ODI

COVID-19: Identifying and managing ethical data issues

Open Data Institute

Contents

About

This report has been researched and produced by the Open Data Institute, and published in June 2020. Dr David Tarrant is the lead author, with contributions by Renate Sampson and Violeta Mezeklieva. If you want to share feedback or would like to get in touch, contact the ODI Learning Team at learning@theodi.org.

OPEN FOR FEEDBACK

How can it be improved? We welcome suggestions from the community in the comments.

Covid-19: Identifying and managing ethical issues around data

A crisis is not a time to throw ethics out of the window – it is a time to do things fast and do them right. When data is used to make decisions, it is vital that people are protected from any possible harmful impacts, such as exclusion or profiling.

This guide is designed to help those collecting, sharing and using data identify and manage ethical issues.

Introduction

The Covid-19 pandemic has thrown unprecedented pressure on our services to do everything possible to prevent needless deaths. While many professions have existing ethical guidelines, notably guidance for healthcare professionals stemming from the Hippocratic Oath, new dilemmas will undoubtedly arise where more specific thought is needed.

The use of technology to help track symptoms, trace contacts or provide immunity passports has been a key area of debate. Data is a central part of all of these technologies. Many questions and concerns have already been raised about how these systems will be developed, how they interact with legal and ethical regulations, and who is set to benefit from them.

The Data Ethics Canvas is a tool developed by us at the Open Data Institute (ODI) that can help those collecting, sharing and using data identify and manage ethical issues.

What is data ethics?

The Open Data Institute defines data ethics as:

> A branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use.

Data ethics relates to good practice around how data is collected, used and shared. It is especially relevant when data activities have the potential to affect people and society, directly or indirectly.

Data ethics is not just about personal data

Although data ethics is about how data activities have the potential to affect people, this isn't always a result of personal data being collected.

An example of this is the game PokemonGo. Within days of the game being released many felt excluded. It was discovered that many areas were deprived of PokeStops and these appeared to be in lower-income and predominantly ethnic minority neighbourhoods. Investigations found that PokeStops had been crowdsourced from a previous game called 'Ingress' which was primarily played by people from white and higher-income neighbourhoods This limitation in the data led to accusations of bias, even though the data was not personal data.

Data ethics is not just about the use of data

Data ethics should be addressed at all stages of the data lifecycle:

The ethics of collecting data:

Before even considering the use of the data, you should ask questions about both the legal right and the ethics of collecting data.

For example, in a recent case, the police passed on images of seven people to a [private company using facial recognition technology](#) in a public space, without first disclosing the details of the image-sharing agreement. This highlights the ethical issues currently facing such technologies.. A lack of legislation or regulation over the use of such technology combined with a lack of transparency about when and where the technology was being being used resulted in a loss of trust[1].

---

[1] The Guardian (2019), Facial recognition technology scrapped at King's Cross site,

The ethics of using data:

> Data use is perhaps the clearest area where data ethics applies and many organisations already have codes of practice in place. However, many governments and organisations fail to inform people how the data will be used, what the results are, and how decisions have been made. We have published guides on [publishing models](#) and data that can help here.

The ethics of sharing data:

> At the ODI we believe that data should be available equitably and fairly to those who need it. Not sharing data can therefore be as unethical as sharing it.

Data ethics is not optional when free help is offered

During times of crisis, many technology companies offer pro-bono data services to public health organisations, governments, and communities. While we cannot be sure of the motivations of these companies, it is important that those accepting a pro-bono offer still carry out the standard legal and ethical checks.

Organisations considering accepting such help have to consider the same set of things regardless, to retain public trust, safeguard its own future operations, and consider the market effects it generates. Our Vice President Jeni Tennison has [written a blog post](#) considering this issue[2].

---

[2] Jeni Tennison (2020), <u>What to do about pro-bono data service offers</u>

Four focus areas in data ethics

At the ODI, we believe our Data Ethics Canvas can help those collecting, sharing and using data identify and manage ethical issues.

