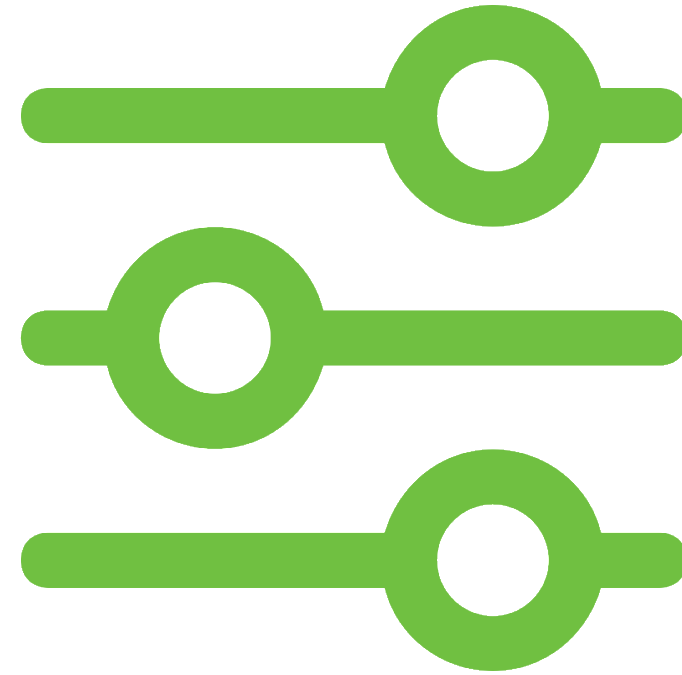– *Edward Snowden* (Reddit comment)

What is the GDPR *really* about?
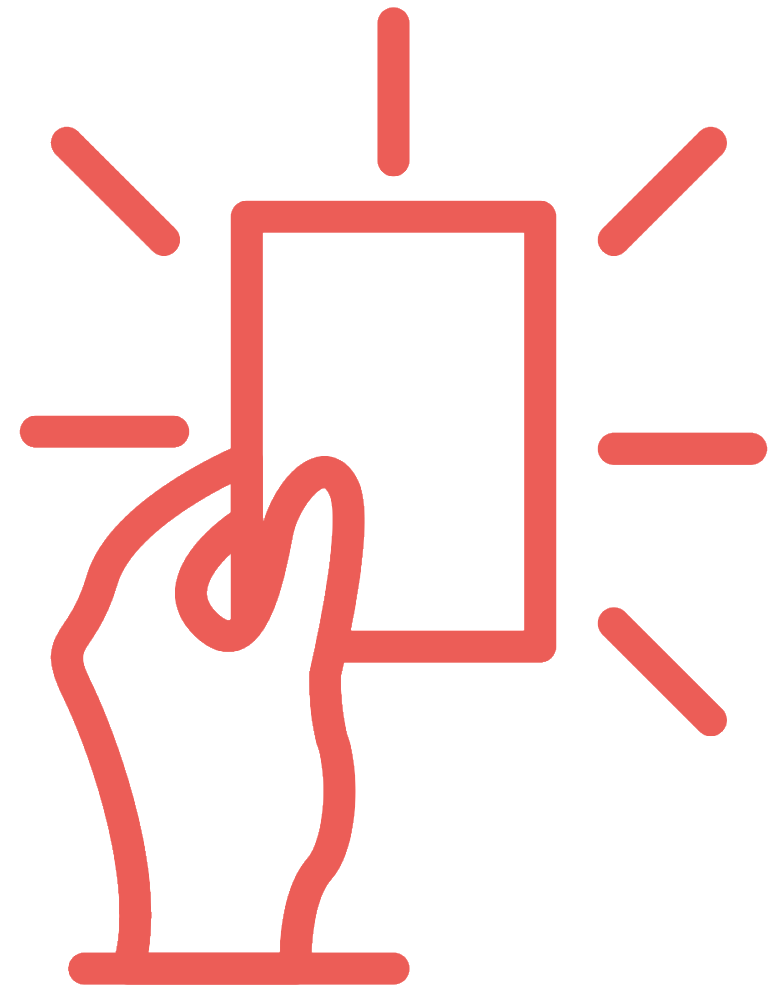
Penalties? **Stifling innovation?** **Reputation?**
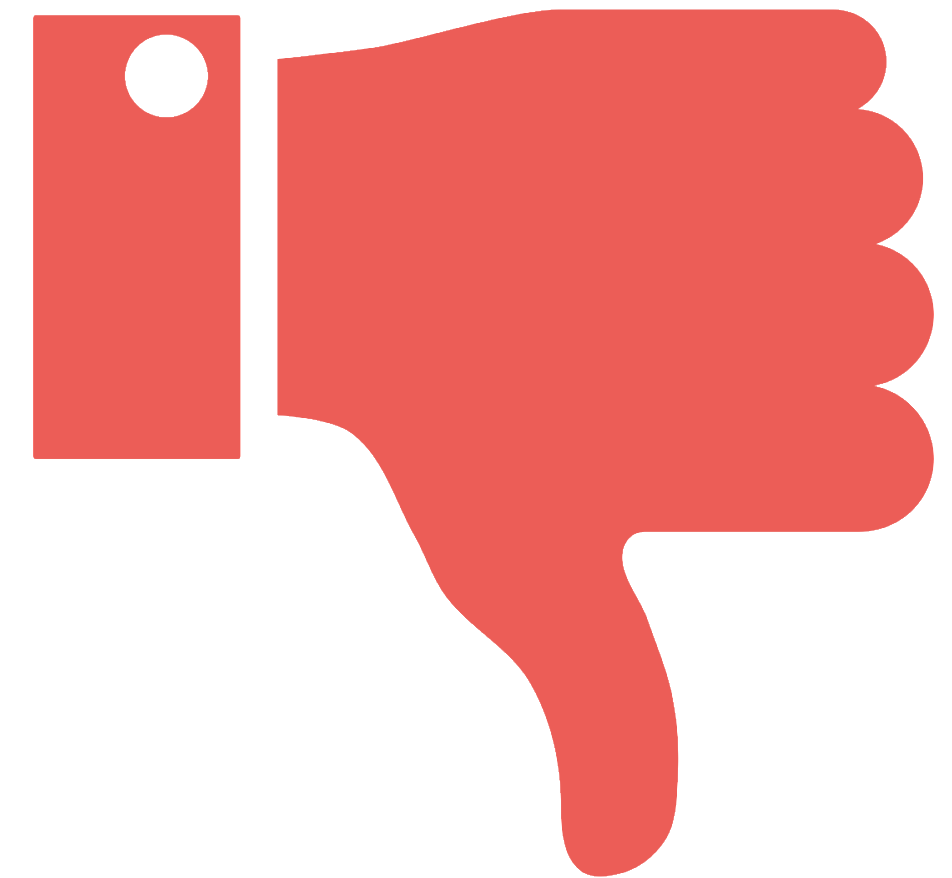
What is the GDPR *really* about?

