

Personal Data for Common Good.
Towards sustainable Big Data

Nadya Purtova
22 May 2017, Ljubljana

Focus

- (Re)use of (personal) data sets rather than AI;
- Offer a conceptual framework of commons to:
 - Define what sustainable Big Data / data analytics means;
 - conceptualize / articulate / systematize the **problems** of personal data (re)use, beyond privacy & data protection, and
 - outline a roadmap to dealing with these problems
- Two econ analysis of data protection law papers:
 - 2015 Illusion of personal data as noone's property, *LIT*
 - 2017 (Health) data for common good, Springer

Some background

		SUBTRACTABILITY OF USE	
		<i>High</i>	<i>Low</i>
DIFFICULTY OF EXCLUDING	<i>High</i>	Common good	Public good
	<i>Low</i>	Private good	Club

Subtractability:

- less(er quality) of a resource available due to use;
- not limited to physical extinction / depletion;
- also **socially constructed**.

Value of the commons discourse

In the context of the knowledge commons, speaking in terms of the commons helps “articulate ... concerns and provide a public vernacular for talking about the politics of creativity and knowledge.”

D. Bollier (2007)

‘The Growth of the Commons Paradigm’



EMMA WATSON **TOM HANKS**

THE CIRCLE

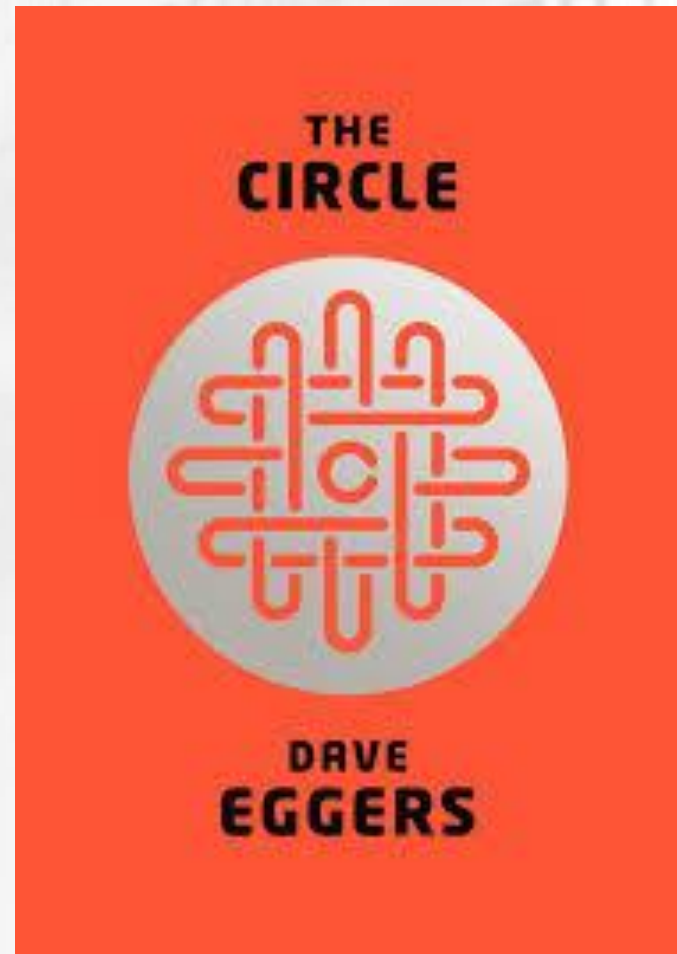
**KNOWING IS GOOD.
KNOWING EVERYTHING IS BETTER.**

4.28.17

The **promise** of Big Data Analytics

“We can cure any disease, end hunger, everything, because we won’t be dragged down by ... our petty secrets, our hoarding of information and knowledge. We will finally reach our full potential.”

Dave Eggers, *“The Circle”*





‘Sharing is caring!’

