UMC Utrecht
Julius Center

Big Data in Research, Privacy and Data Protection

M. (Menno) Mostert, LLM, PhD Candidate

Julius Center, University Medical Center Utrecht
Email: m.mostert-2@umcutrecht.nl

Warming Up

Mark Zuckerberg:
"The age of privacy is over"

Privacy Is Completely And Utterly Dead, And We Killed It

08/20/2014 08:27 am ET | Updated Aug 20, 2014

Jacob Morgan
Forbes

Proposition:

Privacy is dead

Privacy is not dead

- *" (…) if we think about privacy as the question of what rules should govern the use of personal information, then privacy has never been more alive.*
- *In fact, it is perhaps the most important and most vital issue we face as a society today."**

Value of privacy?

- Important to individuals (and society)
  - Identity, relationships, sense of personal well-being, freedom

Objective of personal data protection?

- Respect for individual rights and interests
- Social and economical progress (including scientific progress)

(*Richards, Neil M., Four Privacy Myths (April 22, 2014). Available at SSRN: http://ssrn.com/abstract=2427808)

3. Privacy

Stelling

People with nothing to hide have nothing to fear

(Richards, 2014)

3. Privacy

The 'nothing to hide' argument is false:

1. *"(..) all of us have "something to hide"*
2. *"reducing privacy to an individual right to hide dark secrets ignores the power effects of privacy."*
3. *"it reduces privacy to an individual's right to hide big secrets. Such a crude reduction of the issue ignores both the complexity of privacy, as well as the social value that comes from living in a society that not everything about us is publicly available all of the time." (…)*

   Privacy is thus essential for individuality and self-determination, with substantial benefits for society."

(Richards, 2014; Cohen, 2013)

Koppie Koppie®

Iemands kind op uw favoriete mok

Dreamy Daughter
Jan de Graaf
€ 15.95

First day of school
Bruno de Regge
€ 16.95

Brother & Sisters
Remy Snippe
€ 15.95

Three poor kids, locked up in a cold dungeon
Eddy Van 3000
€ 15.95

Volkskrant:

China kent elke burger score toe - ook voor internetgedrag

China's Sociaal Krediet Systeem

ARTIKEL 'Het is wat Amazon doet, maar met een orwelliaanse politieke draai eraan.' China werkt aan een systeem om alle burgers een 'sociale kredietcode' te geven, mede gebaseerd op hun gedragingen op internet.

Door: Michael Persson, Marije Vlaskamp, Fokke Obbema 25 april 2015, 12:00

(...)

> Wie gelooft in socialistische kernwaarden, gedraagt zich beter
>
> — Professor Wang Shuqin, een onderzoeker die meewerkt aan de ontwikkeling van het systeem

Introduction

What is the right to 'privacy'?

- European Court of Human Rights
  - *"Although the notion of private life is not considered susceptible for exhaustive definition, it is "encompassing the physical, psychological and moral aspects of the personal integrity, identity and autonomy of individuals."*

- EU Court of Justice
  - the right to privacy is directly and specifically affected, when the *"data, taken as a whole, may allow very precise conclusions to be drawn concerning the private lives of the persons whose data has been retained,*
  - *such as the habits of everyday life, permanent or temporary places of residence, daily or other movements, the activities carried out, the social relationships of those persons and the social environments frequented by them."*

Introduction

How is Big Data in Research regulated?

- Fundamental rights
  - Privacy, data protection, non-discrimination, etc.
- Regulations
  - International treaties
  - General Data Protection Regulation (GDPR)
  - National laws
- Non-binding guidelines
  - Declaration of Taipei *(health research)*
  - CIOMS Guidelines *(health research)*

Introduction

BMJ 2013;346:f3534 doi: 10.1136/bmj.f3534 (Published 31 May 2013)

Proposed EU data protection regulation is a threat to medical research

A suggested amendment would make most epidemiological and health research impossible

M C Ploem *associate professor of health law* , M L Essink-Bot *professor of social medicine*, K Stronks *professor of social medicine*

Introduction

THE LANCET

A step forward for data protection and biomedical research

To update previous correspondence to this journal,[1] we are pleased to report that, in large part through the efforts of the biomedical research community, a European General Data Protection Regulation that is favourable for research was agreed by Member States and Parliament in December, 2015.[2] Although the Regulation will probably not apply until mid-2018, now is an appropriate time to highlight the implications for biomedical research.