Respect

Control

Trust

GDPR vs. Big Data

The popular definition of 'Big Data'…

Laney's 'three V's': volume, velocity, variety*

My definition of 'Big Data'…

"Collecting more data than you [think you] need, because you can" (*n*=ALL**)

* Source: Doug Laney/Meta Group Inc '3D Data Management' (Feb 2001)

** Source: Mayer-Schönberger/Cukier *'Big Data'* (2013)

GDPR vs. Big Data

Data Protection Act and General Data Protection Regulation

Big data, artificial intelligence, machine learning and data protection

ico.
Information Commissioner's Office

"…many instances of big data analytics do not involve personal data at all…"

ICO Guidance

GDPR vs. Big Data

Round One: Purpose limitation

"Personal data shall be collected for **specified**, **explicit** and **legitimate purposes** and n**ot further processed** in a manner that is i**ncompatible with those purposes**"

Article 5(1)(b), GDPR

GDPR vs. Big Data

Round One: Purpose limitation

- Value of Big Data often in secondary not primary purposes
- Secondary purposes must be compatible taking into account:
  - nature of personal data and linkage between purposes
  - relationship between controller and data subject
  - possible consequences of further processing for data subjects
  - appropriate safeguards (e.g. encryption or pseudonymisation)

GDPR vs. Big Data

Round Two: Notice and consent

"Personal data shall be processed **lawfully**, **fairly** and in a **transparent** manner in relation to the data subject"

Article 5(1)(a), GDPR

GDPR vs. Big Data

Round Two: Notice and consent

- Data subjects should not be surprised by any processing: about expectations <u>and</u> effects

- Data subjects must be informed about what data is being collected, why it is being collected and what the lawful basis for processing is (privacy notices)

- Notices must be concise, intelligible and written in clear and plain language (eek!)

- What about 'legitimate interests'?

GDPR vs. Big Data

Round Three: Profiling and automated decisions

"any form of **automated processing** of personal data consisting of the use of personal data to **evaluate certain personal aspects** relating to [an individual], in particular to **analyse or predict aspects**…"

Article 4(4), GDPR

GDPR vs. Big Data

Round Three: Profiling and automated decisions

"[Individuals] have the right not to be subject to a decision based on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her"

Article 22(1), GDPR

GDPR vs. Big Data

Round Three: Profiling and automated decisions

- Only prohibited if it results in significant decisions
- Lawful basis for processing still needed
- Must use "appropriate mathematical or statistical procedures for profiling"
- Data inaccuracies should be corrected, the risk of errors should be minimised and the risk of discriminatory effects mitigated (Recital 71, GDPR)

