DATAETHICS

Principles and Guidelines for
Companies, Authorities
& Organisations

The Independent Thinkdotank

DATAETHICS – Principles and Guidelines for Companies, Authorities & Organisations
1. Edition 2018

Authors: Pernille Tranberg, Gry Hasselbalch, Birgitte Kofod Olsen & Catrine Søndergaard Byrne
Book layout: Spintype.com
Cover design: Paws Fabrik
Translated by: Focuspr.dk
Printed by: AKAPRINT A/S
Published with: Spintype.com

Isbn print: 9788771920475
Isbn pdf: 9788771920482
Isbn epub: 9788771920499

CONTENTS

CHAPTER 1

INTRO

The independent thinkdotank DataEthics.eu has developed a set of data ethics principles and guidelines that may help the integration of data ethics in your data processing activities. Here, we present the principles, a detailed questionnaire and a FAQ on data ethics. We acknowledge that nobody is perfect, that everything is in beta - also with data ethics. The important thing is that we have started the process and get better at it for every step we take.

The principles and guidelines may be reproduced freely as long as DataEhics.eu is clearly cre-

dited with a link to dataethics.eu/en/data-ethics-principles/

Find more information, tools and inspiration on www.dataethics.eu

All the very best,

Pernille Tranberg
Gry Hasselbalch
Birgitte Kofod Olsen
Catrine Søndergaard Byrne

September 2018

CHAPTER 2

DEFINITION OF DATA ETHICS

Data ethics is about responsible and sustainable use of data. It is about doing the right thing for people and society. Data processes should be designed as sustainable solutions benefitting first and foremost humans.

Data ethics refer and adhere to the principles and values on which human rights and personal data protection laws are based. It's about honest and genuine transparency in data management. To actively develop privacy-by-design and privacy-enhancing products and infrastructures. To treat someone else's personal information as you wish your own, or your children's, treated.

Data ethics is the step further than mere compliance with personal data protection laws: All data processing therefore respects as a minimum the requirements set out in the EU's General Data Protection Regulation (GDPR), the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights.

CHAPTER 3

DATA ETHICS PRINCIPLES

THE HUMAN BEING AT THE CENTRE

Human interests always prevail for institutional and commercial interests. People are not computer processes or pieces of software, but unique with empathy, self- determination , unpredictability, intuition and creativity and therefore have a higher status than machines. The human being is at the centre and have the primary benefit of data processing.

INDIVIDUAL DATA CONTROL

Humans should be in control of their data and empowered by their data. A person's self-determination should be prioritised in all data processes and the person should be actively involved in regards to the data recorded about them. The individual has the primary control over the usage of their data, the context in which his/her data is processed and how it is activated.

TRANSPARENCY

Data processing activities and automated decisions must make sense for the individual. They must be truly transparent and explainable. The purpose and interests of data processing must be clearly understood by the individual in terms of understanding risks, as well as social, ethical and societal consequences.

ACCOUNTABILITY

Accountability is an organisation's reflective, reasonable and systematic use and protection of personal data. Accountability is an integral part of all aspects of data processing, and efforts are being made to reduce the risks for the individual and to mitigate social and ethical implications. Sustainable personal data processing is embedded throughout the organisation and ensures ethical accountability in the short, medium and long term. An organisation's accountability should also apply to subcontractor's and partners' processing of data.

EQUALITY

Democratic data processing is based on an awareness of the societal power relations that data systems sustain, reproduce or create. When processing data, special attention should be paid to

vulnerable people, who are are particularly vulnerable to profiling that may adversely affect their self-determination and control or expose them to discrimination or stigmatisation, for example due to their financial, social or health related conditions. Paying attention to vulnerable people also involves working actively to reduce bias in the development of self-learning algorithms.

CHAPTER 4

QUESTIONNAIRE

These questions can be used in combination with the FAQ to work with data ethics dilemmas in your organisation. You can for example use your discussion of the questions as a basis for preparing data ethics guidelines.