DOI: http://dx.doi.org/10.1016/S0140-6736(16)30078-2

Introduction

Key rule

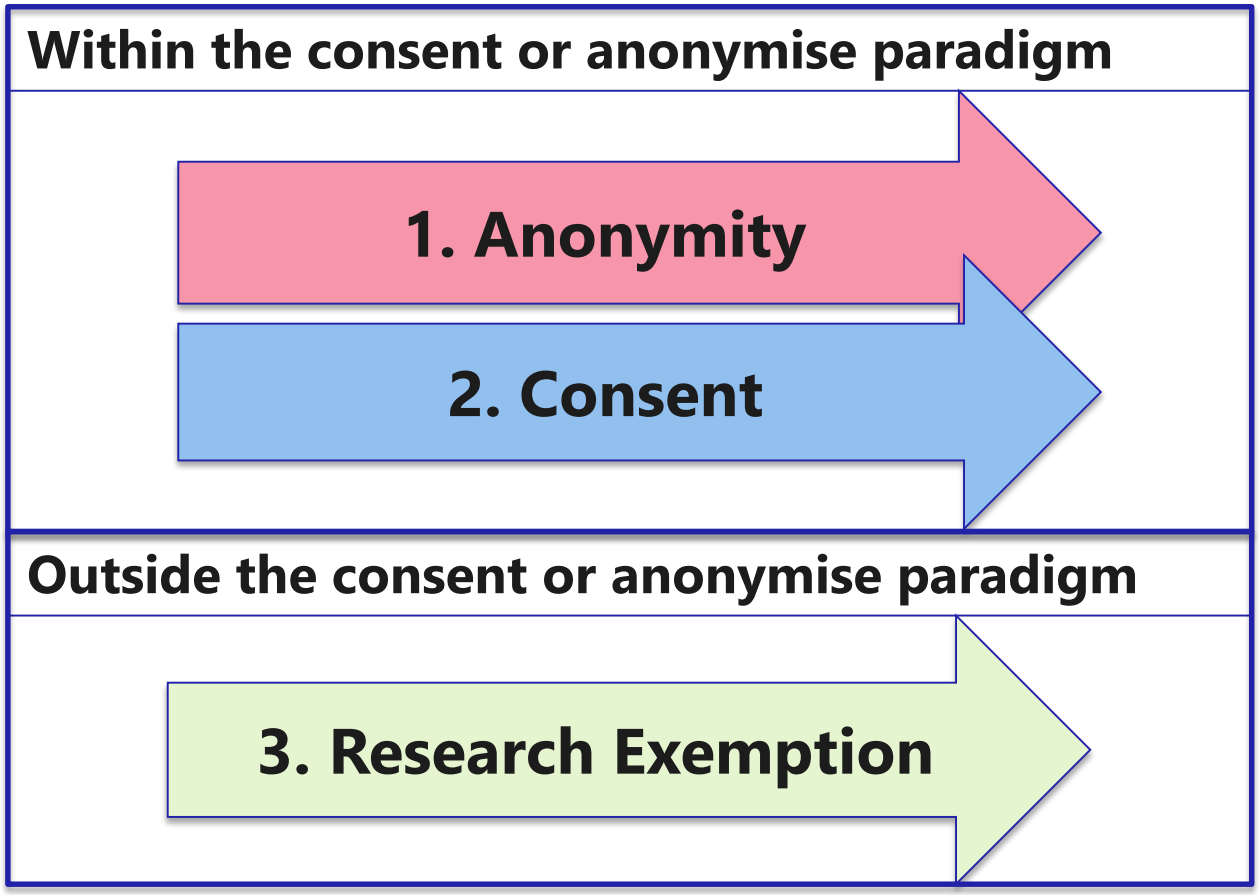
General Data Protection Regulation (GDPR) (2016/679/EU)

- Article 9
  - Processing of **special categories of personal data** shall be **prohibited**. (e.g. health-related and genetic data)
  - This prohibition does not apply
    a) if the data subject has given **explicit consent**
    b) if an appropriate **research exemption** from this prohibition is laid down in national (or EU) law

Introduction

Main question

- How is the consent or anonymise approach challenged in Big Data health research? Appropriate ways forward?



