

THE ETHICS OF OPEN DATA

Sharing data can lead to great new insights — but it can cause problems too!



Open data is becoming an increasingly important type of data, with organisations, governments and even some companies increasingly publishing and relying upon it. Open data, in its simplest terms, can be freely used, re-used and redistributed by anyone. This means that the data has to be made available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet.

Open data should also be useable by anybody. This means everyone should not only have access but should also be able to use, re-use and redistribute — there should be no discrimination against fields of endeavour or against persons or groups. This means that companies, individuals and charities all have the same rights to the data and to using it.

Open data can help us more clearly understand the world around us, and because so many people can use the data, it helps ensure that the data is used in more representative ways — for example, communities might use open data to provide evidence of governments not supplying sufficient funding for schools.

Open data, though, can have significant ethical issues. Data from which it is possible to identify individuals or where they live could have serious consequences for people.

KEY RESOURCES

Open Knowledge Foundation. (2022). [Open Data Handbook](#) (Online) From: opendatahandbook.org

Politzer, M. (2021). [Why we need to think about ethics when using satellite data for development](#). DevEx (Online) From: bit.ly/3L9ik5B

Specht, D. (2021). [Responsible and ethical use of location data](#). [Ask the Geographer Podcast](#) (Online) From: bit.ly/3L80fos

Troullinou, P. (2017). [Rethinking Privacy and Freedom of Expression in the Digital Era: An interview with Mark Andrejevic](#) (Online) From: bit.ly/3IW0tvb

Open Data and Sacred Forests

Let's look at an example of where open geographic data can cause harm, even when people are trying to do good.

Imagine a forest. It is made up of hundreds or thousands of trees that together produce a forest scape. This forest holds special meaning for a community of people who live near it and who regard some of the trees as sacred. The forest, though, is also seen by a logging company as having significant economic value.

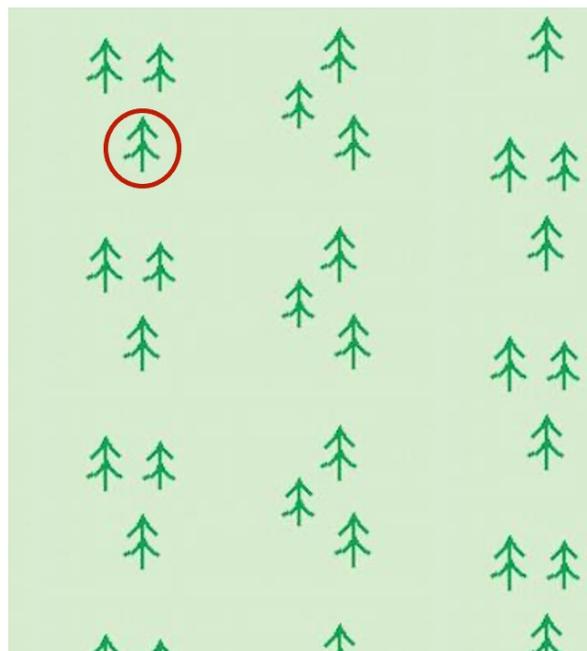
It isn't possible for the logging company to cut down the trees, because the sacred trees are protected by law. Some people were worried that the logging company might ignore this, so they set out to map all the sacred trees in the forest so they could prove the forest shouldn't be cut down. Once they had mapped all the trees, they decided to make this information open data. That way, they thought, the whole world would know how important this forest is, and they would help protect it for the people.

Open data, though, is open for everyone. And that means it was open for the logging company too. They were able to use the openness to their advantage and cut down all the trees that were not listed as sacred. This meant the logging company acted legally, but the community lost their forest — the single sacred trees did not have the same meaning when not part of the wider landscape.

This example demonstrates why we might not always want data to be available to everyone on equal terms, but getting to decide which data stays secret and which is made open is a deeply ethical question.

The majority of research data — even sensitive data — can be shared ethically and legally if researchers employ strategies of informed consent and anonymisation and control access to data. This means that when you are collecting data you should only collect the data that you need and that everyone you collect it from should be aware of your intention to make the data open to everybody.

Open data is normally shared through specific databases, all of which will have their own standards and guidelines to help ensure datasets are not going to harm people. Using data from these kinds of sites will also ensure that you are using secondary data that has been carefully checked for any ethical issues.



ACTIVITIES

1. Visit the Humanitarian Data Exchange (data.humdata.org). At the bottom of the page go to 'QA Process'. Here you will find out about how HDX checks data before it is public.
 - i. Do you think these processes would have helped stop the forest problem?
 - ii. If yes, how? If no, what extra steps would you add to the process?
2. At the top of HDX website, choose 'Dataviz' and the 'dataviz gallery'. Browse the gallery until you find something of interest to you. Then answer the following:
 - i. What is the main story this dataviz is trying to tell?
 - ii. What kind of data does the dataviz contain? Is any of it 'sensitive data'?
 - iii. Is there data you think is missing? Why do you think this is? How could we collect that data?
 - iv. Do you think anyone would dislike this dataviz? Why or why not?