Dave Eggers, *“The Circle”*

2015: 'Donate personal data to science via Apple ResearchKit'



2016: **Google** DeepMind for better healthcare (1.6 mln patients)

**New
Scientist**

HOME NEWS TECHNOLOGY SPACE PHYSICS HEALTH EARTH HUMANS LIFE TOPICS EVENTS JOBS

Advertisement

[Home](#) | [News](#) | [Health](#) | [Technology](#)

TECHNOLOGY NEWS 29 April 2016

Revealed: Google AI has access to huge haul of NHS patient data

A data-sharing agreement obtained by **New Scientist** shows that Google DeepMind's collaboration with the NHS goes far beyond what it has publicly announced

GDPR more allowing re reuse?

Recital 157: “By coupling information from registries, researchers can obtain new knowledge of great value with regard to widespread *medical* conditions such as cardiovascular disease, cancer and depression. [...] Within *social science*, research on the basis of registries enables researchers to obtain essential knowledge about the long-term correlation of a number of social conditions such as unemployment and education with other life conditions. Research results [...] provide [...] knowledge which can provide the basis for the formulation and implementation of knowledge-based policy, improve the quality of life for a number of people and improve the efficiency of social services.”

'Open data'

Public data — 'all the information that public bodies in the European Union produce, collect or pay for' — open for re-use (commercial and non-commercial).

European Commission. 2011. "Open data: An engine for innovation, growth and transparent governance"

Directive 2003/98/EC on the re-use of public sector information (2013 Amendments)

Language of **the commons**
emphasis on **sharing**

The commons so far

- ‘Traditional commons’ - natural resources:
 - Agricultural land;
 - Fisheries;
 - Water basins
- ‘New commons’:
 - Cyberspace
 - Scientific knowledge (Ostrom & Hess)
 - Social norms (Karen Yeung)

We need to **account for the dilemmas**
of the personal data commons for
sustainable data sharing & (re)use

Two types of commons dilemmas

Provision problem

- Destruction / exhaustion / corruption of the common resource;
- Lack of incentives to preserve a resource;
- Unsustainable *appropriation rate*, or
- Unsustainable *appropriation tools*.



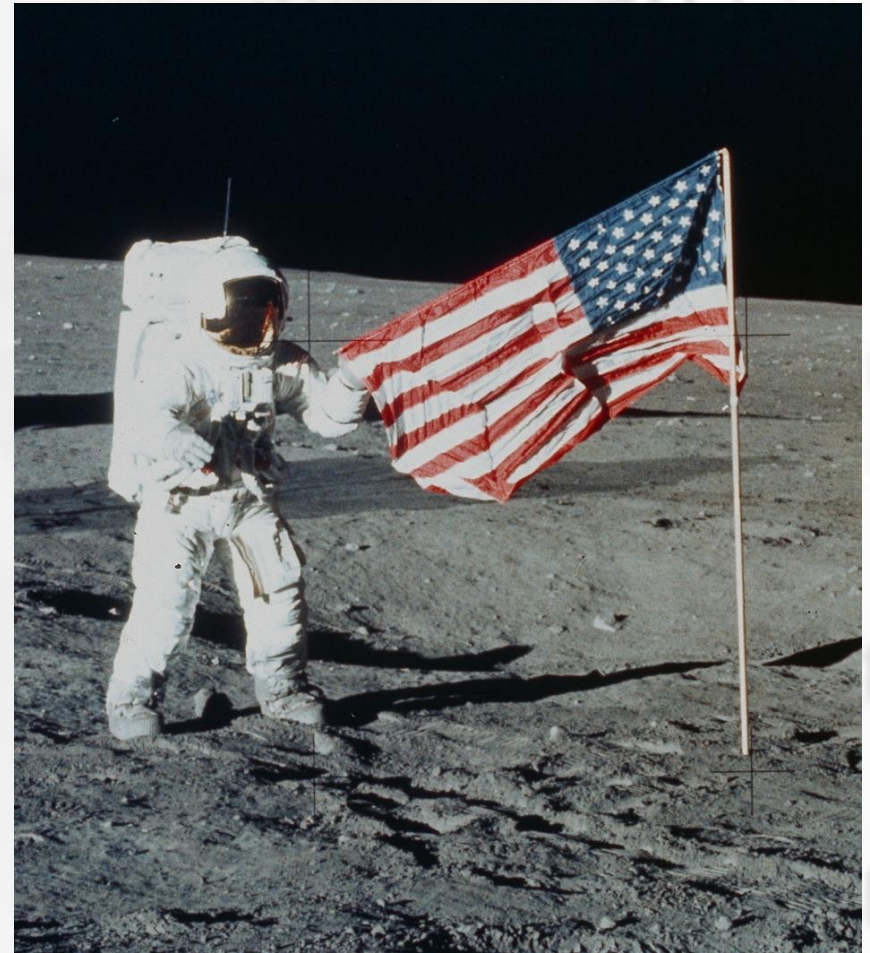
Appropriation problem (1/2)