THE HUMAN BEING AT THE CENTRE

- *Is your data processing based on the fact that you borrow data from the users (not owner of their data)?*
- *Do you ensure that the user's rights are prioritised, rather than commercial or institutional interests?*

- *Do you ensure that primarily users benefit from their own data – not just the organisation?*
- *Do you use privacy-by-design principles, and can you describe them clearly and transparently?*

INDIVIDUAL DATA CONTROL

On-device processing

- *Do you ensure that users' data - as far as possible - is processed directly on the users' own device(s)?*
- *When the processing of data is necessary other than on the user's own devices, such as your server or a cloud solution, is collected data not related to an identifiable person?*

Profiling

- *Do you use profiling? If so, do you allow the user to influence and determine the values, rules and input that underlie the profiling?*

Predictions

- *Do you use data to predict individual-level behaviour or only patterns?*

TRANSPARENCY

Data Storage

- *In which country is your data stored?*
- *Where is the storage solutions provider headquartered?*
- *Does the transmission of data go through countries outside of the EU?*

Artificial Intelligence

- *Do you use machine learning / artificial intelligence? If so, can you explain the algorithms - the criteria and parameters?*

Behavioural Design

- *Do you use personal data to influence user behaviour?*
- *Do you ensure that it is transparent when the use of personal data may influence a user's behaviour?*
- *Do you ensure that the design does not create addiction and thus influences the person's self-determination and empowerment?*

Open Source

- *Do you operate with open source software, so others can use it and possibly develop it further ?*

ACCOUNTABILITY

Anonymity

- *When do you anonymise personal data?*
- *Do you use end-to-end encryption of data?*
- *Do you minimise the use of metadata and explain how it is done?*

Zero-knowledge

- *Do you use zero knowledge as a design principle?*

Sales of Data

- *Do you sell data to third parties?*
- *Do you sell data as personal identifiable data?*
- *Do you sell data as patterns on an aggregated level?*
- *If you sell data, are you making sure that it is fully anonymised information only describing patterns, not individuals?*

Data Sharing

- *Do you use third-party cookies?*
- *Does this include SoMe (social media) cookies and SoMe logins?*
- *Do you use Google Analytics or similar tracking tools?*
- *If you use third-party cookies, are your users fully aware that your cookie use leads to sharing of data about your users with third parties and do they agree with it?*

Data Enrichment

- *Do you enrich data with external data, such as social media data, bought data or web scraping?*
- *Does this enrichment occur in response to, or in cooperation with, your users?*

Organisational Anchoring

- *Do you have an individual or a department responsible for the ethical managing of data?*
- *How is the work with data ethics embedded in the organisation?*
- *How do you ensure that your data ethics guidelines are respected?*

External Control

- *Can the processing of data be audited by an independent third party?*
- *Do you require and and control the data ethics of your subcontractors and partners?*

EQUALITY

Public Platforms

- *Do you engage in dialogue with your users on a public platform?*

- *Do you have guidelines for using the platform?*
- *Do you moderate the platform in order to remove sensitive personal data?*
- *If your services are offered to children, do you ensure parental consent?*

Reuse of data

- *Is data used to develop or train an algorithm?*
- *Do you ensure that the use of data does not lead to discrimination?*
- *Do you ensure that the use of data does not expose the vulnerabilities of individuals?*

Artificial Intelligence

- *Do you ensure that the use of artificial intelligence / machine learning is to the benefit of the individual and does not cause physical, psychological, social or financial harm to the individual?*

CHAPTER 5

FAQ ON DATA ETHICS

Below are a number of Frequently Asked Questions, in alphabetical order, regarding data ethics.

Active Party

What does it mean to be an active party?

If a doctor writes something about you in a patient journal or a teacher records something about your child, or something about you as a parent, and it is your belief that they have not considered the nuances of the matter, you can contribute additional information that is visible to anyone who has access to your data. Or an insurance company gives you access to com-

ment on the conclusions they have reached based on your data.

Anonymization

What is ethically responsible anonymisation?

Pseudonymisation means that you cannot directly see the individuals the information concerns, however, there is still an opportunity to establish the identity of those individuals. Anonymisation is a step further. No one should be able to recreate the identity of a person. In pseudonymisation as well as anonymisation of personal data, it is important to document this and allow a third party to examine the internal machinery and to verify and possibly certify the fact. There are not many external third parties offering this service today, but it will be an important step forward, see "External Control"

Artificial Intelligence (AI)

What is the best way to manage artificial intelligence?