Introduction

European Journal of Human Genetics (2016) 24, 956–960

www.nature.com/ejhg

POLICY

Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach

Menno Mostert*, Annelien L Bredenoord, Monique CIH Biesaart and Johannes JM van Delden

1. Anonymity

1. Anonymity

Technical perspective

- Might be impossible to guarantee absolute anonymity

Legal perspective

→ Next slide

http://www.technologyreview.com/news/514351/has-big-data-made-anonymity-impossible/

1. Anonymity

GDPR (recital 26)

- To determine whether a natural person is identifiable,
  - account should be taken of **all the means reasonably likely to be used**, **such as singling out**,
  - either by the controller **or by another person**
  - to identify the natural person **directly or indirectly**

- Means reasonably likely to be used?
  - account should be taken of all objective factors, such as the costs of and the amount of time required for identification,
  - taking into consideration the available technology at the time of the processing and technological developments

Discussion: does my dataset contain identifiable data?

1. Anonymity

Irreversible anonymisation (conform legal standards)

- Implicates extensive stripping, reduces data quality
- Excludes data linkage

Ways forward?

Regarding two-way coded data as de-identified

- Pseudonymisation not a method of anonymisation
- GDPR, Recital 28: *pseudonymisation can reduce risks, but does not preclude the applicability of data protection law*

Anonymisation should be avoided

No obligations to protect data or interests, while risks remain :

- Risk of re-identification
- Group risks, discrimination or stigmatisation

(References in: Mostert et al, 2016)

1. Anonymity

Key messages

I. Anonymisation is, on its own, an unsafe strategy for ensuring protection of individual rights and interests

II. In addition, anonymisation is difficult in practice without compromising the utility of the data set. In Big Data research this problem increases

III. Pseudonymisation is a useful method to reduce privacy risks, but it does not preclude the applicability of the law

2. Consent

2. Consent

BMJ 2013;346:f3534 doi: 10.1136/bmj.f3534 (Published 31 May 2013)

Proposed EU data protection regulation is a threat to medical research

A suggested amendment would make most epidemiological and health research impossible

M C Ploem *associate professor of health law* , M L Essink-Bot *professor of social medicine*, K Stronks *professor of social medicine*

'"Processing of personal data concerning health which is necessary for (..) scientific research purposes, shall be permitted only with the consent of the data subject. And it leaves limited room for exceptions (..)'

Consent

- *"With each sign of failure of privacy self-management, however, the typical response by policymakers, scholars, and others is to call for more and improved privacy self-management."*
- *"Although privacy self-management is certainly a laudable and necessary component of any regulatory regime, I contend that it is being tasked with doing work beyond its capabilities."*

(Solove, 2013)

Consent

Informational self-determination, problems

- **Cognitive**
  - *"(1) people do not read privacy policies;*
  - *(2) if people read them, they do not understand them*
  - *(3) if people read and understand them, they often lack enough background knowledge to make an informed choice, and;*
  - *(4) if people read them, understand them, and can make an informed choice, their choice might be skewed by various decisionmaking difficulties."*

- **Structural**
  - Information overload

(Solove, 2013)

2. Consent

Project-specific consent

- Costly and time-consuming
- Poor recruitment, consent bias
- Often impracticable or impossible

Ways forward?

Consent at a one-off event

- Legal validity?
- Ethical acceptability?
- Impossible to foresee or comprehend consequences
- Insufficient to ensure meaningful control

Dynamic consent

- Using IT, Consent as an ongoing process
- Participatory approach
- Practical difficulties?
- Information overload for the individual?

(References in: Mostert et al, 2016)

2. Consent

Broad consent, GDPR (Recital 33)

- *"It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection.*
- *Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research."*