GDPR vs. Big Data

Further reading

- ICO, [Big Data, AI, ML and data protection guidance](#)
- UKAN [Anonymisation Decision-making Framework](#)
- Article 29 WP [Statement on impact of Big Data](#)
- Article 29 WP [Guidelines on automated decision-making and profiling](#)
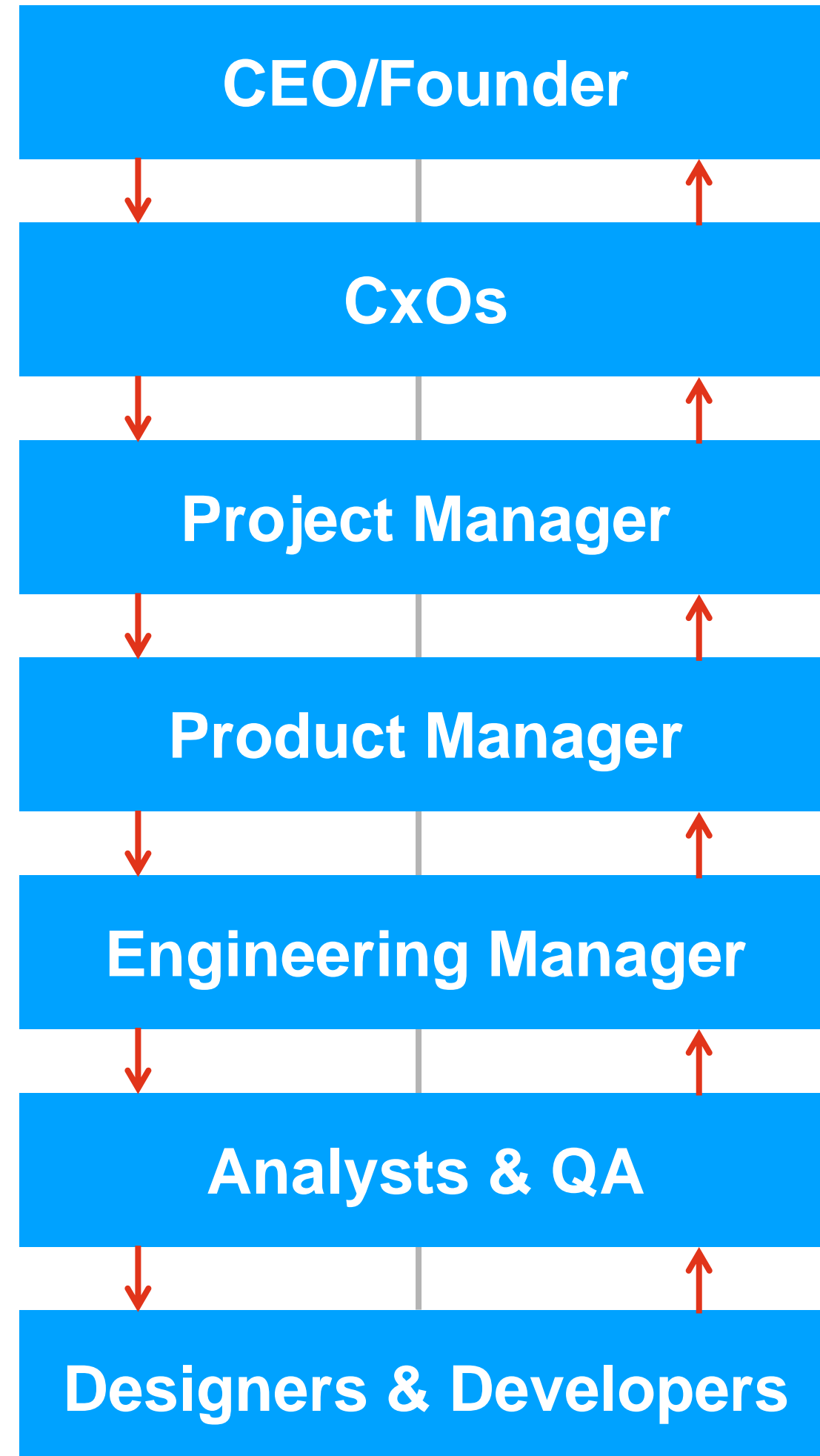
What does the GDPR mean for data science?

What does the GDPR mean for data science?

- Conflict of interests and objectives

- Need a standardised way to document and explain model logic in a non-technical way

- Technical challenges in eliminating bias/discrimination

- ML can be a problem *and* solution

- Data Protection by Design, Data Protection by Default

Towards a new Data Ethics

- Data is innocent – collection and interrogation *can* be constrained

- Risks of confirmation bias and "naïve realism"

- GDPR as an ethical framework rather than a threat framework

- Ethics requires a simple pro-active framework

Command-line ethics?

CEO/Founder

CxOs

Project Manager

Product Manager

Engineering Manager

Analysts & QA

Designers & Developers

- Dialogue between each level

- Accountability between each level

- A 'silver ethical thread' from top-to-bottom

- Becomes part of agile frameworks

- Culturally embedded

- Promotes Data Protection by Design