To help people get started with the canvas, this guide is broken down into four key areas.

1. Know the data

Key questions to ask:

- What data do you need, is it available?
- Does the data have limitations, gaps, inconsistencies or biases?
- Are there any existing ethical or legislative contexts you need to consider?
- Are there any issues with data rights that need to be addressed?

This focus area is concerned with the existing environment and how data is acquired. Establishing the usability of data, including your rights, legal obligations and current ethical practices is essential in any activity involving data. For example if the data is deemed personal (and related to European citizens) then it is likely that a Data Privacy Impact Assessment will be required.

Following this it is important to understand the limitations, gaps, inconsistencies or biases in the data. This is a critical stage to understand if decisions will be made equitably based on the data or if other techniques also need to be employed. We have already seen that Covid-19 is affecting black, Asian and minority ethnic (BAME) communities more than white British communities.We need to understand this and ensure that any data collected is used to inform decisions can help those who need it most[3].

2. Explore impacts

Key questions to ask:

- What are you trying to achieve by collecting, sharing or using data?
- What positive impacts will there be?
- What negative impact might there be?
- How will you minimise negative impact?

---

[3] News Chain (2020), Coronavirus outbreak's impact on BAME communities to be assessed by experts

Key to this focus area is to consider the trade-off between positive and negative impacts, both in the short and long term. This is especially the case during times of crisis when the short-term benefits appear to take precedence over the longer-term implications, however both should be carefully considered.

One of the key areas of discussion during the Covid-19 outbreak has been with contact-tracing applications and how the data is collected, stored and shared and for how long and for what purposes. There has been great debate about the merits of centralised and decentralised approaches to such efforts. While both approaches are able to perform slightly different granularities of contact tracing, one of the benefits governments favouring a centralised approach have articulated is that it may enable health providers and others to analyse and learn from data to manage future events. But many have questioned who the data will be shared with, for how long data will be collected and shared for, and how ethical use will be enforced[4].

Considering all the potential negative impacts or unintentional consequences can be challenging, It can help to think like a nefarious actor and try to envisage what uses they might have for the data. Cues can be taken from data protection regulations regarding data minimisation, for example, is it possible to achieve the same positive impact with a smaller amount of collected, used and shared data.

3. Plan your engagement

Key questions to ask:
- How do you ensure you are as open as possible about this project?
- How will you communicate with people?
- How will you allow people to engage if they have concerns?
- Will you be publishing data as part of this process?

Communication is key to helping people understand what it is you are trying to achieve, why and how.

Additionally, communication shouldn't stop once the ethical 'paperwork' has been completed. One of the biggest criticisms is often that communication is clear at the start of a project, but that in later stages, as more questions arise these go unanswered. Without clear communication it can be difficult to tell if the ethical practices are being followed and the positive impacts being observed.

As part of this process, think also about what data and models could be published transparently. Our guides on data and model publication can help here.

Engagement is all about building trust, and should be a high priority when dealing with crisis events.

4. Integrate ethical practices

---

[4] Aarogya Setu (2020): Why India's Covid-19 contact tracing app is controversial

Key questions to ask:

- How will systems and processes support the ethical practices?

- How often will these be reviewed?

- How will ongoing data ethics issues be measured, monitored, discussed and actioned?

- Who will be responsible for the integration and who must be involved?

Ethics is not just about filling in a piece of paperwork at the beginning of the work, it is a practice that should be embedded and thought about continuously. Unintended consequences may still occur, and it is important to have a process in place to discover, evaluate and take action when they do. Being open about this process, about how people can raise concerns, and about who is responsible are all necessary steps to build trust.

During a crisis hard decisions have to be made and compromises found that society will accept. Widespread action, such as contact tracing, can only succeed when society accepts those decisions and trusts that the right compromise has been found. Ultimately, the more trust that is built between those dealing with the crisis and those who need to follow guidelines, the more effective the response.

The data ethics canvas

To help guide people, our learning team have created an adaptation of the [ODI Data Ethics Canvas](#) to show the four focus areas clearly and help people get started.