2. Consent

Key messages

I. Seeking (broad) consent remains an important requirement to show respect for persons, also in Big Data research

II. Difficulties in seeking consent could hamper scientific progress

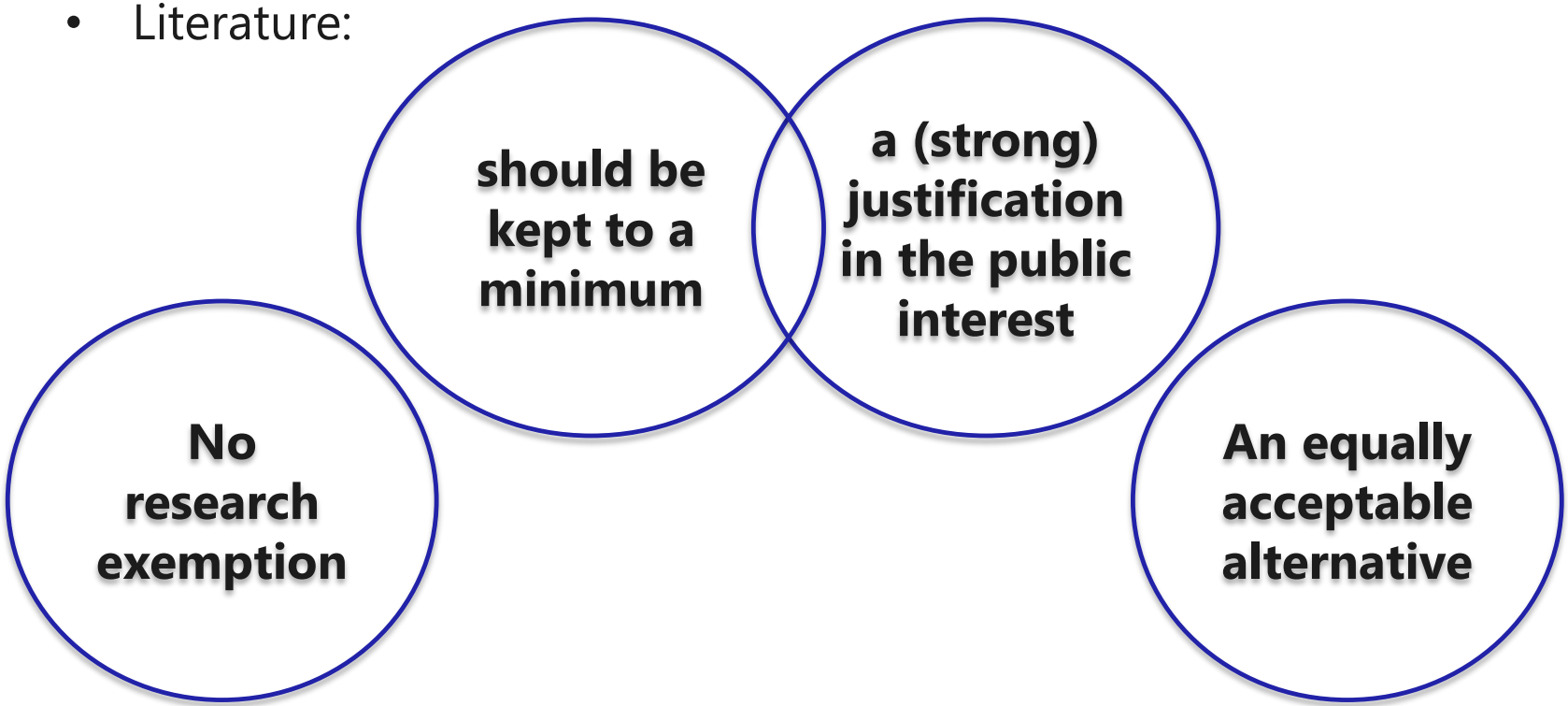
III. Consent does not, in itself, reduce risks to individuals

3. Research Exemption

3. Research Exemption

Scope of another legal basis than consent?

- Literature:



- GDPR, art. 9(2i) and 89(1)
  - processing is necessary for scientific research purposes

(References in: Mostert et al, 2016)

3. Research Exemption

The importance of appropriate safeguards

- **Literature**
  - Limiting data access and use
  - Opt-out registration
  - Pseudonymisation
  - Authorisation by ethics committee
  - Engaging in public participation, etc.

- **GDPR**
  - Research exemption should be subject to certain additional appropriate safeguards (technical and organisational measures, e.g. pseudonymisation if possible)
  - Privacy by design, good information governance

(References in: Mostert et al, 2016)

4. Principles of data protection

- http://fra.europa.eu/en/publication/2018/handbook-european-data-protection-law

General principles (5 GDPR)
Lawfulness
Fairness
Transparency
Purpose limitation
Data minimisation
Data accuracy
Storage limitation
Data security
Accountability

