

A Research Perspective on the GDPR 2018

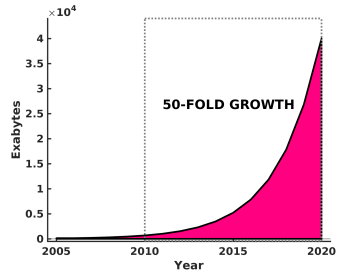
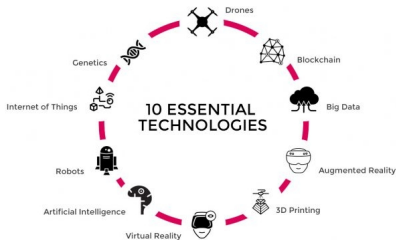
Legal Constraints on Data-driven SSH Research

Kristoffer L Nielbo
knielbo@sdu.dk
knielbo.github.io

datacube | Department of History | University of Southern Denmark

PROGRAM

0.00	Data, OD & initial concerns	deluge, challenges and copyright
0.10	GDPR and scientific research	safeguards, exemptions, pseudonymization
0.20	Discussion	...



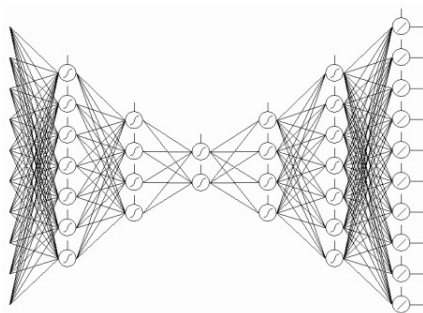
- the data deluge is transforming knowledge discovery and understanding in every domain of human inquiry
- knowledge discovery depends critically on advanced computing capabilities

a large part of these data are unstructured and fundamentally cultural

- we have limited experience with computational approaches and lack common standards for managing first- and third-party data

- with the advent of machine learning data combined with rich domain knowledge are our most valuable assets
- ‘generic’ benchmarking data sets are being replaced by domain relevant datasets
- probabilistic notions of ‘fairness’ (i.e., parity definitions) are challenged by impossibility results

GDPR grants users the right to a logical explanation of how algorithms use their data



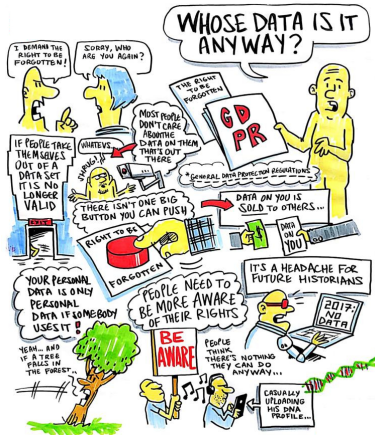
DATA-RELATED CHALLENGES

data obligations

- personal data protection policies (e.g., GDPR) are basically a good thing
- intellectual property rights (e.g., copyright) can also be a good thing
- BUT both can be a hindrance to research

cultural issues

- develop an organisational culture of analytics
- computational and data literacy in SSH
- data silos at private and public data controllers



Pseudonymization

definition

- data that cannot be attributed to data subject without additional data
- iow. removal of direct identifiers

anonymization

- not GDPR relevant
- removal of direct and indirect identifiers

why

- facilitate data re-use
- safeguard for scientific, historical and statistical purposes
- essential “by-design” feature
- used to meet data security requirements
- “delete and provide” is not necessary if data subject is not identifiable



SoMe example

data

- web-based and predominately unstructured data
- majority are publicly available
- often available through API

principles

- consent/opt-in policy
- right to be forgotten (delete and provide)
- parent consent

strategies

- 'walled garden' of Terms and Conditions
- 'Silence, pre-ticked boxes or inactivity' is inadequate to confer consent
- exempt from notice requirement, if data are from publicly available sources

THANK YOU

knielbo@sdu.dk
knielbo.github.io

& THANKS TO

Culture Analytics @ Institute of Pure and Applied Mathematics (UCLA)
Tina Eliassi-Rad