- relationship between input of effort & yield;
- Cost of production $>$ returns;
- result of (mis)allocation of or exclusion from the resource (e.g. 'inequality of arms')



Appropriation (2/2)

- **'Grabbing'** of the previously publicly available resource by elites → have-nots excluded from benefits;
- **New technologies enable "capture"** of resources previously unowned (minerals on Mars);



What it mean for personal data (re)use

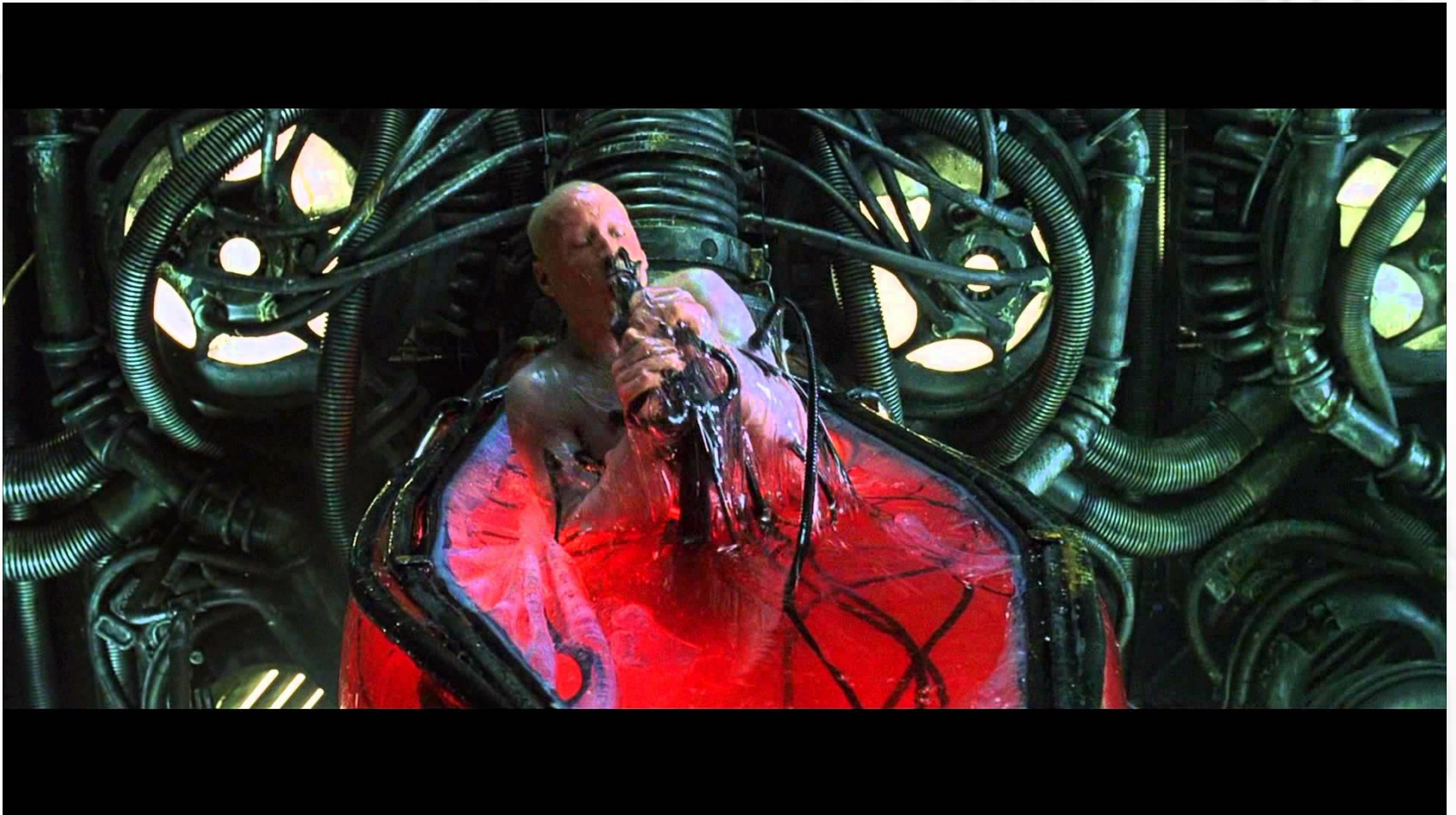
The background features a light gray, semi-transparent grid of binary code (0s and 1s) arranged in a perspective that recedes into the distance. Overlaid on this are several thin, dark, curved lines that sweep across the frame, creating a sense of motion and depth.