4. Principles of data protection

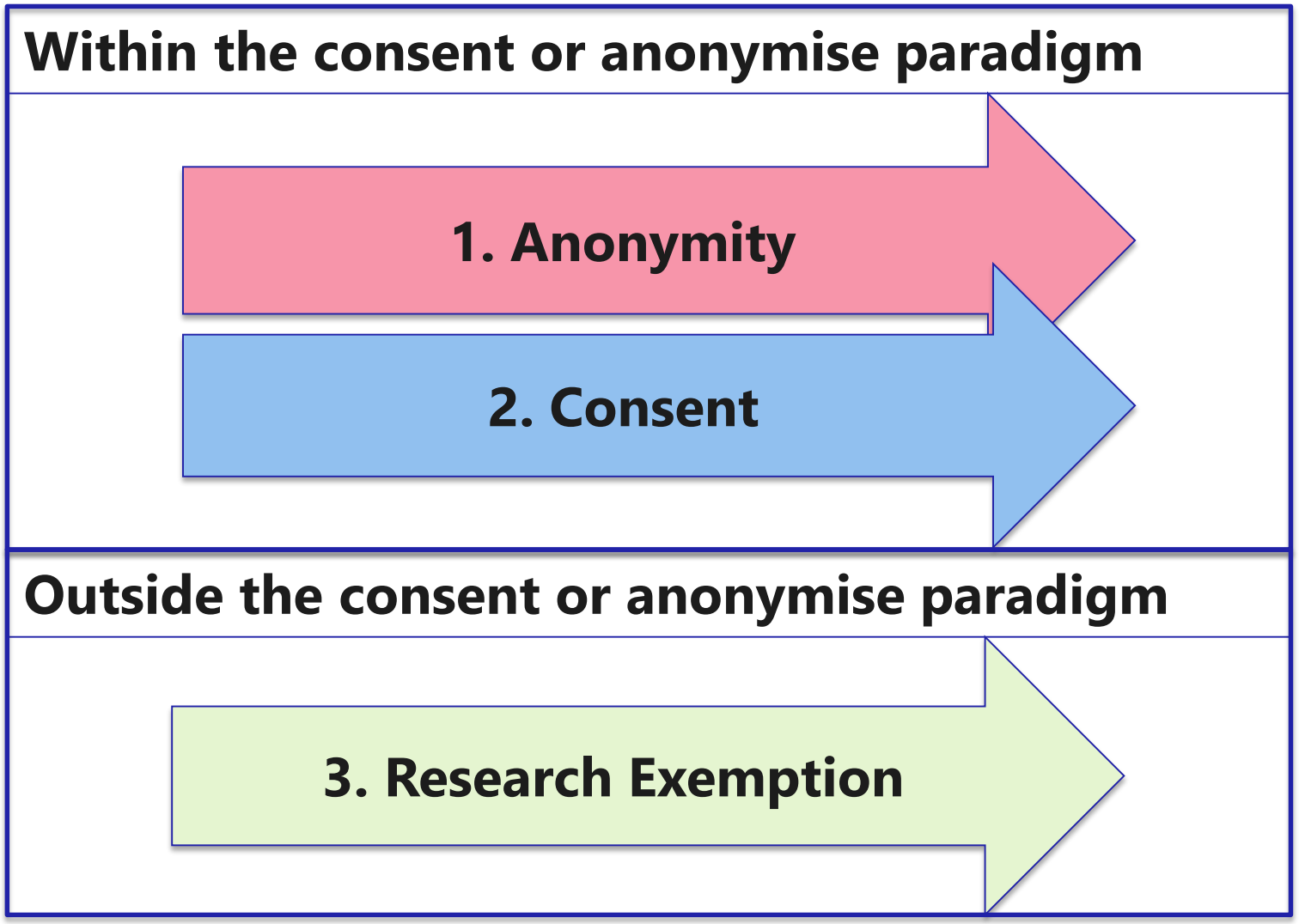
Research exemptions

- Purpose limitation
  - "further processing for (..) scientific (..) research purposes (..) shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes"
  - This exemption applies to further processing by the controller
- Storage limitation
  - "personal data may be stored for longer periods insofar as the personal data will be processed (..) for (..) scientific (..) research purposes or statistical purposes in accordance with Article 89(1) (..)"

Article 89(1) GDPR

- " (..) safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation (..) [or anonymisation]"

5. Discussion



5. Discussion

- Limitations to the consent or anonymise approach should be recognised
- It is a continuing moral duty to protect the rights and interests of those who have provided personal data, irrespective of the legal basis for the use of this data
- Privacy by design strategies and good information governance are vital for this protection

Trust takes a long time to be earned and a short time to be lost

Further reading

- Nuffield Council on Bioethics: *The collection, linking and use of data in biomedical research and health care: ethical issues*, 2015. Available at http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf

- Mittelstadt BD, Floridi L. *The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts*. Sci Eng Ethics. 2016;22(2):303-41.