We do so by ensuring human control. The Department of Applied Mathematics and Computer Science at the Technical University of Denmark, has formulated a number of fine Safe AI principles:

- **Safe AI is safe:** has passed tests and verification and is robust against systematic and expert attacks
- **Safe AI is self-conscious:** understands its own role and uncertainty and can, for example, refuse to act
- **Safe AI can keep a secret:** privacy protection and privacy by design, is built-in
- **Safe AI has well-defined values:** is cleansed of stereotypes, bias, and understands emotions

- **Safe AI has social skills:** understands social relations and understands the user's knowledge and skills
- **Safe AI understands power:** understands the data and related action contexts and its consequences
- **Safe AI is documented:** transparent and communicative, offering the right to explanation
- **Safe AI is open source:** methods, code and test results are available to all

Source: Professor Lars Kai Hansen, DTU Compute

Behavioural Design

What is behavioural design in a data ethics perspective?

Use of personal data to influence user behaviour may be manipulative if the user's control is not at the centre, but the design primarily is developed to create dependence, increase use of a ser-

vice and user numbers or simply to effect more sales. Behavioural design should be transparent to the user, and aim not to have discriminatory effects or be addictive. The individual must be empowered and be able to preserve their self-determination.

Bias

What is bias in design?

Bias is built-in prejudices and negative stereotyping. Bias may occur in training data, the historical data used to develop a self-learning algorithm. Bias may also occur in the design of an algorithm that can categorise and label people in a way that discriminates between, for example, population groups. Bias can be reduced by, among others, manual sorting and cleanup of data. It can also be diminished by making sure that the algorithm can be explained and interpreted and is open to auditing. For example, the winners of the first beauty con-

test, judged by a self-learning algorithm, were virtually all white, because the algorithm had been trained mostly with images of white people. Bias was here in the training data that did not include many images of other races.

Explainability

What is explainability?

Algorithms must be explained in a way that individuals understand them. They should not only provide basic information about the data processing but must be documented and be able to explain how a given algorithmic decision has been taken, including the criteria and parameters for a decision e.g. regarding credit rating, insurance premium or allocation of social benefits.

Data Activation

What is Data Activation?

The GDPR gives individuals the right to con-

trol their own data and the right to 'portability', that is to get data easily transferred at the users' request. However, individual data control is not enough in the long run, individuals will increasingly also need to be empowered to activate their own data and utilise it for the purpose of enriching their own finances, health and everyday lives. This will also be beneficial to the company or institution providing new services where the individual can activate their own data.

Data Enrichment

What is web scraping and can you be ethically responsible?

Is it possible to enrich data with web scraping from websites, including public sections of social media. However it is controversial, because even though data is publicly available, it has ethical implications. Therefore, make sure

that it is done at the request, and with the informed consent, of the users.

Data Sharing

When is it unethical to use third party cookies?

If a company or organisation is dealing with children and others considered vulnerable, it is not ethical to allow third party cookies on one's website, for the purposes of sharing identifiable, sensitive data with third parties. If you have health data or data about political opinions or affiliation, sexual or religious orientation or other sensitive data, it is also unethical to allow third party cookies on the website. It is also unethical for the Public Sector to share data about citizens' behaviour via third party cookies - including SoMe cookies. Few consumers and citizens understand data sharing and perceive it as covert even though they have explicitly agreed to it via a pop-up. It may be legal, but it is regarded as disturbing and over-

stepping boundaries by most people and must therefore be considered unethical.

Data Storage

When is data storage ethically problematic?

It may be lawful to store data in a country outside the EU, however, just because it is legal, doesn't mean it is ethical. For example, it could be discussed whether it is ethically justifiable to store data with a company based in countries that practice and allow digital dictatorships or data monopolies. We don't think so. However, storing data on one's own servers, where you always have control over the data, or in a cloud provider with its registered office in the EU/EEA, can be considered as data ethical.

Earning money

Can you earn money on data processing?

Yes, you can, as long as the individual's control is at the centre. Financial interests can never

override human rights, such as the right to privacy, self-determination and not to be discriminated against or stigmatised. This can be achieved, for example, through Privacy-by-Design.

End-to-end encryption

What is encryption?

One thing is to encrypt traffic, so nobody can intercept data in transit, another is end-to-end encryption, where no one other than sender and recipient can see the content, not even the company that owns the platform where the communication takes place.

External Control

Why independent auditing?

It is important - at least prospectively - that data processing can withstand being reviewed and verified by an external independent auditor. As with the environment, child labour and IT security, more and more users need to know

that what you say is what you actually do. Today, only few credible schemes exist to verify or certify, such as ISO and EuroPriSe. However in the wake of GDPR, the EU has announced a decisive European privacy certification system, as there are no doubt more certification schemes to come.