Drawing boundaries of a shared resource

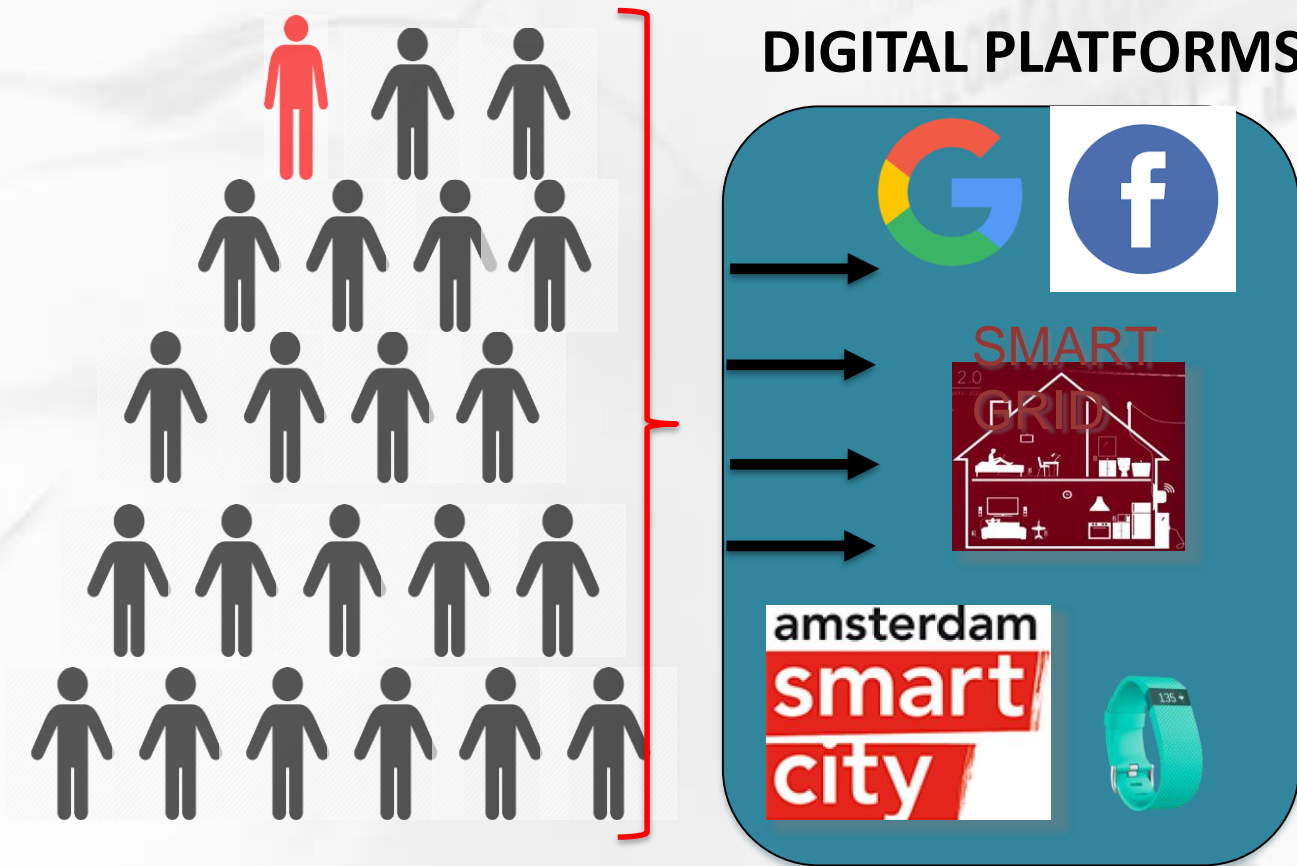
Determining the **boundary** of the resource