How to use the canvas

1. All areas should be addressed by a cross-cutting team of experts with diverse social backgrounds.

2. Accept that people will have different moral beliefs. Addressing all the questions should help to find an ethical approach as a group.

3. The canvas should have an equal amount of detail and actions related to all focus areas.

4. The canvas should be regularly reviewed and updated as the project progresses.

Data Ethics Canvas

2020-06

Data sources

Name/describe your project's key data sources, whether you're collecting data yourself or accessing via third parties.

Is any personal data involved, or data that is otherwise sensitive?

Rights around data sources

Where did you get the data from? Is it produced by an organisation or collected directly from individuals?

Was the data collected for this project or for another purpose? Do you have permission to use this data, or another basis on which you're allowed to use it? What ongoing rights will the data source have?

Limitations in data sources

Are there limitations that could influence your project's outcomes?

Consider:

> bias in data collection, inclusion/exclusion, analysis, algorithms

> gaps or omissions in data

> provenance and data quality

> other issues affecting decisions, such as team composition

Ethical and legislative context

What existing ethical codes apply to your sector or project? What legislation, policies, or other regulation shape how you use data? What requirements do they introduce?

Consider: the rule of law; human rights; data protection; IP and database rights; anti-discrimination laws; and data sharing, policies, regulation and ethics codes/frameworks specific to sectors (eg health, employment, taxation).

Ongoing implementation

Are you routinely building in thoughts, ideas and considerations of people affected by your project? How?

What information or training might be needed to help people understand data issues?

Are systems, processes and resources available for responding to data issues that arise in the long-term?

Your reason for using data

What is your primary purpose for collecting and using data in this project?

What are your main use cases? What is your business model?

Are you making things better for society? How and for whom?

Are you replacing another product or service as a result of this project?

Positive effects on people

Which individuals, groups, demographics or organisations will be positively affected by this project? How?

How are you measuring and communicating positive impact? How could you increase it?

Negative effects on people

Who could be negatively affected by this project?

Could the way that data is collected, used or shared cause harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access (eg exclusive arrangements)?

How are limitations and risks communicated to people? Consider: people whom the data is about, people impacted by its use and organisations using the data.

Minimising negative impact

What steps can you take to minimise harm?

How could you reduce any limitations in your data sources? How are you keeping personal and other sensitive information secure?

How are you measuring, reporting and acting on potential negative impacts of your project?

What benefits will these actions bring to your project?

Reviews and iterations

How will ongoing data ethics issues be measured, monitored, discussed and actioned?

How often will your responses to this canvas be reviewed or updated? When?

Engaging with people

How can people engage with you about the project?

How can people correct information, appeal or request changes to the product/service? To what extent?

Are appeal mechanisms reasonable and well understood?

Communicating your purpose

Do people understand your purpose – especially people whom the data is about or who are impacted by its use?

How have you been communicating your purpose? Has this communication been clear?

How are you ensuring more vulnerable individuals or groups understand?

Openness and transparency

How open can you be about this project? Could you publish your methodology, metadata, datasets, code or impact measurements?

Can you ask peers for feedback on the project? How will you communicate it internally?

Will you publish your actions and answers to this canvas openly?

Sharing data with others

Are you going to be sharing data with other organisations? If so, who?

Are you planning to publish any of the data? Under what conditions?

Your actions

What actions will you take before moving forward with this project? Which should take priority?

Who will be responsible for these actions, and who must be involved?

Will you openly publish your actions and answers to this canvas?

Learn more

Whitepaper: helping organisations navigate ethical concerns in their data practices

This paper explores the relationship between data ethics and legal compliance, some existing data ethics frameworks and ethical considerations in data collection, sharing and use.

Download the paper

Course: Introduction to Data Ethics and the Data Ethics Canvas

In this half-day interactive course run by our experts you will be given an introduction to the concept of data ethics, and how to help your organisation use the ODI Data Ethics Canvas to help make better and more ethical decisions about data

Find out more and book your place