Metadata

What is metadata?

Metadata is data about data or a type of declaration for data that can provide information about data sets and services. For example, there may be data about who has sent an email, to whom, and when, but this metadata shows nothing about the content of the email. There have been discussions about whether metadata is personal information, i.e. whether it is possible to identify a person based on metadata. However, there is no doubt that metadata is a powerful tool for mapping behaviour, circle of interests,

habits, etc., as well as for profiling. The data ethical question is, whether to identify and inform users regarding the use of metadata, and whether the user can gain insight and access to metadata.

On-Device Processing

What is on-device processing?

Instead of collecting and storing personal data on your own server or cloud, you can process data directly on users' devices. Apple does it with Siri, as does the secure Swiss messaging app Wire and the secure German browser Cliqz. When there is a need for processing of data on a server, data should be collected anonymously without being identifiable.

Open Source

Why is open source a good thing?

Open source means that the source code in your applications is freely available and others can

improve, test, debug and secure the system - there is worldwide support and security vulnerabilities are often captured quickly. It does not mean it's free, but often cheaper and allows you to develop independent of the supplier. Then there's the transparency: Your users - or third parties, can find out exactly what the product contains.

Organisational Anchoring

What is organisational anchoring of data ethics?

Data ethics is not a "one-man job". It's a broad approach covering everything from product development, innovation and marketing to strategic development. The data ethics approach should therefore be driven entirely from top management and considered as a value among employees. It can be done in different ways, for example, Apple's privacy experts have been involved from the beginning in all product development teams, and French AXA (insurance)

has an advisory board of independent experts, that meets in Paris twice a year, to discuss data ethics dilemmas.

Predictions

Can predictions based on data be done ethically?

It may be acceptable to do individual predictions with, for example, personalised medicine and treatment. The data ethics question concerns whether the individual has insight and the power to object, to say no, or to choose whether to accept or not.

Privacy by Design

What is Privacy by Design?

Privacy by Design (PbD) means that the setting of a service is private by default and that it is designed and developed with privacy as a starting point. The first PbD principles were developed in the 1990s by Dr. Ann Cavoukian, former Information and Privacy Commissioner in

Canada, and have continued to evolve. You can also choose to see PbD as a business philosophy, that is, an innovative approach to digital business development, where privacy is the starting point for the different innovative business processes a company starts, from design and technological development to human resource development, CSR and marketing. The PbD principles thus become a general guide to building alternatives to the data-driven public-by-default enterprise.

Profiling

What is profiling?

Profiling is the automatic processing of data that analyses certain personal aspects regarding a person, so as to make predictions about them, for example, their work, financial situation or health. Profiling is a data ethics challenge because it is the most intense form of personal data processing, so you must first determine

whether there is a legal basis for profiling. An ethical data profile will always be developed for the benefit of the individual, and the individual will have the opportunity to influence and determine values, rules and input used in the profiling.

Sales of data

Can you sell data to third parties in an ethically responsible manner?

It is illegal to sell data to third parties unless you explicitly have consent to do it. A majority of Danes e.g. believe that organisations are selling data to third parties. It may not necessarily be so, but it is important to explain what, perhaps, you take for granted. You can sell data, with ethical responsibility, to third parties in the case of fully anonymized information based on patterns among a group of citizens.

Transparency

What is transparency?

It's not enough simply having a Transparency Report that shares how many times, for example, regulatory bodies asks for access to a company's data with a court order. Transparency is also about personal data processes in one's own organisation.

Transparency is a step further than basic legal requirements for consent and the possibility of insight and objection. Data systems and processes are designed to support the individual's trust and to give the individual real opportunity to object to the processing of their data. Individuals must be able to object to data processing without losing rights.

Vulnerable

Who is particularly vulnerable to data processing?

More data is often collected regarding particu-

larly vulnerable people, such as refugees, people with physical disabilities, people with mental illnesses, the socially disadvantaged, the unemployed, prisoners, etc. For example, Udbetaling Danmark, the authority responsible for the collection, disbursement and control of public benefits in Denmark, collects a lot of information about beneficiaries, their partners and households and assumed partners.

Zero-knowledge

What is the zero-knowledge principle?

According to GDPR, no data must be stored longer than is necessary for the purpose. You can choose to go beyond the legislation and delete data before the required date, so that you are not unnecessarily at risk by holding particularly sensitive data. This can be by means of auto-deletion, but also, by never having access to data in the first place.