- Personal data
 - Advances in data analytics make all data (potentially) personal
- Health data
 - all data can potentially reveal current and future health status (mobile phone metadata revealed that a caller had multiple sclerosis (Stanford study, May 2016))
- Specific data pools?
 - Smart cities; NHS records;

'Human livestock'



(Personal) data is a complex system resource, people at the core



Individual or group?

Individuals?

- personal data is no longer strictly personal
- Genetic data relates to family members
- It does not have to be your data to affect you:
 - “marketers ... do not care about snooping into particular people’ s lives.” (Solove)
 - Group profiles are as valuable; identification is not a must.
- Your privacy choices deteriorate privacy protection for all (‘network effects’)

Communities (groups) contributing & affected? (e.g. smart cities)

- Problem of coordinating collective action in large groups
- Overlapping lower- and higher levels of groups
- The larger the data pool, the larger the group to coordinate
- Dynamic communities (e.g. website visitors)

Appropriation: 'data grabs' & enclosure

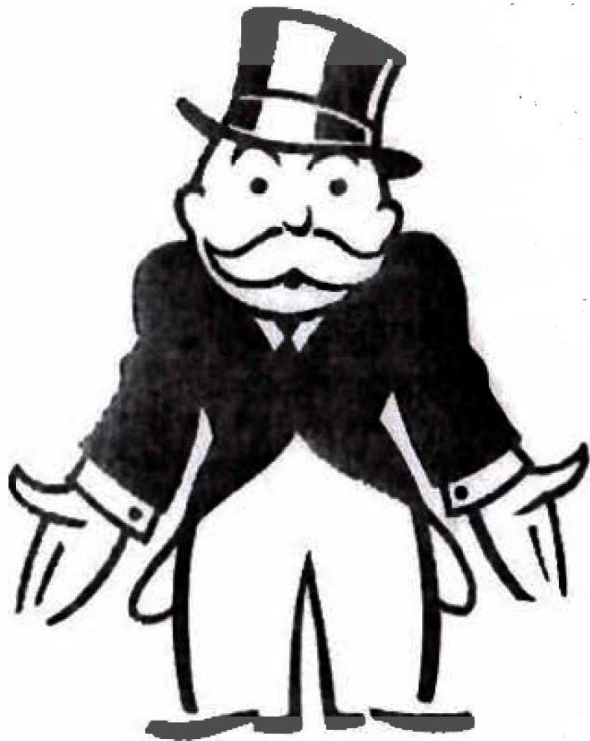
'Data grab' by (tech.) elites → data have's & have-nots.

- Unique digital platforms enable collection and appropriation;
- Contrary to Google, their search data is not for everyone to collect;
- A few have advanced AI tools to benefit from Big Data → gatekeeps to data-driven knowledge?

Distribution of data wealth:

- will patients contributing data to pharmaceutical companies have share in the benefits? Prices of medication?

Data inequality



URNS OUT
~~**INCOME**~~ **data**
INEQUALITY
IS A THING

Provision dilemmas: old & new 'privacy harms'

Given that the resource is people,

Not physical extinction of humanity, but

Long-term corruptive effects of the use of data commons on social/community values:

- democracy, non-discrimination, due process, individual autonomy, filter bubble, manipulation of citizens & consumers, etc.



So what?

Preliminary conclusions

- Sustainable Big Data accounts for provision & appropriation dilemmas:
 - Break Google monopoly on search data?
 - Ensure equal access to data-driven knowledge.
- Data is inseparable from people & communities →
- What ‘fair personal data use’ means should not be limited to individual impact;
- Learn from the successful cases of governance of commons (*few* conclusive recommendations):
 - Involvement of the community / all ‘appropriators’ in data governance;
 - Necessary limits on access to data

Thank you!

Nadya Purtova, TILT
n.n.purtova@uvt.nl